Structure, Vol. 9, 115–124, February, 2001, ©2001 Elsevier Science Ltd. All rights reserved. PII S0969-2126(01)00573-1

Crystal Structure of the Calcium-Loaded Spherulin 3a Dimer Sheds Light on the Evolution Of the Eye Lens βγ-Crystallin Domain Fold

Naomi J. Clout,* Michael Kretschmar,† Rainer Jaenicke,† and Christine Slingsby*‡
*Department of Crystallography
Birkbeck College
Malet Street
London WC1E 7HX
United Kingdom
†Institut für Biophysik und Physikalische Biochemie
Universität Regensburg
D-93040 Regensburg
Germany

Summary

Background: The βγ-crystallins belong to a superfamily of two-domain proteins found in vertebrate eye lenses, with distant relatives occurring in microorganisms. It has been considered that an eukaryotic stress protein, spherulin 3a, from the slime mold *Physarum polycephalum* shares a common one-domain ancestor with crystallins, similar to the one-domain 3-D structure determined by NMR.

Results: The X-ray structure of spherulin 3a shows it to be a tight homodimer, which is consistent with ultracentrifugation studies. The (two-motif) domain fold contains a pair of calcium binding sites very similar to those found in a two-domain prokaryotic βγ-crystallin fold family member, Protein S. Domain pairing in the spherulin 3a dimer is two-fold symmetric, but quite different in character from the pseudo-two-fold pairing of domains in βγ-crystallins. There is no evidence that the spherulin 3a single domain can fold independently of its partner domain, a feature that may be related to the absence of a tyrosine corner.

Conclusion: Although it is accepted that the vertebrate two-domain βγ-crystallins evolved from a common one-domain ancestor, the *mycetezoan* single-domain spherulin 3a, with its unique mode of domain pairing, is likely to be an evolutionary offshoot, perhaps from as far back as the one-motif ancestral stage. The spherulin 3a protomer stability appears to be dependent on domain pairing. Spherulin-like domain sequences that are found within bacterial proteins associated with virulence are likely to bind calcium.

Introduction

Proteins from complex organisms that perform specialized functions can evolve from prototypes found in much simpler life forms. For example, the vertebrate eye lens is made largely from proteins belonging to two superfamilies, the α- and βγ-crystallins [1, 2, 3], with distant relatives being found in prokaryotes [4]. α-Crystallin is a member of the ubiquitous small heat shock protein family [5, 6, 7] in which members appear to have the common function of protecting against stress. However, members of the βγ-crystallin fold family are more thinly spread phylogenetically, presumably reflecting a more specialized role. A particular feature of the βγ-crystallin domain fold that makes it an intriguing subject for tracing its evolutionary origins is the presence of internal symmetry [8], reflecting an ancient modular design from an approximate 40 residue "Greek key" motif (Figure 1a). Here, we use X-ray crystallography to define the symmetrical structural features of a βγ-crystallin domain from a primitive organism and compare it with the modern lens βγ-crystallin domain fold in order to address issues of protein design and the evolutionary origin of the vertebrate βγ-crystallins.

The superfamily of lens βγ-crystallins comprises around 13 polypeptide chains that each fold into 4 similar "Greek key" motifs, with 2 successive motifs pairing to form a domain [9]. An alignment of the sequences of the motifs of γB- and βB2-crystallin clearly illustrates that a sequence signature codes for the structural motifs, although the sequence identity is moderate at approximately 35% (Figure 2). Based on their low sequence identity, it was proposed that Protein S of the (prokaryotic) bacterium *Myxococcus xanthus* would have a two-domain structure similar to βγ-crystallins [6]. Solution NMR and crystal structures have shown that the Protein S domains are remarkably similar to the domains of the lens βγ-crystallins [10, 11]. There are two types of motifs (referred to as A and B) in the lens βγ-crystallins, and a motif of each type is found to make up a domain, with a tyrosine corner [12] occurring in the B type motif [4,13] (Figure 1a). However, the order of the two motifs in the Protein S domains is permuted (BA) compared to the lens βγ-crystallins (AB) (Figure 1b).

The βγ-crystallin domains share around 15% sequence identity with spherulin 3a of the eukaryotic slime mold *Physarum polycephalum*. Spherulin 3a is a 102 amino acid polypeptide that contains two copies of the βγ-crystallin sequence fingerprint (Figure 2) and was predicted to have a one-domain βγ-crystallin fold [4]. The sequence is so distantly related that it was not possible to ascertain whether or not one of the motifs had a tyrosine corner. Spherulin 3a has since been shown by solution NMR spectroscopy to form a single γ-crystallin-like domain but with an additional short β strand at the N-terminal extension, contributing to the first β sheet of the Greek key fold [14] (Figure 3a). The stability of spherulin 3a is highly dependent on calcium binding [15, 16], and the NMR structure showed their approximate positions.

A common function of the ancient members of the βγ-crystallin fold family appears to involve high protein concentrations that confer stress resistance. *Physarum polycephalum* is a true plasmodial slime mold belonging

‡To whom correspondence should be addressed (e-mail: c.slingsby@mail.cryst.bbk.ac.uk).

Key words: calcium-binding sites; greek key; pathogenicity islands; *Physarum polycephalum*; stress proteins; tyrosine corner

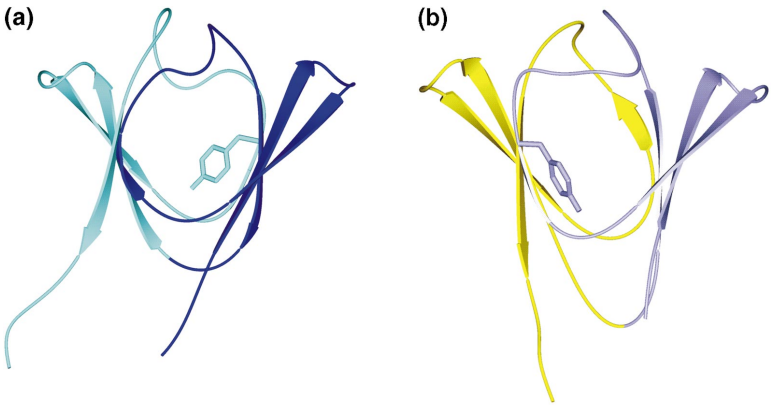

Figure 1. Circular Permutation in the βγ-Crystallin Fold Family Single Domain

(a) The N-terminal domain of bovine γB-crystallin (4GCR), with motif 1 (A type) shown in dark blue and motif 2 (B type) shown in light blue. The view is approximately perpendicular to the pseudo-two-fold axis relating the two motifs. The c strand of one motif contributes to the a, b, and d strands of the partner motif to make a pair of β sheets. The tyrosine corner residue is appended in light blue on the B type motif.

