UNIVERSITY OF CALIFORNIA,
IRVINE

Exploring Cold-Adapted Eye Lens Proteins and Discovery of an Antimicrobial Protein from a Carnivorous Plant

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Jan C. Bierma

Dissertation Committee:
Professor Rachel W. Martin, Chair
Professor Melanie J. Cocco
Professor Gregory A. Weiss
Professor Douglas J. Tobias

2019

DEDICATION

To those who were forgotten.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank Professor Rachel Martin for accepting me into her lab and giving me the freedom to work on a variety of projects. My degree would not be possible without the opportunities she provided. Also, to Professor Carter Butts for at times being like a second advisor

I also want to thank Professor Melanie Cocco, Professor Gregory Weiss and Professor Douglas Tobias for agreeing to be on my thesis committee and writing recommendation letters for me.

I should also recognize the other Martin lab members for being generally helpful and being the Martin lab members that they are.

I would also like to thank those from other labs for sharing reagent, equipment, etc. You should know who you are.

Thanks to Dmitry Fishman for all of his help and excellent management of the UCI Optical Spectroscopy Facility. Felix Grun and Ben Katz for excellent management of the UCI Laser Spectroscopy Labs and for assistance with mass spectrometry data collection and analysis.

This work was made possible by the Genomic High Throughput Facility Shared Resource of the Cancer Center Support Grant (CA-62203) at the University of California, Irvine and NIH shared instrumentation grants 1S10RR025496-01 and 1S10OD010794-01. A portion of this work was performed at the National High Magnetic Field Laboratory, which is supported by the National Science Foundation Cooperative Agreement No. DMR-1644779 and the State of Florida. This work was supported by the National Science Foundation awards DMR 1410415, DMS 1361425, and CHE 1308231 and National Institutes of Health awards 1R01EY021514 and 1R01EY025328 to RWM, CTB, and collaborators.

CURRICULUM VITAE

Jan C. Bierma

EDUCATION

Doctor of Philosophy in Molecular Biology & Biochemistry **2013-2019**
University of California, Irivne *Irvine, CA*

Bachelor of Science in Biochemistry & Biophysics **2008-2012**
Oregon State University *Corvallis, OR*

RESEARCH EXPERIENCE

Structural Biology **2014–2019**
University of California, Irvine *Irvine, California*

Devlopmental Biology **2013–2014**
University of California, Irvine *Irvine, California*

Microbiology **2010–2012**
Oregon State University *Corvallis, Oregon*

PUBLICATIONS

1. **Controlling Liquid-Liquid Phase Separation of Cold-Adapted Crystallin Proteins from the Antarctic Toothfish** J.C. Bierma, K. Roskamp, A. Ledray, A.J. Kiss, C.H.C. Cheng, and R.W. Martin, *J. Mol. Biol.* **2018**, *430*, 5151-5168

2. **Protein refractive index increment is determined by conformation as well as composition** D. Khago, J.C. Bierma, K.W. Roskamp, N. Kozlyuk, and R.W. Martin, *J. Phys. Condens. Matter* **2018**, *30*, 435101.

3. **Calcium binding dramatically stabilizes an ancestral crystallin fold in tunicate betagamma-crystallin** N. Kozlyuk, S. Sengupta, J.C. Bierma, and R.W. Martin, *Biochemistry.* **2016**, *55*, 6961-6968.

4. **Novel proteases from the genome of the carnivorous plant *Drosera capensis*: Structural prediction and comparative analysis.** C.T. Butts, J.C. Bierma, and R.W. Martin, *Proteins.* **2016**, *84*, 1517-33.

5. **Stability of protein-specific hydration shell on crowding.** K.Y. Huang, C.N. Kingsley, R. Sheil, C.Y. Cheng, J.C. Bierma, K.W. Roskamp, D. Khago, R.W. Martin, and S. Han, *J. Am. Chem. Soc.* **2016**, *138*, 5392-5402.

6. **Dissecting fission yeast shelterin interactions via MICro-MS links disruption of shelterin bridge to tumorigenesis.** J. Liu, C. Yu, X. Hu, J.K. Kim, J.C. Bierma, H.I. Jun, S.D. Rychnovsky, L. Huang, and F. Qiao, *Cell. Rep.* **2015**, *12*, 2169-2180.

7. **γs-crystallin proteins from the antarctic nototheniid toothfish: A model system for investigating differential resistance to chemical and thermal denaturation.** C.N. Kingsley, J.C. Bierma, V. Pham, and R.W. Martin, *J. Phys. Chem. B.* **2014**, *118*, 13544-13553.

8. **Characterization of a grape class IV chitinase.** S. Vincenzi, J. Bierma, S.I. Wickramasekara, A. Curioni, D. Gazzola, and A.T. Bakalinsky *J. Agric. Food. Chem.* **2014**, *62*, 5660-5668.

9. **Acetic acid inhibits nutrient uptake in *Saccharomyces cerevisiae*: auxotrophy confounds the use of yeast deletion libraries for strain improvement.** J. Ding, J. Bierma, M.R. Smith, E. Poliner, C. Wolfe, A.N. Hadduck, S. Zara, M. Jirikovic, K. van Zee, M.H. Penner, J. Patton-Vogt, and A.T. Bakalinsky, *Appl. Microbiol. Biotechnol.* **2013**, *97*, 7405-7416.

PRESENTATIONS

Oral Presentations:

1. Uncovering the role of surface residues and buffer composition in liquid-liquid phase separation of eye lens crystallins from an Antarctic toothfish. 63rd Biophysical Society Meeting, March, 2019.

2. Discovery and characterization of a membrane interacting protein from the carnivorous plant *D. capensis*. 2018 Southern California Users of Magnets, May, 2018.

3. Determination of cold adaptation in Antarctic toothfish lens proteins by structural comparison. 251st American Chemical Society National Meeting, March, 2016.

Poster Presentations:

1. Uncovering the role that surface residues play in liquid-liquid phase separation of Antarctic toothfish eye lens crystallins. 62nd Biophysical Society Meeting, February, 2018

2. Characterization of cold adaptation in Antarctic toothfish eye lens proteins. 253st American Chemical Society National Meeting, April, 2017

3. Structural studies of exotic eye lens crystallins. 58st Experimental Nuclear Magnetic Resonance Conference, March, 2017.

4. Determination of cold adaptation in Antarctic toothfish lens proteins by structural comparison. 56nd Experimental Nuclear Magnetic Resonance Conference, April, 2015.

ABSTRACT OF THE DISSERTATION

Exploring Cold-Adapted Eye Lens Proteins and Discovery of an Antimicrobial Protein from a Carnivorous Plant

By

Jan C. Bierma

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2019

Professor Rachel W. Martin, Chair

This thesis explores the eye lens proteins from the Antarctic toothfish *Dissostichus mawsoni* and how they have evolved to the subfreezing temperatures of their environment. This includes showing that toothfish γS1- and γS2-crystallins are less stable than their homologous human counterparts and working toward solving their structures using solution-state NMR. By making structural and biophysical comparisons, inferences can be made about how these crystallins have become cold adapted. Another unique adaptation of *D. mawsoni* is the ability to completely resist cold cataract, a liquid-liquid phase separation of the proteins. By studying γM-crystallins from *D. mawsoni* that are susceptible to phase separation and performing site specific mutagenesis it was discovered the temperature of phase separation could be controlled by simply swapping between lysine and arginine residues. These results hint at hydration effects and salt bridges as being a major factor influencing the crystallin's propensity to self associate into a separate liquid phase. To further characterize these proteins their functional role of providing refractive power was measured. The results of these measurements demonstrated that all the crystallins tested had a refractive index much higher than would be predicted by their amino acid composition. This suggests that protein conformation has a large impact on protein refractivity and the assumed models used to predict protein refractivity must be approached much more carefully.

Also described is the D1-PSI discovered in the genome of the carnivorous plant *Drosera capensis*. As a saposin-like protein, it has demonstrated the ability to interact with membranes in a way that can exhibit anti-microbial growth. Results show that this PSI is able to disrupt membranes, but seems to lack bias for which lipid head groups it interacts with in the context of a stable lipoprotein complex. To better understand what is structurally happening to the D1-PSI while interacting with a membrane, solid-state NMR experiments are ongoing. Such information can inform the mechanism by which D1-PSI is capable of disrupting membranes.

Chapter 1

The γS-crystallin proteins from the Antarctic notothenioid toothfish: a model system for investigating differential resistance to chemical and thermal denaturation

1.1 Background

Proteins require optimized stability and flexibility to perform their biological roles, including interacting with binding partners, responding to their environment and resisting aggregation. Protein function depends on both structure and dynamics; making the folded state of a protein more stable by rigidifying it does not necessarily lead to enhanced functionality.[1] Comparisons among homologous proteins from thermophilic, mesophilic, and psychrophilic

organisms have often shown that these proteins are comparably flexible when each is considered at its physiologically relevant temperature, even though in thermodynamic terms adaptation to higher temperature often correlates with higher stability.[2] The crystallins, the structural proteins of the eye lens, are unusually stable and thus present an attractive model system for studying questions of protein stability. These proteins create the high refractive index necessary for this specialized tissue to focus light on the retina. Unlike in land animals, where the air-water interface at the cornea provides a significant amount of focusing power, in aquatic organisms the crystallins alone produce the refractive capability of the eye. Fish lenses therefore have both increased protein concentrations and greater refractivity of the proteins themselves in comparison to their mammalian counterparts. While the crystallin concentrations in mammalian lenses can reach up to 450 mg/mL, lenses belonging to aquatic organisms reach up to 1,000 mg/mL [3, 4].

In vertebrates there are two common types of lens proteins: the $\beta\gamma$-crystallins, which are primarily structural, and the α-crystallins, which have an additional function of binding damaged structural proteins and preventing aggregation[5]. These two protein families have different evolutionary histories and distinct structures[6]. The α-crystallins are believed to have resulted from the gene duplication of an ancestral α-crystallin domain; these proteins are closely related to each other and to other chaperone proteins[7, 8]. The structural crystallins, including the taxon-specific crystallins found in many organisms, and the $\beta\gamma$-crystallins that are the focus of this study have been recruited from diverse abundant, soluble proteins via gene sharing or duplication often followed by selection for increased refractivity of the protein itself[9, 10, 11]. The $\beta\gamma$-crystallin superfamily, which is characterized by two double Greek key domains, is thought to be derived from a calcium-binding motif that existed before the evolution of eye lenses,[12] as evidenced by similarity to calcium-binding proteins in sequences from archaea,[13] slime mold,[14] and urochordate[15]. The urochordate (*Ciona intestinalis*) protein in particular is highly similar to the vertebrate members of the family, but with two notably different features: it has only one domain rather than two, and it contains two

calcium binding sites. Despite this common evolutionary history, functional mammalian lens proteins lack calcium-binding activity [16]. A study of the structure and dynamics of zebrafish (*Danio rerio*) γM7-crystallin by solution-state NMR has revealed a potentially general unfolding pathway for all $\beta\gamma$-crystallin domains [17]. Although this protein has the same overall fold as the mammalian $\beta\gamma$-crystallins, there are significant primary sequence differences, including enhanced methionine content as well as the absence of some of the tryptophan residues that are strongly conserved in mammals. In teleost fishes, the γM-crystallins are the most common structural proteins in the lens, while β-crystallins are more common in humans [18]. Although γS-crystallins are a relatively minor subclass in both cases, they were chosen as the focus of this study because their amino acid sequences are strongly conserved among all vertebrates and because of their ability to resist cold cataract [19].

Crystallins, particularly in fish lenses, are enriched in highly polarizable amino acids and depleted in aliphatic amino acids as a result of their selection for high refractive index [20]. This selective pressure can sometimes work against the selective pressure for cold tolerance, which favors a relatively high proportion of hydrophobic residues. In comparing crystallin proteins from different environments, two notions of protein stability are relevant [21]. The thermodynamic stability, ΔG° of unfolding, is the difference in Gibbs free energy between the folded and unfolded states, measured by reversible denaturation of the protein [22]. The unfolding temperature, T_m, which is measured by (usually irreversible) direct thermal denaturation [23]. Although they are not directly comparable, the thermal and chemical stabilities of similar proteins are often positively correlated; in a series of homologs or variants, it is common for the ordering of thermal and chemical denaturation resistance to follow the same ordinal ranking. In the case of the eye lens crystallins from the Antarctic toothfish (*Dissostichus mawsoni*), comparing the thermal and chemical stabilities of two closely related proteins provides insight into the different sequence characteristics related to their two major functions; cold tolerance and high refractive index.

The Antarctic toothfish lives in the cold waters of the Southern Ocean, where temperatures can be as cold as -2 °C. This large fish has a relatively long lifespan (≈ 50 years). Its lens proteins are therefore resistant to both age-related loss of solubility and the formation of cold cataract [24]. *D. mawsoni* has two γS-crystallin paralogs, γS1- and γS2-crystallin (abbreviated γS1 and γS2 throughout), with a sequence identity of 60%. Protein turnover is very low in the eye lens, requiring the crystallins to maintain their stability and solubility over the whole lifespan of the organism. In mesophilic organisms, high stability corresponds to a high thermal denaturation temperature and high ΔG° of unfolding. Quantitative thermodynamic studies of cold-adapted proteins so far have generally found decreased thermodynamic stability [25, 26, 27, 28]. In general, psychrophilic proteins are characterized by decreased core hydrophobicity, depleted isoleucine content, increased surface hydrophobicity, fewer total charged residues, increased surface charge, a lower arginine/lysine ratio, weaker interdomain and intersubunit interactions, decreased secondary structure content, more and longer loops, more glycine residues, fewer and weaker metal-binding sites, fewer disulfide bonds, fewer electrostatic interactions, and increased conformational entropy of the unfolded state [29]. Some of these adaptations conflict with the primary optical function of the γ-crystallins, for which highly polarizable amino acids are selected. Quantitatively, this is described by the refractive index increment (dn/dc); the change in refractive index with concentration. Although this can be empirically determined, for proteins it is often assumed to be described by a simple additive model in which only the amino acid content is important for determining dn/dc for the entire protein molecule [10].

Here we compare the thermal and chemical stabilities of *D. mawsoni* γS1- and γS2- crystallins in light of these competing functions. Although these proteins have comparable values for ΔG° of unfolding, surprisingly, γS1 is more susceptible to thermal denaturation while γS2 is more readily unfolded with urea. For related proteins, these quantities are typically positively correlated with each other and with overall thermodynamic stability. The differential resistance to thermal and chemical unfolding in this system underscores the different

mechanisms of unfolding involved, and the intramolecular interactions involved in resistance to them.

1.2 Materials and Methods

1.2.1 Gene construction, expression, and purification

Plasmids containing the cDNA sequences of the human γS-crystallin (hγS), *D. mawsoni* γS1 (GenBank, DQ143971.1) and γS2 (GenBank, DQ143972.1) genes [30] were purchased from Blue Heron Biotech, LLC. (Bothell, WA). Each gene was flanked by regions containing restriction sites for NcoI and XhoI, an N-terminal 6x His tag, and a TEV cleavage sequence (ENLFQG) with the N-terminal methionine of hγS, γS1 and γS2 replaced by the final glycine in the cleavage sequence. The toothfish crystallin genes were amplified using oligonucleotide primers purchased from Sigma-Aldrich (St. Louis, MO) and the resulting gene products were individually cloned into pET28a(+) vectors (Novagen, Darmstadt, Germany). hγS, γS1 and γS2 were overexpressed in Rosetta (DE3) *Escherichia coli* using standard IPTG-induced overexpression protocols at 25 °C in standard Luria broth (LB). Cells were allowed to grow for 16-24 hours post induction. The cells were lysed by sonication and cell debris was removed by centrifugation. His-TEV-hγS, His-TEV-γS1 and His-TEV-γS2 were purified on a Ni-NTA column (Applied Biosystems, Foster City, CA). The pure protein was collected from the column elution fraction and then dialyzed extensively against 10 mM phosphate buffer, pH 6.9, for all experiments.

1.2.2 Circular dichroism

Purified γS1 and γS2 was diluted to 0.125 mg/mL with 10 mM phosphate buffer at pH 6.9 for the collection of full circular dichroism (CD) spectra, and to 0.25 mg/mL with 10 mM phosphate buffer at pH 6.9, 150 mM NaCl, and 1 mM DTT for unfolding experiments. Measurements were taken on a J-810 spectropolarimeter (JASCO, Easton, MD) equipped with a thermal controller. For unfolding measurments, the samples were heated at a rate of 2 °C/min. For thermal denaturation curves, the CD at 218 nm was monitored and the curves were fit to a two-state equilibrium unfolding model to determine the thermal denaturation temperature (T_m).

1.2.3 Fluorescence spectroscopy

UV fluorescence measurements were made on γS1 and γS2 at a concentration of 0.075 mg/mL in 10 mM phosphate buffer, pH 6.9. Samples for chemical unfolding curves were prepared with increasing concentrations of 10 M urea (Fisher Scientific, Waltham, MA). Urea stock solutions were prepared as outlined by Pace *et al.* [31]. Samples were allowed to equilibrate for at least 24 hours before absorption-emission fluorescence spectra were obtained using a F4500 fluorescence spectrophotometer (Hitachi, Tokyo, Japan) with a λ_{ex} of 280 nm. The ratio of baseline-corrected emission intensities at 360 nm and 320 nm was used for analysis. To determine the thermodynamic parameters (ΔG°_{w} and m-values), $\Delta G_{[urea]}$ was calculated from the normalized equilibrium unfolding data and a linear least squares fit was performed in Mathematica to the line:

$$\Delta G^{\circ}_{w} = m[urea] \tag{1.1}$$

where ΔG°_{w} is the value of ΔG at 25 ° C, extrapolated to zero concentration of denaturant, and m is a measure of the dependence of ΔG on denaturant concentration.

1.2.4 Dynamic light scattering

Dynamic light scattering (DLS) measurements were obtained with a Zetasizer Nano ZS (Malvern Instruments, Malvern, UK) on γS1 and γS2 at a concentration of 1.0 mg/mL in 10 mM phosphate buffer, pH 6.9. At each temperature, the sample was allowed to equilibrate for 2 min before measurements were obtained, after which scattering measurements were performed in triplicate, resulting in a heating rate of ∼0.5 °C/min.

1.2.5 Transmittance

Transmittance was obtained using a Cary 4000 UV-Vis (Agilent Technologies, Santa Clara, US) on hγS, γS1 and γS2 at a concentration of 10.0 mg/mL in 10 mM phosphate buffer, pH 6.9 from 25 °C to 5°C.

1.3 Results and Discussion

1.3.1 Primary sequence analysis suggests adaptation for high refractivity

The structural βγ-crystallins share a common fold consisting of paired homologous double Greek key domains, each with two sets of four adjacent antiparallel β-strands linked by short loops. This protein architecture has been identified as contributing to the very high stability of the βγ-crystallins [32]. The stability of the fold is in part due to the tight interdomain

interface that contains several critical hydrophobic interactions [33]. Homology models based on the solution structure of hγS(PDBID 2M3T) [34] were constructed using SwissModel [35]. and are shown in Fig. 1.1.

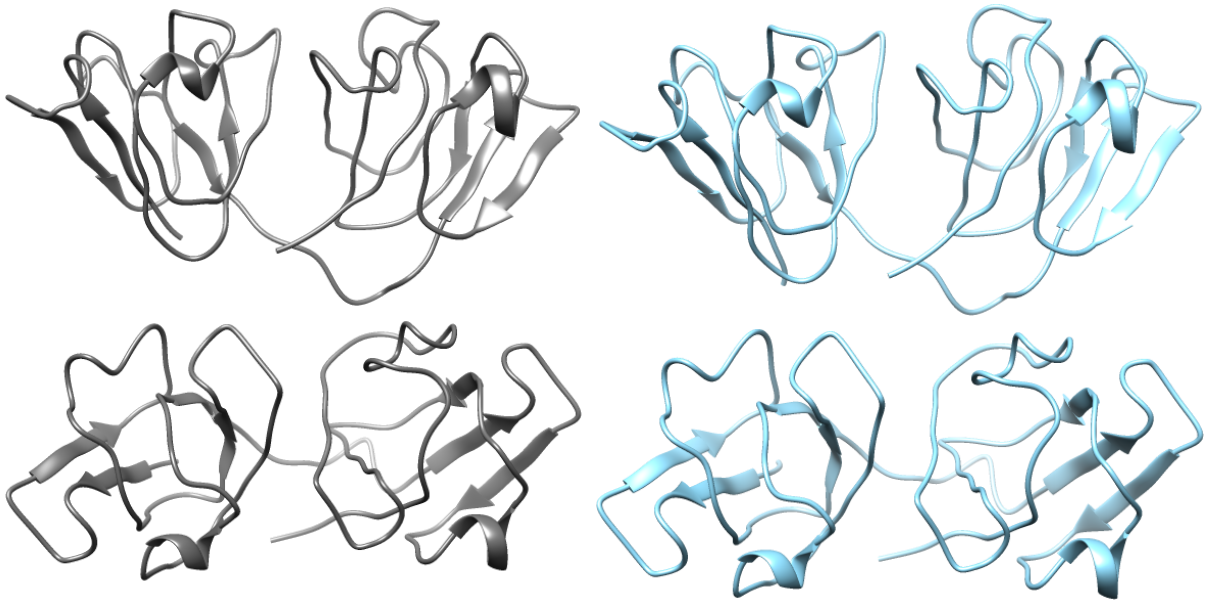

Figure 1.1: Homology models of *D. mawsoni* γS1 (grey) and γS2 (blue) based upon the solution structure of human γS-crystallin.

γS-crystallin is found in many vertebrates, and its primary sequence is highly conserved accross diverse species including fish, mammals, and birds [36, 37, 38, 39]. A sequence alignment for hγS, γS1, and γS2 is shown in Fig. 1.2. Both γS1 and γS2 have moderate sequence identity to hγS, (57% and 53%, respectively). Conserved residues include the four tryptophans, several glycines in loops, and other structurally important residues.

In general, the amino acid composition of proteins is relatively constant, as it primarily depends on factors such as the codon redundancy and mutation tolerance [40, 41], such that deviation from the average amino acid frequency is often indicative of selection for a particular function or environmental adaptation. For example, thermophilic proteins are often enriched in arginine because of its importance in forming stabilizing salt bridges [42]. Many psychrophilic proteins are depleted in proline content, possibly because of the reduced low-

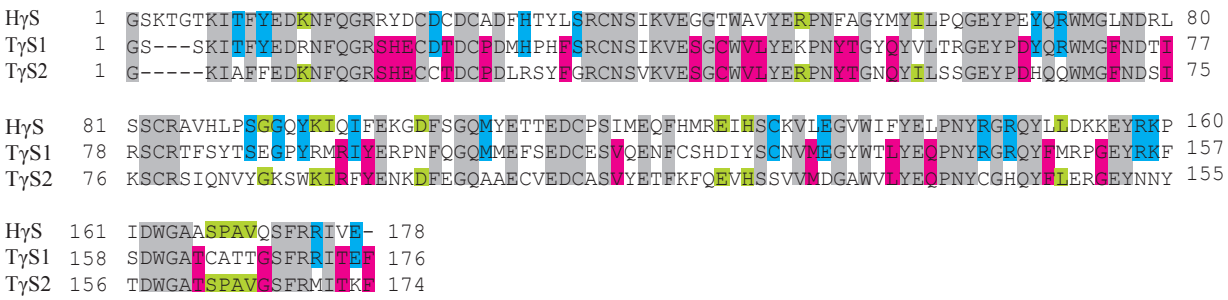

Figure 1.2: Sequence alignment of *D. mawsoni* γS1- and γS2-crystallin protein sequences with human γS-crystallin. *D. mawsoni* γS1 and γS2 have overall sequence identities of 57% and 53% with γS-crystallin, respectively, and 60% sequence identity to one another. Residues are highlighted according to their pattern of conservation among the three crystallin sequences examined. Gray: residues that are conserved in all three proteins. Blue: residues that are identical in hγS and γS1. Green: residues that are identical in hγS and γS2. Magenta: residues that are identical in γS1 and γS2.

temperature activity of the peptidyl-prolyl cis-trans isomerases that isomerize the peptide bond preceding the proline [29, 43]. Eye lens proteins, which have been selected for their high refractive index increments are enriched in highly polarizable amino acids such as Trp, Tyr, Phe, Arg, Met, and Cys, and depleted in aliphatic residues [11]. The amino acid frequencies for hγS, γS1, and γS2, are given in Table 1.1 along with the average values for vertebrate proteins and the contribution of each amino acid type to dn/dc.

The Grand AVerage of hydropathicitY index (GRAVY) predicts the hydrophobic character of the proteins where more positive values indicate higher hydrophobicity. Table 1.2 summarizes the sequence characteristics of γS1, γS2, and hγS as calculated from the ExPASy ProtParam tool [44]. γS2 is the most hydrophobic, yet it has a lower aliphatic index than hγS, reflecting its enhanced content of aromatic residues. The aliphatic index measures the relative volume occupied by aliphatic side chains (non-polar, hydrophobic) including alanine, valine, isoleucine and leucine. Of the three, hγS has much higher aliphatic character followed by γS2 and lastly, γS1 with the least aliphatic character. Alanine, valine, isoleucine and leucine are expected to be selected against in lens proteins because these amino acids have low refractive index increments. Consistent with this idea, γS1 has the lowest alipha-

Table 1.1: Refractive index increment by amino acid type

amino acid type	dn/dc from ref. [10] (mL/g)	frequency average (%)	frequency γS1 (%)	frequency γS2 (%)	frequency hγS (%)
Ala (A)	0.167	7.4	1.1	4.0	3.9
Arg (R)	0.206	4.3	9.1	4.6	7.3
Asn (N)	0.192	4.4	4.5	5.7	2.8
Asp (D)	0.197	5.9	5.1	5.2	5.6
Cys (C)	0.206	3.3	5.1	5.2	3.9
Gln (Q)	0.186	3.7	4.5	5.2	5.1
Glu (E)	0.183	5.8	8.5	8.0	7.9
Gly (G)	0.175	7.4	6.8	7.5	8.4
His (H)	0.219	2.9	2.3	2.3	2.2
Ile (I)	0.179	3.8	3.4	3.4	5.6
Leu (L)	0.173	7.6	1.7	2.9	5.1
Lys (K)	0.181	7.2	2.3	5.2	5.6
Met (M)	0.204	1.8	4.5	2.3	2.8
Phe (F)	0.244	4.0	6.8	6.9	5.1
Pro (P)	0.165	5.0	4.5	2.9	4.5
Ser (S)	0.170	8.1	8.5	8.6	6.2
Thr (T)	0.172	6.2	6.8	3.4	3.9
Trp (W)	0.277	1.3	2.3	2.9	2.2
Tyr (Y)	0.240	3.3	9.1	7.5	7.9
Val (V)	0.172	6.8	2.8	6.3	3.9

Table 1.2: Sequence analysis of γS1, γS2, and hγS, crystallins

	GRAVY	aliphatic index	# -ve charged residues	# +ve charged residues	predicted dn/dc	predicted pI	Charge neutral pH
Human γS	-0.685	56.97	24	23	0.1983	6.4	-0.9
Toothfish γS1	-0.957	29.32	24	20	0.2020	5.6	-4.0
Toothfish γS2	-0.651	47.01	23	17	0.2002	5.2	-6.0

tic index, as well a higher predicted dn/dc than hγS. γS1 has more hydrophobic residues than hγS while γS2 has fewer hydrophobic residues than either hγS or γS1 according to the GRAVY index. However, the lower aliphatic character and depletion of valine, alanine, isoleucine, and leucine in the sequences of γS1 and γS2 are also consistent with the need for these proteins to have increased flexibility in order to maintain cold tolerance. Selection for cold tolerance often leads to decreased numbers of hydrophobic residues in the protein core, and more on the surface, reflecting the reduced entropy cost of exposing hydrophobic groups to solvent at low temperatures. However, selection for increased refractive index leads to depletion of aliphatic residues in favor of those with more polarizable side chains [10, 11]. This would lead to the expectation that the *D. mawsoni* crystallins would have an unusually high proportion of aromatic amino acids to compensate, which is the case; γS1, which has the highest predicted dn/dc, has 9.1% Tyr, almost three times the average value, and is also enriched in Phe (6.8% vs. the average of 4.0%). Similarly, γS2 and hγS are enriched in Phe, Tyr and Trp, although to a lesser extent than γS1.

Other highly polarizable residues that would be expected to be enriched in lens proteins include Arg, Met, Cys, and His. However, Arg is often depleted in cold-adapted proteins due to its role in the formation of stabilizing salt bridges. In this case, both γS1 and hγS are significantly enriched in Arg, while γS2 has approximately the average value. This may provide insight into the denaturation behavior of the *D. mawsoni* crystallins described in the next section. In contrast to the zebrafish γM-crystallin, in which methionine content is greatly enriched, only γS1 (4.5%) has a greater level than the average (1.8%), while γS2

and hγS have more moderate levels (2.3% and 2.8%, respectively). Taken together, these observations suggest that rather than any one residue being critical for determining the high dn/dc values required for function in fish lens proteins, this function can be acquired by different combinations of high-refractivity amino acids, within the constraints imposed by the additional selection for cold tolerance in the *D. mawsoni* proteins.

The toothfish crystallins have lower isoelectric point (pI) values with respect to hγS, meaning that they have much more negative charge at neutral pH. This is consistent with the previous observation that psychrophilic proteins often have more acidic pI values than homologous mesophilic or thermophilic proteins, possibly because negatively charged residues are important for mediating interactions with the solvent and hence maintaining flexibility in cold environments [25].

1.3.2 γS1 and γS2 are both folded, with primarily β-sheet secondary structure

Using PsiPred, a secondary structure prediction software, [45, 46] the sequences for both γS1 and γS2 were predicted to have primarily β-sheet secondary structures. The prediction results are shown in Fig. 1.3.

Circular dichroism (CD) spectra were collected for both γS1 and γS2 to assess the overall general secondary structures of the proteins (Fig. 1.4). A comparison of the circular dichroism spectra of γS1 and γS2 at 25°C indicates that both proteins have primarily β-sheet secondary structures. The negative ellipticities of γS1 and γS2 occur at 216 nm and 217 nm, respectively. These values are in the range that is typical of β-sheet proteins and is consistent with the predicted secondary structure results and with the experimental results for hγS [47].

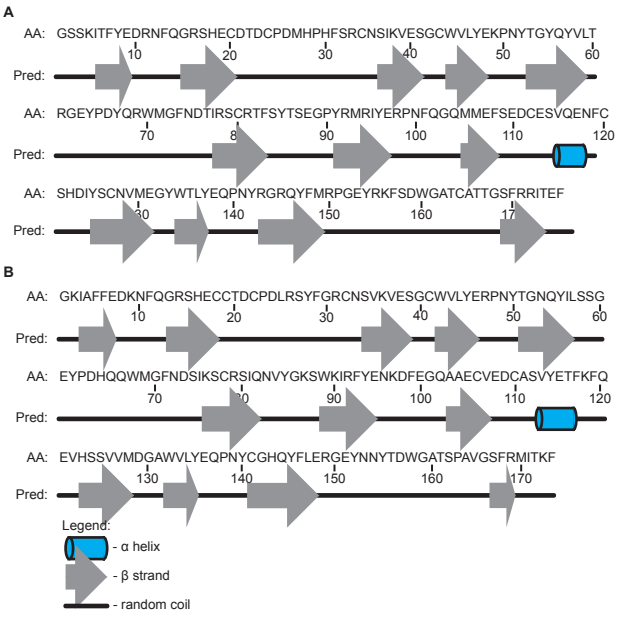

Figure 1.3: Predicted secondary structures of *D. mawsoni* (A) γS1 and (B) γS2.

The intrinsic fluorescence of crystallin proteins is an important indication of the degree of folded structure in the double Greek key domains, as the fluorescence is primarily due to four highly conserved tryptophan residues in the protein core. Both γS1 and γS2 share these conserved tryptophans; γS2 also has a fifth tryptophan. In mammalian eye lenses, the positioning of the conserved tryptophan side chains is essential for the rapid quenching of UV fluorescence hypothesized to protect crystallins from photochemical degradation in species that are subject to strong UV light exposure [48]. Their conservation in this aquatic species may be due to their contributions to the dn/dc, hydrophobic packing, or both. UV fluorescence spectra for *D. mawsoni* γS1 and γS2 are shown in Fig. 1.4. The emission maxima in the fluorescence spectra for excitation at 280 nm are 331 nm and 336 nm for γS1 and γS2, respectively. γS2 has increased fluorescence intensity in comparison to γS1 due to the presence of the additional tryptophan in its sequence. The reported λ_{max} for tryptophan fluorescence of hγS is 326 nm [47]. Typically, tryptophan fluorescence emission maxima are in the range of 300 to 350 nm. Tryptophans that are exposed to water have emission maxima

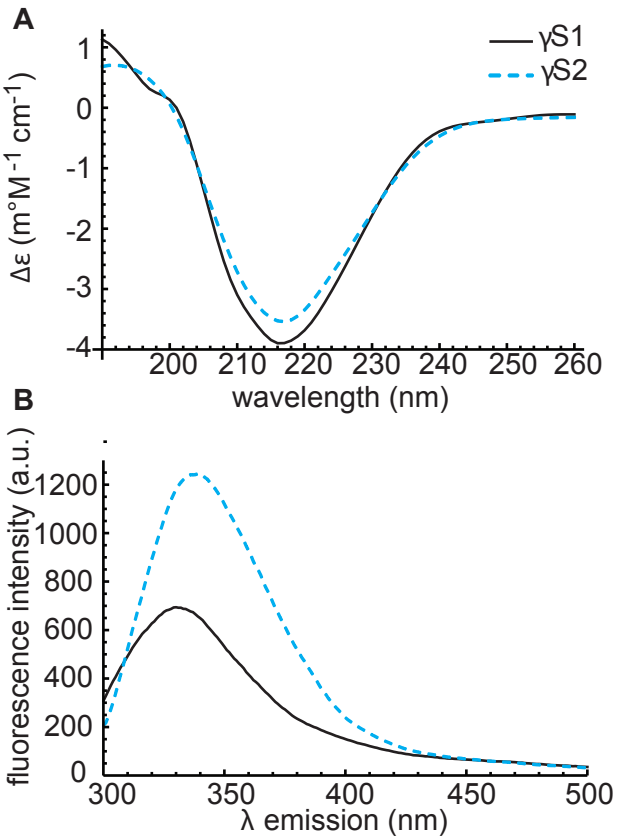

Figure 1.4: A. Circular dichroism spectra of *D. mawsoni* γS1- and γS2-crystallins. γS1 displays a negative ellipticity at 216 nm while γS2 displays a negative ellipticity at 217 nm. Both of these values are indicative of primarily β-sheet secondary structures. B. Tryptophan fluorescence emission spectra of γS1 and γS2. γS1 has an emission maximum at 331 nm and the emission maximum for γS2 is 336 nm.

between 340–350 nm whereas completely buried tryptophans have maxima around 330 nm. The slight red shifts of both γS1 and γS2 with respect to hγS indicate that the tryptophans in both toothfish proteins are more exposed to water, suggesting that γS1 and γS2 are less compactly folded and more structurally flexible than hγS, as expected.

Cold denaturation, protein unfolding due to the decreased energetic cost of exposing hydrophobic residues to solvent, does not occur for most globular proteins until well below the freezing point of water. Except in special cases, experimental studies of cold denaturation have required the use of chemical denaturants, high pressure [49], encapsulation in reverse

micelles [50], or limiting the sample volume to small capillaries [51] to study the unfolding intermediates. Thus, cold cataract, the low-temperature opacity of many protein solutions such as mammalian lenses, results from liquid-liquid phase separation rather than protein unfolding. The toothfish eye lens does not undergo cold cataract until -12 °C, in contrast to mammalian lenses, which form them at much higher temperatures (≈20 °C) [24]. γS-crystallins in general are resistant to cold cataract, and are thought to play an important role in maintaining solubility in multicomponent crystallin mixtures; e.g. bovine γS-crystallin has a theoretical liquid-liquid phase separation temperature of −28 °C [52] and its presence in concentrated solutions of other γ-crystallins results in a lowered phase separation temperature [53]. Transmission measurements at 600 nm were taken as a function of temperature in order to establish that solutions of our *in vitro*- generated protein constructs are transparent over the relevant temperature range (i.e., cold cataract does not occur). These results, summarized in Table 1.3 indicate that hγS, γS1, and γS2 all remain transparent at 5°C, consistent with previous measurements of γS-crystallins. Although solutions of each of the γ-crystallins studied here remain transparent at 5 °C, further investigations will be needed to assess their ability to stabilize mixtures of other crystallins as a function of temperature. Enhanced low-temperature solubilizing activity has been previously observed in cold-adapted αA-crystallins, which have greater hydrophobic character, and are better able to maintain their chaperone activity at lower temperatures than their mesophilic orthologs at the cost of high thermal stability [54]. Having two γS-crystallins with varying degrees of hydrophobic surface character may be an important adaptation for avoiding cold cataract in the toothfish lens.

Table 1.3: Transmittance of hγS, γS1 and γS2 from 25 °C to 5 °C.

Transmittance (λ = 600 nm)	hγS	γS1	γS2
25 °C	0.93	0.98	0.97
20 °C	0.95	0.93	0.97
15 °C	0.95	0.91	0.97
10 °C	0.94	0.88	0.97
5 °C	0.93	0.87	0.97

1.3.3 γS1 and γS2 have different relative stabilities under chemical and thermal denaturation

The overall thermodynamic stability of the γS-crystallin fold is highly relevant to the biological function of the eye lens, because of the lack of protein turnover in the lens; the crystallins must remain stable and soluble for decades. In general, cold-stable proteins are generally more susceptible to chemical denaturation than their higher-temperature counterparts [29]. Factors affecting overall protein stability include hydrophobic interactions, hydrogen bonds, and conformational entropy.

γS1 and γS2 were subjected to chemical denaturation with increasing concentrations of urea, while fluorescence spectroscopy was used to monitor unfolding. Each sample was allowed to equilibrate for at least 24 hours before fluorescence measurements were collected. The excitation wavelength was 280 nm and emission spectra were collected between 300–500 nm. Fluorescence maximum intensities were normalized by taking the F360/320 ratio at each concentration of denaturant and plotted as fraction unfolded vs. denaturant concentration (Fig. 1.5)A. The data points indicated with open circles in Fig. 1.5A represent dilution of the samples to 2M urea after full unfolding at 7M urea, demonstrating the reversibility of this transition. Fig. 1.5B is a plot of ΔG vs. denaturant concentration for the transition regions of the unfolding curves used to extrapolate values for ΔG°_{w}. γS1 is more susceptible to unfolding by urea; the $[\text{urea}]_{1/2}$ of γS1 is equal to 3.8 M and 5.6 M for γS2. ΔG°_{w} is 13.35 kJ mol^{-1} and 18.74 kJ mol^{-1} for γS1 and γS2, respectively. The thermodynamic parameters

Table 1.4: Thermodynamic parameters from chemical denaturation of γS1 and γS2.

	γS1	γS2	hγS
$[urea]_{1/2}$ (M)	3.8	5.6	6.3
ΔG°_{w} (kJ/mol)	13.35	18.74	27.12
m (kJ/mol·M)	3.51	3.35	4.33
T_m (°C)	68.5 ± 0.1	58.0 ± 0.1	72.0 ± 0.1 [47]

calculated from these denaturation curves are summarized in Table 1.4. Urea unfolding is thought to be driven by a combination of indirect and direct mechanisms [55]. It weakens the hydrophobic effect by disrupting the hydrogen bonding network of the solvent, as well as stabilizing unfolded states via direct electrostatic and hydrogen bonding interactions with side chain and backbone groups [56, 57, 58]. More urea may be needed to unfold γS2 because it has nearly twice as many aliphatic residues as γS1 and fewer hydrophilic amino acid residues. The slope, m, describing the dependence of ΔG on denaturant concentration, is very similar for both γS1 and γS2.

