The expanding small heat-shock protein family, and structure predictions of the conserved .. crystallin domain

Caspers, G.J.M.; Leunissen, J.A.M.; de Jong, W.W.

DOI
10.1007/BF00163229

Publication date
1995

Published in
Journal of molecular evolution

J Mol Evol (1995) 40:238–248

JOURNAL OF MOLECULAR EVOLUTION

The Expanding Small Heat-Shock Protein Family, and Structure Predictions of the Conserved "α-Crystallin Domain"

Gert-Jan Caspers,[1] Jack A.M. Leunissen,[2] Wilfried W. de Jong[1,3]

[1] Department of Biochemistry, University of Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands
[2] CAOS/CAMM Center, University of Nijmegen, Nijmegen, The Netherlands
[3] Institute of Taxonomic Zoology, University of Amsterdam, Amsterdam, The Netherlands

Received: 8 July 1994 / Accepted: 28 September 1994

Abstract. The ever-increasing number of proteins identified as belonging to the family of small heat-shock proteins (shsps) and α-crystallins enables us to reassess the phylogeny of this ubiquitous protein family. While the prokaryotic and fungal representatives are not properly resolved, most of the plant and animal shsps and related proteins are clearly grouped in distinct clades, reflecting a history of repeated gene duplications. The members of the shsp family are characterized by the presence of a conserved homologous "α-crystallin domain," which sometimes is present in duplicate. Predictions are made of secondary structure and solvent accessibility of this domain, which together with hydropathy profiles and intron positions support the presence of two similar hydrophobic β-sheet-rich motifs, connected by a hydrophilic α-helical region. Together with an overview of the newly characterized members of the shsp family, these data help to define this family as being involved as stable structural proteins and as molecular chaperones during normal development and induced under pathological and stressful conditions.

Key words: Molecular chaperones — Molecular phylogeny — Molecular evolution — Structural domains — Structure prediction

Correspondence to: W.W. de Jong

Introduction

The small heat-shock proteins (shsps) are one of the four most common groups of heat-shock proteins (Lindquist and Craig 1988; Hendrick and Hartl 1993; Morimoto et al. 1994). These low-molecular-weight proteins are evolutionarily related to the vertebrate lens protein α-crystallin, as was first noted by Ingolia and Craig (1982). The shsps are a diverse family of proteins of 15–30 kDa which tend to form large aggregates. Most species have multiple shsp genes, although in yeast and chicken only one gene was detected (Susek and Lindquist 1989; Miron et al. 1991). In plants, the shsp genes are most numerous (Vierling 1991). Apart from two multigene families of cytoplasmic shsps (Raschke et al. 1988), plants have also nucleus-encoded shsps that localize to chloroplasts (Chen and Vierling 1991). Recently, a fourth group of plant shsps was found which are localized in the endomembrane system, most likely in the endoplasmic reticulum (Helm et al. 1993).

α-Crystallin is an abundant eye lens protein in vertebrates. It is usually found as large aggregates, consisting of two types of subunits, αA and αB. (For reviews see Wistow and Piatigorsky 1988; Groenen et al. 1994.) It has a structural function in the lens, warranting proper refractive properties and transparency (Tardieu and Delaye 1988). αB- and to a lesser extent αA-crystallin have been shown to occur also in various tissues outside the lens (Bhat and Nagineni 1989; Kato et al. 1991). Both subunits are encoded by single-copy genes in humans. Functionally, the shsps and α-crystallins share the prop-

erty of being molecular chaperones (Horwitz 1992; Jakob et al. 1993), and both are able to convey thermotolerance (Landry et al. 1989; Klemenz et al. 1991).

Apart from the shsps and α-crystallins, this family also includes more disparate members, like certain surface antigens in parasitic eukaryotes and bacteria (Nene et al. 1986; Nerland et al. 1988; Verbon et al. 1992). All members of the family are characterized by the presence of a homologous sequence of about 80 residues, which has been dubbed the "α-crystallin domain" and probably forms a distinct structural and functional unit. This domain is preceded by an N-terminal region of variable length, which shows little or no similarity between the various branches of the family. A short and variable sequence, but containing a conserved motif, extends C-terminally from the "α-crystallin domain" (Wistow 1985; de Jong et al. 1988). Unfortunately, no direct information about the three-dimensional structure of any member of the shsp family is available, nor is there any deeper insight into their structural roles and functional activities.

The α-crystallin/small hsp family is not related to any other protein family, although Lee et al. (1993) noticed some minor sequence similarities between the "α-crystallin domain" and two separate regions in the hsp70 family. This might reflect the sharing of some structural features, possibly pertaining to the shared chaperone functioning, rather than an evolutionary relationship.

Recently, a considerable number of sequences of diverse new members of the shsp family have been reported. A comparison of the sequences and the structural and functional properties of all members of the family should contribute to a more detailed view of the divergent evolution of this family. The objective of the present paper is to give an update of the phylogenies presented earlier (Plesofsky-Vig et al. 1992; de Jong et al. 1993), as well as to analyze the structural and functional variation and similarity of the proteins. The latter information might be important for the future unravelling of the structure–function of especially the conserved "α-crystallin domain."

New Members of the Family

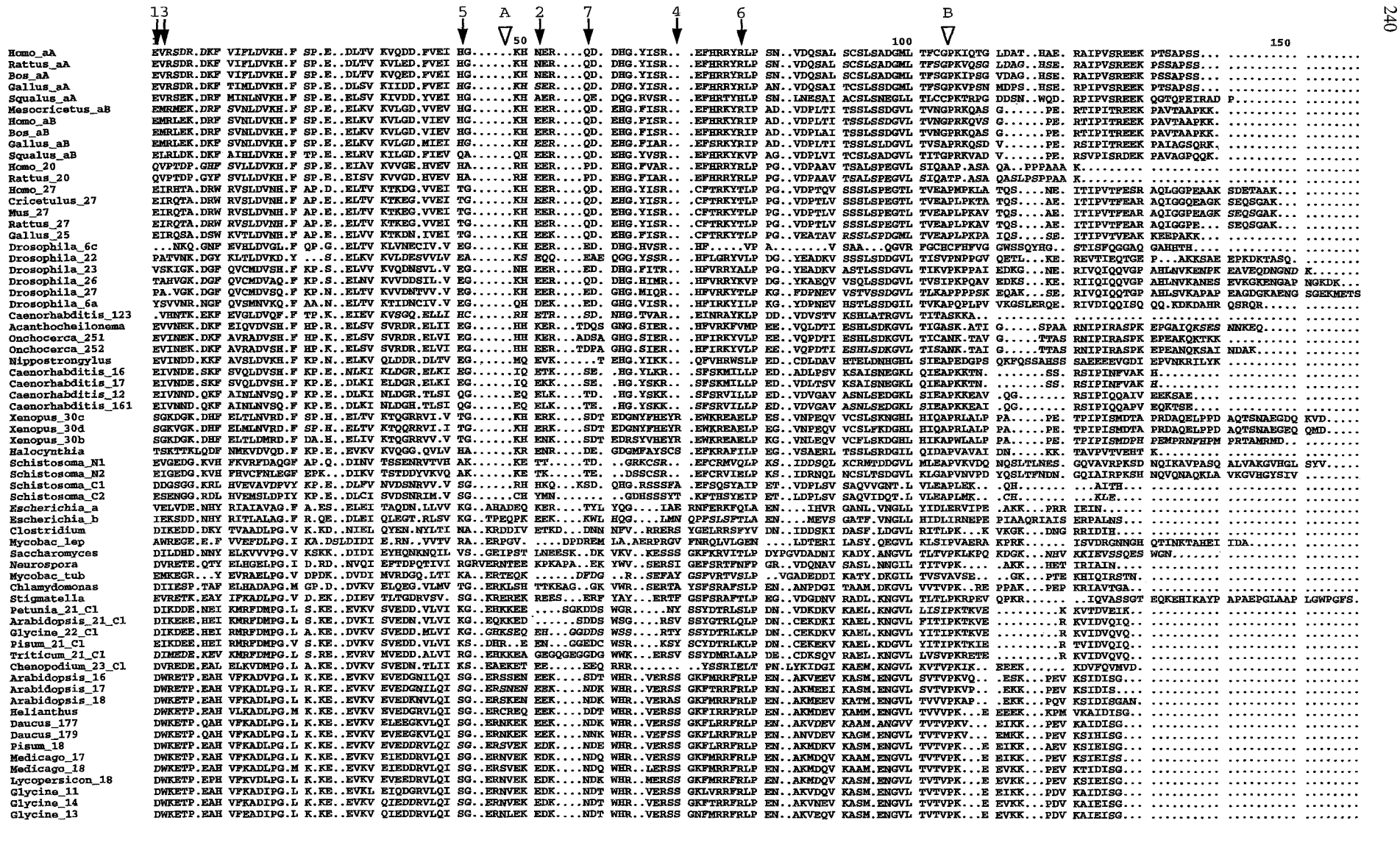
A description of the newly characterized members of the shsp family further highlights the variety of features presented by this family, as revealed already earlier (de Jong et al. 1993). The 16-kDa *Escherichia coli* IbpA and IbpB proteins, which are 52.2% identical, were found to be induced in response to expression of several heterologous proteins (Allen et al. 1992). They are tightly associated with inclusion bodies formed during the expression of these foreign proteins. The presence of a high level of certain unfolded heterologous proteins may in fact be responsible for induction of IbpA and IbpB. Also, induction occurred in wild-type *Escherichia* after heat shock. A putative heat-shock promoter is indeed located upstream from the genes, which are separated by 110 bp on the chromosome.

