Supplementary Materials for

Eye Patches: Protein Assembly of Index-gradient Squid Lenses

J. Cai, J.P. Townsend, T.C. Dodson, P.A. Heiney, and A.M. Sweeney

correspondence to: alisonsw@physics.upenn.edu

This PDF file includes:

Materials and Methods
Figs. S1 to S4
References 33–43

Materials and methods:

Specimen collection

Specimens of the squid *Doryteuthis pealeii* were obtained from the Marine Biological Laboratory at Woods Hole, MA. Lenses were excised from the suspending tissue and either used fresh or stored at -80 ºC either alone for use in SAXS or in RNAlater for use in molecular biology experiments.

Refractive index and protein concentration:

To measure the refractive index at the center of the lens (n_c), the refractive index at the edge of the lens (n_e), and the radial refractive index distribution for the lens as a whole, we dissected a mature *D. pealeii* lens into seven concentric layers. The refractive index of each layer was estimated by submerging the pieces of tissue in refractive index matching fluids of 1.45 and 1.55 (Cargille, Cedar Grove, NJ). We quickly photographed each layer in index-matching oil and then recorded the radial position of the tissue sample that was invisible in each of the two index-standard fluids. With measurements of exact refractive index at two radial positions, we can estimate the refractive index at all radial positions in the lens (Eq. 1). We also put additional bounds on this index gradient using independent measures of packing fraction from SAXS and a direct measurement of density via determination of buoyancy solutions of potassium iodide of known density; ultimately, we estimate n_e to be 1.33 and n_c to be 1.62.

There are two possible mechanisms by which this gradient index could be generated. If the change in refractive index with change in concentration is the same for all proteins, then increasing refractive index must be generated by an increase in protein concentration. Alternatively, if different S-crystallin isoforms have different refractive increments, then protein content can be modulated such that a given increase in refractive index is achieved with smaller increases in protein concentration throughout the lens. To check whether either or both of these mechanisms is active in squid lens, we calculated the refractive index, n_p, of dry S-crystallin proteins using the Lorentz-Lorenz formula

$$n_p = \sqrt{\frac{2R_p + \bar{v}_p}{\bar{v}_p - R_p}}$$

where R_p is the refraction per gram of the protein, and $\bar{v}_p$is the mass-averaged partial specific volume of all 20 amino acids (*33, 34*). We calculated R_p for all the proteins in the lens based on the amino acid sequences derived from our RNAseq data.

From there, we can estimate the increase in refractive index with increase in protein concentration (refractive increment, or d*n*/d*c*) for S-crystallins using the Wiener equation:

$$\frac{dn}{dc} = \frac{3}{2}\bar{v}_p n_o \frac{n_p^{\,2} - n_o}{n_p^{\,2} + 2n_o^{\,2}}$$

(*34*).

Here, n_0 is the refractive index of the medium, which was considered to be pure water at 1.334, and we used a wavelength of 578 nm. We calculated d*n*/d*c* for all the S-crystallin proteins expressed in the lens, and compared the d*n*/d*c* for proteins preferentially expressed in the core versus the periphery of the lens. We found that all S-crystallin proteins had roughly the same d*n*/d*c* and that there was no correlation between d*n*/d*c* and the likely position of maximum expression of a given S-crystallin isoform in the lens.

SAXS

Small angle X-ray scattering experiments were performed both using the multi-angle X-ray scattering system (MAXS) in our labs at University of Pennsylvania, and using the X9 beam line of the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory. For both systems, we used silver behenate for detector calibration.

Measurements at Penn employed Cu Kα radiation (λ = 1.542 Å) from a Bruker-Nonius FR591 generator operated at 3.4 kW. Collimation was accomplished using Osmic Max-Flux confocal optics and circular pinholes. The scattered X-rays were collected using a Bruker Hi-Star area detector. Data were collected at sample-to-detector distances of 54 cm and 150 cm, and the data merged to create a single profile over the range $q \in [0.008, 0.35]$ Å^{-1} (with $q=(4\pi sin\theta)/\lambda$, where θ is the half-angle between the beam and the detector). To minimize background, an integral vacuum with a pressure of <0.3 mbar was maintained along the entire flight path. The sample temperature was maintained at 10 ºC by a TMS 94 and LNP 94/2 temperature control unit (Linkam Scientific Instruments Ltd, Waterfield, Tadworth, United Kingdom).

