Pharmacological Reversal of Cataracts Caused by an Amyloid-Prone Small Heat Shock Protein

Leah N. Makley[1,†], Kathryn A. McMenimen[1,†], Brian T. DeVree[3], Joshua W. Goldman[2], Brittany N. McGlasson[2], Ponni Rajagopal[4], Bryan M. Dunyak[1], Thomas J. McQuade[5], Andrea D. Thompson[1], Roger Sunahara[3], Rachel E. Klevit[4], Usha P. Andley[2], and Jason E. Gestwicki[1,5]

SUPPLEMENTAL INFORMATION

Materials and Methods

Protein purification. N-terminal hexahistidine-tagged human heat shock protein 27 (His_6Hsp27) was expressed in *Escherichia coli* BL21 cells using the pET 28b vector. Human, wild-type αB-crystallin and R120G αB-crystallin were purified in *E. coli* Rosetta cells using the pMCSG7 vector. Cultures of transformed cells were grown to an optical density of 0.8 AU at 600 nm in lysogeny broth media supplemented with 50 μg/mL kanamycin (pET 28b) or ampicillin (pMCSG7) at 37 °C. Cultures were induced with 500 μM IPTG and incubated at 30 °C for 8 to 16 hours. After incubation, cells were harvested by centrifugation and suspended in 1/25 culture volume of lysis buffer (20 μM Tris pH 8.0, 100 mM NaCl, 6 M urea, 5 mM β-mercaptoethanol, 15mM imidazole). The suspension was sonicated on ice and insoluble material removed by centrifugation. The supernatant was loaded onto Ni-NTA resin (10 mL/L culture volume) and washed with 10 column volumes of wash buffer (lysis buffer containing 30 mM imidazole). The proteins were eluted with elution buffer (lysis buffer with 150 mM imidazole) and the fractions containing protein were pooled. EDTA was added to 5 mM, and the unfolded protein was concentrated to ~20 mg/mL in 10 kDa MWCO Amicon centrifugal filter units. The proteins were refolded by injecting 1 mL samples of concentrated protein onto a Superdex 200 HR 10/30 column equilibrated with refolding buffer (20 mM sodium phosphate, pH 7.2, 100 mM NaCl) at room temperature. The refolded protein eluted as an oligomer (~24-mer) with a diameter of ~165 Å, consistent with previous literature values (*1*). Protein was concentrated to ~10 mg/mL before flash freezing and storing at -80 °C. The melting temperature (T_m) of Hsp27 by both circular dichroism and differential scanning calorimetry was 72°.

High throughput DSF screen. The screen was performed using a Thermofluor® instrument (a generous gift from Johnson & Johnson) in Abgene black 384-well PCR plates. Each well received 7 µL of 10 µM Hsp27 in 50 mM $NaPO_4$, pH 7.4, 700 mM NaCl, 50 mM LiCl, and 100 µM bis-ANS, with test compounds added via pin tool (~200 nL) to achieve final concentrations between 20 to 40 µM (1% DMSO). The chemical library was composed of the Spectrum MS2000 and NCC collections of known bioactives. Reactions were covered with silicon oil to limit evaporation. The plates were measured in up/down mode, which was empirically determined to give better signal-to-noise than the continuous ramp mode. Plates were heated from 65 °C to 80 °C in 1 °C increments, equilibrated for 130 seconds at each high temperature, cooled to 25 °C, and held for 10 sec at 25 °C prior to imaging with a single 10 sec exposure for each temperature reading. Plate uniformity tests measured the T_m of Hsp27 at 72.3 ± 0.16 °C. The Z factor was calculated to vary between ~0.59 and 0.71 and the CVs were 8%. Treatments that reduced the apparent T_m by 3SD were considered actives.

HTS data analysis. The output of the ThermoFluor® assay is a set of fluorescence intensities at increasing temperatures for each well. We found that a single curve-fitting algorithm will occasionally misidentify the T_m for a well. Accordingly, we minimized this problem by applying three distinct data fitting routines to the same data and using the resulting T_m values as pseudo-replicates. Two curve fitting routines were performed using the Thermofluor® analysis software (version 1.3.7). A sigmoidal curve fit option was used with non-constant y-upper and y-lower slopes and constant dCp = 2500. A derivative curve fit was also used with a polynomial degree = 2 and filter points = 3. The 'detect dropouts' filter was used for both curve fits with a maximum deviation multiplier of 2. The first, sigmoidal curve fit uses a non-linear least squares regression to fit the equation below, taken from Matulis *et al.* (*2*)

$$Y(T) = Y_u + m_u T + \frac{(Y_f - Y_u) + (m_f - m_u)T}{1 + e^{\left(\frac{-\Delta H_u}{R}\right)\left(\frac{1}{T} - \frac{1}{T_m}\right) + \left(\frac{C_{pu}}{R}\right)\left(\left[\frac{T_m}{T} - 1\right] + \ln\left[\frac{T}{T_m}\right]\right)}} \quad \textbf{(1)}$$

where $Y(T)$ is the response at temperature T, Y_u and m_u are the intercept and slope, respectively, for the fully unfolded protein baseline, Y_f and m_f are the intercept and slope for the

folded protein baseline, ΔH_u is the enthalpy of unfolding, T_m is the melting temperature, and Cp_u is the heat capacity of unfolding. This equation tended to be difficult to fit to data sets with non-horizontal baselines. The second method approximates the derivative at each data point using a least-squares fit polynomial of specified order and finds the T_m as the maximum value of the derivative. This fit is sensitive to noise and to multiple inflection points in a given data set. The third method used an iteratively reweighted non-linear least squares regression to fit the equation below:

$$Y(T) = mT + c\left(\frac{T^n}{T_m^n + T^n}\right) + Y(T_m) - mT_m - \frac{c}{2} \quad \textbf{(2)}$$

where m is the slope of the parallel folded and unfolded protein baselines, c is a constant multiplier, n is the Hill slope, and other symbols are as in equation (1). This is a version of the Hill equation that allows for non-horizontal, parallel baselines. The fitting routine was coded in Matlab and iteratively fit the equation to the window of temperatures 6.5 degrees above and below the T_m according to the relationship $Weight(T_i) = \cos^{1/2}\frac{\pi(T_i - T_m)}{2 \times 6.5}$, where T_i is the temperature of each data point within the window and the weight for all other data points was set to 0. The weighted data was then fit to the Hill equation using least-squares nonlinear regression and the resulting T_m estimate was used to update the weights and re-run the curve fit until a sum of squares tolerance of 0.01% or 20 iterations was reached. This approach decreased the curve fit's sensitivity to non-ideal baselines, giving better estimates in many of the cases where the first, sigmoidal curve fit performed poorly.