A *Clostridium acetobutylicum* hsp18 was shown to be induced at the mRNA level by heat shock and the onset of solventogenesis, a metabolical shift in which excreted butyric acid and residual sugar are converted to acetone and butanol shortly before entry into the stationary phase (Sauer and Dürre 1993). The shift to solventogenesis seems to be connected with the heat-shock response, but the underlying molecular mechanisms are still largely unknown. In the myxobacterium *Stigmatella aurantiaca* a 21-kDa protein, SP21, is synthesized during heat shock, fruiting body formation, and stress induced by oxygen limitation (Heidelbach et al. 1993a,b). The protein sedimented with the membrane fraction. As there were no indications of a direct interaction of SP21 with the membranes, this suggests that SP21 forms aggregates or is organized in larger complexes.

In the eukaryotic realm, several new plant sequences were added to the set, including shsps of *Arabidopsis thaliana* (Bartling et al. 1992), maize *Zea mays* (Jorgensen and Nguyen 1994), wheat *Triticum aestivum* (Weng et al. 1991) and rice *Oryza sativa* (Tseng et al. 1992; R. Nishi et al., data base acc. nr. P31673). Most of these sequences were determined at the cDNA level. The gene of the *Pharbitis nil* 17.1-kDa shsp was induced by light treatment and by heat shock, whereas its 18.8-kDa gene was induced only by heat shock. Both are encoded by one open reading frame. In the noncoding region of both genes several heat-shock elements were found (Krishna et al. 1992). The *Chenopodium rubrum* 18.3-kDa shsp belongs to the cytosolic shsp subfamily (class I). A structure with "positive DNA-binding regulatory properties" was predicted that would form a helix-turn-helix region (Knack et al. 1992).

Maximal accumulation of the *Helianthus annuus* shsp mRNA is detected in dry seeds and during embryo mid-maturation stage in the absence of exogenous stress (Almoguera and Jordano 1992). In seedlings, mRNA accumulation to lower levels is found in response to osmotic stress and abscisic acid treatments. The protein has a predicted weight of 17.6 kDa. A 22-kDa shsp from the soybean *Glycine max* was found to be endomembrane localized. It possesses an amino-terminal signal peptide and a carboxyl-terminal sequence characteristic of an endoplasmic reticulum retention signal (Helm et al. 1993). The expression of meiotic prophase repeat proteins in *Lilium* sp. is developmentally induced in absence of stress. A cDNA sequence for one of these proteins was determined, and demonstrated it to be a shsp (Bouchard 1990).

Among the invertebrates, an additional *Schistosoma mansoni* egg antigen sequence, p40-2 (Cao et al. 1993), has 57% amino acid identity to the previously reported p40 antigen (Nene et al. 1986), and likewise contains a duplicate "α-crystallin domain." These stage-specific


13
5
A
2
7
4
6
B
50
100
150
Homo_aA
Rattus_aA
Bos_aA
Gallus_aA
Squalus_aA
Mesocricetus_aB
Homo_aB
Bos_aB
Gallus_aB
Squalus_aB
Homo_20
Rattus_20
Homo_27
Cricetulus_27
Mus_27
Rattus_27
Gallus_25
Drosophila_6c
Drosophila_22
Drosophila_23
Drosophila_26
Drosophila_27
Drosophila_6a
Caenorhabditis_123
Acanthochoilonema
Onchocerca_251
Onchocerca_252
Nippostrongylus
Caenorhabditis_16
Caenorhabditis_17
Caenorhabditis_12
Caenorhabditis_161
Xenopus_30c
Xenopus_30d
Xenopus_30b
Halocynthia
Schistosoma_N1
Schistosoma_N2
Schistosoma_C1
Schistosoma_C2
Escherichia_a
Escherichia_b
Clostridium
Mycobac_lep
Saccharomyces
Neurospora
Mycobac_tub
Chlamydomonas
Stigmatella
Petunia_21_C1
Arabidopsis_21_C1
Glycine_22_C1
Pisum_21_C1
Triticum_21_C1
Chenopodium_23_C1
Arabidopsis_16
Arabidopsis_17
Arabidopsis_18
Helianthus
Daucus_177
Daucus_179
Pisum_18
Medicago_17
Medicago_18
Lycopersicon_18
Glycine_11
Glycine_14
Glycine_13

