J Mol Evol (1990) 30:140–145

Journal of
Molecular Evolution

Evolution of a Protein Superfamily: Relationships between Vertebrate Lens Crystallins and Microorganism Dormancy Proteins

Graeme Wistow

Room 204, Building 6, LMDB, National Eye Institute, National Institutes of Health, Bethesda, Maryland 20982, USA

Summary. A search of sequence databases shows that spherulin 3a, an encystment-specific protein of *Physarum polycephalum,* is probably structurally related to the β- and γ-crystallins, vertebrate ocular lens proteins, and to Protein S, a sporulation-specific protein of *Myxococcus xanthus.* The β- and γ-crystallins have two similar domains thought to have arisen by two successive gene duplication and fusion events. Molecular modeling confirms that spherulin 3a has all the characteristics required to adopt the tertiary structure of a single γ-crystallin domain. The structure of spherulin 3a thus illustrates an earlier stage in the evolution of this protein superfamily. The relationship of β- and γ-crystallins to spherulin 3a and Protein S suggests that the lens proteins were derived from an ancestor with a role in stress-response, perhaps a response to osmotic stress.

Key words: Lens crystallins — *Physarum polycephalum* — Spherulins — Protein evolution — Stress

Introduction

The lenses of vertebrates contain high concentrations of soluble structural proteins upon whose intermolecular interactions the transparency of the tissue depends (Wistow and Piatigorsky 1988). Surprisingly, many of these proteins have turned out to be common enzymes with important non-structural roles in other tissues (Wistow and Piatigorsky 1988; Piatigorsky and Wistow 1989). However, the β- and γ-crystallins, members of a single superfamily (the βγ-crystallin superfamily), seem to be truly lens-specific and are represented in all vertebrate lenses. X-ray analysis has revealed a highly symmetrical structure for bovine γII-crystallin (γB) (see Bloemendal et al. 1989 for nomenclature), with two similar domains, each of which is composed of two very similar supersecondary structures or motifs (Blundell et al. 1981; Wistow et al. 1983). Protein S, a development-specific protein of the bacterium *Myxococcus xanthus* and a major component of the bacterial spore, seems to be an evolutionarily distant, non-lens member of this superfamily (Wistow et al. 1985). It seems likely that the βγ-crystallin tertiary fold represents an ancient and useful structure that has been used and reused many times during protein evolution and consequently is widely distributed. With this in mind sequence databases have been surveyed for other related structures, revealing an interesting relationship with spherulin 3a of the slime mold *Physarum polycephalum.*

Spherulins are development-specific proteins synthesized in response to various kinds of stress leading to encystment and dormancy (Raub and Aldrich 1982; Bernier et al. 1987). The product of the major spherulation-specific mRNA is spherulin 3a (Bernier et al. 1987). Examination of the sequence of this protein strongly suggests that it has a tertiary structure similar to that of the symmetrical, repeated domains of the βγ-crystallins (Wistow and Piatigorsky 1988). It has been proposed that the βγ-crystallin superfamily arose by successive gene duplication and fusion events (Fig. 1), resulting in multiple genes encoding proteins with two very similar globular domains (Driessen et al. 1980; Blundell et al. 1981; Wistow et al. 1981). Clearly this model requires that ancestral one-domain forms should

