Evolution and Diversity of Prokaryotic Small Heat Shock Proteins

Guido Kappé[1], Jack A.M. Leunissen[2], and Wilfried W. de Jong[1]

1
Introduction

To understand the evolutionary mechanisms that led to the diversification of the various types of heat shock proteins(Hsps) and their functioning in multichaperone networks is a great challenge (Feder and Hofmann 1999). Considerable information is already available on the evolution of the Hsp60 and Hsp70 families (e.g., Gupta 1995; Budin and Philippe 1998; Karlin and Brocchieri 1998; Macario et al. 1999; Archibald et al. 2000; Brocchieri and Karlin 2000). Relatively less is known about the early evolution of the small heat shock proteins (sHsps), which are considerably more divergent in structure and function than the Hsp60s and Hsp70s.

The sHsps are found in bacteria, archaea and eukaryotes. They range in monomer size between 12 and 43 kDa, and are characterized by a conserved "α-crystallin" domain of about 80 residues (reviewed in Plesofsky-Vig et al. 1992; Caspers et al. 1995; Waters and Vierling 1999). The N-terminal domains and C-terminal extensions, if present, are not conserved in sequence and length. The sHsps generally form high molecular weight complexes, between 150 and 800 kDa in size, although some occur as dimers or tetramers. The only published crystal structure of any sHsp is as yet that of *Methanococcus jannaschii* Hsp16.5 (Kim et al. 1998). This archaeal sHsp forms a hollow sphere, ~120 Å in outer diameter, and composed of 24 subunits with an Ig-like β-sandwich folding. The characteristic property of most sHsps is their ability to suppress the in vitro aggregation of denaturing proteins, while in vivo their expression protects cells during stress (Ehrnsperger et al. 1997b). For detailed information about the broad variety of structural and functional properties of the sHsps we refer to the other chapters in this Volume.

It is the purpose of this chapter to describe the evolutionary diversity of prokaryotic sHsps. In terms of sequence divergence, this diversity is broader than that amongst plant or animal sHsps (de Jong et al. 1998; Waters and

[1] Department of Biochemistry, University of Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands
[2] Centre for Molecular and Biomolecular Informatics, University of Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands

Vierling 1999). In fact, both plant and animal sHsps are suggested to be monophyletic groupings, which may have originated separately from different ancestral prokaryotic sHsps. Knowing the evolutionary history of the prokaryotic sHsps, and comparing their properties, may also help to understand the structure-function relationships of eukaryotic sHsps. This will provide an insight in the pathways along which they diversified, and how they acquired their characteristic structural and functional properties.

2
sHsps in Prokaryotes

Most prokaryotic sHsps function as chaperone-like proteins in the cytoplasm, but some have become structural components of the spore coat, as for *Bacillus subtilis* COTM (Henriques et al. 1997), while others associate with membranes, like *Mycobacterium tuberculosis* Hsp16.3 (Cunningham and Spreadbury 1998) and *Oenococcus oenii* Hsp (Jobin et al. 1997). In addition to rigidifying and stabilizing membranes under heat stress, such associations may constitute a feedback mechanism for the regulation of heat shock genes, as has been proposed for *Synechocystis sp.* SP21 (Horváth et al. 1998). As in eukaryotes, many prokaryotic sHsps are developmentally regulated and inducible by heat and other stress. Their expression can reach high levels, e.g., up to 22% of total protein for *Streptococcus thermophilus* Hsp16.4 (González-Márques et al. 1997). This Hsp16.4 and other sHsps are plasmid-encoded, which may help to achieve such high expression levels.

There is good evidence that sHsps facilitate the refolding of denatured proteins in conjunction with other chaperones (Ehrnsperger et al. 1997b; Lee et al. 1997). Indeed, *Escherichia coli* IBPB interacts in a multichaperone network with Hsp60 and Hsp70 (Veigner et al. 1998). Such a functional relation is also suggested by a heat-shock cluster comprising genes for sHsps and GroES in *Bradyrhizobium japonicum* (Narberhaus et al. 1996), by an operon encoding an sHsp as well a DnaK in *Thermotoga maritima* (Michelini and Flynn 1999), and by an operon encoding DnaK and two sHsps in *Porphyromonas gingivalis* (Yoshida et al. 1999). If such multichaperone networks are a universal requirement for cellular functioning, one might expect to find the essential components in all organisms. Yet, the gene for Hsp70 is lacking in several Archaea (Macario et al. 1999). Although it is generally taken for granted that sHsps occur in all organisms, this remains to be established. The availability of a rapidly increasing number of complete prokaryote genome sequences now makes it possible to determine whether this is indeed the case. This also enables us to assess the number and variety of sHsp genes in different bacteria and archaea, as described in the following sections.