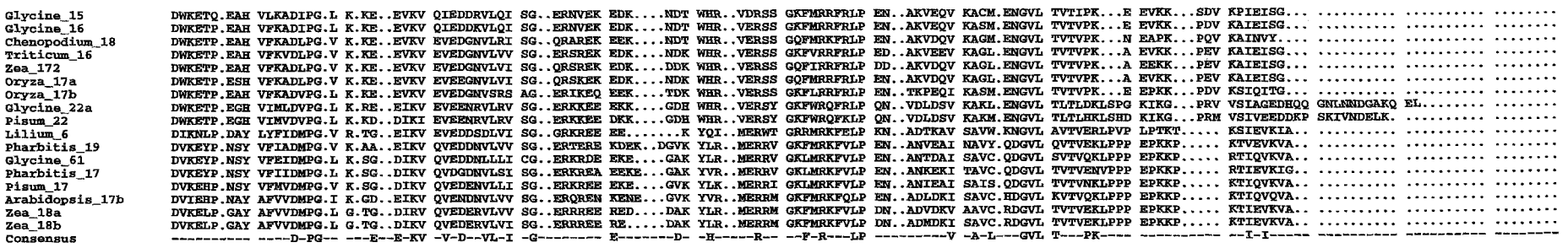


Fig. 1. Alignment of the homologous C-terminal regions of the shsps and related proteins. Sequences are designated by *genus name* and, for the small hsps, the *numbers* as used in the databases. These numbers can be different from those used in the literature. *aA* and *aB* = αA- and αB-crystallin sequences, respectively. *Schistosoma-N1, -N2, -C1,* and *-C2* = N- and C-terminal copies, respectively, of the duplicated α-crystallin domains in the egg antigen p40 and p40-2. *Mycobac lep* and *Mycobac tub* = *Mycobacterium leprae* and *M. tuberculosis*. *Cl* = chloroplast-localized hsps. The positions corresponding to the introns in the genes for *Caenorhabditis elegans* hsps (*1, 2*), *Halocynthia roretzi* HR-29 (*3, 4*), mammalian hsp27 (*5, 6*), and αA- and αB-crystallins (*1, 7*) are indicated with *arrows* above the alignment. All introns fall between codons (phase 0), apart from #4, which is between the second and third nucleotides in the codon for residue 70 (phase 2). "*Consensus*" indicates residues that are present in 50% or more of the entries at a given position. To the alignment of de Jong et al. (1993) the following sequences are added: *Gallus gallus* αA-crystallin (database accession number P02504, residues 63–173), *G. gallus* αB-crystallin (Q05713, 66–174), *Homo sapiens* p20 (Kato et al. 1994, 66–160), *Rattus norvegicus* p20 (Kato et al. 1994, 66–162), *R. norvegicus* hsp27 (R.R. Gilmont and M.J. Welsh, M86389, 91–206), *Nippostrongylus brasiliensis* hsp20 (S33416, 57–172), *Cae norhabditis elegans* hsp12.3 (L15188, translation of residues 36,675–36,752, 37,383–37,496, 37,904–38,044), *Acanthocheilonema viteae* AV25 (S29691, 62–179), *Onchocerca volvulus* OV25-1 (S29692, 61–173), *O. volvulus* OV25-2 (S29693, 49–165), *Halocynthia roretzi* HR-29 (JX0258, 147–252), *Schistosoma mansoni* p40-2 (M96866, 128–253 and 254–342), *Escherichia coli* IbpA (A45245, 39–137), *E. coli* IbpB (B45245, 37–142), *Clostridium acetobutylicum* hsp18 (S25534, 48–151), *Stigmatella aurantiaca* SP21 (M94510, 57–188), *Arabidopsis thaliana* hsp17.6 (P29830, 49–155), *Zea mays* hsp17.2 (X65725, 48–152), *Triticum aestivum* hsp26.6 (Q00445, 134–236), *Oryza sativa* hsp17a (P27777, 46–150), *O. sativa* hsp17b (P31673, 50–154), *Pharbitis nil* hsp17 (Q01544, 48–155), *P. nil* hsp19 (Q01545, 59–167), *Chenopodium rubrum* hsp18 (Q05832, 58–161), *Helianthus annuus* hsp18 (P30693, 48–153), *Glycine max* hsp22 (P30236, 70–192), *Lilium* sp. EMPR6 (A61054, 43–146). The *Homo sapiens* αA-crystallin sequence underwent a minor correction at positions 114–118 of the alignment according to Takemoto and Emmons (1991) and Caspers et al. (1994). The *Mus* hsp27 has been corrected as described by Gaestel et al. (1993) and Merck et al. (1993b). To avoid undue lengthening of the alignment, two unique sequences are left out (marked by *open arrowheads*): a 64-residue insert after position 47 in *Neurospora* hsp30 (*A*), and a 15-residue insert after position 104 in *Lilium* EMPR6 (*B*).

antigens are soluble calcium-binding proteins (Moser et al. 1992). Expression of the gene for the shsp of the gastrointestinal nematode *Nippostrongylus brasiliensis* is developmentally regulated, but not upregulated by heat shock or other stress conditions (Tweedie et al. 1993). Another new nematode shsp sequence was found on chromosome III of *Caenorhabditis elegans* (Wilson et al. 1994). Also two sequences of *Onchocerca volvulus* and one of *Acanthocheilonema viteae* are now available (W. Hoefle, acc. nrs. S29691–3).

A new shoot of the family is represented by the 29-kDa protein HR-29 of the ascidian *Halocynthia roretzi* (Takagi et al. 1993). The protein is abundant in body wall muscle and localizes close to the plasma membrane. In the physiological condition it forms oligomers, observed by electron microscopy as globular aggregates with a diameter of 16.6 nm, as is also the approximate size for α-crystallin and shsps.

Finally, an additional shsp was recently reported in mammals. Kato et al. (1994) found high levels of a 20-kDa protein, p20, in rat soleus muscle, heart, and diaphragm. These tissues also contain high levels of αB-crystallin and hsp27. The p20 protein is present at lower levels in other tissues. It occurs both as high-molecular-weight aggregates and in dissociated forms. Upon heating at 45°C of rat diaphragm in vitro, p20 was redistributed from the cytoplasm to the insoluble fraction, and dissociation of the aggregated p20 to the small form was enhanced. The primary structure of rat p20 and its human homologue were determined at the amino acid level.

Sequence Comparisons and Gene Structure

To ensure that all available shsps and related sequences were included in the present paper, a thorough database search was performed. To that end a previous alignment of 57 sequences of "α-crystallin domains" (de Jong et al. 1993) was used to create a profile (Gribskov et al. 1990) with which the NBRF, SwissProt, and EMBL databases were searched to identify new sequences with similarities to the shsps. These were included in the expanded alignment. As for the extensive set of available α-crystallin sequences, only those for chicken αA- and αB-crystallin (de Jong et al. 1984; Sawada et al. 1992) were added to the previous alignment. Initial alignments of the expanded set of 85 entries were made with a multiple alignment program, followed by manual improvements, as described earlier (de Jong et al. 1993). Because a clear similarity among all members of the family is restricted to the C-terminal parts of the proteins, only this region could be aligned satisfactorily, as shown in Fig. 1. In the expanded shsp family more gaps needed to be introduced in the alignment of this conserved region. The region presented in Fig. 1 corresponds with the putative C-terminal structural domain and extending tail of the α-crystallins and shsps as proposed by Wistow (1985). A considerable number of residues is highly conserved throughout the family (indicated as "consensus" on the bottom line in Fig. 1), and the conservative nature of many other positions is conspicuous. The demarcation between domain and tail is around position 110.

The positions of the conserved "α-crystallin domain" in the various typical representatives of the shsp family are schematically indicated in Fig. 2. Both *Schistosoma* egg antigen sequences contain duplicate C-terminal domains. The C-terminal domain of the *Neurospora* shsp is interrupted by an insertion of 64 amino acids (after position 47 in the alignment). This figure also reveals the length variation in the family. The length of the N-terminal regions, preceding the α-crystallin-like domains, varies from only 25 residues in the *Caenorhabditis* 123 gene that was found in the course of the chromosome III sequencing project (Wilson et al. 1994) to 148 residues in the *Halocynthia roretzi* 29-kDa body-wall protein. Although no sequence similarity can be detected between the N-terminal regions of all members of the family, there are still some minor similarities between the N-termini of the α-crystallins, higher vertebrate shsps, and *Drosophila* shsps, as well as between plant class I shsps and vertebrate shsps (de Jong et al. 1988). It must be concluded that the rate of evolution has been higher in the N-termini than in the C-terminal regions.

In plants, yeast, and most invertebrates, the shsp genes are encoded by intronless genes. Four shsp genes of *Caenorhabditis* contain only one intron, coinciding precisely with the first intron of the α-crystallin genes (Figs. 1 and 2). The aforementioned *Caenorhabditis* 123 gene contains two introns, of which the second coincides approximately with the second intron of the α-crystallin genes. The two introns of mammalian hsp27 are located at different positions. The *Halocynthia* body-wall protein gene contains three introns, of which the second is located three nucleotides downstream from intron #1 in *Caenorhabditis* shsps and in α-crystallins, thus directly preceding the C-terminal domain. The other two introns are located at different positions than those in other genes of the family and are phase 2, rather than phase 0 as are all other ones (Takagi et al. 1993). Although introns #1 and 3 in Fig. 2 demarcate the boundary between the putative N- and C-terminal domains, there is little evidence for the assumption that exon shuffling has played a role in the evolution of this protein family (Pathy 1994).

