SUPERIOR TO ONE OF GLASS: NATURAL GRADIENT INDEX LENSES VIA PATCHY PARTICLE SELF-ASSEMBLY

Jing Cai

A DISSERTATION

in

Physics and Astronomy

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2015

Supervisor of Dissertation

_____

Alison M. Sweeney, Assistant Professor of Physics and Astronomy

Graduate Group Chairperson

_____

Marija Drndic, Professor of Physics and Astronomy

Dissertation Committee

Randall D. Kamien, Professor of Physics and Astronomy

Paul A. Heiney, Professor of Physics and Astronomy

Mark Goulian, Professor of Physics and Astronomy

Burt A. Ovrut, Professor of Physics and Astronomy

SUPERIOR TO ONE OF GLASS: NATURAL GRADIENT INDEX LENSES VIA PATCHY PARTICLE SELF-ASSEMBLY

ACKNOWLEDGMENTS

I have been fortunate to have the opportunity to complete this dissertation at the Department of Physics and Astronomy at University of Pennsylvania. I was encouraged, guided and supported by the teachers, colleagues, and friends in this place. I never would have made it this far without all of you by my side.

Three years ago, I was lost and searching for my research directions. Alison Sweeney impressed me with an amazing new world of marine biology. After talking to her, I decided to pursue my Ph.D. under her guidance. During these years, she was extremely encouraging, patience, and instructive. She guided me in every single step for my research and eventually fostered me into a mature and confident scientist. She is always around and always willing to help, from teaching me how to perform SDS-PAGE to discussing with me the future directions. I appreciate all her help and guidance, and I would like to express my extreme gratitude to her. She is a both great adviser and a life-long friend. I was very fortunate to have her as my adviser, and without her, this dissertation would not have even started.

Of course, research is a collaborative process. During my school years, I am grateful to all the guidance and assistance from professors, postdocs, and graduate students.

When my project started, I was searching a method to measure directly the proteins in the intact squid lens. Professor Paul Heiney gave me endless help on that. Not only did he taught me how to use our in-house x-ray equipment, he also went to the Brookhaven National Lab three times with me to help me collecting data. Though this work was very intensive and tiring, Paul helped me for every detail for the whole trip, including application, data collection, and data analysis. I would like to thank Paul for all of his effort, hard work, and support to make this dissertation possible!

I would like to thank the rest of my dissertation committee members: Prof. Randall Kamien, Prof. Mark Goulian and Prof. Burt Ovrut. Thank all of you for the discussions, questions, critisims, and advice! I also want to thank Prof. Adam Lidz for discussing with me the three-dimensional Fourier transform.

The x-ray scattering experiment is a big part of my thesis. To perform the experiment, I got a lot of help from Lin Yang at the Brookhaven National Lab, Scott Barton at the Molmex Scientific Inc., and Hao-Jan Sun here at UPenn. Their effort and support made my experiments possible, and I would like to thank all of them.

My research is focusing on studying the eyes of squids. I would also want to thank all the squids who sacrifice their lives for science!

Through these years during my years in graduate school, I received a lot of assistance from many other students and postdocs as well. I am very grateful to all members in the Sweeney's lab, who performed additional experimental analysis and participated in numerous scientific discussions. In particular, I would like to thank James Townsend for conducting experiment at Brookhaven National Lab, for useful discussions, and for English help. I would like to thank Tom Dodson for discussion, advise, and English help. I would also want to express my gratitude to the Yodh's soft matter group members. Thank Dr. Ye Xu, Dr. Tim Still, and Weishao Wei for allowing me to 'borrow' their lab support, and providing me with guidance on rheology measurements, light scattering experiments, and cross linking PDMS. I was also grateful to my friend Matt Malloy for Matlab and English help, Xiaqing Zhao for helping me with the evolution theory, and Annemarie Exarhos for Latex help.

I would like to thank my boyfriend Yue Cao for always being encouraging and supportive for my research.

Last but not least, I would like to thank my families. My parents Yi Cai and Aiying Dong raised me and foster my curiosity to explore the world. They have been supporting

me for the longest and the most. I can never fully thank them for everything they have done for me.

ABSTRACT

SUPERIOR TO ONE OF GLASS: NATURAL GRADIENT INDEX LENSES VIA PATCHY PARTICLE SELF-ASSEMBLY

Jing Cai

Alison M. Sweeney

To achieve acute and sensitive vision in a camera-like eye in the ocean, a graded refractive index spherical lens is required to maximize the photon flux on the retina, while leaving an eye structure that fits in the head of an animal. This biological lens must also maintain low protein density fluctuation at the length-scale of a wavelength of visible light in order to maintain transparency. In squids, this sophisticated optical design emerges from the properties of a single protein fold, the S-crystallin. In this thesis, I study the material properties and the self-assembly of the squid lens system. I show that squids have evolved graded index and low density fluctuation in a spherical lens using a suite of proteins that can act as patchy colloids with specific, low valence ($M = 2$ or $M = 3$) with geometric flexibility in bond angles. We conducted small x-ray scattering (SAXS) at different radial positions of the lens, and performed a Monte Carlo simulation to estimate structures consistent with the SAXS result. This analysis suggests that lens proteins may form a gel with gradient density throughout the cellular lens structure, with density mediated by a tightly controlled protein coordination number in each region of the organ. Patchy colloid theory may therefore explain both the graded refractive index lens and the transparency evolved in the lens. I also studied the Chinese century egg, which appears to be a physically analogous system of a protein-based, low-valence patchy colloidal gel that was developed in prehistoric Chinese culinary culture as a method of egg preservation. I compare the structure and material properties of these two systems.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Have you seen snowflakes? They are the graceful six-arm fairies falling from the sky. If you watch closely, you will find that a snowflake is most likely to be transparent, and it prefers to have a six-arm structure. These elegant flakes are single crystals with molecules packing in long-range order. Their macroscopic six-arm shape reflects the structure of the water molecules: the angle between two hydrogen-oxygen bonds is 104.5° [67]. This angle is close to 120°, which is the angle required for a hexagonal lattice. This is why the water molecules adapt the hexagonal form to construct their crystal. As a result, the ice Bravais lattice, which is the minimum repeating unit in a crystal, is a hexagonal lattice with 6-fold symmetry. The snowflakes carry this symmetry, showing the splendid six-arm shape.

However, the ordered patterns observed in snowflakes are generated from initially random and disordered water molecules in the liquid phase. These water molecules were originally homogeneously distributed in space before they froze. Then, when the temperature drops or nucleation occurs, more-ordered hydrogen bonds between the water molecules start to form. When the hydrogen bonds are so numerous that the thermal movement of the molecules are restricted, the elegant snowflakes start to appear. During this process, ordered patterns are developed from initially disordered molecules. This kind of organization,

where systematic patterns form from initially disordered particles with certain interactions, is self-assembly.

1.1 Self-assembly

Self-assembly is the process where ordered patterns form from initially disordered elements [18]. This process abounds in nature at all scales, from crystal formation to planetary systems [127]. Because it happens so often, it has attracted more and more attention from researchers. Attention has been focused on different aspects of self-assembly throughout time. Nowadays, self-assembly research is mostly focused on a few branches [127], which include 1. designing and building artificial materials with controlled properties of the constituent particles, and 2. understanding pattern formations using new theories of self-assembly.

With self-assembly as a guide, a vast number of new materials with novel properties have been discovered and fabricated [3, 26, 52, 60, 76, 84, 91, 111, 132]. One example is self-assembling quantum dots which can be synthesized with a variety of materials [3]. These dots have special fluorescence properties [84] and can be used in living cells for diagnostics and imaging [76]. Further, long-range order in nanocrystal superlattices can be self-assembled by quantum dots with careful control of the size of monodispersed nanoparticles [111] or a mixture of nanoparticles with different sizes [26]. This superlattice provides a tuning method for the interactions between the quantum dots, and collective behaviors between different quantum dot components have been observed. With novel magnetic properties, this material is potentially of great use in industry.

Besides making new materials with self-assembly properties, a more fascinating aspect of self-assembly is to understand the patterns that nature presents us [91]. Specifically, life, in some sense, is a completely self-assembled result of systematically ordering and

combining amino acids, lipid molecules, ribonucleic acids, *etc.* [105]. In the dynamical system of the earth, the sun constantly pumps in energy, and when the first robust self-replicated form appeared, life and evolution might have started there. Through billions of years, life evolved into forms in which a complicated body is constructed with a huge diversity of small units — the cells [115]. A variety of cell properties and functions have been studied under the idea of self-assembly [115]. Each cell is a living factory which uses glucose as the energy source and achieves specific functions. Roughly speaking, they are composed with a similar structure: a cell membrane provides the cell with a proper chemical and physical environment; the cell nucleus stores all the genetic informations; the cell ribosomes and other organelle fabricate proteins based on the transcript from the genetic code.

Every single part of the cell can be viewed as a self-assembled system [62, 127], and these systems coordinate to make the cell a living entity with constant energy and material exchange with the environment. Even outside of living cells, in an aqueous environment, molecules with hydrophilic and hydrophobic moieties on either end, such as the phospholipids that compose cell membranes, will automatically assemble into the bilayer shape as observed in a living cell [6, 59]. DNA uses four simple sugar bases to carry enormous genetic information. The duplication and interpretation are all done systematically and accurately by self-assembly. Proteins are formed by self-assembly with specific folding rules, and perform a variety of functions in cells [29, 37, 82]. Therefore, to better understand life on Earth, one approach is to turn to new theories of self-assembly.

Not only is self-assembly necessary for us to understand life, it can guide us to harness what we have learned to fabricate complicated materials with sophisticated properties that existed initially only in biological systems [69, 91, 128, 132]. Biological systems can produce materials with an enormous amount of properties which can be useful in a lot of fields such as aeronautics, biotechnology, *etc.*. Instead of using the biomaterials directly, an

alternative way is to study the underlying physics which causes the material properties of these biological systems [45, 63]. Next, one can fabricate artificial materials with the same material properties and underling physics but using non-protein materials. The *Papilio blumei* butterfly wing has extraordinary reflective properties achieved by having multilayer stacks, which are mimicked using inorganic particles with a combination of layer deposition techniques [63]. Red roses' petal surfaces have hierarchical micropapillae to achieve superhydrophobic property, which has been mimicked using PVA film and nano-embossed structures duplicated from the petal's surface [45]. Therefore, it is beneficial to study the physics of self-assembly, and people can ultimately harness it to fabricate biomimic materials.

In short, we are living in a self-assembled world. Studying these mechanisms helps us understand ourselves as a form of life, and to fabricate new materials to better our life.

1.1.1 Patchy particle system

A self-assembled pattern reflects the properties of the particles that compose the pattern [1, 27]. Therefore, the key to understanding self-assembly is to study the properties of the particles. A vast majority of the particles that compose the ordered structures described above are attracted to each other. In most cases, these attractions are anisotropic: they are attracted to each other through only a few 'sticky spots' on them, but not the entire surface of the particle. These spots are like patches, and when they are considered theoretically, particles like this are called the 'patchy particles'.

Patchy particles are a general class of particles which interact with each other under strong anisotropic and directional interactions, which are mostly attractive [12]. These interactions are usually achieved by a few 'patches' on the surface of the particles [133]. In the definition, it is essential for a patchy particle to have anisotropic interactions [39]. For example, nanoparticles can be ligated with long sticky polymers to prevent them from

aggregation. The ligands themselves are anisotropically attractive to each other [40], and therefore the system can be treated as a patchy particle system. At a smaller length scale, water molecules have strong hydrogen bonds between the hydrogen and oxygen atoms. These bonds are strongly anisotropic and directional, serving as patches of the molecules. A patchy particle can also be self-assembled via a multiblock copolymer chain [130]. In contrast, particles with isotropic interactions, such as in a Lennard-Jones liquid [38], do not belong to the group of patchy particles.

Patchy particles can also be actively designed and fabricated so that they can self-assemble into artificially designed patterns [80, 134]. A Bernal spiral was self-assembled using patch-antipatch particles [81]. Colloidal spheres have been induced to self-assemble into a colloidal kagome lattice [27]. A new field of DNA-made patches is showing a promising future [44, 102].

In summary, patchy particle systems are a subset of self-assembled systems that show prodigious physical and chemical novelty. Studying the self-assembly process of the patchy particle system can be beneficial to manufacturing designed patterns for industrial uses.

1.1.2 Proteins as patchy particles

Life is assembled with an abundance varieties and numbers of proteins. The proteins are "folded" in their final form [37, 82], and the folding mechanism is strictly regulated in a cell [89]. When the polypeptide chain is fabricated on a ribosome [34], some part of it may be folded into α-helix or β-sheet [55]. These structures are called the protein secondary structure [88]. From there, these structural elements interact to form the final shapes of the single polymer protein monomer (tertiary structure), and these folded monomers may form dimers or oligomers (quaternary structure). All these hierarchical structural elements require attractive interactions between their substructures. One major group of proteins are the globular proteins, whose final, folded form is roughly spherical. These proteins

can have attractions to each other through a variety of biophysical mechanisms, such as hydrophobic interactions, Yukawa potential, *etc.*. Since these anisotropically distributed attractive interactions come from clusters of certain amino acids on the surface of the protein, these amino acid clusters can be treated as patches. The resulting protein system can be studied as a self-assembly of patchy particles, with patterns formed that may serve for different protein functions for biological systems.

These patchy proteins are so important to life that no cells can survive without them. For example, the muscle contraction process is achieved by having lines of myosins anchored to M-line moving in the opposite direction of lines of actin (F-actin) anchored to the Z-line inside sarcomere, which is the contraction unit of myofibril in skeletal muscles [22, 47, 107]. The critical part of the unit, the actin line, is the self-assembled result of the globular protein actin (Figure 1.1a) [30, 36, 57, 71, 92, 93, 106]. Each actin protein has hydrophobic patches and salt bridge patches which are attractive to the corresponding patches on another actin protein. Under this patchy interaction, actin proteins form a chain structure with a flat surface [87] that a myosin protein can walk on. The formation of the F-actin is also crucial for filopodia [72], which is a component of lamellipodia that control the movement of a cell [66].

Similarly, a microtubules in the cytoskeleton are self-assembled by tubulin, making a dynamic example of patchy particles. In a living cell, microtubules are polymerizing and depolymerizing constantly [35, 58, 85]. They are the cytoskeletal structure that is the major component of cilia and flagella, and are essential for mitosis. The protein component of the microtubules, the tubulin [100], has a similar structure to actin, except that the protein quaternary structure is a $\alpha\beta$ heterodimer and it forms hollow cylinders, as is shown in Figure 1.1b. This heterodimer is a unit cell to form the protofilaments. A microtubule is an assembly of about 13 parallel protofilaments, which is found *in vivo* [2]. The structures of α - and β - tubulin are very similar: each monomer is composed by two β-sheets in the

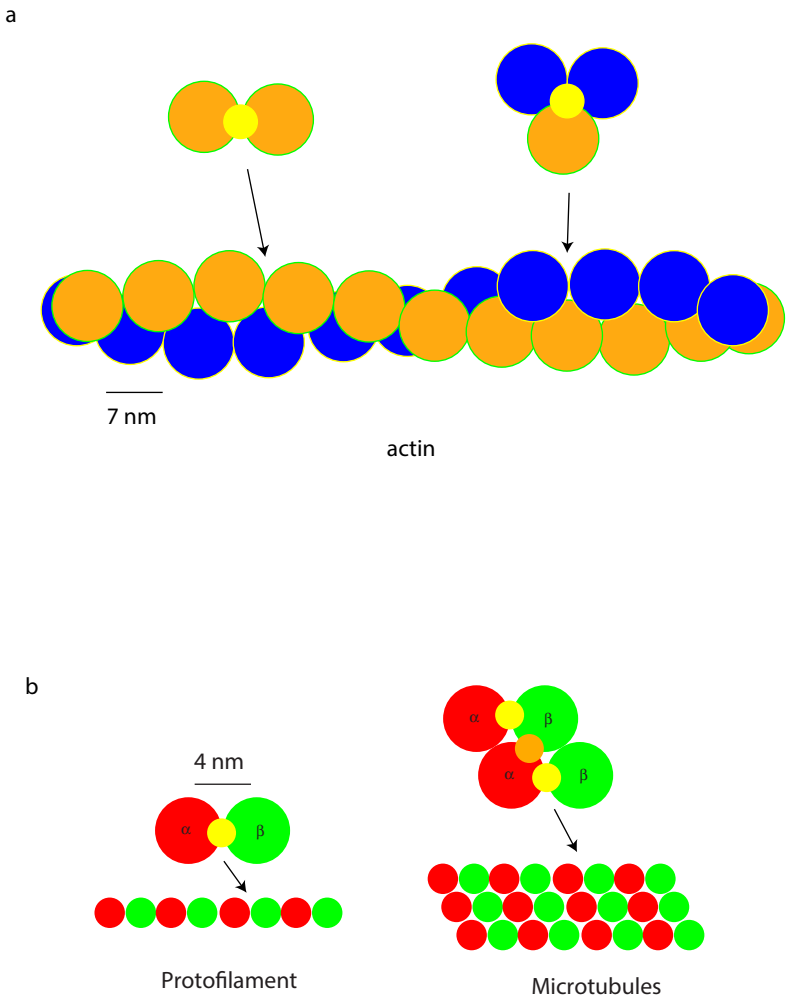


Figure 1.1: Actin and microtubules. Globular proteins with anisotropic attractive interactions can often be treated as patchy particles. Actins (a) are self-assembled by lines of myosins, which is shown as the blue and orange circles. The patches are shown in yellow. Similarly, microtubules (b) are self-assembled by α- and β- tubulins with patches in yellow. The tubulins form protofilaments, and the protofilaments form microtubules, with the patches shown in orange. Both actin and tubulin are globular proteins, and the patch interaction involves salt bridge, coil-coil interactions, *etc.*.

core, surrounded by α-helices. The oligomerization is guided by the α - helical coiled-coil interactions [2, 86], though the details vary between tubulin proteins. This interaction can be treated as sticky patches. These tubulin patchy patches assemble to protofilaments. On top of the protofilament structure, a microtubule is formed by two possible binding ways between the protofilaments: α to β tubulin, or α to α and β to β tubulin [35]. In any case, these binding site are the lateral contacts between protofilaments. In other words, each protofilament can also be treated as a patchy particle with the lateral binding sites as the patches, and the microtubule is the resulting patchy particle self-assembly. In fact, self-polymerization has been found in vitro with a high concentration of a mixture of the α-β heterodimers [21].

One way to study the self-assembly of proteins as patchy particles is by computational methods. Patchy particle simulations were performed by controlling the area of hydrophobic patches on otherwise hydrophilic molecules [131]. The result of this study shows that the system forms chain-like structures under low hydrophobic coverage, whereas a membrane-like structure is favored at high hydrophobic coverage. This simulation provides an insight on possible configurations of globular proteins under hydrophobic interactions.

1.1.3 Phase behavior of a patchy particle system

For any system composed of interacting particles, the particle behaviors depend on the particle interactions, temperature and the density of the system. These distinct behaviors can be described as the phases of the system. Stable or metastable phases are defined to exist when the equilibrium of the homogeneous fluid does not change within a long experimental observation time. For a system with isotropic interactions, the phase diagram with respect to the density and temperature is shown in Figure 1.2a [24, 101]. The stable, homogeneous fluid phase for these isotropically interacting particle systems is limited by the “glass line” on the high density side, and by the gas-liquid or liquid-liquid “spinodal

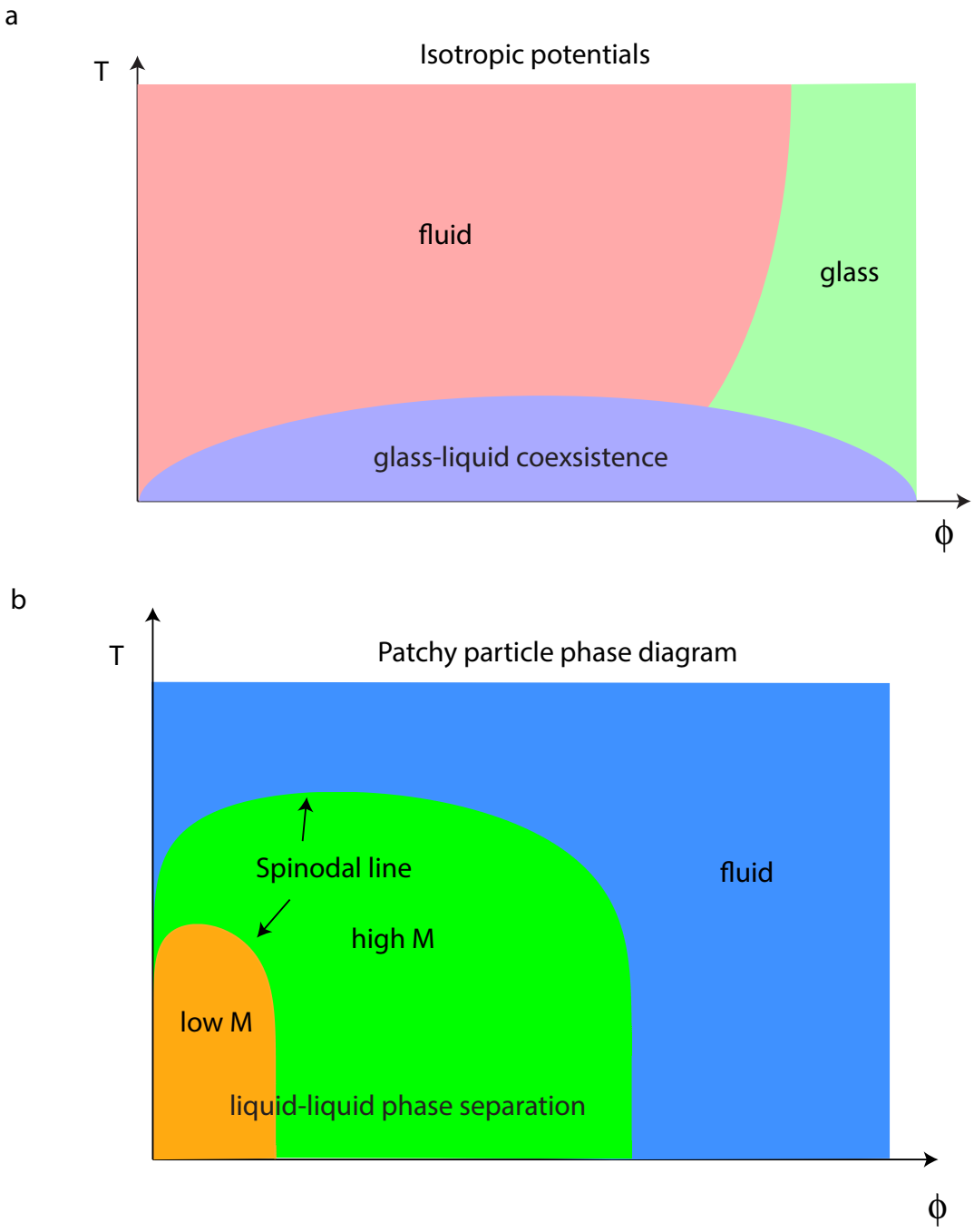


Figure 1.2: Phase diagram of systems of particles with attractive interactions. The phase diagram plotted on temperature (T) v.s. packing fraction (ϕ) of particles with isotropic potentials (a) shows a gas-liquid coexistence phase at low packing fractions. However, this glass-liquid coexistence phase may disappear in the phase diagram of patchy particle systems (b). The spinodal line of the patchy system moves towards a lower ϕ with decreasing average coordinate number $< M >$, making it possible to have a gel at all packing fractions [32].

line" at low temperature side. The glass line intersects with the gas-liquid spinodal line for these systems. Therefore, at low temperature, it is not possible to obtain a disordered, homogeneous arrested state; all states will exhibit phase separation. What will the phase diagram be for the anisotropically interacting patchy particle systems?

Bianchi and colleagues [12] approached this question using a Monte Carlo computational modeling method. They studied the phase behaviors of patchy particle systems with different numbers of patches with broad density and temperature ranges. They modeled the interactions between patchy particles by an isotropic repulsive hard sphere interaction and an anisotropic attractive patch interaction. The anisotropic interaction is exerted at certain spots on the particle surface, and the number of these spots per particle is called the coordinate number, M. Their simulation was performed on systems with different average coordinate numbers, $< M >$.

This patchy particle system shows novel physical phase behaviors, especially novel gelation properties [32, 96, 97, 99]. Most interesting is the behavior of spinodal line drawn in a phase diagram with temperature and concentration (Figure 1.2b). The spinodal line is a border which separates a stable phase from an unstable phase. To the left of the spinodal line, there is a liquid-liquid separated phase, while to the right of the spinodal line is a volume-spanning material. In some systems, this material to the left of the spinodal line is one with droplets of the solute that gather in one liquid phase. These droplets are suspended in the solute with another liquid phase of solvent. In other solutions, the liquid-liquid phase separation can be so strong that it results in a dense aggregation in the liquid. Sciortino and his collaborators found that the spinodal line of the patchy particle system systematically moves towards a lower solute density with a decrease of $< M >$ in the system [12]. With $< M >= 2.05$, the spinodal line is located at the packing fraction $\phi \approx 0.005$. This leaves most of the $\phi > 0.005$ regions of the diagram in a stable, volume-spanning, homogeneous phase. For $< M >= 3$, the spinodal line moves to $\phi \approx 0.1$, leaving the range in $\phi <$

0.1 as an unstable non-homogeneous phase. This patchy particle phase diagram has been experimentally verified using the interactions between DNA oligomers as the patches [13, 68, 94]. Given this theoretical understanding, new materials with bizarre phase properties were predicted and discovered, such as reaching a reversible gelled state from a disordered state by heating [95] and the discovery of a gelled, "empty liquid" in a clay [99].

To sum up, more and more study has focused on the patchy particle systems, both experimental and computational/theoretical. This is not only because these patchy particle systems are a rich new theoretical tool for studying the fundamental building blocks of biological systems, but they also show new physics with applicable properties. Novel phase behaviors have been found, and self-assembly of attractive particles with isotropic interactions are studied both for non-biological particles and for proteins.

1.2 Underwater vision

As we have seen above, patchy particle systems can create a huge variety of patterns and structures, and have attracted more and more attention recently. This is particularly 'hot' in biological systems, where our understanding has exploded dramatically in past decades. In this dissertation, I will focus on the self-assembly of a specific biological system, squid eye lenses. I will then compare the lens system to a separate but physically similar material, preserved Chinese "century eggs", or pidan in Chinese. The goal is to understand the self-assembly mechanism, and to propose a theory to explain it. Since the proteins in a squid lens are globular proteins, I will engage patchy particle theory to help understand the construction of the lens.

But before I go through details about the proteins and the structure of the squid lens, let us understand a little more about the lens system, and what is required to have an eye with high resolution and sensitivity, and the requirement of these underwater visual systems.

1.2.1 Eyes, lenses, and underwater vision

With the sun shining on earth from the beginning of life, life has evolved to capture the image from objects around in order to survive in the living environment. The structure of the image capturing systems vary in a continuum, from photon receptors of a few cells to complex camera-like eyes. In fact, it is not obvious how to define what an eye is, because nearly every imaginable intermediate exists between an eye with high resolution and the simple light sensor. However, the big distinction between an eye and a light-sensitive cell is whether the system is capable of spatial vision [73, 75].

Spatial vision describes a kind of vision which perceives light intensity with spatial distribution. It requires some mechanism to focus incident photons onto a set of photoreceptors that are overall independent from each other. Given these, a camera-like structure of an eye is a good candidate, and has been found in a lot of species. This eye structure requires a lens to focus light, and the appearance of a lens marks a giant leap in the eye evolution. Nillson and Pelger have studied the course of the evolution of lenses by using computational methods [83]. They showed that an animal might have used 200,000 generations to evolve a pin-hole-like structure from a flat layer of photon sensitive cells. A lens could plausibly have appeared after 38,000 additional generations. After that, the eye and the lens could have modified themselves and reached the configuration similar to a modern eye within 120,000 generations. During this process, the appearance of a lens is crucial because a pin-hole structured eye can only reach a limited vision in an evolutionary dead-end, but a lens can break the bottleneck and form a camera-like eye to achieve better spatial vision.

For these kind of eyes, there are two aspects to describe good vision: resolution and sensitivity. Resolution is the minimum angle distance between two incident photons that the eye can tell apart. It is limited by the diffraction of light and the distance between the independent cells. Sensitivity is proportional to the number of photons that a receptor

receives. This is crucial to underwater animals because the total number of photons in the aquatic environment is low. In fact, the number of photons in water is attenuated roughly by a factor of 10 for every 70 meters [73, 75], while animals have good spatial vision down to abyssal depths. For animals like squid whose habitats are at 300 meters deep, the irradiance is 10,000 times dimmer compared to the land. Thus, underwater animal eyes have adapted to this darker environment by having a bigger lens with a shorter focal length in their eyes.

1.2.2 Spherical lens

To maximize the amount of light reaching the retina, a spherical lens is optimal for camera-like eyes because it is the shape to achieve highest sensitivity. There are two features which determine the sensitivity of an eye: the lens diameter (for underwater eyes) or the pupil size (for terrestrial animals) D, and the solid angle Ω which each receptor covers to accept light. The size of a lens or a pupil behaves like an aperture of a camera and the number of photons getting through is in proportion to the area of the aperture, which is D^2. The number of photons that a receptor accepts is also in proportion to Ω, which is in reverse proportion to the square of the focal length f, as is shown in Figure 1.3. The sensitivity is expressed as

$$S \propto D^2 \Omega. \tag{1.1}$$

Therefore, to increase the sensitivity underwater, it is required to have a bigger lens and a smaller focal length. Note that the focal length is in proportion to the radius of curvature of the lens, so there is evolutionary pressure for a lens with smaller radius of curvature. On the other hand, an eye cannot be too big compared to the size of the animal, which also limits the maximum diameter of the lens. So given that the diameter of the lens cannot exceed a certain size, having a smaller focal length requires the lens to have the smallest radius of curvature. Thus, sensitivity requires a spherical lens for sea animals compared to

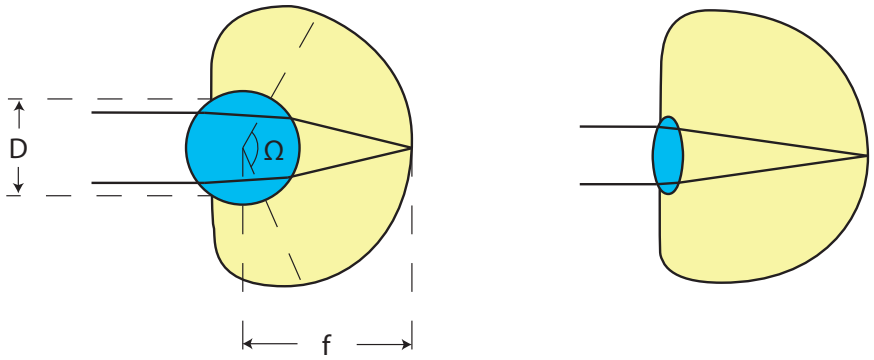


Figure 1.3: Spherical lenses for underwater vision. A spherical lens is required to maximize the photon flux for an underwater eye. Having a spherical lens can increase the aperture and decrease the focal length. Both of these result in an increase of the photo flux. The eye aperture D is the opening that allows photons getting through. The focal length is f, and the solid angle from the center of the lens to the retina is labelled as Ω.

a 'lenticular' shaped as is found in human eyes. This lens design has evolved in parallel in both fishes and squids.

1.2.3 Spherical aberration

But there is a problem with spherical lenses. If the whole lens is composed of material with the same refractive index, the images the lens forms will suffer from spherical aberration. For example, a glass bead will focus light which is incident on the center of the lens to a further point compared to light incident on the periphery [114]. Thus, the image from a glass bead is not acute. Spherical lens animals have evolved lenses which have a graded refractive index (GRIN): The refractive index is high in the core of the lens and low in the periphery. This gradient compensates for the difference of the focal lengths for light incident at different part of the lens.

1.2.4 Maxwell's fish lens equation

The refractive index at different radial positions of a spherical lens can be calculated using geometric optics. The relationship between a refractive index and a lens radius that will correct for spherical aberration can be calculated using the Maxwell's fish lens equation [23]

$$n(r) = \frac{n_c}{1 + \left(\frac{r}{a}\right)^2}. \tag{1.2}$$

Here, r is the radial position, n_c is the refractive index of the core, and a is the radius of a reference sphere (Figure 1.4b). This equation describes a distribution with spherical symmetry where refractive index increases with decreasing r, and the maximum of refractive index appears at the very center of the lens, with the value of n_c. With this refractive index distribution, a spherical lens can focus light into a point, as is shown in the following: in a medium with spherical symmetry, the equation of a ray path can be written in this form:

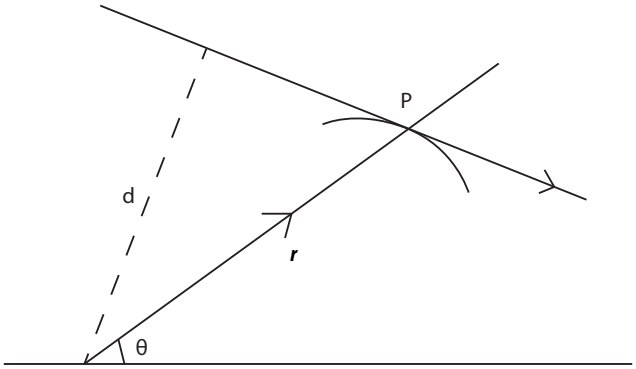
$$\theta = c \int^r \frac{dr}{r\sqrt{n^2(r)r^2 - c^2}}, \tag{1.3}$$

where θ is azimuthal angle and c is a constant. The details can be found in Figure 1.4a. Plugging in Equation 1.3, we have

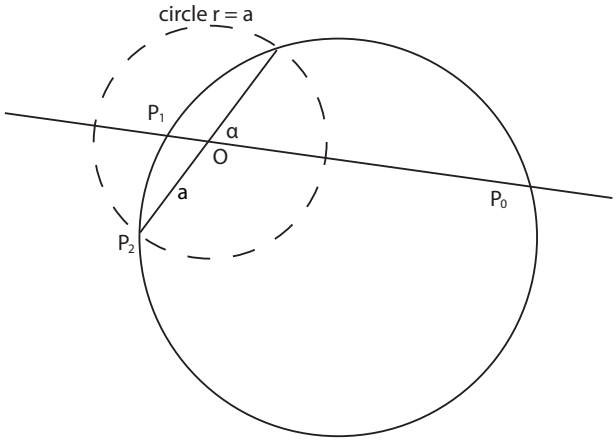
$$\frac{r^2 - a^2}{r\sin(\theta - \alpha)} = \frac{r_0^2 - \alpha^2}{r_0 \sin(\theta_0 - \alpha)}, \tag{1.4}$$

where $P_0(r_0, \theta_0)$ is a fixed point and α is a constant. The physical meaning of this equation is shown in Figure 1.4b. For any incident light passing through a point P_1, they will meet at P_0 which is at the line joining P_0 to O with , and the paths of the rays are arcs of circles. Hence, a medium with its refractive index distributed as in Equation 1.2 focuses each point on the spherical surface to the opposite point on the same surface.

Thus, to correct for spherical aberration in a spherical lens, the predicted refractive



a



b

Figure 1.4: A refractive index gradient can eliminate spherical aberration in a spherical lens. The ray path of in a medium with spherical symmetry is shown in (a). The position of any point P in the ray path is described as the vector r and the azimuthal angle θ. Figure (b) shows that a Maxwell's fish lens can focus light into a point, where $P_0(r_0, \theta_0)$ is a fixed point and α is a constant. Any incident light passing a point P_1 will meet at P_0.

index is described in the Maxwell's fish lens equation. This equation requires that the refractive index increases monotonically from the periphery to the core of the lens.

1.3 Vertebrate eye lens

All vertebrate lenses which have acute vision are graded refractive index lenses. The shape of an underwater vertebrate eye lens is a sphere. In the development of a vertebrate lens, they are generated from ectodermal cells. The mitotically active cells are located only in the outer epithelial monolayer around the lens [90]. There is a region in the epithelium where cells elongate into fiber-like structures. After the elongation starts, nuclei and mitochondria are ejected from the lens fibers. Therefore, only cells in the epithelial monolayer can duplicate and create lens proteins. The cells are laid down in layers after the elongation from the epithelium at the periphery of the lens. Throughout the growth of the lens, these cells are displaced toward the center. The resulting structure of the lens has distinct layers of cells which are maintained in the mature organ. The lens grows throughout the whole animal's life, but the rate of the lens growth is only rapid in the embryonic development and in early life after birth [90].

1.3.1 Vertebrate lens protein

Proteins are one of the structural building blocks for different living tissues. This is true with the lenses as well. A lens cell is like a bag of water filled with proteins. The primary structural material of the lens is water-soluble proteins called "crystallins", which make up more than 90% percent of the total protein content of the lens [17, 90]. The concentration of these proteins is responsible for the refractive index gradient. In vertebrate lenses, the crystallins can be grouped into α-crystallin, β-crystallin, and γ-crystallin. These crystallin proteins were first discovered in vertebrate lenses by Bhat and colleagues in 1989 [11]. α

crystallins are evolutionarily related to stress proteins [5]. There is also a 35 kDa crystallin found in a frog lens [119].

α-crystallins are related to small heat-shock protein family [56] and are also expressed in heart, skin, brain, retina and other tissues. They are polydisperse oligomers with the molecular weight between 800 to 900 kDa, containing more than 40 subunits. There are two α-crystallin genes, αA and αB, sharing around 60% sequence identity [16]. αA-crystallin is expressed only in the lens [108] and αB-crystallin is stress-inducible. The interactions between α-crystallins are repulsive in a physiological solution [116]. The interactions involve hard-core and screened Coulomb interactions, which are responsible for lens transparency [25, 46]. α-crystallins are essential to keep transparency of the vertebrate lens, given that they are presumed to be function both as structural proteins and as protein chaperones [17].

β-crystallins also form oligomers, and monomers range from 22 to 28 kDa in molecular weight. γ-crystallins are mostly monomers, with molecular weight about 21 kDa. The main difference between β- and γ- crystallins in sequence is that compared to γ-crystallins, β-crystallins have N-terminal extensions. The basic β-crystallins have C-terminal extensions as well [8, 53, 65]. The interactions between β-crystallins are repulsive, and the interactions between γ-crystallins are attractive [116].

1.3.2 Transparency of vertebrate lens

Lens transparency in vertebrates is achieved by complex interactions and assembly of the constituent proteins. Physically, transparency requires that the protein density fluctuation is maintained low at the length scales of visible light wavelengths. There are two major ways by which a cataract happens in a vertebrate lens: protein unfolding and phase separation.

Protein unfolding can result in protein aggregation or a change in the interaction with other crystallins resulting in lower solubility. To prevent protein unfolding, α-crystallins

has been found to act as a chaperone [90]. Experimental results show that α-crystallins suppress thermally-induced aggregation of various enzymes, including β- and γ- crystallins [56]. However, a cataract may still happen when an immature lens fibre cell is overwhelmed by stress if there is misfolding of a mutant crystallin, or when a mature lens fibre cell uses up the chaperone capacity of α-crystallins [90].

Another kind of cataract, however, happens without protein misfolding. It is caused by changes in solubility or attractive properties in lens proteins, which leads to a phase separation [90]. Due to this liquid-liquid phase separation, big density fluctuations result in scattering of light. For a mixture with attractive interactions between particles at low temperature and low concentration, the mixture is separated into two phases at mesoscopic scale with different solute concentrations. This separation can be droplets of a liquid phase of the solution component suspending inside another liquid phase of the same solution component [20, 79], or even sediment to the bottom of the container. These two liquid phases have different refractive index, and scattering comes from the interface of these two phases. With millions of interfaces between the two phases acting as little scatterers, opacity occurs. Concentrated γ-crystallins, which have van der Waals and other attractive forces, have been observed with liquid-liquid phase separation [70]. This is why lowering temperature of a bovine lens results in droplet-like liquid-liquid phase separation, and opacity in the lens, a so-called "cold cataract" [79]. For an underwater lens, turbidity is more likely in the periphery, where the protein packing fraction is similarly low [9, 79, 114]. A little protein aggregation can result in strong density fluctuations. In the core of an underwater lens, the ultradense gelled condition suppresses of the density fluctuation, reducing the unwanted light scattering [79].

In the periphery of a vertebrate lens, opacity is avoided by having a short range order of the crystallin proteins, perhaps coming from α-crystallin repulsive interactions [46]. SAXS on α-crystallins in solution shows a decrease in scattering intensity at small angles

with the increase of α-crystallin concentration in the solution [124], indicating a repulsive interaction between the α-crystallins. But this repulsion is weak, and with the combination of this repulsion and depletion, a short-range liquid-like order forms for α-crystallins. The predicted quaternary structure of these proteins displays tetrahedral symmetry [124]. Since the molecular weight of α-crystallins can be large, this short-range ordered size is close to visible light and this is how a vertebrate lens keeps transparency.

The short-range order disappears at the core of a vertebrate lens. No peaks have been seen in SAXS experiment at the core of either a bovine or a fish lens. This disappearance is achieved by the increase of polydispersity in the sizes of the crystallins and the increase in concentration of the small γ-crystallins which occupy the void space [79].

1.4 Squid eye lens

The material evolutionary story for a squid lens is quite different from a fish lens. The lenses of cephalopod eyes are composed of independently developing anterior and posterior hemispheres that later fuse to become the mature, spherical lens. These hemispheres differentiate from an overlying layer of ectodermal cells [117]. Both segments are developed from separate groups of ectodermal cells referred to as lentigenic cells [126]. S-crystallins accumulate first in the posterior segment and protein synthesis begins in the anterior segment two days later. The cephalopod lens continues to grow throughout maturity, and S-crystallins are made throughout the squid's life [126]. However, as in the vertebrate lens, mature lens cells lack nuclei and mitochondria; there can be no protein turnover after a lens cell matures and is laminated onto the existing spherical lens structure.

Similar to a vertebrate lens cell, the shape of a squid lens cell is a long cylinder with dimensions of approximately $10 \times 1 \times 1 \mu$m, as shown in Figure 1.5. The cells are packed with their membranes tightly fused together to avoid light scattering in between the cells.

In fact, when we look at the lens under microscope, a lens cell has to be isolated from the others to have enough refractive index difference with the environment for us to observe it, as is shown in Figure 1.5d. The lens structural proteins are contained inside these cells.

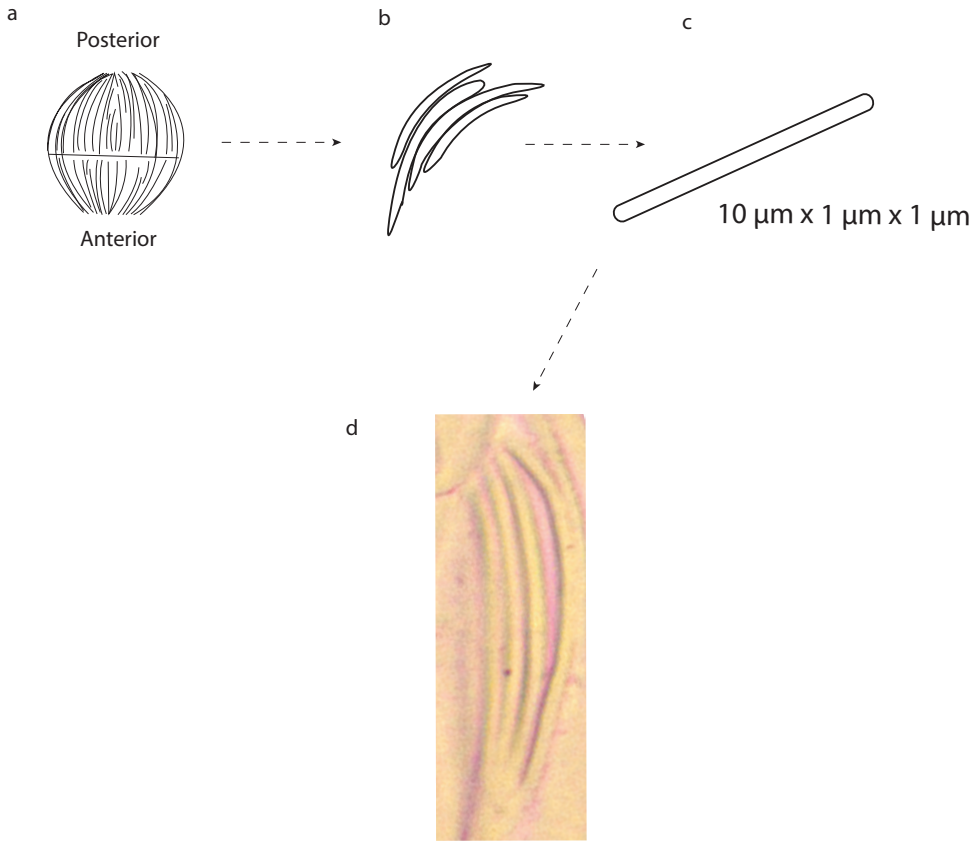


Figure 1.5: Cellular structure of a lens. The lens is composed of cells which are laid down parallel to each other on a path from the zenith to the nadir of the sphere. The lens has a membrane which separates the lens into the posterior and anterior halves (a). The cells have filament-like shape next to each other (b,c). The cell image under an optical microscope is shown in (d).

