UC Irvine
UC Irvine Electronic Theses and Dissertations

Title
Understanding the Molecular Basis of Transparency and Refraction in the Eye Lens

Permalink
https://escholarship.org/uc/item/1vh7r6wk

Author
Khago, Domarin

Publication Date
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Understanding the Molecular Basis of Transparency and Refraction in the Eye Lens

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Chemistry

by

Domarin Khago

Dissertation Committee:
Professor Rachel W. Martin, Chair
Professor Douglas J. Tobias
Professor Jennifer A. Prescher

2016

DEDICATION

To my parents, Emanuel and Linda,
for their constant love and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I must thank my advisor, Dr. Rachel W. Martin. I came to Rachel as a naive first year student who had no experience in biology but she did not let that stop me. She encouraged me to join her lab and gave me the freedom to explore the unknown. She also provided me with every opportunity to succeed by handing me challenging problems in the lab to solve, training new students, attending conferences to share my work with other scientists in the field, and even sending me off to Grenoble, France not once, but twice to work with our collaborators on some exciting projects. I can never thank Rachel enough for teaching me and continuing to show me how to be a great scientist.

I also would like to thank several members of the Martin lab, former and current, to making my time here enjoyable: Dr. Carolyn N. Kingsley for initially training me (sorry if I messed up!) and joining me often for the little adventures to grab caffeinated-sugary drinks and exotic lunches just to escape from lab a bit; Dr. Amanda J. Holton for giving me the easy teaching assignments and occasionally grabbing coffee just so she could listen to me vent; Pia Dizon, my first undergrad that I got to train (a month after I was trained), that was patient and learned the material along with me and truly made me enjoy teaching her how to be a great scientist; Dr. Suvrajit Sengupta for teaching me almost EVERYTHING about protein NMR and always making me think about the little details you typically wouldn't think about; Dr. Kelsey A. Collier and John E. Kelly for going on coffee adventures and especially for keeping our magnets functioning; and Kyle W. Roskamp for the football talk as well as our silly inside jokes that made each day fun and for jumping on my projects to finish them up once I leave (I know they are in great hands!). And to the rest of the Martin lab members not mentioned here, I thank you for listening to my crazy suggestions and sitting through some pretty interesting practice talks the last several years.

With much gratitude and respect, I would like to thank my dissertation committee members, Drs. Douglas J. Tobias and Jennifer A. Prescher for all the advice and feedback on my research throughout the years. I would also like to thank Drs. Melanie Cocco, Andrej Lupták, and James S. Nowick for the conversations we have shared, which have always been encouraging and helpful.

I would like to thank all my collaborators; Dr. Carter T. Butts who was so instrumental to getting my data cleaned up with his brilliant statistical program; Dr. J. Alfredo Freites for his assistance in moving the light scattering project forward; Dr. Sudipta Majumdar for the insightful discussions on light scattering and always providing helpful suggestions; Dr. Hartmut Oschkinat who so kindly found me at each conference to discuss my work; Dr. Anne Diehl who so graciously provided me protein every single time I had asked for it; Drs. Martin Weik and Yann Fichou for inviting me to Grenoble, help with SAXS experiments, and for showing me around such a beautiful city; and lastly, Eric K. Wong and Vera Prytkova for all their hard work and efforts to getting their simulations for our crystallin project. I

would also like to thank my undergraduate advisor, Dr. Amy C. Moskun, for being in her office that day I decided to become a chemist. The talks we shared about chemistry and life in general made me feel like I could tackle anything that came my way. The support and encouragement she showed me as an undergraduate will always stay with me.

I would like to thank the facility managers, Drs. Dmitry A. Fishman and Philip R. Dennison for their constant help and advice as well as keeping the instruments us students use daily functioning beautifully. I would also like to thank Dr. Stephen H. White for allowing me to use his light scattering instruments while taking over some of his lab space.

Lastly, I would like to thank all my friends and family who have been their for me every step of the way. My friends from my hometown, my friends from Whittier College, and my friends from UCI have been there for me, through the good and the bad, and have always shown me a good time when we're together. I especially would like to thank my family; my sister, Dorothy Khago, for being the level-headed one the last couple years and always telling me that there was nothing I couldn't do; my "brother" Brynner Agassi who always calls to check in on me, take me out to dinner, discuss real life issues, and quite literally, step in to the big brother role that was missing from my life; and my parents, Emanuel and Linda Khago, who have never stopped encouraging me to pursue my goal. Thank you for all the love and support you have shown me. I hope I have made you both proud.

CURRICULUM VITAE

Domarin Khago

EDUCATION

Doctor of Philosophy in Chemistry   **2016**
University of California, Irvine   *Irvine, California*

Bachelor of Arts in Chemistry   **2011**
Bachelor of Arts in Economics   **2011**
Whittier College   *Whittier, California*

RESEARCH EXPERIENCE

Ph.D. Candidate   **2013–2016**
University of California, Irvine   *Irvine, California*

Graduate Research Assistant   **2011–2013**
University of California, Irvine   *Irvine, California*

Undergraduate Research Assistant   **2010–2011**
Whittier College   *Whittier, California*

TEACHING EXPERIENCE

Graduate Student Instructor, General Chemistry Lecture III   **Spring 2014**
University of California, Irvine   *Irvine, California*

Graduate Student Instructor, General Chemistry Laboratory   **Summer 2013**
University of California, Irvine   *Irvine, California*

Graduate Student Instructor, General Chemistry Lecture III   **Summer 2013**
University of California, Irvine   *Irvine, California*

Graduate Student Instructor, Organic Chemistry Laboratory III   **Spring 2013**
University of California, Irvine   *Irvine, California*

Graduate Student Instructor, Honors General Chemistry   **Fall 2012**
University of California, Irvine   *Irvine, California*

Graduate Student Instructor, Organic Chemistry Laboratory III　　Spring 2012
University of California, Irvine　　　　　　　　　　　　　　　　　　*Irvine, California*

Graduate Student Instructor, Organic Chemistry Laboratory II　　Winter 2012
University of California, Irvine　　　　　　　　　　　　　　　　　　*Irvine, California*

Graduate Student Instructor, Organic Chemistry Laboratory I   **Fall 2011**
University of California, Irvine   *Irvine, California*

PUBLICATIONS

- Increased hydrophobic surface exposure in the cataract-related G18V variant of human γS-crystallin. **Khago, D.**; Wong, E. K.; Kingsley, C. N.; Frietas, J. A.; Tobias, D. J.; Martin, R. W. Biochimica et Biophysica Acta (2016) 1860, 325–332.

- Stability of protein-specific hydration shell on crowding. Huang, K.-Y.; Kingsley, C. N.; Sheil, R.; Cheng, C.-Y.; Bierma, J. C.; Roskamp, K. W.; **Khago, D.**; Martin, R. W., Han, S. Journal of the American Chemical Society (2016) 138: 5392–5402.

- Multi-conformation Monte Carlo: a method for introducing flexibility in efficient simulations of many-protein systems. Heyden, M.; Prytkova, V.; *Khago, D.*; Frietes, J. A.; Martin, R. W.; Tobias, D. J. Journal of Physical Chemistry B (2016) 120: 8115–8126.

- Protein refractive index increment is determined by conformation as well as composition. **Khago, D.**; Bierma, J. C.; Roskamp, K. W.; Kozlyuk, N.; Martin, R. W. *Manuscript in preparation*

- Probing peptide-protein interactions in human γS-crystallin. **Khago, D.**; Kingsley, C. N.; Wong, E. K.; Frietas, J. A.; Tobias, D. J.; Martin, R. W. *Manuscript in preparation*

- Bayesian analysis of static light scattering data for globular proteins. Butts, C. T.; **Khago, D.**; Martin, R. W. *Manuscript in preparation*

- Stability and refractive index increment of the box jellyfish eye lens protein J2-crystallin. **Khago, D.**; Kingsley, C. N.; Collier, K. A.; Valentic, T. R.; Dizon, P. A.; Martin, R. W. *Manuscript in preparation*

CONFERENCE PRESENTATIONS

Biophysical characterization and solution NMR structure of J2-crystallin: A novel eye lens protein   **March 2016**
251st American Chemical Society National Meeting, San Diego, California

Examining the Surface Hydrophobicity Present Among Human γS-Crystallin and the Cataract Variant G18V   **April 2015**
56th Experimental Nuclear Magnetic Resonance Conference, Asilomar, California

Biophysical Characterization and Structural Study of J2-Crystallin: A Novel Eye Lens Crystallin   **April 2013**
54th Experimental Nuclear Magnetic Resonance Conference, Asilomar, California

Photophysics of 5-methyl-2-[(1-pyrenylmethylene)amino]-phenol   **March 2011**

241st American Chemical Society National Meeting, Anaheim, California

ABSTRACT OF THE DISSERTATION

Understanding the Molecular Basis of Transparency and Refraction in the Eye Lens

By

Domarin Khago

Doctor of Philosophy in Chemistry

University of California, Irvine, 2016

Professor Rachel W. Martin, Chair

Crystallins are water-soluble proteins that are necessary for focusing light on the retina. In mammalian lenses, there are two classes of crystallins; α-crystallins (molecular chaperone proteins), and $\beta\gamma$-crystallins (structural proteins). My research has focused on γS-crystallin, which is the main structural component of the human eye lens cortex and contributes to the lens fiber organization. Present up to 450 mg/mL in the human lens, γS-crystallin depends on its long-term ability to remain stable and retain a high refractive index. The G18V variant of γS-crystallin (γS-G18V), associated with hereditary childhood-onset cataract, shows decreased stability and increased aggregate formation. My work explores the intermolecular interactions contributing to the increased aggregation propensity of γS-G18V relative to wild-type (γS-WT). By titration of a hydrophobic chemical probe (ANS) and by titration of tripeptides (a library composed of the crystallin sequence), residue-specific binding was observed via NMR chemical shift perturbations (CSP). Protein-protein interactions have also been measured for the direct determination of the second virial coefficient, which describes the intermolecular interactions as being either repulsive or attractive. The second virial coefficient for lysozyme under physiological conditions will give better insight into how the individual molecules are interacting with one another in solution. The control experiments with lysozyme will be applied to both γS-WT and a deamidated variant, N15D, to investigate how prone the crystallins are to aggregation. Additionally, in order to investigate

the functionality of crystallins from aquatic species, the biophysical characterization of the J2-crystallin from box jellyfish (Tripedalia cystophora) will be presented. J2-crystallin is an excellent example of convergent evolution developing a protein to perform the refractive function needed in the eye lenses and thus a deeper understanding of how these crystallins are related to one another. Lastly, the refractive index increment (dn/dC) of various crystallins from aquatic and terrestrial species will be reported to fully understand the functionality of the eye lens.

Chapter 1

The molecular basis of transparency and refraction in the eye lens

1.1 Cataract affects transparency in the eye lens

One of the leading causes of blindness globally is cataract disease [13, 14]. Cataract is the opacification (cloudiness) of the eye lens [13, 15, 16] which, in turn, causes visible light passing through the lens to scatter [13]. The opacification of the lens is due to high molecular-weight protein aggregates [15], reducing the sharpness of the image reaching the retina [16]. Forms of cataract include congenital cataract, caused by inherited mutations, and age-related cataract, caused by cumulative damage of the protein due to environmental [17]. Surgery remains the current treatment for cataract. It has been proven to be quick and effective, however it does have a risk of infection or bleeding that can lead to complete vision loss [16, 17]. Surgery is also expensive and people in developing countries do not hold access. Efforts set forth in understanding the eye lens and its transparency can result in alternative methods towards treating cataract as well as an understanding the mechanism of cataract

1
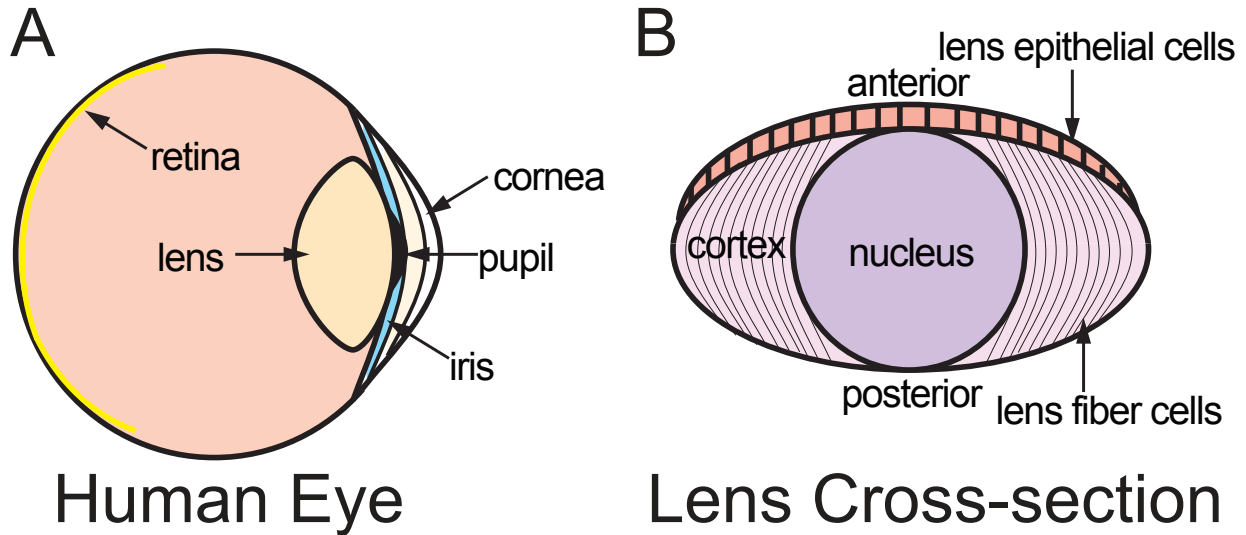
Figure 1.1: Schematic representation of the human eye lens. (A) Light passes through the cornea, iris, and lens to refract to the back of the retina. (B) Lens cross section containing the lens epithelial cells and lens fiber cells.

formation at the molecular level.

1.2 The lens's role in vision

1.2.1 Focusing power in on the human eye lens

The human eye is an important sensory organ giving us the ability to see due to the refraction of light. Starting from the outer most layer working in, the eye (Figure 1.1A) consists of the cornea, providing a protective membrane layer against infection and structural damage, the iris, which controls the size of the pupil, thus the amount of visible light reaching to the retina to form an image [18, 19]. The three transparent structures surrounding the ocular layers are the aqueous humor, the vitreous humor, and the lens [18].

The lens itself contains three important components: the capsule, the epithelial cells, and the fiber cells [20, 18, 21]. The lens capsule acts as a protective membrane layer for the

lens, as well as controlling the forces necessary to adjust the shape of the lens [20]. The lens epithelial cells are located inside the anterior capsule where the cells divide and migrate towards center of the lens (Figure 1.1B) [21]. During maturation, epithelial cells elongate to form fiber cells that stretch lengthwise in concentric layers from the posterior poles to the anterior poles of the lens [20, 21]. Before the lens has reached maturity, the fiber cells express structural proteins called crystallins in concentrations varying across the lens tissue, creating a refractive index gradient increasing from the cortex towards the nucleus [20]. The lens has no protein turnover, meaning when the lens has matured, the fiber cells lose their organelles and are unable to express more crystallins [21].

1.2.2 Crystallins: Robust proteins for life

The crystallins make up to 90% of total protein in the tissue [22] where concentrations are as high as 450 mg/mL in humans and even higher for some other species [23]. These crystallins are split into two classes: α-crystallins and $\beta\gamma$-crystallins [23, 24]. The α-crystallins are molecular chaperone proteins that belong to the small heat-shock protein superfamily [25, 26]. The $\beta\gamma$-crystallins are highly stabile structural proteins necessary for maintaining transparency and refractivity in the lens [21, 23, 27]. $\beta\gamma$-Crystallins share structural similarity where each domain is comprised of two intercalated antiparallel β-sheet Greek key motifs [21, 28, 3]. The Greek key motif gives high thermodynamic and kinetic stability between $\beta\gamma$-crystallins [29], where γ-crystallins are known to be more stable than the β-crystallins [21].

The focus of this thesis is on human γS-crystallin, which is expressed abundantly in the lens cortex (periphery) [21, 30, 31]. γS-Crystallin is less stable than other branches of γ-crystallins [21]. There are four cataract variants associated with γS-crystallin: γS-V42M, where the methionine mutation disrupts the compact β-sheet packing causing bilateral dense
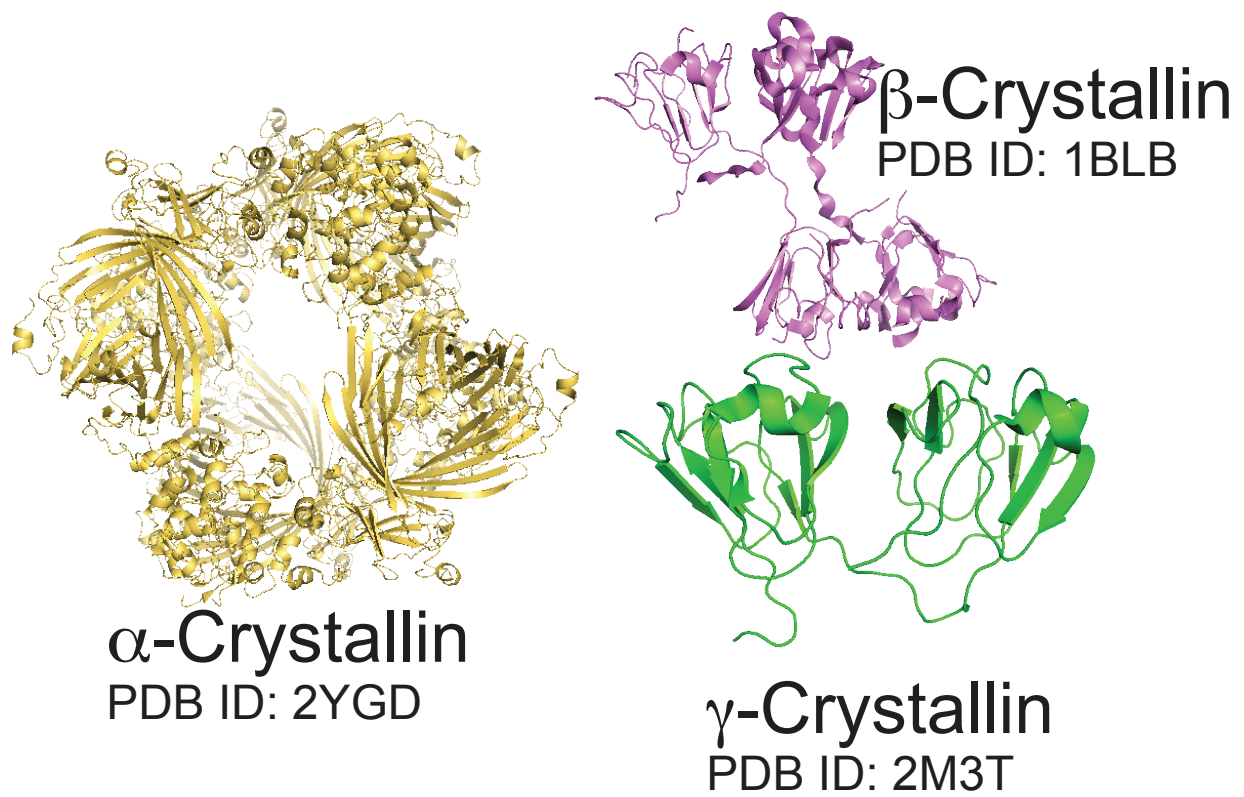
Figure 1.2: Structures of example α-, β-, and γ-crystallins. The structure of α-crystallin is of the human αB (PDB ID: 2YGD) [1]. The structure of β-crystallin is of bovine βB2 (PDB ID: 1BLB) [2]. The structure of γ-crystallin is of human γS-crystallin (PDB ID: 2M3T) [3].

cataract in children [32]; γS-D26G, where tertiary structural changes lead to changes in net intermolecular interactions causing congenital Coppock-type cataract [33]; γS-S39C, where preliminary data suggests a minor disruption in one of the Greek key folds leading to congenital cataract and microcornea [34]; and γS-G18V, which causes childhood-onset cortical cataract [21]. Comparing wild-type γS-crystallin (γS-WT) with γS-G18V, the two structures are fairly similar where the variant shows slight structural changes in the N-terminal domain, particularly near the mutation site [3]. Although the structural changes are minor, significant perturbations occur throughout the N-terminal domain resulting in altered intermolecular interactions leading to aggregation, and ultimately, cataract formation.

1.2.3 Understanding protein aggregation at the molecular level

Protein aggregation is a chemical phenomenon where protein molecules interact in such a way to form large insoluble complexes. These interactions are often tied to protein folding and stability, where disruption is observed in the high order structure [35]. Mechanisms of aggregation include depletion of chaperone activity [3, 36], hydrophobic interactions [37], electrostatic interactions [38], disulfide linkages [39], salt-bridge formation [39], thermal destabilization [36], deamidation [40], and other post-translational modifications. In particular to the crystallins, these proteins must last a lifetime while maintaining transparency and high refractive power to properly focus light. Crystallins must be stable at high concentrations and in compact organizations, making them susceptible to aggregation if they are damaged. Efforts must be made to help understand and develop a mechanistic model of protein aggregation which can lead to aiding pharmaceutical drug development into reversing cataract disease.

As we understand them, crystallins are highly stable proteins which exist in elevated concentrations within the eye lens. Crystallins are packed in order to maintain stablity and

solubility for an organisms lifetime. Due to their high refractive power, measuring the crystallins refractive index increment (dn/dC) can provide a valuable information enabling us to further understand the functions of these protein. However, minimal literature exists regarding the measurement of any proteins dn/dc, as lysozyme studied as a standard almost exclusively. Human γS-cyrstallin is found in the cortical lens epithelial cells, and previous work has shown that γS-WT is thermally stable and less aggregation prone that γS-G18V [4]. Chaperone studies involving αB-crystallin shows strong binding to the cataract variant γS-G18V, whereas no binding is observed in γS-WT [3]. These studies have concluded that the cataract variant of γS-crystallin is not only aggregation prone, but interacts strongly with its chaperone protein. Although there are no evident mechanisms describing potential aggregation pathways for γS-crystallin, I have plenty of biophysical methods available that could provide a better understanding in the solubility of the crystallins.

Thus, the goals of my thesis are as follows:

- Accurately measure the refractive index increment (dn/dC) for crystallins of aquatic and terrestrial species to compare among one another

- Develop a working model to understand what factors influence a dn/dC value by comparing experimental values with the present theoretical model

- Characterize the changes in exposed hydrophobic surface area between wild-type γS-crystallin and the cataract variant G18V to help elucidate the mechanisms of aggregation propensity and chaperone recognition in the context of cataract formation

- Investigate the interaction potential of the N-terminal extension loops of γS-WT and γS-G18V and to help understand how the N-terminal extension loop plays a role in cataract formation

- Use experimental and theoretical approaches to characterize protein-protein and protein-solvent interactions

- Express, purify, characterize an eye lens protein from the box jellyfish (*Tripedalia cystophora*), which is believed to have a novel fold

Overall, this thesis uses methods to investigate protein aggregation through protein-protein interactions. Probing into these interactions will ultimately help understand lens transparency and refractivity.

Chapter 2

Protein refractive index increment is determined by conformation as well as composition

2.1 Background

The crystallins of the eye lens are extremely stable, soluble proteins responsible for maintaining its transparency and providing sufficient refractive power to focus light onto the retina [20]. The crystallins' solubility and and stability are possibly why these proteins were selected for transparency [41].The high refractivity of the eye lens results from two major contributions; the high protein concentration (between 700 to 1000 mg/mL in some aquatic species) [42, 43, 44, 45] and the high refractive indices of the crystallin proteins themselves [46]. The latter is often expressed as the change in refractive index with concentration, or refractive index increment (dn/dC). In lens proteins, where the increased refractive index is functionally important, dn/dC is generally higher than for proteins not selected for this function [47].

Two major classes of crystallins, α-crystallins and $\beta\gamma$-crystallins, make up 90% of the vertebrate eye lens [48, 49]. α-Crystallins are small heat shock proteins that are most abundant in mammalian species [50]. $\beta\gamma$-Crystallins are structural proteins that share a Greek key motif [51, 50] and are the predominant class of crystallins in fish lenses [48, 49], where they contribute to the high concentration in the nucleus and cortical region [51, 33]. In particular, γ-crystallins are the more stable class of crystallins [52]. Mammalian γ-crystallins contain a significantly larger amount of methionine and/or cysteine residues [49] as well as a set of conserved tryptophan residues [51]. The amount of sulfur-containing residues is even higher in fish γM-crystallins [51]. The high content of methionine and cysteine are believed to mediate intermolecular interactions that could enhance solubility and dense packing in lenses [51].

Crystallins from most species, such as cephalopods (S-crystallins) [53] and jellyfish (J1- and J3-crystallins) [42], are derived from proteins involved in gene-sharing [42, 53]. For example, S-crystallin shared high sequence homology with the metabolic enzyme glutathione S-transferase [43] while J1- and J3-crystallins show similarity to ADP-ribosylglycohydrolases and vertebrate saposins, respectively[42]. In contrast, J2-crystallin has shares no sequence homology with other proteins and is preferentially expressed in the eye lens [42]. Comparing the eye lenses among vertebrate versus invertebrate and terrestrial versus aquatic species provides better insight the crystallin proteins. A refractive index increment is a measurement capable of comparing the refractive power of the crystallins.

The refractive index of a solution is given by the relationship between the angle of incident light and its refraction when passing through two different isotropic media, otherwise known as Snell's Law [54]. Taking solutions at varying concentrations and reporting their refractive indices yields the dn/dC value (the change in refractive index over the change

in concentration) [44, 55, 56], measured using a differential refractometer. Experimentally, protein or polymer solutions are injected at increasing concentrations through a flow cell. A laser source is used to shine light on the sample in a split flow cell, the reference solution in one half and the sample solution flowed continuously in the other, where the refraction proportional to the difference between the sample and reference solutions is measured. The dn/dC value measured is used for analytical untracentrifugation, surface plasmon resonance (SPR), and protein molecular mass from light scattering measurements [44, 55, 57]. Typically, a general dn/dC value used is for all proteins (0.185 g/mL) [58]. However, dn/dC values have been measured for a few proteins such as bovine serum albumin (BSA) and hen egg white lysozyme (HEWL). Although BSA and HEWL are typical protein standards used for calibration, their dn/dC values can vary due to solvent (in this case buffered solutions), wavelength of light, and even temperature [56, 59, 60], resulting in conflicting data in the literature. These conflicting data may also rise due to using narrow sample ranges used (between 0.5 to 5 mg/mL). Expressing and purifying these protein samples at similar concentrations with the appropriate volumes can prove to be labor intensive and require large amounts of sample, making it difficult for measurement of other proteins.

Here we report the measured dn/dC values of several vertebrate and invertebrate crystallins for the first time and compare the refractive power of these proteins across different species. Since these eye lens proteins' purpose is to refract light, we hypothesize that their dn/dC values will be significantly higher than other proteins, where the average value is 0.185 mL/g. To our knowledge, the literature has not shown dn/dC measurements for lens crystallins. However, a prediction model using the weighted average of individual dn/dC residue measurements has been reported [44]. Under the assumption that all proteins have a narrow dn/dC value, Zhao and coworkers developed a calculator to predict dn/dC values by taking the weighted average of known amino acid dn/dC values using unmodified protein sequences[44]. Contra the prevailing model, we find that for the lens crystallins that the

measured dn/dC values are much higher than those predicted using amino acid composition alone. To accurately measure the dn/dC, approximately 750 mg of each crystallin protein was prepared at larger concentration ranges, making this method labor intensive, yet thorough. With these dn/dC measurements of various crystallins, we will able to show that the current calculated model does not match for proteins selected for high refractive power and that other factors may be involved.

2.2 Methods Investigating the crystallin refractive index increment

2.2.1 Plasmid construction

Plasmids containing the cDNA sequence of human γS-, toothfish γS1-, toothfish γS2-, toothfish γM8b-, toothfish γM8c-, toothfish γM8d-, and box jellyfsh J2-crystallin were purchased from Blue Heron (Bothell, WA). Primers for human γS-, toothfish γS1-, toothfish γS2-, and J2-crystallins were purchased from Sigma-Aldrich (St. Louis, MO) while primers for γM8b-, γM8c-, and γM8d-crystallins were purchased from Integrated DNA Technologies (Coralville, IA). The crystallin genes were amplified with primers containing flanking restriction sites for NcoI and XhoI, an N-terminal 6x His tag, and a TEV cleavage sequence (ENLYFQ). The polymerase chain reaction product was cloned into a pET28(+)a vector, purchased from Novagen (Darmstadt, Germany). Human γS-crystallin was prepared by William D. Brubaker as previously described [4]. Toothfish γS1- and γS2-crystallin were prepared by Carolyn N. Kingsley as previously described [61].

2.2.2 Expression and purification of eye lens crystallins

A 50 mL starter culture of Luria Broth (LB) media was innoculated with a single colony of Rosetta (DE3) *Escherichia coli* containing the vector of either human γS-, toothfish γS1-, toothfish γS2-, toothfish γM8b-, toothfish γM8c-, toothfish γM8d-, or box jellyfsh J2-crystallin. Starter cultures were grown at 37 °C for 16 hours shaking at 225 RPM in a New Brunswick Scientific Innova-42R incubator Shaker (Hauppauge, NY). For human γS-crystallin and box jellyfish J2-crystallin, the individual starter cultures were used to innoculate a 1 L culture of LB with a starting OD_{600} of approximately 0.20. The cultures were grown at 37 °C with shaking at 225 RPM until an OD_{600} of 0.60 was reached. Protein overexpression was induced using IPTG (Gold Biotechnology) at a final concentration of 0.10 mM at 37 °C for 8 hours (γS-WT) and 25 °C for 18 hours (J2-crystallin). The toothfish crystallins were overexpressed using the Studier autoinduction protocol at 25 °C for 24 hours. Cells were collected via centrifugation with a Beckman Coulter Avanti-JE centrifuge (Brea, CA) spinning at 6000 RPM for 20 minutes. The cell pellets were resuspended in 40 mL of 50 mM sodium phosphate buffer with 300 mM sodium chloride, 10 mM imidazole, 0.05% sodium azide at pH 7.4. Cells were lysed by sonication in 10 second intervals for a total of 30 minutes, followed by centrifugation at 15000 RPM for 90 minutes. The supernatant was filtered with through a 0.22 μm filter (Millipore) before being loaded onto a Bio-Rad Duo-Inject FPLC system (Hercules, CA). The His-tagged crystallins were purified using a Ni-IDA column (Bio-Rad) and cut by a His-tagged TEV protease (produced in-house). The TEV protease and His-tag were removed by a second application to a Ni-IDA column. The crystallins were then run on a HiLoad 16/600 Superdex 75 PG gel filtration column from GE (Pittsburgh, PA) using10 mM sodium phosphate buffer, 100 mM sodium chloride, 0.05% sodium azide at pH 6.9. All of the toothfish crystallins were expressed and purified by Jan C. Bierma.

Tunicate (*Ciona intestinalis*) Ci-$\beta\gamma$-crystallin was expressed and purified as previously de-

Table 2.1: Table of extinction coefficients for various crystallins.

Crystallin	ϵ at 280 nm (mL*mM^{-1}cm^{-1})
Human γS	2.05
Toothfish γS1	2.16
Toothfish γS2	2.30
Toothfish γm8b	1.19
Toothfish γm8c	1.06
Toothfish γm8d	1.06
Ci-$\beta\gamma$	1.50
Hen Egg White Lysozyme	2.64
Box jellyfish J2	0.329

scribed in Kozlyuk *et al.* [62].

2.2.3 Sample preparation and refractive index increment experiments

Lyophilized lysozyme from chicken egg white (Cat. No. 195303) was purchased from MP Biomedicals (Solon, OH). Lysozyme was dissolved 10 mM sodium phosphate buffer, 100 mM sodium chloride, 0.05% sodium azide at pH 6.9 for a final concentration of 50 mg/mL. Crystallin proteins were concentrated down to a stock solution of 50 mg/mL. Stock solutions for all proteins were used to make samples with concentrations ranging from 0.5 to 30 mg/mL. The concentrations were checked by UV absorbance measurements at 280 nm using the following extinction coefficients [63]:

Refractive index increments were measured following the batch-mode technique using an Optilab rEX refractive index detector (Wyatt Technology, Santa Barbara, CA) configured with a 685 nm fiber-optic laser diode source.

2.2.4 Calculation of dipole moments

For proteins where a structure has been solved, PDB files were used. For proteins without a solved structure, the structures were predicted using the Robetta server [64, 65]. To calculate protonation states, the PDB files were uploaded to the H++ server, version 3.2 [66, 67, 68], with a pH set at 6.9, salinity at 100 mM and the internal and external dielectric constants set at 10 and 80, respectively. The resulting structures were opened in VMD [11] and the Dipole Watcher plugin, version 1.0 was used to calculated the magnitude of the dipole [69].

