**[Frontiers in Bioscience 9, 1318-1332, May 1, 2004]**

# EVOLUTION OF ASSISTED PROTEIN FOLDING: THE DISTRIBUTION OF THE MAIN CHAPERONING SYSTEMS WITHIN THE PHYLOGENETIC DOMAIN ARCHAEA

**Alberto J. L. Macario [1, 2], Mona Malz [1, 3], and Everly Conway de Macario [1, 2]**

*[1]Wadsworth Center, New York State Department of Health, Division of Molecular Medicine; and [2]Department of Biomedical Sciences, School of Public Health, The University at Albany (SUNY), Albany, New York, USA, [3] Current address: Johannes Gutenberg University, Mainz, 65321 Heidenrod, Germany*

## TABLE OF CONTENTS



## 1. ABSTRACT

Newly made proteins must achieve a functional shape, the native configuration, before they can play their physiological roles in the cell. Proteins must also travel to the locale (*e.g.*, the mitochondrion) in the cell where their functions are required. In these processes of folding into the native configuration and translocation to the place of work, proteins may be assisted by molecules called molecular chaperones. Stressors can unfold (denature) proteins, and genetic defects can cause misfolding and, in addition, both abnormalities can lead to polypeptide aggregation. Chaperones play a role in assisting refolding of partially denatured or misfolded proteins, thus preventing aggregation. Clearly, molecular chaperones are key cell components under normal, physiological circumstances, as well as in potentially harmful situations resulting from environmental or inherited factors. Hence, molecular chaperones constitute attractive targets for a variety of efforts aiming at improving the cell's performance, particularly under stress, to prevent disease, or at least to slow down its progression and to contain the deleterious effects of stress. In our efforts in this direction, we have undertaken to investigate the chaperoning systems of cells belonging to the phylogenetic domain Archaea. The findings reported here pertain to the distribution of the molecular chaperone machine, the chaperonins, and the prefoldins, among archaea. The genes *hsp70(dnaK)*, *hsp40(dnaJ)*, and *grpE* encoding the components of the molecular chaperone machine were present only in some archeaeal species: this contrasts with bacteria and eucarya, which do have the genes with no known exception. The group I, or bacterial, chaperonin-genes *groEL* and *groES* occured in the genomes of *Methanosarcina* species but were not found in any of the other archaea whose genomes have been sequenced. While all the archaea studied had between one and three chaperonins of group II (thermosome subunits), *Methanosarcina acetivorans* was exceptional since it had five of these chaperonins. This is the largest number of group II chaperonins ever found in a prokaryote. Furthermore, two of the *M. acetivorans* chaperonins were different from, albeit related to, the other known archaeal and eucaryal chaperonins of group II. Prefoldins were found in all archaea examined. Overall, the results provide clues to the evolution of the chaperoning systems, which must have played a critical role in survival since life started. Also, the data suggest new avenues of research for elucidating the evolution of assisted protein folding and for uncovering roles and interactions not yet described for these molecules.

## 2. INTRODUCTION

Protein production includes synthesis and folding of the new polypeptides, which yields the final products with a functional shape. The latter is termed the native configuration and is achieved through a series of steps of varying complexity depending on the organism and the type of protein being produced (1). The folding and translocation of many proteins are assisted by molecular chaperones (1,2-5). Several chaperones have been described; among the best studied are the GroEL/S complex, archaeal thermosome, eukaryotic CCT, prefoldins, and the components of the molecular chaperone machine (6-19).

The molecular chaperone machine in prokaryotes is composed of three key molecules: Hsp70(DnaK), Hsp40(DnaJ), and GrpE. This machine occurs in all bacteria, eukaryotic-cell organelles of bacterial ancestry,

and some archaea (20-27). In the cytosol of eukaryotic cells, the functions of GrpE are carried out by other proteins (17,28-34). An important question still under investigation concerns the actual distribution of the machine among organisms of the phylogenetic domain Archaea. The present work addresses this question.

Other points dealt with in this work are the distribution among the Archaea of the chaperonins of group I, GroEL and GroES, the chaperonins of group II or thermosome subunits, and the prefoldins.

All of these molecular chaperones play important roles in the normal physiology of the cell; most importantly, they are part of the cellular anti-stress mechanisms (9,35,36). A variety of stressors cause cell stress, whose central consequence is protein denaturation, namely protein unfolding. The molecular chaperoning systems intervene to prevent denaturation, to restore the native configuration of proteins reversibly unfolded by stress, and to degrade those proteins that have been irreversibly damaged (29,36-38).

Most likely, the chaperoning systems have played a crucial role in evolution, and still play it today as they maintain cellular integrity and health. It is, therefore, of great interest to learn about all aspects of these systems: they offer opportunities for developing means to improve cellular performance under stress, and survival. One approach is to investigate how organisms in a variety of environments deal with stress, and to elucidate the components of their chaperoning systems. This approach should reveal what structures and mechanisms the various extant cell types, which live in various ecosystems differing widely in temperature, pH, salt concentration, barometric pressures, population density, etc., have evolved to counteract the effects of stress and survive. In turn, this information should help in the development of new anti-stress mechanisms, and in the improvement of existing ones, through manipulation of pertinent molecular chaperone genes and their products.

Research in various laboratories over the last few years has shown that molecular chaperones participate in a number of physiological processes above and beyond those strictly pertinent to the folding of nascent polypeptides and to the refolding of partially denatured proteins. In parallel, and as a consequence of the uncovering of the multiple roles of chaperones, it is becoming evident that defective chaperone molecules may cause disease, or at least contribute to pathogenesis (reviewed in 39). Thus, the more we understand about the evolution and function of the chaperoning systems, the easier it will be to understand their role in health and disease, and to addres the problems caused by their failure.

The main goals of the work reported here were to elucidate the distribution of the chaperoning systems and their components among organisms of the phylogenetic domain Archaea, correlate the findings with the organisms' optimal temperatures for growth, and compare the findings with data on the equivalent systems from representatives of the other two domains, Bacteria and Eucarya.

## 3. ORGANISMS STUDIED

The lists of organisms studied belonging to the phylogenetic domains Archaea and Bacteria are displayed in tables 1, 2, and 3, containing Euryarchaeota, Crenarchaeota, and bacteria, respectively. Pertinent physiological information, *i.e.*, optimal temperature for growth (OTG) is shown for the archaeal and bacterial species examined. Also, the pH optimum for growth is indicated for the archaeal organisms, since it is important, along with the OTG, for demonstrating the organisms' degree of "extremophilicity" (or lack of it). The tables also include the genome size, for those organisms whose genomes have been sequenced, and the methods that were applied to determine the occurrence of the molecular chaperone genes in the genomes.

## 4. WEB SITES

Most of the Websites visited for this work are listed below:

- http://gib.genes.nig.ac.jp/
- http://gib.genes.nig.ac.jp/single/index.php?spid=Aful_DSM4304
- http://gib.genes.nig.ac.jp/single/index.php?spid=Aper_K1
- http://gib.genes.nig.ac.jp/single/index.php?spid=Halo_NRC1
- http://gib.genes.nig.ac.jp/single/index.php?spid=Mace_C2A
- http://gib.genes.nig.ac.jp/single/index.php?spid=Mjan_DSM2661
- http://gib.genes.nig.ac.jp/single/index.php?spid=Mkan_AV19
- http://gib.genes.nig.ac.jp/single/index.php?spid=Mmaz_GOE1
- http://gib.genes.nig.ac.jp/single/index.php?spid=Mthe_DELTAH
- http://gib.genes.nig.ac.jp/single/index.php?spid=Paby_ORSAY
- http://gib.genes.nig.ac.jp/single/index.php?spid=Paer_IM2
- http://gib.genes.nig.ac.jp/single/index.php?spid=Pfur_DSM3638
- http://gib.genes.nig.ac.jp/single/index.php?spid=Phor_OT3
- http://gib.genes.nig.ac.jp/single/index.php?spid=Ssol_P2
- http://gib.genes.nig.ac.jp/single/index.php?spid=Stok_7
- http://gib.genes.nig.ac.jp/single/index.php?spid=Taci_DSM1728
- http://gib.genes.nig.ac.jp/single/index.php?spid=Tvol_GSS1
- http://gib.genes.nig.ac.jp/single/index.php?spid=Cace_ATCC824
- http://gib.genes.nig.ac.jp/single/index.php?spid=Bsub_168
- http://gib.genes.nig.ac.jp/single/index.php?spid=Ccre_CB15
- http://gib.genes.nig.ac.jp/single/index.php?spid=Scoe_A3
- http://gib.genes.nig.ac.jp/single/index.php?spid=Tten_MB4T
- http://gib.genes.nig.ac.jp/single/index.php?spid=Drad_R1
- http://gib.genes.nig.ac.jp/single/index.php?spid=Syne_PCC6803
- http://gib.genes.nig.ac.jp/single/index.php?spid=Tmar_MSB8
- http://gib.genes.nig.ac.jp/single/index.php?spid=Aaeo_VF5
- http://genome.ornl.gov/microbial/mmar/
- http://genome.ornl.gov/microbial/mbur/
- http://genome.ornl.gov/microbial/mbar/
- http://genome.ornl.gov/microbial/tfus/
- http://genome.ornl.gov/microbial/faci/
- http://www-genome.wi.mit.edu/annotation/microbes/methanosarcina/
- http://us.expasy.org/sprot/
- http://www.ncbi.nlm.nih.gov/
- http://www.g2l.bio.uni-goettingen.de/
- http://www.tigr.org/
- http://www.jgi.doe.gov/
- http://www.jgi.doe.gov/JGI_microbial/html/methanococcoides/methanoc_mainpage.html
- http://www.jgi.doe.gov/JGI_microbial/html/methanosarcina/methano_mainpage.html
- http://bahama.jgi-psf.org/prod/bin/blast.tfus.cgi
- http://www.jgi.doe.gov/JGI_microbial/html/ferroplasma/ferro_mainpage.html
- http://www.genome.ad.jp/kegg/

## 5. PROGRAMS

The programs used were those in the Wisconsin GCG package, for example, Gap, Seqed, Stringsearch,

