

The functional significance of high cysteine content in eye lens γ -crystallins

Eugene Serebryany ^{1,2*}, Rachel W. Martin ^{3,4*} and Gemma R. Takahashi ⁴

¹ Department of Physiology & Biophysics, Stony Brook University, SUNY, Stony Brook NY 11794

² Laufer Center for Physical & Quantitative Biology, Stony Brook University, SUNY, Stony Brook NY 11794

³ Department of Chemistry, UCI Irvine; Irvine CA 92697-2025

⁴ Department of Molecular Biology & Biochemistry, UCI Irvine; Irvine CA 92697-3900

* Correspondence: ES: eugene.serebryany@stonybrook.edu; RWM: rwmartin@uci.edu

Abstract: Cataract disease is strongly associated with progressively accumulating oxidative damage to the extremely long-lived crystallin proteins of the lens. Cysteine oxidation affects crystallin folding, interactions, and light-scattering aggregation especially strongly due to formation of disulfide bridges. Minimizing crystallin aggregation is crucial for lifelong lens transparency, so one might expect the ubiquitous lens crystallin superfamilies (α and $\beta\gamma$) to contain little cysteine. Yet, the Cys content of γ -crystallins is well above the average for human proteins. We review literature relevant to this longstanding puzzle and take advantage of expanding genomic databases and improved machine learning tools for protein structure prediction to investigate it further. We observe remarkably low Cys conservation in the $\beta\gamma$ -crystallin superfamily; however, in γ -crystallin the spatial positioning of Cys residues is clearly fine-tuned by evolution. We propose that the requirements of long-term lens transparency and high lens optical power impose competing evolutionary pressures on lens $\beta\gamma$ -crystallins, leading to distinct adaptations: high Cys content in γ -crystallins but low in $\beta\beta$ -crystallins. Aquatic species need more powerful lenses than terrestrial ones, which explains the high methionine content of many fish γ - (and even β -) crystallins. Finally, we discuss synergies between sulfur-containing and aromatic residues in crystallins and suggest future experimental directions.

Keywords: eye lens, cataract, crystallin, cysteine, methionine, disulfide, protein misfolding, protein aggregation, refractive index, protein evolution

Citation: To be added by editorial staff during production.

Academic Editor: Firstname Last-name

Received: date

Revised: date

Accepted: date

Published: date



Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Cysteine is the rarest canonical amino acid in proteins; its frequency increases with organismal complexity, from about 0.5–1% in prokaryotes to 2% in mammals, all of which are low compared to the expected frequency of about 3.3% for an amino acid with two codons [1, 2]. Thus, it appears that cysteine is frequently selected against by evolution, potentially because of its reactivity. It is uniquely capable of readily reversible redox reactions: the formation and breakage of disulfide bonds, along with several other types of oxidative modifications. It is a plausible hypothesis that the importance and sensitivity of redox chemistry in an organism's life cycle plays an increasingly important role roughly in proportion to the percentage of Cys residues in its proteome. For example, *E. coli* can tolerate wide variations in its cytoplasmic redox potential – a feature that has been harnessed to great advantage for the heterologous expression of disulfide-containing proteins [3, 4]. By contrast, alteration of the redox potential setpoint in human cells, in either direction, are

strongly associated with the development of cancer [5-7]. In general, accumulating evidence links proteomic redox balance to the aging process [8-10].

Cataract, caused by light-scattering aggregation of natively highly soluble eye lens proteins, crystallins, is one of the most common diseases of aging, afflicting the majority of Americans over 65 (self-reported diagnoses [11]) and accounting for 50% of all cases of blindness in low- and middle-income countries [12]. It is a striking example of the nexus of molecular biophysics and biochemistry with organismal aging. Most cataract occurs in the central (or nuclear) region of the eye lens, but peripheral (cortical) cataract is also common [13]. The proteome of terminally differentiated lens fiber cells is dominated by highly concentrated α -, β -, and γ -crystallins – soluble monomeric or oligomeric proteins that are never replaced [14, 15]. There are two superfamilies [16] of lens crystallins in humans: (1) the α -crystallins are small heat-shock proteins that act as molecular chaperones [17, 18], and (2) the β - and γ -crystallins, which are the primary structural and refractive proteins [19]. The α and $\beta\gamma$ crystallin families are present in all vertebrates; additional taxon-specific lens crystallin families exist, such as the δ -crystallins of many avians and reptiles [20, 21]. Our focus here is on the $\beta\gamma$ -crystallins, which share a common double-Greek key fold and a two-domain structure held together by an interdomain interface [22-24], and on the distinctions between the β - and γ -crystallin families. The β -crystallins are found in domain-swapped [25] or face-en-face dimers [26] and can form higher-order oligomers [27] in contrast to the monomeric γ -crystallins [28]. In addition to these crystallins, which are common to all vertebrates, many organisms (although not humans) also have taxon-specific crystallins recruited from diverse small enzyme families [29].

The cytoplasmic redox potential of the extremely long-lived lens fiber cells drifts predictably with age toward increasingly oxidizing values, leading to formation of many disulfide bonds in the crystallin proteome [30-35]. The severity of age-onset cataract depends on the extent of crystallin disulfide bonding and on which specific Cys residues are involved [31, 36, 37]. *In vitro* experiments by ourselves and others have demonstrated clear evidence that γ -crystallins, in particular, can act as oxidoreductases, exchanging disulfides among themselves, forming a de-facto protein-based redox buffer to compensate for the depletion of the glutathione redox buffer in the aging lens [38, 39].

While some disulfide bonds can form in native or native-like conformations of the γ -crystallins, others trap non-native, aggregation-prone conformations [38, 40]. Exposure of buried Cys residues can also create new binding sites for transition metals, which accumulate in the lens with age and particularly with cataractogenesis [41-45]. In turn, there is evidence of extensive regulation of the reactivity of Cys residues in the aging lens, ranging from partial methylation of exposed Cys residues [46] to formation of irreversible oxidation products at “sacrificial” Cys residues [47]. Moreover, the lens metabolome appears to have evolved to suppress the conformational dynamics of the γ -crystallins that would otherwise lead to the formation of non-native disulfide bonds favoring misfolded, aggregation-prone γ -crystallin conformations [40].

In light of the evolution of the $\beta\gamma$ -crystallin family, however, all the above observations present a surprising paradox. The γ -crystallins and β -crystallins are closely related both evolutionarily and structurally, and they coexist in the cytoplasm of the same lens fiber cells at high concentrations. Yet, neither the number, nor the positions of Cys residues is conserved within this protein family: as shown in **Figure S1**, the most conserved residue types are Gly, Ser, and aromatics – all crucial for the highly compact double-Greek key fold – but not Cys. Although γ -crystallins are universally cysteine-rich, β A-crystallins have relatively typical content for proteins of their length, while β B-crystallins are cysteine-poor (see **Table 1** and **Table S1**). As a result, multiple sequence alignment using the DeepMSA2 algorithm [48, 49] shows quite low conservation of Cys residues in the $\beta\gamma$ -

crystallins overall, compared to the high conservation of Ser, Gly, and aromatic residues important for proper folding of the topologically complex Greek key motifs (**Figure 1**). Given the highly deleterious consequences of disulfide-trapped misfolding and subsequent light-scattering aggregation leading to cataract, why did the γ -crystallins not evolve to be largely or entirely Cys-free, as the β B-crystallins have done? We propose two potential hypotheses: (1) Cys is selected for because it increases refractivity, even at the cost of deleterious aggregation later in life; and (2) The spatial positioning of Cys residues in the crystallin structure is more important than the total number of Cys residues for maintaining the monomeric state of γ -crystallins.

A common misconception regarding diseases of aging is that human life expectancy in the era before modern medicine was much shorter than today and did not extend much past peak reproductive years; therefore, evolution has not had a chance to “solve” protein misfolding diseases or diseases of aging in general. This persistent belief relies on incorrect assumptions. First, average life expectancy in premodern societies is an irrelevant metric of adult human lifespan because the average is heavily skewed by high infant and child mortality. It is well-documented, both in ancient sources and by observations of modern hunter-gatherer societies and societies with little or no modern medical care, that humans who survived childhood and had proper nutrition frequently lived to their 70s and even beyond [50, 51]. Second, grandparents contribute substantially to the evolutionary fitness of their grandchildren [52, 53]. For these reasons, our species has likely faced strong evolutionary pressure to maintain clarity of vision well into old age. In this way, cataract disease is akin to, e.g., Huntington’s disease: the protein misfolding is nearly inevitable from the physico-chemical standpoint, but the proteins and their environment have both been tuned by evolution to delay such an outcome for a lifetime [54]. Furthermore, other long-lived vertebrates are reproductively active until late in life and lack the human social structures that can allow for survival of blind individuals, so we should not consider old-age lens transparency as beyond the reach of evolution. As pointed out already by W. D. Hamilton, evolution tends to select for early-life fitness at the cost of late-life senescence, yet human evolution has clearly selected for a long post-reproductive lifespan, too [55].

Mere genetic drift clearly does not provide the explanation, either. Evolutionary pressure on γ -crystallin sequences has been high, as evidenced by their highly unusual amino acid composition. For example, the ratio of Lys to Arg residues in human proteins is typically ~1:1 [56], but this ratio within the abundant human γ -crystallins (γ C, γ D, and γ S) it is ~1:4. Meanwhile, the ratio of Ala to Cys in human proteins is ~3:1 [56], but in the γ -crystallins it is ~2:3. Fish γ -crystallins have even more extreme compositions; several crystallins from the Antarctic toothfish (*Dissostichus mawsoni*) contain no alanine and are instead highly enriched in highly polarizable residues, including Phe, Tyr, Arg, and Met [57]. This leads to unusual biophysical properties, including differential resistance to thermal and chemical denaturation [58], as well as resistance to cold cataract far below temperatures encountered by the fish [59]. One likely biophysical explanation for this lies with the differential propensity of Lys-enriched vs. Arg-enriched γ -crystallin variants to undergo liquid-liquid phase separation at the very high (100s of mg/ml) protein concentrations involved [59, 60].

Why are Cys residues within the eye lens proteome so concentrated in the γ -crystallins and exiguous in many of the closely related β -crystallins? Here we discuss the γ -crystallins’ high Cys content in relation to two major evolutionary requirements for eye lens function: (1) high long-term transparency, i.e., minimization of light-scattering aggregation, which is achieved via thermodynamic and kinetic stability, plus oxidoreductase activity; and (2) high optical power, i.e., maximization of the refractive index increment of

the protein to facilitate focusing of light, which is achieved via highly unusual amino acid compositions, plus tertiary structural features. We use multiple sequence alignments and phylogenetic clustering at both the amino acid and the nucleotide levels to draw evolutionary relationships among lens crystallins. We use the most successful current protein structure prediction algorithm, D-I-TASSER [61], to compile a dataset of predicted γ -crystallin structures from phylogenetically representative species. This combination of methods reveals potential evolutionary signatures of thiol/disulfide chemistry in the eye lens. We hypothesize that the requirements of high optical power (refractive index increment) and high long-term transparency (aggregation avoidance) are sometimes in tension with each other and may have pushed the β and γ families of lens crystallin to opposite extremes of Cys content.

2. Molecular etiology of cataract and the paradox of high Cys content

2.1. Thermodynamic and kinetic stability of $\beta\gamma$ -crystallins

Human γ -crystallins are among the most thermodynamically stable in the body [62]. The most stable of these is γ D-crystallin: a well-structured, monomeric, two-domain, double-Greek-key, predominantly β -sheet protein that cannot be unfolded by a saturated solution of urea at neutral pH [63]. This protein is also thermostable up to $\sim 80^\circ\text{C}$ [37, 64]. Other human γ -crystallins are also highly thermodynamically stable proteins, and they also show exceptional kinetic stability, with unfolding half-lives of minutes to hours even in 4 M guanidinium [65, 66]. Notably, γ -crystallins from mice appear to be less thermodynamically stable than their human counterparts [67]. How crystallin stability varies among animal species has not yet been systematically investigated.

By contrast, the β -crystallins, which share the same fold and high sequence identity to the γ -crystallins, are more thermodynamically labile. Although they are still highly stable compared to the average human protein, members of this family are urea-denaturable [68, 69] and have melting temperatures some $10\text{--}20^\circ\text{C}$ below the γ -crystallins [70]. Unlike the monomeric γ -crystallins, the β -crystallins are also known to form a variety of oligomeric complexes [68, 71]. Nevertheless, β -crystallins are more highly abundant than γ -crystallins in the human lens [72]. One important reason may be their lower redox sensitivity: as noted in the Introduction above, cysteine residues are the Achilles' heel of γ -crystallins; although they are likely beneficial under the reducing conditions of young lens cytoplasm, later in life they acquire disulfide bonds that distort the native structure and promote aberrant oligomerization in the more oxidizing cytoplasm of aged lenses [32, 36, 73]. The β B-crystallins, being exiguous in Cys, are accordingly less susceptible to the effects of changing thiol chemistry with age.

2.2 Aggregation propensity of γ -crystallins

Cataract is increased turbidity (visible light scattering) of the lens [14, 74]. It is typically a disease of aging, although other factors, such as eye trauma, can promote cataract development [75, 76]. Once the size of protein aggregates becomes comparable to the wavelength of visible light, these particles scatter increasing amounts of visible light photons before they can reach the retina. The result is the blurring of vision and a red-shift in the perceived color palette [77], because the shorter blue wavelengths of light are scattered more strongly by small aggregate particles than the longer red wavelengths. However, due to the enormous challenge of *in vivo* structural biology, particularly of protein aggregates, there remains some controversy over the exact nature of protein aggregation during

typical age-onset cataract disease, as discussed below. We will briefly summarize available evidence from *in vitro*, *in vivo*, and *ex vivo* studies, although a comprehensive overview of the crystallin aggregation literature is beyond the scope of the present focused review.

