J Mol Evol (1992) 35:537–545

Journal of Molecular Evolution

Phylogeny of the α-Crystallin-Related Heat-Shock Proteins

Nora Plesofsky-Vig,[1,2] Jesse Vig,[1] and Robert Brambl[1,3]

[1] Department of Plant Biology
[2] Department of Genetics and Cell Biology
[3] Plant Molecular Genetics Institute, University of Minnesota, St. Paul, MN 55108, USA

Summary. Phylogenetic relationships were examined among 35 α-crystallin-related heat-shock proteins from animals, plants, and fungi. Approximately one-third of the aligned amino acids in these proteins were conserved in 74% of the proteins, and three blocks of consensus sequence were identified. Relationships were established by maximum parsimony and distance matrix analyses of the aligned amino acid sequences. The inferred phylogeny trees show the plant proteins clearly divided into three major groups that are unrelated to taxonomy: the chloroplast-localized proteins and two groups that originate from a common ancestral plant protein. The animal proteins, in contrast, branch in accordance with taxonomy, the only clear exception being the α-crystallin subgrouping of vertebrates. This analysis indicates that the small heat-shock proteins of animals have diverged more widely than have the plant proteins, one group of which is especially stable.

Key words: Heat-shock proteins — α-crystallin — Phylogeny — Maximum parsimony—Distance matrix

Introduction

Among the common classes of heat-shock proteins are two well-conserved high-molecular-weight classes, hsp70 and hsp83/90, and a diverse class of low-molecular-weight proteins that are unified by limited homology and by shared structural features (Lindquist 1986). A trademark of this last group of proteins is their unexpected homology to vertebrate eye lens α-crystallin (Ingolia and Craig 1982). Aside from this 38-amino-acid region (Susek and Lindquist 1989) near their carboxy termini, however, the sequence differences among the proteins are more striking than their similarities. In the present study we have compared 35 of these α-crystallin-related proteins across the fungal, plant, and animal kingdoms. The aligned amino acid sequences have been extended to include up to 89 residues.

One of our objectives in deriving phylogenetic trees for these proteins was to determine the relationship of the small heat-shock protein of *Neurospora crassa* (Plesofsky-Vig and Brambl 1990) to the small heat-shock proteins of other species. Among other fungi, only the sequence of the *Saccharomyces cerevisiae* heat-shock protein hsp26 has been determined (Susek and Lindquist 1989; Bossier et al. 1989). Several small heat-shock-protein genes have been sequenced from animals, but these include relatively few different animal species, that is, *Xenopus, Caenorhabditis, Drosophila,* chicken, mouse, and human. On the other hand, multiple genes from several different species of flowering plants have now been sequenced and, in the present study, they yielded reasonably secure classifications based on parsimony and distance analysis.

Materials and Methods

For parsimony analysis and tree construction, we used the sequences of 35 small heat-shock proteins (including hsp18 of the prokaryote *Mycobacterium*) or 34 proteins (excluding the myco-