To define active molecules, the three curve-fitting methods were to analyze each melting curve. To normalize any systematic bias between the fits, we calculated ΔT_m by subtracting the average of the 32 negative control wells on a plate from the T_m calculated from the same method for each well. This resulted in three directly comparable estimates of the ΔT_m for each well. These ΔT_m values were used to calculate an estimate of the mean and standard deviation for a normal distribution of both the positive and negative standards on each plate. We used a Bayesian inference scoring method written in Matlab using the Statistics Toolbox add-on package. The posterior probability that a compound belonged to the 'Hit' hypothesis (H_{Hit}) as opposed to the 'Null' (H_{null}) or 'Fail' (H_{fail}) hypotheses given the data (D) was calculated according to Bayes' theorem as shown below.

$$P(H_{Hit} \mid D) = \frac{P(H_{Hit}) \cdot P(D \mid H_{Hit})}{P(H_{Hit}) \cdot P(D \mid H_{Hit}) + P(H_{Null}) \cdot P(D \mid H_{Null}) + P(H_{Fail}) \cdot P_{average}(Standards \mid H_{Hit}\ or\ H_{Null})}$$

This Bayesian inference method tested the three ΔT_m values from each well against three hypotheses: 1) the ΔT_m measurements indicate that the drug in that well did not change the T_m of the protein (abbreviated H_{Null}) 2) the ΔT_m measurements indicate that the compound in the well changed the T_m of the protein (H_{Hit}) and 3) the ΔT_m values did not belong well either of the two hypotheses, suggesting that the well simply failed (H_{Fail}). The prior probabilities for each hypothesis were set the values $P(H_{Null}) = .979$, $P(H_{Hit}) = .02$, and $P(H_{Fail}) = .001$. The likelihood of observing the data D given hypothesis H_{Hit} was calculated by assuming the three ΔT_m values were independent and identically distributed samples taken from a normal distribution with a mean equal to the mean of the data itself and a standard deviation equal to the estimate calculated from the positive standards on the plate from which the data was taken. The likelihood was normalized to a value between 1 and 0 by division by the maximum likelihood value for that particular distribution (i.e. the likelihood of observing three data points exactly equal to the mean). The likelihood for observation of the data D given hypothesis H_{Null} was calculated similarly, but used a normal distribution with a mean and standard deviation equal to that of the negative standards on the plate from which the data was taken. Because of the inability to reliably calculate the likelihood that a particular well failed from only three pseudo-replicate ΔT_m measurements, we chose instead to introduce a constant into the denominator that represented an ever-present possibility of failure that was not dependent on the data for that particular well. The constant is composed of the prior probability of hypothesis H_{fail} and the average likelihood that the three ΔT_m values from each positive/negative standard on the plate were taken from the overall normal distribution that was fit to all the positive/negative standards on that plate, indicated as $P_{average}(Standards|H_{Hit} or H_{Null})$. Using this likelihood value automatically scales the constant to be comparable with the likelihoods used to evaluate the H_{Hit} and H_{Null} hypothesis, since they are also computed with the same standard deviations and similar data sets. Thus, when the neither H_{Hit} and H_{Null} describe the data well, the denominator is dominated by the constant and both H_{Hit} and H_{Null} are given low posterior probabilities. The final score was calculated by multiplying the posterior probability that H_{Hit} is true by the absolute value of the ΔT_m, times 100. This method rewarded wells that showed large temperature changes and penalized wells that produced widely scattered ΔT_ms.

Electron microscopy. Purified recombinant R120G αB-crystallin was thawed on ice, diluted to 15 μM and treated with either solvent (1% DMSO) or 100 μM compound **29**, then incubated for 30 min at ambient temperature before applying to glow-discharged Formvar/carbon 300-mesh copper grids (Electron Microscopy Sciences) and staining with 2% uranyl formate. Grids were visualized on a Phillips CM-100 transmission electron microscope at an accelerating voltage of 80 kV with magnification settings ranging from 10500–92000×. Shown are representative images and at least 12 random fields were imaged per treatment condition. Scale bars represent 1 μm. Disaggregation experiments were performed similarly. R120G αB-crystallin was incubated at ambient temperature for 30 minutes and then a sample of this solution was visualized by EM to ensure fibril formation. The aggregated cryAB was treated with either solvent (1% DMSO) or compound 29 (100 μM) and incubated an additional 44 hours at ambient temperature. No change in amyloid structure or prevalence was observed in the vehicle control between the 44 hour and 30 minute time points.

