*Acta Cryst.* (1996). D**52**, 712–721

# Structure of the Bovine Eye Lens γD (γIIIb)-Crystallin at 1.95 Å

YU. N. CHIRGADZE,[a] H. P. C. DRIESSEN,[b] G. WRIGHT,[b] C. SLINGSBY,[b] R. E. HAY[c] AND P. F. LINDLEY[d]

[a]*Laboratory of Structure Analysis, Institute of Protein Research, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia,* [b]*Laboratory of Molecular Biology and Imperial Cancer Research Fund Unit, Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, England,* [c]*Department of Ophthalmology and Visual Sciences, Washington University School of Medicine, Campus Box 8096, 660 S. Euclid Avenue, St Louis, MO 63110, USA, and* [d]*CCLRC Daresbury Laboratory, Warrington WA4 4AD, England*

*(Received 3 November 1995; accepted 5 January 1996)*

## Abstract

The crystal structure of bovine lens γIIIb-crystallin at 2.5 Å resolution previously reported was interpreted using a consensus sequence derived from related vertebrate sequences on the assumption that γIIIb-crystallin derived from the γC-crystallin gene. It has recently been shown that γIIIb is a product of the bovine γD gene. The structure of γIIIb has now been refined with the bovine γD sequence using new 1.95 Å resolution synchrotron data. The crystallographic *R* factor was 20.4% for all 33 104 reflection data between 8.0 and 1.95 Å measured at 277(1) K. The electron density fully supported the assignment of the γD sequence to γIIIb. The crystal belongs to space group $P2_12_12_1$ with two molecules of molecular mass 20 749 Da in the asymmetric unit in which 219 water molecules were located. The two-domain four-Greek-key motif highly symmetrical protein is very similar in structure to γB-crystallin (81% sequence identity). There is a single amino-acid deletion in γD in the linker region connecting the two domains. The intermolecular oganization in the crystal lattice is quite different from γB as a result of key mutations involving surface residues Leu51, Ile103 and His155. These point mutations will contribute to the intermolecular behaviour of the γ-crystallins in the eye lens, where they are major components of the densely packed, high refractive index regions of the lens.

## 1. Introduction

Crystallins are intracellular structural proteins which form the transparent body of the vertebrate eye lens. All lenses comprise representatives of the three major classes, α-, β- and γ-crystallins (Wistow & Piatigorsky, 1988; Bloemendal & de Jong, 1991), although a growing number of metabolically related proteins have been found to be over-expressed in lenses of certain lineages, such as δ-crystallin in the avian lens (Piatigorsky & Wistow, 1991). α-Crystallins, which also function as small heat-shock proteins (Klemenz, Fröhli, Steiger, Schäfer & Aoyama, 1991), are expressed in many other tissues whereas the βγ-crystallin superfamily has so far only been found in the vertebrate lens. γ-Crystallins are monomeric with a molecular mass of around 21 000. They comprise two similar domains connected by a single covalent linker and possess a short C-terminal extension (Blundell *et al.*, 1981; Chirgadze, Sergeev, Fomenkova & Oreshin, 1981) whereas the related β-crystallins are oligomeric as a result of domain exchange mediated by a conformational change in the linker (Bax *et al.*, 1990). There are five to seven different gene products of γ-crystallins depending on the species which are diferentially expressed during development resulting in gradients of distribution along the optical axis of the adult lens (Lubsen, Aarts & Schoenmakers, 1988). At present the γ-crystallin family of vertebrates consists of 25 sequences including fish, frog, mouse, rat, calf and human (Chang, Jiang, Chiou & Chang, 1988; Tomarev *et al.*, 1984; Gause, Tomarev, Zinovieva, Arutyunyan & Dolgelvich, 1986; Breitman *et al.*, 1984; Den Dunnen, Moormann, Lubsen & Schoenmakers, 1986; Bhat & Spector, 1984; Hay, Woods, Church & Petrash, 1987; Hay, Andley & Petrash, 1994; Meakin, Du, Tsui & Breitman, 1987).

Crystallins function by providing a dense transparent body mediated by intermolecular protein–protein interactions, which have short-range order (Delaye & Tardieu, 1983). γ-Crystallins are typically found in the region of highest index of refraction and hence must be good packing proteins (Slingsby, 1985). Protein interactions that unbalance the short-range order cause discontinuities in the refractive index, as displayed in 'cold cataract' where a low-temperature phase separation of specific γ-crystallins into two phases of unequal density occurs (Benedek, 1971; Siezen, Fisch, Slingsby & Benedek, 1985). The value of the critical temperature, Tc, for phase separation of γ-crystallins was found to be a useful measure of the protein interaction (Siezen, Christophe & Benedek, 1985) with γB (γII) being described as a low-Tc protein. γ-Crystallins with a high

Tc are potentially cataractogenic as phase separation is likely to occur at body temperature. $\gamma$-Crystallins have a high cysteine content and it has been shown recently that thiol oxidation in $\gamma$B results in a dimer with a greatly increased value of Tc (Pande *et al.*, 1995). These observations underline the need to understand intermolecular interactions of $\gamma$-crystallins. The direct observation of protein–protein interactions in the condensed state is only possible by studying the spatial structure of protein crystals at atomic resolution.

Up to the present time the three-dimensional refined structures of bovine eye-lens $\gamma$-crystallins have been reported as follows: $\gamma$B at 1.47 Å (Najmudin *et al.*, 1993), $\gamma$-IIIb at 2.5 Å (Chirgadze *et al.*, 1991) and $\gamma$-IVa at 2.3 Å (White, Driessen, Slingsby, Moss & Lindley, 1989). The intermolecular interactions in the crystal medium of $\gamma$IIIb were considered in detail (Sergeev *et al.*, 1988). The exact amino-acid sequence is known for $\gamma$B whereas the atomic models of $\gamma$-IIIb and $\gamma$-IVa crystallins were built on consensus sequences of gene products $\gamma$C and $\gamma$E, respectively. This was based on the premise that $\gamma$C in rat and human are low-Tc proteins and rat $\gamma$E is a high-Tc protein (Siezen, Thomson, Kaplan & Benedek, 1987; Siezen, Wu, Kaplan, Thomson & Benedek, 1988), and that bovine $\gamma$IIIb and $\gamma$IVa are low- and high-Tc proteins, respectively (Broide, Berland, Pande, Ogun & Benedek, 1991). Recently the complete cDNA bovine sequences of $\gamma$C and $\gamma$D have been determined, their protein products expressed and the $\gamma$D product, rather than that of $\gamma$C, was shown to correlate with $\gamma$IIIb-crystallin (Hay *et al.*, 1994).

We have now collected higher resolution data from our $\gamma$IIIb crystals. We show here that refinement of the structure using the $\gamma$D sequence confirms the assignment of $\gamma$IIIb as a product of the $\gamma$D gene. Both $\gamma$D and $\gamma$B are low-Tc proteins yet the $\gamma$D crystal lattice is much less densely packed than $\gamma$B. The intermolecular interactions are reanalysed in the light of the correct $\gamma$D sequence and compared with $\gamma$B in order to elucidate critical point mutations that distinguish the packing interactions of these two proteins.

