Phylogenetic and Biochemical Studies Reveal a Potential Evolutionary Origin of Small Heat Shock Proteins of Animals from Bacterial Class A

Xinmiao Fu,[1,2,3] **Wangwang Jiao,**[3] **Zengyi Chang**[1,2]

[1] State Key Laboratory of Protein Engineering and Plant Genetic Engineering, Peking University, Beijing 100871, China
[2] College of Life Science, Peking University, Beijing 100871, China
[3] Department of Biological Sciences and Biotechnology, Tsinghua University, Beijing 100084, China

Abstract. Small heat shock proteins (sHSPs), as one subclass of molecular chaperones, are important for cells to protect proteins under stress conditions. Unlike the large HSPs (represented by Hsp60 and Hsp70), sHSPs are highly divergent in both primary sequences and oligomeric status, with their evolutionary relationships being unresolved. Here the phylogenetic analysis of a representative 51 sHSPs (covering the six subfamilies: bacterial class A, bacterial class B, archae, fungi, plant, and animal) reveals a close relationship between bacterial class A and animal sHSPs which form an outgroup. Accumulating data indicate that the oligomers from bacterial class A and animal sHSPs appear to exhibit polydispersity, while those from the rest exhibit monodispersity. Together, the close evolutionary relationship and the similarity in oligomeric polydispersity between bacterial class A and animal sHSPs not only suggest a potential evolutionary origin of the latter from the former, but also imply that their oligomeric polydispersity is somehow a property determined by their primary sequences.

Key words: Chaperone — Small heat shock protein — Oligomer — Polydisperse — Monodisperse

Introduction

Heat shock proteins (HSPs) are expressed constitutively or/and in response to stress conditions in living organisms, playing important or even essential physiological roles (Lindquist and Craig 1988). The HSPs are commonly categorized into five families: Hsp100, Hsp90, Hsp70, chaperonin (e.g., Hsp60), and small heat shock proteins (sHSPs) (Narberhaus 2002). The structures and biological functions of large HSPs like Hsp90, Hsp70, and Hsp60 are usually highly conserved from prokaryotes to eukaryotes (Bukua and Howrich 1998; Hartl and Hayer-Hartl 2002) and the evolutionary relationships of members in each family have been well elucidated via phylogenetic analyses (Gupta et al. 1993; Gupta 1995; Brocchieri and Karlin 2000).

Unlike the large HSPs, the sHSPs are highly divergent in both primary sequences and oligomeric status (Lindquist and Craig 1988; Vierling 1991; Plesofsky-Vig et al. 1992; Kappe et al. 2000). Although sHSPs are characterized by having a relatively conserved α-crystallin domain spanning about 100 residues, their sequence similarity is far lower than that for the large HSPs (Vierling 1991; Plesofsky-Vig et al. 1992; de Jong 1993). Due to the lack of high sequence similarity, tracing the evolutionary relationships of members of the sHSP family becomes difficult (Vierling 1991; Plesofsky-Vig et al. 1992; Kappe et al. 2000; de Jong et al. 1993, 1998; Waters 1995; Waters et al. 1996). Furthermore, unlike the large HSPs that usually exist as well-defined

Correspondence to: Zengyi Chang; *email:* changzy@pku.edu.cn

oligomers, the sHSPs from different organisms exist as oligomers that vary in size, shape, and subunit inter-action dynamics, with some of them even exhibiting polydispersity (meaning the proteins exist simulta-neously as oligomers of various sizes; reviewed by Narberhaus [2002], Hartl and Hayer-Hartl [2002], Vierling [1991], and Waters [1995]).

This study was carried out in an attempt to under-stand the evolutionary relationships of sHSPs. Our phylogenetic analysis revealed that bacterial sHSPs such as *E.coli* IbpA/IbpB and animal sHSPs each forms an independent monophyletic group that cluster together to form an outgroup, which is highly distinct from the sHSPs of plant, fungi, archaea, and other bacteria. Further analyses demonstrate that these bacterial sHSPs are members of the bacterial class A, according to the reports of Narberhaus and co-work-ers (Studer et al. 2000; Munchbach et al. 1999). Con-sistent with such an evolutionary assignment are the accumulating observations that the bacterial class A and animal sHSPs are often found to exist as polydis-perse oligomers, while the other sHSPs exist as rela-tively well-defined monodisperse oligomers. In view of these revelations, it is hypothesized that the animal sHSPs were evolutionarily originated from the sHSPs of bacterial class A, extending the previous observa-tions that animal sHSPs are linked with some pro-karyote sHSPs (de Jong et al. 1998).

Materials and Methods

Sequence Alignments

All the available amino acid sequences of sHSPs were downloaded from the mirror website of SWISSPROT and TREMBL databases in China (available at http://cn.expasy.org). Partial and duplicated sequences were discarded. The remaining 344 unique and complete sequences, each indicated by the accession number of either SWISSPROT or TREMBL (see Supplementary Table S1), were grouped as bacterial class A, bacterial class B (grouped according to Munchbach et al. 1999), archaea, fungi, plant, animal, and other eukaryotes.

To perform sequence alignment and phylogenetic analysis, a representative of 51 sequences (as shown in Fig. 1) were chosen from the total 344 sequences. The exact number of sequences se-lected from each subfamily for the alignment was roughly pro-portional to its total number of sequences so far accumulated (for instance, 13 of the 115 plant sequences and 11 of the 79 animal sequences were chosen). The criteria in choosing certain particular sequences from each subfamily were to represent as many species and classes as possible. For instance, chosen to represent the higher plant were three, two, two, two, and 2 sequences from classes I, II, and IV, chloroplast, and mitochondria respectively; chosen to represent the lower plants (moss) were 1 sequence each from classes I and II. A partial list of species selected to represent each sub-family is as follows: bacteria—*Mycobacterium tuberculosis, Streptococcus thermophilus, Bradyrhizobium japonicum, and Escherichia coli*; archaea—*Methanococcus jannaschii, Halobacteri-um* sp., *and Sulfolobus solfataricus;* fungi— *Schizosaccharomyces pomb,; Saccharomyces cerevisiae,* and *Laccaria bicolor*; plants—*Arabidopsis thaliana, Lycopersicon esculentum, Pisum sat-*

ivum, and *Funaria hygrometrica*; and animals—human, frog, fish, fruit fly, and nematode.

