

Published in final edited form as:

J Mol Biol. 2011 August 19; 411(3): 680–699. doi:10.1016/j.jmb.2011.06.007.

The molecular refractive function of lens γ -crystallins

Huaying Zhao[&], Patrick H. Brown[#], M. Teresa Magone[&], and Peter Schuck^{&,*}

[&]Dynamics of Macromolecular Assembly Section, Laboratory of Cellular Imaging and Macromolecular Biophysics, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, U.S.A

[#]Biomedical Engineering and Physical Sciences Shared Resource, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, U.S.A

Abstract

γ -crystallins constitute the major protein component in the nucleus of the vertebrate eye lens. Present at very high concentrations, they exhibit extreme solubility and thermodynamic stability to prevent scattering of light and the formation of cataracts. However, functions beyond this structural role have remained mostly unclear. Here, we calculate molecular refractive index increments of crystallins. We show that all lens γ -crystallins have evolved a significantly elevated molecular refractive index increment, which is far above those of most proteins, including non-lens members of the $\beta\gamma$ -crystallin family from different species. The same trait has evolved in parallel in crystallins of different phyla, including in the S-crystallins of cephalopods. A high refractive index increment can lower the crystallin concentration required to achieve a suitable refractive power of the lens, and thereby reduce their propensity to aggregate and form cataract. To produce a significant increase of the refractive index increment, a substantial global shift in the amino acid composition is required, which can naturally explain the highly unusual amino acid composition of γ -crystallins and their functional homologues. This function provides a new perspective for interpreting their molecular structure.

Keywords

crystallin; protein structure function; protein refractive index; excluded volume

Introduction

Crystallins are the major protein component in the cytoplasm of fiber cells in the lens, a tissue where biochemical, thermodynamic, and optical properties of its constituent proteins closely intertwine^{1–3}. In order to refract and focus light onto the retina, crystallins need to be soluble and present in extremely high concentrations for the lens to have a high tissue refractive index above those of the aqueous and vitreous humor. The high protein concentration also serves to provide transparency of the lens due to the suppression of long-range concentration fluctuations that would be a source of light scattering^{4,5}. In order to diminish light scattering further, fiber cells also lose their nucleus and organelles⁶, and

© 2011 Published by Elsevier Ltd.

*address for correspondence: Peter Schuck, NIH, Bldg 13, Rm 3N17, 13 South Drive, Bethesda, MD 20892, U.S.A., Phone: 301-435-1950, Fax: 301-480-1242, schuckp@mail.nih.gov.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

consequently there is no protein turn-over and the crystallin molecules have to last the life-time of the individual organism¹. Light scattering arising from crystallin aggregation, crystallization, as well as liquid-liquid phase transitions, can cause cataracts^{4,7}, which are seemingly unavoidable in human lenses at higher age, and are the leading cause of blindness worldwide⁸.

This poses extraordinary stability and solubility requirements for crystallins, which have been well studied^{1,3,9-14}. A common structural feature within the $\beta\gamma$ -crystallin family is a pair of domains containing two tightly packed Greek key beta sheet motifs. This fold, with additional contributions from domain interactions, provides for an unusually high thermodynamic stability¹⁵. High solubility is thought to be governed by weak inter-particle interactions conferred by surface properties of the proteins^{11,16,17}. However, crystallin functions beyond these structural aspects, as well as the purpose of their unusual amino acid compositions, have remained largely unclear.

Information about possible crystallin functions has been gathered from their molecular evolution, which has been of high interest also in the context of the evolution of the eye¹⁸⁻²¹. Most crystallins appear to have been co-opted from proteins originally serving functions unrelated to eyes^{1,21,22}. The major crystallin families in vertebrates, the ubiquitous α -, β -, and γ -crystallins, have been linked to stress-related proteins¹². α -crystallin is a member of the small heat-shock protein superfamily and can function as a chaperone²³⁻²⁵. β -, and γ -crystallins belong to a superfamily characterized structurally by the Greek key motif²⁶, which is also found, for example, in a spore coat protein S from *Myxococcus xanthus*²⁷, and in an encystment-specific protein spherulin 3A from *Physarum polycephalum*²⁸.

In addition to the ubiquitous crystallins, many vertebrate species have developed taxon-specific crystallins¹. These show a remarkable similarity or are even identical to metabolic enzymes, and often continue to be expressed to lesser extent in other tissues^{29,30}. In some cases, these enzyme crystallins are derived by gene sharing, with continued expression of the same protein in other tissues without loss of the enzymatic function, while in other cases gene duplication took place¹. Examples include δ -crystallin in duck and chicken, derived from argininosuccinate lyase³¹, ϵ -crystallin in birds and crocodiles derived from lactate dehydrogenase B³², ν -crystallin in platypus derived from lactate dehydrogenase A³³, and λ -crystallin in rabbit derived from L-gulonate 3-dehydrogenase³⁴.

Through parallel evolution, many invertebrates have developed camera-like eyes with structures superficially very similar to those of vertebrates, which also exhibit different classes of crystallins³⁵. For example, the main protein component in the lenses of squid and octopus is S-crystallin. Cephalopod S-crystallins form a large and diverse family of proteins, and their evolution has also been studied in detail³⁶. Remarkably, the same strategy of recruitment of metabolic enzymes as lens proteins has been used in cephalopods: S-crystallins have a high similarity to squid glutathione-S-transferase (GST), from which they emerged by gene duplication and now have little or no enzymatic activity³⁷.

This naturally poses again the question which property qualifies certain proteins to be recruited to serve as lens crystallins, and – if gene sharing is followed by gene duplication – which further adaptation to a lens specific function takes place that conflicts with the original function^{33,38,39}. For example, it was recently shown how a chromophore binding protein CRBP was recruited to serve as a taxon-specific crystallin (ι -crystallin) in the gecko, and adapted to bind a suitable chromophore to help cope with intense UV exposure of this species⁴⁰. A similar function has been proposed for some other taxon-specific crystallins⁴¹. Unfortunately, such specific crystallin functions, beyond their purely structural properties

governing thermodynamic stability and solubility, are largely unclear for most crystallins, notably including the ubiquitous vertebrate $\beta\gamma$ -crystallins^{2,3,42}.

Motivated by the consideration that the refraction of light is the most fundamental function of the lens tissue, we have recently embarked on a study of a possible molecular refractive function of crystallins. As a first stage providing biophysical methodology, we have recently developed a computational tool to predict the molecular refractive index increments of proteins, abbreviated dn/dc , from their amino acid composition⁴³, based on the method by McMeekin^{44,45}. Previously, we have applied this to the entire set of predicted proteins of different species, and found that they are normally distributed and fall in a remarkably narrow range⁴³, but with exception of a few protein classes including crystallins. Second, with a biophysical model of macromolecular crowding in the lens tissue, we have explored the possible relevance of variations of molecular refractive indices⁴⁶. Specifically, we have analyzed the thermodynamic consequences of small reductions in protein concentration that become possible from elevated crystallin refractive index increments. We found that, in the most dense lens tissues, even only 5 – 10% higher dn/dc values can, in a non-linear fashion, translate into very significant reductions in the osmotic pressure and in the propensity for aggregation⁴⁶. We hypothesized that this can provide a driving force for the evolution of high refractive index increment crystallins.

The present communication is focused on the detailed structure and function of γ crystallins and their homologues from the perspective of their refractive index increment dn/dc . To this end, we examined the known sequences and evolutionary relationships between members of the different crystallin families. Our results show that the molecular refractive index is consistently and significantly elevated in all those crystallins prevalent in the highest density lens tissues, but not for non-lens crystallins, suggesting the molecular refractive index increment to be a tissue-specific crystallin function. To test this idea, we examined the refractive index properties of unrelated crystallins from other phyla. The results from the available sequence data on S-crystallins indicate that they have continuously evolved towards a high dn/dc . Studying this protein function in more detail, we found that no single amino acids and no single class of amino acids can be held responsible for the elevated dn/dc , which instead is generated by a global shift in amino acid composition. We show in the present paper how these findings can naturally explain the previously noted highly unusual amino acid compositions of crystallins, as well as newly identified anomalies in their composition. Finally, we discuss the implications of this dn/dc function for our understanding of the protein structure of γ crystallins, and the role of the Greek key motif to accommodate a variety of different amino acids.

Results

Refractive index increments of vertebrate crystallins

For light in the visible frequency range, the protein refractive index depends essentially only on its atomic composition and local chemical groups, and it is well-known to be independent of long-range structural details^{47,48}. Accordingly, it is possible to estimate the protein refractive index increment dn/dc from the amino acid composition, using known amino acid refractivities^{43–45}. We applied this method to the protein sequences of the ubiquitous vertebrate crystallins, as shown in Table 1. Figure 1 shows as vertical bars the dn/dc values computed for the human α -crystallin (blue), β -crystallin (green), and γ -crystallin genes (red), superimposed to the dn/dc distribution previously calculated for all predicted human proteins (UCSC genome browser Feb. 2009 (GRCh37/hg19) assembly)⁴³. (Not shown are the dn/dc values determined for different bovine γ -crystallin genes, which range from 0.199 to 0.202 ml/g, in good agreement with the human values and the experimental value of 0.203 measured by Pierscionek *et al.*⁴⁹.)

While values of α -crystallin are close to average, β -crystallins have elevated dn/dc , and those of γ -crystallin are very high. Since the distribution of dn/dc for all predicted proteins – most of which will not have a refractive index function – is close to a Gaussian with an average value of 0.1899 ml/g and a small standard deviation of ± 0.0030 ml/g (with nearly identical distributions obtained from several other species)⁴³, we can calculate a probability that a given dn/dc value will be achieved or exceeded if this were a just random variable. For γ -crystallin, these probabilities all are within the range from $P = 0.01$ to $P = 0.0004$. The dn/dc values computed for various γ M-crystallins from different aquatic vertebrates are still higher (horizontal orange bar), ranging from 0.199 to 0.209 ml/g, corresponding to probabilities $P < 10^{-9}$ for them to be achieved by chance. An extended list including known γ -crystallins from frog, newt, lizard, and several fish species can be found in Supplementary Table 1, consistently showing dn/dc values three standard deviations above the average.

This significantly elevated dn/dc values of all lens β - and γ -crystallins poses the question to what extent the high refractive index increment is an attribute that already existed in the ancient precursor to the $\beta\gamma$ -crystallin family and predisposed it to serve as major lens protein, and/or to what extent this trait has evolved as specialization in the lens. In order to address this, next, we have examined sequences of current non-lens members of the $\beta\gamma$ -crystallin family (Table 1). We found that all non-lens $\beta\gamma$ -crystallins had low dn/dc with an average of 0.1901, nearly exactly identical to the average of all proteins. For example, *Myxococcus xanthus* protein S has a value of 0.187 ml/g. An exception is the recently discovered archaeal M-crystallin⁵⁰, which exhibited a moderately high value of 0.1961 ml/g ($P = 0.02$). Of special interest is the protein likely most closely related to vertebrate $\beta\gamma$ -crystallin, the urochordate Ci- $\beta\gamma$ -crystallin⁵¹, for which we computed a value of 0.1918 ml/g that is well within the spread of dn/dc of average proteins ($P = 0.26$) (see below). This suggests that the high refractive index increment is a trait that likely did not pre-exist but was specifically acquired in conjunction with expression in the lens, to fulfill a specific function.

Among the different crystallins, the ranking of the crystallin dn/dc mimics conspicuously the ranking of the tissue refractive indices where these crystallin molecules are found: A common feature of eye lenses is a radial refractive index gradient that minimizes spherical aberration⁵². A higher refractive index is found in the lens nucleus near the optical axis and lower refractive index in the cortex. This gradient is achieved by a variation of protein concentration and crystallin species^{53–55}. In human and bovine lenses, α and β crystallins are located mostly in the cortex with lower refractive index and lower protein concentration (200 – 300 mg/ml^{56,57}), and α crystallin was found to be excluded from the lens nucleus⁵⁸. As shown in Figure 1 and Table 1, α and β crystallins also have the lowest dn/dc values in the group of ubiquitous crystallins. In contrast, γ crystallins, which exhibit the highest dn/dc values, are most abundant in the center with highest lens refractive index and highest protein concentration (440 – 500 mg/ml and higher^{56,59,60}). In the innermost regions of bovine and toad lenses, it was even found to provide nearly all of the soluble protein^{61,62}. Generally, γ crystallins are also particularly prevalent in hard, high refractive index-lenses which are more closely packed, and relatively dehydrated^{63,64}. In particular, this is true for fish eyes, which lack the refractive power of the cornea provided by the interface with air, and have hard lenses with the highest known protein concentration of any tissue (estimated in excess of 1,000 mg/ml¹²). Correspondingly, as shown in Figure 1 and Table 1, we found that γ M-crystallins of aquatic vertebrates exhibit the highest dn/dc of all.