## 2. Experimental

### 2.1. *Data collection*

The crystals of calf $\gamma$D were prepared as described earlier (Chirgadze, Nikonov, Garber & Reshetnikova, 1977; Chirgadze *et al.*, 1991) with the modification that dithiothreitol was used as the reducing agent rather than glutathione. They belong to space group $P2_12_12_1$ with unit-cell dimensions $a = 57.81(3)$, $b = 70.03(3)$, $c = 117.25(3)$ Å with two protein molecules of molecular mass 20749 Da each in the asymmetric unit, corresponding to a solvent content in the crystal of 57% by volume (Matthews, 1968). The crystals are prismatic in habit, elongated about the $c$ axis with typical dimensions 0.4 × 0.4 × 1.2 mm.

Intensity data were collected using the synchrotron radiation source (SRS), at Daresbury Laboratory, operating in multi-bunch mode at an energy of 2GeV and with an average circulating current of 200 mA. Two separate data-collection experiments were undertaken and in each case a crystal was mounted so that the $c$ axis was approximately parallel to a quartz capillary tube axis and cooled to around 277 (1) K. The first experiment utilized station 9.5 equipped with a Pt-coated fused quartz toroidal mirror and a channel-cut Si(111) monochromator set to a wavelength of 0.92 Å. One crystal was used, translated into five different positions to minimize the effect of radiation damage, to collect data through 64.5° starting with the $a$ axis along the X-ray beam. A 180 mm diameter MAR Research image-plate system set at a distance of 170 mm from the crystal was used to record 1.5° oscillations with three separate passes and exposure times varying between 70 and 90 $s^{\circ -1}$ depending on the beam current. The images were processed with *DENZO* (Otwinowski, 1993), followed by scaling and merging (*ROTAVATA* and *AGROVATA*, *CCP*4 program suite, Collaborative Computational Project, Number 4, 1994) to yield 27481 independent reflections, 79.4% of a complete data set to 1.95 Å resolution, from a total of 112001 useable measurements with an overall $R_{\rm merge}$ of 0.097 for 26828 reflections measured more than once.

The second data collection utilized the high-intensity station 9.6 at the SRS equipped with a Pt-coated fused quartz cylindrical mirror and a bent triangular Si(111) monochromator set to give a wavelength of 0.89 Å. A second crystal, translated through four different positions, was used to collect data from $\varphi = -18$ to $+32°$ with the $b$ axis along the beam at $\varphi = 0$. The detector was a Rigaku $R$-AXIS II image-plate system set at a distance of 200 mm from the crystal and 2° oscillations were used with exposure times ranging from 60 to 75 $s^{\circ -1}$. Processing of these data yielded 17081 unique reflections, 45.0% of a complete data set to 1.95 Å resolution, from 72639 useable measurements with an overall $R_{\rm merge}$ of 0.048 for 16344 reflections measured more than once The lower value of $R_{\rm merge}$ probably reflects the higher intensity available from station 9.6 relative to station 9.5.

Scaling and merging of the two data sets simultaneously with *ROTAVATA* and *AGROVATA* yielded 33667 unique reflections, 95.2% of a complete data set (multiplicity 4.3) at 1.95 Å resolution with an overall $R_{\rm merge}$ of 0.081 for 33150 reflections measured more than once. Overall 87.6% of this unique data set had $I > 3\sigma(I)$. The data were corrected for Lorentz and polarization factors, but no correction was made for absorption. A summary of the crystallographic details is given in Table 1.

Table 1. *Summary of the crystallographic data*

| | |
|---|---|
| Unit-cell parameters (Å) | |
| *a* | 57.81 (3) |
| *b* | 70.03 (3) |
| *c* | 117.25 (3) |
| Wavelengths used (Å) | 0.89, 0.92 |
| Nominal resolution (Å) | 1.95 |
| No. of reflections measured | 144932 |
| No. of unique reflections | 33667 |
| Multiplicity | 4.3 |
| Reflections with $I > 3\sigma(I)$ (%) | 87.6 |
| Completeness (%) | 95.2 |
| Merging residual (%) | 8.1 |

## 2.2. *Structure redetermination*

Initially the structure of bovine γIIIb-crystallin was solved at 3.0 Å resolution by the method of multiple heavy-atom isomorphous derivatives with the use of anomalous scattering and the main chain traced (Chirgadze *et al.*, 1981). The structure was then refined at 2.5 Å resolution (Chirgadze *et al.*, 1986) and analysed (Chirgadze *et al.*, 1991). An atomic model was built using a tentative amino-acid sequence that was derived from a consensus sequence of the γC gene product of human (Siezen *et al.*, 1987) and rat (den Dunnen *et al.*, 1986) and the chemically derived bovine γIII sequence (Croft, 1973). Some uncertainties caused by the lack of an exact amino-acid sequence and an electron-density map of limited resolution influenced the quality of the protein model. It has now been established that the sequence of bovine γIIIb corresponds to gene product D (Hay *et al.*, 1994). The structure has therefore been redetermined using new X-ray diffraction data at the limit of resolution for this crystal, 1.95 Å. Initially, the structure redetermination of γD was performed starting from the atomic model of the related bovine γB whose structure has been refined at 1.47 Å resolution (Najmudin *et al.*, 1993). A rotation function using *AMoRe* (Collaborative Computational Project, Number 4, 1994) with *E* values at 2.0–20.0 Å resolution with a Patterson search radius of 25 Å gave two peaks with $\Delta S/\sigma$ of 5.3 (*A*) and 3.9 (*B*). A translation function using *TFFC* (Collaborative Computational Project, Number 4, 1994) at 2.0–20.0 Å resolution gave single peaks with $\Delta S/\sigma$ of 11.6 for *A* not knowing *B*, and with $\Delta S/\sigma$ of 28.2 for *B* knowing *A* (sum of crystallographic and non-crystallographic translation function). These results confirmed the location and orientation of the two molecules in the asymmetric unit of γD according to Chirgadze *et al.* (1986). Rigid-body refinement, two cycles, was then performed using *RESTRAIN* (Driessen *et al.*, 1989), for the two molecules and each of their two domains. This rigid-body model was further refined using *X-PLOR* (Brünger, 1988) giving a structure with a crystallographic *R* factor = 29.6% for the range 8.0–2.3 Å with 20 802 reflections taken with a 1.5 σ cutoff.