Offprint requests to: N. Plesofsky-Vig

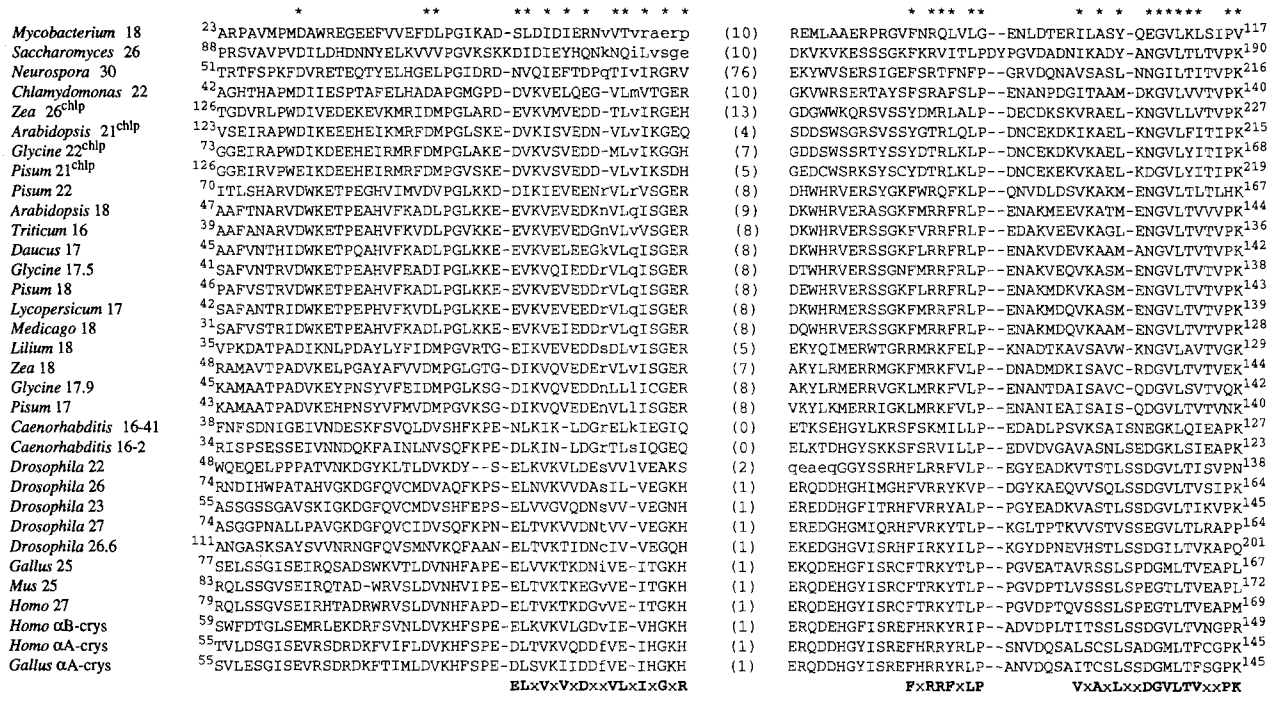

Fig. 1. Amino acid sequence alignment of 33 small heat-shock proteins. The uppercase letters indicate the aligned residues, and asterisks denote the residues that are identical or similar (conservatively substituted) in at least 74% of the proteins. Three blocks of identified consensus sequence are shown at the bottom. The amino acids (in parentheses) were considered conservative amino acid substitutions in the blocks of consensus sequence: block 1, E(D)L(I/V)xVxV(I/L)xDxxV(I)L(V)xI(V)xGxR(H); block 2, FxRR(K)F(Y)xLP; Block 3, V(I)xA(S)xL(M/I/V)xxD(N)GVLT(S)V(I/L)xxPK. The number of amino acids in the omitted variable-length region is shown in parentheses for each protein. Two additional *Pharbitis* sequences (unpublished information, Krishna et al.) were used for this analysis, but are not presented in this alignment.

bacterial protein). The 34 eukaryotic sequences include proteins from vertebrate animals (*Gallus, Mus,* and *Homo*) and invertebrate animals (*Caenorhabditis* and *Drosophila*), monocotyledonous plants (*Zea, Triticum,* and *Lilium*) and dicotyledonous plants (*Arabidopsis, Glycine, Pisum, Daucus, Lycopersicum, Medicago, Pharbitis*), two ascomycetous fungi (*Saccharomyces* and *Neurospora*), and a green alga (*Chlamydomonas*). Most of the sequences that were available to us for the α-crystallin-related small heat-shock proteins were included in the analysis. Not included were sequences that were nearly identical to included proteins of the same or closely related species, proteins whose complete sequences were not known, and a few proteins with highly divergent sequences that appeared to distort the phylogeny trees. α-crystallin sequences were also included in the analysis; the B chain of these proteins has been shown to be a heat-shock protein (Klemenz et al. 1991).

The amino acid sequences were aligned by visual inspection with the aid of the Multiple Aligned Sequence Editor (MASE 3.1) alignment program (Faulkner and Jurka 1988). Lowercase letters (Fig. 1) indicate sequences that could not be aligned unambiguously and were not included in the parsimony or distance analysis. The distance analysis did not include sequence gaps; however, these gaps were included in the parsimony analysis, except for gaps at two positions that seemed ambiguous (corresponding to lowercase letters in the aligned sequences). Especially notable is a gap at the 12th position from the end (Fig. 1) that appears in all non-animal proteins. The asterisks at the top of selected amino acids indicate those residues that are identical or similar in at least 74% of the proteins (appearing in 26 out of the 35 sequences). This threshold ensured that the amino acid indicated (or a conservative substitution) was present in proteins of both plant and animal origin.

The trees of maximum parsimony were constructed with the PROTPARS and BOOT programs of the Phylogenetic Inference Package (PHYLIP 3.4; Felsenstein 1981, 1985). The sequences of 17 of these proteins were analyzed by the distance matrix method (Fitch and Margoliash 1967) of the FITCH program of PHYLIP; distances were based on the number of nucleotide changes required to convert one amino acid to the other.

In this study, proteins are designated by their predicted molecular weight or by published names that are based on their predicted or apparent molecular weight. The analyzed protein sequences are: *Mycobacterium leprae* hsp18 (Nerland et al. 1988), *Saccharomyces cerevisiae* (yeast) hsp26 (Susek and Lindquist 1989; Bossier et al. 1989), *Neurospora crassa* hsp30 (Plesofsky-Vig and Brambl 1990), *Chlamydomonas reinhardtii* hsp22 (Grimm et al. 1989), *Zea mays* (maize) hsp26 (Nieto-Sotelo et al. 1990), *Arabidopsis thaliana* hsp21 (Chen and Vierling 1991), *Glycine max* (soybean) hsp22 (Vierling et al. 1988), *Pisum sativum* (pea) hsp21 (Vierling et al. 1988), *P. sativum* hsp22 (hsp22.7, Vierling 1991), *A. thaliana* hsp18 (JQ0352, Takahashi and Komeda 1989), *Triticum aestivum* (wheat) hsp16 (C5-8, McElwain and Spiker 1989), *Daucus carota* (carrot) hsp17 (hsp17.7, Darwish et al. 1991), *G. max* hsp17.5 (hsp17.5-M, Nagao et al. 1985), *P. sativum* hsp18 (hsp18.1, DeRocher et al. 1991), *Lycopersicum* sp. (tomato) hsp17 (pTOM66, Fray et al. 1990), *Medicago sativa* (alfalfa) hsp18 (hsp18.1, Györgyey et al. 1991), *Lilium* sp. hsp18 (pLEc6, Bouchard 1990), *Z. mays* hsp18 (hsp18-9, Goping et al. 1991; Dietrich et al. 1991), *G. max* hsp17.9 (hsp17.9-D, Raschke et al. 1988), *Pharbitis nil* hsp17 (shsp-1, Krishna P, Felsheim RF, Larkin JC, Das A, personal communication), *P. nil* hsp18 (shsp-2, Krishna et al., personal communication), *P. sativum* hsp17 (hsp17.7, Lauzon et al. 1990), *Caenorhabditis elegans* hsp16-41 (Jones et al. 1986), *C. elegans*

hsp16-2 (Jones et al. 1986), *Drosophila melanogaster* hsp22 (Ingolia and Craig 1982), *D. melanogaster* hsp26 (Ingolia and Craig 1982), *D. melanogaster* hsp23 (Ingolia and Craig 1982), *D. melanogaster* hsp27 (Ingolia and Craig 1982), *D. melanogaster* hsp26.6 (67B1, Ayme and Tissières 1985), *Gallus gallus* (chicken) hsp25 (25-kD IAP, Miron et al. 1991), *Mus musculus* (mouse) hsp25 (Gaestel 1989), *Homo sapiens* hsp27 (Hickey et al. 1986), *H. sapiens* αB-crystallin (Kramps et al. 1977), *H. sapiens* αA-crystallin (de Jong et al. 1975), and *G. gallus* αA-crystallin (de Jong et al. 1984).