2.3 Results and discussion

The species of crystallins for which dn/dC was measured included human, Antarctic toothfish, box jellyfish, and tunicate. Table 2.2 contains the dn/dC values measured for the crystallins. Lysozyme dn/dC measurements were performed as a control where we compared our results to literature values, which did indeed match [70]. However, the crystallin dn/dC measurements agreed with our hypothesis of these lens proteins having a higher dn/dC value rather than the prevailing model. With the exception of Ci-$\beta\gamma$-crystallin, the terrestrial and aquatic lenses have a dn/dC of 0.20 mL/g or above, significantly higher than the 0.185 mL/g general average for proteins. The toothfish crystallins showed some of the highest dn/dC values, consistent with the fact that these are aquatic species and must have lenses with a higher refractive power to make up for the loss of the air/water interface terrestrial species have.

We compared our measured results with the predicted dn/dC values using a dn/dC calculator [44]. Zhao and coworkers developed a model to predict the dn/dC value of a protein based on the weighted average dn/dC of the residues within the protein. Figure 2.1 contains

Table 2.2: Table representing the experimentally measured dn/dC values and predicted dn/dC values of crystallins from various species. All of the crystallins measured have dn/dC values significantly larger than the average dn/dC for proteins.

Crystallin	Calculated dn/dC (mL/g)	Measured dn/dC (mL/g)
Human γS	0.1985	0.2073 ± 0.0014
Toothfish γS1	0.2020	0.2183 ± 0.0014
Toothfish γS2	0.2002	0.2168 ± 0.0014
Toothfish γM8b	0.2003	0.2158 ± 0.0015
Toothfish γM8c	0.2003	0.2061 ± 0.0014
Toothfish γM8d	0.1995	0.2041 ± 0.0014
Ci-$\beta\gamma$	0.1917	0.1985 ± 0.0012
Hen Egg White Lysozyme	0.1963	0.1970 ± 0.0010
Box jellyfish J2	0.1920	0.2037 ± 0.0012

the histogram of the predicted dn/dC values of the human proteome using the calculator developed by Zhao and coworkers [44]. The 62,378 protein sequences were gathered from the University of California Santa Cruz (UCSC) genome browser for the human February 2009 assembly (GRCh37/hg19) [44]. Superimposed on this histogram are the predicted and measured dn/dC values are plotted to compare with the predicted dn/dC model. As it was presented by Zhao and coworkers, the distribution of the human proteome predicted dn/dC values is close to Gaussian with the average dn/dC reported at 0.1899 mL/g and a small standard deviation of 0.0030 mL/g [44]. In the case of hen egg white lysozyme, the predicted dn/dC value matched the experimentally measured dn/dC value. The eye lens crystallins show something significantly different; not only are the predicted dn/dCs higher than the predicted average, which are located at the upper tail of the distribution, but the experimentally measured dn/dC values exceed their predicted values. In fact, the spread of the the measured value spans the width of the distribution. Experimental dn/dC values conclude that amino acid composition is not the sole contributor to a crystallin's dn/dC value, since the crystallins are selected for high refractivity in the eye lens. The distribution of the dn/dC values show that the current model of predicting dn/dC values must be updated, involving significant components of a protein's structure.
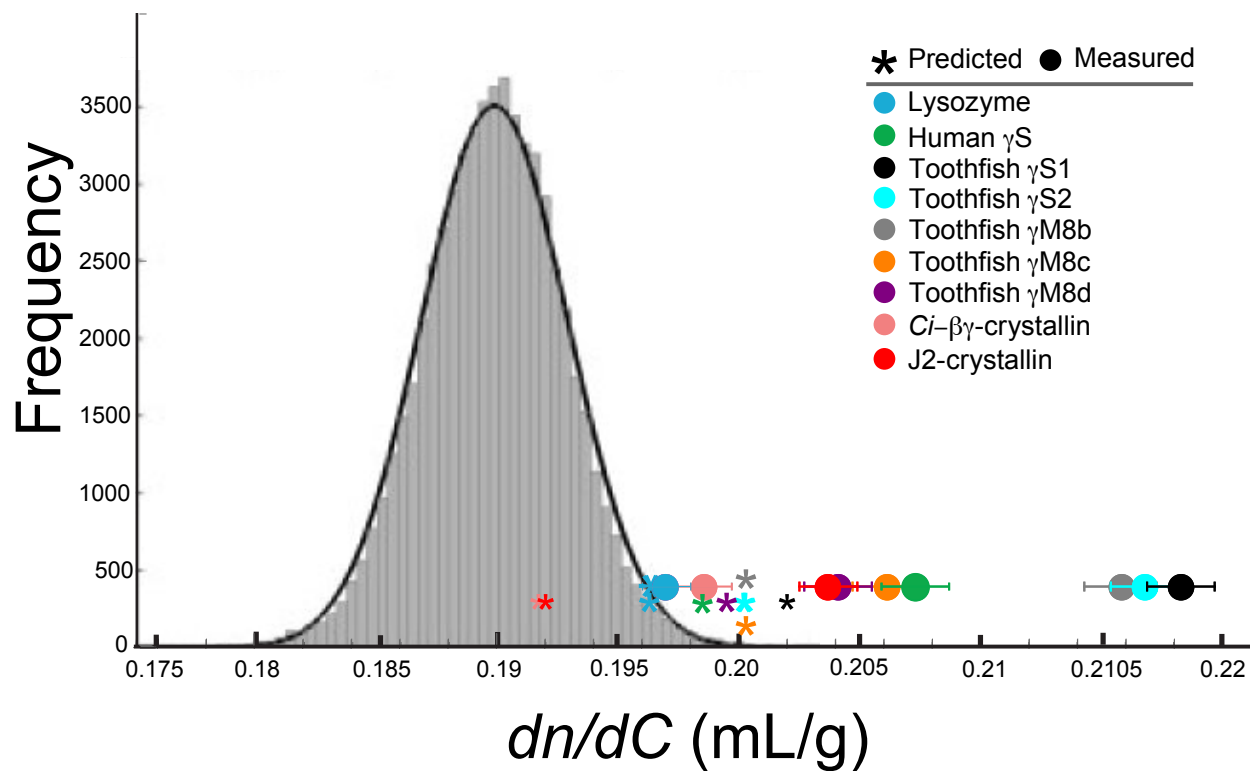
Figure 2.1: Histogram of predicted dn/dC values for the 62,378 proteins in the human proteome, gathered from the University of California Santa Cruz (UCSC) genome browser February 2009 assembly (GRCh37/hg19). Predicted dn/dC values and experimentally measured dn/dC values of crystallins are overlayed on the histogram indicated in an asterisk and closed circle, respectively. Predicted and measured crystallin dn/dC values are significantly higher than the average dn/dC of 0.1899 mL/g. The experimentally measured dn/dC values surpass their predicted dn/dC values for crystallin, which fall within the highest end of the Gaussian distribution.

Based on our results, we conclude that amino acid composition is not the sole determinant of a protein's dn/dC value. Instead, we suggest that the 3-dimensional (3D) structure and polarization could contribute to the overall dn/dC value. The 3D structure of a protein consists a main chain of covalently linked amino acids, which twists and bends to a low-energy state stabilized by intramolecular interactions among the side-chain groups. The sum of these intramolecular interactions creates an electric field, which can be visualized as simple distribution of charge [71]. These interactions can include hydrogen bonding interactions, hydrophobic interactions, electrostatic interactions, and pi-pi interactions in addition to other attractive/repulsive interactions [71]. Hydrogen bonding is often described as a dipole-dipole interaction [71]. Aromatic residues in close proximity exhibit resonance energy transfer amongst a donor molecule and acceptor molecule following Förster's theory describing molecular interactions [72]. This resonance energy transfer introduces larger dipole moments within the protein itself, potentially providing a larger influence on the protein's dn/dC. Figure 2.2 (right panel) highlights the aromatic resides that could contribute towards Förster resonance transfer for all crystallins measured. Aromatic residues make up close to 20% of human γS-, toothfish γS1- and γS2-crystallins while on average, proteins contain between to 11-12% [73].

Additionally, the protein's own electric field forms a dynamic response due to reorientation of its molecules to optimize its interactions with the electric field, which could result in the formation of large dipoles [74, 75, 76]. Figure 2.2 (left panel) shows the overall dipole moments for all crystallins measured using VMD [11] following the methods from Dixit *et al.* [12]. Although these predicted dipole moments are based on comparative structures, the large dipole moments calculated for *Ci-$\beta\gamma$*-crystallin (469.9 D), toothfish γS2 (463.8 D), toothfish γS1 (383.6 D), and human γS (259.1 D). Table 2.3 contains the calculated dipole moments for the crystallins. It has been suggested that a protein, or even a class of proteins, with a large dipole moment is considered to have a highly specific role [76]. In the case of crystallins, that role is to refract light in the eye lens.
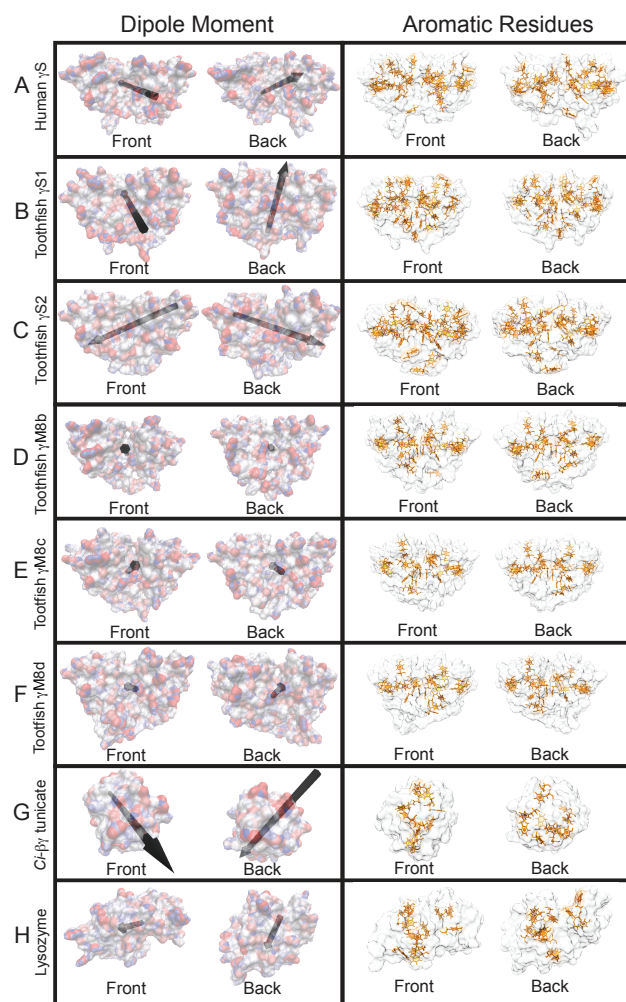
Figure 2.2: Molecular surface representations of (A) human γS-crystallin and (B) toothfish γS1-crystallin, (C) γS2-crystallin, (D) toothfish γM8b-crystallin, (E) toothfish γM8c-crystallin, (F) toothfish γM8d-crystallin, (G) Ci-$\beta\gamma$-crystallin, and (H) lysozyme. The left panel portrays the calculated overall dipole moment. The right panel portrays the highlighted aromatic residues in orange which are possibly contributing towards Förster resonance transfer resulting in dipole-dipole interactions.

Table 2.3: Table containing the calculated dipole moment of crystallins used in this study. The overall dipole moments for all crystallins were measured using VMD [11] following the methods from Dixit *et al.* [12].

Crystallin	Estimated Dipole Moments (D)
Human γS	259.1
Toothfish γS1	383.6
Toothfish γS2	463.8
Toothfish γm8b	209.4
Toothfish γm8c	163.7
Toothfish γm8d	108.2
Ci-$\beta\gamma$	173.9
Hen Egg White Lysozyme	383.6

2.4   Conclusion

For the first time, I have accurately measured the refractive index increment of crystallins from aquatic and terrestrial species. These measured dn/dC values finally allow for comparison among terrestrial and aquatic crystallins, which are proteins responsible in refracting light. Although the current model developed by Zhao *et al.* of calculating dn/dC values using amino acid composition is very quick and efficient, to really understand the differences among crystallins and comparing them among other species, we put forth the effort to experimentally measure. Using a wide concentration range, our results showed that these crystallins not only have higher dn/dC values than predicted, but that they vary by quite a large amount indicating that other components must be taken into account to determine a protein's dn/dC. Future studies will include the measured dn/dC values of an expanded crystallin library that includes more terrestrial and aquatic species as well as including crystallins that are expressed in only one area of the lens (nucleus versus cortex). We would also like to explore how the dn/dC value is affected by measuring the crystallins in a partially unfolded environment, giving much insight on how these intramolecular interactions really play a role.

Chapter 3

Increased hydrophobic surface exposure in the cataract-related G18V variant of human γS-crystallin

3.1 Background

The solution-state NMR structure of wild-type γS-crystallin has been determined [3], revealing a double Greek key architecture for each of the two domains, consistent with the structures of other $\beta\gamma$-crystallins [28, 77]. The childhood-onset cataract variant G18V (γS-G18V) is structurally similar to γS-WT, but it has dramatically lower thermal stability and solublity [4, 21], as well as strong, specific interactions with αB-crystallin, the holdase chaperone of the lens [3]. Despite the well-documented aggregation propensity and reduced stability of γS-G18V, the particular intermolecular interactions leading to its aggregation are as yet unknown. Protein self-aggregation leading to cataract can occur due to an increase in net hydrophobic interactions, as previously shown in the congenital Coppock-type

cataract variant D26G γS-crystallin [33], the cerulean cataract variant P23T γD-crystallin [78], acetylation of the G1 and K2 residues in γD-crystallin [23], and the lamellar cataract variant D140N αB-crystallin [79]. All of these mutations introduce altered conformations that produce lowered solubility by exposure of hydrophobic patches on the surface, even though the structural differences from their wild-type counterparts are relatively subtle. γS-G18V is no exception; the mutation does not cause large-scale unfolding or rearrangement into a misfolded conformation, but rather produces altered intermolecular interactions with itself and with αB-crystallin [3].

The fluorescent probe 1-anilinonaphthanlene-8-sulfonate (ANS), which has both negatively charged and hydrophobic moieties, is often used to quantify exposed hydrophobic surface patches in proteins by introducing known concentrations of ANS into a protein solution and measuring its emission spectrum [78, 80, 81]. Two types of protein-ANS interactions are required for fluorescence enhancement: hydrophobic interactions between the conjugated ring system of ANS and the protein surface [82], and electrostatic interactions between the sulfonate group and positively charged side chains at the binding site [83]. An increase in fluorescence intensity indicates that either more ANS is binding to the protein surface, or that it is bound more tightly, correlating with higher surface hydrophobicity. This method has been used to characterize exposed hydrophobic surface in a number of protein systems, including the mitochondrial chaperone protein Atp11p, which recognizes its client proteins via hydrophobic interactions [84], and aggregation-prone variants of superoxide dismutase-1 (SOD1), an essential cellular enzyme whose aggregation is associated with amyotrophic lateral sclerosis (ALS) [85, 86]. Despite the utility of ANS binding as a probe of hydrophobic surface exposure, and the sensitivity afforded by using fluorescence as a reporter, this assay is limited by the lack of detailed information about which amino acid residues, or even general regions of the protein, are taking part in the dye-binding interaction. NMR chemical shift perturbation (CSP) mapping can forge a link between fluorescence enhancement upon dye

binding and the corresponding changes in the local chemical environment of specific residues in the protein.

Nuclear Magnetic Resonance (NMR) is an analytical technique where structural information is gathered based on the nuclear spin characteristics of the protein of interest. By using a 2-D correlation experiment such as ^{1}H-^{15}N HSQC, chemical shifts are assigned that represent each N-H correlation within the protein, thereby identifying each residue or sidechain. Of course, there must be an existing protein structure to use the HSQC assignments. The addition of a ligand allows for determination of binding sites within the protein [87]. Using ANS titration experiments, comparisons between wild-type and variant proteins can then be used to compare differences in exposure of hydrophobic residues on the surface under particular solution conditions. CSP mapping is a commonly used technique for investigating protein-protein or protein-ligand interactions and interfaces [87], and is the basis of the "SAR by NMR" methodology that is indispensable in the identification of active pharmaceutical agents [88].

Molecular docking, a computational technique used widely to model the conformation of protein-ligand complexes, enables experimental perturbations to be analyzed in atomistic detail. Bound ligand conformations, or poses, are ranked using an empirical scoring function designed to evaluate intermolecular interactions using minimal computational time. Conventionally, knowledge of the active site is used to guide the pose generation, often in the context of screening large libraries of compounds against known protein structures [89, 90, 91, 92]. However, docking protocols without prior knowledge of the active site (blind docking) [93], have successfully identified putative allosteric binding sites of drugs, leading to the design of novel allosteric modulators [94], and fluorescent dyes [95, 80]. Bis-ANS binding sites found by docking, validated with steady-state and time-resolved fluorescence assays, have

22

been used to identify hydrophobic patches in a lipase from *Bacillus subtilis* [96]. Using ANS binding mapped by NMR in conjunction with molecular docking, we focus on determining whether the cataract-related G18V variant of human γS-crystallin has increased exposure of hydrophobic residues that could explain its aggregation propensity and/or recognition by αB-crystallin.

3.2 Materials and methods used to investigate γS-crystallin interacting with ANS

3.2.1 Plasmid construction

Plasmids containing the cDNA sequences γS-WT and γS-G18V crystallin were purchased from Blue Heron (Bothell, WA) while primers were purchased from Sigma-Aldrich (St. Louis, MO). The crystallin genes were amplified with primers containing flanking restriction sites for NcoI and XhoI, an N-terminal 6x His tag, and a TEV cleavage sequence (ENLYFQ). The polymerase chain reaction product was cloned into a pET28(+)a vector, purchased from Novagen (Darmstadt, Germany). The γS-crystallin proteins were prepared by William D. Brubaker as previously described [4].

3.2.2 Expression and purification of natural abundance γS-WT and γS-G18V

A 50 mL starter culture of LB media was innoculated with a single colony of Rosetta (DE3) *E. coli* containing a pET28(+)a vector, purchased from Novagen (Darmstadt, Germany). The vector containing either γS-WT or γS-G18V gene inserts was grown at 37 °C for 16 hours at 225 RPM in a New Brunswick Scientific Innova-42R incubator Shaker (Hauppauge, NY). The individual starter cultures were used towards a 1 L culture of LB with a starting OD_{600} at approximately 0.20. The cultures were grown at 37 °C with shaking at 225 RPM until an OD_{600} of 0.60 was reached. Protein overexpression was induced using IPTG (Gold Biotechnology) at a final concentration of 0.10 mM at 37 °C for 8 hours (γS-WT) and 25 °C for 18 hours (γS-G18V). Cells were collected via centrifugation with a Beckman Coulter Avanti-JE centrifuge (Brea, CA) spinning at 6000 RPM for 20 minutes. The cell pellets were resuspended in 40 mL of 50 mM sodium phosphate buffer with 300 mM sodium chloride, 10 mM imidazole, 0.05% sodium azide at pH 7.4. Cells were lysed by sonication in 10 second intervals for a total of 30 minutes, followed by centrifugation at 15000 RPM for 90 minutes. The supernatant was filtered with through a 0.22 μm filter (Millipore) before being loaded onto a Bio-Rad Duo-Inject FPLC system (Hercules, CA). The His-tagged crystallins were purified using a Ni-IDA column (Bio-Rad) and cut by a His-tagged TEV protease (produced in-house). The TEV protease and His-tag were removed by a second application to a Ni-IDA column. The purified crystallins were dialyzed into 10 mM sodium phosphate buffer with 0.05% sodium azide at pH 6.9.

3.2.3 Expression and purification of ^{15}N-labeled γS-WT and γS-G18V

A 50 mL starter culture of LB media innoculated with a single colony Rosetta (DE3) *E. coli* cell containing a pET28(+)a vector with either γS-WT or γS-G18V gene inserts was grown at 37 °C for 16 hours at 225 RPM. The individual starter cultures were used to innoculate a 1 L culture of LB with a starting OD_{600} at approximately 0.20. The cultures were grown at 37 °C at 225 RPM until an OD_{600} of 0.60 was reached. The cells were then collected in 500 mL batches by centrifugation at 3000 RPM for 30 minutes and each 500 mL batch was resuspended in 1 L ^{15}N-labeled minimal media cultures. The 1 L minimal media cultures were grown for an additional 2 hours at 37 °C at 225 RPM. Protein overexpression was induced using IPTG (Gold Biotechnology) at a final concentration of 0.10 mM at 25 °C for 30 hours for both γS-WT and γS-G18V. Cells were collected via centrifugation, spinning at 6000 RPM for 20 minutes. The cell pellets were resuspended in 40 mL of 50 mM sodium phosphate buffer with 300 mM sodium chloride, 10 mM imidazole, and 0.05% sodium azide at pH 7.4. Cells were lysed by sonication in 30 second intervals for a total of 10 minutes, followed by centrifugation at 15000 RPM for 90 minutes. The supernatant was filtered with through a 0.22 μm filter (Millipore) before being loaded onto a Bio-Rad Duo-Inject FPLC system (Hercules, CA). The His-tagged crystallins were purified using a Ni-IDA (Bio-Rad) and cut by a His-tagged TEV protease (produced in-house). The TEV protease and His-tag were removed by a second application to a Ni-IDA column. The purified ^{15}N-labeled crystallins were dialyzed into 10 mM sodium phosphate buffer with 0.05% sodium azide at pH 6.9.

3.2.4 ANS fluorescence assay

Fluorescence spectra were collected as a function of ANS concentration for γS-WT and γS-G18V with a F4500 Hitachi fluorescence spectrophotometer with slits set to 5 nm. The excitation and emission wavelengths were 390 nm and 500 nm, respectively. Protein concentrations for both γS-WT and γS-G18V were approximately 1 mg/mL in 10 mM sodium phosphate buffer and 0.05% sodium azide at pH 6.9. ANS concentrations ranging from 5 μM to 2 mM were measured using $\epsilon = 4.95$ mM^{-1} cm^{-1} at 350 nm [97].

3.2.5 Dynamic Light Scattering (DLS) Measurements

Dynamic light scattering (DLS) measurements were collected for γS-WT and γS-G18V with a Malvern Zetasizer ZS Nano DLS. Protein concentrations for both γS-WT and γS-G18V were approximately 0.3 mM in 10 mM sodium phosphate buffer and 0.05% sodium azide at pH 6.9. ANS at concentration ratios of 1:0, 1:0.5, 1:1, and 1:2 of γS:ANS were measured to determine the aggregation state of crystallin.

3.2.6 NMR experiments

Experiments were performed on a Varian $^{\text{Unity}}$INOVA spectrometer (Agilent Technologies, Santa Clara, CA) operating at 800 MHz and equipped with a ^{1}H–^{13}C–^{15}N 5 mm tri-axis PFG triple-resonance probe, using an 18.8 Tesla superconducting electromagnet (Oxford Instruments). Decoupling of ^{15}N nuclei was performed using the GARP sequence [98]. ^{1}H chemical shifts were referenced to TMSP, and ^{15}N shifts were referenced indirectly to TMSP. NMR data were processed using NMRPipe [99] and analyzed using CcpNMR Analysis [100].

Center operating frequencies and (unless otherwise stated) center frequency offsets were as follows:

Center	[1]H: 799.8056964 MHz	[13]C: 201.1282461 MHz	[15]N: 81.0504078 MHz
Offset	[1]H: -294.932 Hz (4.8 ppm)	[13]C: -9863.17 Hz (43 ppm)	[15]N: 2400 Hz (116.7 ppm)

3.2.7 Calculation of chemical shift perturbations

^{1}H-^{15}N HSQC spectra of γS-WT and γS-G18V were collected in the presence and absence of ANS at concentration ratios of 1:0, 1:0.5, 1:1, and 1:2 of γS:ANS [101]. Resonances were identified and assigned based on chemical shift data previously collected by our group. The full NMR spectra can be is shown in Figures 3.1 and 3.2. Resonances showed perturbations that are indicative of ANS binding. The change in chemical shift for each peak in the 2D spectrum upon ANS binding was calculated using the following chemical shift perturbation (CSP) equation:

$$\Delta\delta_{avg} = \sqrt{\frac{(\Delta\delta_N/5)^2 + (\Delta\delta_H)^2}{2}} \tag{3.1}$$

A strong-binding threshold for each set of conditions was set at two times the root mean square (RMS) of the calculated CSP, while the weak-binding threshold was set at half the RMS to determine which residues had strong or weak binding with ANS. These calculated thresholds are shown in Table 3.1.

3.2.8 Binding site search by rigid receptor docking

Protein coordinates were obtained from the NMR structures of γS-WT and γS-G18V crystallins (PDB ID: 2M3T and 2M3U) [3]. Autodock Tools [102] was used to prepare both the
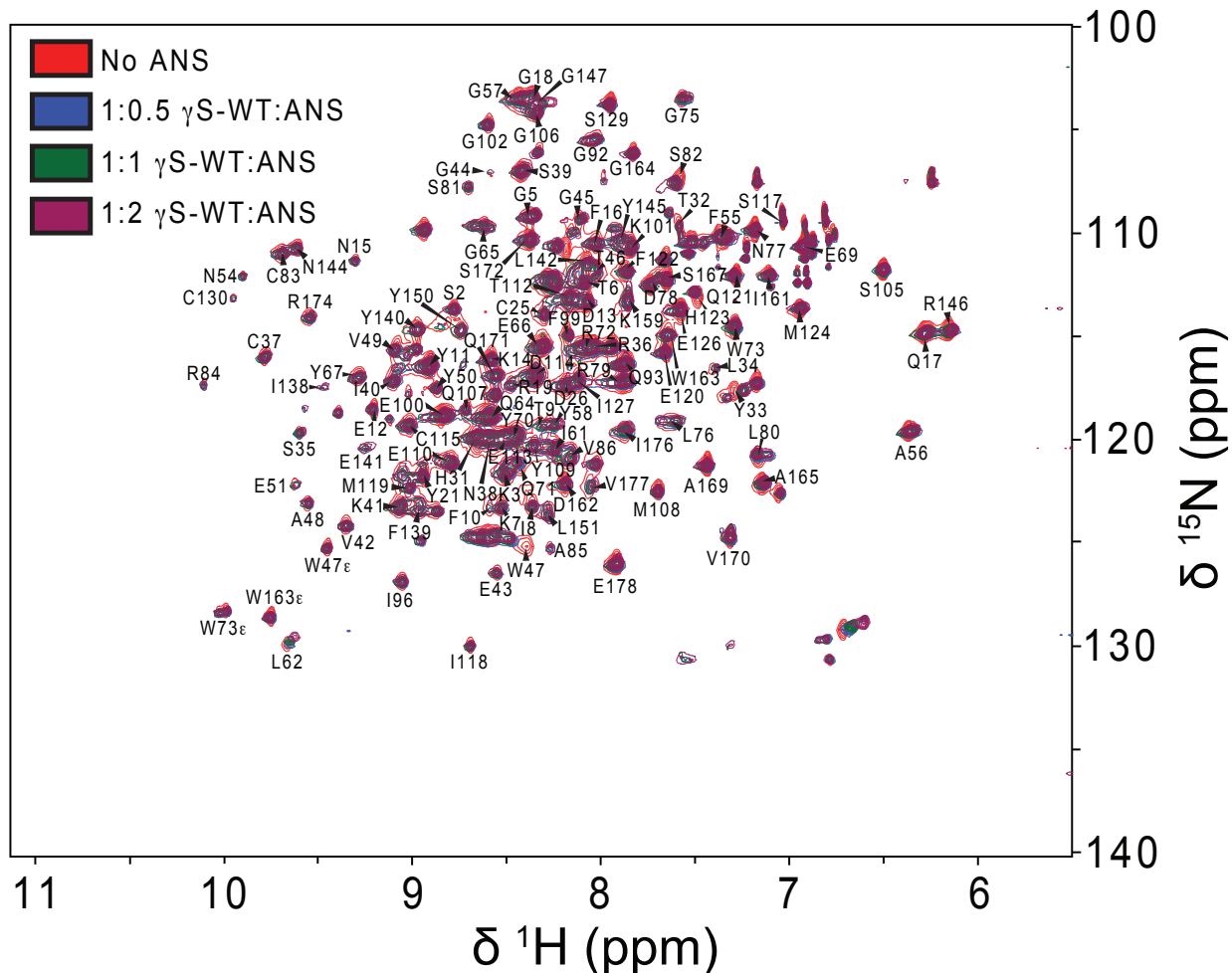
Figure 3.1: ^{1}H-^{15}N HSQC spectra of ^{15}N labelled γS-WT with increasing concentrations of ANS. Ratios of γS:ANS were 1:0, 1:0.5, 1:1, and 1:2 where the concentration of protein was approximately 0.3 mM. Spectra were acquired at 25 °C. Residues were assigned based on previous assignments of γS-WT [4].

Table 3.1: Root mean square (RMS) values used for ANS-residue contact frequency calculations and CSP calculations. The Boltzmann weighted-contact frequencies had a strong-binding threshold set at two times the RMS of the calculated ANS-residue contact frequency, while the weak-binding threshold was set at half the RMS value. Following the CSP calculations, a strong-binding threshold for each set of conditions of γS:ANS was set at two times the RMS of the calculated CSP, while the weak-binding threshold was set at half the RMS.