(b) The N-terminal domain of Protein S (1NPS), with motif 1 (B type) shown in dark blue and motif 2 (A type) shown in yellow. The tyrosine corner residue is appended in dark blue on the B type motif. The figure was drawn using the program Setor [42].

to the *mycetozoa*, which are thought to be late emerging, multicellular eukaryotes more closely related to the animal-fungal clade than are green plants [17]. This creature has three lifestyles: microscopic amoeba; gigantic, crawling, multi-nucleate plasmodium; and the fruiting body sporangium. During stress, such as starvation and darkness, the plasmodium can divide into smaller dehydrated spherules, each containing several nuclei that overexpress specific proteins, the most abundant of which is spherulin 3a [18]. These proteins are encased within a hard wall in the presence of calcium, which is in sufficient amounts to occupy the two calcium binding sites on the spherulin 3a domain [15]. Protein S also has a stress connection. Upon starvation, cells differentiate into highly durable myxospores that are resistant to desiccation, heat, and ultraviolet radiation, enabling them to be viable for years. They are protected by a multilayered spore coat consisting of polysaccharides and proteins that includes the calcium binding Protein S. The process of spore coat formation requires calcium [19, 20]. It seems that the ancestral-like βγ-crystallin domains use calcium to stabilize the domain fold.

Other one-domain structures have membership claims on the βγ-crystallin fold family. The killer toxin from the yeast *Williopsis mraki* [21] and a metalloproteinase inhibitor from the prokaryotic *Streptomyces nigrescens* [22] have the same topology as βγ-crystallin domain folds, although they may be examples of convergent evolution [23]. Here, we show by X-ray crystallography that spherulin 3a and Protein S have very similar two-fold symmetric calcium binding sites. As Protein S is a well established member of the βγ-crystallin fold family, this further supports spherulin 3a being a bona fida single-domain member of this fold family. However, spherulin 3a does not exist as a single domain in solution but as a dimer with tight binding, as dissociation cannot be accomplished without full denaturation [15, 16, 24]. For a one-domain crystallin to be the prototype of a βγ-crystallin ancestor, it would be expected to self-associate to form a noncovalent dimer with a similar domain interface as that between the N and C domains in the βγ-crystallins. The spherulin 3a dimer was not revealed by solution NMR spectroscopy. Here, we use X-ray crystallography to define the dimer structure and show that it is unlikely to be the prototype single-domain ancestor of the βγ-crystallin fold family.

Results and Discussion

The Domain Structure

The crystal structure of the spherulin 3a monomer confirms the fold of the monomer found by solution NMR

```
Sp  N-arm   1      SVCKGVSGNPAK

S-type motif
Sp1        13      GEVFLYKHVNFQGDSWKVT--GNVYDFRS-VSG-L-NDVVSSVKVGPN       55
Vc2       215      WSVRFYEHGDYQGRYWTRD-ASG--N-E---SG-F-NDVISSIEILK       253
Vc1       175      NVVRLYADHNYTGHYIDIE-NST--KF---LHG-F-NDTLSSWTIPHG      214
Sp2        56      TKAFIFKDDRFNGNFIRLEESSQVTDLTT—-RN-L-NDAISSIIVATFESA   102
                          $                                  *$ *
B-type motif
Ps1         1    MANITVFYNEDFQGKQVDLP-PGNYTRAQLAALGIE-NNTISSVKVPPG       47
gB2        40      GCWMLYERPNYQGHQYFLR-RGDYPDYQQ-WMG-F-NDSIRSCRLIPQHTGT   87
gB4       129      GSWVLYEMPSYRGRQYLLR-PGEYRRYLD-WGA-M-NAKVGSLRRVM       171
bB2        40      GPWVGYEQANCKGEQFVFE-KGEYPRWDS-WTSSRRTDSLSSLRPIKVDSQE  87
bB4       129      GTWVGYQYPGYRGLQYLLE-KGDYKDSGD-FGA-P-QPQVQSVRRIR       171

A-type motif
Ps2        48      VKAILYQNDGFAGDQIEVV-A-NAEELGP-----L-NNNVSSIRVIS       86
gB1         1      GKITFYEDRGFQGHCYECS-S—DCPNLQP-Y—-F-S-RCNSIRVDS        39
gB3        88      FRMRIYERDDFRGQMSEIT-D-DCPSLQDRFH--L-T-EVHSLNVLE       128
bB1         1      PKIIIFEQENFQGHSHELN-G-PCPNLKE-TG--V-E-KAGSVLVQA        39
bB3        88      HKITLYENPNFTGKKMEVIDD-DVPSFHA-HG—-YQE-KVSSVRVQS       128
```

Figure 2. A Structural Alignment, by Motifs, of Spherulin 3a, the N-terminal Domain of Protein S, and the N- and C-terminal Domains of Bovine γB-Crystallin and βB2-Crystallin, Indicated by the Abbreviations Sp, Ps, gB, and bB, Respectively

The Ps, gB, and bB motifs are clustered together to emphasize their two types of motifs (B and A) and are arranged to emphasize the circular permutation of Ps motifs compared with crystallin motifs. The residue numbers corresponding to the first and last motif residues are indicated. The two spherulin-like motifs found within a larger protein from the pathogenicity island of *Vibrio cholerae* (Vc) were then added to the alignment and clustered with spherulin, as S-type motifs. The two Vc motifs share around 35% sequence identity, with Vc2 sharing around 30% identity with Sp1. The residues shown in bold indicate the motif-conserved glycines, serines, and aromatics that are associated with the stabilization of the Greek key folded hairpin. The residues indicated by "*" and "$" are the side chains and backbone sites, respectively, that are the observed calcium binding ligands in the motifs of spherulin 3a and Protein S. The residues in the tyrosine corners are underlined (in B type motifs), but note text for discussion of Sp1. The sequence of spherulin lacks the N-terminal methionine.

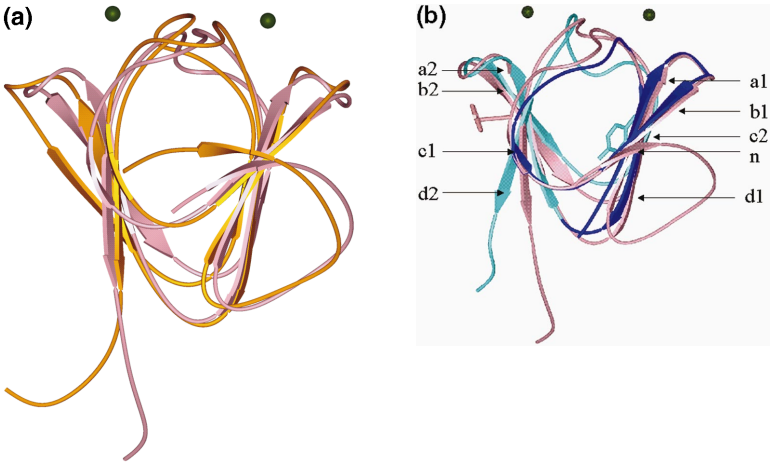

Figure 3. Structural Comparison of Spherulin 3a and γ-Crystallin Single Domains

(a) Superposition of the X-ray structure of the spherulin S3a domain (pink) on the NMR solution structure (yellow), 1AG4.

(b) Superposition of the X-ray structure of the spherulin S3a domain (pink) on the X-ray structure of the N-terminal domain of bovine γB-crystallin color coded as in Figure 1a.

