

# The P23T Cataract Mutation Causes Loss of Solubility of Folded $\gamma$ D-Crystallin

P. Evans<sup>1</sup>, K. Wyatt<sup>2</sup>, G. J. Wistow<sup>2</sup>, O. A. Bateman<sup>1</sup>, B. A. Wallace<sup>1,3</sup> and C. Slingsby<sup>1\*</sup>

<sup>1</sup>Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, UK

<sup>2</sup>National Eye Institute, Room 201, Building 7, National Institutes of Health, Bethesda MD 20892, USA

<sup>3</sup>Centre for Protein & Membrane Structure & Dynamics, Daresbury Laboratory Warrington WA4 4AD Cheshire, UK

Mutations in the human  $\gamma$ D-crystallin gene have been linked to several types of congenital cataracts. In particular, the Pro23 to Thr (P23T) mutation of human  $\gamma$ D crystallin has been linked to cerulean, lamellar, coralliform, and fasciculiform congenital cataracts. We have expressed and purified wild-type human  $\gamma$ D, P23T, and the Pro23 to Ser23 (P23S) mutant. Our measurements show that P23T is significantly less soluble than wild-type human  $\gamma$ D, with P23S having an intermediate solubility. Using synchrotron radiation circular dichroism spectroscopy, we have determined that the P23T mutant has a slightly increased content of  $\beta$ -sheet, which may be attributed to the extension of an edge  $\beta$ -strand due to the substitution of Pro23 with a residue able to form hydrogen bonds. Neither of the point mutations appears to have reduced the thermal stability of the protein significantly, nor its resistance to guanidine hydrochloride-induced unfolding. These results suggest that insolubility, rather than loss of stability, is the primary basis for P23T congenital cataracts.

© 2004 Elsevier Ltd. All rights reserved.

**Keywords:** cataract; crystallin; synchrotron radiation; CD spectroscopy; proline; protein solubility

\*Corresponding author

## Introduction

Cataract is a major cause of blindness in humans,<sup>1</sup> leading to an estimated 20 million cases worldwide.<sup>2</sup> Cataracts arise when the refractive index of the protein-packed lens cytoplasm is disrupted over distances comparable to the wavelength of light, resulting in light-scattering with loss of lens transparency.<sup>3</sup> The major proteins expressed in the mammalian eye lens are the crystallins, comprising the  $\alpha$  and  $\beta\gamma$ -crystallin families. Short-range order between the concentrated components of the lens results in optimum refraction concomitant with minimal light-scattering.<sup>4,5</sup>

The  $\gamma$ -crystallins share a common topology of two domains, each consisting of two intercalated Greek key motifs. In the fetal human lens, mRNAs of  $\gamma$ A,  $\gamma$ B,  $\gamma$ C, and  $\gamma$ D have been reported, with  $\gamma$ C and  $\gamma$ D predominating.<sup>6</sup> Lampi and co-workers have measured levels of crystallin protein through

two-dimensional electrophoresis of lens extract, and identified  $\gamma$ C,  $\gamma$ D, and  $\gamma$ S as the major  $\gamma$ -crystallins of the young human lens.<sup>7</sup> In recent years, a number of point mutations have been mapped to the human CRYG genes, a cluster of four genes ( $\gamma$ A- $\gamma$ D) and two pseudogenes ( $\gamma$ E and  $\gamma$ F), which encode the  $\gamma$ -crystallins.<sup>8</sup> These point mutations cause opacity of the lens in regions of  $\gamma$ -crystallin expression, giving rise to cataract. The molecular basis of several of these point mutations has been investigated. In human  $\gamma$ D, the R14C mutation has been shown to cause a progressive juvenile-onset cataract, which may be mediated through thiol-mediated aggregation, as the normally monomeric  $\gamma$ -crystallins assemble into oligomers.<sup>9</sup> R58H and R36S mutants of human  $\gamma$ D are believed to cause cataract through their spontaneous crystallization under physiological conditions.<sup>10</sup> Crystals of R36S have been extracted from the lens of a patient with congenital cataract.<sup>11</sup> The high-resolution crystal structures of wild-type human  $\gamma$ D and R58H have revealed that the structural changes associated with the increased rate of R58H crystallization are very small.<sup>12</sup> Recently, the substitution of proline by threonine

Abbreviation used: SRCD, synchrotron radiation CD spectroscopy.