## 2.3. *Structure refinement*

Restrained least-squares refinement of the atomic coordinates and isotropic temperature factors of the two protein molecules and crystal water molecules were carried out using *RESTRAIN* (Driessen *et al.*, 1989) on a Convex C220 computer. Model building was carried out with the graphics program *FRODO* (Jones, 1978, as modified by P. Evans) on Evans and Sutherland Picture Systems and the *O* program suite (Jones, Zou, Cowan & Kjeldgaard, 1991) using a Silicon Graphics Indigo Extreme. Electron-density maps were calculated using the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). A structure-factor weighting scheme with modified SIM weights (Read, 1986) was used to calculate electron-density maps. Details of refinement statistics are given in Table 2.

The electron-density map of γD was rebuilt first at 2.3 Å resolution by analyzing two superimposed maps: $2F_o - F_c$ (1.0 r.m.s. contour level) and $F_o - F_c$ (3.0 r.m.s. contour level). From the difference electron-density map it was clearly seen that in most cases the electron density matched the side chains of sequence γD better than the starting model sequence of γB and, importantly, better than the γC sequence. The alignment of the γB, γC and γD sequences (Fig. 1), shows that there are 33 sequence differences between γD and γB. All of them were replaced and built in the electron density. Tracing of the main chain of residues 83–87 of the connecting peptide (shortened by one residue compared with γB) was easily followed and rebuilt. A total of 67 water molecules were included. However, the side chains of a few residues were not clearly defined and so were assigned temporarily to alanines in order to obtain an omit map.

As the last two residues from the C-terminal extensions of both molecules were ill defined in the maps they were kept in the same configuration as in γB and restrained least-squares refinement was then performed. After one cycle of geometry regularization about 40 cycles of refinement were carried out expanding sequentially the limit of resolution from 3.0 to 2.5, 2.2 and finally up to 1.95 Å. Initially the atomic

Table 2. *Details of the refinement of γD-crystallin*

| | |
|---|---|
| Resolution range (Å) | 8.0–1.95 |
| *R* factor for all data in range (%) | 20.4 |
| No. of reflections | 33104 |
| Temperature factors | Isotropic |
| Average $B_{iso}$ of protein atoms (Å$^2$) | 28.3 |
| No. of protein atoms | 2928 |
| No. of water molecules | 219 |
| R.m.s. deviations bond lengths (Å) | 0.014 |
| R.m.s. deviations angles (°) | 2.92 |
| Overall *G* factor | −0.1 |
| Residues in most favoured regions of Ramachandran plot (%) | 89.7 |
| Outliers in Ramachandran plot | None |

positional parameters and overall scale and temperature factors were refined, weak $F_{hkl}$ less than 0.3 $F_{mean}$ and less than $3\sigma$ were omitted. Finally, after one cycle of regularization a refinement of individual temperature factors for all non-H atoms was carried out. The resulting value of $R = 23.7\%$ was obtained for the range 8.0–1.95 Å which includes the structure amplitudes of all available 33 104 independent reflections.

As the electron density improved the temporary alanines were replaced by the $\gamma$D sequence. The remaining ill defined side-chain positions were: molecule *A* 38D, 94E, 95R, 99R, 113Q, 115R, 119N, 140R, 172D; and for molecule *B*: 96D, 140R, 142R, 152R, 153R, 163K, 172D. In addition the electron density of the two C-terminal residues from both molecules remained undefined. The total number of water molecules found increased up to 219. After the next 66 cycles of restrained refinement (including some cycles of geometry regularization) the *R* factor fell to 20.4% for all the measured reflections. The geometry of both molecules was satisfactory with all main- and side-chain stereochemical parameters deviating from the Engh–Huber parameters within the limits expected for the resolution (Table 2). The two C-terminal residues from both molecules remained disordered and the following side chains remained ill defined: A99R, A140R, B152R and B153R.

### 2.4. *Programs used for analysis*

Analysis of the peptide and side-chain torsional angles was carried out using the program *PROCHECK* (Laskowski, MacArthur, Moss & Thornton, 1993), which was also used to evaluate the secondary structure. Calculation of intermolecular lattice contacts (less than 4.0 Å) was carried out with the program *MODEL* by H. Driessen (unpublished work). Conformations were compared and rotational parameters calculated using the superposition program *BESTFIT* using the *Kearsley* algorithm (Kearsley, 1989). Surface accessibilities were calculated using *RESAREA* (Collaborative Computational Project, Number 4, 1994) with a 1.4 Å probe. The model was displayed using *O* (Jones *et al.*, 1991),

```
N-terminal domain

                        ♠♠      ♠♠       ♥♠    ♥          ♥  ♠   ♠ ♠              ♠
                        **      **                          *   * *
Gamma D  GKITFYEDRGFQGRHYECSSDHSNLQP-Y-LGRCNSVRVDS|GCWMIYEQPNYLGPQYFLRRGDYPDYQQWMGLNDSIRSCRLIP
Gamma C  GKITFYEDRGFQGRCYQCSSDCPNLQP-Y-FSRCNSIRVDS|GCWMLYERPNYQGHQYFLRRGDYPDYQQWMGFNDSIRSCCLIS
Gamma B  GKITFYEDRGFQGHCYECSSDCPNLQP-Y-FSRCNSIRVDS|GCWMLYERPNYQGHQYFLRRGDYPDYQQWMGFNDSIRSCRLIP
         1                  20                    40                  60                  80

connecting peptide

         ♠♠♠♠♠
          * *
Gamma D  HAG-S
Gamma C  DTS-S
Gamma B  QHTGT

C-terminal domain

         ♥ ♥ ♥   ♠ ♠    ♠   ♠  ♠       ♠♠ ♥                ♠ ♠                ♠              ♠
         *         *    *      *       *                                       *
Gamma D  HRLRLYEREDYRGQMIEITEDCSSLQDRFHFNEIHSLNVLE|GSWVLYELPNYRGRQYLLRPGEYRRYHDWGAMNAKVGSLRRVI
Gamma C  HRLRLYEREDQKGLIAELSEDCPCIQDRFRLSEVRSLHVLE|GCWVLYEMPNYRGRQYLLRPQEYRRYQDWGAVDAKAGSLRRVV
Gamma B  FRMRIYERDDFRGQMSEITDDCPSLQDRFHLTEVHSLNVLE|GSWVLYEMPSYRGRQYLLRPGEYRRYLDWGAMNAKVGSLRRVM
                  100                 120                 140                 160

C-terminal extension

          ♠

Gamma D  DIY
Gamma C  DLY
Gamma B  DFY
         174
```

Fig. 1. The alignment of bovine $\gamma$D-, $\gamma$C- and $\gamma$B-crystallin sequences. The sequences are aligned as four blocks: the N-terminal domain (residues 1–82), the connecting peptide (residues 83–87), the C-terminal domain (residues 88–171) and the C-terminal extension (residues 172–174). The vertical lines indicate the division of the domains into motifs; the sequences of the two domains are arranged such that topologically equivalent residues are aligned vertically. The numbering is based on $\gamma$B, the numbered residue refers to the last digit. Changes involving surface residues are indicated by ♠, changes to buried residues by ♥, and replacement residue 118 which contributes to both core and surface ◆. The most significant replacements are indicated additionally by *.