Binding to cryAB by biolayer interference (BLI). R120G αB-crystallin (100 μL of 2.2 mg/mL) was reacted with a twenty fold excess of NHS-PEG_4-biotin (Thermo Scientific) in DMSO for two hours at 4 °C with gentle rocking. The excess biotinylation reagent was then removed by dialysis against phosphate buffer (20 mM Na_2PO_4 pH 7.2, 100 mM NaCl) at 4 °C for 20 hours. To measure the affinity of compounds for this protein, the biotinylated R120G αB-crystallin was then immobilized on super-streptavidin pins (ForteBio) as follows: pins were first equilibrated in phosphate buffer for ~10 min before moving them to wells containing either 200 μL 150 μg/mL biotinylated His_6-R120G αB-crystallin or 200 μL 100 μg/mL biocytin for 60 min. The pins were washed in buffer for 5 min, free streptavidin sites were quenched with 100 μg/mL biocytin for 10 min, and the pins were washed again for at least 5 min prior to beginning an experiment. All immobilization steps were carried out on-line on a ForteBio Octet Red96 instrument with 1000 rpm rotary shaking throughout. Approximately 7 nm of biolayer interference shift was consistently obtained from R120G αB-crystallin immobilization under these conditions. For each experiment, compounds were first diluted into phosphate buffer and the final DMSO concentration was held constant at 1%. Compounds were incubated with the pin surface for 5 min and tested from high to low concentration and then the experiments repeated from low to

high concentration. Between experiments, pins were stored in phosphate buffer at 4 °C, and 50 μM compound **29** was used as a positive control at the beginning of each experiment to ensure the quality of the surface. Binding data was analyzed using the ForteBio software, version 7.0. Binding curves were aligned to the association phase from 0.2-0.4 s and either a matched concentration, biocytin-blocked pin response or a DMSO-only. The R120G cryAB-pin response was subtracted for each compound concentration to correct for drift and bulk solvent effects. The association response from 290-300 sec was averaged for each binding curve and used to estimate the equilibrium binding affinity. Data were fit in GraphPad Prism, version 5.04, using a one-site specific binding model. The weighted mean and weighted standard deviation of the apparent K_d values across multiple experiments was calculated using $\bar{x} = \frac{\sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$ and $\sigma_{\bar{x}}^2 = \frac{1}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$.

NMR. ^{15}N-R120G-ACD was expressed and purified as described previously (*3*). Resonances in R120G-ACD have been assigned by standard NMR techniques. HSQC spectra were acquired on a Bruker Avance III 500 MHz spectrometer equipped with a triple resonance, z-axis gradient probe. 10 mM stock solutions of compounds **29** and **16** were made in deuterated DMSO. Spectra were acquired on a sample containing 100 μM R120G-ACD and either 1 mM compound 29 or compound 16. Since these samples contained 20% DMSO, the R120G-ACD plus compound spectra were compared to a control spectrum of ^{15}N-R120G-ACD containing 20% DMSO. 256 scans were acquired per t_1 value for a total of 200 t_1s at a temperature of 32 °C. Spectral widths of 6009 Hz and 1419 Hz were used in the ^{1}H and ^{15}N dimensions, respectively. Spectra were processed with NMRPipe (*4*) and analyzed with NMRView (*5*). Docking was performed on cryAB as previously described (*6*) Briefly, the alphaB-crystallin dimeric unit was isolated from the crystal structure (PDB: 2WJ7) and prepared for docking with UCSF Chimera. Low energy ligand conformations were generated in Szybki and partial charges were calculated with the AM1-BCC force field. Flexible ligand docking was then performed in UCSF Dock (v6.6) using the iterative anchor-and-grow method.

Animals and lenses. R120G cryAB knock-in mice expressing an A to G point mutation in codon 120 of mouse cryab which results in the substitution of arginine 120 with glycine (R120G) were generated as described previously (*7*). These mice are in the C57BL/6J background (Jackson

Laboratories, Bar Harbor ME). Mice were maintained at Washington University School of Medicine in the Division of Comparative Medicine by trained veterinary staff. All protocols and procedures involving mice were approved by the Washington University Animal Studies Committee and performed by Mouse Genetics Core. Mice were euthanized by CO_2 inhalation and lenses were immediately dissected and used for analysis.

Treatment of mice. Wild type or R120G cryAB knock-in mice were treated with vehicle (40% cyclodextrin in PBS) or compound **29** dissolved in vehicle to make 5 mM solutions. The drops were administered in six doses over two weeks, coinciding with each Monday, Wednesday, and Friday during the period of experimentation. To the left eye of each animal was administered one drop of vehicle so that the drop coated the eye. One drop of compound **29** was administered to the right eye of each animal in the same manner. R49C cryAA knock-in mice were treated in a similar manner. The concentration of vehicle and compound **29** was reduced 5-fold in some experiments as indicated in the text and legends to the figures.

Slit lamp biomicroscopy. Mice were examined by slit lamp biomicroscopy on the day following final dosing. Prior to observation, pupils were dilated with 1% Topicamide Ophthalmix Solution, USP and 10% Phenylephrine Hydrochloride Solution, USP in a 9:1 ratio. The slit was positioned directly orthogonal to the mouse for best delineation of opacities and clear areas in the lens. Videos from each eye were recorded on a Sony DCR-DVD403 3MP DVD Handycam Camcorder w/10x Optical Zoom with Carl Zeiss Vario-Sonnar Super Steady Shot (Sony, Tokyo) using the camcorder attachment on the slit lamp. Night-mode was used for best identification of opacities. Still images were then extracted from each video using Topaz Moment (Topaz Labs, Dallas TX) or Adapter (Macroplant, McKinney TX) and selected images were analyzed for intensity of light (as gray value across a selected cross section) using the ImageJ program (http://rsb.info.gov/ij). Plot profiles were produced along horizontal and vertical cross sections of each image. Improvements in transparency were qualitatively scored by visual analysis and assigned a grade based on the following criteria, modified from the LOCS III grading system (*7, 8*): stage 0, clear lens; stage 1, loss of normal appearance of anterior, nuclear, and posterior lenses as well as prominence of y-suture line; stage 2, discrete anterior changes accompanied by distinct nuclear opacity; stage 3, increased involvement of nuclear and cortical regions of the

lens accompanied by increasing opacity; stage 4, completely mature cataract involving the cortex and nucleus.