Phylogeny

Residues 1–104 of the alignment in Fig. 1, encompassing the "α-crystallin domain," were used to construct a phylogenetic tree, using the neighbor-joining program (Saitou and Nei 1987) from the PHYLIP package (Felsenstein 1993). The "α-crystallin domain" was considered

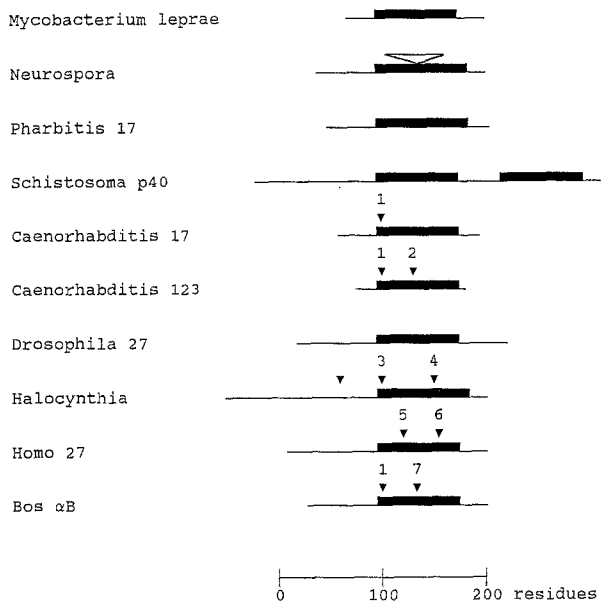


Fig. 2. Location of the homologous domains and intron positions in representative members of the shsp family. The homologous C-terminal "α-crystallin domains" are indicated by *solid bars.* N-terminal domains and C-terminal extensions are indicated by *lines,* corresponding in length to the number of residues. Intron positions in the corresponding genes are marked by *arrowheads.* Whether introns are present in the *Schistosoma* p40 genes is not known. Intron positions are more precisely indicated in Fig. 1 and its legends. The position of the insertion in *Neurospora* hsp30 is indicated by the *open triangle.* Names of the sequences are as in Fig. 1.

not to extend beyond residue 104, because after this position a 15-residue insert in the *Lilium* sequence breaks up the alignment. Since the structures of all members of the family are not known at the DNA level, the amino acid sequences rather than the nucleotide sequences were used in tree construction. The resulting tree is shown in Fig. 3, where bootstrap values of 75% and higher are indicated. Because of the large number of sequences, the tree is depicted in three parts, comprising prokaryotes and lower eukaryotes (Fig. 3a), and the higher plant and animal subtrees (Fig. 3b and c, respectively). The sequences at the deepest branches of the tree, the prokaryote and fungal proteins, are very divergent and not reliably resolved (Fig. 3a). Only the two *Escherichia* Ibp proteins group significantly together. The *Mycobacterium* antigens do not form a sister group, and neither do the *Neurospora* and *Saccharomyces* shsps. However, these proteins need not be orthologues; it is likely that only a small subset of the existing shsp variation has been discovered yet. The present findings indicate that reconstruction of the earliest divergent evolution of the shsps will be problematic.

The angiosperm sequences group together, although at a low bootstrap value (Fig. 3b). A first divergence occurs between the chloroplast sequences and the cytoplasmic shsps. The latter divide highly significantly into the class I and class II sequences (Vierling 1991), which must have emerged before the divergence of monocots and dicots. Repeated duplications have occurred, especially of class I genes. The *Lilium* meiotic prophase repeat protein appears to be most closely related to the class II proteins. The endomembrane-localized *Pisum* and *Glycine* shsps represent an early offshoot of the class I proteins, as was also noted by Helm et al. (1993).

A monophyletic origin of the animal sequences is hardly supported (Fig. 3c). The duplicated N- and C-domains within the two *Schistosoma* egg antigens are very divergent. The two antigens are apparently the result of a more recent gene duplication, long after the emergence of the first antigen with two "α-crystallin domains."

All nematode sequences group together, as do all *Drosophila* sequences, albeit poorly supported. It was noted earlier that the *Caenorhabditis* shsp genes duplicated at an early stage, followed by a more recent, further duplication (Candido et al. 1989). The new *Caenorhabditis* sequence (Wilson et al. 1994) appears to represent a different lineage, as it is more closely related to the *Onchocerca* and *Acanthocheilonema* sequences. It was tabulated as most similar to αB-crystallin by Wilson et al. (1994), but the present phylogenetic analysis reveals no special relationship. The two *Onchocerca* shsps appear to have arisen from a very recent gene duplication.

As was noted before, the *Xenopus* shsps are not closely related to other vertebrate shsps, and may in fact be paralogues of these (de Jong et al. 1988). Interestingly, the body-wall protein of the ascidian *Halocynthia* appears to be most closely related to the *Xenopus* shsps. The other vertebrate shsps form a clade with the αA- and αB-crystallins. In this clade, the recently discovered rat and human p20 proteins form a sister group to the monophyletic αA- and αB-crystallins. They are not closer to αB-crystallins, as Kato et al. (1994) assumed. Three gene duplications thus have led to the expression of four paralogous genes in higher vertebrates—hsp27, p20, and αA- and αB-crystallin. These duplications preceded the earliest vertebrate radiation, since αA- and αB-crystallins were already present before the dogfish *Squalus* diverged from the other vertebrates. Additional duplications must have occurred in the hsp27 lineage. The human hsp27 gene family consists of four members (Hickey et al. 1986), although only a single sequence has been reported. Cooper and Uoshima (1994) found multiple hsp27 transcripts in murine osteoblasts, which may be encoded by separate genes. Also a mouse hsp27 pseudogene has been characterized (Fröhli et al. 1993).

Secondary Structure and Surface Residues of the "α-Crystallin Domain"

Insight into the properties of the shsp family is seriously hampered by the lack of structural information. Using advanced prediction methods on the accumulated se-

