Published in final edited form as:
J Mol Biol. 2007 September 7; 372(1): 205–222.

Biophysical Properties of γC-Crystallin in Human and Mouse Eye Lens: The Role of Molecular Dipoles

Andrew G. Purkiss[1,*], **Orval A. Bateman**[1,*], **Keith Wyatt**[2], **Phillip A. Wilmarth**[3], **Larry L. David**[3], **Graeme J. Wistow**[2], and **Christine Slingsby**[1,¶]

1*Birkbeck College, Department of Crystallography, Institute of Structural Molecular Biology, University of London, Malet Street, London WC1E 7HX, UK.*

2*Section on Molecular Structure and Functional Genomics, National Eye Institute, Bethesda, MD 20892-0703, USA.*

3*Oregon Health Science University, School of Dentistry, Dept. Integrated Biosciences, 611 SW Campus Dr, Portland, OR 97239, USA.*

Summary

The eye lens is packed with soluble crystallin proteins, providing a lifetime of transparency and light refraction. γ-crystallins are major components of the dense, high refractive index central regions of the lens and generally have high solubility, high stability and high levels of cysteines. Human γC belongs to a group of γ-crystallins with a pair of cysteines at positions 78 and 79. Unlike other γ-crystallins it has relatively low solubility, whereas mouse γC, which has the exposed C79 replaced with arginine, and a novel mouse splice variant, γCins, are both highly soluble. Furthermore, human γC is extremely stable, while the mouse orthologs are less stable. Evolutionary pressure may have favoured stability over solubility for human γC and the reverse for the orthologs in the mouse. Mutation of C79 to R79, in human γC, greatly increased solubility, however neither form produced crystals. Remarkably, when the human γD R36S crystallization cataract mutation was mimicked in human γC-crystallin, the solubility of γC was dramatically increased, although it still did not crystallize. The highly soluble mouse γC-crystallin did crystallize. Its X-ray structure was solved and used in homology modelling of human γC, and its mutants C79R and R36S. The human γD R36S mutant was also modelled from human γD coordinates. Molecular dynamics simulation of the six molecules in the solution state showed that the human γCs differed from γDs in domain pairing, behaviour that correlates with interface sequence changes. When the fluctuations of the calculated molecular dipoles, for the six structures, over time were analysed, characteristic patterns for soluble γC and γD proteins were observed. Individual sequence changes that increase or decrease solubility correlated well with changes in the magnitude and direction of these dipoles. It is suggested that changes in surface residues have allowed adaptation for the differing needs of human and mouse lenses.

Keywords

crystallin; cysteine; eye lens; protein solubility; solution interactions

[*] These authors contributed equally to the work.
[¶] corresponding author

Introduction

The optical properties and molecular composition of vertebrate lenses vary depending on the organism's habitat.[1-3] The refractive power of the lens derives largely from mixtures of crystallins, proteins that exist in a crowded environment inside cells with little capacity for protein renewal where they have to last a lifetime. Crystallins come in a wide range of assembly sizes but in most mammals they are products of two (unrelated) gene families: the small heat shock proteins/α-crystallins and the βγ-crystallins.[4] A comparison of the repertoire of expressed proteins in human and rodent lens indicates patterns of specific adaptations to the requirements of different species.[5] The human lens cytoplasm is largely composed of 12 gene product crystallins[6] in ratios that can now be quantified using mass spectrometry, showing that newborn lens is composed of equal amounts of α-, β- and γ-crystallins.[7] Monomeric γ-crystallins are synthesized early in life and remain in the central regions of the mature and ageing eye lens. As the human lens is required to remain transparent for decades, human γ-crystallins must have evolved structures with remarkable biophysical properties.

The underlying evolutionary mechanism for providing γ-crystallins to suit particular lifestyles involves rounds of gene loss and expansion in the different classes of vertebrates.[8] In placental mammals, there is a cluster of six closely related γ-crystallin genes: γA-F, along with two more distantly related members, γS and γN, although in humans expression of some of these genes has been lost.[8,9] Considerable efforts have been expended on characterizing several mammalian γ-crystallins, including 3D structures of: bovine γB-crystallin pdb 1amm[10], bovine γD-crystallin pdb 1elp[11], human γD-crystallin pdb 1hk0[12], rat γE-crystallin pdb 1a5d[13] and pdb 1zie[14], bovine γE-crystallin (pdb 1m8u, Mayer et al., 2003, unpublished), bovine γF-crystallin pdb 1a45[13], and mouse γS-crystallin by NMR, pdb 1zwm[15]. Structural biology of a major human lens protein, γD-crystallin, has shown that in common with most characterized vertebrate crystallins it is very soluble and stable.[16-18] The selection pressure for optimization of solubility in γD is evident from several childhood cataracts that result from surface point mutations, often replacing arginines.[19-22] Surprisingly, γ-crystallin sequences are rich in the reactive amino acid cysteine. It has been argued that the high local concentration of protein cysteine, higher than the cytoplasmic reducing agent GSH, may contribute to their long-term stability by scavenging free radicals.[23]

The major γ-crystallin expressed in the young human lens is γC-crystallin.[6,7] Human γC differs from the other γ-crystallins that have been studied in detail in its distribution of cysteine residues; it belongs to a group of γB/γC sequences in different mammals that have cysteine at position 79 rather than arginine. In contrast, the mouse ortholog has the more common arginine at 79. Here we compare and contrast the human and mouse γC recombinant proteins and also describe a variant of mouse γC, γCins, which is the result of alternative splicing and which does not occur in the human lens. These proteins turn out to have significantly different properties that may relate to the adaptation of γC-crystallin for the accommodating, long-lived human lens as compared with a shorter-lived, short-sighted mouse. The properties of human γC-crystallin also give us new insight into a major crystallin in the centre of the human lens and into the determinants of stability and solubility in this family.

Results

(a) The sequences of human, mouse, dog, bovine and guinea pig γCs and γ-crystallins with 79C

Complete sets of sequences for γB and γC crystallins were derived from EST analysis of several mammalian species [5] and unpublished data available through http://neibank.nei.nih.gov. Analysis of multiple cDNA clones from mouse eye cDNA libraries revealed the existence of a novel form of γC at a frequency of about 14%. This variant, designated γCins, contains an

additional residue (Gln) in the linker region as a result of alternative splice site usage at the 5′ end of exon 3. Interestingly, a similar potential alternative splice site is also present in the human γC gene. However, if used in human this alternative splice would include a stop codon (TAG) rather than the CAG found in mouse. This would produce a one-domain version of human γC, however there is no present genomic or proteomic evidence for such an expressed variant.

Figure 1a shows these sequences, with the 2-domain γ-crystallins divided into their component modules (Greek key motifs, linker, and C-terminal extension), together with human γA and human γD-crystallin (for which the 3D structure and biophysical data are available). A buried cysteine at position 78 is a conserved feature of motif 2 in γ-crystallins, but its neighbour at position 79, although commonly an arginine (present also in topologically equivalent positions in motifs 1 and 4), can also be a cysteine. (The highly expressed human γD-crystallin and γS-crystallins have arginine at position 79). Of the species examined, only primates (human, chimp and rhesus monkey) have a cysteine at this position in both γB and γC sequences. Human and mouse γC-crystallins were characterized to assess the effect of Cys79, as well as some other sequence variations.

(b) Protein levels in mouse and human lens

Lack of resolution of mouse γC and γCins precludes quantitation of the relative abundance of these crystallins from two-dimensional electrophoresis (2-DE) gels. However, evidence that the two mouse γC-crystallin sequences are synthesized in the lens comes from mass spectrometry whereby deconvoluted spectra of co-eluting mouse γC and mouse γC insert proteins from a cation exchange column have masses matching the ones predicted from the sequences (Figure 2). These data also give an estimate of the relative abundance of the two mouse crystallins as 86% γC and 14% γCins, which agrees exactly with the abundance levels predicted from EST data. The percentage that γC and γCins comprise the total soluble protein in mouse lens is more difficult to estimate because γC, γCins, and γB co-migrate on 2-DE gels and could not be resolved by cation exchange chromatography.[24] However, the combined abundance of γCs and γB in the soluble fraction of 4-week-old mouse was estimated from 2-DE gels as 8.2% of the total soluble fraction.[24] To determine the relative proportion of γCs and γB in young mouse lens, a shotgun analysis of a complex mixture of trypsin digested mouse soluble protein was analyzed by two-dimensional liquid chromatography mass spectrometry, and the number of spectra assigned to γCs and γB was used to estimate the relative abundance of the two proteins. The number of spectra assigned to each protein was similar (1003 counts for γC and 815 counts for γB), suggesting roughly equal amounts of γB and γC (approximately 4% each of the total soluble protein of mouse lens). In contrast mouse γD comprised 9% of the soluble protein in mouse lens.[24]

In young human lens, both two-dimensional electrophoresis (2-DE) data and more recent high-resolution mass spectral analysis indicate that the relative abundance of γC and γD-crystallins range from 12-16% and 3-8 % mole percent of the total soluble protein, respectively.[7] Thus, in young human lens, the approximate ratio of γC/γD in the soluble protein fraction is around 2.5, compared with 0.4 in 18-28 day old mice.

(c) Human γC-crystallin behaves differently from the C79R mutant and mouse orthologs during purification

Human γC-crystallin was highly expressed in the soluble cell extract of *E. coli* following induction, as gauged by the presence of a major protein band around 21 kDa on Coomassie stained SDS gels, as has already been shown by Fu and Liang.[25] The identity of the protein was confirmed by mass spectrometry. Attempts to purify the recombinant human γC-crystallin protein following standard chromatography methods for isolating lens and recombinant γ-

crystallins[13] led to very low protein recoveries. The protein could not be dialyzed, or desalted for further column treatment. Performing column chromatography procedures in urea-containing and glycine-containing buffers still resulted in very poor recoveries. Overall a protein concentration of no more than 1 mg ml^{-1} was achieved. Salt precipitation was therefore used as a first protein purification step. Conditions ranging from 20 to 70% ammonium sulphate precipitation indicated that the bulk of the crystallin protein could be recovered from a single precipitation from the soluble fraction of the cell lysate with 50% saturated ammonium sulphate, and could then be resolubilized to a limiting (total) protein concentration of around 50 mg ml^{-1}. A single gel filtration column purification procedure was possible under optimized conditions of 200 mM ammonium acetate, although even under these conditions, protein recoveries were low and human γC-crystallin did not survive re-running on the gel filtration column. Protein oxidation was unlikely to be the cause as increasing the concentration of DTT from 2 mM to 10 mM DTT did not increase yields. Human γC-crystallin elutes with the same elution volume as the closely related family member human γD-crystallin indicating that it is a monomer. By combining around 10 gel filtration runs, 2.5 ml of the human γC-crystallin at a protein concentration of around 0.7 mg ml^{-1} was prepared, and concentrated in a VivaSpin 500 10MWCO to a limiting protein concentration of around 5mg ml^{-1}.

In contrast to human γC, both mouse γC and γCins purified easily using standard chromatographic procedures with good recoveries.

(d) Comparison of protein solubilities

To determine whether the marked difference in behavior during isolation between human and mouse proteins was related to C79, the human sequence was mutated to R79 and expressed. Unlike the wild-type sequence, this produced a highly soluble protein similar to mouse. Other natural mutations in exposed residues are known to affect solubility in human γD-crystallin. We asked whether similar mutations would further decrease the solubility of human γC-crystallin. As a first step, the effect of a congenital cataract mutation found in human γD-crystallin R36S (this residue in motif 1 is topologically equivalent to R79 in motif 2, Figure 1a), was tested in the context of the wild-type human γC-crystallin sequence. Solubility for all these proteins was measured by concentration.

Concentration of wild-type human γC-crystallin was difficult. Freeze-drying was attempted but failed to yield soluble protein. Attempts to increase the concentration using Microcons in low salt buffers resulted in the final protein concentration being orders of magnitude less than that of the initial solution. Using optimized conditions whereby the protein was maintained in 200 mM ammonium acetate, an attempt was made to concentrate the protein solution using a VivaSpin 500 10MWCO which had been treated by first spinning through some 5% PEG 6K in an attempt to block protein interaction sites, but this did not increase the yield. However, when the protein was maintained in 200 mM ammonium acetate, a limiting protein concentration of ~5 mg/ml was obtained at the expense of a ~50% loss of protein.

In contrast, the human γC-crystallin C79R mutant, the ortholog mouse γC-crystallin and mouse γCins-crystallin could all be concentrated by ultrafiltration to around 90, 80 and 100 mg/ml, respectively. Lyophilization was not attempted for these soluble crystallins, although a previous study on human γD-crystallin indicated that higher solubilities could be achieved using this method.[17]

Unexpectedly, the solubility of the human γC-crystallin R36S mutant, a sequence that has C79, was found to be substantially higher, reaching around 200 mg/ml. This surprising result indicates that dramatic changes in solubility may be due to changes in the overall surface properties of the proteins rather than simply to the presence or absence of a surface cysteine.

