FOLDING, STABILITY AND AGGREGATION OF THE LONG-LIVED EYE LENS PROTEIN HUMAN GAMMA D CRYSTALLIN

by

Shannon L. Flaugh

B.S. Biochemistry
Colorado State University, 2001

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

June 2006

Signature of Author_____
Department of Biology
April 2006

Certified by_____
Jonathan King
Thesis Supervisor

Accepted by_____
Steven P. Bell
Chairman, Department Committee on Graduate Students

FOLDING AND STABILITY OF THE LONG-LIVED EYE LENS PROTEIN HUMAN GAMMA D CRYSTALLIN

by

Shannon L. Flaugh

Submitted to the department of Biology at the Massachusetts Institute of Technology on April 11, 2006 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biology

ABSTRACT

Human γD crystallin (HγD-Crys) is a monomeric, two domain, primarily β-sheet protein found in high concentrations in the human eye lens. HγD-Crys and other crystallins are found in insoluble protein inclusions associated with the eye disease cataract. HγD-Crys is expressed *in utero* and does not regenerate during life, thus necessitating high stability and solubility. Covalent damage, including glutamine deamidation, of the lens crystallins increases with age and as a result of exposure to environmental insults. Such covalent damage may cause partial-unfolding into aggregation-prone confomations that cause cataract.

The *in vitro* stability of HγD-Crys was analyzed in the denaturant guanidine hydrochloride at pH 7.0 and 37°C. An off-pathway aggregation reaction that competed with refolding was previously reported when HγD-Crys was refolded to less than 1 M GuHCl. Equilibrium transitions of HγD-Crys were best fit to a three-state model suggesting the presence of a partially-folded intermediate that likely had a structured C-terminal domain (C-td) and unstructured N-terminal domain (N-td). Similarly, previous analyses revealed a sequential domain refolding pathway where the C-td refolds first followed by the N-td. These findings suggest that the inter-domain interface of HγD-Crys is important in both folding and stability.

The domain interface of HγD-Crys contains a central hydrophobic cluster of six residues and two pairs of peripheral interacting residues. To test this importance of these residues in folding and stability, site-directed alanine mutants were constructed at all ten positions and properties of the mutant proteins were analyzed. Single mutations of hydrophobic domain interface residues caused a decrease in stability of the N-td, but did not affect stability of the C-td. Similarly, stability of the N-td but not the C-td was reduced as a result of single and double mutations of peripheral interface residues. Minimal to no interaction energy was observed for the peripheral residues suggesting they contribute to stability indirectly, perhaps by shielding the central hydrophobic cluster from solvent.

Both the hydrophobic and peripheral domain interface alanine mutants also had reduced rates of productive refolding for the N-td while refolding rates for the C-td were unchanged. These results suggest a productive folding pathway where the C-td refolds first and domain interface residues of the structured C-td act as a nucleating center for refolding of the N-td. Effects on N-td refolding rates were most prominent for the

hydrophobic residues indicating the importance of proper hydrophobic burial during refolding.

The peripheral domain interface residues of HγD-Crys include a pair of two glutamines that are targets for covalent damage during aging. Deamidation mimics at these sites were constructed by site directed mutagenesis of glutamine to glutamate. Properties of the mutants were analyzed to assess the affects of deamidation on stability and folding. Similar to the alanine mutants at these sites, the deamidation mutants had a destabilized N-td but not C-td at pH 7.0. In contrast, stabilities of the mutants were indistinguishable from wild type at pH 3.0. The N-td of the deamidation mutants also unfolded faster than that of wild type during kinetic unfolding. These results indicate that deamidation of domain interface glutamines destabilizes HγD-Crys and lowers the kinetic barrier to unfolding. A reduction in the thermodynamic and kinetic stability as a result of domain interface deamidation could result in the population of partially-unfolded conformations in the lens that may aggregate through mechanisms such as domain swapping or loop-sheet insertion.

Thesis Supervisor: Jonathan King, Professor of Biology

ACKNOWLEDGEMENTS

It is not surprising that this, the most important section of my thesis, has also been the most difficult to write. It is nearly impossible to communicate the gratitude I have for influences and actions that have not only made this work possible but have changed my life forever.

Professor Jonathan King. When I first met you in a graduate school interview in March of 2001 I knew immediately that I liked you. Your office was cluttered with books, journals and plants. You talked to me in a clear and frank manner and asked me questions about who I was outside of academics. Your attitude told me that you were interested in me as a person rather than a number and upon choosing you for my thesis advisor my initial impressions were confirmed. Early on I told you that I was in Biology because of a deep and long-standing love of animals. Your support and encouragement to stay connected with this motivation has truly moved me. In the end, the most important things you have taught me have nothing to do with protein folding and aggregation. Jon, thank you for taking the job of mentor so seriously and for guiding me through all of the rough spots. Your integrity and commitment will forever amaze and inspire me.

Dr. Melissa Kosinski-Collins. Dr. K-C. Octopussy. Of all the people I've had the pleasure of getting to know during my time at M.I.T., you have influenced me the most. You welcomed me onto the crystallin project with more warmth and enthusiasm than I could have ever hoped for. In the early days you taught me about equilibrium experiments with all of the talent of a seasoned expert and later, you taught me about the more important things in life. You taught me how to be a friend, how to listen, how to teach, how to discuss, and most importantly how to love, accomplishing this last great feat through your superior match-making skills! Melissa, thank you for everything. You know that I will always love you and miss those days we spent battling over the color of the tape ball!

Ishara Mills. My student host, turned roommate, turned lab-mate, turned great friend. As my host during a graduate school interview, you convinced me to come to M.I.T. not by extolling the virtues of the department or the science being conducted here, but by being genuine. Ishara, thank you for all of the fun times we've spent together! From digging your car out of a snow bank, to cleaning out the basement and doing kinetic experiments, you can turn even the most arduous task into a pleasurable experience. I look forward to a life-long friendship with you!

Thank you to all of "Team Crystallin", past and present! Veronica Zepeda, you set the bar for UROP students. Your hard work, insight and positive attitude has and will continue to amaze me for years to come. Jiejin Chen, thank you for bringing both a grand sense of humor and grand work ethic to the lab. Just remember, Mr. Ratburn says that good scientists must be nosy! Ligia Acosta, you bring great things to the world of α-crystallin. Good luck and happy chaperoning. Robin Nance and Cecilie Lin, you were both a delight to have around the lab. Thank you for bringing me back to reality with

your wonderful laughter. Abby Bushman, I know that you'll appreciate knowing Mr. Ratburn says scientists must also be patient. Katerina Papanikolopoulou, thank you for your insight, ideas and all of your hard work during your short stay. Thank you also to two of my classmates-turned-friends, Shamsah Ebrahim and Soraya Yekta. Your humor and wisdom have helped keep me sane and happy! To all members of the King Lab, thank you for making the lab a fun and interesting place to work. Dr. Peter Weigele, thank you for encouraging and inspiring me to follow my heart. Cammie Haase-Pettingell, the lab would not function without you! Thank you also to Ryan Simkovsky, Welkin Pope, Sean Decatur, Jacqueline Piret, Phillip Campbell, Kristen Cook, Claire Ting, David Gossard, and Cindy Woolley.

Thank you to my thesis committee, Professors Amy Keating, Bob Sauer, Jack Liang and David Housman. I know that your time is valuable and I appreciate all that you have done to help me through the final stage of my graduate career.

Thank you to my whole family. Bart and Kika, you make life worth living. Thank you for your patience, humor, love, and support. I can't wait to see our dreams become realities! Thank you to my sister Erin, my brother Keegan, and their families. I am so lucky to be a part of your lives and to be an Auntie to your wonderful children. Thank you to my dad, Don Flaugh for loving and accepting me. Grandma and Grandpa, I cannot express how much your love has meant to me over the years. I am happy that you are proud I will be a doctor, but don't forget that I am first and foremost your little ladybug. Thank you to Alan Ziff, so much more than my stepdad. You are and will always be a tremendous role model.

Finally, thanks to my mom, Dr. Lana Carter. My mom started her Ph.D. in psychology when I was 13 years old. She was a single parent supporting three teenagers on student loans and teaching assistantships. We were poor in money but rich in love. My sister and I spent countless hours playing "Mission Impossible" in the basement of the Clark Building at Colorado State University. I now know that this was when the seed was planted in my mind that I too would someday go to graduate school. Little did I know that getting a Ph.D. is not quite as much fun as sneaking around hallways with your sister! Mom, you have done so much for me. I would not have accomplished this if it weren't for your unwavering love and support. Thank you for being mother, my inspiration and most of all, my friend. I will wear your gown with pride!

This research was supported by NIH grant GM17980 awarded to Jonathan King and a Cleo and Paul Schimmel Fellowship awarded to Shannon L. Flaugh.

"You must be the change you wish to see in the world."

Mohandas Ghandi

"You've got to know when to hold 'em, know when to fold 'em, know when to walk away, know when to run."

Kenny Rogers "The Gambler"

BIOGRAPHICAL NOTE

Shannon L. Flaugh

Education

Ph.D. Expected 2006	Massachusetts Institute of Technology, Department of Biology, Cambridge, MA.
BS May 2001	Colorado State University, Major Biochemistry, Minor Chemistry, *Cum Laude*, Fort Collins, CO.

Research and Professional Experience

2001 to 2006	Graduate Research Assistant in the laboratory of Professor Jonathan King, MIT Department of Biology, Cambridge, MA.
April 2005	Wildlife Health Intern, New England Wildlife Center, Hingham, MA.
Summer 2001	Research Intern in the Department of Vaccine Research, Merck Pharmaceuticals, West Point, PA.
2000 to 2001	Undergraduate Research Assistant in the laboratory of Professor Kevin J. Lumb, Colorado State University Department of Biochemistry and Molecular Biology, Fort Collins, CO.
1998 to 2000	Undergraduate Research Technician, Colorado State University Department of Radiological Health Sciences, Fort Collins, CO.
1995 to 2000	Animal Care and Education Volunteer, Rocky Mountain Raptor Program, Fort Collins, CO.

Publications

Flaugh, S.L., Kosinski-Collins, M.S., & King, J.A. 2005. "Inter-domain side chain interactions in human γD crystallin influencing folding and stability." *Protein. Sci.* 14: 2030-2043.

Flaugh, S.L., Kosinski-Collins, M.S., & King, J.A. 2005. "The role of the hydrophobic interface residues in folding and stability of human γD crystallin." *Protein. Sci.* 14: 569-581.

Kosinski-Collins, M.S., Flaugh, S.L., & King, J.A. 2004. "Probing folding and fluorescence quenching in human γD crystallin Greek key domains using triple tryptophan mutant proteins." *Protein Sci.* 13: 2223-2235.

Flaugh S.L., & Lumb, K.J. 2001. "Effects of macromolecular crowding on the intrinsically disordered proteins c-Fos and p27(Kip1)" *Biomacromolecules* 2: 538-540.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE:

INTRODUCTION

A. THE PROTEIN FOLDING PROBLEM

Proteins are synthesized in cells on the ribosome through the step-wise polymerization of amino acids to construct long, linear polypeptide chains. These linear chains must then fold to form the unique three-dimensional structure determined by the protein's specific sequence of amino acids. The sequence control of protein folding was first described by Christian Anfinsen in 1961 (Haber and Anfinsen 1961; Anfinsen 1973). Many proteins fold spontaneously without aid, while other proteins require chaperones or chain modifications in order to fold. Ultimately, the unique shape, size and chemistry of the folded protein collectively confer function to the molecule.

The difficulty in understanding the process by which a polypeptide chain folds to adopt a specific and unique three-dimensional structure is the protein folding problem. Anfinsen originally proposed that the native state is the conformation with the lowest possible free energy (Anfinsen 1973; Anfinsen and Scheraga 1975). However, for some proteins this state is inaccessible due to a high kinetic barrier and thus they will fold into a conformation that represents a local free energy minimum (Baker et al. 1992). Many globular proteins are only marginally stable due to the multitude of weak attractive and repulsive forces that are present in the native state. Thermodynamic stability is determined by differences in the free energies of the native and unfolded states. Kinetic stability is conferred by a high kinetic barrier to unfolding. Several exceptionally stable enzymes from thermophilic bacteria display kinetic stabilization (Jaenicke and Bohm 1998). Many proteins that display kinetic stability are oligomeric β-sheet proteins with rigid structure (Manning and Colon 2004).

If protein folding occurred by a random search of all possible conformations accessible to a specific amino acid sequence, finding the free energy minimum would require millions of years. This phenomenon is known as Levinthal's paradox (Levinthal 1968; 1969). Instead, proteins generally fold in less than minutes both *in vivo* and *in vitro*. Rapid and efficient folding is enabled by protein folding pathways, where folding into the final native state occurs via a series of partially folded intermediates. The free energy landscapes of folding reactions have been described as funnels where an ensemble of unfolded states may transition into multiple different intermediates determined by the

kinetic and thermodynamic properties of the folding protein under the specific reaction conditions (Bryngelson et al. 1995). The partially folded intermediates are usually populated transiently and are under kinetic control. Lack of knowledge of the conformations of partially-folded intermediates is one of the barriers to solving the protein folding problem.

The utility of genomic sequence data from a variety of organisms and pathogens would be greatly increased if it were possible to predict the three-dimensional structures of all encoded proteins. At present, three-dimenstional protein structure predictions rely on homology modeling to proteins of known structure (Aloy et al. 2005). This method requires a high degree of sequence and structural similarity with the protein of known structure. In contrast, true de novo structure prediction based solely on amino acid sequence remains elusive. Deciphering the rules that determine the *pathway* by which a protein folds may eventually allow for this true de novo structure prediction.

Furthering understanding of protein folding mechanisms will also advance knowledge of a class of diseases that are caused by protein misfolding and aggregation (Horwich 2002). A common feature of experimentally characterized protein aggregation mechanisms is that the aggregation-prone precursor conformations are partially folded states. These states exist either on the productive folding pathway or are off-pathway but still accessible by the folding polypeptide chain (Haase-Pettingell and King 1988; Mitraki 1989; Horwich 2002). Understanding protein folding pathways will therefore inform investigations of protein aggregation associated with disease.

Historically, protein folding was studied using model proteins that were generally small, single domain, and primarily α-helical. In order to extract descriptive thermodynamic parameters from the unfolding and refolding transitions, off-pathway reactions such as aggregation were suppressed, even if it meant studying folding under conditions far from physiological. Deciphering the rules governing the folding of larger, more complex proteins has proven to be a difficult task.

In this thesis I describe experiments that explore the folding, stability and aggregation of human γD crystallin (HγD-Crys), a small, two-domain, primarily β-sheet protein located in the human eye lens and present in insoluble protein inclusions associated with mature-onset cataract. Below I review background information on

protein folding and unfolding, mechanisms of aggregation, and the lens crystallins with the intention of setting a framework for the studies performed herein.

1. *β-sheet protein folding*

Understanding the process by which β-sheet proteins fold is complicated both conceptually and experimentally. In contrast, consider the process of α-helix folding, where residues that are nearby in space are also nearby in sequence. Both theoretical and experimental results support general mechanisms of α-helix folding that involve a local helix nucleation event followed by growth or propagation. Helix nucleation would require formation of initial contacts that are entropically unfavorable due to constraints of the peptide backbone, but still have a reasonably high probability of occurring given their sequence proximity (Laurents and Baldwin 1998; Clarke et al. 1999). While the site of nucleation almost certainly differs depending on amino acid sequence, helix length and other factors, many if not all α-helices likely fold by such a mechanism.

Kinetic studies of α-helical peptides and small proteins have revealed folding on the millisecond time-scale (Williams et al. 1996; Gilmanshin et al. 1997; Clarke et al. 1999). Detection of partially folded intermediates in small, single-domain proteins has proven difficult presumably due to such extremely fast folding rates. The inability to detect intermediates has led to two-state descriptions of folding which invoke direct transition between the unfolded (U) and native (N) states (Jackson 1998). Partially-folded intermediates such as the α-helix nucleus are almost certainly polulated by these alleged "two-state folders", but detection and description of the intermediates will require more sensitive experimental measures.

In contrast to α-helix folding, β-sheet folding has the added complexity that while individual strands represent sequential residues, strands that hydrogen-bond in a sheet are typically distant in sequence. A universal pathway of β-sheet folding is more difficult to formulate due to the obligate formation of short-range contacts between remote sequence elements. Numerous small (<90 amino acids) β-sheet proteins have been described as undergoing a two–state folding mechanism (N⇔U). These proteins include Src homology 3 domain, cold shock protein B (Capaldi and Radford 1998) and

apoplastocyanin (Mizuguchi et al. 2003). Further studies of these proteins with more sensitive measures will likely reveal the presence of partially-folded intermediates. Other β-sheet proteins have been shown to populate partially folded intermediates in equilibrium experiments, kinetic experiments or both. Examples include the Fyn SH3 domain (Mittermaier et al. 2005), the DNA binding domain of the E2 transcriptional regulator from human papillomavirus (de Prat-Gay et al. 2005), cellular retinoic acid-binding protein I (CRABPI) (Clark et al. 1996), intenstinal fatty acid-binding protein (Burns et al. 1998), Cobrotoxin (Hsieh et al. 2006), and interleukin 1-β (IL1-β) (Varley et al. 1993; Finke and Jennings 2002). For many of these proteins, conformations of the partially-folded states remain elusive.

The conformation of partially-folded intermediates on the folding pathway of IL1-β have been studied in detail by hydrogen-deuterium exchange NMR, circular dichroism (CD) and fluorescence (Varley et al. 1993). An early folding intermediate was detected by CD that had 90% the β-sheet content of native IL1-β. In contrast, stable intermediates detectable by NMR were not observed until later times. Thus, the conformation of the chain was ill-defined prior to sheet formation detected by NMR. The intermediates detected by H-D exchange had native-like β-sheets in two regions that contributed non-polar residues to the hydrophobic core (Varley et al. 1993). The authors propose a folding pathway where the early intermediates detected by CD have non-native β-sheets that rearrange to form intermediates with native-like secondary and tertiary structure. This intermediate then undergoes final stabilization into the compact native state (Varley et al. 1993). Similarly, IL1-β unfolds through intermediates (Roy and Jennings 2003). However, instead of populating discrete stable intermediates, IL1-β populates a continuum of states during unfolding (Roy and Jennings 2003). First, tertiary structure is lost, followed by central β-sheet unfolding and finally, helix and β-turn unfolding (Roy and Jennings 2003).

Experimental and theoretical work has indicated the importance of β-turns and -hairpins in the folding and unfolding of many β-sheet proteins. These structures are often the first structures to form during folding and last to denature during unfolding (Katou et al. 2001; Walkenhorst et al. 2002; Rotondi and Gierasch 2003; Rotondi et al. 2003). In the case of CRABPI, peptides corresponding to only two of the seven turns are able to

form native-like structure in isolation (Rotondi and Gierasch 2003). These turns participate in a network of conserved long-range interactions important for folding and stability of CRABPI (Rotondi et al. 2003). Local sequences contributing to β-turns or –hairpins may reduce the conformational freedom of intermediates during folding thus accelerating rates of sheet formation.

2. *Multidomain protein folding*

Domains are discrete units of a protein that exhibit independence in one of a number of characteristics, including structure, function, folding or genetic inheritance (Jaenicke 1999). At least some multidomain proteins are believed to have evolved through gene fusion events that resulted in covalently linked single domain proteins. This phenomenon can generate countless new proteins without requiring evolution of new independent folds (Jaenicke 1999).

The individual domains of some multidomain proteins are often capable of independent folding and unfolding (Wetlaufer and Ristow 1973; Corbett et al. 1986; Beechem et al. 1995). During a folding reaction *in vivo* off the ribosome, or *in vitro* out of denaturant, domains may fold synchronously or sequentially. Synchronous domain folding reduces the lifetime of partially-structured folding intermediates, while sequential domain folding results in the transient population of partially-folded conformations (Jaenicke 1999). These partially-folded species are susceptible to proteolytic degradation and may be prone to aggregation through mechanisms such as domain swapping, described in detail below (Liu and Eisenberg 2002).

Independent domain unfolding has been observed for numerous proteins (Jaenicke 1999). This phenomenon has most often been observed as the presence of multiple calorimetric transitions during the thermal denaturation of multidomain proteins (Novokhatny et al. 1992; Kurochkin et al. 1995). For instance, thermal unfolding of the plasma transglutaminase, factor III, proceeds via five independent transitions reflecting independent unfolding of the five domains (Kurochkin et al. 1995). Similarly, thermal unfolding of the heparin binding domain of fibronectin revealed independent unfolding of the three domains (Novokhatny et al. 1992). Analysis of the isolated domains allowed

for assignments of the transitions. The melting temperature of one of the domains was lower in isolation than in the full-length protein, indicative of stabilization by domain-domain interactions (Novokhatny et al. 1992).

Synergy in domain stability has also been observed for a number of other multi-domain interactions, such as recombinant antibody fragments (Rothlisberger et al. 2005). Mutual thermodynamic and kinetic stabilization was observed for interaction of variable heavy (V_H) and light chains (V_L) as well as interaction of constant heavy (C_H) and light chains (C_L). However, no stabilization was observed for interactions of V_L and C_L chains (Rothlisberger et al. 2005). Differences in the domain interfaces were likely responsible for this disparity. First, the interface of the V_L and C_L chains is much less extensive than the V_H-V_L and C_H-C_L interfaces; second, an intermolecular disulfide bond exists between the C_H-C_L domains (Rothlisberger et al. 2005).

Independent kinetic domain folding has also been described for multidomain proteins such as phosphoglycerate kinase and PapD (Beechem et al. 1995; Sherman et al. 1995; Bann et al. 2002; Bann and Frieden 2004). The periplasmic pili subunit chaperone PapD is a two-domain, primarily β-sheet protein that assists folding of its substrate through transient donation of one of its β-stands (Choudhury et al. 1999; Sauer et al. 1999; Barnhart et al. 2000). Kinetic folding of 6-fluorotryotophan-labeled PapD was probed by ^{19}F-NMR (Bann et al. 2002). A partially-folded intermediate formed early, followed by folding into the final native state. A folding model was proposed where the C-terminal domain (C-td) rapidly folded into an intermediate, and subsequently the N-terminal domain (N-td) folded simultaneous with the final readjustment of the C-td (Bann et al. 2002). Subsequent analysis of the PapD kinetic refolding mechanism expanded the description to include the formation of domain-domain interactions (Bann and Frieden 2004). According to this analysis, the C-td and N-td both folded prior to the formation of domain interface interactions (Bann and Frieden 2004).

A designed multidomain protein also displayed independent kinetic domain folding (Zhou et al. 2005). Zhou et al. (2005) constructed a fusion of two small α-helical proteins and analyzed its structure and folding. The designed protein folded into a two-domain structure. With the exception of one α-helix, structures of the domains were consistent with the structures of the parent proteins. Stability of the designed protein was

similar to the parent proteins indicating there was minimal energetic synergy of the domains. However, the designed protein folded faster than one of the parent proteins. This observation suggests that domain interactions in the fusion protein stabilized the transition state during kinetic refolding (Zhou et al. 2005).

The examples provided above demonstrate that domains are capable of independent unfolding/refolding and that the role of domain-domain interactions in folding and stability depend on individual characteristics of the domains and the domain interfaces. To determine common characteristics of domain interfaces, Jones et al. (2000) surveyed domain interactions in both multidomain and oligomeric proteins of the Protein Data Bank (PDB). The amino acid composition of intra-chain domain interactions more closely resembles that of protein surfaces rather than protein cores (Jones et al. 2000). Despite this, hydrophobic residues are still highly prevalent in both inter- and intra-chain domain interactions (Jones et al. 2000).

3. *Protein unfolding pathways*

Just as many proteins fold via multi-step pathways, the unfolding of proteins also occurs by specific pathways. Complex *in vitro* unfolding transitions have been observed for many multimeric and multidomain proteins (Chen and Smith 2000; Simmons et al. 2004; Slaughter et al. 2005). For example, single-molecule unfolding of the two domain protein calmodulin identified a single major intermediate that may have one domain folded and one unfolded (Slaughter et al. 2005). Similarly, sequential domain unfolding has been observed for the three domain protein, human serum albumin (Santra et al. 2005). However, these investigations did not address the order in which structural elements of the domains unfolded. Studies of the coiled-coil dimeric protein GCN4 have identified unfolding intermediates at a smaller scale (Dragan and Privalov 2002). First, the N-termini of the helices unfolded or frayed. Second, remainders of the helices unfolded but did not dissociate, followed finally by chain dissociation (Dragan and Privalov 2002). As described above, detailed analysis of IL1-β unfolding has identified a rugged unfolding free-energy landscape controlled by β-turns (Roy and Jennings 2003).

The unfolding experiments described above utilized chemical denaturants or temperature to induce unfolding. Recently, force-induced or mechanical unfolding has been investigated by atomic force microscopy (AFM). Two distinct unfolding pathways were observed in one such study of the β-sandwich type III domain of fibronectin, (Li et al. 2005). The two pathways differed in the order that β-strands were unfolded prior to complete unfolding (Li et al. 2005). The presence of parallel unfolding pathways has also been observed for O^6-methylguanine-DNA methyltransferase during *in vitro* chemical-induced denaturation (Nishikori et al. 2005). Parallel *folding* channels have been described for the α-subunit of tryptophan synthetase (Bilsel et al. 1999).

In vivo, protein unfolding is important in degradation, aggregation and transport (Prakash and Matouschek 2004). As described in detail below, aberrant *in vivo* unfolding may result in population of partially unfolded states that may be susceptible to aggregation. For instance, partial unfolding is sufficient for amyloid formation by transthyretin, a protein implicated in the disease Senile Systemic Amyloidosis (Colon and Kelly 1992). I*n vivo* protein unfolding is generally mediated by unfoldase enzymes that catalyze the reaction by pulling on substrate proteins thus causing local unfolding events (Prakash and Matouschek 2004). Local stabilities of the substrate proteins are significant in determining these catalytic unfolding pathways. Therefore, *in vitro* unfolding pathways may have limited relevance to *in vivo* unfolding.

4. *Protein aggregation and disease*

Many human diseases are associated with the accumulation of insoluble protein deposits or inclusions. Examples of such diseases and their associated proteins include Aβ in Alzheimer's disease (AD), α-synuclein in Parkinson's disease, the prion protein in Creutzfeld-Jacob disease (CJD), and the crystallin proteins in cataract (Horwich 2002). In the case of neurodegenerative diseases such as AD, Parkinson's and CJD, it is currently unknown if protein aggregation causes neuronal damage or acts as a protective mechanism (Caughey and Lansbury 2003). It has been suggested that soluble protofibrils or other non-native forms of the proteins may cause neurodegeneration, and that aggregation into the large mature protein deposits functions as a protective mechanism to

remove the damaging species (Caughey and Lansbury 2003). However, for other diseases such as sickle-cell anemia, it is the mature protein polymers and not smaller precursors that cause the disease phenotype.

Protein aggregation often occurs through interaction of partially folded proteins (Horwich 2002). An exception to this is the polymerization of hemoglobin in sickle cell anemia that is described in detail below. For proteins that do aggregate from a partially folded state, in order for aggregation and disease to occur the problematic conformation must first be populated. Experimental evidence indicates that interactions between protein molecules in aggregates are specific and that they occur through precise mechanisms of association (Speed et al. 1996). Outlined below are several general mechanisms of protein aggregation that have been well-studied.

a. *Amyloidosis*

Amyloid deposits are insoluble protein inclusions found in a variety of diseased tissues. Many human diseases, such as AD, transmissible spongiform encephalopathy (TSE), senile systemic amyloidosis, and type 2 diabetes are associated with the presence of amyloid plaques formed from unique proteins (Sipe 1994). The proteins associated with these diseases have dissimilar primary, secondary and tertiary native structures, suggesting that the amyloid state is accessible by a multitude of proteins (Horwich 2002). In fact, many proteins not associated with amyloid diseases have been shown to form amyloid fibers under mild denaturing conditions (Chiti et al. 2000; Fandrich et al. 2001). It has also been suggested that amyloid fibers are a generic structure of the polypeptide backbone that all proteins may adopt given the right conditions (Dobson 1999).

Amyloid deposits are composed of long, unbranched fibers that bind the dye Congo Red. Further structural analyses revealed that amyloid fibers have a cross-β structure with a characteristic X-ray diffraction pattern due to the regular arrangement of β-strands perpendicular to the fiber axis (Eanes and Glenner 1968; Bonar et al. 1969). Recent structural analyses of two amyloidgenic peptides have furthered the structural

analysis to atomic resolution (Luhrs et al. 2005; Nelson et al. 2005). The peptide Aβ(1-42), associated with AD, was studied by quenched H-D exchange NMR (Luhrs et al. 2005). The peptide forms a strand-turn-strand structure with 17 unstructured residues at the N-terminus. These structures interact to form elongated, twisting parallel β-sheets that resemble amyloid fibers observed by transmission electron microscopy (Luhrs et al. 2005). Similarly, the structure of a peptide fragment of the yeast prion Sup35 was studied by X-ray crystallography (Nelson et al. 2005). The peptide formed both amyloid fibers and closely related microcrystals, which were used to determine an atomic-resolution crystal structure. The peptides of the microcrystals adopt extended conformations that interacted to form long parallel sheets which laterally associated into a double β-sheet (Nelson et al. 2005). Amino acid side chains of the peptide were regularly stacked and in register, a characteristic also observed for parallel β-helix proteins (Jenkins and Pickersgill 2001) and critical for folding of the β-helical protein, P22 tailspike (Simkovsky and King 2006). In fact, the parallel β-helix is also a proposed model of a structure that meets the cross-β amyloid criteria (Wetzel 2002).

The *in vitro* aggregation of amyloidgenic proteins follows nucleation-growth kinetics (Harper and Lansbury 1997). Accordingly, monomer dominates below a critical protein concentration, and at higher protein concentrations the concentration of monomer stays the same while that of the polymer increases. This is indicative of cooperative nucleus formation (Harper and Lansbury 1997). A simple schematic describing a possible mechanism to explain nucleation-growth kinetics is shown in Figure 1-1. Although the aggregation kinetics of many amlyoidgenic proteins have been well-described, how the proteins transition into the amyloidgenic state and the process of nucleus formation is still being elucidated. Understanding these critical steps in fiber formation may lead to the development of novel therapeutics to combat these debilitating diseases. Described below are two specific cases where substantial progress is being made to understand these processes.

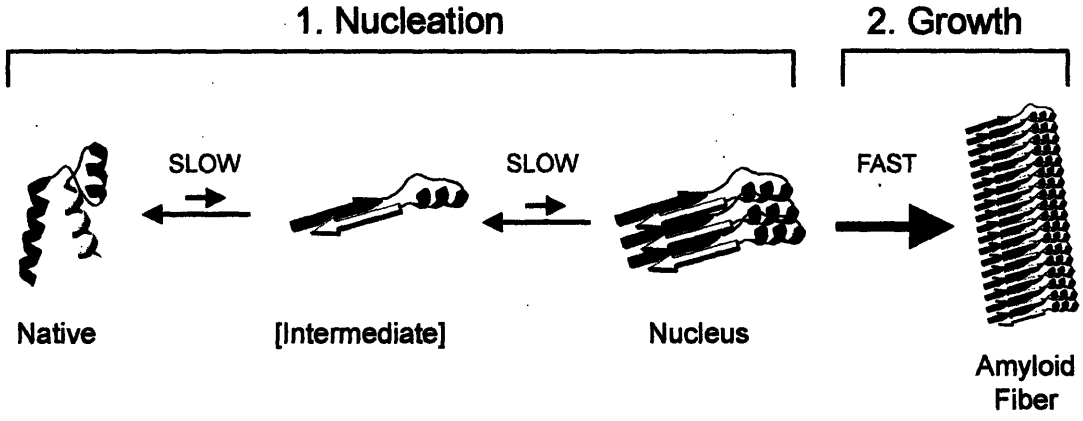

Figure 1-1. Schematic model of a simple mechanism following nucleation-growth aggregation kinetics. Structural rearrangements into the amyloidgenic intermediate and nucleus formation correspond to slow events, whereas fiber growth is fast.

The prion protein, which is associated with TSE, bovine spongiform encephalopathy, CJD, and sheep Scrapie, exists in both a soluble cellular (PrP^C) and aggregation-prone disease form (PrP^{Sc}). Transition from PrP^C to PrP^{Sc} involves structural transformations from a largely α-helical to β-sheet conformation (Nguyen et al. 1995; Zhang et al. 1995). This transition is limited by a high kinetic barrier that can be overcome by familial mutations, or seeding, which causes the disease to be transmissible (Prusiner 1998). A monomeric intermediate on the *in vitro* kinetic refolding pathway of PrP^C has been identified (Apetri and Surewicz 2002). This compact intermediate is stabilized at acidic pH, where transition into a PrP^{Sc}-like form is enhanced, suggesting that the productive folding intermediate may correspond to the amyloidgenic state of PrP (Apetri and Surewicz 2002). This intermediate state is stabilized by familial disease-causing mutations further supporting its role in amyloid formation (Apetri et al. 2004). However, the conformation of this intermediate has yet to be described and its *in vivo* role in amyloid formation is unclear.

The mechanism of amyloid fiber formation in AD has also been studied in detail. The amyloidgenic peptides, Aβ(1-42) and Aβ(1-40) are proteolytic cleavage products of a larger transmembrane protein, the amyloid precursor protein (APP). There are two proteolytic processing pathways of APP, the nonamyloidgenic pathway where benign cleavage products are produced, and the amyloidgenic pathway which produces one of two amyloidgenic peptides, Aβ(1-42) or Aβ(1-40) (Buchet and Pikula 2000). On the amyloidgenic pathway, APP is first cleaved by β-sectretase, which targets the extracellular domain, and then by γ-secretase, which cuts one of two sites in the transmembrane domain. While Aβ(1-40) is produced at higher levels than Aβ(1-42), the amyloid plaques of AD patients contain primarily Aβ(1-42) and *in vitro*, the longer peptide forms fibers more rapidly (Selkoe 1991; Jarrett and Lansbury 1993; Harper and Lansbury 1997). The identity of the additional two amino acids in Aβ(1-42), and not just the extra length, are responsible for these differences. Kim and Hecht (2005) created a library of Aβ(1-42) peptides with random amino acids present at the last two sites and analyzed their aggregation potential. Peptides with residues that were hydrophobic or had high β-sheet propensity aggregated more readily than those with hydrophilic residues

or β-sheet breakers (Kim and Hecht 2005). This indicates that aggregation into an amyloid state occurs by a precise process of specific interactions.

b. *Domain swapping*

Domain swapping is a general mechanism of protein oligomerization defined by two or more proteins exchanging identical structural elements to form dimers or high order oligomers (Liu and Eisenberg 2002). The exchanged structural elements of domain swapped proteins range in complexity from single β-strands or α-helices, both of which are observed in different domain swapped dimers of RNase A, to large domains that constitute half of the protein, as is the case for βB2-crystallin (Liu and Eisenberg 2002). Distinct domain-swapped oligomers exist in closed conformations where the structure of each monomer is satisfied by interaction with another subunit such that there are no unpaired domains (Fig. 1-2).

Domain swapping may be a general method of protein aggregation and amyloid fiber formation (Jaskolski 2001). For example, domain swapping may occur where an open conformation is produced that has unsatisfied domains on the oligomer termini. These open subunits would act as sites of monomer addition, resulting in long aggregated species (Fig. 1-2). The human prion protein and cystatin C, both of which form amyloid fibers *in vivo*, have been shown to dimerize by domain swapping (Janowski et al. 2001; Staniforth et al. 2001). Four human cystatin C dimers pack together in the crystal structure into an octamer species that has significant β-sheet content (Janowski et al. 2001). Linkers of the domain swapped dimers act as sites of octamer interaction, thus interconnecting an unlimited network of molecules. This structure may be related to the amyloidgenic structure of human cystatin C (Janowski et al. 2001). The domain swapping model of amyloid formation is not inconsistent with the amyloid structures already known, as these structures are of short peptide fragments, not the whole disease-related proteins. Domain swapping may help explain the conformation of protein segments that do not directly contribute to the cross-β amyloid structure (Sambashivan et al. 2005).

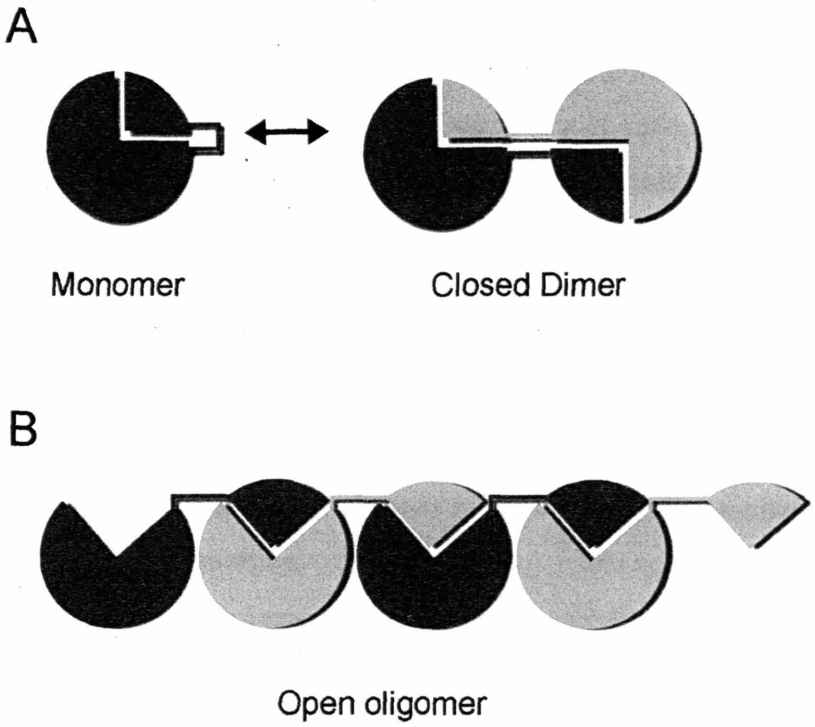

Figure 1-2. A) Schematic model of dimerization by domain swapping where two monomers exchange identical subunits to form a Closed dimer. B) Unsatisfied domains at domain-swapper oligomer termini may result in Open ends that will act as sites for monomer addtion.

The kinetic mechanism by which domain swapping occurs has been studied in-depth for only a few proteins. For example, a domain-swapped dimeric mutant of the immunoglobulin binding domain of streptococcal protein G transitions between the dimeric and monomeric states via a partially folded intermediate (Byeon et al. 2003; 2004). This intermediate has molten globule characteristics and resembles the wild-type monomer rather than the monomer conformation in the dimer. In contrast, for the cell cycle regulatory protein p13suc1, transition from the monomer to domain-swapped dimer requires complete unfolding (Rousseau et al. 2001). The dimer and monomer of p13suc1 have two separate folding pathways and the β-strand that exchanges in the dimer is a critical part of the folding nucleus, suggesting that association is an early event in dimer folding (Schymkowitz et al. 2000). Rousseau et al. (2001) also observed correlation between domain swapping and mild thermal aggregation of p13suc1, signifying a common mechanism. Discrepancies between these studies indicate that there is not a universal kinetic mechanism of domain swapping, and instead the process depends on the properties of the proteins under specific reaction conditions.

c. *Loop-sheet insertion*

Loop-sheet insertion is a mechanism of protein polymerization and aggregation utilized by the serpins (serine protease inhibitors). The serpins are a superfamily of metastable proteins found in diverse organisms (Silverman et al. 2001). These α/β proteins adopt a highly conserved fold that includes a reactive loop of ~17 residues that is important in both protease inhibition and serpin polymerization (Fig. 1-3). The serpins utilize a suicide mechanism to inhibit their substrate proteases. First, the serpin binds to its substrate protease, at which point the protease cleaves the serpin backbone in the reactive loop leading to a covalent ester linkage between the serpin and protease (Silverman et al. 2001). The reactive loop attached to the protease then inserts into a native β-sheet of the serpin thereby translocating the protease ~70 Å and altering the order of β-strands in the sheet. The protease is distorted and inactivated in this

conformation due to constraints imparted by the reactive loop (Silverman et al. 2001). In a parallel reaction, if loop-sheet insertion does not occur before the covalent attachment between the serpin and protease is broken, the serpin is transformed into an inactive form and the active protease is released. In the inactive form, the reactive loop is inserted into the β-sheet but not attached to the protease (Silverman et al. 2001). The conformation of α1-antitrypsin (α1-AT) in the active (Elliott et al. 2000) and inactive states (Engh et al. 1989) as well as covalently bond to a substrate trypsin (Huntington et al. 2000) are shown in Figure 1-3. The active form of the serpins is a metastable state with a melting temperature (T_m) of ~60°C while the inactive state is very stable with a T_m of greater than 100°C (Carrell and Gooptu 1998).

Intermolecular loop-sheet insertion and polymerization has also been observed for a number of serpins due to naturally occurring mutations (Fig. 1-4). This causes disease by either loss of serpin function or pathological effects of the aggregates. For instance, mutations in α1-AT lead to formation of aggregates that cause emphysema because α1-AT function is diminished, and alternatively cirrhosis because of mal-effects of the intracellular inclusions (Sifers 1995). The folding rate of the mutant α1-AT was severely reduced compared to wild type but stability of the mutant was unchanged (Yu et al. 1995). From this observation it was postulated that a folding intermediate was the aggregation prone state (Yu et al. 1995). The intermediate may be common to both the productive folding and aggregation pathways.

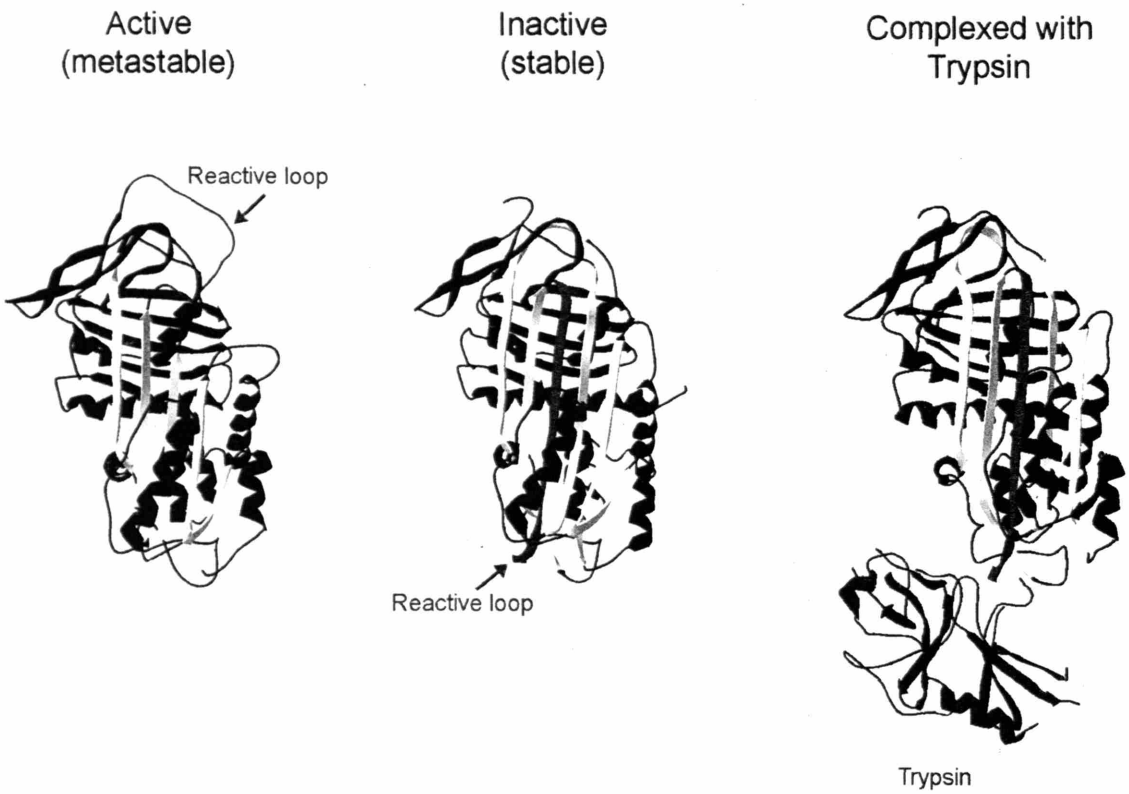

Figure 1-3. Ribbon structures of α1-AT: 1) in the active state with a mobile reactive loop (PDB 1QLP), 2) in the inactive state with a cleaved reactive loop inserted in the β-sheet (PDB 7API) and 3) in complex with trypsin (PDB 1EZX).

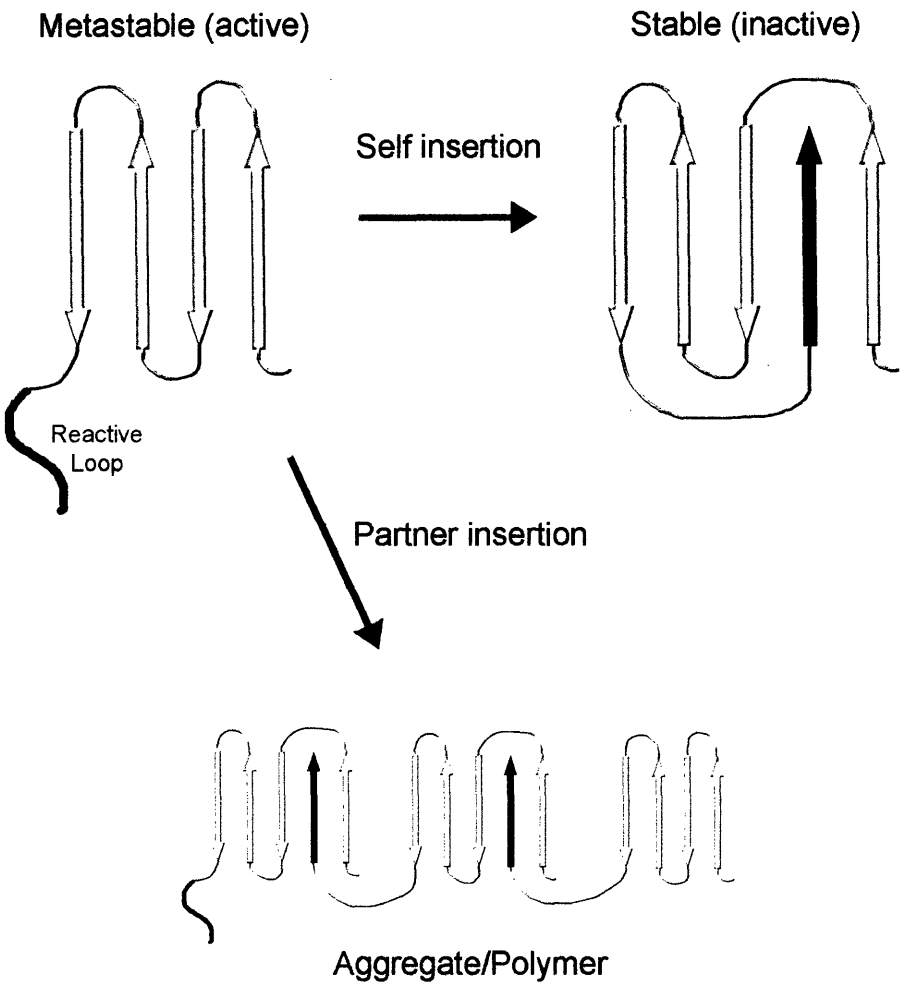

Figure 1-4. Schematic model of loop-sheet insertion. The Reactive Loop of one monomer may either insert as a β-strand into its own β-sheet or insert into the β-sheet of another molecule thereby causing polymerization.

d. *Native state polymerization*

In contrast to the aggregation mechanisms described above, the aggregation of mutant hemoglobin in sickle cell anemia occurs through interaction of native protein molecules. Sickle cell anemia is disease of erythrocytes that is caused by a single amino acid substitution, E6V, of hemoglobin (Ingram 1956). The Glu6 side chain is located on the surface of the wild-type hemoglobin β-subunits. Substitution with the hydrophobic valine side chain causes the native hemoglobin molecules to polymerize by interaction with a surface hydrophobic pocket made of Phe85 and Leu88 on the β-subunit of another molecule (Bihoreau et al. 1992). The mutant side chain of the second protein is then able to interact with the hydrophobic pocket of a third protein and these interactions proceed ad nauseam to form long polymers of native hemoglobin that stretch the erythrocyte membranes into a sickle shape.