**Table 1.** Euryarchaeota studied

| Order | Family | Organism | OTG (° C)[a] | Genome size (Mb) | Method[b] | pH[c] |
|---|---|---|---|---|---|---|
| *Methanosarcinales* | *Methanosarcinaceae* | *Methanosarcina mazeii* S-6 | 37 | n.d.[d] | S, N, W, seq. | 6.8-7.2 |
| | | *Methanosarcina mazeii* JC3 | 37 | n.d. | N | 6.8-7.2 |
| | | *Methanosarcina mazeii* LYC | 37 | n.d. | N | 6.8-7.2 |
| | | *Methanosarcina mazeii* Goe1 | 37 | 4.1 | Seq. | 6.8-7.2 |
| | | *Methanosarcina* sp. strain JVC | 37 | n.d. | N | 6.8-7.0 |
| | | *Methanosarcina acetivorans* C2A | 37 | 5.7 | N, Seq. | 6.5-7.0 |
| | | *Methanosarcina barkeri* | 37 | 2.7 | S, Seq. | 7.0 |
| | | *Methanosarcina thermophila* TM-1 | 50 | 2.7 | S, N, Seq. | 6.0-7.0 |
| *Methanomicrobiales* | *Methanospirillaceae* | *Methanospirillum hungateii* | 37 | n.d. | S | 6.8-7.2 |
| *Methanobacteriales* | *Methanobacteriaceae* | *Methanobacterium thermoautotrophicum*[e] | 65 | 1.7 | Seq. | 7.0-8.0 |
| | *Methanothermaceae* | *Methanothermus fervidus* | 85-88 | n.d. | S, P | 6.5 |
| *Methanococcales* | *Methanococcaceae* | *Methanococcus voltae* | 37 | n.d. | S, W | 6.5-8.0 |
| | | *Methanococcus vannielii* | 37 | n.d. | S, P | 6.5-8.0 |
| | | *Methanococcus jannaschii* DSM 2661 | 85 | 1.7 | S, Seq. | 6.0 |
| | | *Methanococcus maripaludis* | 38 | n.d. | Seq. | 6.5-8.0 |
| | | *Methanococcoides burtonii* | 23 | 3.0 | Seq. | 7.7 |
| *Methanopyrales* | *Methanopyraceae* | *Methanopyrus kandleri* AV19 | 100-110 | 1.6 | Seq. | 6.5 |
| *Thermoplasmatales* | *Thermoplasmataceae* | *Thermoplasma acidophilum* DSM 1728 | 59 | 1.6 | Seq., P | 2.0 |
| | *Ferroplasmaceae* | *Ferroplasma acidarmanus* fer1 | 40 | 2.0 | Seq. | 0-3.0 |
| | | *Thermoplasma volcanium* GSS1 | 60 | 1.6 | Seq. | 2.0 |
| *Archaeoglobales* | *Archaeoglobaceae* | *Archaeoglobus fulgidus* DSM4304 | 83 | 2.2 | Seq., P | 6.0-7.0 |
| *Thermococci* | *Thermococcaceae* | *Thermococcus tenax* | 88 | n.d. | S, P | 6.8-7.0 |
| | | *Pyrococcus furiosus* DSM 3638 | 100 | 1.9 | Seq. | 6.8-7.0 |
| | | *Pyrococcus woesei* | 100 | 1.9 | S, P | 6.8-7.0 |
| | | *Pyrococcus horikoshii* (shinkaj) OT3 | 98-100 | 1.7 | Seq. | 6.8-7.0 |
| | | *Pyrococcus abyssi* GE5 | 103 | 1.8 | Seq. | 6.8-7.0 |
| *Halobacteriales* | *Halobacteriaceae* | *Halobacterium* sp. NRC-1 | 37 | 2.6 | Seq. | 6.8-7.0 |
| | | *Halobacterium marismuorti* | 37 | n.d. | seq. | 6.8-7.0 |
| | | *Halobacterium cutirubrum* | 37 | n.d. | seq. | 6.8-7.0 |

[a]OTG, optimal temperature for growth. Supplementary information can be found in 26,40-45, and the Websites listed in the text. [b]S, N, and W, Southern, Northern, and Western blotting respectively; P, PCR; seq., or Seq., sequencing of gene or genome, respectively. [c]pH, or pH range, reported to support the best growth, as compared to other pH values tested. [d]n.d., not determined. [e]*Methanothermobacter thermoautotrophicus* delta-H.

**Table 2.** Crenarchaeota studied

| Order | Family | Organism | OTG (° C)[a] | Genome size (Mb) | Method[b] | pH[c] |
|---|---|---|---|---|---|---|
| *Sulfolobales* | *Sulfolobaceae* | *Sulfolobus solfataricus* P2 | 80 | 2.9 | S, P, Seq. | 2.0-4.5 |
| | | *Sulfolobus* sp. | 70 | n.d.[d] | S | 1.0-5.5 |
| | | *Sulfolobus tokodaii* strain 7 | 80 | 2.6 | Seq. | 2.0-3.0 |
| | | *Sulfolobus acidocaldarius* | 75 | n.d. | seq. | 1.0-5.0 |
| *Desulfurococcales* | *Desulfurococcaceae* | *Desulfurococcus mobilis* | 85 | n.d. | S, P | 1.0-5.0 |
| | | *Aeropyrum pernix* K1 | 95 | 1.7 | Seq. | 6.8-7.2 |
| *Thermoproteales* | *Thermoproteaceae* | *Pyrobaculum aerophilum* IM2 | 100 | 2.2 | Seq. | 6.8-7.2 |

[a]OTG, optimal temperature for growth. Supplementary information can be found in 26,42,44,45, and the Websites listed in the text. [b]S, N, and W, Southern, Northern, and Western blotting respectively; P, PCR; seq., or Seq., sequencing of gene or genome, respectively. [c]pH, or pH range, reported to support the best growth as compared to other pH values tested. [d]n.d., not determined.

Fetch, and Pileup. Other programs were available in the Internet and in the various genome Websites (see above), such as Blast, JGI Blast, NCBI Blast, and ORNL Microbial Blast Server.

## 6. THE MOLECULAR CHAPERONE MACHINE: Hsp70(DnaK), Hsp40(DnaJ), AND GrpE

A summary of the organisms investigated in the phylogenetic domains Archaea and Bacteria in our search for the molecular chaperone-machine genes is presented in table 4, and the distribution of *hsp70(dnaK)* is shown in table 5. Basic information on all of the archaeal Hsp70(DnaK) proteins available in databases and genome Websites is provided in table 6. In all cases, when *hsp70(dnaK)* was present in a genome, the *hsp40(dnaJ)* and *grpE* genes were also present (data not shown).

The newly studied archaeal Hsp70(DnaK) proteins lack a segment of 23-24 amino acids by comparison with the homologs from Gram negative bacteria, confirming the observation made when the first archaeal Hsp70(DnaK) sequence was described several years ago (20,47). This 23-24 amino-acid "deletion" appears between positions 83-84, or 106-108, depending on the organism, in the Hsp70(DnaK)s from archaea and

**Table 3.** Bacteria studied

| Type according to OTG, and Name [a] | OTG (°C) | Genome size (Mb) |
|---|---|---|
| **Psychrotolerants (16-35)[b]** | | |
| *Acidithiobacillus ferrooxidans* ATCC 23270 *(Thiobacillus ferrooxidans)* | 30-35 | 2.9 |
| *Geobacter sulfurreducens* | 26 | 2.5 |
| *Magnetococcus* MC1 | 20-27 | 4.5 |
| *Magnetospirillum magnetotacticum* MS-1 (ATCC 31632) | 30 | 4.5 |
| *Methylobacterium extorquens* | 30 | 6 |
| *Streptomyces griseus* | 28 | n.d.[c] |
| *Streptomyces coelicolor* A3 (a) | 28 | 8.7 |
| *Caulobacter crescentus* CB15 | 28 | 4.0 |
| **Mesophiles (36-45)** | | |
| *Bacillus anthracis* Ames | 30-40 | 4.5 |
| *Desulfitobacterium hafniense* DCB-2 | 37-38 | 4.6 |
| *Deinococcus radiodurans* R1 | 37-40 | 2.6 (total 3.3) |
| *Bacillus subtilis* | 37-40 | 4.2 |
| *Clostridium acetobutylicum* ATCC824 | 37-40 | 3.9 (total 4.1) |
| *Escherichia coli* | 37-40 | 4.6 |
| *Synechocystis* sp. PCC 6803 | 38-40 | 3.6 |
| **Thermophiles (46-70)** | | |
| *Thermus thermophilus* HB27 | 70 | 1.8 |
| *Thermobifida fusca* | 55 | 3.6 |
| **Hyperthermophiles (71 and higher)** | | |
| *Thermotoga maritima* MSB8 | 80 | 1.8 |
| *Aquifex aeolicus* VF5 | 83 | 1.6 |
| *Aquifex pyrophilus* | 83 | n.d. |
| *Thermoanaerobacter tengcongensis* MB4T | 75 | 2.6 |
| *Thermomicrobium roseum* | 80 | n.d. |

[a]OTG, optimal temperature for growth. Supplementary information can be found in 23,26,45,46, and the Websites listed in the text. Species names are in italics while strain designations are in romans. [b]OTG range in degrees Centigrade within parentheses. [c]n.d., not determined; genome sequence not available.

**Table 4.** Summary of data source

| Organisms studied | | Method | |
|---|---|---|---|
| **Phylogenetic domain** | **Number of species** | **Genome sequence** | **Other[a]** |
| **Archaea** | | | |
| Euryarchaeota | 29 | 17 | S, N, W, seq. |
| Crenarchaeota | 6[b] | 4 | S, P |
| **Bacteria** | 22 | 19 | seq. |

[a]S, N, and W, Southern, Northern, and Western blotting respectively; P, PCR; seq., or Seq., sequencing of gene or genome, respectively. [b]In **Table 2** seven species are mentioned, but only six were part of the set studied in search of the *hsp70(dnaK)* gene; *Sulfolobus acidocaldarius* is part of the prefoldin study only.

**Table 5.** *hsp70*(*dnaK*) in the organisms studied

| Organism | Total studied | With the gene |
|---|---|---|
| **Euryarchaeota** | | |
| Psychrotolerants | 1 | 1 (100%) |
| Mesophiles | 15 | 11 (73%) |
| Thermophiles | 4 | 4 (100%) |
| Hyperthermophiles | 9 | 0 (0%) |
| **Crenarchaeota (70-100)[a]** | 6 | 0 (0%) |
| **Bacteria** | | |
| Psychrotolerants | 8 | 8 (100%) |
| Mesophiles | 7 | 7 (100%) |
| Thermophiles | 2 | 2 (100%) |
| Hyperthermophiles | 5 | 5 (100%) |

[a]Optimal temperature for growth, range, in degrees Centigrade.

Gram-positive bacteria when they are aligned together with those from Gram negative bacteria.

Phylogenetic analyses revealed that archaeal Hsp70(DnaK)s form three clusters that are related to the Hsp70(DnaK)s from the Gram-positive bacteria with low G+C contents, the Gram-positive bacteria with high G+C contents, and with the Thermotogales-Aquificales-Deinococci-Green NS bacteria-Cyanobacteria-chloroplast group, respectively (45, and data not shown).

## 7. THE GROUP I CHAPERONINS GroEL AND GroES

The occurrence of the gene encoding GroEL in *Methanosarcina* species has been demonstrated (48). The

**Table 6.** Hsp70(DnaK) in organisms of the phylogenetic domain Archaea

| Organism | Accession number for SwissProt database | Total number of amino acids |
|---|---|---|
| *Methanosarcina mazeii* Goe1 | AE013494-5[a] | 619 |
| *Methanosarcina mazeii* S-6 | P27094 | 619 |
| *Methanosarcina acetivorans* | Q8TQR2 | 617 |
| *Methanosarcina barkeri* | Contig1869 Gene 1996[a] | 620 |
| *Methanosarcina thermophila* | Y17862 | 610 |
| *Methanobacterium thermoautothrophicum*[b] | O27351 | 596 |
| *Methanococcoides burtonii* | Scaffold2 Gene 967[a] | 620 |
| *Thermoplasma acidophilum* | L35529[c] | 613 |
| *Ferroplasma acidarmanus* | Contig151 Gene 56[a] | 565 |
| *Thermoplasma volcanium* | Q97BG8 | 613 |
| *Halobacterium* sp. NRC-1 | Q9HRY2 | 629 |
| *Halobacterium marismortui* | Q01100 | 635 |
| *Halobacterium cutirubrum* | L35530[c] | 629 |

[a]Number in genome Website. [b]*Methanothermobacter thermoautotrophicus* delta-H. [c]Accession number for GenBank database

*Methanosarcina acetivorans* GroEL was at least 50% identical to the bacterial counterparts examined (table 7). The Annotations confirmed that these bacterial proteins detected by Blast with *M. acetivorans* GroEL as a query, and compared with it by GAP alignments, are indeed GroEL proteins.