The measurements at the NSLS were made at an incident X-ray energy of 8.00 keV. These measurements were done in air due to constraints of the facility's sample holder geometry. Each measurement was made at a single sample position but with two MAR-CCD detectors with wave vector ranges of $q \in [0.002, 0.08]$ Å^{-1} and $q \in [0.08, 0.6]$ Å^{-1}. These two measurements were then merged to create a single profile.

For all X-ray scattering measurements, dissected pieces of lens tissue were mounted in custom-built o-ring-sealed sample holders with clear ruby mica windows (7.5 mm and 12 mm in diameter, Attwater Group, Preston, England). For these measurements, a transparent and ice-cold *D. pealeii* lens was carefully dissected into four concentric layers and the dissected pieces were individually loaded under a stereoscope so as to completely fill sample holders with lens tissue to a total volume of approximately 45 µL. The sample holders were carefully sealed and no leaks were observed during the scan. The measurement time was 30 minutes at our home source, and 5 minutes at NSLS, for each sample at each angle range. We did not observe time dependence in the scattering patterns for any sample. Primary data reduction was performed using Datasqueeze (www.datasqueezesoftware.com). $I(q)$ data were then normalized to the value at maximum q or to the volume fraction of the sample as predicted by its refractive index; both approaches yielded very similar results.

Form factor and structure factor

The spherically averaged protein form factor was estimated as a function of wave vector, q using the atom positions from our homology models and the Debye formula as follows:

$$I_q = \sum_{i,j} \frac{f_i(q) f_j(q)}{\sin(qd_{ij})(qd_{ij})}$$

(*35*).

Here, f_i is the form factor of the *i*th atom, which is approximated by the number of electrons in the atom. The variable *d* is the pair distance between atom *i* and *j*. We then obtained an averaged form factor for all transcripts in the lens by taking the average of this representation for each individual protein isoform. Given the range of isoforms of S-crystallin, we used an average form factor for 20 different homology models with roughly equal quality scores as determined by SwissModel (http://swissmodel.expasy.org/docs/references), however form factors for all homology models were quite similar. Then, the structure factor of lens tissue at various radial positions was determined by dividing the scattering intensity as a function of angle by this average of form factors.

DAMMIF

We used the algorithm DAMMIF (v. 1.1.2 (r3709)) to study the spatial distribution and configuration of S-crystallins at different radial positions in the lens (*26*). Importantly, the only input into DAMMIF is X-ray scattering intensity as a function of wave vector *q,* there are no other assumptions or inputs about protein or molecular structure available to the algorithm. The software component GNOM calculates a radial distribution function from scattering intensity as a function of wave vector *q*; we used program parameters of zero condition at $r=r_{min}$ and non-zero condition at $r=r_{max}$, with r_{max} = 200 Å. A single run of DAMMIF uses the output from GNOM alone to calculate a likely three-dimensional protein spatial distribution via a Monte Carlo fitting algorithm; DAMMIF assumes that these structures are spatially continuous, and we considered this to be a valid assumption in the length range considered given our result from SAXS. For DAMMIF, we used a simulation sphere diameter of 200 Å and particle diameter 4.2 Å, consistent with a single particle representing a single amino acid residue; no symmetry was presumed.

Because DAMMIF is a Monte Carlo algorithm, such that the structures generated by individual runs of the program are approximately equally likely but not unique, it was possible to generate an ensemble of many similar and equally likely protein configuration geometries. We reasoned that this ensemble of configurations would be similar to an ensemble of gel fragments, each with the diameter of the largest D-spacing measured by our SAXS experiments, that together compose the larger gelled system. To find this ensemble of lens gel fragments, we ran DAMMIF 100 times with the same input from a single one-dimensional SAXS measurement from each region of the lens, a process which generated a slightly different predicted network configuration each time. We then manually counted the coordination numbers at furcations in the predicted network, and calculated the average and the standard deviation of network coordination number. To test whether the measured differences in network coordination number were significantly different between different samples, we used a partition test. This test assumes that all observations come from the same distribution, and then asks if the two input distributions can be recapitulated by resampling half of the input observations.