1.4.1 Squid lens protein

The protein composition of cephalopod lenses is completely different from the vertebrate lens proteins, and the entire system is independently derived. Squid lenses contain only one type of protein: S-crystallin [31, 114, 118, 121–123]. These are relatively homogeneous,

soluble proteins [19]. There are dozens of different S-crystallin forms expressed in a single lens that are 46-99% identical to each other by amino acid sequence [123]. S-crystallins are evolutionarily derived from glutathione S-transferase (GSQ), a liver detoxification enzyme, which may also play a role in molecular chaperoning, similar to αA crystallin in vertebrates [64]. For S-crystallins, all the isoforms except lops4 (Figure 1.6b) have a few inserted peptides with various lengths in the sequence of GSQ [117, 123]. These exons correspond to a disordered, loop-like structure protruding from the folded surface of an S-crystallin. The presence of the loop is correlated with a loss of GSQ enzymatic activity in a given protein [117]. A single S-crystallin is a dimeric structure, with each monomer containing a mixture of β and α structure in the N-terminal domain, and all α-helix in the C-terminal domain. Comparing the octopus S-crystallins to the squid S-crystallins shows that most of the members of this lens protein family in squids expanded after the divergence of these two taxa 200-300 mya [117, 120], and the major S-crystallins seem to have undergone positive evolutionary selection correlated with a lens structural role [120].

The proteins were cataloged by Sweeney and colleagues into two groups based on the loop length: long-loop proteins (molecular weight $\approx$ 27 kDa) and short-loop proteins ($\approx$ 25 kDa), as is shown in Figure 1.6. They used homology modeling to calculate the surface charge of these S-crystallin isoforms [114]. They found that the long-loop proteins have higher net charge (+4.8) compared to the short-loop proteins (+2.8).

The evolutionary relationships of these S-crystallins' DNA sequences were studied by Sweeney, *et al.* [114]. They constructed a gene trees using the non-loop part of these proteins in Figure 1.7, where the underlying DNA sequences can be aligned. Their result shows that the amino acids which are positively selected are the ones which contribute to the stability of the S-crystallin tertiary and quaternary structures. These include the residues which are responsible for the dimer interface, and the ones that stabilize the end of helices in the C-terminal domain of the protein. Further, their result shows that the long-loop proteins

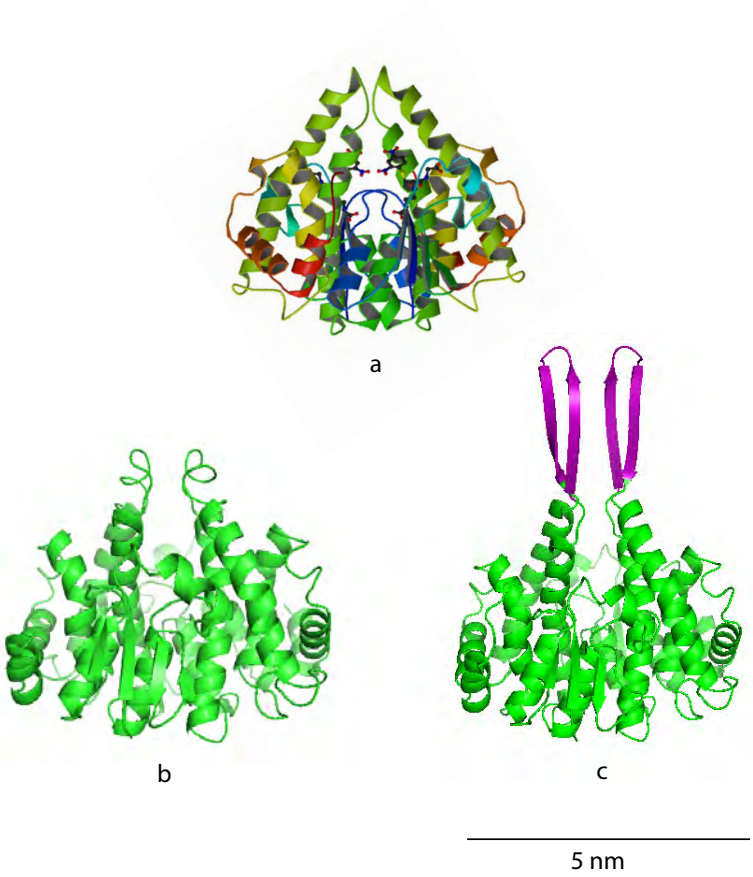


Figure 1.6: S-crystallins and glutathione S-transferase (1GSQ). The protein structure of 1GSQ (a), a short-loop S-crystallin (b), and a long-loop S-crystallin (c) are shown. The protein structure 1GSQ is from the liver-expressed enzyme related to the lens-expressed S-crystallins. The short-loop S-crystallin and long-loop S-crystallins expressed in the lens differ in loop length. Lops4 is shown here as an example of short-loop S-crystallin (b).

are more recently diverged from the rest of the family compared to the short-loop proteins. These highly-charged long-loop proteins are under more-intense positive selection, and their appearance is approximately simultaneous with the appearance of the complex GRIN lens. These results suggest that long-loop proteins are crucial to form the GRIN lens.

Comparing vertebrate lenses and cephalopod lenses, the protein components between vertebrate lenses and cephalopod lenses are completely different and unrelated. Squid lenses contain many isoforms of S-crystallin, whereas there are three crystallin families with large polydispersity in a vertebrate lens. In addition, graded index in squid also

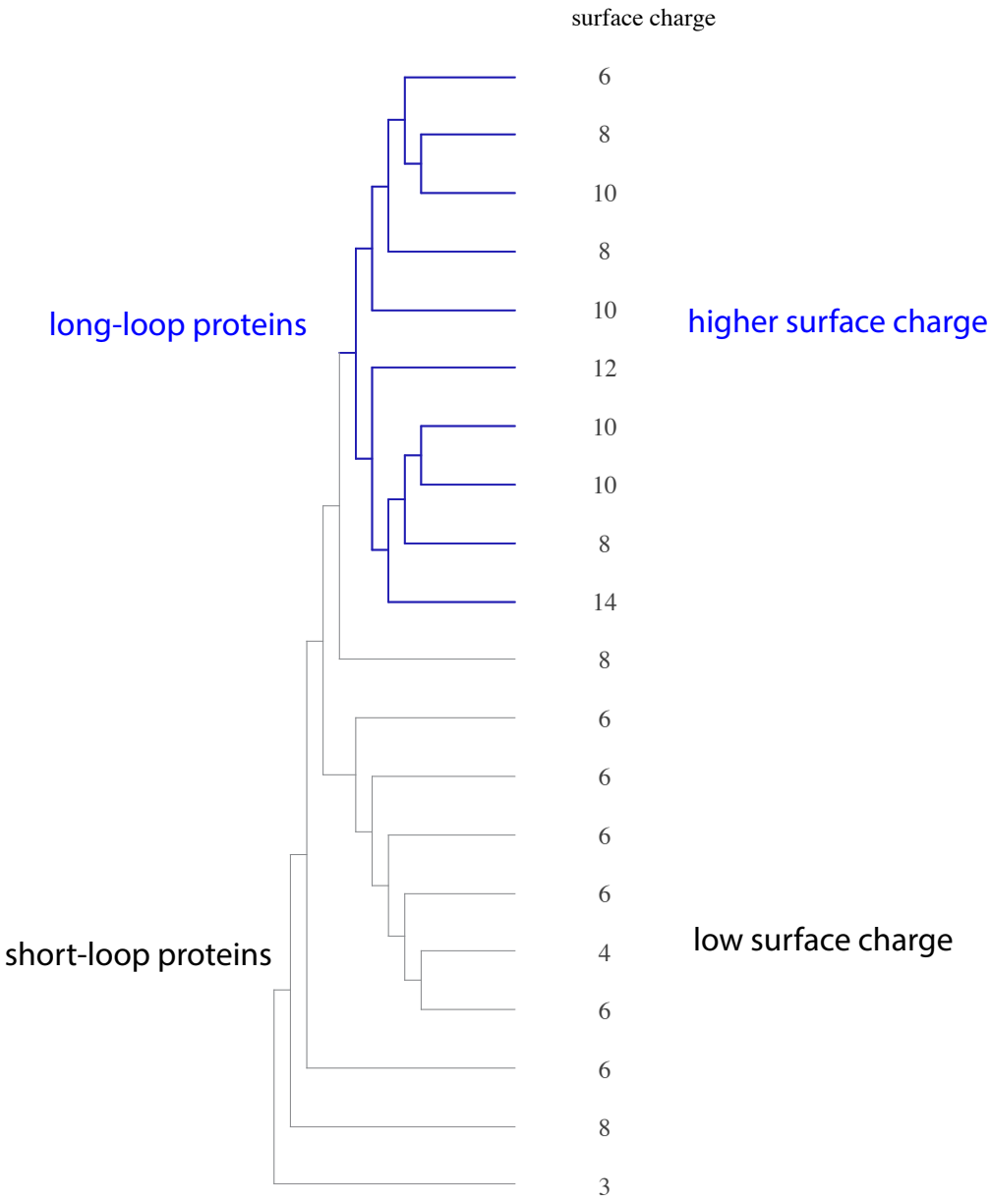


Figure 1.7: Gene tree of S-crystallins. The gene tree of S-crystallins (taken from Sweeney *et al*. 2007 [114]) shows evolutionary relationships of S-crystallin proteins as inferred by DNA sequences. The long-loop S-crystallins are labeled in blue and the short-loop in black. This gene tree shows that the long-loop S-crystallins form a clade, and the clade has a higher surface charge compared to the short-loop S-crystallins.

evolved independently from vertebrate lenses. All of the above suggest that there is another mechanism of how a squid lens builds a gradient index and maintains transparency.

1.4.2 Previous hypothesis for lens transparency

To study the possible mechanism of how a squid lens builds and maintains graded index and transparency, Sweeney and colleagues studied the S-crystallin distribution in the lens with respect to the lens's optical properties [114]. They found that the long-loop proteins are expressed more in the peripheral, low-index regions of the lens. At the core of the lens, it is mostly composed of short-loop proteins. Further, they found that long-loop proteins are highly charged, with much higher positive surface charge compared to short-loop proteins. Therefore, they hypothesized that the mechanism for a squid lens to keep transparent in the periphery of the lens is to have proteins here highly charged, and the repulsive screened coulomb interaction between these proteins prevents them from aggregating. Then short range order forms in this mechanism, similar to the repulsive mechanism for transparency of vertebrate lenses, and no aggregations or accumulations are allowed by the screened Coulomb repulsion. This hypothesis also predicts that there is no larger structure formed at length-scales larger than a few protein diameters, and proteins will be homogeneously distributed. In this hypothesis, thermal fluctuations are reduced by a repulsive jamming in the system, and proteins in the periphery form a repulsive glass. In the core of a squid lens, they predicted that since the protein density is so high, the space is all packed with proteins, so there is little density fluctuation in the core of the lens. The increase in concentration of short-loop proteins in the center would be due to the requirement of high packing fraction — smaller proteins without long-loops can be packed denser compared to long-loop proteins. The core of the lens is also jammed, required by the high protein density.

In short, the old hypothesis for invertebrate lens transparency predicts that the whole

lens is a jammed system, with charge-charge repulsion between jammed proteins in the periphery of the lens, and close-packing interactions in the core.

In this thesis, I will conduct experiments to investigate in our previous hypothesis. I will check whether the old hypothesis is consistent with my experimental results. If not, what is the mechanism that the proteins assemble in the squid lens to achieve the GRIN lens with such a large ϕ span? Are the proteins in the lens patchy particles? What is the phase of the lens system? Can the phase diagram provide us with a solution to the assembly of the lens system? Further, are there any other protein systems that are similar to the squid lens system with proteins as the patchy particle? My research will focus on these questions.

1.5 Outline of this work

In this thesis, I study the material properties and the self-assembly of transparency and gradient index of the squid eye lens. I start from the protein density requirement of the spherical lens in Chapter 2. I also use biological methods to analyze the protein properties of the lens described in Chapter 2. These methods provided some initial insights into the larger structure of the self-assembling protein system. After that, I describe the small angle x-ray scattering experiments on intact lens tissue. The result provides information about the possible phase of the lens, and checks whether our old hypothesis is correct. This is explained in Chapter 3. In Chapter 4, I convert the x-ray scattering result in Fourier space to real space by fitting the scattering intensity using a Monte Carlo simulation. This probes a real space protein configuration in an intact lens. I also discuss the possible assembly mechanisms of the lens system. In Chapter 5, I introduce another patchy particle protein system, the Chinese century egg, and compare this system to the lens system. Finally, in Chapter 6, I summarize what I have done and propose future work.

Chapter 2

Proteins in the squid lens

Just as the non-water content of most biological tissues is predominantly protein, the major component of a squid lens is protein. As was mentioned in Chapter 1, squid lens protein composition is very different from vertebrate lens protein composition. Therefore, to understand how the squid lens is assembled, and what makes the lens transparent, it is essential to systematically study the proteins inside the lens. This includes the protein component in the lens, the spatial distribution of the protein components, the molecular weight, the charge of the protein component, and the native form. In this chapter, I describe how I conducted several biological experiments to study these properties of the protein in a squid lens. Specifically, I begin from the refractive index requirements of the lens, and then estimate the protein concentration necessary to achieve the observed refractive index distribution. Then I describe gel electrophoresis on the lens proteins to study the composition and concentration of protein at different radial positions in the lens. To understand how the general molecular weight distribution of S-crystallins maps to the proteins' atomic structure, I estimated the three-dimensional the protein structure of S-crystallins using homology modeling. From there, I predict the circular dichroism absorbance (CD) of the folded protein, and perform CD on S-crystallins to verify that the proteins structure consis-

tent with my homology models, and are well-folded in the lens. At the end of this chapter, I studied the surface charge of the protein by conducting isoelectric focusing experiments (IEF). All of these properties from the lens proteins will be essential to construct my hypothesis to explain the lens as a self-assembling system of graded refractive index in the subsequent two chapters.

2.1 Introduction - protein identification

Squid lens cells are essentially membranous bags of S-crystallin proteins which are laminated together and organized into a sphere. The major protein components in a squid lens are S-crystallins, as was mentioned in Chapter 1. To understand the lens system, I need to know what isoforms of S-crystallins are contained in the lens, and where they are expressed. The exact protein composition of *Loligo* lenses was studied using Illumina technology RNA sequencing by my labmate Tom Dodson. The sequencing assembly results showed a total number of 53 unique S-crystallins expressed in the lens, 14 of which were assembled as full-length constructs. These proteins are very similar to S-crystallin sequences currently in genbank from other species, and thus, we identify them as S-crystallins. The major differences between the proteins in our dataset appear mainly in the region of the sequences that encode a loop-like structure in the center of the two folding sub-domains of the conserved enzyme. In fact, we identified 39 unique "loop" sequences of variable length, and in all cases they are predicted to encode an unstructured, soluble loop protruding from the folded core of the homodimeric protein. This loop accounts for the observed differences in molecular weight between S-crystallin isoforms.

In short, there are a whole spectrum of S-crystallin isoforms expressed in the squid lens. In the following chapters, I will study their physical interactions and spatial distribution in the lens.

2.2 Sample collection and lens layers

Loligo forbesii squid lens tissue was obtained from the Marine Biological Laboratory at Woods Hole, MA, and coastal campus of University of Connecticut at Avery Point. *Sthenotuethis* lens tissue was obtained with squid jigs from the R/V Kilo Moana from various sites in offshore Hawaii. Lenses were excised from the suspending tissue, and fixed immediately for TEM, or stored at -80 °C either alone or in RNA Later (Life Technologies #AM7020, Grand Island, NY 14072) for use in other experiments. Lenses were thawed on ice and divided into anterior and posterior hemispheres along the naturally-occurring membrane separating the two. For most of our experiment, the posterior hemispheres were dissected into four concentric hemispherical layers, shown in Figure 2.1. The thickness of each layer is somewhat arbitrary and varies with different lens samples. Therefore, these dissections were digitally photographed to make a record of the radial position of each tissue subsample. On average, the peripheral layer usually ranges from 80% to 100% of the total lens radius, and we call it the 100% layer; the next layer, the 80% layer, ranges from 60% to 80%; the 60% layer ranges from 40% to 60%, and the radius of the core sample is typically less than 40% of the total radius of the lens.

2.3 Refractive index matching experiment

The squid lens is interesting because a graded refractive index sphere is required to have acute vision underwater (see Chapter 1). To start, we experimentally studied the refractive index at different positions in a squid lens and fit these data to theoretical expectations about refractive index distribution. We did several independent experiments and confirmed that *Loligo* lenses have the expected density and refractive index distribution with radius that is described to first order by the Maxwell's fish lens equation.

The refractive index of the squid lens varies at different radial positions to eliminate

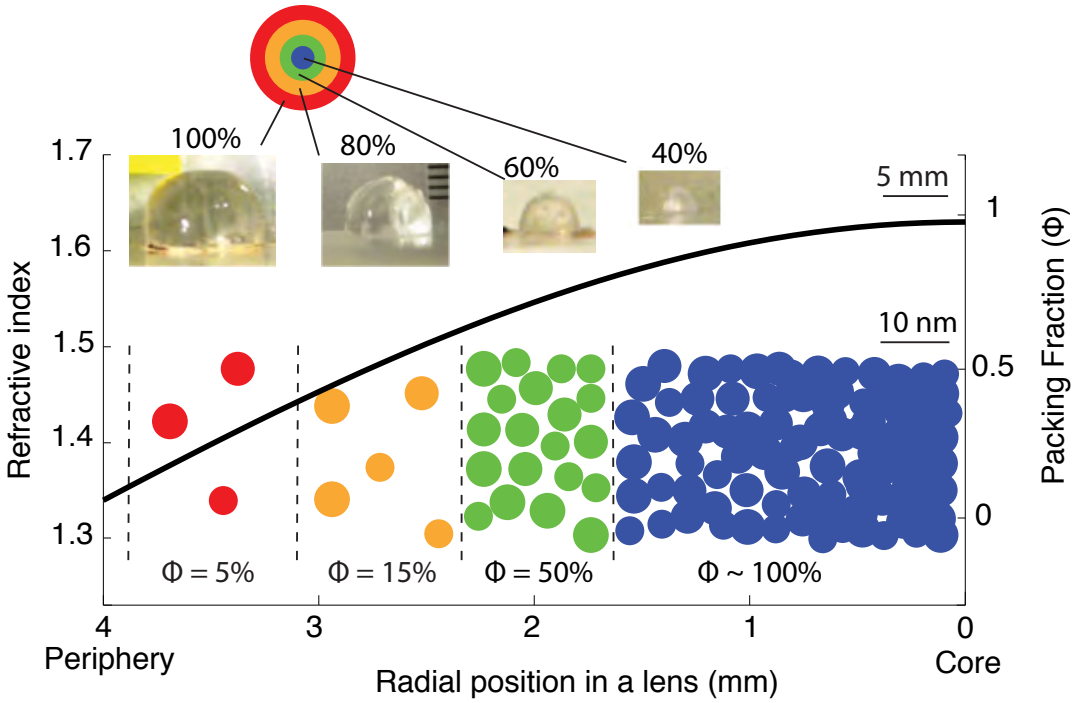


Figure 2.1: Dissection of the lens, refractive index, and packing fraction distribution. Hemispheres of lens tissue were dissected into four concentric layers labelled as 100%, 80%, 60%, and 40% layers (shown in the images); the exact thickness of each layer varies with different lens samples. The relationships between refractive index, protein packing fraction and lens radius are plotted in the black curve. This relation is obtained by fitting the Maxwell's fish lens relation using my index oil matching result, as explained in the text. Colored circles show the approximate protein density required at each position to generate the required refractive index gradient.

spherical aberration (Chapter 1). Using geometric optics, the refractive index at different radial positions of a ratio lens can be calculated using the Maxwell's fish lens equation:

$$n(r) = \frac{n_c}{1 + \left(\frac{r}{a}\right)^2}, \tag{2.1}$$

where r is the radial position, n_c is the refractive index of the core and a is the radius of the focus sphere in the lens. To estimate n_c and a, we dissected a *Loligo* lens into seven concentric layers, and the refractive index of each dissected layer was estimated by submerging the pieces of tissue in refractive index matching oils of 1.45 and 1.55 (Cargille

labs, Cedar grove, NJ). Immediately after immersion, we photographed each layer in index-matching oil, and using these photographs, we recorded the radial position of the tissue sample that was invisible in each of the two index-standard fluids. With this information about exact refractive index at two radial positions, we can estimate the refractive index at all radial positions in the lens with the Maxwell's fish lens equation (Eq. 2.1).

The result of the index matching experiment shows that for a single example lens, tissue samples from approximately $r = 3$ mm radial position disappear in the $n = 1.45$ standard, and samples from $r = 1.9$ mm disappear in the n = 1.55 standard. Using equation (1), we have $n_c = 1.63$ and $a = 8.6$ mm, and the predicted refractive index at different radial position of the lens is

$$n(r) = \frac{1.63}{1 + \left(\frac{r}{8.6}\right)^2} \, . \tag{2.2}$$

The distribution of the refractive index is shown in Figure 2.1. From this equation, the inferred refractive index at the outer most radial point in the lens is then 1.37, close to that of seawater with $n = 1.34$. In the core of the lens, the inferred refractive index is 1.63. See Table 2.1 for the refractive index at other positions in the lens.

2.4 Protein concentration in the lens

Having estimated the refractive index distribution, we want to find out what protein properties are responsible for this gradient refractive index. There are two possible scenarios: 1. The lens is composed of proteins with different protein refractive indices (n_p) at different radial positions, or 2. the lens is composed of proteins with the same protein refractive indices, but the concentration of the proteins are different. In this section, I describe the experiment which was conducted to answer this question, and estimate the protein concentrations required for that refractive index.

2.4.1 The gradient of refractive index is achieved by a gradient in the protein concentration

To check whether the refractive index gradient was achieved by having a graded protein concentration, I calculated the refractive indices n_p of the dry proteins in the lens using the Lorentz-Lorenz formula

$$n(p) = \sqrt{\frac{2R_p + \bar{v}_p}{\bar{v}_p - R_p}}\ , \tag{2.3}$$

where R_p is the refraction per gram of the protein, and $\bar{v}_p$ is the protein partial specific volume, which is the weighted average of the partial specific volume of each amino acid [74]. I calculated R_p for all the proteins in the lens based on the amino acid sequences in our sequencing data. These proteins have been studied before [114] evolutionarily, and our previous study presents the evolutionary tree of these proteins. From there, I can estimate the refractive increment, or increase in refractive index with increase in protein concentration (c) at 578 nm, using the Wiener equation [135]:

$$\frac{dn}{dc} = \frac{3}{2}\bar{v}_p n_0 \frac{n_p^2 - n_0^2}{n_p^2 + 2n_0^2}\ . \tag{2.4}$$

In this equation, n_0 is the refractive index of the medium, which was set to be 1.334. I calculated refractive index increment dn/dc for all the proteins in the lens, and compared the dn/dc for the new emerged proteins and the old proteins. Then I used Zhao's method [135] to calculate the predicted dn/dc as the increment of protein concentration.

For all of the S-crystallins, my calculation shows that the mean of dn/dc is 0.196 ml/g, with a standard deviation of 0.003 ml/g. The refractive index increment of the long loop proteins, which are relatively recently diverged from the rest of the lens proteins, is $dn/dc = 0.197 \pm 0.002$ ml/g. The short loop proteins have $dn/dc = 0.195 \pm 0.004$ ml/g. Thus, there

are no major differences in refractive increment among proteins in the squid lens, and therefore no evidence that squid lenses increase refractive increment in the center of the lens by changing the amino acid composition of the crystallin proteins, as was shown by Zhao *et al* for vertebrates [135].

In short, the refractive index gradient in a squid lens is achieved by having a gradient in the protein concentration. I now want to estimate the predicted protein concentration. Given that we have dn/dc of the protein assembly and the refractive index distribution, we need to have one data point with both known refractive index and known concentration to estimate the whole concentration distribution in the lens. I measured both these properties at the core of the lens, where the protein concentration is the highest across the lens.

2.4.2 Core of the lens, measurement and theoretical estimation

To find the protein concentration in the lens core, I measured its density by comparing it to fluids of known density. I did this by submerging additional samples from the very core of the lens in potassium iodide solutions of known density; a tissue sample will float in less-dense fluids, and will sink in more-dense fluids. I can also infer the refractive index of the core from the Maxwell's fish lens equation. As reported below, these two methods of estimating the core refractive index were in good agreement.

For the core of the lens, our density measurement result is 1.28 g/ml. The average partial specific volume of S-crystallins is 0.73 ml/g. Thus, the average density of S-crystallins is 1.37 g/ml. Assuming that protein is the only material in the lens, the packing fraction of the core is calculated to be 93%. That is, there is little water in the core. I checked this by exposing a 4.22 mg sample of lens core to vacuum for approximately two hours, and the mass of the sample remained unchanged after this exposure. Additionally, the average calculated refractive index of the dry protein from amino acid sequence shows that $n_p = 1.63$. This is the same as the core refractive index n_c, which I determined by

extrapolating from direct index measurements using Maxwell's fish lens equation. The fact that the inferred core refractive index is the same as dry lens protein is another piece of evidence which indicates that there is no water in the core of the lens.

To sum up, our density experiment at the core of the lens shows that the core almost entirely composed of proteins with little water in it. This is required to achieve the high refractive index. This conclusion is further supported by our small angle x-ray scattering (SAXS) result as will be discussed in Chapter 3. Next, I will estimate the protein concentration at different radial positions in the lens.

2.4.3 Predicted protein concentration of a squid lens at different radial positions

The concentration and the packing fraction ϕ of the protein in the lens at all positions can be estimated: Given a radial position in the lens, the refractive index was calculated using the Maxwell's fish lens equation. The required protein concentration can be calculated using the refractive index increase increment equation 2.4.

Given both the core density and dn/dc, the calculated protein packing fraction at different radial positions of the lens is:

$$\phi = 0.75c = \frac{6.1}{1+(r/8.6)^2} - 5.1 \ . \tag{2.5}$$

The result is shown in Table 2.1 at several radial positions of the lens, with the bold refractive indices the measured data. The packing fraction of protein in lens tissue spans from $\sim 5\%$ in the periphery to 100% in the core of the lens. This result is also shown in Figure 2.1.

In short, our results show that the refractive index gradient in a squid lens is achieved by having a gradient of protein concentration in the lens, with the core 20 times more

Radius(mm)	RI	Packing Fraction
3.7	1.37	0.04
3.0	**1.45**	0.33
1.9	**1.55**	0.71
1.3	1.59	0.86
1.0	1.60	0.92
0 (inferred)	1.63	1.0

Table 2.1: Refractive indices at different radial positions in a squid lens. The radial position of a lens, refractive index (RI), and the packing fraction are listed. The bold numbers in RI are obtained directly from the index oil matching experiment, and the rest are calculated.

concentrated than the periphery of the lens. Throughout this span of protein concentration, transparency is required at all positions in the lens. The question is how does the lens maintain a low protein density fluctuation at such a large protein concentration span? How does the lens avoid liquid-liquid phase separation throughout the lens, especially in the periphery where protein concentration is low? To approach the answer, we need to know the distribution of S-crystallin isoforms at different positions in the lens.

2.5 Relative protein abundance distribution

Studying the relative abundance of the S-crystallin isoforms throughout the lens is important to understanding how the lens maintains transparency and graded refractive index because the different S-crystallin isoforms may have different properties in assembling the lens system. Determining the S-crystallin isoform distribution will help us to develop a hypothesis for the function of the S-crystallin loops. To start, I performed gel-electrophoresis (SDS-PAGE) on squid lenses.

2.5.1 Relative protein abundance at different radial positions in the lens

Polyacrylamide gel electrophoresis (SDS-PAGE) separates the proteins by their molecular weight. Proteins unfolded by SDS are loaded into a gel, and an electric field is applied to the gel. Proteins with higher molecular weight will migrate slower in the gel compared to the smaller proteins.

In this experiment, lenses were dissected into four concentric layers and stored in RNA later at -80 °C. Then the subsamples were homogenized with plastic micropestles in 1.5 ml microcentrifuge tubes in XT sample buffer (Bio-Rad Laboratories, Hercules, CA, #161-0791) with XT reducing agent (Bio-Rad 161-0792) and protease inhibitor cocktail (#P8849-5ml, Sigma Aldrich, St. Louis, MO) added according to the manufacturer's instructions . These samples were heated to 95 °C for 5 minutes to denature the constituent proteins. Then the solution was centrifuged at 14,000 rcp for 1 minute to remove any remaining solids. This procedure resulted in no visible pellet, suggesting that lens proteins were entirely solubilized. I used a BioRad 10% precast polyacrylamide 12+2 well gel (Bio-Rad #345-0111), with XT MOPS running buffer (Bio-Rad #161-0788) for electrophoresis. The gels were stained with Invitrogen SafeStain according to manufacturer's instructions (Life Technologies, #LC6060). Then the gels were scanned in transparency mode on a desktop scanner (HP Scanjet G4050). I calibrated both the quantities of protein loaded in the gel, and the images from the scanner such that the resulting grayscale intensities of protein bands in the scanned images were linearly proportional to the concentration of protein in the band.

I calibrated molecular weights of proteins in the samples using protein molecular weight ladder (Bio-Rad Precision Plus Protein Kaleidoscope Standards #161-0375). The relative migration of a given protein band to the dye front was fitted as an exponential function

of the molecular weights of the proteins in the standard. The gaps between protein bands were treated as the intensity baseline and were corrected by the built-in MATLAB function 'msbackadj'. The greyscale intensities of the samples were then normalized relative to total S-crystallin protein content by dividing by the sum of all values in the range of 20 to 40 kDa for a single sample.

At all radial positions, greater than 90% of the protein in the lens was found in two overlapping, broad bands migrating on the gel. These broad bands represent two major molecular weight classes of S-crystallin, one with majority molecular weight at around 24.5 kDa and the other with majority molecular weight around 26.5 kDa, as is shown in Figure 2.2. It was not possible to completely separate these two bands, which are the composite migrations of many S-crystallin isoforms, and seem to represent the mean migration position of the long-loop and short-loop categories of S-crystallin proteins. Therefore, based on information from our RNA sequence data, I chose to define the boundary between the 24.5 kDa (short-loop) and 26.5 kDa (long-loop) protein migration bands to be 25 kDa. The ratio between 26.5 kDa band to 24.5 kDa band decreases significantly from 1.0 in the periphery of the lens to 0.06 in the core, shown in Figure 2.2. There are also two minor bands in each sample, at 32 and 36 kDa, similar to proteins in a frog lens [119]. The very high molecular weight S-crystallins ($\sim$ 36 kDa) behave differently from the 26.5 kDa and 24.5 kDa molecular weight S-crystallins. They are at a minimum in the periphery of the lens, and increase in abundance with decreasing lens radius to a maximum of 7% of total protein at a lens radius of 2 mm. The abundance of these proteins then decreases again to a prevalence of 1% to 3% in the lens core.

In short, the SDS PAGE result shows a complex set of proteins composing the majority of the non-water mass of the lens, consistent with our RNA sequencing result. In the molecular weight range from 23 kDa to 28 kDa, the average molecular weights shift towards a lower molecular weight from the periphery to the core of the lens. This trend provides a

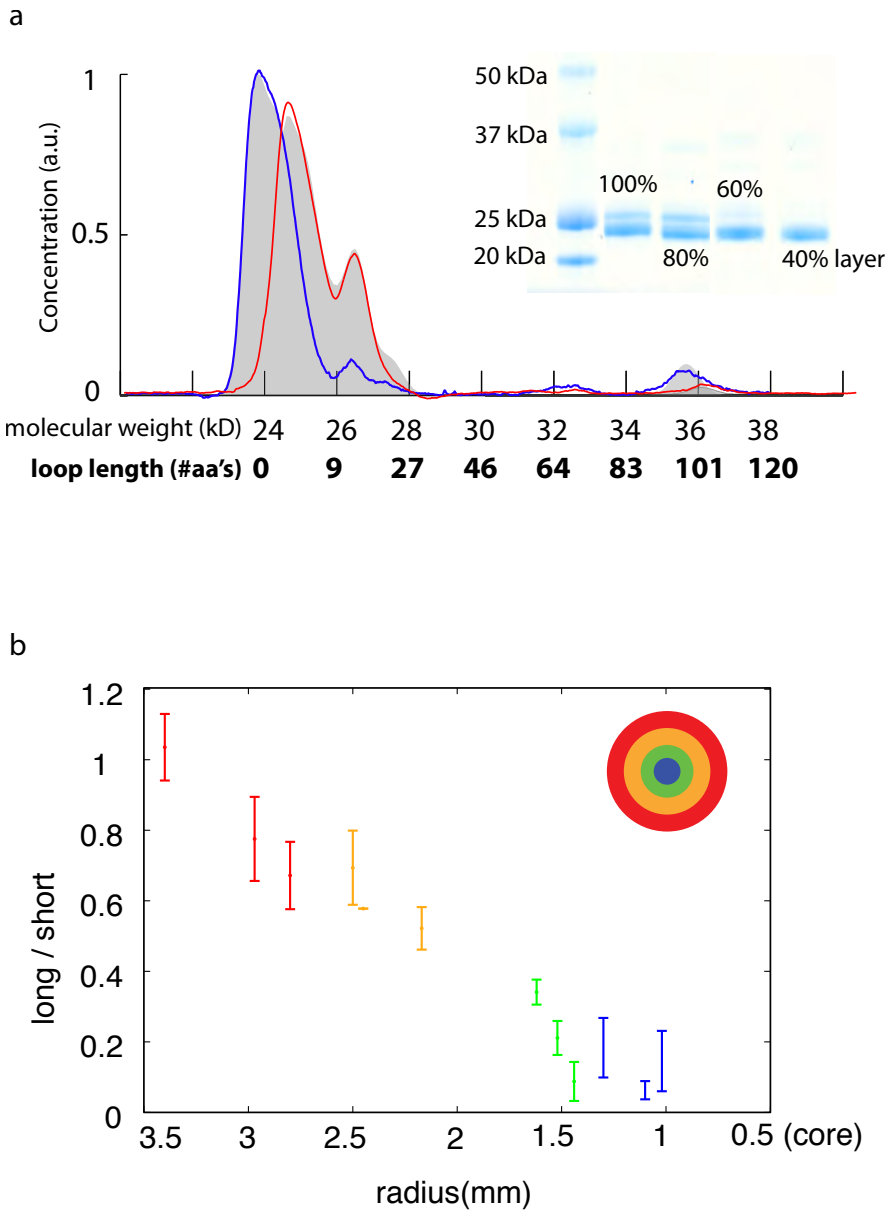


Figure 2.2: Protein concentration of different molecular weight. The relative concentration of proteins of different molecular weights (a) is plotted from the peripheral layer (red) and the core (blue) of the lens. The red and blue curves are data from SDS-PAGE with background adjusted, as described in the text. The grey regions are the fit to the raw data using peak decomposition. The inset shows the gel image. The ratio of the concentration of long-loop S-crystallin to short-loop S-crystallins with error bar at different radial positions in the lens (b) decreases from periphery to the core, with the 100% layer in red, 80% layer in orange, 60% layer in green, and 40% layer in blue.

clue on how the lens assembles, as will be discussed at the end of this section and chapter 4. For now, I further investigate the protein composition by comparing the RNA sequencing result to the SDS-PAGE result.

2.5.2 Peak decomposition

As was mentioned in the introduction section of this chapter, both RNA sequencing and SDS-PAGE can be used to determine the protein composition of a squid lens. This section compares these techniques together to further our understanding about the proteins in the squid lens. RNA sequencing results show that the loop lengths for 24.5, 26.5, and 36 kDa proteins are about 14, 25, and 106 amino acids, respectively. The width of each of the peaks in SDS-PAGE suggests that these peaks result from the summation of a mixture of S-crystallin isoforms of slightly different molecular weights. Given the large number of S-crystallin-encoding transcripts in our RNA sequencing data, and the difficulty of fully separating these isoforms using biochemical techniques, I wanted to infer the relative abundance of each predicted protein as a function of radial position in the lens using our SDS-PAGE data. This was achieved by deconvolving the total protein migration pattern observed in SDS-PAGE as a sum of individual protein isoforms with molecular weights encoded by RNA transcripts observed in our sequencing data, for all radial positions in the lens. I estimated the contribution of each predicted protein transcript to the total migration pattern by modeling the observed SDS-PAGE migration pattern as the sum of many Gaussian curves, with each individual transcript contributing a Gaussian distribution of density of fixed width consistent with a monodisperse protein, and a height as determined by the following fitting algorithm. The positions of the Gaussian curves were fixed to have center positions within 1 kDa of a predicted transcript with a maximum width of 0.5 kDa for transcripts of 23.5 kDa to 28 kDa. For proteins between 35 to 37 kDa, I fit with one peak with the position ranges from 35 to 37 kDa, and the maximum width of 0.5 kDa.

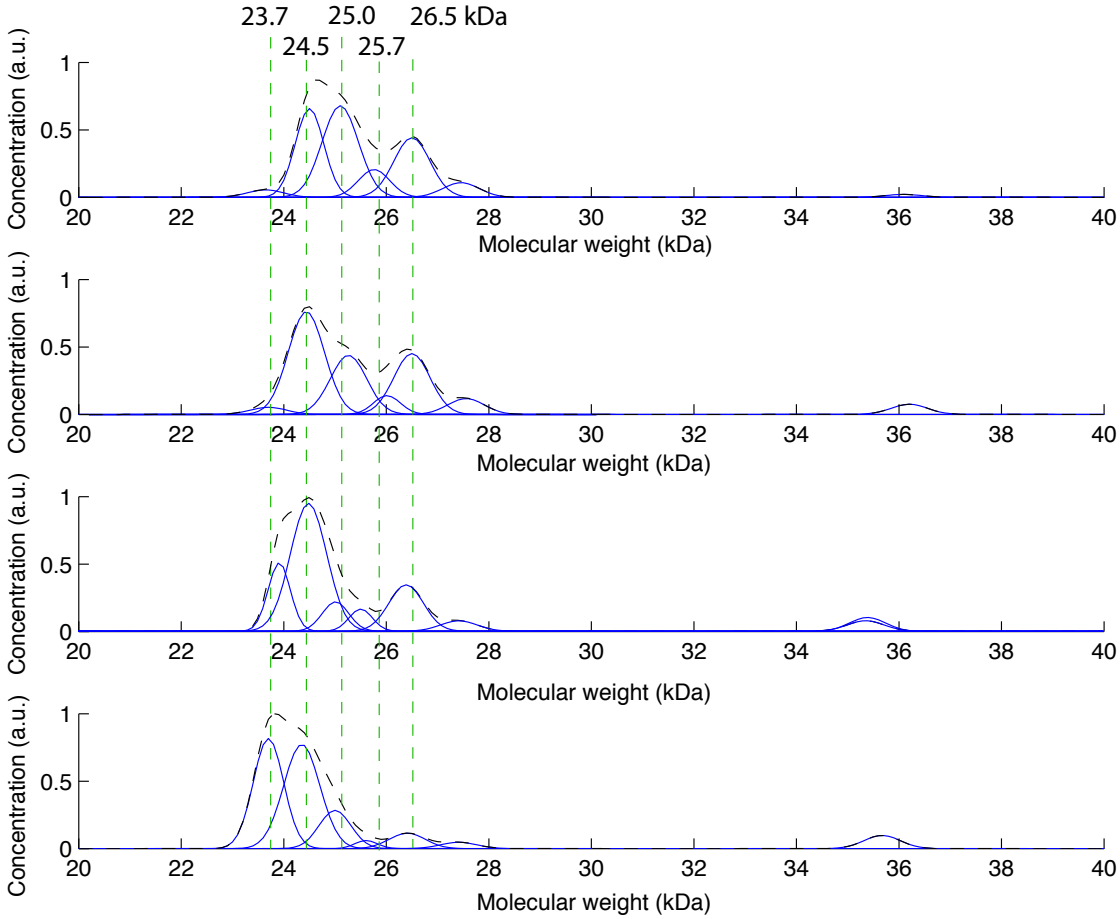


Figure 2.3: Protein abundance decomposition. The decomposition of the result from SDS-PAGE combined with RNA sequencing result at different radial positions is shown in these four panels. From top to bottom: 100% layer, 80% layer, 60% layer, and 40% layer. Each solid line shows the fitted peak representing a single isoform, and the dotted curve is the sum of these peaks. The green dotted lines denote the peak positions for isoforms with molecular weights of 23.7, 24.5, 25.0, 25.7, and 26.5 kDa. The decomposition the aggregate electrophoresis data also shows that the average molecular weight of S-crystallins in the lens decreases from periphery to the core.

The result showed that although similar-looking broad peaks are present at all lens radii, S-crystallin isoforms in each major broad composite peak continuously shift to a larger molecular weight as a function of lens radius (Figure 2.3). Though the exact position of the sub-peaks vary at different radial positions in the lens, the wide composite peaks centered at 24.5 and 26.5 kDa are mainly comprised of S-crystallins with molecular weight 23.7, 24.5, 25.0, 25.7, and 26.5 kDa.

The close fitting of SDS-PAGE data to transcripts found in RNA sequencing, along with previous published data about squid lens composition [114] suggests that the lens is composed of many variations of a single glutathione s-transferase-related protein fold. The most salient difference between these closely related proteins is the length and composition of a loop inserted between the two folding domains of the protein. There is a clear increase of the relative abundance of the low molecular weight S-crystallin isoforms from the periphery to the core of the lens. The lowest two major sub-peaks in molecular weight are comprised of S-crystallins with molecular weight 24.5 and 25.1 kDa in the periphery of the lens, whereas in the core, it is comprised of 23.7 and 24.35 kDa. This shift shows a clear trend of decreasing average molecular weight from the periphery to the core of the lens. At the high molecular weight side, the 26.5 kDa sub-peak component comprises 22% of the total S-crystallin composition in the periphery, and this proportion decreases to 6% in the core. This trend also appears in the 25.7 kDa sub-peak. The percentage of this isoform drops from 9% to 2% from periphery to the core of the lens.

In conclusion, the SDS PAGE result is consistent with RNA sequencing data in that both of datasets show a complex S-crystallin composition of many isoforms. The peripheral layer of the lens contains more proteins with a broad peak at molecular weight 26.5 kDa compared to that in the core. It is very likely that these 26.5 kDa proteins play a role in maintaining the lens transparency at the periphery of the lens. Next, we want to know what is the difference in the folded structure between these long loop proteins and other proteins.

I used the protein structure estimation technique called homology modeling to estimate the atomic-scale protein structure of the ensemble of S-crystallins.

2.6 Homology modeling of the proteins in the lens

Homology modeling provides a method to estimate atomic structure and therefore the surface characteristics of the different S-crystallin isoforms, as well as a plausible initial conformation of the apparently unstructured cytoplasmic loops.