Thermal denaturation provides complementary information regarding protein stability and aggregation propensity. For some proteins. e.g. lysozyme [59], the thermally denatured state has been shown to differ form that induced by chemical denaturation. Furthermore, this process is often irreversible, making the calculation of thermodynamic quantities problematic; however the midpoint of the unfolding transition (T_m) is itself a useful measure of protein stability. The thermal denaturation of γS1 and γS2 were monitored by circular dichroism at 218 nm (Fig. 1.6) and fit to a two-state equilibrium unfolding model. The CD melting curves obtained provide information about the overall stability of the protein folds. The difference in thermal stabilities between the two proteins is quite different; γS1 has a T_m of 68.5 ± 0.1°C, while γS2 had a T_m of 58.0 ± 0.1°C. For comparison, human γS-crystallin has a T_m of 72.0 ± 0.1°C under the same conditions [47]. The relationship between thermal stability, hydrophobicity, and aliphatic index is not immediately clear for these proteins; γS2, the most hydrophobic crystallin studied, has the lowest T_m while γS1, the least hydrophobic, does not have the highest T_m value. The highest T_m belongs to hγS,

which has a hydrophobic content in between those of γS1 and γS2. The explanation for this surprising observation may be a result of selection for high dn/dc. γS1, but not γS2, is enriched in Arg, which is known for increasing stability in thermophilic proteins, due to the ability of the guanidinium group to form stabilizing salt bridges. Although increased arginine content should stabilize the protein with respect to thermal denaturation, as seen for hγS and γS1, chemical denaturation by urea should affect these salt bridges the same way as any other polar interaction, meaning that other factors such as hydrophobicity come into play in γS2.

Aggregation under thermal stress was also measured as a function of temperature using dynamic light scattering (DLS) for both toothfish γS-crystallins to ascertain their aggregation propensity. Aggregation propensity is not always directly correlated with thermal stability [47]. At even moderately high concentrations, protein aggregates can form well below the thermal denaturation temperature as a result of interactions between transiently exposed groups in conformationally mobile protein monomers. The measurements were made in 10 mM phosphate buffer with no additional reducing agents to avoid interfering with any attractive intermolecular forces that may be responsible for aggregation. For each data point, taken in triplicate, the sample was allowed to equilibrate for 2 minutes prior to measurement at a given temperature. Three scans of % abundance by number were averaged at each temperature and then fit to a Gaussian function using nonlinear regression. The average apparent particle size is plotted as a function of temperature in Fig. 1.7. γS1 remains monomeric until 35.0 °C where intermediate aggregates that range 30–100 nm begin to form before quickly transitioning into larger aggregates up to 1200 nm in size at 47.5 °C. γS2 follows a similar trend, but intermediate sized aggregates begin forming at 34.0 °C and large aggregates at 42.0 °C. In comparison, intermediate aggregates of hγS do not begin forming until around 49.0 °C while larger aggregates appear at 58.5 °C [47]. Both γS1 and γS2 are less thermally stable and more aggregation prone than hγS.

In summary, the biophysical characterization of γS1 and γS2 showed that while both proteins have the primarily β-sheet secondary structures characteristic of γ-crystallins, they appear to have slightly less overall β-sheet character than their human homolog. γS1 and γS2 also appear to have greater structural flexibility as observed in the red shifted tryptophan fluorescence spectra indicating that the structurally conserved tryptophans in the core of both proteins are more accessible to water. Of the two crystallins, the less structurally rigid γS2 has the lowest thermal stability as determined by thermal denaturation despite having a higher ΔG_w. Nevertheless, both toothfish crystallins have lower thermal stabilities than hγS and begin forming high molecular weight aggregates at lower temperatures. Many organisms only have one γS-crystallin while the toothfish eye lens contain two of these crystallins with varying thermal stabilities,, possibly because of the conflicting demands of this structural proteins two distinct roles in providing a high refractive index and maintaining the low-temperature solubility of crystallin mixtures in the lens. The biophysical characterization of γS1 and γS2 demonstrates an unusual set of protein homologs in which thermal stability does not directly correlate with ΔG°, and provides a useful model system for future structural and mutagenesis studies pinpointing the molecular determinants of protein solubility, thermal stability, and denaturation resistance.

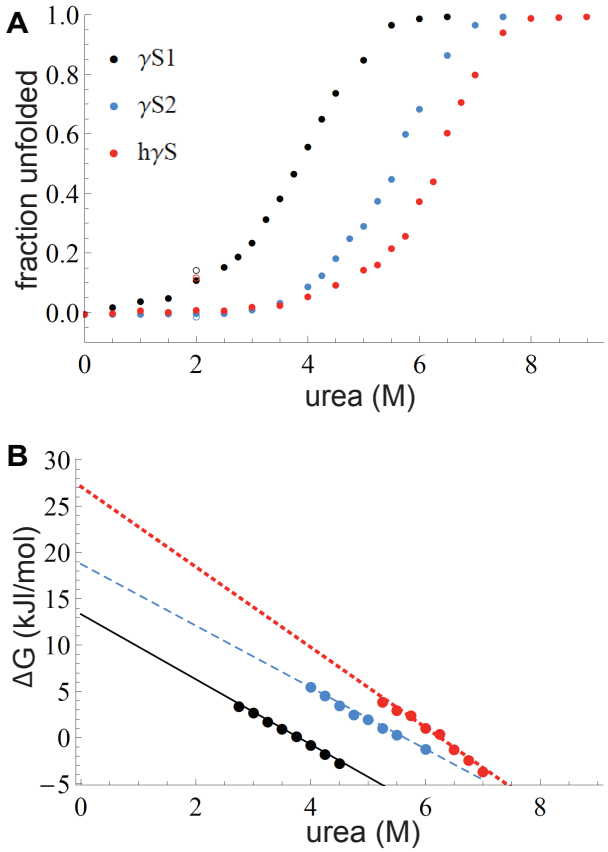

Figure 1.5: A. Chemical denaturation curves of hγS, *D. mawsoni* γS1 and γS2 with varying amounts of urea, measured by fluorescence spectroscopy and plotted as fraction unfolded. All three proteins exhibit two-state equilibrium unfolding behavior by urea denaturation. Data points designated with open circles represent samples of hγS, γS1 and γS2 that were first unfolded with 7 M urea and then diluted to 2 M urea to indicate that urea denaturation is reversible. B. ΔG vs denaturant concentration for the transition regions of the chemical unfolding curves used to extrapolate values for $\Delta G(H_2O)$. γS1 is more susceptible to unfolding by urea at lower concentrations where $\Delta G(H_2O)$ is 13.35 kJ/mol and 18.74 kJ/mol, for γS1 and γS2, respectively.

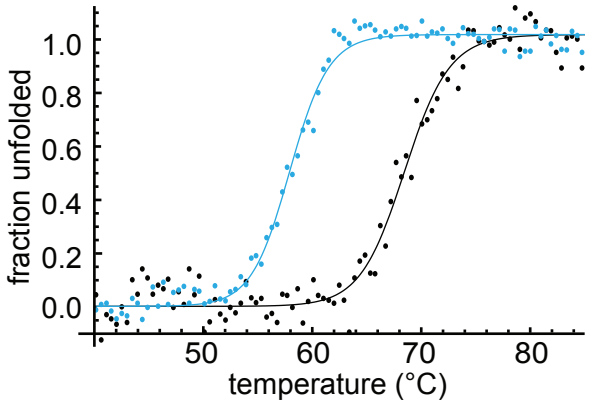

Figure 1.6: Thermal unfolding curves of γS1 and γS2 measured by monitoring the circular dichroism signal at 218 nm, with best-fit unfolding curves. T_m values for γS1 and γS2 are 68.5 °C and 58.0 °C, respectively. Both proteins exhibit two-state equilibrium unfolding behavior.

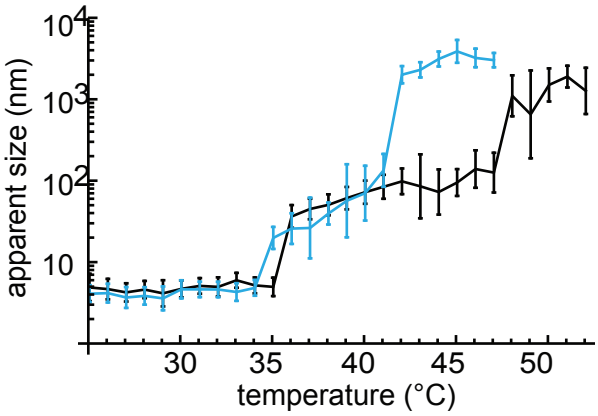

Figure 1.7: DLS measurements of thermally induced aggregation of γS1 and γS2. γS1 is monomeric until 35.0 °C where it begins to form intermediate aggregates and then finally forming large aggregates around 48.0 °C. γS2 behaves similarly but because it is not as thermally stable, it forms intermediate sized aggregates at 34.0 °C and large aggregates 42.0 °C.

Chapter 2

Structural determination of γS-crystallins from an Antarctic toothfish using solution-state NMR

2.1 Background

The Antarctic toothfish *Dissostichus mawsoni* lives in the frigid waters of the Southern Ocean and is adapted to handle freezing cold temperatures. A prominent way this fish, among other Antarctic species, has adapted is through the presence of antifreeze glycoproteins which protect it from ice crystal formation. Another noteworthy adaptation to such a cold environment is the ability to completely resist the phenomenon known as cold cataract, a phase separation of the lens proteins, whereas the eye lens of many other animals are susceptible. One feature common to fish eye lenses is the abundance of γ-crystallins as compared to mammals where the small heat shock proteins known as α-crystallins are the predominant protein species [24, 30]. The higher prevalence of γ-crystallins may be a way

of modulating the protein interactions within the eye lens to accommodate the need for higher protein concentrations in aquatic lenses. This higher protein concentration is a way to increase the refractivity of the lens due to the lack of an air cornea interface providing additional refractive power [60]. Studying the γ-crystallins from *Dissostichus mawsoni* could provide valuable insight in comparison to other fish and mammalian crystallins, to explore adapted features specific to aquatic organisms versus features related to cold adaptation and the ability resist cold cataract. Such research could also provide valuable insight into features of protein-protein interaction and solubility in general.

One technique particularly well suited to tackling these questions is solution-state nuclear magnetic resonance (NMR). Due the ability keep the protein in solution for structural determination the sample better reflects the functional adaption of these proteins to stay soluble. Another advantage to NMR, besides structural determination, is the ability to access information regarding protein dynamics and hydration [61, 62]. To compare the toothfish crystallins to other fish and mammalian crystallins, γS1- and γS2-crystallin have been subjected to NMR experiments for structural and biophysical determination. The results of this information will provide clues to functional difference between cold adapted and aquatic crystallins in regard to how they affect protein interactions and solubility.

2.2 Materials and Methods

2.2.1 Gene construction, expression, and purification

Plasmids containing the cDNA sequences for *Dissotichus mawsoni* γS1 and γS2 were purchased from Blue Heron (Bothell, WA). Each gene was flanked by regions containing restriction sites for NcoI and XhoI. The gene was amplified using oligonucleotide primers purchased from Intergrated DNA Technologies (Coralville, IA), and the resulting gene products were

cloned into pET28a(+) vector (Novagen, Darmstadt, Germany). Isotopically ^{15}N labeled toothfish crystallins were overexpressed in Rosetta (DE3) *Escherichia coli* using an optimized high-cell-density IPTG-induction minimal medium protocol [63] where cells were grown for 24 hours at 25 °C. Cells were lysed by sonication and cell debris was removed by centrifugation. Untagged γS1 and γS2 were dialyzed in 10 mM Tris, 0.05% sodium azide, pH 8 then purified by anion exchange on an UNOsphere Q column purchased from Bio-Rad (Hercules, CA) using a 1 M sodium chloride gradient. The final purification step for both crystallins was application to a HiLoad 16/600 Superdex 75 PG gel filtration column from GE (Pittsburgh, PA) using 10 mM sodium phosphate buffer, 100 mM sodium chloride, 0.05% sodium azide at pH 6.9. Proteins were dialyzed into 10 mM phosphate, 0.05% NaN_3, pH 6.9, concentrated and adjusted to 10 % D2O and 2 mM TMSP at a final concentration of approximately 1 mM.

2.2.2 NMR experiments

Experiments were performed on a Varian Unity INOVA spectrometer (Agilent Technologies, Santa Clara, CA) operating at 800 MHz and equipped with a ^{1}H-^{13}C-^{15}N 5 mm tri-axis PFG triple-resonance probe, using an 18.8 Tesla superconducting electromagnet (Oxford Instruments). Decoupling of ^{15}N nuclei was performed using the GARP sequence [64]. ^{1}H chemical shifts were referenced to TMSP, and ^{15}N shifts were referenced indirectly to TMSP. NMR data were processed using NMRPipe [65] and analyzed using CcpNMR Analysis [66].

2.3 Results and Discussion

Solving the structure of γS1 and γS2 would allow for the comparison to both other fish and mammalian homologs that could give insight into features that are important for being

adapted to a fish eye lens or in the case of *D. mawsoni*, cold temperatures. Using solution state NMR is an excellent method due to the functional adaptation of crystallins to be highly soluble. To assess how suitable these proteins are for solution state NMR, ^{15}N isotopically labeled samples were prepared. As seen in both Fig. 2.1 and Fig. 2.2 both ^{1}H-^{15}N heteronuclear single-quantum correlation (HSQC) experiments display well-dispersed peaks that represent the N-H pairs in the protein, which in the case of proteins are mostly the backbone amides. Dispersion of the peaks indicates that the protein is well folded under the conditions in which the sample was prepared. Additionally, this allows for determination of the signal to noise which is important for considering further 3D experiments which are needed to assign the individual resonances.

In addition to being useful for evaluating the chemical shift dispersion, signal sensitivity, and being used as a reference for resonance assignments, HSQC experiments can provide additional information about the protein of interest. One such experiment involves taking HSQC spectra as a function of temperature. As the temperature changes so does the chemical shift of some of the N-H pairs in the protein even in absence of large structural changes. Often this is due to the hydrogen bonding which the protein is participating in. For example an amide that is hydrogen bonded with water in the solvent will experience a change in chemical shift as the thermal fluctuations of the water increases. However, an amide involved in an intramolecular hydrogen bond within the protein will not experience as large of a change to its chemical shift due to the large size of the protein leading to less intense thermal fluctuations of the protein relative to small water molecules [62]. As seen in Fig. 2.3, γS2 shows both small and large changes in the chemical shifts of its N-H pairs. Once the resonances are fully assigned this information allows the inference of which parts of the protein are buried as opposed to exposed to solvent. It is noteworthy that the usable spectra are able to be acquired up to 55 °C, which is remarkable considering that this is only about 5 °C below the melting temperature of the protein and would be expected to show more perturbation due to unfolding [67]. By comparing this data between the *D. mawsoni* crystallins and their fish

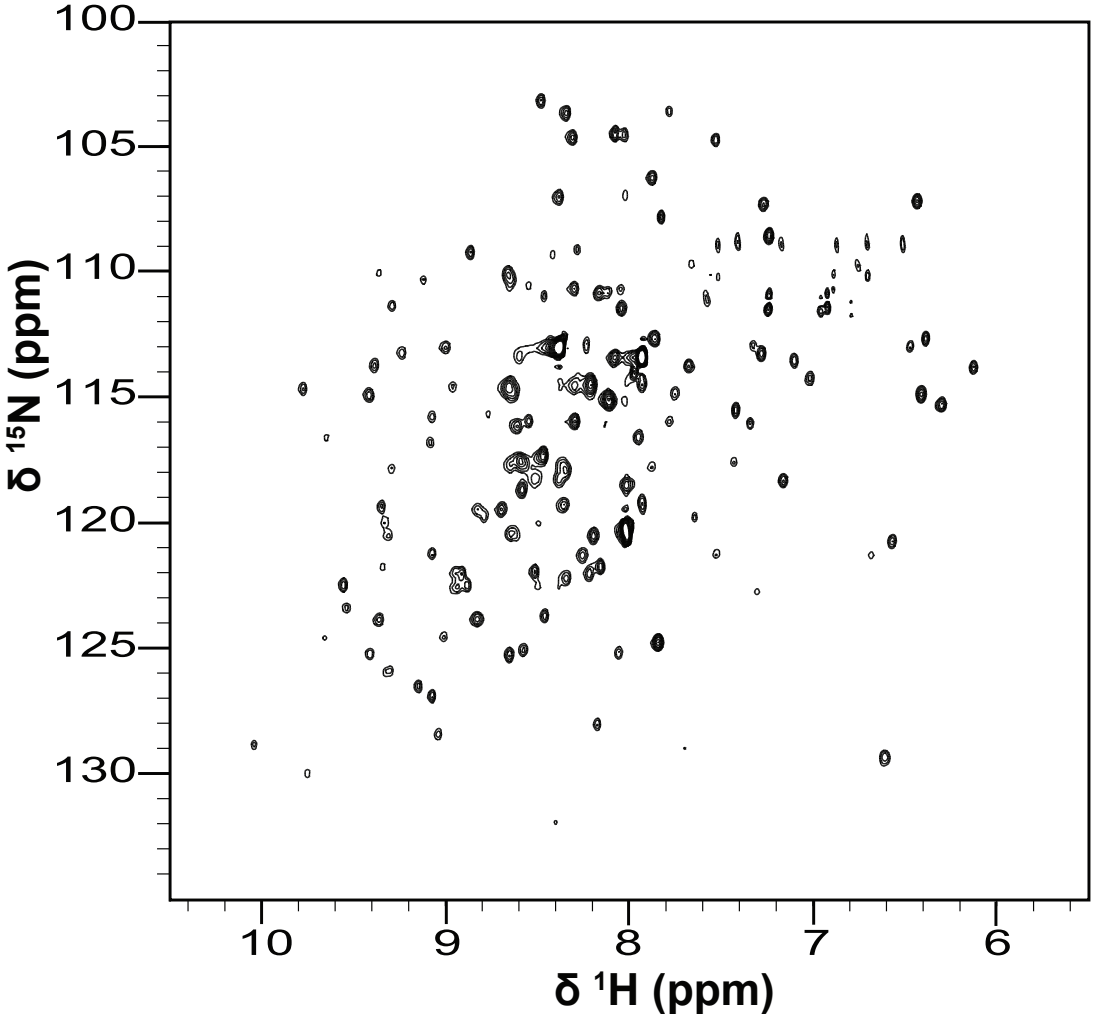

Figure 2.1: ^{1}H-^{15}N HSQC spectrum of γS1-crystallin at 25 °C. The presence of well dispersed peaks indicates that the protein is well folded and does not participate in higher order oligomerization.

and mammalian counterparts a better model of how the hydration of these crystallins differ and how that relates to their function of remaining soluble can be formed.

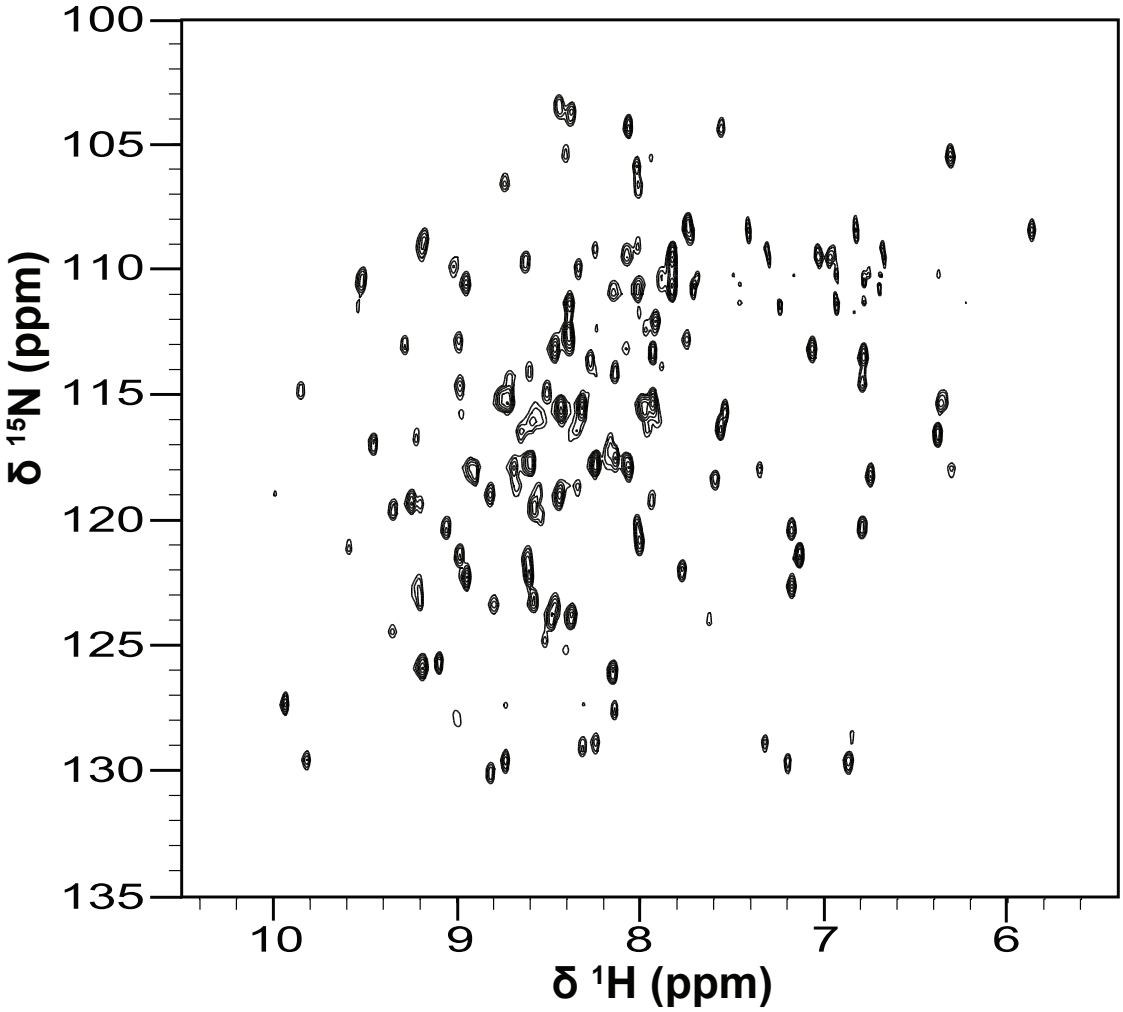

Figure 2.2: ^{1}H-^{15}N HSQC spectrum of γS2-crystallin at 25 °C. The presence of well dispersed peaks indicates that the protein is well folded and does not participate in higher order oligomerization.

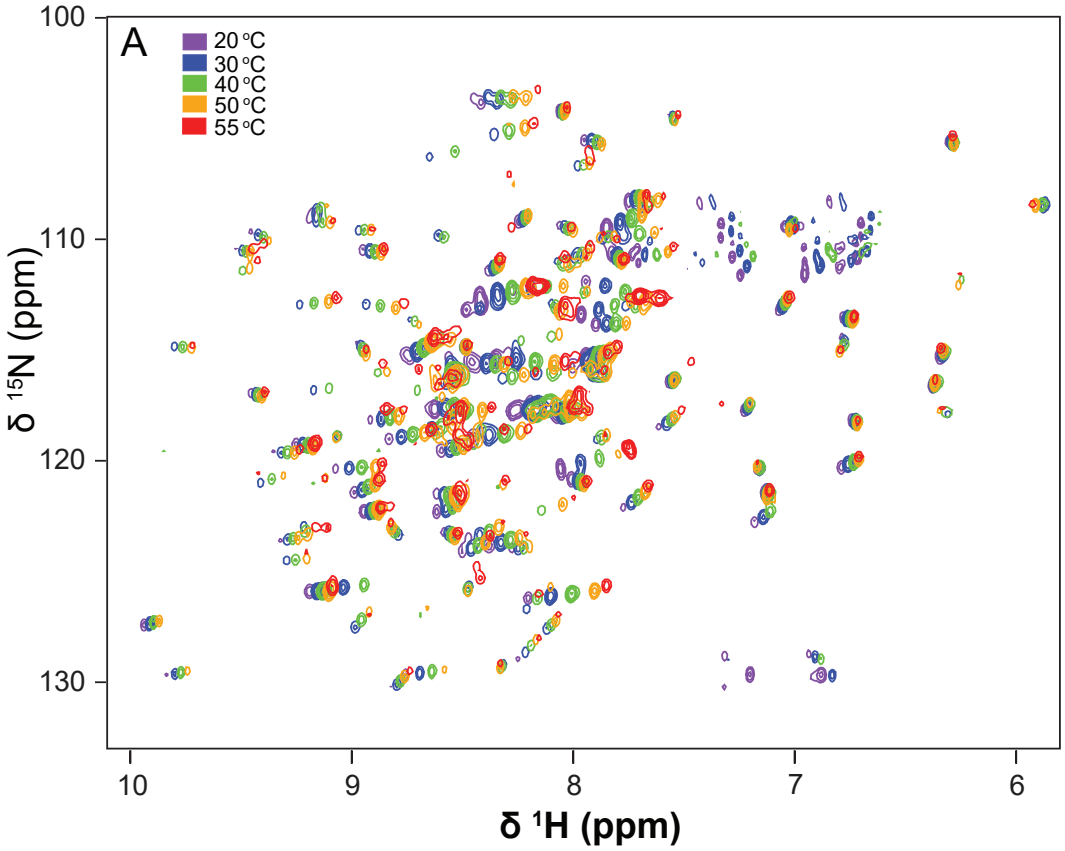

Figure 2.3: ^{1}H-^{15}N HSQC spectrum of γS2-crystallin at variable temperature indicates which residues are likely involved in solvent vs intramolecular hydrogen bonding as observed by the large and small changes in chemical shift respectively.

Chapter 3

Controlling liquid-liquid phase separation of cold-adapted crystallin proteins from the Antarctic toothfish

3.1 Background

The coexistence of multiple liquid phases is critical for formation of membraneless organelles including stress granules [68], nuage [69], and nucleoli [70]. These concentrated protein droplets function to localize particular cellular components [71], increase the local concentrations of key reactants [72], and sequester reactive or aggregation-prone species from the rest of the cell [73]. LLPS is frequently observed in proteins with low-complexity sequences, particularly those containing relatively high proportions of aromatic and charged residues. Some proteins with intrinsically disordered regions (IDRs) form protein-rich droplets as a precursor to aggregation, as in the case of the tau protein found in neurofibrillary tangles [74], and the proteins TDP-43 [75], FUS [76], and hnRNPA1 [71], all of which aggregate in

amyotrophic lateral sclerosis (ALS). In some cases, sequence elements such as the recently discovered low-complexity aromatic-rich kinked segments (LARKS), reversibly assemble into easily-disrupted protofilaments, allowing dynamic response to stimuli [77]. Much of the recent literature on LLPS has focused on IDRs; however, there are also examples where interactions between compactly folded domains are critical for producing protein-rich droplets, as in the dimerization of human Edc3 during P-body formation [78].

Another such case is that of "cold cataract," a reversible opacification of vertebrate eye lenses upon cooling, caused by LLPS of the structural crystallin proteins [3]. LLPS is also observed in the lenses of cataract-prone mice as a precursor to protein precipitation [79]. For several isolated lens proteins, LLPS occurs under conditions where the protein is metastable with respect to crystallization, highlighting its potential importance to understanding protein solubility more broadly [80, 81]. Many mutations associated with hereditary cataract do not result in large-scale unfolding or the creation of IDRs, but instead produce only subtle variation in structure, coupled with altered protein-protein interactions. In the human lens, this has been observed for the R76S [82] and P23T [83, 84, 85] variants of γD-crystallin and the G18V [34] variant of human γS-crystallin. Similarly, cold cataract occurs despite the compactly-folded, globular structures of the lens crystallins, which typically lack IDRs.

Cold cataract, with its reversible formation of protein-rich droplets, is distinct from cold denaturation, which is driven by temperature-dependent interactions between water and the normally buried hydrophobic residues in the protein core [86]. Although assessing the full protein stability curve requires the observation of cold denaturation, for most globular proteins the temperature at which this occurs is below the freezing point of water and thus not readily accessible in the laboratory [87]. One exception is yeast frataxin, which has been extensively studied at and near its cold denaturation temperature, yielding valuable insights into the residual secondary structure [88] and unfolding mechanism [89]. For most other proteins, accessing the cold denatured state in the laboratory requires high pressure [90],

the addition of denaturants [91], or specialized experimental methods such as encapsulation in reverse micelles [92, 93]. Cold cataract has been studied for decades, with initial work focused on determining the scattering elements of intact lenses and characterizing the phase separation behavior of individual crystallins [94, 95]. More recent studies have progressed toward better understanding of crystallin mixtures [53, 96, 97]. In all cases reported so far, the opacity observed at low temperature is reversible and is not consistent with cold denaturation.

Here we investigate cold cataract via the cold-adapted eye lens crystallins of the Antarctic toothfish (*Dissostichus mawsoni*). This large, long-lived fish is adapted to the cold waters of the Southern Ocean, where the year-round temperature is -1.9°C, near or at the freezing point of seawater. As in many other vertebrate eye lenses, the major protein components of the *D. mawsoni* lens belong to two conserved superfamilies, the chaperone-like α-crystallins and the structural $\beta\gamma$-crystallins [6]. In addition to the γS- and γN-crystallins shared with mammals, the eye lenses of fish contain a unique class of γ-crystallins known as γM-crystallins, named for their elevated content of the highly refractive amino acid methionine. These γM-crystallins are the most abundant proteins in fish lenses, comprising over 50% by mass [98]. In contrast to the eyes of terrestrial animals, where incident light refraction mostly occurs at the air/cornea interface, in aquatic animals the combined refraction and focusing is accomplished by the lens alone. Therefore, fish lens crystallins are present in concentrations up to 1 g/mL, and are strongly selected for high refractive index increments [99]. Despite this extremely high protein density, the *D. mawsoni* eye lens is completely transparent at -1.9°C and retains clarity down to at least -12°C [24]. The eye lens crystallins of *D. mawsoni* are subject to competing selection pressures for cold adaptation and high refractive index increment, leading to unusual biophysical properties.

Despite the complete resistance of Antarctic toothfish lenses to cold cataract, not all of its component proteins are equally cold-tolerant. Here we probe the cold tolerance of the indi-

vidual γM-crystallins from cold-vulnerable sub-fractions of the *D. mawsoni* eye lens. Our results show that the critical temperatures (T_c; the temperature at which a protein solution undergoes LLPS, where both phases contain equal proportions of protein) of all the recombinant crystallins are below -1.9°C, however substantial variation in T_c is observed. We tested several alternative hypotheses about which properties of the γM-crystallins underly their differential resistance to cold cataract: 1. Selection for cold tolerance leads to reduced thermal or chemical stability; 2. Resistance to cold cataract is negatively correlated with affinity to bovine α-crystallin in a cross-species chaperone assay; 3. Cold-cataract resistance is determined by surface properties mediated by non-conserved residues; and 4. The key property is the number and strength of the potential intermolecular interactions formed. These hypotheses were tested using a variety of biophysical techniques as well as site-directed mutagenesis. Our results indicate that resistance to cold cataract does not correlate with thermal or chemical stability, chaperone affinity, or generalized surface properties, but instead depends, in part, on the type of positively charged residues exposed at the surface, even in the absence of a change in overall surface charge. Substitution of arginine residues to lysine at surface-exposed positions allowed tuning of the T_c, enabling us to turn γM8b, one of the most cold-susceptible of the toothfish γM-crystallins, into most cold-resistant of the proteins tested. Conversely, mutation of three lysine residues to arginine (K18, K85, K172) increased the T_c of γM8b by an equivalent amount. The altered cold resistance produced by a small number of targeted mutations provides a unique opportunity to investigate the evolution of protein solubility and stability within highly crowded, protein-dense environments. Furthermore, it suggests mechanisms for using protein structural and surface properties to control liquid-liquid phase separations, which are used in biological systems to produce subcellular compartments of locally elevated protein concentration, e.g. in membraneless organelles [100, 101].

3.2 Materials and Methods

3.2.1 Lens collection and fractionation

Lenses of the giant Antarctic toothfish (*Dissostichus mawsoni*) were collected from live specimens caught in McMurdo Sound, Antarctica and processed as described previously [24]. Size exclusion chromatography (SEC) using Sephacryl 200 High Resolution (S200HR) resin (Sigma, St Louis, MO) was performed to fractionate the soluble lens protein components of *D. mawsoni*. Fresh or frozen lenses were first solubilised by stirring overnight at 8°C in three volumes of solubilisation buffer (10 mM potassium phosphate, 100 mM KCl, 0.05% NaN_3, pH 7.6). Solubilised lens homogenate was centrifuged at 20,198g for 20 minutes set at 4°C to sediment the insoluble components. Between 0.5 g to 1.0 g of soluble lens proteins (supernatant) were loaded (maximum volume 5 mL) on a S200HR column (2.5 cm ID x 116 cm L) equilibrated with the solubilisation buffer, and protein elution was carried out with the same buffer at room temperature with a flow rate of 36 mL per hour. The heterogeneous toothfish γ-crystallin fraction was pooled and partially resolved into its component crystallins by low-pressure cation exchange chromatography. Low-pressure ion-exchange chromatography was performed at 8°C using a strong cation SP-Sepharose FF resin (Amersham Biosciences, Piscataway, NJ) in a (40 cm L X 1.5 cm ID) column. Toothfish γ-crystallins were dialysed from the SEC buffer (10 mM Potassium phosphate, 100 mM KCl, 0.05% NaN_3, pH 7.6) in SpectraPor4 MWCO 12,000 14,000 membrane tubing (Spectrum Medical Supplies, Inc., Los Angeles, CA) against the starting buffer (25 mM Sodium acetate, 0.05% Sodium azide, pH 5.0). After dialysis, the protein solution was centrifuged at 20,198g at 4°C to sediment any aggregated protein. The sample was further filtered through a 0.2 micrometer polycarbonate Nucleopore filter (Nucleopore Corp., Pleasanton, CA, USA). The A280 was read prior to loading between 100 to 150 mg (60 to 70 mLs) onto the column. Elution flow rate was maintained using a Gilson Minipuls2 (Gilson Inc., Middleton, WI)

generating a constant flow rate of 36 mL per hour. One column volume of starting buffer was applied to wash unbound protein before elution of the toothfish γ-crystallins with a linear salt gradient of 0.0 to 0.2 M NaCl followed by a 2M NaCl wash.

3.2.2 Gene construction, expression, and purification

Plasmids containing the cDNA sequences of the *D. mawsoni* γM1- (GenBank, DQ143973), γM4- (GenBank, DQ143976), γM7- (GenBank, DQ143979), γM8b- (GenBank, DQ143981), γM8c- (GenBank DQ143982), and γM8d-crystallin (GenBank, DQ143983) genes were purchased from Blue Heron Biotech, LLC. (Bothell, WA). Each gene was flanked by regions containing restriction sites for NcoI and XhoI. γM4, γM8b, γM8c, and γM8d contained an N-terminal 6x His tag, and a TEV cleavage sequence (ENLYFQG) with the N-terminal methionine replaced by the final glycine in the cleavage sequence. γM1 and γM7 lack the N-terminal extensions found in the other toothfish γM-crystallins, making TEV protease cleavage of the His tag ineffective for these proteins. Therefore, these proteins were cloned without an N-terminal His tag and purified using ion exchange chromatography as described below. The toothfish crystallin genes were amplified using oligonucleotide primers purchased from Intergrated DNA Technologies (San Diego, CA), and the resulting gene products were individually cloned into pET28a(+) vectors (Novagen, Darmstadt, Germany). Variants of γM8b were created using standard site directed mutageneis PCR. γM1, γM4, γM7, γM8b, γM8c, and γM8d were overexpressed in Rosetta (DE3) *Escherichia coli* using standard autoinduction protocols [102] at 25°C. Cells were allowed to grow for 24 hours, lysed via sonication, and cell debris was removed by centrifugation. His-TEV-γM4, His-TEV-γM8b, His-TEV-γM8c, and His-TEV-γM8d were purified on a Ni-charged Bio-Scale Mini Profinity IMAC Cartridges (Bio-Rad, Hercules, CA) and cut with the use of a His-tagged TEV protease (produced in-house). The TEV protease and His tag were removed by a second application to a Ni-charged column, and the pure product was collected in the column

flow-through. γM1 and γM7 was purified by dialysis in 10 mM MES, 10 mM succinate, 0.05% NaN_3, pH 6.0. The dialyzed lysate was applied to a UNOsphere^{TM} S cation exchange column (Bio-Rad, Hercules, CA) and eluted using a 1 M NaCl gradient. The resulting eluate fractions containing the recombinant protein were pooled. The final purification step for all proteins consisted of applying the sample to a HiLoad 16/600 Superdex 75 pg gel filtration column (GE, Pittsburgh, PA) in 10 mM phosphate buffer. Protein purity was analyzed by SDS-PAGE and electrospray ionization mass spectrometry as seen in Fig. A.15 and A.16. All proteins were dialyzed into 10 mM phosphate, 0.05% NaN_3, pH 6.9, for all experiments unless otherwise stated.

3.2.3 Isoelectric focusing

Fractions from cation exchange chromatography were desalted by dialysis and lyophilised until dry and then resuspended in sample buffer (7 M urea, 2 M thio-urea, 4% CHAPS). Between 10 to 40 micrograms of protein was loaded onto 7 cm, non-linear pH 3-10 IPG strips (Amersham Biosciences) in a re-hydration buffer consisting of 9M urea, 2% CHAPS, 2.5 mg/mL DTT and IPG buffer (Amersham Biosciences). The strips were allowed to re-hydrate for 12 hours followed by isoelectric focusing for 28,000 V hours.

3.2.4 Mass spectrometery

Mass spectrometry analysis was performed using standard addition regression, as previously described for quantitation of proteins [103] and peptides [104]. Known aliquots of trypsin-digested recombinant proteins were added to the digested lens sample (fraction III) in order to estimate amounts of the proteins expected to be present (γM1, γM7, γM8b, γM8c, and γM8d). Samples were run on a Xevo G2-XS QTof spectrometer (Waters, Milford, MA) in MS^E mode, with an in-line ACUITY UPLC BEH C18 Column (Waters, Milford, MA), where

peptides were eluted with a water, acetonitrile gradient. Peptide precursor and product ions in the mass spectrum were identified using BiopharmaLynx (Waters, Milford, MA). Each protein were considered to be present if the regression of detected peptide intensities yielded a $p \leq 0.05$, at least 7 beta and gamma fragments per sample, and a regression intercept greater than the detection cutoff (2000 counts). Fragments matching other, unexpected, γ-crystallin sequences were considered to be above the detection limit if at least 5 beta and gamma fragments were observed in each run, and each ion detection count from three replicates was greater than 2000 counts.

The unique sequence coverage of each protein was calculated. The total number of unique amino acids of a protein was determined by cross-referencing the theoretical trypsin digests of all toothfish α-, β-, and γ-crystallins to remove any peptides with a duplicated sequence or mass.

3.2.5 Circular dichroism

Purified γM1, γM4, γM7, γM8b, γM8c and γM8d were diluted to 0.125 mg/mL with 10 mM phosphate buffer at pH 6.9 for the collection of full circular dichroism (CD) spectra, at room temperature, or at 0.25 mg/mL in 10 mM phosphate buffer, 150 mM NaCl, and 1 mM DTT at pH 6.9 for unfolding experiments. Measurements were taken on a J-810 spectropolarimeter (JASCO, Easton, MD) equipped with a thermal controller. During unfolding measurements, the CD signal at 218 nm monitored while the samples were heated at a rate of 2 °C/min. The data were fit to a two-state equilibrium unfolding model to determine the thermal denaturation temperature (T_m). All measurements were taken in triplicate.