E-mail address of the corresponding author: c.slingsby@mail.cryst.bbk.ac.uk

DSSP	1HK0	EEEEETTTTEEEEEESS	BS	TT-T- S	SEEEEE-EEEEETTTTEEEEEESSGGGGT	SSS	EEEE-	S-					
hGD	P07320	1HK0			GKITYEDRGRFQGRHYECSSDHPNLQP-Y-LSRCSNARVDS-GCWMLYEQPNYSGLQYFLRRGDYADHQQWMLGSDSVRSCLRI-PHSG-S				86				
bGD	P08209	1ELP			GKITYEDRGRFQGRHYECSSDHPNLQP-Y-LGRCSNARVDS-GCWMLYEQPNYLGQYFLRRGDYADHQQWMLGSDSVRSCLRI-PHAG-S				86				
hGA	P11844				GKITYEDRGRFQGRHYECSSDHPNLQP-Y-FSRCSNIRVDS-GCWMLYERPNYQGHQYFLRRGKYPDYQHWMLGSDSVQSCRLI-PHTS-S				86				
hGB	P07316				GKITYEDRGRFQGRHYECSSDHPNLQP-Y-FSRCSNIRVES-GCWMLYERPNYQGHQYFLRRGKYPDYQHWMLGSDSVRSCLLI-PPHSGA				87				
hGC	P07315				GKITYEDRGRFQGRHYECSSDHPNLQP-Y-FSRCSNIRVES-GCWMLYERPNYQGHQYFLRRGKYPDYQHWMLGSDSVRSCLLI-PQTV-S				86				
bGB	P02526	1AMM			GKITYEDRGRFQGRHYECSSDHPNLQP-Y-FSRCSNIRVDS-GCWMLYERPNYQGHQYFLRRGDYADHQQWMLGSDSVRSCLRI-PQHTGT				87				
bGE	Q28088	1M8U			GKITYEDRGRFQGRHYECSSDHPNLQP-Y-FSRCSNIRVDS-GCWMLYEQPNFQGPYFLRRGDYADHQQWMLGSDSVRSCLRI-PHTS-S				86				
bGF	P23005	1A45			GKITYEDRGRFQGRHYECSSDHPNLQP-Y-FSRCSNIRVDS-GCWMLYEQPNFQGPYFLRRGDYADHQQWMLGSDSVRSCLRI-PHTG-S				86				
rGA	P10065				GKITYEDRGRFQGRHYECSSDHPNLQP-Y-FSRCSNIRVDS-GCWMLYERPNYQGHQYFLRRGDYADHQQWMLGSDSVRSCLRI-PYTS-S				86				
rGB	P10066				GKITYEDRGRFQGRHYECSSDHPNLQP-Y-FSRCSNIRVDS-GCWMLYERPNYQGHQYFLRRGDYADHQQWMLGSDSVRSCLRI-PQHSST				87				
rGC	P02529				GKITYEDRGRFQGRHYECSSDHPNLQP-Y-FSRCSNIRVDS-GCWMLYERPNYQGHQYFLRRGDYADHQQWMLGSDSVRSCLRI-PHTG-S				86				
rGD	P10067				GKITYEDRGRFQGRHYECSSDHPNLQP-Y-FSRCSNIRVDS-GCWMLYEQPNFTGCQYFLRRGDYADHQQWMLGSDSVRSCLRI-PHAG-S				86				
rGE	P02528	1A5D			GKITYEDRGRFQGRHYECSSDHPNLQP-Y-FSRCSNIRVDS-GCWMLYEQPNFTGCQYFLRRGDYADHQQWMLGSDSVRSCLRI-PHSS-S				86				
rGF	P10068				GKITYEDRGRFQGRHYECSSDHPNLQP-Y-FSRCSNIRVDS-GCWMLYEQPNFTGCQYFLRRGDYADHQQWMLGSDSVRSCLLI-PHSS-S				86				
mGA	P04345				GKITYEDRGRFQGRHYECSSDHPNLQP-Y-FSRCSNIRVDS-GCWMLYERPNYQGHQYFLRRGDYADHQQWMLGSDSVRSCLRI-PYTS-S				86				
mGB	P04344				GKITYEDRGRFQGRHYECSSDHPNLQP-Y-FSRCSNIRVDS-GCWMLYERPNYQGHQYFLRRGKYPDYQHWMLGSDSVRSCLLI-PQHSST				87				
mGC	Q61597				GKITYEDRGRFQGRHYECSSDHPNLQP-Y-FSRCSNIRVDS-GCWMLYERPNYQGHQYFLRRGKYPDYQHWMLGSDSVRSCLRI-PHAG-S				87				
mGD	P04342				GKITYEDRGRFQGRHYECSSDHPNLQP-Y-FSRCSNIRVDS-GCWMLYEQPNFAGCQYFLRRGDYADHQQWMLGSDSVRSCLRI-PHAG-S				86				
mGE	P26999				GKITYEDRGRFQGRHYECSSDHPNLQP-Y-FSRCSNIRVDS-GCWMLYEQPNFTGCQYFLRRGDYADHQQWMLGSDSVRSCLLI-PHSS-S				86				
β-strands					-----								
					a	b	c	d					
DSSP	1HK0	EEEEEEGGGEEEEESS	BS	GGGTSS	SB	EEEEEE-S	EEEEETTTTEEEEEESS	SEEE	SGGGT	SS	B	EEEE-	
hGD	P07320	1HK0			HRIRLYEREDYRGMVEITDDCPSLQDRFHLSEIHSNLVLE-GSWVLYELSNYRGRQYLLRPGDYRRYHDWGMANAKVGSLLRRV-IDFS								173
bGD	P08209	1ELP			HRIRLYEREDYRGMVEITDDCPSLQDRFHLSEIHSNLVLE-GSWVLYELSNYRGRQYLLRPGDYRRYHDWGMANAKVGSLLRRV-IDYI								173
hGA	P11844				HKLRLYERDDYRGLMSELTDCCACVPELFLRPEIYSLHVLVLE-GCWVLYEMPNYRGRQYLLRPGDYRRYHDWGMANAKVGSLLRRV-TDLY								173
hGB	P07316				YRMKIYDRDELQGMSELTDCCACVPELFLRPEIYSLHVLVLE-GSWVLYEMPNYRGRQYLLRPGDYRRYHDWGMANAKVGSLLRRV-MDLY								174
hGC	P07315				HRIRLYEREDYRGMVEITDDCPSLQDRFHLSEIHSNLVLE-GCWVLYELSNYRGRQYLLRPGDYRRYHDWGMANAKVGSLLRRV-VLDY								173
bGB	P02526	1AMM			FRMRIYERDDYRGMVEITDDCPSLQDRFHLSEIHSNLVLE-GSWVLYEMPNYRGRQYLLRPGDYRRYHDWGMANAKVGSLLRRV-MDFY								174
bGE	Q28088	1M8U			HRIRLYEREDYRGMVEITDDCPSLQDRFHLSEIHSNLVLE-GWVWLYEMPNYRGRQYLLRPGDYRRYHDWGMANAKVGSLLRRV-VDFY								173
bGF	P23005	1A45			HRIRLYEREDYRGMVEITDDCPSLQDRFHLSEIHSNLVLE-GWVWLYEMPNYRGRQYLLRPGDYRRYHDWGMANAKVGSLLRRV-VDFY								173
rGA	P10065				HRIRLYERDDYRGLMSELTDCCACVPELFLRPEIYSLHVLVLE-GSWVLYEMPNYRGRQYLLRPGDYRRYHDWGMANAKVGSLLRRV-MDLY								173
rGB	P10066				YRMRIYERDDYRGMVEITDDCPSLQDRFHLSEIHSNLVLE-GCWVLYEMPNYRGRQYLLRPGDYRRYHDWGMANAKVGSLLRRV-MDFY								174
rGC	P02529				HRIRLYEREDYRGMVEITDDCPSLQDRFHLSEIHSNLVLE-GCWVLYEMPNYRGRQYLLRPGDYRRYHDWGMANAKVGSLLRRV-VLDY								173
rGD	P10067				HRIRLYEREDYRGMVEITDDCPSLQDRFHLSEIHSNLVLE-GCWVLYEMPNYRGRQYLLRPGDYRRYHDWGMANAKVGSLLRRV-MDFY								173
rGE	P02528	1A5D			HRIRLYEREDYRGMVEITDDCPSLQDRFHLSEIHSNLVLE-GYVWLYEMPNYRGRQYLLRPGDYRRYHDWGMANAKVGSLLRRV-MDFY								173
rGF	P10068				HRIRLYEREDYRGMVEITDDCPSLQDRFHLSEIHSNLVLE-GYVWLYEMPNYRGRQYLLRPREYRRYHDWGMANAKVGSLLRRV-MDYI								173
mGA	P04345				HRIRLYERDDYRGLMSELTDCCACVPELFLRPEIYSLHVLVLE-GCWVLYEMPNYRGRQYLLRPGDYRRYHDWGMANAKVGSLLRRV-MDLY								173
mGB	P04344				YRMRIYERDDYRGMVEITDDCPSLQDRFHLSEIHSNLVLE-GCWVLYEMPNYRGRQYLLRPGDYRRYHDWGMANAKVGSLLRRV-MDFY								174
mGD	P04342				HRIRLYEREDYRGMVEITDDCPSLQDRFHLSEIHSNLVLE-GCWVLYEMPNYRGRQYLLRPGDYRRYHDWGMANAKVGSLLRRV-VLDY								174
mGD	P04342				HRIRLYEREDYRGMVEITDDCPSLQDRFHLSEIHSNLVLE-GCWVLYEMPNYRGRQYLLRPGDYRRYHDWGMANAKVGSLLRRV-MDFY								173
mGE	P26999				HRIRLYEREDYRGMVEITDDCPSLQDRFHLSEIHSNLVLE-GYVWLYEMPNYRGRQYLLRPGDYRRYHDWGMANAKVGSLLRRV-MDFY								173
β-strands					-----								
					a	b	c	d	a	b	c	d	

**Figure 1.** Alignment of human  $\gamma$ D with other  $\gamma$ -crystallins, numbered according to PDB coordinates. The  $\beta$ -strands a1-d4 are annotated according to standard bovine  $\gamma$ B classification. DSSP codes: E,  $\beta$ -strand; T, hydrogen bonded turn; B, isolated  $\beta$ -bridge; S, bend; G, 3/10 helix. Pro23 is highlighted in red. Ser23 is highlighted in yellow. The N-terminal residues that show correlation with either Pro/Ser23 include 15, 20, 47, 50 and 75. With the exception of the human P23S mutant and bovine  $\gamma$ D, placement of serine at position 23 is accompanied by Phe50 instead of Tyr50.