Aggregation of γ -crystallins has been studied *in vitro* for a half-century but continues to reveal new and surprising biophysical phenomena. Lens crystallins natively exist at high concentrations and have clearly evolved to push the biophysical envelope on solubility, stability, and a plethora of fascinating mechanisms of aggregation-avoidance. Yet, they are not invulnerable. In fact, *in vitro* biophysical and biochemical research on γ -crystallins has revealed every conceivable mode of aggregation in these highly natively stable proteins. Rare cataract-associated variants of γ -crystallins have been shown to aggregate via crystallization [78], amyloid formation [79], amorphous aggregation without major conformational change [80], and redox-driven aggregation from partially unfolded states via a domain-swap-like mechanism [37, 81–83]. They can also form liquid droplets in their folded states, thus undergoing liquid-liquid phase separation [59, 60], or form aggregates wherein the monomers are either bridged or misfolded (or both) by transition metal cations [84–86]. Predominantly amorphous (and some amyloid) aggregation can be induced by exposure to UV light [47, 87, 88], which is a physiologically relevant stressor for the lenses of terrestrial animals.

In-vivo and *ex-vivo* studies of crystallin aggregation likewise have a long history, and it has been a tremendous yet motivating challenge to extend the tools of structural biology and molecular biophysics from the test-tube to the unique (and uniquely accessible) environment of the lens. For example, the transparency of the lens has enabled pioneering work on *in vivo* dynamic light scattering for improved diagnostics of cataract, and potentially even Alzheimer's, progression [89–91], as well as the first *in vivo* application of two-dimensional infrared spectroscopy (2D-IR) that is highly attuned to detecting amyloid-like β -structures [92], while electron microscopy of lens slices revealed increased cytoplasmic texturing that could account for increased light scattering without fibrils [93]. The phenomenon of cold cataract was the first to be extensively researched, and indeed temperature-driven "cold precipitation" (later understood to arise via protein condensate formation) in the lens crystallins was reported already in 1964 [94]. Crystals extracted directly from the lenses of patients carrying the rare point mutations of human γ D-crystallin that lead to crystal cataracts yielded some of the highest-resolution protein crystal structures at the time [78]. Other research has emphasized that proteomes evolve at the edge of insolubility [95, 96], and broad insolubilization may sometimes occur in the lens [97].

However, crystallin aggregates extracted from age-onset cataract lenses were neither crystalline nor temperature-reversible; an early demonstration that the combination of strong denaturant and reducing agent could disassemble these aggregates established age-onset cataract as a protein misfolding disease with an important component of redox chemistry [31]. Since then, proteomic studies have revealed a large and growing number of post-translational modifications in lens proteins as a function of age, including deamidation; Asp isomerization, oxidation of Met, Cys, and Trp; Cys methylation; Lys succinylation; and much more [46, 98–100]. Notably, the formation of a diffusion barrier roughly coincident with the boundary between the lens nucleus and the lens cortex has been observed, and it correlates with the drift in the redox potential of lens fiber cell cytoplasm to increasingly oxidizing levels beginning in midlife [34, 101–103]. For the Cys-rich γ -crystallins, but also for those α - and β -crystallins that contain at least two Cys residues, formation of intramolecular disulfide bonds correlates strongly with cataract progression [36]. Intermolecular disulfides have also been observed in the lens [73], and both intra- and intermolecular disulfides have been shown to promote aggregation [37, 38, 40, 81, 104, 105] or inhibit chaperone function [106] *in vitro*. Sustained research efforts have therefore focused

on the chemistry of crystallin Cys residues as an important contributor to the etiology and progression of age-onset cataract [32, 107–109].

3. Evolutionary and structural analysis of chordate $\beta\gamma$ -crystallins

To better understand the evolutionary origin and importance of high Cys content in γ -crystallins, we compared γ -crystallins from representative chordates to one another and, as a group, to β A- and β B-crystallins from the same organisms. For each organism, we selected every protein annotated in UniProt as a γ -crystallin (or as β A- and β B-crystallin, respectively) and containing either two (for the single-domain $\beta\gamma$ -crystallin from *Ciona intestinalis*) or four (all other organisms) complete Greek key domains. A small number of truncated crystallins were excluded, as were some homologs of absent-in-melanoma-1 (AIM-1), which has several $\beta\gamma$ -crystallin domains but is not a lens protein [110].

Evolutionary comparisons allow us to test several hypotheses as to why lens γ -crystallins are cysteine-rich. First, if Cys residues at specific positions are important in stabilizing the double-Greek key domains, then those Cys positions should be conserved among these and homologous proteins. Second, if sulfur-containing residues are favored in lens crystallins due to their high polarizability and consequently high refractive index, then the lens crystallins of aquatic animals should be especially Cys-rich (as well as Met-rich) compared to land-dwelling animals who are able to derive significant optical power from the cornea. Third, if avoidance of disulfide-driven aggregation or changes in short-range packing is a major evolutionary pressure, then the natively monomeric γ -crystallins should have no more than one solvent-exposed Cys-rich site per molecule: enough to facilitate redox chemistry but not enough to undergo disulfide-driven “daisy-chain” type aggregation from the native state. All these hypotheses can be tested by a combination of evolutionary and structural analysis that is now possible thanks to the rapid growth in available genomic data on the one hand and machine-learning driven protein structural prediction methods on the other. We note, however, that these hypotheses will ultimately require direct experimental testing, some of which is already underway in our labs.

3.1 Lack of Cys conservation in $\beta\gamma$ -crystallins

We used the DeepMSA2 algorithm [49] to generate residue conservation logos across the sequence, for four representative crystallins: the human γ D- and γ S-crystallins, a fish γ M-crystallin, and an ancestral-like $\beta\gamma$ -crystallin from tunicates (**Supplementary Figure S1**). By default, this tool measures residue conservation across all protein-coding sequences in metagenomic databases that can be successfully aligned to the sequence of interest (typically thousands or tens of thousands of sequences). Considering that γ -crystallins, and even many β -crystallins, are enriched in cysteine, the result is surprising: there is very low conservation of Cys residues. Instead, Ser, Gly, and aromatic residues, all of which are likely critical for the proper folding and stability of crystallin domains, are by far the most conserved residues. Several charged residues are also conserved. While this result may be skewed by high abundance of microbial $\beta\gamma$ -crystallins in the metagenomic databases, it indicates at least that crystallin domains generally do not require Cys residues.

Why, then, are Cys residues so abundant in the γ -crystallins? To gain more fine-grained understanding of this question, we carried out phylogenetic clustering of γ -crystallin sequences and structural analysis of Cys and Met positions and solvent exposure. Here we primarily focus on γ -crystallins because these are the most enriched in Cys residues, and also the most refractive proteins, in the mammalian lens [111]; the abundance of Cys and Met in β -crystallins is provided for comparison in **Supplementary Table S1**.

3.2 Phylogenetic clustering of γ -crystallins

The protein set chosen for this discussion consists of all γ -crystallin isoforms reported in UniProt[112] for the following vertebrates: *Petromyzon marinus* (sea lamprey), *Latimeria chalumnae* (coelacanth), *Chiloscyllium indicum* (slender bamboo shark), *Danio rerio* (zebrafish), *Xenopus laevis* (African clawed frog), *Ornithorhynchus anatinus* (platypus), *Macropus fuliginosus* (western gray kangaroo), *Mus musculus* (mouse), *Bos taurus* (cow), *Homo sapiens* (human), *Anolis carolinensis* (green anole), *Alligator mississippiensis* (American alligator), and *Gallus gallus* (chicken). Additional $\beta\gamma$ -crystallins from the chordates *Eptatretus burgeri* (inshore hagfish), *Ciona intestinalis* (tunicate) and *Branchiostoma floridae* (lancelet) were included for comparison. The same organisms' β -crystallin sequences were likewise retrieved from UniProt for comparison.

Protein sequences from the crystallins were aligned using Clustal Omega v1.2.4 [113] and custom scripts (github.com/grtakaha/protein_alignment_tool) written in and tested on Python v3.7.4 [114]. Sequence alignments for the γ - and β -crystallin are shown in **Supplementary Figures S2-S5** and **S7-S10**, respectively. Conserved residues from large numbers of related sequences were identified using Deep MSA [48, 49]. Residue conservation was visualized using WebLogo [115].

Nucleotide sequences (exons only) were retrieved using the following methods depending on availability of information in UniProt [112].

Method 1: extracted using exon coordinates and an NCBI nucleotide ID (Nucleotide database) [116] from the "Genomic Coordinates" tab of the given UniProt entry.

Method 2: identified within an NCBI reference sequence entry (Nucleotide database)[116] from the "Sequence databases" section of the main page of the given UniProt entry.

Method 3: retrieved from an Ensembl [117] ID (gene or transcript), found in the "Genome annotation databases" section of the main page of the given UniProt entry; only "cds:protein coding" sequences were used.

Method 4: retrieved from a "PROTEIN SEQUENCE" EMBL [118] link to the European Nucleotide Archive (ENA) [119] found in the "Sequence databases" section of the main page of the given UniProt entry.

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

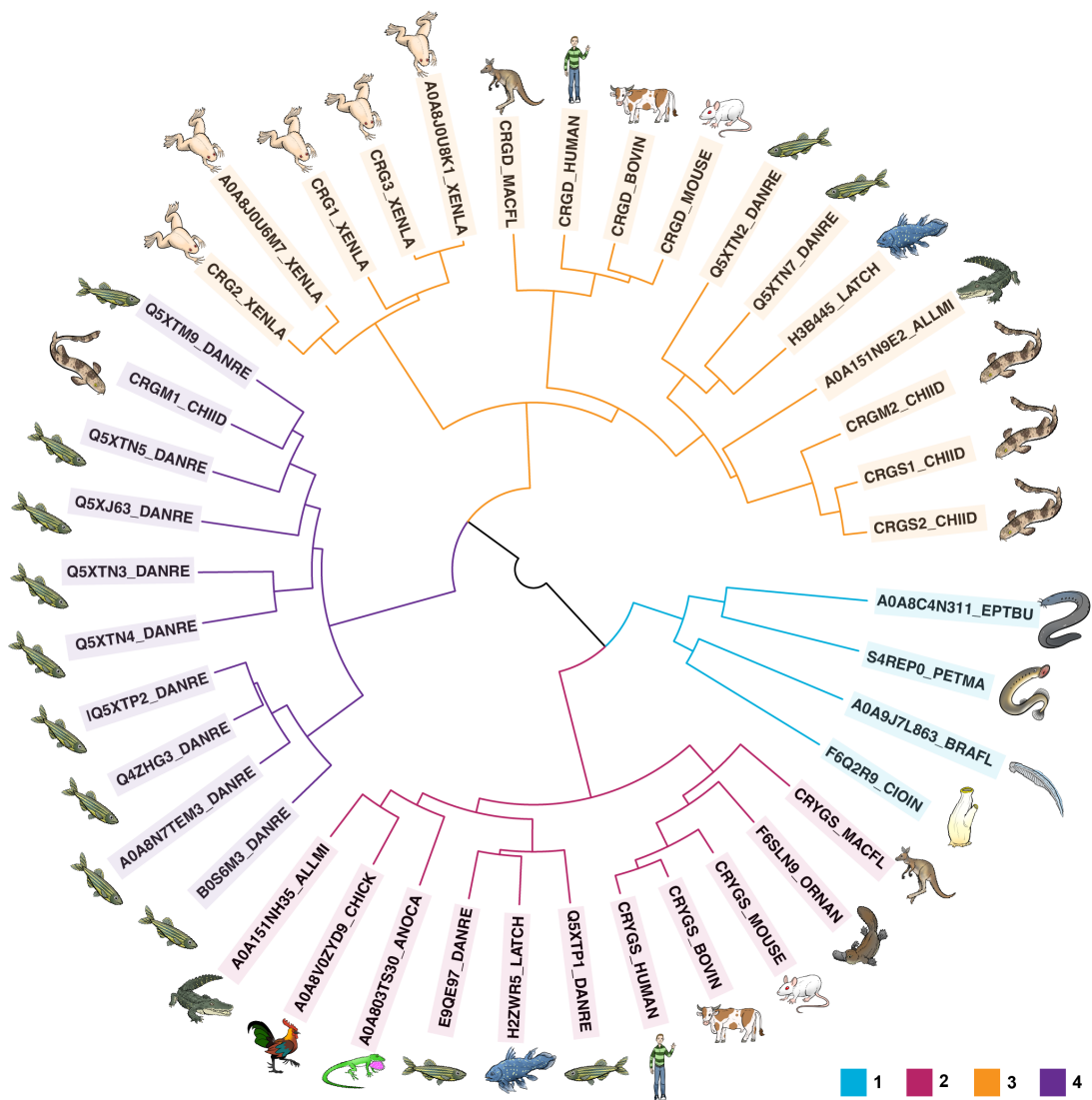


Figure 1. Protein sequence similarity among $\beta\gamma$ - and γ -crystallins. Crystallin sequences from chordeates (tunicate, lancelet, hagfish) cluster with that of the sea lamprey, a representative of the most basal vertebrates (Cluster 1). The next major group of crystallins to split (Cluster 2) include γ S-crystallins from many organisms. From there, two more large groups of γ -crystallins split off. Group 3 includes mammalian γ D-crystallins (kangaroo, human, cow, mouse). as well as other γ -crystallins found in lobe-finned fish (coelacanth), chondrichthyans (bamboo shark), bony fish (zebrafish), crocodilians (alligator), and amphibians (clawed frog). Finally, group 4 comprises fish-specific γ M-crystallins from the zebrafish and bamboo shark.