Measurement of water-soluble lens crystallins. The control and drug eyes from each mouse were removed and the lenses excised. Lenses were either placed in a unique microcentrifuge tube with 1 mL PBS and 1:1000 protease inhibitor cocktail (Sigma, St. Louis MO), or in two instances were pooled with lenses from other mice. The lenses were homogenized using disposable grinders before centrifugation for one hour at 10,000 rpm. The samples were kept at 4 °C for the entire process. After centrifugation, the soluble fraction (supernatant) was collected filtered through a 0.22 μm filter, transferred to a vial and analyzed by gel permeation chromatography using the TDA305 detector with GPXMax injection system (Malvern, Worcestershire, UK). The system provided refractive index, ultraviolet absorbance at 280 nm, viscosity, and right and low-angle-light-scattering readings after each sample passed through a P3000 and P2500 column (Malvern). PBS was used as the mobile phase and the system was calibrated with bovine serum albumin (Sigma) as a standard. The results were displayed using OmniSec 4.7 software (Malvern). Gel permeation chromatography data were analyzed using SAS version 9.3 (SAS Institute; Cary, NC). The effect of compound **29** on the level of soluble α-crystallins was expressed as the percent difference in area between the treated and untreated fellow eyes. To control for differences in assay sensitivity among animals, the area under the α-crystallin peak in the treated eye was compared to the area under the curve for the corresponding, untreated eye and expressed as percent difference. The percent difference between treated and untreated was compared using the Wilcoxon signed rank test for the null hypothesis of no difference. The association between age and percent difference was calculated using the Spearman rank correlation coefficient. Non-parametric statistical models were used to protect from departures from normality.

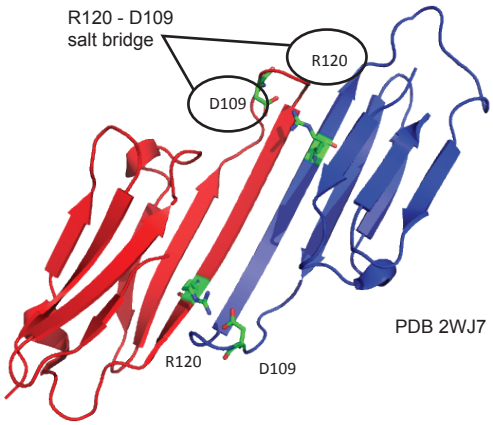
Treatment and analysis of human lenses. Human lenses (HL) used in this study were obtained from unidentified patients (70-80 year old) undergoing extracapsular cataract surgery. The

material had been classified as non-human discarded material. Preoperative clinical exam was used to classify the cataract as mixed cortical and nuclear 3-4 l for HL1. Each lens was cut into four nearly equal parts and placed in individual tubes. To each tube was added 50 µL solution of vehicle (8% cyclodextrin in PBS containing 1:1000 protease inhibitor cocktail) or compound **29** (1 mM) in vehicle which covered the lens tissue completely. Lens tissue was incubated in these solutions for 6 days in the dark at room temperature. On the last day, 150 µL of PBS containing 1:1000 protease inhibitor cocktail was added, the tissue was homogenized and an aliquot was removed and centrifuged for one hour at 10,000 rpm at 4 ℃ and the supernatant was collected. Protein content in the supernatant and total homogenate was measured using the BCA kit.

Citations:

1. I. V. Dudich *et al.*, Dimer structure as a minimum cooperative subunit of small heat-shock proteins. *Biochim Biophys Acta* **1253**, 163 (Dec 6, 1995).
2. D. Matulis, J. K. Kranz, F. R. Salemme, M. J. Todd, Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor. *Biochemistry* **44**, 5258 (Apr 5, 2005).
3. S. Jehle *et al.*, alphaB-crystallin: a hybrid solid-state/solution-state NMR investigation reveals structural aspects of the heterogeneous oligomer. *J Mol Biol* **385**, 1481 (Feb 6, 2009).
4. F. Delaglio *et al.*, NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277 (Nov, 1995).
5. B. A. Johnson, R. A. Blevins, NMR View: A computer program for the visualization and analysis of NMR data. *J Biomol NMR* **4**, 603 (Sep, 1994).
6. X. Li *et al.*, Validation of the Hsp70-Bag3 Protein-Protein Interaction as a Potential Therapeutic Target in Cancer. *Mol Cancer Ther*, (Jan 6, 2015).
7. U. P. Andley, P. D. Hamilton, N. Ravi, C. C. Weihl, A knock-in mouse model for the R120G mutation of alphaB-crystallin recapitulates human hereditary myopathy and cataracts. *PLoS ONE* **6**, e17671 (2011).
8. L. T. Chylack, Jr. *et al.*, The Lens Opacities Classification System III. The Longitudinal Study of Cataract Study Group. *Archives of ophthalmology* **111**, 831 (Jun, 1993).

A. Structure of the crystallin domain of cryAB



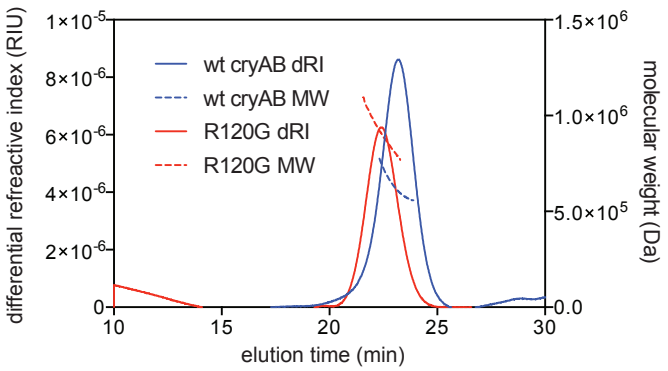
C. Alignment of the crystallin domains

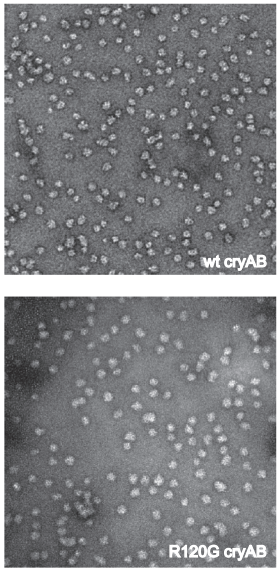
```
Hsp27        EIRHTADRWRVSLDVNHFAPDELTVKT  115
cryAB    54  EMRLEKDRFSVNLDVKHFSPEELKVKV   98
cryAA    63  EVRSDRDKFVIFLDVKHFSPEDLTVKV  105

Hsp27   116  KDGVVEITGKHEERQDEHGYISRCFTRKYTLPPG  151
cryAB    99  LGDVIEVHGKHEERQDEHGFISREFHRKYRIPAD  134
cryAA   106  QDDFVEIHGKHNERQDDHGYISREFHRRYRLPSN  141

Hsp27   152 VDPTQVSSSLSPEGTLTVEAPMP 172
cryAB   135 VDPLTITSSLSSDGVLTVNGPRK 151
cryAA   142 VDQSALSCSLSADGMLTFCGPKI 160
```