We also computed the refractive index increments for taxon-specific crystallins (Table 2). In contrast to the lens γ crystallins, their refractive index increment, as well as those of their related enzymes, are only average (Table 2). Thus, enzymes seem not to have been recruited for an amino acid composition generating unusually high dn/dc . One explanation for this

could be that enzyme crystallins continue to be expressed to a lesser extent in other tissue^{29,30}, and that, therefore, constraints from non-lens functions may have prevented an adaptation to develop a high molecular refractive index increment. On the other hand, many taxon-specific crystallins do not constitute major fractions of the total protein content comparable to those of the highly conserved ubiquitous crystallins⁶⁵, and therefore the impact of their refractive index increment on the local tissue refractive index would be minimal. A notable exception is δ -crystallin which replaces γ -crystallin in the soft lenses of birds⁴¹. Interestingly, the comparatively low protein content in these lenses, which might be only between 200 – 300 mg/ml⁶⁶, in conjunction with the low refractive index increment of these δ -crystallins fits well to the observed correlation between molecular and tissue refractive index.

Refractive index increments of invertebrate crystallins

In lieu of an experimental test of the hypothesis that the crystallin molecular refractive index increment has specifically evolved towards higher than average values in high density lens tissue, we next studied crystallins from other phyla, which have evolved independently. Through parallel evolution, some invertebrates have developed camera-like eyes with structures superficially very similar to those of vertebrates⁵² but with very different crystallins. Amino acid compositions of invertebrate crystallins are known for squid, octopus, jellyfish, and scallop species.

S-crystallins represent the major component of the hard lenses of cephalopods. From the refractive index in the core of the cephalopod eye lens of 1.485, we can estimate the protein concentration to be ~760 mg/ml⁶⁷, which is comparable to the total protein concentration in lenses of fish¹. S-crystallins are also thought to have lens (and cornea) specific expression, where they provide 70–90% of all soluble protein⁴¹. Like γ M-crystallins, S-crystallins have acquired extremely high methionine and tyrosine content (see below), and similar to γ -crystallins the specific functions of S-crystallin other than their stability at high concentrations are unknown. S-crystallins form a large and diverse family of proteins, and their evolution has been studied in detail³⁶. Thus, they provide an excellent model system for the present purpose of testing our refractive index hypothesis. The calculated refractive index increments of different S-crystallins and their relatives are shown in Table 3.

In the squid *Loligo opalescens*, there are 24 isoforms of S-crystallins with known sequences (termed Lops) that are 46–99% identical to each other, and exhibit 36–44% sequence identity to squid glutathione-S-transferase (GST), from which they emerged by gene duplication, but now lack enzymatic activity³⁶. We found that the ancestral enzyme, squid GST, has a predicted dn/dc of 0.191 ml/g, similar to those of GST proteins from other organisms (0.192 ± 0.002) ml/g. In contrast, the known sequences of S-crystallins from *Loligo opalescens* have an average value of (0.199 ± 0.002) ml/g. Lops12 has been identified a species of major abundance, does not exhibit GST activity³⁶, and has a dn/dc value of 0.198 ml/g. The S-crystallin species thought to be closest to GST is termed Lops4, exhibits similar enzymatic activity as GST, and has the lowest dn/dc value of 0.195 ml/g.

In order to investigate the relationships between the different Lops molecules and their refractive properties, a phylogenetic tree was constructed with the neighbor-joining method (Figure 2 **Top**). (The resulting tree is very similar to the Bayesian tree reported by Sweeney *et al.*⁶⁸.) It can be discerned that both Lops4 and Lops20 are closest to 1GSQ, with relatively low but elevated dn/dc . Then, the tree branches off into an upper cluster with relatively high dn/dc values that contains the most abundant Lops12 (Lops102 – Lops13), and a lower cluster (Lops16 and below) that exhibits the highest dn/dc values. In order to more directly visualize the relationship between dn/dc and the sequence distance from 1GSQ (which we assume is closest of all to the ancestral precursor enzyme), we plotted dn/dc

versus the sequence distance as extracted from the distance matrix in the software PHYLIP (Figure 2 **Bottom**). A strong correlation between sequence distance and dn/dc can be discerned. Indicated in black are the isoforms in the lower cluster of the tree (Lops16 and below), which have the largest sequence distance. This is consistent with the notion that the advantages from optimizing the refractive index were at least one of the driving forces in the evolution of Lops.

A consistent pattern was found in other squid and octopus species, for which less detail is known (Table 3): For S-crystallins of the squid *Nototodarus gouldi*⁶⁹, high refractive increments (0.198 ml/g) were obtained, although no sequences from intermediates are known. In the squid *Ommastrephes sloani*, there are four known members of S-crystallins that exhibit 42–44% sequence identity to squid glutathione-S-transferase (GST)³⁷. Their average dn/dc is (0.196 ± 0.002) ml/g. Mirroring the findings from *Loligo opalescens*, in *Ommastrephes sloani* the most abundant and putatively most recent member, SL20 (exhibits 80% sequence identity to Lops12) has no enzymatic activity and exhibits a value of 0.199 ml/g. One difference between GST and most S-crystallins are repeats of an insert of up to 193 residues with several repeats of RGDGGYXVQG/S³⁷, a sequence with relatively high value of 0.196 ml/g. The S-crystallin species thought to be closest to GST and orthologous to Lops4 is termed SL11, and has similarly high enzymatic activity. It does not have an insert, and is thought to be an evolutionary intermediate between GST and the fully developed SL20³⁶. We found it to have the lowest dn/dc value at 0.195 ml/g, close to the value of squid GST. The correlation between decrease of enzyme activity and increase in dn/dc implies that the lens-specific S-crystallins in the two types of cephalopods have undergone the adaptation into structural proteins from their ancestor to provide high refractivity in eye.

S-crystallin sequences have also been determined for two octopus species^{70,71}. Like in squid, they are related to GST (20–25% sequence similarity⁷⁰), and exhibit a 48–58% sequence similarity with squid SL20 and 42–80% similarity between the octopus species⁷¹. We calculated a high dn/dc , with values of ~0.197 ml/g (Table 3). In addition, octopus has taxon-specific crystallins in low abundance that show only slightly elevated dn/dc values (Table S3), including Ω -crystallin related to aldehyde dehydrogenase (~10% of lens proteins)^{70,72}, as well as an O-crystallin (3%-5% of lens proteins)⁷³. An orthologue of O-crystallin is also found in the soft lens of scallop⁷⁴, but with an only average dn/dc of 0.190 ml/g. Although it is the sole crystallin in the scallop, it is also found in significant amounts in non-ocular tissue⁷⁴.

In a different phylum, box jellyfish also have sophisticated eyes with relatively high refractive index lenses⁷⁵. Their crystallins are J-crystallins, which may also serve some non-lens function³⁰. They have a 43–50% sequence similarity to saposins⁷⁶ (which have a predicted dn/dc of 0.188 ml/g), and 36% identity to the ADP-ribosylation enzyme SelJ⁷⁷ (with 0.192 ml/g). The average dn/dc of the known J-crystallins was calculated to be 0.195 ml/g.

In summary, our results show that the development of a high dn/dc trait of the vertebrate γ -crystallins from low dn/dc non-lens precursors seems to be repeated independently in different evolutionary lineages. This signifies a functional role of the molecular refractive index in lens crystallins.

Amino acid composition

It is interesting to study the protein amino acid composition and structure from this perspective. As mentioned above, the overall distribution of protein dn/dc values is very narrow with a standard deviation of 0.003 ml/g. This is due to the fact that the amino acid

refractivities are little correlated with their chemical properties⁴³. Amino acids with high polarizability and refractive index increment are those containing aromatic rings, sulfur, or double-bonds in the R-group, the highest ones being tryptophan, phenylalanine, tyrosine, histidine, cysteine, arginine, and methionine. However, for achieving a high refractive index increment through amino acid substitution, the most drastic single amino acid change possible (a substitution of a proline to a tryptophan) would change the γ -crystallin dn/dc only by 0.0009 ml/g, and most single substitutions would change it much less (for example, a single lysine to arginine substitution would change dn/dc by only 0.0002 ml/g). This is small compared to the difference in refractive index of γ -crystallins to average proteins, which can be achieved only by a major overall change towards highly unusual composition.

The amino acid composition from the perspective of its impact on dn/dc is illustrated in Figure 3 **Top Panel** for several vertebrate γ -crystallins in comparison with the putative ancient non-lens relative *M. Xanthus* protein S. In the abscissa of Figure 3 are shown all amino acids, ranked by their refractivity. The ordinate shows the relative abundance of each amino acid (in %) as a difference to the normally expected percentage of each amino acid, the latter determined from the set of all predicted human proteins (UCSC Feb. 2009 (GRCh37/hg19) assembly). In order to visualize how unusual a certain deviation from the average abundance is, their full frequency distribution of % content for each amino acid is shown as a color contour plot in the background. It highlights in red to white the extremely rare percentages for each amino acid (frequencies of lower than 0.05 and higher than 0.95 of the cumulative frequency distribution, as shown in the color bar). Examples of the full frequency distribution of % abundance and the observed percentages for individual amino acids are shown in the insets.

In this representation, we can discern several aspects of the origin of the high dn/dc of vertebrate γ -crystallins. First, the fact that the predominant population of the symbols are in the lower left and upper right quadrant of this plot shows that there is not just one or two specific amino acids that are in high abundance and serve to alter the dn/dc , but rather there is an overall shift in amino acid composition. Second, the elevation of the relative abundance of highest dn/dc residues, such as aromatic amino acids, cysteine, arginine, and methionine, is accompanied by the elimination of amino acids with low refractivity, such as alanine, valine, threonine, and leucine. Third, the comparison with the natural distribution of these amino acids' abundance in human proteins indicates that several of the deviations are very significant. For example, for human γ B-crystallin highlighted by the red line, this can be discerned from the data points being located in the red to white bands. In contrast, the putative ancient relative *M. Xanthus* protein S (magenta line), even though it does not have a perfectly average content of all amino acids, exhibits examples for both high contents of some low dn/dc amino acids as well as low content of high dn/dc amino acids.

This pattern is exacerbated in fish, which has the highest density lenses and the highest dn/dc crystallins. For example, highlighted in Figure 3 **Top Panel** by the green line are the data from lip shark γ M2. In fact, several fish γ -crystallins are very extreme in that they contain no alanine, lysine, and/or threonine at all, but unusually high fractions of phenylalanine, tyrosine, arginine, and methionine. The most extreme example is the Mexican tetra, which has very few low dn/dc residues – a single alanine, a single leucine, no lysine and no threonine, but extremely high numbers of dn/dc ones with 4.2 % tryptophan, 10.2% tyrosine, 10.8% arginine, and 12.6% methionine. In a similar presentation, Figure 3 **Bottom Panel** shows the deviation from average amino acid composition for the different human γ -crystallins (blue symbols) in comparison with their closest known ancient relative, the urochordate Ci- β γ -crystallin (magenta) which has a dn/dc value of 0.1918 ml/g. In this plot, γ S is highlighted (blue line), which was found to be expressed in the cortex⁷⁸ and has a dn/dc

dc value of 0.1983 ml/g, as well as γ D (green line), which is a major component of the lens nucleus with a dn/dc value of 0.1992 ml/g.

For Lops proteins, a pattern emerges similar to that of vertebrate γ -crystallins (Figure 4). While squid GST shows relatively evenly distributed values, Lops crystallins have systematically lower fractions of low dn/dc residues, especially serine and leucine, accompanied by high fractions of high dn/dc residues, such as asparagines, methionine, tyrosine, and phenylalanine. Further, one can discern that the evolution from the most GST-like Lops4 to the currently Lops12 has caused, with few exceptions, the systematic exacerbation of this pattern.

It is interesting to look more quantitatively at the contributions of the abundances of each amino acid in the crystallin sequence towards the cumulative elevation of the crystallin dn/dc above that of average proteins. As outlined in the Methods section Eqs. 1 – 3, dn/dc is not simply additive. Nevertheless, we can calculate the change in dn/dc for a hypothetical sequence where, separately for each amino acid, their relative abundance in crystallin is adjusted to match that of average proteins. This will quantify how important the contribution of that amino acid is towards maintaining the elevated crystallin dn/dc . The results are listed in Table 4, normalized by the total difference in dn/dc between the crystallin and the average protein. The values in parenthesis quantify how, from a background of an average protein, the average dn/dc will increase if a particular amino acid abundance is adjusted to the level occurring in the crystallin. Both viewpoints provide slightly different numbers, especially for those amino acids with dn/dc in between those of the average protein and crystallin (which could potentially play a role an evolutionary path).

From either perspective, the data highlight that a global shift in amino acid composition is required for achieving the high molecular refractive index increment of lens crystallins. Although the tyrosine residues make the biggest contribution, they provide little more than a third of the refractive index elevation of human γ B, and substantially less in the other species examined. In contrast, the cumulative contribution from the relative lack of low dn/dc residues is larger than the effect from tyrosine or any other high dn/dc residues. For example, in the γ M from Mexican tetra the extremely low leucine content alone has almost the same effect on dn/dc as the elevated tyrosine content.