A comprehensive search for the *groEL* gene in all available archaeal genomes demonstrated that it occurs exclusively in *Methanosarcina* species, whose GroEL proteins are very similar to one another with over 90% identity (table 8, top two lines). The other archaea have Hsp60 proteins that produce Blast hits with GroEL from *M. acetivorans*, but GAP alignments show these proteins to be chaperonin (thermosome) subunits, with over 50% identity to Hsp60-1 (a thermosome subunit) from *M. acetivorans*. These subunits are just 30%, or less, identical to *M. acetivorans* GroEL.

The GAP results in table 8, obtained with *M. acetivorans* GroEL as standard, must be compared with those obtained with Hsp60-1 (shown within parentheses). When the query for Blast searches of genomes was *M. acetivorans* GroEL, the best Blast hits in *M. barkeri* and *M. mazeii* Goe1 genomes gave also high I (identity percent) and S (similarity percent) values in GAP alignments with *M. acetivorans* GroEL, indicating that the *Methanosarcina* proteins are GroEL. This was confirmed by phylogenetic analyses (48). In contrast, the best hits in the other archaea gave low I and S values in GAP alignments with *M. acetivorans* GroEL, showing that they are not GroEL proteins, but rather chaperonin (thermosome) subunits, homologs of the *M. acetivorans* Hsp60-1 and Hsp60-2. This is demonstrated by the GAP results shown within parentheses in table 8: the best hits in *M. barkeri* and *M. mazeii* Goe1 when *M. acetivorans* GroEL was used as query gave lower I and S values with Hsp60-1 than with GroEL in the alignments, because the hit proteins are GroEL, not chaperonin subunits. In contrast, the best hits in the other archaea, when *M. acetivorans* GroEL was used as query, gave higher I and S values with *M. acetivorans* Hsp60-1 (figures within parentheses in table 8) than with GroEL in GAP alignments because the hit proteins are chaperonin subunits, not GroEL (also demonstrated by phylogenetic analyses, Maeder *et al.*, 2003, in preparation).

The results pertaining to the *groES* gene mirror those for *groEL*. The *M. acetivorans* GroES is at least 31% identical to the bacterial homologs tested (table 9). Also, the Annotations confirmed that these bacterial proteins, detected by Blast with *M. acetivorans* GroES as a query, are GroES proteins.

None of the proteins detected in archaeal genomes by Blast with *M. acetivorans* GroES as a query are GroES, except for the proteins detected in the *Methanosarcina* genomes – which, as seen above, also contain the GroEL gene (table 10). All other genes detected by Blast encode proteins different from GroES, as shown by the Annotations. Furthermore, these proteins have considerably higher numbers of amino acids than does GroES. The lack of any resemblance of these proteins with GroES obviated the need to do GAP alignments of them with with *M. acetivorans* or *M. mazeii* Goe1 GroES proteins.

## 8. Hsp60 CHAPERONINS

The fact that *M. acetivorans* has five Hsp60 proteins belonging to the chaperonin family has been demonstrated (Maeder *et al.*, 2003, in preparation). Here, a comprehensive search for subunits Hsp60-4 and Hsp60-5 that were previously found in *M. acetivorans* has been carried out in other archaea (table 11). The proteins detected by Blast using *M. acetivorans* Hsp60-4 as a query are not Hsp60-4, because, in all archaeal species studied, the I and S values obtained by GAP alignments were considerably higher with Hsp60-1 than with Hsp60-4, both from *M. acetivorans*. Furthermore, phylogenetic analyses demonstrated that Hsp60-4 and Hsp60-5 are different from, albeit related to, the other three chaperonin group II subunits (data not shown). From this it may be concluded that *M. acetivorans* is unique in having two extra Hsp60 subunits, Hsp60-4 and Hsp60-5, which are not present in any other known archaea.

The chaperonin (thermosome) subunits known to occur in archaea are listed in table 12. They were detected using stringsearch (GCG), or by searching (Blast methods and text searches of Annotations) the respective genome Websites. The proteins that were the best Blast hits when

**Table 7.** Similarity between the *Methanosarcina acetivorans* and bacterial GroEL proteins

| Organism | Method | | | |
|---|---|---|---|---|
| | Blast[a] | | GAP | |
| | Gene/Protein[b] | Hit score (E) | I[c] | S |
| *Deinococcus radiodurans* R1 | Drad_R1: 01 \|614\| DR0607 (548 aa)[d] | 469 ($e^{-133}$) | 52.8 | 65.8 |
| *Synechocystis sp.* PCC 6803 | Syne_PCC6803: \|842\| *groEL* (541 aa) | 443 ($e^{-125}$) | 50.9 (52.6) | 64.1 (66.2) |
| *Thermotoga maritima* | Tmar_MSB8: \|521\| TM0506 (538 aa) | 501 ($e^{-143}$) | 56.5 | 68.7 |
| *Aquifex aeolicus* VF5 | Aaeo_VF5: \|1571\| *mopA* (545 aa) | 503 ($e^{-144}$) | 56.7 | 67.0 |
| *Thermobifida fusca* | Contig 63 Gene 4372 (541 aa) | 540 ($e^{-154}$) | 53.2 | 63.7 |
| *Thermoanaerobacter tengcongensis* | Tten_MB4T: \|1543\| *GroL* (540 aa) | 486 ($e^{-138}$) | 57.5 | 66.3 |
| *Bacillus subtilis* | Bsub_168: \|830\| *groEL* (544 aa) | 494 ($e^{-141}$) | 56.5 | 66.8 |
| *Streptomyces albus* | m76657 (540 aa) | n.d.[e] | 51.2 | 62.3 |

[a]Query: *Methanosarcina acetivorans* GroEL (536 aa). [b]Annotation: DR0607, encodes the groEL protein; *groEL* , encodes GroEL; TM0506, encodes GroEL; *mopA*, encodes GroEL; Gene 4372, encodes a chaperonin 2; *GroL*, encodes GroEL, *groEL*, encodes GroEL; m76657, encodes GroEL. [c]I, Percent identity and S, percent similarity (identities plus conservative substitutions), obtained by GAP (GCG) alignment with the *M. acetivorans* GroEL. [d]Total number of amino acids. [e]n.d., not determined.

**Table 8.** GroEL in archaea: Present in *Methanosarcina* species but absent in the others

| Organism | Method | | | |
|---|---|---|---|---|
| | Blast[a] | | GAP | |
| | Gene/Protein[b] | Hit score (E) | I[c] | S |
| *Methanosarcina barkeri* | 2351479_fasta. screen. Contig1865 Gene 1925 (536 aa)[d] | 837 (0.0) | 90.5 (25.6) | 94.4 (38.1) |
| *Methanosarcina mazeii* Goe1 | Mmaz_GOE1: \|1844\| *groEL* (536 aa) | 868 (0.0) | 94.6 (26.2) | 96.3 (38.7) |
| *Methanobacterium thermoautotrophicum* | Mthe_DELTAH: \|1158\| MTH794 (538 aa) | 102 ($2e^{-23}$) | 31.1 (60.4) | 42.4 (72.8) |
| *Methanococcus jannaschii* DSM 2661 | Mjan_DSM2661: \|1058\| MJ0999 (542 aa) | 109 ($3e^{-25}$) | 32.8 (60.7) | 43.9 (71.8) |
| *Methanococcus maripaludis* | Contig 1 Gene 1089 (543 aa) | 141 ($8e^{-35}$) | 27.8 | 40.2 |
| *Methanococcoides burtonii* | Scaffold_13 11184 Gene 856 (537 aa) | 96 ($5e^{-21}$) | 29.2 | 40.9 |
| *Methanopyrus kandleri* AV19 | Mkan_AV19: \|1032\| *groL* MK1006 (545 aa) | 92 ($5e^{-20}$) | 30.9 (62.5) | 43.8 (72.6) |
| *Thermoplasma acidophilum* DSM 1728 | Taci_DSM1728: \|1338\| Ta1276 (543 aa) | 93 ($2e^{-20}$) | 25.7 | 39.0 |
| *Ferroplasma acidarmanus* | 2351485_fasta.screen. Contig149 Gene 11 (542 aa) | 88 ($1e^{-18}$) | 27.7 | 39.8 |
| *Thermoplasma volcanium* | Tvol_GSS1: \|525\| TVG0494466 (544 aa) | 92 ($4e^{-20}$) | 26.3 | 38.9 |
| *Archaeoglobus fulgidus* | Aful_DSM4304: \|1460\| AF1451 (545 aa) | 84 ($1e^{-17}$) | 28.0 | 41.0 |
| *Pyrococcus furious* DSM 3638 | Pfur_DSM3638: \|2076\| PF1974 (549 aa) | 86 ($3e^{-18}$) | 26.8 | 39.1 |
| *Pyrococcus horikoshii* OT3 | Phor_OT3: \|21\| PH0017 (549 aa) | 86 ($3e^{-18}$) | 25.0 | 37.5 |
| *Pyrococcus abyssi* | Paby_ORSAY: \|27\| PAB2341 (550 aa) | 86 ($3e^{-18}$) | 25.2 | 38.8 |
| *Sulfolobus solfataricus* P2 | Ssol_P2: \|270\| *thsB* SSO0282 (557 aa) | 67 ($2e^{-12}$) | 26.9 | 37.9 |
| *Sulfolobus tokodaii* | Stok_7: \|383\| ST0321 (559 aa) | 78 ($7e^{-16}$) | 26.7 | 38.7 |
| *Aeropyrum pernix* K1 | Aper_K1: \|2196\| APE2072 (555 aa) | 77 ($2e^{-15}$) | 28.3 (51.0) | 41.5 (64.5) |
| *Pyrobaculum aerophilum* | Paer_IM2: \|1512\| PAE2117 (549 aa) | 60 ($3e^{-10}$) | 24.0 (51.9) | 38.2 (63.8) |
| *Halobacterium* sp. NRC-1 | Halo_NRC1: \|1632\| *cctB* VNG2096G (656 aa) | 83 ($3e^{-17}$) | 26.9 | 38.2 |

[a]Query: *Methanosarcina acetivorans* GroEL (536 aa). [b]Annotation: Gene 1925, encodes the GroEL protein; *groEL*, encodes a 60 kDa chaperonin (GroEL); MTH794, encodes a chaperonin; MJ0999, encodes a thermosome (ths); Gene 1089, encodes a thermosome subunit (Chaperonin subunit); Gene 856 encodes an Hsp60; *groL*, encodes a HSP60 family chaperonin; Ta1276, encodes a thermosome beta chain; Gene 11, encodes a thermosome, beta subunit; TVG0494466, encodes an archaeal chaperonin [group II]; AF1451, encodes a thermosome, subunit beta; PF1974, encodes a thermosome, single subunit; PH0017, encodes a 549aa long hypothetical thermophilic factor; PAB2341, encodes a thermosome subunit (chaperonin subunit); *thsB*, encodes aThermosome beta subunit; ST0321, encodes a thermosome, beta subunit; APE2072, encodes a 555aa long hypothetical thermosome, subunit; PAE2117, encodes a thermosome (chaperonin) alpha subunit; *cctB*, encodes a thermosome subunit beta. [c]I, Percent identity and S, percent similarity (identities plus conservative substitutions), obtained by GAP (GCG) alignment with *M. acetivorans* GroEL, or Hsp60-1 (a chaperonin subunit of 552 amino acids) for figures within parentheses. [d]Total number of amino acids.