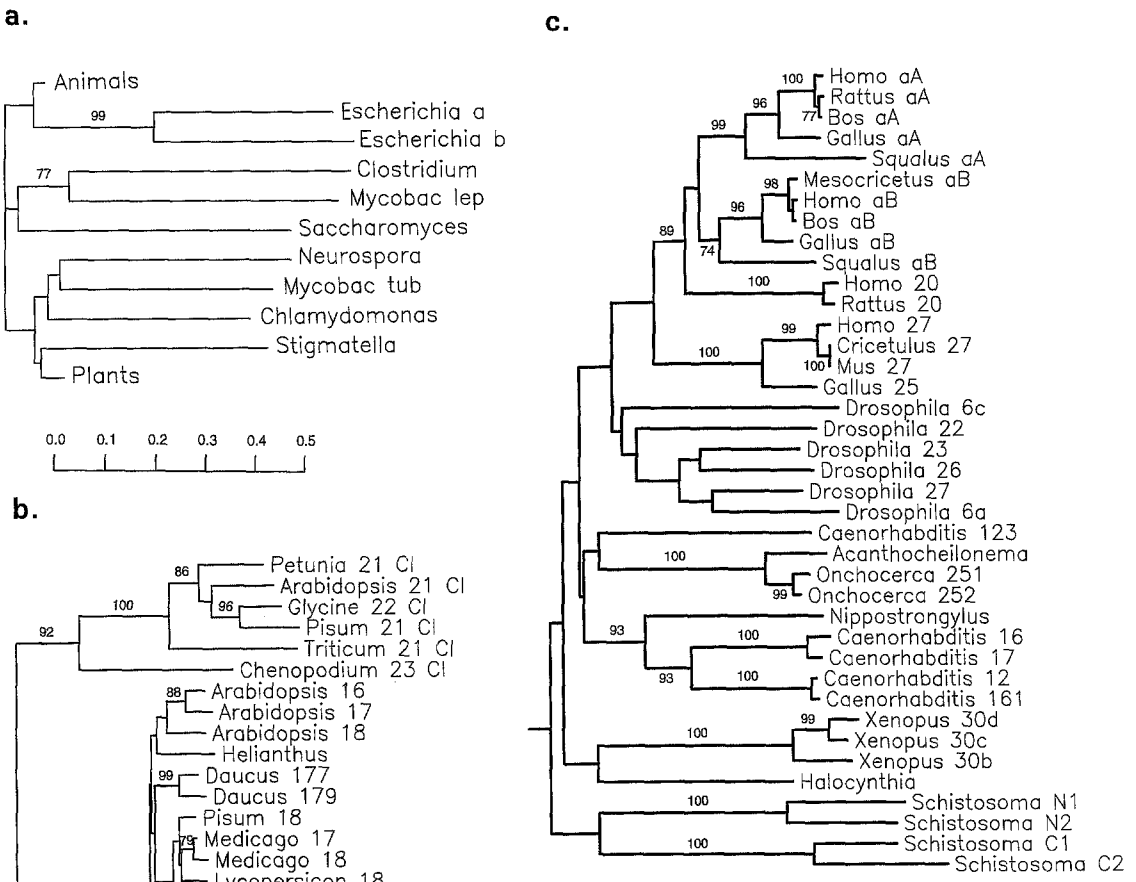


Fig. 3. A phylogenetic tree based on the "α-crystallin domains" of the shsp family (residues 1–104 in Fig. 1). The tree was constructed according to the neighbor-joining method, using the program NEIGHBOR from the PHYLIP package (Felsenstein 1993). Input distance matrices were calculated according to Fitch and Margoliash (1967), excluding gaps from the calculations (program HOMOLOGIES, J.A.M. Leunissen, unpublished). Correction for multiple substitutions was made according to Jukes and Cantor (1969). Relative bootstrap values (program SEQBOOT from the PHYLIP package) of 75% and higher (from 1000 replicates) are indicated. Branch lengths are proportional to the minimum number of mutations per residue (see scale). The tree is depicted in three parts: **(a)** prokaryotes, fungi and green alga, **(b)** higher plants, **(c)** animals. Within plants, *Arabidopsis* 16 to *Oryza* 17b belong to the class I shsps, and *Pharbitis* 19 to *Zea* 18a to the class II shsps. *Rattus* 27 is not included in the tree, being identical to *Mus* 27 and *Cricetulus* 27 in the "α-crystallin domain."

quences in Fig. 1 might allow us to reveal some major structural features of this conserved domain.

Secondary structure predictions of the α-crystallin domain were done for subsets of the alignment, using the program PHD (Rost and Sander 1993; Rost et al. 1994) via the PredictProtein e-mail server at EMBL, Heidelberg. Solvent accessibility prediction was according to Rost and Sander (1994), via the same e-mail server. The results are shown in Fig. 4. For all subsets, the secondary structure predictions start with a short α-helix, followed by three β-sheet stretches, in the first half of the homologous domain. This αβββ motif is repeated in the second half of the domain, although in plant shsps there are two initial α-helices and an α-helix stretch is predicted instead of the second β-sheet stretch. Experimental spectroscopic data support the prevalence of β-sheet conformation in the mammalian shsps and α-crystallins (Li and Spector 1974; Merck et al. 1993a).

The residues of the last β-sheet stretch of both motifs, as well as the residues of the first β-sheet stretch of the first motif, are predicted to be buried, while the α-helix between the motifs is predicted to be in an exposed region. Surface residue predictions of the other parts of the sequence are more equivocal. The location of introns, the large insertion in *Neurospora* hsp30, and the extensive presence of gaps in the region 43–71 (cf. Fig. 1) also make it likely that this is a more flexible and exposed region connecting the two more rigid β-sheet-rich motifs.

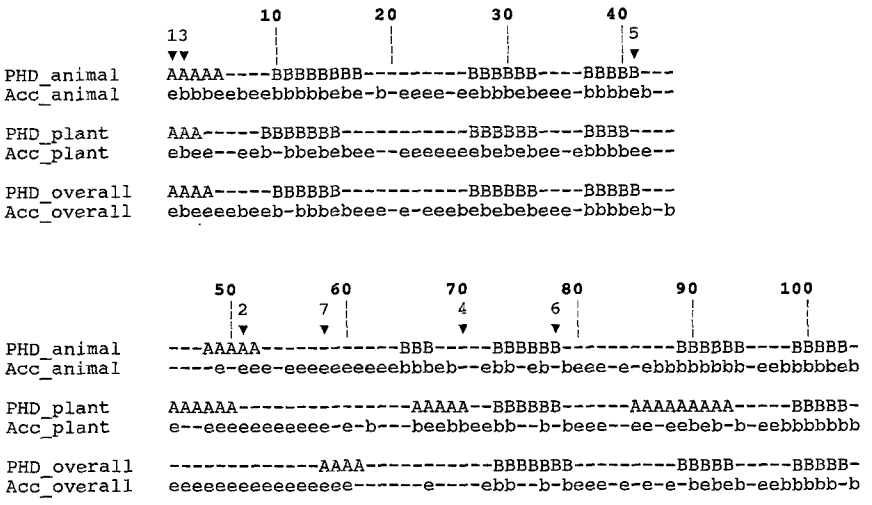


Fig. 4. Secondary structure prediction (*PHD*) and solvent accessibility prediction (*Acc*) by the methods of Rost and Sander (1993, 1994) for the C-terminal "α-crystallin domain" of subsets of the shsp family (animal, animal sequences; plant, plant sequences; overall, bacterial and fungal sequences plus *Chlamydomonas* hsp22, *Petunia* hsp21, *Zea* hsp18a, *Helianthus*, *Schistosoma* p40 N-terminal domain, *Caenorhabditis* hsp12, *Drosophila* hsp27, *Xenopus* hsp30c, *Bos* αB-crystallin, and *Halocynthia* HR-29). Amino acid position numbers are as in Fig. 1. *A*, predicted α-helix; *B*, predicted β-sheet; *e*, predicted exposed residue; *b*, predicted buried residue; -, no prediction made. The positions of introns in the conserved domains are indicated by *arrowheads* as in Figs. 1 and 2.

This is also the position of the helix-turn-helix with DNA-binding properties suggested to be present in *Chenopodium* shsp (Knack et al. 1992). The possible presence of two similar structural motifs in the C-terminal domain of α-crystallin was first proposed by Wistow (1985) and appears to be corroborated by the present data.

The composite hydrophobicity profile (Fig. 5) confirms the two-motif structure of the "α-crystallin domain." Peaks of hydrophobicity concur with the first β-sheet stretches in both motifs, whereas also the other β-sheet stretches are located in the more hydrophobic regions, in agreement with the predicted buried position of these stretches. The exposed α-helix that connects the two motifs locates with the strongly hydrophilic region around position 50. The C-terminal arm, beyond position 110, also has a more hydrophilic character. In α-crystallins, the terminal eight to ten residues indeed form very flexible extensions, as revealed by two-dimensional NMR spectroscopy (Carver et al. 1992).