Offprint requests to: G. Wistow

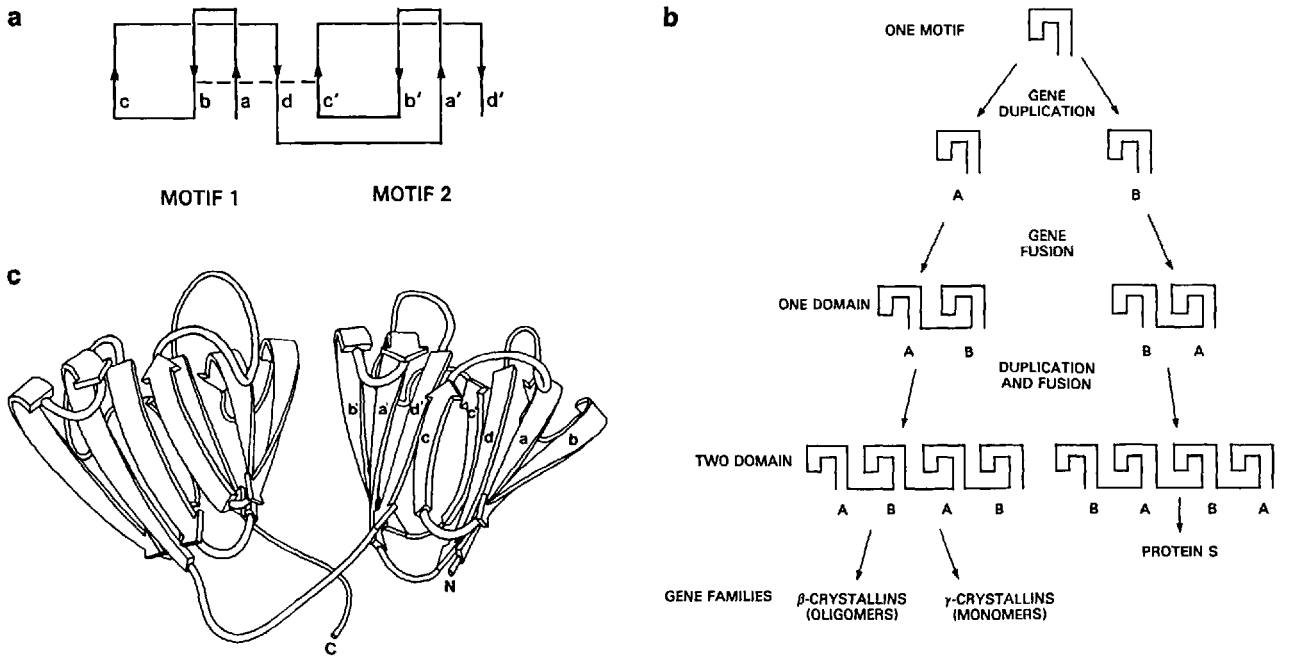

Fig. 1. **a** A schematic representation of the characteristic γ-crystallin folding pattern for one domain, corresponding to two similar structural motifs related by a pseudo twofold axis. Two β-sheets are formed and one, with strands b, a, d, c', is indicated by connecting dashed lines. See also Fig. 3. **b** The evolution of the βγ-crystallin superfamily. A gene coding for a single structural motif arose in the ancient period of inventiveness that laid the basis for modern proteins. This structure would probably have existed only as a dimer. Subsequent duplication and fusion gave rise to two-motif structures analogous to spherulin 3a or single βγ-crystallin domains. Further duplications gave rise to the two-domain structures exemplified by γ-crystallin. From the structure of Protein S it has been suggested that the initial fusion of two motifs may have occurred in both possible arrangements in different lines of descent (Wistow et al. 1985; Wistow and Piatigorsky 1988). **c** The structure of bovine γB-crystallin (Blundell et al. 1981; Wistow et al. 1983). β-strands are represented by arrows. N- and C-termini and the β-strands of the N-terminal domain are marked. Drawing is shown courtesy of J. Richardson (Duke University).

have existed and may still exist as stable structures. Indeed, attempts are currently underway to engineer a one-domain version of a γ-crystallin (Blundell et al. 1988). Spherulin 3a is probably a naturally occurring example of such a structure, analogous to an ancestor of the two-domain members of the βγ-crystallin superfamily. The relationship between β- and γ-crystallins and these encystment or sporulation-specific proteins may suggest the nature of the major functional role of this protein superfamily and the basis for the recruitment of the crystallins as lens structural proteins.

Methods

Database Search. Members of the βγ-crystallin superfamily have a characteristic pattern of amino acid residues (the βγ fold) on which their tertiary structure depends (Blundell et al. 1981; Wistow et al. 1983). This represents a useful template for discrimination of distantly related structures. The basis of the template is the unusual fold-over feature involving a turn between strands a and b and an interaction with residues of strand d of each of the four-stranded structural motifs (Fig. 1). In motif 1 of bovine γB (Fig. 2) the important residues are Tyr6 (an aromatic is required), Phe11 (aromatics are preferred), Gly13 (absolutely conserved), and Ser34 (almost always conserved). Hydrophobic residues contributing to the protein core are required at other positions and there is broad conservation of other residue types. There is some variability in lengths of connecting loops, particularly those between the c and d strands of each motif. Because of the allowed variability, it is most convenient to search databases with known βγ-crystallin protein sequences at a fairly high stringency and then examine the limited number of sequences detected for the presence of the template.

