UNIVERSITY OF CALIFORNIA,
IRVINE

Client-Chaperone Selectivity, Solubility, and Hydrogel Characteristics of Human Crystallins and the Properties of D1-PSI.

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Molecular Biology and Biochemistry

by

Marc Anthony Sprague-Piercy

Dissertation Committee:
Rachel Martin, Chair
Celia Goulding
Douglas Tobias

2021

DEDICATION

Dedicated to The Cat, Jocelyn, Eden, and my Mom

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This work acknowledges the following grants for supporting this work: NIH grant EY021514, NIH-IMSD GM055246, and the HHMI Gilliam Fellowship. I also acknowledge Dima Fishman and the Linear Laser Lab at UCI, Ben Katz and everyone at the UCI Mass Spec facility. Wyeth Gibson and the Patterson Lab for the Cryo-TEM work. I also acknowledge Anna DeAngeles, Rai Ratan, and everyone at the USCD NMR facility.

I also want to thank the following members of the Martin Lab:

Kyle for all of the honest advice and late night science discussions.

Jan and Christian for the memes.

Megha and Megan for the enthusiam.

Suv for the NMR.

Jess for the patience.

Brenna for all her hard work.

Marquees for taking over the Mpro project.

All of the undergrads and high school students I worked with for teaching me as much as I taught them.

And Rachel for everything.

Chapter 1 adapted with permission from: Sprague-Piercy, M.A.; Rocha, M.A.; Kwok, A.O.; Martin, R.W. α-Crystallins in the Vertebrate Eye Lens: Complex Oligomers and Molecular Chaperones. Annual Review of Physical Chemistry 2021 72:1, 143-163

Chapter 2 adapted from: Sprague-Piercy, M.A; Wong, E.; Roskamp, K.W; Fakhoury, J.N; Freites, J.A.; Tobias, D.J.; Martin, R.W. Human αB-crystallin discriminates between aggregation-prone and function-preserving variants of a client protein, Biochimica et Biophysica Acta (BBA) - General Subjects 2020 , 1869.

Chapter 4 adapted from: Sprague-Piercy, M.A.; Bierma, J.C.; Crosby, M.G.; Carpenter, B.P.; Takahashi, G.R.; Paulino, J.; Hung, I.; Zhang, R.; Kelly, J.E.; Kozlyuk, N.; Chen, X.; Butts, C.T.; Martin, R.W. The Droserasin 1 PSI: A Membrane-Interacting Antimicrobial Peptide from the Carnivorous Plant Drosera capensis. Biomolecules 2020, 10, 1069.

VITA

Marc Anthony Sprague-Piercy

Education

- 2016–Current PhD Molecular Biology & Biochemistry, Expected Summer 2021 University of California Irvine
- 2014–2016 B.S. Molecular Biology and Biochemistry University of California Irvine GPA – 3.94 Summa Cum Laude
- 2006-2014 Transferred Fullerton College Fullerton
- 2009 Transferred Cypress College Cypress
- 2003-2006 Transferred Cerritos College Norwalk

Awards and Honors

- 2020. Teaching and Learning (CIRTL) Associate level certificate
- 2019. HHMI Gilliam Fellowship
- 2019. NSF GRFP Honorable Mention
- 2018. NIH-IMSD Graduate award
- 2016.Ayala Fellowship
- 2016. Graduate Opportunity Fellowship
- 2016. NIH-IMSD Graduate award
- 2015. MARC Fellow
- 2015. Sigma Xi nominee
- 2014. Fullerton College Man of Distinction
- 2014. HHMI EXROP Fellow
- 2013. BRIDGES Fellow

Publications

- Netrin-1 directs dendritic growth and connectivity of vertebrate central neurons in vivo. Anastasia N. Nagel, Sonya Marshak, Colleen Manitt, Rommel A. Santos, **Marc A. Piercy**, Sarah D. Mortero, Nicole J. Shirkey-Son, Susana Cohen-Cory. Neural development 10 (1), 1-21. (2015).

- Human αB-crystallin discriminates between aggregation-prone and function-preserving variants of a client protein. **Marc A Sprague-Piercy**, Eric Wong, Kyle W Roskamp, Joseph N Fakhoury, J Alfredo Freites, Douglas J Tobias, Rachel W Martin* Biochimica et Biophysica Acta (BBA)-General Subjects 1864 129502 (2020).

- Human γS-crystallin copper binding helps buffer against aggregation caused by oxidative damage. Kyle W. Roskamp, Sana Azim, Guther Kassier, Brenna Norton-Baker, **Marc A. Sprague-Piercy**, R.J. Dwayne Miller, and R.W. Martin* Biochemistry 59, 25, 23712385 (2020).

- An in silico approach to protease discovery. P. Kadandale , **Marc A. Sprague-Piercy**, and R. W. Martin* Article 10 In: McMahon K, editor. Advances in Biology Laboratory Education. Volume 41. Publication of the 41st Conference of the Association for Biology Laboratory Education (ABLE). (2020)

- The droserasin 1 PSI: A membrane-interacting antimicrobial peptide from the carnivorous plant *Drosera capensis*. **Marc A. Sprague-Piercy**, Jan C. Bierma, Marquise G. Crosby, Brooke P. Carpenter, Gemma R. Takahashi, Joana Paulino, Ivan Hung, Rongfu Zhang, John E. Kelly, Natalia Kozlyuk, Xi Chen, Carter T. Butts, and Rachel W. Martin* Biomolecules 10, 1069 (2020).

- Large-scale recombinant production of the SARS-CoV-2 proteome for high-throughput and structural biology applications. Nadide Altincekic, Sophie Marianne Korn, Nusrat Shahin Qureshi, Marie Dujardin, Martí Ninot-Pedrosa, Rupert Abele, Marie Jose Abi Saad, Caterina Alfano, Fabio CL Almeida, Islam Alshamleh, Gisele Cardoso de Amorim, Thomas K Anderson, Cristiane D Anobom, Chelsea Anorma, Jasleen Kaur Bains, Adriaan Bax, Martin Blackledge, Julius Blechar, Anja Böckmann, Louis Brigandat, Anna Bula, Matthias Bütikofer, Aldo R Camacho-Zarco, Teresa Carlomagno, Icaro Putinhon Caruso, Betül Ceylan, Apirat Chaikuad, Feixia Chu, Laura Cole, Marquise G Crosby, Vanessa de Jesus, Karthikeyan Dhamotharan, Isabella C Felli, Jan Ferner, Yanick Fleischmann, Marie-Laure Fogeron, Nikolaos K Fourkiotis, Christin Fuks, Boris Fürtig, Angelo Gallo, Santosh L Gande, Juan Atilio Gerez, Dhiman Ghosh, Francisco Gomes-Neto, Oksana Gorbatyuk, Serafima Guseva, Carolin Hacker, Sabine Häfner, Bing Hao, Bruno Hargittay, K Henzler-Wildman, Jeffrey C Hoch, Katharina F Hohmann, Marie T Hutchison, Kristaps Jaudzems, Katarina Jovi, Janina Kaderli, Gints Kalni, Iveta Kaepe, Robert N Kirchdoerfer, John Kirkpatrick, Stefan Knapp, Robin Krishnathas, Felicitas Kutz, Susanne Zur Lage, Roderick Lambertz, Andras Lang, Douglas Laurents, Lauriane Lecoq, Verena Linhard, Frank Löhr, Anas Malki, Luiza Mamigonian Bessa, Rachel W Martin, Tobias Matzel, Damien Maurin, Seth W McNutt, Nathane Cunha Mebus-Antunes, Beat H Meier, Nathalie Meiser, Miguel Mompeán, Elisa Monaca, Roland Montserret, Laura Mariño Perez, Celine Moser, Claudia Muhle-Goll, Thais Cristtina Neves-Martins, Xiamonin Ni, Brenna Norton-Baker,

Roberta Pierattelli, Letizia Pontoriero, Yulia Pustovalova, Oliver Ohlenschläger, Julien Orts, Andrea T Da Poian, Dennis J Pyper, Christian Richter, Roland Riek, Chad M Rienstra, Angus Robertson, Anderson S Pinheiro, Raffaele Sabbatella, Nicola Salvi, Krishna Saxena, Linda Schulte, Marco Schiavina, Harald Schwalbe, Mara Silber, Marcius da Silva Almeida, **Marc A Sprague-Piercy**, Georgios A Spyroulias, Sridhar Sreeramulu, Jan-Niklas Tants, Kaspars Trs, Felix Torres, Sabrina Töws, Miguel Á Treviño, Sven Trucks, Aikaterini C Tsika, Krisztina Varga, Ying Wang, Marco E Weber, Julia E Weigand, Christoph Wiedemann, Julia Wirmer-Bartoschek, Maria Alexandra Wirtz Martin, Johannes Zehnder, Martin Hengesbach, Andreas Schlundt. Frontiers in molecular biosciences 8, 89 (2021)

- α-Crystallins in the Vertebrate Eye Lens: Complex Oligomers and Molecular Chaperones. **Marc A. Sprague-Piercy**, Megan A. Rocha, Ashley O. Kwok, Rachel W. Martin*. Annual review of physical chemistry 72, 143-163 (2021).

- Chemical Properties Determine Solubility and Stability in Crystallins of the Eye Lens. Megan A. Rocha, **Marc A. SpraguePiercy**, Ashley O. Kwok, Kyle W. Roskamp, Rachel W. Martin*. ChemBioChem 22 (8), 1329-1346 (2021)

- Deamidation of the human eye lens protein S-crystallin accelerates oxidative aging. Brenna Norton-Baker, Pedram Mehrabi, Ashley O Kwok, Kyle W Roskamp, **Marc A Sprague-Piercy**, David von Stetten, R. J. Dwayne Miller, Rachel W. Martin. Structure. Submitted

Teaching Experience

- January 2020–March 2020. Univeristy of California, Irvine. Teaching and Learning (CIRTL) Associate level certificate
  - Given upon completion of the Developing Teaching Excellence course
  - Course focused on developing coursework based on active learning techniques
  - Received training on developing coursed centered on Student Learning Objectives (SLOs) to clearly explain to students what they will learn and what is expected of them

- March 2019–Current. University of California, Irvine. Developed an active learning activity centered around getting students involved in research and introducing them to techniques in structural biology Sponsor: Dr. Pavan Kadandale.
  - Involved students in research identifying novel cysteine proteases
  - Introduced students to examining protein structures using the program UCSF-Chimera
  - Resulted in a presentation an publication at the Association for Biology Laboratory Education (ABLE) conference and journal

- January 2019–March 2019. University of California, Irvine. Teaching Assistant Biochemistry, BIO SCI 98 Lecture Instructor: Dr. Pavan Kadandale.
  - Graded exams in a timely manner
  - Ran discussion textbfs focused on active learning sessions designed to meet clearly explained SLOs
  - Answered student questions in online discussion boards

- September 2018–Dec. 2018. University of California, Irvine. Teaching Assistant. Molecular Biology Lab, BIO SCI M116L Lab Instructor: Dr. Pavan Kadandale
  - Trained and guide students through lab activities
  - Taught students how to design experiments and interpret data
  - Graded lab reports and provide actionable feedback for students on how to improve writing skills
  - Answered student questions in online discussion boards
  - Hosted office hours to provide additional feedback and guidience for students

- September 2018–Dec. 2018. University of California, Irvine. Teaching Assistant. Virology, BIO SCI M124A Lecture Instructors: Dr. Paul Gershon, Dr. Micheal Buchmeier
  - Graded exams in a timley manner
  - Ran discussions focused on active learning sessions designed to meet clearly explained SLOs
  - Answered student questions in online discussion boards

- January 2018–March 2018. Teaching Assistant. University of California, Irvine Biochemistry Lab, BIO SCI M114L Lab Instructor: Dr. Julia Massimelli
  - Trained and guide students through lab activities
  - Taught students how to design experiments and interpret data
  - Graded lab reports and provide actionable feedback for students on how to improve writing skills
  - Answered student questions in online discussion boards
  - Hosted office hours to provide additional feedback and guidience for students

- 2018–March 2018. Teaching Assistant University of California, Irvine Biochemistry, BIO SCI 98 Lecture Instructor: Dr. Pavan Kadandale

- Graded exams in a timley manner
- Ran discussions focused on active learning sessions designed to meet clearly explained SLOs
- Answered student questions in online discussion boards

Research Experience

2017–Present PhD Candidate PI: Rachel Martin University of California - Irvine Working to understand the mechanism of protein aggregation and client-chaperone interactions at the atomic level using the Human lens proteins, the crystallins, as a model.

- Used both solution and solid-state NMR to build models of client-chaperone interactions
- Used protein structure to rationally generate protein variants to test specific hypotheses on protein function
- Extensive experience with protein expression and purification in bacterial expression systems in standard and isotopically labeled media
- Extensive experience with PCR and cloning to generate novel proteins
- Familiar with NMRpipe, NMRFAM-Sparky, UCSF-Chimera, PyMol, and NIH-Xplor
- Sponsored by The HHMI Gilliam Fellowship

2016 Graduate Student Researcher PI: Markus Ribbe University of California - Irvine Worked toward expressing a functional nitrogenase complex from *M. acetivorans* in *E. coli* as a proof of concept that the complete nitrogenase maturation pathway can be heterologously expressed to allow new organisms to fix nitrogen.

- Expression and purification of oxygen sensitive protein complexes.
- Cloning and transformation of DNA into *E. coli*
- Characterization of iron-sulfur clusters bound to proteins by EPR

2014–2016 Undergraduate Researcher PI: Susana Cohen-Cory University of California - Irvine Worked to develop new assays to characterize the role of Netrin-1 and DSCAM in the development of the retina, using *X. laevis* as a model by generating knock-out mutants using CRISPR-Cas9 on zygotes.

- Sponsored by the MARC (Maximizing Access to Research Careers) program

Summer 2014 Undergraduate Researcher PI: Keiko Torii University of Washington - Seattle Worked to develop a technique to identify downstream signaling partners of the membrane protein ERECTA, a protein important for plant growth, in *A. thaliana* by generating a biotin ligase-ERECTA conjugated protein to pull-down proteins that get tagged with biotin that come into contact with ERECTA.

- Sponsored by The HHMI EXROP fellowship

2013-2014 Undergraduate Researcher PI: Susana Cohen-Cory University of California - Irvine Developed a visually based behavior task to characterize changes in the development of the visual system of *X. laevis* tadpoles in response to changing axonal guidance cues.

- Sponsored by the BRIDGES to Baccalaureate program

Presentations

- February 2020 Identifying the structural features that differentiate client proteins of αB-crystallin. Marc A. Sprague-Piercy, Kyle Roskamp, Rachel W. Martin. Biophysical Society Annual Meeting, San Diego, CA.
- October 2019. Identifying the structural characteristics that determine αB-crystallin chaperone selectivity. Marc A. Sprague-Piercy, Kyle Roskamp, Rachel W. Martin. HHMI Science Meeting, Ashburn, VA.
- June 2019. An in silico approach to protease discovery. Pavan Kadandale, Marc A. Sprague-Piercy, Rachel W. Martin. ABLE. Ottawa, Canada
- February 2018. Investigating the aggregation mechanism of γS-crystallin by examining the G18A variant. Marc A. Sprague-Piercy, Kyle Roskamp, Rachel W. Martin. Lake Arrowhead, CA
- January 2018. Solid-state NMR probe development and characterization of γS-Crystallin as a lens-like hydrogel. Marc A. Sprague-Piercy, Rachel W. Martin. NMR Winter School. Stowe, VT
- October 2015. Marc Piercy, Rommel Santos, Susanaa Cohen-Cory. Methods to Elucidate the Role of DSCAM in the Development of the Vertebrate Retinotectal Circuit. Sigma Xi, Kansas City, MO.
- May 2015, Marc Piercy, Julian Avila. New Methods to Discover Protein-Protein interactions, HHMI EXROP Symposium, Washington D.C.
- November 2013. The Effect of Netron on Visually Guided Behavior. Marc Piercy, Susana Cohen-Cory. SCCUR, Whittier, CA.

- September 2013. The Effect of Netrin on Visually Guided Behavior. Marc Piercy, Susana Cohen-Cory. ABRCMS. Nashville, TN.

Service to the Scientific Community

- November 2019 Organized and hosted the Gilliam symposium at UCI. The symposium aimed to recruit URM students from local Cal States to promote diversity at the graduate level
- Summer 2019 Hosted a journal club for community college students participating in the BRIDGES program at UCI
- May 2018 Organized and hosted the Southern California Users of Magnets (SCUM) conference
- Summer 2018. Hosted a journal club for community college students participating in the BRIDGES program at UCI
- February 2018. Hosted a speaker session of the MB&B department retreat at UCI

ABSTRACT OF THE DISSERTATION

Client-Chaperone Selectivity, Solubility, and Hydrogel Characteristics of Human Crystallins and the Properties of D1-PSI.

By

Marc Anthony Sprague-Piercy

Doctor of Philosophy in Molecular Biology and Biochemistry

University of California, Irvine, 2021

Rachel Martin, Chair

This thesis examines the properties of the human crystallin eye lens proteins from three angles, how the structure of γS-crystallin relates to aggregation propensity, how αB-crystallin select between aggregation prone variants, and the structure of the crystallins in a lens like hydrogel. The human crystallins are normally highly soluble and stable, as an evolved adaptation to maintain the transparency of the eye lens. This is due to the intrinsic stability of the proteins themselves and the chaperone ability of the α-crystallins. By studying the structure and biophysical properties of novel and cataract associated variants of γS-crystallin with respect to aggregation propensity and client-chaperone interactions, it was discovered that αB-crystallin can select between different aggregation prone and function preserving variants, as well as a disconnect between the structural features that lead to a loss in stability and selection as a client. In the eye lens the concenration of the crystallins can reach over 400 mg/mL. By examining pure γS-crystallin as it exists in a highly concentrated hydrogel it was discovered that the gel is formed by the protein forming into larger order oligomers in the form of filaments and amorphous domains, but the domains retain a high degree of liquid like character. Lastly the properties of the plant specific insert (PSI) from *Drosera capensis* were examined. It was discovered that this protein exhibits anti-microbial properties, likely

stemming from the fact that PSI can destabilize cell membranes and readily forms complexes with lipids.

Chapter 1

Introduction

1.1 INTRODUCTION

For the eye lens to function correctly, it must be transparent and refract visible light strongly enough to form an image on the retina. The bulk of this unique tissue is made of layers of densely packed fiber cells filled with very stable proteins called crystallins. Crystallins comprise over 90% of the dry weight of the human eye lens [1]. These proteins exhibit a higher refractive index than average proteins, and their refractive power is not a simple function of the amino acid sequence [2]. During development, all organelles in fiber cell are degraded to minimize light scattering, leaving these cells without the machinery needed to synthesize new proteins. Thus, the crystallin proteins that are made during development remain soluble under crowded conditions for decades, and are used by the organism throughout its life. Crystallin proteins exist at very high concentrations in the lens, up to 400 mg/mL in humans [3], and even higher, over 1000 mg/mL in some aquatic species [4]. If these proteins aggregate, the light-scattering aggregate formed is called a cataract, the most common cause of blindness [5]. To understand how the eye lens works is to understand the physicochemical

properties of crystallins, including the molecular basis of their extraordinary solubility, the relationship between sequence and structure, how protein-protein interactions are mediated under conditions of extreme macromolecular crowding, and what happens when this system fails, resulting in cataract.

The study of lens crystallins tracks with the beginning of modern chemistry. In 1830, Berzelius isolated a gelatinous substance from the eye lens that he called crystallin because of its transparency [6]. This material was then further fractionated to show that it was in fact made up of more than one substance [7], but it was not until 1894 that the individual protein components were isolated and named α-crystallin and β-crystallin by Mörner [8]. Since then great strides have been made in understanding the structure and function of these proteins. The crystallins themselves are categorized into three broad categories in the human lens, the α-, β-, and γ-crystallins. In addition to these ubiquitous vertebrate crystallins, many more taxon-specific crystallins exist in other organisms, but these are beyond the scope of this review. The α-crystallins are chaperone proteins that belong to the small heat-shock protein family [9], while the β- and γ- crystallins are a part of an evolutionarily distinct superfamily of proteins called the $\beta\gamma$-crystallin superfamily [10]; they are mentioned here because they are the most common client proteins for α-crystallins in the lens. This review will discuss some history and recent advances in understanding the physical chemistry underpinnings of the unique and extraordinary properties of α-crystallins, with a particular focus on comparisons between human and fish crystallins. We discuss α-crystallin structures, their complex and dynamic oligomerization states, and most importantly, the intermolecular interactions by which they bind to damaged structural crystallins and keep them in solution. Finally, the sidebars provide an introduction to several techniques that are used to obtain this information.

1.2 α-crystallins are small heat-shock proteins

α-crystallins are small heat-shock proteins (sHSPs) that act as holdase chaperones, meaning they maintain the solubility of damaged client proteins, but are unable to refold them [11, 12]. While αA-crystallin is primarily expressed in the lens [13], αB-crystallin is expressed in many tissues throughout the body [14, 15], and alterations to its solubility or substrate binding competence are implicated in a variety of diseases [16]. Many studies of α-crystallin function have been performed using zebrafish as a model organism, as the vertebrate crystallin functions [17] and expression patterns [18] are strongly conserved. Furthermore, zebrafish embryos have the advantage of being transparent, enabling detailed studies of the developing lens [19]. Mouse α-crystallin promoters were found to drive GFP expression in several organs in the zebrafish, including the lens, notochord, heart, and skeletal muscle, indicating that mammalian α-crystallin promoter activity can also be screened in this model organism [20]. In zebrafish, knocking out αB-crystallin causes both lens defects and reduced cardiac stress tolerance [21], while αA was found to have the same function in the lens as it does in other vertebrates [22].

As in other sSHPs, α-crystallins have a β-sheet rich α-crystallin domain flanked by flexible N-terminal and C-terminal extensions [23]. A schematic view is shown in Figure 1.1. The α-crystallin domain is main active part of the protein, containing the region that is responsible for substrate binding; in fact, relatively short peptides from this region, the mini-α-crystallins, display substantial chaperone activity on their own [24, 25]. However, the tails also play important roles in controlling the activity, oligomeric state, and substrate recognition [26, 27], as illustrated by the different binding surfaces, corresponding to different sequence regions, that recognize amorphous aggregates and amyloid fibrils [28]. Structures of truncated α-crystallins show the primary dimer interface, which is in the middle of an extended β-sheet, but also reveal how the C-terminal extensions can participate in domain-

swapping, leading to oligomers of differing sizes [29]. The N-terminal extension has also been implicated in binding of specific substrates [28].

1.3 Polydispersity is central to α-crystallin function

Understanding how α-crystallins recognize client proteins and maintain their structural integrity requires high-resolution structures not only for the monomeric proteins, but also their complexes. α-crystallins exist as polydisperse mixtures of 10–40-mers [30, 31], with the most common species in the range of 24-28 [32]. Characterizing the equilibrium of oligomeric states is central to gaining full mechanistic insight into α-crystallin activity: NMR experiments have shown that polyhedra of varying sizes form, providing different environments for client proteins [33]. Heterogeneous interactions are observed even in constructs of the isolated α-crystallin domain, as key residues display resonances at more than one chemical shift position, indicating multiple conformations [34]. However, the importance of the C- and N-terminal extensions for controlling oligomerization has been demonstrated in several studies. The IPI motif on the C-terminal domain, embedded in the palindromic sequence "ERTIPITRE" can interact bidirectionally with the $\beta4/\beta8$ surface of the α-domain, and these interactions control both the oligomeric state and the ability of the protein to chaperone amyloid clients [35, 36]. Peptides mimicking this sequence readily bind to the α-crystallin domain in solution, providing support for the importance of this motif for protein-protein interactions [37]. Polydispersity appears to be central to α-crystallin function: not only does the formation of variable oligomers discourage crystallization or other deleterious interactions between the α-crystallin molecules themselves, but the ability of these molecules to self-organize into constantly shifting polyhedra provides a variety of surfaces that are presented to client proteins of differing sizes and shapes.

Solving the structures of α-crystallins in biologically relevant complexes is a challenging endeavor. Their dynamic and polydisperse nature makes them particularly suitable for NMR [38, 39], and in many cases hybrid structural methods are used to integrate information across different length scales. Various combinations of experimental and molecular modeling techniques have been used to model α-crystallin oligomers, sometimes with inconsistent results, possibly reflecting differences in protein constructs or sample preparation methods, or true heterogeneity.

For example, building on solid-state and solution NMR structures of smaller oligomers [40, 41], cryo-EM, solid-state NMR, and small-angle x-ray scattering (SAXS) were used to generate a model of symmetric αB-crystallin 24-mers [42] (Figure 1.2A). These oligomers are built using a hierarchical progression where dimers connected by the β-sheet interface assemble via their C-terminal extensions to form hexamers, which in turn associate via their N-terminal regions to create higher-order structures. Another hybrid structure of αB-crystallin 24-mers was generated using a combination of cryo-EM, solid-state NMR, and molecular modeling also came to the conclusion that the building blocks are dimers, which then associate to form hexamers [43]. This structure also identifies two different dimer arrangements, consistent with the variable interactions between the palindromic sequence of one monomer and hydrophobic residues of the other observed in the crystal structures [35], but in contrast to solution-state NMR data showing only a single set of chemical shifts, suggesting that all monomers adopt the same conformation [33].

In addition to the most common species, the 24-mers, oligomers ranging from 6-mers to 48-mers have also been observed in solution. These data and the closed, spherical shape of the higher-order oligomers suggest that the highly symmetric complexes may be for storage, whereas the binding-competent species are either smaller complexes (perhaps dimers) or incomplete spheres allowing room for client proteins. Solution NMR indicates that ms-timescale dynamics in the C-terminal extensions regulate the oligomerization state [44].

Some details of the dynamic equilibrium among these complexes were revealed using deuteration-assisted small-angle neutron scattering (DA-SANS) in conjunction with electrospray ionization (ESI) native mass spectrometry (nMS): exchange of subunits between complexes was common, however oligomer collapse did not appear to occur, even for a subset of the population [45]. Thus, the dynamic quaternary structure of αB-crystallin appears to be mediated by small oligomers dissociating from and re-associating with larger complexes. Other studies have suggested that monomers are present under physiological conditions [46] and have chaperone activity [32], but are compromised from a solubility standpoint [47]. Beyond the size of the oligomers, the morphology of the α-crystallin oligomers can also affect chaperone function. α-crystallins can retain function after forming amyloid fibrils or amorphous aggregates, with amyloid fibrils even showing enhanced affinity for particular clients [48].

The first structural models of the lens-specific chaperone αA-crystallin in 12-, 16- and 20-mers were generated recently using a combination of cryo-EM, mass spectrometry with cross-linking, NMR spectroscopy, and molecular modeling [49](Figure 1.2B). This model suggests that for αA-crystallin, the building block for oligomerization is a tetramer. As in the case of αB-crystallin, the exact extent of oligomerization is governed in part by the extent of C-terminal domain swapping [50, 51, 52], although the N-terminus is also involved in oligomerization, as its truncation shifts the equilibrium to favor smaller oligomers [53, 54]. Domain-swapped oligomers are found in multimers larger than 12, whereas the dodecameric species can form without any interactions with the C-terminus [49]. These results are consistent with earlier crystal structures of truncated versions of vertebrate αA-crystallins that showed conformations both with [29] and without domain-swapping [55]. Recently, an alternative splicing variant of human αA-crystallin with a truncated N-terminal sequence was discovered. On its own, this isoform has only weak chaperone activity and makes only small oligomers, but it is capable of integrating into larger oligomers of the more common canonical variant, modulating oligomer size and chaperone activity [56]. In

contrast, an alternative splicing variant found in rodents (αA^{ins}) has a longer N-terminal domain, and has enhanced activity relative to the standard isoform [57]

1.4 Cataract-related α-crystallin variants can be aggregation-prone or have altered chaperone activity

Many cataract-associated variants of α-crystallins feature an altered α-domain, which results in structural changes and/or compromised chaperone activity. The αB-R69C and D109H variants are implicated in human disease, including cataract [58, 59]. Both cause considerable structural changes, reduced chaperone activity, and increased aggregation propensity *in vitro* [60]. The D109H variant is particularly disruptive, probably because the loss of the negative charge at position 109 disrupts a critical salt bridge that is required for proper formation of the anti-parallel sheet dimer interface, as does the R120G variant [61]. The R12C variant of αA-crystallin is more prone to aggregation than WT, especially in the presence of calcium ions [62]. Both the αA-G98R and αA-R21Q variants show decreased function, but when mutated together to make an αA-G98R/R12Q double variant, compensating for the difference in charge, much of the function and stability of the protein is rescued [63].

In other cases, the major driving force of cataract formation is how mutations in α-crystallin affect the aggregation of other proteins. For example, the R49C and R116C variants of αA-crystallin have differing affinities for the aggregation-prone client protein I4F γD-crystallin [64]. The higher affinity R49C variant results in increased aggregation of γD-I4F, possibly through simple mass action, as the folded form is sequestered by the chaperone protein, thus pushing the equilibrium toward formation of the denatured form [64]. In mouse models, knocking out αA and αB-crystallin leads to increased abundance and cross-linking of βB2-crystallin [65] the αB-R120G and the αA-R49C variants lead to decreased protein degra-

dation in the lens during development [66], which may lead to an increased propensity for cataract formation, or alternatively might be an effect of cataract formation. However, not all mutations are deleterious; for example the R12C variant of αB-crystallin exhibits increased chaperone activity in addition to changes in the oligomeric state of the protein with increased population of a dimer form that is favored as a result of a disulfide bond [67]. This variant also resists thermal and calcium-induced aggregation compared to WT. However, when exposed to calcium ions the chaperone activity of the variant is decreased, potentially due to structural changes, as the R12C variant shows altered chemical environments of its tryptophan residues, suggesting partial unfolding [67].

1.5 α-crystallins bind divalent cations, with complex effects on their function

Human αB-crystallin binds one equivalent of Cu^{2+}, and its chaperone activity is enhanced by this interaction [68, 69, 70]. αB is able to prevent aggregation and co-precipitation of Cu^{2+} and human γD at all Cu^{2+}:γD ratios, suggesting that αB is acting as a chaperone rather than a chelator in this situation [71]. However, there is also evidence of metal ion homeostasis through sequestration. Increasing the amount of α-crystallin bound to the client protein γD, which would occur naturally with age, results in an increase in free iron and calcium ions [72], which would increase the susceptibility of the eye lens to oxidation. Schiff bases and rutin, a naturally occurring flavonoid, were shown to inhibit copper-induced aggregation of human γD and promote the chaperone activity of αB, suggesting a possible role as a cataract therapeutic [73, 74].

In the case of Zn^{2+}, αB-crystallin was again able to prevent metal ion-induced aggregation, but here it appears to act as a chelator rather than a chaperone, because higher concen-

trations of Zn^{2+} were still able to cause aggregation [75]. Zinc ions mitigated diabetic cataract in mouse models due to their antioxidant properties and positive interactions with α-crystallin [76, 77, 78]. Zn^{2+} interactions with α-crystallin do not restructure the protein's secondary or tertiary arrangement, but they do increase its surface hydrophobicity, thereby enhancing the chaperone activity [76]. In human lens epithelial cells, treatment with Cd^{2+} or Cu^{2+} promoted the expression of αB-crystallin, whereas Cu^{2+} promoted the expression of αA-crystallin [79].

1.6 α-crystallin post-translational modifications can cause age-related cataract

The majority of cases of cataract are caused by aging: most people are born with healthy lens proteins, which accumulate post-translational modifications (PTMs) over time, leading to cataract [80, 81]. UV irradiation is associated with the formation of cataract [82]. In αA-crystallin, UV damage leads to increased hydrophobic exposure, secondary structure changes, and diminished chaperone activity [83]. Backbone cleavage can lead to the accumulation of insoluble peptides in the lens, providing nucleation sites for the growth of aggregates. For example, a 15-residue peptide derived from αA-crystallin aggregates in aging lenses, recruiting full-length protein and forming β-sheet rich fibril structures [84].

One of the most common PTMs in the lens is isomerization [85], which is difficult to detect because it does not change the protein mass. Serine and aspartic acid can epimerize to form their D-counterparts, and Asp can also convert to L- or D-isoAsp [86]. These modifications can alter the protein structure, sometimes causing disruption that goes beyond the mutation site [87]. In the eye lens, Asp isomerization increases with age, with αA-crystallin being more susceptible to isomerization than αB-crystallin [88]. Isomerization in αA changes the relative

populations of different oligomers [89] In αB, isomerization of Asp109 disrupts a critical salt bridge at the dimer interface, negatively impacting its solubility, while epimerization of Ser162 significantly weakens the dimerization interface between the palindromic sequence and the β4-β8 groove [90]. Glycation, particularly of lysine side chains, can also occur in the crystallins, linking cataract and diabetes [91]. In mice, L-lysine treatment has been shown to reduce glycation in α-crystallin after treatment with glucose [92]. PTMs have even been shown to enhance α-crystallin activity. Acetylation of lysine residues have been linked to higher hydrophobic exposure and increased chaperone activity in α-crystallin [93]. Recently, the link between the αA-crystallin oxidation state and oligomerization and function has been characterized. When αA-crystallin is oxidized, a destabilizing disulfide bond is formed, dispersing higher-order oligomers [49].

Another important post-translational modification that impacts the regulation of protein function is phosphorylation, which happens to α-crystallins under stress conditions. In αB-crystallin, phosophorylation occurs in a heterogenous manner, with mixed populations of proteins phosphorylated at different sites. Overall phosphorylation efficiency is low, due to the inability of kinases to act on monomers that are part of a large complex; phosphorylation appears to happen during subunit exchange [94]. Phosphorylation of αB-crystallin is mostly localized to three serine residues: 19, 45, and 59 [95, 96]. When these sites are phosphorylated, αB-crystallin forms smaller oligomeric complexes, and shows preferential binding to different client proteins [97]. An αB-crystallin variant mimicking hyperphosphorylation of these residues showed increased chaperone function relative to wild-type, using insulin as a client [98]. Low levels of phosphorylation appear to improve chaperone activity and reduce aggregation propensity, whereas hyperphosphorylation, which sometimes happens to variant proteins *in vivo*, leads to increased aggregation [99]. Consistent with this observation, extensive phosphorylation at S45 resulted in uncontrolled aggregation in vitro [100]. Epimerization and phosphorylation can also interact to alter the protein's behavior: phosphorylation of Ser59, is precluded by epimerization of this residue, and reduced by the

isomerization of the nearby residue Asp62 [90]. Despite the many advances in understanding α-crystallin function since Horwitz first discovered its molecular chaperone properties, studies such as these underscore the importance of recognizing that α-crystallins may behave differently in dilute, homogenous solutions than they do in the messy, crowded environment of the cell [101].

1.7 Client-chaperone interactions

Recognition of misfolded, partially unfolded, or otherwise solubility-compromised proteins is central to α-crystallin function. Although the exact mechanisms of client recognition are not yet fully understood, the size, charge, size, and exposed hydrophobic surface of the client protein all appear to play a role [102]. All-atom molecular dynamics (MD) simulations were used to probe the interactions of monomers and oligomers of a small peptide implicated in Alzheimer's disease-related plaques ($A\beta_{17-42}$) with an αB-crystallin ACD dimer. The dimer bound more strongly to oligomers of $A\beta_{17-42}$ relative to monomers, as each peptide has only limited number of contacts with the ACD dimer, resulting in weak and transient interactions. However, oligomer formation locks the peptides into a more rigid conformation, resulting in a larger interaction surface [103]. Particular residues in αA, such as D69 [104] and F71 [105] have all been proposed as being critical to chaperone activity, as have the C-terminal [106] and N-terminal tails [107]. An important step in understanding what governs these protein-protein interactions was the discovery of small peptides derived from the α-crystallin domain that are active as chaperones. For example, the peptide "KFVIFLDVKHFSPEDLTVK" from αA-crystallin has chaperone activity and was thus named mini-αA-crystallin [108]. In αB-crystallin a similar peptide, "DRFSVNLDVKHFSPEELKVK" was identified, and has been called mini-αB-crystallin (see Figure 1.1B) [109]. The discovery of these peptides

has allowed for more focused probing of the molecular mechanism of the client-chaperone interaction involving the α-crystallin domain.

A common finding when investigating α-crystallin binding interactions is the importance of hydrophobic interactions. Mini αA-crystallin is notably hydrophobic, and its confirmed site of interaction with human γD-crystallin specifically involves several hydrophobic residues [110]. In human γD-crystallin, both the I4F and V76D variants stabilize an aggregation-prone intermediate, leading to early onset cataract. However αB-crystallin does not recognize either of these variants, whereas it does bind the I4F/V76D double variant [111], suggesting that a minimum amount of unfolding is needed to trigger α-crystallin activity. More evidence that full-length αB-crystallin can select between native-like and strongly aggregation-prone proteins with very similar, folded structures is provided by the difference between αB-crystallin's robust binding to the G18V variant of human γS-crystallin, but not the G18A variant, which is likewise thermally destabilized, albeit to a lesser extent [112]. Similarly, hydrophobic residues in melittin mediate the interaction of full-length αA-crystallin with this peptide [113]. Although hydrophobic interactions are clearly important, there is also more to the story. In αB-crystallin alone there are at least 13 peptide sequences related to chaperone activity [114], and each of these could reveal unique mechanisms that explain αB-crystallin's prodigious ability to chaperone diverse client proteins. A schematic illustrating the general binding modes is given in Figure 1.3A; Figure 1.3B shows the location of the mini-αA peptide and several other key residues on the α-crystallin monomer.