Results

We made comparisons among this diverse group of 35 proteins by identifying similar sequences and extending the amino acid alignments into regions of the proteins that have diverged considerably, especially between the major groups of organisms. This extended alignment was facilitated by the large number of proteins being compared, and it depends upon the presence of conserved amino acids at well-distributed sites in a majority of the proteins (Fig. 1). Within the aligned region, there is a sequence of considerable length diversity, for many of the proteins, that was not included in the comparison. A consistent feature of this variable sequence is that it includes a cluster of charged amino acids. It is notable that this hypervariable region is present at the same position in many of the small heat-shock proteins (Fig. 1). For example, there is an especially large hypothetical insertion of 76 amino acids in the *Neurospora* hsp30, but smaller insertions of 10 and 13 residues appear in the yeast hsp26 and the maize chloroplast protein, respectively. Several of the plant proteins have an eight-amino-acid insertion in this region; and most of the animal proteins, despite extreme sequence divergence elsewhere, have one amino acid. Finally, the sequences of the two *C. elegans* proteins are continuous in this region.

The sequence alignment that is shown in Fig. 1 includes all the protein sequences that were used in the maximum parsimony analysis. In this alignment, 30 amino acids out of the 90 presented are identical or similar in at least 74% of the proteins; almost half of these amino acids (13) precede the hypothetical insertion. This sequence alignment reveals that there are three blocks of widely conserved sequences in the small heat-shock proteins (Fig. 1). Block 1 has 10 out of 18 consecutive amino acids that are conserved, in block 2 there are 6 conserved amino acids out of 8, and in block 3 there are 11 conserved residues out of 16 (or 17).

Distance measurements (Table 1) indicate that the small heat-shock proteins of animals have undergone greater divergence than those of plants. This is seen in the greater distance of individual animal proteins from prokaryotic and fungal proteins, compared with the distance between specific plant proteins and these microbial heat-shock proteins. The range of distances of the animal proteins from the fungal *Neurospora* hsp30 is 104.7 to 120.7, while the range of distances between individual plant proteins and the fungal hsp30 is 91.1 to 98.9 for the slowly evolving class I proteins (discussed below) and 103.4 to 106.7 for the class II plant proteins. The same asymmetrical relationship is shown in the distances of animal and plant proteins to hsp26 of yeast. The animal proteins are also further from the mycobacterial protein, which may be considered an outgroup. The distances for animal proteins range from 122.0 to 127.2, while the distances of the plant proteins to the mycobacterial protein range from 104.8 to 115.5. Much of this sequence divergence appears to have occurred after the establishment of the animal kingdom, since it is also seen in interspecies comparisons between various animal proteins. For example, the maximum measured distance among the animal proteins is 104.7 between *Caenorhabditis* hsp16-2 and *Drosophila* hsp26. In comparison, the maximum distance between any two plant proteins, excluding the chloroplast subgroup, is 84.3 between pea hsp22 and *Lilium* hsp18.

The inferred parsimony trees (Fig. 2) indicate that the plant proteins of any one species fall into one of three discrete groups that include members of various species. This suggests that in the plant ancestry there was an early divergence of the small heat-shock proteins into distinct types that were maintained throughout flowering plant evolution into the modern species. The multiple proteins that are also present in individual animal species, in contrast, do not fall into protein classes that are clearly shared by separate species. Instead, the principal distinctions among the animal proteins appear to correspond to species' distinctions. A clear exception to this generalization is found in the vertebrate clade, where there is a subgroup comprised of α-crystallins from various species that is distinct from the subgroup of traditional heat-shock proteins. Thus, the animal proteins display two properties that contrast with the plant proteins: they have diverged widely from one another and, in general, they do not clearly form protein subgroups. The more extensive evolution of the animal proteins, compared with plant proteins, may have led to the disappearance of some protein types and greater divergence of others. Only one α-crystallin-related protein has been identified and characterized for either of the fungal representatives *Saccharomyces* and *Neurospora;* if related proteins exist in these fungi they have diverged considerably (Susek and Lindquist 1989; Plesofsky-Vig and Brambl 1990). The distance (Table 1) between the charac-