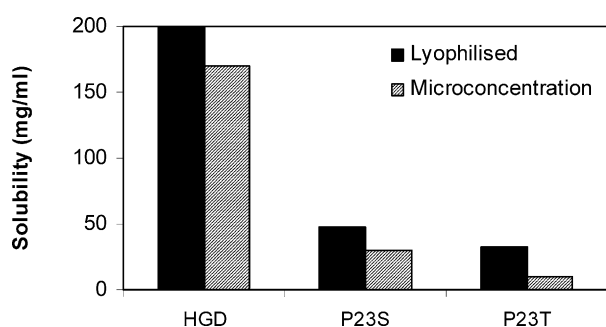
at position 23 of human  $\gamma$ D has been implicated in several different autosomal dominant cataract phenotypes, including cerulean cataract,<sup>13</sup> lamellar cataract,<sup>14</sup> coralliform cataract,<sup>15,16</sup> fasciculiform cataract,<sup>17</sup> and an unclassified silica-like nuclear cataract.<sup>18</sup> Pro23 is in a highly conserved region of the protein. Interestingly, alignments of  $\gamma$ -crystallins from several species show that serine is sometimes found at this position (Figure 1). Given the similarity between serine and threonine side-chains, we have studied both P23T and P23S mutations of human  $\gamma$ D.

In this work, we show that the P23T and P23S mutations are both significantly less soluble than their wild-type counterpart (H $\gamma$ D). We have used synchrotron radiation circular dichroism spectroscopy to study these proteins in solution, and we have detected a small alteration in secondary structure between the wild-type human  $\gamma$ D and the P23T cataract mutant. The P23T and P23S mutant proteins retain the high-level stability of the wild-type as shown by guanidinium hydrochloride (Gdn-HCl)-induced unfolding studies, as well as remaining soluble and folded at high temperatures.

## Results

### Expression and purification of H $\gamma$ D, P23T, and P23S

Recombinant wild-type and mutant H $\gamma$ D crystallins were expressed in *Escherichia coli* and purified from the soluble fraction. During purification, all three proteins behaved in a similar manner, eluting at similar concentration of salt on Q-Sepharose (2% (w/v) NaCl) and Mono-S (7.5% (w/v) NaCl) columns. However, whilst desalting by buffer-exchange between columns, heavy precipitation was observed in the P23T solution. The solution was filtered (2  $\mu$ m pore size) and purified as normal. Whilst the yield of H $\gamma$ D was in the region of 20 mg per litre of bacterial culture, the yields from P23T and P23S purifications were less than 15 mg per litre of bacterial culture. Samples of all three proteins produced single bands upon SDS-PAGE using 20% (w/v) acrylamide homogeneous Phast gels.



**Figure 2.** Relative solubility limits of human  $\gamma$ D crystallin samples, as determined by methods of lyophilisation and microconcentration. Solubility was measured in distilled water, and determined by absorbance spectroscopy at 280 nm. The concentration of H $\gamma$ D exceeded 200 mg/ml.

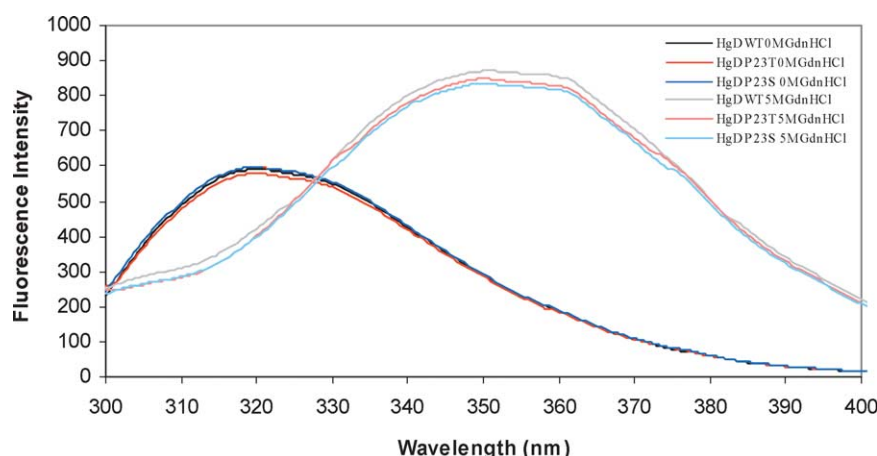
### Maximum solubility of crystallin mutants

To determine the solubility limits of the crystallins, the concentration of samples was raised by two alternative methods: samples were centrifuged in microconcentrators until the volume and

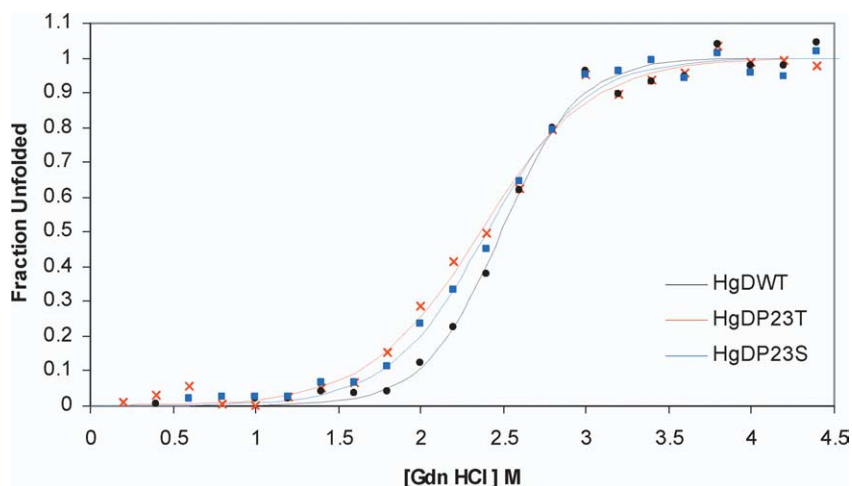
concentration of the solution remained constant. When concentrated by this method, H $\gamma$ D reached significantly greater concentrations than either mutant, 170 mg/ml compared to 28 mg/ml (P23S) and 10 mg/ml (P23T) (Figure 2).

Alternatively, by redissolving lyophilised crystallin samples to increasing concentrations, apparent solubility limits were found to be in excess of 200 mg/ml for H $\gamma$ D, approximately 48 mg/ml for P23S, and 30 mg/ml for P23T. Whilst solubility measurements by the two methods show a similar trend, the absolute concentrations obtained by the ultrafiltration method for each sample are less than those obtained by the dissolution method. This is probably due to a localised increase in protein concentration at the filtration membrane, resulting in protein precipitation blocking the membrane pores.