B. THE HUMAN EYE

Cataract is a major disease of the human eye lens associated with the presence of insoluble protein inclusions that scatter light (Hoenders and Bloemendal 1983). The human eye lens is a complex and highly specialized tissue and is a principal component of the eye. In simple terms, the biological function of the eye is to sense light in order to provide information about an organism's environment. According to this definition, the light-sensitive pigments of unicellular organisms may reflect the ancient ancestors of the complex modern eyes (Oyster 1999). In multicellular organisms the eye is composed of a group of cells that can detect light by way of similar light-sensitive pigments.

1. *Optic systems and vision*

In order for light information to be converted into visual information, the eye must have an array of photosensitive cells and an optic system that allows for discrimination of light direction (Oyster 1999). Animals have evolved many types of eyes, the three general classes being: 1) simple eyes with a single optic system, 2) compound eyes with multiple optic systems that act independently and 3) optical superposition eyes with multiple cooperating optical systems (Oyster 1999). The more complex optic systems are further classified by whether reflection or refraction is utilized for focusing light onto the photosensitive cells.

The human eye is a simple eye that utilizes a refracting optic system (Fig. 1-5). The approximate 100 million photoreceptor cells of the human eye are found in a single layer in the posterior of the eye (Oyster 1999). Underlying the single layer of photoreceptor cells are several layers of neuronal cells that convey visual information to the brain via the optic nerve. Collectively, the photoreceptor and nerve cells make up the retina. The major volume of the eye corresponds to the vitreous humor, which fills the large aqueous vitreous cavity. The vitreous humor is composed of a highly hydrated collagen and proteoglycan matrix with a gel-like consistency (Oyster 1999). The vitreous has little refractive abilities but must still remain transparent to visible light. Most light refraction in the human eye occurs in the cornea. Transparency of the cornea is established by spatially ordered collagen fibrils. The opaque sclera which surrounds the cornea also has high levels of collagen, but the fibrils are not ordered in comparison. The eye lens is positioned posterior to the cornea and is also transparent to visible light. While the cornea is the principal refractive tissue in the eye, the lens is responsible for fine-tune focusing of light on the retina and performs about one third of the refraction in the eye (Oyster 1999).

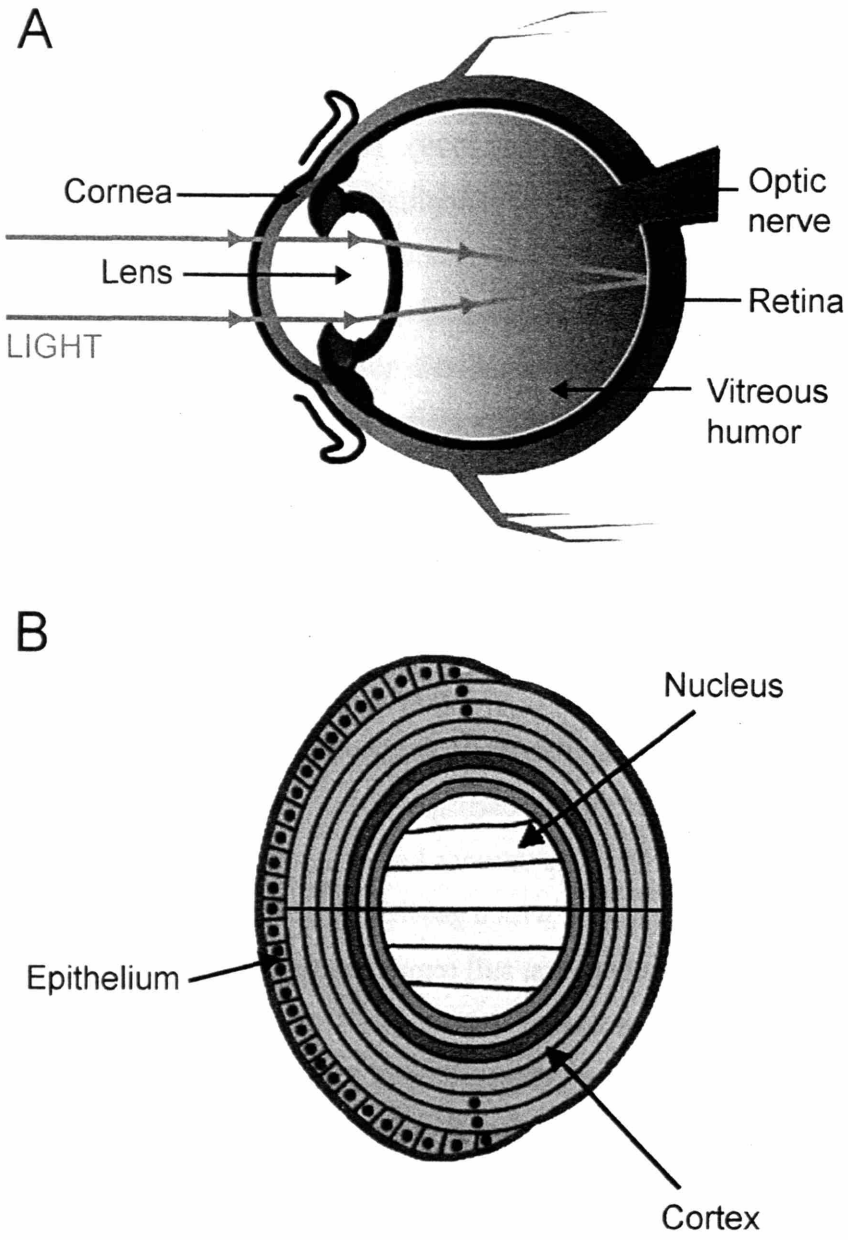

Figure 1-5. A) Schematic drawing of the human eye in cross-section. The path that light takes to reach the retina is shown and the major parts of the eye are labelled. B) Schematic drawing of the human eye lens in cross-section. The lens nucleus, cortex and epithelium are labelled.

2. *The human eye lens*

The human lens has an overall spherical shape and is composed of terminally differentiated, elongated, and enucleated fiber cells (Fig. 1-5). There are three cell types in the lens classified according to age, location and morphology: 1) nuclear cells are in the center of the lens, 2) cortical cells surround the nucleus, and 3) a single layer of epithelial cells covers the anterior portion of the lens. The nuclear and cortical cells are elongated and enucleated while the epithelial cells are nucleated and undifferentiated. The nuclear cells and major mass of cortical cells differentiate *in utero* and during infancy and new layers of cortical cells are added throughout life via differentiation of cells from the epithelial layer. All proteins in the nuclear and cortical cells were synthesized prior to differentiation, and since there is essentially no protein turnover in these cells, most cells of the adult eye contain proteins synthesized *in utero* or during infancy (Oyster 1999).

Crystallins are the major protein component of the lens and are present in fiber cells at concentrations of 200-400 mg/ml. Short-range ordering of crystallins renders the lens transparent to visible light and gives it a high refractive index necessary for focusing light on the photoreceptor cells of the retina (Delaye and Tardieu 1983; Fernald and Wright 1983). In order to maintain lens transparency, the crystallins must remain stable and soluble in the continued presence of oxidative stress, and without the possibility of regeneration for a lifetime. Cataract is an opacification of the lens to visible light.

3. *Cataract*

Cataract is a major global health problem as it is the leading cause of blindness in the world and costs the United States government over $3.4 billion per year by way of surgery (National Eye Insitute (U.S.) 2002). Pathologically, cataract is associated with the presence of insoluble inclusions of crystallin proteins. The prevalence of cataract is similar among both males and females and several different ethnic groups. In the United States, one in six people over age 40 and one in two over age 80 have cataract (National Eye Insitute (U.S.) 2002). A striking feature of cataract is the extreme rise in prevalence

with increasing age. The current method of treatment for cataract is surgery to remove the diseased lens and replacement with a plastic intraocular lens. The side-effects of cataract surgery are minimal, the operation is generally performed as an outpatient procedure, and the cost to the patient is low. These factors make surgery a viable choice for individuals with access to modern health care. However, surgery is generally not an option for afflicted individuals from developing nations, thus making it vitally important to understand the molecular basis of cataract formation.

C. EYE LENS CRYSTALLIN PROTEINS

The classes of crystallins found ubiquitously in vertebrate lenses are the α-, β-, and γ-crystallins. Collectively, theses proteins account for approximately 80-90% of total protein in the lens and are present in concentrations of 200-400 mg/ml (Oyster 1999). In addition to the three ubiquitous crystallins, there are several taxon-restricted crystallins found in a smaller subset of vertebrates that are related to "housekeeping" enzymes (Slingsby 1997). Given the very high concentrations of crystallins in the fibrous lens cells, the presence of multiple proteins that adopt different structures and oligomeric states is vital for preventing protein crystallization.

1. *α-crystallin*

Two unique α-crystallin proteins are present in the lens, αA and αB, which display approximately 60% sequence identity (Bloemendal and de Jong 1991). αA-crystallin is limited to the lens, while αB-crystallin has been found in other tissues including the brain, heart and muscle (Iwaki et al. 1990). In the lens, αA- and αB-crystallin interact to form large polydisperse complexes that range in molecular mass from 300 to 1200 kDa, corresponding to 15 to 55 monomers per complex. Due to the polydisperse nature of the lens α-crystallins, thus far it has not been possible to determine an X-ray crystal structure of the human α-crystallins. However, X-ray structures of small heat shock proteins (sHSPs) have been solved that contain "α-crystallin domains" linked

to variable N-terminal regions (Kim et al. 1998; van Montfort et al. 2001; Stamler et al. 2005). The "α-crystallin domain" fold has two β-sheets arranged into a sandwich that are connected by long loops containing both α-helical and unstructured regions. In the case of wheat HSP 16.9, the monomers associate to form dimers by strand exchange and these dimers further associate into a dodecameric structure (van Montfort et al. 2001).

While the crystal structure of the human α-crystallins remains elusive, some insight into the structure of the multimeric species has been gained using cryo-electron microscopy. These analyses have led to a micellar-like structural model where the α-crystallin subunits interact to form a hollow sphere (Haley et al. 1998; Haley et al. 2000). The shell of the complex has an overall diameter of ~19 nm while the hollow internal cavity has a diameter of ~8 nm.

In addition to a structural role in the lens, the oligomeric α-crystallins possess molecular chaperone activity *in vitro,* which has been observed in interaction with several other crystallins as well as non-lens proteins (Clark and Muchowski 2000; Horwitz 2000; MacRae 2000). The α-crystallins have also been shown to be phosphorylated in lenses, which likely influences their activity as chaperones (Wang et al. 2000; Ueda et al. 2002; Sathish et al. 2004). Chaperone activity of the α-crystallins may be significant in preventing or delaying cataract by binding and sequestering other partially unfolded or damaged crystallin proteins and thus preventing their aggregation.

Several α-crystallin knock-out mice have been generated and their lens phenotypes analyzed to assess the role of α-crystallin in maintaining lens transparency. The αA-crystallin knock-out mouse had premature cataract caused by the presence of insoluble inclusions of αB-crystallin and the γ-crystallins (Brady et al. 1997; Horwitz 2003). The αB-crystallin knock-out mouse had a shortened lifespan thus preventing analysis of effects on mature-onset cataract formation (Brady et al. 2001). The double knock-out mouse of αA/αB has smaller than normal lenses with altered cellular structure suggesting that the α-crystallins may also contribute to lens development (Boyle et al. 2003).

2. *β- and γ-crystallins*

In contrast to the α-crystallins, the β- and γ-crystallins do not have any detectable enzymatic or chaperone activity and are thought to function solely as structural proteins in the lens. The β- and γ-crystallins are primarily β-sheet, two-domain proteins that are evolutionarily related (Wistow and Piatigorsky 1988; Wistow et al. 2005). The domains of the β- and γ-crystallins each have eight β-strands (A-H) that are arranged into two Greek key motifs (Fig. 1-6). The Greek key motifs further associate to form a β-sandwich structure characterized by β-strands wrapped around a central hydrophobic core (Fig. 1-6). This double Greek key β-sandwich fold defines the βγ-crystallin structural superfamily. The Greek keys and domains of the β- and γ-crystallins are hypothesized to have originated from consecutive gene duplication events (Wistow and Piatigorsky 1988).

Despite analogous domain folds, structural differences do exist between the β- and γ-crystallins. The γ-crystallins are strictly monomeric while the β-crystallins form a range of multimeric states ranging from dimers to octamers (Wistow et al. 1983; Bax et al. 1990; Slingsby and Bateman 1990). As shown in Figure 1-6, the domains of the γ-crystallins pair intramolecularly through amino acid side chain interactions across a domain interface, giving the proteins overall pseudo-twofold symmetry (Wistow et al. 1983). The domain interface interactions are characterized by non-covalent contacts between a central hydrophobic cluster and peripheral polar amino acid side chains. As shown in Figure 1-7, the major bovine β-crystallin, βB2-Crys, forms a dimer by domain swapping where the N-td of one monomer packs against the C-td of the other monomer (Bax et al. 1990). The domains of the domain-swapped dimer interact in a manner identical to that described for the domains of the γ-crystallins.

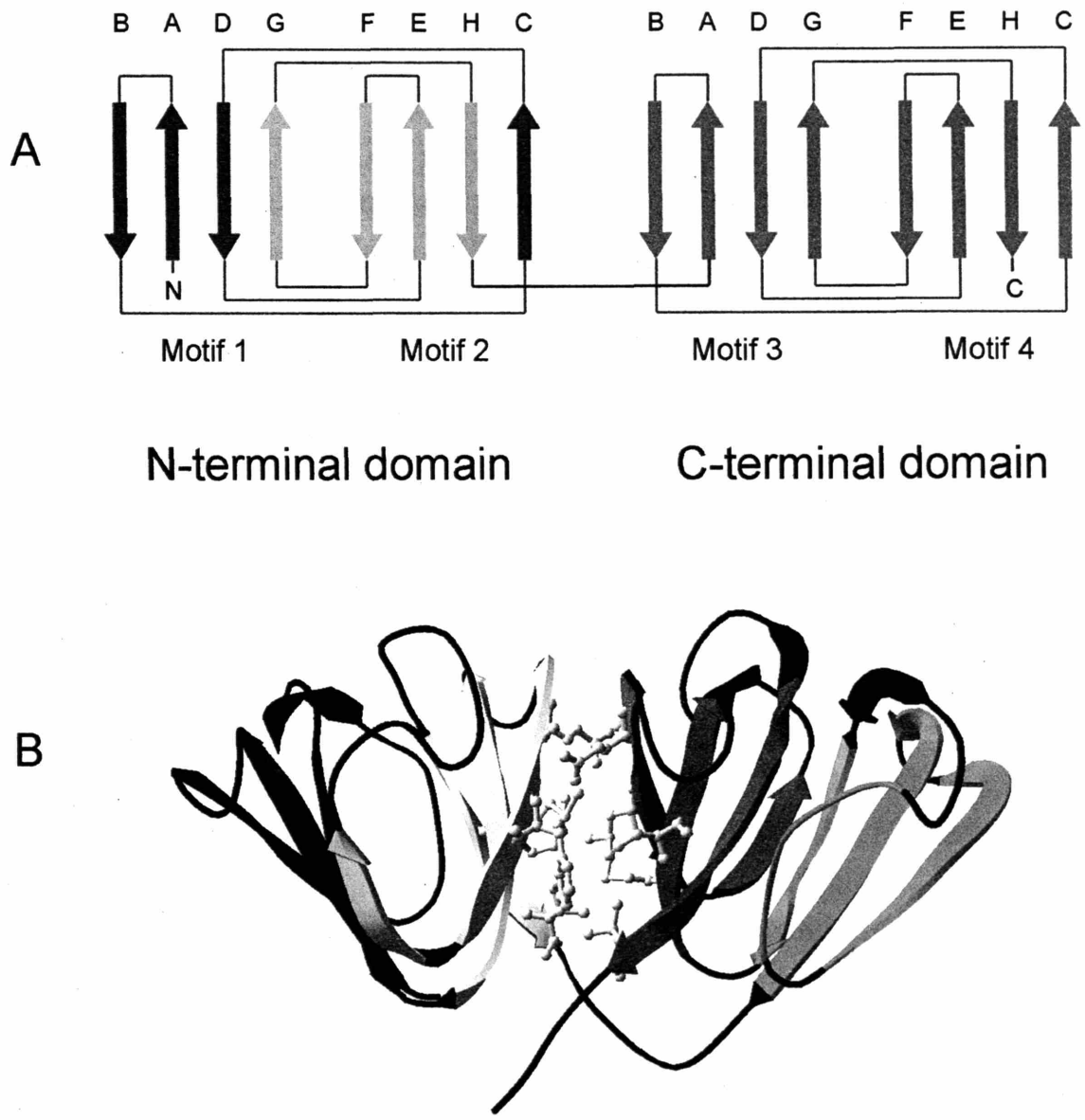

Figure 1-6. A) Two-dimensional topology diagram of the two-domain double Greek key motif fold characteristic of the β- and γ-crystallins. The strands are lettered A through H in each domain. B) Ribbon diagram of monomeric HγD-Crys. Amino acids making contact across the domain interface are shown in stick representation.

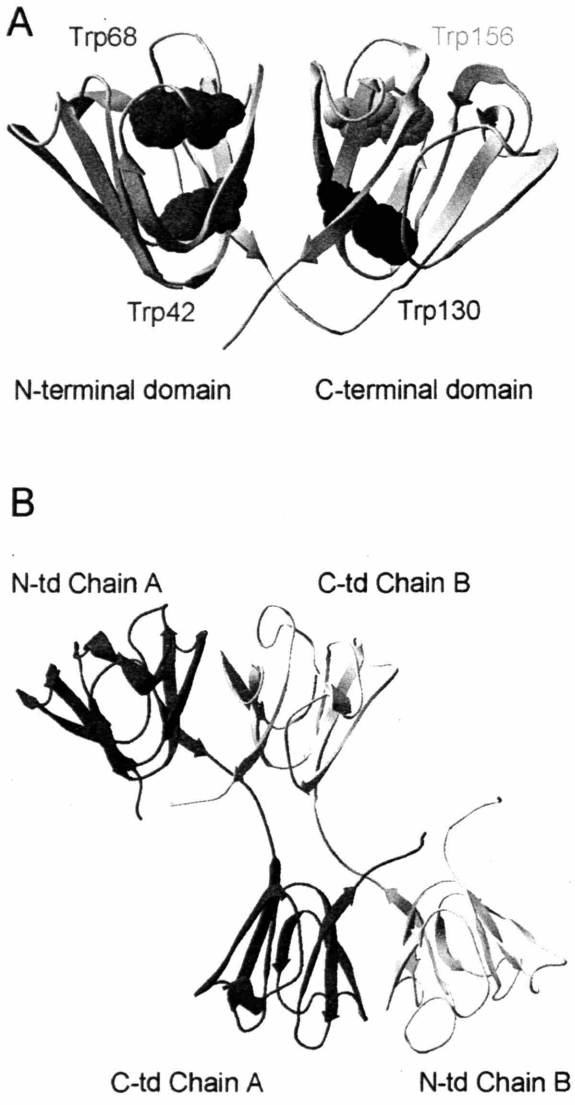

Figure 1-7. (A) Ribbon structure of monomeric HγD-Crys where the domains interact intramolecularly (Basak et al. 2003). Locations of the four buried tryptophans are shown in spacefill representation. (B) Ribbon structure of the domain swapped dimer of BβB2-Crys where the N-td of one monomer pairs with the C-td of the other monomer (Bax et al. 1990).

There are seven γ-crystallins genes in the human genome, the γA-F and γS genes. The γD and γC genes are expressed early in lens development and thus their corresponding proteins are found primarily in the aged lens nucleus. The γS gene is expressed post-natally and thus the γS protein is localized in the lens cortex (Wistow et al. 2002). Neither γA nor γB are expressed at appreciable levels in the human lens and in the human genome the genes for γE and γF are pseudogenes (Brakenhoff et al. 1990). There is also evidence of non-lens and possibly stress-related expression of γS-crystallin (Sinha et al. 1998; Bloemendal et al. 2004).

Several non-lens proteins also belong to the βγ-crystallin structural superfamily. Among these are the two domain spore coat protein, Protein S from *Myxococcus xanthus* and the single domain, homodimeric protein Spherulin 3a from *Physarum polycephalum* (Wistow et al. 1985; Wistow 1990). Other members are the human protein AIM1, the amphibian protein EDSP, yeast-killer toxin WmKT from *Williopsis mrakii* and protease inhibitor SMPI from *Streptomyces nigrescens* (Bloemendal et al. 2004). These non-lens proteins almost certainly share an evolutionary ancestor with the β- and γ-crystallins that possibly had a stress-response function (Wistow 1990). This ancestral protein was likely recruited to the lens in an early vertebrate, where the gene encoding it underwent a series of duplication events giving rise to the vertebrate β- and γ-crystallins recognized today (Shimeld et al. 2005; Wistow et al. 2005).

The β- and γ-crystallins display relatively low overall sequence identity of ~30% (Bloemendal et al. 2004). However, there are regions of high sequence conservation between the proteins that contribute to important structural elements. First, the tyrosine corner is a conserved feature of the double Greek key β-sandwich fold (Hemmingsen et al. 1994). The tyrosine corner is formed by the hydroxyl group of a tyrosine on one side of the β-sandwich forming a hydrogen bond with a backbone atom of a β-strand from the opposite side of the sandwich. There are two tyrosine corners in the β- and γ-crystallins, one each in the N-td and C-td. Tyrosines are conserved in these positions at 100% identity in γ-crystallins and the N-td of the β-crystallins, and 67% identity in the C-td of

the β-crystallins. The tyrosine corners are also conserved in Protein S but not in Spherulin 3a.

The second region of high sequence conservation in the β- and γ-crystallin is the domain interface. As mentioned above, the domain interface residues are made of a central cluster of hydrophobic residues abutted by two pairs of peripheral polar residues (Fig. 1-6). These domain interface residues are highly conserved among β- and γ-crystallins from diverse vertebrates suggesting that they are important in determining and maintaining the folds of these two-domain proteins (Table 1-1 and 1-2). The βγ-crystallin domain interface residues are not well conserved in Protein S and Spherulin 3a, which have different modes of domain interaction (Bagby et al. 1994; Rosinke et al. 1997; Clout et al. 2001). The intramolecular domain interactions in Protein S are assymetrical and occur between strand G from the N-td and B from the C-td. The intermolecular domain interactions in the Spherulin 3a dimer, occur through interaction across the face of an additional β-strand not found in the β- and γ-crystallins (Rosinke et al. 1997).

Residues contributing to the β-hairpins between the A/B and E/F strands of each domain also show high sequence conservation (Fig. 1-6). There are conserved glycines present at the beginning of the B and F strands and conserved serines at the beginning of the D and H strands that hydrogen bond to mainchain atoms of the A and E strands. This hairpin results in precise positioning of two interacting aromatic amino acids. Collectively, these features are believed to stabilize this β-hairpin (Blundell et al. 1981; Wistow et al. 1985; MacDonald et al. 2005). These residues are also conserved to 100% identity in both Protein S and Spherulin 3a.

The β- and γ-crystallins have high numbers of conserved aromatic amino acids, including four tryptophans buried in the domain cores (Fig. 1-7). Additionally, there are many well conserved tyrosines (13 in HγD-Crys), which are primarily surface-exposed and distal from the linker peptide. Interestingly, these aromatics are not well conserved in the non-lens proteins Spherulin 3a and Protein S. There are only six tyrosines and no

tryptophans in Protein S, and only one tryptophan and two tyrosines in Spherulin 3a. The high percentages of aromatics may have evolved in the lens proteins or alternatively, were present in the ancestral protein and lost in the non-lens proteins. In either case, the evolutionary pressure for lens proteins to have high percentages of aromatics may be related to continuous exposure of the lens proteins to visible and UV light.

Table 1-1. Conservation of domain interface residues among 35 vertebrate γ-crystallins determined by sequence alignment (see Appendix A for full alignments).

Residue in human γD crystallin	**% Identity**	**% Similarity**	**Other amino acids[1]**
Met43	74	91	Val, Ala
Gln54	89	97	Met, Pro
Phe56	74	100	Ile, Val, Leu
Arg79	83	91	Lys, His, Cys
Ile81	77	97	Val, Leu, Pro
Val132	54	100	Ile, Leu
Gln143	80	80	Met, Leu
Leu145	71	83	Phe, Tyr
Met147	0	0	Arg, Asp, Glu, Lys
Val170	49	97	Ile, Met, Leu, Ala

[1]Other amino acids found in the position listed by decreasing frequency.

Table 1-2. Conservation of domain interface residues among 32 vertebrate β-crystallins determined by sequence alignment (see Appendix A for full alignments).

Residue in human γD crystallin	% Identity	% Similarity	Other amino acids[1]
Met43	0	100	Val, Leu, Ile
Gln54	78	78	Met
Phe56	0	100	Val, Ile
Arg79	100	100	—
Ile81	53	97	Val, Leu, Phe, Ala
Val132	100	100	—
Gln143	100	100	—
Leu145	34	100	Val, Ile
Met147	0	0	Asp
Val170	22	100	Ile, Leu

[1]Other amino acids found in the position listed by decreasing frequency.

3. *Human congenital cataracts and the γ-crystallins*

There are several cases of heritable congenital cataracts caused by single amino acid substitutions in the gene encoding HγD-Crys and one case caused by a mutation in the gene encoding HγC-Crys (Heon et al. 1999; Stephan et al. 1999; Kmoch et al. 2000; Santhiya et al. 2002). In HγD-Crys the mutations are R14C, P23T, R36S and R58H. Studies of recombinant proteins containing these mutations led to the hypotheses that the R58H and R36S mutants cause cataract by crystallization (Kmoch et al. 2000; Pande et al. 2001; Basak et al. 2003), the R14C mutant by intermolecular disulfide bonding (Pande et al. 2000) and the P23T mutant by precipitation (Evans et al. 2004). Interestingly, none of these mutations appeared to alter the conformation or stability of HγD-Crys, suggesting that polymerization or association of native HγD-Crys occurred in the lenses of affected individuals (Pande et al. 2000; Pande et al. 2001; Evans et al. 2004). In

contrast, the congenital mutant T5P of HγC-Crys caused an altered native state fold, reduced stability and an increased tendency to interact with αA-Crys (Fu and Liang 2002; Liang 2004). The T5P mutation of HγC-Crys may have caused cataract by aggregation of partially unfolded chains.

4. *Folding stability and oligomerization of the β- and γ-crystallins*

The *in vitro* stabilities of the crystallin proteins have been intensively studied as a means of understanding how the proteins remain folded for a lifetime *in vivo*. The monomeric γ-crystallins have relatively high free energies of unfolding in the absence of denaturant (ΔG°) that are generally greater than that of the multimeric α- and β-crystallins (Bloemendal et al. 2004). Additionally, some γ-crystallins have high kinetic barriers to unfolding that may aid in maintenance of their native folds (Das and Liang 1998; Jaenicke and Slingsby 2001). As outlined below, several of the β- and γ-crystallins studied to date have complex equilibrium unfolding/refolding transitions that suggest the population of partially folded intermediates.

a. *γB crystallin*

The initial comprehensive studies addressing the folding and stability of the γ-crystallins focused on bovine γB crystallin (BγB-Crys). A two-state equilibrium unfolding transition was observed for BγB-Crys at pH 7.0 (N⇔U), but at pH 2.0 a three-state transition was observed signifying population of an intermediate (I) in equilibrium with N and U (N⇔I⇔U) (Rudolph et al. 1990). This intermediate was theorized to have one domain folded and the other unfolded (Rudolph et al. 1990; Sharma et al. 1990). To test this, recombinant proteins that corresponded to the isolated domains of BγB-Crys were constructed and their stabilities were analyzed (Mayr et al. 1997).

At pH 7.0, the isolated N-terminal domain (γBN-td) was only marginally more stable than the isolated C-terminal domain (γBC-td). At pH 2.0, however, γBN-td was notably more stable than γBC-td (Mayr et al. 1997). Comparing these data to that of

wild-type BγB-Crys, the authors concluded that in the full-length protein, domain interface interactions stabilized the C-td at pH 2.0. Additionally, it was observed that stability of γBC-td decreased with decreasing pH, while stability of γBN-td was independent of pH (Mayr et al. 1997). γBC-td has three more basic residues than γBN-td so that at pH 2.0 its net charge was +16 versus +13 for γBN-td. It was hypothesized that differences in stabilities of the domains at pH 2.0 was due to the greater number of charged residues on the surface of γBC-td, which may have destabilized the protein by charge repulsion (Mayr et al. 1997).

To confirm that the domain interface interactions of BγB-Crys were stabilizing the C-td, site-directed mutants of the domain interface residue Phe56 were constructed and stabilities of the mutant proteins were analyzed (Palme et al. 1997). Phe56 is one of six hydrophobic residues clustered in the center of the domain interface that is highly conserved among β- and γ-crystallins from various species (Fig. 1-6; Tables 1-1 and 1-2). Mutating this residue to alanine, aspartate or tryptophan caused significant destabilization of the C-td at pH 2.0 (Palme et al. 1997). Crystal structures of the mutant proteins were also determined and contrary to expectations, no changes in domain core structures or domain interface pairing were observed. This suggests that the fold was too rigid to adjust to the mutant residues (Palme et al. 1998).

b. *γS and γC crystallins*

γS crystallin differs from the other γ-crystallins as it is expressed post-natally in the lens, exhibits non-lens expression and has a longer domain linker (Bloemendal et al. 2004). Thus far it has not been possible to crystallize full-length human γS crystallin (HγS-Crys), however, the crystal structure of a HγS-Crys C-td homodimer has been determined and an NMR structure of wild-type murine γS-Crys was also recently solved (Purkiss et al. 2002; Wu et al. 2005). Despite the additional linker residue, the domains of HγS-Crys display the typical intrachain domain interactions characteristic of the γ-crystallins.

The stability of wild-type HγS-Crys, wild-type bovine γS crystallin (BγS-Crys) and their independent domains have been studied by Wenk et al. (2000). The equilibrium transitions of both wild-type HγS-Crys and BγS-Crys were best fit to two-state models (N⇔U) at pH 7.0, suggesting that the domains unfolded cooperatively in the full-length protein (Wenk et al. 2000). However, the isolated N-td of HγS-Crys was less stable that the isolated C-td suggesting that domain interface interactions were stabilizing in the full-length protein (Wenk et al. 2000). The differences in N-td and C-td stability were not as great for BγS-Crys as they were for HγB-Crys.

In contrast to the other γ-crystallins which displayed two-state transitions (N⇔U) at pH 7.0, human γC crystallin (HγC-Crys) displayed a three-state equilibrium transition (N⇔I⇔U) at physiological pH (Fu and Liang 2002). The equilibrium intermediate likely had one domain folded and the other unfolded. However, because stabilities of the isolated domains of HγC-Crys were not studied, it was not possible to predict which domain was folded and which was unfolded in the intermediate.

c. *Protein S and Spherulin 3a*

The domains of Protein S and Spherulin 3a each bind two Ca^{2+} molecules via surface patches. While sequences of the β- and γ-crystallins do not appear to be conserved for Ca^{2+}-binding (Clout et al. 2001), recent investigations have suggested that the γ-crystallins are capable of binding four Ca^{2+} molecules at μM affinities (Rajini et al. 2001) The tertiary structures of Spherulin 3a and Protein S were altered in the absence of Ca^{2+} and the domains were stabilized by Ca^{2+} binding (Kretschmar et al. 1999; Wenk et al. 1999). Spherulin also displayed an increase in kinetic stabilization upon binding Ca^{2+}, and Protein S displayed kinetic stabilization as a result of domain interactions (Wenk et al. 1998; Kretschmar et al. 1999).

The stability of Protein S has been studied in depth as a function of pH in the presence and absence of Ca^{2+}. A partially folded intermediate was observed in equilibrium experiments performed at pH 2.0, at which the protein does not bind Ca^{2+} (Wenk and Mayr 1998). Two major thermal transitions were also observed at both pH

2.0 and 7.0 in the absence of Ca^{2+} suggesting independent unfolding of the two domains (Wenk and Jaenicke 1999). In contrast, at pH 7.0 and in the presence of Ca^{2+}, wild-type Protein S unfolded very cooperatively by both chemical and thermal denaturation (Wenk and Mayr 1998; Wenk and Jaenicke 1999).

The isolated domains of Protein S were studied to further describe the intermediates observed for the full-length protein. No homo- or hetero-dimerization was observed for the isolated domain proteins. The isolated N-td was more stable than wild-type Protein S and the isolated C-td was less stable than wild-type Protein S (Wenk et al. 1999) Both of the isolated domain proteins were stabilized by Ca^{2+}, however, stabilities of the domains responded differently to pH (Wenk et al. 1999; Wenk and Jaenicke 1999). These results suggest that the C-td was stabilized by domain interactions in the full-length protein while the N-td was slightly destabilized in the full-length protein.

The structure of the isolated N-td of Protein S in the absence of Ca^{2+} was also investigated by NMR spectroscopy (Bagby et al. 1998). The protein formed two equilibrium intermediates with slightly different structures. The first intermediate had a core group of residues that were in a native conformation and surrounding residues fluctuating between native and near-native states. In the second intermediate the core and the surrounding residues all appeared to be in a native conformation. This Ca^{2+}-insensitive core region was composed of the tyrosine corner, the E/F β-hairpin and the α-helix found in the loop that connects the C and D strands (Bagby et al. 1998). Although there are differences in the motif arrangement and domain interactions of Protein S and the β- and γ-crystallins, these experiments may provide insight into potential folding nuclei or exceptionally stable regions of the double Greek key folds.

d. *β-crystallins*

Both rat and human βB2-Crys were shown to be less stable than the monomeric γ-crystallins and their equilibrium unfolding transitions were best fit to a three state model (N⇔I⇔U) at physiological pH (Wieligmann et al. 1999; Fu and Liang 2002) For rat βB2-Crys, the first transition was concentration dependent suggesting that dimer dissociation occurs during this step (Wieligmann et al. 1999). Properties of the isolated domains of rat βB2-Crys were also analyzed to probe for differential domain stability.

The isolated N-td, but not C-td was capable of forming homodimers in solution where dimer dissociation was observed at less than 50 μg/ml protein concentrations (Wieligmann et al. 1999). The isolated N-td was also less stable than the isolated C-td (Wieligmann et al. 1999). These data suggested that the partially folded intermediate was a monomeric species with a folded C-td and unfolded N-td. Similar to the BγB-Crys and Protein S results, domain interface interactions stabilized the domain of βB2-Crys with lower intrinsic stability.

Despite analogous domain structures, the β-crystallins form oligomeric structures but the γ-crystallins do not. βB2-Crys dimerizes by domain-swapping where the domain linker sequence is in a extended conformation compared to the bent linker of the γ-crystallins (Bax et al. 1990). The linker of the β-crystallins lack a conserved glycine and proline found in the γ-crystallin linkers. It was originally hypothesized that variation in quaternary structure may be due to differences in the linker peptide connecting the domains or the N- and C-terminal extensions of the β-crystallins. To test this, a βB2-Crys mutant without the N- and C-terminal extensions was constructed, which, unlike wild type formed of both dimeric and tetrameric species (Trinkl et al. 1994). Therefore, the extensions likely function to prevent higher order association and do not control dimerization.

To test the role of linker sequences in dimerization, another mutant was constructed that had the wild-type βB2-Crys domains connected by the linker sequence of BγB-Crys. This protein formed monomers only (Trinkl et al. 1994). In a mirror experiment, a mutant was constructed that had the linker of βB2-Crys connecting the wild-type domains of BγB-Crys (Mayr et al. 1994). This mutant also formed monomers only (Mayr et al. 1994). Together these data indicate that the βB2-Crys linker was necessary but not sufficient for dimer formation. The domain interface interactions of βB2-Crys may also contribute to dimer formation. In contrast, introducing the linker of BγB-Crys into the wild-type βA3-Crys sequence did not prevent oligomerization (Hope et al. 1994; Sergeev et al. 2000). Similarly, deleting the extensions of βA3-Crys did not abolish dimer formation and in fact, the modified proteins formed a tighter dimer than wild type (Sergeev et al. 2000).

e. *γD crystallin*

Folding and stability of HγD-Crys, one of the most abundant γ-crystallins of the human lens, have been intensively studied using fluorescence as a probe of conformation (Kosinski-Collins and King 2003; Evans et al. 2004; Kosinski-Collins et al. 2004). The tryptophans of HγD-Crys are quenched in the native protein such that fluorescence intensity of HγD-Crys is lower in the native state than the unfolded state (Kosinski-Collins and King 2003; Kosinski-Collins et al. 2004). Native state quenching has also been observed for a number of other β- and γ-crystallins suggesting that this phenomenon is a general property of the βγ-crystallin fold (Kim et al. 2002; Bateman et al. 2003).

The equilibrium unfolding transition of HγD-Crys at pH 7.0 and 37°C appeared largely two-state (N⇔U) in the denaturant guanidine hydrochloride (GuHCl) (Kosinski-Collins and King 2003). At 37°C, the equilibrium unfolding and refolding transitions overlaid at concentrations of GuHCl above 1.0 M, indicating that HγD-Crys was able to productively refold under these conditions. However, refolding to lower denaturant concentrations (<1.0 M) resulted in accumulation of a high molecular weight aggregate that competed with productive refolding (Kosinski-Collins and King 2003). The morphology of the aggregate was probed with AFM and found to have an elongated, fibrillar morphology different from amyloid fibers (Kosinski-Collins and King 2003). The aggregates were capable of binding the hydrophobic dye bisANS, indicating the presence of exposed hydrophobic surfaces in the aggregate (Kosinski-Collins and King 2003). Similar aggregation results were observed at 25°C, however, the equilibrium unfolding and refolding transitions at this lower temperature exhibited marked hysteresis (Kosinski-Collins and King 2003). Compared to the 37°C data, the 25°C unfolding transition was shifted to lower concentrations of GuHCl whereas the refolding transition was unchanged. These data suggest that there was a kinetic barrier to unfolding that was exaggerated at 25°C.

The unfolding and productive refolding kinetics of HγD-Crys were also measured at pH 7.0 and 37°C (Kosinski-Collins and King 2003). Both transitions were best fit by multiple exponentials suggesting population of partially folded/unfolded kinetic intermediates. To probe the conformations of the kinetic intermediates, triple tryptophan

mutant proteins were constructed that each had one endogenous tryptophan and the other three tryptophans substituted for phenylalanine (Kosinski-Collins et al. 2004). These mutant proteins, termed W42-only, W68-only, W130-only and W156-only, were studied by equilibrium and kinetic unfolding/refolding to independently monitor structural transitions of the two domains. In contrast to the isolated domain experiments described above for other β- and γ-crystallins, these experiments had the advantage of being able to independently monitor the domains while still maintaining them in the context of the full-length protein.

In equilibrium unfolding/refolding experiments, the mutants maintaining N-td tryptophans (W42-only and W68-only) both had transition midpoints of ~1.3 M GuHCl, while the proteins maintaining C-td tryptophans (W130-only and W156-only) both had transition midpoints of ~2.0 M GuHCl (Kosinski-Collins et al. 2004). It is important to note that transition midpoints of the triple mutants were significantly lower than the transition midpoint of wild-type HγD-Crys (2.8 M GuHCl) indicating the triple mutations were detrimental to stability (Kosinski-Collins et al. 2004). Nonetheless, differences between the transition midpoints of the mutants imply that the N-td of HγD-Crys is less stable than the C-td. These data also suggest that there may be an equilibrium intermediate present during the unfolding/refolding of wild-type HγD-Crys not previously detected.

Differences between the unfolding and refolding kinetics of the triple tryptophan mutants were also observed. During productive kinetic refolding to 1.0 M GuHCl, the mutants with N-td tryptophans refolded slower than the mutants with C-td tryptophans (Kosinski-Collins et al. 2004). Comparing these data to the refolding rates of wild-type HγD-Crys it was observed that refolding rates of the mutants with C-td tryptophans agreed well with the first kinetic transition of wild-type HγD-Crys and refolding rates of the mutants with N-td tryptophans agreed well with the second kinetic transition of wild type (Fig. 1-8) (Kosinski-Collins et al. 2004). Similarly, mutants with N-td tryptophans unfolded faster than mutants with C-td tryptophans (Kosinski-Collins et al. 2004). These observations led to a model of kinetic unfolding and refolding where a single major intermediate is populated that has intact tertiary structure in the C-td but not the N-td (Fig. 1-8).

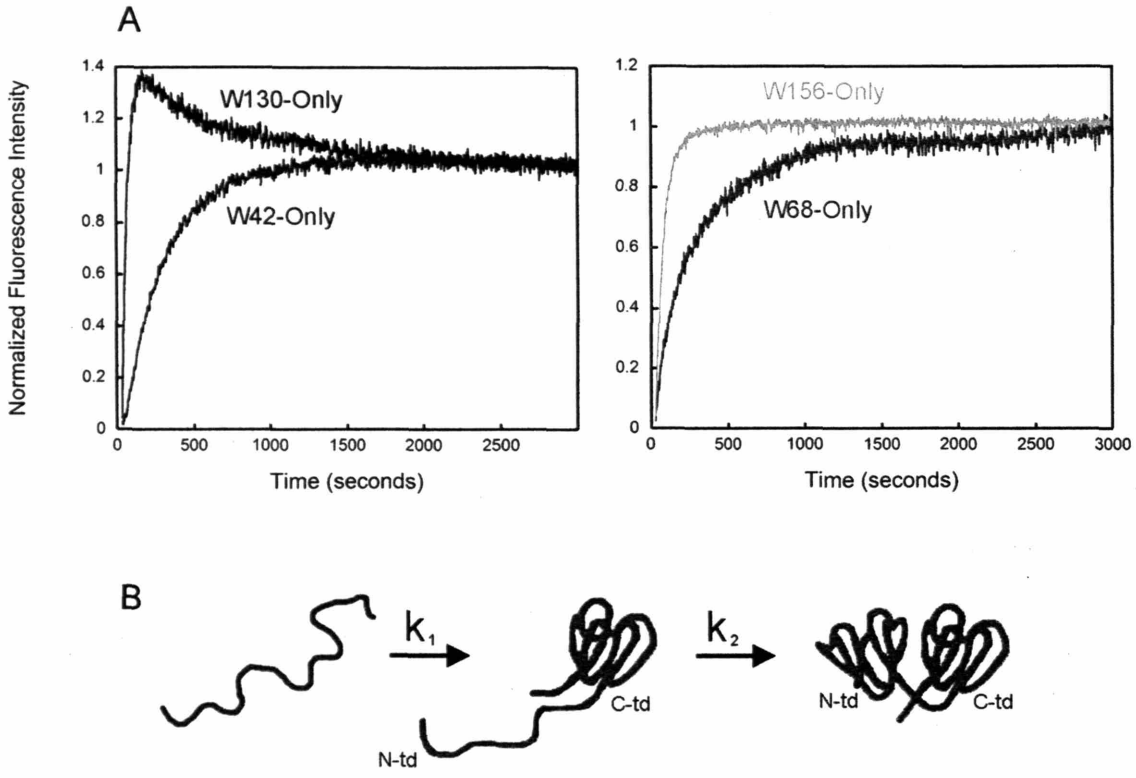

Figure 1-8. (A) Productive refolding kinetics of HγD-Crys triple-tryptophan mutants monitored by fluorescence (Kosinski-Collins et al. 2004).
(B) Schematic diagram of the productive refolding pathway of wild-type HγD-Crys where the C-td refolds first followed by the N-td.

Differences in the equilibrium and kinetic unfolding/refolding behavior of the triple tryptophan mutants suggest that domain interface interactions in HγD-Crys may contribute to folding and stability. This is in accord with previous observations of the importance of domain interactions in the two domain βγ-crystallin fold.

5. *Molecular mechanism of mature-onset cataract formation*

Mature-onset cataract is a unique protein aggregation disease. Protein inclusions of cataract can be localized in any region of lens and by unknown mechanisms aggregation is able to "communicate" across or traverse cell membranes. Additionally, the protein inclusions contain all members of the α-, β-, and γ-crystallins (Hanson et al. 2000; Searle et al. 2005). Structures of crystallins in the insoluble inclusions and the detailed molecular mechanisms of mature-onset cataract formation are completely unknown. Models of cataract formation have been informed by well-studied mechanisms of aggregation and *in vitro* studies of the crystallin proteins.

The lens crystallins do not regenerate or turnover in the lens thus necessitating life-long stability and solubility to ensure lens transparency (Oyster 1999). This task is particularly challenging given the continued presence of oxidative and radiative stress. The crystallin proteins in the aged lens are covalently damaged presumably as a result of age and environmental and cellular insults (Hanson et al. 1998; Hanson et al. 2000; Searle et al. 2005). The major forms of damage in the insoluble crystallins of aged lenses are deamidation, non-native disulfide bonds, methionine oxidation, and truncations (Hanson et al. 2000). Glycation and carbamylation are also common forms of crystallin modifications observed in the lens (Harding 2002). It is currently unknown if covalent damage causes the protein aggregation and insolubilization associated with cataract, or if covalent damage is enhanced in the insoluble inclusions.

Several *in vitro* studies have addressed the effects of covalent damage on properties of the crystallins (Lampi et al. 2001; Kim et al. 2002; Gupta and Srivastava 2004b; 2004a; Harms et al. 2004; Lampi et al. 2006). The effects of such damage ranged from formation of higher order oligomers, increased tendency of thermal aggregation, destabilization and partial unfolding for the β-crystallins (Lampi et al. 2001; Kim et al.

2002; Harms et al. 2004; Lampi et al. 2006), to changes in secondary and tertiary structure and decreased chaperone activity for α-crystallin (Gupta and Srivastava 2004b; 2004a). These results indicate that effects of covalent damage are context dependent, and that some damage may be detrimental enough elicit *in vivo* changes that could cause cataract formation.

One model for the molecular mechanism of cataract formation is crystallin liquid-liquid phase separation (Clark and Clark 2000). High concentration of the crystallins and their short-range order are responsible for lens transparency and refraction. Crystallin phase separation has been observed as a result of congenital cataract mutations in HγD-Crys (Pande et al. 2000; Pande et al. 2001; Pande et al. 2005), age-associated covalent modifications of βB1-Crys (Annunziata et al. 2005), changes in temperature, and as a result of X-ray irradiation (Clark et al. 1983). During these phase transitions, cytoplasmic crystallins separate into micovolumes of differing protein concentrations (Clark et al. 1983). Phase separation may cause light scattering and thus lens opacity and cataract (Clark and Clark 2000). However, the phase separation hypothesis does not explain the presence of insoluble protein fractions pathologically associated with cataract.

An alternative model of cataract formation is protein insolubilization as a result of covalent damage. According to this hypothesis, changes in surface properties of the crystallins may disrupt short-range order in the highly-concentrated lens cells and cause reactions akin to native-state polymerization or association (Delaye and Tardieu 1983). The major types of covalent damage of aged crystallins including deamidation, oxidation, glycation, and carbamylation, are all capable of changing the surface properties of proteins. In further support of this hypothesis, reduced solubility has been observed for the congenital cataract mutant P23T of HγD-Crys (Evans et al. 2004). Similarly, other congenital mutants of HγD-Crys appear to cause in native-state polymerization in the lens cells of affected individuals (Pande et al. 2000; Pande et al. 2001).

Finally, mature-onset cataract formation may be caused by protein aggregation induced by covalent damage. Unlike many protein aggregation diseases, protein insolubilization and cataract occur long after the crystallins were initially synthesized. Therefore, aggregation-prone states of the crystallins that may cause cataract are most likely populated as a result of destabilization and partial unfolding rather than during

productive refolding off of the ribosome. Studies of lysozyme and soybean trypsin inhibitor have demonstrated that protein unfolding and aggregation can result from *in vitro* oxidation of histidine, tryptophan and tyrosine residues (Hawkins and Davies 2005). Potential partially unfolded crystallin species may be prone to aggregate in the densely packed lens cells through mechanisms such as loop-sheet insertion or domain swapping. These two mechanisms are particularly intriguing models given the two-domain nature of the β- and γ-crystallins and the high β-sheet content of the proteins.