Refractive index increments and protein structures

Even though the refractive properties are independent of the three-dimensional structure, it is of interest to inspect the location of the amino acids that are particularly enriched in lens crystallins. To this end, Figure 5 shows a structural comparison of human γ B-crystallin and the homologous spore coat protein from *M. Xanthus* protein S⁷⁹, highlighting the high dn/dc amino acids Met, Arg, Cys, Tyr, and Trp, which are significantly more abundant in the lens protein than in its ancient relative. It is well possible that the increased fraction of at least some of these high dn/dc residues could additionally support other essential functions. In particular, their surface accessibility is significant, as the surface characteristics of γ crystallin will govern inter-particle interactions and solubility. (For simplicity, in Figure 5, we do not show the low dn/dc residues that are more abundant in protein S, the lack of which would be difficult to attribute an alternative function in γ B-crystallin; see below).

In human γ B-crystallin, there are 19 arginines compared to 6 in Protein S (Figure 5, magenta). They are indeed preferentially located at the surface (which may be discerned from the fact that they are ~11% of total amino acids providing ~21% of the total protein surface area), and therefore could fulfill dual roles in contributing both to surface characteristics and dn/dc . However, it appears that the high dn/dc residues overall are rather evenly distributed between surface and interior, contributing to the solvent accessible area

approximately in proportion of their relative abundance, representing 35% of the total amino acids and providing 38% of solvent accessible area in human γ B-crystallin. In particular, the four tryptophan residues of human γ B-crystallin are nearly entirely buried¹⁰, and most of the 15 tyrosine residues are largely solvent inaccessible, and are therefore difficult to imagine contributing significantly to solubility and interaction properties.

A high abundance of methionine in γ -crystallin of aquatic vertebrates has been proposed to fulfill a role in mediating intermolecular interactions that aid in dense packing⁸⁰. Similarly, there are many surface accessible arginine and methionine residues in the S-crystallins of cephalopods. However, a homology model of OctS3⁸¹ shows methionine contributing approximately 10% of the total amino acids while also providing ~10% of the surface area (see Supplemental Figure 1), the even distribution suggesting that they may at least not have an exclusive role in mediating surface properties, consistent with their role in aiding the molecular refractive index, a function for which pragmatically the location of the residues within the three-dimensional protein structure is not relevant.

A more detailed structural comparison of the amino acid replacements from lens and non-lens crystallin is possible from a structure-based sequence alignment of human γ B-crystallin (2JDF) and its closest known ancient relative, the $\beta\gamma$ -crystallin from *C. intestinalis* (2BV2)⁵¹. The high degree of sequence similarity (49%) and structural similarity⁵¹ makes it possible to ask where individual amino acid substitutions occurred that led to the increase from low dn/dc of 0.1918 ml/g in *C. intestinalis* to the high value of 0.1988 ml/g. To this end, Figure 6 shows a structural alignment and a structure-based sequence alignment (γ B-crystallin in yellow, *C. intestinalis* $\beta\gamma$ -crystallin in grey). It is interesting here that the set of identical residues alone (marked blue in the γ B-crystallin sequence and structure) are of a composition with predicted dn/dc value of 0.201 ml/g. Elevation of the dn/dc of the complete protein for the lens crystallin is accomplished by substitution of the majority of remaining residues replacing low with high dn/dc residues (marked in red). In our view, these substitutions are not systematically coincident with structural features.

Discussion

Our data show that, independently in different phyla, lens crystallins from the highest density tissues have specifically evolved a significantly elevated molecular refractive index. The observation of an elevated protein refractive index may be unexpected considering the wide-spread assumption that dn/dc of all proteins can be considered in good approximation to be identical, with values expected to be between 0.18 – 0.19 ml/g^{82–84}. However, this conventional wisdom has been questioned previously by Pierscionek *et al.*, who have measured an unusually high value for bovine γ -crystallin of 0.203 ml/g⁴⁹. Consistent with this, a detailed bioinformatics analysis of the distribution of refractive index increments of predicted proteins from different species generally confirmed the narrow spread of values for most proteins, but with distinct exceptions, including the $\beta\gamma$ -crystallin family⁴³. Another class of proteins with highly elevated dn/dc are reflectins⁴³, which were identified in reflector platelets of squid light organ reflector tissues that serve to provide thin film interference^{85,86}.

Since the increased crystallin refractive index could only be achieved by virtue of a substantial, global shift in amino acid composition, and yet, the relative increase over average protein generally amounts to 10% or less, one can suspect that a high dn/dc must confer a substantial evolutionary advantage. The question arises in which way such a molecular refractive index function would be helpful. We believe the answer to this can be found in the correlation of high dn/dc with the intracellular protein concentration in the lens tissue, and in the observation that with crystallin molecules of higher refractive index

increment, the same lens refractive index can be achieved with a proportionally lower concentration of crystallins.

In a related work we have explored this question from a theoretical perspective⁴⁶. We have shown with a model of macromolecular crowding of non-interacting hard spheres, that in the extreme environment of fiber cells (which in the center of high density lenses may have total protein concentrations 500 mg/ml and above) even a small reduction of protein concentration afforded by the higher macromolecular dn/dc can have profound thermodynamic consequences. With average free distances between crystallins only being approximately half their radius, the cytosol is highly thermodynamically non-ideal. Further increasing the protein concentration is associated with a steeply increasing osmotic pressure and steeply increasing chemical activity and energetic cost to keep the proteins soluble. Using scaled particle theory to capture the effect of volume exclusion in this solution, we have calculated the saving in osmotic pressure from lower protein concentrations that become possible with a 6.5% higher dn/dc crystallin (at the same tissue refractive index). This relative reduction in osmotic pressure is shown in Figure 7 **Top** as a function of total protein lenticular protein concentration. In the highest density lenses, it can amount to 20 – 25%. Similarly, we have calculated the reduction in the chemical activity (Figure 7 **Bottom Inset**), and – in a thermodynamic model originally introduced in the context of hemoglobin sickling^{87,88} – the increase in the stability of crystallin to be soluble relative to it being in a polymer or crystallin phase (Figure 7, **Bottom**). In lenses of greater than 500 mg/ml total protein, this can amount to more than 1 kcal/mol. Although the thermodynamic consequences of volume exclusion are obligatory and will qualitatively affect all species, interactions between different protein species (which were not accounted for in our model⁴⁶) can further modulate their stability in the crowded environment¹⁷.

The salient aspect of these calculations is that, because at higher concentrations it becomes increasingly more difficult to find sufficient space for accommodating more proteins, small changes in the actual protein concentration can be strongly non-linearly amplified into very significant changes in thermodynamic parameters governing reactivity and aggregation. Since the formation of crystalline or polymer phases leads to light scattering and cataract, as shown by many forms of cataract caused by point mutants of γ crystallin that cause reduced solubility and/or crystallization^{89–95}, the improved thermodynamic stability of the soluble state afforded by molecules with higher dn/dc (and at correspondingly lower concentration) could well explain the evolutionary driving force for the adaptation to high dn/dc of crystallins prevalent in the highest density lenses. This hypothesis can explain the correlation between molecular refractive index increment and the tissue density of the ubiquitous vertebrate crystallins, as well as the distribution of different crystallins between lens cortex and lens nucleus observed in several vertebrate species. Further, it is consistent with the absence of adaptation to high dn/dc of taxon-specific crystallins in low density lens tissues (as in the lenses of birds as well as in the cortex of human lenses)⁴⁶Footnote 1

Footnote 1 Only for squid proteins, at first sight our hypothesis appears to be in conflict with the report by Sweeney *et al.*⁶⁸, who have measured the radial concentration distribution of ‘long loop S-crystallins’ (which are those molecules in lower branch in our tree Figure 2 **Top**, indicated by black dots in Figure 2 **Bottom**), and reported them to be located preferentially in the periphery with lower concentrations and to be ‘nearly absent in the centre’. According to our data, these are the isoforms with highest molecular refractive index increment (Figure 2 **Bottom**) (not to be confused with the tissue refractive index discussed by⁶⁸). We cannot resolve this problem, since, unfortunately, no absolute values for the protein concentration measured by Sweeney *et al.* were reported in the paper⁶⁸. On the other hand, no real conflict may exist: In our reading, the MALDI-TOF data shown in the Supplement of⁶⁸ do not allow a clear localization of the molecules; contrary to the conclusions proposed by Sweeney and co-workers, they seem to show tryptic digest peptides stemming from many of the high dn/dc isoforms being present also in the innermost lens regions (including Lops5, Lops7, Lops23, Lops18, Lops10, Lops25, Lops16, and Lops8), even with most of these being present in the SDS gel band increasing in density towards the center of the lens (including Lops5, Lops7, Lops23, Lops18, and Lops16).

From our data it is evident that substitution of a single or a few amino acid residues is insufficient to generate significant increase in dn/dc , but requires an overall shift, replacing many low dn/dc residues with many residues of above-average dn/dc . (This is evident, similarly, in the amino acid composition of reflectins, which are highly enriched in sulfur-containing and aromatic amino acids and contain no alanine, isoleucine, leucine, and lysine⁸⁵.) This provides a natural explanation for the puzzle of an unusual amino acid composition in γ -crystallins and cephalopod S-crystallins that have been noted previously by several groups: Vertebrate γ -crystallins exhibit an unusually high tyrosine content, conserved in most vertebrate γ -crystallins⁴¹, and an unusually high number of cysteine residues⁸⁰. Similarly, an unusually high fraction of arginine relative to lysine in human γ -crystallins⁴¹ would be naturally explained by the higher refractive index increment of the former (0.206 ml/g) compared to that of the latter (0.181 ml/g). The present work led to the identification of another strong anomaly in crystallins, which is the relative lack of alanine, threonine, valine, and leucine.

A prominently noted anomaly in the amino acid composition of γ -crystallin is an extremely high content of methionine residues in γ M-crystallins of fish (e.g., 15% in catfish)^{96,97}, mirrored in S-crystallins of cephalopods^{69,70}, which also have very high fraction of arginine and methionine residues (11.6% and 10.7%, respectively). The purpose of the high methionine content in fish γ -crystallin is unknown, and was speculated to contribute to protein stability⁹⁷, aid solubility and cold adaptation⁹⁸, and to lead to an increased refractive index increment⁶⁵. From the present work, it seems that high methionine content would be suitable to contribute to an increase dn/dc for most proteins.

On the other hand, it is interesting to consider 'extreme' crystallins such as γ M-crystallin from Mexican tetra or Lops22. They have overall refractive index increments similar or in excess of that of a hypothetical polypeptide of methionine (0.204 ml/g). Therefore, the methionine residues do actually not appear to be very helpful anymore, or might even be detrimental, for achieving dn/dc -values in excess of 0.204 ml/g. This points to the fact that more crucial for extremely high dn/dc proteins is the balance between the aromatic amino acids and the elimination of low dn/dc residues (compare Table 4). However, at some point it seems certain that other requirements for γ -crystallins, such as high thermal stability and solubility and low hydration, would be conflicting with such optimization. In this regard, reflectins are noteworthy in that they have a dn/dc substantially higher than crystallins, but exhibit strong self-assembly⁸⁶ that would be most detrimental for a crystallin. The fact that even single amino acid replacements have been identified as the cause of altered phase transition leading to the formation of cataracts^{91,93,99} highlights the delicate balance of forces for maintaining solubility.

The most effective residues for increasing dn/dc appear to be tryptophan and tyrosine. The highly elevated tyrosine content does currently not seem well justified by their UV absorption property, considering that most of the UV-B radiation at wavelengths < 300 nm is filtered out already in the cornea and anterior chamber¹⁰⁰, and that at wavelengths > 295 nm the tyrosine residues do not absorb much. Tryptophan absorption extends to higher wavelengths, and it has been shown that the γ -crystallin structure allows for a very efficient quenching by charge transfer, likely as a protection from photodamage^{101,102}. One could speculate that this would make the acquisition of additional tryptophan residues for the purpose of elevating dn/dc structurally complicated. Generally, photodamage might represent less of a constraint for most aquatic species.

Structurally, the shift in amino acid composition from an ancient precursor protein necessary to produce high dn/dc γ -crystallins poses the problem of accommodating many replacements, the location of which is irrelevant for the refractive index property, but

without compromising the extraordinarily high thermal stability of the Greek key domains¹². This is highlighted, for example, in Figure 6. Already Blundell and colleagues⁹ have remarked on the surprisingly high degree of variability how side chains in the hydrophobic core of these domains can pack while maintaining close three-dimensional homology. Such sequence variability is mirrored also by non-lens members of the $\beta\gamma$ -crystallin family¹⁰³. Thus, one could speculate that this flexibility of the Greek key motif to accommodate different residues lends itself well as a stable scaffold for a lens protein, providing the opportunity of replacing residues of low dn/dc with high ones. In fact, this property has recently been exploited to engineer new proteins^{104,105}, such as affilins from combinatorial libraries randomizing eight solvent-exposed amino acid residues of γ B-crystallins.

In summary, the most compelling explanation for the unusual amino acid composition of crystallin molecules, specifically from the highest refractive index tissue, is a molecular refractive index function. The same trait has evolved in parallel across different phyla, consistent with the notion of universal physical constraints arising in protein solutions at extremely high concentrations, which make solubility easier to maintain at slightly lower concentrations afforded by higher refractive index molecules. Such molecules have evolved by directed amino acid substitutions without altering the three dimensional structure. This shows for the first time that a protein refractive index may be more than a coincidental property, and that the refractive function of the eye lens is manifest on a molecular level.