The program SEQFT of the IDEAS package (Kanehisa 1986) was used to translate and search GenBank v60, corresponding to 26,317 different nucleic acid sequence entries. Bovine γB-crystallin (Bhat and Spector 1984) and *M. xanthus* Protein S (Inouye et al. 1983) were used as probe sequences. Search parameters were MAXD = −65, KTPL = 2, NW = 10, DEL = 8. Using Protein S as a probe, nine matches to distinct translated sequences were obtained. Four of these (two *Escherichia coli* noncoding sequences and open reading frames from a bacterial amylase and a chicken laminin binding protein) did not contain the required structural template. The remaining five all gave a good fit to the template; one was the self match for Protein S and three were crystallins, mouse γD-crystallin and bovine γD- and βB1-crystallins. The fifth was spherulin 3a (Bernier et al. 1987) (called 2b in GenBank), a major encystment-specific protein of *P. physarum*. The sequence of this small protein (103 amino acid residues) contains a twofold repeat of the template required for βγ-crystallin motif structure and is therefore equivalent to a single domain of a γ-crystallin (Fig. 2). Indeed, like γ-crystallin, spherulin 3a is predicted to have mainly β-sheet secondary structure (Bernier et al. 1987; G.W., data not shown).

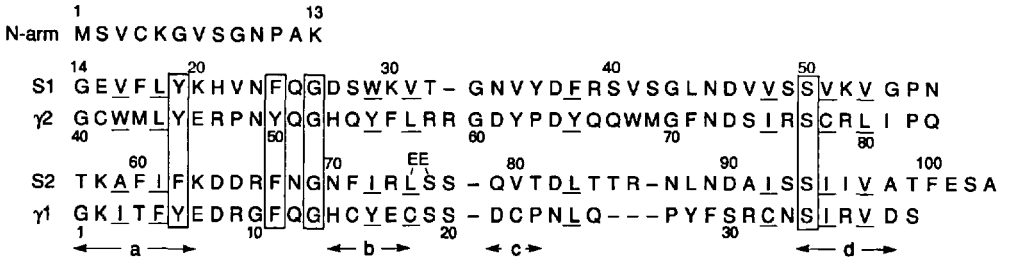

Fig. 2. Comparison of the deduced sequence of spherulin 3a (Bernier et al. 1987) with that of the N-terminal domain of bovine γII-crystallin (Bhat and Spector 1984) (γB). Structurally important residues (Wistow et al. 1983) are boxed and core hydrophobics are underlined. The positions of the β-strands a, b, c, and d of each motif are indicated.

Model Building. The graphics program FRODO (Jones 1978), implemented for the Evans and Sutherland PS300, was used to build a model of spherulin 3a based on the experimentally determined coordinates of bovine γβ (Wistow et al. 1983) (Fig. 3). In the descriptions that follow, named residues (such as Asn56) refer to the spherulin sequence unless otherwise stated.

Results

Because of the high symmetry of γB there were several possible strategies for building the model. When comparing the two motifs of spherulin (s1–s2) with the four γB motifs (γ1–γ4), s1 gave the closest match, in terms of sequence identity and number of residues, with γ2 and γ4, the most similar pair of motifs in the crystallin. Motif s2 gave a poorer match in terms of residue number but had the greatest number of sequence identities with γ1 and γ3. This nonlinear mapping of motifs between the two proteins resembles that which is more strikingly apparent for the comparison of γB and Protein S (Wistow et al. 1985) (see Fig. 1). The model was therefore built by fitting s1 and s2 to γ2 and γ1, respectively. As previously described for the model of Protein S, the pseudo twofold symmetry axis within each γB domain allowed main chain connectivity to be changed simply. Thus, residue 83 of γB, modeled as Asn56 of spherulin, was connected to residue 1 of γB, Thr57 of spherulin. The structure was regularized for bond lengths with residues Val53 and Lys58 anchored. With few exceptions, "conserved" residues were left unmoved, whereas the sidechains of nonidentical residues were manipulated to avoid collisions and poor structural environments. There were some apparently concerted changes. The replacement of hydrophobic cysteine and phenylalanine residues in the γB structure by Glu15 and Lys30 in the spherulin model left the two new sidechains close enough on the surface to form an ion pair, whereas an existing glutamate–histidine interaction was changed to a reversed ion pair between Lys20 and Asp27.

