12-2022

# Light Scattering Study of an Important Eye Lens Protein: Beta Crystallin

Malcolm F. LaRose
mfl3719@rit.edu

# Light Scattering Study of an Important Eye Lens Protein: Beta Crystallin

By

*Malcolm F. LaRose*

A thesis submitted in partial fulfillment of the requirements

for the degree of Master of Science in Physics, in the

College of Science, Rochester Institute of Technology.

December 2022

PHYSICS

COLLEGE OF SCIENCE

ROCHESTER INSTITUTE OF TECHNOLOGY

ROCHESTER, NEW YORK

M.S. THESIS DEFENSE

---

Candidate: Malcolm F. LaRose

Thesis Title: Light Scattering Studies of Mammalian Beta Crystallin Protein

Advisor: Dr. George Thurston

Date of defense: / /2022

The candidate's M.S. Thesis has been reviewed by the undersigned. The Thesis

(a) is acceptable, as presented.

(b) is acceptable, subject to minor amendments.

(c) is not acceptable in its current form.

*Written details of required amendments or improvements have been provided to the candidate.*

Committee:

---

Dr. Michael Kotlarchyk, Committee Member

---

Dr. David Ross, Committee Member

---

Dr. Lishibanya Mohapatra, Committee Member

---

Dr. George Thurston, Thesis Advisor

Please submit form to Physics MS Graduate Program Coordinator

PHYSICS
COLLEGE OF SCIENCE
ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NEW YORK

# CERTIFICATE OF APPROVAL

**M.S. DEGREE THESIS**

The M.S. Degree Thesis of *Malcolm F. LaRose* has been examined and approved by the thesis committee as satisfactory for the thesis requirement for the M.S. degree in Physics.

Dr. Michael Kotlarchyk, Committee Member

Dr. David Ross, Committee Member

Dr. Lishibanya Mohapatra, Committee Member

Dr. George Thurston, Thesis Advisor

Date

## Acknowledgements

I want to thank my advisor and thesis committee for their deep knowledge and insight, as well as their excellent mentorship. I want to acknowledge my family and friends for invaluable moral support and their continued love and acceptance of me and my work. Finally, I want to thank my partner for always listening to me rant and rave about my favorite niche topic.

# Abstract

Light scattering experiments were performed on bovine $\beta_H$ crystallin, an important protein component of the mammalian eye lens. Light scattering results are compared to simplified models for the free energy of mixing of $\beta$ crystallins with water. The static light scattering data are well-represented by hard convex body equations of state having dimensionless non-sphericity coefficients of close to 2, compatible with prolate spherocylinders, and weight-average molecular weights ranging between $2-5\times10^5$ $g/mole$. Additionally, quasi-elastic light scattering data was obtained which shows that $\beta_H$ has a hydrodynamic radius near 7.3 $nm$. These experiments extend light scattering work on $\beta$ crystallins well into the realm of concentrations in which short-range order between proteins dramatically reduces light scattering efficiency, as occurs in the eye lens cytoplasm.

# Contents

# List of Figures

# List of Tables

# 1 Introduction

## 1.1 Anatomy and Lens Protein Overview

Mammalian eye lenses are composed primarily of lens fiber cells collected within a collagen membrane. The lens fibers are extremely long (several millimeters) but only a few microns wide. These lens fiber cells lack many of the usual organelles found in ordinary cells aiding in their transparency. The central 'nuclear' fibers form first, followed by progressive layers forming the 'cortex', leading to the biconvex ellipsoid shape, see Fig.(1). The cytoplasm of these fiber cells is a thick solution ($\sim$ 400 $mg/mL$) of water-soluble crystallin proteins, which come in three principal families in mammals, the $\alpha$, $\beta$, and $\gamma$ crystallins. The nuclear region possesses a higher concentration of crystallins than the cortex, and is enriched in the $\gamma$ crystallins.

The eye lens is nearly fully formed by birth. The transparency of the lens is essential for visual function, however, in the disease known as cataract, the lens scatters enough light to impair vision. One of the sources of light scatter is aggregation or phase separation of eye lens proteins (Thurston, 2006; Bloemendal *et al.*, 2004) producing a characteristic cloudy appearance. To understand the molecular basis of the scattering, it is necessary to understand how each lens component contributes to scattering light as well as how their mixtures behave.

Studies of solutions of physiological mixtures of crystallins show a relationship between the amount of light scattered and concentration where scattering initially increases to a maximum, followed by a decay as concentration further increases. This is due to short-range order (Delaye & Tardieu, 1983) among the proteins, where refers to the fact that increasing concentrations have decreased the nearest-neighbor distance and allowed more molecules to crowd, thereby making the density more uniform. This results in short-range structures for the relative positions of the molecules. However, these structures do not persist over a large length scale such as is the case in a crystal. Such a short range structure is described as 'liquid-like' or 'glass-like' by Delaye and Tardieu (Delaye & Tardieu, 1983; Delaye & Gromiec, 1983). Short range order results in the reduced scattering at higher (more physiological) concentrations, and thus contributes to the transparency of the eye lens (Benedek, 1971). It is

evident that understanding the structure and distribution of proteins in the eye lens is critical to understanding scattering from the lens, especially in the context of understanding diseases like cataract. Furthermore, to gain insight into the light scattering from the mixture it is critical to understand how each of the components behave by themselves, as well as all of the interactions between different components. This thesis focuses on the properties of one of the components, the $\beta$ crystallins.

Much work has already been done regarding the $\alpha$ and $\gamma$ crystallins, including mixtures of the two (Thurston, 2006; Bell *et al.*, 2017; Banerjee *et al.*, 2011; Dorsaz *et al.*, 2011). It was found by Thurston (2006) that mixtures of $\alpha$ and $\gamma$ crystallins can scatter less light than does a linear combination of the components. Thus, the properties of the mixtures are not simple linear combinations of those of the components. Further studies (Dorsaz *et al.*, 2009, 2011; Banerjee *et al.*, 2011; Bell *et al.*, 2017) showed that for the purpose of modelling available light and neutron scattering data, $\alpha$ crystallins can be modelled as hard spheres (repulsive interactions), and $\gamma$ crystallins can be modelled as hard spheres with square well or sticky attractions near the surface. It is interesting to note that the multi-subunit $\alpha$ crystallin particles are on the order of forty times larger in molecular weight than the $\gamma$ crystallins. Also, because the $\gamma$ crystallins are not actually spherical, but have more of a dumbbell shape, higher resolution scattering data is expected to necessitate more detailed models. The $\gamma - \alpha$ interaction was also modelled by short range attractions, and the overall scattering pattern was sensitively and non-monotonically dependent on the strength of those attractions (Dorsaz *et al.*, 2009, 2011; Banerjee *et al.*, 2011). In contrast to the situation with the $\alpha$ and $\gamma$ crystallins, in the case of the $\beta$ crystallins, not enough data are yet available to model the single component light scattering, an important prerequisite for studying their mixtures with $\alpha$ and $\gamma$. This thesis aims to contribute to the needed $\beta$ crystallin data.

A recent study on $\beta$ crystallin at high concentrations modelled the scattering as that due to a polydisperse mixture of hard spheres with a mild attractive force and a “corrugated" surface to account for the difference between idealized shapes and actual protein surfaces (Roosen-Runge *et al.*, 2020). In contrast, the present study finds static and quasielastic light scattering

evidence for a rod-like growth of a key component of the $\beta$ crystallins.

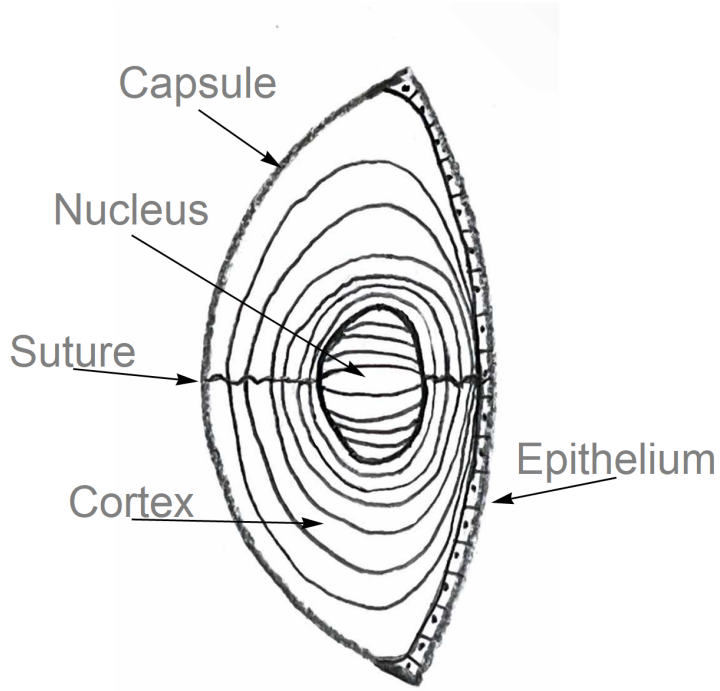



Figure 1: Cross sectional diagram (not to scale) of some of the anatomy of a mammalian eye lens primarily showing the locations of the cortical and nuclear regions. The sutures are where lens fiber cells fused when the lens formed. Note the presence of a thin epithelial membrane on the exterior facing surface of the lens.

The lens crystallins are identified and separated based on elution time in size exclusion chromatography (SEC), described in Sec.(2.2), and characterized by their average molecular weights, given in Table (1). As done in this work, the $\beta$ crystallin family can be subdivided into beta high ($\beta_H$) and beta low ($\beta_L$), which are separable with SEC. Both $\beta_H$ and $\beta_L$, further, are composed from a family of gene products, as described below, for which the bovine members are shown in Table (2). Similarly, the $\alpha$ crystallin is composed of two main types of subunits, and there are many distinct types of $\gamma$ crystallins (Bloemendal *et al.*, 1989).

Table 1: Families of Mammalian Crystallin Proteins

| Protein | $\alpha$ | $\beta_H$ | $\beta_L$ | $\gamma$ |
|---|---|---|---|---|
| Molecular Weight (kDa) | 800 | $\sim 200$ | $\sim 50$ | 20 |

## 1.2 $\beta$ Crystallin Details

As indicated above, the $\beta_H$ and $\beta_L$ crystallins are oligomeric associations of smaller protein subunits. There are six $\beta$ subunits that associate to produce $\beta_H$ and $\beta_L$. The $\beta A$ subunits are so named because they have an excess of acidic residues, and likewise the $\beta B$ subunits have an excess of basic residues. Although the individual subunit types that are present in each of $\beta_H$ and $\beta_L$ have been characterized (Lampi *et al.*, 2014), there is very little information about the quantitative proportions of each of these subunits in each association (Lampi, 2022). There is an x-ray crystallographically determined structure for only one of the subunit types, $\beta B2$ (Bateman & Slingsby, 1992). For the subunits for which experimental structures have not yet been determined, an important recent development is that of AlphaFold, which uses machine learning to predict the tertiary structures of proteins from their amino acid sequence and homology with other proteins (Varadi *et al.*, 2021; Jumper *et al.*, 2021). AlphaFold predictions for $\beta$ crystallin structures show largely globular protein subunits with dangling N-terminal arms. In Table 2 the molecular weights for each of the $\beta_H$ crystallin subunits as determined from their amino acid sequences by UniProt (Consortium, 2020) are shown. Fig.(2) shows the structures predicted by AlphaFold.

Table 2: Molecular Weight of $\beta_H$ crystallin subunits

| Beta High Subunit | Molecular Weight (Da) |
|---|---|
| Beta B1 | 28141 |
| Beta B2 | 23298 |
| Beta B3 | 24328 |
| Beta A2 | 22231 |
| Beta A3 | 25131 |
| Beta A4 | 23860 |

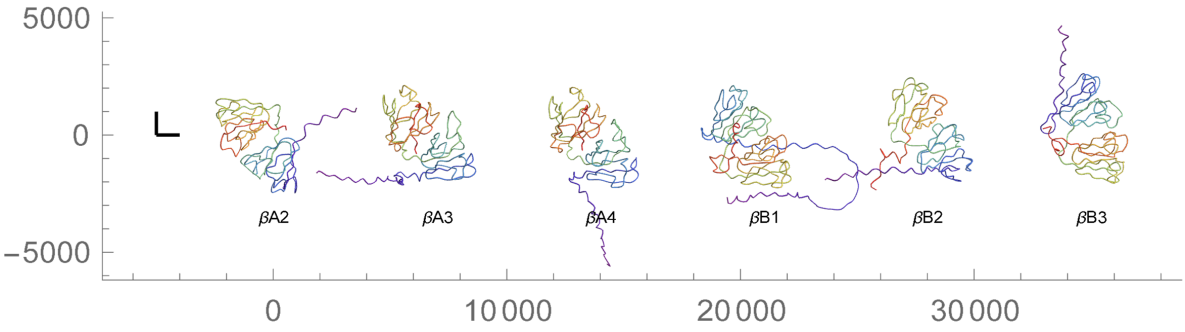



Figure 2: Plotting each of the constituent subunits of $\beta_H$ crystallin on the same figure allows for direct quantitative size comparison. Note that the arms of the black L are each 1 $nm$ in length. The axes are measured in picometers.

Studies have shown that $\beta_H$ crystallin has a concentration dependence on aggregation (Siezen *et al.*, 1986). That is, at very low concentrations, $\beta_H$ appeared to reversibly dissociate into it's constituent subunits, and most notably, a sizeable fraction of $\beta_L$ crystallin. This is to say that $\beta_H$ crystallin exists in an equilibrium with $\beta_L$ and individual $\beta$ monomers, where the dominating state at high concentration is $\beta_H$ crystallin. The exact character of this equilibrium is not known but is critical for understanding the exact role $\beta_H$ crystallin plays in the eye lens.

It has also been noted that $\beta_H$ crystallin is apparently less stable upon purification (Siezen *et al.*, 1986; Herbrink & Bloemendal, n.d.). Purified $\beta_H$ appears to form large molecular weight aggregates far faster than it does as a mixture with $\beta_L$. This is further evidence for

the existence of an equilibrium between $\beta_H$ and its subunits.

## 1.3 Light Scattering

At the microscopic scale, everything is made of atoms which are in constant motion (rotation, vibration, and translation). In a sample placed in a light scattering apparatus, or anywhere else, these motions result in a constantly fluctuating distribution of charges. As a light wave passes through the sample, it interacts with and alters this charge distribution, causing it to oscillate and re-emit light of nearly the same frequency. The fluctuation of charges in dielectric media can be thought of as fluctuations in the electric susceptibility, which is related to the media's refractive index in electromagnetic theory. This is to say that the electric field in the sample is made to fluctuate by the incident laser light, and the resulting scattering in every direction is measured at an angle to the incident beam. At an arbitrary scattering angle, the two primary quantities of interest are the time-averaged intensity of the scattered light (a measure of how much light was scattered) and the time dependence of the fluctuations. The full details of the derivations of various relevant quantities can be found in Ch. 2 of Berne and Pecora (Berne & Pecora, 2000) or Ch. 2 of Chu (Chu, 2007).

In static light scattering (SLS), the time-averaged scattering intensity at a scattering angle $\theta$ is measured. Using the Zimm equation (given in Eq.(2.12)) and making measurements at varying concentrations and scattering angles yields the so-called Zimm plot. As discussed in further detail later, the primary focus of this work is on changes in concentration, which reduces the Zimm equation to a simpler form, the Debye equation given in Eq.(2.13). This Zimm analysis is ubiquitous in light scattering for the determination of the average molecular weight (for a polydisperse sample), the second virial coefficient and the radius of gyration. Thermodynamic models of the light scattering from solutions of particles may be fit to the data, which allows one to take account of scattering in both the high and low concentration regimes. Details of these analyses are in Sec.(2), particularly around Eq.(2.8) and Eq.(2.13).

In quasielastic, or dynamic, light scattering (QELS or DLS), the time dependence of fluctuations in scattered light are measured and analyzed. The size and frequency of the fluctuations

has to do with diffusion of particles in and out of the scattering volume at equilibrium. Analyzing the time correlation function of the measured signal yields the hydrodynamic radius of the particles in the sample. Details of this analysis are in Sec.(2), particularly around Eq.(2.20).

Fig.(6) shows a schematic diagram of our light scattering apparatus, which aids in describing and defining some relevant quantities. It is important to note that in these experiments, the laser light is in the so-called 'V-V geometry', where the incident and scattered waves have electric field components both in the vertical direction. It is also critical to note that the intensity that is measured in light scattering is all due to single scattering events, where incident light was scattered only once while passing through the sample. This is a difficult criterion to quantify, but can be checked for 'experimentally' by looking for a diffuse scattering cone around the main laser beam passing through the sample. Finally, as our particles are significantly smaller than the wavelength of the incident light[1], all of the scattering is in the regime described by Rayleigh scattering (Strutt, 1871).

## 1.4 Outline of Thesis

In Sec.(2) the steps necessary to obtain, purify, and clean samples of crystallin proteins from bovine eye lenses are discussed. This is a lengthy process and deserves detailed discussion that is not readily found in the literature. Next the methods used for each experiment are described and discussed with as much detail as possible. The methods used for analyzing and interpreting collected data are also discussed. In Sec.(3) the direct results from light scattering experiments are graphically and numerically shown. Details from each major experiment are included and help shed light on the variations found between experiments. In Sec.(4) the light scattering results are further analyzed and interpretations on protein characterization and model selection are provided. Finally, the results are gathered and conclusions are made in Sec.(6). Further interpretations and the next steps for future researchers are suggested.

[1] A generously large particle in our sample is about 20 $nm$ wide. The wavelength of the laser is 637 $nm$, so $\frac{20}{637} \sim 0.03 \ll 1$.

# 2 Methods

Before an experiment can be performed, much work is necessary to properly separate the proteins found in an eye lens. This work begins with the dissection of a number of bovine eyes to extract the lenses, followed by chromatographic methods to separate the proteins. Finally, the proteins are reconcentrated and the final concentrations measured with ultraviolet spectrophotometry. Further efforts are required to ensure the sample is viable for light scattering; these efforts mostly revolve around the mitigation of dust, which is a catch-all term to describe anything in our sample that is not the molecule of interest. Average light scattering intensity goes as the fourth power of the average molecular radius (Berne & Pecora, 2000), so small increases in particle size leads to a massive increase in scattering. This causes dust particles (cell debris, human skin or hair particles, aggregated proteins, etc.) to very easily overwhelm the scattering signal from the proteins in our sample. Dust particles very nearly the same size as the proteins were particularly pernicious, as it is more difficult to remove them without affecting the properties of the sample. Proper cleaning of the sample tube and filtration of the sample and buffer used in a dilution series is critical for obtaining usable data. The most successful methods for removing dust combined multiple methods and often repeated steps a handful of times.

Here the methods used at every stage of sample preparation and experiment are described in detail. First, the methods of sample preparation, including the buffers used as a solvent, and the method of eye lens homogenization are described. Also discussed are chromatographic methods, primarily size exclusion chromatography and how it pertains to the specific protein studied in this thesis. This segues into a discussion of the experimental setup, as well as the specifics of cleaning and filtering samples. Finally, the methods used for analysis, including efforts taken while collecting data to mitigate the harmful effects of dust, are discussed.

## 2.1 Buffer Prep and Protein Purification

The primary buffer used for size exclusion chromatography (SEC) and light scattering was a $pH = 6.8$ 50 $mM$ sodium phosphate buffer containing 200 $mM$ sodium chloride to protect

the proteins from binding to the SEC gel, 2.5 $mM$ dithiothreitol (DTT) to prevent oxidation, 1.0 $mM$ EDTA to chelate metal ions and reduce interfacial forces (W. Wang, 2010), and 0.2 % sodium benzoate by mass as a bacteriostatic agent (Perrin & Dempsey, 1974). These components are dissolved into nanopure water, which has a measured resistance of greater than 18.2 $M\Omega$. The buffer is then degassed by applying a vacuum and stirring for several minutes. All chemicals were purchased from Sigma Aldrich, aside from the DTT, which was purchased from GoldBio, and the nanopure water, which was obtained in Gosnell Hall.

A similar $pH = 7.1$ 100 $mM$ sodium phosphate buffer was also used for some of the first light scattering experiments on $\beta_H$ (Thurston, 2006). This buffer includes the same proportion of sodium benzoate as the 50 $mM$ phosphate buffer, however it contains 20 $mM$ DTT instead of 2.5 $mM$. Additionally, this buffer contains no sodium chloride or EDTA. To use this buffer, the sample was diluted over 200 times into the new buffer while using Amicon ultrafiltration to continuously concentrate the sample. After several experiments, it was decided buffer exchanging was unnecessary, as the additional filtration often resulted in lost sample due to aggregation and adsorption. Specific buffers used for each experiment are mentioned in the experiments section.

The protein purification begins with dissecting around 30 eyeballs extracted from newborn or third trimester calves. These lenses are rinsed to remove some excess connective tissue then homogenized in an equal volume of 50 $mM$ phosphate buffer to gently break open the lens fiber cells and solubilize the crystallin proteins. This mixture is then centrifuged at $\sim 30,000\ g$ for $2 - 4$ hours, and the supernatant is separated from the 'puck' of cell debris and insoluble molecules. It was sometimes necessary to centrifuge the lens homogenate for additional time if the solution was not apparently free of debris. The solution was also sometimes passed through a 0.8 $\mu m$ filter, however this was often avoided as syringe filtration tends to reduce the final yield of proteins. The homogenate was stored in the refrigerator near $4^oC$ until it was used for size exclusion chromatography, to be discussed in the next section.

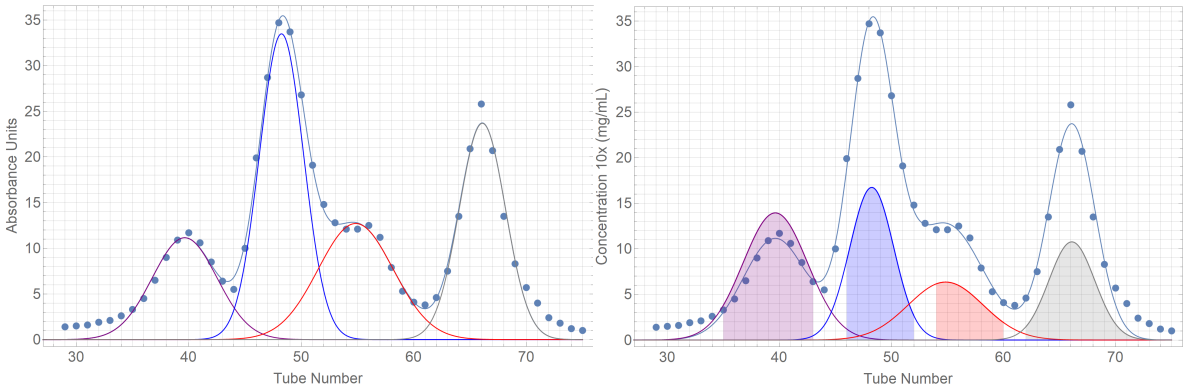
## 2.2 SEC Theory, UV Absorbance, and our System

Size Exclusion Chromatography is a technique used for separating mixtures of biological samples based on molecular size (gel, 1998). Small samples (5 $ml$ homogenate in 20 $mL$ buffer) are passed through a complex gel matrix by a continuous flow of buffer. The gel matrix consists of a solid phase (the gel 'beads') and a liquid phase, which contains the sample dissolved in eluent buffer. The pores in the solid phase form a cave-like network that the sample may diffuse in and out of freely. Small particles can 'explore' a larger number of pores and so take a longer time to diffuse through the column than larger particles. In this way, the particles are separated based on their relative sizes, with the largest particles eluting first and the smallest eluting last. Fig.(3a) below shows an example chromatography run on the eye lens homogenate, with each peak labeled with its respective protein. The first proteins to elute are the large globular $\alpha$ crystallins, followed by the intermediate size $\beta_H$ and $\beta_L$ crystallins. The light $\gamma$ crystallins elute last. Eluent absorbance at 280 $nm$ ultraviolet light with a 0.2 $cm$ path length is measured and recorded versus elution volume with an analog Amersham Pharmacia-Biotech (APB) chart recorder.