GROMACS

Molecular dynamics (MD) simulations were performed in GROMACS (v. 5.0) with the AMBER99SB*-ILDNP forcefield, explicit SPC216 water, and Ewald summation (*36, 37*). We modeled the loops in the sequences DORPE21 and DORPE13, named for the species "*Doryteuthis pealeii*" and the number of amino acids in the loop. To minimize the number of atoms in the atomistic simulation, we split the dimers obtained by homology modeling into monomers, and modeled the loop-loop interaction between two monomers originally located in two different dimers. To obtain an initial conformation for each lens protein, structures predicted using homology modeling were equilibrated in water for 10 ns (*38*). We used the structures present in the final frame of each of these trajectories in a longer simulation of loop-loop interaction. The centers of mass of the monomers were separated by 7.5 nm, as observed in DAMMIF, and at the start of the simulation, no atoms in the loops were closer than 0.8 nm.

The resulting structure was modeled as described above for 100 ns. H-bonding interactions throughout the simulation were analyzed with custom scripts written using BioPython. Hydrogen bonds were identified and their energies computed according to the criteria in the PyMol v. 1.8 protein visualization software (Schrodinger, LLC).

To estimate the relation between the physical loop length in Angstroms versus the number of amino acids composing the loop, we fit the loop length in these simulations to the number of amino acids in the loop, and estimated the ratio using a linear relation.

Simulation of branching networks of particles

To estimate the structure factors of simple networks of particles with varying average coordinate number, we built a computer simulation of branching spheres using MATLAB 2014b (Mathworks, Inc.). Code for this simulation is available upon request from the authors. Our simulation placed spheres of the same diameter but with different ratios of M=2 and M=3 in branching chains and calculated the structure factor of the resulting network. For each simulation, the total number of particles was 10,000. An initial particle was generated as the nucleus of the structure, and then subsequent particles were added using a walk algorithm. This walk algorithm is therefore similar to a three-dimensional self-avoiding random walk, with a few exceptions to better replicate the predicted behavior of proteins in the lens: the position of a new particle position is normally distributed along the x-axis from the previous particle; one simulation step may involve adding either one or two particles; and the angular position of a particle was chosen from a Gaussian distribution to approximate the radius of gyration of an S-crystallin protein. The probability of adding two particles vs. one particle in each simulation depended upon the fixed ratio of M=3:M=2 in each simulation, with two particles added to the growing chain at the frequency of M=3; if two particles were added, a bifurcation in the network resulted. In the subsequent step, both new branches in the network continued adding more particles, with each of the branches having the original probability of further bifurcation. To avoid stalling when no additional self-avoiding growth was possible in a given chain, the algorithm specified the number of times to attempt to add an additional particle to a chain while

avoiding a collision before stopping the growth of that chain. However, the general features of the resulting structure factor were not sensitive to this "attempt" parameter. The particle-particle separation was set to be 36 Å, or roughly two times the radius of gyration of an S-crystallin. The azimuthal angular position of particles added to the network was chosen from a Gaussian distribution of particle-particle angles with a mean of 1 and variance of 0.2, and we used a uniform distribution for the altitudinal angle between particles. The ratio M=3:M=2 was chosen to be 0.05, 0.10, 0.45, and 0.9 in different simulations to capture the range of networks predicted in the lens, up to the density where the structures are no longer well described by bifurcating chains (Fig. 3). To estimate the structure factor of the resulting simulated particle network, we calculated the Fourier transform of the resulting particle positions using the Debye method (Fig. 3C).

Lens transcriptome generation

In this technique, RNA molecules are isolated from whole lens tissue and enriched for protein-coding messenger RNAs (mRNAs) via chemical selection of a polyadenine tail. The mRNAs are then enzymatically fragmented, linker sequences are added, they are bound to a flow cell, enzymatically amplified, then sequenced by serially imaging individual additions of fluorescent nucleotides during another round of enzymatic duplication. The resulting sequences of fragmented mRNAs are then reassembled in silico.

RNA was extracted from tissue samples using the Qiagen RNeasy Mini Kit (Cat. No. 74104) according to manufacturer instructions and normalized to 25 ng/μL. Library construction and sequencing for 100% layer and 60% layer was performed by the University of Utah High Throughput Genomics Core Facility. Libraries were prepared from total RNA using the Illumina truSeq stranded mRNA kit (cat# RS-122-2101, RS-122-2102), with polyA selection to remove ribosomal RNA. Libraries were assessed for fragment length, concentration, and molarity with a NanoDrop spectrophotometer, Agilent 2200 Tape Station using a D1K (cat# 5067-5361 and 5067-5362) assay and KapaBiosystems Kapa Library Quant Kit (cat# KK4824). Libraries were sequenced with the HiSeq 101 cycle paired-end sequencing protocol using read adapters AGATCGGAAGAGCACACGTCTGAACTCCAGTCA (read 1) and AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT (read 2). The samples were pooled with one other samples in a single lane, using barcodes TAGCTT (100% layer) and GGCTAC (60% layer).