One area of current interest is how α-crystallin interacts with UV-damaged client proteins. Current work has shown that α-crystallin forms very large complexes when binding β_L-crystallin, and that the size of the complex is positively correlated with the content of UV-damaged protein [115]. High-resolution structural studies are needed to resolve the details of these complexes. So far studies have all mostly focused on the role and mechanism of the

α-crystallin domain, but the N- and C- terminal extensions are also important, possibly for initial client recognition [116], in addition to their role in mediating oligomerization.

1.7.1 Species-specific adaptations

α-crystallins have species-specific adaptations that enable them to recognize particular client proteins: bovine α-crystallin effectively protects cow γ-crystallins from aggregation, while failing to interact with structural crystallins from the Antarctic toothfish, and even α-crystallins from the bigeye tuna only partially protected toothfish γ-crystallins under heat stress conditions [117]. The cold adaptations required for the lifestyle of the Antarctic toothfish, which lives at a temperature of -2°C year-round provide some clues to this exquisite specificity. Like humans, the toothfish has two α-crystallin paralogs, αA and αB, although its lens protein mixture is more complicated due to the much larger number of γ-crystallin paralogs [118]. Sequence comparison shows that α-crystallins from cold-adapted fish have more hydrophobic residues than their warm-water counterparts, and mutating key residues in zebrafish αA-crystallin with their counterparts in the toothfish protein altered surface hydrophobicity (measured using bis-ANS binding), oligomerization, and chaperone activity [119]. Unlike humans and toothfish, zebrafish have two αB-crystallins, αBa and αBb [120]. αBb is the more similar to human αB, whereas αBa has unusual oligomerization behavior and superior chaperone activity against a panel of destabilized T4 lysozyme variants [57].

1.7.2 Human α-crystallin chaperone activity as as a biomedical and therapeutic target

An important reason for studying the molecular mechanisms of α-crystallin chaperone action and aggregation is the long-term goal of using this information in a therapeutic context. α-

crystallin itself is potentially useful as a therapeutic agent against protein deposition diseases, particularly the α-crystallin domain and component peptides. The α-crystallin domain binds to, and prevents the fibrillization of Aβ_{1-42} [121], a protein found in amyloid plaques in Alzheimer's disease. The ability of α-crystallin to bind amyloid fibrils in a therapeutic context has also been examined in the context of Multiple Sclerosis (MS). MS symptoms are linked to the formation of amyloid fibrils of fibronectin, contributing to the formation of an MS lesion, so maintaining the solubility of these proteins could lead to an effective treatment for MS. In a clinical trial, doses of αB-crystallin were shown to reduce the number of MS lesions by 76% over a 9 month period [122]. Delivery remains a problem with this approach: it is difficult to get α-crystallin into the cell efficiently. αB-crystallin can be fused to the glycoprotein C cell penetrating peptide, followed by stimulation of cell uptake by applying heparan sulfate [123].

Of course α-crystallin is also associated with disease states, and so it is itself a target for therapeutics. Attempts at treating α-crystallin associated cataract typically focus on resolubilizing the protein, as protein turnover in the lens is negligible so degradation in not a viable option. Aspirin nanorods have shown effectiveness in maintaining α-crystallin solubility *in vitro*, allowing for the possibility of an aspirin-based cataract treatment [124]. Mouse models of hereditary and age-related cataract have been successfully treated with small molecules, with one compound in particular, 29, 5-cholesten-3b,25-diol, improving the solubility of α-crystallin by 63% [125]. RNA aptamers can act as molecular switches to control the activity of the targeted proteins. RNA aptamers have been developed that bind specifically to αB-crystallin while not targeting the very similar αA-crystallin [126]. This work marks an important first step toward selectively and specifically controlling the activity of αB-crystallin.

1.8 Conclusion and outlook

We have presented an overview of the current state of knowledge of the molecular mechanisms of α-crystallin oligomerization and chaperone activity. Much of this information has come from combinations of complementary methodologies, notably solid-state and solution NMR with cryo-EM. Future progress toward understanding native α-crystallin oligomers and their interactions with client proteins will require advanced biomolecular simulations. MD simulations provide a detailed picture of how proteins move, leading to insights into protein folding, enzyme activity, and solvent interactions [127, 128]. MD is also an integral part of the process of refining macromolecular structures based on X-ray crystallographic [129] and NMR [130] data. MD simulations are often performed using a detailed structural model containing all the atoms in the protein as well as atomically detailed solvent molecules, although coarse-grained approaches and implicit solvent treatments are also available. Monte Carlo simulations can be used to study the equilibrium behavior of a wide range of physical systems. In this method, the system starts in an arbitrary configuration, and a perturbation is applied. Depending on whether the new configuration is lower or higher in energy, the move is accepted or rejected [131], and a wide range of thermodynamic properties can then be explored. In the context of systems containing many proteins, rigid-body models are often used to reduce the computational cost [132]. Multiconformation Monte Carlo simulations can reintroduce some conformational complexity at low additional cost by using a library of predetermined protein conformations [133]. For systems that are characterized by complex interactions among a large number of monomers, another alternative to computationally expensive all-atom simulations is the network Hamiltonian dynamics approach, a coarse-grained method in which a network Hamiltonian is written strictly in terms of connectivity among the protein monomers, eliding the molecular-level details. The properties being simulated are therefore characteristic of the entire system rather than of the structural details of any particular protein, allowing long-timescale simulation of systems comprising hundreds

to thousands of monomers. This approach has been used to recapitulate all known types of amyloid fibrils observed in the Protein Data Bank [134] and to model the formation kinetics of several different fibrillization pathways [135]. It may also provide a modeling framework for understanding α-crystallin oligomerization and the dynamic equilibrium among oligomers of different sizes, even before detailed all-atom simulations for the full system can be obtained. α-Crystallin forms complex, heterogeneous oligomers that are built up in a modular fashion based on domain-swapping interactions with its N- and C-terminal extensions. The full picture of α-crystallin chaperone behavior is very complicated, with different parts of the chaperone contact- ing its aggregation-prone substrate depending on specific properties of the client. A clear future direction in this area is the continued exploration of α-crystallin interactions with a greater variety of misfolded or destabilized proteins, as the full range of α-crystallin activity has probably not yet been characterized. Many different client protein modifications have been observed, ranging from amino acid substitutions to PTMs such as phosphorylation and isomerization. A pattern that emerges from all of these studies is that there are multiple pathways for aggregation depending on the nature of the modification. As these aggregates are formed by different mechanisms and may have different physicochemical properties, pharmaceutical interventions for medical conditions caused by these modifications will require either tailored treatments or a general solution that encompasses all of the α-crystallin functionalities observed in nature. In contrast, understanding α-crystallin properties and crystallincrystallin interactions can also lead to improved designs for artificial lenses that more closely mimic the functionality of the biological lens.

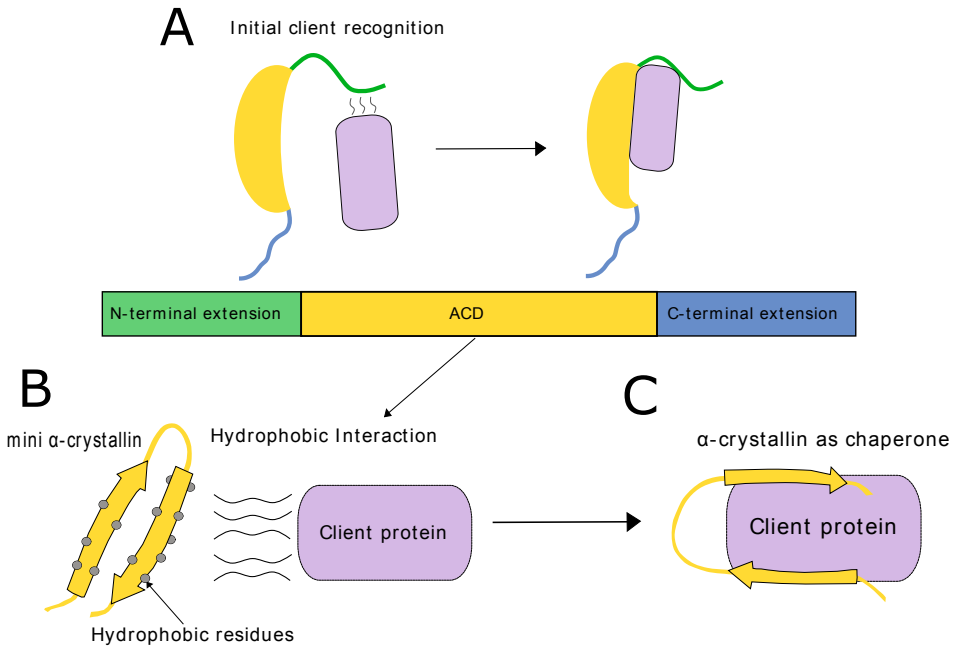

Figure 1.1: **α-crystallin binding to client protein.** (A) The initial recognition of the client protein sometimes occurs at the N- terminal extension of the α-crystallin [116]. (B) Mini αA-crystallin contains many hydrophobic residues that bind to the client protein through hydrophobic interactions [110]. (C) αA-crystallin tightly binds the client protein, keeping it in solution.

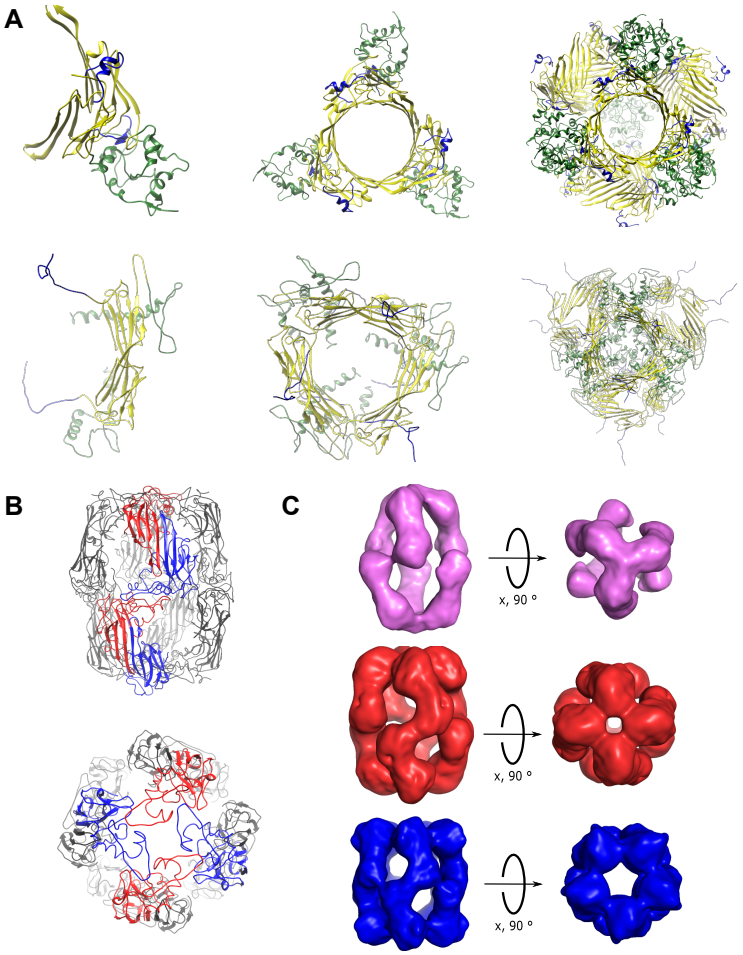

Figure 1.2: Oligomeric states of α-crystallins. (A) The Jehle (top) [31, 42] and Braun (bottom, PDB: 2YGD) [43] αB 24mer pseudoatomic models. Homodimers (left) are formed through ACD (yellow) contacts. Dimers combine to form a hexameric species (middle) through NTR (green) and ACD contacts. The C-terminal domain is highlighted in blue. Hexamers combine to form the dominant 24-mer structures (right). (B) Pseudoatomic models of a 16-mer of wild-type reduced human αA-crystallin[49]. Top: αA-crystallin monomers (blue and red) forming the β7-interface dimer. Dimers via interactions between N-terminal extensions, to form z-shaped tetramers that stack to make up the pillars of the hollow oligomer. Bottom: Monomers (red) binding through N-terminal interactions. (C) Cryo-EM density maps from the Electron Microscopy Database (EMD) [49] of the (top) 12-mer (EMD-4895), (middle) 16-mer (EMD-4894), and (bottom) 20-mer (EMD 4896) with three, four, and five-fold symmetry, respectively, from the apical axis (right). All oligomers are formed from a z-shaped tetrameric building block. Abbreviations: ACD, α-crystallin domain; cryo-EM, cryo-electron microscopy; EMD, Electron Microscopy Database; NTR, N-terminal region; PDB, Protein Data Bank.

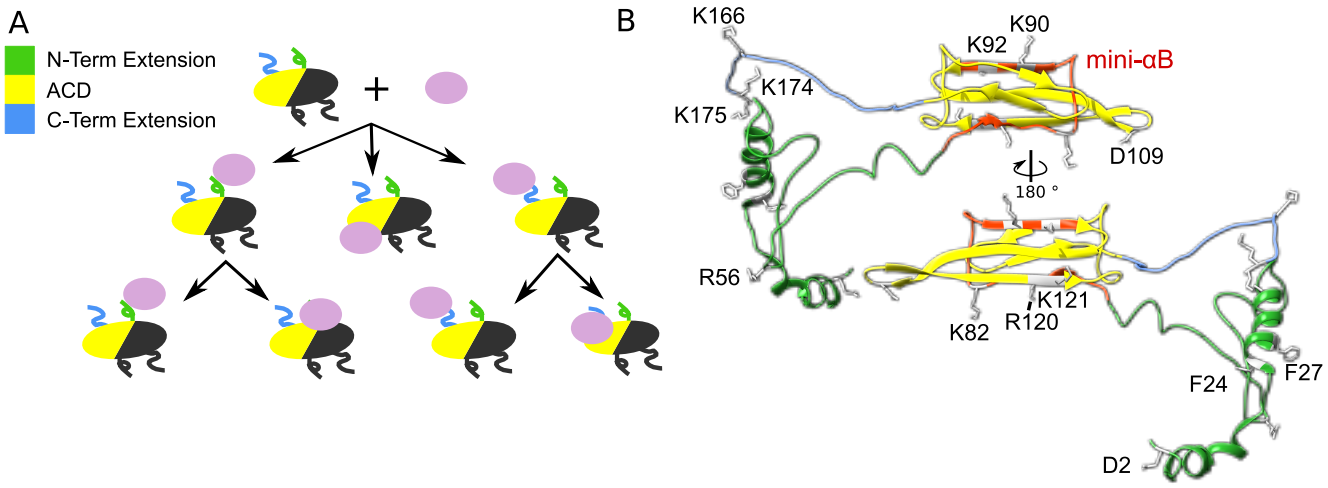

Figure 1.3: **A** A schematic of the interaction modes of αB-crystallin with client proteins. There are several different possible paths for client recognition and chaperone activity. Both the N- (green) and C-terminal extensions (blue) can act as site for initial recognition or even as the active holdase region. In the case of initial recognition, the client initially binds the flexible extension and then becomes bound to the α-crystallin domain. The α-crystallin domain itself can also directly interact with client proteins for holdase activity, without intermediate binding. **B**. Many specific residues and sequence regions are implicated in αB-crystallin holdase activity. The mini-αB peptide is highlighted in red, while the other key residues throughout the protein are labeled. These residues are D2, F24, F27 [107], R56 [136], D109 [59], R120[137], R157 [138], K82, K90, K92, K121, K166, K174, K175 [139]. Abbreviation: PDB, Protein Data Bank.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Our crystallin work is supported by NIH grants 2R01EY021514 to R.W.M and 1R01EY025328 to R.W.M. and D.J. Tobias and NSF grant DMR-2002837 to R.W.M. and D.J. Tobias. M.A.S.P. is supported by the HHMI Gilliam Fellowship. This work builds on discussions at the 2018 Crystallin Satellite Biophysical Society Meeting in San Francisco, and we are grateful to the participants for ideas and suggestions.

4

Chapter 2

Human αB-crystallin discriminates between aggregation-prone and function-preserving variants of a client protein

2.1 Introduction

For many proteins, the solubility and oligomerization state are critical functional properties that are subject to alteration by mutation or post-translational modification. Protein aggregation is associated with many human pathologies, including Alzheimer's and Parkinson's diseases as well as chronic respiratory conditions [140], transthyretin amyloidosis[141], and sickle cell anemia[142]. Protein aggregates can cause disease via loss of function, as in α-antitrypsin deficiency: when α-antitrypsin aggregates, the normal function of this protein to

inhibit proteolysis is lost, leading to tissue disruptions[140]. In other cases pathogenesis is due to direct cytotoxicity of the aggregates, as in transthyretin amyloidosis[141].

In the case of crystallins, their normal function is to maintain lens refractivity and transparency. These proteins are therefore exceptionally stable and soluble, even at high concentrations (above 400 mg/mL in humans)[143, 144]. In the eye lens there is a negligible amount of protein turnover[145], meaning that the crystallins must remain soluble throughout the life of an organism, offering a unique model for the study of protein stability and solubility. With age, however, accumulated damage to the crystallins results in light-scattering aggregates causing an opacification of the lens, a condition known as cataract.

There are two main classes of lens crystallins, α- and $\beta\gamma$-crystallins. The α-crystallins are small holdase chaperone proteins that form very large, multimeric complexes[146, 147]. The $\beta\gamma$-crystallins are highly refractive structural proteins that maintain the refractive index gradient of the eye lens. Post-translational modifications including deamidation[148], glycation[149], and UV-filter adduction[150] have been shown to reduce crystallin solubility. Many mutations in both α- and $\beta\gamma$-crystallins have been associated with early-onset cataract, including the R116C [151] and the R49C variants [152] of αA-crystallin, the R120G variant of αB-crystallin[153], the S39C[154] variant of γS-crystallin, and the P23T[155, 156, 157] variant of γD-crystallin. In particular, the γS-G18V variant of γS-crystallin has been associated with childhood-onset cataract[158].

The structure of γS-crystallin is typical of $\beta\gamma$-crystallins in general, with two double Greek key domains. Although the sequences of these two domains are not identical, their structures are very similar, resulting in a highly symmetric protein. A theme that has emerged from investigation of cataract-associated variants of γD- and γS-crystallin is their structures are often not dramatically different from that of the wild-type protein. In the V24M variant, the core of the N-terminal domain is more closely packed, but the protein maintains the overall β-sheet structure[159]. Similarly, the tertiary structures of the γS-D26G [160] and γS-G18V

[161] variants are only subtly different from that of γS-WT. The structure of γS-crystallin with important features labeled is shown in Figure 2.1.

Although the structures of γS-G18V and γS-WT γS-crystallin are very similar, the holdase chaperone αB crystallin recognizes these subtle changes in structure, preferentially binding to the partially destabilized γS-G18V variant but not the wild-type protein[161]. The unfolding temperature of γS-G18V is 6 °C higher than the aggregation temperature, indicating that the aggregation of the protein is not due to the protein simply unfolding; [162] instead γS-G18V exhibits a slightly destabilized N-terminal domain characterized by an increase in hydrophobic exposure on the surface of the protein [163]. Not all of the known cataract-associated variants of structural lens crystallins exhibit increased hydrophobic surface area or recognition by αB-crystallin. For instance, the cataract-associated I4F and V76D variants of γD-crystallin do not bind to αB to any appreciable degree, even though these variants are noticeably destabilized relative to γD-WT[164].

In order to investigate which local changes in structural crystallins enable recognition by αB-crystallin, we generated the novel G18A variant of γS-crystallin (γS-G18A). This variant was chosen to provide a minimal structural disruption that may or may not be recognized by αB-crystallin, as the mutated residue is located in a tight turn with backbone torsion angles that are canonically favorable only for glycine. We find that although γS-G18A is thermally destabilized and somewhat aggregation-prone relative to γS-WT, it exhibits minimal structural perturbations at physiological temperature and is not recognized by αB-crystallin.

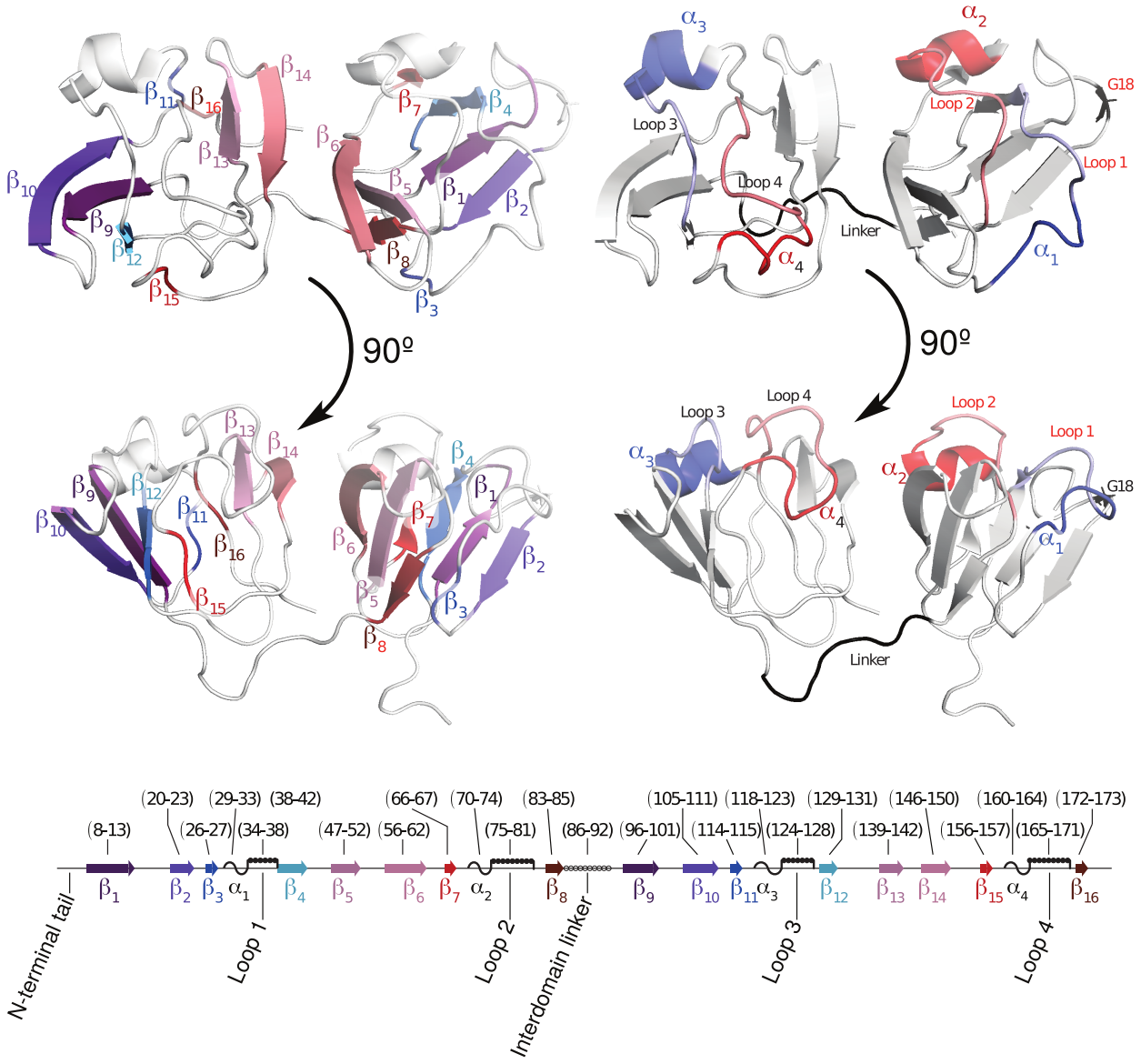

Figure 2.1: A schematic view of human γS-WT (PDB ID: 2M3T), highlighting key sequence features, including β-strands (Left) and α-helices or loops (Right) on the structure. The mutated residue (G18) is labeled as well. The span of each helix, strand, or loop is laid out linearly (Bottom) to indicate how the sequence of γS-WT relates to the structure.

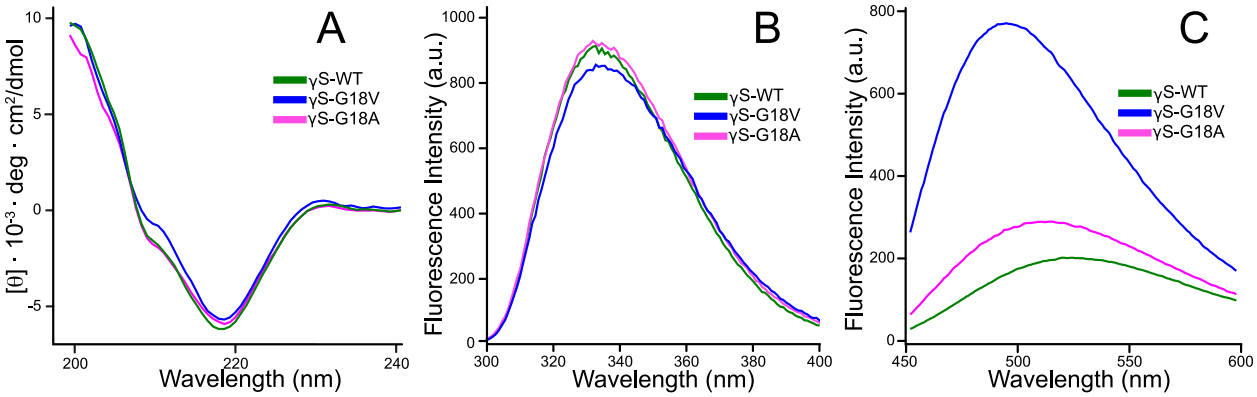

Figure 2.2: CD spectra indicate that the overall fold of γS-G18A is similar to that of γS-WT. (**A**) CD spectra of γS-WT, γS-G18A, and γS-G18V. The spectra of all three proteins are similar overall, however γS-G18V has a shoulder at 200-210 nm not observed in the others. (**B**) Intrinsic tryptophan fluorescence spectra of γS-WT, γS-G18A, and γS-G18V after excitation at 280 nm. The spectrum of the γS-G18V variant exhibits a slight red shift relative to the other two. (**C**) ANS binding assay for γS-WT, γS-G18A, and γS-G18V. The γS-G18V variant has the highest intensity, indicating that the γS-G18V variant has the largest hydrophobic surface area exposed to solvent.

2.2 Results and discussion

2.2.1 Experiments and molecular dynamics simulations suggest that the overall fold of γS-G18A is similar to that of γS-WT

The circular dichroism (CD) spectra of γS-WT, γS-G18A, and γS-G18V were measured to probe changes in secondary structure. The spectrum of γS-G18A exhibits a minimum at 218 nm, indicative of a primarily β-sheet structure (**Fig. 2.2A**). The spectrum of γS-G18A is more similar to γS-WT than γS-G18V; it lacks the shoulder at 200–210 nm observed in the spectrum of γS-G18V.[158, 162]

In vertebrate lens γ-crystallins, two buried tryptophans are ubiquitously conserved within each domain.[165, 166]. Each set acts independently as a FRET pair, enabling the γ-crystallins to quench UV radiation and act as a photoprotectant.[167, 168]. Minor changes to the chemical environment surrounding the tryptophans reduce their quenching efficiency,

providing a sensitive measure to detect minor conformational changes.[169, 170] Increased polarity of the tryptophan environment from solvent exposure results in red-shifting of the fluorescence spectrum[171]. This change can be observed for γS-G18V, in which the fluorescence is broadened and the peak intensity is red-shifted to 328 nm (**Fig. 2.2B**). The emission spectrum of γS-G18A is almost identical to that of γS-WT, consistent with the absence of increased relative solvent accessibility (RSA) observed in the γS-G18A Molecular Dynamics (MD) simulations (see below), and indicates that the chemical environments near the buried tryptophans are minimally perturbed (**Supplementary Figure 1**).

Anilinonaphthalene 8-sulfonate (ANS) is a small-molecule fluorophore often used to investigate hydrophobic surface exposure in proteins.[172, 173] Binding is mediated through both hydrophobic and electrostatic interactions.[174, 175] Increased hydrophobic surface area is observed in some cataract-related variants such as γS-G18V, [163] γS-V42M, and γD-P23T, but not others, including γS-D26G.[160, 173, 159] ANS fluorescence increases moderately in the presence of γS-G18A relative to γS-WT, but the signal is dramatically weaker than for γS-G18V (**Fig. 2.2C**). Similarly, a hypsochromic shift to 513 nm (from 525 nm) is observed for γS-G18A compared to 495 nm for γS-G18V. These changes indicate that γS-G18A exhibits a small increase in hydrophobic surface area relative to γS-WT. We hypothesize that these changes are localized close to the mutation site; in our previous study of G18V the area around this loop opened up to expose more hydrophobic side-chains to solvent[163]. Further, the strong hypsochromic shifting in conjunction with increased fluorescence suggests strong interactions with positively charged residues, such as Arg19 and Arg20.

MD simulations of γS-WT and γS-G18A were performed to investigate mutation-induced changes in protein structure and dynamics. No dramatic changes were observed in the N-terminal domain (NTD) or C-terminal domain (CTD) of γS-WT or γS-G18A as a whole based on C_α root mean square deviation (RMSD). Whole-protein changes are attributable to the flexible N-terminal tail extension and inter-domain linker regions of the protein

(Fig. 2.3). At the residue level, γS-G18A exhibits increased root mean-square fluctuations (RMSFs) within the α-helices and sequential loop sequences of the NTD (**Supplementary Figure 2**). Spatially, these regions are adjacent to the first β-hairpin, which contains residue 18. At the mutation site, no large changes in RMSF are observed. There is no increase in calculated solvent accessibility suggesting that the core of the protein would not be exposed and there would be no additional hydrophobic patches in γS-G18A when compared to WT (**Supplementary Figure 1**). The ψ backbone angle of A18 is identical to that of G18, whereas the ϕ angle is slightly larger and more broadly distributed (**Fig. 2.4**). Although it is difficult to determine the exact significance of this change, alanine and valine exhibit a near identical distribution of ψ-ϕ torsion angles in lens β- and γ-crystallins (**Supplementary Figure 3**). Like V18 of γS-G18V, the torsion angle of A18 is dissimilar to all native β- and γ-crystallin alanines. Therefore, even though it is not as bulky or hydrophobic, alanine may be expected to cause similar changes in the protein backbone conformation at the mutation site. In γS-G18V the torsion angle changes results in altered backbone hydrogen bonding with the subsequent β-strand relative to wild-type. This change, however, is not observed for γ-G18A. The absence of significant structural differences between γS-G18A and γS-WT as observed in the MD simulations does not necessarily preclude the possibility of increased aggregation propensity or α-crystallin binding, as it is well-established that structural changes to the lens γ-crystallins are not necessary to alter protein stability and aggregation in the lens. [157, 162, 159, 160]

2.2.2 NMR chemical shifts of γS-G18A indicate only mild structural perturbations

The ^{1}H-^{15}N-HSQC [176] spectrum of γS-G18A (**Fig. 2.5A**) was measured in order to compare the NMR chemical shifts, a sensitive probe of the local chemical environment, to those of γS-WT and γS-G18V, which were measured previously. Chemical shift assignments were

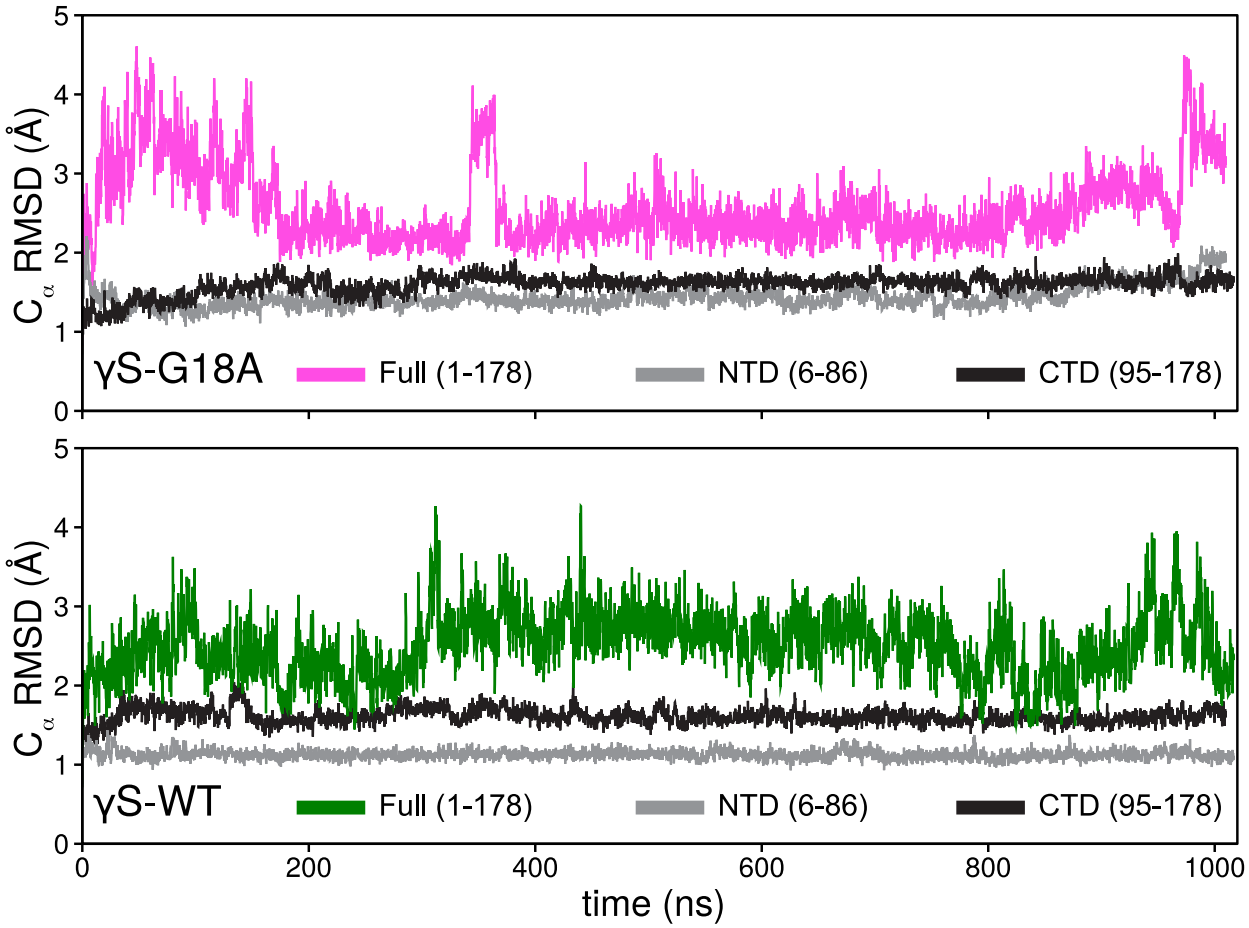

Figure 2.3: C_α RMSD from the initial configuration. The time evolution of γS-G18A predicts a well-folded protein much like γS-WT. Full = RMSD of all α-carbons. CTD = RMSD of residues 95-178. NTD = RMSD of residues 6-86.

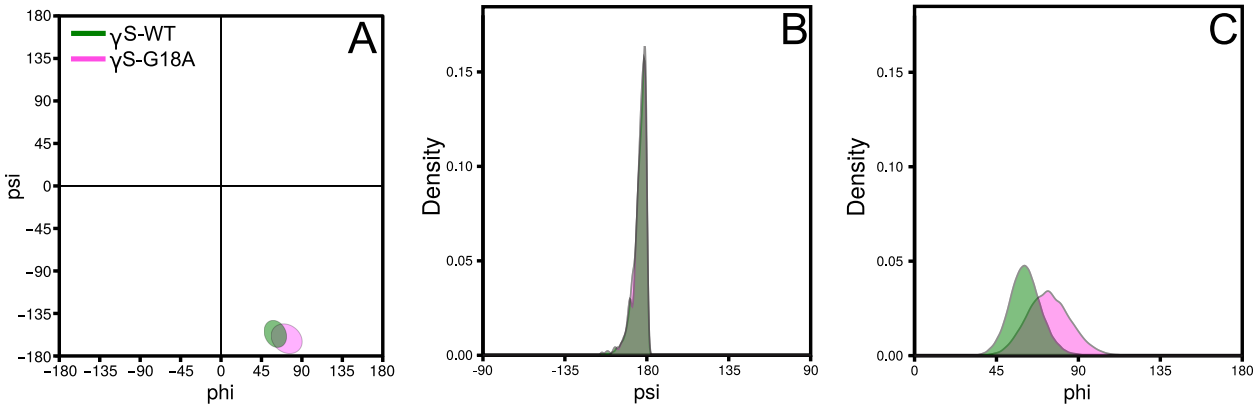

Figure 2.4: Backbone torsion angles for lens γS-crystallin residue 18. (**A**) The ψ-ϕ populations of γS-G18A are similar to those of γS-WT. (**B**) The ψ torsion angle distributions are almost identical, while (**C**) the γS-G18A ϕ torsional angle distribution is broader that in γS-WT.