The Beer-Lambert Law (Beer, 1852; Lambert, 1760) allows us to quantify the concentration $c$ of protein in a given sample if the molar attenuation coefficient $\epsilon$ is known and the absorbance $A$ is measured over a path length $l$.

$$A = \epsilon c l \tag{2.1}$$

The absorbance trace can be used as a tool to quantify the mass of protein in each fraction by integrating over each shaded area in Fig.(3b) while accounting for the differences in UV absorbance per unit mass, which are different for different species in general. Absorbance at the midpoint of each fraction was measured from the chart paper with a ruler and input into Mathematica. Four Gaussians were fit simultaneously to the overall curve to quantify the amounts of each crystallin proteins. Each Gaussian was divided by the appropriate extinction coefficient for the corresponding protein to convert absorbance units to concentration, shown in Fig.(3b).




(a) (b)

Figure 3: The result of size exclusion chromatography on bovine eye lenses. Volume is on the horizontal axis (where one fraction is 22 $mL$) and 10 $x$ absorbance units on the vertical axis. Collected fractions are shown by the shaded areas.

More often, this integral is done as the sum of the concentration in each tube multiplied by its elution volume. For example, the total mass of $\beta_H$ crystallin eluted in the run shown in Fig.(3a) is approximately 196 $mg$. To demonstrate, consider the measured absorbance value in tube 49, $A_{49} = 34$. The full width of the chart paper corresponds to 2 absorbance units, so this is really $0.34 * 2 = 0.68$. Dividing this by the molar attenuation coefficient gives 0.34 $\frac{mg}{mL}$ once again. This is through a 2 $mm$ path length however, so a measurement through a 1 $cm$ path length yields an absorbance that is five times greater. The absorbance must correspondingly be multiplied by this factor to get the final concentration of 1.7 $mg/mL$. Multiplying by the volume gives the mass of protein, which for this tube is 37.4 $mg$. Considering that a dissection of $15 - 30$ eyes yields enough material for $5 - 10$ SEC runs, upwards of 1 $g$ of each of the $\alpha$, $\beta_H$, and $\gamma$ crystallin proteins may be obtained, which allows for multiple experiments if care is taken and time is well managed.

Our SEC system is composed of a Pharmacia-Biotech (PB) XK-50 column with an approximate bed volume of 1 $L$, connected to a pump to allow for the flow of buffer and a collector to automatically collect and separate fractions. The column is packed with Sepharose CL-6B gel filtration media, which allows for fractionation of globular proteins between $10 - 4,000$ $kDA$. Samples are loaded in 25 $mL$ volumes, and approximately 300 $mL$ of buffer is applied to begin the chromatographic process. At this point, fractions of 22 $mL$ are collected automatically for

an additional 1790 $mL$ by a PB GradiFrac system, which also drives the pump and marks the chart recorder when a fraction is collected. The fractions are collected into amber bottles as indicated in the figure above and stored near $4^oC$.

The concentration of all samples were measured using a Shimadzu $UV-2401PC$ UV-Spectrophotometer, which measures ultraviolet light absorbance between $250-300$ $nm$. The spectrophotometer was set to 'slow' mode, with a 0.1 $nm$ wavelength increment. Each protein separated via SEC was measured in a clean quartz glass cuvette with $l = 1$ $cm$ path length. Upon comparison, it was possible to see that each protein may be identified by its absorbance 'trace' shown in the table below. Concentrations were determined by the value of absorbance at 280 $nm$, $A_{280nm}$, multiplied by the dilution factor (how much the sample was diluted for an absorbance measurement) and divided by the appropriate molar attenuation coefficient[2] for each protein, as indicated by Eq.(2.1). The characteristic traces of each major bovine lens crystallin, as well as the corresponding extinction coefficients that were used are shown in Fig.(4).

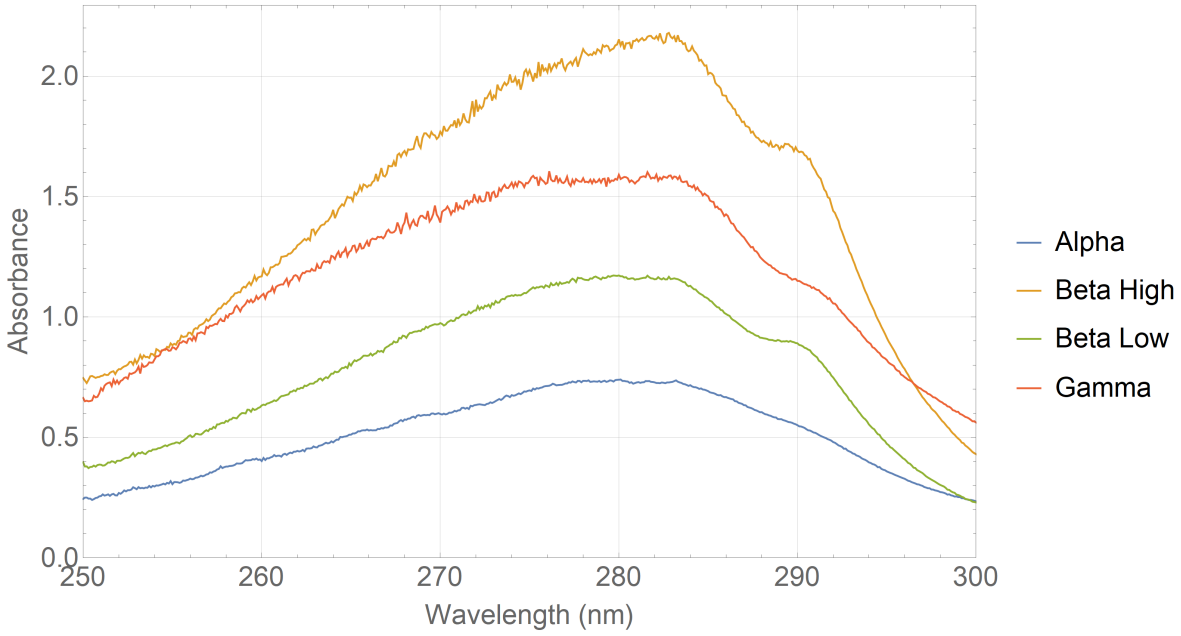



Figure 4: Ultraviolet absorbance traces for each of the principal lens crystallins. Note the qualitative differences between the curves, for example, the strong 'shoulder' near 290 $nm$ that is present in both $\beta$ species, but less so in $\gamma$ and almost nonexistent in $\alpha$.

[2]For $\beta$ crystallin, this is 2 $\frac{mg}{mL}^{-1}cm^{-1}$ (Pierscionek *et al.*, 1987; Siezen, 1984)

To produce reliable UV absorbance measurements, especially of highly concentrated samples, the sample must typically be diluted carefully into a liquid that does not contain UV absorbing components, such as the DTT and sodium benzoate in our typical 50 $mM$ sodium phosphate buffer. The dilution factor was calculated from masses measured on a 5 digit mass balance, accurate to a tenth of a milligram, where it was assumed that 1 $mL = 1$ $g$, i.e. samples of proteins have nearly the same density as pure liquid water. Small volumes of protein were obtained with graduated 5 $\mu L$ glass Drummond micropipettes. The pipettes were wiped with a wet tissue to remove excess protein and repeatedly flushed in the dilute sample, and thus retained the highest degree of precision possible. The sample was then thoroughly mixed using a Vortex Genie to ensure complete homogeneity of the sample.

It is vital to note here that all fractions of $\beta_H$ collected comprise a mixture of proteins. That is to say, there is no point at which only $\beta_H$ elutes from the column, it is always present with some mixture of $\alpha$ and $\gamma$. In Fig.(3b) there is no region in which the blue shaded area lies only under the blue curve, there is some admixture with the purple and red curves representing different crystallin species. This inspired later experiments where efforts were taken to further purify and separate $\beta_H$, and these efforts are discussed in the next section. For these studies, these 'contaminating' proteins likely contributed some to the variability of the results, shown in Sec(3). Future studies may use the above absorbance traces to quantify mixtures of proteins, as the total absorbance is a simple linear combination of the absorbances of the constituents.

## 2.3 Rechromatography

Rechromatography was used to try to further separate $\beta_H$ from the other lens crystallins, particularly $\beta_L$, where fractions of $\beta_H$ were ran a second time through the SEC column. An example chromatography trace is shown in Fig.(5). As explained before briefly and at further length in the results section, our chromatography setup does not have the necessary resolution to perfectly separate $\beta_H$ from the 'nearby' $\alpha$ and $\beta_L$. Using this method essentially increases the resolution of our initial SEC run, effectively spreading out the $\beta_H$ peak over more fractions. The same high and low concentration analysis was performed for rechromatographed $\beta_H$ as

was done for $\beta_H$ that was only ran once on SEC.

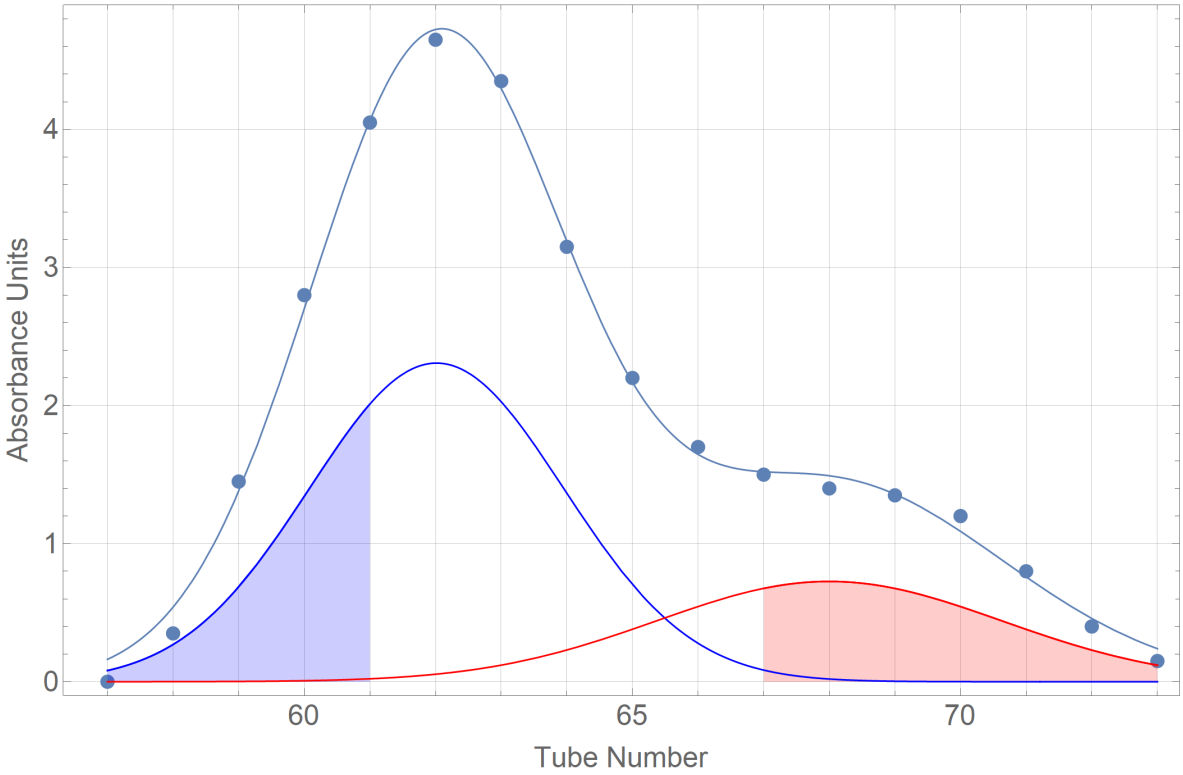



Figure 5: An absorbance trace for rechromatographed $\beta_H$ crystallin. The entire $\beta_H$ fraction from the initial SEC run was loaded, however, $\beta_L$ also eluted. This is likely due to concentration dependent dissociation of $\beta_H$ into $\beta_L$, as discussed by (Bateman & Slingsby, 1992)(Bindels *et al.*, 1981)(Siezen *et al.*, 1986). See the text for further discussion.

## 2.4 LS Sample Prep

After separating $\beta_H$ crystallin from the other lens proteins, more work was needed to produce a viable light scattering sample. After chromatography, the $\beta_H$ existed as a dilute solution ($\sim 1-2\ mg/mL$), but for these studies, it was desired to start with a solution with a concentration around $200-400\ mg/mL$. This allowed a dilution series to be performed for a light scattering experiment, where a known quantity of buffer was successively added to a sample of protein and laser light scattering was measured at a scattering angle of $90^o$. The specifics of the measurement are described in the next section. At each step, the concentration was calculated from the known initial concentration and sample mass, as well as the mass of

buffer introduced at each step[3].

It is preferable to perform a dilution series as opposed to its counterpart, the concentration series. Successively concentrating the sample after each measurement involves removing it from the light scattering tube and returning it to a centrifugal concentrator tube. Afterwards, it would be necessary to measure the concentration directly as it would be extremely difficult to accurately quantify the elution volume. Each step of transferring the sample introduces air and surface borne dust, which has to be dealt with at each step. Transferring the sample back and forth also brings added risk of spillage, and also leaves small volumes of proteins behind on the walls of containers, which is extremely undesirable when working with animal tissue samples, both for the cost and cleanup, and for respect for the animal tissue.

To concentrate the sample initially, ultrafiltration was performed with an Amicon Stirred Cell Concentrator using a 10 $kDA$ cutoff cellulose acetate membrane. Approximately 50 $PSI$ of pressure is provided from an ultra high purity nitrogen tank to force the water and buffer components through the membrane, leaving the heavy protein behind. Gentle stirring is necessary to clear proteins from clogging the membrane, however it may lead to shear induced aggregation (W. Wang, 2010) if care is not taken, especially as higher concentrations (and lower volumes) are reached. Nonetheless, it becomes unfeasible to elute more buffer after a certain concentration is reached ($\sim 100\ mg/mL$ for $\beta_H$ crystallin) so it is loaded into Amicon Ultra 10 $kDa$ centrifugal concentrators and spun between $5000-6000\ g$ until the final concentration is reached. This final concentration is determined ultimately by the amount of starting sample and is limited to a final volume of 650 $\mu L$ to fit in the light scattering tubes. At these high concentrations, direct filtration of the sample to remove dust is impossible, so any syringe filtration must be done with a larger volume and lower concentration. This also reduces loss due to adsorption of proteins to the filter. Evidently, this also means that increasing care must be taken as concentration increases to avoid the introduction of dust.

Glass light scattering tubes were purchased from Brookhaven Instruments and were thoroughly cleaned with the following procedure, adapted from that performed by Donovan *et al.* (1991). The tubes were first rinsed well with several tube volumes of nanopure water and

[3]This is with the conservation of mass expressed as $C_1V_1 = C_2V_2$

additionally dilute solutions of Liquinox lab soap if there are visible stains. They were then placed in a 2 $M$ potassium hydroxide bath whose solvent is equal parts pure water and pure ethanol for at least 24 hours. The tubes were then transferred to a 4 $M$ nitric acid bath for another 24 hours. Finally, the tubes were removed and rinsed with an extremely large volume of nanopure water to remove any trace of the cleaning baths. These steps constitute the rough cleaning, but polishing was still necessary before introducing a sample into the tube. The tube was again rinsed with a few tube volumes of nanopure water, this time using the Vortex Genie to violently rinse the water against the tube walls. Finally, nanopure water was syringe filtered directly into the tubes with a 0.02 $\mu m$ Whatman Anotop syringe filter. This was repeated a few times until the water would sheet evenly down the glass walls. At this point, the tube was dried with syringe filtered air, and was labelled and sealed with plastic film. The mass of this tube was recorded as the dry mass, used to calculate the mass of sample contained within.

Before introducing the sample, the tube was rinsed one final time with syringe filtered water and care was taken to shake out drops of water while limiting the time the tube was open and exposed to the air. After introducing the sample, the concentration was measured (to account for any minor dilution due to remaining water droplets), and the dry mass was subtracted from the total mass to obtain the mass of the sample. Before withdrawing sample for concentration measurement, it was carefully and lengthily mixed to ensure the sample is homogeneous in concentration, avoiding use of the Vortex Genie as it was to vigorous to work directly with light scattering samples. At high concentrations, this was often achieved by stirring with a small clean plastic or glass rod, moving carefully to ensure no protein was splashed and shear was minimized.

For the dilution series, syringe filtered buffer was prepared by filtering through fresh 0.02 $\mu m$ filters into a glass vial which was rinsed several times with the same buffer. The vial was sealed with plastic film (also rinsed as before) and dilution buffer was drawn into a syringe by stabbing a needle through the film. Any excess was flicked off the end, and buffer was once again filtered and added to the sample by similarly stabbing it through it's film seal. By never removing the film and filtering many times, it was found that this method

significantly reduces the amount of dust introduced with each successive dilution. However, when the sample was very concentrated, it could only be adequately mixed with the dilution buffer with a small clean instrument, as heavy shaking may have aggregated the proteins or suspended bubbles that act as scattering particles e.g. dust. Several early experiments did not use all of the steps involved here, specifically the use of a a syringe needle and plastic seal; this is addressed in the experiments section.

## 2.5 Light Scattering Experiment

Static and quasielastic light scattering were performed using a Brookhaven Instruments BI-200SM goniometer equipped with a BI-APDX avalanche photodiode and a Mini-L40 diode laser, which emits red light at 637.6 $nm$ at 20 $mW$. A 100 $\mu m$ pinhole in front of the detector selected a solid angle for measurement. Data were simultaneously processed with a TurboCorr digital autocorrelator to produce a correlation function at extremely small 100 $ns -$ 80 $\mu s$ lag times, and the average photocounts per second were recorded each second. All measurements were performed at a $90^o$ scattering angle. A schematic diagram is shown in Fig.(6) to help define and describe relevant quantities.

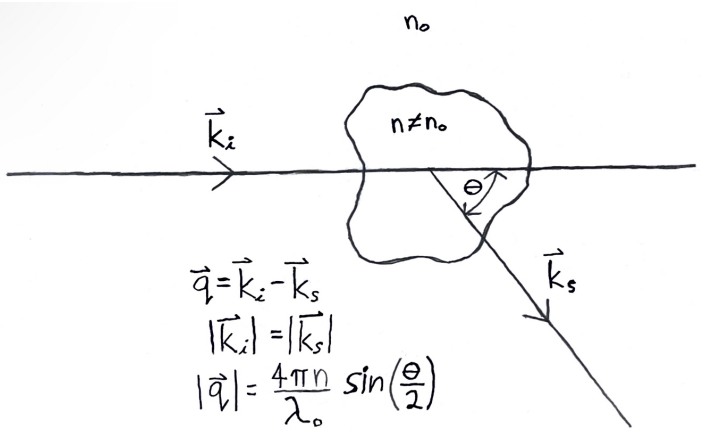



Figure 6: A schematic diagram of light scattering from an arbitrary sample particle whose refractive index differs from its surroundings. Incident light emitted by a laser is denoted by $\boldsymbol{k}_i$ and the light observed by a detector is denoted by $\boldsymbol{k}_s$. For static and quasielastic light scattering, the magnitudes of these wave vectors are equal, and so the resultant scattering vector magnitude is found to be $q = \frac{4\pi n}{\lambda_i} \sin \frac{\theta}{2}$.