In green are the two calcium ions determined from the X-ray structure. The four β strands of motifs one and two are labeled a, b, c, and d, with the additional strand from the N-terminal arm in S3a labelled n. Note that the appended tyrosine of γB-crystallin is inside the domain, whereas in spherulin 3a, the nearest tyrosine (when the sequence is aligned with the sequences of the βγ-crystallins) is in motif one and is on the surface of the domain.

spectroscopy [14]. There are two spherulin 3a monomers per asymmetric unit that will be referred to as domains A and B. Comparing the two domains reveals very little difference between them: they both have an rmsd of 0.33 Å for the Cα backbone residues. When the spherulin 3a domain A and the average solution NMR spectroscopy structure are superimposed, they have an rmsd of 1.73 Å for the Cα backbone residues (Figure 3a). The rmsd between the X-ray structures of spherulin 3a and the C-terminal domain of γB-crystallin is 1.36 Å (Figure 3b).

The fold of spherulin 3a is that of the βγ-crystallin fold family consisting of eight β strands (a_1, b_1, c_1, d_1, a_2, b_2, c_2, and d_2) which form two Greek key motifs that intercalate to form two β sheets (Figure 3b). There is an N-terminal extension that consists of 12 residues (the protein is expressed in *E. coli* and does not retain the N-terminal methionine; the numbering begins with Ser1, Figure 2) and includes an additional β strand (n), residues 2–4, which extends the first β sheet compared to lens βγ-crystallins (Figure 3b) and Protein S. Although the β-crystallins have N-terminal extensions, these have not been shown to extend the first β sheet. There is a short arch connecting β strands b and c, and a long arch connecting β strands c and d between each sheet present in all of the βγ-crystallin family members. There is a short arch joining the two motifs (d_1 to a_2 and d_3 to a_4) (Figure 3b). Most of the insertions and deletions in the βγ-crystallin superfamily occur in the long arches within motifs that are between β sheets. There are three insertions in the c_1d_1 arch of spherulin 3a compared to γB-crystallin N domain, and one less residue in the c_2d_2 arch of spherulin 3a compared to γB-crystallin N domain (Figure 2).

The folded hairpin turn between strands a and b in each β sheet is a conserved feature of the Greek key motif of the βγ-crystallins. It consists of 8 residues (7–14 in lens βγ-crystallins, and their topological equivalents in the other three motifs): residues 19–26 in motif one and 62–69 in motif two of spherulin 3a (Figure 2). The hydrophobic stabilization is provided by an aromatic residue in the hairpin interacting with an aromatic side chain on the a strand (Tyr6 and Phe1 in motif one of γB-crystallin, and their topological equivalents in the other three motifs). The equivalent residues of spherulin 3a are Tyr18 and Phe23 in motif one and Phe61 and Phe66 in motif two. Residues 9–11 of the lens proteins (and their topological equivalents) form a distorted overlapping β turn. The topologically equivalent residues in spherulin 3a are 21–23 in motif one, and 64–66 in motif two. The turn region is stabilized by hydrogen bonds to the c and d β strands within the sheet, involving the conserved serine residue at position 34 (and topological equivalents in the other three motifs) of the lens βγ-crystallins and the equivalents in spherulin 3a are Ser49 and Ser93 (Figure 2).

Tyrosine corners are conserved structural features of most β sheet proteins, and one is indeed present in all B type motifs of the β- and γ- crystallins, but not in the spherulin 3a structure [12]. When the sequences of γB-crystallin and spherulin 3a are aligned linearly, the residue topologically equivalent to Tyr62 from the corner of γB-crystallin is Val79 in spherulin 3a (Figure 2). However, the sequence of spherulin 3a does have a tyrosine close by when the motifs are permuted (Tyr 35), but the residue does not form a tyrosine corner as it does in the βγ-crystallins but points away from the core of the protein (Figure 3b). The domain fold of the eukaryotic slime mold protein in this respect is more distant from the vertebrate βγ-crystallin proteins than is the prokaryotic protein, Protein S.

Calcium Coordination

The stability of the spherulin 3a structure has been shown to be highly dependent on calcium binding [14, 16, 24]. The NMR solution structure showed that one calcium ion is complexed near the turn comprised by residues Asp36–Asn44, while the second is bound near the loop region between Asp45 and Ser49 and the region between Leu82 and Ala90 [14]. The two calcium binding sites per monomer have now been more precisely identified in the X-ray crystal structure and are shown to be very similar to each other. The two binding sites are located on the surface of the protein at positions related by the internal symmetry of the spherulin 3a domain (Figure 4a). This means that topologically equivalent residues from each motif are responsible for calcium binding (Figure 2) and that the two sites are very similar

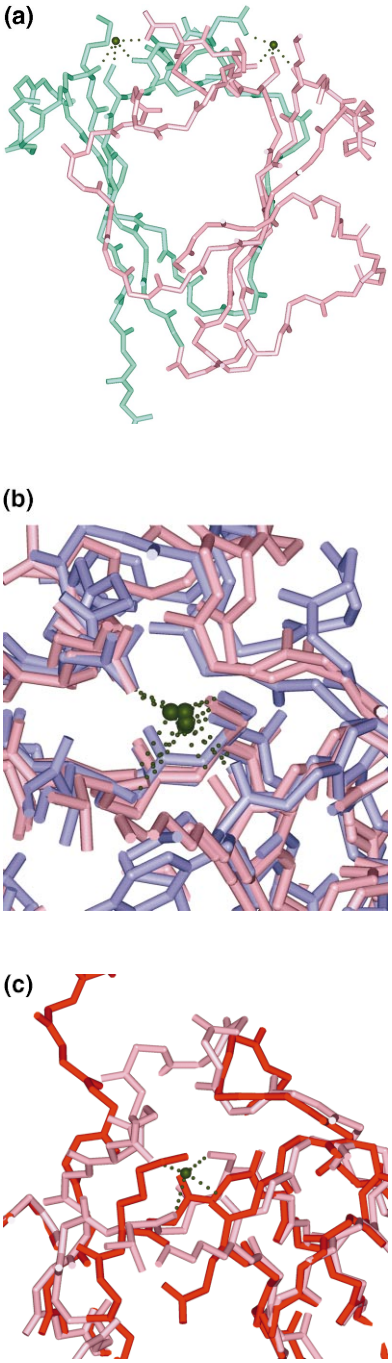

Figure 4. The Calcium Binding Sites in the βγ-Crystallin Fold Family Domain

(a) The polypeptide backbone chain of spherulin 3a domain with motif 1 and n-arm colored in pink and motif 2 in green. Appended are the D/S side chains of the DXXS fingerprint shown to be binding the calcium ligands. The calciums are shown in green. Each site comprises four ligands, two side chains, and two backbone carbonyls. Note that the view is perpendicular to the pseudo-two-fold axis that relates motifs and that each calcium atom is liganded by an aspartate from each motif.

(b) The two calcium binding sites of spherulin 3a (pink) and the two calcium binding sites of the N-terminal domain of Protein S (blue) are all superposed, showing a similar coordination of the calcium ions (green).