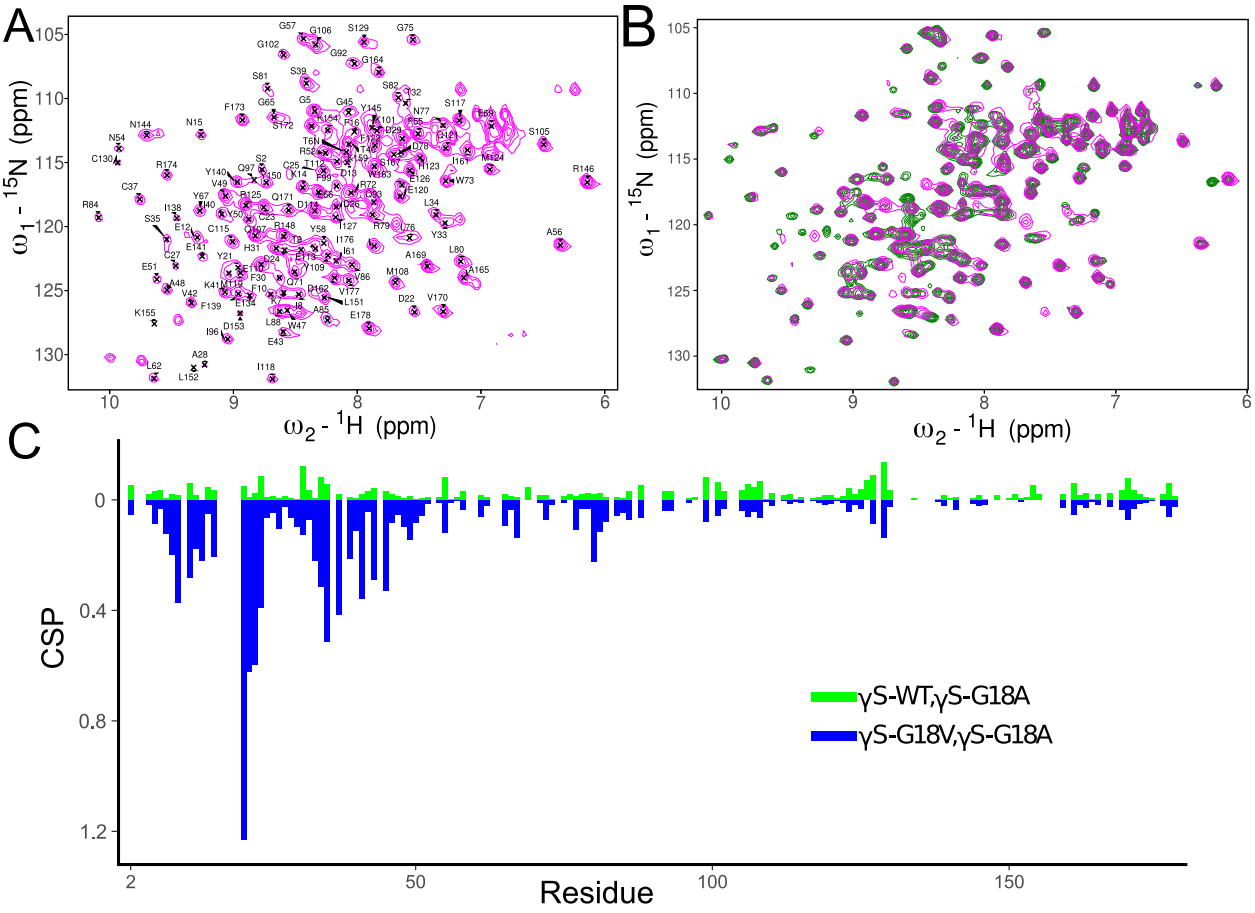

Figure 2.5: (**A**) ^{1}H-^{15}N HSQC spectrum of γS-G18A. Peaks were assigned by comparing the spectrum of γS-G18A to the previously assigned spectra of γS-WT and γS-G18V [177]. (**B**) Overlaid ^{1}H-^{15}N HSQC spectra of γS-WT (green) and γS-G18A (magenta). The two spectra overlay very well but some peaks are in slightly different positions. (**C**) Comparison of the chemical shift differences between γS-G18A and γS-WT (green) and γS-G18A to γS-G18V (blue).

determined by comparison to the γS-WT and γS-G18V assignments from the BioMagResBank (BMRB) (Entries 17576 and 17582), in which 72 % of shifts were assigned. As a whole, the chemical shift perturbations (CSP) for γS-G18A indicate that the structure is much more similar to γS-WT than γS-G18V (**Fig. 2.5B,C**). The most pronounced differences are observed in the first 50 residues of the N-terminal domain. Even for these residues, the γS-G18A chemical shifts are still more similar to γS-WT than to γS-G18V. Residues 17-20 were not assignable, probably due to local structural rearrangement.

	Transition 1			Transition 2		
Protein	Midpoint [Gnd-HCl]	ΔG_{app}(kcal/mol)	m (kcalůmol^{-1}ůM^{-1})	Midpoint [Gnd-HCl]	ΔG_{app}(kcal/mol)	m (kcalůmol^{-1}ůM^{-1})
WT				2.14±0.01 M	4.07±0.01	-1.93±0.01
G18V	1.58±0.01 M	3.73±0.1	-2.36±0.01	2.3 ±0.01 M	4.43±0.01	-2.14± 0.01
G18A				2.11±0.01 M	3.68±0.01	-1.75±0.01

Table 2.1: Thermodynamic parameters as calculated from chemical unfolding in Gnd-HCl. The midpoint of unfolding is similar for all three proteins in the final transition, but γS-G18V shows a stable intermediate. The parameters were calculated from the lines shown in Supplementary Figure 4 where "m" represents the slope of each line. The midpoint of unfolding was read at the x-intercept for each line and ΔG_{app} was read at the y-intercept

2.2.3 γS-G18A is of intermediate stability and aggregation propensity, between γS-WT and γS-G18V

The stability of γS-G18A relative to γS-WT and γS-G18V was measured in response to chemical stress. Chemical denaturation was performed using guanidine hydrochloride and assessed using the 355/325 nm fluorescence ratio (**Fig. 2.6A**). γS-WT and γS-G18A follow a two-state unfolding model with near identical unfolding midpoints at 2.1 M. In γS-G18V, three state-unfolding is observed with an unfolding intermediate between ∼1 M and 2 M. The unfolding midpoint leading up to the intermediate is at 0.9 M and leading up to complete unfolding is 2.3 M. We hypothesize that γS-G18A does not exhibit an intermediate unfolding state because the minimal structural changes near the mutation site are not sufficient to catalyze unfolding of the N-terminal domain. This is corroborated by the "m" parameter for each protein, which is the dependance of ΔG on concentration of denaturant. For the first transition observed in γS-G18V, m is higher than any slope in the second transition suggesting that some part of the protein is more suseptible to denaturation by guanidine hydrochloride. The δG_{app} of unfolding (**Table 2.1**) for each protein was calculated by using a linear approximation of a two-state unfolding curve as previously described[178]. γS-G18V was separated into two, two-state unfolding curves to generate its thermodynamic properties. ΔG_{app} of unfolding was estimated by extrapolating a line of best fit for each data set to 0 [178] (**Supplementary Figure 4**).

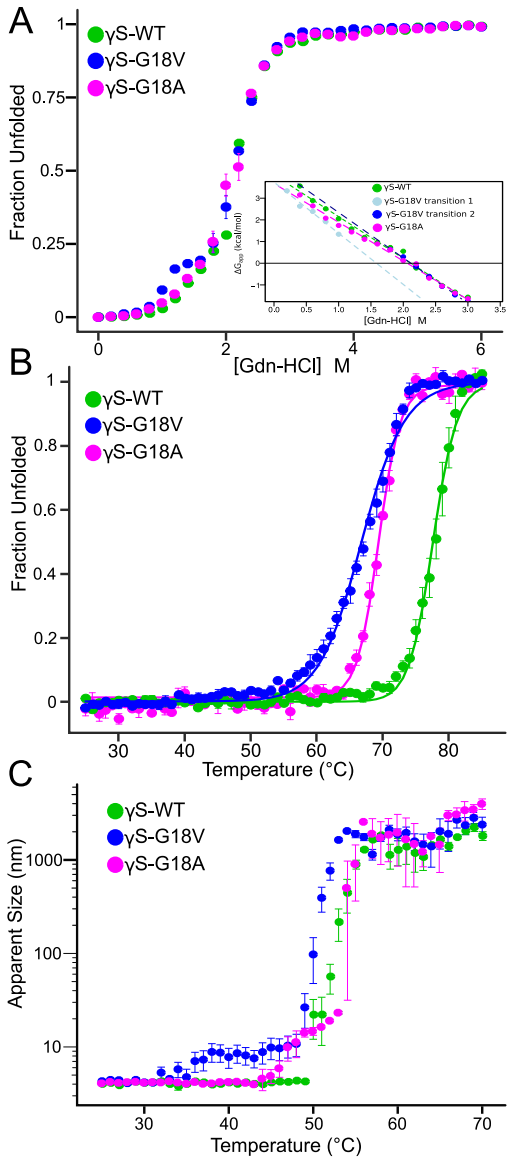

Figure 2.6: The stability and aggregation propensity of γS-G18A are more similar to γS-WT than to γS-G18V, although decreased stability is observed. (**A**) Chemical stability of γS-WT, γS-G18A, and γS-G18V was measured by taking the ratio of 355/325 nm after incubation in increasing concentrations of guanidine hydrochloride. Both γS-WT and γS-G18A show very similar results. However, γS-G18V is the least stable protein, exhibiting a three-stage unfolding pattern. The inset shows the linear extrapolation used to calculate the parameters in Table 1.(**B**) Thermal unfolding curves were generated by monitoring the CD intensity at 218 nm as a function of temperature for γS-WT (green), γS-G18A (magenta), and γS-G18V (blue). γS-WT is the most stable protein with a thermal unfolding midpoint temperature (T_m) of 77.67±0.03 °C. The γS-G18A variant is of intermediate stability, with a T_m of 69.87±0.05 °C, while the γS-G18V variant is the least stable with a T_m of 66.74±0.04 °C. (**C**) Thermal aggregation propensity was measured by DLS. Both γS-G18A and γS-G18V are aggregation-prone relative to wild-type, however, γS-G18V forms small aggregates at physiological temperature, where γS-G18A remains monomeric. Each experiment was repeated three times.

Thermal denaturation as measured by circular dichroism (CD) as a function of temperature (**Fig. 2.6B**). As previously observed, γS-WT is highly stable, with a unfolding midpoint at 77.67±0.03 °C whereas γS-G18V has the lowest unfolding midpoint at 66.74±0.04 °C. γS-G18A is of intermediate stability, with a unfolding midpoint at 69.87±0.05 °C. Although γS-G18A is nearly as robust as γS-WT with respect to chemical denaturation, its thermal stability of γS-G18A is noticably depressed.

Although γS-crystallin has a high thermal unfolding midpoint, thermal aggregation occurs at much lower temperatures, probably due to transiently unfolded states. Using dynamic light scattering (DLS), we measured protein aggregation as a function of temperature to investigate the behavior of γS-G18A (**Fig. 2.6C**). Between 44 °C and 47 °C, aggregates ≤ 10 nm in diameter are observable. These aggregates exceed 20 nm by 53 °C, after which large, insoluble aggregates ≥ 1000 nm are observed. The formation of intermediate size aggregates of γS-G18A is similar to that of γS-G18V, but at elevated temperature. In contrast, γS-WT does not exhibit any aggregation until 51–53 °C, when large, insoluble aggregates rapidly form. The similarity in thermal unfolding and aggregation of γS-G18A to γS-G18V suggests that γS-G18A has a similar aggregation pathway to γS-G18V, but shifted such that its onset occurs above physiological temperature.

2.2.4 αB-crystallin does not bind γS-G18A

Our characterization of γS-G18A indicates that protein stability is reduced compared to γS-WT, however, not as dramatically as it is in γS-G18V. Similarly, the aggregation propensity is intermediate, with an onset temperature around 45 °C. Moreover, γS-G18A does not appear to undergo large structural changes and does not exhibit a large increase in hydrophobic surface exposure. We next investigated how γS-G18A would interact with αB-crystallin to learn more about how mutated, but essentially structurally unmodified, γ-crystallins

behave as client proteins. Specifically, the objective was to determine whether the holdase chaperone αB-crystallin could discriminate between γS-G18A and γS-G18V. To this end, ^{1}H-^{15}N-HSQC experiments were performed using mixed samples of ^{15}N-labeled γS-G18A and natural abundance human αB-crystallin.

αB-crystallin forms high-molecular weight complexes that enable it to keep bound client proteins soluble.[147]. Therefore, if αB-crystallin were able to recognize and bind γS-G18A, then the NMR spectra of the latter should display an increase in line-width and a change in chemical shift for residues that participate in binding to αB-crystallin, as previously observed for γS-G18V.[161] The HSQC spectra (**Fig. 2.7A**) of ^{15}N-labeled γS-G18A in the presence of natural abundance αB do not exhibit these changes, instead showing only modest chemical shift perturbations (**Fig. 2.7B**) compared with γS-G18A alone, even as temperature increases. The vast majority of the residues have a CSP of less than 0.05 ppm at both 25 (**Fig. 2.7C**) and 37 °C (**Fig. 2.7D**). Using a threshold of 2*RMS, the CSP of only 3 residues at 25 °C and only 4 residues at 37 °C are large enough to suggest significant conformational change. In both cases these residues with the highest CSP are in the NTD suggesting that there might be some very weak and transient interaction between the NTD of γS-G18A and αB-crystallin.

When comparing the NMR solved structure of γS-WT to the NMR structure of γS-G18V and the predicted structure of γS-G18A there are some key structural differences that could account for the changes in solubility and αB-crystallin selectivity (**Fig. 2.8**). It has been previously shown that the binding interface for αB-crystallin includes the 3 cysteine residues on γS-G18V [163]. In this regard γS-G18A is much more similar to γS-WT (**Fig. 2.8A**), so it is possible that the conformation of this region of γS-G18A does not promote interactions with αB-crystallin, while the altered conformation seen in γS-G18V does (**Fig. 2.8B**).

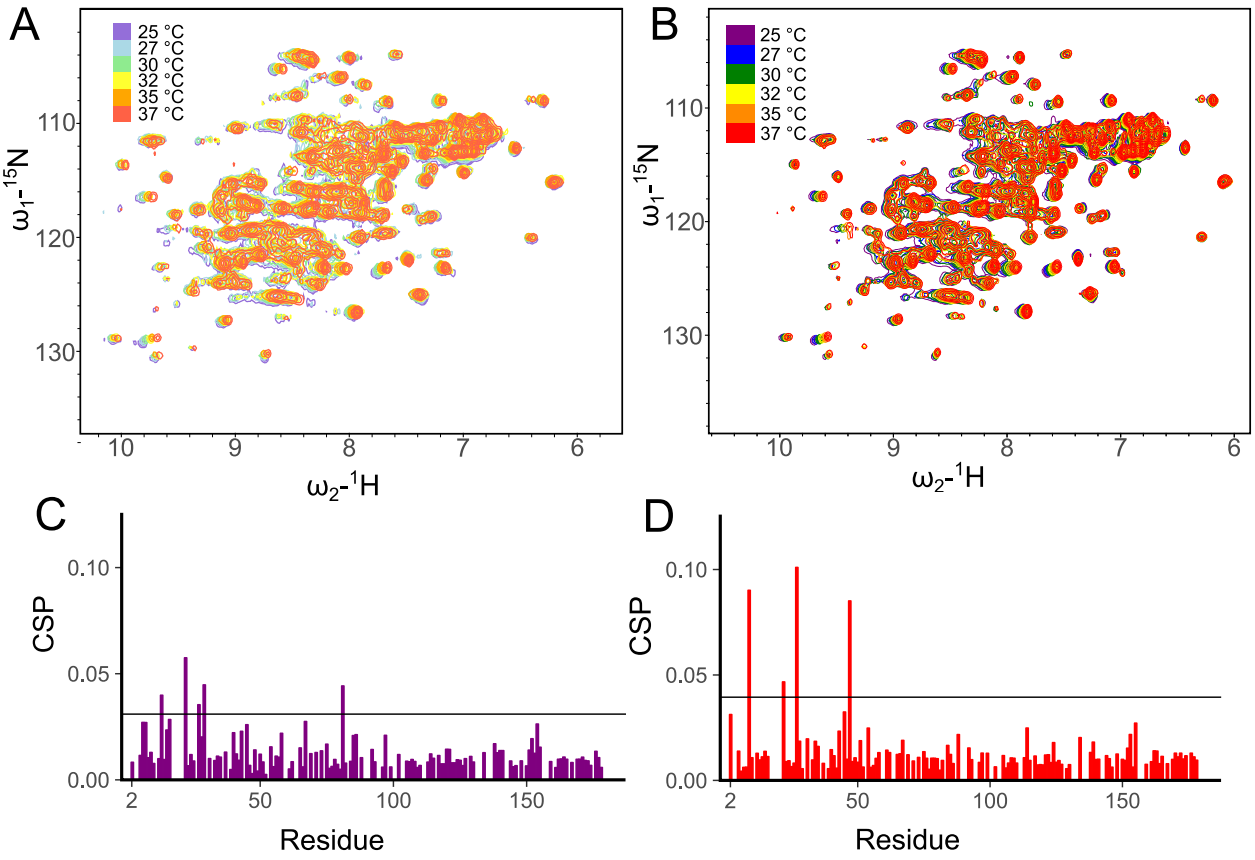

Figure 2.7: αB-crystallin does not specifically bind γS-G18A. (**A**) ^{15}N-HSQC spectra of ^{15}N-labeled γS-G18A as a function of temperature. (**B**) ^{15}N-HSQC spectra of ^{15}N-labeled γS-G18A mixed with natural abundance αB-crystallin. The spectra are almost identical, suggesting little interaction between γS-G18A and αB. (**C**) Observed CSPs for γS-G18A compared to γS-G18A+αB at 25 °C indicate very little change in conformation, suggesting minimal interaction. Only four residues have a CSP greater than $2 * RMS$, residues 13, 22, 29, and 81. (**D**) a similar observation is made at 37 °C where only four residues show a CSP greater than $2 * RMS$, these residues are 9, 22, 27, and 47.

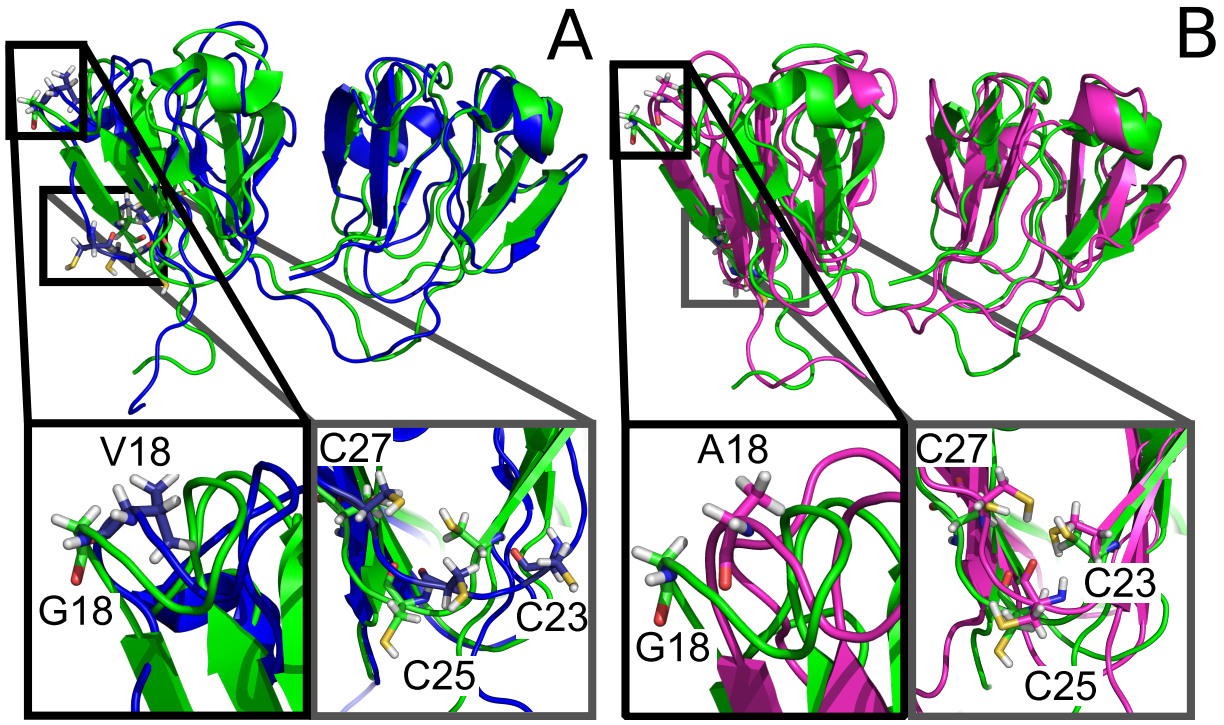

Figure 2.8: The predicted structure of γS-G18A is very similar to that of γS-WT. (**A**) The overall structures of γS-WT (PDB ID: 2M3T) and γS-G18V (PDBID: 2M3U) are very similar but there are some key differences. The loop region around the mutation site exhibits an altered conformation suggesting altered salt-bridging interactions with this loop region. The cysteine residues in γS-G18V are pointed out into the solvent and not turned into the protein like in γS-WT. (**B**) Comparing a predicted structure of γS-G18A generated from molecular dynamics simulations to the NMR structure of γS-WT shows some important similarities. Although γS-G18A has a slightly altered conformation in the loop region around the mutation site, the disturbances are not sufficient enough to alter the conformation of the loop containing cysteines 23, 25 and 27, in contrast to γS-G18V, where these cysteines are more exposed to solvent.

2.2.5 Conclusion

CD spectra and NMR chemical shifts suggest that the structure of γS-G18A is much more similar to γS-WT than to γS-G18V. However, the protein is thermally sensitive and there was a noticeable increase in aggregation propensity upon heating. Although it might be expected that a protein of intermediate aggregation potential between γS-WT and γS-G18V would also bind αB-crystallin to some degree, this is not the case for γS-G18A. This indicates that αB-crystallin is capable of discriminating very subtle structural changes in its client proteins, binding to the aggregation-prone G18V variant while leaving the function-preserving G18A variant untouched.

2.3 Materials and Methods

2.3.1 Protein expression and purification

The G18A variant of human γS-crystallin was generated using standard site directed mutagenesis. The cDNA encoding each protein and containing an N-terminal 6x-His tag was ligated into a pET28a(+) vector (Novagen, Darmstadt, Germany) and transformed into Rosetta *E. coli* for expression. All natural abundance proteins were expressed via β-D-1-thiogalactopyranoside induction. A culture of Lysogeny broth (LB) was an inoculated from a starter culture at OD_{600} of $\sim$0.05 and grown to an OD_{600} of $\sim$1. The growth culture was grown at 37 °C to an OD_{600} of $\sim$1.2 and then induced with 0.5 mM IPTG. After 24 hours at 25 °C the cells were harvested via centrifugation at 4000 rpm. The cells were lysed by sonication and subsequently centrifuged. The lysate supernatant was purified by nickel affinity chromatography and digested using TEV protease (produced in house) to remove the 6x His tag. The TEV protease and 6x His tag were removed by a second round of nickel affinity chromatography. Finally, size exclusion chromatography using a Superdex-75 SEC column was run to confirm purity. Protein masses were further confirmed using a Waters Xevo XS-QTOF.

For ^{15}N-labeled protein expression, a culture of *E. coli* in natural abundance LB was grown to an OD_{600} of $\sim$1.2. The culture was spun down and the pelleted cells were resuspended in an equivalent volume of M9 minimal media. The culture was allowed to grow for 1 h at 16 °C and before being induced with β-D-1- thiogalactopyranoside at a final concentration of 0.5 mM. The cells were then grown at 18 °C for 36 hours. Purification was performed using the same procedure as the unlabeled abundance protein.

2.3.2 ANS Fluorescence

γS-WT, γS-G18V, and γS-G18A samples at 1 mg/mL (10 mM sodium phosphate, pH 6.9) were incubated with 750 μM of 1-Anilinonaphthalene-8-Sulfonic Acid (ANS) for 1 hour at room temperature. Spectra from 450–500 nm were collected on a Cary Eclipse Fluorescence Spectrophotometer (Agilent Technology Inc., Santa Clara California, USA) using a 390 nm excitation and 5 nm slit widths.

2.3.3 Circular dichroism (CD) and thermal unfolding

The circular dichroism of all proteins were collected using a J-810 spectropolarimeter (JASCO, Easton MD). Samples were prepared at 0.1 mg/mL (10 mM sodium phosphate, pH 6.9) and measured over the 190–250 nm wavelength range at RT. Thermal denaturing of proteins was assessed via the absorbance at 218.0 nm. A temperature ramp of 2 °C/min was applied over the temperature range of 25–85 °C using 5 second equilibrations and 1 °C increments. Thermal denaturation experiments were performed at a sample concentration of 0.25 mg/mL (10 mM sodium phosphate, 150 mM NaCl, 1 mM DTT, pH 6.9). The 50% unfolding point was calculated by fitting each curve to a two-state unfolding model as described previously [162].

2.3.4 Tryptophan fluorescence and chemical unfolding

The intrinsic tryptophan fluorescence of all proteins was measured using a SpectraMax GeminiEM spectrometer (Molecular Devices, USA). Samples were excited at 295 nm and emission intensity was recorded from 300–400 nm. All samples were prepared at 0.1 mg/mL in 10 mM sodium phosphate, 50 mM sodium chloride, 0.05% sodium azide, pH 6.9.

The chemical unfolding via guanidine hydrochloride were assessed by 355/325 fluorescence ratio. Protein samples at 0.1 mg/mL were prepared in increasing concentrations of guanidine hydrochloride up to 6 M. The reaction mixtures were equilibrated at room temperature for 48 hours. Unfolding curves were generated by calculating the fluorescence ratio of 355 nm to 325 nm.

2.3.5 Dynamic light scattering

Protein thermal aggregation was measured via dynamic light scattering (DLS) using a Zetasizer Nano-ZS (Malvern Analytical, Malvern, United Kingdom). The number mean was used to assess particle sizes across a temperature range of 25–80 °C. Protein was diluted to 1 mg/mL in 10 mM sodium phosphate, 0.05% sodium azide, pH 6.9.

2.3.6 Solution-state NMR

^{15}N-HSQC experiments were conducted using an ^{1}H -^{13}C-^{15}N 5 mm tri-axis PFG triple resonance probe attached to a Varian $^{\text{textitUnity}}$INOVA spectrometer (Agilent Technologies) operating at 800 MHz (Oxford Instruments). Proteins were concentrated to 35 mg/mL and then supplemented with 2mM TMSP and 10% D_2O. ^{1}H chemical shifts were referenced to TMSP and ^{15}N shifts were referenced indirectly to TMSP. NMR data were processed using NMRPipe[179] and analyzed using NMRFAM-Sparky[180].

2.3.7 αB-crystallin binding assay

Lyophilized αB-crystallin was resuspended in 10 mM sodium phosphate at pH 6.9 and mixed in a 2:1 ratio with the ^{15}N-labeled γS-G18A variant of γS-crystallin at a concentration of 1.5

mM. A series of ^{15}N-HSQC experiments were then performed at 25, 27, 30, 32, 35, and 37 °C. The sample was incubated at the experimental temperature for 1 hour and the ^{15}N-HSQC experiment was run for 2 hours at each temperature. The chemical shifts collected were compared to a sample of a ^{15}N-labeled γS-G18A variant of γS-crystallin at a concentration of 1.5 mM. Chemical shift perturbations (CSP) were calculated using the following formula:

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

2.3.8 Molecular dynamics simulations

The top configuration of the human γS-crystallin solution-state NMR structure[161] (PDB ID 2M3T) was used as the initial protein configuration in the γS-WT simulation and in the γS-G18A simulation after performing the Gly to Ala substitution at residue 18. The γS-WT simulation system consisted of a single protein chain, 11,493 water molecules and a single sodium counterion to neutralize the protein net charge for a total of 37,347 atoms. The initial simulation cell size was 63.573 x 81.480 x 77.315 Å^3. The γS-G18A simulation system consisted of a single protein chain, 12,860 water molecules and a single sodium counterion for a total of 41,451 atoms. The initial simulation cell size was 81.52 x 67.71 x 80.34 Å^3.

The simulations were performed with NAMD 2.9[181]. Each simulation system was subjected to 10,000 steps of conjugate-gradient energy minimization followed by a 1-ns MD run at constant pressure (1 bar) and a constant temperature (300 K) over which the protein was released from its initial configuration in a stepwise manner using harmonic restraints on all backbone heavy atoms. The simulations were then run for 518 ns (γS-WT) and 1048 ns (γS-G18A) at constant temperature (300 K) and pressure (1 bar). The CHARMM36 force field[182, 183, 184] was used for proteins and ions and the TIP3P[185] model was used for

waters. The smooth particle mesh Ewald method[186, 187] was used to calculate electrostatic interactions. Short-range, real-space interactions were cut off at 12 Å by means of a switching function. A reversible, multiple time-step algorithm[188] was used to integrate the equations of motion with a time step of 4 fs for electrostatic forces, 2 fs for short-range nonbonded forces. The integration time step for bonded forces was 1 fs in the γS-WT simulation an 2 fs in the γS-G18A simulations . All bond lengths involving hydrogen atoms were held fixed using the SHAKE[189] and SETTLE[190] algorithms. A Langevin dynamics scheme was used for temperature control, and a Nosé-Hoover-Langevin piston was used for pressure control.[191, 192]

2.3.9 Ramachandran analysis

The observed phi and psi angles for glycine, alanine, and valine were determined by collecting all available native vertebrate lens $\beta\gamma$-crystallins from the PDB (40 structures). Angles were calculated by using the DSSP package and processed using R.[193, 194]

Funding

This work was supported by National Institutes of Health Grants 1R01EY025328 to R.W.M. and D.J. Tobias and 2R01EY021514 to R.W.M. KWR was supported by National Science Foundation Grant DGE-1633631. RWM is a CIFAR fellow.

Acknowledgement

The authors thank Dmitry Fishman for excellent management of the UCI Laser Spectroscopy Labs and Hartmut Oschkinat for helpful discussions. We would also like to thank Anne Diehl and Hartmut Oschkinat for generously providing the αB-crystallin sample.

Chapter 3

Investigations into the Structure of γS-WT at High Concentrations

3.1 Introduction

The crystallins were named by Berzelius in 1830 after the apparent crystalline quality of a substance isolated from eye lenses [6]. This name is misleading as the proteins of the lens do not actually form a crystal and the wild-type crystallins are notoriously difficult to crystallize. What the name does capture is the fact that the lens does form a firm, glassy structure. This structure comes from the proteins of the lens forming a hydrogel due to the high concentration of these proteins in the lens.

The concentration of proteins in the eye lens varies depending on the organism the lens is from and what part of the lens is being examined. For example, mammalian lenses can have protein concentrations up to 400 mg/mL [3], while the lenses from aquatic species exhibit protein concentrations exceeding 1000 mg/mL [195]. Similarly the protein concentration of the cortex is much lower than what is found in the nucleus [196].

The majority of studies performed on the crystallins are done at concentrations far lower than what is found in the lens. This is mainly due to experimental constraints, as performing many kinds of experiments at concentrations greater than 300 mg/mL is not feasible. For example, the first structure of γS-crystallin was solved by solution-state NMR, done at a concentration of 40 mg/mL [197]. Other kinds of biophysical characterization are done at even lower concentrations, in the 0.1–1 mg/mL range. This includes light scattering techniques[112], circular dichroism[112, 162], fluorescent techniques[112], and so on. All of these are at best two orders of magnitude lower in concentration that what is found in the lens. The result of this is that much of what is known about the crystallins is their behavior at very low concentrations, where the effects of molecular crowding are not in effect.

Molecular crowding can have a profound effect on a protein. These effects can range from changing the folding dynamics of the protein to changing its enzymatic proprieties. For example hen egg white lysozyme will refold much faster under crowding conditions as a different folding pathways is preferred [198]. T4 DNA polymerase binds DNA with a much higher affinity under crowding conditions [199]. With regard to the crystallins, both α-crystallin and γB-crystallin exhibit unique protein-protein interactions at high concentration due an increase in particular kinds of anisotropic interactions [200]. α-crystallin exhibits a resistance to aggregation under crowding conditions as the nucleation stage of aggregation is inhibited [201].

This has naturally lead to the question of how do the crystallins behave under very high concentrations, similar to what is found in the intracellular milieu of the lens fiber cells. There is a clear difference in the physical properties of the crystallins as concentration increases. In dilute solution human γS-crystallin are primarily monomeric [202], and weakly repulsive protein-protein interactions dominate [203]. This is in stark contrast to the firm gel that forms in the lens at very high concentrations, where protein-protein interactions would be responsible for holding the gel together.

I hypothesize that there are important changes in the oligomerization state and the structure of the protein to allow for hydrogel formation at high concentrations. To examine this I used both solid-state and solution-state NMR to investigate the structure of γS-crystallin at the residue level and use Cryo-EM to characterize the bulk hydrogel that is formed at high concentrations.

3.2 Results and Discussion

3.2.1 Solution-State NMR of γS-Crystallin

To first understand how concentration can affect the biophysical properties of γS-crystallin we performed a series of ^{15}N-HSQCs at increasing concentration (**Fig. 3.1**). At higher concentrations the spectrum look identical to what is seen at lower concentrations in terms of peak position allowing for easy assignment of the spectra (**Fig. 3.2**). This result suggest that as concentration increases there is little change to the structure of the protein.

While there appears to be little change in structure of γS-crystallin, the more crowded environment found in a higher concentration solution, could result in altered relaxation mechanisms brought on by different protein-protein interactions. For example if the proteins begin to form complexes then you would expect a decrease in T_2. As molecular motions slow and different relaxation mechanisms become available causing the net magnetization vector to return to equilibrium more quickly. Spin-spin interactions become prominent if the local fluctuations in the magnetic field around a nucleus slow below the Larmor frequency of that nucleus resulting in a higher degree of field inhomogeneity and therefore the transverse magnetization of individual spins dephase with each other at a shorter time scale. From the ^{15}N-HSQCs performed we measured the ^{1}H linewidths as decreasing relaxation times would

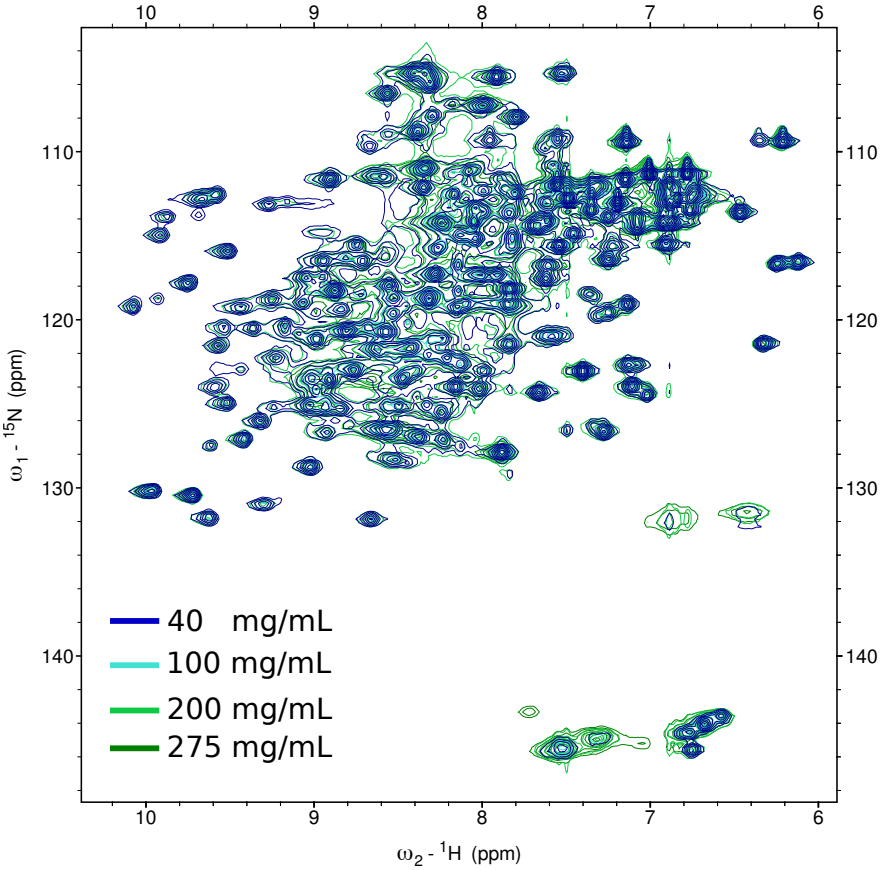

Figure 3.1: **The ^{15}N-HSQC of γS-crystallin is largely unchanged as concentration increases.** A ^{15}-HSQC At (Blue) 40 mg/mL, (Teal) 100 mg/mL, (Light Green) 200 mg/mL, and (Dark Green) 275 mg/mL. There is a high degree of similarity between each of the spectra.