To estimate the structure of S-crystallin isoforms, I used homology modeling which constructs the protein secondary and tertiary structure based on homologous, known protein structures. Specifically, S-crystallins tertiary and quaternary structure were estimated using the Swiss-model (http://swissmodel.expasy.org/). The predicted S-crystallin models were built based on homo-oligomeric structures from selected proteins from the SWISS-MODEL template library. Glutathione S-transferase (1GSQ) was the main known protein structure selected as the template for all of the isoforms of S-crystallins. 1GSQ is the progenitor of the squid lens S-crystallin [61, 123]. It has a dimeric structure. The squid S-crystallin have a few inserted peptides with various lengths in the sequence of 1GSQ. The similarities between 1GSQ and S-crystallin are the reason why 1GSQ are chosen to be the homology model template.

The resulting estimate of S-crystallin structure shows similar configurations as glutathione S-transferase, as is shown in Figure 2.4. S-crystallin monomers form a homodimer. In addition to the conserved globular structure, two disordered loops are inserted in the 1GSQ sequence, taking a flexible, unfolded configuration in the solvent. The major structural variation between different S-crystallin isoforms mostly appears in the details of the predicted configurations of the loops. Most of these loops are predicted to be flexible in shape, and floppy in the sense that they can freely swing in the solvent. For the super long

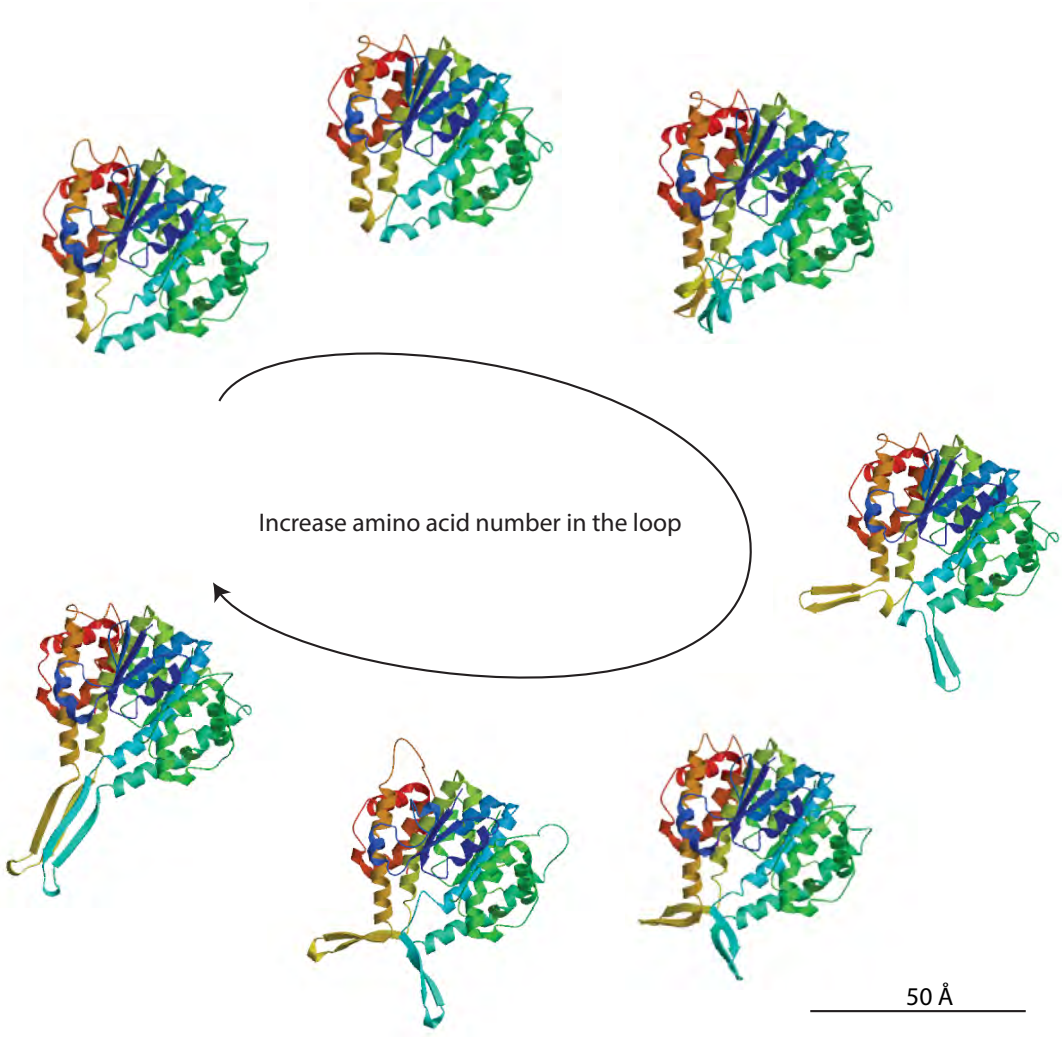


Figure 2.4: S-crystallin isoforms. The S-crystallin isoforms from homology model with different loop lengths are shown with the arrow labels the increase of the loop length. These S-crystallin isoforms have similar dimeric body structure. The loops have various lengths and structures.

loop proteins with 103 and 106 amino acids, there are no templates which are similar in the loop amino acid components. As a result, homology model fails to output a predicted structure for them. If the 106 amino acid loop is extended and unfolded in the solvent like the other shorter loops appear to be, it can easily extend to 300 Å away from the folded body of the protein. In general, basic residues tend to be exposed on the surface of the protein, while the disordered loop inserted in the otherwise conserved 1GSQ sequence tends to take a flexible, unfolded configuration in the solvent.

2.7 Circular dichroism spectroscopy

In order to determine whether S-crystallins are folded in lens cells as predicted according to homology models, I performed circular dichroism spectroscopy. A *Loligo* lens was dissected into four concentric layers, and the 100% layer tissue was placed in 150 mM phosphate buffered saline (PBS) (P4417, Sigma Aldrich, St. Louis, MO) with protease inhibitor cocktail (P8849-5ml, Sigma Aldrich) 1:200 added. This tissue and buffer were homogenized in a Dounce tissue homogenizer (06-434, Fisher Scientific, Pittsburgh, PA), and the resulting fluid was centrifuged at 13,400 rcf for 10 min to remove undissolved proteins. Then the solution was 1:10 diluted in the PBS with the same percentage of protease inhibitor added. All of the above was performed either on ice or in a temperature-controlled system with temperature set to be 4 °C. A Jasco J-810 spectropolarimeter (Easton, MD) continuously purged with nitrogen was used to obtain far-UV circular dichroism spectra (CD) concurrently with far-UV absorption spectra. The UV region was measured between 180 and 400 nm using a 0.1 mm quartz silica cell (International Crystal Labs). The spectra were all acquired using a data pitch of 0.05 nm, and at a continuous scan speed of 200 nm/min, with response time of 0.5 s, and bandwidth of 5 nm. A minimum of 5 accumulations for each sample were averaged at a temperature of 20 °C (±0.05 °C; 293K). The

temperature was controlled by a Peltier heating and cooling system. The analysis of the CD spectrum is performed using the online server CAPITO [129].

The absorbance difference between the two oppositely circularly polarized light sources at different wavelengths is shown in Figure 2.5. Two minima were observed in the absorbance at 213 nm and 223 nm. A maximum was located in the middle of these two minima. The two minima are consistent with a helix structure [28]. This fact is consistent with our homology modeling that the S-crystallins are mostly composed of helical structures.

The spectra values at 200 nm and 222 nm are plotted to estimate the folding properties of the lens protein together with these values from known folded and unfolded proteins, shown in Figure 2.5. The lens protein locates in the middle of the globular proteins. This is consistent with the fact that S-crystallins are globular proteins. The lens protein does not locate among the unfolded proteins. Therefore, the circular dichroism experiment shows that the proteins in the lens are well folded.

2.8 Isoelectric focusing (IEF)

Our previous hypothesis was that the proteins in the lens are highly charged and the repulsive interactions between the proteins prevent them from aggregating (Chapter 1). In this section, I study the net charge of the S-crystallins. To estimate the charge of S-crystallin isoforms, I performed isoelectric focusing (IEF) on the protein solutions. This method separates proteins by their charge in the gel. *Loligo* lenses were dissected into three or four layers, and each layerwas homogenized in a Dounce homogenizer in 360 mg/ml urea and 25% (v/v) glycerol buffer on ice, then centrifuged at 14 000 g at 4 °C for 10 min. The resulting supernatant was loaded in a pH 3–10 gradient Criterion IEF Gel (Bio-Rad # 345-0072). I used 20 mM lysine plus 20 mM arginine (Bio-Rad #161-0762) for cathode buffer,

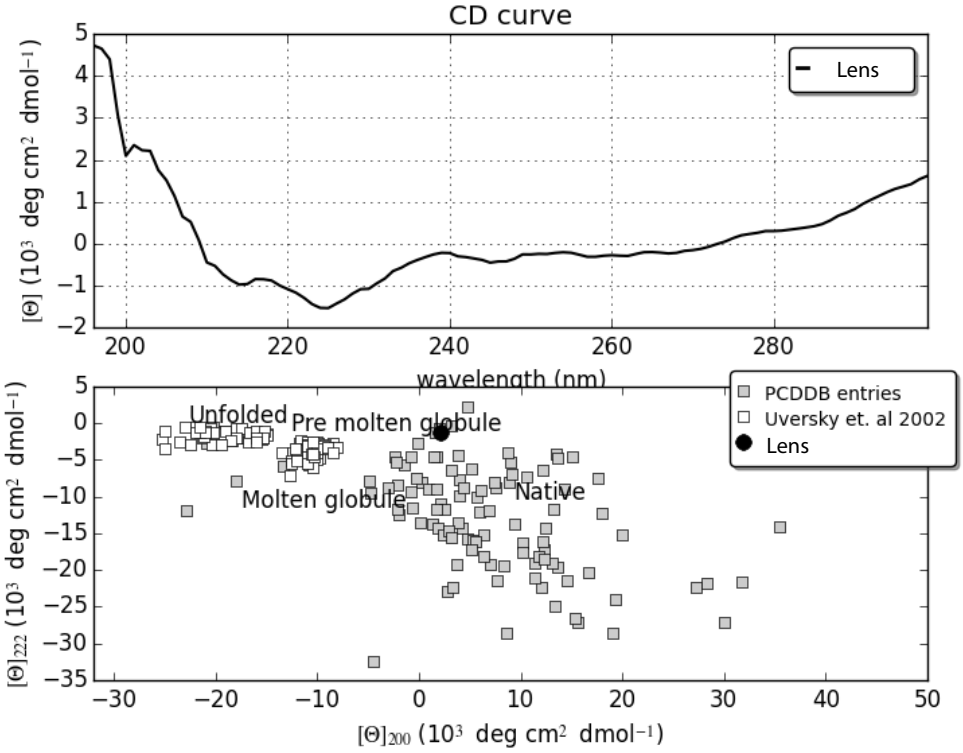


Figure 2.5: Circular dichroism spectroscopy of lens protein. The circular dichroism absorbance (Θ) from a ground lens tissue with buffer subtracted in shown in the top figure. There are two valleys at 209 and 222 nm, and a peak at 192 nm. These features are consistent with alpha-helical structure. The bottom figure shows the absorbance at 200 nm and 222 nm of the lens protein compared to known proteins in their native form and unfolded form using the online server CAPITO [129]. The lens protein locates in the middle of the globular proteins, which is consistent with the structure of S-crystallin. This indicates that the proteins in the lens are well folded.

and 7 mM phosphoric acid for anode buffer (Bio-Rad #161-0761). One gel was run following the manufacture's recommendation on the connection of electrode. To increase the resolution of the basic isoelectric point's of S-crystallin proteins, I ran a second gel with the anode and the cathode buffers switched relative to the manufacture's recommendation, and the electrodes from the power supply also switched accordingly; the protein samples were loaded in the anode side of the gel. Both gels were fixed in 30% methanol, 10% trichloroacetic acid, and 3.5% sulfosalicylic acid for an hour, followed by 4 hours of wash

in 30% methanol and 12% trichloroacetic acid, then stained with silver stain (Bio-Rad Silver Stain Plus Kit #161-0049). The gel results reported here show the data from the first gel for isoelectric point PI < 6 and data from the second, switched electrode gel for PI > 6. The isoelectric point of the 36 kDa S-crystallins are calculated using the online server ExPASy from the amino acid composition [14, 15, 50].

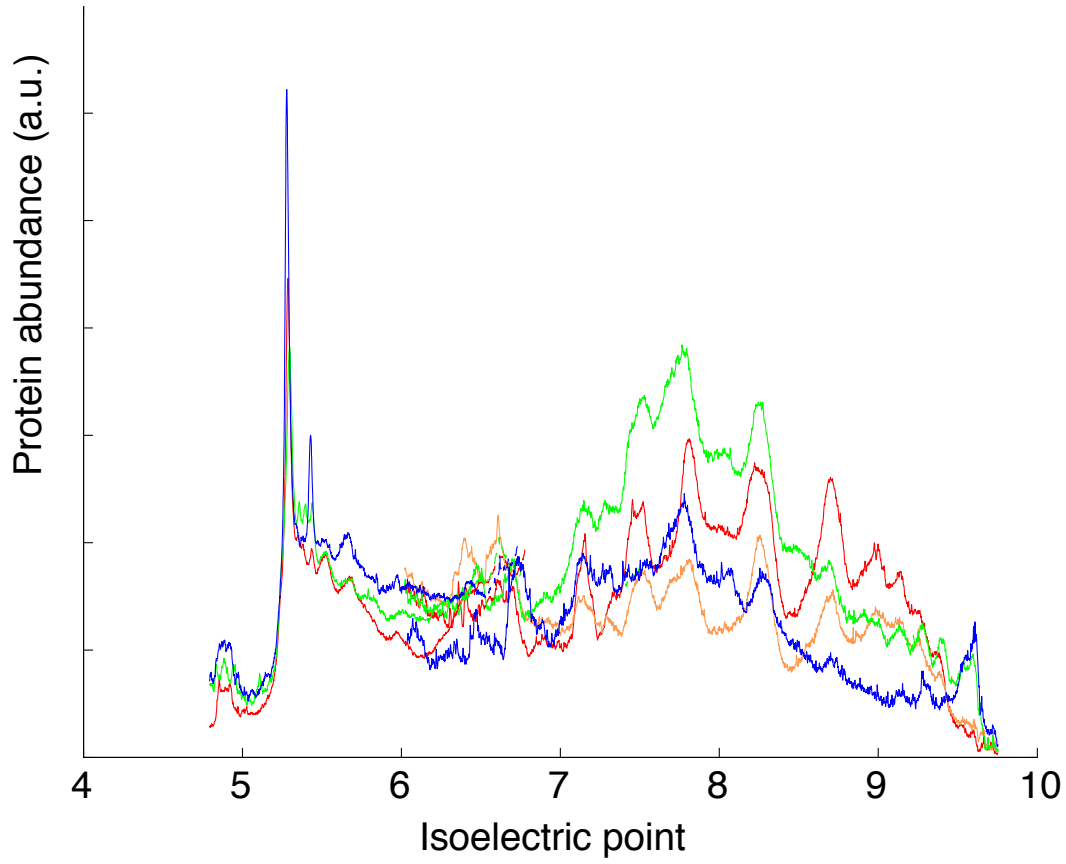


Figure 2.6: IEF-PAGE of lens proteins. Lens proteins are separated by their isoelectric point (PI), and the abundance of each PI is shown in the figure. Curves are colored according to the radial position in the lens the samples were taken from, with red 100%, orange 80%, green 60%, and blue 40%. The sharp peak at isoelectric point of 5.4 is very likely from the 32 and 36 kDa S-crytallin isoforms. The peaks at $PI \approx 8$ may come from the long- and short- loop S-crystallins.

The IEF gel results show a spectrum of proteins ranging from a $5 <$ PI < 9.5, as is shown in Figure 2.6. In the basic range of pH, from $7 <$ PI < 9.5, there is a broad composite

peak at PI $\approx$ 8 for all layers in the lens. I attribute this broad composite peak to the mixture of S-crystallins with molecular weights from 23 to 28 kDa. This spectrum of peaks is consistent with our SDS-PAGE result that the proteins in the lens are a mixture of dozens of S-crystallin isoforms, rather than a binary mixture of two proteins. The acidic part from the IEF-PAGE shows a sharp peak at PI= 5.4 for all layers. This peak is associated with the S-crystallins with super long loops at 36 kDa. The calculated PI is 5.2 for these proteins because the loops of these proteins are highly negatively charged. This result is also consistent with our SDS PAGE result that there are 36 kDa S-crystallins at all positions in the lens.

In short, this IEF-PAGE result shows that majority of S-crystallins are positively charged at all positions in the lens. There is a wide distribution of protein PIs, ranging from 5 to 9.5. This is consistent with our SDS-PAGE result in that the lens is composed of a mixture of dozens of S-crystallins. The super long loop protein with molecular weight 36 kDa is negatively charged because the long loop contains more negatively charged amino acids. Across the different positions in the lens, there is not much difference in the PI of the protein composition.

If the charge of the proteins are high enough and the phase of the protein assembly in the lens is repulsive glass, one would expected to see individual proteins in their native form. This will be addressed in the next section.

2.9 Native PAGE

To study the quaternary structure of the S-crystallins in their native form, I performed Native PAGE on the lens proteins. A *Loligo forbesii* lens was thawed, dissected into four concentric layers and placed into 0.9 ml 150 mM phosphate buffered saline (PBS) (P4417, Sigma Aldrich, St. Louis, MO) with protease inhibitor cocktail (P8849-5ml, Sigma

Aldrich) 1:100 added. Each of these samples was ground in a Dounce tissue homogenizer (06-434, Fisher Scientific, Pittsburgh, PA), and the supernatants were taken after centrifuging at 13.4 k rcf for 10 min. This step removes undissolved proteins. Then these solutions were added 1:1 to the native PAGE sample buffer (#161-0738, Bio-Rad). The running buffer was Tris/Glycine buffer (#161-0734, Bio-Rad) and the gel was a 12-2 well Criterion 12.5% Tris-HCl (#345-0014, Bio-Rad). Cytochrome c (C2506-50MG Sigma Aldrich) was used as the standard. The gel was run at 150 V at 4 °C with the electrodes switched so that positively charged proteins would migrate into the gel. The voltage was removed when the cytochrome c (which is orange in its native form due to the presence of a heme ligand, and therefore visible in the unstained gel) had migrated to the bottom of the gel. Bio-Safe Coomassie G-250 (#161-0406, Bio-Rad) was used as the protein stain, following the manufacturer's instructions. All of these steps were done with careful temperature considerations by either conducting the experiments on ice or in a temperature controlled system set at 4 °C. Gels were imaged and protein bands quantified as described in the SDS-PAGE methods above.

The native PAGE gel shows that the lens proteins barely migrated into the gel, whereas the cytochrome c monomers and dimers migrated to the end of the gel, shown in Figure 2.7. The lens proteins in the gel do not focus into a band, as cytochrome c does. Instead, there is a large smear of proteins that barely migrates into the gel. In native PAGE, proteins migrate according to their charge to mass ratios. The charge to mass ratio of cytochrome c is roughly double that of an average S-crystallin. Therefore, if S-crystallins were soluble dimers in their native form, they should migrate roughly 60% of the distance that cytochrome c migrates. However, our result shows that S-crystallins have barely entered the gel when cytochrome c has migrated the full length of the gel. This is consistent with S-crystallins existing in an aggregated form in their native state.

The native PAGE result suggests that S-crystallins in their native form are bound to

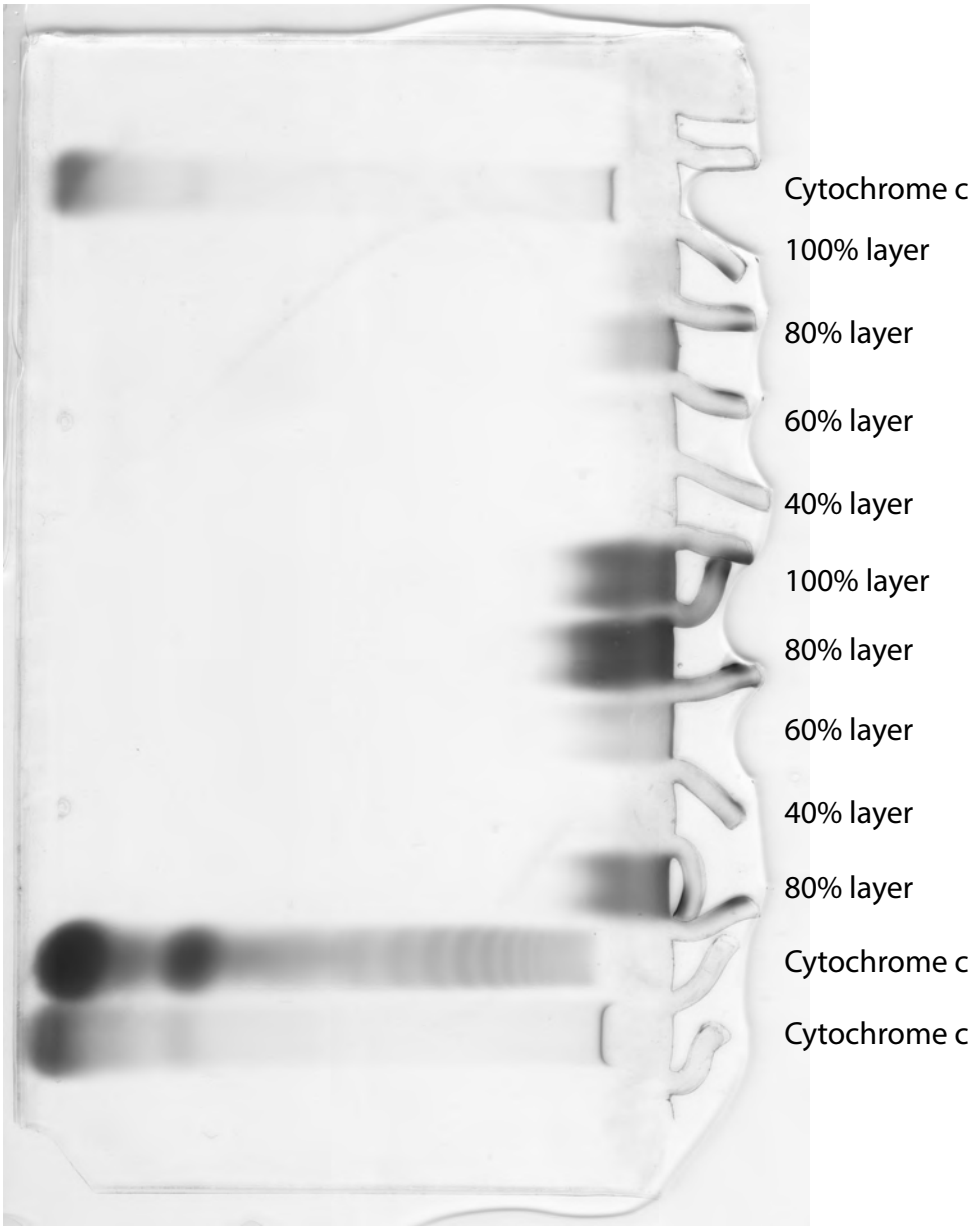


Figure 2.7: Native-PAGE electrophoresis of lens protein, with lane 6-10 a higher amount of protein loaded compared to lane 1-5. Cytochrome c was used for calibration purpose. The lens proteins at all radial positions do not migrate into the gel much. The proteins do not focus into a band, either. This is consistent with S-crystallins existing in an aggregated form in their native state.

each other in a manner that is not disrupted by homogenizing and re-suspending them in an excess of aqueous solvent. Instead, they may form a larger scale structure, which is consistent with a gelled structure. Therefore, our original hypothesis of repulsive glass is not supported by this experiment. I performed one more experiment on the native lens protein assembly to test the original repulsive glass hypothesis.

2.10 Filtration

The filtration experiment was conducted to test whether S-crystallins are gelled in the lens. I filtered aqueous lens solutions, prepared as described above, through 100 kDa and 30 kDa filters. The idea is that if the lens proteins form an assembly with the average molecular weight higher than both of the filters, few proteins are expected to pass through the filter.

A lens was dissected into four concentric layers and ground in 0.7 ml PBS buffer with protease inhibitor cocktail 1:25 added. The sample solution was centrifuged at 14,000 rcf for 10 min and the supernatant was obtained. An additional 0.5 ml PBS buffer (with protease inhibitor 1:25) was added to each tube. The mass of total protein in each solution was estimated separately by measuring the absorbance at 280 nm using a NanoDrop 2000 UV-Vis spectrometer (Thermo Scientific, Wilmington, DE). Dilution was conducted if the protein absorbance was higher than 1.1. Then the supernatant was filtered using Amicon Ultra 0.5 ml centrifugal filters with 100 kDa and 30 kDa molecular weight cutoffs (mwco) with 450 μl of the solution loaded (UFC510024 and UFC503024, EMD Millipore, Billerica, MA). Protein solutions were filtered through these devices by centrifuging at 15,000 rcf for 15 min. Any solution that did not pass through the filter device was collected by placing the device upside down in a fresh tube and centrifuging at 1,000 rcf for 2 minutes. All the above was performed either on ice or at a set temperature of 4 $^{\circ}$C. Then the volume of the protein solution in the filtrates were measured using a micropipettor, and the protein

concentration of the filtrate and the remaining fluid was estimated by absorption at 280 nm. The absorbance of the concentrates, which stayed in the filter after the filtration process, showed a broad peak around 305 nm. So I cannot use the 280 nm absorption to quantify the protein concentration. To check whether there is a high lens protein density in the concentrates, SDS PAGE was performed using another lens treated by the same dissection and grinding method. The SDS-PAGE method is described in section 2.5.1. Both the filtrate and concentrate proteins were loaded into SDS PAGE for comparison. The filtrate protein was diluted 1:2 with Laemmli sample buffer and the concentrate protein was diluted 1:4.

To estimate the protein concentration from absorbance at 280 nm, the extinction coefficients of the lens proteins were estimated using the ProtParam tool from ExPASy (http://web.expasy.org/). The input amino acid sequences come from our RNA sequencing result. The extinction coefficients of all proteins with full sequences were calculated and the average of them was used to estimate the protein concentration. The molecular weight of the proteins was assumed to be 25 kDa. Having estimated the protein concentration, I calculated the total mass of the protein originally in solution and calculated the percentage of the protein which passed through the filters.

In every 450 μl solution, the PBS dissolved protein mass after grinding are 0.55 mg for the solution at 100% layer, 1.45 mg for the 80% layer, 0.086 mg for the 60% layer, and 0.015 for the 40% layer, as is shown in Table 2.2. The proteins dissolve much more in the peripheral two layers compared to the core. After 100 k and 30 k mwco filtration, the mass of protein in the filtrate decreased to 29% relative to the original solution for the 100% layer, 5.1% for the 80% layer, and 12.8% for the 60% layer. This trend holds with the use of 30 k filters. The decrease of the protein concentration is because the size of the protein native form in the PBS solution is bigger than the filter size, which are 100 kDa and 30 kDa. To check whether the proteins were trapped in the filter, I flipped the filter and centrifuged lightly to get the concentrate. SDS-PAGE was performed to verify that the concentrate

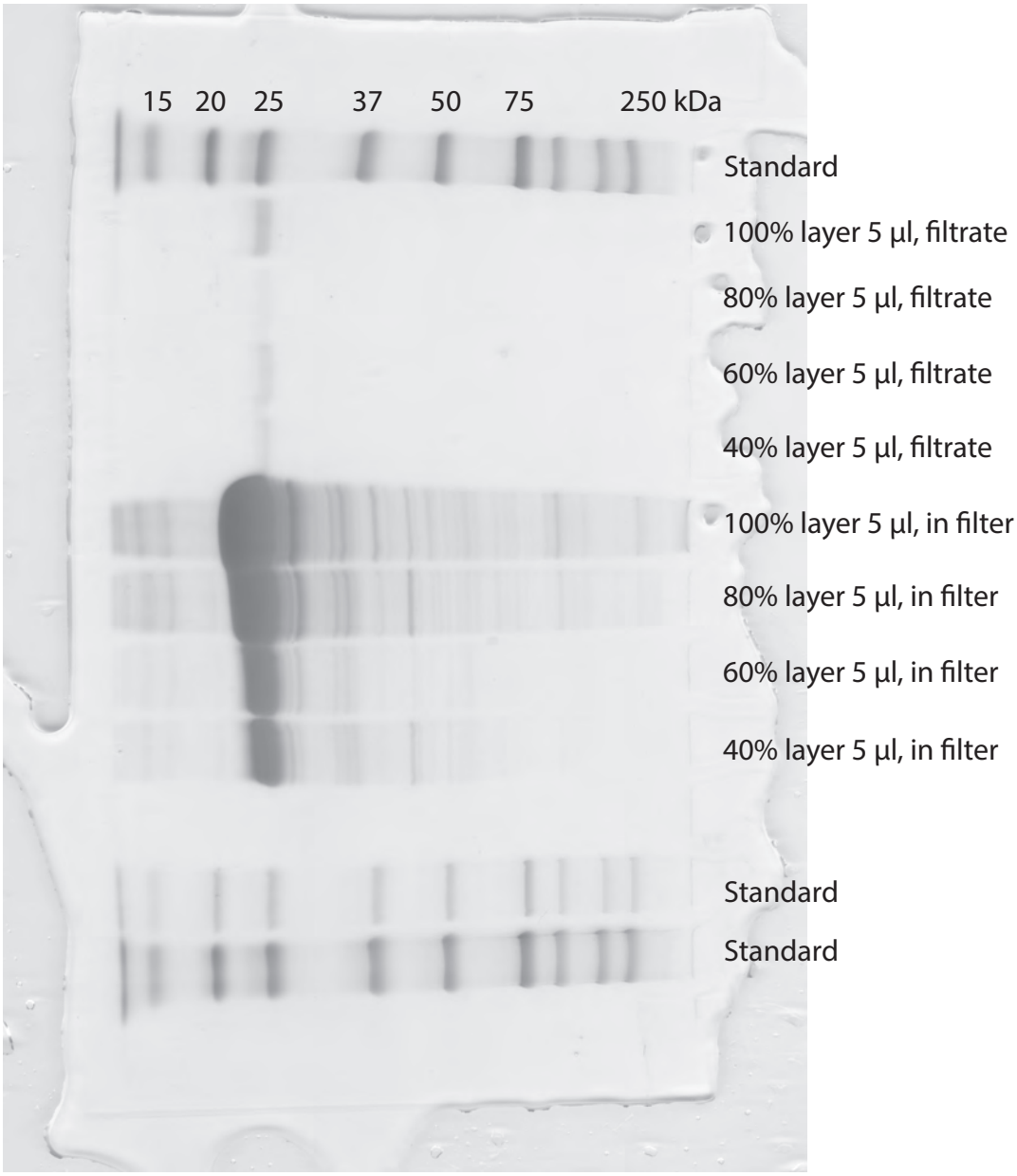


Figure 2.8: Filtration result. The SDS-PAGE was performed after the filtration experiment. The result shows that the proteins in both the filtrate and concentrate have the molecular weight consistent with S-crystallins. The concentration of protein in the concentrate is much higher than in the filtrate.

	Total protein mass (mg) in sample	Protein mass (mg) in filtrate, 100k mwco	% of total protein smaller than 100k mwco	Protein mass (mg) in filtrate, 30k mwco	% of total protein smaller than 30k mwco
100% layer	0.55	0.16	29.1	0.16	29.1
80% layer	1.45	0.074	5.1	0.069	4.8
60% layer	0.086	0.011	12.8	0.0096	11.2

Table 2.2: S-crystallin aggregate size distribution – filtration analysis

is composed of lens protein, shown in Figure 2.8. Both the filtrate and concentrate lanes of four layers show protein bands at 25 - 30 kDa, which is consistent with the molecular weight of S-crystallins. The filtrate proteins (1:2 diluted) are barely visible for all layers after stain whereas the concentrate (1:4 diluted) are massively over loaded. Therefore, the filtration result shows that majority of the lens proteins do not go through either 100 kDa or 30 kDa filter. Thus, the proteins in the lens forms a structure which is bigger than 100 kDa, which is consistent with our native PAGE result that the proteins form a gel structure.

The filtration result gives further support to the hypothesis that the phase of the lens proteins is a gel. This structure will further be discussed in chapter 3 and chapter 4.

2.11 Transmission electron microscopy (TEM)

To give a rough estimation of the protein configurations in the lens, I performed TEM on fixed lens tissue. Osmium tetroxide (2%) was used to fix the peripheral layer of a squid lens. When moderate blackening occurred, samples stored in deionized water. Water was then replaced with hexamethyldisilazane (99.9%) allowed to air dry, and samples were then placed in acetone for subsequent embedding in Spurr's resin (Electron Microscopy Sciences, Cat. 14300, Hatfield, PA). We cut 35 nm thin sections onto copper grids for imaging. Images were obtained using a Tecnai TEM at 8 keV.

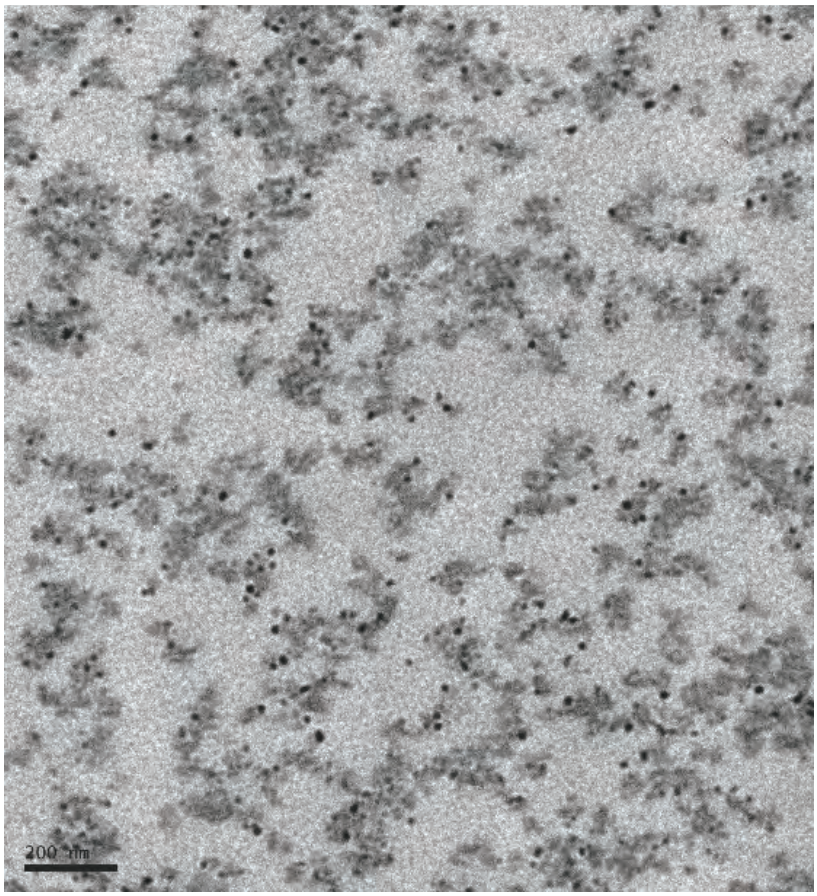


Figure 2.9: Transmission electron microscopy (TEM). The TEM image shows that the proteins appear to be connected to each other and form a network. This network spans the whole space, and it is consistent with a gelled structure as we have seen from several experimental results.

The TEM result is shown in Figure 2.9. The TEM image shows that the proteins appear to be connected to each other and form a network. This network spans the whole space, and it is consistent with a gelled structure as we have seen from several experimental results. This network from TEM also shows some chain-like structures with bifurcations. The interpretation of this structure will be discussed in Chapter 4.

2.12 Conclusion

In this Chapter, I performed a series of experiments attempting to study the physical properties of the lens from the perspective of protein composition. I found that the graded refractive index of the lens comes from a gradient in protein density, and estimated the protein packing fraction from the gradient refractive index requirement. This estimation is consistent with the density measurement at the core of the lens, and the protein packing fraction ranges from 5% to 100% from periphery to the core of the lens.

I conducted another series of experiments to study the protein components and their radial distribution in the lens. I studied the relative abundance of the S-crystallin isoforms at different radial positions of the lens using the SDS-PAGE method. The result shows that the relative abundance of 26.5 kDa S-crystallins compared to other isoforms in the same location is higher in the periphery, whereas the core is dominated by short loop proteins. Decomposition of the SDS-PAGE migration pattern using molecular weights predicted by our RNA sequencing data verifies this. Then, the atomic structures of the S-crystallins isoforms were estimated using homology modeling. This result suggests that the basic S-crystallin structure is similar to 1GSQ, the related enzyme present in squid, but with an additional loop-like structure with various amino acid numbers. Combining these results, these loops are highly correlated with the optical properties of transparency and refractive index in the lens. Long loops and short loops may therefore serve different functions for the optical function of the lens. My circular dichroism study on the lens protein shows that the proteins are folded, and consistent with globular proteins.

I also studied the charge distribution with lens position of these S-crystallin isoforms by IEF-PAGE. The result shows that majority of the proteins are positively charged, but there is not a major difference in average charge of the proteins at different lens radii. Native PAGE and filtration measurements showed that proteins most likely form a larger-scale

structures in their native form. The data in this chapter are therefore not consistent with our first hypothesis that proteins in the lens form a repulsive glass.

The implications of the combined results are twofold: the proteins in the lens are possibly gelled, and the average length of the loop-like structures present on the proteins is correlated with the optical properties of the gel formed at each position in the lens. We hypothesize that these loops may serve as sticky spots on each protein that link to other proteins to form a gel. We will further investigate this hypothesis in the following two chapters.

Chapter 3

Small angle x-ray scattering (SAXS) - a direct measurement of the squid lens proteins

In the previous chapter, I studied the protein components of a squid lens using biological methods. The results suggest that the disordered loops that apparently protrude from the conserved, folded body of the protein play a crucial role in protein assembly, and this assembly is consistent with a gelled material. In this chapter, I will continue to characterize the lens protein structure using SAXS at different layers of the squid lens, both for intact lens tissue and for lens proteins homogenized in solution. SAXS is a direct measurement of the protein configurations in Fourier space, and from this, it is possible to infer the spatial distribution of the proteins in real space. The hierarchical structure of protein assembly in the lens at all observed protein densities will also be constructed. These data will then be compared with the results from last chapter.

3.1 X-ray apparatus

To start, I will introduce the SAXS method in this section. The basic SAXS setup is straight forward: X-rays are generated from a source, and then this beam is collimated. Normally, a single x-ray wavelength is chosen. Then, the light passes through a pinhole and impinges on the sample in the sample chamber. The sample scatters the beam, and this scattering pattern is recorded by a detector. The experimental apparatus is shown in Figure 3.1. The output of the measurement is the x-ray intensity as a function of wave vector q, which is proportional to the scattering angle 2θ for small θ.

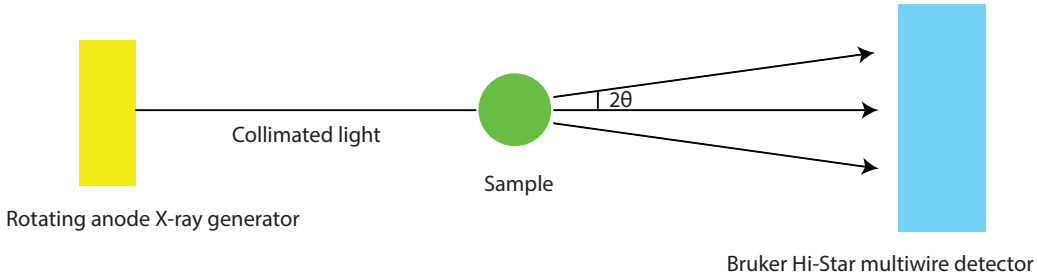


Figure 3.1: Beam path in an X-ray apparatus. An x-ray beam is generated by a rotating anode. The collimated light interacts with sample, and the scattering is recorded by a detector. The angle between the incident beam and the scattered beam is 2θ.

For this thesis, small angle x-ray scattering experiments were performed both in a multi-angle x-ray scattering system (MAXS) at University of Pennsylvania and at the X9 beam line of the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory. For both systems, we used silver behenate powder for calibration of scattering angle. I used two sample cabinets at different positions along the beam, allowing measurement of three angular ranges. The distance of these cabinets from the x-ray generators were 150, and 54 cm, for small and intermediate angle measurements, respectively. The small angle measurements collect data at $q \in \left[0.008, 0.13\right]$ Å^{-1}, whereas the intermediate angle ranges at $q \in \left[0.025, 0.35\right]$ Å^{-1}. If not otherwise described, sample temperature was fixed at 10 $^{\circ}$C for all experiments.

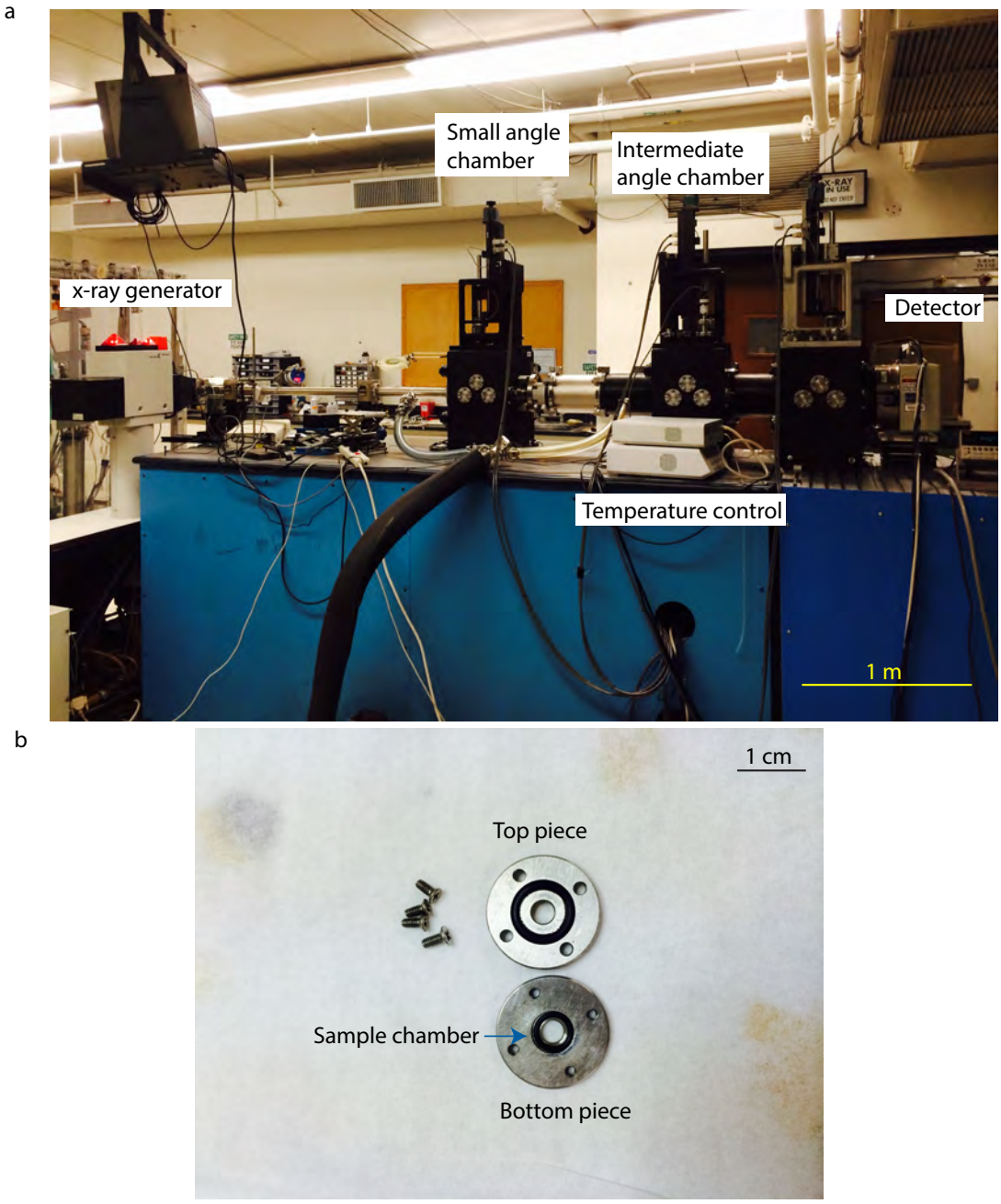


Figure 3.2: Image of the MAXS and the sample cell. The top image (a) shows the MAXS apparatus. The x-ray generator is on the far left of the image. The two black boxes are the sample chambers at the small and intermediate angle ranges. The bottom image (b) shows the top and the bottom pieces of the sample cell. Samples are loaded inside the small o-ring, and mica are used as the window.