Table 1. Distance data and amino acid identity for α-crystallin-related proteins

Organisms	Evolutionary distance/amino acid identity										
	Myc.	*Sac.*	*Neu.*	*Chly.*	*Zea*	*P21*	*P22*	*Ara.*	*Tri.*	*P18*	*Lil.*
Mycobacterium 18		104.8	114.3	107.2	114.5	112.0	106.0	108.3	107.1	104.8	115.5
Saccharomyces 26	27.7		95.3	96.4	107.1	115.5	90.6	101.2	93.0	94.1	102.4
Neurospora 30	21.7	33.3		89.8	95.5	114.4	96.6	98.9	94.4	91.1	103.4
Chlamydomonas 22	27.7	31.0	31.8		87.6	97.8	82.0	76.4	73.0	78.7	90.1
Zea 26^{chlp}	20.5	23.8	33.0	34.1		58.0	101.1	98.9	89.9	87.5	95.4
Pisum 21^{chlp}	24.1	22.6	20.4	27.3	54.5		100.0	95.5	88.8	89.8	95.4
Pisum 22	31.3	40.5	30.7	40.9	29.5	28.4		47.8	48.9	44.4	84.3
Arabidopsis 18	30.1	32.1	31.8	46.6	35.2	31.8	61.4		18.9	19.1	75.0
Triticum 16	28.9	36.9	34.1	45.5	36.4	34.1	62.5	86.4		22.2	70.0
Pisum 18	30.1	36.9	35.2	44.3	40.9	36.4	63.6	86.4	81.8		74.2
Lilium 18	21.7	29.8	31.8	36.4	38.6	30.7	37.5	45.5	44.3	46.6	
Pharbitis 17	28.9	31.0	25.0	40.9	34.1	29.5	40.9	50.0	47.7	51.1	53.4
Caenorhabditis 16-2	14.6	24.1	26.4	23.0	23.0	21.8	24.1	21.8	26.4	24.1	20.7
Drosophila 22	19.7	24.7	19.8	24.7	25.9	27.2	23.5	28.4	30.9	32.1	22.2
Drosophila 26	20.5	17.9	15.9	20.5	23.9	25.0	22.7	26.1	29.5	25.0	22.7
Homo 27	16.9	19.0	23.9	20.5	21.6	19.3	20.5	21.6	22.7	23.9	21.6
Gallus αA-crys	19.3	17.9	25.0	20.5	20.5	23.9	21.6	22.7	21.6	26.1	19.3

The upper-right half of the table presents the distance data (×100) for selected small heat-shock proteins, based on the aligned sequences shown in Fig. 1. The pair-wise distance value is the number of nucleotide changes required to convert one amino acid to the other amino acid, divided by the number of amino acid positions compared. The lower-left half of the table presents the percentage of identical amino acids in the pair of aligned amino acid sequences

terized proteins of these related fungi is large, 95.3, which is comparable to the distance from the *Drosophila* proteins to the human heat-shock protein.

All the small heat-shock proteins of plants belong to one of three distinct groups (Fig. 2). The group that corresponds to class I proteins, as recently defined (Vierling 1991), includes *Arabidopsis* hsp18, wheat hsp16, carrot hsp17, soybean hsp17.5, pea hsp18, tomato hsp17, and alfalfa hsp18. This group shares an ancestral node, by both parsimony and distance analyses (Figs. 2, 3), with the second major group of plant proteins, or class II proteins (Vierling 1991). The class II proteins include lily hsp18, maize hsp18, soybean hsp17.9, two *Pharbitis* heat-shock proteins of 17 and 18 kDa, and pea hsp17. The inferred parsimony and distance trees indicate that pea hsp22, which is reported to be an endomembrane protein (Vierling 1991), is clearly related to the class I plant proteins, from whose line of ancestry it diverged early.

The third major plant group in this phylogenetic analysis consists of the chloroplast-localized small heat-shock proteins, which are nuclear encoded. The position at which these chloroplast-localized proteins diverged, relative to other groups, is not certain. Bootstrap statistical analysis indicates that branching of the plant chloroplast proteins before *Chlamydomonas* hsp22, shown in Fig. 2, is supported by only 13.5 possible parsimony trees out of 50 examined, whereas 10.9 of the possible trees show the *Chlamydomonas* protein grouping with the chloroplast proteins. The inferred distance tree (Fig. 3) shows still a different arrangement. There might have been additional support for the grouping of the chloroplast-localized *Chlamydomonas* protein with the chloroplast proteins of higher plants if a shared gap introduced into their alignments (Fig. 1) had been unambiguous.

The origin of the chloroplast protein clade close to the point of radiation of the ancestors of modern eukaryotic groups (Figs. 2, 3) suggests that the chloroplast proteins had a separate origin from the other plant proteins, not a surprising conclusion in light of the independent genetic origin of chloroplasts. It is likely that the chloroplast small heat-shock proteins, although encoded by nuclear genes, are descended from a protein encoded by the chloroplast progenitor; this gene would subsequently have been transferred to the nuclear DNA. The *Chlamydomonas* protein may have derived from the same endosymbiotic event as the chloroplast proteins of higher plants but diverged far from them, or it may have the same ancestry as the higher-plant class I and II proteins and have acquired a chloroplast function. A parallel process of specialization may be seen in the endomembrane location of hsp22 of pea, which is clearly in the class I protein lineage.

Analysis of fungal small heat-shock proteins is limited by the availability of only two sequences. The grouping of the fungal proteins depended upon whether *Mycobacterium* hsp18 was included in the sequence data used for inferring the phylogenetic trees (Figs. 2, 3). When the mycobacterial protein