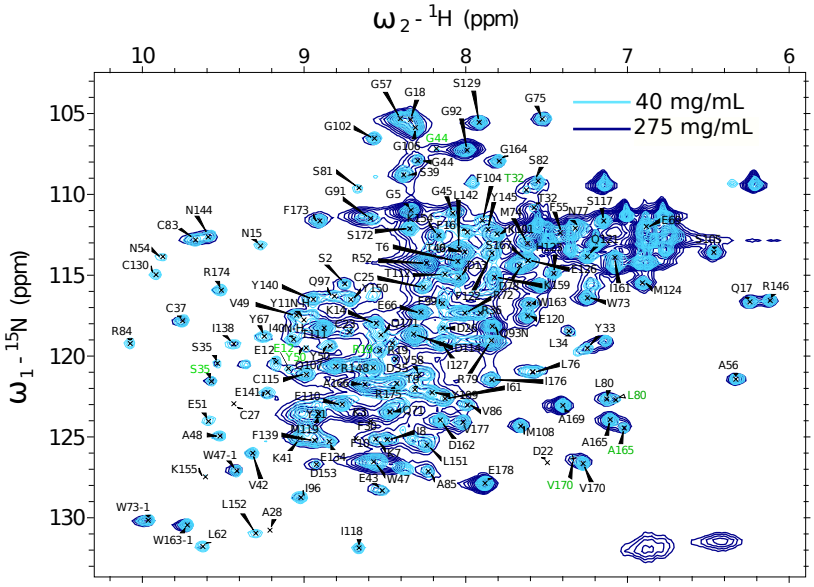

Figure 3.2: **Solution state NMR of γS-crystallin at high and low concentration.** ^{15}N-HSQC of γS-WT at (light blue) 40 mg/mL reveals a well-dispersed peak pattern, expected of a well folded protein. At (dark blue) higher concentrations (275 mg/mL) there is the same overall placement of peaks but there does appear to be an increase in linewidth. Alternate conformations of different residues are labeled in green.

result in an increase in peak width (**Fig. 3.3).** This result is consistent with the γS-crystallin molecules interacting with each other more frequently due to the close proximity.

3.2.2 Solid-State NMR of γS-Crystallin Hydrogel

γS-Crystallin was concentrated to point that a hydrogel was formed, at approximately 475 mg/mL and we examined the hydrogel using magic angle spinning solid-state NMR. Cross polarization (CP) and INEPT experiments were performed to get an idea on the mobility of the protein and individual residues. In CP experiments the chemical shift of nuclei can be measured if the sample being measured exhibits slow molecular motions, due to dipolar-coupling based magnetization transfer. A rigid solid could give a very high resolution spectra when combined with magic angle spinning. INEPT experiments on the other hand transfer magnetization directly through bonds due to J-coupling and require rapid molecular motions

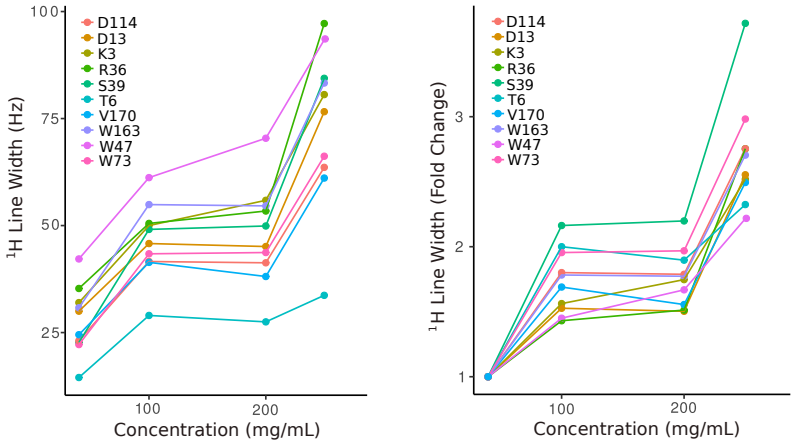

Figure 3.3: **Linewidth increases as concentration increases**. (A) The proton linewidth of selected peaks at 40, 100, 200, and 275 mg/mL. In general the peaks begin to broaden as concentration increases suggesting that the relaxation time for γS-crystallin is increasing. (B) The fold change in line width was plotted as well to better illustrate how the linewidths were changing.

for this transfer mechanism to be favored. A gel with a high degree of liquid like character would how ^{13}C- and ^{15}N-CP experiments reveal very little signal, with any signal noticed simply being an artifact of the rotor (**Fig. 3.4**). As CP experiments highlight polarization transfer through dipolar couplings a highly mobile sample would not show much signal in a CP experiment. In contrast INEPT relies on magnetization transfer through J-couplings and would be easy to measure if dipolar couplings are averaged to zero. The ^{15}N-INEPT performed by MAS-ssNMR shows very little resolution and very broad peaks making it very difficult to make claims on the structure of the protein at the level of individual residues (**Fig. 3.5**). The overall contours of the spectrum overlaps well with the solution state-HSQCs performed and this suggests that the overall state of the protein is similar to that seen in dilute solution. The ^{13}C-INEPT seems to show a higher resolution and better peak separation than the ^{15}N-INEPT upon inspection (**Fig. 3.6**). The higher resolution offered by the ^{13}C-INEPT is possibly do to a higher degree of flexibility in the side chains when compared to the backbone of the protein, as faster molecular motions increase resolution in INEPT experiments. The dispersion of peaks for the ^{13}C-INEPT suggests that the fold of the protein is well folded and similar to the structure found in dilute solution. Similar

to the ^{15}N-INEPT, the overall contours of the spectrum is very similar to what is seen in solution-state NMR. Taken together these suggest that the structure of γS-crystallin in a hydrogel is very similar to what is seen in dilute solution.

3.2.3 Cryo-EM of the Hydrogel Shows Different Phases and Hints at a Maturation Process

What we can gather from the structural information gathered from NMR on γS-crystallin in a hydrogel is that the structure of the protein does not change as protein concentration increases. This is surprising as there is a noticeable change in the bulk properties of the protein at high concentration, when a gel is formed. To better understand what is the topology of this hydrogel, the structure of the gel was examined with cryo-EM.

γS-crystallin was concentrated down to 625 mg/mL and prepared for cryo-TEM. When compared to vitreous ice, the hydrogel shows a collection of filament type structures (**Fig. 3.7**). This suggests that the gel like properties of the sample are due to interactions between these filaments. A similar sample that was left to age for 8 months in the gel state was also prepared for cryo-TEM and revealed a larger particles of higher density protein (**Fig.3.8**).

Taken together this suggests that the formation of a firm gel takes time to mature and under crowding conditions protein-protein interactions of γS-crystallin are altered to promote the formation of filaments and aggregates that maintain a solution like quality and do not scatter light. How these aggregates manage to maintain transparency is still a question that needs active research.

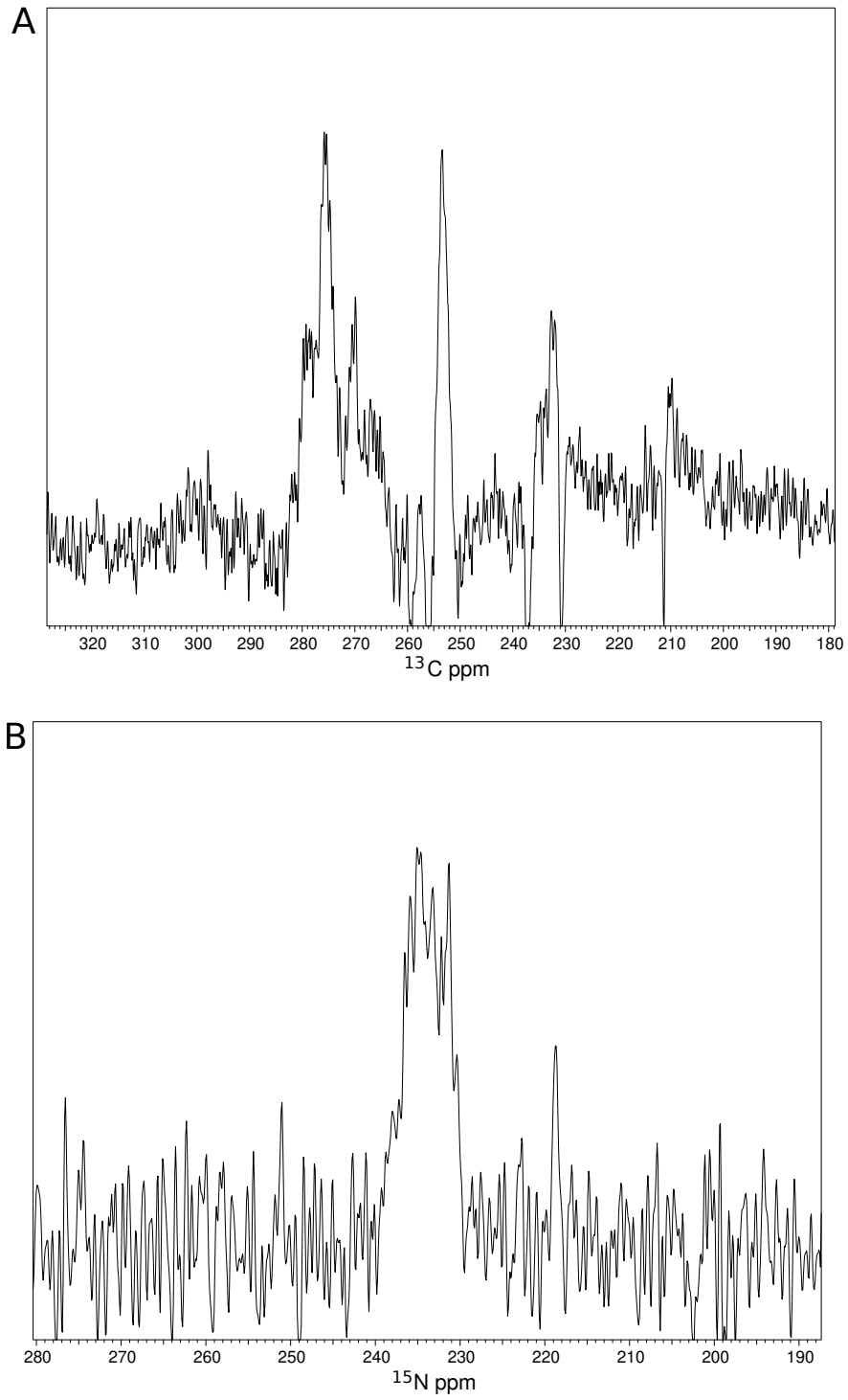

Figure 3.4: **1D Cross polarization experiments of gelled crystallin.** Cross polarization experiments show no signal for either (A) carbon or (B) nitrogen This suggests that there is a high level of mobility in the sample.

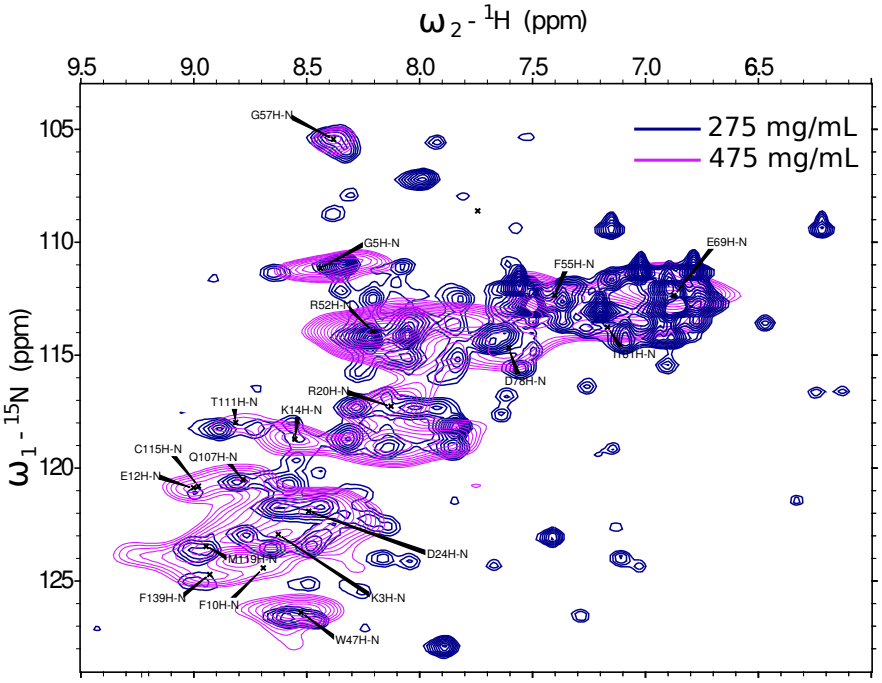

Figure 3.5: **^{15}N-HSQC reveals very broad lines suggesting slower molecular motions.** Compared to the HSQC done in solution (Blue) at 275mg/mL the solid-state INEPT (Magenta) performed at 475 mg/mL has very broad lines and low resolution.

3.3 Conclusion

While there is still much work to be done some important things have been revealed in this study. Firstly the molecular structure of γS-crystallin likely not significantly different in a hydrogel or dilute solution. Second is that the hydrogel itself is not perfectly homogeneous, and instead has an internal structure that changes over time. What still needs to be understood is how the structure of γS-crystallin contributes to the formation of the aggregates and fibrilar structures seen in the gel. This includes how these structures are different from light scattering aggregates seen during cataract formation.

These results mark the first biophysical characterization of γS-crystallin as a hydrogel, providing valuable information on how this protein might behave in the eye lens and as a proof-of-concept for further studies on other lens proteins, aggregation prone variants of these proteins, and client-chaperone interactions at physiologically relevant concentrations.

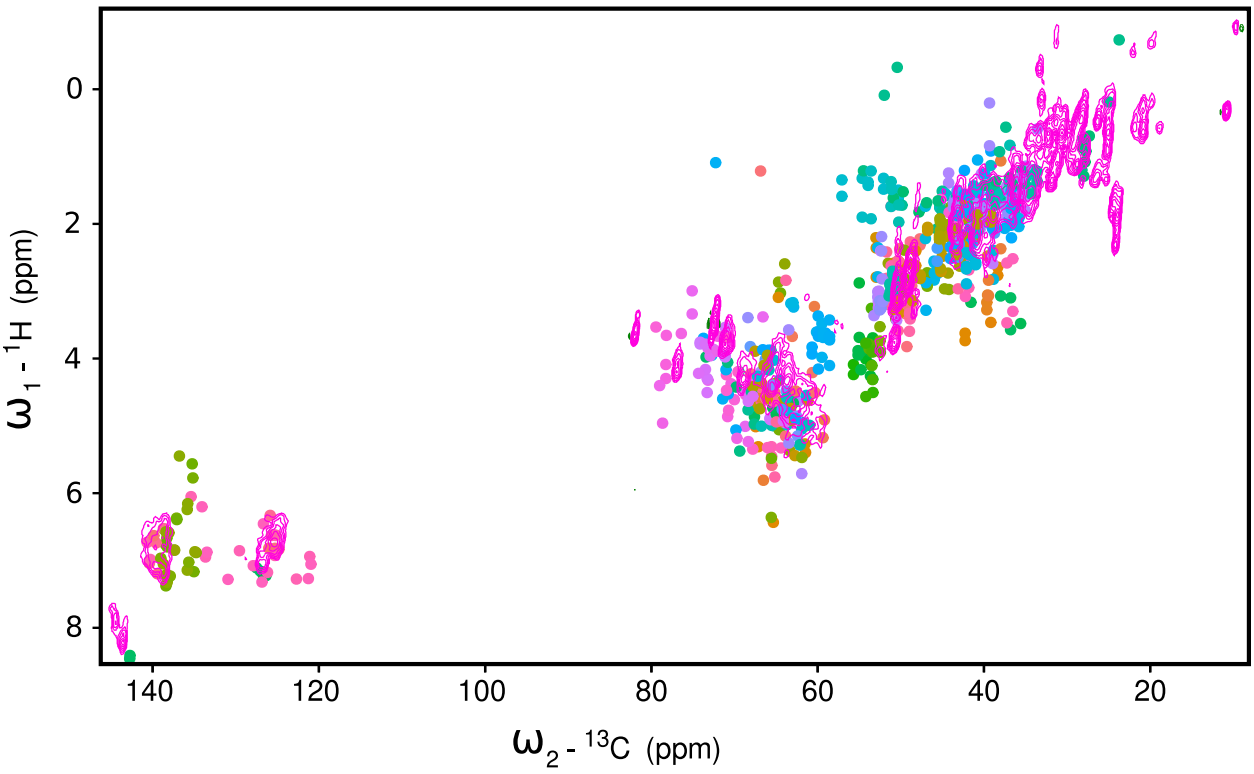

Figure 3.6: **^{13}C-INEPT reveals broad lines and has poor resolution but matches nicely with a simulated ^{13}C-INEPT.** The measured ^{13}C-INEPT (A) is difficult to interpret but when compared to a (B) simulated ^{13}C-INEPT generated from published chemical shifts [177] the two spectra do coincide.

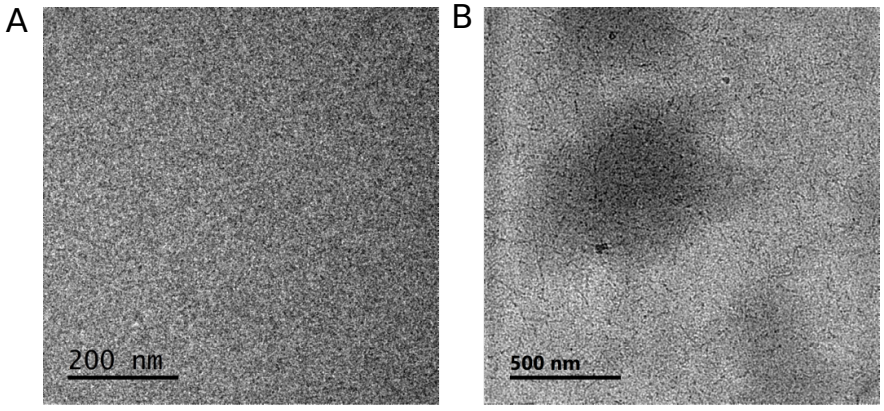

Figure 3.7: **γS-crystallin hydrogel is comprised of filament like structures.** When compared to a preparation of vitreous ice (A), γS-crystallin at a concentration of 625 mg/mL (B) appears to be comprised to a collection of filament-like structures.

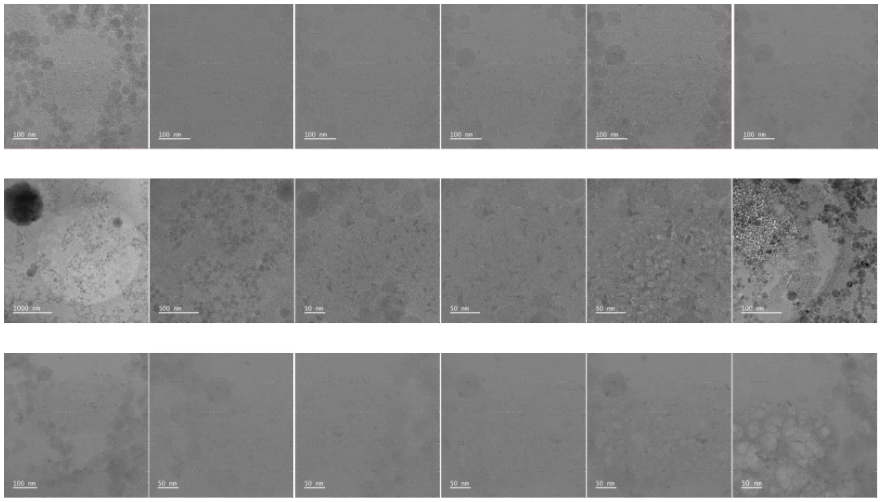

Figure 3.8: **Cryo-EM images of highly concentrated γS-crystallin reveal higher density protein phases.** Looking at a collection of different Cryo-EM images of γS-crystallin hydrogel that has been aged 13-months reveals different phases of protein concentration in the sample that appear to be aggregates.

3.4 Materials and Methods

3.4.1 Expression and Purification of Human γS-Crystallin

Human γS-crystallin in the pET-28+(a) vector was transformed into rosetta cells. After inoculation from an overnight starter culture, the growth culture was grown at 37 °C until an OD_{600} of 1.2 at which point the culture was induced with 1 mM IPTG and grown overnight at 37 °C. For ^{15}N, ^{13}C labeled protein the growth culture was also grown to an OD_{600} at which point it was spun down and resuspended in an equivalent amount of M9 minimal media enriched with ^{15}N ammonium chloride and ^{13}C glucose. This culture was grown at 37 °C for 1 hour and then induced with 1 mM IPTG. This culture was grown at 16 °C for 48 hours. In each case the cells were harvested and lysed through sonication. Purification was performed by a round of nickel affinity chromatography, the 6xHIS tag was cleaved by TEV protease, a round of nickel affinity chromatography to remove leftover TEV protease, and a final round of size exclusion chromatography. The pure protein was

concentrated with Amicon centrifugal filters with a 3000 MWCO and formed into a gel with VIVASPIN 500 centrifugal filters with a 10000 MWCO.

3.4.2 Solid-State NMR

The ^{15}N INEPT + experiment was performed using the following parameters: number of scans = 480, sweep width = 21306.818 Hz, acquisition time = 0.0149248 sec. The protein was concentrated to a hydrogel from a more dilute solution buffered with 10 % D20 10 mM phosphate pH 6.9. The solids data was collected at 25C, spinning at 10K in a 3.2 mm rotor on a 700 MHz Bruker Avance II. Data was processed with TopSpin and assignments were done in NMRFAM-Sparky. All cross-polarization experiments were performed on the same sample with the same rotor, spinning speed and magnet. Data processing was performed in TopSpin

3.4.3 Solution-State NMR

All the ^{15}N-HSQC experiments were performed with the following parameters: water suppression was performed with the 3-9-1-9 watergate program, 10% D2O in 10 mM phosphate pH 6.9, referenced to 2 mM TMSP, number of scans = 16, sweep width = 12019.23 Hz, acquisition time = 0.085 seconds, Temp = 25C, performed on the Varian Inova 500 MHz magnet. Data was processed with NMRPipe and assignments were done with NMRFAM-Sparky

3.4.4 Cryo-TEM

Cryo-TEM samples were prepared using a, Automatic Plunge Freezer ME GP2 (Leica Microsystems) and Quantafoil R2/2 TEM grids (Electron Microscopy Sciences) which were

plasma cleaned for ten seconds in a 50:50 Ar:O2 mixture (Gatan Solarus, model 950). 7 μL of a 625 mg/mL wildtype crystallin sample was deposited on the grid in a controlled environment of 98% relative humidity at 20 °C. Immediately after deposition, the grid was blotted for six seconds and plunged into liquid propane at -180 °C. Grids were loaded into a Gatan Cryo-TEM holder and imaged using a JEM-2100F TEM operating at 200 keV. Images were recorded using DigitalMicrograph (Gatan) software with a Gatan OneView CMOS camera at 4k by 4k resolution. Aged crystallin gel was concentrated until a gel formed and then stored at 4 °C for 13-months before preparation for Cryo-TEM.

Chapter 4

The Droserasin 1 PSI: A Membrane-Interacting Antimicrobial Peptide from the Carnivorous Plant *Drosera capensis*

4.1 Introduction

Plant carnivory has been considered as part of defense mechanisms involving the jasmonate pathway [204]. In most cases, the leaves are modified for capturing prey, using a variety of mechanisms including pitfall traps, the unique snap trap mechanism of the Venus flytrap, and sticky flypaper traps, among others, and all of these plants must perform their digestion without any mechanical breakdown of prey tissues, such as mastication. In species with flypaper traps, such as the *Drosera*, digestion generally occurs in an exposed environment over a prolonged period of time, without the benefit of a pitcher or closed trap. These

plants catch prey in the sticky polysaccharide mucilage of their leaf tentacles, which then wrap around the meal to increase contact with the digestive mucilage [205]. The digestion is thus exposed to variable physical conditions from changes in weather and has increased risk of opportunistic microbial growth from bacteria and fungi that compete for nutrients from the captured prey and potentially infect the plant tissue, causing disease [206]. In the Droseracaea, the evolution of carnivory was accompanied by the loss of many genes common to other plants, and the concomitant expansion of genes specifically related to carnivory [207]. For these reasons, carnivorous plants are a potential source of novel and useful antimicrobial peptides as well as digestive enzymes. Here, we focus on an antimicrobial peptide discovered from the genome of the Cape sundew, *Drosera capensis*. This plant is relatively easily cultivated and has been the target of genome sequencing [208] and enzyme discovery [209, 210, 211] efforts.

Analysis of putative digestive enzyme sequences from *D. capensis* revealed several aspartic proteases that contain a segment of about 100 residues called a plant specific insert (PSI). PSIs are mostly-helical domains that are often cleaved off during maturation and act as independent proteins [212]. Structurally, PSIs are categorized as saposin-like proteins, a protein family whose members have substantial sequence diversity but share a strongly conserved, compact tertiary fold, usually stabilized by three disulfide bonds [213]. Although the *S. tuberosum* PSI is monomeric in solution [214], it was crystallized as a dimer [212] and appears to oligomerize upon interaction with anionic membranes [215, 216]. Membrane permeabilization can induce cytotoxicity and is a common antimicrobial defense. PSI-containing aspartic proteases from *D. capensis* may have been recruited for digestive function due to their ability to digest prey proteins as well as inhibit microbial growth during digestion. These functions may be separate or synergistic: PSIs in the digestive mucilage could simply serve to suppress the growth of pathogens, or may also function to make insect lipids more available for digestion.

In other plants, aspartic proteases containing PSIs are implicated in stress responses, senescence, and pathogen responses [217, 218, 219]. Their overall function involves interacting with lipid membranes in a variety of ways, including membrane localization, increasing the availability of membrane lipids for enzymatic processing, and permeabilizing membranes [220]. For example, in the cardoon, the PSI from the aspartic protease cardosin A plays a role in vacuole localization [221, 222] and induces vesicle leakage below pH 5.5, whether or not it is attached to the parent aspartic protease [223]. Similarly, the well-studied PSI from the potato *Solanum tuberosum* disrupts lipid vesicles and bilayers [212] and has been shown to exhibit antimicrobial activity against plant and human pathogens [224]. Here we report the expression and characterization of recombinantly expressed Droserasin 1 PSI (D1 PSI) from *D. capensis*. Our results show that the D1 PSI forms a compact, stable structure and is indeed capable of disrupting and permeabilizing membranes and inhibiting microbial growth.

4.2 Materials and Methods

4.2.1 Sequence Alignment and Clustering

All sequences from the *Drosera capensis* genome [208] and the *Dionaea muscipula* transcriptome [225] that were previously annotated as coding for MEROPS A1 aspartic proteases using the MAKER-P (v2.31.8) pipeline [226] and a BLAST search against SwissProt (downloaded 8/30/15) and InterProScan [227] were examined for the presence of a PSI. Sequence alignment with the previously- characterized PSI from the *Arabidopsis thaliana* protease APA1_ARATH was used for quality control; proteins that did not contain a full-length PSI were not selected for modeling or further analysis. ClustalOmega was used to produce sequence alignments [228] for putative aspartic protease PSIs. The following settings were used: gap open penalty = 10.0, gap extension penalty = 0.05, hydrophilic residues =

GPSNDQERK, and a BLOSUM weight matrix. This resulted in six complete PSIs from *D. capensis* and two from *D. muscipula*. Three previously-characterized aspartic proteases from other plants are also included as reference sequences [229, 230, 231]. The resulting PSI sequences were clustered by sequence similarity.

4.2.2 Structure Prediction

Structures for the PSIs were predicted using the Robetta [232] implementation of Rosetta [233]. This software uses a combination of comparative modeling and all-atom refinement based on a simplified forcefield, yielding five structures for each protein. For the PSIs, both open and closed conformations were observed in the models: we selected the lowest-energy representative of each type of structure for each PSI. The open conformation was observed as part of a dimer structure, which was also employed for subsequent modeling of the D1 PSI as described below. This process was performed for the *D. capensis* and *D. muscipula* PSIs, as well as a reference PSIs from *Arabidopsis thaliana*. The PDB files corresponding to the predicted structures for the PSIs reported in this manuscript are available in the Supplementary Information. Protein structure figures were generated using UCSF Chimera [234] and VMD [235].

4.2.3 Gene Construction, Expression, and Purification

Plasmids containing the DNA sequence of the *D. capensis* Droserasin 1 PSI (D1 PSI) genes were purchased from Integrated DNA Technologies (San Diego, CA, USA). Each gene was flanked by regions containing restriction sites for NcoI and XhoI and contained an N-terminal 6xHis tag and a TEV protease cleavage sequence (ENLYFQG). The gene was amplified using oligonucleotide primers purchased from Integrated DNA Technologies (Coralville, IA, USA), and the resulting gene product was cloned into pET28a(+) vector (Novagen, Darmstadt,

Germany). D1 PSI was overexpressed in SHuffle T7 *Escherichia coli* (New England Biolabs, Ipswich, MA, USA) using an autoinduction protocol [236] at 25 °C with the modification of adding 50 μM IPTG [237]. Cells were allowed to grow for at least 24 h, lysed via sonication, heated at 70 °C for 20 min to precipitate most *E. coli* proteins, and cell debris was removed by centrifugation. His-TEV-D1 PSI was purified on a Ni-charged Bio-Scale Mini Profinity IMAC Cartridges (Bio-Rad, Hercules, CA, USA) where bound protein was washed with low imidazole wash buffer followed by washing with 40% isopropanol wash buffer mixture and 40% DMSO wash buffer mixture to remove potentially bound lipids before elution. The His tag was removed with the use of a His-tagged TEV protease (produced in-house), over one week, (the time required to obtain a sufficient yield of His tag-free protein), followed by reapplication to Ni-charged Bio-Scale Mini Profinity IMAC Cartridges (Bio-Rad, Hercules, CA, USA) to remove His-tagged TEV protease and uncleaved His-TEV-D1 PSI. The final purification step consisted of applying the sample to a HiLoad 16/600 Superdex 75 pg gel filtration column (GE, Pittsburgh, PA, USA) in 10 mM phosphate buffer. D1 PSI was dialyzed into 10 mM phosphate, 0.05% NaN_3, pH 6.9, for all experiments unless otherwise stated. The mass of the protein was confirmed by electrospray mass spectrometry. For ^{13}C, ^{15}N labeled protein samples, protein was expressed using an optimized high-cell-density IPTG-induction minimal media protocol [238]. Purification was performed in the same manner as for natural abundance protein.

4.2.4 Circular Dichroism

D1 PSI was diluted to 0.125 mg/mL with either 10 mM succinate, acetate, MES, or phosphate buffer at pH 4, 5, 6, and 7, respectively, for the collection of full circular dichroism (CD) spectra between 190 and 260 nm in a 10 mm quartz cuvette. Three accumulations were collected and no smoothing function was applied to the collected data. Additional spectra were taken under the same conditions but where the D1 PSI was reduced in 1 mM DTT

before dilution into the final buffer. Measurements were taken on a J-810 spectropolarimeter (JASCO, Easton, MD, USA) equipped with a thermal controller.

4.2.5 Fluorescence Spectroscopy

UV fluorescence measurements were made on D1 PSI at a concentration of 0.2 mg/mL under the same buffer conditions as for CD spectra including unreduced and DTT-reduced protein for full emission spectra, with an excitation wavelength of 280 nm. Spectra were taken using a Cary Eclipse Fluorescence Spectrophotometer (Agilent, Santa Clara, CA, USA).

4.2.6 Characterization of Oligomeric State

0.5 mg/mL of purified D1-PSI with 6x His-tag were incubated in 50 mM ammonium formate for pH 3.5, 50 mM ammonium acetate for pH 4, 5, and 6 or 50 mM ammonium bicarbonate for pH 7 and 8. Intact protein samples were diluted 5-fold in water and then run on the SYNAPT G2-Si Mass Spectrometer (Waters, Milford, MA, USA) via direct injection in a 80/20% 0.1% formic acid/ACN running buffer. Samples were ionized by electrospray ionization with a capilary voltage of 2.7 kV and then separated by the T-Wave ion mobility collision cell (Waters, Milford, MA, USA). The ion series generated was then deconvoluted using the MaxEnt1 algorithm supplied by the MassLynx (Waters Milford, MA, USA) software package to calculate the mass of the proteins at each pH.

4.2.7 Antimicrobial Assay

Antimicrobial activity of D1 PSI was measured by growing *Pichia pastoris* in standard yeast extract, peptone, dextrose media (YPD) with the addition of 50 mM sodium phosphate and

adjusted to pH 4, 5, 6, or 7, with or without 25 μM D1 PSI, peptide concentration was selected to match previously reported experiments [224]. Two replicates were performed at each pH value. Cells were grown in 15 mL culture tubes in a final culture volume of 1 mL. Cultures were inoculated with a pregrown culture, resulting in a starting optical density (OD) of 0.005 at 600 nm, and were allowed to grow for 48 h at 30 °C. After growth total cell yield was estimated by measuring the OD at 600 nm.

4.2.8 Vesicle Fusion Assay

Lipids used for experiments were extracted from either *E. coli* or *P. pastoris* using the Bligh and Dyer method [239]. The lipids were then dried under a stream of nitrogen gas and polar lipids were extracted using a cold acetone precipitation [240]. The *E. coli* or *P. pastoris* polar lipids were solubilized by repeated heating and cooling, between 42 °C and room temperature, in either 10 mM succinate pH 4, 10 mM acetate pH 5, 10 mM MES pH 6, or 10 mM phosphate pH 7. The solution was then run through a mini extruder equipped with a 100 nm polycarbonate membrane (Avanti Polar Lipids, Alabaster, AL, USA) to create large unilamellar vesicles (LUV) of approximately 100 nm in diameter. Vesicle size was monitored over time at the different pH with or without 50 μM D1 PSI using dynamic light scattering on a Zetasizer Nano ZS (Malvern Instruments, Malvern, U.K.).

4.2.9 Lipid Interaction Quantification

1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (POPC), 1-palmitoyl-2-oleoyl-sn-glycero-3- phosphoethanolamine (POPE), 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) sodium salt (POPG), 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine sodium salt (POPS), and 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphate sodium salt (POPA) were purchased from Avanti Polar Lipids (Alabaster, AL, USA). A solution of 10 mM sodium acetate pH 4.5 containing

75 μM of POPC, POPE, POPG, POPS, and POPA each was divided into two aliquots. One aliquot had His-tagged D1 PSI added to a final concentration of 0.04 mg/mL. Both solutions were sonicated for an hour at room temperature and applied to a Ni-charged Bio-Scale Mini Profinity IMAC Cartridges (Bio-Rad, Hercules, CA, USA). The aliquot lacking D1 PSI had the flow-through collected, while the elution was collected from the aliquot containing D1 PSI. Lipids were extracted from the flow-through and elution respectively using the Bligh and Dyer method [239]. A series of ten-fold dilutions were made for each and the resulting samples were run on a Xevo G2-XS QTof spectrometer (Waters, Milford, MA, USA) in positive mode, with an in-line ACUITY UPLC BEH C18 Column (Waters, Milford, MA, USA), where lipids were eluted with a water/isopropanol gradient containing ammonium formate. Only the sample concentrations within the linear range of ion intensities were used. Ion intensities were tabulated for each lipid species and normalized to the total ion count to estimate relative lipid proportions.

4.2.10 Solid-State NMR

Samples were prepared by evaporating away chloroform from 5 mg of POPC and POPA, respectively, under a stream of nitrogen gas. Lipids were sonicated in 0.5 mL methanol to redissolve lipids then 0.5 mL of water was added followed by 2 mg of ^{13}C, ^{15}N labeled D1 PSI dissolved in water. Protein lipid mixture was briefly sonicated, flash frozen and lyophilized. Lyophilized sample was then hydrated with 10 μL 10 mM acetate, 0.025% sodium azide, pH 4.5 buffer. The resulting sample was cycled between 42 °C and room temperature ten times. Spectra were taken at the National High Magnetic Field Lab (Tallahassee, FL, USA) using the 40 mm bore Series Connected Hybrid (SCH) magnet system, currently the highest-field NMR magnet [241]. Two-dimensional ^{13}C-^{13}C cross polarization [242] dipolar assisted rotational resonance [243] (CP DARR) spectra were obtained at 36 T, with a 2 mm CPMAS

probe tuned to frequencies of ^{1}H, ^{13}C, ^{15}N, with ferroshims, at a temperature of 10 °C, a 100 ms mixing time, and a MAS rate of 24.4 kHz [241].

