doi:10.1016/S0022-2836(03)00375-9 *J. Mol. Biol.* (2003) **328**, 1137–1147

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

High-resolution X-ray Crystal Structures of Human γD Crystallin (1.25 Å) and the R58H Mutant (1.15 Å) Associated with Aculeiform Cataract

Ajit Basak[1], Orval Bateman[1], Christine Slingsby[1]*, Ajay Pande[2] Neer Asherie[2], Olutayo Ogun[2], George B. Benedek[2] and Jayanti Pande[2]*

[1]*Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, UK*

[2]*Department of Physics and Center for Materials Science and Engineering and Materials Processing Center Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139-4307, USA*

**Corresponding authors*

Several human cataracts have been linked to mutations in the γ crystallin gene. One of these is the aculeiform cataract, which is caused by an R58H mutation in γD crystallin. We have shown previously that this cataract is caused by crystallization of the mutant protein, which is an order of magnitude less soluble than the wild-type. Here, we report the very high-resolution crystal structures of the mutant and wild-type proteins. Both proteins crystallize in the same space group and lattice. Thus, a strict comparison of the protein–protein and protein–water intermolecular interactions in the two crystal lattices is possible. Overall, the differences between the mutant and wild-type structures are small. At position 58, the mutant protein loses the direct ion-pair intermolecular interaction present in the wild-type, due to the differences between histidine and arginine at the atomic level; the interaction in the mutant is mediated by water molecules. Away from the mutation site, the mutant and wild-type lattice structures differ in the identity of side-chains that occupy alternate conformations. Since the interactions in the crystal phase are very similar for the two proteins, we conclude that the reduction in the solubility of the mutant is mainly due to the effect of the R58H mutation in the solution phase. The results presented here are also important as they are the first high-resolution X-ray structures of human γ crystallins.

Keywords: crystal cataract; crystallin; eye lens; ion-pairs; protein solubility

Introduction

The γ crystallins are a family of monomeric proteins found in the mammalian lens. They are part of the high concentration mixture of crystallins that is responsible for establishing the refractive index gradient required for transparency.[1] The short-range order among the various components ensures that light-scattering is minimized and transparency maintained.[2,3] A disruption in this order may be caused by different types of protein condensates, such as aggregates, coexisting liquid phases or crystals.[4] Any such disruption results in increased light-scattering and lens opacity or cataract.

The γ crystallin gene cluster, which consists of six genes located on chromosome 2q 33–35 in humans, has been found to be one of the major loci for cataracts of genetic origin.[5,6] In a previous publication[7] we provided a molecular basis for lens opacity in two forms of genetic cataracts caused by mutations in the γD crystallin gene: the aculeiform cataract associated with the Arg58 to His (R58H) mutation[8] and the "crystal cataract" associated with the Arg36 to Ser (R36S) mutation.[9] We showed that both mutant proteins (R58H and R36S) have lower solubilities and crystallize more readily than wild-type human γD crystallin (HGD), and suggested that lens opacity occurs due to the formation of these crystal condensates. Furthermore, we determined from measurements in solution that there was no large difference in

Abbreviations used: HGD, human γD crystallin; BGD, bovine γD crystallin; LLPS, liquid–liquid phase separation.

E-mail addresses of the corresponding authors: c.slingsby@mail.cryst.bbk.ac.uk; jpande@mit.edu

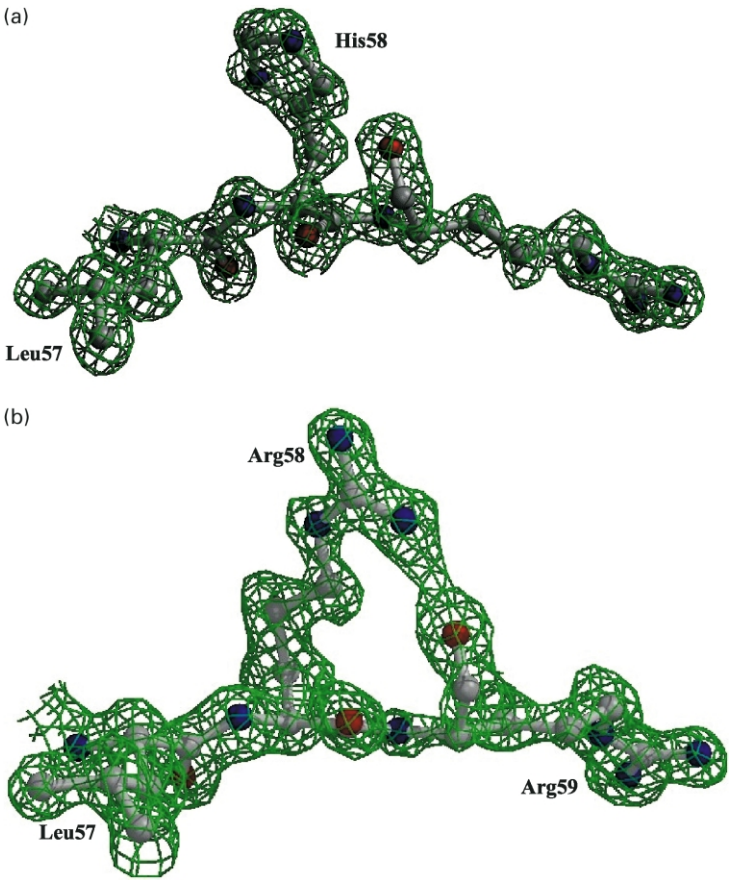

Figure 1. The electron density clearly shows the outline of the two residues, (a) His58 and (b) Arg58, in R58H and HGD, respectively.

the protein structure between the wild-type and the two mutants. Thus, we showed that a conformational change was not required to bring about a cataractogenic change in the mutant proteins.

Kmoch and co-workers[9] have determined the 2.25 Å resolution X-ray crystal structure of R36S and have shown that the mutation leads to many altered crystal contacts relative to wild-type, bovine γD crystallin, (BGD). These authors used the bovine protein as a template, since no X-ray crystal structure was then available for HGD. In order to find out whether the enhanced crystallization of R58H was also due to a modification of crystal contacts, we determined the structure of R58H as well as that of HGD. Here, we present the first high-resolution X-ray crystal structure of wild-type, HGD (at 1.25 Å) and its R58H mutant (at 1.15 Å) and show that in contrast to R36S, crystals of HGD and R58H have an identical space group and therefore very similar lattice contacts. This similarity allows us to conduct a comparison of the interactions present in the two proteins.