Illumina data were assembled with Trinity (version r20140413p1), using Transdecoder to generate putative amino acid sequences (*39*). Predicted S-crystallins were identified with HMMER using a hidden Markov model constructed from known S-crystallins (*40*). Only gene candidates were retained for further analysis.

SDS-PAGE

In this technique, protein is solubilized in an anionic detergent, loaded into a slab of low-density polyacrylamide gel and a voltage is applied to the gel. The proteins unfold and become coated in negative charge, with the result that a protein's mobility due to the applied voltage is proportional to the logarithm of the molecular weight. After visualizing the protein, the mixture of molecular weights in a given protein sample can be deduced.

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

We calibrated molecular weights of these samples using protein molecular weight ladder (Bio-Rad Precision Plus Protein Kaleidoscope Standards). The relative migration to the dye front is fitted as an exponential function of the molecular weight. The part between protein bands were treated as background and this background was subtracted with the Matlab function *msbackadj*. The greyscale intensities of the samples were then normalized relative to total S-crystallin protein content by dividing by the sum of all values in the range of 20 to 40 kDa for a single sample.

To compare the molecular weights of predicted protein transcripts to our SDS-PAGE result, we estimated the contribution of each predicted protein transcript by modeling the observed SDS-PAGE migration pattern as the sum of many Gaussian curves, with each individual transcript contributing a Gaussian distribution of density of fixed width consistent with a single protein transcript, and a height as determined by the following fitting algorithm. The positions of the Gaussian curves were fixed to have center positions within 1 kDa of a predicted transcript a maximum width of 0.5 kDa for transcripts of 23.5 kDa to 28 kDa. For the protein at ~35–37 kDa, we fit with one peak with the position ranging from 35 to 37 kDa, and a maximum width of 0.5 kDa.

Estimate of lens <*M*> via dilution

To study the phase properties of S-crystallins from different radial positions, lenses were dissected into four concentric layers as described above; the dissection and the following procedures were conducted on ice. Each tissue sample was homogenized in a ground-glass tissue grinder in 700 µl molecular biology-grade water with 1:100 protease inhibitor cocktail (P8849-5mL, Sigma Aldrich, St. Louis, MO). Protein concentrations were then measured via absorption at 280 nm using a NanoDrop spectrophotometer and adjusted by adding additional water such that each sample had a final volume fraction of 0.01. Homogenized tissue samples were then transferred to pre-weighed microcentrifuge tubes. The resulting preparations were then centrifuged at 14.8k rcf for 10 min at 10 ºC.

Next, the supernatants that resulted after centrifugation were separated from the pellet using a Pasteur pipet, taking care not to disturb the pellet. The volume of this supernatant sample was measured by withdrawal into a volumetrically calibrated pipetter, and the concentration was again measured via absorption at 280 nm. The remaining pellet wet weight was then measured via microbalance. To determine the protein packing fraction in these pellets, the pellets were lyophilized overnight and then re-weighed on the microbalance. The water volume and mass in the pellets before lyophilization were then calculated using the weight differences of the pellets before and after lyophilization. The total pellet volume was then calculated using the partial specific volume of the S-crystallins of 0.72 ml/g (this value resulted from the protein amino acid sequences, see above). For samples taken from lens radii <60%, granular pellets resulted from this procedure; we assumed that the protein component of these pellets had a packing fraction of 0.65 within the total pellet volume, consistent with random spherical packing of the protein-granule component. We interpolated the intersections of the spinodal lines for each $<M>$ with the x-axis from Bianchi et al., 2006 (*12*) along the x-axis of the phase diagram. We estimated $<M>$ of our pellets by finding the closest spinodal line to the density after this dilution-induced phase transition. We assumed that all proteins in the supernatant fraction have $M = 2$, given that they form unbranching chains. Given the number of particles in the system with $M=2$ in the supernatant, in addition to the number of particles of a higher given $<M>$ from the density of the pellet, we were then able to calculate the effective $<M>$ of the system before dilution as a function of lens radius.