3.2.6 Fluorescence spectroscopy

UV fluorescence measurements were made on γM1, γM4, γM7, γM8b, γM8c and γM8d at a concentration of 0.1 mg/mL in 10 mM phosphate buffer, pH 6.9, at room temperature for full emission spectra, exciting at 280 nm. Samples for chemical unfolding curves were prepared with increasing concentrations of 10 M urea (Fisher Scientific, Waltham, MA). Urea stock solutions were prepared as outlined by [31]. Samples were allowed to equilibrate at room temperature for at least 24 h before emission was measured in triplicate using a Cary Eclipse Fluorescence Spectrophotometer (Agilent, Santa Clara, CA). The ratio of baseline-corrected emission intensities at 360 and 320 nm was used for analysis. To determine the thermodynamic parameters (ΔG_w° and m values), $\Delta G_{[urea]}$ was calculated from the normalized equilibrium unfolding data and a linear least-squares fit was performed in Mathematica where ΔG_w° is the value of ΔG at 25 °C, extrapolated to zero concentration of denaturant, and the slope m is a measure of the dependence of ΔG on denaturant concentration. All measurements were taken in triplicate.

3.2.7 Dynamic Light Scattering

Dynamic light scattering (DLS) measurements for aggregation propensity were obtained with a Zetasizer Nano ZS (Malvern Instruments, Malvern, U.K.) on γM1, γM4, γM7, γM8b, γM8c and γM8d at a concentration of 1.0 mg/mL in 10 mM phosphate buffer, pH 6.9. At each temperature, the sample was allowed to equilibrate for 2 min before measurements were obtained, after which scattering measurements were performed in triplicate. DLS measurements of phase separation were performed at protein concentrations of 150 mg/mL in 10 mM phosphate buffer pH 6.9. The measurements were taken at decreasing temperatures of 1 °C increments with 2 min equilibration times. The input viscosity for DLS processing was obtained from the linear fit of viscosity measurements. All measurements were performed in

triplicate. Apparent size as a function of temperature was plotted using the average diameter of particles normalized according to particle volume.

3.2.8 Viscosity

Viscosity measurements were taken using the AR-G2 magnetic bearing rheometer (TA Instruments, New Castle, DE). The rheometer was fit with a 40 mm, 2 degree stainless steel cone (TA Instruments, New Castle, DE). Protein samples were at a concentration of 150 mg/mL in 10 mM phosphate buffer pH 6.9. Viscosity measurement were taken as a function of temperature with a ramp rate of -0.2 °C/min and a 30 second averaging time for each point. Duplicate measurements were collected, and the data were fit to the Ahrrenius equation [105].

3.2.9 Coexistence curves

Coexistence curves were measured for γM1, γM4, γM7, γM8b, γM8c and γM8d at varying concentrations. Samples were concentrated in an 3,000 NMWL Amicron ultra-15 centrifugal filter unit (EMD Millipore, Darmstadt, Germany). Sample concentrations were measured in triplicate by measuring the absorbance at 280 nm using a Nanodrop 2000 spectrophotometer (ThermoFisher Scientific, Waltham, MA), which has an upper measurement limit of 300 at 1 cm equivalent absorbance units. For the highest concentration solutions a 10 fold dilution was made for measurements to ensure an absorbance below 200 at 1 cm equivalent. Due to increasing viscosity with protein concentration, higher concentration solution absorbance measurements are more error prone. Replicate measurements used did not deviate from each other by more than 10 %. For LLPS measurements the absorbance at 600 nm was measured as a function of temperature at a rate of -0.2 °C/min, resulting in a sigmoidal plot. The halfway point was determined as the cloud point. The data were fit to a binodal function

to determine the critical temperature and concentration [95]. After each measurement the same sample was warmed and mixed, then the process was repeated for a total of the three measurements.

3.2.10 Cross-species chaperone assays

Assays of chaperone protection against heat aggregation were performed in the manner of previously published methods [5, 106, 24]. Bovine α-crystallin used for the cross-species functional assays was separated from β- and γ-crystallins on S200HR resin in the same manner as the toothfish lens sample and taken from the leading edge of the α-crystallin peak as previously described [24]. Crystallin composition of the peaks for the cow has been previously verified by immunoblotting and consists of αA and αB at a a ratio of approximately 3 to 1 [24]. Briefly, recombinant toothfish γ-crystallins (final concentration of 1 mg/mL) were incubated at a temperature of 60 °C for one hour with A360 measurements sampled every five minutes with or without bovine α-crystallin (final concentration of 1 mg/mL) in a total assay volume of 800 μL. All measurements were taken in triplicate.

3.2.11 Comparative models

Comparative models were generated using the Robetta server (http://robetta.bakerlab.org/) [107].

3.3 Results

3.3.1 The toothfish eye lens has a cold-susceptible fraction that undergoes LLPS

The complete resistance of the intact Antarctic toothfish eye lens to cold cataract raises the question of whether the individual proteins and subsets thereof also have the same degree of cold tolerance. The toothfish has at least fourteen γ-crystallin paralogs, accounting for >40% of the proteins in the lens [24, 108]. Fish lenses often contain numerous γ-crystallin paralogs; the abundance of these small proteins is hypothesized to contribute to the dense packing required to form a functional aquatic lens [18]. In this case, the large number of γM paralogs may also play a role in maintaining solubility under the cold conditions of the Antarctic [30]. A phylogram illustrating the amino acid sequence similarity among *D. mawsoni* γ-crystallins and those of selected reference organisms is shown in Fig. 3.1; the corresponding multiple sequence alignment is shown in Supplementary Fig. A.1. The sequence identity values for the six γM-crystallins in this study range from 52.1% for the least similar pair (γM7 and γM8d) to 92.3% for the most similar (γM8c and γM8d) [30]. With the exception of γM7, all of the toothfish γM-crystallins cluster together, in a separate group from the γS- and γN-crystallins from *D. mawsoni* and other species. These crystallins were partially purified from the homogenized lens using size-exclusion and cation-exchange chromatography, yielding eleven fractions enriched in particular γ-crystallin paralogs.

One fraction (FIII) was observed to undergo a LLPS at 4 °C, 6 °C higher than the physiological norm, in contrast to the behavior of the whole lens, which remains transparent at least down to -12 °C (Fig. A.2). Isoelectric focusing (IEF) under denaturing conditions reveals that FIII contains predominantly proteins with isoelectric points (pI) above 6.5 (Fig. A.3). It is potentially significant that the proteins in the LLPS-prone fraction of the toothfish

lens have isoelectric points near neutral pH, as higher net surface charge would be likely to lead to increased intermolecular repulsion in solutions with one protein component. Further mass spectrometry (MS) was then performed on the cold-susceptible fraction (Fig. A.4, Table A.1). Unique peptides from γM1, γM4, γM7, γM8b, and γM8d were detected in the MS experiments, and their expected pI values are consistent with the IEF data (Table A.1). Quantification for γM1, γM7, γM8b, and γM8d revealed that there was only an approximately 4-fold difference in the most abundant γM8d and least abundant γM7. Therefore, these five γ-crystallins were chosen for recombinant expression and further characterization, as was γM8c due to its identical isoelectric point and high sequence identity to γM8d.

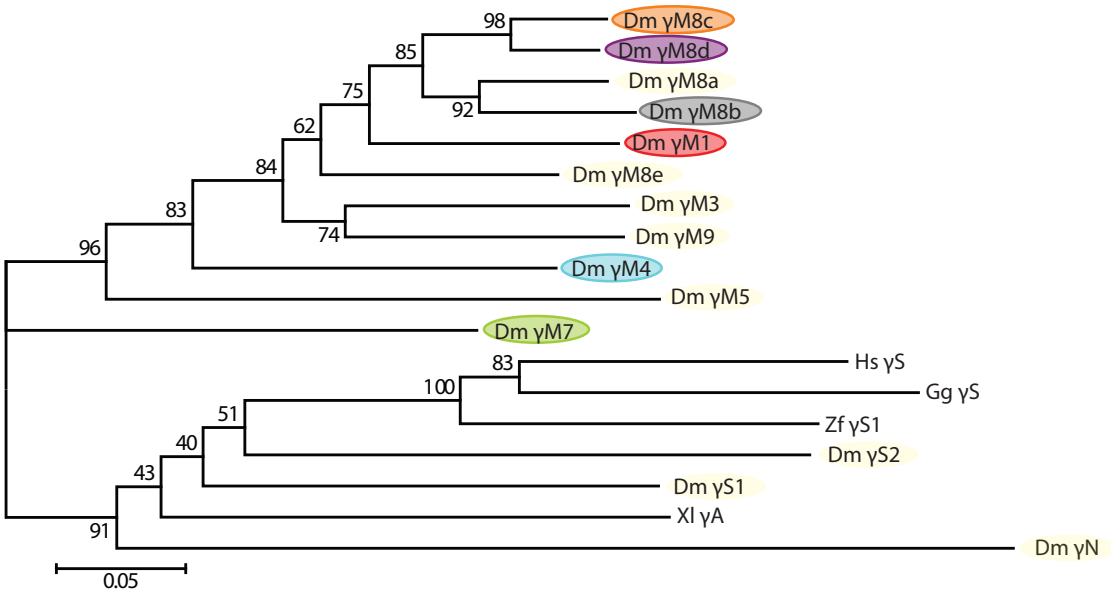

Figure 3.1: The neighbor-joining phylogram was drawn using the MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for the amino acid sequences of the *D. mawsoni* (Dm) γ lens crystallins as well as selected sequences from *Homo sapiens* (Hs), *Gallus gallus* (Gg), *Danio rerio* (Zf), *Xenopus laevis* (Xl). The phylogram was drawn based on the alignment shown in Fig. A.1. Proteins identified as belonging to the LLPS-forming lens fraction and chosen for experimental characterization are highlighted.

3.3.2 The unusual amino acid compositions of the toothfish crystallins hint at their functional adaptations

Having determined that a subset of the toothfish lens proteins does undergo cold cataract, we set out to explore the relationship between sequence adaptations that contribute to the refractive power of the lens and those that are primarily associated with cold tolerance. Both the sequence relationships described in the previous section and the amino acid compositions are necessary to put in context the phase behavior of the individual proteins. Lens proteins are generally enriched in residues with a high refractive index increment (dn/dc) and exiguous in aliphatic residues that do not enhance refractive power [10, 11]. The Antarctic toothfish crystallins are no exception, with very low aliphatic indices and much higher than average proportions of residues with large, polarizable side chains. Their amino acid compositions are tabulated in Fig. 3.2 (see also Table A.2), which is sorted from highest to lowest average amino acid abundance in the γMs relative to the average non-membrane protein from non-mammalian vertebrates [109]. Strikingly, the *D. mawsoni* γM-crystallins have remarkably low counts for some of the most common amino acids: for example, the toothfish γM1-, γM4- γM8b-, γM8c-, and γM8d-crystallins contain no alanine. Conversely, all six proteins are much richer in methionine, normally one of the least abundant amino acids.

3.3.3 Toothfish γM-crystallins are variable in their resistance to cold cataract

To determine the onset temperature of LLPS for individual *D. mawsoni* γM-crystallins, we measured the coexistence curves representing the temperature and concentration dependence of LLPS formation for samples of each recombinantly expressed protein (Fig. 3.3). The measurements were performed by slowly lowering the temperature at a given protein concentration until a reversible phase transition was observed, known as the cloud point

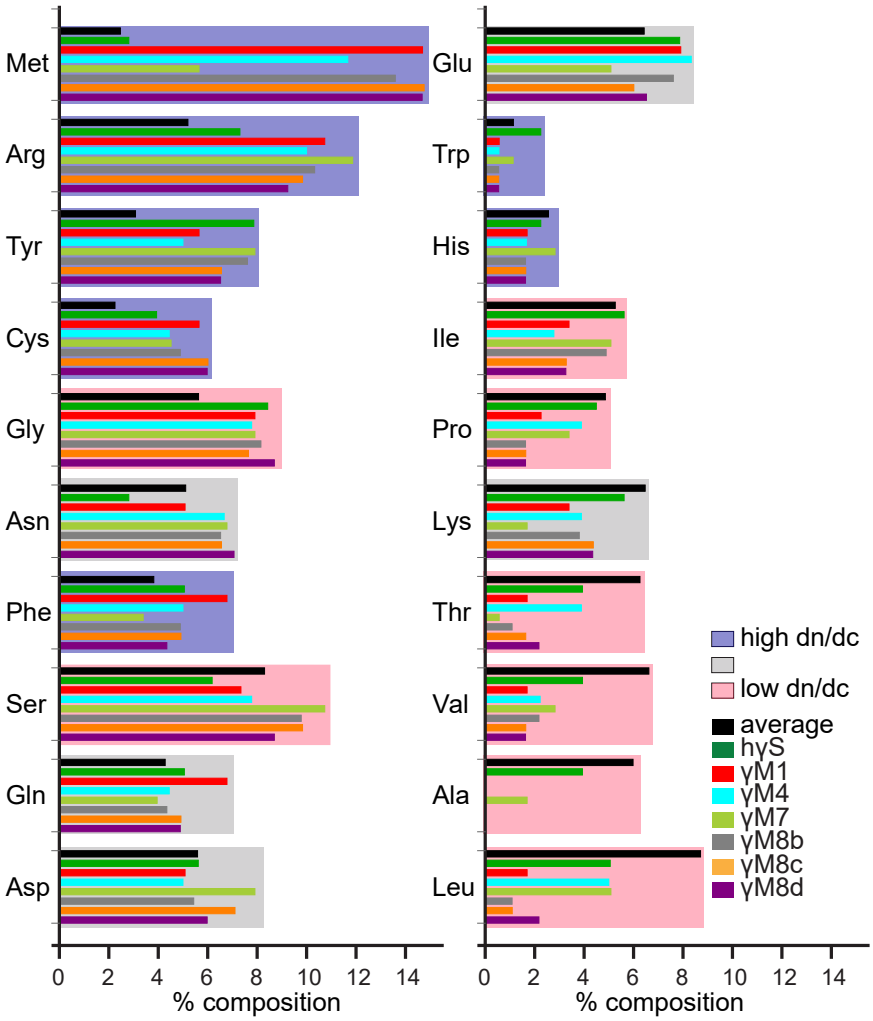

Figure 3.2: Bar chart of amino acid composition of an average non-membrane, non-mammalian, vertebrate protein [109] (black), hγS (dark green), γM1 (red), γM4 (cyan), γM7 (lime green), γM8b (gray), γM8c (orange), and γM8d (purple). Light blue, gray and red shaded boxes represent dn/dc values above 0.2, between 0.2 and 0.18, and below 0.18 mL/mg, respectively. Most crystallins are exiguous in low-dn/dc aliphatic residues. Highly refractive amino acids such as methionine, arginine, and tyrosine are substantially enriched in the γM-crystallins.

[95]. The maximum of the curve is the critical point, defined by the concentration and temperature at which the ratio of protein in each liquid phase is equivalent. The T_c range from -12.3 to -4.4 °C for the six proteins and the critical concentrations between 190 and 300 mg/mL, consistent with previous measurements for γ-crystallins [95, 53]. The proteins undergo reversible LLPS, which settles into two phases with time, rather than precipitation (Fig. A.5). Supplementary Figure A.6 shows the ratio of the intrinsic Trp fluorescence signals at 360/320 nm. This fluorescence is highly sensitive to the local environment of Trp residues, and hence serves as an indicator of whether the protein is folded. The observed Trp fluorescence indicates that all of the toothfish γM-crystallins investigated here remain folded over the entire temperature range where LLPS was observed. For comparison, the fluorescence signal was also measured at 80 °C, where all of the proteins are unfolded.

The two most cold cataract-resistant proteins are γM4 and γM8d, while the least resistant are γM7 and γM8b. Examining the sequence relationships among the outliers in cold cataract resistance (Fig. 3.1) shows that γM4 and γM7 have relatively low sequence identity to the rest of the γM-crystallins, whereas γM8b and γM8d are closely related to each other. In order to better understand the physical basis for the differences in LLPS onset temperature, we characterized the biophysical properties of these proteins.

3.3.4 Toothfish γM-crystallins are highly stable and aggregation-resistant

In addition to cold tolerance, the toothfish crystallins have been selected for refractive index increment, stability and aggregation resistance, as have all lens crystallins [6]. In order to test the hypothesis that selection for cold tolerance would lead to reduced thermal or chemical stability, we measured heat denaturation and urea unfolding of the γM-crystallin paralogs, as well as their thermal aggregation propensities. To provide a baseline for these

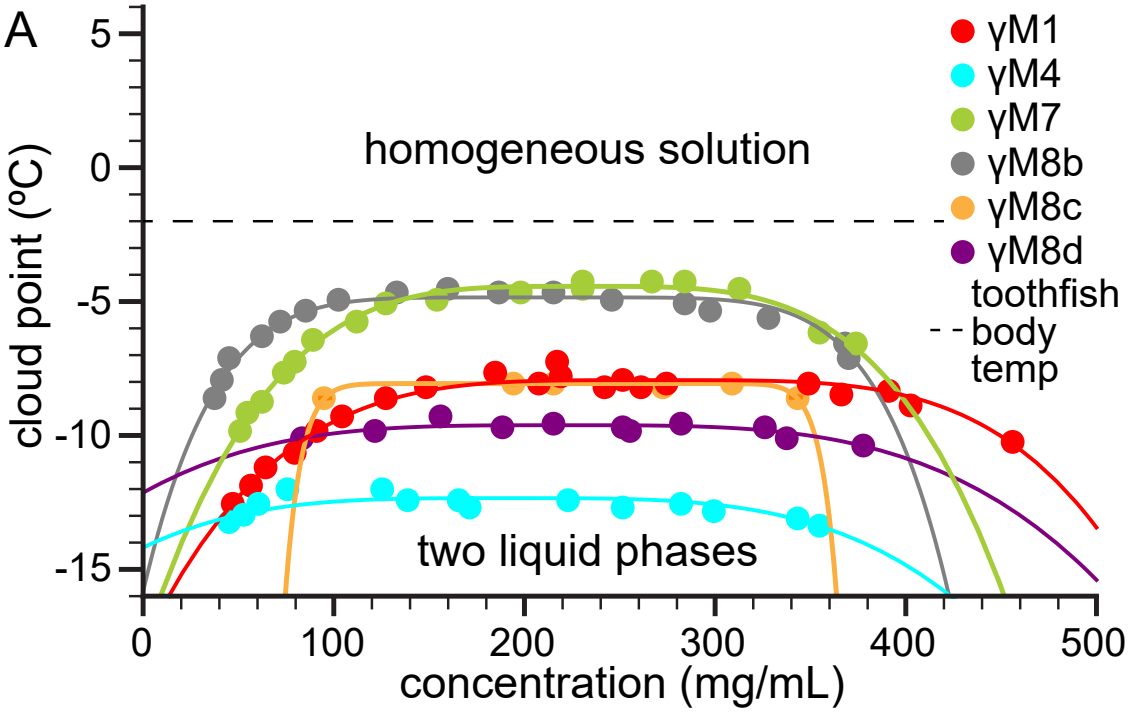

Figure 3.3: Coexistence curves of γM1 (red), γM4 (cyan), γM7 (lime green), γM8b (gray), γM8c (orange), and γM8d (purple). The data were fit to a binodal function [95]. The dashed line represents the average toothfish body temperature. The phase separation is reversible for samples allowing for triplicate measurements. All proteins have similar critical concentrations of 268, 191, 230, 211, 219, and 220 mg/mL, respectively. The T_c are -7.9, -12.4, -4.4, -4.8, -8.1, and -9.6 °C, respectively. γM7 and γM8b show the highest T_c in this set, yet both T_c values are below the average body temperature of *D. mawsoni*. The other γMs show an increased resistance to LLPS with γM4 and γM8d being the most resistant.

measurements, the recombinant γM-crystallins' secondary structure was characterized by circular dichroism (CD) spectroscopy. The resulting spectra had minima between 216 and 220 nm, characteristic of the high β-sheet content of the βγ-crystallin fold (Fig. A.7A). The intrinsic tryptophan fluorescence was measured in order to estimate the degree of tryptophan exposure. All of the γM-crystallins show maximal emission between 318 and 322 nm, indicative of a non-polar environment for the tryptophan side chains, consistent with their effective burial within the protein (Fig. A.7B). Interestingly, although γM7 has an additional tryptophan residue compared to the other γM-crystallin paralogs, it has the lowest emission intensity. This additional γM7 tryptophan is in a position that is conserved in other vertebrate γ-crystallins and has been shown to both be quenched and absorb energy from the other conserved tryptophan in the same domain [48].

All the toothfish γM-crystallins are relatively robust with respect to chemical denaturation: the concentration of urea needed to halfway unfold the γMs ranges from 4.1 M for γM8d to 5.7 M for γM8b, with γM8b surpassing the chemical stability of HγS [47]. With the exception of γM8b, all the γMs denature at a comparable concentration of urea. Urea denaturation is reversible for these proteins, as demonstrated by the open circles in Fig. 3.4A, which represent refolding of the protein after initial denaturation in 8 M urea. This reversibility enables calculation of the ΔG of unfolding (Fig. 3.4B). Thermal unfolding was also measured; this process is not reversible and therefore cannot be used to calculate ΔG of unfolding. Even though these crystallins come from a cold-adapted organism, they all exhibit high thermal stability: γM7 has the lowest thermal denaturation temperature (T_m) of 59 °C, whereas γM8b shows the highest thermal stability with a T_m of 71 °C, comparable to HγS (Fig. 3.4C) [110]. These crystallins also show no sign of cold denaturation down to -10°C as would be expected for these highly stable, cold adapted proteins (Fig. A.6). γM4 has a T_m of 69.3 °C, only two degrees below γM8b, yet the concentration of urea needed to halfway denature the protein most closely corresponds to the rest of the γMs with lower T_ms (Fig. 3.4A and C).

We investigated the thermal aggregation propensity of the γMs using dynamic light scattering (DLS); this experiment measures the size distribution of aggregates formed in solution, which can begin well below the thermal unfolding temperature. For these crystallins, the onset of aggregation begins over a moderate temperature range (Fig. 3.4D). With the exception of γM8b and γM4, the γMs show an inverse trend in their aggregation propensity compared to their thermal stability. γM1 is the most aggregation-prone, forming aggregates at 43 °C, while γM8b has the lowest aggregation propensity, resisting aggregation until 56 °C, comparable to HγS. Interestingly, despite γM4's increased thermal stability compared to all but γM8b, it is no more aggregation-resistant than the other γM-crystallins. This is consistent with previous observations that protein aggregation can occur well below the protein T_m, for example in human γD- [111, 112] and γS-crystallin [47], as well as in fish γM-crystallins [20].

3.3.5 Cold tolerance does not correlate with cross-species chaperone binding

Protein adaptation to the crowded environment of the eye lens represents a tradeoff between the need to avoid protein-protein interactions leading to aggregation or LLPS, while maintaining those required to maintain lens transparency [3]. α-crystallins, which are homologous to small heat shock proteins, act as holdase chaperones in the lens, protecting the structural $\beta\gamma$-crystallins from stress-induced denaturation, aggregation, and the precipitation that causes cataract [5]. The α-crystallins bind their client proteins by preferentially recognizing stress-induced unfolded intermediates or aggregation-prone variants over properly folded, wild-type proteins [34], and at least one cataract-related variant of human γD- crystallin is aggregation-prone specifically due to increased interaction with alpha-crystallins [113].

We previously found that α-crystallins specifically bind structural $\beta\gamma$-crystallins from their

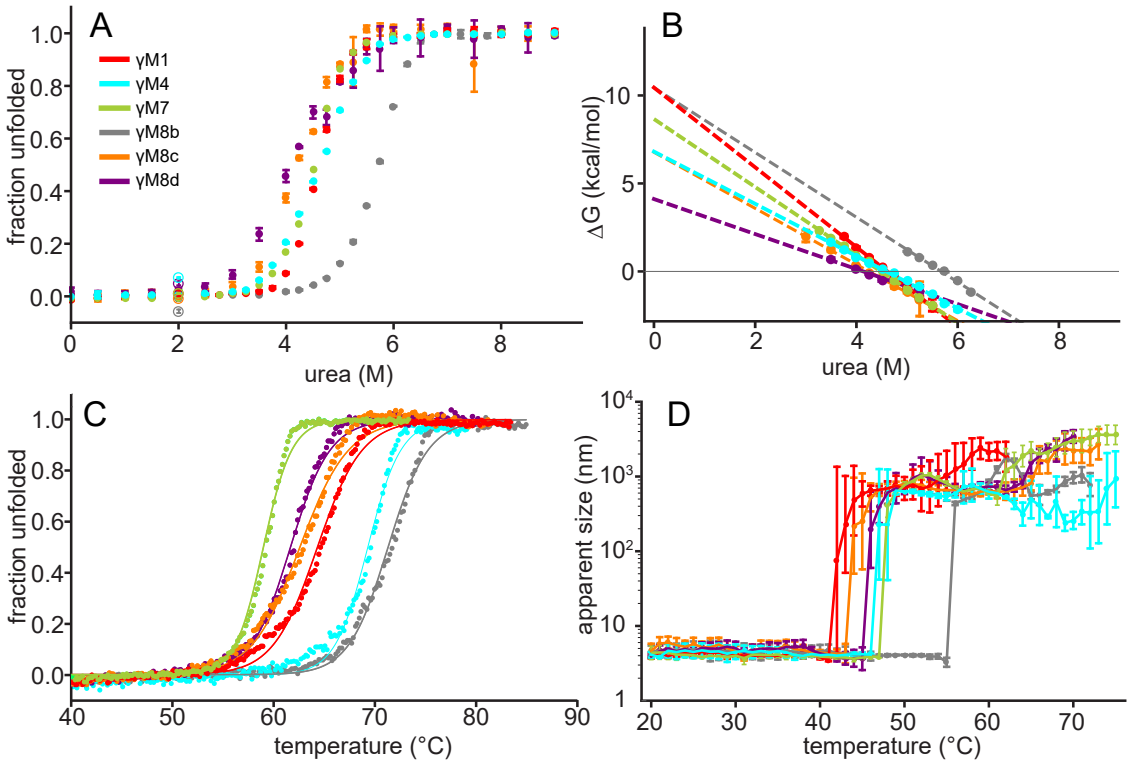

Figure 3.4: (**A**) Urea denaturation for γM1 (red), γM4 (cyan), γM7 (lime green), γM8b (gray), γM8c (orange), and γM8d (purple) as determined by measuring the 360/320 nm fluorescence emission ratio with a 280 nm excitation. The $[urea]_{1/2}$ of each protein was 4.6, 4.6 4.5, 5.7, 4.2, and 4.1 M, respectively. γM8b shows the highest stability with respect to chemical denaturation. Open circles represent dilution of protein samples in 8 M urea to a final concentration of 2 M urea. (**B**) Gibbs free energy plot as a function of urea concentration with ΔG°_{w} of 10.1, 6.8, 8.5, 10.4, 6.7, and 4.2 kcal/mol, respectively. Interestingly, the Gibbs free energies for unfolding vary among the γM8s despite similar half-unfolded urea concentrations. This disparity could be due to small differences in the interactions keeping each protein folded or may be an indication that the unfolding under urea denaturation does not strictly follow a two-state model. (**C**) Thermal denaturation curves were determined by measuring the absorbance of circularly polarized light at 218 nm, which corresponds to the beta sheet minima, as a function of temperature. Thermal unfolding temperatures (T_m) of 64.3, 69.3, 59, 71.3, 62.6, and 61.7 °C, respectively, were observed. γM8b shows an unexpectedly high T_m. (**D**) Thermal aggregation as determined by particle size using DLS. All five γM crystallins show a very sharp transition from monomers to larger aggregates. γM8b exhibits a much lower thermal aggregation propensity than the other γM-crystallins, consistent with its high T_m value, but γM4 does not.

own species under thermal stress, suggesting a co-evolutionary relationship based on physiological, and in the case of stenothermal fishes, environmental conditions [24]. Here we have employed cross-species chaperone assays as a functional, rather than a purely sequence- or structure-based measure, of evolutionary relatedness. We have previously reported that fish crystallins that are more closely related to their mammalian counterparts were better protected against thermal aggregation by bovine α-crystallin than the more distantly related paralogs [24]. The results of the cross-species assays for recombinant toothfish crystallins incubated with bovine eye lens α-crystallin fraction are shown in Fig. 3.5.

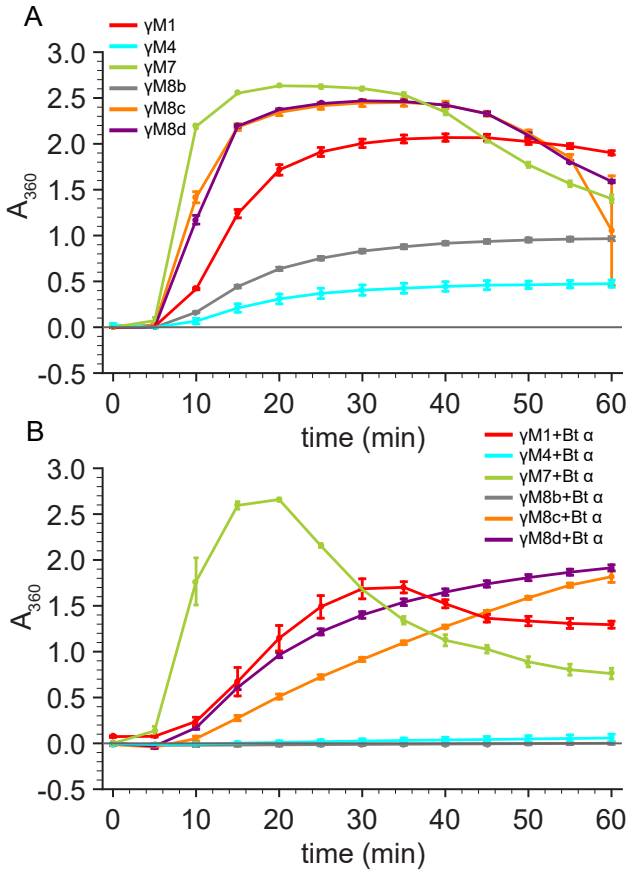

Figure 3.5: This assay measures the extent to which recombinant γM-crystallins from the Antarctic toothfish are protected by bovine α-crystallin from thermally-induced aggregation. (**A**) All the γMs aggregate when incubated at 60 °C, with γM8b (gray) and γM4 (cyan) displaying significantly more aggregation resistance than the other γMs. (**B**) γM8b and γM4 are completely protected by bovine α-crystallin under the experimental conditions tested, whereas γM1, γM8c and γM8d are only partially protected. In contrast, γM7 shows almost no protection from aggregation by bovine α-crystallin.

γM8b and γM4 display elevated aggregation resistance relative to the other γMs (Fig. 3.5A), and both are well protected by bovine α-crystallin, with thermal aggregation being completely suppressed (Fig. 3.5B). The aggregation of the other four γMs (γM1, γM7, γM8c and γM8d) are only partially retarded, with more protection seen for γM8c and γM8d than either γM1 or γM7. Based on these data, the ability to be protected from heat-induced aggregation by bovine α-crystallin does not correlate with resistance to cold cataract; γM4 and γM8b represent the most resistant and one of the two most susceptible to cold cataract, respectively, yet these are the two most well-protected proteins. These results indicate that the surface residues responsible for making the γM-γM interactions leading to LLPS are distinct from those recognized by α-crystallins.

3.3.6 Mutational analysis of γM8b

Given that the onset temperature of LLPS does not correlate straightforwardly with amino acid composition, thermal or chemical stability, aggregation propensity, or binding affinity to bovine α-crystallin (Table A.3), we turned to more specific hypotheses about the molecular origin of LLPS. The pair of proteins with the largest difference in the onset temperature of LLPS is γM4 and γM7, which have only 55% sequence identity (see Fig. A.8), making it difficult to pinpoint critical sequence positions for mutational studies. Therefore, we chose to examine γM8b and γM8d, which exhibit a dramatic difference in resistance to cold cataract despite their much greater (83.6%) sequence identity. Taking γM8b as a starting point, we set out to change this relatively cold-cataract susceptible γM-crystallin paralog into a more resistant one using a small number of mutations.

The sequence alignment shown in Supplementary Fig. A.8A reveals relatively few positions where three or more of the proteins do not share the same residue (colored red). Many of these positions lie in loop regions or at the N- and C-termini, while the core associated with

the Greek key motifs is more highly conserved. In many cases where there are sequence differences, chemical functionality is conserved, such as substitution of a residue with the same charge or that of one hydrophobic side chain for another. The sequence conservation is mapped on a molecular model of γM8b in Fig. 3.6A and 3.7A. The highlighted residues are those chosen for mutagenesis based on their predicted ability to mediate intermolecular interactions leading to LLPS. Inspired by previous findings that low-T_c and high-T_c γ-crystallins are often highly similar, with differences in only a few key positions [95], we mutated surface residues to investigate their effects on T_c. Based on the premise that the propensity of a protein to undergo LLPS in concentrated solutions is due to specific intermolecular interactions that may be abolished by changing the chemical properties of a key site, two hypotheses are proposed. The first – the surface properties hypothesis – is that the residues responsible for the difference in cold-cataract resistance between γM8b and γM8d can be found by following the sequence variation and focusing on residues that are well-conserved overall, but exhibit a significant difference in size or chemical functionality between the two proteins of interest. The second – the interaction strength hypothesis – is that the fundamental principle is the strength and number of the potential intermolecular interactions formed, rather than the chemical properties of the residues themselves.

In order to test the surface properties hypothesis, we examined the sequence alignments and models for surface residues that were highly conserved among the toothfish γM-crystallins, focusing on residues that were conserved in at least four of the six sequences examined here, but differing between γM8b and γM8d. Sequence conservation is mapped onto a comparative model of γM8b made using the Robetta server [107] in Fig. 3.6A. Positions where a lack of consensus suggests that variation is well tolerated from a functional standpoint were excluded from consideration, as were those within two residues of non-consensus sites, and those that are not surface-exposed based on the structural model. This analysis led to the choice of Q122 and S128 (γM8b numbering). In γM8d, the corresponding residues are Met and Asn, respectively, resulting in the γM8b variants Q122M and S128N (highlighted in orange in

Fig. 3.6A and Fig. A.8). Both of these mutations make chemically significant changes to the side chain, potentially disrupting intermolecular interactions and affecting the onset of LLPS. Substitution of Ser for Asn represents a change in the size of the side chain in this surface-exposed position, while substitution of Met for Gln places a hydrophobic residue on the surface instead of hydrophilic one. We hypothesized that one or both of the Q122M and S128N mutations would lower the T_c of γM8b to a value similar to its next closest paralogs (γM8c and γM8d, with T_c of -8.1 °C and -9.6 °C respectively).

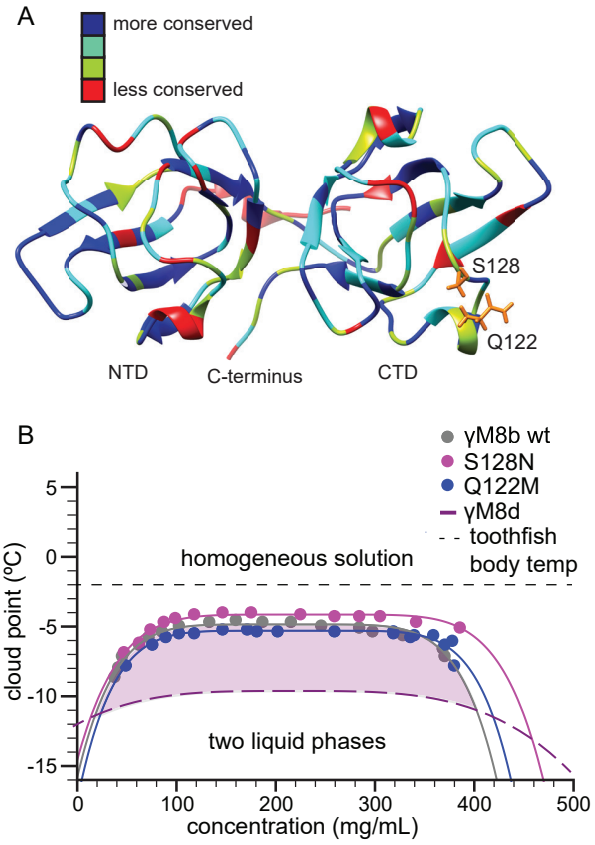

Figure 3.6: (**A**) Structural model of γM8b colored according to sequence conservation among the toothfish γM-crystallins (see alignment in Fig. A.8A). Conservation was determined from the sequence alignment of the six γMs in this study where none, one, two, and three or more sequences that deviate are colored with blue, cyan, green and red respectively. The mutated residues, Q122 and S128, are colored orange. (**B**) Coexistence curves of γM8b (gray) and its variants Q122M (blue), and S128N (magenta). The dashed purple line is the binodal curve of γM8d, while the dashed black line represent the average toothfish body temperature. The region between γM8b and γM8d is shaded light purple. The three γM8b variants show similar critical concentrations at 211, 220, and 232 mg/mL respectively. The Q122M and S128N variants have T_c of -5.3, and -4.1 °C, respectively, similar to wild-type (-4.5 °C).

As expected, biophysical characterization of the variants, performed in the same manner as for the wild-type proteins (heat, urea), indicated only negligible changes in stability relative to wild-type γM8b (Fig. A.9, A.10). However, the change in LLPS onset temperature for these variants was very small and shifted in opposite directions for Q122M and S128N (Fig. 3.6B). For Q122M the T_c dropped by 0.5 °C, while for S128N it increased by 0.7 °C. The latter observation was particularly unexpected, because all the other lower T_c γMs tested have an Asn in the equivalent position. This disparity indicates that despite their overall level of conservation, these two residues play only a small role in the LLPS behavior of the γM crystallins.

3.3.7 Targeted mutation of surface residues enables tuning the LLPS onset temperature

Our initial hypotheses were developed under the assumption that surface charge and polarity would be the major drivers of LLPS in this system. However the lack of significant differences in these properties between γM8b and γM8d, which have a large difference in the onset temperature of LLPS led us to look for other factors. Positions where Arg is substituted for Lys or vice versa between γM8b and γM8d were initially overlooked. Because R and K are both positively charged, this mutation would be expected to have no effect on the surface properties and hence LLPS based on charge alone. However, arginine is capable of making more and stronger electrostatic interactions with the carboxylic acid groups of aspartic acid and glutamic acid because of the resonance structure of the guanidinium group, its higher pKa, and its potential for making bifurcated hydrogen bonds [114]. In addition, the arginine guandinium group has been shown to be weakly hydrated suggesting that it is likely to be excluded from bulk solvent towards weakly hydrated surfaces releasing water molecules to the bulk solvent [115]. The release of weakly hydrating water has also been implicated as a contributing factor in the unusual occurrence of arginine-arginine stacking

despite their electrostatic repulsion [116]. Based on analysis of structures in the Protein Data Bank, arginine is also more likely than lysine to be found in cation-π interactions [117]. Mutations or post-translational modifications of surface-exposed arginine residues are also known drivers of altered intermolecular interactions between α-crystallins and their client lens proteins [118]. Therefore, the first mutation sites chosen to test the interaction strength hypothesis were found by looking for surface-exposed positions where there is an R in γM8b and a K in γM8c or γM8d. There are three such sites, Arg 104, 161, and 164, all located in the C-terminal domain; they are highlighted in black in Fig. 3.7A). The triple variant R104K/R161K/R164K was made to investigate the effect of highly specific intermolecular interactions on cold cataract. The mutations did not change the stability of the protein (Fig. A.9, A.10), however the T_c dropped by 8.8 °C from - 4.9 °C in wild type to -13.7 °C in the triple variant, making it the most cold-resistant of all the proteins tested (Fig. 3.7). The single variants R104K, R161K, and R164K each change the onset temperature of LLPS by approximately one third of the difference observed for the triple variant (Fig. A.11), suggesting that the effect is additive and depends on surface exposure rather than specific sequence position.