Unlike solutions of bovine and rat γE and γF-crystallins[13], the human and mouse γC-crystallins, and their mutants did not show cold-induced precipitation in low ionic strength buffers around neutral pH.

(e) Comparison of protein stabilities

Protein conformational stabilities for human γC-crystallin, human γC-crystallin C79R, mouse γC-crystallin and mouse γCins were estimated by determining their degree of unfolding after incubation in a chemical denaturant over a range of concentrations. An indication that all four crystallins were stable was found when the proteins showed no change in fluorescence measurements when dissolved in urea at concentrations up to 8 M. Therefore the stabilities were estimated using unfolding conditions previously described for the very stable human γD-crystallin.[26,27] Following tryptophan excitation at 280 nm, the tryptophan fluorescence spectra of folded (0 M GdnHCl) and unfolded (6 M GdnHCl) human γC-crystallin showed the expected wavelength shift in emission maximum, although unlike other measured γ-crystallins[28] the fluorescence intensity of the unfolded state was lower than when folded, indicating that tryptophan residues were not quenched (Figure 3a). Similar spectra were obtained for the human γC-crystallin C79R mutant (Figure 3b). The fluorescence spectra for the two mouse crystallins showed that they also unfolded, as evidenced by the shift in wavelength, although in these protein spectra, like other measured γ-crystallins, the fluorescence intensity of the unfolded state was higher than the folded state indicating that tryptophan residues were efficiently quenched (Figures 3c and 3d). The molecular basis for quenching and its possible role in relation to protection from UV damage is currently under investigation for human γD-crystallin.[28]

When these four crystallins were unfolded in increasing amounts of GdnHCl, unfolding curves were measured from the change in tryptophan emission at 350 nm, and the mid point of unfolding estimated (Figure 4) and compared with human γD-crystallin. The wild-type γC-crystallin and C79R γC-crystallin mutant human protein had similar stabilities to human γD-crystallin,[17,26] whereas mouse γC-crystallin and the insert version were less stable. The high stability of human γC-crystallin makes it unlikely that protein denaturation is responsible for its low solubility.

(f) Mouse γC-crystallin structure

Mouse γC-crystallin crystallized in space group P1 and the diffraction statistics are shown in Table 1. Phase information for the dataset was generated by molecular replacement using the program PHASER[29] and using the coordinates of human γD-crystallin (pdb code 1HK0) with residues differing in sequence cut back to alanine. Two molecules are present in the unit cell, with R_{cryst} = 0.430 and R_{free} = 0.445. Initial restrained refinement with REFMAC brought these values down to R_{cryst} = 0.288 and R_{free} = 0.363, before several rounds of manual rebuilding, addition of water molecules, using COOT, and refinement. TLS refinement was then used, with each TLS group being a single γ-crystallin domain, to give a final R_{cryst} = 0.219 and R_{free} = 0.285 (Table 2). The coordinates (id: 2v2u) and structure factors (id: r2v2usf) have been submitted to RCSB PDB.

The similarity of the three-dimensional structures of the domains in mouse γC-crystallin and human γD-crystallin is evident from an all residue main chain rmsd of 0.74 Å. The mouse γC-crystallin asymmetric unit contains two molecules with an intermolecular main chain rmsd of 0.41 Å. The main difference between the two chains is that molecule A has a partial disulphide bond between the sulphurs of Cys 18 and Cys 22. At the resolution of the dataset, it is not possible to clearly define the proportion of oxidised to reduced protein and so this has been set to 50/50. Chain B appears to be fully reduced.

Comparison and superposition of the X-ray coordinates of mouse γC-crystallin with human γD-crystallin, bovine γB-crystallin, bovine γF-crystallin and rat γE-crystallin (Figure 1b) allows refinement of the linker alignment (Figure 1a). In all these proteins, the linker takes a sharp turn at the conserved glycine, which supports placement in the longer linkers (γB-crystallins and mouse γCins) of any extra residue in front of the turn.

(g) Comparison of the mouse experimental structure with modelled human structures

Wild-type human γC-crystallin and the C79R mutant did not crystallize. Therefore homology models for human γC-crystallin and the C79R and R36S mutants were built based on the X-ray determined coordinates of chain A of mouse γC-crystallin. In addition the human γD-crystallin R36S cataract mutant was built based on the X-ray structure of human γD-crystallin. The aim was to compare the homology models and the two X-ray structures of mouse γC-crystallin and human γD-crystallin in the light of their biophysical solution properties. In order to facilitate the extrapolation of static X-ray-derived protein-water coordinates to solution behaviour of protein molecules over a short period of time, computational simulations of the interactions between various proteins with water were performed using molecular dynamics. Furthermore, as human γD-crystallin was crystallized under different solution conditions and in a different packing arrangement in the unit cell from mouse γC-crystallin, these simulations removed bias resulting from intermolecular contacts in the crystal lattice. Accordingly, the two sets of X-ray determined coordinates (chain A of mouse γC-crystallin and human γD-crystallin) and the four homology models (human γC-crystallin, C79R and R36S plus human γD-crystallin R36S mutants), were each subjected to 5 ns of room temperature (300 K) molecular dynamics simulation, in explicit water, allowing equilibration of side chain conformations with water at neutral pH and low ionic strength.

Domain pairing geometry—During the simulations, the individual domains from the six γ-crystallins showed very little change in secondary structure. However, there were some changes seen in the structure, with the domain pairing geometry changing with time. The three proteins with phenylalanine at position 56 (mouse γC-crystallin and the two human γD-crystallins) showed a stable interface, with the distance between the Cβ atoms of residues 43 and 144 (the most central residues in the interface) averaging around 5 Å (see Table 3 and figure 5); whilst the proteins with a leucine at this position (human γC-crystallin and its mutants) showed the distance between these atoms increasing greatly during the simulations. Investigation of the position of the side chain of residue 56 during the simulations showed that the phenylalanine stays packed flat against the N-terminal domain β-sheet, whilst the methyl groups at the end of the leucine side chain moved away from the surface of the domain.

TTTDC loop—Analysis of the region between residues 18-22, where the human γC-crystallin has the unusual sequence TTTDC while the mouse γC-crystallin has CSSDC (with the partial disulphide linking the ends), showed that the backbone conformations are very similar, but differ from human γD-crystallin (sequence CSSDH). This difference seems to stem from altered packing in the core of human γD-crystallin due to the sequence change of Leu 5 in place of the more usual phenylalanine, leading to partial rotation of the Trp 42 side chain, giving room for the change in main chain conformation. Another difference is that the atom NE2 of His 22 is in a position to form a long hydrogen bond (3.0 Å) with the OH of Tyr 16 in human γD-crystallin.

Electrostatics—A key feature of γ-crystallins is that their surfaces are covered in ion pairs. The arginine head groups are the most accessible to solvent water and will therefore have a profound effect on the surrounding water, particularly important in the eye lens. Both wild-type human γC and γD-crystallins have pI's close to neutral whereas mouse γC-crystallin and the human C79R mutant are more basic; both the R36S mutations give proteins with more

acidic pIs (Table 4). Examining the pIs of the individual domains showed that the C-terminal domains all have a similar pI (in the range 6.8 - 7.0), close to neutral, whilst the N-terminal domains have a wider range of pI from the acidic (γC WT and both R36S mutants) to the basic (human γC C79R mutant and mouse γC) (Table 4). Nevertheless, when the electrostatic surfaces of the simulated proteins are shown from a common viewpoint, (Figure 6) the human γD-crystallin C-terminal domain can be seen to have a larger negatively charged region than all the γC-crystallins. There is thus no clear correlation between the solubility of a γ-crystallin and its sequence, its overall charge as calculated by its pI, or the overall surface charge distribution based on visual inspection of the atomic structures.

Analysis of the solvent accessibilities of cysteine residues (Table 5) showed that the extra cysteine at position 79 in human γC-crystallin is in an exposed position. The presence of cysteine at position 79 in human γC-crystallin has the effect of breaking up the large area of positive charge that is present in mouse γC and human C79R, on the N-terminal domain near the interface. The absence of the positively charged arginine in the middle of the interface region allows the negatively charged residues of the C-terminal domain to have a larger influence, and changes to the electrostatics on the surface of the protein were observed after 1ns of molecular dynamics. Several charged residues behave differently when comparing the human wild-type γC-crystallin and C79R mutant, presumably adapting to the presence or absence of R79. For example R147 points away from the N-terminal domain (into space) in the wild-type but across the interface in the C79R mutant, hence giving a more positive interface in the mutant.

Dipole moments—Although γ-crystallins carry little overall charge, we also estimated the range of dipole moments of the various solvated protein molecules. Investigation of the dipole moments of the six proteins modelled showed significant differences between the γC and γD-proteins in the direction of their overall dipole moments. For the γD-crystallins, the dipole moments are aligned roughly along the long axis of the molecule from the N-terminal domain to the C-terminal, whilst for the γC-crystallins, the dipole moments are aligned across the inter-domain interface (top row, Figure 7). For each protein, the magnitude of the dipole moments was calculated during the course of the simulations and Table 4 shows the mean (standard deviation) for each protein (and the individual domains), as well as maximum and minimum values. It can be seen that the mean overall dipole moment of the two proteins, which have low solubility - human γC WT (at 134.4 D) and γD R36S mutant (at 346.1 D), differ somewhat from the other proteins studied (that are in the range 204-231 D). The C-terminal domain dipole moment has a larger magnitude (range 235-430 D) than the N-terminal domain (range 58-140 D) in all the proteins simulated, with the C-terminal domains in the γD-crystallins having a larger dipole moment than in the γC-crystallins (Figure 7 and Table 4). The dipole moment from the N-terminal domain of human γC-crystallin wild-type has the smallest magnitude, but with the greatest variation (range 3 – 106D) and hence it experiences much larger changes in orientation than the dipole moments of the other N-terminal domains. The N-terminal domains of the human γC-crystallin C79R and R36S mutants have dipole moments with properties similar to the soluble mouse γC-crystallin.

Discussion

γ-Crystallins are long-lived proteins of the eye lens and are generally characterized by both high stability and high solubility. Interestingly, they contain relatively high levels of cysteine residues. This would seem to be a liability for the lens, as cysteine is reactive and prone to modification; cysteine oxidation has often been related to cataract formation.[30,31] Presumably cysteine residues in γ-crystallins have important conserved roles in the lens, perhaps contributing to protein stability, to intermolecular interactions that have an impact on the refractive index, or scavenging of free radicals.

Human γC-crystallin is the major γ-crystallin of the young human lens and hence is a major component of the oldest, longest-lived cells of the lens central region. By contrast, the γCs are less prominent in the short-lived mouse lens. Human γC-crystallin has a solvent exposed cysteine at position 79, while the mouse ortholog has arginine at this position, similar to other well-characterized γ-crystallins, such as human γD-crystallin. While recombinant mouse γC-crystallin had the high solubility typical of mammalian γ-crystallins, human γC-crystallin proved to have greatly reduced solubility. Replacement of Cys 79 with arginine dramatically increased the solubility of the human protein, suggesting that cysteine at this position in wild-type human γC-crystallin may be responsible for the altered solution properties.

However, an unexpected observation showed that exposure of Cys 79 alone does not necessarily confer insolubility. Introduction into human γC-crystallin of the R36S mutation (which is associated with congenital cataract in human γD-crystallin and is topologically equivalent to C79) rather than further decreasing solubility actually increased solubility to levels similar to mouse γC or human γD-crystallins. This suggests that the interaction properties of γ-crystallins are determined more by the surface networks of ion pairs and hydrogen bonds than by a single surface cysteine.

What might be the selective advantages of the differences in γC-crystallins between human and mouse, bearing in mind that it is just one component in a complex milieu of other crystallins? The soft, accommodating human lens requires a lower protein density and lower refractive index than the mouse lens. The much greater life span of humans (and other primates) compared with mice may also have selected for changes that enhanced stability, even at the expense of solubility. In contrast, in the short-lived mouse with a very dense lens there may have been selection for solubility at the expense of stability. It is interesting that the mouse has acquired a new γC-crystallin with an insertion in the linker, although what new property this brings to the protein is not yet evident.

One advantage of the high conformational stability of human γC-crystallin may be in protecting buried thiols (Table 5), since unfolding could expose cysteines to oxidation and permit formation of intermolecular disulphide bonds. Exposed cysteines are also potentially vulnerable to oxidation in all γ-crystallins. One way of preventing these from forming intermolecular bonds would be to favour intramolecular disulphides instead. 3D structures can address whether oxidation can lead to local disulphide formation without disruption of the tertiary structure, hence preserving protein solubility. Indeed, from the mouse γC-crystallin structure analysis it became clear that a small loop that comprises the sequence CSSDCP readily forms an intramolecular disulphide with little change to the overall structure. This is similar to the same sequence found in bovine γB-crystallin where a ‘harmless’ disulphide was also found in aged crystals.[32,33] Human γA and γB-crystallins might be able to do the same thing (Figure 1a). However, human γC-crystallin has the sequence TTTDCP so the sole cysteine (C 22) cannot make a local disulphide under oxidizing conditions. Elsewhere, oxidation of Cys 79 with adjacent Cys 78 could only be envisaged after the protein had unfolded. Thus, in spite of its stability, human γC still has vulnerabilities to oxidative damage with disease and ageing.