Methods

Sequence-based analysis of the protein refractive index increment

The experimental measurement of a protein dn/dc requires tens of mg of highly pure and soluble material⁴³, which would pose formidable difficulties for a systematic study of many crystallins using either bacterial over-expression or purification from biological samples. Therefore, the computational prediction of the protein refractive index increment from the amino acid sequence was adopted in the present work, following the method of McMeekin^{44,45}, as recently described⁴³. In brief, the refractivity per gram of protein, R_p , is calculated from its amino acid composition as the weight average

$$R_p = \frac{\sum_a R_a M_a}{\sum_a M_a} \quad (\text{Eq. 1})$$

of the contributions of the amino acid refractivities R_a , experimentally measured and tabulated for 589 nm and 25 °C in⁴⁵ and reproduced in⁴³. The data are supported by ab initio calculations of mean residue polarizabilities^{106–108}. M_a is the residue molecular weight. Analogously, the protein partial specific volume v_p is predicted as a weight-average from composition, using the tables from Cohn & Edsall¹⁰⁹. The Lorentz-Lorenz formula then predicts the protein refractive index as

$$n_p = (2R_p + \bar{v}_p) / (\bar{v}_p - R_p)^{1/2}, \quad (\text{Eq. 2})$$

, and the refractive index increment was obtained with the Wiener equation for dilute solutions as

$$(dn/dc)_p = 1.5\bar{v}_p n_0 (n_p^2 - n_0^2) / (n_p^2 + 2n_0^2) \quad (\text{Eq. 3})$$

¹¹⁰ (with solvent refractive index n_0 taken as 1.3340, corresponding to water at 25 °C with 150 mM sodium chloride).

This calculation assumes unconjugated polypeptides, without any preferential hydration. Appropriate corrections could in principle be applied for contributions from preferential solvation, post-translational modifications, temperature effects, and solution refractive index ^{111–114}. The validity of the computational method was verified experimentally by refractometry and dry weight measurement of common proteins ^{44,45} and peptides specifically constructed to span a large refractive index range ⁴³.

Sequence alignment and phylogenetic analysis of Lops proteins

An initial global sequence alignment was performed with CLUSTAL W2 ¹¹⁵ on all known sequences of GST from *O. pacificus* as well as the S-crystallins and GST from *L. opalescens* found in the NCBI protein database. The BLOSUM model was chosen for the sequence alignment option. Next, a distance matrix was calculated using the protdist program of the PHYLIP package version 3.57c (kindly provided by Dr. J. Felsenstein), using the Jones-Taylor-Thornton substitution matrix ¹¹⁶. The NEIGHBOR program of PHYLIP was then used to construct a neighbor-joining tree ¹¹⁷. A bootstrap analysis with 500 replicates was performed to determine the levels of support for the interior nodes, and a consensus tree was constructed.

Protein structure analysis

For the crystal structure analysis, the solvent accessibility was calculated with the program AREAIMOL in the CCP4 Suite ¹¹⁸. Graphics of protein structures were created with the software PyMOL.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Gu-Gang Chang for providing the coordinates from the homology model of octopus S-crystallin, Dr. Andor Kiss for providing some protein sequences, and Dr. Graeme Wistow for critically reading an initial draft of the manuscript. This work was supported by the Intramural Research Program of the National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health.

Literature Cited

1. Wistow GJ, Piatigorsky J. Lens crystallins: the evolution and expression of proteins for a highly specialized tissue. *Annu Rev Biochem.* 1988; 57:479–504. [PubMed: 3052280]
2. Zigler, JS. Lens proteins. In: Albert, DM.; Jakobiec, FA., editors. *The Principles and Practice of Ophthalmology: Basic Sciences.* W.B. Saunders Co; Philadelphia, PA: 1994. p. 97-112.
3. Bloemendal de Jong W, Jaenicke R, Lubsen NH, Slingsby C, Tardieu A. Ageing and vision: structure, stability and function of lens crystallins. *Progr Biophys Mol Biol.* 2004; 86:407–485.
4. Benedek GB. Theory of transparency of the eye. *Appl Optics.* 1971; 10:459–473.
5. Delaye M, Tardieu A. Short-range order of crystallin proteins accounts for eye lens transparency. *Nature.* 1983; 302:415–7. [PubMed: 6835373]

6. Bassnett S. On the mechanism of organelle degradation in the vertebrate lens. *Exp Eye Res.* 2009; 88:133–9. [PubMed: 18840431]
7. Graw J. Genetics of crystallins: cataract and beyond. *Exp Eye Res.* 2009; 88:173–89. [PubMed: 19007775]
8. WHO. Fact Sheet No 282. Vol. 2011. WHO; 2009. Visual impairment and blindness.
9. Blundell T, Lindley P, Miller L, Moss D, Slingsby C, Tickle I, Turnell B, Wistow G. The molecular structure and stability of the eye lens: x-ray analysis of gamma-crystallin II. *Nature.* 1981; 289:771–7. [PubMed: 7464942]
10. Wistow G, Turnell B, Summers L, Slingsby C, Moss D, Miller L, Lindley P, Blundell T. X-ray analysis of the eye lens protein gamma-II crystallin at 1.9 Å resolution. *J Mol Biol.* 1983; 170:175–202. [PubMed: 6631960]
11. Tardieu A, Veretout F, Krop B, Slingsby C. Protein interactions in the calf eye lens: interactions between beta-crystallins are repulsive whereas in gamma-crystallins they are attractive. *Eur Biophys J.* 1992; 21:1–12. [PubMed: 1516556]
12. Jaenicke R, Slingsby C. Lens crystallins and their microbial homologs: structure, stability, and function. *Crit Rev Biochem Mol Biol.* 2001; 36:435–99. [PubMed: 11724156]
13. Pande A, Annunziata O, Asherie N, Ogun O, Benedek GB, Pande J. Decrease in protein solubility and cataract formation caused by the Pro23 to Thr mutation in human gamma D-crystallin. *Biochemistry.* 2005; 44:2491–500. [PubMed: 15709761]
14. Mills IA, Flaugh SL, Kosinski-Collins MS, King JA. Folding and stability of the isolated Greek key domains of the long-lived human lens proteins gammaD-crystallin and gammaS-crystallin. *Protein Sci.* 2007; 16:2427–44. [PubMed: 17905830]
15. Jaenicke R. Stability and folding of domain proteins. *Progr Biophys Mol Biol.* 1999; 71:155–241.
16. Purkiss AG, Bateman OA, Wyatt K, Wilmarth PA, David LL, Wistow GJ, Slingsby C. Biophysical properties of gammaC-crystallin in human and mouse eye lens: the role of molecular dipoles. *J Mol Biol.* 2007; 372:205–22. [PubMed: 17659303]
17. Dorsaz N, Thurston GM, Stradner A, Schurtenberger P, Foffi G. Phase separation in binary eye lens protein mixtures. *Soft Matter.* 2010; 7:1763–1776.
18. Darwin, C. On the origin of species. Harvard University Press; Cambridge, MA: 1859.
19. De Jong, W. Evolution of lens and crystallins. In: Bloemendal, H., editor. *Molecular and cellular biology of the eye lens.* Wiley; New York: 1981. p. 221-278.
20. Fernald RD. Casting a genetic light on the evolution of eyes. *Science.* 2006; 313:1914–8. [PubMed: 17008522]
21. Weadick CJ, Chang BS. Molecular evolution of the betagamma lens crystallin superfamily: evidence for a retained ancestral function in gamma N crystallins? *Mol Biol Evol.* 2009; 26:1127–42. [PubMed: 19233964]
22. Riyahi K, Shimeld SM. Chordate betagamma-crystallins and the evolutionary developmental biology of the vertebrate lens. *Comp Biochem Physiol B Biochem Mol Biol.* 2007; 147:347–57. [PubMed: 17493858]
23. Ingolia TD, Craig EA. Four small *Drosophila* heat shock proteins are related to each other and to mammalian alpha-crystallin. *Proc Natl Acad Sci U S A.* 1982; 79:2360–4. [PubMed: 6285380]
24. Horwitz J. Alpha-crystallin can function as a molecular chaperone. *Proc Natl Acad Sci U S A.* 1992; 89:10449–53. [PubMed: 1438232]
25. Acosta-Sampson L, King J. Partially folded aggregation intermediates of human gammaD-, gammaC-, and gammaS-crystallin are recognized and bound by human alphaB-crystallin chaperone. *J Mol Biol.* 401:134–52. [PubMed: 20621668]
26. Lubsen NH, Aarts HJ, Schoenmakers JG. The evolution of lenticular proteins: the beta- and gamma-crystallin super gene family. *Prog Biophys Mol Biol.* 1988; 51:47–76. [PubMed: 3064189]
27. Wistow G, Summers L, Blundell T. Myxococcus xanthus spore coat protein S may have a similar structure to vertebrate lens beta gamma-crystallins. *Nature.* 1985; 315:771–3. [PubMed: 3925350]
28. Wistow G. Evolution of a protein superfamily: relationships between vertebrate lens crystallins and microorganism dormancy proteins. *J Mol Evol.* 1990; 30:140–5. [PubMed: 2107329]

29. Li X, Wistow GJ, Piatigorsky J. Linkage and expression of the argininosuccinate lyase/delta-crystallin genes of the duck: insertion of a CR1 element in the intergenic spacer. *Biochim Biophys Acta*. 1995; 1261:25–34. [PubMed: 7893758]
30. Piatigorsky J. Gene sharing, lens crystallins and speculations on an eye/ear evolutionary relationship. *Integr Comp Biol*. 2003; 43:492–499. [PubMed: 21680457]
31. Wistow GJ, Piatigorsky J. Gene conversion and splice-site slippage in the argininosuccinate lyases/delta-crystallins of the duck lens: members of an enzyme superfamily. *Gene*. 1990; 96:263–70. [PubMed: 2269436]
32. Wistow GJ, Mulders JW, de Jong WW. The enzyme lactate dehydrogenase as a structural protein in avian and crocodylian lenses. *Nature*. 1987; 326:622–4. [PubMed: 3561501]
33. van Rheede T, Amons R, Stewart N, de Jong WW. Lactate dehydrogenase A as a highly abundant eye lens protein in platypus (*Ornithorhynchus anatinus*): epsilon (epsilon)-crystallin. *Mol Biol Evol*. 2003; 20:994–8. [PubMed: 12716980]
34. Mulders JW, Hendriks W, Blankesteyn WM, Bloemendal H, de Jong WW. Lambda-crystallin, a major rabbit lens protein, is related to hydroxyacyl-coenzyme A dehydrogenases. *J Biol Chem*. 1988; 263:15462–6. [PubMed: 3170592]
35. Piatigorsky J. Evolution of mollusc lens crystallins: Glutathione S-transferase/S-crystallins and aldehyde dehydrogenase/Omega-crystallins. *Amer Malac Bull*. 2008; 26:73–81.
36. Tomarev SI, Chung S, Piatigorsky J. Glutathione S-transferase and S-crystallins of cephalopods: evolution from active enzyme to lens-refractive proteins. *J Mol Evol*. 1995; 41:1048–56. [PubMed: 8587103]
37. Tomarev SI, Zinovieva RD, Piatigorsky J. Characterization of squid crystallin genes. Comparison with mammalian glutathione S-transferase genes. *J Biol Chem*. 1992; 267:8604–12. [PubMed: 1373730]
38. Piatigorsky J, Wistow G. The recruitment of crystallins: new functions precede gene duplication. *Science*. 1991; 252:1078–9.
39. Wistow G. Lens crystallins: gene recruitment and evolutionary dynamism. *Trends Biochem Sci*. 1993; 18:301–6. [PubMed: 8236445]
40. Werten PJ, Roll B, van Aalten DM, de Jong WW. Gecko iota-crystallin: how cellular retinol-binding protein became an eye lens ultraviolet filter. *Proc Natl Acad Sci U S A*. 2000; 97:3282–7. [PubMed: 10725366]
41. Wistow, G. Molecular biology and evolution of crystallins: Gene recruitment and multifunctional proteins in the eye lens. R.G. Landes Company; Georgetown, TX: 1995.
42. Bhat SP. Transparency and non-refractive functions of crystallins—a proposal. *Exp Eye Res*. 2004; 79:809–16. [PubMed: 15642317]
43. Zhao, H.; Brown, PH.; Schuck, P. *Biophys J*. Vol. 100. 2011. On the distribution of protein refractive index increments; p. 2309-2317.
44. McMeekin, TL.; Groves, ML.; Hipp, NJ. Refractive indices of amino acids, proteins, and related substances. In: Stekol, J., editor. *Amino Acids and Serum Proteins*. American Chemical Society; Washington DC: 1964.
45. McMeekin TL, Wilensky M, Groves ML. Refractive indices of proteins in relation to amino acid composition and specific volume. *Biochem Biophys Res Commun*. 1962; 7:151–156.
46. Zhao H, Magone MT, Schuck P. The role of macromolecular crowding in the evolution of lens crystallins with high molecular refractive index. *Phys Biol*. 2011; 8:046004. [PubMed: 21566271]
47. Tanford, C. *Physical Chemistry of Macromolecules*. Wiley; New York: 1961.
48. Born, M.; Wolf, E. *Principles of Optics*. 7. University Press; Cambridge: 1999.
49. Pierscionek B, Smith G, Augusteyn RC. The refractive increments of bovine alpha-, beta-, and gamma-crystallins. *Vision Res*. 1987; 27:1539–41. [PubMed: 3445487]
50. Barnwal RP, Jobby MK, Devi KM, Sharma Y, Chary KV. Solution structure and calcium-binding properties of M-crystallin, a primordial betagamma-crystallin from archaea. *J Mol Biol*. 2009; 386:675–89. [PubMed: 19138688]