To further establish the role that arginine plays in LLPS, three lysines in predicted loop regions (Lys 18, 85, and 172) were chosen for mutagenesis, resulting in K18R/K85R/K172R in γM8b. The variant protein had biophysical characteristics consistent with wild-type, with a small increase in chemical and thermal stability (Fig. A.9, A.10). However, this mutation produced a change in T_c similar in magnitude to that observed in the triple Arg to Lys variant, but opposite in direction. The T_c of this variant is 3.5 °C, 8.3 °C higher than that of wild-type γM8b, making it by far the least cold-resistant of all the proteins tested (Fig. 3.7). This increase in phase transition temperature allowed us to probe the phase separation by dynamic light scattering (DLS) and optical microscopy. As shown in Fig. 3.7C, as the sample is cooled, phase separation can be detected beginning at at 4 °C, with an increase in particle size and size variance as the temperature continues to drop. The droplets are visualized

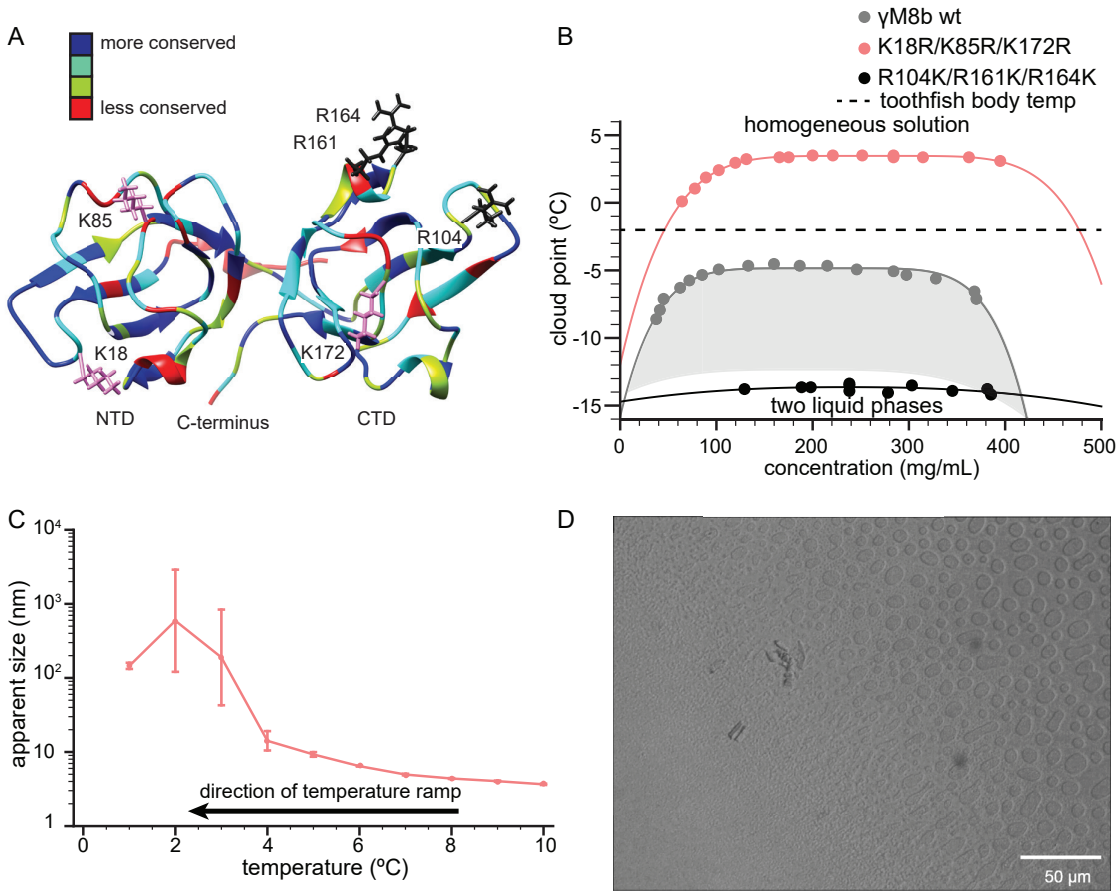

Figure 3.7: (**A**) The positional coloring from the sequence alignment of Fig. A.8A is superimposed onto a comparative model of γM8b; positions where none, one, two, and three or more of the six sequences in this set differ are colored with blue, cyan, green and red respectively. The mutated residues, R104, R161, and R164 are colored black while K18, K85, and K172 are colored pink. (**B**) Coexistence curves of γM8b (gray) and its triple Arg to Lys (R104K/R161K/R164K, black) and triple Lys to Arg (K18R/K85R/K172R, pink) variants. The gray shading is the area between γM8b and γM4 coexistence curves. The black dashed line represents the average body temperature of the toothfish. All proteins show similar critical concentrations at 211, 235, and 261 mg/mL respectively. The triple Arg to Lys and triple Lys to Arg variants exhibit dramatic changes in the LLPS onset temperature of similar magnitude and in opposite directions, with T_c of -13.7 and 3.5 °C respectively. **(C)** Particle size of a 150 mg/mL solution of triple Lys to Arg γM8b was monitored as the temperature was reduced in 1 °C increments. The apparent size is the average diameter of particles normalized according to particle volume. The viscosity correction for analysis of the DLS data were obtained using the viscosity data shown in Fig. A.12. An increase in size variance starts at 4 °C, slightly above the phase transition temperature, and increases for following temperatures with a dramatic increase in particle size occurring around 3 °C. **(D)** Micrograph of a 150 mg/mL solution of triple Lys to Arg γM8b cooled to the phase transition temperature. The image shows the phase transition front where larger droplets have formed closer to the heat sink (top right corner), while smaller droplets are forming farther away where the slide is warmer (bottom left corner).

in Fig. 3.7D; this image shows a sample placed on a slide that is subject to a temperature gradient. Smaller droplets are observed at the LLPS front, while larger droplets are present closer to the heat sink, reaching sizes on the order of 10 μm. Over time, these droplets coalesce and settle out into two separate phases as shown in Fig. A.13.

Interestingly, when the 10 mM phosphate was replaced with MOPS buffer at an equivalent pH, LLPS was unobservable even with the triple Lys to Arg γM8b variant. One possible explanation for this observation is the difference in ionic strength between the two buffers. MOPS at pH 6.9 exists in equilibrium between forms with 0 or -1 overall charge in comparable proportions. In contrast, at this pH the comparable equilibrium for phosphate is between species with -1 and -2 overall charge, thus giving rise to increased ionic strength. This could be important when considering that although these proteins are at most 1 pH unit below their predicted isoelectric points, a net charge that is not sufficiently shielded would lead to repulsion between individual proteins.

Alternatively, the -2 species of phosphate may be participating in bridging interactions between positive charges on different protein monomers. To test the latter hypothesis we increased the phosphate concentration expecting that if phosphate was increased, so too would the T_c. Increasing the phosphate concentration to 50 mM for wild-type γM8b led to little change, with a negligible decrease of the T_c to -5.0 °C, only 0.2 °C lower than the standard buffer (Fig. 3.8). However, when the triple Lys to Arg variant was also tested the T_c increased 3 °C from 3.5 to 6.5 °C.

To further investigate the effects of the solution composition on intermolecular interactions, salts with different cations were added to the standard 10 mM phosphate buffer in moderate concentrations (50 mM)(Fig. 3.8). When guanidinium HCl was added, both the wild-type γM8b and the triple Lys to Arg variant displayed a lowering of their T_c, to -6.8°C and 2.2°C, respectively. However, with the additional of either NaCl or NH_4Cl, the trend was more similar to that seen with increased phosphate in solution. In the case of NaCl the T_c

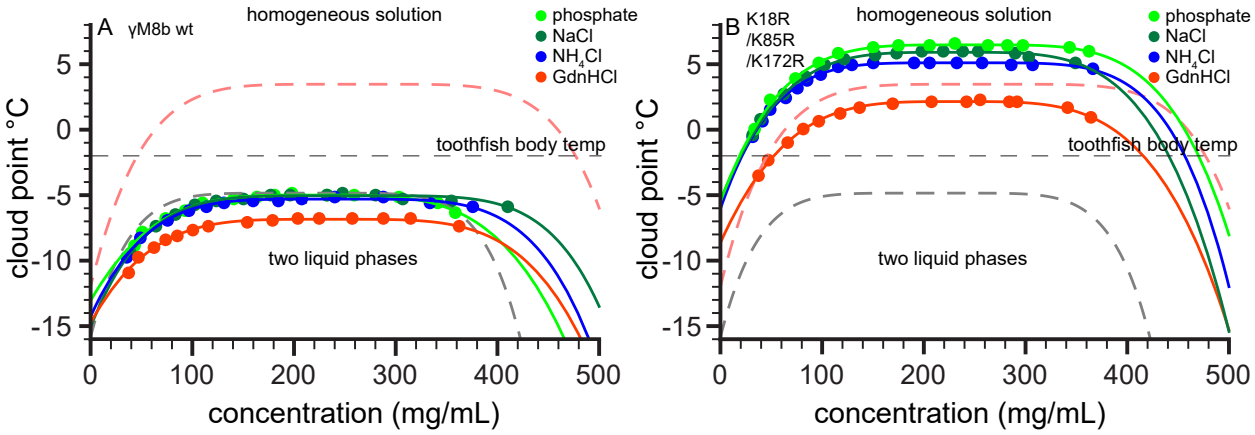

Figure 3.8: LLPS coexistence curves were probed after exchanging the buffer from the standard 10 mM phosphate, pH 6.9. The buffers tested consisted of 50 mM phosphate pH 6.9 (green) and the standard 10 mM phosphate, pH 6.9 buffer plus the addition of either 50 mM NaCl (dark green), NH_4Cl (blue), or guanidinium HCl (red-orange). The dashed gray and pink line represent the coexistence curves of γM8b wt and the triple K to R variant, respectively, in standard buffer. (A) The coexistence curves of γM8b wt was measured in the presence of additional buffer components. The presence of NaCl, NH_4Cl, or increased phosphate had little effect on the critical temperature with values of -5.0, -5.3, and -5.0°C, respectively, compared to the critical temperature of -4.8°C in the standard buffer. The presence of guanidinium HCl had a more pronounced effect, with a critical temperature of -6.8°C. (B) The coexistence curves of γM8b K18R/K85R/K172R variant were also measured in the presence of additional buffer components. The presence of NaCl, NH_4Cl, or increased phosphate raised the critical temperature to 5.9, 5.1 and 6.5°C respectively compared to a critical temperature of 3.5°C in the standard buffer. Similar to wt, guanidinium HCl lowered the triple variant critical temperature to 2.2°C

were -5.0°C and 5.9°C for wild-type and the triple variant, respectively. For NH_4Cl, the T_c were -5.3°C and 5.1°C for wild-type and the triple variant, respectively. The observation that overall, the larger effect of buffer concentration and added salts on the triple Lys to Arg variant as compared wild-type, indicate that protein-buffer and protein-salt interactions in this system are complicated and may depend on the positioning of particular residues.

3.4 Discussion

A major function of the toothfish eye lens crystallins is their ability to resist cold cataract, or LLPS, at low temperatures [24]. LLPS occurs due to attractive interactions between individual proteins, driving coalescence into a separate liquid phase [52, 95]. The observation that *D. mawsoni* eye lens fraction FIII undergoes LLPS at temperatures above the toothfish's body temperature, when this behavior is not observed for either the whole lens or the individual recombinant proteins, is unexpected. In other vertebrate crystallins, large variations in T_c have been observed [119], and T_c values above body temperature are associated with cataract-related mutations [79]. Although the T_c of these γM-crystallins vary considerably, even γM7, which has the highest T_c at -4.43 °C, remains transparent nearly 2 °C lower than the average body temperature of the toothfish. This raises the possibility that there may be a synergistic effect among these proteins that raises the T_c of FIII under our experimental conditions. In the bovine lens, the mixture of paralogs has been shown to affect the T_c. Bovine γS-crystallin not only resists LLPS at temperatures as low as -10 °C, it also prevents LLPS when mixed with more susceptible γ-crystallins [53].

On the other hand, demixing of protein solutions can occur in conjunction with phase separation if the intermolecular interactions between identical protein molecules are significantly more favorable than those between different paralogs [96]. Preventing this type of demixing and its concomitant phase separation may be one reason why the toothfish has so many

γ-crystallin paralogs. Another possible explanation for the difference in LLPS temperature between the susceptible fraction and the individual recombinant proteins is that the coexistence curves were measured using freshly prepared recombinant protein, while the lens fraction came from an adult fish where the protein may have accumulated deleterious post-translational modifications. Even very low modification levels have been shown to result in substantial changes in the T_c of the bulk solution [120]. Although fragmentation ions for conclusive identification of post-translational modification were lacking in the mass spectra of Fraction III, based on studies in human lens crystallins, [121, 122, 123] the most common modifications are expected to be oxidation and deamidation. Methionine oxidation to a sulfoxide or deamidation of asparagine or glutamine would lead to the presence of an additional hydrogen bonding acceptor or negative charge, potentially increasing the interaction potential between proteins based on the additional interactions (e.g. salt bridges and hydrogen bonding) available to each side chain.

The many γM-crystallin paralogs in the toothfish lens share the double-Greek key fold common to vertebrate $\beta\gamma$-crystallins. Despite the ubiquity of this fold, the primary sequences of vertebrate structural crystallins can vary considerably among and within organisms [20]. For example, human γS-crystallin (hγS) has 55% identity to its closest human paralog, hγB. Comparing hγS to its closest orthologs in other vertebrates, we find 71% identity for *Danio rerio* γS1, 72% for *Gallus gallus* γS, and 58% for *Xenopus laevis* γA. These sequence variations are suspected to underlie functional diversity, given the dual structural and optical functions of these proteins. From an evolutionary standpoint, the need for higher refractive index in the lens is addressed both by increasing the protein solubility (and hence the maximum concentration), as well as the refractivity of the proteins themselves. It is clear from the amino acid composition of the toothfish γM-crystallins that this has resulted in an enrichment of highly refractive amino acids with a corresponding decrease in low refractive index amino acids. This necessarily leads to a lower than average composition of aliphatic amino acids, as they have the lowest refractive index increment values. Methionine, which otherwise

behaves like an aliphatic residue in terms of chemical properties, has a substantially higher refractive index (Table A.2). An enriched methionine content serves at least two functions in the lens; it increases the protein refractive index increment and substantially alters the hydration properties of the protein surface [98]. It may also be important for mediating transient protein-protein interactions in the crowded lens environment, due to the flexibility of the methionine side chain compared to other hydrophobic residues [124]. Proline is also noticeably exiguous in these psychrophilic crystallins compared to hγS (Fig. 3.2), consistent with the hypothesis that psychrophilic proteins have evolved to be more flexible [125].

Thermal and chemical denaturation curves were measured for all the toothfish γM-crystallins in order to measure their stability, heat tolerance, and aggregation propensity. Chemical and thermal denaturation provide different information: denaturation under reversible conditions (in this case by adding urea) enables calculation of the ΔG between the folded and unfolded states, while the thermal denaturation is often inextricably linked with aggregation and is characterized by the midpoint temperature of the unfolding curve, T_m. In general, T_m does not correlate reliably with ΔG between the folded and unfolded states [87]. However it does provide a useful point of comparison with the thermal aggregation temperature, enabling assessment of whether the onset of aggregation coincides with large-scale unfolding or if transiently unfolded states begin aggregation below T_m [126]. Despite originating from a cold-adapted organism, these crystallins still exhibit relatively robust thermal stability: γM7 has the lowest thermal denaturation temperature (T_m) at 59 °C, while γM8b shows the highest thermal stability, with a T_m of 71 °C, comparable to HγS (Fig. 3.4C). It is interesting that γM4 has a T_m of 69.3 °C, only two degrees below γM8b, yet the urea concentration needed to halfway unfold the protein most closely corresponds to the rest of the γMs with lower T_ms (Fig. 3.4A and C) highlighting our previous findings that homologous proteins can show differential stabilities depending on the denaturation method [67].

We hypothesize that the increased stability seen in γM8b may be partly due to the N-

terminal extension, which is only present in the γM8s , making contact with the C-terminal domain. Judging from the comparative models γM8b and γM8c are capable of making polar contacts between the N-terminal amine and Asp137, potentially forming a stable salt bridge that could resist the type of initial N-terminal unfolding first observed in *D. rerio* γM7-crystallin and hypothesized to be a general unfolding mode of the $\beta\gamma$-crystallin fold [17]. Additionally, the substitution of Gly5 to Asp in γM8b relative to γM8c lends more helical character to the predicted secondary structure of the N-terminal extension. As a result, the N-terminal extension of γM8b packs better along the interdomain linker than that of γM8c. Taken together, the presumptive salt bridge and the more efficient packing of the N-terminal extension appears to put the two domains more stably in contact with each other. This type of cooperative stabilization leading to increased thermal and chemical tolerance has previously been observed in mammalian γS-crystallins [127].

One striking result of this study is that substituting arginine for lysine and vice versa significantly affects the T_c even though both are positively charged amino acids and therefore contribute equally to the surface charge of the protein. The importance of charged surface residues in general is in line with the current view that biological LLPS is driven by complex coacervation of highly charged biomolecules [128]. The dramatic effect of substituting Arg for Lys is consistent with the over-representation of Arg at protein-protein interfaces [129]. Furthermore, previous sequence analysis of low- and high-T_c γ-crystallins showed that one salient feature of high-T_c γ-crystallins was an Arg in a particular C-terminal domain loop, while the low-T_c crystallins had a lysine instead [130]. This position coincides with two of the residues mutated in γM8b, R161 and R164. Similar effects have been observed in other proteins that undergo LLPS as well. In the stress granule protein *C9orf72*, arginine-rich dipeptide repeats are essential for LLPS and stress granule assembly [131]. Recent *in vitro* and *in vivo* studies of FUS have indicated that arginine methylation inhibits phase separation, [132], which is specifically modulated by cation-π interactions [133], and lies along the pathway toward pathological aggregates.

We hypothesize that the effect of Arg/Lys substitution on T_c results from the differences in geometry and electronic structure in the side chains of Arg and Lys, which in turn lead to the formation of different intermolecular interactions. Furthermore, the weaker hydration of Arg side-chains leads to increased association with weakly hydrated surfaces of other proteins [115, 116]. All but two of the proteins in this set have positively charged residues in the positions occupied by the three highly exposed positively charged residues (indicated in black in Fig. 3.7A) that were mutated. The exceptions are γM8d, which retains only two of the three, and γM7, which has one of the positive positions offset by two residues, although based on the comparative model, the side chain still occupies a similar spatial position, making it equally surface-exposed. γM8b and γM7 contain Arg in all three of their respective positions, γM4 has two Arg and one Lys, γM1 and γM8c have one Arg and two Lys, while γM8d only has two Lys and a Met in the third position. With the exception of γM4, this trend is consistent with the T_cs seen in these five crystallins, where a greater number of arginine residues in these surface positions correlates with a higher T_c. The deviation of γM4 from this trend may be explained by the substitutions of Arg to Lys at positions 50 and 90, which is only observed in γM4. These sequence positions correspond to flexible coil regions, suggesting that they are also exposed at the surface.

The role of arginine in these crystallins is especially interesting when it is considered that these six γM-crystallins show an enrichment of Arg with a concomitant depletion of Lys. In terms of attractive interactions that increase the propensity of phase separation, enriched Arg content would be undesirable. However, within the full context of a crystallin's function, Arg enrichment has the desirable effect of increasing protein refractivity [11, 60]. These two roles appears to have led to a sort of tension in the selection for Arg enrichment, where increased arginine content is desirable to increase refractivity, but not so much as to compromise the lens with the risk of phase separation. Comparing the Arg/Lys content of γM-crystallins of the toothfish to other, warmer-water fish species reveals that all of these crystallins favor Arg over Lys (Fig. A.14). However, many of the toothfish γM-crystallins have lower Arg

and higher Lys content relative to those seen in the other fish species. These results suggest that the reduced Arg content may have been a feature selected for cold tolerance.

The role that dissolved ions play in the LLPS of the these proteins is another important consideration. Here, changing the buffer from phosphate to MOPS reduced the T_c beyond what is experimentally accessible. This may be due to the decreased ionic strength being insufficient to shield charges causing inter-protein repulsion. The alternative hypothesis, that -2 charged phosphate ions may be able to bridge positive charges between proteins, was not well supported by the observation of LLPS in the presence of more phosphate. In this case, wild-type γM8b shows little change in the onset temperature of LLPS, while the triple Lys to Arg γM8b variant has a 3°C increase in the T_c. The addition of NaCl or NH_4Cl followed a similar trend, suggesting the effect is not unique to phosphate or ions with a charge of -2. Addition of guanidinium HCl reduced the T_c for both wild-type and the triple variant, with wild-type type exhibiting a slightly greater reduction in T_c. These observations lead us to hypothesize that, provided the ionic strength of the solution is sufficiently high to shield repulsive net charge between individual proteins, the kosmotropic and chaotropic nature of the ions predominate as far as their modulating effect on the LLPS temperature. This also supports the idea that the increased interaction from arginine is at least in part driven by hydration forces; as kosmotropic ions would more strongly exclude the arginine side chain towards weakly hydrated surfaces. Similar observations have also been made in regard to the phase behavior of hen egg white lysozyme, where different ion species demonstrate varied effects on the phase transition temperature [134, 135, 136]. Such difference require further and more extensive study to better understand the interactions at play between proteins and dissolved ions.

In conclusion, we have characterized the stability and LLPS behavior of six γM-crystallins from the Antarctic toothfish *D. mawsoni* and site-directed variants thereof. The LLPS of these crystallins does not correlate strongly with thermal or chemical stability or ability to

be recognized by bovine α-crystallin. Instead LLPS formation depends on the presence of solvent-exposed residues that have a higher propensity to interact with residues of another individual protein. We also demonstrated that it is possible to tune the LLPS temperature by adjusting the Lys/Arg composition without varying the overall charge. Mutagenesis of three exposed Lys to Arg in γM8b leads to a shift in the T_c of a similar magnitude but opposite direction compared to the triple Arg to Lys variant. It is worth noting that in the triple Arg to Lys variant, the mutations were clustered in two adjacent loops in the C terminal domain, while the triple Lys to Arg variant has two mutations in the N-terminal domain and one in the C-terminal domain. The observation that both triple variants have similar magnitudes of their T_c shift and that the single variants R104K, R161K, and R164K have magnitude shifts approximately one third the triple variant (Fig. A.11) suggest that the effect is additive and not strongly positionally dependent provided that the residues are adequately exposed. The results of this study suggest that LLPS can be controlled, even in well-folded proteins, by manipulating specific interactions between residues on the protein surface, a direction that bears further investigation given the importance of such systems in cell biology and the promise of creating tunable membraneless organelles to serve as microreactors or modulators of solution conditions.

Chapter 4

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

4.1 Background

In organisms with camera-type eyes, the transparent, refractive medium that focuses light on the retina is made up of densely packed crystallin proteins [6]. This specialized tissue is a crowded molecular environment; the protein concentration ranges from 400 mg/mL in the human lens to nearly 700 mg/mL in some aquatic species [137, 10, 98]. In terrestrial organisms, much of the refractive power is provided by the air/water interface at the cornea. In aquatic animals, the lens alone is responsible for refraction; hence their higher protein concentration and, on average, greater refractivity of the proteins themselves. Vertebrates share two conserved lens protein classes, the α-crystallins, which are small heat shock proteins [5], and the structural $\beta\gamma$-crystallins, which are primarily β-sheet proteins with a characteristic

two-domain double Greek key fold [138]. In addition to these crystallin superfamilies, there are a variety of taxon-specific crystallins. For example, the S-crystallins of cephalopods are thought to have evolved from the enzyme glutathione S-transferase [139, 140, 141]. The ϵ crystallin of crocodiles and some birds is identical to lactate dehydrogenase; in the duck lens, it retains its catalytic activity even at very high concentrations [142]. The box jelly *Tripedalia cystophora* has three lens proteins; the J1- and J3-crystallins show similarity to ADP-ribosylglycohydrolases and vertebrate saposins, respectively, whereas J2-crystallin has no apparent sequence homologs [143]. All of these proteins have in common their high stability and solubility, consistent with the necessity for lens proteins to last the lifetime of the organism. The current model of lens protein evolution is that an abundant, soluble protein is recruited to the lens, followed by gene duplication and further selection for stability and aggregation resistance, as well as the refractive index [6, 144].

The refractive index (n) describes how much the path of light is bent when traversing the boundary between one isotropic medium and another. It is defined as $n = \frac{c}{v}$, where c is the speed of light in vacuum and v is the phase velocity of light in the material of interest. For molecules in solution, an important quantity is the refractive index increment, dn/dc [145], in which c refers to solute concentration. dn/dc values are required for data analysis when performing analytical ultracentrifugation, using a refractometer to detect analytes in size-exclusion chromatography [146], and characterizing protein oligomerization via multi-angle light scattering [147]. In many cases, the approximate average value of 0.185 mL/g is used for all proteins [148]. Depending on the application, this approximation may be sufficient, but in the crystallins of the eye lens, dn/dc is generally higher than for proteins not selected for this function. A better approximation is to use the weighted average dn/dc predicted based on the amino acid composition of the protein of interest. Zhao and coworkers developed a dn/dc calculator based on the model that protein refractive index is fully explained by the amino acid composition [10]. Using this model, Mahendiran et al. investigated the effects of protein structure and primary sequence to refractive index increment. This study found

that the predicted dn/dc values for $\beta\gamma$ crystallins are much higher than for non-lens proteins with similar Greek key domain structures, a feature attributed to the higher fraction of polarizable amino acids such as arginine and methionine in the crystallins relative to other proteins [149].

Experimental measurements of dn/dc have been performed for well-characterized proteins such as bovine serum albumin (BSA) [148] and hen egg white lysozyme (HEWL) [150]. Measurements of dn/dc for bovine α-, β, and γ-crystallins have enabled rationalization of the refractive index gradient in the mammalian lens [151]. Recent experiments have shown that protein refractive index also depends on environmental factors such as solvent dielectric, ionic strength, and temperature [152]. Here we report the measured dn/dc values of several vertebrate and invertebrate crystallins. Contrary to the prevailing model, we find that for the lens crystallins the measured dn/dc values are much higher than those predicted using amino acid composition alone.

4.2 Materials and Methods

4.2.1 Protein sample preparation and dn/dc measurements

Lyophilized lysozyme from hen egg white (Cat. No. 195303) was purchased from MP Biomedicals (Solon, OH). Lysozyme was dissolved 10 mM sodium phosphate buffer, 100 mM sodium chloride, 0.05% sodium azide at pH 6.9 for a final concentration of 50 mg/mL. Human γS-crystallin [47] and *Ciona intestinalis*-$\beta\gamma$-crystallin [153] were expressed and purified as previously described. Plasmids containing the cDNA sequences for *Dissotichus mawsoni* γS1, γS2 γM8b-, γM8c-, γM8d-, and J2-crystallin were purchased from Blue Heron (Bothell, WA). All but J2-crystallin oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA); the J2-crystallin primer was purchased from Sigma-Aldrich (St. Louis,

MO). The crystallin genes were amplified with primers containing flanking restriction sites for NcoI and XhoI, an N-terminal 6x His tag, and a TEV cleavage sequence (ENLYFQG) except γS1 and γS2, which lacked an N-terminal 6x His tag, and a TEV cleavage sequence. The polymerase chain reaction product was cloned into a pET28a(+) vector, purchased from Novagen (Darmstadt, Germany). The toothfish crystallins were overexpressed in Rosetta (DE3) *Escherichia coli* using the Studier autoinduction protocol at 25°C for 24 hours. J2-crystallin was overexpressed in Rosetta (DE3) *E. coli* using standard IPTG-induced overexpression protocols at 25 °C for 18 hours. Cells were lysed by sonication and cell debris was removed by centrifugation. His-tagged crystallins were purified on an Ni-IDA column purchased from Bio-Rad (Hercules, CA) and cleaved by a His-tagged TEV protease (produced in-house). The TEV protease and His-tag were removed by a second application to an Ni-NTA column. Untagged γS1 and γS2 were dialyzed in 10 mM Tris, 0.05% sodium azide, pH 8 then purified by anion exchange on an UNOsphere Q column purchased from Bio-Rad (Hercules, CA) using a 1 M sodium chloride gradient. The final purification step for all crystallins was application to a HiLoad 16/600 Superdex 75 PG gel filtration column from GE (Pittsburgh, PA) using 10 mM sodium phosphate buffer, 100 mM sodium chloride, 0.05% sodium azide at pH 6.9. Samples for dn/dc measurements were prepared by serial dilution from a starting concentration of 50 mg/mL, measured using UV absorbance measurements at 280 nm using the extinction coefficients given in Table 1.

Refractive index increments were measured following the batch-mode technique using an Optilab rEX refractive index detector (Wyatt Technology, Santa Barbara, CA) configured with a 685 nm fiber-optic laser diode source. The instrument measures differential refraction using a flow cell where the light path first passes through the sample containing the analyte then through a reference sample. Any difference in refraction between the two solutions results in beam deflection that is detected by an array of photodiodes. The most common sources of experimental error in this type of measurement are caused by temperature fluctuations or inaccuracies in the sample concentration.

4.2.2 Refractive index calculations

The method used here to calculate protein refractive index increment (dn/dc) is adapted from the work of McMeekin and coworkers [154, 155]. This treatment also forms the basis of the dn/dc calculator published by Zhao and colleagues [10, 11].

Protein refractivity (R_P) per gram is calculated from the weight percentages of each amino acid (indexed i) applying the same method used to calculate protein partial specific volume ($\bar{v}_P$). Empirical values for refractivity [155] (R_i) and specific volume ($\bar{v}_i$) of the individual amino acids [156] are summed, as in (4.1) and (4.2), respectively.

$$R_P = \frac{\sum_i M_i R_i}{\sum_i M_i} \qquad (4.1)$$

$$\bar{v}_P = \frac{\sum_i M_i \bar{v}_i}{\sum_i M_i} \qquad (4.2)$$

The refractive index of the protein follows from a rearrangement of the Lorentz-Lorenz equation (4.3), which is itself an expanded Gladstone-Dale expression [157], yielding (4.4).

$$R_P = \bar{v}\frac{n_p^2 - 1}{n_p^2 + 2} \qquad (4.3)$$

$$n_p = \sqrt{\frac{2R_P + \bar{v}}{\bar{v} - R_P}} \qquad (4.4)$$

Assuming volume additivity, the refractive index of protein (n_p) in solution can be calculated from the Wiener equation (from Heller et. al.[158], eq. 7, eq. 17) such that (4.5)

becomes (4.6) when the sample is sufficiently dilute $\mathrm{n}_{solution} \rightarrow \mathrm{n}_{solvent}$.

$$n_p^2 = n_{solv}^2 \frac{\frac{2}{\bar{v}}\frac{dn}{dc}(n_{solv} + n_{soln}) + (n_{soln}^2 + n_{solv}^2)}{(n_{soln}^2 + n_{solv}^2) - \frac{1}{\bar{v}}\frac{dn}{dc}(n_{solv} + n_p)} \tag{4.5}$$

$$\frac{dn}{dc} = n_{solv}\frac{3\bar{v}}{2}\frac{(n_p^2 - n_{solv}^2)}{(n_p^2 + 2n_{solv}^2)} \tag{4.6}$$

Corrections for the wavelength (4.7) and temperature (4.8) were implemented by Zhao et al. based on work by Perlmann and Longsworth [159].

$$\left(\frac{dn}{dc}\right)_{\lambda} = \left(\frac{dn}{dc}\right)_{578nm}\left(0.94 + \frac{20,000\ nm^2}{\lambda^2}\right) \tag{4.7}$$

$$\frac{dn}{dc} = \left(\frac{dn}{dc}\right)\left(1 + (25 - T)\left(\frac{0.0005}{30°C}\right)\right) \tag{4.8}$$

In this implementation, an R script was written to compute predictions of protein (dn/dc) from multiple amino-acid sequence inputs. All calculations were run using $\mathrm{n}_{solvent} = 1.3340$, T = 25°C, and $\lambda = 589.3$ nm (corresponding to the wavelength used to measure the amino acid dn/dc values by McMeekin and coworkers).

4.2.3 Molecular modeling and calculation of solvent accessible surface area

Three-dimensional structural models were obtained for each protein in order investigate potential sources of the deviation from the additive model of dn/dc. Experimentally determined structures were used where they were available: structural models for hen egg white lysozyme (PDB ID: 4WG1 and 4WG7) [160], human γS-crystallin (PDB ID: 2M3T) [34], and *Ciona intestinalis* $\beta\gamma$-crystallin (PDB ID: 2BV2 [15]) were downloaded from the Protein Data Bank (PDB) [161]. For the proteins that lack empirical structures, models were calculated using the Robetta server [107], which predicts three-dimensional structure from a primary sequence input. Robetta uses comparative modeling based on solved PDB structures of similar sequence fragments, followed by all-atom refinement. J2-crystallin was omitted from the structure calculations because it has no known sequence homologs, reducing our confidence in this type of comparative modeling in its case. Solvent accessible surface areas (SASA) were computed in UCSF Chimera [162] using the MSMS package with default settings [163]. This package calculates SASA using a rolling sphere of radius 1.4 Å to approximate a water molecule.

4.2.4 π-pair refractive index correction

The 3D models described above were also used to generate a correction factor to the dn/dc calculations accounting for short-range interactions between pairs of highly polarizable residues, focally tryptophan, phenylalanine, tyrosine, histidine, and arginine. These highly polarizable residues all contain π-bonding systems, thus we refer to this term as the π-pair correction. We estimated the polarizability contribution of each residue to be its dn/dc difference from alanine ($(\frac{dn}{dc})_{Ala} = 0.167$), as this takes into account only contributions from the side chain beyond the β-carbon. The refractive index correction factor for a given protein

(CF) was determined based on the distance (d) between the polarizable side chain centroids of residues (i, j) as follows.

$$CF = \frac{\sum_i \sum_j ((\frac{dn}{dc})_i - 0.167)((\frac{dn}{dc})_j - 0.167)}{d^3}\delta \tag{4.9}$$

Upper bound distance cutoffs for π-π and cation-π interactions were taken as 7 and 6 respectively, where any interaction with arginine considered to be cation-π and all others treated as π-π. Unscaled correction factors were fit to residuals using a multiple linear regression yielding best fits of $\delta_{\pi-\pi}$ = 0.11 and $\delta_{cation-\pi}$ = -0.08.

4.3 Results and Discussion

The proteins investigated here include human γS-crystallin, γM8b-, γM8c-, γM8d-, γS1-, and γS2-crystallins from the Antarctic toothfish (*Dissotichus mawsoni*), J2-crystallin from the box jellyfish (*Tripedalia cystophora*), and $\beta\gamma$-crystallin from the tunicate *Ciona intestinalis*, as well as hen egg white lysozyme (HEWL) as a non-lens control protein. The results are shown in Fig. 4.1. Comparisons to the calculated dn/dc values are summarized in Table 4.1 and Fig. 4.2. Fig. 4.2 shows the predicted (open circle) and measured (filled circle) dn/dc values for each protein in our set. Predictions were performed using the methodology of Zhao and coworkers as described above [10]. In comparison, predicted dn/dc values were also calculated for all known proteins in the the human proteome, comprising 70,940

human proteome sequences gathered from the Uniprot database (organism 9606 and proteome up000005640). The distribution of predicted dn/dc values for the human proteome is approximated well by a Gaussian with an average dn/dc of 0.1897 mL/g and a standard deviation of 0.0037 mL/g. This range provides a benchmark against which to compare the predicted and measured values for the lens proteins.

For lysozyme, which has not been subject to selective pressure for refractivity, the measured value agrees with that calculated based on its amino acid composition, as well as with a previously measured literature value [164]. In contrast, all of the eye lens proteins investigated here have a dn/dc of 0.20 mL/g or higher. This average value is more than 2.5 standard deviations higher than the standard value of 0.185 mL/g often cited as the mean for all proteins. Furthermore, for all the lens proteins, the measured dn/dc values are much higher than the predictions (between 1-4 standard deviations), indicating that amino acid composition alone is not sufficient to explain the high refractivity of lens proteins.

Although all the crystallin proteins exhibit a difference between the predicted and measured dn/dc, the largest discrepancies are observed for the γS1-, γS2-, and γM8b- crystallins from the Antarctic toothfish, *D. mawsoni*. The discrepancy for J2-crystallin from the box jelly is nearly as large, while those for human γS-crystallin and toothfish γM8b- and γM8c-crystallin are more moderate. The differences among dn/dc values for the toothfish proteins are particularly interesting in light of the fact that *D. mawsoni* has at least has thirteen γ-crystallin paralogs [24]. This diversity may be necessary to balance the competing requirements for lens function in the Antarctic habitat of this fish: maintaining a high refractive index while also resisting freezing and cold cataract formation at -2 °C . Taken together, these results raise the question of which other features of lens proteins have evolved to increase refractivity beyond selection for a large fraction of highly polarizable amino acid residues.