Results

The mutation site

As described in Materials and Methods, both wild-type and mutant proteins were grown under very similar, close to physiological, conditions. Both proteins crystallized in the same space group with one molecule in the asymmetric unit. Diffraction datasets to the limit of 1.25 Å and 1.15 Å were collected for HGD and R58H protein crystals, respectively. In Figure 1 we see that the electron density for residue 58 in each of the two structures is well defined. R58H and HGD have almost identical polypeptide structures with an average root-mean-square deviation (rmsd) of 0.3 Å using 173 C^{α} coordinates. Figure 2(a) shows the location of the mutation site at position 58 on the N-terminal domain, close to the interface with the C-terminal domain. A close-up view of the mutation site shows how closely the histidine and arginine side-chains follow each other with only minimal differences between the two structures (Figure 2(b)).

Water replaces the protein–protein intermolecular interaction in R58H

Since both HGD and R58H crystallized in essentially the same space group and lattice, the effect of the mutation on intermolecular interactions could be analysed readily. The most striking difference between the two lattice structures is that R58 in HGD forms a strong ion-pair with symmetry-related D156; both NH1 and NH2 atoms of the arginine residue form hydrogen bonds with the OD1 and OD2 atoms of the aspartate residue (Figure 3(a), left). In contrast, the "shorter" histidine residue in the R58H mutant forms neither

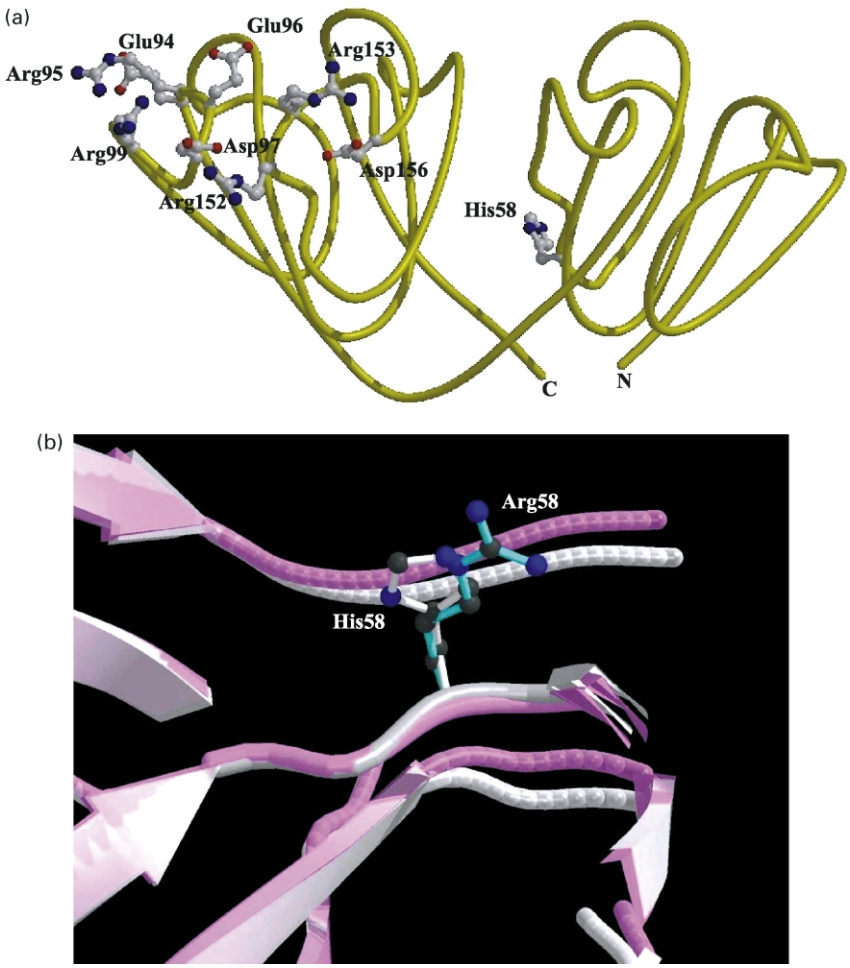

Figure 2. The overall impact of the residue alteration on the structure. (a) The C^{α} backbone trace of the R58H, showing the mutation site, is located close to the interface between the N and C-terminal domains. The component side-chains of the charged cluster in the C-terminal domain of R58H (R152, D97, R99, R95, E94, D156, R153, E96) are appended. (b) A close-up view around the mutation site shows the superposition of the two side-chains that differ by one nitrogen atom only. The side-chains occupy closely equivalent spatial positions within the limit of three (out of four) shared carbon atoms with the NE atom of arginine superposed close to one of the histidine side-chain nitrogen atoms. The spatial atomic structural difference between HGD (turquoise) and R58H (white) is thus confined to one nitrogen atom for the mutant, two nitrogen atoms for the wild-type crystallin, and one carbon atom each.

this contact, nor any other direct protein–protein intermolecular interaction (Figure 3(a), right). Instead, in the mutant, D156 interacts with two water molecules. These Figures clearly show that replacement of R58 in the wild-type with H58 in the mutant results in the loss of a strong, protein–protein, intermolecular, ion-pair interaction at the mutation site. However, the nitrogen atom on the other side of H58 makes an intermolecular contact with D97 in the lattice, which is mediated *via* a single water molecule. This is equivalent to the interaction mediated by two consecutive water molecules in the wild-type between the NE nitrogen atom of R58 and D97 in the lattice (Figure 3(b)). Furthermore, in the mutant, R168 from the C-terminal domain is anchored in an intermolecular interaction with D97, whereas in the wild-type, R168 and D97 are separated by water molecules (Figure 3(b)). These data suggest that compensation for structural differences at the mutation site is gained by the insertion of water molecules between the side-chain and aspartate residues of symmetry-related molecules. This process is accompanied by local conformational changes to side-chains in the immediate vicinity of the mutation site in order to satisfy hydrogen-bonding requirements (Figure 3(a) and (b)). Thus, at the actual mutation site a weaker protein–protein intermolecular interaction occurs in the mutant relative to the wild-type protein (Table 2), and is accompanied by the re-organisation of the protein–water–protein interactions that include R168 in the close vicinity.

