CONTRIBUTIONS OF AROMATIC PAIRS OF HUMAN GAMMA D-CRYSTALLIN TO ITS FOLDING, STABILITY, AGGREGATION, AND INTERACTION WITH HUMAN ALPHA B-CRYSTALLIN

by

Fanrong Kong

B.S. Biological Sciences
Cornell University, 2007

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

Doctor of Philosophy

at the
Massachusetts Institute of Technology

February 2012

Signature of Author _____

Department of Biology
February, 2012

Certified by _____

Jonathan King
Thesis Supervisor

Accepted by _____

Robert T. Sauer
Chairman, Department Committee for Graduate Students

CONTRIBUTIONS OF AROMATIC PAIRS OF HUMAN GAMMA D-CRYSTALLIN TO ITS FOLDING, STABILITY, AGGREGATION, AND INTERACTION WITH HUMAN ALPHA B-CRYSTALLIN

by

Fanrong Kong

Submitted to the Department of Biology at the Massachusetts Institute of Technology
on February 6, 2012 in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Biology

ABSTRACT

Two distinct groups of proteins, α-crystallins and βγ-crystallins, constitute 90% of the vertebrate eye lens soluble proteins. Long-term solubility and stability against unfolding and aggregation are essential properties of crystallins and crucial to the function of the lens. Aggregation of crystallins in the lens causes light scattering and directly contributes to development of cataract, the leading cause of blindness in the world. The amino acid determinants of these biochemical/biophysical properties of crystallins are not entirely understood. Aromatic residues in proteins have been shown to be important determinants of their folding pathways, native-state stability, aggregation propensity and other intermolecular interactions. In this thesis study, I have investigated the contributions of the paired aromatic residues of human γD-crystallin (HγD-Crys) to its folding, stability, aggregation, and interaction with the chaperone human αB-crystallin (HαB-Crys).

HγD-Crys is a highly stable protein that remains folded in the nucleus of the eye lens for the majority of an individual's lifetime. Like other βγ-crystallins, HγD-Crys exhibits two homologous crystallin domains, each containing two Greek key motifs and eight β-strands. Six conserved aromatic pairs (four Tyr/Tyr, one Tyr/Phe and one Phe/Phe) are present in HγD-Crys. Four among them are located at conserved β-hairpins of the Greek key motifs, thus termed "Greek key pairs". The Greek key pairs have the consensus sequence Y/FXXXXY/FXG and are one of the defining features of the βγ-crystallin family. Ultraviolet (UV) damage to these aromatic residues in βγ-crystallins may contribute to unfolding and aggregation of the proteins, leading to development of cataract.

α-Crystallins belong to the small heat shock protein (sHsp) family and have both structural and chaperone functions in the lens. Human α-crystallins form polydisperse oligomers of 15-60 subunits, with αA:αB ratio about 3:1 *in vivo*. The core α-crystallin domain (αCD) of α-crystallins has an immunoglobulin (Ig)-like β-sandwich fold, but the quaternary structure of α-crystallin remains to be fully solved. Like other sHsps, α-crystallins exert their chaperone function by sequestering partially-unfolded and aggregation-prone substrates in an ATP-independent manner, thus preventing their aggregation. The chaperone-substrate interactions of α-crystallins and other sHsps remain poorly understood.

To investigate the roles of the paired aromatic residues in HγD-Crys, mutant proteins with these aromatic residues substituted with alanines were constructed and expressed in *E. coli*. All mutant proteins maintained native-like secondary structures by circular dichroism (CD). Except F115A and F117A, all mutant proteins had lower thermal stability than the wildtype (WT) protein. Equilibrium unfolding/refolding experiments in guanidine hydrochloride (GuHCl) showed that all mutant proteins had lower thermodynamic stability than the WT protein. N-terminal domain (N-td) substitutions shifted the N-td transitions to lower GuHCl concentrations, but the C-terminal domain (C-td) transitions remained unaffected. C-td substitutions led to a more synchronized unfolding/refolding process of the N-td and C-td, and the overall transitions shifted to lower GuHCl concentrations. These results were consistent with a sequential unfolding/refolding model of HγD-Crys, in which the N-td unfolds first and refolds last. The Greek key pairs had larger contributions to both thermal stability and thermodynamic stability than the non-Greek-key pairs. Aromatic-aromatic interaction energy was estimated by double mutant cycles as 1.5-2.0 kcal/mol.

To distinguish the effects in unfolding and refolding, kinetic experiments were also performed. In kinetic unfolding experiments, N-td substitutions accelerated the early phase of unfolding, while C-td substitutions accelerated the late phase. For refolding, only substitutions of the second Greek key pair of each crystallin domain slowed refolding: N-td substitutions Y45A and Y50A affected the late phase while Y133A and Y138A affected the early phase of the overall refolding reaction. The second Greek key may serve as a nucleation site during the folding of the double-Greek-key crystallin domain.

The aggregation pathway that competes with productive refolding *in vitro*, as well as the suppression of aggregation by chaperone HαB-Crys were also investigated for mutant HγD-Crys variants with replacements of the Greek key paired residues. The WT and the mutant HγD-Crys behaved very similarly, in term of aggregation kinetics and final aggregation level, indicating that these aromatic pairs played minimal role in the aggregation process. The efficiencies of aggregation suppression by HαB-Crys, as well as the HαB-bound conformations characterized by the fluorescence of HαB-HγD complexes were also very similar for the WT and mutant HγD-Crys, arguing against a critical role of these aromatic pairs in the chaperone recognition process. Together with the earlier results, it can also be concluded that stability and refolding kinetics of HγD-Crys were not critical determinants for its refolding-induced aggregation as well as HαB-Crys recognition. Both of these processes may rely on features of the folding intermediate in a very early stage of Greek key refolding. The tryptophan fluorescence of the HαB-HγD complexes with WT or mutant HγD-Crys resembled a partially unfolded state of HγD-Crys, consistent with the tryptophans being part of the contact sites with the chaperone HαB-Crys.

Thesis Supervisor: Jonathan King, Professor of Biology.

ACKNOWLEDGEMENTS

The completion of this thesis would not be possible without the guidance, support and encouragement from many people. My advisor, Prof. Jonathan King, you are in no doubt the first person on this list. Your influence covered all aspects of my work and life for the past few years. Thank you for your scientific advices throughout my entire time in the lab, and for guiding this thesis project in a very productive and rewarding direction. Thank you for your efforts in enhancing my writing and presentation skills, your invaluable foresight and advices on my postdoctoral search, as well as your psychological encouragement during my depressed moments. Thank you for the scientific philosophy that you have planted in our minds, the anecdotes you have shared, and the delicious holiday meals.

To Prof. Robert Sauer and Prof. Amy Keating, thank you for serving on my prelim and thesis committee over the last few years. Your insightful comments, advices and criticisms have broadened my vision in protein biochemistry. To Prof. Jonathan King, Prof. Gerald Feigenson and Prof. Amy Keating, thank you for taking your time to write those postdoctoral reference letters for me. They have earned me important interview opportunities, and eventually a great postdoctoral position—not easy under this difficult funding climate. To Prof. Gene Brown, Prof. Michael Yaffe and Prof. Matthew Vander Heiden, thank you for your assistance in solving numerous scientific and logistic complications in the 7.05 class. Without your support, I would not be able to complete my head TA duties.

I would like to thank the entire King Lab, past and current members, for sharing your scientific expertise on protein science, and creating a friendly and cheerful working environment. Dr. Jiejin Chen and Prof. Yongting Wang, thank you for orientating me to the lab environment and teaching me experimental techniques in my rotation and initial period in the lab. Dr. Kate Moreau and Dr. Ligia Acosta-Anastasopoulos, with the tiny space here, I could never thank you enough for your generous sharing of experimental experiences and materials you have generated through hard work. Thank you for tirelessly answering my naïve questions on protocols, data analysis, and much more. Dr. Kelly Knee, Dan Goulet and Jeannie Chew, thank you for creating such a beautiful facet on the chaperonins for the lab. Your expertise on chaperones have enriched my knowledge on protein folding. By the way, I found chaperonins the most beautiful-looking proteins in the world. Jeannie Chew, your swift pace always energized the lab in drowsy summer afternoons. Nathaniel Schafheimer and Oksana Sergeeva, my beloved neighbors in the lab, thank you for sharing funny and/or ridiculous news happening around the world (so I don't have to read the news). Your diligent and upbeat spirit in face of experimental difficulties has always been inspiring me and will continue to benefit me for the years to come. I wish your projects both blossom. Cammie Haase-Pettingell, Dessy Raytcheva, Prof. Jacqueline Piret and Dr. Takumi Takata, thank you for sharing your expertise on phage assembly and maintaining the phage tradition in the lab. Dr. Ishara Mills-Henry, thank you for sharing your insightful knowledge on the folding model of crystallins. They were really helpful in constructing my manuscript. Dr. Ishara Mills-Henry and Lisa Guisbond, thank you for inviting us to the science education events. I admire your passion and dedication in science education. Cammie Haase-Pettingell, Cindy Woolley, and Althea Hill, thank you for all the behind-the-scene work to keep things in order for the lab. I truly appreciate your hard work.

The Ph.D. life would not be bearable without the fun times outside the lab with many science and non-science friends of mine. I really enjoyed the outings to the wilds and the hot-pot gatherings we had. I am sorry to miss many of them recently. A full name list seems impossible here, but my housemates, Albert Cheng and Alan Tang are impossible not to mention. Thank you for your loyal company throughout these years. Thank you for playing video games with me, listening to my bad jokes, and tolerating some of my incomprehensible living habits. I will miss you when I leave Boston.

To my girlfriend Ida Lam, the working hours of a graduate student is hard to understand for outsiders. Thank you for forgiving me setting our dating time around my experiment schedule. The emotional stress of a graduate student is hard to understand for outsiders. Thank you for your listening and encouragement. Whenever I have been depressed or tired, you have been the most effective medicine on earth. Some say the distance between two cities is the worst enemy of a relationship, but together we beat it with lots of 4-hour bus rides. I really look forward to the time after the defense of this thesis, by then we can enjoy lots of carefree time. To my parents Qingmei Kong and Rumei Ruan, you may not understand the science that I am doing at MIT, but you have been giving your unconditional love to me and encouragement to whatever I am doing. Thank you for the tremendous sacrifice you have made for my education. Thank you for the echoic reminders over the phone of staying warm and healthy. Thank you for the delicious dishes you prepared whenever I came home. In nowhere I can find a place better than home.

This research was funded by an NIH grant (GM17980), an NEI grant (EY015843), and an NIH/NIGMS training grant (GM007287).

BIOGRAPHICAL NOTE

Fanrong Kong

Education

Ph. D. Expected 2012	Massachusetts Institute of Technology, Department of Biology, Cambridge, MA.
B. S. May 2007	Cornell University, Department of Molecular Biology and Genetics, Ithaca, NY. Major: Biological Sciences, *Magna Cum Laude.*

Research and Professional Experience

2008 – 2012	Graduate Research Assistant Laboratory of Professor Jonathan King, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA.
2006 – 2007	Undergraduate Research Assistant Laboratory of Professor Gerald W. Feigenson, Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY.
Summer 2005	Undergraduate Research Assistant Laboratory of Professor Nicole S. Sampson, Department of Chemistry, Stony Brook University, Stony Brook, NY.

Publications

Kong, F. and King, J. 2011. Contributions of aromatic pairs to the folding and stability of long-lived human γD-crystallin. *Protein Science*, 20(3): 513–528.

Zhao J., Wu J., Shao H., **Kong F.**, Jain N., Hunt G., Feigenson G. 2007. Phase studies of model biomembranes: Macroscopic coexistence of Lα + Lβ, with light-induced coexistence of Lα + Lo phases. *Biochimica et Biophysica Acta - Biomembranes*, 1768(11): 2777-2786.

Kong, F. and King, J. 2012. Aggregation of Greek key aromatic mutant human γD-crystallins and their interactions with human αB-crystallin. *Protein Science*, (submitted).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abs: absorbance
αCD: α-crystallin domain
BFSP: beaded filament structural protein
bis-ANS: 4,4'-bis(1-anilinonaphthalene 8-sulfonate)
CD: circular dichroism
C-td: C-terminal domain
CV: column volume
DTT: dithiothreitol
EDTA: ethylenediaminetetraacetic acid
EM: electron microscopy
FPLC: fast protein liquid chromatography
FRET: fluorescence resonance energy transfer
GuHCl: guanidine hydrochloride
Hsp: heat shock protein
HαA-Crys: human αA-crystallin
HαB-Crys: human αB-crystallin
HγC-Crys: human γC-crystallin
HγD-Crys: human γD-crystallin
IB: inclusion body
IDP: intrinsic disordered protein
Ig: immunoglobulin
IPTG: isopropyl β-D-1-thiogalactopyranoside
MS: mass spectrometry
MWCO: molecular weight cutoff
Ni-NTA: nickel-nitrilotriacetic acid
NMR: nuclear magnetic resonance
N-td: N-terminal domain
OD: optical density
PCO: posterior capsule opacification
PCR: polymerase chain reaction
pI: isoelectric point
ROS: reactive oxygen species
RT: room temperature
SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC: size exclusion chromatography
sHsp: small heat shock protein
ThT: Thioflavin T
TRiC: TCP-1 Ring Complex
Tris: tris(hydroxymethyl)aminomethane
UV: ultraviolet
V_0: void volume
WT: wildtype

Chapter 1

Introduction

A. Significance of Studying Protein Folding, Stability and Aggregation

1. Protein Folding as a Fundamental Biological Quest

Inside the cell, newly synthesized proteins emerge from the ribosome exit channel as unstructured nascent chains. The versatile functions of most proteins, however, depend on their unique three-dimensional structures. The process by which an unstructured polypeptide chain transforms into a well-defined three-dimensional structure is defined as protein folding (Dobson 2004). Half a century ago, Christian Anfinsen made the classic discovery that the native structure of bovine pancreatic RNAse was encoded solely in the amino acid sequence (Sela *et al.* 1957; Anfinsen 1972; Anfinsen 1973). Subsequently, Cyrus Levinthal postulated that proteins simply could not find their native structures through random research due to the astronomical time needed, known as the "Levinthal's paradox" (Levinthal 1969). This suggested the existence of folding pathways. These two pioneers have inspired decades of research in the rules that govern protein folding thereafter. It is now generally accepted that amino acid sequences determine not only the native structure, but also the folding pathway, as well as the susceptibility to off-pathway aggregation.

For some proteins, the kinetics and/or efficiency of protein folding *in vivo* may be influenced by additional factors. It has been postulated that rare codons on an mRNA sequence can pause the translation process, thus allowing co-translational folding of defined folding units within the protein before undesirable interactions occur (Komar 2009). Interestingly, a strong correlation was found between the locations of these rare codons and the boundaries between structural domains or elements, but not ~30 amino acids downstream of such boundaries if the protection by the ribosome exit tunnel is taken into consideration. Therefore the author suggested that co-translational folding units may not correlate with structural domains or elements of the native state (Komar 2009). Besides mRNA sequence, molecular chaperones can facilitate protein folding through active or passive mechanisms, and they fall into diverse families described in Section D below (Hartl and Hayer-Hartl 2009). The roles of chaperones are largely "catalytic", hence they do not determine the native structures of their substrates.

Several general protein folding mechanisms have been proposed. First, the nucleation growth model suggested existence of an independent folding region, called the nucleus. Folding propagates from one or more local nuclei to the whole protein (Kang and Kini 2009; Wathen and Jia 2010). Second, the framework or diffusion-collision model proposed local secondary structures form autonomously, and then diffuse together. Subsequent stabilizing interactions among the folded secondary structures generate and fine-tune the tertiary structure (Fersht 1997; Kang and Kini 2009). Third, the hydrophobic collapse model suggested that a gross collapse of the hydrophobic residues occurs initially, perhaps to an initial "molten globule" state, to give rise to a native-like tertiary structure. The answer to the question "what folds first" is the key in distinguishing these theories, with the answers being local structure, secondary structure, and tertiary structure, respectively. Based on φ-value analysis on chymotrypsin inhibitor 2 (CI2), Fersht and colleagues further suggested the nucleation-condensation model that seemed to blur the line between the second and third models. This new model called for a new kind of folding nucleus different from the classical local nucleus, i.e., the residues most native-like in the transition state (Otzen *et al.* 1994; Fersht 1997).

In studies pioneered by Dill and colleagues, the "new view" of protein folding described protein folding as a stochastic process on a funnel-like energy landscape, and has modified the traditional view of a single mandatory pathway (Figure 1-1) (Dill and Chan 1997; Dinner *et al.* 2000). The energy landscape concept has the advantage of describing different protein folding mechanisms through modifications of the energy surface, and therefore has the potential of unifying different mechanisms of protein folding.

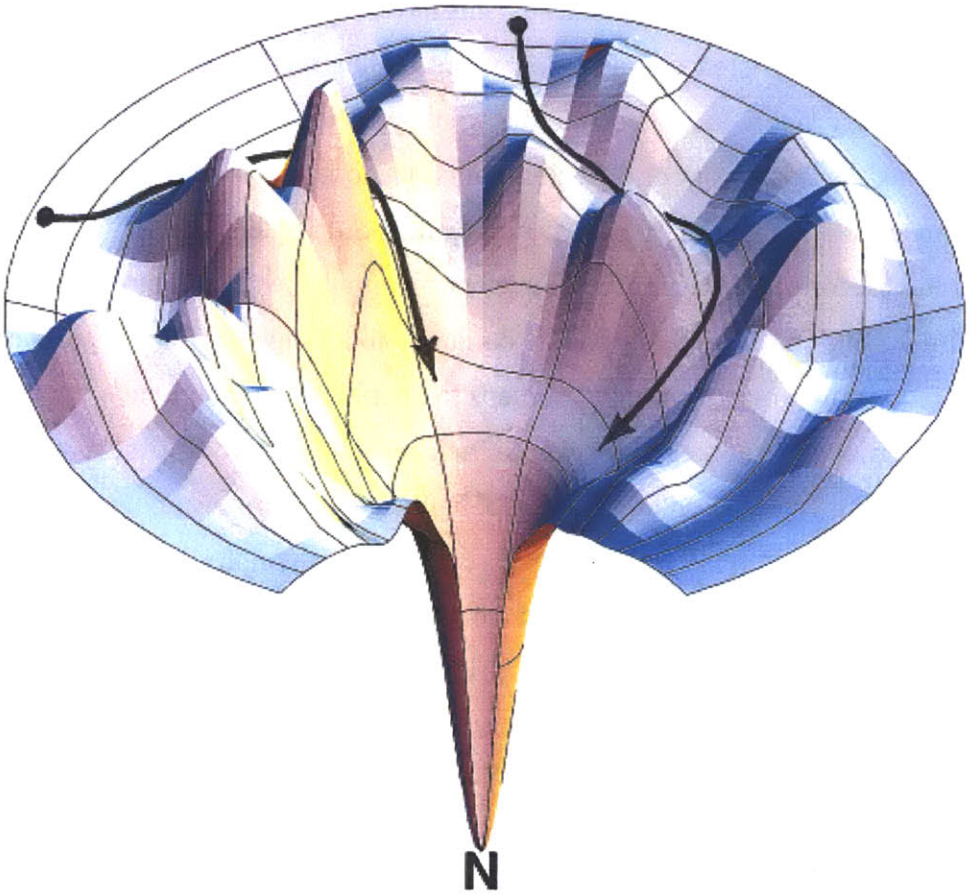


Figure 1-1 Protein Folding Energy Landscape.

A "rugged" landscape probably reflects a more realistic picture of protein folding. The energy wells, depending on the location and depth, may represent on-pathway intermediates or off-pathway misfolded species. The energy barriers may represent the transition states of folding. A protein folds through "some narrow throughway paths to native" (Dill and Chan 1997).

(Image reprinted with permission from Chan and Dill 1998)

The emergence of intrinsic disordered proteins (IDP) has added a twist in the field of protein folding. IDPs are a class of proteins that under physiological condition lack well-defined three-dimensional structures, either in certain regions or in the entire sequence. The widespread appearance and functional importance of this group of proteins have made IDPs a distinct field in structural biology, and have challenged the traditional sequence-structure-function paradigm (Uversky and Dunker 2010; Uversky 2011). α-Synuclein, the key protein in Parkinson's disease is one example of an IDP, while p53, the tumor suppressor involved in 50% of all cancers, has been suggested to have IDP domains. The structures of certain regions of p53 depend on the interacting partners, allowing it to be a signaling hub with many binding partners (Dunker *et al.* 2005). The emergence of IDPs serves as an extension of the protein folding problem. It highlights the urgency to identify amino acid determinants of both protein folding and "non-folding".

β-Sheet proteins are a particularly difficult part of the protein folding problem and progress in their understanding is lagging behind other simpler proteins. Unlike α-helices, β-sheets involve contacts distant in the amino acid sequence. How these distant contacts are brought together precisely is a key question in β-sheet folding mechanisms. Dill *et al.* proposed the hydrophobic zipper model in which β-sheet formation is initiated by local contacts, such as the β-turn region of a β-hairpin, followed by the growth of the regions successively farther away from these local contacts in a zipper-like fashion (Dill *et al.* 1993). In such a model, established earlier contacts bring the later contacts close in space and greatly reduce the large conformational search space for long-distance contacts. Supporting this model, φ-value studies on some β-sheet model domains, src SH3 domain and WW domains, revealed that two β-turns and an adjacent hydrophobic cluster in the SH3 domain, and β-turn 1 in the WW domains formed in the transition states (Grantcharova *et al.* 1998; Deechongkit *et al.* 2006). In addition, β-turn types, as well as the length of the connecting sequences from the turn to the hydrophobic stabilizing elements, had influenced the refolding kinetics of β-sheet peptides (Dyer *et al.* 2004; Kuo *et al.* 2005). These experimental investigations, together with theoretical and computational studies have established a prevalent view on the folding of β-hairpins: early turn formation controls the kinetics, while the subsequent cross-strand contacts provide thermodynamic stabilization (Wathen and Jia 2010). However, β-hairpin/β-sheet formation is a complicated interplay among the β-turns, the extended β-strands, and the cross-strand interacting elements (Dyer *et al.* 2004).

A complete picture of β-hairpin/β-sheet formation remains to be solved. For example, an alternative model different from the zipper model has also been proposed. Computational studies on a β-hairpin peptide from streptococcal protein G suggested an early folding nucleus formed by hydrophobic collapse, from which hydrogen bonds propagated in both directions. The subsequent interconversion between compact conformations was determined to be the rate-limiting step in such a model (Dinner *et al.* 1999).

Despite the progress made, many questions remain. Except for a few relatively straight-forward cases, such as the G-X-Y (X often proline, Y often hydroxyproline) tripeptide repeat for collagen (Kuivaniemi *et al.* 1991), and the heptad repeat for α-helical coiled-coils (Grigoryan and Keating 2008), the fundamental question of how amino acid sequences guide the protein folding process and determine the final structures remains to be fully answered.

2. *Protein Stability: a Delicate Balance between Folding and Unfolding*

After a protein successfully folds into the native structure, it needs to stay folded for an appropriate period of time in order to function. Proteins are only marginally stable—free energy differences between the native and the unfolded states of typical proteins are roughly 5-20 kcal/mol (Dill 1990). A team of non-covalent interactions, including hydrogen bonds, ionic bonds, aromatic interaction, and hydrophobic packing work together to contribute to the intrinsic stability of proteins (Jaenicke and Slingsby 2001). Many extracellular proteins are additionally stabilized by covalent disulfide bonds. Non-covalent forces are often weak individually, but with a large number, they can have a significant effect on protein stability. Specific residues may have distinct roles in folding vs. stability. Studies on yeast phosphoglycerate kinase (yPGK) showed that the loops between secondary structures did not play an active role in folding, but contributed to stability by increasing the unfolding rate (Collinet *et al.* 2001). In the simplest folding model with a single transition state, whether an interaction is formed before or after the transition state determines its specific role in folding vs. unfolding. On the other hand, A334I substitution on phage P22 tailspike accelerated folding, but decreased thermal stability. The authors suggested that the bulky isoleucine improved the packing of the loosely structured early folding intermediate, but could not be accommodated in the compact native structure (Beissinger *et al.*

1995). This study highlighted the possibility for a folding intermediate to have non-native amino acid interactions.

Besides the intrinsic stability encoded in each protein's amino acid sequence, a number of environmental factors can also influence protein stability. Extreme pH, ionic strength, temperature, or pressure can destabilize or unfold proteins. In this sense, maintaining protein stability posts a challenge to extremophiles, organisms living in extreme environments (Jaenicke 1996; Jaenicke and Bohm 1998). Comparing protein homologues from extremophiles and mesophiles, the ΔG's between native and the unfolding states are on the same order of magnitude (Jaenicke and Bohm 1998). Therefore, only a few amino acid changes may be required to confer the extreme stability. For example, two cold shock proteins (Csp) from extremophilic and mesophilic origins differ in sequence at 12 amino acid positions, but 2 positions fully accounted for their difference in thermodynamic and thermal stabilities (Perl *et al.* 2000). Although no single magic strategy of protein stabilization stood out when comparing extremophilic and mesophilic protein homologues, the most common strategies identified were increasing the number of ion pairs and hydrogen bonds, and improving the hydrophobic packing of protein cores and between subunits. Loops and termini shortening/ strengthening, constraint relief by glycines, increased helicity and helix capping were also found frequently (Querol *et al.* 1996; Jaenicke and Bohm 1998).

Conformational destabilization may render proteins susceptible to more severe covalent damages by agents such as reactive oxygen species (ROS) and ultraviolet (UV) light. These covalent damages are irreversible, further destabilizing the proteins. In metabolically active cells, damage proteins are usually degraded by proteases or proteasomes. However, in metabolically inert cells, such as the eye lens fiber cells, where no protein turnover takes place due to loss of organelles, long-term protein stability becomes an essential requirement (Oyster 1999; Bassnett 2002).

Understanding the determinants of protein stability allows us to better predict protein structures and to engineer protein variants with enhanced stability through rational design, although methods that do not require prior knowledge also exist, such as directed evolution and consensus approach. Directed evolution has the advantage of covering a large sequence space, while consensus method offers an immediate solution of potential large-scale stability enhancement (Lehmann and Wyss 2001). Rational design, on the other hand, provides a direct

test of our knowledge on protein stability. In the case of the subunit binding domain of dihydrolipoamide acetyltransferase, replacement of Arg8 with isosteric neutral or negatively charged residues improved thermodynamic stability by ~1 kcal/mol and thermal stability by ~10 °C, offering a verification of the theoretical calculation (Spector *et al.* 2000). The improvement was achieved by alleviation of unfavorable electrostatic interactions on a surface charged patch.

In a cellular environment, however, high stability is not the only goal for all proteins. Enzymes and receptors may need a certain degree of flexibility to make conformational changes upon binding of substrates and ligands. A certain degree of unfolding tendency is also needed for most proteins for degradation at the end of their lives. Therefore, having the right balance between the two opposing forces: unfolding and refolding, is a property of proteins that billions of years of evolution has fine-tuned since their first appearance in the living world.

3. *Misfolding and Aggregation: the Dark Side of Proteins*

Protein folding is a complicated process—errors can happen. Disrupted protein folding pathways and/or reduced stability can be caused by congenital mutations, improper posttranslational modifications, or environmental factors, resulting in partially folded or misfolded species. Failure in protein folding is intimately related to protein aggregation, because the misfolded or partially folded species often expose their hydrophobic cores or other elements normally buried inside the native protein, enhancing the intermolecular interactions (Wetzel 1996). The total accessible states of a polypeptide are numerous, and they form an interconnected network (Figure 1-2) (Dobson 2003). Most conformational conversions in this network are reversible, but the aggregation pathways are largely one-way. The pathways in this network are also closely regulated by molecular chaperones (Dobson 2003). Protein aggregation may occur to cytoplasmic proteins inside the cell or to secreted proteins in the extracellular space.

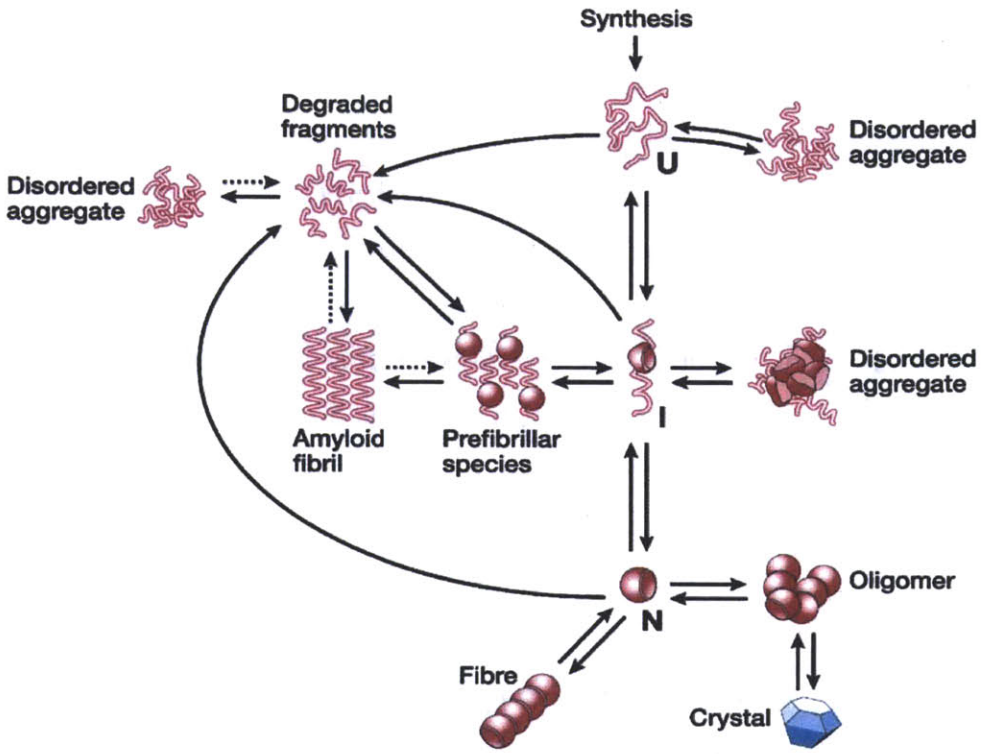


Figure 1-2 A Folding/Aggregation Network of Polypeptides.

Many conformational states of a polypeptide, monomeric or polymeric, can exist after its synthesis. The interconvertions among these states may be reversible or irreversible. The different modes of associations have their corresponding sequence and conformational requirements. In cells, these pathways are closely regulated by molecular chaperones. (Reprinted by permission from Macmillan Publishers Ltd: *Nature Reviews Drug Discovery*, Dobson, copyright (2003)).

In the light of the common structure of amyloid fibrils formed by unrelated amino acid sequences, it was proposed that all polypeptides/proteins possess a "dark side"—the inherent aggregation tendency that largely rely on the polypeptide backbones (Stefani 2004). Aggregation is obviously detrimental to the functions of soluble proteins. Therefore, during evolution, sequences that hinder aggregation must be selected in the first place, followed by their ability to perform biological functions (Stefani 2004). Aggregation can also occur to native-like proteins without substantial structural change (often called polymerization), such as seen in the polymerization of actin, tubulin, or sickle-cell hemoglobin. Polymerization of native-like proteins depends on surface features of proteins. Other modes of aggregations include domain-swap aggregation and amorphous aggregation. Detailed differences among these modes of aggregations will be discussed in Section C of this introduction.

Protein misfolding and aggregation is not just an important biological problem, it is also implicated in many human diseases (Table 1-1) (Dobson 2001; Stefani 2004). The exact causes of many protein misfolding/aggregation diseases are unknown, and may be one of the following: 1) loss of the soluble functional proteins; 2) binding of the small oligomers or large aggregates to other proteins, abolishing or altering their functions (Olzscha *et al.* 2011); 3) altered interactions with the chaperones and/or other protein quality control systems, such as in some forms of cystic fibrosis (Welsh and Smith 1993); 4) the aggregates directly causing the diseases (such as in cataract). It is very important to distinguish these mechanisms. Depending on the role of aggregation in the etiology, intervention that aims to reduce or enhance aggregation may be therapeutically useful (Bodner *et al.* 2006).

Disease	Protein	Site of folding
Hypercholesterolaemia	Low-density lipoprotein receptor	ER
Cystic fibrosis	Cystic fibrosis trans-membrane regulator	ER
Phenylketonuria	Phenylalanine hydroxylase	Cytosol
Huntington's disease	Huntingtin	Cytosol
Marfan syndrome	Fibrillin	ER
Osteogenesis imperfecta	Procollagen	ER
Sickle cell anaemia	Haemoglobin	Cytosol
αl-Antitrypsin deficiency	αl-Antitrypsin	ER
Tay–Sachs disease	β-Hexosaminidase	ER
Scurvy	Collagen	ER
Alzheimer's disease	Amyloid-peptide/tau	ER
Parkinson's disease	α-Synuclein	Cytosol
Scrapie/Creutzfeldt–Jakob disease	Prion protein	ER
Familial amyloidoses	Transthyretin/lysozyme	ER
Retinitis pigmentosa	Rhodopsin	ER
Cataracts	Crystallins	Cytosol
Cancer	p53	Cytosol

Table 1-1 Representative Protein Misfolding/Aggregation Diseases.
(Reprinted with permission from Dobson 2001).

Besides the strong link to human health, protein misfolding/aggregation problem also hampers biomedical research and the pharmaceutical industry. Recombinant protein expression, either small-scale in biomedical research or large-scale in the pharmaceutical industry, may results in intracellular aggregates, called inclusion bodies (IB), that significantly lower the soluble protein yields (Mitraki and King 1989). Overexpression of a β-galactosidase fusion protein showed that IBs tended to occupy the poles of *E. coli* cells with 1-2 IBs per cell. The aggregates consisted mainly of the overexpressed protein, but also contained small amount of contaminants such as ribosomal components and chaperones such as DnaK and GroEL (Carrio *et al.* 1998). High protein concentration and heterologous cellular environment that promotes incorrect modifications and aberrant protein interactions are the major drivers of IB formation. Weaker promoters and lower temperatures are often used to inhibit IB formation, but may also lower soluble protein production. Co-expression of chaperones is another extensively used strategy (Baneyx and Mujacic 2004). If IB formation seems difficult to avoid, complete unfolding of the IBs, followed by a graduate dialysis removal of denaturants has also been used to recover soluble proteins. However, success is not universal. In the case of tryptophanase, the soluble yields of the renaturation process increased significantly with protein concentration and slightly in the presence of a co-factor, but was independent of co-factor concentration and temperature (London *et al.* 1974). Interestingly, an intermediate renaturation step in around 3 M urea, but not higher or lower concentration, significantly decreased the soluble yield, presumably due to a specific aggregation-prone folding intermediate populated in 3 M urea (London *et al.* 1974). Effective strategies to avoid IB formation and to improve soluble protein yields require a thorough understanding of the determinants of both the productive folding and the off-pathway aggregation processes.

4. Functional Aggregates: Make Use of Protein Polymers

Nature exploits every aspect of a polypeptide chain for properties useful for biological functions. Although most proteins function in soluble forms, proteins in the aggregated state can also perform a diverse array of biological functions. Examples can be found in both the microbial world and higher organisms (Table 1-2) (Stefani 2004). Mechanical strength provided by protein polymers is easy to envision. For example, cytoskeletal filaments such as actin and tubulin, are protein polymers that play important roles in cellular morphology, motility, transport, and division. The amyloid-like structure of chorion provides both mechanical and chemical protection for silkmoth oocytes (Iconomidou *et al.* 2000). A second important function for protein aggregates is to quarantine potentially harmful materials. The polar positioning of inclusion bodies in *E. coli* ensures asymmetric inheritance of damaged proteins and may enhance the fitness of bacterial population (Winkler *et al.* 2010). In eukaryotes, amyloid fibrils formed by Pmel17 could act as scaffolds for melanin pigment polymerization, as well as for localizing the highly toxic melanin precursors (Berson *et al.* 2003; Fowler *et al.* 2006). Other quite surprising functions of amyloid fibrils are also found, including cell adhesion, cell killing, self-recognition, vascular sealing and peptide hormone storage (Table 1-2) (Stefani 2004; Maji *et al.* 2009).

Protein	Source	Function
Bacterial toxins?	bacteria	cell killing
Curlin	*E. coli*	substrate adhesion
Chaplins	*S. coelicolor*	aerial hyphae formation
HET-s prion	*P. anserina*	self/non-self recognition
Hydrophobins?	fungi	fungal coat
Chorion proteins	silkmoth	egg and embryo protection
Crystallins	mammals	refractive index of the lens
Pore proteins?	mammals	cell killing/apoptosis
SAA?	mammals	cell killing
Aβ peptides	human	small cerebral vessel sealing
Pmel17 (fragment)	human	melanine polymerization

Table 1-2 Amyloid with Biological Functions.

(Reprinted from *Biochim Biophys Acta.*, 1739(1): 5-25, Stefani, Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world. Copyright (2004), with permission from Elsevier).

B. Aromatic Interactions in Proteins

Aromatic interaction is both hydrophobic and electrostatic in nature. Due to their distinct geometry and charge distribution, aromatic rings and their interactions have some special characters. Four of the twenty amino acids in proteins and all five nucleotides in nucleic acids have aromatic groups. Aromatic interactions have important structural roles in these biological molecules (Meyer *et al.* 2003).

1. Criteria for Aromaticity

From Hückel's criteria for aromaticity, an aromatic molecule must have 1) a planar, cyclic geometry; 2) continuous overlap from a p orbital on each atom on the ring; 3) a total of 4n+2 electrons in the p orbitals (n being non-negative integers) (Brown and Foote 2002). Aromatic molecules may contain a single ring (monocyclic) or fused rings (polycyclic), composed of all carbon atoms (homocyclic) or a mix of carbon and other atoms (heterocyclic). Aromatic rings may also be neutral or charged (Brown and Foote 2002). Benzene is the simplest aromatic molecule. Aromatic molecules are extremely stable chemically because of the high degree of π orbital overlap (resonance). In addition, all the π bonding molecular orbitals are occupied by electron pairs, while all the antibonding and non-bonding orbitals are empty. These π electrons form two highly delocalized "donut-shaped" electron clouds above and below the ring. These partial charges and the geometry of the aromatic rings are key factors in aromatic interactions.

2. Aromatic Amino Acids

According to Hückel's criteria for aromaticity, four of the twenty natural amino acids in proteins have aromatic side chains: phenylalanine, tyrosine, tryptophan, and histidine (Figure 1-3). The side chains of these anatomic amino acids, without the connecting -CH_2- group, are equivalent to aromatic compounds benzene, phenol, indole, and imidazole, respectively (Figure 1-3). His is charged below around pH 6.5, and is considered as a charged amino acid in some classification schemes. The electrostatic interaction from a full charge probably dominates over that from the partial charge of the π electrons. However, protonation does not alter the molecular orbitals and the π electron system, therefore does not affect the aromaticity of His. From an extensive statistical analysis from more than 1000 unrelated proteins, aromatic amino acids were found to have generally low frequencies of occurrence in proteins (McCaldon and Argos 1988). Trp had the frequency of 1.3%, the lowest among all amino acids.

The aromatic amino acids are largely responsible for the UV absorbance and florescence properties of proteins (Creighton 1993). Among them Trp is the most UV absorbing and fluorescent amino acid (Figure 1-4). Photo-oxidation of the aromatic residues can result in irreversible damages such as formation of di-tyrosine, cleavage of the Phe side-chain off the backbone, and opening of the 5-membrer rings on Trp and His (Davies and Truscott 2001). For Trp, irreversible oxidative damage can generate N-formylkynurenine through opening of the indole ring (Creighton 1993).

Aromatic amino acids are synthesized through the shikimate pathway in plants and microorganisms (Brown *et al.* 2003). Human lacks the shikimate pathway, making all aromatic amino acids except for tyrosine essential amino acids for human. Synthesis of Tyr requires Phe, therefore, Tyr is considered as a conditional essential depending on the availability of Phe (Berg *et al.* 2007).

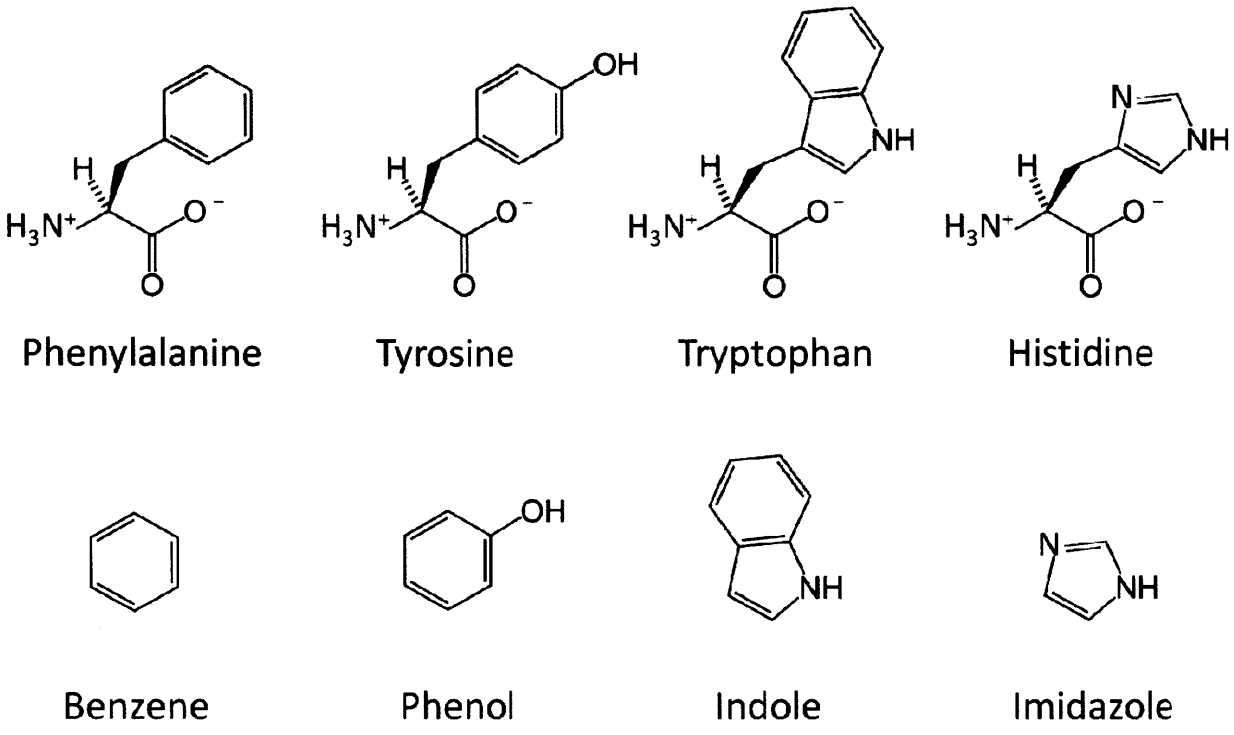


Figure 1-3 Aromatic Amino Acids and Their Corresponding Aromatic Compounds. These molecules are shown as at neutral pH.

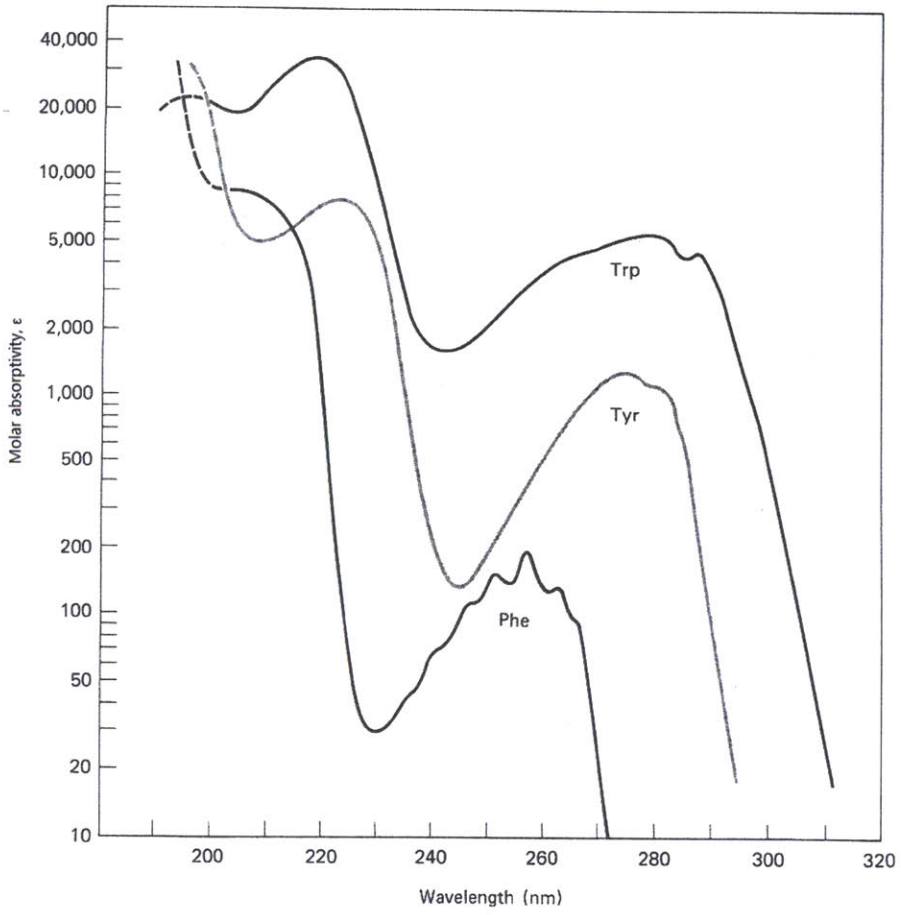


Figure 1-4 UV Absorbance Spectra of Aromatic Amino Acids at pH 6.
(Reprinted from *Advances in Protein Chemistry*, 17: 303-390, Wetlaufer, Ultraviolet Spectra of Proteins and Amino Acids, Copyright (1963), with permission from Elsevier. The image is from Creighton 1993).

3. Aromatic Interactions in Small Compounds

Aromatic molecules interact through van der Waals interaction and electrostatic interaction of the partial charges on the δ^- ring faces and δ^+ ring edges. Benzene dimer is the prototype for aromatic-aromatic interaction and has been extensively studied theoretically and experimentally. Potential energy surface calculations revealed two low-energy states of nearly identical stabilities: parallel displaced and tilted T-shaped orientations. The energy barrier between these two orientations was low, indicating ready conversion between them (Hobza *et al.* 1996; Podeszwa *et al.* 2006) (Figure 1-5). The interaction energy of benzene dimer was estimated to be about 3 kcal/mol. The stability was largely due to van der Waals interaction, while the orientation was dictated by electrostatic interaction (Hobza *et al.* 1996; Podeszwa *et al.* 2006). Both orientations were indeed found in benzene crystal (Cox and Smith 1958).

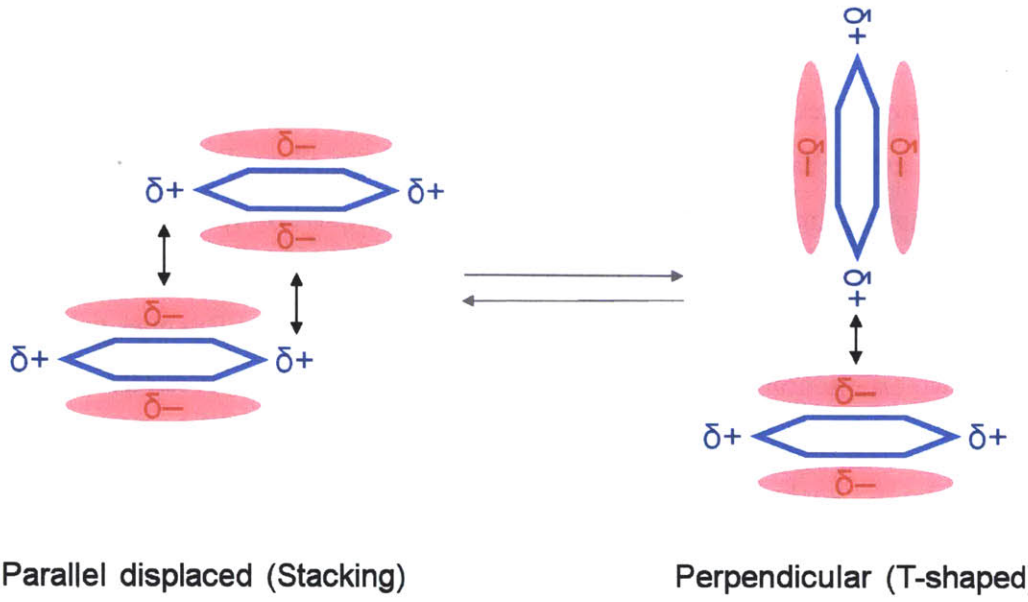


Figure 1-5 Favorable Orientations of Aromatic-aromatic Interaction.

Two low-energy states, parallel displaced (stacking) and perpendicular (T-shaped) orientations have nearly the same energy and can be readily converted from one to the other.

4. Aromatic Interactions in Proteins

A seminal statistical study on aromatic interactions in proteins by Burley and Petsko indicated that aromatic residues tended to cluster into pairs or networks of three or more (Burley and Petsko 1985). The ring centroids preferred a distance of 4.5-7 Å. Preference of nearly perpendicular or parallel-displaced orientations were both observed (Burley and Petsko 1985; McGaughey *et al.* 1998). It was estimated that buried aromatic-aromatic interaction contributed ~1 kcal/mol to protein stability (Burley and Petsko 1985). Positively charged groups or δ^+ side-chain amino groups also preferentially interact with δ^- aromatic ring faces (Burley and Petsko 1986; Ma and Dougherty 1997). These observations indicated that the electrostatic nature of aromatic interaction is conserved from free benzene dimers to amino acid side chains in proteins. The complexity of the tryptophan indole side chain appeared to generate more versatile interactions, including CH-π, S-aromatic, OH/NH-π, and O-edge interactions, in addition to those among aromatic rings (Samanta *et al.* 2000). The interaction between a δ^+ hydrogen with a δ^- aromatic ring face is regarded as similar to a hydrogen bond, with the aromatic rings as hydrogen bond acceptors. The strength of such interaction is about half of a conventional hydrogen bond (Levitt and Perutz 1988). Aromatic-aromatic interactions tend to stabilize elements far away in the primary structure, most often two β-strands (Thomas *et al.* 2002; Thomas *et al.* 2002), although local interactions were also reported (Bhattacharyya *et al.* 2002).