(c) Calcium binding site one of spherulin 3a (pink) superposed on an equivalent region of rat βB2-crystallin (red). The calcium binding side chains of spherulin 3a are not conserved in the lens protein, nor is the loop backbone conformation conserved, resulting in the loss of a calcium binding site.

(Figure 4b). Furthermore, each calcium binding site utilizes residues from both motifs (Figure 4a).

Unlike the lens βγ-crystallins, Protein S has two calcium binding sites formed from residues topologically equivalent to those that bind calcium in the spherulin 3a structure (Figure 4b). In both microbial proteins, there are two calcium binding sites per domain and four calcium binding ligands per calcium binding site. From the sequence alignment, it can be seen that calcium binding domains have the sequence fingerprint D/N-X-X-S at topologically equivalent positions in both motifs (Figure 2). The side chains of aspartate/asparagine and serine provide two of the ligands, and the first X residue provides a backbone carbonyl ligand (Figure 4b). However, the side chains that are in the D/N-X-X-S fingerprint do not contribute to the same calcium binding site but rather one from each fingerprint contributes to each calcium binding site (Figure 4a). The fourth calcium binding ligand comes from the backbone carbonyl of Lys19 of spherulin 3a (and the topological equivalent Lys62) that is distant in sequence from the D/N-X-X-S fingerprint. In lens βγ-crystallins, the equivalent residues to those that provide side chains for calcium binding in spherulin 3a are not strictly conserved. Furthermore, the backbone conformation in this region is insufficiently conserved to provide ligands, in keeping with the presence of insertions and deletions in the cd arches (Figure 4c). In the crystal lattices, there are additional calcium ligands from water molecules, but they are variable in their coordination geometry between the various sites in spherulin 3a and Protein S.

The prokaryote and the vertebrate proteins share similar domain folds in terms of their distinctive motif types, and the two microbial organisms share very similar calcium binding sites. The spherulin 3a domain differs from that of the lens βγ-crystallins in both the absence of a tyrosine corner and the presence of two distinct calcium binding sites.

A distantly related nonlens member of the βγ-crystallin fold family (epidermis differentiation-specific protein, EDSP) has been found in the vertebrate amphibian *Cynops*, and it shows all the sequence characteristics to fold into two domains, each comprising a pair of AB motifs [13]. The first two motifs each show the calcium binding fingerprint D-X-X-S, and so the N-terminal domain is likely to have a pair of calcium binding sites. Although the first motif of the second domain contains the D-X-X-S signal, the second motif has D-X-X-(deletion), so it is unlikely that a pair of calcium binding sites will form. Furthermore, the deletion is likely to disrupt the cd arch conformation in this motif, and consequently the backbone ligands to the partner motif will be lost, resulting in no calcium binding sites for this domain.

It is clear that vertebrate genomes have an extensive family of βγ-crystallin domains mainly involved in lens formation. The *cynops* protein indicates that a primitive

vertebrate possesses a relative that has a different function and intriguingly appears to have a feature, calcium binding, found in microbial βγ-crystallin folds. Further evidence that links the vertebrate βγ-crystallin fold to the microbial forms came with a cDNA sequence from the lowest metazoan phylum, the marine sponge *Geodia cydonium*, which contains a sequence finger print that suggests a four-fold Greek key motif repeat. The Greek key motif fingerprint is more conserved for the putative motif pair of the C-terminal domain [25]. However, there is little conservation of the calcium binding fingerprint, with each motif retaining only one of the calcium binding side chains (motif three has the serine, and motif four has the aspartate). It is possible then that one calcium binding site has been preserved. However, the cd arch in motif three is quite short, and so it may not have retained the backbone conformation to form even one of the calcium binding sites. The presence of a calcium binding site in the *Geodia cydonium* Greek key repeating protein would further substantiate the relationship between the microbial (calcium binding) and vertebrate βγ-crystallin family members.

Domain Pairing in the βγ-Crystallin Family

The next level of organization in the βγ-crystallin fold family is the mode of domain pairing. In the lens βγ-crystallins, the motif organization can be described as ABAB, and the N- and C-terminal domains, connected by a short linker, pair about an approximate two-fold axis. In γ-crystallins, N- and C-terminal domains pair intramolecularly, using topologically equivalent residues from their B type motifs (two and four), giving rise to monomers (Figure 5a). However, in βB2-crystallin, N- and C-terminal domains from two polypeptide chains pair intermolecularly, creating a dimeric molecule [26] in a process known as 3-D domain swapping [27]. In Protein S, where motif organization can be described as BABA, the domain interface is intramolecular, as in γ-crystallins, but is formed by residues from motifs two and three, resulting in a different mode of domain packing that is nonsymmetrical (Figure 5b).

The Spherulin 3a Dimer

Two domains of spherulin 3a in the crystal lattice form a dimer (Figure 5c) around a pseudo two-fold axis of rotation of 178.7°. The orientation of the domains with respect to each other is not the same in spherulin 3a as it is in the lens βγ-crystallins (Figures 5a and 5c). There is an extensive hydrophobic patch that is responsible for the domain pairing (Figure 6). The spherulin 3a domain interface is larger than that of γB-crystallin, with a buried surface area of 843 Å for domain A and 835 Å for domain B, while γB-crystallin has a buried surface area of 779 Å for the N-terminal domain and 788 Å for the C-terminal domain. The residues involved in the interface are listed in Table 1.

The domain-interface residues of motif two at positions 43, 56, and 81 are conserved as hydrophobic in all lens βγ-crystallin N-terminal domains, while the topologically equivalent residues of spherulin 3a are Phe59, Arg72 and Ala97 (Figure 2). These residues of spherulin 3a are all involved in its dimer interface, although in addition many extra hydrophobic and aromatic residues are also involved that are not present in the lens βγ-crystallin interfaces. It is likely that the extensive hydrophobic dimer interface contributes significantly to the stability of the protomer fold in spherulin 3a.

A large contribution to the dimer interface is made by the hydrophobic interaction between Tyr35 and Glu100. Interestingly, Tyr35 of spherulin 3a is almost topologi-