Differences in lattice ion pairs

The direct lattice contact of R58, namely D156, is on the edge of a dense charge cluster[10] that can be described as two rows of charged residues on the C-terminal domain (Table 1 and Figure 2(a)). The majority of the ionic contacts in the crystal lattice are in this charge cluster (Table 2). The side-chain conformations of the component residues of the cluster differ in detail between the two protein structures due to the differing perturbations caused by residue 58 in the lattice. The mutation site has an impact on the overall molecular surface, since it is at a key intermolecular contact. One consequence is that, although the mutant has lost an intermolecular ion-pair at the mutation site, it gains two other intermolecular ion-pairs that are stronger than those in the wild-type (Table 2). As mentioned above, R58H has a stronger lattice ion-pair interaction between D97 and R168 than does HGD, and a stronger lattice ion-pair interaction between R99 and D108 than the corresponding

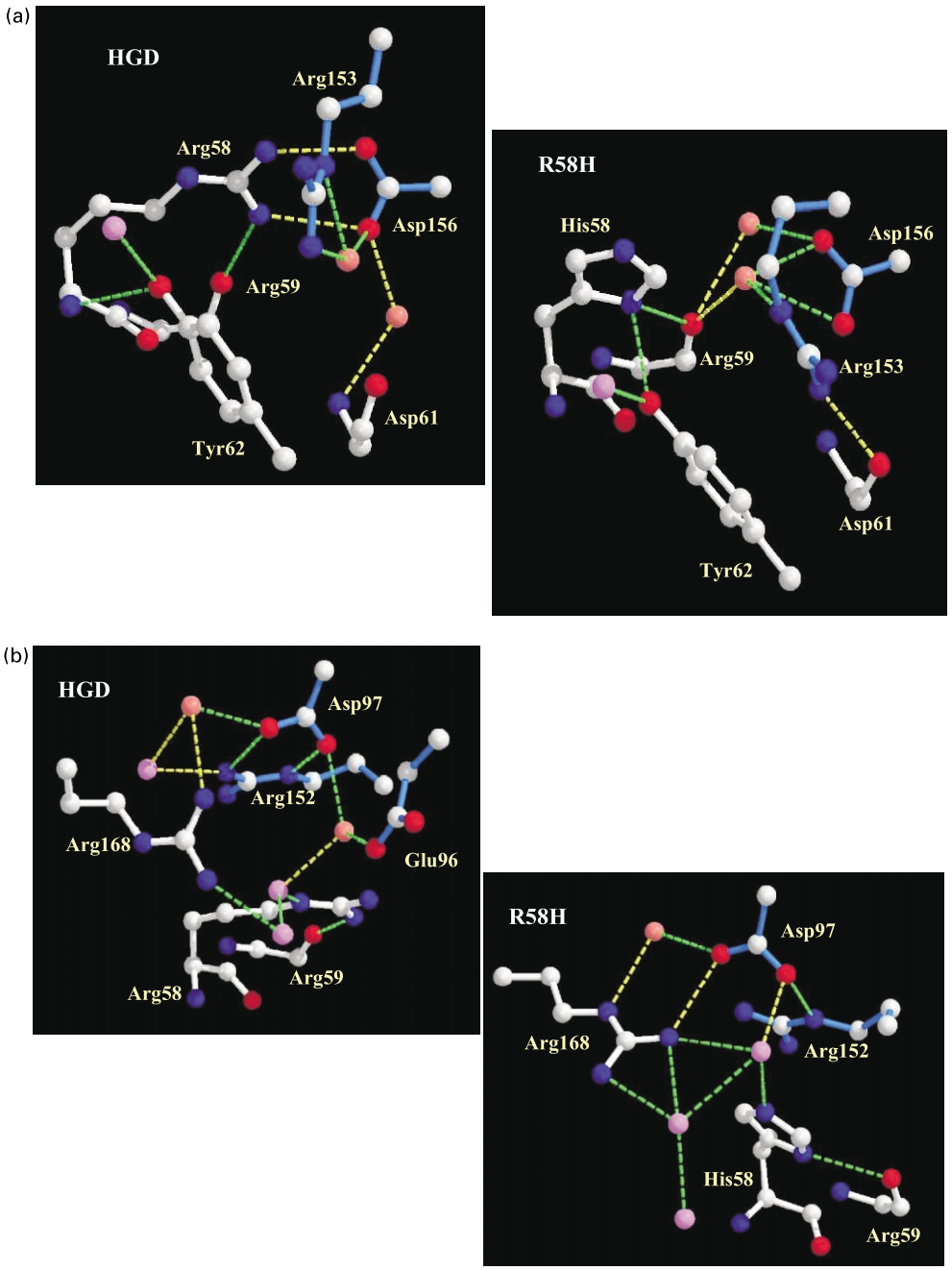

Figure 3. There are more water molecules around the environment of the nitrogen atoms of residue 58 in HGD compared with R58H. (a) Lattice side: in HGD the two NH atoms of R58 are involved in direct hydrogen bonds with two carboxylate side-chain oxygen atoms of D156 from a neighbouring molecule in the crystal lattice. In R58H the side-chain of H58 does not make direct hydrogen bonds with a symmetry-related molecule, instead D156 makes water-mediated hydrogen bonds. Both R58 and H58 make a hydrogen bond between a side-chain N atom and the backbone carbonyl group of R59. Note that in the mutant the side-chain hydroxyl group of Y62 is hydrogen bonding to the same side-chain N of H58, bringing the aromatic ring up close to symmetry-related R153. (b) Intramolecular side: in the mutant H58 interacts with symmetry-related D97 *via* a single water molecule, whereas in HGD the corresponding interaction is *via* two water molecules. In the mutant there is a direct protein interaction between R168 and symmetry-related D97, whereas the corresponding interaction in HGD is *via* a water molecule. In the mutant there is an interaction between R168 and H58 mediated *via* a single water molecule, whereas in HGD the corresponding interaction is *via* two water molecules. Protein carbon, nitrogen and oxygen atoms, are coloured white, blue and red, respectively; white bonds, molecule in asymmetric unit; pale blue bonds, symmetry-related molecules; pink and violet atoms, water molecules from different asymmetric units; broken yellow lines, intermolecular hydrogen bonds; broken green lines, intramolecular hydrogen bonds.