Comparison between thermophilic and mesophilic protein homologues revealed additional or enlarged aromatic networks in the thermophilic proteins, suggesting their importance in protein thermal stability (Kannan and Vishveshwara 2000). Mutagenesis studies have also indicated that aromatic interactions played important roles in stability and folding of diverse proteins. For barnase, aromatic-aromatic interaction of the Tyr13/Tyr17 pair contributed ~1.3 kcal/mol to protein stability. Tyr/Tyr and Phe/Phe pairs made identical stability contributions, indicating that the aromatic rings, instead of the -OH groups, contributed predominantly to their interaction energy (Serrano *et al.* 1991). Gln33Tyr substitution on the N-terminal domain of phage λ repressor created an aromatic cluster from Tyr22, Tyr33, and Phe51. This mutant protein had enhanced thermal stability compared to the WT protein, presumably due to this aromatic cluster (Hecht *et al.* 1984). Aromatic residues were introduced to the N-terminal β-hairpin of ubiquitin, a folding nucleation site in the protein. These substitutions resulted in a

kinetically trapped, stable folding intermediate stabilized by non-native interactions (Rea *et al.* 2008).

In some cases, aromatic residues have important roles in both the folding pathway and stability. *E. coli* cold shock protein CspA has a six-aromatic patch including Phe18, 20, and 31 on the surface that was shown to be responsible for both protein stability and nucleic acid binding (Hillier *et al.* 1998). The same cluster was also important for the folding kinetics, probably by nucleating folding (Rodriguez *et al.* 2000). The core of bovine pancreatic trypsin inhibitor (BPTI) is largely composed of eight Tyr or Phe residues. Substitutions of these residues decreased the stability of the native state by 2-7 kcal/mol (Zhang and Goldenberg 1997). In addition, these substitutions altered the distribution of disulfide bonded folding intermediates, indicating the deferential contributions of these aromatic residues in stabilizing the conformational different early folding intermediates (Zhang and Goldenberg 1997).

Aromatic interactions have also been studied in peptides, and employed in *de novo* peptide design (Waters 2004). As in proteins, cross-stand aromatic pairs and clusters could stabilize β-hairpins peptides (Griffiths-Jones and Searle 2000; Tatko and Waters 2002; Wu *et al.* 2010). In these studies, aromatic-aromatic interactions seemed to provide greater stabilizing effects than comparable aromatic-aliphatic interactions, reminiscent of the aromatic clustering phenomenon observed in proteins (Tatko and Waters 2002; Wu *et al.* 2010). In "Trpzip" peptides, differential energy contributions existed among different aromatic-aromatic pairs (Wu *et al.* 2010). φ-Value analysis on a similar peptide showed that the aromatic pairs were important for the unfolding kinetics while the β-turn region was important for the folding kinetics, suggesting a zipper-like folding/unfolding pathway (Du *et al.* 2006). The distance between the aromatic pairs and the β-turn was probably a key in reaching this conclusion.

Aromatic residues in proteins are also important in intermolecular interactions with drugs, ligands or substrates. Phe10 in P22 Arc repressor was shown to be important for DNA binding specificity (Schildbach *et al.* 1999). An unusual solvent-exposed hydrophobic patch with six highly conserved aromatic residues is located in the ankyrin groove of human TRPV2 channel ankyrin repeat domain (ARD). This aromatic cluster may be important in ligand or protein recognition (McCleverty *et al.* 2006). An aromatic-hydrophobic motif in ClpX interacts with substrates, and transmits the ATP-dependent conformational changes in the pore loops to substrate translocation (Martin *et al.* 2008).

C. Modes of Protein Aggregation

Protein aggregates, in a broader definition, collectively describe insoluble, high-molecular-weight protein assemblies. They fall into several different modes, each with distinct conformation and amino acid sequence requirements. For γB-crystallin, aggregates induced by four agents—heating, cooling, UV exposure and refolding—showed distinct morphological features and secondary structural contents (Fatima *et al.* 2010). This example highlights the importance of investigating different modes of aggregations, even when working with the same protein.

1. Native-like Polymerization

Protein aggregation often requires unfolding, but not always. Protein aggregation can occur with conformations very similar to the native states of the monomers (often referred to as polymerization instead of aggregation). This type of aggregation or polymerization occurs in both pathological and normal conditions.

The best-known example of native protein polymerization resulting in pathology is probably the aggregation of sickle hemoglobin. Sickle hemoglobin arises from genetic mutation of the gene encoding the β chain of hemoglobin that leads to a single amino acid substitution, E6V on the surface of the protein. The amino acid substitution does not affect the native structure or the assembly of hemoglobin tetramer. However, under deoxygenated condition, this substitution creates a hydrophobic protrusion that allows the intermolecular interaction among sickle hemoglobin molecules, resulting in a long, rigid polymer (Figure 1-6) (Harrington *et al.* 1997). The same contact site may be used in both the axial and lateral growth of the polymer. The final rigid fiber may be composed of 14 primary strands (Rotter *et al.* 2005).

Eukaryotic cytoskeleton also assembles through native-like polymerization. Actin filament is a major class of cytoskeleton and a major component in the muscle sarcomere. Monomeric G-actin molecules associate to form the two-stranded twisted polymeric F-actin. The polymerization reaction is regulated by ATP hydrolysis. ATP or ADP binding causes subtle conformational changes that favor the polymeric vs. monomeric states (Kabsch and Vandekerckhove 1992). Microtubules are hollow tubular fibers composed of alternating α- and

β-tubulin molecules. Microtubule assembly mirrors that of actin filaments, but with the nucleotide regulator being GTP/GDP instead of ATP/ADP (Berg *et al.* 2007). In both cases, no unfolding or large structural rearrangement is needed for the polymerization.

2. Amyloid Fibril

Amyloid represents a class of protein aggregates that underlies many human diseases (Sipe and Cohen 2000; Stefani 2004). Under the electron microscope, amyloid aggregates appear as long, twisted, unbranched fibrils. Experimentally, amyloid fibrils lead to increase in fluorescence when stained with thioflavin T (ThT) (LeVine 1999), and apple green birefringence when stained with Congo Red (Klunk *et al.* 1999). The molecular structure of amyloid fibril is equivalent to parallel β-sheets, with the hydrogen bonds parallel to the fibril axis (Figure 1-6) (Wille *et al.* 2002; Makin and Serpell 2005), also known as the "cross-β" structure. This structure can be identified by X-ray diffraction. Goldschmidt *et al.* estimated that proteins or peptides with native β-sheet structures were somewhat more prone to amyloid formation, probably due to the propensity for an extended conformation, but the preference was not strong (Goldschmidt *et al.* 2010). Amino acid sequences natively in α-helices and coil regions are not excluded for forming amyloid fibrils, but the formation requires substantial conformational changes.

The interior packing of amyloid fibrils may be dehydrated or hydrated (mediated through water molecules) (Balbirnie *et al.* 2001; Kishimoto *et al.* 2004), but in either case requires certain amino acid sequences in order to achieve complementary packing of the side chains, although the sequence requirement may be less stringent than for the hydrophobic cores of native proteins. The side chains pointing to the outside of the primary fibril may be important in the lateral growth of the fibrils into large, rigid fibers (Nelson *et al.* 2005). Using computation modeling based on the crystal structure of a fragment of the yeast prion Sup35 NNQQNY, Goldschmidt *et al.* suggested that the population of amylome (all amino acid sequences able to form amyloid fibrils) was quite large, and almost all proteins contained at least one potentially amyloidogenic segment. Therefore, the sequence and structure contexts of the amyloidogenic segments are important in shielding them from amyloid formation (Goldschmidt *et al.* 2010).

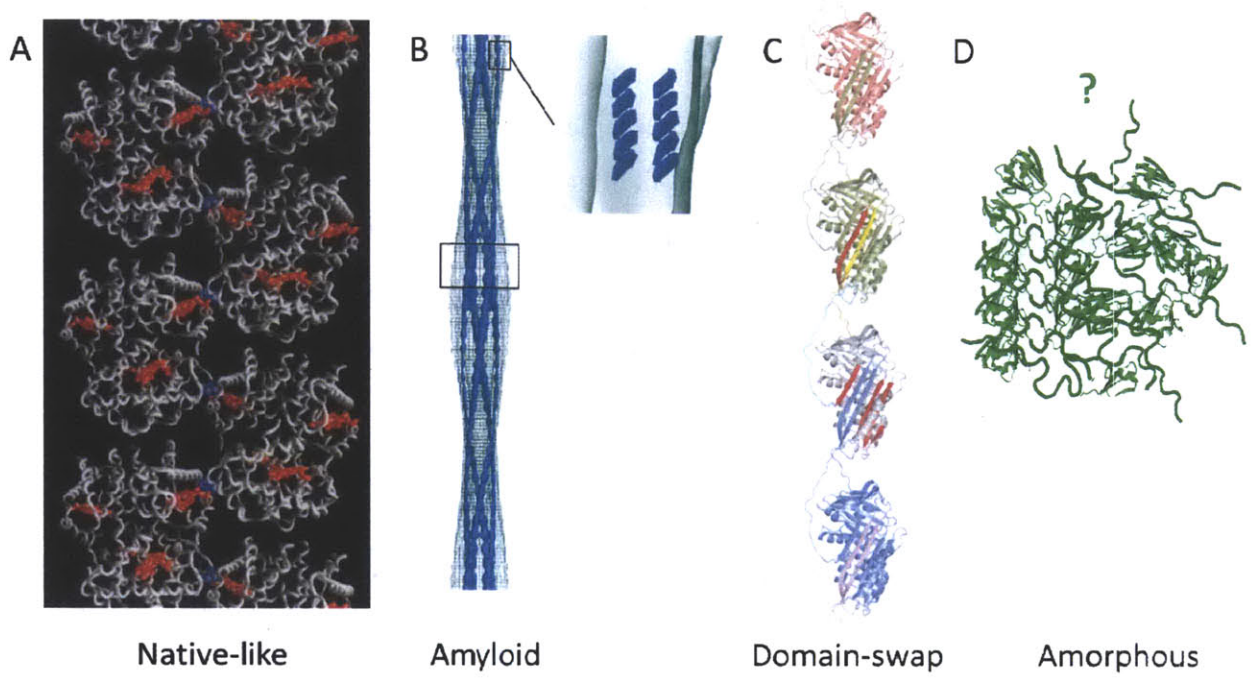


Figure 1-6 Modes of Aggregations.

(A) Crystal structure of sickle-cell hemoglobin polymer (Reprinted from *J Mol Biol.*, 272(3): 398-407, Harrington *et al.*, The high resolution crystal structure of deoxyhemoglobin S. Copyright (1997), with permission from Elsevier). (B) Cryo-EM model of amyloid fibril from the SH3 domain of phosphatidylinositol-3'-kinase. (Reprinted by permission from Macmillan Publishers Ltd: *The EMBO Journal.*, Jimenez *et al.*, copyright (1999)). (C) Structure of open-ended domain-swap polymer of serpin, modeled from the crystal structure of a domain-swap dimer. (Reprinted by permission from Macmillan Publishers Ltd: *Nature*, Yamasaki *et al.*, copyright (2008)). (D) Amorphous aggregation, hypothetical for human γD-crystallin.

3. Domain Swapping

As the name implies, in domain-swap interaction, the interacting proteins exchange their domains or motifs among each other. Native-like interactions between domains or elements in the monomeric state are maintained at the intermolecular interfaces. This requires a selective partial unfolding of a protein—opening of domain/element interfaces while maintaining the domains in native state. The connecting regions of domains or motifs of the protein must be flexible enough to accommodate both the monomeric or domain-swap conformations. The products of domain-swap interactions could be a dimer or oligomers, as in closed domain swapping, or a long polymer, as in open-ended domain swapping (Liu and Eisenberg 2002). Domain-swap dimerization followed by shortening of the connecting peptides and/or formation of secondary intramolecular contacts may be a mechanism for the evolution of protein homodimers (Bennett *et al.* 1995). These oligomers have relatively extensive and symmetrical dimer interfaces, which are remnants of the ancestral intramolecular domain interfaces. In the case where the connecting peptide no longer allows the ancestral monomeric conformation, this type of interaction is called pseudo-domain-swapping.

Upon the pH change accompanying freezing in phosphate buffer, diphtheria toxin formed a classic domain-swap dimer with the receptor binding domains exchanged. Higher oligomers were also observed, probably through similar swapping of the receptor binding domains, but in a linear or cyclic fashion (Bennett *et al.* 1994). In another example, serine protease inhibitor (serpin) formed domain-swap dimer through insertion of a long β-hairpin into the center of a β-sheet of the neighboring molecule. Because of the flexibility of the connecting peptide between the hairpin and the rest of the protein, an open-ended domain-swap polymer model was also proposed, and may be the underlying mechanism of serpin polymerization in pathological conditions (Figure 1-6) (Yamasaki *et al.* 2008).

4. Amorphous Aggregation

As the name implies, in amorphous aggregates, the macromolecular morphology, as well as conformations of the polypeptide chains in the aggregated state are not well defined (Figure 1-6). Nonetheless the intermolecular interaction may be quite specific. The amorphous

macromolecular structure may be a result of the flexibility of certain regions of the aggregated proteins that do not allow the propagation of the rigid structure, or a heterogeneous mix of different forms of aggregates. Most inclusion bodies resulting from the failure of protein folding within bacterial hosts represent amorphous aggregates.

D. Molecular Chaperones

Molecular chaperones are proteins that assist other proteins to achieve their native structures or appropriate oligomeric assemblies, to prevent proteins from aggregating, to refold misfolded proteins, and/or disassemble aggregates. Although chaperones do not determine the native structures of proteins, they are important factors that influence the kinetics of protein folding, and the partitioning between productive folding and aggregation, acting essentially as "folding catalysts". Although many small proteins can fold efficiently *in vitro* without chaperones, the crowded and heterogeneous nature of cellular environment favors protein aggregation and aberrant protein interaction, and thus may hinder the folding of even proteins with simple folds (van den Berg *et al.* 1999). Fortunately, cells have a cohort of chaperones to cope with the protein misfolding and aggregation problem (Hartl and Hayer-Hartl 2009). Chaperones are classified by their functional mechanisms and the stages of the protein biosynthesis pathway they work on (Chang *et al.* 2007; Tang *et al.* 2007). In this section, I will briefly introduce selected common classes of chaperones, then focus on small heat shock proteins.

1. Hsp70/40

The Hsp70/40 system (DnaK/DnaJ in *E. coli*) associates with nascent polypeptide chains coming out of the ribosome in an unfolded or partially unfolded state. Hsp70/40 shields the hydrophobic regions from self-aggregation and aberrant interactions with other proteins. The structure of Hsp70 is organized into the N-terminal ATPase domain and C-terminal peptide-binding domain (Zhu *et al.* 1996; Harrison *et al.* 1997). The crystal structure of the complex of the peptide binding domain of DnaK with a substrate peptide showed that the peptide binding domain consists of an N-terminal β-sandwich sub-domain and several C-terminal α-helices that can act as a "lid" (Zhu *et al.* 1996). The peptide binds to a groove formed by the loops of the β-sandwich sub-domain in an extended conformation, running perpendicular to the β-strands. The substrate binding is achieved by a combination of main-chain hydrogen bonds and side-chain hydrophobic interaction. A hydrophobic pocket in the center of the groove could accommodate a variety of hydrophobic side chains.

Substrate binding and releasing of Hsp70 is regulated by ATP hydrolysis (Mayer *et al.* 2000). The peptide-binding domain is in open state when Hsp70 is bound with ATP. Upon substrate binding in the groove, ATP hydrolysis in the ATPase domain allosterically causes the closure of the lid in the peptide-binding domain, generating a stable chaperone-substrate complex (Hartl and Hayer-Hartl 2009). Hsp40 functions as a co-chaperone and accelerates the ATPase activity of Hsp70. Another group of co-chaperones, such as GrpE in *E. Coli* and Hsp110 in eukaryotes, serve as nucleotide exchange factors (NEF) (Polier *et al.* 2008). ADP-ATP exchange causes lid opening and substrate release. The substrate binding and release cycle maintains the substrates in a folding-competent state and prevents their aggregation. The substrates, after being completely synthesized from the ribosome, may be transferred to more sophisticated chaperones, such as chaperonins and Hsp90.

2. Hsp60/Chaperonin

Hsp60 are a class of ATP-dependent chaperonins with a signature double-ring structure. Each ring consists of 7-9 identical or homologous subunits and forms a barrel with a central cavity (Horwich *et al.* 2007). Group I chaperonins are present in prokaryotes, mitochondria and chloroplasts, with GroEL as the most studied member. Each GroEL ring consists of 7 identical subunits. The co-chaperone GroES also consists of 7 identical subunits and serves as the lid for the chambers formed by GroEL. Group II chaperonins include the archaeal thermosome and the eukaryotic TCP-1 Ring Complex (TRiC or CCT). Compared to Group I chaperonins, Group II chaperonins have more heterogeneous subunit compositions while maintaining a highly symmetrical overall structure. For TRiC, each ring consists of eight different subunits. Group II chaperonins use "built-in" protrusions from the subunits to enclose the cavity instead of using a detachable lid (Horwich *et al.* 2007).

The subunits of chaperonins (the large subunits for Group I) are organized into an equatorial ATPase domain, an intermediate domain and an apical domain (Xu *et al.* 1997). The apical domain is largely responsible for substrate recognition. GroEL preferentially recognizes non-native states of proteins, presumably through hydrophobic interaction between the substrates and the apical domain on the chaperonins (Viitanen *et al.* 1992). GroEL interacts with at least ~250 cytosolic substrates (Hartl and Hayer-Hartl 2009). TRiC seems to have a more limited

substrate range and the eight different subunits may confer some substrate specificity. Tubulin and actin were thought to be the only substrates for TRiC but it was later found that ~10% of newly synthesized proteins interact with TRiC (Horwich *et al.* 2007). Direct TRiC binding site on VHL tumor suppressor protein was mapped to several hydrophobic residues on two spatially adjacent β-strands (Feldman *et al.* 2003). TRiC was also shown to preferentially bind to a specific β-strand on Gβ protein (Kubota *et al.* 2006).

ATP hydrolysis at the equatorial ATPase domain cycles the chaperonins through a series of events including substrate binding, enclosure, and release, alternating between the two rings. This general mechanism seems to be conserved between the two groups of chaperonins. The motion of the subunits of one ring is highly cooperative for GroEL, but there is some evidence for sequential ATP binding and substrate release for the different subunits of one TRiC ring (Horwich *et al.* 2007). How substrates fold inside the chaperonin chamber remains a mystery. Passive separation of the substrates from the cytosol, as well as active molding by the interactions between the interior of the chamber and the substrates, were both proposed as the functional mechanism for chaperonins. Repeated ATP hydrolysis cycles may be needed for a substrate to achieve the completely native state (Horwich *et al.* 2007; Hartl and Hayer-Hartl 2009).

3. Hsp90

The Hsp90 class of ATP-dependent chaperones form homodimers (Wandinger *et al.* 2008). Each subunit consists of an N-terminal ATP binding domain (N-domain), an M-domain, and a C-terminal dimerization domain (C-domain) (Ali *et al.* 2006). Studies on yeast Hsp90 showed that binding of ATP to the N-domain induces conformational change of the N-domain that activates the ATPase site. In addition, the two N-domains dimerize transiently through exchange of an N-terminal segments, resulting in a closed form of the Hsp90 dimer (Richter *et al.* 2006). ATP hydrolysis leads to the dissociation of the N-domains and returns the dimer to the open state. Hsp90 interacts with a large number of co-chaperones. Binding of co-chaperone Sba1 to the N-domain decreases ATP hydrolysis rate and traps the dimer in the closed form (Richter *et al.* 2004; Ali *et al.* 2006). In summary, ATP hydrolysis powers a "scissors-like" motion of Hsp90 that presumably assists the conformational remodeling of substrates.

Compared to other chaperones, Hsp90 acts on a much more selected range of substrates, most of which are steroid hormone receptors/transcription factors and cell-cycle kinases (Buchner 1999). Studies on the cellular and viral forms of Src kinase suggested that stability played an important role in determining the binding of Hsp90 to substrates (Falsone *et al.* 2004). The substrate binding site on Hsp90 is not clearly understood. All three domains of the protein have been proposed to contain the binding sites. Different binding modes, such as binding in between the subunits, as well as asymmetrical binding on individual subunits have been postulated (Wandinger *et al.* 2008).

4. Hsp100

Hsp100 (the Clp family in *E. coli*) are a family of ATP-dependent chaperones, belonging to the AAA+ superfamily. Class 1 members, such as ClpA, ClpB and ClpC, contain two AAA+ modules, while Class 2 members, such as ClpX and ClpY, contain only one ATP binding domain (Lee *et al.* 2004). Hsp100/Clp chaperones assemble into hexameric ring structures, and use energy from ATP hydrolysis to power the translocation of polypeptides through the narrow central opening in the ring (Lee *et al.* 2003). Unfolding of proteins results from the mechanical threading process. Cooperation of ClpB and DnaK is a major force in resolubilizing protein aggregates in the cell (Lee *et al.* 2004; Weibezahn *et al.* 2005). While the threading process of ClpB is necessary for untangling the polypeptides, DnaK is involved in both the early stage of interactions with the aggregated proteins and the handling of the translocated polypeptides from ClpB, reminiscent of its interactions with newly synthesized polypeptide chains (Weibezahn *et al.* 2005). Certain members of Hsp100, such as ClpA, ClpC and ClpX contain a tripeptide element that allows their interactions with barrel-like protease ClpP, forming the AAA+ protease complexes (Kim *et al.* 2001). Coupling protein unfolding and degradation, AAA+ proteases are powerful, but energetically costly protein destruction machines (Baker and Sauer 2006). Recognition of substrates by AAA+ proteases can be mediated by peptide tags or adaptors. Exposure of peptide tags can be regulated by protein cleavage, unfolding, or subunit dissociation, while adaptor-mediated recognition is achieved by increasing the effective local concentration of substrates (Baker and Sauer 2006).

5. sHsp

This section focuses on general information about small heat shock proteins (sHsp). A separate introduction focused on α-crystallins will be provided in Section F.2 in this introduction.

Small heat shock proteins are a group of chaperones with molecular weights of 12-43 kDa, with the majority being 14-27 kDa (Narberhaus 2002). sHsps are widely distributed in most, but not all organisms throughout all three domains of life, indicating that they are generally important, but not universally required in protein quality control (Caspers *et al.* 1995). Consistent with this notion, sHsps seem to be somewhat dispensable in prokaryotes unless combined with additional deletions of other chaperones and/or stress conditions (Narberhaus 2002). For example, deletion of sHsp Ibp in *E. coli* did not show any phenotype at low temperature, but had a detrimental effect on growth when combined with heat stress and deletion of DnaK (Thomas and Baneyx 1998). The number of sHsp genes varies greatly from organism to organism, from none in some bacteria to numerous in plants. The human genome encodes 10 sHsps dispersed over 9 chromosomes, most of which are abundant in various muscles (Kappe *et al.* 2003). Most sHsps are inducible by heat and various stresses. For example, transcription of *E. coli* sHsps IbpA and IbpB increased by ~300 fold at elevated temperature (Richmond *et al.* 1999).

Sequence similarity among sHsps is generally low. The defining element of this group of chaperones is a conserved, primarily β-sheet α-crystallin domain (αCD) of about 90 amino acids. αCD is organized into a β-sandwich with β-strands 2-9 (β2-9), similar to the immunoglobulin (Ig) fold, and also reminiscent of the crystallin domain fold in βγ-crystallins (Haslbeck *et al.* 2005). Flanking the αCD are variable length N- and short C-terminal extensions with limited to no similarity (de Jong *et al.* 1993; Caspers *et al.* 1995). The only conserved amino acids on the extensions are a Phe in the N-terminal extension and an IXI/V motif in the C-terminal extension (Narberhaus 2002).

Formation of large oligomeric complex is a hallmark of sHsps, although the number of subunits varies greatly from different sHsps. For sHsps from lower organisms, the complexes usually have defined numbers of subunits and are highly symmetrical (Figure 1-7). Hsp16.5 from archaeon *Methanococcus jannaschii* and Hsp16.9 from wheat *T. aestivum* are among the first sHsps with available crystal structures (Kim *et al.* 1998; van Montfort *et al.* 2001). These sHsps

formed hollow spherical 24-mers and double-disc 12-mers, respectively. For plant pathogen XaHspA from *Xanthomonas*, four types of 36-mers were proposed, but only one configuration was consistent with the substrate binding ratio (Hilario *et al.* 2011). Tsp36 from the parasitic flatworm *Taenia saginata* contains two αCDs in each chain, and was shown to form dimers or tetramers, a much simpler organization compared to typical sHsps (Stamler *et al.* 2005). Cryo-EM reconstruction suggested sHsp Acr1 from *Mycobacterium tuberculosis* forms a dodecameric structure (Kennaway *et al.* 2005). Yeast Hsp26 appeared to form quite uniform 24-mers (Haslbeck *et al.* 1999). Human α-crystallin formed polydisperse oligomers with 15-60 subunits (Horwitz 1992; Haley *et al.* 1998; Haley *et al.* 2000). From the available structures of sHsps, the following common themes can be found. First, the basic building block appears to a dimer, however, the dimeric interfaces are different between metazoan and lower organisms (Clark *et al.* 2011). Second, the oligomeric assemblies are diverse in geometry, but typical oligomers seem to have a spherical shape, with hollow interior and large solvent-accessible area. The αCDs are responsible for dimer formation while both flanking extensions are responsible for higher oligomer formation (Haslbeck *et al.* 2005).

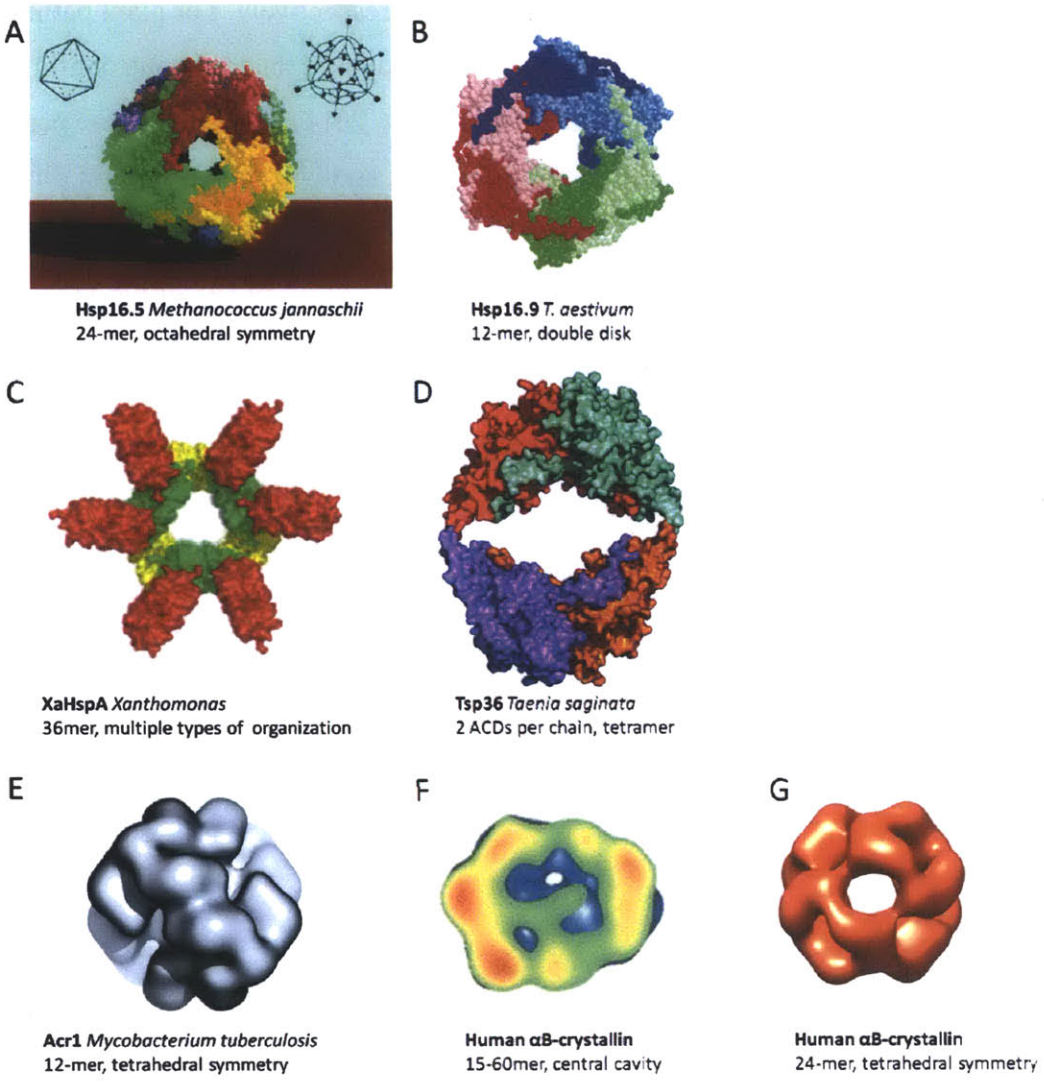


Figure 1-7 Structures of Representative Small Heat Shock Proteins.

The sizes are not drawn to scale. (A-D) Crystal structures. (E-G) Cryo-EM structures. (A) (Reprinted by permission from Macmillan Publishers Ltd: *Nature*, Kim *et al.*, copyright (1998)). (B) (Reprinted by permission from Macmillan Publishers Ltd: *Nat Struct Biol.*, van Montfort *et al.*, copyright (2001)). (C) (Reprinted from *J Mol Biol.*, 408(1): 74-86, Hilario *et al.*, Crystal structures of Xanthomonas small heat shock protein provide a structural basis for an active molecular chaperone oligomer. Copyright (2011), with permission from Elsevier). (D) (Reprinted from *J Mol Biol.*, 353(1): 68-79, Stamler *et al.*, Wrapping the alpha-crystallin domain fold in a chaperone assembly. Copyright (2005), with permission from Elsevier). (E) (Reprinted with permission from Kennaway *et al.* 2005). (F) (Reprinted from *J Mol Biol.*, 298(2): 261-272, Haley *et al.*, Small heat-shock protein structures reveal a continuum from symmetric to variable assemblies. Copyright (2000), with permission from Elsevier). (G) (Reprinted with permission from Peschek *et al.* 2009).

It remains controversial whether the oligomeric assembly of sHsps is important for their chaperone function. Supporting this hypothesis, a group of naturally occurring sHsps, Hsp12.2, Hsp12.3 and Hsp12.6 from *C. elegans* possess very short or no N- or C-terminal extensions. They formed monomers and tetramers instead of large oligomers. Interestingly, they were devoid of *in vitro* chaperone activity (Leroux *et al.* 1997; Kokke *et al.* 1998). N-terminal truncated (but not C-terminal truncated) *C. elegans* Hsp16-2 also lost the large oligomeric structure and chaperone activity simultaneously (Leroux *et al.* 1997). Cataract associated R120G αB-crystallin resulted in larger and irregular complexes with the absence of a normal central cavity. This mutant αB-crystallin had reduced to completely absent chaperone function against aggregations of multiple substrates (Bova *et al.* 1999). Together these studies indicated that a proper oligomeric assembly could be a prerequisite for the chaperone function of sHsps.

On the other hand, a body of conflicting results was also obtained. Isolated αCD from αB-crystallin without the extensions formed dimers instead of large oligomers, but retained significant chaperone activity (Feil *et al.* 2001). Similarly, truncation of 5 C-terminal residues of αA-crystallin decreased the oligomeric molecular weight and exchange rate with αB-crystallin, but had no effect on the chaperone efficiency (Aquilina *et al.* 2005). F71G substitution in the αCD completely abolished the chaperone activity of αA-crystallin without affecting the oligomer size (Santhoshkumar and Sharma 2001). Substitutions of the polar residues on the C-terminal extension of murine Hsp25 substantially reduced its chaperone activity without affecting the overall secondary structure and oligomeric size (Morris *et al.* 2008). From these studies, it appears that oligomeric assembly is neither necessary nor sufficient for the chaperone functions of human sHsps. A clear relationship between oligomeric assembly and chaperone function remains to be established, and it may operate in an sHsp-specific manner.

Instead of the oligomeric assembly *per se* being important for chaperone function, it was also proposed that the oligomeric assembly is merely the storage form of sHsps. Upon induction by stress or heat, the dimer units dissociate from the large oligomers and act as the active chaperone (Nakamoto and Vigh 2007). The main evidence came from the observation that many sHsps oligomers underwent reversible temperature-dependent dissociation, including Hsp27-αB-crystallin hybrid oligomer (Zantema *et al.* 1992), yeast Hsp26 (Haslbeck *et al.* 1999) and wheat Hsp16.9 (Van Montfort *et al.* 2003). Furthermore, in the case of Hsp26, dissociation of complex at heat shock temperatures substantially enhanced the chaperone activity. For wheat Hsp16.9,

temperature-dependent high-molecular-weight complex formation with a model substrate was observed. While the evidence seems strong for the heat-induced dissociation mechanism of chaperone activation, counter-evidence was also reported. Disulfide cross-linked yeast Hsp26 (Franzmann *et al.* 2005) and glutaraldehyde cross-linked bovine α-crystallin (Augusteyn 2004) both lost their ability to dissociate into smaller species. However, chaperone activities remained unchanged for Hsp26, and surprisingly, even became more effective for α-crystallin. While the relationship between oligomer dissociation and chaperone function remains unsolved, additional studies on α-crystallin suggested temperature-dependent tertiary structural transition into a molten-globule-like state, presumably accompanied by exposure of hydrophobic surface, rather than subunit dissociation, as the underlying mechanism of activation, providing a possibility for functional activation while maintaining the oligomeric structure (Raman and Rao 1997).

Similar to Hsp70/40 and chaperonins, sHsps generally have low degree of substrate specificity. Murine Hsp25, human Hsp27, and α-crystallins were shown to suppress the aggregation of a variety of enzymes and lens proteins (Horwitz 1992; Jakob *et al.* 1993). Two yeast sHsps, Hsp26 and Hsp42 suppressed the aggregation of about 30-35% of cytosolic proteins *in vivo*. In addition, the two sHsps showed overlapping but not identical substrate selections, indicating some degree of selectivity (Haslbeck *et al.* 2004). As examined by immunoprecipitation, Cyanobacterial sHsp Hsp16.6 from *Synechocystis* interacted with 42 proteins involved in diverse cellular functions (Basha *et al.* 2004).

It is well established that sHsps function as chaperones by sequestering aggregation-prone substrates, and preventing their off-pathway aggregation in an ATP-independent manner. Reports on positive or negative ATP effects were also found (Muchowski and Clark 1998; Smykal *et al.* 2000). sHsps can form complexes with substrates, but cannot refold substrates *per se*, as exemplified by α-crystallin-bound βB2-crystallin when returned to native condition (Evans *et al.* 2008). However, sHsps can maintain the misfolded substrates in a folding-competent state and pass them to ATP-dependent chaperones for active refolding. Such transfers of substrates have been reported to occur from IbpB to DnaK/DnaJ/GrpE and GroEL/S in *E. coli* (Veinger *et al.* 1998), and from murine Hsp25 to Hsp70 (Ehrnsperger *et al.* 1997). Reactivation of α-crystallin-bound luciferase and citrate synthase was shown to be more efficient by reticulocyte lysate, compared to a system with Hsp70/40 and Hsp60, indicating additional help from other chaperone systems or cellular components (Wang and Spector 2000).

E. Human Eye Lens and Cataract

1. Cellular and Molecular Properties of the Human Eye Lens

The vertebrate eye has one of the most complex and elegant structures found in life. Even Charles Darwin found it difficult to believe that such an organ of "extreme perfection and complication" was a product of natural selection (Darwin 1859). In humans, the eye lens is situated inside a collagen-rich lens capsule, located in the anterior center of the eye, behind the cornea and iris, and in front of the vitreous humor. The lens is central to eye function—it contributes about one third of the total refractive power and is the only variable refractive element in the eye (Oyster 1999). The lens develops from the lens vesicle, a hollow shell of epithelial cells derived from the ectoderm (Figure 1-8) (Gilbert 2000). In the mature lens, the nucleus, the center of the lens, consists of the oldest, primary fiber cells derived from the posterior epithelial cells of the lens vesicle. The anterior layer of epithelial cells serves as "stem cells" for continuing growth of the lens. As lens develops, these anterior epithelial cells divide and migrate to the equatorial region, then differentiate into concentric layers of long, curved secondary fiber cells, forming the cortex of the lens (Figure 1-8) (Sharma and Santhoshkumar 2009). Slightly different definitions of lens nucleus and cortex also exist (Fagerholm *et al.* 1981; Oyster 1999). The ends of these secondary fiber cells join at the anterior and posterior faces of the lens, forming a radiating pattern called the "lens suture". The fiber cells make up the majority of the lens mass. Lens growth never stops throughout the human life, although the rate is 100-fold faster in prenatal and infant period than for the rest of the life (Bron *et al.* 2000).

The function of the lens requires two key features: transparency and refractive power. These in turn, depend on the unique properties of the lens on both cellular and molecular levels. The fiber cells are packed tightly in a highly ordered hexagonal scheme. Adjacent cells are held together by interdigitated "knob and hole" structures (Oyster 1999). An important means to reduce light scattering is the degradation of membrane-based organelles of the fiber cells, including mitochondria, ER, and nucleus (Bassnett 2002). Organelle degradation occurs concurrently with fiber cell differentiation and elongation. The degradation process largely resembles apoptosis, but retains key differences (Bassnett 2002). Unlike apoptotic cells, fiber cells maintain the integrity of actin cytoskeleton and cell membrane, which are important for the

morphology of the fiber cells (Bassnett and Beebe 1992). Degradation of organelles leads to the halt of protein synthesis. Therefore, long-term stability is a critical requirement for the proteins that reside in the lens fiber cells.

To obtain a high refractive index, the fiber cells have to maintain a high concentration of solutes. The dry mass content (about 95% proteins and lipoproteins, 5% lipids and electrolytes) increases from about 0.2 g/mL in the cortex to about 0.45 g/mL in the nucleus (Fagerholm *et al.* 1981). About 90% of the soluble proteins are crystallins. Three classes of crystallins exist and will be subjects of Section F of this introduction. Expression of some crystallins begins in epithelial cells and dramatically increases in a highly regulated manner upon differentiation (Wang *et al.* 2004). Short-range order among the crystallins prevents macroscopic density fluctuation, and furthermore, may prevent the protein molecules from acting as independent light scatterers (Delaye and Tardieu 1983). Both of these effects enhance the transparency of the lens. The molecular nature of short-range order in the lens can be explained by both the self-associations of α- and β-crystallins into oligomers, and Coulombic repulsive/attractive interactions among α-, β- and γ-crystallins (Ponce *et al.* 2006). In addition, the heterogeneity of crystallins is believed to hinder protein crystallization inside the lens.

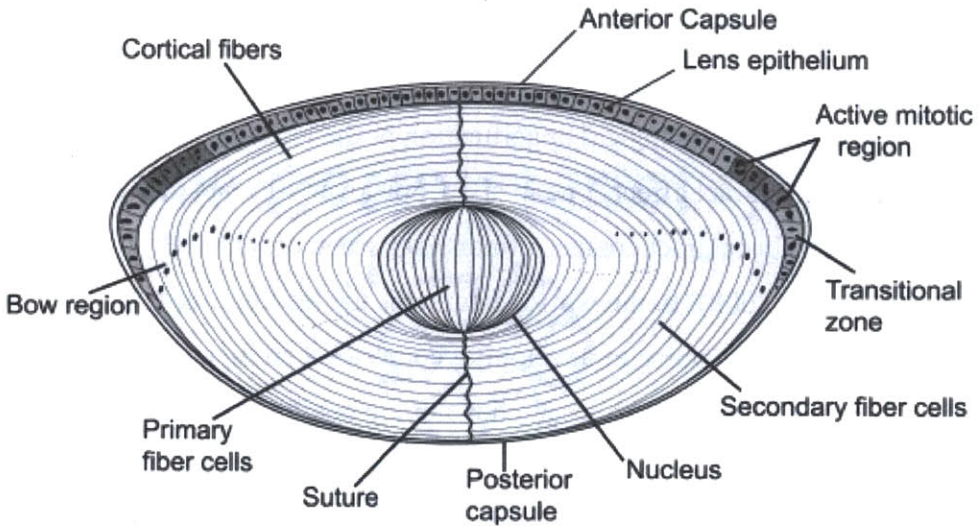


Figure 1-8 Lens Structure.

(A) Lens vesicle. (B-C) Differentiation of posterior epithelial cells into primary fiber cells. (D-E) Differentiation of anterior epithelial cells into secondary fiber cells. The fiber cells lose their nuclei and other organelles upon differentiation.

(Reprinted from *Biochim Biophys Acta.*, 1790(10): 1095-1108, Sharma and Santhoshkumar, Lens aging: effects of crystallins. Copyright (2009), with permission from Elsevier).

2. *Cataract: the Leading Cause of Blindness in the World*

Cataract results from pathological opacification of the lens that leads to the loss of transparency. In the US, cataract affects one in six people over 40 years old, and half of people over 80 years old. Annually cataract creates a financial burden of $6.8 billion to the healthcare system (Shoemaker *et al.* 2008). In the world, cataract is the leading cause of vision impairment and blindness. The situation is worse in developing countries than in developed countries, largely due to insufficient cataract surgical services in developing countries (Resnikoff *et al.* 2004). Currently there is no cure for cataract. The only effective treatment is surgical implant of an artificial intraocular lens after the removal of the cataract lens (Oyster 1999). However, this procedure is invasive, expensive and not without complication. Posterior capsule opacification (PCO) is the most common complication, and occurs in 10% of patients after one year and in about 40% of patients after four years (Wormstone 2002). PCO results from growth of residual lens epithelial cells and requires additional surgical procedures.

Several risk factors have been reported to be associated with cataract, such as smoking, UV exposure and diabetes (Abraham *et al.* 2006). However, a causative relationship or the underlying mechanism is not clear for these factors. We have a much clearer understanding of the immediate cause of cataract. Loss of lens transparency may be caused by disruption of 1) the orderly packed fiber cells, or 2) the short-range order of the intracellular soluble proteins. The leading theory of cataract formation suggests that cataract is caused by light scattering of irreversible, high-molecular-weight aggregates of crystallins (Jedziniak *et al.* 1973; Benedek 1997). Phase separation of the protein-water mixture into protein-rich and water-rich phases in low temperature can cause cold-cataract (Tanaka and Benedek 1975). Other proposed mechanisms include abnormal osmotic pressure due to defect in the active pumps in the lens epithelial cells, leading to influx of water to the lens and the disruption of order on both the molecular and cellular level (Jaenicke and Slingsby 2001). Crystallization of crystallins inside the lens also represents a rare type of cause of cataract (Kmoch *et al.* 2000). Stepping one level upstream, the causes of these abnormalities in the lens could be genetic or environmental: 1) congenital mutations in crystallins or other lens proteins typically cause juvenile cataract, or early-onset cataract, but the occurrence is not common, and 2) acquired damages to crystallins and other lens proteins cause the more common type, age-onset cataract.

a. Juvenile Cataract—Caused by Congenital Mutations

Juvenile cataract is rare compared to age-onset cataract, and is usually caused by congenital mutations in the crystallin genes. As summarized by Graw in 2009, at least 13 mutations in α-crystallins, 10 in β-crystallins and 16 in γ-crystallins have been identified to be associated with cataract (Graw 2009). Novel mutations continue to emerge (Roshan *et al.* 2010; Wang *et al.* 2011). γD-crystallin is the most affected gene among all crystallins. Point mutations that cause single amino acid changes are the most common type, and they include R14C, P23T, P23S, R36S, R36P, R58H, G61C, R76S and E107A (Heon *et al.* 1999; Stephan *et al.* 1999; Kmoch *et al.* 2000; Santhiya *et al.* 2002; Messina-Baas *et al.* 2006; Plotnikova *et al.* 2007; Li *et al.* 2008; Roshan *et al.* 2010; Wang *et al.* 2011). These mutations cluster on the N-terminal domain of γD-crystallin. All affected amino acids are surface exposed and a majority of them are charged resides (Figure 1-9). *In vitro* studies on these mutant γD-crystallins showed that most of these mutant proteins maintained their WT-like native structure. Enhanced aggregation or precipitation seemed to stem from intermolecular disulfide bonds for R14C (Pande *et al.* 2000), insolubility for P23T (Evans *et al.* 2004), and decreased stability for G61C (Zhang *et al.* 2011). The reasons that congenital mutations are biased towards surface residues over buried residues that are also destabilizing remain to be understood. It is possible that α-crystallin is more capable of rescuing aggregation-prone species from hydrophobic core mutant βγ-crystallins rather than surface mutant βγ-crystallins.

Crystallin genes consist of the majority of the cataract-associated mutation targets, but other proteins in the lens have also been reported. Gap junctions, made of the proteins connexins, are channels that permeate small molecules between adjacent cells. Gap junctions present in lens cells make the lens a functional syncytium, and allow the epithelial cells to maintain the precise ionic balance throughout the fiber cells (White and Bruzzone 2000). Mutations in Cx50, one of the major lens connexins, have been identified to be associated with zonular pulverulent cataract (Shiels *et al.* 1998). Cytoskeletal support for fiber cells is provided by a large number of intermediate filaments (IF). Among these IFs, vimentin and beaded filament structural proteins (BFSP) are most abundant. Interaction between BFSP and α-crystallin was important for BFSP morphology and may be important in the spatial organization of α-crystallin (Carter *et al.* 1995).

Mutations in both vimentin and BFSP have also been identified to be associated with cataract (Song *et al.* 2009).

b. Age-onset Cataract—Caused by Covalent Modifications

The prevalence of age-onset cataract increases dramatically starting at around age 50 (Shoemaker *et al.* 2008). The leading hypothesis proposes that a lifetime accumulation of covalent damages to the crystallins lead to their partial unfolding and aggregation. The chaperone α-crystallin may delay the aggregation process by sequestering the aggregation-prone species. However, once the chaperone capacity is reached, α-crystallin can no longer prevent aggregation, and may itself become aggregation-prone. The agents that cause protein damages may be reactive oxygen species (ROS), ultraviolet (UV) radiation, or a combination of both (Grossweiner 1984; Davies and Truscott 2001; Truscott 2005; Brennan and Kantorow 2009). UV absorption in proteins is largely through aromatic amino acids, cysteines and cystines (disulfide bonds). Photo-oxidation of these resides may occur by direct pathway or indirect pathway mediated via oxygen. Formation of di-tyrosine, cleavage of the Phe side chain off the backbone, opening of the 5-member rings on Trp and His are some of the common products of these reactions. These covalent damages may induce structural alterations or unfolding. Cross-linking reactions can also occur between oxidized His and Lys/Cys/His. Intermolecular cross-linking reactions, such as the formation of di-tyrosine, and cross-links between His and Lys/Cys/His can create oligomers and high-molecular-weight aggregates (Davies and Truscott 2001). The evidence of photo-oxidation-induced backbone cleavage has been scarce. In the few cases, they are also mediated by initial radical formation on the aromatic residues (Davies and Truscott 2001).

Although photo-oxidative damage to proteins has been well documented, suggesting a mechanism of cataract development, evidence from epidemiological studies was not strong. For example, studies by Taylor *et al.* showed an association between exposure to UVB and cortical cataract, but not nuclear cataract. No correlation was established between UVA and cataract (Taylor *et al.* 1988; Taylor 1999). The inconclusive results may be due to difficulties in quantifying UV exposure of an individual.

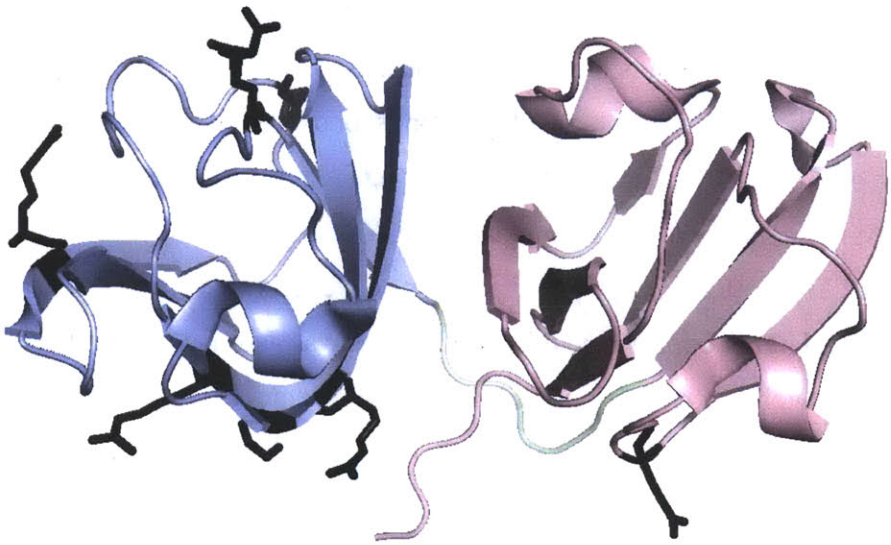


Figure 1-9 Cataract-associated Congenital Mutations on Human γD-Crystallin.
Locations of the amino acids on human γD-crystallin affected by congenital mutations associated with cataract. The affected residues are shown in black sticks. The amino acid changes include R14C, P23T, P23S, R36S, R36P, R58H, G61C, R76S, E107A. Light blue: N-terminal domain (N-td); pink: C-terminal domain (C-td).
(Constructed from crystal structure from Basak *et al.* 2003).