The X-ray structure of mouse γC-crystallin and the ensuing molecular simulations of several molecular models also addressed the geometry of the domain pairing. In humans, the presence of Leu, in place of Phe, at position 56 in the inter-domain interface, seems to lead to greater motion of the domains relative to each other.

Radicalisation or oxygenation of buried thiols could trigger the proteins to unfold, allowing intermolecular disulphide formation and aggregation. Based on a comparison between young and old human lenses, the most abundant age-related cysteine modification in human γC-

crystallin is methylation at accessible thiols.[34,35] This suggests that while conformational stability may protect buried cysteines (32, 78, 108) the most accessible thiols (22 and 79) in the aging lens may be protected by programmed chemical modification.

Although cysteines are well conserved and may have important roles in γ-crystallin function, our data suggest that they do not have the most significant role in determining the biophysical behavior of these proteins. Instead, changes in the charge properties of the molecular surface seem to be much more important. Inspection of the motif sequence alignment shows the acidic residue (indicated in red in Figure 1a) within the TTTDC loop of human γC is conserved in most motifs of all γ-crystallins. In the N-terminal domain this acidic residue (Asp 21) in motif 1 ion pairs with Arg 79 of motif 2 in all X-ray structures of γ-crystallins, while the topologically equivalent acidic in motif 2 (Glu 61) ion pairs with Arg 36 in motif 1 (topologically equivalent to Arg 79). The low solubility of the wild-type human γC-crystallin and the crystallization prone R36S mutant of human γD-crystallin, correlate with the loss of arginine in one or other of these ion pairs suggesting that solubility is strongly effected by an isolated acidic side chain (Figure 8). However, what is remarkable is that the human γC-crystallin sequence can have high solubility if both acidic groups are unpaired. This may be explained by having a single 'free' acid perturbing the dipole moment of the domain, whereas having an unpaired acid on both sides of the domain 'cancels out' the perturbation to the dipole moment. The magnitude of the calculated dipole moments (Table 4) support this explanation, as the dipole moments on the N-terminal domain of both the human γC-crystallin C79R (Asp 21 and Glu 61 each ion-paired) and human γC-crystallin R36S (Asp 21 and Glu 61 each unpaired) are similar, both to each other and to the mouse γC-crystallin N-terminal domain (Asp 21 and Glu 61 each ion-paired). As wild-type human γC-crystallin is the only γ-crystallin in this study in which the N-terminal domain has a dipole moment with a small magnitude and highly variable orientation, this is the likely cause of this protein's insolubility. Whilst a low macroscopic dipole can contribute to low solubility due to reduced electrostatic repulsion, it is less obvious how the direction of the dipole could also play a role. As the solution behaviour of γ-crystallins as a class has been characterized as exhibiting attractive interactions [4], one can only speculate that the orientation of the macroscopic dipole influences the orientation of their solution encounters, in a way that supports local attractive interactions. Furthermore, the large magnitude and constant orientation of the dipole moment for the complete human γD-crystallin R36S molecule may favour the nucleation of this molecule into its crystal lattice, which leads to childhood cataract.

These observations, together with the surface point mutations that are associated with congenital cataract, indicate that the sequences of γ-crystallins are optimized for high solubility, and that minor changes to the surface, particularly changes associated with intermotif ion pairs can dramatically alter solution interaction properties. Bearing in mind that a common point mutation, P23T which drastically reduces the solubility of human γD-crystallin, is also in the region of the inter-motif ion pair Asp 21-Arg 79, this region of the molecular surface of γ-crystallins appears to have a profound effect on the solution interaction properties. In human lens the proteins become increasingly insoluble with age. The most common crystallin post-translational modification is deamidation, and this increases with age and with the level of crystallin water insolubility.[35] Thus the evidence from both congenital cataract point mutations, and the age-related post-translational changes, point to γ-crystallin solubility being sensitive to surface charge effects.

In the eye lens, the various members of the γ-crystallins are in concentrated solutions with the α- and β-crystallins and it is their combined solution interaction properties that create the required refractive index medium.[36] It is assumed that the solubility of purified γ-crystallins has some bearing on this crucial biophysical attribute. For example, the orientation of a protein such as human γC-crystallin will be less affected by the negatively charged alpha-crystallin,

than the human γD-crystallin, which will take up a preferred orientation in line with its large dipole moment, with the other γC-crystallins exhibiting intermediary behaviour.

The vertebrate lens arose in an aquatic ancestor which required a very high refractive index due to the lack of an air/cornea interface.[3] Fish lenses have high levels of γ-crystallins that are remarkable in having high levels of polarizable sulphur containing residues on their surfaces.[37] These crystallins are synthesized early in life and form the very high refractive index core.[38] A particular role of γ-crystallins may have been to allow the formation of a very high concentration of protein while maintaining transparency. This is much less useful in the softer, accommodating lenses of diurnal, terrestrial vertebrates.[39] Indeed, in birds, γ-crystallins have largely been replaced by the enzyme crystallins.[1] With the radiation of mammals, it seems that a surviving γ-crystallin founder gene has undergone several duplications resulting in the present day γA-F crystallin genes.[8,40] As mammals have successfully occupied a wide variety of habitats, their evolution required rapid adaptation of their optics.[2] Gene regulatory mechanisms have already been discovered that appear to facilitate this such as taxon restricted expression of enzyme crystallins,[41] or the switching off of γ-crystallin genes in humans.[9] Here we show that species-specific alterations in critical regions of the surface of the protein structure can also result in major modifications in the interaction properties of γ-crystallins and that different species can rapidly fine-tune their crystallins for the specific requirement of their environment.

Materials and Methods

(a) Cloning, mutagenesis and expression of human and mouse γ-crystallins

(i) Human—The human γC-crystallin coding region, a generous gift from Dr N.H. Lubsen, was isolated from a human lens cDNA library by PCR using the primers:

forward: GCATATGGGGAAGATCACCTTCTATG

reverse: AGGATCCAAAATGGGAAATTGGTAGTG

The PCR product was cloned in the pGEM-T Easy vector (Promega) and sequenced. The insert was then inserted NdeI/BamHI in the pET3a vector. The pET3a recombinants were transformed into BL21(DE3) pLys (Novagen) and grown and induced as described previously.[42] Cells were harvested by centrifugation and resuspended in 10 ml BugBuster (Novagen) containing a protease inhibitor (Pefabloc, Merck). Cells were lysed by sonification (3 × 45 sec), Benzonase (Novagen) was then added at the level of 1 μl/ml of lysate which was then incubated for 30 min at room temperature. The lysate was cleared by centrifugation for 30 min at 18,000 rpm. SDS-Page showed that there was a prominent protein ~21 kDa in the soluble fraction. Prior to mass analysis protein samples were desalted using a BioSpin (BioRad). Mass analysis using a Micromass Platform API electrospray mass spectrometer, performed as described earlier,[43] identified the protein as human γC-crystallin from its measured mass close to 20,748 Da.

The human γC-crystallin C79R and R36S mutants were prepared using the QuikChange XL Site-Directed Mutagenesis Kit (Stratagene) and the mutations confirmed by DNA sequencing. The mutant proteins were expressed as for wild-type and were both found in the soluble fractions with molecular masses measured by electrospray mass spectrometry consistent with the calculated masses of 20,787 Da for C79R and 20,678 da for R36S.

(ii) Mouse—Full-length clones for mouse γC and γCins were identified from the NEIBank database[5] (Wistow 2006). Both were amplified and cloned into pET31b+ using primers MGCNde: CATATGGGGAAGATCACCTTCTTCGAG and MGCXho: CTCGAGTATTTTAGTATAAATCTACCAC.

Recombinant proteins were expressed as described for the human clones. The measured masses of mouse γC-crystallin and γCins were consistent with their calculated masses of 20,786 and 20,914 respectively and were found in the soluble fraction.

(b) Estimation of endogenous γ-crystallin levels by mass spectrometry

Soluble protein from whole decapsulated lenses of 18-28 day old mice was isolated, digested with trypsin, and analyzed by two-dimensional liquid chromatography (cation exchange in the first dimension and reverse phase in the second dimension) as previously described,[35] except data-dependant MS/MS spectra were collected using an LTQ linear ion trap (ThermoFinnigan, San Jose, CA). The MS/MS spectra resulted in a total of 222,935 DTA files that were searched using SEQUEST (version 2.7, rev 12, part of Bioworks 3.2, ThermoFinnigan) using a mouse only subset of the SwissProt databases (12,000 entries, downloaded Dec. 7, 2006). Oxidized Met or Trp and protein N-terminal acetylation were specified as differential modifications, and cysteines were statically modified (+57.02 Da) due to reduction and alkylation with iodoacetamide. Filtering of results and mapping of peptide sequences to protein sequence entries was done using DTASelect (version 1.9, The Scripts Research Institute[44], using the default criteria for 1+, 2+, and 3+ peptides of having Xcor value from SEQUEST in excess of 1.8, 2.5, and 3.8, respectively, and DeltaCN values greater than 0.08. DTASelect results were further processed with a PERL program written in-house to eliminate multiple counting of shared peptides and tally counts of unique peptides (peptides appearing in only one protein sequence in the searched protein database). Counts of shared peptides in mouse γB and γC-crystallins were apportioned to their respective proteins on the basis of the relative counts of spectra assigned to unique peptides in each protein to estimate the net total spectral counts associated with each. The relative abundance of endogenous γC and γCins in mouse lens was measured by isolation of the mouse γC and γCins by gel filtration and cation exchange, followed by mass measurement of the coeluting species using electrospray ionization mass spectrometry as previously described.[24] The ratio of γC and γCins in the coeluting proteins was then determined by integrating the areas of each form in the deconvoluted mass spectrum.

(c) Purification of recombinant proteins

The soluble fractions from *Escherichia coli* whole cell lysate containing wild-type human γC-crystallin, human γC C79R mutant, mouse γC-crystallin and mouse γCins-crystallin were each subjected to anion exchange chromatography using a Q Sepharose (HiPrep 16/10 QFF) column (GE Healthcare) run in 25 mM Tris-HCl, pH 8.5, 1 mM DTT (Buffer A) with a linear gradient (0 – 70 % over 10 column volumes) of Buffer B (Buffer A/1 M NaCl). The peaks were collected and analysed by SDS-PAGE. Candidate peaks were also run on a reversed phase column and analysed using electrospray mass spectrometry. In the case of wild-type human γC-crystallin, the only protein peak found to contain the crystallin eluted at 39 % Buffer B. As protein recovery was very low compared with other γ-crystallin purifications, in an attempt to increase the yield of protein, anion exchange chromatography was performed under dissociating conditions. Solid urea was added to the lysate to make a final concentration of 6 M and this was then loaded onto a Q Sepharose column run in 25 mM Tris-HCl, pH 7.5, 6 M urea, 1 mM DTT (Buffer C) with a linear gradient of 0 – 70 % over 10 column volumes of Buffer D (Buffer C/1 M NaCl). Human γC-crystallin eluted at 21.3% buffer C but there was no improvement in yield. Therefore, solid ammonium sulphate was added to the original soluble cell lysate to give a 50 % saturated solution that precipitated the bulk of the ~21 kDa protein and the precipitate spun down. A saturated protein solution, around 50 mg ml^{-1} as estimated by BioRad Assay, was achieved by dissolving the pellet in 10 mM sodium phosphate, pH 7.0. A 250 μl sample of this concentrated protein solution was chromatographed on a Superose 12 HR 10/30 column equilibrated and run in 200 mM ammonium acetate, pH 6.8, 2 mM DTT. The elution peak in which the A280 exceeded the A260 was pooled and gave a single band on an SDS PAGE gel.

The human γC-crystallin C79R and R36S mutants, and the mouse γC-crystallin eluted after the breakthrough peak but before the start of the salt gradient. Each protein peak was homogeneous as judged by SDS-page. The proteins were each concentrated to around 10 mg/ml and chromatographed on a Superdex 75 10/300 GL, equilibrated and run in 50 mM MES, 200 mM NaCl, pH 6.0, 2 mM DTT, as final purification steps. The mouse γCins-crystallin also eluted before the salt gradient but was further purified. The protein peak from anion-exchange was desalted on a HiPrep 26/10 Desalting column in 50 mM MES, pH 6.0, 1 mM DTT. The desalted protein was then loaded onto a Mono S 10/100 GL column in 50 mM MES, pH 6.0, 2 mM DTT/1 M NaCl when the protein eluted at about 7% Buffer B. The protein peak was then concentrated and subjected to gel filtration as above.