**Table 9.** Similarity between the *Methanosarcina acetivorans* and bacterial GroES proteins

| Organism | Method | | | |
|---|---|---|---|---|
| | Blast[a] | | GAP | |
| | Gene/Protein[b] | Hit score (E) | I[c] | S |
| *Deinococcus radiodurans* R1 | Drad_R1: 01 \|613\| DR0606 (120 aa)[d] | 64 ($3e^{-12}$) | 31.1 | 53.8 |
| *Synechocystis sp.* PCC 6803 | Syne_PCC6803: \|841\| *groES* (106 aa) | 65 ($2e^{-12}$) | 34.7 | 51.5 |
| *Thermotoga maritima* | Tmar_MSB8: \|520\| TM0505 (92 aa) | 89 ($6e^{-20}$) | 44.6 | 57.6 |
| *Aquifex aeolicus* VF5 | Aaeo_VF5: \|1570\| *mopB* (122 aa) | 59 ($5e^{-11}$) | 36.3 | 57.1 |
| *Thermobifida fusca* | Scaffold_2 (103 aa) | 69 ($1e^{-13}$) | 36.1 | 55.7 |
| *Thermoanaerobacter tengcongensis* | Tten_MB4T: \|1540\| *GroS* (94 aa) | 75 ($1e^{-15}$) | 42.2 | 61.1 |
| *Bacillus subtilis* | Bsub_168: \|829\| *groES* (108 aa) | 67 ($6e^{-13}$) | 31.1 | 51.5 |
| *Streptomyces albus* | m76657 (102 aa) | n.d.[e] | 35.7 | 55.1 |

[a]Query: *Methanosarcina acetivorans* GroES (109 aa). [b]Annotation: DR0606, encodes GroES; *groES*, encodes GroES; TM0505, encodes GroES; *mopB*, encodes GroES; Scaffold_2, encodes GroES; *GroS*, encodes GroES; *groES*, encodes GroES; m76657, encodes GroES. [c]I, Percent identity and S, percent similarity (identities plus conservative substitutions) obtained by GAP (GCG) alignment with *M. acetivorans* GroES. [d]Total number of amino acids. [e]n.d., not determined.

**Table 10.** GroES in archaea: Present in *Methanosarcina* species but absent in the others

| Organism | Method | | | |
|---|---|---|---|---|
| | Blast[a] | | GAP | |
| | Gene/Protein[b] | Hit score (E) | I[c] | S |
| *Methanosarcina barkeri* | 2351479_fasta. Screen. Contig 1865 Gene 1924 (92 aa)[d] | 187 ($6e^{-49}$) | 93.5 | 96.7 |
| *Methanosarcina mazeii* Goe1 | Mmaz_GOE1: \|1843\| *groES* (92 aa) | 180 ($2e^{-47}$) | 94.6 | 96.3 |
| *Methanobacterium thermoautotrophicum* | Mthe_DELTAH: \|1915\| MTH1412 (382 aa) | 23 (2.7) | n.d.[e] | n.d. |
| *Methanococcus jannaschii* DSM 2661 | Mjan_DSM2661: \|1000\| MJ0942 (651 aa) | 24 (1.5) | n.d. | n.d. |
| *Methanococcus maripaludis* | Contig1 Gene 1805 (510 aa) | 25 (0.96) | n.d. | n.d. |
| *Methanococcoides burtonii* | Scaffold_1 495139 Gene 559 (600 aa) | 27 (0.39) | n.d. | n.d. |
| *Methanopyrus kandleri* AV19 | Mkan_AV19: \|602\| MK0587 (190 aa) | 25 (0.67) | n.d. | n.d. |
| *Thermoplasma acidophilum* DSM 1728 | Taci_DSM1728: \|330\| Ta0326 (428 aa) | 27 (0.27) | n.d. | n.d. |
| *Ferroplasma acidarmanus* | 2351485_fasta.screen. Contig147 Gene 55 (233 aa) | 22 (9.4) | n.d. | n.d. |
| *Thermoplasma volcanium* | Tvol_GSS1: \|605\| TVG0571194 (428 aa) | 30 (0.024) | n.d. | n.d. |
| *Archaeoglobus fulgidus* | Aful_DSM4304: \|425\| AF0422 (243 aa) | 26 (0.51) | n.d. | n.d. |
| *Pyrococcus furious* DSM 3638 | Pfur_DSM3638: \|1817\| PF1725 (447 aa) | 29 (0.088) | n.d. | n.d. |
| *Pyrococcus horikoshii* OT3 | Phor_OT3: \|1799\| PH1699 (447 aa) | 29 (0.086) | n.d. | n.d. |
| *Pyrococcus abyssi* | Paby_ORSAY: \|575\| PAB2002 (559 aa) | 30 (0.028) | n.d. | n.d. |
| *Sulfolobus solfataricus* P2 | Ssol_P2: \|888\| *nrd* (841 aa) | 26 (0.85) | n.d. | n.d. |
| *Sulfolobus tokodaii* | Stok_7: \|1209\| ST1058 (602 aa) | 27 (0.26) | n.d. | n.d. |
| *Aeropyrum pernix* K1 | Aper_K1: \|2694\| APE2556 (1007 aa) | 23 (3.2) | n.d. | n.d. |
| *Pyrobaculum aerophilum* | Paer_IM2: \|1458\| PAE2051 (358 aa) | 24 (1.9) | n.d. | n.d. |
| *Halobacterium* sp. NRC-1 | Halo_NRC1: \|441\| VNG0557H (226 aa) | 27 (0.33) | n.d. | n.d. |

[a]Query: *Methanosarcina acetivorans* GroES. [b]Annotation: Gene 1924, encodes the 92aa long GroES protein; *groES*, enodes a 92aa long10 kDa chaperonin (GroES); MTH1412, encodes a 382aa long Cdc6 related protein; MJ0942, encodes an 651aa long ATP-dependent DNA helicase DinG, putative (dinG); Gene 1805, encodes a 510aa long conserved hypothetical protein; Gene 559, encodes a 600aa long hypothetical protein; MK0587, encodes an 190aa long uncharacterized conserved protein; Ta0326, encodes a 428aa long fixC protein related; Gene 55, encodes a 233aa long 5'-Methylthioadenosine Phosphorylase; TVG0571194, encodes a 401aa long hypothetical protein; AF0422, encodes a 243aa long uroporphyrin-III C-methyltransferase (cysG-1); PF1725, encodes a 447aa long hypothetical protein; PH1699, encodes a 447aa long hypothetical protein; PAB2002, encodes a 559aa long lig DNA ligase; *nrd*, encodes a 841aa long ribonucleotide reductase (nrd); ST1058, encodes a 602aa long hypothetical phosphoenolpyruvate carboxykinase; APE2556, encodes a 1007aa long hypothetical protein; PAE2051, encodes an 358aa long alcohol dehydrogenase (Zinc); VNG0557H, encodes a 226aa long hypothetical protein. [c]I, Percent identity and S, percent similarity (identities plus conservative substitutions) obtained by GAP (GCG) alignment with *M. acetivorans* GroES. [d]Total number of amino acids. [e]n.d., not done.

we used *M. acetivorans* Hsp60-4 as query (see table 11) are marked with asterisks. The data show that archaea have varying numbers of thermosome subunits, either1, 2, or 3, as reported previously when a smaller sample was examined (49).

The protein of *Methanosarcina barkeri*, annotated as "thermosome subunit 1" (Contig1921; Gene 3128; Table 12) consists of only 156 amino acids, which is very short, both by comparison with the other three annotated thermosome subunits of *M. barkeri*, and by comparison with the thermosome subunits of all other archaeal organisms, which consist of at least 500 amino acids, usually more (table 12). A Pileup with the four annotated *M. barkeri* subunits demonstrated that the short "thermosome subunit 1" is similar to the C-terminal regions of the two longest subunits of this organism (data not shown). Data in table 13 show that the 156-amino acid-long protein annotated as "thermosome subunit 1" shares the highest I and S values with the C-terminal region of the protein annotated as "thermosome subunit beta" (Contig 1923; Gene 3177), consisting of 543 amino acids. The data also show that the 156 amino-acid long protein ("thermosome subunit 1") is not alignable with the C-terminal region of the shorter protein annotated as "thermosome, subunit ß" (400 amino-acids long) since this shorter beta subunit lacks this C-terminal region. The data also demonstrate that the C-terminal region of the longer "thermosome, subunit beta" (543 aa; Contig 1923; Gene 3177), shares the highest I and S values with that of the thermosome subunit with 547 amino acids (Contig 1922; Gene 3148) because both have the C-terminal segment, whereas the shorter thermosome subunit beta (Contig 1921; Gene 3126) has only 400 amino acids and lacks the C-terminal segment common to the other three subunits.

## 9. PREFOLDINS

There is little comprehensive information on the occurrence of prefoldins in archaea. table 14 is a compendium of data from our searches. The prefoldins were found by stringsearch (GCG) in the SwissProt database, or by searching the various genome Websites using Blast methods and text searches of the Annotations. The data show that all organisms whose genomes have been fully sequenced have two prefoldin subunits. The apparent exception listed in table 14, *S. acidocaldarius*, is not necessarily a true exception, because its genome has not been sequenced, and the information available stems from cloning and sequencing of the single gene.

Initially, it was found that *Ferroplasma acidarmanus* had no prefoldin, according to its genome Website Annotations. Interestingly, no significant Blast hits were obtained in the *F. acidarmanus* genome sequence when the *M. acetivorans* prefoldins were used as queries.