Lens protein homology modeling

S-crystallins' tertiary and quaternary structure were estimated using the Swiss-model server (http://swissmodel.expasy.org/docs/references) (*38*). The predicted S-crystallin models were built based on homo-oligomeric structures from selected proteins from SWISS-MODEL template library. The server's algorithm always chose the squid sigma-class glutathione s-transferase structure (1gsq) as the best template for our structural modeling of our S-crystallin sequences from the lens transcriptome (*38, 41*). Since the "loops" present in the sequence are not homologous to any sequences or structures in the Protein Data Bank, we submitted them for secondary structure prediction to the prediction server Jpred (*42*). The ability of the S-crystallin loops to form amyloid fibrils both internal to the loop and with other loops was evaluated using the algorithm in ZipperDB (*43*).

Central protein packing fraction

Our finding that the packing fraction of protein at the center of the lens is very close to 1 was surprising but supported in two ways. First, a sample of tissue from the central core of the lens did not change mass after maceration and exposure to high vacuum overnight (the sample's mass was 4.22 mg before and after an overnight exposure to a pressure of <0.3 mbar at 20ºC) suggesting there is little or no bulk water or other solvent present. Second, we submerged samples from the core of the lens in solutions of potassium iodide of known density and observed their buoyancy. This experiment showed that the density of the lens core is 1.28 g/ml;

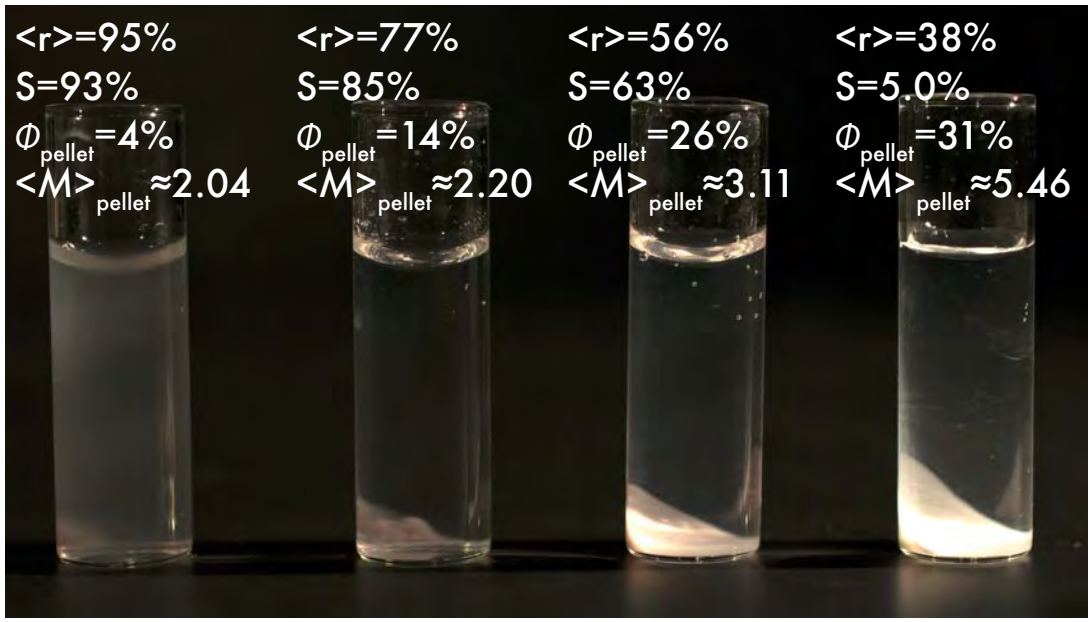
if the partial specific volume of S-crystallin is 0.73 g/ml then the packing fraction of S-crystallins in the core must be 0.93, and the refractive index is 1.63.

Solution SAXS on diluted lens tissue

Squid lenses were dissected into concentric layers as described above. The whole tissue from the regions with a maximum of 100% of the radius and 80% of the radius were placed in one of a few concentrations of phosphate-buffered saline in a 1.5 ml tube. These preparations were then placed on an oscillating mixer on low speed at 4ºC overnight. The resulting preparations were transparent, homogeneous fluids, with no remaining lens tissue, phase separation, or precipitates. After this mixing procedure, the tubes were centrifuged at 14.8k rcf for 10 minutes at 4ºC; a sparse, translucent pellet resulted. The protein concentration of the supernatant was consistently around 10 mg/ml, with the exact concentration depending on the exact size of the lens portion that was dissolved. These solutions were then characterized via SAXS using procedures described above. For our concentration experiment, serial two-fold dilutions were performed and each new dilution measured, until we reached the concentration at which measurements exhibited low signal to noise ratios.