These 51 proteins were initially subjected to sequence alignment by using the clustalW multiple alignment (Thompson et al. 1994) in BioEdit (version 5.06; Department of Microbiology, North Carolina State University). The α-crystallin domain sequences were first identified from the above alignments by referring to the alignments presented by Saitou and Nei (1987) and then realigned using the clustalx software (version 1.7 [Thompson et al. 1994]) after all gaps were removed (as shown in Fig. 1).

Phylogenetic Analysis

The neighbor joining (NJ) (Saitou and Nei 1987) distance tree (Fig. 2), formulated from the realignment of the α-crystallin do-main sequences of the 51 sHSPs, was evaluated by bootstrap (Felsenstein 1985) (the number of trials is 1000). Phylogenetic trees plotted on the basis of either the full or the N-terminal region sequences were also constructed using the NJ method (Supple-mentary Figs. S1A and S1B). The maximum likelihood (Felsenstein and Churchill 1996) and maximum parsimony phylogenetic trees were also constructed based on the realigned α-crystallin domains using the PROTML and PROTPARS softwares in the phylip package 3.6α, respectively (Figs. 3A and B).

Plasmid Construction and Protein Purification

Rat lens αA-crystallin was expressed from the BL21(DE3) *E. coli* strain harboring a pET20b$^+$ expression vector that carries the encoding cDNA gene and purified according to the method de-scribed before (Bova et al. 1997). Recombinant *M. jannaschii* Hsp16.5 protein was expressed in BL21(DE3) cells harboring the expression vector pET21a carrying the encoding gene and purified according to the method described previously (Kim et al. 1998). Wheat Hsp16.9 was expressed in BL21(DE3) cells harboring the expression vector PJC20 carrying the gene and purified according to methods described before (Lee et al. 1995). The gene for Hsp16.3 was subcloned into the pET-9d expression vector as previously described (Chang et al. 1996). Recombinant Hsp16.3 proteins were overexpressed in *E. coli* BL21 (DE3) host cells and purified as previously described (Fu et al. 2003; Gu et al. 2002).

The gene encoding *E. coli* IbpB was inserted into pALTER-Ex1, expressed in JM109(DE3) cells, with the protein purified via a sequential chromatography operations: first with two ion-exchange columns (resins used, DEAE Sepharose FastFlow and subse-quently Source 15Q; buffer, 50 m*M* Tris–HCl, pH 8.5; elution, 0–0.5 *M* NaCl), then with a hydrophobic interaction column (resin, phenyl-Sepharose HP; buffer, 50 m*M* Tris–HCl, pH 8.5, 0.8 *M* [NH$_4$]$_2$SO$_4$; elution, 0.8–0 *M* [NH$_4$]$_2$SO$_4$), and eventually with a preassembled size-exclusion column of Superose 6 HR 10/30 (buffer, 50 m*M* Na$_3$PO$_4$, pH 7.3, 50 m*M* NaCl).

The purified proteins were dialyzed in deionized water, lyoph-ilized, and stored at −20°C before further analysis. Protein con-centrations were determined using the Bio-Rad Protein Assay.

Blue Native Polyacrylamide Gel Electrophoresis (PAGE)

The blue native PAGE, having a 4–25% linear gradient concen-tration of acrylamide, was prepared as described before (Schagger and von Jagow 1991). Electrophoresis was performed at 4°C at a constant electric voltage: 100 V before the samples entered into the separating gel, increasing to 150 V afterward and for another 2 h. Protein bands were visualized by staining with Coomassie brilliant blue G-250.
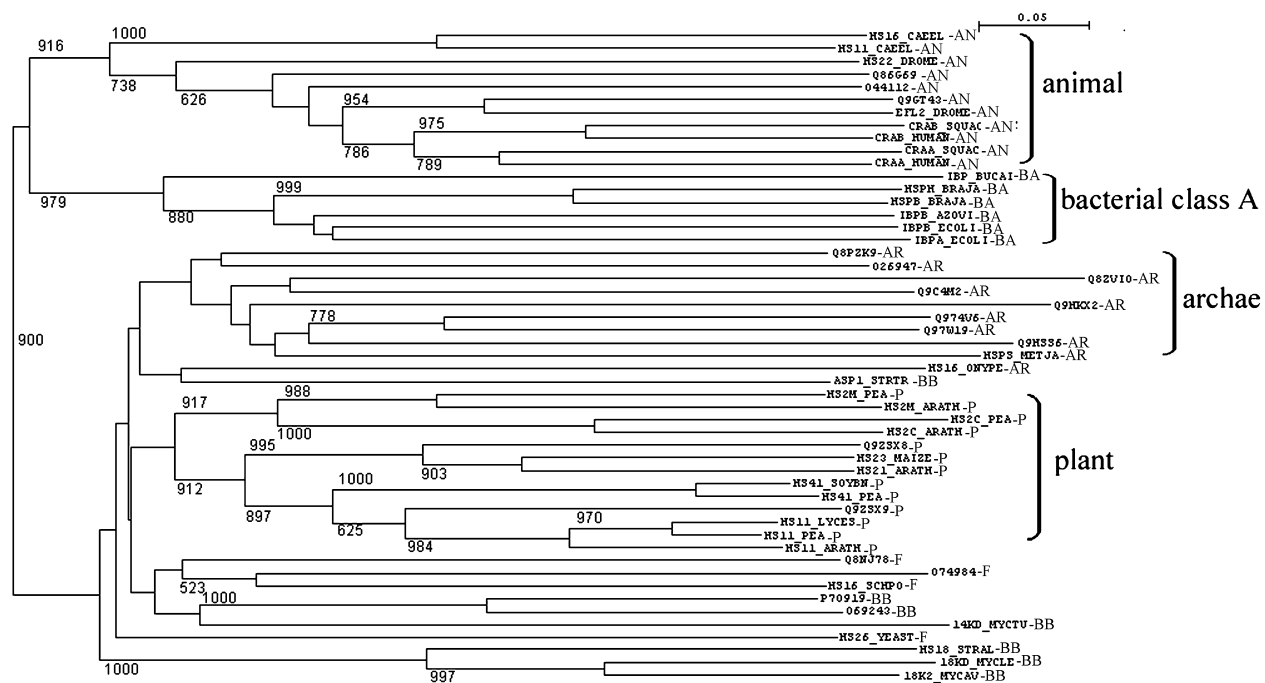
Fig. 1. Sequence alignments of the α-crystallin domains of 51 representative sHSPs. These 51 proteins cover the six subfamilies of sHSPs: bacterial class A, bacterial class B, fungi, archaea, plant, and animal. The number of proteins selected from each subfamily is indicated in brackets. The highly conserved residues are indicated by black backgrounds. The types of secondary structures indicated at the bottom are assigned according to the determined crystal structure of *M. jannaschii* Hsp16.5 (Kim et al. 1998).