B. Characterization of recombinant R120G and wild type cryAB by GPC and EM



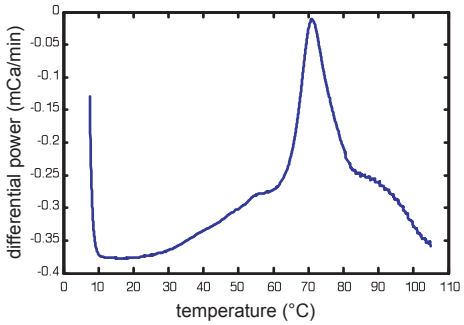


R120G cryAB calculated molecular weight = 950 ± 35 kDa
wt cryAB calculated molecular weight = 570 ± 26 kDa

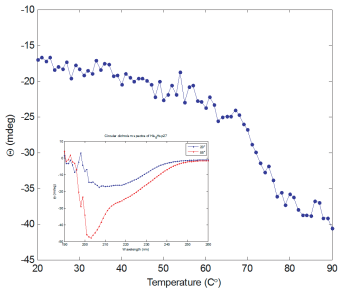
	mean	SEM	n
R120G	946550	35250	2
wt cryAB	573433	25934	3

D. Characterization of recombinant Hsp27 by DSC, CD and gel filtration

i. differential scanning calorimetry



ii. circular dichroism:



iii. gel filtration

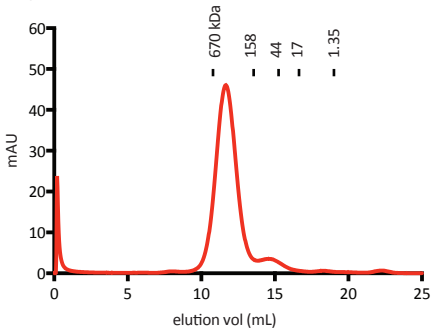
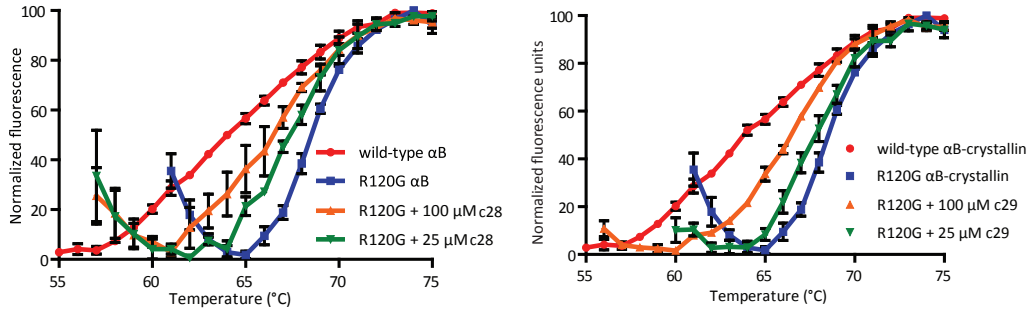
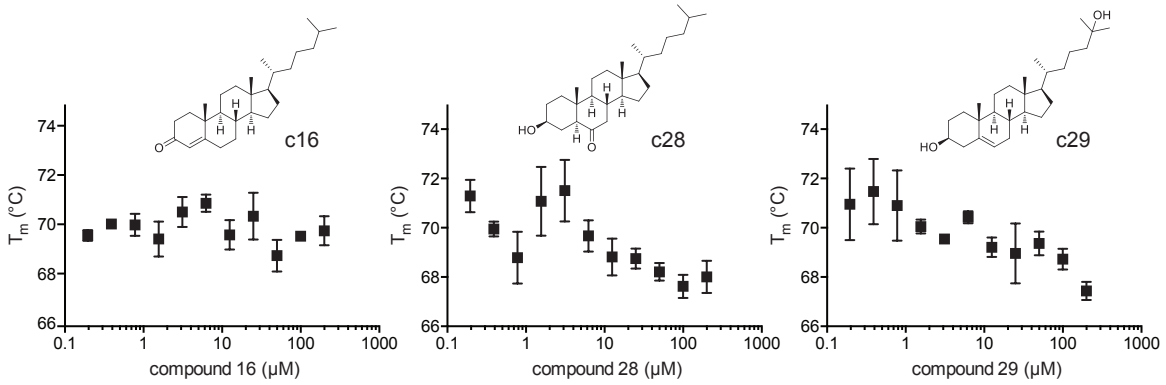


Fig S1. Characterization of cryAB and Hsp27. (A) Structure of the cryAB crystallin domain dimer, highlighting the R120-D109 salt bridge. (B) Characterization of recombinant, purified wild type and R120G cryAB, by GPC and EM. (C) Clustal W alignment of Hsp27, cryAB and R120G cryAB. The cross beta sheet involved in dimer formation is underlined. (D) Characterization of purified human Hsp27 by (i) DSC, (ii) CD and (iii) gel filtration. Both DSC and CD support a melting transition at 72 °C and the gel filtration suggests that purified Hsp27 is an oligomer, consistent with previous reports (Rogalla *et al.* 1999 *J. Biol. Chem.* 274:18947).