Lending support to the damage-aggregation theory of cataractogenesis, a wide range of covalent modifications on crystallins have been identified by proteomic studies on human lens, including non-enzymatic glycosylation, disulfide bonds, phosphorylation, oxidation of Met and Trp, deamidation of Gln and Asn, and backbone cleavage (Chiou *et al.* 1981; Hanson *et al.* 1998; Lampi *et al.* 1998; Ma *et al.* 1998; Hanson *et al.* 2000; Takemoto and Boyle 2000; Harrington *et al.* 2004; Hains and Truscott 2007). The modifications cover all three classes of crystallins (α-, β-, and γ-crystallins). Higher appearance of modifications was associated with water-insoluble protein fractions of protein, as well as with aged lens and cataract lens. Although the support is strong, a direct causative relationship between protein damage and cataract is still lacking.

Among these modifications, deamidation of Asn and Gln seems to be the most common type of modification. Perhaps due to the convenience of generating site-specific deamidation, *in vitro* effects of deamidation to crystallins have been extensively studied in βB1-, B2-, A3-, and γD-crystallins. The results of these studies showed that the effects of deamidation are multi-faceted, including 1) local structural change and increased solvent accessibility (Takata *et al.* 2007; Takata *et al.* 2010); 2) decreased thermodynamic stability (Flaugh *et al.* 2006); 3) decreased or destabilized dimer/oligomer formation (Lampi *et al.* 2006; Takata *et al.* 2009); and 4) increased thermal aggregation and higher α-crystallin requirement for aggregation rescue (Lampi *et al.* 2002; Takata *et al.* 2008; Michiel *et al.* 2010). Deamidation is a relatively mild damage among the list of modifications aforementioned, but it can cause such broad and profound structural perturbations to crystallins. It is very likely that other more severe modifications would induce more substantial changes to the structure and interaction of crystallins.

F. Classes of Lens Crystallins

Crystallins constitute 20-60% of the wet weight of the lens, and 90% of lens soluble proteins (Wistow and Piatigorsky 1988; Horwitz 2000). α-, β- and γ-crystallins are the major classes of crystallins (Bloemendal *et al.* 2004). The nomenclature originated from the order by which they were eluted by size-exclusion chromatography (SEC). β- and γ-crystallins are structurally related. In this section, α-, β- and γ-crystallins, as well as proteins structurally similar to βγ-crystallins are discussed.

In addition to the major classes of crystallins, certain vertebrates may have one to several additional taxon-specific crystallins in their lens. These taxon-specific crystallins are identical or related to metabolic enzymes and were probably recruited to the lens as structural proteins during evolution. Some of them retain enzymatic activity in the lens (Wistow and Piatigorsky 1988). At least 12 classes of taxon-specific crystallins exist. The most widely distributed ones includes δ-crystallin in birds and reptiles that is identical to argininosuccinate lyase, ε-crystallin in some birds and crocodiles that is identical to lactate dehydrogenase, and τ-crystallin in many vertebrates that is related to α-enolase (Wistow and Piatigorsky 1988; Jaenicke and Slingsby 2001). Taxon-specific crystallins are structurally unrelated to the major classes of crystallins, and appear to be lacking in human lens.

1. βγ-Crystallins

a. Expression

The human genome encodes seven γ-crystallins: γA-F on a gene cluster and γS on a separate chromosome, with a somewhat more divergent sequence. The predominant γ-crystallins synthesized in young human lens are γC- and γD-crystallin (Siezen *et al.* 1987). γE and γF are inactive pseudogenes. γA and γB have low expression levels due to low rate of transcription for γA and high rate of mRNA degradation for γB (Brakenhoff *et al.* 1990). Due to their early expression during lens development and late expression during fiber cell differentiation, γC and γD-crystallins are enriched in the lens nucleus. Together, γC and γD-crystallins account for close to 20% of the soluble proteins in young human lens (Lampi *et al.* 1997; Robinson *et al.* 2006).

γS-Crystallin is expressed predominantly during adulthood and present in the lens cortex (Lampi *et al.* 1997; Wistow *et al.* 2000).

β-Crystallins are further divided into two groups: βA(acidic)-crystallins, including βA1-4, and βB(basic)-crystallins, including βB1-3. βA1 and βA3 are products of the same gene, and are identical except the different lengths of the N-terminal extensions, due to alternative translation initiation sites on the same mRNA transcript (Wistow and Piatigorsky 1988). Expression of all β-crystallins except βA2 were confirmed in young human lens. β-Crystallins account for 42.5% of soluble proteins in young human lens (Lampi *et al.* 1997). Only trace amount of βA2-crystallin is expressed, likely due to a block in translation (Bloemendal *et al.* 2004).

The expression of different members of βγ-crystallins is highly regulated temporally and spatially, creating a refractive index gradient for the lens. This gradient corrects the optically imperfect shape of the lens and contributes to the light-focusing function of the lens (Siezen *et al.* 1988).

b. Structure

βγ-Crystallins have homologous structures and are the major structural proteins in the lens. They are about 20-30 kDa in size, and have a double-crystallin-domain, primarily β-sheet structure (Najmudin *et al.* 1993; Basak *et al.* 2003; Wu *et al.* 2005) (Figure 1-10). The two crystallin domains are homologous, each containing two Greek key motifs and eight β-strands. The repeating elements of βγ-crystallins presumably arose from ancient gene duplication and fusion events (Wistow and Piatigorsky 1988). Sequence homology between the homologous Greek keys from different crystallin domains is higher than the two Greek keys from the same crystallin domain, suggesting duplication of individual Greek key within the crystallin domain before the duplication of the crystallin domain. In each crystallin domain, the two Greek key motifs do not fall neatly onto the two β-sheets, leading to a complicated fold that may contribute to the high stability of these proteins (MacDonald *et al.* 2005) (Figure 1-10). The crystallin domain may be viewed as a compressed β-barrel (Branden and Tooze 1999).

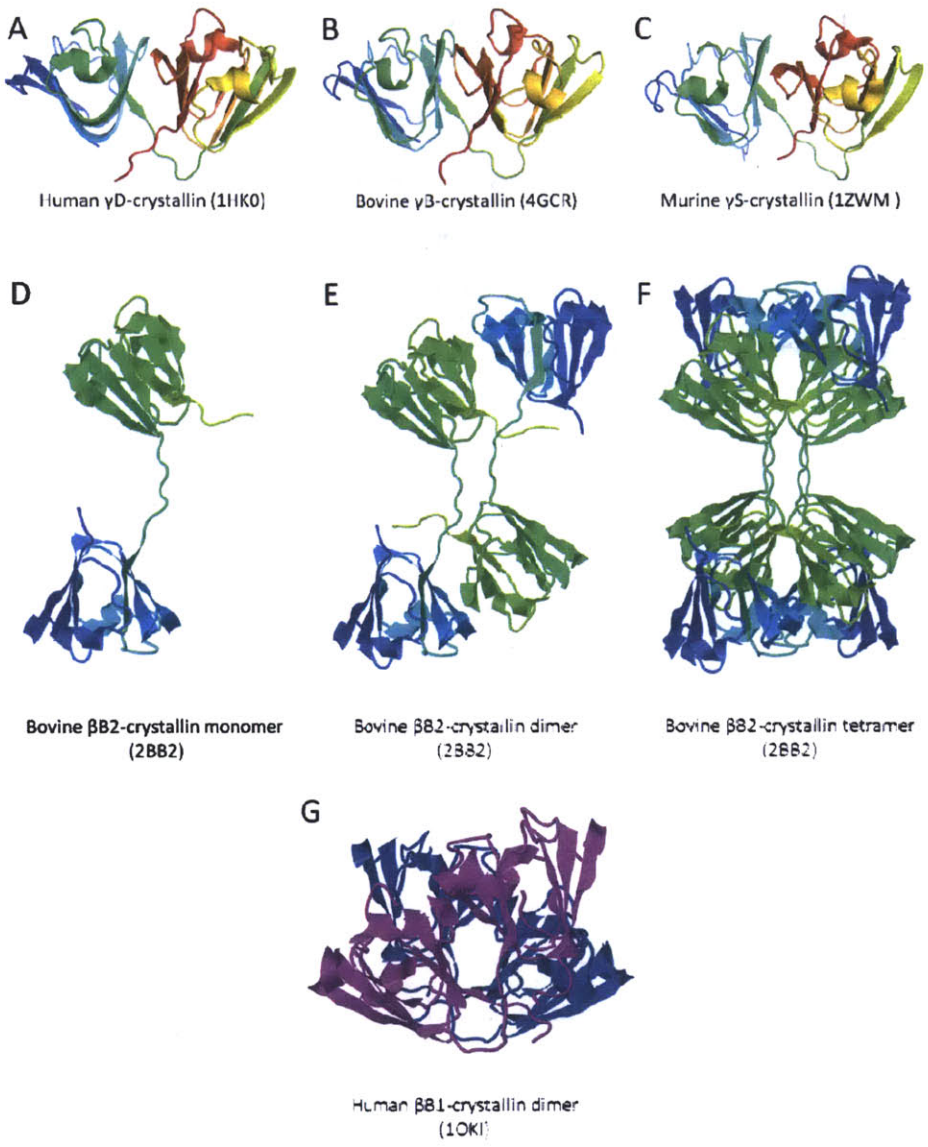


Figure 1-10 Structures of βγ-Crystallins.

(A-C) The structures are shown in rainbow color. Blue end: N-terminus; red end: C-terminus. (A) Crystal structure of human γD-crystallin (Basak *et al.* 2003). (B) Crystal structure of bovine γB-crystallin (Najmudin *et al.* 1993). (C) NMR structure of murine γS-crystallin (Wu *et al.* 2005). (D-F) Crystal structure of bovine βB2-crystallin (Bax *et al.* 1990). The structures are shown in rainbow color. Blue end: N-terminus; green end: C-terminus. (D) Conformation of the monomer in the dimer and tetramer. (E) Dimer. (F) Tetramer. (G) Crystal structure of human βB1-crystallin dimer (Van Montfort *et al.* 2003). Blue and purple colors indicate the two monomers.

β-Sandwich is a commonly observed protein fold. In addition to the aforementioned unique topology, the crystallin domain is defined by several conserved structural motifs. Figure 1-11 illustrates these motifs, using human γD-crystallin as an example. First, each Greek key motif contains an aromatic pair located at the β-hairpin near the turn region, with the consensus sequence Y/FXXXXY/FXG (termed "Greek key pair"). Both displaced stacking and perpendicular orientations of the aromatic rings are observed. Second, a set of conserved Gly, Ser and Arg are also present surrounding each Greek key aromatic pair (Blundell *et al.* 1981). While the flexibility of Gly makes the β-turn possible, Ser forms a hydrogen bond with the backbone nitrogen of the second (towards the C-terminus) member of the Greek key aromatic pair. The Arg packs against the second aromatic ring, possibly making charge-aromatic interaction, further stabilizing this region (Figure 1-11) (Blundell *et al.* 1981). Among the four Greek keys in human γD-crystallin, the third Greek key seems to deviate somewhat from the other three, being the only parallel pair and with the Arg replaced by Asn. Third, one Δ4 tyrosine corner is found in the second Greek key in each crystallin domain. For classic Δ4 tyrosine corners in other Greek key beta-barrel proteins, the sequence consensus is LXPGXY (Hemmingsen *et al.* 1994). In βγ-crystallin, the tyrosine corners seem not to follow this consensus strictly. However, the structure is maintained: the -OH groups on the tyrosines make hydrogen bonds with the backbone of the residue four amino acid ahead in the sequence, bridging the two β-sheets of the crystallin domain (Figure 1-11). Tyrosine corners are important in the stability of Greek key proteins, but may not be important in nucleating folding (Hamill *et al.* 2000; Nicaise *et al.* 2003). Last, a pair of buried tryptophan residues is also conserved at the crystallin level (Figure 1-11). The fluorescence of these tryptophans makes it convenient to follow the folding status of βγ-crystallins. An unusual native state quenching behavior has been observed for the tryptophans in human γD-crystallin. Experimental and simulation studies showed that the quenching was achieved by efficient energy transfer from Trp42 to Trp68 in the N-terminal domain (N-td), and from Trp130 to Trp156 in the C-terminal domain (C-td), and then dissipated through the backbones. Efficient electron transfer to the backbone is mediated by several surrounding charged residues and water molecules (Chen *et al.* 2006; Chen *et al.* 2008; Chen *et al.* 2009).

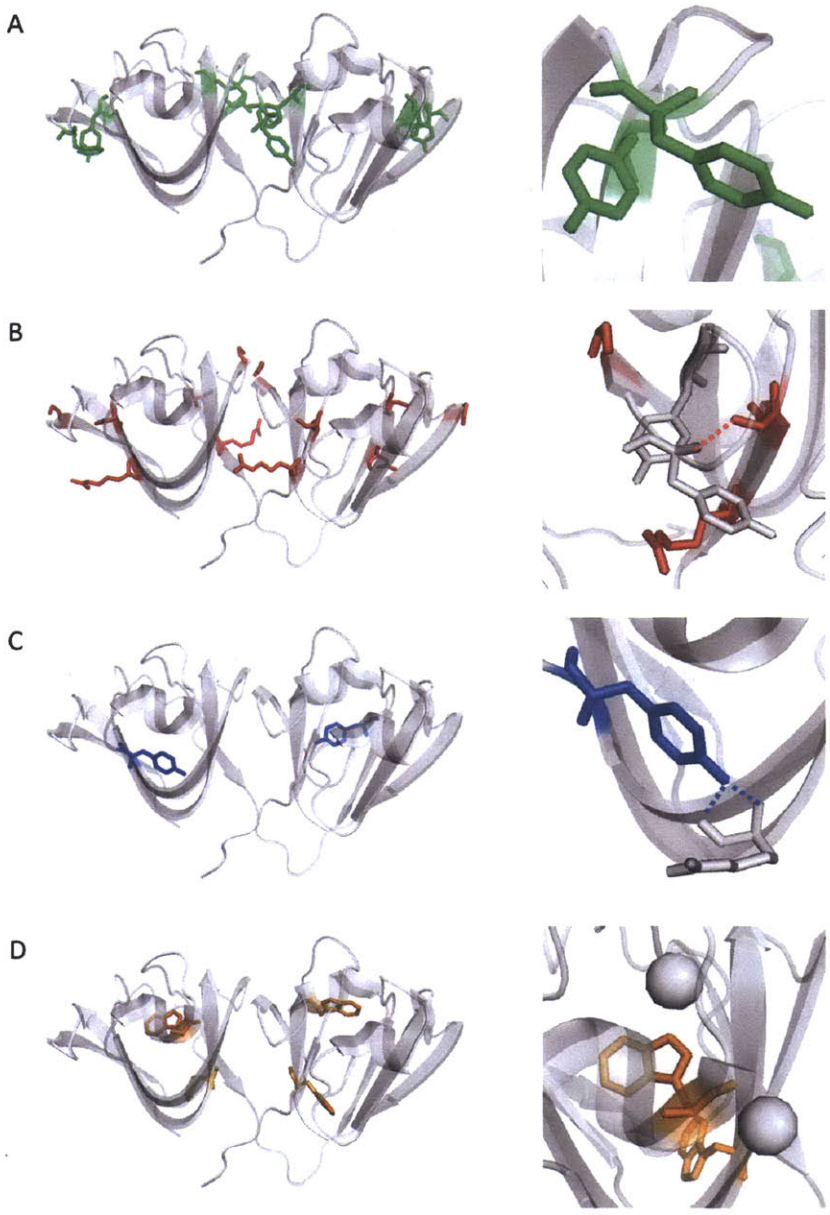


Figure 1-11 Conserved Motifs of βγ-Crystallins.

The motifs are shown in human γD-crystallin as an example. (A) Greek key aromatic pairs. (B) The conserved Gly, Ser and Arg surrounding the Greek key aromatic pair. (C) Tyrosine corners. (D) The Trp responsible for the fluorescence quenching property of human γD-crystallin. Two crystallographic water molecules shown in grey spheres are important in the electron transfer from the Trp to the backbone. Left panels: locations of the motifs in human γD-crystallin. Right panels: close-ups of the motifs.

(Constructed from crystal structure from Basak *et al.* 2003).

γ-Crystallins are monomeric. The available crystal and NMR structures of γ-crystallins (Figure 1-10) revealed an intramolecular crystallin domain interface. While the double-crystallin-domains structure is largely the same as γ-crystallins, two properties set β-crystallins apart from γ-crystallins: the extensions from the core domains and the linker of the two domains. βA-Crystallins have N-terminal extensions, while βB-crystallins have both N-terminal and C-terminal extensions. In truncated βB2-crystallin, these extensions were shown to be important in higher oligomeric states of the proteins (Trinkl *et al.* 1994). Crystal structure of βB2-crystallin revealed a domain-swap dimer formation, with an intermolecular crystallin domain interface. The crystal structure also revealed tetramer formation, using a separate interface on the "side" of the dimer (Figure 1-10) (Bax *et al.* 1990). Interestingly, truncated βB1-crystallin also forms dimers, but with each monomer in γ-crystallin-like conformation, and the crystallin domains making intramolecular interaction. The dimer interface of βB1-crystallin is similar to the tetramer interface seen in βB2-crystallin (Figure 1-10) (Van Montfort *et al.* 2003). It was suggested that several differences in the connecting peptides between β- and γ-crystallins, such as a flexible Gly residue and the peptide length, may dictate the oligomeric state and the type of dimer formation (Bax *et al.* 1990). Swapping the connecting peptides of monomeric γB- and dimeric βB2-crystallins resulted in two monomeric proteins, indicating that features beyond the connecting peptide, such as the domain interfaces may be important in dictating the oligomeric state (Mayr *et al.* 1994; Trinkl *et al.* 1994). β-Crystallins also form higher-ordered homo- and hetero-oligomers, adding another layer of complication to the rules of oligomer formation of β-crystallins (Slingsby and Bateman 1990).

Proteins with similar folds to lens βγ-crystallins are found in lower organisms (Table 1-3). These proteins share from some to no sequence homology with lens βγ-crystallins. On one end of the spectrum, Absent In Melanoma 1 (AIM1) appears to retain most of the conserved motifs, including the Greek key aromatic pairs, tyrosine corners and the characteristic Trp (Figure 1-11) (Ray *et al.* 1997). On the other end of the spectrum, *Streptomyces* metalloproteinase inhibitor (SMPI) has no sequence similarity at all with lens βγ-crystallins (Ohno *et al.* 1998). The number of crystallin domains also varies from protein to protein (Table 1-3). Interestingly, nitrollin and spherulin 3a are single-domain βγ-crystallin-like proteins that form homodimers, but they use alternative interfaces compared to those in lens βγ-crystallins (Aravind *et al.* 2009).

Protein	Organism	Number of Crystallin Domains	Function	Reference
AIM1 (absence in melanoma)	human	6	putative suppressor of human malignant melanoma	(Ray *et al.* 1997)
Ciona crystallin	sea squirt *Ciona intestinalis*	1	expressed in light-sensing system	(Shimeld *et al.* 2005)
EDSP (epidermis differentiation-specific protein)	newt *Cynops pyrrhogaster*	2	development	(Wistow *et al.* 1995)
WmKT (*Williopsis mrakii* killer toxin)	Yeast *Williopsis mrakii*	1	inhibit growth of foreign yeast stains	(Antuch *et al.* 1996)
Spherulin 3a	slime mold *Physarum polycephalum*	1	stress response leading to encystment and dormancy	(Wistow 1990)
Protein S	sporulating bacterium *Myxococcus xanthus*	2	structural protein in spore coat	(Wistow *et al.* 1985)
SMPI (Streptomyces metalloproteinase inhibitor)	bacterium *Streptomyces nigrescens*	1	metalloproteinase inhibitor	(Ohno *et al.* 1998)
Nitrollin	soil bacterium *Nitrosospira multiformis*	1	unknown	(Aravind *et al.* 2009)

Table 1-3 βγ-Crystallin-like Proteins.

The extended βγ-crystallin family contains a calcium binding motif on the "top" loops (orientated as in Figure 1-11) that can potentially bind two calcium ions per crystallin domain (Aravind *et al.* 2009). Calcium binding induces a range of effects from passive binding to active stabilization. However, this motif and thus the calcium binding ability is partially or completely lost in lens βγ-crystallins (Aravind *et al.* 2009). These βγ-crystallin-like proteins have diverse functions outside the visual system, except for *Ciona* crystallin, which was found to be expressed specifically in the light sensing organs (Shimeld *et al.* 2005). The promoter of *Ciona* crystallin could drive green fluorescence protein (GFP) expression in vertebrate visual system, suggesting *Ciona* crystallin as the ancestor of lens βγ-crystallins in higher vertebrates (Shimeld *et al.* 2005).

As summarized above, there exist taxon-specific lens crystallins without the Greek key/crystallin-domain fold, as well as βγ-crystallin-like proteins outside the lens. Therefore, it seems likely that the crystallin-domain β-sandwich fold and the lens optical properties have no strict dependence on each other.

c. Stability and Folding

γ-Crystallins are in general very stable proteins. Bovine γB-crystallin was shown to resist unfolding at pH 1-10, or at 75 °C. It could only be fully unfolded with high concentration of urea combined with low pH or high temperature (Rudolph *et al.* 1990). Human γD-crystallin also has a very high profile of stability, resistant to denaturation by up to 8 M urea (Kosinski-Collins and King 2003) and with a thermal denaturation midpoint of >80 °C (Flaugh *et al.* 2006). Comparison of data obtained from different studies showed that γ-crystallins are in general more stable than β-crystallins (Jaenicke and Slingsby 2001; Bloemendal *et al.* 2004), although counter-examples were also reported (Fu and Liang 2002). Direct comparison of the stability between β- and γ-crystallins may not be meaningful due to the ability of β-crystallins to form dimer and oligomers. For example, dissociation of rat βB2-crystallin dimer occurred simultaneously with the unfolding of the N-td (Wieligmann *et al.* 1999). Hetero-oligomer formation of β-crystallins also means that the molecular context is lost when these proteins are studied in isolation (Bloemendal *et al.* 2004).

Sequential unfolding/refolding has been observed for rat βB2-crystallin, bovine γB-crystallin and human γD-crystallin (Rudolph *et al.* 1990; Wieligmann *et al.* 1999; Kosinski-

Collins *et al.* 2004). In bovine γB-crystallin, the C-td unfolds first, followed by the N-td, while in human γD-crystallin and rat βB2-crystallin, the opposite order was observed, suggesting that the folding/unfolding orders were set after the these proteins diverged. Isolated crystallin domains could fold autonomously, indicating that the two-domain organization did not create any folding dependency between the domains (Rudolph *et al.* 1990; Mayr *et al.* 1997; Wenk *et al.* 2000; Mills *et al.* 2007). The folding/unfolding orders observed probably stemmed from the intrinsic stability difference of the two domains. The domain interface contributed to the stability of the full-length proteins in γB and γD-crystallins (Mayr *et al.* 1997; Mills *et al.* 2007) but not in γS-crystallin (Wenk *et al.* 2000).

Neither the higher order assembly nor the crystallin domain interface universally confer higher stability for βγ-crystallins, suggesting that the intercalated β-strands and other features within individual crystallin domains are the main contributors for the stability of βγ-crystallins (MacDonald *et al.* 2005).

2. *α-Crystallin*

This section focuses on α-crystallin, see Section D.5 in this introduction for general information about small heat shock proteins (sHsp).

a. Expression

α-Crystallins belong to the small heat shock protein (sHsp) family. In humans, there are two members, αA and αB, encoded by genes located at different chromosomes. On the protein level, αA- and αB-crystallins consist of 173 and 175 amino acids, respectively, with 57% sequence identity. α-Crystallins make up more than 30% of the soluble proteins in the lens. Within one year after birth, the expression of αA-crystallin drops from 30% to 24%, while αB-crystallin maintains at about 7% (Robinson *et al.* 2006). α-Crystallins used to be, but are no longer considered truly lens-specific. By immunoassay, αB-crystallin was found to be present in the retina, heart, muscle, skin, brain, spinal cord and lungs (Bhat and Nagineni 1989). αA-Crystallin is largely lens-specific, but was also found in thymus and spleen at relatively high levels, and multiple other organs at low levels (Kato *et al.* 1991). mRNA transcripts of both αA- and αB-crystallins have even wider distributions in various parts of the body (Graw 2009). Controlled by heat-shock elements in the promoter region, the expression of αB-crystallin, but not αA-crystallin, is inducible by heat and various types of stress. (de Jong *et al.* 1993). This is in line with the largely lens-specific expression pattern for αA-crystallin. Response to heat or stress for sHsps probably plays a larger role in tissues outside the lens, since 1) the lens lacks protein synthesis machinery; 2) the expression of α-crystallins is already high in the lens; and 3) alteration of protein concentration would probably disturb the finely established refractive index gradient in the lens.

b. Structure

Like other sHsps, the sequences of α-crystallins are organized into a central β-sandwich α-crystallin domain (αCD) of about 80 amino acids, a long N-terminal extension and a short C-terminal extension. α-Crystallins form dimer units largely through the interactions of the αCDs.

In sHsps from lower organisms, dimer formation processes are probably mediated through swapping of β6 of the αCDs (Kim *et al.* 1998; van Montfort *et al.* 2001; Hilario *et al.* 2011). However, in α-crystallins, β6 and β7 fuse into an extended strand called β6+7. Truncated αCDs of bovine αA-crystallin and human αB-crystallin formed dimers by constructing an extended β-sheet through a symmetrical β6+7/β6+7 interface (Figure 1-12) (Bagneris *et al.* 2009; Laganowsky *et al.* 2010; Clark *et al.* 2011). Additional crystal structures of rat Hsp20, human Hsp27 and Hsp20 suggested this mode of dimerization to be common for mammalian sHsps (Bagneris *et al.* 2009; Baranova *et al.* 2011). The dimer unit builds a shared groove at the dimer interface, lined with positive charges including the cataract-associated arginines R116 in αA and R120 in αB. Suggested from sHsp structures from lower organisms and the topology of the β-sandwich, this groove may be filled by the N-terminal extension (Bagneris *et al.* 2009). On the ends of the α-crystallin dimer, two hydrophobic pockets are created by the β4/β8 interface. These pockets are filled by C-terminal extensions containing the conserved IPI/V motifs from the neighboring dimers (Figure 1-12) (Laganowsky *et al.* 2010).

Comparison of different human sHsp structures also revealed possible mechanisms of polydispersity of α-crystallins (Figure 1-12) (Laganowsky *et al.* 2010). The polydispersity is achieved by 1) bi-directional binding of the IPI/V-motif-containing C-terminal extension to the hydrophobic pocket of a neighboring dimer unit, made possible by the palindromic sequences of the C-terminal extensions; 2) at least three different registrations of the β6+7/β6+7 dimer interface; and 3) the flexibility of the hinge loop connecting the C-terminal extension and the αCD (Laganowsky *et al.* 2010).

A defined quaternary structure of full-length α-crystallins has been lacking. The majority of the evidence showed that α-crystallins form polydisperse oligomers, with undefined number of subunits. The number of subunits ranges widely from 15-60, leading to complexes with molecular weight of 300-1200 kDa, with average of 800 kDa (Horwitz 1992; Haley *et al.* 1998; Haley *et al.* 2000). Cryo-EM structure reconstruction has proposed a highly symmetrical αB-crystallin 24-mer model (Peschek *et al.* 2009). One must interpret this model with care because the cryo-EM structure reconstruction was only possible by selecting a narrow subset of molecules from a highly heterogeneous pool.

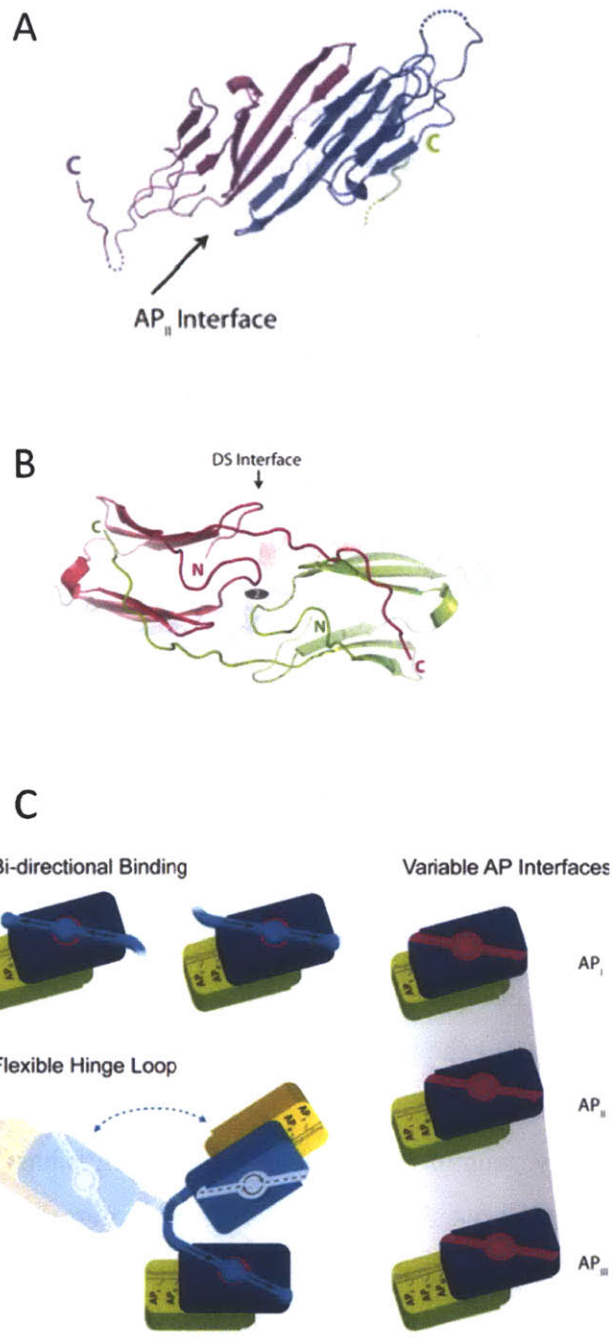


Figure 1-12 Dimer Organization and Mechanisms of Polydispersity of α-Crystallins.
(A) The dimer interface, formed by β6+7/β6+7. (B) Docking of the C-terminal IPI/V motifs into the β4/β8 pockets of neighboring dimer units. (C) Mechanisms of the polydispersity of α-crystallins.
(Reprinted with permission from Laganowsky *et al.* 2010).

Human α-crystallins form hetero-complexes with αA:αB ratio about 3:1 *in vivo*. Purified recombinant αA- or αB-crystallin was shown to be less polydisperse and less stable than the native hetero-complex, indicating that the heterogeneity was optimized for *in vivo* stability (Horwitz *et al.* 1999). The heterogeneity was also suggested to be important for preventing self-aggregation and glycation-induced structural modification of αB-crystallin (Srinivas *et al.* 2010). Recombinant αA- and αB-crystallins had different secondary and tertiary structures. αB-Crystallin was more hydrophobic and had higher chaperone activity against insulin aggregation than αA-crystallin (Sun *et al.* 1997). Overall, purified αA- or αB-crystallin retained the typical complex size and chaperone activity in large part similar to the native complex. For simplicity, recombinant αA- or αB-crystallin is often used in structural and functional studies of α-crystallin.

c. Subunit Exchange

The difficulty in crystallizing α-crystallin is not only due to the polydispersity of α-crystallin, it is also because the subunits of α-crystallin complexes undergo rapid exchange. Bova *et al.* have used fluorescence resonance energy transfer (FRET) to study subunit exchange of α-crystallins. Subunit exchange for αA-crystallin was shown to be markedly increased at elevated temperature, suggesting its role in chaperone function. The exchange reaction was independent of pH and calcium concentration, but decreased at both high and low ionic strength, suggesting involvement of both ionic and hydrophobic interactions in subunit-subunit interactions (Bova *et al.* 1997). Subunit exchange was independent of αA-crystallin concentration, suggesting that subunit dissociation may be the rate-limiting step. Truncation studies showed that residues 20-56 immediately before the αCD, but not the N-terminal 19 or C-terminal 10 residues, were important for subunit exchange or oligomeric assembly of αA-crystallin (Bova *et al.* 2000). Exchanges between αA-crystallin and αB-crystallin or Hsp27 were also reported (Bova *et al.* 2000).

d. Functions

The high concentration of α-crystallins in the lens indicates their structural role—together with βγ-crystallins, to produce a high refractive index for the lens. In line with the high

expression of α-crystallins in the lens, mutations in α-crystallins have been identified associated with cataract in human and mouse (Graw 2009). As sHsps, α-crystallins also have chaperone function. α-Crystallins were first identified as sHsps by sequence homology with four *Drosophila* sHsp by Ingolia and Craig (Ingolia and Craig 1982). Subsequently, it was demonstrated that α-crystallin could indeed suppress aggregations of a wide range of substrates, such as WT βγ-crystallins (Horwitz 1992; Evans *et al.* 2008; Acosta-Sampson and King 2010), cataract-associated T5P γC-crystallin (Liang 2004), and non-lens proteins such as alcohol dehydrogenase, alkaline protease and insulin (Horwitz 1992; Sun *et al.* 1997; Tanksale *et al.* 2002). In some cases, increased recovery of native substrates during refolding in the presence of α-crystallin was also reported. It is generally accepted that like other sHsps, α-crystallins exert their chaperone function by sequestering the partially-unfolded and aggregation-prone substrates in an ATP-independent manner, preventing their aggregation (Tanksale *et al.* 2002; Evans *et al.* 2008; Acosta-Sampson and King 2010), but cannot refold misfolded proteins or disassemble aggregates. Therefore, the increased recovery may be interpreted as increased partitioning of productive refolding pathway over the suppressed aggregation pathway, rather than active refolding by α-crystallin. ATP enhancement of chaperone activity and substrate release was also reported, but the mechanisms were not well established (Muchowski and Clark 1998; Rawat and Rao 1998).

Besides the chaperone and the optical functions in the lens, α-crystallins are also found to interact with cytoskeletal proteins within and beyond the lens. For example, bovine αA-crystallin was found to associate strongly with intermediate filaments and beaded filament in the lens. In addition, α-crystallin was shown to interact with cytoskeletal proteins actin and desmin in mouse heart (Chiesi *et al.* 1990; Longoni *et al.* 1990). In addition to cataract, mutations in αB-crystallin are also associated with muscular diseases desmin-related myopathy and dilated cardiomyopathy (Graw 2009). α-Crystallins are presumably involved in the organization of the cytoskeleton in the cells involved in these diseases.

e. Substrate Binding Sites

Cross-linking experiments on bovine α-crystallins using alcohol dehydrogenase (ADH) indicated that both αA- and αB-crystallins interacted with ADH. Further analysis identified two

13-15 amino acid regions, one in the N-terminal extension and one in the αCD as ADH binding site (Sharma *et al.* 1997). Similar experiments with hydrophobic protein melittin identified a very different set of regions on both αA- and αB-crystallins, also throughout the N-terminal extension and the αCD (Sharma *et al.* 2000). Given the general belief that chaperones recognize misfolded proteins by hydrophobic interactions, photo-labeling using bis-ANS or related hydrophobic dyes was also used to locate the substrate binding sites on α-crystallins. These studies again identified several regions dispersed throughout the N-terminal extension and the αCDs of αA- and αB-crystallins (Smulders and de Jong 1997; Sharma *et al.* 1998). The involvement of these sites in substrate binding is supported by the reduced chaperone activity of the labeled α-crystallin (Smulders and de Jong 1997). From these cross-linking studies, overlapping sets of candidate regions were obtained. Strikingly, synthetic peptide corresponding to one of the candidate regions, residues 70-88 on αA-crystallin, retained bis-ANS binding ability, chaperone activity towards thermal aggregation of ADH, and the ability to form a peptide-substrate complex (Sharma *et al.* 2000). Substitution of a single Phe in this region, Phe71, abolished the chaperone activity of αA-crystallin, further supporting the importance of this region in chaperone activity, likely participating in substrate binding (Santhoshkumar and Sharma 2001).

From these experiments, one can draw the following conclusions. First, there may be multiple regions involved in substrate binding. Second, different substrates may interact with different regions of α-crystallins. Third, substrate binding may involve more than just hydrophobic interactions. In addition to hydrophobic interactions, electrostatic interactions may be helpful for the long-term stability of the chaperone-substrate complex (Narberhaus 2002).

In terms of substrate binding site relative to the oligomeric complex, immunogold studies supported the theory that the substrates bound to the central region of an α-crystallin oligomer (Boyle *et al.* 1993; Boyle and Takemoto 1994). This was consistent with cryo-EM studies that showed when bound with α-lactalbumin, additional density was found in the α-crystallin hollow central cavity (Haley *et al.* 2000).

Chapter 2

Contributions of Aromatic Pairs to the Folding and Stability of Long-lived Human gamma D-Crystallin [1]

[1] **Reprinted and modified with permission from**: Kong, F. and King, J. (2011) Contributions of aromatic pairs to the folding and stability of long-lived human γD-crystallin. *Protein Science*, 20 (3): 513–528.

A. Abstract

Human γD-crystallin (HγD-Crys) is a highly stable protein that remains folded in the eye lens for the majority of an individual's lifetime. HγD-Crys exhibits two homologous crystallin domains, each containing two Greek key motifs with eight b-strands. Six aromatic pairs (four Tyr/Tyr, one Tyr/Phe and one Phe/Phe) are present in the b-hairpin sequences of the Greek keys. Ultraviolet damage to the aromatic residues in lens crystallins may contribute to the genesis of cataract. Mutant proteins with these aromatic residues substituted with alanines were constructed and expressed in E. coli. All mutant proteins except F115A and F117A had lower thermal stability than the WT protein. In equilibrium experiments in guanidine hydrochloride (GuHCl), all mutant proteins had lower thermodynamic stability than the WT protein. N-terminal domain (N-td) substitutions shifted the N-td transition to lower GuHCl concentration, but the C-terminal domain (C-td) transition remained unaffected. C-td substitutions led to a more cooperative unfolding/refolding process, with both the N-td and C-td transitions shifted to lower GuHCl concentration. The aromatic pairs conserved for each Greek key motif (Greek key pairs) had larger contributions to both thermal stability and thermodynamic stability than the other pairs. Aromatic-aromatic interaction was estimated as 1.5–2.0 kcal/mol. In kinetic experiments, N-td substitutions accelerated the early phase of unfolding, while C-td substitutions accelerated the late phase, suggesting independent domain unfolding. Only substitutions of the second Greek key pair of each crystallin domain slowed refolding. The second Greek keys may provide nucleation sites during the folding of the double-Greek-key crystallin domains.

B. Introduction

Crystallins make up over 90% of the eye lens proteins(Bron *et al.* 2000). α-Crystallins are small heat shock proteins (sHsp) with chaperone functions (Horwitz 2000; Horwitz 2003), while βγ-crystallins are the major structural proteins (Bloemendal *et al.* 2004). βγ-crystallins are a family of proteins exhibiting two homologous crystallin domains. Each domain contains two Greek key motifs forming a β-sandwich of eight intercalated β- strands (Basak *et al.* 2003) (Figure 2-1). β-Crystallins form dimers and higher-order oligomers, while γ-crystallins are monomeric (Bloemendal *et al.* 2004). βγ-Crystallin -like proteins are also found in a number of lower organisms lacking the lens tissue (Wistow *et al.* 1985; Wistow 1990; Shimeld *et al.* 2005). Lens βγ-crystallins together with these βγ-crystallin-like proteins form the βγ-crystallin superfamily.

Lens epithelial cells lose their organelles upon differentiation into the mature lens fiber cells, and cannot synthesize crystallins afterwards (Bassnett 2002). The central region of the eye lens, the nucleus, is formed *in utero*. Human γ-crystallins, particularly γC- and γD-crystallins, are enriched in the lens nucleus, representing some of the oldest molecules in the human body (Siezen *et al.* 1987; Brakenhoff *et al.* 1990; Lampi *et al.* 1997). To maintain lens transparency, the crystallins must remain soluble and folded throughout the majority of an individual's lifetime. In fact, human γD-crystallin (HγD-Crys) is resistant to denaturation by up to 8 *M* urea (Kosinski-Collins and King 2003). The half time of the initial unfolding step was about 19 years (Mills-Henry 2007), and the thermal denaturation midpoint was 83.8° C (Flaugh *et al.* 2006). The complex topology of the Greek key motifs may contribute to the high stability of βγ- crystallins (MacDonald *et al.* 2005).

Like all βγ-crystallins, the 173-amino-acid HγD-Crys has two homologous crystallin domains [Figure 2-1 (A)]: N-terminal domain (N-td, residues 1–82), and C-terminal domain (C-td, residues 88–173) and four Greek key motifs (named Greek key 1 through 4 from N- to C-terminus) (Basak *et al.* 2003). Each crystallin domain has two highly conserved tryptophans. The fluorescence spectrum of denatured HγD-Crys becomes higher and red shifted compared to that of the native state (Kosinski-Collins *et al.* 2004; Chen *et al.* 2006; Chen *et al.* 2008; Chen *et al.* 2009). This unusual native-state fluorescence quenching property of HγD-Crys provides a sensitive reporter of the transition between the native and unfolded states.

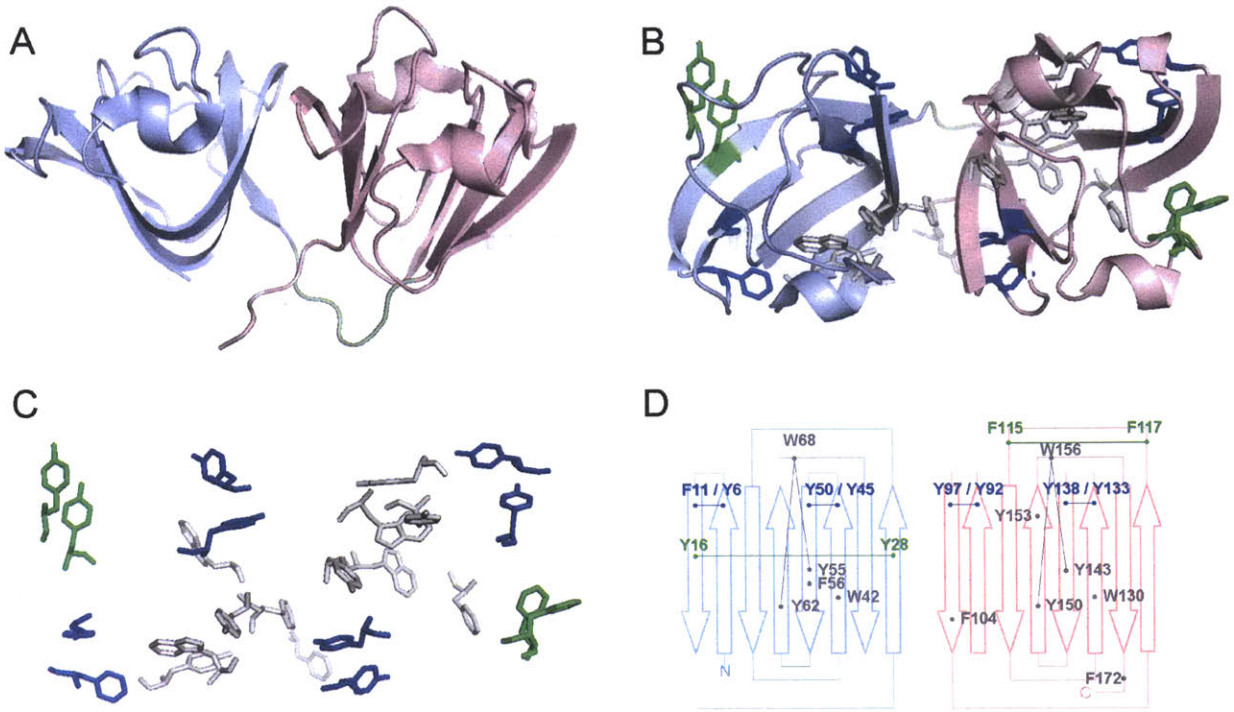


Figure 2-1 Crystal structure of HγD-Crys (1HK0).

Three-dimensional structures were prepared by PyMOL (DeLano Scientific). (A) "Side view" of HγD-Crys, showing the two crystallin domains. Pale blue: N-td; pink: C-td; light green: linker. (B) "Top view" of HγD-Crys. All aromatic residues are shown in stick representation. Blue: four β-hairpin aromatic pairs (the "Greek key pairs"); green: two additional β-hairpin aromatic pairs (the "non-Greek-key pairs"); gray: all other aromatic residues. (C) The aromatic residues in HγD-Crys in the same orientation and color scheme as in panel B. the cartoon representation of the nonaromatic residues are removed for a clearer view. (D) Topology diagram of HγD-Crys and the locations of the aromatic residues. Color scheme is the same as in B. Block arrows represent β-strands. The six aromatic pairs and the two Tyr-Trp-Tyr clusters are indicated by thin lines connecting dots. From (Basak *et al.* 2003)

During protein expression in *E. coli* cytoplasm, a substantial proportion of recombinant HγD-Crys successfully folded into the soluble native state. Purified HγD-Crys also readily refolds into a native-like state *in vitro* at 37°C when diluted from 5 *M* guanidine hydrochloride (GuHCl) to ~1 *M*, without any assistance of chaperones. Below 1 *M*, however, partially folded HγD-Crys forms high-molecular-weight aggregates (Kosinski-Collins and King 2003). In equilibrium experiments in GuHCl, an intermediate with the N-td unfolded and C-td folded was identified in the unfolding/refolding pathway. Mutations of the hydrophobic core, polar pairs interacting across the domain interface, and N-td congenital mutations of HγD-Crys affected the thermodynamic stability and the refolding rate of the N-td but not the C- td (Flaugh *et al.* 2005; Flaugh *et al.* 2005; Flaugh *et al.* 2006; Moreau and King 2009). Kinetic experiments revealed a sequential refolding pathway where the C-td folds first, followed by the N-td (Kosinski-Collins *et al.* 2004; Mills *et al.* 2007). Computational studies using molecular dynamics supported this pathway (Das *et al.* 2009). A more detailed pathway was proposed by Flaugh *et al.*, where the four Greek keys unfold successively from the N- to the C-terminus, and refold successively in the opposite direction (Flaugh *et al.* 2006). However, experimental evidence regarding the unfolding/refolding pathway within each double-Greek-key crystallin domain of HγD-Crys remains scarce.

HγD-Crys contains a high percentage of aromatic residues distributed throughout the protein: 14 Tyr, 6 Phe, and 4 Trp. 18 of these 24 residues have aromatic partners within 5 Å [Figure 2-1 (B–D]. These aromatic residues cluster into several structural elements. The "tyrosine corner" is a classic feature of the Greek key motifs, bridging two β-strands by a hydrogen bond between the tyrosine hydroxyl group and a backbone carboxyl group (Hemmingsen *et al.* 1994). The hydroxyl groups were suggested to be essential for stability, but not for nucleating the folding of Greek key proteins (Hamill *et al.* 2000). HγD-Crys has two homologous tyrosine corners, Y62 in the N-td and Y150 in the C-td.

Besides the tyrosine corners, four homologous aromatic pairs with partners five residues apart in the primary sequence are present (three Tyr/Tyr and one Tyr/Phe, N-td: Y6/F11, Y45/Y50, C-td: Y92/Y97, Y133/Y138, termed "Greek key pairs"). They are conserved at the homologous β-hairpins of the four Greek key motifs. Two additional aromatic pairs (one Tyr/Tyr and one Phe/Phe; N-td: Y16/Y28, C-td: F115/F117, termed "non-Greek-key pairs") are found in other β-hairpins. Two three-residue Tyr-Trp-Tyr clusters are also present, one in each crystallin

domain. Epidemiological studies of humans and laboratory experiments on animals supported ultraviolet (UV) radiation being a causative factor for cataract, a major cause of blindness in the world (McCarty and Taylor 2002). Trp, Tyr, and Phe are the main UV absorbers in proteins and are susceptible to direct and indirect UV damage (Grossweiner 1984; Davies and Truscott 2001). UV damage to these aromatic residues may impair the folding and stability of lens crystallins, such as HγD-Crys, and contribute to the genesis of cataract.

Aromatic molecules can interact through partial charges on the δ(–) ring faces and δ(+) ring edges (Meyer *et al.* 2003). Aromatic residues tend to cluster in proteins, and prefer a distance of ~5.5 Å between phenyl ring centroids (Burley and Petsko 1985). Two frequently observed orientations are perpendicular (Burley and Petsko 1985) and "parallel displaced stacking" (McGaughey *et al.* 1998). Aromatic -aromatic interaction energy was estimated to be about −0.6 to −1.3 kcal/mol (Burley and Petsko 1985). A rich body of experimental evidence suggests that aromatic-aromatic interactions play important roles in stability and/or folding of diverse proteins (Hecht *et al.* 1984; Serrano *et al.* 1991; Kannan and Vishveshwara 2000; Betts *et al.* 2004; Simkovsky and King 2006), especially β-sheet proteins (Smith and Regan 1995; Hillier *et al.* 1998; Rodriguez *et al.* 2000; MacDonald *et al.* 2005; Rea *et al.* 2008). Aromatic interactions have also been studied in peptides, and employed in *de novo* peptide design (Waters 2004). As in proteins, β-hairpin peptides could be stabilized by aromatic pairs and clusters (Yao *et al.* 1994; Griffiths-Jones and Searle 2000; Wu *et al.* 2010).