In order to discover the molecular basis for the deviation from the predicted values, it is necessary to examine the assumptions made in this treatment of refractive index. In parti-

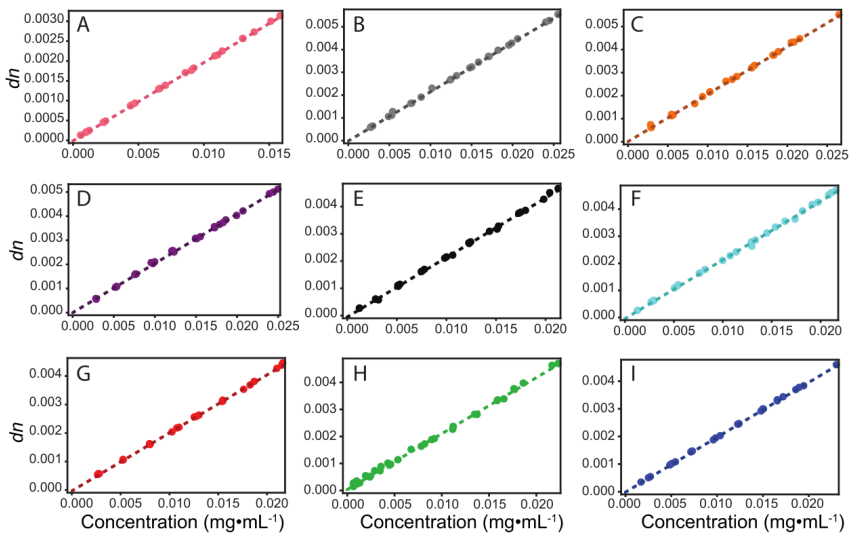

Figure 4.1: A. HEWL (control), B. γM8b-crystallin, C. γM8c-crystallin, D. γM8d-crystallin, E. γS1-crystallin, F. γS2-crystallin, G. J2-crystallin, H. human γS-crystallin, I. tunicate βγ-crystallin. Hen egg white lysozyme was measured as a control protein that has not been selected for high refractivity.

cular, the Gladstone-Dale relation assumes straightforward volume additivity and isotropic polarizability [166]. In this treatment, the refractive volume of one protein molecule in nm^3 is given by $R(M_m)/N_A$, and the refractivity of one protein molecule is given by $R = \frac{4\pi}{3} N_A \alpha$, where M_m is the molecular mass, N_A is Avogadro's number and α is the average polarizability. The fact that α is a second-rank tensor is ignored because the sample in question is an isotropic solution where all orientations of these small globular proteins are assumed to

Table 4.1

Protein	Organism	Calculated dn/dc (mL/g)	Measured dn/dc (mL/g)	Standard Deviations Between Calculated and Measured dn/dc	Extinction coefficient (mL/mg)
lysozyme	*G. gallus*	0.1963	0.1970 ± 0.0010	0.189	2.64
γS	*H. sapiens*	0.1985	0.2073 ± 0.0014	2.38	1.94
βγ	*C. intestinalis*	0.1917	0.1985 ± 0.0012	1.84	1.54
γM8b	*D. mawsoni*	0.2003	0.2158 ± 0.0015	4.19	1.06
γM8c	*D. mawsoni*	0.2003	0.2061 ± 0.0014	1.57	0.957
γM8d	*D. mawsoni*	0.1995	0.2041 ± 0.0014	1.24	1.03
γS1	*D. mawsoni*	0.2020	0.2183 ± 0.0014	4.41	2.15
γS2	*D. mawsoni*	0.2002	0.2168 ± 0.0014	4.49	2.31
J2	*T. cystophora*	0.1920	0.2037 ± 0.0012	3.16	0.283

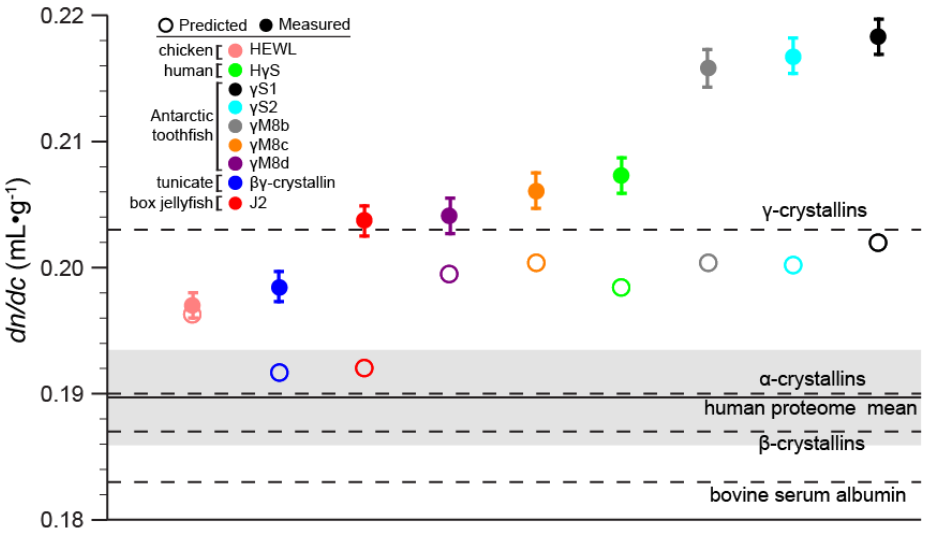

Figure 4.2: The dn/dc of HEWL, human γS-crystallin, toothfish γS1-, γS2-, γM8b-, γM8c-, and γM8d-crystallins, box jelly J2-crystallin and tunicate $\beta\gamma$-crystallin were measured and compared to their predicted values, represented by filled and open circles respectively. The solid line represents the mean dn/dc of the human proteome, with the shaded region representing one standard deviation from the mean. The dashed lines indicate the literature dn/dc values for bovine serum albumin [165], and α-, β, and γ-crystallin fractions from bovine eye lens [151].

be equally represented in the ensemble. The deviation of our measured dn/dc data from the additive model could be explained if the assumption that the protein partial specific volume is equal to the sum of that for its component amino acids is incorrect, or if the anisotropic polarizability of the individual amino acids (or small groups of them) is not negligible in the context of a folded protein.

Volume additivity could play a major role in principle, as the effective volume of an amino acid in solution is affected by hydration, and thus in solution the volume of a compactly folded protein is smaller than that of its component amino acids. However, in practice volume additivity per se does not appear to be the dominant effect, because the additive Gladstone-Dale model works well for many typical proteins, including HEWL, the control protein in this set. This may be because the effect would largely be expected to impact aliphatic amino acid sidechains, which are generally buried inside the protein interior, but have low polarizability and thus contribute little to the refractive index. For surface-exposed

residues, previous experimental studies have found that at room temperature and below, hydration of charged and non-polar sidechains is similar to that observed in small model compounds (e.g. isolated amino acids), while significant deviations are observed for the hydration of polar neutral sidechains [167].

Water molecules bound to the surface of a protein behave differently from those solvating small polar molecules because each water molecule can form hydrogen bonds with multiple polar groups on the protein surface, resulting in a network of ordered solvation water that influences the compressibility more than the specific volume [168, 169, 170, 171]. These hydrogen bond networks line up water molecules, albeit transiently, impacting the electric field at the protein surface. The water layer around a biomolecule is dynamic and heterogeneous, and exact degree of ordering and the timescale of "bound" water dynamics remain controversial [172]. However, it is reasonable to expect protein hydration to have a non-negligible impact on refractive index. Subtle effects of the solution composition have previously been shown to impact refractivity, for example the measured dn/dc of lysozyme depends on the buffer compound used [173].

Examination of the structural models for our protein set revealed a positive correlation between the solvent-exposed surface area (SASA) of hydroxyl groups and the deviation from predicted dn/dc values (Fig. 4.3). Hydroxyl groups are highly polar and can both donate and accept hydrogen bonds, contributing to the formation of the water network at the protein surface. This correlation supports the hypothesis that protein hydration plays a role in the dn/dc discrepancies we observe in lens proteins and provides a rationale for future experimental studies of their hydration properties. Prior experiments have suggested that the hydration shells of eye lens crystallin proteins are particularly robust, highlighting a possible mechanism of selection for high refractivity as well as enhanced solubility [62].

Another factor that may influence the protein dn/dc is that the polarizabilities of individual amino acids in the context of a folded protein need not be isotropic; residues may interact

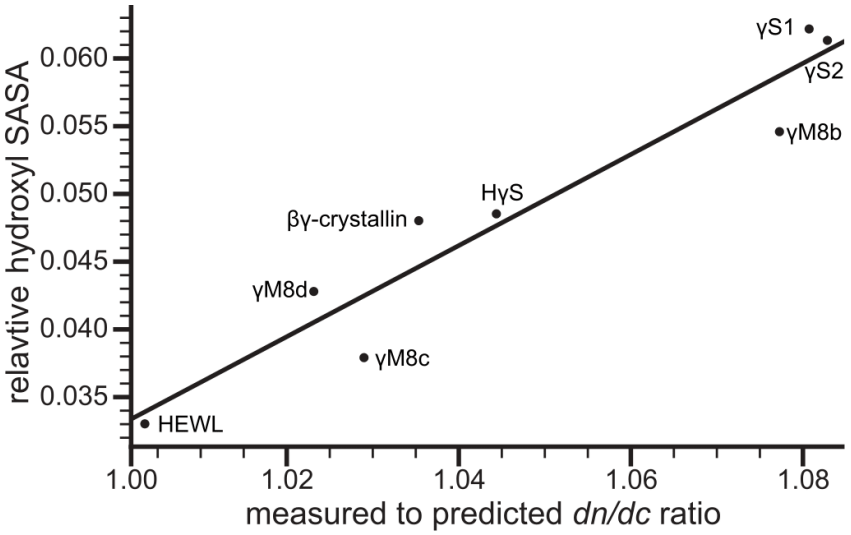

Figure 4.3: The relationship was fit to a linear regression that follows the form $S_{hyd} = 0.336x - 0.304$ with an R^2 of 0.906, in which S_{hyd} and x are the fraction of hydroxyl SASA and measured to predicted dn/dc ratio respectively.

with their neighbors to produce local (much smaller than the wavelength of light) regions of larger polarizability. Some of the most polarizable amino acid side chains, e.g. Trp, Phe, Tyr, His, and Arg, are also highly anisoptropic in shape, raising the possibility of highly specific interactions held in place by the packing of the protein interior. Surface residues may also adopt particular conformations via strong interactions such as salt bridges or cation-π interactions. These effects are probably not independent, as changes in the compressibility of the molecule also influence its polarizability [174]. To approximate the effect of polarizable amino acid interactions, we applied a correction factor to the additive dn/dc model from Zhao et al. In this correction, applying a small positive and negative weight for π-π and cation-π interactions, respectively, improves prediction accuracy (Fig. 4.4). More experimental data is needed to develop a more complete theory of protein refractive index in lens proteins. In particular, this π-pair correction may be refined using the detailed angular information about the relative orientations of key side chains that is only available from high-resolution structures, while experimental measurements of the hydration shell mobility will help to clarify the role of surface hydration.

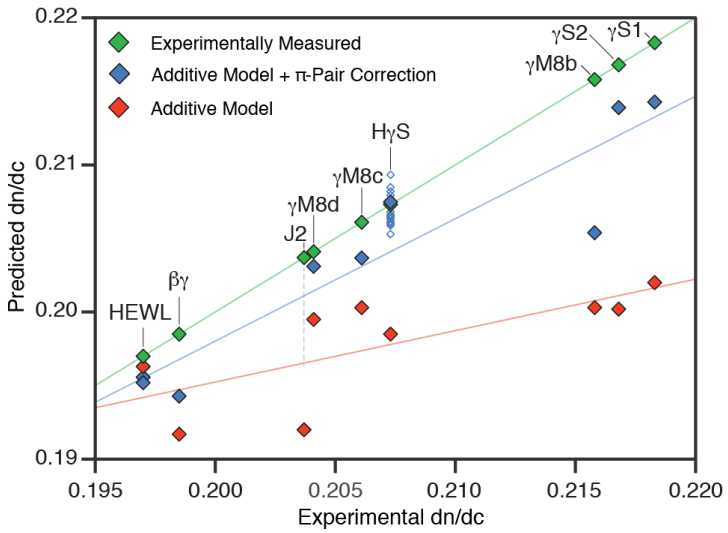

Figure 4.4: Experimental dn/dc values (green), values predicted from the additive model of Zhao et al (red) and additive model values plus the π-pair correction (blue) are plotted as a function of the experimental dn/dc. Corrected predictions are shown as filled diamonds for the lowest energy structure and empty diamonds for alternate confirmations where they are available. Two filled diamonds are shown representing the two lysozyme crystal structures, while no predictions are shown for J2-crystallin, as no previously solved structures were sufficiently similar for confident structural modeling. Additional unfilled diamonds are shown for human γS-crystallin to represent alternate low energy NMR conformations. Regression lines are shown as visual guides for model comparison.

In summary, the measured refractive index values for all the lens crystallins investigated here is higher than that predicted using the prevailing model, indicating that factors other than amino acid composition are involved in producing the high refractivity of lens proteins. The difference is particularly striking for γS1-, γS2-, and γM8b-crystallin from the Antarctic toothfish. We propose two hypotheses for the origin of this effect, which may work independently or in concert. The arrangement of hydroxyl groups on the surface may affect the protein hydration structure and hence the dipole moment and polarizability. Alternatively, the effect of short-range interactions between highly polarizable amino acids may cause local regions of anisotropic polarizability. We propose a correction to the additive model based on the idea that π-π interactions contribute more to the refractive index than the sum of the amino acid polarizabilities. For the proteins studied here, this π-pair model improves agreement between predicted and measured values, although further refinement is needed.

Structure determination efforts for these proteins and experiments probing their hydration shells are expected to provide further insight into the mechanisms underlying the high refractive indices of eye lens proteins. Other remaining questions include the behavior of these proteins at higher concentrations, where self-organization becomes important, as well as the concentration gradient and differential distribution of these proteins in the lens.

Chapter 5

Characterizing D1 Plant-Specific Insert

5.1 Background

Carnivory has independently evolved in the plant kingdom multiple times as a way of acquiring more nutrients [175]. The hypothesis is that these plants evolved the ability to capture and digest prey to supplement a nutrient deficiency, such as available soil nitrogen in wetland type environments [176]. In most cases, the leaves are, in addition to photosynthesis, recruited for capturing prey. Often, this change in function results in loss of photosynthetic capacity of the leaves, in addition to increased energy demands for capturing and digesting prey. Different plants utilize different methods to capture prey ,such as pitfall traps and fly paper traps among others [175, 177]. Once prey has been captured, there are several challenges the plants face for digestion. All of these plants must perform their digestion without any mechanical breakdown of their prey, such as mastication. In addition, digestion often occurs in an exposed environment over a prolonged period of time, meaning the digestion is exposed to variable physical conditions from changes in weather. Exposure of the digestion process also increases the risk of opportunistic microbial growth that compete for

nutrients from the captured prey and possibly infecting the plant tissue. These challenges have lead to the evolution of interesting traits in the proteins of carnivorous plants to meet their environmental demands [178, 179].

For these reasons, carnivorous plants are a potential source of new and interesting digestive enzymes and antimicrobial proteins. These new proteins could be useful in a variety of biotechnological application from proteases that more readily breakdown protease resistant proteins, chitinases that can be repurposed as an insceticide or fungicide, or antimicrobial proteins to inhibit microbial growth for medical uses. One potentially interesting carnivorous plant is *Drosera capensis*, also known as the Cape sundew. This plant is native to the Cape region of South Africa and is relatively easily cultivated. This plant catches prey in the sticky mucilage of its leaf tentacles. It then wraps its leaf around the captured prey to increase contact with the digestive mucilage and enhance nutrient absorption [180]. In order to explore the potentially interesting proteins that have evolved in this plant, we sequenced the genome of *D. capensis*, representing the first carnivorous genome sequenced from the order Caryophyllales [178].

With a whole genome worth of protein sequences from *D. capensis*, proteins that could be involved in digestion were prioritized and discovered by homology. During analysis, aspartic proteases were discovered that contain what is called a plant specific insert (PSI). These PSIs are domains of usually 50-100 amino acids found in plant aspartic proteases that upon maturation are cleaved off and act as independent proteins [178]. PSIs are known as saposin-like proteins, which consists of a family of proteins that have a conserved tertiary fold often containing three disulfide bonds, but can vary in sequence identity [181]. 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 the membrane [182]. Permeabilizing the membrane can be used as an antimicrobial defense to induce cytotoxicity. For example, the well-studied PSI from the

potato *Solanum tuberosum*, which has been shown to exhibit antimicrobial activity against various plant and human pathogens [183]. Similarly, we hypothesize that PSI-containing aspartic proteases from *D. capensis* could have been recruited for digestive function for both its ability to digest prey proteins as well as inhibit microbial growth during digestion. Alternatively, a PSI in the digestive mucilage could function to make insect lipids more available for digestion. Furthermore, PSIs alone could have application in microbial growth inhibition. To this end, we have biochemically and biophysically studied the properties of the Droserasin 1 PSI (D1-PSI) from *D. capensis*. Our results have shown that indeed the D1-PSI is capable of antimicrobial activity and disrupting membranes.

5.2 Materials and Methods

5.2.1 Gene construction, expression, and purification

Plasmids containing the DNA sequence of the *D. capensis* Droserasin 1 PSI (D1-PSI) genes were purchased from Intergrated DNA Technologies (San Diego, CA). Each gene was flanked by regions containing restriction sites for NcoI and XhoI and contained an N-terminal 6x His tag, and a TEV cleavage sequence (ENLYFQG). The gene was amplified using oligonucleotide primers purchased from Intergrated DNA Technologies (Coralville, IA), 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) using autoinduction protocols [102] at 25 °C with the modification of adding 50 μM IPTG [184]. Cells were allowed to grow for at least 24 hours, lysed via sonication, heated at 70 °C for 20 minutes 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) 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, followed by reapplication to Ni-charged Bio-Scale Mini Profinity IMAC Cartridges (Bio-Rad, Hercules, CA) 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) 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 (Fig A.1). For ^{13}C, ^{15}N labeled protein samples, protein was expressed using an optimized high-cell-density IPTG-induction minimal media protocol [63]. Purification was performed in the same manner as unlabeled protein.

5.2.2 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 at different temperatures. Spectra were also taken under the same condition but D1-PSI was reduced in 1 mM DTT before dilution into final buffer. Measurements were taken on a J-810 spectropolarimeter (JASCO, Easton, MD) equipped with a thermal controller.

5.2.3 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, exciting at 280 nm. Spectra were taken using a Cary Eclipse Fluorescence Spectrophotometer (Agilent, Santa Clara, CA)

5.2.4 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. 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) at 600 nm of 0.005 and were allowed to grow for 48 hours at 30 °C. After growth total cell yield was estimated by measuring the OD at 600 nm.

5.2.5 Vesicle fusion assay

Lipids used for experiments were extracted from either *E. coli* or *P. pastoris* using the Bligh and Dyer method [185]. The lipids were then dried under steam of nitrogen gas and polar lipids were extracted using a cold acetone precipitation [186]. 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 an mini extruder equipped with a 100 nm polycarbonate membrane (Avanti Polar Lipds, Alablaster, AL) to create large unilamellar veicles (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.).

5.2.6 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 (Alablaster, AL). 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). 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 [185]. A series of ten fold dilutions were made for each and were run on a Xevo G2-XS QTof spectrometer (Waters, Milford, MA) in positive mode, with an in-line ACUITY UPLC BEH C18 Column (Waters, Milford, MA), where lipids were eluted with a water/isopropanol gradient containing ammonium formate. Only sample concentrations for 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.

5.2.7 Solid-state NMR

Samples were prepared by evaporating away chloroform from 5 mg of POPC and POPA, respectively under a stream of nitrogen. 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 lyopholized. Lypholized 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) using the 40 mm bore series connected hybrid magnet system, currently the largest field NMR magnet [187]. Two dimensional ^{13}C-^{13}C cross polarization [188] dipolar assisted rotational resonance

[189] (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, and a spinning speed of 24.4 kHz [187].

5.3 Results and Discussion

5.3.1 *D. capensis* D1-PSI is highly stable

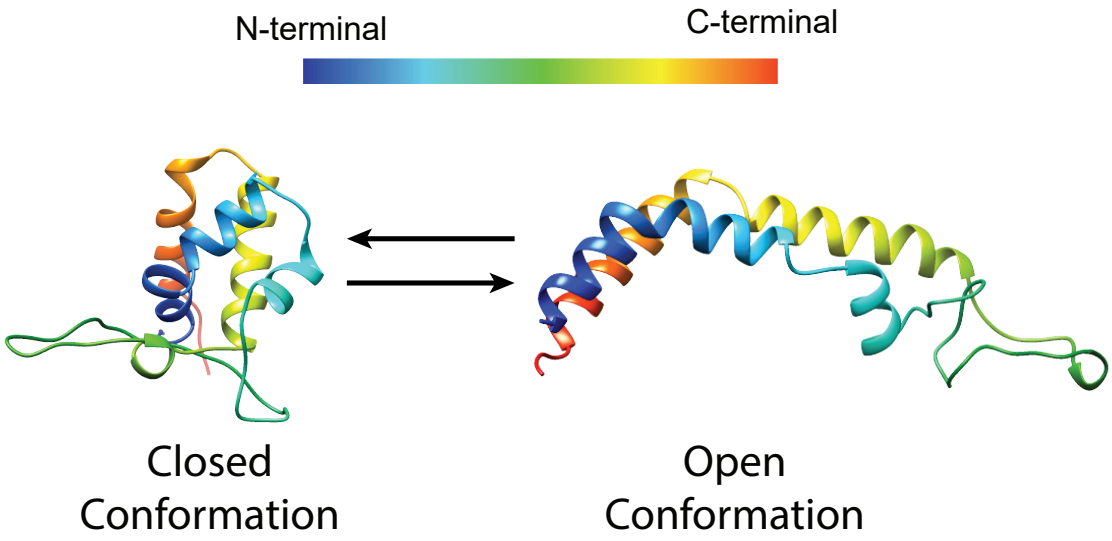

Figure 5.1: Comparative model of mature D1-PSI were predicted using Robetta server [107]. Predictions resulted in models in both the closed and open conformations.

Saposin-like proteins are known to be very stable proteins in part due to the presence of disulfide bonds that lock the tertiary structure into place. This tertiary structure results in a fold is that conserved among saposin-like proteins as seen in Fig. 5.1 [190]. It is thought that saposin-like proteins can exist in two different conformations, the closed and open conformation. The open conformation is proposed to be responsible for membrane interacting activity due the exposure of hydrophobic residues compared to the closed conformation [191]. To gain structural insight and test the stability of D1-PSI, CD spectra were taken under different pH and temperature conditions. Regardless of the pH, there is very little

variance in the CD spectra taken at 20, 55, or 90 °C as seen in Fig. 5.2, 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. The degree to which the tertiary stucture is perturbed under the conditions tested would require data from other experimental techniques such as structural determination or scattering techniques.

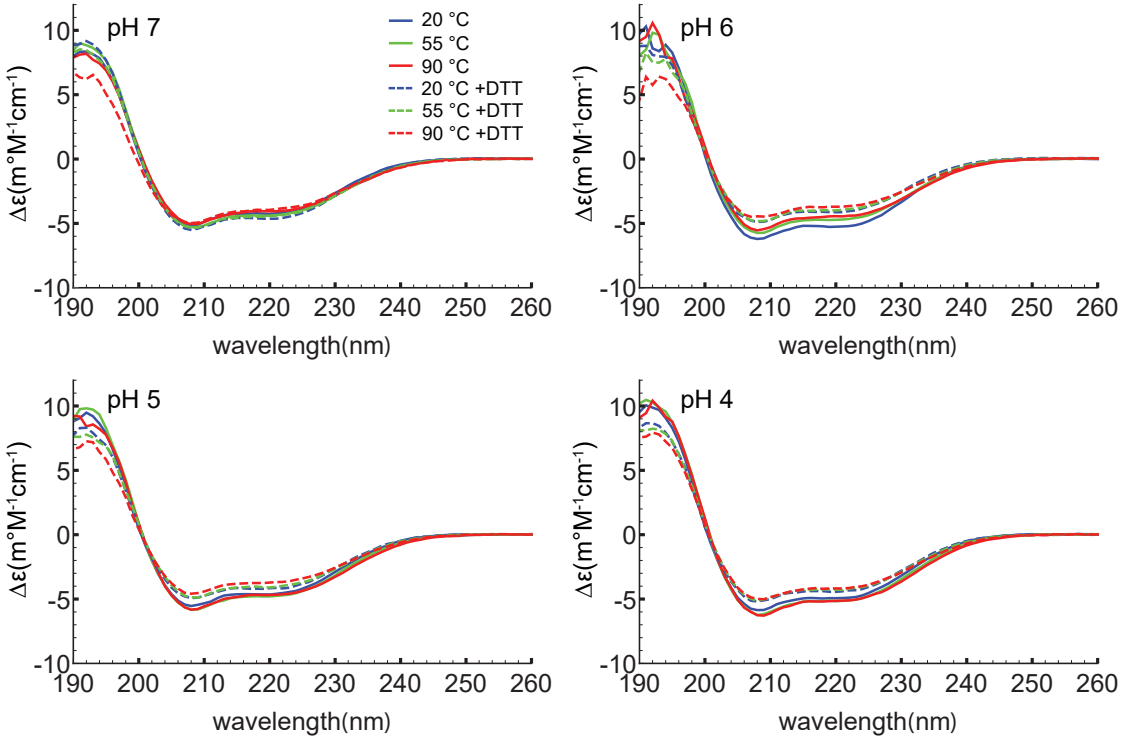

Figure 5.2: Circular dichroism spectra were taken of D1-PSI at pH 4, 5, 6, and 7. Spectra was also taken at 20, 55 and 90 °C with or with out DTT. Spectra indicate that the secondary structure of the protein is highly stable under all condition tested. Reduction of disulfide bond by DTT seems to result in only a small change in the secondary structure.

To further probe the response of the D1-PSI to pH, temperature, and reducing agent, intrinsic tryptophan fluorescence spectroscopy was employed. 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. The intrinsic fluorescence was probed under

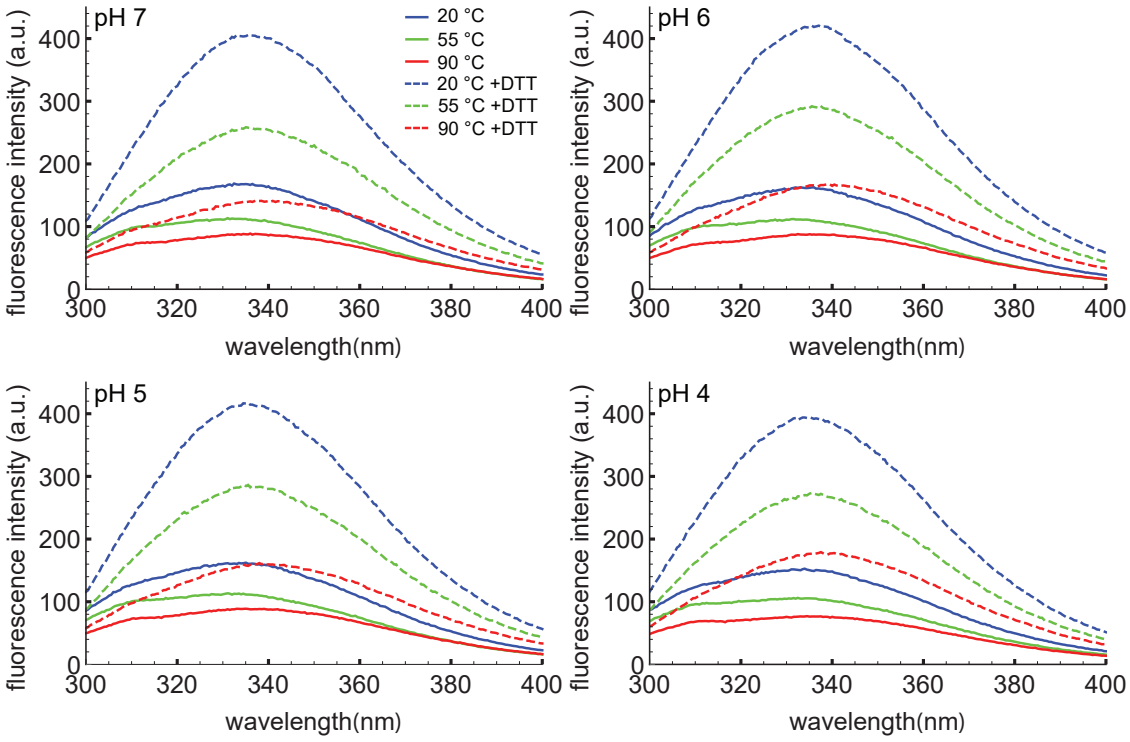

Figure 5.3: Intrinsic tryptophan fluorescence spectra was taken of D1-PSI at pH 4, 5, 6, and 7. Spectra were also taken at 20, 55 and 90 °C with or without out DTT. The excitation wavelength was 280 nm. Under non-reducing conditions the emission maxima are at approximately 334 nm regardless of pH or temperature. While under reducing conditions the emission maximum for most spectra shifts 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 reductions of a nearby disulfide bond that was quenching the tryptophan.

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 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 [192]. 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 the tryptophan is next to a disulfide bond that is capable of quenching it, while once the disulfide is reduced it would release the quenching [193]. This is also supported by the predicted models of the D1-PSI that shows the tryptophan being next to a disulfide bond either in the closed conformation or in a dimerized closed conformation.

5.3.2 D1-PSI is capable of inhibiting microbial growth

Based on the work of Munoz et. al. [183] where a potato PSI was shown to display antimicrobial activity, we hypothesized the D1-PSI could serve a similar function in *D. capensis*. To test this, *E. coli* was grown with or without 25 μM D1-PSI present at four different pH conditions. At pH 7 and 6 there was no difference seen between the treated and untreated final cell yield, while at pH 5 and 4 there was virtually no cell growth in either condition (data not shown). We then tested the growth of the yeast *P. pastoris* at different pH with or without D1-PSI. Again at pH 6 and 7 there was nearly no difference between the treated and untreated cells, similar to what was observed in *E. coli* (Fig. 5.4). However, at pH 5 there was a decrease in cell yield of the treated samples relative to the untreated samples, indicating that the D1-PSI does indeed display antimicrobial activity with a pH dependence. The observation that D1-PSI only shows activity at acidic pH further supports the idea that this protein could be utilized in the digestive mucilage of *D. capensis*, which typically has a pH of approximately 5 [178]. Unfortunately, there was not sufficient yeast cell growth at

pH 4 to further test the pH dependence of the D1-PSI. Testing the antimicrobial activity against other organisms, especially acid tolerant organisms, is an avenue worth exploring to further characterize both pH dependence of activity and the generalizability of the D1-PSI antimicrobial activity against other microbes.

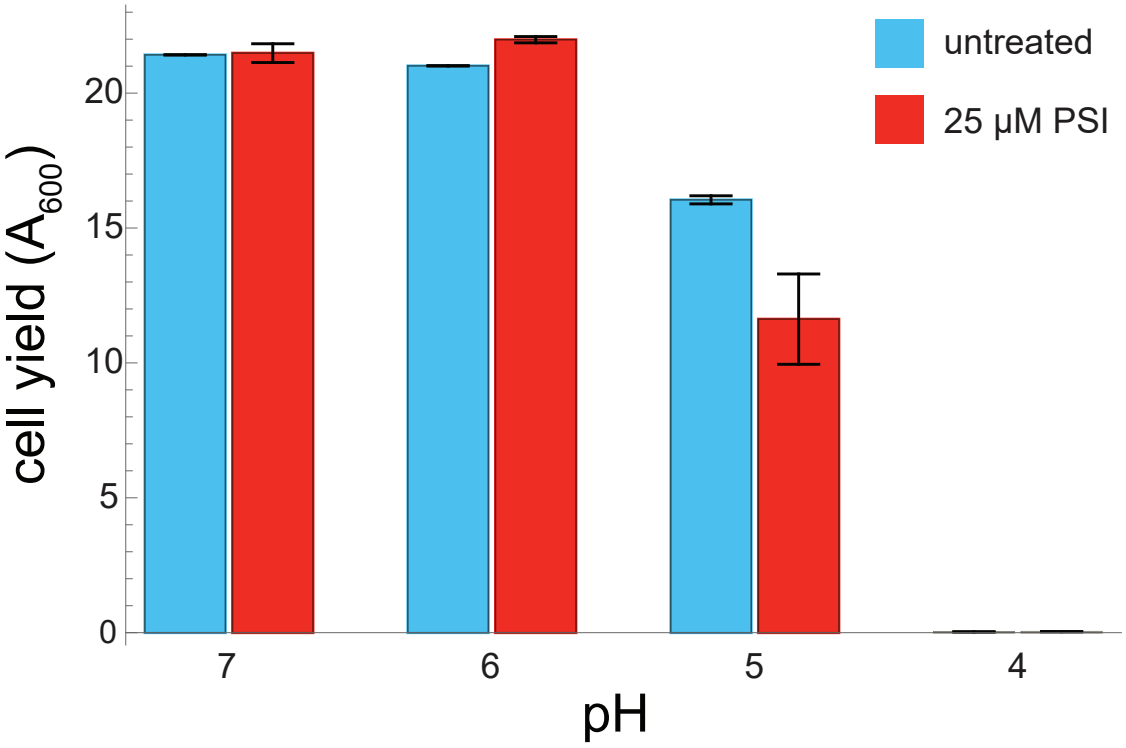

Figure 5.4: The yeast *P. pastoris* was grown at either pH 4, 5, 6, or 7 with or without 25 μM D1-PSI. At close to neutral pH of 6 and 7 there was little to no difference between the treated and untreated cell yield. At the more acidic pH of 5, treated cells showed a reduce overall cell yield, indicating antimicrobial activity of D1-SI. At pH 4, cells did not grow to a measurable amount.

5.3.3 D1-PSI enables vesicle fusion at acidic pH

Based on the results from the antimicrobial assay, we hypothesize that the mechanism of action involves permeablizing the membrane to inhibit cell growth. To test how D1-PSI interacts with membranes, a vesicle fusion assay was utilized. In this study DLS is used to monitor increasing vesicle size due to fusion from membrane disruption. First 100 nm LUVs

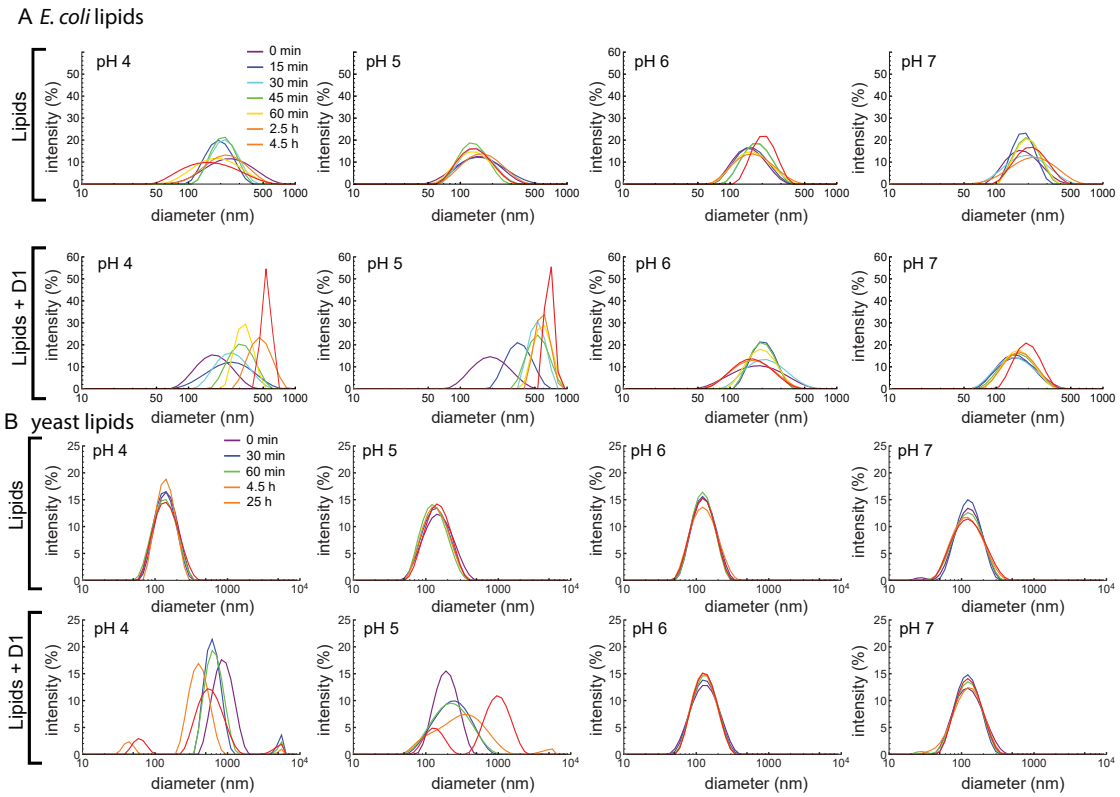

Figure 5.5: **(A)** LUVs of 100 nm were made using an *E. coli* polar lipid extract. LUVs were made in buffered solutions at pH 4, 5, 6, and 7 and size was monitored over time, either with(bottom) or without(top) D1-PSI, using DLS. LUVs are stable in size for all pH values, but upon addition of D1-PSI they fuse into larger vesicles at acidic pH. **(B)** LUVs of 100 nm were made using an yeast polar lipid extract. LUVs were made in buffered solutions at pH 4, 5, 6, and 7 and size was monitored over time, either without(top) or with(bottom) D1-PSI, using DLS. Again, LUVs are stable in size for all pH values, but upon addition of D1-PSI they fuse into larger vesicles at acidic pH.

were prepared using *E. coli* polar lipid extract. At all the pH conditions tested, LUVs were stable in size over time as seen in Fig. 5.5A. Upon addition of D1-PSI the distribution of LUV size begin to shift to larger sizes, but only at pH 4 and 5, while pH 6 and 7 remain stable like the control. DLS of the D1-PSI by itself was also recorded at each pH, where there was only one size peak at about 2 nm that was stable over time(data not shown). When LUVs were made using yeast polar lipid extract instead, again the LUVs are stable over time at each pH (Fig. 5.5B). When D1-PSI is added to the LUVs, again only at pH 4 and 5 does the size distribution increase, just like was seen in with the *E. coli* polar lipids. These results 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 gradual while in the case of the yeast polar lipids the change was very rapid at pH 4, suggesting that the kinetics of this process are accelerated with the yeast polar lipids as compared to the *E. coli* polar lipids. 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.

5.3.4 D1-PSI lacks head group specificity for incorporation into lipoprotein complex

After showing that D1-PSI is able to interact with membrane lipids from a natural source that contains a variety of lipid species, we sought to gain insight into the lipid specificiy of D1-PSI. First, the specificity D1-PSI displays towards different head groups of phospholipids was tested. Five phospholipid species containing different head groups, but the same acyl groups of a singly unsaturated 18-carbon acyl chain and saturated 16-carbon acyl chain,

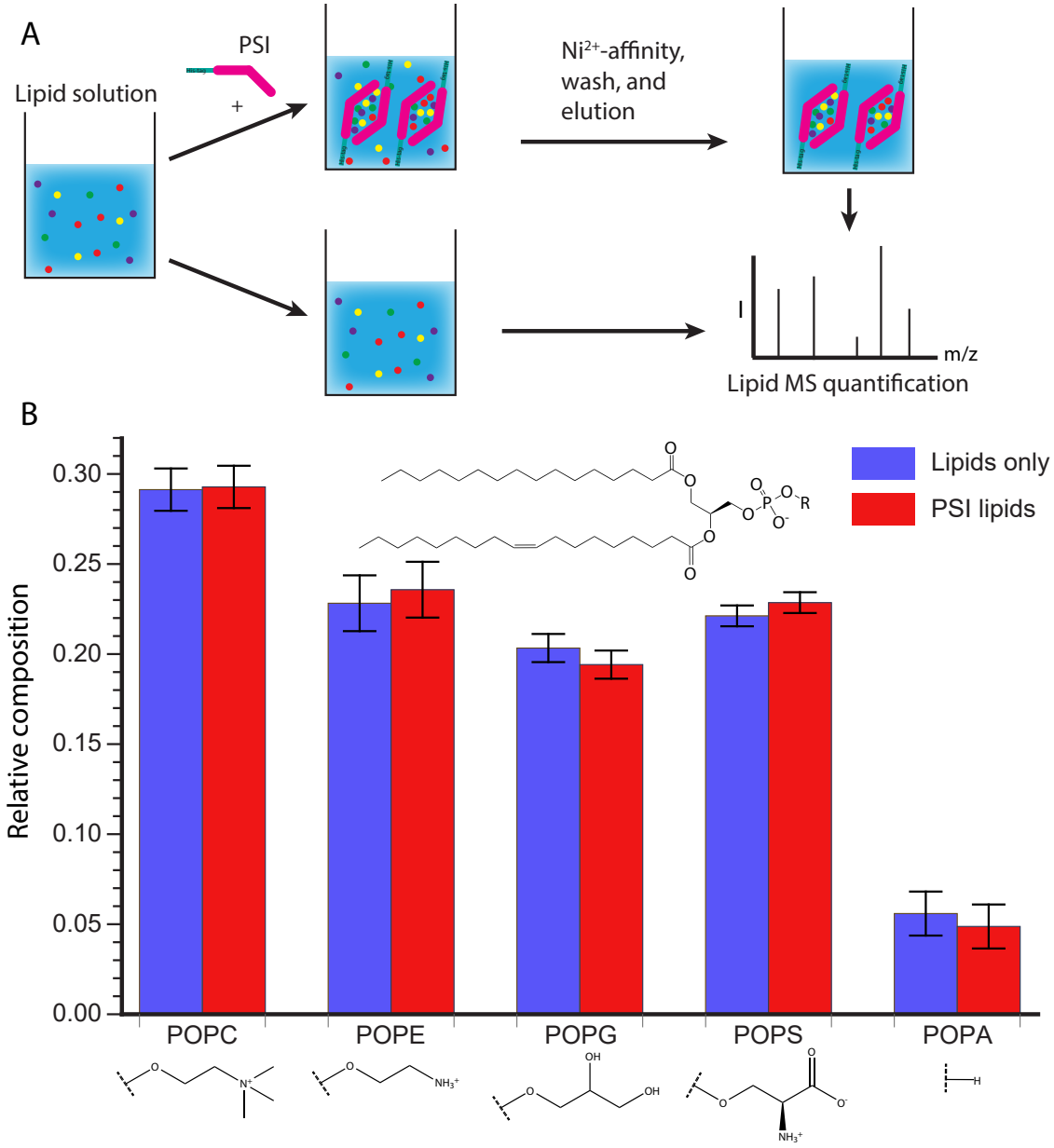

Figure 5.6: **(A)**Schematic representation of the experimental procedure. A lipid solution containing POPC, POPE, POPG, POPS, and POPA was prepared and split into two aliquots. One aliquot was treated with His tagged D1-PSI and allowed to mix before purifying the His tagged PSI and bound lipids from 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.

were tested for their ability to associate with the D1-PSI. A solution containing both the neutrally-charged POPC and POPE as well as negatively-charged POPG, POPS, and POPA was made. Part 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, while any unbound lipids could be washed away. The PSI was eluted and any lipids that associated with it were quantified by MS and compared to the quantification of the original lipid solution (Fig. 5.6A). As can be seen in Fig. 5.6B there is no significant difference between the relative lipid composition between the original lipid solution and the lipids that co-purified with the PSI. This result suggests there is no bias due to acyl chain composition for lipid interaction in the final stable protein-lipid complex. However, it is possible the kinetics of association may be different depending on the lipid composition as hinted with the vesicle fusion assay (Fig. 5.5). During membrane association the PSI must first interact with the head groups, before interacting with the acyl chains buried in the membrane. It is not unreasonable to hypothesize that specific head groups more strongly interact with the PSI which promotes initial association, followed by contact with the acyl chains at which point PSI-acyl chain interactions predominate. This could be tested with a vesicle fusion or leakage assay, where vesicles are composed of different lipid species.