More radical manipulations were needed to accommodate insertions in the structural motifs and to add N- and C-terminal extensions. These parts of the structure are necessarily the most speculative.

Both motifs had different numbers of residues in the strand c–strand d connecting loop when compared with the γB template (Fig. 2). This region is the most variable part of βγ-crystallin sequences, and in γB this region tends to adopt a somewhat disordered helical conformation. For s1, only a single deletion of a γB residue (Arg59) was required. In both γB and spherulin 3a the following residue is glycine, a highly flexible residue that was used to accommodate most of the movement required to fill the gap. Val31 and Phe38, at positions that make important contributions to the hydrophobic core of γB, were anchored and the modified chain regularized.

For spherulin motif s2 it was necessary to insert two residues in the turn between strands b and c of motif γ1, without moving the residues of the c strand. This required movement of the two following residues even though sequence alignment showed them to be conserved (Fig. 2). Residues Leu74 and Leu83, core hydrophobics, were anchored and the chain regularized as before. Modification of the variable c–d loop of γ1 to fit the sequence of s2 also required the insertion of two residues without altering the positions of the important conserved core hydrophobics, Leu83 and Ile92. These residues were anchored and the intervening octapeptide rebuilt de novo as a short stretch of regular α-helix, as this region in γB-crystallin is roughly helical. With limited manipulation this gave an encouragingly good fit. The newly modeled hydrophilic residues Thr85, Arg86, Asn89, and Asp90 were found to be exposed to solvent, with a potential ion pair between Arg86 and Asp90. Asn87 was less well exposed but was in a position to have polar interactions with the carbonyl group of Leu83 or perhaps with the sidechain of Lys63. The hydrophobic Leu88 was conveniently buried in the core.

Finally, the N- and C-terminal extensions of spherulin 3a were added. Secondary structure predictions by the method of Garnier et al. (1978) suggested that these polypeptides have mainly extended structure, with a possible turn region between res-