(d) Solubility measurements

Human wild-type γC-crystallin could be isolated from the gel filtration column in 200 mM ammonium acetate at a concentration of around 0.7 mg/ml and remained in solution for long periods at 4°C. The protein was further concentrated using a Vivaspin 500 10MWCO to a limiting concentration of around 5 mg/ml (with losses of at least 50% of the starting protein amount, as compared with an almost complete loss when attempting to concentrate in the absence of the 200 mM ammonium acetate). Large volumes of human γC-crystallin R36S mutant, human γC-crystallin C79R mutant mouse γC-crystallin, and mouse γCins were concentrated in a Vivacell 70 10,000 MWCO (Vivascience AG) followed by Vivaspin 500 10,000 MWCO for the final concentrations. Protein concentrations were determined using Bio-Rad Protein Assay on suitably diluted samples.

(e) Stability measurements

A stock solution of 6 M Gdn HCl in 10 mM sodium phosphate, pH 7.0 was diluted with the phosphate buffer to produce solutions from 0 to 6 M GdnHCl (at 1 M intervals or from 0 to 4.2 at 0.2 M intervals). Aliquots of solutions of human γC-crystallin, human γC-crystallin C79R, mouse γC-crystallin and mouse γCins-crystallin were then added to each GdnHCl dilution to give solutions with a final crystallin concentration of 0.02 mg/ml. The crystallin/GdnHCl/phosphate solutions were then incubated at 37°C for 6 hours. Fluorescence spectra were collected with a Hitachi F-2500 spectrophotometer with excitation at 280 nm. The slits were set at 10 nm and the emission spectra were collected from 300 to 370 nm in a 1 cm path length cell (Helma Ltd). The emission data at 350 nm were used to compare the unfolding behaviour of the crystallins.

(f) Screening for crystallization

Crystallization screens (Hampton 1 and 2, MDL 1 and 2, Nextal methane pentane diol, ammonium sulphate, PEG/Ion) were set up using a Perkin Elmer Multiprobe II Plus HTEX Robotic Liquid Handling System. Vapour diffusion sitting drops were set up in 96 well plates (Greiner) with 0.5 μl protein solution + 0.5 μl screen solution. For each protein the concentration used for the crystallisation trials was optimised by setting up a trial screen and then increasing or decreasing the protein concentration in order to obtain precipitates in approximately one third of the test screen conditions for Hampton screens 1 and 2. The concentrations used were: human γC-crystallin ∼5 mg ml^{-1}, human mutant C79R ∼10 mg ml^{-1}, and mouse γC-crystallin and γCins ∼6 mg ml^{-1}.

(g) Crystallization of mouse γC-crystallin

Initial crystals were obtained from a protein concentration of 6 mg/ml with a well solution consisting of 0.2 M sodium acetate, 0.1 M sodium cacodylate pH 6.5 and 30% PEG 8000. Conditions were optimized by scanning 1 + 1 μl hanging drops in Nextal 24 well crystallization trays. Well-shaped crystals appeared within 2 days.

(h) Crystal structure determination

A dataset comprising 180 degrees (1 degree per oscillation) collected to a resolution of 1.5 Å was collected at ESRF on beam line ID23-2. No additional cryoprotection was needed. The dataset was integrated with MOSFLM[45] and scaled using the SCALA[46] from the CCP4 suite; [47] all data were integrated, but during refinement, the resolution was cut back to 1.9 Å.

(i) Human homology models and molecular dynamics simulations

The mouse γC-crystallin coordinates were used to generate homology models of human γC-crystallin wild-type and human γC-crystallin C79R and R36S mutants. For comparison, human γD-crystallin coordinates were used to generate a homology model of the γD-crystallin R36S mutant (which forms crystalline cataracts). These simple mutations were undertaken using COOT,[48] with the most probable rotamer chosen, unless this clashed with other atoms. In order to simulate the properties of the protein molecules, solvated in water, over time, the homology model structures plus the experimentally determined mouse γC-crystallin and human γD-crystallin structures were solvated, with TIP3P water molecules and parameterized using the parm99SB force field[49] and using the LEAP program from AMBER.[50] These solvated molecules then underwent energy minimization, before being subjected to molecular dynamics, using the PMEMD program. The molecular dynamics involved a start up protocol, where firstly the solvent molecules were warmed from 50 K to 300 K and allowed to adjust to the protein coordinates; the temperature was then reset to 50 K and all atoms warmed to 300 K. This temperature was held for 100 ps for a brief equilibration of the solvated system, before data for analysis was collected, over 5ns from this point.

The program PMEMD was used for both energy minimization and molecular dynamics. Representative structures, after 4 ns of molecular dynamics, were used for electrostatic mapping of the proteins, using the program APBS.[51] Dipole moments for the domains in the proteins were calculated (at 2.5 ps intervals over the full 5 ns of simulation time) and displayed using the program VMD.[52] Other structural figures were produced using the program PYMOL. [53]

Acknowledgements

We are grateful to our colleagues for helpful advice and discussions. X-ray data were collected at the ESRF Beamline ID23-2, Grenoble. AGP, OAB and CS are grateful to the Medical Research Council (London) for financial support. Work for this study by KW and GW was carried out under NEI intramural project Z01 EY000255 and Z01 EY000320. The work of LLD and PAW was supported by National Institutes of Health (USA) grants NEI EY007755 and EY10572.

References

1. Wistow G. Lens crystallins: gene recruitment and evolutionary dynamism. Trends Biochem Sci 1993;18:301–306. [PubMed: 8236445]
2. Wistow, G. Molecular biology and evolution of crystallins: Gene recruitment and multifunctional proteins in the eye lens. R.G. Landes Company; Austin TX: Springer; New York: 1995.
3. Land, MF.; Nilsson, DE. Animal Eyes. Oxford: Oxford University Press; 2002.
4. Bloemendal H, de Jong W, Jaenicke R, Lubsen NH, Slingsby C, Tardieu A. Ageing and vision: structure, stability and function of lens crystallins. Progress Biophys & Mol Biol 2004;86:407–485.
5. Wistow G. The NEIBank project for ocular genomics: Data-mining gene expression in human and rodent eye tissues. Prog Ret & Eye Res 2006;25:43–77.
6. Lampi KJ, Ma Z, Shih M, Shearer TR, Smith JB, Smith DL, David LL. Sequence analysis of βA3, βB3, and βA4 crystallins completes the identification of the major proteins in young human lens. J Biol Chem 1997;272:2268–2275. [PubMed: 8999933]
7. Robinson NE, Lampi KJ, Speir JP, Kruppa G, Easterling M, Robinson AB. Quantitative measurement of young human eye lens crystallins by direct injection Fourier transform ion cyclotron resonance mass spectrometry. Mol Vis 2006;12:704–711. [PubMed: 16807530]

8. Wistow G, Wyatt K, David L, Gao C, Bateman O, Bernstein S, Tomarev S, Segovia L, Slingsby C, Vihtelic T. Gamma N-crystallin and the evolution of the beta gamma-crystallin superfamily in vertebrates. Febs J 2005;272:2276–2291. [PubMed: 15853812]
9. Brakenhoff RH, Aarts HJM, Reek FH, Lubsen NH, Schoenmakers JGG. Human γ-crystallin genes: A gene family on its way to extinction. J Mol Biol 1990;216:519–532. [PubMed: 2258929]
10. Kumaraswamy VS, Lindley PF, Slingsby C, Glover ID. An eye-lens protein-water structure: 1.2 Å resolution structure of γB-crystallin at 150 K. Acta Crystallogr 1996;D52:611–622.
11. Chirgadze YN, Driessen HPC, Wright G, Slingsby C, Hay RE, Lindley PF. Structure of the bovine eye lens γD (γIIIb)-crystallin at 1.95 Å. Acta Crystallogr 1996;D52:712–721.
12. Basak A, Bateman O, Slingsby C, Pande A, Asherie N, Ogun O, Benedek GB, Pande J. High-resolution x-ray crystal structures of human γD crystallin (1.25Å) and the R58H mutant (1.15Å) associated with aculeiform cataract. J Mol Biol 2003;328:1137–1147. [PubMed: 12729747]
13. Norledge BV, Hay RE, Bateman OA, Slingsby C, Driessen HPC. Towards a molecular understanding of phase separation in the lens: A comparison of the X-ray structures of two high Tc γ-crystallins, γE and γF, with two low Tc γ-crystallins, γB and γD. Exp Eye Res 1997;65:609–630. [PubMed: 9367641]
14. Artero JB, Hartlein M, McSweeney S, Timmins P. A comparison of refined X-ray structures of hydrogenated and perdeuterated rat gamma E-crystallin in H2O and D2O. Acta Crystallogr Section D 2005;61:1541–1549. [PubMed: 16239733]
15. Wu ZR, Delaglio F, Wyatt K, Wistow G, Bax A. Solution structure of gamma S-crystallin by molecular fragment replacement NMR. Prot Sci 2005;14:3101–3114.
16. 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]
17. Evans P, Wyatt K, Wistow GJ, Bateman OA, Wallace BA, Slingsby C. The P23T cataract mutation causes loss of solubility of folded γD-crystallin. J Mol Biol 2004;343:435–444. [PubMed: 15451671]
18. 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;44:2491–2500. [PubMed: 15709761]
19. Heon E, Priston M, Schorderet DF, Billingsley GD, Girard PO, Lubsen N, Munier FL. The gamma-crystallins and human cataracts: A puzzle made clearer. Amer J Hum Gen 1999;65:1261–1267.
20. Stephan DA, Gillanders E, Vanderveen D, Freas-Lutz D, Wistow G, Baxevanis AD, Robbins CM, VanAuken A, Quesenberry MI, Bailey-Wilson J, Juo SHH, Trent JM, Smith L, Brownstein MJ. Progressive juvenile-onset punctate cataracts caused by mutation of the gamma D-crystallin gene. Proc Natl Acad Sci USA 1999;96:1008–1012. [PubMed: 9927684]
21. Kmoch S, Brynda J, Asfaw B, Bezouška K, Novák P, Rezácová P, Ondrová L, Filipec M, Sedlácek J, Elleder M. Link between a novel human γD-crystallin allele and a unique cataract phenotype explained by protein crystallography. Hum Molec Gen 2000;9:1779–1786. [PubMed: 10915766]
22. Nandrot E, Slingsby C, Basak A, Cherif-Chefchaouni M, Benazzouz B, Hajaji Y, Boutayeb S, Gribouval O, Arbogast L, Berraho A, Abitbol M, Hilal L. Gamma-D crystallin gene (CRYGD) mutation causes autosomal dominant congenital cerulean cataracts. J Med Gen 2003;40:262–267.
23. Srikanthan D, Bateman OA, Purkiss AG, Slingsby C. Sulfur in human crystallins. Exp Eye Res 2004;79:823–831. [PubMed: 15642319]
24. Ueda Y, Duncan MK, David LL. Lens proteomics: The accumulation of crystallin modifications in the mouse lens with age. Invest Ophthalmol Vis Sci 2002;45:205–215. [PubMed: 11773033]
25. Fu L, Liang JJN. Spectroscopic analysis of lens recombinant beta B2-and gamma C-crystallin. Mol Vis 2001;7:178–183. [PubMed: 11483894]
26. Kosinski-Collins MS, King J. In vitro unfolding, refolding and polymerization of human γD crystallin, a protein involved in cataract formation. Prot Sci 2003;12:480–490.
27. Kosinski-Collins MS, Flaugh SL, King J. Probing folding and fluorescence quenching in human gamma D crystallin Greek key domains using triple tryptophan mutant proteins. Prot Sci 2004;13:2223–2235.