Amyloid fiber formation is a less likely explanation for crystallin aggregation given the lack of pathological data suggesting the presence of amyloid fibers in mature-onset cataract. However, α-, β- and γ-crystallins have been shown to form amyloid fibers under mild denaturing conditions (Meehan et al. 2004). Similarly, mouse congenital cataract mutations in the gene for γβ-crystallin cause formation of *in vivo* inclusions that are stained by the amyloid-detecting dye, Congo Red (Sandilands et al. 2002). Recombinant proteins with these congenital mutants also formed amyloid fibers *in vitro* under mildly denaturing conditions (Sandilands et al. 2002).

Understanding the molecular basis of mature-onset cataract is a daunting task considering the number of proteins involved and the complexity of the interactions that could form between them. In order to understand the phenomena leading to cataractogenesis, it is first necessary to understand the individual proteins involved. In particular, characterization of partially folded conformations adopted by the crystallins may elucidate potential aggregation-prone conformations that are significant in the aged lens. Additionally, understanding the effects of covalent damage on crystallin stability and aggregation may give insight into the role of damage in cataract formation.

D. SUMMARY OF THESIS

In this thesis I have analyzed the contributions of domain interface residues of HγD-Crys to folding, stability and aggregation (Chapters 2 and 3). Following this is an analysis of the effects of domain interface glutamine deamidation on physiologically important properties of HγD-Crys (Chapter 4). These experiments have important

implications for understanding how HγD-Crys remains stable and folded for a lifetime *in vivo*, and how these properties are affected by covalent damage. These experiments also probe for partially folded states on the unfolding and refolding pathways of HγD-Crys that may provide insight into conformations important in aggregation and cataract formation. More broadly, these studies have general applications to understanding β-sheet protein folding, assembly and aggregation.

CHAPTER TWO:

CONTRIBUTIONS OF HYDROPHOBIC DOMAIN INTERFACE INTERACTIONS TO THE FOLDING AND STABILITY OF HUMAN γD CRYSTALLIN[1]

[1] **Reprinted from:** Flaugh, S.L., Kosinski-Collins, M.S., and King, J. (2005) Contributions of hydrophobic domain interface interactions to the folding and stability of human γD crystallin. *Protein Sci.* 14:569-581.

A. ABSTRACT

Human γD-crystallin (HγD-Crys) is a monomeric eye lens protein composed of two highly homologous β-sheet domains. The domains interact through inter-domain side chain contacts forming two structurally distinct regions, a central hydrophobic cluster and peripheral residues. The hydrophobic cluster contains Met43, Phe56 and Ile81 from the N-terminal domain (N-td) and Val132, Leu145 and Val170 from the C-terminal domain (C-td). Equilibrium unfolding/refolding of wild-type HγD-Crys in guanidine hydrochloride (GuHCl) was best fit to a three-state model with transition midpoints of 2.2 and 2.8 M GuHCl. The two transitions likely corresponded to sequential unfolding/refolding of the N-td and the C-td. Previous kinetic experiments revealed that the C-td refolds more rapidly than the N-td. We have constructed alanine substitutions of the hydrophobic interface residues to analyze their roles in folding and stability. After purification from *E. coli*, all mutant proteins adopted a native-like structure similar to wild type. The transitions of M43A and V170A were similar to wild type. In contrast, the mutants F56A, I81A, V132A, and L145A had a destabilized N-td, causing greater population of the single folded domain intermediate. Compared to wild type, these mutants also had reduced rates for productive refolding of the N-td but not the C-td. These data suggests a refolding pathway where the domain interface residues of the refolded C-td act as a nucleating center for refolding of the N-td. Specificity of domain interface interactions is likely important for preventing incorrect associations in the high protein concentrations of the lens nucleus.

B. INTRODUCTION

The transparency of the human eye lens depends on the stability and solubility of the α-, β- and γ-crystallin proteins (Delaye and Tardieu 1983; Fernald and Wright 1983). Crystallins are present in the enucleated fibrous lens cells at concentrations of 200-400 mg/ml with the β- and γ-crystallins accounting for over 50% of the total protein (Oyster 1999). The β- and γ-crystallins are two domain proteins that structurally define the βγ-crystallin superfamily. The oligomeric α-crystallins exhibit *in vitro* molecular chaperone activity in addition to structural roles in lens transparency (Horwitz 1992; Boyle and Takemoto 1994). The crystallin proteins of the lens nucleus are synthesized early in lens development and do not regenerate during adulthood (Oyster 1999).

Human γD-crystallin (HγD-Crys) is a 173 amino acid protein found in the densely packed lens nucleus. The crystal structure of HγD-Crys has recently been solved to 1.25 Å (Fig. 2-1) and is consistent with the two-domain, primarily β-sheet structure of the βγ-crystallin superfamily (Basak et al. 2003). HγD-Crys is the third most abundant γ-crystallin in young human lenses (Lampi et al. 1997). Within each domain of HγD-Crys are two β-sheet Greek key motifs. The domains are connected by an extended six amino acid peptide and interact non-covalently through inter-domain amino acid side chain contacts that form two structurally distinct regions. These are (1) a central hydrophobic cluster and (2) polar peripheral pairwise interactions surrounding the cluster. The hydrophobic cluster consists of Met43, Phe56 and Ile81 from the N-td and Val132, Leu145 and Val170 from the C-td (Fig. 2-1). Peripheral pairwise interactions are between Gln54/Gln143 and Arg79/Met147.

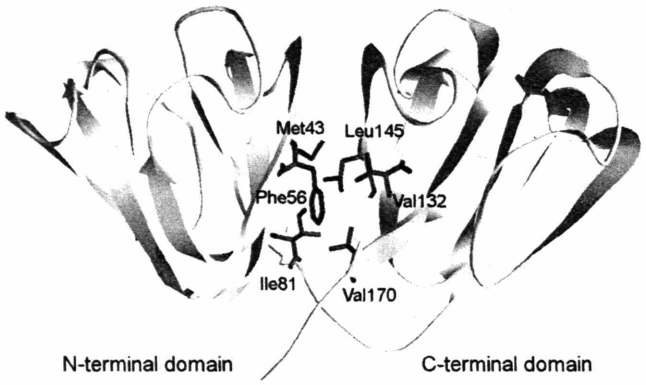

Figure 2-1. The crystal structure of wild-type HγD-Crys depicted in ribbon representation (Basak et al. 2003). Amino acids contributing to the hydrophobic cluster of the domain interface are shown in wire frame.

The mechanisms of aggregation for many protein deposition diseases have been elucidated by studying the *in vitro* unfolding and refolding of their associated proteins (Westermark et al. 1990; DiFiglia et al. 1997). A common feature of these mechanisms is that the aggregation-prone species adopts a partially folded or non-native conformation (Mitraki 1989; Wetzel 1994; Booth et al. 1997; Jiang et al. 2001; Nicholson et al. 2002). The processes that lead to loss of solubility and aggregation of crystallins are less well understood. In contrast to the aggregation mechanisms of some other protein deposition diseases, cataract is likely related to an unfolding and not a folding defect. The rare inherited juvenile-onset cataracts associated with mutations of HγD-Crys are caused by crystallization and intermolecular disulfide bonding of the native state molecules (Pande et al. 2000; Pande et al. 2001). These mechanisms are unlikely to be related to those of mature-onset cataract. Instead, aggregation in the aged lens is probably correlated with destabilization of crystallin proteins. Covalent damage is profuse in the crystallins of aged and cataractous lenses, presumably resulting from a lifetime exposure to UV and oxidative stresses (Hoenders and Bloemendal 1983; Hanson et al. 1998; Hanson et al. 2000). This damage may generate partially unfolded species of crystallins that polymerize through domain swapping, loop-sheet insertion or another unknown mechanism. Though many of the damaged molecules may be bound by α-crystallin, this process appears to break down or become saturated in older adults.

Previous analysis of the unfolding and refolding of HγD-Crys in guanidine hydrochloride (GuHCl) identified an *in vitro* aggregation pathway that may provide a model of crystallin aggregation (Kosinski-Collins and King 2003). The aggregate formed from partially folded species after refolding to concentrations of GuHCl less than 1.0 M. The aggregated protein had ordered morphology resembling polymerized states of globular subunits as seen by atomic force microscopy (Kosinski-Collins and King 2003). Subsequent experiments determined that the C-terminal domain (C-td) of HγD-Crys was more stable than the N-terminal domain (N-td) and the C-td acquired structure more rapidly during kinetic refolding (Kosinski-Collins et al. 2004). These results suggest that the domain interface of HγD-Crys may play a key role in folding and stability.

In this study we analyze the role of the hydrophobic domain interface cluster in folding and stability. Single alanine substitutions of these residues were constructed and the mutant proteins were analyzed for alterations in stability or refolding kinetics. The majority of the mutations affect both thermodynamic unfolding/refolding properties and kinetic refolding properties suggesting that the hydrophobic cluster contributes to stability and acts as a nucleus for domain refolding.

C. MATERIALS AND METHODS

1. *Mutagenesis, expression and purification of recombinant HγD-Crys*

Alanine substitutions of residues Met43, Phe56, Ile81, Val132, Leu145, and Val170 were constructed using PCR-based site-directed mutagenesis. Primers encoding the site-specific alanine substitutions (IDT-DNA) were used to amplify a pQE.1 plasmid encoding the HγD-Crys gene with an N-terminal 6-His tag (Kosinski-Collins et al. 2004). All resulting plasmids were sequenced to verify the substitutions and to ensure no additional mutations were present (Massachusetts General Hospital).

Recombinant wild-type and mutant HγD-Crys proteins were expressed and purified as described by Kosinski-Collins et al. (2004). Briefly, proteins were expressed in *E. coli* and cell lysates were purified to over 98% homogeneity by affinity chromatography with a Ni-NTA resin (Qiagen).

2. *Circular dichroism spectroscopy*

CD spectra of the purified proteins were collected with an AVIV model 202 CD spectrometer (Lakewood, NJ). Proteins were present in concentrations of 100 μg/ml for far-UV CD and 300 μg/ml for near-UV CD. Protein concentrations were determined by absorbance at 280 nm using an extinction coefficient of 41,040 cm^{-1} M^{-1} for wild type and mutant His-tagged proteins. All samples contained 10 mM sodium phosphate, 5 mM

DTT, 1 mM EDTA, pH 7.0. The buffer signal was subtracted from all spectra. Spectra were collected from 200 to 260 nm to monitor secondary structure and from 260 to 340 nm to monitor tertiary structure. An internal Peltier thermo-electric temperature controller was used to maintain the temperature at 37°C.

3. *Fluorescence emission spectroscopy*

Fluorescence emission spectra were recorded with a Hitachi F-4500 fluorimeter. Intrinsic tryptophan fluorescence was measured using an excitation wavelength of 295 nm and monitoring emission from 310 to 400 nm. A slitwidth of 10 nm was used for both excitation and emission. All samples contained 10 μg/ml purified protein in 10 mM sodium phosphate, 5 mM DTT, 1 mM EDTA, pH 7.0, and 5.5 M GuHCl where appropriate. Emission spectra were corrected for the buffer signal. A circulating water bath was used to maintain the temperature at 37°C.

4. *Equilibrium unfolding and refolding*

Equilibrium unfolding experiments were performed by diluting purified proteins to 10 μg/ml in 0 to 5.5 M GuHCl (purchased as an 8.0 M solution from Sigma-Adlrich, Saint Louis, MO). All unfolding samples contained 10 mM sodium phosphate, 5 mM DTT, and 1 mM EDTA, pH 7.0. Unfolding samples were incubated at 37°C for 24 hours to ensure equilibrium had been reached.

Equilibrium refolding experiments were carried out by initially preparing unfolded stock solutions of 100 μg/ml purified protein in 5.5 M GuHCl. The unfolded stock solutions were incubated at 37°C for five hours. The unfolded stocks were then diluted into refolding samples to give a final protein concentration of 10 μg/ml. Refolding samples contained 10 mM sodium phosphate, 5 mM DTT, 1 mM EDTA, pH 7.0, and GuHCl from 0.55 to 5.5 M. The refolding samples were allowed to reach equilibrium by incubation at 37°C for 24 hours.

Fluorescence emission spectra were recorded for each unfolding and refolding sample using a Hitachi F-4500 fluorimeter as described above. The concentration of

GuHCl in the unfolding/refolding samples was determined by measuring the refractive index. Data was analyzed by plotting the concentration of GuHCl for each sample versus the ratio of fluorescence intensities at 360 and 320 nm (FI 360/320 nm). All data were plotted from 0 to 5.0 M GuHCl, instead of 5.5 M GuHCl to improve visual clarity of the transitions. Equilibrium unfolding/refolding experiments of the wild-type and mutant proteins were performed three times each.

Equilibrium unfolding and refolding data were fit to a two-state model by the methods of (Greene and Pace 1974), or a three-state model by the methods of (Clark et al. 1993) using the curve fitting feature of Kaleidagraph (Synergy software). The model that best fit the data was selected based on a random distribution of residuals. Transition midpoints, ΔG_{H2O} and m-values were calculated for all transitions from these fits.

5. *Productive refolding kinetics*

Kinetic refolding experiments were carried out by diluting purified proteins to 100 μg/ml in 5.5 M GuHCl. The solutions were unfolded by incubation at 37°C for five hours. Refolding buffer containing 10 mM sodium phosphate, 5 mM DTT, and 1 mM EDTA, pH 7.0 was equilibrated to 37°C with stirring. Fluorescence emission of the refolding buffer was continually monitored in a Hitachi F-4500 fluorimeter using an excitation wavelength of 295 nm and an emission wavelength of 350 nm. Unfolded stocks were diluted into the refolding buffer using a syringe port injection system to give a final protein concentration of 10 μg/ml in 1.0 M GuHCl. Fluorescence emission of the refolding sample was monitored at 350 nm for three hours. The fluorescence emission spectra of resulting refolded samples were measured to ensure that the proteins had refolded into a native-like conformation. Kinetic refolding data were fit to one, two and three exponentials using the curve fitting feature of Kaleidagraph (Synergy software). The model with the best fit was determined by inspection. Kinetic refolding experiments of the wild-type and mutant proteins were performed two times each.

D. RESULTS

1. *Protein expression and purification*

Single alanine substitutions of the six hydrophobic domain interface residues of HγD-Crys were constructed using PCR-based primer extension. The mutant proteins were expressed at 37°C and purified by Ni-NTA affinity chromatography. Expression levels of all mutant proteins were comparable to wild type. The mutants behaved similarly to wild type during purification and were present in the soluble fraction after cell lysis. All proteins purified to greater than 98% homogeneity as determined by SDS-PAGE (Data not shown).

The proteins used in this study possessed an N-terminal His-tag of the sequence MKHHHHHHQ to aid in purification. Previous analysis of wild-type HγD-Crys with and without the His-tag confirmed that the exogenous peptide did not perceptibly alter the structure of the native state or the thermodynamic and kinetic unfolding/refolding properties (Kosinski-Collins and King 2003; Kosinski-Collins et al. 2004).

2. *Circular dichroism and fluorescence spectroscopy*

Circular dichroism (CD) and fluorescence emission spectroscopy were used to analyze the native state structures of hydrophobic domain interface mutants. The far-UV CD of wild-type HγD-Crys displayed a strong minimum at 218 nm in accord with previous results (Andley et al. 1996; Pande et al. 2000). All hydrophobic domain interface mutants had analogous spectra with a minimum at 218 nm suggesting similar β-sheet content as wild type (Fig. 2-2). These results indicate that the overall secondary structure content of the mutant proteins was similar to wild-type HγD-Crys. Despite the fact that the structures of the mutant proteins appeared to be similar to wild type, dynamic properties of the proteins may have been altered. This phenomenon has been previously observed in a mutational study of bovine pancreatic trypsin inhibitor (Beeser et al. 1997). The hydrophobic domain interface mutants could have had altered domain pairing not

detected by far-UV CD. The near-UV CD spectra of wild-type HγD-Crys and all mutant proteins superimposed suggesting similar aromatic environments (data not shown).

Fluorescence emission spectroscopy was used to further probe the environment of aromatic amino acids in wild-type and mutant HγD-Crys. HγD-Crys has four tryptophan residues, two per domain, buried in the hydrophobic cores of the two domains. Additionally, HγD-Crys has 14 tyrosines, many of which are surface-exposed. All fluorescence experiments performed here use an excitation wavelength of 295 nm to selectively excite the buried tryptophans and thus probe conformation of the domain cores. Wild-type HγD-Crys displayed a native-state emission maximum of 325 nm and an unfolded maximum of approximately 350 nm (Fig. 2-3). The fluorescence emission intensity increased upon unfolding indicating that the tryptophans were quenched in the native fold (Fig. 2-3). This phenomenon has been previously described for several of the β- and γ-crystallins (Kim et al. 2002; Bateman et al. 2003; Kosinski-Collins et al. 2004).

Tryptophan emission of domain interface mutants was measured in an analogous method as wild-type HγD-Crys. All mutant proteins displayed a native emission maximum of 325 nm and an unfolded maximum of about 350 nm (Fig. 2-3). Fluorescence emissions of all proteins were quenched in the native state (data not shown). These results suggest that the hydrophobic domain interface mutations did not disrupt the structure of the native state buried hydrophobic cores. Similar to the CD measurements described above, tryptophan fluorescence of HγD-Crys would not report altered domain pairing.

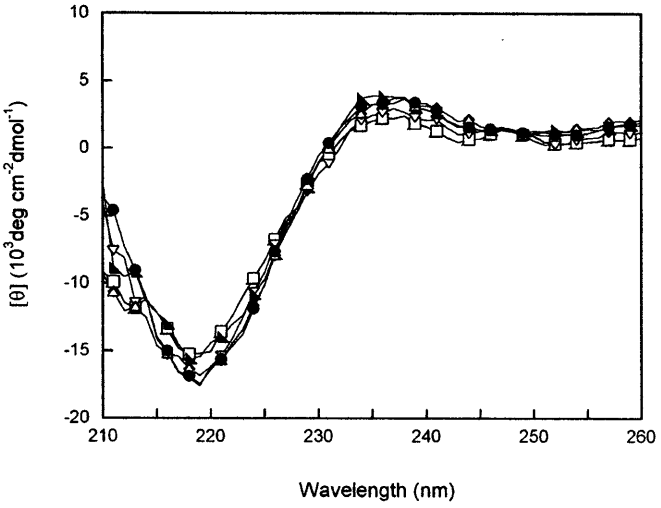

Figure 2-2. Far-UV CD of wild-type (◢), M43A (□), F56A (▼), I81A (◣), V132A (◆), L145A (△) and V170A (●) HγD-Crys. Samples contained 100 μg/ml protein in 10 mM sodium phosphate, 5 mM DTT, 1 mM EDTA, pH 7.0 at 37 C. A 0.25 cm pathlength cuvette was used for all measurements. All spectra were corrected for background buffer signal.

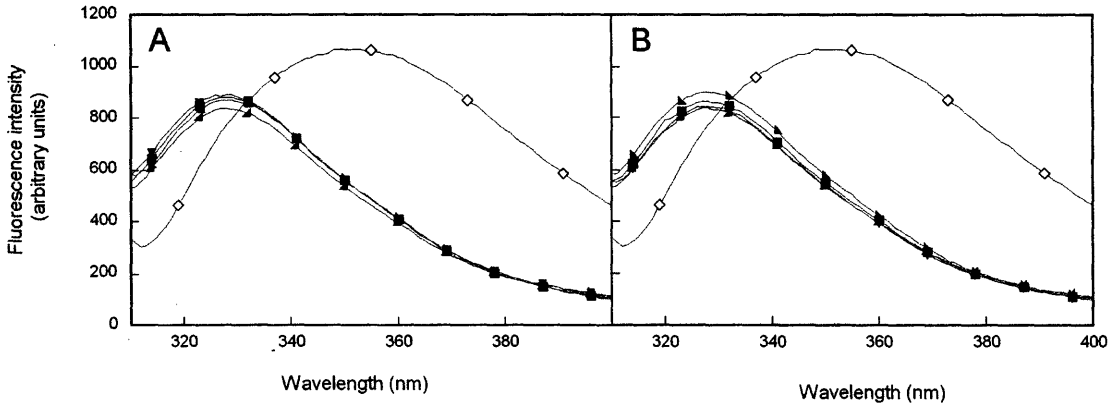

Figure 2-3. A) Fluorescence spectroscopy of native (◢) and unfolded (◇) wild-type HγD-Crys, and native N-terminal domain mutants M43A (■), F56A (▼), and I81A (◣). B) Fluorescence spectroscopy of native (◢) and unfolded (◇) wild-type HγD-Crys, and C-terminal domain mutants V132A (◆), L145A (▲) and V170A (●). Protein was present at 10 μg/ml protein in 10 mM sodium phosphate, 5 mM DTT, 1 mM EDTA, pH 7.0, and GuHCl where appropriate at 37 C. All spectra were corrected for background buffer signal.

3. *Equilibrium unfolding and refolding of wild type*

In order to assess the stability of the wild-type and mutant proteins, equilibrium unfolding/refolding experiments were performed. Tryptophan emission was used to probe the conformation of the domains using GuHCl as a denaturant at 37°C, pH 7.0. To best assess the shape of the transitions, a ratio of fluorescence intensities at 360 and 320 nm (FI 360/320 nm) was plotted as a function of GuHCl concentration.

Equilibrium unfolding/refolding of wild-type HγD-Crys has been previously investigated (Kosinski-Collins and King 2003). The unfolding and refolding samples in this earlier investigation were allowed to equilibrate at 25°C or 37°C for six hours prior to measuring fluorescence emission. The transitions were best fit to a two-state model for both temperatures. At 25°C the unfolding and refolding transitions exhibited significant hysteresis. The unfolding transition had a midpoint of 3.7 M GuHCl while the refolding transition had a midpoint of 2.7 M GuHCl. In contrast, at 37°C the two transitions deviated only slightly and both had midpoints of ~2.7 M GuHCl (Kosinski-Collins and King 2003). These observations suggested that structural transformations were controlled by a high kinetic barrier. Given that the unfolding transition but not the refolding transition changed with temperature, the kinetically controlled step was likely on the unfolding pathway.

In order to test for the presence of a high kinetic barrier to unfolding, we extended the equilibration time for all unfolding and refolding samples to 24 hours. No hysteresis was evident between the unfolding and refolding transitions of wild-type HγD-Crys at these extended equilibration times (Fig. 2-4). Additionally, the increased times caused a shift in the location of the unfolding transition only. This further confirms the presence of a high kinetic barrier during unfolding. The molecular basis of the hysteresis is a subject of current investigations.

Compared to the data collected with a six hour equilibration time, the 24 hour unfolding data exhibited a decreased slope in the transition region possibly reflecting reduced cooperativity of the reaction. The m-value for a two-state fit of six hour data was 3.6 ± 0.1 while an m value of 2.7 ± 0.1 was calculated for a two-state fit of 24 hour data. This change may suggest an unfolding/refolding mechanism for the 24 hour data that is more complicated than the two-state model previously employed. To test this, equilibrium unfolding/refolding transitions of wild-type HγD-Crys were fit to both a two- and three-state model and residuals of the fit were calculated (Fig. 2-4). The two-state model assumes direct transition between the native and unfolded states, while a three-state model allows for population of a partially folded intermediate. When fit to a two-state model the unfolding/refolding transitions had midpoints of 2.8 M GuHCl (Fig. 2-4A). By visual examination, the two-state fit appeared to be valid; however, the residuals displayed a semi-regular pattern, especially in the region of 2 to 3 M GuHCl (Fig. 2-4A). In contrast, fitting to a three-state model yielded a midpoint of 2.15 M for a transition from native to partially folded intermediate and a midpoint of 2.8 M GuHCl for an intermediate to unfolded transition (Fig. 2-4B). The residuals for the three-state fit had an overall lower magnitude and a more random arrangement than those of the two-state fit (Fig. 2-4B). This observation was reproducible in all three iterations of the experiment.

The three-state fit of the equilibrium unfolding data suggested that an intermediate was populated in the region of 2.5 M GuHCl. Given that the two-state fit was particularly poor in this region, it is likely that a three-state model is a better description of the data. The two transitions may correspond to independent unfolding/refolding of the two domains. These results along with apparent ΔG_{H2O} and m values calculated for the transitions are reported in Table 2-1.

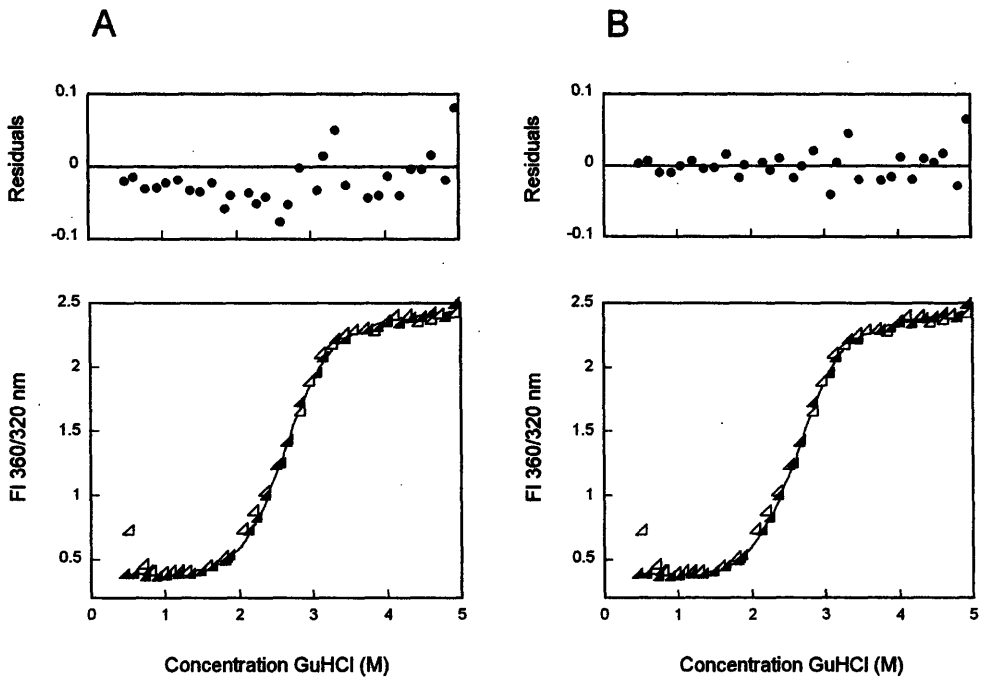

Figure 2-4. Equilibrium unfolding (closed symbols) and refolding (open symbols) of wild-type HγD-Crys in GuHCl probed by fluorescence emission. (A) The solid black line represents a two-state fit of equilibrium unfolding data. Residuals of the fit are shown. (B) Equilibrium unfolding data fit to a three-state model (solid black line), including residuals of the fit. Fluorescence spectra were recorded for each sample using an excitation wavelength of 295 nm. Fluorescence intensity at 360/320 nm was used in order to simultaneously monitor changes in the unfolding and native maxima. Samples were allowed to equilibrate in GuHCl at 37°C for 24 hours prior to recording fluorescence emission spectra. By inspection, both fits appear suitable; however, residuals of the three-state fit are of lower magnitude and more random than those of the two-state fit. This observation along with other factors described in the text suggests that the three-state fit is a better description of the data.

Table 2-1. Equilibrium unfolding/refolding parameters for wild-type and mutant HγD-Crys.

Protein	Transition 1			Transition 2		
	Apparent ΔG_{H2O}[a]	Apparent m value[b]	$[GuHCl]_{1/2}$[c]	Apparent ΔG_{H2O}[a]	Apparent m value[b]	$[GuHCl]_{1/2}$[c]
Wild type	7.7 ± 0.2	3.6 ± 0.1	2.2 ± 0.1	8.9 ± 1.3	3.1 ± 0.4	2.8 ± 0.1
M43A	7.7 ± 0.1	3.7 ± 0.1	2.2 ± 0.1	8.1 ± 0.8	2.8 ± 0.3	2.9 ± 0.1
F56A	6.1 ± 0.7	3.9 ± 0.4	1.6 ± 0.1	9.3 ± 0.7	3.2 ± 0.2	2.9 ± 0.1
I81A	5.3 ± 0.6	3.7 ± 0.4	1.5 ± 0.1	8.7 ± 0.4	3.0 ± 0.2	2.9 ± 0.1
V132A	5.0 ± 0.8	3.7 ± 0.6	1.3 ± 0.1	7.8 ± 0.2	3.1 ± 0.2	2.5 ± 0.1
L145A	6.5 ± 0.5	3.9 ± 0.1	1.6 ± 0.1	7.5 ± 0.1	2.9 ± 0.1	2.6 ± 0.1
V170A	N/A	N/A	N/A	10.2 ± 0.9	4.0 ± 0.3	2.5 ± 0.1

[a] Free energy of unfolding in units of kcal*mol^{-1} calculated assuming a linear dependence on the concentration of GuHCl.

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

[c] Transition midpoints in units of M.

4. *In vitro aggregation*

Consistent with previous results, wild-type HγD-Crys aggregated upon rapid refolding out of 5.5 M GuHCl (Kosinski-Collins and King 2003). A native-like conformation was attained when refolded to 1.0-1.8 M GuHCl. However, refolding to less than 1.0 M GuHCl resulted in the accumulation of a high molecular weight aggregate. This was seen as a sharp increase in FI 360/320 nm due to right angle light scattering by the aggregate (Fig. 2-4). Association by intermolecular disulfide bonding was prevented by inclusion of 5 mM DTT in all refolding samples.

Previous experiments investigated the morphology of the aggregate using atomic force microscopy. The aggregate adopted a fibrillar structure which did not bind Congo red or thioflavin T (Kosinski-Collins and King 2003). Previous results also indicate that the levels of aggregation are consistent over a range of incubation times from 3 to 41 hours (Kosinski-Collins and King, 2003). The levels of aggregation seen here with a 24 hour incubation time were also consistent with those previously reported. Therefore, the effect of the increased incubation time was restricted to a change in the position of the unfolding transition.

All hydrophobic domain interface mutant chains also aggregated upon rapid refolding to less than 1.0 M GuHCl. As with wild type, a sharp increase in FI 360/320 nm values on equilibrium refolding traces was due to light scattering by the aggregate (Figs. 2-5 and 2-6). Since the scattering would mask the presence of productively folded chains, the presence of native-like protein in aggregation samples was tested after centrifugation at 12,000 rpm. The soluble protein remaining after centrifugation displayed fluorescence emission spectra consistent with the native state spectra of the mutants (data not shown). Therefore, these mutant proteins exhibit partitioning between productive refolding and aggregation as previously described for wild type (Kosinski-Collins and King 2003).

5. *Equilibrium unfolding and refolding of N-terminal domain mutants*

The amino acids from the N-td that contribute to the interface hydrophobic cluster are Met43, Phe56 and Ile81 (Fig. 2-1). Equilibrium unfolding/refolding analyses of the single alanine substitution mutants of these residues were performed in a manner analogous to that used for wild-type HγD-Crys. It was possible to fit the equilibrium unfolding/refolding of M43A with both a two- and a three-state model (Fig. 2-5). The two-state fit yielded midpoints of 2.8 M GuHCl for both transitions. The three-state fit yielded midpoints of 2.2 M GuHCl for the first transition and 2.9 M GuHCl for the second transition (Table 2-1). Similar to wild-type HγD-Crys, the residuals of the three-state fit were more random and of lower magnitude suggesting that the three-state model is a better representation of the data (data not shown). Assuming a three-state mechanism, the mutation had minimal effect on both the native to intermediate and intermediate to unfolded transitions.

The equilibrium unfolding/refolding transitions of F56A were significantly different from wild type (Fig 2-5). A noticeable plateau was present in the transition region from approximately 2.0 to 2.3 M GuHCl suggesting greater population of the partially folded intermediate. The intermediate species had a fluorescence signal that was distinct from that of the native and unfolded conformations (Fig 2-5). Equilibrium unfolding/refolding data were best fit to a three-state model with transition midpoints of 1.6 and 2.9 M GuHCl for the first and second transitions, respectively (Table 2-1).

Similar to F56A, the mutant I81A also displayed a plateau from 2.0 to 2.3 M GuHCl where the fluorescence emission spectrum was different from both the native and unfolded states (Fig. 2-5). The unfolding/refolding transitions were best fit to a three-state model with a transition midpoint of 1.5 M GuHCl for the transition from native to intermediate and a midpoint of 2.9 M GuHCl for the intermediate to unfolded transition (Table 2-1).

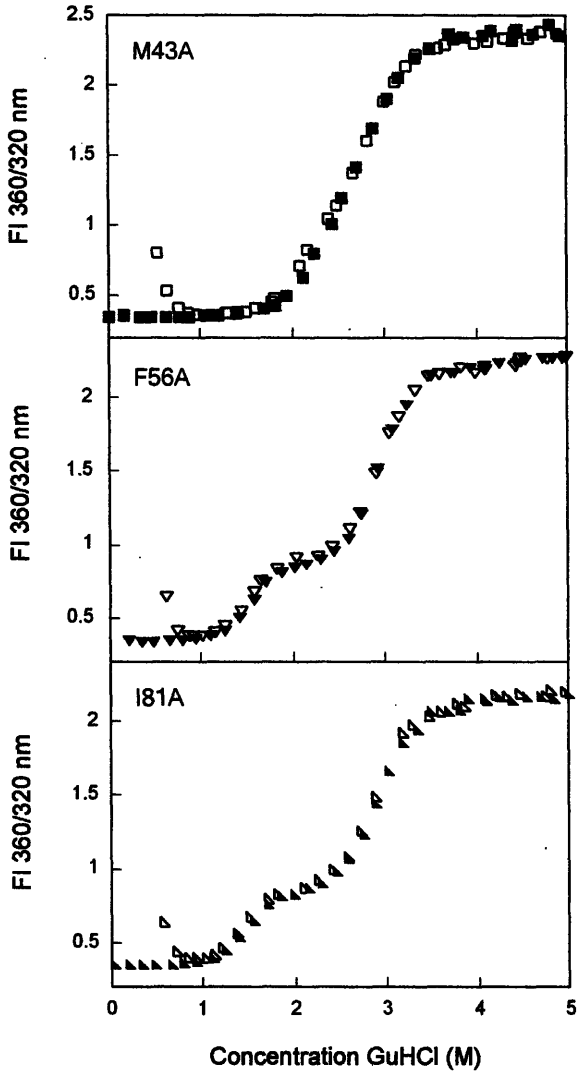

Figure 2-5. Equilibrium unfolding (closed symbols) and refolding (open symbols) of the N-terminal domain mutants M43A (■), F56A (▼), and I81A (◣) as probed by fluorescence emission. Protein was present at 10 μg/ml in 10 mM sodium phosphate, 5 mM DTT, 1 mM EDTA, pH 7.0 and GuHCl from 0 to 5.5 M. All samples were allowed to equilibrate for 24 hours at 37°C before recording fluorescence emission. Data was analyzed by fluorescence intensity at 360/320 nm using an excitation wavelength of 295 nm. Transitions for all mutants were best fit by a three-state model.

6. *Equilibrium unfolding and refolding of C-terminal domain mutants*

The amino acids of the C-td that participate in the interface hydrophobic cluster are Val132, Leu145 and Val170 (Fig. 2-1). The mutant protein V132A displayed transitions similar to those described for F56A and I81A above (Fig. 2-6). The unfolding/refolding transitions were best fit to a three-state model with a partially folded intermediate populated in the region of 2.3 M GuHCl. A transition midpoint of 1.3 M GuHCl was calculated for the native to intermediate transition and a midpoint of 2.5 M GuHCl was calculated for the intermediate to unfolded transitions, respectively (Table 2-1).

Equilibrium unfolding/refolding of L145A also displayed a three-state transition where the intermediate was populated at 2.3 M GuHCl (Fig. 2-6). The native to intermediate transition had a midpoint of 1.6 M GuHCl and the intermediate to unfolded transition had a midpoint of 2.6 M GuHCl (Table 2-1). The partially folded intermediate had a unique fluorescence spectrum similar to the intermediate conformation populated by all other domain interface mutants (data not shown).

The equilibrium unfolding/refolding transitions of V170A were best fit to a two-state transition with a random pattern of residuals (Fig. 2-6). The two transitions overlaid identically and displayed transition midpoints of 2.5 M GuHCl (Table 2-1). Unlike wild type and M43A, it was not possible to fit the transitions of V170A to a three-state model. However, it is not possible to rule out a three-state mechanism based on this observation alone. Further analysis will be performed to elucidate the unfolding/refolding mechanism of this mutant.

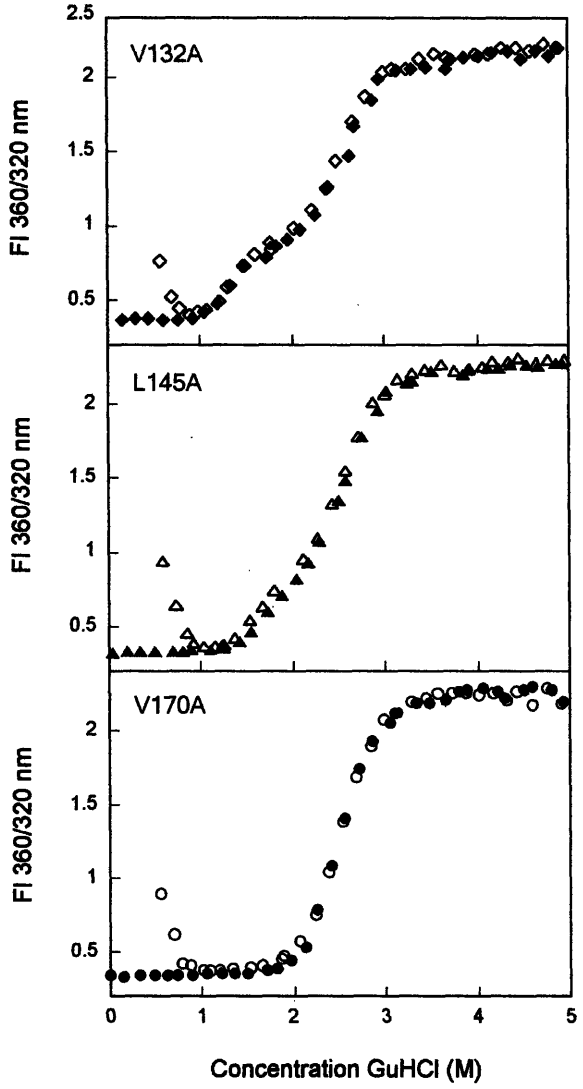

Figure 2-6. Equilibrium unfolding (closed symbols) and refolding (open symbols) of the C-terminal domain mutants V132A (♦), L145A (▲) and V170A (●). Fluorescence spectroscopy was used to probe the solvent accessibility of tryptophans in each sample using an excitation wavelength of 295 nm. Data was analyzed by fluorescence intensity at 360/320 nm. Protein was present at 10 μg/ml in 10 mM sodium phosphate, 5 mM DTT, 1 mM EDTA, pH 7.0 and GuHCl from 0 to 5.5 M. Samples were incubated at 37°C for 24 hours prior to measuring fluorescence emission. Transitions of all mutants, except V170A, were best fit by a three-state model. The transitions of V170A were best by a two-state model.

7. Productive refolding kinetics of wild type

In order to assess the role of hydrophobic domain interface residues in kinetic refolding, structural transformations of the mutant proteins were monitored over time using fluorescence as a probe of conformation. To allow for comparison, these experiments were performed similarly to our previous analyses of kinetic refolding of HγD-Crys (Kosinski-Collins et al. 2004). Unfolded proteins were diluted from 5.5 to 1.0 M GuHCl and the decrease in fluorescence intensity at 350 nm was monitored to follow burial of tryptophan residues. A syringe injection port was used instead of a stopped-flow apparatus since the major structural transformations of HγD-Crys occur on a second timescale (Kosinski-Collins et al. 2004). These experiments did not address folding intermediates that may have been populated on a sub-second timescale.

Previous analysis of wild-type HγD-Crys showed that refolding kinetics were best fit to two exponentials suggesting an intermediate was populated on the kinetic refolding pathway (Kosinski-Collins and King 2003; Kosinski-Collins et al. 2004). The partially folded intermediate was more fluorescent than the native state and less fluorescent than the unfolded state at 350 nm. The transition from unfolded to intermediate occurred with a half-time ($t_{1/2}$) of 15 seconds and the transition from intermediate to native occurred with a $t_{1/2}$ of 190 seconds (Kosinski-Collins et al. 2004).

Triple tryptophan mutant proteins were used to further analyze the structural transformations corresponding to the two kinetic fits of wild-type HγD-Crys (Kosinski-Collins et al. 2004). As mentioned previously, wild-type HγD-Crys has four intrinsic tryptophans, two per domain, buried in the hydrophobic cores. Triple tryptophan mutants were constructed where three of the endogenous tryptophans were substituted with phenylalanines. The four mutant proteins retained one endogenous tryptophan so that unfolding/refolding of the N-td and C-td could be followed independently of each other. This technique has been previously employed to elucidate the folding pathways of cellular retinoic acid binding protein I and phosphoglycerate kinase (Beechem et al. 1995; Sherman et al. 1995; Clark et al. 1998).

Kinetic refolding of mutant proteins retaining a tryptophan in the C-td were best fit to a three-state model (two exponentials) with $t_{1/2}$ values of 30 seconds for the first

transition and 150-300 seconds for the second transition. In contrast, kinetic refolding of mutant proteins retaining a tryptophan in the N-td were best fit to a two-state model with $t_{1/2}$ values of approximately 200 seconds. This data suggests that tryptophans in the C-td were buried first, followed by burial of tryptophans in the N-td.

Productive refolding experiments of wild-type HγD-Crys were repeated here to ensure the measurements were comparable to those reported previously. Kinetic refolding of wild type was best fit by two exponentials where a partially folded intermediate (I) was populated on the productive refolding pathway (Figs. 2-7 and 2-8).

$$\text{Denatured} \xrightarrow{k_1} [\text{I}] \xrightarrow{k_2} \text{Native}$$

In excellent agreement with previous results, a $t_{1/2}$ of 15 seconds was calculated for the transition from unfolded to intermediate and a $t_{1/2}$ of 190 seconds was calculated for the transition from intermediate to native (Kosinski-Collins et al. 2004). The kinetic refolding parameters calculated for wild-type and mutant HγD-Crys are reported in Table 2-2.

Productive kinetic refolding experiments were performed on interface mutants in a manner analogous to that described for wild-type HγD-Crys. Kinetic refolding curves for all mutant proteins were best fit to two exponentials. The kinetic fits in the early part of the curve were not particularly good (data not shown). While the fits were improved with the inclusion of additional exponentials, it was unclear whether these additional variables represented actual intermediates or were due to experimental, instrumental or human error. From the triple tryptophan studies, the spectroscopic changes during kinetic refolding of wild-type HγD-Crys were shown to correspond to sequential domain refolding (Kosinski-Collins and King 2003). Since fitting the data to two exponentials yielded phases that could be defined in terms of major structural transformations, discussion of the refolding kinetics has been limited to these clearly defined intermediates.

Table 2-2. Kinetic refolding parameters for wild-type and mutant HγD-Crys.

Protein	k_1 (seconds^{-1})	$t_{1/2}$ (seconds)	k_2 (seconds^{-1})	$t_{1/2}$ (seconds)
Wild type	0.048 ± 0.001	15 ± 1	0.0037 ± 0.0001	190 ± 10
M43A	0.009 ± 0.001	79 ± 8	0.0004 ± 0.0001	1700 ± 20
F56A	0.030 ± 0.002	23 ± 1	0.0002 ± 0.0001	2700 ± 320
I81A	0.034 ± 0.001	21 ± 1	0.0003 ± 0.0001	2100 ± 80
V132A	0.016 ± 0.002	45 ± 6	0.0005 ± 0.0001	1400 ± 80
L145A	0.014 ± 0.001	52 ± 5	0.0004 ± 0.0001	1600 ± 240
V170A	0.018 ± 0.001	38 ± 1	0.0026 ± 0.0004	300 ± 50

8. *Productive refolding kinetics of N-terminal domain mutants*

Kinetic refolding of the mutant protein M43A was best fit by a three-state model (two exponentials) suggesting the population of a partially folded intermediate similar to wild type (Fig. 2-7). Upon dilution out of GuHCl, the fluorescence intensity of M43A at 350 nm rapidly decreased with a $t_{1/2}$ of 79 seconds (Table 2-1). This change presumably corresponded to a transition from the unfolded to intermediate state. The $t_{1/2}$ value for M43A was less than that calculated for the first transition of wild-type HγD-Crys. After the initial phase, a slower decrease in fluorescence was observed with a $t_{1/2}$ of 1700 seconds. This loss of fluorescence correlated with a transition from the intermediate to the native state. The $t_{1/2}$ value for M43A differed by an order of magnitude from that calculated for the second transition of wild type.

Kinetic refolding of F56A was also best fit by a three-state model. An initial rapid decrease in fluorescence intensity with a $t_{1/2}$ of 23 seconds was followed by a slower decrease in fluorescence with a $t_{1/2}$ of 2700 seconds (Fig. 2-7). By inspection, the transition from unfolded to intermediate for F56A was indistinguishable from that of wild

type and had a $t_{1/2}$ value similar to that calculated for wild type. In contrast, the second transition from intermediate to native did not overlay with wild type and had a $t_{1/2}$ value more than 14 times greater.

I81A also underwent kinetic refolding that was best described by two exponentials. An initial rapid decrease in fluorescence occurred with a $t_{1/2}$ of 21 seconds and was followed by a slower phase that occurred with a $t_{1/2}$ of 2100 seconds (Fig. 2-7). Similar to the results described for M43A and F56A, the transition from unfolded to intermediate was indistinguishable from wild type while the rate of transition from intermediate to native was significantly reduced.

9. *Productive refolding kinetics of C-terminal domain mutants*

Productive refolding kinetics of the mutant protein V132A were best fit to a three-state model (Fig. 2-8). The initial rapid decrease in fluorescence occurred with a $t_{1/2}$ of 45 seconds (Table 2-2). This value differs slightly from that calculated for wild-type HγD-Crys but is still within the same order of magnitude. In contrast, the rate of transition from intermediate to native was greatly reduced compared to that of wild type. The second transition of V132A occurred with a $t_{1/2}$ of 1400 seconds (Table 2-2).

Kinetic refolding of L145A proceeded in a manner identically to that described for V132A and was best fit to a three-state model (Fig. 2-8). The first transition had a $t_{1/2}$ of 52 seconds and the second transition a $t_{1/2}$ of 1600 seconds. The $t_{1/2}$ value for unfolded to intermediate was similar to that for wild type while the $t_{1/2}$ value for intermediate to native was appreciably larger (Table 2-2).

Of all mutants examined here, refolding kinetics of V170A were the most similar to wild-type HγD-Crys. Changes in fluorescence intensity at 350 nm during refolding overlaid that of wild type almost identically (Fig. 2-8). The initial transformation from unfolded to intermediate occurred with a $t_{1/2}$ of 38 seconds and the second transition to native occurred with a $t_{1/2}$ of 300 seconds (Table 2-2).