**Table 11.** Hsp60 in archaea: Hsp60-4 is present only in *Methanosarcina acetivorans*

| Organism | Method | | | | | | |
|---|---|---|---|---|---|---|---|
| | Blast[a] | | | GAP | | | |
| | Gene/Protein[b] | Hit[c] | | With Hsp60-4 | | With Hsp60-1 | |
| | | Score | E | I[d] | S | I[d] | S |
| *Methanosarcina mazeii* Goe1 | RMMZ00858 1289933_1291558 (542 aa)[e] | 276 | $2e^{-75}$ | 34.6 | 45.4 | 70.8 | 78.4 |
| *Methanosarcina barkeri* | 2351479_fasta.screen.Contig1923 (543 aa) | 275 | $1e^{-74}$ | 32.8 | 45.2 | 69.7 | 77.1 |
| *Methanobacterium thermoauto*[f] | Mthe_DELTAH: \|1158\| MTH794 (538 aa) | 288 | $3e^{-79}$ | 33.9 | 47.5 | 60.4 | 72.8 |
| *Methanococcus jannaschii* | Mjan_DSM2661: \|1058\| MJ0999 (542 aa) | 288 | $3e^{-79}$ | 34.1 | 46.9 | 60.7 | 71.8 |
| *Methanococcus maripaludis* | Contig1 Gene 1089 (543 aa) | 261 | $5e^{-71}$ | 33.9 | 46.8 | 61.3 | 72.0 |
| *Methanoccocoides burtonii* | Scaffold_1 495139 Gene 142 (542 aa) | 266 | $2e^{-72}$ | 32.9 | 45.5 | 76.5 | 84.7 |
| *Methanopyrus kandleri* | Mkan_AV19: \|1032\| *groL* MK1006 (545 aa) | 281 | $3e^{-77}$ | 33.7 | 47.7 | 62.5 | 72.6 |
| *Thermoplasma acidophilum* | Taci_DSM1728: \|1035\| Ta0980 (549 aa) | 249 | $2e^{-67}$ | 31.4 | 43.9 | 57.2 | 68.3 |
| *Ferroplasma acidarmanus* | 2351485_fasta.screen. Contig145 Gene 15 (545 aa) | 258 | $6e^{-70}$ | 31.6 | 44.1 | 58.5 | 68.6 |
| *Thermoplasma volcanium* | Tvol_GSS1: \|1186\| TVG1181974 (549 aa) | 259 | $1e^{-70}$ | 31.6 | 44.3 | 57.6 | 68.6 |
| *Archaeoglobus fulgidus* | Aful_DSM4304: \|1460\| AF1451 (545 aa) | 281 | $5e^{-77}$ | 33.5 | 47.2 | 64.5 | 74.9 |
| *Pyrococcus furiosus* | Pfur_DSM3638: \|2076\| PF1974 (549 aa) | 275 | $4e^{-75}$ | 33.7 | 47.8 | 60.0 | 71.6 |
| *Pyrococcus horikoshii* | Phor_OT3: \|21\| PH0017 (549 aa) | 280 | $8e^{-77}$ | 34.2 | 47.0 | 59.9 | 71.0 |
| *Pyrococcus abyssi* | Paby_ORSAY: \|27\| PAB2341 (550 aa) | 274 | $4e^{-75}$ | 33.9 | 46.9 | 60.0 | 70.8 |
| *Sulfolobus solfataricus* P2 | Ssol_P2: \|270\| *thsB* SSO0282 (557 aa) | 250 | $1e^{-67}$ | 34.9 | 46.3 | 50.3 | 62.8 |
| *Sulfolobus tokodaii* | Stok_7: \|383\| ST0321 (559 aa) | 257 | $9e^{-70}$ | 34.0 | 45.3 | 51.7 | 64.1 |
| *Aeropyrum pernix* | Aper_K1: \|978\| APE0907 (557 aa) | 271 | $4e^{-74}$ | 34.1 | 48.0 | 47.9 | 61.2 |
| *Pyrobaculum aerophilum* | Paer_IM2: \|2386\| PAE3273 (553 aa) | 256 | $2e^{-69}$ | 33.9 | 46.6 | 50.0 | 62.5 |
| *Halobacterium* sp. NRC-1 | Halo_NRC1: \|1746\| *cctA* VNG2226G (581 aa) | 252 | $4e^{-68}$ | 31.0 | 44.0 | 58.6 | 66.8 |

[a]Query: *Methanosarcina acetivorans* Hsp60-4. The proteins detected by Blast with Hsp60-4 as a query are not Hsp60-4 but closer to Hsp60-1, considering the Annotations and the identity and similarity percentages provided by Blast, and by the GAP alignments. See also Table 12. [b]Annotation: RMMZ00858, encodes a thermosome, alpha subunit; 2351479_fasta.screen. Contig1923, encodes a thermosome, subunit beta (thsB); MTH794, encodes a chaperonin; MJ0999, encodes a thermosome (ths); Contig1 Gene 1089, encodes a thermosome subunit (Chaperonin subunit); Scaffold_1 495139 Gene 142, encodes a Hsp60; *groL*, encodes a HSP60 family chaperonin; Ta0980, encodes a thermosome, alpha chain; 2351485_fasta.screen. Contig145, encodes a thermosome, alpha subunit; TVG1181974, encodes an archaeal chaperonin [group II]; AF1451, encodes a thermosome, subunit beta (thsB); PF1974, encodes a thermosome, single subunit; PH0017, encodes a hypothetical thermophilic factor; PAB2341, encodes a thermosome subunit (chaperonin subunit); thsB, encodes a thermosome beta subunit; ST0321, encodes a thermosome, beta subunit; APE0907, encodes a hypothetical thermosome subunit; PAE3273, encodes a thermosome (chaperonin) beta subunit; *cctA*, encodes a thermosome subunit alpha. [c]The tabulated Score and E values detected by Blast pertain to Hsp60-4 (not to Hsp60-1). [d]I, Percent identity and S, percent similarity (identities plus conservative substitutions) obtained by GAP (GCG) alignment with *M. acetivorans* Hsp60-4 and Hsp60-1, as indicated. Hsp60-4 is 97% identical to Hsp60-5 meaning that all the results obtained with the former can be taken to accurately represent the results that would have been obtained with Hsp60-5. In conclusion, both genes/proteins, Hsp60-4 and Hsp60-5, occur only in *M. acetivorans*. [e]Total number of amino acids. [f]*Methanobacterium thermoautotrophicum* or *Methanothermobacter thermoautotrophicus* delta-H.

Since *F. acidarmanus* belongs to the same order as *Thermoplasma acidophilum* and *Thermoplasma volcanium*, and to the same family (the *Ferroplasmaceae*) as the latter, we decided to use the prefoldin alpha and beta subunits of *T. acidophilum* and *T. volcanium* as queries in the Blast searches, and significant hits were obtained. Data in table 15 show that the genes detected in this way in the *F. acidarmanus* genome, which had been annotated as "hypothetical" in its Website (table 14), most likely encode prefoldin subunits.

A comparison of the archaeal prefoldin subunits listed in table 14 with one another and with the known eucaryal prefoldin subunits (table 16) showed: a) in the Archaea, the beta subunit was, on average, shorter than the alpha subunit (table 17); b) in eukaryotes, the lengths of the subunits 1 to 6, varied in a progression from the shortest to the longest as follows: subunit 1, 6, 4, 2, 5, and 3; c) in terms of average length, the archaeal alpha was closer to the eucaryal subunit 2, and the archaeal beta was closer to the eucaryal subunit 6; and d) eukaryotic organisms had a maximum of six subunits (table 16), in contrast to the archaeal organisms, which had two (table 14).

## 10. DISCUSSION

The occurrence of the *hsp70(dnaK)* gene in an organism of the phylogenetic domain Archaea was demonstrated, by cloning and sequencing, for the first time in 1991 (47). This finding was soon extended to other archaeal species (20). However, early observations suggested that *hsp70(dnaK)*, which is present in all bacteria and eukaryotes with no known exception, is absent in some archaea (42,50). Later, by using more reliable methods, it was confirmed that the distribution of the gene among the archaea is indeed discontinuous (45). While there was no doubt that the gene and its teammates in the molecular chaperone machine, *hsp40(dnaJ)* and *grpE*, are absent in some archaeal organisms, their actual distribution among the sequenced genomes had not been assessed. The studies reported here demonstrate that the gene occurs in a high percentage of mesophiles and thermophiles, but is absent in hyperthermophiles (OTG equal to, or higher than, 70 degrees Centigrade), and in the Crenarchaeota (all of those examined are hyperthermophiles). In contrast, the gene was found in all bacteria examined, regardless of their OTGs.

**Table 12.** Chaperonin (thermosome) subunits identified in archaeal genomes

| Organism | Protein annotated as: | Number of amino acids | Accession number in genome Website |
|---|---|---|---|
| *Methanosarcina mazeii* Goe1 | Thermosome subunit | 567 | RMMZ02514 |
| | Thermosome, alpha subunit | 542 | RMMZ00858[a] |
| | Thermosome, alpha subunit | 551 | RMMZ01724 |
| *Methanosarcina barkeri* | Thermosome subunit 1 | 156 | Contig1921 Gene 3128 |
| | Thermososme, subunit beta | 400 | Contig1921 Gene 3126 |
| | Thermosome subunit | 547 | Contig1922 Gene 3148 |
| | Thermosome, subunit beta | 543 | Contig1923 Gene 3177[a] |
| *Methanosarcina acetivorans* | Hsp60-1 | 552 | MA0086 |
| | Hsp60-2 | 543 | MA4413 |
| | Hsp60-3 | 547 | MA1682 |
| | Hsp60-4 | 535 | MA4386 |
| | Hsp60-5 | 517 | MA0857 |
| *Methanobacterium thermoautotrophicum*[b] | Chaperonin | 552 | MTH218 |
| | Chaperonin | 538 | MTH794[a] |
| *Methanococcus jannaschii* | Thermosome | 542 | MJ0999[a] |
| *Methanococcus maripaludis* | Thermosome subunit | 543 | Contig1 Gene 1089[a] |
| *Methanoccocoides burtonii* | Hsp60 | 542 | Scaffold1 Gene 142[a] |
| | Hsp60 | 503 | Scaffold1 Gene 427 |
| | Hsp60 | 537 | Scaffold13 Gene 856 |
| *Methanopyrus kandleri* | HSP60 family chaperonin | 545 | MK1006 (*groL*)[a] |
| *Thermoplasma acidophilum* | thermosome, alpha chain | 549 | Ta0980[a] |
| | thermosome beta chain | 543 | Ta1276 |
| *Ferroplasma acidarmanus* | Thermosome alpha-subunit | 545 | Contig145 Gene 15[a] |
| | Thermosome, beta subunit | 542 | Contig149 Gene 11 |
| *Thermoplasma volcanium* | archaeal chaperonin [group II] | 544 | TVG0494466 |
| | archaeal chaperonin [group II] | 549 | TVG1181974[a] |
| *Archaeoglobus fulgidus* | thermosome, subunit alpha (thsA) | 545 | AF2238 |
| | thermosome, subunit beta (thsB) | 545 | AF1451[a] |
| *Pyrococcus furiosus* | thermosome, single subunit | 549 | PF1974[a] |
| *Pyrococcus horikoshii* | hypothetical thermophilic factor | 549 | PH0017[a] |
| *Pyrococcus abyssi* | thermosome subunit (chaperonin subunit) | 550 | PAB2341[a] |
| *Sulfolobus solfataricus* P2 | Thermosome alpha subunit (thermophilic factor 55) | 559 | SSO0862 |
| | Thermosome beta subunit (thermophilic factor 55) | 557 | SSO0282 (*thsB*)[a] |
| | Thermosome gamma subunit (thermophilic factor 55) | 539 | SSO3000 |
| *Sulfolobus tokodaii* | thermosome, alpha subunit | 568 | ST1253 |
| | thermosome, beta subunit | 559 | ST0321[a] |
| | hypothetical thermosome, unidentified subunit | 545 | ST0820 |
| *Aeropyrum pernix* | hypothetical thermosome, subunit | 555 | APE2072 |
| | hypothetical thermosome subunit | 557 | APE0907[a] |
| *Pyrobaculum aerophilum* | thermosome (chaperonin) alpha subunit | 549 | PAE2117 |
| | thermosome (chaperonin) beta subunit | 553 | PAE3273[a] |
| *Halobacterium* sp. NRC-1 | thermosome subunit alpha | 581 | VNG2226G (*cctA*)[a] |
| | thermosome subunit beta | 656 | VNG2096G (*cctB*) |

[a]Proteins that were the best hits when blasted with Hsp60-4 from *M. acetivorans* (see Table 11). [b]*Methanobacterium thermoautotrophicum* or *Methanothermobacter thermoautotrophicus* delta-H.