28. Chen JJ, Flaugh SL, Callis PR, King J. Mechanism of the highly efficient quenching of tryptophan fluorescence in human gamma D-crystallin. Biochemistry 2006;45:11552–11563. [PubMed: 16981715]
29. Read RJ. Pushing the boundaries of molecular replacement with maximum likelihood. Acta Crystallogr 2001;D57:1373–1382.
30. Harding JJ. Viewing molecular mechanisms of ageing through a lens. Ageing Res Reviews 2002;1:465–479.
31. Truscott RJW. Age-related nuclear cataract - oxidation is the key. Exp Eye Res 2005;80:709–725. [PubMed: 15862178]
32. Wistow G, Turnell B, Summers L, Slingsby C, Moss D, Miller L, Lindley P, Blundell T. X-ray analysis of the eye lens protein γ-II crystallin at 1.9 Å resolution. J Mol Biol 1983;170:175–202. [PubMed: 6631960]
33. Najmudin S, Nalini V, Driessen HPC, Slingsby C, Blundell TL, Moss DS, Lindley PF. Structure of the bovine eye lens protein γB(γII)-crystallin at 1.47 Å. Acta Crystallogr 1993;D49:223–233.
34. Lapko VN, Smith DL, Smith JB. Methylation and carbamylation of human gamma-crystallins. Prot Sci 2003;12:1762–1774.
35. Wilmarth PA, Tanner S, Dasari S, Nagalla SR, Riviere MA, Bafna V, Pevzner PA, David LL. Age-related changes in human crystallins determined from comparative analysis of post-translational modifications in young and aged lens: Does deamidation contribute to crystallin insolubility? J Proteome Res 2006;5:2554–2566. [PubMed: 17022627]
36. Tardieu A, Vérétout F, Krop B, Slingsby C. Protein interactions in the calf eye lens: interactions between β-crystallins are repulsive whereas in γ-crystallins they are attractive. Eur Biophys J 1992;21:1–12. [PubMed: 1516556]
37. Chuang MH, Pan FM, Chiou SH. Sequence characterization of gamma-crystallins from lip shark (Chiloscyllium colax): Existence of two cDNAs encoding gamma-crystallins of mammalian and teleostean classes. J Prot Chem 1997;16:299–307.
38. Kröger RHH, Campbell MCW, Munger R, Fernald RD. Refractive index distribution and spherical aberration in the crystalline lens of the African cichlid fish *Haplochromis burtoni*. Vision Res 1994;34:1815–1822. [PubMed: 7941384]
39. Pierscionek BK. Refractive index contours in the human lens. Exp Eye Res 1997;64:887–893. [PubMed: 9301469]
40. Lubsen NH, Aarts HJM, Schoenmakers JGG. The evolution of lenticular proteins: the β- and γ-crystallin super gene family. Prog Biophys Molec Biol 1988;51:47–76. [PubMed: 3064189]
41. Piatigorsky J, Wistow G. The recruitment of crystallins: new functions precede gene duplication. Science 1991;252:1078–1079.
42. Bateman OA, Sarra R, van Genesen ST, Kappé G, Lubsen NH, Slingsby C. The stability of human acidic β-crystallin oligomers and hetero-oligomers. Exp Eye Res 2003;77:409–422. [PubMed: 12957141]
43. Bateman OA, Lubsen NH, Slingsby C. Association behaviour of human βB1-crystallin and its truncated forms. Exp Eye Res 2001;73:321–331. [PubMed: 11520107]
44. Tabb DL, McDonald WH, Yates JR. DTASelect and contrast: Tools for assembling and comparing protein identifications from shotgun proteomics. J Proteome Res 2002;1:21–26. [PubMed: 12643522]
45. Leslie AGW. Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography, No. 26. 1992
46. Evans, PR. “Data reduction”, Proceedings of CCP4 Study Weekend, 1993, on Data Collection & Processing. p. 114-122.
47. Collaborative Computational Project, Number 4. The CCP4 Suite: Programs for Protein Crystallography. Acta Crystallogr 1994;D50:760–763.
48. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr 2004;D60:2126–2132.
49. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins-Structure Function and Bioinformatics 2006;65:712–25.

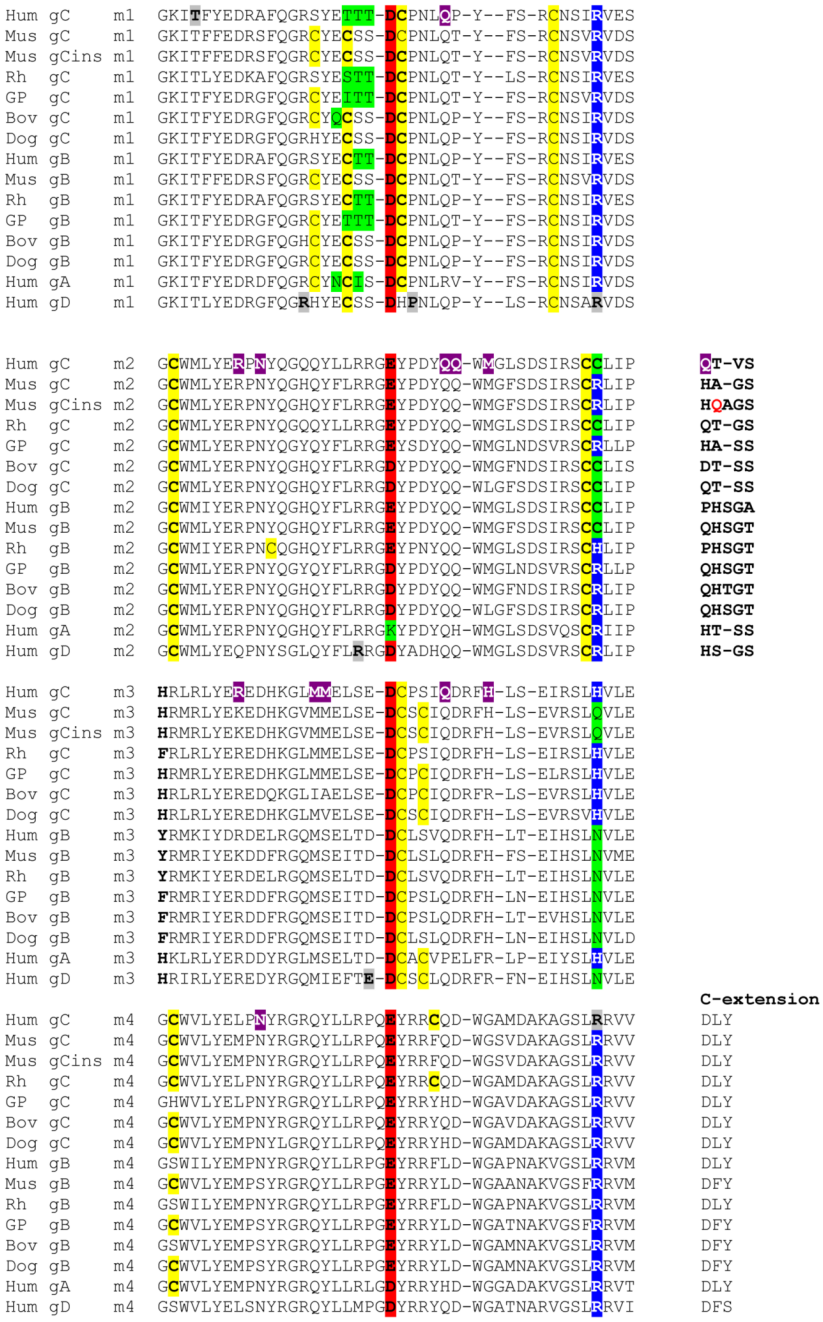
50. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber biomolecular simulation programs. J Computat Chem 2005;26:1668–1688.

51. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci USA 2001;98:10037–1004. [PubMed: 11517324]

52. Humphrey W, Dalke A, Schulten K. VMD - Visual Molecular Dynamics. J Molec Graphics 1996;14:33–38.

53. DeLano, WL. The PyMOL Molecular Graphics System. DeLano Scientific; San Carlos, CA, USA: 2002. http://www.pymol.org

54. Hubbard, SJ.; Thornton, JM. 'NACCESS', Computer Program. Department of Biochemistry and Molecular Biology, University College; London: 1993.



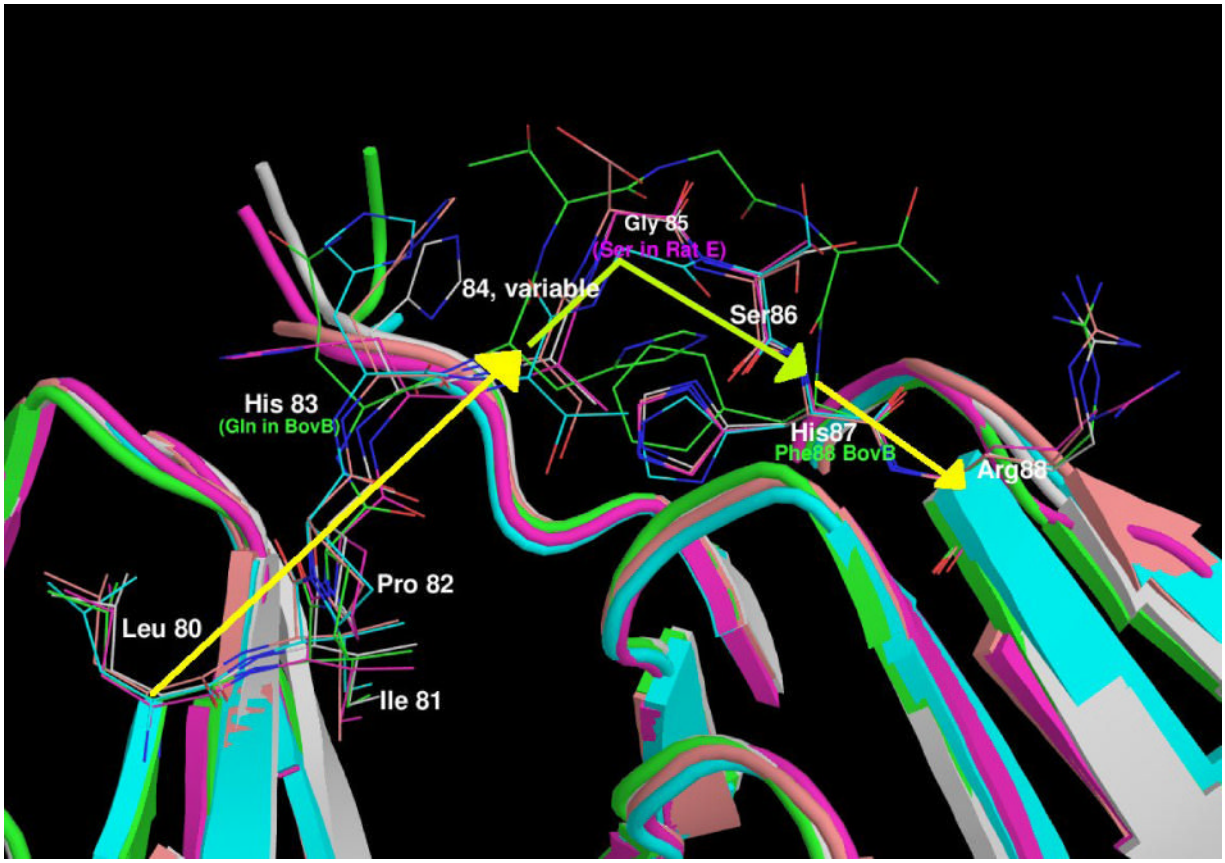


Figure 1.

(a) Motif alignment of selected γ-crystallins.

The sequence of human γC-crystallin [GI:10518338] is aligned with other known mammalian γC-crystallins (bovine [GI:600873], rhesus monkey [EF426305], guinea pig [EF426306], dog [EF426309], C57Bl/6 mouse [EF426301] and C57Bl/6 mouse γCins [EF426302], and with some other members of the γ-crystallin family (human γA [EF426311], human γB [GI:117461], bovine γB [GI:162918], C57Bl/6 mouse γB [EF426303], rhesus monkey γB [EF426304], guinea pig γB [EF426307] and dog γB [EF426308]. The two motifs from the N-terminal domain are aligned above the two from the C-terminal domain. Residues referred to in the text are highlighted. The residues highlighted in red and blue in each motif form an inter-motif ion pair, with the acidic residue from one motif closely associated with the basic residue in the other. Both the basic residues on the N-terminal domain are altered in proteins that have solubility problems: in wild-type human γC-crystallin the motif 2 Arg is replaced by Cys and in a human γD-crystallin cataract mutant, the motif 1Arg is Ser. Both these changes disrupt the tight ion pair and leave an unpartnered acidic side chain. The occurrence of an amide containing side chain in motif 3 does not seem to affect the interaction, with the amide group hydrogen bonding to the acidic side chain in motif 4.

Abbreviations: Hum - Human, Mus - Mouse, Rh - Rhesus Monkey, GP - Guinea Pig, Bov - Bovine.