4.2.11 Molecular Modeling and Analysis

Predicted monomer and dimer structures of D1 PSI as described above were modeled within lipid bilayers using atomistic molecular dynamics (MD) simulations. Open monomer and PSI dimer structures were first adjusted by adding disulfide bonds based on homology to the potato (*S. tuberosum*) PSI, with protonation states corrected for pH 5 using PROPKA3 [244]. For each respective structure, a POPC membrane patch was prepared using the VMD membrane plugin; for the monomer structure, a 100 Å square patch was used, with a 150 Å patch employed for the PSI dimer. The membrane patch was then centered within a TIP3P water box [245] of dimension 80 Å normal to the patch, the PSI was added within the membrane center, and 0.1M NaCl was added. (All structure preparation was performed using VMD [235].) The prepared system was then equilibrated as follows. Initially, with all atoms other than those of the lipid tails were held fixed, the system was minimized for 5000 iterations and simulated at 300 K for 0.5 ns. (All simulations were performed in NAMD [246] using the CHARMM36 force field [247] under periodic boundary conditions in an NpT ensemble at 1 atm pressure with Nosé-Hoover Langevin piston pressure control [248, 249] and Langevin temperature control (damping coefficient 1/ps).) Water and lipid were then released (with the protein held fixed), and the system was then minimized for 1000 iterations and simulated for an additional 0.5 ns at 300 K; during this time, water was excluded from the lipid layer using the script of [250]. Following this, all atoms were released, and the system was equilibrated for 0.5 ns at 310 K. The final state of this trajectory was then used as the initial state for a production run of 100 ns at 310 K with frames collected at 10 ps intervals. This protocol was employed for both monomer and dimer structures.

Following simulation, water flow through PSI-induced channels was measured by counting transitions of water molecules through the membrane during each 10 ps observation window using a custom R [251] script using the `bio3d` library [252]. Underlying flow rates were estimated by Bayesian inference using an autoregressive latent rate model defined as

$$\lambda_i \sim t_\nu(\rho\lambda_{i-1}, \sigma) \qquad\qquad Y_i \sim \text{NegBin}(\lambda_i, \phi),$$

where λ_i is the log expected flow rate in window i (in units of counts per window), Y_i is the number of water molecules transitioning in window i, ν and σ are the degree of freedom and scale parameters for drift in the latent rate function, ρ is an autocorrelation parameter, and ϕ is the overdispersion parameter of the negative binomial distribution. (Student's t and negative binomial distributions were respectively employed because of the extremely bursty character of the flow process.) Weakly informative priors were used for all parameters, based on physically plausible limits for the processes in question; specifically, we take $1/\phi \sim \text{HalfCauchy}(0, 5)$, $\sigma \sim \text{HalfCauchy}(0, 5)$, $\nu \sim \text{Gamma}(2, 0.1)$, $\rho \sim \text{Normal}(0, 1)$, and $\lambda_1 \sim \text{Cauchy}(0, 5)$. Posterior inference was performed using No-U-Turn Hamiltonian Monte Carlo sampling [253] for 2000 draws using four independent Markov chains; simulation was performed using the `rstan` package [254]. In addition to rate estimation, marginal distributions of transition counts over all observation windows were also obtained; these were compared to the distributions arising from Poisson distributions with equivalent expected values to assess overdispersion.

4.3 Results and Discussion

4.3.1 *D. capensis* and Related Plants Contain Several Aspartic Proteases with PSIs

The genome of *D. capensis* contains at least six aspartic proteases with moderate sequence identity to mammalian pepsin (Droserasins 1–6) [208]. These proteases are classified as belonging to the MEROPS A1 class [255], which also includes pepsin and the nepethesins found in the digestive fluid of pitcher plants of the related genus *Nepenthes* [256, 257]. The PSIs (without the catalytic domains) from *D. capensis* and *D. muscipula* were clustered by protein sequence similarity (Figure 4.1). Three well-characterized PSI sequences from other plants were also included for reference: *C. cardunculus* Cardosins A and B (Uniprot IDs CARDA_CYNCA and CARDB_CYNCA, respectively) and *A. thaliana* APA1 (Uniprot ID APA1_ARATH) [229, 230, 231]. Previous studies have shown that recombinantly expressed APA1_ARATH is maximally efficient at pH 5.3, and has a highly specific cleavage profile with respect to the insulin β-chain [258]. The full-length droserasins share important functional sequence features with the vacuolar protein APA1_ARATH and the cardosins, including the active site residues, the disulfide bonding pattern, and the PSI. Protein sequence alignments comparing sequences of the PSIs described here can be found in Supplementary Figure S1. All the PSI sequences share a relatively high degree of sequence identity, but the two from *C. cardunculus* are much more similar to each other than to the *D. capensis*, *D. muscipula*, and *A. thaliana* sequences. The latter group further clusters into two subtrees, with three representatives from *D. capensis* and one from *D. muscipula* in each group. The Droserasin 1 PSI (D1 PSI) was chosen for further modeling and experimental characterization as a representative of the group that is less related to the previously characterized APA1_ARATH.

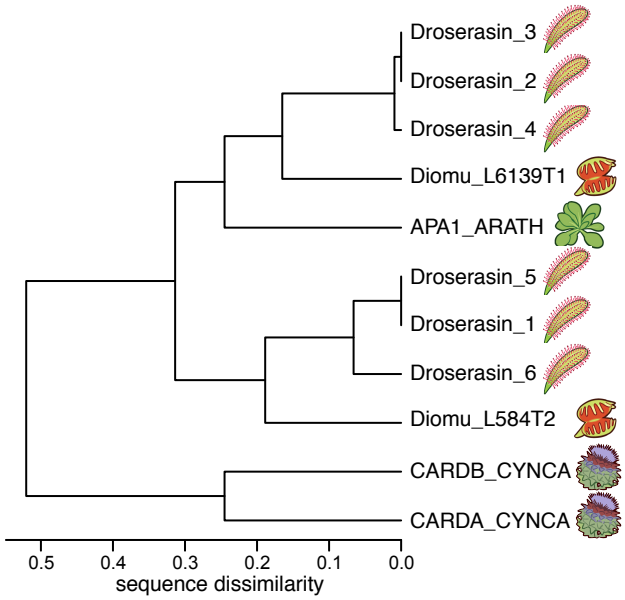

Figure 4.1: PSIs from *D. capensis* and *D. muscipula* clustered according to protein sequence similarity. Reference sequences from *Arabidopsis thaliana* and *Cynara cardunculus* are also included for comparison.

Saposin-like proteins, including plant aspartic protease PSIs, are very stable, in part due to the presence of disulfide bonds that lock the tertiary structure into place [259]. This fold has two characteristic conformations, a compact closed form (Figure 4.2A) and an extended open form (Figure 4.2B), both of which can be adopted by human saposin C [260, 261]. The crystal structure of the *S. tuberosum* PSI captures the open structure, observed as a domain-swapped dimer [212]. The corresponding model for the D1 PSI is shown in Figure 4.2C. The open conformation is proposed to be responsible for membrane-interacting activity due its increased exposure of hydrophobic residues compared to the closed conformation; in Section 4.3.7 we show via MD simulation that the open conformation of D1 PSI is indeed compatible with embedding in lipid bilayers. As we show, dimers formed from open-conformation monomers are also lipid-compatible, however; they have a distinct stabilization mechanism. The experiments and simulations that follow were performed in order to characterize the biophysical and membrane-interacting properties of the D1 PSI as a first step toward understanding its mode of antimicrobial activity.

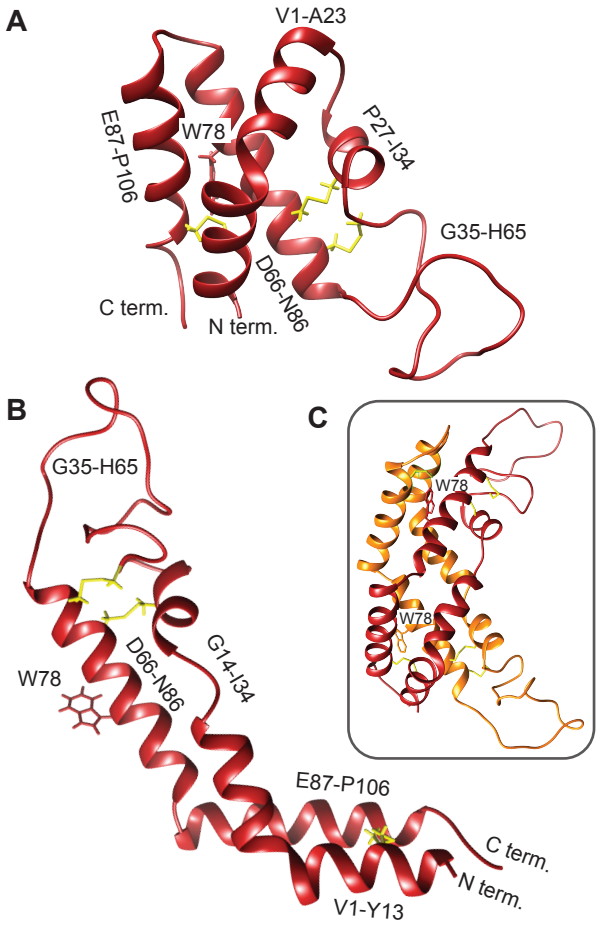

Figure 4.2: Comparative models of mature D1 PSI were predicted using the Robetta server [232, 233]. (**A**) Closed conformation. (**B**) Open conformation. (**C**) Predicted dimer in solution.

4.3.2 *D. capensis* D1 PSI Is Highly Stable

In order to test the thermal stability of D1 PSI, experiments were performed with protein recombinantly expressed in *E. coli*. Typical yields of 20 mg/L of bacterial growth were achieved. Purification and characterization data are shown in Supplementary Figures S2 and S3. CD spectra were collected under different pH and temperature conditions. Regardless of the pH, the CD spectra collected at 20, 55 or 90 °C show very little variance (Supplementary Figure S4), indicating that at least the secondary structure of this protein is extremely thermostable. Furthermore, after pretreating the PSI with DTT to reduce the disulfide bonds, only a marginal reduction of signal is observed, indicating that the secondary structure is perturbed very little even after its three presumptive disulfide bonds have been reduced.

To further probe the response of the D1 PSI to pH, temperature, and reducing agent, intrinsic tryptophan fluorescence spectroscopy was employed ((Figure 4.3). This technique can provide information on the local environment of the tryptophan as a more hydrophobic environment leads to a blue-shifted emission relative to a polar environment. The D1 PSI only has one tryptophan, allowing that particular position to be probed. Here, the intrinsic fluorescence was measured under the same conditions used for CD experiments. Under non-reducing conditions the wavelength of maximal emission is approximately 334 nm, regardless of temperature or pH, indicating that the single tryptophan (W78) is moderately exposed but insensitive to both pH and temperature. The emission intensity does decrease as a function of temperature, a known phenomenon in proteins [262]. When the D1 PSI has been reduced, there appears to be a slight shift in the emission maxima to approximately 337 nm, and more strikingly, the emission intensity increased relevant to the non-reduced PSI at equivalent temperature. A likely explanation could be that in the oxidized form, the tryptophan is next to a disulfide bond that is capable of quenching it, as in the closed conformation shown in Figure 4.2A but not the open monomer shown in Figure 4.2B, where the single Trp is solvent-exposed. Reducing the adjacent disulfide would release the quenching [263].

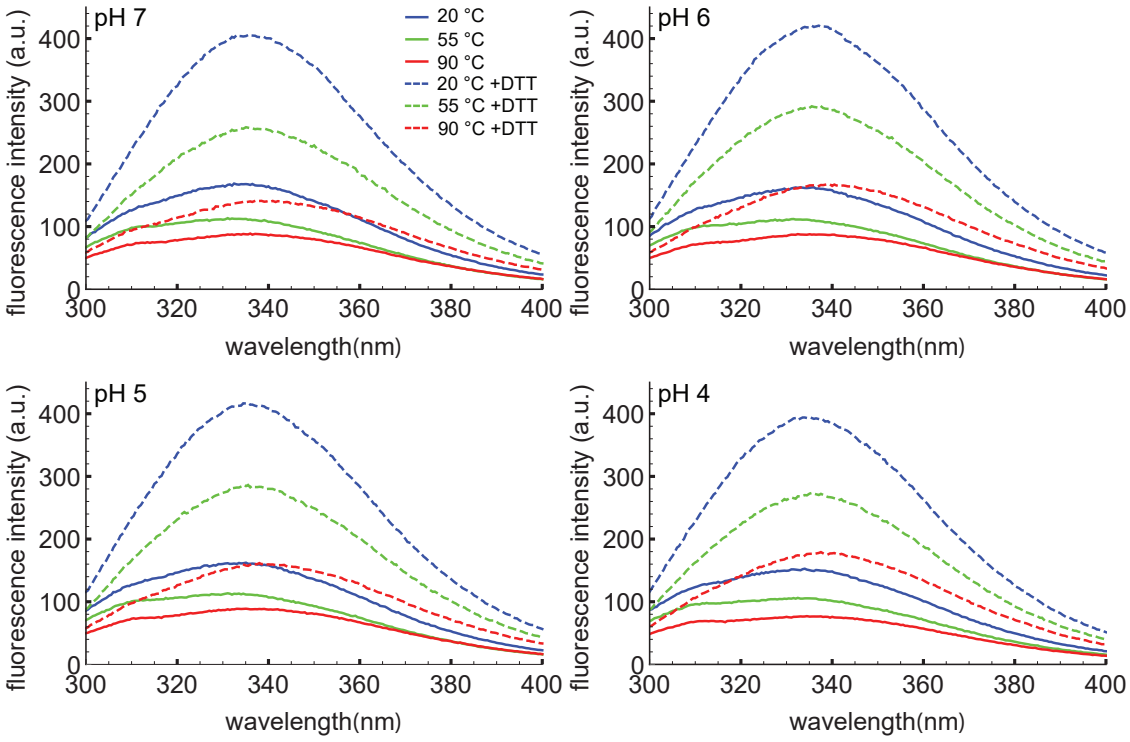

Figure 4.3: Intrinsic tryptophan fluorescence spectra of D1 PSI at pH 4, 5, 6, and 7. Spectra were collected at 20, 55 and 90 °C in the presence (dashed lines) or absence (solid lines) of DTT. The excitation wavelength was 280 nm. Under non-reducing conditions the emission maxima are at approximately 334 nm regardless of pH or temperature. Under reducing conditions the emission maxima for most of the spectra shift slightly to about 337 nm, indicating slightly more exposure of the single tryptophan. Reducing the protein also leads to an increase in the emission intensity, likely due to the reduction of a nearby disulfide bond that was quenching the tryptophan.

4.3.3 D1 PSI Is Monomeric in Solution over a Wide pH Range

There is experimental evidence that PSI of *Solanum tuberosum* forms dimers under low pH conditions [216], leading to the hypothesis that the activation of *D. capensis* D1 PSI requires dimerization at low pH values. To test this hypothesis, the presence of monomeric and dimeric D1 PSI was measured by intact protein mass spectrometry at a range of pH values, from 3 to 8 (Supplementary Figure S5). A dimeric form of D1 PSI would be expected to predominate at pH 3.5 and 4 but not at pH 7 and above if there is pH-dependent dimerization. For each pH value measured, D1 PSI was found to exist predominately in the monomeric form. Notably even at low pH values (pH 3.5 and 4) the monomeric form of the protein predominates. Also, the relative amounts of monomer to dimer do not change much as pH changes. Taken together we conclude that the oligomeric state of D1 PSI in aqueous solution is primarily monomeric, and is not strongly pH-dependent. This suggests that either the active form of D1 PSI is the monomeric state or that lipid bilayer interactions are required for dimerization.

4.3.4 D1 PSI Enables Vesicle Fusion at Acidic pH

Based on previous studies in the literature, the simulations described in Section 4.3.7, and preliminary results consistent with inhibition of microbial growth at pH 5 (Supplementary Figure S6), we hypothesize that the PSI permeabilizes the membrane. We therefore used a vesicle fusion assay to test whether and how the D1 PSI interacts with membranes. Dynamic light scattering (DLS) was used to monitor increasing vesicle size due to fusion from membrane disruption. First, 100 nm large unilamellar vesicles (LUVs) were prepared using *E. coli* polar lipid extract. At all pH conditions tested, LUVs were stable in size over time as shown in Figure 4.4A. Upon addition of D1 PSI the distribution of LUV size begin to shift to larger sizes, but only at pH 4 and 5: the pH 6 and 7 samples do not change in size. DLS

measurements of the D1 PSI alone were also recorded at each pH. For each sample, only one peak was present, corresponding to a size of about 2 nm, which was stable over time. When LUVs were made using yeast polar lipid extract instead, again the LUVs are stable over time at each pH (Figure 4.4B). When D1 PSI is added to the LUVs, only at pH 4 and 5 does the size distribution increase, similar to the results for the *E. coli* polar lipids. These findings are consistent with the antimicrobial assay in that D1 PSI interacts with membranes in a pH-dependent manner where it is only active at acidic pH. Another noteworthy observation is that with LUVs composed of *E. coli* polar lipids the rate of vesicle fusion was slow and gradual while in the case of the yeast polar lipids the change was very rapid at pH 4. Not only is the rate faster in this case, but for the pH 4 yeast polar lipid condition, smaller peaks appear at the latest time points. A potential hypothesis to explain the presence of these smaller particles after extended time is that it is possible that the D1 PSI is actually able to extract some of the lipids from the vesicle into small lipoprotein particles, consistent with the role of some saposins as surfactants.

4.3.5 D1 PSI Is Able to Interact with Lipids Having Diverse Head Groups

After showing that D1 PSI was able to interact with membrane lipids from a natural source that contains a variety of lipid species, we sought to gain insight into whether the D1 PSI preferentially binds specific polar lipids. The *S. tuberosum* PSI was found to interact with both anionic and neutral lipids, although its specific mode of membrane interaction does depend on the head groups, with more membrane disruption observed in negatively charged membranes [215]. Furthermore, the presence of cholesterol inhibits membrane fusion activity, making the PSI non-toxic to animal cells [214]. First, the affinity of D1 PSI for different phospholipid head groups was tested. Five phospholipid species containing different head groups, but the same acyl groups (a singly unsaturated 18-carbon acyl chain and saturated

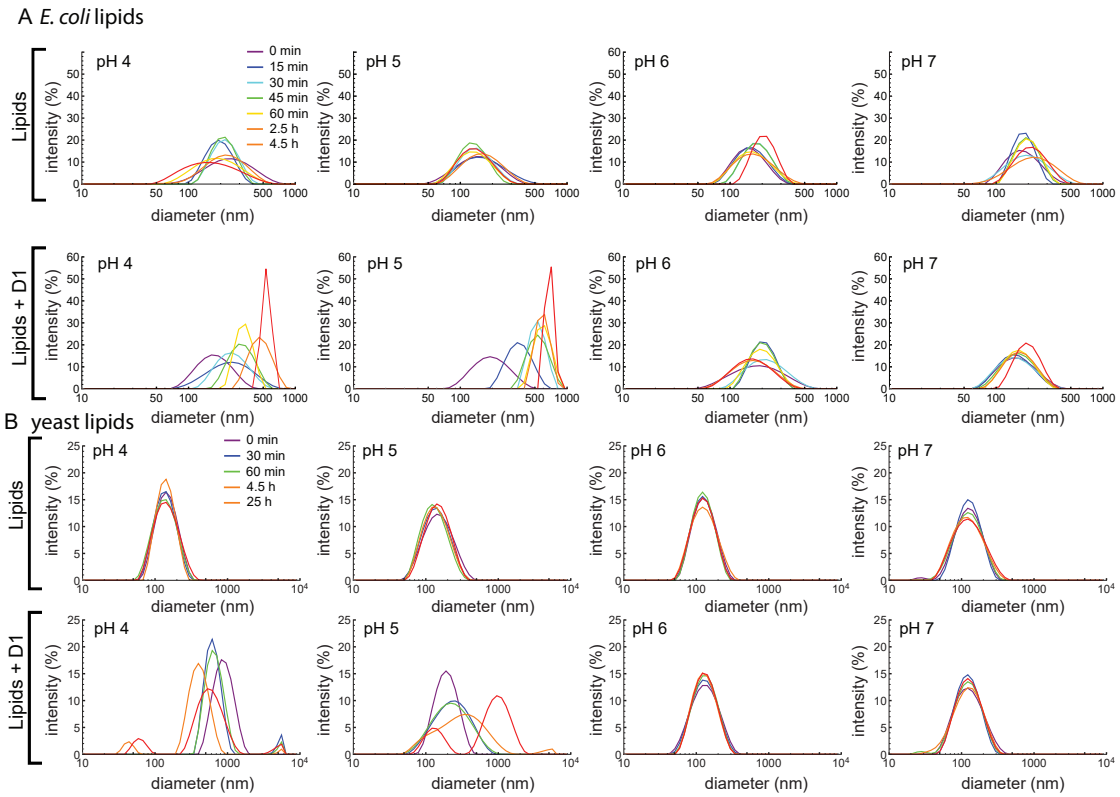

Figure 4.4: (**A**) LUVs of 100 nm were made using an *E. coli* polar lipid extract. LUVs were made in buffered solutions at pH 4, 5, 6, and 7. Vesicle size was monitored over time using DLS, either with (bottom) or without (top) D1 PSI. LUVs made with lipids alone are stable in size for all pH values, but upon addition of D1 PSI they fuse into larger vesicles at acidic pH. (**B**) LUVs of 100 nm were made using a yeast polar lipid extract. LUVs were made in buffered solutions at pH 4, 5, 6, and 7 and size was monitored using DLS over time, either without (top) or with (bottom) D1 PSI. Again, LUVs are stable in size for all pH values, but upon addition of D1 PSI they fuse into larger vesicles at acidic pH.

16-carbon acyl chain), were tested for their ability to associate with the D1 PSI. A solution was prepared containing both the neutrally-charged POPC and POPE as well as negatively-charged POPG, POPS, and POPA. One aliquot of this lipid solution had His-tagged D1 PSI added and was mixed, allowing time for lipid-protein interactions to occur. Then the lipid-protein mixture was applied to Ni^{2+} resin where the PSI would bind, bringing with it any associated lipids, while unbound lipids were washed away.

The PSI was eluted and any lipids that associated with it were quantified by MS and compared to the composition of the original lipid solution (Figure 4.5A). As shown in Figure 4.5B there is no significant difference in relative lipid composition between the original lipid solution and the lipids that co-purified with the PSI. However, it is possible the kinetics of association may be different depending on the lipid composition as hinted with the vesicle fusion assay (Figure 4.4). During membrane association the PSI must first interact with the head groups, before interacting with the acyl chains buried in the membrane. It is possible that specific head groups more strongly interact with the PSI, promoting initial association, followed by contact with the acyl chains, at which point PSI-acyl chain interactions predominate. The lack of sensitivity to the lipid head group makes the D1 PSI a promising candidate for making lipoprotein nanoparticles of different sizes for NMR studies of other membrane proteins, as previously demonstrated for saposin A [264, 265, 266].

4.3.6 Solid-State NMR Shows That D1 PSI Is Ordered and Strongly Bound to the Membrane

Solid-state NMR (ssNMR) is often used to solve the structures of membrane-associated proteins whose complexes with lipids are too large to tumble isotropically in solution. Because ssNMR makes use of magic angle spinning (MAS) to average out chemical shift anisotropy and dipolar coupling interactions it does not suffer the size limitations of solution-state NMR.

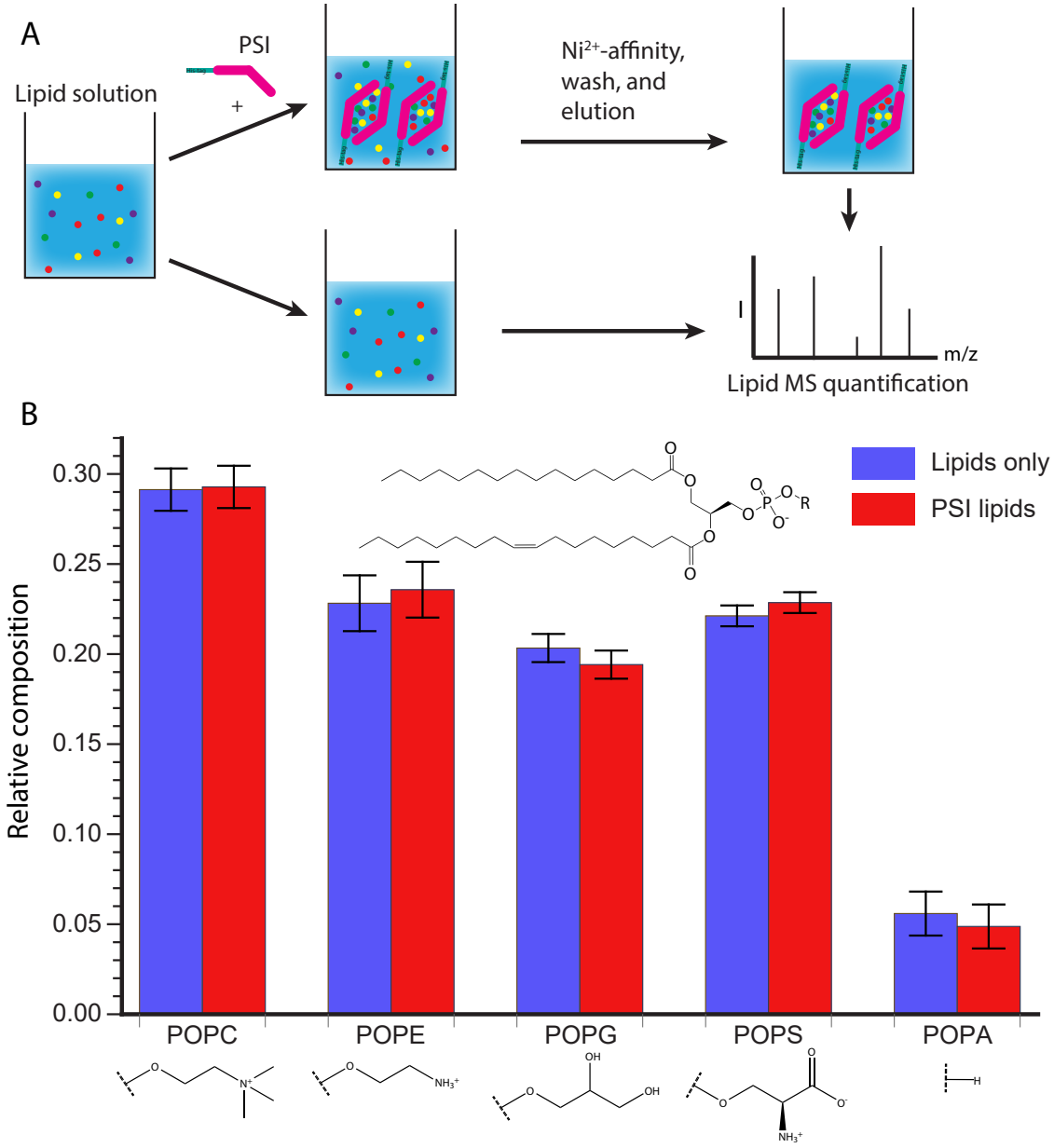

Figure 4.5: (**A**) Schematic representation of the experimental procedure. A lipid solution containing POPC, POPE, POPG, POPS, and POPA was prepared and split into two aliquots. One aliquot was treated with His-tagged D1 PSI and allowed to mix before purifying the His-tagged PSI and bound lipids from the unbound lipids. Lipids were then extracted from the D1 PSI and lipid solution and relative lipid species were quantified using mass spectrometry. (**B**) Results of lipid quantification. Blue represents lipids from the initial lipid mixture while red represents lipids that were bound to D1 PSI. The data indicate that there is no difference between the initial composition of lipids and the composition of lipids bound to the D1 PSI.

This allows for the use of vesicles or bicelles as a membrane system to study membrane proteins [267]. Because of its similarity to saposins, the D1 PSI may be useful for stabilizing lipid nanodiscs, which are composed of a small section of lipid bilayer that is encircled by a membrane scaffold protein [268]. For ssNMR studies, ^{13}C, ^{15}N isotopically labeled D1 PSI was expressed, purified and mixed with a mixture of 1: 1 POPC and POPA at pH 4.5. The sample was sent to the National High Magnetic Field Laboratory in Tallahassee, FL to obtain spectra.

Our NMR investigations of this protein begin with 2D ^{13}C-^{13}C correlation spectra collected using the dipolar-assisted rotational resonance (DARR) experiment [269, 243]. Early experiments using nano- or microcrystalline preparations of small, globular model proteins such as BPTI [270], ubiquitin [271], GB1 [272], and the α-spectrin SH3 domain [273] demonstrated that this type of homonuclear correlation can be used to provide partial assignments. In the usual procedure, well-resolved resonances with distinctive chemical shift values are identified, followed by mapping of spin systems via a "walk" among the proximal ^{13}C atoms of the sidechain. Full assignments can then be obtained using further 2D, 3D and 4D experiments [274, 275], followed by the measurement of through-space distance restraints (often using these same simple homonuclear correlations), and finally, structure determination [273, 276, 277]. This methodology is also fully applicable to membrane proteins [278, 279, 280], most readily in cases where the protein adopts a well-ordered conformation in the membrane. The ^{13}C-^{13}C CP DARR spectrum (Figure 4.6) shows that there are clearly-defined peaks that are reasonably well-dispersed, demonstrating that the PSI is not overly mobile and is not structurally heterogeneous in the sample so as to distribute the signal over multiple chemical shifts. Additional DARR spectra were collected with mixing times of 50 and 400 ms (Supplementary Figure S7.) These spectra are similarly well-resolved, and comparing the three spectra shows the expected increase in the number of cross-peaks with increasing mixing time, as correlations between more distant spin pairs are observed.

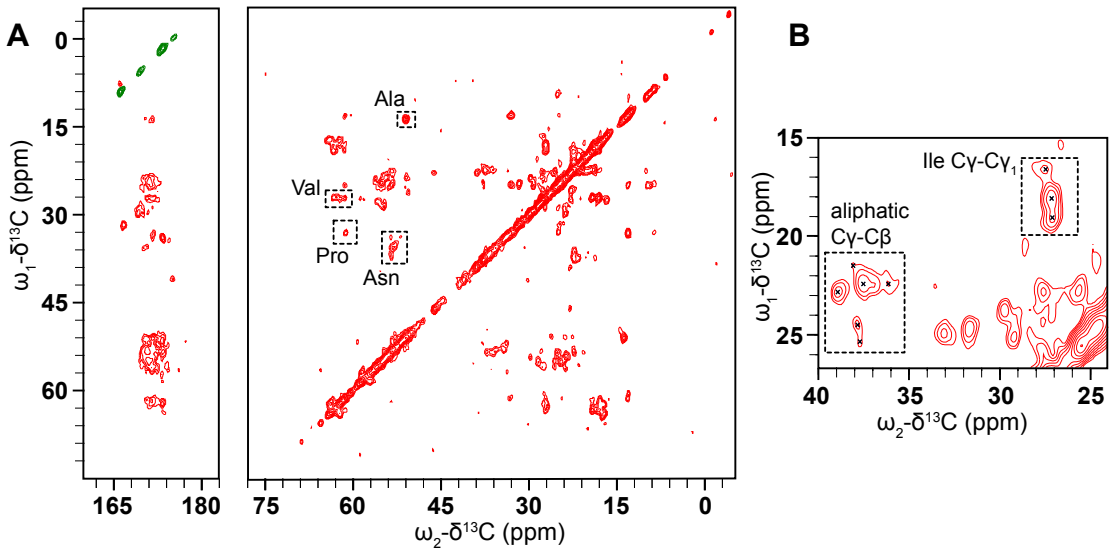

Figure 4.6: (**A**) CO (left) and aliphatic (right) regions of a ^{13}C-^{13}C CP DARR spectrum of D1 PSI in a 1:1 POPC, POPA membrane system was taken at 10 °C and spinning at 24.4 kHz. The spectrum displays well-resolved off-diagonal peaks of reasonable intensity indicating that the PSI is suitable for ssNMR structure determination. Tentative assignments are labeled by residue type. (**B**) The inset region shows some representative cross-peaks in the aliphatic region.

These preliminary spectra indicate that this protein is amenable to structure determination by ssNMR. Based on common ^{13}C chemical shifts we can tentatively assign some residues based on correlations between the Cα and Cβ carbons. For example, 1 Ala, 3 Val, 1 Pro, and 4 Asn residues were tentatively assigned. This suggests that a backbone walk is feasible for D1-PSI in a POPC and POPA membrane system. Carbon chemical shifts were compared to the average chemical shifts from all the proteins in the BioMagResBank [281], gathered from `http://www.bmrb.wisc.edu/ref_info/statful.htm` as published on 5/14/2020. 3D experiments will be needed for complete resonance assignments. Further sample preparation, such as optimization of lipid composition and protein-lipid ratios, is also needed to increase the signal of the sample and to explore how the D1 PSI responds to different lipids in the membrane system.

4.3.7 Molecular Modeling Suggests Potential Stability of Both Monomeric and Dimeric D1 PSI within Membranes, and Indicates That both Induce Permeability

To further explore the potential interaction of the D1 PSI with membranes, in particular the question of whether the membrane-interacting form is likely to be monomeric or dimeric, we performed all-atom molecular dynamics (MD) simulations of the PSI within POPC bilayers. As both monomeric and dimeric conformations have been proposed to be biologically relevant for other PSIs, we examined both cases. After 100 ns (following initial stabilization and equilibration) at 310K, we see that both monomeric and dimeric forms retain stable-but quite distinct-conformations within the bilayer. Figure 4.7B provides a schematic depiction of the monomeric case. The PSI assumes the open "L" conformation, with the terminal helices V1–Y13 and E87–P106 forming a partially solvated "pontoon" that sits parallel to the lipid surface and the central helix pair (residues G14-I34 and D66-N86) spanning the bilayer at an angle of roughly 120 degrees with respect to the "pontoon". On the opposite side, the protein is anchored by a mostly unstructured, hydrophilic loop region (G35-H65) that extends well into the solvent (see Figure 4.7C). This conformation appears to be stable, and indeed the "hinge" between the pontoon and the spanning helices would appear to allow the PSI to accommodate substantial deformation in membrane curvature or thickness without extensive conformational change.

In the case of the dimer, we observe a very different anchoring strategy. Figure 4.7E shows a schematic of the D1 PSI within the bilayer. In this case, the PSI forms a symmetric homodimer that anchors to both sides of the bilayer by respective copies of the hydrophilic anchor loop (G35-H65). Interestingly, the hydrophobic residues of the central helices that are lipid-exposed in the monomeric case serve here as the core of the homodimer, remaining stable despite burial in a hydrophobic environment. By contrast, the nominally lipid-exposed

residues about the core (which form the "back" of the open "L" in the monomer, and the pontoon) are largely hydrophilic. This does not appear to destabilize the dimer; instead, these residues appear to facilitate the maintenance of a water channel, as described below. As with the monomeric case, the anchor loop is aggressively solvated, with a water layer extending several nanometers beyond the lipid head groups (Figure 4.7F). The highly flexible nature of the anchoring loops suggests a substantial entropic contribution to the free energy of the lipid-embedded dimer (more so than for the monomer), possibly allowing the dimer to remain anchored within the bilayer at higher temperatures or at higher ionic concentrations.

Both monomeric and dimeric D1 PSI are observed to induce local changes in membrane conformation, creating "depressions" in the membrane surface (Figure 4.7C,F) that are linked to transmembrane water channels (Figure 4.7G,H). In the monomeric case, a single channel is formed that follows the hydrophilic residues on the "back" of the spanning helices, bridging to the hydrophilic residues on the sides and "bottom" of the pontoon (Figure 4.7G). For the dimer, the larger available hydrophilic surface area tends to produce multiple locally solvated regions (Figure 4.7H), supporting a broader channel structure. Examination of water transport across the membrane confirms that these apparent channels do indeed induce membrane permeability. Figure 4.8A shows posterior estimates of water transport rates (molecules/ns) across both simulated MD trajectories. As suggested by the broader channel structure, the PSI dimer supports a higher and somewhat more consistent mean flow rate (apx 37 molecules/ns) than the monomer (apx 7 molecules/ns), though the former is still somewhat "bursty" and irregular. Flow through the monomer channel is extremely irregular, with infrequent bursts and substantial changes in the overall flow rate over longer time scales (the former being evident from the roughness of the estimated rate function). Additional confirmation of the bursty nature of water flow in both cases can be seen by comparing the marginal distribution of observed transition events over short (10 ps) intervals to an equivalent Poisson process, as shown in Figure 4.8B. While the upper tail of the transition counts (i.e., numbers of water molecules observed to transition through the channel) is

heavier than the Poisson in both cases, the departure is especially profound for the monomer trajectory. This may arise from the relative asymmetry of the channel structure (as shown in Figure 4.7G), which may create "pools" of water that can only pass when the monomer assumes specific conformations. Whether or not this is the case, the substantially higher mean flow rates for dimeric D1 PSI suggest that enhanced concentrations of PSI within membranes will produce greater than linear increases in membrane permeability, potentially contributing to toxicity in biological settings.