Secondary structure regions: region II, region I. Secondary structure elements (bottom): β2, β3, β4, β5, β6, β7, α1, β8, β9.

Groups and sequences:

bacterial class B (8): 14KD_MYCTU, 18K2_MYCAV, 18KD_MYCLE, ASP1_STRTR, HS16_ONYPE, HS18_STRAL, O69243, P70919

bacterial class A (6): IBPA_ECOLI, IBPB_AZOVI, IBPB_ECOLI, IBP_BUCAI, HSPB_BRAJA, HSPH_BRAJA

archaea (9): HSPS_METJA, Q9C4M2, Q9HSS6, Q9HKX2, Q97W19, Q974V6, Q8ZVI0, O26947, Q8PZK9

fungi (4): HS16_SCHPO, HS26_YEAST, Q8NJ78, O74984

plant (13): HS11_ARATH, HS11_PEA, HS11_LYCES, HS21_ARATH, HS23_MAIZE, HS2C_ARATH, HS2C_PEA, HS2M_ARATH, HS2M_PEA, HS41_PEA, HS41_SOYBN, Q9ZSX9, Q9ZSX8

animal (11): CRAA_HUMAN, CRAA_SQUAC, CRAB_HUMAN, CRAB_SQUAC, EFL2_DROME, HS22_DROME, HS11_CAEEL, HS16_CAEEL, O44112, Q86G69, Q9GT43

Consensus, sec.16.5

Size-Exclusion Chromatography

Size-exclusion chromatography (SEC) was performed on a ÄKTA FPLC system connected with a water bath, using an XK 16/75 column that was self-packed with Superdex 200 pregrade medium (all from Amersham Pharmacia Biotech). For each analysis, a protein sample, after being cleaned by centrifugation (10,000g, 10 min), of 500 µl was loaded onto the column. Protein samples were then eluted with 50 mM phosphate sodium buffer (containing 0.15 M NaCl, pH 7.0) at a flow rate of 0.8 ml/min.

Results

Bacterial Class A and Animal sHSPs Are Closely Related as an Evolutionary Outgroup

Optimal alignments for the 51 sequences, essential for the reliable phylogeny reconstruction, were performed by discarding the sequences of the highly variable N-terminal regions and C-terminal extensions, by referring to the previous alignment results of sHSPs (Kappe et al. 2000; Kim et al. 1998), after the first round of alignments based on the full sequences. The realigning results of the sequences of the α-crystallin domains are presented in Fig. 1.

The phylogenetic tree presented in Fig. 2 was constructed by applying the NJ method (Saitou and Nei 1987) to the aligned sequences of the α-crystallin domains (Fig. 1) and evaluated by bootstrapping (Felsenstein 1985). Data displayed in Fig. 2 unambiguously revealed the following phenomenon: the bacteria class A and animal sHSPs cluster together to form an outgroup that is highly distinct from the other sHSPs (the bootstrap score is 900/1000).

0.05

HS16_CAEEL-AN
HS11_CAEEL-AN
HS22_DROME-AN
Q86G69-AN
O44112-AN
Q9GT43-AN
EFL2_DROME-AN
CRAB_SQUAC-AN
CRAB_HUMAN-AN
CRAA_SQUAC-AN
CRAA_HUMAN-AN

animal

IBP_BUCAI-BA
HSPH_BRAJA-BA
HSPB_BRAJA-BA
IBPB_AZOVI-BA
IBPB_ECOLI-BA
IBPA_ECOLI-BA

bacterial class A

Q8P2K9-AR
O26947-AR
Q82VI0-AR
Q9C4M2-AR
Q9HKX2-AR
Q974V6-AR
Q97W19-AR
Q9HS56-AR
HSPS_METJA-AR
HS16_ONYPE-AR

archae

ASP1_STRTR-BB
HS2M_PEA-P
HS2M_ARATH-P
HS2C_PEA-P
HS2C_ARATH-P
Q92SX8-P
HS23_MAIZE-P
HS21_ARATH-P
HS41_SOYBN-P
HS41_PEA-P
Q92SX9-P
HS11_LYCES-P
HS11_PEA-P
HS11_ARATH-P

plant

Q8NJ78-F
O74984-F
HS16_SCHPO-F
P70917-BB
069243-BB
14KD_MYCTU-BB
HS26_YEAST-F
HS18_STRAL-BB
18KD_MYCLE-BB
18K2_MYCAV-BB

916
1000
738
626
954
975
786
789
979
999
880
778
900
988
917
995
1000
903
912
897
625
970
984
523
1000
1000
997

Fig. 2. The neighbor-joining distance tree constructed from the α-crystallin domain. The NJ distance tree was first constructed from the realigned α-crystallin domains of the 51 sHSPs (as shown in Fig. 1) using the program clustalx and then evaluated by bootstrap (Felsenstein 1985) (the number of trials being 1000). The number of bootstrap replicates (of a total of 1000) for a branch to appear is indicated above the branch, with values below 500 not indicated. The scale bar indicates the number of substitutions per site for a unit branch length. Indicated on the right are the sHSPs subfamilies that appear as monophyletic groups. To show the subfamily of the sequence, each name of the sequence is followed by the abbreviation of the subfamily (BA, bacterial class A; BB, bacterial class B; AR, Archae; F, fungi; P, plant; AN, animal), and the same indication is used in the other phylogenetic trees.