Efforts were made to crystallize the protein. The first approach was directed towards conditions similar to those used for crystallization of wild-type  $\gamma$ D crystallin and the R58H mutant, which are similar to the solution phase separation conditions of wild-type  $\gamma$ -crystallins.<sup>10,12,19</sup> The poor solubility of the P23T mutant results in precipitation of the



**Figure 3.** Tryptophan fluorescence spectra of human  $\gamma$ D crystallin samples at 37 °C. Protein concentrations were 0.02 mg/ml, in 10 mM sodium phosphate buffer (pH 7.0), 5 mM DTT, 1 mM EDTA. Unfolded samples were equilibrated in 5 M Gdn-HCl for seven hours.



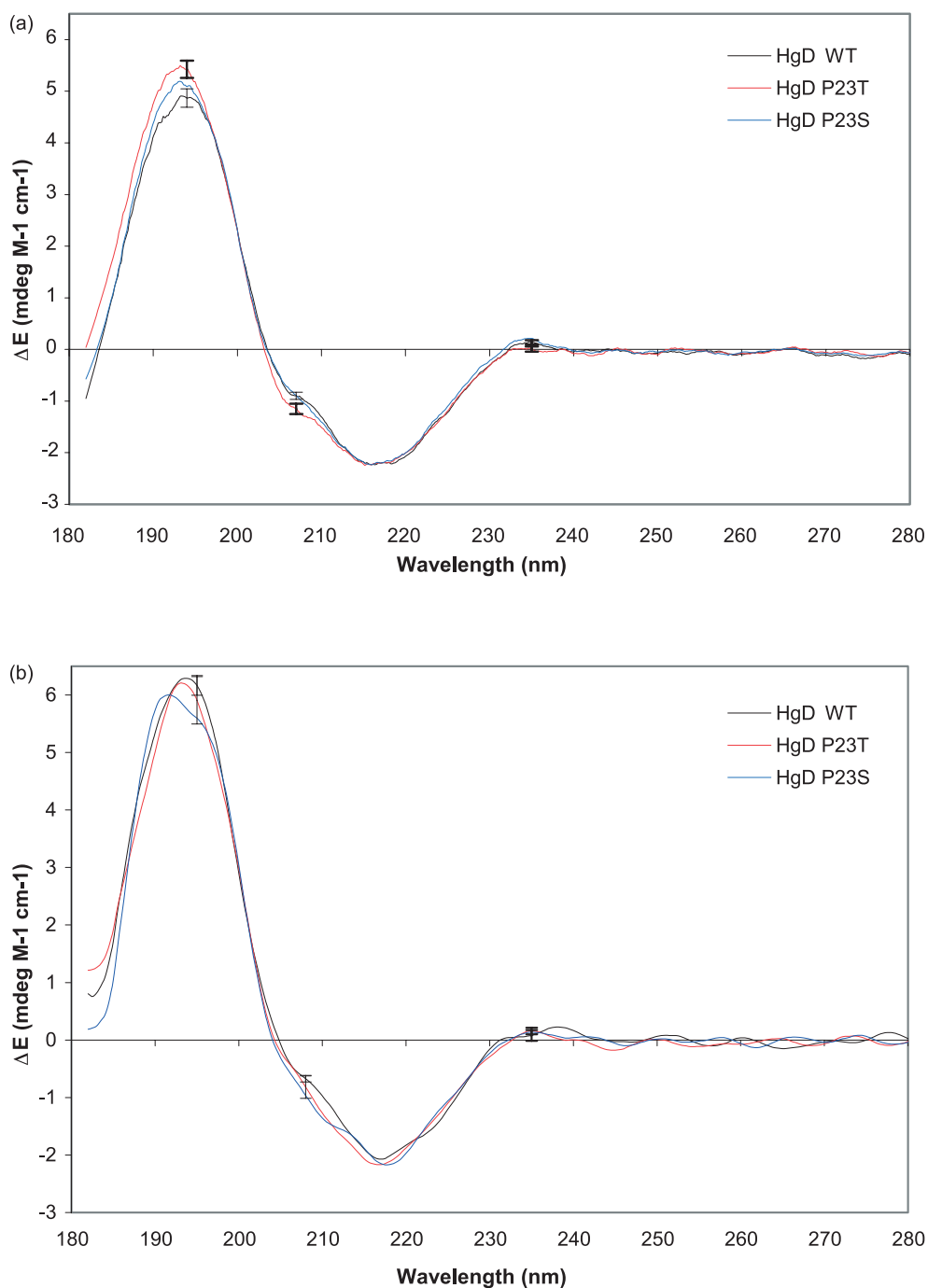
**Figure 4.** Gdn-HCl-induced unfolding of H $\gamma$ D, P23T and P23S at 37 °C, pH 7.0. Unfolding was monitored by emission at 350 nm, following tryptophan excitation at 280 nm. Samples of 0.02 mg/ml concentration were unfolded in increasing concentration of Gdn-HCl for seven hours.

protein into a non-crystalline form as judged by light microscopy. Wider screening of the P23T mutant under a variety of conditions did not yield single crystals.

### Gdn-HCl-induced unfolding

Following tryptophan excitation at 280 nm, the fluorescence emission spectra of P23T and P23S

were found to be identical with that of H $\gamma$ D, with an emission maximum at 320 nm, demonstrating that the tryptophan residues in all three proteins are buried. When fully unfolded in 5 M Gdn-HCl, all three crystallins produced identical spectra with emission maximum of 350 nm (Figure 3). The unfolded states showed a substantial increase in fluorescence intensity relative to the folded state, suggesting that several of the four tryptophan



**Figure 5.** (a) Far-UV CD spectra of H $\gamma$ D, P23T and P23S collected at 4 °C. (b) Far-UV CD spectra of H $\gamma$ D, P23T and P23S at 70 °C. All spectra were collected at 0.05 mg/ml in 10 mM sodium phosphate buffer (pH 7.0). Error bars representing one standard deviation in the repeated measurements from three sequential scans are shown on WT and P23T spectra. These spectra demonstrate that H $\gamma$ D, P23T and P23S retain their native secondary structure at high temperature. WT is represented by a black line, P23T by a red line, and P23S by a blue line.

residues in H $\gamma$ D undergo significant fluorescence quenching in the native state.

To establish if the point mutations P23T and P23S had altered the conformational stability of the crystallins, samples were unfolded in increasing amounts of Gdn-HCl at close to physiological buffer conditions of pH 7.0, 37 °C, in the presence of 1 mM EDTA and 5 mM DTT. Unfolding was measured by the increase in tryptophan emission at 350 nm, the  $\lambda_{\text{max}}$  of fully unfolded  $\gamma$ D-crystallin. Under these conditions, H $\gamma$ D underwent a single unfolding transition between  $\sim$ 2–3 M Gdn HCl (Figure 4). P23T and P23S also exhibited very similar transitions upon unfolding. The mutant transitions were very slightly broader, and midpoints were shifted by  $<$ 0.1 M Gdn (with the P23S mutant being intermediate between wild-type and P23T). These studies indicate that the P23T and P23S point mutations have retained the high level of conformational stability of the wild-type protein to chaotropic agents, although the broader transition of the P23T mutant provides some evidence that the domain bearing the mutation is destabilised marginally and unfolding at lower concentrations of Gdn-HCl than in wild-type.