A multiple sequence alignment of all γ -crystallin protein sequences was further generated using Clustal Omega [113], and the results were used to generate a tree based on protein sequence identity. Specifically, the percent identity matrix from this alignment was used to make a dissimilarity matrix (100 - percent identity), which was then used as input for hierarchical clustering via the ward.D2 method of the hclust function in the base R v4.1.1 Stats package [120]. Clusters were visualized in a circular dendrogram using circize v0.4.15 [121] and dendextend v1.17.1 [122] (Figure 4). This type of tree does not show any information about evolutionary relationships: it does not account for synonymous mutations at the nucleic acid level or convergent evolution toward similar sequences. The

crystallin sequences cluster into four groups. The most basal cluster (Cluster 1) contains crystallins from the tunicate, lancelet, and hagfish, consistent with the positioning of these organisms near the base of the chordate/vertebrate family tree. The next group to split (Cluster 2) contains γ S-crystallins, which are found in most vertebrates, including hagfish, bony fish, lobe-finned fish, mammals, reptiles, and birds. From there, the remaining crystallins are divided into two groups, one containing mammalian γ D-crystallins as well as similar proteins from other organisms (Cluster 3), and the other (Cluster 4) containing fish γ M-crystallins. In this set, most of the γ M-crystallins come from the zebrafish, *Danio rerio*, with one sequence belonging to the slender bamboo shark.

For the γ -crystallins listed in **Figure 5**, nucleotide sequences were aligned by codon using MUSCLE [123] through the MEGA X interface [124] (Gap Open=−2.90, Gap Extend=0.00, Hydrophobicity Multiplier=1.20, Max Iterations=16, Genetic Code=Standard, Cluster Methods=UPGMA, Min Diag Length (Lambda)=24). This alignment was used to generate a tree (protein coding option) in MEGA [124] using the Neighbor-Joining method [125] with a bootstrap test [126] of 1000 replicates. Distances (base substitutions per site) were calculated using the Maximum Composite Likelihood method [127]. Ambiguous comparisons were handled by pairwise deletion.

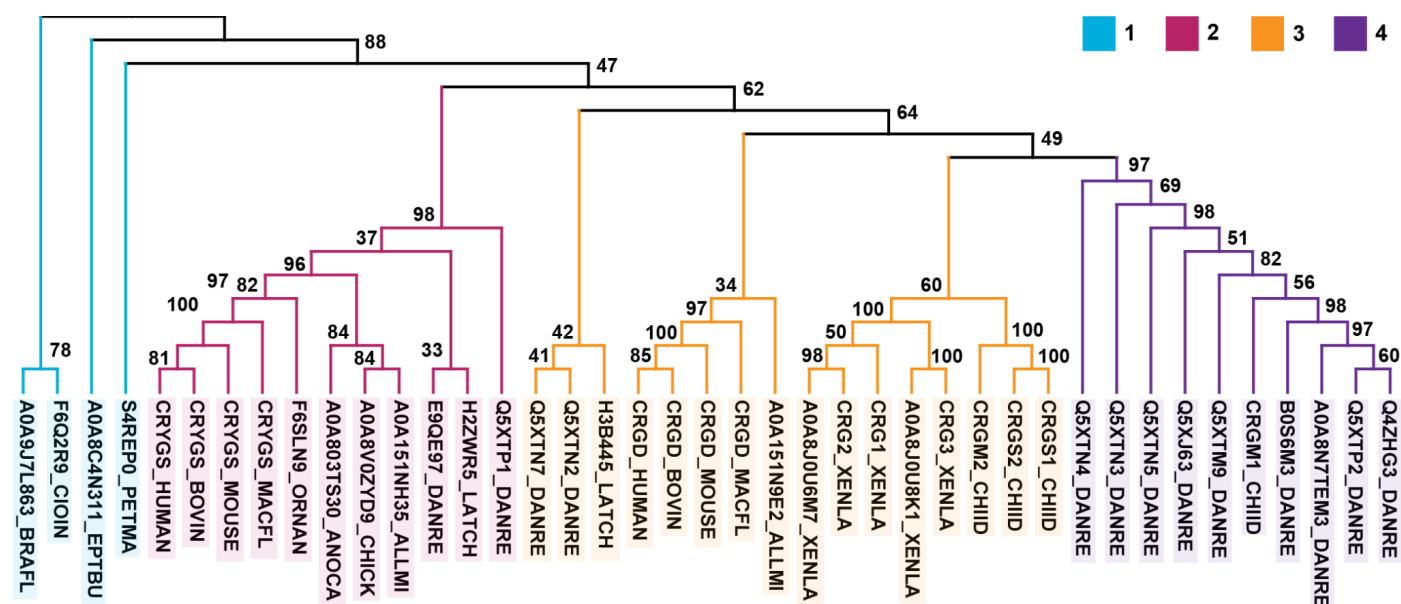


Figure 2. Phylogenetic tree for β - γ and γ -crystallins. Nucleic acid sequences coding for crystallin proteins from the tunicate and lancelet are highly similar. The labels are color-coded according to the protein sequence clusters described in Figure 4.

Surprisingly, many proteins that were previously annotated as γ M or γ S cluster with the γ D-crystallins, including four proteins from the zebrafish and three from the slender bamboo shark. The zebrafish proteins most strongly resemble orthologs from either the coelacanth or the alligator. The mammalian γ D-crystallins are strongly similar to one another but are otherwise unremarkable with respect to the other proteins in Cluster 4; they are located in the middle of the larger cluster. Fish are a polyphyletic group, with three distinct lineages, the chondrichthians (including sharks and rays), the teleosts such as the zebrafish, and the lobe-finned fish like the coelacanth, which are more closely related to the tetrapods [128, 129]. The *Xenopus* γ -crystallins cluster with this group as well, raising the possibility that the hypothesis of Bloemendal et al. that these proteins arose from a distinct gene duplication relative to γ S may also apply to other γ D-like crystallins [23].

A phylogenetic tree based on the nucleic acid sequences coding for these same proteins is shown in **Figure 5**. At the nucleic acid level, these sequences do not cluster as

straightforwardly. The tunicate and lancelet sequences are relatively similar to each other, but well-separated from the rest of the sequences. Although the exons are highly similar, Kappé et al. have reported that the intron placement is quite different [130]. In the *Ciona intestinalis* protein, as in vertebrate γ -crystallins, the introns are located between the regions coding for the Greek key motifs, whereas the introns are found within these motifs in the lancelet protein [130]. The latter also lacks the short N-terminal extension encoded by a separate exon in the tunicate protein and vertebrate γ S-crystallins. The hagfish and lamprey crystallins are well-separated from each other and from the next major cluster, which contains the γ S-crystallins. The sequences found in Cluster 3 of the protein tree are found in three subclusters, containing γ M-crystallins from the zebrafish and coelacanth, the mammalian and alligator γ D-crystallins, and γ -crystallins from the African clawed frog and slender bamboo shark, respectively. Despite their annotations as γ S1, γ S2, and γ M1, all three shark crystallins are highly similar to each other, and all have a methionine content of 9–10%. The presence of these three subclusters, as well as the groupings of crystallins from highly divergent organisms (e.g., zebrafish and coelacanth, frog and shark) and the inclusion of the alligator sequence with the mammalian γ D-crystallins may suggest that γ -crystallin gene loss has commonly occurred during crystallin evolution. Finally, the rest of the γ M-crystallin sequences are related by a highly chained structure.

The most striking observation from the phylogenetic clustering is that Cys positions seem to be a stronger predictor of which γ -crystallin partitions to which cluster than even the proteins' current classification in the proteomic databases. Crystallins containing the "DCDCDC" loop (or its close homologs) found in human γ S-crystallin clustered together in Cluster 2. By contrast, even proteins that are currently classified as γ S-crystallins but that lack this signature sequence partitioned to Cluster 3 along with γ D-crystallins, as shown in **Figures S2–S5**. Thus, even though there is very low overall conservation of Cys within the $\beta\gamma$ -crystallin superfamily, specific Cys-containing motifs appear to be linked to functional specialization within the γ -crystallin family. This not only reveals somewhat surprising evolutionary relationships but suggests more appropriate annotations for many of these proteins. For example, the clustering of fish and frog crystallins with mammalian γ D-crystallins suggests that γ D-like crystallins are more common and widespread than was previously supposed. Furthermore, the positioning of the fish γ M-crystallins in the nucleic acid tree indicates that they may have been derived from γ D-like crystallins rather than γ S-crystallins.

3.3 Structural analysis of crystallin homologs

The $\beta\gamma$ -crystallin fold is an ancient motif that first arose in microbial metal ion-binding proteins [131, 132]. Each Greek key consists of four interleaved β -strands (**Figure 3**). In order to explore the structural significance of high Cys content in γ -crystallins using our phylogenetically representative set of γ -crystallins, we downloaded experimentally determined structures from the Protein Data Bank for cases where they are available; for the other cases, structures were predicted using D-I-TASSER, the current most accurate machine-learning algorithm for protein structure prediction from multiple sequence alignments [61].

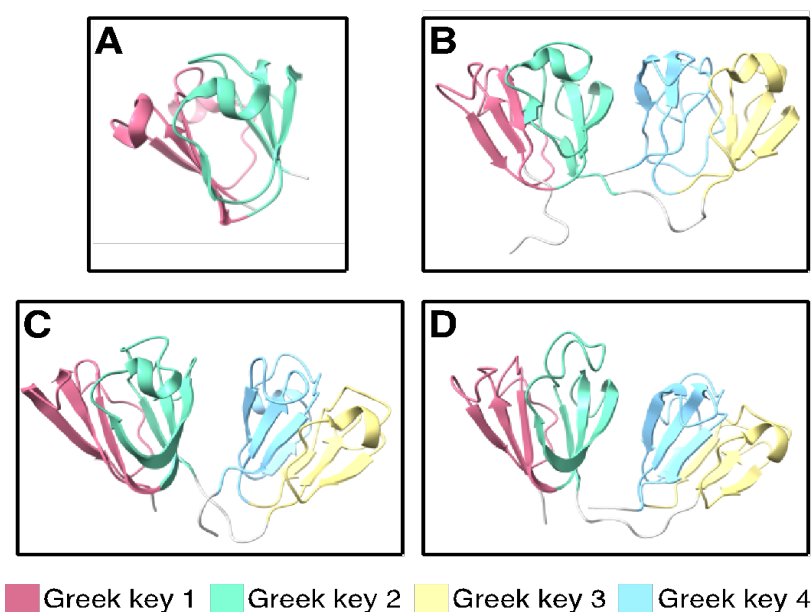


Figure 3. Experimentally determined structures of representative $\beta\gamma$ - and γ -crystallins. The double Greek key fold is common to $\beta\gamma$ -crystallins. **A** Tunicate (*Ciona intestinalis*) $\beta\gamma$ -crystallin (PDB ID: 2BV2) [133]. **B** Human γ S-crystallin (PDB ID: 2M3T) [58]. **C** Human γ D-crystallin (PDB ID: 2KLJ) [134]. **D** Zebrafish γ M7-crystallin (PDB ID: 2M3C) [135]. The individual Greek key motifs are colored in pink, green, yellow, and blue from N- to C-terminus.

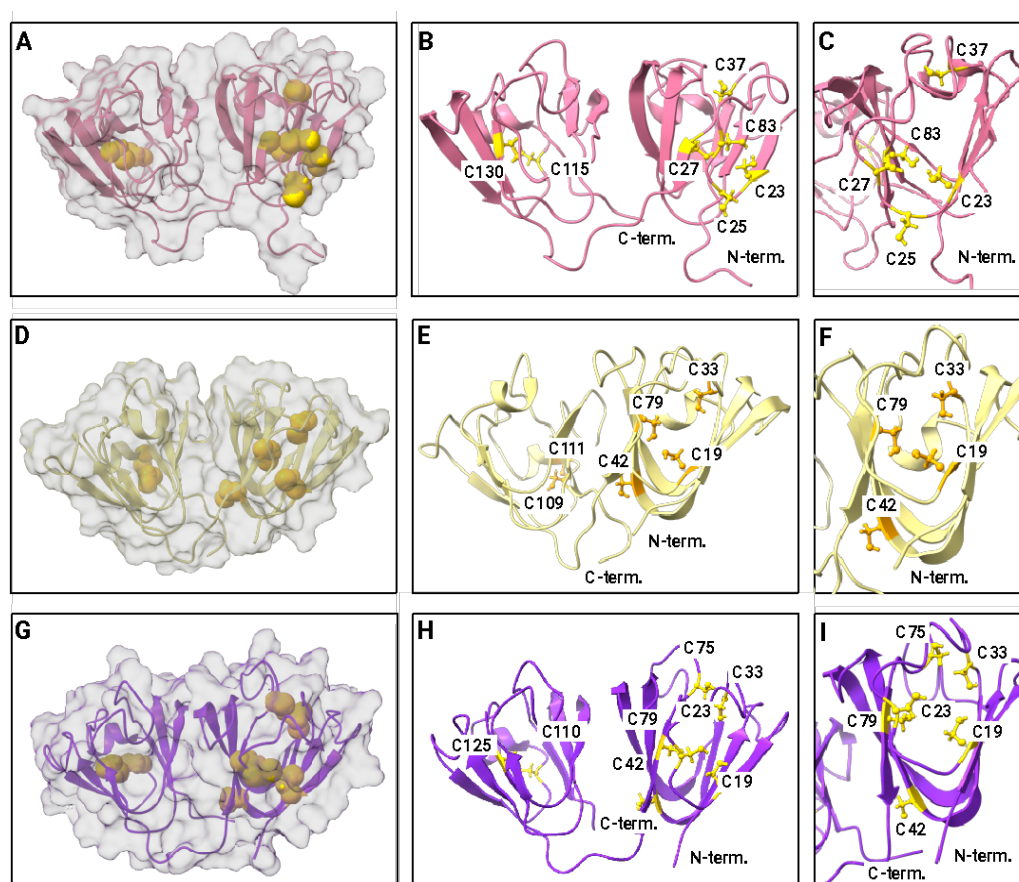


Figure 4. Cysteine in representative γ -crystallins. **A–C** human γ S-crystallin (PDB ID: 2M3T) [58]. Inset of the cysteine loop, formed by C23, C25, and C27 along with surrounding residues. **D–F** human γ D-crystallin (PDB ID: 2KLJ) [134]. **G–I** zebrafish γ M7-crystallin (PDB ID: 2M3C) [135]. Note that we use the UniProt convention for numbering γ -crystallin residues in this paper, counting Met1 as the first residue even though it is absent from the mature form of the protein. In other papers we and others have used the PDB numbering scheme of Basak et al. [136] or the traditional scheme based on alignment to human γ B-crystallin.