	Contact Frequencies		1:0.5	1:1	1:2
WT	strong-binding	0.02699	0.02420	0.02699	0.03382
	weak-binding	0.01349	0.00605	0.00675	0.00846
G18V	strong-binding	0.02345	0.02342	0.02345	0.02345
	weak-binding	0.01172	0.00586	0.00586	0.00586
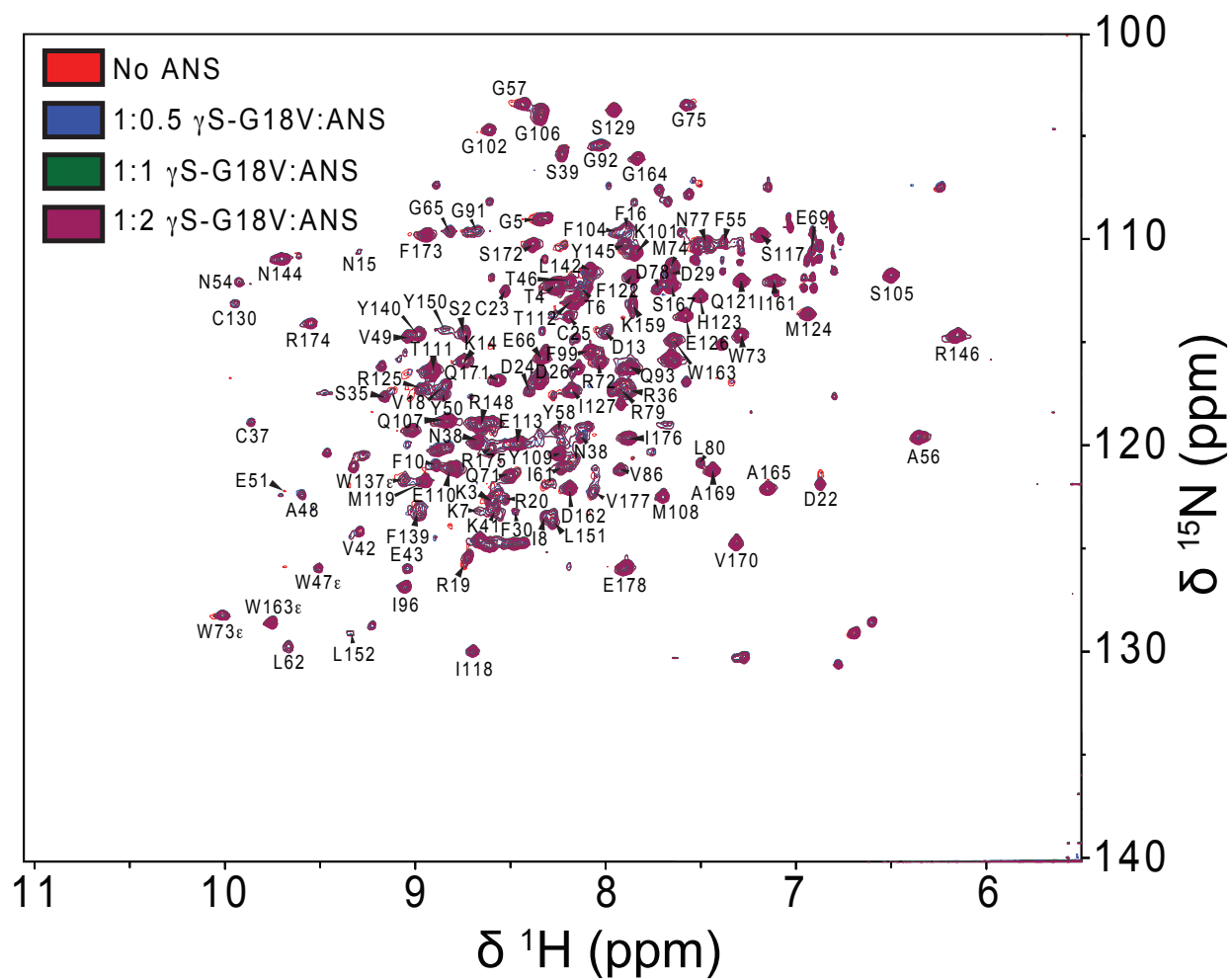
Figure 3.2: ^{1}H-^{15}N HSQC spectra of ^{15}N labelled γS-G18V with increasing concentrations of ANS. Ratios of γS:ANS were 1:0, 1:0.5, 1:1, and 1:2 where the concentration of protein was approximately 0.3 mM. Spectra were acquired at 25 °C. Residues were assigned based on previous assignments of γS-G18V [4].
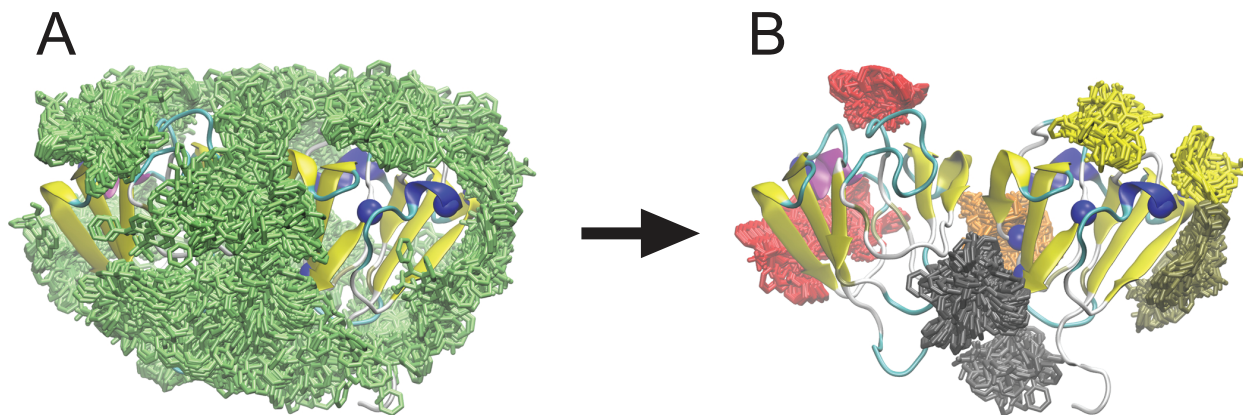
Figure 3.3: Clustering of rigid docking results to define flexible binding sites (γS-WT shown). A. Set of poses containing both hydrophobic and electrostatic contacts necessary for ANS fluorescence enhancement upon binding. The resulting set covers nearly the entire surface of the protein. B. Clustering of the screened docking set resulted in 20 binding sites. Pose clusters near highly perturbed residues (shown as blue spheres) were picked to define search spaces for flexible docking. The pose cluster colors distinguish different search spaces.

receptor (crystallin) and ligand (ANS) by merging non-polar hydrogens atoms into united heavy atoms. The docking simulations, run by Eric K. Wong with the Tobias Lab at UCI, was performed over each one of twenty solution-state NMR conformations for either γS-WT or γS-G18V [37]. The resulting poses were screened to ensure that both electrostatic and hydrophobic interactions required for ANS fluorescence enhancement upon binding were present. Docked poses that did not include both interactions within the first coordination shell of the ANS-protein radial distribution function were considered non-fluorescent and removed from the docked set. The screened docked set covers most of the protein surface as shown in Figure 3.3.

3.2.9 Calculation of residue contacts

To compare the screened docked set with the residue-based CSP data, ANS-residue contact frequencies were calculated by summing the Boltzmann weights of all the poses in contact
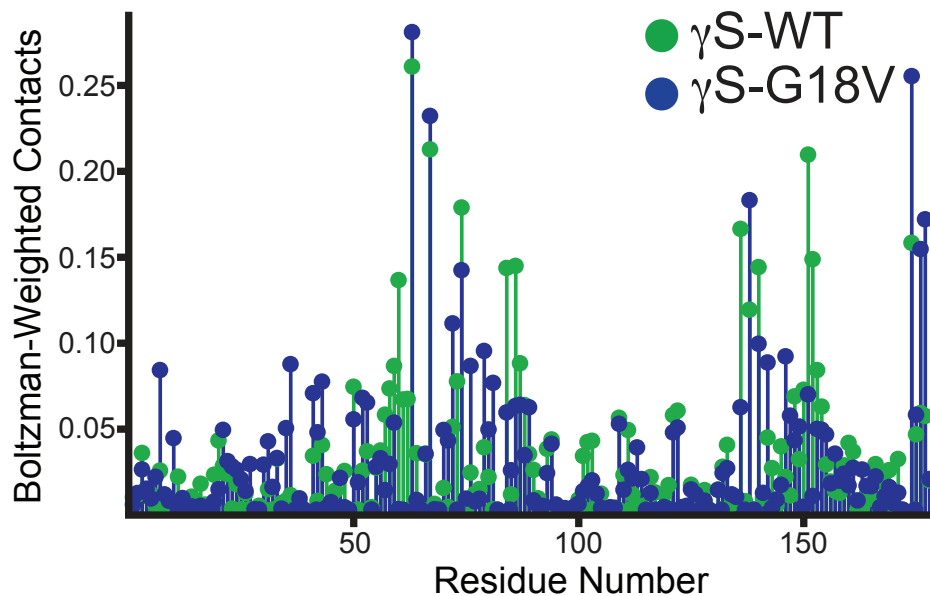
Figure 3.4: ANS-residue contact frequencies for γS-WT (green) and γS-G18V (blue) from docking simulations. Although non-specific binding is observed for both proteins, the contact frequencies show more ANS binding for γS-G18V than γS-WT, with maximum binding localized near the interdomain interface.

with a given residue. The Boltzmann weight of a given docked pose was calculated according to

$$w_i = \frac{exp(-E_i/k_B T)}{\sum_i exp(-E_i/k_B T)} \tag{3.2}$$

where i is the index of the docked pose, E_i is the pose binding energy, k_B is the Boltzmann constant, and T is the absolute temperature. The residue contact frequencies for each protein, calculated by Eric K. Wong [37], are shown in Figure 3.4. Following the CSP analysis, to determine which residues had strong- or weak-binding with ANS, a strong-binding threshold was set at two times the RMS of the calculated ANS-residue contact frequency, while the weak-binding threshold was set at half the RMS value. These values appear along with the CSP thresholds in Table 3.1.

3.2.10 Flexible refinement of binding sites

A flexible docking refinement, run by Eric K. Wong, was performed near all the highly perturbed residues identified using the strong-binding cutoff on the CSP data [37]. Docking search spaces were defined by clustered conformations of ANS from the screened docked set used to calculate the ANS-residue contact frequencies. Clustered poses were grouped into potential binding sites near the experimentally perturbed residues as shown in Figure 3.3B. Residues with an experimental CSP above the weak-binding cutoff were considered as flexible. A total of five potential binding sites were used to dock ANS to either flexible γS-WT or γS-G18V. The resulting poses were clustered again, and the locations and interactions of each pose were compared visually.

3.3 Results and Discussion

3.3.1 ANS fluorescence indicates that the relative surface hydrophobicity of γS-G18V is higher than that of γS-WT

Dye-binding assays were performed on γS-WT and the aggregation prone variant, γS-G18V. The ANS fluorescence measurements for γS-WT and γS-G18V, shown in Figure 3.5, indicate more exposed hydrophobic surface in γS-G18V compared to its wild type counterpart. These data also allow determination of the lowest ANS concentration required to produce the maximum emission before saturation, which was 1.5 mM for γS-WT and 1 mM for γS-G18V. The lower concentration required to saturate γS-G18V is consistent with the observation that it binds ANS more readily than wild-type.
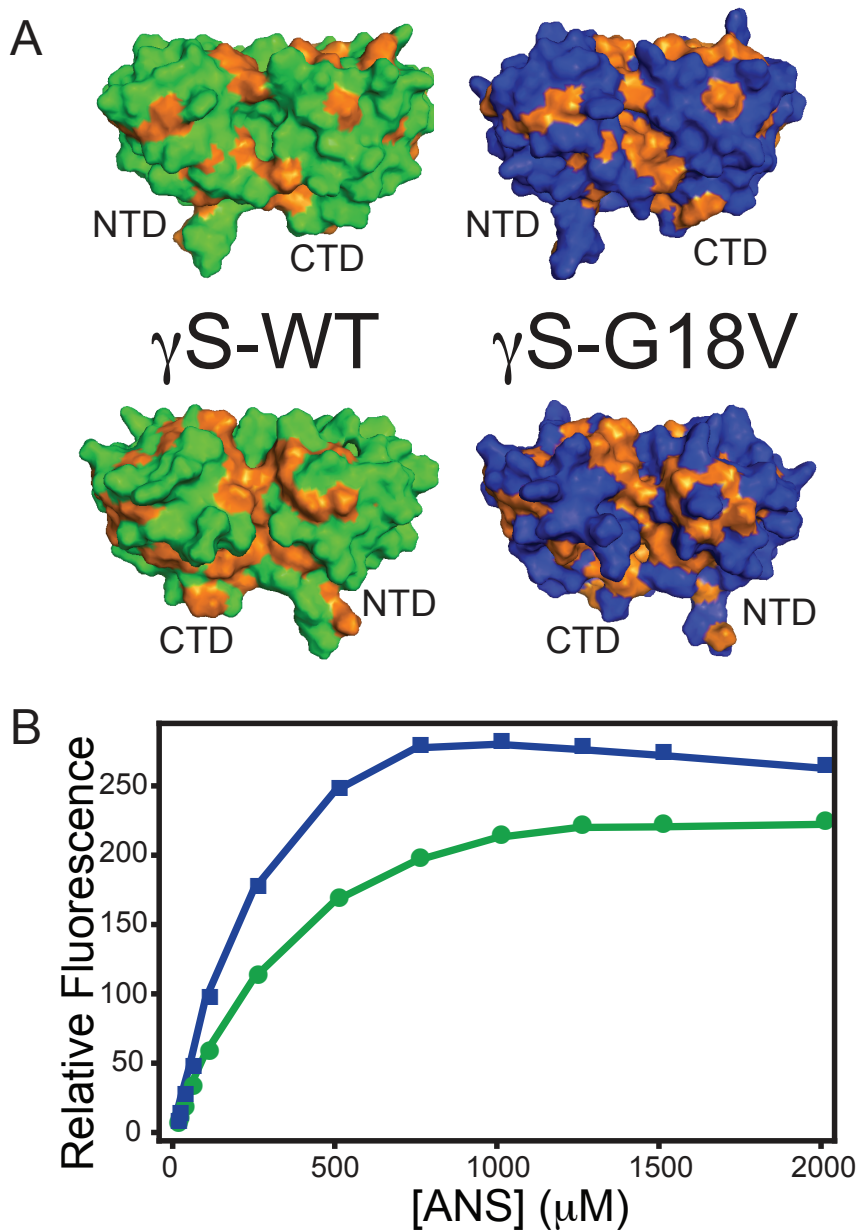
Figure 3.5: (A) Molecular surface representation of γS-WT (green) and γS-G18V (blue) based on the solution-state NMR structures (PDB ID 2M3T and 2M3U, respectively). Hydrophobic residues are highlighted in orange. (B) Fluorescence spectra representing ANS binding monitored at 500 nm using γS-WT and γS-G18V crystallins. Protein concentrations for both γS-WT and γS-G18V were approximately 1 mg/mL. Saturation occurred at 1.5 mM ANS for γS-WT and 1 mM ANS for γS-G18V. Higher emission was observed for γS-G18V, indicating more hydrophobic surface area exposed than for γS-WT.

3.3.2 Chemical shift perturbation mapping reveals the residues involved in ANS binding and the relative strengths of the interactions

Binding interactions between ANS and γS-WT or γS-G18V were measured at concentration ratios of 1:0, 1:0.5, 1:1, and 1:2 of γS:ANS, using CSP mapping via ^{1}H-^{15}N HSQC spectra. Selected regions of the NMR spectra where resonances show perturbations indicative of ANS binding are shown in Figure 3.6. The change in chemical shift for each peak in the 2D spectrum upon ANS binding was calculated using Equation 3.1. Because molecular docking simulations were performed using 1:1 ANS binding conditions, the focus for this chapter will be for 1:1 ANS binding conditions. While nonspecific binding is observed throughout the surfaces of both proteins, γS-G18V binds ANS more strongly in the N-terminal domain (approximately the first 100 residues). As seen in the 1:1 ANS binding shown in Figure 3.7,the maximum ANS binding occurs within residues 15 through 50, which is close to the glycine-to-valine mutation site. These observations are mapped onto the protein structures in Figure 3.8 (left panel) where the residues exhibiting strong (CSP at least two times the RMS) and weak (CSP at least half the RMS) ANS binding are highlighted. For γS-WT, strong-binding residues are highlighted in bright green and weak-binding residues in pale green. For γS-G18V, strong-binding residues are highlighted in dark blue and weak-binding residues in pale blue. Although some strong-binding residues are observed in both proteins near the mutation site, (e.g. G18 in γS-WT and D22 in γS-G18V), G18V displays more ANS binding, both strong and weak, in the N-terminal domain. Strong binding is also seen in the interdomain interface of γS-WT (residues L62, S82, and H123) and γS-G18V (residues L62, W73, H87, L88, and G91). These results are consistent with the observation that αB-crystallin strongly binds near the N-terminmal domain and the interdomain interface in γS-G18V, but not γS-WT [3]. Thus, the ANS-binding data support the hypothesis that the chaperone may be recognizing an exposed hydrophobic patch in this region of γS-G18V.
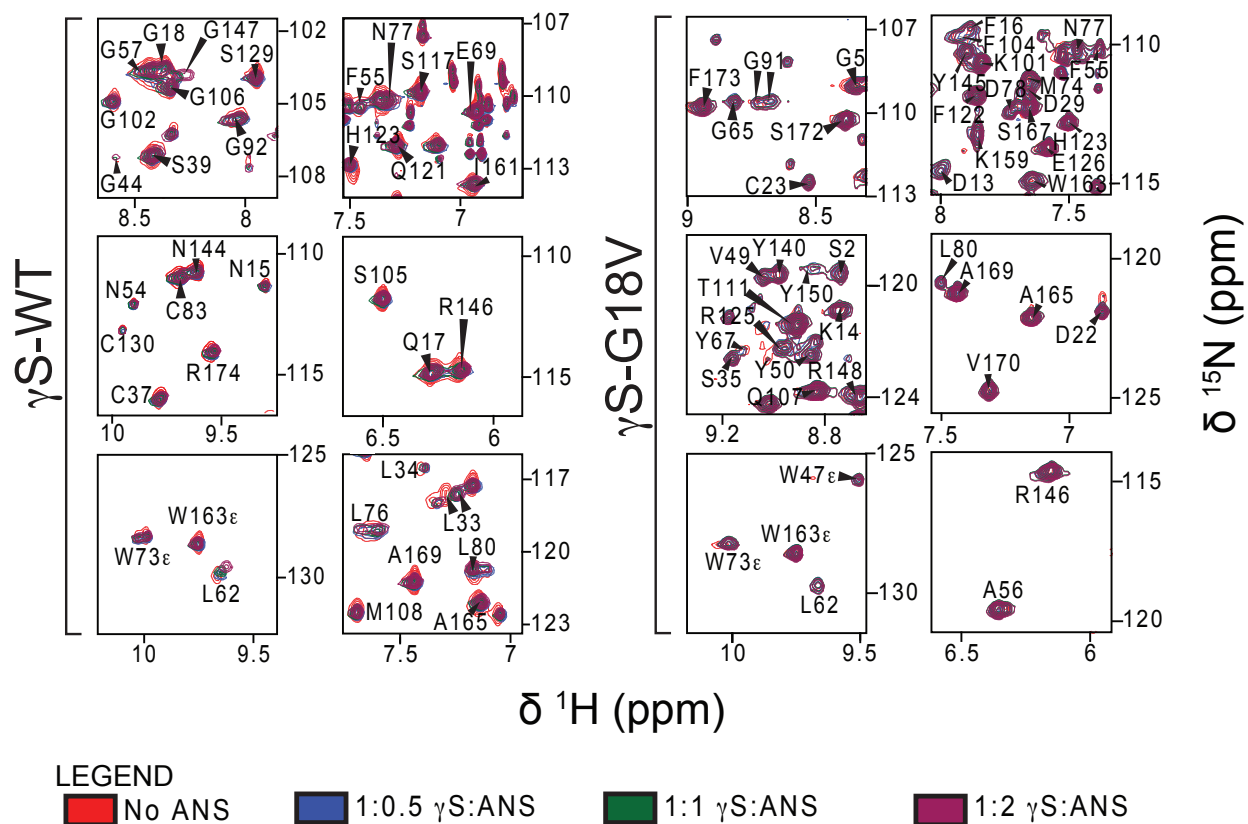
Figure 3.6: Selected portions of the overlaid ^{1}H-^{15}N HSQC spectra of γS-WT and γS-G18V. Experiments were carried out using ratios of 1:0, 1:0.5, 1:1, and 1:2 of γS:ANS. Spectra were acquired at 25 °C with concentrations of all γS-WT and γS-G18V samples at 0.3 mM. Resonances having a change in chemical shift indicate ANS binding to specific residues, which is quantified using the CSP equation (Equation 1).

Although we are focused on the 1:1 ANS binding conditions, the full CSP calculation was performed for the 1:0.5 and 1:2 conditions featured in Figure 3.9. Again, nonspecific binding is observed primarily within the N-terminal domain. Amongst residues 1 through 50 are residues experiencing maximum ANS binding, surrounding the mutation site. Again, this observation supports our previously stated hypothesis that the αB-crystallin strongly binds near the N-termimal domain and the interdomain interface in γS-G18V, but not γS-WT [3].
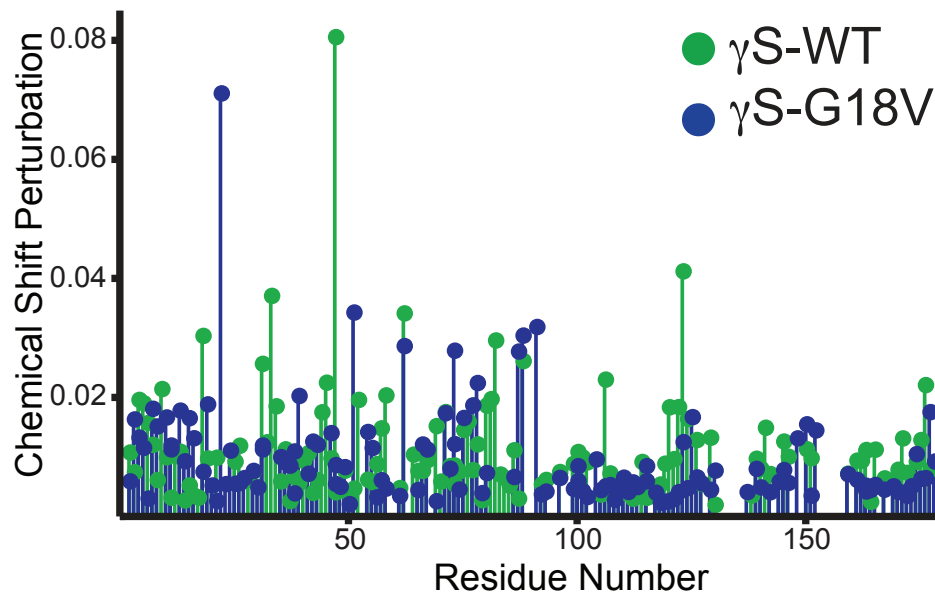
Figure 3.7: Average chemical shift perturbation (CSP) of γS-WT (green) and γS-G18V (blue). Nonspecific binding, with maximum perturbation in the N-terminal domain, is observed in both proteins. However, in γS-G18V more of the CSPs are localized to the N-terminal domain, particularly between residues 15 to 50, in the cysteine loop near the mutation site. Inspection of the structures confirms that this region is exposed to solvent in γS-G18V but not γS-WT.

3.3.3 Presence of ANS Does Not Affect Aggregation State of γS-crystallin

In order to characterize the aggregation states of γS-WT and γS-G18V, DLS data were acquired for both proteins under the same solution conditions used in the NMR experiments (shown in Figures 3.1 and 3.2). As observed in previous studies of γS-G18V[3, 4], Figure 3.10 shows the γS-WT solution contains only monomers, while γS-G18V shows a slightly broader range of sizes consistent with transient formation of dimers and potentially other small oligomers. However, the NMR spectra rule out the presence of a significant stable population of large aggregates; the linewidths for representative peaks in the HSQC spectra are comparable to the corresponding peaks in γS-WT. Linewidth comparisons for representative peaks in the spectra of γS-WT and γS-G18V are tabulated in Table 3.2. Despite the presence of transient oligomers in the γS-G18V sample, consistent with its increased aggregation propensity, it is clear from the linewidth data that the chemical shift changes
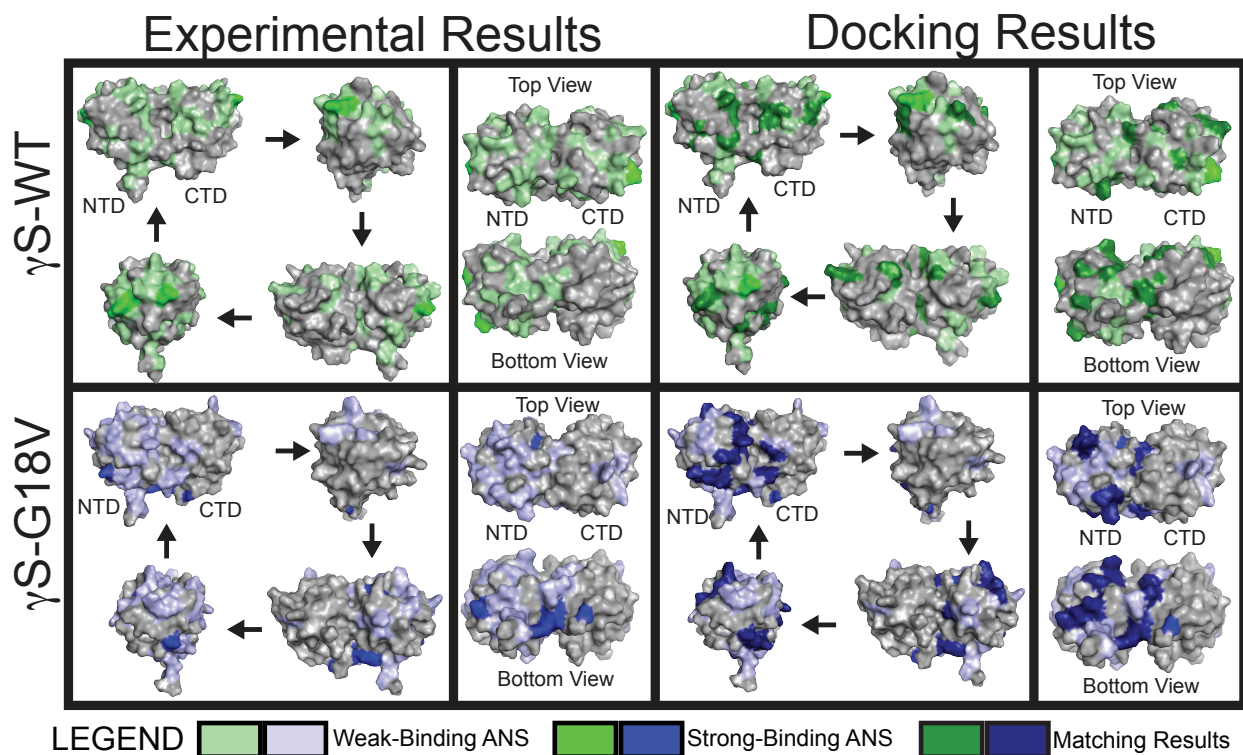
Figure 3.8: ANS interactions with γS-WT and γS-G18V. The strong-binding threshold and weak-binding threshold were defined as two times the RMS and half the RMS, respectively. Experimental CSP values indicate that ANS binding occurs throughout the N- and C-terminal domains for γS-WT, (strong binding residues in green and weak binding residues in pale green), while in γS-G18V ANS binding mainly occurs at the N-terminal domain (strong binding residues in blue and weak binding residues in pale blue). Some strong binding is observed in the N-terminal domain for both proteins near the mutation site, e.g. G18 in γS-WT and D22 in γS-G18V. However, G18V displays more ANS binding (both strong and weak) overall in the N-terminal domain. Strong binding is also observed in the interdomain interface of γS-WT, residues L62, S82, and H123, and γS-G18V, residues L62, W73, H87, L88, and G91. G18V exhibits more binding (strong and weak) within that interdomain interface suggesting that this variant has higher surface hydrophobicity localized to the N-terminal domain near the mutation site and the interdomain interface. Coverage of both strong- and weak-binding residues are nearly identical between experimental and docking results, highlighted in dark green for γS-WT and dark blue for γS-G18V, indicating that the docking results are in good agreement with the experimental data.
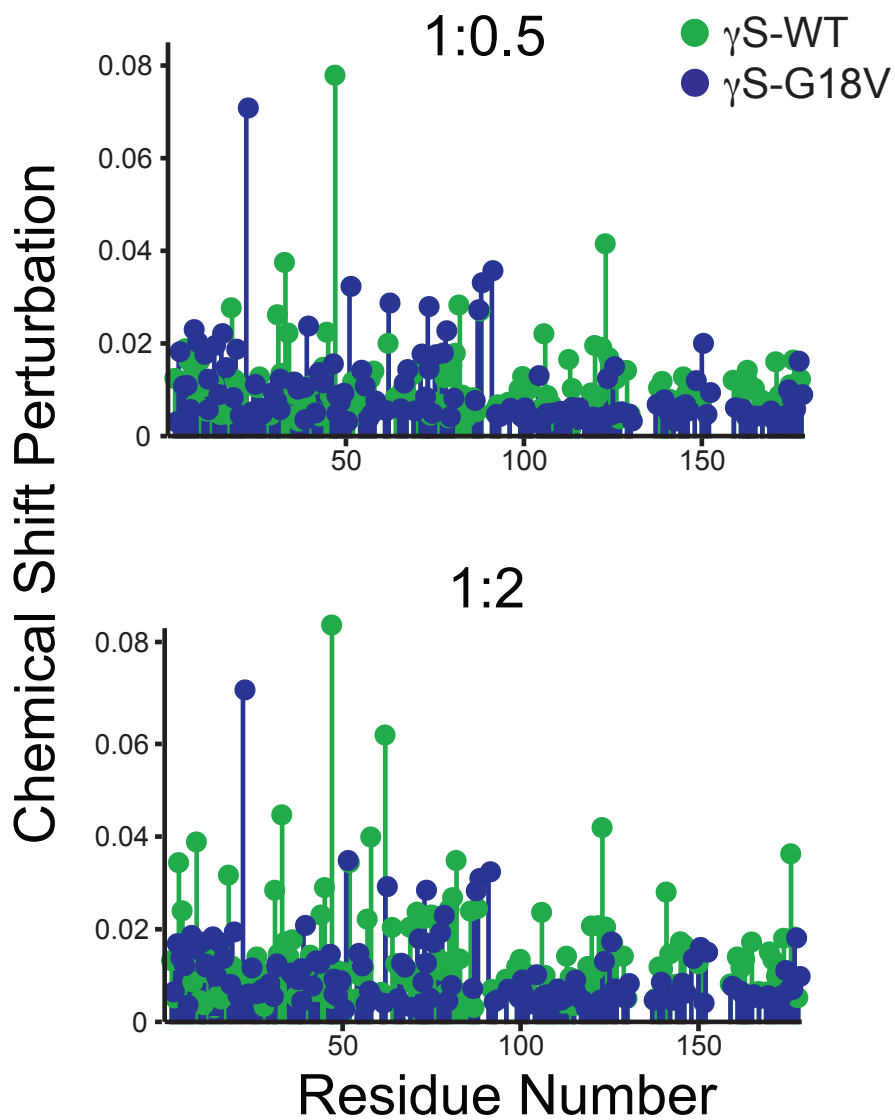
Figure 3.9: Average CSP of γS-WT (green) and γS-G18V (blue) for 1:0.5 and 1:2 γS:ANS. Maximum perturbation for both proteins is in the N-terminal domain and the interdomain interface. At the 1:2 γS:ANS ratio, γS-WT exhibits more perturbation, likely because γS-G18V is already saturated at 1 mM ANS.

Table 3.2: A table of selected line widths taken at half height for several representative residues in the HSQC spectra of γS-WT and γS-G18V for the 1:1 γS:ANS mixtures. The line widths (reported in Hz) are comparable for both proteins.

Residue	γS-WT		γS-G18V	
	^{1}H (Hz)	^{15}N (Hz)	^{1}H (Hz)	^{15}N(Hz)
C37	33.782	36.110	32.238	31.454
L62	36.003	44.881	37.979	35.617
W73ϵ	32.349	48.174	30.599	37.302
G102	32.392	33.081	32.082	37.695
F122	31.787	34.271	31.279	34.540
A165	29.652	30.582	31.966	28.605

upon addition of ANS are due to dye binding rather than a change in aggregation state. Although the oligomerization states of the starting solutions were not identical, this is accounted for by the chemical shift differences between γS-WT and γS-G18V in the absence of ANS, while the chemical shift perturbations reflect binding of each protein to ANS. If stable, large complexes were present in the NMR samples, the increased aggregation would be expected to cause significant line-broadening and disappearance of signals, as was observed for mixtures of γS-G18V with αB-crystallin [3], where large complexes were formed and TROSY techniques were required to observe the NMR signals. Although it is possible to prepare purely monomeric samples of γS-G18V at low pH, for the current study, neutral pH was chosen in order to investigate intermolecular interactions under more physiologically realistic conditions.