The goniometer was filled with an index matching fluid (decahydronaphthalene) that closely matched the refractive index of the light scattering tube and glass vat walls, greatly simplifying the optics of the system. It was also equipped with a temperature controlled water bath, which kept the sample and system at $25^oC$ for all experiments. The system was allowed to reach and maintain this temperature for at least an hour before an experiment begins. This also ensures the laser has warmed up, as shown in Fig(7). Samples placed into the system take only a moment to reach thermal equilibrium, as their mass is negligible compared to the system. When following this procedure, thermal fluctuations are not a significant source of error in our measurements. Toluene[4] was used as a reference (Kaye & McDaniel, 1974; Coumou, 1960) for converting from average intensity to excess Rayleigh ratio, a measure of the efficiency of scattered light. Eqs.(2.2 - 2.4) are used to do this conversion (Kirkwood & Goldberg, 1950; Coumou, 1960). The bracketed terms represent time averaged intensities of samples of the protein solution, samples of the buffer it is dissolved in treated as a single component, samples of spectrophotometric grade toluene, and the average intensity of the intrinsic noise of the detector, obtained with a closed shutter and called the dark count.

$$\Delta R(90) = \frac{\Delta\langle I(90)\rangle}{\langle I_{ref}(90)\rangle} R_{ref}(\frac{n}{n_{ref}})^2 \tag{2.2}$$

$$\Delta\langle I(90)\rangle = \langle I(90)_{solution}\rangle - \langle I(90)_{solvent}\rangle \tag{2.3}$$

$$\langle I(90)_{ref}\rangle = \langle I(90)_{toluene}\rangle - \langle I(90)_{dark}\rangle \tag{2.4}$$

Before measurement, samples were placed in the system with the laser on and examined briefly with the naked eye at a safe (close to $90^o$) angle to safely look for the presence of dust. Dust particles, even those too small to see normally, scatter light much more than the surrounding solution (even when quite concentrated), and were visible as bright 'speckles' moving in and out of the path of the beam. A clean sample (see Fig.(8a)) had few dim speckles that

[4]These experiments used a value of $R_{tol}^{633\ nm} = 14 \times 10^{-6}\ cm^{-1}$

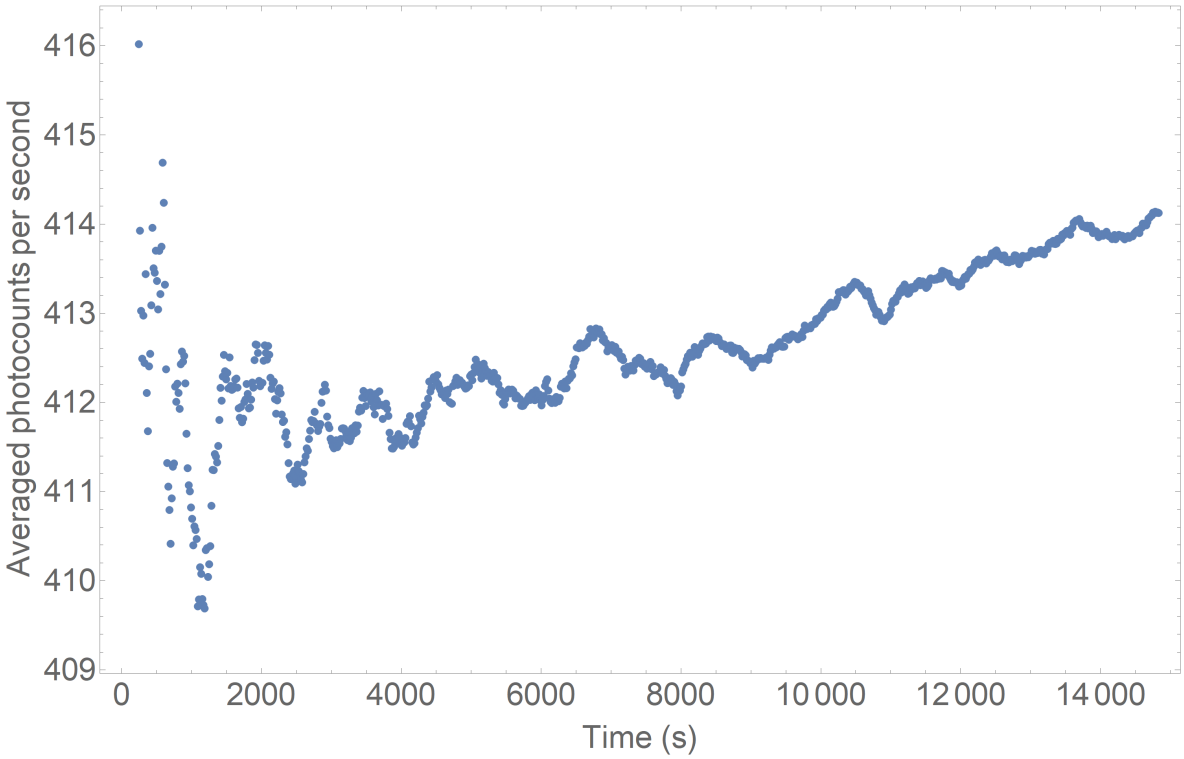
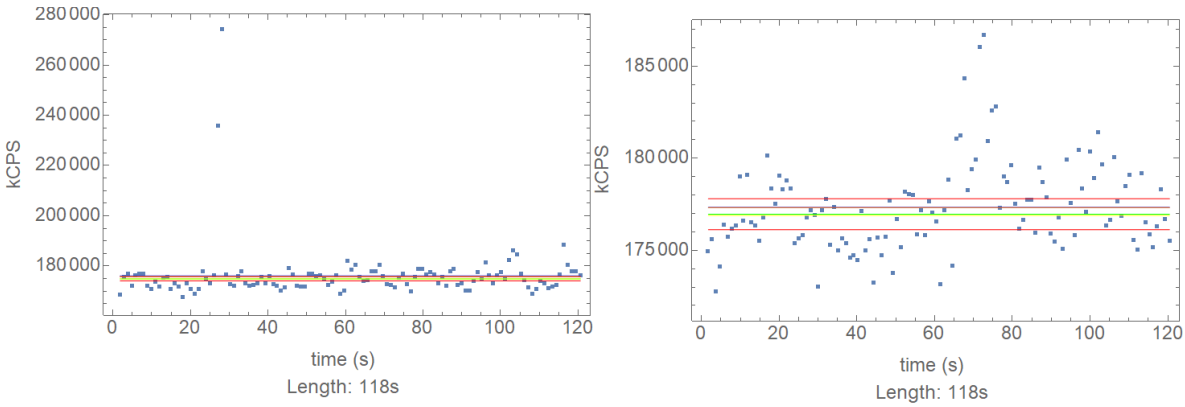



Figure 7: A running average of light scattering photocounts from toluene while the laser is allowed to reach thermal equilibrium. The vertical axis represents the number of photons detected averaged over about 16 seconds, and the horizontal axis represents time in seconds. Note that the fluctuations have largely died down by about 3200 $s$, indicating that about an hour is adequate time to reach thermal equilibrium. The scale of the vertical axis also indicates that the slight positive slope is insignificant, as it only accounts for about a 0.5% increase over an hour.

rarely intersect the beam, and may even appear as a uniform line passing through the sample. For a dirty sample (see Fig.(8b)), many bright speckles moved in and out of the beam, and it often appeared as a chain of bright beads. These samples were not viable for light scattering, and had to be cleaned further, resulting in lost protein. As mentioned in the previous section, syringe filtering highly concentrated samples is impossible, so this also resulted in a considerable loss of time as the samples had to be centrifuged to 'sediment' the comparatively large dust particles. Estimates of the time necessary to sediment a particle of a given size and density in a fluid of a certain viscosity can be made from free-body considerations. In practice, it was found that centrifuging at the maximum acceptable speed to use with the glass light scattering tubes ($\sim 2000\ g$) would produce viable samples from dusty ones in about an hour. This varies based on the size and quantity of dust present, however, and was difficult to quantify. Additionally, the samples then need to be transferred to clean light scattering tubes and the concentration remeasured. Alternatively, the sample could remain in the tube that was spun, but additional spinning is necessary after each dilution step.

First, a long measurement ( $5-7\ min$) was taken to check whether there was a drift in the average counts per second, which is one of the key quantities measured in a light scattering experiment. A drift usually indicated the sample needed more mixing, but if it just came from the refrigerator, it may only need more time to thermally equillibrate. Example 'count rate histories' (photocounts accumulated in a second versus time) were shown in Fig.(8). Ideally, there was very little dust and no drift, and there was a very clear 'baseline'. This allowed for longer measurements, which help accumulate a stronger correlation function. When some dust was present, and especially when the concentration was low, it was difficult to 'build up' an adequate correlation function, limiting the analysis of the measurement to static light scattering. Often, an average could still be obtained in the presence of considerable dust, if one was patient to wait for a long enough[5] baseline amidst dust peaks. For most observations, the arithmetical mean was a reasonable estimation of the average. For dusty data, the median tended to do a little better at finding an average closer to the observed baseline.

[5] An average could be obtained in as few as $\sim 10\ s$ if there was sufficient observation to establish it as the 'true' average.




(a) (b)

Figure 8: Two count rate histories measured from a $2.15\frac{mg}{mL}$ $\beta_H$ crystallin solution. The difference in vertical scales is caused by the presence of a large scatterer in Fig.(8a), which did not spend much time in the path of the laser. The various colored bars represent different forms of averages for estimating the average photocounts per second. Most importantly, the red bars represent the standard deviation given by Poisson statistics. See subcaptions or text for details.

In addition to the light scattering experiments performed at a scattering angle of $90^o$, an effort was made to determine the viability of measuring scattered light at greater or lesser angles. A measurement of the distribution of scattered light as a function of angle is called the 'angular dissymmetry' or 'angular anisotropy' of the sample, and this information yields the radius of gyration, $R_g$, discussed in Sec.(2.6). Unfortunately, it was found that the light scattering apparatus was out of alignment. This was not an issue for comparing measurements made at the same angle, but proved prohibitive for obtaining angular scattering information. To obtain angular data, the device would require realignment, which is a difficult and time consuming task. In the interest of the time of the researchers who need to use the instrument, it was decided to postpone realignment until it is necessary and there is appropriate time.

To try to measure the radius of gyration with a different technique it was attempted to perform small-angle X-ray scattering using Dr. Michael Pierce's x-ray system. The x-ray source was molybdenum with an average beam energy of 17.4 $keV$.

Samples were prepared in $1-2$ $mm$ borosilicate glass capillary tubes manufactured by Borokapillaren. The capillary tubes were extremely fragile, so great care was taken while preparing samples. Unfortunately, no significant signal was detected from samples of proteins,

so this was not a viable method for measuring the radius of gyration.

## 2.6 Methods of Analyzing Light Scattering Data

### 2.6.1 Static Light Scattering

For this study, three main analyses were used for our data of average photocounts per second, or average intensity. These analyses constitute static light scattering. The first analysis starts with a model for the Gibbs free energy of a system of particles given a few parameters, such as those defining the particle shape, and uses the following equation to derive the excess Rayleigh ratio for such a system. This equation is valid for scattering from multi-component isotropic liquid systems (Kirkwood & Goldberg, 1950; Ross *et al.*, 2008).

$$\Delta R = (\frac{\pi^2 k_B T}{\lambda^4}) \nabla_\rho^T \epsilon \cdot H_\rho[g]^{-1} \cdot \nabla_\rho \epsilon \tag{2.5}$$

In this equation, $H_\rho[g]^{-1}$ is the inverse of the Hessian matrix of second partial derivatives of the dimensionless Gibbs free energy $g$ with respect to the number densities of each component. It can be thought of as the 'size' of thermal fluctuations in the sample, which are exactly the fluctuations that lead to scattering. The dimensionless Gibbs free energy is defined as $g = \frac{G}{V}$, which is the usual Gibbs free energy divided by the volume of the system. $\nabla_\rho \epsilon$ represents the gradient of the dielectric constant with respect to the number densities, and the superscript $T$ represents the matrix transpose. These terms connect the fluctuations in $g$ to electromagnetic theory. Thus, fluctuations in the dielectric constant caused by fluctuations in $g$ lead to light scattering. The prefactor contains $\pi$, the ratio of a circle's circumference to its diameter, Boltzmann's constant $k_B$, the absolute temperature $T$, and the wavelength of laser light used, $\lambda$. The specifics of deriving $\Delta R$ from a thermodynamic model are in Appx.(6).

The experimental results are compared to the Boublik and Nezbeda model of mixtures of hard convex bodies (Boublik & Nezbeda, 1986). This model accounts for the statistical behavior of solutions of bodies with a strong repulsive force (such that the particles never overlap), that are composed completely out of convex shapes (e.g. spheres, ellipsoids, drops, etc.). Interactions between the particles and the solvent are not accounted for by this model,

and neither are attractive interactions between proteins. Furthermore, this model neglects orientational contributions to the intermolecular pair potential and only considers the radial separation $r$. A diagram of the intermolecular pair potential for hard-core spherical particles is shown in Fig.(9), and is described by Eq.(2.6) below.

$$V(r) = \begin{cases} \infty & |r| \leq \frac{a}{2} \\ 0 & otherwise \end{cases} \tag{2.6}$$

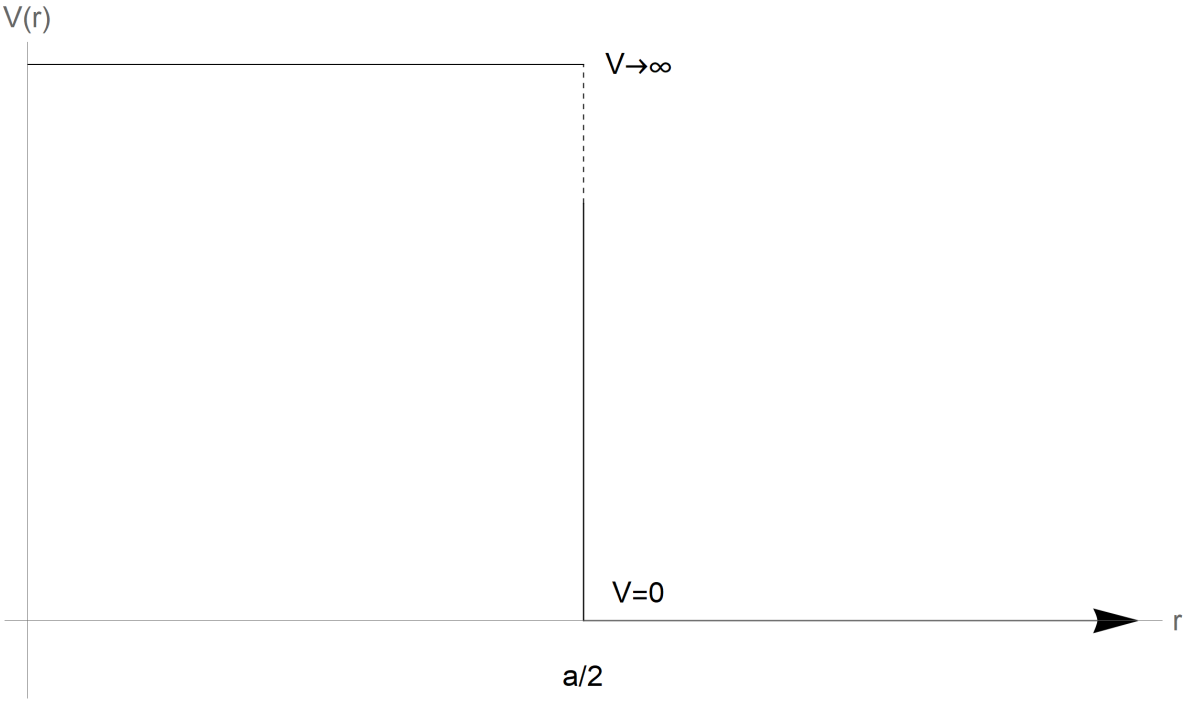



Figure 9: The hard core potential between two bodies of diameter $a$. When $|r| \leq a/2$, the potential is infinite, and therefore the particles can never overlap as overlapping particles experience an infinite repulsive force. When $|r| > a/2$, the particles do not interact.

The mixing free energy model used here, Eq.(4.105) of Boublik & Nezbeda (1986), is given in terms of a compressibility factor:

$$\frac{\beta P}{\rho} = \frac{1}{1-\phi} + \frac{3\alpha\phi}{(1-\phi)^2} + \frac{3\alpha^2\phi^2}{(1-\phi)^3} \tag{2.7}$$

Here, $\phi$ is the fraction of solution volume occupied by all of the proteins, defined as

$\phi = \frac{\Omega_p N_p}{V}$, the volume of a single protein times the number of protein molecules divided by the total solution volume. The shape factor $\alpha$ is defined as $\alpha = \frac{\mathcal{RS}}{3\mathcal{V}}$, the $\frac{1}{4\pi}$ multiple of the mean surface radius of curvature, $\mathcal{R}$, times the average protein surface area, $\mathcal{S}$, divided by three times the average protein volume, $\mathcal{V}$. When $\alpha = 1$, the objects represented by the model are spheres. $\alpha$ is restricted to values greater than or equal to one for convex bodies. The left-hand side is the compressibility factor, composed of $\beta = \frac{1}{k_b T}$, the pressure $P$, and the number density $\rho$. The compressibility factor has the value 1 in the case of the ideal gas. In addition, Eq.(2.7) reduces to the Carnahan-Starling equation of state for hard spheres (Carnahan & Starling, 1969) when $\alpha = 1$. It is important to note that Boublik and Nezbeda studied additional models for the equation of state of convex bodies; analysis of their possible applicability is left for future work.

A critical piece of the derivation of the excess Rayleigh ratio that follows from Eq.(2.7) is that it is assumed that our sample has a single solute component, in this case, $\beta_H$ crystallin. Consideration of the combined effects of polydispersity and non-spherical shapes at high concentrations is also left for future work. The details of the derivation of the excess Rayleigh ratio may be found in Appendix A. The final result is given by Eq.(2.8).

$$\Delta R = \frac{4\pi^2 n^2 (\frac{dn}{dc})^2 m_p^2}{N_A \lambda_o^4} \left[ \frac{1}{\phi(1-\phi)^2} + \frac{6\alpha}{(1-\phi)^3} + \frac{9\phi\alpha^2}{(1-\phi)^4} \right] \tag{2.8}$$

The term in square brackets on the right hand side is the second derivative of our model of Gibbs free energy with respect to the volume fraction $\phi$ of proteins. The leading coefficient is the optical coefficient $K$ times the mass of the proteins, $m_p$, squared. The optical coefficient is equal to $4\pi^2 n^2 \frac{dn}{dc}^2 / N_A \lambda^4$ and relates the size of fluctuations in the volume fraction to the resulting fluctuations in the dielectric constant.

Eq.(2.8) permits direct comparison with excess Rayleigh ratios determined with Eq.(2.2). Measurements of the needed values of $\frac{dn}{dc}$ are presented in Section (4). Because the data are in the form of a list of concentrations with their respective measured excess Rayleigh ratios, concentrations must be converted to volume fractions for use with Eq.(2.8). Concentrations were converted to volume fractions by making an assumption about the volume occupied by

a single $\beta_H$ crystallin. It is assumed that $\beta_H$ has a partial specific volume appropriate for globular proteins, which is typically very close to 0.71 $mL/g$ (Akasaka & Matsuki, 2015)[6]. This value was used because it is similar to that reported for $\gamma$ crystallins (Schurtenberger *et al.*, 1989), which are highly homologous to the $\beta$ crystallins. The conversion is given below in Eq.(2.9) for concentrations, $c$, in milligrams per milliliter.

$$\phi = c\frac{0.71}{1000}\frac{ml}{mg} \tag{2.9}$$

The measurements of $\Delta R$ also permit us to estimate the molecular weight and second osmotic virial coefficient independently of the Boublik and Nezbeda model. The second virial coefficient comes from the virial expansion in number density $\rho$ of the ideal gas law expressed in terms of the compressibility factor.

$$\frac{\beta P}{\rho} = 1 + A_2\rho + A_3\rho^2 + \cdots \tag{2.10}$$

The $A_i$ are the $i^{th}$ virial coefficients. When all $A_i$ are zero, Eq.(2.10) reduces to the equation of state for an ideal gas.

The second virial coefficient may be derived in terms of the configuration integral of the classical canonical partition function, as in the textbook by McQuarrie (1976). To account for asymmetric molecules, one must integrate over all possible orientations of the molecules:

$$A_2(T) = -\frac{1}{32\pi^2}\int \{exp\left[-\beta u(r_{1,2}, \boldsymbol{\omega}_1, \boldsymbol{\omega}_2)\right] - 1\}\, 4\pi\, r_{1,2}^2\, dr_{1,2}\, d\boldsymbol{\omega}_1\, d\boldsymbol{\omega}_2 \tag{2.11}$$

Placing one molecule at the origin simplifies the problem considerably, as implicit in Eq.(2.11). Instead of integrating over the positions $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$, the center to center separation $\boldsymbol{r}_{1,2} = \boldsymbol{r}_2 - \boldsymbol{r}_1$ is used, the magnitude of which is $r_{1,2}$. In the case of cylindrically symmetric molecules, as assumed in Eq.(2.11), each $\boldsymbol{\omega}_i$ denotes the orientation of molecule $i$[7]. The differential of $\omega_i$ is given as $d\omega_i = \sin\theta_i d\theta_i d\phi_i$. $u(r_{1,2}, \boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$ is the potential due to the separation and orientations of the molecules. The term in curly braces is also known as

[6] Commonly reported to be between $0.70 - 0.74\ mL/g$

[7] For spherically symmetric molecules, these integrate to each give a factor of $4\pi$.

the Mayer f-function. It is important to note that the Mayer f-function goes to zero as the separation between particles approaches infinity. It should also be noted that $\beta = \frac{1}{k_b T}$ appears in the integrand in Eq.(2.11), so $A_2$, and therefore $B_2$, are functions of temperature.

Higher order virial coefficients may also be calculated in a similar manner, as shown by McQuarrie (1976).

For the second method of analysis, the data is analyzed in the low concentration regime ($c \lessapprox 40\ mg/mL$). This is the portion of Fig.(13) near the origin that appears linear, as shown in Fig.(14). The Zimm equation (Zimm, 1948) describes scattering from a monodisperse (meaning all the scatterers have the same diffusion coefficient, i.e. they have the same size and shape) solution of particles. For the measurements made for this thesis, we used a scattering angle of $\theta = 90^o$, for which the excess Rayleigh ratio is given to leading order by Eq.(2.12):

$$\frac{\Delta R(\theta)}{Kc} = \frac{M_w^{app}(90)}{1 + 2A_2 M_w^{app}(90)c + \mathcal{O}(c^2)} \frac{1}{1 + \frac{1}{3}(qR_g)^2 + \mathcal{O}(q^4)} \tag{2.12}$$

In this equation, $K$ is the optical coefficient and $c$ is the concentration of a sample. The usual molecular weight term $M_w$, which represents the mass of a single molecule is replaced with $M_w^{app}(90)$. The reason for using this term is that the rigorous weight-average molecular weight $M_w$ would be obtained from measurements which permitted the simultaneous extrapolation of Eq.(2.12) to $q = 0$ and $c = 0$, whereas the current measurements were all done at $90^o$. $q$ is the scattering vector magnitude, defined in Fig.(6), and $A_2$ is the second virial coefficient given in Eq.(2.11). $R_g$ represents the radius of gyration, defined in the text before Eq.(**??**).

If $R_g$ in Eq.(2.12) is neglected and $c$ is allowed to be small, a simplification valid for the low concentration regime known as the Debye equation is obtained, given in Eq.(2.13). If Eq.(2.13) is plotted against concentration, the intercept with the vertical axis represents the inverse of the weight of a single protein. Inverting this and multiplying by Avogadro's number yields the mass per mole of the protein. The slope of this graph yields twice the second virial coefficient. Once again, $A_2$ is made dimensionless by dividing by the quantity 2 times the partial specific volume and multiplying by the mass of a single protein, as indicated by Eq.(2.9).

$$\frac{Kc}{\Delta R(\theta)} = \frac{1}{M_w^{app}(90)} + 2A_2c \tag{2.13}$$

To compare the second virial coefficient to that obtained from other methods or experiments we make it dimensionless with Eq.(2.14) given as

$$B_2 = \frac{A_2 M_w^{app}(90)c}{\phi} \tag{2.14}$$

For error analysis, the standard error of the mean count rate was used to obtain errors in the measured excess Rayleigh ratio of a sample. Error in concentration was estimated at 0.2% of the measured concentration by repeated measurements of the same sample. Earlier experiments may involve more error in concentration, as the technique was not yet refined, however the exact error is difficult to quantify. Directly measuring the concentration at each step of dilution was not practical, as some sample would be lost and dust would potentially be introduced. Since all dilution steps were accomplished with the aforementioned five digit balance, whose precision is greater than this estimated concentration error, we did not expect error to accumulate with each dilution.

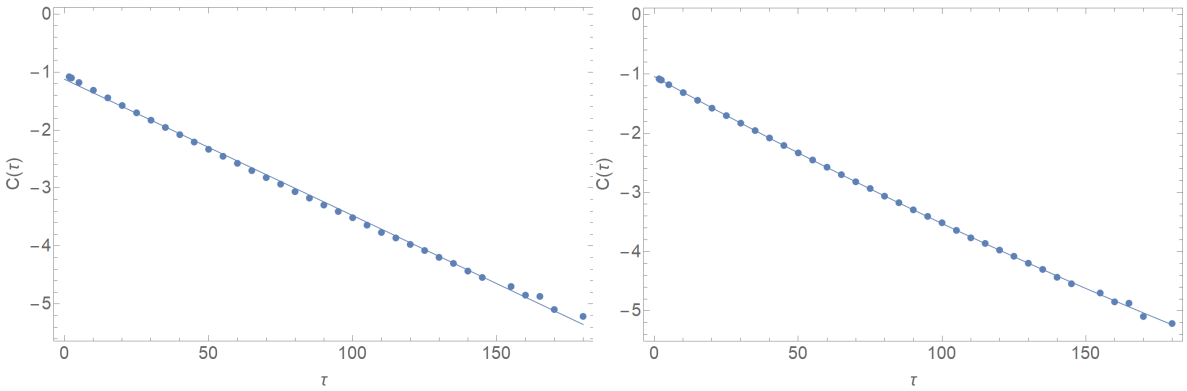
Parameter error was calculated from the 95% confidence intervals for each parameter provided by the fitting routine in Mathematica as one half the absolute magnitude of the difference between the fit value and the maximum value predicted by the confidence interval.