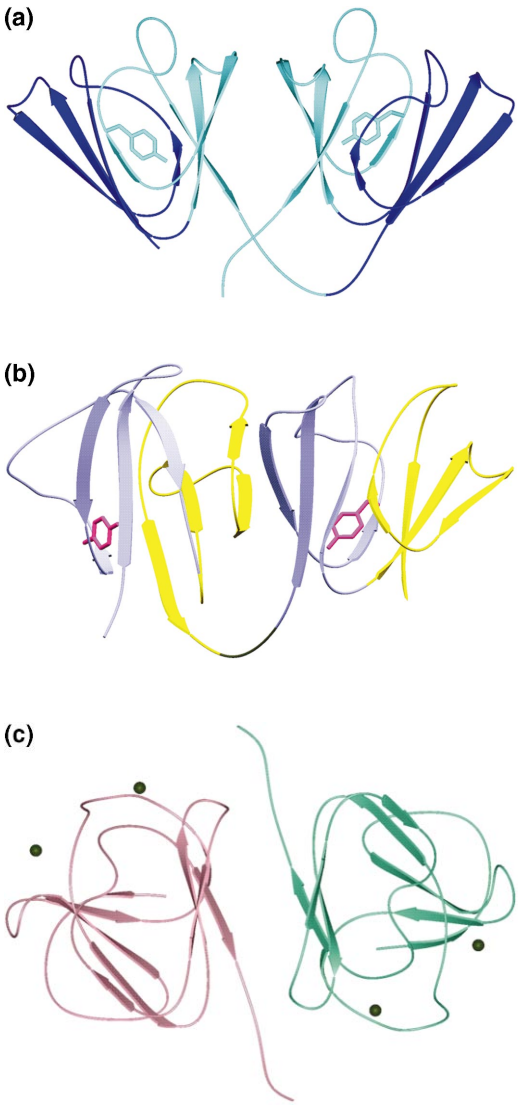

Figure 5. The Different Geometries of βγ-Crystallin Fold Family Domain Pairing

(a) Pseudo-two-fold symmetric domain pairing in bovine γB-crystallin involves topologically equivalent residues from B type motifs (light blue). The view is approximately perpendicular both to the axis relating domains and to the pseudo-two-folds relating motifs within each domain.

(b) Nonsymmetric domain interactions in the NMR solution structure of Protein S (1PRS), showing that residues from motif two (A type, yellow) interact with nontopologically equivalent residues from motif three (B type, blue).

(c) Two-fold symmetric domain pairing between identical domains of spherulin 3a viewed down the dyad axis (which is not in the same plane as the intradomain pseudo-two-folds relating motifs).

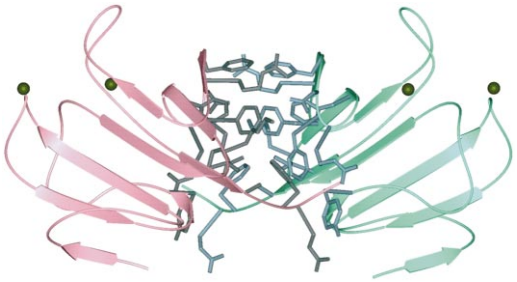

Figure 6. The Extensive Hydrophobic Interface Between Domains of the Spherulin s3a Homodimer

The appended side chains are listed in Table 1.

cally equivalent to Tyr62 of γB-crystallin when the sequences are compared with the motifs permuted (Figure 2). It is this residue that forms the tyrosine corners in B type domains in lens βγ-crystallins and Protein S (Figures 1 and 5). It is interesting that the presence of one tyrosine corner per domain is correlated with the folding of domains that are capable of existing independently [28]. It may be that, in the absence of a tyrosine corner, the stabilization of spherulin 3a is provided through the dimerization of the domains. It would be interesting to engineer the spherulin 3a domain so that a tyrosine is in an equivalent position to that of the B type motifs of a lens βγ-crystallin, or alternatively, to place a tyrosine in a lens βγ-crystallin domain that is in an equivalent position to Tyr35 of spherulin 3a.

The Spherulin 3a Tetramer

The sequence of spherulin 3a contains one cysteine that is located on the N-terminal extension at position 3 in the additional β strand, "n." Solution experiments had shown that in the absence of added reducing agent, higher order aggregates of spherulin 3a can form, although the dimeric state was not sensitive to reducing conditions [24]. This was confirmed when the cysteine was replaced by serine and all higher oligomer formation in the mutant protein was abolished [24]. The crystal lattice now shows that the two cysteines (one per monomer) are responsible for dimer–dimer interactions due to the formation of two disulphide bonds between dimers (Figure 7). The two dimers are related by an approximate two-fold rotation of 165°.

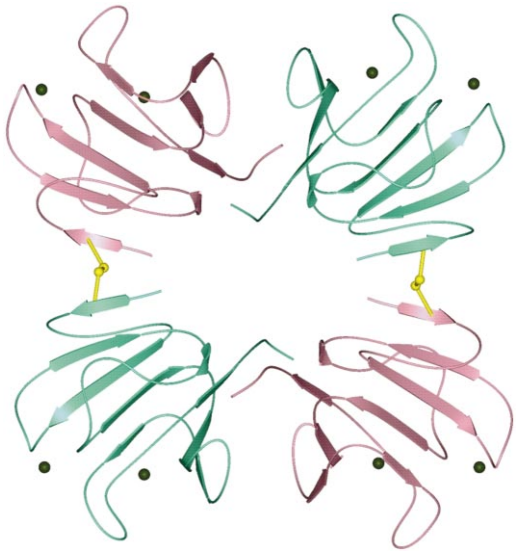

Figure 7. The Spherulin 3a Tetramer Viewed Down the Distorted Noncrystallographic Dyad

The upper and lower pair of domains are the homodimers. The dimer–dimer interaction involves two disulphide bridges (yellow) between the single cysteine, per monomer. Note that the pairs of calcium binding sites on each monomer are all pointing away from each other.

Conclusion

Spherulin 3a, from a primitive eukaryote, is a natural single domain member of the βγ-crystallin fold family. It has been hypothesized that the one-domain spherulin 3a could be the prototype ancestor of the vertebrate two-domain eye lens βγ-crystallins [4, 14]. In lens two-domain βγ-crystallins, the N- and C-terminal domains always associate as pairs about a good, local two-fold

Table 1. Residues Involved in the Spherulin 3a Dimer Interface

Residue from Domain A			Residue from Domain B				
Number	Type	Atom	Number	Type	Atom	Type of Interaction	Distance/Å
35	Tyr	CB	99	Phe	CB	Hydrophobic	3.46
35	Tyr	CZ	100	Glu	CG	Hydrophobic	3.58
57	Lys	NZ	65	Arg	O	Hydrogen Bond	3.17
57	Lys	NZ	67	Asn	OD1	Hydrogen Bond	2.63
59	Phe	CZ	61	Phe	CE2	Hydrophobic	3.40
65	Arg	NE	100	Glu	O	Hydrogen Bond	2.44
66	Phe	CE1	97	Ala	CB	Hydrophobic	3.90
66	Phe	O	72	Arg	NH1	Hydrogen Bond	2.86
68	Gly	O	72	Arg	NH1	Hydrogen Bond	2.94
70	Phe	CD2	72	Arg	NH2	Amino Aromatic	3.36
70	Phe	CE1	70	Phe	CE1	Hydrophobic	3.53
95	Ile	CG2	95	Ile	CG2	Hydrophobic	3.81
95	Ile	CG1	97	Ala	CB	Hydrophobic	4.09

Residues from A and B are listed only once, although the same interactions occur between equivalent residues from B and A due to the approximate two-fold symmetry.

Table 2. Data Collection Statistics for Spherulin 3a

Data Type	Peak	Inflection	Remote
Wavelength/Å	0.97885	0.9791	0.9500
Resolution/Å	2.3	2.3	2.3
No. Observations	54,400	57,284	27,155
No. Unique Reflections	9,236	9,688	10,284
Completeness/%			
Overall	90.9	95.3	91.2
Last shell	72.4	91.6	70.8
R_{sym}/%			
Overall	7.0	5.7	5.1
Last Shell	13.5	12.0	10.4
Ave I/σI			
Overall	25.7	26.4	20.6
Last Shell	14.0	14.8	12.2

$R_{sym} = \Sigma\Sigma|I(h)_i - <I(h)>|/\Sigma<I(h)>$, where $I(h)_i$ is the observed intensity of the $_i$th measurement of a reflection (h), and $<I(h)>$ is the mean intensity of reflection h over the i measurements.