Table 1. Intermolecular contacts mediated by the C-terminal charge cluster

R58H Cluster **Cluster Contacts 3 Molecules**

Y98$ F173$
D150* R152%…..D97*....R99* R95…..E94 R163* F173% R168* D108*

D156…..R153% E96 H58 Y62%

HGD Cluster

Y98$ F173$
D150* R152%…..D97* R99* R95….E94 R163* F173% R168* D108*

D156* R153% E96* R58* Y62%

-1 0 0 1 0 0 1 -1 0

The residues in the cluster are all contiguous on the C-terminal domain of one molecule. These residues make lattice interactions with three different molecules (colour-coded) in the lattice. Interacting residues for a given pair of molecules are in the same colour and can be correlated with Table 2. The following symbols indicate the nature of the interaction: stops…, a strong intramolecular ion-pair; *, an intermolecular ion-pair; $, a hydrophobic intermolecular interaction; %, an amino–aromatic intermolecular interaction.

Table 2. Intermolecular lattice contacts

ION-PAIRS

R58H (6) **HGD** (12)

Å

OD1 38….NH1 117 3.6
OD2 38….NH1 117 3.1
OD2 38….NH2 117 2.5

Water mediated contacts

NH1 58….OD1 156 3.0 1 0 0
NH1 58….OD2 156 3.3 1 0 0
NH2 58….OD1 156 3.6 1 0 0
NH2 58….OD2 156 2.9 1 0 0

Å

OD1 97….. NH2 168 3.9 -1 0 0 OE2 96…NH2 168 3.7 -1 0 0
OD2 97….. NH2 168 2.9 -1 0 0 OD2 97…NH2 168 3.1 -1 0 0
OD2 97….. NE 168 3.9 -1 0 0

NH1 99…..OD2 108 3.2 -1 0 0 NH2 99…OD2 108 2.5 -1 0 0
NH2 99…..OD2 108 2.5 -1 0 0

OD1 150…..NH1 163 3.2 1 -1 0 OD1 150…NH1 163 3.2 1 -1 0
OD2 150…..NH1 163 3.4 1 -1 0 OD2 150…NH1 163 3.5 1 -1 0

High temperature factors
Distances between interacting atoms are given under column labeled Å

HYDROPHOBIC ($) AND AMINO-AROMATIC (%) INTERACTIONS

$ Y98….F173 (4) -1 0 0 Y 98…F173 (3) -1 0 0
% R152….F173 (4) -1 0 0 R 99…Y139 (4) R152…Y173 (3) -1 0 0
% Y62…. R153 (6) 1 0 0 Y 62….R153 (2) 1 0 0

AMINO- SULPHUR

NH2 R14… SD M102 1 1 0 NH2 R14…SD M102 1 1 0

The numbers in parentheses indicate the number of contacts estimated from the program MODEL: Lattice Contact Calculation, version 5.08 (H.P.C. Driessen, 1997).

interaction in HGD. The weaker intermolecular ion-pairs in HGD are due, in part, to higher disorder in the side-chains of R99 and R168, as outlined below. Both lattices share a similar set of intermolecular ion-pairs between D150 and R163, but all others differ in detail. Residues outlined in red in Table 2 indicate that they are disordered to some extent in the HGD lattice. Interactions involving residues D97, R99, D108 and R168 bind two protein molecules together in the lattice. These interactions are stronger in R58H.

Comparison of the overall lattice contacts

While there are differences in detail between the number and types of intermolecular ion-pairs in HGD compared with R58H, how important are these ion-pairs relative to all the interactions that bind the molecules together in their lattices? In Table 2 we list the intermolecular contacts that involve side-chain to side-chain, hydrophobic, amino to aromatic and amino to sulphur atoms, as well as ion-pairs, and note that arginine is involved in many interactions. The molecules interact in a head-to-tail fashion, involving R14 and M102 at each end of the long axis of the molecule (1 1 0). These interactions are similar in both HGD and R58H (Table 2). However, the charge cluster from the C-terminal domain of a single molecule interacts with three other molecules in the lattice (−1 0 0, 1 0 0, 1 −1 0), as shown in Table 1. Although there is a major hydrophobic interaction between Y98 and F173, arginine participates in interactions with all three molecular sites, with two of the sites involving amino–aromatic interactions (R152 and F173, and R153 and Y62).

When considering a single molecule, the mutation site (position 58) is adjacent to Y62 (Figure 3(a)) and is in the region of F173. We will now consider the lattice interactions that involve these residues from the perspective of the charge cluster in the C-terminal domain. In HGD, D156 and R153 in the lattice make direct intermolecular interactions with R58 and Y62 from one molecule, while Y98 makes an intermolecular hydrophobic interaction with F173 from another molecule in the lattice (Table 1). In R58H, the interaction between D156 and H58 is weakened considerably, while that between R153 and Y62 is strengthened compared to HGD; the Y98 and F173 interaction remains the same. The side-chain conformation of R153 differs between mutant and wild-type proteins (Figure 3(a)), with more intermolecular protein contacts with Y62 in R58H (Table 2).

The increased protein–protein intermolecular interactions in the mutant compared with wild-type, two ion-pairs (D97-R168, R99-D108) and one amino–aromatic interaction (R153-Y62) are all in the vicinity of the mutation site lattice contact. In the mutant, molecules related by the 100 symmetry operator (contacts shown in pink) have a stronger amino–aromatic interaction and a weaker ionic interaction compared with wild-type (Tables 1 and 2). In the mutant, the large contact site between molecules related by the −100 symmetry operator (contacts show in green) has stronger ion pairs than wild-type (Tables 1 and 2). These alterations to the 100 and −100 lattice contact sites may contribute towards differences in crystal nucleation rate of the two proteins.