**Table 13.** Chaperonin (thermosome) subunits in *Methanosarcina barkeri*

| Pairs of thermosome subunits that were compared[a] | | GAP | |
|---|---|---|---|
| | | I[b] | S |
| Thermosome, subunit deta (543 aa)[c] | Thermosome subunit 1 (156 aa) | 67.4 | 73.8 |
| Thermosome subunit (547 aa) | Thermosome subunit 1 (156 aa) | 43.0 | 53.0 |
| Thermosome, subunit beta (400 aa) | Thermosome subunit 1 (156 aa) | N.A.[d] | N.A.[d] |
| Thermosome, subunit beta (543 aa) | Thermosome subunit (547 aa) | 49.3 | 62.1 |
| Thermosome, subunit beta (543 aa) | Thermosome, subunit beta (400 aa) | N.A.[d] | N.A.[d] |
| Thermosome subunit (547 aa) | Thermosome, subunit beta (400 aa) | 20.7 | 34.5 |

[a]See Table 12. Comparisons were done first using Pileup: it was established that the 156-amino acid long "thermosome subunit 1" aligned with the last 156 C-terminal amino acids of the other *M. barkeri* subunits, except the shorter (400 amino acids) "thermosome subunit beta." GAP alignments were then run between "thermosome subunit 1" and the last 156 C-terminal amino acids of the other subunits, and between the last C-terminal amino acids of each of the other subunits: and the results are displayed in this Table. [b]I, Percent identity and S, percent similarity (identities plus conservative substitutions), obtained by GAP. [c]Total number of amino acids of the entire molecule. [d]N.A., not alignable.

**Table 14.** Prefoldin subunits in archaeal genomes

| Organism | Protein annotated as: | Number of amino acids | Accession number for SwissProt database |
|---|---|---|---|
| **Euryarchaeota** | | | |
| *Methanosarcina mazeii* Goe1 | Prefoldin alpha subunit | 145 | AE008384_809[a] |
| | Prefoldin beta subunit | 117 | AE008384_674[a] |
| *Methanosarcina acetivorans* | Prefoldin alpha subunit (GimC alpha subunit) | 142 | Q8tin6 |
| | Prefoldin beta subunit (GimC beta subunit) | 117 | Q8tjd5 |
| *Methanosarcina barkeri* | Prefoldin alpha subunit (GimC alpha subunit) | 144 | Contig1947 Gene 3963[a] |
| | Hypothetical protein (prefoldin beta subunit) | 117 | Contig1934 Gene 3465[a] |
| *Methanobacterium thermoauto*[b] | Prefoldin alpha subunit (GimC alpha subunit) | 141 | O27646 |
| | Prefoldin beta subunit (GimC beta subunit) | 121 | O26774 |
| *Methanococcus jannaschii* | Prefoldin alpha subunit (GimC alpha subunit) | 142 | Q58362 |
| | Prefoldin beta subunit (GimC beta subunit) | 113 | Q58394 |
| *Methanococcus maripaludis* | Conserved hypothetical protein / Prefoldin alpha subunit | 144 | Conti1 Gene 845[a] |
| | Prefoldin beta subunit (GimC beta subunit) | 113 | Contig1 Gene 1744[a] |
| *Methanococcoides burtonii* | Prefoldin, subunit alpha | 138 | Scaffold7 Gene 3461[a] |
| | Prefoldin, subunit beta | 117 | Scaffold5 Gene 2870[a] |
| *Methanopyrus kandleri* | Prefoldin alpha subunit (GimC alpha subunit) | 157 | Q8tuy7 |
| | Prefoldin beta subunit (GimC beta subunit) | 120 | Q8tyc7 |
| *Thermoplasma acidophilum* | Prefoldin alpha subunit (GimC alpha subunit) | 130 | Q9hj94 |
| | Prefoldin beta subunit (GimC beta subunit) | 124 | Q9hj36 |
| *Ferroplasma acidarmanus* | Conserved hypothetical protein (alpha subunit)[c] | 133 | Contig131 Gene 18[a] |
| | Hypothetical protein (beta subunit)[c] | 127 | Contig130 Gene 12[a] |
| *Thermoplasma volcanium* | Prefoldin alpha subunit (GimC alpha subunit) | 130 | Q97bc5 |
| | Prefoldin beta subunit (GimC beta subunit) | 124 | Q979c4 |
| *Archaeoglobus fulgidus* | Prefoldin alpha subunit (GimC alpha subunit) | 137 | O28216 |
| | Prefoldin beta subunit (GimC beta subunit) | 116 | O29115 |
| *Pyrococcus furiosus* | Prefoldin alpha subunit (GimC alpha subunit) | 146 | Q8u3t0 |
| | Prefoldin beta subunit (GimC beta subunit) | 117 | Q8u3s3 |
| *Pyrococcus horikoshii* | Prefoldin alpha subunit (GimC alpha subunit) | 148 | O58263 |
| | Prefoldin beta subunit (GimC beta subunit) | 117 | O58268 |
| *Pyrococcus abyssi* | Prefoldin alpha subunit (GimC alpha subunit) | 148 | Q9uyi4 |
| | Prefoldin beta subunit (GimC beta subunit) | 117 | Q9uyj4 |
| *Halobacterium* sp. | Prefoldin alpha subunit (GimC alpha subunit) | 154 | Q9hmn2 |
| | Prefoldin beta subunit (GimC beta subunit) | 125 | Q9hsh0 |
| **Crenarchaeota** | | | |
| *Sulfolobus solfataricus* | Probable prefoldin alpha subunit (GimC alpha subunit) | 147 | P58179 |
| | Prefoldin beta subunit (GimC beta subunit) | 126 | Q9uxb8 |
| *Sulfolobus tokodaii* | Prefoldin alpha subunit (GimC alpha subunit) | 151 | Q971i6 |
| | Prefoldin beta subunit (GimC beta subunit) | 125 | Q975h2 |
| *Aeropyrum pernix* | Probable prefoldin alpha subunit (GimC alpha subunit) | 154 | Q9yd28 |
| | Prefoldin beta subunit (GimC beta subunit) | 123 | Q9yc11 |
| *Pyrobaculum aerophilum* | Prefoldin alpha subunit (GimC alpha subunit) | 132 | Q8ztt9 |
| | Prefoldin beta subunit (GimC beta subunit) | 126 | Q8zvn4 |
| *Sulfolobus acidocaldarius* | Probable prefoldin alpha subunit (GimC alpha subunit) | 146 | P38617 |

[a]Number in genome Website. [b]*Methanobacterium thermoautotrophicum* or *Methanothermobacter thermoautotrophicus* delta-H. [c]Detected by Blast with subunits from *T. acidophilum* and *T. volcanium* subunits as queries, and identified by GAP with these subunits: see Table 15. *M. acetivorans* subunits failed to produce hits in Blast.

**Table 15.** Prefoldin subunits in *Ferroplasma acidarmanus*: Comparison with the subunits from two related archaeal species

| GAP with: | ***F. acidarmanus* Gene/Protein** | | | |
|---|---|---|---|---|
| | **Contig 131 Gene 18** | | **Contig 130 Gene 12** | |
| | **I[a]** | **S** | **I** | **S** |
| *Thermoplasma acidophilum* | | | | |
| alpha | 31.8 | 48.8 | | |
| beta | | | 53.2 | 68.5 |
| *Thermoplasma volcanium* | | | | |
| alpha | 31.8 | 48.8 | | |
| beta | | | 51.6 | 71.0 |

[a]I, Percent identity and S, percent similarity (identities plus conservative substitutions), obtained by GAP (GCG) alignment with the *T. acidophilum* and *T. volcanium* subunits, as shown.

**Table 16.** Prefoldin subunits in eukaryotic genomes

| Organism | Protein annotated as: | Number of amino acids | Accession number for SwissProt database |
|---|---|---|---|
| *Caenorhabditis elegans* | Probable prefoldin subunit 1 | 117 | Q17827 |
| | Probable prefoldin subunit 2 | 141 | Q9N5M2 |
| | Probable prefoldin subunit 3 | 185 | O18054 |
| | Probable prefoldin subunit 4 | 126 | Q17435 |
| | Probable prefoldin subunit 5 | 152 | Q21993 |
| | Probable prefoldin subunit 6 | 126 | P52554 |
| *Homo sapiens* | Prefoldin subunit 1 | 104 | O60925 |
| | Prefoldin subunit 2 | 154 | Q9UHV9 |
| | Prefoldin subunit 3 | 185 | Q15765 |
| | Prefoldin subunit 4 | 134 | Q9NQP4 |
| | Prefoldin subunit 5 | 154 | Q99471 |
| | Prefoldin subunit 6 | 129 | O15212 |
| *Mus musculus* | Prefoldin subunit 1 | 122 | Q9CWM4 |
| | Prefoldin subunit 2 | 154 | O70591 |
| | Prefoldin subunit 5 | 154 | Q9WU28 |
| | Prefoldin subunit 6 | 127 | Q03958 |
| *Drosophila melanogaster* | Probable prefoldin subunit 2 | 143 | Q9VTE5 |
| | Probable prefoldin subunit 3 | 185 | Q9VGP6 |
| | Probable prefoldin subunit 4 | 138 | Q9VRL3 |
| | Probable prefoldin subunit 5 | 168 | Q9VCZ8 |
| | Probable prefoldin subunit 6 | 125 | Q9VW56 |
| *Schizosaccharomyces pombe* | Probable prefoldin subunit 1 | 112 | O14334 |
| | Probable prefoldin subunit 2 | 114 | Q9UTC9 |
| | Probable prefoldin subunit 3 | 169 | Q10143 |
| | Probable prefoldin subunit 4 | 123 | Q9UTD4 |
| | Probable prefoldin subunit 5 | 154 | O94307 |
| | Probable prefoldin subunit 6 | 114 | O14450 |
| *Saccharomyces cerevisiae* | Prefoldin subunit 1 | 109 | P46988 |
| | Prefoldin subunit 2 | 123 | P40005 |
| | Prefoldin subunit 3 | 199 | P48363 |
| | Prefoldin subunit 4 | 129 | P53900 |
| | Prefoldin subunit 5 | 163 | Q04493 |
| | Prefoldin subunit 6 | 114 | P52553 |
| *Arabidopsis thaliana* | Probable prefoldin subunit 2 | 148 | Q9LJ98 |
| | Probable prefoldin subunit 3 | 195 | P57741 |
| | Probable prefoldin subunit 4 | 128 | Q9M4B5 |
| | Probable prefoldin subunit 5 | 151 | P57742 |
| *Avena fatua* | Probable prefoldin subunit 4 | 126 | Q9M4C4 |

**Table 17.** Number of amino acids in the prefoldin subunits from organisms of the phylogenetic domains Archaea and Eucarya[a]

| Phylogenetic Domain | Subunit | Number of amino acids | |
|---|---|---|---|
| | | Arithmetic mean | Range |
| **Archaea** | alpha | 143.3 (n = 21) | 130 (*T. acidophilum* and *T. volcanium*) |
| | | | 157 (*M. kandleri*) |
| | beta | 120.1 (n = 20) | 113 (*M. jannaschii* and *M. maripaludis*) |
| | | | 127 (*F. acidarmanus*) |
| **Eucarya** | 1 | 112.8 (n = 5) | 104 (*H. sapiens*) |
| | | | 122 (*M. musculus*) |
| | 2 | 139.6 (n = 7) | 114 (*S. pombe*) |
| | | | 154 (*H. sapiens* and *M. musculus*) |
| | 3 | 186.3 (n = 6) | 169 (*S. pombe*) |
| | | | 199 (*S. cerevisiae*) |
| | 4 | 129.1 (n = 7) | 123 (*S. pombe*) |
| | | | 138 (*D. melanogaster*) |
| | 5 | 156.6 (n = 7) | 151 (*A. thaliana*) |
| | | | 168 (*D. melanogaster*) |
| | 6 | 122.5 (n = 6) | 114 (*S. pombe* and *S. cerevisiae*) |
| | | | 129 (*H. sapiens*) |

[a]See Tables 14 and 16.