In order to subtract the scattering generated by the buffer from these measurements, we normalized measurements of both the empty buffer and the protein solutions by the exposure time of the measurement and the beam intensity at q=0, then subtracted the resulting normalized signal of the buffer from that of the protein solutions.

These SAXS measurements from the high concentration solutions were input into the algorithm GNOM, which estimates a $g(r)$ from the input $I(q)$ data (*26*). The output from GNOM was then analyzed by the algorithm AMBIMETER as implemented via the ATSAS-online server hosted by the European Molecular Biology Laboratory–Hamburg (https://www.embl-hamburg.de/biosaxs/atsas-online/, accessed 01/17).



<r>=average radial position of sample
S=fraction of total protein in the percolating fluid
Φ_{pellet}=measured volume fraction of gel-like pellet
$\langle M \rangle_{pellet}$=estimated <M> of the pellet phase given Φ_{pellet}

Figure S1. Concentric sampling of a single lens, each concentric sample diluted to a constant volume fraction of 0.01. This photograph shows the entirety of a single lens, divided into four concentric samples, all diluted to a constant volume fraction of 0.01. A like volume of each subsample is shown here for ease of presentation. As the radial position from which the sample was taken decreases, the proportion of total protein in the pellet material increases, although the mass of protein in each vial is the same. The data for average radial position ($\langle r \rangle$), fraction of total protein in the supernatant (S), the volume fraction of the pellet (Φ_{pellet}) and associated coordinate number of the pellet ($\langle M \rangle_{pellet}$) are shown superinposed on each vial.

