Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Structural Biology 146 (2004) 79–89
Journal of Structural Biology
www.elsevier.com/locate/yjsbi

Chaperone networks in bacteria: analysis of protein homeostasis in minimal cells

Philip Wong and Walid A. Houry*

Department of Biochemistry, University of Toronto, 1 King's College Circle, Medical Sciences Building, Toronto, Ont., Canada M5S 1A8

Received 15 August 2003, and in revised form 30 October 2003

Abstract

The prevention of aberrant behavior of proteins is fundamental to cellular life. Protein homeostatic processes are present in cells to stabilize protein conformations, refold misfolded proteins, and degrade proteins that might be detrimental to the cell. Molecular chaperones and proteases perform a major role in these processes. In bacteria, the main cytoplasmic components involved in protein homeostasis include the chaperones trigger factor, DnaK/DnaJ/GrpE, GroEL/GroES, HtpG, as well as ClpB and the proteases ClpXP, ClpAP, HslUV, Lon, and FtsH. Based on recent genome sequencing efforts, it was surprising to find that the Mycoplasma, a genus proposed to include a minimal form of cellular life, do not contain certain major members of the protein homeostatic network, including GroEL/GroES. We propose that, in mycoplasmas, there has been a fundamental shift towards favoring processes that promote protein degradation rather than protein folding. The arguments are based on two different premises: (1) the regulation of stress response in Mycoplasma and (2) the unique characteristics of the Mycoplasma proteome.

Keywords: Chaperones; Proteases; Protein homeostasis; Mycoplasmas

1. Introduction

Protein homeostatic mechanisms are a complex set of processes that affect the levels and conformational stability of proteins in cells. Maintenance of proper protein homeostasis is essential for the viability and growth of cells. The proper functioning of proteins depends on their structural integrity; inherent instability and external stresses may cause proteins to misfold and aggregate. Protein homeostatic processes combat these problems by stabilizing proteins that carry out beneficial functions and by refolding or degrading proteins that are detrimental to the cell. Molecular chaperones and proteases perform a significant subset of these processes. Molecular chaperones promote protein folding and prevent protein misfolding and aggregation, while certain proteases function primarily to degrade improperly folded proteins (Dougan et al., 2002a; Hartl and Hayer-Hartl, 2002). The main cytoplasmic chaperones identified so far include trigger factor, DnaK/DnaJ/GrpE, GroEL/GroES, HtpG, and ClpB, while the main cytoplasmic proteases include ClpXP, ClpAP, HslUV, Lon, and FtsH.

In *Escherichia coli*, trigger factor (Tig) is a modular ATP-independent chaperone consisting of an N-terminal ribosomal-binding domain, an FKBP-type peptidyl-prolyl isomerase (PPIase) middle domain, and a C-terminal domain implicated in enhancing the binding of Tig to unfolded proteins. Tig interacts with nascent chains and binds closely to the nascent chain exit site of the ribosomal complex (Blaha et al., 2003; Maier et al., 2003). Data suggest that trigger factor is involved in the prevention of the misfolding and aggregation of nascent chains as they are translated by the ribosome (Hartl and Hayer-Hartl, 2002).

Unlike trigger factor, DnaK is an ATP-dependent chaperone, consisting of an N-terminal ATPase domain, a substrate-binding domain, and a C-terminal regulatory domain. DnaK functions with the cochaperone DnaJ, and a nucleotide exchange factor GrpE. DnaK recognizes hydrophobic sequences in extended polypeptide chains. Through cycles of binding and release regulated by ATP, DnaJ, and GrpE, DnaK facilitates

* Corresponding author. Fax: +1-416-978-8548.
E-mail address: walid.houry@utoronto.ca (W.A. Houry).

1047-8477/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2003.11.006

substrate folding to the native state (Hartl and Hayer-Hartl, 2002). Like Tig, DnaK also binds to nascent chains (Deuerling et al., 1999; Teter et al., 1999), implying an overlap in the function of the two chaperones.