Evolution of Function

It is logical to assume that the conserved C-terminal domain would be responsible for the common structural and functional properties of the α-crystallin/shsp family. (For review of the latter, see Arrigo and Landry 1994.) Deletion of the last 42 amino acids of this domain in *Drosophila* hsp27 indeed abolishes its capacity to protect cells against heat stress (Mehlen et al. 1993). However, in α-crystallins and mouse hsp25 this separate domain with its C-terminal extension has no chaperone activity in vitro (Merck et al. 1993b). It thus seems likely that also the N-terminal domain is required for proper functioning, and may moderate more specific functions for various subgroups of shsps. Considering the multitude of functions among the members of the shsp family, one might wonder what the primordial function of the common ancestor was. The fact that also in bacteria, like *Escherichia* and *Clostridium*, shsps are now found to be induced by heat shock, suggests that these proteins, too, function in chaperoning other proteins under conditions of stress. The *Escherichia* Ibp proteins also appear to be induced by a high level of unfolded proteins (Allen et al. 1992), which further strengthens the notion that chaperoning is the primordial function of the family, rather than a secondarily developed feature.

The abundant expression of shsps as surface antigens in mycobacteria and in *Schistosoma* must then be one of the various derived functions in the family. For the SP21 protein of the myxobacterium *Stigmatella* it is speculated that it might help pack mRNAs necessary for the early steps of germination or might protect mRNAs of housekeeping genes during periods of development during which other mRNAs are degraded (Heidelbach et al. 1993b). A tight association during heat shock between shsps and a specific subset of mRNAs has earlier been reported for tomato cell cultures (Nover et al. 1989).

Developmental regulation is also well documented in many other members of the family, notably in animal shsps and αB-crystallins (de Jong et al. 1993; Gernold et al. 1993; Marin et al. 1993). Expression of the gene for the shsp of the gastrointestinal nematode *Nippostrongylus brasiliensis* is developmentally regulated (Tweedie et al. 1993). mRNA peak expression is concomitant with the onset of immune damage in the host, but expression is regulated independently of stress stimuli, apparently according to a strict developmental program. In plants, like for *Helianthus* (Almoguera and Jordano 1992) and alfalfa (Györgyey et al. 1991), the pronounced expression of shsp mRNA occurs during embryo midmaturation. Shsp mRNAs are stored for long periods of time in dry seeds of sorghum (Howarth 1990), wheat (Helm and

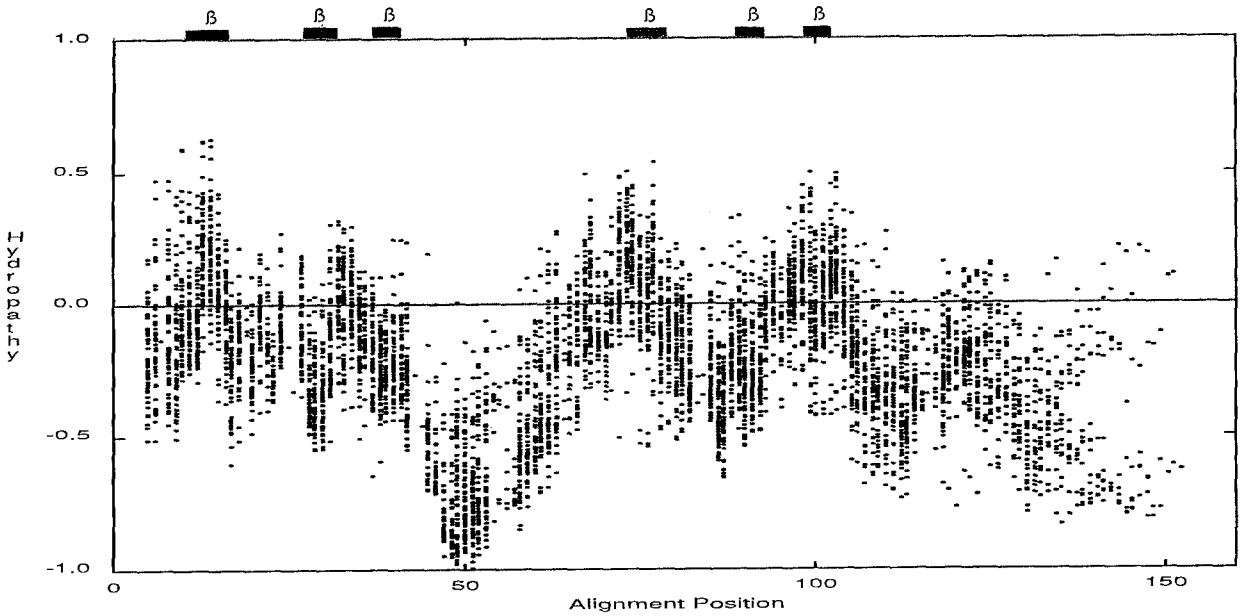


Fig. 5. Composite hydrophobicity profiles of all "α-crystallin domains" of the shsp family, using the hydrophobicity scale of Sweet and Eisenberg (1983), as produced by the program CAMELEON (version 3.0, Oxford Molecular Ltd, Oxford, UK). Amino acid position numbers are as in Fig. 1. Hydrophobicity increases toward the top of the figure; hydrophilicity toward the bottom. Predicted β-sheets (see Fig. 4) are indicated by *bars*.

Abernethy 1990), and pea (Vierling and Sun 1987). All this supports the notion that developmental expression of shsps is quite universal (see Linquist and Craig 1988). The shsps are apparently involved in cytomorphological rearrangements, which may relate to their influence on actin polymerization (Miron et al. 1991) and intermediate filament assembly (Nicholl and Quinlan 1994).

As pointed out by Bouchard (1990), we find an additional non–stress-related function in meiotic cells of three eukaryotic kingdoms. *Lilium* meiotic prophase repeat protein expression is developmentally induced (Bouchard 1990). Two of the four *Drosophila* shsps are expressed in the egg chamber of the ovary during the meiotic period of oogenesis (Zimmerman et al. 1983). The controlling regions for the heat shock and the ovarian induction are different (Hoffman et al. 1987). The *Saccharomyces* shsp is induced during sporulation and meiosis (Kurtz et al. 1986).

Gene duplication is a general mechanism to acquire more functions. For example, in *Pharbitis nil,* one of the shsp genes is induced by heat shock only, the other by light as well (Krishna et al. 1992). The expression of smaller hsp27 transcripts in murine osteoblasts, which may be encoded by separate genes, was facilitated by estrogen treatment prior to heat shock, whereas expression of the normal, longer, hsp27 was induced by heat shock alone (Cooper and Uoshima 1994). Gene duplication must also be the mechanism which enabled an ancestral α-crystallin gene to be recruited for its function as a lens structural protein. Another structural function may be found for the HR-29 protein of *Halocynthia roretzi,* which is localized in body-wall muscle (Shirakata et al. 1986). It may be a component of myofibrils and may act as a stabilizing protein like αB-crystallin does in skeletal muscle (Atomi et al. 1991).

This brief and necessarily fragmentary account makes it clear that gene duplications and divergent evolution have produced a broad array of structural and functional properties in the shsp family, using the common "α-crystallin domain" as an essential building block. The major challenge for the years to come is to elucidate the tertiary and quaternary structure of members of this family and relate these to their various intriguing functions.

Acknowledgments. Part of this work was carried out under the auspices of the Netherlands Foundation for the Life Sciences (SLW) with financial aid from the Netherlands Organization for Scientific Research (NWO). The use of the services and facilities of the Dutch CAOS/CAMM Center is gratefully acknowledged.