Table 1. Extended

Evolutionary distance/amino acid identity					
Phb.	*Cae.*	*Dr22*	*Dr26*	*Hom.*	*Gal.*
106.0	127.2	122.2	125.6	126.8	122.0
95.3	103.7	112.2	120.5	120.5	114.5
106.7	104.7	114.0	120.7	109.2	105.7
87.6	109.3	109.4	116.3	112.8	115.1
93.2	108.8	107.1	102.3	116.3	114.0
102.3	110.6	101.2	103.5	116.3	109.3
78.9	110.3	107.0	109.2	114.9	110.3
69.7	111.6	102.4	106.9	114.9	112.6
68.9	108.0	98.8	96.6	110.3	106.9
66.3	109.3	97.6	106.9	116.1	106.9
52.3	116.5	106.0	109.3	114.0	118.4
	118.6	101.2	108.0	119.5	110.3
16.1		98.8	104.7	87.2	82.6
27.2	29.6		67.1	95.3	85.9
20.5	30.7	47.6		93.1	89.7
14.8	37.5	35.4	41.6		60.9
20.5	39.8	36.6	40.4	52.8	

was included in the parsimony and distance analyses, the prokaryotic protein grouped with *Saccharomyces* hsp26, leaving the *Neurospora* protein to originate independently. However, when the prokaryotic protein was excluded from the analysis, the two fungal proteins formed a clade with a common node. The divergence between proteins in either group occurred very close to their common node (Fig. 3). The inclusion of the mycobacterial protein in the phylogenetic trees did not change other branching patterns (data not shown).

For the animal proteins, the tree of maximum parsimony (Fig. 2) accords with accepted evolutionary relationships (Cedergren et al. 1988), showing the proteins of invertebrates diverging from the main phylogenetic line before the vertebrate proteins. The nematode proteins branch off first, followed by divergence of the insect proteins (represented by *Drosophila*) from the clade of vertebrate proteins. Among the vertebrate small heat-shock proteins, the chicken protein diverges before the two mammalian proteins. The α-crystallins form a distinct subgroup, within the vertebrate clade, that shares a common node with the small heat-shock proteins. The α-crystallin B-chain, which is heat-shock-inducible in non-lens tissue (Klemenz et al. 1991), groups separately from the A-chains. Distance matrix analysis yields a different topology, in which the insect proteins branch off from the main line before the nematode protein. The distance between the two *Drosophila* proteins included in the analysis is 67.1 (Table 1), approximately the same as the distance between class I and class II proteins of plants, but no general classes of animal heat-shock proteins have been identified or are evident from this analysis.

Within the three classes or groups of plant proteins (Fig. 2), the order of protein divergence shows varying degrees of conformation with taxonomic relationships. There is clear conformation within the chloroplast lineage, where the monocot maize protein diverges from the clade of dicot proteins; within the dicots, proteins of pea and soybean, members of the Leguminoseae, group together leaving the *Arabidopsis* (Cruciferae) protein to branch separately. There is looser taxonomic agreement in the branching pattern of the class II proteins. Proteins of the monocots, lily and maize, diverge early and consecutively from the class II lineage, but they do not form a clade. Furthermore, the pea and soybean class II proteins do not group together. Instead, pea hsp17 groups with *Pharbitis* hsp18 (supported by 36.3 parsimony trees out of 50). A possible grouping of soybean hsp17.9 with *Pharbitis* hsp17 is supported by 13.0 trees, compared with 24.5 that separate them in the consensus tree. Aside from the divergent hsp22 of pea, the proteins within class I are all closely related to one another and give no evidence of taxonomic groupings. The only subgroup within these class I proteins, supported by 25.8 parsimony trees, contains *Arabidopsis* hsp18 and wheat hsp16. The proteins of tomato, alfalfa, soybean, pea, and carrot show no particular grouping pattern. The low amount of divergence within the plant class I proteins, as compared to the class II and the chloroplast proteins, is also illustrated by the distance tree (Fig. 3). As shown in Table 1, the distance between the class I proteins of wheat (hsp16) and pea (hsp18) is only 22.2, while the distance between the maize and pea chloroplast proteins is 58 and the distance between the lily and *Pharbitis* class II proteins is 52.3.

Discussion

In this study we analyzed sequences of related but widely divergent proteins, the α-crystallin-related heat-shock proteins, from a broad spectrum of organisms. The sequence data that were employed for inferring phylogenetic relationships among these proteins ranged from 14.6% amino acid identity to 86.4% identity (Table 1). Overall, these proteins have evolved rapidly, especially when compared to other heat-shock proteins of the hsp70 and hsp83/90 families (Lindquist 1986). Nevertheless, the small heat-shock proteins appear to be induced by the same stimuli in all organisms, i.e., by high-temperature stress and during sexual reproduction or embryogenesis (Lindquist 1986). The higher-order particulate structure of these proteins also appears to be a constant feature (Arrigo et al. 1988). However, it is not known if these proteins have the