In order to investigate whether ANS changes the oligomerization states of γS-αB complexes and interferes with binding of αB-crystallin to γS-G18V, gel filtration chromatography was performed. Samples of γS-WT and γS-G18V were prepared at 1 mg/mL and compared to equivalent samples in the presence of αB-crystallin (1:1) and both αB-crystallin and ANS (1:1:1), shown in Figure 3.11. For γS-WT alone, the sample is mostly monomeric with a small amount of dimers. Upon addition of αB-crystallin, a population of larger oligomers at about 160 kDa appears, at the expense of the populations of both the monomeric and
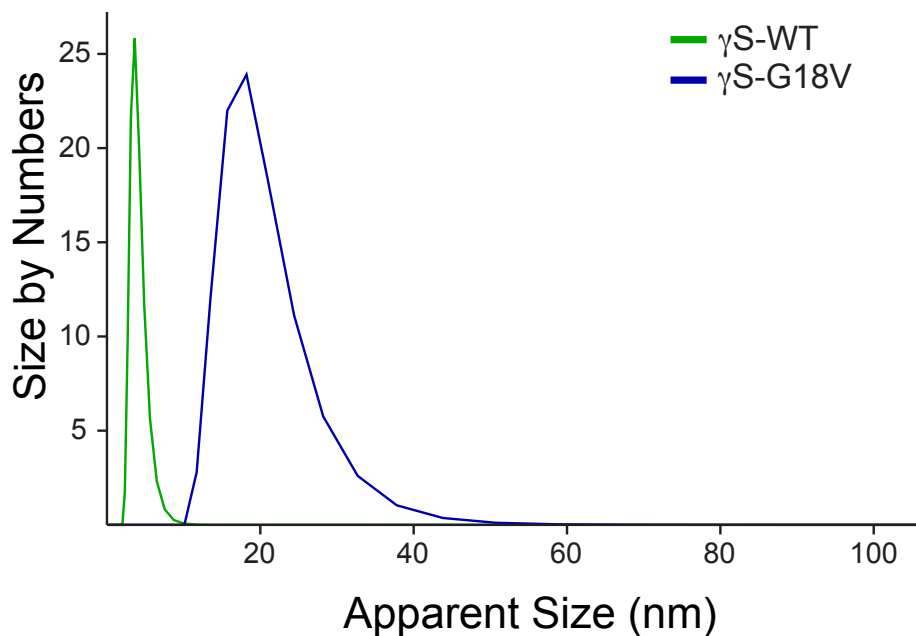
Figure 3.10: DLS data for γS-WT and γS-G18V in 10 mM phosphate at pH 6.9 displayed as the distribution of particle size by number. The concentrations of both proteins were at 0.3 mM, as in the NMR experiments. γS-WT is fully monomeric under these conditions, however γS-G18V forms a mixture of transient dimers and small oligomers, as observed in previous studies.

dimeric states. Addition of ANS to this mixture slightly increases the proportion of large aggregates. In the case of γS-G18V, both the initial oligomerization states and the effect of adding ANS is different. Initially, although much of the sample is monomeric, small populations of dimers and large oligomers exist. The peak at 160 kDa is much broader than in γS-WT, suggesting greater polydispersity. In the presence of αB-crystallin, the main effect is a dramatic narrowing of the peak corresponding to large oligomers, indicating a more uniform population. Addition of ANS to this mixture produces both further narrowing and an increase in the population of monomers, suggesting that interaction with ANS does disrupt the αB-γS complex formation. The full chromatogram including molecular weight standards is also provided in Figure 3.12.
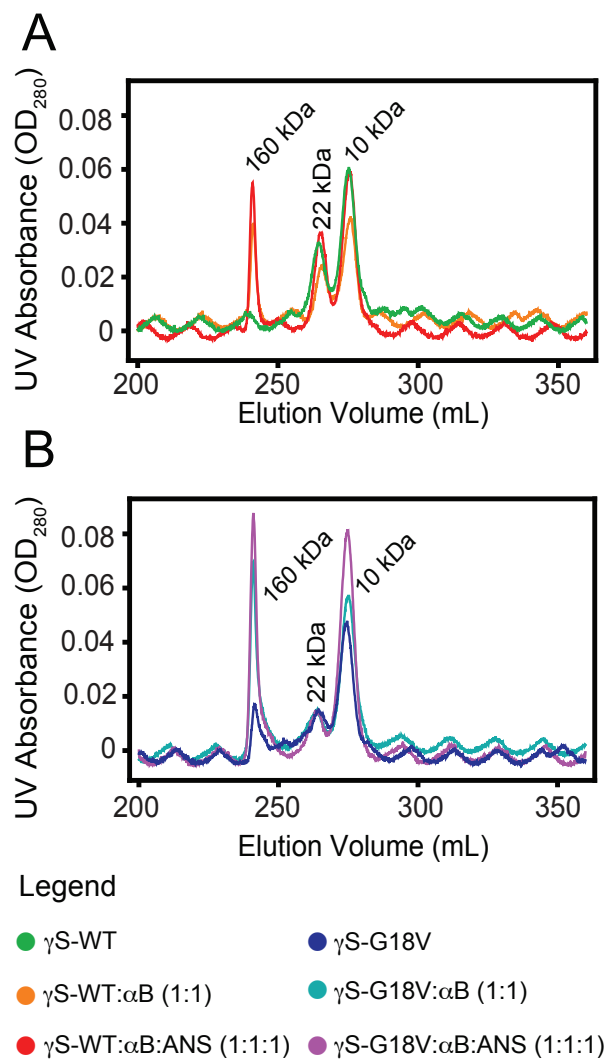
Figure 3.11: Gel filtration chromatograms for γS-WT (A) and γS-G18V (B) in the presence and absence of αB-crystallin and ANS. (A) For γS-WT alone (green), the sample is mostly monomeric (10 kDa) with a small amount of dimers (22 kDa). Upon addition of αB-crystallin (orange), a population of larger oligomers (160 kDa) appears, at the expense of the populations of both the monomeric and dimeric states. Addition of ANS to this mixture (red) slightly increases the proportion of large aggregates. (B) For γS-G18V alone (blue), much of the sample is monomeric, although small populations of dimers and large oligomers exist. The peak at 160 kDa is much broader than in γS-WT, suggesting greater polydispersity. In the presence of αB-crystallin (cyan), the main effect is a dramatic narrowing of the peak corresponding to large oligomers, indicating a more uniform population. Addition of ANS to this mixture (purple) produces both further narrowing and an increase in the population of monomers, suggesting that interaction with ANS does disrupt the αB-γS complex formation.
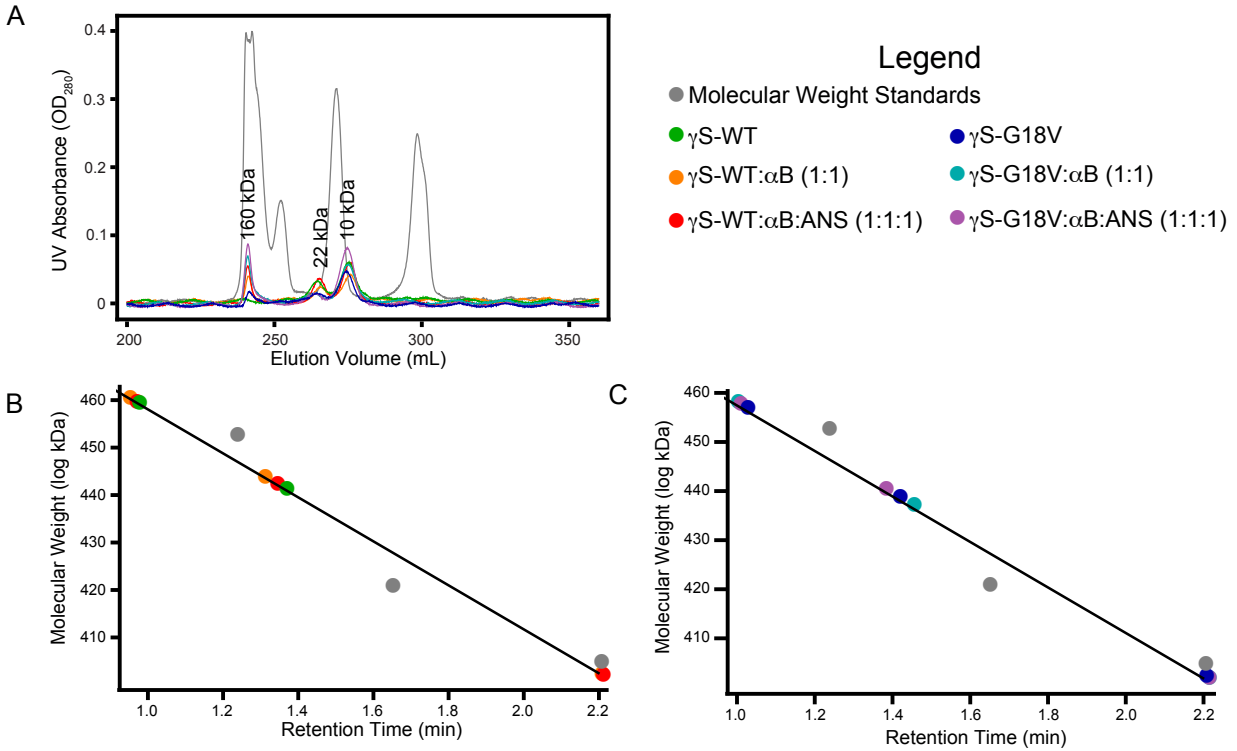
Figure 3.12: Gel filtration chromatogram for γS-WT and γS-G18V at 1 mg/mL with and without the presence of αB-crystallin and ANS. The samples were run in 10 mM sodium phosphate at pH 6.9. (A) Chromatogram representing γS-WT (green), γS-WT in the presence of αB-crystallin at a ratio of 1:1 (orange), γS-WT in the presence of αB-crystallin and ANS at a ratio of 1:1:1 (red), γS-G18V (blue), γS-G18V in the presence of αB-crystallin at a ratio of 1:1 (cyan), and γS-G18V in the presence of αB-crystallin and ANS at a ratio of 1:1:1 (purple). Molecular weight standards were also run and shown in light gray. (B) Graph showing the calibration of molecular weight standard as well as the position and relative size of the peak for γS-WT. For γS-WT (green), calculated sizes were 9.3 kDa and 23 kDa. For γS-WT in the presence of αB-crystallin at a ratio of 1:1 (orange), calculated sizes were 8.8 kDa, 20.1 kDa, and 158.4 kDa. For γS-WT in the presence of αB-crystallin and ANS at a ratio of 1:1:1 (red), calculated sizes were 9.1 kDa, 21.7 kDa, and 159.7 kDa. (C) Graph showing the calibration of molecular weight standard as well as the position and relative size of the peak for γS-G18V. For γS-G18V (blue) calculated sizes were 10.4 kDa, 25.8 kDa, and 158.4 kDa. For γS-G18V in the presence of αB-crystallin at a ratio of 1:1 (cyan), calculated sizes were 9.9 kDa, 28.0 kDa, and 161.0 kDa. For γS-G18V in the presence of αB-crystallin and ANS at a ratio of 1:1:1 (purple), calculated sizes were 10.0 kDa, 23.8 kDa, and 161.0 kDa.

3.3.4 Docking of ANS on the protein surface predicts more binding sites on γS-G18V than γS-WT and allows interpretation of the CSP data

Rigid receptor docking resulted in a total of 4860 docked poses (27 search spaces × 20 NMR conformations × 9 poses/search space). After screening for poses consistent with ANS fluorescence enhancement upon binding, 3423 poses and 3367 poses remained for γS-WT and γS-G18V, respectively as shown in Figure 3.3A. Filtered poses covered nearly the entire surface of the protein and exhibit a broad range of scores (from -2 kcal/mol to -7 kcal/mol, with a mean of -4.5 kcal/mol). Due to the pocket-like shape of the interdomain interface, ANS preferentially bound to the large hydrophobic pocket between the N- and C-terminal domains. However, sites were identified near all highly perturbed residues with comparable binding scores, as shown in Figure 3.4. Flexible docking poses located near the highly perturbed residues according to the CSP data had binding scores between -4.5 kcal/mol and -6.0 kcal/mol, consistent with a stronger preference for ANS to bind near the perturbed residues. A total of ten binding sites were found for γS-G18V and nine binding sites for γS-WT using flexible docking. Most of these binding sites were very similar in both γS-WT and γS-G18V. However, three binding modes were unique to γS-G18V. The first and most populated binding mode is located in the hydrophobic cavity at the interface between the N- and C-terminal domains, shown in Figure 6A and 6D [77, 103]. Although this binding site was found in both γS-WT and γS-G18V, the presence of the R84-D153 salt-bridge blocks the exposure of the hydrophobic surface in γS-WT. In contrast, γS-G18V lacks this salt-bridge interaction, which exposes the interdomain hydrophobic cavity and allows the entry of ANS into the interdomain binding site. This finding is consistent with the experimental NMR data, which indicate that chemically perturbed residues, H87 and L88, located near the interdomain pose, interact strongly with ANS only in γS-G18V (Figure 3.13). The second and third binding sites are located close to residues 20 through 30, which includes

a loop region containing three cysteine residues (C23, C25, and C27). As a result of the G18V mutation, C23 and C27 become solvent exposed, suggesting possible formation of intermolecular disulfide bridges, consistent with the observation that an excess of reducing agents abrogates the formation of small oligomers [104]. Previous studies suggested that the exposure of these cysteines results from a disruption in secondary structure due to the burial of V18 side chain [3]. As a result of this cysteine exposure and concomitant structural changes, a new hydrophobic pocket is uncovered as the second ANS binding site. Although ANS binds this cysteine loop in γS-WT after flexible docking refinement, it is not in direct contact with any hydrophobic surface, suggesting that the pose may not be consistent with enhancement in ANS fluorescence (Figure 3.13B). In contrast, when the hydrophobic pocket is exposed, as it is in γS-G18V, ANS becomes buried deep within the pocket (Figure 3.13E). This conformation provides both the hydrophobic interactions necessary for fluorescence as well as reduced quenching due to water exposure [82]. In addition to cysteine exposure, the third binding site shows additional hydrophobic surface exposure due to the cysteine loop separating from the main Greek key motif. This binding site was not found in γS-WT using the same docking search space, indicating that this hydrophobic patch is a unique characteristic of γS-G18V (Figure 3.13F). Additionally, the CSP data shows local perturbation of the backbone amides of the residues involved in these three binding sites only for γS-G18V. The presence of these γS-G18V-specific binding sites can explain the higher ANS fluorescence intensity of the variant protein over WT, and they also identify exposed hydrophobic patches which may potentially serve as protein-protein interfaces in crystallin aggregates, and which can be targeted in future mutagenesis studies.

The CSP data and the ANS-residue contact data from the docking simulations show generally good agreement in that the same protein regions were observed to bind ANS, shown in Figure 3.8. In some cases, the specific residues classified as strong binding vary between experimental and docking results, but coverage of both strong- and weak-binding residues
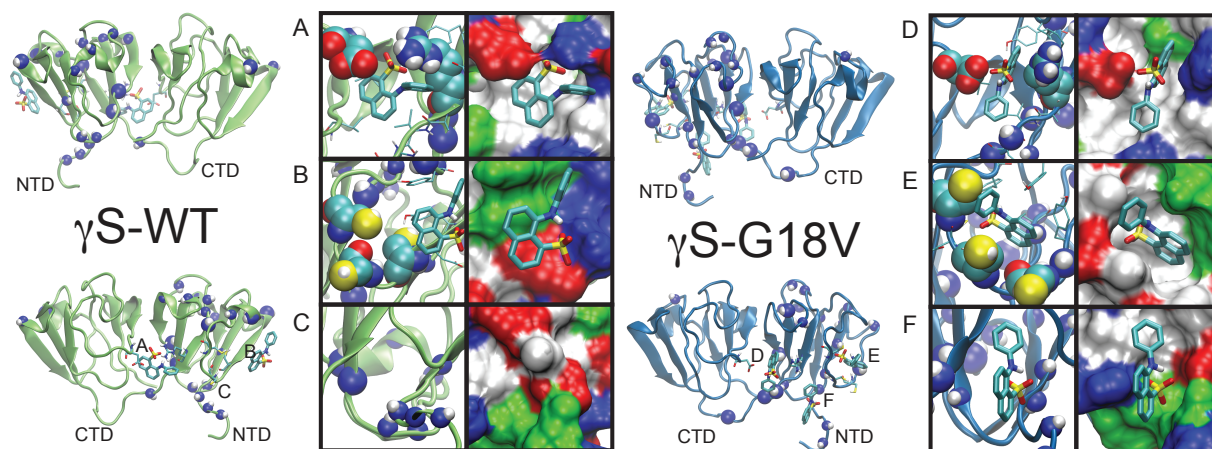
Figure 3.13: Docking poses of ANS bound to a flexible γS-WT and γS-G18V receptor. Three binding sites were found to be unique to γS-G18V. The protein surface was generated with MSMS[5]; red, blue, green, and white correspond to negative, positive, polar and hydrophobic regions, respectively. ANS is shown in licorice representation. The R84-D153 salt bridge and cysteine residues (C23, C25, and C27) critical to the hydrophobic patch availability are shown in space-filling representation. In the left-hand panels, residues defined as strong/weak binding by CSP data have their backbone amides represented as spheres. Large spheres represent strongly-binding residues, and small spheres represent weakly-binding residues. Atoms are colored by element (carbon, cyan; nitrogen, blue; oxygen, red; hydrogen, white; sulfur, yellow). (A & D) At site 1, the R84-D153 salt bridge separates to expose the hydrophobic cavity at the interdomain interface. Although a pose is generated in both proteins, the lack of perturbed residues at the binding site, according to the CSP data, indicates that the binding site is inaccessible to ANS in γS-WT. (B & E) At site 2, the docked pose of ANS shifts from a polar surface in γS-WT to inserting into a hydrophobic cavity in γS-G18V. Due to specific backbone torsions propagating from the G18V mutation site that keep the V18 buried, C23 and C27 become solvent exposed and reveal a hydrophobic cavity. This pose is located near the largest perturbed residue in γS-G18V according to the CSP data. (C & F) At site 3, the disordered cysteine loop separates from the Greek key motif of the N-terminal domain resulting in exposure of an additional hydrophobic patch. No equivalent pose could be generated on the γS-WT structure.

are nearly identical (highlighted in dark green for γS-WT, and dark blue γS-G18V in the right panel of Figure 3.8). This outcome is to be expected because the docking scoring function is more effective at identifying binding sites than distinguishing more subtle changes in binding energy: the standard error of the Autodock Vina scoring function [105] is larger than the variation among scored poses. The agreement between rigid protein docking results and experimental ANS binding results suggests that there is no major change in protein conformation upon binding of ANS, supporting the hypothesis that hydrophobic patches on the surface are involved in intermolecular interactions. Good agreement between the experimental and docking results further confirms that ANS binding is localized near the mutation site in the N-terminal domain for γS-G18V, consistent with the CSP data. Experimental and simulation results are also consistent on the binding of ANS to the exposed interdomain hydrophobic surface located in the interdomain interface between the two domains due to the breaking of the R84-D153 salt bridge in γS-G18V. Exposure of this hydrophobic patch facilitates ANS binding and may be involved in hydrophobic protein-protein interactions.

3.4   Conclusion

In this study, I have used ANS fluorescence and solution-state NMR chemical shift perturbation mapping along with molecular docking (performed by Eric K. Wong) to investigate the differences in exposed hydrophobic surface between human γS-crystallin and its cataract-related γS-G18V variant. The experimental results indicated that both proteins have a fairly high level of nonspecific binding, but both the fluorescence and NMR measurements indicate more ANS binding to γS-G18V, particularly in the N-terminal domain near the mutation site. The docking studies, in agreement with the NMR data, found three binding modes that are unique to γS-G18V that were not found in γS-WT: one in the exposed hydrophobic patch in the interdomain interface and two binding modes in the exposed hydrophobic

pocket formed when the cysteine loop becomes solvent exposed. Using docking and binding assays, hydrophobic surface patches were identified that may be responsible for some of the intermolecular interactions between crystallins that promote aggregation in the lens. In conjunction with DLS measurements and gel filtration chromatography, the addition of ANS did not disrupt the αB-γS complex formation. To further investigate αB-γS complexes in the presence of ANS, CSP mapping must be performed to understand whether or not ANS is displacing αB-crystallin at higher concentrations. The results of this chapter may guide the design of future mutagenesis and drug-binding studies to further investigate the importance of such intermolecular interactions in mediating γS-crystallin solubility and aggregation resistance in the healthy eye lens.

Chapter 4

Probing Protein-Peptide Interactions of Human γS-Crystallin

4.1 Background

The crystallins of the human lens belong to two superfamilies: the α-crystallins, which are molecular chaperones related to small heat-shock proteins, and the structural $\beta\gamma$-crystallins. The $\beta\gamma$ crystallins share a common Greek key fold where it is hypothesized that non-specific surface interactions are repelled, however, subtle structural differences between them could disprove that idea. The fold of most γ-crystallins is very compact, with the N- and C-termini hidden within the protein. In contrast, β-crystallins possess longer N- and C-terminal extensions as well as a distinctive linker sequence between the two domains [106]. γ-crystallins usually exist as monomers whereas β-crystallins often form dimers, tetramers, and higher-order complexes [107, 108, 109], in the form of both homo- and heterooligomers [106]. Previous studies of interactions among β-crystallins have suggested that they are often mediated by association of the N-terminal extensions, which are highly variable due to differences in

length and flexibility. Ajaz *et al.* have reported β-crystallin aggregates in the human eye lens containing βB1-crystallin with varying levels of truncations at the N-terminus, along with other β-crystallins (βA1, βA3, βA4, and βB2). Using mass spectrometry and chromatography, Ajaz *et al.* observed that βB1-crystallin with a variety of N-terminal truncations form aggregates *in vivo* [106]. The sizes of the aggregates were correlated with the length of the N-terminus, where the largest aggregates contained intact βB1 or very short truncations, while the smallest aggregates contained more extensive truncations [106]. These truncations, which involve deletion of anywhere between 15 to 45 residues from the N-terminus of βB1 *in vivo*, appear to play a role in β-crystallin heterooligomerization, as tested with βB1 and βB1ΔN41 [110]. Further understanding of the N-terminal extension was enhanced by experimentation on various truncated forms of βB1, using truncations of up to 50 residues [111]. By monitoring kinetics of isolated βB1, Leng *et al.* observed that the N-terminal extension may act as an intramolecular chaperone via weak binding of the N-terminal extension to the main body of βB1 or to exposed hydrophobic patches on the interior of βB1, either of which may lead to stabilization of the native state or avoidance of the misfolding/aggregation pathway [111]. Although the mechanisms of the N-terminal extension loop for β-crystallins is not fully understood, it is clear that intermolecular interactions with the N-terminal extension has a large impact on the aggregation propensity. Leng *et al.* proposed an intramolecular chaperone mechanism from refolding experiments, suggesting that the N-terminal extension of βB1 is important for the co-refolding of β-crystallin heteromers and the chaperone function of αA-crystallin. Unlike most γ-crystallins, human γS-crystallin contains an N-terminal extension similar to that of βB1 [109]. Although wild-type γ-crystallins are not known to form dimers or have proteolytic sites on the N-terminal extension as β-crystallins do, interactions with the N-terminal extension in γ-crystallins could potentially lead to the formation of aggregates in the lens. However, Ray *et al.* proposed that the N-terminal extension loop of γS-crystallin could regulate dimer formation in aggregation-prone variants based on significant chemical shift differences between γS-WT and its deamidated variant, N76D [27].

In this chapter, I test this idea by investigating possible intermolecular interactions among human γS-crystallin and its N-terminal extension loop towards aggregation propensity.

Crystallin protein self-aggregation leads to cataract formation, in many cases mediated by subtle structural changes. Experiments on deamidated variants in γS-crystallin presented by Pande *et al.* showed that although the secondary and tertiary structures were similar to wild-type, the significant changes were observed in the weakened net repulsive interactions of the deamidated variants [40]. Crystal structures of two congenital-cataract related γD-crystallin variants, γD-W42R [112] and γD-P23T [113], as well as a solution NMR structure, γD-R76S [114], confirmed no major structural alterations. Potential aggregation sites caused by the mutations were not visible compared to the previous solution NMR structures of γD-WT [115]. Ji *et al.* proposed that aggregation behavior of proteins is best observed in the solution environment [112, 113] as proteins are dynamic in solution and behave as they would in a physiological environment.

Here we investigate the potential intermolecular interactions between γS-WT and γS-G18V molecules using a library of tripeptides derived from the human γS-crystallin sequence. CSP mapping using solution NMR identify some of the intermolecular interactions between crystallins that may promote aggregation in the lens. However, in agreement with molecular docking simulations, non-specific binding at the surface or both γS-WT and γS-G18V does indeed occur, disproving the hypothesis that the very low aggregation propensity of γS-WT is due to unusually good resistance to non-specific binding. We find that although both proteins are susceptible to non-specific binding, many of the tripeptides investigated bind more strongly to γS-G18V. In particular, two peptides derived from the N-terminal extension showed strong interactions with γS-G18V that are different from those available to γS-WT. The N-terminal extension is of particular interest because it is unstructured and exposed to solvent in both the γS-WT and γS-G18V structures, such that intermolecular interactions are possible in the context of the full protein. Binding of disordered regions

to structured domains is an important mechanism for mediating intra- and intermolecular interactions, leading to stable interactions in solution or inducing aggregation [116]. In light of the previous finding that the N-terminus of bovine γS-crystallin is acetylated [117], the N-terminus may act as a mediator of inter-crystallin interactions and is sometimes found in proteins involved in interactions with cytosolic chaperones. Intra- and intermolecular interactions in βB-crystallins are sometimes mediated by the N-terminal extensions, leading to the hypothesis that the short N-terminal extension of γS-crystallin is involved in mediating intermolecular interactions between itself and other crystallins.

The SPOT-synthesis method of producing peptide arrays on cellulose supports uses standard Fmoc chemistry based on solid phase peptide synthesis [118]. Despite varying coupling yields across peptide "spots," detection of peptide-protein interactions via fluorescence of the dye-labeled target protein is relatively simple and reliable [119, 120]. Peptide-binding assays using SPOT synthesis peptide arrays provide a semiquantitative approach to studying protein-protein interactions [118]. This method has been used to study binding interfaces between Src homology 2 (SH2) domains and protein kinases,[121] to monitor interactions among Src homology 2 domains and phosphorylated immunoreceptor tyrosine-based regulatory motifs (ITRMs) necessary for immune signaling,[122] to map the interactions between yeast Ire1 ER-lumenal domain (Ire1) to a constitutively misfolded mutant of carboxypeptidase Y (CPY*) [123], and to establish the αA-crystallin recognition site on the surface of αB-crystallin [26]. However, these assays neither provide a detailed view of the binding interfaces nor produce any structural data regarding possible conformational changes. In contrast, NMR chemical shift perturbation (CSP) mapping can provide a detailed picture of which residues in the protein are involved in ligand binding [87]. Short peptide sequences in previous studies have been shown to form aggregates, particularly sequences that are 4-10 residues long [124, 125]. These short peptide sequences have been shown to recapitulate interactions of full proteins [126]. SPOT synthesis peptide binding assays are typically performed with peptides composed of 2-5 amino acid residues; here tripeptides were chosen because tripeptides have

been proposed as the minimum functional unit of protein aggregation [127], and have been shown to display hydrogen-bonding functionality similar to proteins [128].

4.2 Materials investigating γS-crystallin interacting with the N-terminal extension loop

4.2.1 Design and SPOT-synthesis of tripeptides

Tripeptide libraries were synthesized using the methodology of Hilpert *et al.*[129] for screening with γS-WT and γS-G18V. The synthesis was performed by Carolyn N. Kingsley [130]. Libraries were made covering the sequence of γS-WT in three reading frames beginning with S2, K3, or T4. A fourth random tripeptide library was also designed to incorporate a broad range of charged and hydrophobic amino acids.

4.2.2 Binding assay

Wild-type and G18V γS-crystallins were expressed and purified as described previously [4]. Binding assays using non-specifically Oregon Green-labeled γS-WT and γS-G18V were carried out and binding of the protein to the peptide spots was visualized using a fluorescence laser scanner (λ_{ex}= 488 nm, λ_{em}= 520 nm). The peptide spot emission intensities were analyzed using ImageQuantTL. After correction for background emission, the change in emission intensity was calculated for each peptide spot. All four tripeptide libraries and their respective binding assays appear in Figures 4.1.
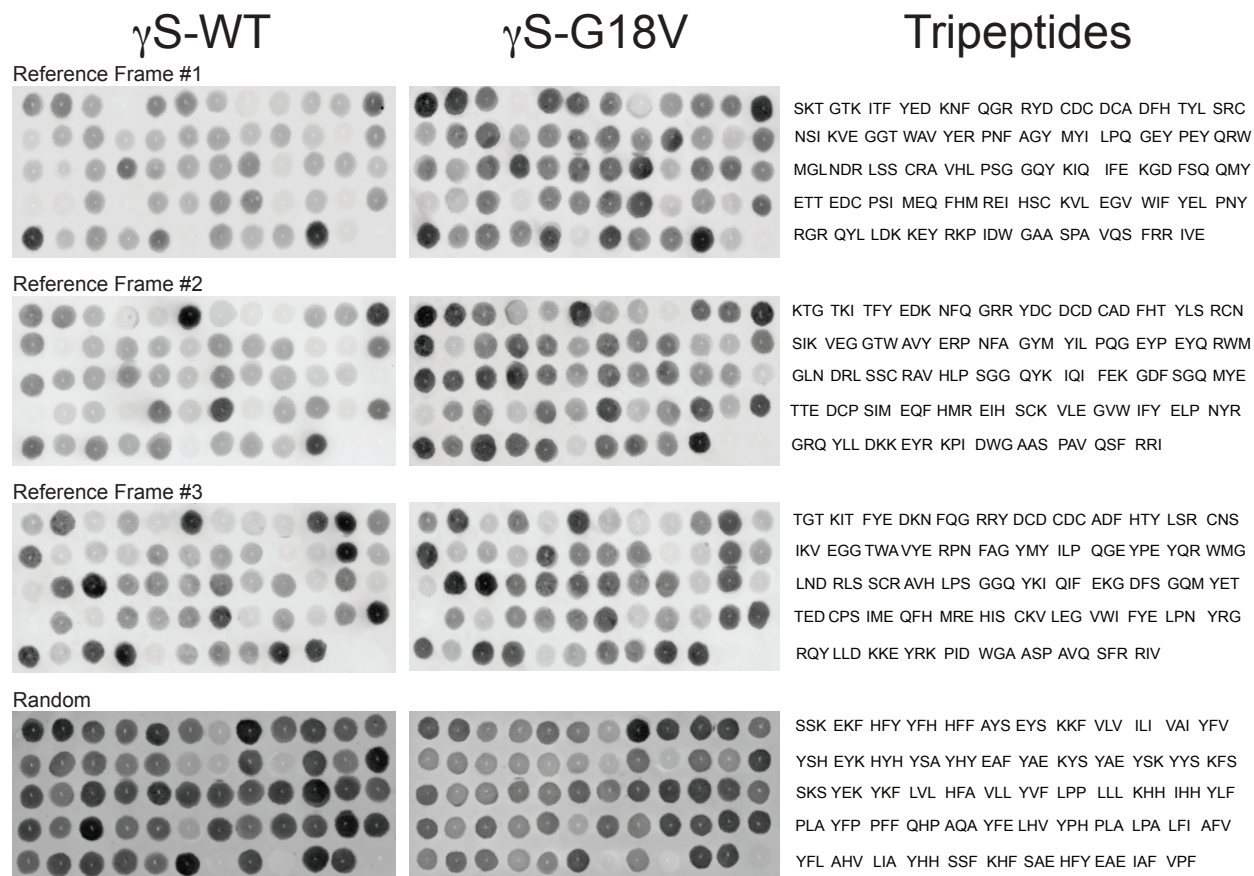
Figure 4.1: Tripeptide binding assay data set utilizing SPOT peptide libraries with Oregon Green-labeled γS-WT and γS-G18V Tripeptide libraries contain 3 reference frames taken from the γS-crystallin sequence and a random tripeptide library for control.

4.2.3 Mass spectrometry of tripeptides

The tripeptides SKT, KTG, and RLS were purchased from NeoBioLab (Woburn, MA). The tripeptides were dissolved in 10 mM sodium phosphate buffer and 0.05% sodium azide at pH 6.9 to a final concentration of 300 mM and characterized by mass spectrometry on an ESI LC-TOF Micromass LCT 3 to verify their molecular masses and purity before carrying out the CSP NMR experiments. Mass spectrometry data for tripeptides RLS, SKT, and KTG are shown in Figure 4.2.

4.2.4 NMR sample preparation

[15]N-labeled γS-WT and γS-G18V were expressed and purified as described previously [4]. Purified γS-crystallin was concentrated and supplemented with 2 mM TMSP, 10% D_2O, and 0.05% sodium azide. The final concentration of all γS-WT and γS-G18V samples was 1.5 mM. Tripeptides KTG, SKT, and RLS were titrated into the protein samples to give final molar ratios of 1:0, 1:1, 1:2, 1:5 and 1:10 of γS:tripeptide.

4.2.5 NMR experiments

Experiments were performed on a Varian [Unity]INOVA spectrometer (Agilent Technologies, Santa Clara, CA) operating at 800 MHz and equipped with a [1]H–[13]C–[15]N 5 mm tri-axis PFG triple-resonance probe, using an 18.8 Tesla superconducting electromagnet (Oxford instruments). Decoupling of ^{15}N nuclei was performed using the GARP sequence [98]. [1]H chemical shifts were referenced to TMSP, and ^{15}N shifts were referenced indirectly to TMSP. NMR data were processed using NMRPipe [99] and analyzed using CcpNMR Analysis [100]. Center operating frequencies and (unless otherwise stated) center frequency offsets were as follows:

Figure 4.2: Mass spectrometry of tripeptides RLS, SKT, and KTG using an ESI LC-TOF Micromass LCT 3. The major peaks of the RLS spectra are 335 and 357 m/Z while those of the SKT spectra are 375 and 397 m/Z . The actual molecular masses of RLS and SKT are 374.4 334.4 g·mol^{-1}. The molecular mass of KTG is 304.4 g·mol^{-1}. The mass spectrometry data shows two peaks at 305 and 327 m/Z. The second peak is +23 m/Z, the molecular weight of a sodium ion from the sodium phosphate buffer. Although minor impurities were present, they are not concentrated enough to interfere with the NMR assays.

Table 4.1: Table featuring weak and strong binding thresholds for 1:10 γS:tripeptide. The weak-binding threshold was set at half the RMS while the strong-binding threshold was set at two times the root mean square (RMS) of the calculated CSP to determine which residues had strong or weak binding with the individual tripeptides

	γS-WT		γS-G18V	
	Weak	Strong	Weak	Strong
RLS	0.00728192	0.0291277	0.01180599	0.04722398
SKT	0.00883259	0.03533037	0.01287653	0.05150614
KTG	0.0099113	0.0396452	0.01583817	0.06335269

Center	[1]H: 799.8056964 MHz	[13]C: 201.1282461 MHz	[15]N: 81.0504078 MHz
Offset	[1]H: -294.932 Hz (4.8 ppm)	[13]C: -9863.17 Hz (43 ppm)	[15]N: 2400 Hz (116.7 ppm)

4.2.6   Calculation of chemical shift perturbations

^{1}H-^{15}N HSQC spectra of γS-WT and γS-G18V were collected in the presence and absence of tripeptides RLS, SKT, and KTG at concentration ratios of 1:0, 1:1, 1:2, 1:5 and 1:10 of γS:tripeptide. Resonances were identified and assigned based on previously collected chemical shift data (PDB ID: 2M3T and 2M3U) [3]. The full NMR spectra is shown in Figure 4.3. The chemical shifts for each resonance in the 2D spectra upon tripeptide binding were recorded and the differences were calculated using the Equation 3.1. A strong-binding threshold for each set of conditions was set at two times the root mean square (RMS) of the calculated CSP, while the weak-binding threshold was set at half the RMS to determine which residues had strong or weak binding with the individual tripeptides. The CSP values for the 1:10 γS:tripeptide concentrations were used for analysis in flexible docking. The values used for 1:10 γS:tripeptide appear in the Table 4.1.