The β-hairpin aromatic pairs are a prominent feature of the Greek keys in βγ-crystallins. Assessing the contributions of these aromatic pairs to the folding and stability of HγD-Crys is important in advancing our understanding of the folding and stability of this and other Greek key and β-sheet proteins, as well as other proteins containing aromatic pairs. In this study, we used site-directed mutagenesis, equilibrium and kinetic experiments, and both thermal and GuHCl denaturation to dissect the roles of the six β-hairpin aromatic pairs in the folding and stability of HγD-Crys.

C. Materials and Methods

1. Site-specific Mutagenesis

N-terminally 6xHis tagged WT HγD-Crys pQE.1 plasmid (Qiagen) was amplified by PCR-based mutagenesis using primers encoding the desired mutations. The primers were designed from PrimerX <http://www.bioinformatics.org/primerx/> and ordered from Integrated DNA Technologies. The mutant plasmids were amplified in *E. coli* Nova Blue, purified by miniprep and sequenced at Massachusetts General Hospital to confirm the correct mutations. Double mutants were constructed using the confirmed single mutant plasmids as templates in a two-step procedure.

2. Expression and Purification

E. coli M15(pREP4) were transformed with the purified plasmids, cultured on shakers at 37°C until OD_{600} reached ~1.0, then induced with 1 m*M* IPTG, and allowed for expression for 3 hr. The cells were pelleted and resuspended in 300 m*M* NaCl, 50 m*M* NaH_2PO_4/Na_2HPO_4, 18 m*M* imidazole, pH 8.0. The cells were lysed by sonication and centrifuged at 20,000*g* for 30 min. The supernatant was applied to a Ni-NTA column (Qiagen) by an FPLC system (GE Healthcare), eluted with a linear gradient of imidazole from 18 to 250 m*M* at 4°C. Purity was ensured by SDS-PAGE. Fractions exceeding about 95% purity were pooled, dialyzed with 10 m*M* ammonium acetate, pH 7.0, at 4°C, and concentrated with Amicon centrifuge filters with 10 kDa MWCO. Protein concentrations were measured by A_{280} in 5.5 *M* GuHCl, using extinction coefficient calculated from Expasy <http://ca.expasy.org/tools/protparam.html.>

3. Structural Assessment by CD

Far-UV CD spectra were obtained with an AVIV Model 202 CD spectrophotometer. The samples were prepared as follows: 100 μg/mL protein in 10 m*M* NaH_2PO_4/Na_2HPO_4, pH 7.0, 0.2 μm filtered, degassed for >1 hr. Each protein was scanned at 190–260 nm, 5 sec averaging time of signal, at 25°C. Triplicates of independent experiments were done for each protein.

4. Thermal Denaturation

Sample preparation and the instrument were the same as in the Structural Assessment section above. Temperature was raised from 25 to 90°C in 1°C steps, equilibrated for 1 min at each temperature, CD signal at 218 nm was measured for every °C, 10 sec averaging time. Triplicates of independent experiments were done for each protein. The data were normalized to the maxima and minima, transition midpoints were determined by the intersections of the traces and the 50% lines.

5. Equilibrium Unfolding/Refolding

The buffer system was as follows: 100 m*M* NaH_2PO_4/Na_2HPO_4, 5 m*M* DTT, 1 m*M* EDTA, pH 7.0, 0.2 μm filtered. Unfolding samples were prepared by diluting native protein into the buffer above, with final protein concentration 10 μg/mL, as well as 38 different concentrations of GuHCl in the range of 0–5.5 *M*. The samples were incubated at 37°C for 24 hr. Refolding samples were prepared similarly, except that the native protein was first denatured at protein concentration of 100 μg/mL, in 5.5 *M* GuHCl at 37°C for 5 hr, before being diluted into a 0.55–5.5 *M* GuHCl gradient.

6. Kinetic Unfolding/Refolding

The buffer system used was the same as in the Equilibrium Unfolding/Refolding section above. For unfolding, buffer with GuHCl at a final concentration of 5.5 *M* was incubated in a cuvette at 37°C (or 18°C), and stirred with a mini spin bar. Fluorescence was monitored by a Hitachi F-4500 fluorometer with excitation wavelength 295 nm and emission wavelength 350 nm. Native protein at 100 μg/mL (10×) was syringe injected through a small injection port into the cuvette. The final protein concentration was 10 μg/mL. Fluorescence was monitored until a plateau was reached. Triplicates of independent experiments were done for each protein. Refolding experiments were similar to unfolding, except that the 10× protein was incubated with 5.5 *M* GuHCl at 37°C for 5 hr before being injected into the buffer containing 0.5 *M* GuHCl. Final GuHCl concentration was 1.0 *M*.

D. Results

1. Conservation of the Aromatic Pairs in βγ-Crystallins

Sequence comparisons between the lens βγ-crystallins and the βγ-crystallin-like relatives in lower organisms were previously conducted for protein S (Wistow *et al.* 1985), spherulin 3a (Wistow 1990), and Ciona crystallin (Shimeld *et al.* 2005). In these three βγ-crystallin-like proteins, when Phe or Tyr was considered to be conserved, the Greek key aromatic pairs were maintained, while the non-Greek-key pairs were not. The Greek key pair consensus Y/FXXXXY/FXG, was used for defining the βγ-crystallin superfamily (Aravind *et al.* 2009). To evaluate the conservation of the aromatic pairs in lens βγ-crystallins, we performed a sequence alignment with 43 β-crystallins and 36 γ-crystallins from human and seven other vertebrates, using ClustalW2 (Larkin *et al.* 2007) (Figure 2-9, see Supplementary Material at the end of this chapter). Residues homologous to the six aromatic pairs in HγD-Crys were examined for their conservation.

When either Phe or Tyr was considered to be conserved in each position, conservation are >80% for the Greek key pairs, and ~20–70% for the non-Greek-key pairs in the 79 βγ-crystallins selected (Table 2-1). For the two less conserved non-Greek-key pairs, most γ-crystallins, but none of the β-crystallins, maintain the pair Y16/Y28; all of the γD-, γE-, γF-crystallins, but almost none of the other γ-, or β-crystallins, maintain both residues of the pair F115/F117 (Figure 2-9). For the positions not occupied by Tyr or Phe, there was no obvious pattern for amino acid usage. The conservation of the Greek key pairs in the lens βγ-crystallins and βγ-crystallin-like relatives, in addition to the well-defined primary sequence and three-dimensional structure, suggest that for the functions these aromatic pairs serve in βγ-crystallins, the Greek key pairs are much more important than the non-Greek-key pairs.

Table 2-1 Conservation of Aromatic Pairs in Selected βγ-Crystallin Sequences.

Position [a]	Usage of Tyr or Phe [b]
Y6	82%
F11	100%
Y16	44%
Y28	44%
Y45	100%
Y50	91%
Y92	100%
Y97	90%
F115	68%
F117	19%
Y133	91%
Y138	99%

[a] Amino acid positions are as in HγD-Crys.

[b] Percentages were calculated from the sequence alignment result shown in Figure 2-9.

2. Mutant Protein Construction, Expression, and Purification

To assess experimentally the roles of the six aromatic pairs in HγD-Crys, 12 alanine single mutants of each of the nine Tyr and three Phe residues were constructed by site-directed mutagenesis. Each mutant coding sequence had an N-terminal 6xHis tag. The recombinant mutant proteins were expressed in *E. coli* at 37°C and purified using a Ni-NTA column. The purification protocol was originally optimized for the wildtype (WT) HγD-Crys, but was reasonably efficient with the mutant proteins. Soluble protein yields for 10 out of the 12 single mutant proteins were comparable to the WT protein. Both the soluble and insoluble fractions of the lysate contained considerable amounts of the recombinant proteins by SDS-PAGE. However, two single mutant proteins Y133A and Y138A of Greek key 4 remained largely in the insoluble inclusion body fractions, and the soluble fractions yielded less than 1/10 of the WT amount. Expression at 18°C for longer periods of time failed to improve the soluble yields of these two mutant proteins. Nonetheless, sufficient amounts of these two single mutant proteins were recovered in soluble form for the subsequent experiments. Three double mutants were also constructed, and used for double mutant cycle analysis described in a later section. All other experiments involved only the 12 single mutant proteins.

3. Structural Assessment

Far-UV circular dichroism (CD) was used for gross secondary structural assessment of the purified WT and mutant HγD-Crys. Overall, the spectra of the mutant proteins were very similar to that of the WT protein, displaying a major peak at 218 nm with similar intensity (Figure 2-2). This indicated that the mutant proteins maintained the folded β-sheet structures very similar to the WT protein. A small inflection at 208 nm may reflect a minimal amount of helical content for the WT protein. This feature was also retained for all the mutant proteins, although it appeared less obvious for some of the mutant proteins (Figure 2-2).

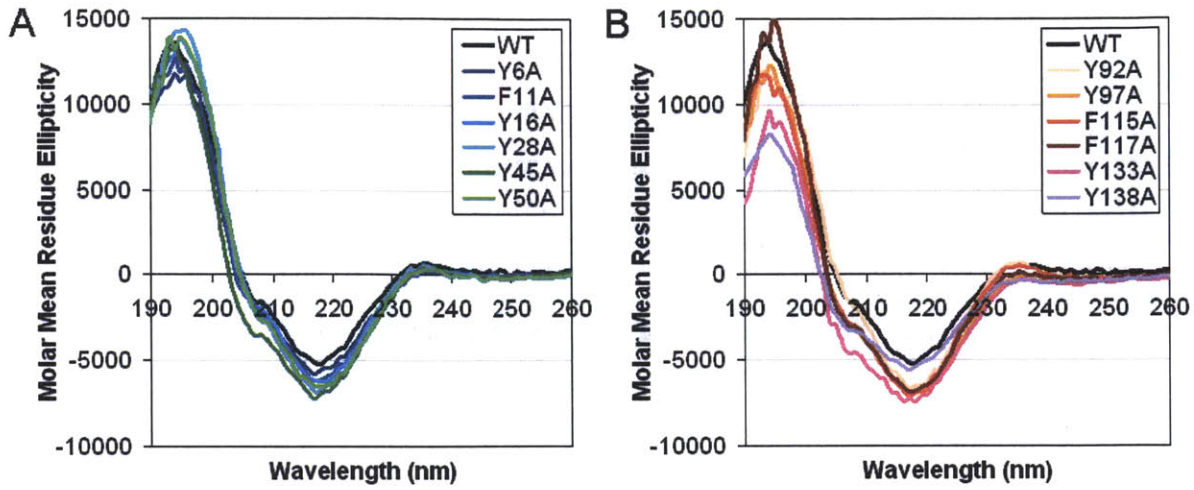


Figure 2-2 Circular Dichroism for WT and Mutant HγD-Crys.

The samples were prepared as: 100 μg/mL protein in buffer containing 10 m*M* NaH_2PO_4/Na_2HPO_4, pH 7.0. Each protein was scanned from 190–260 nm, at 25°C. Molar mean residue ellipticity was calculated from the raw signal. Averages of independent triplicates for each protein are shown. (A) N-td mutant proteins. (B) C-td mutant proteins.

4. Thermal Denaturation

As an initial test for stability, thermal denaturation experiments were performed on the WT and mutant HγD-Crys. Ellipticity at 218 nm was monitored while the samples were heated slowly from 25 to 95°C. Wavelength 218 nm is the characteristic wavelength of antiparallel β-sheet structure, and also the major peak wavelength found in the initial structural assessment.

The results showed that for the WT and all mutant proteins, ellipticity at 218 nm became less negative as temperature increased (Figure 2-3). At the end of each experiment, the protein solution became turbid from particulate aggregates. The decrease in ellipticity signal was therefore likely due to both the loss of secondary structure and aggregate formation that increased light scattering. All WT and mutant HγD-Crys had similar two-state thermal denaturation traces and similar degrees of cooperativity, with no obvious intermediates. The transition midpoint for the WT protein was 80.9°C. For the mutant proteins, the transition midpoints varied from 71.3 to 81.1°C (Figure 2-3; Table 2-2). Mutant proteins F115A and F117A had transition midpoints indistinguishable from the WT protein. The other 10 mutant proteins had transition midpoints lower than the WT protein to different degrees, indicating that the corresponding five aromatic pairs but not F115/F117 contributed to the thermal stability of HγD-Crys.

Substitutions of the pairs Y6/F11, Y45/Y50, and Y133/Y138 had the largest effects on thermal stability: the transition midpoints ranged from 71.3 to 73.4°C, substantially lower than 80.9°C of the WT protein (Table 2-2). Within each crystallin domain, substitutions of the Greek key pairs (N-td: Y6/F11, Y45/Y50; C-td: Y92/Y97, Y133/Y138) had larger effects than substitutions of the non-Greek-key pairs (N-td: Y16/Y28; C-td: F115/F117), indicating that the Greek key pairs are especially important for the thermal stability of the proteins. For the first Greek key pair in each crystallin domain, substitution of the first residue (N-td: Y6; C-td: Y92) had a larger effect than substitution of the second residue (N-td: F11; C-td: Y97).

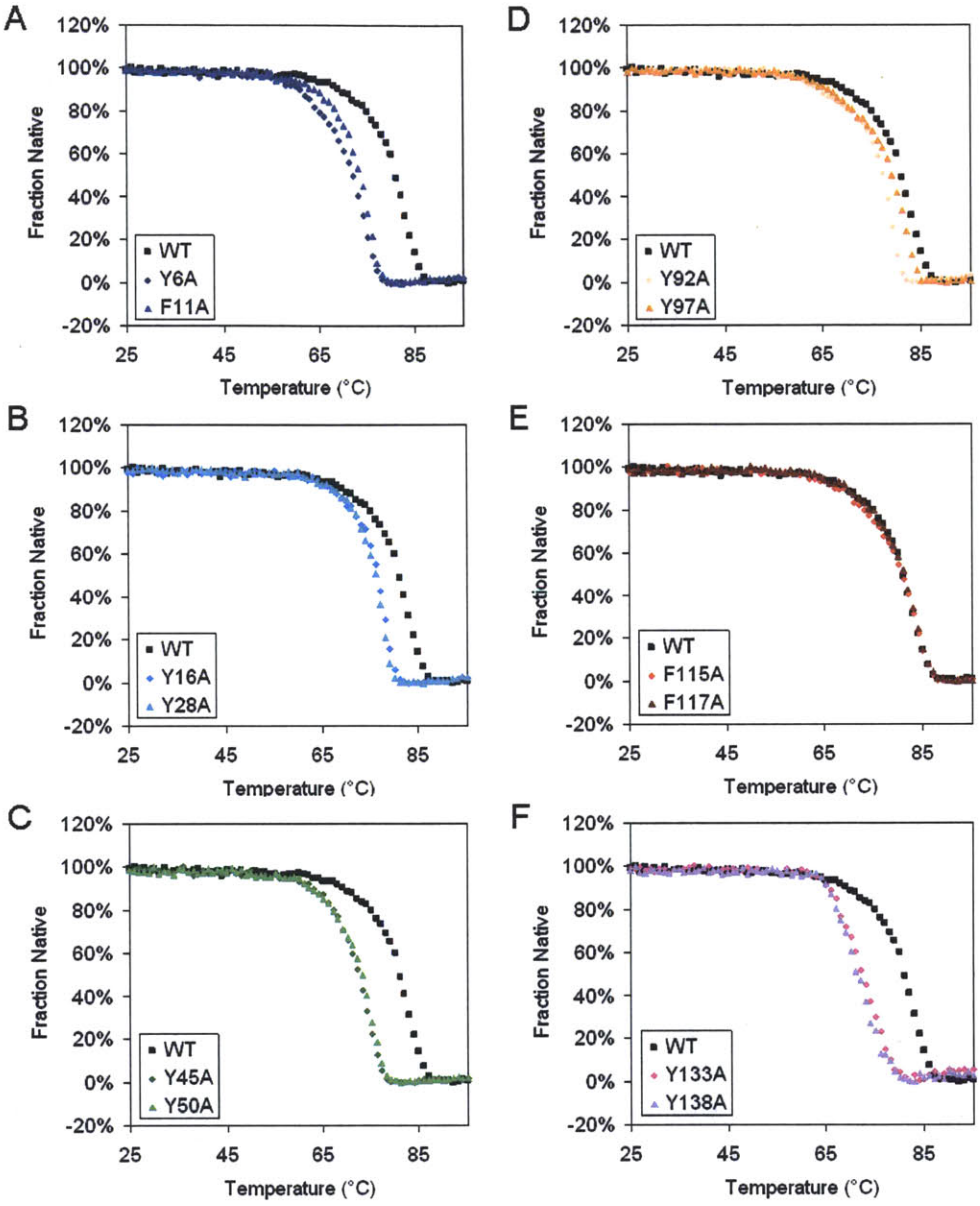


Figure 2-3 Thermal Denaturation for WT and Mutant HγD-Crys.

The samples were prepared as: 100 μg/mL purified protein in 10 m*M* NaH_2PO_4/Na_2HPO_4, pH 7.0. Temperature was slowly raised from 25 to 90°C in 1°C steps, and equilibrated for 1 min at each temperature; CD at 218 nm was measured for every °C. Averages of at least two independent repeats for each protein are shown. The data were normalized to the highest and lowest data points. (A) Y6A, F11A. (B) Y16A, Y28A. (C) Y45A, Y50A. (D) Y92A, Y97A. (E) F115A, F117A. (F) Y133A, Y138A.

Table 2-2 Thermal Denaturation and Equilibrium Unfolding Parameters for WT and Mutant HγD-Crys.

Protein	Thermal Denaturation	Equilibrium Transition 1			Equilibrium Transition 2		
	T_m [a]	C_m [b]	Apparent *m* value [c]	Apparent $\Delta\Delta G_0$ [d]	C_m [b]	Apparent *m* value [c]	Apparent $\Delta\Delta G_0$ [d]
N-td							
WT	80.9±0.3	2.21±0.05	4.6±0.4	0.0±0.8	2.99±0.03	2.9±0.3	0.0±1.3
Y6A	72.0±0.1	1.12±0.03	4.5±0.4	-4.8±0.3	3.02±0.03	2.8±0.1	-0.3±0.5
F11A	73.4±0.2	1.47±0.05	4.2±0.3	-3.7±0.3	3.09±0.09	2.6±0.8	-1.1±2.2
Y16A	76.4±0.1	1.85±0.03	4.4±0.1	-1.7±0.2	3.00±0.06	2.8±0.3	-0.8±0.7
Y28A	76.0±0.0	1.77±0.05	4.1±0.8	-2.5±1.3	2.91±0.09	3.0±0.4	-0.2±1.1
Y45A	72.2±0.1	1.13±0.03	3.8±0.2	-5.5±0.2	2.99±0.02	2.9±0.3	-0.3±0.8
Y50A	72.9±0.1	1.20±0.02	4.4±0.5	-4.6±0.6	2.94±0.06	2.8±0.5	-0.9±1.3
C-td							
WT	80.9±0.3	2.81±0.02	0.6±0.0	0.0±0.0	-	-	-
Y92A	77.0±0.2	1.80±0.04	3.0±0.5	3.6±0.7	-	-	-
Y97A	79.0±0.2	2.21±0.04	1.6±0.4	1.7±0.9	-	-	-
F115A	80.6±0.1	2.50±0.02	1.1±0.1	1.0±0.2	-	-	-
F117A	81.1±0.0	2.58±0.07	1.0±0.1	0.9±0.2	-	-	-
Y133A	72.1±0.2	1.93±0.01	1.7±0.1	1.4±0.3	-	-	-
Y138A	71.3±0.6	1.98±0.01	2.2±0.1	2.6±0.2	-	-	-

[a] Thermal denaturation transition midpoints in units of °C.

[b] Equilibrium unfolding transition midpoints in units of M GuHCl.

[c] Apparent *m* values in units of $kcal \cdot mol^{-1} \cdot M^{-1}$.

[d] Difference of free energy change extrapolated to 0 M GuHCl, as compared to the WT protein, in units of $kcal \cdot mol^{-1}$. A negative number means mutant is less stable than the WT.

"-" not applicable. Errors are standard deviations of three trials.

5. Equilibrium Unfolding/Refolding

To assess the effects of the Tyr or Phe substitutions on thermodynamic stability of HγD-Crys, equilibrium unfolding/refolding experiments were performed on the WT and mutant proteins. The samples were prepared in GuHCl at pH 7.0, and incubated at 37°C for 24 hr. The fluorescence ratio 360/320 was used to assess the unfolded/native state. Upon refolding at very low GuHCl concentration, WT HγD-Crys aggregation competes with refolding, and interferes with the fluorescence signal (Kosinski-Collins and King 2003). This behavior was also observed for all 12 single mutant proteins. At 1.0 *M* or higher GuHCl concentration, the equilibrium refolding traces overlapped very well with the unfolding traces. Therefore, the refolding traces were used mainly to ensure that the samples were fully equilibrated. The unfolding traces are used for the following analysis.

For WT HγD-Crys, a three-state equilibrium model best fit the equilibrium trace (Figure 2-4): black traces in each panel. The bottom baseline at low GuHCl concentration corresponded to the native state, while the top baseline at high GuHCl concentration corresponded to the unfolded state. The intermediate at fluorescence ratio ~1.0 (a small kink in the trace) was previously identified as a single-domain-folded species with the N-td unfolded and C-td folded. The N-td transition midpoint was 2.21 *M* GuHCl, and the C-td transition midpoint was 2.99 *M* (Table 2-2).

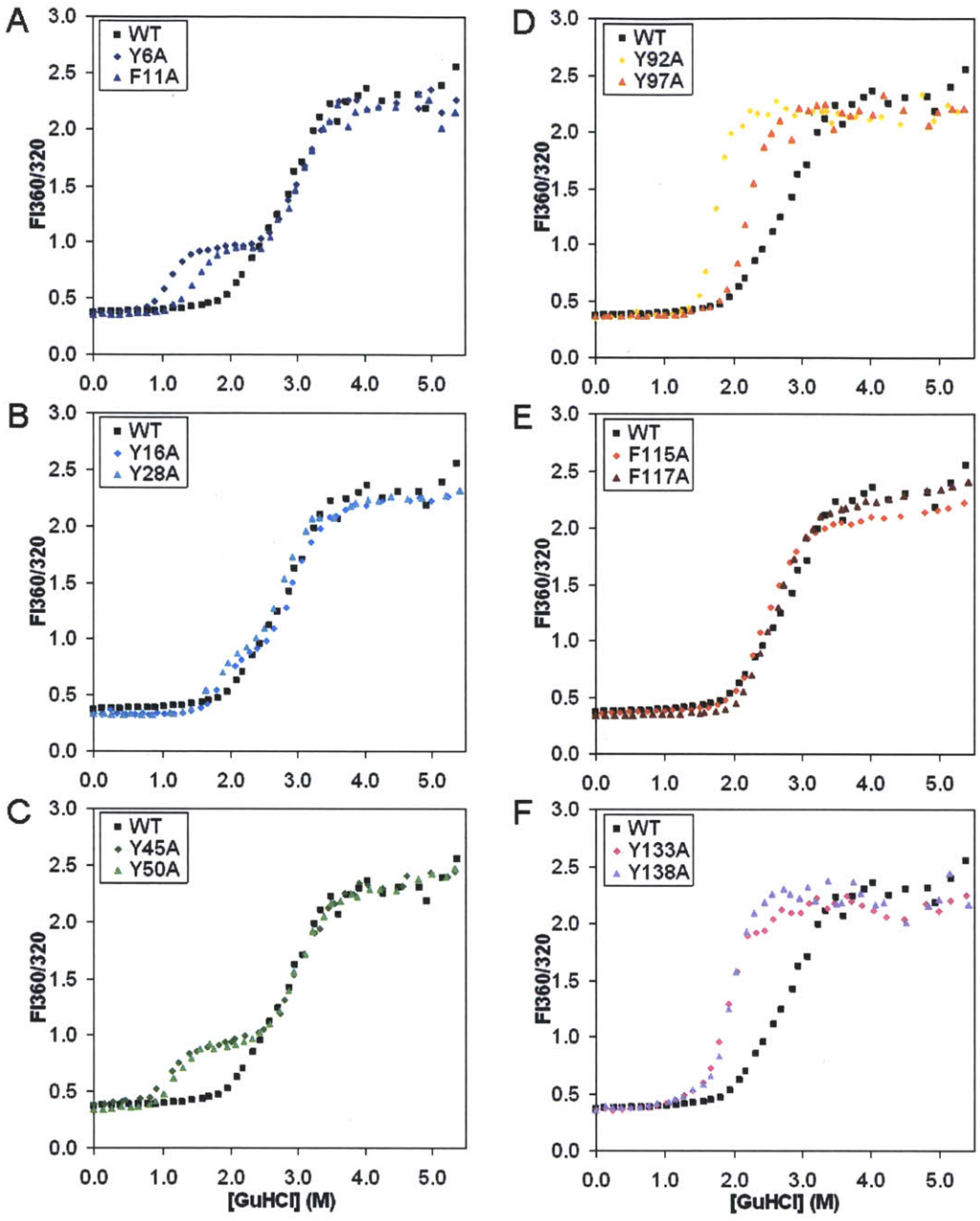


Figure 2-4 Equilibrium Unfolding for WT and Mutant HγD-Crys.

The samples were prepared as: 10 μg/mL protein, in 100 m*M* NaH_2PO_4/Na_2HPO_4, 5 m*M* DTT, 1 m*M* EDTA, in varying [GuHCl] at pH 7.0, 37°C, incubated for 24 hr. Fluorescence spectra were measured at excitation wavelength 295 nm. Ratios of fluorescence intensity of 360/320 are extracted and plotted against [GuHCl]. Representative traces of independent triplicates for each protein are shown. (A) Y6A, F11A. (B) Y16A, Y28A. (C) Y45A, Y50A. (D) Y92A, Y97A. (E) F115A, F117A. (F) Y133A, Y138A.

Each of the mutant proteins had a different equilibrium profile compared to the WT (Figure 2-4). From the equilibrium traces, parameters including transition midpoints, apparent *m* values and $\Delta\Delta G$'s compared to the WT protein were calculated (Table 2-2). Figure 2-4 (A–C) shows that compared to the WT protein, each of the N-td mutant proteins had the N-td transition shifted to lower GuHCl concentration, indicating a destabilized N-td. The C-td transitions remained unaffected. Thus, an increased population of the single-domain-folded intermediate was revealed. A three-state equilibrium model was best fit to these data. Two transition midpoints were calculated: the N-td transition midpoints ranged from 1.12 to 1.85 *M*, significantly lower than 2.21 *M* for the WT protein. The C-td transition midpoints ranged from 2.91 to 3.09 *M*, closely centered around 2.99 *M* for the WT protein. Figure 2-4 (D–F) showed that compared to the WT protein, C-td mutant proteins had both the N-td and C-td destabilized. Also, the N-td and C-td transitions were closer and inseparable, indicating a more cooperative unfolding/refolding process. A two-state equilibrium model was best fit to these data. One transition midpoint was calculated. This ranged from 1.80 to 2.58 *M*, significantly lower than 2.81 *M* of the WT protein, if the WT transition was also considered two-state. These results clearly showed that all the aromatic pairs were important for the thermodynamic stability of HγD-Crys.

Comparing $\Delta\Delta G$'s and transition midpoints yielded similar results for the N-td mutant proteins. However, for the C-td mutant proteins, because the N-td and C-td transitions partially overlap, the apparent $\Delta\Delta G$ values probably reflect the level of separation of the two transitions, rather than the genuine thermodynamic stability of the mutant proteins. Therefore, our analysis focuses on the transition midpoints. Comparing the mutant proteins carrying substitutions within the same crystallin domain, two patterns regarding the extents of destabilization emerged, and they were consistent with the patterns of thermal denaturation. (1) The Greek key pairs (N-td: Y6/F11, Y45/Y50; C-td: Y92/Y97, Y133/Y138) had larger contributions to the thermodynamic stability than the non-Greek-key pairs (N-td: Y16/Y28, C-td: F115/F117). (2) For the first Greek key pair in each crystallin domain, the first residue (N-td: Y6, C-td: Y92) had a larger contribution than the second residue (N-td: F11, C-td: Y97).

6. Kinetic Unfolding/Refolding at 37°C

Decreased thermodynamic stability of the mutant HγD-Crys pointed to changes in unfolding and/or refolding rates. To further distinguish the effects on unfolding vs. refolding, kinetic unfolding and refolding experiments were performed on the WT and mutant proteins. Initial kinetic experiments were performed at 37°C, with one final GuHCl concentration each for unfolding and refolding. For unfolding, native protein was mixed into 5.5 *M* GuHCl. For refolding, fully denatured protein in 5.5 *M* GuHCl was diluted to a final concentration of 1.0 *M* GuHCl. One molar was chosen because at concentration below 1.0 *M*, refolding HγD-Crys species formed aggregates that interfered with the fluorescence signal. Fluorescence at 350 nm was monitored for 1 hr. The results showed that the WT protein exhibited an early/fast and a late/slow phase, corresponding to the two crystallin domains as previously suggested [Figure 2-5 (A) and 6(A)] (Kosinski-Collins *et al.* 2004; Flaugh *et al.* 2006). The fast phase was not resolved by the instrument at 37°C, so our initial analysis focuses on the slow phase.

For unfolding, the N-td mutant proteins had no observable difference compared to the WT protein [Figure 2-5 (B–D)]. In contrast, C-td mutant proteins unfolded significantly faster than the WT protein [Figure 2-5 (E–G)]. In particular, mutant proteins of the Greek key pairs unfolded extremely fast, with the fluorescence signals reaching their final levels in less than 1 min, compared to about 15 min for the WT protein [Figure 2-5 (E,G)]. Mutant proteins of the non-Greek-key pair unfolded slightly faster than the WT protein [Figure 2-5 (F)]. These results showed that the C-td aromatic pairs made important contributions to the kinetic stability of HγD-Crys, while the N-td aromatic pairs were less important.

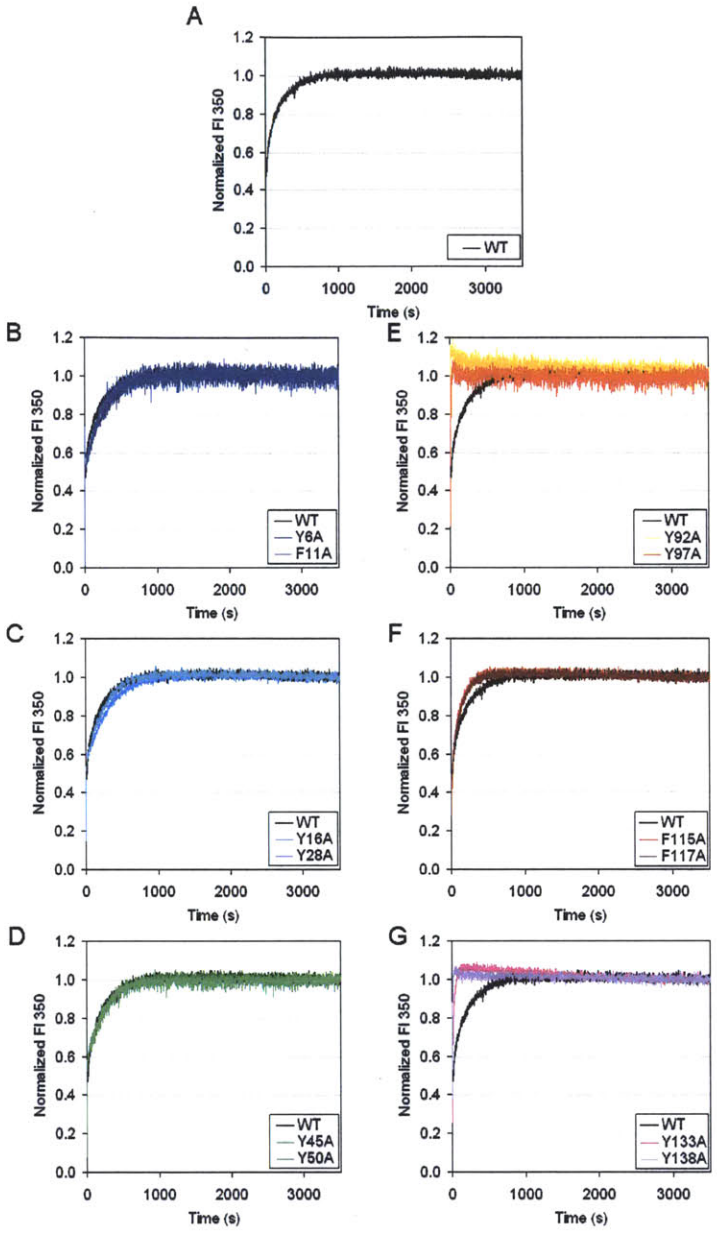


Figure 2-5 Kinetic Unfolding for WT and Mutant HγD-Crys at 37°C.

The samples were prepared in 100 m*M* NaH_2PO_4/Na_2HPO_4, 5 m*M* DTT, 1 m*M* EDTA, pH 7.0, at 37°C. Unfolding was initiated by mixing native protein into 5.5 *M* GuHCl. Fluorescence at 350 nm was monitored. Fluorescence intensity was normalized to the native and unfolded protein controls. Representative traces of independent triplicates for each protein are shown. (A) WT. (B) Y6A, F11A. (C) Y16A, Y28A. (D) Y45A, Y50A. (E) Y92A, Y97A. (F) F115A, F117A. (G) Y133A, Y138A.

For refolding, mutant proteins of the second Greek key pair in each crystallin domain refolded significantly slower than the WT protein [Figure 2-6 (D,G)], but affected different phases. For Y45A, Y50A proteins, the initial 70% of the fluorescence changes occurred at similar rates as for the WT protein, but the later 30% occurred at slower rates. This result was similar to the refolding kinetics of the interface mutant HγD-Crys (Flaugh *et al.* 2005; Flaugh *et al.* 2005; Flaugh *et al.* 2006). Unlike Y45A and Y50A proteins, the refolding of Y133A and Y138A proteins were slower than the WT protein from the beginning of the reaction. From these results, it appeared that the pair Y133/Y138 was important for the early phase in the refolding process, while the pair Y45/Y50 was important for the late phase. Except for these two pairs, all other mutant proteins had no observable differences compared to the WT protein [Figure 2-6 (B,C,E,F)]. These results indicated that the second Greek key pair in each crystallin domain contributed to the refolding of HγD-Crys, while the other pairs were less important.

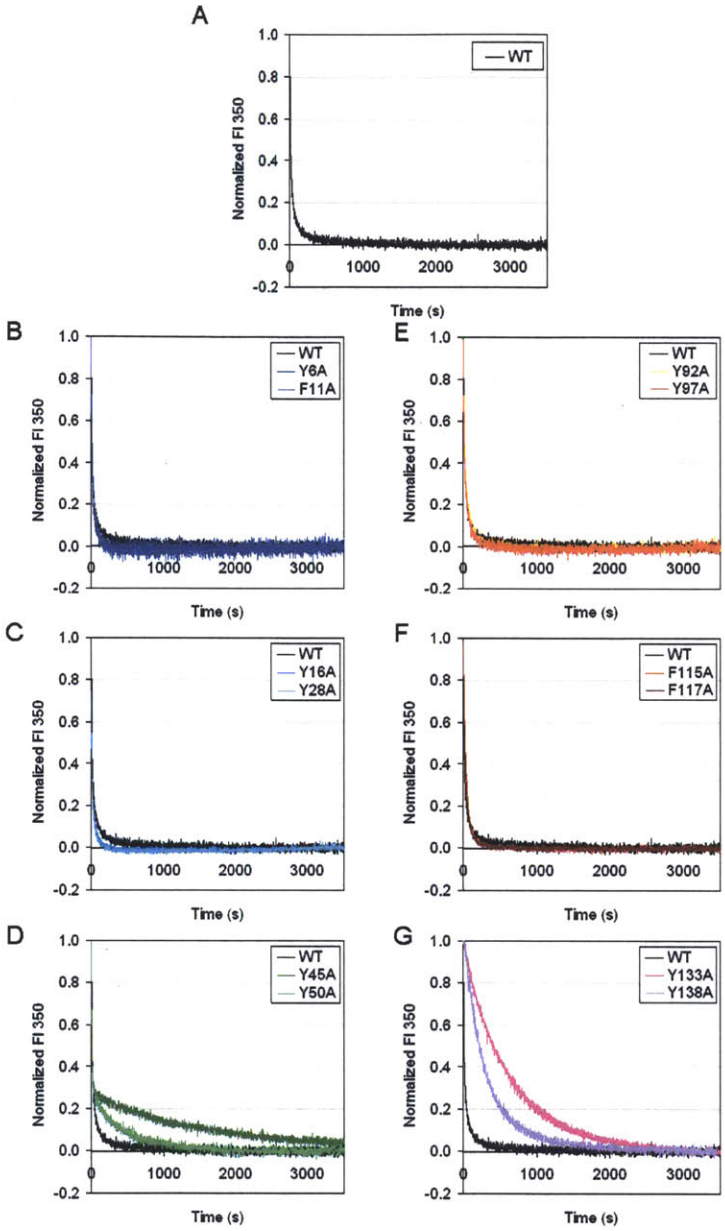


Figure 2-6 Kinetic Refolding for WT and Mutant HγD-Crys at 37°C.

The samples were prepared in 100 m*M* NaH_2PO_4/Na_2HPO_4, 5 m*M* DTT, 1 m*M* EDTA, pH 7.0, at 37°C. Refolding was done by diluting fully unfolded protein in 5.5 *M* GuHCl into a final GuHCl concentration of 1.0 *M*. Fluorescence at 350 nm was monitored. Fluorescence intensity was normalized to the native and unfolded protein controls. Representative traces of independent triplicates for each protein are shown. (A) WT. (B) Y6A, F11A. (C) Y16A, Y28A. (D) Y45A, Y50A. (E) Y92A, Y97A. (F) F115A, F117A. (G) Y133A, Y138A.

7. Kinetic Unfolding/Refolding at 18°C

In the kinetic experiments at 37°C reported above, it was puzzling to see that four mutant HγD-Crys, Y6A, F11A, Y16A, and Y28A, had no obvious differences compared to the WT protein in either unfolding or refolding. To slow down the fast phase of the kinetic traces that were not resolved by the instrument at 37°C, the kinetic experiments were repeated under identical conditions to those described above, only at 18°C instead of 37°C.

The unfolding process of the WT and the four mutant proteins tested was much slower at 18°C than at 37°C, lasting for about 3 hr before reaching the final level, compared to about 15 min at 37°C [Figure 2-7 (A–C insets)]. As a whole, the unfolding traces of the WT and the mutant proteins overlapped very well. However, a close examination of the beginning of the traces revealed a "burst" for the mutant proteins, where the mutant proteins unfolded slightly but significantly faster than the WT protein [Figure 2-7 (A–C main panels)]. The mutant proteins were then caught up by the WT protein about 10 min into the experiment and at about 50% of the final fluorescence level. All four mutant proteins behaved similarly. This result indicated that these N-td aromatic pairs determined the unfolding rate of the early phase; however, the late phase was independent of these N-td aromatic pairs.

For refolding, lowering the temperature from 37 to 18°C did not slow down the refolding process for neither the WT nor mutant proteins [Figure 2-7 (D–F)]. The refolding processes reached the final level in about 10 min at both temperatures. The refolding traces of the WT and mutant proteins still overlapped very well at 18°C, like at 37°C [Figure 2-7 (E,F)]. This result further supported our conclusion that the refolding kinetics of each crystallin domain depended on only the second Greek key pair but not any other pairs.

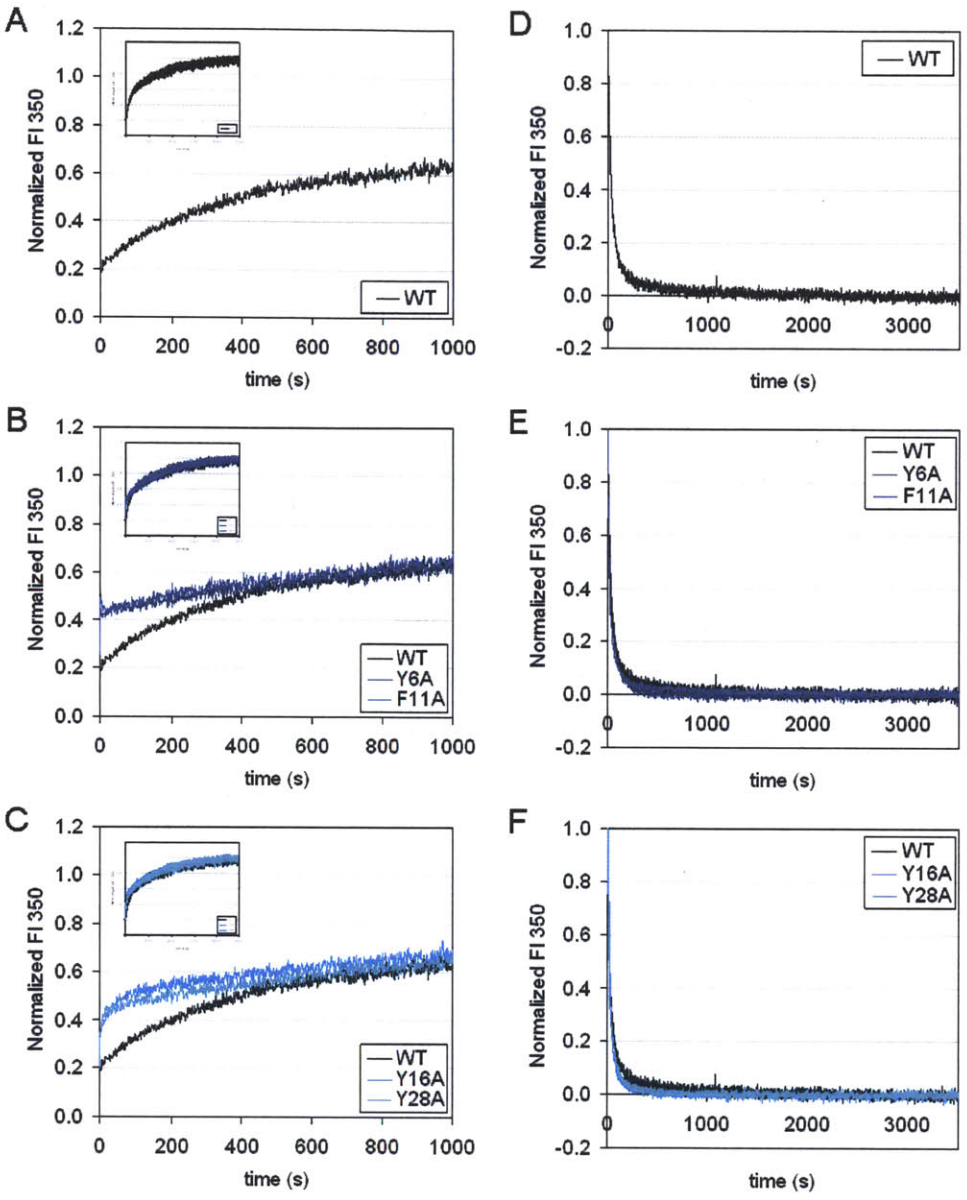


Figure 2-7 Kinetic Unfolding/Refolding for WT and Mutant HγD-Crys at 18°C.
Experiments were performed under identical conditions to those shown in Figure 2-5 and Figure 2-6, except that temperature was maintained at 18°C, instead of 37°C. Left panels (A)–(C): unfolding. The main panels show the first 20 min, and the insets show the full-length experiments for 3 hr (*x*-axis maximum: 10,000 sec). Right panels (D)–(F): refolding. Top panels (A) and (D): WT. Middle panels (B) and (E): Y6A, F11A. Bottom panels (C) and (F): Y16A, Y28A.

8. Double Mutant Cycle

Aromatic-aromatic interaction was estimated to be ~1.3 kcal/mol in the case of barnase by double mutant cycle analysis (Serrano *et al.* 1991). We employed this method for the three N-td aromatic pairs, because the N-td mutant proteins had separated, well-defined N-td and C-td equilibrium transitions. Three additional mutants of HγD-Crys: double mutants Y6A/F11A, Y16A/Y28A, and Y45A/Y50A were constructed, expressed and subjected to equilibrium unfolding/refolding experiments described above. ΔG's were calculated and compared with those of the corresponding pair of single mutant proteins. Aromatic-aromatic interaction energies for pairs Y6/F11, Y16/Y28, Y45/Y50 were −1.6, −1.5, and −2.0 kcal/mol, respectively. These numbers were within reasonable range of the literature values (Burley and Petsko 1985; Serrano *et al.* 1991), consistent with the notion that aromatic interaction is a "weak force." The results also suggested that the contributions of these aromatic residues to thermodynamic stability originated partially from the pair-wise aromatic-aromatic interaction, and partially from the interactions of these residues with other neighboring residues.

E. Discussion

The stability of crystallins is critical for maintenance of the transparency and appropriate refractive index of the vertebrate eye lens. The amino acid determinants of the stability of the βγ-crystallins probably reflect the folding/unfolding of the intertwined β-strands of the double Greek keys. The UV-damage model of cataractogenesis suggests that the aromatic residues could be the key to the folding and stability of lens crystallins. HγD-Crys contains six Tyr/Tyr, Tyr/Phe, or Phe/Phe pairs: four highly conserved Greek key pairs, and two moderately conserved non-Greek-key pairs. They are all located at structurally crucial β-hairpins, bridging two neighboring β-strands. Five of these six pairs (except for Y16/Y28) have partners five or fewer residues apart in the primary sequence. Burly and Petsko in their pioneering work suggested that aromatic-aromatic interactions may be nucleation sites for protein folding (Burley and Petsko 1985). The proximity of the partners in the primary sequence makes these aromatic pairs in HγD-Crys very likely to be such nucleation sites during the early steps of the folding process. After the protein matures into the native structure, these aromatic pairs may also act as "clasps" to stabilize local conformations, and thus contribute to the overall stability of the protein. Using site-directed mutagenesis followed by thermal, equilibrium, and kinetic experiments, we performed initial tests of these hypotheses. The results allowed us to confirm and expand upon the previously proposed HγD-Crys unfolding/refolding pathway.

1. Mutant Proteins Maintain WT-like Structures

Soluble protein yields and far-UV CD spectra showed that loss of one or two aromatic residues did not prevent most of the mutant HγD-Crys from folding into near-native structures. However, pair Y133/Y138 seemed essential for folding, since newly synthesized HγD-Crys was largely insoluble without either of these two residues. This was presumably due to intracellular aggregation into inclusion bodies that overwhelmed a defective folding process. For the small fractions of Y133A and Y138A proteins that did fold up and were soluble, they were able to maintain the near-native structure in buffer. These results indicated that these six aromatic pairs were not required for dictating the Greek key fold. However, the subsequent results indicated

that the stability and unfolding/refolding kinetics were compromised by substitutions of specific subsets of these aromatic pairs.

2. Thermal Denaturation

During thermal denaturation, substitutions at various positions throughout HγD-Crys generated very similar two-state transition traces without obvious intermediates, but with transition midpoints shifted to lower temperatures. This result suggested that thermal denaturation of HγD-Crys was probably a highly cooperative process. All the stabilizing factors within the protein collectively determine a single "melting point," at which different parts of the protein unfold simultaneously. However, aggregation occurred during the thermal denaturation process, and it was not clear when aggregation started to dominate during thermal denaturation. It is possible that the observed two-state process simply reflected the nature of the aggregation process, but not the unfolding process. In this case, more complicated unfolding features such as a three-state transition might be masked. The aggregation may be a result of the heating procedure, higher protein concentration, buffer with low ionic strength, or a combination of these. Nonetheless, the transition midpoints should be valid indicators of the general stability. The results indicated that five out of the six pairs (except F115/F117) clearly contributed to the thermal stability of HγD-Crys. However, the lack of an identifiable intermediate prevented the inference of a sequential pathway of HγD-Crys thermal unfolding.

β-Crystallins generally lack the two non-Greek-key pairs and can therefore be viewed as "natural" aromatic mutants of the γ-crystallins. Human βB1-crystallin lacks three of the four non-Greek-key paired residues while maintaining all of the Greek key pairs. In experiments similar to those in this study, Lampi *et al.* showed that the thermal transition temperature of βB1-crystallin was about 67 °C (Lampi *et al.* 2002). This transition temperature was lower than any of the single aromatic mutant HγD-Crys, despite the possible dimeric state that could enhance the stability of βB1-crystallin. The aromatic residues may be the key to the different thermal stabilities of β- and γ-crystallins.

3. Destabilization Patterns

Comparing the relative stabilities of the WT and mutant HγD-Crys carrying substitutions within the same crystallin domain, we observed two patterns common in both thermal and equilibrium experiments. (1) Substitutions of the Greek key pairs had larger effects than substitutions of the non-Greek-key pairs. This result was not surprising considering the higher level of conservation of the Greek key pairs, across the four Greek keys in HγD-Crys and across the βγ-crystallin superfamily, in terms of both primary sequence and three-dimensional structure. It can be concluded that one of the functions of these Greek key pairs is maintenance of protein stability. (2) For the first Greek key pair in each crystallin domain, substitution of the first residue had larger effects than substitution of the second residue. This may be because the first residue is slightly more buried, and thus has slightly more interaction with the core of the motif.