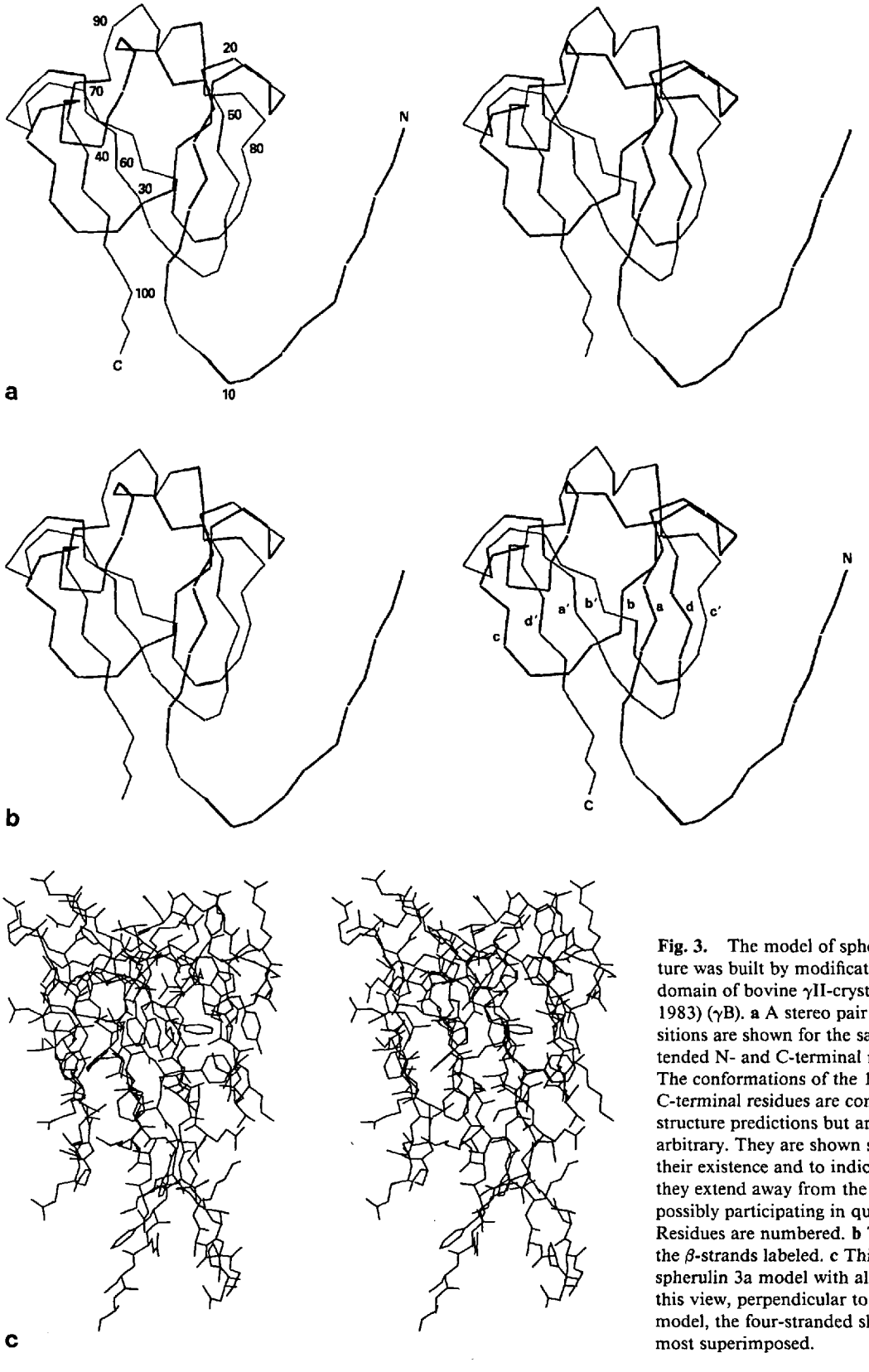

Fig. 3. The model of spherulin 3a. The structure was built by modification of the N-terminal domain of bovine γII-crystallin (Wistow et al. 1983) (γB). **a** A stereo pair in which only Cα positions are shown for the sake of clarity. The extended N- and C-terminal regions are apparent. The conformations of the 13 N-terminal and 4 C-terminal residues are consistent with secondary structure predictions but are otherwise completely arbitrary. They are shown simply to illustrate their existence and to indicate the possibility that they extend away from the core of the structure, possibly participating in quaternary interactions. Residues are numbered. **b** The same view with the β-strands labeled. **c** This stereo pair shows the spherulin 3a model with all residues included. In this view, perpendicular to the β-sheets of the model, the four-stranded sheets can be seen almost superimposed.

idues 6 and 11 at the N-terminus. These structures are reminiscent of the N- and C-terminal arms of the β-crystallins, which are thought to be involved in oligomerization and may perhaps play the same role in spherulins. This means that their structures in an isolated monomer may be moot. Here they were built in for completeness in accord with secondary structure predictions. To prevent steric col-

lisions and to mimic the way polypeptides enter and leave γB domains, the arms were swung away from the globular domain. For the N-arm this was achieved by making use of the flexibility of Gly14. For the short C-arm the added sequence was rotated away at Thr99. These manipulations minimized any disruption to the β-sheet structure of the model.

Discussion

The 103-residue sequence of spherulin 3a has all the features required for the characteristic tertiary fold of the βγ-crystallin superfamily. The model shows that the sequence of spherulin 3a is indeed completely consistent with a tertiary structure like that of a single domain bovine γB-crystallin. As was also observed for Protein S, the residues of the hydrophobic core are generally less bulky than in the crystallins, for example two γB tryptophans are replaced by smaller valines. In this case there is some steric compensation in the substitution of Trp29 for tyrosine (a tryptophan at this position is also seen in motif 3 of all βA1/A3-crystallins, see Wistow and Piatigorsky 1988 for references). However, it is likely that the two β-sheets of spherulin are more closely packed than those of γB. It seems that the high content of aromatic residues in β- and γ-crystallins is characteristic only of the lens protein members of this superfamily. This may perhaps be related to the way in which the crystallins interact with UV and visible light (Wistow et al. 1983).