4.2.7   Binding site search by blind docking

Docking receptor coordinates were obtained from the NMR structures of γS-WT and γS-G18V (PDB ID: 2M3T and 2M3U) [3]. Autodock Tools [131] was used to prepare both

Figure 4.3: Chemical shift perturbation mapping of ^{15}N-labeled γS-WT and γS-G18V with natural-abundance peptides RLS, SKT, and KTG monitored by ^{1}H-^{15}N HSQCs. γS protein sample conditions consist of 10 mM phosphate buffer at pH 6.9 supplemented with 10% D_2O, 0.05% sodium azide and 2 mM TMSP for referencing. Protein concentrations for both γS-WT and γS-G18V were 1.5 mM. The γS:tripeptide ratios used were 1:0, 1:1, 1:2, 1:5, and 1:10.

receptor and tripeptide ligands by adding Gasteiger charges [132] and merging non-polar hydrogen atoms into united heavy atoms. Docking simulations for this study were performed by Eric K. Wong in the Tobias Lab at UCI [133].

4.2.8   Calculation of residue contacts

To compare the filtered docked set with the residue-based CSP data, tripeptide-residue contact frequencies were calculated by summing the Boltzmann weights of all the poses in contact with a given residue. The calculations for this study were performed by Eric K. Wong in the Tobias Lab at UCI. The Boltzmann weight of a given docked pose was calculated according Equation 3.2. A strong-binding threshold was set at two times the RMS of the calculated tripeptide-residue contact frequency, while the weak-binding threshold was set at the RMS value. A weighted contact frequency cutoff of 0.0457 and 0.0114 were used to identify strong- and weak-binding residues for comparison to CSP in Figure 4.4.

4.3   Results and discussion

4.3.1   Tripeptide libraries based on the γS-crystallin sequence show low affinity overall for γS-WT, slightly higher for γS-G18V

Of the 233 peptides screened, 81% (189 peptides) interacted more strongly with γS-G18V than with γS-WT. A schematic representation of the tripeptide-binding assay is shown in Figure 4.5. Although the NMR structures of the proteins have revealed that the G18V mutation in γS-crystallin leads to some minor structural changes, while leaving the overall fold intact, the binding assays presented here suggest that surface interactions are more strongly affected, consistent with previous data on binding of the hydrophobic dye ANS [37].
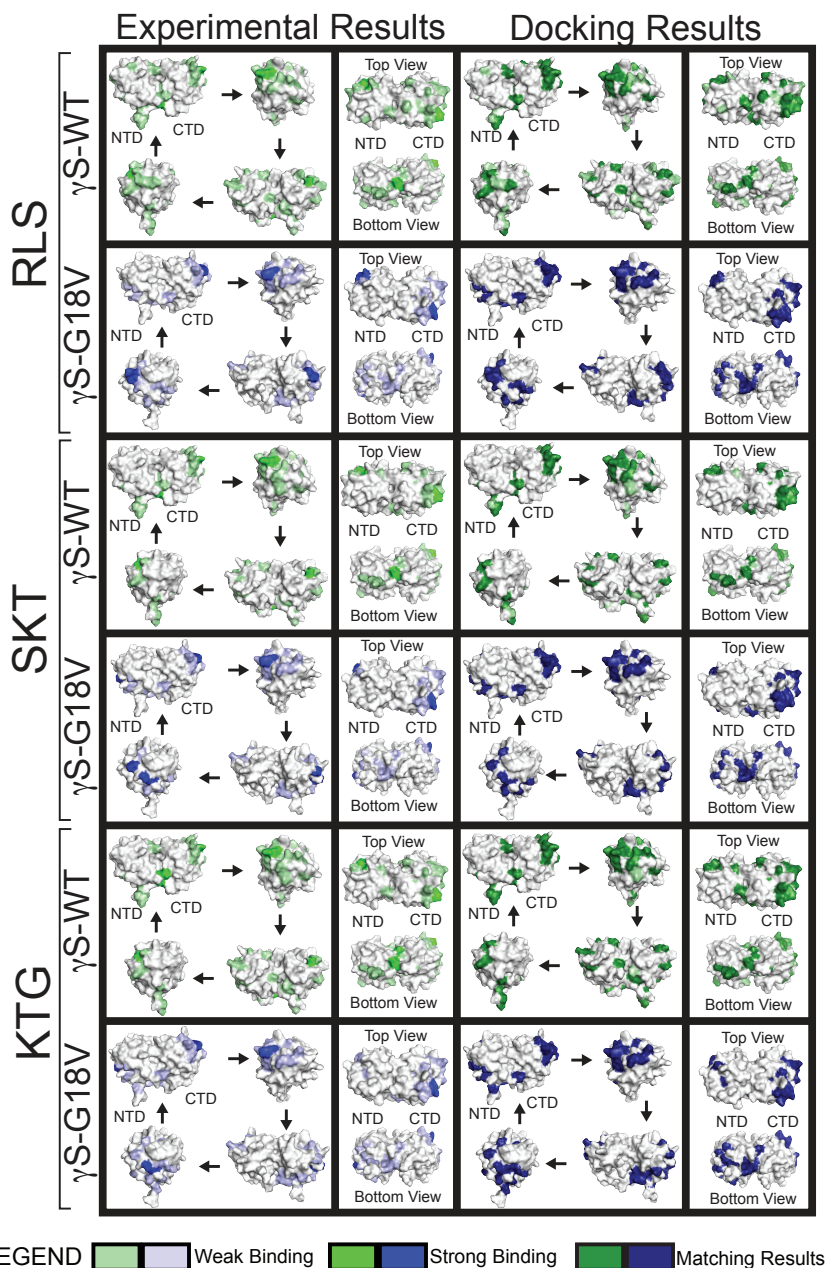
Figure 4.4: Molecular surface representation of tripeptide interactions with γS-WT and γS-G18V. A strong-binding threshold and weak-binding threshold were defined as two times the RMS and half the RMS, respectively, using experimental CSP values and rigid docking simulations. The right panel displays the experimental results based on the NMR titration studies, while the left panel displays the matching rigid docking results. The left panel shows that tripeptide binding occurs throughout both the N- and C-terminal domains for γS-WT (strong binding residues in green and weak binding residues in pale green) and in γS-G18V (strong binding residues in blue and weak binding residues in pale blue). Coverage of both strong and weak binding residues are nearly identical between experimental and docking results, highlighted in dark green for γS-WT and dark blue for γS-G18V, indicating that the docking results are in good agreement with the experimental data.

When all four libraries are plotted by their change in emission intensities between γS-WT and γS-G18V, shown in Figure 4.6, the first three libraries (purple) have more preferential binding towards γS-G18V while the random library (salmon) has binding to both proteins. Tripeptides having a particularly low binding affinity to γS-WT include MEQ, SAE, TED, EAE, and YED. Similarly, SAE, TED, EAE, QGE had low binding affinity for γS-G18V, potentially suggesting that negative charges may be involved in disrupting potentially harmful surface interactions. The tripeptides that had the highest degree of binding to γS-G18V are KTG, SKT, RLS and VLV. VLV proved to be insoluble in water and was therefore not investigated further. Interestingly, KTG and SKT both contain subsequences located near the N-terminus (GSKTGTKITF). Strong binding from subsequences in the N-terminus raises a possible mechanism for intermolecular interactions. To test the hypothesis that the short N-terminal extension of γS-crystallin is involved in mediating intermolecular interactions between itself and other crystallins, NMR chemical shift perturbation experiments were performed using the tripeptides SKT and KTG, both found in the N-terminus. The tripeptide RLS, located in a β-sheet in the C-terminal domain was used as a control.

4.3.2 Chemical shift perturbation mapping shows that many of the same residues are involved in interactions with all three tripeptides.

Chemical shift perturbation assays reveal that many of the same residues, in particular those located at the surface of the proteins, interact with tripeptides RLS, SKT, and KTG. Selected portions of the NMR data are shown in Figure 4.7. The chemical shift perturbation for each peak in the 2D spectrum upon binding each tripeptide was calculated using Equation 3.1. Both γS-WT and γS-G18V show non-specific binding, validated through CSP mapping in Figures 4.8 - 4.10. In the interest of comparing the CSP data to the molecular docking
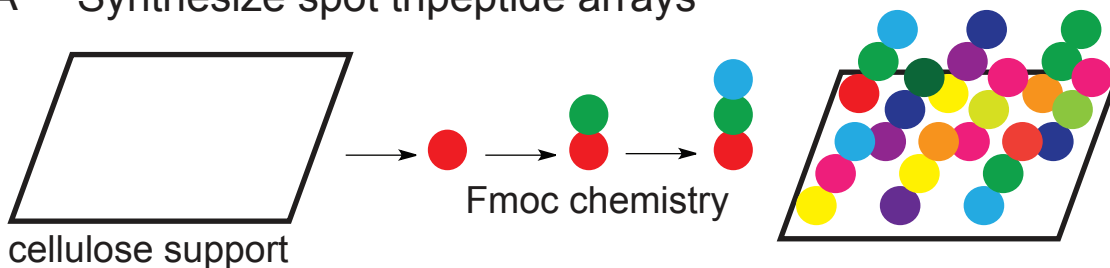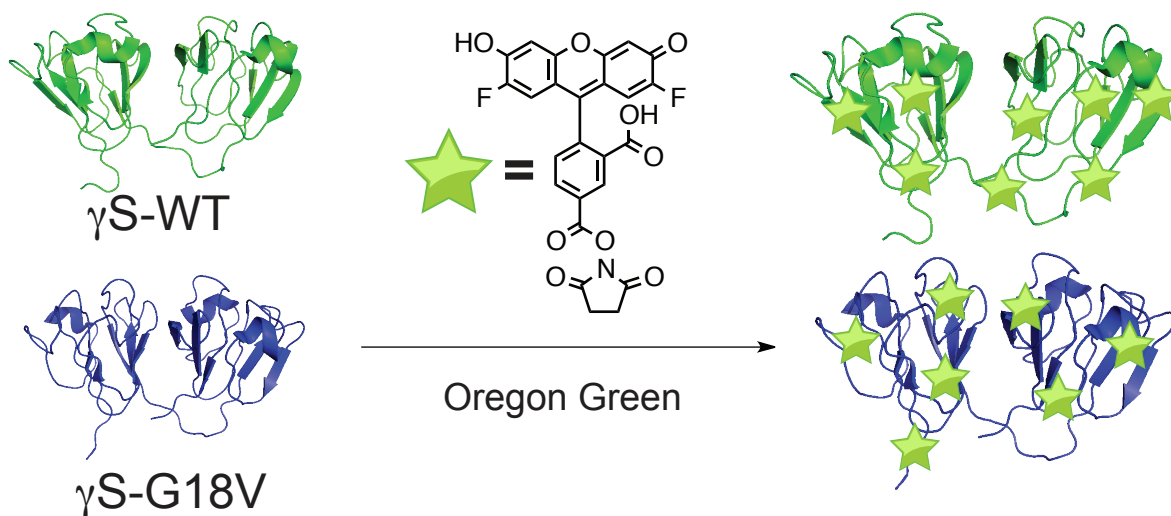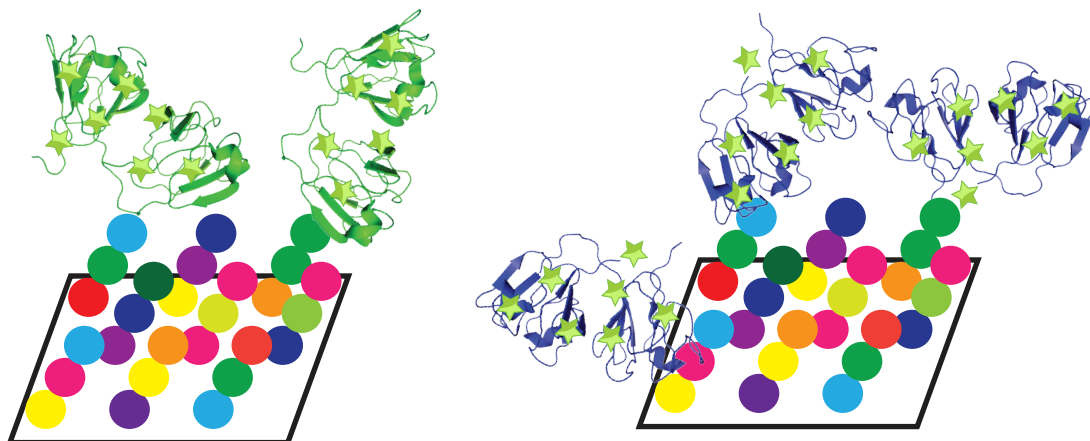
Figure 4.5: Representative tripeptide binding assay utilizing SPOT peptide libraries with Oregon Green dye. (A) SPOT synthesis of tripeptides using standard Fmoc chemistry. Libraries were made covering the sequence of γS-WT. (B) γS-WT and γS-G18V labeled with Oregon Green dye. (C) Samples were excited at 488 nm by a fluorescence laser scanner and emission intensities were recorded at 520 nm where emission intensities represented amount of protein binding to the tripeptides.
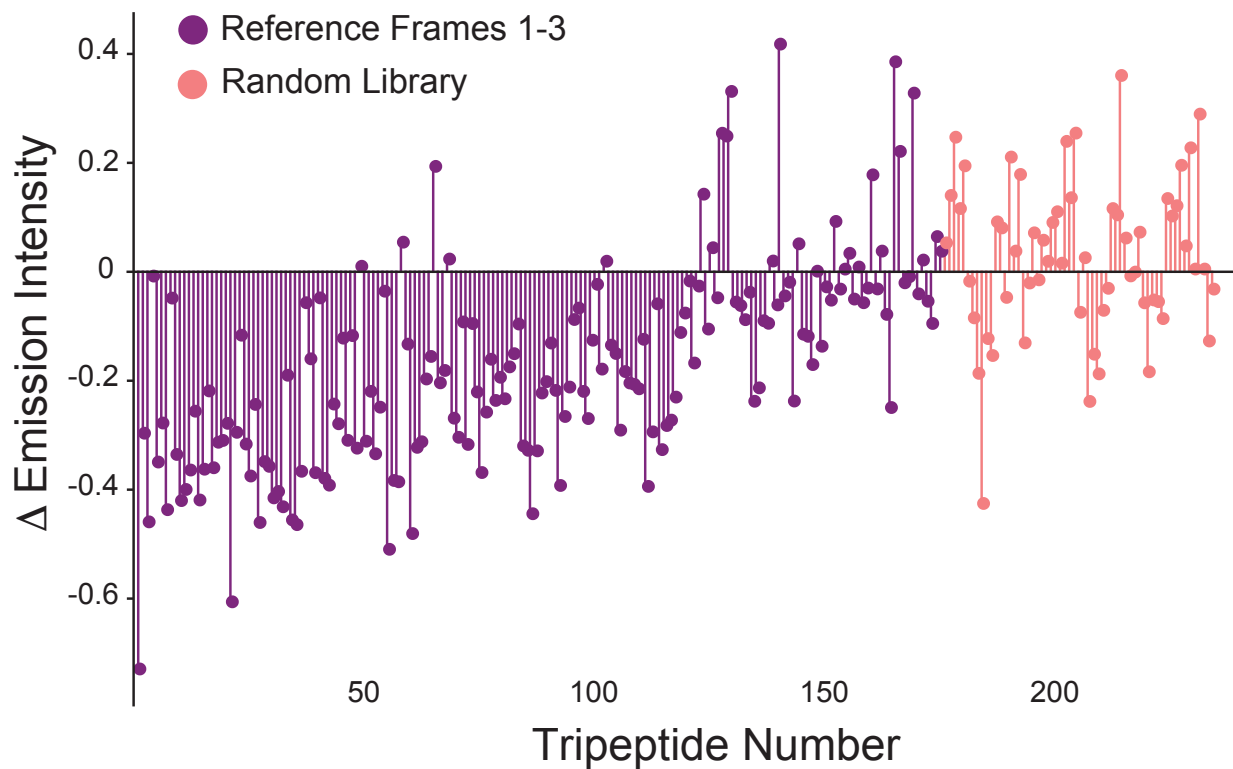
Figure 4.6: Graphic representation of the change in emission intensity between γS-WT and γS-G18V. The three tripeptide libraries using the first, second, and third reading frame are in purple while the random library is indicated in salmon. The three individual reading frames find more preferential binding towards γS-G18V while the random library does seem to bind to both proteins.

simulations, we focus on the 1:10 γS:tripeptide CSP data, shown in Figure 4.11. For γS-WT, all three tripeptides show strong binding at the disordered loop in the N-terminal domain, the N-terminal Greek key motif, the interdomain linker, and the surface of the C-terminal domain loop, particularly at residues H31, W47, L88, and H123. KTG shows additional strong binding near the C-terminal extension loop at residue V177. γS-G18V shows strong binding for all three tripeptides at the N-terminal Greek key motif and the surface of the C-terminal domain loop, in particular W47, W47ϵ, and H123 respectively. Additionally, for γS-G18V, the tripeptides RLS, SKT, and KTG show strong binding at the surface of the N-terminal domain helical loop, specifically residues H31 for RLS, T32 for RLS and SKT, and Y21 for SKT and KTG. Overall, tripeptides RLS, SKT, and KTG have shown to bind stronger to similar regions throughout both γS-WT and γS-G18V, with γS-G18V showing slightly higher perturbations overall.

Molecular surface representations for both γS-WT and γS-G18V are shown with weak- and strong-binding residues highlighted for all three tripeptides in Figure 4.12. For γS-WT, weak-binding residues are highlighted in pale green while strong-binding residues are highlighted in dark green, while for γS-G18V, weak-binding residues are highlighted in pale blue while strong-binding residues are highlighted in dark blue. Based on the calculated CSP, binding occurs throughout the N- and C-terminal domains for both proteins, but with subtle differences. For γS-WT, weak binding is observed at the N-terminal extension for all three tripeptide cases, something not observed in in the case of γS-G18V. However, in both γS-WT and γS-G18V, binding is observed around the inter-domain interface, consistent in all three tripeptide cases. We can compare these results with the rigid docking studies, where the low Boltzman-weighted contact frequency is used as a cutoff and residues above that cutoff have been highlighted in dark green for γS-WT and dark blue for γS-G18V. For RLS, SKT, and KTG, the experimental results where the strong- and weak-binding residues are highlighted are in strong agreement with the rigid docking results. More interestingly, γS-G18V is in complete agreement with the rigid docking results, highlighted in dark blue. This represen-
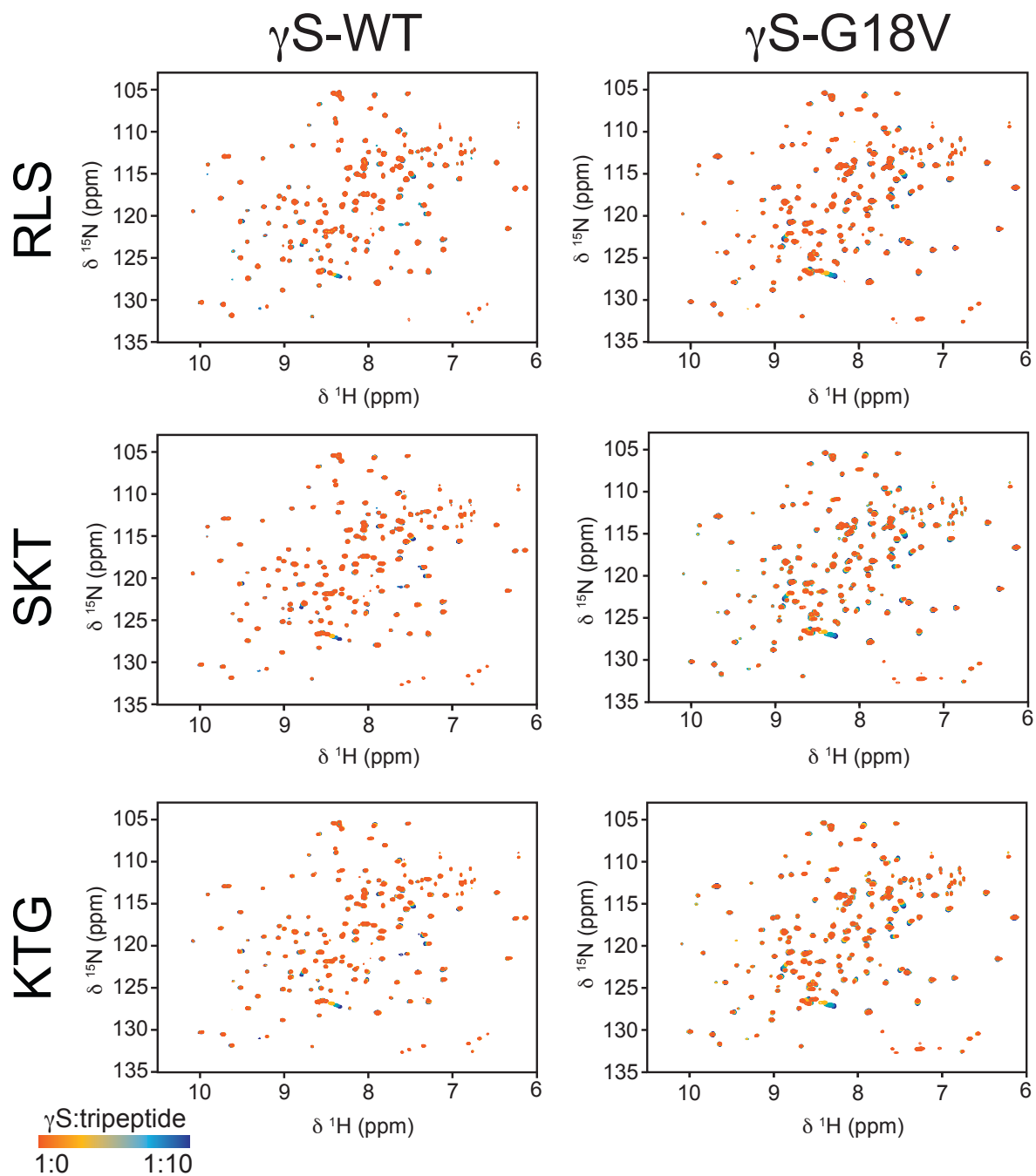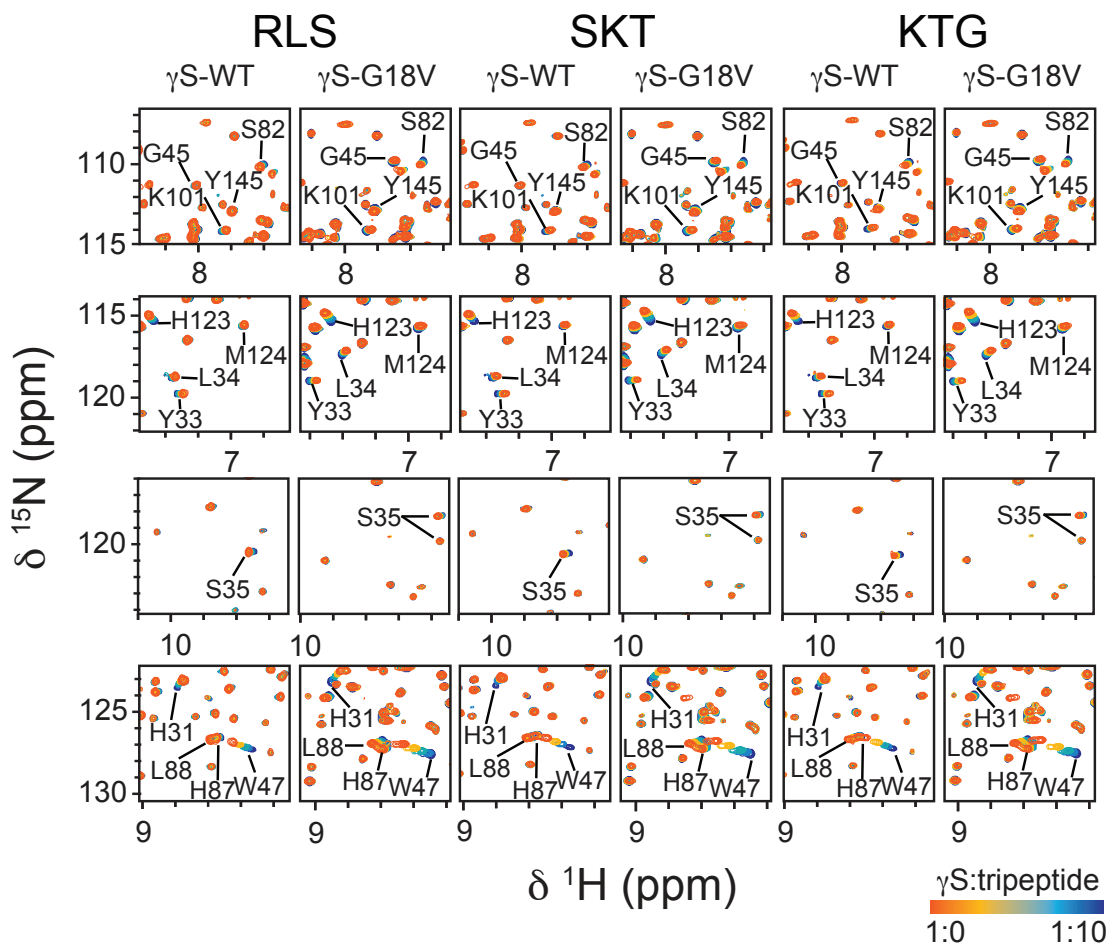
Figure 4.7: Chemical shift perturbation mapping of the selected ^{15}N-labeled γS-WT and γS-G18V with natural-abundance peptides RLS, SKT, and KTG, monitored by ^{1}H-^{15}N HSQCs. γS protein sample conditions consist of 10 mM phosphate buffer at pH 6.9 supplemented with 10% D$_2$O, 0.05% sodium azide and 2 mM TMSP for referencing. Protein concentrations for both γS-WT and γS-G18V were 1.5 mM. The γS:tripeptide ratios used were 1:0, 1:1, 1:2, 1:5, and 1:10. Panels showing representative portions of the spectra for tripeptides RLS, SKT, and KTG binding to γS-crystallin show both large chemical shift differences, such as residues H31 and W47, and smaller chemical shift differences, such as S35 and G45.
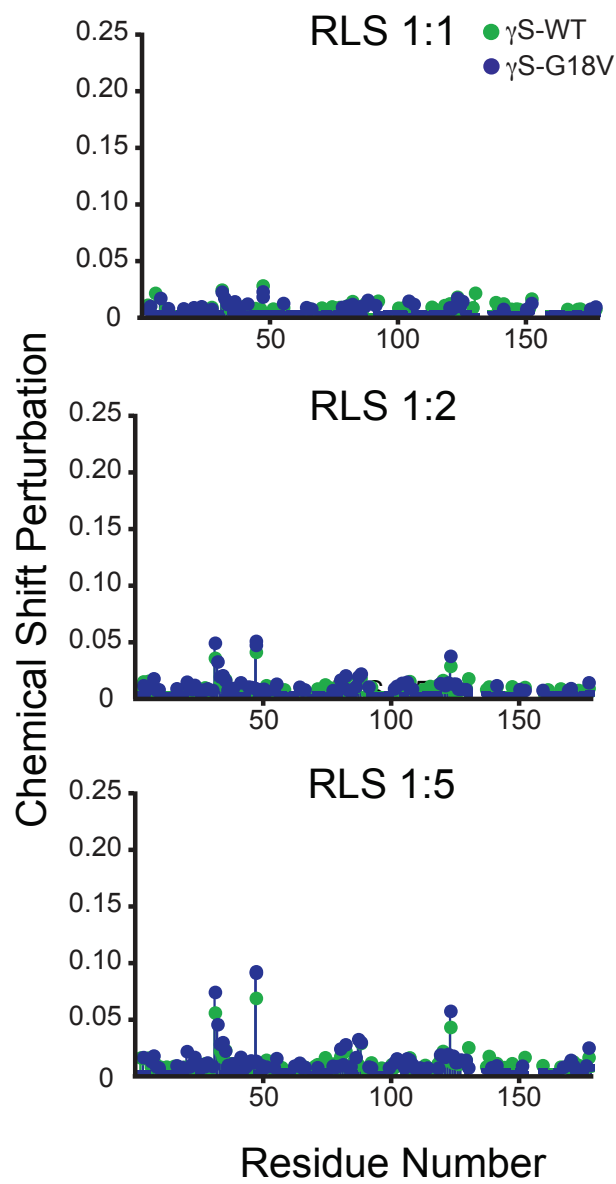
Figure 4.8: Average CSP of tripeptide RLS in the presence of γS-WT (green) and γS-G18V (blue).
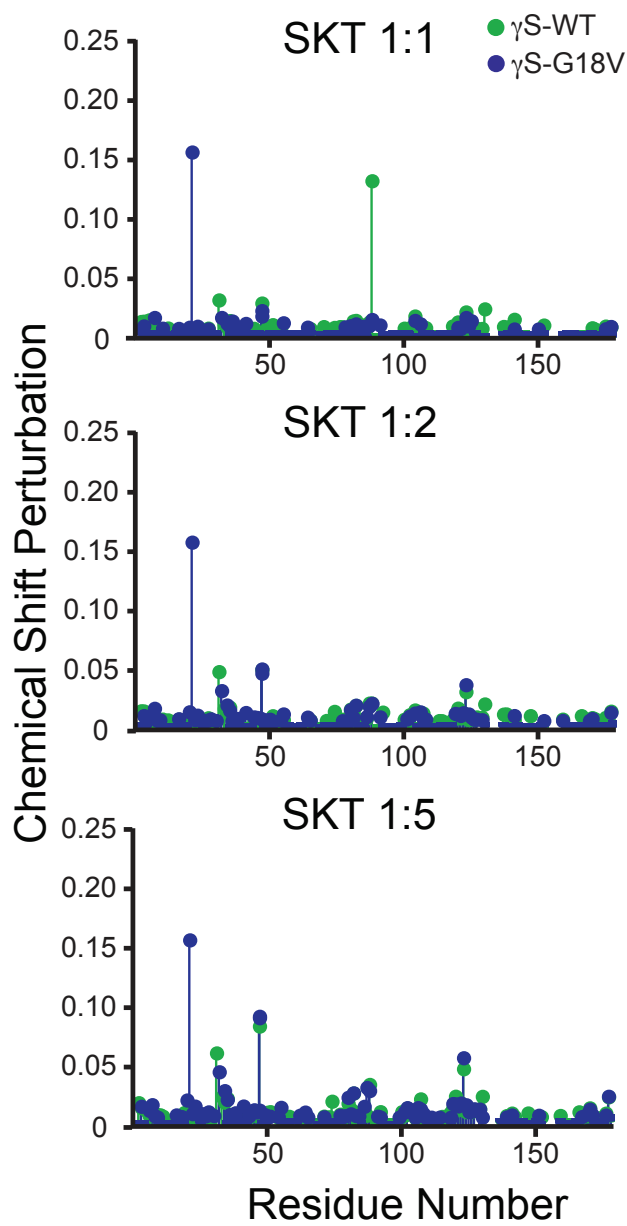
Figure 4.9: Average CSP of tripeptide SKT in the presence of γS-WT (green) and γS-G18V (blue).
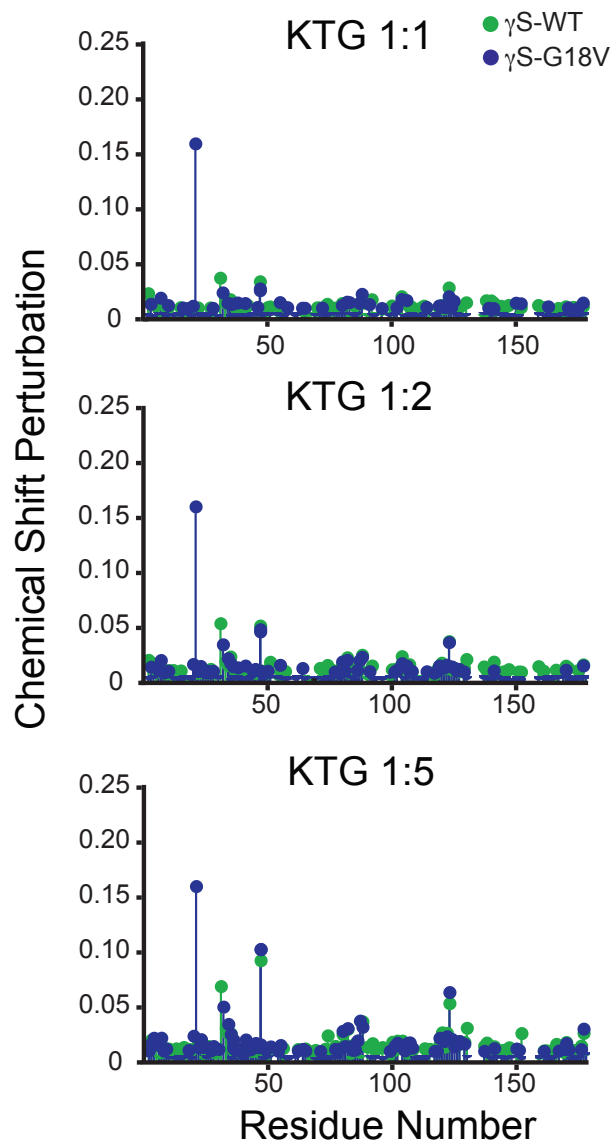
Figure 4.10: Average CSP of tripeptide KTG in the presence of γS-WT (green) and γS-G18V (blue).
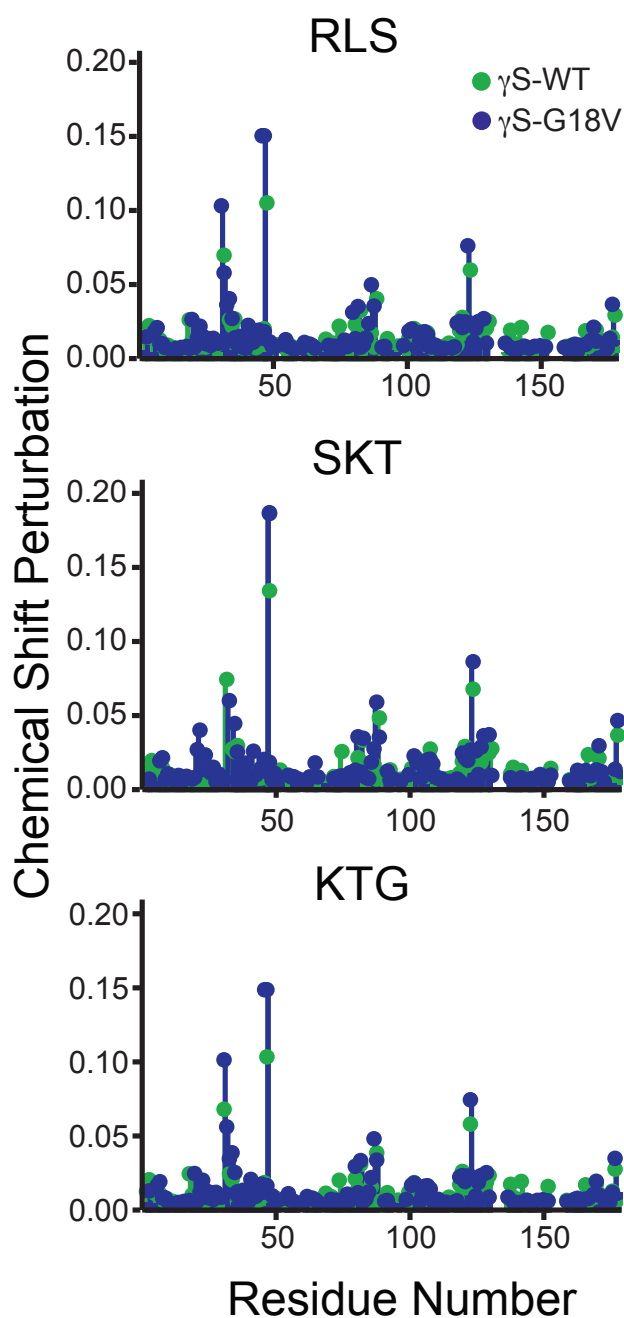
Figure 4.11: Average chemical shift perturbation (CSP) of γS-WT (green) and γS-G18V (blue). For γS-WT, all three tripeptides show strong binding to residues H31, W47, L88, and H123 with KTG showing strong binding additionally at residue V177. For γS-G18V, all three tripeptides show strong binding at residues W47, W47ε, and H123. Additionally, both RLS and SKT show strong binding at residue T32 while both SKT and KTG show strong binding at residue Y21. Lastly, RLS alone shows strong-binding towards residue H31, similar to γS-WT. All three tripeptides show strong binding to similar regions among both γS-WT and γS-G18V.

tation shows a qualitative measurement of where on the protein surface tripeptide binding occurs.