4.4 Conclusions

In conclusion, the D1 PSI is a mostly α-helical, highly thermostable peptide that maintains its secondary structure up to 90 °C, even under conditions where its three disulfide bonds are reduced. It causes vesicle fusion in lipid mixtures from both yeast and *E. coli*, but only at a pH $\leq$ 5, consistent with the physiological pH of *D. capensis* digestive mucilage. Mass spectrometry of bound lipids upon extraction from lipid mixtures indicate that the PSI interacts with a wide range of lipids, independent of the charge on the head groups. Solid-state NMR data indicate that it strongly interacts with the membrane in a bicelle mixture, consistent with its robust lipid interactions in vesicle fusion assays. Furthermore, these data suggest that the PSI adopts an ordered conformation when interacting with the membrane, making structural studies in the membrane-bound state feasible. A major question left to be resolved is the oligomeric state of the PSI in membranes. In aqueous solution, it is predominantly monomeric over a wide range of pH values from 3–8. However, MD simulations show that either the monomeric or dimeric state may interact with and permeabilize membranes. In both the monomeric an dimeric states, the PSI in its open conformation can span the membrane, inducing local changes in membrane curvature and allowing the passage of water molecules from one side to the other. Further experimental

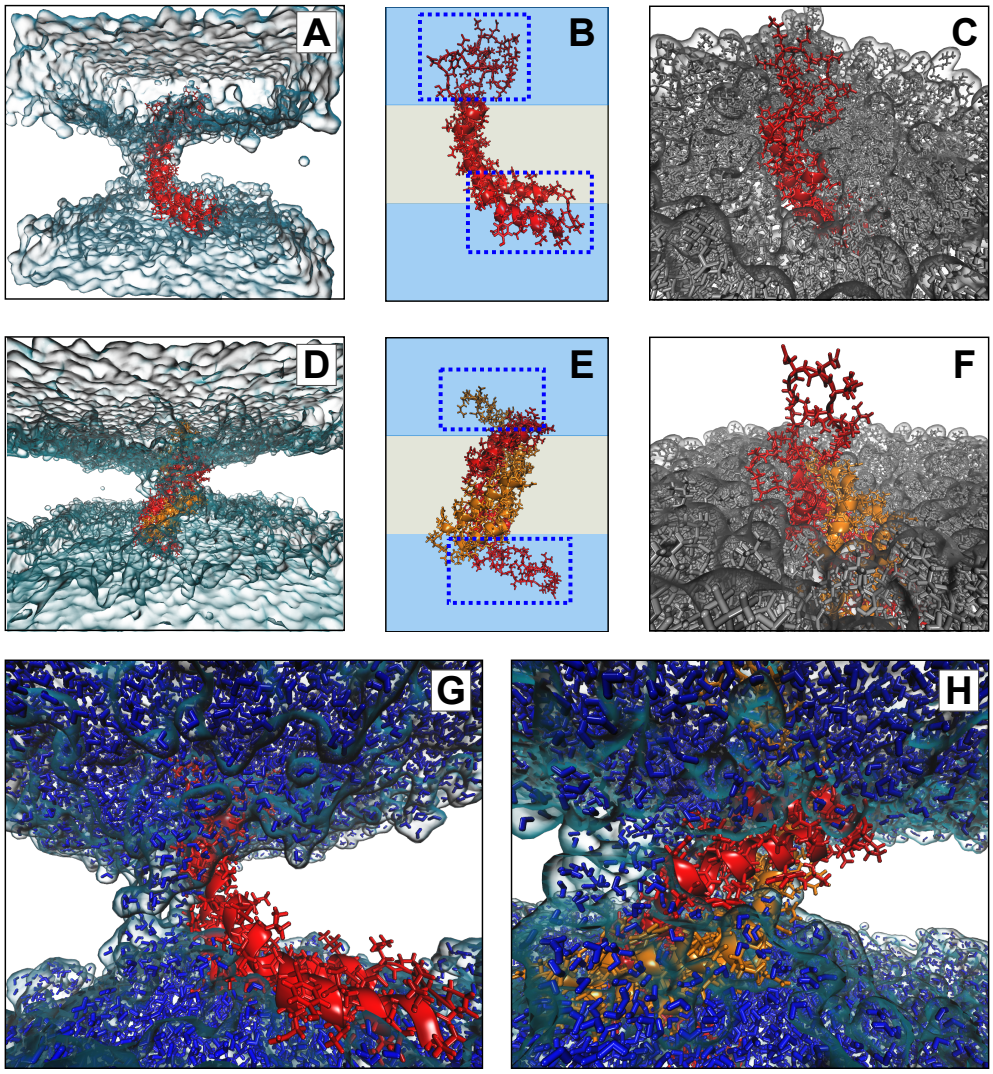

Figure 4.7: Conformations of monomeric (**A**–**C**) and dimeric (**D**–**F**) D1 PSI in POPC bilayer after 100ns. Both monomer and dimer assume stable but distinct conformations within the membrane (**A**,**D**), anchored by loop regions and (for the monomer) a helical "pontoon" as shown schematically in (**B**,**E**); blue regions indicate solvent, whereas grey regions indicate lipid. Loops extend substantially into solvent in both cases (**C**,**F**), while a depression forms in the surrounding bilayer. Bilayer depressions are linked to channels along the hydrophilic regions of the protein (**G**,**H**), through which water diffuses.

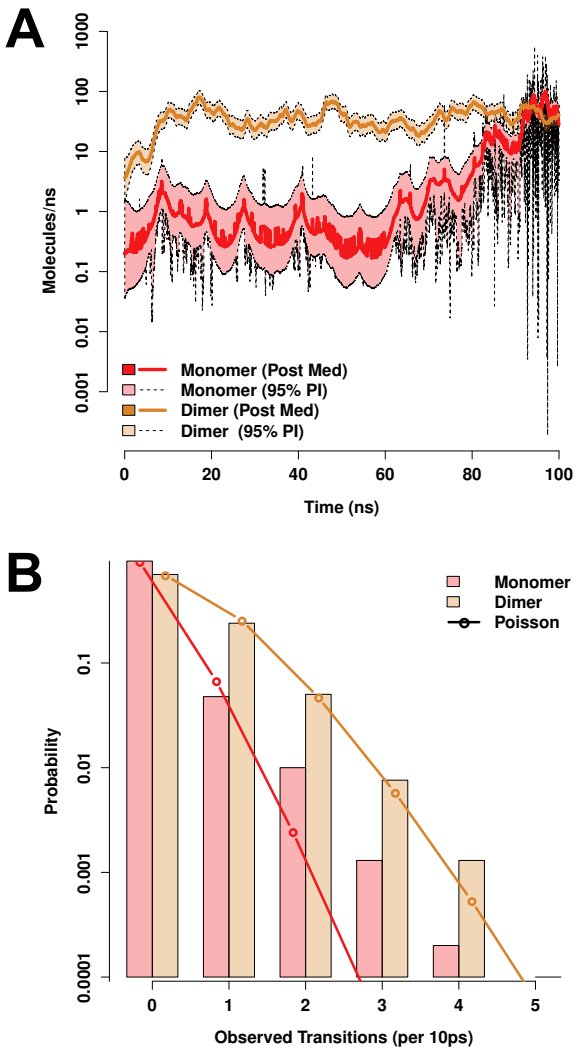

Figure 4.8: Both monomeric and dimeric D1 PSI conformations induce permeability in lipid bilayers, as revealed by MD trajectories. (**A**) Posterior median estimates (solid lines) and 95% posterior intervals (shaded areas) of trans-membrane water flow as a function of simulation time. Transport is extremely "bursty" in both conformations, with substantially larger fluctuations in mean rate for the PSI monomer trajectory. (**B**) Marginal distribution of transport events (cross-membrane water molecule transitions) per 10ns observation interval (bars). Observed distributions are heavier-tailed than Poisson distributions with equivalent expectation (lines), indicating high levels of "burstiness' even on short timescales.

characterization is needed to elucidate the mechanism of membrane interaction, which may involve monomers, dimers, or even larger pore-forming complexes.

4.5 Author Contributions

M.A.S.-P., J.C.B., M.G.C., G.R.T., B.P.C., N.K., and X.C. prepared protein samples, performed biophysical experiments and analyzed data. J.E.K., C.T.B., and R.W.M. performed protein sequence analysis. J.P., I.H., and R.Z. collected solid-state NMR data. M.A.S.P., J.C.B., and R.W.M. analyzed solid-state NMR data. C.T.B. conceptualized and performed molecular dynamics simulations and analyzed MD data. R.W.M designed the experimental studies. M.A.S.-P., J.C.B., M.G.C., C.T.B., and R.W.M. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

4.6 funding

This research was supported by NSF awards DMS-1361425 to C.T.B and R.W.M, and CHE-1308231 to R.W.M., ARO award W911NF-14-1-0552 to C.T.B, and the CIFAR Molecular Architecture of Life program. This work was made possible, in part, through access to the Genomics High Throughput Facility Shared Resource of the Cancer Center Support Grant (P30CA-062203) at the University of California, Irvine and NIH shared instrumentation grants 1S10RR025496-01, 1S10OD010794-01, and 1S10OD021718-01. The National High Magnetic Field Laboratory (NHMFL) is supported by the National Science Foundation through NSF/DMR-1644779 and the state of Florida. Development of the 36-T series connected hybrid magnet and NMR instrumentation was supported by NSF (DMR-1039938 and DMR-0603042) and NIH (BTRR 1P41 GM122698).

4.7 Acknowledgements

The authors acknowledge Dmitry Fishman of the UCI Laser Spectroscopy Labs for assistance with optical spectroscopy, Ben Katz and Felix Grun of the UCI Mass Spectrometry Facility for assistance with mass spectrometry, Tim Cross for helpful discussions and facilitating SCH use, and Veronika Cesar for preparing the plant diagrams in Figure 1. R.W.M. is a CIFAR fellow.

4.8 Conflict of Interests

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

D1	Droserasin 1
DMSO	dimethylsulfoxide
DTT	dithiothreitol
IPTG	Isopropyl β-D-1-thiogalactopyranoside
PSI	plant-specific insert
MES	2-(N-morpholino)ethanesulfonic acid
POPC	1-palmitoyl-2-oleoyl-glycero-3-phosphocholine
POPE	1-palmitoyl-2-oleoyl-sn-glycero-3- phosphoethanolamine
POPG	1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol)
POPS	1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine
POPA	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphate
LUV	large unilamellar vesicles
CD	circular dichroism
CP	cross-polarization
DARR	dipolar-assisted rotational resonance
MAS	magic angle spinning
MD	molecular dynamics
MS	mass spectrometry
NMR	nuclear magnetic resonance

Bibliography

[1] Keith J. Dilley and John J. Harding. Changes in proteins of the human lens in development and aging. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 386(2):391–408, April 1975.

[2] Domarin Khago, Jan C Bierma, Kyle W Roskamp, Natalia Kozlyuk, and Rachel W Martin. Protein refractive index increment is determined by conformation as well as composition. *Journal of Physics: Condensed Matter*, 30(43):435101, 2018.

[3] Mireille Delaye and Annette Tardieu. Short-range order of crystallin proteins accounts for eye lens transparency. *Nature*, 302(5907):302415a0, March 1983.

[4] Ronald H. H. Kröger, Melanie C. W. Campbell, Rejean Munger, and Russell D. Fernald. Refractive index distribution and spherical aberration in the crystalline lens of the African cichlid fish *Haplochromis burtoni*. *Vision Research*, 34(14):1815–1822, 1994.

[5] Chandler R. Dawson and Ivan R. Schwab. Epidemiology of cataract—a major cause of preventable blindness. *Bull World Health Organ*, 59(4):493–501, 1981.

[6] J.J. Berzelius. Lärobok i kemien. *Stockholm, P. A. Norstedt & Söner*, pt. 2:p. 512, 1830.

[7] Johann Franz Simon. *Animal Chemistry: With Reference to the Physiology and Pathology of Man*. Lea and Blanchard, Philadelphia, 1846.

[8] C.T. M\:orner. Untersuchung der Proteïnsubstanzen in den lichtbrechenden Medien des Auges. *Ztschr. f. physiol. Chem.*, 16:80, 1894.

[9] Wilfried W. de Jong, Gert-Jan Caspers, and Jack A. M. Leunissen. Genealogy of the α-crystallin—small heat-shock protein superfamily. *International Journal of Biological Macromolecules*, 22(3):151–162, May 1998.

[10] C. Slingsby, G. J. Wistow, and A. R. Clark. Evolution of crystallins for a role in the vertebrate eye lens. *Protein Sci.*, 22:367–380, 2013.

[11] J. Horwitz. Alpha-crystallin can function as a molecular chaperone. *Proceedings of the National Academy of Sciences of the United States of America*, 89(21):10449–10453, 1992.

[12] M. Haslbeck and E. Vierling. A first line of stress defense: Small heat shock proteins and their function in protein homeostasis. *Journal of Molecular Biology*, 427(7):1537–1548, 2015.

[13] Stephanie Runkle, Julie Hill, Marc Kantorow, Joseph Horwitz, and Mason Posner. Sequence and spatial expression of zebrafish (*Danio rerio*) αA-crystallin. *Mol. Vis.*, 8:45–50, 2002.

[14] A.B. Chepelinsky, J. Piatigorsky, M.M. Pisano, R.A. Dubin, G. Wistow, T.I. Limjoco, J.F. Klement, and C.J. Jaworski. Lens protein gene expression: Alpha-crystallins and MIP. *Lens and Eye Toxicity Research*, 8(2-3):319, 1991.

[15] A.N. Srinivasan, C.N. Nagineni, and S.P. Bhat. Alpha a-crystallin is expressed in non-ocular tissues. *Journal of Biological Chemistry*, 267(32):23337–23341, 1992.

[16] Junna Hayashi and John A. Carver. The multifaceted nature of αB-crystallin. *Cell Stress and Chaperones*, pages https://doi.org/10.1007/s12192–020–01098–w, 2020.

[17] Jason M Dahlman, Kelli L Margot, Linlin Ding, Joseph Horwitz, and Mason Posner. Zebrafish α-crystallins: Protein structure and chaperone-like activity compared to their mammalian orthologs. *Molecular Vision*, 11:88–96, 2005.

[18] M. Posner, M. Hawke, C. Lacava, C. J. Prince, N. R. Bellanco, and R. W. Corbin. A proteome map of the zebrafish (*Danio rerio*) lens reveals similarities between zebrafish and mammalian crystallin expression. *Molecular Vision*, 14:806–814, 2008.

[19] T.M.S Greiling, M. Aose, and J.I. Clark. Cell fate and differentiation of the developng ocular lens. *Investigative Ophthalmology & Visual Science*, 51:1540–1546, 2010.

[20] Mason Posner, Kelly L. Murray, Matthew S. McDonald, Hayden Eighinger, Brandon Andrew, Amy Drossman, Zachary Haley, Justin Nussbaum, Larry L. David, and Kirsten J. Lampi. The zebrafish as a model system for analyzing mammalian and native α-crystallin promoter function. *PeerJ*, 5:e4093, 2017.

[21] Sanjay Mishra, Shu-Yu Wu, Alexandra W. Fuller, Zhen Wang, Kristie L. Rose, Kevin L. Schey, and Hassane S. Mchaourab. Loss of αB-crystallin function in zebrafish reveals critical roles in the development of the lens and stress resistance of the heart. *J. Biol. Chem.*, 293(2):740–753, 2018.

[22] Ping Zou, Shu-Yu Wu, Hanane A. Koteiche, Sanjay Mishra, Daniel S. Levic, Ela Knapik, Wenbiao Chen, and Hassane S. Mchaourab. A conserved role of αA-crystallin in the development of the zebrafish embryonic lens. *Experimental Eye Research*, 138:104–113, 2015.

[23] K. K. Sharma, H. Kaur, and K. Kester. Functional elements in molecular chaperone -crystallin: identification of binding sites in bcrystallin. *Biochem. Biophys. Res. Commun.*, 239:217–222, 1997.

[24] J. Bhattacharyya, E. G. Padmanabha Udupa, J. Wang, and K. K. Sharma. Mini-αB-crystallin: a functional element of αB-crystallin with chaperone-like activity. *Biochemistry*, 45:3069–3076, 2006.

[25] Murugesan Raju, Puttur Santhoshkumar, and K. Krishna Sharma. Alpha-crystallin-derived peptides as therapeutic chaperones. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1860(1, Part B):246–251, 2016.

[26] J. G. Ghosh, A. K. Jr. Shenoy, and J. I. Clark. N- and C-terminal motifs in human αB crystallin play an important role in the recognition, selection, and solubilization of substrates. *Biochemistry*, 45:13847–13854, 2006.

[27] Hassane S. McHaourab, Jared A. Godar, and Phoebe L Stewart. Structure and mechanism of protein stability sensors: chaperone activity of small heat shock proteins. *Biochemistry*, 48(18):3828–3837, 2009.

[28] Andi Mainz, Jirka Peschek, Maria Stavropoulou, Katrin C Back, Benjamin Bardiaux, Sam Asami, Elke Prade, Carsten Peters, Sevil Weinkauf, Johannes Buchner, and Bernd Reif. The chaperone αB-crystallin uses different interfaces to capture an amorphous and an amyloid client. *Nature Structural & Molecular Biology*, 22:898–905, 2015.

[29] 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 αA- and αB-crystallins reveal structural mechanisms of polydispersity important for eye lens function. *Protein Science*, 19:1031–1043, 2010.

[30] John I Clark and Paul J Muchowski. Small heat-shock proteins and their potential role in human disease. *Current opinion in structural biology*, 10(1):52–59, 2000.

[31] Scott P Delbecq and Rachel E Klevit. One size does not fit all: the oligomeric states of αb crystallin. *FEBS letters*, 587(8):1073–1080, 2013.

[32] J. Andrew Aquilina, Justin L. P. Benesch, Orval A. Bateman, Christine Slingsby, and Carol V. Robinson. Polydispersity of a mammalian chaperone: Mass spectrometry reveals the population of oligomers in αB-crystallin. *Proc. Natl. Acad. Sci. U. S. A.*, 100(19):10611–10616, 2003.

[33] Andrew J. Baldwin, Hadi Lioe, Gillian R. Hilton, Lindsay A. Baker, John L. Rubinstein, Lewis E. Kay, and Justin L. P. Benesch. The polydispersity of αB-crystallin is rationalized by an interconverting polyhedral architecture. *Structure*, 19(12):1855–1863, 2011.

[34] Stefan Jehle, Barth van Rossum, Joseph R. Stout, Satoshi M. Noguchi, Katja Falber, Kristina Rehbein, Hartmut Oschkinat, Rachel E. Klevit, and Ponni Rajagopal. αB-crystallin: A hybrid solid-state/solution-state NMR investigation reveals structural aspects of the heterogeneous oligomer. *Journal of Molecular Biology*, 385(5), 2009, Pages 1481-1497.

[35] C. Bagneris, O.A. Bateman, C.E. Naylor, N. Cronin, W.C. Boelens, N.H. Keep, and C. Slingsby. Crystal structures of α-crystallin domain dimers of αB-crystallin and Hsp20. *J. Mol. Biol.*, 392:1242–1252, 2009.

[36] Zhenying Liu, Chuchu Wang, Yichen Li, Chunyu Zhao, Tongzhou Li, Dan Li, Shengnan Zhang, and Cong Liu. Mechanistic insights into the switch of αB-crystallin chaperone activity and self-multimerization. *The Journal of Biological Chemistry*, 293(38):14880–14890, September 2018.

[37] Scott P. Delbecq, Stefan Jehle, and Rachel Klevit. Binding determinants of the small heat shock protein, αB-crystallin: recognition of the 'IxI' motif. *EMBO Journal*, 31:4587–4594, 2012.

[38] R.W. Martin. *NMR studies of eye lens crystallins.* eMagRes (Encyclopedia of NMR). John Wiley & Sons, Ltd., The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom, 2014.

[39] Björn M. Burmann and Sebastian Hiller. Chaperones and chaperone–substrate complexes: Dynamic playgrounds for NMR spectroscopists. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 86–87:41–64, 2015.

[40] Stefan Jehle, Barth van Rossum, Joseph R. Stout, Satoshi M. Noguchi, Katja Falber, Kristina Rehbein, Hartmut Oschkinat, Rachel E. Klevit, and Ponni Rajagopal. αB-crystallin: a hybrid solid-state/solution-state NMR investigation reveals structural aspects of the heterogeneous oligomer. *Journal of Molecular Biology*, 385(5):1481–1497, 2009.

[41] Stefan Jehle, Ponni Rajagopal, Benjamin Bardiaux, Stefan Markovic, Ronald Kühne, Joseph R. Stout, Victoria A. Higman, Rachel E. Klevit, Barth-Jan van Rossum, and Hartmut Oschkinat. Solid-state nmr and saxs studies provide a structural basis for the activation of αb-crystallin oligomers. *Nature Structural & Molecular Biology*, 17(9):1037, 2010.

[42] Stefan Jehle, Breanna S. Vollmar, Benjamin Bardiaux, Katja K. Dove, Ponni Rajagopal, Tamir Gonen, Hartmut Oschkinat, and Rachel E. Klevit. N-terminal domain of αB-crystallin provides a conformational switch for multimerization and structural heterogeneity. *Proc. Natl. Acad. Sci. U. S. A.*, 108(16):6409–6414, 2011.

[43] Nathalie Braun, Martin Zacharias, Jirka Peschek, Andreas Kastenmüller, 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, December 2011.

[44] Andrew J. Baldwin, Gillian R. Hilton, Hadi Lioe, Claire Bagnéris, Justin L. P. Benesch, and Lewis E. Kay. Quaternary dynamics of αB-crystallin as a direct consequence of localised tertiary fluctuations in the C-terminus. *J. Mol. Biol.*, 413(2):310–320, 2011.

[45] Rintaro Inoue, Takumi Takata, Norihiko Fujii, Kentaro Ishii, Susumu Uchiyama, Nobuhiro Sato, Yojiro Oba, Kathleen Wood, Koichi Kato, Noriko Fujii, and Masaaki Sugiyama. New insight into the dynamical system of αB-crystallin oligomers. *Scientific Reports*, 6:29208, 2016.

[46] Ponni Rajagopal, Eric Tse, Andrew J. Borst, Scott P. Delbecq, Lei Shi, Daniel R. Southworth, and Rachel E. Klevit. A conserved histidine modulates HSPB5 structure to trigger chaperone activity in response to stress-related acidosis. *eLife*, 4:e07304, 2015.

[47] T. Reid Alderson, Julien Roche, Heidi Y. Gastall, David M. Dias, Iva Pritišanac, Jinfa Ying, Ad Bax, Justin L. P. Benesch, and Andrew J. Baldwin. Local unfolding of the HSP27 monomer regulates chaperone activity. *Nature Communications*, 10:1068, 2019.

[48] Megan Garvey, Heath Ecroyd, Nicholas J. Ray, Juliet A. Gerrard, and John A. Carver. Functional amyloid protection in the eye lens: Retention of α-crystallin molecular chaperone activity after modification into amyloid fibrils. *Biomolecules*, 7(3):67, 2017.

[49] Christoph JO Kaiser, Carsten Peters, Philipp WN Schmid, Maria Stavropoulou, Juan Zou, Vinay Dahiya, Evgeny V Mymrikov, Beate Rockel, Sam Asami, Martin Haslbeck, et al. The structure and oxidation of the eye lens chaperone αa-crystallin. *Nature Structural & Molecular Biology*, 26(12):1141–1150, 2019.

[50] P. Thampi and E. C. Abraham. Influence of the C-terminal residues on oligomerization of αA-crystallin. *Biochemistry*, 42:11857–11863, 2003.

[51] A. Aziz, P. Santhoshkumar, K. K. Sharma, and E. C. Abraham. Cleavage of the C-terminal serine of human αA-crystallin produces αA1-172 with increased chaperone activity and oligomeric size. *Biochemistry*, 46:2510–2519, 2007.

[52] T. M. Treweek, A. Rekas, M. J. Walker, and J. A. Carver. A quantitative NMR spectroscopic examination of the flexibility of the C-terminal extensions of the molecular chaperones, αA- and αB-crystallin. *Exp. Eye Res.*, 91:691–699, 2010.

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

[54] M. Kundu, P. C. Sen, and K. P. Das. Structure, stability and chaperone function of αA-crystallin: role of N-terminal region. *Biopolymers*, 86:177–192, 2007.

[55] A. Laganowsky and D. Eisenberg. Non-3D domain swapped crystal structure of truncated zebrafish αA crystallin. *Protein Science*, 19(10):1978–1984, 2010.

[56] Waldemar Preis, Annika Bestehorn, Johannes Buchner, and Martin Haslbeck. An alternative splice variant of human αA-crystallin modulates the oligomer ensemble and the chaperone activity of α-crystallins. *Cell Stress and Chaperones*, 22:541–552, 2017.

[57] H. A. Koteiche, D. P. Claxton, S. Mishra, R. A. Stein, E. T. McDonald, and H. S. McHaourab. Species-specific structural and functional divergence of α-crystallins: Zebrafish αBa- and rodent αA ins-crystallin encode activated chaperones. *Biochemistry*, 54:5949–5958, 2015.

[58] A.K. Panda, S.K. Nandi, A. Chakraborty, R.H. Nagaraj, and A. Biswas. Differential role of arginine mutations on the structure and functions of α-crystallin. *Biochimica et Biophysica Acta*, 1860(1 Pt B):199–210, 2015.

[59] Sabrina Sacconi, Léonard Féasson, Jean Christophe Antoine, Christophe Pécheux, Rafaelle Bernard, Ana Maria Cobo, Alberto Casarin, Leonardo Salviati, Claude Desnuelle, and Andoni Urtizberea. A novel CRYAB mutation resulting in multisystemic disease. *Neuromuscul. Disord.*, 22(1):66–72, 2012.

[60] Maryam Ghahramani, Reza Yousefi, Alexey Krivandin, Konstantin Muranov, Boris Kurganov, and Ali Akbar Moosavi-Movahedi. Structural and functional characterization of D109H and R69C mutant versions of human αB-crystallin: The biochemical pathomechanism underlying cataract and myopathy development. *International Journal of Biological Macromolecules*, 2019.

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

[62] Sadaf Saba, Maryam Ghahramani, and Reza Yousefi. A comparative study of the impact of calcium ion on structure, aggregation and chaperone function of human αa-crystallin and its cataract-causing r12c mutant. *Protein and peptide letters*, 24(11):1048–1058, 2017.

[63] Ashutosh S. Phadte, Sundararajan Mahalingam, Puttur Santhoshkumar, and Krishna K. Sharma. Functional rescue of cataract-causing αA-G98R-crystallin by targeted compensatory suppressor mutations in human αA-crystallin. *Biochemistry*, 58(40):4148–4158, October 2019.

[64] Shu-Yu Wu, Ping Zou, Sanjay Mishra, and Hassane S. Mchaourab. Transgenic zebrafish models reveal distinct molecular mechanisms for cataract-linked αA-crystallin mutants. *PloS One*, 13(11):e0207540, 2018.

[65] Usha P. Andley, James P. Malone, Paul D. Hamilton, Nathan Ravi, and R. Reid Townsend. Comparative proteomic analysis identifies age-dependent increases in the abundance of specific proteins after deletion of the small heat shock proteins αA- and αB-crystallin. *Biochemistry*, 52(17):2933–2948, 2013.

[66] Usha P Andley and Joshua W Goldman. Autophagy and upr in α-crystallin mutant knock-in mouse models of hereditary cataracts. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1860(1):234–239, 2016.

[67] M. Ragerdi Kashani, R. Yousefi, M. Akbarian, M. M. Alavianmehr, and Y. Ghasemi. Structure, Chaperone Activity, and Aggregation of Wild-Type and R12C Mutant αB-Crystallins in the Presence of Thermal Stress and Calcium Ion - Implications for Role of Calcium in Cataract Pathogenesis. *Biochemistry. Biokhimiia*, 81(2):122–134, February 2016.

[68] Md Faiz Ahmad, Devendra Singh, Aftab Taiyab, Tangirala Ramakrishna, Bakthisaran Raman, and Ch Mohan Rao. Selective cu2+ binding, redox silencing, and cytoprotective effects of the small heat shock proteins αa-and αb-crystallin. *Journal of molecular biology*, 382(3):812–824, 2008.

[69] Maria Luisa Ganadu, Michaela Aru, Giovanni Maria Mura, Alessio Coi, Piotr Mlynarz, and Henryk Kozlowski. Effects of divalent metal ions on the αb-crystallin chaperone-like activity: spectroscopic evidence for a complex between copper (ii) and protein. *Journal of inorganic biochemistry*, 98(6):1103–1109, 2004.

[70] Andi Mainz, Benjamin Bardiaux, Frank Kuppler, Gerd Multhaup, Isabella C Felli, Roberta Pierattelli, and Bernd Reif. Structural and mechanistic implications of metal binding in the small heat-shock protein αb-crystallin. *Journal of Biological Chemistry*, 287(2):1128–1138, 2012.

[71] Liliana Quintanar, José A Domínguez-Calva, Eugene Serebryany, Lina Rivillas-Acevedo, Cameron Haase-Pettingell, Carlos Amero, and Jonathan A King. Copper and zinc ions specifically promote nonamyloid aggregation of the highly stable human γD- crystallin. *ACS Chemical Biology*, 11(1):263–272, 2016.

[72] Kalyan S Ghosh, Ajay Pande, and Jayanti Pande. Binding of γ-crystallin substrate prevents the binding of copper and zinc ions to the molecular chaperone α-crystallin. *Biochemistry*, 50(16):3279–3281, 2011.

[73] Priyanka Chauhan and Kalyan S Ghosh. Inhibition of copper-induced aggregation of human γd-crystallin by rutin and studies on its role in molecular level for enhancing the chaperone activity of human αa-crystallin by using multi-spectroscopic techniques. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 218:229–236, 2019.

[74] Priyanka Chauhan, Sai Brinda Muralidharan, Anand Babu Velappan, Dhrubajyoti Datta, Sanjay Pratihar, Joy Debnath, and Kalyan Sundar Ghosh. Inhibition of copper-mediated aggregation of human γd-crystallin by schiff bases. *JBIC Journal of Biological Inorganic Chemistry*, 22(4):505–517, 2017.

[75] José Antonio Dominguez-Calva, Cameron Haase-Pettingell, Eugene Serebryany, Jonathan Alan King, and Liliana Quintanar. A histidine switch for Zn-induced aggregation of γ-crystallins reveals a metal-bridging mechanism that is relevant to cataract disease. *Biochemistry*, 57(33):4959–4962, 2018.

[76] Ashis Biswas and Kali P Das. Zn2+ enhances the molecular chaperone function and stability of α-crystallin. *Biochemistry*, 47(2):804–816, 2008.

[77] A Biswas, S Karmakar, A Chowdhury, and KP Das. Interaction of α-crystallin with some small molecules and its effect on its structure and function. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1860(1):211–221, 2016.

[78] Susmita Barman and Krishnapura Srinivasan. Zinc supplementation ameliorates diabetic cataract through modulation of crystallin proteins and polyol pathway in experimental rats. *Biological trace element research*, 187(1):212–223, 2019.

[79] John R Hawse, Jonathan R Cumming, Brian Oppermann, Nancy L Sheets, Venkat N Reddy, and Marc Kantorow. Activation of metallothioneins and α-crystallin/shsps in human lens epithelial cells by specific metals and the metal content of aging clear human lenses. *Investigative ophthalmology & visual science*, 44(2):672–679, 2003.

[80] S. R. Hanson, A. Hasan, D. L. Smith, and J. B. Smith. The major in vivo modifications of the human waterinsoluble lens crystallins are disulfide bonds, deamidation, methionine oxidation and backbone cleavage. *Experimental Eye Research*, 71:195–207, 2000.

[81] Roger JW Truscott and Michael G Friedrich. The etiology of human age-related cataract. proteins don't last forever. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1860(1):192–198, 2016.

[82] Stefan Löfgren. Solar ultraviolet radiation cataract. *Experimental Eye Research*, 156:112–116, March 2017.

[83] Afrooz Anbaraki, Maryam Ghahramani, Konstantin O. Muranov, Boris I. Kurganov, and Reza Yousefi. Structural and functional alteration of human αA-crystallin after exposure to full spectrum solar radiation and preventive role of lens antioxidants. *International Journal of Biological Macromolecules*, 118(Pt A):1120–1130, October 2018.

[84] P. Santhoshkumar, M. Raju, and K. K. Sharma. αA-crystallin peptide SDRDKFVIFLDVKHF accumulating in aging lens impairs the function of α-crystallin and induces lens protein aggregation. *PLoS ONE*, 6:e19291, 2011.

[85] M. Y. Hooi and R. J. Truscott. Racemisation and human cataract. D-Ser, D-Asp/Asn and D-Thr are higher in the lifelong proteins of cataract lenses than in age-matched normal lenses. *Age (Dordr.)*, 33:131–141, 2011.

[86] Dana W. Aswad, Mallik V. Paranandi, and Brandon T. Schurter. Isoaspartate in peptides and proteins: Formation, significance, and analysis. *Journal of Pharmaceutical and Biomedical Analysis*, 21(6):1129–1136, 2000.

[87] Shuji Noguchi. Structural changes induced by the deamidation and isomerization of asparagine revealed by the crystal structure of Ustilago sphaerogena ribonuclease U2B. *Biopolymers*, 93(11):1003–1010, 2010. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bip.21514.

[88] Yana A. Lyon, Georgette M. Sabbah, and Ryan R. Julian. Differences in α-Crystallin isomerization reveal the activity of protein isoaspartyl methyltransferase (PIMT) in the nucleus and cortex of human lenses. *Experimental Eye Research*, 171:131–141, June 2018.

[89] T. Takata and N. Fujii. Isomerization of Asp residues plays an important role in αA-crystallin dissociation. *FEBS J.*, 283:850–859, 2016.

[90] Yana A. Lyon, Miranda P. Collier, Dylan L. Riggs, Matteo T. Degiacomi, Justin L. P. Benesch, and Ryan R. Julian. Structural and functional consequences of age-related isomerization in α-crystallins. *The Journal of Biological Chemistry*, 294:7546–7555, 2019.

[91] V J Stevens, C A Rouzer, V M Monnier, and A Cerami. Diabetic cataract formation: Potential role of glycosylation of lens crystallins. *Proceedings of the National Academy of Sciences of the United States of America*, 75(6):2918–2922, June 1978.

[92] Fereshteh Bahmani, S. Zahra Bathaie, S. Javid Aldavood, and Arezou Ghahghaei. Prevention of α-crystallin glycation and aggregation using l-lysine results in the inhibition of in vitro catalase heat-induced-aggregation and suppression of cataract formation in the diabetic rat. *International Journal of Biological Macromolecules*, 132:1200–1207, July 2019.

[93] Sandip K. Nandi, Rooban B. Nahomi, Peter S. Harris, Cole R. Michel, Kristofer S. Fritz, and Ram H. Nagaraj. The absence of SIRT3 and SIRT5 promotes the acetylation of lens proteins and improves the chaperone activity of α-crystallin in mouse lenses. *Experimental Eye Research*, 182:1–9, May 2019.

[94] L. K. Muranova, M. V. Sudnitsyna, and N. B. Gusev. αB-crystallin phosphorylation: Advances and problems. *Biochemistry (Moscow)*, 83(10):1196–1206, October 2018.

[95] Md Faiz Ahmad, Bakthisaran Raman, Tangirala Ramakrishna, and Ch Mohan Rao. Effect of phosphorylation on alpha B-crystallin: Differences in stability, subunit exchange and chaperone activity of homo and mixed oligomers of alpha B-crystallin and its phosphorylation-mimicking mutant. *Journal of Molecular Biology*, 375(4):1040–1051, January 2008.

[96] J. B. Smith, Y. Sun, D. L. Smith, and B. Green. Identification of the posttranslational modifications of bovine lens alpha B-crystallins by mass spectrometry. *Protein Science : A Publication of the Protein Society*, 1(5):601–608, May 1992.

[97] Michela Ciano, Simona Allocca, Maria Camilla Ciardulli, Lucrezia della Volpe, Stefano Bonatti, and Massimo D'Agostino. Differential phosphorylation-based regulation of αB-crystallin chaperone activity for multipass transmembrane proteins. *Biochemical and Biophysical Research Communications*, 479(2):325–330, October 2016.

[98] Kashmeera D. Baboolall, Yusrah B. Kaudeer, Anne Gershenson, and Patricia B. O'Hara. ph dependence of oligomerization and functional activity of α b crystallin. *Biophysical Journal*, 118(3):510a, February 2020.

[99] Raman Bakthisaran, Kranthi Kiran Akula, Ramakrishna Tangirala, and Ch. Mohan Rao. Phosphorylation of αB-crystallin: Role in stress, aging and patho-physiological conditions. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1860(1, Part B):167–182, January 2016.

[100] J. Andrew Aquilina, Justin L. P. Benesch, Lin Lin Ding, Orna Yaron, Joseph Horwitz, and Carol V. Robinson. Phosphorylation of αB-crystallin alters chaperone function through loss of dimeric substructure. *The Journal of Biological Chemistry*, 279(27):28675–28680, 2004.