51. Shimeld SM, Purkiss AG, Dirks RP, Bateman OA, Slingsby C, Lubsen NH. Urochordate betagamma-crystallin and the evolutionary origin of the vertebrate eye lens. *Curr Biol*. 2005; 15:1684–9. [PubMed: 16169492]
52. Land, MF.; Nilsson, D-E. *Animal eyes*. Oxford University Press; Oxford: 2002.
53. Campbell MC. Measurement of refractive index in an intact crystalline lens. *Vision Res*. 1984; 24:409–15. [PubMed: 6740962]
54. Veretout F, Tardieu A. The protein concentration gradient within eye lens might originate from constant osmotic pressure coupled to differential interactive properties of crystallins. *Eur Biophys J*. 1989; 17:61–8. [PubMed: 2766998]
55. Keenan J, Orr DF, Pierscionek BK. Patterns of crystallin distribution in porcine eye lenses. *Mol Vis*. 2008; 14:1245–53. [PubMed: 18615203]
56. Amore JE, Bartley W, Van Heyningen R. Distribution of sodium and potassium within cattle lens. *Biochem J*. 1959; 72:126–33. [PubMed: 13651147]
57. Stevens A, Wang SX, Caines GH, Schleich T. 13C-NMR off-resonance rotating frame spin-lattice relaxation studies of bovine lens gamma-crystallin self association: effect of 'macromolecular crowding'. *Biochim Biophys Acta*. 1995; 1246:82–90. [PubMed: 7811735]
58. Grey AC, Schey KL. Distribution of bovine and rabbit lens alpha-crystallin products by MALDI imaging mass spectrometry. *Mol Vis*. 2008; 14:171–9. [PubMed: 18334935]
59. Huizinga A, Bot ACC, De Mul FFM, Vrensen GF, Greve J. Local variation in absolute water content of human and rabbit eye lenses measured by raman microspectroscopy. *Exp Eye Res*. 1989; 48:487–496. [PubMed: 2714410]
60. Uhlhorn SR, Borja D, Manns F, Parel JM. Refractive index measurement of the isolated crystalline lens using optical coherence tomography. *Vision Res*. 2008; 48:2732–8. [PubMed: 18824191]
61. Blundell T, Lindley PF, Miller LR, Moss DS, Slingsby C, Turnell WG, Wistow G. *Lens Res*. 1983; 1:109–131.
62. Keenan J, Elia G, Dunn MJ, Orr DF, Pierscionek BK. Crystallin distribution patterns in concentric layers from toad eye lenses. *Proteomics*. 2009; 9:5340–9. [PubMed: 19813212]
63. Slingsby C. Structural variation in lens crystallins. *TIBS*. 1985; 10:281–284.
64. White HE, Driessen HP, Slingsby C, Moss DS, Lindley PF. Packing interactions in the eye-lens. Structural analysis, internal symmetry and lattice interactions of bovine gamma IVa-crystallin. *J Mol Biol*. 1989; 207:217–35. [PubMed: 2738925]
65. Kappe G, Purkiss AG, van Genesen ST, Slingsby C, Lubsen NH. Explosive expansion of betagamma-crystallin genes in the ancestral vertebrate. *J Mol Evol*. 2010; 71:219–30. [PubMed: 20725717]
66. Simpson A, Moss D, Slingsby C. The avian eye lens protein delta-crystallin shows a novel packing arrangement of tetramers in a supramolecular helix. *Structure*. 1995; 3:403–12. [PubMed: 7613869]
67. Sivak JG. Optical properties of a cephalopod eye. *J Comp Physiol*. 1982; 147:323–327.
68. Sweeney AM, Des Marais DL, Ban YE, Johnsen S. Evolution of graded refractive index in squid lenses. *J R Soc Interface*. 2007; 4:685–98. [PubMed: 17293312]
69. Siezen RJ, Shaw DC. Physicochemical characterization of lens proteins of the squid *Nototodar* *gouldi* and comparison with vertebrate crystallins. *Biochim Biophys Acta*. 1982; 704:304–20. [PubMed: 7104371]
70. Tomarev SI, Zinovieva RD, Piatigorsky J. Crystallins of the octopus lens. Recruitment from detoxification enzymes. *J Biol Chem*. 1991; 266:24226–31. [PubMed: 1721068]
71. Chiou SH, Yu CW, Lin CW, Pan FM, Lu SF, Lee HJ, Chang GG. Octopus S-crystallins with endogenous glutathione S-transferase (GST) activity: sequence comparison and evolutionary relationships with authentic GST enzymes. *Biochem J*. 1995; 309 (Pt 3):793–800. [PubMed: 7639695]
72. Chiou SH. A novel crystallin from octopus lens. *FEBS Lett*. 1988; 241:261–4. [PubMed: 3197835]
73. Zinovieva RD, Piatigorsky J, Tomarev SI. O-Crystallin, arginine kinase and ferritin from the octopus lens. *Biochim Biophys Acta*. 1999; 1431:512–7. [PubMed: 10350626]

74. Piatigorsky J, Kozmik Z, Horwitz J, Ding L, Carosa E, Robison WG Jr, Steinbach PJ, Tamm ER. Omega -crystallin of the scallop lens. A dimeric aldehyde dehydrogenase class 1/2 enzyme-crystallin. *J Biol Chem.* 2000; 275:41064–73. [PubMed: 10961997]
75. Nilsson DE, Gislen L, Coates MM, Skogh C, Garm A. Advanced optics in a jellyfish eye. *Nature.* 2005; 435:201–5. [PubMed: 15889091]
76. Piatigorsky J, Norman B, Dishaw LJ, Kos L, Horwitz J, Steinbach PJ, Kozmik Z. J3-crystallin of the jellyfish lens: similarity to saposins. *Proc Natl Acad Sci U S A.* 2001; 98:12362–7. [PubMed: 11675486]
77. Castellano S, Lobanov AV, Chapple C, Novoselov SV, Albrecht M, Hua D, Lescure A, Lengauer T, Krol A, Gladyshev VN, Guigo R. Diversity and functional plasticity of eukaryotic selenoproteins: identification and characterization of the SelJ family. *Proc Natl Acad Sci U S A.* 2005; 102:16188–93. [PubMed: 16260744]
78. Wenk M, Herbst R, Hoeger D, Kretschmar M, Lubsen NH, Jaenicke R. Gamma S-crystallin of bovine and human eye lens: solution structure, stability and folding of the intact two-domain protein and its separate domains. *Biophys Chem.* 2000; 86:95–108. [PubMed: 11026675]
79. Bagby S, Harvey TS, Eagle SG, Inouye S, Ikura M. Structural similarity of a developmentally regulated bacterial spore coat protein to beta gamma-crystallins of the vertebrate eye lens. *Proc Natl Acad Sci U S A.* 1994; 91:4308–12. [PubMed: 8183906]
80. Srikanthan D, Bateman OA, Purkiss AG, Slingsby C. Sulfur in human crystallins. *Exp Eye Res.* 2004; 79:823–31. [PubMed: 15642319]
81. Chuang CC, Wu SH, Chiou SH, Chang GG. Homology modeling of cephalopod lens S-crystallin: a natural mutant of sigma-class glutathione transferase with diminished endogenous activity. *Biophys J.* 1999; 76:679–90. [PubMed: 9929473]
82. Babul J, Stellwagen E. Measurement of protein concentration with interference optics. *Anal Biochem.* 1969; 28:216–221. [PubMed: 5781411]
83. Barer R, Josephs R. Refractometry of living cells. *Quart J Microscop Sci.* 1954; 95:399–423.
84. Wen J, Arakawa T, Philo JS. Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions. *Anal Biochem.* 1996; 240:155–66. [PubMed: 8811899]
85. Crookes WJ, Ding LL, Huang QL, Kimbell JR, Horwitz J, McFall-Ngai MJ. Reflectins: the unusual proteins of squid reflective tissues. *Science.* 2004; 303:235–8. [PubMed: 14716016]
86. Kramer RM, Crookes-Goodson WJ, Naik RR. The self-organizing properties of squid reflectin protein. *Nat Mater.* 2007; 6:533–8. [PubMed: 17546036]
87. Eaton WA, Hofrichter J. Sick cell hemoglobin polymerization. *Adv Protein Chem.* 1990; 40:63–279. [PubMed: 2195851]
88. Liu Z, Weng W, Bookchin RM, Lew VL, Ferrone FA. Free energy of sickle hemoglobin polymerization: a scaled-particle treatment for use with dextran as a crowding agent. *Biophys J.* 2008; 94:3629–34. [PubMed: 18212015]
89. Heon E, Priston M, Schorderet DF, Billingsley GD, Girard PO, Lubsen N, Munier FL. The gamma-crystallins and human cataracts: a puzzle made clearer. *Am J Hum Genet.* 1999; 65:1261–7. [PubMed: 10521291]
90. Kmoch S, Brynda J, Asfaw B, Bezouska K, Novak P, Rezacova P, Ondrova L, Filipec M, Sedlacek J, Elleder M. Link between a novel human gammaD-crystallin allele and a unique cataract phenotype explained by protein crystallography. *Hum Mol Genet.* 2000; 9:1779–86. [PubMed: 10915766]
91. Pande A, Pande J, Asherie N, Lomakin A, Ogun O, King J, Benedek GB. Crystal cataracts: human genetic cataract caused by protein crystallization. *Proc Natl Acad Sci USA.* 2001; 98:6116–6120. [PubMed: 11371638]
92. Basak A, Bateman O, Slingsby C, Pande A, Asherie N, Ogun O, Benedek GB, Pande J. 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; 328:1137–47. [PubMed: 12729747]

93. Evans P, Wyatt K, Wistow GJ, Bateman OA, Wallace BA, Slingsby C. The P23T cataract mutation causes loss of solubility of folded gammaD-crystallin. *J Mol Biol.* 2004; 343:435–44. [PubMed: 15451671]
94. McManus JJ, Lomakin A, Ogun O, Pande A, Basan M, Pande J, Benedek GB. Altered phase diagram due to a single point mutation in human gamma-D crystallin. *Proc Natl Acad Sci USA.* 2007; 104:16856–16861. [PubMed: 17923670]
95. Moreau KL, King J. Hydrophobic core mutations associated with cataract development in mice destabilize human gammaD-crystallin. *J Biol Chem.* 2009; 284:33285–95. [PubMed: 19758984]
96. Pan FM, Chang WC, Chao YK, Chiou SH. Characterization of gamma-crystallins from a hybrid teleostean fish: multiplicity of isoforms as revealed by cDNA sequence analysis. *Biochem Biophys Res Commun.* 1994; 202:527–34. [PubMed: 8037758]
97. Chang T, Jiang YJ, Chiou SH, Chang WC. Carp gamma-crystallins with high methionine content: cloning and sequencing of the complementary DNA. *Biochim Biophys Acta.* 1988; 951:226–9. [PubMed: 3191133]
98. Kiss AJ, Cheng CH. Molecular diversity and genomic organisation of the alpha, beta and gamma eye lens crystallins from the Antarctic toothfish *Dissostichus mawsoni*. *Comp Biochem Physiol Part D Genomics Proteomics.* 2008; 3:155–71. [PubMed: 20483216]
99. Lee S, Mahler B, Toward J, Jones B, Wyatt K, Dong L, Wistow G, Wu Z. A single destabilizing mutation (F9S) promotes concerted unfolding of an entire globular domain in gammaS-crystallin. *J Mol Biol.* 399:320–30. [PubMed: 20382156]
100. Sliney DH. How light reaches the eye and its components. *Int J Toxicol.* 2002; 21:501–9. [PubMed: 12537646]
101. Chen J, Callis PR, King J. Mechanism of the very efficient quenching of tryptophan fluorescence in human gamma D- and gamma S-crystallins: the gamma-crystallin fold may have evolved to protect tryptophan residues from ultraviolet photodamage. *Biochemistry.* 2009; 48:3708–16. [PubMed: 19358562]
102. Xu J, Chen J, Toptygin D, Tcherkasskaya O, Callis P, King J, Brand L, Knutson JR. Femtosecond fluorescence spectra of tryptophan in human gamma-crystallin mutants: site-dependent ultrafast quenching. *J Am Chem Soc.* 2009; 131:16751–7. [PubMed: 19919143]
103. Aravind P, Wistow G, Sharma Y, Sankaranarayanan R. Exploring the limits of sequence and structure in a variant betagamma-crystallin domain of the protein absent in melanoma-1 (AIM1). *J Mol Biol.* 2008; 381:509–18. [PubMed: 18582473]
104. Ebersbach H, Fiedler E, Scheuermann T, Fiedler M, Stubbs MT, Reimann C, Proetzel G, Rudolph R, Fiedler U. Affilin-novel binding molecules based on human gamma-B-crystallin, an all beta-sheet protein. *J Mol Biol.* 2007; 372:172–85. [PubMed: 17628592]
105. Kapoor D, Singh B, Subramanian K, Guptasarma P. Creation of a new eye lens crystallin (Gambeta) through structure-guided mutagenic grafting of the surface of betaB2 crystallin onto the hydrophobic core of gammaB crystallin. *Febs J.* 2009; 276:3341–53. [PubMed: 19438717]
106. Hansen T, Jensen L, Åstrand PO, Mikkelsen KV. Frequency-dependent polarizabilities of amino acids as calculated by an electrostatic interaction model. *J Chem Theory Comput.* 2005; 1:626–633.
107. Millefiori S, Alparone A, Millefiori A, Vanella A. Electronic and vibrational polarizabilities of the twenty naturally occurring amino acids. *Biophys Chem.* 2008; 132:139–47. [PubMed: 18055096]
108. Kassimi NEB, Thakkar AJ. A simple additive model for polarizabilities: Application to amino acids. *Chem Phys Lett.* 2009; 472:232–236.
109. Cohn, EJ.; Edsall, JT. Density and apparent specific volume of proteins. In: Cohn, EJ.; Edsall, JT., editors. *Proteins, Amino Acids and Peptides.* Van Nostrand-Reinhold; Princeton, NJ: 1943. p. 370-381.
110. Heller W. Remarks on the refractive index mixture rules. *J Phys Chem.* 1965; 69:1123–1129.
111. Perlman GE, Longworth LG. The specific refractive increment of some purified proteins. *J Am Chem Soc.* 1948; 70:2719–2724. [PubMed: 18876982]
112. Pittz EP, Lee JC, Bablouzian B, Townend R, Timasheff SN. Light scattering and differential refractometry. *Methods Enzymol.* 1973; 27:209–256. [PubMed: 4773285]