As for the molecular chaperone machine discussed above, the distribution of the other chaperoning systems had not been determined among the archaeal genomes now sequenced. Previous surveys had shown that conserved homologs of the bacterial co-chaperones NAC and trigger factor, and the eucaryal co-chaperones BAG, Hop, and Hip, are not present in archaea, with the probable exception of the NAC alpha subunit (36).

The chaperonins GroEL and GroES had been classified as group I. or bacterial, chaperonins, based on the belief that they existed only in bacteria and eukaryotic-cell organelles derived from bacteria (18,27,51,52). This dogma was refuted when it was discovered that *Methanosarcina* species do have the genes encoding GroEL and GroES (48,53,54). In this work we report the results of extensive searches for these two genes among the sequenced archaeal genomes. The conclusion is that they occur only in *Methanosarcina* species.

Another related chaperoning system is constituted of the chaperonin of group II, considered to be typical of the Archaea and the eukaryotic-cell cytosol (7,8,10,15,19,49,51,52,55-57). Although this chaperonin system had been investigated in various archaeal species, its distribution among organisms had not been mapped. Here we report that all sequenced archaeal genomes have at least one subunit and some have two or three, thus extending previous observations when less genome sequences were available (49,57). We also report that *Methanosarcina acetivorans* has five subunits, the highest number ever found in an archaeon, approaching the value of 8-9 that is typical of eukaryotes (10,19,55,56). Interestingly, another species of *Methanosarcina*, *M. barkeri*, has three subunits plus what appears to be a piece of a fourth.

The *M. acetivorans* chaperonin subunits were named Hsp60-1 through 5 (53), and recently it was established that Hsp60-4 and Hsp60-5 are very closely related and seem to be unique, despite the fact that they are clearly related to the other three subunits (Maeder *et al*., 2003, in preparation). Blast searches, with *M. acetivorans* Hsp60-4 as query, produced hits in all archaeal genomes. However, the results revealed that *M. acetivorans* is the only archaeal species that has Hsp60-4 (and, consequently, Hsp60-5; see below). In the rest of the archaeal species studied, the hit proteins showed higher identity when *M. acetivorans* Hsp60-1 instead of Hsp60-4 was used as a query: this was confirmed by GAP alignments of the Blast-hit proteins with the *M. acetivorans* Hsp60-1 and Hsp60-4 (table 11). The Annotations confirmed that the detected proteins are not Hsp60-4, and phylogenetic analyses showed that Hsp60-4 and Hsp60-5 are unique, and exist only in *M. acetivorans*.

The results obtained with Hsp60-4 (535 amino acids) by extension demonstrated that Hsp60-5 is absent in all archaea except *M. acetivorans*. Why? Because Hsp60-4, which was used as a query for all Blast searches, and as standard for GAP alignments, is 97% identical to Hsp60-5 (517 amino acids). Hence all results obtained with Hsp60-4 apply to Hsp60-5, its almost identical twin (the results obtained with Hsp60-4 can be predicted to be the same as those that would be obtained using Hsp60-5 instead as query and as standard for GAP alignments).

Prefoldins were discovered in eukaryotic organisms (14,16) and later identified in archaea (58), but they have never been reported to occur in bacteria. Our data confirm this absence in the bacterial domain (results not shown) and demonstrate that prefoldins exist in all archaea whose genomes have been sequenced. Two subunits were present in every archaeal especies investigated for which a complete genome sequences is available. The exception, *Sulfolobus acidocaldarius*, with only one prefoldin subunit reported, cannot be definitively considered a true exception because its genome has not yet been sequenced.

While the eucaryal prefoldin subunits number five in most organisms investigated and vary in length both within and between organisms, the two archaeal prefoldin subunits are about the same length in all species.

## 11. CONCLUSIONS AND PERSPECTIVES

A comprehensive survey of archaeal genomes conducted in order to map the distribution and composition of four chaperoning systems, revealed a number of remarkable features. 1) The gene *hsp70(dnaK)*, encoding the main component of the molecular chaperone machine, Hsp70(DnaK), is absent in a considerable proportion of genomes; 2) Whenever *hsp70(dnaK)* occurs in a genome, the genes *hsp40(dnaJ)* and *grpE*, which encode the other components of the machine, are also present; conversely, absence of one of these three genes means that the others are also absent (59,60); 3) In *Methanosarcina mazeii* S-6, the three genes are organized 5'-*grpE-hsp70(dnaK)-hsp40(dnaJ)*-'3 (61), as in many bacteria, but in contrast to the bacterial genes, the *M. mazeii*'s are transcribed individually, not as a single operon (62); 4) All hyperthermophilic archaea with an OTG of 70 degrees Centigrade or higher lack the three chaperone machine genes; 5) All archaea belonging to the Crenarchaeota also lack these three genes; 6) In contrast, all bacteria, regardless of OTG, have the three genes; 7) All archaeal Hsp70(DnaK) molecules lack a segment of 23-24 amino acids in their N-terminal quadrant by comparison with the homologs from Gram negative bacteria; 8) The genes *groEL* and *groES*, encoding the chaperonins of group I, also called the bacterial chaperonins because it was thought that they did not exist in archaea, were found in the genomes of *Methanosarcina* species, and in these species only; 9) Genes encoding the chaperonins of group II, considered typical of the eukaryotic- and archaeal-cell cytosols, were found in all species examined, without exception; 10) The total number of chaperonin genes (subunits) varies between 1 and 3; 11) *Methanosarcina acetivorans* is exceptional in that it has five genes encoding chaperonin subunits, two of which are unique to this species and are different from previously described subunits; 12) All archaeal genomes examined had two genes encoding prefoldin subunits, without exception; 13) The data in this report and in the literature suggest that the

*hsp70(dnaK)* gene, and by extension *hsp40(dnaJ)* and *grpE*, which are always found to co-exist in archaeal genomes with *hsp70(dnaK)*, were received by lateral transfer from bacteria; 14) In contrast, lateral gene transfer does not seem to have been the origin of the archaeal *groEL* and *groEL*; if lateral transfer of these two genes did occur, it must have been very early in evolution, or at least not recently enough to be demonstrable with currently available methods for evolutionary analyses. In this regard, *M. acetivorans* occupies a unique evolutionary niche that warrants further investigation. It remains to be clarified whether the *M. acetivorans groEL* and *groES* genes were inherited directly from this species' ancestors or are the result of horizontal gene transfer occurring very early; 15) Another feature that makes *M. acetivorans* of special interest, from the evolutionary and molecular biological standpoints, is the fact that its complement of chaperonin subunits approaches that of eukaryotes. Is the *M. acetivorans* thermosome more similar to the eukaryotic CCT (its functional and structural equivalent) than to the thermosomes of other archaea that have fewer chaperonin subunits? This is a tantalizing question that merits experimental investigation.

## 12. ACNOWLEDGEMENTS

We thank Yimin Dong, Tara Garcia-Collins, David A. Gross, and Anthony J. Hickey for help and useful comments, and John Leigh and Maynard Olson for providing unpublished data pertaining to the sequence of the genome of *Methanococcus maripaludis*. M. M. was partially supported by the Margret scholarship.

## 13. REFERENCES

1. Bukau B, E. Deuerling, C. Pfund & E. A. Craig: Getting newly synthesized proteins into shape. *Cell* 101, 119-122 (2000)

2. Ellis R. J. & S. M. van der Vies: Molecular chaperones. *Annu Rev Biochem* 60, 321-347 (1991)

3. Pfanner N. & M. Meijer: Mitochondrial biogenesis: The Tom and Tim machine. *Curr Biol* 7, R100-R103 (1997)

4. Ellis R. J.: Molecular chaperones: Pathways and networks. *Curr Biol* 9, R137-R139 (1999)

5. Young J. C., N. J. Hoogenraad & F. U. Hartl: Molecular chaperones Hsp90 and Hsp70 deliver preproteins to the mitochondrial import receptor Tom70. *Cell* 112, 41-50 (2003)

6. Lindquist S. & E. A. Craig: The heat-shock proteins. *Annu Rev Genet* 22, 631-677 (1988)

7. Phipps B. M., A. Hoffmann, K. O. Stetter & W. Baumeister: A novel ATPase complex selectively accumulated upon heat shock is a major cellular component of thermophilic archaebacteria. *EMBO J* 10, 1711-1722 (1991)

8. Trent J. D., E. Nimmesgern, J. S. Wall, F. U. Hartl & A. L. Horwich: A molecular chaperone from a thermophilic archaebacterium is related to the eukaryotic protein t-complex polypeptide-1. *Nature* 354, 490-493 (1991)

9. Georgopoulos C. & W. J. Welch: Role of the major heat shock proteins as molecular chaperones. *Annu Rev Cell Biol* 9, 601-634 (1993)

10. Rommelaere H., M. Van Troys, Y. Gao, R. Melki, N. J. Cowan, J. Vandekerckhove & C. Ampe: Eukaryotic cytosolic chaperonin contains t-complex polypeptide 1 and seven related subunits. *Proc Natl Acad Sci* USA 90, 11975-11979 (1993)

11. Rommelaere H., M. De Neve, R. Melki, J Vandekerckhove & C. Ampe: The cytosolic class II chaperonin CCT recognizes delineated hydrophobic sequences in its target proteins. *Biochemistry* 38, 3246-3257 (1999)

12. Hartl F. U. & J. Martin: Molecular chaperones in cellular protein folding. *Curr Op Struct Biol* 5, 92-102 (1995)

13. Cheetham M. E. & A. J. Caplan: Structure, function and evolution of DnaJ, conservation and adaptation of chaperone function. *Cell Stress Chap* 3, 28-36 (1998)

14. Geisler S., K. Siegers & E. Schiebel: A novel protein complex promoting formation of functional alpha- and gamma-tubulin. *EMBO J* 17, 952-966 (1998)

15. Klumpp M. & W. Baumeister: The thermosome: archetype of group II chaperonins. *FEBS Lett* 430, 73-77 (1998)

16. Vainberg I. E., S. A. Lewis, H. Rommelaere, C. Ampe, J. Vandekerckhove, H. L. Klein & N. J. Cowan: Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. *Cell* 93, 863-873 (1998)

17. Houry W. A.: Chaperone-assisted protein folding in the cell cytoplasm. *Curr Prot & Pep Sci* 2, 244-227 (2001)

18. Levy-Rimler G., R. E. Bell, N. Ben-Tal & A. Azem: Type I chaperonins: Not all are created equal. *FEBS Lett* 529, 1-5 (2002)

19. Valpuesta J. M., J. Martin-Benito, P. Gomez-Puertas, J. L. Carrascosa & K. R. Willison: Structure and function of a protein folding machine: The eukaryotic cytosolic chaperoninm CCT. *FEBS Lett* 529, 11-16 (2002)

20. Gupta R. S. & B. Singh: Cloning of the Hsp70 gene from *Halobacterium marismortui***:** relatedness of archaebacterial Hsp70 to its eubacterial homologs and a model for the evolution of the Hsp70 gene. *J Bacteriol* 174, 4594-4605 (1992)