*GRASP* (Nicholls, Bharadwaj & Honig, 1993) and *SETOR* (Evans, 1993).

## 3. Results and discussion

### 3.1. *Confirmation that crystals of γIIIb are derived from the gene product γD*

The refined maps clearly showed that the gene *D* sequence (Hay *et al.*, 1994) fitted the electron density. Several examples of side-chain electron density from the N and C domains and linker show how the density distinguishes the γD from the γC sequence (Fig. 2). Superposition of molecules *A* or *B* of γD on γB shows that when the sequence of the linker His-Ala-Gly-Ser is fitted to the density then structurally the deletion is residue Gly86 of γB; accordingly the γD chain is numbered Gly85, Ser87 (Fig. 3).

### 3.2. *General description of the structure*

The refined set of atomic coordinates of the two protein molecules *A* and *B* consists of 2 × 1464 non-H atoms and 219 water O atoms. The mean value of atomic isotropic temperature factors $B_{iso}$ for all protein atoms listed above is equal to 28.3 Å$^2$. The region of main chain with the highest mobility is the C-terminal extension. The atomic temperature factors of the end groups of surface charged residues are higher, particularly certain arginine side chains and equal to about 30–50 Å$^2$. The r.m.s. difference between molecules *A* and *B* of γD is equal to 0.39 Å for the 170 Cα atoms of residues 1–171, excluding the three C-terminal residue outliers (there is no residue 86 to keep the numbering scheme consistent with γB). All peptide planes are in the *trans* conformation. A Ramachandran plot (Ramachandran & Sasisekharan, 1968) of the main-chain torsion angles showed that the majority of points, 70%, occupy the β-extended region, 21% are in the α-R-helical region, 5% in the α-L-helical region and 4% are in the glycine specific region. 89.7% of all residues are in the most favourable regions, and the rest of them are in the additional allowed regions.

The polypeptide-chain backbone tracing of γD is very similar to γB, the main difference being in the connecting peptide. The r.m.s. differences between the N-terminal domains of molecules *A* and *B* of γD with the N-terminal domain of γB are 0.30 and 0.33 Å, respectively, for the 80 equivalent Cα atoms less than 3 Å apart. For the C-terminal domains of molecules *A* and *B* of γD the r.m.s. differences with the C-terminal domain of γB are 0.30 nd 0.30 Å, respectively, also for 80 Cα-atoms less than 3 Å apart. Each molecule of γD consists of an N-terminal domain, comprising residues 1–82, a connecting peptide (residues 83–87), a C-terminal domain comprising residues 88–171 and a short C-terminal extension (residues 172–174). The γB, γC and γD sequences of these modules are aligned in Fig. 1. Each domain comprises two Greek-key motifs related by a pseudo-dyad (Fig. 4*a*) and the N- and C-terminal domains are related by a further pseudo-dyad (Fig. 4*b*). Each motif has topology **d** of a simple 'Greek key' and the domain topology can be described as **d h d** with the internal overlapping hairpin of **h**-type topology (Chirgadze, 1987). Combination of two sequential motifs gives rise to a domain structure with eight extended β-strands which form two well ordered antiparallel twisted β-sheets.

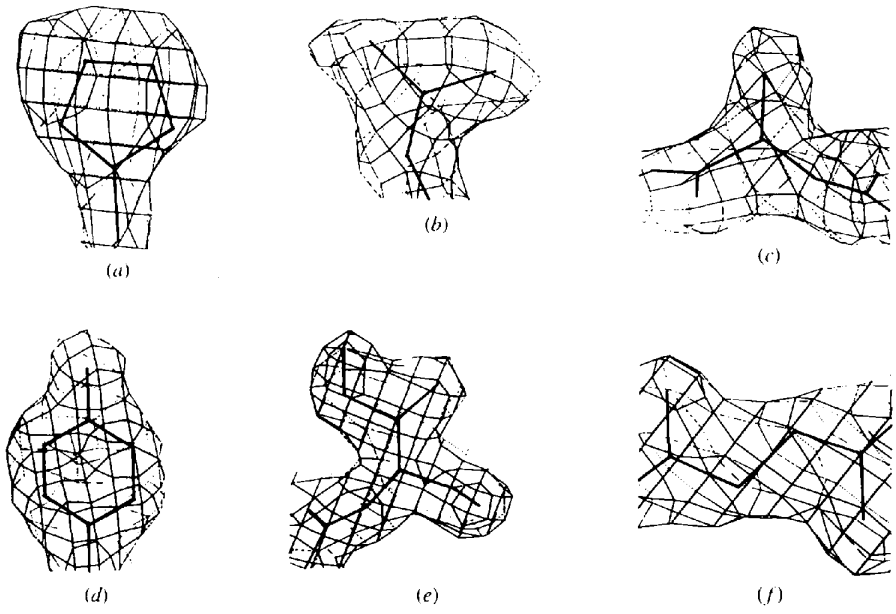



Fig. 2. The sequence of γD fits the electron density. Selected regions of the molecule *A* map, calculated with coefficients $(2|F_o| - |F_c|)$, contoured at 1.0 r.m.s. (*a*) His15(Cys), (*b*) Leu29(Phe), (*c*) Ala84(Thr), (*d*) Tyr98(Gln), (*e*) Ile103(Ala), (*f*) Gly149(Gln), where the residues in brackets refer to those from the γC sequence.

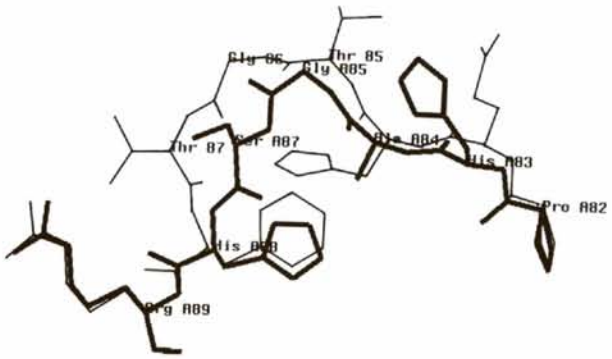



Fig. 3. Conformational differences between the linkers of γD and γB. γ-Crystallins with short linkers (because of a deletion) generally have a histidine at position 88 and a short side chain at position 84. In γD (thick line) the side chain of His88 interacts with the C-terminal residues of the C-terminal domain (not shown), making a side-chain hydrogen bond with O170 and van der Waals with the side chain of Ile171. In γB (thin line) Phe88 makes van der Waals interactions with His84, which in turn hydrogen bonds with 87 O; these interactions will stabilize the conformation of the longer linker.