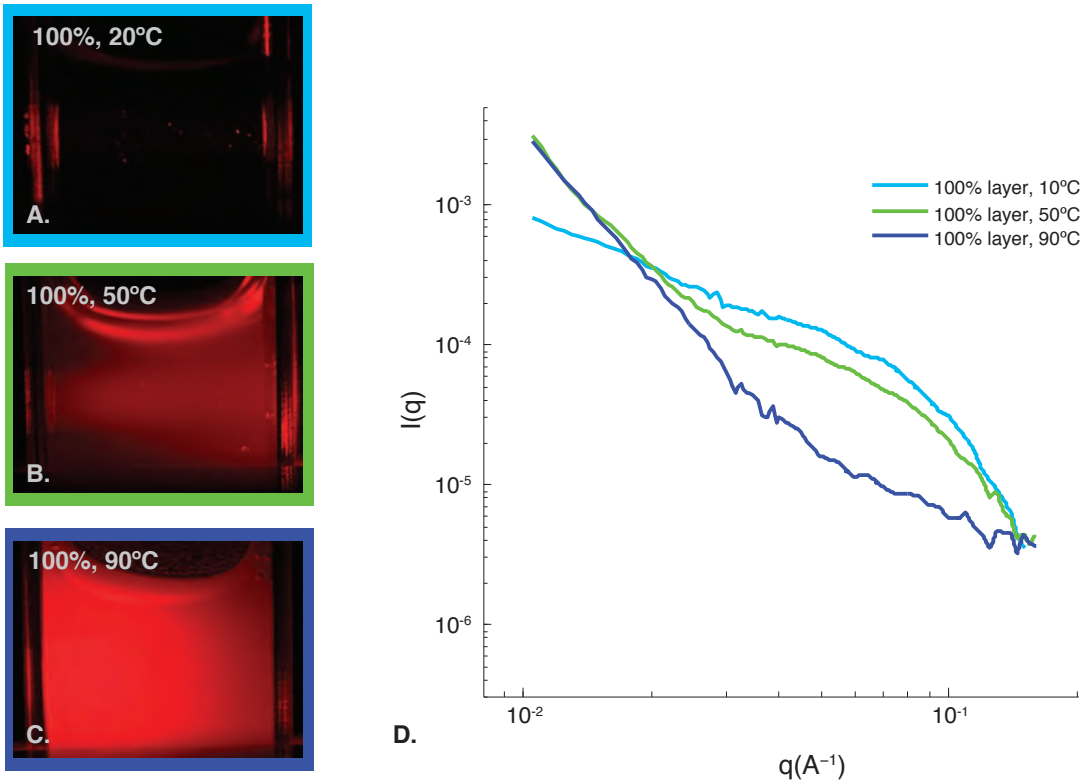


Figure S2. The fluid portion of phase-separated lens tissue results in a transparent solution but denatures and scatters upon heating. A) Peripheral layer lens tissue diluted in 20 mM tris buffer and centrifuged for 10 minutes at 13k relative centifugation factor, isolating the fluid phase shown in figure S1. Sample is in a 1-cm pathlength cuvette and illuminated by a laser on the left surface of the cuvette. Little or no scattering of the laser from the protein sample is visible. **B)** Sample in **A**, now heated to 50ºC. Significant side-scattering of the laser from the protein in the sample is apparent. **C)** Sample in **A** and **B**, now heated to 90ºC. Sample is highly scattering, nearly opaque to the laser. **D)** Small-angle X-ray scattering from samples in **A**, **B**, and **C**. Scattering pattern in **A** (at 10ºC) is consistent with chains of spheres consisting of folded protein (see figure S2 for more detail). Scattering in **B** (at 50ºC) is consistent wiith clusters of spheres consisting of folded protein. Scattering in **C** (at 90ºC) is consistent with clusters of unfolded protein.