With the successful results of the rigid docking results, the flexible docking refinement was pursued to give a better understanding of how the tripeptides were binding at those interaction sites. However, due to the flexible nature of the tripeptides, the flexible docking results are inconsistent with consistent interaction sites. Therefore, to truly investigate the nature of the N-terminal extension loop of γS-crystallin, future experiments will incorporate preparing truncated variants of γS-crystallin where the N-terminal extension loop will be removed. DLS experiments will reveal whether or not the truncated variants of γS-crystallin form oligomeric complexes.

4.4 Conclusion

Here we comprehensively test the binding of representative tripeptides in γS-crystallin. Using CSP mapping and rigid docking simulations, we show that the tripeptides selectively bind to the cataract-related γS-G18V. Particularly, the tripeptides SKT and KTG represents four residues in the N-terminal extension loop, thus enforcing the idea that the N-terminal extension loop play a significant role in aggregate formation. To further investigate the interaction potential of the N-terminal extension loops of γS-WT and γS-G18V, truncated variants of these crystallins missing the N-terminal extension loop will be prepared. Whether or not these truncated variants form oligomeric species will ultimately reveal a better glimpse into how the N-terminal extension loop plays a role in cataract formation.
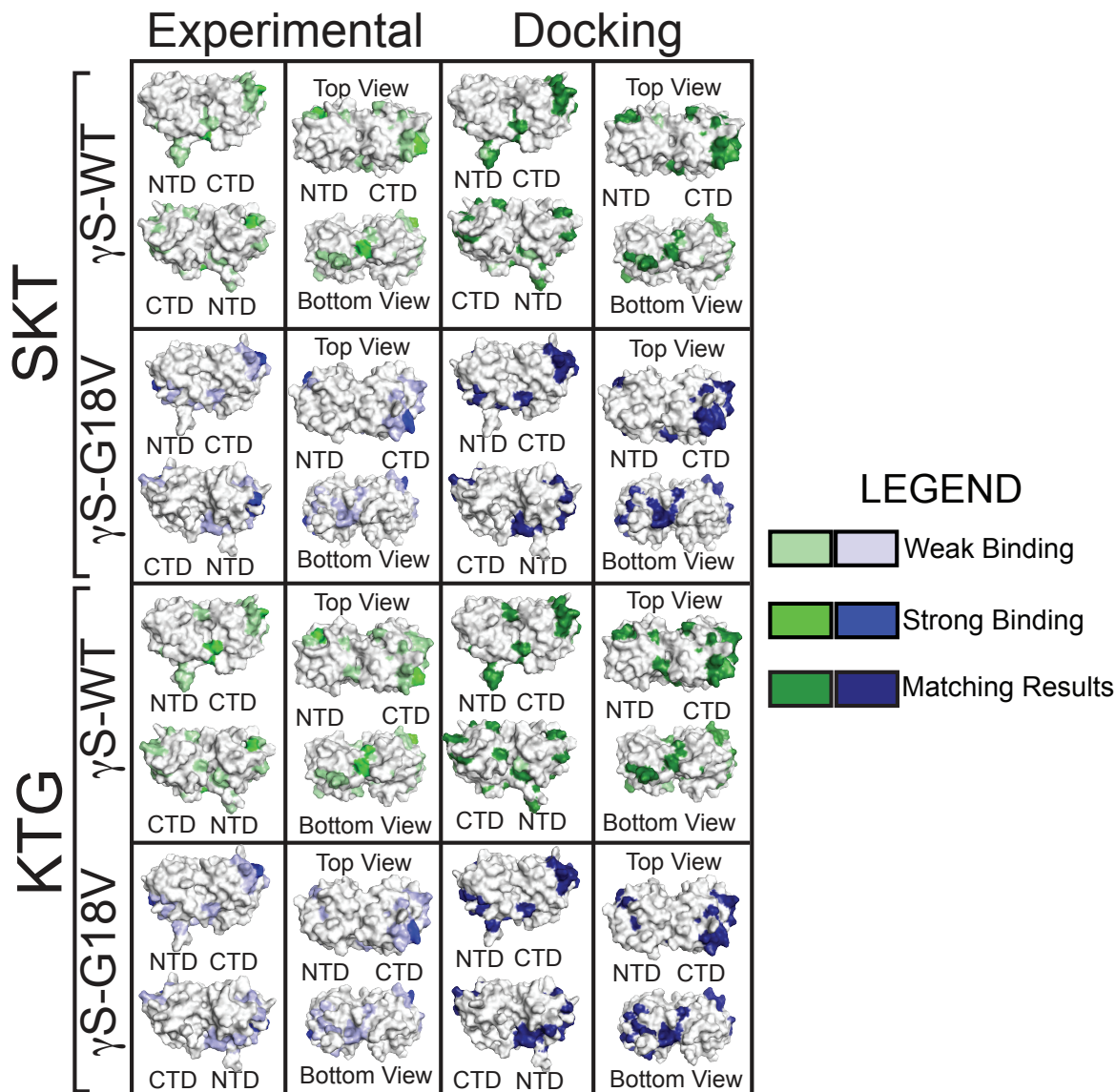
Figure 4.12: Molecular surface representation of tripeptide interactions with γS-WT and γS-G18V for SKT and KTG. A strong- and weak-binding threshold were defined as two times the RMS and half the RMS, respectively, using experimental CSP values and rigid docking simulations. The right panel displays the experimental results based on the NMR titration studies, while the left panel displays the matching rigid docking results. The left panel shows that tripeptide binding occurs throughout both the N- and C-terminal domains for γS-WT (strong-binding residues in green and weak-binding residues in pale green) and in γS-G18V (strong-binding residues in blue and weak-binding residues in pale blue). Coverage of both strong- and weak-binding residues are nearly identical between experimental and docking results, highlighted in dark green for γS-WT and dark blue for γS-G18V, indicating that the docking results are in good agreement with the experimental data.

Chapter 5

MALS

5.1 Background

Protein fold, solubility, and function can dictate protein stability in solution [35]. A change in pH, temperature, or ionic strength can cause proteins to unfold or precipitate, resulting to changes in solvent accessible surface area [134]. Controlling protein aggregation proves to be a challenge due to the numerous pathways, making it difficult to predict a protein's aggregation propensity. Current models of aggregation have found two factors contributing towards stability: colloidal and conformational stability [135, 134]. Colloidal stability is determined by the balance of repulsive and attractive intermolecular interactions between protein molecules in solution, applicable towards measuring the second osmotic virial coefficient [135, 134].

The second osmotic virial coefficient (B_2) is a thermodynamic property describing the pairwise interactions between proteins in a particular solvent [38, 136, 137, 134]. Positive B_2 values dictate predominantly mutual repulsive interactions, i.e. a protein molecule will interact with buffer, whereas negative B_2 values indicates predominantly attractive interactions,

i.e. a protein molecule will interact with another protein molecule making it possible for the protein to aggregate [38, 135, 137]. Measuring B_2 of protein solutions can be useful in modeling protein-protein and protein-solvent interactions which can be achieved through experimentation (Multi-angle light scattering) and simulation (multi-conformational Monte Carlo method).

Multi-angle light scattering (MALS) is a biophysical technique used to characterize biomacro-molecules by measuring the average molecular weight, the average radius, and intermolecular interactions. Samples prepared at various concentrations are injected into a flow cell where it is irradiated with light, typically from a laser source. The sample will scatter light, directly proportional to the molecular weight and concentration while detectors placed at different angles will measure the scattering intensity. By measuring the scattering intensity as a function of the scattering angle (θ), molecular weight and size can be determined. Zimm's second-order expansion of light scattering intensity with respect to the concentration for particles in dilute solution leads to the approximation [136, 138, 139]:

$$\frac{Kc}{R_\theta} \approx \frac{1}{MP_\theta} + 2B_2c \tag{5.1}$$

where K is an optical constant given by the equation [136, 138, 139]:

$$K = \frac{4\pi^2 n^2 (dn/dC)^2}{N_A \lambda^4} \tag{5.2}$$

and n is the refractive index of the solvent, dn/dC is the refractive index increment for the sample/solvent pair (mL/g), N_A is Avogadro's number (mol$^-$1, λ is the wavelength of the incident vertically polarized light in a vacuum (cm), c is the concentration of the sample (g/mL), R_θ is the Rayleigh ratio at a given scattering angle (cm$^-$1), M is the average molecu-lar weight of the sample(g/mol), and B_2 is the second osmotic virial coefficient (mol*mL*g$^-$2) [139]. MALS can experimentally determine protein-protein and protein-solvent interactions.

Monte Carlo (MC) simulations use random sampling and statistical modeling to investigate aggregation propensity in protein solutions. These simulations involve translational and rotational motions of known structures to determine protein-protein interaction potentials in solution [136, 140, 141]. However, proteins are dynamic, meaning that a single protein molecule exists in various orientations and conformations. The Tobias Lab at UCI developed a new method known as the multi-conformational Monte Carlo (mcMC) simulation which implements a conformational swap in the simulation method [136]. The conformational swap uses all possible protein conformations to better simulate protein-solvent interactions [136]. To test the mcMC method, experimental MALS measurements of hen egg white lysozyme were used as a control to compare with single-conformation Monte Carlo (scMC) simulations. The optimization of the mcMC method allows for simulating B_2 values for γS-crystallin and other variants to understand and compare their aggregation propensities.

5.2 Methods for investigating protein-protein interactions

5.2.1 Hen egg white lysozyme sample preparation

Lyophilized hen egg white lysozyme (Cat No. 195303) was purchased from MP Biomedicals (Solon, OH). Lysozyme was dissolved in 10 mM sodium phosphate with 0.05% sodium azide (pH 4.7 and 6.9) and sodium chloride concentrations at 50, 75, 100, 125, 150, 200, 250, and 300 mM for a final concentration of 50 mg/mL. Lysozyme samples were prepared analytically with concentrations ranging from 2.5 to 50 mg/mL. These concentrations were checked by UV absorbance measurements using $\epsilon = 2.64$ mL mM^{-1} cm^{-1} at 280 nm.

5.2.2 Multi-angle light scattering experiments

Light scattering experiments were performed in combination with a Dawn HELEOS multi-angle light scattering (LS) detector and an Opitlab rEX refractive index (RI) detector from Wyatt Technology (Santa Barbara, CA). Samples were injected using the batch-mode technique from lowest to highest concentrations. Filters (0.1 μm) were used to ensure monodispersity as samples are injected through the flow cell. Scattering intensity data from each concentration was processed to remove artifacts caused by the sample injection by use fo a clustering algorithm developed by Carter T. Butts at UCI [142]. The median of the remaining observations were used as the scattering intensity measurements for each detector [136, 142].

5.2.3 Estimation of second virial coefficients for lysozyme

Zimm's second-order expansion of light scattering intensity (Equation 5.1) was used to determine B_2 values for lysozyme [136, 142].

5.2.4 Single-conformational and multi-conformational Monte Carlo simulations

Monte Carlo simulations were performed on lysozyme solutions under the same conditions used for MALS measurements on three different structures obtained from neutron scattering (PDB ID: 1IO5 and 1LZN) [6, 7] and solution NMR (PDB ID: 1E8L, model 1) [8] experiments [136]. All three structures contained information on protonation states and proton coordinates as well as information regarding heavy atom coordinates. The Monte Carlo simulations were performed by Vera Prytkova from the Tobias Lab [136].

5.3　Results and Discussion

5.3.1　Batch-mode collection of lysozyme shows artifacts in scattering intensity data

MALS data were collected for lysozyme in 10 mM sodium phosphate and 0.05% sodium azide (pH 4.7 and 6.9) at varying ionic strengths. Using batch-mode collection, samples were injected through a flow cell using a syringe pump starting with buffer only, followed by lysozyme samples (lowest to highest conentration),and finally ending with buffer only for baseline correction. Batch-mode collection does introduce potential artifacts causing some inconsistencies in the scattering intensities. These artifacts are results of air bubbles or back pressure from the injection of a new sample during data collection. Scattering intensities were recorded for each sample concentration and Figure 5.1 portrays some of the artifacts observed in black. With the use of a clustering algorithm, the artifacts were removed, allowing the real scattering intensities shown in Figure 5.2 to be used in the analysis.

5.3.2　Experimentally determined B_2 agrees with multi-conformation Monte Carlo simulations

Single-conformation and multi-conformation Monte Carlo simulations were performed to determine B_2 values. Three structures of lysozyme included neutron scattering structures (PDB ID: 1IO5 and 1LZN) [6, 7] and a solution NMR structure (PDB ID: 1E8L, model 1) [8]. All lysozyme measurements show that B_2 decreases as the ionic strength increases, meaning that lysozyme exhibits attractive interactions. The increasing concentrations of sodium chloride in solution causes an overall decrease in repulsive electrostatic interactions of charged proteins, which is why sodium chloride is often viewed as a precipitating agent
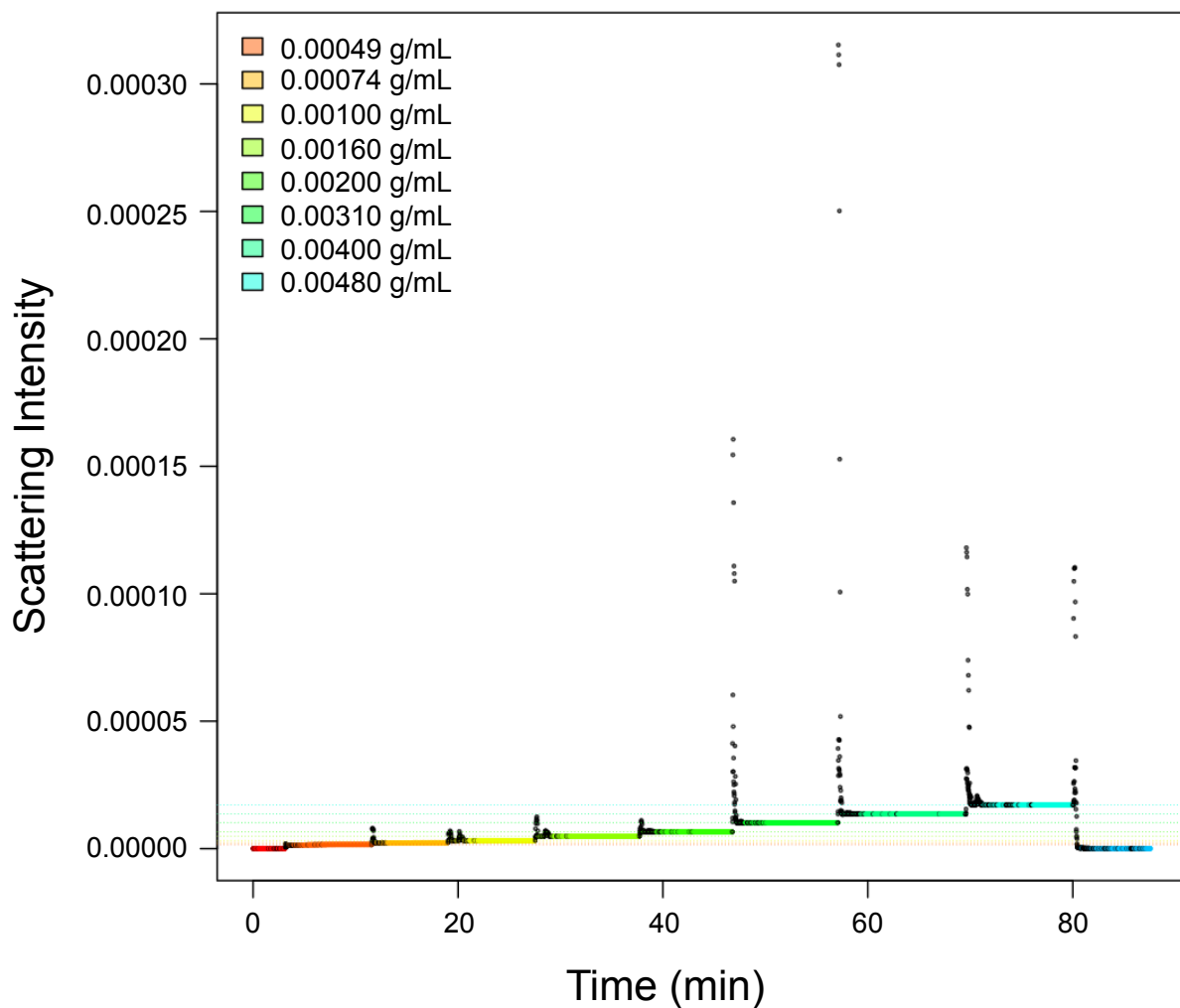
Figure 5.1: Light scattering data of lysozyme in 10 mM phosphate, 100 mM sodium chloride, 0.05% sodium azide at pH 6.9. Scattering intensity is recorded over the time of the experiment. Each color represents a particular concentration of lysozyme being injected into the MALS instrument with the first and last being buffer for baseline correction. Black areas indicate artifacts introduced by the sample injection.
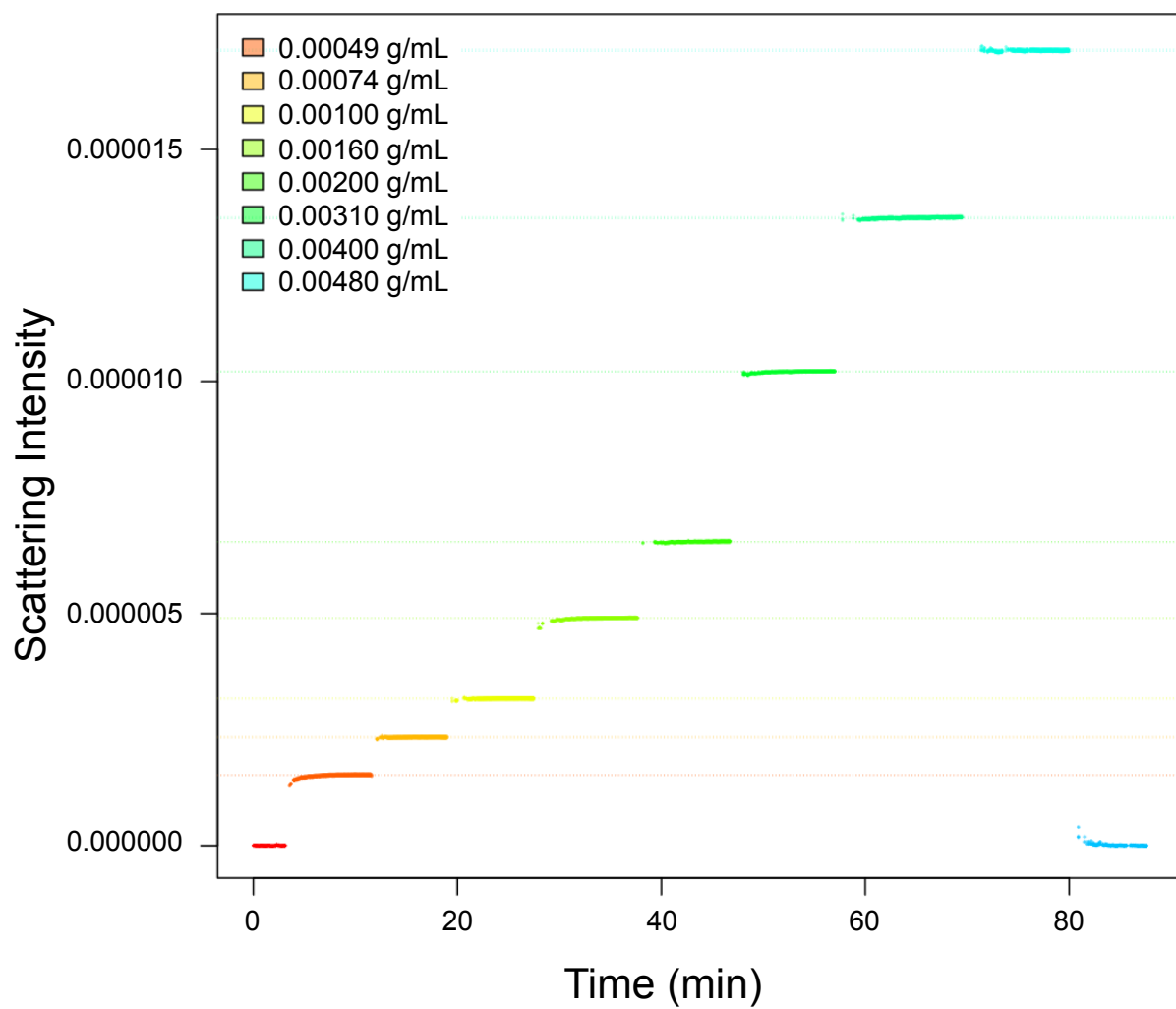
Figure 5.2: Light scattering data of lysozyme in 10 mM phosphate, 100 mM sodium chloride, 0.05% sodium azide at pH 6.9 where artifacts have been removed. Median scattering intensity used as measurement from each detector.

[143]. When comparing single-conformation Monte Carlo simulations shown in Figure 5.3, experimental values slightly match the solution NMR structure of lysozyme chosen. The two neutron scattering structures of lysozyme have overall slightly higher B_2 values than the solution NMR structure used in the scMC simulations. The solution NMR structure of lysozyme B_2 values closely match the experimentally measured B_2 values. Including a conformational swap into the simulation method allows for more protein conformations to be included, better simulating protein-protein/protein-solvent interactions overall. Using the multi-conformation Monte Carlo method with the solution NMR structure of lysozyme, B_2 values better match the experimental values for lysozyme, as shown in Figure 5.4.

5.4   Conclusion

I have successfully collected lysozyme light scattering data with increasing ionic strengths for different pH conditions using Carter T. Butts's clustering algorithm to clean up light scattering data and Vera Prytkova's multi-conformation Monte Carlo simulation method. Comparing single-conformation Monte Carlo simulations with experimental data shows that matching experimental results with simulation results depends on which lysozyme structure is chosen for the simulation method. Including several conformations of a protein structure in the simulation method, the multi-conformation Monte Carlo method, better simulates protein-protein/protein-solvent interactions. The mcMC method better simulated B_2 values that closely agreed with experimentally measured B_2 values for lysozyme at two different pH conditions. With the success of better simulating B_2 values using the mcMC method, we will now apply this method towards γS-crystallin and other variants to determine the aggregation propensity for these systems.
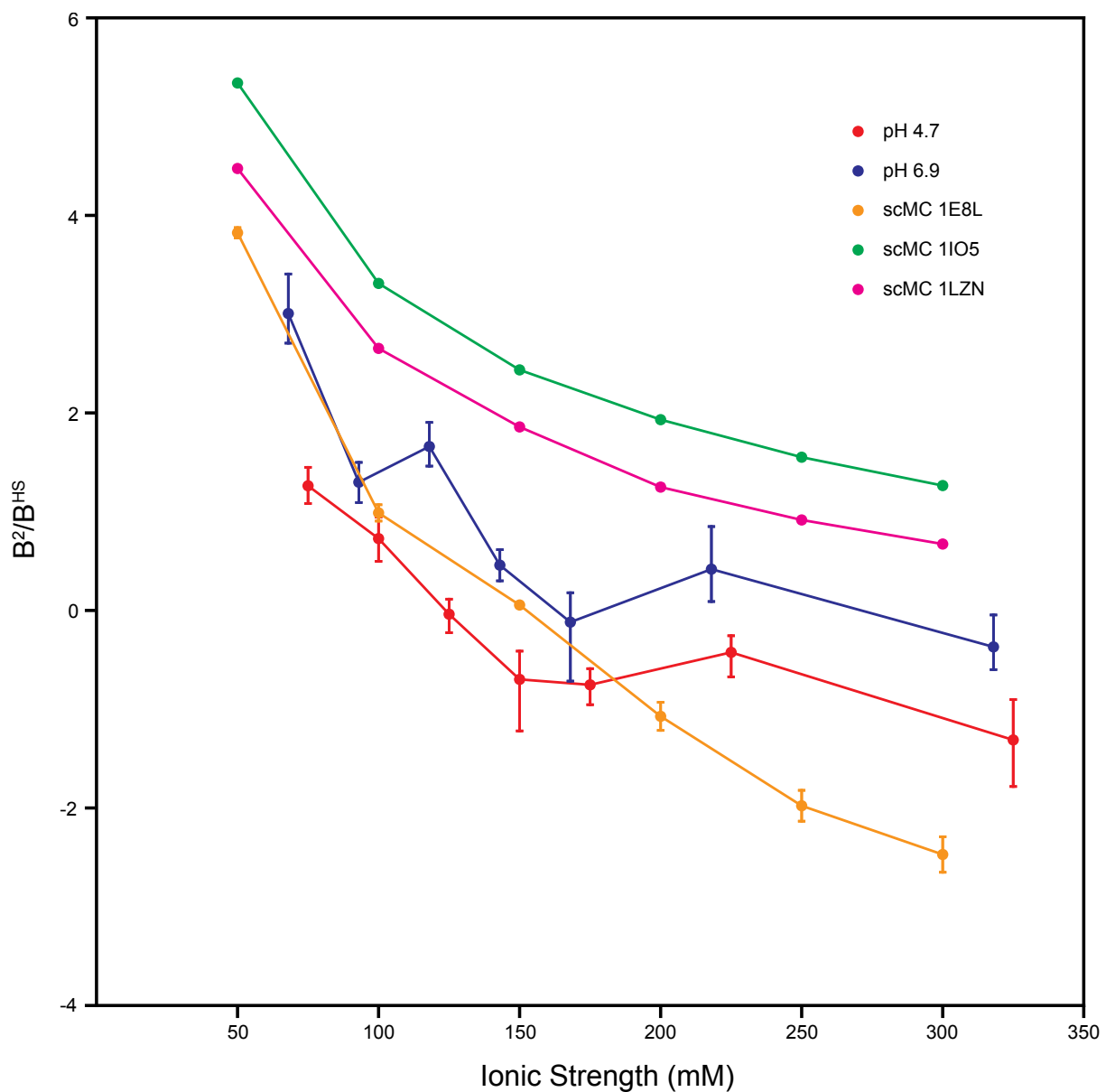
Figure 5.3: B_2 values for lysozyme, which decrease as ionic strength increases. Experimentally measured B_2 using MALS: lysozyme in 10 mM phosphate, 100 mM sodium chloride, 0.05% sodium azide at pH 4.7 (red) and pH 6.9 (blue). Single-conformation Monte Carlo simulations used three lysozyme structures, including neutron scattering structures (PDB ID: 1IO5 and 1LZN, green and pink, respectively) [6, 7] and a solution NMR structure (PDB ID: 1E8L, model 1, orange) [8]. The scMC simulated B_2 values using the solution NMR structure closely resemble the experimentally gathered B_2 values.
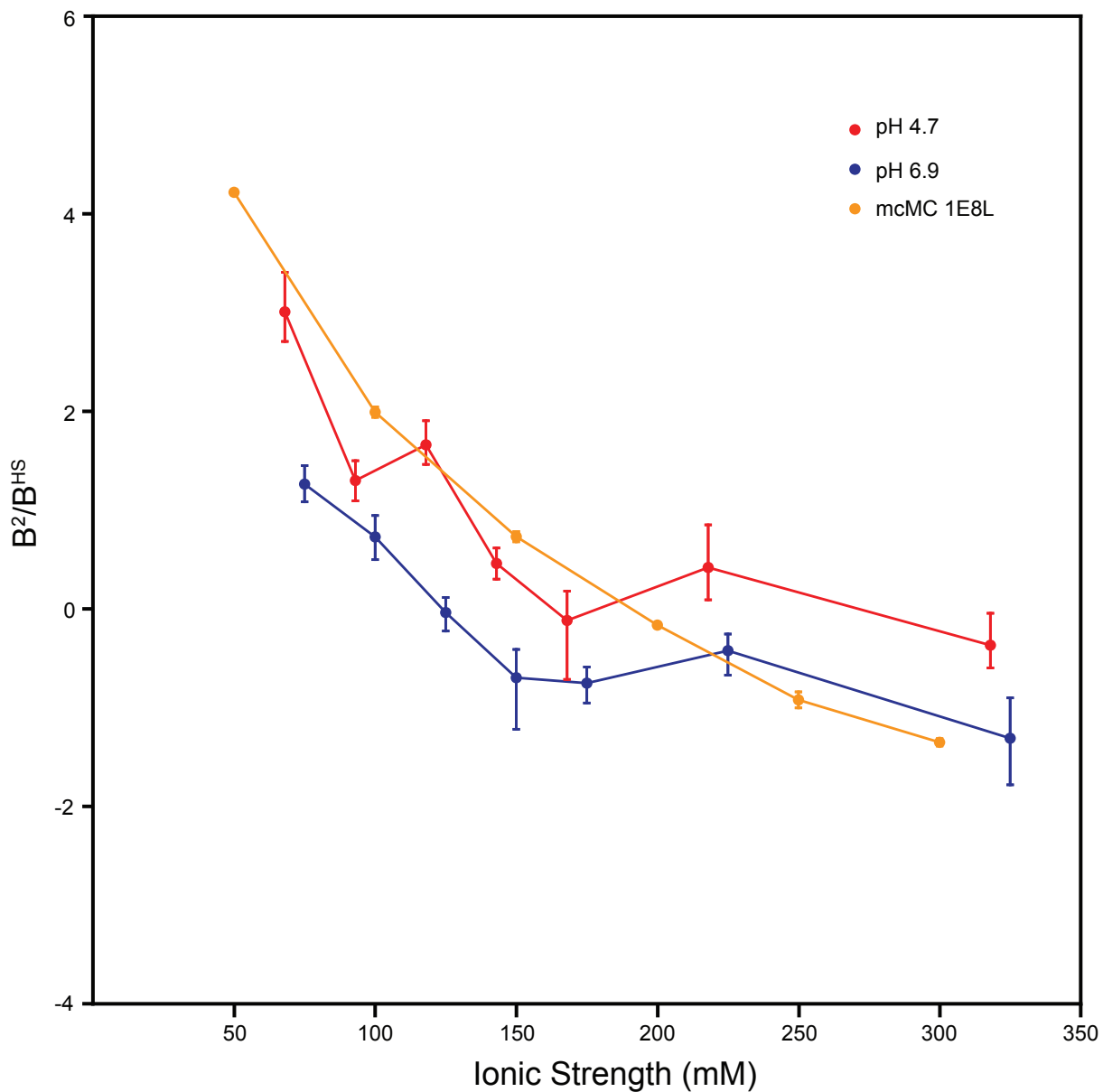
Figure 5.4: B$_2$ values for lysozyme, which decrease as ionic strength increases. Experimentally measured B$_2$ using MALS: lysozyme in 10 mM phosphate, 100 mM sodium chloride, 0.05% sodium azide at pH 4.7 (red) and pH 6.9 (blue). Multi-conformation Monte Carlo simulations included a solution NMR structure (PDB ID: 1E8L, model 1, orange) [8]. The mcMC simulated B$_2$ values have better agreement with the experimentally gathered B$_2$ values, compared to Figure 5.3.