Differences detected in disorder and the water structure

In general, the specific location of side-chain disorder differs between HGD and R58H. For example, in the mutant protein, H83 and F116 have well-resolved density for two equally populated conformers, whereas the corresponding residues in the wild-type are in well-ordered, single conformations (Figure 4(a) and (b)). By contrast, R76 in HGD can be refined in two conformations, whereas in R58H it is in a single conformation. If we consider H88, both protein structures have side-chain density for a single conformation, although the temperature factors are much higher in HGD. The temperature factors for several other basic side-chains (K2, R9, R95, R99, R115, R117, R168 and R169) are higher in HGD than in R58H, indicating that these residues are more disordered. Of particular note are those side-chains in HGD with higher temperature factors that are involved in intermolecular ion-pairs; namely, R99, R117 and R168.

Discussion

The X-ray data show that the differences between the structures of HGD and R58H are small, and suggest that for the crystal to form readily the mutation need not make a direct protein–protein contact. This is in contrast to R36S, for which position 36 is at a crystal contact.[9] The many similarities between the crystal structures of HGD and R58H lead us to believe the marked difference in solubility stems principally from the solution states of the two proteins. We discuss below how the solution state may contribute to the lowered solubility of the mutant protein.

The solubility is the concentration of the liquid which is in equilibrium with the crystal at a given temperature.[11] We may therefore write:[12]

$$\ln\left(\frac{C^{M}}{C^{WT}}\right) = \frac{\Delta G_s - \Delta G_l}{RT} \quad (1)$$

Here C^{M}, C^{W} are the solubilities of the mutant and wild-type protein respectively at temperature T, and ΔG_s is the difference between the molar free energies of the mutant and wild-type in the solid phase, while ΔG_l is this difference in the liquid phase; R is the universal gas constant. Equation (1) assumes that the liquid phase is dilute, an assumption that holds for the ranges of concentration we have studied.[7,13]

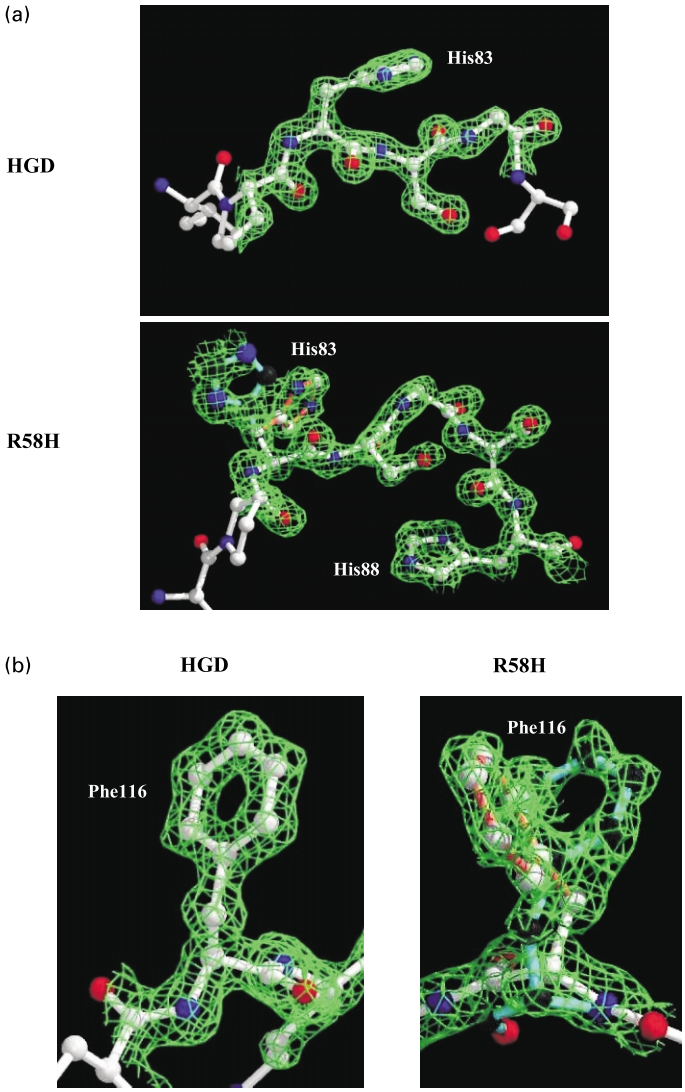

Figure 4. Residues that show differences in static disorder tend to be specific to mutant or wild-type structures. (a) In the linker region connecting the N and C-terminal domains, H83 occurs in a single conformation in HGD, whereas in the mutant H83 occurs in two conformations. (b) In the HGD structure, F116 is in a single conformation, whereas in R58H it is in two conformations. Where two conformations occur they are indicated by blue and pink bonds.

If the structures of both the wild-type and mutant proteins are known, then the free energies can, in principle, be calculated by using these structures with models for the protein interaction energies.[14,15] Free energy calculations of this type have been carried out for phenomena that involve only one phase, e.g. to compute the second virial coefficient of bovine chymotrypsinogen A[16] or to determine the binding energies of mutants of hen egg-white lysozyme to an antibody.[17] Such calculations, however, are computationally intensive and, as yet, do not produce quantitative agreement with experiment. Also, for the calculation of protein solubilities, the structures in both the liquid and solid phases must be known. Since we do not have sufficient information for an exact calculation, here we will discuss only in general terms whether there is one dominant factor of the R58H mutation that leads to the lower solubility.

As stated previously, the X-ray data show that there are no large global changes in the solid (crystal) phase. This is consistent with the data from circular dichroism, fluorescence and calorimetric measurements, which reveal that the mutation does not introduce any significant global change in the protein structure in solution either.[7] We conclude that the free energies in Equation (1) are determined mainly by the effects of small, local changes in protein structure brought about by the mutation.