113. Eisenberg, H. Biological macromolecules and polyelectrolytes in solution. Clarendon Press; Oxford: 1976.
114. Eisenberg H. Halophilic malate dehydrogenase - a case history of biophysical investigations: ultracentrifugation, light-, x-ray- and neutron scattering. *Biochem Soc Symp.* 1992; 58:113–125. [PubMed: 1445401]
115. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* 2003; 31:3497–500. [PubMed: 12824352]
116. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A.* 1992; 89:10915–9. [PubMed: 1438297]
117. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 1987; 4:406–25. [PubMed: 3447015]
118. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr.* 1994; 50:760–3. [PubMed: 15299374]
119. Ogawa M, Takahashi TC, Takabatake T, Takeshima K. Isolation and characterization of a gene expressed mainly in the gastric epithelium, a novel member of the ep37 family that belongs to the betagamma-crystallin superfamily. *Dev Growth Differ.* 1998; 40:465–73. [PubMed: 9783472]
120. Liu SB, He YY, Zhang Y, Lee WH, Qian JQ, Lai R, Jin Y. A novel non-lens betagamma-crystallin and trefoil factor complex from amphibian skin and its functional implications. *PLoS One.* 2008; 3:e1770. [PubMed: 18335045]
121. Ray ME, Wistow G, Su YA, Meltzer PS, Trent JM. AIM1, a novel non-lens member of the betagamma-crystallin superfamily, is associated with the control of tumorigenicity in human malignant melanoma. *Proc Natl Acad Sci U S A.* 1997; 94:3229–34. [PubMed: 9096375]
122. Jobby MK, Sharma Y. Calcium-binding crystallins from *Yersinia pestis*. Characterization of two single betagamma-crystallin domains of a putative exported protein. *J Biol Chem.* 2005; 280:1209–16. [PubMed: 15536081]
123. Di Maro A, Pizzo E, Cubellis MV, D'Alessio G. An intron-less betagamma-crystallin-type gene from the sponge *Geodia cydonium*. *Gene.* 2002; 299:79–82. [PubMed: 12459254]
124. Wistow GJ, Lietman T, Williams LA, Stapel SO, de Jong WW, Horwitz J, Piatigorsky J. Tau-crystallin/alpha-enolase: one gene encodes both an enzyme and a lens structural protein. *J Cell Biol.* 1988; 107:2729–36. [PubMed: 2462567]
125. Graham C, Hodin J, Wistow G. A retinaldehyde dehydrogenase as a structural protein in a mammalian eye lens. Gene recruitment of eta-crystallin. *J Biol Chem.* 1996; 271:15623–8. [PubMed: 8663049]
126. Reiss H, Frisch HL. Statistical mechanics of rigid spheres. *J Chem Phys.* 1959; 31:369–380.

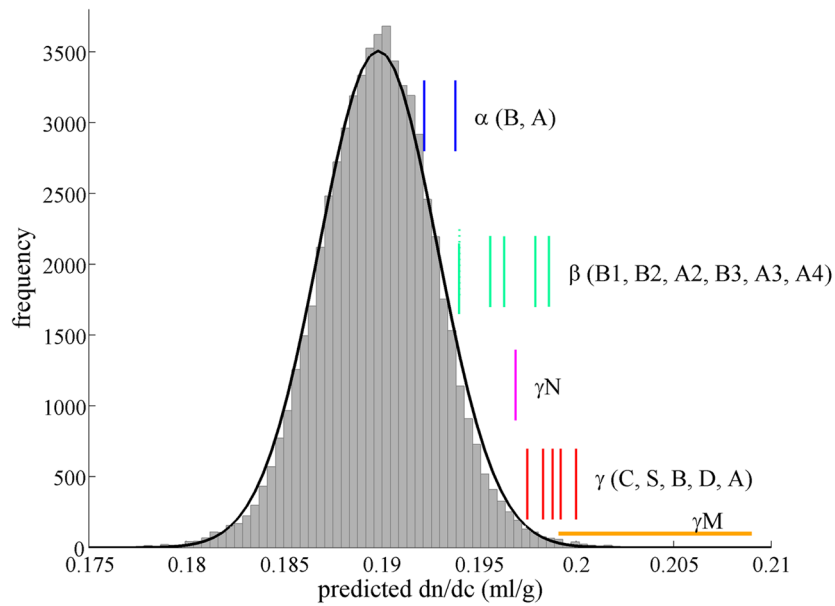
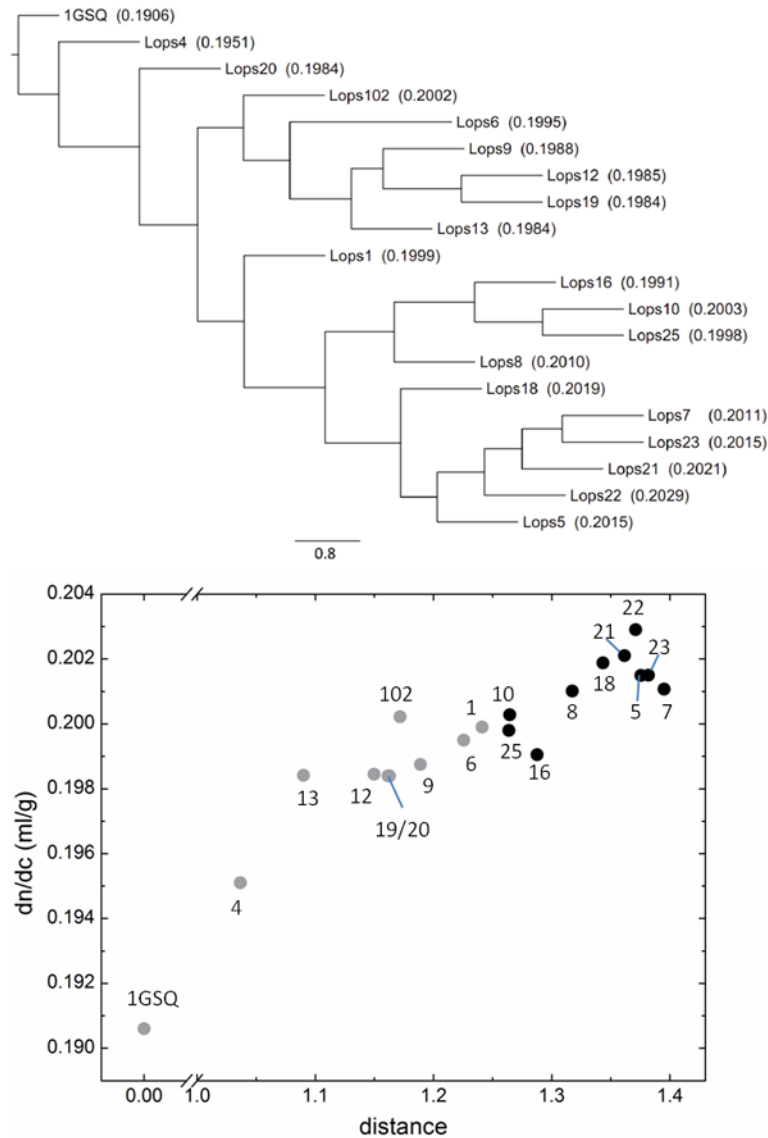


Figure 1.

Refractive index increment of human crystallin genes (vertical bars): human α -crystallin (blue), β -crystallin (green), γ N-crystallin (magenta), and γ -crystallin (red). They are drawn superimposed to the dn/dc distribution of all human proteins (histogram for the predicted proteins from the UCSC human genome browser for the human Feb. 2009 (GRCh37/hg19) assembly)⁴³. The best-fit Gaussian to the distribution of all proteins is indicated as black solid line, with mean of 0.1899 ml/g and standard deviation of 0.0031 ml/g. Also shown is the range of values computed for different fish γ M-crystallins (orange horizontal bar), as listed in Table 1.



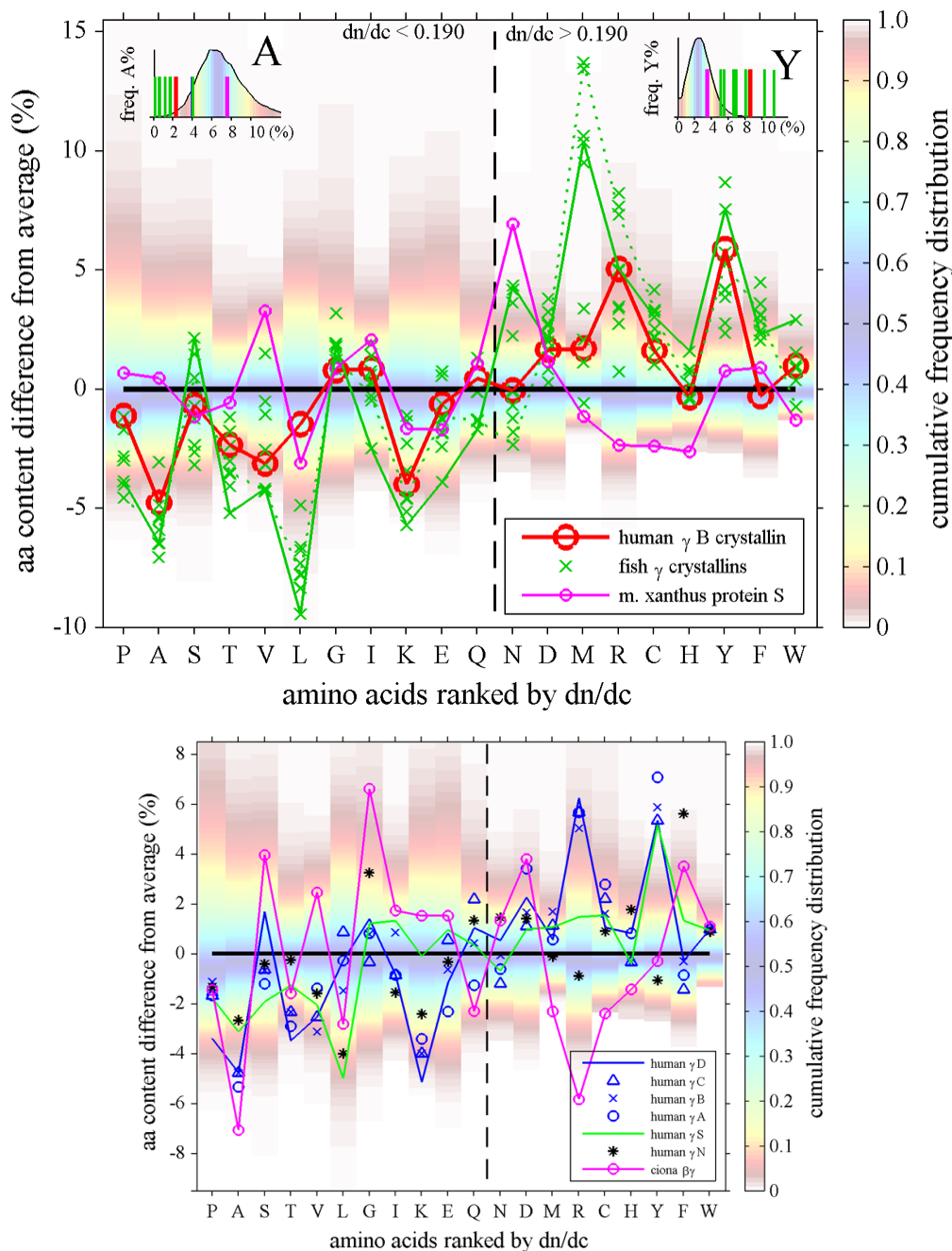


Figure 3. Top Panel: Amino acid composition of γ -crystallins (crosses and triangles) and *M. xanthus* protein S (magenta line with small circles), plotted as difference of their percent % content from the average content observed in human proteins. Each amino acid is represented in one vertical column. In the background of each column, the color patches indicate the spread of the distribution of fractional amino acid content in all human proteins (Feb. 2009 (GRCh37/hg19) assembly), using the color temperature scale on the right. For example, the range between the red shaded areas (from yellow to blue to yellow) encompasses the content of the amino acid found in approximately the central 80 percent of all proteins, whereas dark red shade fading to white indicates unusually high and low amino acid content among the lowest