4. An Unfolding/Refolding Model of HγD-Crys

The result that all mutant HγD-Crys had destabilized equilibrium profiles indicated that all six aromatic pairs must contribute to some aspect(s) of unfolding and/or refolding kinetics. Surprisingly, for the four N-td mutant proteins Y6A, F11A, Y16A, and Y28A, neither kinetic unfolding nor refolding experiment revealed any obvious difference compared to the WT protein at 37°C. The mixing system used in all the kinetic experiments in this study had a dead time on the order of seconds. Therefore, we suspected that there might be differences in the fast phase that were not detected due to this limitation. By repeating the kinetic experiments at 18°C, we successfully captured a subtle but significant "burst" for the four N-td mutant proteins Y6A, F11A, Y16A, and Y28A compared to the WT at the beginning of the unfolding traces. These four N-td substitutions accelerated the early phase, but did not subsequently accelerate the late phase. This indicated that the fast phase, although occurred before the slow phase, was not a prerequisite for the slow phase to occur. At 37°C, only C-td but not N-td substitutions accelerated the late phase. Combining these results at both temperatures, we propose that the unfolding of the two crystallin domains of HγD-Crys are independent, but appear to be sequential due to intrinsic stability difference [Figure 2-8 (A)]. All β-hairpin aromatic pairs contribute to the kinetic stability of each crystallin domain. Double mutant cycle analysis showed that aromatic-aromatic interactions of the pairs was one source of stability. However, our results

did not show evidence regarding the sequential unfolding of the two Greek key motifs within a crystallin domain.

The refolding rate of HγD-Crys was only significantly slowed down by substitutions of the second Greek key pair in each crystallin domain but not any other pair, even though the two Greek key pairs in a crystallin domain are homologous in sequence and structure. Lowering the temperature to 18°C neither slowed down the refolding process of the WT and mutant proteins Y6A, F11A, Y16A, Y28A, nor revealed any differences in refolding among the WT and these mutant proteins. These results indicated that unlike unfolding, refolding had limited temperature dependency, and also confirmed the lesser role of the corresponding residues in refolding. Flaugh *et al.* proposed that unfolding/refolding of a crystallin domain is reversible through an intermediate with the first Greek key unfolded and the second at Greek key folded (Flaugh *et al.* 2006). This proposal with details at the Greek key level was largely based on data that showed that the second Greek key had stabilizing effects (Bagby *et al.* 1998; MacDonald *et al.* 2005), but might overlook other structural elements in the first Greek key that are equally stabilizing (Bagby *et al.* 1998) [Figure 2-4 (A, D)]. Our experimental evidence of unequal contributions to the refolding kinetics from the two homologous Greek key pairs in the same crystallin domain strongly argues that the second Greek keys are involved in the rate-determining step of crystallin domain folding. Based on previous studies suggesting that aromatic-aromatic interactions may nucleate protein folding (Burley and Petsko 1985; Bagby *et al.* 1998), we propose that this rate-determining step is a nucleation event from the aromatic-aromatic interaction of the second Greek key pair, and then the first Greek key folds using the folded second Greek key as a template [Figure 2-8 (B)]. However, we cannot rule out the possibility that this rate-determining step is a late step that is vulnerable to amino acid substitution. Yet another alternative is that the apparent unequal contributions may be due to the asymmetric positions of tryptophans within the crystallin domain, leading to different sensitivity of the fluorescence signal to structural changes in different parts of the crystallin domain.

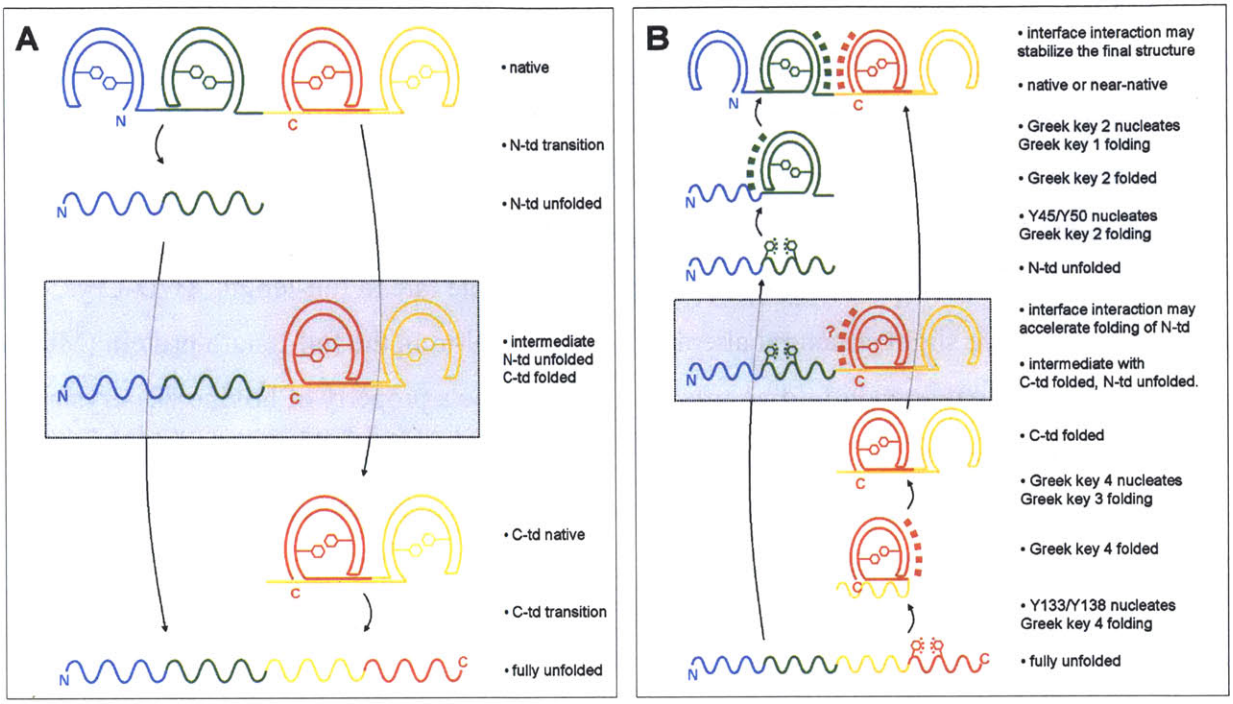


Figure 2-8 Schematic Diagram of the Proposed HγD-Crys Unfolding and Refolding Model. The four Greek keys are colored blue, green, yellow, and red from N- to C-terminus. Horseshoe-shape symbols represent folded Greek keys, while the wavy lines represent unfolded chains. Hexagonal symbols represent Greek key aromatic pairs. Square dots indicate molecular events in action. See text next to each panel for more explanation. The species in gray boxes are the observed, but probably not obligate intermediates, due to independent domain unfolding/refolding. (A) Unfolding. (B) Refolding.

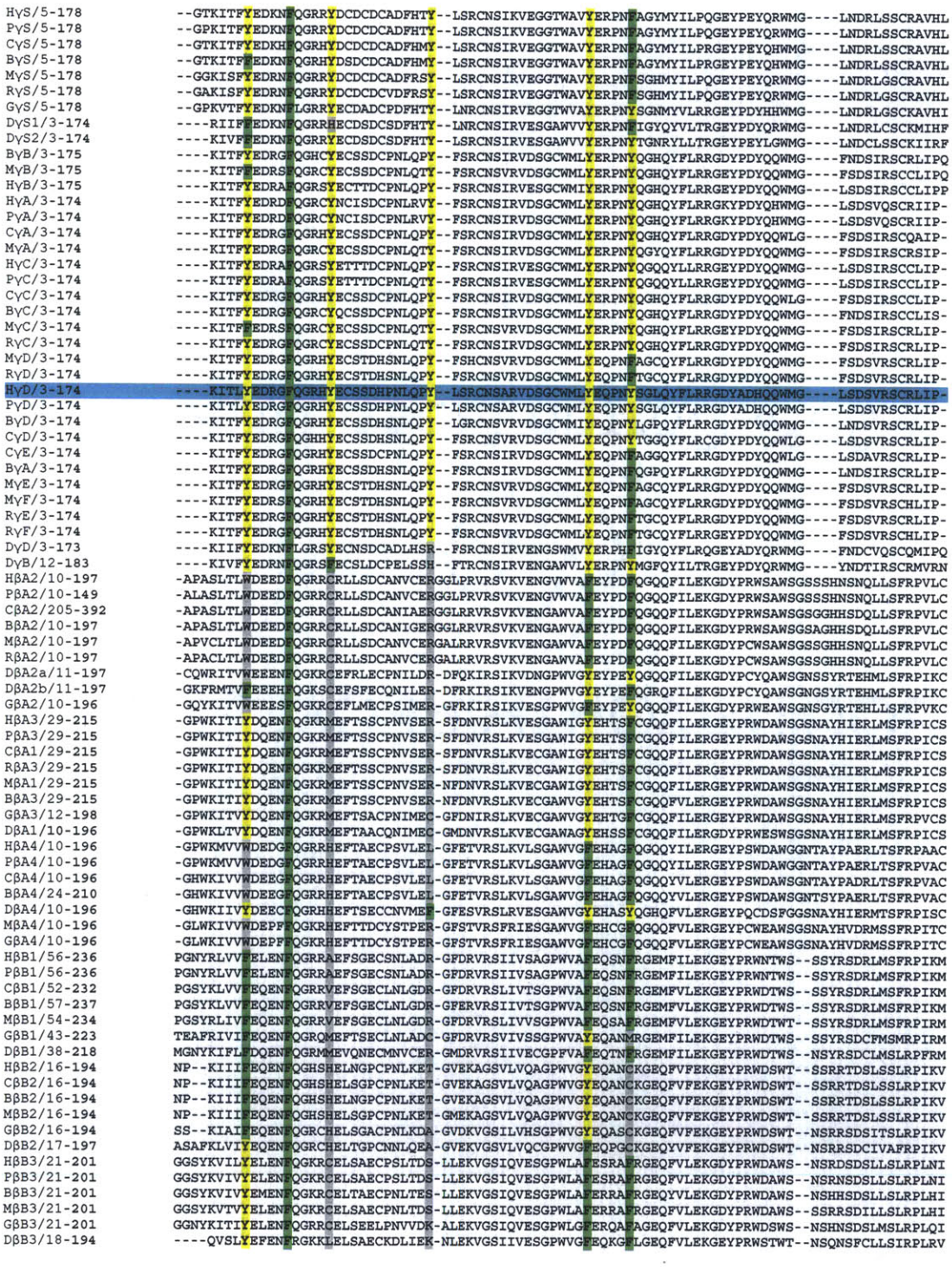
At the crystallin domain level, although it is well-established that the C-td folds first, followed by the N-td, during HγD-Crys refolding [Figure 2-6 (D,G)] (Kosinski-Collins *et al.* 2004; Flaugh *et al.* 2006; Mills *et al.* 2007), the role of the domain interface in folding remains unclear. While Flaugh *et al.* showed that substitutions of interface residues slowed down the refolding of the N-td (Flaugh *et al.* 2005; Flaugh *et al.* 2005; Flaugh *et al.* 2006), Mills *et al.* found that isolated N-td was able to refold with a comparable rate to full-length. HγD-Crys, although it may have slight structural alterations from the N-td in the full-length protein (Mills *et al.* 2007). Independent domain folding was also shown to be a property of bovine γB- crystallin (Rudolph *et al.* 1990). In HγD-Crys, the interface interactions may not be required for the initial refolding process of the N-td, but could be important for stabilizing the final conformation of the N-td after it folds up [Figure 2-8 (B)].

5. Connection with Computational Studies

A number of computational studies offered insights on the HγD-Crys folding pathway. Using a structure-based model, Itoh and Sasai recaptured the sequential folding model of Flaugh *et al.* (Flaugh *et al.* 2006; Itoh and Sasai 2008). Their results also showed that the connectivity of HγD-Crys determined the folding sequence of the two crystallin domains. Among the six aromatic pairs, the Greek key pair Y133/Y138 is particularly interesting because substitutions of this pair affected all the parameters tested, including thermal stability, thermodynamic stability, unfolding, and refolding kinetics. These results helped to explain the insolubility of mutant proteins Y133A and Y138A during expression. Interestingly, Das *et al.* identified a salt bridge, E134-R141, as a nucleation site for folding by molecular dynamics simulation (Das *et al.* 2009), and the location coincided with the pair Y133/Y138, our proposed nucleation site of the initial folding of HγD-Crys. The region around the β-hairpin at Greek key 4 seems to be highly crucial for the folding and structural integrity of HγD-Crys.

F. Supplementary Materials

N-td



C-td

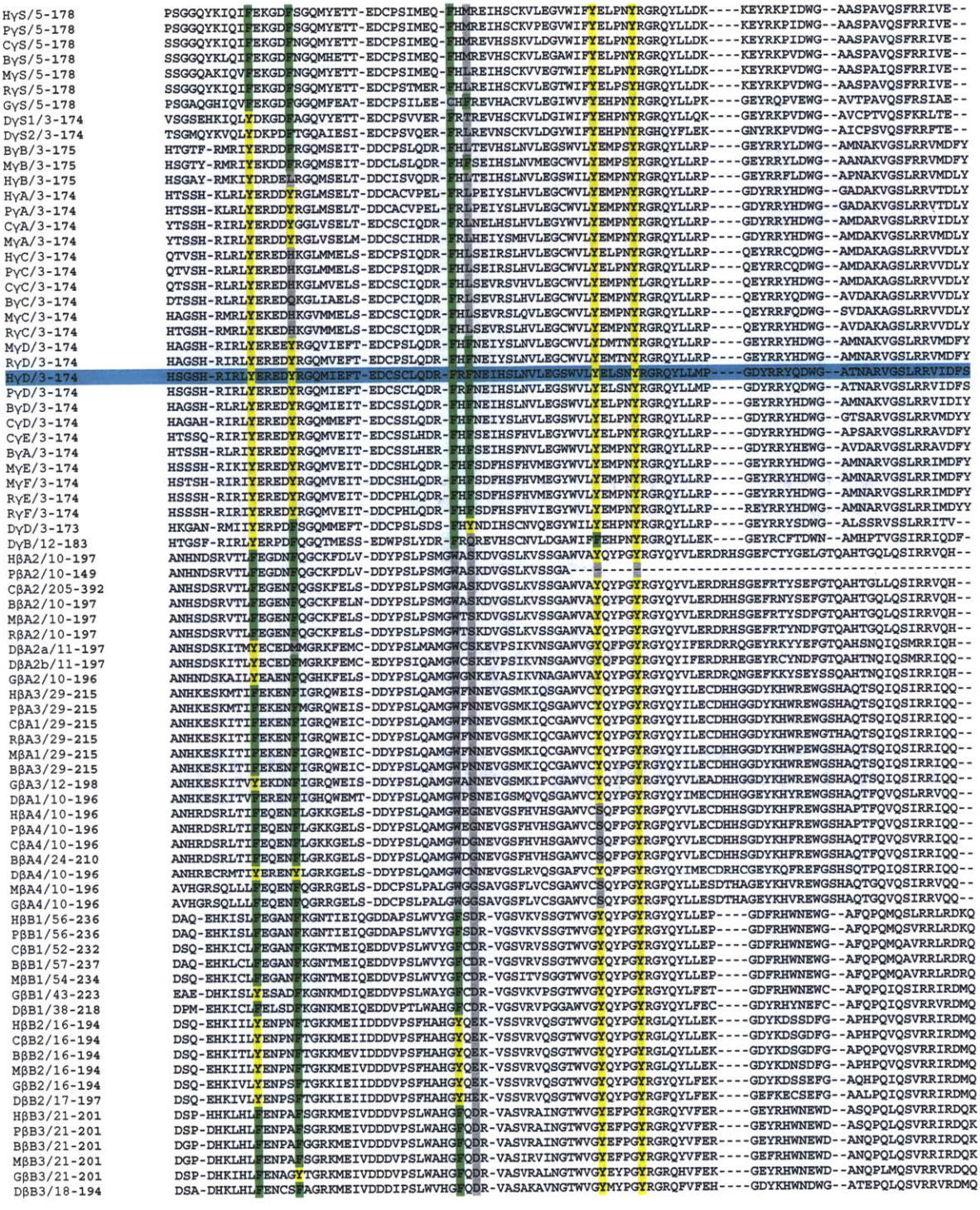

HγS/5-178 PSGGQYKIQIFEKGDFSGQMYETT-EDCPSIMEQ-FHMREIHSCKVLEGVWIFYELPNYRGRQYLLDK----KEYRKPIDWG--AASPAVQSFRRIVE--
PγS/5-178 PSGGQYKIQIFEKGDFSGQMYETT-EDCPSIMEQ-FHMREIHSCKVPEGVWIFYELPNYRGRQYLLDK----KEYRKPIDWG--AASPAVQSFRRIVE--
CγS/5-178 SSGGQYKIQIFEKGDFNGQMYETT-EDCPSIMEQ-FHMREVHSSKVLDGVWIFYELPNYRGRQYLLDK----KEYRKPIDWG--AASPAVQSFRRIVE--
BγS/5-178 SSGGQYKLQIFEKGDFNGQMHETT-EDCPSIMEQ-FHMREVHSCKVLEGAWIFYELPNYRGRQYLLDK----KEYRKPVDWG--AASPAVQSFRRIVE--
MγS/5-178 SSGGQAKIQVFEKGDFNGQMYETT-EDCPSIMEQ-FHLREIHSCKVVEGTWIFYELPNYRGRQYLLDK----KEYRKPVDWG--AASPAIQSFRRIVE--
RγS/5-178 SSGGQYKIQIFEKGDFSGQMYETT-EDCPSTMER-FHLREIHSCKVLEGTWIFYELPSYHGRQYLLDK----KEYRKPVDWG--AASPAVQSFRRIVE--
GγS/5-178 PSGAQGHIQVFEKGDFGGQMFEAT-EDCPSILEE-CHFREVHACRVLEGIWVFYEHPNYRGRQYLLPK----GEYRQPVEWG--AVTPAVQSFRSIAE--
DγS1/3-174 VSGSEHKIQLYDKGDFAGQVYETT-EDCPSVVER-FRTREVHSCKVLDGIWIFYEHPNYRGRQYLLQK----GEYRKPVDWG--AVCPTVQSFKRLTE--
DγS2/3-174 TSGMQYKVQLYDKPDFTGQAIESI-EDCPSVQER-FRLREVNSCKVLDGYWIFYEHPNYRGHQYFLEK----GNYRKPVDWG--AICPSVQSFRRFTE--
BγB/3-175 HTGTF-RMRIYERDDFRGQMSEIT-DDCPSLQDR-FHLTEVHSLNVLEGSWVLYEMPSYRGRQYLLRP----GEYRRYLDWG--AMNAKVGSLRRVMDFY
MγB/3-175 HSGTY-RMRIYEKDDFRGQMSELT-DDCLSLQDR-FHFSEIHSLNVMEGCWVLYEMPSYRGRQYLLRP----GEYRRYLDWG--AANAKVGSFRRVMDFY
HγB/3-175 HSGAY-RMKIYDRDELRGQMSELT-DDCISVQDR-FHLTEIHSLNVLEGSWILYEMPNYRGRQYLLRP----GEYRRFLDWG--APNAKVGSLRRVMDLY
HγA/3-174 HTSSH-KLRLYERDDYRGLMSELT-DDCACVPEL-FRLPEIYSLHVLEGCWVLYEMPNYRGRQYLLRP----GDYRRYHDWG--GADAKVGSLRRVTDLY
PγA/3-174 HTSSH-KLRLYERDDYRGLMSELT-DDCACVPEL-FRLPEIYSLHVLEGCWVLYEMPNYRGRQYLLRP----GDYRRYHDWG--GADAKVGSLRRVTDLY
CγA/3-174 YTSSH-RIRLYERDDYGGLVSELT-EDCSCIQDR-FRLNELHSLHVLEGYWVLYEMPNYRGRQYLLRP----GDYRRYHDWG--AMDARVGSLRRVIDLY
MγA/3-174 YTSSH-RIRLYERDDYRGLVSELM-DDCSCIHDR-FRLHEIYSMHVLEGCWVLYEMPNYRGRQYLLRP----GDYRRYHDWG--AMDAKVGSLRRVMDLY
HγC/3-174 QTVSH-RLRLYEREDHKGLMMELS-EDCPSIQDR-FHLSEIRSLHVLEGCWVLYELPNYRGRQYLLRP----QEYRRCQDWG--AMDAKAGSLRRVVDLY
PγC/3-174 QTVSH-RLRLYEREDHKGLMMELS-EDCPSIQDR-FHLSEIRSLHVLEGCWVLYELPNYRGRQYLLRP----QEYRRCQDWG--AMDAKAGSLRRVVDLY
CγC/3-174 QTSSH-RLRLYEREDHKGLMVELS-EDCSCIQDR-FHLSEVRSVHVLEGCWVLYEMPNYLGRQYLLRP----QEYRRYHDWG--AMDAKAGSLRRVVDLY
BγC/3-174 DTSSH-RLRLYEREDQKGLIAELS-EDCPCIQDR-FRLSEVRSLHVLEGCWVLYEMPNYRGRQYLLRP----QEYRRYQDWG--AVDAKAGSLRRVVDLY
MγC/3-174 HAGSH-RMRLYEKEDHKGVMMELS-EDCSCIQDR-FHLSEVRSLQVLEGCWVLYEMPNYRGRQYLLRP----QEYRRFQDWG--SVDAKAGSLRRVVDLY
RγC/3-174 HTGSH-RMRLYEKEDHKGVMMELS-EDCSCIQDR-FHLSEVRSLHVLEGCWVLYEMPNYRGRQYLLRP----QEYRRYHDWG--AVDAKAGSLRRVVDLY
MγD/3-174 HAGSH-RIRLYEREEYRGQVIEFT-EDCPSLQDR-FHFNEIYSLNVLEGCWVLYDMTNYRGRQYLLRP----GEYRRYHDWG--AMNAKVGSLRRVMDFY
RγD/3-174 HAGSH-RIRLYEREDYRGQMVEFT-EDCPSLQDR-FHFNEIYSLNVLEGCWVLYEMTNYRGRQYLLRP----GEYRRYHDWG--AMNARVGSLRRVMDFY
HγD/3-174 HSGSH-RIRLYEREDYRGQMIEFT-EDCSCLQDR-FRFNEIHSLNVLEGSWVLYELSNYRGRQYLLMP----GDYRRYQDWG--ATNARVGSLRRVIDFS
PγD/3-174 HSGSH-RIRLYEREDYRGQMIEFT-EDCSCLQDR-FRFNEIHSLNVLEGSWVLYELSNYRGRQYLLMP----GDYRRYQDWG--ATNARVGSLRRVIDFS
BγD/3-174 HAGSH-RLRLYEREDYRGQMIEIT-EDCSSLQDR-FHFNEIHSLNVLEGSWVLYELPNYRGRQYLLRP----GEYRRYHDWG--AMNAKVGSLRRVIDIY
CγD/3-174 HAGAH-RIRLYEREDYRGQMMEFT-EDCSSLQDR-FHFNEIYSLHVLEGSWVLYELPNYRGRQYLLRP----GDYRRYHDWG--GTSARVGSLRRVMDYY
CγE/3-174 HTSSQ-RIRIYEREDYRGQMVEIT-EDCSSLHDR-FHFSEIHSFHVLEGYWVLYELPNYRGRQYLLRP----GDYRRYHDWG--APSARVGSLRRAVDFY
BγA/3-174 HTSSH-RLRIYEREDYRGQMVEIT-EDCSSLHER-FHFSEIHSFNVLEGWWVLYEMPNYRGRQYLLRP----GDYRRYHEWG--AVDARVGSLRRAVDFY
MγE/3-174 HSSSH-RIKIYEREDYRGQMVEIT-DDCSHLQDR-FHFSDFHSFHVMEGYWVLYEMPNYRGRQYLLRP----GEYRRYHDWG--AMNARVGSLRRIMDFY
MγF/3-174 HSTSH-RIRIYEREDYRGQMVEIT-DDCSHLQDR-FHFSDFHSFHVMEGYWVLYEMPNYRGRQYLLRP----GEYRRYHDWG--AMNARVGSLRRIMDFY
RγE/3-174 HSSSH-RIRIYEREDYRGQMVEIT-DDCPHLQDR-FHFSDFHSFHVMEGYWVLYEMPNYRGRQYLLRP----GEYRRYHDWG--AMNARVGSLRRIMDFY
RγF/3-174 HSSSH-RIRIYEREDYRGQMVEIT-DDCPHLQDR-FHFSDFHSFHVIEGYWVLYEMPNYRGRQYLLRP----REYRRYHDWG--AMNARVGSLRRIMDYY
DγD/3-173 HKGAN-RMIIYERPDFSGQMMEFT-DDCPSLSDS-FHYNDIHSCNVQEGYWILYEHPNYRGRQYLLRP----GEYRRYSDWG--ALSSRVSSLRRITV--
DγB/12-183 HTGSF-RIRLYERPDFQGQTMESS-EDWPSLYDR-FRQREVHSCNVLDGAWIFFEHPNYRGRQYLLEK----GEYRCFTDWN--AMHPTVGSIRRIQDF-
HβA2/10-197 ANHNDSRVTLFEGDNFQGCKFDLV-DDYPSLPSMGWASKDVGSLKVSSGAWVAYQYPGYRGYQYVLERDRHSGEFCTYGELGTQAHTGQLQSIRRVQH--
PβA2/10-149 ANHNDSRVTLFEGDNFQGCKFDLV-DDYPSLPSMGWASKDVGSLKVSSGA----------
CβA2/205-392 ANHSDSRVTLFEGENFQGSKFELS-DDYPSLPSMGWASKDVGSLKVSSGAWVAYQYPGYRGYQYVLERDRHSGEFRTYSEFGTQAHTGLLQSIRRVQH--
BβA2/10-197 ANHSDSRVTLFEGENFQGCKFELN-DDYPSLPSMGWASKDVGSLKVSSGAWVAYQYPGYRGYQYVLERDHHSGEFRNYSEFGTQAHTGQLQSIRRVQH--
MβA2/10-197 ANHSDSRVTLFEGENFQGCKFELS-DDYPSLPSMGWTSKDVGSLKVSSGAWVAYQYPGYRGYQYVLERDRHSGEFRTYSDFGTQAHTGQLQSIRRVQH--
RβA2/10-197 ANHSDSRVTLFEGENFQGCKFELS-DDYPSLPSMGWTSKDVGSLKVSSGAWVAYQYPGYRGYQYVLERDRHSGEFRTYNDFGTQAHTGQLQSIRRVQH--
DβA2a/11-197 ANHSDSKITMYECEDMMGRKFEMC-DDYPSLMAMGWCSKEVPSIKVNSGAWVGYQFPGYRGYQYIFERDRRQGEYRKYYEFGTQAHSNQIQSMRRIQH--
DβA2b/11-197 ANHSDSKITLYECEDEMGRKFEMC-EDYPSIQAMGWCSKEVPSIKVNSGAWVAYQYPGYRGYQYVLERDRQNGEFKKYSEYSSQAHTNQIQSMRRIQH--
GβA2/10-196 ANHNDSKAILYEAENFQGHKFELS-DDYPSLQAMGWGNKEVASIKVNAGAWVAYQYPGYRGYQYVLERDRQNGEFKKYSEYSSQAHTNQIQSIRRIQH--
HβA3/29-215 ANHKESKMTIFEKENFIGRQWEIS-DDYPSLQAMGWFNNEVGSMKIQSGAWVCYQYPGYRGYQYILECDHHGGDYKHWREWGSHAQTSQIQSIRRIQQ--
PβA3/29-215 ANHKESKMTIFEKENFMGRQWEIS-DDYPSLQAMGWFNNEVGSMKIQSGAWVCYQYPGYRGYQYILECDHHGGDYKHWREWGSHAQTSQIQSIRRIQQ--
CβA1/29-215 ANHKESKITIFEKENFIGRQWEIC-DDYPSLQAMGWFNNEVGSMKIQCGAWVCYQYPGYRGYQYILECDHHGGDYKHWREWGSHAQTSQIQSIRRIQQ--
RβA3/29-215 ANHKESKITIFEKENFIGRQWEIC-DDYPSLQAMGWFNNEVGSMKIQCGAWVCYQYPGYRGYQYILECDHHGGDYKHWREWGTHAQTSQIQSIRRIQQ--
MβA1/29-215 ANHKESKITIFEKENFIGRQWEIC-DDYPSLQAMGWFNNEVGSMKIQCGAWVCYQYPGYRGYQYILECDHHGGDYKHWREWGSHAQTSQIQSIRRIQQ--
BβA3/29-215 ANHKESKITIFEKENFIGRQWEIC-DDYPSLQAMGWPNNEVGSMKIQCGAWVCYQYPGYRGYQYILECDHHGGDYKHWREWGSHAQTSQIQSIRRIQQ--
GβA3/12-198 ANHKESKITVYEKDNFIGRQWEIS-DDYPSLQAMGWANNEVGSMKIPCGAWVCYQYPGYRGYQYVLEADHHGGDYKHWREWGSHAQTSQIQSIRRIQQ--
DβA1/10-196 ANHKESKITVFERENFIGHQWEMT-DDYPSLQAMGWPSNEIGSMQVQSGAWVCYQYPGYRGYQYIMECDHHGGEYKHYREWGSHAQTFQVQSLRRVQQ--
HβA4/10-196 ANHRDSRLTIFEQENFLGKKGELS-DDYPSLQAMGWEGNEVGSFHVHSGAWVCSQFPGYRGFQYVLECDHHSGDYKHFREWGSHAPTFQVQSIRRIQQ--
PβA4/10-196 ANHRDSRLTIFEQENFLGKKGELS-DDYPSLQAMGWEGNEVGSFHVHSGAWVCSQFPGYRGFQYVLECDHHSGDYKHFREWGSHAPTFQVQSIRRIQQ--
CβA4/10-196 ANHRDSRLTIFEQENFLGKKGELS-DDYPSLQAMGWDGNEVGSFHVHSGAWVCSQFPGYRGFQYVLECDHHSGDYKHFREWGSHAQTFQVQSVRRIQQ--
BβA4/24-210 ANHRDSRLTIFEQENFLGRKGELS-DDYPSLQAMGWDGNEVGSFHVHSGAWVCSQFPGYRGFQYVLECDHHSGDYKHFREWGSHAQTFQVQSIRRIQQ--
DβA4/10-196 ANHRECRMTIYERENYLGRKGELS-DDYPSLQAMGWCNNEVGSLRVQSGAFVCYQFPGYRGYQYIMECDRHCGEYKQFREFGSHSQTPQIQSIRRIQQ--
MβA4/10-196 AVHGRSQLLLFEQENFQGRRGELS-DDCPSLPALGWGGSAVGSFLVCSGAWVCSQYPGYRGFQYLLESDTHAGEYKHVREWGSHAQTGQVQSIRRVQQ--
GβA4/10-196 AVHGRSQLLLFEQENFQGRRGELS-DDCPSLPALGWGGSAVGSFLVCSGAWVCSQYPGYRGFQYLLESDTHAGEYKHVREWGSHAQTGQVQSIRRVQQ--
HβB1/56-236 DAQ-EHKISLFEGANFKGNTIEIQGDDAPSLWVYGFSDR-VGSVKVSSGTWVGYQYPGYRGYQYLLEP----GDFRHWNEWG--AFQPQMQSVRRLRDKQ
PβB1/56-236 DAQ-EHKISLFEGANFKGNTIEIQGDDAPSLWVYGFSDR-VGSVKVSSGTWVGYQYPGYRGYQYLLEP----GDFRHWNEWG--AFQPQMQSVRRLRDKQ
CβB1/52-232 DSQ-EHKICLFEGANFKGKTMEIQEDDVPSLWVYGFCDR-VGSVRVSSGTWVGYQYPGYRGYQYLLEP----GDFRHWNEWG--AFQPQMQAVRRLRDRQ
BβB1/57-237 DAQ-EHKLCLFEGANFKGNTMEIQEDDVPSLWVYGFCDR-VGSVRVSSGTWVGYQYPGYRGYQYLLEP----GDFRHWNEWG--AFQPQMQAVRRLRDRQ
MβB1/54-234 DSQ-EHKICLFEGANFKGNTMEIQEDDVPSLWVYGFCDR-VGSITVSGGTWVGYQYPGYRGYQYLLEP----GDFRHWNEWG--AFQPQMQAVRRLRDRQ
GβB1/43-223 EAE-DHKISLYESADFKGNKMDIQEDDVPSLWAYGFCDR-VGSVKVPSGTWVGYQYPGYRGYQYLFET----GDFRHWNEWC--AFQPQIQSIRRIRDMQ
DβB1/38-218 DPM-EHKICLFELSDFKGNKMEIQEDDVPTLWAHGFCDR-VGSVRVPGGAWVGYQYPGYRGYQYLFEC----GDYRHYNEFC--AFQPQIQSVRRVRDMQ
HβB2/16-194 DSQ-EHKIILYENPNFTGKKMEIIDDDVPSFHAHGYQEK-VSSVRVQSGTWVGYQYPGYRGLQYLLEK----GDYKDSSDFG--APHPQVQSVRRIRDMQ
CβB2/16-194 DSQ-EHKIILYENPNFTGKKMEIIDDDVPSFHAHGYQEK-VSSVRVQSGTWVGYQYPGYRGLQYLLEK----GDYKDSGDFG--APHPQVQSVRRIRDMQ
BβB2/16-194 DSQ-EHKITLYENPNFTGKKMEVIDDDVPSFHAHGYQEK-VSSVRVQSGTWVGYQYPGYRGLQYLLEK----GDYKDSGDFG--APQPQVQSVRRIRDMQ
MβB2/16-194 DSQ-EHKIILYENPNFTGKKMEIVDDDVPSFHAHGYQEK-VSSVRVQSGTWVGYQYPGYRGLQYLLEK----GDYKDNSDFG--APHPQVQSVRRIRDMQ
GβB2/16-194 DSQ-EHKIVLYENPSFTGKKIEIIDDDVPSFHAHGYQEK-VSSVRVQSGTWVGYQYPGYRGYQYLFEK----GDYKDSSEFG--AQHPQIQSVRRIRDMQ
DβB2/17-197 DSQ-EHKIVLYENPSFTGKKIEIIDDDVPSFHAHGYHEK-VSSVRVQSGTWVGYQYPGYRGFQYLFEK----GEFKECSEFG--AALPQIQSVRRIRDMQ
HβB3/21-201 DSP-HHKLHLFENPAFSGRKMEIVDDDVPSLWAHGFQDR-VASVRAINGTWVGYEFPGYRGRQYVFER----GEYRHWNEWD--ASQPQLQSVRRIRDQK
PβB3/21-201 DSP-DHKLHLFENPAFSGRKMEIVDDDVPSLWAHGFQDR-VASVRAINGTWVGYEFPGYRGRQYVFER----GEYRHWNEWD--ASQPQLQSVRRIRDQK
BβB3/21-201 DGP-DHKLHLFENPAFGGRKMEIVDDDVPSLWAHGFQDR-VASVRAINGTWVGYEFPGYRGRQYVFER----GEYRHWNEWD--ANQPQLQSVRRIRDQK
MβB3/21-201 DGP-DHKLHLFENPAFSGRKMEIVDDDVPSLWAHGFQDR-VASIRVINGTWVGYEFPGYRGRQYVFER----GEFRHWNEWD--ANQPQLQSVRRIRDQK
GβB3/21-201 DSP-DHKIHLFENAGYTGRKMEIVDDDVPSLWAHGFQDR-VASVRALNGTWVGYEYPGYRGRQHVFEK----GEYRHWNEWD--ANQPLMQSVRRVRDQQ
DβB3/18-194 DSA-DHKLHLFENCSFAGRKMEIVDDDIPSLWVHGFQDR-VASAKAVNGTWVGYMYPGYRGRQFVFEH----GDYKHWNDWG--ATEPQLQSVRRVRDMQ

Figure 2-9 Sequence Alignment of Selected βγ-Crystallin Sequences.
79 sequences of βγ-crystallins, including human βA1-4, βB1-3, γA-F, γS, and homologues from chimpanzee, dog, cattle, rat, mouse, chicken and zebrafish were compiled and aligned using ClustalW2 [45]. HγD is highlighted with turquoise color, Homologous residues of the six HγD-Crys aromatic pairs (Y6, F11; Y16, Y28; Y45, Y50; Y92, Y97; F115, F117; Y133, Y138) in other βγ-crystallins are in bold face. For these residues, tyrosines (Y) are highlighted with yellow; phenylalanines (F) are highlighted with green; all other residues are highlighted with grey. For the nomenclature of the proteins, the first upper letters indicate the organisms where they come from, H: human; P: chimpanzee; C: dog; B: cattle; R: rat; M: mouse; G: chicken; D: zebrafish. The numbers after the name indicate the fragments of the proteins that are shown here. N-td and C-td extensions, if any, are not shown for clarity. The roster is ordered by relatedness in sequence. The alignment is separated into N-td and C-td panels.

Chapter 3

Aggregation of Greek Key Aromatic Mutant Human γD-Crystallins and Their Interactions with Human αB-Crystallin [1]

[1] **Modified from manuscript submitted to *Protein Science* in January, 2012.**

A. Abstract

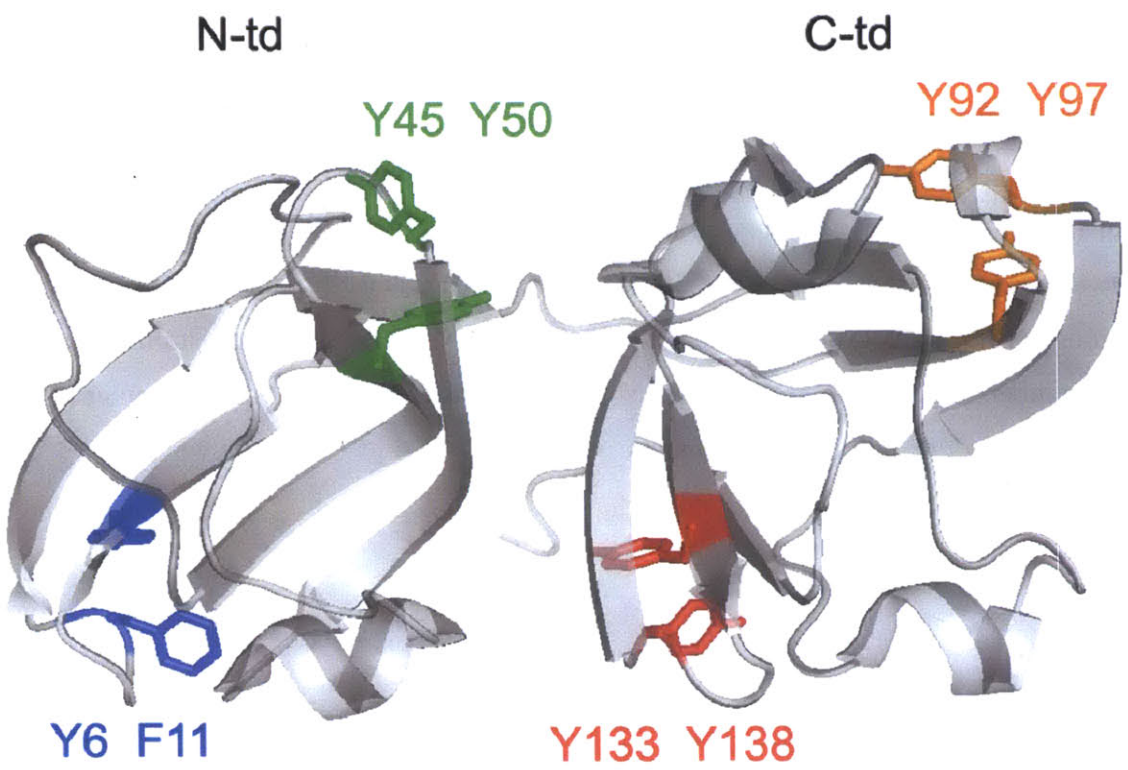
Lens crystallins include the major structural proteins βγ-crystallins and the chaperone α-crystallin—a small heat shock protein that sequesters partially unfolded and aggregation-prone substrates. Intrinsic solubility and stability, as well as chaperone-assisted resistance to aggregation are essential properties of crystallins. Aggregated crystallins within the lens scatter light and directly contribute to the development of cataract, a worldwide leading cause of blindness. Human γD-crystallin (HγD-Crys) is an abundant protein in the lens nucleus. Four Tyr-Tyr or Tyr-Phe pairs located at the Greek key β-hairpins in HγD-Crys play important roles in the stability and unfolding/folding kinetics of the protein. Since these aromatic pairs are only fully exposed in partially folded or unfolded states, they are candidates for interactions involved in HγD-Crys off-pathway aggregation and chaperone-substrate recognition by α-crystallin. Using alanine substitutions of each of the eight paired Tyr or Phe residues, we investigated refolding-induced aggregation, as well as suppression of aggregation by the chaperone human αB-crystallin (HαB-Crys). Overall, the WT and mutant HγD-Crys had very similar aggregation kinetics and final aggregation levels, indicating that these aromatic pairs played minimal role in the aggregation process. The aggregation suppression efficiencies by HαB-Crys chaperone, as well as the HαB-bound conformations characterized by the fluorescence of HαB-HγD complexes were also very similar for the WT and mutant HγD-Crys, arguing against a critical role of these aromatic pairs in the chaperone recognition process. The fluorescence of the tryptophans resembled a partially unfolded state, consistent with the tryptophans being part of the contact site with the chaperone.

B. Introduction

Protein folding pathways and/or the stability of native protein folds are often challenged by mutations, incorrect post-translational modifications, or environmental insults, resulting in misfolded/partially unfolded conformations that are prone to off-pathway aggregation (Wetzel 1996). Such aggregated states are implicated in many human diseases (Stefani 2004). Well-known examples include neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease, and Huntington's disease (Bossy-Wetzel *et al.* 2004); systemic light chain amyloidosis (Baden *et al.* 2009); the human form of prion syndrome— Creutzfeldt-Jakob disease; and the eye disease cataract. Unfortunately for many of the protein aggregation diseases, it remains unclear whether the aggregates are the cause of the diseases, by-product of the pathology, or part of a protective mechanism (Bodner *et al.* 2006). For cataract on the other hand, it is clear that high-molecular-weight aggregates from the lens proteins crystallins can scatter visible light and directly cause the loss of visual acuity (Benedek 1997). However, for many of the protein aggregation diseases, cataract included, the sites of polypeptide chain that initiate the aggregation pathways or maintain the aggregated states remain elusive. In addition to understanding and controlling disease processes, combating protein aggregation is also a major mission in biomedical research and biotechnology industry, during small- or large-scale recombinant protein expression (Mitraki and King 1989).

Protein aggregation can be classified into two broad categories: 1) amyloid and 2) non-amyloid/amorphous forms (Wetzel 1996). Typical amyloid aggregates appear as long, twisted and unbranched fibers under electron microscopy, and cause spectral alteration when stained with fluorescence dyes Thioflavin T (ThT) or Congo Red. The cross-β structure identified by X-ray diffraction is the signature of amyloid. In the case of non-amyloid/amorphous aggregation, the macromolecular morphology, as well as the conformations of the polypeptide chains in the aggregated state are not well defined. Nonetheless the intermolecular interaction may be quite specific, such as those in a domain-swap or loop-sheet insertion manner (Yamasaki *et al.* 2008). Most inclusion bodies resulting from failure of folding of recombinant proteins within bacterial hosts are classic representatives of amorphous aggregates, though amyloid forms have also been reported (Wang *et al.* 2008).

The crystallins represent 90% of the soluble proteins in the eye lens (Bron *et al.* 2000). They fall into two major classes: α-crystallins and βγ-crystallins (Bloemendal *et al.* 2004). βγ-Crystallins are the major structural proteins and share a common structure: two homologous crystallin domains, each domain with two Greek key motifs and 8 intercalating β-strands. The β-hairpin aromatic pair with consensus sequence Y/FXXXXY/FXG (Figure 3-1) and nearby glycine and serine at each Greek key (Blundell *et al.* 1981), as well as the tyrosine corner (Hemmingsen *et al.* 1994) and tryptophan pair (Chen *et al.* 2009) at each crystallin domain, signify membership of this class of proteins. β-Crystallins form dimeric and oligomeric structures while γ-crystallins are monomers.




```
N-td
1   GKITLYEDRGFQGRHYECSSDHPNLQPYLSRCNSARVDS
40  GCWMLYEQPNYSGLQYFLRRGDYADHQQWMGLSDSVRSCRLIPHSGS
C-td
87  HRIRLYEREDYRGQMIEFTEDCSCLQDRFRFNEIHSLNVLE
128 GSWVLYELSNYRGRQYLLMPGDYRRYQDWGATNARVGSLRRVIDFS
```

Figure 3-1 The Greek Key Aromatic Pairs in HγD-Crys.

Structure and sequence of HγD-Crys and the conserved Greek key aromatic residues. Blue: Y6, F11; green: Y45, Y50; orange: Y92, Y97; red: Y133, Y138. These residues were substituted with alanines in this study. Upper panel: Crystal structure of HγD-Crys (PDB ID: 1HK0) (Basak *et al.* 2003) showing the two homologous crystallin domains: N-terminal domain (N-td) and C-terminal domain (C-td). The Greek key aromatic pairs are shown in stick representation. The three-dimensional representation was prepared by PyMOL (DeLano Scientific). Lower panel: amino acid sequence of HγD-Crys, segmented by the four Greek key motifs, but without further alignment.

In lens development, mature lens fiber cells lose their organelles upon differentiation from the epithelial cells, thus lose their protein regeneration ability (Bassnett 2002). For this reason, maintaining high solubility and resistance to aggregation over the majority of an individual's lifetime are essential requirements of lens crystallins. Aggregation of crystallins contribute to development of cataract, the leading cause of blindness in the world.

Human γD-crystallin (HγD-Crys) is an abundant protein in the nucleus of the lens (Siezen *et al.* 1987; Brakenhoff *et al.* 1990; Lampi *et al.* 1997). In addition to the aforementioned "Greek key" aromatic pairs, two additional aromatic pairs are also present in HγD-Crys. All six aromatic pairs were shown to be important in the stability of HγD-Crys. Specific subsets of them contributed to the unfolding/refolding kinetics of HγD-Crys (Kong and King 2011). Upon refolding from high concentration of guanidine hydrochloride (GuHCl) to below 1 M, aggregation of HγD-Crys competes with the productive refolding pathway, leading to filamentous, high-molecular-weight aggregates (Kosinski-Collins and King 2003). Although these aggregates lacked the unbranched, long and twisted characteristics of genuine amyloid fibril, the filamentous morphology suggested that they could originate from an ordered polymerization of defined subunits (Kosinski-Collins and King 2003). This form of aggregate may be related to the cataract aggregate. Given that the aromatic pairs contribute to the folding and stability of the native state, we wanted to determine if they also participated in the off-pathway aggregation reaction.

α-Crystallin belongs to the small heat shock protein (sHsp) family (Ingolia and Craig 1982), and has both structural and chaperone functions in the lens (Horwitz 1992; Horwitz 2003). Human αB-crystallin (HαB-Crys) also has diverse functions in many organs and tissue types outside the lens (Graw 2009). Crystal structures of sHsps from lower organisms revealed monodisperse oligomeric structures with defined numbers of subunits (Kim *et al.* 1998; van Montfort *et al.* 2001; Hilario *et al.* 2011). A highly symmetrical 24-mer structure has also been modeled for αB-crystallin (Peschek *et al.* 2009). However, the majority of evidence supported that α-crystallin forms polydisperse oligomers of ~15-60 subunits (Haley *et al.* 2000; Laganowsky *et al.* 2010). Human α-crystallin forms complex with an αA:αB ratio of 3:1 *in vivo* (Horwitz *et al.* 1999; Srinivas *et al.* 2010). The quaternary structure is not yet fully resolved. Like other sHsps, α-crystallin cannot refold misfolded proteins or disassemble aggregates. It exerts its chaperone function by sequestering the partially-unfolded and aggregation-prone

substrates in an ATP-independent manner, preventing their aggregation (Tanksale *et al.* 2002; Evans *et al.* 2008; Acosta-Sampson and King 2010). Several regions distributed throughout the N-terminal extension and the α-crystallin domain on α-crystallin have been suggested to be substrate binding sites by cross-linking studies (Sharma *et al.* 1997; Smulders and de Jong 1997; Sharma *et al.* 2000). Based on the crystal structure of αB-crystallin, Slingsby and colleagues proposed that the shared groove and side pocket on the dimer unit worked together as potential substrate binding sites (Clark *et al.* 2011).

The conformation of α-crystallin-bound substrates is also enigmatic. A wide range of conformations from native-like to significantly unfolded have been reported, indicating some degree of substrate specificity (Das *et al.* 1996; Das *et al.* 1999; Claxton *et al.* 2008). We explored the hypothesis that the conserved Greek key aromatic pairs on HγD-Crys, when exposed in unfolded or partially unfolded state, serve as a binding site for α-crystallin recognition.

All three types of crystallins (α-, β-, and γ-crystallins) were able to form homogenous or hybrid amyloid aggregation in vitro after treated with heat or low pH (Weinreb *et al.* 2000; Meehan *et al.* 2004; Papanikolopoulou *et al.* 2008; Wang *et al.* 2010), however, evidence that amyloid aggregation is involved in cataract formation in vivo has been scarce (Sandilands *et al.* 2002). Cataract aggregates may resemble more closely those associated with inclusion bodies (Ramirez-Alvarado *et al.* 2010). The bottom line is, little is known regarding what form of aggregation is most relevant to cataract. Identifying amino acid determinants of crystallin aggregation would greatly contribute to our understanding of cataract formation.

In this study, we substituted individual residues of the highly conserved Greek key aromatic pairs on HγD-Crys with alanines, and used refolding-induced aggregation experiments to investigate whether these aromatic pairs are important in the aggregation pathway and/or the chaperone-substrate recognition with human HαB-Crys. Furthermore, HγD-Crys also contains four conserved, buried tryptophans, whose fluorescence properties differ sharply between the native state and the unfolded state (Kosinski-Collins *et al.* 2004; Chen *et al.* 2009). These tryptophans remained intact in all the mutant HγD-Crys used in this study, thus allowed the use of a mutant HαB-Crys without tryptophans, together with fluorescence spectroscopy, to probe the conformations of the substrates on the HαB-HγD complexes.