Additional data supporting such a close relationship between the bacterial class A and animal sHSPs include the following. First, such a close relationship was also revealed when the complete sequences of the 51 proteins were phylogenetically analyzed (see Supplementary Fig. S1A) or when the phylogenetic tree was constructed based on the method of maximum likelihood (Fig. 3A). The close relationship between the two subfamilies of sHSPs was, however, not revealed in the phylogenetic tree constructed by applying maximum parsimony (Fig. 3B). Similar inconsistency among the three methods for phylogenetic analysis has also been reported previously (Gupta 1995; Nikoh et al. 1994) and it might be attributed to the fact that small errors in distance may disrupt the phylogenetic tree for the maximum parsimony method (Bruno et al. 2000). As a result, when this method was applied in analyzing the sequences of sHSPs which are so diverse between each other, the close evolutionary relationship between the animal and bacterial class A sHSPs failed to be revealed.

Consistent with previous reports (Plesofsky-Vig et al. 1992; de Jong et al. 1998; Munchbach et al. 1999; Waters and Vierling 1999), animal sHSPs (Fig. 2 and Supplementary Figs. S1A and S2) as well as plant sHSPs (Figs. 2 and 3A) were all found to exhibit monophyletic grouping. The bacterial class A sHSPs were also found to exhibit monophyletic grouping (Figures 2 and 3A and Supplementary Fig. S2). Such monophyletic properties for the animal, bacterial class A, and plant sHSPs, however, were not detected in the NJ tree on the basis of the N-terminal regions alone (Fig. S1B), reflecting the extreme variation in amino acid sequences of this region among sHSPs (de Jong et al. 1993). Interestingly, the archaea sHSPs were found to exhibit monophyletic grouping with a marginal tendency in our analysis (Figs. 2 and 3A), suggesting that this subfamily may either have diverged after being evolved from one ancestral gene or have formed by convergent evolution from more than one ancestral genes.

Bacterial Class A and Animal sHSPs Exist as Highly Polydisperse Oligomers, and the Others as Relatively Monodisperse Oligomers

The information for the native folding of one polypeptide chain is in principle encoded in its amino acid sequence (Anfinsen 1973; Dobson and Karplus 1999). The fact that sHSPs commonly exist as oligomers, as well as the close evolutionary relationship between bacteria class A and animal sHSPs (Fig. 2), as deduced from a comparison of their amino acid sequences, all prompted us to pursue whether they also share certain features in their patterns of oligomeri-
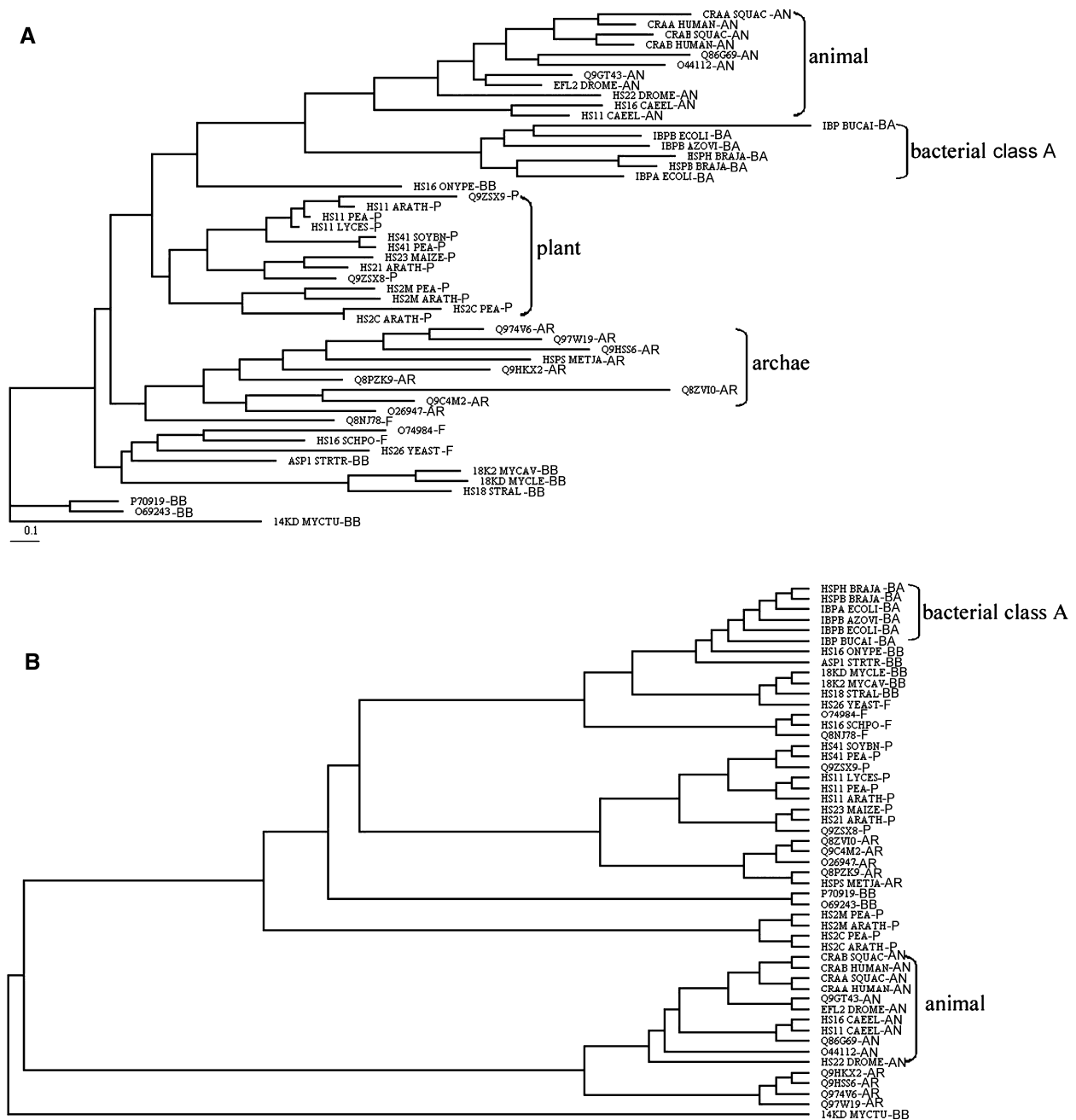
A