Table 3. The Heavy Atom Sites Found by SOLVE for the Selenomethionine Spherulin 3a Mutant

Site	Occupancy	X	Y	Z	B factor
1	1.1515	0.3338	0.1704	0.0753	15.000
2	1.2886	0.9652	0.5029	0.5362	18.3681

axis, regardless of whether they are two-domain monomers or domain-swapped dimers [26]. Recently, the X-ray structure of a single domain of βB2-crystallin that forms a solution dimer has been solved, and it showed that the same kind of domain pairing can be recreated by a homodimer [29]. This suggests that the single-domain βγ-crystallin ancestor would have been capable of forming homodimers. Spherulin 3a forms a homodimer. Although the similarity between the βγ-crystallin and spherulin 3a protomers is striking, their mode of domain pairing is different. An ancestral molecule like the spherulin 3a homodimer did not give rise to the two-domain β- and γ-crystallin polypeptides.

The generic lens βγ-crystallin domain comprises two Greek key motifs organized about a local two-fold axis, reflecting an assumed early gene duplication of an ancestral single motif, followed by divergence and fusion. There exists a protein fold within prokaryotes that has greater similarity to the lens βγ-crystallin fold than does the eukaryotic spherulin 3a, namely, Protein S of *Myxococcus xanthus*, although it occurs as a two-domain polypeptide. The two Protein S domains have very similar folds and probably very similar calcium coordination sites [30], suggesting that even within prokaryotes there has been a one-domain βγ-crystallin ancestral form. Sequence and structural comparisons of the βγ-crystallin domains show that the two motifs within a domain have diverged into two distinctive types (A and B) [4] and that the motifs in Protein S have been permuted compared to the lens βγ-crystallins [10]. This means that even before prokaryotes and eukaryotes diverged, the motif gene fusion has occurred in two ways (AB and BA). The one-domain eukaryotic spherulin 3a does not show this motif specialization and therefore represents an evolutionary offshoot that has either never acquired or has lost this motif specialization that is intimately connected with the possession of a tyrosine corner (Figure 3). However, spherulin 3a has a pair of very similar calcium binding sites to the pair found in the prokaryotic Protein S, thus strengthening its claim of being a member of the βγ-crystallin fold family (Figure 4). The calcium binding sites have, however, been lost from vertebrate lens βγ-crystallins (Figure 4).

In summary, the lens βγ-crystallins, Protein S, and spherulin 3a most likely have independent histories of genetic duplication and fusion events before any domain pairing had occurred, i.e., while the ancestral proteins were still only one-domain proteins. In the case of Protein S, the motifs are permuted in comparison with higher eukaryotic βγ-crystallins, indicating independent fusion events of the ancestral single motifs rather than a more recent lateral gene transfer. It is possible that the β-crystallins, γ-crystallins, and spherulin 3a did all evolve from a one-domain common ancestor, but that ancestor is not like spherulin 3a.

It is interesting to consider the present day outcomes of these distant evolutionary events. The βγ-crystallin domain that has survived to the vertebrates has undergone a massive expansion. There are 13 crystallin polypeptides corresponding to 26 domains. The nonlens human AIM 1 has 6 domains [31], and another two nonlens domains have turned up in amphibians [13, 25]. Thus, this kind of AB domain has been successful both in the company of similar domains or fused with non-βγ-crystallin sequences. In the lens, the domain assemblies build up in complexity to form the high concentration media required for high refractive index while maintaining the favorable phenotypes of high stability, solubility, and polydispersity. This success appears to have been dependent on specific two-fold domain pairing (but with variability in the connections), a route not followed by the BA prototype, Protein S, where duplication was followed by nonisologous domain interactions (Figure 5). Families of putative calmodulin binding sequences have been found in ciliates harboring modules very distantly related to βγ-crystallin sequences, and these modules have been suggested to play a role in multimerisation [32]. In the case of spherulin 3a, it appears that its own protomer stability is dependent on isologous domain pairing. This would place an additional constraint on the choice of prospective partners, as it must both domain pair and stabilize the protomer fold. There may be a role for the tyrosine corner in the evolution of independently folded β sandwich proteins. Although the tyrosine corner has been proposed as a possible folding nucleus [12, 33], this has not been universally supported by further protein engineering experiments [34].

Interestingly, there are spherulin-like domain sequences embedded within ORFs encoding potential large bacterial proteins associated with virulence. For example, in *Vibrio cholerae* there is a tandem pair of motifs between residues 175 and 253 of a 1111 residue putative inner membrane protein (GenBank database, GI: 3,004,928) found within a pathogenicity island [35]. Like spherulin 3a, they are likely to have calcium binding motifs and no tyrosine corner and so have been called S type motifs (Figure 2). However, both their potential

and their site for dimerization are harder to predict, as the equivalent spherulin interface residues are only moderately conserved. These spherulin-like domains may have potentially three interrelated functions: facilitate dimerization of the larger protein molecule, contribute a calcium binding module to a larger assembly, and confer additional stability to the pathogenic form to enable adaptation to the human intestine. The unifying functional theme so far for the origin of proteins with the βγ-crystallin fold, and for lens crystallins in general, is one that is associated with cellular stress.

Biological Implications

Tracing the structural origin of vertebrate proteins can give clues as to their range of functions. The eye lens crystallins are good examples of proteins that have been reemployed in vertebrates for a new function, whereby their role of providing light refraction over a long lifetime presumably benefits from an earlier function as a stress protein. The primitive βγ-crystallin fold considered here is extremely distant from the lens crystallins and uses symmetric calcium binding sites to increase fold stability in a strikingly similar way to those found in a prokaryotic stress protein. Analysis of the 3-D calcium binding motif allows the definition of a weak sequence fingerprint that can indicate other members of the βγ-crystallin fold family having a calcium binding role and stress function, such as the spherulin-like domain sequence found within a protein associated with the emergence of pandemic strains of cholera.

Comparison of the higher order domain pairing of the vertebrate and microbial βγ-crystallin domains has indicated the importance of having independently folded domains for modular assembly, allowing the massive expansion of the two-domain lens crystallin polypeptides. It is possible that the evolutionary success of the independently folded βγ-crystallin domain has arisen through the acquisition of a structural feature found in many β sandwich proteins, namely, the tyrosine corner. This feature is associated with the diversification of the ancient ancestral motif, although the details are missing due to lack of "fossil evidence." Solving more 3-D structures of these ancient members of the βγ-crystallin fold family will help us understand the early stages of protein evolution, which in turn will inform our own attempts of protein design.