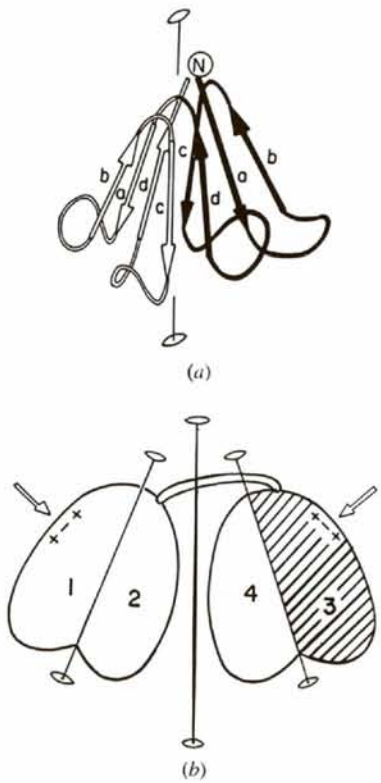
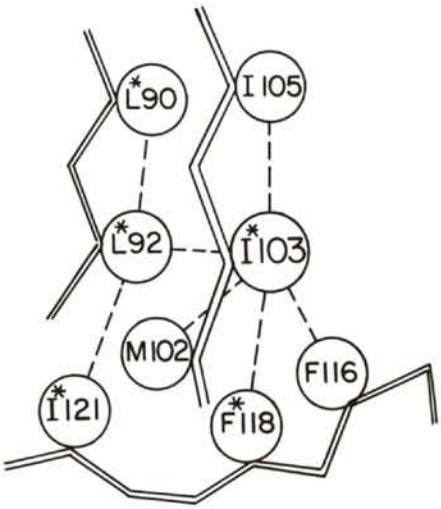
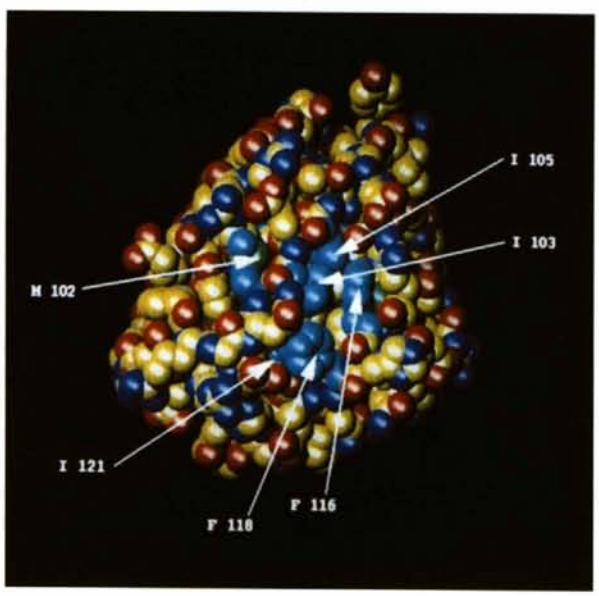



Fig. 4. Schematic spatial model of γD-crystallin. (*a*) Pseudo-symmetry and topology of the peptide backbone in one domain. Two Greek-key motifs are related by an approximate dyad; the first key motif is depicted in black. Each motif comprises four antiparallel β-sheet strands labelled *abcd*. (*b*) Domain organization of the whole protein molecule. Motif 1 and motif 2 form the N-terminal domain while motif 3 and motif 4 form the C-terminal domain. An interdomain pseudo twofold symmetry rotation axis is shown relating the domains about motifs 2 and 4. The most variable third motif is hatched.

Rotation of the N-terminal domain of γD (molecule *A*) onto the C-terminal domain of γD (molecule *A*) gives a rotation angle of 178.7° with an r.m.s. deviation of 0.81 Å for the 80 equivalent Cα atoms less than 3 Å apart. For molecule *B* the rotation is 176.4° with an r.m.s. of 0.82 Å for the same Cα atoms. Another way of




(*a*)




(*b*)

Fig. 5. The surface hydrophobic patch associated with Ile103 in γD-crystallin. (*a*) Schematic diagram showing hydrophobic cluster around surface 'key' residue Ile103 in the variable C-terminal domain. Surface residues Met102, Ile103, Phe116 and Phe118 are in contact with partially exposed residues Ile105 and Ile121 and completely buried residues Leu90, Leu92. Residues marked by an asterisk are replaced in bovine γB. (*b*) Space-filling representation of the C-terminal domain in the same orientation as in (*a*) in which the side chains depicted in (*a*) are coloured in cyan, and the rest of the molecule is shown with C, N, O and S atoms coloured yellow, blue, red and green. Leu90 and Leu92 are completely buried in the core and not visible.

looking at this is a direct fit of the N-terminal domains of molecules *A* and *B* of γD onto each other, and examining the relative positions of the C-terminal domains. The relative r.m.s. difference for the 80 equivalent Cα-atoms of the C-terminal domains mentioned is then 1.2 Å and the angle which would be required to fit the C-terminal domains is 3.8°. Although the numbers calculated vary somewhat with the method of calculation, the suggestion is one of domain movements around the inter-domain dyad.

### 3.3. *Side-chain distribution, comparison with γB*

An earlier structure analysis of bovine IIIb-crystallin had been undertaken on the assumption that it was homologous to the γC crystallins and a consensus sequence was used to interpret the map (Chirgadze *et al.*, 1991). Now that higher resolution data are available and the γIIIb is known to be the gene product γD (Hay *et al.*, 1994), details of the previous structure are here revised and the newly refined γD structure compared with γB.

Sequence identity between γD and γB is 81.0%, there being 32 point mutations and one deletion. Of the four motifs, the most conserved is motif 4 with four replacements and the most variable is motif 3 with 11 mutations (Fig. 1). The majority of replacements are to surface residues (♠ in Fig. 1) and 13 of these can be described as significant (* in Fig. 1). Only seven replacements involve buried residues (♥ in Fig. 1), while residue 118 contributes to both core and surface (◆ in Fig. 1). His88 in γD and Phe88 in γB is a critical sequence change that appears to be correlated with the length of the linker (Fig. 3). The close proximity of this histidine in the C-terminal domain core of γD to the buried Trp131 may have a small effect on the protein's fluorescence spectroscopic properties.

3.3.1. *Replacements that effect surface hydrophobicity/hydrophilicity.* There are 12 positions where side-chain replacements effect the relative hydrophilicity of surface residues. The seven replacements His15Cys His22Cys, Ser23Pro, His88Phe, Tyr98Phe, Ser110Pro and His155Leu represent a situation where γD is more polar than γB, and the five replacements Leu51Gln, Pro53His, Ala84His, Ile103Ser and Phe118Leu are where γD is more hydrophobic than γB. Two of the γD hydrophobic positions, Leu51, Ile103, and one of the hydrophilic γD positions, His155, are involved in lattice contacts (see §3.4).