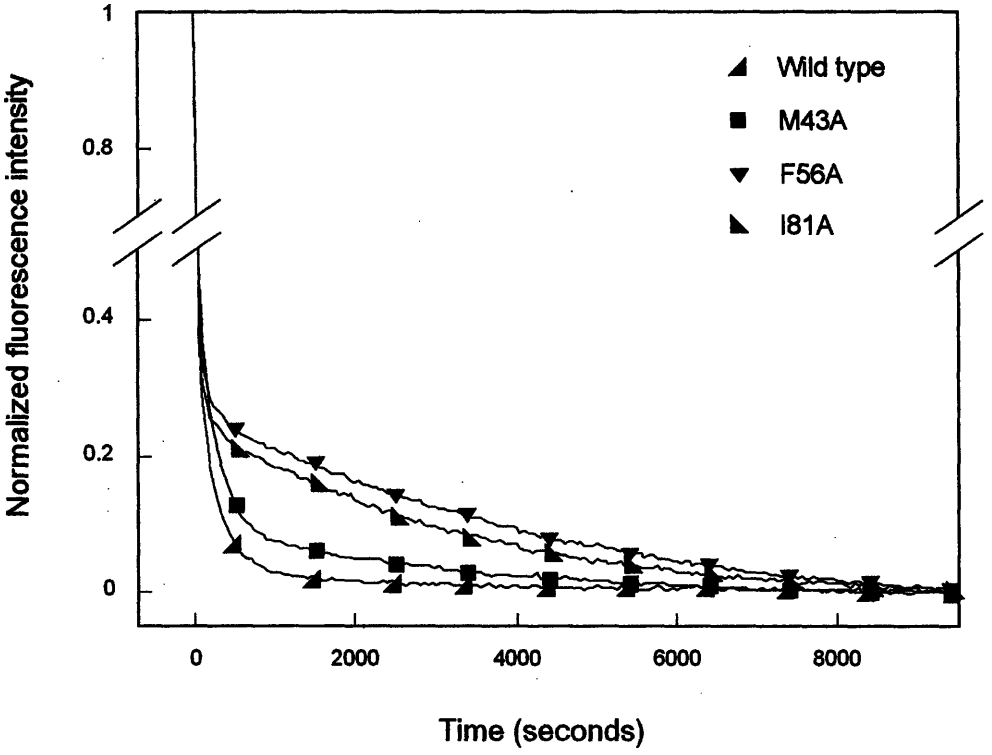

Figure 2-7. Productive kinetic refolding of wild-type (◢) HγD-Crys and N-terminal domain mutants M43A (■), F56A (▼), and I81A (◣). Black lines represent data points taken every 1 second with symbols included for ease of viewing. Fluorescence emissions were normalized for ease of viewing and comparison. Wild-type and mutants proteins were initially unfolded for three hours at a concentration of 100 μg/ml in 5.5 M GuHCl, 37 C. Refolding was initiated by dilution of unfolded proteins into 10 mM sodium phosphate, 5 mM DTT, and 1 mM EDTA, pH 7.0 using a syringe injection port. The final concentration of GuHCl in refolding samples was 1.0 M and the final protein concentration was 10 μg/ml. The temperature was maintained at 37°C during refolding using a circulating water bath. Refolding was monitored by changes in fluorescence emission at 350 nm for three hours. The data were fit to two exponentials to calculate rate constants. Fits were improved by inclusion of additional exponentials; however, it was unclear whether these additional variables actually represented population of further intermediates.

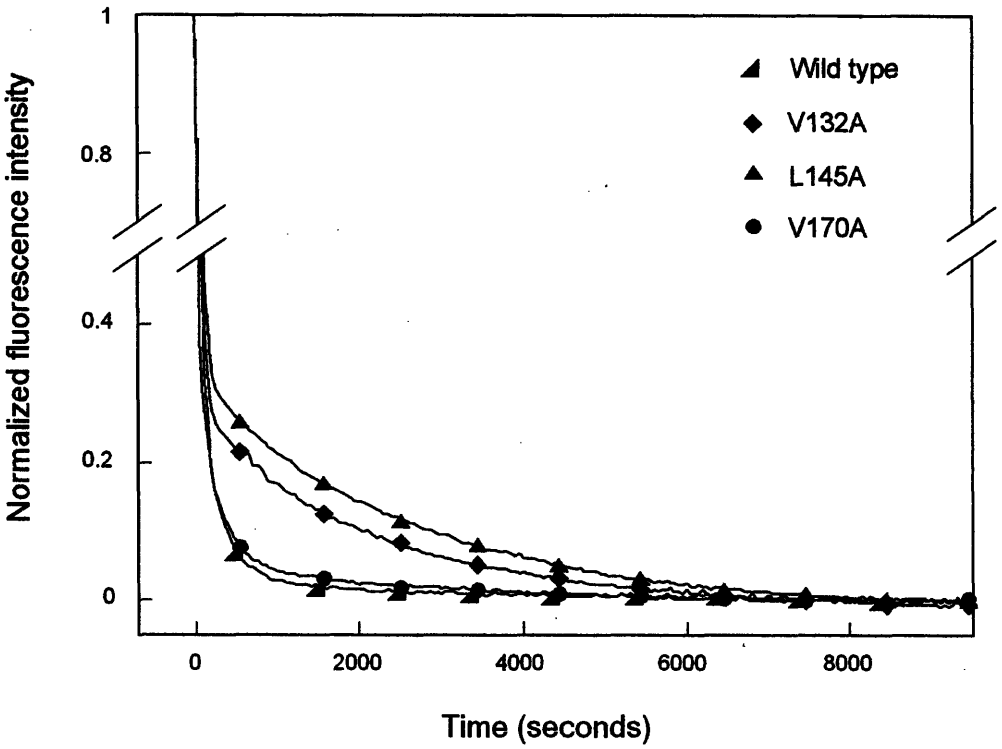

Figure 2-8. Productive kinetic refolding of wild type (◢) HγD-Crys and C-terminal domain mutants V132A (◆), L145A (▲) and V170A (●). Black lines represent data points taken every 1 second with symbols included for ease of viewing. Fluorescence emissions were normalized for ease of viewing and comparison. The proteins were first unfolded at 100 μg/ml in 5.5 M GuHCl, 37 C for three hours. Refolding was initiated by dilution of unfolded proteins into 10 mM sodium phosphate, 5 mM DTT, and 1 mM EDTA, pH 7.0 using a syringe injection port, to give a final concentration of 1.0 M GuHCl and 10 μg/ml protein. Structural changes during refolding were monitored by changes in fluorescence emission at 350 nm for three hours. The temperature was maintained at 37°C during refolding using a circulating water bath. The data were fit to two exponentials to calculate rate constants. Inclusion of extra exponentials improved the fits. It was not possible to determine if the extra exponentials characterized the genuine population of further intermediates.

E. DISCUSSION

HγD-Crys is two-domain protein of the eye lens that must remain stable for an entire human lifetime without the possibility of regeneration. Along with other lens crystallins, HγD-Crys is found in the insoluble aggregates associated with mature-onset cataract. HγD-Crys is composed of two domains that share approximately 34% sequence identity and adopt highly similar, primarily β-sheet folds (Fig. 2-1). The two domain nature of HγD-Crys and the other β- and γ-crystallins is hypothesized to have arisen from a gene duplication event (Wistow et al. 1983).

The domains of HγD-Crys interact non-covalently through interdomain side chain contacts forming two structurally distinct regions (Basak et al. 2003). These include a highly conserved hydrophobic cluster of Met43, Phe56, Ile81, Val132, Leu145, and Val170 and polar peripheral pairwise interactions between Gln54/Gln143 and Arg79/Met147 (Fig. 2-1). The domain interface of HγD-Crys is relatively similar in amino acid composition to domain interfaces of other proteins. Overall, the amino acid composition of domain interfaces more closely resembles that of protein surfaces rather than protein cores (Jones et al. 2000). However, hydrophobic residues are still highly prevalent in both inter- and intra-chain domain interactions (Jones et al. 2000).

1. *Differential domain stability of the β- and γ-crystallins*

Many of the β- and γ-crystallins studied to date exhibit differential domain stability. Bovine γB-crystallin (BγB-Crys) displayed a three-state transition in urea at pH 2.0, 20°C (Mayr et al. 1997). At pH 2.0, the isolated N-td of BγB-Crys was much more stable than the C-td presumably due to the abundance of positively charged residues on the surface of the C-td at this acidic pH. At pH 7.0 the domains had similar stabilities and equilibrium unfolding of the full-length protein was two-state (Mayr et al. 1997). Bovine βB2-crystallin also displayed differential domain stability (Wieligmann et al.

1999). The crystal structure of βB2-crystallin is a domain-swapped dimer where the N-td of one monomer pairs with the C-td of the other (Bax et al. 1990). The isolated C-td of bovine βB2-crystallin was significantly more stable than the isolated N-td (Wieligmann et al. 1999).

Studies of triple tryptophan mutants of HγD-Crys suggest that the C-td is more stable than the N-td (Kosinski-Collins et al. 2004). The transition midpoint of proteins retaining tryptophans in the N-td was 1.3 M GuHCl while the midpoint of mutants retaining tryptophans in the C-td was 2.0 M GuHCl. These data suggest that an intermediate would be populated during equilibrium unfolding/refolding of the wild-type protein. Previous analyses using an incubation time of six hours could not distinguish a three-state transition for wild-type HγD-Crys (Kosinski-Collins and King 2003; Kosinski-Collins et al. 2004). In the experiments described here, the incubation time was increased to 24 hours which altered the equilibrium transition and eliminated the unfolding/refolding hysteresis (Fig. 2-4). At these extended equilibration times, the transition had reduced cooperativity which may reflect a more complex transition such as would be expected for a three-state mechanism. When fit to a three-state model the first transition had a midpoint of 2.2 M GuHCl and the second transition a midpoint of 2.8 M GuHCl. Overall, the three-state model was a better description of the data than the two-state model, especially in the region of 2.0 to 2.5 M GuHCl. From these observations, we hypothesize that wild-type HγD-Crys populates a partially folded intermediate in equilibrium unfolding/refolding experiments at approximately 2.3 M GuHCl. Based on the stabilities of the domains elucidated by triple tryptophan mutants, we hypothesize that these two transitions correspond to unfolding/refolding of the N-td (at lower concentrations of GuHCl) and the C-td (at high concentrations of GuHCl). The transition midpoints of triple tryptophan mutant proteins were considerably lower than these values potentially due to loss destabilizing effects of the triple mutations.

2. *Domain interface interactions are crucial for stability*

Contribution of domain interface interactions to the stability of BγB-Crys has been previously studied (Palme et al. 1997; Palme et al. 1998). The domain interface of BγB-Crys is comprised of a central hydrophobic cluster including a phenylalanine at position 56 and peripheral pairwise interactions (Wistow et al. 1983). Palme et al. mutated Phe56 to alanine, aspartate or tryptophan and analyzed the effects on the structure and stability of BγB-Crys in urea at pH 2 and 20 °C (1997; 1998). All proteins displayed reduced stability of the C-td that varied with the mutation. Substitution with aspartate or alanine resulted in considerable destabilization while substitution with tryptophan had less of an affect (Palme et al. 1997). Domain core structures of the mutants were indistinguishable from the wild-type protein and local structure around residue 56 was unchanged (Palme et al. 1998). These results suggest that, despite the destabilizing effects, the global structure of BγB-Crys was too rigid to adjust to the altered size or hydrophobicity of the mutations.

Similar to the results described above for BγB-Crys, the hydrophobic domain interface residues of HγD-Crys are also critical for stability. All mutants except V170A were best fit to a three-state model where the first transitions likely corresponded to unfolding/refolding of the N-td and the second transition to unfolding/refolding of the C-td. According to this hypothesis, C-td unfolding/refolding for the mutants F56A, I81A, V132A, and L145A, occurred with a midpoint between 2.5 and 2.9 M GuHCl. These values are relatively similar to that calculated for the second transition of wild type (2.8 M GuHCl) further supporting a three-state mechanism for the wild-type protein. Substitution of Met43, Phe56 and Ile81 from the N-td resulted in an increased midpoint of 2.9 M GuHCl for transitions of the C-td. If the intermediate had a folded C-td and unfolded N-td as hypothesized, interface residues from the N-td would not be expected to stabilize the intermediate. Increased stability of the intermediate as is seen with M43A, F56A and I81A may be due to a favorable decrease in solvent exposed hydrophobics compared to wild type because of the alanine substitutions. In contrast, mutation of Val132 and Leu145 resulted in decreased stability of the intermediate. The domain

interface of the C-td is likely structured in the intermediate conformation. Consequently, mutations that disrupt correct intra-domain hydrophobic packing would be expected to decrease stability of the intermediate.

In contrast to the marginal affect on stability of the C-td, the N-td was significantly destabilized by mutations of Phe56, Ile81, Val131 and Leu145. For these mutants, unfolding/refolding of the N-td occurred with a midpoint between 1.3 and 1.6 M GuHCl (Table 2-1). These values were significantly lower than that calculated for the first transition of wild type when fit to a three-state model (2.2 M GuHCl). Therefore, mutation of residues located in the C-td affected stability of the N-td but not vice versa. From these results we postulate that the stability of the N-td is dependent on correct domain interface contacts while stability of the C-td is not enhanced by domain pairing. Destabilization of the N-td resulted in population of the intermediate over a greater range of GuHCl concentrations than was seen for the wild-type protein.

Mutation of Met43 and Val170 had significantly different effects on the stability of HγD-Crys. The equilibrium unfolding/refolding transitions of M43A did not differ dramatically from wild type, and had ΔG_{H2O}, m-values and transition midpoints very similar to wild type. In contrast, the transitions of V170A were best fit to a two-state model with midpoints of 2.5 M GuHCl. This may suggest that unfolding/refolding of V170A occurs with a direct transition between the native and unfolded states. An alternative interpretation is that the mutant did unfold/refold by a three-state mechanism, but that the two transitions were not discernible. This may have been caused by a shift of the first transition to higher concentrations of GuHCl or shift of the second transition to lower concentrations of GuHCl. One of these phenomena or a combination of the two would effectively merge the transitions. From comparison to the transitions of wild type, we hypothesize that mutation of Val170 had both effects described above. That is, the first transition was stabilized relative to wild type and the second transition was destabilized relative to wild type. Further experiments will be done determine which, if either of these hypotheses can explain this data.

3. *Kinetic refolding pathway of HγD-Crys*

We have previously studied the productive refolding pathway of HγD-Crys using triple tryptophan mutant proteins (Kosinski-Collins et al. 2004). Previous investigations of the β- and γ-crystallins analyzed behavior of the domains by studying polypeptides that corresponded to isolated N- and C-td's of the proteins (Sharma et al. 1990; Mayr et al. 1994; Wieligmann et al. 1999; Wenk et al. 2000). This approach does not reveal properties of the domains in context of the full-length proteins. This limitation has been circumvented in our analysis of triple tryptophan mutants since we were able to independently follow folding of the two domains while still maintaining a full-length protein.

During kinetic refolding of HγD-Crys the tryptophans of the C-td were buried before those of the N-td (Kosinski-Collins et al. 2004). These results suggest a productive refolding pathway where the C-td refolds first followed by the N-td. Given that the domain interface of HγD-Crys does not contain any fluorescent amino acids, it was impossible to determine when the domain interface became structured in these experiments.

Three simple models are illustrated in Figure 2-9 to describe the potential role of domain interface residues in the sequential domain refolding pathway of HγD-Crys. In the first model, the domain interface residues initially collapse and form a nucleating center for refolding of the C-td and subsequently the N-td. If HγD-Crys refolded by this first pathway, domain interface mutations would be expected to decrease the refolding rates for both the C-td and the N-td. In the second model, the two domains refold independently and subsequently come together to form interface contacts. Mutations of domain interface residues would likely not reduce refolding rates of the N-td or the C-td

in this model. In the third model, the C-td refolds first and the interface amino acids of the C-td act as a nucleating center for refolding of the N-td. By this model, mutations of domain interface residues would likely result in decreased rates for refolding of the N-td only.

Kinetic refolding of all interface mutants was best fit to a three-state model similar to wild type. Comparing these data to the triple tryptophan mutant data, it was possible to correlate the two transitions to sequential refolding of the C-td and N-td (Kosinski-Collins et al. 2004). Assuming that the domains of the mutant proteins refolded in the same order as wild type, the mutants M43A, F56A, I81A, V132A, and L145A all had similar $t_{1/2}$ values for refolding of the C-td but greatly increased $t_{1/2}$ values for refolding of the N-td. These results suggested a productive refolding pathway consistent with model three described above in which the interface residues of the refolded C-td act as a nucleating center for refolding of the N-td (Fig. 2-9).
From the perspective of *in vitro* domain folding, a sequential nucleation-based folding pathway as was exhibited by HγD-Crys may be considered detrimental for productive folding. This is because sequential domain folding results in an increased lifetime of partially folded intermediates that may partition into kinetically-trapped aggregates (Jaenicke 1999). The single folded domain conformer populated during equilibrium and kinetic unfolding/refolding experiments is an attractive target for the aggregation-prone species in the *in vitro* aggregation pathway of HγD-Crys (Kosinski-Collins and King 2003; Kosinski-Collins et al. 2004).

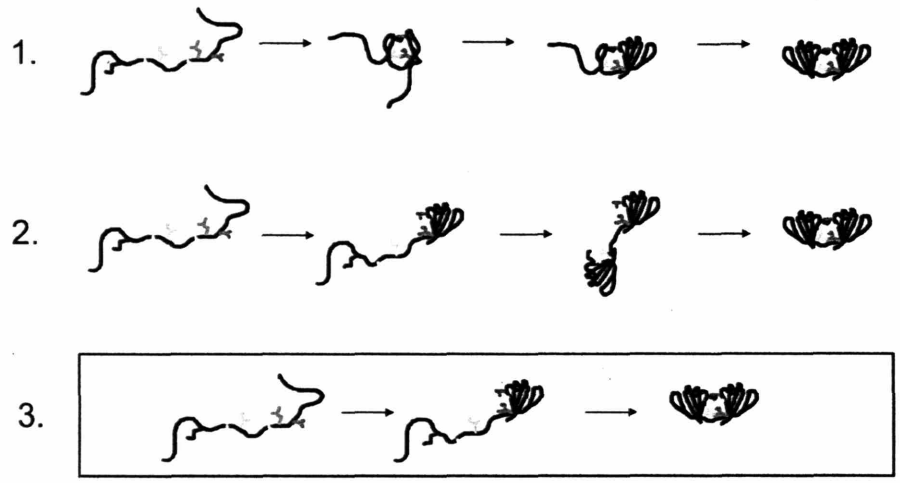

Figure 2-9. Schematic diagram of three models to describe the role of the domain interface during sequential domain refolding of wild-type HγD-Crys. Model three was consistent with the data presented here. The C-td of HγD-Crys refolded first, resulting in an intermediate that had an unpaired, solvent-exposed domain interface. This surface likely acted as a nucleating center for refolding of the N-td.

4. *Implications for understanding stability and oligomerization in the lens*

In this investigation the hydrophobic domain interface residues were substituted with alanine resulting in a loss of hydrophobic surface area. If residues of the cluster contributed uniformly to stability by hydrophobic burial, differences in effects of the mutations would be expected to correlate with buried accessible surface area of the wild-type residue (Rose et al. 1985; Zhou and Zhou 2004). Greater accessible surface area was buried for Phe56 than Val132, but the mutant protein V132A was more destabilized than the mutant F56A. This indicates that the locations of the hydrophobic domain interface residues are important in determining their role in stability and folding of HγD-Crys. This may reflect different packing densities around the residues. Additionally, unique identity and position of domain interface residues are thought to be important in determining the oligomeric states of the β- and γ-crystallins (Hope et al. 1994; Mayr et al. 1994; Trinkl et al. 1994).

The two domain β- and γ-crystallins comprise over 50% of the protein in the lens nucleus and adopt highly similar domain topology with domain interactions that are predominated by a central hydrophobic cluster (Lapatto et al. 1991; Slingsby 1997; Oyster 1999). Protein concentration in the lens nucleus is extremely high (200-400 mg/ml) necessitating precise control of folding and oligomerization in order to prevent aberrant intermolecular associations. Due to the hydrophobic nature of the domain interfaces and the ability of β- and γ-crystallins to adopt single folded domain conformers, the domain interfaces may be regions of the molecules that are particularly prone to incorrect protein-protein interactions. Given these factors, it is reasonable to assume that the β- and γ-crystallins may have evolved for specificity of domain interface residues in folding, stability and oligomerization in order to prevent incorrect domain interactions in the crowded lens nucleus.

5. *Aggregation and cataract*

It is currently unknown which conformations of crystallin proteins are the aggregation-prone species that lead to cataract. Crystallins in the lens nucleus are subject to a lifetime of oxidative and radiative stress. The crystallin proteins of both young and old lenses are covalently damaged as a result of these insults (Hoenders and Bloemendal 1983; Hanson et al. 1998; Hanson et al. 2000). Covalent damage may result in destabilization and partial unfolding into the aggregation-prone species that are precursors to cataract formation. These partially folded species may be sequestered by α-crystallin in order to prevent aggregation. In fact, it has been shown that partially structured conformations of β- and γ-crystallins are capable of binding to α-crystallin *in vitro* as a result of mutation or after exposure to heat (Lampi et al. 2002; Liang 2004; Sathish et al. 2004). We are currently testing whether α-crystallin is able to bind the single folded domain conformer of HγD-Crys described here. The age-onset nature of non-congenital forms of cataract may reflect accumulation of sufficient levels of damage to induce unfolding or saturation of α-crystallin.

CHAPTER THREE:

INTER-DOMAIN SIDE CHAIN INTERACTIONS IN HUMAN γD CRYSTALLIN INFLUENCING FOLDING AND STABILITY[2]

[2] **Reprinted from:** Flaugh, S.L., Kosinski-Collins, M.S., and King, J. (2005) Interdomain side chain interactions in human γD-crystallin influencing folding and stability. *Protein Sci.* 14:2030-2043.

A. ABSTRACT

Human γD crystallin (HγD-Crys) is a two domain, β-sheet eye lens protein that is synthesized early in lens development and must remain soluble throughout life. Single amino acid substitutions of HγD-Crys are associated with juvenile-onset cataracts. Features of the interface between the two domains are conserved among γ-crystallins, and are likely important in folding and stability. The domain interface of HγD-Crys includes a central hydrophobic cluster and two pairs of interacting residues peripheral to the cluster, Gln54/Gln143 above and Arg79/Met147 below. We previously reported the contribution of the hydrophobic cluster residues to protein stability (Flaugh et al. 2005a). In this study, single and double alanine substitutions of the peripheral residues were constructed to assess their contributions to folding and stability. Equilibrium unfolding/refolding experiments at 37°C revealed a plateau in the transitions at 2.3 M guanidine hydrochloride for the wild-type and mutant proteins, suggesting population of a partially folded intermediate with a folded C-terminal domain (C-td) and unfolded N-terminal domain (N-td). The N-td was destabilized by substituting residues from both domains. In contrast, the C-td was not significantly affected by substitutions of either domain. The Arg79/Met147 pair did not display favorable interaction energy, while the Gln54/Gln143 pair had a small interaction energy of 0.7 kcal*mol^{-1}. We suggest that these residues act as structural boundaries shielding the central hydrophobic cluster from solvent. Refolding rates of the N-td were significantly decreased for mutants in both domains. In contrast, refolding rates for the C-td of mutants from both domains were similar to wild type. Therefore, domain interface residues of the folded C-td probably nucleate refolding of the N-td. Glutamine and methionine side chains are among the residues covalently damaged in aged and cataractous lenses. Damage to Gln54, Gln143 or Met147 may generate partially unfolded, aggregation-prone conformations of HγD-Crys that could be significant in cataract.

B. INTRODUCTION

The α-, β- and γ-crystallins are structural proteins of the vertebrate eye lens whose solubility and stability are required to maintain eye lens transparency throughout life. The β- and γ-crystallins act solely as structural proteins in the lens, while the α-crystallins are related to small heat shock proteins and probably have an additional *in situ* function as passive chaperones (Horwitz 1992; Boyle and Takemoto 1994). The α-crystallin subunits associate in the lens to form polydisperse, high molecular weight complexes (Haley et al. 1998; Aquilina et al. 2005).

Lens transparency is established via short-range ordering of natively folded crystallins in the fibrous cells of the lens (Delaye and Tardieu 1983; Fernald and Wright 1983). The protein concentration in these cells approaches 70% g/g wet weight, with the crystallins accounting for over 90% of the total protein (Oyster 1999). The major mass of crystallin proteins in the lens are expressed *in utero*, do not turnover during adulthood, and thus must remain soluble and stable in the continued presence of environmental stresses for a lifetime. The crystallins in aged and cataractous lenses are covalently damaged, presumably as a result of age and exposure to radiative and oxidative stress (Hoenders and Bloemendal 1983). Covalent damage found in the crystallins of aged and cataractous lenses include non-native disulfide bonding, backbone cleavage, methionine oxidation and glutamine and asparagine deamidation (Hanson et al. 1998; Lampi et al. 1998); (Ma et al. 1998; Hanson et al. 2000; Lampi et al. 2001); (Lampi et al. 2002; Harms et al. 2004). The eye disease mature-onset cataract is associated with the aggregation and condensation of such damaged crystallin proteins (Hoenders and Bloemendal 1983).

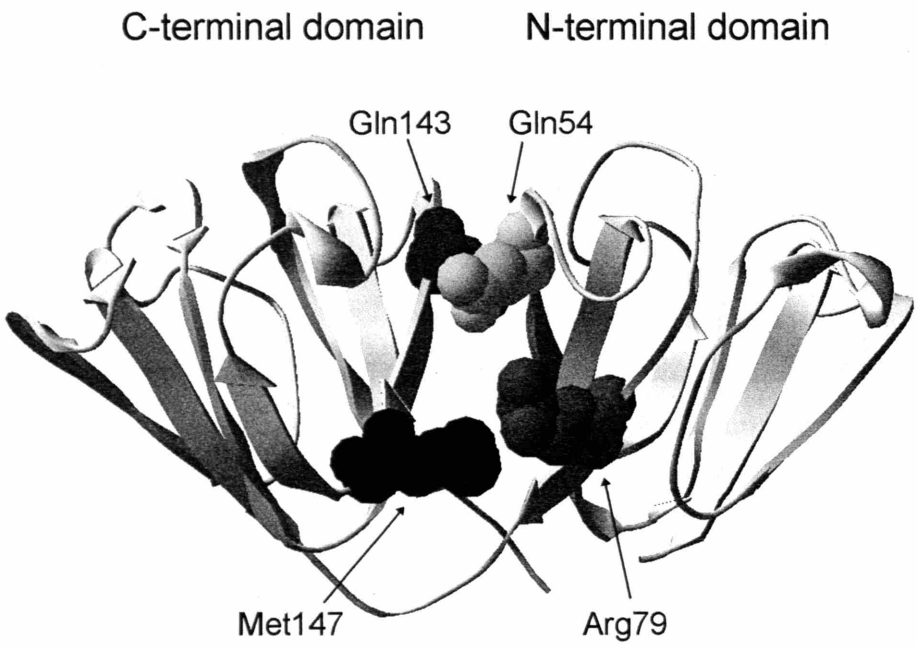

Figure 3-1. The crystal structure of wild-type HγD-Crys (Basak et al. 2003) shown in ribbon representation with the peripheral interface residues in spacefill (PDB code 1HK0).

The βγ-crystallin superfamily comprises greater than 50% of the crystallins in the vertebrate lens (Wistow et al. 1983; Slingsby 1997). Members of this superfamily exhibit a conserved two-domain, primarily β-sheet fold that likely arose from a gene duplication event (Lubsen et al. 1988; Wistow and Piatigorsky 1988; Norledge et al. 1996). The highly homologous N- and C-terminal domains are composed of two Greek key motifs each and exhibit pseudo-2-fold pairing. The domains are covalently linked by a six to eight amino acid linker and interact non-covalently through inter-domain side chain contacts. The γ-crystallins are monomeric, with the N- and C-terminal domains pairing intra-molecularly, while the β-crystallins associate into multimers (Wistow et al. 1983) (Bax et al. 1990; Slingsby and Bateman 1990). For instance, bovine βB2 crystallin forms dimers by domain swapping where the N- and C-terminal domains pair in a similar manner as in the γ-crystallins, except that paring occurs between domains from different subunits (Bax et al. 1990). Despite differences in oligomeric state, side chains that make contact across the domain interfaces of the β- and γ-crystallins play a critical role in intra- or inter-molecular domain interactions.

Human γD crystallin (HγD-Crys) is a monomeric, 173 amino acid protein and a member of the βγ-crystallin superfamily. HγD-Crys is present at highest concentrations in the lens nucleus, the region formed earliest in development. Mutations resulting in single amino acid substitutions in HγD-Crys are associated in juvenile-onset cataract (Stephan et al. 1999; Pande et al. 2001; Santhiya et al. 2002). The crystal structure of HγD-Crys has been solved to 1.25 Å resolution (Basak et al. 2003) and displays the βγ-crystallin superfamily's characteristic two-domain fold, with the domains interacting intra-molecularly (Fig. 3-1). The two domains of HγD-Crys are covalently joined by an extended six residue linker and the side chains of ten amino acids interact non-covalently across the domain interface. Buried in the center of the interface is a hydrophobic cluster of Met43, Phe56, and Ile81 from the N-terminal domain (N-td), and Val132, Leu145 and Val170 from the C-terminal domain (C-td). Additionally, there are pair wise interactions on the periphery of the hydrophobic cluster between Gln54 (N-td) and Gln143 (C-td) at the top of the interface, and Arg79 (N-td) and Met147 (C-td) nearer to the linker peptide (Fig. 3-1). These peripheral interactions flank the central hydrophobic cluster and

appear to shield it from solvent. Chemical properties of the amino acids in these interface positions are highly conserved among γ-crystallins from diverse organisms suggesting that they are critical for either folding, stability or solubility. An exception to this is Met147 in the human sequence, as most γ-crystallins have an arginine in this position. The amino acids Gln54, Gln143 and Met147 are all potential sites of covalent damage in aged lenses.

Folding and stability of the β- and γ-crystallins have been extensively studied and in general the γ-crystallins display higher intrinsic stability than the β-crystallins (Bloemendal et al. 2004). During refolding, rat βB2 crystallin, bovine γB crystallin (BγB-Crys) and HγD-Crys display competing aggregation reactions (Rudolph et al. 1990; Jaenicke 1999; Kosinski-Collins and King 2003). Aggregation may have also been observed but not reported for other β- and γ-crystallins. The domain stability of several β- and γ-crystallins have also been studied both in the context of the full-length proteins as well as in isolation (Rudolph et al. 1990; Sharma et al. 1990; Mayr et al. 1997; Wieligmann et al. 1999). The domains of these proteins are often less stable in isolation than when paired with their partner domain in the full-length proteins.

We previously investigated wild-type HγD-Crys unfolding/refolding in near-physiological conditions (37°C, and phosphate buffer, pH 7.0) to best relate *in vitro* and *in vivo* properties (Kosinski-Collins and King 2003; Kosinski-Collins et al. 2004; Flaugh et al. 2005a). Refolding to concentrations of GuHCl below 1.0 M revealed an aggregation pathway that competed with productive refolding, which may provide an *in vitro* model for the involvement of HγD-Crys in cataract (Kosinski-Collins and King 2003). In productive kinetic refolding to 1.0 M GuHCl, HγD-Crys exhibited a sequential domain refolding pathway (Fig. 3-2) where the C-td refolded first followed by the N-td (Kosinski-Collins et al. 2004). Subsequently, a partially-folded intermediate was detected in equilibrium unfolding/refolding experiments that likely had a folded C-td and unfolded N-td (Flaugh et al. 2005a). Since partially folded species are often involved in off-pathway aggregation, the single folded domain species is an attractive candidate for the species that aggregates during refolding (Fig. 3-2).

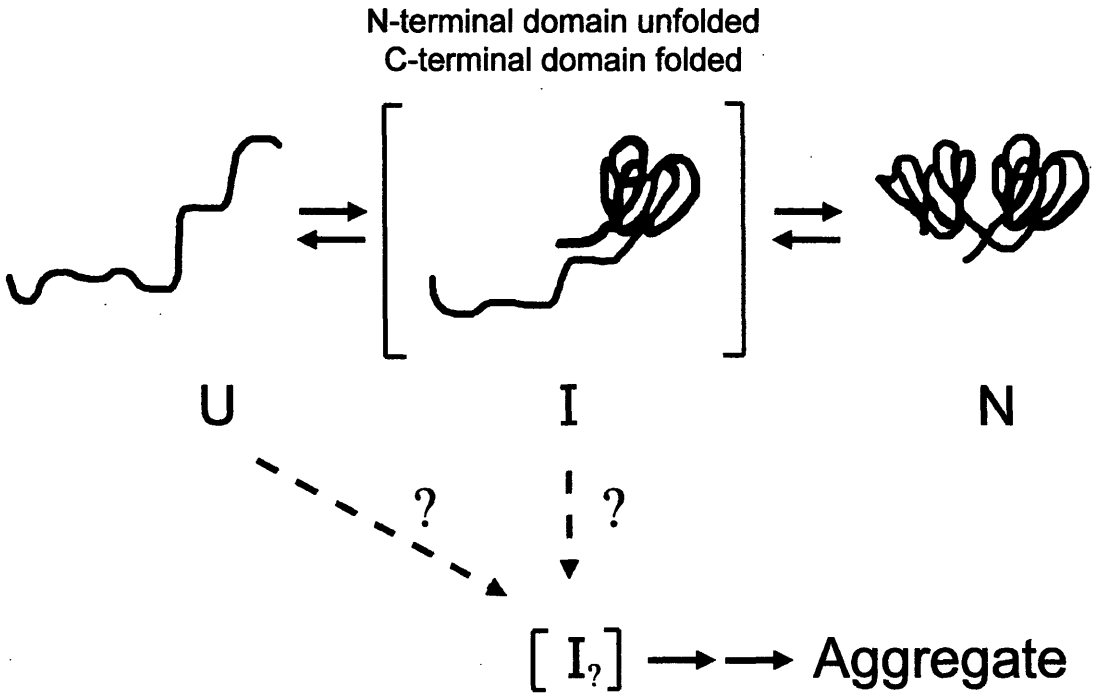

Figure 3-2. A schematic illustration of the in vitro kinetic refolding pathway of HγD-Crys. The C-td refolds first followed by the N-td (Kosinski-Collins et al., 2004). The conformation of the aggregation-prone intermediate is currently unknown, but may be related to the single folded domain conformer.

Given that the two domains of HγD-Crys refold sequentially and display differential stability, association of the residues that make contact across the domain interface may be critical in folding and stability. This has been confirmed by a mutational study of hydrophobic domain interface residues of HγD-Crys (Flaugh et al. 2005a). Several site directed alanine mutants of the hydrophobic domain interface residues displayed significantly reduced refolding rates for the N-td but not the C-td. Additionally, these mutants also displayed reduced midpoints for unfolding/refolding of the N-td (Flaugh et al. 2005a). Similarly, mutating domain interface residues of BγB-Crys and human βB1 crystallin (HβB1-Crys) destabilizes the proteins (Palme et al. 1997; Kim et al. 2002).

In this study we use site-specific mutagenesis to study the two pairs of peripheral interface residues in HγD-Crys that flank the hydrophobic cluster, Gln54/Gln143 and Arg79/Met147. Similar to results of the hydrophobic interface residues, substituting these residues with alanine caused selective destabilization of the N-td and a decrease in the rate of refolding for the N-td but not the C-td.

C. MATERIALS AND METHODS

1. *Mutagenesis, expression and purification of recombinant HγD-Crys*

Site-directed single alanine substitutions of residues Gln54, Arg79, Gln143, and Met147 were constructed using site-directed mutagenesis. Primers encoding the respective alanine substitutions (IDT-DNA) were used to amplify a pQE.1 plasmid encoding the HγD-Crys gene with an N-terminal 6-His tag (Kosinski-Collins et al. 2004). The double mutation Q54A/Q143A was constructed by using the primer encoding the Q143A substitution to amplify the Q54A mutant plasmid. The double mutant R79A/M147A was created in an analogous manner using the M147A primer with the R79A plasmid. DNA sequencing of all constructs was performed to verify the substitutions and to ensure no additional mutations were present (Massachusetts General Hospital).

Recombinant wild-type and mutant HγD-Crys proteins were prepared as previously described (Kosinski-Collins et al. 2004). Briefly, proteins extracted from *E. coli* cell lysates were purified to over 98% homogeneity by affinity chromatography with a Ni-NTA resin (Qiagen).

2. *Circular dichroism and fluorescence spectroscopy*

CD spectra of recombinant mutant proteins were collected with an AVIV model 202 CD spectrometer (Lakewood, NJ) using a 0.25 cm pathlength cuvette. The temperature was maintained at 37 °C with an internal Peltier thermo-electric temperature controller. The buffer conditions for all experiments were 10 mM sodium phosphate, 5 mM DTT, 1 mM EDTA, pH 7.0. Protein was present at 100 μg/ml. Protein concentrations were determined by absorbance at 280 nm using an extinction coefficient of 41,040 cm^{-1} M^{-1} for wild type and mutant His-tagged proteins. Spectra were collected from 200 to 260 nm to monitor secondary structure and the buffer signal was subtracted from all spectra.

Fluorescence emission spectra were collected at 37 °C using a Hitachi F-4500 fluorimeter equipped with a circulating water bath to control temperature. Intrinsic tryptophan fluorescence was measured in the range of 310 to 420 nm using an excitation wavelength of 295 nm. All samples contained 10 μg/ml purified protein in 10 mM sodium phosphate, 5 mM DTT, 1 mM EDTA, pH 7.0, and GuHCl where appropriate. All spectra were corrected for the buffer baseline.

3. *Equilibrium unfolding/refolding*

Equilibrium unfolding experiments were carried out by diluting purified mutant proteins to 10 μg/ml into increasing concentrations of GuHCl from 0 to 5.5 M (purchased as an 8.0 M solution from Sigma-Aldrich, Saint Louis, MO). Buffer conditions for all unfolding samples were 10 mM sodium phosphate, 5 mM DTT, 1 mM EDTA, pH 7.0. The unfolding samples were incubated at 37 °C for 24 hours, by which time equilibrium had been reached.

Equilibrium refolding experiments were carried out by initially preparing an unfolded stock solution containing 100 μg/ml purified protein in 5.5 M GuHCl. The unfolded stock solution was incubated at 37 °C for six hours and then diluted into refolding buffer to a final protein concentration of 10 μg/ml. Refolding buffer contained 10 mM sodium phosphate, 5 mM DTT, 1 mM EDTA, pH 7.0, and GuHCl from 0.55 to 5.5 M. The refolding samples were allowed to reach equilibrium by incubation at 37 °C for 24 hours.

Fluorescence emission spectra were recorded for all unfolding/refolding samples using an excitation wavelength of 295 nm and monitoring emission from 310 to 420 nm. GuHCl concentrations were determined by measuring the refractive indexes of all samples. Data was analyzed by plotting concentration of GuHCl versus fluorescence intensity at 360 nm divided by fluorescence intensity at 320 nm (FI 360/320 nm). The ratio of fluorescence intensities at these wavelengths was chosen for the analysis in order to simultaneously monitor changes in the native and unfolded maxima. Equilibrium unfolding/refolding experiments were performed a minimum of three times for each protein.

Equilibrium unfolding/refolding data were analyzed to determine transition midpoints, $\Delta G°$ and *m* values by fitting to a three-state model using the method of Clark et al. (1993) with the curve fitting feature of Kaleidagraph (Synergy software). Averages and standard deviations were determined for the parameters of each protein from the fits of three separate experiments.

All transitions were fit to a three-state model described by equations (2), (3) and (4),

$$Y=((Y_N+S_N*[GuHCl])+(Y_I*K_1)+((Y_U+S_U*[GuHCl])*K_1*K_2))/(1+K_1+K_1*K_2) \quad (2)$$

$$K_1=\exp((m_1*[GuHCl]-\Delta G_1)/(R*T)) \quad (3)$$

$$K_2=\exp((m_2*[GuHCl]-\Delta G_2)/(R*T)) \quad (4)$$

where Y is the observed FI 360/320 nm signal, Y_N and Y_U are the intercepts of the native and unfolded baselines, S_N and S_U are the slopes of the native and unfolded baselines, and Y_I is the signal of the intermediate. Additionally, m_1 and ΔG_1 are the *m* value and $\Delta G°$

for the native to intermediate transition and m_2 and ΔG_2 are the *m* value and $\Delta G°$ for the intermediate to unfolded transition. T is temperature in Kelvin and R is the gas constant in units of kcal*mol^{-1}*K^{-1}.

4. *Productive refolding kinetics*

Kinetic refolding experiments were carried out by initially preparing unfolded stock solutions of the mutant proteins at 100 μg/ml in 5.5 M GuHCl. The unfolded stock solutions were incubated at 37 °C for three hours to ensure complete unfolding. Refolding buffer containing 10 mM sodium phosphate, 5 mM DTT, 1 mM EDTA, pH 7.0 was equilibrated at 37 °C. Unfolded protein was injected into refolding buffer using a syringe port injection system to give a final protein concentration of 10 μg/ml. The refolding samples were continuously excited at 295 nm and fluorescence emission monitored at 350 nm for three hours. The fluorescence spectra of refolded samples were subsequently measured to ensure that the proteins had refolded into a native-like conformation. Kinetic refolding data were fit to one, two and three exponentials using the curve fitting feature of Kaleidagraph (Synergy software) and the model with the best fit was determined by inspection. Productive kinetic refolding experiments were performed three times for each protein, from which and averages and standard deviations were calculated for the fitted parameters.

D. RESULTS

1. *Protein expression and purification*

All of the proteins used in this study had an exogenous N-terminal peptide with the sequence MKHHHHHHQ to aid in purification. Previous studies confirmed that the addition of the His-tag did not perceptibly affect the structure of the native protein or its thermodynamic or kinetic refolding properties (Kosinski-Collins and King 2003; Kosinski-Collins et al. 2004).

The mutant proteins expressed at levels comparable to wild type and behaved similarly to wild type during purification. The proteins were found primarily in the soluble fraction after cell lysis (greater than 90%) and were purified by Ni-NTA affinity chromatography to greater than 98% homogeneity (data not shown). As described further below, the proteins were in soluble native-like conformations. Thus, these residues were not required for the *in vivo* folding of HγD-Crys in *E. coli.*

2. *Circular dichroism and fluorescence spectroscopy*

The structures of single and double alanine mutants of HγD-Crys were probed by circular dichroism (CD) and fluorescence spectroscopy. In accord with previous results, the far-UV CD spectrum of wild-type HγD-Crys at 37 °C was reminiscent of a primarily β-sheet protein with a prominent minimum at 218 nm (Andley et al. 1996; Pande et al. 2000). The far UV-CD spectra of the single and double mutants, Q54A, and R79A/M147A closely resembled that of wild-type HγD-Crys (Fig. 3-3A). In contrast, the ellipticity intensities of Q143A, Q54A/Q143A, R79A and M147A differed slightly from wild-type HγD-Crys and the other mutants. Despite the difference in intensity, the minima of all proteins were indistinguishable. The discrepancy in intensity may reflect disturbances of the domain interface or other structural rearrangements that do not grossly disrupt overall β-sheet content.

Fluorescence spectra of the wild-type and mutant proteins were measured using an excitation wavelength of 295 nm with emission monitored from 310 to 420 nm. Consistent with previous results, the spectrum of wild-type HγD-Crys had an emission maximum of 325 nm (Kosinski-Collins and King 2003). The fluorescence spectra of the mutant proteins had emission maxima and similar intensities as wild-type HγD-Crys (Fig. 3-3B). These results suggest that the tryptophan side chains are buried in the hydrophobic cores of the mutant proteins and are located in environments similar to that of the wild-type protein. Unfolding the proteins in 5.5 M GuHCl shifted the emission maxima to 350 nm and increased the emission intensities (data not shown). This native state quenching has been previously described for HγD-Crys and other β- and γ-crystallins (Kim et al. 2002; Bateman et al. 2003; Kosinski-Collins et al. 2004).

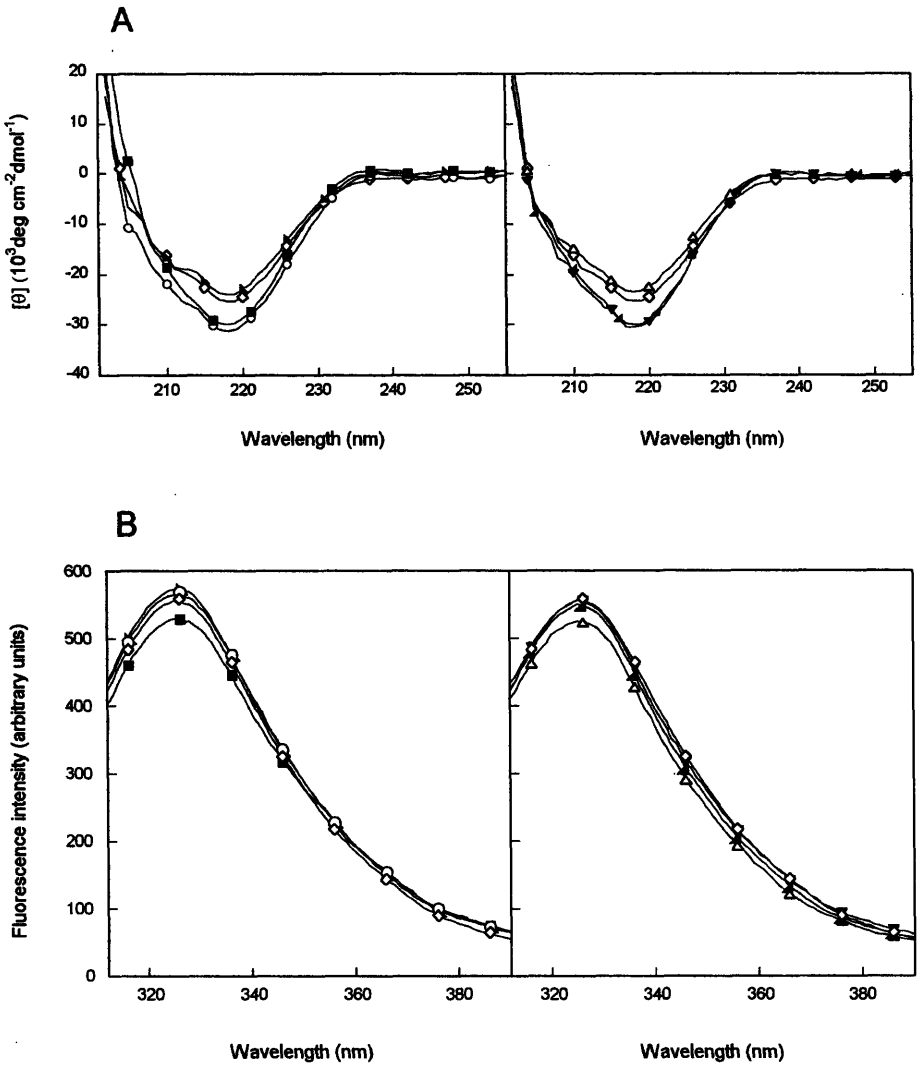

Figure 3-3. (A) Far-UV CD spectra of wild-type (◇), R79A (○), M146A (■), R79A/M146A (◣), Q54A (△), Q143A (◢) and Q54A/Q143A (▼) HγD-Crys. Samples contained 100 μg/ml protein in 10 mM sodium phosphate, 5 mM DTT, 1 mM EDTA, pH 7.0 at 37 C. A 0.25 cm pathlength cuvette was used for all measurements. (B) Fluorescence spectroscopy of wild-type and mutant HγD-Crys. Symbols are the same as in (A). Samples contained 10 μg/ml protein in 10 mM sodium phosphate, 5 mM DTT, 1 mM EDTA, pH 7.0 at 37 C.

3. *Equilibrium unfolding/refolding of wild-type HγD-Crys*

Previous equilibrium unfolding/refolding experiments of wild-type HγD-Crys used fluorescence spectroscopy to monitor structural changes in GuHCl (Kosinski-Collins and King 2003; Kosinski-Collins et al. 2004; Flaugh et al. 2005a). The fluorescent residues of HγD-Crys include four buried tryptophans and fourteen primarily surface-exposed tyrosines. An excitation wavelength of 295 nm was used to selectively monitor exposure of the tryptophan side chains to solvent. Structural transitions observable by this approach include complete unfolding of the molecule or structural transitions of a single domain.

Previous equilibrium unfolding/refolding experiments of HγD-Crys were performed at 37°C using an equilibration time of 24 hours, and the data was analyzed by plotting the concentration of GuHCl versus a ratio of fluorescence intensities at 360 and 320 nm (Flaugh et al. 2005a). Using these parameters, a slight inflection in the transitions was observed at approximately 2.3 M GuHCl suggesting population of a partially folded intermediate (Fig. 3-4). The transitions were best fit by a three-state model where population of a partially folded intermediate was assumed to occur in equilibrium with the native and unfolded states. When fit to a three-state model, the native to intermediate transition had a midpoint of 2.2 M GuHCl and the intermediate to unfolded transition had a midpoint of 2.8 M GuHCl (Table 3-1). The first transition likely corresponded to unfolding/refolding of the N-td and the second transition to unfolding/refolding of the C-td (Flaugh et al. 2005a).