To select between any two models representing the same quantity with different numbers of parameters, the Akaike information criterion (AIC) was used with the correction appropriate for a finite sample size (Akaike, 1974; Hurvich & Tsai, 1989) (AICc). This statistic compares the goodness of fit of a model to how many parameters the model used, rewarding fewer parameters and stronger fits. The equation used for calculating the AICc is presented in Eq.(2.15),

$$AICc = 2k - 2\ln\hat{L} + \frac{2k^2 + 2k}{n - k - 1}, \tag{2.15}$$

where $k$ is the number of parameters used in the model, $n$ is the number of data points, and

$\hat{L}$ is the maximized value of the likelihood function of the model. The likelihood function is an expression of the probability of measuring a value in a data set within a range of values predicted by a model (Ward & Ahlquist, 2018). The model whose AICc value is the smallest is selected. Fig.(10) describes an example of this. Note that the absolute magnitude of the value is not important, eg. a model with a more negative AICc than another is more desirable.




(a) (b)

Figure 10: These data are from cumulant fits to quasielastic correlation functions obtained from a sample of $\beta_H$ crystallin. It is clear to see that the quadratic fit in Fig.(10b) fits the data better than the linear fit in Fig.(10a). The reported AICc also suggests that the quadratic fit is the better fit. $DW$ for the linear fit is reported to be 0.4, indicating positive correlation of the residuals. $DW$ for the quadratic fit shown is 3.2, indicating negative correlation in the residuals. However, it is evident that the last five points contribute to alternating residuals; omitting these points lowers $DW$ to 1.7, more typical of a good fit. The cubic fit (not shown) did not improve significantly upon the quadratic fit.

If two different models report nearly the same AICc, the one with fewer parameters was chosen, as in Fig.(11). Additionally, the models selected did not change on using the AIC without the finite sample correction, however AICc is reported as it is more appropriate for the sample. For an infinitely sampled process, the third term in Eq.(2.15) vanishes and the usual expression for the AIC is recovered.

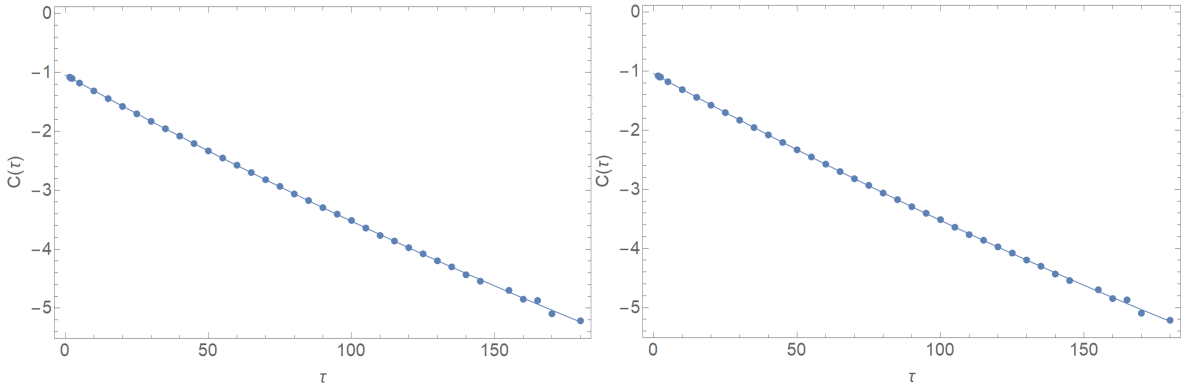



Figure 11: Comparison of quadratic and cubic fits which are nearly identical and whose AICc values are also nearly the same. These data are from the cumulant fit to quasielastic correlation functions obtained from a sample of $\beta_H$ crystallin. Quadratic and cubic fits often report similar AICc for this type of data, but the quadratic fit was selected to reduce the number of model parameters.

As another discriminator between models, the Durbin-Watson test statistic (Durbin & Watson, 1950, 1951) was calculated. This test measures the autocorrelation of successive residuals. The ideal model would have no correlation between residuals, so a significant correlation in the residuals can indicate a statistically significant error in the fit. The Durbin-Watson statistic is calculated from the residuals as in Eq.(2.16), where $r_i$ represents the $i^{th}$ residual, and $T$ is the total number of data points.

$$DW = \frac{\sum_{i=2}^{T}(r_i - r_{i-1})^2}{\sum_{i=1}^{T} r_i^2} \tag{2.16}$$

A value of $DW$ close to 2 indicates there is little to no significant correlation between the residuals. When it is less than 2, it indicates a positive correlation, and when it is greater, a negative correlation.

### 2.6.2 Quasielastic Light Scattering

In addition to the average photocounts per second, all experiments also measured a homodyne (also called the self-beating) (Berne & Pecora, 2000; Chu, 2007) intensity-intensity correlation function, which is defined as:

$$G^{(2)}(\tau) = \langle I(t)I(t+\tau)\rangle \tag{2.17}$$

Here, $I$ denotes the intensity measured at time $t$ and the bracket represents a time-average over small $t$ for fixed $\tau$. A very short time after taking a measurement (here $\tau \sim 100\ ns$), one should expect to measure a value of $I$ similar to the starting measurement. The longer one waits to make a second measurement, the less sure one can be about a prediction of the value of the second measurement. If one waits very long between measurements relative to the time-scale relevant for thermodynamic fluctuations, and averages over increasing delay times, one obtains the square of the average photocounts per second. Fig.(12) outlines this concept for a general non-periodic process.

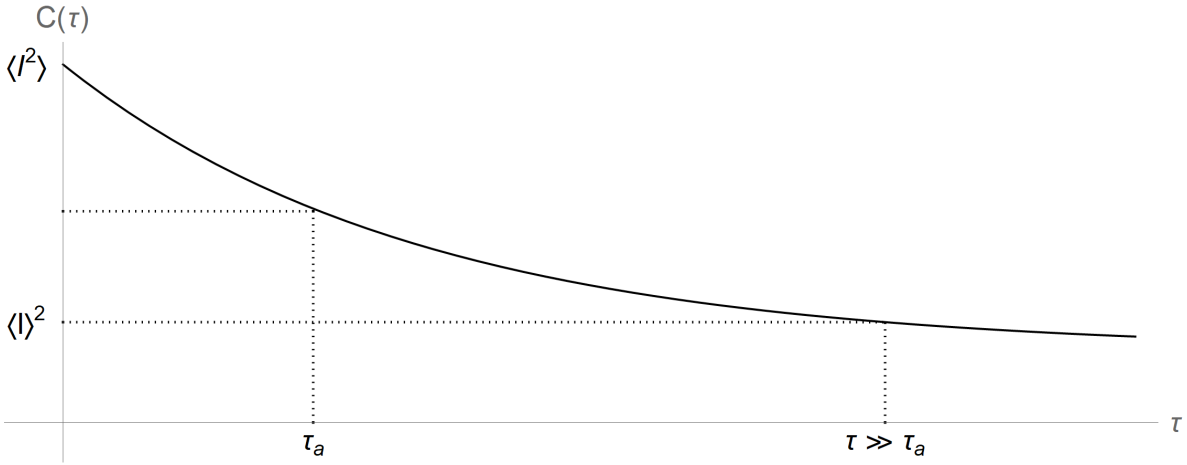



Figure 12: A time correlation function for a general non-periodic process. The correlation function begins with an initial value $< I^2 >$ and decays over time to $< I >^2$ which is the square of the baseline intensity.

In the case of a single diffusing species, Eq.(2.17) becomes, where A is the signal amplitude in excess of the baseline:

$$G^{(2)}(\tau) = \langle I(t)I(t+\tau)\rangle = \langle I\rangle^2[1 + Ae^{-2q^2D\tau}] \tag{2.18}$$

Two methods were primarily used to obtain the slope from the correlation function, the cumulant analysis and a direct fit to a squared sum of exponential functions (Berne & Pecora,

2000; Chu, 2007). The cumulant analysis attempts to find the moments of the diffusion coefficient of the particles in a sample. This is equivalent to finding the moments of the size distribution of the particles in the sample via Eq.(2.26). We follow the method provided by Koppel (1972) laid out in Berne & Pecora (2000), and present the relevant equations and parameters here.

First, the correlation function is normalized by subtracting the baseline (the square of the average counts per second) and dividing by the same. This yields

$$N(q,\tau) = g_2(q,\tau) = \frac{\langle I(q,\tau)I(q,0)\rangle}{|\langle I(q,0)\rangle|^2} = \langle e^{-q^2 D\tau}\rangle \tag{2.19}$$

Here, two different notations are provided for the same concept to assuage any confusion. Taking the natural logarithm and expanding in a power series in $\tau$, it is found

$$ln(N(q,\tau)) = -K_1\tau + \frac{1}{2}K_2\tau^2 - \frac{1}{3!}K_3\tau^3 + \mathcal{O}(\tau^4) \tag{2.20}$$

Where the $K_n$ are the so-called cumulants of the distribution. They are defined in terms of the scattering vector magnitude, $q$, and the collective diffusion coefficient $D$.

The formula for calculating the cumulants is readily found in Berne and Pecora (Berne & Pecora, 2000) or a similar text. The first few cumulants are provided here, as those are the ones most relevant to this study.

$$K_1 = \langle q^2 D\rangle \tag{2.21}$$

$$K_2 = \langle (q^2 D - \langle q^2 D\rangle)^2\rangle \tag{2.22}$$

Note that $K_2$ is related to the variance in the diffusion coefficient. Additionally, it may be assumed that the polarizability per unit mass is the same for every particle in solution. This is a reasonable assumption to make for the samples, as the relevant species are all $\beta$ crystallin and therefore should all have nearly the same polarizability per unit mass[8]. This

[8]Recall that $\beta_H$ and $\beta_L$ are assembled from the same subunits.

alters slightly the definition of the above cumulants where we need to introduce a 'z-averaged' (or intensity-weighted) diffusion coefficient in Eq.(2.23) given by Koppel (1972). $N_i$ and $m_i$ are the number and mass of the species $i$ in a sample with diffusion coefficient $D_i$.

$$\langle D \rangle_z = \frac{\sum_i N_i m_i^2 D_i}{\sum_i N_i m_i^2} \tag{2.23}$$

$$K_1 = q^2 \langle D \rangle_z \tag{2.24}$$

$$K_2 = q^4 \langle \delta D \rangle_z \tag{2.25}$$

The decay of the correlation function is related to the rate of diffusion, or diffusion coefficient, of a particle through the solvent medium (Berne & Pecora, 2000; Chu, 2007). Using the Stokes-Einstein equation (McQuarrie, 1976) given below, the diffusion coefficient can be converted to a hydrodynamic radius when the sample is sufficiently dilute. For samples of higher concentration, this relationship may still be used, however the diffusion coefficient no longer represents the self diffusion, but the collective diffusion. Therefore, the hydrodynamic radius becomes a sort of average.

$$R_h = \frac{k_b T}{6\pi\eta D} \tag{2.26}$$

When performing fits to correlation function data, it is often necessary to identify and remove 'dusty' data. First, the count rate history of each measurement was examined for the presence of dust moving in and out of the scattering volume, evidenced by spikes in measured intensity. Large spikes well above the expected standard deviation of the mean in Poisson statistics dramatically change the correlation function, sometimes completely ruining the quasielastic data for a particular measurement. Such observations were omitted from the final fits and ensuing results. Second, not all of the correlation functions were well measured over the entire range of lag times. Frequently, the lag times closer to the baseline experienced

greater levels of variation, as did the few channels at extremely small lag times. Omitting these channels often improved the fit by reducing residual error and bringing the results closer in line to those obtained from other measurements.

The presence of two or more unlike scatterers in a light scattering experiment leads to multiple slopes in the intensity-intensity correlation function. To account for these, fits were attempted with the squared sum of two exponentials, the form of which is given below, which assumes there are two unlike scattering species present. This assumption is reasonable to make if it is expected that there are only two unlike species in a sample. For the experiments, these would be $\beta_H$ crystallin, with the other term representing perhaps $\beta_L$ crystallin or another unknown scattering species. It was expected that there would be a small presence of $\beta_L$ based on chromatography fractions.

$$\left(a_1 e^{-a_2 \tau} + (1 - a_1) e^{-b_2 \tau}\right)^2 \tag{2.27}$$

Unfortunately, it was difficult to systematically analyze the correlation functions with this method. Fitting Eq.(2.27) to a correlation function constitutes an ill-posed problem, as there are potentially a number of non-unique solutions for the parameters. For the purposes of this thesis, analysis with this method was forgone, but its description remains above for completeness.

The quantities obtained from light scattering are summarized in Table (3).

Table 3: Quantites Obtained from Light Scattering

| Analytical Regime | Analysis Type | Results |
|---|---|---|
| High Concentration Static Light Scattering | Boublik and Nezbeda | $M_w$, $\alpha$ |
| Low Concentration Static Light Scattering | Zimm/Debye | $M_w$, $B_2$ |
| Quasielastic Light Scattering | Cumulant | $R_h$, $D_{collective}$ |

## 2.7 Data Rejection

Not all of the experiments performed yielded usable data. The primary reason data were rejected was because the experiment did not cover the entire desired concentration range from high concentration to low concentration. Without data in both regimes, comparisons of the various derived quantities could not be made. Some experiments were halted before the entire concentration range was covered, as the presence of dust (often large protein aggregates) made it infeasible to continue. Out of 6 experiments that attempted to cover the whole concentration range, only 4 covered enough or were consistent enough to include in this thesis.

Within each experiment, some data were also rejected. The presence of dust, changes in temperature or atmospheric pressure, variations in laser intensity, and many other factors played a part in moving the average photocounts per second away from its 'true' value. Single spikes in intensity caused by dust moving in and out of the scattering volume were relatively easily dealt with for determining an average, as the measurement could be truncated around the spike, effectively erasing it. An average that changed significantly over time was immediate cause for rejecting data, as there was no systematic way to account for this.

## 2.8 A Note on Technique

After months of performing experiments, not all of which produced viable data capable of being analyzed, it became apparent that good and careful technique is critical to performing consistent light scattering experiments. There were many areas where small improvements to technique could lead to large savings in time or error. For example, practice with cleaning and loading samples into light scattering tubes introduced less dust initially and with each subsequent dilution. Practice was also necessary to obtain reliable concentration measurements, and further experience allowed one to better determine whether collected data was acceptable for analysis. Perhaps most critically, with practice it became possible to perform an entire light scattering experiment in a single day, rather than over the course of several. This helped reduce variations in the sample caused by the delay time between experiments, where there was ample time for the protein to change in a hard-to-predict way.

To reiterate, two major factors played a part in the consistency of the results of an experiment. The first are steps taken to mitigate sample exposure to dust. After many experiments, it became clear that the most dust was introduced during the dilution steps. Variations on this step, including extra filtration and injection via syringe through plastic film, greatly helped reduce the concentration of dust in a sample. The other major factor was alacrity in performing the experiment. Taking several days to perform an experiment meant the sample had to go in the refrigerator ($\sim 4^oC$) or stay on the lab bench ($\sim 20^oC$) overnight. Performing an experiment in a single day had the advantage of limiting these time or temperature dependent effects. However, it also had the disadvantage that measurements had to be made quicker and with less regard to the correlation functions. That is because dust had a greater effect on the correlation function than on the average photocounts per second, and it was often infeasible for measurements to be retaken to secure apparently dust-free lengths of time.

This is all to say that the variations between experiments can be explained (but not yet quantified) by three things. The first is improvement in technique and dust mitigation. The second is time from dissection to experiment, particularly the time taken to do the experiment once the sample is at high concentration. Lastly, and perhaps most importantly, variations are caused by the proportion of $\beta_H$ to $\alpha$ and $\beta_L$ crystallin. This changed from experiment to experiment due to inconsistencies in the SEC process. Rechromatography may be used to further isolate $\beta_H$ but comes with the added costs of aggregation and time.

# 3 Experiments and Data

Here we describe the light scattering experiments performed on $\beta_H$ crystallin.

## 3.1 Beta High Light Scattering

### 3.1.1 Dilution Series

Four dilution series light scattering experiments were carried out on $\beta_H$ crystallin for which measurements were made over a large concentration range. For an overview of dilution series light scattering experiments, see Sec.(2). In the following paragraphs, any particularities about each of the experiments are explained. These data were analyzed over both the whole concentration range and at low concentration ($\lesssim 40\ \frac{mg}{mL}$). All of the experiments took place at a scattering angle of $90^o$ and a temperature of $25^oC$. The laser was set to emit with 25 $mW$ of power, unless otherwise stated. A pinhole size of 100 $\mu m$ was used to limit the light reaching the detector to result in fewer than two hundred thousand counts per second. For each sample, an appropriate average photocounts per second (static light scattering) was measured. Correlation functions (quasielastic light scattering) were also measured, however only Experiment A resulted in viable correlation functions.

**Experiment A** was performed on a sample of $\beta_H$ crystallin in July of 2021. This sample was exchanged into the 100 $mM$ sodium phosphate buffer described in the methods section. Exchange into this buffer allows for comparison with other studies, which commonly use this buffer (Berland *et al.*, 1992). This was the first experiment performed on $\beta_H$ crystallin for this study, and this inexperience may have lead to some of the variations in the data, as discussed briefly in the beginning of this section. This experiment covered a range of concentrations from $321 - 4.9\ mg/mL$, extending well past the point where short range order dominates the system.

**Experiment B** was the final experiment performed on $\beta_H$ crystallin for which the sample was exchanged into the 100 $mM$ sodium phosphate buffer. This experiment was performed in September of 2021. The concentration range covered by this experiment was $223 - 0.99\ mg/mL$. After this experiment, it was decided to skip the buffer exchange, as

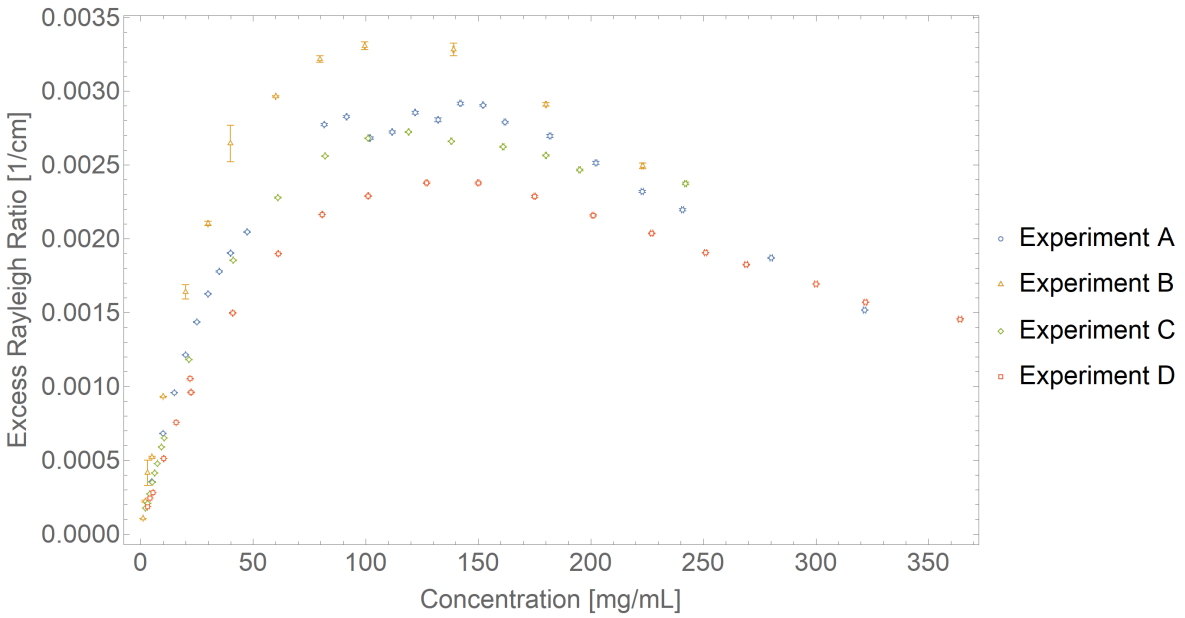



Figure 13: Measured excess Rayleigh ratio vs. concentration for four experiments on $\beta_H$ crystallin. Fits are described later and shown in Fig.(21). Note that the symbols are small to accommodate the error bars.

it appeared the excess time and exposure to stirring and changing concentrations in the Amicon was often leading to unacceptable levels of aggregation. Removing the aggregates via centrifugation or filtration was necessary to continue the experiment. Siezen *et al.* (1986) has shown that $\beta_H$ crystallin becomes less stable upon separation from other lens components, and great care was taken to eliminate any other sources of dust. It was unlikely that the 'dusty' data was caused by actual dust, but rather was caused by large protein aggregates.

**Experiment C** was performed on a sample of $\beta_H$ that was rechromatographed, as described in Section (2), and measured in the 50 $mM$ sodium phosphate buffer, also described in Section (2). This experiment was performed in May of 2022 and covered a concentration range of $242 - 2.04\ mg/mL$. Rechromatography had the effect of increasing the relative concentration of $\beta_H$ in the sample compared to primarily $\beta_L$ and small proportions of other lens proteins. It was difficult to obtain decent average intensities for this sample, as it appeared that the sample accumulated dust or aggregates quite quickly (Siezen *et al.*, 1986). This experiment was not performed all in one day, and after a night in the refrigerator, fibre-like protein aggregates would appear and the data appeared more dusty. These aggregates were removed

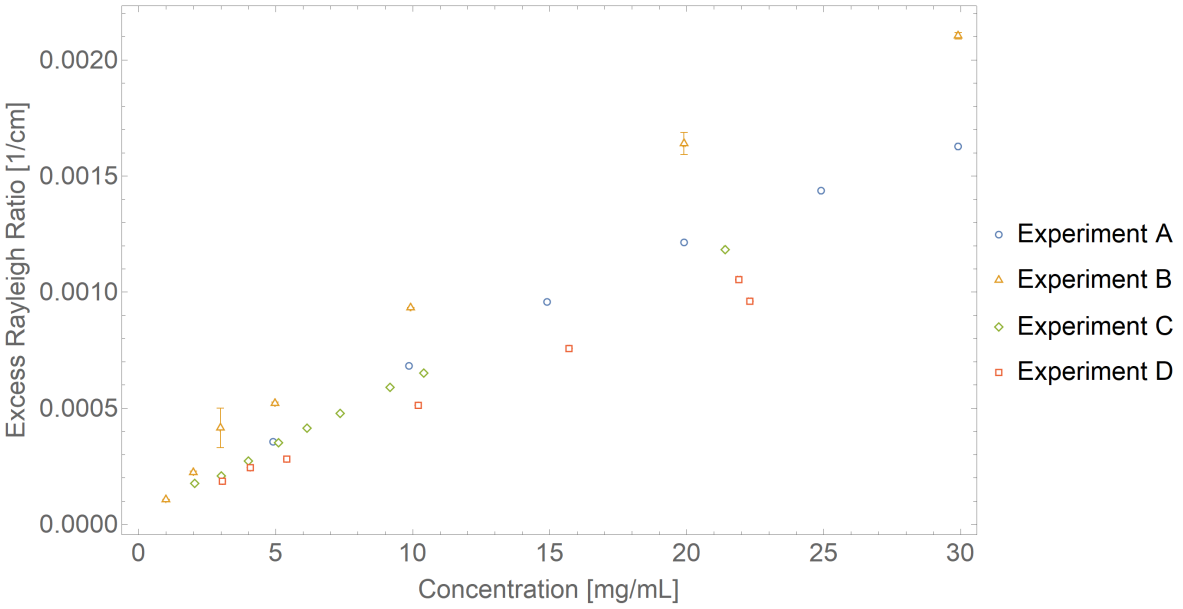



Figure 14: Same as Fig.(13) but zoomed into the region of the graph whose concentration is below 30 $\frac{mg}{mL}$. Fits are described later and shown in Fig.(22).

via centrifugation and the concentration remeasured. It was later decided that attempts to purify $\beta_H$ via rechromatography were not worth the increased instability of the sample. More work should be done, however, to investigate this fully.