5.3.5 D1-PSI is amenable to structure determination by solid state NMR

Knowing that D1-PSI interacts with membranes, we next wanted to investigate the change in protein conformation upon interaction with the membrane lipids. For high-resolution structural data, X-ray diffraction or NMR are the most common methods utilized. Crystallizing proteins in a membrane system is notoriously challenging and would likely be a tedious endeavor [194]. Solution-state NMR does not require growing a crystal and instead

structure determination experiments can be performed under more native-like conditions. The drawback of solution-state NMR is that proteins must be relatively small (<50 kDa) to ensure that the protein is tumbling fast enough to average out the chemical shift anisotropy and dipolar couplings [195]. In the case of D1-PSI this is not a problem with a monomeric molecular weight of 12 kDa. However, as the goal is to obtain structural information while interacting with a membrane, the particles in solution would be much larger and become inaccessible by solution state NMR. One way to overcome this problem for membrane proteins is by using nanodiscs. Nanodiscs are composed of a small section of a membrane that is encircled by a membrane scaffold protein [196]. While this method has been successfully employed for some membrane proteins, using this system for the D1-PSI could be potentially problematic. One of the potential problems is that when the PSI is interacting with the membrane it unknown whether it oligomerizes as is the case for some pore forming membrane proteins [197]. This could lead to the nanodicscs becomeing too big or the nanodiscs lacking enough space to accommodate an oligomerized form of D1-PSI. Another potential problem, highlighted by the vesicle fusion assay, is that D1-PSI appears to disrupt membranes. If the PSI disrupts the nanodisc membrane it could potentially lead to nanodiscs interacting with each other in a way that perturbs the uniform size expected from the nanodisc, putting them outside the size that is acceptable for solution state-NMR.

An alternate method is using solid-state NMR (ssNMR) to solve the structure of D1-PSI while interacting with a membrane. Since 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. This allows for the use of vesicles or bicelles as a membrane system to study membrane proteins [198]. For this reason we have decided to make use of ssNMR. To do this ^{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 lab to obtain spectra. The ^{13}C-^{13}C CPDARR spectrum (Fig. 5.7) 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. These preliminary spectra indicate that this protein is amenable to structure determination by ssNMR. More experiments, including 3D experiments, will be needed for 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.

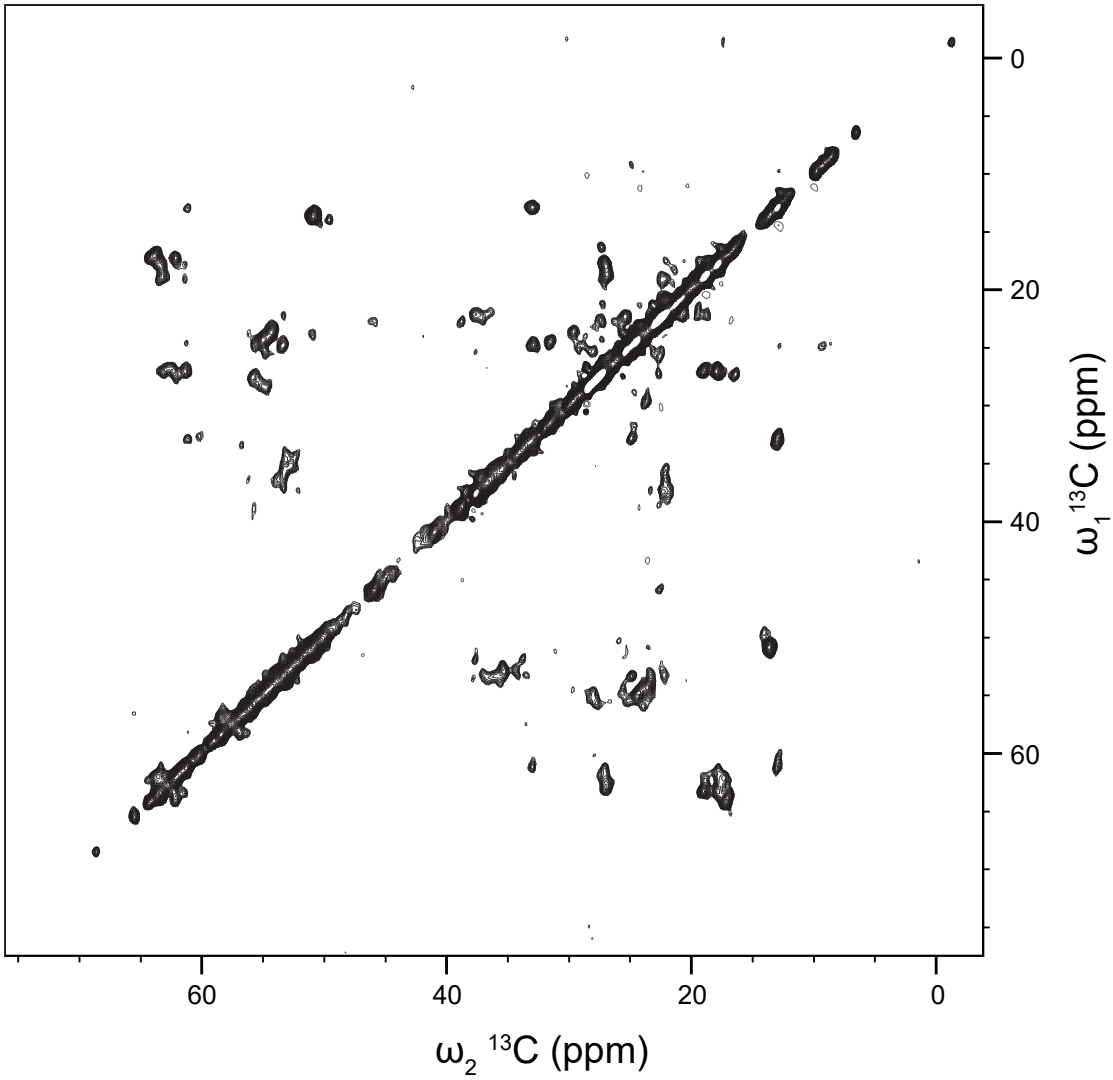

Figure 5.7: ^{13}C-^{13}C CPDARR spectrum of D1-PSI in a 1:1 POPC, POPA membrane system was taken at 10 °C and spinning at 24.4 kHz. The spectrum displays well-resolved off-diagonal peaks of reasonable intensity indicating that the PSI is suitable for ssNMR structure determination.

Chapter 6

Conclusion

This thesis work covers a diverse array of protein systems to better understand how they fundamentally function. Studying the eye lens proteins has provided insight into how they perform their function of remaining stable and soluble while resisting phase transition. This is especially relevant in the case of the Antarctic toothfish, where phase transitions is a particularly pressing challenge. Furthermore, by measuring the functionally relevant refractive indices of individual eye lens proteins, it was shown that refractive index is determined by more than amino acid composition. Protein conformation also appears to be important through yet to be explored mechanisms that cause the measured refractivities to deviate from predictions. Additionally, the D1-PSI from the carnivorous Cape sundew was expressed for the first time, characterized and shown to be able to interact with membranes and exert antimicrobial activity. The next step is to solve the solid state NMR structure while the PSI is interacting with a membrane system to better understand the structural consequences of membrane interaction.

Bibliography

[1] K. Teilum, B. B. Olsen, B. B. Kragelund, Protein stability, flexibility and function, Biochimica et Biophysica Acta 1814 (2011) 969–976.

[2] R. Jaenicke, G. G. Bohm, The stability of proteins in extreme environments, Current Opinion in Structural Biology 8 (1998) 738—748.

[3] M. Delaye, A. Tardieu, Short-range order of crystallin proteins accounts for eye lens transparency, Nature 302 (1983) 425–417.

[4] R. H. Kroger, M. C. Campbell, R. Munger, R. D. Fernald, Refractive index distribution and spherical aberration in the crystalline lens of the african cichlid fish haplochromis burtoni, Vision Research 34 (14) (1994) 1815–1822.

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

[6] C. Slingsby, G. J. Wistow, A. R. Clark, Evolution of crystallins for a role in the vertebrate eye lens, Protein Science 22 (4) (2013) 367–380.

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

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

[9] J. Piatigorsky, G. J. Wistow, Enzyme crystallins - gene sharing as an evolutionary strategy, Cell 57 (2) (1989) 197–199.

[10] H. Zhao, P. H. Brown, P. Schuck, On the distribution of protein refractive index increments, Biophysical Journal 100 (2011) 2309–2317.

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

[12] P. Aravind, A. Mishra, S. K. Suman, M. K. Jobby, R. Sankaranarayanan, Y. Sharma, The $\beta\gamma$-crystallin superfamily contains a universal motif for binding calcium, Biochemistry 48 (51) (2009) 12180–12190.

[13] R. Barnwal, M. K. Jobby, K. M. Devi, Y. Sharma, K. V. Chary, Solution structure and calcium-binding properties of m-crystallin, a primordial betagamma-crystallin from archaea, Journal of Molecular Biology 386 (2009) 675–689.

[14] N. J. Clout, M. Kretschmar, R. Jaenicke, C. Slingsby, Crystal structure of the calcium-loaded spherulin 3a dimer sheds light on the evolution of the eye lens betagamma-crystallin domain fold, Structure 9 (2) (2001) 115–124.

[15] S. M. Shimeld, A. G. Purkiss, R. P. Dirks, O. A. Bateman, C. Slingsby, N. H. Lubsen, Urochordate betagamma-crystallin and the evolutionary origin of the vertebrate eye lens, Current Biology 15 (2005) 1684–1689.

[16] S. K. Suman, A. Mishra, L. Yeramala, I. Das Rastogi, Y. Sharma, Disability for function: Loss of ca^{2+}-binding is obligatory for fitness of mammalian $\beta\gamma$-crystallins, Biochemistry 52 (50) (2013) 9047–9058.

[17] B. Mahler, Y. Chen, J. Ford, C. Thiel, G. Wistow, Z. Zu, Structure and dynamics of the fish eye lens protein, γM7-crystallin, Biochemistry 52 (2013) 3579–3587.

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

[19] R. J. Siezen, M. R. Fisch, C. Slingsby, G. B. Benedek, Opacification of gamma-crystallin solutions from calf lens in relation to cold cataract formation, Proceedings of the National Academy of Science U S A 82 (1985) 1701–1705.

[20] Y. Chen, H. Zhao, P. Schuck, G. Wistow, Solution properties of γ-crystallins: Compact structure and low frictional ratio are conserved properties of diverse γ-crystallins, Protein Science 23 (1) (2014) 76–87.

[21] C. R. Cantor, P. R. Schimmel, Biophysical Chemistry, Part III: The behavior of biological macromolecules, W. H. Freeman and Company, New York, 1980.

[22] J. K. Myers, C. N. Pace, J. M. Scholtz, Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding, Protein Science 4 (10) (1995) 2138—2148.

[23] C. N. Pace, Conformational stability of globular proteins, Trends in Biochemical Sciences 15 (1) (1990) 14—17.

[24] A. J. Kiss, A. Y. Mirarefi, S. Ramakrishnan, C. F. Zukoski, A. L. DeVries, C.-H. C. Cheng, Cold-stable eye lens crystallins of the Antarctic nototheniid toothfish *Dissostichus mawsoni* Norman, Journal of Experimental Biology 207 (2004) 4633–4649.

[25] G. Feller, D. A'mico, C. Gerday, Thermodynamic stability of a cold-active α-amylase from the antarctic bacterium *Alteromonas haloplanctis*, Biochemistry 38 (1999) 4613–4619.

[26] B. sgeisson, J. B. Hauksson, H. Gunnarsson, Dissociation and unfolding of cold-active alkaline phsophatase from atlantic cod in the prescence of guanidinium chloride, European Journal of Biochemistry 267 (2000) 6403–6412.

[27] S. D'Amico, J. C. Marx, C. Gerday, G. Feller, Activity–stability relationships in extremophilic enzymes, Journal of Biological Chemistry 278 (2003) 7891–7896.

[28] O. García-Arribas, R. Mateo, M. M. Tomczak, P. L. Davies, M. G. Mateu, Thermodynamic stability of a cold-adapted protein, type III antifreeze protein, an energetic contribution of salt bridges, Protein Science 16 (2009) 227–238.

[29] K. S. Siddiqui, R. Cavicchioli, Cold-adapted enzymes, Annual Review of Biochemistry 75 (2006) 403–433.

[30] A. J. Kiss, C.-H. C. 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 3 (2008) 155–171.

[31] C. N. Pace, J. M. Scholtz, Measuring the conformational stability of a protein. In: *Protein structure–a practical approach*, Oxford University Press, Oxford, New York, 1997.

[32] J. T. Macdonald, A. G. Purkiss, M. A. Smith, P. Evans, J. M. Goodfellow, C. Slingsby, Unfolding crystallins: The destabilizing role of a β-hairpin cysteine in β b2-crystallin by simulation and experiment, Protein Science 14 (5) (2005) 1282–1292.

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

[34] C. N. Kingsley, W. D. Brubaker, S. Markovic, A. Diehl, A. J. Brindley, H. Oschkinat, R. Martin, Preferential, specific binding of human αB-crystallin to a cataract-related variant of γS-crystallin, Structure 21 (2013) 2221–2227.

[35] N. Guex, M. C. Peitsch, Swiss-model and the swiss-pdbviewer: An environment for comparative protein modeling., Electrophoresis 18 (1997) 2714–2723.

[36] T. Chang, W. C. Chang, Cloning and sequencing of a carp beta S-crystallin cDNA, Biochimica Et Biophysica Acta 910 (1987) 89–92.

[37] Y. Quax-Jeuken, H. Driessen, J. Leunissen, W. Quax, W. de Jong, H. Bloemendal, Beta S-crystallin: structure and evolution of a distinct member of the $\beta\gamma$-superfamily, European Molecular Biology Organization Journal 4 (1985) 2597–2602.

[38] G. L. van Rens, W. W. de Jong, H. Bloemendal, One member of the γ-crystallin gene family, γS, is expressed in birds., Experimental Eye Research 53 (1991) 135–138.

[39] G. L. van Rens, J. M. Raats, H. P. Driessen, M. Oldenburg, J. T. Wijnen, P. M. Khan, W. W. de Jong, H. Bloemendal, Structure of the bovine eye lens γS-crystallin gene (formerly βS), Gene 78 (1989) 225–233.

[40] J. U. Bowie, J. F. Reidhaar-Olson, W. A. Lim, R. T. Sauer, Deciphering the message in protein sequences: tolerance to amino acid substitutions, Science 247 (1990) 1306–1310.

[41] S. Hormoz, Amino acid composition of proteins reduces deleterious impact of mutations, Scientific Reports 3 (2013) Article number: 2919.

[42] A. Szilágyi, P. Závodszky, Structural differences between mesophilic, moderately thermophilic, and extremely thermophilic protein subunits: results of a comprehensive survey, Structure 8 (5) (2000) 493–503.

[43] C. Budiman, Y. Koga, K. Takano, S. Kanaya, FK506-Binding Protein 22 from a psychrophilic bacterium, a cold shock-inducible peptidyl prolyl isomerase with the ability to assist in protein folding, International Journal of Molecular Sciences 12 (8) (2011) 5261–5284.

[44] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. WIlkins, R. D. Appel, A. Bairoch, The Proteomics Protocols Handbook, Humana Press, 2005, Ch. Protein Identification and Analysis Tools on the ExPASy Server, pp. 571–607.

[45] D. T. Jones, Protein secondary structure prediction based on position-specific scoring matrices, Journal of Molecular Biology 292 (1999) 195–202.

[46] D. W. A. Buchan, F. Minneci, T. C. O. Nugent, K. Bryson, D. T. Jones, Scalable web services for the PSIPRED protein analysis workbench, Nucleic Acids Research 41 (W1) (2013) W340–W348.

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

[48] J. Chen, P. Callis, J. 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 (2009) 3708–3716.

[49] N. Vajpai, L. Nisius, M. Wiktor, S. Grzesiek, High-pressure NMR reveals close similarity between cold and alcohol protein denaturation in ubiquitin, Proceedings of the National Academy of Sciences of the United States of America 110 (5) (2013) E368–E376.

[50] C. Babu, V. Hilser, A. Wand, Direct access to the cooperative substructure of proteins and the protein ensemble via cold denaturation, Nature Structural & Molecular Biology 11 (2004) 352–357.

[51] M. Jaremko, Ł. Jaremko, H.-Y. Kim, M.-K. Cho, C. D. Schwieters, K. Giller, S. Becker, M. Zweckstetter, Cold denaturation of a protein dimer monitored at atomic resolution, Nature Chemical Biology 9 (2013) 264–270.

[52] O. Annunziata, O. Ogun, G. B. Benedek, Observation of liquid-liquid phase separation for eye lens γS-crystallin, Proceedings of the National Academy of Sciences of the United States of America 100 (2003) 970–974.

[53] C. Liu, N. Asherie, A. Lomakin, J. Pande, O. Ogun, G. Benedek, Phase separation in aqueous solutions of lens γ-crystallins: special role of γS, Proceedings of the National Academy of Sciences of the United States of America 93 (1996) 377–382.

[54] M. Posner, A. J. Kiss, J. Skiba, A. Drossman, M. B. Dolinska, J. F. Hejtmancik, Y. V. Sergeev, Functional validation of hydrophobic adaptation to physiological temperature in the small heat shock protein αA-crystallin, PLoS One 7 (2012) e34438.

[55] B. J. Bennion, V. Daggett, The molecular basis for the chemical denaturation of proteins by urea, Proceedings of the National Academy of Sciences of the United States of America 100 (9) (2003) 5142—5147.

[56] O. D. Monera, C. M. Kay, R. S. Hodges, Protein denaturation with guanidine hydrochloride or urea provdes a differenct estimate of stability depending on the contributions of electrostatic interactions, Protein Science 3 (1994) 1984–1991.

[57] E. P. O'Brien, R. I. Dima, B. Brooks, D. Thirumalai, Interactions between hydrophobic and ionic solutes in aqueous guanidinium chloride and urea solutions: lessons for protein denaturation mechanism, Journal of the American Chemical Society 129 (2007) 7346–7353.

[58] T. Koishi, K. Yasuoka, S. Y. Willow, S. Fujikawa, X. C. Zeng, Molecular insight into different denaturing efficiency of urea, guanidinium, and mathanol: a comparative simulation study, Journal of Chemical Theory and Computation 9 (2013) 2540–2551.

[59] A. Hédoux, S. Krenzlin, L. Paccou, Y. Guinet, M.-P. Flamentad, J. Siepmann, Influence of urea and guanidine hydrochloride on lysozyme stability and thermal denaturation; a correlation between activity, protein dynamics and conformational changes, Physical Chemistry Chemical Physics 12 (2010) 13189–13196.

[60] D. Khago, J. Bierma, K. Roskamp, N. Kozlyuk, R. Martin, Protein refractive index increment is determined by conformation as well as composition,, Journal of Physics: Condensed Matter 30 (43) (2018) 435101.

[61] I. R. Kleckner, M. P. Foster, An introduction to NMR-based approaches for measuring protein dynamics, Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1814 (8) (2011) 942–968. doi:10.1016/j.bbapap.2010.10.012.

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

[63] A. Sivashanmugam, V. Murray, C. Cui, Y. Zhang, J. Wang, Q. Li, Practical protocols for production of very high yields of recombinant proteins usingEscherichia coli, Protein Science 18 (5) (2009) 936–948. doi:10.1002/pro.102.

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

[65] F. Delaglio, S. Grzesiek, G. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, Journal of Biomolecular NMR 6 (1995) 277–293.

[66] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, E. D. Laue, The CCPN data model for NMR spectroscopy: Development of a software pipeline., Proteins 59 (4) (2005) 687–696.

[67] C. N. Kingsley, J. C. Bierma, V. Pham, R. W. Martin, γS-crystallin proteins from the Antarctic nototheniid toothfish: A model system for investigating differential resistance to chemical and thermal denaturation, The Journal of Physical Chemistry B 118 (47) (2014) 13544–13553.

[68] A. Molliex, J. Temirov, J. Lee, M. Coughlin, H. J. Kanagaraj, Anderson P.and Kim, T. Mittag, P. Taylor, Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization, Cell 163 (1) (2015) 123–133.

[69] T. J. Nott, E. Petsalaki, P. Farber, D. Jervis, E. Fussner, A. Plochowietz, D. P. Craggs, Timothy D.and Bazett-Jones, T. Pawson, J. D. Forman-Kay, A. J. Baldwin, Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles, Molecular Cell 57 (5) (2015) 936–947.

[70] J. Berry, S. Weber, N. Vaidya, M. Haataja, C. Brangwynne, RNA transcription modulates phase transition-driven nuclear body assembly, Proceedings of the National Academy of Sciences of the United States of America 112 (2015) E5237–E5245.

[71] Y. Lin, D. S. Protter, M. K. Rosen, R. Parker, Formation and maturation of phase-separated liquid droplets by RNA-binding proteins, Molecular Cell 60 (2) (2015) 208–219.

[72] M. Feric, N. Vaidya, T. S. Harmon, D. M. Mitrea, L. Zhu, T. M. Richardson, R. W. Kriwacki, R. V. Pappu, C. P. Brangwynne, Coexisting liquid phases underlie nucleolar subcompartments, Cell 165 (2016) 1686–1697.

[73] L. Guo, H. Kim, H. Wang, J. Monaghan, F. Freyermuth, J. C. Sung, K. O'Donovan, C. M. Fare, Z. Diaz, N. Singh, Z. C. Zhang, M. Coughlin, E. A. Sweeny, M. E. DeSantis, M. E. Jackrel, C. B. Rodell, J. A. Burdick, O. D. King, A. D. Gitler, C. Lagier-Tourenne, U. B. Pandey, Y. M. Chook, J. P. Taylor, J. Shorter, Nuclear-import receptors reverse aberrant phase transitions of RNA-binding proteins with prion-like domains, Cell 173 (2018) 677–692.

[74] S. Wegmann, B. Eftekharzadeh, K. Tepper, K. M. Zoltowska, R. E. Bennett, S. Dujardin, P. R. Laskowski, D. MacKenzie, T. Kamath, C. Commins, C. Vanderburg, A. D.

Roe, Z. Fan, A. M. Molliex, A. Hernandez–Vega, D. Muller, A. A. Hyman, E. Mandelkow, J. P. Taylor, B. T. Hyman, Tau protein liquid–liquid phase separation can initiate tau aggregation, The EMBO Journal (2018) e98049.

[75] A. E. Conicella, G. H. Zerze, J. Mittal, N. L. Fawzi, ALS mutations disrupt phase separation mediated by α-helical structure in the TDP-43 low complexity C-terminal domain, Structure 24 (9) (2016) 1537–1549.

[76] Z. Monahan, V. H. Ryan, A. M. Janke, K. A. Burke, S. N. Rhoads, G. H. Zerze, R. O'Meally, G. L. Dignon, A. E. Conicella, W. Zheng, R. B. Best, R. N. Cole, J. Mittal, F. Shewmaker, N. L. Fawzi, Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity, The EMBO Journal 36 (20) (2017) 2951–2967.

[77] M. Hughes, M. Sawaya, B. D.R., L. Goldschmidt, J. Rodriguez, D. Cascio, L. Chong, T. Gonen, D. Eisenberg, Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks, Science 359 (6376) (2018) 698–701.

[78] S. Ling, C. Decker, M. Walsh, M. She, R. Parker, H. Song, Crystal structure of human Edc3 and its functional implications, Molecular and Cellular Biology 28 (2008) 5965–5976.

[79] J. I. Clark, D. Carper, Phase separation in lens cytoplasm is genetically linked to cataract formation in the Philly mouse, Proceedings of the National Academy of Sciences of the United States of America 84 (1987) 122–125.

[80] J. A. Thomson, P. Schurtenberger, G. M. Thurston, G. B. Benedek, Binary liquid phase separation and critical phenomena in a protein/water solution, Proceedings of the National Academy of Sciences of the United States of America 84 (1987) 7079–7083.

[81] P. Schurtenberger, R. A. Chamberlin, G. M. Thurston, J. A. Thomson, G. B. Benedek, Observation of critical phenomena in a protein-water solution, Physical Review Letters 63 (1989) 2064–2067.

[82] F. Ji, J. Jung, A. M. Gronenborn, Structural and biochemical characterization of the childhood cataract-associated R76S mutant of human γD-crystallin, Biochemistry 51 (12) (2012) 2588–2596.

[83] A. Pande, J. Zhang, P. R. Banerjee, S. S. Puttamadappa, A. Shekhtman, J. 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) (2009) 196–199.

[84] J. Jung, I.-J. L. Byeon, Y. Wang, J. King, A. M. Gronenborn, The structure of the cataract-causing P23T mutant of human γD-crystallin exhibits distinctive local conformational and dynamic changes, Biochemistry 48 (12) (2009) 2597–2609.

[85] P. R. Banerjee, S. S. Puttamadappa, A. Pande, A. Shekhtman, J. 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) (2011) 647–659.

[86] P. Privalov, Cold denaturation of proteins, Critical Reviews in Biochemistry and Molecular Biology 25 (4) (1990) 281–305.

[87] D. Sanfelice, P. A. Temussi, Cold denaturation as a tool to measure protein stability, Biophysical Chemistry 208 (2016) 4–8.

[88] M. Adrover, V. Esposito, G. Martorell, A. Pastore, P. A. Temussi, Understanding cold denaturation: The case study of Yfh1, Journal of the American Chemical Society 132 (45) (2010) 16240–16246.

[89] D. Sanfelice, E. Morandi, A. Pastore, N. Niccolai, P. A. Temussi, Cold denaturation unveiled: Molecular mechanism of the asymmetric unfolding of yeast frataxin, ChemPhysChem 16 (17) (2015) 3599–3602.

[90] S. Kunugi, N. Tanaka, Cold denaturation of proteins under high pressure, Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology 1595 (1–2) (2002) 329–344.

[91] K.-B. Wong, S. M. V. Freund, A. R. Fersht, Cold denaturation of barstar: ^{1}H, ^{15}N and ^{13}C NMR assignment and characterisation of residual structure, Journal of Molecular Biology 259 (1996) 805–818.

[92] R. Babu, Charles, V. J. Hilser, A. J. Wand, Direct access to the cooperative substructure of proteins and the protein ensemble via cold denaturation, Nature Structural & Molecular Biology 11 (2004) 352–357.

[93] M. S. Pometun, R. W. Peterson, C. R. Babu, A. J. Wand, Cold denaturation of encapsulated ubiquitin, Journal of the American Chemical Society 128 (33) (2006) 10652–10653.

[94] M. Delaye, J. L. Clark, G. B. Benedek, Identification of the scattering elements responsible for the lens opacification in cold cataracts, Biophysical Journal 37 (1982) 647–656.

[95] M. L. Broide, C. R. Berland, J. Pande, O. O. Ogun, G. B. Benedek, Binary-liquid phase separation of lens protein solutions, Proceedings of the National Academy of Sciences of the United States of America 88 (1991) 5660–5664.

[96] Y. Wang, A. Lomakin, J. J. McManus, O. Ogun, G. B. Benedek, Phase behavior of mixtures of human lens proteins γD and βB1, Proceedings of the National Academy of Sciences of the United States of America 107 (30) (2010) 13282–13287.

[97] G. M. Thurston, Liquid-liquid phase separation and static light scattering of concentrated ternary mixtures of bovine α and γB-crystallins, Journal of Chemical Physics 124 (2006) doi:10.1063/1.2168451.

[98] H. Zhao, Y. Chen, L. Rezabkova, Z. Wu, G. Wistow, P. Schuck, Solution properties of γ-crystallins: Hydration of fish and mammal γ-crystallins, Protein Science 23 (1) (2014) 88–99.

[99] A. J. Kiss, A. L. Devries, R. M. Morgan-Kiss, Comparative analysis of crystallins and lipids from the lens of Antarctic toothfish and cow, Journal of Comparative Physiology B 180 (7) (2010) 1019–1032.

[100] V. N. Uversky, Intrinsically disordered proteins in overcrowded milieu: Membraneless organelles, phase separation, and intrinsic disorder, Current Opinion in Structural Biology 44 (2017) 18–30.

[101] T. J. Nott, T. D. Craggs, A. J. Baldwin, Membraneless organelles can melt nucleic acid duplexes and act as biomolecular filters, Nature Chemistry 8 (2016) 569–575.

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

[103] J. Zhang, G. R. Kinsel, Quantification of protein-polymer interactions by matrix-assisted laser desorption/ionization mass spectrometry, Langmuir 18 (2002) 4444–4448.

[104] M. C. L. Gay, I. Mullaney, D. Trinder, J. K. Olynyk, R. D. Trengove, Quantitative assay of urinary hepcidin using MALDI-TOF mass spectrometry, Analytical Methods 2 (2010) 268–274.

[105] K. Monkos, Temperature behaviour of viscous flow with proteins, General Physiology and Biophysics 30 (Special Issue) (2011) 121–129.

[106] J. Liao, J. Lee, S. Chiou, Distinct roles of αA- and αB-crystallins under thermal and UV stresses, Biochemical and Biophysical Research Communications 295 (2002) 854–861.

[107] D. E. Kim, D. Chivian, D. Baker, Protein structure prediction and analysis using the Robetta server, Nucleic Acids Research 32 (suppl 2) (2004) W526–W531.

[108] A. J. Kiss, Functional, biochemical and molecular analyses of the cold stable eye lens crystallins from the Antarctic toothfish *Dissostichus mawsoni*, Ph.D. thesis, University of Illinois at Urbana-Champaign (2005).

[109] R. K. Gaur, Amino acid frequency distribution among eukaryotic proteins, The IIOAB Journal 5 (2) (2014) 6.

[110] Z. Ma, G. Piszczek, P. Wingfield, Y. Sergeev, J. Hejtmancik, The G18V CRYGS mutation associated with human cataracts increases gammaS-crystallin sensitivity to thermal and chemical stress, Biochemistry 48 (2009) 7334–7341.

[111] J. W. Wu, M.-E. Chen, W.-S. Wen, W.-A. Chen, C.-T. Li, C.-K. Chang, C.-H. Lo, H.-S. Liu, S. S.-S. Wang, Comparative analysis of human γD-crystallin aggregation under physiological and low pH conditions, PloS ONE 9 (11) (2014) e112309.

[112] E. Serebryany, T. Takata, E. Erickson, N. Schafheimer, Y. Wang, J. A. King, Aggregation of Trp > Glu point mutants of human gamma-D crystallin provides a model for hereditary or UV-induced cataract, Protein Science 25 (2016) 1115—1128.

[113] P. Banerjee, A. Pande, J. Patrosz, G. Thurston, J. Pande, Cataract-associated mutant E107A of human γD-crystallin shows increased attraction to α-crystallin and enhanced light scattering, Proceedings of the National Academy of Sciences of the United States of America 108 (2011) 574–579.

[114] L. Shimoni, J. Glusker, Hydrogen bonding motifs of protein side chains: Descriptions of binding of arginine and amide groups, Protein Science 4 (1995) 65–74.

[115] P. E. Mason, G. W. Neilson, C. E. Dempsey, A. C. Barnes, J. M. Cruickshank, The hydration structure of guanidinium and thiocyanate ions: Implications for protein stability in aqueous solution, Proceedings of the National Academy of Sciences 100 (8) (2003) 4557–4561. doi:10.1073/pnas.0735920100.

[116] J. VondVorášek, P. E. Mason, J. Heyda, K. D. Collins, P. Jungwirth, The molecular origin of like-charge arginine-arginine pairing in water, The Journal of Physical Chemistry B 113 (27) (2009) 9041–9045. doi:10.1021/jp902377q.

[117] J. Gallivan, D. Dougherty, Cation-π interactions in structural biology, Proceedings of the National Academy of Sciences of the United States of America 96 (1999) 9459–9464.

[118] A. K. Panda, S. K. Nandi, A. Chakraborty, R. H. Nagaraj, A. Biswas, Differential role of arginine mutations on the structure and functions of α-crystallin, Biochimica et Biophysica Acta - General Subjects 1860 (2016) 199–210.

[119] R. J. Siezen, E. Wu, E. D. Kaplan, J. A. Thomson, G. B. Benedek, Rat lens γ-crystallins: Characterization of the six gene products and their spatial and temporal distribution resulting from differential synthesis, Journal of Molecular Biology 199 (1988) 475–490.

[120] M. Quinn, N. Gnan, S. James, A. Ninarello, F. Sciortino, E. Zaccarelli, J. McManus, How fluorescent labelling alters the solution behaviour of proteins, Physical Chemistry Chemical Physics 17 (46) (2015) 31177–31187.

[121] P. A. Wilmarth, S. Tanner, S. Dasari, S. R. Nagalla, M. A. Riviere, V. Bafna, P. A. Pevzner, L. L. David, Age-related changes in human crystallins determined from comparative analysis of post-translational modifications in young and aged lens: Does deamidation contribute to crystallin insolubility?, Journal of Proteome Research 5 (10) (2006) 2554–2566.

[122] P. G. Hains, R. J. W. Truscott, Post-translational modifications in the nuclear region of young, aged, and cataract human lenses, Journal of Proteome Research 6 (2007) 3935–3943.

[123] O. Srivastava, K. Srivastava, J. Chaves, A. Gill, Post-translationally modified human lens crystallin fragments show aggregation in vitro, Biochemistry and Biophysics Reports 10 (2017) 94–131. doi:10.1016/j.bbrep.2017.01.011.

[124] S. H. Gellman, On the role of methionine residues in the sequence-independent recognition of nonpolar protein surfaces, Biochemistry 30 (1991) 6633–6636.

[125] G. Feller, Psychrophilic enzymes: From folding to function and biotechnology, Scientifica 2013 (2013) Article ID 512840.

[126] A. Schön, B. R. Clarkson, M. Jaime, E. Freire, Temperature stability of proteins: Analysis of irreversible denaturation using isothermal calorimetry, Proteins: Structure, Function, and Bioinformatics 85 (11) (2017) 2009–2016k.

[127] M. Wenk, R. Herbst, D. Hoeger, M. Kretschmar, N. H. Lubsen, R. Jaenicke, Gamma S-crystallin of bovine and human eye lens: solution structure, stability and folding of the intact two-domain protein and its separate domains, Biophysical Chemistry 86 (2-3) (2000) 95–108.

[128] C. W. Pak, M. Kosno, A. S. Holehouse, S. B. Padrick, A. Mittal, R. Ali, A. A. Yunus, D. R. Liu, R. V. Pappu, M. K. Rosen, Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein, Molecular Cell 63 (2016) 72–85.

[129] P. B. Crowley, A. Golovin, Cation–π interactions in protein–protein interfaces, Proteins: Structure, Function, and Bioinformatics 59 (2) (2005) 231–239.

[130] B. Norledge, R. Hay, O. Bateman, C. Slingsby, H. Driessen, Towards a molecular understanding of phase separation in the lens: A comparison of the x-ray structures of two high Tc γ-crystallins, γE and γF, with two low Tcγ-crystallins, γB and γD, Experimental Eye Research 65 (5) (1997) 609–630.

[131] S. Boeynaems, E. Bogaert, D. Kovacs, A. Konijnenberg, E. Timmerman, A. Volkov, M. Guharoy, M. De Decker, T. Jaspers, V. H. Ryan, A. M. Janke, P. Baatsen, T. Vercruysse, R. Kolaitis, D. Daelemans, J. P. Taylor, N. Kedersha, P. Anderson, F. Impens, F. Sobott, J. Schymkowitz, F. Rousseau, N. L. Fawzi, W. Robberecht, P. Van Damme, P. Tompa, L. Van Den Bosch, Phase separation of *C9orf72* dipeptide repeats perturbs stress granule dynamics, Molecular Cell 65 (2017) 1044–1055.

[132] M. Hofweber, S. Hutten, B. Bourgeois, E. Spreitzer, A. Niedner-Boblenz, M. Schifferer, M.-D. Ruepp, M. Simons, D. Niessing, T. Madl, D. Dormann, Phase separation of FUS is suppressed by its nuclear import receptor and arginine methylation, Cell 173 (2018) 706–719.

[133] S. Qamar, G. Wang, S. J. Randle, F. S. Ruggeri, J. A. Varela, J. Q. Lin, E. C. Phillips, A. Miyashita, D. Williams, F. Ströhl, W. Meadows, R. Ferry, V. J. Dardov, G. G. Tartaglia, L. A. Farrer, G. S. Kaminski Schierle, C. F. Kaminski, C. E. Holt, P. E. Fraser, G. Schmitt-Ulms, D. Klenerman, T. Knowles, M. Vendruscolo, P. St George-Hyslop, FUS phase separation is modulated by a molecular chaperone and methylation of arginine cation-π interactions, Cell 173 (2018) 720–734.

[134] V. G. Taratuta, A. Holschbach, G. M. Thurston, D. Blankschtein, G. B. Benedek, Liquid-liquid phase separation of aqueous lysozyme solutions: effects of pH and salt identity, The Journal of Physical Chemistry 94 (5) (1990) 2140–2144. doi:10.1021/j100368a074.

[135] J. Grigsby, H. Blanch, J. Prausnitz, Cloud-point temperatures for lysozyme in electrolyte solutions: effect of salt type, salt concentration and pH, Biophysical Chemistry 91 (3) (2001) 231–243. doi:10.1016/s0301-4622(01)00173-9.

[136] F. Platten, J. Hansen, J. Milius, D. Wagner, S. U. Egelhaaf, Additivity of the specific effects of additives on protein phase behavior, The Journal of Physical Chemistry B 119 (48) (2015) 14986–14993. doi:10.1021/acs.jpcb.5b08078.

[137] G. J. Wistow, J. Piatigorsky, Lens crystallins: The evolution and expression of proteins for a highly specialized tissue, Annual Review of Biochemistry 57 (1988) 479–504.

[138] C. J. Weadick, B. S. W. Chang, Molecular evolution of the $\beta\gamma$ lens crystallin superfamily: evidence for a retained ancestral function in γn crystallins, Molecular Biology and Evolution 26 (2009) 1127—1142.

[139] J. Piatigorsky, G. Wistow, The recruitment of crystallins: new functions precede gene duplication, Science 252 (1991) 1078–1079.

[140] S. Tomarev, S. I. snd Chung, J. Piatigorsky, Glutathione S-transferase and S-crystallins of cephalopods: evolution from active enzyme to lens-refractive proteins, Journal of Molecular Evolution 41 (1995) 1048–1056.