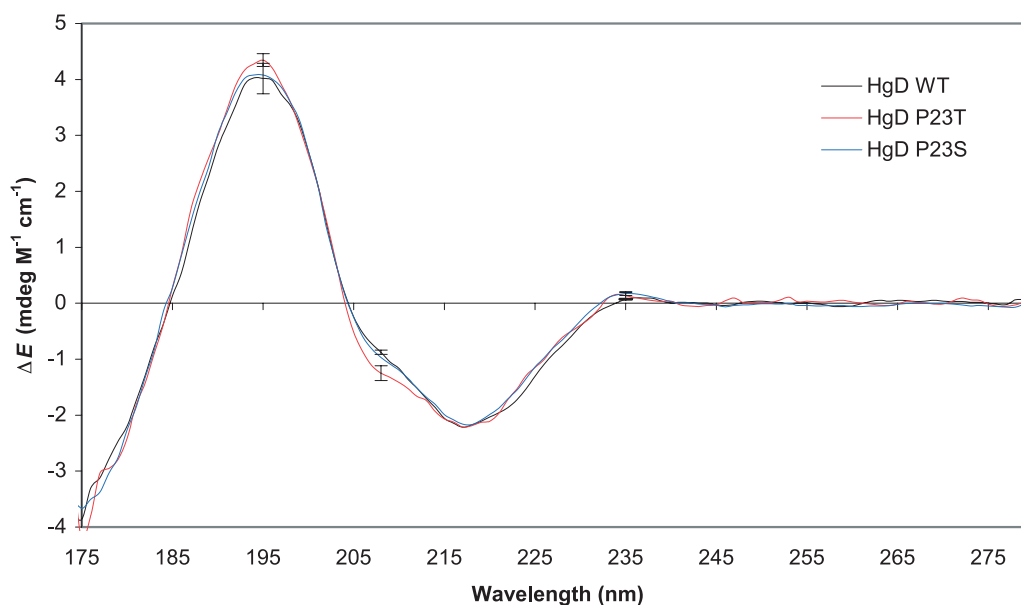
### Circular dichroism (CD) spectroscopy of H $\gamma$ D crystallin and mutants

The far-UV CD spectrum of a 0.05 mg/ml solution of H $\gamma$ D, collected at 4 °C on a conventional CD instrument, displayed the characteristic weak signal of a  $\beta$ -sheet protein with a broad negative trough centred at 218 nm and a more intense, positive peak at 194 nm (see Figure 5). The

spectrum of P23T collected under the same conditions was very similar, with maxima at 218 nm and 194 nm. The magnitude of the peak at 194 nm exceeded that of H $\gamma$ D by approximately 10%. In addition, a pronounced shoulder centred at 208 nm was observed in this spectrum. By comparison, the spectrum of P23S superposes well with that of H $\gamma$ D, having maxima and minima at 219 nm and 194 nm, respectively, and a weak shoulder at 208 nm. This shoulder is larger than that of the wild-type protein, but less intense than that of P23T, as is the magnitude of the 194 nm peak. However, as the photon flux from a conventional CD light source declines rapidly at lower wavelengths, error levels for signals collected in this region are significant. In particular, the magnitude of peaks at 194 nm is influenced heavily by noise.

### Synchrotron radiation CD (SRCD) spectroscopy

The differences between the CD spectra of P23T and P23S, and that of H $\gamma$ D are at the lower limits of significance, due to the low light flux at these wavelengths resulting in low signal-to-noise levels in conventional CD instruments. To determine if the shoulder observed in P23T at 208 nm is a reproducible and significant feature, spectra were collected on two separate SRCD instruments (SRS, Daresbury, UK, and ISA, Aarhus, Denmark<sup>20</sup>) that have considerably higher light fluxes, and thus produce spectra with improved signal-to-noise levels. SRCD spectra show that the shoulder at 208 nm is a feature of P23T that differs from H $\gamma$ D well in excess of experimental variances (Figure 6), although the magnitude difference at 194 is still



**Figure 6.** SRCD spectra of H $\gamma$ D, P23T and P23S. Sample concentrations were 2 mg/ml, in 10 mM sodium phosphate buffer (pH 7.0). Error bars representing one standard deviation in the repeated measurements from three sequential scans are shown on WT and P23T spectra. Mutant and wild-type spectra have been measured using protein samples prepared from different bacterial growth batches, and at different synchrotrons; the distinction between mutant and wild-type is always maintained.

(just) within the experimental error.<sup>20</sup> The magnitude of the 208 nm shoulder of P23S is smaller than that of P23T, falling within one standard deviation of the H $\gamma$ D measurement, and so it cannot be considered to be significantly different from wild-type. Similar results were obtained on two different SRCD instruments, using two different sets of protein preparations.

### Thermal stability of H $\gamma$ D, P23T, and P23S measured by CD spectroscopy

Wild-type human  $\gamma$ D remains folded at temperatures greatly in excess of physiological conditions. To determine if the point mutations had affected the thermal stability of H $\gamma$ D, samples of H $\gamma$ D, P23T, and P23S, were equilibrated at 70 °C for ten minutes, and CD spectra in the far-UV region were collected. The resulting H $\gamma$ D spectra were compared to equivalent spectra obtained at 4 °C (Figure 5). The spectra of wild-type protein at the higher temperature resembled closely those collected at 4 °C. Significantly, P23T and P23S CD spectra collected at 70 °C appear unchanged at the elevated temperature, indicating that these point mutations retain the high-level thermal stability of the wild-type protein.

### Analyses of CD spectra using a custom crystallin-specific reference database

SRCD spectra collected from the SRS facility were analyzed using the SELCON3 program<sup>21</sup> to provide information on the structural basis of the spectral changes in structure resulting from P23T and P23S mutations. To perform these analyses, a custom version of a 29 protein reference set<sup>22</sup> was created, so that it would incorporate features more representative of the types of secondary structures and folds found in crystallin proteins. The CD spectra of two crystallins whose crystal structures were known, human  $\beta$ B2 (M. A. Smith & C.S. unpublished results) and bovine  $\gamma$ B,<sup>23</sup> were added to the standard reference set, and the spectrum of bovine  $\gamma$ B, which appears to be erroneous in the standard reference database was deleted, to improve the accuracy of the crystallin deconvolutions. This improved the analyses (as indicated by the goodness-of-fit parameter normalised root mean square deviation) and the fit of wild-type to the known crystal structure.<sup>12</sup> Analysis of the wild-type H $\gamma$ D data produced a secondary structure of 10% helix, 40%  $\beta$ -strand, 19% turn, with  $\sim$ 30% classified as "other". These values can be compared with the DSSP<sup>24</sup> analysis of the H $\gamma$ D crystal structure (1HK0): 8% helix, 49%  $\beta$ -strand, 14% turn, and 29% other. The secondary structural analysis of P23T produced 9% helix, 43%  $\beta$ -strand, 18% turn, and  $\sim$ 30% other. The  $\beta$ -strand content of P23T is slightly but significantly ( $>1\%$ ) different from that of wild-type, with a concomitant reduction in turn: this would correspond to an increase in several amino acid residues in a  $\beta$ -strand conformation in

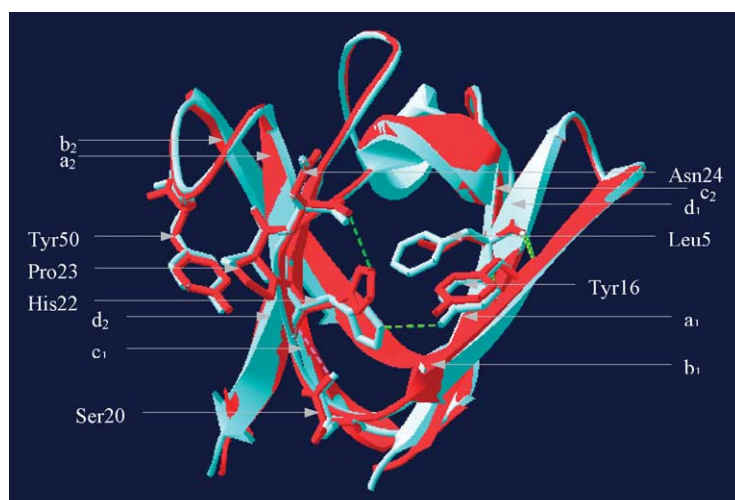
the mutant. The differences in the calculated values for P23S, whilst intermediate between those of P23T and wild-type, were below the significance level.