C. Materials and Methods

1. Site-specific Mutagenesis

N-terminally 6xHis tagged WT HγD-Crys pQE.1 plasmid (Qiagen) was amplified by PCR-based mutagenesis using primers encoding the desired mutations. The primers were designed from PrimerX <http://www.bioinformatics.org/primerx/> and ordered from Integrated DNA Technologies. The mutant plasmids were transformed into *E. coli* Nova Blue, purified by miniprep and sequenced at Massachusetts General Hospital to confirm the correct mutations. Untagged WT and mutant HαB-Crys pAED4 plasmids were generous gifts from Dr. Ligia Acosta-Anastasopoulos.

2. Protein Expression and Purification

For HγD-Crys, transformed *E. coli* M15 (pREP4) was cultured on shakers at 37 °C until OD_{600} reached ~1.0, then induced with 1 mM IPTG, and allowed for expression for 3 hr. The cells were pelleted, and resuspended in 300 mM NaCl, 50 mM NaH_2PO_4/Na_2HPO_4, 18 mM imidazole, pH 8.0. The cells were lysed by sonication and centrifuged at 20,000g for 30 min. The supernatant was applied to a Ni-NTA column (Qiagen) by an FPLC system (GE Healthcare), eluted with a linear gradient of imidazole from 18 to 250 mM at 4 °C. Purity was checked by SDS-PAGE. Fractions exceeding about 95% pure were pooled, dialyzed with 10 mM ammonia acetate, pH 7.0, at 4 °C, and concentrated with Amicon centrifuge filter with 10 kDa MWCO. Protein concentrations were measured by A_{280} in 5.5 M GuHCl, using extinction coefficient calculated from Expasy <http://ca.expasy.org/tools/protparam.html>. For HαB-Crys, a published protocol was followed (Acosta-Sampson and King 2010). Briefly, transformed *E. coli* BL21 (DE3) was cultured and allowed for overnight expression at 16 °C. The cells were prepared in 50 mM Tris at pH 8.5, treated with lysozyme, Dnase, deoxycholic acid, and lysed by sonication. Two rounds of ion-exchange chromatography and one round of SEC were used for purification. The protein was stored in SEC buffer (50 mM NaH_2PO_4/Na_2HPO_4, 150 mM NaCl, pH 7.0) at 4 °C and dialyzed with 10 mM ammonia acetate at pH 7.0 before use.

3. Aggregation and Aggregation Suppression Assays

WT or mutant HγD-Crys were unfolded at 500 μg/mL protein, in 5.0 M GuHCl, 100 mM NaH_2PO_4/Na_2HPO_4, 5 mM DTT, 1 mM EDTA, at pH 7.0, 37 °C, for > 5 hrs or overnight. the reactions were initiated by a 10-fold dilution of the unfolded proteins to refolding buffer with the same buffer components but without GuHCl, pre-incubated at 37 °C. Final concentrations were 50 μg/mL protein in 0.5 M GuHCl. Mixing was achieved by the initial pipetting without further stirring. The reactions proceeded at 37°C and was monitored by Abs_{350} with a Varian Cary 50 Bio UV–Visible spectrophotometer. Aggregation suppression experiments were done identically as the aggregation experiments, except that HαB-Crys was introduced to the refolding buffer at final concentration of 250 μg/mL (mass of HγD:HαB=1:5). Refolding buffer containing HαB-Crys were incubated at 37 °C for 20 min before the reaction. All experiments were done in three or more replicates.

4. HαB-HγD Complex Isolation and Fluorescence Measurement.

Published protocols were followed for these experiments (Acosta-Sampson and King 2010). Briefly, samples were collected at the end of the aggregation suppression experiments and maintained at 37 °C for a total reaction time of 2 hr. Control samples were prepared by mixing the same amount of HγD-Crys and HαB-Crys directly into buffer containing 0.5 M GuHCl. All 0.5-mL samples were added with 0.2 mL of 3.5x SEC buffer. This effectively introduced the 1x SEC buffer components (50 mM NaH_2PO_4/Na_2HPO_4, 150 mM NaCl, pH 7.0) to the samples. The samples were then 0.22 μm filtered, and analyzed by SEC (GE Life Sciences Superose 6 GL 10/300, CV = 24 mL, V_0 = 7.6 mL). Fractions containing the HαB-HγD complexes were eluted at the void volume and collected for fluorescence characterization.

Fluorescence spectra of the samples were obtained using a Hitachi F-4500 fluorometer with the following specifications: excitation wavelength 295 nm, emission wavelength 310-400 nm, scan speed 60 nm/min, excitation and emission slit band width both 10nm. HγD-only samples were prepared in 10 μg/mL protein, 100 mM NaH_2PO4/Na_2HPO4, 5 mM DTT, 1 mM EDTA, equilibrated in varying [GuHCl] at 37 °C for 24 hr. HαB-only samples were prepared in

the aforementioned SEC buffer system at 50 μg/mL protein. All samples were equilibrated to ambient temperature before fluorescence measurement.

D. Results

1. Protein Expression and Purification

Based on the N-terminally 6xHis-tagged wildtype (WT) HγD-Crys sequence, we constructed eight aromatic-to-alanine single mutant proteins on the Greek key aromatic residues (Y6A, F11A, Y45A, Y50A, Y92A, Y97A, Y133A, and Y138A) using site-directed mutagenesis (Figure 3-1). These WT and mutant HγD-Crys were expressed in *E. coli* and purified by Ni-NTA chromatography with a fast protein liquid chromatography (FPLC) system. Most mutant HγD-Crys behaved very similarly to the WT protein during expression and purification, except Y133A and Y138A HγD-Crys. These two single substitutions caused the mutant proteins to accumulate largely in the insoluble fractions and significantly lowered the soluble yields. This observation was consistent with our previous report (Kong and King 2011). However, sufficient yields of these two mutant proteins were recovered from the soluble fractions to be characterized. They maintained a β-sheet fold very similar to the WT protein. For HαB-Crys, the untagged WT and W9F/W60F (no-Trp) forms were also expressed in *E. coli* and purified by ion-exchange and size exclusion chromatography (SEC) with an FPLC system, following previously established protocols (Acosta-Sampson and King 2010). Except the anticipated difference in Abs_{280} signal, both forms behaved very similarly during expression and purification.

2. Aggregation Induced by Refolding

Upon refolding from high concentration of GuHCl to buffer with less than 1.0 M GuHCl, HγD-Crys formed high-molecular-weight filamentous aggregates (Kosinski-Collins and King 2003). As the concentration of protein increases, the partitioning between refolded native molecules and high-molecular-weight aggregates shifts to the aggregated state. This aggregation reaction can be quantified conveniently by a turbidity assay that measures the light scattering of high-molecular-weight aggregates (Acosta-Sampson and King 2010). Since the paired aromatic residues of HγD-Crys would only be fully exposed in partially folded intermediates, and might form intermolecular contacts, they are candidates for the sites involved in initiation and/or growing phases of the aggregation pathway. To test this hypothesis, refolding-induced aggregation experiments were performed on the WT and the eight aromatic mutant HγD-Crys. The experiments were done by diluting the fully unfolded HγD-Crys from 5.0 M GuHCl to 0.5 M GuHCl at pH 7.0, 37°C, at a final protein concentration of 50 μg/mL. Amount of aggregation was followed by absorbance at 350 nm.

The results showed that for WT HγD-Crys, Abs350 level reached the maximum of ~0.5 in about 200 seconds at 37 °C (Figure 3-2), consistent with previous report (Acosta-Sampson and King 2010). The results for the all eight aromatic mutant HγD-Crys were very similar to the WT protein in terms of both the kinetics and the final levels (Figure 3-2 and Figure 3-3). We initiated the experiments by pipette-mixing and the reactions proceeded without stirring. The dead time was estimated to be on the order of seconds. Considering these procedures, the kinetics of the WT and the mutant proteins were very similar. Average final aggregation levels of the mutant proteins ranged from 0.44-0.55, centered around 0.49 of the WT protein within the error range. We considered the differences between the WT and the mutant proteins insignificant. These results suggested the polypeptide chain interactions during refolding-induced HγD-Crys aggregation do not depend on any of the Greek key aromatic residues. In addition, since these mutant proteins were destabilized to different degrees (Kong and King 2011), the results also indicated that protein stability did not correlate with aggregation propensity under our experimental condition.

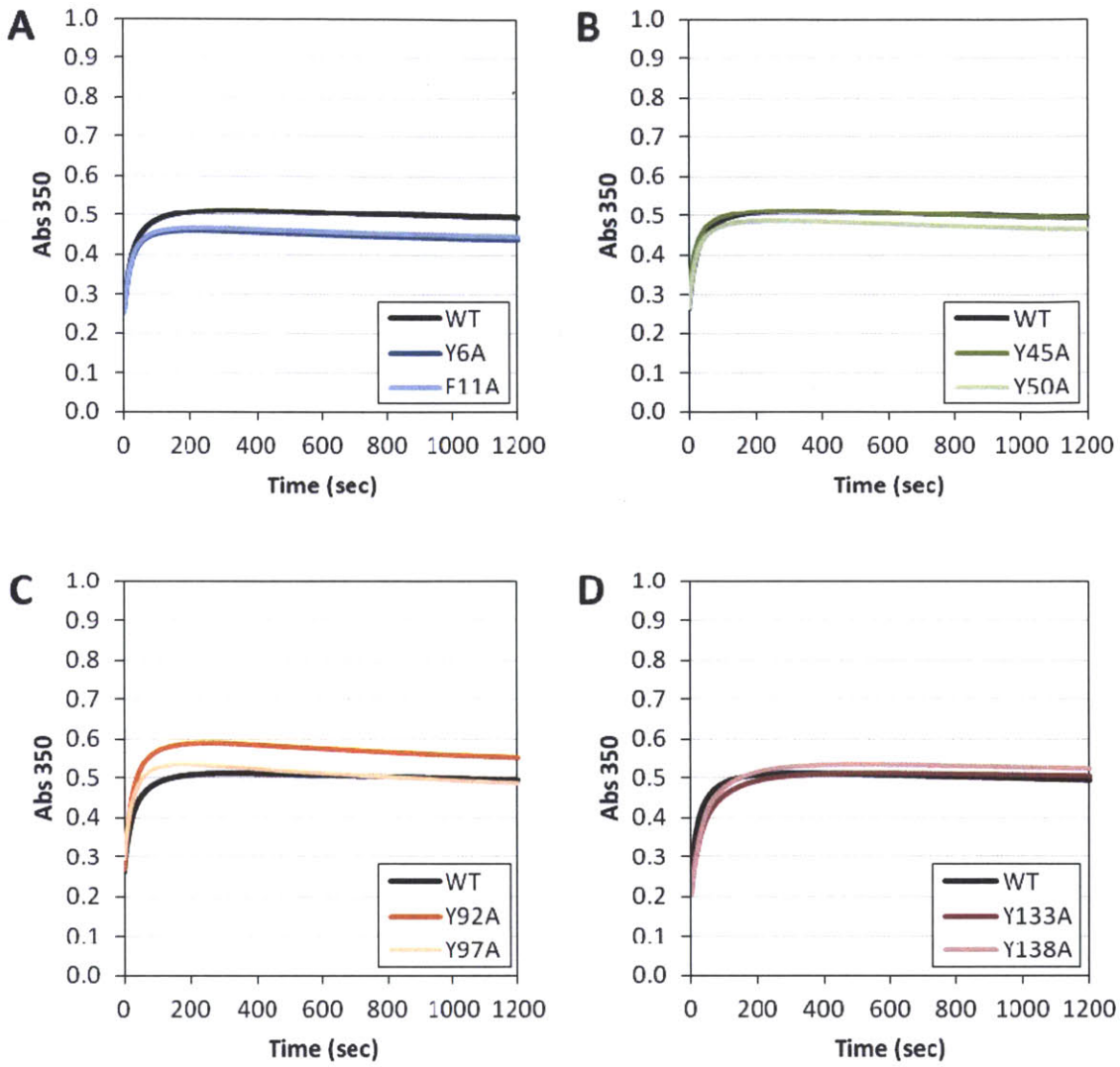


Figure 3-2 Aggregation Kinetics of WT and Mutant HγD-Crys.

The proteins were fully unfolded at 500 μg/mL protein, in 5.0 M GuHCl, 100 mM NaH_2PO_4/Na_2HPO_4, 5 mM DTT, 1 mM EDTA, at pH 7.0, 37 °C. Upon a 10-fold dilution of the unfolded proteins to refolding buffer without GuHCl, the reactions were monitored by Abs_{350}. Final concentrations were 50 μg/mL protein in 0.5 M GuHCl. Representative traces of three or more replicates for each protein are shown. (A) Y6A, F11A; (B) Y45A, Y50A; (C) Y92A, Y97A; (D) Y133A, F138A. WT traces are shown in black in each panel.

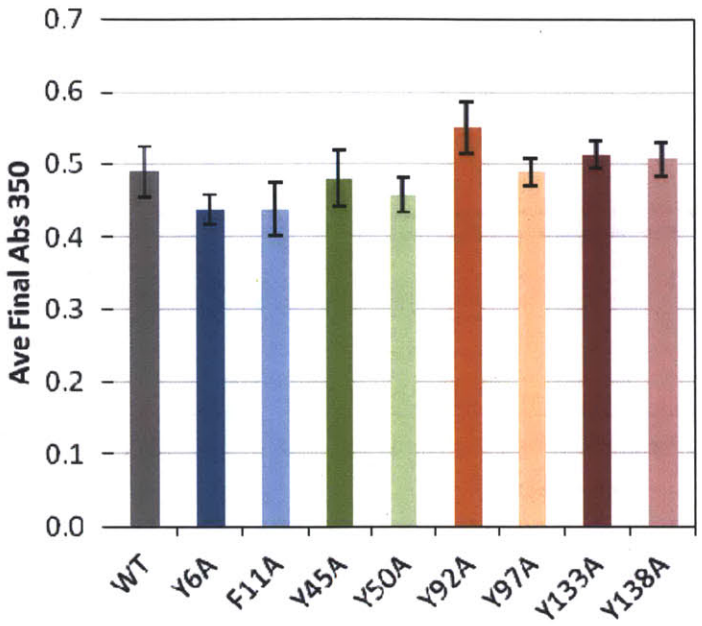


Figure 3-3 Aggregation Final Levels of WT and Mutant HγD-Crys.

Comparison of the final Abs_{350} levels of the aggregation experiments shown in Figure 3-2. Averages of three or more replicates are shown. Error bars are standard deviations. Grey: WT; blue and green: N-td mutant proteins; orange and red: C-td mutant proteins.

3. Aggregation Suppression by α-Crystallin

Purified HαB-Crys alone had many of the features of the naturally-occurring α-crystallin oligomer, including the ability to form polydisperse oligomers with comparable size and to suppress protein aggregation (Sun *et al.* 1997; Haley *et al.* 1998; Acosta-Sampson and King 2010). HαB-Crys appeared to be more hydrophobic and have greater chaperone activity than human αA-crystallin (HαA-Crys) (Sun *et al.* 1997). HαB-Crys was employed in our study to test the hypothesis that the unfolded or partially unfolded state of HγD-Crys allows the aromatic residues to be a feature for α-crystallin to recognize. The aggregation experiments described in the previous section were performed in the presence of mass ratio of 1:5 HγD:HαB. The refolding buffer was added with HαB-Crys, and then used to dilute the unfolded HγD-Crys in 5 M GuHCl.

The results showed that HαB-Crys effectively suppressed the aggregation of all the mutant proteins, with slightly different efficiencies (Figure 3-4). Unlike the aggregation of HγD-Crys alone, the initial rising of Abs_{350} was not observed in the aggregation with HαB-Crys. Figure 3-5 compared the suppression efficiencies across the WT and different mutant proteins in term of final Abs_{350} levels. WT HγD-Crys suppressed aggregation was reduced to 0.15, about 31% of the unsuppressed aggregation, consistent with previous report (Acosta-Sampson and King 2010). Suppressed aggregation levels of the C-terminal domain (C-td) mutant proteins ranged from 0.14-0.16, very similar to the WT level. For the N-terminal domain (N-td) mutant proteins, the suppressed aggregation levels ranged from 0.16-0.19, with Y6A protein having the highest value 0.19. Although the N-td mutant HγD-Crys, had slightly less effective aggregation suppression compared to the WT or C-td mutant protein, the differences were mostly within error range and considered not significant. Suppression efficiency was somewhat lower for Y6A HγD-Crys compared to the WT protein, but the effect, if significant, was small. These results indicated that the in general, recognition of partially unfolded HγD-Crys by HαB-Crys during suppression of aggregation did not depend on the Greek key aromatic residues.

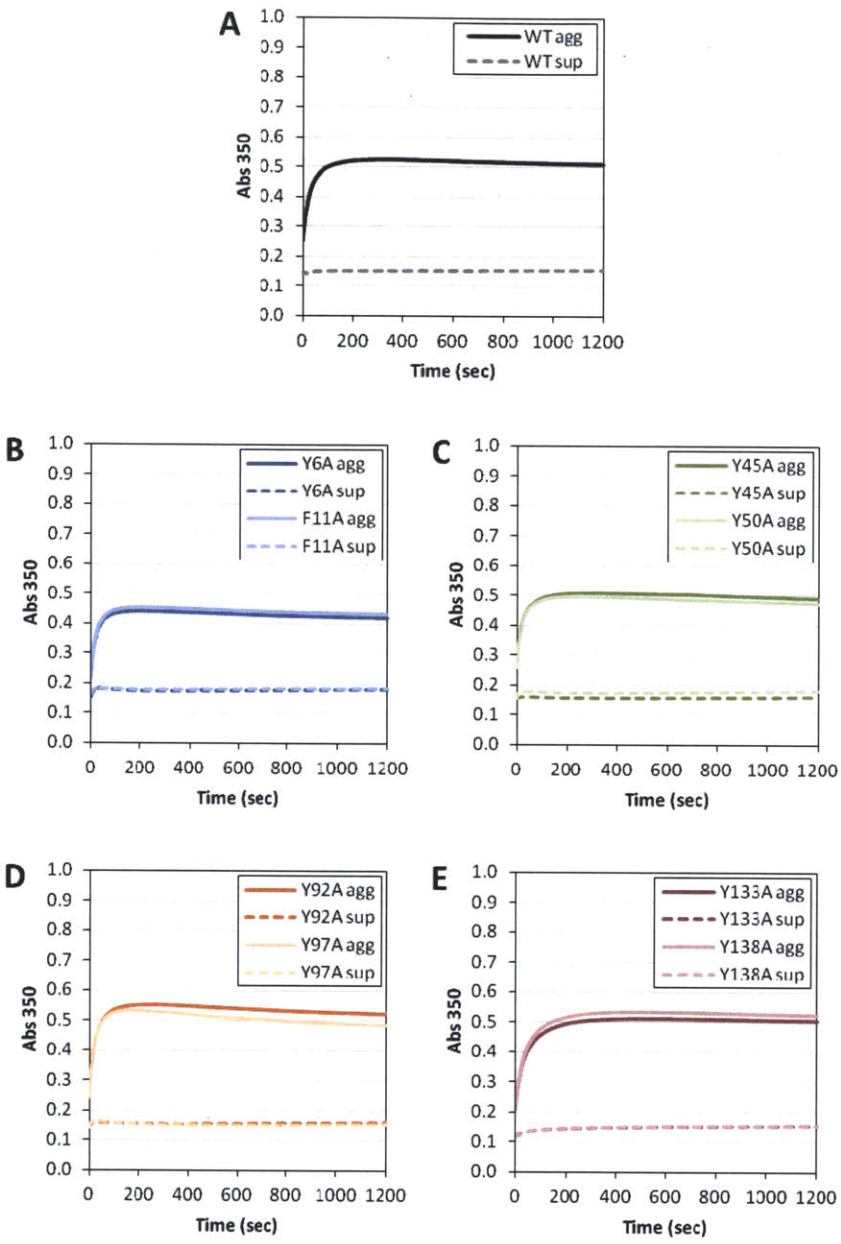


Figure 3-4 Suppression of Aggregation of WT and Mutant HγD-Crys by HαB-Crys Chaperone. The proteins were fully unfolded at 500 μg/mL protein, in 5.0 M GuHCl, 100 mM NaH_2PO_4/Na_2HPO_4, 5 mM DTT, 1 mM EDTA, at pH 7.0, 37 °C. Upon a 10-fold dilution of the unfolded proteins to refolding buffer without GuHCl, but containing HαB-Crys at a of mass ratio of 1:5 HγD:HαB. The reactions were monitored by Abs_{350}. Final concentrations were 50 μg/mL HγD-Crys, 250 μg/mL HαB-Crys in 0.5 M GuHCl. Representative traces of three or more replicates for each protein were shown. (A) WT; (B) Y6A, F11A; (C) Y45A, Y50A; (D) Y92A, Y97A; (E) Y133A, F138A. Solid traces: HγD-Crys alone; dotted traces: HγD + HαB.

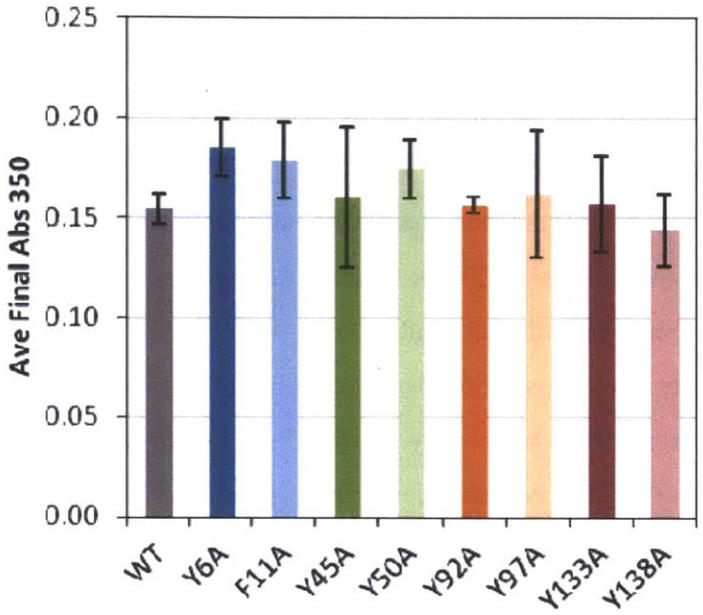


Figure 3-5 Final Suppressed Aggregation Levels of WT and Mutant HγD-Crys.
Comparison of the final Abs_{350} levels of suppression of aggregation experiments shown in Figure 3-4. Averages of three or more replicates are shown. Error bars are standard deviations. Grey: WT; blue and green: N-td mutant proteins; orange and red: C-td mutant proteins.

4. Fluorescence Analysis of the α-γ Complex

As an ATP-independent sHsp, α-crystallin cannot refold misfolded or unfolded substrates. Instead, α-crystallin binds to the substrates in the aggregation-prone conformations, leading to a chaperone-substrate complex (Boyle *et al.* 1993; Haley *et al.* 2000; Acosta-Sampson and King 2010). Intrinsic tryptophan fluorescence has been routinely used to monitored the conformation of HγD-Crys (Kosinski-Collins *et al.* 2004; Chen *et al.* 2009). α-Crystallin lacking the tryptophans ("no-Trp") has been demonstrated to be useful in probing the conformation of the chaperone-bound substrates (Acosta-Sampson and King 2010). When bound on no-Trp αA-crystallin, γ-crystallin resembled a partially unfolded state (Das *et al.* 1996). W9F and W60F HαB-Crys had some tertiary but no secondary structural changes compared to WT HαB-Crys (Liang *et al.* 1999). W9F/W60F (no-Trp) HαB-Crys retained its chaperone function but did not show any interfering tryptophan fluorescence (Figure 3-6) (Acosta-Sampson and King 2010). The same result was obtained using excitation wavelength 295 nm or 300 nm. When bound on no-Trp HαB-Crys, WT HγD-Crys was shown to be in either a partially folded conformation, or a state with its tryptophans directly interacting with HαB-Crys (Acosta-Sampson and King 2010). We took advantage of this methodology to investigate whether the aromatic substitutions on HγD-Crys affected the chaperone-bound conformations of HγD-Crys.

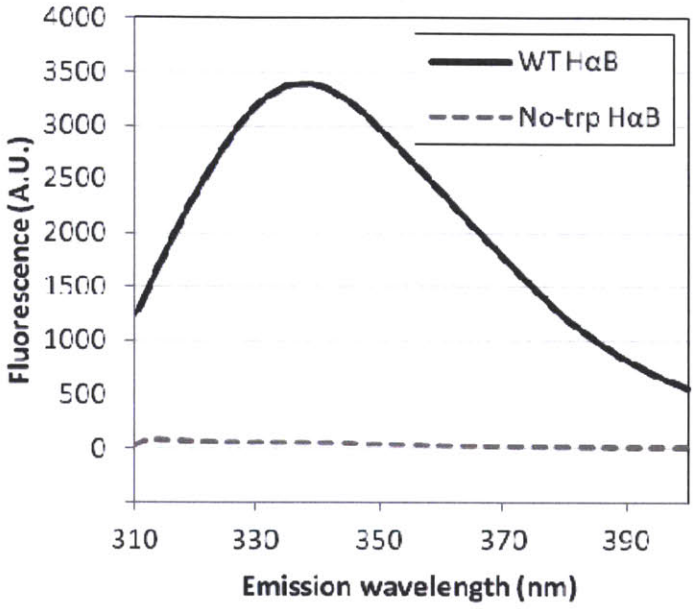


Figure 3-6 Fluorescence Spectra of WT and W9F/W60F (no-Trp) HαB-Crys.
WT and no-Trp HαB-Crys were incubated at 50 μg/mL protein, in 50 mM NaH_2PO_4/Na_2HPO_4, 150 mM NaCl, pH 7.0 at ambient temperature. Fluorescence spectra were measured at excitation wavelength 295 nm.

By SEC analysis, all eight aromatic mutant HγD-Crys formed complexes with no-Trp HαB-Crys at the end of the aggregation suppression experiments, similarly to the WT HγD-Crys. In control experiments, native WT and mutant HγD-Crys were incubated with no-Trp HαB-Crys in 0.5 M GuHCl in matching final condition of the aggregation suppression experiment, without going through the refolding/aggregation process. No complexes formed in the control experiments when analyzed by SEC, indicating that similar to the WT protein, complex formation depended on the refolding process for these aromatic mutant HγD-Crys. These complexes were purified by SEC and analyzed by fluorescence spectroscopy.

The fluorescence spectra of the complexes formed by HαB-Crys and mutant HγD-Crys were very similar to that formed by WT HγD-Crys (Figure 3-7). They all showed a characteristic maximum wavelength at 335-336 nm. The variation of amplitudes across different mutant proteins was likely due to imperfect alignment of the elution peak and fraction collection, since the fractions were not pooled. It also remains possible that this variation reflected the different numbers of substrate molecules on each HαB-mutant-HγD complex. However, this is unlikely given the similar suppression efficiency of HαB-Crys on different mutant HγD-Crys. Our results showed that substitutions of the aromatic residues on HγD-Crys had little effect on its conformation bound by HαB-Crys. Therefore, it is unlikely that these paired aromatic residues are the features recognized by HαB-Crys. Compared to HγD-Crys alone that were equilibrated in different concentrations of GuHCl, the HαB-bound conformations of the aromatic mutant HγD-Crys were clearly not in native or completely unfolded state, but rather resembled a partially unfolded state (Figure 3-7). These fluorescence spectra also indicated that the tryptophans were neither in native state, nor solvent exposed. One possible interpretation is that the tryptophans were in contact with the chaperone binding surface, and thus neither native, nor fully solvent exposed.

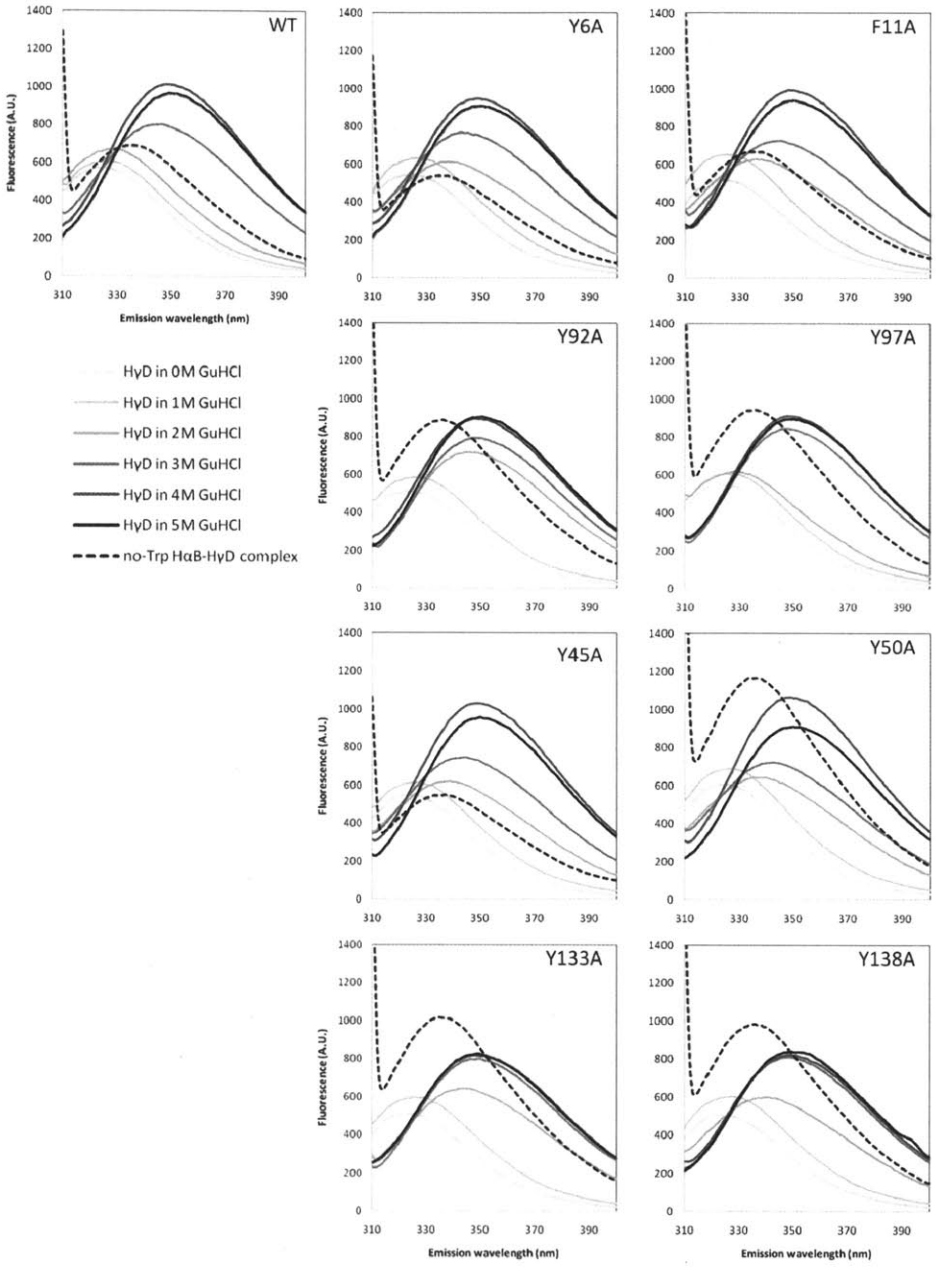


Figure 3-7 Fluorescence Spectra of the no-Trp-HαB-HγD Complexes

HαB-HγD complexes were generated by aggregation suppression experiments, isolated by SEC, and analyzed immediately. Complexes were in arbitrary protein concentration, eluted with 50 mM NaH_2PO_4/Na_2HPO_4, 150 mM NaCl, pH 7.0, (dotted traces). HγD-only samples were in 100 mM NaH_2PO_4/Na_2HPO_4, 5 mM DTT, 1 mM EDTA, fully equilibrated in varying [GuHCl]. Lighter solid traces: lower [GuHCl], darker solid traces: higher [GuHCl]. All fluorescence spectra were measured at excitation wavelength 295 nm at ambient temperature.

E. Discussion

The off-pathway refolding-induced aggregation of HγD-Crys requires unknown features of the protein that appear only in the partially unfolded intermediate. HαB-Crys does not recognize the native state of HγD-Crys, but does suppress off-pathway aggregation, presumably also through recognizing sites or sequences in partially folded substrates. As noted in the Introduction section, the partially buried aromatic pairs of the crystallin Greek key folds were candidates for such features. Our experimental results showed that partially unfolded HγD-Crys with each of the eight conserved, partially buried aromatic residues disrupted by alanine substitutions, exhibited aggregation kinetics, HαB-Crys suppression efficiencies, and chaperone-bound conformations that were all similar to the WT protein. These results make it very unlikely that the aromatic pairs are critically involved either in off-pathway aggregation or as signals for chaperone recognition.

1. Chaperone Recognition Sites

Sharma and colleagues have identified sequences in α-crystallin involved in substrate binding (Sharma *et al.* 1997; Sharma *et al.* 2000). However, the substrate features recognized by sHsps remain poorly defined. There may be specific amino acid interactions between some chaperones and substrates, including examples involving aromatic residues. Peptide binding of gp96/GRP94, an ER paralogue of Hsp90, required several aromatic residues in the peptide binding pocket and may involved specific aromatic stacking interaction (Linderoth *et al.* 2001). Periplasmic chaperone SurA preferentially bound substrates rich in aromatic residues and these aromatic residues mediated majority of the binding interactions (Xu *et al.* 2007). A single Phe substitution abolished the chaperone activity of αA-crystallin and this residue may be involved in substrate binding (Sharma *et al.* 2000; Santhoshkumar and Sharma 2001). In addition, the aggregations of specific glutamine deamidation mutant βB2-crystallins were only partially rescued or not rescued at all by α-crystallin, suggesting a potentially direct role of these glutamines in chaperone recognition (Michiel *et al.* 2010). Due to very different operating mechanisms as well as the broad substrate spectra of many chaperones, chaperone-substrate recognition may require investigation on a case-by-case basis.

In addition to the amino acid determinants, investigations on the conformation of chaperone-bound substrates are also informative to chaperone-substrate interactions. For α-crystallin, a variety of methods have been developed to separate signals from the chaperone vs. the substrates. However, diverging results have been obtained, from significantly unfolded T4 lysozyme, to partially unfolded βγ-crystallin, to native-like rhodanese (Das *et al.* 1996; Das *et al.* 1999; Claxton *et al.* 2008). In our case, the isolated HαB-WT-HγD complex showed that the substrate resembled neither the native or the completely unfolded states in solution. Acosta-Sampson and King offered a simple interpretation that the tryptophans normally buried within the crystallin domains were interacting directly with the chaperone (Acosta-Sampson and King 2010). All the Greek key aromatic mutant HγD-Crys studied here contained the four conserved, normally buried tryptophans. Since none of our aromatic substitutions altered the fluorescence spectrum of the HαB-HγD complex, this is consistent with a model in which the tryptophans are part of the binding site recognized by the chaperone. It remains possible that the recognition or complex stabilization requires additional motifs on HγD-Crys, or general features of a partially unfolded polypeptide. Although no single aromatic residue dramatically decreased the chaperone recognition and the formation of chaperone-substrate complex, the aromatic residues and additional hydrophobic residues may as a group contribute considerably to the recognition by HαB-Crys.

2. Aggregation Sites

Elimination of the Greek key aromatic residues calls for a reexamination of candidate features on a partially unfolded HγD-Crys that contribute aggregation sites. In our previous report, we suggested a HγD-Crys refolding model in which the two Greek keys in each crystallin domain fold under different kinetics, leading to intermediates with individual Greek keys (Kong and King 2011). In computational simulations of the HγD-Crys folding pathway, individual Greek keys also appeared in intermediates that leads to domain-swap aggregation (Das *et al.* 2011). A Greek key motif rarely occurs in isolation in a native protein. The side chain spacing on an exposed surface of Greek key motifs or hairpins would provide a distinct site for aggregation, given that these surfaces are buried in the native state. This is reminiscent of immunoglobulin light chain association implicated in light chain amyloidosis, where the promiscuous dimer

interface may provide novel contacts in amyloid fibril formation (Peterson *et al.* 2010). Works on single-domain HγD-Crys and unfolding GuHCl concentration requirement for aggregation suggested that the C-td of HγD-Crys was more likely than the N-td to contain aggregation determinants (Mills *et al.* 2007; Acosta-Sampson and King 2010).

The pH 7 refolding-induced aggregates of HγD-Crys showed a filamentous morphology, though lacking the features of amyloid fiber. However, amyloid-like aggregation of human γC- and γD-crystallins can also be induced by low pH (Papanikolopoulou *et al.* 2008; Wang *et al.* 2010). The paired aromatic residues remains possible determinants for this alternative aggregation mode, through the stacking stabilization of the aromatic rings (Gazit 2002).

3. Stability-Aggregation Relationship

Productive folding and aggregation are competing pathways in a network of accessible conformations for proteins (Dobson 2004). Therefore, besides direct participation in the chain-to-chain interaction, destabilization of the native state may be a second mechanism through which residues could affect aggregation propensity. In light chain amyloidosis, residue replacements increasing monomer and dimer stability protected the protein against amyloidogenesis (Baden *et al.* 2008). On the other hand, data that showed a uncorrelated or even anti-correlated relationship between instability and aggregation propensity have been reported for human γC-, γD-, and γS-crystallin (Talla *et al.* 2006; Brubaker *et al.* 2011; Sahin *et al.* 2011), blurring the stability-aggregation relationship. In the case of a mutant bovine growth hormone, impeded refolding kinetics accompanied enhanced aggregation (Brems *et al.* 1988), suggesting slow refolding rate may be a more accurate predictor of high aggregation propensity. The paired aromatic residues on HγD-Crys have been shown to be important in the stability and unfolding/refolding kinetics of HγD-Crys (Kong and King 2011). In particular, Y133A and Y138A substitutions decreased refolding rate and rendered majority of the protein insoluble during production. Therefore it was initially thought that the off-pathway aggregation for these aromatic mutant HγD-Crys could easily out-compete productive refolding in our aggregation assays. Surprisingly, none of the aromatic substitutions significantly changed the *in vitro* aggregation kinetics and final levels of HγD-Crys. Therefore, the refolding-induced aggregation

process in our condition may depend on an early refolding intermediate even before the formation of the first Greek key motif.

4. Concluding Remarks

Identifying protein aggregation determinants and chaperone-substrate recognition sites has been a long-standing problem. We eliminated the Greek key aromatic pairs in HγD-Crys as involved in such molecular events, and provided some evidence for the buried tryptophans as potential chaperone recognition candidates. Detailed studies are needed to further pinpoint the features for recognition of partially folded βγ-crystallin substrates by α-crystallin chaperone.

Chapter 4

Concluding Discussion

A. Concluding Discussion

When I first joined the laboratory of Jonathan King, simply driven by curiosity, I spent a significant time examining the crystal structure of HγD-Crys in PyMol. The large number of aromatic residues in this protein create an unmistakable dense network (Figure 2-1). I subsequently grouped these aromatic residues in space and mapped them on the topological diagram of the HγD-Crys. Interestingly, four highly conserved pairs, in addition to two seemingly random pairs, began to emerge. Believing that a high degree of conservation must accompany some type of function, I started to make mutant proteins and used the standard techniques to characterize the roles of these aromatic residues. This marked the beginning of this thesis. Upon further examination of the literature, I learned that the conservation of the Greek key paired aromatic residues in βγ-crystallins was actually described before I was born (Blundell *et al.* 1981). Together with several surrounding residues, this motif is one of the defining features of βγ-crystallins. Surprisingly, systematic studies on these conserved residues appeared to be absent. One reason could be that attention has been concentrated on the cataract-associated residues, which have almost no overlap with these conserved residues (see Chapter 1.F.1.b).

But what functions could these aromatic residues serve? HγD-Crys is a protein in the lens with long-term stability against unfolding and aggregation. For this reason, I began with the experiments measuring the stability and unfolding/refolding kinetics. The results were positive in confirming the important roles of these aromatic pairs in stability and unfolding/refolding kinetics. Most importantly, the results displayed several interesting patterns that allowed improvement of the established unfolding/refolding model of HγD-Crys (Figure 2-8). This finding contributed to β-sheet protein folding, a difficult area in protein folding. In particular, I found that only the second Greek key aromatic pair in each crystallin domain, but not the first one, was important in the refolding kinetic of HγD-Crys.

The Greek key aromatic residues are located at the turn region of a β-hairpin in each Greek key. As mentioned in the introduction, a widely held view of β-hairpin formation suggests that early β-turn formation controls the folding kinetics (Wathen and Jia 2010). According to this view, it was not surprising to find that some of these Greek key aromatic residues played important roles in the refolding kinetics. The two Greek key sequences within each crystallin domain are more divergent from each other than from their respective sister Greek keys in the

other crystallin domain. In a hypothetical evolutionary scenario, an ancestral Greek key fold depended on the aromatic pair for nucleating folding. After the duplication of the Greek keys within a crystallin domain, the first Greek key aromatic pair was no longer essential for nucleating the folding of the first Greek key. The folding of the first Greek key may be assisted by newly evolved interactions between the two Greek keys. The second Greek key may become a nucleation site for the folding of the entire crystallin domain. The differential roles of the two Greek key aromatic pairs was the most unexpected and exciting finding of the first half of this thesis. This novel experimental evidence has enhanced the resolution of the refolding model of HγD-Crys from the crystallin domain level to the Greek key level.

Due to the link to cataract, aggregation propensity is also a major concern for lens crystallins. The experiments on stability and unfolding/refolding kinetics set a solid stage for investigating the aggregation and chaperone interaction of the Greek key aromatic mutant HγD-Crys. Since the earlier results highlighted the importance of the Greek key aromatic pairs in stability and unfolding/refolding of HγD-Crys compared to the other two pairs and other unpaired aromatic residues (also see Chapter 6 for additional data). I focused on the four Greek key pairs in the subsequent experiments. As highly conserved structural motifs in the protein, the Greek key aromatic residues may also hold the secrets of the aggregation behaviors of HγD-Crys. The rationale was two-fold. First, given that the intramolecular aromatic-aromatic interaction contributed to the folding and stability of the monomer, the growth and stability of the aggregated polymer may be similarly determined by intermolecular aromatic-aromatic interactions, assuming the aggregated state retains some sorts of native-like structure. Second, because our aggregation process requires an unfolding-refolding procedure, it is possible that the conformation of the aggregated species resembles the single-domain-folded intermediate that appeared in the equilibrium unfolding/refolding experiments. Mutant proteins with slow refolding kinetics may populate this intermediate and thus become more aggregation-prone (Y45A and Y50A in Figure 2-4). Mutant proteins with an overall slow refolding kinetics are also interesting candidates for a similar rationale (Y133A and Y138A in Figure 2-4). These two hypotheses predict decreased and increased aggregations for the aromatic mutant proteins, respectively. One can apply the same rationale to the chaperone-substrate interaction between HαB-Crys and HγD-Crys and predict altered aggregation suppression efficiency by HαB-Crys.

To my surprise, the results were negative in differentiating the WT and the mutant HγD-Crys in either the aggregation or the suppression of aggregation experiments (Figure 3-2 and Figure 3-4). Nonetheless, important insights were obtained through these results. First, the aggregation process was not sensitive to changes of refolding kinetics, and thus was likely to initiate at a very early stage of Greek key refolding when the entire polypeptide chain was still largely unfolded. Second, no Greek-key-stabilizing aromatic residue was required for maintaining the aggregated structure, therefore the structural perturbation accompanying the aggregation must be beyond the Greek keys, suggesting a quite extensively non-native conformation. Third, the Greek key aromatics are unlikely to be the recognition site for HαB-Crys. Protein stability and refolding kinetics were irrelevant to HαB-Crys rescue as well. The results are consistent with the tryptophan as part of the binding sites.

In the end, much was learned about the structural roles of these Greek key aromatic residues in HγD-Crys (Figure 4-1). The residues studied in this thesis, instead of cataract-associated residues, captured my attention purely from their sequence conservation and interesting structural characteristics. Eventually they have proved to be rewarding to study, and should encourage structural studies rooted from a pure scientific curiosity. One can find many directions to expand on from this thesis study and some of them are discussed in the next section.

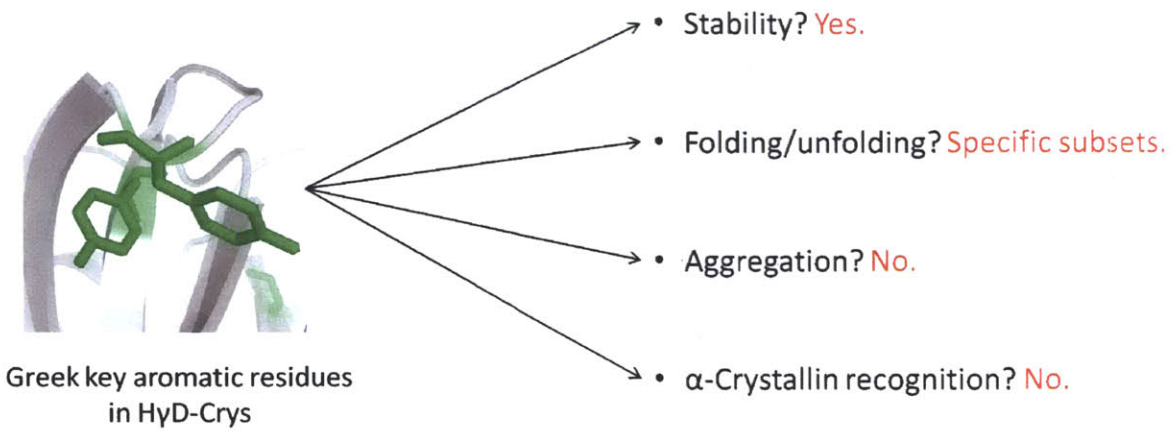


Figure 4-1 Structural Roles of the Greek Key Aromatic Residues in HγD-Crys.
This figure illustrates the key questions and answers of this thesis, centered on the Greek key paired aromatic residues.
(Image constructed from crystal structure from Basak *et al.* 2003).

B. Remaining Questions and Future Directions

1. Protein Structure under Native Condition

It was clearly demonstrated that the aromatic mutant HγD-Crys were destabilized. Although thermal denaturation and equilibrium experiments are standard methods to measure protein stability, the denaturation agents used (heat and GuHCl) were far outside the physiological range. In neutral buffer without denaturant at room temperature, the mutant HγD-Crys all showed native-like structures by far-UV CD (Figure 2-2). However, the measurement was rough and limited to secondary structure. Additional methods can be employed to gain further structural information under physiological condition. Near-UV CD can probe the tertiary structures and complement the far-UV CD measurement. Hydrogen-deuterium exchange experiments, followed by protease digestion and MS, can detect increased solvent-accessible regions. Ultimately, detailed structures including local flexibility can be obtained by NMR. The NMR results in particular can answer: 1) what kinds of structural perturbations are brought to the protein by the lack of the Greek key aromatic pairs, and 2) whether the structural perturbations (if any) are limited to local regions or propagated to the whole Greek key or even the whole crystallin domain.

2. The Folding Pathway of Crystallin Domain

Based on the kinetic refolding experiments, I proposed a HγD-Crys refolding model: the two crystallin domains refold independently, with the second Greek keys folding by nucleation of the second Greek key aromatic pairs, followed by folding of the first Greek key independent of the first Greek key aromatic pairs. An assay on the order of folding of the two Greek keys can be a test of this proposal. To achieve this, it is crucial to distinguish the folding signals of the two Greek keys within a crystallin domain, but this may not be an easy task.

Each crystallin domain only has one set of tryptophans in the second Greek key. Therefore, it is not possible to manipulate the tryptophan fluorescence to report the folding of individual Greek key. This may be overcome by fitting a multi-step exponential equation to the overall fluorescence change during folding. Isolating the N- and C-terminal crystallin domains

also simplifies the situation. Mills *et al.* from our laboratory has done such analysis and successfully isolated two kinetic folding phases for single crystallin domains. They were speculated to correspond to the folding of the two Greek keys (Mills *et al.* 2007). However, one must interpret these kinetic data with care. The injection mixing system we used had a dead time on the order of seconds, thus the half-times of the first kinetic phase for N-td and C-td, at 3 and 4.5 seconds, may not be meaningful. Due to this resolution problem, while the kinetic traces presented in Figure 2-5 and Figure 2-6 were perfectly reproducible, they limited the analysis to the qualitative level. A kinetic technique with higher resolution, possibly with a stopped-flow instrument would help to generate traces suitable for more accurate separation of the kinetic phases. Given a high-quality kinetic experimental and analytical paradigm is available, testing which kinetic phases (if any) were affected by the aromatic substitutions on isolated crystallin domains may provide a test of my proposed refolding model. Another prerequisite is unambiguous assignment of the kinetic phases to the folding of the two Greek keys.

Fluorescence resonance energy transfer (FRET) is a popular method in protein folding research. One technical caveat to FRET is that the introduced fluorophores may alter the native structure and/or folding pathway. Intrinsic tryptophan fluorescence is far superior than FRET as a reporter of protein conformation. Now the intrinsic tryptophan fluorescence is insufficient, the pros and cons of FRET should be reconsidered. 1) The locations to introduce the fluorophores should be carefully selected to be solvent exposed, thus perturbing the native structure as little as possible. 2) The candidate fluorophore sites can be further analyzed by the kinetic experiments monitored by the intrinsic tryptophan fluorescence, and selected for those least perturbing the folding kinetics. 3) Multiple sets of fluorophores can be used to ensure the "safety" of a particular site. With these precautions, the risks of using FRET would be minimal. By setting up FRET at multiple sites, the refolding order of multiple regions of the crystallin domain, in particular the two Greek keys, can then be delineated as a direct test of the proposed refolding model of crystallin domain.