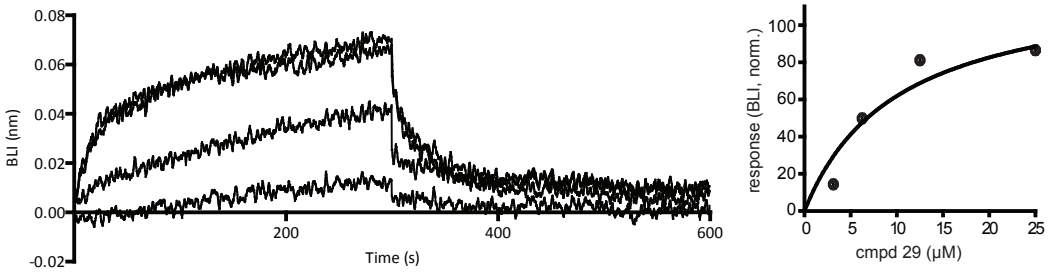
A. Compound **28** and **29** recover Tm in R120G cryAB by DSF



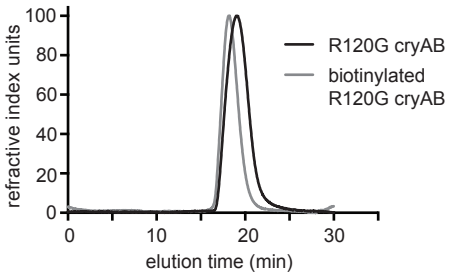
B. Compound **28** and **29** have dose dependent activity against R120G cryAB by DSF



C. Compound **29** binds R120G cryAB by BLI



D. Biotinylation of R120G cryAB does not dramatically change its apparent molecular mass

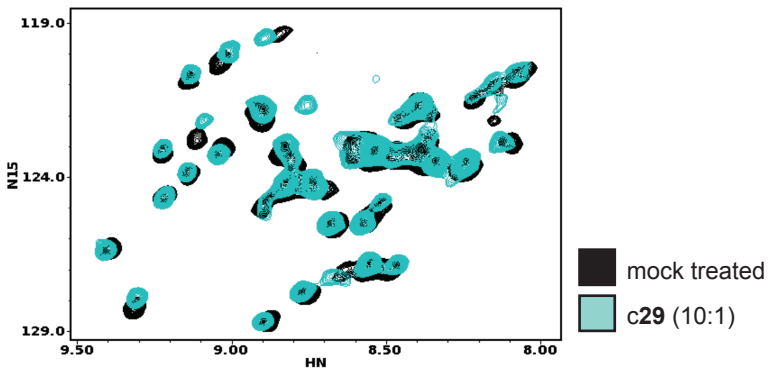


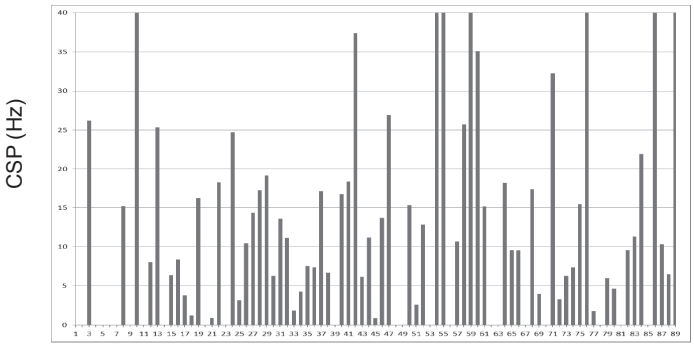
E. Compound **29** partially restores normal oligomeric structure to aggregated R120G cryAB

sample	Mn (g/mol)	approx. # subunits	PDI
wt cryAB + DMSO	616300	27	1.006
R120G cryAB + DMSO	985400	47	1.417
R120G cryAB + c29	843700	37	1.515

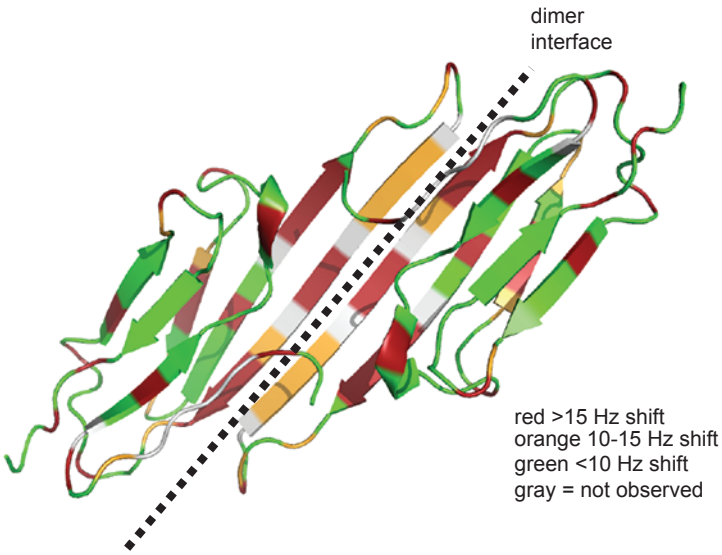
Fig. S2 Sterols bind to R120G cryAB and recover wild type stability. (A) Representative results for R120G cryAB treated with compounds **28** or **29**. Mock treated wild type cryAB is shown for comparison. Both sterols partially recover wild type behavior in the melting curves. Results are the average of duplicate experiments performed in quadruplicate and error bars represent SEM. (B) Compounds **28** and **29**, but not **16**, restoring wild type-like melting transitions in R120G cryAB by DSF. Results are the average of experiments performed twice in quadruplicate and error bars represent SEM. (C) Binding of compound **29** to immobilized R120G cryAB by BLI. Raw plots (left) and analysis of the binding at equilibrium (right) are shown. Results are representative of experiments performed in duplicate. (D) SEC-MALS profiles of recombinant R120G cryAB and biotinylated R120G cryAB; the oligomeric conformation of the protein is largely unchanged. (E) Treatment of cryAB R120G with compound 29 partially restores the normal oligomer distribution, as measured by GPC.