Chapter 6

Biophysical characterization of J2-crystallin: A novel eye lens protein

6.1 Background

Many crystallins are derived from either metabolic enzymes or physiological stress proteins and appear to have undergone selection for increased refractive index after gene duplication [46]. Mammalian lenses consist of two strongly conserved classes of crystallins, the chaperone α-crystallins, and the structural $\beta\gamma$-crystallins [42, 43, 53, 46, 144, 145] whereas taxon-specific crystallins with diverse structural properties are found in other species, including the ϵ-crystallin in avians and reptiles and the S-crystallins in cephalopods [42, 53]. Unlike other non-cephalopod invertebrates, in which typical visual systems consist of simple ocelli and/or compound eyes made up of an array of ommatidia [146, 147], the box jellyfish, *Tripedalia cystophora*, has camera-type eyes similar to those of vertebrates and cephalopods [146]. Its lenses are capable of forming sharp images, although the distance between the lens and the retina is not optimized for maximum visual acuity, since the animal uses vision primarily for

navigation rather than detection of detailed objects [148]. *T. cystophora* has a total of 24 eyes split among four rhopalia, around the bell of the medusa, each containing two camera-type eyes as well as an array of simpler pigment cup eyes [149, 147]. The camera-type eye lenses are composed of the J1-, J2- and J3-crystallin proteins [149]. J1- and J3-crystallins are homologous to known enzymes, (ADP-ribohydroglycosylase and saposins, respectively) [149], whereas J2-crystallin (J2) has no known homologs [148, 24]; a BLAST search of the Protein Data Bank (PDB) found no proteins above 36% similarity [150]. To our knowledge, J2 has not previously been expressed and characterized. Here we focus on the box jellyfish eye lens protein J2-crystallin. This 157-amino acid protein has a molecular weight of 18.2 kDa [42] and a theoretical isoelectric point (pI) of 9.25 [63].

6.2 Materials and Methods

6.2.1 Gene Construction

Plasmids containing the cDNA sequence of box jellyfsh J2-crystallin was purchased from Blue Heron (Bothell, WA). Primers were purchased from Sigma-Aldrich (St. Louis, MO) and were amplified with primers containing flanking restriction sites for NcoI and XhoI, an N-terminal 6x His tag, and a TEV cleavage sequence (ENLYFQ). The polymerase chain reaction product was cloned into a pET28(+)a vector, purchased from Novagen (Darmstadt, Germany). The gene construction was performed by Timothy R. Valentic.

6.2.2 Expression and purification of J2-crystallin

A 50 mL starter culture of LB media was innoculated with a single colony of Rosetta (DE3) *E. coli* containing the vector of box jellyfsh J2-crystallin. Starter cultures were grown at 37

°C for 16 hours at 225 RPM in a New Brunswick Scientific Innova-42R incubator Shaker (Hauppauge, NY). The individual starter cultures were used to innoculate a 1 L culture of LB with a starting OD_{600} at approximately 0.20. The cultures were grown at 37 °C with shaking at 225 RPM until an OD_{600} of 0.60 was reached. Protein overexpression was induced using IPTG (Gold Biotechnology) at a final concentration of 0.10 mM at 25 °C for 18 hours (J2-crystallin). Cells were collected via centrifugation with a Beckman Coulter Avanti-JE centrifuge (Brea, CA) spinning at 6000 RPM for 20 minutes. The cell pellets were resuspended in 40 mL of 50 mM sodium phosphate buffer with 300 mM sodium chloride, 10 mM imidazole, 0.05% sodium azide at pH 7.4. Cells were lysed by sonication in 10 second intervals for a total of 30 minutes, followed by centrifugation at 15000 RPM for 90 minutes. The supernatant was filtered with through a 0.22 μm filter (Millipore) before being loaded onto a Bio-Rad Duo-Inject FPLC system (Hercules, CA). The His-tagged crystallin were purified using a Ni-IDA column (Bio-Rad) and cut by a His-tagged TEV protease (produced in-house). The TEV protease and His-tag were removed by a second application to a Ni-IDA column. The purified crystallin were dialyzed into 10 mM sodium phosphate buffer with 0.05% sodium azide at pH 6.9.

6.2.3   Expression and purification of ^{15}N-labeled J2-crystallin

A 50 mL starter culture of LB media innoculated with a single colony of Rosetta (DE3) *E. coli* cells containing a pET28(+)a vector with J2-crystallin gene inserts was grown at 37 °C for 16 hours with shaking at 225 RPM. The individual starter cultures were used to innoculate a 1 L culture of LB with a starting OD_{600} at approximately 0.20. The cultures were grown at 37 °C with shaking at 225 RPM until an OD_{600} of 0.60 was reached. The cells were then collected in 500 mL batches by centrifugation at 3000 RPM for 30 minutes and each 500 mL batch was resuspended in 1 L ^{15}N-labeled minimal media cultures. The

1 L minimal media cultures were grown for an additional 2 hours at 37 °C at 225 RPM. Protein overexpression was induced using IPTG (Gold Biotechnology) at a final concentration of 0.10 mM at 25 °C for 30 hours. Cells were collected via centrifugation, spinning at 6000 RPM for 20 minutes. The cell pellets were resuspended in 40 mL of 50 mM sodium phosphate buffer with 300 mM sodium chloride, 10 mM imidazole, and 0.05% sodium azide at pH 7.4. Cells were lysed by sonication in 30 second intervals for a total of 10 minutes, followed by centrifugation at 15000 RPM for 90 minutes. The supernatant was filtered with through a 0.22 μm filter (Millipore) before being loaded onto a Bio-Rad Duo-Inject FPLC system (Hercules, CA). The His-tagged crystallin was purified using a Ni-IDA (Bio-Rad) and cut by a His-tagged TEV protease. The TEV protease and His-tag were removed by a second application to a Ni-IDA column. The purified ^{15}N-labeled crystallin was dialyzed into 10 mM sodium phosphate buffer with 0.05% sodium azide at pH 6.0. The final J2-crystallin sample was prepared in 10 % D_2O and 2 mM TMSP at a final concentration of 1.8 mM.

6.2.4   Electrospray ionization mass spectrometry of J2-crystallin

Purified J2-crystallin was prepared at 1.07 mg/mL was dialyzed into nanopure water for a period of 48 hours where the water was changed every 12 hours to remove as much salt as possible. J2-crystallin was injected into a Waters Micromass QTOF2 (Milford, MA) in the presence of 50% acetonitrile and 0.1% formic acid (Micromass LCT).

6.2.5   Far-UV circular dichroism

Purified J2-crystallin samples were prepared at 0.083 mg/mL in 10 mM sodium phosphate buffer with 0.05% sodium azide at pH 6.9. Measurements were taken with a JASCO: J-18 Spectropolarimeter (Easton, MD) with a temperature control unit JASCO: PFD-425S

(Easton, MD). Scans were performed with high sensitivity detection from 260 nm to 180 nm using the step-scanning mode. Data pitch was set at 1 nm at a 16-second response time, bandwidth set at 1 nm, and accumulation of four scans. Data was adjusted using Savitzky-Golay smoothing [10].

6.2.6 Differential scanning fluorometry

Purified J2-crystallin samples were prepared at a final concentration of 5 μM in in 10 mM sodium phosphate buffer with 0.05% sodium azide at pH 6.9 in the presence of 40 μM SYPRO Orange dye. Differential scanning fluorometry (DSF) with samples of J2-crystallin were assessed using an Agilent Mx3005P QPCR machine (Santa Clara, CA). The excitation and emission wavelengths were 492 nm and 610 nm, respectively. Fluorescence readings were recorded from 2595 °C with a temperature ramp of 1 °C/min and recorded in duplicate [151]. The data were fit by a nonlinear regression analysis using the Boltzmann function and melting temperature values were determined by the inflection point of the resulting sigmoidal curves defined from 25-87 °C.

6.2.7 NMR experiments

Experiments were performed on a Varian $^{\text{Unity}}$INOVA spectrometer (Agilent Technologies, Santa Clara, CA) operating at 800 MHz and equipped with a ^{1}H–^{13}C–^{15}N 5 mm tri-axis PFG triple-resonance probe, using an 18.8 Tesla superconducting electromagnet (Oxford Instruments). Decoupling of ^{15}N nuclei was performed using the GARP sequence [98]. ^{1}H chemical shifts were referenced to TMSP, and ^{15}N shifts were referenced indirectly to TMSP. NMR data were processed using NMRPipe [99] and analyzed using CcpNMR Analysis [100]. Center operating frequencies and (unless otherwise stated) center frequency offsets were as follows:

Center	^{1}H: 799.7988955 MHz	^{15}N: 81.04252684 MHz
Offset	^{1}H: 4.81 ppm	^{15}N: 118.70 ppm

6.3   Results and Discussion

6.3.1   J2-Crystallin optimal expression and purification

Rosetta cells transformed with a pET 28a(+) vector containing DNA encoding N-terminal His-J2 were expressed at 25 °C for 24 hours. Expression and purification conditions were monitored via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) to examine if the isolation of J2-crystallin was successful. Referring to Figure 6.1, the blue band around 18 kDa becomes darker starting from lane 2 going through lane 4, indicating that the protein was expressed successfully. As the purification steps progressed, the subsequent lanes show bands that were not J2-crystallin disappearing and the band that was J2-crystallin become noticeably narrower, indicating that the purification of J2-crystallin was successful.

For further verification in isolating J2-crystallin, electrospray ionization mass spectrometry (ESI-MS) was used providing for the molecular mass of the protein. ESI-MS uses electrical energy to transfer ions from a solution phase to a gaseous phase through a capillary tube at a high, positive voltage [152]. Mass spectra were generated for J2-crystallin, shown in Figure 6.2, one being the cluster of multiple-charged ions (A) and the other being the deconvoluted spectrum showing the true mass scale after processing (B). The spectra containing the cluster of ions are due to the protein picking up proton charges from the number of accessible basic sites [152]. Since each peak only differs by one proton charge, one can set up two simple equations and solve for two unknowns: the charge (z) and the mass of the protein (m). In relation to the J2-crystallin spectrum, the mass of the protein was solved to be 18251.65 Da
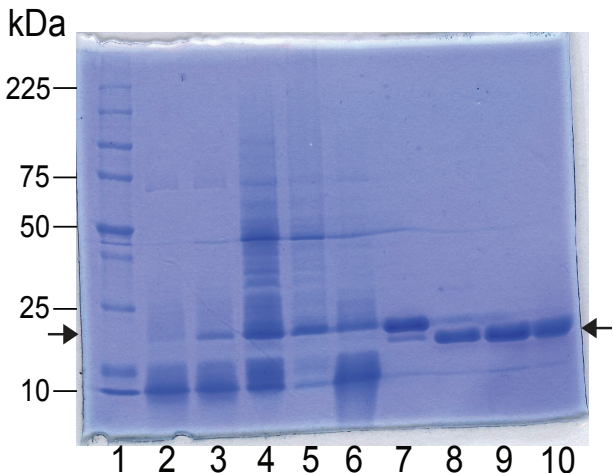
Figure 6.1: SDS-PAGE gel used to determine the presence of J2-crystallin. Samples of the cell culture (1 mL) were collected before induction, 1 hour after induction, and 24 hours after induction. Samples of protein (50 μL) were collected after subsequent purification steps to determine the isolation of J2-crystallin. Each lane of the gel represents the molecular weight marker (Promega) (1), cells pre-induction (2), cells 1 hour post-induction (3), cells 24 hours post-induction (4), insoluble cell pellet (5), cell-free extract (6), J2-crystalin with polyhistidine tag (7), TEV digest (8), J2-crystallin after TEV removal (9), purified J2-crystallin (10). The arrow indicates the presence of J2-crystallin at approximately 18 kDa.

and the charge was solved to be 8.00.

A secondary structure prediction was performed for J2-crystallin, using the Self Optimized Method (SOPM) via the ULC Bioinformatics PSPIRED Protein Structure Prediction Server and is shown in Figure 6.3 [9]. J2-crystallin is predicted to contain primarily α-helices (42%), unlike vertebrate α- and $\beta\gamma$-crystallins, which typically contain less than 10% α-helical secondary structure [46, 24]. This protein appears to have a novel fold that is not obviously related to any known enzymes or taxon-specific crystallins.
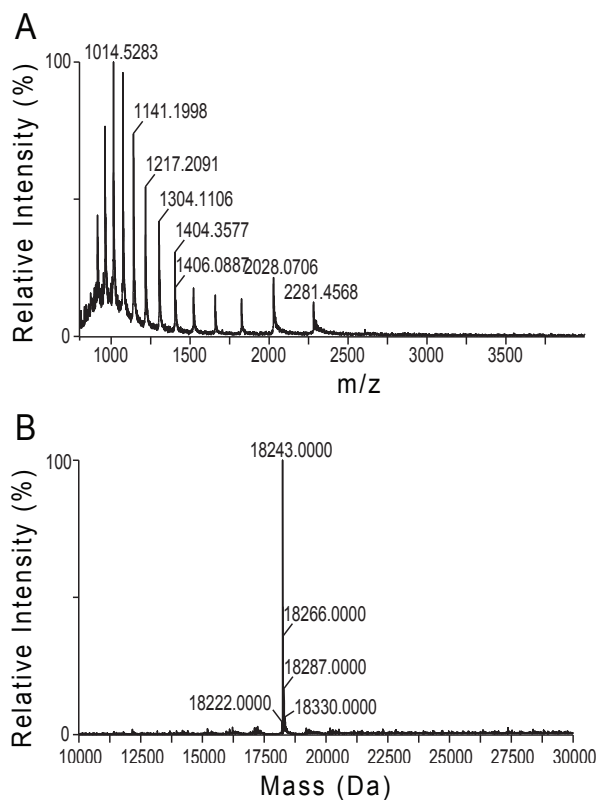
Figure 6.2: ESI-MS spectra of purified J2-crystallin prepared at 1.07 mg/mL in 10 mM sodium phosphate at pH 6.9 dialyzed against nanopure water for 48 hours. The water was changed every 12 hours to remove as much salt as possible. The sample of protein was injected in the presence of 50% acetonitrile and 0.1% formic acid, denaturing the protein. The spectrum in panel A represents the raw data showing the cluster of ions while panel B represents the deconvoluted spectrum of J2-crystallin on the true mass scale after processing. The mass of J2-crystallin was found to be 18251.65 Da and the charge was found to be 8.00.
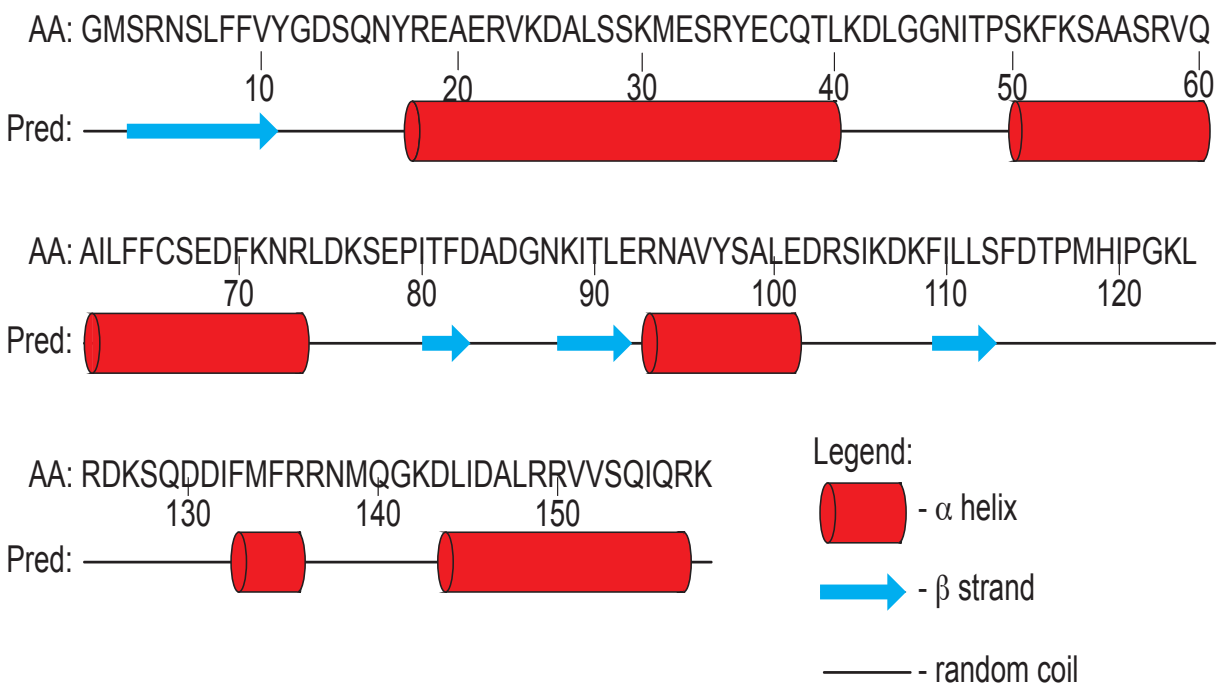
Figure 6.3: PSPIRED secondary structure prediction results for J2-crystallin [9]. The amino acid sequence (AA) is represented by each residues one letter code and the prediction (Pred) is presented as α-helix (red rod), β-strand (cyan arrow), or random coil (black line). The majority of the protein is predicted to contain α-helices.
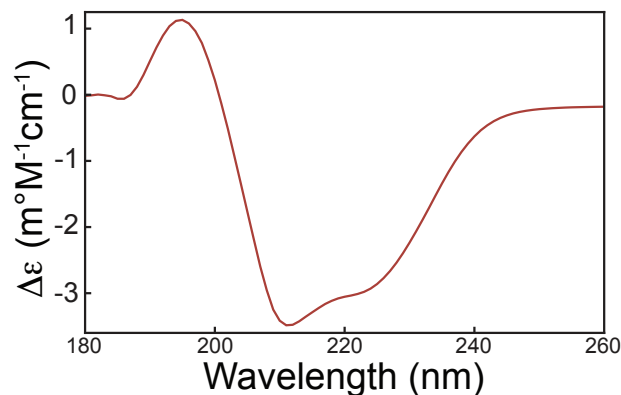
Figure 6.4: Far-UV CD spectrum of J2-crystallin at 0.083 mg/mL. Minimum ellipticities at 211 nm and 218 nm are consistent with α-helical secondary structure. Measurements were taken at 10 °C. Scans were performed with high sensitivity detection from 260 to 180 nm using the step-scanning mode. The data pitch was set at 1 nm at a 16-second response time, bandwidth set at 1 nm, and accumulation of four scans. Data was adjusted using Savitzky-Golay smoothing [10].

6.3.2 Biophysical characterization of J2-crystallin reveals α-helical content

Various biophysical techniques were used to investigate the structure of J2-crystallin. The overall structure of J2-crystallin was qualitatively measured by far-UV circular dichroism (CD) spectroscopy while conformational properties of J2-crystallin were characterized by fluorescence methods. CD spectroscopy provides a qualitative measure of protein secondary structure by measuring the amide-amide orientation of the protein backbone that absorb into the far UV region [153]. CD measurements support the secondary structure prediction, indicating that J2-crystallin is primarily α-helical. The far-UV CD spectra of J2-crystallin in Figure 6.4 shows minimum ellipticities at 211 nm and 218 nm, indicative of a protein with α-helical content, a result fairly uncommon for eye lens crystallins [153].

The thermostabilty of J2-crystallin was determined by DSF. The SYPRO Orange fluorescent probe was used to monitor the changes in intensity over the temperature range from 25 to 95°C [151]. The probe binds to hydrophobic pockets of the protein and as the temperature increases, the protein unfolds resulting in the emission intensity of the probe to increase.
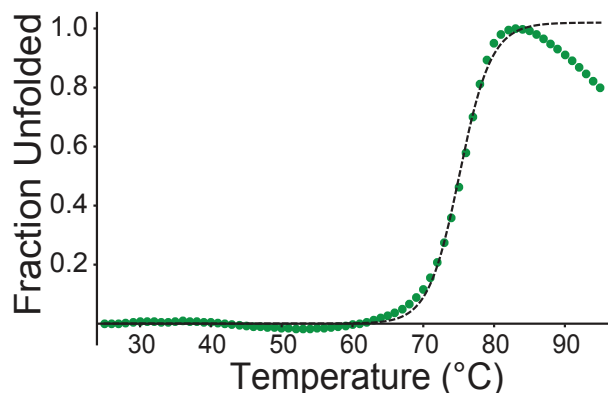
Figure 6.5: Thermal unfolding curve of J2-crystallin using DSF with best-fit unfolding curves. The Tm-value was found to be 75.2 °C. The excitation and emission wavelengths were 492 nm and 610 nm, respectively. Fluorescence readings were recorded from 25-95 °C with a temperature ramp of 1 °C/min and recorded in duplicate.The data were fit by a nonlinear regression analysis using the Boltzmann function and unfolding temperature values were determined by the inflection point of the resulting sigmoidal curves defined from 25-87 °C

For J2-crystallin, the resulting changes in fluorescence yields a sigmoidal curve signifying a two-state transition where the midpoint determines the melting temperature. The unfolding temperature (Tm) for J2-crystallin was found to be 75.2 ± 0.1 °C as shown in Figure 6.5.

To qualitatively examine exposed surface hydrophobic patches, structural changes in J2-crystallin must be observed by use of a hydrophobic probe. Using SYPRO Tangerine dye in conjunction with fluorescence spectroscopy, I collected emission spectra of J2-crystallin prepared at 25, 50, and 75 °C, as plotted in Figure 6.6. At 75 °C, J2-crystallin exhibits the most exposed surface hydrophobic patches while at 50 °C it is 5% less hydrophobic, and approximately 17% less hydrophobic at 25 °C.

6.3.3 Pursuing structural studies for J2-crystallin

To assess the suitability of J2-crystallin for structural studies, a ^{1}H–^{15}N heteronuclear single-quantum correlation experiment (HSQC) was collected using samples of ^{15}N-labelled J2-crystallin. The first of many solution-state NMR experiments in a structure determination
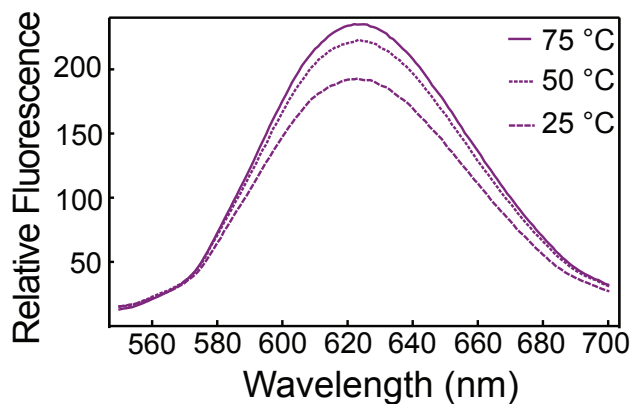
Figure 6.6: Emission spectra of J2-crystallin at various temperatures. Samples were prepared at 0.95 mg/mL and allowed to incubate with SYPRO Tangerine dye (1% v/v) at various temperatures for 5 minutes. Samples were excited at 492 nm and emission was recorded from 550 to 700 nm. J2-crystallin exhibits the most exposed surface hydrophobic patches at 75 °C and is 5% less hydrophobic at 50 °C followed by approximately 17% less hydrophobic at 25 °C.

effort, this experiment determines how well-folded the protein is by the distribution of cross peaks. The cross peaks are representative of the amide N-H pairs of the protein backbone and sidechains. Clean and separated cross peaks allow for distinction among the 158 amino acid residues that should be seen in the spectrum. An HSQC also can indicate the signal-to-noise ratio at the particular pH and concentration of the sample. The right pH and concentration will allow for further 3D experiments that can take longer and are less sensitive. The spectrum of J2-crystallin, Figure 6.7 shows well-separated cross peaks in both spectral dimensions, indicating that the protein is folded and monomeric.

6.4 Conclusion

Along with the ^{1}H–^{15}N HSQC experiment performed earlier, further characterization of J2-crystallin will be done to ultimately solve the structure. As this is a structural protein, the structure is of paramount importance in determining its function. J2-crystallin is a novel eye lens protein that has been expressed recombinantly and purified in high yield. In agreement with the secondary structure prediction and CD spectroscopy, J2-crystallin has primarily
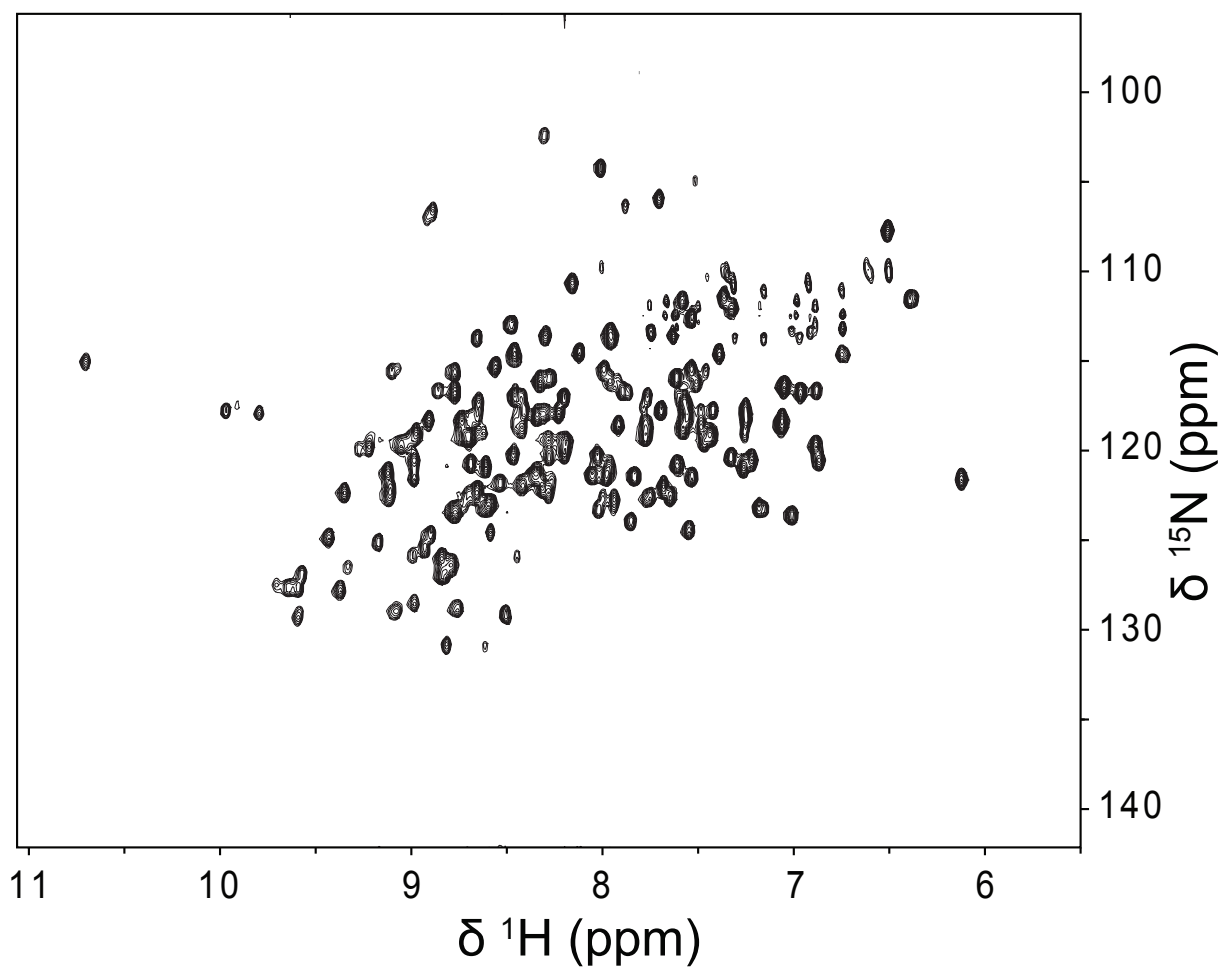
Figure 6.7: ^{1}H–^{15}N HSQC spectrum of ^{15}N-labelled J2-crystallin acquired at 25 °C, indicating that the protein is folded and monomeric. The crystallin sample was prepared in 10 % D_2O and 2 mM TMSP at a final concentration of 1.8 mM.

α-helical character and thermal studies have concluded a melting temperature of 75.2 °C. A ^{1}H–^{15}N HSQC shows the protein is well-folded, allowing for further NMR experiments to be performed.

Bibliography

[1] Nathalie Braun, Martin Zacharias, Jirka Peschek, Andreas Kastenmueller, Juan Zou, Marianne Hanzlik, Martin Haslbeck, Juri Rappsilber, Johannes Buchner, and Sevil Weinkauf. Multiple molecular architectures of the eye lens chaperone αB-crystallin elucidated by a triple hybrid approach. *Proceedings of the National Academy of Sciences of the United States of America*, 108(51):20491–20496, 2011.

[2] V Nalini, B Bax, H Driessen, DS Moss, PF Lindley, and C Slingsby. Close packing of an oligomeric eye lens beta-crystallin induces loss of symmetry and ordering of sequence extensions. *Journal of molecular biology*, 236(4):1250—1258, March 1994.

[3] Carolyn N. Kingsley, William D. Brubaker, Stefan Markovic, Anne Diehl, Amanda J. Brindley, Hartmut Oschkinat, and R. W. Martin. Preferential, specific binding of human αB-crystallin to a cataract-related variant of γS-crystallin. *Structure*, 21:2221—2227, 2013.

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

[5] M. F. Sanner, A. Olsen, and J. Spehner. Fast and robust computation of molecular surfaces. In *Proceedings of the 11th ACM symposium on computational geometry*, pages C6—C7, 1995.

[6] Nobuo Niimura, Yoshiaki Minezaki, Takamasa Nonaka, Jean-Charles Castagna, Florent Cipriani, Peter H⊘gh⊘j, Mogens S. Lehmann, and Clive Wilinson. Neutron laue diffractometry with an imaging plate provides an effective data collection regime for neutron protein crystallography. *Nature Structural & Molecular Biology*, 4(11):909—914, 1997.

[7] Cécile Bon, Mogens S. Lehmann, and Clive Wilkinson. Quasi-Laue neutron-diffraction study of the water arrangement in crystals of triclinic hen egg-white lysozyme. *Acta Crystallographica Section D*, 55(5):978—987, May 1999.

[8] Harald Schwalbe, Shaun B. Grimshaw, Andrew Spencer, Matthias Buck, Jonathan Boyd, Christopher M. Dobson, Christina Redfield, and Lorna J. Smith. A refined solution structure of hen lysozyme determined using residual dipolar coupling data. *Protein Science*, 10(4):677—688, 2001.

[9] D. Jones. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*, 292:195—202, 1999.

[10] Abraham Savitsky and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8):1627—1639, 1964.