3. Alternative Aggregation Pathways

The form(s) of aggregates represented by cataract has been a mystery. It is believed to resemble the amorphous bacterial inclusion bodies, but amyloid-like of aggregates were also

reported. Nothing is known about the detailed molecular structure. The *in vivo* characterization of cataract aggregates is important but difficult. On the *in vitro* track, it may be useful to learn as much as possible from different types of aggregation experiments. Besides the refolding-induced aggregation, heat and low pH have also been used to induce the aggregation of crystallins. Because the aggregation mechanisms are very different when induced by different treatments, the Greek key aromatic residues remain possible determinants of these alternative types of aggregation processes. Low-pH-induced aggregation is of particular interest because amyloid-like aggregates were formed under this condition (Papanikolopoulou *et al.* 2008; Wang *et al.* 2010). Amyloid aggregation depends on ordered side-chain packing and should respond to amino acid substitution in a definite manner. However, preliminary data comparing the pH-3-induced amyloid aggregation of the WT and the aromatic mutant HγD-Crys was inconclusive (also see Chapter 6 for additional data). Successful investigation may require very carefully designed protocols not only for the experiments but also for routine protein storage and handling.

4. Native Aggregation and Interaction with α-Crystallin

The issue of concerning whether experiment design best captures what happens under physiological condition is a recurring concern. α-Crystallin's suppression of refolding-induced aggregation experiments were convenient to conduct and did capture certain aspects of chaperone-substrate interaction. However, such dramatic unfolding and refolding processes do not happen in the lens fiber cells. Furthermore, such an abrupt change may mask the relatively mild differences between the WT and the mutant HγD-Crys. The molecular events associated with age-onset cataract start with the native states of crystallins. Both aggregation and chaperone rescue probably occur in a slow time scale. Moreau from our laboratory has performed prolonged incubation and α-crystallin interaction experiments on the cataract-associated V75D and I90F HγD-Crys under native-like condition and successful captured the differential behaviors among the WT, V75D and I90F HγD-Crys (Moreau 2011). Both V75D and I90F HγD-Crys maintained native-like secondary structures by CD, like the aromatic mutant HγD-Crys. Residues V75 and I90 are located in the buried core of the protein. Surprisingly, once incubated at 37 °C for four weeks, the soluble fraction of V75D HγD-Crys diminished, while I90F HγD-Crys formed complex with HαB-Crys (Moreau 2011). Although not associated with cataract, the

Greek key aromatic pairs are semi-exposed and are critical structural elements. It is worth testing whether their perturbations would induce aggregation and interaction with α-crystallin under native conditions.

Chapter 5

References

Abraham, A. G., N. G. Condon and E. West Gower (2006). The new epidemiology of cataract. *Ophthalmol Clin North Am.* **19**(4): 415-425.

Acosta-Sampson, L. and J. King (2010). Partially folded aggregation intermediates of human gammaD-, gammaC-, and gammaS-crystallin are recognized and bound by human alphaB-crystallin chaperone. *J Mol Biol.* **401**(1): 134-152.

Ali, M. M., S. M. Roe, C. K. Vaughan, P. Meyer, B. Panaretou, P. W. Piper, C. Prodromou and L. H. Pearl (2006). Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature.* **440**(7087): 1013-1017.

Anfinsen, C. B. (1972). The formation and stabilization of protein structure. *Biochem J.* **128**(4): 737-749.

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science.* **181**(96): 223-230.

Antuch, W., P. Guntert and K. Wuthrich (1996). Ancestral beta gamma-crystallin precursor structure in a yeast killer toxin. *Nat Struct Biol.* **3**(8): 662-665.

Aquilina, J. A., J. L. Benesch, L. L. Ding, O. Yaron, J. Horwitz and C. V. Robinson (2005). Subunit exchange of polydisperse proteins: mass spectrometry reveals consequences of alphaA-crystallin truncation. *J Biol Chem.* **280**(15): 14485-14491.

Aravind, P., A. Mishra, S. K. Suman, M. K. Jobby, R. Sankaranarayanan and Y. Sharma (2009). The betagamma-crystallin superfamily contains a universal motif for binding calcium. *Biochemistry.* **48**(51): 12180-12190.

Aravind, P., S. K. Suman, A. Mishra, Y. Sharma and R. Sankaranarayanan (2009). Three-dimensional domain swapping in nitrollin, a single-domain betagamma-crystallin from Nitrosospira multiformis, controls protein conformation and stability but not dimerization. *J Mol Biol.* **385**(1): 163-177.

Augusteyn, R. C. (2004). Dissociation is not required for alpha-crystallin's chaperone function. *Exp Eye Res.* **79**(6): 781-784.

Baden, E. M., B. A. Owen, F. C. Peterson, B. F. Volkman, M. Ramirez-Alvarado and J. R. Thompson (2008). Altered dimer interface decreases stability in an amyloidogenic protein. *J Biol Chem.* **283**(23): 15853-15860.

Baden, E. M., L. A. Sikkink and M. Ramirez-Alvarado (2009). Light chain amyloidosis - current findings and future prospects. *Curr Protein Pept Sci.* **10**(5): 500-508.

Bagby, S., S. Go, S. Inouye, M. Ikura and A. Chakrabartty (1998). Equilibrium folding intermediates of a Greek key beta-barrel protein. *J Mol Biol.* **276**(3): 669-681.

Bagneris, C., O. A. Bateman, C. E. Naylor, N. Cronin, W. C. Boelens, N. H. Keep and C. Slingsby (2009). Crystal structures of alpha-crystallin domain dimers of alphaB-crystallin and Hsp20. *J Mol Biol.* **392**(5): 1242-1252.

Baker, T. A. and R. T. Sauer (2006). ATP-dependent proteases of bacteria: recognition logic and operating principles. *Trends Biochem Sci.* **31**(12): 647-653.

Balbirnie, M., R. Grothe and D. S. Eisenberg (2001). An amyloid-forming peptide from the yeast prion Sup35 reveals a dehydrated beta-sheet structure for amyloid. *Proc Natl Acad Sci U S A.* **98**(5): 2375-2380.

Baneyx, F. and M. Mujacic (2004). Recombinant protein folding and misfolding in Escherichia coli. *Nat Biotechnol.* **22**(11): 1399-1408.

Baranova, E. V., S. D. Weeks, S. Beelen, O. V. Bukach, N. B. Gusev and S. V. Strelkov (2011). Three-dimensional structure of alpha-crystallin domain dimers of human small heat shock proteins HSPB1 and HSPB6. *J Mol Biol.* **411**(1): 110-122.

Basak, A., O. Bateman, C. Slingsby, A. Pande, N. Asherie, O. Ogun, G. B. Benedek and J. Pande (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**(5): 1137-1147.

Basha, E., G. J. Lee, L. A. Breci, A. C. Hausrath, N. R. Buan, K. C. Giese and E. Vierling (2004). The identity of proteins associated with a small heat shock protein during heat stress in vivo indicates that these chaperones protect a wide range of cellular functions. *J Biol Chem.* **279**(9): 7566-7575.

Bassnett, S. (2002). Lens organelle degradation. *Exp Eye Res.* **74**(1): 1-6.

Bassnett, S. and D. C. Beebe (1992). Coincident loss of mitochondria and nuclei during lens fiber cell differentiation. *Dev Dyn.* **194**(2): 85-93.

Bax, B., R. Lapatto, V. Nalini, H. Driessen, P. F. Lindley, D. Mahadevan, T. L. Blundell and C. Slingsby (1990). X-ray analysis of beta B2-crystallin and evolution of oligomeric lens proteins. *Nature.* **347**(6295): 776-780.

Beissinger, M., S. C. Lee, S. Steinbacher, P. Reinemer, R. Huber, M. H. Yu and R. Seckler (1995). Mutations that stabilize folding intermediates of phage P22 tailspike protein: folding in vivo and in vitro, stability, and structural context. *J Mol Biol.* **249**(1): 185-194.

Benedek, G. B. (1997). Cataract as a protein condensation disease: the Proctor Lecture. *Invest Ophthalmol Vis Sci.* **38**(10): 1911-1921.

Bennett, M. J., S. Choe and D. Eisenberg (1994). Domain swapping: entangling alliances between proteins. *Proc Natl Acad Sci U S A.* **91**(8): 3127-3131.

Bennett, M. J., M. P. Schlunegger and D. Eisenberg (1995). 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci.* **4**(12): 2455-2468.

Berg, J. M., J. L. Tymoczko and L. Stryer (2007). *Biochemistry*. New York, W.H. Freeman.

Berson, J. F., A. C. Theos, D. C. Harper, D. Tenza, G. Raposo and M. S. Marks (2003). Proprotein convertase cleavage liberates a fibrillogenic fragment of a resident glycoprotein to initiate melanosome biogenesis. *J Cell Biol.* **161**(3): 521-533.

Betts, S., C. Haase-Pettingell, K. Cook and J. King (2004). Buried hydrophobic side-chains essential for the folding of the parallel beta-helix domains of the P22 tailspike. *Protein Sci.* **13**(9): 2291-2303.

Bhat, S. P. and C. N. Nagineni (1989). alpha B subunit of lens-specific protein alpha-crystallin is present in other ocular and non-ocular tissues. *Biochem Biophys Res Commun.* **158**(1): 319-325.

Bhattacharyya, R., U. Samanta and P. Chakrabarti (2002). Aromatic-aromatic interactions in and around alpha-helices. *Protein Eng.* **15**(2): 91-100.

Bloemendal, H., W. de Jong, R. Jaenicke, N. H. Lubsen, C. Slingsby and A. Tardieu (2004). Ageing and vision: structure, stability and function of lens crystallins. *Prog Biophys Mol Biol.* **86**(3): 407-485.

Blundell, T., P. Lindley, L. Miller, D. Moss, C. Slingsby, I. Tickle, B. Turnell and G. Wistow (1981). The molecular structure and stability of the eye lens: x-ray analysis of gamma-crystallin II. *Nature.* **289**(5800): 771-777.

Bodner, R. A., T. F. Outeiro, S. Altmann, M. M. Maxwell, S. H. Cho, B. T. Hyman, P. J. McLean, A. B. Young, D. E. Housman and A. G. Kazantsev (2006). Pharmacological promotion of inclusion formation: a therapeutic approach for Huntington's and Parkinson's diseases. *Proc Natl Acad Sci U S A.* **103**(11): 4246-4251.

Bossy-Wetzel, E., R. Schwarzenbacher and S. A. Lipton (2004). Molecular pathways to neurodegeneration. *Nat Med.* **10 Suppl**: S2-9.

Bova, M. P., L. L. Ding, J. Horwitz and B. K. Fung (1997). Subunit exchange of alphaA-crystallin. *J Biol Chem.* **272**(47): 29511-29517.

Bova, M. P., H. S. McHaourab, Y. Han and B. K. Fung (2000). Subunit exchange of small heat shock proteins. Analysis of oligomer formation of alphaA-crystallin and Hsp27 by fluorescence resonance energy transfer and site-directed truncations. *J Biol Chem.* **275**(2): 1035-1042.

Bova, M. P., O. Yaron, Q. Huang, L. Ding, D. A. Haley, P. L. Stewart and J. Horwitz (1999). Mutation R120G in alphaB-crystallin, which is linked to a desmin-related myopathy, results in an irregular structure and defective chaperone-like function. *Proc Natl Acad Sci U S A.* **96**(11): 6137-6142.

Boyle, D., S. Gopalakrishnan and L. Takemoto (1993). Localization of the chaperone binding site. *Biochem Biophys Res Commun.* **192**(3): 1147-1154.

Boyle, D. and L. Takemoto (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**(1): 9-15.

Brakenhoff, R. H., H. J. Aarts, F. H. Reek, N. H. Lubsen and J. G. Schoenmakers (1990). Human gamma-crystallin genes. A gene family on its way to extinction. *J Mol Biol.* **216**(3): 519-532.

Branden, C.-I. and J. Tooze (1999). *Introduction to protein structure*. New York, Garland Pub.

Brems, D. N., S. M. Plaisted, H. A. Havel and C. S. Tomich (1988). Stabilization of an associated folding intermediate of bovine growth hormone by site-directed mutagenesis. *Proc Natl Acad Sci U S A.* **85**(10): 3367-3371.

Brennan, L. A. and M. Kantorow (2009). Mitochondrial function and redox control in the aging eye: role of MsrA and other repair systems in cataract and macular degenerations. *Exp Eye Res.* **88**(2): 195-203.

Bron, A. J., G. F. Vrensen, J. Koretz, G. Maraini and J. J. Harding (2000). The ageing lens. *Ophthalmologica.* **214**(1): 86-104.

Brown, K. A., E. P. Carpenter, K. A. Watson, J. R. Coggins, A. R. Hawkins, M. H. Koch and D. I. Svergun (2003). Twists and turns: a tale of two shikimate-pathway enzymes. *Biochem Soc Trans.* **31**(Pt 3): 543-547.

Brown, W. H. and C. S. Foote (2002). *Organic chemistry*. Forth sic Worth, Harcourt College Publishers.

Brubaker, W. D., J. A. Freites, K. J. Golchert, R. A. Shapiro, V. Morikis, D. J. Tobias and R. W. Martin (2011). Separating instability from aggregation propensity in gammaS-crystallin variants. *Biophys J.* **100**(2): 498-506.

Buchner, J. (1999). Hsp90 & Co. - a holding for folding. *Trends Biochem Sci.* **24**(4): 136-141.

Burley, S. K. and G. A. Petsko (1985). Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science.* **229**(4708): 23-28.

Burley, S. K. and G. A. Petsko (1986). Amino-aromatic interactions in proteins. *FEBS Lett.* **203**(2): 139-143.

Carrio, M. M., J. L. Corchero and A. Villaverde (1998). Dynamics of in vivo protein aggregation: building inclusion bodies in recombinant bacteria. *FEMS Microbiol Lett.* **169**(1): 9-15.

Carter, J. M., A. M. Hutcheson and R. A. Quinlan (1995). In vitro studies on the assembly properties of the lens proteins CP49, CP115: coassembly with alpha-crystallin but not with vimentin. *Exp Eye Res.* **60**(2): 181-192.

Caspers, G. J., J. A. Leunissen and W. W. de Jong (1995). The expanding small heat-shock protein family, and structure predictions of the conserved "alpha-crystallin domain". *J Mol Evol.* **40**(3): 238-248.

Chan, H. S. and K. A. Dill (1998). Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics. *Proteins.* **30**(1): 2-33.

Chang, H. C., Y. C. Tang, M. Hayer-Hartl and F. U. Hartl (2007). SnapShot: molecular chaperones, Part I. *Cell.* **128**(1): 212.

Chen, J., P. R. Callis and J. King (2009). Mechanism of the very efficient quenching of tryptophan fluorescence in human gammaD- and gammaS-crystallins: the gamma-crystallin fold may have evolved to protect tryptophan residues from ultraviolet photodamage. *Biochemistry.* **48**(17): 3708-3716.

Chen, J., S. L. Flaugh, P. R. Callis and J. King (2006). Mechanism of the highly efficient quenching of tryptophan fluorescence in human gammaD-crystallin. *Biochemistry.* **45**(38): 11552-11563.

Chen, J., D. Toptygin, L. Brand and J. King (2008). Mechanism of the efficient tryptophan fluorescence quenching in human gammaD-crystallin studied by time-resolved fluorescence. *Biochemistry.* **47**(40): 10705-10721.

Chiesi, M., S. Longoni and U. Limbruno (1990). Cardiac Alpha-Crystallin .3. Involvement during Heart Ischemia. *Molecular and Cellular Biochemistry.* **97**(2): 129-136.

Chiou, S. H., L. T. Chylack, Jr., W. H. Tung and H. F. Bunn (1981). Nonenzymatic glycosylation of bovine lens crystallins. Effect of aging. *J Biol Chem.* **256**(10): 5176-5180.

Clark, A. R., C. E. Naylor, C. Bagneris, N. H. Keep and C. Slingsby (2011). Crystal structure of R120G disease mutant of human alphaB-crystallin domain dimer shows closure of a groove. *J Mol Biol.* **408**(1): 118-134.

Claxton, D. P., P. Zou and H. S. McHaourab (2008). Structure and orientation of T4 lysozyme bound to the small heat shock protein alpha-crystallin. *J Mol Biol.* **375**(4): 1026-1039.

Collinet, B., P. Garcia, P. Minard and M. Desmadril (2001). Role of loops in the folding and stability of yeast phosphoglycerate kinase. *Eur J Biochem.* **268**(19): 5107-5118.

Cox, E. G. C., D. W. J. and J. A. S. Smith (1958). The Crystal Structure of Benzene at -3C. *Proceedings of the Royal Society of London. Series A.*(A247): 1-21.

Creighton, T. E. (1993). *Proteins : structures and molecular properties*. New York, W.H. Freeman.

Darwin, C. (1859). *On the origin of species by means of natural selection.* London,, J. Murray.

Das, K. P., L. P. Choo-Smith, J. M. Petrash and W. K. Surewicz (1999). Insight into the secondary structure of non-native proteins bound to a molecular chaperone alpha-crystallin. An isotope-edited infrared spectroscopic study. *J Biol Chem.* **274**(47): 33209-33212.

Das, K. P., J. M. Petrash and W. K. Surewicz (1996). Conformational properties of substrate proteins bound to a molecular chaperone alpha-crystallin. *J Biol Chem.* **271**(18): 10449-10452.

Das, P., J. A. King and R. Zhou (2009). beta-strand interactions at the domain interface critical for the stability of human lens gammaD-crystallin. *Protein Sci.* **19**(1): 131-140.

Das, P., J. A. King and R. Zhou (2011). Aggregation of gamma-crystallins associated with human cataracts via domain swapping at the C-terminal beta-strands. *Proc Natl Acad Sci U S A.* **108**(26): 10514-10519.

Davies, M. J. and R. J. Truscott (2001). Photo-oxidation of proteins and its role in cataractogenesis. *J Photochem Photobiol B.* **63**(1-3): 114-125.

de Jong, W. W., J. A. Leunissen and C. E. Voorter (1993). Evolution of the alpha-crystallin/small heat-shock protein family. *Mol Biol Evol.* **10**(1): 103-126.
Deechongkit, S., H. Nguyen, M. Jager, E. T. Powers, M. Gruebele and J. W. Kelly (2006). Beta-sheet folding mechanisms from perturbation energetics. *Curr Opin Struct Biol.* **16**(1): 94-101.
Delaye, M. and A. Tardieu (1983). Short-range order of crystallin proteins accounts for eye lens transparency. *Nature.* **302**(5907): 415-417.
Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry.* **29**(31): 7133-7155.
Dill, K. A. and H. S. Chan (1997). From Levinthal to pathways to funnels. *Nat Struct Biol.* **4**(1): 10-19.
Dill, K. A., K. M. Fiebig and H. S. Chan (1993). Cooperativity in protein-folding kinetics. *Proc Natl Acad Sci U S A.* **90**(5): 1942-1946.
Dinner, A. R., T. Lazaridis and M. Karplus (1999). Understanding beta-hairpin formation. *Proc Natl Acad Sci U S A.* **96**(16): 9068-9073.
Dinner, A. R., A. Sali, L. J. Smith, C. M. Dobson and M. Karplus (2000). Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem Sci.* **25**(7): 331-339.
Dobson, C. M. (2001). The structural basis of protein folding and its links with human disease. *Philos Trans R Soc Lond B Biol Sci.* **356**(1406): 133-145.
Dobson, C. M. (2003). Protein folding and disease: a view from the first Horizon Symposium. *Nat Rev Drug Discov.* **2**(2): 154-160.
Dobson, C. M. (2004). Principles of protein folding, misfolding and aggregation. *Semin Cell Dev Biol.* **15**(1): 3-16.
Du, D., M. J. Tucker and F. Gai (2006). Understanding the mechanism of beta-hairpin folding via phi-value analysis. *Biochemistry.* **45**(8): 2668-2678.
Dunker, A. K., M. S. Cortese, P. Romero, L. M. Iakoucheva and V. N. Uversky (2005). Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* **272**(20): 5129-5148.
Dyer, R. B., S. J. Maness, E. S. Peterson, S. Franzen, R. M. Fesinmeyer and N. H. Andersen (2004). The mechanism of beta-hairpin formation. *Biochemistry.* **43**(36): 11560-11566.
Ehrnsperger, M., S. Graber, M. Gaestel and J. Buchner (1997). Binding of non-native protein to Hsp25 during heat shock creates a reservoir of folding intermediates for reactivation. *EMBO J.* **16**(2): 221-229.
Evans, P., C. Slingsby and B. A. Wallace (2008). Association of partially folded lens betaB2-crystallins with the alpha-crystallin molecular chaperone. *Biochem J.* **409**(3): 691-699.
Evans, P., K. Wyatt, G. J. Wistow, O. A. Bateman, B. A. Wallace and C. Slingsby (2004). The P23T cataract mutation causes loss of solubility of folded gammaD-crystallin. *J Mol Biol.* **343**(2): 435-444.
Fagerholm, P. P., B. T. Philipson and B. Lindstrom (1981). Normal human lens - the distribution of protein. *Exp Eye Res.* **33**(6): 615-620.
Falsone, S. F., S. Leptihn, A. Osterauer, M. Haslbeck and J. Buchner (2004). Oncogenic mutations reduce the stability of SRC kinase. *J Mol Biol.* **344**(1): 281-291.
Fatima, U., S. Sharma and P. Guptasarma (2010). Structures of differently aggregated and precipitated forms of gamma B crystallin: an FTIR spectroscopic and EM study. *Protein Pept Lett.* **17**(9): 1155-1162.

Feil, I. K., M. Malfois, J. Hendle, H. van Der Zandt and D. I. Svergun (2001). A novel quaternary structure of the dimeric alpha-crystallin domain with chaperone-like activity. *J Biol Chem.* **276**(15): 12024-12029.

Feldman, D. E., C. Spiess, D. E. Howard and J. Frydman (2003). Tumorigenic mutations in VHL disrupt folding in vivo by interfering with chaperonin binding. *Mol Cell.* **12**(5): 1213-1224.

Fersht, A. R. (1997). Nucleation mechanisms in protein folding. *Curr Opin Struct Biol.* **7**(1): 3-9.

Flaugh, S. L., M. S. Kosinski-Collins and J. King (2005). Contributions of hydrophobic domain interface interactions to the folding and stability of human gammaD-crystallin. *Protein Sci.* **14**(3): 569-581.

Flaugh, S. L., M. S. Kosinski-Collins and J. King (2005). Interdomain side-chain interactions in human gammaD crystallin influencing folding and stability. *Protein Sci.* **14**(8): 2030-2043.

Flaugh, S. L., I. A. Mills and J. King (2006). Glutamine deamidation destabilizes human gammaD-crystallin and lowers the kinetic barrier to unfolding. *J Biol Chem.* **281**(41): 30782-30793.

Fowler, D. M., A. V. Koulov, C. Alory-Jost, M. S. Marks, W. E. Balch and J. W. Kelly (2006). Functional amyloid formation within mammalian tissue. *PLoS Biol.* **4**(1): e6.

Franzmann, T. M., M. Wuhr, K. Richter, S. Walter and J. Buchner (2005). The activation mechanism of Hsp26 does not require dissociation of the oligomer. *J Mol Biol.* **350**(5): 1083-1093.

Fu, L. and J. J. Liang (2002). Unfolding of human lens recombinant betaB2- and gammaC-crystallins. *J Struct Biol.* **139**(3): 191-198.

Gazit, E. (2002). A possible role for pi-stacking in the self-assembly of amyloid fibrils. *FASEB J.* **16**(1): 77-83.

Gilbert, S. F. (2000). *Developmental biology*. Sunderland, Mass., Sinauer Associates.

Goldschmidt, L., P. K. Teng, R. Riek and D. Eisenberg (2010). Identifying the amylome, proteins capable of forming amyloid-like fibrils. *Proc Natl Acad Sci U S A.* **107**(8): 3487-3492.

Grantcharova, V. P., D. S. Riddle, J. V. Santiago and D. Baker (1998). Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nat Struct Biol.* **5**(8): 714-720.

Graw, J. (2009). Genetics of crystallins: cataract and beyond. *Exp Eye Res.* **88**(2): 173-189.

Griffiths-Jones, S. R. and M. S. Searle (2000). Structure, Folding, and Energetics of Cooperative Interactions between the β-Strands of a de Novo Designed Three-Stranded Antiparallel β-Sheet Peptide. *J Am Chem Soc.* **122**: 8350-8356.

Grigoryan, G. and A. E. Keating (2008). Structural specificity in coiled-coil interactions. *Curr Opin Struct Biol.* **18**(4): 477-483.

Grossweiner, L. I. (1984). Photochemistry of proteins: a review. *Curr Eye Res.* **3**(1): 137-144.

Hains, P. G. and R. J. Truscott (2007). Post-translational modifications in the nuclear region of young, aged, and cataract human lenses. *J Proteome Res.* **6**(10): 3935-3943.

Haley, D. A., M. P. Bova, Q. L. Huang, H. S. McHaourab and P. L. Stewart (2000). Small heat-shock protein structures reveal a continuum from symmetric to variable assemblies. *J Mol Biol.* **298**(2): 261-272.

Haley, D. A., J. Horwitz and P. L. Stewart (1998). The small heat-shock protein, alphaB-crystallin, has a variable quaternary structure. *J Mol Biol.* **277**(1): 27-35.

Hamill, S. J., E. Cota, C. Chothia and J. Clarke (2000). Conservation of folding and stability within a protein family: the tyrosine corner as an evolutionary cul-de-sac. *J Mol Biol.* **295**(3): 641-649.

Hanson, S. R., A. Hasan, D. L. Smith and J. B. Smith (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**(2): 195-207.

Hanson, S. R., D. L. Smith and J. B. Smith (1998). Deamidation and disulfide bonding in human lens gamma-crystallins. *Exp Eye Res.* **67**(3): 301-312.

Harrington, D. J., K. Adachi and W. E. Royer, Jr. (1997). The high resolution crystal structure of deoxyhemoglobin S. *J Mol Biol.* **272**(3): 398-407.

Harrington, V., S. McCall, S. Huynh, K. Srivastava and O. P. Srivastava (2004). Crystallins in water soluble-high molecular weight protein fractions and water insoluble protein fractions in aging and cataractous human lenses. *Mol Vis.* **10**: 476-489.

Harrison, C. J., M. Hayer-Hartl, M. Di Liberto, F. Hartl and J. Kuriyan (1997). Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK. *Science.* **276**(5311): 431-435.

Hartl, F. U. and M. Hayer-Hartl (2009). Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol.* **16**(6): 574-581.

Haslbeck, M., N. Braun, T. Stromer, B. Richter, N. Model, S. Weinkauf and J. Buchner (2004). Hsp42 is the general small heat shock protein in the cytosol of Saccharomyces cerevisiae. *EMBO J.* **23**(3): 638-649.

Haslbeck, M., T. Franzmann, D. Weinfurtner and J. Buchner (2005). Some like it hot: the structure and function of small heat-shock proteins. *Nat Struct Mol Biol.* **12**(10): 842-846.

Haslbeck, M., S. Walke, T. Stromer, M. Ehrnsperger, H. E. White, S. Chen, H. R. Saibil and J. Buchner (1999). Hsp26: a temperature-regulated chaperone. *EMBO J.* **18**(23): 6744-6751.

Hecht, M. H., J. M. Sturtevant and R. T. Sauer (1984). Effect of single amino acid replacements on the thermal stability of the NH2-terminal domain of phage lambda repressor. *Proc Natl Acad Sci U S A.* **81**(18): 5685-5689.

Hemmingsen, J. M., K. M. Gernert, J. S. Richardson and D. C. Richardson (1994). The tyrosine corner: a feature of most Greek key beta-barrel proteins. *Protein Sci.* **3**(11): 1927-1937.

Heon, E., M. Priston, D. F. Schorderet, G. D. Billingsley, P. O. Girard, N. Lubsen and F. L. Munier (1999). The gamma-crystallins and human cataracts: a puzzle made clearer. *Am J Hum Genet.* **65**(5): 1261-1267.

Hilario, E., F. J. Martin, M. C. Bertolini and L. Fan (2011). Crystal structures of Xanthomonas small heat shock protein provide a structural basis for an active molecular chaperone oligomer. *J Mol Biol.* **408**(1): 74-86.

Hillier, B. J., H. M. Rodriguez and L. M. Gregoret (1998). Coupling protein stability and protein function in Escherichia coli CspA. *Fold Des.* **3**(2): 87-93.

Hobza, P., H. L. Selzle and E. W. Schlag (1996). Potential Energy Surface for the Benzene Dimer. Results of ab Initio CCSD(T) Calculations Show Two Nearly Isoenergetic Structures: T-Shaped and Parallel-Displaced. *J Phys Chem.* **100**(48): 18790-18794.

Horwich, A. L., W. A. Fenton, E. Chapman and G. W. Farr (2007). Two families of chaperonin: physiology and mechanism. *Annu Rev Cell Dev Biol.* **23**: 115-145.

Horwitz, J. (1992). Alpha-crystallin can function as a molecular chaperone. *Proc Natl Acad Sci U S A.* **89**(21): 10449-10453.

Horwitz, J. (2000). The function of alpha-crystallin in vision. *Semin Cell Dev Biol.* **11**(1): 53-60.

Horwitz, J. (2003). Alpha-crystallin. *Exp Eye Res.* **76**(2): 145-153.
Horwitz, J., M. P. Bova, L. L. Ding, D. A. Haley and P. L. Stewart (1999). Lens alpha-crystallin: function and structure. *Eye (Lond).* **13 (Pt 3b)**: 403-408.
Iconomidou, V. A., G. Vriend and S. J. Hamodrakas (2000). Amyloids protect the silkmoth oocyte and embryo. *FEBS Lett.* **479**(3): 141-145.
Ingolia, T. D. and E. A. Craig (1982). Four small Drosophila heat shock proteins are related to each other and to mammalian alpha-crystallin. *Proc Natl Acad Sci U S A.* **79**(7): 2360-2364.
Itoh, K. and M. Sasai (2008). Cooperativity, connectivity, and folding pathways of multidomain proteins. *Proc Natl Acad Sci U S A.* **105**(37): 13865-13870.
Jaenicke, R. (1996). Stability and folding of ultrastable proteins: eye lens crystallins and enzymes from thermophiles. *Faseb J.* **10**(1): 84-92.
Jaenicke, R. and G. Bohm (1998). The stability of proteins in extreme environments. *Curr Opin Struct Biol.* **8**(6): 738-748.
Jaenicke, R. and C. Slingsby (2001). Lens crystallins and their microbial homologs: structure, stability, and function. *Crit Rev Biochem Mol Biol.* **36**(5): 435-499.
Jakob, U., M. Gaestel, K. Engel and J. Buchner (1993). Small heat shock proteins are molecular chaperones. *J Biol Chem.* **268**(3): 1517-1520.
Jedziniak, J. A., J. H. Kinoshita, E. M. Yates, L. O. Hocker and G. B. Benedek (1973). On the presence and mechanism of formation of heavy molecular weight aggregates in human normal and cataractous lenses. *Exp Eye Res.* **15**(2): 185-192.
Jimenez, J. L., E. J. Nettleton, M. Bouchard, C. V. Robinson, C. M. Dobson and H. R. Saibil (2002). The protofilament structure of insulin amyloid fibrils. *Proc Natl Acad Sci U S A.* **99**(14): 9196-9201.
Kabsch, W. and J. Vandekerckhove (1992). Structure and function of actin. *Annu Rev Biophys Biomol Struct.* **21**: 49-76.
Kang, T. S. and R. M. Kini (2009). Structural determinants of protein folding. *Cell Mol Life Sci.*
Kannan, N. and S. Vishveshwara (2000). Aromatic clusters: a determinant of thermal stability of thermophilic proteins. *Protein Eng.* **13**(11): 753-761.
Kappe, G., E. Franck, P. Verschuure, W. C. Boelens, J. A. Leunissen and W. W. de Jong (2003). The human genome encodes 10 alpha-crystallin-related small heat shock proteins: HspB1-10. *Cell Stress Chaperones.* **8**(1): 53-61.
Kato, K., H. Shinohara, N. Kurobe, S. Goto, Y. Inaguma and K. Ohshima (1991). Immunoreactive alpha A crystallin in rat non-lenticular tissues detected with a sensitive immunoassay method. *Biochim Biophys Acta.* **1080**(2): 173-180.
Kennaway, C. K., J. L. Benesch, U. Gohlke, L. Wang, C. V. Robinson, E. V. Orlova, H. R. Saibil and N. H. Keep (2005). Dodecameric structure of the small heat shock protein Acr1 from Mycobacterium tuberculosis. *J Biol Chem.* **280**(39): 33419-33425.
Kim, K. K., R. Kim and S. H. Kim (1998). Crystal structure of a small heat-shock protein. *Nature.* **394**(6693): 595-599.
Kim, Y. I., I. Levchenko, K. Fraczkowska, R. V. Woodruff, R. T. Sauer and T. A. Baker (2001). Molecular determinants of complex formation between Clp/Hsp100 ATPases and the ClpP peptidase. *Nat Struct Biol.* **8**(3): 230-233.
Kishimoto, A., K. Hasegawa, H. Suzuki, H. Taguchi, K. Namba and M. Yoshida (2004). beta-Helix is a likely core structure of yeast prion Sup35 amyloid fibers. *Biochem Biophys Res Commun.* **315**(3): 739-745.

Klunk, W. E., R. F. Jacob and R. P. Mason (1999). Quantifying amyloid by congo red spectral shift assay. *Methods Enzymol.* **309**: 285-305.

Kmoch, S., J. Brynda, B. Asfaw, K. Bezouska, P. Novak, P. Rezacova, L. Ondrova, M. Filipec, J. Sedlacek and M. Elleder (2000). Link between a novel human gammaD-crystallin allele and a unique cataract phenotype explained by protein crystallography. *Hum Mol Genet.* **9**(12): 1779-1786.

Kokke, B. P., M. R. Leroux, E. P. Candido, W. C. Boelens and W. W. de Jong (1998). Caenorhabditis elegans small heat-shock proteins Hsp12.2 and Hsp12.3 form tetramers and have no chaperone-like activity. *FEBS Lett.* **433**(3): 228-232.

Komar, A. A. (2009). A pause for thought along the co-translational folding pathway. *Trends Biochem Sci.* **34**(1): 16-24.

Kong, F. and J. King (2011). Contributions of aromatic pairs to the folding and stability of long-lived human gammaD-crystallin. *Protein Sci.* **20**(3): 513-528.

Kosinski-Collins, M. S., S. L. Flaugh and J. King (2004). Probing folding and fluorescence quenching in human gammaD crystallin Greek key domains using triple tryptophan mutant proteins. *Protein Sci.* **13**(8): 2223-2235.

Kosinski-Collins, M. S. and J. King (2003). In vitro unfolding, refolding, and polymerization of human gammaD crystallin, a protein involved in cataract formation. *Protein Sci.* **12**(3): 480-490.

Kubota, S., H. Kubota and K. Nagata (2006). Cytosolic chaperonin protects folding intermediates of Gbeta from aggregation by recognizing hydrophobic beta-strands. *Proc Natl Acad Sci U S A.* **103**(22): 8360-8365.

Kuivaniemi, H., G. Tromp and D. J. Prockop (1991). Mutations in collagen genes: causes of rare and some common diseases in humans. *FASEB J.* **5**(7): 2052-2060.

Kuo, N. N., J. J. Huang, J. Miksovska, R. P. Chen, R. W. Larsen and S. I. Chan (2005). Effects of turn stability on the kinetics of refolding of a hairpin in a beta-sheet. *J Am Chem Soc.* **127**(48): 16945-16954.

Laganowsky, A., J. L. Benesch, M. Landau, L. Ding, M. R. Sawaya, D. Cascio, Q. Huang, C. V. Robinson, J. Horwitz and D. Eisenberg (2010). Crystal structures of truncated alphaA and alphaB crystallins reveal structural mechanisms of polydispersity important for eye lens function. *Protein Sci.* **19**(5): 1031-1043.

Lampi, K. J., K. K. Amyx, P. Ahmann and E. A. Steel (2006). Deamidation in human lens betaB2-crystallin destabilizes the dimer. *Biochemistry.* **45**(10): 3146-3153.

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

Lampi, K. J., Z. Ma, S. R. Hanson, M. Azuma, M. Shih, T. R. Shearer, D. L. Smith, J. B. Smith and L. L. David (1998). Age-related changes in human lens crystallins identified by two-dimensional electrophoresis and mass spectrometry. *Exp Eye Res.* **67**(1): 31-43.

Lampi, K. J., Z. Ma, M. Shih, T. R. Shearer, J. B. Smith, D. L. Smith and L. L. David (1997). Sequence analysis of betaA3, betaB3, and betaA4 crystallins completes the identification of the major proteins in young human lens. *J Biol Chem.* **272**(4): 2268-2275.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins (2007). Clustal W and Clustal X version 2.0. *Bioinformatics.* **23**(21): 2947-2948.

Lee, S., M. E. Sowa, J. M. Choi and F. T. Tsai (2004). The ClpB/Hsp104 molecular chaperone-a protein disaggregating machine. *J Struct Biol.* **146**(1-2): 99-105.

Lee, S., M. E. Sowa, Y. H. Watanabe, P. B. Sigler, W. Chiu, M. Yoshida and F. T. Tsai (2003). The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state. *Cell.* **115**(2): 229-240.

Lehmann, M. and M. Wyss (2001). Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. *Curr Opin Biotechnol.* **12**(4): 371-375.

Leroux, M. R., B. J. Ma, G. Batelier, R. Melki and E. P. Candido (1997). Unique structural features of a novel class of small heat shock proteins. *J Biol Chem.* **272**(19): 12847-12853.

Leroux, M. R., R. Melki, B. Gordon, G. Batelier and E. P. Candido (1997). Structure-function studies on small heat shock protein oligomeric assembly and interaction with unfolded polypeptides. *J Biol Chem.* **272**(39): 24646-24656.

LeVine, H., 3rd (1999). Quantification of beta-sheet amyloid fibril structures with thioflavin T. *Methods Enzymol.* **309**: 274-284.

Levinthal, C. (1969). How to fold graciously. *Mossbauer Spectroscopy in Biological Systems Proceedings*. University of Illinois at Urbana. **67:** 22-24.

Levitt, M. and M. F. Perutz (1988). Aromatic rings act as hydrogen bond acceptors. *J Mol Biol.* **201**(4): 751-754.

Li, F., S. Wang, C. Gao, S. Liu, B. Zhao, M. Zhang, S. Huang, S. Zhu and X. Ma (2008). Mutation G61C in the CRYGD gene causing autosomal dominant congenital coralliform cataracts. *Mol Vis.* **14**: 378-386.

Liang, J. J. (2004). Interactions and chaperone function of alphaA-crystallin with T5P gammaC-crystallin mutant. *Protein Sci.* **13**(9): 2476-2482.

Liang, J. J., T. X. Sun and N. J. Akhtar (1999). Spectral contribution of the individual tryptophan of alphaB-crystallin: a study by site-directed mutagenesis. *Protein Sci.* **8**(12): 2761-2764.

Linderoth, N. A., M. N. Simon, J. F. Hainfeld and S. Sastry (2001). Binding of antigenic peptide to the endoplasmic reticulum-resident protein gp96/GRP94 heat shock chaperone occurs in higher order complexes. Essential role of some aromatic amino acid residues in the peptide-binding site. *J Biol Chem.* **276**(14): 11049-11054.

Liu, Y. and D. Eisenberg (2002). 3D domain swapping: as domains continue to swap. *Protein Sci.* **11**(6): 1285-1299.

London, J., C. Skrzynia and M. E. Goldberg (1974). Renaturation of Escherichia coli tryptophanase after exposure to 8 M urea. Evidence for the existence of nucleation centers. *Eur J Biochem.* **47**(2): 409-415.

Longoni, S., S. Lattonen, G. Bullock and M. Chiesi (1990). Cardiac Alpha-Crystallin .2. Intracellular-Localization. *Molecular and Cellular Biochemistry.* **97**(2): 121-128.

Ma, J. C. and D. A. Dougherty (1997). The Cationminus signpi Interaction. *Chem Rev.* **97**(5): 1303-1324.

Ma, Z., S. R. Hanson, K. J. Lampi, L. L. David, D. L. Smith and J. B. Smith (1998). Age-related changes in human lens crystallins identified by HPLC and mass spectrometry. *Exp Eye Res.* **67**(1): 21-30.

MacDonald, J. T., A. G. Purkiss, M. A. Smith, P. Evans, J. M. Goodfellow and C. Slingsby (2005). Unfolding crystallins: the destabilizing role of a beta-hairpin cysteine in betaB2-crystallin by simulation and experiment. *Protein Sci.* **14**(5): 1282-1292.

Maji, S. K., M. H. Perrin, M. R. Sawaya, S. Jessberger, K. Vadodaria, R. A. Rissman, P. S. Singru, K. P. Nilsson, R. Simon, D. Schubert, D. Eisenberg, J. Rivier, P. Sawchenko, W. Vale and R. Riek (2009). Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science.* **325**(5938): 328-332.

Makin, O. S. and L. C. Serpell (2005). Structures for amyloid fibrils. *Febs J.* **272**(23): 5950-5961.

Martin, A., T. A. Baker and R. T. Sauer (2008). Pore loops of the AAA+ ClpX machine grip substrates to drive translocation and unfolding. *Nat Struct Mol Biol.* **15**(11): 1147-1151.

Mayer, M. P., S. Rudiger and B. Bukau (2000). Molecular basis for interactions of the DnaK chaperone with substrates. *Biol Chem.* **381**(9-10): 877-885.

Mayr, E. M., R. Jaenicke and R. Glockshuber (1994). Domain interactions and connecting peptides in lens crystallins. *J Mol Biol.* **235**(1): 84-88.

Mayr, E. M., R. Jaenicke and R. Glockshuber (1997). The domains in gammaB-crystallin: identical fold-different stabilities. *J Mol Biol.* **269**(2): 260-269.

McCaldon, P. and P. Argos (1988). Oligopeptide biases in protein sequences and their use in predicting protein coding regions in nucleotide sequences. *Proteins.* **4**(2): 99-122.

McCarty, C. A. and H. R. Taylor (2002). A review of the epidemiologic evidence linking ultraviolet radiation and cataracts. *Dev Ophthalmol.* **35**: 21-31.

McCleverty, C. J., E. Koesema, A. Patapoutian, S. A. Lesley and A. Kreusch (2006). Crystal structure of the human TRPV2 channel ankyrin repeat domain. *Protein Sci.* **15**(9): 2201-2206.

McGaughey, G. B., M. Gagne and A. K. Rappe (1998). pi-Stacking interactions. Alive and well in proteins. *J Biol Chem.* **273**(25): 15458-15463.

Meehan, S., Y. Berry, B. Luisi, C. M. Dobson, J. A. Carver and C. E. MacPhee (2004). Amyloid fibril formation by lens crystallin proteins and its implications for cataract formation. *J Biol Chem.* **279**(5): 3413-3419.

Messina-Baas, O. M., L. M. Gonzalez-Huerta and S. A. Cuevas-Covarrubias (2006). Two affected siblings with nuclear cataract associated with a novel missense mutation in the CRYGD gene. *Mol Vis.* **12**: 995-1000.

Meyer, E. A., R. K. Castellano and F. Diederich (2003). Interactions with aromatic rings in chemical and biological recognition. *Angew Chem Int Ed Engl.* **42**(11): 1210-1250.

Michiel, M., E. Duprat, F. Skouri-Panet, J. A. Lampi, A. Tardieu, K. J. Lampi and S. Finet (2010). Aggregation of deamidated human betaB2-crystallin and incomplete rescue by alpha-crystallin chaperone. *Exp Eye Res.* **90**(6): 688-698.

Mills-Henry, I. (2007). Stability, unfolding, and aggregation of the gamma D and gamma S human eye lens crystallins. *Department of Biology*, Massachusetts Institute of Technology.

Mills, I. A., S. L. Flaugh, M. S. Kosinski-Collins and J. A. King (2007). Folding and stability of the isolated Greek key domains of the long-lived human lens proteins gammaD-crystallin and gammaS-crystallin. *Protein Sci.* **16**(11): 2427-2444.

Mitraki, A. and J. King (1989). Protein Folding Intermediates and Inclusion Body Formation. *Nat Biotechnol.* **7**(7): 690-697.

Moreau, K. L. (2011). The characterization of human [gamma]D-crystallin mutants and their differential interactions with the lens chaperone [alpha]B-crystallin. *Department of Biology.* Cambridge, MA, Massachusetts Institute of Technology.

Moreau, K. L. and J. King (2009). Hydrophobic core mutations associated with cataract development in mice destabilize human gammaD-crystallin. *J Biol Chem.* **284**(48): 33285-33295.

Morris, A. M., T. M. Treweek, J. A. Aquilina, J. A. Carver and M. J. Walker (2008). Glutamic acid residues in the C-terminal extension of small heat shock protein 25 are critical for structural and functional integrity. *FEBS J.* **275**(23): 5885-5898.

Muchowski, P. J. and J. I. Clark (1998). ATP-enhanced molecular chaperone functions of the small heat shock protein human alphaB crystallin. *Proc Natl Acad Sci U S A.* **95**(3): 1004-1009.

Najmudin, S., V. Nalini, H. P. Driessen, C. Slingsby, T. L. Blundell, D. S. Moss and P. F. Lindley (1993). Structure of the bovine eye lens protein gammaB(gammaII)-crystallin at 1.47 A. *Acta Crystallogr D Biol Crystallogr.* **49**(Pt 2): 223-233.

Nakamoto, H. and L. Vigh (2007). The small heat shock proteins and their clients. *Cell Mol Life Sci.* **64**(3): 294-306.

Narberhaus, F. (2002). Alpha-crystallin-type heat shock proteins: socializing minichaperones in the context of a multichaperone network. *Microbiol Mol Biol Rev.* **66**(1): 64-93; table of contents.

Nelson, R., M. R. Sawaya, M. Balbirnie, A. O. Madsen, C. Riekel, R. Grothe and D. Eisenberg (2005). Structure of the cross-beta spine of amyloid-like fibrils. *Nature.* **435**(7043): 773-778.

Nicaise, M., M. Valerio-Lepiniec, N. Izadi-Pruneyre, E. Adjadj, P. Minard and M. Desmadril (2003). Role of the tyrosine corner motif in the stability of neocarzinostatin. *Protein Eng.* **16**(10): 733-738.

Ohno, A., S. Tate, S. S. Seeram, K. Hiraga, M. B. Swindells, K. Oda and M. Kainosho (1998). NMR structure of the Streptomyces metalloproteinase inhibitor, SMPI, isolated from Streptomyces nigrescens TK-23: another example of an ancestral beta gamma-crystallin precursor structure. *J Mol Biol.* **282**(2): 421-433.

Olzscha, H., S. M. Schermann, A. C. Woerner, S. Pinkert, M. H. Hecht, G. G. Tartaglia, M. Vendruscolo, M. Hayer-Hartl, F. U. Hartl and R. M. Vabulas (2011). Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. *Cell.* **144**(1): 67-78.

Otzen, D. E., L. S. Itzhaki, N. F. elMasry, S. E. Jackson and A. R. Fersht (1994). Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding. *Proc Natl Acad Sci U S A.* **91**(22): 10422-10425.

Oyster, C. W. (1999). *The human eye : structure and function.* Sunderland, Mass., Sinauer Associates.

Pande, A., J. Pande, N. Asherie, A. Lomakin, O. Ogun, J. A. King, N. H. Lubsen, D. Walton and G. B. Benedek (2000). Molecular basis of a progressive juvenile-onset hereditary cataract. *Proc Natl Acad Sci U S A.* **97**(5): 1993-1998.

Papanikolopoulou, K., I. Mills-Henry, S. L. Thol, Y. Wang, A. A. Gross, D. A. Kirschner, S. M. Decatur and J. King (2008). Formation of amyloid fibrils in vitro by human gammaD-crystallin and its isolated domains. *Mol Vis.* **14**: 81-89.

Perl, D., U. Mueller, U. Heinemann and F. X. Schmid (2000). Two exposed amino acid residues confer thermostability on a cold shock protein. *Nat Struct Biol.* **7**(5): 380-383.

Peschek, J., N. Braun, T. M. Franzmann, Y. Georgalis, M. Haslbeck, S. Weinkauf and J. Buchner (2009). The eye lens chaperone alpha-crystallin forms defined globular assemblies. *Proc Natl Acad Sci U S A.* **106**(32): 13272-13277.

Peterson, F. C., E. M. Baden, B. A. Owen, B. F. Volkman and M. Ramirez-Alvarado (2010). A single mutation promotes amyloidogenicity through a highly promiscuous dimer interface. *Structure.* **18**(5): 563-570.

Plotnikova, O. V., F. A. Kondrashov, P. K. Vlasov, A. P. Grigorenko, E. K. Ginter and E. I. Rogaev (2007). Conversion and compensatory evolution of the gamma-crystallin genes and identification of a cataractogenic mutation that reverses the sequence of the human CRYGD gene to an ancestral state. *Am J Hum Genet.* **81**(1): 32-43.

Podeszwa, R., R. Bukowski and K. Szalewicz (2006). Potential energy surface for the benzene dimer and perturbational analysis of pi-pi interactions. *J Phys Chem A.* **110**(34): 10345-10354.

Polier, S., Z. Dragovic, F. U. Hartl and A. Bracher (2008). Structural basis for the cooperation of Hsp70 and Hsp110 chaperones in protein folding. *Cell.* **133**(6): 1068-1079.

Ponce, A., C. Sorensen and L. Takemoto (2006). Role of short-range protein interactions in lens opacifications. *Mol Vis.* **12**: 879-884.

Querol, E., J. A. Perez-Pons and A. Mozo-Villarias (1996). Analysis of protein conformational characteristics related to thermostability. *Protein Eng.* **9**(3): 265-271.