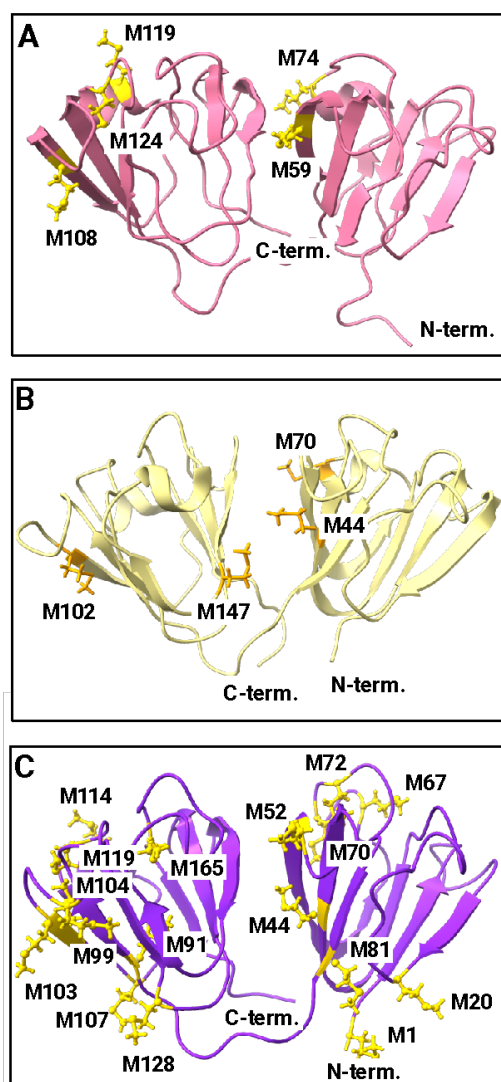


Figure 5. Methionine in representative γ -crystallins. **A** human γ S-crystallin (PDB ID: 2M3T) [58]. **B** human γ D-crystallin (PDB ID: 2KLJ) [134]. **C** zebrafish γ M7-crystallin (PDB ID: 2M3C) [135].

Protein structure visualizations were generated using ChimeraX [137, 138]. The lowest-energy predicted structures of all the $\beta\gamma$ -crystallins examined here show the same overall fold (models provided in the online **Supplementary Information**). Although ion binding dramatically stabilizes tunicate $\beta\gamma$ -crystallin, the two Ca^{2+} ions are coordinated to the loops internal to each Greek key, where they do not perturb the β -strand structure. Multiple sequence alignment of a large number of related sequences using DeepMSA (**Supplementary Figure S1**) showed that the strongly conserved residues – those most likely to be critical for maintaining the fold – are Ser and Gly, along with the tryptophans

commonly found in the hydrophobic core of each Greek key domain, the tyrosines associated with the tyrosine corners, and some surface-exposed charged residues.

Although they are not strongly conserved among crystallins generally, Cys and Met are highly important for lens crystallin function. The Cys and Met content of the crystallins investigated here are given in **Table 1 and Supplementary Table S1**. Initiator methionines are assumed to be cut off in the mature form of eukaryotic proteins and are therefore not included in the total. The positioning of cysteines in solution NMR structures for representative γ -crystallins is shown in **Figure 4**. Human γ S-crystallin, human γ D-crystallin, and zebrafish γ M-crystallin were chosen because of the availability of solution NMR structures for these three proteins. The top row shows human γ S-crystallin: **Panel A** shows the protein surface, revealing that all three cysteines in the N-terminal domain are at least partially exposed. The two cysteines in the C-terminal domain are fully buried. **Panel B** shows a ribbon diagram with all of the cysteines as ball and stick models. Their sequence positions are labeled; C25 is the most prominently exposed to solvent, although C23 and C27 are partially exposed as well. **Panel C** shows a view of the N-terminal Cys loop with the protein rotated 90° relative to the other views. The middle row shows human γ D-crystallin. **Panel A** shows the protein surface. Only C19 (C18 in the PDB numbering scheme of Basak et al. [51]) is present in a position corresponding to the Cys loop of γ S; it is fully buried. In contrast, the two C-terminal cysteines are partially exposed. **Panel B** shows a ribbon diagram with the cysteine residues labeled. **Panel C** shows a rotated view of the N-terminal Cys. The bottom row (**Panels G-I**) shows zebrafish γ M7-crystallin. Like human γ D-crystallin, it has only one Cys residue (C19) in the solvent loop; however, it is partly solvent exposed while its C-terminal cysteines are buried.

The positions of Met residues for the same proteins are shown in **Figure 5**. Human γ S (**Panel A**) and γ D-crystallin (**Panel B**) have five and four methionines, respectively, representing an abundance of about 5% in each case. In contrast, zebrafish γ M7-crystallin (**Panel C**) has 16 Met residues or 9.2%, a value that is high compared to typical proteins, but on the low side for fish γ M-crystallins, which can have a Met content of 20% or higher. The highest Met content for the proteins examined here is 22.5% for γ M1-crystallin from *Chiloscyllium indicum* (slender bamboo shark).

As discussed in Section 2, high Cys content puts lens γ -crystallins at high risk of disulfide-driven aggregation, though it also allows them to function as oxidoreductases. Using our set of solution NMR and predicted D-I-TASSER structures, we therefore analyzed solvent accessibility of Cys residues. Buried Cys are less likely to form disulfides, unless or until the protein adopts a misfolded conformation for a sufficiently long time. By contrast, natively solvent-exposed Cys residues present a more immediate risk: if there are solvent-exposed Cys residues in at least two distinct regions of the protein's surface, then head-to-tail aggregates may form via intermolecular disulfides. By contrast, if there is only one solvent-exposed Cys residue or residue cluster on the protein's surface, then aggregation will stop at the head-to-head dimer stage.

Our analysis revealed a striking observation, shown in **Figure 6**: the γ -crystallins in our set had solvent-exposed Cys residues in the N-terminal domain (NTD) or in the C-terminal domain (CTD), but not both. The only exception was γ M2b-crystallin from zebrafish. Of course, domain-level Cys accessibility analysis is still a relatively crude metric. The data overall strongly suggest that avoidance of multiple intermolecular disulfides per molecule, given the ensuing light-scattering aggregation, has been a significant evolutionary constraint of the γ -crystallin family.

4. Discussion

4.1 Evolutionary pressure for low light scattering by lens γ -crystallins over a lifetime

For our discussion of γ -crystallins, we focus on the three most common types in extant vertebrates: γ S-crystallins, γ D-crystallins, and γ M-crystallins. In humans, γ D-crystallin is concentrated primarily in the lens core, whereas γ S-crystallin is primarily in the lens cortex, where protein expression continues throughout life [139]. Based on our analyses of the sequences and predicted structures of these proteins, we propose that the most easily identifiable difference between γ S- and γ D-like crystallins, which have traditionally been differentiated based on features such as the hydrophobicity of the interdomain interface [66], is whether the DCD CD motif is present in the N-terminal domain. In humans, γ S- and γ D-crystallins are at the extreme ends of the spectrum in terms of NTD and CTD Cys exposure, respectively (Figure 6), and γ S-crystallins in general have at least two of these DC pairs (Supplementary Figure S3).

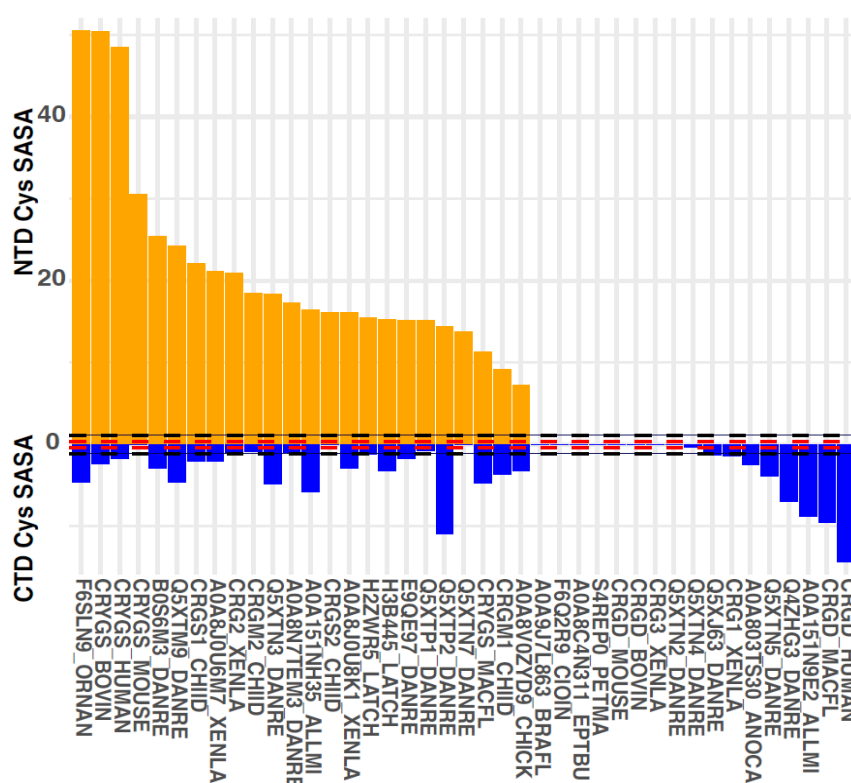


Figure 6. Combined solvent-accessible surface area (SASA), in square Angstroms, of Cys residues by domain in solution NMR structures or D-I-TASSER predicted structures of representative γ -crystallins. Protein nomenclature follows UniProt. Orange = N-terminal cysteines; blue = C-terminal cysteines. Red dashed lines indicate the mean SASA for a control buried Cys that is conserved in almost all of the γ -crystallins, e.g., C83 in human γ S (see Figure 2C). Black dashed lines indicate the mean + standard deviation for the control buried SASA.

In other vertebrates, the number of γ -crystallin paralogs and their distribution varies considerably. Teleost fish have many paralogs of γ M-crystallins, which share the same double-Greek key fold with γ S and γ D but are distinguished by their unusually high methionine content. Unlike the lenses of terrestrial organisms, which derive some optical power from an air-water interface at the cornea, the crystallins in fish lenses must provide all the refraction needed to focus an image on the retina. The Met-rich γ M-crystallins are thought to contribute to this function due to the very high refractivity of the Met side chain. The multitude of paralogs may also aid in maintaining solubility at the extreme concentrations found in the fish lens, much in the same way that frustration prevents crystallization in an ionic liquid. Another hypothesis that has been proposed is that the large number of

surface-exposed Met residues is responsible for mediating solvent interactions, making γ M-crystallins the least strongly hydrated of the structural crystallins, a feature that may aid in packing them in the crowded fish lens [140].

4.2. Evolutionary pressure for high optical power of lens γ -crystallins

Refractive index depends on lens shape, hydration, and protein structure, as well as post-translational modifications [100, 141]. The lens has a gradient of both protein concentration and distribution, with the highest protein concentration overall as well as the highest abundance of γ -crystallins near the central nucleus of the lens [142]. The optical properties of the lens are also influenced by the refractivity of the proteins themselves. The traditional understanding of protein refractivity is based on the additive Gladstone-Dale model, where the refractivity of a protein is determined by the weighted average of the refractivities of its component amino acids [19, 143, 144]. This model does not account for the differences in surface hydration between a collection of separate amino acids and a folded protein, nor potential interactions between pairs of aromatic residues, which are very highly enriched in lens crystallins relative to other proteins. Experimental measurements have shown that for lens crystallins, the Gladstone-Dale model significantly underestimates the refractivity based on sequence alone; a simple correction accounting for interactions between pairs of aromatic residues can correct for about two-thirds of this error [145]. Aside from structural factors, analysis of crystallin sequences reveals that they have very unusual amino acid compositions. They are highly enriched in the most polarizable residues (e.g., Arg, Cys, Met, Phe, Tyr, and Trp), at the expense of aliphatic residues that contribute little to refractivity.