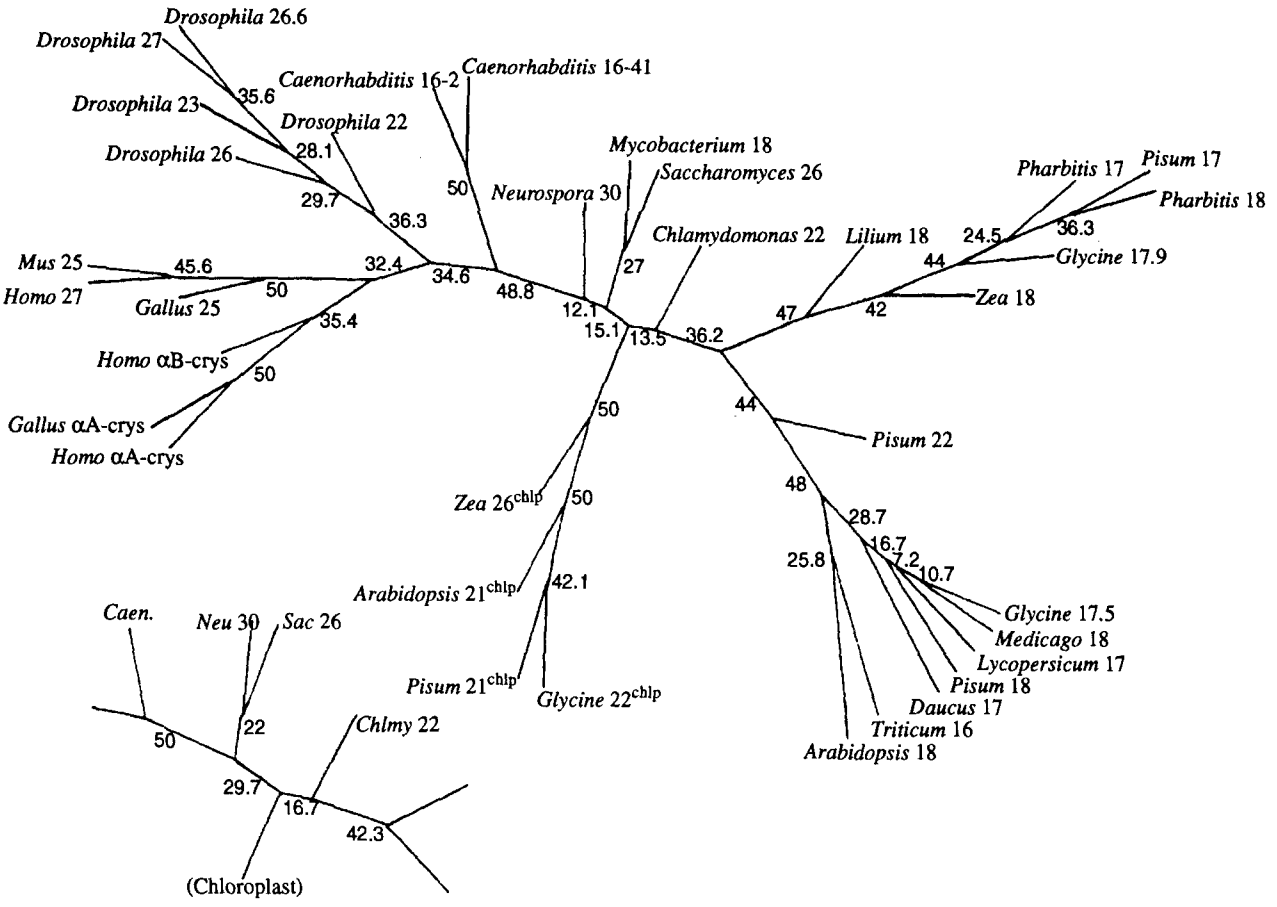

Fig. 2. Inferred phylogenetic tree of the α-crystallin-related heat-shock proteins, based on maximum parsimony analysis of the aligned sequences in Fig. 1. The bootstrap value for each branch indicates how many parsimony trees, out of 50 analyzed, specify the division of groups defined by the branch. Fractional values were assigned when multiple trees of equal parsimony were generated for any one of the 50 samples. The inset shows a portion of the same analysis of the data without the *Mycobacterium* hsp18 sequence. The branch lengths are not proportional to distance.

same functions in all organisms or if, even within the same organism, related proteins might have different functions. The sequence alignments of disparate proteins may highlight islands of widely conserved amino acids that are important for the structure or fundamental function of the protein.

Phylogenetic comparisons of gene or protein sequences can reveal relationships between different taxonomic groups, or such comparisons may reveal relationships between the proteins that reflect differentiation of function. The phylogenies inferred in this study, by both maximum parsimony and distance matrix methods, show the fungal and animal small heat-shock proteins generally conforming to established taxonomic relationships. An important exception is the division within the vertebrate clade between the heat-shock proteins and the α-crystallins. The topology of the plant proteins, in contrast, does not conform to taxonomic relationships, with the exception of the clade of chloroplast-localized proteins. The other two major classes of plant proteins, which derive from a common ancestral node, are not based on taxonomy; nor are possible subgroups within them taxonomically based.

Another distinction, revealed by this study, between the animal and the plant small heat-shock proteins is that the animal proteins have diverged much more extensively from one another and from the prokaryotic and fungal proteins than have the plant proteins. The plant class I proteins especially have been extremely stable. The small heat-shock proteins do not appear to be unique, however, in showing more divergence among animal species than among plants. In studies of small-subunit rRNA gene sequences (Elwood et al. 1985), the distance between two vertebrate animals, rat and *Xenopus*, was shown to be 0.044, while the distance between two monocot plants, rice and maize, was only 0.023. In addition, the distance between rat and *Saccharomyces*, 0.265, was larger than the corresponding distance between rice and *Saccharomyces*, 0.207 (Elwood et al. 1985). A similar pattern was shown by cytochrome *c* sequences (Kemmerer et al. 1991). The relatively large protein distance