**Experiment D** was performed on a sample of $\beta_H$ crystallin in June of 2022, also measured in the 50 $mM$ sodium phosphate buffer. As the latest of the experiments performed, much experience was gained in the handling and treatment of the sample to ensure stability and cleanliness. Additionally, the experiment was performed entirely in one day, avoiding any solubility changes that may be associated with time in the refrigerator and increased experiment time. The concentration range covered by this experiment was $338 - 5.7$ $mg/mL$. The light scattering tubes were sealed with plastic film for the duration of the experiment, and buffer was injected through the seal with a clean syringe needle. This greatly reduced the amount of dust apparently introduced at each dilution step compared to previous experiments, and assisted in being able to perform the experiment in a single day. Nevertheless, there was enough 'dust' to alter the quasielastic light scattering, while the part of the signal affected by dust could be eliminated for the purpose of finding the underlying static light scattering

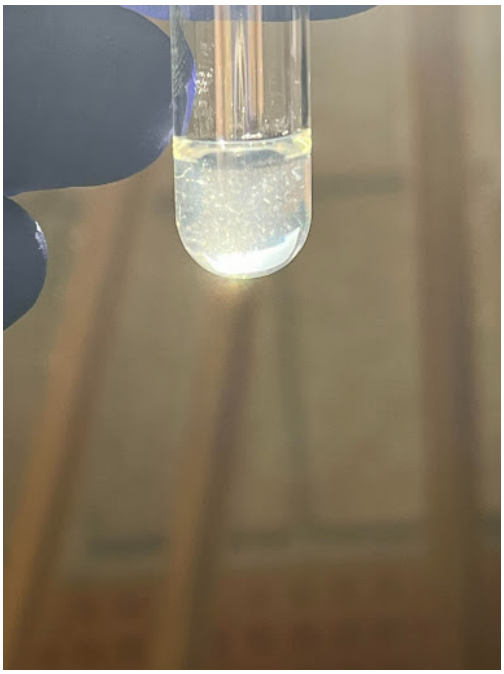


Figure 15: A picture of the aggregates that formed in a solution of rechromatographed $\beta_H$ crystallin described in Experiment C. It is impossible for aggregates of this size to pass through the filters used for cleaning dilution buffer, so it is most likely that these formed in solution from dissociating $\beta$ subunits.

intensity.

## 3.2 Quasielastic Light Scattering

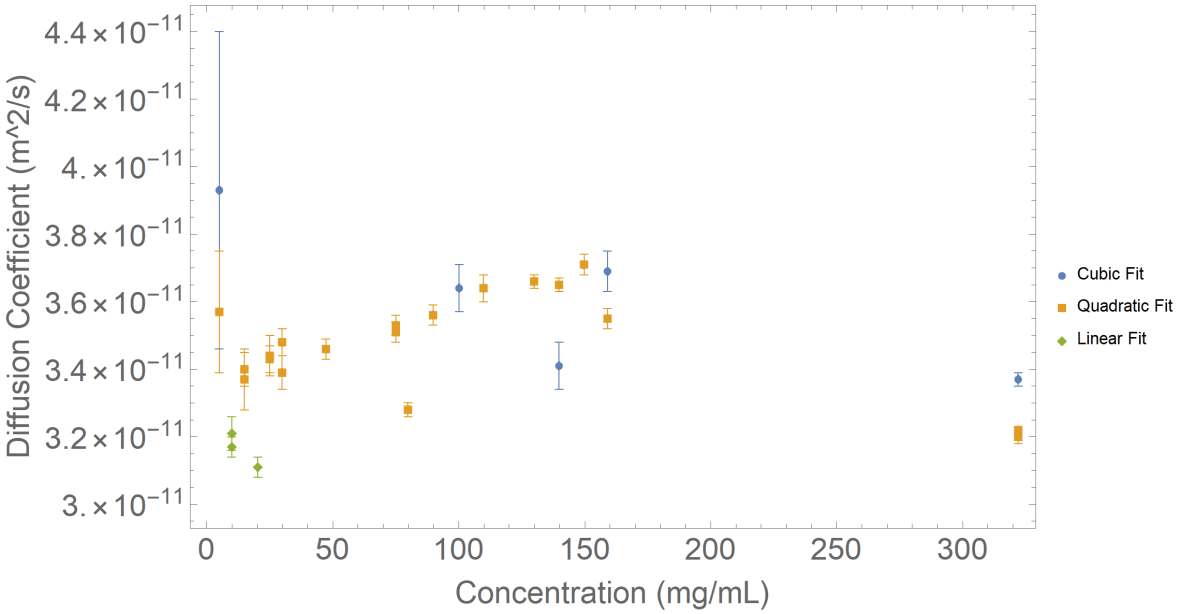



Figure 16: Quasielastic light scattering results for Experiment A before data rejection described in the text. Values for the diffusion coefficient were obtained from the cumulant analysis with varying numbers of parameters, indicated by the legend. Fit type was selected by using the Akaike information criterion (Eq.(2.15)) and the Durbin-Watson statistic (Eq.(2.16)).

Intensity-intensity correlation functions were measured simultaneously with average counts per second for as many measurements as possible. Correlation functions were rejected if the count rate history showed spikes in intensity or if a steady average was not measured. Unlike with static light scattering analysis methods, the correlation function could not be adjusted after an experiment to account for dust spikes, and since the lag times are much shorter than the averaging times in the count rate history the correlation function could not be rebuilt. Nearly every measurement from Experiments B, C, and D were rejected, thus only the correlation functions from Experiment A were analyzed. Fig.(16) shows a summary of the analysis where the method of cumulants (see Section (2)) was used to determine the diffusion coefficients.

## 3.3 Exploratory Low Concentration Studies of Beta Crystallin

After SEC, $\beta_H$ crystallin elutes from the column at $\sim 1-2\ mg/mL$. Static light scattering dilution series experiments were performed on tubes in the $\beta_H$ fraction (see Fig.(3)). The purpose of these experiments was to see how the measured average molecular weight changes throughout the fraction as a result of changing concentrations of $\alpha$ and $\beta_L$ crystallin. As anticipated, the molecular weight increases with greater proportions of a heavier species and decreases with greater proportions of a lighter species. This experiment served to verify qualitatively that fitting Gaussians as in Fig.(3) was an adequate way of showing the relative proportions of each species in each fraction.

More interestingly, upon dilution in these experiments, a reduction in molecular weight was noted at concentrations between $0.4-0.6\ mg/mL$, as shown in Fig.(17). It was chosen to investigate this further by performing light scattering experiments on whole pooled fractions of $\beta_H$ crystallin. This concentration dependent change in apparent molecular weight has been well noted in the literature, primarily by Bateman & Slingsby (1992), Bindels *et al.* (1981), and Siezen *et al.* (1986). However, the quantitative nature of the concentration dependence has not yet been elucidated in the literature. Thus, it would be interesting to follow up on this preliminary experiment in the future.

Dilution series light scattering experiments were performed on the whole mixed $\beta_H$ fraction rather than individual tubes, shown in Fig.(18). Furthermore, this experiment was repeated, whereas the experiments on the individual tubes were not. Light scattering tubes were prepared as described in Sec.(2). The dilute $\beta_H$ solution was passed through a $0.2\ \mu m$ syringe filter to diminish the concentration of dust. The concentration was measured with a UV spectrophotometer. The different colors represent differing settings for light scattering. These settings were the pinhole size and laser intensity, which were changed to keep the measured average photocounts per second below $\sim 200\ kCPS$, but also so that the sample count rate was significantly above the buffer count rate. Increasing the laser intensity or widening the pinhole increases the count rate, but reduces the ratio of the initial maximum ($G_2(0)$) to baseline ($G_2(\infty)$). It is interesting that the effective signal-to-noise ratio remains the same (Clark

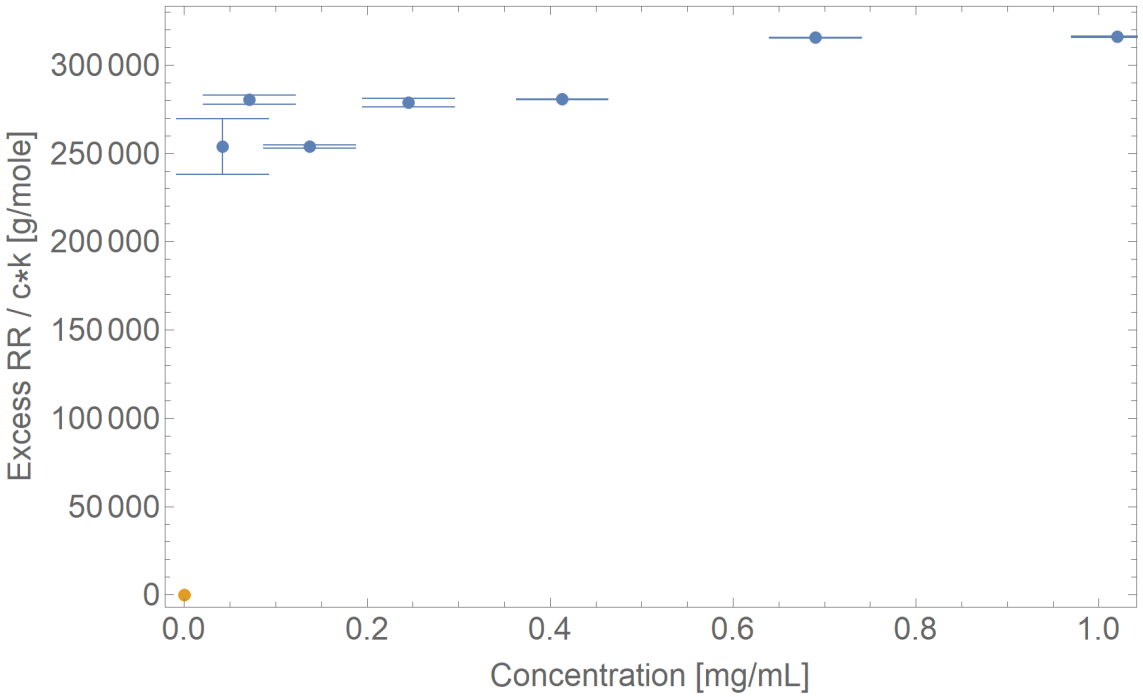
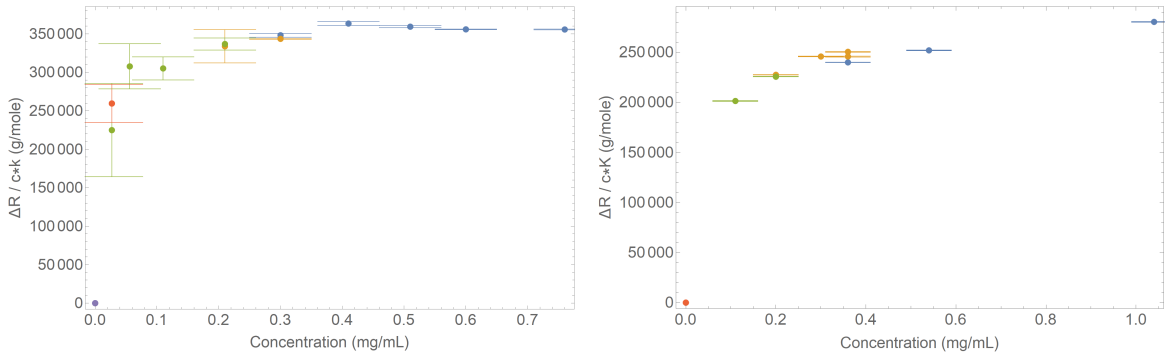



Figure 17: Low concentration static light scattering analysis of $\beta_H$ crystallin on an individual SEC fraction. Here the focus is on the apparent change in molecular weight. The point at the origin is to fix the scale of the graph so it is more easily compared with others.

*et al.*, 1970).




(a) (b)

Figure 18: Molecular weight versus concentration for low concentration $\beta_H$ crystallin. The two datasets were obtained from samples taken from the same pool of $\beta_H$ about one week apart. The different colors represent differing settings for light scattering, see text for details. Perhaps most provocatively, the changes in molecular weight appear to be in integer multiples of $\sim 25\ kDa$, consistent with the average weight of crystallin subunits comprising $\beta_H$. See Table (2) for subunit molecular weights.

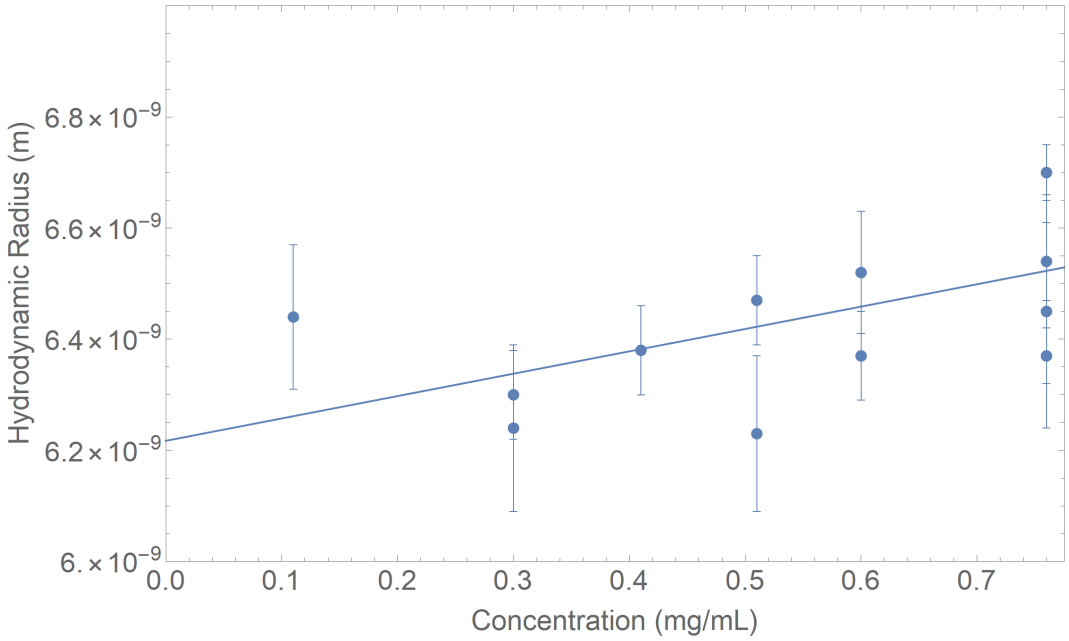



Figure 19: Hydrodynamic radius vs. concentration for low concentration $\beta_H$ crystallin. These data were obtained simultaneously with those from Fig.(18a), however many of the data were rejected. The fitted line is to indicate the trend in the data. The fit was also weighted to account for the errors. Repeated measurements were not averaged and are thus plotted directly.

It was found that the hydrodynamic radius in this extremely dilute regime approaches a

value smaller than that measured at higher concentrations, see Fig.(23) for further discussion.

Experiments in this regime allow for a direct comparison of the molecular weight obtained from static light scattering analysis to the hydrodynamic radius obtained from quasielastic light scattering analysis. As described in Sec.(2), this allows for a constraint on the possible shapes allowed for a protein of a given density.

## 3.4 Refractive Index Increment

The refractive index of solutions changes with concentration. It is critical to measure the index of refraction for various known concentrations of protein to obtain the refractive index increment with concentration, $\frac{dn}{dc}$, which enters into the optical constant, $K$. Previously, studies done in this lab have measured a value of 0.17 $\frac{mL}{g}$, however, this needed to be confirmed. A Bausch and Lomb Abbe 2L Refractometer was used to measure the refractive index of samples of $\beta_H$ crystallin at $25^oC$, which was maintained with a temperature controlled water bath. These were plotted versus concentration, and linear and quadratic functions were fitted to the data. The slope of these functions yielded the refractive index increment. The initial experiment began at low concentration ($\leq 20$ $\frac{mg}{mL}$), and initially it appeared that a quadratic would best fit the data. However, in subsequent experiments much higher concentrations were included which made it apparent that a linear function best fit the data. This was confirmed with use of the Akaike information criterion. Furthermore, the intercept provided by the linear fit was close to the value of the refractive index measured here for pure 50 $mM$ sodium phosphate buffer, which was $1.3347 \pm 0.0004$. The final fit is shown in Fig.(20) along with the measured refractive index increment. Systematic errors are difficult to quantify but may include changes in concentration due to effects like the degree of evaporation of the small sample before measurement. However, it only took seconds to prepare a sample on the refractometer, which was cleaned carefully between measurements. Due to the fragile nature of the instrument, this cleaning step had to be done with caution, thus there may have been some cross-contamination between samples.

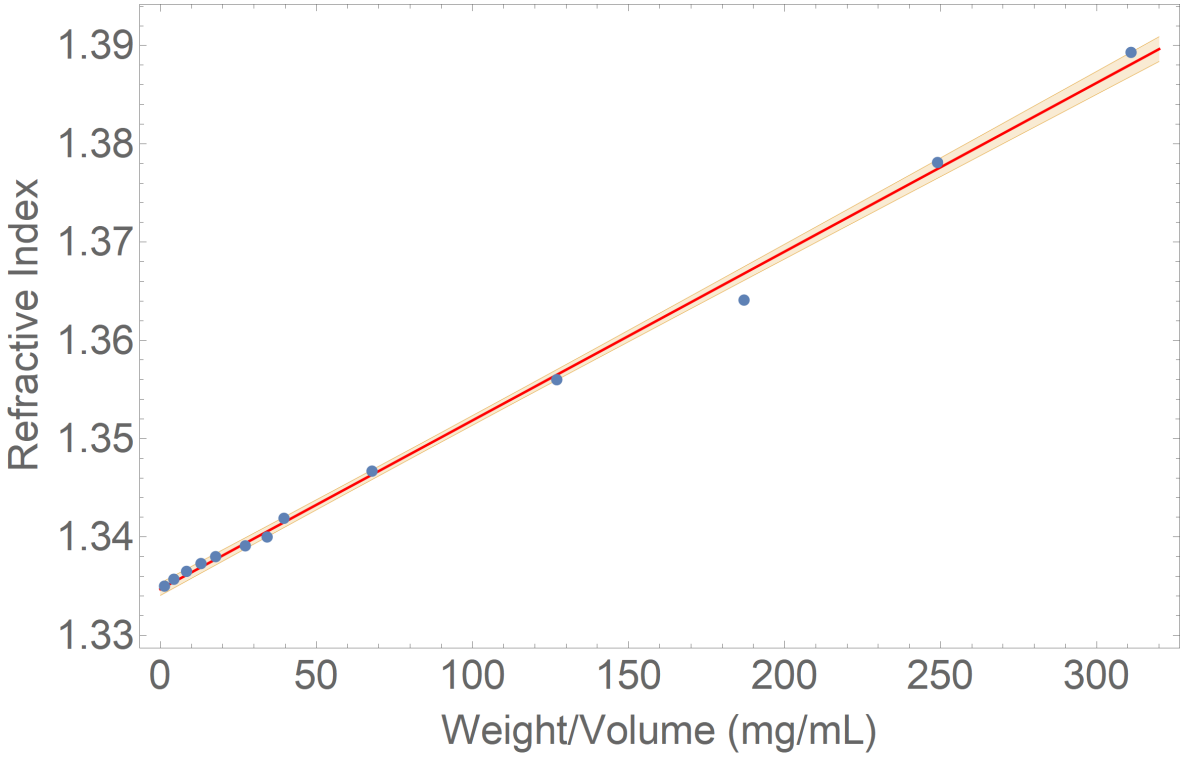



Figure 20: The measured refractive index increment $\frac{dn}{dc}$ was found to be $(0.172 \pm 0.003)\ \frac{mL}{g}$ for the samples of $\beta_H$ crystallin produced here. The red bands represent the 90% confidence interval for the fit.

# 4 Results From Model Fitting

Here the parameters that follow from the experiments discussed in Section (3) are presented. These include the particle shape parameter $\alpha$ introduced in Section (2), the weight-average apparent molecular weight ($M_{app}(90)$), the dimensionless second osmotic virial coefficient ($B_2$), and for Experiment A, the hydrodynamic radius ($R_h$) as a function of low concentrations, as well as the collective diffusion coefficient for all concentrations. The table below summarizes these parameters. The mixing free energy model used in this work is one described by Boublik & Nezbeda (1986) for fluid mixtures of hard convex particles at low concentrations as well as high concentrations (see Eq.(2.8)). In the table, subscripts ZD for Zimm/Debye and BN for Boublik and Nezbeda are included to facilitate later discussion of these results, which appears in Section (5).

| | Dilute SLS | | | Full Range SLS Model | |
|---|---|---|---|---|---|
| Exp. | $M_{app,ZD}(90)$ (kDa) | $B_{2,ZD}$ | Exp. | $M_{app,BN}(90)(kDa)$ | $\alpha_{BN}$ |
| A | 380±3 | 10.3±0.3 | A | 358±2 | 2.2 |
| B | 590±33 | 17±3 | B | 510±4 | 2.8 |
| C | 370±14 | 14±3 | C | 322±2 | 2.0 |
| D | 290±14 | 11±3 | D | 256±2 | 1.8 |

Table 4: The results of analysing the light scattering data obtained from $\beta_H$ crystallin in both the low and high concentration regimes. Note that $B_2$ and $\alpha$ are related for hard convex bodies as $B_2 = 1 + 3\alpha$.