References

Allen SP, Polazzi JO, Gierse JK, Easton AM (1992) Two novel heat shock genes encoding proteins produced in response to heterologous protein expression in *Escherichia coli.* J Bacteriol 174:6938–6947

Almoguera C, Jordano J (1992) Developmental and environmental concurrent expression of sunflower dry-seed-stored low-molecular-weight heat-shock protein and Lea mRNAs. Plant Mol Biol 19: 781–792

Arrigo A-P, Landry J (1994) Expression and function of the low-molecular-weight heat shock proteins. In: Morimoto RJ, Tissières A, Georgopoulos C (eds) The biology of heat shock proteins and molecular chaperones. Cold Spring Harbor Laboratory Press, New York, pp 335–373

Atomi Y, Yamada S, Strohman R, Nonomura Y (1991) αB-Crystallin in skeletal muscle: purification and localization. J Biochem 110: 812–822

Bartling D, Bülter H, Liebeton K, Weiler EW (1992) An *Arabidopsis thaliana* cDNA clone encoding a 17.6 kDa clas II heat shock protein. Plant Mol Biol 18:1007–1008

Bhat SP, Nagineni CN (1989) αB subunit of lens-specific protein α-crystallin is present in other ocular and non-ocular tissues. Biochem Biophys Res Commun 158:319–325

Bouchard RA (1990) Characterization of expressed meiotic prophase repeat transcript clones of *Lilium:* meiosis-specific expression, relatedness, and affinities to small heat shock protein genes. Genome 33:68–79.

Candido EPM, Jones D, Dixon DK, Graham RW, Russnak RH, Kay RJ (1989) Structure, organization, and expression of the 16-kDa heat shock gene family of *Caenorhabditis elegans.* Genome 31:690–697

Cao M, Chao H, Doughty BL (1993) Cloning of a cDNA encoding an egg antigen homologue from *Schistosoma mansoni.* Mol Biochem Parasitol 58:169–172

Carver JA, Aquilina JA, Truscott RJW, Ralston GB (1992) Identification by ^{1}H NMR spectroscopy of flexible extensions in bovine lens α-crystallin. FEBS Lett 311:143–149

Caspers G-J, Pennings J, de Jong WW (1994) A partial cDNA sequence corrects the human αA-crystallin primary structure. Exp Eye Res 59:125–126

Chen Q, Vierling E (1991) Analysis of conserved domains identifies a unique structural feature of a chloroplast heat shock protein. Mol Gen Genet 226:425–431

Cooper LF, Uoshima K (1994) Differential estrogenic regulation of small M_1 heat shock protein expression in osteoblasts. J Biol Chem 269:7869–7873

de Jong WW, Leunissen JAM, Leenen PJM, Zweers A, Versteeg M (1988) Dogfish α-crystallin sequences: comparison with small heat shock proteins and *Schistosoma* egg antigen. J Biol Chem 263: 5141–5149

de Jong WW, Leunissen JAM, Voorter CEM (1993) Evolution of the α-crystallin/small heat-shock protein family. Mol Biol Evol 10: 103–126

de Jong WW, Zweers A, Versteeg M, Nuy-Terwindt EC (1984) Primary structures of the α-crystallin A chains of twenty-eight mammalian species, chicken and frog. Eur J Biochem 141:131–140

Felsenstein J (1993) PHYLIP: phylogeny inference package, version 3.5c. Distributed by the author, Department of Genetics, University of Washington, Seattle WA

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155:279–284

Fröhli E, Aoyama A, Klemenz R (1993) Cloning of the mouse hsp25 gene and an extremely conserved hsp25 pseudogene. Gene 128: 273–277

Gaestel M, Gotthardt R, Miller T (1993) Structure and organization of a murine gene encoding small heat-shock protein Hsp25. Gene 128:279–283

Gernold M, Knauf U, Gaestel M, Stahl J, Kloetzel P-M (1993) Development and tissue-specific distribution of mouse small heat shock protein hsp25. Dev Genet 14:103–111

Gribskov M, Luethy R, Eisenberg D (1990) Profile analysis. Methods Enzymol 193:146–159

Groenen PJTA, Merck KB, de Jong WW, Bloemendal H (1994) Structure and modifications of the junior chaperone α-crystallin: from lens transparency to molecular pathology. Eur J Biochem 225:1–19

Györgyey J, Gartner A, Nemeth K, Magyar Z, Hirt H, Heberle-Bors E, Dudits D (1991) Alfalfa heat shock genes are differentially expressed during somatic embryogenesis. Plant Mol Biol 16:999–1007

Heidelbach M, Skladny H, Schairer HU (1993a) Purification and characterization of SP21, a development-specific protein of the myxobacterium *Stigmatella aurantiaca.* J Bacteriol 175:905–908

Heidelbach M, Skladny H, Schairer HU (1993b) Heat shock and development induce synthesis of a low-molecular-weight stress-responsive protein in the myxobacterium *Stigmatella aurantiaca.* J Bacteriol 175:7479–7482

Helm KW, Abernethy RH (1990) Heat shock proteins and their mRNAs in dry and early imbibing embryos of wheat. Plant Physiol 93:1626–1633

Helm KW, LaFayette PR, Nagao RT, Key JL, Vierling E (1993) Localization of small heat shock proteins to the higher plant endomembrane system. Mol Cell Biol 13:238–247

Hendrick JP, Hartl FU (1993) Molecular chaperone functions of heat-shock proteins. Annu Rev Biochem 62:349–384

Hickey E, Brandon SE, Potter R, Stein G, Stein J, Weber LA (1986) Sequence and organization of genes encoding the human 27 kDa heat shock protein. Nucleic Acids Res 14:4127–4145

Hoffman EP, Gerring SL, Corces VG (1987) The ovarian, ecdysterone, and heat-shock-responsive promoters of the *Drosophila melanogaster* hsp27 gene react very differently to perturbations of DNA sequence. Mol Cell Biol 7:973–981

Horwitz J (1992) α-Crystallin can function as a molecular chaperone. Proc Natl Acad Sci USA 89:10449–10453

Howarth C (1990) Heat shock proteins in *Sorghum bicolor* and *Pennisetum americanum.* II. Stored RNA in sorghum seed and its relationship to heat shock protein synthesis during germination. Plant Cell Environ 13:57–64

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

Jakob U, Gaestel M, Engel K, Buchner J (1993) Small heat shock proteins are molecular chaperones. J Biol Chem 268:1517–1520

Jorgensen JA, Nguyen HT (1994) Isolation, sequence and expression of a cDNA encoding a Class I heat shock protein (HSP17.2) in maize. Plant Sci 97:169–175

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, pp 21–132

Kato K, Goto S, Inaguma Y, Hasegawa K, Morishita R, Asano T (1994) Purification and characterization of a 20-kDa protein that is highly homologous to αB crystallin. J Biol Chem 269:15302–15309

Kato K, Shinohara H, Kurobe N, Goto S, Inaguma Y, Ohshima K (1991) Immunoreactive αA-crystallin in rat non-lenticular tissues detected with a sensitive immunoassay method. Biochim Biophys Acta 1080:173–180

Klemenz R, Fröhli E, Steiger RH, Schäfer R, Aoyama A (1991) αB-Crystallin is a small heat shock protein. Proc Natl Acad Sci USA 88:3652–3656

Knack G, Liu Z, Kloppstech K (1992) Low molecular mass heat-shock proteins of a light-resistant photoautotrophic cell culture. Eur J Cell Biol 59:166–175

Krishna P, Felsheim RF, Larkin JC, Das A (1992) Structure and light-induced expression of a small heat-shock protein gene of *Pharbitis nil.* Plant Physiol 100:1772–1779

Kurtz S, Rossi J, Petko L, Lindquist S (1986) An ancient developmental induction: heat-shock proteins induced in sporulation and oogenesis. Science 231:1154–1157

Landry J, Chrétien P, Lambert H, Hickey E, Weber LA (1989) Heat shock resistance conferred by expression of the human HSP27 gene in rodent cells. J Cell Biol 109:7–15

Lee DC, Kim RY, Wistow GJ (1993) An avian αB-crystallin: non-lens expression and sequence similarities with both small (HSP27) and large (HSP70) heat shock proteins. J Mol Biol 232:1221–1226