For the MAXS system, temperature was controlled by a TMS 94 and LNP 94/2 temperature control unit (Linkam Scientific Instruments Ltd, Waterfield, Tadworth, United Kingdom). An incident beam with energy of 8.03 keV was generated using a Bruker Nonius FR591 rotating-anode source (BRUKER AXS, Inc. Madison, WI). The beam was collimated and pinholes were placed. X-ray scattering patterns were then collected with a multiwire detector (Bruker Hi-Star). At this facility, vacuum between the generator and the detector was maintained below 1 Torr to reduce air scattering (Molmex Scientific Inc., Northampton, MA). The setup is shown in Figure 3.2. For the corresponding wave vector range of the squid lens proteins and pickled quail eggs, two measurements were done, at small and intermediate angle ranges. These data were spliced together to create a single measurement with a larger angular range than either chamber separately.

The procedure for performing scattering measurements using the MAXS system is summarized here: to begin an experiment, the first thing is to raise the voltage and current of the x-ray generator since the intensity of output x-rays may fluctuate a little within one hour after reaching appropriate levels for a measurement. Before turning on the detector, the beam block was adjusted to prevent possible detector damage. Then, the thermal cell holder was connected to the temperature control unit, and the horizontal and vertical positions of the holder were adjusted to maximize the output intensity. Calibration of the beam was repeated every 24 hours during an experiment.

For all measurements at NSLS, calibration was done with the incident x-ray energy at 14 keV. Then the energy was lowered to 8 keV for measurements, and the corresponding scattering angle was shifted accordingly. Measurements at the NSLS were done in air due to constraints of the sample holder geometry at that facility. Similar to the MAXS system, for each experiment, two scans were performed at the wave vector range of 0.002 - 0.08 and 0.08 - 0.6 Å^{-1} using two detectors. Measurements for the same sample from these two angle ranges were concatenated after the experiment.

For x-ray scattering measurements, dissected pieces of lens tissue or pidan were mounted in custom-built o-ring sealed sample holders with clear ruby mica windows (7.5 mm and 12 mm, Attwater Group, Preston, England). This cell is shown in Figure 3.2b. To ensure that the proteins are not denatured and the x-ray scattering signal comes from the sample, the procedure that I developed to use this sample cell is the following: A big o-ring (McMasterCarr 9262K445) and a small o-ring (McMasterCarr 9262K421) were cleaned with deionized (DI) water first. I then cleaned the piece of mica with DI water and compressed air (McMasterCarr 7437K35) if necessary. Samples were normally loaded under a stereoscope. Solid samples were loaded so that the height of the sample at the top of the sample did not exceed the top of the small o-ring in the sample cell. For liquid samples, the volume of the liquid for a good seal was adjusted so that the top of the liquid was just a little higher than small o-ring. Once the liquid was loaded, it was immediately sealed with a big mica cover to prevent leaks. To ensure a good seal under vacuum, the screws (Figure 3.2) at the top of the top half of the sample holder were tightened uniformly so that the side of the sample cell (top and bottom) were parallel to each other.

3.2 Kinematical scattering on non-crystal materials

In the previous section, I introduced the SAXS methods used in this thesis. Now, I will briefly introduce the meaning of the SAXS intensity as a function of wave vector q, and also explain how I calculated the S-crystallin form factor.

3.2.1 SAXS measures the Fourier transform of the particle radial distribution function

X-rays are electromagnetic waves with a wavelength on the order of magnitude of 1 Å. The x-ray photons interact with electrons of matter. Therefore, x-rays are a convenient tool for

characterizing the spatial distribution of electrons in a sample. The SAXS measurement records the scattering intensity at different wave vector q, where q is defined as

$$q = 4\pi \sin(\theta)/\lambda. \tag{3.1}$$

In this equation, θ is the half-angle change (Figure 3.1), and λ is the x-ray wavelength. The wave vector q is in reciprocal space, and it corresponds to a scale (r) in real space with the relation,

$$q = 2\pi/r\ . \tag{3.2}$$

Therefore, in x-ray scattering data, smaller wave vectors correspond to larger spatial scales.

The scattering intensity is the modulus squared of the scattering amplitude. For a mono-molecular system, the scattering intensity can be written as

$$I(q) = f(q)^2 \sum_n \exp(iqr_n) \sum_m \exp(-iqr_m), \tag{3.3}$$

where $f(q)$ is estimated to be the number of electrons in an atom, n and m are indices to label particles. Because SAXS measures the particle correlations, the relation between the radial distribution function $g(r)$, which is the probability of finding a particle at the position r, can be calculated as

$$g(r) = 1 + \frac{1}{2\pi^2 r \rho_{ave}} \int_0^\infty q\Big[\frac{I(q)}{Nf(q)^2} - 1\Big] \sin(qr) dq, \tag{3.4}$$

where N is the total number of particles and ρ_{ave} is the average density. Therefore, the result of a SAXS measurement is the Fourier transform of the radial distribution function.

3.2.2 Form factor of S-crystallin

In any material, there are always interactions between particles filling a volume. Under these interactions, patterns form between the particles and the particles are correlated. What SAXS probes, then, is the sum of these correlations within the hierarchically ordered structure. But for practical purposes, such as in our squid lens case, the structures of interest are the larger-scale structures that result from the interactions of S-crystallins, apart from the aqueous solvent background. Because these proteins have a well-characterized structure, the contribution of individual proteins to the scattering intensity at length-scales larger than a single protein can be calculated. To do this, we deconvolve the contribution to scattering from the internal structure of S-crystallins from contribution to scattering from the assembly of many S-crystallins. Now, the scattering intensity from the individual proteins is called the form factor, and the larger structure they make is called the structure factor. In this section, I will focus on the methods to obtain form factors for the lens proteins and pidan proteins. Later, in section 3.4, the structure factor for the intact squid lens will be calculated and analyzed.

Form factor is a measure of the scattering amplitude of x-rays from an isolated protein. It corresponds to the inner structure of the protein and its scattering envelope. To obtain the form factor of a system, it can either be measured directly by SAXS, or calculated from the positions of atoms in the protein. The SAXS measurement to obtain a form factor is normally performed on the protein of interest in a low density in solution, so that particle-particle interactions are negligible and proteins are isolated from each other. In the scenario when the protein structure is known but it is hard to dilute the protein, the form factor can be calculated using the Debye formula [33]:

$$f(q) = \sum_i \sum_j N_i^e(q) N_j^e(q) \frac{\sin(q d_{ij})}{q d_{ij}}. \tag{3.5}$$

Here, $N_i^e(q)$ is the number of electron in the i^{th} atom. The variable d is the pair distance between atom i and j.

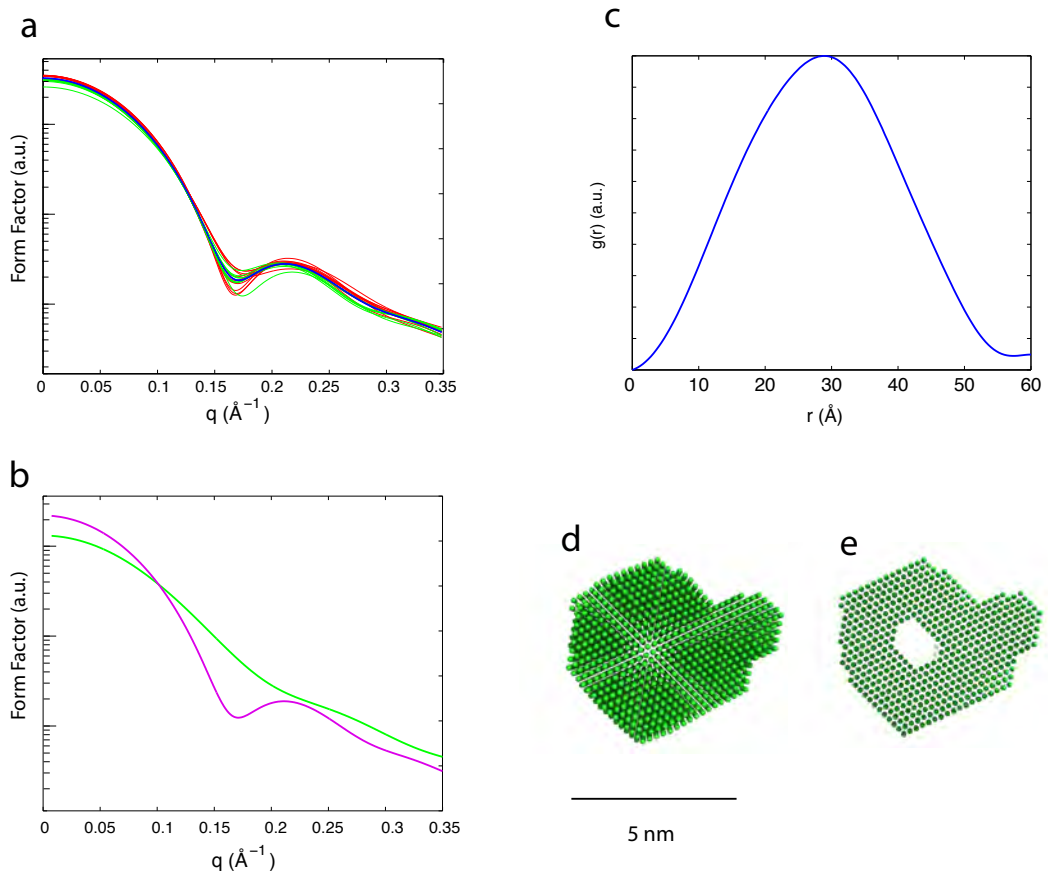


Figure 3.3: The form factors of S-crystallin isoforms calculated from homology models are plotted (a) with red as the long-loop S-crystallins and green short-loop S-crystallins. They show very similar results. The average calculated form factor of all isoform dimers is plotted in pink in (b). The green in (b) is the form factor from a S-crystallin monomer. The peak at $q = 0.21\text{Å}^{-1}$ comes from the dimeric structure. The calculated RDF ($g(r)$) is plotted in (c). One likely spatial configuration of the average S-crystallin obtained from the average form factor is plotted in (d) and (e) at different cross sections.

We found that the natural S-crystallin network was almost impossible to stably disassemble into individual S-crystallin subunits (see sections 2.8 and 2.9); therefore, we used the strategy of calculating form factor from atomic positions inferred from a crystal structure [99]. Since the lens tissue is predominately composed of S-crystallin proteins, we used S-crystallin homology models to calculate the form factor of S-crystallin proteins. We as-

sumed for simplicity that all S-crystallins in the lens are homodimers. The predicted form factors of all S-crystallin homodimers are similar, as is shown in Figure 3.3a: they are flat at small values of $q < 0.05$ Å^{-1}, then decrease to a minimum near $q = 0.17$ Å^{-1}, with a secondary peak at $q \approx 0.21$ Å^{-1}. The main difference between isoforms is the minimum value near q = 0.17 Å^{-1}. The ratio between the peak at $q = 0.21$ Å^{-1} and the valley at q = 0.17 Å^{-1} ranges from 1.2 to 2.3. The average form factor for all isoforms has all these features, and from this we calculated an average radius of gyration (R_g) for S-crystallin of 15.8 Å.

In Figure 3.3b, the average peak in the form factor at $q = 0.21$ Å^{-1} over all S-crystallin isoforms corresponds to the size of the dimer structure, formed by two interlocking monomers. This peak is absent in form factor of an S-crystallin monomer (Figure 3.3b, pink). Further, the radial distribution function of the S-crystallin dimer was estimated using the software GNOM (for the method, see Chapter 4), and the RDF shows only one peak at 28 Å in Figure 303c. The reciprocal space associated with this peak is estimated by $2\pi/28$ Å$= 0.22$Å^{-1}, so this real space RDF peak corresponds to the $q = 0.21$ Å^{-1} peak in reciprocal space. The distance 28 Å is approximately the distance between the centers of the two folded monomer chains within the homodimer. To show the spatial configuration associated with this peak, we turn to the software DAMMIF (for the method, see Chapter 4). The DAMMIF output on the averaged form factor shows the approximately spherical structure of an S-crystallin dimer. Therefore, in the average form factor of S-crystallins, the peak at $q = 0.21$ Å^{-1} corresponds to the presence of the dimeric structure.

Having calculated the form factor of S-crystallins, it is possible to study the protein-protein interaction from a SAXS experiment on intact lens tissue.

3.3 Intensity versus wave vector of S-crystallins

We performed SAXS on the intact lens tissue. Each lens was dissected into four radially concentric layers. The tissue was loaded into the thermal cell as described above. The volume of the tissue was sufficient to fill the whole sample chamber for the outer three layers, and was positioned in the middle of the x-ray beam for the core of the lens sample, since the volume of this sample was smaller than that of the sample chamber. The measurements were carefully examined to make sure that the scattering comes from the lens tissue, not the background: At low q, the background scattering from the air shows a straight line with the slope sharper than -3 in a log-log plot. The scattering from lens tissues show a slope around -2 (shown in section 3.5), and this signal is considered as the lens protein scattering. The absolute output scattering intensities were adjusted so that they overlap at very large q.

Figure 3.4 shows that the scattering intensity as a function of wave vector I(q) is very similar for each layer of the lens, even though the layers transition from $\phi = 0.05$ to $\phi = 1.0$: At small q, the scattering intensity for all layers decreases with increasing q and there is a shoulder or peak for all layers near $q = 0.02$ Å^{-1}. At the intermediate range, there is another shoulder or a peak around $q = 0.15$ Å^{-1}. Then at a large q, there is another peak, which corresponds to the dimer configuration at $q = 0.21$ Å^{-1}.

In all layers, in the range of 0.006Å$^{-1} < q < 0.01$ Å^{-1}, the scattering intensity decreases for all layers with increasing q. This feature is consistent with the larger-scale structure forming a gelled state [7].

In the ranges of $0.01 < q < 0.02$ Å^{-1} and $0.04 < q < 0.13$ Å^{-1}, the scattering intensity is greater in tissue samples taken from the peripheral layers of the lens relative to a sample taken from the core (Figure 3.4). The intensity at the core of the lens at all measured q is lower than in any other layers in the lens. It is likely that at the core, the packing fraction is so high that it is not possible to have high density fluctuation.

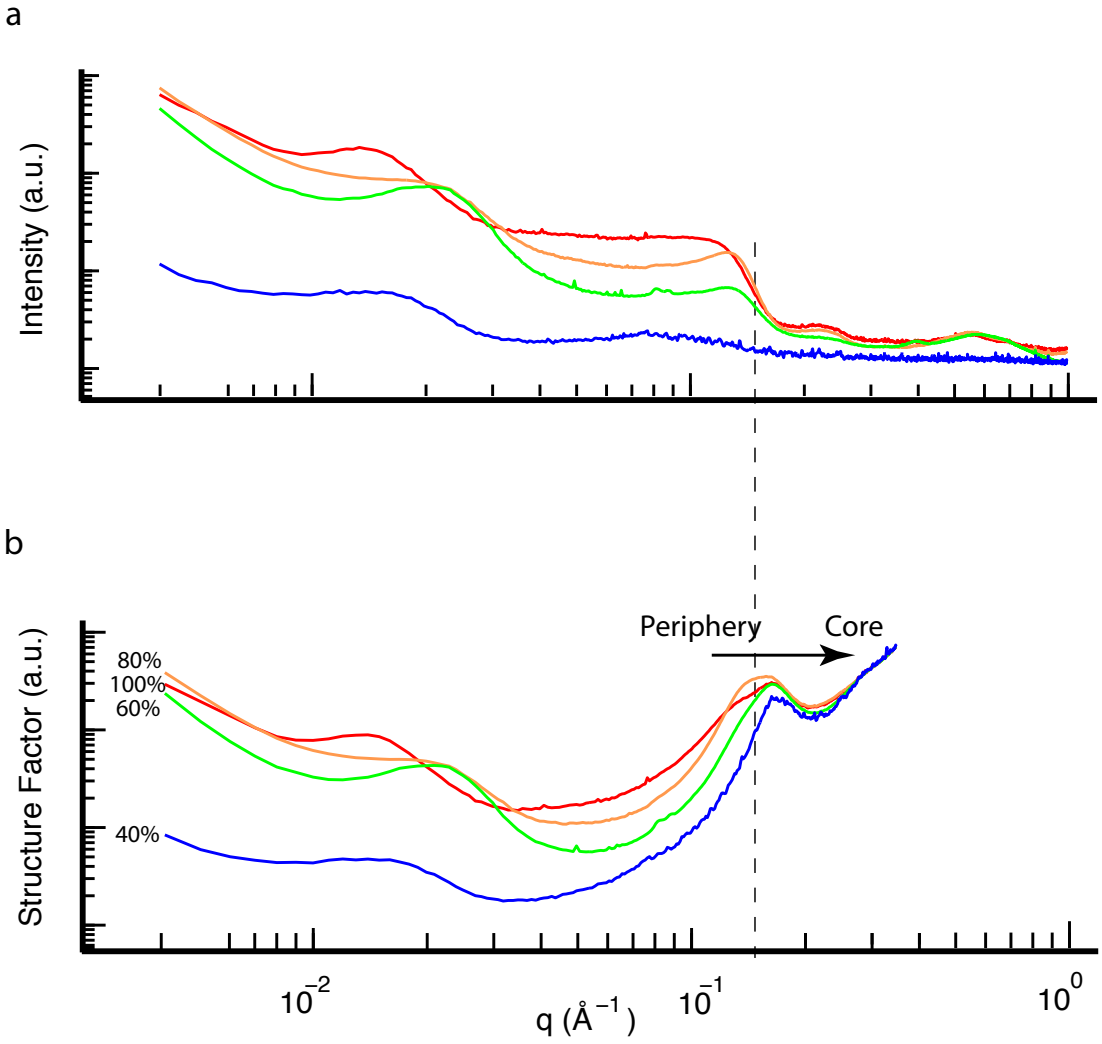


Figure 3.4: SAXS result. The measured intensity (a) and the structure factor (b) at different lens radial positions are shown. Red labels 100% layer, orange 80%, green 60%, and blue 40%.

The SAXS curves show a peak at $0.01 < q < 0.03$ Å^{-1} in all four layers in Figure 3.4. This peak is possibly coming from the protein network, or from another aspect of the cellular nature of the tissue, such as the cell membranes. In section 3.6, I will discuss the possible interpretations of this peak.

At the range of $0.07 < q < 0.2$ Å^{-1} in Figure 3.4a, there is a shoulder or peak next to a plateau for all layers in the lens. This scale is in the range of the protein-protein interaction. The scattering intensity of the plateau decreases from the periphery to the core of the lens. The sharp increase with the decrease of q at $0.1 < q < 0.2$ Å^{-1} is very similar to the scattering intensity from spheres. This observation suggests that in this q range, which is 30 - 60 Å in real space, it is very likely that there are sphere-like structures in real space. Comparing this to what we learned in the previous chapter, it is natural to associate these spheres with the roughly spherical S-crystallin dimers. This will be discussed more in section 3.4 with the structure factor.

At the range of $q > 0.2$, there is one peak at $q = 0.21$ Å^{-1} and another peak at $q = 0.6$ Å^{-1} (Figure 3.4). The $q = 0.21$ Å^{-1} peak corresponds to the dimeric structure of the S-crystallins, and the $q = 0.6$ Å^{-1} corresponds to the inner structure of an S-crystallin monomer. These two peaks are more apparent for the outer three layers of the lens than for the core.

To carefully examine the SAXS result at different scales and to analyze the structure resulting from protein-protein interactions compared to the protein inner structure, I next calculated and analyzed the system's structure factor.

3.4 Structure factor of S-crystallins

The structure factor is the Fourier transform of the radial distribution function of 5 nm protein dimers. It neglects the inner structure of the protein, and only considers the spatial

distribution of these proteins. Knowing the structure factor can help us understand the spatial configuration of the protein assembly. The measured x-ray scattering intensity as a function of wave vector, q, is the product of the form factor and the structure factor. Therefore, the structure factor for both pidan (see Chapter 6) and the lens tissue can be determined by dividing the scattering intensity by the appropriate form factor, calculated as described above.

The structure factors of the intact lens samples are shown in Figure 3.4b. They are similar at all radial positions of the lens. Because of these similarities, I will discuss the four prominent scattering features seen in samples taken from all four relative radial lens positions at increasing q: A large scale gelation feature at $q < 0.01$ Å^{-1}; a shoulder present at $0.01 < q < 0.03$ Å^{-1}; a trough in scattering at $0.03 < q < 0.1$ Å^{-1}; a peak in structure factor due to direct protein-protein interactions at $q = 0.15$ Å^{-1}.

At spatial scales from $0.006 < q < 0.01$ Å^{-1}, the scattering intensity decreases in a straight line in a log-log plot. This indicates there is a fractal character of the protein-protein interactions of S-crystallins. This fractal nature also reveals that there is a larger-scale structure of the S-crystallins in the lens. There are only subtle differences in the x-ray scattering pattern between different layers, which indicates that the underlying structures at different radial positions are similar. This larger-scale structure will be discussed in section 3.5.

At intermediate scales of $0.01 < q < 0.03$ Å^{-1}, there is a shoulder with spatial scale around 400 Å and then a sharp decrease. This feature will be discussed in section 3.6.

Then at a smaller spatial scale, a local minimum appears at $q \sim 0.05$ Å^{-1}. The depth of the minimum increases from outer to the core of the lens, from 20 to 110 calculated by using the protein-protein contact peak at $q = 0.13$ Å^{-1} to divide by the minimum of the valley. A further analysis about this feature will be conducted in section 4.6.2.

A peak at the size of S-crystallin diameter appears at $q = 0.13$ Å^{-1}, which is about

50 Å in real space. This is the same as the diameter of the S-crystallin dimers. This means that the majority of the signal in structure factor arises from pairwise protein-protein interactions. This peak shifts towards a higher q from periphery to the core, indicating that the nearest neighbor distance decreases from the periphery to the core of the lens. This is consistent with the sharp decrease at $0.1 < q < 0.2$ Å^{-1} in the intensity in Figure 3.4. Both of these indicate that the protein dimer structure is observed in SAXS, and these S-crystallin dimers are the unit from which larger structures within the lens tissue are assembled.

In order to see the protein behavior in the very center of the lens, I performed SAXS on a tissue sample from the lens core, with the radius of the dissected tissue sample as small as possible. It shows a straight line at $q < 0.15$ Å^{-1}, indicating that the proteins are so densely packed that they are not distinguishable from each other in space. The spatial structure is so compact, it is consistent with dehydrated protein homogeneously spanning the whole core of the lens with no bulk water remaining. This suggests that in order to form such a dense state, possibly the protein surfaces are all surrounded and tightly packed with other proteins.

In this section, I described the results from my SAXS experiments. I showed that at spatial scales larger than a single S-crystallin dimer, there is a structural network. The protein-protein interaction peak at $q = 0.13$ Å^{-1} is clearly observed as well. This indicates that the pair-wise interaction between the S-crystallin dimers is crucial for the assembling of the lens system, and that individual 50-ÅS-crystallin dimers are likely to be the unit building blocks for this assembly.

3.5 Gel structure at large scale

As is mentioned from Chapter 2, the native PAGE and filtration experiment suggests that the lens proteins form a gel structure in the living tissue. In this section, I will demonstrate

evidence from SAXS that the lens is structured as a gel.

Gels are a state of matter which are solid but with very low density of the composite particles. There is no strict definition of a gel. The common characteristics of a gel structure are: interactions between particles are primarily attractive, and the system has yield stress, meaning that the composite particles are connected, or percolated. However, a system may not be called a gelled system if the density of the particles is either too high or too low. The density of a gel is very similar to a liquid. A gel at high density may be eventually jammed and becomes a solid state, whereas a low density system may be in a gas phase. However, there are no strict requirements delineating the physical definition of a gel.

Gels have some interesting properties. A gelled system is mostly adhesive and sticky. This comes from the attractive interaction between the components. Only under attractive interaction can large-scale structures form. Gelled systems also do not flow, and exhibit elasticity. This is because the network formed by the attractive particles spans the whole volume, and the particles in a gel are not free to move. Therefore, when a small stress is exerted on a gel, the percolated system reacts elastically at first, because the particles are all connected, and moving some particles requires breaking bonds, which is energetically costly. But when the applied stress is big, bonds start to break and the material response is not elastic any more.

In Chapter 2, I showed that the lens material is sticky, meaning that it sticks to the forceps. This is consistent with the attractive interactions of gelled systems. The conclusion that proteins are gelled within the lens tissue is further supported by our SAXS data. The structure factors increase with the decrease of q for all layers in the lens at the range of $q < 0.01$ Å^{-1}. A higher structure factor means that the spatial density fluctuation increases. If the system does not have any attractive interaction between the particles, the larger the spatial scale one probes, the more homogeneous the system appears. Therefore, the increase in structure factor with decreasing q indicates that there are larger scale structures

in the tissue, requiring attractive interactions between particles. Thus, structure factor at low q shows that there are attractive interaction between the proteins. Our old hypothesis [114] mentioned in Chapter 1, that the proteins in a squid lens are dominated by repulsive interactions is therefore very unlikely based on this SAXS result.

At $0.006 < q < 0.01$ Å^{-1}, the structure factor decreases as a power law with a slope in a log-log plot close to -2 for all layers in the lens. This slope indicates the fractal dimension d_f in the spatial scale that one probes. Fractal dimension is the power of the number of particles n as a function of the radius r in the region. One considers: $n(r) \propto r^{d_f}$. For a chain, the number of particles in a region with radius r is linear with r, and thus, it has $d_f = 1$. For a compact three dimensional ball formed by particles, the number of particle contained inside it is proportional to the cube of the radius, and $d_f = 3$. For a three-dimensional random walk model, it has $d_f = 2$. A polymer chain system is often described using the three dimensional random walk model and has $d_f = 2$ [7].

For the lens system, the fractal dimension d_f at $0.006 < q < 0.01$ Å^{-1} is approximately 2. This is higher than d_f of a chain, and is smaller than the d_f of a ball. Therefore, we can rule out the chain structure and ball structure for the bigger scale structure. The d_f of the lens system is very similar to the d_f of a three-dimensional random walk. The picture of the the three-dimensional random walk model looks like a chain-like structure with lots of bifurcations. The particles are all connected in the whole network, and the system can be a gel if the density of the particles are neither too big nor too small. Therefore, $d_f = 2$ in the lens system strongly suggests that proteins in the lens system are connected together and form a percolating, gelled network of bifurcating chains composed of pairwise S-crystallin interactions.

In short, the structure factor at large scales shows that the folded proteins are very likely to be linked together, forming a percolated network at large scales. The structure is more consistent with gel structure than with a jammed structure as previously hypothesized

[114]. There are only subtle differences in the x-ray scattering pattern between different layers, which indicates that the underlying structures at different radial positions share similarities and are all gel-like.

3.6 The peak at $0.01 < q < 0.03$ Å^{-1}

The SAXS curves or structure factors show a peak at $0.01 < q < 0.03$ Å^{-1} in all the four layers in Figure 3.4. There are a few possible explanations for this peak.

It is possibly corresponding to the imperfect homogeneity of the gelled system. In Shibayama's paper [100], he showed that a gelled system may have some expansions and contractions under the interaction of van der Waals, electrostatic, or hydrophobic interactions. In the lens system, the degree of density fluctuation consistent with optical transparency is required to be low, but not necessarily zero. This remaining density fluctuation may be captured by SAXS as the shoulder at $0.03 < q < 0.07$Å^{-1}. The possible fluctuation small enough that there is no clear peak in the SAXS curve. This is likely consistent with the observed transparency in the lens.

The density fluctuation responsible for this peak could also be due to the nodes of the protein network. At each bifurcation, there are more particles around the network nodes compared to particles within the chains that constitute most of the network. The Sciortino group's work on simulating networks of particles shows that nodes in these networks are typically complex and composed of many particles, and not simply one particle with three nearest neighbors as might be the simplest expectation [98]. This more complex picture of multi-particle network nodes might explain the shoulder at $q \sim 0.03$.

Another possible explanation is that this peak is not from S-crystallins in the cell, but instead from some feature of cell membranes or remaining nuclei, since the lens tissue is cellular. The thickness of the cell membrane is about 40 Å, but is likely to have arrays

of densely packed integral membrane proteins, possibly with length-scales around 300 Å. Although lens cells eject their nuclei to avoid intracellular scattering, there are still some stray, compacted nuclei remaining within the mature tissue, visible in light microscopy (data not shown) - this subtle feature at 300 Å in SAXS could also be some feature of tightly packed DNA remaining in the tissue.

To better understand this feature at 300 Å , we compared the structure of the squid lens system to the Chinese century egg (see Chapter 5). In the SAXS intensity of pidan, there is also a bump in the structure factor at low q, but there are no cell membranes present in the egg system. This observation suggests that an explanation for this feature in our SAXS data related to complex, multiparticle network nodes is more likely than one stemming from features of cell membranes.

It is also possible that the S-crystallin dimers assemble and form larger particles which are about 300 Å in diameter, just as the α-crystallins assemble into larger particles in vertebrate lenses. However, if this were the case, we would have expected to see these particles in our native PAGE experiment. Instead, our native PAGE result in S-crystallins shows a smear of proteins which does not have a clear peak at very large molecular weight, as is shown in section 2.9. Therefore, it is unlikely that the S-crystallin dimers form a discrete larger particle.

Therefore, the peak at $0.01 < q < 0.03$ Å^{-1} seems to be more likely coming from subtle inhomogeneity of the network and/or the structure of nodes in the network, similar to those described by Shibayama [104].

3.7 SAXS on diluted S-crystallins

To study the effects of dilution on the network of the S-crystallin dimers in the lens tissue, I performed SAXS on the dissolved lens tissue sample in different solvent conditions. I

used a high ionic strength buffer, phosphate buffered saline (PBS) (P4417, Sigma Aldrich, St. Louis, MO) with different concentrations, and a low ionic strength buffer, Tris-HCl, since interactions like hydrogen bonds behave differently at different ion concentrations. *Loligo* lenses were dissected into four concentric layers, and the 100% and 80% layers were homogenized in 650 mM PBS, 150 mM PBS, 30 mM PBS, and 20 mM Tris-HCl with pH = 7.4 respectively. All the sample buffers had protease inhibitor cocktail (P8849-5ml, Sigma Aldrich) at 1:100 v/v added, and the homogenization was carried out on ice. Then the protein solutions were centrifuged at 14 k rcp for 10 min at 4 °C to remove insoluble proteins, and supernatant was used for SAXS measurements. SAXS was done on each sample in both the small and intermediate angle chambers as described in section 3.1. Clean sample buffer for each sample was used as background, and background subtraction was done with the scattering intensities normalized with the exposure time and the intensity at $q = 0$.

In Figure 3.5, results show that at a small angle with $q < 0.12$ Å^{-1}, scattering intensities from solution-suspended S-crystallins still decrease, as was observed for measurements on intact lens tissue. This decrease indicates that the proteins are still assembled into large scale networks, even in homogenized, centrifuged solution. Hence, the protein network is robust enough so that grinding the lens tissue in buffers with different ion concentrations does not break the network, and/or the protein network quickly reassembles *in vitro* after this disruption. The peak at $q = 0.02$ Å^{-1}, possibly consistent with network nodes or another subtle inhomogeneity, is not visible in all of the dilution scattering intensities. It is likely that the dilution of the proteins decreases the spatial frequency of bifurcation of the protein chain-like structure, and decreases the inhomogeneity of the protein network as well. Alternatively, this observation is consistent with this peak emerging from membrane or nuclear components of the intact tissue, which would be removed by the high-speed centrifugation.

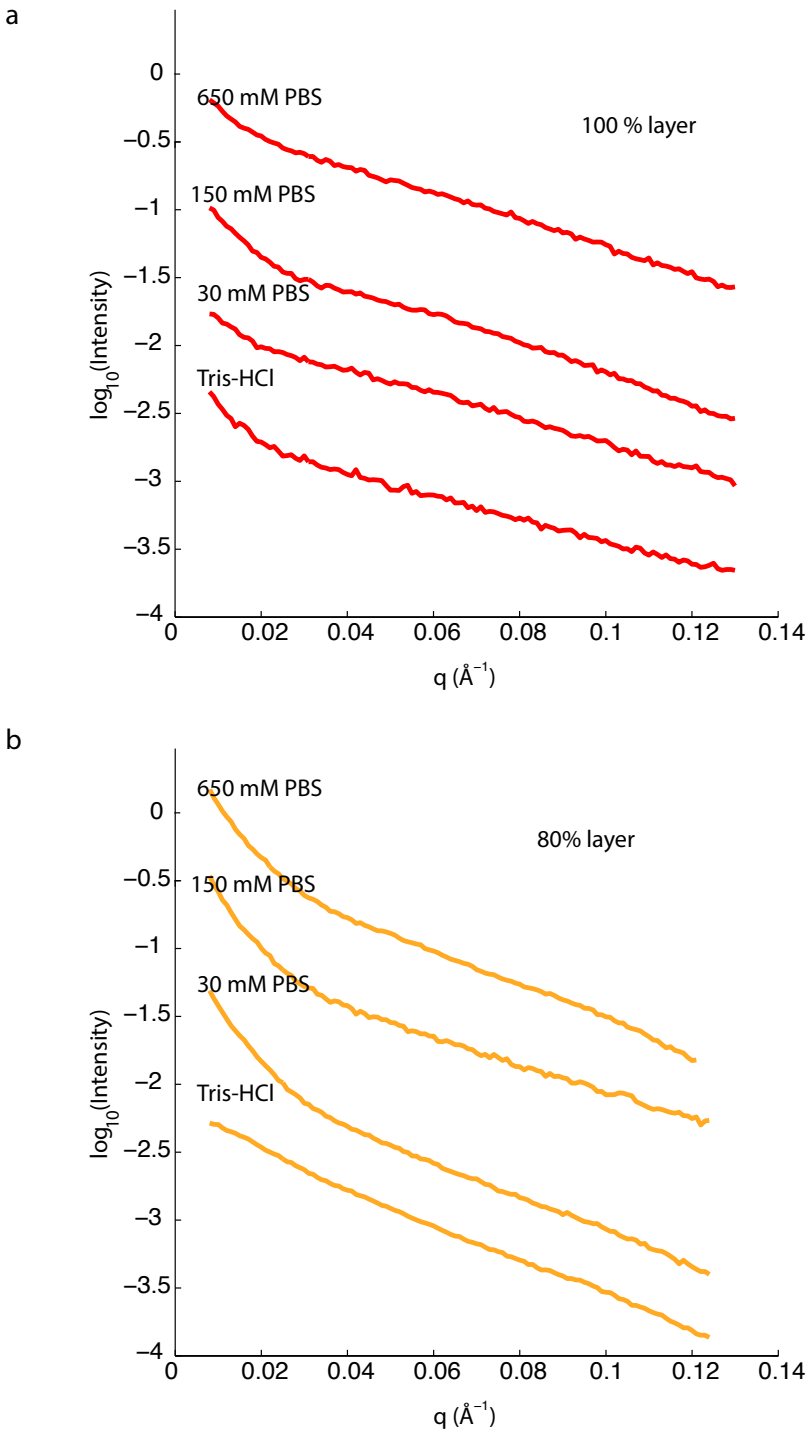


Figure 3.5: SAXS on ground lens tissue. The x-ray scattering on homogenized, soluble lens tissue in different buffer conditions, with red the 100% layer (a), and orange 80% layer (b). Vertical offset is arbitrary to show differences between the curves. The scattering intensities are similar, and they decrease with the increase of q. This indicates that the protein network structures are robust under a wide range of ionic conditions.

There is not a large difference in SAXS measurements in different sample buffers in the 100% layer. However, for the 80% layer, lens proteins in PBS solutions show that at $q < 0.03$ Å^{-1}, the decrease of the scattering intensity is faster than $0.03 < q < 0.12$ Å^{-1}. But for Tris-HCl solution, the scattering intensity shows a straight line. Having a less-sharp decrease at low q indicates a lower density fluctuation at the range. This suggests that the lack of ions in the Tris-HCl solution results in more homogeneously distributed network fragments in solution than in the high ionic strength buffers. Therefore, though the effect is subtle, ions play a role in the interactions between the proteins, and in the assembly of the network system.

The scattering intensity in the intermediate chamber at $q < 0.32$ Å^{-1} in different buffer conditions is very similar between different buffers (results not shown). There are no obvious peaks at $q = 0.13$ Å^{-1}, either due to the disruption of some of the nodes or second-order interactions in the protein-protein network, or due to the removal of other cellular components such as membranes and residual nuclei.

Therefore, S-crystallin network is robust, in that homogenizing the network in different buffers either does not destroy the network, or the network rapidly reassembles *in vitro*. However, ion concentration plays a subtle role in the network structure, suggesting that polar and electrostatic interactions play a role in assembling and maintaining the structure. At a smaller spatial scale, the SAXS intensity in these diluted lens samples suggests that the proteins are diluted homogeneously.

3.8 SAXS on uncooked and cooked squid lens

In the last section of this chapter, I will describe a temperature study that induces opacity in lens tissue, to study what features of the protein network are required for transparency. To do this, SAXS on heat-denatured squid lenses was performed and compared to the

scattering intensity of fresh tissue, described above. A typical procedure began with a SAXS scan on a healthy, transparent lens, which was carefully handled to prevent protein denaturation (as is described in Chapter 3). The lens was then heated inside the MAXS system to 50 °C, and a SAXS measurement was taken at this temperature. Then the lens was heated to 90 °C, and another SAXS scan was taken. After these heat treatments, the lens became opaque and white in appearance. Data reported in this section come from squid lenses of *Sthenotheuthis*. These lenses are usually bigger than *Loligo* lenses, which were used before. The lens used in the small angle range in the SAXS in this section has the radius of 4.5 mm, whereas the *Loligo* lenses usually have a radius of around 3 mm.

3.8.1 Cooked and fresh tissue in the periphery of the lens

Figure 3.6 shows the SAXS result of the fresh and heated 100% layer in a lens at both small (a) and intermediate (b) angle range. Different heat treatments are labelled as different colors. The scattering profiles at the small angle range ($q < 0.13$ Å^{-1}) show very different behaviors between different heat treatments. The scattering profiles at the intermediate range with $0.13 < q < 0.35$ Å^{-1} show very similar behaviors.

At the intermediate angle range, lens tissues at all heat treatments appear to be very similar to each other. The scattering intensity decreases slowly at $0.05 < q < 0.1$ Å^{-1}, and then decreases sharply from $q = 0.1$ Å^{-1}, reaching a minimum at 0.16 Å^{-1}. These features are consistent with the scattering from spherical objects. I used DAMMIN to construct spatial configuration models (method in Chapter 4) from the SAXS intensity profile at 90 °C. One likely spatial configuration is composed of spheres with diameter about 50 Å. These spheres are consistent with the S-crystallins in the squid lens. So even at 90 °C, where the lens tissue is completely opaque and white, the spherical structure of S-crystallin is still preserved.

A peak at $q = 0.21$ Å^{-1} is present in all these three heat treatments. This peak cor-

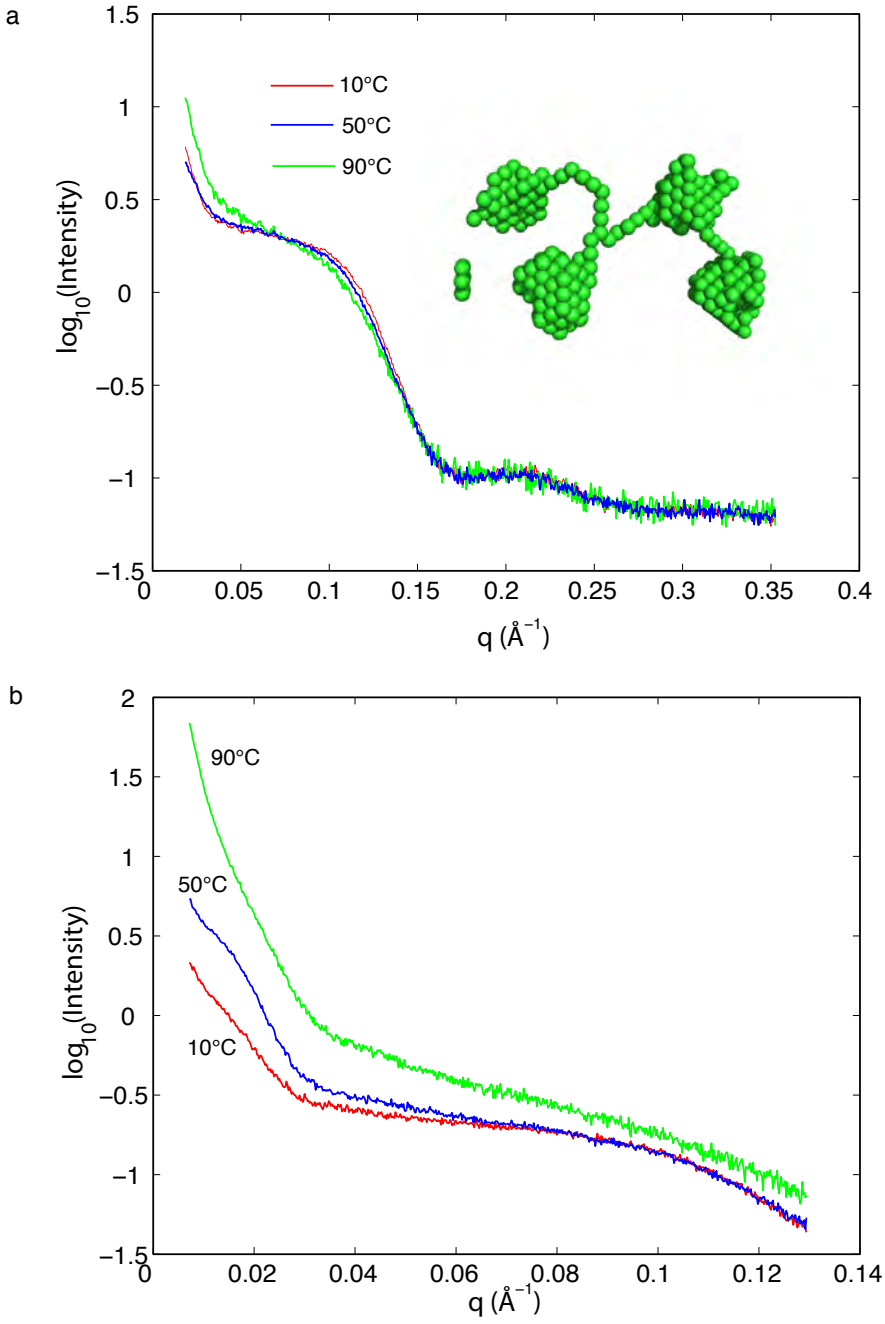


Figure 3.6: Comparison of fresh lens to heated lens. The peripheral layer of the lens was treated by heating at 50 °C and 90 °C. The SAXS measurements were taken to compare the structures under different heat treatment. The intermediate angle range (a) shows very similar x-ray scattering intensities. The green image shows the DAMMIN output of the 90 °C configuration. At the small angle range (b), heated samples have higher scattering intensity, indicating that the protein network has higher density fluctuation.

responds to the dimeric structure of the S-crystallin. There is no difference in SAXS intensities between different heats at this peak. This suggests that the S-crystallin dimeric structure does not change at all of these temperatures. The inner structure of an S-crystallin dimer remains unchanged as well.