The dual selection of crystallins for refractivity and solubility leads to some inherent tension, as the residues that best promote solubility (e.g., small polar and/or charged residues) are not the ones that most increase refractivity. The model of W. D. Hamilton [55] predicts that improved refractive power in younger age should be selected by evolution even at the cost of comparable increase in light scattering due to aggregation in old age – and that this should hold true even in fish or other species whose reproductive fitness increases with age.

Both cysteine and methionine significantly enhance refraction in the lens, a phenomenon that is best illustrated by the unusually high content of both, but especially methionine, in the fish γ M-crystallins. In fish, the lens provides all the focusing power of the eye, so the evolutionary pressure for high refractive index increment is especially intense; accordingly, fish crystallins are packed to concentrations of up to 1000 mg/ml and are highly refracting [23]. For the γ -crystallins found in Table 1, the average Cys content is 5.6% and the Met content is 10.9% for γ M-crystallins, compared to 3.6% and 2.4% for $\beta\gamma$ -crystallins, 3.9% and 2.2% for γ S-crystallins, and 1.7% and 2% for eukaryotic proteins in general [2]. Each of their side chains has a large, polarizable sulfur atom that contributes greatly to refractivity; however, these moieties are also susceptible to oxidation and chemical damage that may compromise solubility.

In particular, Cys plays many roles in proteins, such as functional redox chemistry and stabilizing disulfide bonds, but it can also powerfully promote misfolding and aggregation if incorrect intra- or intermolecular disulfides are formed. The only way to avoid Cys while retaining the high polarizability of sulfur is to replace it with Met; however, Met is even more readily oxidized than Cys, and the resulting sulfoxides could disrupt the core

packing and/or the local hydration environment of the protein. Comparison of the predicted Gladstone-Dale refractive index increment (the change in refractive index per concentration increment, dn/dc) for lens crystallins to other proteins with Greek key motifs suggests that the lens crystallins are more refractive than others with similar secondary structure, even though other Greek key proteins are also highly stable [146]. This hints at the tension between the dual functions of lens crystallins; amino acids that are selected for stability and solubility do not necessarily provide high refractivity and vice versa. These authors also found that the K/R ratio in $\beta\gamma$ -crystallins was lower than for comparable proteins, which they attribute to selection for stabilizing salt bridges. Alternatively, it may be due to selection for Arg itself, which has a higher reactive index increment than Lys. In fact, substituting Arg ($dn/dc = 0.206$ mL/g) for Lys ($dn/dc = 0.181$ mL/g) represents a major increase in refractivity without changing the charge of the residue.

Our examination of the abundance and positioning of cysteine and methionine in vertebrate γ -crystallins and their relatives in invertebrate chordates suggests that these residues provide the key to understanding the differences among types of γ -crystallins and generating hypotheses about their functionality in the lens. Based on the requirement for higher refractivity in the aquatic lens than its terrestrial counterpart, we would expect crystallins from aquatic organisms to be more enriched in Cys and Met than their terrestrial homologs. This is most dramatically exemplified by the large enrichment in these amino acids in the fish-specific γ M-crystallins, but it can also be observed for other aquatic crystallins in this data set. Notably, among the β -crystallins, which are assumed to have low refractivity in mammals, many fish sequences are highly enriched in Cys and Met even when their terrestrial homologs are not. For example, several of the β B1-crystallins from both *Danio rerio* and *Latimeria chalumnae* have Met contents in the 7% range, compared to 2% for mouse and 1.5% for human. This is notable because the zebrafish (ray-finned fish) and coelacanth (lobe-finned fish) are not close relatives, suggesting that the similarity is due to convergence in function. The increased abundance of Cys and Met concentration in the lenses of aquatic organisms (and those that require good vision, as opposed to e.g., the tunicate), may be selected due to increased refractivity, altered hydration properties, or both. These hypotheses require experimental validation, some of which is underway in our own and other labs.

4.3 Synergies between sulfur-containing and aromatic residues in lens γ -crystallins

Much of our discussion above highlights the evolutionary tension between the requirements for high optical power and high long-term transparency of the lens. We have speculated that these competing evolutionary pressures may have pushed γ -crystallins and β B-crystallins toward opposite extremes of Cys (and Met) abundance or depletion. Yet, these two broad phenotypic requirements do not always or necessarily conflict. Sulfur-aromatic interactions may contribute to the thermodynamic stability of these proteins [147, 148], but they may also contribute to anomalously high refractive index increments that are beyond what can be calculated from primary structure alone, as is the case for pi-pi and cation-pi interactions. Accordingly, convergent evolution has favored high sulfur and aromatic content not only in crystallins but also in reflectins, another protein family with exceptionally high refractive index [149]. Additionally, sulfur atoms may protect aromatic residues from irreversible damage by channeling away high-energy excitons produced by UV irradiation [47, 150], while the aromatics, by absorbing those photons and channeling them to the Cys residues, in particular, could photochemically reduce disulfide bonds [32, 151, 152]. Strategically positioned reversible disulfides might not only turn

crystallins into a proteinaceous redox buffer but also render the crystallin domains “self-healing,” capable of resisting the deleterious effects of certain denaturing stresses.

5. Conclusion

Much remains to be learned about how evolution has harnessed and balanced the many unique properties of the sulfur-containing amino acids to optimize and preserve vision. We have reviewed available evidence from experimental literature, genomic datasets, and ML-based protein structure prediction. We have found support for our hypotheses that lens crystallin families may have evolved to solve a trade-off between high refractive power and low late-life aggregation propensity and that the evolutionary record contains a signature of optimized spatial positioning of Cys residues to maximize useful redox activity while minimizing deleterious light-scattering aggregation. In addition, we found reason to suggest re-evaluating the evolutionary relationships among γ -crystallin subfamilies, as the γ M and γ D crystallin subfamilies in our dataset are unexpectedly similar with respect to their evolutionarily tuned Cys residue distributions. Based on our present review and investigation, we may hazard several experimentally testable hypotheses to focus future research in this area. Do crystallin thermodynamic and kinetic parameters vary predictably with an organism’s lifespan? Have crystallins in aquatic vertebrates faced a different balance of selective pressures for refractivity and transparency from those of land-dwelling ones? Do sulfur-aromatic interactions couple redox chemistry and photochemistry in the eye lens? We highlight the need for systematic comparisons of the biophysical and biochemical properties of lens crystallins across species and across crystallin subfamilies as one way of testing these hypotheses. We have previously offered conceptual models of disulfide flow and redox buffering in lens γ -crystallins [32, 38, 86]. Improved empirical understanding of the origins, benefits, and drawbacks of high Cys content in the γ -crystallin family will improve future therapeutic strategies to postpone or prevent the progression of turbidity in the aging lens.

Table 1. Cysteine and methionine content of $\beta\gamma$ - and γ -crystallins in representative chordates.

639

organism	protein	UniProt ID	method	length	Cys (no./%)	Met (no./ %)		
<i>Branchiostoma floridae</i>	S-crystallin	C3YKG6_BRAFL	1	128	6/4.7	4/3.1		
<i>Ciona intestinalis</i>	By-crystallin	F6Q2R9_CIOIN	3	84	0/0	0/0		
<i>Eptatretus burgeri</i>	By-crystallin	A0A8C4N311_EPTBU	3	153	7/4.6	3/2.0		
<i>Petromyzon marinus</i>	γ S-crystallin	S4REP0_PETMA	3	176	9/5.1	8/4.5		
<i>Danio rerio</i>	γ S1-crystallin	E9QE97_DANRE	1	178	8/4.5	2/1.1		
<i>Danio rerio</i>	γ S2-crystallin	Q5XTP1_DANRE	1	174	8/4.6	2/1.1		
<i>Latimeria chalumnae</i>	γ S-crystallin	H2ZWR5_LATCH	3	178	7/3.9	6/3.4		
<i>Ornithorhynchus anatinus</i>	γ S-crystallin	F6SLN9_ORNAN	1	178	6/3.4	3/1.7		
<i>Macropus fuliginosus</i>	γ S-crystallin	CRYGS_MACFL	4	178	8/4.5	4/2.2		
<i>Mus musculus</i>	γ S-crystallin	CRYGS_MOUSE	1	178	7/3.9	4/2.2		
<i>Bos taurus</i>	γ S-crystallin	CRYGS_BOVIN	1	178	6/3.4	6/3.4		
<i>Homo sapiens</i>	γ S-crystallin	CRYGS_HUMAN	1	178	7/3.9	5/2.8		
<i>Anolis carolinensis</i>	γ S-crystallin	A0A803TS30_ANOCA	1	179	5/2.8	5/2.8		
<i>Alligator mississippiensis</i>	γ S-crystallin	A0A151NH35_ALLMI	2	182	7/3.8	4/2.2		
<i>Gallus gallus</i>	γ S-crystallin	A0A8V0ZYD9_CHICK	1	175	7/4.0	3/1.7		
<i>Macropus fuliginosis</i>	γ D-crystallin	CRGD_MACFL	4	174	8/4.6	8/4.6		
<i>Mus musculus</i>	γ D-crystallin	CRGD_MOUSE	1	174	7/4.0	7/4.0		
<i>Bos taurus</i>	γ D-crystallin	CRGD_BOVIN	4	174	5/2.9	5/2.9		
<i>Homo sapiens</i>	γ D-crystallin	CRGD_HUMAN	1	174	6/3.4	5/2.9		
<i>Alligator mississippiensis</i>	γ D-crystallin	A0A151N9E2_ALLMI	2	175	6/3.4	7/4.0		
<i>Xenopus laevis</i>	γ -crystallin 1	CRG1_XENLA	2	175	5/2.9	5/2.9		

<i>Xenopus laevis</i>	γ-crystallin 2	CRG2_XENLA	4	175	7/4.0	6/3.4		
<i>Xenopus laevis</i>	γ-crystallin 3	CRG3_XENLA	2	175	6/2.4	5/2.9		
<i>Xenopus laevis</i>	γ-crystallin 4	A0A8J0U6M7_XENLA	1	175	7/4.0	6/3.4		
<i>Xenopus laevis</i>	γ-crystallin 5	A0A8J0U8K1_XENLA	1	175	6/3.4	5/2.9		
<i>Chiloscyllium indicum</i>	γS1-crystallin	CRGS1_CHIID	4	173	7/4.0	10/5.8		
<i>Chiloscyllium indicum</i>	γS2-crystallin	CRGS2_CHIID	2	173	6/3.5	9/5.2		
<i>Chiloscyllium indicum</i>	γM2-crystallin	CRGM2_CHIID	4	176	8/4.5	9/5.1		
<i>Latimeria chalumnae</i>	γM2-crystallin	H3B445_LATCH	3	176	9/5.1	4/2.3		
<i>Danio rerio</i>	γS3-crystallin	Q5XTN7_DANRE	1	183	6/3.3	7/3.8		
<i>Danio rerio</i>	γS4-crystallin	Q5XTN2_DANRE	1	176	7/4.0	2/1.1		
<i>Chiloscyllium indicum</i>	γM1-crystallin	CRGM1_CHIID	4	120	10/8.3	27/22.5		
<i>Danio rerio</i>	γM1-crystallin	Q5XTN6_DANRE	1	178	8/4.5	19/10.7		
<i>Danio rerio</i>	γM2-crystallin	A0A8N7TEM3_DANRE	1	174	10/5.7	22/12.6		
<i>Danio rerio</i>	γM2a-crystallin	Q4ZHG3_DANRE	1	181	10/5.5	24/13.3		
<i>Danio rerio</i>	γM2c-crystallin	Q5XTP2_DANRE	1	175	10/5.7	22/12.6		
<i>Danio rerio</i>	γM2d1-crystallin	B0S6M3_DANRE	2	175	10/5.7	23/13.1		
<i>Danio rerio</i>	γM3-crystallin	Q5XTM9_DANRE	1	174	10/5.7	13/7.5		
<i>Danio rerio</i>	γM4-crystallin	Q5XTN5_DANRE	1	174	9/5.2	10/5.7		
<i>Danio rerio</i>	γM5-crystallin	Q5XJ63_DANRE	1	177	11/6.2	12/6.8		
<i>Danio rerio</i>	γM6-crystallin	Q5XTN4_DANRE	2	177	7/4.0	11/6.2		
<i>Danio rerio</i>	γM7-crystallin	Q5XTN3_DANRE	1	174	8/4.6	16/9.2		

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Figure S1: title; Table S1: title; Video S1: title.

640
641
642
643
644

Author Contributions: Conceptualization, E.S. and R.W.M.; methodology, E.S., R.W.M, and G.R.T.; investigation, E.S. and R.W.M; data curation, E.S., G.R.T and R.W.M.; writing—original draft preparation, E.S. G.R.T and R.W.M; writing—review and editing, E.S. G.R.T. and R.W.M.; visualization, G.R.T. and R.W.M.; supervision, E.S. and R.W.M.; project administration, E.S. and R.W.M.; funding acquisition, E.S. and R.W.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Science Foundation grant BMAT DMR-2002837 to RWM and D.J. Tobias. G.R.T. was supported by GAAN P200A210024.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: The sequences and structures used in this work were downloaded from publicly available repositories, including UniProt, the NCBI database, and the Protein Data Bank. Structural models predicted using D-I-Tasser are provided as a zip file in the Supplementary Information.