5 or highest 95 percentile of all proteins. The insets for A and Y highlight the relationship between the observed amino acid content in crystallins (bars) and the histogram of the frequency distribution of that amino acid content in all proteins (in absolute %). The abscissa shows all amino acids ranked by increasing dn/dc . We can discern that in γ -crystallins the content of low dn/dc residues is mostly lower than average, whereas their content of high dn/dc residues is mostly higher than average. For example, human γ B-crystallin (red circles and solid red line) shows values predominantly in the lower left and upper right quadrant. This behavior is particularly evident for A, V, L, T and M, R, Y residues in γ -crystallins of fish (green crosses, showing data from carp, teleostean fish, lip shark, Mexican tetra, and zebra fish), which have contents far outside the normal distribution. Highlighted by the green solid line are the data for (green solid line) and lip shark γ -M2 (green dotted line), which have the highest and second highest values, respectively. In contrast, while Protein S (magenta line and small circles) exhibits unusual content of N and H, it shows values scattered more evenly for both high and low dn/dc residues. Bottom Panel: In the same comparison with the average abundances of the amino acids, shown are the data for different human γ -crystallins (blue symbols, γ S is highlighted by a blue line and γ D highlighted by a green line) and its closest known ancestor, the urochordate Ci- $\beta\gamma$ -crystallin (magenta).

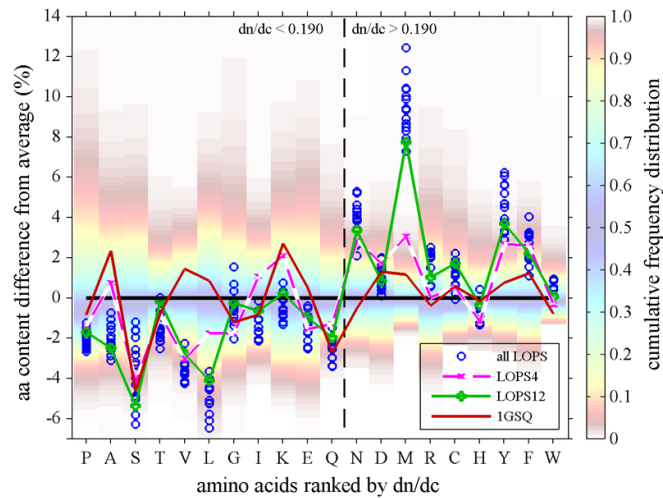


Figure 4.

Amino acid composition of different S-crystallins isoforms from *Loligo opalescens* (symbols) and of 1GSQ (red solid line), plotted as difference of their percent % content from the average content observed in human proteins, in the same presentation as in Figure 4. Highlighted are the most ancient S-crystallin (Lops4, magenta dashed line), and the most abundant and more recent S-crystallin (Lops12, green solid line). With the exception of F and D, all high dn/dc residues in the more recent molecule Lops12 are higher, and with the exception of G and E all low dn/dc residues in the more recent molecule are lower than in the most ancient Lops4. This results in the increase of refractive index in the order from GSQ to Lops4 to Lops12.

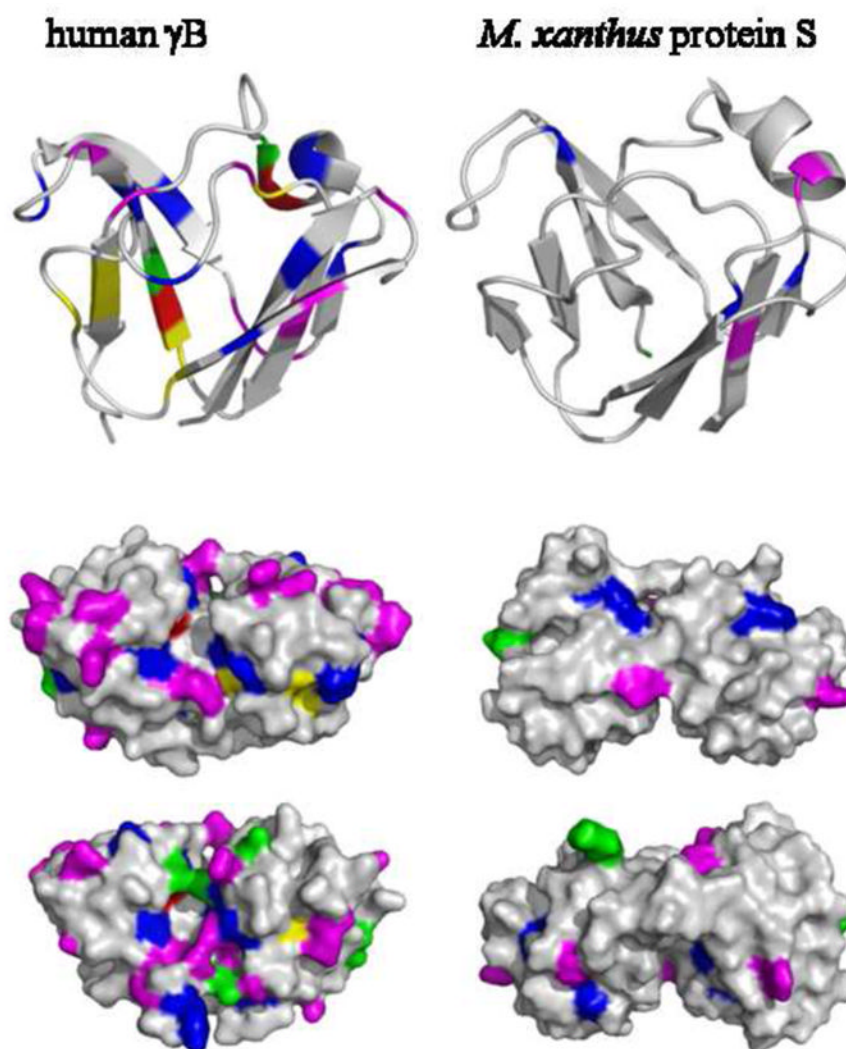


Figure 5. Top: Ribbon representation of the N-terminal domain of human γ B-crystallin (2JDF) and *M. xanthus* protein S (1PRR). Highlighted residues are in high relative abundance in γ B-crystallin with high refractive index increment: methionine (green), arginine (magenta), cysteine (yellow), tryptophan (red), and tyrosine (blue). Middle and Bottom: Surface representation of the whole molecules containing two domains, viewed from two opposite angles.

```

γB_human      1  G S T I L F E D V E F G K K I E L E T S V S D L N V E G S F N D I V S S I V E S G 40
C.intestinalis 1  G S T I L F E D V E F G K K I E L E T S V S D L N V E G S F N D I V S S I V E S G 42

γB_human      41  T W F V E D D E G H S G P S Y K L T F G K Y P N P G S W G G N L D E L S S V K Q Q 81
C.intestinalis 43  T W F V E D D E G H S G P S Y K L T F G K Y P N P G S W G G N L D E L S S V K Q Q 83
    
```

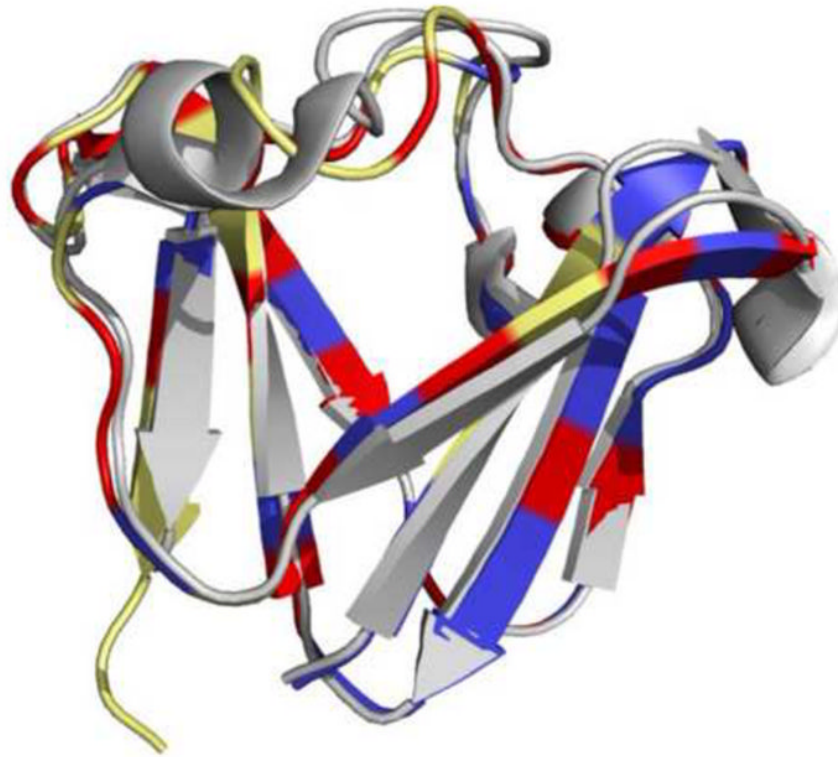


Figure 6. Sequence and structure comparison of N-terminal domain of human γ B crystallin (2JDF) and $\beta\gamma$ -crystallin from *C. intestinalis* (2BV2). (A) Sequence alignment of the two polypeptides. The identical residues are in blue, and the residues with higher dn/dc in human γ B crystallin are highlighted in red. (B) Structural overlay of the two constructs. The N-terminal domain of human γ B crystallin is in yellow and $\beta\gamma$ -crystallin from *C. intestinalis* is in gray. The residues highlighted in blue are the identical ones. The residues in human γ B crystallin which have higher dn/dc values compared to those at the same positions in $\beta\gamma$ -crystallin *C. intestinalis* are highlighted in red.

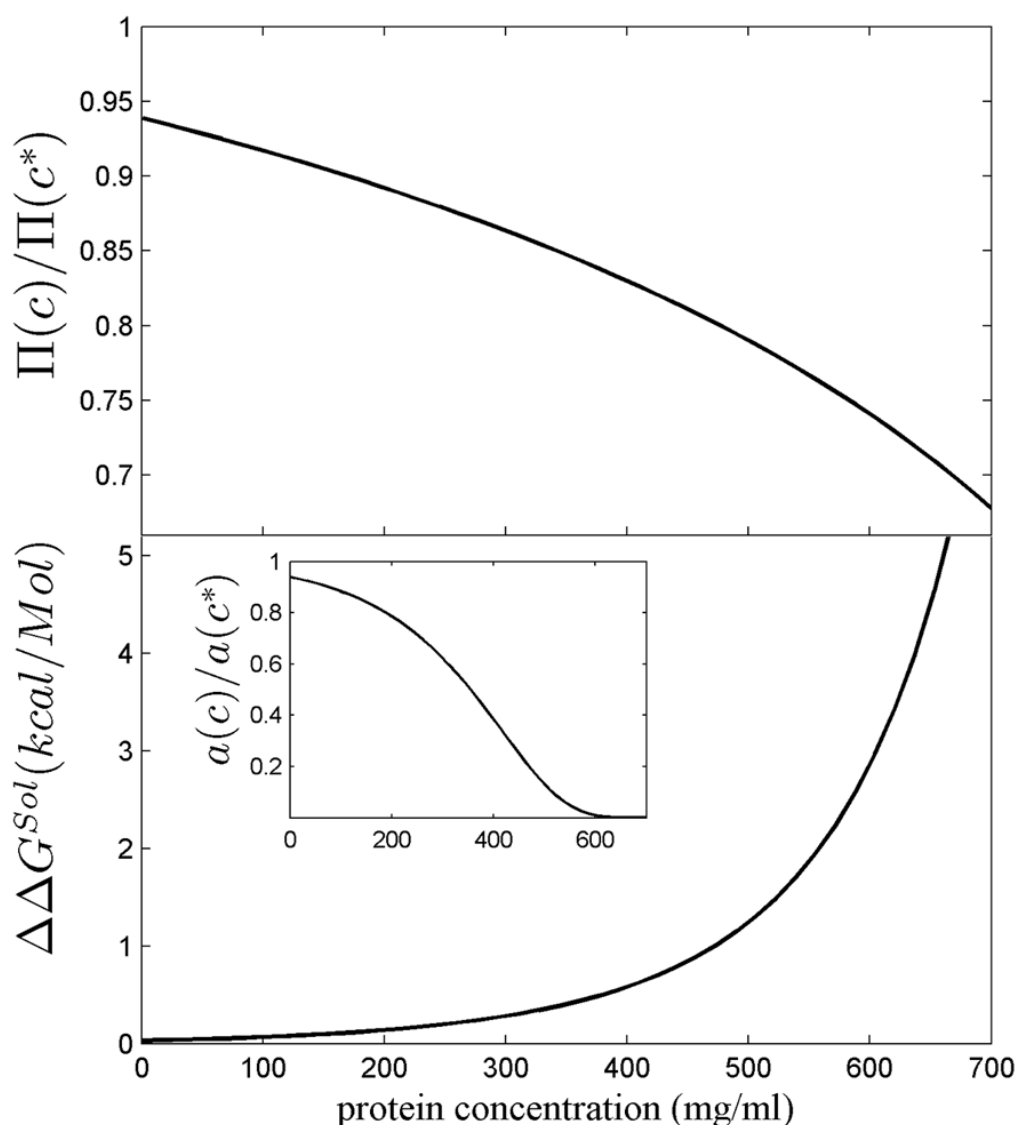


Figure 7.