The model of spherulin 3a shows some interesting surface features. In γB the fourth residue of strand a in motifs 1 and 3, which are exposed to solvent, are hydrophilic, whereas the equivalent residues of motifs 2 and 4 are hydrophobic and contribute to interdomain packing. Both equivalent residues in spherulin are hydrophobic (Phe17 and Phe60). Indeed there is a striking hydrophobic patch (Phe 60, 62, 67, Tyr36, Ile96) exposed on one face of the subunit. This suggests that spherulin is capable of extensive intermolecular interactions, oligomerizing or binding other cell structures. Such interactions may involve the N-terminal arm as has been suggested for the oligomeric β-crystallins (Wistow et al. 1981).

This structure supports both the model of successive duplications in the history of the βγ-crystallin superfamily (by demonstrating that a single domain structure can exist) and the suggestion that proteins with this tertiary fold are widespread in prokaryotes and eukaryotes (Wistow and Piatigorsky 1988). It will be interesting to see whether an even more ancient structure, the single motif, has any existing analogues.

Although convergent evolution in tertiary structure cannot be ruled out, it is likely that proteins with very similar architecture do indeed share origins. A genetic element encoding a structure similar to a single γ-crystallin motif could have arisen at an early stage in protein evolution (Fig. 1). The single motif would probably not have been stable as a monomer, instead forming homodimers. Subsequent duplication at the genetic level would have led to slightly divergent structures and the formation of heterodimers. Genetic fusions would have then given rise to fused heterodimers, single-domain proteins similar to spherulin 3a. This could have occurred several times and with different combinations of motifs. For example, sequence comparisons suggest that two ancestral motifs fused in a different order in the ancestor of Protein S compared with the ancestors of the β- and γ-crystallins (Wistow et al. 1985; Wistow and Piatigorsky 1988). The same kind of nonlinear mapping, although less marked, is also apparent for spherulin 3a. The origins of this superfamily must predate the separation of prokaryotes and eukaryotes and many such proteins (both one-domain and two-domain) may exist in both groups.

Although the β- and γ-crystallins seem to have specialized for the lens environment, they were undoubtedly derived from non-lens ancestors of as yet unknown function. In this regard, it is interesting that both Protein S and spherulin 3a are produced as part of a response to stress and participate in the formation of dormant states, spores, or cysts. It is particularly intriguing to note that specific similarities have been observed in the induction of dormancy in *P. polycephalum* (a eukaryote) and *M. xanthus* (a prokaryote). In both cases, and in contrast to other sporulating organisms, these otherwise unrelated species sporulate or encyst in response to polyols such as mannitol and sorbitol (Chet and Rusch 1969), and, as described here, it seems that in both organisms part of the response to such challenge involves the synthesis of proteins related to γ-crystallins. Although these proteins might simply be inert coat proteins, it is also possible that they have a more general protective role, reminiscent of other stress-induced proteins. This would provide a functional connection with the α-crystallins, which are related to the small heat-shock proteins (Ingolia and Craig 1982). In this case it may be that the members of the βγ-crystallin superfamily have evolved to play a role in cellular responses to some kinds of stress, perhaps including osmotic stress. Indeed some connection to stress or cell proliferation may be a common feature relating many of the crystallins, even including the recruited enzymes (Wistow and Piatigorsky 1987, 1988; de Jong et al. 1988; Wistow et al. 1988). Stress proteins are likely to be thermodynamically stable and may also be

able to pack at high concentration, both useful properties in lens. They may also have genes that are either already transcriptionally active or else that are easy to activate in many cell types. With the evolution of the vertebrate lens, such genes, perhaps together with those of certain "housekeeping" proteins such as enzymes of carbohydrate metabolism, may have required only minor modification to give high expression of stable proteins in the new tissue (Piatigorsky and Wistow 1989).