## Discussion

From the results of solubility experiments, it is apparent that the P23T mutation results in human  $\gamma$ D being less soluble than the wild-type protein. Whilst P23T reached a maximum concentration of 33 mg/ml, lyophilised wild-type protein could be solubilized at concentrations exceeding 200 mg/ml. This is consistent with the previous observation (A. Pande and co-workers) that by melting crystals of human  $\gamma$ D into phosphate buffer, it is possible to reach concentrations of up to 300 mg/ml.<sup>10</sup> Interestingly, the P23S mutation also shows a substantial, but less dramatic reduction in solubility than P23T, reaching a maximum concentration of 48 mg/ml. However, other  $\gamma$ -crystallins do contain serine in this position (Figure 1) without apparent reductions in solubility, for example bovine  $\gamma$ D ( $\gamma$ IIIb), bovine  $\gamma$ E ( $\gamma$ IIIa) and bovine  $\gamma$ F ( $\gamma$ IVa).<sup>19</sup>

It is assumed frequently that crystallin mutations resulting in cataract cause a significant destabilisation of the protein.<sup>25</sup> For example, the T5P mutation of human  $\gamma$ C, associated with Coppock-like cataract, renders the protein insoluble in overexpression and is accompanied by a conformational change and reduction in thermal stability.<sup>26</sup> The effects of P23T and P23S mutations on the stability of human  $\gamma$ D, however, were very small. Wild-type human  $\gamma$ D has previously been shown to be a stable protein, capable of resisting denaturation at high concentrations of urea.<sup>27</sup> In this study, Gdn-HCl-induced unfolding of human  $\gamma$ D at 37 °C results in an apparent two-state process with a transition mid-point occurring at approximately 2.5 M Gdn-HCl. These observations agree with the previous work on H $\gamma$ D by J. King and co-workers.<sup>27</sup> In the presence of this chaotropic agent, P23T and P23S mutants show stability similar to that of wild-type human  $\gamma$ D. Bovine  $\gamma$ D shows very similar behavior, appearing slightly less stable than the human proteins (data not shown). Furthermore, comparisons of the far-UV CD spectra collected at 70 °C and 4 °C have demonstrated the substantial thermal stability of human  $\gamma$ D and both of the mutants, and indicated that neither point mutation had induced a significant decrease in thermal stability.

The protein secondary structure was considered as a potential source of difference between the cataract-inducing mutant and wild-type human  $\gamma$ D. Simplistically, it might have been thought that replacement of the conformation-restricted proline residue with threonine could change the local conformation, by perhaps removing a bend in the polypeptide backbone, which might result in the cataract-susceptibility, although in several  $\gamma$ -crystallin structures serine in this position has the same torsion angles. The CD spectrum of P23T showed a



**Figure 7.** Overlay of the N-terminal domains of H $\gamma$ D (1HK0) and bovine  $\gamma$ D (1ELP). H $\gamma$ D is coloured red, bovine  $\gamma$ D is coloured blue. Pro/Ser23 side-chains are shown, as well as Leu/Phe5, His22, Tyr16, Ser20 and Asn24. His22 of H $\gamma$ D hydrogen bonds with Asn24 and forms a weak hydrogen bond with Ser20, whereas His22 of bovine  $\gamma$ D hydrogen bonds only with Tyr16. The side-chain oxygen atom of Tyr50 in B $\gamma$ D points directly towards the backbone nitrogen atom of Ser23, and in H $\gamma$ D the side-chain oxygen atom of Tyr50 is directed towards the backbone nitrogen atom of Pro23, so the mutation to Ser/Thr may allow an

increase in association between residue 23 and Tyr50. The Ser23 side-chain oxygen atom is directed out into solvent, whilst the carbonyl oxygen atom hydrogen bonds with the backbone nitrogen atom of Asn49. Images were prepared using Swiss PDB viewer.

small but significant difference compared with wild-type. In order to make certain these differences were greater than the variations in the measurements, it was necessary to use SRCD. Differences in a single shoulder in the spectrum at 208 nm, observed in data from both SRCD facilities, fall significantly outside the (one standard deviation) error bars of wild-type human  $\gamma$ D spectra. Deconvolution of these spectra with an extended reference set has predicted several small alterations in the secondary structure. Most notable is a small decrease in residues in a turn conformation, with a commensurate increase in  $\beta$ -sheet content. The high-resolution crystal structure of human  $\gamma$ D<sup>12</sup> reveals that Pro23 is located on the edge of the  $\beta$ -sheet formed by  $\beta$ -strands b2-a2-d2-c1 (Figure 7). DSSP<sup>24</sup> analysis of these crystal coordinates indicates that His22 forms an isolated  $\beta$ -bridge, whilst Pro23 is in a bend leading into the turn formed by residues 26–28, which crosses back to  $\beta$ -sheet 1 (Figure 1). Given the prominence with which proline is observed at the edges of  $\beta$ -sheets<sup>28</sup> where, lacking an NH group, it is unable to form hydrogen bonds,<sup>29,30</sup> Pro23 may be crucial in maintaining the correct conformation of this region in human  $\gamma$ D. In addition, studies of  $\beta$ -sheet proteins have shown that the inclusion of a proline residue in the edge strand of a  $\beta$ -sandwich is a strategy used commonly to minimize the potential for edge-to-edge aggregation events.<sup>31</sup> In the light of the CD spectroscopy results, a plausible explanation for the altered solution behavior of the mutant protein is that the replacement of proline with threonine has facilitated the extension of the c1  $\beta$ -strand on the edge of a  $\beta$ -sheet.