Acknowledgments: The authors thank Ethan Kocak for preparation of the animal drawings used in Figure 4.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Supplementary information

The Supplementary information contains DeepMSA plots and sequence alignments for γ - and β -crystallins as well as a table of Cys and Met content for selected chordate β -crystallins.

References

- Miseta, A. and P. Csutora, *Relationship between the occurrence of cysteine in proteins and the complexity of organisms*. Molecular Biology and Evolution, 2000. **17**(8): p. 1232-1239.
- Castillo-Villanueva, A., H. Reyes-Vivas, and J. Oria-Hernández, *Comparison of cysteine content in whole proteomes across the three domains of life*. PLoS One, 2023. **18**(11): p. e0294268.
- Bessette, P.H., F. Aslund, J. Beckwith, and G. Georgiou, *Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm*. Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(24): p. 13703-13708.
- Lobstein, J., C.A. Emrich, C. Jeans, M. Faulkner, P. Riggs, and M. Berkmen, *SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm*. Microbial Cell Factories, 2012. **11**: p. 16.
- Bakalova, R., Z. Zhelev, I. Aoki, and T. Saga, *Tissue Redox Activity as a Hallmark of Carcinogenesis: From Early to Terminal Stages of Cancer*. Clinical Cancer Research, 2013. **19**(9): p. 2503-2517.
- Jorgenson, T.C., W.X. Zhong, and T.D. Oberley, *Redox Imbalance and Biochemical Changes in Cancer*. Cancer Research, 2013. **73**(20): p. 6118-6123.
- Tasdogan, A., J.M. Ubellacker, and S.J. Morrison, *Redox Regulation in Cancer Cells during Metastasis*. Cancer Discovery, 2021. **11**(11): p. 2682-2692.
- Xiao, H., M.P. Jedrychowski, D.K. Schweppe, E.L. Huttlin, Q. Yu, D.E. Heppner, J. Li, J. Long, E.L. Mills, J. Szpyt, et al., *A Quantitative Tissue-Specific Landscape of Protein Redox Regulation during Aging*. Cell, 2020. **180**(5): p. 968-983 e24.

9. Maldonado, E., S. Morales-Pison, F. Urbina, and A. Solari, *Aging Hallmarks and the Role of Oxidative Stress. Antioxidants* (Basel), 2023. **12**(3). 689 690
10. Wishart, T.F.L., M. Flokis, D.Y. Shu, S.J. Das, and F.J. Lovicu, *Hallmarks of lens aging and cataractogenesis*. *Exp Eye Res*, 2021. **210**: p. 108709. 691 692
11. Centers for Disease Control and Prevention, V.H.I. 2019. 693
12. Mencucci, R., S. Stefanini, E. Favuzza, M. Cennamo, C. De Vitto, and E. Mossello, *Beyond vision: Cataract and health status in old age, a narrative review*. *Front Med* (Lausanne), 2023. **10**: p. 1110383. 694 695
13. Klein, B.E., R. Klein, and K.E. Lee, *Incidence of age-related cataract: the Beaver Dam Eye Study*. *Arch Ophthalmol*, 1998. **116**(2): p. 219-25. 696 697
14. Wride, M.A., *Lens fibre cell differentiation and organelle loss: many paths lead to clarity*. *Philos Trans R Soc Lond B Biol Sci*, 2011. **366**(1568): p. 1219-33. 698 699
15. Lynnerup, N., H. Kjeldsen, S. Heegaard, C. Jacobsen, and J. Heinemeier, *Radiocarbon dating of the human eye lens crystallines reveal proteins without carbon turnover throughout life*. *PLoS One*, 2008. **3**(1): p. e1529. 700 701
16. Wistow, G., *The human crystallin gene families*. *Hum Genomics*, 2012. **6**(1): p. 26. 702
17. Horwitz, J., *Alpha-crystallin can function as a molecular chaperone*. *Proceedings of the National Academy of Sciences of the United States of America*, 1992. **89**(21): p. 10449-10453. 703 704
18. Sprague-Piercy, M.A., M.A. Rocha, A.O. Kwok, and R.W. Martin, *α -crystallins in the vertebrate eye lens: Complex oligomers and molecular chaperones*. *Annual Review of Physical Chemistry*, 2020. **72**: p. 6.1-6.21. 705 706
19. Zhao, H., P.H. Brown, M.T. Magone, and P. Schuck, *The molecular refractive function of lens γ crystallins*. *Journal of Molecular Biology*, 2011. **411**: p. 680-699. 707 708
20. Wistow, G. and J. Piatigorsky, *Recruitment of Enzymes as Lens Structural Proteins*. *Science*, 1987. **236**(4808): p. 1554-1556. 709 710
21. Sampaleanu, L.M., A.R. Davidson, C. Graham, G.J. Wistow, and P.L. Howell, *Domain exchange experiments in duck delta-crystallins: functional and evolutionary implications*. *Protein Sci*, 1999. **8**(3): p. 529-37. 711 712
22. Vendra, V.P.R., I. Khan, S. Chandani, A. Muniyandi, and D. Balasubramanian, *Gamma crystallins of the human eye lens*. *Biochimica et Biophysica Acta - General Subjects*, 2016. **1860**(1): p. 333-343. 713 714
23. Bloemendal, H., W.d. Jong, R. Jaenicke, N. H. Lubsen, C. Slingsby, and A. Tardieu, *Ageing and vision: structure, stability and function of lens crystallins*. *Prog. Biophys. Mol. Biol.*, 2004. **86**(3): p. 407-485. 715 716
24. Rocha, M.A., M.A. Sprague-Piercy, A.O. Kwok, K.W. Roskamp, and R.W. Martin, *Chemical properties determine solubility and stability in β -gamma-crystallins of the eye lens*. *ChemBioChem*, 2021. **22**: p. 1329-1346. 717 718
25. Smith, M.A., O.A. Bateman, R. Jaenicke, and C. Slingsby, *Mutation of interfaces in domain-swapped human β B2-crystallin*. *Protein Sci.*, 2007. **16**(4): p. 615-625. 719 720
26. Xi, Z., M.J. Whitley, and A.M. Gronenborn, *Human β B2-crystallin forms a face-en-face dimer in solution: An integrated NMR and SAXS study*. *Structure*, 2017. **25**: p. 496-505. 721 722
27. Rolland, A.D., T. Takata, M.T. Donor, K.J. Lampi, and J.S. Prell, *Eye lens β -crystallins are predicted by native ion mobility-mass spectrometry and computations to form compact higher-ordered heterooligomers*. *Structure*, 2023. **31**(9): p. P1052-1064.E3. 723 724 725
28. D'Alessio, G., *The evolution of monomeric and oligomeric $\beta\gamma$ -type crystallins*. *Eur. J. Biochem.*, 2002. **269**(13): p. 3122-3130. 726 727
29. Piatigorsky, J. and G. Wistow, *The recruitment of crystallins: New functions precede gene duplication*. *Science*, 1991. **252**(5009): p. 1078-1079. 728 729
30. Lou, M.F., *Redox regulation in the lens*. *Prog Retin Eye Res*, 2003. **22**(5): p. 657-82. 730
31. Spector, A. and D. Roy, *Disulfide-linked high molecular weight protein associated with human cataract*. *Proc Natl Acad Sci U S A*, 1978. **75**(7): p. 3244-8. 731 732
32. Serebryany, E., D.C. Thorn, and L. Quintanar, *Redox chemistry of lens crystallins: A system of cysteines*. *Exp Eye Res*, 2021. **211**: p. 108707. 733 734
33. Truscott, R.J.W. and R.C. Augusteyn, *OXIDATIVE CHANGES IN HUMAN LENS PROTEINS DURING SENILE NUCLEAR CATARACT FORMATION*. *Biochimica Et Biophysica Acta*, 1977. **492**(1): p. 43-52. 735 736
34. Sweeney, M.H. and R.J. Truscott, *An impediment to glutathione diffusion in older normal human lenses: a possible precondition for nuclear cataract*. *Exp Eye Res*, 1998. **67**(5): p. 587-95. 737 738

35. Hains, P.G. and R.J. Truscott, *Proteomic analysis of the oxidation of cysteine residues in human age-related nuclear cataract lenses*. *Biochim Biophys Acta*, 2008. **1784**(12): p. 1959-64. 739 740
36. Fan, X., S. Zhou, B. Wang, G. Hom, M. Guo, B. Li, J. Yang, D. Vaysburg, and V.M. Monnier, *Evidence of Highly Conserved beta-Crystallin Disulfidome that Can be Mimicked by In Vitro Oxidation in Age-related Human Cataract and Glutathione Depleted Mouse Lens*. *Mol Cell Proteomics*, 2015. **14**(12): p. 3211-23. 741 742 743
37. Serebryany, E., J.C. Woodard, B.V. Adkar, M. Shabab, J.A. King, and E.I. Shakhnovich, *An Internal Disulfide Locks a Misfolded Aggregation-prone Intermediate in Cataract-linked Mutants of Human gammaD-Crystallin*. *J Biol Chem*, 2016. **291**(36): p. 19172-83. 744 745 746
38. Serebryany, E., S. Yu, S.A. Trauger, B. Budnik, and E.I. Shakhnovich, *Dynamic disulfide exchange in a crystallin protein in the human eye lens promotes cataract-associated aggregation*. *J Biol Chem*, 2018. **293**(46): p. 17997-18009. 747 748
39. Norton-Baker, B., P. Mehrabi, A.O. Kwok, K.W. Roskamp, M.A. Rocha, M.A. Sprague-Piercy, D.v. Stetten, R.J.D. Miller, and R.W. Martin, *Deamidation of the human eye lens protein γ S-crystallin accelerates oxidative aging*. *Structure*, 2022. **30**(5): p. 763-776. 749 750 751
40. Serebryany, E., S. Chowdhury, C.N. Woods, D.C. Thorn, N.E. Watson, A. McClelland, R.E. Klevit, and E.I. Shakhnovich, *A native chemical chaperone in the human eye lens*. *eLife*, 2022. **11**: p. e76923. 752 753
41. Quintanar, L., J.A. Dominguez-Calva, E. Serebryany, L. Rivillas-Acevedo, C. Haase-Pettingell, C. Amero, and J.A. King, *Copper and Zinc Ions Specifically Promote Nonamyloid Aggregation of the Highly Stable Human gamma-D Crystallin*. *ACS Chem Biol*, 2016. **11**(1): p. 263-72. 754 755 756
42. Dominguez-Calva, J.A., M.L. Perez-Vazquez, E. Serebryany, J.A. King, and L. Quintanar, *Mercury-induced aggregation of human lens gamma-crystallins reveals a potential role in cataract disease*. *J Biol Inorg Chem*, 2018. **23**(7): p. 1105-1118. 757 758 759
43. Palomino-Vizcaino, G., J.A. Domínguez-Calva, E. Martínez-Jurado, N. Schuth, O. Rodríguez-Meza, E. Serebryany, J.A. King, T. Kroll, M. Costas, and L. Quintanar, *Free radical chemistry in copper-induced aggregation of human lens gamma-crystallins and its relevance to cataract disease*. In Preparation, 2021. 760 761 762
44. Roskamp, K.W., N. Kozlyuk, S. Sengupta, J.C. Bierma, and R.W. Martin, *Divalent cations and the divergence of β / γ -crystallin function*. *Biochemistry*, 2019. **58**(45): p. 4505-4518. 763 764
45. Ramirez-Bello, V., J. Martinez-Seoane, A. Fernández-Silva, and C. Amero, *Zinc and copper ions induce aggregation of human β -crystallins*. *Molecules*, 2022. **27**(9): p. 2970. 765 766
46. Lapko, V.N., D.L. Smith, and J.B. Smith, *Methylation and carbamylation of human γ -crystallins*. *Protein Science*, 2003. **12**(8): p. 1762-1774. 767 768
47. Schafheimer, N., Z. Wang, K. Schey, and J. King, *Tyrosine/Cysteine Cluster Sensitizing Human gamma D-Crystallin to Ultraviolet Radiation-Induced Photoaggregation in Vitro*. *Biochemistry*, 2014. **53**(6): p. 979-990. 769 770
48. Zhang, C., W. Zheng, S.M. Mortuza, Y. Li, and Y. Zhang, *DeepMSA: Constructing deep multiple sequence alignment to improve contact prediction and fold-recognition for distant-homology proteins*. *Bioinformatics*, 2020. **36**: p. 2105-2112. 771 772 773
49. Zheng, W., Q. Wuyun, Y. Li, C. Zhang, P.L. Freddolino, and Y. Zhang, *Improving deep learning protein monomer and complex structure prediction using DeepMSA2 with huge metagenomics data*. *Nature Methods*, 2024. **21**: p. 279-289. 774 775 776
50. Batrinos, M.L., *The length of life and eugeria in classical Greece*. *Hormones (Athens)*, 2008. **7**(1): p. 82-3. 777
51. Gurven, M. and H. Kaplan, *Longevity among hunter-gatherers: A cross-cultural examination*. *Population and Development Review*, 2007. **33**(2): p. 321-365. 778 779
52. Hawkes, K. and J.E. Coxworth, *Grandmothers and the evolution of human longevity: a review of findings and future directions*. *Evol Anthropol*, 2013. **22**(6): p. 294-302. 780 781
53. Lieberman, D.E., T.M. Kistner, D. Richard, I.M. Lee, and A.L. Baggish, *The active grandparent hypothesis: Physical activity and the evolution of extended human healthspans and lifespans*. *Proc Natl Acad Sci U S A*, 2021. **118**(50). 782 783
54. Chen, S., F.A. Ferrone, and R. Wetzel, *Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation*. *Proc Natl Acad Sci U S A*, 2002. **99**(18): p. 11884-9. 784 785
55. Hamilton, W.D., *The moulding of senescence by natural selection*. *J Theor Biol*, 1966. **12**(1): p. 12-45. 786
56. Brune, D., M.A. Andrade-Navarro, and P. Mier, *Proteome-wide comparison between the amino acid composition of domains and linkers*. *BMC Res Notes*, 2018. **11**(1): p. 117. 787 788