3.3.2. *Replacements of cysteine residues.* Bovine γD contains five cysteines as compared with seven in γB. The N-terminal domain has four cysteines; Cys32 and Cys78 are completely buried, Cys18 and Cys41 are almost completely buried (surface accessibility of around 2 Å$^2$). Cys109 in the C-terminal domain has a limited accessibility (~7 Å$^2$). It has been shown that in γB Cys18 and Cys22 can partly oxidize to form an intramolecular disulfide bridge (Wistow *et al.*, 1983; Najmudin *et al.*, 1993) although the native protein is completely reduced (Lindley *et al.*, 1993). In γD, as the surface-accessible positions corresponding to Cys15 and Cys22 in γB are replaced by histidines, the molecule should be more resistant to oxidation as the intramolecular S—S bond cannot form and the molecule will be less prone to intermolecular cross-linking.

3.3.3. *Significant replacement Ile103.* It was shown earlier that the molecular packing of γB and γD in the crystal medium is different (Segeev *et al.*, 1988). This is connected mainly with a few so-called 'key' substitutions. One of them is Ile103 (γD) which is replaced by Ser103 in γB. Ile103 is situated in the middle of a large hydrophobic region on the surface of the molecule. Schematically it is shown in Fig. 5(*a*) in which residues Met102, Ile103, Phe116 and Phe118 contribute to the surface structure and conjugate with the underlying hydrophobic core residues Leu90, Leu92, Ile105 and Ile121. It is interesting to note that this hydrophobic cloud is formed by five replaced residues: Met90, Ile92, Ser103, Leu118, Val121 in γB are replaced in γD with Leu90, Leu92, Ile103, Phe118, Ile121. These five replacements constitute a major part of the variability associated with motif 3. The radical Ile103 mutation extends a hydrophobic surface patch on the γ-crystallin C-terminal domain surface (Fig. 5*b*) and its presence there may be stabilized by conservative mutations in the surrounding hydrophobic cloud.

3.3.4. *Charged residues and charged clusters.* γD, like γB, has 22 acidic residues and 22 basic residues but has two extra histidines making a total of seven. Replacement of Gln47 (γD) instead of Arg47 (γB) leads to the loss of one positively charged group situated on the surface of the molecule in the region of an interdomain contact. Overall this charge substitution is partially compensated by the replacement of Arg14 (γD) instead of His14 (γB) which is situated on the exposed part of the N-terminal domain.

Clustering of charged side chains on the surface of the molecule is a very pronounced feature of γ-crystallins of vertebrates (Chirgadze, 1987). The 'hydrophilic potential' of the charged clusters is assumed to be lower compared to a similar number of isolated charged residues or ionic pairs. This provides a higher stability of the condensed protein phase with a lower water content. Two large clusters in γD are localized at opposite sides of the molecule: the N-domain cluster, Lys2, Glu17, Arg36, Asp38, Arg59, Asp61 and the C-domain cluster: Arg89, Glu104, Arg91, Glu128, and they are shown in Fig. 4(*b*) by the arrows. The charged cluster residues are very conserved among different gene products of γ-crystallins of vertebrates (Chirgadze, 1987) and in

accordance with this we have observed the same clusters in the structures of both γB and γD. However, γ-crystallins do have a group of conserved arginines in the interdomain region, namely Arg58, Arg79, Arg147, Arg168 and Arg142.

### 3.4. *Crystal packing and intermolecular contacts*

There are two protein molecules, *A* and *B* in the asymmetric unit and these are arranged in layers in the crystal. Parallel protein chains, 12–20 Å apart, can be visualized within a layer approximately along direction [110] (Fig. 6*a*). The two protein molecules in the asymmetric unit are related to each other by a non-crystallographic pseudo $2_1$ screw axis, also situated approximately along [110] so that the N- and C-terminal domains of the molecules form a chain of 'head-to-tail' contacts. Four layers cross the *z* axis of the unit cell at about 1/8, 3/8, 5/8 and 7/8 being parallel to the *xy* plane. Analysis of the intermolecular contacts using the program *MODEL* and a consideration of the reduction of surface accessibility on forming lattice contacts, show that about 90% of all intermolecular contacts are grouped in four main surface areas on each molecule. A contact region on the N-terminal domain of molecule *A* covering ~260 Å$^2$ (involving residues Arg14, Tyr16, Asn24, Pro27, Tyr28) interacts with a patch on the C-terminal domain of molecule *B* covering a surface area of ~240 Å$^2$ (involving residues Gln101, Met102, Ile103, Glu104, Arg115 and Phe116). A very similar interaction occurs between equivalent contacts involving the symmetry-related molecules and these make up the chain of 'head-to-tail' contacts [designated *A*(C)···*B*(N) and *A*(N)···*B*(C) in Fig. 6*a*]. The 'key' residue Ile103 plays an important role in these intralayer contacts as it completes a large hydrophobic patch on the C-terminal domain as compared with γB (Fig. 5).

The contact areas between layers (Fig. 6*b*) can be grouped into two regions *A*(N)···*B*(N,C) and *A*(N,C)···*B*(N), where the symbol (N,C) designates the surface area in the region of the cleft between N- and C-terminal domains: these regions cover about twice the surface area as the 'head-to-tail' contacts. Residues from the N-terminal domain of molecule *A* in the asymmetric unit (involving residues Arg9, Gly10, Phe11, Gln12, Pro63, Asp64, Tyr65, Gln66, Gln67, Leu71, Asp72, Asp73 and Ser74) covering a surface area of ~560 Å$^2$ interact with a symmetry-related molecule *B* in a different layer (involving residues Asn49, Tyr50, Leu51, Arg79, Arg147, Glu150, Tyr151, Arg152, His155, Asp156 and Gly158) covering a surface area of ~540 Å$^2$. Similar residues in the *A*(N,C)···*B*(N) contact are used except that in molecule *B* there is an additional contribution from Asp8 and Arg31, and a lesser contribution from Leu71, whereas in molecule *A* there is a greater contribution from His155 and less from Arg152 (due to distortion arising from the pseudo $2_1$ axis relating the two molecules in the asymmetric unit). Compared with the 'head-to-tail' chains the interlayer contacts have a greater involvement of polar–polar interactions including a few ionic charge interactions such as between Asp64 and Arg79, Arg147 (Fig. 6*c*). Here we see another two 'key' residues, His155 and Leu51, involved in making intermolecular interactions.

An important difference between the crystal packing of the two proteins is related to the much smaller area buried in intermolecular contacts in γD which is 15% of the total accessible area compared with 33% for γB. Three critical sequence differences between the two proteins are present in this relatively limited contact area and presumably make a strong contribution to the intermolecular interactions in the condensed crystal medium.