Experimental Procedures

Crystallization

Spherulin 3a was crystallized by vapor diffusion from a modified solution 24 from the Crystal Screen (Hampton Research). Hanging drops of 1 μl protein (~30mg/ml) and 1 μl well solution were suspended at room temperature over a well solution containing 0.2 M calcium chloride (pH 4.6), 0.2 M sodium acetate, and 20% w/v 2-propanol. Tetragonal crystals grew overnight. The space group was $P4_32_12$ or $P4_12_12$; unit cell dimensions were a = b = 42.30 Å, c = 213.85 Å, and α = β = γ = 90; and there were most likely two molecules per asymmetric unit. Native data were collected at room temperature using a rotating anode generator to a resolution of 2.6 Å.

Phasing

All attempts to solve the structure using molecular replacement or preparation of heavy atom derivatives failed. As there is no natural methionine in the sequence, Ile94 was engineered to methionine in order to effect structure solution using MAD phasing. The protein was expressed in the methionine auxotroph strain *E. coli* B84 (DE3) in minimal medium supplemented with selenomethionine [36]. The protein was purified to homogeneit, and the substitution and derivatization was checked by electrospray mass spectrometry. The measured monomer molecular mass of 11,217.5 Da was consistent with the calculated molecular mass of 11,218.4 Da for a 102 residue polypeptide chain in which the initiating N-terminal methionine is removed, an isoleucine residue exchanged by methionine, and a methionine residue derivatized to selenomethionine. The selenomethionine derivative crystallized under the same conditions as native protein, indicating that the mutation had had little effect on the structure of the protein.

Data Collection and Reduction

A selenomethionine spherulin 3a crystal was flash cooled to 100 K in a liquid nitrogen gas stream. Glycerol was added to the crystallization buffer (25% v/v) as a cryoprotectant. MAD data were collected by the reverse beam method at three wavelengths (peak at λ_1 = 0.978850 Å, rise at λ_2 = 0.97905 Å, and remote at λ_3 = 0.95 Å) at station X12C at NSLS at Brookhaven National Laboratory on a CCD area detector (Brandeis B4). Data collection protocols aimed at reducing the effect of systematic error were adopted, collecting all observations contributing to an individual phase determination as close together in time as possible. Data were collected rapidly from a single crystal at 20 s per image. Each data set required $\phi = 40^\circ$ collected with $\Delta\phi = 0.5^\circ$ and a resolution limit of 2.3 Å.

The data were indexed and reduced using the d*TREK suite of programs [37]. The space group was $P4_12_12/P4_32_12$ with cell dimensions a = b = 41.35 Å, c = 213.64 Å, and α = β = γ = 90°. The effective resolution of all the data is 2.2 Å. The data from each wavelength (two sets per wavelength, collected 180° apart in ϕ) were merged using dtreflnmerge of d*TREK. The data processing statistics are shown in Table 2.

Structure Solution

The phases were solved using the package SOLVE [38]. SOLVE was able to locate two selenium sites with an overall Z score for the solution of 69.8 and figure of merit of 0.69; the values of the peaks' heights were 25.9 σ and 26.6 σ for each of the heavy atom sites. The heavy atom sites are shown in Table 3. Density modification was performed using the program DM [39]. Electron density maps viewed using the graphics program O [40] indicated that the space group is $P4_32_12$.

Model Building and Refinement

The map was "skeletonized" using MAPMAN, and the corresponding "bones" and electron density map were read into the graphics and modeling package O [40]. This skeleton was used as a basis for fitting the spherulin 3a sequence into the electron density. It was possible to trace the whole molecule using the original map (based on the calculated phases), apart from four residues in a loop region (85–88) and the last four residues in the sequence (99–102). Once the Cα backbone was distinguished, it was possible to identify most of the side chain atoms. CNS was used to perform the macromolecular structure refinement using all data from 25.0–2.2 Å resolution, and a bulk solvent correction was applied [41]. The refinement consisted of several rounds of simulated annealing (at 1500 K), followed by manual rebuilding in O [40]. Five percent of the data were excluded for calculating R_{free}. Individual B factors were initially set to 20 Å^2 for protein atoms. Water molecules were excluded from this early refinement. Final refinement included the water molecules and anisotropic B factor refinement. The final structure contains 1574 nonhydrogen atoms and 71 water molecules in the asymmetric unit. The final R factor is 25.1%, with an R_{free} factor of 28.9%. All of the residues lie within allowed regions of a Ramachandran plot. The deviation, from ideal, of the bond lengths is 0.007 Å, and the deviation of bond angles is 1.27°. Residues 1–100 were defined by the electron density.

Accession Numbers

The coordinates have been deposited in the PDB with accession code 1hdf.

Acknowledgments

Diffraction data for this study were collected at Brookhaven National Laboratory in the Biology Department single-crystal diffraction facility at beamline X12C in the National Synchrotron Light Source. This facility is supported by the United States Department of Energy Offices of Health and Environmental Research and of Basic Energy Sciences, and by the National Institutes of Health National Center for research sources. The spherulin 3a MAD data were collected, processed, and the phases solved on the Data Col 99 course organized by Robert Sweet. Much gratitude is owed to all those involved in teaching the course, including Robert Sweet and Jim Pflugrath. Many thanks are owed to Jason Yano for his help during the DataCol 99 course. Orval Bateman, David Moss, and Burkhard Rosinke are kindly acknowledged. The financial support of the Medical Research Council is gratefully acknowledged as is that of the Deutsche Forschungsgemeinschaft (grant ja78/33). The work has also been supported by an EU BioMed grant (BMH4-CT98-3895).

Received: September 5, 2000
Revised: November 16, 2000
Accepted: November 28, 2000