Thermodynamic effect of a crystallin refractive index 6.5% above that of average proteins, corresponding to the value of bovine γ -crystallin. Results are based on a non-interacting hard sphere model of crystallin calculating the excluded volume by scaled particle theory¹²⁶, as described in detail in⁴⁶, and presented as a function of total protein concentration c . In order to reach the same tissue refractive index as high dn/dc molecules at concentration c , molecules with average dn/dc would have to be present at a higher concentration c^* , in the present case $c^* = 1.065 c$. Top Panel: The ratio of the osmotic pressure $\Pi(c)/\Pi(c^*)$, corresponding to the reduction in osmotic pressure arising from the high dn/dc . Bottom: The difference in the free energy of transfer of a crystallin molecule from the soluble phase into a polymer phase, $\Delta G_{Sol}(c)$, to the same free energy at a higher crystallin concentration, $\Delta G_{Sol}(c^*)$, with $\Delta\Delta G^{Sol}(c) = \Delta G_S(c) - \Delta G_S(c^*)$. The calculations are based on a ‘crystal approximation’^{87,88}. The positive sign of $\Delta\Delta G^{Sol}$ means that at the lower concentration afforded by the increase in dn/dc makes the transition into the polymer phase less favorable. The insert shows the ratio of the chemical activity (i.e. ratio of ‘effective’ concentrations) arising from a 6.5% lower actual protein concentration.

Table 1Refractive index increment of members of the $\beta\gamma$ -crystallin family

protein	accession #	dn/dc (ml/g)	$P^{\&}$
human alpha-A-crystallin	CAG28619	0.1938	0.097
human alpha-B-crystallin	NP_001876	0.1922	0.22
human beta-A1/A3 crystallin	NP_005199	0.1979	0.0038
human beta-A2-crystallin	NP_476435	0.1956	0.029
human beta-A4-crystallin	NP_001877	0.1986	0.0019
human beta-B1-crystallin	NP_001878	0.1940	0.086
human beta-B2-crystallin	AAB39700	0.1940	0.086
human beta-B3-crystallin	NP_004067	0.1963	0.016
human gamma-A-crystallin	NP_055432	0.2000	3.8E-4
human gamma-B-crystallin	NP_005201	0.1988	0.0015
human gamma-C-crystallin	NP_066269	0.1975	0.0056
human gamma-D-crystallin	NP_008822	0.1992	9.7E-4
human gamma-S-crystallin	NP_060011	0.1983	0.0026
human gamma-N-crystallin	NP_653328	0.1969	0.0098
gamma-crystallin M1 (carp)	P10043	0.2037	2.1E-6
gamma-crystallin M1-2 (teleostean fish)	AAA50960	0.2017	4.2E-5
gamma-crystallin M2-2 (teleostean fish)	AAA50961	0.2041	1.1E-6
gamma-crystallin M2-3 (teleostean fish)	AAA50962	0.1995	6.9E-4
gamma-crystallin M1 (lip shark)	P48648	0.2030	7.3E-6
gamma-crystallin M2 (lip shark)	P48649	0.2050	2.4E-7
gamma-crystallin I (bamboo shark)	CAA55810	0.2031	6.3E-6
gamma-crystallin M (Mexican tetra)	AAF07205	0.2091	1.2E-9
gamma-crystallin MX (zebra fish)	NP_001013280	0.2003	2.6E-4
gamma-crystallin S1 (zebra fish)	NP_001013294	0.1991	0.0011
gamma-crystallin N1 (zebra fish)	Q5XTN8	0.2001	3.4E-4
gamma-crystallin N2 (zebra fish)	Q6DGY7	0.2014	6.3E-5
EDSP of the amphibious newt ¹¹⁹	BAA01991	0.1901	0.47
$\beta\gamma$ -component of the protein complex $\beta\gamma$ -CAT in toad skin ¹²⁰	ABV56830	0.1882	0.71
AIM1 ¹²¹	AAB53792	0.1822	0.995
protein S of <i>Myxococcus xanthus</i> ²⁷	AAA25407	0.1874	0.80
spherulin 3A of <i>Physarum polycephalum</i> ²⁸	1AG4_A	0.1911	0.34
Yersina crystallin <i>Yersina pestis</i> CO92 ¹²²	YP_002347819	0.1917	0.27
$\beta\gamma$ -crystallin <i>Geodia cydonium</i> ¹²³	CAA70024	0.1923	0.21
archaeal M-crystallin ⁵⁰	AAM05909	0.1961	0.020
the urochordate $\beta\gamma$ -crystallin ⁵¹	2BV2_B	0.1918	0.26

$\&$ P is the probability for obtaining just by chance a dn/dc value as large or larger, if it was a normally distributed random variable as shown in the solid line of Figure 1.

Table 2

Refractive index increment of taxon-specific crystallins

species	related protein/enzyme	dn/dc (ml/g)	crystallin	dn/dc (ml/g)
lampreys, fish, reptiles, birds ¹²⁴	α -enolase ^a	0.189	τ	0.189
duck ³²	lactate dehydrogenase B ^b	0.187	ϵ	0.187
duck/chicken ³¹	argininosuccinate lyase ^c	0.185	δ	0.185–0.187
rabbit ³⁴	L-gulonate 3-dehydrogenase ^d	0.188	λ	0.188
elephant shrew ¹²⁵	aldehyde dehydrogenase ^e	0.190	η	0.189
gecko ⁴⁰	retinol binding protein I ^h	0.195	ι	0.195
platypus ³³	lactate dehydrogenase A ^j	0.189	υ	0.188
octopus ⁷⁰	aldehyde dehydrogenase ^f	0.190	Ω	0.193
scallop ³⁰	aldehyde dehydrogenase ^f	0.190	Ω	0.192
octopus ⁷³	Lipid-binding protein PEBP ^g	0.191	O	0.194

Prediction based on sequences from accession number:

^a AAA49218^b AAA49221^c NP_001025885^d P14755.3^e AAC48588^f AAF73122^g P13696.2 GI:1352725^h NP_036865^j P00340

Table 3

Cephalopod S-crystallins and glutathione-S-transferases

protein	accession number	<i>dn/dc</i> (ml/g)	P
S-crystallin (SL4)	AAA91343	0.1954	0.033
S-crystallin (SL11)	AAA29403	0.1940	0.086
S-crystallin (SL18)	AAA29404	0.1965	0.014
S-crystallin (SL20)	AAA29402	0.1987	0.0017
Lops1	AAA97541	0.1900	0.49
Lops4	AAA97540	0.1951	0.042
Lops5	AAA97542	0.2015	5.5E-05
Lops6	AAA97543	0.1995	6.9E-04
Lops7	AAA97544	0.2011	9.4E-05
Lops8	AAA97547	0.2010	1.1E-04
Lops9	AAA97545	0.1988	0.0015
Lops10	AAA97546	0.2003	0.00030
Lops12	AAA97548	0.1985	0.0021
Lops13	AAA97549	0.1984	0.0023
Lops16	AAA97550	0.1991	1.1E-03
Lops18	AAB01053	0.2019	3.2E-05
Lops19	AAB01054	0.1984	0.0023
Lops20	AAB01055	0.1984	0.0023
Lops21	AAB01056	0.2021	2.4E-05
Lops22	AAB01057	0.2029	7.3E-06
Lops23	AAB01058	0.2015	5.5E-05
Lops25	AAB01059	0.1998	4.8E-04
Lops102	AAA97551	0.2002	3.0E-04
OctS1	P27013	0.1969	0.0098
OctS2	P27014	0.1975	0.0056
OctS3	Q25626	0.1975	0.0056
OL1	P27009	0.1967	0.012
OL2	P27010	0.1963	0.016
OL3	P27011	0.1977	0.0047
OL4	P27012	0.1968	0.011
S I (squid <i>Nototodarus gouldi</i>)	N/A *	0.1964	0.015
S II (squid <i>Nototodarus gouldi</i>)	N/A *	0.1982	0.0028
S III (squid <i>Nototodarus gouldi</i>)	N/A *	0.1994	7.7E-04
glutathione-S-transferases			
1GSQ [<i>Ommastrephes sloani pacificus</i> (squid)]	1GSQ_A	0.1906	0.41
1GNE [<i>Schistosoma Japonicum</i>]	1GNE	0.1942	0.076

protein	accession number	<i>dn/dc</i> (ml/g)	P
1GUH [human]	1GUH_A	0.1899	0.50
6GST [Rattus rattus (black rat)]	6GST_A	0.1955	0.031
2GSR [Sus scrofa (pig)]	2GSR_A	0.1911	0.34

[&] P is the probability for obtaining just by chance a *dn/dc* value as large or larger, if it was a normally distributed random variable as shown in the solid line of Figure 1.

* determined from the amino acid composition reported in ⁶⁹.

Table 4

Contribution of different amino acids to the elevation of crystallin refractive indices

amino acid	$(dn/dc)_a$ (ml/g) ⁽¹⁾	normalized dn/dc contribution of each amino acid (%) ⁽²⁾		
		γ B ⁽³⁾	γ M ⁽⁴⁾	Lops22 ⁽⁵⁾
Pro	0.166	3.6 (2.9)	6.5 (2.0)	5.3 (2.2)
Ala	0.167	9.5 (7.7)	8.2 (2.9)	4.3 (1.3)
Ser	0.170	1.3 (1.0)	-3.6 (-4.2)	5.6 (2.2)
Thr	0.172	5.8 (4.3)	8.3 (2.5)	5.2 (1.7)
Val	0.172	8.5 (6.5)	6.6 (1.4)	7.0 (2.9)
Leu	0.173	4.8 (3.5)	15.9 (7.4)	14.0 (7.9)
Gly	0.175	-0.7 (-0.5)	-1.5 (-3.0)	0.5 (-1.5)
Ile	0.179	-1.2 (-0.8)	3.5 (-0.8)	3.9 (0.3)
Lys	0.181	8.5 (5.0)	8.1 (0.8)	1.6 (-1.0)
Glu	0.183	1.1 (0.6)	5.8 (-0.5)	4.3 (0.0)
Gln	0.186	-0.9 (-0.2)	2.1 (-2.0)	4.6 (-0.7)
Asn	0.192	-0.4 (0.1)	-3.6 (-1.8)	-3.7 (-1.1)
Asp	0.197	-0.4 (1.2)	-1.1 (-1.7)	-0.8 (-0.9)
Met	0.204	1.1 (3.1)	-3.2 (6.0)	1.5 (12.6)
Arg	0.206	5.5 (11.7)	-1.2 (3.0)	0.6 (1.9)
Cys	0.206	1.1 (2.7)	-0.5 (-0.1)	0.3 (0.2)
His	0.219	-1.5 (-2.2)	1.0 (0.9)	-1.8 (-5.4)
Tyr	0.240	38.2 (42.4)	18.2 (2.2)	25.7 (30.1)
Phe	0.244	-3.5 (-4.5)	5.3 (6.4)	8.7 (10.1)
Trp	0.278	15.7 (18.1)	16.7 (19.3)	7.9 (8.0)

⁽¹⁾ Calculated for a hypothetical poly-peptide of that amino acid.

⁽²⁾ Difference between the dn/dc calculated from the protein sequence $(dn/dc)_P$ and a sequence where the relative abundance of amino acid X is changed to that of average proteins $(dn/dc)_{PX_{avg}}$, normalized by the total difference $(dn/dc)_P - 0.190$ ml/g. The values in parenthesis are the difference between $(dn/dc)_{avgXP}$ of an average protein with relative abundance of X set to that in crystallin minus 0.190, normalized by the total difference $(dn/dc)_P - 0.190$ ml/g.

⁽³⁾ human γ B crystalline (NP_005201).

⁽⁴⁾ Mexican tetra γ M-crystallin (AAF07205).

⁽⁵⁾ squid S-crystallin (AAB01057). The solid horizontal line separates amino acids with lower than average from those with higher than the dn/dc of average proteins. The horizontal lines in the columns of the individual proteins similarly separates the amino acids lower and higher relative to the dn/dc of that crystallin.