Colour Key

D and R: Conserved Acid and Basic residues forming inter-motif ion-pair.
C - Cysteine Residues
N - Non-consensus residue
Q - Extra residue in Mouse gC insert
R - Site of cataract causing point mutations in human sequences
Hum gC sites: T5P and R168W
Hum gD sites: R14C, P23T, R36S, R58H & E106A
Hum gS sites: G18V
N - Hum gC site modified on aging

(b) A view of the linkers of all the γ-crystallins with solved structures.
The domains are shown in cartoon format, N-terminal domain to the left and the C-terminal domain on the right with the intervening residues shown in stick format. The proteins are coloured as follows: Mouse γC in pink, human γD in grey, bovine γB in green, bovine γF in turquoise and rat γE in orange. It can be seen that the relatively conserved LIPHX section of the intervening sequence continues from β-strand D2 of the N-terminal domain, whilst the SHR section continues into β-strand A3 in the C-terminal domain. In all the proteins, the consensus G (S in the case of Rat γE) separates the two ‘straight’ parts of the linker with a sharp turn. The presence of the extra T in bovine γB pushes the C-terminal part of the linker away, but in all the proteins it is the consensus glycine where the linker takes a sharp turn. Therefore, structurally the gap in the shorter sequences lies in the N-terminal end of the linker before the location of the turn.

```
Mus gC  LIPHA-GSHR
Hum gD  LIPHS-GSHR
Bov gB  LIPQHTGTFR
Bov gF  LIPHT-GSHR
Rat gE  LIPHS-SSHR
```

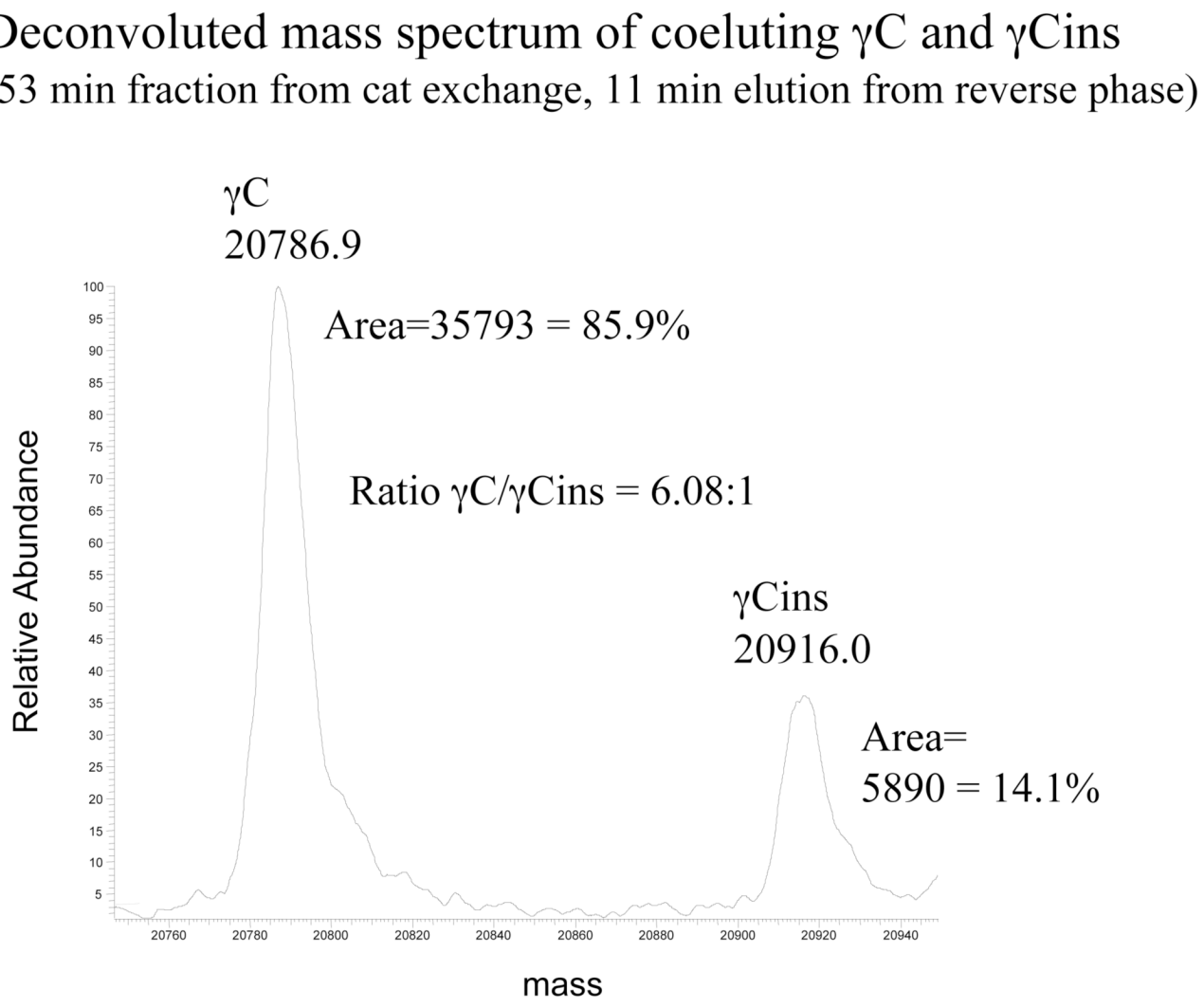
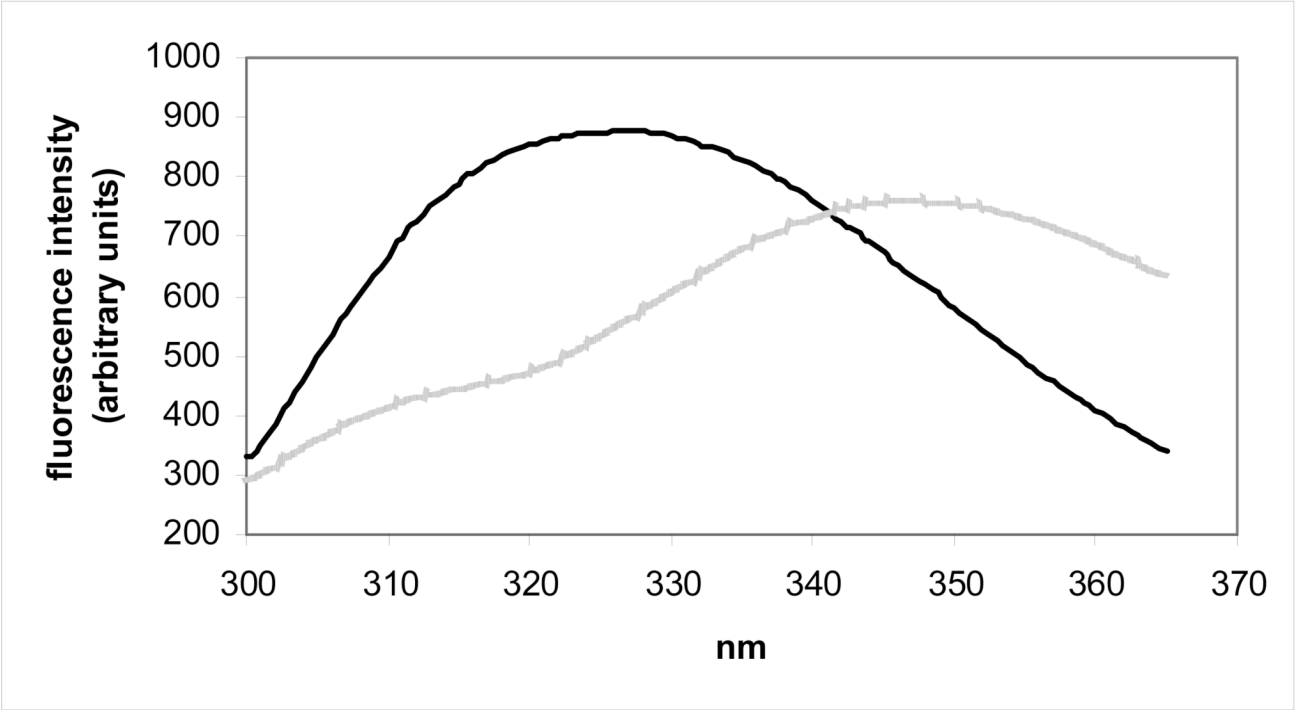
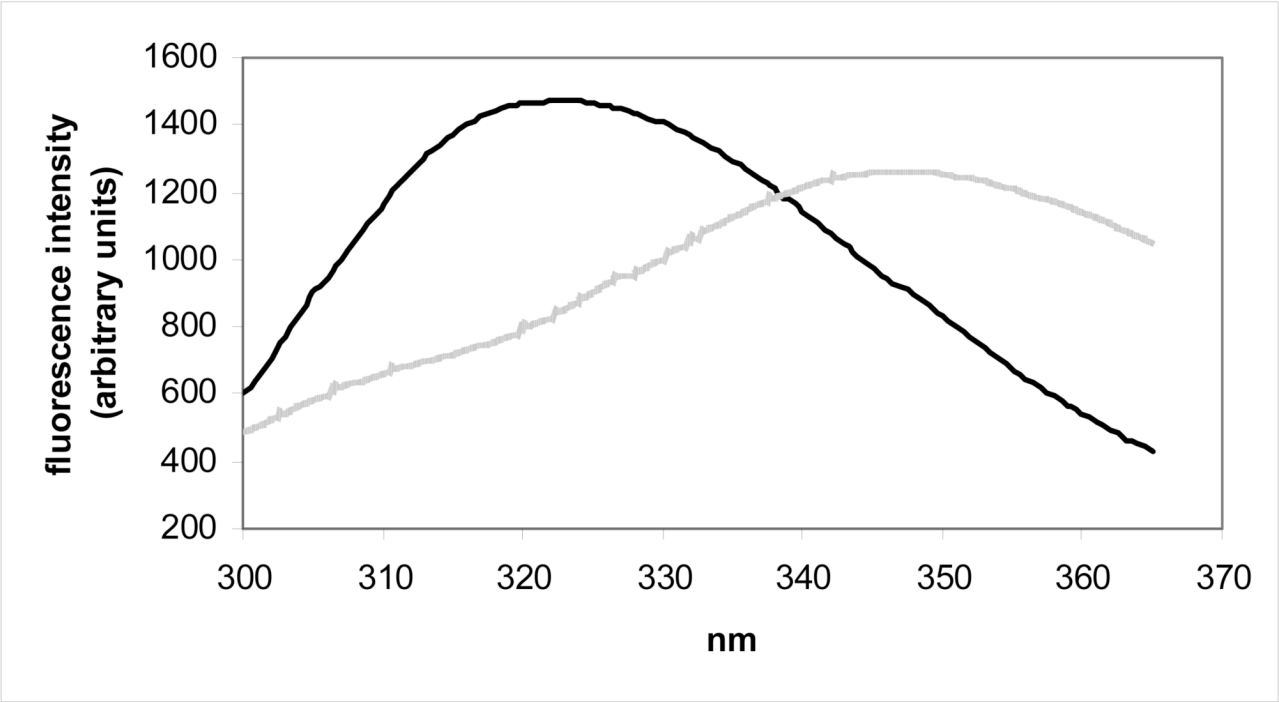
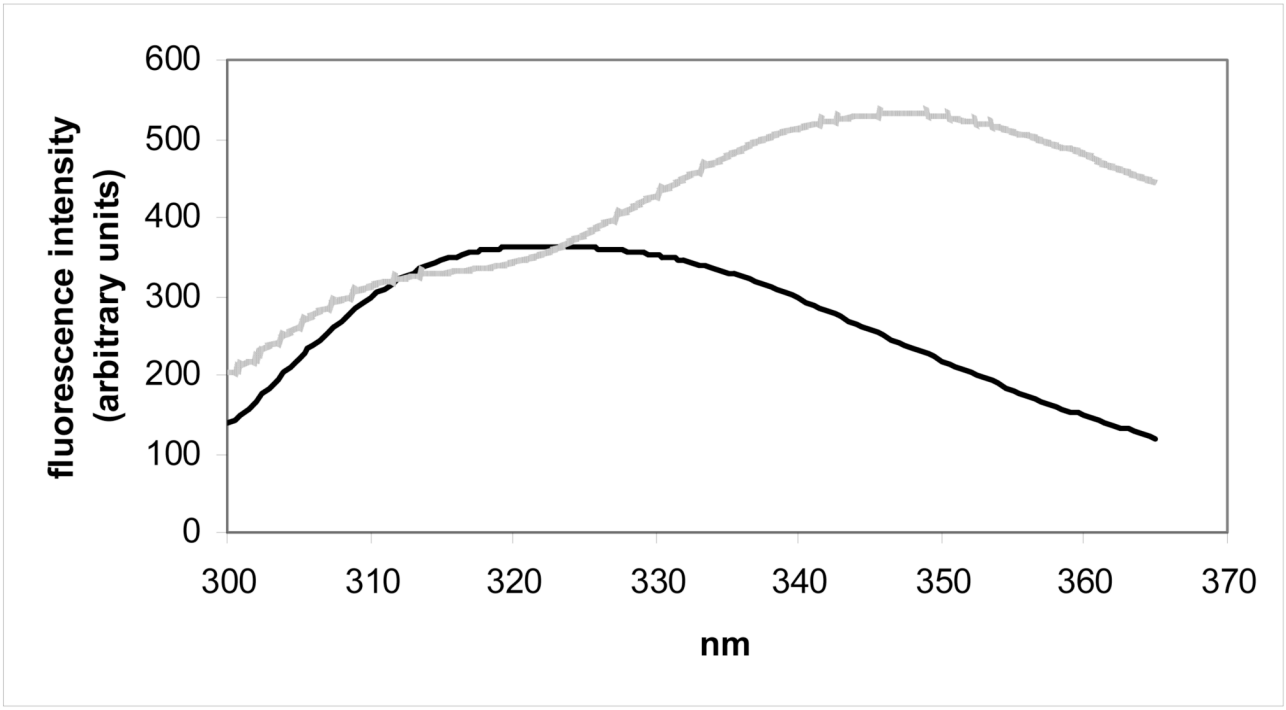


Figure 2.
Relative abundance of γC and γCins in mouse lens. Mouse γCs were copurified by cation exchange chromatography and analyzed by electrospray ionization mass spectrometry during their coelution by reverse phase chromatography. The deconvoluted spectrum was then used to estimate the relative abundance of the two forms of mouse γC by integration of the areas for each protein species.