Although surprisingly, at small and intermediate angle ranges, heat treatment to 90^{o} does not significantly change the sample, in the small angle range where $q < 0.13$ Å^{-1}, SAXS intensities differ between different heat-treated samples. Although all of the scattering intensities show a sharp decrease at $q < 0.03$ Å^{-1} and a less-sharp decrease at $0.03 < q < 0.13$ Å^{-1}, the slopes of these sharp decreases vary in a log-log plot. The fresh eye lens at 10 °C shows a slope with the absolute value less than 2, while it is 3 for the 90 °C sample. This slope is associated with the larger-scale structure of the lens system. It shows that the 90 °C sample has a fractal dimension of 3, which is higher than that in a random walk model, and it is consistent with that of a 3-dimensional bulk. The fractal dimension of the 50 °C sample is in between the 90 °C sample and fresh lens. There is also a shoulder or peak at $q \sim 0.02$ Å^{-1} in the 50 °C sample. This peak indicates that there are spatial fluctuations in the protein network. The fresh lens sample has a less-pronounced peak compared to the 50 °C sample. The fractal dimension of the fresh lens is less than that of a random walk model. Thus, our SAXS at $q < 0.03$ Å^{-1} shows that at a larger-scale, the lens network differs between the heated and fresh lens tissue: They have different fractal dimensions at all temperatures tested, so the whole structure of the network is different in each case, while the secondary and tertiary structure of S-crystallin proteins is apparently maintained. At 90 °C, there is a large density fluctuation at large spatial scale, and this fluctuation apparently results in the opacity of the heat-treated lens.

These intensities at different heat treatments are also different in the range of $0.03 < q < 0.13$ Å^{-1}. This range is correlated with the dimer-dimer pair-wise interaction distance. The 90 °C scattering is higher in this range compared to the other two, indicating that

the 90 °C sample has a higher protein density fluctuation here. This may be due to the heat treatment breaking the bonds of the dimer-dimer interaction present in the fresh lens. This shift at a temperature of 50 - 90 °C is roughly consistent with interaction dimer-dimer interaction consisting of hydrogen bonds, which are likely to break between 50 - 90 °C. This temperature range is lower than the bonding energy, although because the lens is hydrated, one has to consider the entropy of the system as well.

In short, SAXS on the heated samples show that the major features of the secondary and tertiary protein structure do not change at high temperature. However, heat treatment does change the interaction between protein dimers, and alters the network at a larger scale. This change in the network increases the spatial density fluctuation, and is associated with opacity of the sample. This change appears in the range of 50 - 90 °C, suggesting that the protein-protein interaction is consistent with hydrogen bonds between S-crystallin dimers.

3.8.2 Heated samples at all lens layers

Figure 3.7 shows the 90 °C heat-treated samples compared to untreated samples at all lens layers. The features of the heated samples at different lens positions are similar to those observed in the periphery of the lens. At the range of $q < 0.04$ Å^{-1}, the heat-treated lens scattering intensity decreased much faster with increasing q compared to the intensity from a fresh lens in the region of the outer three layers. The steep slope indicates that the protein network at a large spatial scale has been changed to be more inhomogeneous after heat treatment. The peak at $q = 0.02$ Å^{-1} from the uncooked lens at 80%, 60% and 40% layers disappears after heating. At the core of the lens, the heated and unheated measurements are very similar, except that the peak at 0.02 Å^{-1} disappears in the heated sample. The absence of this peak suggests that the homogeneous protein network is broken and the inhomogeneity is so large that the proteins are randomly aggregated.

At the intermediate angle range, the heated and unheated lenses appear to be similar,

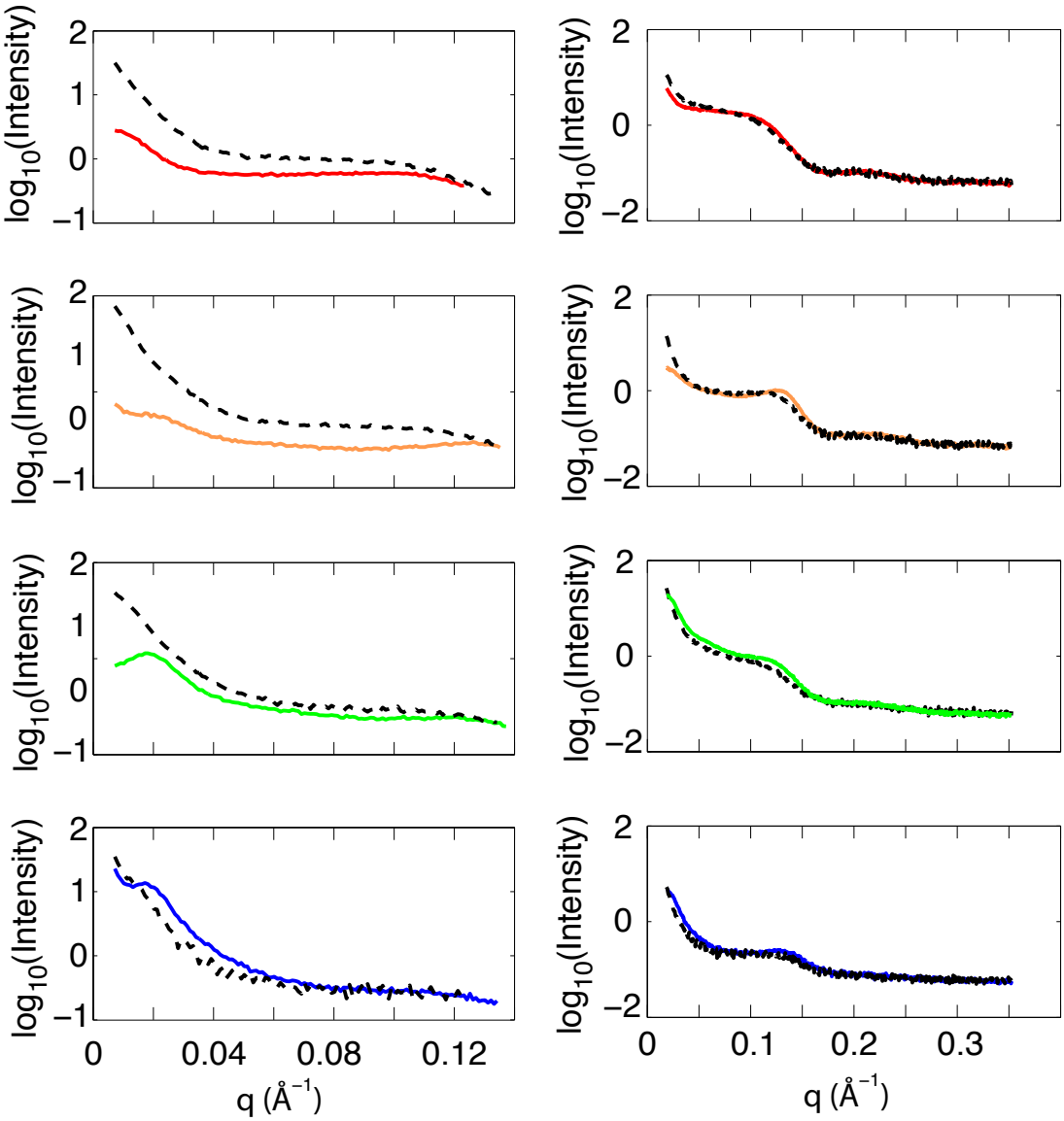


Figure 3.7: Fresh and heated lens sample at all layers. The x-ray scattering on both fresh squid lenses and heated lenses at different q range are shown on the left column and right column. The solid lines are data from fresh lenses, and the dotted black lines are data from lenses heated at 90 °C. Red labels 100% layer, orange 80%, green 60%, and blue 40%.

except that the shoulder or peak at $q = 0.15$ Å^{-1} is less apparent in the cooked lens. It suggests that for all layers in the lens, the change in pair-wise interactions results in a total change of the network configuration observed in the small angle range discussed in the previous paragraphs.

In short, the lens tissue from different radial positions behaves very similarly with heat treatment. The protein dimeric structures are not changed at all parts of the lens, but we observed large changes in the interactions between the S-crystallin dimers and in the associated protein network

3.9 Conclusion

In this chapter, I studied the lens protein assembly by performing SAXS on both intact and homogenized lens tissues in buffers of different ionic strength. The intact lens SAXS scattering intensity shows that there is a large scale structure, which is assembled by the attractive interactions between the S-crystallin proteins. I also discuss this large scale structure in comparison with a gelled phase. This large scale structure is very likely to be gelled because: there are attractive interactions between the proteins; it is very likely that the network is percolated; the density of the protein network is similar or higher to the density of a liquid from the periphery of the lens to the core. Therefore, it is very likely that most of the lens is gelled, and then transitions into a jammed solid at very small lens radii. This gelled network appears to have a structure of bifurcating chains of S-crystallin dimers. The nodes associated with these bifurcations may produce a shoulder-like structure in the intensity at $q \sim 0.02$ Å^{-1}. The large scale network appears in the buffer-homogenized sample as well, indicating that this network is robust to being dissolved in aqueous solution, or readily reassembles after disassembly.

Finally, we compared the SAXS intensities from a heat-denatured lens tissue to fresh

lens tissue. The result suggests that a change in the nature of the protein pair-wise interactions leads to a totally different network configuration. The cooked lens appears to have still-folded proteins now in random, rather the chain-like aggregations, resulting in big protein density fluctuations, and that is why the cooked lens is opaque.

Chapter 4

Estimating three-dimensional spatial configurations of lens tissue from SAXS data

In the previous chapter, I described how, based on a reciprocal space analysis, the organization of proteins in squid lens is much more consistent with a protein network rather than soluble protein in cytoplasm as previously thought [114]. S-crystallins in the squid lens are very likely to form a gelled network at large scale in the periphery of the lens, which gradually shifts into a possible jammed state towards the core of the lens. This network can be viewed as a bifurcating chain-like structure in the peripheral layer. SAXS evidence suggests that the network is built up through pairwise dimer-dimer interactions of S-crystallin proteins. But SAXS measurements, which show reciprocal space, do not provide us with a clear picture of the possible molecular mechanisms of protein-protein linkages within the larger structure. In this chapter, I am going to use the SAXS data from the previous chapter to study the squid lens system in real space. I will build up a three-dimensional protein configuration which is a good fit to the SAXS data, and study the protein-protein interac-

tions implied in these configurations. Then I will propose a mechanism to explain how the lens generates a transparent material at all possible packing fractions using patchy particle theory.

4.1 Radial distribution function (RDF) of squid lens proteins

The first thing I did to convert the SAXS intensity to real space is to obtain the radial distribution function of the S-crystallins. The radial distribution function describes the probability of finding a particle at the distance r. The scattering intensity from SAXS is the Fourier transform of the radial distribution function of all the component particles. Therefore, I performed an indirect Fourier transform of the SAXS intensity to get back to the radial distribution function of the proteins in the lens. This is achieved using a software GNOM [112] in a series of data analysis software ATSAS.

4.1.1 Indirect Fourier transform of SAXS intensity to get RDF

GNOM calculates the radial distribution function from scattering intensity as a function of the wave vector q using an indirect method with regularizations [112]. The input of this program is the x-ray scattering intensity $I = I(q)$. The relation between the scattering curve and the radial distribution function in real space $g = g(r)$ is

$$I(q) = \int_{D_{min}}^{D_{max}} K(q,r)g(r)dr, \tag{4.1}$$

where $K(q,r)$ or K is an kernel of integral transform which is performed in the interval $D_{min} < r < D_{max}$. The disadvantage in using the equation only is that small errors in $I(q)$ will result in large errors in $g(r)$. Therefore, regularization methods have been employed

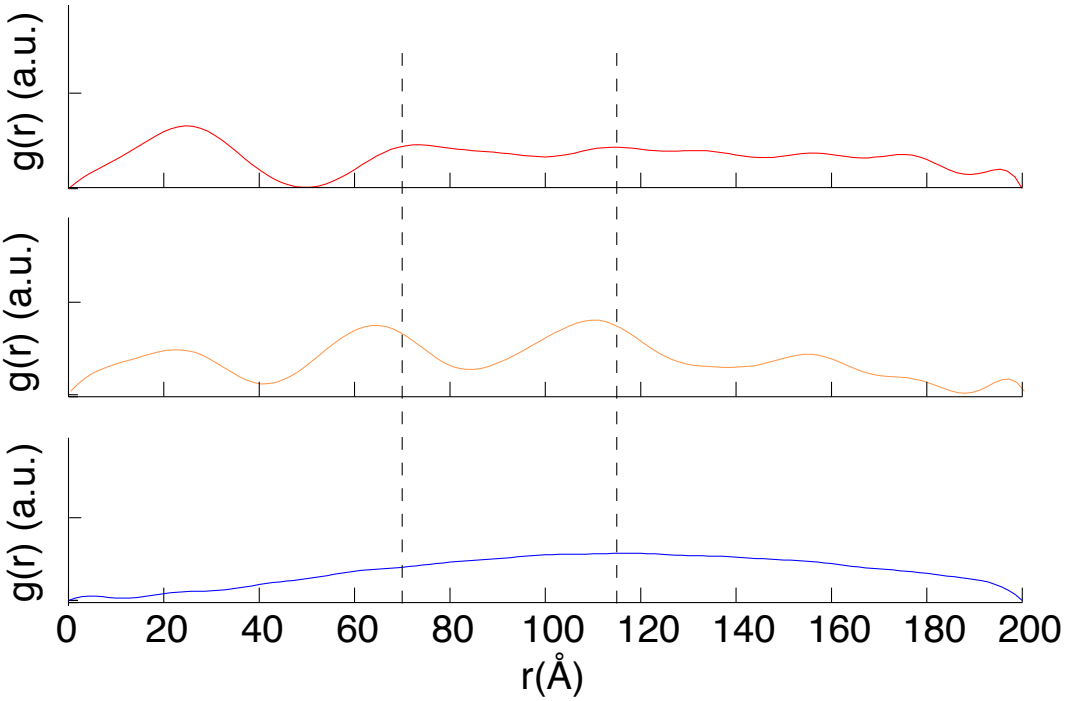


Figure 4.1: Radial distribution function ($g(r)$) for different layers of the lens. The 100% layer (red) shows the dimeric structure peak at $r \sim 20$ Å and the nearest neighbor distance at ~ 70 Å. The 80% layer RDF (orange) shows similar trend as the 100% layer. These features are not preserved in the 40% layer (blue). Vertical lines label the first and second nearest-neighbor distance.

to solve this issue. GNOM outputs a result which minimizes the functional

$$T_\alpha[g] = ||I - Kg||^2 + \alpha\Omega[g]. \tag{4.2}$$

Here, $||I - Kp||$ denotes the norm in reciprocal space, with I as the measured intensity. α is the weight, and Ω is a stabilizer. These regularizations are to correct for oscillations, systematic deviations, discrepancy, stability, positivity, and validity in the center of the system [112]. For the case of SAXS on the lens, we used $r_{max} = 200$ Å, which generally outputs around 7 globular proteins.

4.1.2 Calculation of the radial distribution function

The radial distribution function (RDF) shows the probability of finding a particle at the distance r from one particle and could therefore reveal the properties of S-crystallin proteins packing in space. We also studied RDF using the software GNOM. I used program parameters of zero condition at $r = r_{min}$ and non-zero condition at $r = r_{max}$, with $r_{max} = 200$ Å. The RDF of the lens proteins calculated from $I(q)$ at different radial positions of the lens are shown in Figure 4.1.

In the peripheral layer, there are two major peaks in the RDF. There is one peak located at about 25 Å, and another peak at 70 Å. There is also a weak peak at 120 Å that is less pronounced than the other two. The distance between two S-crystallin monomers within the folded dimer is approximately 25 Å. So the 25 Å peak is likely to correspond to the inner structure of the S-crystallin dimer. This peak is also associated with the $q = 0.21$ Å^{-1} peak in the form factor, and is consistent with our homology models.

The peaks at $r > 70$ Å in the peripheral layer correspond to the spatial distribution of S-crystallin dimers. These peaks are associated with the features in the structure factor that correspond to the nearest-neighbor distance of S-crystallins. The major peak at 70 Å indicates that in the periphery of the lens, the average nearest neighbor distance of the protein is 70 Å. Therefore, the picture one can build up from studying the RDF in the peripheral layer is that each S-crystallin monomer forms a dimer with another S-crystallin monomer, then these dimers link to other dimers with an average center-center distance of ~ 70 Å. Our homology modeling on S-crystallin shows that the S-crystallin forms a dimeric structure with diameter 50 Å; meanwhile, the SAXS results suggest that there are attractive interactions between the proteins. How, then, can proteins with diameter of 50 Å attract each other and yet form a structure dominated by 70 Å center-center distances? One possible explanation is that the as-yet-unexplained unstructured regions, or "loops" in the proteins play a role in the linkage between the proteins. If two loops from 26.5 kDa

S-crystallin can be linked up together, the protein-protein distance will be able to extend to 70 Å. Therefore, a loop-loop linkage mechanism for protein-protein interaction building the network is at least consistent with the real-space data. This hypothesis will be discussed in the next section.

In the 80% layer, this protein-protein distance decreases to ~ 60 Å. This decrease of the nearest neighbor distance indicates that the S-crystallins are packed more closely compared to the peripheral layer, and this is consistent with the fact that the packing fraction in the 80% layer is higher than in the periphery of the lens.

At the core of the lens, the RDF is very different from the peripheral layers. The peak at 25 Å disappears and there are no sharp peaks from this tissue layer. This result is consistent with our hypothesis that the protein density at the core of the lens is so high that the proteins are squeezed, and the network of the proteins at the core has little free space, so there are no significant features captured by the RDF.

In short, what we learned from the RDF calculations supported our analysis in the SAXS study that the proteins are linked together through a pair-wise interaction. The proteins in the peripheral layers are separated on average by 70 Å, which is larger than the diameter of the core of the folded protein dimer; these observations suggest that the loops may play a role in the protein-protein linkage.

4.2 Three-dimensional configurations

Now that we have the radial distribution function in real space and the SAXS intensity in reciprocal space, is it possible to construct a protein configuration in three-dimensional real space? This configuration could help us better understand whether the proteins are linked by loops, and if so, what the structure of the interacting dimers look like. In this section, I

used a Monte Carlo method to achieve this at different radial positions in the lens with the program DAMMIF [49].

4.2.1 Construction of the three-dimensional protein configuration

We used the algorithms coded in DAMMIN and DAMMIF structure prediction software to study the structure of protein networks in the lens. DAMMIN returns one likely three-dimensional protein spatial distribution by the program via a Monte Carlo fitting algorithm with the input of the GNOM output only [113]. Specifically, DAMMIN implements 'dummy atoms' to simulate real structures in an effort to generate a real structure that is consistent with scattering data. In the case of this work, "dummy atoms" are consistent in real size with "dummy amino acids". DAMMIF also uses a Monte Carlo algorithm to fit the SAXS intensity and output one possible configuration in three-dimension real space, but implements a new algorithm to speed up the estimate.

In DAMMIN, each dummy atom has two phases associated with it: the protein phase and the solvent phase. The scattering intensity is only calculated from the protein phase [113]. The algorithm then minimizes the function

$$f(x) = \chi^2 + \sum_k \alpha_k P_k(x) \ , \tag{4.3}$$

where α_k is the weight of the k$^{\text{th}}$ penalty, $P_k(x)$ is the penalty, and χ^2 is the discrepancy between the experimental and calculated data: Given a set of $M \geq 1$ curves from experiment $I^{(i)}_{exp}(q)$, $i = 1,...,M$ [49],

$$\chi^2 = \frac{1}{M} \sum_{i=1}^{M} \sum_{j=1}^{N(i)} \left[\left(I^i_{exp}(q_j) - I^i_{cal}(q_j) \right) / \sigma(q_j) \right]^2 , \tag{4.4}$$

where $N(i)$ is the number of points in the i$^{\text{th}}$ curve, and $\sigma(q_i)$ denotes the experimental

error. The calculated intensity $I(q)$ is expressed as [113]

$$I(q) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^{l} |A_{lm}(q)|^2 \,, \tag{4.5}$$

$$A_{lm}(q) = i^l (2/\pi)^{1/2} v_a \sum_{j=1, X(j)=1}^{M} j_l(qr_j) Y_{lm}^*(\omega_j). \tag{4.6}$$

In these equations, $A_{lm}(q)$ is the partial scattering amplitudes, (r_j, ω_j) are the polar coordinates of the positions of the dummy atoms, v_a is the displaced volume per dummy atom, $Y_{lm}(\omega_j)$ is the corresponding spherical harmonics and $j_l(qr_j)$ denotes the spherical Bessel functions. The shape of this dummy atom model is described by a binary configuration vector X, with each dummy atom either belongs to the particle phase with index = 1, or belongs to the solvent phase with index = 0. In each Monte Carlo step, DAMMIN will randomly choose an atom and flip its phase. Whether the change will be accepted or rejected is determined by simulated annealing, with the probability of accepting the change $\exp(-\Delta/T)$ [113]. The temperature is set to be high in the beginning of the simulation, where the changes in the dummy atom phases are almost random. Then temperature is lowered and the configuration of dummy atoms in protein phase is selected. In this method, the global minimum of the possible configuration is most likely to be selected.

DAMMIF is a similar but faster program to DAMMIN; it runs 25-40 times faster under equivalent conditions [49]. The difference in algorithm and result between these two programs are as follows. In DAMMIN, there is a penalty on producing an unconnected shape, but it is not necessary that the output shape has to be connected. In contrast, DAMMIF always outputs a connected configuration for fitting the SAXS curve, since an assumption of the algorithm is that the real space structure is continuous. Therefore, computational speed in DAMMIF is increased mostly by this assumption of connectivity because DAMMIF tests the connectivity first and denies the disconnected configuration before moving on to

the processor-consuming process of updating the scattering intensity. Other major differences involve unlimited search volume for DAMMIF and the difference in penalty settings. For DAMMIF, we used a simulation sphere diameter of 200 Å and particle diameter of 4.2 Å, consistent with a single particle representing a single amino acid residue; no symmetry was presumed.

Note that the resulting protein distribution from either DAMMIN or DAMMIF may not necessarily mirror the exact structure of proteins in the lens system. This is because SAXS detects the density fluctuation of the system, not the absolute density. The intensity as a function of q can be written as

$$I(q) = Nf(q)^2\Big[1 + \frac{N}{V}\int_0^\infty 4\pi r^2\big(P(r)-1\big)\frac{\sin(qr)}{qr}dr\Big], \tag{4.7}$$

where $P(r)$ is the distribution function which approaches to 1 at large r. The first term $Nf(q)^2$ is the form factor term. For the second term, we have

$$\big(P(r)-1\big) \sim \frac{\Delta\rho}{\rho_{ave}}, \tag{4.8}$$

where $\Delta\rho$ is the density fluctuation. If the proteins are homogeneously distributed in real space at the scale that we measure, these proteins cannot be detected by SAXS or by DAMMIF because their density fluctuation is low.

4.2.2 Spatial configuration of S-crystallins

As we have seen in the previous chapters, the proteins in the lens form a connected network which is presumably gelled. Since both native PAGE and a filtration experiment suggest that lens proteins are bound to each other, I used DAMMIF to estimate the spatial structure in the lens from SAXS data. I used a simulation sphere diameter of 200 Å and particle

diameter 4.2 Å, consistent with a single particle representing a single amino acid residue; no symmetry was presumed.

For the outermost layer of the lens, DAMMIF's closest fits of a real space structure show 35 - 50 Å spheres connected by bridges. These connected bridges are $\sim$ 20 Å in width, $\sim$ 40 Å long and the angle between the bridges ranges from 60° to 120°. These bridges are geometrically consistent with the characteristics of the unstructured loops protruding from the otherwise folded protein observed in our sequencing data and homology models. Therefore, each sphere observed in DAMMIF is consistent with the folded body of S-crystallin, and the two lines connecting spheres are consistent with the loops (Figure 4.2). Most of the 50 Å spheres in the DAMMIF prediction show symmetry consistent with dimers formed from structurally similar monomers (i.e., homodimers, as described in Chapter 2).

Because DAMMIF is a Monte Carlo algorithm and the solution is not unique, it was possible to get the equally-likely protein configuration geometries predicted by the software by repeating it several times. The dimeric proteins observed in the periphery of the lens from these repeats are connected to a minimum of one and a maximum of four other dimers. I describe the number of other 5 nm spheres connected to a single sphere as the coordinate number, M, following Sciortino [12]. The result from the peripheral layer shows that most of the proteins are $M = 2$ particles, with an occasional particle of $M = 3$, as is shown in Figure 4.3.

Note that DAMMIN outputs similar result to DAMMIF, as is shown in Figure 4.3b. DAMMIN does not necessarily output a connected configuration, but given data from the lens, the output of DAMMIN configuration is also connected, providing more evidence that the proteins in the lens form a gel.

The results of DAMMIF fitting for the 80% layer configuration is very similar to the 100% layer of the lens (Figure 4.2). For this layer, there are also spherical proteins which

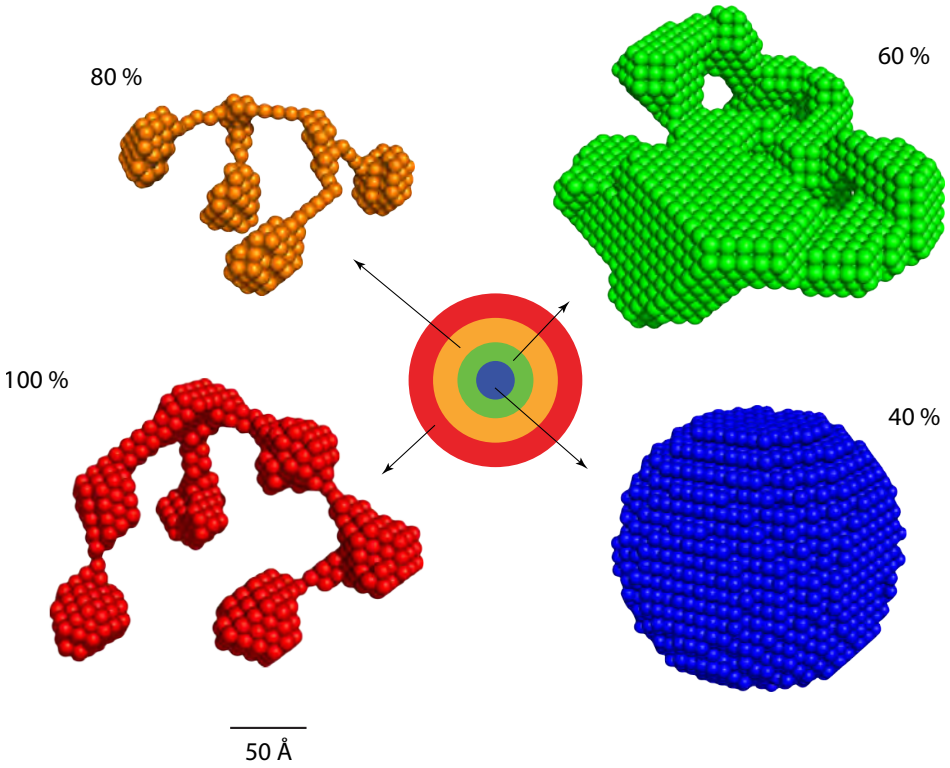


Figure 4.2: Output from DAMMIF at different radial positions in the lens. The patterns composed from spheres are the output from DAMMF at different radial positions. The pattern from 100% layer (red) and 80% layer (orange) shows that 5-nm proteins are connected to each other. The 60% layer (green) protein configuration shows sheet-like structure. In the 40% layer (blue) of the lens, there is very little empty space in the configuration.

are connected by lines. The spheres in the 80% layer are ~ 40 Å in diameter, and the connecting lines are longer compared to the 100% layer.

In the 60% layer, DAMMIF predicts a lamellar configuration of the proteins. The thickness of these sheets is around 50 Å, which is consistent with the size of the S-crystallin, suggesting that at this radial position, S-crystallin proteins form ramifying sheets in the cells.

In the core of the lens, the protein packing fraction is so high that DAMMIF predicts that space is nearly entirely filled with protein, with little or no bulk water. This is consistent with our SAXS result that the peak at $q = 0.21$ Å^{-1} disappears, as is explained in Chapter 3. This peak corresponds to the monomer-monomer interface of an S-crystallin dimer (see

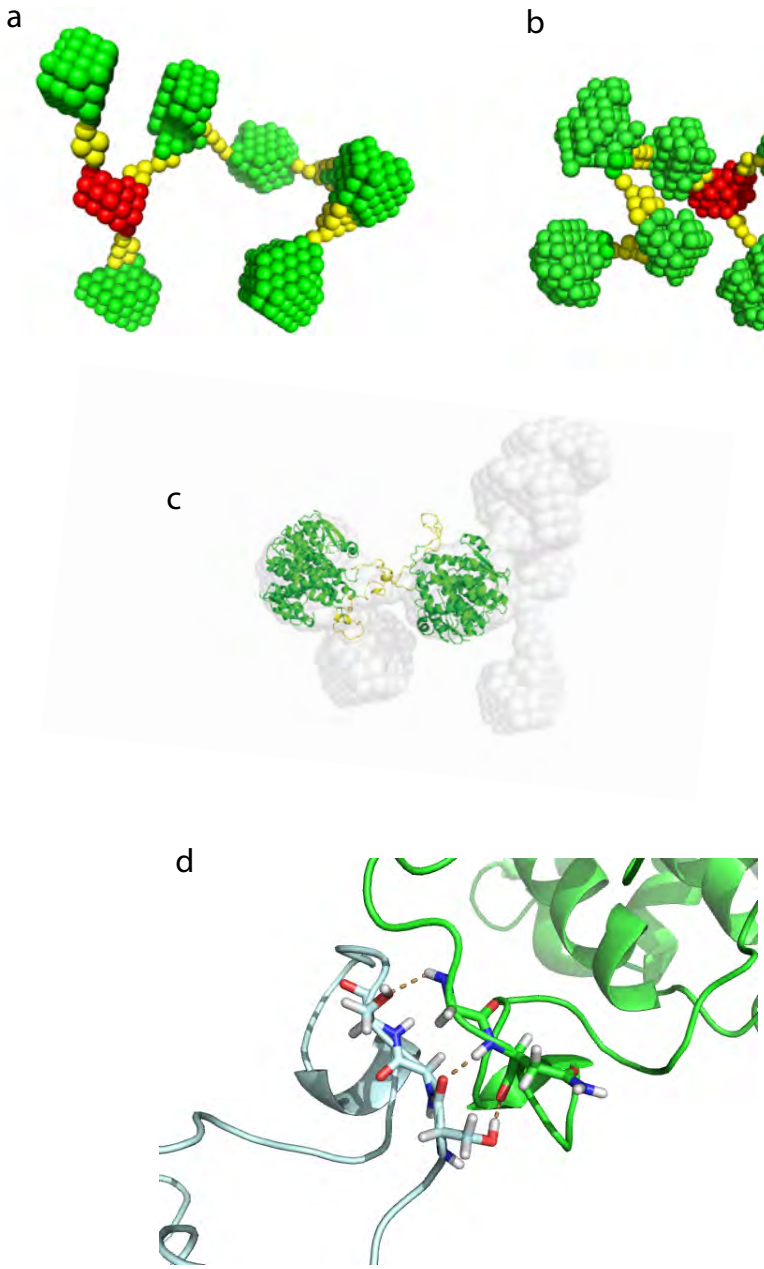


Figure 4.3: DAMMIF output and the loop-loop interaction. The top left figure (a) shows the output from DAMMIF of the 100% layer of a squid lens. Yellow labels the patches; green labels $M = 2$ particles; red $M = 3$ particles. This is compared to the output from DAMMIN (b). DAMMIN outputs a very similar configuration as DAMMIF. The comparison between DAMMIF output to the protein structure from homology modeling is shown in (c). It suggests that the proteins are linked via their loops. The bottom figure (d) shows that GROMACS results in loop-loop interaction via hydrogen bonds.

form factor). The absence of the peak suggests that the proteins are so tightly packed in the core that this empty space is no longer present, as shown by the DAMMIF result.

4.2.3 Loop-loop interaction

The previous section shows that in the peripheral two layers of a squid lens, the proteins may be linked by their loops. In this section, I collaborated with my labmate Tom Dodson to study loop-loop interactions using the molecular dynamics package GROMACS. We manually docked two proteins with their center-center distance at 70 Å, with the loops oriented at the same angles relative to the centers of mass as observed in SAXS. The result from GROMACS suggests a possible mechanism for the interaction between two long-loop proteins. In our molecular dynamics simulations, a few hydrogen bonds form between two long-loops, as is shown in Figure 4.3c. Over the course of the 10 ns simulation, the number of the hydrogen bonds varies from 3 to 4 between two long-loop lops8 S-crystallins. Interestingly, although there are many amino acids that participate in hydrogen bonding over 10 ns, there are never fewer than 3 bonds formed at any given time step. These simulations suggest that two long-loop dimers can interact through hydrogen bonds between their loop segments. We interpret these data to mean that the loops of the "long-loop" proteins (approximately 26.5 kDa) can be treated as patches with attractive interaction.

4.2.4 The linkage via loops is consistent with our RNA sequencing result

In this section, I will compare what we learned from our RNA sequencing result to the S-crystallin real space configuration. RNA sequencing allows us to infer the amino acid sequences of the loops in S-crystallin. From this, we can compute the histogram of loop lengths in number of amino acids. The RDF provides us with the histogram of protein-

protein distances in the tissue in angstroms. So here, I am going to compare the distribution of the amino acid number in the loops to the distribution of dimer-dimer distances as inferred from the RDF.

To estimate the relation between the physical loop length in Angstroms versus the number of amino acids in the loop, we fit loop length to the number of amino acids, and estimated the ratio between amino acid number and physical length using a linear relation. The physical loop length was initially estimated from the loop configuration observed in our homology models. To convert from number of amino acids in the loop to the physical length in angstroms, we found the following relationship: loop length $=$ $0.76 \times$ amino acid number $+ 10.45$.

We then compared the physical length of the loops in our homology models to the dimer-dimer distance measured in our SAXS data, as seen in the RDF. The distance between the edges of the two dimers in the homology model is about 32 Å, and the non-overlapping length of each single loop length 23 Å. So an estimated ratio between one loop length and the distance between the folded edges of two dimers is 1.4. This ratio is used to estimate the distance between two S-crystallin dimers in the periphery of the lens for all of the isoforms. In the RDF from SAXS in Figure 4.4, which can be viewed as the histogram of dimer-dimer separation distances, the majority of S-crystallin distances are at about 80 Å. This peak overlaps with the second peak of $g(r)$, which corresponds to the nearest neighbor distance between the proteins. The overlap makes a reasonably good correspondence between the physical interaction observed in a molecular dynamics model and the RDF calculated from SAXS data. At larger protein interaction distances, longer loops separate S-crystallin dimers, corresponding to more peaks at $g(r)$.

Therefore, the good correspondence between the loop length histogram from our inferred protein sequence to the RDF obtained from SAXS is further evidence showing that the proteins are linking up to each other through their loops. The dimer-dimer distance can

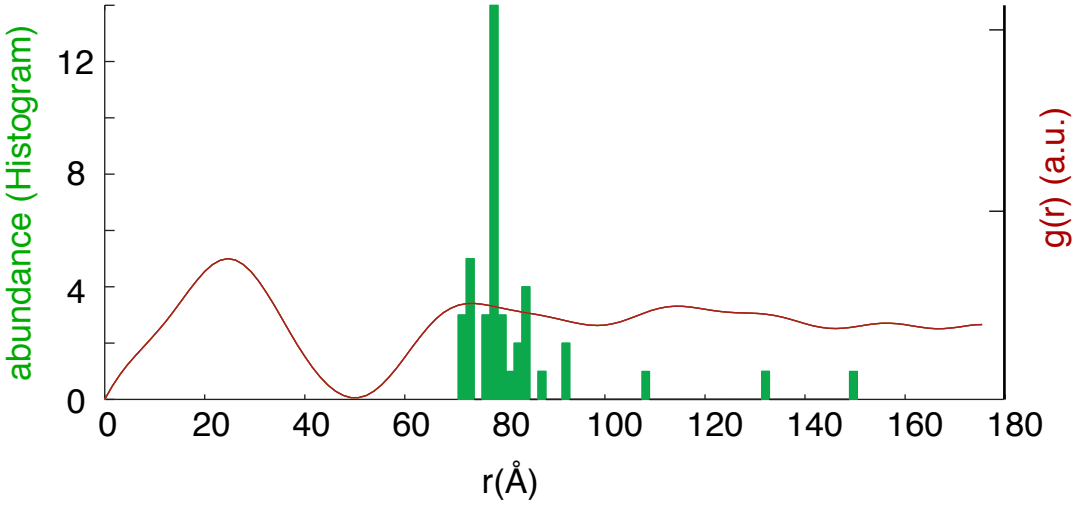


Figure 4.4: Comparison between the RDF and the number of unique loops in RNA sequencing. The red curve shows the RDF from the 100% layer, which is calculated from our SAXS data. The green bars shows the histogram of the estimated protein-protein distance from our RNA sequencing result, assuming the loop-loop interaction. The consistency of these data indicate that the protein loop-loop interaction links S-crytallins together.

be very well approximated from the amino acid sequence by assuming that the proteins are linked via their loops. This estimate matches very well to our RDF data. This attractive loop-loop interaction is highly anisotropic, and can be treated as patches on the S-crystallin surface.

4.3 S-crystallins as patchy particles — mapping to the patchy particle phase diagram

In Chapter 1, I described generic patchy particle systems containing a bag of particles attracting each other via several geometrically constrained spots on the surface. Globular proteins in general can often be treated as patchy particles because their interaction is usually attractive as a result of the non-homogeneous distribution of amino acid groups on

their surfaces. S-crystallins in the squid lens are no exception of this. Our data are consistent with the idea that S-crystallins attract and bind to each other via disordered loops protruding from the surface of the folded proteins, serving as the "patches". In chapter 1, I introduced the phase diagram of a patchy particle system. The spinodal line in the phase diagram separates the stable and possibly gelled phase from a liquid-liquid phase separated state at a lower packing fraction. Liquid-liquid phase separation is bad for the lens because the possible droplet formed in another liquid state contains the boundaries which scatters light [79]. The details have been described in Chapter 1. In this section, I am going to map the lens system to the phase diagram and study the S-crystallins as patchy particles.

Proteins expressed at different radial positions of the lens can be mapped to the phase diagram as is shown in Figure 4.5. From the periphery to the core of the lens, the protein packing fraction gradually increases. The strength of interaction between the proteins in the peripheral two layers are estimated using the energy of hydrogen bonds as observed in our molecular dynamics simulations and consistent with our SAXS data ($\sim$ 60 kJ/mol for 3 hydrogen bonds). The strength of the interactions in the core two layers are harder to estimate since the details of the protein-protein interaction are not clear, and I therefore plotted them using a bigger range, spanning the possible binding energies of proteins generally. The spinodal lines of $M < 3$ pass through the area in the phase diagram bounded by the outer two layers, whereas in the core two layers, the average coordinate number is not crucial because even a high $< M >$ system will have a spinodal line located at a packing fraction much less than the packing fraction of the lens system. Therefore, the lens system is particularly sensitive to average coordinate number in the peripheral regions of the lens to avoid liquid-liquid phase separation, and not particularly sensitive to coordinate number in the center of the lens.

Therefore, the mapping of the radial positions of the lens to the phase diagram of patchy particles requires proteins with low coordinate numbers in the peripheral two layers. The

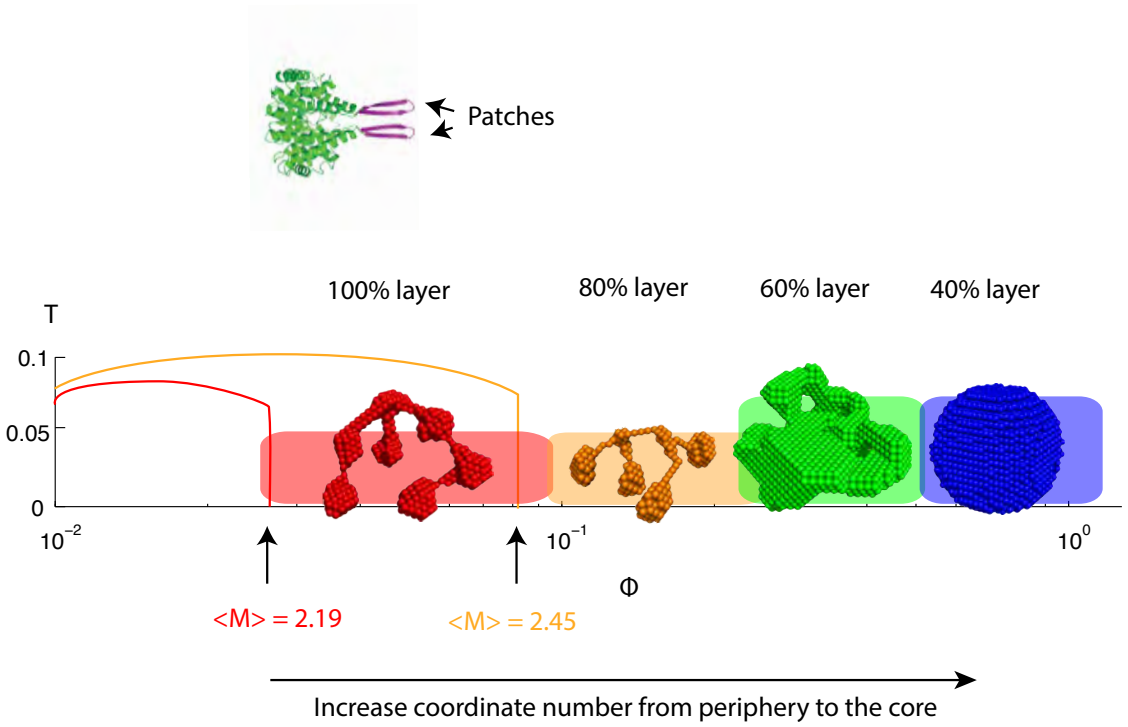


Figure 4.5: Phase diagram of the S-crystallin in the lens. The map of the lens system at different radial positions relative to the patchy particle phase diagram is shown. The colored rectangles are the regions that the proteins at different radial positions occupy in the lens, with red 100% layer, orange 80%, green 60%, and blue 40%. The estimated $< M >$ for 100% layer is 2.19, and 80% layer is 2.45. These estimations map well to the phase diagram. The images are the output from DAMMIF.

purpose of this section is to examine the estimated coordinate number of these two layers and compare to the phase diagram.

4.3.1 "Bootstrapping" coordinate number from DAMMIF result

Because DAMMIF is a Monte Carlo algorithm and the solution to each new iteration of the fitting process is not unique, it was possible to estimate the average configuration of many equally-likely protein configuration geometries predicted by the software. This should bear some similarity to the average configuration of fragments of the real protein network. To estimate the ensemble of configurations in the bulk network, we ran DAMMIF 100 times with the same one-dimensional SAXS measurement input, generating a different predicted network configuration each time. We then manually counted the coordination numbers at each non-edge protein within the predicted network, and calculated the average network coordination number. The result of this "bootstrapping" approach to DAMMIF showed an average coordinate number $< M >$ of 2.19 for the 100% layer, and 2.45 for the 80% layer.

To test whether the differences predicted in DAMMIF in network coordination number were significantly different between different layers in a single lens, we used a permutation test. This test assumes that all observations come from the same distribution, and then asks if the two input distributions can be recapitulated by resampling half of the input observations. This test showed a standard deviation of the averaged coordinate number in the peripheral layer of 0.0377. Therefore, the difference in $< M >$ between the outer two layers is 6.9 times higher than the standard deviation of a single layer, and the difference in $< M >$ fit by DAMMIF is significant.

The results from repeating DAMMIF suggest that the 100% layer S-crystallins are patchy particles with a coordinate number of 2.19, which is 0.26 less than the same result for the 80% layer. These coordinate numbers were then mapped to the phase diagram (Figure 4.5). The protein packing fraction of the 100% layer ranges from 5% to 10%,

which is higher than where the spinodal line for $M = 2.19$ is located. Similarly, the calculated $< M >$ of 80% layer corresponds to a spinodal line at the packing fraction $\sim 6\%$, which is also smaller to the packing fraction of the proteins in this layer. Therefore, the estimated $< M >$ of the peripheral two layers maps very well to the phase diagram of patchy particles.