The change of free energy in the liquid phase includes a contribution due to solvation effects. This contribution can be calculated from the values obtained experimentally for amino acid solvation energies. The difference in solvation energies between an arginine and a histidine side-chain (at pH 7 and 25 °C), ΔG(His) − ΔG(Arg), lies in the range of 1–1.7 kcal/mol (1 cal = 4.184 J), as determined from a variety of experiments on amino acid analogs and peptides.[18–21] This corresponds to a contribution of 1.7–2.9 to $\Delta G_1/RT$.

There will, of course, be other contributions to ΔG_1: for example, the residues near the mutation site will interact differently with arginine than with histidine (see Results). However, we believe that the solvation free energy is the dominant

contribution, since it accounts for most of the change in solubility (at pH 7 and 25 °C):[7]

$$\ln(C^{M}/C^{WT}) = -2.8$$

Figure 3(a) illustrates that there is little change in the local environment of the mutation. While it is true that H58 in the mutant does not have the same direct intermolecular contact with D156 as does R58 in the wild-type, this loss is compensated, in part, by extra contacts with water molecules. Therefore, we believe that $\Delta G_s/RT$ will be a small positive number, much less than 0.7, the value expected for an R–D salt-bridge.[22]

Our conclusion that the reduction in solubility of the mutant is mainly due to changes in the solution phase could be tested if we could show that the mutation brings about a change in the intermolecular interactions of the protein in solution. One property of the γ crystallins that is sensitive to these interactions is liquid–liquid phase separation (LLPS). LLPS is a temperature-dependent separation of a homogeneous protein solution into two coexisting liquid phases of unequal concentration. This transition is the mechanism for many animal models of cataract, including cold cataract.[4] LLPS is metastable with respect to crystallization and can be measured only if crystal formation is slow, as is the case for several native γ crystallins;[23] or if crystallization can be suppressed, as is the case for the Cys18 to Ser mutation of bovine γB crystallin.[24]

We were unable to measure LLPS for the R58H mutant, because the tendency of this protein to crystallize readily does not allow the transition to be measured. Our attempts to suppress the crystallization of R58H in order to observe the metastable LLPS were unsuccessful. In fact, even in mixtures with α and β crystallins (which were designed to mimic the environment of the lens) the R58H mutant crystallizes readily.

Most of the genetic cataracts studied thus far are due to mutations at arginine residues.[8,9,25] This suggests that these residues play an important role in maintaining the solubility of the γ crystallins. Indeed, when compared with most proteins, the slightly basic γ crystallins have a high ratio of Arg to Lys. The arginine residues in the γ crystallins form extensive ion-pair networks over the surface of the molecules that extend between molecules in the crystal lattice.[26] The substitution of one of these arginine residues for another residue perturbs the ion-pair-water network in the local crystal lattice. In the case of R58H the effect of the substitution on part of the arginine network (R58, R59, R152 and R168) is shown in Figure 3(b). The importance of the arginine residues may explain why they are conserved across many mammalian species.[27] It should be noted, however, that not all of the arginine residues in the γ crystallins appear to be equally important. For example, the R47H mutation, which arises from a single nucleotide polymorphism in human γC crystallins, does not lead to cataract.[6] This residue is not conserved across species.[27]

Since the high-resolution structure of HGD was not available until now, previous analysis of the behavior of this protein was based on the known structure of BGD.[10] The level of sequence identity between the two proteins is high (87%)[27] and they have very similar phase diagrams.[7,28,29] The 3D structures are also very similar; wild-type HGD and BGD structures superpose with an rmsd of 0.8 Å (for all 173 C^{α} atoms) and both structures have I103 contributing to a surface hydrophobic patch.[10] Although these proteins crystallize in the same space group, the asymmetric units differ, in that there are two protein molecules in BGD and one in HGD. Furthermore, the protein is more densely packed in crystals of HGD, which have a solvent content of only 40% (v/v) compared with 57% (v/v) in the case of BGD. It is difficult to pinpoint exactly which of the 22 sequence differences between HGD and BGD are critical for their different crystal lattice interactions. The sequence changes at positions 51, 147 and 155 are probably important, since changes at these positions are responsible for the different crystal packing between BGD and bovine γB-crystallin.[10]

The work presented here confirms our previous conclusion, that large conformational changes in either the crystal or in solution are not required for cataract formation. Furthermore, our work shows that cataract can occur even though the native and mutant proteins crystallize in essentially identical crystal lattices. It is therefore important to evaluate both the solution behavior and the crystal structure when studying crystal cataracts. Finally, the high-resolution X-ray structure of HGD we have presented here provides an important benchmark in the study of human cataract.

Materials and Methods

Cloning, expression and isolation of proteins

Recombinant HGD was prepared by the amplification of the coding sequence from a human fetal lens cDNA library as described.[29] Over-expression of HGD and R58H, and isolation and purification of the proteins were all done as described.[7,29] The crystallins folded efficiently and fractionated almost exclusively (>95%) into the soluble fraction. The purified proteins were characterized by electrospray ionization mass spectrometry and gave an average mass of 20,610 (±2) Da (HGD) and 20,588 (±2) Da (R58H), respectively. The concentrations of HGD and R58H were determined by measuring absorbance at at 280 nm and using an extinction coefficient of 41.4 mM^{-1} cm^{-1}.[30]

Crystallization

A slurry of fine crystals of HGD or R58H, grown initially by batch crystallization in 100 mM sodium phosphate (pH 7) containing 20 mM DTT, was suspended

Table 3. Data collection statistics

Data collection	HGD	R58H
Resolution range (Å)	28.17–1.25	28.33–1.15
No of observations		
Measured	212,732	463,412
Unique	39,021	55,087
Completeness (%)	86.8 (45.5)	95.3 (97.6)
R_{merge} (%)[a]	0.048 (0.189)	0.050 (0.109)
Multiplicity	2.9 (1.9)	4.6 (2.7)
$I/\sigma(I)$	7.4 (3.7)	9.2 (6.3)

[a] $R_{merge} = \sum|I_I - I_m|/\sum I_m$, where I_I and I_m are the observed intensity and mean intensity of related reflections, respectively.

in a small test-tube (50 mm × 6.5 mm) in a small aliquot of the phosphate buffer containing 5mM DTT. The suspension was warmed to 35 °C to dissolve the crystals and the tubes immediately placed in a Styrofoam container that was incubated at 4 °C. The cooled solution produced overnight a large number of discrete crystals suitable for X-ray diffraction measurements.