A. Compound **29** binds the crystallin domain of cryAB by NMR





B. Residues sensitive to compound 29 cluster at the dimer interface



C. Compound 29 binds a groove between the cryAB protomers

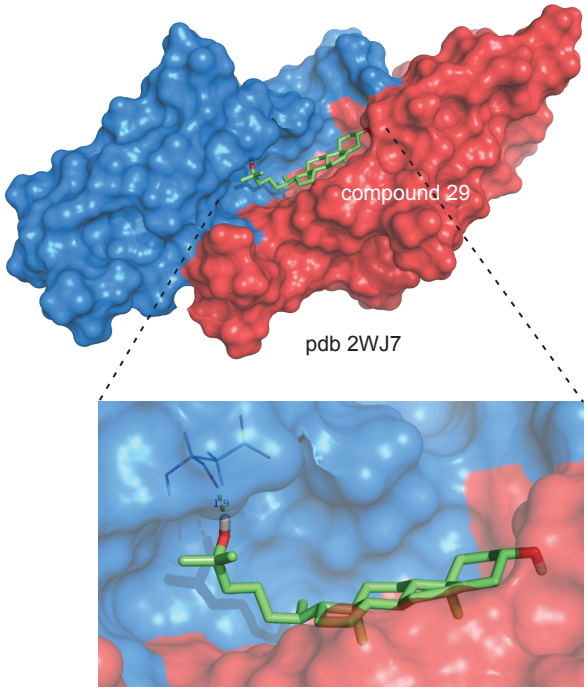


Fig S3. Compound **29** binds the dimer interface of R120G cryAB. (A) HSQC NMR titrations were performed on the ^{15}N labelled crystallin domain of cryAB (100 μM). Residues that underwent a significant chemical shift perturbation (>15 Hz) in the presence of 1 mM compound **29** were noted. CSP: $|\Delta\delta HN + \Delta\delta 15N|$. Compound **16** did not perturb residues with CSP >15 Hz (data not shown). (B) Plotting the sensitive residues revealed a cluster around the dimer interface. (C) Docking of compound **29** into the cryAB core domain, based on the experimental CSPs. A close up of the contact site shows a hydrogen bond with one protomer.

	Tm (DMSO)	Tm (c29)	ΔTm
wt cryAB	63 ± 0.0	63 ± 0.1	0.0
R120G cryAB	68 ± 0.2	67 ± 0.2	-1.7
R120G cryAB (N.P.)	68 ± 0.1	67 ± 0.4	-1.2
wt cryAA	59 ± 0.2	60 ± 1.2	1.0
R49C cryAA	63 ± 0.5	61 ± 0.4	-2.5
F71L cryAA	48 ± 0.0	46 ± 0.3	-2.3
R116C cryAA	70 ± 0.0	68 ± 0.3	-2.0
wt cryGD	72 ± 0.3	72 ± 0.3	0.5
P23T cryGD	69 ± 0.2	70 ± 0.1	0.5
W42R cryGD	65 ± 0.2	66 ± 0.1	0.4

Fig S4. Compound **29** interacts with damaged cryAB and cryAA, but not gamma-crystallin. DSF assay results for c29 (100 μM). Results are the average of at least three experiments performed in triplicate and error is SEM. N.P. = Native Purification; this protein was not denatured and refolded during purification.