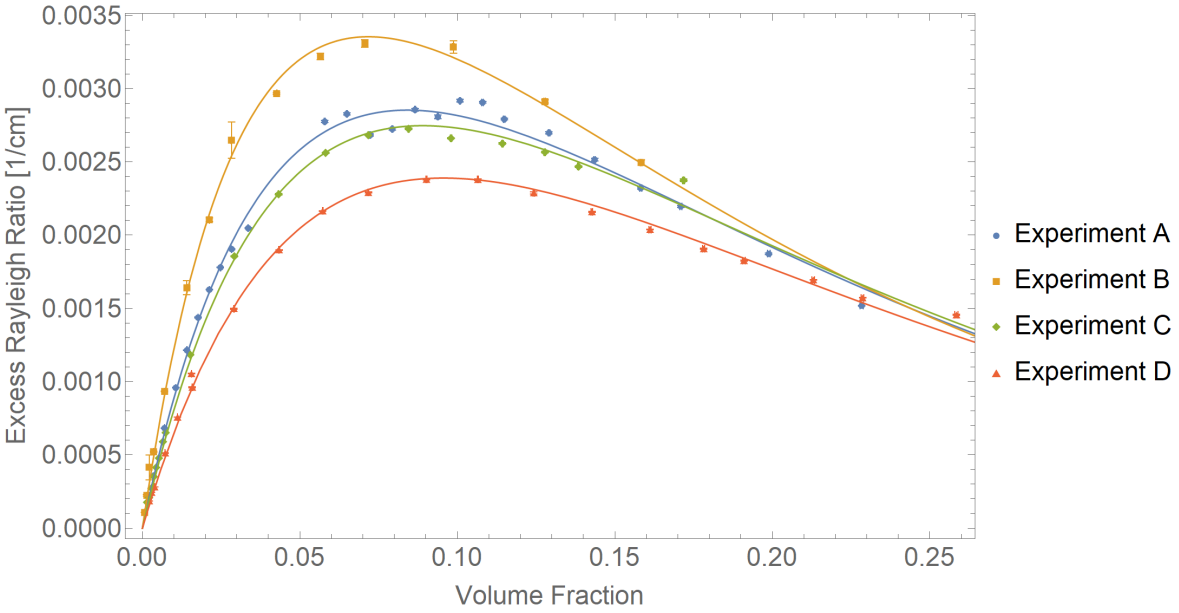



Figure 21: A graph of excess Rayleigh ratio versus volume fraction for several experiments on beta high crystallin. Points are the measured values and the curves are fits to a Boublik and Nezbeda model of hard convex bodies, see Eq.(2.8). Each curve shows the characteristic increase in scattering intensity in the low concentration regime until interactions between proteins become significant enough to reduce the scattering.

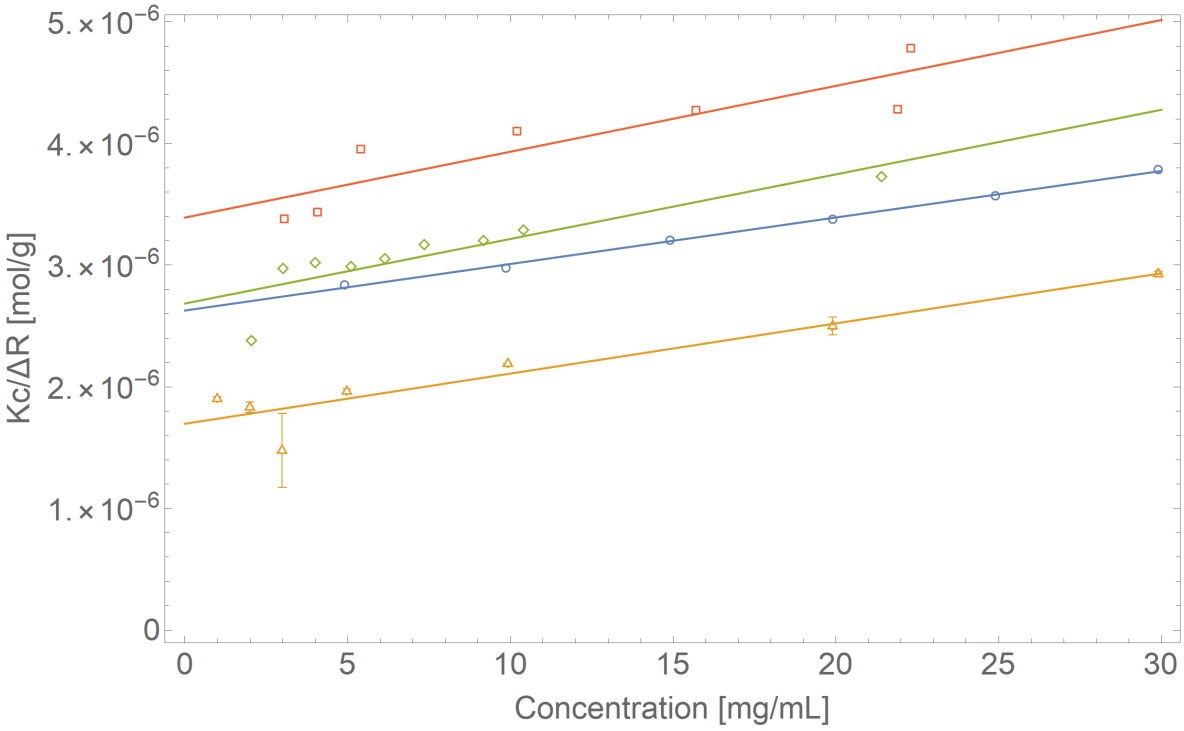
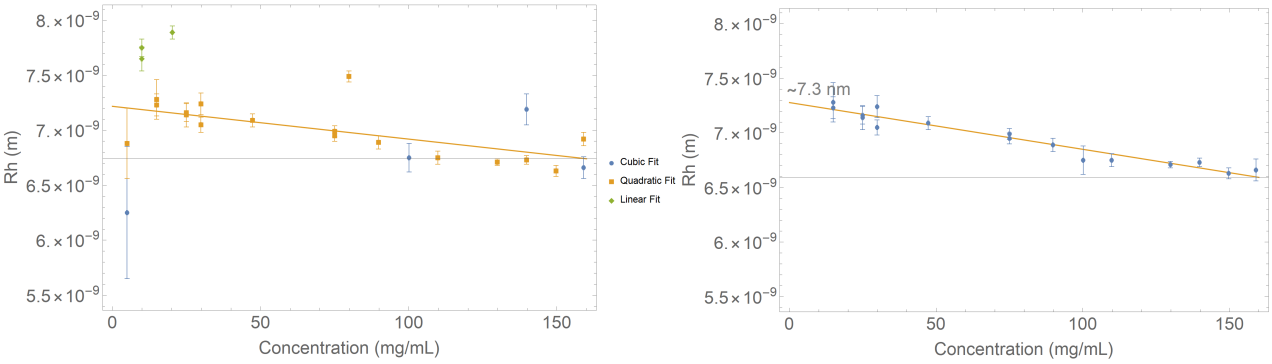
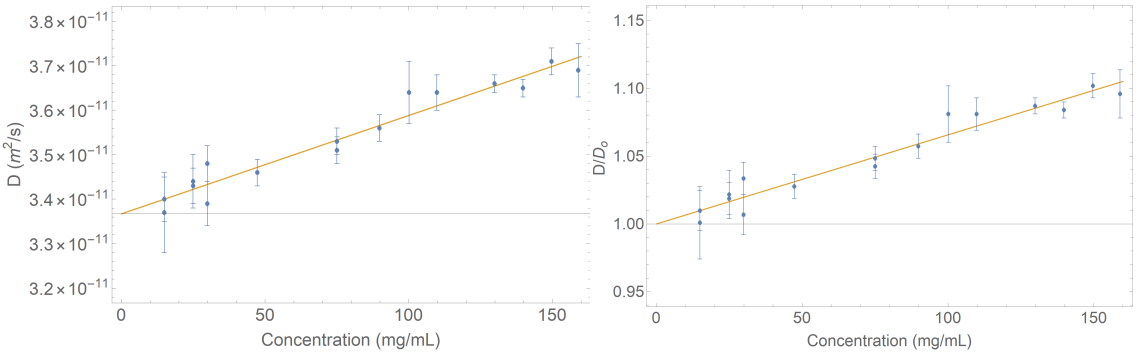



Figure 22: The same data as in Fig.(21) truncated to concentrations below 30 $\frac{mg}{mL}$, in the form of $\frac{Kc}{\Delta R}$, as shown in Sec.(2), specifically Eq.(2.12).




(a) (b)

Figure 23: Results from quasielastic light scattering Experiment A on $\beta_H$ crystallin. (a): The results from before filtering bad data as described in Section (2). (b): The results after omission of measurements compromised by dust. The omission of the bad data from the first figure allows for a much more robust linear fit. Hydrodynamic radii at higher concentrations are apparent, and are more correctly expressed in terms of the collective diffusion coefficient, see Fig.(24).




(a) (b)

Figure 24: Diffusion coefficients, $D$, from quasielastic light scattering Experiment A on $\beta_H$ crystallin. (a): $D$ vs concentration. (b): $D/D_o$ vs concentration.

Results at higher concentrations should not be directly interpreted as the hydrodynamic radius, but as the collective diffusion coefficient. They are included here to show how the hydrodynamic radius arises. The equation of the linear fit as a function of concentration in Fig.(23b) is

$$\left(7.3 \times 10^{-9}\right) m - \left(4.28 \times 10^{-12}\right) \frac{mL \times m}{mg} c \tag{4.28}$$

where $m$ are meters and the concentration is measured in $mg/mL$.

The equation of the linear fit as a function of concentration in Fig.(24b) is

$$1 + \left(6.57 \times 10^{-4}\right) \frac{mL}{mg} c \tag{4.29}$$

Further analysis is made in Section (5).

# 5 Analyses

Here the analyses of the experimental results discussed in Section (4) are presented. These results were used to estimate parameters for candidate particle shape models, molecular weights, virial coefficients, and hydrodynamic radii. The mixing free energy models used are those described by Boublik & Nezbeda (1986). These models are for fluid mixtures of hard convex particles at low concentrations as well as high concentrations that are comparable to those found in the eye lens.

First, the first table in Section (4) shows that the molecular weights and virial coefficients determined from the low concentration light scattering are systematically larger than those inferred from the Boublik and Nezbeda fits. There are several possible reasons for these discrepancies. First, the low concentration light scattering experiments are affected by polydispersity. In particular, polydispersity affects the average expressed by $M_w$, whereas the Boublik and Nezbeda model used here assumes a single component. Second, polydispersity will also affect the measured apparent virial coefficient. The low and high concentration light scattering experiments also are affected by the attractive and repulsive inter-protein interactions beyond those envisioned by the Boublik and Nezbeda convex monodisperse particle model. Studying the reasons for the observed discrepancy in a quantitative fashion is anticipated to be a large undertaking and will likely necessitate the inclusion of attractive interactions and alteration of the Boublik and Nezbeda model to account for polydispersity, as well as consideration of pure and mixed second virial coefficients for a polydisperse system of rods, see discussion below.

Experiment A is the only experiment for which all of the analyses could be applied, as it is the only experiment for which static light scattering data exists in both the high and low concentration regimes as well as quasielastic light scattering data over most of the concentration range. For experiments for which low concentration data were lacking, the data were used to fit our chosen model to ascertain molecular weight ($M_w$) and a shape factor ($\alpha$); for these data the second osmotic virial coefficient ($A_2$) was inferred from $\alpha$. For experiments for which there were only low concentration data, the Zimm/Debye analysis yielded the same information as the fit to a thermodynamic model, though over a smaller range of data. At

all concentration ranges it was attempted to obtain quasielastic light scattering data, however only for experiment A were there sufficient data for analysis.

## 5.1 Is $\beta_H$ Crystallin a Prolate Spherocylinder?

First, a plot of molecular weights versus shape factors as inferred from the fits to the Boublik and Nezbeda model for each of the four high concentration experiments (see Fig.(21)) was made. As shown in Fig.(25) below, there is an approximately linear relationship between the molecular weight and the shape factor. This suggested that the proteins might be undergoing rod-like growth as molecular weight increases.

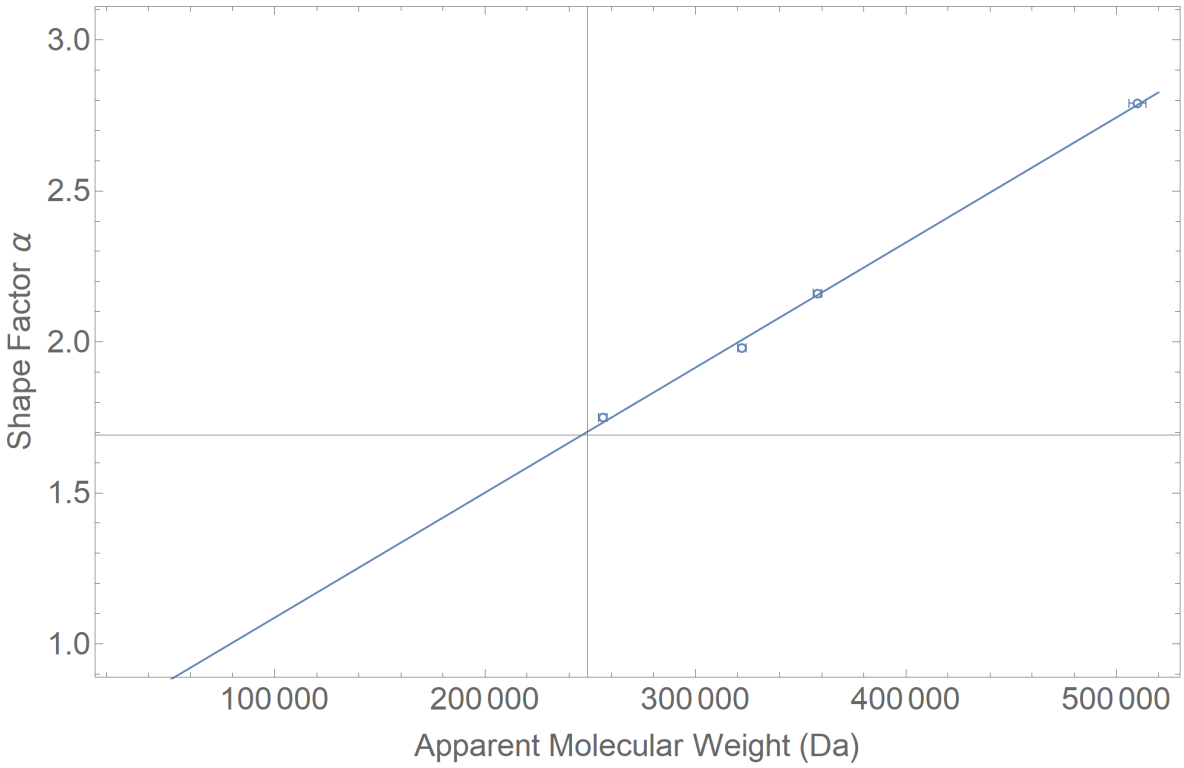



Figure 25: The measured shape factor appears to increase approximately linearly with apparent molecular weight as measured at a scattering angle of $90^o$. The equation of the fitted line is: $\alpha(m_w) = 0.67 + (4.1 \times 10^{-6})m_w$. Note that a shape factor less than one does not occur for convex bodies.

This suggested that it would be interesting to plot the aspect ratios, the ratio of maximum length to diameter for convex spherocylinders, versus the molecular weights. As described further below, Boublik and Nezbeda provide expressions for $\mathcal{R}$, $\mathcal{S}$, and $\mathcal{V}$ for such spherocylinders

so that the corresponding $\alpha$ can be related to the aspect ratio, which is denoted here by $\xi$. The resulting plot is shown in Fig.(26). Again, the inferred aspect ratio appears to increase approximately linearly with apparent molecular weight.

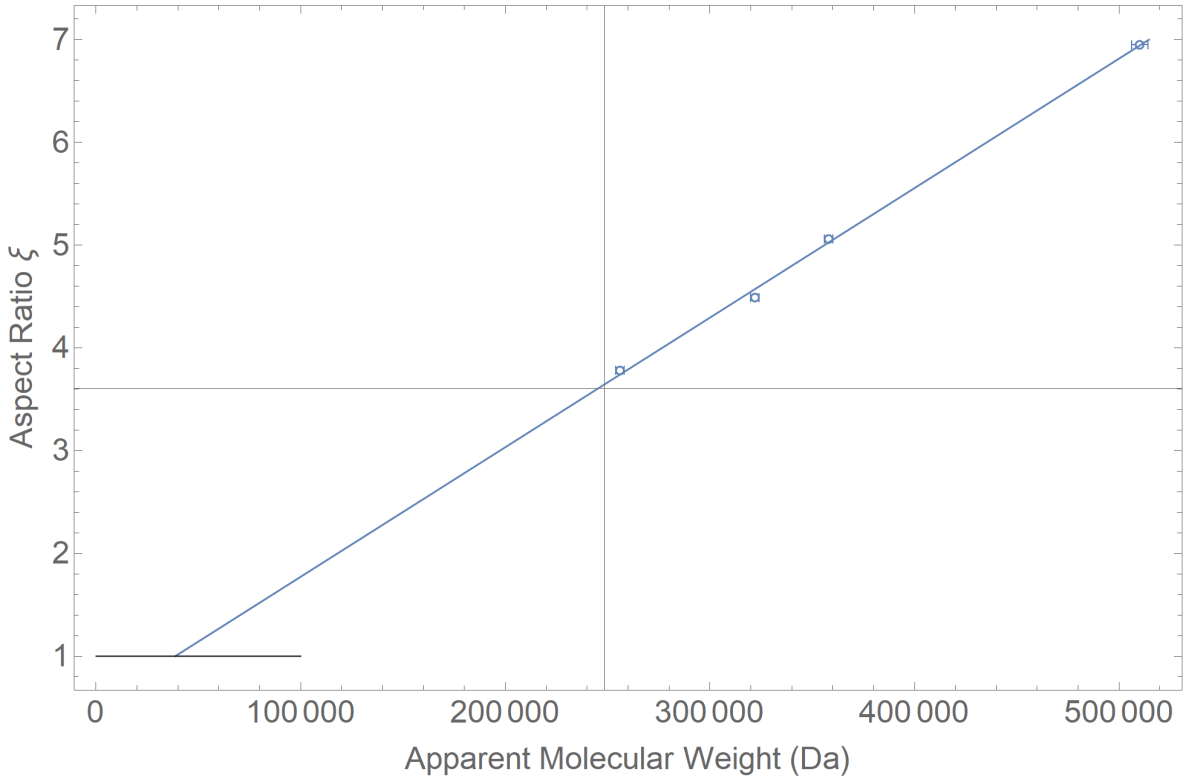



Figure 26: The aspect ratios for prolate spherocylinders that correspond to the inferred shape factors, $\alpha$, appear to increase approximately linearly with apparent molecular weight as measured at a scattering angle of $90^o$. The equation of the fitted line is: $\xi(m_w) = 0.51 + (1.26 \times 10^{-5})m_w$.

To investigate this in more detail, the dependence of $\alpha$ on parameters that describe a prolate spherocylinder is examined. By doing so, the goal is to ascertain the actual widths and lengths of the appropriate model prolate spherocylinders so as to be able to compare them with current knowledge about the size and shape of $\beta$ crystallin subunit types.

Denote the radii by $b$ and overall end-to-end lengths by $a$. Because this assumed shape is prolate, $\xi = \frac{a}{2b} > 1$. According to Boublik & Nezbeda (1986), $\alpha$ for a prolate spherocylinder is given by Eq.(5.30). Note that $a$ and $b$ enter into $\alpha$ only in terms of $\xi$.

$$\alpha\left(\frac{a}{2b}\right) = \frac{\mathcal{RS}}{3\mathcal{V}} = \frac{a\left(1+\frac{a}{2b}\right)}{2b\left(\frac{3a}{2b}-1\right)} = \frac{\xi\left(1+\xi\right)}{3\xi-1} \tag{5.30}$$

It is interesting that if Eq.(5.30) is expanded in terms of $\frac{1}{\xi}$ and only lowest order terms are kept, alpha becomes a linear equation in $\xi$:

$$\alpha\left(\frac{a}{2b}\right) = \frac{4}{9} + \frac{a}{6b} + \mathcal{O}\left(\frac{b^2}{a^2}\right) = \frac{4}{9} + \frac{\xi}{3} + \mathcal{O}\left(\frac{1}{\xi}\right) \tag{5.31}$$

A comparison between the full expression for $\alpha$ and the first two terms in the expansion is shown in Fig.(27)

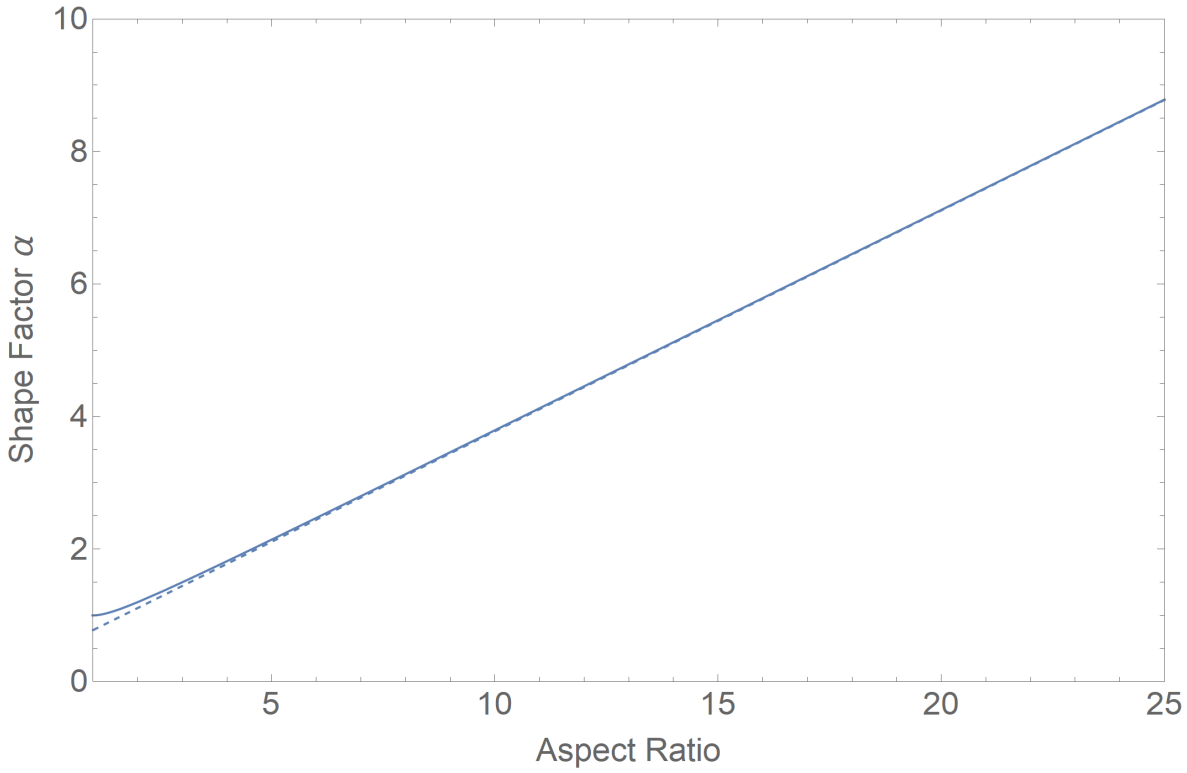



Figure 27: Shape factor for a prolate spherocylinder as a function of aspect ratio for fixed unit radius, as shown in Eq.(5.30). Values below 1 are not valid for this analysis and are thus not shown. The dashed line is the first order expansion shown in Eq.(5.31). See the text for details.