3
Are sHsps Dispensable in Some Pathogenic Bacteria?

We searched the complete, known genomes of 15 bacteria and 4 archaea for sequence coding for sHsps. These genomes were selected to represent a broad variety of prokaryotic lineages. For comparison, we also included the genome of the eukaryote *Saccharomyces cerevisiae*. To minimize the chance of overlooking sHsp-related sequences, five different search profiles were used, based on the α-crystallin domain of known sHsps. Four of these profiles were calculated from a previously published alignment of the α-crystallin domains of a broad array of sHsps (de Jong et al. 1998). The alignment was divided into two subalignments, one comprising the animal sHsps and the other all remaining sHsps, because considerable differences were observed between these two groups of sequences. To calculate the profiles, two different programs were used on each subalignment, PROFILEMAKE v4.4 (Wisconsin Package Version 8.1) and PROFILEWEIGHT v2.1 (Thompson et al. 1994a). These profiles are available from the CMBI FTP server (ftp://ftp.cmbi.kun.nl/pub/molbio/-kappe/). The fifth profile was downloaded from the PROSITE database (accession no. PS01031). This profile has been calculated using 140 sHsps from a wide range of organisms, including animals, plants, fungi and prokaryotes.

The completely sequenced genomes that were used in the search for sHsps are listed in bold in Table 1. All open reading frames (ORFs) that code for at least 50 amino acids were extracted from these genomes. These ORFs were taken as a database that was searched with the five sHsp profiles as a query, using Compugen's implementation of the PROFILESEARCH program (GenCore users manual, Compugen). Only those ORFs with a similarity over at least 60 positions, with fewer than 11 gaps in the region of similarity, and with more than 25% identity to a profile were considered as possible sHsps. These ORFs were compared with the SWISSPROT database, using Compugen's implementation of the FASTA program (GenCore users manual, Compugen), and were only considered as genuine sHsps when the best scoring proteins turned out to be known sHsps.

Interestingly, no sHsp-related sequences could be detected in eight of the bacterial genomes (indicated by n.d. in Table 1). In four bacterial genomes a single sHsp gene was found, *E. coli* and *M. tuberculosis* have two copies, and *B. subtilis* three. Three of the archaeal genomes had one sHsp gene, the fourth (*Archaeoglobus fulgidus*) had two copies. In yeast, as expected, only the genes for the known Hsp26 and Hsp42 were found. All 16 detected prokaryotic sHsp-related sequences are present in the protein databases, although *Pyrococcus horikoshii* 172aa had not been recognized as an sHsp; this demonstrates the sensitivity of our search procedure.

Since sHsp genes have been reported to also reside on bacterial plasmids (see Table 1), we similarly searched the plasmids of *Borrelia burgdorferi*, the extra-chromosomal element 1 of *Aquifex aeolicus*, and the small and large

Table 1. Prokaryotic sHsps found in completely sequenced genomes and in protein databases