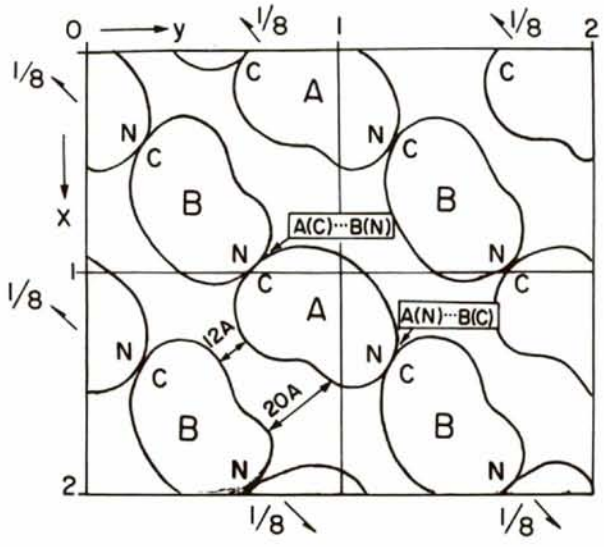
### 3.5. *Conclusions*

γ-Crystallins are a family of closely related structural proteins that are particularly enriched in the densely packed regions of the eye lens. As there are two distinct classes of γ-crystallins in terms of phase separation (Broide *et al.*, 1991), it was anticipated that analysis of the crystal contacts of high- and low-Tc γ-crystallins might show sequence specific intermolecular contact regions. This approach has in the past been severely hampered by incomplete bovine γ-crystallin sequences and a lack of conservation of phase-temperature behaviour between orthologous γ-crystallins from different species (Hay *et al.*, 1994). Analysis of the intermolecular interactions of two low-Tc bovine crystallins, γD and γB, shows that a determining factor in the two quite different packing arrangements can be ascribed to residues Leu51, Ile103 and His155 in γD being replaced with Gln, Ser, and Leu in γB. Although in concentrated solutions crystallin molecules will participate in a range of intermolecular interactions, pairs of molecules interacting as in the more extensive lattice contacts are likely to be favoured. In the case of γD this will be as shown in Fig. 6(*c*) where the positively charged domain interface region interacts with an acidic group (Asp64) on the N-terminal domain of another molecule. This interdomain region is also involved in intermolecular contacts in γB but this time with Asp73 (Najmudin *et al.*, 1993). We, therefore, conclude that the positively charged perimeter of the domain interface presents a preferred site for intermolecular interactions. Further work comparing lattice structures of native γ-crystallins of known sequence and measured Tc, along with closely related mutant proteins of different Tc, will be

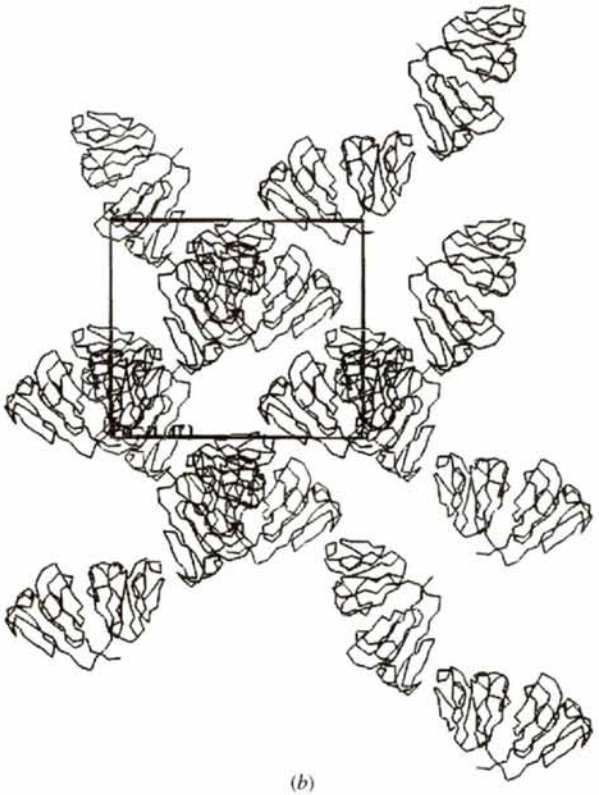
required to address the structural basis of phase separation in γ-crystallins and its role in transparency.*

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1ELP, R1ELPSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: HE0144).

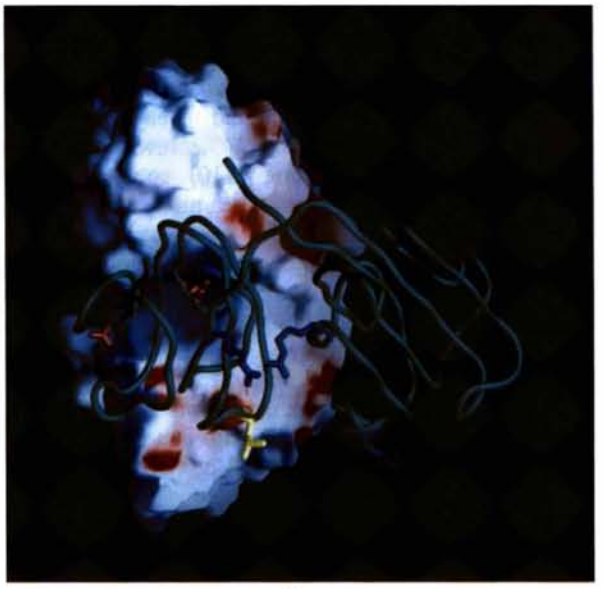
We wish to thank Dr David Moss for many useful discussions. We would like to thank the Royal Society and the Medical Research Council (London) for financial support. One of us (YNC) had additional support from the Russian Basic Research Fund and America ACA/ISCN Fund. We also acknowledge the Wellcome Foundation for provision of computer graphics equipment. REH thanks the National Eye




(a)



(b)



(c)

Fig. 6. Lattice structure of γD. (*a*) Schematic view of the packing of the protein molecules into layers. The two molecules *A* and *B* of the asymmetric unit are related by a non-crystallographic pseudo $2_1$ screw axis. The plane of the layer is situated approximately parallel to the *xy* plane and crosses the *z* axis at 1/8. Symbols N and C designate the corresponding domains. (*b*) Stacking of layers. Cα backbone trace of two parallel chains of 'head-to-tail' γD molecules from one layer interacting with two chains from the next layer, rotated by around 90°. (*c*) Detail of the electrostatic characteristics of the *A*(N,C)···*B*(N) extensive contact that stacks layers together. A space-filling model of molecule *A* is shown interacting with the N-terminal domain of molecule *B*, depicted by the green trace. Appended to the Cα backbone chain of molecule *B* are the positively charged side chains (blue) of Arg79, His155 and Arg147, and the negatively charged (red) side chains of Asp8 and Asp64. The positively charged Arg79, His155, Arg147 of molecule *A* form the large blue patch at the interdomain region. This site interacts with the acidic residues in the N-terminal domain of molecule *B*. Also buried in this contact is the key residue Leu51 of molecule *A* (shown in yellow on molecule *B*). In the calculation of the charge distribution of molecule *A* using the program *GRASP*, all aspartates and glutamates were given one negative charge, lysines and arginines were given one positive charge, His88 was given zero charge as it is largely buried and all other histidines were given 0.5 positive charge.