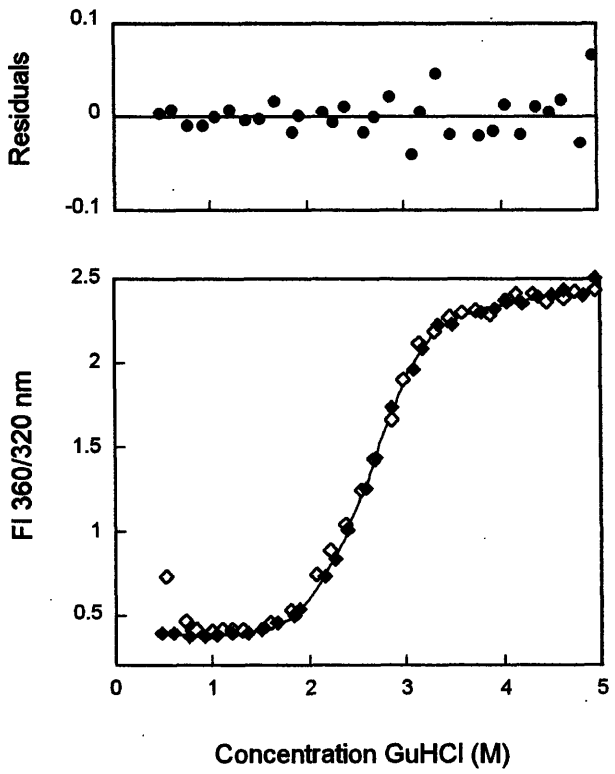

Figure 3-4. Equilibrium unfolding (closed symbols) and refolding (open symbols) of wild-type HγD-Crys in GuHCl probed by fluorescence emission. Fluorescence spectra were collected using an excitation wavelength of 295 nm and fluorescence intensity at 360/320 nm was used for data analysis. Protein was present at 10 μg/ml in 10 mM sodium phosphate, 5 mM DTT, 1 mM EDTA, pH 7.0 and GuHCl from 0 to 5.5 M at 37 C, and samples were incubated for 24 hours prior to measurement.. The solid line is a three-state fit of the unfolding data. Residuals of the three-state fit are shown.

Table 3-1. Equilibrium unfolding/refolding parameters for wild-type and mutant HγD-Crys.

Protein	Transition 1					Transition 2		
	$[GuHCl]_{1/2}$[a]	Apparent *m* value[b]	Apparent $\Delta G°$[c]	$\Delta\Delta G°$[d]	$\Delta\Delta G_{Int}$[e]	$[GuHCl]_{1/2}$[a]	Apparent *m* value[b]	Apparent $\Delta G°$[c]
Wild type[f]	2.2 ± 0.1	3.6 ± 0.1	7.7 ± 0.2	-		2.8 ± 0.1	3.1 ± 0.4	8.9 ± 1.3
R79A	1.8 ± 0.1	3.4 ± 0.4	6.7 ± 1.6	-1.0		2.9 ± 0.1	2.7 ± 0.2	7.6 ± 0.6
M147A	1.8 ± 0.1	3.7 ± 0.4	6.6 ± 1.3	-1.1		2.8 ± 0.1	3.1 ± 0.3	8.9 ± 1.0
Q54A	2.0 ± 0.1	3.8 ± 0.2	7.5 ± 0.2	-0.2		2.9 ± 0.1	3.4 ± 0.3	9.6 ± 1.1
Q143A	1.9 ± 0.1	3.8 ± 0.2	7.2 ± 0.3	-0.5		3.0 ± 0.1	3.0 ± 0.2	9.0 ± 0.7
R79A/M147A	1.7 ± 0.1	3.2 ± 0.4	5.6 ± 0.5	-2.1	0.0	2.8 ± 0.1	3.2 ± 0.3	8.9 ± 0.8
Q54A/Q143A	1.8 ± 0.1	3.5 ± 0.1	6.3 ± 0.1	-1.4	-0.7	3.1 ± 0.1	3.1 ± 0.2	9.6 ± 0.7

[a] Transition midpoints in units of M.

[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 $kcal*mol^{-1}$.

[d] $\Delta\Delta G° = \Delta G°_{Mutant} - \Delta G°_{Wild\text{-}type}$ in units of $kcal*mol^{-1}$.

[e] Free energy of interaction for peripheral amino acids calculated using the equation: $\Delta\Delta G_{Int} = \Delta G°_{Wild\text{-}type} + \Delta G°_{XY} - \Delta G°_{X} - \Delta G°_{Y}$ in units of $kcal*mol^{-1}$, where X, Y and XY are the single and double mutants of the putatively interacting side chains. A negative value represents a favorable interaction energy in the native state.

4. *Equilibrium unfolding/refolding of Arg79/Met147 mutants*

In order to assess the contributions of peripheral domain interface amino acids to the stability of HγD-Crys, equilibrium unfolding/refolding experiments were performed. These experiments were performed on all alanine substitution mutants using analogous methodologies as those described for wild-type HγD-Crys.

The equilibrium unfolding/refolding transitions of R79A deviated slightly from that observed for wild-type (Fig. 3-5). Similar to wild type, a plateau was evident in the unfolding/refolding curves at approximately 2.3 M GuHCl. However, the range of GuHCl concentrations over which the intermediate was populated was increased for R79A. The unfolding/refolding transitions were best fit to a three-state model similar to wild type with a native to intermediate transition midpoint of 1.8 M GuHCl and an intermediate to unfolded transition midpoint of 2.9 M GuHCl (Table 3-1). Alanine substitution of Arg79 decreased the free energy of unfolding ($\Delta G°$) of the native to intermediate transition by approximately 1.0 kcal*mol^{-1} compared to wild-type HγD-Crys (Table 3-1).

The equilibrium unfolding/refolding transitions of M147A also displayed a significant plateau suggesting population of a partially folded intermediate (Fig. 3-5). Similar to R79A, the intermediate was populated over a larger range of GuHCl concentrations than wild-type HγD-Crys. When fit to a three-state model, the native to intermediate transition had a midpoint of 1.8 M GuHCl and the intermediate to unfolded transition had a midpoint of 2.8 M GuHCl (Table 3-1). The $\Delta G°$ of the native to intermediate transition was decreased by 1.1 kcal*mol^{-1} compared to wild type (Table 3-1).

The double alanine mutant R79A/M147A also populated a partially folded intermediate during equilibrium unfolding/refolding (Fig. 3-5). The native to intermediate transition was calculated to have a midpoint of 1.7 M GuHCl and the transition from intermediate to unfolded had a transition of 2.8 M GuHCl (Table 3-1). The $\Delta G°$ of the native to intermediate transition was decreased by approximately 2.1 kcal*mol^{-1} compared to wild-type HγD-Crys (Table 3-1).

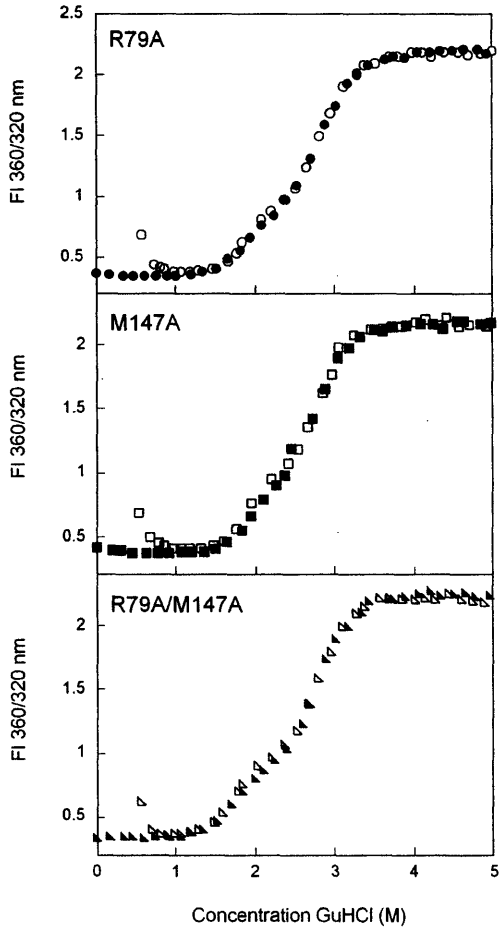

Figure 3-5. Equilibrium unfolding/refolding of R79A (●), M147A (■), and R79A/M147A (◣) HγD-Crys in GuHCl probed by fluorescence emission. Data was analyzed by fluorescence intensity at 360/320 nm using an excitation wavelength of 295 nm. Protein was present at 10 μg/ml in 10 mM sodium phosphate, 5 mM DTT, 1 mM EDTA, pH 7.0 and GuHCl from 0 to 5.5 M at 37 C. Samples were equilibrated for 24 hours. The unfolding transitions are shown as closed symbols and the refolding transition as open symbols.

5. *Equilibrium unfolding/refolding of Gln54/Gln143 mutants*

Equilibrium unfolding/refolding of Q54A HγD-Crys also occurred by an apparent three-state mechanism with population of an intermediate at approximately 2.3 M GuHCl (Fig. 3-6). The transition from native to intermediate had a midpoint of 2.0 M GuHCl while the intermediate to unfolded transition had a midpoint of 2.9 M GuHCl (Table 3-1). This corresponded to a decrease in the native to intermediate $\Delta G°$ of 0.2 kcal*mol^{-1} relative to wild type. In the crystal structure of HγD-Crys (Basak et al. 2003), the side chain of Gln54 participates in a hydrogen bond with the main chain nitrogen of Leu145.

The side chain of Gln143 lies in very close proximity to that of Gln54. However, Gln143 does not participate in hydrogen bonding to any main chain or side chain atoms as observed in the crystal structure. There was a plateau in equilibrium unfolding/refolding curves of Q143A at 2.3 M GuHCl (Fig. 3-6). When fit to three-state model, the native to unfolded transition had a midpoint of 1.9 M GuHCl and the intermediate to unfolded transition had a midpoint of 3.0 M GuHCl (Table 3-1). The native to intermediate transition had a $\Delta G°$ 0.5 kcal*mol^{-1} less than that for wild type.

The double mutant protein, Q54A/Q143A also displayed a plateau in the equilibrium unfolding/refolding transitions (Fig. 3-6). The presence of a partially unfolded intermediate is more obvious for Q54A/Q143A than for any of the other mutants, as it appears to be populated over a range of 2.2 to 2.6 M GuHCl. The two transitions were calculated to have midpoints of 1.8 and 3.1 M GuHCl (Table 3-1). The $\Delta G°$ of the native to intermediate transition was decreased approximately 1.4 kcal*mol^{-1} relative to wild type.

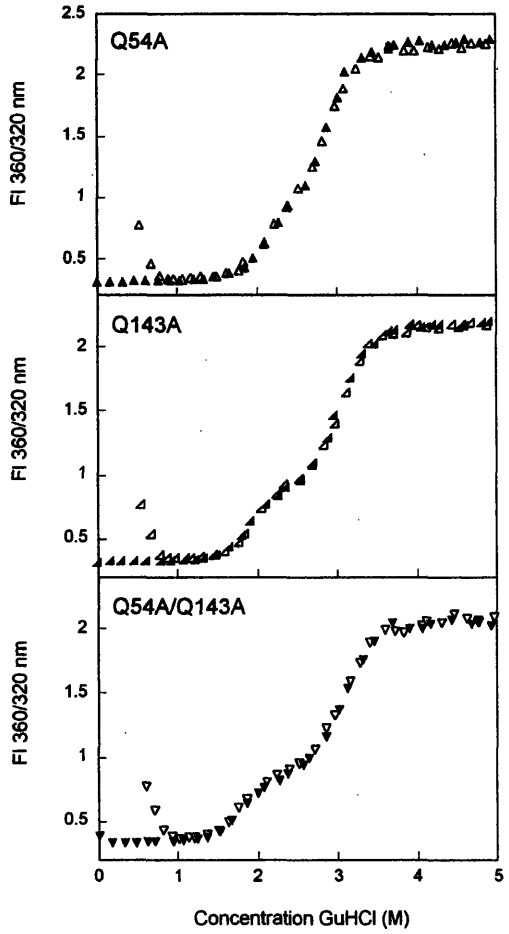

Figure 3-6. Equilibrium unfolding/refolding of Q54A (▲), Q143A (◢) and Q54A/Q143A (▼) HγD-Crys in GuHCl probed by fluorescence emission. The unfolding transitions are shown as closed symbols and the refolding transition as open symbols. Data was analyzed by fluorescence intensity at 360/320 nm using an excitation wavelength of 295 nm. Protein was present at 10 μg/ml in 10 mM sodium phosphate, 5 mM DTT, 1 mM EDTA, pH 7.0 and GuHCl from 0 to 5.5 M at 37 C, with a 24 hour equilibration time.

6. *In vitro aggregation*

HγD-Crys did not significantly aggregate or self-associate in the transition region during unfolding and refolding. However, consistent with previous results, when refolded from 5.5 M GuHCl to less than 1.0 M GuHCl, wild-type HγD-Crys aggregated into a high molecular weight species that significantly scattered light (Kosinski-Collins and King 2003). This causes a sharp increase in the FI 360/320 nm values of samples refolded to less than 1.0 M GuHCl due to right-angle light scattering by the aggregates (Fig. 3-4). All of the peripheral domain interface mutants displayed similar aggregation behavior (Figs. 3-5 and 3-6).

Refolding to the native state below 1.0 M GuHCl is masked by the light scattering from the aggregated chains. The aggregation samples of all the mutant proteins were tested for the presence of native-like protein by fluorescence spectroscopy after centrifugation at 12,000 rpm. The fluorescence spectra of the soluble protein present in the supernatant after centrifugation were consistent with the presence of native protein (data not shown). Thus, the productive refolding and aggregation pathways compete under these conditions. All spectra displayed an emission maximum of approximately 325 nm and decreased fluorescence intensity reflecting the loss of protein molecules into the aggregate.

7. *Productive refolding kinetics of wild-type HγD-Crys*

To analyze the role of peripheral domain interface residues in kinetic refolding of HγD-Crys, productive kinetic refolding experiments were performed analogous to previous experiments with wild type (Kosinski-Collins et al. 2004). Proteins were fully unfolded in 5.5 M GuHCl and subsequently diluted into refolding buffer. Burial of the tryptophan residues was monitored by observing a decrease in fluorescence intensity at 350 nm over time. Rapid dilution was performed with a syringe-injection port instead of a stopped-flow apparatus because the major transitions in HγD-Crys refolding have been shown to occur on a second, and not millisecond time scale (Kosinski-Collins et al.

2004). Potential refolding intermediates populated on a millisecond time scale were not addressed in these experiments.

Previous experiments determined that kinetic refolding of wild-type HγD-Crys was best fit with two exponentials suggesting population of a partially folded intermediate (Kosinski-Collins et al. 2004). Upon dilution into refolding buffer the fluorescence intensity at 350 nm rapidly decreased with a $t_{1/2}$ of 15 seconds. This first phase presumably corresponded to refolding into an intermediate conformation. The partially folded intermediate was more fluorescent than the native state and less fluorescent than the unfolded state at 350 nm. Following the fast phase, a second phase was observed where the fluorescence intensity decreased slowly with a $t_{1/2}$ of 190 seconds. This transition presumably corresponded to refolding into the native conformation.

$$\text{Denatured} \xrightarrow{k_1} \text{Intermediate} \xrightarrow{k_2} \text{Native} \qquad (1)$$

8. *Productive refolding kinetics of Arg79/Met147 mutants*

Kinetic refolding of R79A is shown in Figure 3-7. Consistent with previous results of wild-type HγD-Crys, the curve was best fit with a three-state model. An initial rapid decrease in fluorescence intensity at 350 nm occurred with a $t_{1/2}$ of 19 seconds. A second slower phase occurred with a $t_{1/2}$ of 890 seconds (Fig. 3-7, Table 3-2). Similar to wild-type HγD-Crys, the intermediate had a fluorescence emission signal at 350 nm unique from both the native and unfolded conformations (Fig. 3-7). The $t_{1/2}$ for transition from the denatured to intermediate state was approximately equal to that of wild-type HγD-Crys while the $t_{1/2}$ for the intermediate to native transition was increased more than four fold.

Refolding of M147A was also best fit with a three-state model (Fig. 3-7). A $t_{1/2}$ of 21 seconds was calculated for the initial rapid decrease in fluorescence corresponding to a transition from the denatured to intermediate state. The second transition from partially folded intermediate to native was significantly slower with a $t_{1/2}$ of 680 seconds (Table 3-

2). Similar to the results for R79A, the $t_{1/2}$ for the unfolded to intermediate transition of M147A was approximately equal to that of wild-type HγD-Crys and the $t_{1/2}$ for the intermediate to native transition was notably increased.

The double mutant, R79A/M147A displayed the most significantly altered refolding kinetics (Fig. 3-7). The denatured to partially folded intermediate transition was similar to that of wild-type HγD-Crys with a $t_{1/2}$ of 23 seconds. Conversely, the partially folded intermediate to native transition was markedly slower, occurring with a $t_{1/2}$ of 1700 seconds (Table 3-2).

Table 3-2. Kinetic refolding parameters for wild-type and mutant HγD-Crys.

Protein	k_1 (seconds^{-1})	$t_{1/2}$ (seconds)	k_2 (seconds^{-1})	$t_{1/2}$ (seconds)
Wild type[a]	0.048 ± 0.001	15 ± 1	0.0037 ± 0.0001	190 ± 10
R79A	0.037 ± 0.006	19 ± 3	0.0008 ± 0.0001	890 ± 50
M147A	0.032 ± 0.002	21 ± 1	0.0010 ± 0.0001	680 ± 20
Q54A	0.040 ± 0.001	17 ± 1	0.0023 ± 0.0002	310 ± 30
Q143A	0.041 ± 0.003	17 ± 1	0.0016 ± 0.0001	430 ± 10
R79A/M147A	0.030 ± 0.004	23 ± 3	0.0004 ± 0.0001	1700 ± 20
Q54A/Q143A	0.033 ± 0.004	20 ± 2	0.0012 ± 0.0001	600 ± 40

[a] From Flaugh (Flaugh et al. 2005a)

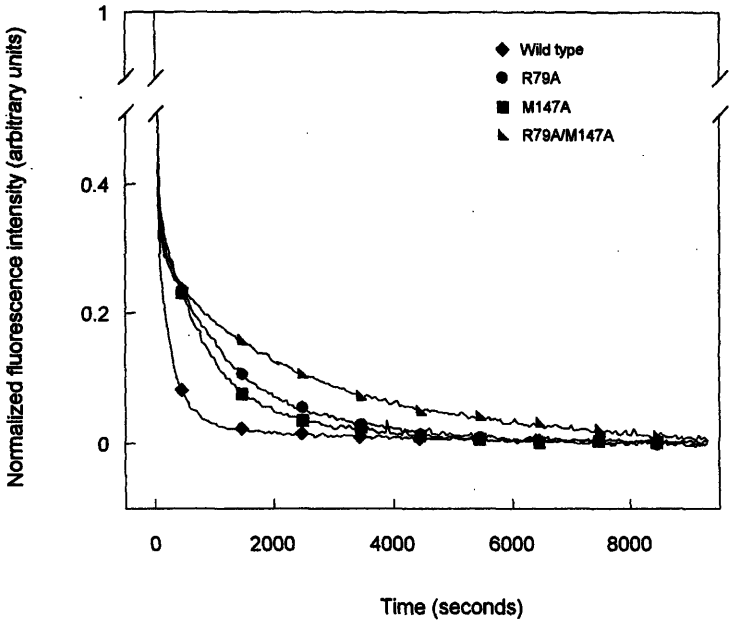

Figure 3-7. Productive kinetic refolding of wild-type (♦), R79A (●), M147A (■) , and R79A/M147A (■) HγD-Crys. Proteins were initially unfolded at a concentration of 100 μg/ml in 5.5 M GuHCl at 37 C. Refolding was initiated by dilution of unfolded proteins into refolding buffer to give a final concentration of 10 μg/ml. Refolding buffer contained 10 mM sodium phosphate, 5 mM DTT, 1 mM EDTA, pH 7.0, and 1.0 M GuHCl at 37 °C. Refolding was monitored by changes in fluorescence emission at 350 nm.

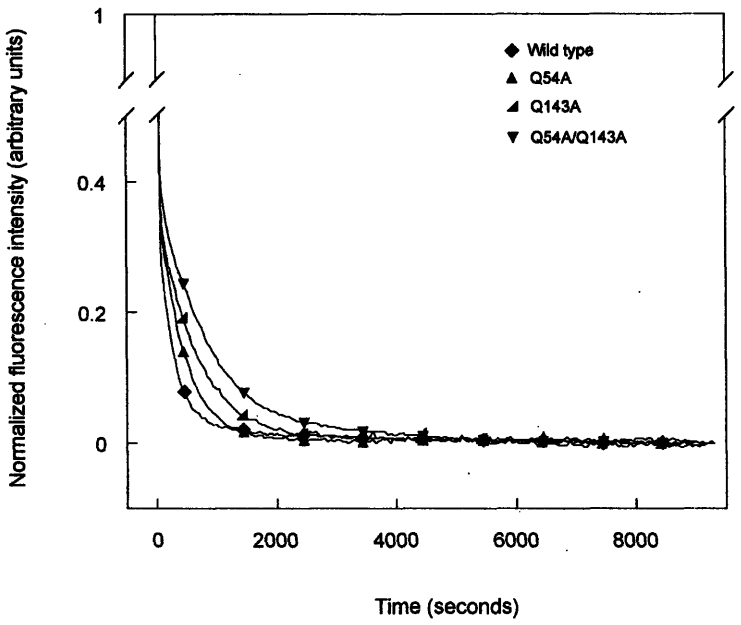

Figure 3-8. Productive kinetic refolding of wild-type (♦), Q54A (▲), Q143A (◢) and Q54A/Q143A (▼) HγD-Crys. Unfolded stock solutions of proteins were initially prepared at a concentration of 100 μg/ml in 5.5 M GuHCl at 37 C. Refolding was initiated by diluting the unfolded stock solutions into refolding buffer to give a final protein concentration of 10 μg/ml. Refolding buffer contained 10 mM sodium phosphate, 5 mM DTT, 1 mM EDTA, pH 7.0, and 1.0 M GuHCl at 37 °C. Refolding was monitored by changes in fluorescence emission at 350 nm.

9. *Productive refolding kinetics of Gln54/Gln143 mutants*

Productive kinetic refolding of the mutant proteins Q54A, Q143A and Q54A/Q143A are shown in Figure 3-8. The data were best fit with a three-state model similar to wild type. Changes in fluorescence intensity at 350 nm upon refolding were similar to those described above for wild type. The first phase was characterized by a rapid decrease in fluorescence intensity and was followed by a second phase with a slower decrease in intensity. The transition from denatured to partially folded intermediate occurred with $t_{1/2}$ values of 17 seconds for both Q54A and Q143A and 20 seconds for Q54A/Q143A (Table 3-2). These values are not significantly different than that measured for wild-type HγD-Crys. Conversely, the $t_{1/2}$ values for the transition from partially folded intermediate to native were slightly increased compared to wild-type HγD-Crys. The $t_{1/2}$ values were 310 seconds for Q54A, 430 seconds for Q143A, and 600 seconds for Q54A/Q143A (Table 3-2).

E. DISCUSSION

Based on crystal structure and sequence alignment data, most domain interfaces of the β- and γ- crystallins are composed of a central hydrophobic cluster surrounded by peripheral paired interactions. The paired residues of HγD-Crys, identified by proximity in the crystal structure, are Gln54/Gln143, and Arg79/Met147 (Fig. 3-1). The side chains of these amino acids are in close proximity to both the exterior of the protein and the interface hydrophobic cluster. The side chains of Arg79 and Met147 are located towards the bottom of the domain interface and interact by packing the hydrophobic side chain of Met147 against the β, γ and δ methylene groups of Arg79 (Fig. 3-1). The hydrophobic portions of the side chains of Arg79 and Met147 are also in close contact with the central hydrophobic cluster of the domain interface. Gln54 and Gln143 are located at the top of the domain interface and are in proximity of the hydrophobic cluster and the exterior of the protein (Fig. 3-1).

1. *Destabilizing effects of alanine substitutions*

The far-UV CD minima of all single and double alanine substitution mutants of peripheral interface residues were analogous to that of wild type. These data suggest that the mutants folded into a native-like structure with similar β-sheet content as wild-type HγD-Crys (Fig. 3-3A). However, the far-UV CD spectra of the mutants R79A, M147A, Q143A, and Q54A/Q143A deviated moderately in overall intensity from that of wild type and the other mutants. Mutations of domain interface residues may disrupt or alter domain pairing which could cause these differing CD signals. An alternative interpretation is that the differences in the spectra are due to discrepancies in individual solution conditions and not actual changes in the native state structures.

The fluorescence emission spectra of the mutant proteins indicated that they adopted native-like conformations similar to wild-type HγD-Crys. All single and double alanine mutants displayed a fluorescence emission maximum of approximately 325 nm, and native-state quenching similar to the wild-type protein. These results suggest that the tryptophan side chains are buried in the hydrophobic cores of all mutants. However, native state fluorescence intensities of the mutants differed slightly from that of wild type (Fig. 3-3B). This may be due to a structural rearrangement or relaxation around the tryptophans responsible for the quenching (Kosinski-Collins et al. 2004). Conclusive evidence for structural modifications will require obtaining high resolution structures of the mutant proteins. The overall behavior of these mutant proteins suggests that, individually, these side chains are not critical determinants of the native state fold.

Stabilities of the mutant proteins were determined by equilibrium unfolding/refolding in GuHCl. An inflection in the unfolding/refolding curves was evident at 2.3 M GuHCl for the wild-type and mutant proteins, suggesting population of a partially folded intermediate in equilibrium with the native and unfolded states. Previous investigations of triple tryptophan mutants of HγD-Crys indicated that the C-td is more stable than the N-td (Kosinski-Collins et al. 2004). Preliminary studies of the isolated N- and C-terminal domains of HγD-Crys have confirmed this observation (I.A. Mills, S.L. Flaugh and J. King unpubl.). Thus, we postulate that the intermediate populated during

equilibrium unfolding/refolding of wild-type HγD-Crys is a single folded domain conformer with a folded C-td and unfolded N-td (Flaugh et al. 2005a). According to this hypothesis, the native to intermediate transition corresponded to unfolding/refolding of N-td and the intermediate to unfolded transition corresponded to unfolding/refolding of the C-td. From here on we discuss the results according to these terms.

All single and double alanine substitution mutants displayed unfolding/refolding transitions that differed from that of wild type. The midpoint of N-td unfolding/refolding was consistently decreased for all mutants, thus causing the partially folded intermediate to be populated over a larger range of GuHCl concentrations (Figs. 3-5, 3-6; Table 3-1). Similar effects were observed with alanine substitution mutants of hydrophobic domain interface residues (Flaugh et al. 2005a). Therefore, stability of the N-td was dependent on correct association of both peripheral and hydrophobic domain interface residues.

In contrast to the consistent destabilization of the N-td, midpoints of C-td unfolding/refolding remained the same or increased slightly with the mutations. Thus, the C-td of full-length wild-type HγD-Crys was not stabilized by domain interface contacts. The mutants Q143A had a midpoint of 3.0 M GuHCl for C-td unfolding/refolding and Q54A/Q143A had a midpoint of 3.1 M GuHCl. If the partially folded intermediate has a structured C-td and unstructured N-td, in the intermediate conformation the side chain of Gln143 would be in close proximity to hydrophobic domain interface residues of the structured C-td. Substituting this side chain for alanine may have increased the stability of the C-td by eliminating potentially unfavorable interactions between the closely positioned polar and hydrophobic side chains. Two of the other mutants, R79A and Q54A also displayed intermediate to unfolded transition midpoints slightly greater than that of wild type (Table 3-1). These differences may be due to slight changes in the *m* values of the mutants or other effects on the stability of the intermediate that would be difficult to identify without a more detailed description of the conformation.

Double-mutant cycle analysis has been extensively used as a method to evaluate interaction energy between two amino acid side chains in close proximity in the native state conformation of a protein (Horovitz 1996). By comparing the $\Delta G°$ for the single and double alanine substitution mutants of Arg79/Met147 and Gln54/Gln143, it was

possible to estimate their interaction energies in the native state. Destabilization of the N-td exhibited by the single mutants R79A (-1.0 kcal*mol^{-1}) and M147A (-1.1 kcal*mol^{-1}) summed to equal the destabilization of the N-td exhibited by the double mutant R79A/M147A (-2.1 kcal*mol^{-1}). This suggests that the interaction of Arg79 and Met147 does not contribute significantly to overall stability of the native state. In contrast, destabilization of the N-td by the single mutants Q54A (-0.2 kcal*mol^{-1}) and Q143A (-0.5 kcal*mol^{-1}) did not sum to equal the destabilization of the N-td by the double mutant Q54A/Q143A (-1.4 kcal*mol^{-1}). Instead, these side chains displayed a free energy of interaction of approximately 0.7 kcal*mol^{-1}. It is important to note that double mutant cycle analyses are reliable measures of interaction energy only if the native state conformations of the mutant proteins are identical to the wild-type protein. While CD and fluorescence spectroscopy studies suggest that the conformations of peripheral domain interface mutants are similar to wild type, these techniques have relatively low sensitivity and would not likely detect altered domain pairing. Additionally, as seen in Table 3-1, the error associated with the $\Delta G°$ values is high in some cases (R79A and M147A), which limits the accuracy of calculated free energies of interaction. However, interaction energies also appear minimal when estimated by transition midpoints, which have much lower error. Together, these results suggest minimal to no interaction energy between these side chains.

Despite the minimal interaction energies of peripheral domain interface side chains, single and double alanine substitutions of these residues resulted in a significant decrease in the native to intermediate $\Delta G°$ from 0.2 to 2.1 kcal*mol^{-1} (Table 3-1). As described above, the hydrophobic domain interface residues are also critical for maintaining stability of the N-td by precise positioning and burial of hydrophobic surface area (Flaugh et al. 2005a). While both the peripheral and hydrophobic domain interface residues act to stabilize the N-td, they likely do so by very different means. Given that the peripheral residues are in close proximity to the interface hydrophobic cluster, we suspect that these side chains function to shield the cluster from solvent. In this way, a decrease in the size and polarity of the side chains by mutation to alanine would cause an unfavorable exposure of the hydrophobic residues to solvent and thus destabilization of the N-td through a loss of favorable domain interface contacts.

2. *Effect of interface substitutions on the kinetic refolding pathway*

At 37°C and pH 7.0, wild-type HγD-Crys refolded to a native-like conformation at concentrations of GuHCl above 1.0 M. Previous studies found that productive kinetic refolding of wild-type HγD-Crys was best fit to two exponentials with a $t_{1/2}$ value of 15 seconds for the first phase and 190 seconds for the second phase (Kosinski-Collins et al. 2004). The structural transitions that corresponded to these two kinetic phases were elucidated using engineered triple tryptophan mutant proteins (Kosinski-Collins et al. 2004). The triple tryptophan mutant proteins had three of the four native tryptophans of HγD-Crys mutated to phenylalanine, so that refolding of the two domains could be independently monitored by fluorescence spectroscopy. Refolding rates of the two mutants containing tryptophans from the C-td were comparable to the first exponential fit of wild type, with $t_{1/2}$ values of 30 seconds for both mutants. Similarly, refolding rates of the two mutant proteins with tryptophans in the N-td were comparable to the second exponential fit of wild type, with $t_{1/2}$ values of 190 and 210 seconds. Together these data suggest that the wild type protein refolded through a short lived kinetic intermediate, in which the C-td was largely folded and the N-td was largely unfolded (Kosinski-Collins et al. 2004).

Kinetic refolding of the single and double alanine mutants of peripheral interface residues were also best fit to two exponentials suggesting a similar sequential domain refolding pathway as wild type (Kosinski-Collins et al. 2004). By this model, the C-td of the mutant proteins refolded first followed by the N-td. Under this assumption, refolding of the C-td was not notably affected by mutations of the peripheral interface residues as seen by similar $t_{1/2}$ values as wild type. Conversely, all of the mutations resulted in a decreased refolding rate for the N-td. These results suggest that for full-length wild-type HγD-Crys, refolding of the C-td does not depend on correct contacts between peripheral domain interface residues while refolding of the N-td does. Previous experiments on single amino acid substitutions of hydrophobic domain interface residues also indicated reduced refolding rates for the N-td but not the C-td (Flaugh et al. 2005a).

Slow kinetic transformations during protein refolding that occur on a second to minute time scale are often attributed to cis-trans proline isomerization (Brandts et al. 1975). Wild-type HγD-Crys has five proline residues, all of which are found in the trans conformation in the native state. Three of the prolines are in the N-td, one is in the peptide linking the two domains and the final proline is in the C-td. The N-td of HγD-Crys refolds slower than the C-td suggesting that the slow refolding of the N-td may be due to cis-trans isomerization of prolines in the connecting peptide and the N-td. The data presented here makes this unlikely. If the slow intermediate to native transition in HγD-Crys refolding was dependent on cis-trans proline isomerization, mutations in the domain interface would not be expected to decrease refolding rates to the extent that is observed here. Instead, we suggest that the slow kinetic step of HγD-Crys refolding corresponds to domain pairing accompanied by refolding of the N-td. These results suggest a refolding pathway where the solvent-exposed domain interface of the folded C-td likely acts as a nucleating center for refolding of the N-td (Flaugh et al. 2005a).

For the mutant R79A/M147A, kinetic refolding of the N-td occurred with a $t_{1/2}$ nine times greater than that of wild type. Interestingly, the single and double mutations of Gln54 and Gln143 did not have as significant an effect on the refolding rates as mutations of Arg79 and Met147. The double mutant, Q54A/Q143A had a $t_{1/2}$ for the intermediate to native transition of 600 seconds compared to 190 seconds for wild type. Similarly, Q54A and Q143A both had $t_{1/2}$ values for intermediate to native transition that were increased less than three-fold over that for wild type. These results suggest that the side chains of Gln54 and Gln143 are not as critical in the kinetic refolding pathway as those of Arg79 and Met147. The positioning of Gln54 and Gln143 may occur later during the kinetic pathway after the hydrophobic cluster and Arg79 and Met147 are brought together. By this model, mutations of Gln54 and Gln143 could still have a significant effect on stability without severely altering kinetic refolding behavior, as is observed here.

3. *Domain stability and interactions*

The two domains of the β- and γ-crystallins, composed of two Greek key motifs each, are thought to result from a gene duplication event (Wistow et al. 1983). Folding and stability of full-length β- and γ-crystallins as well as proteins corresponding to their isolated domains have been extensively studied (Bloemendal et al. 2004). The individual domains of many of the β- and γ-crystallins exhibit distinctly different stabilities. As described above, the C-td of HγD-Crys is more stable than the N-td at pH 7.0 (Kosinski-Collins et al. 2004). In contrast to this, the N-td of BγB-Crys is more stable than the C-td at pH 2.0 (Rudolph et al. 1990; Mayr et al. 1997). This difference is presumably due to a greater number of acidic amino acids on the surface of the C-td of BγB-Crys, which, at pH 2.0 destabilizes the domain by charge repulsion. The different domain stabilities exhibited by BγB-Crys and HγD-Crys may be due to discrepancies in experimental conditions and not inherent differences in the proteins.

For some of the β- and γ-crystallins that display differential domain stability, the domains are more stable in the full-length protein than in isolation (Sharma et al. 1990; Mayr et al. 1997; Wieligmann et al. 1999). This suggests that for these proteins, domain interface contacts play a significant role in the stability. Indeed, mutating domain interface residue of BγB-Crys, HβB1-Crys or HγD-Crys destabilizes the proteins (Palme et al. 1997; Kim et al. 2002; Flaugh et al. 2005a). Additional evidence indicates that domain interface interactions are important in determining oligomeric states of these proteins (Hope et al. 1994; Mayr et al. 1994; Trinkl et al. 1994).

4. *Implications for understanding aggregation and cataract*

Aged human lenses contain both water soluble and insoluble crystallin. The amount of protein in the water insoluble fraction increases with age and in cataract (Ringens et al. 1982). Analyses of the insoluble crystallin from aged or cataractous lenses have confirmed the presence of covalent damage, including methionine oxidation, and glutamine or asparagine deamidation (Hanson et al. 1998; Lampi et al. 1998; Ma et al. 1998; Hanson et al. 2000).

Erroneous protein aggregation is associated with a variety of human diseases, in addition to mature-onset cataract (Sato et al. 1996; Harper et al. 1997; Prusiner 1998; Uversky et al. 2001). A common characteristic of known aggregation processes is polymerization from a partially folded or non-native conformation (Mitraki 1989; Wetzel 1994; Booth et al. 1997; Jiang et al. 2001). *In vivo*, such partially folded species often represent incompletely folded polypeptide chains released from the ribosome (Mitraki 1989; Wetzel 1994). However, this phenomenon cannot explain the presence of aggregation-prone crystallin species in the lens, as crystallin aggregation likely occurs late in life, long after the proteins were initially synthesized. Instead, the aggregation-prone conformations are probably generated by destabilization of the native state and partial unfolding induced by covalent damage. The effects of deamidation on the structure, stability and other *in vitro* properties of HβB1-Crys have been comprehensively studied by Lampi and colleagues (Lampi et al. 2001; Kim et al. 2002; Lampi et al. 2002; Harms et al. 2004). These studies demonstrated that deamidation can cause altered structure, altered oligomer conformation and reduced stability of HβB1-Crys, *in vitro*. These alterations may be significant in mature-onset cataractogenesis.

Some cases of rare juvenile-onset cataracts in humans are associated with single amino acid substitutions of the γ-crystallins. The mutations R14C, P23T, R36S and R58H of HγD-Crys all result in childhood cataract (Heon et al. 1999; Stephan et al. 1999; Kmoch et al. 2000; Santhiya et al. 2002). Recombinant proteins with the R14C, R36S and R58H mutations have similar native state conformations and stabilities as wild-type HγD-Crys, but do have reduced phase transition barriers *in vitro* (Pande et al. 2000; Pande et al. 2001; Basak et al. 2003). These mutations probably cause cataract by crystallization (R58H and R36S) and inter-molecular disulfide bonding (R14C) of the native molecules in the lenses of affected individuals (Kmoch et al. 2000; Pande et al. 2000; Pande et al. 2001). These processes are unlikely to account for the formation of mature-onset cataracts from aged proteins that are of a wild-type sequence aside from covalent damage. The P23T recombinant mutant of HγD-Crys also had similar stability as the wild-type protein, but had greatly reduced solubility *in vitro*, and was hypothesized to cause congenital cataract by precipitation in the lens (Evans et al. 2004). It is unclear how this relates to the protein insolubility found in mature-onset cataract.

The single amino acid substitution, T5P of human γC crystallin (HγC-Crys) is associated with Coppock-like cataract with causes a dust-like opacity in the lens nucleus of newborns (Heon et al. 1999; Santhiya et al. 2002). In contrast to the congenital mutants of HγD-Crys, Fu and Liang (2002) showed that the T5P mutation of HγC-Crys alters the native state conformation and destabilizes the protein. Liang went on to further show that the T5P mutant also had an increased propensity to interact with the lens chaperone, human αA crystallin (Liang 2004). The mechanism of inherited cataract for the T5P mutant of HγC-Crys may be due to aggregation caused by the altered structure and decreased stability or alternatively loss of native state interactions with other crystallins due to the altered conformation (Liang 2004). Either way, the mechanism of cataract for the mutant of HγC-Crys may more closely reflect that of mature-onset cataract than the R14C, R36S and R58H mutants of HγD-Crys.

Our interest in identifying partially structured intermediates during HγD-Crys unfolding/refolding reflects the possibility that such species may be related to aggregation-prone precursors of cataract formation in the eye lens. Given that the peripheral residues Gln54, Gln143 and Met147 are potential sites of covalent damage, partially unfolded intermediates populated as a result of their modification are of particular interest for understanding aggregation and cataract. The results reported here indicate that these interface residues are critical for stability, and thus, covalent damage of the side chains in aged lenses may result in native state destabilization of HγD-Crys. This would effectively increase the probability of populating the partially unfolded conformer with a folded C-td and unfolded N-td. One model for interactions between such partially unfolded conformations includes domain swapping, a mechanism of protein oligomerization that has been previously described by Liu and Eisenberg (2002). In younger adults these damaged, partially unfolded species would likely be scavenged by α-crystallins. However, at some point these chaperones may become saturated resulting in the late onset of aggregation reactions from damaged crystallins.

CHAPTER FOUR:

EFFECTS OF GLUTAMINE DEAMIDATION ON THE STABILITY AND AGGREGATION OF HUMAN γD CRYSTALLIN[3]

[3] This work has been submitted to the Journal of Biological Chemistry

Collaborator Note: Ishara A. Mills performed all pH 3.0 equilibrium unfolding/refolding experiments.

A. ABSTRACT

The transparency of the eye lens requires life-long stability and solubility of the crystallin proteins. The crystallins of the mature lens exhibit high levels of covalent damage, including glutamine and asparagine deamidation, which may initiate protein aggregation or loss of solubility leading to cataract formation. Human γD crystallin (HγD-Crys) is a two-domain β-sheet protein located in the lens nucleus. The two domains interact through side chain contacts across an interdomain interface, including a pair of glutamines (Gln54 and Gln143) distal from the linker peptide connecting the domains. Gln54 and Gln143 are both important for stability and folding of the N-terminal domain (N-td) of HγD-Crys (Flaugh et al. 2005b). To test the effects of interface deamidation on stability, folding and aggregation of HγD-Crys, single and double glutamine to glutamate substitutions were constructed at these positions. Equilibrium unfolding/refolding experiments of wild type and the deamidation mutants were performed in guanidine hydrochloride at pH 7.0, 37°C or Urea at pH 3.0, 20°C. The single and double deamidation mutants were destabilized at pH 7.0 compared to wild type. The proteins populated a partially unfolded intermediate that likely had a structured C-terminal domain (C-td) and unstructured N-td. When the pH was decreased to 3.0, equilibrium unfolding transitions of the deamidation mutants were indistinguishable from those of wild type. In contrast, the double alanine mutant Q54A/Q143A was destabilized at both pH 7.0 and 3.0. Thermal stabilities of the deamidation mutants were also reduced at pH 7.0. The deamidation mutants also lowered the kinetic barrier to unfolding of the N-td as determined by kinetic unfolding and refolding experiments. The results presented here indicate that deamidation of interface glutamines decreased the thermodynamic stability of HγD-Crys and lowered the kinetic barrier to unfolding as the result of introducing a negative charge into the structurally-critical domain interface. Effects such as these may be significant in causing cataract by inducing protein aggregation or insolubility.

B. INTRODUCTION

Transparency of the human eye lens depends on high concentrations and short-range order of the crystallin proteins (Delaye and Tardieu 1983; Fernald and Wright 1983). The lens develops early in life, during which time the cells elongate and lose all organelles, including nuclei. Thus, mature lens cells do not have the ability to synthesize new protein and consequently the crystallin proteins of nuclear and cortical lens cells are as old as the lens itself. This unique phenomenon necessitates life-long stability and solubility despite elevated protein concentrations and continued exposure to environmental stresses. As the crystallin proteins age, they accumulate high levels of covalent damage.

Cataract is the leading cause of blindness worldwide and affects one in six people over age 40 in the US (National Eye Insitute (U.S.) 2002). Pathologically, cataract is associated with the presence of insoluble light scattering bodies composed of the crystallin proteins. Formation of light scattering bodies is likely due to loss of crystallin solubility and/or crystallin aggregation. A striking feature of cataract is the extreme rise in prevalence with increasing age (National Eye Insitute (U.S.) 2002), suggesting that covalent damage of the crystallins may cause or contribute to disease onset.

The three classes of crystallin proteins ubiquitous in all vertebrate lenses are the α-, β- and γ-crystallins. The α-crystallins associate to form large polydisperse multimers that possess *in vitro* molecular chaperone activity (Horwitz 1992; Boyle and Takemoto 1994). In contrast, the β- and γ-crystallins function solely as structural proteins of the lens. The β-crystallins form a range of multimeric states while the γ-crystallins are monomeric (Wistow et al. 1983; Bax et al. 1990; Slingsby and Bateman 1990). Despite differences in quaternary structure, the β- and γ-crystallins adopt analogous two domain β-sheet folds that are composed of four Greek key motifs, two in each of the homologous domains. The two domains of the γ-crystallins interact intramolecularly via side chain contacts across a domain interface. In contrast, the domains of βB2-Crys interact intermolecularly to form a pseudo domain-swapped dimer (Bax et al. 1990). Despite whether they occur intramolecularly or intermolecularly, domain interface interactions

are important for folding, stability and oligomerization of these two domain proteins (Hope et al. 1994; Mayr et al. 1994; Palme et al. 1997; Flaugh et al. 2005a; 2005b).

The wild-type γ-crystallins are remarkably stable and generally have free energies of unfolding (ΔG°) that are higher than the wild-type α- and β-crystallins (Bloemendal et al. 2004). However, the crystallins of aged lenses are not identical to the wild-type sequences because of the high levels of covalent damage and thus it is important to study crystallins that have been covalently modified in ways that mimic this damage.

Glutamine and asparagine deamidation is a particularly pervasive form of covalent damage that has been observed in all of the major crystallin proteins recovered from cataractous lenses (Groenen et al. 1994; Lund et al. 1996; Lampi et al. 1998). At the atomic level, deamidation can cause backbone isomerization and introduces a negative change at physiological pH because an amide group is replaced with a carboxyl group. Deamidation of the crystallins may cause changes in structure, stability or solubility that could instigate or contributes to cataract formation. For example, it has been suggested that deamidation of Asn143 in human γS crystallin is specifically associated with mature-onset cataract formation (Takemoto and Boyle 2000). Similarly, deamidated β-crystallins isolated from human lenses have an increased tendency to associate into non-covalent aggregates (Zhang et al. 2003).

Lampi and colleagues have carefully studied the effects of deamidation on the structure and stability of human β-crystallins (Lampi et al. 2001; Kim et al. 2002; Harms et al. 2004; Lampi et al. 2006). The effects of glutamine deamidation of the β-crystallins vary according to the site of damage. Deamidation of a glutamine in the connecting peptide of βB2-Crys caused the protein to form larger multimers and increased the tendency for thermal aggregation (Harms et al. 2004). In contrast, deamidation of glutamines in the domain interfaces of dimeric βB1-Crys and βB2-Crys decreased stabilities of the proteins and caused them to populate partially unfolded intermediates in equilibrium experiments (Kim et al. 2002; Lampi et al. 2002; Lampi et al. 2006). Similar studies on the effects of deamidation on α-crystallin structure and function have been carried out by Gupta and Srivastava (2004a, 2004b). The α-crystallin studies also displayed context-dependent effects, including decreased *in vitro* chaperone activity, changes in secondary and tertiary structures and altered oligomerization properties

(Gupta and Srivastava 2004b; 2004a). Specific effects of covalent damage on the structures and stabilities of the γ-crystallins have not been previously addressed.

Human γD crystallin (HγD-Crys) is one of the most abundant γ-crystallins of the lens nucleus and covalently damaged forms accumulate in aged lenses (Lampi et al. 1997; Hanson et al. 1998; Hanson et al. 2000; Searle et al. 2005). HγD-Crys adopts the typical two domain fold of the γ-crystallins where the two domains interact intramolecularly through side chain contacts across a domain interface (Basak et al. 2003). Stability and folding of HγD-Crys has been previously studied using fluorescence as a probe of conformation (Kosinski-Collins and King 2003; Evans et al. 2004; Kosinski-Collins et al. 2004; Flaugh et al. 2005a; 2005b). HγD-Crys has four tryptophans, two per domain, which are buried in analogous positions in the hydrophobic cores of the two domains. In equilibrium unfolding/refolding experiments at pH 7.0 and 37°C, full-length wild-type HγD-Crys populates a partially folded intermediate that likely has a folded C-terminal domain (C-td) and unfolded N-terminal domain (N-td) (Flaugh et al. 2005a). Off-pathway aggregation competes with productive refolding when HγD-Crys is diluted out of high concentrations of guanidine hydrochloride (GuHCl) into buffer (Kosinski-Collins and King 2003).

As shown in Figure 4-1, the amino acid side chains that interact across the domain interface of HγD-Crys include a pair of buried glutamines that are within hydrogen bonding distances to four water molecules present in the crystal structure (Basak et al. 2003). A cluster of six hydrophobic residues that are buried in the domain interface are also in close proximity to these apical glutamines. A previous alanine mutagenesis study found that while neither Gln54 nor Gln143 are critical determinants of structure, they are both vital for the folding and stability of HγD-Crys (Flaugh et al. 2005b). Specifically, single and double alanine mutations destabilized the N-td and also slowed its rate of refolding *in vitro* but did not perceptibly alter overall secondary or tertiary structure (Flaugh et al. 2005b). Glutamine is conserved in these positions among β- and γ-crystallins from diverse species at ~78% identity for Gln54 and ~80% identity for Gln143.