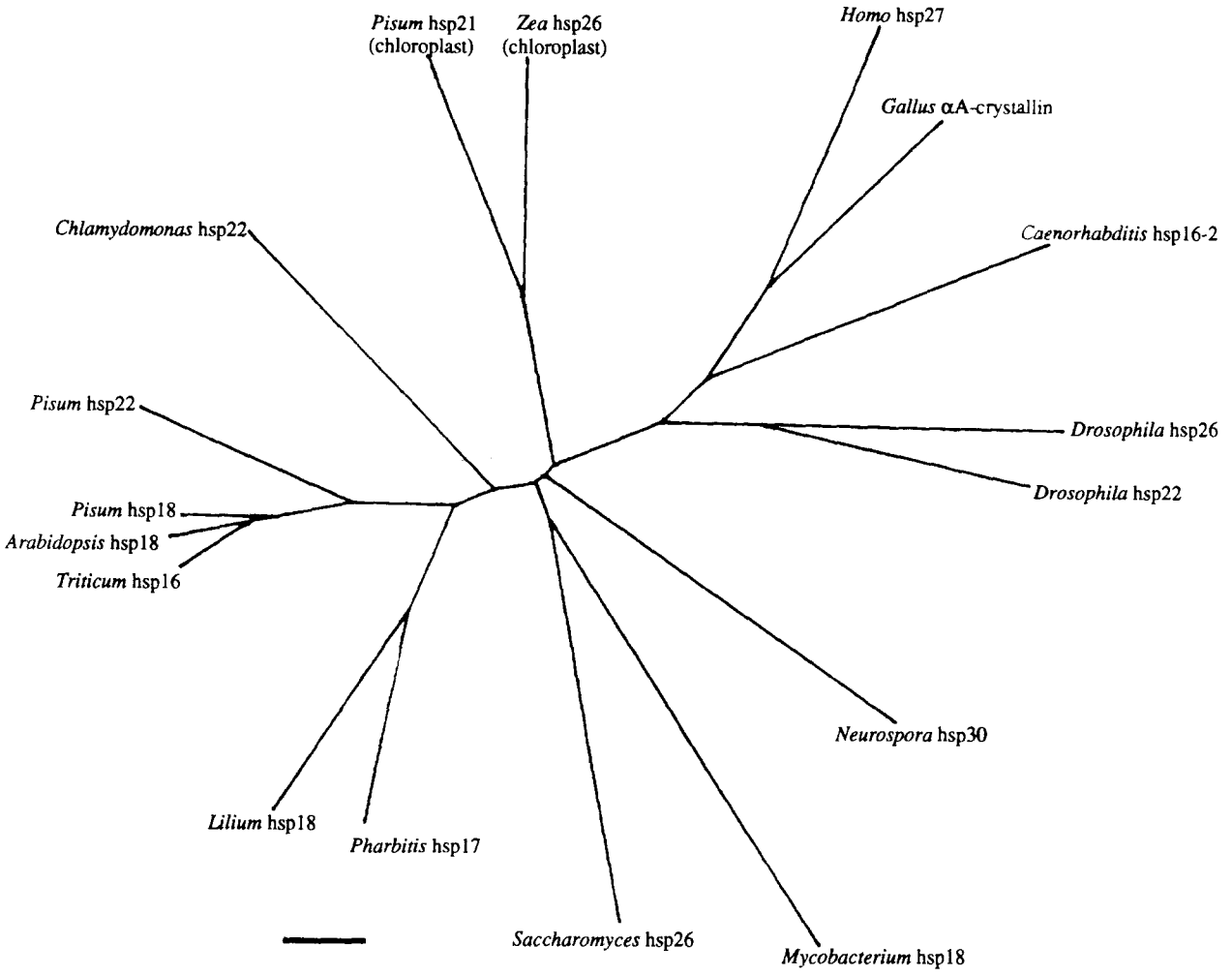

Fig. 3. Inferred phylogenetic tree of 17 α-crystallin-related heat-shock proteins, based on distance matrix analysis (Table 1) of the aligned sequences in Fig. 1. The lengths of the branches are proportional to distance.

seen in this study between *Saccharomyces* and *Neurospora,* which are both ascomycetous fungi, is supported by other phylogenetic analyses of small-subunit rRNA gene sequences (Greenwood et al. 1991) and chitin synthase genes (Bowen et al. 1992).

The vertebrate α-crystallins are assumed to have evolved from heat-shock protein genes, as other lens crystallins have evolved from functionally distinct proteins (Piatigorsky and Wistow 1989). Nevertheless, the distance from chicken α-crystallin is smaller than the distance from the human heat-shock protein to 13 of the 15 nonvertebrate heat-shock proteins in the matrix (Table 1, Fig. 3). This reduced distance may reflect a constraint imposed on the α-crystallins due to their acquisition of a new, non-heat-shock, function. Among the plant heat-shock proteins, the extremely slow rate of class I protein evolution may indicate that the class I proteins have acquired a novel, constraining function. Two heat-shock proteins of *Pharbitis,* which group separately within the class II proteins, appear to be subject to different modes of induction. Hsp17 of *Pharbitis* is expressed in response to light (Krishna et al. personal communication), while preliminary results suggest that light does not induce hsp18. The complex topology of the plant small heat-shock proteins may guide investigators to look for unique functions that are characteristic of particular groups or subgroups of these proteins.

Like other families of heat-shock proteins, the α-crystallin-related proteins are synthesized during specific developmental stages, under non-heat-shock conditions (Lindquist 1986). A dual role might be intrinsic to the proteins, since their joint association with environmental stress and sexual/embryonic development appears to be widespread, appearing in organisms as diverse as yeast (Kurtz et al. 1986), lily (Bouchard 1990), and *Drosophila* (Zimmerman et al. 1983). There is no obvious correlation between the branching patterns within the plant heat-shock proteins and a division of these two functions. Representatives of both class I and

class II plant proteins have been found to be expressed during development, as well as in response to heat shock.

Acknowledgments. This research was supported by the National Institute of General Medical Sciences (GM-19398) and the USDA Competitive Research Grants Office.