[101] Martin Haslbeck, Jirka Peschek, Johannes Buchner, and Sevil Weinkauf. Structure and function of α-crystallins: Traversing from in vitro to in vivo. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1860(1 Part B):149–166, 2016.

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

[103] Payel Das, Seung gu Kang, Sally Temple, and Georges Belfort. Interaction of amyloid inhibitor proteins with amyloid beta peptides: Insight from molecular dynamics simulations. *PloS One*, 9(11):e113041, 2014.

[104] R. H. Smulders, K. B. Merck, J. Aendekerk, J. Horwitz, L. Takemoto, C. Slingsby, H. Bloemendal, and W. W. De Jong. The mutation Asp69–>Ser affects the chaperone-like activity of α A-crystallin. *European Journal of Biochemistry*, 232(3):834–838, September 1995.

[105] P. Santhoshkumar and K. K. Sharma. Phe71 is essential for chaperone-like function in alpha A-crystallin. *The Journal of Biological Chemistry*, 276(50):47094–47099, December 2001.

[106] L. Takemoto, T. Emmons, and J. Horwitz. The C-terminal region of alpha-crystallin: Involvement in protection against heat-induced denaturation. *The Biochemical Journal*, 294 (Pt 2):435–438, September 1993.

[107] M. L. Plater, D. Goode, and M. J. Crabbe. Effects of site-directed mutations on the chaperone-like activity of αB-crystallin. *The Journal of Biological Chemistry*, 271(45):28558–28566, November 1996.

[108] K. Krishna Sharma, R. Senthil Kumar, G. Suresh Kumar, and P. Thomas Quinn. Synthesis and Characterization of a Peptide Identified as a Functional Element in αA-crystallin. *Journal of Biological Chemistry*, 275(6):3767–3771, February 2000.

[109] Jaya Bhattacharyya, E. G. Padmanabha Udupa, Jing Wang, and K. Krishna Sharma. Mini-alphaB-crystallin: A functional element of alphaB-crystallin with chaperone-like activity. *Biochemistry*, 45(9):3069–3076, March 2006.

[110] Priya R. Banerjee, Ajay Pande, Alexander Shekhtman, and Jayanti Pande. Molecular mechanism of the chaperone function of mini-α-crystallin, a 19-residue peptide of human α-crystallin. *Biochemistry*, 54(2):505–515, January 2015.

[111] Sanjay Mishra, Richard A. Stein, and Hassane S. Mchaourab. Cataract-linked γD-crystallin mutants have weak affinity to lens chaperones α-crystallins. *FEBS Letters*, 586(4):330–336, 2012.

[112] Marc A. Sprague-Piercy, Eric Wong, Kyle W. Roskamp, Joseph N. Fakhoury, J. Alfredo Freites, Douglas J. Tobias, and Rachel W. Martin. Human αB-crystallin discriminates between aggregation-prone and function-preserving variants of a client protein. *Biochimica Et Biophysica Acta. General Subjects*, 1864(3):129502, March 2020.

[113] Lisa M. Ramirez, Alexander Shekhtman, and Jayanti Pande. Hydrophobic residues of melittin mediate its binding to αA-crystallin. *Protein Science*, 29(2):572–588, 2020.

[114] John I. Clark. Functional sequences in human alphaB crystallin. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1860(1, Part B):240–245, January 2016.

[115] K. O. Muranov, N. B. Poliansky, N. A. Chebotareva, S. Yu Kleimenov, A. E. Bugrova, M. I. Indeykina, A. S. Kononikhin, E. N. Nikolaev, and M. A. Ostrovsky. The mechanism of the interaction of α-crystallin and UV-damaged βL-crystallin. *International Journal of Biological Macromolecules*, 140:736–748, November 2019.

[116] John A. Carver, Aidan B. Grosas, Heath Ecroyd, and Roy A. Quinlan. The functional roles of the unstructured N- and C-terminal regions in αB-crystallin and other mammalian small heat-shock proteins. *Cell Stress & Chaperones*, 22(4):627–638, July 2017.

[117] 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.

[118] Andor J. Kiss and C.-H. Christina Cheng. Molecular diversity and genomic organisation of the α, β, and γ eye lens crystallins from the Antarctic toothfish *Dissostichus mawsoni*. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, 3(2):155–171, 2008.

[119] Mason Posner, Andor J. Kiss, Jackie Skiba, Amy Drossman, Monika B. Dolinska, J. Fielding Hejtmancik, and Yuri V. Sergeev. Functional validation of hydrophobic adaptation to physiological temperature in the small heat shock protein αA-crystallin. *PloS One*, 7(3):e34438, 2012.

[120] A. A. Smith, K. Wyatt, J. Vacha, T. S. Vihtelic, J.S. Jr. Zigler, G. J. Wistow, and M. Posner. Gene duplication and separation of functions in αB-crystallin from zebrafish (*Danio rerio*). *FEBS Journal*, 273:481–490, 2006.

[121] Georg K. A. Hochberg, Heath Ecroyd, Cong Liu, Dezerae Cox, Duilio Cascio, Michael R. Sawaya, Miranda P. Collier, James Stroud, John A. Carver, Andrew J. Baldwin, Carol V. Robinson, David S. Eisenberg, Justin L. P. Benesch, and Arthur Laganowsky. The structured core domain of αB-crystallin can prevent amyloid fibrillation and associated toxicity. *Proceedings of the National Academy of Sciences of the United States of America*, 111(16):E1562–E1570, 2014.

[122] Johannes M van Noort, Malika Bsibsi, Peter J Nacken, Richard Verbeek, and Edna HG Venneker. Therapeutic intervention in multiple sclerosis with *alpha* b-crystallin: A randomized controlled phase iia trial. *PloS one*, 10(11):e0143366, 2015.

[123] Steven Droho, Mitchell E. Keener, and Niklaus H. Mueller. Heparan sulfate mediates cell uptake of αB-crystallin fused to the glycoprotein C cell penetration peptide. *Biochimica Et Biophysica Acta. Molecular Cell Research*, 1865(4):598–604, April 2018.

[124] Anjali Bisht, Manju Sharma, Shikha Sharma, Md Ehesan Ali, and Jiban Jyoti Panda. Carrier-free self-built aspirin nanorods as anti-aggregation agents towards alpha-crystallin-derived peptide aggregates: potential implications in non-invasive cataract therapy. *Journal of Materials Chemistry B*, 7(44):6945–6954, 2019.

[125] Leah N. Makley, Kathryn A. McMenimen, Brian T. DeVree, Joshua W. Goldman, Brittney N. McGlasson, Ponni Rajagopal, Bryan M. Dunyak, Thomas J. McQuade, Andrea D. Thompson, Roger Sunahara, Rachel E. Klevit, Usha P. Andley, and Jason E. Gestwicki. Pharmacological chaperone for α-crystallin partially restores transparency in cataract models. *Science*, 350(6261):674–677, November 2015.

[126] Prabhat K. Mallik, Hua Shi, and Jayanti Pande. RNA aptamers targeted for human αA-crystallin do not bind αB-crystallin, and spare the α-crystallin domain. *Biochem. Biophys. Res. Commun.*, 491(2):423–428, September 2017.

[127] Martin Karplus and J. Andrew McCammon. Molecular dynamics simulations of biomolecules. *Nature Structural & Molecular Biology*, 9:646–652, 2002.

[128] M. Karplus and J. Kuriyan. Molecular dynamics and protein function. *Proceedings of the National Academy of Sciences of the United States of America*, 102(19):6679–6685, 2005.

[129] A. T. Brunger and P. D. Adams. Molecular dynamics applied to x-ray structure refinement. *Accounts of Chemical Research*, 35:404–412, 2002.

[130] J. H. Chen, H. S. Won, Im W. P., H. J. Dyson, and C. L. Brooks. Generation of native-like protein structures from limited NMR data, modern force fields and advanced conformational sampling. *Journal of Biomolecular NMR*, 31:59–64, 2005.

[131] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, and Augusta H. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087, 1953.

[132] Anl Kurut, Björn A. Persson, Torbjörn Åkesson, Jan Forsman, and Mikael Lund. Anisotropic interactions in protein mixtures: Self assembly and phase behavior in aqueous solution. *Journal of Physical Chemistry Letters*, 3(6):731–734, 2012.

[133] Vera Prytkova, Matthias Heyden, Domarin Khago, J. Alfredo Freites, Carter T. Butts, Rachel W. Martin, and Douglas J. Tobias. Multi-conformation Monte Carlo: A method for introducing flexibility in efficient simulations of many-protein systems. *The Journal of Physical Chemistry B*, 120(33):8115–8126, 2016.

[134] Gianmarc Grazioli, Yue Yu, Megha H. Unhelkar, Rachel W. Martin, and Carter T. Butts. Network-based classification and modeling of amyloid fibrils. *The Journal of Physical Chemistry B*, 123(26):5452–5462, 2019.

[135] Gianmarc Grazioli, Yue Yu, Megha H. Unhelkar, R.W. Martin, and Carter T. Butts. Network hamiltonian models reveal pathways to amyloid fibril formation. *Scientific Reports*, in press, 2020.

[136] Ashis Biswas, Jeffery Goshe, Antonia Miller, Puttur Santhoshkumar, Carol Luckey, Manjunatha B. Bhat, and Ram H. Nagaraj. Paradoxical effects of substitution and deletion mutation of Arg56 on the structure and chaperone function of human alphaB-crystallin. *Biochemistry*, 46(5):1117–1127, February 2007.

[137] Teresa M. Treweek, Agata Rekas, Robyn A. Lindner, Mark J. Walker, J. Andrew Aquilina, Carol V. Robinson, Joseph Horwitz, Ming Der Perng, Roy A. Quinlan, and John A. Carver. R120G alphaB-crystallin promotes the unfolding of reduced alpha-lactalbumin and is inherently unstable. *The FEBS journal*, 272(3):711–724, February 2005.

[138] Evgeniia S. Gerasimovich, Sergei V. Strelkov, and Nikolai B. Gusev. Some properties of three αB-crystallin mutants carrying point substitutions in the C-terminal domain and associated with congenital diseases. *Biochimie*, 142:168–178, November 2017.

[139] Steven Droho, Mitchell E. Keener, and Niklaus H. Mueller. Changes in function but not oligomeric size are associated with αB-crystallin lysine substitution. *Biochemistry and Biophysics Reports*, 14:1–6, July 2018.

[140] D Lomas and H Parfrey. a1-Antitrypsin deficiency • 4: Molecular pathophysiology. *Thorax*, 59(6):529–535, June 2004.

[141] Frederick L. Ruberg and John L. Berk. Transthyretin (TTR) Cardiac Amyloidosis. *Circulation*, 126(10):1286–1300, September 2012.

[142] S. D. Roseff. Sickle cell disease: A review. *Immunohematology*, 25(2):67–74, 2009.

[143] Mireille Delaye and Annette Tardieu. Short-range order of crystallin proteins accounts for eye lens transparency. *Nature*, 302(5907):302415a0, March 1983.

[144] A. Huizinga, A. C. Bot, F. F. de Mul, G. F. Vrensen, and J. Greve. Local variation in absolute water content of human and rabbit eye lenses measured by Raman microspectroscopy. *Experimental Eye Research*, 48(4):487–496, April 1989.

[145] Niels Lynnerup, Henrik Kjeldsen, Steffen Heegaard, Christina Jacobsen, and Jan Heinemeier. Radiocarbon Dating of the Human Eye Lens Crystallines Reveal Proteins without Carbon Turnover throughout Life. *PLoS ONE*, 3(1), January 2008.

[146] J. Horwitz. Alpha-crystallin can function as a molecular chaperone. *Proceedings of the National Academy of Sciences of the United States of America*, 89(21):10449–10453, Nov 1992.

[147] Jirka Peschek, Nathalie Braun, Titus M. Franzmann, Yannis Georgalis, Martin Haslbeck, Sevil Weinkauf, and Johannes Buchner. The eye lens chaperone α-crystallin forms defined globular assemblies. *Proceedings of the National Academy of Sciences*, 106(32):13272–13277, August 2009.

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

[149] Susmitnarayan Chaudhury, Pooja Ghosh, Sultana Parveen, and Swagata Dasgupta. Glycation of human γB-crystallin: A biophysical investigation. *International Journal of Biological Macromolecules*, 96:392–402, March 2017.

[150] J. A. Aquilina and R. J. W. Truscott. Identifying sites of attachment of UV filters to proteins in older human lenses. *Biochimica Et Biophysica Acta*, 1596(1):6–15, April 2002.

[151] M. Litt, P. Kramer, D. M. LaMorticella, W. Murphey, E. W. Lovrien, and R. G. Weleber. Autosomal dominant congenital cataract associated with a missense mutation in the human alpha crystallin gene CRYAA. *Human Molecular Genetics*, 7(3):471–474, March 1998.

[152] Donna S. Mackay, Usha P. Andley, and Alan Shiels. Cell death triggered by a novel mutation in the αA-crystallin gene underlies autosomal dominant cataract linked to chromosome 21q. *European journal of human genetics: EJHG*, 11(10):784–793, October 2003.

[153] Michael P. Bova, Orna Yaron, Qingling Huang, Linlin Ding, Dana A. Haley, Phoebe L. Stewart, and Joseph Horwitz. Mutation R120G in αB-crystallin, which is linked to a desmin-related myopathy, results in an irregular structure and defective chaperone-like function. *Proceedings of the National Academy of Sciences of the United States of America*, 96(11):6137–6142, May 1999.

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

[155] Jennifer C. Boatz, Matthew J. Whitley, Mingyue Li, Angela M. Gronenborn, and Patrick C. A. van der Wel. Cataract-associated P23T γD-crystallin retains a native-like fold in amorphous-looking aggregates formed at physiological pH. *Nature Communications*, 8:15137, May 2017.

[156] Ajay Pande, Jianchao Zhang, Priya R. Banerjee, Shadakshara S. Puttamadappa, Alexander Shekhtman, and Jayanti Pande. NMR study of the cataract-linked P23T mutant of human γD-crystallin shows minor changes in hydrophobic patches that reflect its retrograde solubility. *Biochemical and biophysical research communications*, 382(1):196–199, April 2009.

[157] Jinwon Jung, In-Ja L. Byeon, Yongting Wang, Jonathan King, and Angela M. Gronenborn. The structure of the cataract-causing P23T mutant of human γD-crystallin exhibits distinctive local conformational and dynamic changes. *Biochemistry*, 48(12):2597–2609, March 2009.

[158] Zhiwei Ma, Grzegorz Piszczek, Paul T. Wingfield, Yuri V. Sergeev, and J. Fielding Hejtmancik. The G18V CRYGS mutation associated with human cataracts increases γS-crystallin sensitivity to thermal and chemical stress. *Biochemistry*, 48(30):7334–7341, August 2009.

[159] 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):e51401, December 2012.

[160] Srinivasu Karri, Ramesh Babu Kasetti, Venkata Pulla Rao Vendra, Sushil Chandani, and Dorairajan Balasubramanian. Structural analysis of the mutant protein D26G of human γS-crystallin, associated with Coppock cataract. *Molecular Vision*, 19:1231–1237, 2013.

[161] Carolyn N. Kingsley, William D. Brubaker, Stefan Markovic, Anne Diehl, Amanda J. Brindley, Hartmut Oschkinat, and Rachel W. Martin. Preferential and specific binding of human αB-crystallin to cataract-related variant of γS-crystallin. *Structure (London, England : 1993)*, 21(12):2221–2227, December 2013.

[162] William D. Brubaker, J. Alfredo Freites, Kory J. Golchert, Rebecca A. Shapiro, Vasilios Morikis, Douglas J. Tobias, and Rachel W. Martin. Separating Instability from Aggregation Propensity in γS-Crystallin Variants. *Biophysical Journal*, 100(2):498–506, January 2011.

[163] Domarin Khago, Eric K Wong, Carolyn N Kingsley, J Alfredo Freites, Douglas J Tobias, and Rachel W Martin. Increased hydrophobic surface exposure in the cataract-related g18v variant of human γs-crystallin. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1860(1):325–332, 2016.

[164] Sanjay Mishra, Richard A. Stein, and Hassane S. Mchaourab. Cataract-linked γD-crystallin mutants have weak affinity to lens chaperones α-crystallins. *FEBS Letters*, 586(4):330–336, February 2012.

[165] S. Zarina, C. Slingsby, R. Jaenicke, Z. H. Zaidi, H. Driessen, and N. Srinivasan. Three-dimensional model and quaternary structure of the human eye lens protein gamma S-crystallin based on beta- and gamma-crystallin X-ray coordinates and ultracentrifugation. *Protein Science : A Publication of the Protein Society*, 3(10):1840–1846, October 1994.

[166] Yongting Wang, Sarah Petty, Amy Trojanowski, Kelly Knee, Daniel Goulet, Ishita Mukerji, and Jonathan King. Formation of Amyloid Fibrils In Vitro from Partially Unfolded Intermediates of Human γC-Crystallin. *Investigative Ophthalmology & Visual Science*, 51(2):672–678, February 2010.

[167] Jianhua Xu, Jiejin Chen, Dmitri Toptygin, Olga Tcherkasskaya, Patrik Callis, Jonathan King, Ludwig Brand, and Jay R. Knutson. Femtosecond Fluorescence Spectra of Tryptophan in Human γ-Crystallin Mutants: Site-Dependent Ultrafast Quenching. *Journal of the American Chemical Society*, 131(46):16751–16757, November 2009.

[168] Jiejin Chen, Patrik R. Callis, and Jonathan King. Mechanism of the Very Efficient Quenching of Tryptophan Fluorescence in Human γD- and γS-Crystallins: The γ-Crystallin Fold May Have Evolved To Protect Tryptophan Residues from Ultraviolet Photodamage. *Biochemistry*, 48(17):3708–3716, May 2009.

[169] Melissa S Kosinski-Collins, Shannon L Flaugh, and Jonathan King. Probing folding and fluorescence quenching in human γd crystallin greek key domains using triple tryptophan mutant proteins. *Protein Science*, 13(8):2223–2235, 2004.

[170] Shannon L Flaugh, Melissa S Kosinski-Collins, and Jonathan King. Contributions of hydrophobic domain interface interactions to the folding and stability of human γd-crystallin. *Protein Science*, 14(3):569–581, 2005.

[171] C. Pokalsky, P. Wick, E. Harms, F. E. Lytle, and R. L. Van Etten. Fluorescence resolution of the intrinsic tryptophan residues of bovine protein tyrosyl phosphatase. *The Journal of Biological Chemistry*, 270(8):3809–3815, February 1995.

[172] 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, September 2011.

[173] Ajay Pande, Kalyan S. Ghosh, Priya R. Banerjee, 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(29):6122–6129, July 2010.

[174] A. Bothra, A. Bhattacharyya, C. Mukhopadhyay, K. Bhattacharyya, and S. Roy. A fluorescence spectroscopic and molecular dynamics study of bis-ANS/protein interaction. *Journal of Biomolecular Structure & Dynamics*, 15(5):959–966, April 1998.

[175] Andrea Hawe, Marc Sutter, and Wim Jiskoot. Extrinsic Fluorescent Dyes as Tools for Protein Characterization. *Pharmaceutical Research*, 25(7):1487–1499, July 2008.

[176] G. Bodenhausen and D.J. Ruben. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters*, 69(1):185–189, 1980.

[177] William D. Brubaker and Rachel W. Martin. ^{1}H, ^{13}C, and ^{15}N assignments of wild-type human γS-crystallin and its cataract-related variant γS-G18V. *Biomolecular NMR Assignments*, 6(1):63–67, April 2012.

[178] Raymond F. Greene and C. Nick Pace. Urea and Guanidine Hydrochloride Denaturation of Ribonuclease, Lysozyme, α-Chymotrypsin, and β-Lactoglobulin. *Journal of Biological Chemistry*, 249(17):5388–5393, October 1974.

[179] Frank Delaglio, Stephan Grzesiek, Geerten W. Vuister, Guang Zhu, John Pfeifer, and Ad Bax. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, 6(3):277–293, November 1995.

[180] Woonghee Lee, Marco Tonelli, and John L. Markley. NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics*, 31(8):1325–1327, April 2015.

[181] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kalé, and Klaus Schulten. Scalable molecular dynamics with namd. *J Comp Chem*, 26(16):1781–802, dec 2005.

[182] R. B. Best, J. Mittal, M. Feig, and Jr. MacKerell, A. D. Inclusion of many-body effects in the additive charmm protein cmap potential results in enhanced cooperativity of alpha-helix and beta-hairpin formation. *Biophys J*, 103(5):1045–51, 2012.

[183] Jr. MacKerell, A.D., M. Feig, and II Brooks, C.L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing conformational distributions in molecular dynamics simulations. *J.Comput.Chem.*, 25:1400–1415, 2004.

[184] Jr. MacKerell, A.D., D. Bashford, M. Bellott, Jr. Dunbrack, R.L., J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, III Reiher, W.E., B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J.Phys.Chem.B*, 102(18):3586–3616, 1998.

[185] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein. Comparison of simple potential functions for simulating liquid water. *J.Chem.Phys.*, 79(2):926–935, 1983.

[186] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen. A smooth particle mesh ewald method. *The Journal of Chemical Physics*, 103(19):8577–8593, 1995.

[187] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An n log(n) method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993.

[188] H. Grubmüller, H. Heller, A. Windemuth, and K. Schulten. Generalized verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Molecular Simulation*, 6(1-3):121–142, 1991.

[189] J.-P. Ryckaert, G. Ciccotti, and H.J.C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23:327–341, 1977.

[190] Shuichi Miyamoto and Peter A. Kollman. Settle: An analytical version of the shake and rattle algorithm for rigid water models. *Journal of Computational Chemistry*, 13(8):952–962, 1992.

[191] G.J. Martyna, D.J. Tobias, and M.L. Klein. Constant-pressure molecular-dynamics algorithms. *J.Chem.Phys.*, 101:4177–4189, 1994.

[192] Scott E. Feller, Yuhong Zhang, Richard W. Pastor, and Bernard R. Brooks. Constant pressure molecular dynamics simulation: The langevin piston method. *The Journal of Chemical Physics*, 103(11):4613–4621, 1995.

[193] W. Kabsch and C. Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, December 1983.

[194] Robbie P. Joosten, Tim A.H. te Beek, Elmar Krieger, Maarten L. Hekkelman, Rob W.W. Hooft, Reinhard Schneider, Chris Sander, and Gert Vriend. A series of PDB related databases for everyday needs. *Nucleic Acids Research*, 39(Database issue):D411–D419, January 2011.

[195] Ronald H. H. Kröger, Melanie C. W. Campbell, Rejean Munger, and Russell D. Fernald. Refractive index distribution and spherical aberration in the crystalline lens of the African cichlid fish haplochromis burtoni. 34(14):1815–1822.

[196] A. Huizinga, A. C. Bot, F. F. de Mul, G. F. Vrensen, and J. Greve. Local Variation in Absolute Water Content of Human and Rabbit Eye Lenses Measured by Raman Microspectroscopy. 48(4):487–496.

[197] Carolyn N. Kingsley, William D. Brubaker, Stefan Markovic, Anne Diehl, Amanda J. Brindley, Hartmut Oschkinat, and Rachel W. Martin. Preferential and specific binding of human B-crystallin to a cataract-related variant of S-crystallin. 21(12):2221–2227.

[198] Bert van den Berg, Rachel Wain, Christopher M. Dobson, and R.John Ellis. Macromolecular Crowding Perturbs Protein Refolding Kinetics: Implications for Folding inside the Cell. 19(15):3870–3875.

[199] S B Zimmerman and B Harrison. Macromolecular Crowding Increases Binding of DNA Polymerase to DNA: An Adaptive Effect. 84(7):1871–1875.

[200] Saskia Bucciarelli, Jin Suk Myung, Bela Farago, Shibananda Das, Gerard A. Vliegenthart, Olaf Holderer, Roland G. Winkler, Peter Schurtenberger, Gerhard Gompper, and Anna Stradner. Dramatic influence of patchy attractions on short-time protein diffusion under crowded conditions. 2(12):e1601432.

[201] Natalia A. Chebotareva, Dmitrii O. Filippov, and Boris I. Kurganov. Effect of crowding on several stages of protein aggregation in test systems in the presence of -crystallin. 80:358–365.

[202] Nicholas J. Ray, Damien Hall, and John A. Carver. Deamidation of N76 in human γS-crystallin promotes dimer formation. *Biochimica Et Biophysica Acta*, 1860(1 Pt B):315–324, January 2016.

[203] Ajay Pande, Natalya Mokhor, and Jayanti Pande. Deamidation of human γS-crystallin increases attractive protein interactions: implications for cataract. *Biochemistry*, 54(31):4890–4899, August 2015.

[204] M. Paniw, E. Gil-Cabeza, and F. Ojeda. Plant carnivory beyond bogs: reliance on prey feeding Drosophyllum lusitanicum (Drosophyllaceae) in dry Mediterranean heathland habitats. *Annals of Botany*, page mcw247, 2017.

[205] Tetsuo Kokubun. Occurrence of myo-inositol and alkyl-substituted polysaccharide in the prey-trapping mucilage of drosera capensis. *The Science of Nature*, 104(9-10), 2017.

[206] Elżbieta Król, Bartosz J. Płachno, Lubomír Adamec, Maria Stolarz, Halina Dziubińska, and Kazimierz Trębacz. Quite a few reasons for calling carnivores 'the most wonderful plants in the world'. *Annals of Botany*, 109(1):47–64, 2011.

[207] Gergo Palfalvi, Thomas Hackl, Niklas Terhoeven, Tomoko F. Shibata, Tomoaki Nishiyama, Markus Ankenbrand, Dirk Becker, Frank Förster, Matthias Freund, Anda Iosip, Ines Kreuzer, Franziska Saul, Chiharu Kamida, Kenji Fukushima, Shuji Shigenobu, Yosuke Tamada, Lubomir Adamec, Yoshikazu Hoshi, Kunihiko Ueda, Traud Winkelmann, Jörg Fuchs, Ingo Schubert, Rainer Schwacke, Khaled Al-Rasheid, Jörg Schultz, Mitsuyasu Hasebe, and Rainer Hedrich. Genomes of the Venus flytrap and close relatives unveil the roots of plant carnivory. *Current Biology*, page https://doi.org/10.1016/j.cub.2020.04.051, 2020.

[208] Carter T. Butts, Jan C. Bierma, and R. W. Martin. Novel proteases from the genome of the carnivorous plant *Drosera capensis*: structural prediction and comparative analysis. *Proteins: Structure, Function, and Bioinformatics*, 84(10):1517–1533, 2016.

[209] Carter T Butts, Xuhong Zhang, John E Kelly, Kyle W Roskamp, Megha H Unhelkar, J Alfredo Freites, Seemal Tahir, and Rachel W Martin. Sequence comparison, molecular modeling, and network analysis predict structural diversity in cysteine proteases from the Cape sundew, *Drosera capensis*. *Computational and Structural Biotechnology Journal*, 14:271–282, 2016.

[210] Megha H. Unhelkar, Vy T. Duong, Kaosoluchi N. Enendu, John E. Kelly, Seemal Tahir, Carter T. Butts, and Rachel W. Martin. Structure prediction and network analysis of chitinases from the Cape sundew, *Drosera capensis*. *Biochimica et Biophysica Acta*, 1861:636–643, 2017.

[211] Vy T. Duong, Megha H. Unhelkar, John E. Kelly, Suhn H. Kim, Carter T. Butts, and Rachel W. Martin. Network analysis provides insight into active site flexibility in esterase/lipases from the carnivorous plant *Drosera capensis*. *Integrative Biology*, 10:768–779, 2018.

[212] B.C. Bryksa, P. Bhaumik, E. Magracheva, D.C. De Moura, M. Kurylowicz, A. Zdanov, J.R. Dutcher, A. Wlodawer, and R.Y. Yada. Structure and mechanism of the saposin-like domain of a plant aspartic protease. *Journal of Biological Chemistry*, 286(32):28265–28275, 2011.

[213] Heike Bruhn. A short guided tour through functional and structural features of saposin-like proteins. *Biochemical Journal*, 389(2):249–257, 2005.

[214] Fernando Muñoz, M. Francisca Palomares-Jerez, Gustavo Daleo, José Villalaín, and M. Gabriela Guevara. Cholesterol and membrane phospholipid compositions modulate the leakage capacity of the swaposin domain from a potato aspartic protease (StAsp-PSI). *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1811(12):1038–1044, d 2011.

[215] Fernando F. Muñoz, M. Francisca Palomares-Jerez, Gustavo Daleo, José Villalaín, and M. Gabriela Guevara. Possible mechanism of structural transformations induced by StAsp-PSI in lipid membranes. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1838(1B):339–347, 2014.

[216] Brian C. Bryksa and Rickey Y. Yada. Protein structure insights into the bilayer interactions of the saposin-like domain of *Solanum tuberosum* aspartic protease. *Scientific Reports*, 7:16911, 2017.

[217] K. D'Hondt, S. Stack, S. Gutteridge, J. Vandekerckhove, E. Krebbers, and S. Gal. Aspartic proteinase genes in the Brassicaceae *Arabidopsis thaliana* and *Brassica napus*. *Plant Molecular Biology*, 33(1):187–192, 1997.

[218] Yiji Xia, Hideyuki Suzuki, Justin Borevitz, Jack Blount, Zejian Guo, Kanu Patel, Richard A. Dixon, and Chris Lamb. An extracellular aspartic protease functions in *Arabidopsis* disease resistance signaling. *The EMBO Journal*, 23:980–988, 2004.

[219] I. Simoes and C. Faro. Structure and function of plant aspartic proteinases. *European Journal of Biochemistry*, 271(11):2067–2075, 2004.

[220] Paul C. White, Maria C. Cordeiro, Daniele Arnold, Brodelius Peter E., and John Kay. Processing, activity, and inhibition of recombinant cyprosin, an aspartic proteinase from cardoon (*Cynara cardunculus*). *Journal of Biological Chemistry*, 274:16685–16693, 1999.

[221] Patricia Duarte, Jose Pissarra, and Ian Moore. Processing and tracking of a single isoform of the aspartic proteinase cardosin A on the vacuolar pathway. *Planta*, 227:1255–1268, 2008.

[222] Cláudia Pereira, Susana Pereira, Béatrice Satiat-Jeunemaitre, and José Pissarra. Cardosin A contains two vacuolar sorting signals using different vacuolar routes in tobacco epidermal cells. *The Plant Journal*, 76(1):87–100, 2013.

[223] C. Egas, N. Lavoura, R. Resende, R.M. Brito, Pires E., M.C. de Lima, and C. Faro. The saposin-like domain of the plant aspartic proteinase precursor is a potent inducer of vesicle leakage. *Journal of Bioloical Chemistry*, 275(49):38190–38196, 2000.

[224] Fernando F. Muñoz, Julieta R. Mendieta, Mariana R. Pagano, Roberto A. Paggi, Gustavo R. Daleo, and María G. Guevara. The swaposin-like domain of potato aspartic protease (StAsp-PSI) exerts antimicrobial activity on plant and human pathogens. *Peptides*, 31(5):777–785, 2010.

[225] Michael W. Risør, Line R. Thomsen, Kristian W. Sanggaard, Tania A. Nielsen, Ida B. Th/ogersen, Marie V. Lukassen, Litten Rossen, Irene Garcia-Ferrer, Tibisay Guevara, Carsten Scavenius, Ernst Meinjohanns, F. Xavier Gomis-Rüth, and Jan J. Enghild. Enzymatic and structural characterization of the major endopeptidase in the Venus flytrap digestion fluid. *The Journal of Biological Chemistry*, 291(5):2271–2287, 2016.

[226] M. Campbell, M. Law, C. Holt, J. Stein, G. Moghe, D. Hufnagel, J. Lei, R. Achawanantakun, D. Jiao, C. J. Lawrence, D. Ware, S. H. Shiu, K. L. Childs, Y. Sun, N. Jiang, , and M Yandell. MAKER-P: A tool-kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiology*, 164:513–524, 2013.

[227] E. Quevillon, V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler, and R. Lopez. InterProScan: Protein domains identifier. *Nucleic Acids Research*, 33:W116–W120, 2005.

[228] Fabian Sievers, Andreas Wilm, David Dineen, Toby J. Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, Hamish McWilliam, Michael Remmert, Johannes Söding, Julie D. Thompson, and Desmond G. Higgins. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7:539–539, 2011.

[229] S.V. Silva and Malcata F.X. Proteolysis of ovine caseins by cardosin A, an aspartic acid proteinase from *Cynara cardunculus* L. *Lait*, 78:513–519, 1998.

[230] S.V. Silva and F.X. Malcata. On the activity and specificity of cardosin B, a plant proteinase, on ovine caseins. *Food and Chemical Toxicology*, 67:373–378, 1999.

[231] K. Takahashi, H. Niwa, N. Yokota, K. Kubota, and H. Inoue. Widespread tissue expression of nepenthesin-like aspartic protease genes in Arabidopsis thaliana. *Plant Physiology and Biochemistry*, 46(7):724–729, 2008.

[232] D.E. Kim, D. Chivian, and D. Baker. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research*, 32(Supplement 2):W526–31, 2004.

[233] 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*, 77(Suppl 9):89–99, 2009.

[234] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng, and Thomas E Ferrin. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004.

[235] W Humphrey, A Dalke, and K Schulten. VMD: visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, 27–28, February 1996.

[236] F. William Studier. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, 41(1):207–234, 2005.

[237] Georg Faust, Alexandra Stand, and Dirk Weuster-Botz. IPTG can replace lactose in auto-induction media to enhance protein expression in batch-cultured *Escherichia coli*. *Engineering in Life Sciences*, 15(8):824–829, jun 2015.

[238] Arun Sivashanmugam, Victoria Murray, Chunxian Cui, Yonghong Zhang, Jianjun Wang, and Qianqian Li. Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Science*, 18(5):936–948, may 2009.

[239] E. G. Bligh and W. J. Dyer. A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology*, 37(8):911–917, 1959.

[240] V. Vandana, M. S. L. Karuna, P. Vijayalakshmi, and R. B. N. Prasad. A simple method to enrich phospholipid content in commercial soybean lecithin. *Journal of the American Oil Chemists' Society*, 78(5):555–556, may 2001.

[241] Mark D. Bird, William W. Brey, Timothy A. Cross, Iain R. Dixon, A. Griffin, Scott T. Hannahs, John Kynoch, Ilya M. Litvak, Jefferey L. Schiano, and Jack Toth. Commissioning of the 36 t series-connected hybrid magnet at the NHMFL. *IEEE Transactions on Applied Superconductivity*, 28(3):1–6, 2018.

[242] A. Pines, M. G. Gibby, and J. S. Waugh. Proton-enhanced nuclear induction spectroscopy. a method for high resolution NMR of dilute spins in solids. *Journal of Chemical Physics*, 56(4):1776–1777, 1972.

[243] K. Takegoshi, S. Nakamura, and T. Terao. 13C–1H dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chemical Physics Letters*, 344(5-6):631–637, 2001.

[244] Mats HM Olsson, Chresten R. Sondergaard, Michal Rostkowski, and Jan H. Jensen. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *Journal of Chemical Theory and Computation*, 7(2):525–537, 2011.

[245] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.

[246] James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kalé, and Klaus Schulten. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16):1781–1802, 2005.

[247] Robert B. Best, Xiao Zhu, Jihyun Shim, Pedro E.M. Lopes, Jeetain Mittal, Michael Feig, and Alexander D. Mackerell, Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles. *J Chem Theory Comput*, 8(9):3257–3273, Sep 2012.

[248] Glenn J. Martyna, Douglas J. Tobias, and Michael L. Klein. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics*, 101(5):4177–4189, September 1994.

[249] Scott E. Feller, Yuhong Zhang, Richard W. Pastor, and Bernard R. Brooks. Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics*, 103(11):4613–4621, September 1995.

[250] Alek Aksimentiev. Script to exclude water from lipid bilayers. Software File, 2016.

[251] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.

[252] Grant B.J., Rodrigues A.P.C., ElSawy K.M., McCammon J.A., and Caves L.S.D. Bio3d: An r package for the comparative analysis of protein structures. *Bioinformatics*, 22:2695–2696, Nov 2006.

[253] Matthew D. Homan and Andrew Gelman. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, January 2014.

[254] Stan Development Team. RStan: the R interface to Stan, 2020. R package version 2.19.3.

[255] N.D. Rawlings, M. Waller, A.J. Barrett, and A. Bateman. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, 42:D503–D509, 2014.

[256] K. Takahashi, S.B. Athauda, K. Matsumoto, S. Rajapakshe, M. Kuribayashi, M. Kojima, N. Kubomura-Yoshida, A. Iwamatsu, C. Shibata, and H. Inoue. Nepenthesin, a unique member of a novel subfamily of aspartic proteinases: enzymatic and structural characteristics. *Current Protein and Peptide Science*, 6(6):513–525, 2005.

[257] F. Buch, W. E. Kaman, F. J. Bikker, A. Yilamujiang, and A. Mithöfe. Nepenthesin protease activity indicates digestive fluid dynamics in carnivorous *Nepenthes* plants. *PLoS ONE*, 10(3):e0118853, 2015.