Institute and Research to Prevent Blindness, Inc. (USA) for financial support.

## References


Bax, B., Lapatto, R., Nalini, Driessen, H., Lindley, P. F., Mahadevan, D., Blundell, T. L. & Slingsby, C. (1990). *Nature (London)*, **347**, 776–780.

Benedek, G. B. (1971). *Appl. Optics*, **10**, 459–473.

Bhat, S. P. & Spector, A. (1984). *DNA*, **3**, 287–295.

Bloemendal, H. & de Jong, W. W. (1991). *Prog. Nucleic Acid Res. Mol. Biol.* **41**, 259–281.

Bloemendal, H., Piatigorsky, J. & Spector, A. (1989). *Exp. Eye Res.* **48**, 465–466.

Blundell, T., Lindley, P., Miller, L., Moss, D., Slingsby, C., Tickle, I., Turnell, B. & Wistow, G. (1981). *Nature (London)* **289**, 771–777.

Breitman, M., Lok, S., Wistow, G., Piatigorsky, J., Tréton, J. A., Gold, R. & Tsui, L. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 7762–7765.

Broide, M. L., Berland, C. R., Pande, J., Ogun, O. O. & Benedek, G. B. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 5660–5664.

Brünger, A. T. (1988). *J. Mol. Biol.* **203**, 803–816.

Chang, T., Jiang, Y., Chiou, Sh. & Chang, W. (1988). *Biochim. Biophys. Acta*, **951**, 226–229.

Chirgadze, Y. N. (1987). *Acta Cryst.* A**43**, 405–417.

Chirgadze, Y., Nevskaya, N., Fomenkova, N., Nikonov, S., Sergeev, Y., Brazhnikov, E., Garber, M., Lunin, V., Urzhumtsev, A. & Vernoslova, E. (1986). *Dokl. Acad. Nauk SSSR*, **209**, 492–495.

Chirgadze, Y., Nevskaya, N., Vernoslova, E., Nikonov, S., Sergeev, Y., Brazhnikov, E., Fomenkova, N., Lunin, V. & Urzhumtsev, A. (1991). *Exp. Eye Res.* **53**, 295–304.

Chirgadze, Y. N., Nikonov, S. V., Garber, M. B. & Reshetnikova, L. S. (1977). *J. Mol. Biol.* **110**, 619–924.

Chirgadze, Y. N., Sergeev, Y. V., Fomenkova, N. A. & Oreshin, V. D. (1981). *FEBS Lett.* **131**, 81–84.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Croft, L. R. (1973). *Ciba Found. Symp.* **19**, 207–226.

Delaye, M. & Tardieu, A. (1983). *Nature (London)*, **302**, 415–417.

Den Dunnen, J., Moormann, R., Lubsen, H. & Schoenmakers, J. (1986). *J. Mol. Biol.* **189**, 37–46.

Driessen, H., Haneef, I., Harris, G., Howlin, B., Khan, G. & Moss, D. (1989). *J. Appl. Cryst.* **22**, 510–516.

Evans, S. V. (1993). *J. Mol. Graphics*, **11**, 134–138.

Gause, G., Tomarev, S., Zinovieva, R., Arutyunyan, K. & Dolgelevich, S. M. (1986). *Topics Aging Res. Eur.* **6**, 171–179.

Hay, R. E., Andley, U. P. & Petrash, J. M. (1994). *Exp. Eye Res.* **58**, 573–584.

Hay, R. E., Woods, W., Church, R. & Petrash, J. M. (1987). *Biochim. Biophys. Res. Commun.* **146**, 332–338.

Jones, T. (1978). *J. Appl. Cryst.* **11**, 268–272.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, (1991). *Acta Cryst.* A**47**, 110–119.

Kearsley, S. K. (1989). *Acta Cryst.* A**45**, 208–210.

Klemenz, R., Fröhli, E., Steiger, R. H., Schäfer, R. & Aoyama, A. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 3652–3656.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lindley, P., Najmudin, S., Bateman, O., Slingsby, C., Myles, D., Kumaraswamy, V. S. & Glover, I. (1993). *J. Chem. Soc. Faraday Trans.* **89**, 2677–2682.

Lubsen, N. H., Aarts, H. J. M. & Schoenmakers, J. G. G. (1988). *Prog. Biophys. Mol. Biol.* **51**, 47–76.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Meakin, S., Du, R., Tsui, L. & Breitman, M. L. (1987). *Mol. Cell Biol.* **7**, 2671–2679.

Najmudin, S., Nalini, V., Driessen, H. P. C., Slingsby, C., Blundell, T. L., Moss, D. S. & Lindley, P. F. (1993). *Acta Cryst.* D**49**, 223–233.

Nicholls, A., Bharadwaj, R. & Honig, B. (1993). *Biophys. J.* **64**, A116.

Otwinowski, Z. (1993). *Data Collection and Processing, Proceedings of the CCP4 Study Weekend*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.

Pande, J., Lomakin, A, Fine, B, Ogun, O., Sokolinski, I. & Benedek, G. B. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 1067–1071.

Piatigorsky, J. & Wistow, G. J. (1991). *Science*, **252**, 1078–1079.

Ramachandran, G. N. & Sasisekharan, V. (1968). *Adv. Protein Chem.* **23**, 283–437.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Sergeev, Yu. V. Chirgadze, Yu. N., Mylvaganam, S. E., Driessen, H. P. C., Slingsby, C. & Blundell, T. L. (1988). *Proteins Struct. Func. Genet.* **4** 137–147.

Siezen, R. J., Christophe, M. & Benedek, G. B. (1985). *Biochim. Biophys. Res. Commun.* **133**, 239–247.

Siezen, R. J., Fisch, M. R., Slingsby, C. & Benedek, G. B. (1985). *Proc. Natl Acad. Sci. USA*, **82**, 1701–1705.

Siezen, R. J., Thomson, J. A., Kaplan, E. D. & Benedek, G. B. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 6088–6092.

Siezen, R. J., Wu, E., Kaplan, E. D., Thomson, J. A. & Benedek, G. B. (1988). *J. Mol. Biol.* **199**, 475–490.

Slingsby, C. (1985). *Trends. Biochem. Sci.* **10**, 281–284.

Tomarev, S., Zinovieva, R., Chalovka, P., Krayev, A., Skryabin, K. & Gause, G. (1984). *Gene*, **27**, 301–308.

White, H. E., Driessen, H. P. C., Slingsby, C., Moss, D. S. & Lindley, P. F. (1989) *J. Mol. Biol.* **207**, 217–235.

Wistow, G. J., Turnell, W., Summers, L. J., Slingsby, C., Moss, D. S., Miller, L. R., Lindley, P. F. & Blundell, T. L. (1983). *J. Mol. Biol.* **170**, 175–202.

Wistow, G. & Piatigorsky, J. (1988). *Ann. Rev. Biochem.* **57**, 479–504.