B

Fig. 3. Phylogenetic trees constructed on the basis of α-crystallin domains by maximum likelihood (**A**) and maximum parsimony (**B**). The trees were constructed on the basis of the alignment of α-crystallin domains of 51 sHSPs (Fig. 1). (A) The scale bar indicates the number of substitutions per site for a unit branch length. The animal, bacterial class A, archaeal, and plant sHSPs, respectively, which form monophyletic groups are designated. (B) The animal and bacterial class A sHSPs, respectively, which form monophyletic groups are designated.

zation. To our great amazement, the bacterial class A and animal sHSPs were all found to exclusively exist as polydisperse oligomers (i.e., the protein exists simultaneously as oligomers of various sizes), as revealed by previously published data that are summarized in Table 1 (Kim et al. 1998; Lee et al. 1995; Haley et al. 1998, 2000; Vanhoudt et al. 2000; Ehrnsperger et al. 1997, 1999; van de Klundert et al. 1998; Leroux et al. 1997; Sobott et al. 2002; Lee et al. 1997; Suzuki et al. 1998; Haslbeck et al. 1999). In contrast, all the other sHSPs, including those from plants, fungi, archaea, and bacterial class B, exist exclusively as monodisperse oligomers (Table 1).

Partial confirmation of this extremely interesting revelation comes from the data presented in Fig. 4A, which demonstrate that bovine α-crystallin, rat αA-crystallin, and *E. coli* IbpB (a member of bacterial class A) all exist as polydisperse oligomers, while *M. tuberculosis* Hsp16.3 (a member of bacterial class B), *M. jannaschii* Hsp16.5 (archaea), and wheat

Table 1. Polydispersity vs monodispersity for sHSP oligomers

Subfamily	Reported members	Dispersity	Ref(s).
Animal	αA-Crystallin, αB-crystallin, murine Hsp25, human Hsp27, human Hsp20, *C. elegans* HSP16-2, *Artemia franciscana* P26	Poly	Haley et al. (1998, 2000), Vanhoudt et al. (2000), Ehrnsperger et al. (1997, 1999), van de Klundert et al. (1998), Leroux (1997), Sun et al. (2004)
Plant	Wheat Hsp16.9, pea Hsp18.1, pea Hsp17.7, pea Hsp 21, *A. thaliana* Hsp21	Mono[a]	Lee et al. (1995), Sobott et al. (2002), Lee et al. (1997), Suzuki et al. (2002), van Montfort et al. (2001)
Fungi	*S. cerevisiae* Hsp26, *S. cerevisiae* Hsp42	Mono[b]	Haslbeck et al. (1999, 2004a, b)
Archaea	*M. jannaschii* Hsp 16.5	Mono	Kim et al. (1998)
Bacterial class A	*E. coli* IbpB, *E. coli* IbpA, *B. japonicum* HspH	Poly	Veigner et al. (1998), Shearstone et al. (1999), Studer et al. (2002)
Bacterial class B	*M. tuberculosis* Hsp16.3, *B. japonicum* HspF, *Synechocystis* Hsp 16.6	Mono	Munchbach et al.(1999), Chang et al. (1996), Studer et al. (2002)

[a] Plant sHSPs so far examined all appear to be dodecamers.
[b] SEC, native PAGE, and electronic microscopy analysis indicated that Hsp42 existed as a monodisperse oligomer, but blue native PAGE showed the presence of an additional larger oligomer (Haslbeck 2004).

Hsp16.9 exist as monodisperse oligomers. The results presented here are highly consistent with the previous individual analysis for each of these six proteins, using various other methods (Kim et al. 1998; Chang et al. 1996; Haley et al. 1998, 2000; van Montfort et al. 2001; Shearstone and Baneyx 1999).

Subunit Exchange Does Not Occur Between E. coli IbpB (a Bacteria Class A sHSP) and Bovine α-Crystallin (an Animal sHSP)

The homo-oligomeric sHSPs have been found to undergo dynamic dissociation and reassociation (Gu et al. 2002), thus allowing subunit exchange to occur between the oligomers. Similar subunit exchange has also been reported to occur between the sHSPs of different species (Studer and Narberhaus 2000; Bova et al. 1997, 2000, 2002; Sobott et al. 2002; Datta and Rao 2000), indicating the compatibility of interfaces for oligomeric assembly between them. In view of these observations and the close phylogenetic relation between bacterial class A and animal sHSPs (Fig. 2), it was then asked whether sHSPs from the two groups are able to undergo such subunit exchange. For this purpose, bovine α-crystallin and *E. coli* IbpB were selected as representative of each group. Temperature-controlled size exclusion chromatography (SEC) in combination with SDS-PAGE was applied to detect whether hetero-oligomers could be formed between the two sHSPs via subunit exchange. Such hetero-oligomers were not detected after the two proteins were coincubated at 25°C for 3 h (data not shown). To exclude the possibility that subunit exchange might be inefficient at such a low temperature (25°C) (Bova et al. 1997), the two proteins were coincubated at 45°C for 1 hr before applied for the detection of hetero-oligomers, as shown by data presented in Fig. 4B. Although our data demonstrate that bovine α-crystallin and *E. coli* IbpB sHSPs are essentially incompatible for their subunit interfaces, the nonexhaustive nature of subunit exchange results presented here does not allow us to make a conclusion that subunit exchange does not occur between any members of the bacterial class A and animal sHSPs. Our observations are at least consistent with certain previously reported results, in which subunit exchange was not detected to occur between rat α-crystallin and *M. jannaschii* Hsp16.5 (an archaea) (Bova et al. 2000) or even between bacterial class A and class B sHSPs (Studer et al. 2000).