[258] Miguel A. Mazorra-Manzano, Takuji Tanaka, Derek R. Dee, and Rickey Y. Yada. Structure-function characterization of the recombinant aspartic proteinase A1 from *Arabidopsis thaliana*. *Phytochemistry*, 71(5):515–523, 2010.

[259] Marialetizia Motta, Massimo Tatti, and Rosa Salvioli. Autophagic dysfunction in Gaucher disease and its rescue by cathepsin B and D proteases. In *Autophagy: Cancer, Other Pathologies, Inflammation, Immunity, Infection, and Aging*, pages 131–146. Elsevier, 2014.

[260] V.E. Ahn, P. Leyko, J.R. Alattia, L. Chen, and G.G. Prive. Crystal structures of saposins A and C. *Protein Science*, 15:1849–1857, 2006.

[261] Maxim Rossmann, Robert Schultz-Heienbrok, Joachim Behlke, Natascha Remmel, Claudia Alings, Konrad Sandhoff, Wolfram Saenger, and Timm Maier. Crystal structures of human saposins C and D: Implications for lipid recognition and membrane interactions. *Structure*, 16(5):809–817, 2008.

[262] J.A. Gally and G.M. Edelman. The effect of temperature on the fluorescence of some aromatic amino acids and proteins. *Biochimica et Biophysica Acta*, 60(3):499–509, jul 1962.

[263] Jens Hennecke, Alain Sillen, Martina Huber-Wunderlich, Yves Engelborghs, and Rudi Glockshuber. Quenching of tryptophan fluorescence by the active-site disulfide bridge in the DsbA protein from Escherichia coli. *Biochemistry*, 36(21):6391–6400, may 1997.

[264] Jens Frauenfeld, Robin Löving, Jean-Paul Armache, Andreas F-P Sonnen, Fatma Guettou, Per Moberg, Lin Zhu, Caroline Jegerschöld, Ali Flayhan, John A G Briggs, Henrik Garoff, Christian Löw, Yifan Cheng, and Pär Nordlund. A saposin-lipoprotein nanoparticle system for membrane proteins. *Nature Methods*, 13:345–351, 2016.

[265] Chih-Ta Henry Chien, Lukas R. Helfinger, Mark J. Bostock, Andras Solt, Yi Lei Tan, and Daniel Nietlispach. An adaptable phospholipid membrane mimetic system for solution NMR studies of membrane proteins. *Journal of the American Chemical Society*, 139(42):14829–14832, 2017.

[266] Ali Flayhan, Haydyn D.T. Mertens, Yonca Ural-Blimke, Maria Martinez Molledo, Dmitri I. Svergun, and Christian Löw. Saposin lipid nanoparticles: A highly versatile and modular tool for membrane protein research. *Structure*, 26(2):345–355, 2018.

[267] Vladimir Ladizhansky. Applications of solid-state NMR to membrane proteins. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1865(11):1577–1586, nov 2017.

[268] Ilia G Denisov and Stephen G Sligar. Nanodiscs for structural and functional studies of membrane proteins. *Nature Structural & Molecular Biology*, 23(6):481–486, jun 2016.

[269] K. W. Zilm. Oral presentation at the 40th experimental nmr conference. In *Orlando, FL*, 1999.

[270] Ann McDermott, Tatyana Polenova, Anja Böckmann, Kurt. W. Zilm, Eric K. Paulson, Rachel W. Martin, and Gaetano T. Montelione. Partial nmr assignments for uniformly (^{13}C, ^{15}N)-enriched BPTI in the solid state. *Journal of Biomolecular NMR*, 16:209–219, 2000.

[271] Tatyana I. Igumenova, Ann E. McDermott, Kurt W. Zilm, Rachel W. Martin, Eric K. Paulson, and A. Joshua Wand. Assignments of carbon NMR resonances for microcrystalline ubiquitin. *Journal of the American Chemical Society*, 126:6720–6727, 2004.

[272] W. Trent Franks, Donghua H. Zhou, Benjamin J. Wylie, Brian G. Money, Daniel T. Graesser, Heather L. Frericks, Gurmukh Sahota, and Chad M. Rienstra. Magic-angle spinning solid-state NMR spectroscopy of the β1 immunoglobulin binding domain of protein G (GB1): ^{15}N and ^{13}C chemical shift assignments and conformational analysis. *Journal of the American Chemical Society*, 127:12291–12305, 2005.

[273] Federica Castellani, Barth van Rossum, Annette Diehl, Mario Schubert, Kristina Rehbein, and Hartmut Oschkinat. Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature*, 420(6911):99–102, November 2002.

[274] Adam Lange, Stefan Becker, Karsten Seidel, Karin Giller, Olaf Pongs, and Marc Baldus. A concept for rapid protein-structure determination by solid-state NMR spectroscopy. *Angewandte Chemie*, 44(14):2089–2092, 2005.

[275] Lindsay J. Sperling, Deborah A. Berthold, Terry L. Sasser, Victoria Jeisy-Scott, and Chad M. Rienstra. Assignment strategies for large proteins by magic-angle spinning NMR: the 21-kDa disulfide bond forming enzyme DsbA. *Journal of Molecular Biology*, 399(2):268–282, 2010.

[276] Theofanis Manolikas, Torsten Herrmann, and Beat H. Meier. Protein structure determination from ^{13}C spin-diffusion solid-state NMR spectroscopy. *Journal of the American Chemical Society*, 130(12):3959–3966, 2008.

[277] Patrick C. A. van der Wel. New applications of solid-state NMR in structural biology. *Emerging Topics in Life Sciences*, 2(1):57–67, 2018.

[278] Lichi Shi, Mumdooh A.M. Ahmed, Wurong Zhang, Gregg Whited, Leonid S. Brown, and Vladimir Ladizhansky. Three-dimensional solid-state NMR study of a seven-helical integral membrane proton pump–Structural insights. *Journal of Molecular Biology*, 386(4):1078–1093, 2009.

[279] Shakeel Ahmad Shahid, Benjamin Bardiaux, W. Trent Franks, Ludwig Krabben, Michael Habeck, Barth-Jan van Rossum, and Dirk Linke. Membrane-protein structure determination by solid-state NMR spectroscopy of microcrystals. *Nature Methods*, 9(12):1212–1217, 2012.

[280] Venkata S. Mandala, Jonathan K. Williams, and Mei Hong. Structure and dynamics of membrane proteins from solid-state NMR. *Annual Review of Biophysics*, 47:201–222, 2018.

[281] Eldon L. Ulrich, Hideo Akutsu, Jurgen F. Doreleijers, Yoko Harano, Yannis E. Ioannidis, Jundong Lin, Miron Livny, Steve Mading, Dimitri Maziuk, Zachary Miller, Eiichi Nakatani, Christopher F. Schulte, David E. Tolmie, R. Kent Wenger, Hongyang Yao, and John L. Markley. Biomagresbank. *Nucleic Acids Research*, 36:D402–D408, 2008.

Appendix A

A.1 Supplement for Human αB-crystallin discriminates between aggregation-prone and function-preserving variants of a client protein.

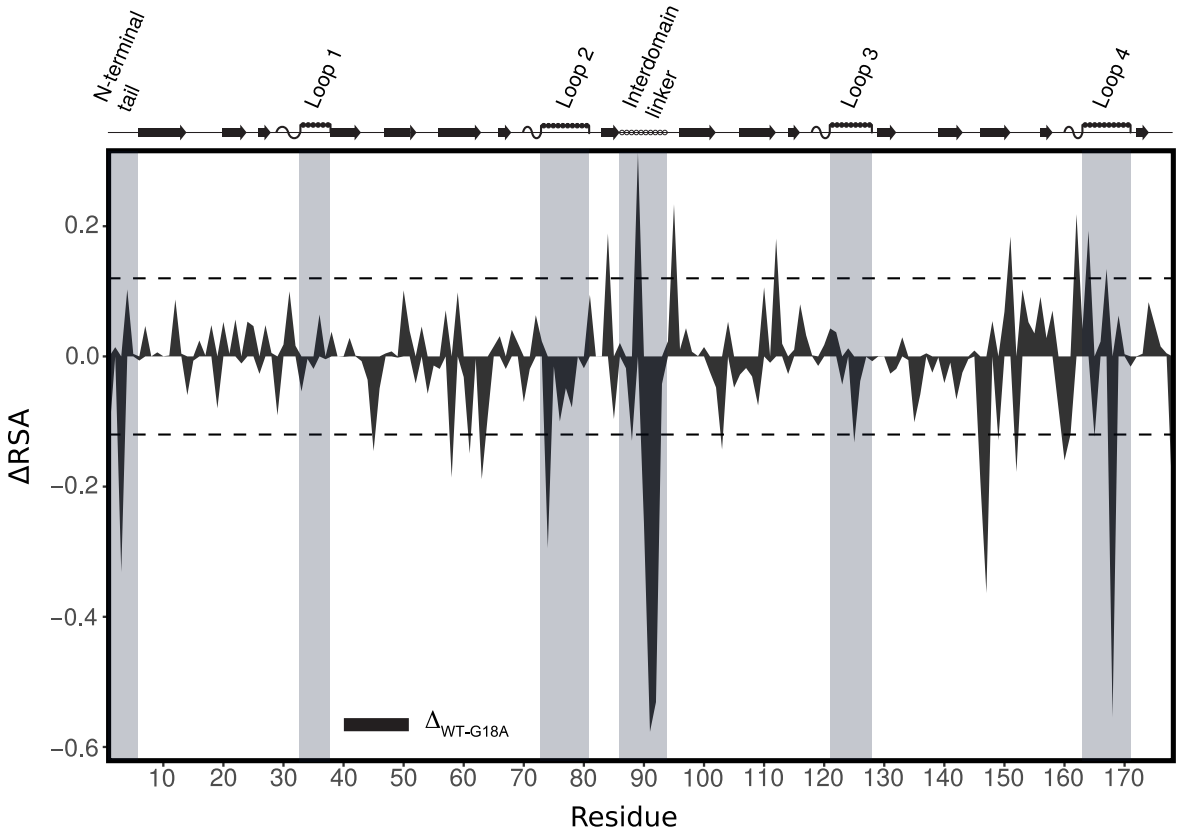

Figure A.1: The difference between relative solvent accessibility of γS-WT and γS-G18A on a per residue basis. There is no large scale difference in relative solvent accessibility between γS-G18A and γS-WT with only a few residues falling outside of the RMS. The dotted lines represent the RMS = ± 0.115

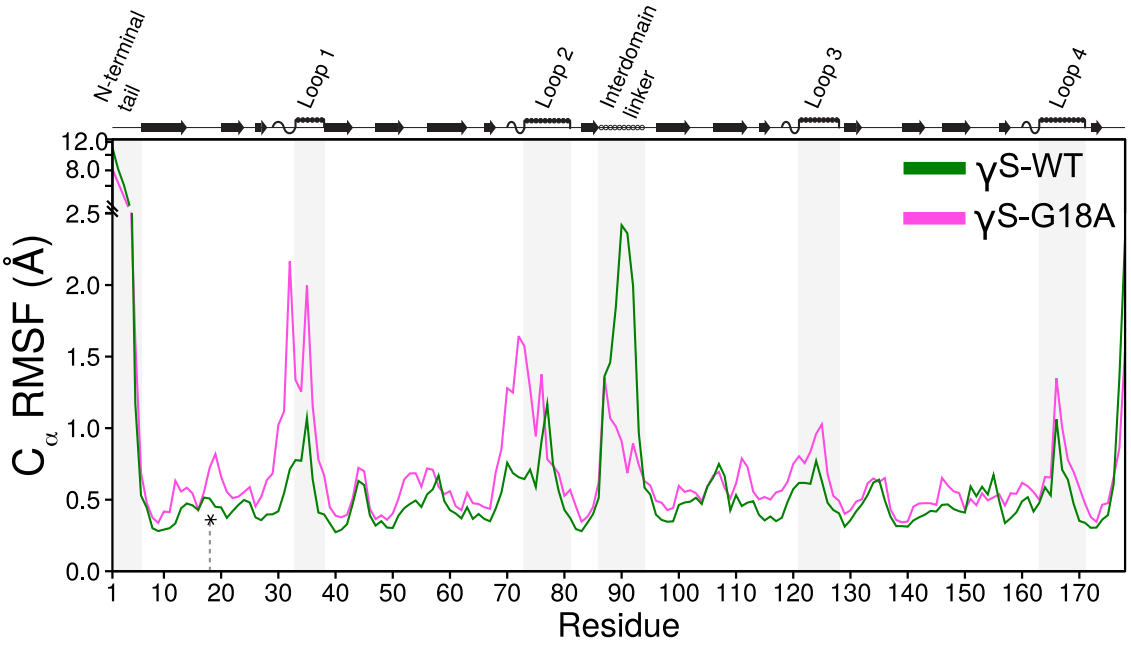

Figure A.2: C_α root mean squared fluctuation (RMSF) per residue. γS-G18A exhibits increased thermal backbone fluctuations in the loops 30-35 and 65-75. The asterisks indicates the position of residue 18. β-strand, helix, and loop secondary structure features are indicated above the graph.

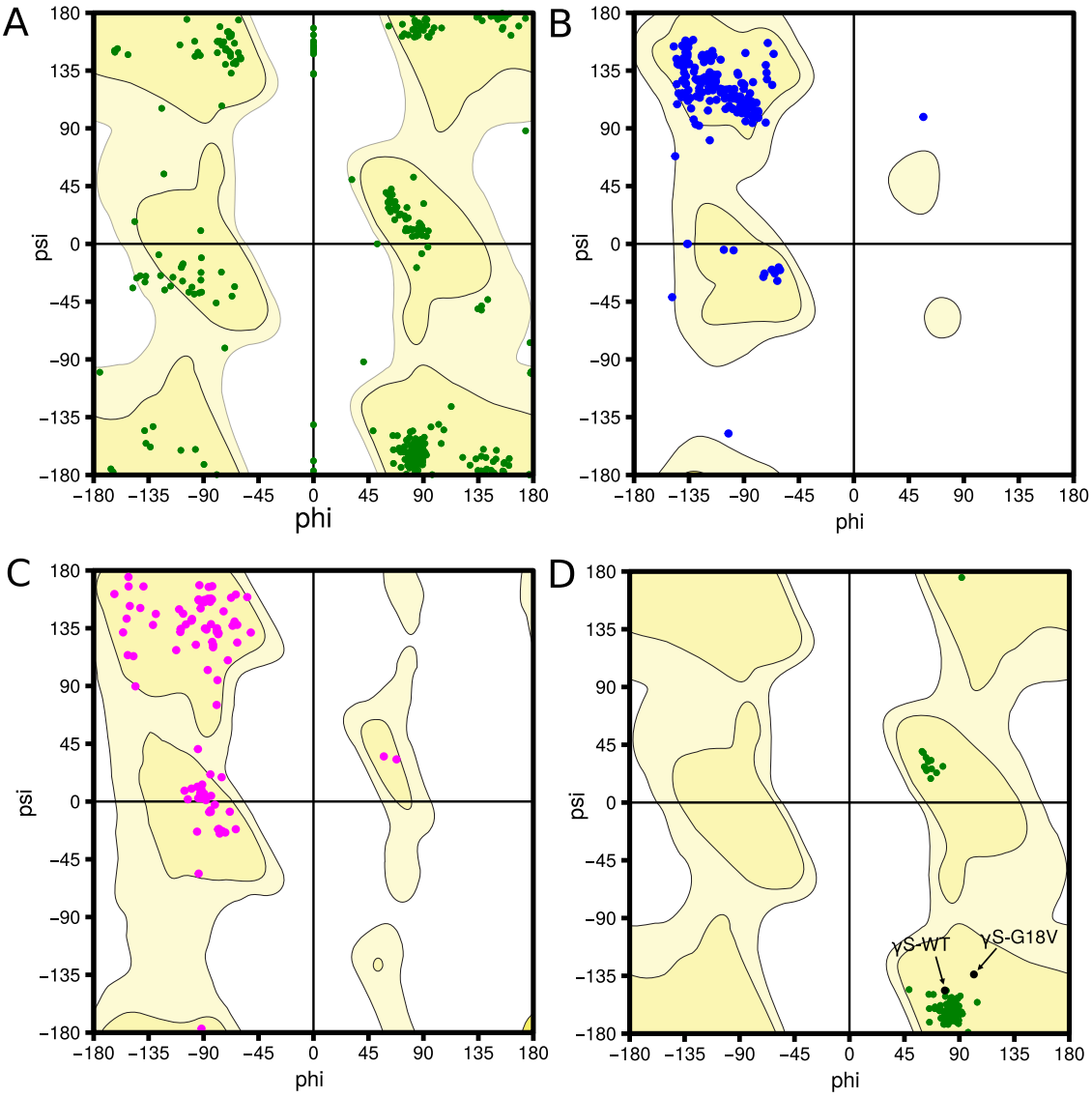

Figure A.3: Ramachandran plots for all native (**A**) glycines, (**B**) valines , and (**C**) alanines in all the available lens $\beta\gamma$-crystallin structures. (**D**) Ramachandran plots for all the glycines in the homologous beta-turn position to γS-WT G18 from all available γ-crystallin structures (shown in green), as well as G18 in γS-WT and V18 in γS-G18V(shown in black). In general, Ala and Val ψ-ϕ populations are highly similar compared to Gly. Therefore, we might expect the A18 torsion angles to be more comparable to V18 than G18.

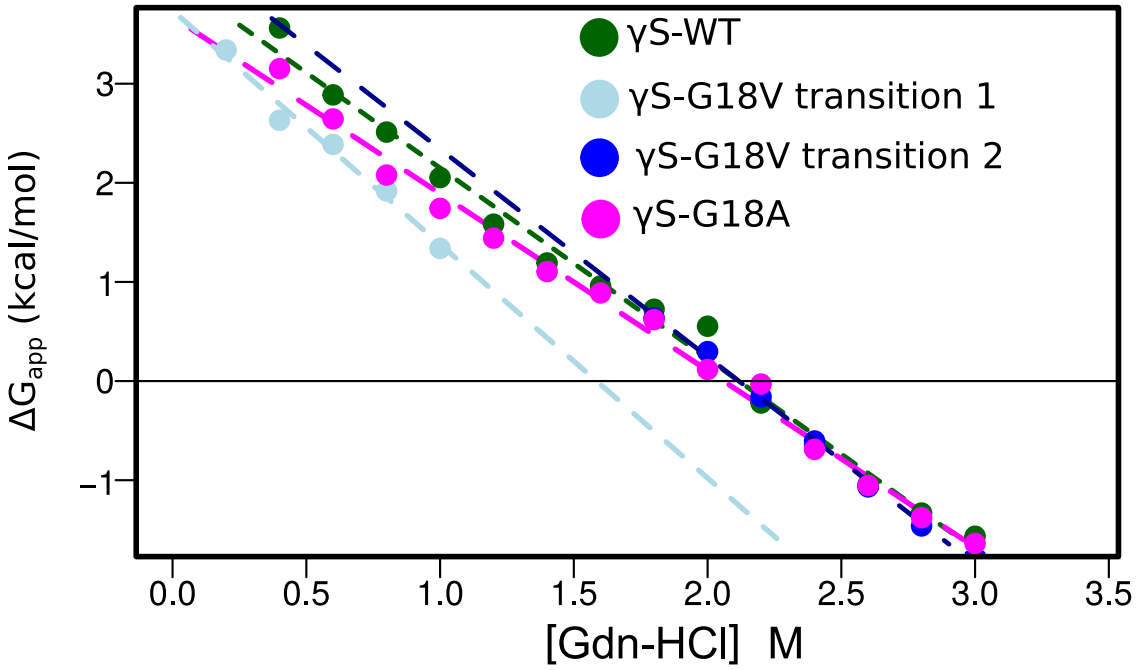

Figure A.4: The ΔG_{app} of γS-WT (green), γS-G18A (magenta), the first transition of γS-G18V (light blue), and the second transition of γS-G18V (dark blue). A line was fitted to each data set and lines were extrapolated to 0 to calculate the ΔG of unfolding for each transition.

Appendix B

B.1 Supplement The Droserasin 1 PSI: A Membrane-Interacting Antimicrobial Peptide from the Carnivorous Plant Drosera capensis.

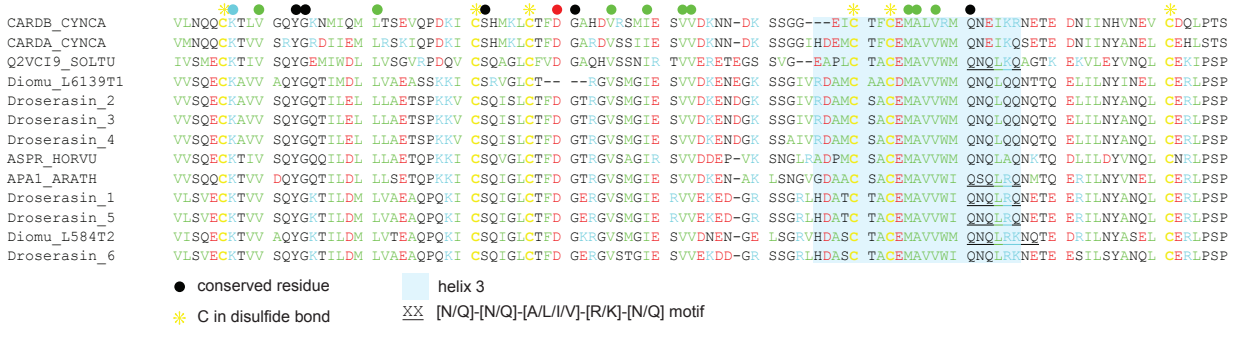

Figure B.1: Sequence alignment of PSIs from *D. capensis* and *D. muscipula*, along with reference sequences from *Arabidopsis thaliana* and *Cynara cardunculus*.

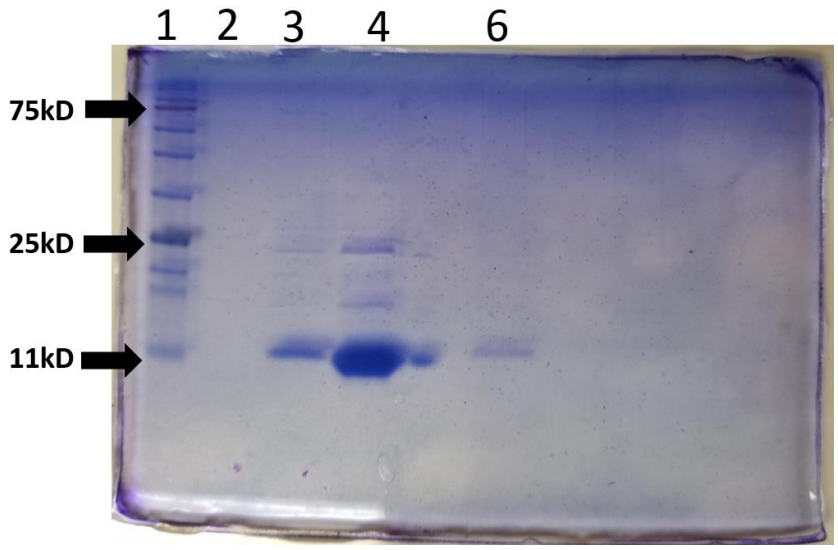

Figure B.2: This 12% SDS-PAGE shows D1-PSI purified from an anion exchange column in 50% ethanol. Lane 1 shows Blue Stain Protein Ladder from Gold Biotechnology (Olivette, MO), Lane 2 is the flow through of the cell lysate from the column. Lanes 3, 4 and 6 are D1-PSI eluted off the column with 250mM NaCl, 500mM NaCl and 1M NaCl, respectively.

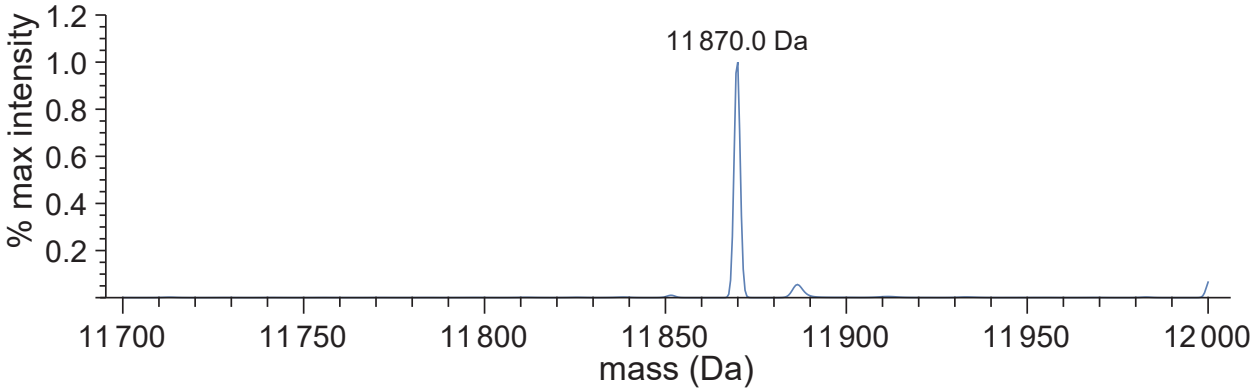

Figure B.3: Mass of D1-PSI was obtained via electrospray mass spectrometry with the Waters (Micromass) Quattro Premier XE machine. We confirm a mass size of 11.87kD for D1-PSI.

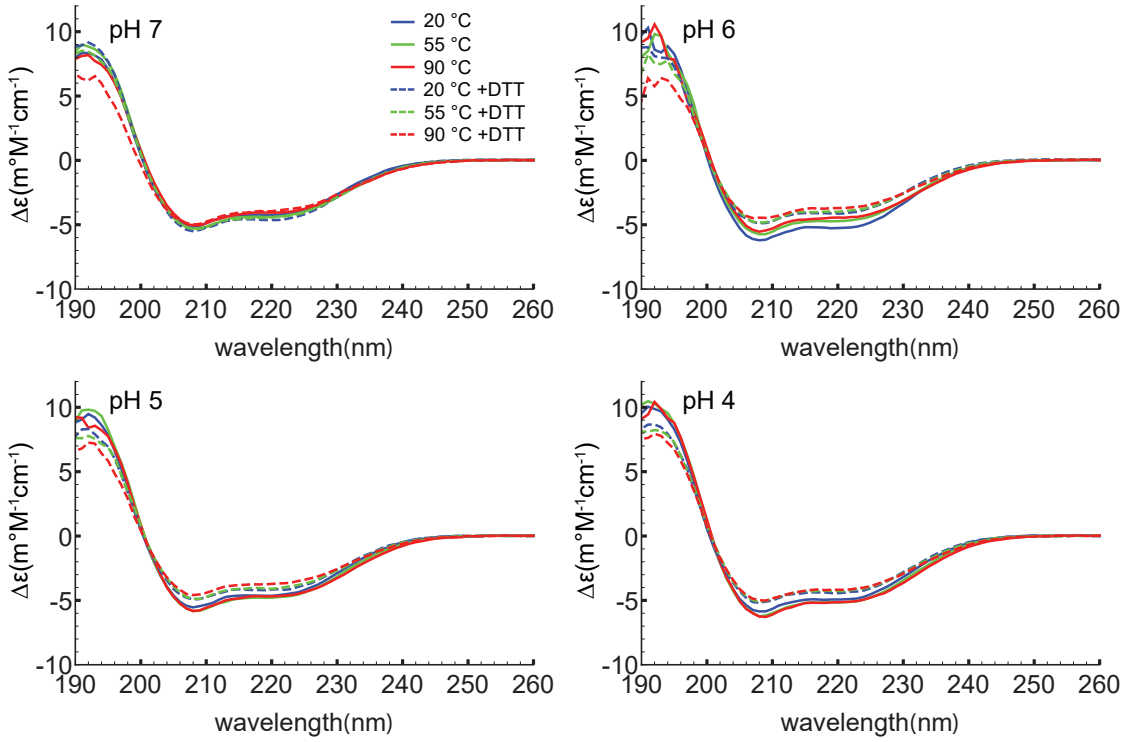

Figure B.4: Circular dichroism spectra of the D1 PSI at pH 4, 5, 6, and 7. Spectra were collected at 20, 55 and 90 °C in the presence (dashed lines) or absence (solid lines) of DTT. Spectra indicate that the secondary structure of the protein is highly stable under all conditions tested. Furthermore, reduction of disulfide bonds by DTT apparently results in only small changes to the secondary structure.

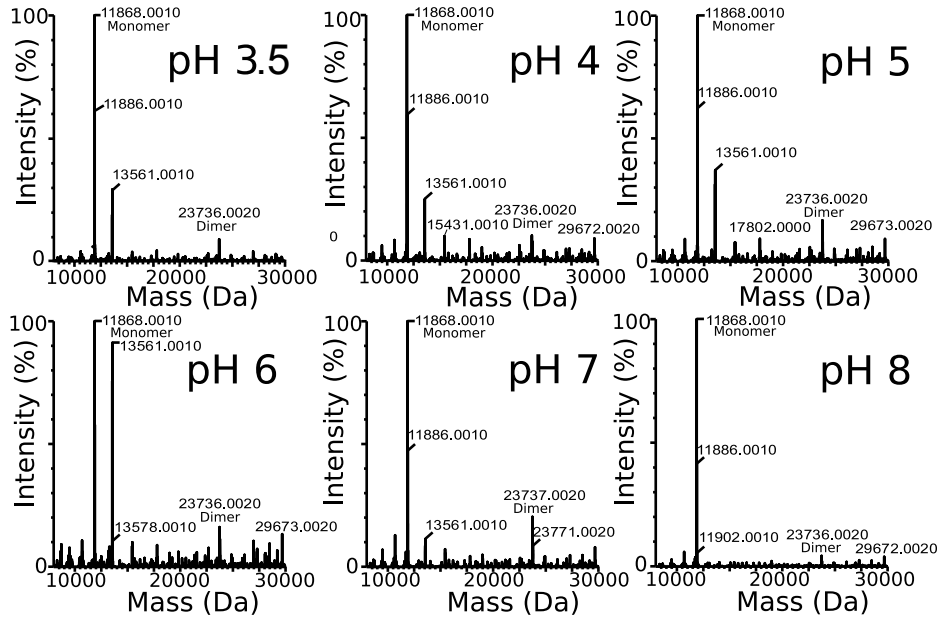

Figure B.5: Mass spectrometry of intact D1-PSI at pH 3.5, 4, 5, 6, 7, 8 as calculated using the maximum entropy deconvolution algorithm. At each pH tested, D1-PSI exists predominately as a monomer (11868 Da) , with a minor dimer peak appearing at each pH (23737 Da). The relative amounts of monomer to dimer does not change significantly as the pH of the solution changes and the monomer peak predominates at pH values where D1 PSI would be active.

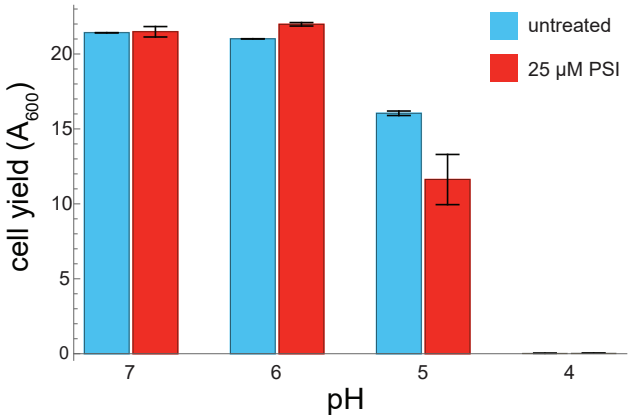

Figure B.6: Pilot test of antimicrobial activity. The yeast *P. pastoris* was grown at pH 4, 5, 6, 7 with or without 25 μM D1 PSI. Two replicates were performed. Error bars indicate standard errors. At close to neutral pH of 6 and 7 there was little to no difference between the treated and untreated cell yield (pH 7: Mean difference (treatment - control): 0.064, t=0.184, p=0.884 95% CI for mean difference: -4.355, 4.483. pH 6: Mean difference (treatment - control): 0.969, t=8.338, p=0.075 95% CI for mean difference: -0.483, 2.421.). At pH 4, very little cell growth was observed (Mean difference (treatment - control): 0.006, t=3.674, p=0.067 95% CI for mean difference: -0.001, 0.014). At pH 5, where the D1 PSI is hypothesized to be active, treated cells showed a reduced overall cell yield, indicating antimicrobial activity of D1-SI (Mean difference (treatment - control): -4.422, t=-2.631, p=0.228, 95% CI for mean difference: -25.000, 16.152.) The results from this pilot experiment are qualitatively compatible with our hypothesis that the D1 PSI inhibits microbial growth at its active pH, but the number of replicates is too small to reliably draw strong conclusions (results not significant at 0.05 under a two-sample t-test). Future studies will include more replicates and a finer gradation of pH values between 5 and 6.

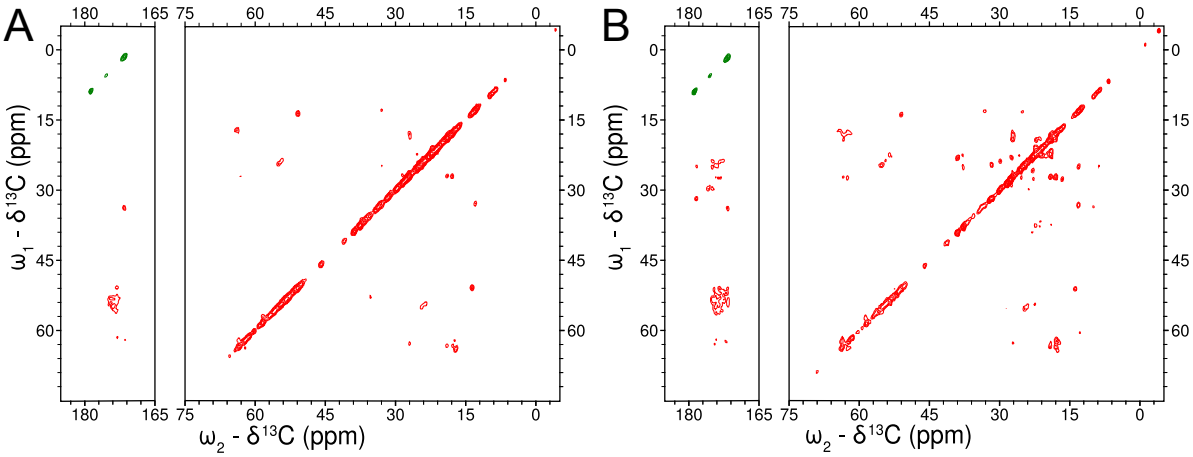

Figure B.7: CO (left) and aliphatic (right) regions of ^{13}C-^{13}C CP DARR spectra of D1 PSI in a 1:1 POPC, POPA membrane system collected at 10 °C and spinning at 24.4 kHz with (**A**) a 50 ms mixing time and (**B**) a 300 ms mixing time. Both spectra show well-resolved off-diagonal peaks, although the intensities are different, reflecting the different mixing times.

Supplementary Table 1: Aspartic protease PSI PDB files available for download.

protein	organism	conformation	file name
APA1_ARATH	*A. thaliana*	open	APA1_ARATH_PSI_xmonomer_m1.pdb
APA1_ARATH	*A. thaliana*	closed	APA1_ARATH_PSI_smonomer_m2.pdb
Diomu_L6139T1	*D. muscipula*	open	Diomu_L6139T1_PSI_xmonomer_m2.pdb
Diomu_L6139T1	*D. muscipula*	closed	Diomu_L6139T1_PSI_smonomer_m1.pdb
Diomu_L584T2	*D. muscipula*	open	Diomu_L584T2_PSI_xmonomer_m1.pdb
Diomu_L584T2	*D. muscipula*	closed	Diomu_L584T2_PSI_smonomer_m2.pdb
Droserasin 1	*D. capensis*	open	Droserasin1_PSI_xmonomer_m1.pdb
Droserasin 1	*D. capensis*	closed	Droserasin1_PSI_smonomer_m2.pdb
Droserasin 1	*D. capensis*	dimer	Droserasin1_PSI_dimer.pdb
Droserasin 2	*D. capensis*	open	Droserasin2_PSI_xmonomer_m1.pdb
Droserasin 2	*D. capensis*	closed	Droserasin2_PSI_smonomer_m2.pdb
Droserasin 3	*D. capensis*	open	Droserasin3_PSI_xmonomer_m1.pdb
Droserasin 3	*D. capensis*	closed	Droserasin3_PSI_smonomer_m2.pdb
Droserasin 4	*D. capensis*	open	Droserasin4_PSI_xmonomer_m1.pdb
Droserasin 4	*D. capensis*	closed	Droserasin4_PSI_smonomer_m2.pdb
Droserasin 5	*D. capensis*	open	Droserasin5_PSI_xmonomer_m1.pdb
Droserasin 5	*D. capensis*	closed	Droserasin5_PSI_smonomer_m2.pdb
Droserasin 6	*D. capensis*	open	Droserasin6_PSI_xmonomer_m1.pdb
Droserasin 6	*D. capensis*	closed	Droserasin6_PSI_smonomer_m2.pdb