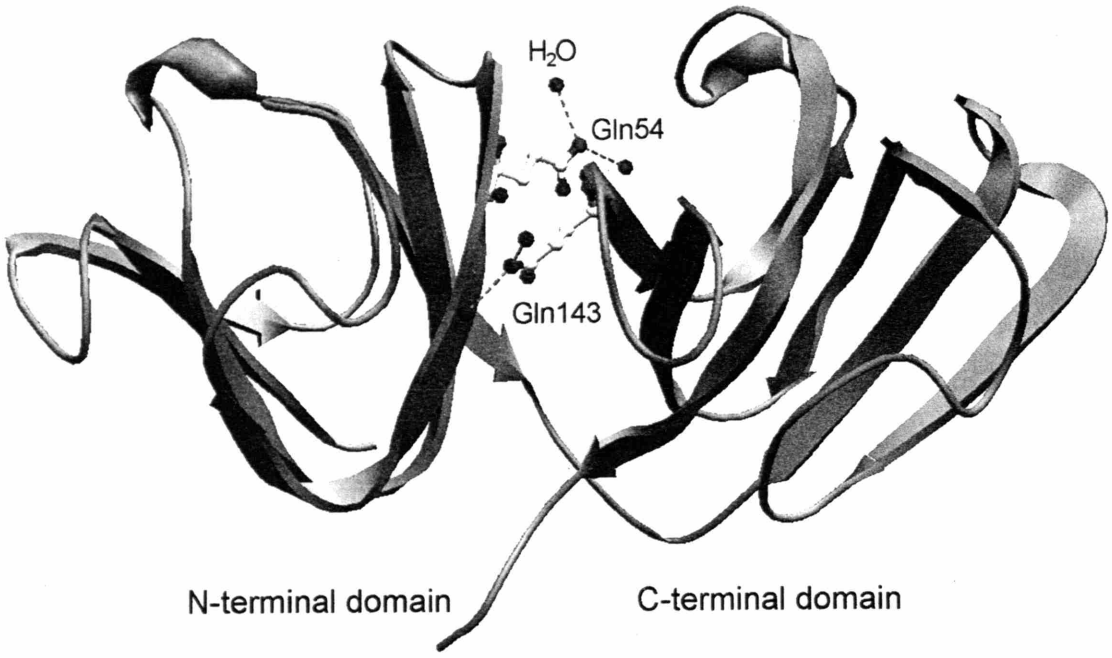

Figure 4-1. Structure of wild-type HγD-Crys (Basak et al. 2003) depicted in ribbon representation showing the location of interface residues Gln54 and Gln143 and four nearby crystallographic waters that form hydrogen bonds to the glutamine side chains (PDB code 1HK0).

Given the importance of Gln54 and Gln143 in stability and folding of HγD-Crys, and that deamidation of analogously positioned glutamines in the β-crystallins causes destabilization, we suspected that deamidation of these residues would destabilize the protein due to introduction of a negative charge at pH 7.0. To test this, single and double glutamine to glutamate mutants were constructed and the *in vitro* properties of the mutant proteins were studied. The deamidation mutations destabilized the proteins at pH 7.0 but not pH 3.0. Additionally, the mutations reduced the rate of refolding and increased the rate of unfolding of the N-td resulting in a decreased kinetic barrier to unfolding.

C. MATERIALS AND METHODS

1. *Mutagenesis, expression and purification of recombinant HγD-Crys*

Single and double glutamine to glutamate substitutions of residues Gln54 and Gln143 were constructed using site-directed mutagenesis. Mutant primers (IDT-DNA) were used to amplify the gene for HγD-Crys with an N-terminal His(6)-tag in a pQE.1 plasmid (Kosinski-Collins et al. 2004). Substitutions were confirmed by DNA sequencing of all amplified plasmids (Massachusetts General Hospital).

Wild-type and mutant HγD-Crys proteins were expressed and purified as described by Kosinski-Collins et al. (2004). The proteins were expressed in *E. coli* and cell lysates were purified by affinity chromatography using a Ni-NTA resin (Qiagen) and were dialyzed into 10 mM ammonium acetate (pH 7.0). Concentrations of purified proteins were calculated from absorbance data at 280 nm using an extinction coefficient of 41,040 cm^{-1} M^{-1} for all proteins.

2. *Calculating solvent accessible surface areas*

Solvent accessible surface areas of Gln54 and Gln143 were calculated from the crystal structure of HγD-Crys, PDB ID 1HK0 (Basak et al. 2003) using the program GETAREA 1.1 (Fraczkiewicz and Braun 1998).

3. *Circular dichroism spectroscopy*

CD spectra of the wild-type and mutant proteins were recorded with an AVIV model 202 CD spectrometer (Lakewood, NJ). Proteins were present at 100 μg/ml in 10 mM sodium phosphate (pH 7.0) and 37°C. The temperature was maintained at 37°C using an internal Peltier thermo-electric temperature controller. Far-UV CD spectra were collected from 195 to 260 nm in a one mm cuvette. Buffer signal was subtracted from all spectra, after which mean residual ellipticity was calculated.

4. *Fluorescence emission spectroscopy*

Fluorescence spectra of the wild-type and mutant HγD-Crys proteins were measured with a Hitachi F-4500 fluorimeter. Samples at pH 7.0 and 37°C contained 10 μg/ml purified protein in 100 mM sodium phosphate, 5 mM DTT, 1 mM EDTA (pH 7.0), and 5.5 M GuHCl where appropriate. Samples at pH 3.0 and 20°C contained 10 μg/ml protein in 100 mM sodium citrate (pH 3.0). Temperature was maintained at 37°C or 20°C using a circulating water bath. An excitation wavelength of 295 nm was used to selectively monitor tryptophan fluorescence. Emission spectra were recorded over a range of wavelengths from 310 to 400 nm using slit widths of 10 nm for both excitation and emission. Fluorescence emission spectra were corrected for the buffer signal.

5. *Equilibrium unfolding and refolding*

Equilibrium unfolding/refolding experiments at pH 7.0 were performed in GuHCl at 37°C. For equilibrium unfolding experiments, purified protein was diluted into solutions containing 0 to 5.5 M GuHCl (purchased as an 8.0 M solution from Sigma-Aldrich, Saint Louis, MO) to give a final protein concentration of 10 μg/ml. All pH 7.0 unfolding samples contained 100 mM sodium phosphate, 5 mM DTT, and 1 mM EDTA (pH 7.0). Unfolding samples were incubated at 37°C for 24 hours to ensure equilibrium had been reached. For equilibrium refolding experiments, purified proteins were first

unfolded in 5.5 M GuHCl at a protein concentration of 100 μg/ml. These unfolded stocks were incubated at 37°C for 5 hours to ensure complete unfolding. The unfolded stock solutions were then diluted into refolding samples, which contained 100 mM sodium phosphate, 5 mM DTT, 1 mM EDTA (pH 7.0) and GuHCl from 0.55 to 5.5 M. Refolding samples were allowed to reach equilibrium by incubation at 37°C for 24 hours prior to recording fluorescence spectra.

Equilibrium unfolding/refolding experiments at pH 3.0 were performed in Urea at 20°C. Samples were set up in a manner identical to that described above for pH 7.0 experiments where the final protein concentration was 10 μg/ml. Unfolding and refolding samples were buffered with 100 mM sodium citrate (pH 3.0) and did not contain DTT or EDTA due to concerns of low solubility at pH 3.0. A 24 hour incubation time was used for both unfolding and refolding samples.

Fluorescence emission spectra were recorded for each unfolding and refolding sample using a Hitachi F-4500 fluorimeter as described above. The concentration of GuHCl or Urea in the unfolding/refolding samples was determined by measuring the refractive index of each sample. Data was analyzed by plotting the concentration of GuHCl for each sample versus the fluorescence intensity at 360 nm and the concentration of GuHCl versus the ratio of fluorescence intensities at 360 and 320 nm (FI 360/320 nm). Equilibrium unfolding/refolding experiments of the wild-type and mutant proteins were performed three times each.

Equilibrium unfolding and refolding data were fit to a two-state model by the methods of Greene and Pace (1974), or a three-state model by the methods of Clark et al. (1993) using the curve fitting feature of Kaleidagraph (Synergy software). The model that best fit the data was selected based on a random distribution of residuals. Transition midpoints, $\Delta G°$ and m values were calculated for all transitions and averaged over the three trials.

6. *Thermal denaturation*

Thermal denaturation experiments were performed using an AVIV model 202 CD spectrometer equipped with an internal Peltier thermo-electric temperature controller

(Lakewood, NJ). All samples contained 100 μg/ml protein in 10 mM sodium phosphate (pH 7.0). The solution conditions differed from equilibrium experiments due to the optical interference of EDTA and DTT and higher concentrations of sodium phosphate. A four mm cuvette with an air-tight screw cap was used for the measurements to prevent loss of sample at high temperatures due to evaporation or boiling over. Changes in molar ellipticity at 218 nm were monitored every 1°C from 25 to 90°C. The samples were allowed to equilibrate at each temperature for one minute before measuring ellipticity over a three second averaging time. The fraction of native (F_N) protein at each temperature was calculated according to equation 1,

$$F_N = (y - y_U)/(y_N - y_U) \quad (1)$$

where y is the ellipticity at 218 nm, y_U is the unfolded/aggregated baseline and y_N is the native baseline. Melting temperatures were calculated by determining the midpoints of the thermal transitions. The experiments were repeated three times for each protein and melting temperatures of the three trials were averaged.

7. *Productive refolding kinetics*

Kinetic refolding experiments were performed by initially unfolding in 5.5 M GuHCl at a protein concentration of 100 μg/ml. These unfolded stocks were incubated at 37°C for five hours to ensure complete unfolding. The unfolded stocks were injected into refolding buffer containing 10 mM sodium phosphate, 5 mM DTT, and 1 mM EDTA (pH 7.0) at 37°C with stirring. A syringe-port injection system with a dead-time of 1 second was used. Changes in fluorescence intensity at 350 nm were monitored with a Hitachi F-4500 fluorimeter over time using an excitation wavelength of 295 nm. The final protein concentration of refolding samples was 10 μg/ml in 1.0 M GuHCl. The fluorescence emission spectra of the samples were measured to ensure that the proteins had completely refolded into native-like conformations. Kinetic refolding data were fit to two and three exponentials using the curve fitting feature of Kaleidagraph and residuals of the fits were calculated (Synergy software). The model with the most random distribution of low-

magnitude residuals was selected as the best fit. Kinetic refolding experiments of the wild-type and mutant proteins were performed three times each and parameters of the fits were averaged.

8. *Unfolding kinetics*

Kinetic unfolding experiments were performed by diluting purified proteins into 5.0 M GuHCl and monitoring fluorescence emission at 350 nm over time. The kinetic unfolding samples contained proteins at final concentrations of 10 μg/ml in 5.0 M GuHCl, 10 mM sodium phosphate, 5 mM DTT, 1 mM EDTA (pH 7.0) at 37°C. Fluorescence emission spectra were recorded at the end of each experiment to ensure that the proteins had fully unfolded. Kinetic unfolding data were fit to two and three exponentials and residuals of the fits were calculated using the curve fitting feature of Kaleidagraph (Synergy software). The best fit was chosen by a random arrangement of residuals. Experiments were performed three times for each protein and parameters were averaged.

D. RESULTS

1. *Protein purification and structure characterization*

To probe the site-specific effects of domain interface glutamine deamidation in HγD-Crys, single and double glutamine to glutamate substitution mutants of Gln54 and Gln143 were constructed. The mutant proteins were purified out of *E. coli* and the stabilities and folding properties of the proteins were determined *in vitro*.

The wild-type and mutant proteins all had exogenous N-terminal His(6)-tags that were utilized during affinity purification. Previous studies of wild-type HγD-Crys with and without the N-terminal His(6)-tag confirmed that the additional sequence did not discernibly affect the structure or thermodynamic and kinetic unfolding/refolding

properties of the protein (Kosinski-Collins and King 2003; Kosinski-Collins et al. 2004). Therefore, the His(6)-tag was not removed for these investigations.

The single and double deamidation mutants of residues Gln54 and Gln143 expressed at levels similar to wild type and behaved like wild type during affinity purification. The mutant proteins were found primarily in the soluble fractions of cell lysates indicating that the mutations did not prevent *in vivo* folding into soluble native-like conformations.

Structures of wild-type HγD-Crys, Q54E, Q143E and Q54E/Q143E were probed with CD and fluorescence spectroscopy. These experiments addressed gross changes in secondary and tertiary structure and were not aimed at distinguishing atomic resolution differences in conformation. Similar to previous results, the Far-UV CD spectrum of wild-type HγD-Crys at pH 7.0 displayed a distinct minimum at 218 nm, indicative of high β-sheet content, and a small shoulder at 208 nm (Andley et al. 1996; Pande et al. 2001). The single and double deamidation mutants all displayed similar minima at 218 nm and a shoulder at 208 nm (Fig. 4-2). CD spectra of the single mutants Q54E and Q143E were indistinguishable from that of wild type in the region analyzed, while the spectra of Q54E/Q143E differed about 20% in intensity at lower wavelengths (Fig. 4-2). This difference may be due to actual changes in conformation or slight differences in solution conditions due to experimental error.

Domain tertiary structures of wild-type HγD-Crys and of the deamidation mutants were surveyed using tryptophan fluorescence emission. HγD-Crys has four intrinsic tryptophans, two per domain, which are buried in the hydrophobic cores of the domains. HγD-Crys also has 14 tyrosines located throughout the protein. Tryptophan fluorescence was selectively monitored by using an excitation wavelength of 295 nm and monitoring emission from 310 to 400 nm. In accord with previous results, native wild-type HγD-Crys displayed a fluorescence emission maximum at 325 nm (Fig. 4-2) and upon denaturing in 5.5 M GuHCl, the fluorescence emission maximum shifted to 350 nm and increased in intensity (Kosinski-Collins and King 2003). The native fluorescence emission spectrum of Q54E was indistinguishable from wild type (Fig. 4-2). In contrast, while the spectra of Q143E and Q54E/Q143E had identical emission maxima as wild type, the overall intensities were increased approximately 10% (Fig. 4-2).

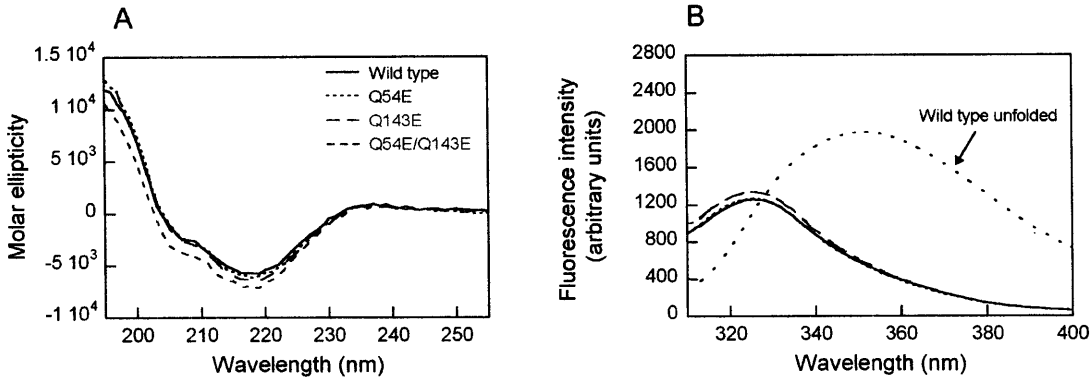

Figure 4-2. (A) Far-UV CD spectra of native wild-type HγD-Crys (solid black), Q54E (short dotted blue), Q143E (long dashed green), Q54E/Q143E (short dashed red). All samples contained 100 μg/ml protein in 10 mM sodium phosphate (pH 7.0) at 37 C. (B) Fluorescence spectra of native wild type and deamidation mutants at pH 7.0 (lines are the same as in (A)) and wild-type H D-Crys denatured in 5.5 M GuHCl (short dotted black). Samples contained 10 μg/ml protein in 100 mM sodium phosphate, 5 mM DTT, 1 mM EDTA (pH 7.0) at 37°C.

Solvent accessible surface areas (SASA) of Gln54 and Gln143 in the crystal structure of HγD-Crys (Basak et al. 2003) were calculated with the program GETAREA 1.1 (Fraczkiewicz and Braun 1998). The SASA of the Gln54 side chain was 20.22 $Å^2$ and the SASA of Gln143 was 3.2 $Å^2$. Relative to the total surface area of the glutamine side chain (143.7 $Å^2$), 86% of the side chain surface area of Gln54 is buried and 97% of the side chain surface area of Gln143 is buried. Despite being highly buried, the side chain amide nitrogen of Gln54 and the side chain amide oxygen of Gln143 form hydrogen bonds to two water molecules each as seen in the crystal structure (Basak et al. 2003).

2. *Equilibrium unfolding/refolding of wild type at pH 7.0*

The thermodynamic stability of wild-type HγD-Crys was analyzed by equilibrium unfolding/refolding using the chemical denaturant GuHCl, at pH 7.0 and 37°C. GuHCl was used in these experiments instead of Urea because wild-type HγD-Crys has been previously shown to resist denaturation in up to 8 M Urea at pH 7.0 (Kosinski-Collins and King 2003). The conformation of proteins in equilibrium experiments was probed with tryptophan fluorescence. To this end, the fluorescence intensity changes at 360 nm and changes in the ratio of intensities at 360 and 320 nm (FI 360/320 nm) were used for data analysis. The FI 360/320 nm ratio data is shown in Fig. 4-3 instead of changes in the intensity at 360 nm because it was visually more descriptive. Nevertheless, parameters derived from the fits of both sets of data were indistinguishable.

Consistent with previous results, the unfolding/refolding transitions of wild-type HγD-Crys were best fit to a three-state model suggesting the presence of a partially folded intermediate in equilibrium with the native and unfolded states (Flaugh et al. 2005a). This intermediate was apparent as a small inflection in the transitions at approximately 2.3 M GuHCl (Fig. 4-3). The first transition had a midpoint (C_m) of 2.2 M GuHCl and an apparent $\Delta G°$ of 7.7 kcal*mol^{-1} (Table 3-1). The second transition had a midpoint of 2.8 M GuHCl and $\Delta G°$ of 8.9 kcal*mol^{-1}. The first transition likely corresponded to unfolding/refolding of the N-td and the second transition to

unfolding/refolding of the C-td (Flaugh et al. 2005a). At pH 7.0 and 37°C, the ΔG° of wild-type HγD-Crys was approximately 16.6 kcal*mol^{-1}.

Also consistent with previous results, wild-type HγD-Crys aggregated upon refolding into buffer (Kosinski-Collins and King 2003). The aggregate was evident by the sharp increase in fluorescence at low concentrations of GuHCl due to right angle light scattering by the aggregate (Fig. 4-3).

3. *Equilibrium unfolding/refolding of deamidation mutants at pH 7.0*

Thermodynamic stabilities of the single and double deamidation mutants were analyzed by the same method described above for wild-type HγD-Crys. Similar to wild type, data for all mutants were analyzed by changes in fluorescence intensity at 360 nm as well as changes in FI 360/320 nm. Parameters derived from fits of both analyses agreed within the error of the experiments. FI 360/320 nm data for the single and double deamidation mutants is shown in Fig. 4-3. As with wild type, the transitions of all mutants were best fit to a three-state model indicating the population of a partially folded intermediate. The presence of an intermediate is more apparent for the deamidation mutants than the wild-type protein as a prominent plateau is present in all transitions. ΔG°'s of the first transition were 5.8 kcal*mol^{-1}*M^{-1} for Q54E, 6.1 kcal*mol^{-1}*M^{-1} for Q143E and 5.3 kcal*mol^{-1}*M^{-1} for Q54E/Q143E (Table 3-1). These values are substantially less than the ΔG° of the first transition for wild type (7.7 kcal*mol^{-1}*M^{-1}). In contrast, the ΔG°'s of the second transition were very similar for wild type and all three deamidation mutants (Table 3-1). Similarly, C_m's for the first transition were decreased for all mutants compared to wild type, while C_m's for the second transition were identical (Table 3-1). Also similar to wild type, all deamidation mutants aggregated upon refolding into buffer (Fig. 4-3).

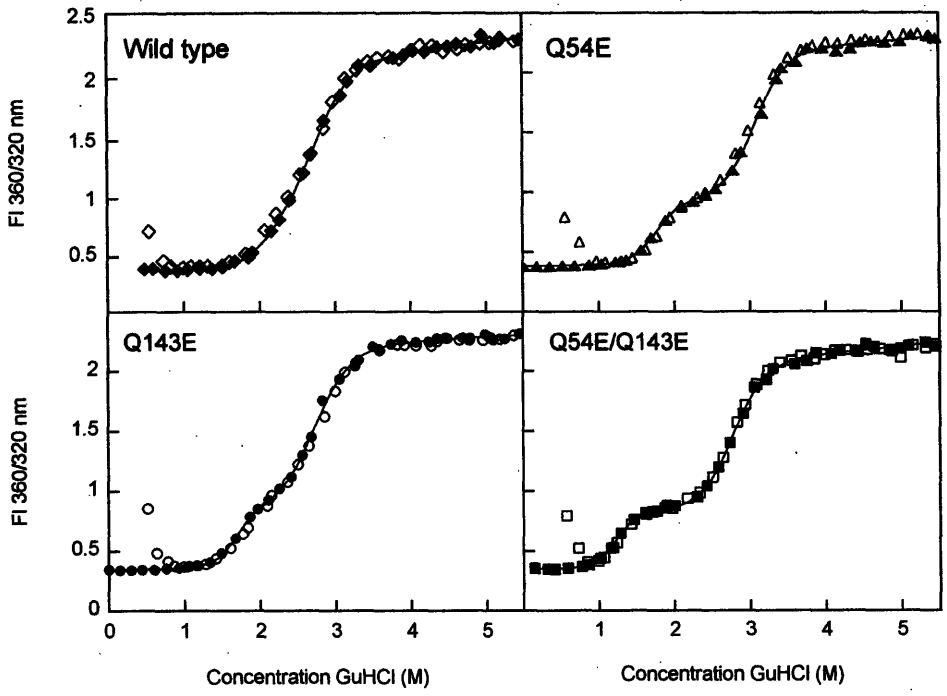

Figure 4-3. Equilibrium unfolding (solid symbols) and refolding (open symbols) of wild-type HγD-Crys (◆), Q54E (▲), Q143E (●) and Q54E/Q143E (■) at pH 7.0 and 37°C in GuHCl. Conformation was probed by fluorescence spectroscopy using an excitation wavelength of 295 nm. Changes in the ratio of fluorescence intensities at 360/320 nm are shown. All samples contained 10 μg/ml in 100 mM sodium phosphate, 5 mM DTT, 1 mM EDTA (pH 7.0) and GuHCl from 0 to 5.5 M at 37°C. Solid lines represent three-state fits of the unfolding data.

Table 4-1. Equilibrium unfolding/refolding and thermal unfolding parameters for wild type and deamidation mutants at pH 7.0

Protein	Thermal Unfolding Transition		Equilibrium Transition 1			Equilibrium Transition 2			Equilibrium Transitions 1 and 2
	T_m [a]	ΔT_m [b]	C_m [c]	Apparent *m* value [d]	Apparent $\Delta G°_{N-I}$ [e]	C_m [c]	Apparent *m* value [d]	Apparent $\Delta G°_{I-U}$ [e]	$\Delta\Delta G_{N-U}$ [f]
Wild type[g]	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	-
Q54E	77.4 ± 0.2	6.4	1.5 ± 0.1	3.8 ± 0.9	5.8 ± 1.7	2.8 ± 0.1	2.9 ± 0.1	8.3 ± 0.1	2.5
Q143E	75.7 ± 0.1	8.1	1.7 ± 0.1	3.6 ± 0.1	6.1 ± 0.3	2.8 ± 0.1	3.0 ± 0.2	8.5 ± 0.9	2.0
Q54E/Q143E	71.0 ± 0.1	12.8	1.2 ± 0.1	4.4 ± 0.4	5.3 ± 0.6	2.8 ± 0.1	2.9 ± 0.1	8.3 ± 0.3	3.0

[a] Midpoints of melting transitions monitored by CD in units of °C.

[b] $\Delta T_m = T_{m\ \text{Wild type}} - T_{m\ \text{Mutant}}$ 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] $\Delta\Delta G_{N-U} = (\Delta G°_{N-I\ \text{Wild type}} + \Delta G°_{I-U\ \text{Wild type}}) - (\Delta G°_{N-I\ \text{Mutant}} + \Delta G°_{I-U\ \text{Mutant}})$ in units of kcal*mol^{-1}*M^{-1}.[g] Equilibrium parameters of wild type are from Flaugh et al. 2005a.

4. *Thermal denaturation at pH 7.0*

Given that the ionic character of GuHCl may mask or interfere with the effects of the deamidation charge change, stabilities of wild-type HγD-Crys and the deamidation mutants were further probed in a low ionic strength buffer by thermal denaturation. These experiments assessed stabilities by monitoring changes in ellipticity at 218 nm every 1°C from 25 to 90°C. Wild-type HγD-Crys and all deamidation mutants aggregated at high temperatures as seen by the presence of protein precipitate subsequent heating to 90°C. Therefore, loss of CD signal at high temperatures was probably due to a combination of unfolding and aggregation initiated by unfolding. Regardless of the origin of the signal change, increase in ellipticity at 218 nm was interpreted as a loss of native structure and therefore a satisfactory measure of stability.

Under the conditions employed here the thermal denaturation transition of wild-type HγD-Crys appeared two-state with no significant evidence of a stable intermediate (Fig. 4-4). The melting temperature (T_m) of wild type was 83.8 °C (Table 4-1). Similarly, the thermal denaturation transitions of all deamidation mutants also appeared two-state (Fig. 4-4). However, all of the mutants unfolded at lower temperatures than wild type with T_m values of 77.4 °C for Q54E, 75.7 °C for Q143E and 71.0 °C for Q54E/Q143E (Table 4-1). Thermal unfolding of the deamidation mutants occurred over a narrower temperature range than wild type suggesting that the mutants may have unfolded more cooperatively.

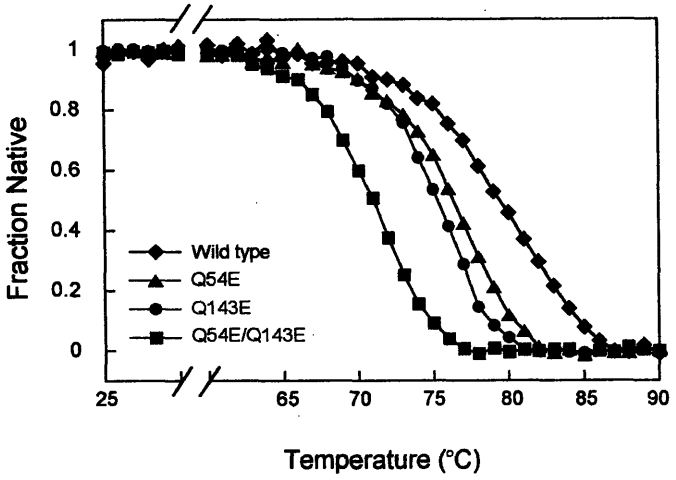

Figure 4-4. Thermal denaturation transitions for wild-type HγD-Crys (◆), Q54E (▲), Q143E (●) and Q54E/Q143E (■) as monitored by changes in ellipticity at 218 nm at increasing temperature. All samples contained 100 μg/ml protein in 10 mM sodium phosphate (pH 7.0).

5. Equilibrium unfolding/refolding wild type at pH 3.0

Equilibrium experiments of wild type, Q54E, Q143E and Q54E/Q143E were performed at pH 3.0 to test the role of charge in destabilization at pH 7.0. The thermodynamic stability of wild-type HγD-Crys at pH 3.0 was determined by equilibrium unfolding/refolding experiments in Urea at 20°C. These experiments were performed at 20°C rather than 37°C because HγD-Crys has been seen to form amyloid fibers at pH 3.0 and 37°C (Katerina Papanikiolopoulou personal communication). Urea was used instead of GuHCl because at pH 3.0 and 20°C, low concentrations of GuHCl also caused polymerization into an amyloid state (K. Papanikiolopoulou, S.L. Flaugh, I.A. Mills and J.King, unpubl.). Finally, pH 3.0 was chosen because wild-type HγD-Crys was partially unfolded at lower pH even in the absence of denaturant (data not shown).

There was no evidence of aggregation or amyloid fiber formation in unfolding samples under the conditions employed here (100 mM sodium citrate (pH 3.0) plus Urea from 0 to 6 M at 20°C). However, all proteins aggregated upon refolding into concentrations of Urea below 2.5 M at pH 3.0 (data not shown). Fluorescence spectra of soluble protein present in the aggregation samples subsequent a high speed spin did not match spectra of the native proteins at pH 3.0 (data not shown). The addition of 5 mM DTT and 1 mM EDTA did not result in recovery of native protein in refolding samples (data not shown). The morphology of the pH 3.0 refolding aggregates and the irreversibility of the reactions will be the subject of future investigations. Given the irreversibility of the reaction at pH 3.0, $\Delta G°$ values were not calculated and instead transition midpoints were used as an estimate of stability (Table 4-2). To reduce complexity of the plots, equilibrium refolding data is not shown in Figure 4-5.

The equilibrium unfolding transition of wild-type HγD-Crys at pH 3.0 was best fit to a two-state model that assumes direct transition between the native and unfolded states (Fig. 4-5). The C_m of the pH 3.0 equilibrium unfolding transition was 2.4 M Urea (Table 4-2). Discrepancy between the number of states observed for the pH 7.0 and pH 3.0 data

is likely due to different effects of pH on the stabilities of the N- and C-tds. For instance, if the stability of each domain of wild-type HγD-Crys are more similar at pH 3.0 than pH 7.0, the partially folded intermediate would not be populated over as wide a range of denaturant concentrations if at all, and thus would not be observable at pH 3.0.

6. *Equilibrium unfolding/refolding of deamidation mutants at pH 3.0*

Stabilities of the mutants Q54E, Q143E and Q54E/Q143E were also measured at pH 3.0, which is below the pK_a of glutamate in aqueous solution. However, the introduced glutamate side chains are in proximity to polar main chain and side chain atoms as well as water molecules that may alter pK_a and cause the glutamate side chains to not be fully protonated at pH 3.0 (Fitch et al. 2002; Forsyth et al. 2002). The pH 3.0 equilibrium unfolding transitions of the deamidation mutants were best fit to two-state models similar to wild type (Fig. 4-5). The transition midpoints of the deamidation mutants were identical or comparable to that of wild-type HγD-Crys. The C_m for Q54E was 2.4 M, the C_m for Q143E was 2.3 M and the C_m for Q54E/Q143E was 2.1 M (Table 4-2).

Equilibrium unfolding of the double alanine mutant, Q54A/Q143A, was also measured at pH 3.0 as a control protein that is destabilized at pH 7.0 but has the same number of acidic residues as wild type (Flaugh et al. 2005b). At pH 3.0, the unfolding transition of Q54A/Q143A deviated significantly from that of wild type as it was best fit to a three state model with C_m's of 1.3 and 2.5 M for the first and seconds transitions, respectively (Fig. 4-5, Table 4-2).

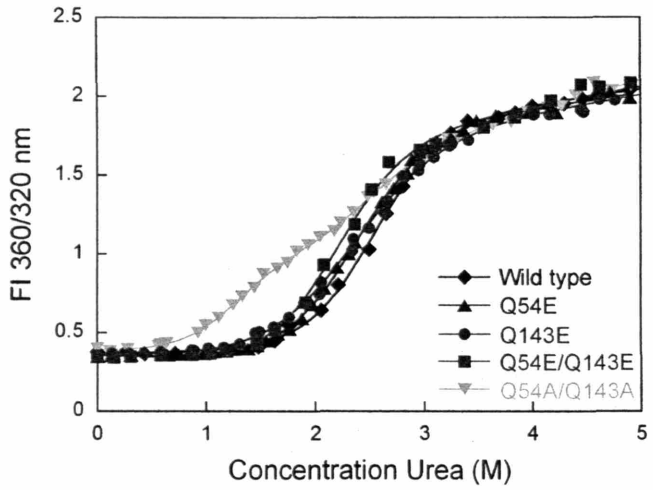

Figure 4-5. Equilibrium unfolding of wild-type HγD-Crys (◆), Q54E (▲), Q143E (●), Q54E/Q143E (■) and Q54A/Q143A (▼) in Urea at pH 3.0 and 20°C. Conformation was probed by fluorescence emission. Data is presented as the ratio of fluorescence intensities at 360/320 nm using an excitation wavelength of 295 nm. Protein was present at 10 μg/ml in 100 mM sodium citrate (pH 3.0) and Urea from 0 to 5.0 M at 20°C.

Table 4-2. Equilibrium unfolding/refolding parameters for wild-type and mutant HγD-Crys at pH 3.0 and 20°C in Urea.

	pH 3.0 Transition 1	pH 3.0 Transition 2
Protein	C_m[a]	C_m[a]
Wild type	2.4 ± 0.1	-
Q54E	2.4 ± 0.1	-
Q143E	2.3 ± 0.1	-
Q54E/Q143E	2.1 ± 0.1	-
Q54A/Q143A	1.3 ± 0.1	2.5 ± 0.1

[a] Transition midpoints of equilibrium unfolding transitions in units of M Urea.

7. *Kinetic refolding of wild type and deamidation mutants at pH 7.0*

The effects of glutamine deamidation on the refolding kinetics of HγD-Crys were determined by performing productive kinetic refolding experiments at pH 7.0. Proteins were diluted from 5.0 to 1.0 M GuHCl and burial of the tryptophans was monitored as a decrease in the fluorescence intensity at 350 nm over time. Proteins were refolded to 1.0 M GuHCl as this was the lowest concentration of GuHCl where aggregation did not compete with productive refolding (Kosinski-Collins and King 2003). A syringe-port injection system was used instead of a stopped-flow apparatus because the major refolding transitions of HγD-Crys occur on a second and not millisecond time scale (Kosinski-Collins et al. 2004). Potential sub-second refolding intermediates were not addressed in these experiments.

Previous analyses of wild-type HγD-Crys refolding kinetics revealed the presence of a single major intermediate that likely has a folded C-td and unfolded N-td (Kosinski-Collins et al. 2004). It was further found that the domain interface of the folded C-td

nucleates refolding of the N-td (Flaugh et al. 2005a; 2005b). Refolding kinetics of the deamidation mutants did not fit well to two exponentials (three states) as was previously described for wild type. This observation prompted a more in-depth analysis of wild-type HγD-Crys refolding kinetics. The refolding kinetics of wild type were fit to two and three exponentials and residuals were examined to select the best fit (Fig. 4-6). Fitting to three exponentials (four states) resulted in a more random arrangement of residuals (Fig. 4-6). Compared to the two exponential fit, the residuals of the three exponential fit were improved for data in the range of 0 to 200 seconds, where the major structural transformations take place. As depicted in equation 2, wild-type HγD-Crys populated two major intermediates during refolding.

$$\text{Unfolded} \xrightarrow{k_{R1}} I_{R1} \xrightarrow{k_{R2}} I_{R2} \xrightarrow{k_{R3}} \text{Native} \qquad (2)$$

First, a rapid decrease in fluorescence was observed with a half-time ($t_{1/2}$) of 8 seconds which corresponded to refolding from the unfolded (U) state into the first intermediate (I_{R1}). Refolding into the second intermediate (I_{R2}) then occurred with a $t_{1/2}$ of 35 seconds and finally refolding into the native (N) conformation occurred with a $t_{1/2}$ of 130 seconds (Table 4-3).

Refolding rates of the deamidation mutants were analyzed in an analogous manner as that described for wild type. All of the mutants exhibited a rapid decrease in fluorescence upon diluting into 1.0 M GuHCl followed by a slower decrease in fluorescence (Fig. 4-7). All of the mutant proteins required significantly longer times than wild type to refold completely and were best fit by three exponentials similar to wild type (Fig. 4-7). For all mutants, the rates of refolding into I_{R1} and I_{R2} were similar to wild-type while the rate of the final transition from I_{R2} to N was decreased (Table 4-3). The $t_{1/2}$ of the I_{R2} to N transition was increased more than 19 times for Q54E/Q143E (2500 seconds) as compared to wild type (130 seconds).

Table 4-3. Productive kinetic refolding parameters for wild-type HγD-Crys and deamidation mutants.

Protein	Kinetic refolding transition 1			Kinetic refolding transition 2			Kinetic refolding transition 3		
	k_{R1}[a]	$\Delta\Delta G_{\ddagger R1}$[b]	$t_{1/2}$[c]	k_{R2}[a]	$\Delta\Delta G_{\ddagger R2}$[b]	$t_{1/2}$[c]	k_{R3}[a]	$\Delta\Delta G_{\ddagger R3}$[b]	$t_{1/2}$[c]
Wild type	0.10 ± 0.04	-	8 ± 3	0.023 ± 0.011	-	35 ± 17	0.0053 ± 0.0003	-	130 ± 10
Q54E	0.06 ± 0.01	0.32	13 ± 2	0.013 ± 0.003	0.35	54 ± 11	0.00038 ± 0.00006	1.62	1900 ± 270
Q143E	0.05 ± 0.02	0.43	17 ± 5	0.030 ± 0.014	-0.16	29 ± 17	0.0021 ± 0.0001	0.57	350 ± 80
Q54E/Q143E	0.04 ± 0.01	0.56	16 ± 3	0.011 ± 0.002	0.45	67 ± 12	0.00028 ± 0.00006	1.81	2500 ± 480

[a] Kinetic rate constants in units of sec^{-1}.

[b] $\Delta\Delta G_{\ddagger R} = -RT*\ln(k_{R\ Mutant}/k_{R\ Wild\ type})$ in units of $kcal*mol^{-1}*M^{-1}$.

[c] Half-time in units of sec.

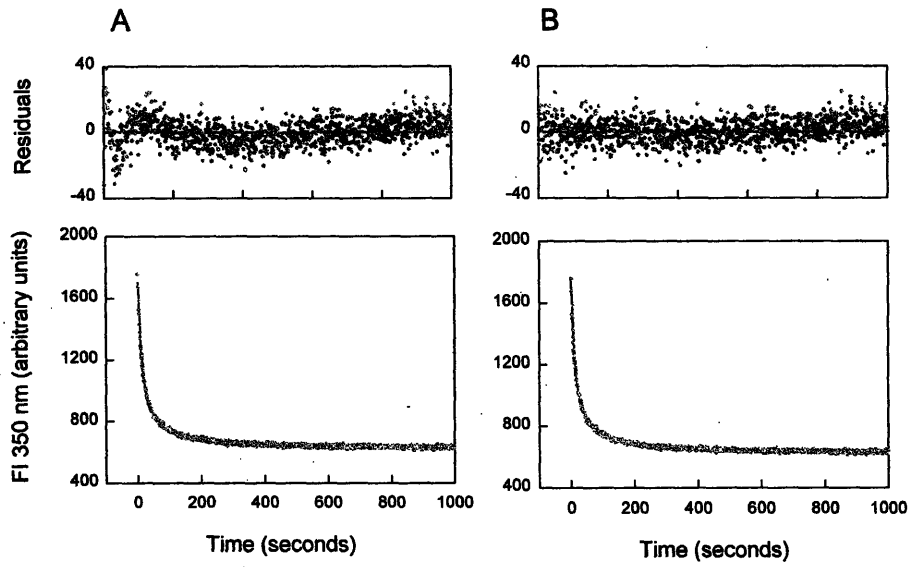

Figure 4-6. Productive kinetic refolding data of wild-type HγD-Crys fit to two (A) and three exponentials (B). Protein was initially unfolded in 5.5 M GuHCl and diluted into 100 mM sodium phosphate, 5 mM DTT, 1 mM EDTA (pH 7.0) at 37°C to give a final GuHCl concentration of 1.0 M. Changes in fluorescence intensity at 350 nm were monitored over time using an excitation wavelength of 295 nm. Fits of the data are shown as black lines and residuals of the fits are shown above.

8. *Kinetic unfolding of wild type and deamidation mutants at pH 7.0*

The effects of glutamine deamidation on the unfolding kinetics of HγD-Crys were analyzed by performing kinetic unfolding experiments. Native proteins were rapidly diluted into 5.0 M GuHCl at pH 7.0 and the decrease in fluorescence intensity at 350 nm was monitored over time in order to follow the solvent-exposure of buried tryptophans. As with the kinetic refolding experiments, a syringe-port injection system was used because major transitions during unfolding occur on a second time-scale (Kosinski-Collins et al. 2004). It was previously shown that the kinetic unfolding transitions of wild-type HγD-Crys were best fit to three exponentials suggesting an unfolding pathway with two major intermediates as depicted in equation 3 (Kosinski-Collins et al. 2004).

$$\text{Native} \xrightarrow{k_{U1}} I_{U1} \xrightarrow{k_{U2}} I_{U2} \xrightarrow{k_{U3}} \text{Unfolded} \qquad (3)$$

The kinetic unfolding transitions of wild type measured here agreed well with the previous analysis. A rapid increase in fluorescence at 350 nm corresponding to the N to the first intermediate (I_{U1}) transition was followed by a slower increase corresponding to unfolding into the second intermediate (I_{U2}) and finally U (Fig. 4-7). The unfolding transitions were best fit to three exponentials with $t_{1/2}$ values of 0.79, 33 and 200 seconds for the three phases, respectively (Table 4-4).

Kinetic unfolding rates of Q54E, Q143E and Q54E/Q143E were measured in a manner analogous to that described for wild-type HγD-Crys. Kinetic unfolding transitions for all mutants exhibited a rapid increase in fluorescence followed by a slower increase, which were best fit to three exponentials (Fig. 4-7). The $t_{1/2}$ values for unfolding into I_{U1} and I_{U2} were increased for all mutants compared to wild type. In contrast, the $t_{1/2}$ values for the I_{U2} to U transition were similar for the mutant and wild-type proteins (Table 4-4).

Table 4-4. Kinetic unfolding parameters for wild-type HγD-Crys and deamidation mutants.

	Kinetic unfolding transition 1			Kinetic unfolding transition 2			Kinetic unfolding transition 3		
Protein	k_{U1}[a]	$\Delta\Delta G_{\ddagger U1}$[b]	$t_{1/2}$[c]	k_{U2}[a]	$\Delta\Delta G_{\ddagger U2}$[b]	$t_{1/2}$[c]	k_{U3}[a]	$\Delta\Delta G_{\ddagger U3}$[b]	$t_{1/2}$[c]
Wild type	0.88 ± 0.01	-	0.79 ± 0.02	0.024 ± 0.01	-	33 ± 13	0.0035 ± 0.0001	-	200 ± 10
Q54E	1.5 ± 0.3	0.32	0.47 ± 0.08	0.11 ± 0.01	0.94	6.6 ± 0.4	0.0035 ± 0.0001	0.04	200 ± 10
Q143E	1.9 ± 0.5	0.47	0.38 ± 0.10	0.19 ± 0.08	1.27	4.2 ± 1.7	0.0038 ± 0.0002	-0.05	180 ± 10
Q54E/Q143E	1.7 ± 0.5	0.41	0.43 ± 0.09	0.35 ± 0.10	1.65	2.1 ± 0.7	0.0038 ± 0.0004	-0.05	180 ± 20

[a] Kinetic rate constants in units of sec^{-1}.

[b] $\Delta\Delta G_{\ddagger U} = -RT*\ln(k_{U\ Wild\ type}/k_{U\ Mutant})$ in units of $kcal*mol^{-1}*M^{-1}$.

[c] Half-time in units of sec.

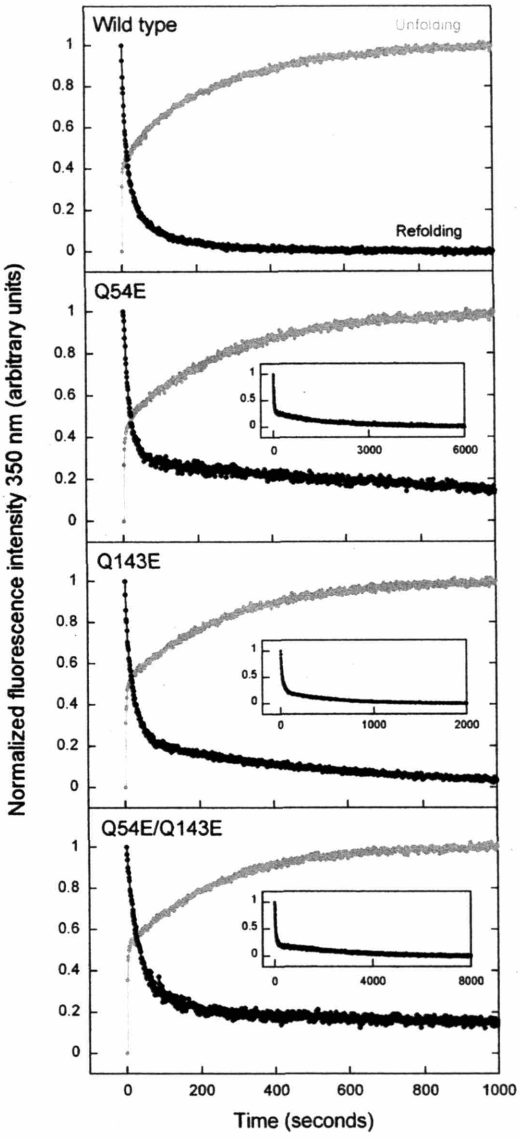

Figure 4-7. Normalized kinetic unfolding (gray) and productive kinetic refolding (black) transitions of wild-type H D-Crys and deamidation mutants. For unfolding, native proteins were diluted into 5.0 M GuHCl, and buffer described in Fig. 6 at 37°C. For productive refolding, proteins were first unfolded in 5.5 M GuHCl and subsequently diluted into buffer described in Fig. 6 to give a final GuHCl concentration of 1.0 M at 37°C. Insets shown for Q54E, Q143E and Q54E/Q143E display the refolding transitions over the extended times that were required for complete refolding.

E. DISCUSSION

Glutamine and asparagine deamidation are very common forms of covalent damage found in the crystallins of aged and cataractous lenses. Chemically, deamidation results in replacement of an amide group with a carboxyl group, and may cause side chain racemization and backbone isomerization. Four products are derived from the deamidation of L-asparagine, L- and D-aspartate and L- and D-isoaspartate (Takemoto et al. 2001). The ratio of products formed depends on steric hindrance induced by the protein backbone (Takemoto et al. 2001). Deamidation of glutamine likely occurs by a similar mechanism where the four products formed are L- and D-glutamate and L- and D-isoglutamate.

The extent of deamidation at different sites in the crystallins is highly dependent on the structural context of the residues, such as solvent accessibility of the asparagine or glutamine and the size and flexibility of the amino acid immediately following it (Geiger and Clarke 1987; Lapko et al. 2002). In human γS crystallin, a close relative of HγD-Crys, high percentages of deamidation were seen for residues with surface accessibilities greater than 80 $Å^2$ and low percentages for residues with accessibilities less than 80 $Å^2$ (Lapko et al. 2002).

Many sites of deamidation have been identified in HγD-Crys from cataractous and non-cataractous lenses of various ages (Hanson et al. 1998; Hanson et al. 2000; Searle et al. 2005). Gln54 was identified as a potential site of deamidation in the water-soluble fraction of non-cataractous lenses, reaching 65% deamidation in a 45 year old lens (Hanson et al. 1998). In contrast, another study found no damage in HγD-Crys from the water-insoluble fraction of non-cataractous lenses (Hanson et al. 2000). A third study of the proteins from a 93-year old cataractous lens found deamidation of HγD-Crys at a variety of sites, some of which agreed with previous results and others which were newly identified (Searle et al. 2005). Deamidation is a particularly difficult modification to detect because it results in a mass change of only one Da. Given the discrepancy between studies, the low numbers of sample sizes, and the difficulty in identifying deamidation, it is plausible that further sites of deamidation in HγD-Crys have yet to be identified. We focused our investigation on glutamines in the domain interface of HγD-

Crys, a structurally significant location of the protein. While Gln54 and Gln143 are highly buried in the native state, they are each within hydrogen bonding distances to two water molecules, critical reactants in non-enzymatic deamidation.