The DAMMIF result for the 60% layer shows ramifying sheet-like structures. This structure is also potentially consistent with an assembly mechanism based in patchy particle theory. Simulations on the assembly of patchy particle systems have been performed by other researchers. Zhang and colleagues [131] found that when the patches cover a large proportion of the surface, the patchy particle assembly became sheet-like, rather than chain-like. This sheet-like structure looks very similar to the structure of our DAMMIF output from the 60% layer in the lens. Therefore, our hypothesis for the lens proteins in the 60% layer is that patches are either very large or numerous so that they cover most of the surface of the dimeric protein. These large patches stick together, forming a sheet-like structure.

The DAMMIF output for the core (40% layer) shows almost no free space, consistent with nearly solidly packed proteins. Distinct particles were not detectable, and each particle apparently has many nearest contact neighbors; the overall structure is predicted to be continuous (Figure 4.2).

Therefore, DAMMIF predicts a trend that the average coordinate number of the S-crystallins increases from the periphery to the core of the lens. This result maps very well to the theoretical predicted phase diagram of patchy particles: for each layer in the lens, as defined by ϕ, there are spinodal lines in the phase diagram located at packing fractions which are lower than the packing fraction of the proteins in that layer. This indicates that the lens can be in a gelled state at any radial position and corresponding concentration in the lens, with the packing fraction spanning from 5% to 100%.

There is further evidence showing that the $< M >$ increases from the periphery to the

core of the lens, as will be discussed in the following section.

4.3.2 The peak at $q = 0.13$ Å^{-1} in SAXS intensity is consistent with the increase of $< M >$

In our SAXS result, the protein-protein pair-wise interaction peak at $q = 0.13$ Å^{-1} moves towards a higher q from periphery to the core of the lens. This trend is consistent with the increase of $< M >$.

Particle systems with different ratios of $M = 2$ and $M = 3$ particles at a constant packing fraction were studied by Sciortino's group [98]. Their result is shown in Figure 4.6a, assuming the particle diameter to be 50 Å. For all of their simulations, they found that the system can be gelled even at low temperature with the resulting state of a fully connected network. In the network, the $M = 3$ particles provide a branching point, and $M = 2$ particles increase the length of the filament. Then the structure factor of every configuration was also calculated, as is shown in Figure 4.6b. The structure factors for different $M = 2$ and $M = 3$ ratios show similar structures: a sharp decrease at low q, and a peak at the particle pair-wise interaction range. However, this peak associated with the particle-particle interaction shifted toward a higher q with increasing average coordinate number in the simulation.

The structure factors of the patchy particle systems in this study are also very similar to the structure factor of the lens. There is a sharp decrease at small q, and a peak at $q = 0.13$ Å^{-1} indicating the pair-wise interaction between two proteins. In the simulation, this peak moves towards a higher q from the periphery to the core of the lens, similar to our SAXS results on the lens in Chapter 3. Figure 4.6 shows a comparison of the structure factors from our experiment and from the Sciortino group's simulations. The similarity between the structure factors from Sciortino *et al*'s patchy particle simulations and the structure factors of the different layers of the lens suggests that the coordinate number of the lens

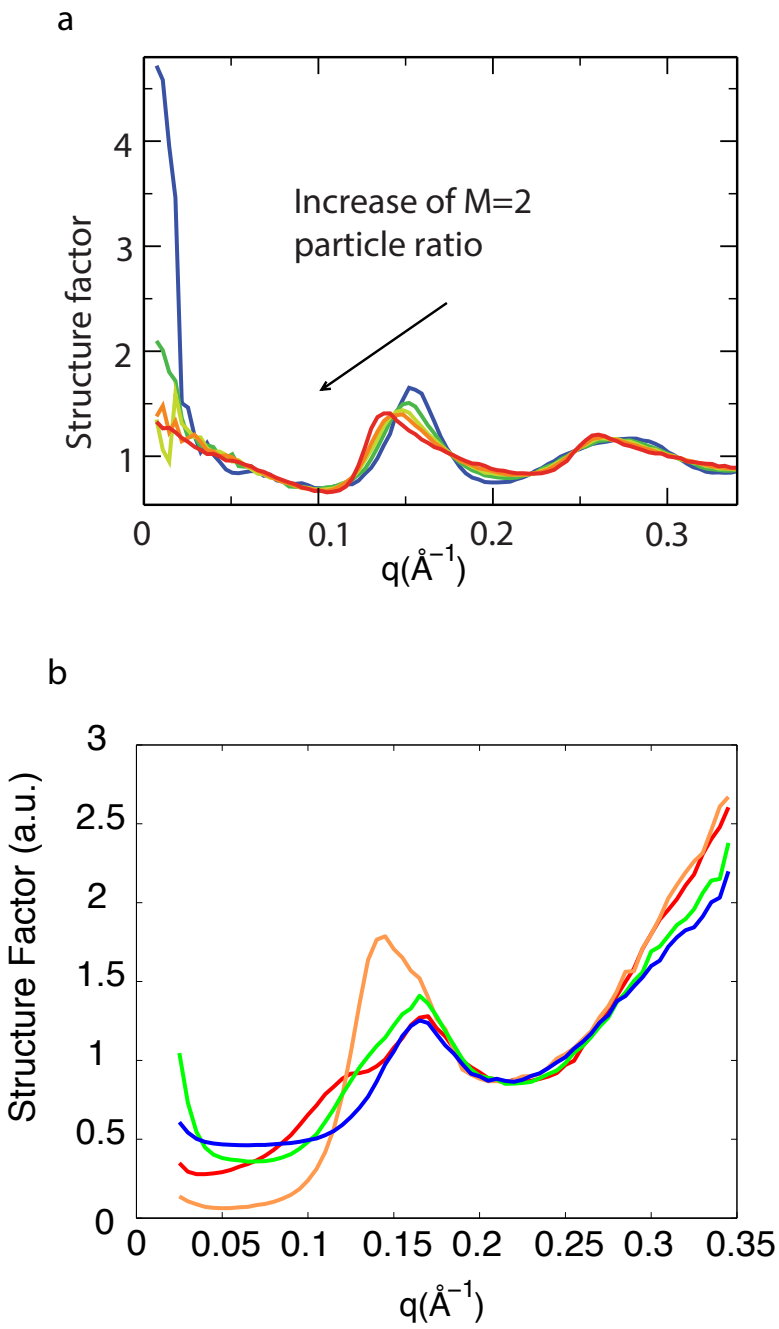


Figure 4.6: Comparison of the pair-wise peak in the structure factor. The top figure (a) is obtained from Sciortino's simulation, which shows that with decreasing $< M >$, the peak positions shift to a lower q [98]. From red to blue, the color labels the $M = 2$ particle to all particle ratio of 0.2, 0.5, 0.66, 0.75, and 0.9. The bottom figure (b) shows the structure factor of S-crystallins in the intact lens, where we see a similar trend. The color labels different radial positions of the lens, with 100% layer in red, 80% layer in orange, 60% layer in green, and 40% layer in blue.

proteins increases from periphery to the core of the lens.

In short, evidence from DAMMIF and SAXS indicate that the average coordinate number S-crystallin proteins in the lens increases from periphery to the core of the lens tissue. The average coordinate number at different radial positions, together with the packing fractions and the estimated bond energies map very well to the theoretical patchy particle phase diagram. This suggests that the lens generates is gradient refractive index by expressing low coordinate number proteins in the periphery and high coordinate number proteins in the core. Using this system, the periphery of the lens may be able to maintain a low density fluctuation by having a possibly gelled state without liquid-liquid phase separation.

4.4 S-crytallins with different $< M >$

We have seen that the lens distributes low coordinate number proteins in the periphery. The next question is: which proteins in the lens have low coordinate numbers?

In Chapter 2, I described our SDS PAGE and RNA sequencing results. SDS-PAGE showed that the vast majority of the protein in the lens is of a molecular weight consistent with S-crystallins, and RNA sequencing provided many new candidate S-crystallin sequences. We found more than 50 isoforms of S-crystallin in the lens, and the major difference between the isoforms is the amino acid composition and the length of the loops. We found that in the periphery of the lens, there are more 26.5 kDa S-crystallins, whereas in the core, the 24.5 kDa S-crystallin is most abundant. The abundance of 26.5 kDa S-crystallin gradually decreases from the periphery to the core. In addition, there are isoforms with super-long loops of about 100 amino acids. The abundance of these isoforms increases first and then decreases from periphery to the core. In this section, I will link what I have learned from the S-crystallin sequence analysis to patchy particle assembly.

In our DAMMIF output from the peripheral layer, the $< M >= 2$ particles are usually

spherical with the two loops stretching out to a distance of about 40 Å. This is consistent with the size of the loop from the 26.5 kDa S-crystallin isoforms. Therefore, we associate the 26.5 kDa S-crystallin isoforms with the low coordinate number proteins in the lens. The decrease of the abundance of the 26.5 kDa isoforms in the core is consistent with the increase of the $< M >$ from periphery to the core. Our GROMACS simulation between the two long-loop S-crystallins also supports this hypothesis that the 26.5 kDa isoforms have a low average coordinate number. Together with our IEF-PAGE result that the S-crystallins are positively charged, the data suggest that the loop-loop interaction is the only attractive interaction between the two 26.5 kDa isoforms, and the rest of the protein surface is repulsive due to the high charge. Thus, we posit that the 26.5 kDa S-crystallins are low coordinate number particles, which are very likely to be $M = 2$ particles under the combined attractive interaction of the loops and a screened Coulomb repulsion from the folded bodies of the proteins.

As we go one layer towards the core, in the 80% layer, the DAMMIF result shows that some of the proteins are connected via a long, thin chain of amino acids. This result is consistent with the SDS-PAGE result that there is an increase in the abundance of the super long-loop proteins in the 80% layer. This loop is very likely to be negatively charged, as is suggested by our IEF page result. Thus, this negatively charged loop can bind to a few long and short-loop S-crystallins which are positively charged. This is one possible way for the 80% layer to contain proteins with a higher $< M >$ than the peripheral layer. The super-long-loop S-crystallins are very likely to be a particle of $M > 2$, and by having a high abundance of the super long-loop S-crystallins, the average coordinate number may be increased in the 80% layer.

For the core two layers, the 24.5 kDa S-crystallin isoforms have a high abundance, and $< M >$ is also high in these two layers. These isoforms have short-loops and low surface charge [114], and therefore, it is possible that these proteins may be able to link to equal to

or more than 3 other proteins and form a $M \geq 3$ patchy particle. The short-loop proteins can also help in building a high packing fraction system as required by the observed refractive index.

In short, the low $< M >$ in the periphery of the lens observed in DAMMIF could be achieved by having relatively more 26.5 kDa S-crystallin isoforms, which could be $M = 2$ patchy particles. In the 80% layer, the increase of $< M >$ is likely due to a combination of the increase in the abundance of the short-loop and the super long-loop S-crystallins, both of which could be high M patchy particles. Towards the core, the short-loop proteins are more abundant and since the $< M >$ of the short-loop proteins are high, the core of the lens has a high $< M >$.

4.5 Gelled network of the lens

In the previous sections, I presented data and analysis suggesting that the squid lens is a patchy particle system which can form a stable, gelled material at any packing fraction. This phenomenon emerges from a set of S-crystallin proteins that has evolved variable coordinate number. These proteins are expressed in different mixtures in cells located at different radial positions in the lens. The spatial configuration of network will be further discussed in this section.

4.5.1 Transmission electron microscopy (TEM)

Sciortino and his collaborators have performed molecular dynamic simulations on a mixture of $M = 2$ and $M = 3$ patchy particles [104]. By varying the ratio of these two kinds of patchy particles, the resulting configuration in real space and the structure factor vary as well. The spatial configurations of their systems with different ratios of $M = 2$ and $M = 3$ are shown in Figure 4.7. The particle spatial positions are shown with $M = 3$ particles in

red; $M = 2$ particles in green; and yellow represents the sticky spots on the particles' surfaces. The figures on top and bottom differ in the $M = 3$ over $M = 2$ particle ratio. In either figure, the structure of the particle network shows a chain-like structure with bifurcations. This chain-like structure spans over the whole space, resulting in a homogeneous particle assembly.

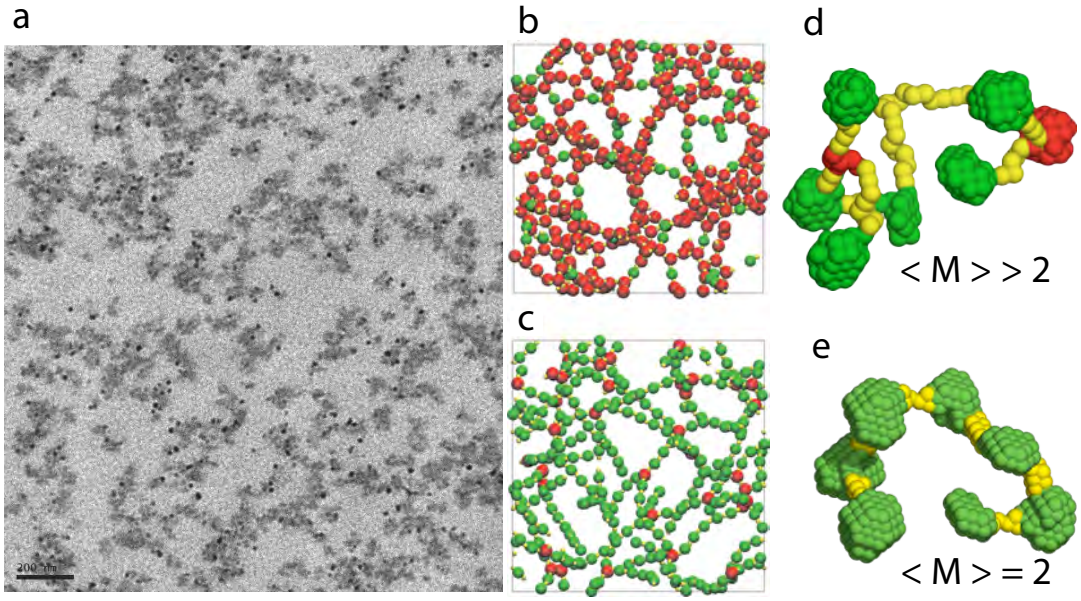


Figure 4.7: TEM and the simulated protein network. Our TEM image (a) of the 100% layer in the squid lens shows that the proteins form a network. This is compared to the spatial configurations from Sciortino's simulation (b,c) [98], with the red $M = 3$ particles; green $M = 2$ particles; and yellow the patches. The similarity between our TEM image and the predicted configuration in simulations of patchy particles supports the hypothesis that the proteins can be treated as patchy particles in a gelled system. This is further supported by result from the DAMMIF fit (d,e).

To compare this picture to our lens protein structure, I performed transmission electron microscopy (TEM). The experimental details of this procedure are described in Chapter 2. The TEM image shows that the protein assembly is very similar to the particle spatial positions from the simulations described above [104]. In TEM images, the proteins form chain-like structures with bifurcations. The assembly is overall connected and homogeneously distributed throughout the cellular space. There seem to be more chain-like

structures than bifurcations in the image. Therefore, gross observation of our TEM images suggests that the peripheral layer of the lens contains more $M = 2$ particles. This TEM image is also similar to the Sciortino's simulation result with a mixture of $M = 2$ and $M = 3$ particles. In addition, the network spans the whole space, which suggests that the S-crystallin network in the periphery of the lens is percolated and forms a gel. All of these observations are consistent with our SAXS intensity at low q and our native PAGE and filtration results suggesting that the proteins in the lens form a larger scale structure which may be a gel.

In this section, I compared the TEM image from the squid lens to a patchy particle system with a mixture of $M = 2$ and $M = 3$ particles. The TEM images contain more chain-like structures than branching structures, suggesting that the majority of S-crystallin molecules in the peripheral layer of the lens are $M = 2$ particles. This result is another piece evidence which supports the conclusion that in the peripheral layer of the lens, the averaged coordinate number is low, which is required to avoid liquid-liquid phase separation.

4.5.2 Ramification simulation

In the previous section, I showed that the squid lens assembly can be treated as a patchy particle system where the average coordinate number increases from the periphery of the lens to the core. In this section, I will further explore this idea by simulating a bifurcating network, and compare the structure factor from the squid lens to the simulated assembly. Specifically, I performed a ramification simulation to study the effect of $< M >$ on the spatial configuration of a resulting network with some similarities to the patchy particle networks described above. By ramification, I mean particles form chain-like structures with bifurcation.

The computer simulation creates different $< M >$ assemblies and calculates the structure factor. The simulation performed on the system is composed of a combination of

$M = 2$ and $M = 3$ particles. All particles have the same diameter. For each simulation, the total number of particles is 100,000 . Particles are generated at in each step of the walk. The walk algorithm is similar to three dimensional self-avoiding random walk, except that the particle positions come from a normal distribution with reference to the x-axis of the previous particle. One simulation step may involve adding either one or two particles, where the probability of adding two particles for each step is a fixed input to the simulation. When two particles are added, each one continues to add more particles with the same original probabilities. The position vector of the newly added particle is chosen to be 1 for radial component. A Gaussian distribution is used for the cosine of the polar angles $\cos(\theta)$ with the expectation set to be 1 and variance set to be 0.2. For any $\cos(\theta) > 1$, $\cos(\theta)$ is set to be $2 - \cos(\theta)$ to avoid complex numbers. The definition of θ is followed in Matlab's definition of spherical coordinates. A uniform distribution is assigned for the azimuthal angles ϕ in the position vector from the previous particle. The percentage of $M = 3$ particle is chosen to be 0, 0.03, 0.1, 0.2, and 0.5 respectively for the 5 particle systems we created. The positions of all these particles were Fourier transformed using the Debye method to calculate the structure factor.

In the Figure 4.8, I show the result of our ramification simulation for different ratios of $M = 2$ and $M = 3$ patchy particles. The plot shows the spatial positions of each system with different $< M >$. These structures are shown in the x-y plane with a projection in the z-direction. Note that x- and y- directions have different scales. The results show that with an increased ratio of $M = 3$ particles, there are more bifurcations present, and the resulting network is more compact. This in turn means that these systems have higher packing fractions.

The structure factors of these configurations shows similar trend to the structure factor of the lens at large q. At $q\sigma = 2\pi$, corresponding to $q = 0.15$ Å^{-1} for the lens protein, there is a peak which corresponds to the nearest neighbor distance. There is also a minimum in

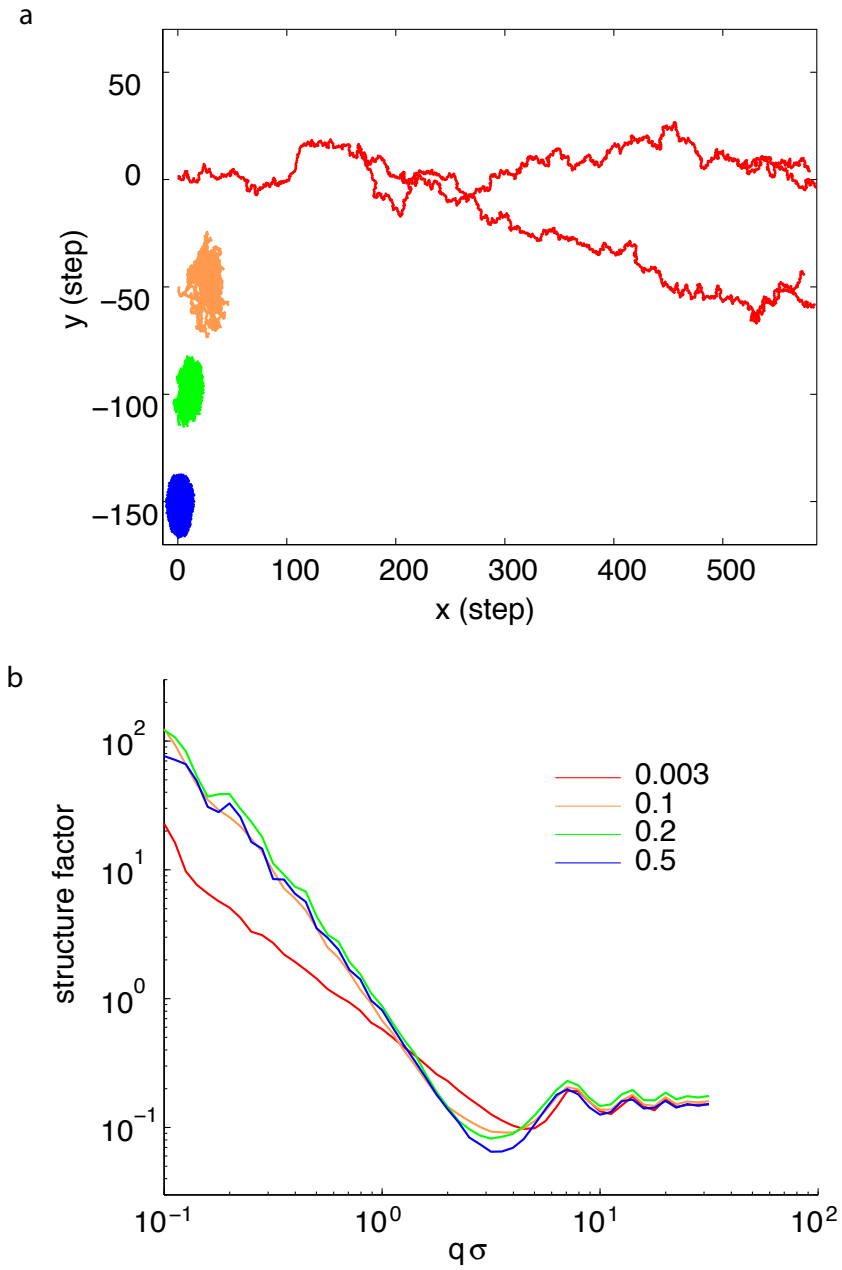


Figure 4.8: Ramification result. The configurations (a) result from varying percentage of $M = 2$ and $M = 3$ particles: The red curve shows the configuration (a) and the calculated structure factor (b) from a simulated system with the ratio of the number of $M = 2$ particle to all particle = 0.003, orange 0.1, green 0.2, and blue 0.5. σ is the diameter of the particle. The resulting networks are denser with the in the fraction of $M = 3$ particles. Only the first 2000 steps from the simulation are shown in (a). The calculated structure factor of these configurations are shown in (b). The depth of the minimum at $q\sigma = 3.5$ increases with increasing $M = 3$ particle ratio.

structure factor at $q\sigma \sim 3$ ($q \sim 0.07$ Å^{-1} in the lens). The structure factor also increases with decreasing q at low values of q.

The depth of the minimum at $1 < q\sigma < 2\pi$ ($0.03 < q < 0.1$ Å^{-1}) increases with an increase of $< M >$ for the configuration of particles. The ratio of the depth for $< M >= 2$ particle system and $< M >= 2.5$ particle system is approximately 3. The decreasing minimum indicates a decrease of the spatial density fluctuation at this position. This increase can be explained by having a higher packing fraction of particles in real space, and a higher packing fraction is a result of having a higher $< M >$.

When compared to our SAXS experiments, there are interesting similarities between the lens and this simple ramification simulation. Next to the nearest neighbor peak, there is a minimum in the structure factor at $q \sim 0.05$ Å^{-1} for all of the layers in the lens, and the depth of this minimum decreases from the periphery to the core. The similarity of the structure factors in this simulation and in measurements of the lens suggests that the lens system can be approximated by chains of two-coordinate particles with nodes of three-coordinate particles. In addition, in our structure factor from the lens, the ratio of the depth of the minimum in 40% layer and the 100% layer is about 5, which is greater than the ratio of $< M >= 3$ to $< M >= 2$ in the simulation result. This suggests that the lens system has a bigger span of $< M >$ than our ramification simulation which mixes particles with $M = 2$ and $M = 3$. In another words, comparing the ramification simulation results to the structure factors from the squid lens suggests that $< M >$ in the core of the lens is higher than 3.

4.6 Self-assembly of the squid lens - a unified view of the data

Up to this point, I have studied the protein composition and structure in a squid lens using biochemical approaches, and the larger scale structure of the lens using SAXS. I have

proposed a mechanism about how the lens is assembled. In this section, I will summarize what I have learned from the lens and explain clearly the mechanism of the assembly of the squid lens.

Contrary to previous expectations [114], our SAXS measurements of an intact squid lens shows that proteins are attractive and in contact with one another. Both the SAXS intensity and the structure factor decrease at $q < 0.05$ Å^{-1}, corresponding to a larger scale structure in the lens, inconsistent with all repulsive interactions. This large-scale structure, on the other hand, is still dominated by protein-protein pair-wise interaction at all lens positions, shown by the presence of the nearest neighbor peak at the protein diameter length-scale at $q = 0.13$ Å^{-1}. This is different from SAXS measurements of fish lenses, where the α-crystallin nearest-neighbor peak is absent at the core [79]. In addition, the lens system cannot have liquid-liquid phase separation such as protein aggregation or precipitation because the difference in refractive index between the two fluid phases would scatter light, resulting in lens opacity [4, 10, 41, 79]. In fact, most systems with isotropic attractive interactions are very likely to have liquid-liquid phase separation at low packing fraction [24]. In contrast, the phase diagram of patchy colloids shows a self-assembled route to avoiding liquid-liquid phase separation at all possible packing fraction. The theoretical basis for this idea comes from work done by Sciortino and colleagues showing that the spinodal line moves towards a smaller packing fraction with a decrease of the average coordinate number $< M >$. A patchy particle gel with no liquid-liquid phase separation can be achieved at systematically smaller packing fractions by reducing the coordinate number of the patchy colloids [12]. This result of stable, gelled, nearly empty systems at low $< M >$ has been corroborated experimentally in non-living systems [99].

Our experimental and analytical results suggest that S-crystallin coordination number in the gel network increases as a function of decreasing lens radius, effectively making it a system of patchy particles. Our DAMMIF fitting at high lens radii clearly shows 50 Å

spheres, consistent with folded S-crystallin proteins. We identify these spheres as canonically folded S-crystallins, as is shown in Figure 4.2. These S-crystallins connect with different numbers of other S-crystallins, the frequency of which was estimated using repeated Monte Carlo simulation in DAMMIF. This analysis showed that the average coordinate number $< M >$ of the particles increases from 2.19 to 2.45 from the 100% layer to 80% layer. This increase of $< M >$ is 6.9 times higher than the standard deviation. Towards the core of the lens, lens radius of 60% shows lamellar-type structures similar to those observed in the simulation study of high surface area patchy-type particles [131]. In the core of the lens, the protein packing fraction is so high that DAMMIF predicts that space is nearly entirely filled with protein. The absence of the $q = 0.21$ Å^{-1} peak in the core suggests that the proteins are so tightly packed the in core that water is possibly no longer present.

Figure 4.5 shows the inferred phase diagram of S-crystallins in the lens at different positions along the radius of the lens. Having $< M >= 2.19$ in the periphery of the lens results in a spinodal line located around $\phi = 2.5\%$ [12], which is a slightly smaller packing fraction than that in the outermost layer of the squid lens. For the 80% layer, the inferred spinodal line of $< M >= 2.45$ is located at $\phi = 8\%$, smaller than the packing fraction of the proteins in this region of the lens. Thus, by decreasing the average coordinate number $< M >$ of S-crystallins in the periphery of the lens, the lens managed to place the spinodal line at a smaller ϕ than the actual ϕ of S-crystallins in the lens at different regions, leaving the whole lens in a stable, and possibly gelled state without liquid-liquid phase separation.

The combination of our results from fitting our SAXS data and our biochemical data suggest a biophysical mechanism for generating a low M particle. In the periphery of the lens, DAMMIF shows 50 Å spheres mostly joined by two 25-35 Å long bridges. These bridge dimensions are consistent with a doubled-back chain of amino acids that is 10-20 amino acids long. The angular separation of these bridges relative to the 50 Å spheres ranges from 60^{o} to 120^{o}, which is consistent with the predicted positions of the evolu-

tionarily novel, disordered domains in S-crystallins. These DAMMIF results suggest that S-crystallins with 10-20 amino acids in the loop are $M = 2$ particles. Further, SDS-PAGE shows that proteins with loops that are 10-20 amino acids long are dominant in the outermost region of the lens, composing 60% of the protein mass of the lens found in the outer regions. The fact that both DAMMIF and SDS-PAGE results show that S-crystallins with 10-20 amino acid loops dominate the lens periphery supports our hypothesis that these proteins can be treated as $M = 2$ patchy particles. Moreover, our GROMACS result suggests a mechanism for the linkage between these 10-20 amino acid loop S-crystallins. A stable but dynamic patch of 3-4 hydrogen bonds is likely to form between the polar residues in the loop when the centers of mass of the proteins are positioned according to experimental data. The energy of this predicted interaction ($\sim$ 60 kJ/mol) suggests that these bonds can be stable at room temperature. Hence, each loop of these 10-20 amino acid loop S-crystallins can be treated as a patch which connects to other S-crystallin loops , making these S-crystallins $M = 2$ particles. DAMMIF fitting on the outer layers occasionally shows loops interacting with the folded near-spherical body of other proteins, suggesting a mechanism for occasional interactions of $M = 3$ particles in the low-density regions of the lens.

DAMMIF results also suggest a mechanism for generating $M = 3$ particles from S-crystallins. DAMMIF fits from lens 80% layer show 3-fold coordination in the narrow bridges between 50 Å spheres. We refer to proteins having this bridge as super-long-loop proteins, with 103 amino acids in the loop as observed from SDS-PAGE and RNA sequencing. These very long disordered loops may be allowed to form multiple hairpins, with each hairpin allowed to interact through a multiple-hydrogen-bonding interaction with other hairpins. Super long-loops have lots of charged and polarized amino acids: out of 103 amino acids, 20 of them are negatively charged, 8 positively charged, and 27 polarized. All of these are consistent with these structures' ability to form multiple polar interacting

patches, and it is possible that these proteins have a very high coordinate number. Although these proteins are present at relatively low concentrations in the lens (1% to 3% of total protein mass), this would be enough to shift $< M >$ by up to + 0.36 assuming these proteins have $M = 20$. At lens radii $< 60\%$, where $M > 3$, protein-protein interactions seem to shift from being mediated by unstructured loops to apparently relatively non-specific large-area docking-type interactions on the surface of the proteins, more typical of better studied systems.

Chapter 5

Chinese century egg and squid lens

The Chinese century egg (pidan) has been a traditional Chinese food for hundreds of years. It was recorded that in the early Ming dynasty (1368–1644) a duck egg was laid in lime and two months later, the egg white was gelled and the yolk turned to black. Nowadays, pidan is normally made by alkali-treated fermentation in clay. The clay is usually composed of sodium carbonate (Na_2CO_3), water (H_2O), and calcium oxide (CaO). Sodium hydroxide (NaOH) is the product from the reaction of the ingredients. This high base environment causes a physico-chemical change of the texture and color in the egg white [125]. The color of the egg white turns to a mostly-transparent yellow or transparent tea-brown color, shown in Figure 5.1. Pidan can be preserved at room temperature for months. It remains a popular food among people in Asian countries.

The physical transformation in the pickling process is that the proteins in the egg white partially unfold and form filamentous structures [43]. The proteins have short-range attraction and long-range repulsion. Under these interactions, they form a transparent, stable gel.

The transparent material resulting from the egg white in pidan shares quite a few similarities with the squid lens system. They are both composed of globular proteins, and these

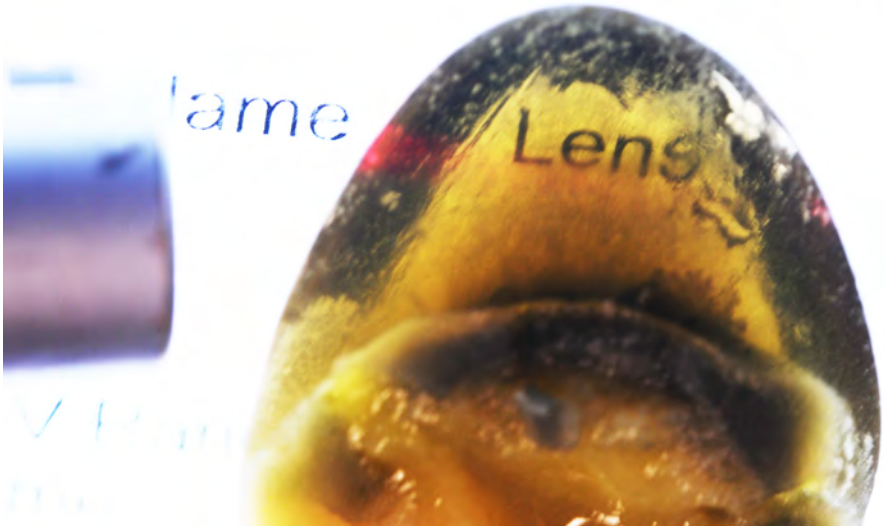


Figure 5.1: Image of Pidan. The pickled quail egg shows transparent tea-brown color. A laser beam injects into pidan and gets out from the other side, showing that pidan is transparent.

proteins form gels that are transparent. The packing fraction of the egg white is very similar to the periphery of the squid lens. In this chapter, I study the physical properties of pidan. Then I compare the pidan to a healthy lens system. Further, the heat denatured state of both systems are studied: I also performed SAXS on hard boiled eggs. To denature the lens system, I incubated the lens at 50 °C. The similarities between the heat denatured assembly of these two systems are discussed. Finally, I discuss both systems in the context of patchy particle theory.

5.1 Egg white protein

For an avian egg, the egg white accounts for about 58% of the total egg mass [77]. Protein is the major component of egg white, which makes up 9.7-10.6% of the composition by

weight [78]. Other components includes carbohydrates (0.5-0.6%), glucose (0.5%), and lipids (0.01%). The major proteins in the egg white are ovalbumin, ovotransferrin, ovomucoid, ovomucin and lysozyme [78, 103]. Briefly speaking, ovotransferrin is an antimicrobial, iron binding protein which constitutes 12% of the egg white proteins. Ovomucoid is a glycoprotein which is known for its trypsin inhibitory activity. Ovomucin is a structurally important glycoprotein which constitutes1.5-3.5% of the total egg white protein. The function of lysozyme is to lyse bacteria and it constitutes 3.4% of the egg white protein [78].

Ovalbumin is the most abundant protein in the egg white. It composes about 55% of the total egg white proteins [110]. There are three isoforms, A1, A2, and A3, which have different numbers of bound phosphate residues [42]. Ovalbumin belongs to the serpin superfamily, but unlike other serpins, it does not act as a protease inhibitor [51]. Across avian species — hen, pheasant, quail and duck — ovalbumin shows similar electrophoretic profiles, with quail and pheasant ovalbumin showed the most basic pI of 4.5–4.6 [77]. The estimated molecular mass across the species is approximately 45 kDa for ovalbumin monomer. The ovalbumin from hen eggs contains 385 residues.

5.2 Chinese century egg pickling

Chinese century eggs (pidan) are traditionally made from raw duck eggs in China. But for the purposes of my experiment, I used raw quail eggs to make my own pidan because they are much smaller. These raw quail eggs were purchased from a local Chinese supermarket. The raw quail eggs were carefully stored and pickled before the expiration date. To make pidan, whole quail eggs were soaked in an excess of 0.9 M NaOH (Sigma-Aldrich ACS reagent, $>$ 97.0% pellets, PO# 221465) and 0.5 M NaCl (Sigma-Aldrich SLBB9000V) at room temperature. After one to two days of soaking, the pigmented patterns on the shell began to fade and the shells became fragile. I washed the surface of the egg shell

and discarded the broken eggs. The rest of the eggs were soaked in the same solution for another 13 - 16 days at room temperature. After that, the eggs were taken out from the solution, dried, and kept at 4^{o}C for future use. A new egg waws peeled right before each experiment.

Under these conditions, raw egg white gelled after two weeks. It was transparent, with a slight yellow pigmentation. The pidan is very sticky, in the sense that sometimes it is hard to get a piece out from tweezers. The pidan is also flexible. Pressing on it gently without smashing it resulted in distortion, but when the pressure is released, it relaxed to the original shape. Unlike raw egg white, this pickled egg white was insoluble in water. Homogenizing the pickled egg in water first and then centrifuging still resulted in a lot of pellet. This pidan did not rot or decompose after a year and a half stored in a 4^{o}C refrigerator. But the color darkened during the storage, with the yellow pigmentation turning very dark brown. It was also less transparent compared to a freshly pickled egg white.

5.3 Density, gelation and viscosity

In order to compare to the lens system, I studied the physical properties of the pidan to understand its gelation. Measurements were conducted to estimate the density and viscosity of pidan. Then I studied the gelation conditions *in vitro* by adding different amount of NaOH into various concentrations of ovalbumin.

5.3.1 Density of pidan

The density of the pickled quail egg white was measured by submerging the gelled pidan samples into potassium iodide (Sigma-Aldrich, ReagentPlus, 99%) solutions with known density. A tissue sample will float in less-dense fluids, and will sink in more-dense fluids. The dense potassium iodide solution was made by adding 6g KI into 10 ml deionized water

and then making serial dilutions. The dilutions are 1, 1:2, 1:4, 3:16, 5:32, 19:128, 9:64, 1:8. For each dilution, the solution was separated into two 1.5 ml tubes. A piece of pickled egg white with diameter $\sim$ 3mm was placed into one of the solutions. Whether it floats or sinks was immediately observed. To find the density of the solution, I weighed the mass of 400 μL solution from the other tube, and density was calculated from that.

The result shows that pidan floats in 1, 1:2, 1:4, 3:16, and 5:32 solutions, whereas it sinks in 19:128, 9:64, 1:8 and water. The mass of 0.4 ml KI solution for 5:32 dilution is 0.4233g, and for 19:128 is 0.4199 g, corresponding to density of 1.0582 g/ml and 1.0498 g/ml. Therefore, the density of the quail egg white lies in between, and I estimated it as 1.05 g/ml. Assuming that the quail egg proteins are all ovalbumin, the density of which is 1.296 g/ml [54], then the estimated packing fraction for pickled quail egg white is 4%. This is approximately the same packing fraction as the outermost layer of a squid lens.

5.3.2 Viscosity of ovalbumin in basic solution

In the traditional alkali method of preparing pidan, the clay contains sodium hydroxide (NaOH), which leads to a strongly basic environment. To study the physical changes that egg proteins undergo in a similar alkaline environment, I performed viscosity measurement in this section and gelation study in the next section *in vitro*, using the main protein in the egg white, ovalbumin, and the strong base solution of NaOH.

For ovalbumin solution with low NaOH concentration, the liquid seems to be much more viscous than the liquid with no base added. I studied the viscosity with a viscometer I built. The basic idea for the viscometer is that a small object will fall more slowly in a more viscous environment. I applied this idea to the design of the viscometer: A 2 ml plastic pipette was used as the sample container, and a cylinder-shaped stir bar with rounded ends was dropped down the length of the pipette. The size of the stir bar was such that the diameter is just smaller than the inner diameter of the 2ml pipette, but the length of the

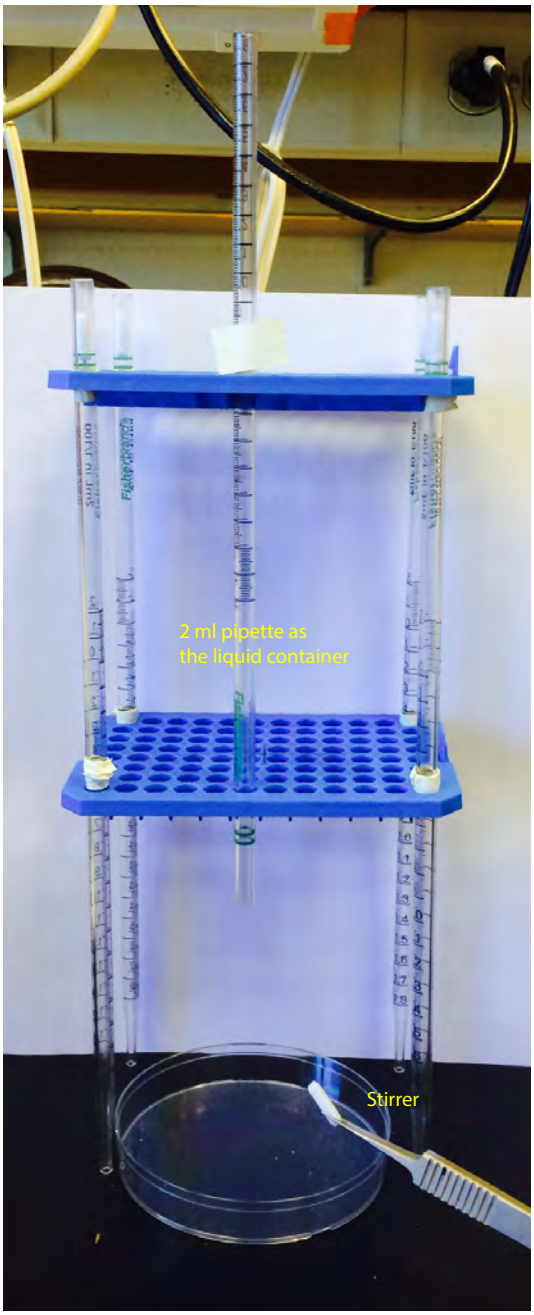


Figure 5.2: Viscometer. The viscometer I built is composed of a 2 ml pipette and a stir bar. The sample to be measured is placed in the pipette. The stir bar is dropped from the top of the pipette and travels down to the bottom with the time recorded by a stopwatch. The viscosity of the liquid will be linearly related to the travel time of the stir bar.

stir bar was much greater than the pipette diameter. Therefore, the stir bar can only move down inside the pipette with its long side parallel to the pipette. There was no rotation observed when the stir bar was moving downward. The pipette was set to be vertical. The experiments were performed by filling the pipette with the target sample, and then dropping the stir bar from the top. I measured the time it took the stir bar to move to the bottom of the pipette with a stopwatch. The apparatus is shown in Figure 5.2.

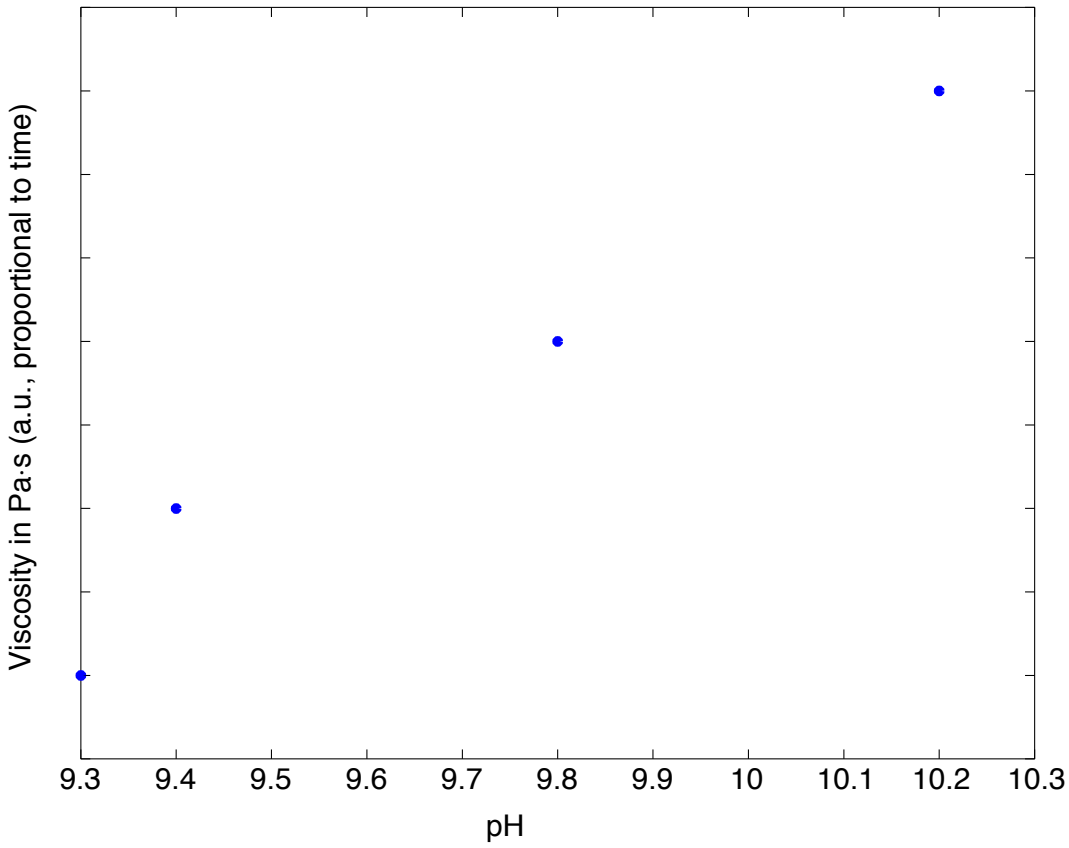


Figure 5.3: Viscosity estimated from the time that the stir bar took in the solution of 2.8 mM ovalbumin with different volume of 2M NaOH added was measured by my viscometer. The viscosity of the ovalbumin solution increases with the increase of the base concentration.