Li L-K, Spector A (1974) Circular dichroism and optical rotatory dispersion of the aggregates of purified polypeptides of alpha-crystallin. Exp Eye Res 19:49–57

Lindquist S, Craig EA (1988) The heat shock proteins. Annu Rev Genet 22:631–677

Marin R, Valet JP, Tanguay RM (1993) Hsp23 and hsp26 exhibit distinct spatial and temporal patterns of constitutive expression in *Drosophila* adults. Dev Genet 14:69–77

Mehlen P, Briolay J, Smith L, Diaz Latoud C, Fabre N, Pauli D, Arrigo AP (1993) Analysis of the resistance to heat and hydrogen peroxide

stresses in COS cells transiently expressing wild type or deletion mutants of the *Drosophila* 27-kDa heat-shock protein. Eur J Biochem 215:277–284

Merck KB, Groenen PJTA, Voorter CEM, de Haard-Hoekman WA, Horwitz J, Bloemendal H, de Jong WW (1993a) Structural and functional similarities of bovine α-crystallin and mouse small heat-shock protein: a family of chaperones. J Biol Chem 268:1046–1052

Merck KB, Horwitz J, Kersten M, Overkamp P, Gaestel M, Bloemendal H, de Jong WW (1993b) Comparison of the homologous carboxy-terminal domain and tail of α-crystallin and small heat shock protein. Mol Biol Rep 18:209–215

Miron T, Vancompernolle K, Vanderkerckhove J, Wilchek M, Geiger B (1991) A 25-kD inhibitor of actin polymerization is a low molecular weight mass heat shock protein. J Cell Biol 114:255–261

Morimoto RI, Tissières A, Georgopoulos C (eds) (1994) The biology of heat shock proteins and molecular chaperones. Cold Spring Harbor Laboratory Press, New York

Moser D, Doenhoff MJ, Klinkert MQ (1992) A stage-specific calcium-binding protein expressed in eggs of *Schistosoma mansoni.* Mol Biochem Parasitol 51:229–238

Nene V, Dunne DW, Johnson KS, Taylor DW, Cordingley JS (1986) Sequence and expression of a major egg antigen from *Schistosoma mansoni:* homologies to heat shock proteins and α-crystallins. Mol Biochem Parasitol 21:179–188

Nerland AH, Mustafa AS, Sweetser D, Godal T, Young RA (1988) A protein antigen of *Mycobacterium leprae* is related to a family of small heat shock proteins. J Bacteriol 170:5919–5921

Nicholl ID, Quinlan RA (1994) Chaperone activity of α-crystallin modulates intermediate filament assembly. EMBO J 13:945–953

Nover L, Scharf K-L, Neumann D (1989) Cytoplasmic heat shock granules are formed from precursor particles and are associated with a specific set of mRNAs. Mol Cell Biol 9:1298–1308

Patthy L (1994) Introns and exons. Curr Opin Struct Biol 4:383–392

Plesofsky-Vig N, Vig J, Brambl R (1992) Phylogeny of the α-crystallin-related heat-shock proteins. J Mol Evol 35:537–545

Raschke E, Baumann G, Schöffl F (1988) Nucleotide sequence analysis of soybean small heat shock protein genes belonging to two different multigene families. J Mol Biol 199:549–557

Rost B, Sander C (1993) Prediction of protein secondary structure at better than 70% accuracy. J Mol Biol 232:584–599

Rost B, Sander C (1994) Conservation and prediction of solvent accessibility in protein families. Proteins (in press)

Rost B, Sander C, Schneider R (1994) PHD—an automatic mail server for protein secondary structure prediction. Comp Appl Biosci 10: 53–60

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sauer U, Dürre P (1993) Sequence and molecular characterization of a DNA region encoding a small heat shock protein of *Clostridium acetobutylicum.* J Bacteriol 175:3394–3400

Sawada K, Agata K, Eguchi G (1992) Crystallin gene expression in the process of lentodogenesis in cultures of chicken lens epithelial cells. Exp Eye Res 55:879–887

Shirakata M, Takagi T, Konishi K (1986) Isolation and characterization of a 29,000 dalton protein from ascidian (*Halocynthia roretzi*) body wall muscle. Comp Biochem Physiol 85B:71–76

Susek RE, Lindquist SL (1989) HSP26 of *Saccharomyces cerevisiae* is related to the superfamily of small heat shock proteins but is without a demonstrable function. Mol Cell Biol 9:5265–5271

Sweet M, Eisenberg D (1983) Correlation of sequence hydrophobicities measures similarities in three-dimensional protein structures. J Mol Biol 171:479–488

Takagi T, Yasunaga H, Nakamura A (1993) Structure of 29-kDa protein from ascidian (*Halocynthia roretzi*) body wall muscle. J Biochem 113:321–326

Takemoto L, Emmons T (1991) Truncation of αA-crystallin from the human lens. Exp Eye Res 53:811–813

Tardieu A, Delaye M (1988) Eye lens proteins and transparency: from light transmission theory to solution X-ray structural analysis. Annu Rev Biophys Chem 17:47–70

Tseng TS, Yeh KW, Yeh CH, Chang FC, Chen YM, Lin CY (1992) Two rice (*Oryza sativa*) full-length cDNA clones encoding low-molecular-weight heat-shock proteins. Plant Mol Biol 18:963–965

Tweedie S, Grigg ME, Ingram L, Selkirk ME (1993) The expression of a small heat shock protein homologue is developmentally regulated in *Nippostrongylus brasiliensis.* Mol Biochem Parasitol 61:149–154

Verbon A, Hartskeerl RA, Schuitema A, Kolk AHJ, Young DB, Lathigra R (1992) The 14,000-molecular-weight antigen of *Mycobacterium tuberculosis* is related to the alpha-crystallin family of low-molecular-weight heat shock proteins. J Bacteriol 174:1352–1359

Vierling E (1991) The roles of heat shock proteins in plants. Annu Rev Plant Physiol Plant Mol Biol 42:579–620

Vierling E, Sun A (1987) Developmental expression of heat shock proteins in higher plants. In: Cherry J (ed) Environmental stress in plants: biochemical and physiological mechanisms associated with environmental stress tolerance in plants. Springer-Verlag, New York, pp 343–354

Weng J, Wang Z-F, Nguyen HT (1991) Nucleotide sequence of a *Triticum aestivum* cDNA clone which is homologous to the 26 kDa chloroplast-localized heat shock protein gene of maize. Plant Mol Biol 17:255–258

Wilson R, Ainscough R, Anderson K, Baynes C, Berks M, Bonfield J, Burton J, Connel M, Copsey T, Cooper J, Coulson A, Craxton M, Dear S, Du Z, Durbin R, Favello A, Fraser A, Fulton L, Gardner A, Green P, Hawkins T, Hillier L, Jier M, Johnston L, Jones M, Kershaw J, Kirsten J, Laisster N, Latrielle P, Lightning J, Lloyd C, Mortimore B, O'Callaghan M, Parsons J, Percy C, Rifken L, Roopra A, Saunders D, Shownkeen R, Sims M, Smaldon N, Smith A, Smith M, Sonnhammer E, Staden R, Sulston J, Thierry-Mieg J, Thomas K, Vaudin M, Vaughan K, Waterston R, Watson A, Weinstock L, Wilkinson-Sproat J, Wohldman P (1994) 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans.* Nature 368:32–38

Wistow G (1985) Domain structure and evolution in α-crystallins and small heat-shock proteins. FEBS Lett 181:1–6

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

Zimmerman JL, Petri W, Meselson M (1983) Accumulation of a specific subset of *D. melanogaster* heat shock mRNAs in normal development without heat shock. Cell 32:1161–1170