Discussion

Although a phylogenetic linkage between animal sHSPs and certain prokaryote sHSPs was reported (de Jong et al. 1998), the evolutionary relationship of the animal sHSPs and the two classes of bacterial sHSPs is still elusive. The present study based on the sequence comparison of the most conserved α-crystallin domain reveals that the animal sHSPs might have their evolutionary origin from the bacterial class A sHSPs (see Figs. 2 and 3A and supplementary Fig. S1A). Such a claim is further strengthened by the fact that members of these two subfamilies of sHSPs exhibit polydispersity in their oligomeric structures, while sHSPs from the other subfamilies exhibit monodispersity (see Table 1 and Fig. 4A). To our knowledge, this is the first report where a close phylogenetic relationship revealed from sequence comparison is also exhibited at the level of quaternary
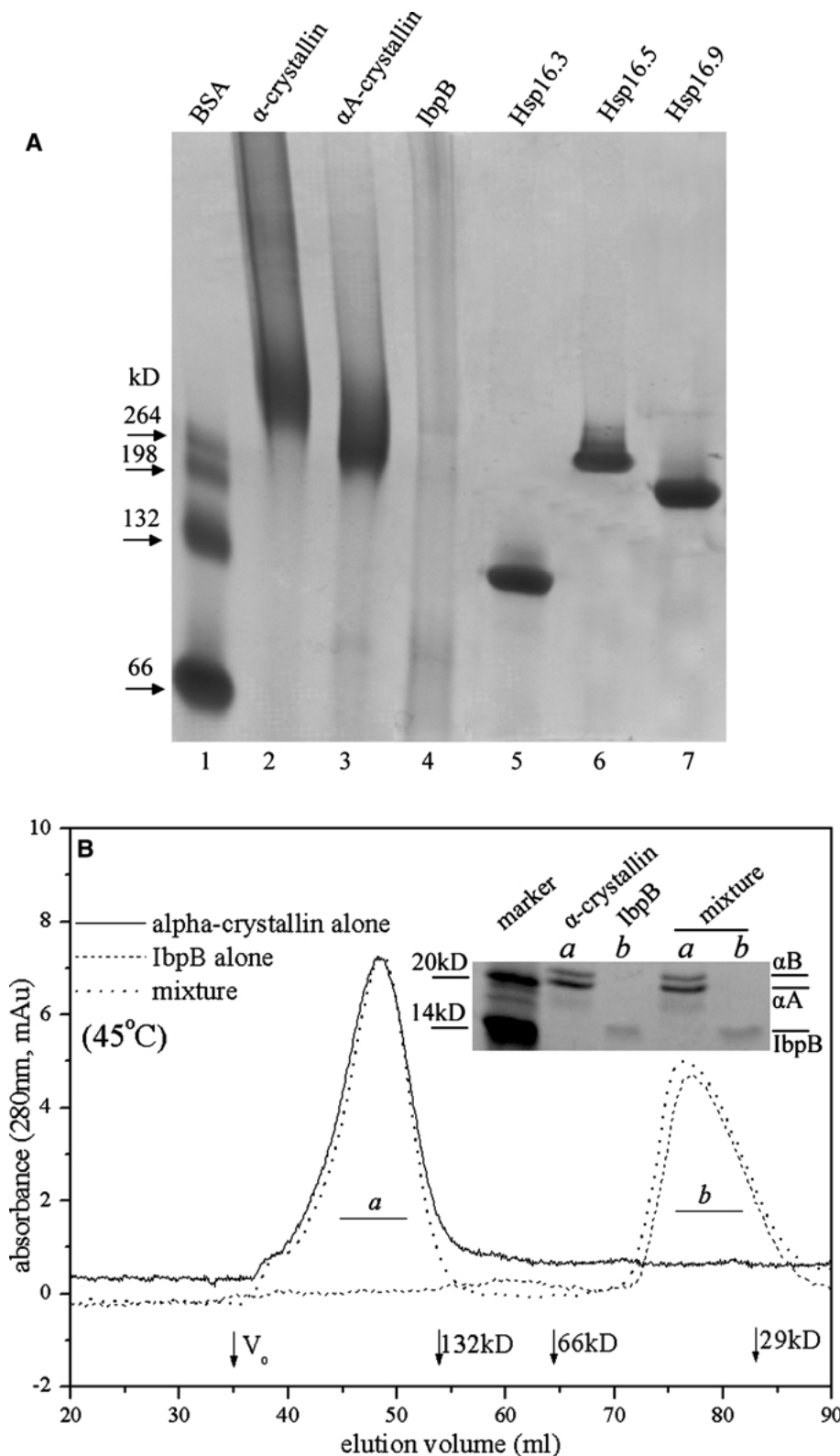
Fig. 4. Characterization of the oligomeric status of representative sHSPs. **A** Blue native pore gradient (4–10%) PAGE analysis results for the following six proteins (with 20 μg protein loaded for each): bovine α-crystallin (lane 2), rat αA-crystallin (lane 3), IbpB (lane 4), *M. tuberculosis* Hsp16.3 (lane 5), *M. jannaschii* Hsp16.5 (lane 6), and wheat Hsp16.9 (lane 7). BSA was used as a marker (lane 1). Electrophoresis was first performed at 100 V and 4°C, then continued at 150 V for another 2 hr after the samples entered the separating gel. **B** The subunit exchange test between α-crystallin and IbpB, examined by temperature-controlled SEC analysis performed at 45°C using a XK 16/70 column self-packed with Superdex 200 prep grade medium. The α-crystallin (2 mg/ml bovine) or *E. coli* IbpB (0.5 mg/ml) protein or their mixture was preheated at 45°C for 1 h before being loaded for chromatography analysis. *Inset* SDS-PAGE analysis results for fractions (corresponding to peaks *a* and *b*) collected from chromatography (concentrated by trichloroacetic acid precipitation before being applied for electrophoresis).

structure. Given that the information for the native tertiary structure of one protein is encoded in its amino acid sequence according to the Anfinsen principle [Anfinsen 1973]), a reasonable hypothesis is that the information for the oligomeric assembly of a protein (i.e., for the quaternary structure) is also

encoded in its amino acid sequence. This hypothesis is apparently supported by our data that the close evolutionary relationship between bacterial class A and animal sHSPs as revealed by analysis of their amino acid sequences is also reflected at the quaternary structure level, by exhibiting the polydispersity in their

oligomers, distinguishing themselves from the other sHSP subfamilies, which all exhibit monodispersity.