21. Conway de Macario E., M. Clarens & A. J. L. Macario: Archaeal *grpE*: transcription in two different morphologic

stages of *Methanosarcina mazei* and comparison with *dnaK* and *dnaJ*. *J Bacteriol* 177, 544-550 (1995)

22. Gupta R. S. & G. B. Golding: The origin of the eukaryotic cell. *Trends Biochem Sci* 21, 166-171 (1996)

23. Gupta R. S., K. Bustard, M. Falah & D. Singh: Sequencing of heat shock protein 70 (DnaK) homologs from *Deinococcus proteolyticus* and *Thermomicrobium roseum* and their integration in a protein-based phylogeny of prokaryotes. *J Bacteriol* 179, 345-357 (1997)

24. Gupta R. S.: Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol Mol Biol Rev* 62, 1435-1491 (1998a)

25. Gupta R. S.: What are archaebacteria: life's third domain or monoderm prokaryotes related to Gram-positive bacteria? A new proposal for the classification of prokaryotic organisms. *Mol Microbiol* 29, 695-707 (1998b)

26. Macario A. J. L., M. Lange, B. K. Ahring & E. Conway de Macario: Stress genes and proteins in the Archaea. *Microbiol Mol Biol Rev* 63, 923-967 (1999)

27. Emelyanov V. V.: Phylogenetic relationships of organellar Hsp90 homologs reveal fundamental differences to organellar Hsp70 and Hsp60 evolution. *Gene* 299, 125-133 (2002)

28. Hohfeld J., Y. Minami & F. U. Hartl: Hip, a novel cochaperone involved in the eukaryotic Hsc70/Hsp40 reaction cycle. *Cell* 83, 589-598 (1995)

29. Levy E. J., J. McCarty, B. Bukau & W. J. Chirico: Conserved ATPase and luciferase refolding activities between bacteria and yeast Hsp70 chaperones and modulators. *FEBS Lett.* 368, 435-440 (1995)

30. Minami Y., J. Hohfeld, K. Ohtsuka & F. U. Hartl: Regulation of the heat-shock protein 70 reaction cycle by the mammalian DnaJ homolog, Hsp40. *J Biol Chem* 271, 19617-19624 (1996)

31. Laufen T., M. P. Mayer, C. Beisel, D. Klostermeier, A. Mong, J. Reinstein & B. Bukau: Mechanism of regulation of Hsp70 chaperones by DnaJ cochaperones. *Proc Natl Acad Sci* USA 96, 5452-5457 (1999)

32. Nollen E. A., A. E. Kabakov, J. F. Brunsting, B. Kanon, J. Hohfeld & H. H. Kampinga: Modulation of in vivo HSP70 chaperone activity by Hip and Bag-1. *J Biol Chem* 276, 4677-4682 (2001)

33. Hartl F. U. & M. Hayer-Hartl: Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* 295, 1852-1858 (2002)

34. Caplan A. J.: What is a co-chaperone? *Cell Stress Chap* 8, 105-107 (2003)

35. Lindquist S.: Heat-shock proteins and stress tolerance in microorganisms. *Curr Op Genet Dev* 2, 748-755 (1992)

36. Macario A. J. L. & E. Conway de Macario: The molecular chaperone system and other anti-stress mechanisms in archaea. *Front. Biosci.* 6, d262-283 (2001) http://www.bioscience.org/2001/v6/d/macario/fulltext.htm

37. Horwich A. L., E. U. Weber-Ban & D. Finley: Chaperone rings in protein folding and degradation. *Proc Natl Acad Sci* USA. 96, 11033-11040 (1999)

38. Ben-Zvi A. P. & P. Goloubinoff: Mechanisms of disaggregation and refolding of stable protein aggregates by molecular chaperones. *J Struct Biol* 135, 84-93 (2001)

39. Macario A. J. L. & E. Conway de Macario: Sick chaperones and ageing: A perspective. *Ageing Res Rev* 1, 295-311 (2002)

40. Franzmann P. D., N. Springer, W. Ludwig, E. Conway de Macario & M. A. Rohde: A methanogenic archaeon from Ace Lake, Antarctica: *Methanococcoides burtonii* sp. nov. *Syst Appl Microbiol* 15, 573?581 (1992)

41. Clarens M., J. J. Cairo, J. M. Paris, A. J. L. Macario & E. Conway de Macario: Characterization and forms of JC3, a new *Methanosarcina* isolate: Comparison with *Methanosarcina mazei* strains S-6$^T$, MC3 and LYC. *Curr Microbiol* 26, 167-174 (1993)

42. Conway de Macario E. & A. J. L. Macario: Heat-shock response in Archaea. *Trends Biotechnol.* 12, 512-518 (1994)

43. Hofman-Bang J. P., M. Lange, E. Conway de Macario, A. J. L. Macario & B. K. Ahring: The genes coding for the *hsp70(dnaK)* molecular chaperone machine occur in the moderate thermophilic archaeon *Methanosarcina thermophila* TM-1. *Gene* 238, 387-395 (1999)

44. Bergey's Manual of Systematic Bacteriology. Volume One. Eds: David R. Boone, Richard W. Castenholz & George M. Garrity. Springer-Verlag, New York, NY, USA (2001)

45. Gribaldo S., V. Lumia, R. Creti, E. Conway de Macario, A. Sanangelantoni & P. Cammarano: Discontinuous occurrence of the *hsp70(dnaK)* gene among archaea and sequence features of HSP70 suggest a novel outlook on phylogenies inferred from this protein. *J Bacteriol* 181, 434-443 (1999)

46. Mazodier P., G. Guglielmi, J. Davies & C. J. Thompson: Characterization of the groEL-like genes in *Streptomyces albus*. *J Bacteriol* 173, 7382-7386 (1991)

47. Macario A. J. L., C. B. Dugan & E. Conway de Macario: A *dnaK* homolog in the archaebacterium *Methanosarcina mazei* S6. *Gene* 108, 133-137 (1991)

48. Conway de Macario E., D. L. Maeder & A.J. L. Macario: Breaking the mould: Archaea with all four chaperoning systems. *Biochem Biophys Res Commun* 301, 811-812 (2003)

49. Archibald J. M., J. M. Logsdon Jr & W. F. Doolittle: Recurrent paralogy in the evolution of archaeal chaperonins. *Curr Biol* 9, 1053-1056 (1999)

50. Hebert A. M., A. M. Kropinski & K. F. Jarrell: Heat shock response of the archaebacterium *Methanococcus voltae*. *J Bacteriol* 173, 3224-3227 (1991)

51. Gupta R. S.: Evolution of the chaperonin families (Hsp60, Hsp10 and Tcp-1) of proteins and the origin of eukaryotic cells. *Mol Microbiol* 15, 1-11 (1995)

52. Ranson N. A., H. E. White & H. R. Saibil: Chaperonins. *Biochem J* 333, 233-242 (1998)

53. Galagan J. E., C. Nusbaum, A. Roy, M. G. Endrizzi, P. Macdonald, W. FitzHugh, C. Calvo, R. Engels, S. Smirnov, D. Atnoor, A. Brown, N. Allen, J. Naylor, N. Stange-Thomann, K. DeArellano, R. Johnson, L. Linton, P. McEwan, K. McKernan, J. Talamas, A. Tirrell, W. Ye, A. Zimmer, R. D. Barber, I. Cann, D. E. Graham, D. A. Grahame, A Guss, R. Hedderich, C. Ingram-Smith, H. D. Kuettner, J. A. Krzycki, J. A. Leigh, W. Li, J. Liu, B. Mukhopadhyay, J. N. Reeve, K. Smith, T. Springer, L. A. Umayam, O. White, R. H. White, E. Conway de Macario, J. G. Ferry, K. F. Jarrell, H. Jing, A. J. L. Macario, I. Paulsen, M. Pritchett, K. R. Sowers, R. V. Swanson, S. H. Zinder, E. Lander, W. W. Metcalf, and B Birren: The genome of *Methanosarcina acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res.* 12, 532-542 (2002)

54. Deppenmeier U., A. Johann, T. Hartsch, R. Merkl, A. R. Schmitz, R. Martinez-Arias, A. Henne, A. Wiezer, S. Baumer, C. Jacobi, H. Brueggemann, T. Lienard, A. Christmann, M. Boemeke, S. Steckel, A. Bhattacharyya, A. Lykidis, R. Overbeek, H. P. Klenk, R. P. Gunsalus, H. J. Fritz & G. Gottschalk: The genome of *Methanosarcina mazei*: Evidence for lateral gene transfer between bacteria and archaea. *J Mol Microbiol Biotechnol* 4, 453-461 (2002)

55. Willison K. R.: Composition and function of the eukaryotic cytosolic chaperonin-containing TCP-1. In: Molecular chaperones and folding catalysis. Ed: B. Bukau. Harwood Academic Publishers, Australia. pp.555-571 (1999)

56. Archibald J. M., C. Blouin & W. F. Doolittle: Gene duplication and the evolution of group II chaperonins: Implications for structure and function. *J Struct Biol* 135, 157-169 (2001)

57. Archibald, J. M. & A. J. Roger: Gene duplication and gene conversion shape the evolution of archaeal chaperonins. *J Mol Biol* 316, 1042-1050 (2002)

58. Leroux M. R., M. Faendrich, D. Klunker, K. Siegers, A. N. Lupas, J. R. Brown, E. Schiebel, C. M. Dobson & F. U. Hartl: MtGimC, a novel archaeal chaperone related to the eukaryotic chaperonin cofactor GimC/prefoldin. *EMBO J* 18, 6730-6743 (1999)

59. Macario A. J. L., C. B. Dugan, M. Clarens & E. Conway de Macario: *dnaJ* in Archaea. *Nucl Acids Res* 21, 2773 (1993)

60. Macario A. J. L. & E. Conway de Macario: The archaeal molecular chaperone machine: peculiarities and paradoxes. *Genetics* 152, 1277-1283 (1999)

61. Conway de Macario E., C. B. Dugan & A. J. L. Macario: Identification of a *grpE* heat-shock gene homolog in the archaeon *Methanosarcina mazei*. *J Mol Biol* 240, 95-101 (1994)

62. Clarens M., A. J. L. Macario & E. Conway de Macario: The archaeal *dnaK-dnaJ* gene cluster: organization and expression in the methanogen *Methanosarcina mazei*. *J Mol Biol* 250, 191-201 (1995)

**Abbreviations:** CCT: chaperonin containing tcp-1 (tailess complex polypeptide-1); OTG: optimal temperature for growth; NAC: nascent chain-associated complex; BAG: Bcl2-associated athanogen; Hop: Hsp70-Hsp90 organizing protein; Hip: Hsp70 interacting protein; S: Southern; N: Northern; W: western

**Key Words:** Stress, Anti-Stress Mechanisms, Chaperoning Systems, Chaperones, Archaea, Chaperonins, Prefoldins, *Methanosarcina acetivorans*, *Methanosarcina mazeii* S-6, Hsp70(DnaK), Hsp40(DnaJ), GrpE, GroEL, GroES, Hsp60 chaperonins, thermosome, thermosome subunits, chaperonin subunits, Review

**Send correspondence to:** Alberto J. L. Macario, M.D., Wadsworth Center, Room B-749, New York State Department of Health, Empire State Plaza, P.O. Box 509, Albany, New York 12201-0509, USA, Tel.: 518-474-2781; Fax: 518-474-1213, E-mail: macario@wadsworth.org