Species[a]	sHsp[b]	Accession number	Subunit size[c]	Complex size[d]	Functional properties[e]	References[f]
Bacteria						
Aquifex aeolicus (1.55)	HspcI	O67316	144/17.2			Deckert et al. (1998)
Azotobacter vinelandii	IBPB	P96193	147/16.3			
Bacillus subtilis (4.21)	COTM	Q45058	130/15.2			Kunst et al. (1997) ; Henriques et al. (1997)
	YDFT	P96698	143/17.0			
	YOCM	O34321	158/18.4			
Borrelia burgdorferi (1.23)	n.d.					Fraser et al. (1997)
Bradyrhizobium japonicum	HspA	P70917	152/17.2		ind+	Narberhaus et al. (1996)
	HspB	P70918	153/17.2		th-, ind+	Narberhaus et al. (1996)
	HspC	P70919	166/18.6		th-, ind+	Narberhaus et al. (1996)
	HspD	O69241	151/17.3		ind+	Münchbach et al. (1999)
	HspE	O69242	150/17.2		ind+	Münchbach et al. (1999)
	HspF	O69243	163/18.6		ind+	Münchbach et al. (1999)
	HspH	O86110	151/17.1		ind+	Münchbach et al. (1999)
Buchnera aphidicola	HspA*	O31288	153/18.0			van Ham et al. (1997)
Chlamydia pneumoniae (1.23)	n.d.					Kalman et al. (1999)
Chlamydia trachomatis (1.04)	n.d.					Kalman et al. (1999)
Clostridium acetobutylicum	Hsp18	Q03928	151/17.7		ind+	Bahl et al. (1995)
Escherichia coli (4.64)	IBPA	P29209	137/15.8		th-, ind+	Blattner et al. (1997); Thomas and Baneyx (1998)
	IBPB	P29210	142/16.1	600*	ch+, th-, ind+	Veigner et al. (1998); Thomas and Baneyx (1998); Shearstone and Baneyx (1999)
Haemophilus influenzae Rd (1.83)	n.d.					Fleischmann et al. (1995)
Helicobacter pylori 26695 (1.67), J99 (1.64)[g]	n.d.					Alm et al. (1999)
Lactobacillus delbrueckii	Hsp	P94867	123/13.6			

Organism	Protein	Accession	Residues/kDa	Oligomer	Expression	References
Legionella pneumophila	GSPA	S49042	166/19.0		ind+	Abu Kwaik et al. (1997)
Mycobacterium avium	18 kDa1	P46729	147/16.5			
	18 kDa2	P46731	138/15.9			
Mycobacterium intracellulare	18 kDa1	P46730	149/16.5			
	18 kDa2	P46732	140/15.7			
Mycobacterium leprae	Hsp167	P12809	148/16.7		ind+	Nerland et al. (1988); Booth et al. (1993)
	18 kDa[h]	E980239	147/16.5			
Mycobacterium tuberculosis (4.41)	Hsp163	P30223	143/16.1	145 (9)*	ch+, ind+	Cole et al. (1998); Chang et al. (1996); Yang et al. (1999)
	Hsp20p	O53673	159/17.8			Fraser et al. (1995)
Mycoplasma genitalium (0.58)	n.d.					Himmelreich et al. (1996)
Mycoplasma pneumoniae (0.82)	n.d.					
Oenococcus oenii	Hsp	P94898	148/16.9		ind+	Guzzo et al. (1997)
Rickettsia prowazekii (1.11)	Hsp22	E71682	163/18.7			Andersson et al. (1998)
Stigmatella aurantiaca	SP21	Q06823	188/21.3	(2)	ind+	Heidelbach et al. (1993)
Streptococcus thermophilus	ASP	P80485	142/16.4		ind+	González-Márquez et al. (1997)
	Hsp164*	O30851	142/16.4		th–, ind+	Somkuti et al. (1998)
	1SThsp*	O52190	150/17.4		ind+	O'Sullivan et al. (1999)
	SEChsp*[h]	O52192	142/16.4		ind+	O'Sullivan et al. (1999)
Streptomyces albus G	Hsp18	Q53595	143/16.1	Granules	th+, ind+	Servant and Mazodier (1996)
Synechococcus vulcanus	HspA	O82825	145/16.7	400 (24)	ch+, ind+	Roy et al. (1999)
Synechocystis sp. (3.57)	SP21	P72977	146/16.6		th+, ind+	Kaneko et al. (1996); Horváth et al. (1998)
Thermotoga maritima (1.86)[i]	Hsp	Q9ZFD1	142/17.0	400–450 (23–26)*	ch+	Nelson et al. (1999); Michelini and Flynn (1999)
Treponema pallidum (1.14)	n.d.					Fraser et al. (1998)
Archaea						
Archaeoglobus fulgidus (2.18)	Hsp201	O28973	174/20.4			Klenk et al. (1997)
	Hsp202	O28308	140/16.5			Klenk et al. (1997)
Methanobacterium thermoautotrophicum (1.75)	HspcI	O26947	145/17.1			Smith et al. (1997)
Methanococcus jannaschii (1.66)	Hsp165	Q57733	147/16.5	400 (24)*	ch+	Bult et al. (1996); Kim et al (1998)
Pyrococcus horikoshii (1.74)	172aa	O59514	172/20.1			Karawabayasi et al. (1998)