However, it is very difficult to experimentally determine the amino acids being involved in the oligomeric polydispersity. We thus explored an alternative logic strategy as following: if the similarity in the oligomeric polydispersity between bacterial class A and animal sHSPs is indeed correlated with their close relation in primary sequences, then amino acids showing a close relation, similarly to the α-crystallin domain, can be reasonably considered to contribute to the oligomeric polydispersity. Previous studies (Kappe et al. 2000; Waters et al. 1996) reported that region I and region II (as indicated at the top of Fig. 1) are conserved in all animal and in most plant sHSPs, respectively. Our phylogenetic analysis results on the basis of each region (I or II), as shown in Supplementary Fig. S2, indicate that region II seems to support the close relation between animal and bacterial class A better than region I does (compare the tree in Fig. 2 with those in Supplementary Figs. S2A and S2B). This observation suggests that region II may play a more important role than region I for these two subfamilies of sHSPs to exhibit oligomeric polydispersity. Structure determination studies of *M. jannaschii* Hsp16.5, on the other hand, indicate that region II is primarily involved in dimerization, and region I in higher oligomerization (Kim et al. 1998). These observations seem to imply that the dimerization interfaces of such sHSPs also play a role for them to exist as polydisperse oligomers. Interestingly, the monophyletic grouping of plant sHSPs is also observed in the evolutionary tree constructed from region II (Supplementary Fig. S2B) but not that from region I (Supplementary Fig. S2A). This might result from the higher diversity of region II compared to region I, allowing a better distinction of subfamilies when sequences of region II are applied for such evolutionary analysis.

If animal sHSPs evolved from a bacterial class A ancestor, it might be assumed that this evolutionary event occurred via lateral gene transfer. The ancestral sHSP gene might have transferred to the animal nuclear genome from the genomes of either an endosymbiotic mitochondria or a bacterial pathogen. The subsequent rapid duplication and divergence of the transferred gene may allow animal sHSPs to become highly divergent from the ancestral gene, as well as from each other (Plesofsky-Vig et al. 1992; de Jong et al. 1993, 1998). Given that there is a second subfamily of bacterial sHSPs (class B), and in view of the close ecological relationships between plant and Rhizobia, a similar logic would suggest that the plant sHSPs might have been evolved from those present in certain bacteria like Rhizobia. The sequences of sHSPs presently available for Rhizobia, however, do not allow us to test such a hypothesis.

Given that functionally important structural features are often highly conserved for proteins during evolution, our data strongly implicate that the exhibition of polydispersity for the bacterial class A and animal sHSPs may somehow allow them to act, for example, in responding to stressful conditions and in forming complexes with denatured substrate proteins, differently from those sHSPs that exhibit monodispersity. The details of this aspect merit further exploration. In addition, does the conclusion that oligomers of bacterial class A and animal sHSPs exhibit polydispersity and those of other subfamilies exhibit monodispersity hold true in general? The tests await further examination of oligomeric structures for more sHSPs.

Supplementary Materials

For manuscript named "Phylogenetic and biochemical studies reveal a potential evolutionary origin of animal small heat shock proteins from bacterial class A" by Fu *et al.*

References

Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181:223–230

Bova MP, Ding LL, Horwitz J, Fung BK-K (1997) Subunit Exchange of alpha-crystallin. J Biol Chem 272:29511–29517

Bova MP, Mchaourab HS, Han Y, Fung BK-K (2000) Subunit exchange of small heat shock proteins. J Biol Chem 275:1035–1042

Bova MP, Huang Q, Ding L, Horwitz J (2002) Subunit exchange, conformational stability, and chaperone-like function of the small heat shock protein 165 from Methanococcus jannaschii. J Biol Chem 277:38468–38475

Brocchieri L, Karlin S (2000) Conservation among Hsp60 sequences in relation to structure, function and evolution. Protein Sci 9:476–486

Bruno WJ, Socci ND, Halpern AL (2000) Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. Mol Biol Evol 17:189–197

Bukau B, Horwich AL (1998) The Hsp70 and Hsp60 chaperone machines. Cell 92:351–366

Chang Z, Primm TP, Jakana J, Lee IH, Serysheva I, Chiu W, Gilbert HF, Quiocho FA (1996) Mycobacterium tuberculosis 16-kDa antigen (Hsp163) functions as oligomeric structure in vitro to suppress thermal aggregation. J Biol Chem 271:7218–7224

Datta SA, Rao CM (2000) Packing-induced conformational and functional changes in the subunits of alpha-crystallin. J Biol Chem 275:41004–41010

de Jong WW, Leunissen JAM, Voorter CEM (1993) Evolution of the a-crystallin/small heat-shock protein family. Mol Biol Evol 10(1):103–126

de Jong WW , Caspers GJ, Leunissen JA (1998) Genealogy of the alpha-crystallin—small heat-shock protein superfamily. Int J Biol Macromol 22:151–162

Dobson CM, Karplus M (1999) The fundamentals of protein folding: bringing together theory and experiment. Curr Opin Struct Biol 9:92–101

Ehrnsperger M, Graber S, Gaestel M, Buchner J (1997) Binding of non-native protein to Hsp25 during heat shock creates a reservoir of folding intermediates for reactivation. EMBO J 16:221–229

Ehrnsperger M, Lilie H, Gaestel M, Buchner J (1999) The dynamics of Hsp25 quaternary structure. J Biol Chem 274:14867–14874

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

Felsenstein J, Churchill GA (1996) A hidden Markov model approach to variation among sites in rate of evolution. Mol Biol Evol 13(1):93–104

Fu X, Li W, Mao Q, Chang Z (2003) Disulfide bonds convert small heat shock protein Hsp163 from a chaperone to a non-chaperone: implications for the evolution of cysteine in molecular chaperones. Biochem Biophys Res Commun 308:627–635