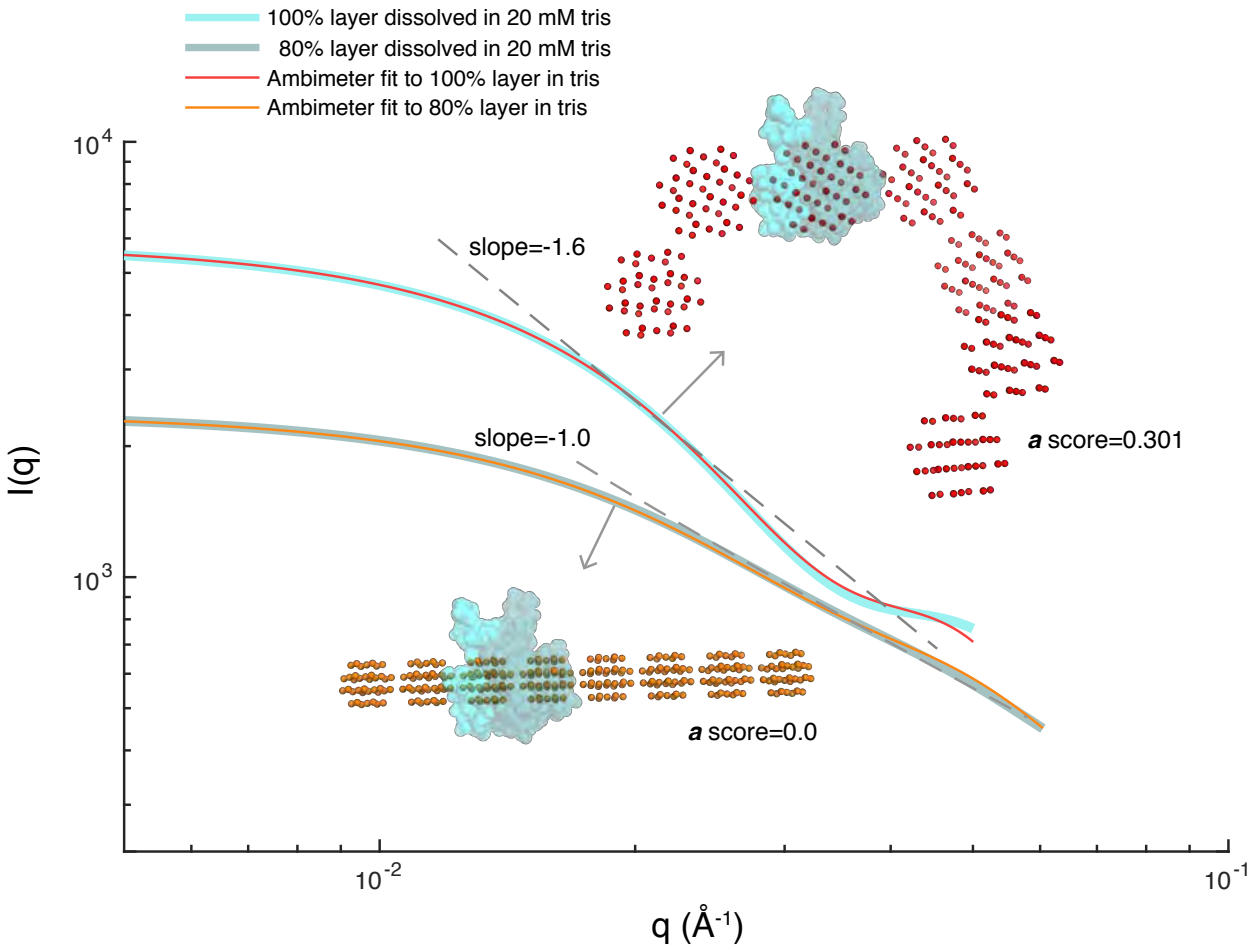


Figure S3. Solution scattering and structural fits of squid lens tissue. Squid lens tissue from the peripheral, sparse layers of the lens was agitated in 700 µl of 20 mM tris buffer, and SAXS measurements were performed on the resulting solutions. Crystal structures of S-crystallins are shown in transparent blue for reference. Light blue curve shows intensity as a function of scattering angle for a sample from the peripheral, 100% radius region of the lens. Dark blue curve shows scattering intensity as a function of angle for the 80% radius region of the lens. Red curve shows AMBIMETER fit for the 100% region, orange curve shows this fit for the 80% region. Red spheres show AMBIMETER structural prediction for the 100% region of the lens, while orange spheres show structural fit for the 80% region of the lens. Dotted guidelines show log-log slope relationships for reference. "*a*-scores" are ambiguity scores reported by the AMBIMETER algorithm; any a-score of less than 0.8 is a likely unambiguous assignment of structure.

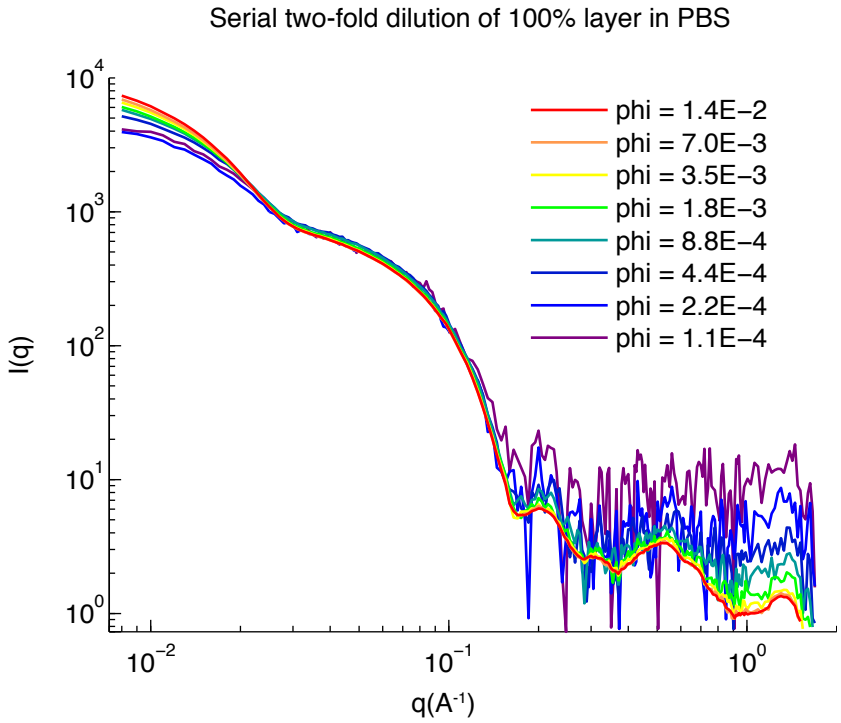


Figure S4. Serial dilution of dissolved lens tissue. Curves show X-ray scattering from serial two-fold dilutions of solutions of the proteins from the 100% radius region of the lens. The calculated form factor of S-crystallin from the crystal structure coordinates is shown for reference. Decreasing S-crystallin volume fraction from a maximum of 1.4e-2 to 1.1e-4 does not alter the overall structure of the soluble protein in the sample, and all measurements are consistent with chains of S-crystallins linked in a pairwise fashion. This result is predicted by the patchy colloidal phase diagram, which shows that these systems will transition from equilibrium gels to equilibrium percolating fluids upon dilution.