To calibrate this apparatus, I measured three liquids of known viscosity: methanol (5.9×10^{-4} Pa $\cdot$ s), deionized water (8.9×10^{-4} Pa $\cdot$ s), and ethanol (1.2×10^{-3} Pa $\cdot$ s). The times that the stir bar took to fall the length of the pipette in each liquid are 0.8, 1.3,

and 2.6 seconds respectively.

I used this viscometer to study the relation between NaOH concentration and the viscosity of my ovalbumin protein samples. For each measurement, I used 125 mg ovalbumin dry protein dissolved in 1000 μL DI water (125 mg/ml or 2.8 mM), with a variable volume of 2M NaOH added. The final concentration of NaOH ranged from 0.2 mM to 6.6 mM. The pH of the final solution was also measured.

The time spent for the stir bar to get to the bottom of the pipette is linearly related to the viscosity of the liquid. The time spent for the stir bar to reach the bottom of the pipette ranges from 1.7 seconds (pH = 9.3) to 2.4 seconds (pH = 10.2). Figure 5.3 shows the viscosity estimated from the time. There is a general trend that the higher the NaOH concentration, the slower the stir bar moves. That is, the liquid is more viscous with higher pH. At concentrations of NaOH added greater than 6.6 mM, gelation occured, and the stirbar no longer moved through the sample. The viscosity of the ovalbumin solution in weak base is higher than water, which is 8.9×10^{-4} Pa $\cdot$ s, but lower than ethanol's 1.2×10^{-3} Pa $\cdot$ s.

5.3.3 Gelation process

As I showed in the previous section, when the concentration of NaOH is high enough, gelation happens. In this section, I study the conditions under which gelation occurs. This process is studied by having different concentrations of ovalbumin adding in different concentrations of base [43]. Ovalbumin constitutes 55% of the proteins in the egg white. I made ovalbumin solutions at concentrations of 1.4 mM (62.5 mg/ml), 2.8 mM (125 mg/ml), and 5.6 mM (250 mg/ml) separately. Then different volumes of 2 M NaOH were added to the ovalbumin solutions. The total volume of the final solutions are 700 μl. The final NaOH concentration ranges from 0.05 M to 0.5 M.

Whether the protein gels or not in a tube is judged by the 'tube flip' method, following

a study by Sciortino and colleagues [99]. The sample tube is flipped upside down for a few seconds. If there is no visible movement of the protein sample in the tube, then I call this sample gelled.

These ovalbumin gels were transparent and colorless, and were qualitatively very similar to the squid lens. For a low protein concentration of 1.4 mM (62.5 mg/ml), ovalbumin solution gelled in a range of NaOH concentration from 0.1 M (pH $\sim$ 13) to 0.25 M (pH $\sim$ 13.4. It did not gel at NaOH concentration higher than 0.3 M (pH $\sim$ 13.5). For a higher concentration of ovalbumin at 2.8 mM (125 mg/ml), it did not gel at very low base concentration of 0.05 M (pH $\sim$ 12.7), whereas it gels at all base concentrations greater than 0.1 M, including 0.5 M (pH $\sim$ 13.7), at which the low ovalbumin concentration did not gel. At the very high ovalbumin concentration of 5.6 mM (250 mg/ml), the protein gelled at the base concentration of both 0.05 M and 0.5 M. Therefore, for a given ovalbumin concentration, it was likely not to gel at either a very low or very high base concentration. Given a fixed base concentration, a higher ovalbumin concentration solution was more likely to gel, especially when the base concentration is high or low.

Given that ovalbumin gels under strong base condition, I explored whether ovalbumin can gel under acidic conditions. I added 125 mg albumin in the following three conditions: 1. 0.5 M HCl with ovalbumin concentration 62.5 mg/ml (1.4 mM). 2. 0.3 M HCl with 43.1 mg/ml (0.96 mM) ovalbumin, and 3. 0.25 M HCl with 62.5 mg/ml (1.4 mM).

The 62.5 mg/ml ovalbumin in 0.5 M HCl experiment resulted in white precipitate at the bottom of the tube, with no gelation observed. The ovalbumin added in 0.3 M HCl resulted in white precipitate as well. But this precipitate is not as dense as in the previous case. For the case of ovalbumin in 0.25 M HCl, turbidity started after ovalbumin was added. After half an hour, it gelled in the tube. However, this gel was milky in appearance, not as transparent as the gels in base, and possibly therefore gelled through a distinct physical mechanism.

To summarize, my experiment of the physical properties of pidan shows that the pidan density is very similar to the periphery of the squid lens. *In vivo*, alkaline conditions increased the viscosity of ovalbumin solutions when the pH of the solution was less than 10.2. As I increased the base concentration, gelation occured. The lowest base concentration required for gelation was higher for a lower ovalbumin concentration. But if the base concentration was too high, the ovalbumin solution at low density did not gel, either. This gelled ovalbumin material seemed to have many similarities to the healthy squid lens system. For both cases, the gels are formed by globular proteins, the gels are transparent, and gelation occurs at low protein packing fraction.

5.4 Comparing SAXS on eggs and lens

MAXS was performed on raw quail egg, pidan, boiled egg using the in-house equipment. The experimental details are the same as described in Chapter 3. In addition, SAXS on squid lenses was also measured.

5.4.1 SAXS on egg

SAXS experiments were performed on pickled quail egg white, raw quail egg white, and boiled egg white. Pickled or boiled quail egg whites were cut into small cuboids before loading into button sample cells. I carefully adjusted the position of the cuboid samples so that the x-ray beam was centered in the sample. The SAXS intensity as a function of scattering angle is shown in Figure 5.4. These three curves show a general trend of decrease with increase of the wave vector.

To generate a form factor for pidan, I measured the raw quail egg white. Since the concentration is low (5% packing fraction), I assumed that there are no interactions between

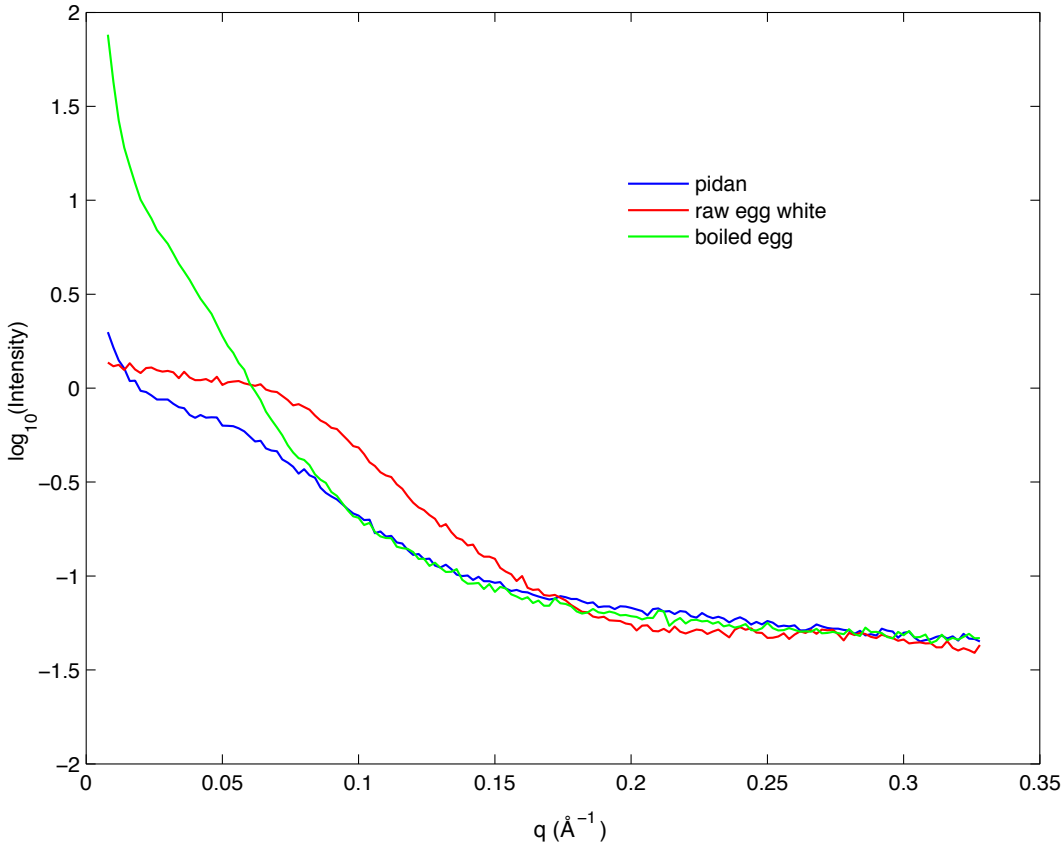


Figure 5.4: SAXS on pidan, raw egg white, and boiled egg white. X-ray scattering intensity as a function of wave vector for pidan, raw egg white, and boiled egg white are shown in different colors. The major difference between these intensities is the decrease at low q, and the protein nearest neighbor peak.

the proteins in the raw quail egg white. I used the scattering curve for raw egg white as the form factor for gelled pidan.

In the region of $q < 0.03$ Å^{-1}, the scattering curves for pidan, raw egg white and boiled egg white behave very differently. The raw egg white shows a flat shape at this range. This flat curve indicates that the monomeric proteins in a raw egg white do not aggregate into larger-scale structures. There are no strong attractive interactions between the proteins. Pidan intensity at this range, however, decreases slowly as the wave vector increases with the slope of ~ -0.6. This decrease denotes the possible larger structure forming in this range, and the assembly of this structure requires attractive interactions between the raw

egg proteins. This structure suggests that by pickling the raw quail egg in a strong base and a high salt solution, the proteins are altered in such a way that there are attractive interactions generated between the proteins. For the case of a boiled egg, there is a sharp decrease at small wave vector with the slope of -2. This sharp decrease indicates that there is a big real-space fluctuation in this range. This spatial fluctuation comes from the boiling treatment, where egg white proteins are completely denatured and aggregated into disordered clumps.

In the range of $0.03 < q < 0.15$ Å^{-1}, there is a shoulder in $I(q)$ for the raw egg white and pickled egg white, but no shoulder observed for the hard boiled egg. This shoulder corresponds to the size of the raw egg white protein. For the boiled egg, proteins are apparently fully unfolded, and there is no globular protein monomer structure remaining. So there is no peak in this range. For the scattering from pidan, there is a shoulder, which indicates that overall globular protein structure is still intact after the harsh treatment in strong base.

At $q = 0.135$ Å^{-1}, there is no peak observed from either the raw egg white or the pidan scattering curves. This suggests that the unit of protein assembly is a monomer, not the interaction of monomers. Figure 5.5 shows the calculated form factor from ovalbumin monomer together with the tetramer found in the crystallized protein as an example of the interaction of monomers [109]. The major difference between these two form factors is that there is a peak at $q = 0.135$ Å^{-1}. This peak corresponds to the size of the subunit that the interacting ovalbumin monomer forms. Thus, if the monomers form some higher structure subunits, it would be expected to have a peak at a q lower than the nearest-neighbor q. For both my pidan and raw egg white scattering curve, I do not observe any of these peaks. This absence of features at q smaller than the monomer-monomer interaction length indicates that the single protein monomer unit is the building block of pidan's structure.

At a smaller spatial scale, $q > 0.15$ Å^{-1}, there is a shoulder at $q = 0.27$ Å^{-1} for raw

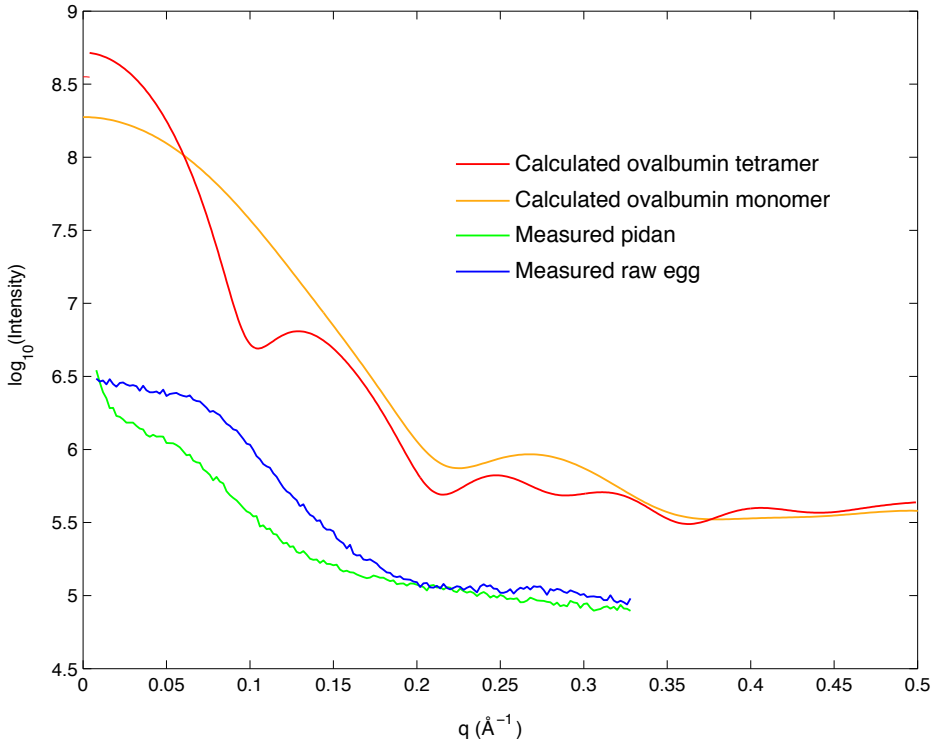


Figure 5.5: Form factor of pidan. Possible form factors for pidan were estimated by calculating the Fourier transform of the crystal structures of ovalbumin monomer and ovalbumin tetramer. The ovalbumin tetramer data show a peak at $q = 0.135$ Å^{-1}, where there is no peak in the same calculation for ovalbumin monomer. Comparison of estimates of form factor from crystallography to the SAXS data on raw egg white and pidan indicates that the ovalbumin monomer is the structural unit in pidan, and was therefore used as a form factor for pidan.

egg white, but no peak is found for pidan or boiled egg white. This shoulder corresponds to the length scale of 23 Å and therefore it comes from the form factor of proteins in raw egg. Since in the egg white, there is a mixture of different proteins with various sizes and inner structures, this polydispersity may have caused a shoulder at 0.27 Å^{-1}, rather than a clear peak. For the cases of pidan and the hard-boiled egg, the strong base or the heat treatment increases the polydispersity or completely unfolds the protein, resulting in the disappearance of the shoulder.

5.4.2 Comparing the structure factor of pidan to the lens

As described in Chapter 3, SAXS measurements of heat treated and untreated lens samples are very similar to the intensities of boiled egg and pidan, respectively. The cooked lens sample and boiled egg white show a disappearance of a peak and a sharp drop at a very small angle, which are different in a fresh lens and a pidan.

The structure factors from both the pidan and from the healthy squid lens are calculated by dividing by a form factor, or measurement of the isolated component protein. The intensity of the raw egg white was used as the form factor for pidan, and the average form factor from our S-crystallin homology models was used for the untreated lens, as described in Chapter 3.

To compare the structure factors from pidan to that of the untreated lens, I plotted them together in Figure 5.6 with q in the unit of the respective protein monomer diameters for both cases, σ. The diameter is chosen to be 30 Å for pidan monomers and 40 Å for lens proteins. The structure factors of pidan and the lens are very similar to each other. There are generally three features for both cases, as is explained in the following:

At a large spatial scale, with $q\sigma < 2$, there is a decrease in the intensity, which is sharper in the lens layers than pidan. This decrease is associated with the attractive interactions in the larger gelled structure in both cases.

In the intermediate range where $2 < q\sigma < 6$, there is a trough in both cases. This trough is due to the impenetrability of the proteins which leads to suppression in the density fluctuation at a scale a bigger than the protein diameter in real space. For pidan, the minimum is at one third the intensity of the nearby plateau. In contrast, the lens minimum is relatively deeper. The intensity at this minimum in the lens ranges from 20 to 110 times smaller compared to the peak at $q = 0.13$ Å^{-1}. This trough also appears in my ramification simulation result described in Chapter 4. The deeper the local minimum, the higher the average coordinate number of the proteins are predicted to be. Thus, the proteins in pidan have a

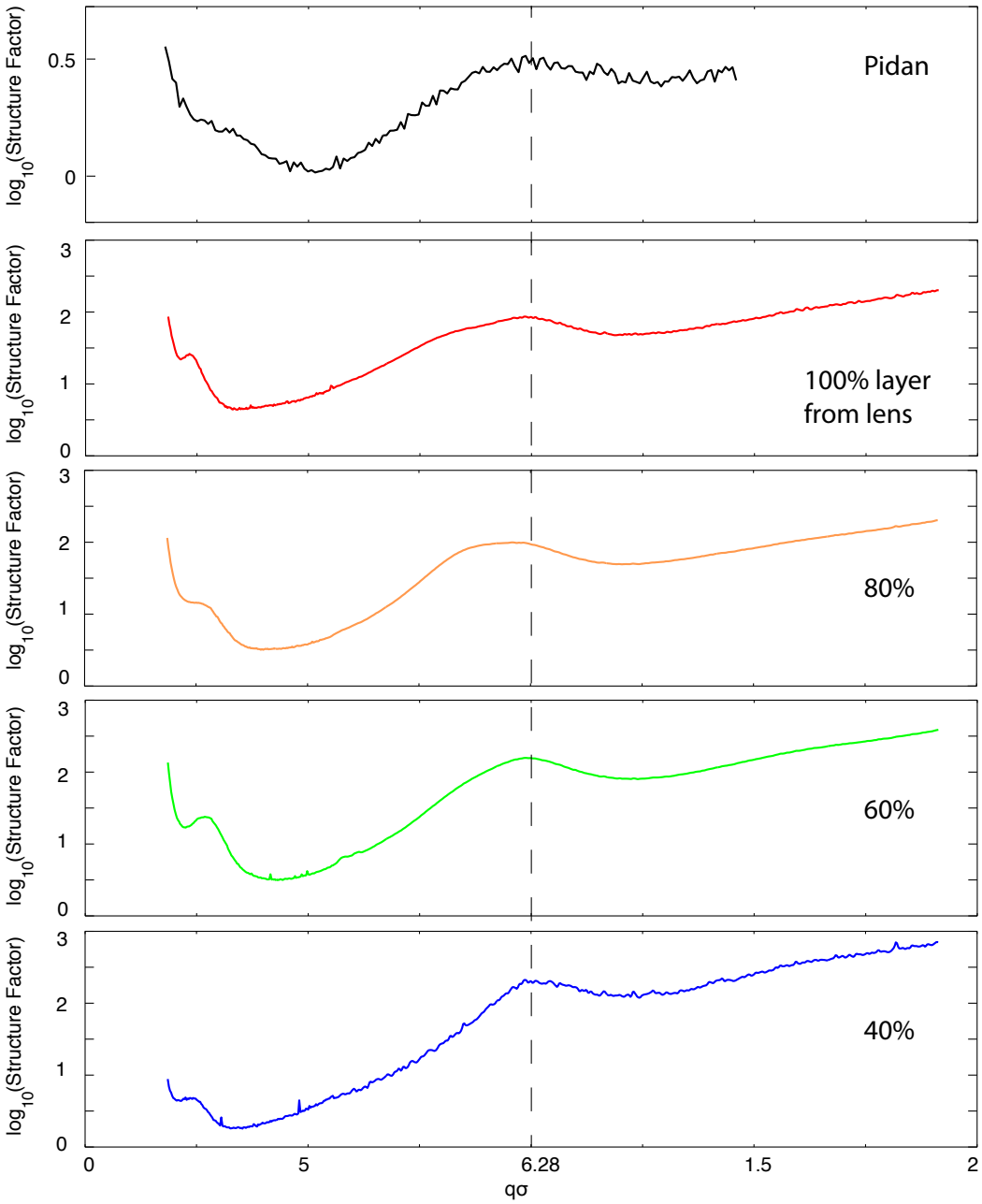


Figure 5.6: Structure factor of pidan and the squid lens. The structure factor of pidan was also estimated by using the intensity of a raw egg white as the form factor. Note that top panel has different vertical axis. Vertical line labels at $q\sigma = 2\pi$. The similarity between the structure factor of pidan and all layers in a squid lens suggests that they have similar underlying physics. The small jump (or Zing) in the intensity at 60% and 40% layer in the lens is a common feature for x-ray scattering data, which is likely to come from cosmic rays.

lower coordinate number compared to most radial positions in the lens, which is consistent with the fact that the density of pidan is also lower than most of the lens.

At $q\sigma = 6$, there is a peak at the real size position of protein-protein interaction for both cases. The protein unit is S-crystallin dimers for the lens, and ovalbumin monomers for pidan. This peak indicates that the individual, folded proteins are present for both cases, and the globular proteins are the basic unit of the structure in these two gels.

5.5 Spatial configuration

I also studied the real-space properties of pidan using techniques similar to those for the lens described earlier. The radial distribution function of pidan was calculated using GNOM, and a real-space structure of a protein as constructed using DAMMIN. These results were compared to the properties of a lens.

5.5.1 Radial distribution function (RDF)

As in the lens system described in Chapter 4, GNOM was used to estimate the radial distribution functions of the pidan, raw egg and the peripheral layer of the squid lens. The details of GNOM can be found in Chapter 3.

The RDFs of both the pidan and the raw egg white show some similarities, shown in Figure 5.7. There is a peak located at 26 Å for both curves, and this distance is consistent with the monomer-monomer distance. At a larger distance, the RDF of a raw egg shows a peak at 95 Å, whereas this peak moves to 125 Å for pidan. The height of this peak is much smaller than the first peak at 26 Å for both cases. For both the raw egg white and pidan, this second peak corresponds to a larger scale structure resulting from the interactions of monomers. The increase of the peak position from raw egg to pidan suggests that during

the pickling process, the a partial unfolding of the egg protein causes it to expand into a somewhat larger scale structure.

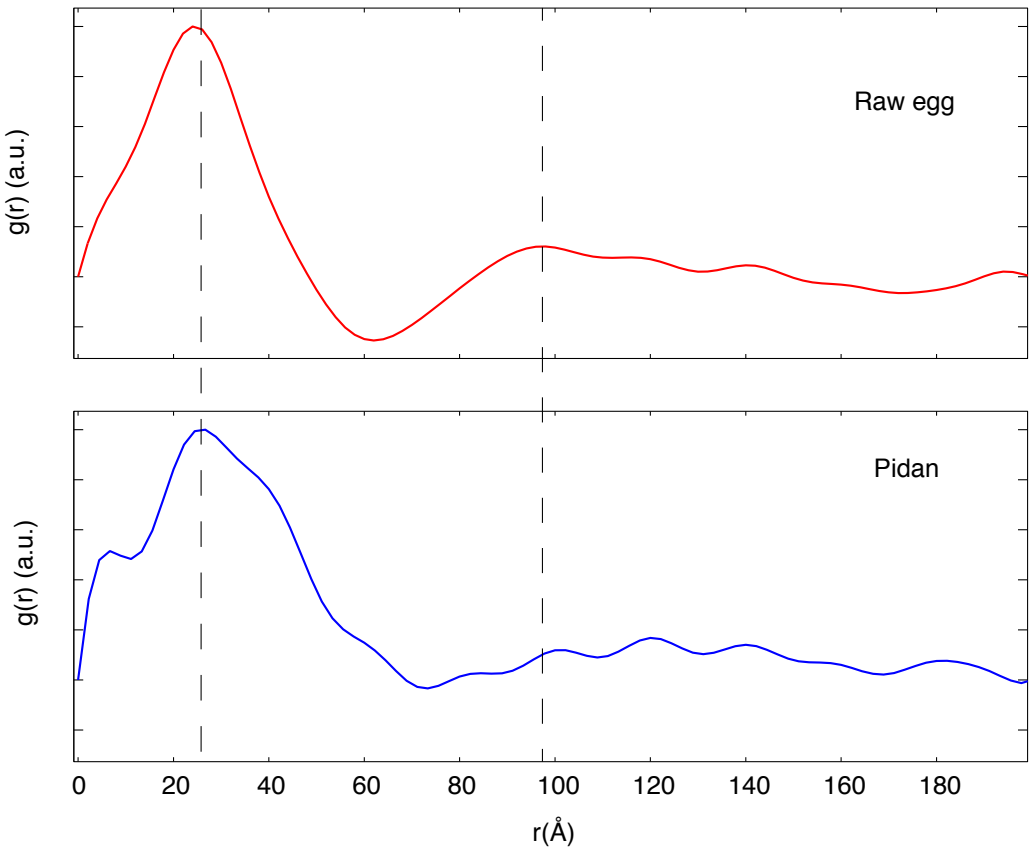


Figure 5.7: Radial distribution function (RDF) from the eggs. The RDFs from a raw egg white (red) and a pidan (blue) are plotted. They both show a peak at 26 Å and a second peak at $r > 90$ Å. The 26 Å peak corresponds to the monomeric unit of an ovalbumin protein. The second peak corresponds to the nearest neighbor distance of the monomers. The increase in the position of this second peak from raw egg white to pidan suggests a gelled structure of pidan build from interactions of partially unfolded ovalbumin monomers. Vertical lines locates at 26 and 97 Å. Data at $r < 18$ Å are under the resolution of the SAXS, which is not trustable.

The RDFs for the squid lens in the peripheral layer are very similar to the RDF of pidan. The lens RDF shows a peak at 25 Å in the lens (Figure 4.1), which corresponds to the size of a monomer S-crystallin. At larger ranges, there is an obvious peak appearing at 71 Å and

a subtle peak at 130 Å. Both of these correspond to the large structure that the S-crystallins assemble in an intact lens. The 71 Å peak comes from the average distance of dimers. The 130 Å peak may reflects the second-nearest neighbor distance.

Comparing the RDF from pidan to that of the lens shows several similarities between the two curves. Firstly, they both show similar patterns with a major peak which corresponding to the protein-protein monomer interaction. This similarity emerges because in both cases, larger structures were composed from a single protein building block, relatively monodisperse in size. The protein building block is an ovalbumin monomer in the case of pidan, and a S-crystallin dimer in the case of a squid lens. Then at a larger scale, there is a distinct peak, which is 71 Å for S-crystallins and 125 Å for pidan. This peak reflects protein quaternary structure. The 71 Å for S-crystallin dimer-dimer distance is the basic unit which is very likely to span the whole volume in the lens periphery. In the pidan case, the shift of this peak from 95 Å in a raw egg to 125 Å of pidan indicates a possible shift in the ovalbumin protein fold during the gelling process.

5.5.2 Construct the spatial configuration of pidan

To estimate a likely three dimensional spatial configuration of proteins in pidan, I used my SAXS data in DAMMIN. I used a fast mode simulation with the maximum diameter of particles set to be 200 Å. No symmetry was presumed. The output configuration shows a mostly connected object with size $65 \times 45 \times 30$ Å in three dimensions, as is shown in Figure 5.8. This size is very similar to the size and aspect ratio of an ovalbumin monomer, which is approximately $70 \times 45 \times 50$ Å. There is one dummy atom which is isolated from the rest of the dummy atoms, located apart from the volume-filling region of the other dummy atoms. The distance between the isolated dummy atom and the rest of the structure is 32 Å. The possible role of this structure in the gelation process of the quail egg white will be discussed in Section 5.6. The rest of the atoms are mostly arranged in a 29 Å-thick diamond-shaped

plane. This is thinner than 36 Å, the thickness of the ovalbumin monomer. This suggests that the main part of the ovalbumin monomer is still folded, but possibly some segments detach from the main monomer.

The shape of the diamond is very similar to the shape of the ovalbumin monomer. In the bottom of the diamond, there are three dummy atoms lined up in the top-left direction at a length of 30 Å. In the ovalbumin monomer, the distance between the left and the bottom vertex is 30 Å as well. Together with the dummy atom on bottom right, DAMMIN output shows the bottom of the output pattern like an arrow pointing downward. This feature is very similar to the ovalbumin monomer in the structure formed by residues number 200 - 300 in the sequence. In the middle left of the DAMMIN output, the dummy atoms form a dent, with the narrow part located in the very middle of the whole shape. This is very similar to the concave shape in ovalbumin formed by residues 290 - 300 and 335 - 340. The DAMMIN output also shows a region 45 Å wide at the top of the folded protein, similar to the top of the ovalbumin monomer where the length is 45 Å across the protein.

5.6 A possible gelation mechanism of pidan

Gelation of the proteins in a quail egg white happens after soaking the egg into a high base environment. During this process, proteins are negatively charged and partially unfolded. An estimation of the charge of the ovalbumin can be approximated based on the amino acid sequence and the pK_a of each amino acid. This is done using an online server PROTEIN CALCULATOR v3.4 (http://protcalc.sourceforge.net/). The estimation shows that under pH $= 13$ environment, which is the egg white pH after soaking the egg in base for 20 days [43], the estimated charge is negative 63. Thus, the ovalbumin in pidan is highly negatively charged.

Our SAXS data and further analysis indicate a possible gelation mechanism of the

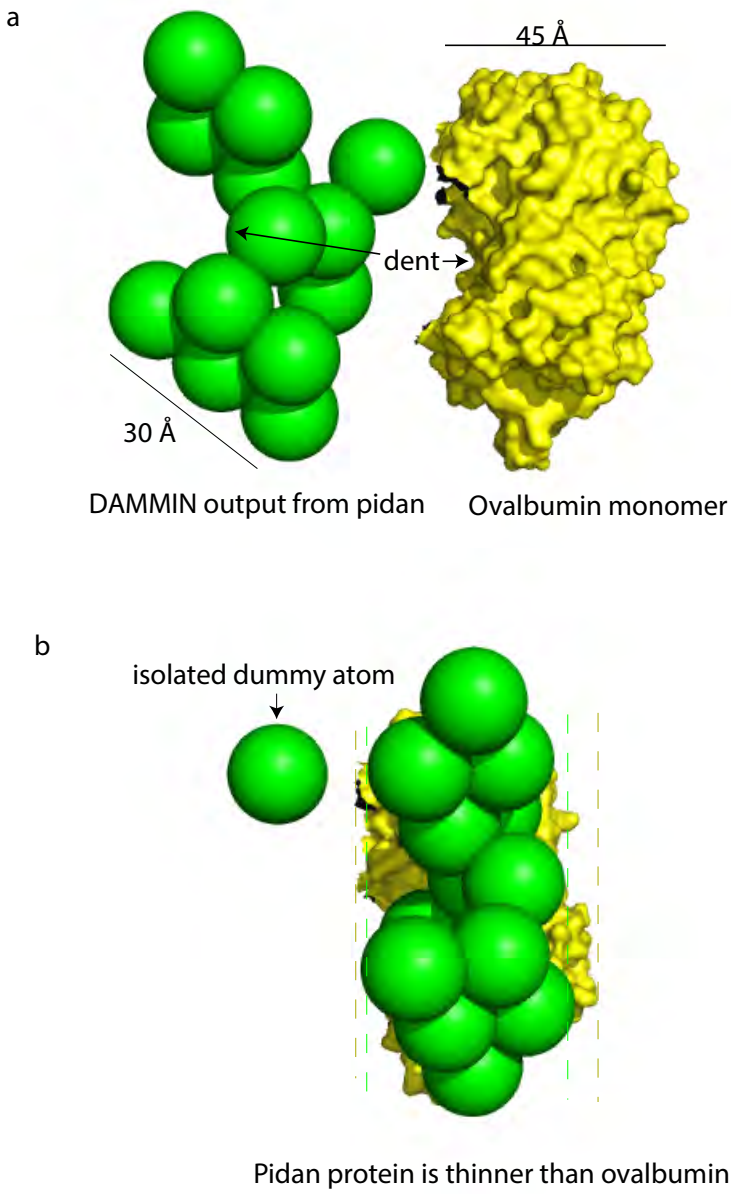


Figure 5.8: DAMMIN output from pidan. DAMMIN outputs a likely spatial configuration from pidan (green spheres). This configuration is very similar to an ovalbumin monomer (yellow). Different view angles are shown in (a) and (b). In (b), there is some structure sticking out from the plane of the other ovalbumin. This structural features may be the linkage from one ovalbumin to another.

pickled quail egg as the following: During the pickling process, quail egg proteins are locally unfolded, with most of them still in shape and recognizable. The unfolded part links to other partially unfolded proteins and forms a gel. This is consistent with the mechanism proposed by Eiser and colleagues [43]. DAMMIN result shows that the configuration of proteins in pidan is very similar to the ovalbumin monomer, except that the gelled protein is narrower and has a dummy atom protruding from the structure, possibly indicative of a local unfolding event. This partially unfolded structure provides an insight into the mechanism of how egg white proteins are linked together. The local unfolding of a loop may serve as a sticky spot which can link to another protein. Comparing DAMMIN and ovalbumin structure, the unfolded tail in DAMMIN output is likely to come from amino acids 87-95. Among these 24 amino acids, 12 are either charged or polarized. So one possible mechanism of the linkage between the pidan proteins is that these charged or polarized amino acids are likely to have attractive interactions to link to other partially unfolded proteins. Further, the GNOM result shows a shift of the second peak from 95 Å for raw egg white to 125 Å for pidan. This suggests that the increase of the average distance between subunits of the protein assembly is probably due to the requirement for the egg proteins to link to each other and form a gel. It is likely that the protein packing fraction is so low that the average subunit distance has to be increased in order to allow most of the proteins to reach out to other proteins and form the links. This result is consistent with the DAMMIN result that proteins are partially unfolded. The unfolded part stretches out, and results in an increase of the average distance between the proteins.

Further study on the role of amino acids 87-95 in the ovalbumin structure shows that the majority of amino acid (87B to 87E) in this region form a helix, which is labelled as helix C2 in the original crystallographic study describing ovalbumin structure [109]. The main-chain temperature factor for each ovalbumin molecule was obtained from x-ray crystallography of ovalbumin, in which ovalbumin crystallized as a tetramer [109]. There

is a peak at residue 87A to 91 for ovalbumin subunits A, C, and D (data for ovalbumin B is missing). This peak for ovalbumin subunits C and D is the highest peak among all residues. A peak in temperature factor means that the structure is more flexible. Therefore, I hypothesize that under strong base treatment, this helix may swing freely with one end attaching to the main body of the ovalbumin, and the rest dangling around until it links to another ovalbumin. This could be the mechanism of how the egg white proteins link together and form a gel.

Specifically, the well-folded ovalbumin shows that the helix C2 is bound to helix E (residue 128 to 137) through a disulfide bond between two cysteines located at residue 87H and 133, as is shown in Figure 5.9. The strong base treatment is very likely to break this disulfide bond, resulting in helix C2 detaching from helix E. In fact, the first-order rate constant of the cleavage of the disulfide bond in ovalbumin is 112 min^{-1} in pH $=$ 13 environment (0.2 M NaOH)[48], meaning that 63% of the disulfide bonds are broken within 0.5 second. This disulfide bond breaking would be crucial to change helix C2 to an unfolded structure extending from the body of the protein, which may eventually attach to other proteins in the egg white. I hypothesize that the breaking of this disulfide bond is the key for a raw egg to gel and transition to pidan.

This mechanism is consistent with my *in vitro* experiments studying the gelation process with ovalbumin. For the case of base treatment, in low pH environment, the ovalbumin solution is more viscous, suggesting adding NaOH increases the attractive interaction between the proteins. This interaction may be from the detached helix C2 of ovalbumin forming linkages between other ovalbumin molecules. At pH less than 10.2, perhaps there are not enough broken disulfide bonds to completely gel the system. With an increase of the pH sufficient to break most of the disulfide bonds in the system, more and more ovalbumin molecules have helix C2 detached and free to interact with other monomers. This is a possible mechanism for gel formation in this system. When the pH is very high, it is likely that

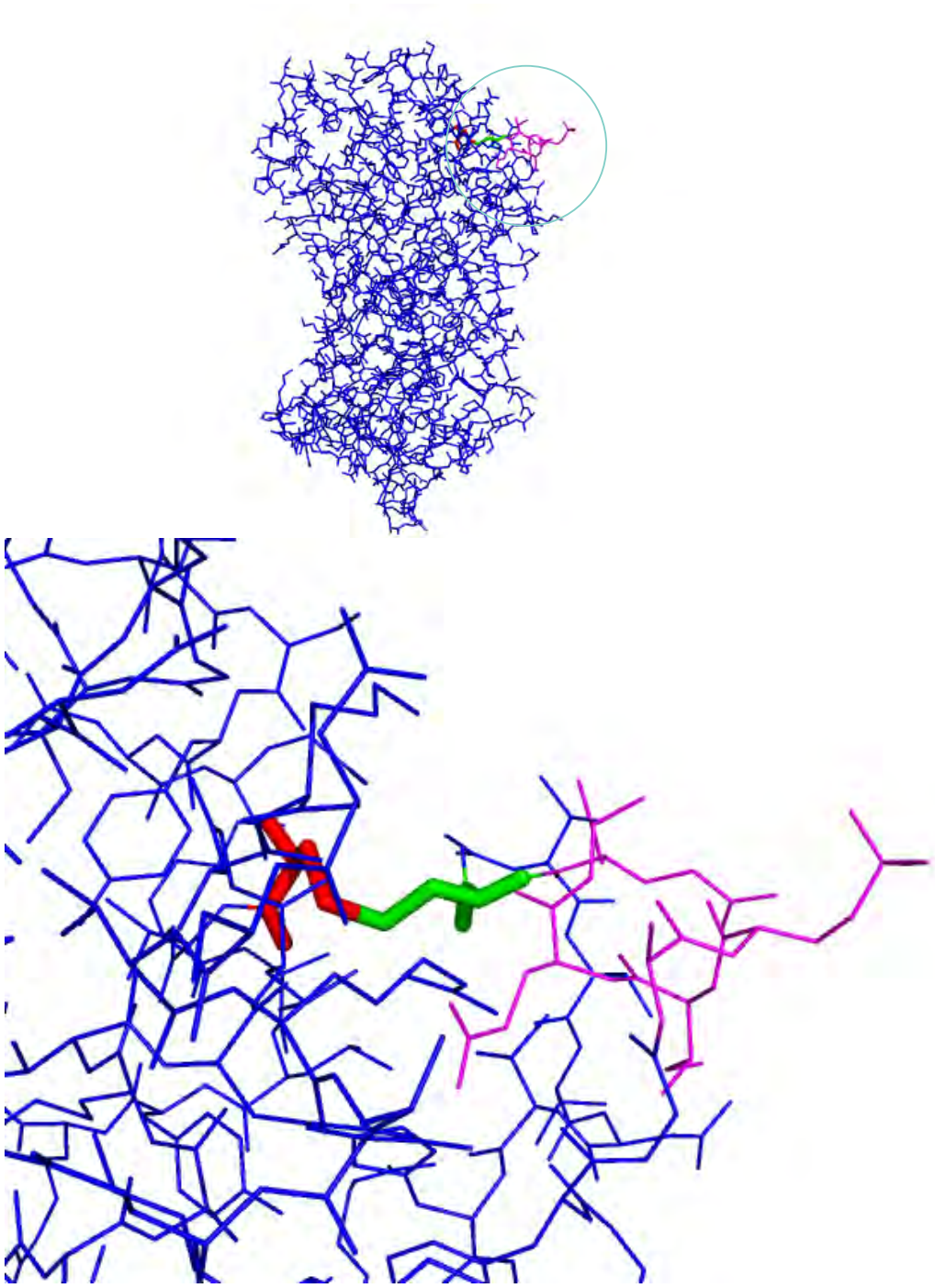


Figure 5.9: Disulfide bond. The disulfide bond between residue 87H (green) to 133 (red) are shown. Our DAMMIN result shows that in the process of pidan gelation, this bond is likely to break, resulting in helix C2 (pink) to stretch out and serving as the patches to link to other ovalbumin.

the proteins are so highly charged that the attractive interaction is not strong enough to link molecules at low packing fraction. My *in vitro* experiment of adding HCl to ovalbumin does not result in a transparent gel. This indicates that under acid treatment, disulfide bonds may not be broken, but unfolding or new interactions are introduced through different mechanisms. Under acidic conditions, each individual ovalbumin molecule may become a patchy particle with a high coordinate number, leading to liquid-liquid phase separation.

5.7 Pidan and squid lens as patchy particles

The pidan gel is very similar to the outer layer of a squid lens in that they are both transparent gels with low protein packing fractions ($\sim 5\%$), and formed by globular proteins. Based on my hypothesis, they are both linked up by disordered amino acid chains in the protein structures. In this section, I will explore the similarity between both systems and answer how they keep transparency at such a low packing fraction of the protein with no protein turnover. Patchy particle theory helps us to understand this phenomenon. Sciortino and colleagues proposed that at low packing fraction, gelation can happen without liquid-liquid phase separation if the average coordinate number of the patchy particles are low. I hypothesize that this gelation at low density due to low-valency patchy particles occurs in both pidan and the squid lens. The structure of proteins in pidan suggests that disorder induced in helix C2 allows it to extend from the folded body of the protein and link to another protein. In addition, pickling quail egg at high pH makes the proteins highly negatively charged, and the screened Coulomb repulsion possibly prevents the proteins from randomly aggregating, making a highly polar, partially unfolded segment the only sticky spot that can link to other proteins. Therefore, the pickling process on the quail egg may make the proteins low valency patchy particles with a patch-antipatch arrangement. Similarly, in the periphery of the squid lens, DAMMIF predicts the average coordinate

number to be $< M >= 2.19$, which is also low enough to avoid liquid-liquid phase separation. Comparing the SAXS intensities on transparent samples to the SAXS intensities on opaque samples for both the egg and the lens, the rapid decrease at small q for opaque tissues suggests that the proteins are randomly aggregated, and this aggregation results in a dramatic increase of the coordinate numbers for the system. At low packing fraction, this leads to density fluctuation at the wavelength scale of visible light so that opacity occurs.

To sum up, the underlying physics between a pickled quail egg white and the outer-most layer of a squid lens seem to be very similar. In both cases, the proteins may have one or two disordered loop-like structures which can freely swing in a relatively wide angle range. For the case of pidan, we hypothesize that this structure comes from breaking the disulfide bond between helix C2 and helix E. Breaking this bond and forming a disordered loop may be induced in pidan by the high base treatment traditionally used to preserve the eggs and induce gel formation. In the lens system, however, the disordered loop is not achieved by artificial treatment, but has evolved in concert with the ability to make a gradient index lens. In squid lens, and also possibly in pidan, the disordered loops can link to other proteins, and serve as the attractive patches in a system of proteins as patchy colloids. Therefore, these proteins can be treated as patchy particles and the systems are transparent because the average coordinate number is low for both cases, resulting in a stable gelled phase without phase separation.

Chapter 6

Summary and future directions

In this thesis, I studied the physical and material properties of a squid lens, and used patchy particle theory to propose an explanation for how the lens maintains transparency and thermodynamic stability while the protein density spans a huge range throughout the lens radius. Then, I compared the squid lens system to the preservation process of the Chinese "century egg", which is another system of patchy particles resulting from globular proteins. Here, in this chapter, I will summarize these results and discuss the possible next steps for this research.