[141] W.-H. Tan, S.-C. Cheng, Y.-T. Liu, C.-G. Wu, M.-H. Lin, C.-C. Chen, C.-H. Lin, C.-Y. Chou, Structure of a highly active cephalopod S-crystallin mutant: New molecular evidence for evolution from an active enzyme into lens-refractive protein, Scientific Reports 6 (2016) 31176.

[142] G. J. Wistow, M. J. W. M., W. W. de Jong, The enzyme lactate dehydrogenase as a structural protein in avian and crocodilian lenses, Nature 326 (1987) 622–624.

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

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

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

[146] J. Wen, T. Arakawa, J. S. Philo, Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions, Analytical Biochemistry 240 (2) (1996) 155–166.

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

[148] R. Barer, S. Joseph, Refractometry of living cells, The Quarterly Journal of Microscopical Science 95 (1954) 399—423.

[149] K. Mahendiran, C. Elie, J.-C. Nebel, A. Ryan, B. K. Pierscionek, Primary sequence contribution to the optical function of the eye lens, Scientific Reports 4 (2014) 5195.

[150] W. Fredericks, M. Hammonds, S. Howard, F. Rosenberger, Density, thermal expansivity, viscosity and refractive index of lysozyme solutions at crystal growth concentrations, Journal of Crystal Growth 141 (1994) 183–192.

[151] B. Pierscionek, G. Smith, R. C. Augusteyn, The refractive increments of bovine α-, β-and γ-crystallins, Vision Research 27 (9) (1987) 1539–1541.

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

[153] N. Kozlyuk, S. Sengupta, J. C. Bierma, R. W. Martin, Calcium binding dramatically stabilizes an ancestral crystallin fold in tunicate $\beta\gamma$-crystallin, Biochemistry 55 (50) (2016) 6961–6968.

[154] T. L. McMeekin, M. Wilensky, M. L. Groves, Refractive indices of proteins in relation to amino acid composition and specific volume, Biochemical and Biophysical Research Communications 7 (2) (1962) 151–156.

[155] T. L. McMeekin, M. L. Groves, N. J. Hipp, Refractive indices of amino acids, proteins, and related substances, in: Amino Acids and Serum Proteins, ACS Publications, 1964, Ch. 4, pp. 54–66.

[156] E. J. Cohn, J. T. Edsall, Density and apparent specific volume of proteins, in: Proteins, amino acids, and peptides as ions and dipolar ions, Reinhold Publishing Corporation, New York, 1943, Ch. 16, pp. 370–381.

[157] R. Barer, S. Joseph, Refractometry of living cells: Part i. basic principles, Journal of Cell Science 3 (32) (1954) 399–423.

[158] W. Heller, Remarks on refractive index mixture rules, The Journal of Physical Chemistry 69 (4) (1965) 1123–1129.

[159] G. E. Perlmann, L. Longsworth, The specific refractive increment of some purified proteins, Journal of the American Chemical Society 70 (8) (1948) 2719–2724.

[160] N. Coquelle, A. Brewster, U. Kapp, A. Shilova, B. Weinhausen, M. Burghammer, J. Colletier, Raster-scanning serial protein crystallography using micro- and nano-focused synchrotron beams, Acta Crystallographica, Section D: Biological Crystallography 71 (2015) 1184–1196.

[161] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, The protein data bank, Nucleic Acids Research 28 (2000) 235–242.

[162] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera–a visualization system for exploratory research and analysis, Journal of Computational Chemistry 25 (13) (2004) 1605–1612.

[163] M. F. Sanner, A. J. Olson, J.-C. Spehner, Reduced surface: An efficient way to compute molecular surfaces, Biopolymers 38 (3) (1996) 305–320. doi:10.1002/(sici)1097-0282(199603)38:3<305::aid-bip4>3.0.co;2-y.

[164] T. Gibaud, F. Cardinaux, J. Bergenholtz, A. Stradner, P. Schurtenberger, Phase separation and dynamical arrest for particles interacting with mixed potentials-the case of globular proteins revisited, Soft Matter 7 (2011) 857–860. doi:10.1039/C0SM01175D. URL http://dx.doi.org/10.1039/C0SM01175D

[165] J. Kratohvil, G. Dezelic, D. N., On the refractive index increment of bovine plasma albumin at low concentrations, Archives of Biochemistry and Biophysics 106 (1964) 381–385.

[166] C. Böttcher, Theory of electric polarization, 2nd Edition, Vol. 1, Elsevier Science Publishers B.V., Sara Burgerhartstraat 25, PO Box 211, 1000 AE Amsterdam, The Netherlands, 1973.

[167] T. V. Chalikian, M. Totrov, R. Abagyan, K. J. Breslauer, The hydration of globular proteins as derived from volume and compressibility measurements: Cross correlating thermodynamic and structural data, Journal of Molecular Biology 260 (1996) 588–603.

[168] D. P. Kharakoz, Volumetric properties of proteins and their analogs in diluted water solutions. 1. Partial volumes of amino acids at 15–55 °C, Biophysical Chemistry 34 (1989) 5634–5642.

[169] D. P. Kharakoz, Volumetric properties of proteins and their analogs in diluted water solutions. 2. Partial adiabatic compressibilities of amino acids at 15–70 ° C, J. Phys. Chem. 95 (1991) 5634–5642.

[170] D. P. Kharakoz, Partial molar volumes of molecules of arbitrary shape and the effect of hydrogen bonding with water, Journal of Solution Chemistry 21 (1992) 569–595.

[171] D. P. Kharakoz, A. P. Sarvazyan, Hydrational and intrinsic compressibilities of globular proteins, Biopolymers 33 (1993) 11–26.

[172] D. Laage, T. Elsaesser, J. T. Hynes, Water dynamics in the hydration shells of biomolecules, Chemical Reviews 117 (2017) 10694âĹŠ10725.

[173] V. Ball, J. J. Ramsden, Buffer dependence of refractive index increments of protein solutions, Biopolymers 46 (7) (1998) 489–492. doi:10.1002/(sici)1097-0282(199812)46:7<489::aid-bip6>3.0.co;2-e.

[174] K. J. Donald, Electronic compressibility and polarizability: Origins of a correlation, Journal of Physical Chemistry A 110 (6) (2006) 2283–2289.

[175] A. Pavlovič, M. Saganová, A novel insight into the cost–benefit model for the evolution of botanical carnivory, Annals of Botany 115 (7) (2015) 1075–1092. doi:10.1093/aob/mcv050.

[176] M. Paniw, E. Gil-Cabeza, F. Ojeda, Plant carnivory beyond bogs: reliance on prey feeding inDrosophyllum lusitanicum(drosophyllaceae) in dry mediterranean heathland habitats, Annals of Botany (2017) mcw247doi:10.1093/aob/mcw247.

[177] A. M. Ellison, N. J. Gotelli, Energetics and the evolution of carnivorous plants-darwin's 'most wonderful plants in the world', Journal of Experimental Botany 60 (1) (2009) 19–42. doi:10.1093/jxb/ern179.

[178] C. T. Butts, J. C. Bierma, R. W. Martin, Novel proteases from the genome of the carnivorous plantDrosera capensis: Structural prediction and comparative analysis, Proteins: Structure, Function, and Bioinformatics 84 (10) (2016) 1517–1533. doi:10.1002/prot.25095.

[179] E. Król, B. J. Płachno, L. Adamec, M. Stolarz, H. Dziubińska, K. Trębacz, Quite a few reasons for calling carnivores 'the most wonderful plants in the world', Annals of Botany 109 (1) (2011) 47–64. doi:10.1093/aob/mcr249.

[180] T. Kokubun, Occurrence of myo-inositol and alkyl-substituted polysaccharide in the prey-trapping mucilage of drosera capensis, The Science of Nature 104 (9-10). doi:10.1007/s00114-017-1502-4.

[181] H. Bruhn, A short guided tour through functional and structural features of saposin-like proteins, Biochemical Journal 389 (2) (2005) 249–257. doi:10.1042/bj20050051.

[182] B. C. Bryksa, D. A. Grahame, R. Y. Yada, Comparative structure-function characterization of the saposin-like domains from potato, barley, cardoon and arabidopsis aspartic proteases, Biochimica et Biophysica Acta (BBA) - Biomembranes 1859 (5) (2017) 1008–1018. doi:10.1016/j.bbamem.2017.02.007.

[183] F. F. Muñoz, J. R. Mendieta, M. R. Pagano, R. A. Paggi, G. R. Daleo, M. G. Guevara, The swaposin-like domain of potato aspartic protease (StAsp-PSI) exerts antimicrobial activity on plant and human pathogens, Peptides 31 (5) (2010) 777–785. doi:10.1016/j.peptides.2010.02.001.

[184] G. Faust, A. Stand, D. Weuster-Botz, IPTG can replace lactose in auto-induction media to enhance protein expression in batch-culturedEscherichia coli, Engineering in Life Sciences 15 (8) (2015) 824–829. doi:10.1002/elsc.201500011.

[185] E. G. Bligh, W. J. Dyer, A RAPID METHOD OF TOTAL LIPID EXTRACTION AND PURIFICATION, Canadian Journal of Biochemistry and Physiology 37 (8) (1959) 911–917. doi:10.1139/o59-099.

[186] V. Vandana, M. S. L. Karuna, P. Vijayalakshmi, R. B. N. Prasad, A simple method to enrich phospholipid content in commercial soybean lecithin, Journal of the American Oil Chemists' Society 78 (5) (2001) 555–556. doi:10.1007/s11746-001-0303-2.

[187] M. D. Bird, W. W. Brey, T. A. Cross, I. R. Dixon, A. Griffin, S. T. Hannahs, J. Kynoch, I. M. Litvak, J. L. Schiano, J. Toth, Commissioning of the 36 t series-connected hybrid magnet at the NHMFL, IEEE Transactions on Applied Superconductivity 28 (3) (2018) 1–6. doi:10.1109/tasc.2017.2781727.

[188] A. Pines, M. G. Gibby, 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) (1972) 1776–1777.

[189] K. Takegoshi, S. Nakamura, T. Terao, 13C–1H dipolar-assisted rotational resonance in magic-angle spinning NMR, Chemical Physics Letters 344 (5-6) (2001) 631–637. doi:10.1016/S0009-2614(01)00791-6.
URL http://www.sciencedirect.com/science/article/pii/S0009261401007916

[190] M. Motta, M. Tatti, R. 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, Elsevier, 2014, pp. 131–146. doi:10.1016/b978-0-12-405529-2.00009-3.

[191] B. C. Bryksa, P. Bhaumik, E. Magracheva, D. C. D. Moura, M. Kurylowicz, A. Zdanov, J. R. Dutcher, A. Wlodawer, R. Y. Yada, Structure and mechanism of the saposin-like domain of a plant aspartic protease, Journal of Biological Chemistry 286 (32) (2011) 28265–28275. doi:10.1074/jbc.m111.252619.

[192] J. Gally, G. Edelman, The effect of temperature on the fluorescence of some aromatic amino acids and proteins, Biochimica et Biophysica Acta 60 (3) (1962) 499–509. doi:10.1016/0006-3002(62)90869-7.

[193] J. Hennecke, A. Sillen, M. Huber-Wunderlich, Y. Engelborghs, R. Glockshuber, Quenching of tryptophan fluorescence by the active-site disulfide bridge in the DsbA protein fromEscherichia coli†, Biochemistry 36 (21) (1997) 6391–6400. doi:10.1021/bi963017w.

[194] V. Cherezov, M. Caffrey, Membrane protein crystallization in lipidic mesophases. a mechanism study using x-ray microdiffraction, Faraday Discussions 136 (2007) 195. doi:10.1039/b618173b.

[195] T. Sugiki, N. Kobayashi, T. Fujiwara, Modern technologies of solution nuclear magnetic resonance spectroscopy for three-dimensional structure determination of proteins open avenues for life scientists, Computational and Structural Biotechnology Journal 15 (2017) 328–339. doi:10.1016/j.csbj.2017.04.001.

[196] I. G. Denisov, S. G. Sligar, Nanodiscs for structural and functional studies of membrane proteins, Nature Structural & Molecular Biology 23 (6) (2016) 481–486. doi:10.1038/nsmb.3195.

[197] R. J. Gilbert, M. D. Serra, C. J. Froelich, M. I. Wallace, G. Anderluh, Membrane pore formation at protein–lipid interfaces, Trends in Biochemical Sciences 39 (11) (2014) 510–516. doi:10.1016/j.tibs.2014.09.002.

[198] V. Ladizhansky, Applications of solid-state NMR to membrane proteins, Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1865 (11) (2017) 1577–1586. doi:10.1016/j.bbapap.2017.07.004.

[199] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, et al., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Molecular Systems Biology 7 (1) (2011) 539.

[200] T. N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions, Nature Methods 8 (10) (2011) 785–786. doi:10.1038/nmeth.1701.

[201] S. Heiss, V. Puxbaum, C. Gruber, F. Altmann, B. Gasser, D. Mattanovich, Multistep processing of the secretion leader of the extracellular protein epx1 in pichia pastoris and implications for protein localization, Microbiology 161 (7) (2015) 1356–1368. doi:10.1099/mic.0.000105.

[202] K. D. Schutter, N. Callewaert, Pichia surface display: A tool for screening single domain antibodies, in: Methods in Molecular Biology, Humana Press, 2012, pp. 125–134. doi:10.1007/978-1-61779-968-6_8.

[203] M. Weidner, M. Taupp, S. J. Hallam, Expression of recombinant proteins in the methylotrophic yeast pichia pastoris, Journal of Visualized Experiments (36). doi:10.3791/1862.

LIST OF FIGURES IN APPENDIX

Appendix A

Supplementary data for Crystallin LLPS

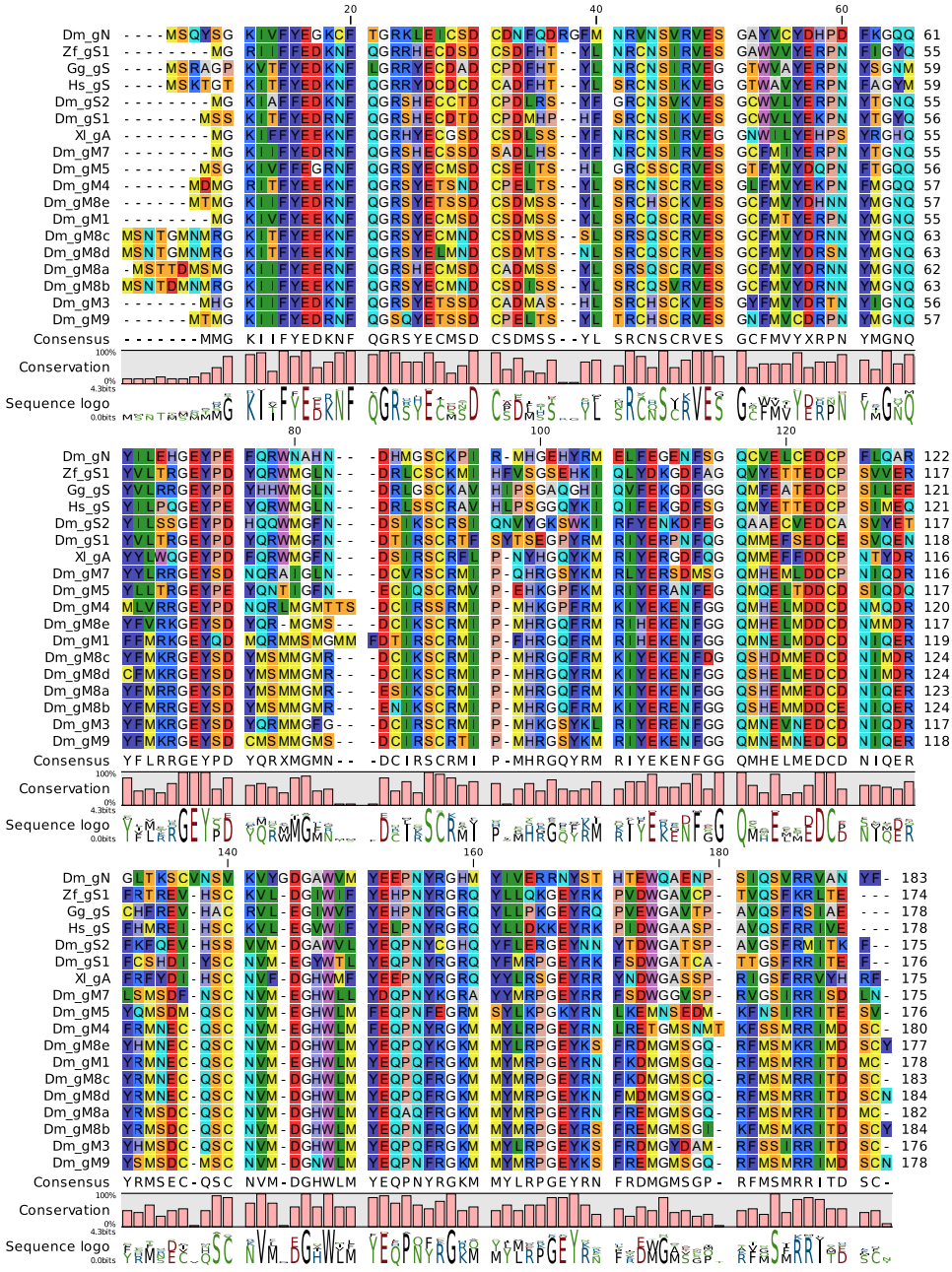

Figure A.1: MUSCLE v3.8.425 based alignment of primary sequence of the *D. mawsoni* (Dm) γ lens crystallins as well as selected sequences from *Homo sapiens* (Hs), *Gallus gallus* (Gg), *Danio rerio* (Zf), *Xenopus laevis* (Xl). Sequences are background colored per RASMOL with consensus sequence, conservation graph and sequence logo indicated. MUSCLE clusters sequences most similar in production of the alignment.

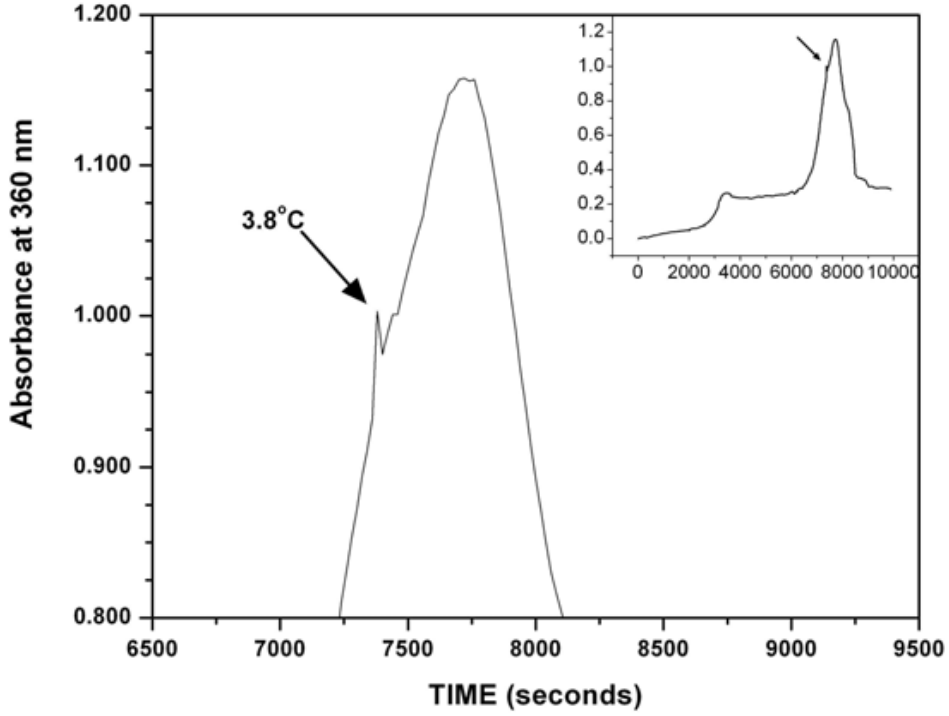

Figure A.2: Reversible cold sensitivity of toothfish γ-crystallin FIII. Increase in A_{360} is interpreted as an increase in turbidity caused by LLPS upon slow cooling of the sample. The initial temperature was 25 °C with a cooling rate of approximately 0.6 °C per minute. The inset shows the full time course, while the regular plot shows zoomed in portion of A_{360} increase. Black arrow indicates the point (3.8 °C) at which there was a rapid increase in turbidity. Turbidity A_{360} began to increase at approximately 6 °C (6060sec). There was a sharp spike in A360 at 7380 seconds (123 minutes) corresponding to 3.8 °C. As cooling continued, the turbidity reached a maximum corresponding to an A_{360} of 1.16 at 2 °C at 7720 seconds (128 minutes) before the turbidity of FIII begins to reverse upon warming. Turbidity gradually decreased when the FIII sampled was warmed from 2 °C to 4 °C with no rapid decrease around 4 °C. During warming to ≈20 °C the turbidity continued to decrease, but did not completely disappear with a final A_{360} of 0.294.

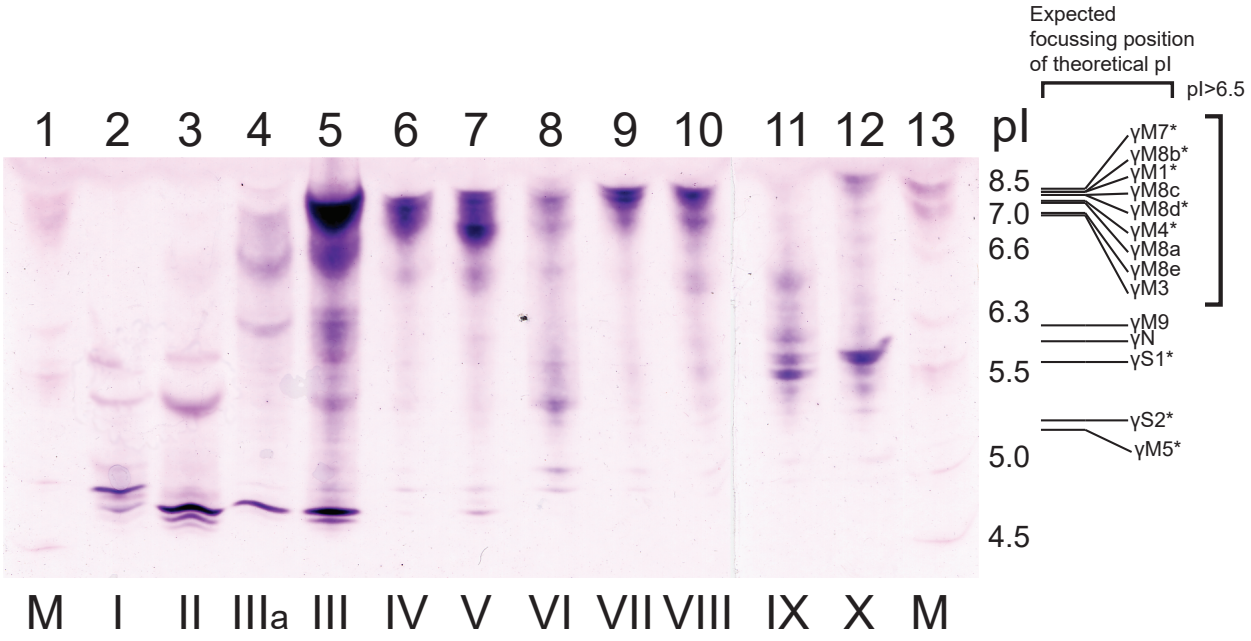

Figure A.3: Lanes denoted with an M represent protein standards and Roman numerals denote specific fractions obtained from the cation exchange. Fraction III undergoes phase separation around 4 °C. The expected focusing positions of the toothfish γ-crystallins are shown based on their predicted pI. γ-crystallins marked with an asterisks are those that showed peptides in the MS analysis.

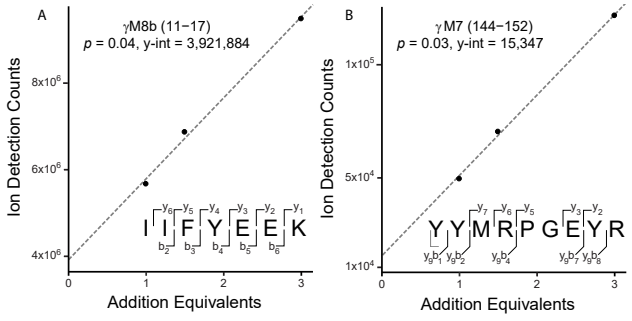

Figure A.4: (A) γM8b peptide fragment 11-17 and (B) γM7 peptide fragment 144-152 detected by mass spectrometry. The b and y fragments of each respective protein detected by MS/MS are shown in the lower right of each panel.

Table A.1:

protein	molecular weight (kD)	# -ve charged residues	# +ve charged residues	predicted pI	pI above 6.5	MS detected unique peptides	standard addition concentration (μg/mL)
γM7	20.7	22	24	8.15	X	X	2.60
γM8b	22.4	24	26	8.04	X	X	9.53
γM1	21.8	23	25	8.02	X	X	3.84
γM8c	22.1	24	26	7.97	X		NA
γM8d	22.1	23	25	7.97	X	X	10.7
γM4	21.5	24	25	7.54	X	X	NA
γM8a	22.0	24	25	7.53	X		NA
γM8e	21.3	23	23	7.01	X		NA
γM3	20.1	24	24	6.98	X		NA
γM9	21.2	23	22	6.15			NA
γN	21.7	25	20	5.9			NA
γS1	21.2	24	20	5.61		X	NA
γS2	20.4	23	17	5.22		X	NA
γM5	20.7	25	19	5.16		X	NA

Table A.2:

amino acid type	dn/dc from ref. [10] (mL/g)	frequency average (%)	frequency HγS human (%)	frequency γM1 toothfish (%)	frequency γM4 toothfish (%)	frequency γM7 toothfish (%)	frequency γM8b toothfish (%)	frequency γM8c toothfish (%)	frequency γM8d toothfish (%)
Ala (A)	0.167	6.0	3.9	0.0	0.0	1.7	0.0	0.0	0.0
Arg (R)	0.206	5.2	7.3	10.0	10.7	11.9	10.3	9.8	9.2
Asn (N)	0.192	5.1	2.8	6.7	5.1	6.8	6.5	6.6	7.1
Asp (D)	0.197	5.6	5.6	5.0	5.1	7.9	5.4	7.1	6.0
Cys (C)	0.206	2.3	3.9	4.4	4.5	4.5	4.9	6.0	6.0
Gln (Q)	0.186	4.3	5.1	4.4	6.8	4.0	4.3	4.9	4.9
Glu (E)	0.183	6.4	7.9	8.3	7.9	5.1	7.6	6.0	6.5
Gly (G)	0.175	5.6	8.4	7.8	7.9	7.9	8.2	7.7	8.7
His (H)	0.219	2.6	2.2	1.7	1.7	2.8	1.6	1.6	1.6
Ile (I)	0.179	5.3	5.6	2.8	3.4	5.1	4.9	3.3	3.3
Leu (L)	0.173	8.7	5.1	5.0	1.7	5.1	1.1	1.1	2.2
Lys (K)	0.181	6.5	5.6	3.9	3.4	1.7	3.8	4.4	4.3
Met (M)	0.204	2.5	2.8	11.7	14.7	5.6	13.6	14.8	14.7
Phe (F)	0.244	3.8	5.1	5.0	6.8	3.4	4.9	4.9	4.3
Pro (P)	0.165	4.9	4.5	3.9	2.3	3.4	1.6	1.6	1.6
Ser (S)	0.170	8.3	6.2	7.8	7.3	10.7	9.8	9.8	8.7
Thr (T)	0.172	6.3	3.9	3.9	1.7	0.6	1.1	1.6	2.2
Trp (W)	0.277	1.1	2.2	0.6	0.6	1.1	0.5	0.5	0.5
Tyr (Y)	0.240	3.1	7.9	5.0	5.6	7.9	7.6	6.6	6.5
Val (V)	0.172	6.6	3.9	2.2	1.7	2.8	2.2	1.6	1.6

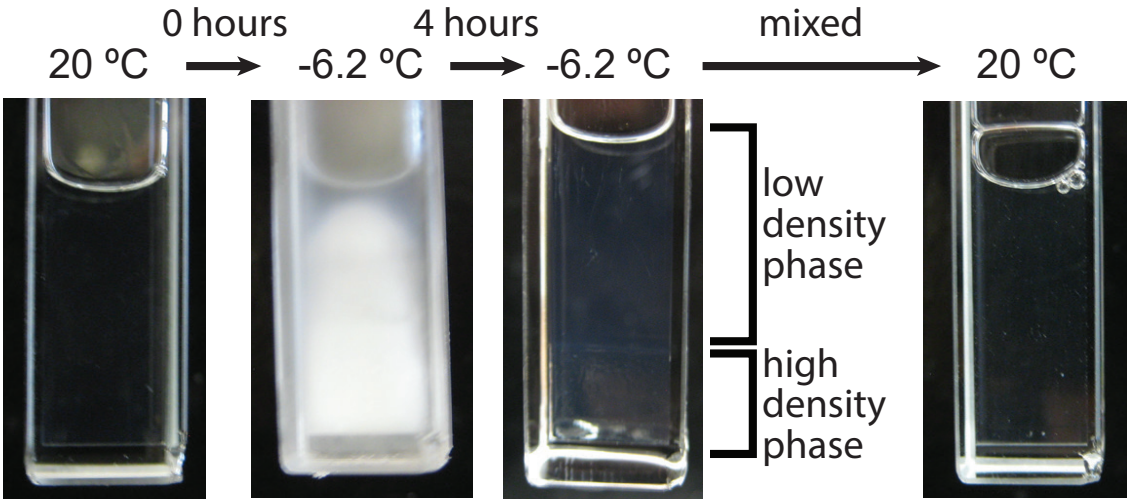

Figure A.5: Images of 150 mg/mL wild-type γM8b. Left image is sample incubated at 20 °C, where it is completely transparent. Middle left image is the sample incubated at -6.2 °C for a few minutes. The solution rapidly becomes turbid once phase separation occurs. Middle right image is the sample after 4 hours incubation at -6.2 °C. Here, the solution loses almost all turbidity as the high-density phase settles to the bottom. Right image is the sample warmed to 20 °C and mixed. The solution becomes transparent again, displaying the reversibility of the phase transition.

Table A.3: First three rows are calculated from urea denaturation, which is shown to be reversible. The fourth row is calculated from the thermal denaturation plot. The last two rows are calculated from the binodal fit to the LLPS data.

	HγS	γM1	γM4	γM7	γM8b	γM8c	γM8d
$[urea]_{1/2}$ (M)	6.3[67]	4.6	4.6	4.5	5.7	4.2	4.1
ΔG°_{w} (kcal/mol)	6.5	10.1	6.8	8.5	10.4	6.7	4.2
m (kcal/mol/M)	1.03	2.27	1.48	1.93	1.84	1.60	0.99
T_m (°C)	72.0[47]	64.3	69.3	59.0	71.3	62.6	61.7
T_c (°C)	-	-7.9	-12.4	-4.4	-4.8	-8.1	-9.6
C_c (mg/mL)	-	268	191	230	211	219	220

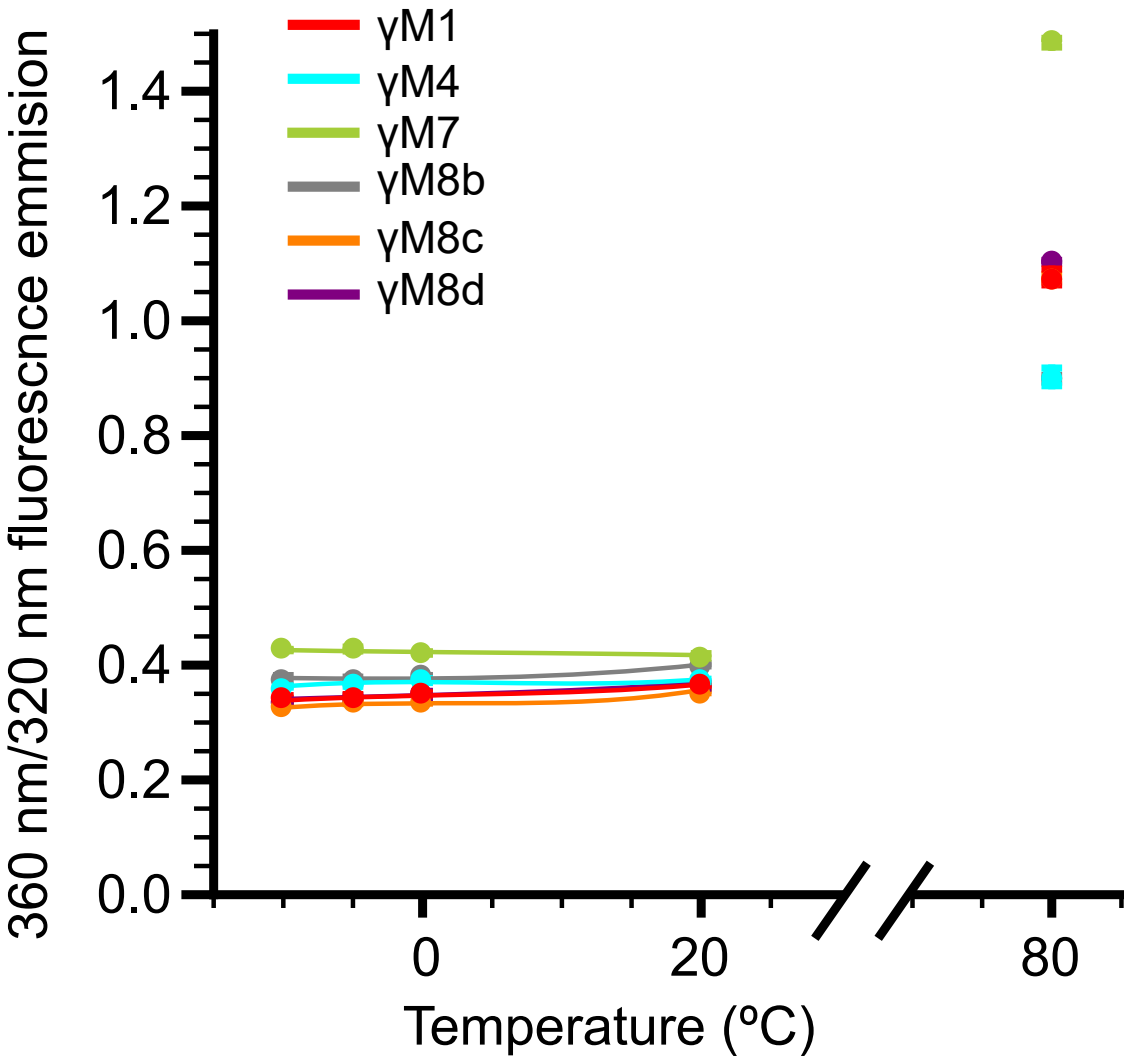

Figure A.6: Intrinsic fluorescence emission (360 nm/320 nm ratio) of γM1 (red), γM4 (cyan), γM7 (lime green), γM8b (gray), γM8c (orange), and γM8d (purple) taken at different temperatures. The 360/320 emission ratio is considerably higher at high temperatures where the protein is unfolded and the tryptophans are exposed to solvent. However, from 20°C down to -10°C each γM retains a stable 360/320 ratio between 0.3 and 0.5 indicating that the proteins are not undergoing denaturation over the experimentally relevant temperature range.

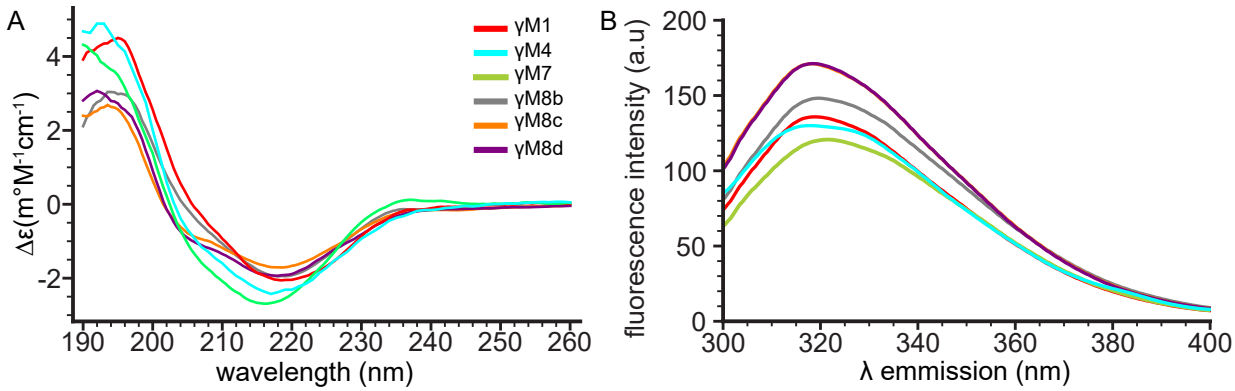

Figure A.7: (A) Circular dichroism spectra of γM1 (red), γM4 (cyan), γM7 (lime green), γM8b (gray), γM8c (orange), and γM8d (purple). These spectra contain minima between 216 and 220 nm, indicative of primarily β-sheet secondary structure, which is consistent with the $\beta\gamma$ crystallin fold. (B) Intrinsic fluorescence emission spectra were collected with an excitation wavelength of 280 nm. The emission maxima are 319, 318, 322, 320 318, and 319 nm, respectively. An emission wavelength closer to 320 than 360 nm indicates that the tryptophans are buried within a hydrophobic environment. Interestingly, γM7 has comparable emission to the other γMs despite having an additional Trp relative to the other γMs. This is consistent with findings in other γ-crystallins, which display quenching of one of the two tryptophans found in each domain. Furthermore, the quenched tryptophan has been implicated as an energy transfer acceptor from the other tryptophan in the same domain. Quenching is hypothesized to result from electron transfer from tryptophan to a backbone amide [48]. In the case of γM7, the additional Trp is in the same position as the quenched Trp observed in other γ-crystallins.

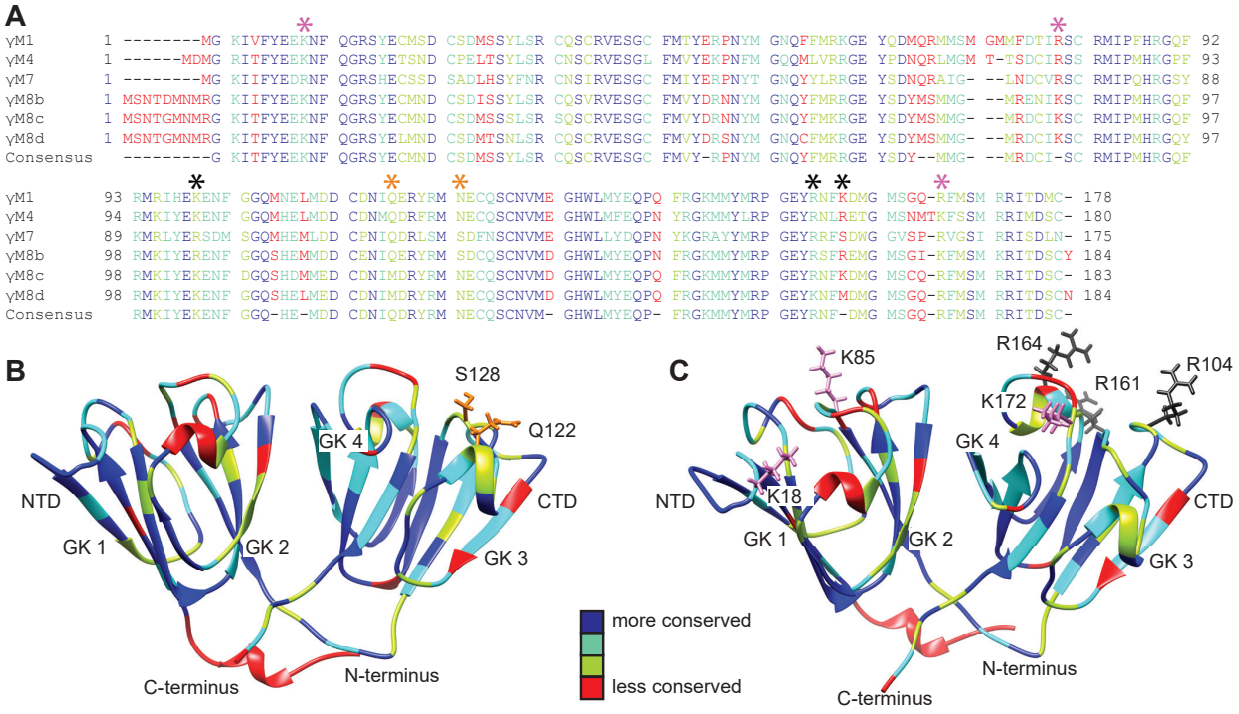

Figure A.8: (**A**) Amino acid sequence alignments of γM1, γM4 γM7, γM8b, γM8c, and γM8d (performed using Clustal Omega [199]). The sequences and models are color-coded as follows: blue: strictly conserved, cyan: only one sequence deviates, green: two sequences deviate, red: three or more sequences deviate from the consensus. The residues chosen for mutagenesis in γM8b are shown highlighted with orange, black and pink. (**B**) The positional coloring from the sequence alignment was superimposed onto a comparative model of γM8b [107]. Representation are depicted with the C-terminal domain (CTD) and N-terminal domain (NTD), as well as the four Greek key motifs (GK 1-4) denoted. The mutated residues, Q122 and S128 are colored orange (**C**) Same model as seen in B but K18, K85, and K172 are colored pink and R104, R161, and R164 are colored black.