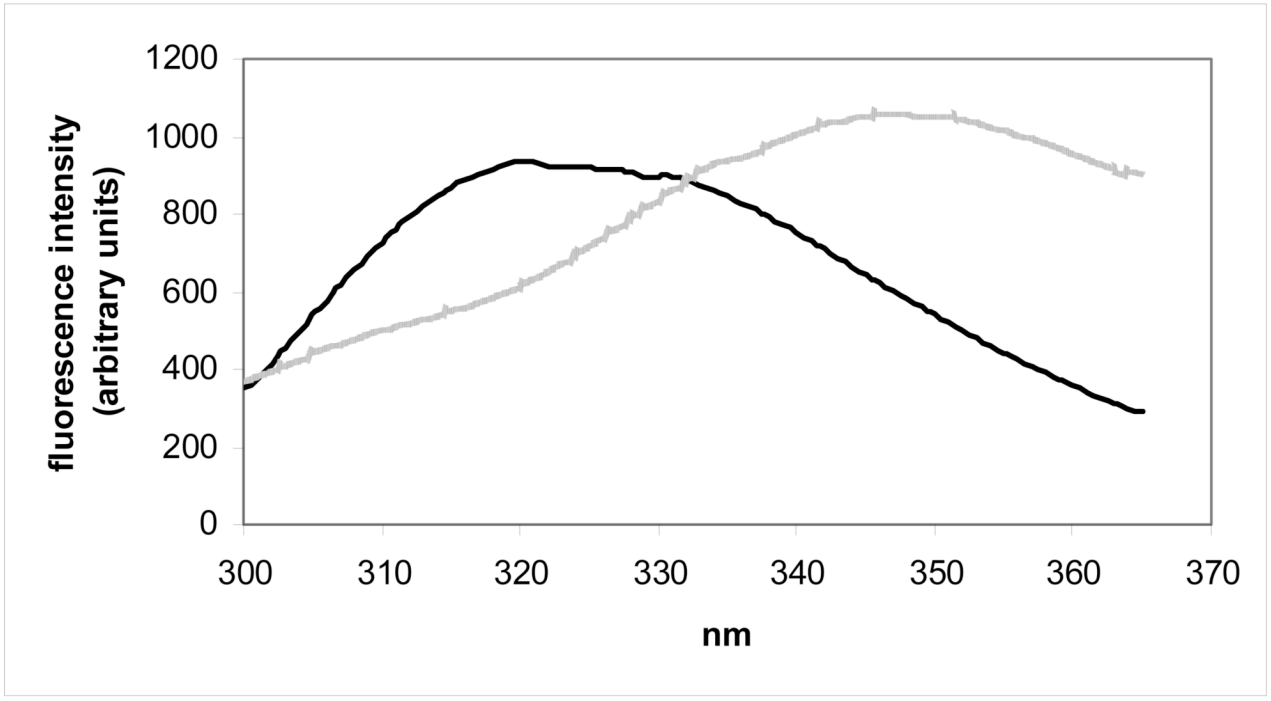


Figure 3.
Tryptophan fluorescence spectra of native (dark lines) human and mouse γC-crystallins compared with their unfolded forms (grey lines) at 37 °C. Unfolded samples were equilibrated

in 6M Gdn-HCl for six hours. (a) human γC-crystallin (b) human γC-crystallin C79R (c) mouse γC-crystallin (d) mouse γCins.

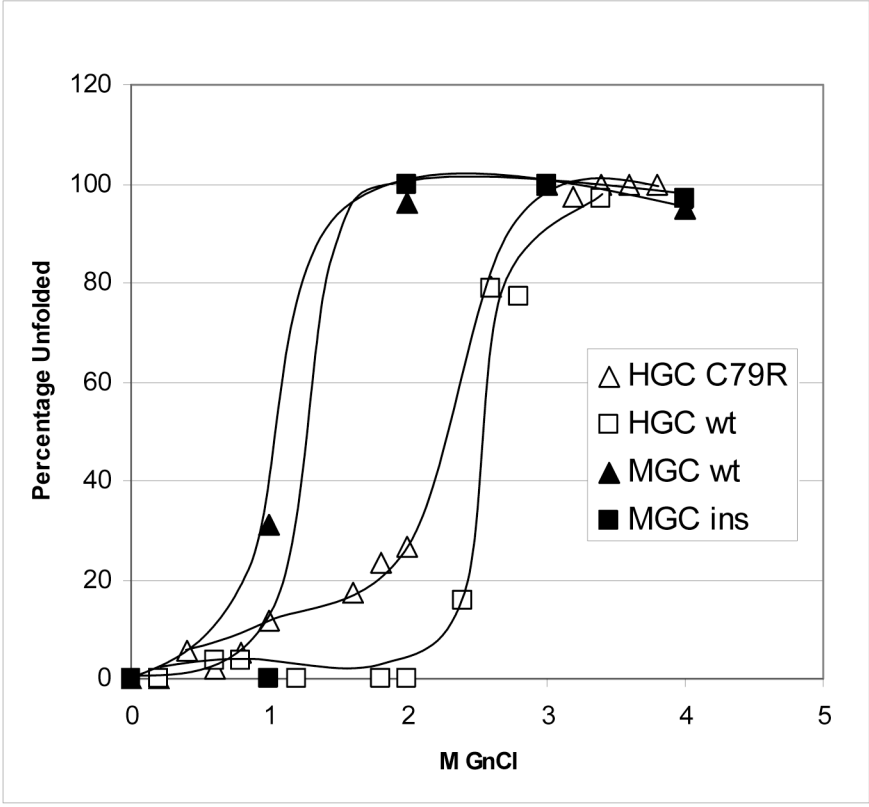


Figure 4.
Gdn-HCl induced unfolding of human and mouse γC-crystallins at 37 °C. Unfolding was monitored by emission at 350 nm, following tryptophan excitation at 280 nm.

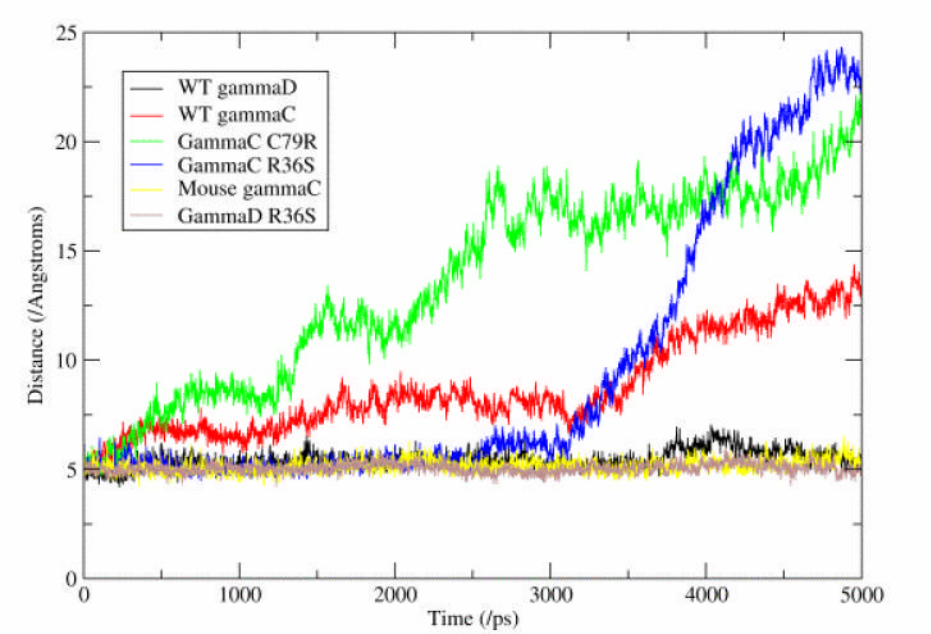


Figure 5. Distance across domain-domain interface

This figure represents the changes seen in the hydrophobic interface between the N- and C-terminal domains. This interface consists of six conserved residues (Met 43, Phe/Leu 56 and Ile 81 from the N-terminal domain and Val 131, Ile 144 and Val 169 from the C-terminal domain). The graph shows the distance between the Cβ atoms of residues Met 43 and Leu 144, these two residues being the most 'central' in the interface. It can be seen that the three proteins with a Phe at position 56 show very little change, whilst the three proteins with a Leu at position 56 show a great increase in the distance.

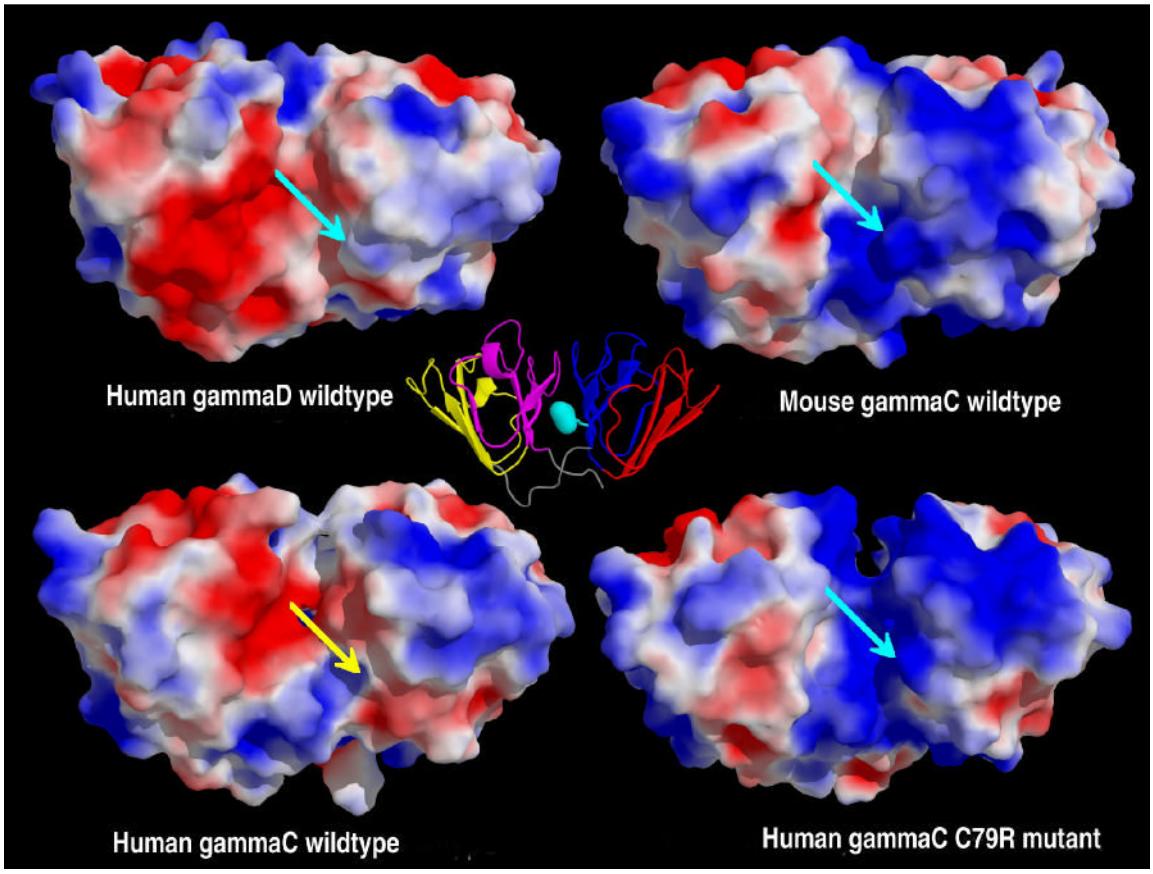


Figure 6. Comparison of electrostatic surfaces

The molecular surfaces of the four crystallins after 1ns of molecular dynamics coloured by the electrostatic potential at the molecular surface; the potential is scaled from a maximum of $+2k_BT/e_c$ (blue) to a minimum of $-2\ k_BT/e_c$ (red). The molecules have been superimposed before the calculation of the electrostatics, and are displayed with the C-terminal domain on the left hand side. Proteins with an Arg at position 79, have this highlighted with a turquoise arrow, whilst the wild-type human γC-crystallin has a yellow arrow signifying Cys 79. A cartoon version of mouse γC-crystallin (motif 1 in red, motif 2 in blue, motif 3 in yellow and motif 4 in magenta) is displayed at the center, as a reference, with the location of Arg 79 highlighted in turquoise. From this view, human γD-crystallin has a mainly negatively charged C-terminal domain, with the N-terminal domain being slightly positively charged, the interface between the domains being slightly negative. The mouse γC-crystallin has a mixed C-terminal domain, with the N-terminal domain being mainly positive, the interface being partially positive. The model of the human γC-crystallin wild-type shows the C-terminal domain being a mix of potentials and the N-terminal domain being predominantly positive, the interface is a mix of positive and negative. The human γC-crystallin C79R mutant again has a mix of potentials on the C-terminal domain, but the N-terminal domain and the interface are both positively charged. Note: The human γC-crystallin R36S mutant is not shown, as this mutation is on the opposite side of the protein.