1. *Effects of interface glutamine deamidation on structure*

Single and double glutamine to glutamate substitutions of the interface residues, Gln54 and Gln143 did not prevent *in vivo* folding into soluble native-like conformations. Additionally, the CD and fluorescence spectra of the mutant proteins were similar to wild type suggesting that they had similar secondary and tertiary structures. Despite similar emission maxima, native fluorescence emission spectra of Q143E and Q54E/Q143E were increased about 10% in intensity compared to wild type (Fig. 4-2). This change in fluorescence intensity may be due to relaxation of the anomalous native state quenching phenomenon that causes the fluorescence intensity of HγD-Crys to be lower in the native state than the unfolded state. Native-state fluorescence quenching has been previously described for HγD-Crys and other β- and γ-crystallins (Kim et al. 2002; Bateman et al. 2003; Kosinski-Collins and King 2003; Kosinski-Collins et al. 2004). Increase in the fluorescence emission of Q143E and Q54E/Q143E may have been caused by relaxation of the native state quenching phenomenon, structural changes or deviations in solution conditions.

Overall, the data observed here suggests that deamidation of Gln54 and Gln143 into L-glutamate does not cause considerable changes in the structure of HγD-Crys. High resolution crystal structures of the mutant proteins will be necessary to determine subtle changes in conformation not detectable by the spectroscopic methods employed here.

2. *Effects of interface glutamine deamidation on stability*

At pH 7.0 and 37°C, wild-type HγD-Crys exhibits three-state equilibrium unfolding/refolding transitions with an intermediate populated that likely has a structured C-td and unstructured N-td (Kosinski-Collins et al. 2004; Flaugh et al. 2005a).

According to this model the first transition represents unfolding/refolding of the N-td and the second transition represents unfolding/refolding of the C-td (Flaugh et al. 2005a).

At pH 7.0 the single and double interface deamidation mutants were all destabilized compared to wild type. The $\Delta G°$ and C_m of the first transition were decreased for the mutants, but the second transition was not affected. Therefore, deamidation of the interface glutamines appeared to destabilize the N-td but not the C-td. Single and double alanine mutagenesis of Gln54 and Gln143 also caused selective destabilization of the N-td (Flaugh et al. 2005b). Double mutant cycle analysis revealed that Gln54 and Gln143 have very low interaction energy of ~0.7 kcal/mol. However, mutating the residues singly and doubly still resulted in notable destabilization. These results suggest that the residues are not positioned at the top of the domain interface to energetically clasp the domains together, but may instead function to shield the central domain interface hydrophobic cluster from solvent (Flaugh et al. 2005b).

According to the model described above, mutating Gln54 and Gln143 to glutamate should not have a large effect on stability because the polar nature of the residues is maintained. However, significant destabilization was observed for the deamidation mutants, probably due to the introduction of negative charges. To test this hypothesis, we performed equilibrium unfolding/refolding experiments at acidic pH, below the pK_a of the free glutamate side chain in aqueous solution. The lowest practical pH for these experiments was 3.0, because wild-type HγD-Crys partially unfolded at lower pH even in the absence of denaturant. In addition, Urea was used instead of GuHCl and lower temperatures were employed for the pH 3.0 experiments to avoid amyloid fiber formation that was observed in the presence of GuHCl (K. Papanikiolopoulou, S.L. Flaugh, I.A. Mills and J. King, unpubl.).

The pH 3.0 equilibrium unfolding transition of wild type appeared two-state under the conditions employed here. Equilibrium unfolding transitions of Q54E, Q143E and Q54E/Q143E at pH 3.0 were comparable to wild type with similar transition midpoints. Thus, the glutamine to glutamate mutations did not notably affect stability when the side chains were protonated. In contrast, the double alanine mutant Q54A/Q143A was destabilized at both pH 7.0 and 3.0 where it was best fit to a three-state transition. These results suggest that wild type and the deamidation mutants do populate an equilibrium

intermediate at pH 3.0, but it is not readily observable due to similarities in the transition midpoints of the two domains at pH 3.0.

The results observed here indicate that subsets of the glutamates were protonated at pH 3.0. These data suggest that the positively charged guanidinium ions present in denaturation samples did not completely mask the negatively charged side chains of the introduced glutamates as has been previously described for acetylated ferricytochrome c and several coiled coil proteins (Hagihara et al. 1994; Monera et al. 1994b; 1994a; Kohn et al. 1995). The inability of guanidinium ions to completely or partially mask the glutamate side chains may be due to the low solvent accessibilities of these residues in the native state. In all probability, destabilization of HγD-Crys by deamidation of the interface glutamines occurred due to introduction of a negative charge into these buried locations. Introducing charged residues into the hydrophobic interior of staphylococcal nuclease has been previously shown to destabilize the protein in a pH-dependent manner (Stites et al. 1991; Nguyen et al. 2004)

3. *Thermal stability of wild-type HγD-Crys and deamidation mutants*

To avoid ionic effects of GuHCl, we studied stability under low ionic conditions using thermal denaturation. The γ-crystallins exhibit high thermal stabilities (Sen et al. 1992; Fu and Liang 2002; Evans et al. 2004). Sen et al. (1992) reported a T_m of approximately 73°C for bovine γD crystallin and Evans et al. (2004) found that HγD-Crys resisted thermal aggregation after incubating at 70°C for up to ten minutes. We measured thermal stabilities of wild-type HγD-Crys and the deamidation mutants by observing changes in ellipticity at 218 nm at increasing temperatures using a CD spectometer. By this method wild-type HγD-Crys had a very high T_m of 83.8°C. The deamidation mutants all had lower T_m's where Q54E was the least destabilized, Q143E was in between and Q54E/Q143E was the most destabilized.

The thermal denaturation data contradict the equilibrium data at pH 7.0 and 37°C where, within error, Q54E and Q143E were destabilized similarily and Q54E/Q143E was most destabilized. This discrepancy may be due to differences in the response of the N- and C-tds to thermal denaturation and GuHCl-induced unfolding. The thermal

denaturation transitions of the wild-type and mutant proteins all appeared two-state while the pH 7.0 equilibrium transitions were clearly three-state, suggesting that the protein unfolds more cooperatively in response to heat than GuHCl. This may be caused by the N-td and C-tds having very similar T_m's or a cooperative unfolding mechanism not demonstrated in GuHCl equilibrium experiments. Therefore, deamidation could have very different effects on stabilities of the domains under the two denaturation conditions. Alternatively, it is still possible that the differences are due to guanidium ion interference in the equilibrium experiments. Similarly, unfolding and aggregation both occur in thermal experiments, which may obscure detection of intermediates.

Thermal stabilities of wild type and the deamidation mutants were not measured at pH 3.0 because wild-type HγD-Crys has been observed to form amyloid fibers at pH 3.0 and temperatures higher than 20°C (K. Papanikiolopoulou, S.L. Flaugh, I.A. Mills and J. King, unpubl.).

4. *Kinetic unfolding and refolding*

High thermodynamic stability is a critical property of the lens crystallins, which allows the proteins to remain folded despite advanced age. It has furthermore been suggested that the lens crystallins also utilize *kinetic stability* to prevent unfolding, where a there exists an exceptionally high kinetic barrier to unfolding (Das and Liang 1998; Jaenicke and Slingsby 2001; MacDonald et al. 2005). The effects of deamidation on the free-energy barrier between the native and intermediate states of HγD-Crys were examined by performing productive kinetic refolding and unfolding experiments. As described above the kinetic unfolding/refolding pathway of HγD-Crys has been investigated in detail by fluorescence (Kosinski-Collins et al. 2004; Flaugh et al. 2005a; 2005b).

Kinetic *refolding* transitions were best fit to three transitions suggesting two major intermediates and rates of the third transitions were the notably different for the mutants (Table 4-3). Given the previously described nucleation-dependent sequential domain refolding pathway, we hypothesize that the first and second transitions monitored refolding of the C-td and the third transition monitored refolding of the N-td. Kinetic

unfolding transitions of wild type and the deamidation mutants were all best fit by three exponentials and the deamidation mutants had significant effects on rates of the second unfolding transition only (Table 4-4). These data suggest that the first and second transitions were N-td unfolding and the third transition was C-td unfolding (Kosinski-Collins et al. 2004).

Previous analyses of proteins adopting the βγ-crystallin domain fold have established that the β-strands contributing to the Greek key motif nearest the domain interface, as well as a β-hairpin between these strands are highly stabilizing and may act as a nucleus for folding (Bagby et al. 1998; MacDonald et al. 2005). Given this, the partially folded intermediate (I_{R1}) that had a partially folded C-td may have be structured in the region near the domain interface and the intermediate (I_{U1}) with a partially folded N-td may have been unstructured in the Greek key motif distant from the domain interface (Fig. 4-8).

According to the hypotheses described above, and contrary to what might be expected for a reversible folding reaction, neither I_{U1} and I_{R2} nor I_{U2} and I_{R1} were equivalent. I_{U1} likely had a partially-folded N-td where domain interface contacts were maintained while I_{R2} likely had a folded C-td and unfolded N-td. The I_{U2} species was probably analogous to the I_{R2} species. Finally, I_{R1} likely had a partially folded C-td and fully unfolded N-td. The reason for these discrepancies may be due to differences in relative sizes of kinetic barriers on the forward and reverse pathways. For instance, during refolding, the U to I_{R1} transition was slow enough to detect, but during unfolding the I_{R1} to U transition may be too fast to detect. The reverse may be true of the N to I_{U1} transition explaining why it was only observed in during unfolding. A schematic diagram describing potential qualitative relationships between the unfolding and refolding intermediates is shown in Fig. 4-8. It is also possible that the discrepancies are due to inherent differences in the unfolding and refolding pathways, such as the presence of off-pathway intermediates or parallel folding channels as has been previously described for the alpha-subunit of tryptophan synthase, an α/β-barrel protein (Bilsel et al. 1999).

Comparing the kinetic unfolding and refolding data to thermodynamic stabilities determined by equilibrium experiments, it was possible to describe the effects of interface deamidation on the various stages of the unfolding and refolding pathways (Fig.

4-8). Consistent with what would be expected for mutations that disrupt the domain interface, the most significant effects were observed for transitions of the N-td. The native state was markedly destabilized by the deamidation mutants ($\Delta\Delta G_{N-U}$ = 2.0-3.0 kcal*mol^{-1}*M^{-1}). It was not possible to determine the stability of the kinetic unfolding intermediate I_{U1}. However, given that the domains were likely still making contact, we expect that the mutants decreased the stability of I_{U1} similar to the native state. The height of the free energy barrier between N and I_{U1} was decreased slightly for the deamidation mutants ($\Delta\Delta G_{\ddagger 1}$ = 0.32-0.47 kcal*mol^{-1}*M^{-1}). In contrast, the height of the free energy barrier between I_{U1} and I_{U2} was markedly decreased for the mutants ($\Delta\Delta G_{\ddagger 2}$ = 0.94-1.65 kcal*mol^{-1}*M^{-1}), suggesting that this was the transition during which domain interface contacts were lost. On the refolding pathway, the free energy barrier between I_{R2} and N was increased considerably for the mutants (0.57-1.81 kcal*mol^{-1}*M^{-1}). Since it was not possible to unambiguously correlate intermediates on the unfolding and refolding pathways, we were not able to quantitatively compare changes in their respective free energy barriers. Regardless, the data here indicate that the deamidation mutants decreased the free energy barrier between N and I_{U2}/I_{R2} enough to reduce the rate of refolding, but not as much as the native state was destabilized so that an increase in the rate of unfolding was also observed. In other words, in the model drawn in Fig. 4-8, $\Delta\Delta G_{\ddagger 2}$ was less than $\Delta\Delta G_{N-U}$ for all of the mutants.

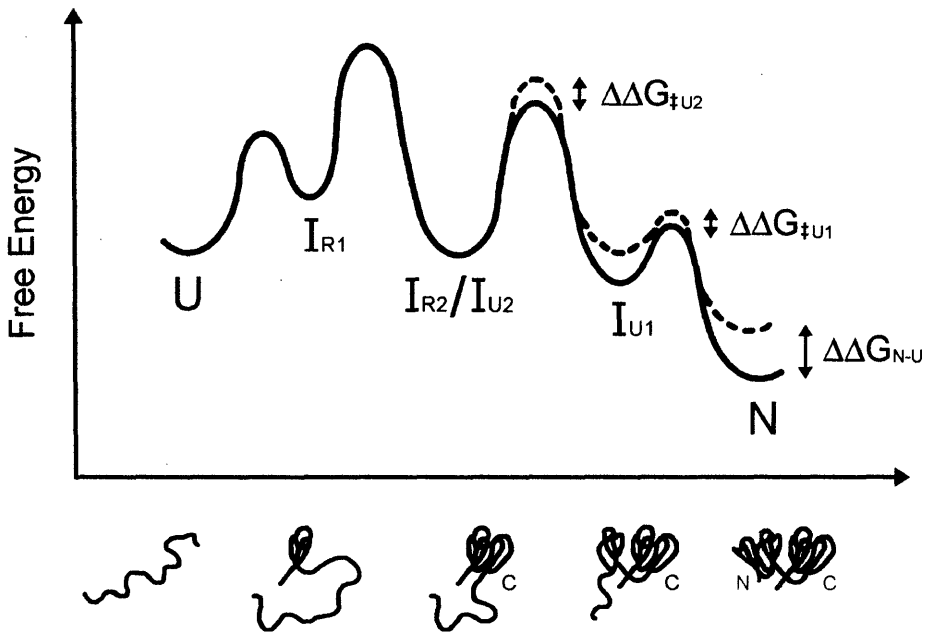

Figure 4-8. Schematic reaction coordinate diagram describing qualitative relationships between the equilibrium and kinetic unfolding/refolding intermediates of HγD-Crys. Theorized effects of the interface deamidation mutations are shown as dashed lines. Line drawings of kinetic intermediates are speculative conformations when the N- and C-terminal domains were partially folded.

5. *Covalent damage and lens transparency*

Cataract is associated with the presence of insoluble inclusions of covalently-damaged crystallin proteins. Studying the effects of covalent damage on important physiological properties of the crystallins may give insight into *in vivo* alterations that could induce aggregation or insolubility. A general feature of many protein deposition diseases is aggregation from a partially-folded or non-native conformation (Mitraki 1989; Wetzel 1994; Booth et al. 1997; Jiang et al. 2001). In order for aggregation to occur, it is first necessary to populate the problematic conformation. For the crystallins, destabilization of the native states and lowering the kinetic unfolding barrier by covalent damage may cause partial unfolding into aggregation-prone states. Interestingly, deamidation has also been shown to induce aggregation of amyloidgenic peptides (Nilsson et al. 2002; Nilsson and Dobson 2003). Alternatively, covalent damage may alter surfaces of the crystallins in such a way that the short-range order is disrupted or insolubility is induced. Congenital cataracts caused by mutations in the gene encoding HγD-Crys instigate cataract formation through such mechanisms where native HγD-Crys associates by intermolecular disulfide bond formation, crystallization or precipitation (Kmoch et al. 2000; Pande et al. 2000; Pande et al. 2001; Evans et al. 2004). A congenital cataract mutant of human γC crystallin (T5P) had both altered conformation and was destabilized, suggesting that association of non-native conformations caused cataract (Fu and Liang 2002). While the exact means by which these mutations cause cataract almost certainly do not explain the mechanism of mature-onset cataract formation, these examples do provide evidence that alteration of surface properties, or native-state destabilization can promote self-association or aggregation resulting in cataract.

Previous analyses have determined that the effects of deamidation on the structures, stabilities and solubilities of the crystallins are dependent on the structural context of the damage (Kim et al. 2002; Lampi et al. 2002; Gupta and Srivastava 2004b; 2004a; Lampi et al. 2006). In the β-crystallins, deamidation of domain interface glutamines has a large effect on stability where a monomeric intermediate is populated in equilibrium experiments (Kim et al. 2002; Lampi et al. 2006). Human βB1 crystallin

deamidated in the domain interface also had an increased tendency to aggregate at high temperatures and required more α-crystalllin to suppress aggregation (Lampi et al. 2002). Here we show that deamidation of the HγD-Crys domain interface glutamines destabilizes the protein and lowers the kinetic barrier to unfolding. These *in vitro* experiments generated deamidation mimics by site-directed substitution of L-glutamine or –asparagine with L-glutamate or –aspartate. The effects of backbone isomerization and side chain racemization that may also occur as a result of deamidation are expected to have even severer effects on stabilities and kinetic unfolding barriers of the crystallins.

The effects of interface glutamine deamidation in HγD-Crys effectively increased the probablility of populating partially-unfolded conformations under conditions that favor the native state. These partially unfolded conformations may be prone to aggregation through mechanisms such as domain swapping or loop-sheet insertion (Carrell et al. 1994; Liu and Eisenberg 2002). For example, the single folded domain conformer may be susceptible to aggregation by domain swapping where the unfolded N-td of one monomer would use the C-td domain interface of another monomer during templated refolding. The intermediate with a folded C-td and partially unfolded N-td may be prone to aggregation through a reaction akin to loop-sheet insertion where an unstructured loop of one monomer inserts as a β-strand into the β-sheet of another monomer.

Rescue or repair of deamidated crystallins may be possible through the actions of α-crystallin and the enzyme isoaspartyl protein carboxyl methyltransferase. This enzyme, which functions to repair racemized aspartly groups, has been identified in the lens where decreased expression appears to correlate with cataract (McFadden et al. 1983; McFadden and Clarke 1986; Kodama et al. 1995). It is unclear if there is a corresponding enzyme in the lens that may repair damaged glutamines. The onset of cataract may correspond to damage at regions of the crystallins particularly sensitive to chemical or structural alterations, such as the domain interfaces of the β- and γ-crystallins. Alternatively, onset may not depend on damage of these *hot-spots* but instead may occur when an overall threshold of damage has been exceeded at which time the α-crystallin chaperone complexes may be saturated or in an inactive aggregated state themselves.

CHAPTER FIVE:

CONCLUDING DISCUSSIONS

A. DISCREPANCIES BETWEEN IN VITRO AND IN VIVO CONDITIONS

A major goal of these investigations was to identify partially structured intermediates on the *in vitro* unfolding/refolding pathways of the eye lens protein human γD crystallin (HγD-Crys) that may be common to the *in vitro* aggregation pathway (Kosinski-Collins and King 2003). HγD-Crys is primarily localized in the densely packed lens nucleus. Other crystallins present in the lens nucleus include human γC crystallin and members of the α- and β-crystallins. The crystallins are present in the nuclear lens cells in concentrations of 200-400 mg/ml, where they account for 90% of the total protein (Oyster 1999). Heterogeneity in the surface properties and oligomeric states of the crystallins are vital for preventing protein crystallization in the lens cells.

Crystallins of the lens nucleus are expressed *in utero* and must remain stable and soluble for the human lifetime to maintain lens transparency and refraction. A gradient of protein concentrations exists in the lens that establishes the gradient of refraction necessary for fine-tune light focusing (Delaye and Tardieu 1983). Covalent damage of the crystallins that increases with age may be significant in causing or accelerating the formation of insoluble crystallin inclusions that cause cataract.

Several *in vitro* studies of HγD-Crys have elucidated properties of the protein that may be physiologically important in maintaining stability and solubility (Kosinski-Collins and King 2003; Evans et al. 2004; Kosinski-Collins et al. 2004; Flaugh et al. 2005a; 2005b). HγD-Crys is a very stable protein that aggregates when refolded out of denaturant (Kosinski-Collins and King 2003). The domains of HγD-Crys display differential stability where the N-td is less stable than the C-td and is stabilized in the full-length protein by inter-domain interactions (Flaugh et al. 2005a; 2005b). The C-td also folds first during refolding and the structured domain interface of the C-td acts as a nucleating center for refolding of the N-td (Kosinski-Collins et al. 2004; Flaugh et al. 2005a). Deamidation of domain interface glutamines decreases stability of the N-td and reduces the kinetic barrier to unfolding (Chapter 3). Therefore, wild-type HγD-Crys displays both thermodynamic and kinetic stability that are dependent on favorable interactions between the domains and are reduced as a result of covalent damage to domain interface glutamines (Chapter 3).

While these experiments have elucidated important *in vitro* properties of HγD-Crys, they were performed in dilute solutions that did not reflect *in vivo* conditions of the lens. Therefore the question remains, to what extent are the *in vitro* refolding pathway and partially-folded intermediates relevant in the lens? Features of the lens fiber cells expected to have significant effects on properties of the crystallins are 1) general excluded volume effects due to molecular crowding, and 2) continued exposure to environmental stresses and lack of protein turnover.

1. *Excluded volume and molecular crowding*

Excluded volume effects of the crowded cellular environment favor native protein conformations. These effects have been mimicked *in vitro* through use of crowding agents such as the polysaccharides dextran, ficoll and polyethylene glycol. Experimentally, macromolecular crowding has been shown to affect enzyme activity (Minton and Wilf 1981), association constants (Reddy et al. 1995), protein stability (Sasahara et al. 2003), folding kinetics (van den Berg et al. 2000), and aggregation kinetics (van den Berg et al. 1999). For example, the stability of hen egg white lysozyme (HEWL) increased in the presence of ~100 g/L dextran (Sasahara et al. 2003). Similarly, the presence of ficoll accelerated the fast folding pathway and retarded the slow folding pathway of HEWL (van den Berg et al. 2000). Finally, crowding increased aggregation of HEWL during refolding under reducing conditions (van den Berg et al. 1999).

In vitro experiments performed with crowding agents reflect *in vivo* conditions to some extent. However, they still rely on the assumptions that crowding agents do not specifically bind to the proteins of interest, and that the crowding agents truly mimic conditions in the cytoplasm. A method was recently developed to assess "true" *in vivo* stability by monitoring changes in fluorescence of an exogenous fluorophore bound to a tetra-Cys motif in a protein of interest (Ignatova and Gierasch 2004). These experiments

took advantage of the fact that *E. coli* cells are Urea-permeable and are viable in up to ~3 M Urea. Stabilities of cellular retinoic acid binding protein I (CRABPI) measured by fluorophore binding in the presence of increasing Urea were similar both *in vivo* in *E. coli* and *in vitro* in dilute solution (Ignatova and Gierasch 2004). Nevertheless, shapes of the transitions were different suggesting different equilibrium unfolding mechanisms. Aggregation kinetics of a CRABPI mutant displayed similar nucleation-growth trends both *in vivo* and *in vitro* (Ignatova and Gierasch 2004). While these experiments represent a significant advance in understanding *in vivo* stability and aggregation, they are unavoidably limited in that all cellular proteins and structure were affected simultaneously with the test protein. Structural changes of the overall protein population may affect results for the test protein by causing aggregation, altering crowding potential of the cytosol, or inducing other specific attractive or repulsive forces.

Given the very high protein concentrations in the lens, the *in vivo* stability of HγD-Crys is probably higher than that measured *in vitro*. Specific interactions mediated by the surface properties of crystallins may also be important in the excluded volume environment of the lens (Bloemendal et al. 2004). There are numerous potential salt bridges in native HγD-Crys, all of which are located on the surface of the protein (Fig. 5-1). In addition to several pairwise interactions, a network of potential salt-bridges exists in each domain that may contribute to stability and short-range order (Salim and Zaidi 2003). Another unique feature of the γ-crystallins is the high numbers of aromatic amino acids. HγD-Crys has four buried tryptophans, fourteen tyrosines and six phenylalanines (Fig. 5-2). Several of these are surface-exposed where they may interact with aromatics from other monomers through edge-face or face-face ring associations.

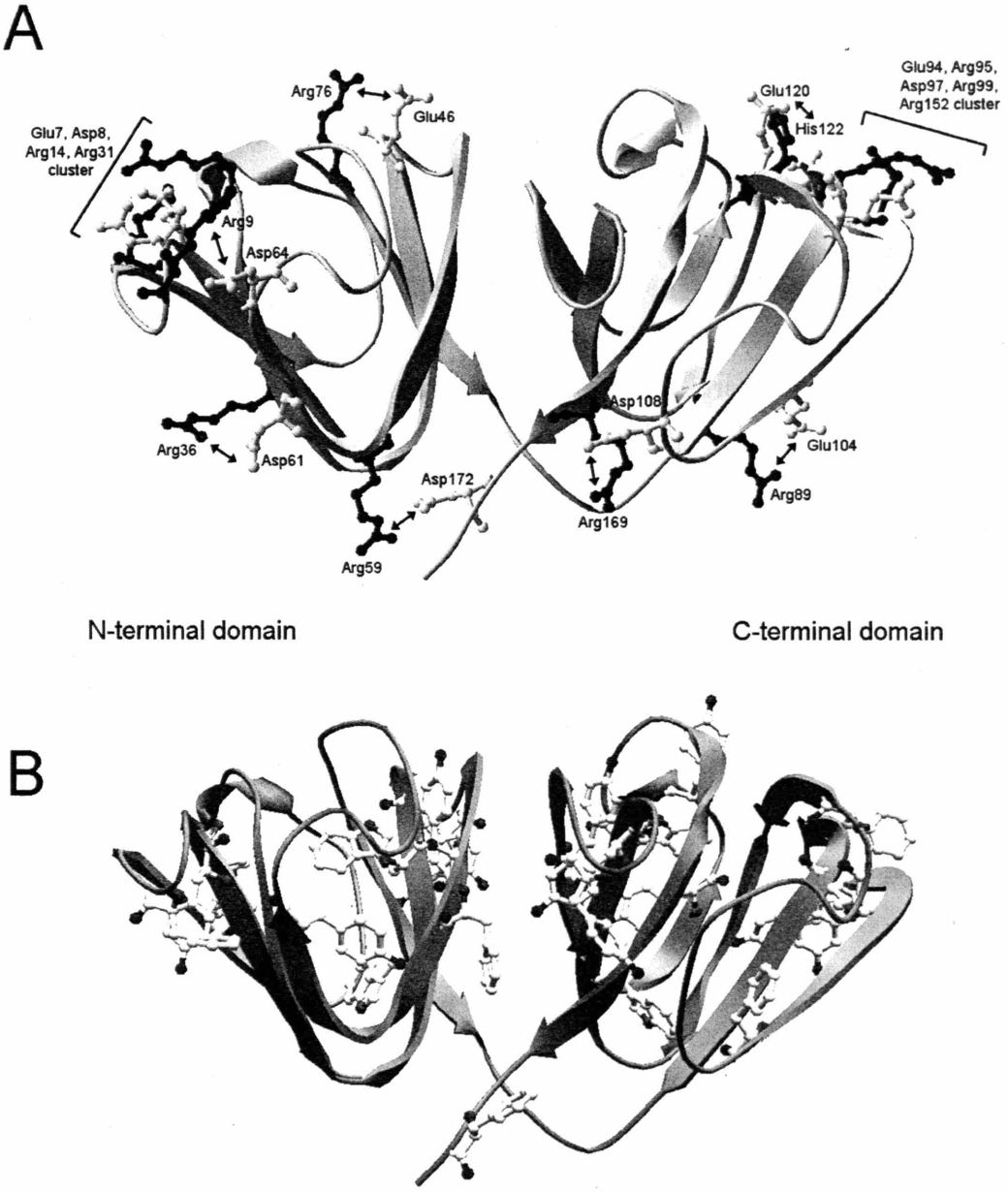

Figure 5-1. (A) Ribbon structure of HγD-Crys with potential salt-bridges shown in stick representation. Pariwise salt-bridges and the two ion-pair networks are labeled. (B) Ribbon structure of HγD-Crys with aromatic amino acids shown in stick representation.

2. *Covalent damage and protein aggregation*

Crystallins in the insoluble protein inclusions associated with mature-onset cataract have high levels of covalent damage. Intermolecular disulfide bonded forms of the crystallins accumulate with age. Truscott (2005) proposed that this is due to loss of the protective reducing agent glutathione. Deamidation of the crystallins also increases with age (Hanson et al. 1998). Oxidation of tryptophan, tyrosine, histidine and methionine may derive from photooxidation induced by UV irradiation (Hott and Borkman 1992). Backbone cleavage and truncation has been proposed to occur through the enzymatic activity of largely unidentified proteases (David et al. 1993; Chaerkady and Sharma 2004). The full complement of enzymes in the ubiquitin-proteosome pathway is present in the lens. However, it is unclear if they function in aged lens fiber cells or if they are simply leftover from differentiation (Pereira et al. 2003). This would be consistent with the notion that complete protein degradation does not occur in the lens.

Covalent damage of the crystallins may instigate aggregation by inducing partial unfolding into aggregation-prone conformations. However, it is important to note that covalent damage begins early in life, long before cataract formation, suggesting that a threshold of damage or damage at crucial sites may be necessary to elicit ill-effects.

The results of Chapter 3 and those of Lampi and colleagues (Kim et al. 2002; Lampi et al. 2006) suggest that deamidation of the β- and γ-crystallins can cause destabilization into partially-unfolded conformations. Two partially-unfolded conformations of deamidated HγD-Crys were identified during kinetic unfolding (Chapter 3). The first had a structured C-td and unstructured N-td. This conformation may be particularly susceptible to aggregation by domain swapping where the unfolded N-td would refold using the C-td of another monomer as a template or nucleus (Fig. 5-2). The second conformation had a folded C-td and partially-folded N-td that likely maintained structure in the region near the domain interface. This species may be susceptible to aggregation through a reaction akin to loop-sheet insertion where an unstructured loop would insert as a β-strand into a β-sheet of another monomer (Fig. 5-2).

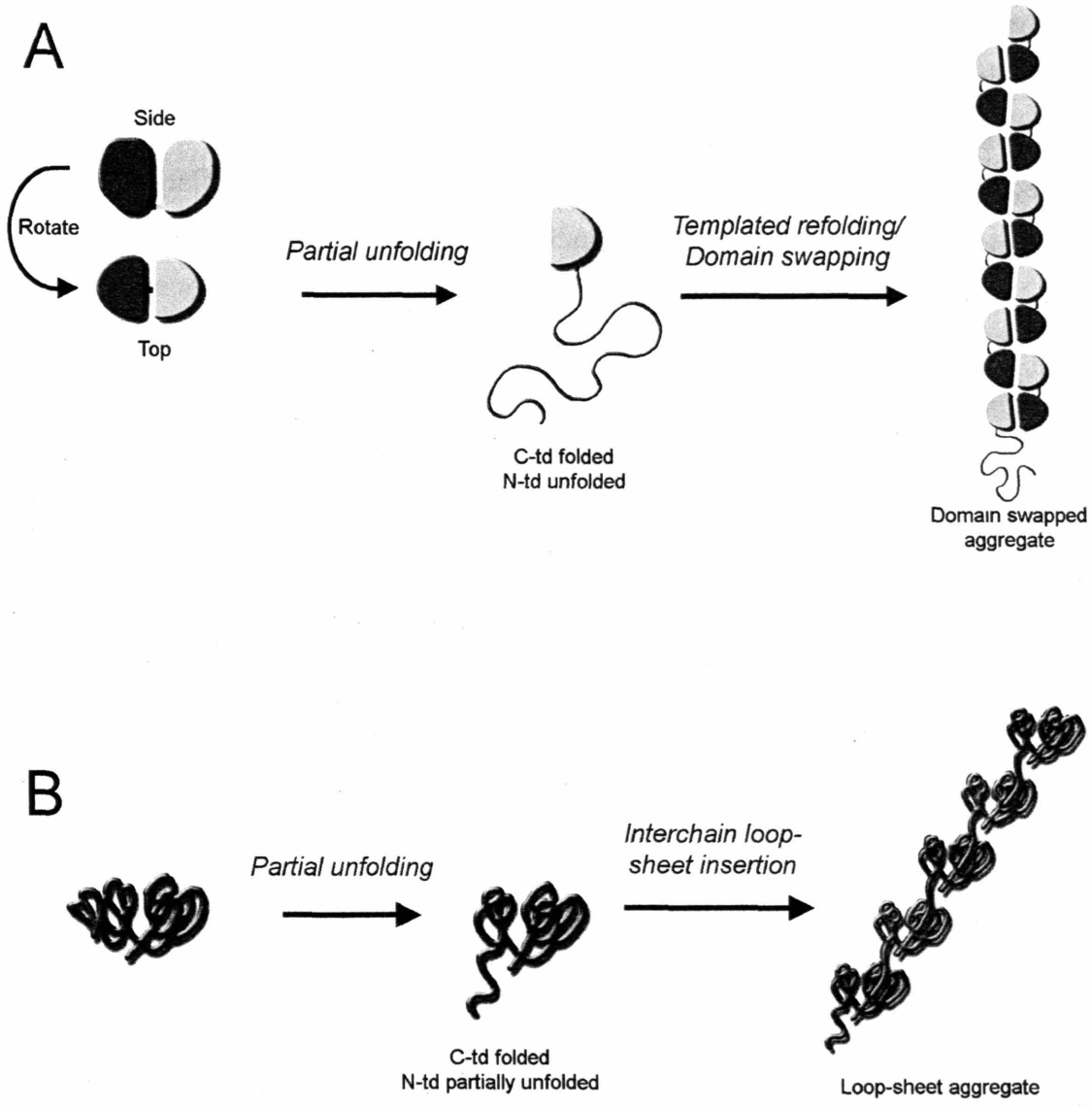

Figure 5-2. (A) Schematic model of HγD-Crys aggregation by domain swapping. The N-td refolds using the structured C-td of another monomer as a template. (B) Schematic model of HγD-Crys aggregation by loop-sheet insertion where a loop from one monomer inserts into the sheet of another.

The two models of aggregation described above could occur strictly between molecules of HγD-Crys, or between HγD-Crys and other β- and γ-crystallins. Structural similarities of the β- and γ-crystallins would be expected to promote incorrect intermolecular interactions of such partially-unfolded conformations. Mixed *in vitro* refolding experiments were performed with HγD-Crys and HγS-Crys, which does not aggregate during *in vitro* refolding (Kosinski-Collins 2004). When refolded together the two proteins formed a mixed aggregate indicating that HγD-Crys was able to recruit HγS-Crys into the aggregate (Kosinski-Collins 2004).

To fully understand interactions of covalently-damaged crystallins in the lens, mixed *in vitro* experiments should also be performed with damaged crystallins. For HγD-Crys this would include a commonly observed 9 kDa truncation product (residues 87 to 173) that is found in cataractous inclusions and is capable of cross-linking with α-, β- and γ-crystallins *in vitro* (Srivastava et al. 1992; Srivastava and Srivastava 2003).

3. *Interaction with α-crystallin*

Another major potential interaction ignored during *in vitro* experiments of single crystallin species is binding or association with α-crystallin. The mechanism by which α-crystallin recognizes and binds substrates in the lens in not well understood. *In vitro*, α-crystallin has been shown to bind partially-folded or –unfolded conformations of substrate proteins, possibly recognizing exposed hydrophobics (Horwitz 1992; Sathish et al. 2004). It is probable that α-crystallin recognizes partially-folded aggregation-prone substrates in the lens through similar means. Alternatively, α-crystallin may recognize particular chemical changes that occur as a result of covalent damage and may precede partial-unfolding.

Previous investigations demonstrated that α-crystallin is capable of preventing off-pathway competing aggregation during *in vitro* refolding of bovine β- and γ-crystallins (Horwitz 2003). The same may be true for HγD-Crys. Experiments should also be performed to investigate potential interactions of α-crystallin and HγD-Crys deamidation mutants. Similar experiments were performed with a domain interface mutant of βB2-Crys that caused domain destabilization (Sathish et al. 2004). The α-

crystallins preferentially bound the destabilized mutant suggesting that the single-folded domain conformer accessible to many β- and γ-crystallins is a specific substrate of α-crystallin.

B. CONCLUDING REMARKS

The morphologies of cataractous crystallin aggregates have not been well investigated. The ophthalmology community generally describes cataractous protein inclusions as amorphous "clumps" devoid of defined morphology. However, *in vitro* experiments have demonstrated that the crystallins can polymerize into well-defined ordered structures, including amyloid fibers (Kosinski-Collins and King 2003; Meehan et al. 2004). It is possible that such well-defined structures also exist in insoluble inclusions of cataractous lenses. More detailed descriptions of conformations of the crystallins present in these inclusions will give insight into the aggregation precursors. Ultimately, the goal of such investigations would be to identify the specific structures that cause cataract formation so that strategies for cataract prevention or treatment may be developed.

CHAPTER SIX: REFERENCES

Aloy, P., Pichaud, M., and Russell, R.B. 2005. Protein complexes: structure prediction challenges for the 21st century. *Curr Opin Struct Biol* **15:** 15-22.

Andley, U.P., Mathur, S., Griest, T.A., and Petrash, J.M. 1996. Cloning, expression, and chaperone-like activity of human alphaA-crystallin. *J Biol Chem* **271:** 31973-31980.

Anfinsen, C.B. 1973. Principles that govern the folding of protein chains. *Science* **181:** 223-230.

Anfinsen, C.B., and Scheraga, H.A. 1975. Experimental and theoretical aspects of protein folding. *Adv Protein Chem* **29:** 205-300.

Annunziata, O., Pande, A., Pande, J., Ogun, O., Lubsen, N.H., and Benedek, G.B. 2005. Oligomerization and phase transitions in aqueous solutions of native and truncated human beta B1-crystallin. *Biochemistry* **44:** 1316-1328.

Apetri, A.C., Surewicz, K., and Surewicz, W.K. 2004. The effect of disease-associated mutations on the folding pathway of human prion protein. *J Biol Chem* **279:** 18008-18014.

Apetri, A.C., and Surewicz, W.K. 2002. Kinetic intermediate in the folding of human prion protein. *J Biol Chem* **277:** 44589-44592.

Aquilina, J.A., Benesch, J.L., Ding, L.L., Yaron, O., Horwitz, J., and Robinson, C.V. 2005. Subunit exchange of polydisperse proteins: Mass spectrometry reveals consequences of alpha A-crystallin truncation. *J Biol Chem.*

Bagby, S., Go, S., Inouye, S., Ikura, M., and Chakrabartty, 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. 1994. 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.

Baker, D., Sohl, J.L., and Agard, D.A. 1992. A protein-folding reaction under kinetic control. *Nature* **356:** 263-265.

Bann, J.G., and Frieden, C. 2004. Folding and domain-domain interactions of the chaperone PapD measured by 19F NMR. *Biochemistry* **43:** 13775-13786.

Bann, J.G., Pinkner, J., Hultgren, S.J., and Frieden, C. 2002. Real-time and equilibrium (19)F-NMR studies reveal the role of domain-domain interactions in the folding of the chaperone PapD. *Proc Natl Acad Sci U S A* **99:** 709-714.

Barnhart, M.M., Pinkner, J.S., Soto, G.E., Sauer, F.G., Langermann, S., Waksman, G., Frieden, C., and Hultgren, S.J. 2000. PapD-like chaperones provide the missing information for folding of pilin proteins. *Proc Natl Acad Sci U S A* **97:** 7709-7714.

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 A) and the R58H mutant (1.15 A) associated with aculeiform cataract. *J Mol Biol* **328:** 1137-1147.

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. 1990. X-ray analysis of beta B2-crystallin and evolution of oligomeric lens proteins. *Nature* **347:** 776-780.

Beechem, J.M., Sherman, M.A., and Mas, M.T. 1995. Sequential domain unfolding in phosphoglycerate kinase: fluorescence intensity and anisotropy stopped-flow kinetics of several tryptophan mutants. *Biochemistry* **34:** 13943-13948.

Beeser, S.A., Goldenberg, D.P., and Oas, T.G. 1997. Enhanced protein flexibility caused by a destabilizing amino acid replacement in BPTI. *J Mol Biol* **269:** 154-164.

Bihoreau, M.T., Baudin, V., Marden, M., Lacaze, N., Bohn, B., Kister, J., Schaad, O., Dumoulin, A., Edelstein, S.J., Poyart, C., et al. 1992. Steric and hydrophobic determinants of the solubilities of recombinant sickle cell hemoglobins. *Protein Sci* **1:** 145-150.

Bilsel, O., Zitzewitz, J.A., Bowers, K.E., and Matthews, C.R. 1999. Folding mechanism of the alpha-subunit of tryptophan synthase, an alpha/beta barrel protein: global analysis highlights the interconversion of multiple native, intermediate, and unfolded forms through parallel channels. *Biochemistry* **38:** 1018-1029.

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.

Bloemendal, H., and de Jong, W.W. 1991. Lens proteins and their genes. *Prog Nucleic Acid Res Mol Biol* **41:** 259-281.

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.

Bonar, L., Cohen, A.S., and Skinner, M.M. 1969. Characterization of the amyloid fibril as a cross-beta protein. *Proc Soc Exp Biol Med* **131:** 1373-1375.

Booth, D.R., Sunde, M., Bellotti, V., Robinson, C.V., Hutchinson, W.L., Fraser, P.E., Hawkins, P.N., Dobson, C.M., Radford, S.E., Blake, C.C., et al. 1997. Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature* **385:** 787-793.

Boyle, D., and Takemoto, L. 1994. Characterization of the alpha-gamma and alpha-beta complex: evidence for an in vivo functional role of alpha-crystallin as a molecular chaperone. *Exp Eye Res* **58:** 9-15.

Boyle, D.L., Takemoto, L., Brady, J.P., and Wawrousek, E.F. 2003. Morphological characterization of the Alpha A- and Alpha B-crystallin double knockout mouse lens. *BMC Ophthalmol* **3:** 3.

Brady, J.P., Garland, D., Duglas-Tabor, Y., Robison, W.G., Jr., Groome, A., and Wawrousek, E.F. 1997. Targeted disruption of the mouse alpha A-crystallin gene induces cataract and cytoplasmic inclusion bodies containing the small heat shock protein alpha B-crystallin. *Proc Natl Acad Sci U S A* **94:** 884-889.

Brady, J.P., Garland, D.L., Green, D.E., Tamm, E.R., Giblin, F.J., and Wawrousek, E.F. 2001. AlphaB-crystallin in lens development and muscle integrity: a gene knockout approach. *Invest Ophthalmol Vis Sci* **42:** 2924-2934.

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.

Brandts, J.F., Halvorson, H.R., and Brennan, M. 1975. Consideration of the Possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. *Biochemistry* **14:** 4953-4963.

Bryngelson, J.D., Onuchic, J.N., Socci, N.D., and Wolynes, P.G. 1995. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21:** 167-195.

Buchet, R., and Pikula, S. 2000. Alzheimer's disease: its origin at the membrane, evidence and questions. *Acta Biochim Pol* **47:** 725-733.

Burns, L.L., Dalessio, P.M., and Ropson, I.J. 1998. Folding mechanism of three structurally similar beta-sheet proteins. *Proteins* **33:** 107-118.

Byeon, I.J., Louis, J.M., and Gronenborn, A.M. 2003. A protein contortionist: core mutations of GB1 that induce dimerization and domain swapping. *J Mol Biol* **333:** 141-152.

Byeon, I.J., Louis, J.M., and Gronenborn, A.M. 2004. A captured folding intermediate involved in dimerization and domain-swapping of GB1. *J Mol Biol* **340:** 615-625.

Capaldi, A.P., and Radford, S.E. 1998. Kinetic studies of beta-sheet protein folding. *Curr Opin Struct Biol* **8:** 86-92.

Carrell, R.W., and Gooptu, B. 1998. Conformational changes and disease--serpins, prions and Alzheimer's. *Curr Opin Struct Biol* **8:** 799-809.

Carrell, R.W., Whisstock, J., and Lomas, D.A. 1994. Conformational changes in serpins and the mechanism of alpha 1-antitrypsin deficiency. *Am J Respir Crit Care Med* **150:** S171-175.

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.

Chaerkady, R., and Sharma, K.K. 2004. Characterization of a bradykinin-hydrolyzing protease from the bovine lens. *Invest Ophthalmol Vis Sci* **45:** 1214-1223.

Chen, J., and Smith, D.L. 2000. Unfolding and disassembly of the chaperonin GroEL occurs via a tetradecameric intermediate with a folded equatorial domain. *Biochemistry* **39:** 4250-4258.

Chiti, F., Taddei, N., Bucciantini, M., White, P., Ramponi, G., and Dobson, C.M. 2000. Mutational analysis of the propensity for amyloid formation by a globular protein. *Embo J* **19:** 1441-1449.

Choudhury, D., Thompson, A., Stojanoff, V., Langermann, S., Pinkner, J., Hultgren, S.J., and Knight, S.D. 1999. X-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic Escherichia coli. *Science* **285:** 1061-1066.

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 Clark, J.M. 2000. Lens cytoplasmic phase separation. *Int Rev Cytol* **192:** 171-187.

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.

Clark, J.I., Neuringer, J.R., and Benedek, G.B. 1983. Phase separation and lens cell age. *J Gerontol* **38:** 287-292.

Clark, P.L., Liu, Z.P., Zhang, J., and Gierasch, L.M. 1996. Intrinsic tryptophans of CRABPI as probes of structure and folding. *Protein Sci* **5:** 1108-1117.

Clark, P.L., Weston, B.F., and Gierasch, L.M. 1998. Probing the folding pathway of a beta-clam protein with single-tryptophan constructs. *Fold Des* **3:** 401-412.

Clarke, D.T., Doig, A.J., Stapley, B.J., and Jones, G.R. 1999. The alpha-helix folds on the millisecond time scale. *Proc Natl Acad Sci U S A* **96:** 7232-7237.

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.

Corbett, R.J., Ahmad, F., and Roche, R.S. 1986. Domain unfolding and the stability of thermolysin in guanidine hydrochloride. *Biochem Cell Biol* **64:** 953-961.

Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16:** 10881-10890.

Das, B.K., and Liang, J.J. 1998. Thermodynamic and kinetic characterization of calf lens gammaF-crystallin. *Int J Biol Macromol* **23:** 191-197.

David, L.L., Shearer, T.R., and Shih, M. 1993. Sequence analysis of lens beta-crystallins suggests involvement of calpain in cataract formation. *J Biol Chem* **268:** 1937-1940.

de Prat-Gay, G., Nadra, A.D., Corrales-Izquierdo, F.J., Alonso, L.G., Ferreiro, D.U., and Mok, Y.K. 2005. The folding mechanism of a dimeric beta-barrel domain. *J Mol Biol* **351:** 672-682.

Delaye, M., and Tardieu, A. 1983. Short-range order of crystallin proteins accounts for eye lens transparency. *Nature* **302:** 415-417.

DiFiglia, M., Sapp, E., Chase, K.O., Davies, S.W., Bates, G.P., Vonsattel, J.P., and Aronin, N. 1997. Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science* **277:** 1990-1993.

Dobson, C.M. 1999. Protein misfolding, evolution and disease. *Trends Biochem Sci* **24:** 329-332.

Dragan, A.I., and Privalov, P.L. 2002. Unfolding of a leucine zipper is not a simple two-state transition. *J Mol Biol* **321:** 891-908.

Eanes, E.D., and Glenner, G.G. 1968. X-ray diffraction studies on amyloid filaments. *J Histochem Cytochem* **16:** 673-677.

Elliott, P.R., Pei, X.Y., Dafforn, T.R., and Lomas, D.A. 2000. Topography of a 2.0 A structure of alpha1-antitrypsin reveals targets for rational drug design to prevent conformational disease. *Protein Sci* **9:** 1274-1281.

Engh, R., Lobermann, H., Schneider, M., Wiegand, G., Huber, R., and Laurell, C.B. 1989. The S variant of human alpha 1-antitrypsin, structure and implications for function and metabolism. *Protein Eng* **2:** 407-415.

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.

Fandrich, M., Fletcher, M.A., and Dobson, C.M. 2001. Amyloid fibrils from muscle myoglobin. *Nature* **410:** 165-166.

Fernald, R.D., and Wright, S.E. 1983. Maintenance of optical quality during crystalline lens growth. *Nature* **301:** 618-620.

Finke, J.M., and Jennings, P.A. 2002. Interleukin-1 beta folding between pH 5 and 7: experimental evidence for three-state folding behavior and robust transition state positions late in folding. *Biochemistry* **41:** 15056-15067.

Fitch, C.A., Karp, D.A., Lee, K.K., Stites, W.E., Lattman, E.E., and Garcia-Moreno, E.B. 2002. Experimental pK(a) values of buried residues: analysis with continuum methods and role of water penetration. *Biophys J* **82:** 3289-3304.