X-ray data collection

Data for crystals of HGD and R58H were collected from crystals flash-frozen following immersion in mother liquor containing 30% (v/v) glycerol as a cryoprotectant.[31] Both data sets were collected at station ID14-1 of the Synchrotron Radiation Source at ESRF, Grenoble, France, on a 1200 × 1200 ADSC Quantum4 CCD detector. Initial indexing of the diffraction patterns indicated that both crystals were in primitive orthorhombic symmetry, $P2_12_12_1$, with one molecule in the asymmetric unit and with unit cell dimensions: HGD, $a = 33.22$ Å, $b = 52.97$ Å, and $c = 90.47$ Å; R58H, $a = 33.40$ Å, $b = 53.18$ Å, and $c = 89.59$ Å, with $V_M = 1.98$ Å^3/Da solvent content of approximately 40% (v/v).[32]

All data sets were processed with the HKL program suite.[33] Scaling and merging was performed using the CCP4 program SCALA.[34] Subsequent data manipulation used the CCP4 program package.[35] Statistics for the data sets are given in Table 3.

Molecular replacement

The structure of wild-type BGD (PDB ID 1ELP) was used to generate initial phases using the program package AMoRe.[36] The same reflections were selected for cross-validation in both the HGD and R58H data sets. Rigid body refinement was initially carried out with CNS[37] using reflections between 8.0 Å and 2.5 Å for each data set. Following this, the *R*-factor and R_{free} were 35.2% and 40.5%, respectively, for HGD and 35.1% and 38.4%, respectively, for R58H. The phases were further improved using solvent flattening and density modification with the program DM[38] prior to the calculation of SigmaA weighted maps[39] for manual rebuilding of the model. The initial difference Fourier maps for the bovine and human proteins (BGD and HGD, respectively) showed differences in electron density supporting the expected amino acid sequence differences between the two proteins.

Model building and refinement

Manual model building was performed for both structures using the program O.[40] The initial refinement to 2.2 Å resolution was performed using the program CNS[37] by iterative cycles of simulated annealing with torsion-angle dynamics and model building. Subsequently, REFMAC[41] and ARP/wARP[42] were used to refine HGD and R58H to 1.25 Å and 1.15 Å resolution, respectively. Water molecules were added in both structures at the end of the refinement, where there was a peak of 3 rms or higher in the difference electron density and a 1 rms peak in the $2F_{obs} - F_{calc}$ map. Statistics for the final model for both structures are listed in Table 4.

Table 4. Refinement statistics

	HGD		R58H	
$N_{reflections}$ (work)	35,783		50,465	
$N_{reflections}$ (free)	3238		4565	
R-factor[a] (%)	16.36		15.55	
R-free[b] (%)	20.78		18.22	
$R_{(Work+test)}$ (%)	16.71		15.76	
Protein non-H atoms				
Water molecules	282		276	
Average *B* (Å^2)	MC	SC	MC	SC
Protein atoms	13.67	15.97	7.95	10.30
Water molecules	27.89		13.65	
rmsd				
Bonds (Å)	0.786		0.751	
Angles (deg.)	1.529		1.453	
Most favoured in Ramachandran plot (%)	91.5		90.8	

MC, main-chain; SC, side-chain.
[a] $R = \sum|F_{obs} - F_{calc}|/\sum F_{obs}$ for the 91% and 92% of the data included in the refinement.
[b] *R*-free $= \sum|F_{obs} - F_{calc}|/\sum F_{obs}$ for the 8% of the data randomly selected and excluded from refinement

Protein data bank accession codes

Coordinates and structure factors have been deposited in the PDB: accession codes 1hk0 and r1hk0sf for HGD; 1h4a and r1h4asf for R58H.

Acknowledgements

We are grateful to Dr Nicolette Lubsen for the gift of the HGD clone and to Professor Jonathan King for providing his laboratory facilities for the expression of the recombinant proteins. We thank the station staff at the ESRF, Grenoble, France. We gratefully acknowledge the financial support of the Medical Research Council (London) and the National Institutes of Health (USA) grants EY05127 to G.B.B. and EY10535 to J.P.

References

1. Zigler, J. S. Jr (1994). Lens proteins. In *Principles and Practice of Ophthalmology* (Alberts, E. M. & Jacobiec, F. A., eds), pp. 97–113, WB Saunders, Philadelphia.
2. Benedek, G. B. (1971). Theory of transparency of the eye. *Appl. Optics.* **10**, 459–473.
3. Delaye, M. & Tardieu, A. (1983). Short-range order of