References

Arrigo A-P, Suhan JP, Welch WJ (1988) Dynamic changes in the structure and intracellular locale of the mammalian low-molecular-weight heat shock protein. Mol Cell Biol 8:5059–5071

Ayme A, Tissières A (1985) Locus 67B of *Drosophila melanogaster* contains seven, not four, closely related heat shock genes. EMBO J 4:2949–2954

Bossier P, Fitch IT, Boucherie H, Tuite MF (1989) Structure and expression of a yeast gene encoding the small heat-shock protein Hsp26. Gene 78:323–330

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

Bowen AR, Chen-Wu JL, Momany M, Young R, Szaniszlo PJ, Robbins PW (1992) Classification of fungal chitin synthases. Proc Natl Acad Sci USA 89:519–523

Cedergren R, Gray MW, Abel Y, Sankoff D (1988) The evolutionary relationships among known life forms. J Mol Evol 28:98–112

Chen Q, Vierling E (1991) Erratum: analysis of conserved domains identifies a unique structural feature of a chloroplast heat shock protein. Mol Gen Genet 228:328

Darwish K, Wang L, Hwang CH, Apuya N, Zimmerman JL (1991) Cloning and characterization of genes encoding low molecular weight heat shock proteins from carrot. Plant Mol Biol 16:729–731

de Jong WW, Terwindt EC, Bloemendal H (1975) The amino acid sequence of the A chain of human α-crystallin. FEBS Lett 58:310–313

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

DeRocher AE, Helm KW, Lauzon LM, Vierling E (1991) Expression of a conserved family of cytoplasmic low molecular weight heat shock proteins during heat stress and recovery. Plant Physiol 96:1038–1047

Dietrich PS, Bouchard RA, Casey ES, Sinibaldi RM (1991) Isolation and characterization of a small heat shock protein gene from maize. Plant Physiol 96:1268–1276

Elwood HJ, Olsen GJ, Sogin ML (1985) The small-subunit ribosomal RNA gene sequences from the hypotrichous ciliates *Oxytricha nova* and *Stylonychia pustulata*. Mol Biol Evol 2:399–410

Faulkner DV, Jurka J (1988) Multiple aligned sequence editor (MASE). Trends Biochem Sci 13:321–322

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

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

Fray RG, Lycett GW, Grierson D (1990) Nucleotide sequence of a heat-shock and ripening-related cDNA from tomato. Nucleic Acids Res 18:7148

Gaestel M, Gross B, Denndorf R, Strauss M, Schunk W-H, Kraft R, Otto A, Bohm H, Stahl J, Drabsch H, Bielka H (1989) Molecular cloning, sequencing and expression in *Escherichia coli* of the 25-kDa growth-related protein of Ehrlich ascites tumor and its homology to mammalian stress proteins. Eur J Biochem 179:209–213

Goping IS, Frappier JRH, Walden DB, Atkinson BG (1991) Sequence, identification and characterization of cDNAs encoding two different members of the 18 kDa heat shock family of *Zea mays* L. Plant Mol Biol 16:699–711

Greenwood SJ, Sogin ML, Lynn DH (1991) Phylogenetic relationships within the class Oligohymenophorea, phylum Ciliophora, inferred from the complete small subunit rRNA gene sequences of *Colpidium campylum, Glaucoma chattoni,* and *Opisthonecta henneguyi*. J Mol Evol 33:163–174

Grimm B, Ish-Shalom D, Even D, Glaczinski H, Ottersbach P, Ohad I, Kloppstech K (1989) The nuclear-coded chloroplast 22-kDa heat-shock protein of *Chlamydomonas*. Eur J Biochem 182:539–546

Györgyey J, Gartner A, Németh 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

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

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

Jones D, Russnak RH, Kay RJ, Candido EPM (1986) Structure, expression, and evolution of a heat shock gene locus in *Caenorhabditis elegans* that is flanked by repetitive elements. J Biol Chem 261:12006–12015

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

Kemmerer EC, Lei M, Wu R (1991) Structure and molecular evolutionary analysis of a plant cytochrome *c* gene: surprising implications for *Arabidopsis thaliana*. J Mol Evol 32:227–237

Kramps JA, de Man BM, de Jong WW (1977) The primary structure of the B2 chain of human α-crystallin. FEBS Lett 74:82–84

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

Lauzon LM, Helm KW, Vierling E (1990) A cDNA clone from *Pisum sativum* encoding a low molecular weight heat shock protein. Nucleic Acids Res 18:4274

Lindquist S (1986) The heat-shock response. Annu Rev Biochem 55:1151–1191

McElwain EF, Spiker S (1989) A wheat cDNA clone which is homologous to the 17 kd heat-shock protein gene family of soybean. Nucleic Acids Res 17:1764

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

Nagao RT, Czarnecka E, Gurley WB, Schöffl F, Key JL (1985) Genes for low-molecular-weight heat shock proteins of soybeans: sequence analysis of a multigene family. Mol Cell Biol 5:3417–3428

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

Nieto-Sotelo J, Vierling E, Ho T-HD (1990) Cloning, sequence analysis, and expression of a cDNA encoding a plastid-localized heat shock protein in maize. Plant Physiol 93:1321–1328

Piatigorsky J, Wistow GJ (1989) Enzyme/crystallins: gene sharing as an evolutionary strategy. Cell 57:197–199

Plesofsky-Vig N, Brambl R (1990) Gene sequence and analysis of hsp30, a small heat shock protein of *Neurospora crassa* which associates with mitochondria. J Biol Chem 265:15432–15440

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

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

Takahashi T, Komeda Y (1989) Characterization of two genes encoding small heat-shock proteins in *Arabidopsis thaliana*. Mol Gen Genet 219:365–372

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

Vierling E, Nagao RT, DeRocher AE, Harris LM (1988) A heat shock protein localized to chloroplasts is a member of a eukaryotic superfamily of heat shock proteins. EMBO J 7:575–581

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

Received March 10, 1992/Revised and Accepted May 30, 1992