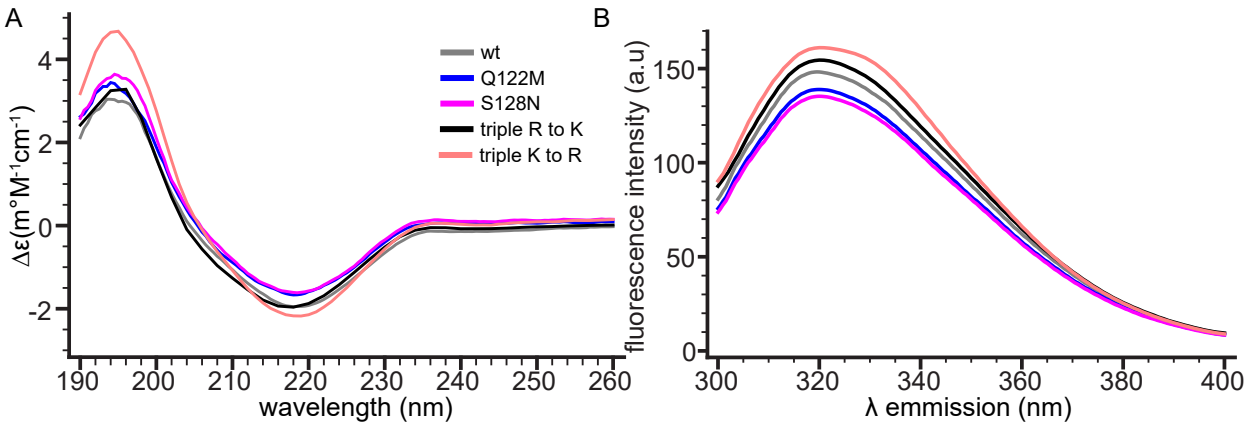

Figure A.9: (A) Circular dichroism spectra of wild-type γM8b (gray) and its variants Q122M (blue), S128N (magenta), triple R to K (black), and triple K to R (pink). The spectra of wt and the variants have very similar profiles, indicating that the mutations cause very little perturbation in the secondary structure in all cases. (B) Fluorescence emission spectra of wild-type γM8b compared to its variants excited at 280 nm. The spectra of the wt and variant proteins have approximately the same maxima and comparable intensities indicating that the tryptophan environments are minimally perturbed by the mutations.

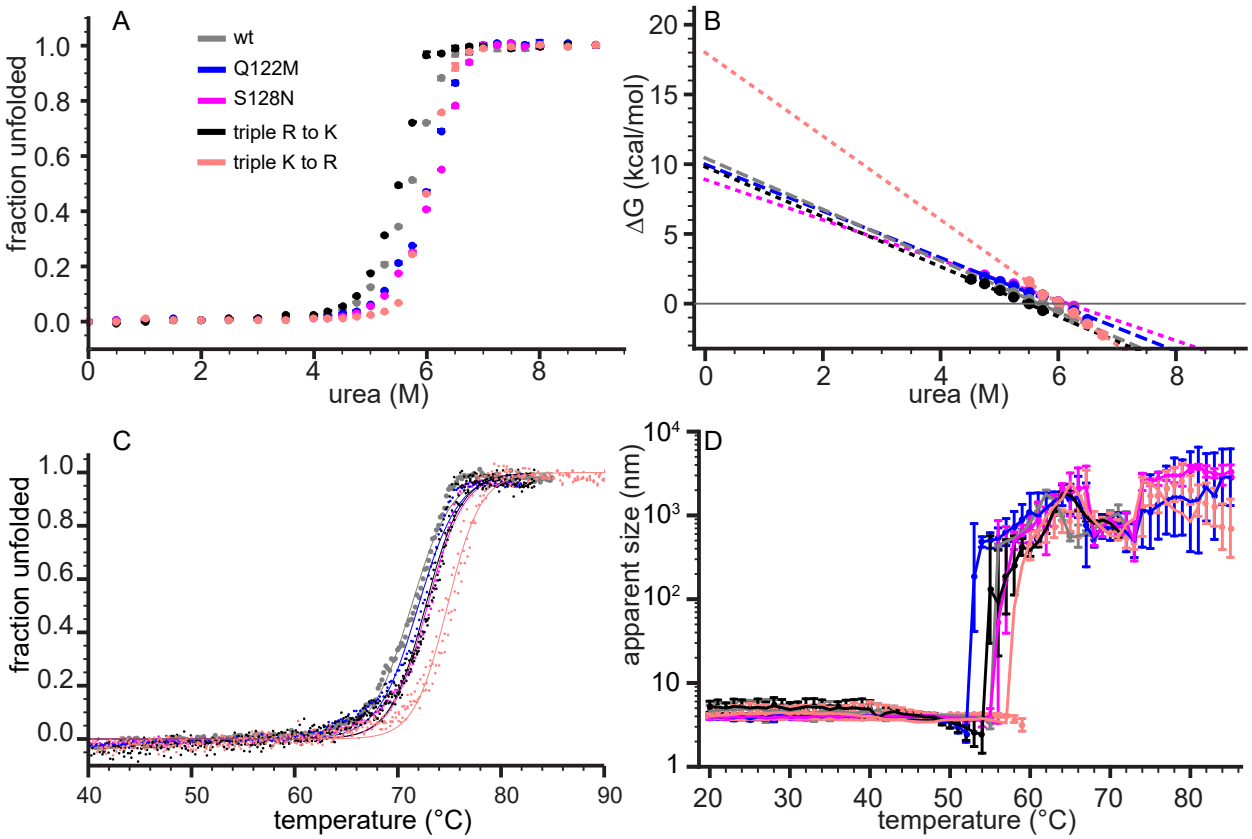

Figure A.10: (A) Urea denaturation of wild-type γM8b (gray) and its variants Q122M (blue), S128N (magenta), triple R to K (black), and triple K to R (pink). The wt and variants show similar urea denaturation curves, with $[urea]_{1/2}$ concentrations of 5.7, 6.0, 6.2, 5.5, and 6.0 M, respectively. (B) Gibbs free energy plot as a function of urea concentration with ΔG°_{w} 10.4, 10.0, 9.9, 9.8, and 18.0 kcal/mol, respectively. (C) Thermal denaturation of γM8b and its variants. The denaturation temperatures of the wt and variant proteins are 71.3, 72.0, 72.9, 72.7 and 74.8 °C, respectively. (D) Thermal aggregation of γM8b and its variants. Wild-type and S128N begin to aggregate around 56 °C, while the Q122M and triple R to K variants aggregate at slightly lower temperatures: 53 and 55 °C respectively.

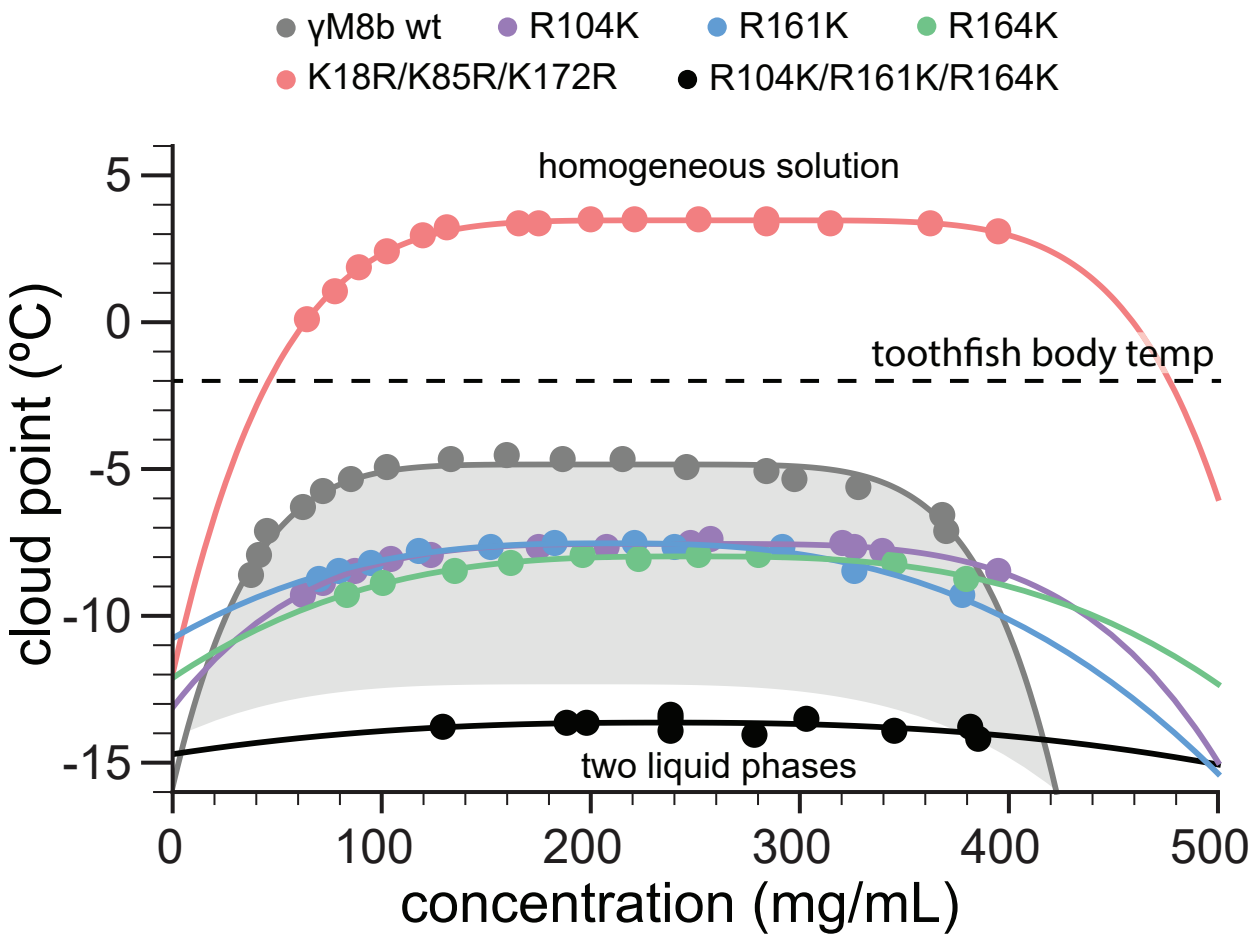

Figure A.11: The same coexistence curves as in Fig 7B are shown, with the addition of the single variants R104K (light purple), R161K (blue) and R164K (light green). All the single variants have similar critical concentrations: 241, 208, and 248 mg/mL respectively. Their critical temperatures are -7.6, -7.6 and -8.0 °C respectively; the change in magnitude for each one is approximately one third of that observed for the triple variant.

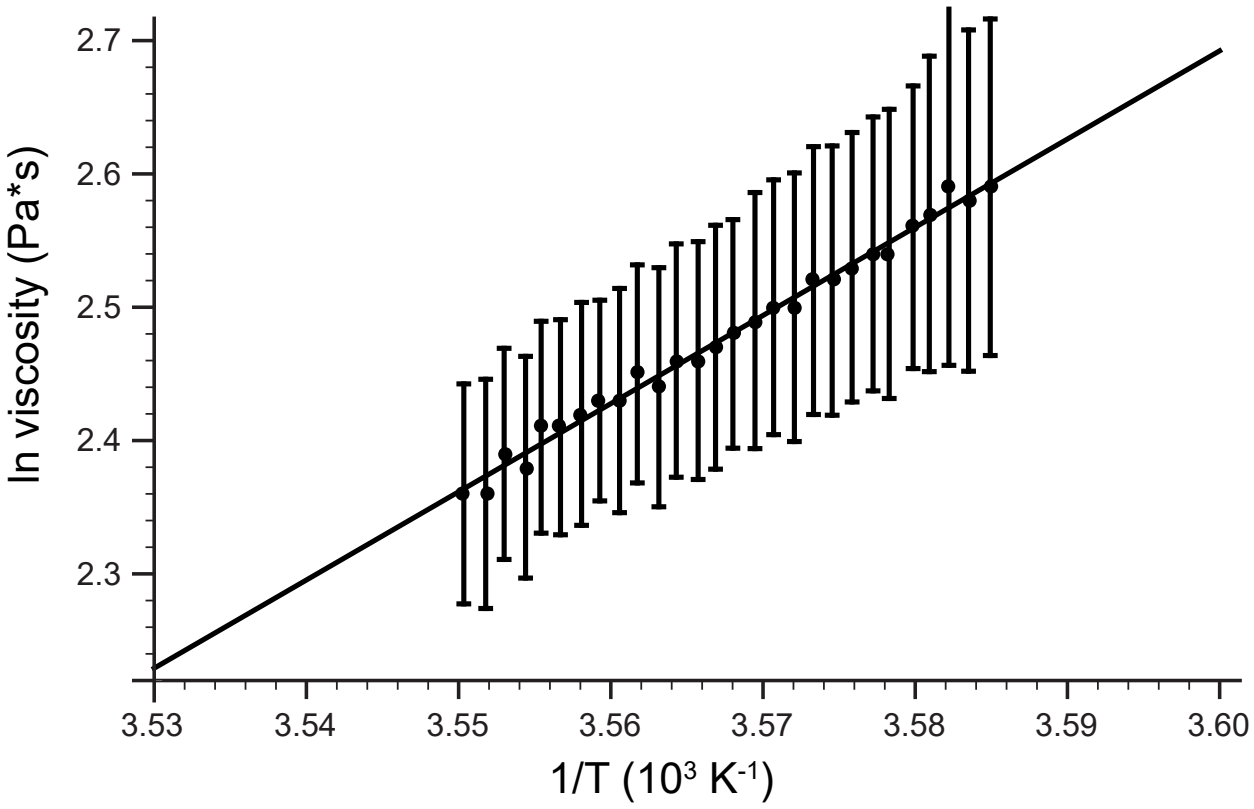

Figure A.12: The natural logarithm of viscosity of a 150 mg/mL solution of γM8b triple K to R variant plotted as a function of 1/T. The data was fit to the Arrhenius equation [105].

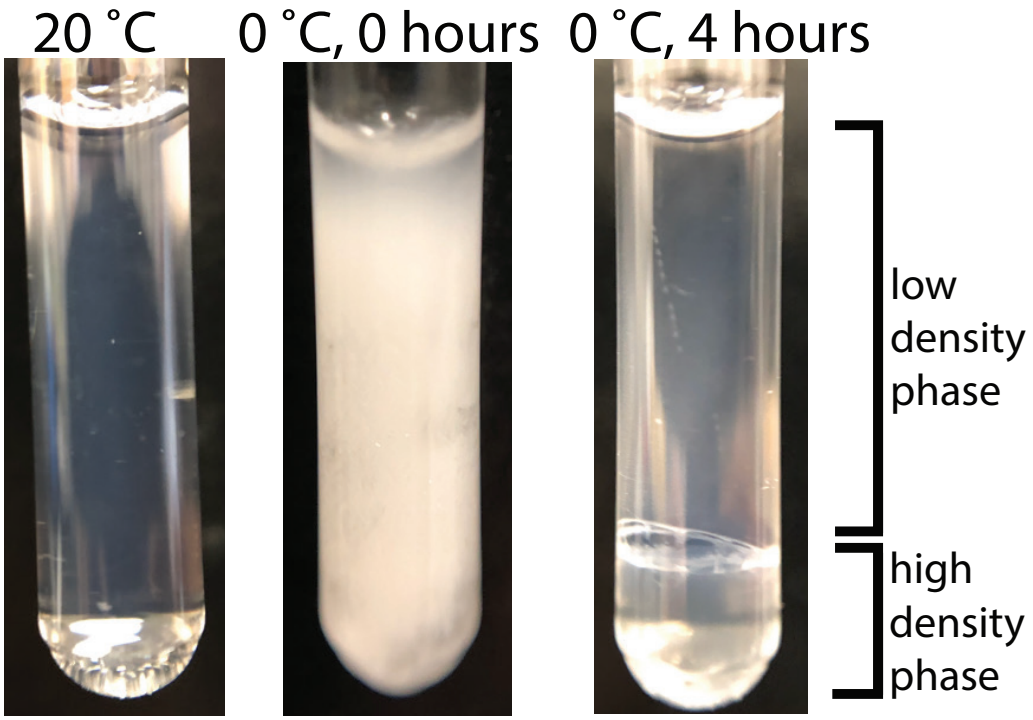

Figure A.13: Images of 150 mg/mL γM8b triple Lys to Arg solution. Left image is sample incubated at 20 °C, where it is completely transparent. Middle image is the sample incubated at 0 °C for a few minutes. The solution rapidly becomes turbid once phase separation occurs. Right image is the sample after 4 hours incubation at 0 °C. Here, the solution loses almost all turbidity as the high-density phase settles to the bottom.

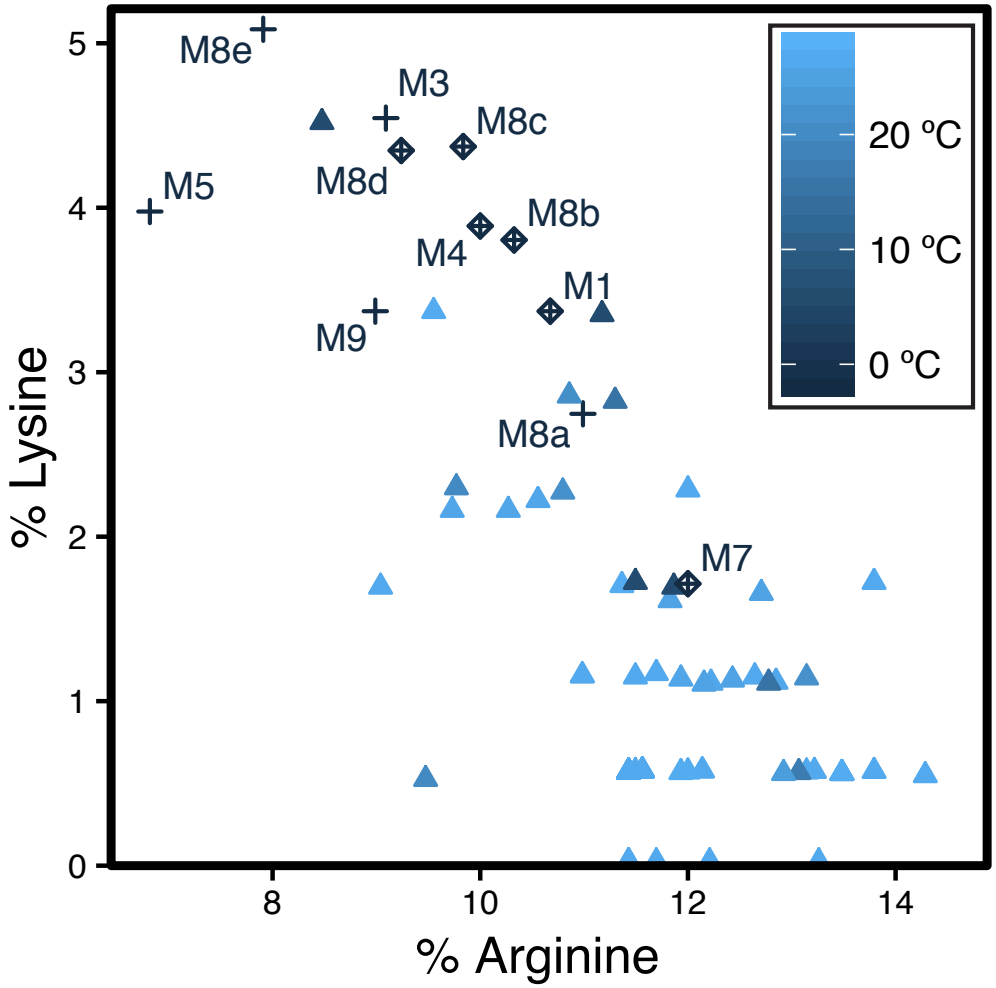

Figure A.14: Arginine and lysine content of γM-crystallins from 18 fish species were analyzed. The sequences were were collected from UniProt (https://www.uniprot.org/) with filters for proteins between 160 and 200 amino acids. The entries were manually assessed to eliminate incomplete sequences, fragments, "like" proteins, and non-fish γM-crystallins. The coloring denotes the approximate environmental temperature for each fish species. Crosses represent so far uncharacterized sequences from *D. mawsoni*, diamonds represent *D. mawsoni* proteins for which LLPS has been measured, and triangles represent proteins from other fish species. In all cases Arg is more predominant than lysine. Many of the toothfish γM-crystallins show a shift towards higher lysine content compared to other fish species; however there is not a strong correlation of Arg/Lys content with environmental temperature.

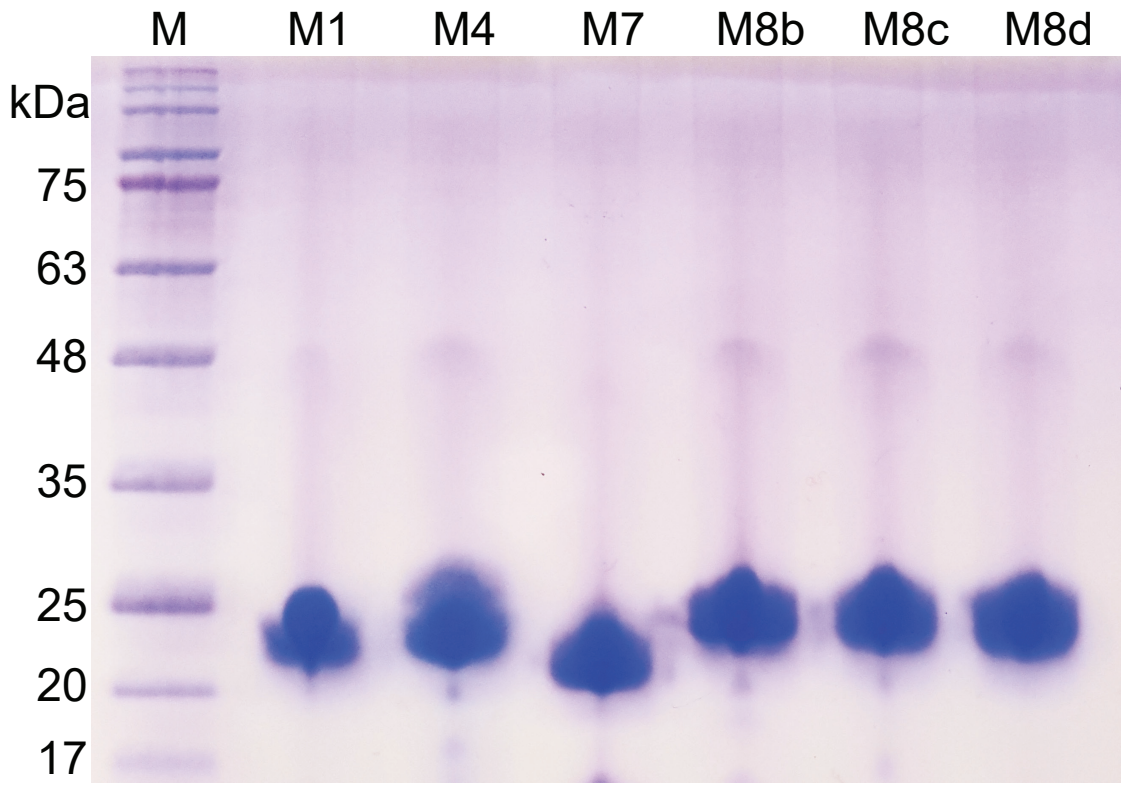

Figure A.15: Purified γM-crystallins were run on tris-glycine SDS-PAGE and stained with Coomassie brilliant blue. All proteins are at the expected size with little other intensity in the lanes.

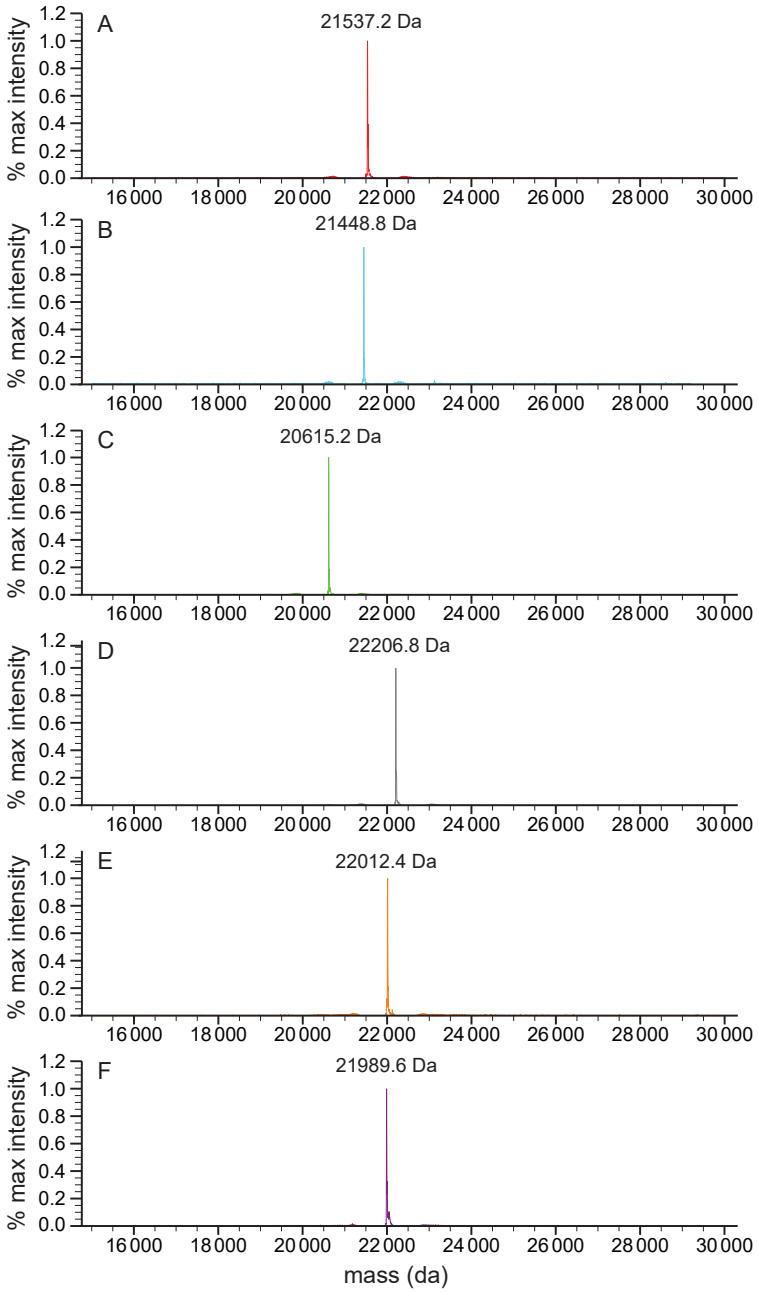

Figure A.16: Purified γM-crystallins were applied to a Xevo G2-XS QTof spectrometer (Waters, Milford, MA) and run in positive mode. The resulting charge envelopes were deconvoluted using the maxent1 function of MassLynx software (Waters, Milford, MA). (A) γM1 measured mass of 21537.2 Da is in good agreement with the predicted 21538.81 Da. (B) γM4 measured mass of 21448.8 Da is in good agreement with the predicted 21450.53 Da. (C) γM7 measured mass of 20615.2 Da is in good agreement with the predicted 20617.97 Da. (D) γM8b measured mass of 22206.8 Da is in good agreement with the predicted 22208.37 Da. (E) γM8c measured mass of 22012.4 Da is in good agreement with the predicted 22014.16 Da. (F) γM8d measured mass of 21989.6 Da is in good agreement with the predicted 21991.21 Da.

Appendix B

D. capensis D1-PSI

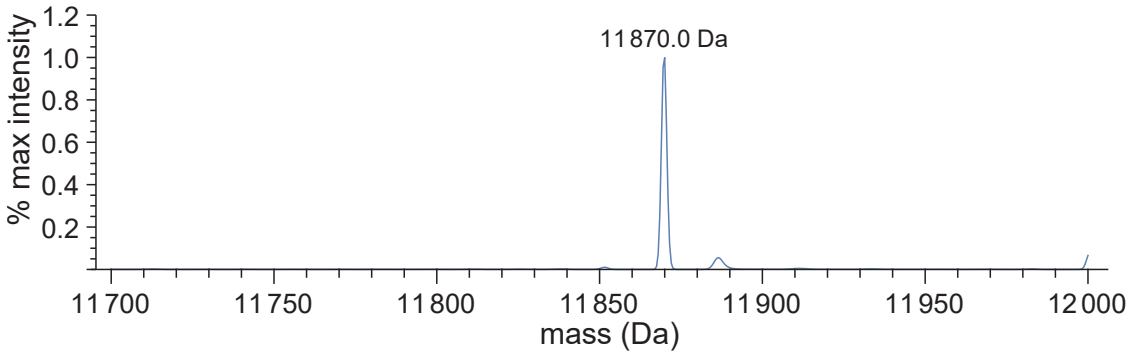

Figure A.1: Purified D1-PSI was applied to a Xevo G2-XS QTof spectrometer (Waters, Milford, MA) and run in positive mode. The resulting charge envelope was deconvoluted using the maxent1 function of MassLynx software (Waters, Milford, MA). The measured mass of 11870.0 Da is in good agreement with the predicted mass of 11868.5 Da for the pure protein containing three disulfide bonds.

Appendix C

Other *D. capensis* proteins

C.1 Cysteine protease 0624

C.1.1 Gene Construction, Expression, and Purification

Plasmids containing the DNA sequences for *Dissotichus mawsoni* cysteine protease 0624 was purchased from Intergrated DNA Technologies (Coralville, IA). The gene encodes for positions 29 to 326 of the full length protease sequence to exclude predicted plant signal sequence which was predicted by SignalP4.1 server [200]. The N-terminal sequence of the gene was appended with the signal sequence (MKLSTNLILAIAAASAVVSA) from *P. pastoris* protein EPX1 (GenBank, CCA40103)[201] and the kozac sequence of *P. pastoris* AOX1 (GenBank, CAA26484). The gene also contained flanking N-terminal and C-terminal EcoRI and XhoI restrictions sites respectively. The gene was amplified using oligonucleotide primers purchased from Intergrated DNA Technologies (Coralville, IA), and the resulting gene product was cloned into pPICZ A vector (Thermo Fisher Scientific, Waltham, MA). The vector was transformed into competent DH5-alpha *Escherichia coli* (Thermo Fisher Scientific, Walt-

ham, MA) for vector amplification. Plasmid vector was linearized using PmeI endonuclease (New England Biolabs, Ipswich, MA) and transformed into electrocompetent Superman$_5$ *Pichia pastoris* cells (Biogrammatics, Carlsbad, CA) that were prepared according to the protocol of De Schutter and Callewaert[202]. Electroporation was carried out in a 2 mm electroporation cuvette at 1.5 kV, 200 ω and 25 μF before being suspended 1 mL 1 M sorbitol, 1% yeast extract, 2% peptone, and 2% dextrose (YPD) plus 50 μg/mL ampicillin and then let sit at 30 °C for 30 minutes. The cells were then allowed to recover with shaking for an additional 5 hours before being plated on YPD plates containing Zeocin. Plasmid integration was confirmed by colony PCR and high expressing colonies were selected for. For protein expression cells were grown in 1% yeast extract, 2% peptone, and 2% glycerol (YPG) at 30 °C until an OD of approximately 1, after which cells were spun down at 1,500 RCF and resuspended in an equivalent volume of BMMY mediaciteWeidner2010, pH 7.5 at a final concentration of 1% methanol. Cells were allowed to grow for 72 hours at 28 °C with feeding of 1% methanol and 1 mM ammonium sulfate every 12 hours. Cells were then pelleted by centrifugation and the spent media was dialyzed against dry PEG 8000 to concentrate the media. Protein was then precipitated using 70% saturated ammonium sulfate precipitation. The precipitated protein was then resuspended and dialyzed against 10 mM imidazole, 0.025% NaN_3, pH 7.5 followed by anion exchange chromatography on an UNOsphere Q column purchased from Bio-Rad (Hercules, CA) using a 1 M sodium chloride gradient. The final purification step was application to a HiLoad 16/600 Superdex 75 PG gel filtration column from GE (Pittsburgh, PA) using 10 mM sodium phosphate buffer, 100 mM sodium chloride, 0.05% sodium azide at pH 6.9.

C.2 Chitinase 0106

C.2.1 Gene Construction, Expression, and Purification

Plasmids containing the DNA sequences for *Dissotichus mawsoni* chitinase 0106 was purchased from Intergrated DNA Technologies (Coralville, IA). The gene encodes for positions 26 to 300 of the full length chitinase sequence to exclude predicted plant signal sequence which was predicted by SignalP4.1 server[200]. The N-terminal sequence of the gene was appended with the signal sequence (MKLSTNLILAIAAASAVVSA) from *P. pastoris* protein EPX1 (GenBank, CCA40103)[201] and the kozac sequence of *P. pastoris* AOX1 (GenBank, CAA26484). The gene also contained flanking N-terminal and C-terminal EcoRI and XhoI restrictions sites, respectively. The gene was amplified using oligonucleotide primers purchased from Intergrated DNA Technologies (Coralville, IA), and the resulting gene product was cloned into pPICZ A vector (Thermo Fisher Scientific, Waltham, MA). The vector was transformed into competent DH5-alpha *Escherichia coli* (Thermo Fisher Scientific, Waltham, MA) for vector amplification. Plasmid vector was linearized using PmeI endonuclease (New England Biolabs, Ipswich, MA) and transformed into electrocompetent Superman$_5$ *Pichia pastoris* cells (Biogrammatics, Carlsbad, CA) that were prepared according to the protocol of De Schutter and Callewaert[202]. Electroporation was carried out in a 2 mm electroporation cuvette at 1.5 kV, 200 ω and 25 μF before being suspended 1 mL 1 M sorbitol, 1% yeast extract, 2% peptone, and 2% dextrose (YPD) plus 50 μg/mL ampicillin and then let sit at 30 °C for 30 minutes. The cells were then allowed to recover with shaking for an additional 5 hours before being plated on YPD plates containing Zeocin. Plasmid integration was confirmed by colony PCR and high expressing colonies were selected for. For protein expression cells were grown in 1% yeast extract, 2% peptone, and 2% glycerol (YPG) at 30 °C until an OD of approximately 1, after which cells were spun down at 1,500 RCF and resuspended in an equivalent volume of BMMY mediaciteWeidner2010, pH 6 at

a final concentration of 1% methanol. Cells were allowed to grow for 72 hours at 28 °C with feeding of 1% methanol and 1 mM ammonium sulfate every 12 hours. Cells were then pelleted by centrifugation and the spent media was dialyzed against dry PEG 8000 to concentrate the media. Protein was then precipitated using 70% saturated ammonium sulfate precipitation. The precipitated protein was then resuspended and dialyzed against 10 mM imidazole, 0.025% NaN_3, pH 7 followed by anion exchange chromatography on an UNOsphere Q column purchased from Bio-Rad (Hercules, CA) using a 1 M sodium chloride gradient. The final purification step was application to a HiLoad 16/600 Superdex 75 PG gel filtration column from GE (Pittsburgh, PA) using 10 mM sodium phosphate buffer, 100 mM sodium chloride, 0.05% sodium azide at pH 6.9.

C.3 Chitinase 4817

C.3.1 Gene Construction, Expression, and Purification

Plasmids containing the DNA sequences for *Dissotichus mawsoni* chitinase 4819 was purchased from Intergrated DNA Technologies (Coralville, IA). The gene encodes for positions 21 to 325 of the full length chitinase sequence to exclude predicted plant signal sequence which was predicted by SignalP4.1 server[200]. The N-terminal sequence of the gene was appended with the signal sequence (MKLSTNLILAIAAASAVVSA) from *P. pastoris* protein EPX1 (GenBank, CCA40103)[201] and the kozac sequence of *P. pastoris* AOX1 (GenBank, CAA26484). The gene also contained flanking N-terminal and C-terminal EcoRI and XhoI restrictions sites, respectively. The gene was amplified using oligonucleotide primers purchased from Intergrated DNA Technologies (Coralville, IA), and the resulting gene product was cloned into pPICZ A vector (Thermo Fisher Scientific, Waltham, MA). The vector was transformed into competent DH5-alpha *Escherichia coli* (Thermo Fisher Scientific, Walt-

ham, MA) for vector amplification. Plasmid vector was linearized using PmeI endonuclease (New England Biolabs, Ipswich, MA) and transformed into electrocompetent Superman$_5$ *Pichia pastoris* cells (Biogrammatics, Carlsbad, CA) that were prepared according to the protocol of De Schutter and Callewaert[202]. Electroporation was carried out in a 2 mm electroporation cuvette at 1.5 kV, 200 ω and 25 μF before being suspended 1 mL 1 M sorbitol, 1% yeast extract, 2% peptone, and 2% dextrose (YPD) plus 50 μg/mL ampicillin and then let sit at 30 °C for 30 minutes. The cells were then allowed to recover with shaking for an additional 5 hours before being plated on YPD plates containing Zeocin. Plasmid integration was confirmed by colony PCR and high expressing colonies were selected for. For protein expression cells were grown in 1% yeast extract, 2% peptone, and 2% glycerol (YPG) at 30 °C until an OD of approximately 1, after which cells were spun down at 1,500 RCF and resuspended in an equivalent volume of BMMY media[203], pH 7 at a final concentration of 1% methanol. Cells were allowed to grow for 72 hours at 28 °C with feeding of 1% methanol and 1 mM ammonium sulfate every 12 hours. Cells were then pelleted by centrifugation and the spent media was dialyzed against dry PEG 8000 to concentrate the media followed by dialysis against 10 mM acetate, 0.025% NaN_3, pH 5. The chitinase from the resulting media was subjected to cation exchange chromatography on an UNOsphere S column purchased from Bio-Rad (Hercules, CA) using a 1 M sodium chloride gradient. The final purification step was application to a HiLoad 16/600 Superdex 75 PG gel filtration column from GE (Pittsburgh, PA) using 10 mM sodium phosphate buffer, 100 mM sodium chloride, 0.05% sodium azide at pH 6.9.