Raman, B. and C. M. Rao (1997). Chaperone-like activity and temperature-induced structural changes of alpha-crystallin. *J Biol Chem.* **272**(38): 23559-23564.

Ramirez-Alvarado, M., J. W. Kelly and C. M. Dobson (2010). *Protein misfolding diseases : current and emerging principles and therapies*. Hoboken, N.J., Wiley.

Rawat, U. and M. Rao (1998). Interactions of chaperone alpha-crystallin with the molten globule state of xylose reductase. Implications for reconstitution of the active enzyme. *J Biol Chem.* **273**(16): 9415-9423.

Ray, M. E., G. Wistow, Y. A. Su, P. S. Meltzer and J. M. Trent (1997). AIM1, a novel non-lens member of the betagamma-crystallin superfamily, is associated with the control of tumorigenicity in human malignant melanoma. *Proc Natl Acad Sci U S A.* **94**(7): 3229-3234.

Rea, A. M., E. R. Simpson, J. K. Meldrum, H. E. Williams and M. S. Searle (2008). Aromatic residues engineered into the beta-turn nucleation site of ubiquitin lead to a complex folding landscape, non-native side-chain interactions, and kinetic traps. *Biochemistry.* **47**(48): 12910-12922.

Resnikoff, S., D. Pascolini, D. Etya'ale, I. Kocur, R. Pararajasegaram, G. P. Pokharel and S. P. Mariotti (2004). Global data on visual impairment in the year 2002. *Bull World Health Organ.* **82**(11): 844-851.

Richmond, C. S., J. D. Glasner, R. Mau, H. Jin and F. R. Blattner (1999). Genome-wide expression profiling in Escherichia coli K-12. *Nucleic Acids Res.* **27**(19): 3821-3835.

Richter, K., S. Moser, F. Hagn, R. Friedrich, O. Hainzl, M. Heller, S. Schlee, H. Kessler, J. Reinstein and J. Buchner (2006). Intrinsic inhibition of the Hsp90 ATPase activity. *J Biol Chem.* **281**(16): 11301-11311.

Richter, K., S. Walter and J. Buchner (2004). The Co-chaperone Sba1 connects the ATPase reaction of Hsp90 to the progression of the chaperone cycle. *J Mol Biol.* **342**(5): 1403-1413.

Robinson, N. E., K. J. Lampi, J. P. Speir, G. Kruppa, M. Easterling and A. B. Robinson (2006). Quantitative measurement of young human eye lens crystallins by direct injection Fourier transform ion cyclotron resonance mass spectrometry. *Mol Vis.* **12**: 704-711.

Rodriguez, H. M., D. M. Vu and L. M. Gregoret (2000). Role of a solvent-exposed aromatic cluster in the folding of Escherichia coli CspA. *Protein Sci.* **9**(10): 1993-2000.

Roshan, M., P. H. Vijaya, G. R. Lavanya, P. K. Shama, S. T. Santhiya, J. Graw, P. M. Gopinath and K. Satyamoorthy (2010). A novel human CRYGD mutation in a juvenile autosomal dominant cataract. *Molecular Vision.* **16**: 887-896.

Rotter, M. A., S. Kwong, R. W. Briehl and F. A. Ferrone (2005). Heterogeneous nucleation in sickle hemoglobin: experimental validation of a structural mechanism. *Biophys J.* **89**(4): 2677-2684.

Rudolph, R., R. Siebendritt, G. Nesslauer, A. K. Sharma and R. Jaenicke (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**(12): 4625-4629.

Sahin, E., J. L. Jordan, M. L. Spatara, A. Naranjo, J. A. Costanzo, W. F. Weiss, A. S. Robinson, E. J. Fernandez and C. J. Roberts (2011). Computational design and biophysical characterization of aggregation-resistant point mutations for gammaD crystallin illustrate a balance of conformational stability and intrinsic aggregation propensity. *Biochemistry.* **50**(5): 628-639.

Samanta, U., D. Pal and P. Chakrabarti (2000). Environment of tryptophan side chains in proteins. *Proteins.* **38**(3): 288-300.

Sandilands, A., A. M. Hutcheson, H. A. Long, A. R. Prescott, G. Vrensen, J. Loster, N. Klopp, R. B. Lutz, J. Graw, S. Masaki, C. M. Dobson, C. E. MacPhee and R. A. Quinlan (2002). Altered aggregation properties of mutant gamma-crystallins cause inherited cataract. *EMBO J.* **21**(22): 6005-6014.

Santhiya, S. T., M. Shyam Manohar, D. Rawlley, P. Vijayalakshmi, P. Namperumalsamy, P. M. Gopinath, J. Loster and J. Graw (2002). Novel mutations in the gamma-crystallin genes cause autosomal dominant congenital cataracts. *J Med Genet.* **39**(5): 352-358.

Santhoshkumar, P. and K. K. Sharma (2001). Phe71 is essential for chaperone-like function in alpha A-crystallin. *J Biol Chem.* **276**(50): 47094-47099.

Schildbach, J. F., A. W. Karzai, B. E. Raumann and R. T. Sauer (1999). Origins of DNA-binding specificity: role of protein contacts with the DNA backbone. *Proc Natl Acad Sci U S A.* **96**(3): 811-817.

Sela, M., F. H. White, Jr. and C. B. Anfinsen (1957). Reductive cleavage of disulfide bridges in ribonuclease. *Science.* **125**(3250): 691-692.

Serrano, L., M. Bycroft and A. R. Fersht (1991). Aromatic-aromatic interactions and protein stability. Investigation by double-mutant cycles. *J Mol Biol.* **218**(2): 465-475.

Sharma, K. K., H. Kaur and K. Kester (1997). Functional elements in molecular chaperone alpha-crystallin: identification of binding sites in alpha B-crystallin. *Biochem Biophys Res Commun.* **239**(1): 217-222.

Sharma, K. K., G. S. Kumar, A. S. Murphy and K. Kester (1998). Identification of 1,1'-bi(4-anilino)naphthalene-5,5'-disulfonic acid binding sequences in alpha-crystallin. *J Biol Chem.* **273**(25): 15474-15478.

Sharma, K. K., R. S. Kumar, G. S. Kumar and P. T. Quinn (2000). Synthesis and characterization of a peptide identified as a functional element in alphaA-crystallin. *J Biol Chem.* **275**(6): 3767-3771.

Sharma, K. K. and P. Santhoshkumar (2009). Lens aging: effects of crystallins. *Biochim Biophys Acta.* **1790**(10): 1095-1108.

Shiels, A., D. Mackay, A. Ionides, V. Berry, A. Moore and S. Bhattacharya (1998). A missense mutation in the human connexin50 gene (GJA8) underlies autosomal dominant "zonular pulverulent" cataract, on chromosome 1q. *Am J Hum Genet.* **62**(3): 526-532.

Shimeld, S. M., A. G. Purkiss, R. P. Dirks, O. A. Bateman, C. Slingsby and N. H. Lubsen (2005). Urochordate betagamma-crystallin and the evolutionary origin of the vertebrate eye lens. *Curr Biol.* **15**(18): 1684-1689.

Shoemaker, J. A., N. Congdon, J. Kempen, J. M. Tielsch and D. S. Friedman (2008). Vision Problems in the U.S. Prevalence of Adult Vision Impairment and Age-Related Eye Disease in America, National Eye Institute, Prevent Blindness America.

Siezen, R. J., J. A. Thomson, E. D. Kaplan and G. B. Benedek (1987). Human lens gamma-crystallins: isolation, identification, and characterization of the expressed gene products. *Proc Natl Acad Sci U S A.* **84**(17): 6088-6092.

Siezen, R. J., E. Wu, E. D. Kaplan, J. A. Thomson and G. B. Benedek (1988). Rat lens gamma-crystallins. Characterization of the six gene products and their spatial and temporal distribution resulting from differential synthesis. *J Mol Biol.* **199**(3): 475-490.

Simkovsky, R. and J. King (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.* **103**(10): 3575-3580.

Sipe, J. D. and A. S. Cohen (2000). Review: history of the amyloid fibril. *J Struct Biol.* **130**(2-3): 88-98.

Slingsby, C. and O. A. Bateman (1990). Quaternary interactions in eye lens beta-crystallins: basic and acidic subunits of beta-crystallins favor heterologous association. *Biochemistry.* **29**(28): 6592-6599.

Smith, C. K. and L. Regan (1995). Guidelines for protein design: the energetics of beta sheet side chain interactions. *Science.* **270**(5238): 980-982.

Smulders, R. H. and W. W. de Jong (1997). The hydrophobic probe 4,4'-bis(1-anilino-8-naphthalene sulfonic acid) is specifically photoincorporated into the N-terminal domain of alpha B-crystallin. *FEBS Lett.* **409**(1): 101-104.

Smykal, P., J. Masin, I. Hrdy, I. Konopasek and V. Zarsky (2000). Chaperone activity of tobacco HSP18, a small heat-shock protein, is inhibited by ATP. *Plant J.* **23**(6): 703-713.

Song, S., A. Landsbury, R. Dahm, Y. Liu, Q. Zhang and R. A. Quinlan (2009). Functions of the intermediate filament cytoskeleton in the eye lens. *J Clin Invest.* **119**(7): 1837-1848.

Spector, S., M. Wang, S. A. Carp, J. Robblee, Z. S. Hendsch, R. Fairman, B. Tidor and D. P. Raleigh (2000). Rational modification of protein stability by the mutation of charged surface residues. *Biochemistry.* **39**(5): 872-879.

Srinivas, P., A. Narahari, J. M. Petrash, M. J. Swamy and G. B. Reddy (2010). Importance of eye lens alpha-crystallin heteropolymer with 3:1 alphaA to alphaB ratio: stability, aggregation, and modifications. *IUBMB Life.* **62**(9): 693-702.

Stamler, R., G. Kappe, W. Boelens and C. Slingsby (2005). Wrapping the alpha-crystallin domain fold in a chaperone assembly. *J Mol Biol.* **353**(1): 68-79.

Stefani, M. (2004). Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world. *Biochim Biophys Acta.* **1739**(1): 5-25.

Stephan, D. A., E. Gillanders, D. Vanderveen, D. Freas-Lutz, G. Wistow, A. D. Baxevanis, C. M. Robbins, A. VanAuken, M. I. Quesenberry, J. Bailey-Wilson, S. H. Juo, J. M. Trent, L. Smith and M. J. Brownstein (1999). Progressive juvenile-onset punctate cataracts caused by mutation of the gammaD-crystallin gene. *Proc Natl Acad Sci U S A.* **96**(3): 1008-1012.

Sun, T. X., B. K. Das and J. J. Liang (1997). Conformational and functional differences between recombinant human lens alphaA- and alphaB-crystallin. *J Biol Chem.* **272**(10): 6220-6225.

Takata, T., J. T. Oxford, T. R. Brandon and K. J. Lampi (2007). Deamidation alters the structure and decreases the stability of human lens beta Alpha 3-crystallin. *Biochemistry.* **46**(30): 8861-8871.
Takata, T., J. T. Oxford, B. Demeler and K. J. Lampi (2008). Deamidation destabilizes and triggers aggregation of a lens protein, beta A3-crystallin. *Protein Science.* **17**(9): 1565-1575.
Takata, T., J. P. Smith, B. Arbogast, L. L. David and K. J. Lampi (2010). Solvent accessibility of betaB2-crystallin and local structural changes due to deamidation at the dimer interface. *Exp Eye Res.* **91**(3): 336-346.
Takata, T., L. G. Woodbury and K. J. Lampi (2009). Deamidation alters interactions of beta-crystallins in hetero-oligomers. *Molecular Vision.* **15**(24): 241-249.
Takemoto, L. and D. Boyle (2000). Increased deamidation of asparagine during human senile cataractogenesis. *Mol Vis.* **6**: 164-168.
Talla, V., C. Narayanan, N. Srinivasan and D. Balasubramanian (2006). Mutation causing self-aggregation in human gammaC-crystallin leading to congenital cataract. *Invest Ophthalmol Vis Sci.* **47**(12): 5212-5217.
Tanaka, T. and G. B. Benedek (1975). Observation of protein diffusivity in intact human and bovine lenses with application to cataract. *Invest Ophthalmol.* **14**(6): 449-456.
Tang, Y. C., H. C. Chang, M. Hayer-Hartl and F. U. Hartl (2007). SnapShot: molecular chaperones, Part II. *Cell.* **128**(2): 412.
Tanksale, A., M. Ghatge and V. Deshpande (2002). Alpha-crystallin binds to the aggregation-prone molten-globule state of alkaline protease: implications for preventing irreversible thermal denaturation. *Protein Sci.* **11**(7): 1720-1728.
Tatko, C. D. and M. L. Waters (2002). Selective aromatic interactions in beta-hairpin peptides. *J Am Chem Soc.* **124**(32): 9372-9373.
Taylor, H. R. (1999). Epidemiology of age-related cataract. *Eye (Lond).* **13 (Pt 3b)**: 445-448.
Taylor, H. R., S. K. West, F. S. Rosenthal, B. Munoz, H. S. Newland, H. Abbey and E. A. Emmett (1988). Effect of ultraviolet radiation on cataract formation. *N Engl J Med.* **319**(22): 1429-1433.
Thomas, A., R. Meurisse and R. Brasseur (2002). Aromatic side-chain interactions in proteins. II. Near- and far-sequence Phe-X pairs. *Proteins.* **48**(4): 635-644.
Thomas, A., R. Meurisse, B. Charloteaux and R. Brasseur (2002). Aromatic side-chain interactions in proteins. I. Main structural features. *Proteins.* **48**(4): 628-634.
Thomas, J. G. and F. Baneyx (1998). Roles of the Escherichia coli small heat shock proteins IbpA and IbpB in thermal stress management: comparison with ClpA, ClpB, and HtpG In vivo. *J Bacteriol.* **180**(19): 5165-5172.
Trinkl, S., R. Glockshuber and R. Jaenicke (1994). Dimerization of beta B2-crystallin: the role of the linker peptide and the N- and C-terminal extensions. *Protein Sci.* **3**(9): 1392-1400.
Truscott, R. J. (2005). Age-related nuclear cataract-oxidation is the key. *Exp Eye Res.* **80**(5): 709-725.
Uversky, V. N. (2011). Intrinsically disordered proteins from A to Z. *Int J Biochem Cell Biol.*
Uversky, V. N. and A. K. Dunker (2010). Understanding protein non-folding. *Biochim Biophys Acta.* **1804**(6): 1231-1264.
van den Berg, B., R. J. Ellis and C. M. Dobson (1999). Effects of macromolecular crowding on protein folding and aggregation. *EMBO J.* **18**(24): 6927-6933.

van Montfort, R. L., E. Basha, K. L. Friedrich, C. Slingsby and E. Vierling (2001). Crystal structure and assembly of a eukaryotic small heat shock protein. *Nat Struct Biol.* **8**(12): 1025-1030.

Van Montfort, R. L., O. A. Bateman, N. H. Lubsen and C. Slingsby (2003). Crystal structure of truncated human betaB1-crystallin. *Protein Sci.* **12**(11): 2606-2612.

Veinger, L., S. Diamant, J. Buchner and P. Goloubinoff (1998). The small heat-shock protein IbpB from Escherichia coli stabilizes stress-denatured proteins for subsequent refolding by a multichaperone network. *J Biol Chem.* **273**(18): 11032-11037.

Viitanen, P. V., A. A. Gatenby and G. H. Lorimer (1992). Purified chaperonin 60 (groEL) interacts with the nonnative states of a multitude of Escherichia coli proteins. *Protein Sci.* **1**(3): 363-369.

Wandinger, S. K., K. Richter and J. Buchner (2008). The Hsp90 chaperone machinery. *J Biol Chem.* **283**(27): 18473-18477.

Wang, K. and A. Spector (2000). alpha-crystallin prevents irreversible protein denaturation and acts cooperatively with other heat-shock proteins to renature the stabilized partially denatured protein in an ATP-dependent manner. *Eur J Biochem.* **267**(15): 4705-4712.

Wang, L., X. Chen, Y. Lu, J. Wu, B. Yang and X. Sun (2011). A novel mutation in gammaD-crystallin associated with autosomal dominant congenital cataract in a Chinese family. *Molecular Vision.* **17**: 804-809.

Wang, L., S. K. Maji, M. R. Sawaya, D. Eisenberg and R. Riek (2008). Bacterial inclusion bodies contain amyloid-like structure. *PLoS Biol.* **6**(8): e195.

Wang, X., C. M. Garcia, Y. B. Shui and D. C. Beebe (2004). Expression and regulation of alpha-, beta-, and gamma-crystallins in mammalian lens epithelial cells. *Invest Ophthalmol Vis Sci.* **45**(10): 3608-3619.

Wang, Y., S. Petty, A. Trojanowski, K. Knee, D. Goulet, I. Mukerji and J. King (2010). Formation of amyloid fibrils in vitro from partially unfolded intermediates of human gammaC-crystallin. *Invest Ophthalmol Vis Sci.* **51**(2): 672-678.

Waters, M. L. (2004). Aromatic interactions in peptides: impact on structure and function. *Biopolymers.* **76**(5): 435-445.

Wathen, B. and Z. Jia (2010). Protein beta-sheet nucleation is driven by local modular formation. *J Biol Chem.* **285**(24): 18376-18384.

Weibezahn, J., C. Schlieker, P. Tessarz, A. Mogk and B. Bukau (2005). Novel insights into the mechanism of chaperone-assisted protein disaggregation. *Biol Chem.* **386**(8): 739-744.

Weinreb, O., A. F. van Rijk, A. Dovrat and H. Bloemendal (2000). In vitro filament-like formation upon interaction between lens alpha-crystallin and betaL-crystallin promoted by stress. *Invest Ophthalmol Vis Sci.* **41**(12): 3893-3897.

Welsh, M. J. and A. E. Smith (1993). Molecular mechanisms of CFTR chloride channel dysfunction in cystic fibrosis. *Cell.* **73**(7): 1251-1254.

Wenk, M., R. Herbst, D. Hoeger, M. Kretschmar, N. H. Lubsen and R. Jaenicke (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**(2-3): 95-108.

Wetlaufer, D. B. (1963). Ultraviolet Spectra of Proteins and Amino Acids. *Advances in Protein Chemistry.* **17**: 303-390.

Wetzel, R. (1996). For protein misassembly, it's the "I" decade. *Cell.* **86**(5): 699-702.

White, T. W. and R. Bruzzone (2000). Intercellular communication in the eye: clarifying the need for connexin diversity. *Brain Res Brain Res Rev.* **32**(1): 130-137.

Wieligmann, K., E. M. Mayr and R. Jaenicke (1999). Folding and self-assembly of the domains of betaB2-crystallin from rat eye lens. *J Mol Biol.* **286**(4): 989-994.
Wille, H., M. D. Michelitsch, V. Guenebaut, S. Supattapone, A. Serban, F. E. Cohen, D. A. Agard and S. B. Prusiner (2002). Structural studies of the scrapie prion protein by electron crystallography. *Proc Natl Acad Sci U S A.* **99**(6): 3563-3568.
Winkler, J., A. Seybert, L. Konig, S. Pruggnaller, U. Haselmann, V. Sourjik, M. Weiss, A. S. Frangakis, A. Mogk and B. Bukau (2010). Quantitative and spatio-temporal features of protein aggregation in Escherichia coli and consequences on protein quality control and cellular ageing. *EMBO J.* **29**(5): 910-923.
Wistow, G. (1990). Evolution of a protein superfamily: relationships between vertebrate lens crystallins and microorganism dormancy proteins. *J Mol Evol.* **30**(2): 140-145.
Wistow, G., C. Jaworski and P. V. Rao (1995). A non-lens member of the beta gamma-crystallin superfamily in a vertebrate, the amphibian Cynops. *Exp Eye Res.* **61**(5): 637-639.
Wistow, G., L. Sardarian, W. Gan and M. K. Wyatt (2000). The human gene for gammaS-crystallin: alternative transcripts and expressed sequences from the first intron. *Mol Vis.* **6**: 79-84.
Wistow, G., L. Summers and T. Blundell (1985). Myxococcus xanthus spore coat protein S may have a similar structure to vertebrate lens beta gamma-crystallins. *Nature.* **315**(6022): 771-773.
Wistow, G. J. and J. Piatigorsky (1988). Lens crystallins: the evolution and expression of proteins for a highly specialized tissue. *Annu Rev Biochem.* **57**: 479-504.
Wormstone, I. M. (2002). Posterior capsule opacification: a cell biological perspective. *Exp Eye Res.* **74**(3): 337-347.
Wu, L., D. McElheny, T. Takekiyo and T. A. Keiderling (2010). Geometry and efficacy of cross-strand Trp/Trp, Trp/Tyr, and Tyr/Tyr aromatic interaction in a beta-hairpin peptide. *Biochemistry.* **49**(22): 4705-4714.
Wu, Z., F. Delaglio, K. Wyatt, G. Wistow and A. Bax (2005). Solution structure of (gamma)S-crystallin by molecular fragment replacement NMR. *Protein Sci.* **14**(12): 3101-3114.
Xu, X., S. Wang, Y. X. Hu and D. B. McKay (2007). The periplasmic bacterial molecular chaperone SurA adapts its structure to bind peptides in different conformations to assert a sequence preference for aromatic residues. *J Mol Biol.* **373**(2): 367-381.
Xu, Z., A. L. Horwich and P. B. Sigler (1997). The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature.* **388**(6644): 741-750.
Yamasaki, M., W. Li, D. J. Johnson and J. A. Huntington (2008). Crystal structure of a stable dimer reveals the molecular basis of serpin polymerization. *Nature.* **455**(7217): 1255-1258.
Yao, J., H. J. Dyson and P. E. Wright (1994). Three-dimensional structure of a type VI turn in a linear peptide in water solution. Evidence for stacking of aromatic rings as a major stabilizing factor. *J Mol Biol.* **243**(4): 754-766.
Zantema, A., M. Verlaan-De Vries, D. Maasdam, S. Bol and A. van der Eb (1992). Heat shock protein 27 and alpha B-crystallin can form a complex, which dissociates by heat shock. *J Biol Chem.* **267**(18): 12936-12941.
Zhang, J. X. and D. P. Goldenberg (1997). Mutational analysis of the BPTI folding pathway: I. Effects of aromatic-->leucine substitutions on the distribution of folding intermediates. *Protein Sci.* **6**(7): 1549-1562.

Zhang, J. X. and D. P. Goldenberg (1997). Mutational analysis of the BPTI folding pathway: II. Effects of aromatic-->leucine substitutions on folding kinetics and thermodynamics. *Protein Sci.* **6**(7): 1563-1576.

Zhang, W., H. C. Cai, F. F. Li, Y. B. Xi, X. Ma and Y. B. Yan (2011). The congenital cataract-linked G61C mutation destabilizes gammaD-crystallin and promotes non-native aggregation. *PLoS One.* **6**(5): e20564.

Zhu, X., X. Zhao, W. F. Burkholder, A. Gragerov, C. M. Ogata, M. E. Gottesman and W. A. Hendrickson (1996). Structural analysis of substrate binding by the molecular chaperone DnaK. *Science.* **272**(5268): 1606-1614.

Chapter 6

Appendices

A. Equilibrium Unfolding/Refolding on Additional Aromatic Mutant HγD-Crys

Additional aromatic mutant HγD-Crys were constructed and their thermodynamic stabilities were measured by equilibrium unfolding/refolding experiments as described in Chapter 2. Similar to the data presented in Chapter 2, the unfolding and refolding traces of all the additional mutant HγD-Crys in this section overlapped very well. Refolding, however, induced aggregation and interfered with the data analysis. Therefore, transition midpoints, apparent m values and ΔΔG's were calculated only from the unfolding traces and summarized in Table 6-1.

1. Tyr-to-Phe Mutant HγD-Crys

To distinguish the contributions from the aromatic rings vs. the hydroxyl groups of the Greek key tyrosine residues, single and double Tyr-to-Phe substitutions were introduced to a representative pair Y45/Y50. The results showed that these substitutions caused very slight destabilization of the N-td, and no effect on the C-td, as measured by transition midpoints and ΔΔG's (Figure 6-1and Table 6-1). A search for hydrogen bonding partners for the hydroxyl groups of Y45 and Y50 in the crystal structure of HγD-Crys identified the backbone nitrogens of H22 and P23 for Y50, but the distances between the hydroxyl oxygen and the nitrogens (3.5 and 3.2 Å) were somewhat longer than a typical hydrogen bond, thus making the possible hydrogen bonds unlikely or very weak. No candidate hydrogen bond partner was found for Y45 within 3.5 Å from the hydroxyl oxygen. In summary, the hydroxyl groups of Y45 and Y50 made minimal, if any, contribution to the thermodynamic stability of HγD-Crys, probably through weak electrostatic interactions from the partial charges with neighboring groups. Therefore, the Greek key aromatic residues interact with the neighboring residues largely through the aromatic rings.

2. Aromatic-to-Ala Double Mutant HγD-Crys

N-td aromatic-to-Ala double mutant HγD-Crys were constructed for the purpose of estimating the aromatic-aromatic interaction energy by double mutant cycle analysis. The results were discussed in Chapter 2 without showing the primary data. The equilibrium traces and

calculated parameters are shown in Figure 6-2 and Table 6-1. The aromatic-aromatic interaction energies for pairs Y6/F11, Y16/Y28, Y45/Y50 were −1.6, −1.5, and −2.0 kcal/mol, respectively.

3. Mutant HγD-Crys on the Unpaired Aromatic Residues

HγD-Crys contains 14 Tyr and 6 Phe residues distributed throughout the protein (Figure 2-1): 12 in pairs and 8 unpaired. The 6 pairs, especially the 4 Greek key pairs has been described in details in Chapter 2 and 3. Except F104 and F172, most of these unpaired aromatic residues are conserved among βγ-crystallins, some further at the crystallin domain level. Similar to the paired residues, 7 of these 8 unpaired aromatic residues are partially-buried: with at least some solved accessible area but not fully exposed. F172 is located at the short C-terminal extension and is fully exposed to solvent. Y55, Y62 at the N-td and Y143, Y150 at the C-td are conserved at the crystallin domain level. Y62 and Y150 are the "tyrosine corners" (Hemmingsen *et al.* 1994). In the N-td, Y55 and Y62 are not in direct contact, but both form contacts with the conserved W68 with a tilted T-shaped orientation. This "aromatic triad" is conserved at the crystallin domain level, and in the C-td, formed by Y143, Y150 and W156. Residues F56, F104, Y153 and F172 do not have aromatic counterparts in the other crystallin domain. To measure the contributions of these unpaired residues to the thermodynamic stability of HγD-Crys, aromatic-to-Ala mutant proteins were constructed and subject to the equilibrium unfolding/refolding experiments.

As shown in Figure 6-3 and Table 6-1, the destabilization effects of all N-td mutant proteins were limited to the N-td transition, consistent with the sequential unfolding/refolding model of HγD-Crys, namely, the N-td unfolds first and refolds last. All substitutions except Y172A caused destabilization to different degrees. As measured by the transition midpoints. For the N-td mutant proteins, the substitutions on the triad tyrosines and Phe 56 had smaller effects compared to the Greek key aromatic pairs (compare with Table 2-2). Phe 56 is involved in stabilizing the domain interface (Flaugh *et al.* 2005). For the C-td mutant proteins, substitutions on the triad tyrosines had effects comparable to the Greek key pairs. F104A and Y153A had intermediate effects between the Greek key pairs and the non-Greek key pairs. In each crystallin domain, the tyrosine corners (Y62 and Y150) always had larger effects than the other triad tyrosines (Y55 and Y143). The equilibrium profile of F172A was nearly identical to that of the

WT proteins, consistent with the fact that F172 does not interact with any other residue in the crystal structure. F172 may be involved in intermolecular interaction. In conclusion, the Greek key aromatic residues had the largest stabilizing effects among all the Tyr and Phe residues in HγD-Crys.

Table 6-1 Equilibrium Unfolding Parameters for WT and Additional Mutant HγD-Crys.

Protein	Equilibrium Transition 1			Equilibrium Transition 2		
	C_m [a]	Apparent *m* value [b]	Apparent $\Delta\Delta G_0$ [c]	C_m [a]	Apparent *m* value [b]	Apparen t $\Delta\Delta G_0$ [c]
WT						
three-state	2.21±0.05	4.6±0.4	0.0±0.8	2.99±0.03	2.9±0.3	0.0±1.3
two-state	2.81±0.02	0.6±0.0	0.0±0.0	-	-	-
Y→F						
Y45F	2.06±0.01	4.6±0.5	-0.2±1.2	2.99±0.01	2.7±0.2	-0.9±0.7
Y50F	2.12±0.03	4.6±0.9	-0.1±1.7	2.97±0.03	2.7±0.5	-1.2±1.6
Y45F/Y50F	1.98±0.11	4.9±1.7	-0.3±2.8	2.99±0.01	2.8±1.3	-0.8±3.9
N-td aromatic→A double						
Y6A/F11A	0.73±0.08	4.0±0.9	-6.9±0.9	3.01±0.09	3.2±1.1	0.4±3.0
Y16A/Y28A	1.60±0.03	4.5±0.2	-2.6±0.3	3.01±0.04	2.8±0.0	-0.8±0.1
Y45A/Y50A	0.36±0.20	4.3±1.4	-8.1±1.2	2.98±0.08	2.6±0.9	-1.5±2.5
conserved "triad" Y→A						
Y55A	1.77±0.08	5.1±1.7	-0.8±2.6	3.09±0.12	2.4±1.4	-1.9±4.1
Y143A	2.17±0.02	1.7±0.1	2.0±0.2	-	-	-
tyrosine corner Y→A [d]						
Y62A	1.50±0.10	n/a	n/a	3.00±0.10	n/a	n/a
Y150A	2.00±0.10	n/a	n/a	-	-	-
non-conserved Y→A						
Y56A [e]	1.60±0.06	3.9±0.4	-1.6±0.7	2.90±0.03	3.2±0.2	0.4±0.7
F104A	2.31±0.03	1.8±0.4	2.3±0.8	-	-	-
Y153A	2.63±0.01	0.9±0.0	0.5±0.1	-	-	-
Y172A	2.80±0.03	0.6±0.0	-0.2±0.0	-	-	-

[a] Equilibrium unfolding transition midpoints in units of M GuHCl.

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

[c] Difference of free energy change extrapolated to 0 M GuHCl, as compared to the WT protein, in units of kcal·mol^{-1}. A negative number means mutant is less stable than the WT.

[d] Data from J. Chen, personal communication.

[e] Data from Flaugh *et al.*(Flaugh *et al.* 2005).

"-" not applicable.

"n/a" not available.

Errors are standard deviations of three trials.

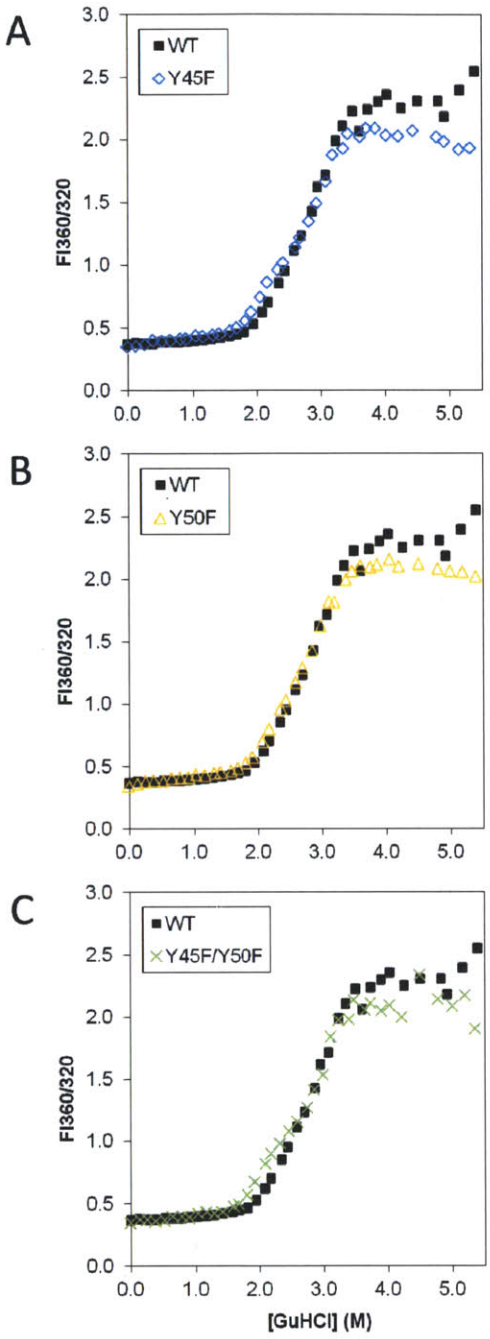


Figure 6-1 Equilibrium Unfolding for WT and Tyr-to-Phe Mutant HγD-Crys.

The experiments were performed as described in Figure 2-4 and Materials and Methods section in Chapter 2. (A) Y45F; (B) Y50F; (C) Y45F/Y50F.

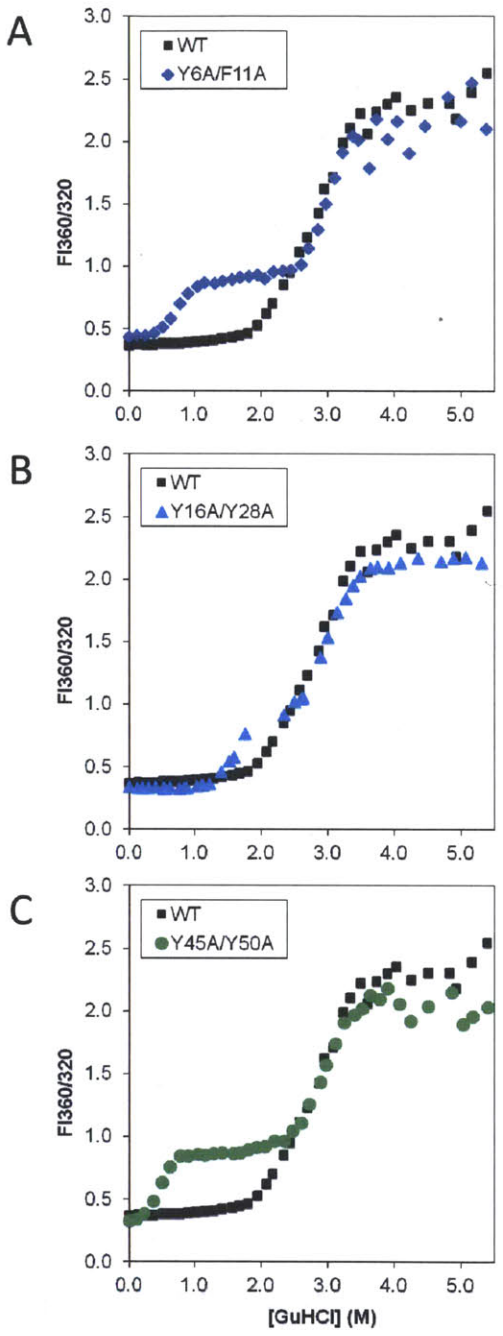


Figure 6-2 Equilibrium Unfolding for WT and N-td Aromatic-to-Ala Double Mutant HγD-Crys. The experiments were performed as described in Figure 2-4 and Materials and Methods section in Chapter 2. (A) Y6A/F11A; (B) Y16A/Y28A; (C) Y45A/Y50A.

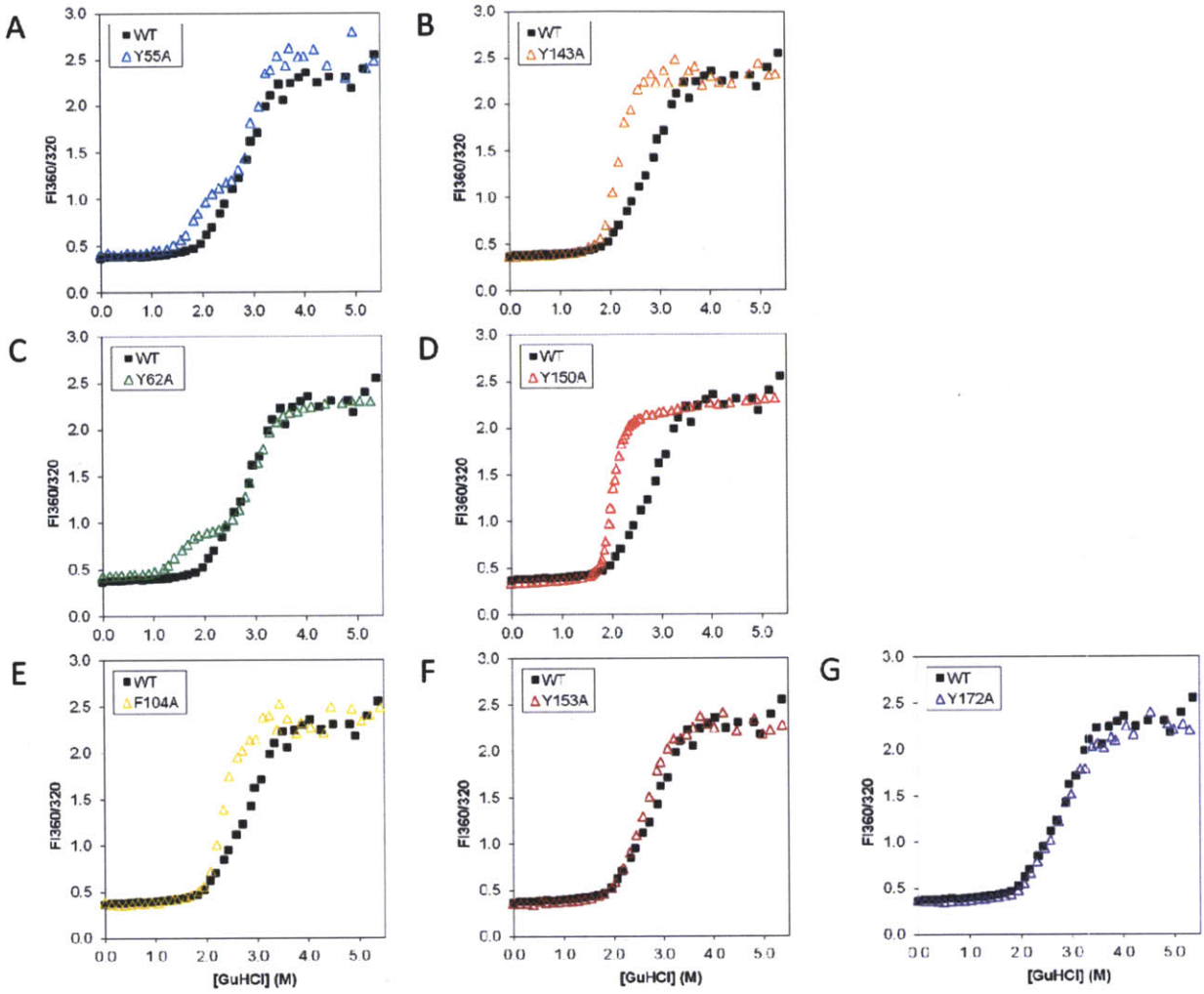


Figure 6-3 Equilibrium Unfolding for WT and Unpaired Aromatics Mutant HγD-Crys. The experiments were performed as described in Figure 2-4 and Materials and Methods section in Chapter 2. (A-B) Substitutions on homologous residues on each crystallin domain: (A) Y55A; (B) Y143A. (C-D) Substitutions on the "tyrosine corners". Experiments on these two mutant proteins were performed by J. Chen. Data were obtained from unpublished presentation by J. Chen. (C) Y62A; (D) Y150A. (E-G) Substitutions on the non-conserved residues: (E) F104A; (F) Y153A; (G) Y172A.

B. pH 3 Amyloid Aggregation

Amyloid-like fibril formation can be induced by low pH for human γC- and γD-crystallins (Papanikolopoulou *et al.* 2008; Wang *et al.* 2010). Both the isolated N-td and C-td of HγD-Crys retained the ability to form amyloid fibrils to lesser extents, indicating that multiple regions were involved in such process. Except this information, the HγD-Crys sequences responsible for amyloid formation are essentially unknown, l*et al*one the molecular structure of the fibril. To test whether the aromatic residues on HγD-Crys are involved in such interaction, pH3-induced amyloid aggregation experiments were performed on the WT and selected aromatic mutant HγD-Crys (Y6A, F11A, Y45A, Y50A).

1. Materials and Methods

This protocol was modified from published work on HγC-Crys (Wang *et al.* 2010). ThT stock was prepared in 1 mM in H_2O, 0.2 μm filtered, wrapped in aluminum foil and stored in dark. ThT working solution was prepared by diluting ThT stock 40x freshly to 25 μM in ThT dilution buffer (50 mM $NaPO_4$ buffer, pH 7, 0.2 μm filtered). WT or mutant HγD-Crys stocks were concentrated to >10 mg/mL, then diluted to 10 mg/mL with 10 mM ammonium acetate, and 0.2 μm spin filtered just before the experiment. To initiate the aggregation reaction, dilute the proteins at 10 mg/mL 10x with pH 3 aggregation buffer (100 mM sodium acetate, 100 mM NaCl, 1 mM EDTA, pH 3, 0.2 μm filtered), to give final protein concentration of 1 mg/mL. ThT dilution buffer was used to replace the pH 3 aggregation buffer as control. Pre-warm the buffer and maintain the reaction at 37 °C. The reaction was monitored by turbidity assay at Abs_{350} as well as by ThT fluorescence. At pre-determinant time-points, mix 8 μL aggregation reaction samples in 392 μL ThT working solution. Invert tubes several times to ensure thorough mix. Final protein concentration: 20 μg/mL, ThT concentration: 24.5 μM. Fluorescence measurement was done with a Hitachi F-4500 fluorometer at RT, with excitation wavelength 444 nm, PMT voltage 950 V. Emission wavelength 470-570 nm was scanned to detect abnormal spectra, 485 nm was used for analysis.

2. *Results and Discussion*

Unlike HγC-Crys, WT HγD-Crys did not show any significant increase in turbidity measurement after incubated for 3 hours, at both pH 3 and pH 7 (Figure 6-4). These results were consistent with parallel experiments conducted by K. Moreau (personal communication). Turbidity measurements on the tested mutant HγD-Crys (Y6A, F11A, Y45A, Y50A) were essential the same as the WT protein. These results suggest that the amyloid fibrils formed by HγD-Crys, if any, were of smaller size or altered structure compared to those of HγC-Crys, making them undetectable by the turbidity assay.

For ThT fluorescence, increase on the order of 100 A.U. was observed in 3 hours for WT HγD-Crys incubated at pH 3 (Figure 6-5). This value was substantially lower than around 500 A.U. previously observed by K. Moreau (personal communication). For the mutant HγD-Crys (Y6A, F11A, Y45A, Y50A), increases from 1,200 to 2,200 A.U. were observed in the same time period, much higher than either WT value. No increase was observed in control experiments at pH 7 (Figure 6-5). The kinetics of amyloid fibril formation (or whether fibrils form at all), is sensitive to small differences in experimental procedures or trace amounts of contaminants in the reaction solution. The contaminants may seed the polymerization or promote seed formation, thus significantly altering the aggregation kinetics. The discrepancy seen in the WT HγD-Crys kinetics was probably due to subtle differences in the practice by the two researchers, although identical protocols were used. Despite this discrepancy, the ThT fluorescence increases for the mutant proteins were significantly higher than the WT value from either researcher, and were likely to reflect real differences in the fibril forming behaviors between the WT and the mutant proteins. Because the mutant proteins promoted fibril formation, residues Y6, F11, Y45 and Y50 probably do not participate directly in the intermolecular interaction in amyloid fibril formation. Instead, the destabilization effects from these substitutions may make accessible the regions involved in such intermolecular interaction.

Further efforts should be made to carefully standardize the experiment protocol. Since both the N-td and C-td of HγD-Crys could form fibril, C-td Greek key aromatic residues can also be tested. The amyloid-responsible regions or sequences may be narrowed down by using successively shorter peptides from HγD-Crys. Computational algorithms may also aid the identification and suggest possible molecular structures (Goldschmidt *et al.* 2010).

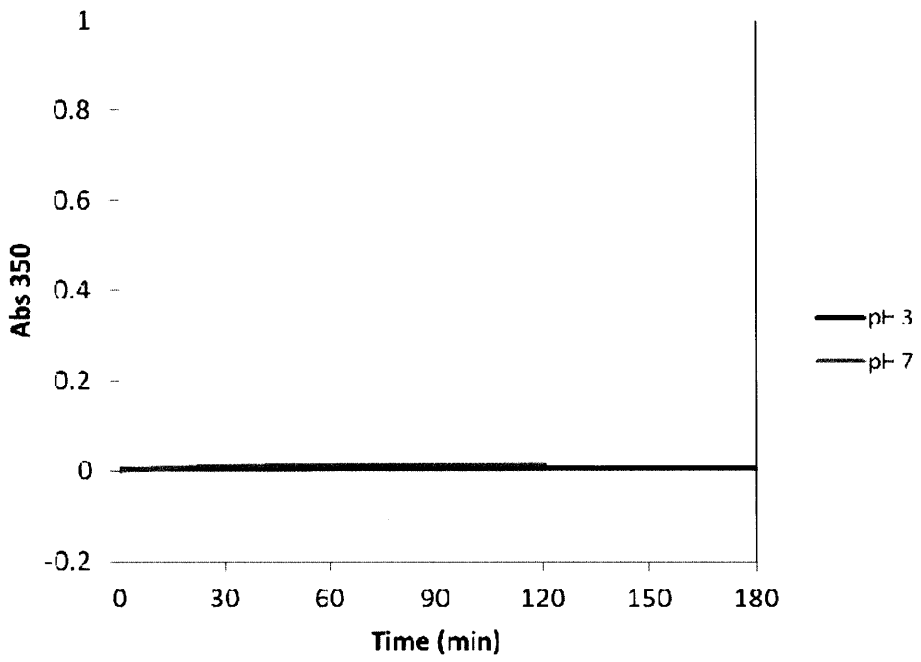


Figure 6-4 Turbidity Measurements on WT HγD-Crys incubated at pH 3 and pH 7.
The experiments was conducted as described in the Materials and Methods section. WT HγD-Crys was incubated at 37 °C. The values were blanked with buffers at each pH with no proteins.

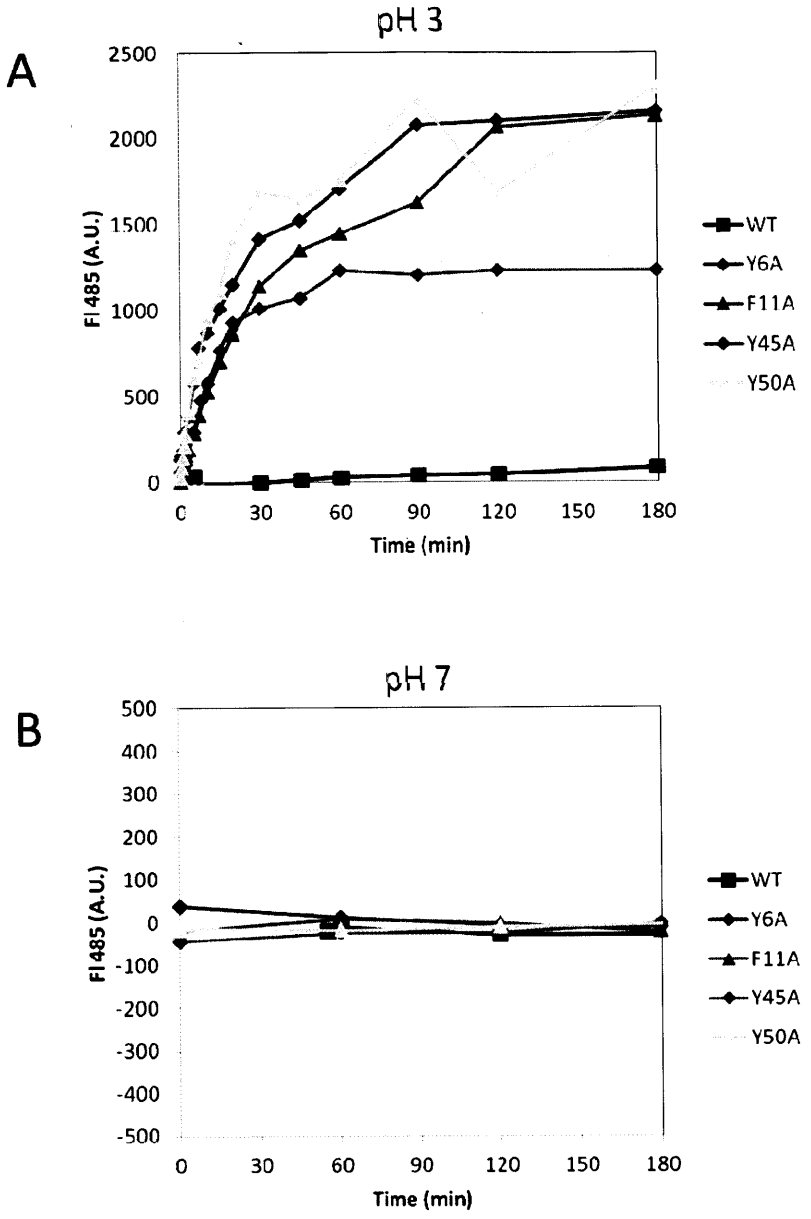


Figure 6-5 ThT Fluorescence of WT and Mutant HγD-Crys incubated at pH 3 and pH 7. The experiments was done as described in the Materials and Methods section. Note that after the proteins were incubated at each pH at 37 °C after certain time, they were mixed with ThT at pH 7. The time axes indicate the mixing times, but not ThT incubation times. The values were blanked with ThT working solutions mixed with pH 3 or pH 7 aggregation buffer with no protein. (A) pH 3. (B) pH 7.