The reduction in solubility of P23S, however, is surprising, as serine is found in this position in other  $\gamma$ -crystallins (Figure 1) without any apparent solubility problems. Bovine  $\gamma$ D, for instance, is soluble at concentrations up to several hundred mg/ml,<sup>19</sup> and the main-chain structure superposes

closely with human  $\gamma$ D in this region (Figure 7). The P23S mutant shows a similar but smaller trend in secondary structure alteration than does the P23T mutant; however, these differences are near the level of significance in the measurements. Bovine  $\gamma$ D, on the other hand, which has ten sequence differences in the N-terminal 82 residues, has a CD spectrum distinct from both the wild-type and mutant  $\gamma$ D discussed here (data not shown). Sequence comparisons demonstrate that, except for human  $\gamma$ D, a proline residue at position 23 in mammalian  $\gamma$ -crystallins is accompanied by a cysteine residue at position 22 (Figure 1). Human  $\gamma$ D coordinates show a weak hydrogen bond between backbone atoms of Ser20 and His22, with the His22 side-chain hydrogen bonding with the Asn24 backbone (Figure 7). In contrast, coordinates of bovine  $\gamma$ D,  $\gamma$ E, and  $\gamma$ F, and rat  $\gamma$ E, all of which have His22-Ser23 pairings, show a loss of the Ser20-residue22 hydrogen bond and rotation of the His22 side-chain into a position where it hydrogen bonds with Tyr16 (Figure 7). In addition, with the exception of bovine  $\gamma$ D, a serine or proline residue at position 23 is accompanied at position 50 by phenylalanine or tyrosine, respectively, which may affect the hydrogen bonding patterns (Figure 7). In the vicinity of these residues, human  $\gamma$ D differs specifically in the core residue Leu5, which in other crystallins is conserved as phenylalanine (Figure 7). Thus, whilst the c1 strand of human  $\gamma$ D does not differ significantly from bovine  $\gamma$ D, side-chain hydrogen bonding changes caused by sequence differences at positions 5, 22 and 50 in the locality of core residue Leu5 may explain why serine at position 23 of human  $\gamma$ D is associated with loss of solubility. Without crystal structures of P23S and P23T it is not possible to elucidate the conformational changes in further detail, but this work suggests that the structural context of human  $\gamma$ D does not permit the replacement of serine or

threonine on an edge  $\beta$ -sheet strand without further changes and loss of solubility.

In conclusion, we have demonstrated that the observed autosomal dominant congenital cataracts<sup>13–18</sup> associated with the P23T mutation result from the significantly reduced solubility of P23T, which represents another instance of a cataract mutation that exhibits reduced solubility without a major loss of stability. We have suggested that a small increase in the fraction of  $\beta$ -sheet content in P23T, perhaps located at the periphery of the protein, may contribute to the physical basis for the precipitation of this protein. Our work on P23T demonstrates that it is possible to have both a decrease in solubility, and a local alteration of secondary structure, whilst maintaining protein stability.

## Materials and Methods

### Expression and purification of proteins

Clones for the complete coding sequence of P23T and P23S mutants were synthesized by PCR, using a wild-type cDNA obtained through NEIBank<sup>32</sup> as template. P23T and P23S point mutations were introduced into the respective clones through the 5' primers. PCR products were subcloned into the TOPO PCR II vector (Invitrogen, Carlsbad, CA), and confirmed by restriction digest and DNA sequencing before being subcloned into pET-17b (Novagen, Madison WI), and transformed into *E. coli* strain BL21(DE3)pLysS. Initial purification was by Q-Sepharose anion-exchange column chromatography (Amersham Biosciences) in 25 mM Tris-HCl (pH 8.0), 1 mM DTT, eluted on 0–1 M NaCl gradient. This was followed with a buffer desalt by pressure-filtration through a 10 kDa cutoff Amicon membrane, into 50 mM Mes (pH 6.05), 1 mM DTT. The final purification stage was an 8 ml Mono-S cation-exchange column (Amersham Biosciences), with fractions eluted by a 0–1 M NaCl gradient.

### Solubility measurements

Relative solubility limits of H $\gamma$ D, P23T, and P23S were measured by two methods. Firstly, samples were concentrated in Amicon Microcon10 micro-concentrators, centrifuged at 4000g at 20 °C, until the retained volume remained constant. Samples were agitated regularly to prevent membrane blockage. Alternatively, samples were concentrated by re-dissolving lyophilized protein aliquots into distilled water (MilliQ) at increasing concentration. Concentrations of H $\gamma$ D, P23T, and P23S were determined by measuring the absorption at 280 nm and using an extinction coefficient of 4.14 mM<sup>-1</sup> cm<sup>-1</sup>.<sup>33</sup>

### CD and SRCD spectroscopy

CD spectra were obtained on an Aviv 62DS spectrometer (Aviv Biomedical, Lakewood, USA), with data collected at 0.2 nm intervals over a wavelength range from 180 nm to 300 nm. Samples were measured at a protein concentration of 0.05 mg/ml, at 4 °C, in 10 mM sodium phosphate (pH 7.0), using a 0.1 cm pathlength Suprasil cuvette (Hellma Ltd). Three scans of each spectrum were averaged, smoothed, and baselines (of

buffer alone) subtracted. SRCD spectra were collected on Station CD12 at the SRS (Daresbury Laboratory, UK), and at Station UV1 at ISA (Aarhus, Denmark). Spectra were collected at 2 mg/ml in cuvettes with pathlengths of 0.05 cm, at a temperature of 4 °C in 10 mM sodium phosphate (pH 7.0). Spectra were averaged over three scans, and baselines subtracted. Spectra collected at 70 °C were under sample conditions identical with those used at 4 °C, and were equilibrated at 70 °C for ten minutes. Both conventional CD and SRCD instruments were calibrated using camphor sulphonic acid.<sup>34</sup> Spectra were converted to delta epsilon units using a mean residue weight value of 119.1. Sample concentrations were determined by quantitative amino acid analysis. Data processing was performed using the CDTool software suite (J. Lees *et al.*<sup>35</sup>).

### Secondary structure analyses

Secondary structure analyses were performed on the Daresbury SRCD data using the self-consistent method SELCON3,<sup>21</sup> available as part of the CDPRO software package. A modified set of 30 reference proteins was used, based on 29 reference proteins from Johnson *et al.*<sup>22</sup> with the addition of SRCD spectra of human  $\beta$ B2-crystallin and bovine  $\gamma$ B crystallin (4GCR). The original CD spectrum of bovine  $\gamma$ B crystallin included with the CDPRO package appeared to be incorrect (shifted by 10 nm relative to the actual spectrum), and so was removed from the reference dataset. All other original reference spectra were examined and found to be free of this error.

### Guanidine hydrochloride-induced unfolding

A stock solution of Gdn-HCl was prepared in 10 mM sodium phosphate buffer (pH7.0) containing 5 mM DTT and 1 mM EDTA. The concentration of Gdn-HCl was determined by measuring the refractive index on an ABBE mark II refractometer (Reichert, USA). Crystallin samples were diluted into a range of concentrations of Gdn-HCl from 0 M to 5 M, to a final concentration of crystallin of 0.02 mg/ml. Samples were equilibrated for seven hours at 37 °C. Fluorescence spectra were obtained with a Hitachi F-2500 spectrometer equipped with a thermostatically controlled circulating waterbath, with excitation at 280 nm. The excitation and emission slits were set to 10 nm, and emission spectra were collected from 300 nm to 450 nm in a 1 cm pathlength cell (Hellma Ltd). Spectra were corrected to a rhodamine B standard to compensate for instrumental response, and baseline spectra of Gdn-HCl solutions from 0 M to 5 M were subtracted. Emission intensity at 350 nm was used for data analysis.

## Acknowledgements

We thank J. Lees, A. Miles, Dr F. Wien (Birkbeck College) and Dr Søren Vrønning Hoffmann (ISA) for their help with SRCD data collection. We thank M. Smith for providing samples of human  $\beta$ B2, Dr R. Sarra and J. Macdonald (all Birkbeck College) for helpful discussions, and P. Sharett of the Cambridge PNAC Facilities for quantitative amino acid analyses. This work was supported by BBSRC grants to

B.A.W. and beamtime grants from the CLRC, and the European Community–Research Infrastructure Action under the FP6 “Structuring the European Research Area” (to S. P. Møller, ISA). P.E. was supported by a BBSRC studentship, and C.S. and O.A.B. acknowledge the financial support of the MRC.