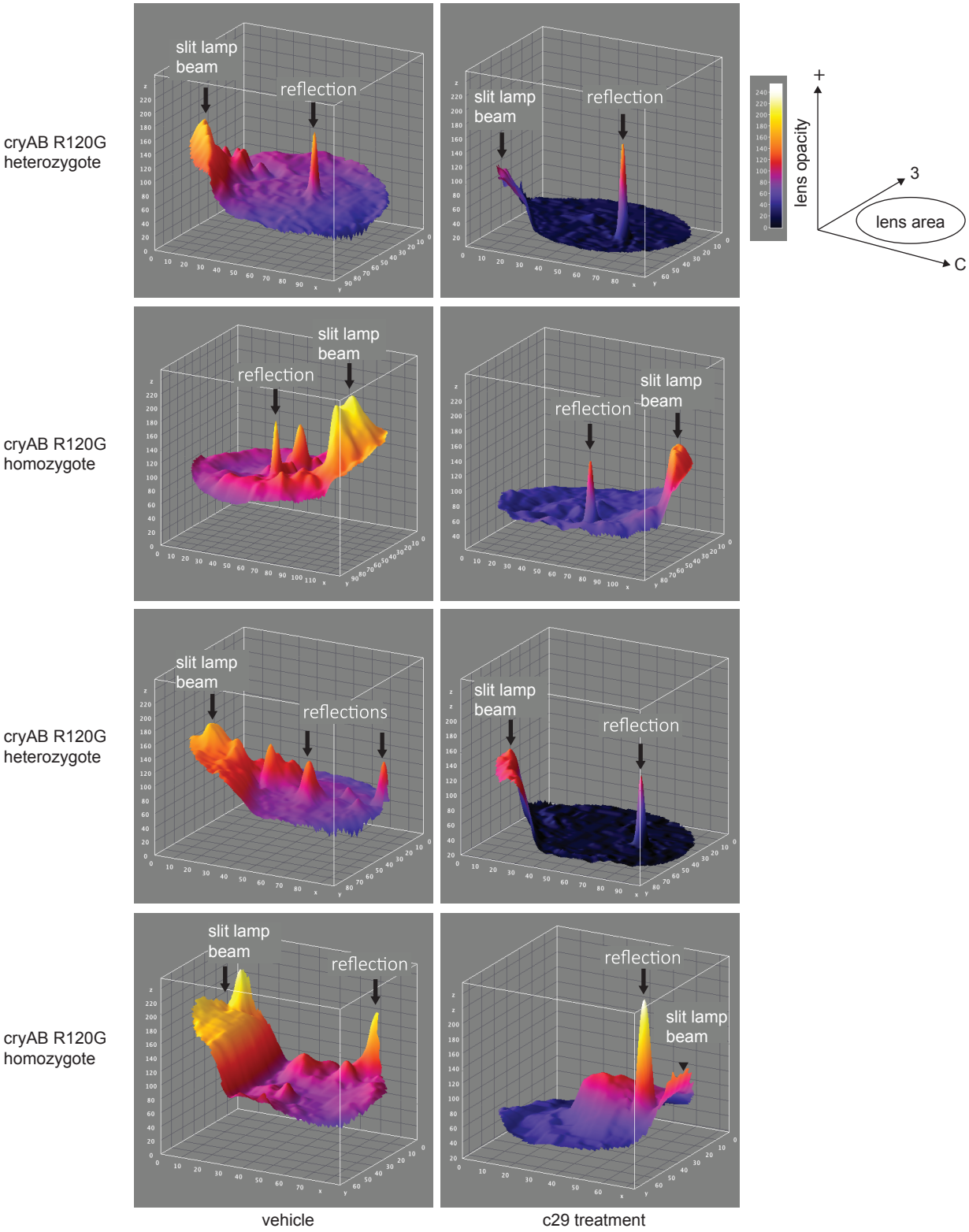
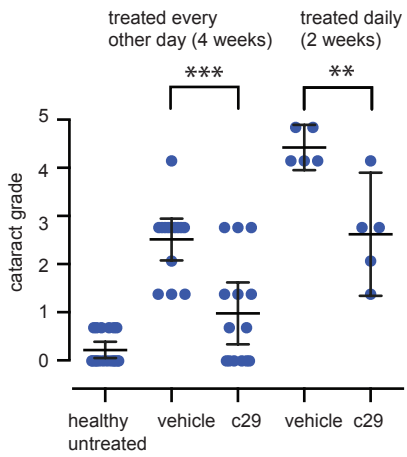
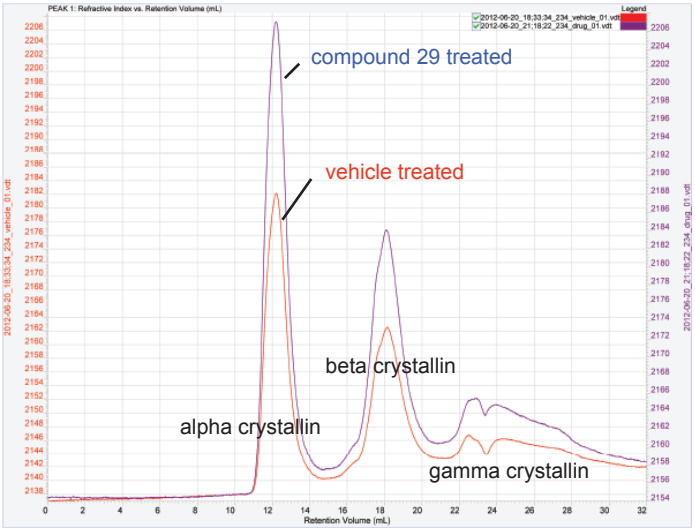


Fig S5: Examples of densitometry analysis of frames taken from slit lamp biomicrosopy videos. The area of the lens is shown on the xy-plane, and the opacity of the lens (as measured by pixel intensity) is plotted on the z-axis according to the heat map scale. In each frame, the locations of the slit lamp beam and the reflection/glare point(s) are indicated with arrows.

A. Compound **29** also recovers transparency in cryAA R49C (+/-) mice, a second model of cataract



B. Compound **29** improves the solubility of crystallins in the mouse lens



C. Compound **29** enhances protein solubility in R120G cryAB mice and is more potent in older mice

Percent increase in solubility

	R120G mouse age (days)						
	66	79	184	192	200	266	All
N	3	1	2	3	4	4	17
Mean	180	98	61	23	39	1400	380
Std Dev	350		2.7	68	26	1800	980
Median	-20	98	61	56	32	680	63

D. Solubility is increased in aged human lens tissue

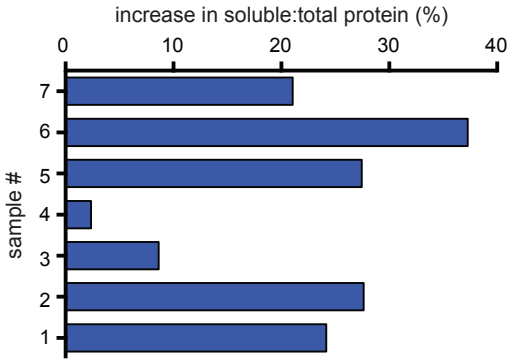


Fig S6. Compound **29** partially recovers the solubility of crystallins in animal models. (A) Cataract grade from biomicroscopy assays on treated cryAA R49C (+/-) mice and wild type controls. Wild type mice aged 60-240 days; cryAA R49C (+/-) aged 48-181 days, treated every other day for 4 weeks (1 mM **29**); cryAA R49C (+/-) aged 77 days treated daily for 2 weeks. (B) The solubility of crystallins is significantly improved in the treated versus untreated eyes. Lens tissue from treated R120G (+/-) mice was centrifuged to remove insoluble material and the crystallins separated by GPC. (C) Quantification of the GPC experiments, showing that compound **29** significantly increased the solubility of the crystallin proteins (Wilcoxon signed rank p-value = 0.001). In addition, older animals (>200 days) tended to have higher percent improvements (Spearman rank correlation = 0.44, p=0.075). (C) (D) Compound **29** improves protein solubility in human lens samples. Individual lens samples were quartered and treated with compound **29** (5 mM) or a mock vehicle control for 6 days in the dark. These samples were homogenized, centrifuged and the amount of protein in the soluble and insoluble fractions measured by BCA assays in triplicate. The ratio of soluble vs total protein was then calculated. Protein solubility was improved in each sample. See the Supporting Information for additional details.