References

1. Lubsen, N.H., Aarts, H.J.M., and Schoenmakers, J.G.G. (1988). The evolution of lenticular proteins: the β- and γ-crystallin super gene family. Prog. Biophys. Mol. Biol. *51*, 47–76.
2. Wistow, G., and Piatigorsky, J. (1988). Lens crystallins - the evolution and expression of proteins for a highly specialized tissue. Annu. Rev. Biochem. *57*, 479–504.
3. de Jong, W.W., Lubsen, N.H., and Kraft, H.J. (1994). Molecular evolution of the eye lens. Prog. Ret. Eye Res. *13*, 391–442.
4. Wistow, G. (1990). Evolution of a protein superfamily: relationships between vertebrate lens crystallins and micro-organism dormancy proteins. J. Mol. Biol. *30*, 140–145.
5. Ingolia, T., and Craig, E. (1982). Four small Drosophila heat shock proteins are related to each other and to mammalian alpha-crystallin. Proc. Natl. Acad. Sci. USA *79*, 2360–2364.
6. Wistow, G., Summers, L., and Blundell, T. (1985). *Myxococcus xanthus* spore coat protein S may have a similar structure to vertebrate lens beta-gamma-crystallins. Nature *315*, 771–773.
7. de Jong, W.W., Caspers, G.J., and Leunissen, J.A.M. (1998). Genealogy of the α-crystallin – small heat shock protein superfamily. Int. J. Biol. Macromol. *22*, 151–162.
8. Blundell, T., et al., and Wistow, G. (1981). The molecular structure and stability of the eye lens: X-ray analysis of γ-crystallin II. Nature *289*, 771–777.
9. Slingsby, C., et al., and Bax, B. (1997). X-ray diffraction and structure of crystallins. Prog. Ret. Eye Res. *16*, 3–29.
10. Bagby, S., Harvey, T., Eagle, S., Inouye, S., and Ikura, M. (1994). NMR-derived three-dimensional solution structure of Protein S complexed with calcium. Structure *2*, 107–122.
11. Wenk, M., Baumgartner, R., Holak, T.A., Huber, R., Jaenicke, R., and Mayr, E.M. (1999). The domains of Protein S from *Myxococcus xanthus*: structure, stability and interactions. J. Mol. Biol. *286*, 1533–1545.
12. Hemmingsen, J.M., Gernert, K.M., Richardson, J.S., and Richardson, D.C. (1994). The tyrosine corner: a feature of most Greek key β-barrel proteins. Protein Sci. *3*, 1927–1937.
13. Wistow, G., Jaworski, C., and Rao, P.V. (1995). A non-lens member of the βγ-crystallin superfamily in a vertebrate, the amphibian *Cynops*. Exp. Eye Res. *61*, 637–639.
14. Rosinke, B., Renner, C., Mayr, E., Rainer, J., and Holak, T. (1997). The solution structure of calcium loaded spherulin 3a. J. Mol. Biol. *272*, 1–11.
15. Kretschmar, M., and Jaenicke, R. (1999). Stability of a homodimeric Ca^{2+}-binding member of the βγ-crystallin superfamily: DSC measurements on spherulin 3a from *Physarum polycephalum*. J. Mol. Biol. *291*, 1147–1153.
16. Kretschmar, M., Mayr, E.M., and Jaenicke, R. (1999). Kinetic and thermodynamic stabilization of the βγ-crystallin homolog spherulin 3a from *Physarum polycephalum* by calcium binding. J. Mol. Biol. *289*, 701–705.
17. Baldauf, S.L., and Doolittle, W.F. (1997). Origin and evolution of the slime molds (Mycetozoa). Proc. Natl. Acad. Sci. USA *94*, 12007–12012.
18. Bernier, F., Seligy, V.L., Pallotta, D., and Lemieux, G. (1986). Changes in gene expression during spherulation in *Physarum polycephalum*. Biochem. Cell Biol. *64*, 337–343.
19. Kaiser, D., Manoil, C.C., and Dworkin, M. (1979). Myxobacteria: cell interactions, genetics and development. Annu. Rev. Microbiol. *33*, 595–639.
20. Inouye, S., Inouye, M., McKeever, B., and Sharma, R. (1980). Preliminary crystallographic data for protein S, a development specific protein of *Myxococcus xanthus*. J. Mol. Chem. *255*, 3713–3714.
21. Antūch, W, Guntert, P., and Wuthrich, K. (1996). Ancestral beta-gamma-crystallin precursor structure in a yeast killer toxin. Nat. Struct. Biol. *3*, 662–665.
22. Ohno, A., et al., and Kainosho, M. (1998). NMR structure of the Streptomyces metalloproteinase inhibitor, SMPI, isolated from *Streptomyces nigrescens* TK-23: another example of an ancestral βγ-crystallin precursor structure. J. Mol. Biol. *282*, 421–433.
23. Clout, N.J., Slingsby, C., and Wistow, G.J. (1997). An eye on crystallins. Nat. Struct. Biol. *4*, 685.
24. Kretschmar, M., Mayr, E.M., and Jaenicke, R. (1999). Homodimeric spherulin 3a: a single-domain member of the βγ-crystallin superfamily. Biol. Chem. *380*, 89–94.
25. Krasko, A., Müller, I.M., and Müller, W.E.G. (1997). Evolutionary relationships of the metazoan βγ-crystallins, including that from the marine sponge Geodia cydonium. Proc. R. Soc. Lond. B. Biol. Sci. *264*, 1077–1084.
26. Bax, B., et al., and Slingsby, C. (1990). X-ray analysis of beta-B2-crystallin and evolution of oligomeric lens proteins. Nature *347*, 776–780.
27. Bennett, M.J., Schlunegger, M.P., and Eisenberg, D. (1995). 3D Domain swapping - a mechanism for oligomer assembly. Protein Sci. *4*, 2455–2468.
28. Mayr, E.-M., Jaenicke, R., and Glockshuber, R. (1994). Domain interactions and connecting peptides in lens crystallins. J. Mol. Biol. *235*, 84–88.
29. Clout, N.J., Basak, A., Wieligmann, K., Bateman, O.A., Jaenicke, R., and Slingsby, C. (2000). The N-terminal domain of βB2-crystallin resembles the putative ancestral homodimer. J. Mol. Biol. *304*, 253-257.
30. Wenk, M., and Mayr, E.M. (1998). *Myxococcus Xanthus* spore coat Protein S, a stress induced member of the βγ-crystallin superfamily, gains stability from binding of calcium ions. Eur. J. Biochem. *255*, 604–610.
31. Ray, M.E., Wistow, G., Su, Y.A., Meltzer, P.S., and Trent, J.M. (1997). AIM1, a novel non-lens member of the βγ-crystallin superfamily, is associated with the control of tumorigenicity in human malignant melanoma. Proc. Natl. Acad. Sci. USA. *94*, 3229–3234.
32. Chan, C.W.M., Saimi, Y., and Kung, C. (1999). A new multigene family encoding calcium-dependent calmodulin-binding membrane proteins of *Paramecium tetraurelia*. Gene *231*, 21–32.
33. Bagby, S., Go, S., Inouye, S., Ikura, M., and Chakrabartty, A. (1998). Equilibrium folding intermediates of a Greek-key beta-barrel protein. J. Mol. Biol. *276*, 669–681.
34. Hamill, S.J., Cota, E., Chothia, C., and Clarke, J. (2000). Conservation of folding and stability within a protein family: the tyrosine corner as an evolutionary cul-de-sac. J. Mol. Biol. *295*, 641–649.
35. Karaolis, D.K.R., Johnson, J.A., Bailey, C., Boedeker, E.C., Kaper, J.B., and Reeves, P.R. (1998). A *Vibrio cholerae* pathogenicity island associated with epidemic and pandemic strains. Proc. Natl. Acad. Sci. USA *95*, 3134–3139.
36. Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J., and Huber, R. (1995). High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in *E.coli*. Eur. J. Biochem. *230*, 788–796.
37. Pflugrath, J.W. (1997). Diffraction data processing of electronic detectors: theory and practice. Methods Enzymol. *276*, 286–307.

38. Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta. Crystallogr. D *55*, 849–861.
39. Cowtan, K.D. (1994). DM: an automated procedure for phase improvement by density modification. Joint CCP4 ESF-EACMB Newslett. Protein Crystallogr. *31*, 34–38.
40. Jones, T., and Kjeldgaard, M. (1997). Electron-density map interpretation. Methods Enzymol. *277*, 173–208.
41. Brünger, A.T., et al., and Warren, G.L. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D *54*, 905–921.
42. Evans, S.V. (1993). Setor: hardware lighted three-dimensional solid model representations of macromolecules. J. Mol. Graphics *11*, 134–138.