crystallin proteins accounts for eye lens transparency. *Nature*, **302**, 415–417.
4. Benedek, G. B. (1997). Cataract as a protein condensation disease. *Invest. Ophthalmol. Vis. Sci.* **38**, 1911–1921.
5. Hejtmancik, J. F. (1998). The genetics of cataract: our vision becomes clearer. *Am. J. Genet.* **62**, 520–525.
6. Santhiya, S. T., Shyam Manohar, M., Rawlley, D., Vijayalakshmi, P., Namperumalsamy, P., Gopinath, P. M. *et al.* (2002). Novel mutations in the γ-crystallin genes cause autosomal dominant congenital cataracts. *J. Med. Genet.* **39**, 352–358.
7. Pande, A., Pande, J., Asherie, N., Lomakin, A., Ogun, O., King, J. & Benedek, G. B. (2001). Crystal cataracts: human genetic cataract caused by protein crystallization. *Proc. Natl Acad. Sci. USA*, **98**, 6116–6120.
8. Héon, E., Priston, M., Schorderet, D. F., Billingsley, G. D., Girard, P. O., Lubsen, N. & Munier, F. L. (1999). The γ-crystallins and human cataracts: a puzzle made clearer. *Am. J. Hum. Genet.* **65**, 1261–1267.
9. Kmoch, S., Brynda, J., Asfaw, B., Bezouška, K., Novák, P., Rezácová, P. *et al.* (2000). Link between a novel human γD-crystallin allele and a unique cataract phenotype explained by protein crystallography. *Hum. Mol. Genet.* **9**, 1779–1786.
10. Chirgadze, Y. N., Driessen, H. P. C., Wright, G., Slingsby, C., Hay, R. E. & Lindley, P. F. (1996). Structure of the bovine eye lens γD (γIIIb)-crystallin at 1.95 Å. *Acta Crystallog. sect. D*, **52**, 712–721.
11. McPherson, A. (1999). *Crystallization of Biological Macromolecules*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
12. van Holde, K. E. (1985). *Physical Biochemistry*, 2nd edit., Prentice-Hall, Englewood Cliffs.
13. Lomakin, A., Asherie, N. & Benedek, G. B. (1996). Monte Carlo study of phase separation in aqueous protein solutions. *J. Chem. Phys.* **104**, 1646–1656.
14. Eisenberg, D. & McLachlan, A. D. (1986). Solvation energy in protein folding and binding. *Nature*, **319**, 199–203.
15. Sitkoff, D., Sharp, K. A. & Honig, B. (1994). Accurate calculation of hydration free-energies using macroscopic solvent models. *J. Phys. Chem.* **98**, 1978–1988.
16. Neal, B. L., Asthagiri, D. & Lenhoff, A. M. (1998). Molecular origins of osmotic second virial coefficients of proteins. *Biophys. J.* **75**, 2469–2477.
17. Sharp, K. A. (1998). Calculation of HyHel10-Lysozyme binding free energy changes: effect of ten point mutations. *Proteins: Struct. Funct. Genet.* **33**, 39–48.
18. Wolfenden, R., Andersson, L., Cullis, P. M. & Southgate, C. C. B. (1981). Affinities of amino acid side-chains for solvent water. *Biochemistry*, **20**, 849–855.
19. Fauchère, J.-L. & Pliska, V. (1983). Hydrophobic parameters π of amino-acid side-chains from the partitioning of *N*-acetyl-amino-acid amides. *Eur. J. Med. Chem.* **18**, 369–375.
20. Radizcka, A. & Wolfenden, R. (1988). Comparing the polarities of amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol and neutral aqueous solution. *Biochemistry*, **27**, 1664–1670.
21. Wimley, W. C., Creamer, T. P. & White, S. H. (1996). Solvation energies of amino acid side-chains and backbone in a family of host–guest pentapeptides. *Biochemistry*, **35**, 5109–5124.
22. Creighton, T. E. (1993). *Proteins: Structures and Molecular Properties*, 2nd edit., W.H. Freeman, New York.
23. Broide, M. L., Berland, C. R., Pande, J., Ogun, O. O. & Benedek, G. B. (1991). Binary-liquid phase separation of lens protein solutions. *Proc. Natl Acad. Sci. USA*, **88**, 5660–5664.
24. Asherie, N., Pande, J., Pande, A., Zarutskie, J. A., Lomakin, J., Lomakin, A. *et al.* (2001). Enhanced crystallization of the Cys18 to Ser mutant of bovine γB crystallin. *J. Mol. Biol.* **314**, 663–669.
25. Stephan, D. A., Gillanders, E., Vanderveen, D., Freas-Lutz, D., Wistow, G., Baxevanis, A. D. *et al.* (1999). Progressive juvenile-onset punctate cataracts caused by mutation of the γD-crystallin gene. *Proc. Natl Acad. Sci. USA*, **96**, 1008–1012.
26. Kumaraswamy, V. S., Lindley, P. F., Slingsby, C. & Glover, I. D. (1996). An eye-lens protein-water structure: 1.2 Å resolution structure of γB-crystallin at 150 K. *Acta Crystallog. sect. D*, **52**, 611–622.
27. Norledge, B. V., Hay, R. E., Bateman, O. A., Slingsby, C. & Driessen, H. P. C. (1997). Towards a molecular understanding of phase separation in the lens: a comparison of the X-ray structures of two high T_c γ-crystallins γE and γF, with two low T_c γ-crystallins, γB and γD. *Expt. Eye Res.* **65**, 609–630.
28. Berland, C. R., Thurston, G. M., Kondo, M., Broide, M. L., Pande, J., Ogun, O. & Benedek, G. B. (1992). Solid-liquid phase boundaries of lens protein solutions. *Proc. Natl Acad. Sci. USA*, **89**, 1214–1218.
29. Pande, A., Pande, J., Asherie, N., Lomakin, A., Ogun, O., King, J. A. *et al.* (2000). Molecular basis of a progressive juvenile-onset hereditary cataract. *Proc. Natl Acad. Sci. USA*, **97**, 1993–1998.
30. Andley, U. P., Mathur, S., Griest, T. A. & Petrash, J. M. (1996). Cloning, expression, and chaperone-like activity of human αA-crystallin. *J. Biol. Chem.* **271**, 31973–31980.
31. Garman, E. F. & Schneider, T. R. (1997). Macromolecular cryocrystallography. *J. Appl. Crystallog.* **30**, 211–237.
32. Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.
33. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
34. Evans, P. R. (1993). Data reductions. In *Proceedings of CCP4 Study Weekend on Data Collection and Processing* (Sawyer, L., Isaacs, N. & Bailey, S., eds), pp. 114–122, Daresbury Laboratory, Warrington.
35. Collaborative computational project No. 4 (1994). The CCP4 suite: programs for the protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
36. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A*, **50**, 157–163.
37. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P. *et al.* (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
38. Cowtan, K. (1994). 'dm': an automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallog.* **31**, 34–38.
39. Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallog. sect. A*, **42**, 140–149.
40. Jones, T. A. & Kjelgaard, M. (1997). Electron density map interpretation. *Methods Enzymol.* **277**, 173–208.

41. Murshadov, G., Vagin, A. & Dodson, E. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.

42. Lanzim, V. S. & Wilson, K. S. (1997). Automated refinement for protein crystallography. *Methods Enzymol.* **277**, 269–305.

Edited by Sir A. Klug

(Received 12 November 2002; received in revised form 12 March 2003; accepted 14 March 2003)