Gu L, Abulimiti A, Li W, Chang Z (2003) Monodisperse Hsp163 nonamer exhibits dynamic dissociation and reassociation, with the nonamer dissociation prerequisite for chaperone-like activity. J Mol Biol 319:517–526

Gupta RS (1995) Phylogenetic analysis of the 90 kD heat shock family of protein sequences and an examination of the relationship among animals, plants, and fungi species. Mol Biol Evol 12:1063–1073

Gupta RS, Golding B (1993) Evolution of Hsp70 gene and its implications regarding relationships between archaebacteria, eubacteria, and eukyrotes. J Mol Evol 37:573–582

Haley DA, Horwitz J, Stewart PL (1998) The small heat-shock protein, alphab-crystallin, has a variable quaternary structure. J Mol Biol 277:27–35

Haley DA, Bova MP, Huang QL, Mchaourab HS, Stewart PL (2000) Small heat-shock protein structures reveal a continuum from symmetric to variable assemblies. J Mol Biol 298:261–272

Hartl FU, Hayer-Hartl M (2002) Molecular chaperones in the cytocol: from nascent chain to folded protein. Science 295:1852–1858

Haslbeck M, Walke S, Stromer T, Ehrnsperger M, White H, Chen S, Raibil H, Buchner J (1999) Hsp26: a temperature-regulated chaperone. EMBO J 18:6744–6751

Haslbeck M, Braun N, Stromer T, Richter B, Model N, Weinkauf S, Buchner J (2004a) Hsp42 is the general small heat shock protein in the cytosol of Saccharomyces cerevisiae. EMBO J 23:638–649

Haslbeck M, Ignatiou A, Saibil H, Helmich S, Frenzl E, Stromer T, Buchner J (2004b) A domain in the N-terminal part of Hsp26 is essential for chaperone function and oligomerization. J Mol Biol 343:445–455

Kappe G, Leunissen JA, de Jong WW (2000) Evolution and diversity of prokaryotic small heat shock proteins. Prog Mol Subcell Biol 28:1–17

Kim KK, Kim R, Kim SH (1998a) Crystal structure of a small heat-shock protein. Nature 394:595–599

Kim KK, Yokota H, Kim S-H (1998b) Purification, crystallization and preliminary X-ray crystallographic analysis of small heat shock protein homolog from Methanococcus jannaschii. J Struct Biol 121:76–80

Lee GJ, Pokala N, Vierling E (1995) Structure and in vitro molecular chaperone activity of cytosolic small heat shock proteins from pea. J Biol Chem 270:10432–10438

Lee GJ, Roseman AM, Saibil HR, Vierling E (1997) A small heat shock protein stably binds heat-denatured model substrates and can maintain a substrate in a folding-competent state. EMBO J 16:659–671

Leroux MR, Melki R, Gordon B, Batelier G, Candido EPMJ (1997) Structure-function studies on small heat shock protein oligomeric assembly and interaction with unfolded polypeptides. J Biol Chem 272:24646–24656

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

Munchbach M, Nocker A, Narberhaus F (1999) Multiple small heat shock proteins in Rhizobia. J Bacteriol 181:83–90

Narberhaus F (2002) Alpha-crystallin-type heat shock proteins: socializing minichaperones in the context of a multichaperone network. Microbiol Mol Biol Rev 66(1):64–93

Nikoh N, Hayase N, Iwabe N, Kuma K, Miyata T (1994) Phylogenetic relationship of the kingdoms animalia, plantae and fungi, inferred from 23 different protein sequences. Mol Biol Evol 11:762–768

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

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

Schagger H, von Jagow G (1991) Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. Anal Biochem 199:223–231

Shearstone JR, Baneyx F (1999) Biochemical characterization of the small heat shock protein IbpB from Escherichia coli. J Biol Chem 274:9937–9945

Sobott F, Benesch JLP, Vierling E, Robinson CV (2002) Subunit exchange of multimeric protein complexes. J Biol Chem 277:38921–38929

Studer S, Narberhaus F (2000) Chaperone activity and homo- and hetero-oligomer formation of bacterial small heat shock proteins. J Biol Chem 275:37212–37218

Studer S, Obrist Lentze M N, Narberhous F (2002) A critical motif for oligomerization and chaperone activity of bacterial a-heat shock proteins. Eur J Biochem 269:3578–3586

Sun Y, Mansour M, Crack JA, Gass GL, MacRae TH (2004) Oligomerization, chaperone activity, and nuclear localization of p26, a small heat shock protein from Artemia franciscana. J Biol Chem 279:39999–40006

Suzuki TC, Krawitz DC, Vierling E (1998) The chloroplast small heat-shock protein oligomer is not phosphorylated and does not dissociate during heat stress in vivo. Plant Physiol 116:1151–1161

Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

van de Klundert FA , Smulders RH, Gijsen ML, Lindner RA, Jaenicke R, Carver JA, de Jong WW (1998) The mammalian small heat-shock protein Hsp20 forms dimers and is a poor chaperone. Eur J Biochem 258:1014–1021

Vanhoudt J, Abgar S, Alerts T, Clauwaert J (2000) Native quaternary structure of bovine alpha-crystallin. Biochemistry 39:4483–4492

van Montfort RLM , Basha E, Fieldrich KL, Viering E (2001) Crystal structure and assembly of a eukaryotic small heat shock protein. Nature Struct Biol 8:1025–1030

Veigner L, Diamant S, Buchner J, Goloubinoff P (1998) The small heat-shock protein IbpB from Escherichia coli stabilizes stress-denatured proteins for subsequent refolding by a multichaperone network. J Biol Chem 273:11032–11037

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

Waters ER (1995) The molecular evolution of the small heat-shock proteins in plants. Genetics 141:785–795

Waters ER, Vierling E (1999) The diversion of plant cytosolic small heat shock proteins preceded the divergence of mosses. Mol Biol Evol 16:127–139

Waters ER, Lee G, Vierling E (1996) Evolution, structure and function of the small heat shock proteins in plants. J Exp Bot 47:325–338