[11] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33—38, 1996.

[12] Karuna Dixit, Ajay Pande, Jayanti Pande, and Siddhartha P. Sarma. Nuclear magnetic resonance structure of a major lens protein, human γc-crystallin: Role of the dipole moment in protein stability. *Biochemistry*, 55:3136—3149, 2016.

[13] Kate L. Moreau and Jonathan A. King. Protein misfolding and aggregation in cataract disease and prospects for prevention. *Trends in Molecular Medicine*, 18(5):273—282, 2012.

[14] A. Foster and S. Resnikoff. The impact of vision 2020 on global blindness. *Eye*, 19:1133—1135, 2005.

[15] A Shiels and J F Hejtmancik. Genetics of human cataract. *Clinical Genetics*, 84(2):120—127, 2013.

[16] NEI. National eye institute - facts about cataract, September 2015.

[17] A. Shiels and J. Hejtmancik. Genetic origins of cataract. *Archives of Ophthalmology*, 125(2):165—173, 02 2007.

[18] Colin E Willoughby, Diego Ponzin, Stefano Ferrari, Aires Lobo, Klara Landau, and Yadollah Omidi. Anatomy and physiology of the human eye: effects of mucopolysaccharidoses disease on structure and function – a review. *Clinical & Experimental Ophthalmology*, 38:2—11, 2010.

[19] Pablo Artal, Antonio Benito, and Juan Tabernero. The human eye is an example of robust optical design. *Journal of Vision*, 6(1):1—7, 2006.

[20] K. Mahendiran, C. Elie, J.-C. Nebel, A. Ryan, and B. K. Pierscionek. Primary sequence contribution to the optical fuction of the eye lens. *Scientific Reports*, 4:5195, 2014.

[21] Z. Ma, G. Piszczek, P.T. Wingfield, Y.V. Sergeev, and J.F. Hejtmancik. The G18V CRYGS mutation associated with human cataracts increases γs-crystallin sensitivity to thermal and chemical stress. *Biochemistry*, 48:7334–7341, 2009.

[22] E. Serebryany and J. A. King. Human γd-crystallin promotes aggregation of its oxidation-mimicking, misfolding-prone w42q mutant. *Journal of Biological Chemistry*, 290:11491—11503, 2015.

[23] M. A. DiMauro, S. K. Nandi, C. T. Raghavan, R. K. Kar, B. Wang, A. Bhuania, R. H. Nagaraj, and A. Biswas. Acetylation of gly1 and lys2 promotes aggregation of human γd-crystallin. *Biochemistry*, 53:7269—7282, 2014.

[24] M. Bloemendal and A. Toumadje. Bovine lens crystallins do contain helical structure. *Biochimica et Biophysica Acta*, 1432:234—238, 1999.

[25] R. Kannan, P. Sreekumar, and D. Hinton. Novel roles for α-crystallins in retinal function and disease. *Progress in Retinal and Eye Research*, 31:576—604, 2012.

[26] Yellamaraju Sreelakshmi, Puttur Santhoshkumar, Jaya Bhattacharyya, and K Krishna Sharma. αA-Crystallin Interacting Regions in the Small Heat Shock Protein, αB-Crystallin . *Biochemistry*, 43(50):15785—15795, dec 2004.

[27] N. J. Ray, D. Hall, and J. A. Carver. Deamidation of n76 in human γs-crystallin promotes dimer formation. *Biochimica et Biophysica Acta*, 1860:315—324, 2016.

[28] Rainer Jaenicke and Christine Slingsby. Lens crystallins and their microbial homologs: Structure, stability, and function. *Critical Reviews in Biochemistry and Molecular Biology*, 36(5):435—499, 2001.

[29] H. Zhao, M. T. Magone, and P. Schuck. The role of macromolecular crowding in the evolution of lens crystallins with high molecular refractive index. *Physical Biology*, 8:046004, 2011.

[30] H. J. Aarts, N. H. Lubsen, and J. G. Schoenmakers. Crystallin gene expression during rat lens development. *European Journal of Biochemistry*, 183:31–36, 1989.

[31] C. N. Nagineni and S. P. Bhat. Lens fiber cell differentiation and expression of crystallins in co-cultures of human fetal lens epithelial cells and fibroblasts. *Experimental Eye Research*, 54:193–200, 1992.

[32] Venkata Pulla Rao Vendra, Sushil Chandani, and Dorairajan Balasubramanian. The mutation v42m distorts the compact packing of the human γs-crystallin molecule, resulting in congenital cataract. *PLOS ONE*, 7(12):1—9, 12 2012.

[33] S. Karri, R. B. Kasetti, V. P. R. Vendra, S. Chandani, and D. Balasubramanian. Structural analysis of the mutant protein D26G of human γS-crystallin, associated with coppock caraeact. *Molecular Vision*, 19:1231—1237, 2013.

[34] Venkata Pulla Rao Vendra, Ismail Khan, Sushil Chandani, Anbukkarasi Muniyandi, and Dorairajan Balasubramanian. Gamma crystallins of the human eye lens. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1860(1, Part B):333 —343, 2016. SI: Crystallin Biochemistry.

[35] Hongping Ye. Simultaneous determination of protein aggregation, degradation, and absolute molecular weight by size exclusion chromatography–multiangle laser light scattering. *Analytical Biochemistry*, 356(1):76 —85, 2006.

[36] Anthony L. Fink. Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Folding and Design*, 3(1):R9—R23, 1998.

[37] D Khago, E K Wong, C N Kingsley, J Alfredo Freites, D J Tobias, and R W Martin. Increased hydrophobic surface exposure in the cataract-related G18V variant of human gammaS-crystallin. *Biochim Biophys Acta*, 1860(1):325–332, 2015.

[38] Yingfang Ma, Diana M. Acosta, Jon R. Whitney, Rudolf Podgornik, Nicole F. Steinmetz, Roger H. French, and V. Adrian Parsegian. Determination of the second virial coefficient of bovine serum albumin under varying ph and ionic strength by composition-gradient multi-angle static light scattering. *Journal of Biological Physics*, 41(1):85—97, 2015.

[39] Mu Yang, Colina Dutta, and Ashutosh Tiwari. Disulfide-bond scrambling promotes amorphous aggregates in lysozyme and bovine serum albumin. *The Journal of Physical Chemistry B*, 119(10):3969—3981, 03 2015.

[40] A. Pande, N. Mokhor, and J. Pande. Deamidation of human γs-crystallin increases protein interactions: Implications for cataract. *Biochemistry*, 54:4890—4899, 2015.

[41] C. Slingsby and Graeme J. Wistow. Functions of crystallins in and out of lens: Roles in elongated and post-mitotic cells. *Progress in Biophysics and Molecular Biology*, 115:52—67, 2014.

[42] Z. Kozmik, K. Shivalingappa, J. Ruzickova, K. Jonasova, V. Paces, C. Vlcek, and J. Piatigorsky. Cubazoan crystallins: evidence for convergent evolution of pax regulatory sequences. *Evolution & Development*, 10:52—61, 2008.

[43] F. Yin, Y. Chen, C. Yu, Y. Pon, and H. Lee. Kinetic refolding barrier of guanidinium chloride denatured goose δ-crystallin leads to regular aggregate formation. *Biophysical Journal*, 93:1235—1245, 2007.

[44] H. Zhao, P. H. Brown, and P. Schuck. On the distribution of protein refractive index increments. *Biophysical Journal*, 100:2309–2317, 2011.

[45] Huaying Zhao, Yingwei Chen, Lenka Rezabkova, Zhengrong Wu, Graeme Wistow, and Peter Schuck. Solution properties of -crystallins: Hydration of fish and mammal -crystallins. *Protein Science*, 23(1):88—99, 2014.

[46] J. Horwitz. Alpha-crystallin. *Experimental Eye Research*, 76:145—153, 2003.

[47] H. Zhao, P. H. Brown, M. T. Magone, and P. Schuck. The molecular refractive function of lens γ-crystallins. *Journal of Molecular Biology*, 411(3):680–699, 2011.

[48] Andor J. Kiss, Arthur L. Devries, and Rachael M. Morgan-Kiss. Comparative analysis of crystallins and lipids from the lens of antarctic toothfish and cow. *Journal of Comparative Physiology B*, 180(7):1019–1032, 2010.

[49] Yingwei Chen, Huaying Zhao, Peter Schuck, and Graeme Wistow. Solution properties of -crystallins: Compact structure and low frictional ratio are conserved properties of diverse -crystallins. *Protein Science*, 23(1):76—87, 2014.

[50] Andor J. Kiss, Amir Y. Mirarefi, Subramanian Ramakrishnan, Charles F. Zukoski, Arthur L. DeVries, and Chi-Hing C. Cheng. Cold-stable eye lens crystallins of the antarctic nototheniid toothfish *Dissostichus mawsoni* norman. *Journal of Experimental Biology*, 207:4633—4649, 2004.

[51] Bryon Mahler, Yingwei Chen, Jason Ford, Caleb Thiel, Graeme Wistow, and Zhengrong Wu. Structure and dynamics of the fish eye lens protein, γM7-crystallin. *Biochemistry*, 52(20):3579–3587, 2013.

[52] H. Sun, Z. Ma, Y. Li, B. Liu, Z. Li, X. Ding, Y. Gao, W. Ma, X. Tang, X. Li, and Y. Shen. γs-crystallin gene (crygs) mutation causes dominant progressive cortical cataract in humans. *Journal of Medical Genetics*, 42:706—710, 2005.

[53] Stanislav I. Tomarev, Sambath Chung, and Joram Piatigorsky. Glutathione s-transferase and s-crystallins of cephalopods: Evolution from active enzyme to lens-refractive proteins. *Journal of Molecular Evolution*, 41:1048—1056, 1995.

[54] John V. Champion, Gerald H. Meeten, and Malcom State. Refractive index of particles in the colloidal state. *Journal of the Chemical Society, Faraday Transactions 2*, 74:1319—1329, 1978.

[55] Tathyana Tumolo, Lucio Angnes, and Mauricio S. Baptista. Determination of the refractive index increment [dn/dc] of molecule and macromolecule solutions by surface plasmon resonance. *Analytical Biochemistry*, 333:273—279, 2004.

[56] M. B. Huglin. Specific refractive index increments of polymer solutions. part i. literature values. *Journal of Applied Polymer Science*, 9:3963—4001, 1965.

[57] Gilbert Smithson Adair and Muriel Elaine Robinson. The specific refraction increments of serum-albumin and serum-globulin. *Biochemical Journal*, 24(4):993—1011, 1930.

[58] R. Barer and S. Joseph. Refractometry of living cells. *The Quarterly Journal of Microscopical Science*, 95:399—423, 1954.

[59] Larry J. W. Miercke, A. Robbins, Rebecca, and Robert M. Stroud. Tetra detector analysis of membrane proteins. *Current Protocols in Protein Science*, 29.10:1—30, 2014.

[60] C.-Y. Tan and Y.-X. Huag. Dependence of refractive index on concentrations and temperature in electrolyte solution, polar solution, nonpolar solution, and protein solutions. *Journal of Chemical & Engineering Data*, 60:2827—2833, 2015.

[61] Carolyn N. Kingsley, Jan C. Bierma, Vyvy Pham, and Rachel W. Martin. γs-crystallin proteins from the antartic nototheniid toothfish: A model system for investigating differential resistance to chemical and thermal denaturation. *The Journal of Physical Chemistry B*, 118(47):13544—13553, 2014.

[62] Natalia Kozlyuk, Suvrajit Sengupta, Jan C. Bierma, and Rachel. W. Martin. Calcium binding dramatically stabilizes an ancestral crystallin fold in tunicate $\beta\gamma$-crystallin. Submitted, 2016.

[63] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, and A. Bairoch. *The Proteomics Protocols Handbook*, chapter Protein identification and analysis tools on the ExPASy Server. Humana Press, 2005.

[64] Srivatsan Raman, Robert Vernon, James Thompson, Michael Tyka, Ruslan Sadreyev, Jimin Pei, David Kim, Elizabeth Kellogg, Frank DiMaio, Oliver Lange, Lisa Kinch, Will Sheffler, Bong-Hyun Kim, Rhiju Das, Nick V. Grishin, and David Baker. Structure prediction for casp8 with all-atom refinement using rosetta. *Proteins: Structure, Function, and Bioinformatics*, 77(S9):89—99, 2009.

[65] Yifan Song, Frank DiMaio, Ray Yu-Ruei Wang, David Kim, Chris Miles, TJ Brunette, James Thompson, and David Baker. High-resolution comparative modeling with rosettacm. *Structure*, 21(10):1735—1742, 2013.

[66] John C. Gordon, Jonathan B. Myers, Timothy Folta, Valia Shoja, Lenwood S. Heath, and Alexey Onufriev. H++: a server for estimating pkas and adding missing hydrogens to macromolecules. *Nucleic Acids Research*, 33(suppl 2):W368—W371, 07 2005.

[67] Jonathan Myers, Greg Grothaus, Shivaram Narayanan, and Alexey Onufriev. A simple clustering algorithm can be accurate enough for use in calculations of pks in macromolecules. *Proteins: Structure, Function, and Bioinformatics*, 63(4):928—938, 2006.

[68] Ramu Anandakrishnan, Boris Aguilar, and Alexey V. Onufriev. H++ 3.0: automating pk prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Research*, 40(W1):W537—W541, 07 2012.

[69] Axel Kohlmeyer. Dipole watcher plugin, version 1.0.

[70] Thomas Gibaud, Frederic Cardinaux, Johan Bergenholtz, Anna Stradner, and Peter Schurtenberger. Phase separation and dynamical arrest for particles interacting with mixed potentials-the case of globular proteins revisited. *Soft Matter*, 7:857–860, 2011.

[71] Monique Laberge. Intrinsic protein electric fields: basic non-covalent interactions and relatioship to protein-induce stark effects. *Biochimica et Biophysica Acta*, 1386:305—330, 1998.

[72] Gregory D. Scholes, Xanthipe J. Jordanides, and Graham R. Fleming. Adapting the förster theory of energy transfer for modeling dynamics in aggregated molecular assemblies. *Journal of Physical Chemistry B*, 105:1640—1651, 2001.

[73] Franck Baud and Samuel Karlin. Measures of residue density in protein structures. *Proceedings of the National Academy of Sciences of the United States of America*, 96:12494—12499, 1999.

[74] Kim A. Sharp and Barry Honig. Electrostatic interactions in macromolecules: theory and applications. *Annual Review of Biophysics and Biophysical Chemistry*, 19:301—332, 1990.

[75] Amit Kessel and Nir Ben-Tal. Introduction to proteins: structure, functions, and motions, 2011.

[76] Clifford E. Felder, Jaime Prilusky, Israel Silman, and Joel. L. Sussman. A server and database for dipole moments of proteins. *Nucleic Acids Research*, 35:W512—W521, 2007.

[77] I. A. Mills, S. L. Flaugh, M. S. Kosinski-Collins, and J. A. King. Folding and stability of the isolated Greek key domains of the long-lived human lens proteins γD-crystallin and γS-crystallin. *Protein Science*, 16(11):2427–2444, 2007.

[78] Ajay Pande, Kaylan S. Ghosh, Priya R. Banjeree, and Jayanti Pande. Increase in surface hydrophobicity of the cataract-associated P23T mutant of human γD-crystallin is responsible for its dramatically lower retrograde solubility. *Biochemistry*, 49:6122—6129, 2010.

[79] Y. Liu, X. Zhang, L. Luo, M. Wu, R. Zeng, G. Cheng, B. Hu, B. Liu, J.J. Liang, and F. Shang. A novel αB-crystallin mutation associated with autosomal dominant congenital lamellar cataract. *Investigative Ophthalmology & Visual Science*, 47(3):1069—1075, 2006.

[80] Priya R. Banerjee, Shadakshara S. Puttamadappa, Ajay Pande, Alexander Shekhtman, and Jayanti Pande. Increased hydrophobicity and decreased backbone flexibility explain the lower solubility of a cataract-linked mutant of γD-crystallin. *Journal of Molecular Biology*, 412(4):647–659, 2011.

[81] S. V. Bharat, A. Shekhtman, and J. Pande. The cataract-associated V41M mutant of human γS-crystallin shows specific structural changes that directly enhance local surface hydrophobicity. *Biochemical and Biophysical Research Communications*, 443:110—114, 2014.

[82] Daumantas Matulis and Rex Lovrien. 1-anilino-8-naphthalene sulfonate anion-protein binding depends primarily on ion pair formation. *Biophysical Journal*, 74(1):422–429, 1998.

[83] J J Ory and L J Banaszak. Studies of the ligand binding reaction of adipocyte lipid binding protein using the fluorescent probe 1, 8-anilinonaphthalene-8-sulfonate. *Biophysical Journal*, 77(2):1107–1116, 1999.

[84] D. Sheluho and S. H. Ackerman. An accessible hydrophobic surface is a key element of the molecular chaperone action of atp11p. *Journal of Biological Chemistry*, 43:39945—39949, 2001.

[85] C. Münch and A. Bertolotti. Exposure of hydrophobic surfaces initiates aggregation of diverse als-causing superoxde dismutase-1 mutants. *Journal of Molecular Biology*, 399:512—525, 2010.

[86] A. Tiwari, A. Liba, S. H. Sohn, S. V. Seetharanab, O. Bilsel, C. R. Matthews, P. J. Hart, J. S. Valentine, and L. J. Hayward. Metal deficiency increases aberrant hydrophobicity of mutant superoxide dismutases that cause amyotophic lateral sclerosis. *Journal of Biological Chemistry*, 284(27746—27758), 2009.

[87] M. P. Williamson. Using chemical shift perturbation to characterize ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 73:1–16, 2013.

[88] S. B. Shuker, P. J. Hajduk, R. P. Meadows, and SW Fesik. Discovering high affinity ligands for proteins: Sar by nmr. *Science*, 274:1531—1534, 1996.

[89] T. ten Brink, C. Aguirre, T. E. Exner, and I. Krimm. Performance of protein-ligand docking with simulated chemical shift perturbations. *Journal of Chemical Information and Modeling*, 55:275—283, 2015.

[90] S. Huang and X. Zou. Efficient molecular docking of nmr structures: Application to hiv-1 protease. *Protein Science*, 16:43—51, 2007.

[91] Jaime L. Stark and Robert Powers. Application of nmr and molecular docking in structure-based drug discovery. *Topics in Current Chemistry*, 326:1—34, 2011.

[92] A. J. van Dijk, R. Boelens, and A. M. Bonvin. Data-driven docking for the study of biomolecular complexes. *FEBS Journal*, 272:293—312, 2005.

[93] Csaba Hetényi and David van der Spoel. Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein science : a publication of the Protein Society*, 11(7):1729–1737, 2002.

[94] Bogdan Iorga, Denyse Herlem, Elvina Barré, and Catherine Guillou. Acetylcholine nicotinic receptors: Finding the putative binding site of allosteric modulators using the "blind docking" approach. *Journal of Molecular Modeling*, 12(3):366–372, 2006.

[95] T. Konuma, Y.-H. Lee, Y. Goto, and K. Sakurai. Principal component analysis of chemical shift perturbation data of a multiple-ligand-binding system for elucidation of respective binding mechanism. *Proteins*, 81:107–118, 2013.

[96] Md Zahid Kamal, Jamshaid Ali, and Nalam Madhusudhana Rao. Binding of bis-ANS to Bacillus subtilis lipase: a combined computational and experimental investigation. *Biochimica et biophysica acta*, 1834(8):1501—1509, August 2013.

[97] G. Weber and L. B. Young. Fragmentation of bovine serum albumin by pepsin. *The Journal of Biological Chemistry*, 239(5):1415—1423, 1964.

[98] A. J. Shaka, P. B. Barker, and R. J. Freeman. Computer-optimized decoupling scheme for wideband applications and low-level operation. *Journal of Magnetic Resonance*, 64(547—552), 1985.

[99] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfiefer, and A. Bax. Nmrpipe: A multidimensional spectral processing system based on unix pipes. *Journal of Biomolecular NMR*, 6:277—293, 1995.

[100] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pjon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, and E. D. Laue. The ccpn data model for nmr spectroscopy: Development of a software pipeline. *Proteins*, 59(4):687—696, 2005.

[101] W. D. Brubaker and R. W. Martin. ^{1}H, ^{13}C, and ^{15}N assignments of wild-type human γS-crystallin and its cataract-related variant γS-G18V. *Biomolec. NMR Assign.*, 6:63–67, 2011.

[102] Garrett M. Morris, Ruth Huey, William Lindstrom, Michel F. Sanner, Richard K. Belew, David S. Goodsell, and Arthur J. Olson. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 30(16):2785–2791, December 2009.

[103] S. L. Flaugh, M. S. Kosinski-Collins, and J. King. Interdomain side-chain interactions in human γD-crystallin influencing folding and stability. *Protein Science*, 14:2030—2043, 2005.

[104] B. G. Mohr, C. M. Dobson, S. C. Garman, and M. Muthukumar. Electrostatic origin of in vitro aggregation of human gamma-crystallin. *Journal of Chemical Physics*, 139:121914, 2013.

[105] Oleg Trott and Arthur J. Olson. Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 2010.

[106] M. S. Ajaz, Z. Ma, D. L. Smith, and J. B. Smith. Size of human lens β-crystallin aggregates are distinguished by n-terminal truncation of βb1. *The Journal of Biological Chemistry*, 272(17):11250—11255, 1997.

[107] P. J. L. Werten, R. A. Lindner, J. A. Carver, and W. W. de Jong. Formation of beta a3/beta b2-crystallin mixed complexes: involvement of n- and c-terminal extensions. *Biochimica et biophysica acta-Protein Structure and Molecular Enzymology*, 1432(2):286—292, Jul 1999.

[108] K. J. Lampi, J. T. Oxford, H. P. Bachinger, T. R. Shearer, L. L. David, and D. M. Kapfer. Deamidation of human βb1 alters the elongated structure of the dimer. *Experimental Eye Research*, 72:279—288, 2001.

[109] M. B. Dolinska, Y. V. Sergeev, M. P. Chan, I. Palmer, and P. T. Wingfield. The n-terminal extension of βb1-crystallin: identification of a critical region which modulates protein interactions with βa3-crystallin. *Biochemistry*, 48(40):9684—9695, 2009.

[110] O. Annunziata, A. Pande, J. Pande, O. Ogun, N. H. Lubsen, and G. B. Benedek. Oligomerization and phase transitions in aqueous solutions of native and truncated human βb1-crystallin. *Biochemistry*, 44:1136—1328, 2005.

[111] X. Leng, S. Wang, N. Cao, L. Qi, and Y. Yan. The n-terminal extension of βb1-crystallin chaperones β-crystallin folding and cooperates with αa-crystallin. *Biochemistry*, 53:2464—2473, 2014.

[112] Fangling Ji, Jinwon Jung, Leonardus M. I. Koharudin, and Angela M. Gronenborn. The human W42R γD-crystallin mutant structure provides a link between congenital and age-related cataracts. *Journal of Biological Chemistry*, 288(1):99–109, 2013.

[113] Fangling Ji, Leonardus M. I. Koharudin, Jinwon Jung, and Angela M. Gronenborn. Crystal structure of the cataract-causing P23T γD-crystallin mutant. *Proteins - Structure, Function and Bioinformatics*, 81(9):1493–1498, 2013.

[114] Fangling Ji, Jinwon Jung, and Angela M. Gronenborn. Structural and biochemical characterization of the childhood cataract-associated R76S mutant of human γD-crystallin. *Biochemistry*, 51(12):2588–2596, 2012.

[115] Jinbu Wang, Xiaobing Zuo, Ping Yu, In-Ja L. Byeon, Jinwon Jung, Xiaoxia Wang, Marzena Dyba, Soenke Seifert, Charles D. Schwieters, Jun Qin, Angela M. Gronenborn, and Yun-Xing Wang. Determination of multicomponent protein structures in solution using global orientation and shape restraints. *Journal of the American Chemical Society*, 131(30):10507—10515, 2009.

[116] W. Khan, F. Duffy, G. Pollastri, D. Shields, and C. Mooney. Predicting binding within disordered protein regions to structurally characterised peptide-binding domains. *PLOS ONE*, 8:e72838, 2013.

[117] P. G. Cooper, J. A. Carver, J. A. Aquilina, G. B. Ralston, and R. J. Truscott. A ^{1}H NMR spectroscopic comparison of γS- and γB-crystallins. *Exp. Eye Res.*, 59:211–220, 1994.

[118] Ronald Frank. The SPOT-synthesis technique. *Journal of Immunological Methods*, 267(1):13–26, sep 2002.

[119] F Molina, D Laune, C Gougat, B Pau, and C Granier. Improved performances of spot multiple peptide synthesis. *Peptide research*, 9(3):151—155, 1996.

[120] Armin a Weiser, Michal Or-Guil, Victor Tapia, Astrid Leichsenring, Johannes Schuchhardt, Cornelius Frömmel, and Rudolf Volkmer-Engert. SPOT synthesis: reliability of array-based measurement of peptide binding affinity. *Analytical biochemistry*, 342(2):300–11, jul 2005.

[121] Maria Rodriguez, Xiaochun Yu, Junjie Chen, and Zhou Songyang. Phosphopeptide Binding Specificities of BRCA1 COOH-terminal (BRCT) Domains. *Journal of Biological Chemistry*, 278(52):52914–52918, 2003.

[122] Huadong Liu, Lei Li, Courtney Voss, Feng Wang, Juewen Liu, and Shawn S-C Li. A comprehensive immunoreceptor phosphotyrosine-based signaling network revealed by reciprocal protein-peptide array screening. *Molecular & cellular proteomics : MCP*, 14:1846–1858, 2015.

[123] B. M. Gardner and P. Walter. Unfolded Proteins Are Ire1-Activating Ligands That Directly Induce the Unfolded Protein Response. *Science*, 333(6051):1891–1894, sep 2011.

[124] L. Goldschmidt, P. K. Teng, R. Riek, and D. Eisenberg. Identifying the amylome, proteins capable of forming amyloid-like fibrils. *Proceedings of the National Academy of Sciences of the United States of America*, 107:3487—3492, 2010.

[125] P. K. Teng and D. Eisenberg. Short protein segments can drive a non-fibrilizing protein into the amyloid state. *Protein Engineering, Design & Selection*, 22:531—536, 2009.

[126] L. Salwinksi and D. Eisenberg. Computational methods of analysis of protein-protein interactions. *Current Opinion in Structural Biology*, 13:377—382, 2003.

[127] Arthur Laganowsky, Justin L P Benesch, Meytal Landau, Linlin Ding, Michael R Sawaya, Duilio Cascio, Qingling Huang, Carol V Robinson, Joseph Horwitz, and David Eisenberg. Crystal structures of truncated alphaA and alphaB crystallins reveal structural mechanisms of polydispersity important for eye lens function. *Protein Science*, 19(5):1031–1043, mar 2010.

[128] J. S. Nowick, D. M. Chung, K. Maitra, S. Maitra, K. D. Stigers, and Y. Sun. An unnatural amino acid that mimics a tripetide β-sheetlike hydrogen-bonded dimers. *J. Am. Chem. Soc.*, 1222:7654—7661, 2000.

[129] K. Hilpert, D. F. Winkler, and R. E. Hancock. Peptide arrays on cellulose support: Spot synthesis, a time and cost efficient method for synthesis of large numbers of peptides in a parallel and addressable fashion. *Nature Protocols*, 2:1333—1349, 2007.

[130] Carolyn N. Kingsley. *Investigating Protein-Protein Interactions Between Eye Lens Crystallins*. PhD thesis, University of California, Irvine, 2013.

[131] Amie M. Morris, Teresa M. Treweek, J. .. . A. .. Aquilina, John A. Carver, and Mark J. Walker. Glutamic acid residues in the c-terminal extension of small heat shock protein 25 are critical for structural and functional integrity. *FEBS JOURNAL*, 275(23):5885–5898, Dec 2008.

[132] J. Gasteiger and M. Marsili. A New Model for Calculating Atomic Charges in Molecules. *Tetrahedron Lett.*, 34:3181–3184, 1978.

[133] Domarin Khago, Carolyn N. Kingsley, Kyle W. Roskamp, Eric K. Wong, J. Alfredo Freites, Douglas J. Tobias, and Rachel W. Martin. Probing peptide-protein interactions in human γs-crystallin. Manuscript in preparation, 2016.

[134] Diwakar Shukla, Curtiss P. Schneider, and Bernhardt L. Trout. Molecular level insight into intra-solvent interaction effects on protein stability and aggregation. *Advanced Drug Delivery Reviews*, 63(13):1074 —1085, 2011. Formulating biomolecules: mechanistics insights in molecular interactions.

[135] A Quigley and D.R. Williams. The second virial coefficient as a predictor of protein aggregation propensity: A self-interaction chromatography study. *European Journal of Pharmaceutics and Biopharmaceutics*, 96:282—290, 2015.

[136] V. Prytkova, M. Heyden, D. Khago, J. A. Freites, C. Butts, R. W. Martin, and D. J. Tobias. Multi-conformation monte carlo: a method for introducing flexibility in efficient simulations of many-protein systems. *Journal of Physical Chemistry B (submitted)*, 2016.

[137] Marco A. Blanco, Erinc Sahin, Yi Li, and Christopher J. Roberts. Reexamining protein-protein and protein-solvent interactions from kirkwood-buff analysis of light scattering in multi-component solutions. *The Journal of Chemical Physics*, 134:225103, 2011.

[138] Philip J. Wyatt. Light scattering and the absolute characterization of macromolecules. *Analytica Chimica Acta*, 272(1):1—40, 1993.

[139] W. William Wilson. Light scattering as a diagnostic for protein crystal growth—a practical approach. *Journal of Structural Biology*, 142:56—65, 2003.

[140] Xuan-Yu Meng, Yu Xu, Hong-Xing Zhang, Mihaly Mezei, and Meng Cui. Predicting protein interaction by brownian dynamics simulations. *Journal of Biomedicine and Biotechnology*, 2012:11, 2012.

[141] Št ěpán Růžička and Michael P. Allen. Collective translational and rotational monte carlo cluster move for general pairwise interactions. *Physical Review E*, 90(3), 2014.

[142] Carter T. Butts, Domarin Khago, and Rachel W. Martin. Bayesian analysis of static light scattering data for globular proteins. In Preperation, 2016.

[143] Johannes Möller, Martin A. Schroer, Mirko Erlkamp, Sebastian Grobelny, Michael Paulus, Sebastian Tiemeyer, Florian J. Wirkert, Metin Tolan, and Roland Winter. The effect of ionic strength, temperature, and pressure on the interaction potential of dense protein solutions: From nonlinear pressure response to protein crystallization. *Biophysical Journal*, 102(11):2641—2648, 6 2012.

[144] T.D. Ingolia and E.A. Craig. Four small drosophila heat shock proteins are related to each other and to mammalian alpha-crystallins. *Proceedings of the National Academy of Sciences of the United States of America*, 79:2360–2364, 1982.

[145] W. W. de Jong, J. A. M. Leunissen, and C. E. M. Voorter. Evolution of the alpha-crystallin/small heat shock protein family. *Molecular Biology and Evolution*, 10:103–126, 1993.

[146] J. Piatigorsky and Z. Kozmik. Cubozoan jellyfish: An evo/devo model for eyes and other sensory systems. *International Journal of Developmetnal Biology*, 48:719—729, 2004.

[147] Z. Kozmik, J. Ruzickova, K. Jonasova, Y. Matsumoto, P. Vopalensky, I. Kozmikoza, H. Strnad, S. Kawamura, J. Piatigorsky, V. Paces, and C. Vlcek. Assembly of the cnidarian camera-type eye from vertebrate-like components. *Proceedings of the National Academy of Sciences of the United States of America*, 105:8989—8993, 2008.

[148] J. Piatigorsky, J. Horwitz, and B. Norman. J1-crystallins of the cubomedusan jellyfish lens constitute a novel family encoded in at least three intronless genes. *Journal of Biological Chemistry*, 268:11894—11901, 1993.

[149] J. Piatigorsky, B. Norman, L. Dishaw, L. Kos, J. Horwitz, P. Steinbach, and Z. Kozmik. J3-crystallin of the jellyfish lens: Similarity to saposins. *Proceedings of the National Academy of Sciences of the United States of America*, 28:12362—12367, 2001.

[150] S. Altschul, A. Madden, T. Madden, A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. Lipman. Gapped blast and psi-blast;a new generation of prtein database search programs. *Nucleic Acids Research*, 25:3389—3402, 1997.

[151] Frank H Niesen, Helena Berglund, and Masoud Vedadi. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols*, 2(9):2212—2221, 09 2007.

[152] C. Ho, C. Lam, M. Chan, R. Cheung, L. Law, L. Lit, N. Ng, M. Suen, and H. Tai. Electrospray ionisation mass spectrometry: Principles and clinical applications. *The Clinical Biochemist Reviews*, 24:3—12, 2003.

[153] W. Johnson. Protein secondary structure and circular dichroism: A practical guide. *Proteins*, 7:205—214, 1990.