57. Kiss, A.J., A.Y. Mirarefi, S. Ramakrishnan, C.F. Zukoski, A.L. DeVries, and C.-H.C. Cheng, *Cold-stable eye lens crystallins of the Antarctic nototheniid toothfish <i>Dissostichus mawsoni</i> Norman*. Journal of Experimental Biology, 2004. **207**: p. 4633-4649. 789-791
58. Kingsley, C.N., J. Bierma, V. Pham, and R.W. Martin, *The γ S-crystallin proteins from the Antarctic Nototheniid toothfish: A model system for investigating differential resistance to chemical and thermal denaturation*. Journal of Physical Chemistry B, 2014. **118**(47): p. 13544-13553. 792-794
59. Bierma, J.C., K.W. Roskamp, A.P. Ledray, A.J. Kiss, C.H.C. Cheng, and R.W. Martin, *Controlling liquid-liquid phase separation in cold-adapted crystallin proteins from the Antarctic toothfish*. Journal of Molecular Biology, 2018. **430**(24): p. 5151-5168. 795-797
60. Benedek, G.B., J.I. Clark, E.N. Serrallach, C.Y. Young, L. Mengel, T. Sauke, A. Bagg, and K. Benedek, *Light scattering and reversible cataracts in the calf and human lens*. Philos. Trans. R. Soc., A, 1979. **293**(1402): p. 329-340. 798-799
61. Zheng, W., Q. Wuyun, Y. Li, Q. Liu, X. Zhou, Y. Zhu, P.L. Freddolino, and Y. Zhang, *Integrating deep learning potentials with I-TASSER for single- and multi-domain protein structure prediction*. Submitted, 2023. 800-801
62. Serebryany, E. and J.A. King, *The $\beta\gamma$ -crystallins: Native state stability and pathways to aggregation*. Progress in Biophysics & Molecular Biology, 2014. **115**: p. 32-41. 802-803
63. Kosinski-Collins, M.S. and J. King, *In vitro unfolding, refolding, and polymerization of human gammaD crystallin, a protein involved in cataract formation*. Protein Sci, 2003. **12**(3): p. 480-90. 804-805
64. Kong, F. and J. King, *Contributions of aromatic pairs to the folding and stability of long-lived human gammaD-crystallin*. Protein Sci, 2011. **20**(3): p. 513-28. 806-807
65. Mills-Henry, I.A., S.L. Thol, M.S. Kosinski-Collins, E. Serebryany, and J.A. King, *Kinetic Stability of Long-Lived Human Lens gamma-Crystallins and Their Isolated Double Greek Key Domains*. Biophys J, 2019. **117**(2): p. 269-280. 808-809
66. Mills, I.A., S.L. Flaugh, M.S. Kosinski-Collins, and J.A. King, *Folding and stability of the isolated Greek key domains of the long-lived human lens proteins γ D-crystallin and γ S-crystallin*. Protein Sci., 2007. **16**(11): p. 2427-2444. 810-811
67. Purkiss, A.G., O.A. Bateman, K. Wyatt, P.A. Wilmarth, L.L. David, G.J. Wistow, and C. Slingsby, *Biophysical properties of gammaC-crystallin in human and mouse eye lens: the role of molecular dipoles*. J Mol Biol, 2007. **372**(1): p. 205-22. 812-814
68. Bateman, O.A., R. Sarra, S.T. van Genesen, G. Kappe, N.H. Lubsen, and C. Slingsby, *The stability of human acidic beta-crystallin oligomers and hetero-oligomers*. Exp Eye Res, 2003. **77**(4): p. 409-22. 815-816
69. Takata, T., J.T. Oxford, T.R. Brandon, and K.J. Lampi, *Deamidation alters the structure and decreases the stability of human lens betaA3-crystallin*. Biochemistry, 2007. **46**(30): p. 8861-71. 817-818
70. Sen, A.C. and B. Chakrabarti, *Effect of acetylation by aspirin on the thermodynamic stability of lens crystallins*. Exp Eye Res, 1990. **51**(6): p. 701-9. 819-820
71. Lampi, K.J., P.A. Wilmarth, M.R. Murray, and L.L. David, *Lens beta-crystallins: the role of deamidation and related modifications in aging and cataract*. Prog Biophys Mol Biol, 2014. **115**(1): p. 21-31. 821-822
72. Su, S., P. Liu, H. Zhang, Z.J. Li, Z. Song, L. Zhang, and S. Chen, *Proteomic Analysis of Human Age-related Nuclear Cataracts and Normal Lens Nuclei*. Invest Ophthalmol Vis Sci, 2011. **52**(7): p. 4182-4191. 823-824
73. Ramkumar, S., X. Fan, B. Wang, S. Yang, and V.M. Monnier, *Reactive cysteine residues in the oxidative dimerization and Cu(2+) induced aggregation of human gammaD-crystallin: Implications for age-related cataract*. Biochim Biophys Acta Mol Basis Dis, 2018. **1864**(11): p. 3595-3604. 825-827
74. Delaye, M. and A. Tardieu, *Short-range order of crystallin proteins accounts for eye lens transparency*. Nature, 1983. **302**: p. 415-417. 828-829
75. Wong, T.Y., B.E. Klein, R. Klein, and S.C. Tomany, *Relation of ocular trauma to cortical, nuclear, and posterior subcapsular cataracts: the Beaver Dam Eye Study*. Br J Ophthalmol, 2002. **86**(2): p. 152-5. 830-831
76. Smith, M.P., M.H. Colyer, E.D. Weichel, and R.D. Stutzman, *Traumatic cataracts secondary to combat ocular trauma*. J Cataract Refract Surg, 2015. **41**(8): p. 1693-8. 832-833
77. Ao, M., X. Li, W. Qiu, Z. Hou, J. Su, and W. Wang, *The impact of age-related cataracts on colour perception, postoperative recovery and related spectra derived from test of hue perception*. BMC Ophthalmol, 2019. **19**(1): p. 56. 834-835
78. Pande, A., J. Pande, N. Asherie, A. Lomakin, O. Ogun, J. King, and G.B. Benedek, *Crystal cataracts: Human genetic cataract caused by protein crystallization*. Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**: p. 6116-6120. 836-838

79. Lee, S., B. Mahler, J. Toward, B. Jones, K. Wyatt, L.J. Dong, G. Wistow, and Z.R. Wu, *A Single Destabilizing Mutation (F9S) Promotes Concerted Unfolding of an Entire Globular Domain in gamma S-Crystallin*. *Journal of Molecular Biology*, 2010. **399**(2): p. 320-330. 839-841
80. Boatz, J.C., M.J. Whitley, M. Li, A.M. Gronenborn, and P.C.A.v.d. Wel, *Cataract-associated P23T γ D-crystallin retains a native-like fold in amorphous-looking aggregates formed at physiological pH*. *Nature Communications*, 2017. **8**: p. Article-number 15137. 842-844
81. Serebryany, E. and J.A. King, *Wild-type human gammaD-crystallin promotes aggregation of its oxidation-mimicking, misfolding-prone W42Q mutant*. *J Biol Chem*, 2015. **290**(18): p. 11491-503. 845-846
82. Serebryany, E., T. Takata, E. Erickson, N. Schafheimer, Y. Wang, and J.A. King, *Aggregation of Trp > Glu point mutants of human gamma-D crystallin provides a model for hereditary or UV-induced cataract*. *Protein Sci*, 2016. **25**(6): p. 1115-28. 847-849
83. Diessner, E.M., J.A. Freites, D.J. Tobias, and C.T. Butts, *Network Hamiltonian Models for Unstructured Protein Aggregates, with Application to gammaD-Crystallin*. *J Phys Chem B*, 2023. **127**(3): p. 685-697. 850-851
84. Quintanar, L., C. Domínguez, J. A. E. Serebryany, L. Rivillas-Acevedo, C. Haase-Pettingell, C. Amero, and J.A. King, *Copper and zinc ions specifically promote nonamyloid aggregation of the highly stable human γ D crystallin*. *ACS Chemical Biology*, 2016. **11**(1): p. 263-272. 852-854
85. Dominguez-Calva, J.A., C. Haase-Pettingell, E. Serebryany, J.A. King, and L. Quintanar, *A Histidine Switch for Zn-Induced Aggregation of gamma-Crystallins Reveals a Metal-Bridging Mechanism That Is Relevant to Cataract Disease*. *Biochemistry*, 2018. **57**(33): p. 4959-4962. 855-857
86. Roskamp, K.W., S. Azim, G. Kassier, B. Norton-Baker, M.A. Sprague-Piercy, R.J.D. Miller, and R.W. Martin, *Human γ S-crystallin copper binding helps buffer against aggregation caused by oxidative damage*. *Biochemistry*, 2020. **59**(25): p. 2371-2385. 858-860
87. Roskamp, K.W., D.M. Montelongo, C.D. Anorma, D.N. Bandak, J.A. Chua, K. Malecha, and R.W. Martin, *Thermal-, pH-, and UV-induced aggregation of human γ S-crystallin and its aggregation-prone G18V variant*. *Investigative Ophthalmology and Visual Science*, 2017. **58**: p. 2397-2405. 861-863
88. Moran, S.D., T.O. Zhang, S.M. Decatur, and M.T. Zanni, *Amyloid fiber formation in human gammaD-Crystallin induced by UV-B photodamage*. *Biochemistry*, 2013. **52**(36): p. 6169-81. 864-865
89. Ansari, R.R. and M.B. Datile, 3rd, *Use of dynamic light scattering and Scheimpflug imaging for the early detection of cataracts*. *Diabetes Technol Ther*, 1999. **1**(2): p. 159-68. 866-867
90. Sarangi, S., O. Minaeva, D.M. Ledoux, D.S. Parsons, J.A. Moncaster, C.A. Black, J. Hollander, Y. Tripodis, J.I. Clark, D.G. Hunter, and L.E. Goldstein, *In vivo quasi-elastic light scattering detects molecular changes in the lenses of adolescents with Down syndrome*. *Exp Eye Res*, 2024. **241**: p. 109818. 868-870
91. Moncaster, J.A., R.D. Moir, M.A. Burton, O. Chadwick, O. Minaeva, V.E. Alvarez, M. Ericsson, J.I. Clark, A.C. McKee, R.E. Tanzi, and L.E. Goldstein, *Alzheimer's disease amyloid-beta pathology in the lens of the eye*. *Exp Eye Res*, 2022. **221**: p. 108974. 871-873
92. Alperstein, A.M., J.S. Ostrander, T.O. Zhang, and M.T. Zanni, *Amyloid found in human cataracts with two-dimensional infrared spectroscopy*. *Proc Natl Acad Sci U S A*, 2019. **116**(14): p. 6602-6607. 874-875
93. Costello, M.J., A. Burette, M. Weber, S. Metlapally, K.O. Gilliland, W.C. Fowler, A. Mohamed, and S. Johnsen, *Electron tomography of fiber cell cytoplasm and dense cores of multilamellar bodies from human age-related nuclear cataracts*. *Exp Eye Res*, 2012. **101**: p. 72-81. 876-878
94. Zigman, S. and S. Lerman, *A Cold Precipitable Protein in the Lens*. *Nature*, 1964. **203**: p. 662-3. 879
95. Vecchi, G., P. Sormanni, B. Mannini, A. Vandelli, G.G. Tartaglia, C.M. Dobson, F.U. Hartl, and M. Vendruscolo, *Proteome-wide observation of the phenomenon of life on the edge of solubility*. *Proc Natl Acad Sci U S A*, 2020. **117**(2): p. 1015-1020. 880-882
96. de Graff, A.M., M.J. Hazoglou, and K.A. Dill, *Highly Charged Proteins: The Achilles' Heel of Aging Proteomes*. *Structure*, 2016. **24**(2): p. 329-36. 883-884
97. Schmid, P.W.N., N.C.H. Lim, C. Peters, K.C. Back, B. Bourgeois, F. Pirolt, B. Richter, J. Peschek, O. Puk, O.V. Amarie, et al., *Imbalances in the eye lens proteome are linked to cataract formation*. *Nat Struct Mol Biol*, 2021. **28**(2): p. 143-151. 885-887
98. Hains, P.G. and R.J. Truscott, *Post-translational modifications in the nuclear region of young, aged, and cataract human lenses*. *J Proteome Res*, 2007. **6**(10): p. 3935-43. 888-889