Flaugh, S.L., Kosinski-Collins, M.S., and King, J. 2005a. Contributions of hydrophobic domain interface interactions to the folding and stability of human {gamma}D-crystallin. *Protein Sci* **14:** 569-581.

Flaugh, S.L., Kosinski-Collins, M.S., and King, J. 2005b. Interdomain side-chain interactions in human gammaD crystallin influencing folding and stability. *Protein Sci* **14:** 2030-2043.

Forsyth, W.R., Antosiewicz, J.M., and Robertson, A.D. 2002. Empirical relationships between protein structure and carboxyl pKa values in proteins. *Proteins* **48:** 388-403.

Fraczkiewicz, R., and Braun, W. 1998. Exact and effiecient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem.* **19:** 319-333.

Fu, L., and Liang, J.J. 2002. Conformational change and destabilization of cataract gammaC-crystallin T5P mutant. *FEBS Lett* **513:** 213-216.

Geiger, T., and Clarke, S. 1987. Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. *J Biol Chem* **262:** 785-794.

Gilmanshin, R., Williams, S., Callender, R.H., Woodruff, W.H., and Dyer, R.B. 1997. Fast events in protein folding: relaxation dynamics of secondary and tertiary structure in native apomyoglobin. *Proc Natl Acad Sci U S A* **94:** 3709-3713.

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.

Groenen, P.J., Merck, K.B., de Jong, W.W., and Bloemendal, H. 1994. Structure and modifications of the junior chaperone alpha-crystallin. From lens transparency to molecular pathology. *Eur J Biochem* **225:** 1-19.

Gupta, R., and Srivastava, O.P. 2004a. Deamidation affects structural and functional properties of human alphaA-crystallin and its oligomerization with alphaB-crystallin. *J Biol Chem* **279:** 44258-44269.

Gupta, R., and Srivastava, O.P. 2004b. Effect of deamidation of asparagine 146 on functional and structural properties of human lens alphaB-crystallin. *Invest Ophthalmol Vis Sci* **45:** 206-214.

Haase-Pettingell, C.A., and King, J. 1988. Formation of aggregates from a thermolabile in vivo folding intermediate in P22 tailspike maturation. A model for inclusion body formation. *J Biol Chem* **263:** 4977-4983.

Haber, E., and Anfinsen, C.B. 1961. Regeneration of enzyme activity by air oxidation of reduced subtilisin-modified ribonuclease. *J Biol Chem* **236:** 422-424.

Hagihara, Y., Tan, Y., and Goto, Y. 1994. Comparison of the conformational stability of the molten globule and native states of horse cytochrome c. Effects of acetylation, heat, urea and guanidine-hydrochloride. *J Mol Biol* **237:** 336-348.

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.

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.

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.

Hanson, S.R., Smith, D.L., and Smith, J.B. 1998. Deamidation and disulfide bonding in human lens gamma-crystallins. *Exp Eye Res* **67:** 301-312.

Harding, J.J. 2002. Viewing molecular mechanisms of ageing through a lens. *Ageing Res Rev* **1:** 465-479.

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.

Harper, J.D., and Lansbury, P.T., Jr. 1997. Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. *Annu Rev Biochem* **66:** 385-407.

Harper, J.D., Lieber, C.M., and Lansbury, P.T., Jr. 1997. Atomic force microscopic imaging of seeded fibril formation and fibril branching by the Alzheimer's disease amyloid-beta protein. *Chem Biol* **4:** 951-959.

Hawkins, C.L., and Davies, M.J. 2005. The role of aromatic amino acid oxidation, protein unfolding, and aggregation in the hypobromous acid-induced inactivation of trypsin inhibitor and lysozyme. *Chem Res Toxicol* **18:** 1669-1677.

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.

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.

Hoenders, H.J., and Bloemendal, H. 1983. Lens proteins and aging. *J Gerontol* **38:** 278-286.

Hope, J.N., Chen, H.C., and Hejtmancik, J.F. 1994. Aggregation of beta A3-crystallin is independent of the specific sequence of the domain connecting peptide. *J Biol Chem* **269:** 21141-21145.

Horovitz, A. 1996. Double-mutant cycles: a powerful tool for analyzing protein structure and function. *Fold Des* **1:** R121-126.

Horwich, A. 2002. Protein aggregation in disease: a role for folding intermediates forming specific multimeric interactions. *J Clin Invest* **110:** 1221-1232.

Horwitz, J. 1992. Alpha-crystallin can function as a molecular chaperone. *Proc Natl Acad Sci U S A* **89:** 10449-10453.

Horwitz, J. 2000. The function of alpha-crystallin in vision. *Semin Cell Dev Biol* **11:** 53-60.

Horwitz, J. 2003. Alpha-crystallin. *Exp Eye Res* **76:** 145-153.

Hott, J.L., and Borkman, R.F. 1992. Analysis of photo-oxidized amino acids in tryptic peptides of calf lens gamma-II crystallin. *Photochem Photobiol* **56:** 257-263.

Hsieh, H.C., Kumar, T.K., Sivaraman, T., and Yu, C. 2006. Refolding of a small all beta-sheet protein proceeds with accumulation of kinetic intermediates. *Arch Biochem Biophys*.

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.

Ignatova, Z., and Gierasch, L.M. 2004. Monitoring protein stability and aggregation in vivo by real-time fluorescent labeling. *Proc Natl Acad Sci U S A* **101:** 523-528.

Ingram, V.M. 1956. A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin. *Nature* **178:** 792-794.

Iwaki, T., Kume-Iwaki, A., and Goldman, J.E. 1990. Cellular distribution of alpha B-crystallin in non-lenticular tissues. *J Histochem Cytochem* **38:** 31-39.

Jackson, S.E. 1998. How do small single-domain proteins fold? *Fold Des* **3:** R81-91.

Jaenicke, R. 1999. Stability and folding of domain proteins. *Prog Biophys Mol Biol* **71:** 155-241.

Jaenicke, R., and Bohm, G. 1998. The stability of proteins in extreme environments. *Curr Opin Struct Biol* **8:** 738-748.

Jaenicke, R., and Slingsby, C. 2001. Lens crystallins and their microbial homologs: structure, stability, and function. *Crit Rev Biochem Mol Biol* **36:** 435-499.

Janowski, R., Kozak, M., Jankowska, E., Grzonka, Z., Grubb, A., Abrahamson, M., and Jaskolski, M. 2001. Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping. *Nat Struct Biol* **8:** 316-320.

Jarrett, J.T., and Lansbury, P.T., Jr. 1993. Seeding "one-dimensional crystallization" of amyloid: a pathogenic mechanism in Alzheimer's disease and scrapie? *Cell* **73:** 1055-1058.

Jaskolski, M. 2001. 3D domain swapping, protein oligomerization, and amyloid formation. *Acta Biochim Pol* **48:** 807-827.

Jenkins, J., and Pickersgill, R. 2001. The architecture of parallel beta-helices and related folds. *Prog Biophys Mol Biol* **77:** 111-175.

Jiang, X., Smith, C.S., Petrassi, H.M., Hammarstrom, P., White, J.T., Sacchettini, J.C., and Kelly, J.W. 2001. An engineered transthyretin monomer that is nonamyloidogenic, unless it is partially denatured. *Biochemistry* **40:** 11442-11452.

Jones, S., Marin, A., and Thornton, J.M. 2000. Protein domain interfaces: characterization and comparison with oligomeric protein interfaces. *Protein Eng* **13:** 77-82.

Katou, H., Hoshino, M., Kamikubo, H., Batt, C.A., and Goto, Y. 2001. Native-like beta-hairpin retained in the cold-denatured state of bovine beta-lactoglobulin. *J Mol Biol* **310:** 471-484.

Kim, K.K., Kim, R., and Kim, S.H. 1998. Crystal structure of a small heat-shock protein. *Nature* **394:** 595-599.

Kim, W., and Hecht, M.H. 2005. Sequence determinants of enhanced amyloidogenicity of Alzheimer A{beta}42 peptide relative to A{beta}40. *J Biol Chem* **280:** 35069-35076.

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.

Kmoch, S., Brynda, J., Asfaw, B., Bezouska, K., Novak, P., Rezacova, P., Ondrova, L., Filipec, M., Sedlacek, J., and Elleder, M. 2000. Link between a novel human gammaD-crystallin allele and a unique cataract phenotype explained by protein crystallography. *Hum Mol Genet* **9:** 1779-1786.

Kodama, T., Mizobuchi, M., Takeda, R., Torikai, H., Shinomiya, H., and Ohashi, Y. 1995. Hampered expression of isoaspartyl protein carboxyl methyltransferase gene in the human cataractous lens. *Biochim Biophys Acta* **1245:** 269-272.

Kohn, W.D., Monera, O.D., Kay, C.M., and Hodges, R.S. 1995. The effects of interhelical electrostatic repulsions between glutamic acid residues in controlling the dimerization and stability of two-stranded alpha-helical coiled-coils. *J Biol Chem* **270:** 25495-25506.

Kosinski-Collins, M.S. 2004. Characterization of the unfolding, refolding, and aggregation pathways of two proteins implicated in cataractogenesis: human gamma D and human gamma S crystallin. In *Biology*, pp. 217. Massachusetts Institute of Technology, Cambridge, MA.

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.

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.

Kretschmar, M., Mayr, E.M., and Jaenicke, R. 1999. Kinetic and thermodynamic stabilization of the betagamma-crystallin homolog spherulin 3a from Physarum polycephalum by calcium binding. *J Mol Biol* **289:** 701-705.

Kurochkin, I.V., Procyk, R., Bishop, P.D., Yee, V.C., Teller, D.C., Ingham, K.C., and Medved, L.V. 1995. Domain structure, stability and domain-domain interactions in recombinant factor XIII. *J Mol Biol* **248:** 414-430.

Lampi, K.J., Amyx, K.K., Ahmann, P., and Steel, E.A. 2006. Deamidation in Human Lens betaB2-crystallin Destabilizes the Dimer. *Biochemistry*.

Lampi, K.J., Kim, Y.H., Bachinger, H.P., Boswell, B.A., Lindner, R.A., Carver, J.A., Shearer, T.R., David, L.L., and Kapfer, D.M. 2002. Decreased heat stability and increased chaperone requirement of modified human betaB1-crystallins. *Mol Vis* **8:** 359-366.

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.

Lampi, K.J., Ma, Z., Shih, M., Shearer, T.R., Smith, J.B., Smith, D.L., and David, L.L. 1997. Sequence analysis of betaA3, betaB3, and betaA4 crystallins completes the

identification of the major proteins in young human lens. *J Biol Chem* **272:** 2268-2275.

Lampi, K.J., Oxford, J.T., Bachinger, H.P., Shearer, T.R., David, L.L., and Kapfer, D.M. 2001. Deamidation of human betaB1 alters the elongated structure of the dimer. *Experimental Eye Research* **72:** 279-288.

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., Purkiss, A.G., Smith, D.L., and Smith, J.B. 2002. Deamidation in human gamma S-crystallin from cataractous lenses is influenced by surface exposure. *Biochemistry* **41:** 8638-8648.

Laurents, D.V., and Baldwin, R.L. 1998. Protein folding: matching theory and experiment. *Biophys J* **75:** 428-434.

Levinthal, C. 1968. Are there pathways for protein folding? *J Chem Phys* **65:** 44-45.

Levinthal, C. 1969. *How to fold graciously*. University of Illinois Press, Urbana, pp. 22-24.

Li, L., Huang, H.H., Badilla, C.L., and Fernandez, J.M. 2005. Mechanical unfolding intermediates observed by single-molecule force spectroscopy in a fibronectin type III module. *J Mol Biol* **345:** 817-826.

Liang, J.J. 2004. Interactions and chaperone function of {alpha}A-crystallin with T5P {gamma}C-crystallin mutant. *Protein Sci* **13:** 2476-2482.

Liu, Y., and Eisenberg, D. 2002. 3D domain swapping: as domains continue to swap. *Protein Sci* **11:** 1285-1299.

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.

Luhrs, T., Ritter, C., Adrian, M., Riek-Loher, D., Bohrmann, B., Dobeli, H., Schubert, D., and Riek, R. 2005. 3D structure of Alzheimer's amyloid-beta(1-42) fibrils. *Proc Natl Acad Sci U S A* **102:** 17342-17347.

Lund, A.L., Smith, J.B., and Smith, D.L. 1996. Modifications of the water-insoluble human lens alpha-crystallins. *Exp Eye Res* **63:** 661-672.

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.

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.

Mayr, E.M., Jaenicke, R., and Glockshuber, R. 1994. Domain interactions and connecting peptides in lens crystallins. *J Mol Biol* **235:** 84-88.

Mayr, E.M., Jaenicke, R., and Glockshuber, R. 1997. The domains in gammaB-crystallin: identical fold-different stabilities. *J Mol Biol* **269:** 260-269.

McFadden, P.N., and Clarke, S. 1986. Protein carboxyl methyltransferase and methyl acceptor proteins in aging and cataractous tissue of the human eye lens. *Mech Ageing Dev* **34:** 91-105.

McFadden, P.N., Horwitz, J., and Clarke, S. 1983. Protein carboxyl methyltransferase from cow eye lens. *Biochem Biophys Res Commun* **113:** 418-424.

Meehan, S., Berry, Y., Luisi, B., Dobson, C.M., Carver, J.A., and MacPhee, C.E. 2004. Amyloid fibril formation by lens crystallin proteins and its implications for cataract formation. *J Biol Chem* **279:** 3413-3419.

Minton, A.P., and Wilf, J. 1981. Effect of macromolecular crowding upon the structure and function of an enzyme: glyceraldehyde-3-phosphate dehydrogenase. *Biochemistry* **20:** 4821-4826.

Mitraki, A., and King, J. 1989. Protein folding intermediates and inclusion body formation. *Biotechnology* **7:** 690-697.

Mittermaier, A., Korzhnev, D.M., and Kay, L.E. 2005. Side-chain interactions in the folding pathway of a Fyn SH3 domain mutant studied by relaxation dispersion NMR spectroscopy. *Biochemistry* **44:** 15430-15436.

Mizuguchi, M., Kroon, G.J., Wright, P.E., and Dyson, H.J. 2003. Folding of a beta-sheet protein monitored by real-time NMR spectroscopy. *J Mol Biol* **328:** 1161-1171.

Monera, O.D., Kay, C.M., and Hodges, R.S. 1994a. Electrostatic interactions control the parallel and antiparallel orientation of alpha-helical chains in two-stranded alpha-helical coiled-coils. *Biochemistry* **33:** 3862-3871.

Monera, O.D., Kay, C.M., and Hodges, R.S. 1994b. Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions. *Protein Sci* **3:** 1984-1991.

National. Eye Insitute (U.S.) 2002. "Vision Problems in the US". Prevent Blindness in America.

Nelson, R., Sawaya, M.R., Balbirnie, M., Madsen, A.O., Riekel, C., Grothe, R., and Eisenberg, D. 2005. Structure of the cross-beta spine of amyloid-like fibrils. *Nature* **435:** 773-778.

Nguyen, D.M., Leila Reynald, R., Gittis, A.G., and Lattman, E.E. 2004. X-ray and thermodynamic studies of staphylococcal nuclease variants I92E and I92K: insights into polarity of the protein interior. *J Mol Biol* **341:** 565-574.

Nguyen, J., Baldwin, M.A., Cohen, F.E., and Prusiner, S.B. 1995. Prion protein peptides induce alpha-helix to beta-sheet conformational transitions. *Biochemistry* **34:** 4186-4192.

Nicholson, E.M., Mo, H., Prusiner, S.B., Cohen, F.E., and Marqusee, S. 2002. Differences between the prion protein and its homolog Doppel: a partially structured state with implications for scrapie formation. *J Mol Biol* **316:** 807-815.

Nilsson, M.R., and Dobson, C.M. 2003. Chemical modification of insulin in amyloid fibrils. *Protein Sci* **12:** 2637-2641.

Nilsson, M.R., Driscoll, M., and Raleigh, D.P. 2002. Low levels of asparagine deamidation can have a dramatic effect on aggregation of amyloidogenic peptides: implications for the study of amyloid formation. *Protein Sci* **11:** 342-349.

Nishikori, S., Shiraki, K., Fujiwara, S., Imanaka, T., and Takagi, M. 2005. Unfolding mechanism of a hyperthermophilic protein O(6)-methylguanine-DNA methyltransferase. *Biophys Chem* **116:** 97-104.

Norledge, B.V., Mayr, E.M., Glockshuber, R., Bateman, O.A., Slingsby, C., Jaenicke, R., and Driessen, H.P. 1996. The X-ray structures of two mutant crystallin domains shed light on the evolution of multi-domain proteins. *Nat Struct Biol* **3:** 267-274.

Novokhatny, V., Schwarz, F., Atha, D., and Ingham, K. 1992. Domain structure and domain-domain interactions in the carboxy-terminal heparin binding region of fibronectin. *J Mol Biol* **227:** 1182-1191.

Oyster, C.W. 1999. *The human eye: structure and function*. Sinauer Associates, Inc, Sunderland, MA.

Palme, S., Jaenicke, R., and Slingsby, C. 1998. 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.

Pereira, P., Shang, F., Hobbs, M., Girao, H., and Taylor, A. 2003. Lens fibers have a fully functional ubiquitin-proteasome pathway. *Exp Eye Res* **76:** 623-631.

Prakash, S., and Matouschek, A. 2004. Protein unfolding in the cell. *Trends Biochem Sci* **29:** 593-600.

Prusiner, S.B. 1998. Prions. *Proc Natl Acad Sci U S A* **95:** 13363-13383.

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.

Rajini, B., Shridas, P., Sundari, C.S., Muralidhar, D., Chandani, S., Thomas, F., and Sharma, Y. 2001. Calcium binding properties of gamma-crystallin: calcium ion binds at the Greek key beta gamma-crystallin fold. *J Biol Chem* **276:** 38464-38471.

Reddy, M.K., Weitzel, S.E., Daube, S.S., Jarvis, T.C., and von Hippel, P.H. 1995. Using macromolecular crowding agents to identify weak interactions within DNA replication complexes. *Methods Enzymol* **262:** 466-476.

Ringens, P.J., Hoenders, H.J., and Bloemendal, H. 1982. Effect of aging on the water-soluble and water-insoluble protein pattern in normal human lens. *Exp Eye Res* **34:** 201-207.

Rose, G.D., Geselowitz, A.R., Lesser, G.J., Lee, R.H., and Zehfus, M.H. 1985. Hydrophobicity of amino acid residues in globular proteins. *Science* **229:** 834-838.

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.

Rothlisberger, D., Honegger, A., and Pluckthun, A. 2005. Domain interactions in the Fab fragment: a comparative evaluation of the single-chain Fv and Fab format engineered with variable domains of different stability. *J Mol Biol* **347:** 773-789.

Rotondi, K.S., and Gierasch, L.M. 2003. Role of local sequence in the folding of cellular retinoic abinding protein I: structural propensities of reverse turns. *Biochemistry* **42:** 7976-7985.

Rotondi, K.S., Rotondi, L.F., and Gierasch, L.M. 2003. Native structural propensity in cellular retinoic acid-binding protein I 64-88: the role of locally encoded structure in the folding of a beta-barrel protein. *Biophys Chem* **100:** 421-436.

Rousseau, F., Schymkowitz, J.W., Wilkinson, H.R., and Itzhaki, L.S. 2001. Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues. *Proc Natl Acad Sci U S A* **98:** 5596-5601.

Roy, M., and Jennings, P.A. 2003. Real-time NMR kinetic studies provide global and residue-specific information on the non-cooperative unfolding of the beta-trefoil protein, interleukin-1beta. *J Mol Biol* **328:** 693-703.

Rudolph, R., Siebendritt, R., Nesslauer, 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.

Sambashivan, S., Liu, Y., Sawaya, M.R., Gingery, M., and Eisenberg, D. 2005. Amyloid-like fibrils of ribonuclease A with three-dimensional domain-swapped and native-like structure. *Nature* **437:** 266-269.

Sandilands, A., Hutcheson, A.M., Long, H.A., Prescott, A.R., Vrensen, G., Loster, J., Klopp, N., Lutz, R.B., Graw, J., Masaki, S., et al. 2002. Altered aggregation properties of mutant gamma-crystallins cause inherited cataract. *Embo J* **21:** 6005-6014.

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.

Santra, M.K., Banerjee, A., Rahaman, O., and Panda, D. 2005. Unfolding pathways of human serum albumin: evidence for sequential unfolding and folding of its three domains. *Int J Biol Macromol* **37:** 200-204.

Sasahara, K., McPhie, P., and Minton, A.P. 2003. Effect of dextran on protein stability and conformation attributed to macromolecular crowding. *J Mol Biol* **326:** 1227-1237.

Sathish, H.A., Koteiche, H.A., and McHaourab, H.S. 2004. Binding of destabilized betaB2-crystallin mutants to alpha-crystallin: the role of a folding intermediate. *J Biol Chem* **279:** 16425-16432.

Sato, S., Ward, C.L., Krouse, M.E., Wine, J.J., and Kopito, R.R. 1996. Glycerol reverses the misfolding phenotype of the most common cystic fibrosis mutation. *J Biol Chem* **271:** 635-638.

Sauer, F.G., Futterer, K., Pinkner, J.S., Dodson, K.W., Hultgren, S.J., and Waksman, G. 1999. Structural basis of chaperone function and pilus biogenesis. *Science* **285:** 1058-1061.

Schymkowitz, J.W., Rousseau, F., Irvine, L.R., and Itzhaki, L.S. 2000. The folding pathway of the cell-cycle regulatory protein p13suc1: clues for the mechanism of domain swapping. *Structure* **8:** 89-100.

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.

Selkoe, D.J. 1991. The molecular pathology of Alzheimer's disease. *Neuron* **6:** 487-498.

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.

Sergeev, Y.V., Wingfield, P.T., and Hejtmancik, J.F. 2000. Monomer-dimer equilibrium of normal and modified beta A3-crystallins: experimental determination and molecular modeling. *Biochemistry* **39:** 15799-15806.

Sharma, A.K., Minke-Gogl, V., Gohl, P., Siebendritt, R., Jaenicke, R., and Rudolph, R. 1990. Limited proteolysis of gamma II-crystallin from calf eye lens. Physicochemical studies on the N-terminal domain and the intact two-domain protein. *Eur J Biochem* **194:** 603-609.

Sherman, M.A., Beechem, J.M., and Mas, M.T. 1995. Probing intradomain and interdomain conformational changes during equilibrium unfolding of phosphoglycerate kinase: fluorescence and circular dichroism study of tryptophan mutants. *Biochemistry* **34:** 13934-13942.

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.

Sifers, R.N. 1995. Defective protein folding as a cause of disease. *Nat Struct Biol* **2:** 355-357.

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., et al. 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.

Simkovsky, R., and King, J. 2006. An elongated spine of buried core residues necessary for in vivo folding of the parallel {beta}-helix of P22 tailspike adhesin. *Proc Natl Acad Sci U S A.*

Simmons, D.A., Wilson, D.J., Lajoie, G.A., Doherty-Kirby, A., and Konermann, L. 2004. Subunit disassembly and unfolding kinetics of hemoglobin studied by time-resolved electrospray mass spectrometry. *Biochemistry* **43:** 14792-14801.

Sinha, D., Esumi, N., Jaworski, C., Kozak, C.A., Pierce, E., and Wistow, G. 1998. Cloning and mapping the mouse Crygs gene and non-lens expression of [gamma]S-crystallin. *Mol Vis* **4:** 8.

Sipe, J.D. 1994. Amyloidosis. *Crit Rev Clin Lab Sci* **31:** 325-354.

Slaughter, B.D., Unruh, J.R., Price, E.S., Huynh, J.L., Bieber Urbauer, R.J., and Johnson, C.K. 2005. Sampling unfolding intermediates in calmodulin by single-molecule spectroscopy. *J Am Chem Soc* **127:** 12107-12114.

Slingsby, C., and Bateman, O.A. 1990. Quaternary interactions in eye lens beta-crystallins: basic and acidic subunits of beta-crystallins favor heterologous association. *Biochemistry* **29:** 6592-6599.

Slingsby, C., Norledge, B., Simpson, A., Bateman, O.A., Wright, G., Driessen, H.P.C., Lindley, P.F., Moss, D.S., and Bax, B. 1997. X-ray diffraction and structure of crystallins. *Prog. Retin. Eye Res.* **16:** 3-29.

Speed, M.A., Wang, D.I., and King, J. 1996. Specific aggregation of partially folded polypeptide chains: the molecular basis of inclusion body composition. *Nat Biotechnol* **14:** 1283-1287.

Srivastava, O.P., McEntire, J.E., and Srivastava, K. 1992. Identification of a 9 kDa gamma-crystallin fragment in human lenses. *Exp Eye Res* **54:** 893-901.

Srivastava, O.P., and Srivastava, K. 2003. Crosslinking of human lens 9 kDa gammaD-crystallin fragment in vitro and in vivo. *Mol Vis* **9:** 644-656.

Stamler, R., Kappe, G., Boelens, W., and Slingsby, C. 2005. Wrapping the alpha-crystallin domain fold in a chaperone assembly. *J Mol Biol* **353:** 68-79.

Staniforth, R.A., Giannini, S., Higgins, L.D., Conroy, M.J., Hounslow, A.M., Jerala, R., Craven, C.J., and Waltho, J.P. 2001. Three-dimensional domain swapping in the folded and molten-globule states of cystatins, an amyloid-forming structural superfamily. *Embo J* **20:** 4774-4781.

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., et al. 1999. Progressive juvenile-onset punctate cataracts caused by mutation of the gammaD-crystallin gene. *Proc Natl Acad Sci U S A* **96:** 1008-1012.

Stites, W.E., Gittis, A.G., Lattman, E.E., and Shortle, D. 1991. In a staphylococcal nuclease mutant the side-chain of a lysine replacing valine 66 is fully buried in the hydrophobic core. *J Mol Biol* **221:** 7-14.

Takemoto, L., and Boyle, D. 2000. Increased deamidation of asparagine during human senile cataractogenesis. *Mol Vis* **6:** 164-168.

Takemoto, L., Fujii, N., and Boyle, D. 2001. Mechanism of asparagine deamidation during human senile cataractogenesis. *Exp Eye Res* **72:** 559-563.

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.

Truscott, R.J. 2005. Age-related nuclear cataract-oxidation is the key. *Exp Eye Res* **80:** 709-725.

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.

Uversky, V.N., Li, J., and Fink, A.L. 2001. Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *J Biol Chem* **276:** 10737-10744.

van den Berg, B., Ellis, R.J., and Dobson, C.M. 1999. Effects of macromolecular crowding on protein folding and aggregation. *Embo J* **18:** 6927-6933.

van den Berg, B., Wain, R., Dobson, C.M., and Ellis, R.J. 2000. Macromolecular crowding perturbs protein refolding kinetics: implications for folding inside the cell. *Embo J* **19:** 3870-3875.

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.

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.

Walkenhorst, W.F., Edwards, J.A., Markley, J.L., and Roder, H. 2002. Early formation of a beta hairpin during folding of staphylococcal nuclease H124L as detected by pulsed hydrogen exchange. *Protein Sci* **11:** 82-91.

Wang, K., Gawinowicz, M.A., and Spector, A. 2000. The effect of stress on the pattern of phosphorylation of alphaA and alphaB crystallin in the rat lens. *Exp Eye Res* **71:** 385-393.

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.

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., 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., and Mayr, E.M. 1998. Myxococcus xanthus spore coat protein S, a stress-induced member of the betagamma-crystallin superfamily, gains stability from binding of calcium ions. *Eur J Biochem* **255:** 604-610.

Westermark, P., Sletten, K., Johansson, B., and Cornwell, G.G., 3rd. 1990. Fibril in senile systemic amyloidosis is derived from normal transthyretin. *Proc Natl Acad Sci U S A* **87:** 2843-2845.

Wetlaufer, D.B., and Ristow, S. 1973. Acquisition of three-dimensional structure of proteins. *Annu Rev Biochem* **42:** 135-158.

Wetzel, R. 1994. Mutations and off-pathway aggregation of proteins. *Trends Biotechnol* **12:** 193-198.

Wetzel, R. 2002. Ideas of order for amyloid fibril structure. *Structure* **10:** 1031-1036.

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.

Williams, S., Causgrove, T.P., Gilmanshin, R., Fang, K.S., Callender, R.H., Woodruff, W.H., and Dyer, R.B. 1996. Fast events in protein folding: helix melting and formation in a small peptide. *Biochemistry* **35:** 691-697.

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., 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., 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., Turnell, B., Summers, L., Slingsby, C., Moss, D., Miller, L., Lindley, P., and Blundell, T. 1983. X-ray analysis of the eye lens protein gamma-II crystallin at 1.9 A resolution. *J Mol Biol* **170:** 175-202.

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.

Wistow, G.J., and Piatigorsky, J. 1988. Lens crystallins: the evolution and expression of proteins for a highly specialized tissue. *Annu Rev Biochem* **57:** 479-504.

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.

Yu, M.H., Lee, K.N., and Kim, J. 1995. The Z type variation of human alpha 1-antitrypsin causes a protein folding defect. *Nat Struct Biol* **2:** 363-367.

Zhang, H., Kaneko, K., Nguyen, J.T., Livshits, T.L., Baldwin, M.A., Cohen, F.E., James, T.L., and Prusiner, S.B. 1995. Conformational transitions in peptides containing two putative alpha-helices of the prion protein. *J Mol Biol* **250:** 514-526.

Zhang, Z., Smith, D.L., and Smith, J.B. 2003. Human beta-crystallins modified by backbone cleavage, deamidation and oxidation are prone to associate. *Exp Eye Res* **77:** 259-272.

Zhou, H., and Zhou, Y. 2004. Quantifying the effect of burial of amino acid residues on protein stability. *Proteins* **54:** 315-322.

Zhou, Z., Feng, H., Zhou, H., Zhou, Y., and Bai, Y. 2005. Design and folding of a multidomain protein. *Biochemistry* **44:** 12107-12112.

CHAPTER SEVEN

APPENDICES

APPENDIX A: SEQUENCE ALIGNMENTS

1. *β-crystallin domain interface residues*

Domain interface residues of vertebrate β-crystallins were determined by sequence alignment using MultAlin (Corpet 1988).

Species and crystallin type	Position in human γD crystallin									
	43	54	56	79	81	132	143	145	147	170
Human γD	M	Q	F	R	I	V	Q	L	M	V
Human βB2	V	Q	V	R	I	V	Q	L	E	I
Mouse βB2	V	Q	V	R	I	V	Q	L	E	I
Rat βB2	V	Q	V	R	I	V	Q	L	E	I
Bullfrog βB2	V	Q	V	R	I	V	Q	L	E	I
Rat βB3	L	Q	V	R	L	V	Q	V	E	I
Mouse βB3	L	Q	V	R	L	V	Q	V	E	I
Bovine βB3	L	Q	V	R	L	V	Q	V	E	I
Human βB3	L	Q	V	R	L	V	Q	V	E	I
Chicken βB3	L	Q	V	R	L	V	Q	V	E	V
Mouse βB1	V	M	V	R	I	V	Q	L	E	L
Rat βB1	V	M	V	R	I	V	Q	L	E	L
Bovine βB1	V	M	V	R	I	V	Q	L	E	L
Human βB1	V	M	I	R	I	V	Q	L	E	L
Chicken βB1	V	M	I	R	I	V	Q	L	E	I
Frog βB1	V	M	I	R	F	V	Q	L	E	V
Zebrafish βB1	V	M	I	R	I	V	Q	L	E	V
Rat βA3	I	Q	I	R	I	V	Q	I	E	I
Bovine βA1	V	Q	V	R	I	V	Q	I	E	I
Human βA3	I	Q	I	R	I	V	Q	I	E	I
Chicken βA3	V	Q	I	R	V	V	Q	V	E	I
Bullfrog βA3	I	Q	V	R	I	V	Q	I	E	I
Bullfrog βA1	I	Q	V	R	I	V	Q	I	E	I
Mouse βA1	I	Q	I	R	I	V	Q	I	E	I
Mouse βA4	V	Q	V	R	V	V	Q	I	E	I
Rat βA4	V	Q	V	R	V	V	Q	V	E	I
Human βA4	V	Q	I	R	A	V	Q	V	E	I
Bovine βA4	V	Q	V	R	V	V	Q	V	E	I
Chicken βA4	V	Q	V	R	I	V	Q	L	E	V
Mouse βA2	V	Q	I	R	V	V	Q	V	E	V
Bovine βA2	V	Q	I	R	V	V	Q	V	E	V
Human βA2	V	Q	I	R	V	V	Q	V	E	V
Chicken βA2	V	Q	I	R	V	V	Q	V	E	I

2. *γ-crystallin domain interface residues*

Domain interface residues of vertebrate γ-crystallins were determined by sequence alignment using MultAlin (Corpet 1988).

Species and crystallin type	Position in human γD crystallin									
	43	54	56	79	81	132	143	145	147	170
Human γD	M	Q	F	R	I	V	Q	L	M	V
Bovine γD	M	Q	F	R	I	V	Q	L	R	V
Rat γD	M	Q	F	R	I	V	Q	L	R	V
Mouse γD	M	Q	F	R	I	V	Q	L	R	V
Human γA	M	Q	F	R	I	V	Q	L	R	V
Rat γA	M	Q	F	R	I	V	Q	L	R	V
Mouse γA	M	Q	F	R	I	V	Q	L	R	V
Human γB	M	Q	F	C	I	I	Q	L	R	V
Bovine γB	M	Q	F	R	I	V	Q	L	R	V
Rat γB	M	Q	F	R	I	V	Q	L	R	V
Mouse γB	M	Q	F	R	I	V	Q	L	R	V
Human γC	M	Q	L	C	I	V	Q	L	R	V
Bovine γC	M	Q	F	C	I	V	Q	L	R	V
Rat γC	M	Q	F	R	I	V	Q	L	R	A
Mouse γC	M	Q	F	R	I	V	Q	L	R	V
Rat γE	M	Q	F	R	I	V	Q	L	R	I
Mouse γE	M	Q	F	R	I	V	Q	L	R	I
Rat γF	M	Q	F	H	I	V	Q	L	R	I
Mouse γF	M	Q	F	R	I	V	Q	L	R	I
Human γS	A	M	I	R	V	I	Q	L	D	I
Bovine γS	A	M	I	R	V	I	Q	L	D	I
Mouse γS	A	M	I	R	V	I	Q	L	D	I
Human γN	V	Q	I	R	V	V	Q	L	E	V
Mouse γN	V	Q	I	R	V	V	Q	L	E	V
Lip shark γM1	V	Q	F	R	I	I	Q	F	R	I
Bullfrog γM1	M	P	F	R	L	I	Q	L	R	V
Catfish γM1	M	Q	F	R	I	L	M	Y	R	I
Carp γM1	M	Q	F	R	I	L	M	Y	R	I
Catfish γM2	M	Q	F	R	I	L	M	Y	R	I
Carp γM2	M	Q	F	R	I	L	M	Y	R	I
Lip shark γS1	V	Q	V	R	L	I	Q	F	R	I
Catfish γS1	M	Q	F	R	I	L	M	Y	R	M
Carp γS	V	Q	V	K	I	I	Q	L	E	L
Lip shark γS2	V	Q	V	R	P	I	L	F	R	I
Catfish γS2	M	Q	F	R	I	L	M	Y	R	I
Frog γII	M	Q	F	K	I	I	Q	F	K	V

APPENDIX B: PROTEIN PARAMETERS

Physical parameters of HγD-Crys mutants calculated using the ProtParam tool from ExPASy (http://us.expasy.org/tools/protparam.html).

Protein	Molecular Weight (Da)	ε_{280} (nm^{-1})
WT HγD-Crys	21,817	41.04
M43A HγD-Crys	21,757	41.04
Q54A HγD-Crys	21,760	41.04
Q54E HγD-Crys	21,818	41.04
F56A HγD-Crys	21,741	41.04
R79A HγD-Crys	21,732	41.04
I81A HγD-Crys	21,775	41.04
V132A HγD-Crys	21,789	41.04
Q143A HγD-Crys	21,760	41.04
Q143E HγD-Crys	21,818	41.04
L145A HγD-Crys	21,775	41.04
M147A HγD-Crys	21,757	41.04
V170A HγD-Crys	21,789	41.04
R79A/M147A HγD-Crys	21,672	41.04
Q54A/Q143A HγD-Crys	21,703	41.04
Q54E/Q143E HγD-Crys	21,819	41.04

APPENDIX C: PRIMERS FOR MUTAGENESIS

Oligonucleotide primers used during site-directed mutagenesis of HγD-Crys.

Name	Oligonucleotide sequence	T_M (°C)
M43A coding	5'-gga cag cgg ctg ctg ggc gct cta tga gca gcc c-3'	77
M43A noncoding	5'-ggg ctg ctc ata gag cgc cca gca gcc gct gtc c-3'	77
Q54A coding	5'-cgg gcc tcg cgt act tcc tgc gcc gcg gc-3'	78
Q54A noncoding	5'-gcc gcg gcg cag gaa gta cgc gag gcc cg -3'	78
Q54E coding	5'-cca act act cgg gcc tcg agt act tcc tgc gcc g-3'	73
Q54E noncoding	5'-cgg cgc agg aag tac tcg agg ccc gag tag ttg g-3'	73
F56A coding	5'-cct cca gta cgc cct gcg ccg cgg cga cta tgc-3'	77
F56A noncoding	5'-gca tag tcg ccg cgg cgc agg gcg tac tgg agg-3'	77
R79A coding	5'-ccg ctc ctg cgc cct cat ccc cca ctc tgg-3'	67
R79A noncoding	5'-cca gag tgg ggg atg agg gcg caa ggg cgc-3'	67
I81A coding	5'-cgc tcc tgc cgc ctc gcc cct cac tct ggc-3'	77
I81A noncoding	5'-gcc aga gtg agg ggc gag gcg gca gga gcg-3'	77
V132A coding	5'-gct gga ggg ctc ctg gga cct cta cga gct gtc c-3'	76
V132A noncoding	5'-gga cag ctc gta gag ggc cca gga gcc ctc cag c-3'	76
Q143A coding	5'-cca act acc gag gac ggg cgt acc tgc tga tgc c-3'	73
Q143A noncoding	5'-ggc atc agc agg tac gcc cgt cct cgg tag ttg g-3'	73
Q143E coding	5'-cca act acc gag gac ggg agt acc tgc tga tgc c-3'	72
Q143E noncoding	5'-ggc atc agc agg tact cc cgt cct cgg tag ttg g-3'	72
L145A coding	5'-cga gga cgg cag tac gcg ctg atg cca ggg-3'	74
L145A noncoding	5'-ccc tgg cat cag cgc gta ctg ccg tcc tcg-3'	74
M147A coding	5'-cgg cag tac ctg ctg gcg cca ggg gac tat agg c-3'	68
M147A noncoding	5'-gcc tat agt ccc ctg gcg cca gca ggt act gcc g-3'	68
V170A coding	5'-cca gag tgg gct ctc tga gga gag cca tag att tct cc-3'	70
V170A noncoding	5'-gga gaa atc tat ggc tct cct cag aga gcc cac tct gg-3'	70

APPENDIX D: ANALYSIS OF EQUILIBRIUM UNFOLDING/REFOLDING DATA

1. *Calculating guanidine hydrochloride concentrations*

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

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

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

2. *Two-state equilibrium unfolding/refolding*

The following is a derivation of the equation that describes two-state equilibrium unfolding/refolding transitions.

Assume a two-state mechanism where the protein undergoes a direct transition between the native (N) and unfolded (U) states:

$$N \underset{}{\overset{K_1}{\Leftrightarrow}} U$$

K_1 is the equilibrium constant for the reaction and can be expressed in terms of the concentration of native and unfolded protein:

$$(1) \qquad K_1 = [U]/[N]$$

The fraction of native (f_N) or fraction of unfolded (f_U) protein present at any point in the curve is equal to the concentration of native or unfolded protein divided by the total protein concentration (P_T).

$$f_N = [N]/[P_T]$$

$$f_U = [U]/[P_T]$$

or

(2) $$[N] = f_N*[P_T]$$

(3) $$[U] = f_U*[P_T]$$

We also know that the f_N and f_U sum to equal one:

(4) $$f_N + f_U = 1$$

Substituting (2) and (3) into (1), we now have an expression for the equilibrium constant in terms f_N and f_U:

$$K_1 = ([f_U]*[P_T])/([f_N]/[P_T])$$

Canceling out $[P_T]$, we see that:

$$K_1 = [f_U]/[f_N]$$

or

(5) $$f_U = K_1*f_N$$

(6) $$f_N = f_U/K_1$$

Substituting (5) and (6) into (4) and rearranging, we now have expressions for f_N and f_U in terms of K_1:

(7) $$f_N = 1/(1 + K_1)$$

(8) $$f_U = 1/(1/K_1+1)$$

The spectroscopic signal (Y) of a mixture of native and unfolded protein is equal to the f_N times the signal of the native protein (Y_N) plus the f_U times the signal of the unfolded protein (Y_U):

(9) $Y = Y_N * f_N + Y_U * f_U$

Y_N and Y_U vary linearly with concentration of denaturant ([den]). These lines are known as the native and unfolded baselines.

(10) $Y_N = Y_N° + S_N * [den]$

(11) $Y_U = Y_U° + S_U * [den]$

$Y_N°$ 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.

Substituting (10) and (11) into (9) yields the following relationship:

(12) $Y = (Y_N° + S_N * [den]) * (1/(1 + K_1)) + (Y_U° + S_U * [den]) * (1/(1/K_1 + 1))$

In equilibrium unfolding/refolding experiments, free energy of unfolding (ΔG_1) is determined in the transition region where the native and unfolded states are in equilibrium. ΔG_1 varies with denaturant concentration in a roughly linear fashion. Therefore, free energy of unfolding in the absence of denaturant ($\Delta G_1°$) can be determined by plotting ΔG_1 in the transition region versus the concentration of denaturant, fitting the data to a straight line and extrapolating back to 0 M denaturant. The slope of this line is known as the *m* value (m).

(13) $\Delta G_1 = \Delta G_1° - m[den]$

The free energy of unfolding in the absence of denaturant ($\Delta G_1°$) is related to K_1, the gas constant (R) and the temperature (T) as follows:

(14) $\Delta G_1° = - RT * \ln(K_1)$

Substituting (13) into (14) and rearranging, we get the following expression:

(15) $K_1 = \exp((m*[den] - \Delta G_1)/RT)$

Finally, substituting (15) into (12), we arrive at an expression for Y in terms of the native and unfolded baselines, and the $\Delta G_1°$ and *m* value of the transition:

(16) $Y = (Y_N° + S_N*[den])*(1/(1 + \exp((m*[den] - \Delta G_1)/RT))) + (Y_U° + S_U*[den])*(1/(1/\exp((m*[den] - \Delta G_1)/RT) + 1))$

Data were fit to the two-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 $cal*mol^{-1}*K^{-1}$.

```
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*

The following is a derivation of the equation that describes three-state equilibrium unfolding/refolding transitions.

Assume a three-state mechanism where a partially folded intermediate (I) is populated in equilibrium with the native (N) and unfolded (U) states:

$$N \underset{}{\overset{K_1}{\Leftrightarrow}} I \overset{K_2}{\Leftrightarrow} U$$

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:

(1) $K_1 = [I]/[N]$

(2) $K_2 = [U]/[I]$

The fraction protein found in the native (f_N), intermediate (f_I) or unfolded (f_U) conformation is related to the total protein (P_T) concentration as follows:

$$f_N = [N]/[P_T]$$
$$f_I = [I]/[P_T]$$
$$f_U = [U]/[P_T]$$

or

(3) $[N] = f_N*[P_T]$

(4) $[I] = f_I*[P_T]$

(5) $[U] = f_U*[P_T]$

The sum of f_N, f_I and f_U is 1:

(6) $f_N + f_I + f_U = 1$

Substituting (3) and (4) into (1), and (4) and (5) into (2), yields expressions for K_1 and K_2 in terms of f_N, f_I and f_U:

$$K_1 = (f_I*[P_T])/(f_N*[P_T])$$
$$K_2 = (f_U*[P_T])/(f_I*[P_T])$$

Canceling $[P_T]$ out of both expressions, we get the following relationships:

$$K_1 = f_I/f_N$$
$$K_2 = f_U/f_I$$

or

(7) $f_I = K_1*f_N$

(8) $f_U = K_2*f_I$

Substituting (7) and (8) into (6) and rearranging, we obtain expressions for f_N, f_I and f_U in terms of K_1 and K_2:

(9) $f_N = 1/(1 + K_1 + K_1*K_2)$

(10) $f_I = K_1/(1 + K_1 + K_1*K_2)$

(11) $f_U = K_1*K_2/(1 + K_1 + K_1*K_2)$

As with the two-state mechanism described above, the spectroscopic signal (Y) of a mixture of N, I and U is equal to the f_N times the signal of the native protein (Y_N) plus the f_I times the signal of the intermediate protein (Y_I) plus the f_U times the signal of the unfolded protein (Y_U):

(12) $Y = f_N*Y_N + f_I*Y_I + f_U*Y_U$

Y_N and Y_U vary linearly with concentration of denaturant ([den]). These lines are known as the native and unfolded baselines.

(13) $Y_N = Y_N° + S_N*[den]$

(14) $Y_U = Y_U° + S_U*[den]$

$Y_N°$ 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.

Substituting (9), (10), (11), (13), and (14) into (12) yields the following relationship:

(15) $Y = (Y_N° + S_N*[den])*(1/(1 + K_1 + K_1*K_2)) + Y_I*(K_1/(1 + K_1 + K_1*K_2)) + (Y_U° + S_U*[den])*((K_1*K_2)/(1 + K_1 + K_1*K_2))$

Free energies of unfolding for the native to intermediate (ΔG_1) and intermediate to unfolded transition (ΔG_2) are determined in the transition regions where the different conformations are in equilibrium. ΔG_1 and ΔG_2 vary with denaturant concentration in a roughly linear fashion. Free energy of unfolding in the absence of denaturant ($\Delta G_1°$ and $\Delta G_2°$) can be determined by plotting ΔG_1 and ΔG_2 in the transition region versus the concentration of denaturant, fitting the data to a straight line and extrapolating back to 0 M denaturant. The slope of this line is known as the *m* value (m).

(16) $\Delta G_1 = \Delta G_1° - m_1[den]$

(17) $\Delta G_2 = \Delta G_2° - m_2[den]$

$\Delta G_1°$ and $\Delta G_2°$ are related to K_1, K_2, the gas constant (R) and the temperature (T) as follows:

(18) $\Delta G_1° = - RT*\ln(K_1)$

(19) $\Delta G_2° = - RT*\ln(K_2)$

Substituting (16) and (17) into (18) and (19) and rearranging, we get the following expressions:

(20) $K_1 = \exp((m_1*[den] - \Delta G_1)/RT)$

(21) $K_2 = \exp((m_2*[den] - \Delta G_2)/RT)$

Finally, substituting (20) and (21) into (15), we arrive at an expression for Y in terms of the native and unfolded baselines, the signal of the intermediate, $\Delta G_1°$ $\Delta G_2°$ and *m* values of the transitions:

(16) $Y = (Y_N° + S_N*[den])(1/(1 + \exp((m_1*[den] - \Delta G_1)/RT) + \exp((m_1*[den] - \Delta G_1)/RT)* \exp((m_2*[den] - \Delta G_2)/RT)) + Y_I*(\exp((m_1*[den] - \Delta G_1)/RT)/(1 + \exp((m_1*[den] - \Delta G_1)/RT) + \exp((m_1*[den] - \Delta G_1)/RT)* \exp((m_2*[den] - \Delta G_2)/RT))) + (Y_U° + S_U*[den])((\exp((m_1*[den] - \Delta G_1)/RT)* \exp((m_2*[den] - \Delta G_2)/RT))/(1 + \exp((m_1*[den] - \Delta G_1)/RT) + \exp((m_1*[den] - \Delta G_1)/RT)* \exp((m_2*[den] - \Delta G_2)/RT))$

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 E: ANALYSIS OF KINETIC DATA

1. *Two-state kinetics*

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

$$A \xrightarrow{k_1} B$$

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

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:

$$A \xrightarrow{k_1} B \xrightarrow{k_2} C$$

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

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:

$$A \xrightarrow{k_1} B \xrightarrow{k_2} C \xrightarrow{k_3} D$$

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

Data were fit to the four-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;
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\;
```

"Did it get funny? Because the part I saw was just weird."