## References

- Vijaya, R., Gupta, R., Panda, G., Ravishaknar, K. & Kumaramanickavel, G. (1997). Genetic analysis of adult-onset cataract in a city-based ophthalmic hospital. *Clin. Genet.* **52**, 427–431.
- Foster, A. (1999). Cataract—a global perspective: output, outcome and outlay. *Eye*, **13**, 449–453.
- Benedek, G. B. (1997). Cataract as a protein condensation disease: the Proctor lecture. *Invest. Ophthalmol. Vis. Sci.* **38**, 1911–1921.
- Benedek, G. B. (1971). Theory of transparency of the eye. *Appl. Optics*, **10**, 459–473.
- Delaye, M. & Tardieu, A. (1983). Short-range order of crystallin proteins accounts for eye lens transparency. *Nature*, **302**, 415–417.
- Brakenhoff, R. H., Aarts, H. J. M., Reek, F. H., Lubsen, N. H. & Schoenmakers, J. G. G. (1990). Human gamma-crystallin genes—a gene family on its way to extinction. *J. Mol. Biol.* **216**, 519–532.
- Lampi, K. J., Ma, Z., Shih, M., Shearer, T. R., Smith, J. B., Smith, D. L. & David, L. L. (1997). Sequence analysis of  $\beta$ A3,  $\beta$ B3 and  $\beta$ A4 crystallins completes the identification of the major proteins in young human lens. *J. Biol. Chem.* **272**, 2268–2275.
- Hejtmancik, J. F. (1998). The genetics of cataract: our vision becomes clearer. *Am. J. Genet.* **62**, 520–525.
- 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.
- 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.
- Kmoch, S., Brynda, J., Asfaw, B., Bezouska, K., Novak, P., Rezacova, P. *et al.* (2000). Link between a novel human  $\gamma$ D-crystallin allele and a unique cataract phenotype explained by protein crystallography. *Hum. Mol. Genet.* **9**, 1779–1786.
- Basak, A., Bateman, O., Slingsby, C., Pande, A., Asherie, N., Ogun, O. *et al.* (2003). High-resolution X-ray crystal structures of human gamma D crystallin (1.25 Å) and the R58H mutant (1.15 Å) associated with aculeiform cataract. *J. Mol. Biol.* **328**, 1137–1147.
- Nandrot, E., Slingsby, C., Basak, A., Cherif-Chefchaoui, M., Benazzouz, B., Hajaji, Y. *et al.* (2003). Gamma-D crystallin gene (CRYGD) mutation causes autosomal dominant congenital cerulean cataracts. *J. Med. Genet.* **40**, 262–267.
- Santhiya, S. T., Shyam Manohar, M., Rawlley, D., Vijayalakshmi, P., Namperumalsamy, P., Gopinath, P. M. *et al.* (2002). Novel mutations in the  $\gamma$ -crystallin genes cause autosomal dominant congenital cataracts. *J. Med. Genet.* **39**, 352–358.
- Xu, W. Z., Zheng, S., Xu, S. J., Huang, W., Yao, K. & Zhang, S. Z. (2004). Localization and screening of autosomal dominant coralliform cataract associated gene. *Zhonghua Yo Xue Yi Chuan Xue Za Zhi*, **21**, 19–22.
- Mackay, D. S., Andley, U. P. & Shiels, A. (2004). A missense mutation in the  $\gamma$ D crystallin gene (CRYGD) associated with autosomal dominant “coral-like” cataract linked to chromosome 2q. *Mol. Vis.* **10**, 155–162.
- Shentu, X., Yao, K., Xu, W., Zheng, S., Hu, S. & Gong, X. (2004). Special fasciculiform cataract caused by a mutation in the  $\gamma$ D-crystallin gene. *Mol. Vis.* **10**, 233–239.
- Burdon, K. P., Wirth, M. G., Mackey, D. A., Russell-Eggitt, I. M., Craig, J. E., Elder, J. E. *et al.* (2004). Investigation of crystallin genes in familial cataract, and report of two disease associated mutations. *Br. J. Ophthalmol.* **88**, 79–83.
- 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.
- Wallace, B. A., Wien, F., Miles, A. J., Lees, J. G., Hoffmann, S. V., Evans, P. *et al.* (2004). Biomedical applications of synchrotron radiation circular dichroism spectroscopy: identification of mutant proteins associated with disease and development of a reference database for fold motifs. *Faraday Discuss.* **126**, 237–243.
- Sreerema, N. & Woody, R. W. (1993). A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal. Biochem.* **209**, 32–44.
- Manavalan, P. & Johnson, W. C., Jr (1987). Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Anal. Biochem.* **167**, 76–85.
- Najmudin, S., Nalini, V., Driessen, H. P. C., Slingsby, C., Blundell, T. L., Moss, D. S. & Lindley, P. F. (1993). Structure of the bovine eye lens protein gamma-B-crystallin at 1.47 Å. *Acta Crystallog.* **49**, 223–233.
- Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure-pattern-recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
- Héon, E., Priston, M., Schorderet, D. F., Billingsley, G. D., Girard, P. O., Lubsen, N. & Munier, F. L. (1999). The  $\gamma$ -crystallins and human cataracts: a puzzle made clearer. *Am. J. Hum. Genet.* **65**, 1261–1267.
- Fu, L. & Liang, J. J. (2002). Conformational change and destabilization of cataract gammaC-crystallin T5P mutant. *FEBS Letters*, **513**, 213–216.
- Kosinski-Collins, M. S. & King, J. (2003). *In vitro* unfolding, refolding, and polymerization of human gamma D crystallin, a protein involved in cataract formation. *Protein Sci.* **12**, 480–490.
- Reiersen, H. & Rees, A. R. (2001). The hunchback and its neighbours: proline as an environmental modulator. *Trends Biochem. Sci.* **26**, 679–684.
- Chou, P. Y. & Fasman, G. D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Advan. Enzymol. Relat. Areas Mol. Biol.* **47**, 45–148.
- Chou, P. Y. & Fasman, G. D. (1978). Empirical predictions of protein conformation. *Annu. Rev. Biochem.* **47**, 251–276.
- Richardson, J. S. & Richardson, D. C. (2002). Natural  $\beta$ -sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl Acad. Sci. USA*, **99**, 2754–2759.
- Wistow, G. (2002). A project for ocular bioinformatics: NEIBank. *Mol. Vis.* **8**, 161–163.
- 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, PA.

- 
34. Miles, A. J., Wien, F., Lees, J. G., Rodger, A., Janes, R. W. & Wallace, B. A. (2003). Calibration and standardisation of synchrotron radiation circular dichroism and conventional circular dichroism spectrophotometers. *Spectroscopy*, **17**, 653–661.
35. Lees, J. G., Smith, B. R., Wien, F., Miles, A. J. & Wallace, B. A. (2004). *CDTool*—An integrated software package for circular dichroism spectroscopic data processing, analysis and archiving. *Anal. Biochem.* **332**, 285–289.

*Edited by Sir A. Klug*

*(Received 15 March 2004; received in revised form 2 June 2004; accepted 10 August 2004)*