99. Cantrell, L.S. and K.L. Schey, *Proteomic characterization of the human lens and Cataractogenesis*. Expert Rev Proteomics, 2021. **18**(2): p. 119-135. 890 891
100. Quinlan, R.A. and J.I. Clark, *Insights into the biochemical and biophysical mechanisms mediating the longevity of the transparent optics of the eye lens*. Journal of Biological Chemistry, 2022. **298**(11): p. 102537. 892 893
101. Borchman, D. and M.C. Yappert, *Age-related lipid oxidation in human lenses*. Invest Ophthalmol Vis Sci, 1998. **39**(6): p. 1053-8. 894 895
102. Deeley, J.M., J.A. Hankin, M.G. Friedrich, R.C. Murphy, R.J.W. Truscott, T.W. Mitchell, and S.J. Blanksby, *Sphingolipid distribution changes with age in the human lens*. Journal of Lipid Research, 2010. **51**(9): p. 2753-2760. 896 897
103. Bejarano, E., J. Weinberg, M. Clark, A. Taylor, S. Rowan, and E.A. Whitcomb, *Redox Regulation in Age-Related Cataracts: Roles for Glutathione, Vitamin C, and the NRF2 Signaling Pathway*. Nutrients, 2023. **15**(15). 898 899
104. Thorn, D.C., A.B. Grosas, P.D. Mabbitt, N.J. Ray, C.J. Jackson, and J.A. Carver, *The Structure and Stability of the Disulfide-Linked gammaS-Crystallin Dimer Provide Insight into Oxidation Products Associated with Lens Cataract Formation*. J Mol Biol, 2019. **431**(3): p. 483-497. 900 902
105. Vetter, C.J., D.C. Thorn, S.G. Wheeler, C. Mundorff, K. Halverson, T.E. Wales, U. Shinde, J.R. Engen, L.L. David, J.A. Carver, and K.J. Lampi, *Cumulative deamidations of the major lens protein gammaS-crystallin increase its aggregation during unfolding and oxidation*. Protein Sci, 2020. 903 904 905
106. Kaiser, C.J.O., C. Peters, P.W.N. Schmid, M. Stavropoulou, J. Zou, V. Dahiya, E.V. Mymrikov, B. Rockel, S. Asami, M. Haslbeck, et al., *The structure and oxidation of the eye lens chaperone alphaA-crystallin*. Nat Struct Mol Biol, 2019. **26**(12): p. 1141-1150. 906 907 908
107. Srikanthan, D., O.A. Bateman, A.G. Purkiss, and C. Slingsby, *Sulfur in human crystallins*. Exp Eye Res, 2004. **79**(6): p. 823-31. 909 910
108. Truscott, R.J., *Age-related nuclear cataract-oxidation is the key*. Exp Eye Res, 2005. **80**(5): p. 709-25. 911
109. Shu, D.Y., S. Chaudhary, K.S. Cho, A. Lennikov, W.P. Miller, D.C. Thorn, M. Yang, and T.B. McKay, *Role of Oxidative Stress in Ocular Diseases: A Balancing Act*. Metabolites, 2023. **13**(2). 912 913
110. Aravind, P., B. Rajini, Y. Sharma, and R. Sankaranarayanan, *Crystallization and preliminary X-ray crystallographic investigations on a \beta\gamma-crystallin domain of absent in melanoma 1 (AIM1), a protein from <i>Homo sapiens</i>*. Acta Crystallographica Section F-Structural Biology and Crystallization Communications, 2006. **62**: p. 282-284. 914 915 916 917
111. Pierscionek, B., G. Smith, and R.C. Augusteyn, *The refractive increments of bovine α -, β - and γ -crystallins*. Vision Research, 1987. **27**(9): p. 1539-1541. 918 919
112. Consortium, T.U., A. Bateman, M.-J. Martin, S. Orchard, M. Magrane, S. Ahmad, E. Alpi, E.H. Bowler-Barnett, R. Britto, H. Bye-A-Jee, et al., *UniProt: The Universal Protein Knowledgebase in 2023*. Nucleic Acids Research, 2023. **51**(D1): p. D523-D531. 920 921 922
113. Madeira, F., M. Pearce, A.R.N. Tivey, P. Basutkar, J. Lee, O. Edbali, N. Madhusoodanan, A. Kolesnikov, and R. Lopez, *Search and Sequence Analysis Tools Services from EMBL-EBI in 2022*. Nucleic Acids Research, 2022. **50**(W1): p. W276-W279. 923 924 925
114. Van Rossum, G. and F.L. Drake, *Python 3 Reference Manual*. 2009, Scotts Valley, CA: CreateSpace. 926
115. Crooks, G.E., G. Hon, J.M. Chandonia, and S.E. Brenner, *WebLogo: A sequence logo generator*. Genome Research, 2004. **14**: p. 1188-1190. 927 928
116. Ew, S., B. Ee, B. Jr, C. K, C. J, C. Dc, C. R, F. K, K. C, K. S, et al., *Database Resources of the National Center for Biotechnology Information*. Nucleic acids research, 2022. **50**(D1). 929 930
117. Martin, F.J., M.R. Amode, A. Aneja, O. Austine-Orimoloye, A.G. Azov, I. Barnes, A. Becker, R. Bennett, A. Berry, J. Bhai, et al., *Ensembl 2023*. Nucleic Acids Research, 2023. **51**(D1): p. D933-D941. 931 932
118. Thakur, M., A. Bateman, C. Brooksbank, M. Freeberg, M. Harrison, M. Hartley, T. Keane, G. Kleywegt, A. Leach, M. Levchenko, et al., *EMBL's European Bioinformatics Institute (EMBL-EBI) in 2022*. Nucleic Acids Research, 2023. **51**(D1): p. D9-D17. 933 934 935
119. Harrison, P.W., A. Ahamed, R. Aslam, B.T.F. Alako, J. Burgin, N. Buso, M. Courtot, J. Fan, D. Gupta, M. Haseeb, et al., *The European Nucleotide Archive in 2020*. Nucleic Acids Research, 2021. **49**(D1): p. D82-D85. 936 937
120. Team, R.C., *R: A Language and Environment for Statistical Computing*. 2021, Vienna, Austria: R Foundation for Statistical Computing. 938 939

121. Z, G., G. L, E. R, S. M, and B. B, *Circlize Implements and Enhances Circular Visualization in R*. Bioinformatics (Oxford, England), 2014. **30**(19). 940-941
122. T, G., *Dendextend: An R Package for Visualizing, Adjusting and Comparing Trees of Hierarchical Clustering*. Bioinformatics (Oxford, England), 2015. **31**(22). 942-943
123. Edgar, R.C., *MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput*. Nucleic acids research, 2004. **32**(5). 944-945
124. Kumar, S., G. Stecher, M. Li, C. Knyaz, and K. Tamura, *MEGA X: Molecular evolutionary genetics analysis across computing platforms*. Molecular biology and evolution, 2018. **35**(6). 946-947
125. Saitou, N. and M. Nei, *The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees*. Molecular Biology and Evolution, 1987. **4**(4): p. 406-425. 948-949
126. Felsenstein, J., *Confidence limits on phylogenies: An approach using the bootstrap*. Evolution; International Journal of Organic Evolution, 1985. **39**(4). 950-951
127. Tamura, K., M. Nei, and S. Kumar, *Prospects for inferring very large phylogenies by using the neighbor-joining method*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(30). 952-953
128. Donoghue, P.C.J. and J.N. Keating, *Early vertebrate evolution*. Frontiers in Palaeontology, 2014. **57**(5): p. 879-893. 954
129. Smith, N.C., M.L. Rise, and S.L. Christian, *A comparison of the innate and adaptive immune systems in cartilaginous fish, ray-finned fish, and lobe-finned fish*. Frontiers in Immunology, 2019. **10**: p. 2292. 955-956
130. Kappe, G., A.G. Purkiss, S.T.v. Genesen, C. Slingsby, and N.H. Lubsen, *Explosive expansion of \beta \gamma-crystallin genes in the ancestral vertebrate*. Journal of Molecular Evolution, 2010. **71**: p. 219-230. 957-958
131. Jaenicke, R. and C. Slingsby, *Lens crystallins and their microbial homologs: structure, stability, and function*. Crit. Rev. Biochem. Mol. Biol., 2001. **36**(5): p. 435-499. 959-960
132. Mishra, A., B. Krishnan, S.S. Srivastava, and Y. Sharma, *Microbial $\beta\gamma$ -crystallins*. Prog. Biophys. Mol. Biol., 2014. **115**(1): p. 42-51. 961-962
133. Shimeld, S.M., A.G. Purkiss, R.P.H. Dirks, O.A. Bateman, C. Slingsby, and N.H. Lubsen, *Urochordate $\beta\gamma$ -crystallin and the evolutionary origin of the vertebrate eye lens*. Curr. Biol., 2005. **15**(18): p. 1684-1689. 963-964
134. Wang, J., X. Zuo, P. Yu, I.J. Byeon, J. Jung, X. Wang, M. Dyba, S. Seifert, C.D. Schwieters, A.M. Gronenborn, and Y.X. Wang, *Determination of multicomponent protein structures in solution using global orientation and shape restraints*. Journal of the American Chemical Society, 2009. **131**: p. 10507-10515. 965-966
135. Mahler, B., Y. Chen, J. Ford, C. Thiel, G. Wistow, and Z. Wu, *Structure and dynamics of the fish eye lens protein, γ M7-crystallin*. Biochemistry, 2013. **52**(20): p. 3579-3587. 967-968
136. Basak, A., O. Bateman, C. Slingsby, A. Pande, N. Asherie, O. Ogun, G.B. Benedek, and J. Pande, *High-resolution X-ray crystal structures of human gammaD crystallin (1.25 Å) and the R58H mutant (1.15 Å) associated with aculeiform cataract*. J Mol Biol, 2003. **328**(5): p. 1137-47. 969-970
137. Pettersen, E.F., T.D. Goddard, C.C. Huang, E.C. Meng, G.S. Couch, T.I. Croll, J.H. Morris, and T.E. Ferrin, *UCSF ChimeraX: Structure visualization for researchers, educators, and developers*. Protein Science, 2021. **30**(1): p. 70-82. 971-972
138. Meng, E.C., T.D. Goddard, E.F. Pettersen, G.S. Couch, Z.J. Pearson, J.H. Morris, and T.E. Ferrin, *UCSF ChimeraX: Tools for structure building and analysis*. Protein Science, 2023. **32**(11): p. e4792. 973-974
139. Anderson, D.M., M.G. Nye-Wood, K.L. Rose, P.J. Donaldson, A.C. Grey, and K.L. Schey, *MALDI imaging mass spectrometry of beta- and gamma-crystallins in the ocular lens*. Journal of Mass Spectrometry, 2020. **55**(4): p. 11. 975-976
140. Zhao, H., Y. Chen, L. Rezabkova, Z. Wu, G. Wistow, and P. Schuck, *Solution properties of γ -crystallins: Hydration of fish and mammal γ -crystallins*. Protein Science, 2014. **23**(1): p. 88-99. 977-978
141. Donaldson, P.J., Y. Chen, R.S. Petrova, A.C. Grey, and J.C. Lim, *Regulation of lens water content: Effects on the physiological optics of the lens*. Progress in Retinal and Eye Research, 2023. **95**: p. 101152. 979-980
142. Pierscionek, B.K. and J.W. Regini, *The gradient index lens of the eye: An opto-biological synchrony*. Prog Retin Eye Res, 2012. **31**: p. 332-349. 981-982
143. McMeekin, T.L., M. Wilensky, and M.L. Groves, *Refractive indices of proteins in relation to amino acid composition and specific volume*. Biochem. Biophys. Res. Commun., 1962. **7**: p. 151-156. 983-984
144. Zhao, H., P.H. Brown, and P. Schuck, *On the distribution of protein refractive index increments*. Biophysical Journal, 2011. **100**(9): p. 2309-2317. 985-986
145. Khago, D., J.C. Bierma, K.W. Roskamp, N. Kozlyuk, and R.W. Martin, *Protein refractive index increment Is determined by conformation as well as composition*. Journal of Physics: Condensed Matter, 2018. **30**(43): p. 435101. 987-988

146. Mahendiran, K., C. Elie, J.C. Nebel, A. Ryan, and B.K. Pierscionek, *Primary sequence contribution to the optical function of the eye lens*. Scientific Reports, 2014. **4**: p. Article-number: 5195. 991
992
147. Reid, K.S.C., P.F. Lindley, and J.M. Thornton, *Sulphur-aromatic interactions in proteins*. FEBS Letters, 1985. **190**(2): 993
p. 209-213. 994
148. Kojasoy, V. and D.J. Tantillo, *Impacts of noncovalent interactions involving sulfur atoms on protein stability, structure, folding, and bioactivity*. Organic and Biomolecular Chemistry, 2023. **21**: p. 11-23. 995
996
149. Levenson, R., D.G. DeMartini, and D.E. Morse, *Molecular mechanism of reflectin's tunable biophotonic control: Opportunities and limitations for new optoelectronics*. Apl Materials, 2017. **5**(10). 997
998
150. Weininger, S., M. Neudorf, S. Groger, E. Plato, R. Broneske, K. Saalwachter, U. Weininger, and J. Balbach, *Early Stage UV-B Induced Molecular Modifications of Human Eye Lens gammaD-Crystallin*. Macromol Biosci, 2023. **23**(5): 999
p. e2200526. 1000
1001
151. Gammelgaard, S.K., S.B. Petersen, K.F. Haselmann, and P.K. Nielsen, *Direct Ultraviolet Laser-Induced Reduction of Disulfide Bonds in Insulin and Vasopressin*. Acs Omega, 2020. **5**(14): p. 7962-7968. 1002
1003
152. Jake A Hill, Y.N., Sam Horrell, David von Stetten, Danny Axford, Robin L Owen, Godfrey Beddard, Arwen R Pearson, Helen Mary Ginn, Briony A Yorke, *An ultraviolet-driven rescue pathway for oxidative stress to eye lens protein human gamma-D crystallin*. Commun Chem, 2024. **7**(81). 1004
1005
1006
1007

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content. 1008
1009
1010