Acknowledgments. I thank the staff of the Advanced Scientific Computing Laboratory and Dr. Alex Wlodawer of the Department of Crystallography at the Frederick Cancer Research Facility, Frederick, Maryland, for help and access to computing, graphics, and printing facilities. I also thank Rick Lonchiarich of the Molecular Modelling Laboratory, DCRT, NIH, for help with figures and Jane Richardson of Duke University for the drawing of γ-crystallin.

References

Bernier F, Lemieux G, Pallotta D (1987) Gene families encode the major encystment-specific proteins of *Physarum polycephalum* plasmodia. Gene 59:265–277

Bhat SP, Spector A (1984) Complete nucleotide sequence of a cDNA derived from calf lens γ-crystallin mRNA: presence of Alu I-like DNA sequences. DNA 3:287–295

Bloemendal H, Piatigorsky J, Spector A (1989) Recommendations for crystallin nomenclature. Exp Eye Res 48:465–466

Blundell T, Lindley P, Miller L, Moss D, Slingsby C, Tickle I, Turnell B, Wistow G (1981) The molecular structure and stability of the eye lens: x-ray analysis of γ-crystallin II. Nature 289:771–777

Blundell T, Carney D, Gardner S, Hayes F, Howlin B, Hubbard T, Overington J, Singh J (1988) Knowledge-based protein modelling and design. Eur J Biochem 172:513–520

Chet I, Rusch HP (1969) Induction of spherule formation in *Physarum polycephalum* by polyols. J Bacteriol 100:673–678

de Jong WW, Leunissen JAM, Hendriks W, Bloemendal H (1988) Molecular evolution of α-crystallin in quest of a function. In: Piatigorsky J, Shinohara T, Zelenka PS (eds) Molecular biology of the eye: genes, vision and ocular disease. Liss, New York, pp 149–158

Driessen HP, Herbrink P, Bloemendal H, de Jong WW (1980) The β-crystallin Bp chain is internally duplicated and homologous to γ-crystallin. Exp Eye Res 31:243–246

Garnier J, Osguthorpe DJ, Robson B (1978) Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J Mol Biol 120: 97–120

Ingolia TD, Craig EA (1982) Four small *Drosophila* heat shock proteins are related to each other and to mammalian α-crystallin. Proc Natl Acad Sci USA 79:2360–2364

Inouye M, Franceschini T, Inouye M (1983) Structural similarities between the development-specific protein S from a gram-negative bacterium, *Myxococcus xanthus,* and calmodulin. Proc Natl Acad Sci USA 80:6829–6833

Jones TA (1978) A graphics model building and refinement system for macromolecules. J Appl Crystallogr 11:268–272

Kanchisa M (1986) IDEAS user manual. Frederick Cancer Research Facility, Frederick MD

Piatigorsky J, Wistow G (1989) Enzyme-crystallins: gene sharing as evolutionary strategy. Minireview. Cell 57:197–199

Raub TJ, Aldrich HC (1982) Sporangia, spherules and microcysts. In: Aldrich HC, Daniel JW (eds) Cell biology of *Physarum* and *Didymium.* Academic Press, New York, pp 21–75

Wistow G, Piatigorsky J (1987) Recruitment of enzymes as lens structural proteins. Science 236:1554–1556

Wistow G, Piatigorsky J (1988) Lens crystallins: evolution and expression of proteins for a highly specialized tissue. Annu Rev Biochem 57:479–504

Wistow G, Slingsby C, Blundell T, Driessen H, de Jong W, Bloemendal H (1981) Eye-lens proteins: the three-dimensional structure of β-crystallin predicted from monomeric γ-crystallin. FEBS Lett 133:9–16

Wistow G, Turnell B, Summers L, Slingsby C, Moss D, Miller L, Lindley P, Blundell T (1983) X-ray analysis of the eye lens protein γ-II crystallin at 1.9 A resolution. J Mol Biol 170:175–202

Wistow G, Summers L, Blundell T (1985) *Myxococcus xanthus* spore coat protein S may have a similar structure to vertebrate lens βγ-crystallins. Nature 315:771–773

Wistow GJ, Lietman T, Piatigorsky J (1988) The origins of crystallins. In: Piatigorsky J, Shinohara T, Zelenka PS (eds) Molecular biology of the eye: genes, vision and ocular disease. Liss, New York, pp 139–147

Received May 23, 1989/Revised August 28, 1989