For experiment A, the molecular weight was measured to be $358 \pm 2\ kDa$ and the shape factor to be 2.2. Note that because a shape factor of $\alpha = 1$ indicates perfectly spherical

particles, the measured value of $\alpha$, under this monodisperse convex body assumption, appears to rule out the possibility of $\beta_H$ existing as spheres. From Fig.(27), the measured shape factor of $\sim 2.2$ corresponds to a prolate spherocylinder that is on the order of 5 times longer than it is wide.

To estimate $a$ and $b$, an expression is needed for the volume of a prolate spherocylinder in terms of the inferred molecular weight. As explained in Sec.(2), the partial specific volume defined as $\Omega_p = 1/\rho$ may be used with the measured mass of a single particle ($M_w$, the molecular weight, divided by $N_A$, Avogadro's number) to infer a volume for a particle; see Eq(5.33). Here, $\Omega_p = 0.71\frac{mL}{g}$ is used, which is a very typical partial specific volume of globular proteins (Van Holde, 1985). The expression for volume of a prolate spherocylinder in combination with the expression for $\alpha$ can then be used to construct two equations with two unknowns which will permit separate determination of $a$ and $b$. For a prolate spherocylinder, the volume is simply the volume of a sphere of radius $b$ plus the volume of a cylinder of the same radius and length $a - 2b$. Thus

$$V = \frac{4}{3}\pi b^3 + \pi b^2(a - 2b) = \pi b^3\left(\frac{a}{b} - \frac{2}{3}\right) \tag{5.32}$$

$$\rho = \frac{M_w}{N_A V} \rightarrow V = \frac{M_w}{N_A \rho} = \frac{M_w \Omega_p}{N_A} \tag{5.33}$$

Plugging in Eq.(5.33) and isolating $a$ and $b$, we have

$$b^3\left(\frac{a}{b} - \frac{2}{3}\right) = \frac{M_w \Omega_p}{\pi N_A} \tag{5.34}$$

This is now a system of two equations (Eq.(5.30) and Eq.(5.34)) and two unknowns, the parameters $a$ and $b$. This system is now solved graphically and numerically. Since numerical solutions can be particularly sensitive to initial trial values, we use the graphical method to inform the numerical method of good starting parameters. First, a contour plot is made (Fig.(28)) of both equations as functions of $a$ and $b$. Additionally, the contours that correspond to our measured values are plotted, the shape factor and the scaled volume given by the right

hand side of Eq.(5.34) [9]. For a molecular weight of $358 \pm 2\ kDa$, the volume of a corresponding prolate spherocylinder with the density of a globular protein is $422 \pm 5\ nm^3$. The intersection of the thick contours corresponds to a solution to the system of equations described above. The intersection is estimated by eye, and these values are used as the starting values in the default root-finding algorithm provided by Mathematica.

[9] This is simply the volume divided by the constant prefactor $2\pi$

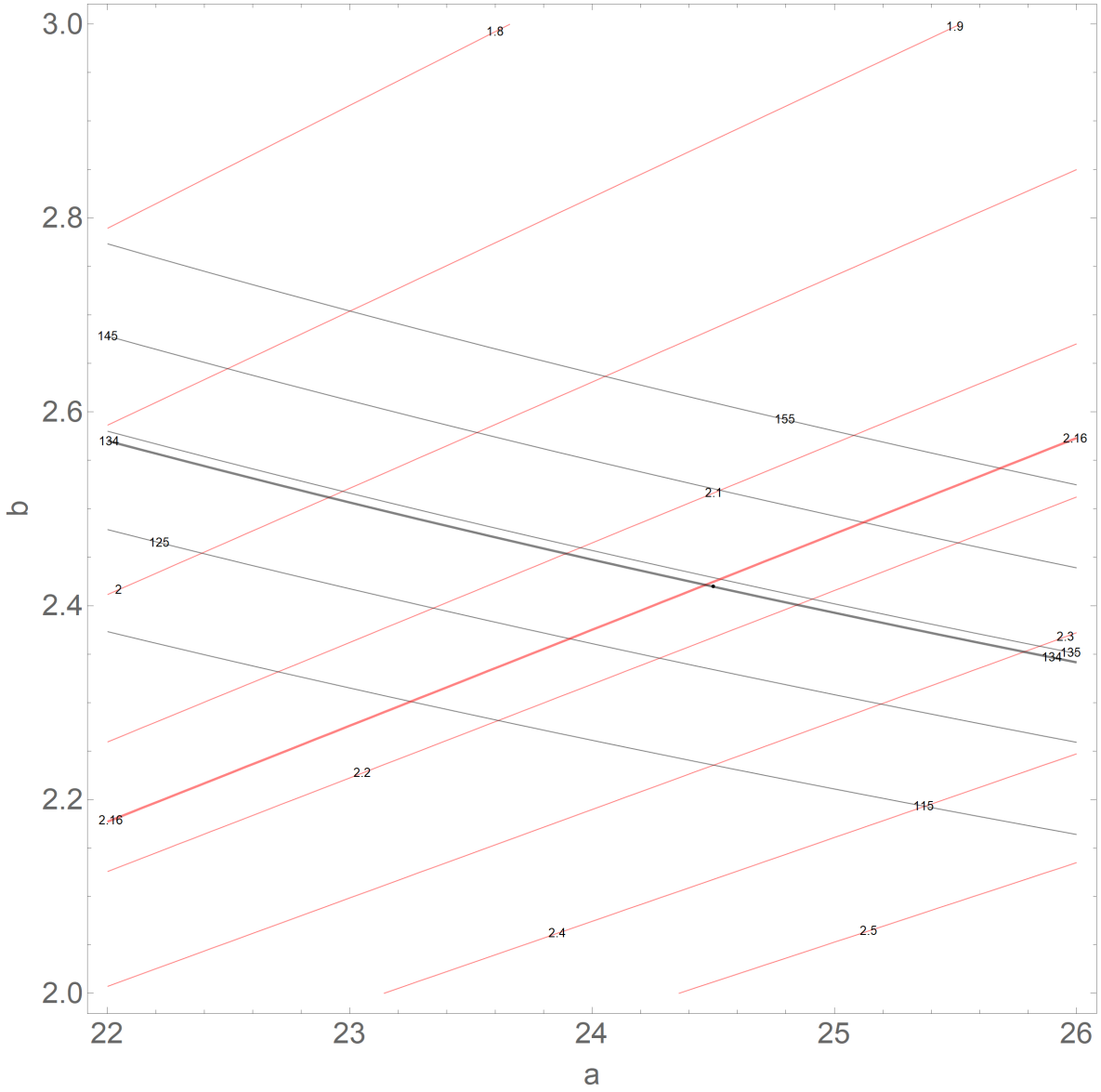



Figure 28: The red contours represent values of the shape factor given in Eq.(5.30) and the black contours represent values of the scaled volume given in Eq.(5.34) as functions of the prolate spherocylinder shape parameters $a$ and $b$. The thick contours represent measured values, and the intersection is a solution to the system of equations given in the preceding text. The black dot is the solution found via numerical methods.

For the above system of equations for a prolate spherocylinder, it was found that the shape parameters $a$ and $b$ are 24.5 $nm$ and 2.42 $nm$ respectively for Experiment A. For the

remaining experiments, the parameters resulting from a similar analysis are summarized in Table(5) below. The fact that the inferred values of $b$ are nearly the same for the different experiments is further evidence that the $\beta_H$ crystallins may be rod-like in shape. That is to say that $\beta_H$ crystallins grow in length as molecular weight increases, but the breadth remains nearly the same. The corresponding shapes are depicted for direct comparison in Fig.(29).

Table 5: Summary of Prolate Spherocylinder Parameters in nm

| Experiment | a | b | $\xi = \frac{a}{2b}$ |
|---|---|---|---|
| A | 24.5 | 2.42 | 5.06 |
| B | 33.9 | 2.44 | 6.95 |
| C | 21.9 | 2.44 | 4.49 |
| D | 18.2 | 2.41 | 3.78 |

The plausibility of such spherocylinders, whose breadth are approximately $\sim 4.8\ nm$ are briefly considered. Simulated models of $\beta$ crystallin subunits were downloaded from AlphaFold except for $\beta B2$, the only subunit for which experimental data exists. From Fig.(2), it is clear to see that all of the subunits have about the same size, so only one subunit type will be considered for this analysis. The arms are likely intrinsically disordered portions which have high flexibility. AlphaFold reports low confidence on the simulation for these arm substructures.

It is intriguing that the $\beta_H$ crystallin subunits' globular parts are measured to be nearly $4.8\ nm$ across their longest axes, close to twice the inferred value of $b$ as is now described. The dimensions of the $\beta$ crystallin subunit structures were measured using the software ChimeraX (Goddard *et al.*, 2018; Pettersen *et al.*, 2021). The structures inferred from x-ray crystallography and AlphaFold for each subtype have two globular components. It was found that for $\beta B2$ the maximum distances across these subunits range from $2.1 - 2.7\ nm$.

An attempt was made to pack copies of $\beta A3$ into a prolate spherocylindrical hulls with the dimensions we obtained via light scattering. In this inital attempt, the default orientation provided by AlphaFold was not altered; future work can relax this assumption and include

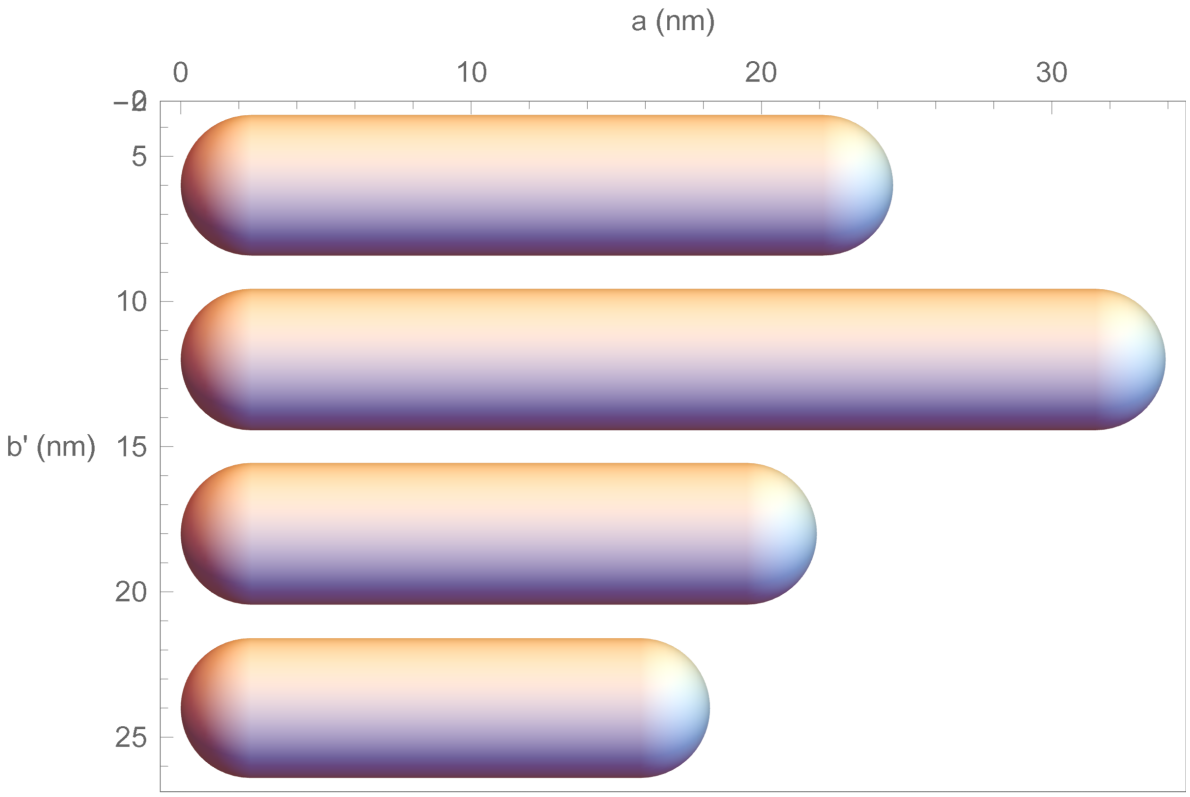
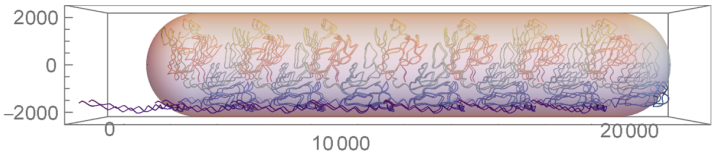


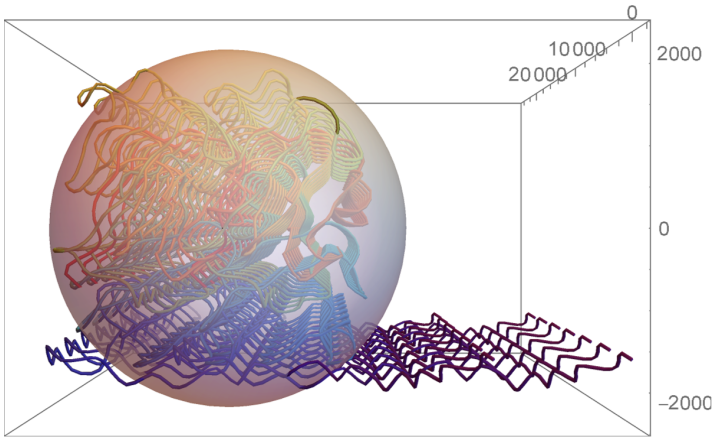

Figure 29: Plotting the results from the prolate spherocylinder shape fitting shows that the radius is nearly the same between different experiments, further supporting the idea that $\beta_H$ is rodlike. These results are in descending alphabetical order for the experiments A-D. The symbol $b'$ denotes the prolate spherocylinder radius $b$ plus the separation applied for the figure.

other subtypes.

Quite provocatively, playing with the exact spacing and positioning of these subunits allowed for up to 14 subunits to fit quite nicely, as shown in Fig.(30). The number 14 was chosen because $14 \times 25\ kDa = 350\ kDa$, approximately the molecular weight inferred from Experiment A. More detailed packing investigations are left for future work.




(a)




(b)

Figure 30: Two views of a packing of fourteen identical $\beta$ crystallin subunits in a prolate spherocylindrical hull with dimensions obtained from light scattering. Subunit structures were packed so as to minimize overlap; this is, however, a preliminary investigation. The axes represent picometers.

However the orientations here are arbitrary and not informed by experiment or theory. In reality, interactions between the proteins would cause their structures to alter slightly and pack in a more complicated way than simply translating and rotating the structures in space. A more systematic way of elucidating the internal structure of $\beta_H$ crystallin is needed.

### 5.1.1 Hydrodynamic Radius Considerations

The limiting value of the hydrodynamic radius at low concentrations in Experiment A was found to be 7.3 $nm$, as presented above in Section (4). A model for the hydrodynamic radius for prolate spherocylinders is examined here in light of the preceding discussion.

Determining the hydrodynamic radius for an arbitrary three dimensional shape is a difficult fluid mechanics problem that must account for the differing drag forces as the shape changes orientation (Dhont, 1996). This is simple for objects such as spheres, that have no dependence on orientation, but quickly becomes considerably more difficult with more com-

plicated shapes. In particular, coupling of translational and orientational diffusion becomes increasingly important for larger aspect ratios of prolate particles (Dhont, 1996).

An expression for the diffusion coefficient of ellipsoids was derived by Perrin (1936). This is the only readily available analytically-obtained expression for the diffusion coefficients (or hydrodynamic radius) of prolate particles. Diffusion coefficients for other objects have been modelled by a variety of methods. A useful resource, which gathers many of these results, is that provided by Mansfield & Douglas (2008).

de la Torre & Bloomfield (1981) proposed the following expression for rigid cylindrical rods, which agreed very well with previous experimental data:

$$\frac{3\pi\eta_o L D_t}{k_b T} = \ln\xi + \gamma(\xi) = \frac{L}{2R_h} \tag{5.35}$$

where $L$ is the length of the cylinder, $b$ is the radius, and $\xi = L/2b$ is the aspect ratio. $\gamma$ is given by:

$$\gamma(\xi) = 0.312 + \frac{0.565}{\xi} + \frac{0.100}{\xi^2} \tag{5.36}$$

Thus, $R_h$ is given in terms of $L$ and $\xi$ by:

$$R_h = \frac{L}{2}\frac{1}{\ln\xi + \gamma(\xi)} \tag{5.37}$$

Using the values of $L = a = 24.5\ nm$ and $\xi = 5.06$ from Table (5) with Eq.(5.37) predicts a corresponding $R_h = 6.0\ nm$, slightly below the dilute extrapolated experimental value, $7.3\ nm$.

Assessing the influence of polydispersity, rod flexibility, and coupling of rotational and translational diffusion on this comparison is left for future work.

In Fig.(24) diffusion coefficients vs concentration are presented. Analysis of these data is also left for future work. To do so will involve a combination of considering the influence of the second virial coefficient on the osmotic compressibility, that replaces the factor $k_bT$ in the Stokes-Einstein relationship (see Eq.(2.26)), combined with the effect of hydrodynamic

interactions between dilute solutions of rods, which alter the friction factor from its dilute value $6\pi\eta_o R_h$. In the meantime, it is worth nothing that $\frac{D}{D_o}$, within the concentration range investigated, is quite linear as a function of concentration.

# 6 Conclusions

In this thesis, light-scattering properties of $\beta_H$ crystallin have been investigated. Different experimental preparations resulted in different measured molecular weights. Analysis of the light scattering results strongly suggested that $\beta$ crystallin associates into rod-like structures. A free energy model of convex prolate spherocylinders was found to be compatible with the data. To assess these results further, consideration of the effects of polydispersity, attractive interactions, and hydrodynamic interactions is needed. Nevertheless, the present evidence for rod-like particles will also be interesting to examine in light of the original motivation for studying its influence on the thermodynamics of mixtures of $\alpha$, $\beta$, and $\gamma$ crystallin.

In addition to the above considerations, future work on $\beta_H$ crystallin should test the applicability of other possible shapes, such as oblate particles and bent rods. To do so, experimental techniques in addition to light scattering are likely to be instructive.

Future work may also try more physically-informed packing methods for filling a candidate shape with simulated subunit structures. These methods would consider the full pair potential between subunits as a function of not only radial separation, but of relative orientations. Furthermore, these models should consider any necessary changes in conformation that occur when a subunit goes from being free to being bonded to other subunits as an aggregate.

Beyond $\beta_H$ crystallin, to fully understand the scattering from mammalian lenses $\beta_L$ crystallin must also be studied. The same procedures and analyses used in this thesis may be applied almost identically to $\beta_L$ crystallin. Future experiments would also assess the scattering of mixtures of different proportions of $\beta_H$ and $\beta_L$ and study their equilibrium characteristics.

This thesis also obtained intriguing evidence of subunit-by-subunit dissociation of $\beta_H$ crystallin at low concentrations, which bears further study.

Finally, as mentioned above, this work directly supports future work on mixtures of the three primary lens crystallins. A better understanding of $\beta_H$ crystallin is necessary to understanding the total scattering from the eye lens.

# Appendices

# 1 Derivation

Here we lay out the steps involved in deriving Eq.(2.8) from Eq.(2.5). This is the process of incorporating a model for the Gibbs free energy of mixing into an equation that relates the free energy to the scattered light at a given angle. The first major simplification is to establish the form of this equation when only one scattering component, our protein, is considered. This greatly reduces the complexity contained within the inverse of the Hessian matrix and in the gradient of the electrical permittivity of the sample. The inverse Hessian matrix reduces down to a single term

$$H_\rho[g]^{-1} = \left( \frac{\partial^2 \frac{G}{V}}{\partial \rho^2} \right)^{-1} \tag{1.38}$$

And the matrix-transpose of the gradient of the electrical permittivity (recalling that the refractive index $n = \sqrt{\epsilon}$ for non-magnetic materials) can be written as the row vector with only one component

$$\nabla_\rho^T \epsilon = \left[ \frac{\partial(n^2)}{\partial \rho_1} \quad \dots \quad \frac{\partial(n^2)}{\partial \rho_n} \right] = \frac{\partial(n^2)}{\partial \rho_1} \tag{1.39}$$

Where we take $\rho_1 = \rho$ to be the number density of our protein in solution. We also make the substitution (using the chain rule) $\rho = \frac{c}{m_p}$, where $m_p$ is the mass of a single protein molecule, and c is the mass per unit volume concentration of the sample. This yields

$$\frac{\partial(n^2)}{\partial \rho} = 2 n m_p \frac{\partial n}{\partial c} \tag{1.40}$$

The term $\frac{\partial n}{\partial c}$ is the refractive index increment with concentration, which, as mentioned in the Methods section, is measurable in the lab.

It should be noted that $\rho$ can also be expressed as $\rho = \frac{N_p}{V} = \frac{\phi}{\Omega_p}$. This relationship is easily shown from the definition of the volume fraction, $\phi$, given in the Analyses section. $\Omega_p$ is the partial specific volume of a protein molecule, or the volume occupied by a single gram of protein. It will also be necessary to express $\frac{G}{V}$ in a dimensionless form with the following equation

$$g = \frac{\Omega_w G}{k_B T V} \tag{1.41}$$

We are finally equipped to begin tackling the main problem, as we have reduced Eq.(1.38) into a form valid for our system, expressed the gradients of the electrical permittivity as something measurable in the lab, and converted the Gibbs free energy into a more convenient form.