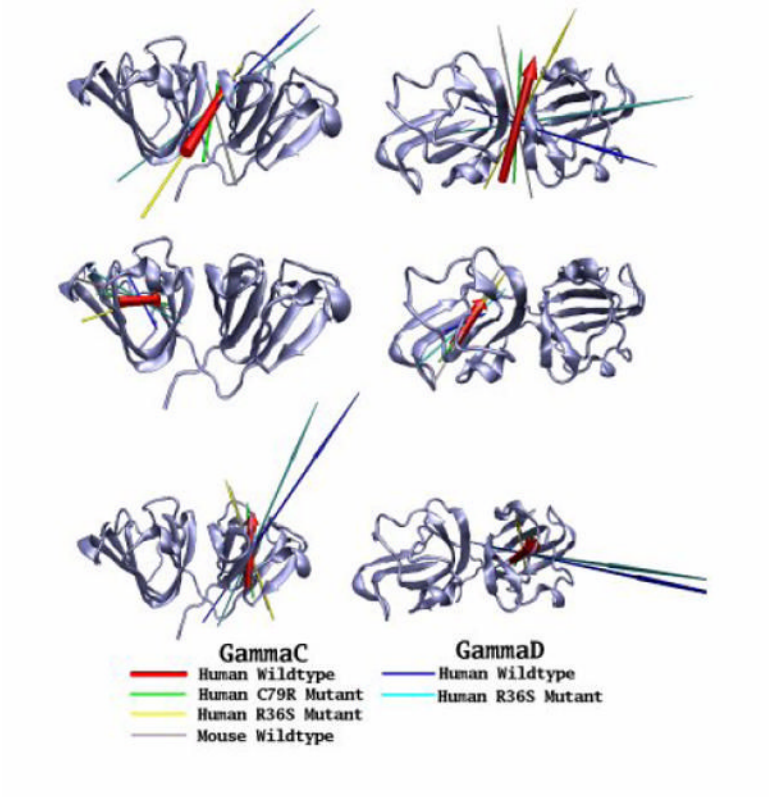


Figure 7. The dipole moments of the γ-crystallins in this study

The dipole moments of six γ-crystallins (human wild-type γC, C79R, R36S and mouse γC; human wild-type γD and R36S mutant) have been calculated at the start of their simulation trajectories. Dipole moments of the whole protein (top line), N-terminal domain (middle line) and C-terminal domain (bottom line) are shown as arrows, with the length proportional to the magnitude, the underlying molecules have been superimposed using the group of residue concerned, with the secondary structure of the human wild-type γC-crystallins used as a reference point. Two views of the proteins are presented, from the ‘side’ and from the ‘top’ with the N-terminal domain on the left.

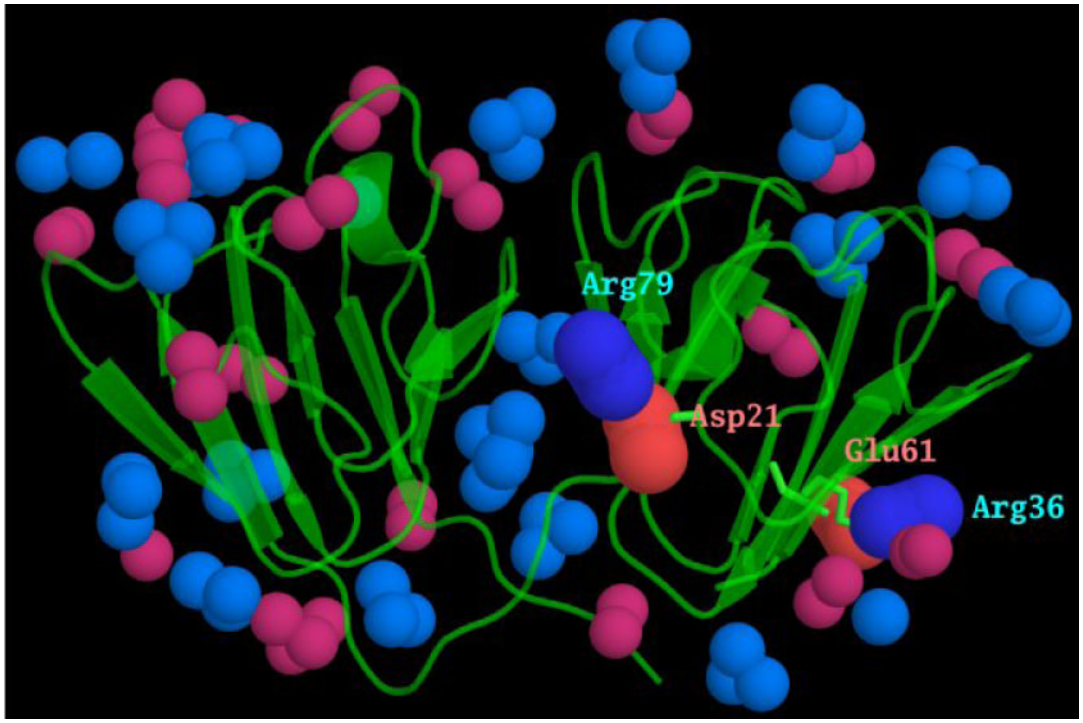


Figure 8. The conserved ion pairs between motifs

The surface of γ-crystallins is covered with ion pairs. Motifs 1, 2 and 4 have a conserved acidic and basic residue, highlighted in red and blue, in the alignment in Fig 1a. As a consequence of each domain being formed from motifs related by a pseudo-symmetrical twofold axis, these charged residues form inter-motif ion pairs. These are shown here, from the mouse γC-crystallin x-ray coordinates, with the N-terminal domain ion pairs between Asp 21 and Arg 79, and between Arg 36 and Glu 61 labelled. The atoms at the ends of the other acidic and basic side chains are shown in lighter colours. It can be seen that each of these inter motif ion pairs lies in the middle of a β-sheet, rather than at the “top” or “bottom” of the wedge-shaped domain where most of the other charged residues and ion pairs reside.

Table 1

X-ray Diffraction Statistics for mouse γC-crystallin

Parameter	Value	
Spacegroup	P1	
Unit Cell	a) 34.9Å b) 46.3Å c) 53.3Å [b)] α = 108.5°, β = 94.1° γ = 99.4°	
Wavelength	0.871Å (ID23-2, ESRF)	
Temperature of data collection	100K	
Molecules in asu.	2	
Matthews Coefficient	1.91	
Solvent Content	35.7%	
	All Data (43.1 – 1.90Å)	High Resolution Bin (2.00 – 1.90Å)
Number of Reflections	149314	4626
Number of Unique Reflections	22725	2313
Multiplicity	1.9	1.8
Completeness	92.0%	76.5%
R_{merge}	8.5%	55.9%
I/Sig I	5.1	0.9

Table 2

Refinement Statistics for mouse γC-crystallin

Parameter	Value
Number of reflections working (test) set	21170 (1149)
R_{cryst} (R_{Free}) before refinement	43.0% (44.5%)
Final R_{cryst} (R_{Free})	21.5% (28.9%)
Overall Figure of Merit	0.777
Number of atoms in final refinement round	3084
Rmsd bond lengths	0.008Å
Rmsd bond angles	1.98O
Rmsd main chain B-factor bond (angle)	0.44Å2 (0.67$Å^2$)
Rmsd side chain B-factor bond (angle)	1.10Å2 (1.72$Å^2$)
Ramachandran plot: Most favoured region	90.3%
Ramachandran plot: Additionally allowed region	9.7%

Table 3

Distances between hydrophobic interface residues throughout the simulations

Residue Pair	**43 – 144**		**56 – 131**		**81 – 169**	
Protein	**Mean (SD)**	**Max / Min**	**Mean (SD)**	**Max / Min**	**Mean (SD)**	**Max / Min**
Human γC WT (L56)	8.7 (2.2)	14.4 / 4.9	6.9 (1.5)	10.7 / 4.2	6.4 (1.1)	10.3 / 4.7
Human γC C79R (L56)	13.6 (4.4)	22.3 / 4.7	13.1 (5.4)	23.6 / 4.3	7.6 (1.4)	11.7 / 5.0
Human γC R36S (L56)	9.4 (6.2)	24.3 / 4.4	8.1 (4.7)	20.6 / 4.5	7.5 (3.9)	17.1 / 4.5
Mouse γC WT (F56)	5.2 (0.3)	6.5 / 4.2	5.2 (0.3)	6.8 / 4.4	5.5 (0.3)	6.6 / 4.5
Human γD WT (F56)	5.4 (0.4)	7.0 / 4.2	5.2 (0.4)	6.6 / 4.2	5.4 (0.4)	6.9 / 4.5
Human γD R36S (F56)	5.1 (0.3)	6.0 / 4.2	6.5 (0.8)	9.0 / 4.5	5.1 (0.3)	6.7 / 4.2

The Cβ – Cβ distances (Å) for residues across the domain – domain interface during the 5 ns long simulations. The stability of the interface in those proteins with Phe 56 (mouse γC and human γD) can be seen when compared to human γC which has Leu 56.

Table 4

Dipole moment data during the simulations

	N-domain			C-domain			All residues		
Protein	**pI**	**Mean (SD)**	**Min / Max**	**pI**	**Mean (SD)**	**Min / Max**	**pI**	**Mean (SD)**	**Min / Max**
Human γC WT	6.4	57.6 (19.6)	2.9 / 105.5	7.0	234.5 (21.0)	177.3 / 296.2	7.0	134.4 (27.3)	52.9 / 208.8
Human γC C79R	7.8	101.9 (15.3)	54.7 / 158.3	7.0	293.4 (27.8)	200.9 / 371.5	7.7	204.4 (26.0)	114.6 / 277.8
Human γC R36S	5.2	129.1 (14.5)	83.7 / 176.1	7.0	269.1 (22.2)	184.2 / 333.8	6.5	231.0 (53.7)	74.8 / 329.7
Mouse γC WT	7.8	139.4 (16.3)	80.7 / 186.1	6.9	292.1 (19.7)	224.4 / 361.9	7.7	209.8 (26.7)	128.6 / 312.8
Human γD WT	7.0	84.5 (16.5)	38.6 / 136.9	6.8	394.9 (22.7)	329.8 / 466.5	7.2	205.4 (24.8)	123.4 / 284.5
Human γD R36S	6.3	121.2 (16.2)	75.7 / 177.2	6.8	427.9 (30.8)	331.4 / 486.5	6.7	346.1 (44.2)	214.3 / 440.5
Human γD WT Crystal	7.0	123.4		6.8	311.8		7.2	196.4	
Human γD P23T Model	7.0	124.3		6.8	311.8		7.2	197.4	
Human γD R58H Crystal	6.4	149.1		6.8	326.1		6.7	211.0	

Statistics for the dipole moment (/Debye) and pI (calculated from http://www.expasy.ch/tools/pi_tool.html) for the individual domains and the whole protein from each of the six simulations, plus the static values from the crystal structures of wild-type human γD-crystallin and R58H mutant and the model structure for the P23T mutant. It can be seen that the simulation of human γD-crystallin wild-type allows the dipole moments to relax from the values seen in the crystal lattice.

Table 5

Solvent Accessibilities

Residue	Human γD	Human γC	C79R	Mouse γC
15	-	-	-	48.2% (53.7Å^2)
18	8.8% (0.3Å^2)	-	-	14.9% (14.1Å^2)*
22	-	15.7% (17.3Å^2)	14.9% (16.4Å^2)	11.9% (12.5Å^2)*
32	2.0% (1.1Å^2)	0% (0.0Å^2)	1.2% (0.0Å^2)	0% (0.0Å^2)
41	3.6% (2.2Å^2)	0% (0.0Å^2)	6.9% (8.0Å^2)	4.9% (4.7Å^2)
78	0% (0.0Å^2)	0% (0.0Å^2)	0% (0.0Å^2)	0% (0.0Å^2)
79	-	21.3% (20.6Å^2)	-	-
108	13.3% (10.6Å^2)	4.1% (5.1Å^2)	0% (0.0Å^2)	11.0% (11.6Å^2)
110	49.6% (43.3Å^2)	-	-	40.4% (24.4Å^2)
129	-	7.5% (10.1Å^2)	1.2% (1.6Å^2)	3.9% (4.2Å^2)
153	-	13.7% (17.0Å^2)	5.3% (6.8Å^2)	-

The solvent accessibilities of the cysteine residues in four simulated proteins, calculated using the program NACCESS.[54] The relative total accessibility of the residue (compared to the sequence GCG) is shown, with the absolute accessibility of the SG atom in brackets.

The two residues with a * are partially disulphide bonded in the X-ray structure, but have been simulated in their reduced form.