6.1 Summary

Many species of midwater squids are common around 300 meters depth in the ocean, where the irradiance is one ten thousandth less than the irradiance on land. For vision that is simultaneously acute and sensitive, this environment requires a particular eye design for camera vision. In squids, the eye is a camera-like structure, just like our eyes. But the lens in the eye has to be spherical to both maximize the photon flux and the retina and to create an eye volume that will fit in the organism's head. From there, the lens has to have graded refractive index in order to eliminate spherical aberration. The refractive

index distribution with radius follows Maxwell's fish lens equation, which has been verified in squid by our oil-matching experiment described in Chapter 2. In squid, the graded refractive index is achieved by having a graded protein concentration. Specifically, our theoretical estimation of the protein concentration results in about 5% in the periphery of the lens, whereas it is about 100% in the core. Meanwhile, transparency requires that spatial fluctuations in protein density are minimized throughout the whole span of the protein concentration. The squid lens achieved all of these using only one type of protein —- the S-crystallins. Our SDS-PAGE and RNA sequencing results show that more than 95% of the lens proteins are S-crystallins at any given positions in the lens. The squid lens is especially interesting because this protein composition is different from that in vertebrate lenses, which are composed of three types of proteins with various molecular weights, abundances, and interactions. In a squid lens, there are dozens of S-crystallin isoforms, but the molecular weight, shape and the protein secondary structure are very similar between these isoforms. In this thesis, I am studying how variations on a single protein fold may assemble into a graded refractive index lens with low density fluctuation at every possible density of protein.

A previous study examined the surface charge of the S-crystallin isoforms using several simulation methods [114]. The results showed that the S-crystallins in the periphery of the lens have higher positive surface charge compared to the ones in the core. This result is supported by my IEF PAGE experiment which shows that the proteins in the lens are highly charged. Therefore, the old hypothesis for the assembly of the S-crystallins in the lens was that the S-crystallins experience repulsive interactions in the peripheral layer to maintain transparency, and the phase of the lens is a repulsive glass.

However, this hypothesis is not supported by my SAXS results. The scattering intensities show that the S-crystallins in an intact lens have attractive interactions between them! This result seems paradoxical for a living, transparent lens: there is always liquid-liquid

phase separation at low packing fraction in an assembly with particles that have isotropic attractive interactions [101]. This liquid-liquid phase separation seems completely inconsistent with a living, transparent lens because the boundaries between these two phases scatter light, and would almost certainly result in opacity in the lens. How then, can the squid lens proteins avoid this deleterious liquid-liquid phase separation? It turned out that the patchy particle theory provides us with a recipe to build a graded refractive index lens from variations on a single type of particle.

The phase diagram of a patchy particle system shows that the system may be gelled at any packing fraction if the average coordinate number is low [12]. A gelled phase could be compatible with a living, transparent lens because not only there is no phase separation, but the thermal fluctuation in a gel is reduced as well. In our lens system, we have an estimation of the packing fractions, the protein-protein interactions, and the coordinate numbers at all layers. All of these map well to predictions from the patchy particle phase diagram. Our measurements suggest that the average coordinate number of the proteins in the peripheral two layers are low enough so that the phase of these layers can be gelled while avoiding liquid-liquid phase separation. Therefore, our results suggest that squid lens maintains transparency by preferentially expressing low coordinate number proteins in the periphery of the lens, and in this way, the whole lens can possibly be gelled. This finding of a gelled phase is also consistent with the SAXS, native PAGE, filtration, and TEM results.

How do the proteins combine together and form this connected large-scale structure? Our SDS PAGE result shows that the majority of proteins in the lens occupy two molecular weight classes, with average molecular weights of 24.5 kDa and 26.5 kDa, with two minor bands at 32 and 36 kDa. Our homology modeling suggests that the 26.5 kDa S-crystallins have longer disordered regions protruding from the folded body of the protein compared the the 24.5 kDa S-crystallins. The peripheral layer of the lens contains more long-loop proteins compared to the core. The DAMMIF result suggests that the long-loop proteins

can link to each other through the loops. Our GROMACS result shows that there are three or four hydrogen bonds formed between the two loops in long-loop proteins. Thus, the long-loop proteins are very likely to be the patchy particles with low coordinate number, and the higher abundance of this molecular weight class results in a low average coordinate number in the periphery of the lens.

In short, I propose that the mechanism of how a squid lens maintains transparency is to preferentially express the low coordinate number proteins in the periphery of the lens. The state of the lens is very likely to be gelled, even at low packing fraction in the periphery, and at all radial positions irrespective of density.

I also studied the Chinese century egg (pidan), which is a system similar to the squid lens. Pidan are eggs that are preserved with strong base in clay using a very old traditional method. After a few months in strong base, the egg white changes into a transparent gel. This gel is very qualitatively similar to the peripheral layers of the squid lens. I performed SAXS on pidan to study this gelled network and to compare this biomaterial to the squid lens system.

I found that the underlying physics between pidan and the peripheral layer of a squid lens are very similar. For both systems, the proteins which comprise the systems are globular proteins. In the squid lens, and also possibly in pidan, these proteins have disordered regions of flexible chains of amino acids that may bind to other proteins in the system. In this way, both the pidan and the squid lens form a network of proteins that seem to be gelled. This network system has low $< M >$ at low density, which in both systems is correlated in a stable gelled phase without liquid-liquid phase separation.

The detailed mechanism of the flexible, disordered amino acid chain (the patch) for both cases are discussed: Ovalbumins in pidan are likely to have a broken disulfide bond, resulting in one end of helix C2 available to combine with other proteins. The S-crystallins in a squid lens achieve this via the disordered loops encoded in a novel exon. Therefore,

both ovalbumins and S-crystallins are patchy particles, with the "patch" physically manifested in a disordered, polar loop protruding from the surface of a folded protein. The systems are maintained transparent because $< M >$ is low for both cases.

6.2 Significance of this work

In this thesis, I described my research on the self-assembly routes evolved in a squid lens. Not only is this study interesting itself since it is a mechanistic description of a spherical lens that has no spherical aberration and with a large span of density, there is some other significance of this work, as discussed below.

6.2.1 First biological sample of evolved proteins as patchy particles

The evolutionary process leading to a graded refractive index lens may follow the path to select low coordinate number proteins and place them in the periphery of the lens. Our previous work [114] studied the gene tree of different S-crystallin isoforms in the lens. Long-loop proteins emerged later and duplicated under positive selection. This process is occurred in correlation with the sophisticated, graded refractive index lens in evolutionary time.

Since long-loop proteins have lower M, these proteins are required at the periphery of the lens to keep the periphery both transparent and at a low packing fraction ($\sim 5\%$). Therefore, once these proteins emerged, they should be positively selected and expressed in the periphery, and this is what we have observed. In other words, the driving force of evolution is very likely to be selecting for S-crystallins with continuously lower coordinate number, and expressing them at the appropriate position in the lens to achieve a gradient index. To our knowledge, the squid lens system is the first biological system in which

evolution has "discovered" the physical properties of patchy particles and used this physical toolbox directly to form a structure.

6.2.2 First protein gel system found at all packing fraction

Theory predicts that patchy particle systems can reach a gelled state at all packing fractions by decreasing the average coordinate number of the system [12]. Experiments have been designed to realize the gelled phase at very low packing fractions, but to date, the experimental materials are mostly inorganic [99], or engineered DNA [44]. To our knowledge, the squid lens system is the first system in nature that deliberately uses proteins as the patchy particles to achieve gelled states at low packing fraction. What is more, unlike microtubules and microfilaments, which are very monodisperse, and are near-crystalline when assembled, S-crystallins assemble into a network that is amorphous. The S-crystallins are evolved to assemble into a low-density-fluctuation protein network, but each network in each individual cell is structurally different from the neighboring cell in the sense that each combination of protein linkages will vary. Only the statistical properties of the network remain, including the density, the density fluctuation, $< M >$, *etc.*. We have not seen any other work with proteins as patchy particles that achieve uniform, average properties in a network, rather than specific, designed patterns.

6.2.3 How a squid lens assembles

Last but not the least, a mechanism on how the squid lens is maintained transparency at a large span of packing fraction is proposed. I studied the lens system experimentally. The results support our proposed mechanism. This mechanism is very different from vertebrate animal lenses, and it may provide with an insight to fabricating an artificial lens with graded refractive index using patchy particles.

6.3 Future work

6.3.1 The development of the squid lens and water movement

All of our experiments were performed on mature squid lens, but the physical characteristics of the lens during development have never been studied. It is not known whether the gradient density is always present and must grow with the squid even as new cellular layers are added, or whether the core of the larval lens is high in density and each subsequent new layer decreases in density. In the scenario where the gradient must grow with the lens during development, each newly emerged layer may have different osmotic pressure. This pressure may potentially cause water movement, resulting in a higher protein density in the core and providing a mechanism to scale up the gradient in size while maintaining its properties during growth. Studying the osmotic pressure, density and protein distribution during lens development will potentially answer how the lens achieves such a high packing fraction (100%) at the core of the lens.

6.3.2 Artificial self-assembled lens with graded refractive index

Knowing the assembly mechanism of the squid lens, it will be interesting and potentially useful in industry to create an artificial graded refractive index lens by self-assembly. Further work will focus on the design of patchy nanoparticles with different M, and possible mechanisms for distributing the particles into different radial positions in an artificial lens. Different coordinate number patchy particles with the designed density will be placed into some segments, as the first step of our biomimetic system from the squid lens. In this system, the density will be a given and is not a self-assembly result, as it seems to be in the living lens. Next, after studying the osmotic pressure, the system will be designed with

the density as a self-assembly result as well. Thus, we may be able to fabricate a totally self-assembled system with only the protein composition as an input.

Bibliography

[1] Pinar Akcora, Hongjun Liu, Sanat K. Kumar, Joseph Moll, Yu Li, Brian C. Benicewicz, Linda S. Schadler, Devrim Acehan, Athanassios Z. Panagiotopoulos, Victor Pryamitsyn, Venkat Ganesan, Jan Ilavsky, Pappanan Thiyagarajan, Ralph H. Colby, and Jack F. Douglas. Anisotropic self-assembly of spherical polymer-grafted nanoparticles. *NATURE MATERIALS*, 8(4):354–U121, APR 2009.

[2] Anna Akhmanova and Michel O. Steinmetz. Tracking the ends: a dynamic protein network controls the fate of microtubule tips. *NATURE REVIEWS MOLECULAR CELL BIOLOGY*, 9(4):309–322, APR 2008.

[3] AP Alivisatos. Semiconductor clusters, nanocrystals, and quantum dots. *SCIENCE*, 271(5251):933–937, FEB 16 1996.

[4] DB Ameen, MF Bishop, and T McMullen. Theory of transparency of the eye using photonic band structure. *BIOPHYSICAL JOURNAL*, 70(2, 2):WP309, FEB 1996.

[5] Usha P. Andley. Crystallins in the eye: Function and pathology. *PROGRESS IN RETINAL AND EYE RESEARCH*, 26(1):78–98, JAN 2007.

[6] TH Bayburt and SG Sligar. Self-assembly of single integral membrane proteins into

soluble nanoscale phospholipid bilayers. *PROTEIN SCIENCE*, 12(11):2476–2481, NOV 2003.

[7] G Beaucage. Small-angle scattering from polymeric mass fractals of arbitrary mass-fractal dimension. *JOURNAL OF APPLIED CRYSTALLOGRAPHY*, 29(2):134–146, APR 1 1996.

[8] GAM BERBERS, WA HOEKMAN, H BLOEMENDAL, WW DEJONG, T KLEINSCHMIDT, and G BRAUNITZER. HOMOLOGY BETWEEN THE PRIMARY STRUCTURES OF THE MAJOR BOVINE BETA-CRYSTALLIN CHAINS. *EUROPEAN JOURNAL OF BIOCHEMISTRY*, 139(3):467–479, 1984.

[9] F A Bettelheim and E L Siew. Effect of change in concentration upon lens turbidity as predicted by the random fluctuation theory. *BIOPHYSICAL JOURNAL*, 41(1):29 – 33, 1983.

[10] FA BETTELHEIM and EL SIEW. EFFECT OF CHANGE IN CONCENTRATION UPON LENS TURBIDITY AS PREDICTED BY THE RANDOM FLUCTUATION THEORY. *BIOPHYSICAL JOURNAL*, 41(1):29–33, 1983.

[11] SP BHAT and CN NAGINENI. ALPHA-B SUBUNIT OF LENS-SPECIFIC PROTEIN ALPHA-CRYSTALLIN IS PRESENT IN OTHER OCULAR AND NON-OCULAR TISSUES. *BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS*, 158(1):319–325, JAN 16 1989.

[12] Emanuela Bianchi, Julio Largo, Piero Tartaglia, Emanuela Zaccarelli, and Francesco Sciortino. Phase diagram of patchy colloids: Towards empty liquids. *PHYSICAL REVIEW LETTERS*, 97(16), OCT 20 2006.

[13] Silvia Biffi, Roberto Cerbino, Francesca Bomboi, Elvezia Maria Paraboschi, Rosanna Asselta, Francesco Sciortino, and Tommaso Bellini. Phase behavior and

critical activated dynamics of limited-valence DNA nanostars. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 110(39):15633–15637, SEP 24 2013.

[14] B BJELLQVIST, GJ HUGHES, C PASQUALI, N PAQUET, F RAVIER, JC SANCHEZ, S FRUTIGER, and D HOCHSTRASSER. THE FOCUSING POSITIONS OF POLYPEPTIDES IN IMMOBILIZED PH GRADIENTS CAN BE PREDICTED FROM THEIR AMINO-ACID-SEQUENCES. *ELECTROPHORESIS*, 14 (10):1023–1031, OCT 1993.

[15] B BJELLQVIST, B BASSE, E OLSEN, and JE CELIS. REFERENCE POINTS FOR COMPARISONS OF 2-DIMENSIONAL MAPS OF PROTEINS FROM DIFFERENT HUMAN CELL-TYPES DEFINED IN A PH SCALE WHERE ISOELECTRIC POINTS CORRELATE WITH POLYPEPTIDE COMPOSITIONS. *ELECTROPHORESIS*, 15(3-4):529–539, MAR-APR 1994.

[16] H BLOEMENDAL and WW DEJONG. LENS PROTEINS AND THEIR GENES. *PROGRESS IN NUCLEIC ACID RESEARCH AND MOLECULAR BIOLOGY*, 41: 259–281, 1991.

[17] H Bloemendal, W de Jong, R Jaenicke, NH Lubsen, C Slingsby, and A Tardieu. Ageing and vision: structure, stability and function of lens crystallins. *PROGRESS IN BIOPHYSICS & MOLECULAR BIOLOGY*, 86(3):407–485, NOV 2004.

[18] Robert Bogwe. Self-assembly: a review of recent developments. *ASSEMBLY AUTOMATION*, 28(3):211–215, 2008.

[19] Samir K. Brahma and Massimo Lancieri. Isofocusing and immunological investigations on cephalopod lens proteins. *Experimental Eye Research*, 29(6):671 – 678, 1979.

[20] ML BROIDE, CR BERLAND, J PANDE, OO OGUN, and GB BENEDEK. BINARY-LIQUID PHASE-SEPARATION OF LENS PROTEIN SOLUTIONS. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 88(13):5660–5664, JUL 1991.

[21] Karen Hopkin Alexander D Johnson Alexander Johnson Julian Lewis Martin Raff Keith Roberts Peter Walter Bruce Alberts, Dennis Bray. Essential cell biology. *Garland Science*.

[22] WA BUSCH, A SUZUKI, MH STROMER, and DE GOLL. CA2+-SPECIFIC REMOVAL OF Z LINES FROM RABBIT SKELETAL-MUSCLE. *JOURNAL OF CELL BIOLOGY*, 52(2):367–&, 1972.

[23] Maxwell J C. Solutions of problems. *Camb Dublin Math J*, 8:188, 1854.

[24] PJ Camp. Phase diagrams of hard spheres with algebraic attractive interactions. *PHYSICAL REVIEW E*, 67(1, 1), JAN 2003.

[25] AI Campbell, VJ Anderson, JS van Duijneveldt, and P Bartlett. Dynamical arrest in attractive colloids: The effect of long-range repulsion. *PHYSICAL REVIEW LETTERS*, 94(20), MAY 27 2005.

[26] Jun Chen, Angang Dong, Jing Cai, Xingchen Ye, Yijin Kang, James M. Kikkawa, and Christopher B. Murray. Collective Dipolar Interactions in Self-Assembled Magnetic Binary Nanocrystal Superlattice Membranes. *NANO LETTERS*, 10(12):5103–5108, DEC 2010.

[27] Qian Chen, Sung Chul Bae, and Steve Granick. Directed self-assembly of a colloidal kagome lattice. *NATURE*, 469(7330):381–384, JAN 20 2011.

[28] YH CHEN, JT YANG, and KH CHAU. DETERMINATION OF HELIX AND BETA-FORM OF PROTEINS IN AQUEOUS-SOLUTION BY CIRCULAR-DICHROISM. *BIOCHEMISTRY*, 13(16):3350–3359, 1974.

[29] Margaret S. Cheung. Where soft matter meets living matter - protein structure, stability, and folding in the cell. *CURRENT OPINION IN STRUCTURAL BIOLOGY*, 23(2):212–217, APR 2013.

[30] Ekta Seth Chhabra and Henry N. Higgs. The many faces of actin: Matching assembly factors with cellular structures. *NATURE CELL BIOLOGY*, 9(10):1110–1121, OCT 2007.

[31] CC Chuang, SH Wu, SH Chiou, and GG Chang. Homology modeling of cephalopod lens S-crystallin: A natural mutant of sigma-class glutathione transferase with diminished endogenous activity. *BIOPHYSICAL JOURNAL*, 76(2):679–690, FEB 1999.

[32] Silvia Corezzi, Cristiano De Michele, Emanuela Zaccarelli, Daniele Fioretto, and Francesco Sciortino. A molecular dynamics study of chemical gelation in a patchy particle model. *SOFT MATTER*, 4(6):1173–1177, 2008.

[33] P. Debye. Zerstreuung von Rontgenstrahlen. *Annalen der Physik*, 301:809–823, 1915.

[34] Wayne A Decatur and Maurille J Fournier. rrna modifications and ribosome function. *Trends in Biochemical Sciences*, 27(7):344 – 351, 2002.

[35] A Desai and TJ Mitchison. Microtubule polymerization dynamics. *ANNUAL REVIEW OF CELL AND DEVELOPMENTAL BIOLOGY*, 13:83–117, 1997.

[36] A Disanza, A Steffen, M Hertzog, E Frittoli, K Rottner, and G Scita. Actin polymerization machinery: the finish line of signaling networks, the starting point of cellular movement. *CELLULAR AND MOLECULAR LIFE SCIENCES*, 62(9):955–970, MAY 2005.

[37] CM Dobson. Protein folding and misfolding. *NATURE*, 426(6968):884–890, DEC 18 2003.

[38] C Donati, SC Glotzer, PH Poole, W Kob, and SJ Plimpton. Spatial correlations of mobility and immobility in a glass-forming Lennard-Jones liquid. *PHYSICAL REVIEW E*, 60(3):3107–3119, SEP 1999.

[39] Jianzhong Du and Rachel K. O'Reilly. Anisotropic particles with patchy, multicompartment and Janus architectures: preparation and application. *CHEMICAL SOCIETY REVIEWS*, 40(5):2402–2416, 2011.

[40] Fabien Dubois, Benoit Mahler, Benoit Dubertret, Eric Doris, and Charles Mioskowski. A versatile strategy for quantum dot ligand exchange. *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, 129(3):482–483, JAN 24 2007.

[41] A Ducruix, JP Guilloteau, M RiesKautt, and A Tardieu. Protein interactions as seen by solution X-ray scattering prior to crystallogenesis. *JOURNAL OF CRYSTAL GROWTH*, 168(1-4):28–39, OCT 1996. 6th International Conference on Crystallization of Biological Macromolecules, HIROSHIMA, JAPAN, NOV 12-17, 1995.

[42] B EGELANDSDAL. HEAT-INDUCED GELLING IN SOLUTIONS OF OVALBUMIN. *JOURNAL OF FOOD SCIENCE*, 45(3):570–&, 1980.

[43] Erika Eiser, Caroline S. Miles, Nienke Geerts, Peter Verschuren, and Cait E. MacPhee. Molecular cooking: physical transformations in Chinese 'century' eggs. *SOFT MATTER*, 5(14):2725–2730, 2009.

[44] Lang Feng, Remi Dreyfus, Ruojie Sha, Nadrian C. Seeman, and Paul M. Chaikin. DNA Patchy Particles. *ADVANCED MATERIALS*, 25(20):2779–2783, MAY 28 2013.

[45] Lin Feng, Yanan Zhang, Jinming Xi, Ying Zhu, Nue Wang, Fan Xia, and Lei Jiang. Petal effect: A superhydrophobic state with high adhesive force. *LANGMUIR*, 24 (8):4114–4119, APR 15 2008.

[46] S Finet and A Tardieu. alpha-crystallin interaction forces studied by small angle X-ray scattering and numerical simulations. *JOURNAL OF CRYSTAL GROWTH*, 232 (1-4):40–49, NOV 2001. 8th International Conference on Crystallization of Biological Macromolecules, SANDESTIN, FLORIDA, MAY 14-19, 2000.

[47] DA FISCHMAN. AN ELECTRON MICROSCOPE STUDY OF MYOFIBRIL FORMATION IN EMBRYONIC CHICK SKELETAL MUSCLE. *JOURNAL OF CELL BIOLOGY*, 32(3):557–&, 1967.

[48] TM FLORENCE. DEGRADATION OF PROTEIN DISULFIDE BONDS IN DILUTE ALKALI. *BIOCHEMICAL JOURNAL*, 189(3):507–520, 1980.

[49] Daniel Franke and Dmitri I. Svergun. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *JOURNAL OF APPLIED CRYSTALLOGRAPHY*, 42:342–346, APR 2009.

[50] Gattiker A. Duvaud S. Wilkins M.R. Appel R.D. Bairoch A. Gasteiger E., Hoogland C. Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook*, 2005.

[51] P GETTINS, PA PATSTON, and M SCHAPIRA. THE ROLE OF CONFORMATIONAL CHANGE IN SERPIN STRUCTURE AND FUNCTION. *BIOESSAYS*, 15(7):461–467, JUL 1993.

[52] Sharon C. Glotzer and Michael J. Solomon. Anisotropy of building blocks and their assembly into complex structures. *NATURE MATERIALS*, 6(8):557–562, AUG 2007.

[53] P. Herbrink, Hanske van Westreenen, and H. Bloemendal. Further studies on the polypeptide chains of β-crystallin. *Experimental Eye Research*, 20(6):541 – 548, 1975.

[54] J HILL and DJ COX. SEDIMENTATION EQUILIBRIUM OF OVALBUMIN IN CONCENTRATED CESIUM CHLORIDE. *JOURNAL OF PHYSICAL CHEMISTRY*, 70(9):2946–&, 1966.

[55] WGJ HOL, LM HALIE, and C SANDER. DIPOLES OF THE ALPHA-HELIX AND BETA-SHEET - THEIR ROLE IN PROTEIN FOLDING. *NATURE*, 294 (5841):532–536, 1981.

[56] J HORWITZ. ALPHA-CRYSTALLIN CAN FUNCTION AS A MOLECULAR CHAPERONE. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 89(21):10449–10453, NOV 1 1992.

[57] P Hotulainen and P Lappalainen. Stress fibers are generated by two distinct actin assembly mechanisms in motile cells. *JOURNAL OF CELL BIOLOGY*, 173(3): 383–394, MAY 8 2006.

[58] J Howard and AA Hyman. Dynamics and mechanics of the microtubule plus end. *NATURE*, 422(6933):753–758, APR 17 2003.

[59] JN ISRAELACHVILI, DJ MITCHELL, and BW NINHAM. THEORY OF SELF-ASSEMBLY OF HYDROCARBON AMPHIPHILES INTO MICELLES AND BILAYERS. *JOURNAL OF THE CHEMICAL SOCIETY-FARADAY TRANSACTIONS II*, 72:1525–1568, 1976.

[60] V Jayawarna, M Ali, TA Jowitt, AE Miller, A Saiani, JE Gough, and RV Ulijn. Nanostructured hydrogels for three-dimensional cell culture through self-assembly of fluorenylmethoxycarbonyl-dipeptides. *ADVANCED MATERIALS*, 18(5):611+, MAR 3 2006.

[61] XH JI, EC VONROSENVINGE, WW JOHNSON, SI TOMAREV, J PIATIGORSKY, RN ARMSTRONG, and GL GILLILAND. 3-DIMENSIONAL STRUCTURE, CATALYTIC PROPERTIES, AND EVOLUTION OF A SIGMA-CLASS GLUTATHIONE TRANSFERASE FROM SQUID, A PROGENITOR OF THE LENS S-CRYSTALLINS OF CEPHALOPODS. *BIOCHEMISTRY*, 34(16):5317–5328, APR 25 1995.

[62] M KIRSCHNER and T MITCHISON. BEYOND SELF-ASSEMBLY - FROM MICROTUBULES TO MORPHOGENESIS. *CELL*, 45(3):329–342, MAY 9 1986.

[63] Mathias Kolle, Pedro M. Salgard-Cunha, Maik R. J. Scherer, Fumin Huang, Pete Vukusic, Sumeet Mahajan, Jeremy J. Baumberg, and Ullrich Steiner. Mimicking the colourful wing scale structure of the Papilio blumei butterfly. *NATURE NANOTECHNOLOGY*, 5(7):511–515, JUL 2010.

[64] EV KOONIN, AR MUSHEGIAN, RL TATUSOV, SF ALTSCHUL, SH BRYANT, P BORK, and A VALENCIA. EUKARYOTIC TRANSLATION ELONGATION-FACTOR 1-GAMMA CONTAINS A GLUTATHIONE TRANSFERASE DOMAIN - STUDY OF A DIVERSE, ANCIENT PROTEIN SUPERFAMILY USING MOTIF SEARCH AND STRUCTURAL MODELING. *PROTEIN SCIENCE*, 3(11):2045–2054, NOV 1994.

[65] KJ Lampi, ZX Ma, M Shih, TR Shearer, JB Smith, DL Smith, and LL David. Sequence analysis of beta A3, beta B3, and beta A4 crystallins completes the identi-

fication of the major proteins in young human lens. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 272(4):2268–2275, JAN 24 1997.

[66] DA Lauffenburger and AF Horwitz. Cell migration: A physically integrated molecular process. *CELL*, 84(3):359–369, FEB 9 1996.

[67] AC LEGON and DJ MILLEN. ANGULAR GEOMETRIES AND OTHER PROPERTIES OF HYDROGEN-BONDED DIMERS - A SIMPLE ELECTROSTATIC INTERPRETATION OF THE SUCCESS OF THE ELECTRON-PAIR MODEL. *CHEMICAL SOCIETY REVIEWS*, 16(4):467–498, DEC 1987.

[68] YG Li, YD Tseng, SY Kwon, L D'Espaux, JS Bunch, PL Mceuen, and D Luo. Controlled assembly of dendrimer-like DNA. *NATURE MATERIALS*, 3(1):38–42, JAN 2004.

[69] JS LINDSEY. SELF-ASSEMBLY IN SYNTHETIC ROUTES TO MOLECULAR DEVICES - BIOLOGICAL PRINCIPLES AND CHEMICAL PERSPECTIVES - A REVIEW. *NEW JOURNAL OF CHEMISTRY*, 15(2-3):153–180, FEB-MAR 1991.

[70] M Malfois, F Bonnete, L Belloni, and A Tardieu. A model of attractive interactions to account for fluid-fluid phase separation of protein solutions. *JOURNAL OF CHEMICAL PHYSICS*, 105(8):3290–3300, AUG 22 1996.

[71] A Mallavarapu and T Mitchison. Regulated actin cytoskeleton assembly at filopodium tips controls their extension and retraction. *JOURNAL OF CELL BIOLOGY*, 146(5):1097–1106, SEP 6 1999.

[72] Pieta K. Mattila and Pekka Lappalainen. Filopodia: molecular architecture and cellular functions. *NATURE REVIEWS MOLECULAR CELL BIOLOGY*, 9(6):446–454, JUN 2008.

[73] Emil Wolf Max Born. Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light. *Cambridge University Press; 7th edition*, book:page 157, 1999.

[74] TL MCMEEKIN, ML GROVES, and M WILENSKY. REFRACTIVE INDICES OF PROTEINS IN RELATION TO AMINO ACID COMPOSITION AND SPECIFIC VOLUME. *BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS*, 7(2):151, 1962.

[75] D-E Nilsson M.F.Land. Animal Eyes. *Oxford animal biology series*, book:page 51, 2004.

[76] X Michalet, FF Pinaud, LA Bentolila, JM Tsay, S Doose, JJ Li, G Sundaresan, AM Wu, SS Gambhir, and S Weiss. Quantum dots for live cells, in vivo imaging, and diagnostics. *SCIENCE*, 307(5709):538–544, JAN 28 2005.

[77] M Miguel, MA Manso, R Lopez-Fandino, and M Ramos. Comparative study of egg white proteins from different species by chromatographic and electrophoretic methods. *EUROPEAN FOOD RESEARCH AND TECHNOLOGY*, 221(3-4):542–546, AUG 2005.

[78] Y MINE. RECENT ADVANCES IN THE UNDERSTANDING OF EGG-WHITE PROTEIN FUNCTIONALITY. *TRENDS IN FOOD SCIENCE & TECHNOLOGY*, 6(7):225–232, JUL 1995.

[79] Amir Y. Mirarefi, Sebastien Boutet, Subramanian Ramakrishnan, Andor J. Kiss, Chi-Hing C. Cheng, Arthur L. DeVries, Ian K. Robinson, and Charles F. Zukoski. Small-angle X-ray scattering studies of the intact eye lens: Effect of crystallin composition and concentration on microstructure. *BIOCHIMICA ET BIOPHYSICA ACTA-GENERAL SUBJECTS*, 1800(6):556–564, JUN 2010.

[80] Priti S. Mohanty, Anand Yethiraj, and Peter Schurtenberger. Deformable particles with anisotropic interactions: unusual field-induced structural transitions in ultrasoft ionic microgel colloids. *SOFT MATTER*, 8(42):10819–10822, 2012.

[81] John W. R. Morgan, Dwaipayan Chakrabarti, Nicolas Dorsaz, and David J. Wales. Designing a Bernal Spiral from Patchy Colloids. *ACS NANO*, 7(2):1246–1256, FEB 2013.

[82] A NICHOLLS, KA SHARP, and B HONIG. PROTEIN FOLDING AND ASSOCIATION - INSIGHTS FROM THE INTERFACIAL AND THERMODYNAMIC PROPERTIES OF HYDROCARBONS. *PROTEINS-STRUCTURE FUNCTION AND GENETICS*, 11(4):281–296, 1991.

[83] DE NILSSON and S PELGER. A PESSIMISTIC ESTIMATE OF THE TIME REQUIRED FOR AN EYE TO EVOLVE. *PROCEEDINGS OF THE ROYAL SOCIETY B-BIOLOGICAL SCIENCES*, 256(1345):53–58, APR 22 1994.

[84] M Nirmal, BO Dabbousi, MG Bawendi, JJ Macklin, JK Trautman, TD Harris, and LE Brus. Fluorescence intermittency in single cadmium selenide nanocrystals. *NATURE*, 383(6603):802–804, OCT 31 1996.

[85] E Nogales and HW Wang. Structural intermediates in microtubule assembly and disassembly: how and why? *CURRENT OPINION IN CELL BIOLOGY*, 18(2): 179–184, APR 2006.

[86] E Nogales, SG Wolf, and KH Downing. Structure of the alpha beta tubulin dimer by electron crystallography. *NATURE*, 391(6663):199–203, JAN 8 1998.

[87] Toshiro Oda, Mitsusada Iwasa, Tomoki Aihara, Yuichiro Maeda, and Akihiro Narita. The nature of the globular-to fibrous-actin transition. *NATURE*, 457(7228):441–445, JAN 22 2009.

[88] J PALAU, P ARGOS, and P PUIGDOMENECH. PROTEIN SECONDARY STRUCTURE - STUDIES ON THE LIMITS OF PREDICTION ACCURACY. *INTERNATIONAL JOURNAL OF PEPTIDE AND PROTEIN RESEARCH*, 19(4):394–401, 1982.

[89] T Pawson and P Nash. Assembly of cell regulatory systems through protein interaction domains. *SCIENCE*, 300(5618):445–452, APR 18 2003.

[90] C PERACCHIA. THE VERTEBRATE EYE LENS - AN EXCITING MODEL FOR GAP JUNCTION STRUCTURE-FUNCTION STUDIES. *FEDERATION PROCEEDINGS*, 39(6):1657, 1980.

[91] D Philp and JF Stoddart. Self-assembly in natural and unnatural systems. *ANGEWANDTE CHEMIE-INTERNATIONAL EDITION*, 35(11):1155–1196, JUN 17 1996.

[92] R POLLACK, M OSBORN, and K WEBER. PATTERNS OF ORGANIZATION OF ACTIN AND MYOSIN IN NORMAL AND TRANSFORMED CULTURED-CELLS. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 72(3):994–998, 1975.

[93] TD Pollard and GG Borisy. Cellular motility driven by assembly and disassembly of actin filaments. *CELL*, 112(4):453–465, FEB 21 2003.

[94] Young Hoon Roh, Roanna C. H. Ruiz, Songming Peng, Jong Bum Lee, and Dan Luo. Engineering DNA-based functional materials. *CHEMICAL SOCIETY REVIEWS*, 40 (12):5730–5744, 2011.

[95] Sandalo Roldan-Vargas, Frank Smallenburg, Walter Kob, and Francesco Sciortino. Gelling by Heating. *SCIENTIFIC REPORTS*, 3, AUG 16 2013.

[96] Flavio Romano, Piero Tartaglia, and Francesco Sciortino. Gas-liquid phase coexistence in a tetrahedral patchy particle model. *JOURNAL OF PHYSICS-CONDENSED MATTER*, 19(32), AUG 15 2007.

[97] Flavio Romano, Eduardo Sanz, and Francesco Sciortino. Phase diagram of a tetrahedral patchy particle model for different interaction ranges. *JOURNAL OF CHEMICAL PHYSICS*, 132(18), MAY 14 2010.

[98] John Russo, Piero Tartaglia, and Francesco Sciortino. Reversible gels of patchy particles: Role of the valence. *JOURNAL OF CHEMICAL PHYSICS*, 131(1), JUL 7 2009.

[99] Barbara Ruzicka, Emanuela Zaccarelli, Laura Zulian, Roberta Angelini, Michael Sztucki, Abdellatif Moussaid, Theyencheri Narayanan, and Francesco Sciortino. Observation of empty liquids and equilibrium gels in a colloidal clay. *NATURE MATERIALS*, 10(1):56–60, JAN 2011.

[100] Linda Sandblad, Karl Emanuel Busch, Peter Tittmann, Heinz Gross, Damian Brunner, and Andreas Hoenger. The Schizosaccharomyces pombe EB1 homolog Mal3p binds and stabilizes the microtubule lattice seam. *CELL*, 127(7):1415–1424, DEC 29 2006.

[101] F. Sciortino. Gel-forming patchy colloids and network glass formers: thermodynamic and dynamic analogies. *EUROPEAN PHYSICAL JOURNAL B*, 64(3-4):505–509, AUG 2008.

[102] NC Seeman. DNA in a material world. *NATURE*, 421(6921):427–431, JAN 23 2003.

[103] Maira Segura-Campos, Roberto Perez-Hernandez, Luis Chel-Guerrero Arturo Castellanos-Ruelas, Santiago Gallegos-Tintore, and David Betancur-Ancona.

Physicochemical and functional properties of dehydrated japanese quail (coturnix japonica) egg white. *Food and Nutrition Sciences*, 04(03):10, 2013.

[104] M Shibayama. Spatial inhomogeneity and dynamic fluctuations of polymer gels. *MACROMOLECULAR CHEMISTRY AND PHYSICS*, 199(1):1–30, JAN 1998.

[105] Kai Simons and Julio L. Sampaio. Membrane Organization and Lipid Rafts. *COLD SPRING HARBOR PERSPECTIVES IN BIOLOGY*, 3(10), OCT 2011.

[106] JV SMALL and JE CELIS. FILAMENT ARRANGEMENTS IN NEGATIVELY STAINED CULTURED-CELLS - ORGANIZATION OF ACTIN. *CYTOBIOLOGIE*, 16(2):308–325, 1978.

[107] JA SPUDICH and S WATT. REGULATION OF RABBIT SKELETAL MUSCLE CONTRACTION .1. BIOCHEMICAL STUDIES OF INTERACTION OF TROPOMYOSIN-TROPONIN COMPLEX WITH ACTIN AND PROTEOLYTIC FRAGMENTS OF MYOSIN. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 246(15): 4866–&, 1971.

[108] AN SRINIVASAN, CN NAGINENI, and SP BHAT. ALPHA-A-CRYSTALLIN IS EXPRESSED IN NONOCULAR TISSUES. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 267(32):23337–23341, NOV 15 1992.

[109] PE STEIN, AGW LESLIE, JT FINCH, and RW CARRELL. CRYSTAL-STRUCTURE OF UNCLEAVED OVALBUMIN AT 1.95 A RESOLUTION. *JOURNAL OF MOLECULAR BIOLOGY*, 221(3):941–959, OCT 5 1991.

[110] Y Sugimoto, S Sanuki, S Ohsako, Y Higashimoto, M Kondo, J Kurawaki, HR Ibrahim, T Aoki, T Kusakabe, and K Koga. Ovalbumin in developing chicken eggs migrates from egg white to embryonic organs while changing its conformation

and thermal stability. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 274(16):11030–11037, APR 16 1999.

[111] SH Sun, CB Murray, D Weller, L Folks, and A Moser. Monodisperse FePt nanoparticles and ferromagnetic FePt nanocrystal superlattices. *SCIENCE*, 287(5460):1989–1992, MAR 17 2000.

[112] DI SVERGUN. DETERMINATION OF THE REGULARIZATION PARAMETER IN INDIRECT-TRANSFORM METHODS USING PERCEPTUAL CRITERIA. *JOURNAL OF APPLIED CRYSTALLOGRAPHY*, 25(4):495–503, AUG 1 1992.

[113] DI Svergun. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing (vol 76, pg 2879, 1999). *BIOPHYSICAL JOURNAL*, 77(5):2896, NOV 1999.

[114] Alison M. Sweeney, David L. Des Marais, Yih-En Andrew Ban, and Sonke Johnsen. Evolution of graded refractive index in squid lenses. *JOURNAL OF THE ROYAL SOCIETY INTERFACE*, 4(15):685–698, AUG 22 2007.

[115] JW Szostak, DP Bartel, and PL Luisi. Synthesizing life. *NATURE*, 409(6818): 387–390, JAN 18 2001.

[116] A TARDIEU, F VERETOUT, B KROP, and C SLINGSBY. PROTEIN INTERACTIONS IN THE CALF EYE LENS - INTERACTIONS BETWEEN BETA-CRYSTALLINS ARE REPULSIVE WHEREAS IN GAMMA-CRYSTALLINS THEY ARE ATTRACTIVE. *EUROPEAN BIOPHYSICS JOURNAL WITH BIOPHYSICS LETTERS*, 21(1):1–12, MAR 1992.

[117] SI Tomarev and J Piatigorsky. Lens crystallins of invertebrates - Diversity and re-

cruitment from detoxification enzymes and novel proteins. *EUROPEAN JOURNAL OF BIOCHEMISTRY*, 235(3):449–465, FEB 1 1996.

[118] SI TOMAREV and RD ZINOVIEVA. SQUID MAJOR LENS POLYPEPTIDES ARE HOMOLOGOUS TO GLUTATHIONE S-TRANSFERASES SUBUNITS. *NATURE*, 336(6194):86–88, NOV 3 1988.

[119] SI TOMAREV, RD ZINOVIEVA, SM DOLGILEVICH, SV LUCHIN, AS KRAYEV, KG SKRYABIN, and GG GAUSE. A NOVEL TYPE OF CRYSTALLIN IN THE FROG EYE LENS - 35-KDA POLYPEPTIDE IS NOT HOMOLOGOUS TO ANY OF THE MAJOR CLASSES OF LENS CRYSTALLINS. *FEBS LETTERS*, 171(2):297–302, 1984.

[120] SI TOMAREV, RD ZINOVIEVA, and J PIATIGORSKY. CRYSTALLINS OF THE OCTOPUS LENS - RECRUITMENT FROM DETOXIFICATION ENZYMES. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 266(35):24226–24231, DEC 15 1991.

[121] SI TOMAREV, RD ZINOVIEVA, and J PIATIGORSKY. CHARACTERIZATION OF SQUID CRYSTALLIN GENES - COMPARISON WITH MAMMALIAN GLUTATHIONE-S-TRANSFERASE GENES. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 267(12):8604–8612, APR 25 1992.

[122] SI TOMAREV, RD ZINOVIEVA, K GUO, and J PIATIGORSKY. SQUID GLUTATHIONE-S-TRANSFERASE - RELATIONSHIPS WITH OTHER GLUTATHIONE S-TRANSFERASES AND S-CRYSTALLINS OF CEPHALOPODS. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 268(6):4534–4542, FEB 25 1993.

[123] SI Tomarev, S Chung, and J Piatigorsky. Glutathione S-transferase and S-crystallins of cephalopods: Evolution from active enzyme to lens-refractive proteins. *JOURNAL OF MOLECULAR EVOLUTION*, 41(6):1048–1056, DEC 1995.

[124] F VERETOUT, M DELAYE, and A TARDIEU. MOLECULAR-BASIS OF EYE LENS TRANSPARENCY - OSMOTIC-PRESSURE AND X-RAY-ANALYSIS OF ALPHA-CRYSTALLIN SOLUTIONS. *JOURNAL OF MOLECULAR BIOLOGY*, 205(4):713–728, FEB 20 1989.

[125] J Wang and DYC Fung. Alkaline-fermented foods: A review with emphasis on pidan fermentation. *CRITICAL REVIEWS IN MICROBIOLOGY*, 22(2):101–138, 1996.

[126] JA WEST, JG SIVAK, J PASTERNAK, and J PIATIGORSKY. IMMUNOLOCALIZATION OF S-CRYSTALLINS IN THE DEVELOPING SQUID (LOLIGO-OPALESCENS) LENS. *DEVELOPMENTAL DYNAMICS*, 199(2):85–92, FEB 1994.

[127] George M. Whitesides and Bartosz Grzybowski. Self-assembly at all scales. *Science*, 295(5564):2418–2421, 2002.

[128] GM Whitesides and M Boncheva. Beyond molecules: Self-assembly of mesoscopic and macroscopic components. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 99(8):4769–4774, APR 16 2002.

[129] Christoph Wiedemann, Peter Bellstedt, and Matthias Goerlach. CAPITO-a web server-based analysis and plotting tool for circular dichroism data. *BIOINFORMATICS*, 29(14):1750–1757, JUL 15 2013.

[130] Jing Zhang, Zhong-Yuan Lu, and Zhao-Yan Sun. A possible route to fabricate patchy nanoparticles via self-assembly of a multiblock copolymer chain in one step. *SOFT MATTER*, 7(21):9944–9950, 2011.

[131] Jing Zhang, Zhong-Yuan Lu, and Zhao-Yan Sun. Self-assembly of amphiphilic

patchy particles with different cross-linking densities. *SOFT MATTER*, 8(26):7073–7080, 2012.

[132] SG Zhang. Fabrication of novel biomaterials through molecular self-assembly. *NATURE BIOTECHNOLOGY*, 21(10):1171–1178, OCT 2003.

[133] ZL Zhang and SC Glotzer. Self-assembly of patchy particles. *NANO LETTERS*, 4 (8):1407–1413, AUG 2004.

[134] ZL Zhang, AS Keys, T Chen, and SC Glotzer. Self-assembly of patchy particles into diamond structures through molecular mimicry. *LANGMUIR*, 21(25):11547–11551, DEC 6 2005.

[135] Huaying Zhao, Patrick H. Brown, and Peter Schuckt. On the Distribution of Protein Refractive Index Increments. *BIOPHYSICAL JOURNAL*, 100(9):2309–2317, MAY 4 2011.