We introduced the form of the Boublik and Nezbeda model for mixtures of hard convex bodies in Eq.(2.7), repeated again here below

$$\frac{\beta P}{\rho} = \frac{1}{1-\phi} + \frac{3\alpha\phi}{(1-\phi)^2} + \frac{3\alpha^2\phi^2}{(1-\phi)^3} \tag{1.42}$$

given in terms of a compressibility factor. We need to re-express this equation in terms of the dimensionless Gibbs free energy, however, we must first make some substitutions. If we assume our solution to be incompressible, the colloid osmotic pressure, $\Pi$, as a function of the protein volume fraction is defined as

$$\Pi(\phi)\Omega_w = \mu_w^o - \mu_w \tag{1.43}$$

where $\Omega_w$ is the volume occupied by a single water molecule, $\mu_w^o$ is the chemical potential of pure water (which we assume to represent our buffer), and $\mu_w$ is the chemical potential of the solution at finite volume fractions of protein. Now, we make the substitutions $\rho \to \frac{\phi}{\Omega_p}$, $\beta \to \frac{1}{k_b T}$, and $P \to \Pi(\phi)$ in Eq.(1.42) yielding

$$\frac{\Pi(\phi)\Omega_p}{\phi k_b T} = \frac{1}{1-\phi} + \frac{3\alpha\phi}{(1-\phi)^2} + \frac{3\alpha^2\phi^2}{(1-\phi)^3} \tag{1.44}$$

The density per unit volume of both proteins and water may be expressed in several ways:

$$\rho_i = \frac{N_i}{V} \tag{1.45}$$

$$\rho_p = \rho = \frac{1}{\Omega_p}\phi. \tag{1.46}$$

Since the volume fraction of water is $1-\phi$, we may write

**Derivation**

$$\rho_w = \frac{1}{\Omega_w}(1-\phi) = \frac{1}{\Omega_w}(1-\rho\Omega_p). \tag{1.47}$$

If we define a factor $\gamma$ to be the ratio of the partial specific volumes of the protein and water

$$\gamma = \frac{\Omega_p}{\Omega_w} \tag{1.48}$$

then

$$\rho_w = \frac{1}{\Omega_w} - \gamma\rho. \tag{1.49}$$

Therefore, we may define a Gibbs free energy per unit volume

$$g = \frac{G}{V} = \rho_w\mu_w + \rho\mu_p = \left(\frac{1}{\Omega_w} - \gamma\rho\right)\mu_w + \rho\mu_p. \tag{1.50}$$

If we assume that the partial specific volume of water doesn't change with the density of protein, or that water is incompressible, ($\frac{\partial\Omega_w}{\partial\rho} = 0$), and the same for the ratio $\gamma$, ($\frac{\partial\gamma}{\partial\rho} = 0$), then

$$\frac{\partial g}{\partial\rho} = -\gamma\mu_w + \frac{1}{\Omega_w}\frac{\partial\mu_w}{\partial\rho} + \mu_p + \rho\frac{\partial\mu_p}{\partial\rho} - \gamma\rho\frac{\partial\mu_w}{\partial\rho} \tag{1.51}$$

The Gibbs-Duhem relation follows from:

$$dG = N_w d\mu_w + \mu_w dN_w + N_p d\mu_p + \mu_p dN_p \tag{1.52}$$

$$dG = -SdT + VdP + \mu_w dN_w + \mu_p dN_p \tag{1.53}$$

where the entropy, $S$, is the only quantity not previously defined. If Eq.(1.52) is subtracted from Eq.(1.53), we have

$$0 = -SdT + VdP - N_w d\mu_w - N_p d\mu_p. \tag{1.54}$$

If it is assumed that there is no change in temperature or pressure, that is $dT = dP = 0$, then

$$N_w d\mu_w + N_p d\mu_p = 0 \tag{1.55}$$

At thermal equilibrium, pressure and the chemical potentials of water and protein only depend on $\rho$

$$d\mu_w = \frac{\partial \mu_w}{\partial \rho} d\rho \quad , \quad d\mu_p = \frac{\partial \mu_p}{\partial \rho} d\rho \tag{1.56}$$

and, using Eq.(1.55),

$$N_w \frac{\partial \mu_w}{\partial \rho} + N_p \frac{\partial \mu_p}{\partial \rho} = 0. \tag{1.57}$$

If we divide by volume, we have

$$\frac{N_w}{V} \frac{\partial \mu_w}{\partial \rho} + \rho \frac{\partial \mu_p}{\partial \rho} = 0 \tag{1.58}$$

or

$$\left( \frac{1}{\Omega_w} - \gamma \rho \right) \frac{\partial \mu_w}{\partial \rho} + \rho \frac{\partial \mu_p}{\partial \rho} = 0. \tag{1.59}$$

Substituting this into Eq.(1.51), we obtain

$$\frac{\partial g}{\partial \rho} = -\gamma \mu_w + \mu_p, \tag{1.60}$$

and taking another derivative with respect to protein number density

$$\frac{\partial^2 g}{\partial \rho^2} = -\frac{1}{\Omega_w \rho} \frac{\partial \mu_w}{\partial \rho}. \tag{1.61}$$

**Derivation**

For constant $\Omega_w$,

$$\Pi\Omega_w = \mu_w^o - \mu_w \tag{1.62}$$

where $\Pi$ and $\mu_w^o$ stand for the osmotic pressure of the sample and chemical potential of pure water respectively, and

$$\Omega_w \frac{\partial \Pi}{\partial \rho} = -\frac{\partial \mu_w}{\partial \rho}. \tag{1.63}$$

Therefore,

$$\frac{\partial^2 g}{\partial \rho^2} = \frac{1}{\rho}\frac{\partial \Pi}{\partial \rho}. \tag{1.64}$$

Recalling that the protein number density may be expressed as in Eq.(1.46), and recalling that $\Omega_p$ is assumed to be constant

$$\frac{\partial^2 g}{\partial \phi^2} = \frac{1}{\phi}\frac{\partial \Pi}{\partial \phi}. \tag{1.65}$$

Finally, plugging in Eq.(1.44) for osmotic pressure $\Pi$ yields the result

$$\frac{\partial^2 g}{\partial \phi^2} = \frac{k_b T}{\Omega_p}\left(\frac{1}{\phi(1-\phi)^2} + \frac{6\alpha}{(1-\phi)^3} + \frac{9\phi\alpha^2}{(1-\phi)^4}\right). \tag{1.66}$$

Combining this result with Eq.(1.38) and Eq.(1.40) gives our final result.

$$\Delta R = \frac{4\pi^2 n^2 (\frac{dn}{dc})^2 m_p^2}{N_A \lambda_o^4}\left(\frac{1}{\phi(1-\phi)^2} + \frac{6\alpha}{(1-\phi)^3} + \frac{9\phi\alpha^2}{(1-\phi)^4}\right). \tag{1.67}$$

## 2 Table of Symbols Used

Care was taken to avoid the reuse of symbols to avoid confusion, however, some symbols still show up for different things in several places where it was perceived that it would be difficult to mistake them. Even so, the relevant equations are tabulated here as well.

| Symbol | Meaning | Equations Used |
|---|---|---|
| $A$ | Absorbance | 2.1 |
| $\epsilon$ | Molar attenuation coefficient | 2.1 |
| $c$ | Concentration | 2.1, 2.9, 2.12, 2.13, 2.14 |
| $l$ | Path length | 2.1 |
| $\Delta R(90)$ | Excess Rayleigh ratio measured at 90 degrees | 2.2, 2.12, 2.13, 2.14, 2.15, 2.16 |
| $\Delta\langle I(90)\rangle$ | Measured average intensity of sample at 90 degrees subtracted by the same from solvent | 2.2, 2.3 |
| $\langle I_{ref}(90)\rangle$ | Measured average intesnity of toluene at 90 degrees in excess of dark count | 2.2, 2.4 |
| $R_{ref}$ | Excess Rayleigh Ratio of toluene | 2.2 |
| $k_b$ | Boltzmann constant | 2.5, 2.19, 5.35 |
| $T$ | Absolute temperature | 2.5, 2.11, 2.19, 2.26, 5.35 |
| $\lambda$ | Wavelength | 2.5, 2.8 |
| $\epsilon$ | Dielectric coefficient | 2.5 |
| $G$ | Gibbs free energy | |
| $g$ | Gibbs free energy per unit volume $g = G/V$ | 2.5 |
| $\phi$ | Protein volume fraction | 2.7, 2.8, 2.9, 2.14 |
| $\Omega_p$ | Protein partial specific volume | 5.33, 5.34 |
| $V$ | Volume | 5.31, 5.32 |

| | | |
|---|---|---|
| $\alpha$ | Boublik and Nezbeda shape factor | 2.7, 5.29, 5.30, 5.31 |
| $\beta$ | Boltzmann factor $= 1/k_bT$ | 2.7, 2.10, 2.11, 2.13 |
| $P$ | Pressure | 2.7, 2.10 |
| $\rho$ | Number density | 2.5, 2.7, 2.12, 5.33 |
| $n$ | Index of refraction | 2.2, 2.8 |
| $n_{ref}$ | Refractive index of toluene | 2.2 |
| $\frac{dn}{dc}$ | Refractive index increment with concentration | 2.8 |
| $m_p$ | Mass of a single protein molecule | 2.8, 5.32, 5.33 |
| $N_A$ | Avogadro's number | 2.8, 5.33, 5.34 |
| $K$ | Optical coefficient | 2.8, 2.12, 2.13 |
| $A_2$ | Osmotic second virial coefficient | 2.10, 2.11, 2.12, 2.13, 2.14 |
| $B_2$ | Dimensionless osmotic second virial coefficient | 2.14 |
| $\mathcal{V}$ | Volume of a single protein | 2.10, 2.11, 5.30 |
| $\mathcal{S}$ | Surface area of a single protein | 5.30 |
| $\mathcal{R}$ | 1/4pi multiple of mean surface radius of curvature | 5.30 |
| $M_w$ | Molar mass of a particle measured by static light scattering | 2.14, 5.33 |
| $M_w^{app}(90)$ | The apparent molar mass of a particle as measured by static light scattering at a fixed scattering angle of 90 degrees | 2.12, 2.13, 2.14 |
| $R_g$ | Radius of gyration | 2.12, 2.14, 2.16 |
| $R_h$ | Hydrodynamic radius | 2.19, 5.35, 5.37 |

| | | |
|---|---|---|
| $q$ | Scattering vector magnitude | Fig.(6), 2.12, 2.14, 2.16, 2.18, 2.19, 2.20, 2.21, 2.22, 2.24, 2.25, 2.26, 2.27 |
| $AICc$ | Akaikie informaiton criterion corrected for finite sampling | 2.15 |
| $\hat{L}$ | Maximum likelihood | 2.15 |
| $k$ | Number of parameters | 2.15 |
| $n$ | Number of data points | 2.15 |
| $DW$ | Durbin Watson statistic | 2.16 |
| $r_i$ | Value of the ith residual | 2.16 |
| $\eta$ | Viscosity | 2.19 |
| $D$ | Collective diffusion coefficient | 2.18, 2.19, 2.21, 2.22, 2.23, 2.24, 2.26 |
| $G_2$ | Measured homodyne correlation function | 2.17, 2.18 |
| $I(t)$ | Measured intensity at time $t$ | 2.17, 2.18 |
| $\tau$ | Lag time for correlation function $t$ | 2.17, 2.18, 2.19, 2.20, 2.27 |
| $g_2$ or $N$ | Normalized homodyne correlation function | 2.19, 2.20 |
| $K_i$ | Cumulants of homodyne correlation function | 2.20, 2.21, 2.22, 2.24, 2.25 |
| $\langle D \rangle z$ | z-avg diffusion coefficient | 2.23, 2.24, 2.25 |
| $a$ | Full length of prolate spherocylinder | 5.29, 5.30, 5.31, 5.32, 5.34 |

**Table of Symbols Used**

| | | |
|---|---|---|
| $b$ | Radius of prolate spherocylinder | 5.29, 5.30, 5.31, 5.32, 5.34 |
| $\xi = \frac{a}{2b}$ | Prolate spherocylinder aspect ratio | 5.30, 5.31, 5.35, 5.36, 5.37 |

# Bibliography

1998. *Gel Filtration: Principles and Methods*. 8 edn. Amersham Pharmacia Biotech. 2.2

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**(6), 716–723. 2.6.1

Akasaka, Kazuyuki, & Matsuki, Hitoshi (eds). 2015. *High Pressure Bioscience: Basic Concepts, Applications and Frontiers*. Subcellular Biochemistry, vol. 72. Dordrecht: Springer Netherlands. 2.6.1

Banerjee, Priya R., Pande, Ajay, Patrosz, Julita, Thurston, George M., & Pande, Jayanti. 2011. Cataract-associated mutant E107A of human D-crystallin shows increased attraction to -crystallin and enhanced light scattering. *Proc. Natl. Acad. Sci. U.S.A.*, **108**(2), 574–579. 1.1

Bateman, O. A., & Slingsby, C. 1992. Structural studies on H-crystallin from bovine eye lens. *Experimental Eye Research*, **55**(1), 127–133. (document), 1.2, 5, 3.3

Beer. 1852. Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. *Annalen der Physik*, **162**(5), 78–88. 2.2

Bell, Michael M., Ross, David S., Bautista, Maurino P., Shahmohamad, Hossein, Langner, Andreas, Hamilton, John F., Lahnovych, Carrie N., & Thurston, George M. 2017. Statistical-thermodynamic model for light scattering from eye lens protein mixtures. *J. Chem. Phys.*, **146**(5), 055101. Publisher: American Institute of Physics. 1.1

Benedek, G. B. 1971. Theory of Transparency of the Eye. *Appl. Opt.*, **10**(3), 459. 1.1

Berland, Carolyn R, Thurston, George M, Kondo, Mamoru, Broide, Michael L, Pande, Jayanti, Ogun, Olutayo, & Benedek, George B. 1992. Solid-liquid phase boundaries of lens protein solutions. *Proceedings of the National Academy of Sciences*, **89**(4), 1214–1218. 3.1.1

Berne, Bruce J., & Pecora, Robert. 2000. *Dynamic light scattering: With applications to chemistry, biology and physics*. Dover Publications, Inc. 1.3, 2, 2.6.2, 2.6.2, 2.6.2, 2.6.2

Bindels, Jacques G., Koppers, Adri, & Hoenders, Herman J. 1981. Structural aspects of bovine -crystallins: Physical characterization including dissociation-association behavior. *Experimental Eye Research*, **33**(3), 333–343. (document), 5, 3.3

Bloemendal, Hans, Piatigorsky, Joram, & Spector, Abraham. 1989. Recommendations for crystallin nomenclature. *Experimental Eye Research*, **48**(4), 465–466. 1.1

Bloemendal, Hans, de Jong, Wilfried, Jaenicke, Rainer, Lubsen, Nicolette H., Slingsby, Christine, & Tardieu, Annette. 2004. Ageing and vision: structure, stability and function of lens crystallins. *Progress in Biophysics and Molecular Biology*, **86**(3), 407–485. 1.1

Boublik, T., & Nezbeda, I. 1986. P-V-T Behaviour of Hard Body Fluids. *Collect. Czech. Chem. Commun.*, **51**(1), 2301–2432. 2.6.1, 2.6.1, 4, 5, 5.1

Carnahan, Norman F., & Starling, Kenneth E. 1969. Equation of State for Nonattracting Rigid Spheres. *The Journal of Chemical Physics*, **51**(2), 635–636. 2.6.1

Chu, Benjamin. 2007. *Laser light scattering: Basic principles and Practice*. 2nd edn. Dover Publications. 1.3, 2.6.2, 2.6.2, 2.6.2

Clark, Noel A, Lunacek, Joseph H, & Benedek, George B. 1970. A Study of Brownian Motion Using Light Scattering. 12. 3.3

Consortium, The UniProt. 2020. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, **49**(D1), D480–D489. 1.2

Coumou, DJ. 1960. Apparatus for the measurement of light scattering in liquids: Measurement

of the Rayleigh factor of benzene and of some other pure liquids. *Journal of Colloid Science*, **15**(5), 408–417. 2.5

de la Torre, José García, & Bloomfield, Victor A. 1981. Hydrodynamic properties of complex, rigid, biological macromolecules: theory and applications. *Quarterly reviews of biophysics*, **14**(1), 81–139. 5.1.1

Delaye, M., & Gromiec, A. 1983. Mutual diffusion of crystallin proteins at finite concentrations: A light-scattering study. *Biopolymers*, **22**(4), 1203–1221. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bip.360220413. 1.1

Delaye, Mireille, & Tardieu, Annette. 1983. Short-range order of crystallin proteins accounts for eye lens transparency. *Nature*, **302**(5907), 415–417. Number: 5907 Publisher: Nature Publishing Group. 1.1

Dhont, Jan KG. 1996. *An introduction to dynamics of colloids*. Elsevier. 5.1.1

Donovan, JM, Timofeyeva, N, & Carey, MC. 1991. Influence of total lipid concentration, bile salt: lecithin ratio, and cholesterol content on inter-mixed micellar/vesicular (non-lecithin-associated) bile salt concentrations in model bile. *Journal of lipid research*, **32**(9), 1501–1512. 2.4

Dorsaz, N, Thurston, GM, Stradner, Anna, Schurtenberger, Peter, & Foffi, G. 2009. Colloidal characterization and thermodynamic stability of binary eye lens protein mixtures. *The Journal of Physical Chemistry B*, **113**(6), 1693–1709. 1.1

Dorsaz, Nicolas, Thurston, George M., Stradner, Anna, Schurtenberger, Peter, & Foffi, Giuseppe. 2011. Phase separation in binary eye lens protein mixtures. *Soft Matter*, **7**(5), 1763–1776. 1.1

Durbin, J., & Watson, G. S. 1950. TESTING FOR SERIAL CORRELATION IN LEAST SQUARES REGRESSION. I. *Biometrika*, **37**(3-4), 409–428. 2.6.1

Durbin, J., & Watson, G. S. 1951. TESTING FOR SERIAL CORRELATION IN LEAST SQUARES REGRESSION. II. *Biometrika*, **38**(1-2), 159–178. 2.6.1

Goddard, Thomas D., Huang, Conrad C., Meng, Elaine C., Pettersen, Eric F., Couch, Gregory S., Morris, John H., & Ferrin, Thomas E. 2018. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci*, **27**(1), 14–25. 5.1

Herbrink, Paul, & Bloemendal, Hans. STUDIES ON//-CRYSTALLIN. 13. 1.2

Hurvich, Clifford M., & Tsai, Chih-Ling. 1989. Regression and time series model selection in small samples. *Biometrika*, **76**(2), 297–307. 2.6.1

Jumper, John, Evans, Richard, Pritzel, Alexander, Green, Tim, Figurnov, Michael, Ronneberger, Olaf, Tunyasuvunakool, Kathryn, Bates, Russ, Žídek, Augustin, Potapenko, Anna, Bridgland, Alex, Meyer, Clemens, Kohl, Simon A. A., Ballard, Andrew J., Cowie, Andrew, Romera-Paredes, Bernardino, Nikolov, Stanislav, Jain, Rishub, Adler, Jonas, Back, Trevor, Petersen, Stig, Reiman, David, Clancy, Ellen, Zielinski, Michal, Steinegger, Martin, Pacholska, Michalina, Berghammer, Tamas, Bodenstein, Sebastian, Silver, David, Vinyals, Oriol, Senior, Andrew W., Kavukcuoglu, Koray, Kohli, Pushmeet, & Hassabis, Demis. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**(7873), 583–589. 1.2

Kaye, Wilbur, & McDaniel, JB. 1974. Low-angle laser light scattering—Rayleigh factors and depolarization ratios. *Applied Optics*, **13**(8), 1934–1937. 2.5

Kirkwood, John G., & Goldberg, Richard J. 1950. Light Scattering Arising from Composition Fluctuations in Multi-Component Systems. *The Journal of Chemical Physics*, **18**(1), 54–57. 2.5, 2.6.1

Koppel, Dennis E. 1972. Analysis of Macromolecular Polydispersity in Intensity Correlation Spectroscopy: The Method of Cumulants. *J. Chem. Phys.*, **57**(11), 4814–4820. Publisher: American Institute of Physics. 2.6.2, 2.6.2

Lambert, Iohann Heinrich. 1760. *photometria, SIUE, de Mensura et gradibus luminis, colorum et umbrae*. Sumptibus Viduae Eberhardi Klett, typis Christophori Petri Detleffsen. 2.2

Lampi, Kirsten. 2022. personal communication. 1.2

Lampi, Kirsten J., Wilmarth, Phillip A., Murray, Matthew R., & David, Larry L. 2014. Lens -crystallins: The role of deamidation and related modifications in aging and cataract. *Progress in Biophysics and Molecular Biology*, **115**(1), 21–31. 1.2

Mansfield, Marc L., & Douglas, Jack F. 2008. Transport Properties of Rodlike Particles. *Macromolecules*, **41**(14), 5422–5432. 5.1.1

McQuarrie, Donald A. 1976. *Imperfect Gasses*. Harper amp; Row. Page 226–228. 2.6.1, 2.6.1, 2.6.2

Perrin, D. D., & Dempsey, Boyd. 1974. *Buffers for ph and metal ion control by D.D. Perrin and Boyd Dempsey*. Chapman and Hall; New York. 2.1

Perrin, F. 1936. Mouvement Brownien D'Un Ellipsoide (II). 10,11. 5.1.1

Pettersen, Eric F., Goddard, Thomas D., Huang, Conrad C., Meng, Elaine C., Couch, Gregory S., Croll, Tristan I., Morris, John H., & Ferrin, Thomas E. 2021. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci*, **30**(1), 70–82. 5.1

Pierscionek, Barbara, Smith, George, & Augusteyn, Robert C. 1987. The refractive increments of bovine $\alpha$-, $\beta$-and $\gamma$-crystallins. *Vision Research*, **27**(9), 1539–1541. 2

Roosen-Runge, Felix, Gulotta, Alessandro, Bucciarelli, Saskia, Casal-Dujat, Lucía, Garting, Tommy, Skar-Gislinge, Nicholas, Obiols-Rabasa, Marc, Farago, Bela, Zaccarelli, Emanuela, Schurtenberger, Peter, & Stradner, Anna. 2020. Crowding in the Eye Lens: Modeling the Multisubunit Protein -Crystallin with a Colloidal Approach. *Biophysical Journal*, **119**(12), 2483–2496. 1.1

Ross, David S., Thurston, George M., & Lutzer, Carl V. 2008. On a partial differential equation method for determining the free energies and coexisting phase compositions of ternary mixtures from light scattering data. *J. Chem. Phys.*, **129**(6), 064106. Publisher: American Institute of Physics. 2.6.1

Schurtenberger, Peter, Chamberlin, Richard A, Thurston, George M, Thomson, John A, & Benedek, George B. 1989. Observation of critical phenomena in a protein-water solution. *Physical review letters*, **63**(19), 2064. 2.6.1

Siezen, Roland J. 1984. Interactions of lens proteins: Ultrafiltration is unsuitable to detect self-or mixed-association. *Biophysical chemistry*, **19**(1), 49–55. 2

Siezen, Roland J., Anello, Robert D., & Thomson, John A. 1986. Interactions of lens proteins. Concentration dependence of -crystallin aggregation. *Experimental Eye Research*, **43**(3), 293–303. (document), 1.2, 5, 3.1.1, 3.3

Strutt, Hon. J.W. 1871. LVIII. On the scattering of light by small particles. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **41**(275), 447–454. 1.3

Thurston, George M. 2006. Liquid-liquid phase separation and static light scattering of concentrated ternary mixtures of bovine ⌣ and B crystallins. *J. Chem. Phys.*, 10. 1.1, 2.1

Van Holde, KE. 1985. *Sedimentation*. 2 edn. Prentice-Hall. Page 117. 5.1

Varadi, Mihaly, Anyango, Stephen, Deshpande, Mandar, Nair, Sreenath, Natassia, Cindy, Yordanova, Galabina, Yuan, David, Stroe, Oana, Wood, Gemma, Laydon, Agata, Žídek, Augustin, Green, Tim, Tunyasuvunakool, Kathryn, Petersen, Stig, Jumper, John, Clancy, Ellen, Green, Richard, Vora, Ankur, Lutfi, Mira, Figurnov, Michael, Cowie, Andrew, Hobbs, Nicole, Kohli, Pushmeet, Kleywegt, Gerard, Birney, Ewan, Hassabis, Demis, & Velankar, Sameer. 2021. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, **50**(D1), D439–D444. 1.2

W. Wang, S. Nema, D. Teagarden. 2010. Protein aggregation-Pathways and influencing factors. *International Journal of Pharmaceutics*. 2.1, 2.4

Ward, M.D., & Ahlquist, J.S. 2018. *Maximum Likelihood for Social Science: Strategies for Analysis*. Analytical Methods for Social Research. Cambridge University Press. 2.6.1

Zimm, Bruno H. 1948. The Dependence of the Scattering of Light on Angle and Concentration in Linear Polymer Solutions. *J. Phys. Chem.*, **52**(1), 260–267. 2.6.1