The chaperonin GroEL is a homo-oligomer of 14 subunits that arrange themselves into two heptameric rings, resulting in a cylindrical structure with two large cavities (Hartl and Hayer-Hartl, 2002). Each subunit of GroEL can be divided into apical, intermediate, and equatorial domains. The stacking interface between the two rings occurs at the equatorial domains. The intermediate domain forms a hinge region between the equatorial and the apical domains. Substrates at an intermediate stage of folding bind to the apical domain at the top of the GroEL cylinder. Substrates are then displaced into the central cavity by the binding of GroES to GroEL in an ATP-dependent manner. Protected inside this cavity, substrates can fold while minimizing the possibility of aggregation. Subsequently, GroES dissociates from GroEL, and substrates are then released into the cytoplasm (Keskin et al., 2002). The GroEL/GroES functional cycle is regulated by ATP binding and hydrolysis.

Little is known about the role of HtpG (Hsp90) in bacteria, even though this chaperone is highly expressed upon heat shock. In *E. coli*, strains devoid of HtpG behave like wild-type strains and show no specific phenotypes (Bardwell and Craig, 1988).

The Clp family of ATPases can be divided into two classes according to the number of nucleotide-binding domains (NBD) in the respective proteins. ClpA and ClpB contain two NBDs; ClpX and HslU, only one. ClpX and ClpA associate with the serine protease ClpP, while HslU associates with the threonine protease HslV (Dougan et al., 2002b). ClpX, ClpA, and HslU are thought to be involved in the recognition, unfolding, and translocation of substrates into the respective proteases (Wickner et al., 1999). On the other hand, ClpB has not been found to associate with a protease subunit but, instead, appears to cause resolubilization of protein aggregates in cooperation with the DnaK system (Mogk et al., 2001).

Lon and FtsH are two other ATP-dependent proteases; Lon is a DNA-binding protease that degrades regulatory and abnormal proteins (Fu et al., 1997; Takaya et al., 2002), and FtsH is a hexameric membrane-anchored protease that degrades both membrane and cytoplasmic proteins (Herman et al., 2003). The mechanism of substrate recognition by Lon and FtsH is not yet known.

Chaperone and protease families are highly conserved across genomes, suggesting that their functions are essential for cellular life (Jordan et al., 2002). However, we observed that, as more genomes are being sequenced, many chaperones and proteases are absent from the Mycoplasma genus. Mycoplasmas are generally characterized by their small genome size—usually less than 1 Mbp—and the lack of a cell wall (Razin, 1995). Their small genome sizes have raised questions as to whether this genus contains an organism with the smallest gene set required for life (Peterson and Fraser, 2001). Consequently, the study of protein homeostatic mechanisms in mycoplasmas may yield new insights into the role of chaperones and proteases in these processes. In this speculative paper, we suggest that protein homeostasis in mycoplasmas has shifted towards favoring processes that lead to protein degradation rather than protein folding. We present several arguments to support this hypothesis.

2. Materials and methods

The existence or absence of proteins in the different genomes was detected from the available annotations and confirmed by BLAST (Altschul et al., 1997). Detection of COGs of known chaperone and protease families across genomes, and the building of the *E. coli* chaperone network, were accomplished using STRING (June, 2003 version) (von Mering et al., 2003). The network was built from ClpX and DnaK seed COGs. Additional COGs were connected to these COGs whenever the confidence threshold of association from STRING was high—75% or greater. The process was repeated until no more COGs could be added. The 75% confidence provided a threshold which excluded many proteins that are not chaperones or proteases.

The search for proteins in mycoplasmas with the same SCOP or CATH architecture as that of GroEL was done using data obtained from PEDANT (Frishman et al., 2003) or Gene3D (Buchan et al., 2003), respectively. The search for putative CIRCEs in Mycoplasma genomes was carried out using the sequences TTAGCAC-N{10–18}-GTGCTAA and TTAG CA-N{12–20}-TGCTAA, where N is any nucleotide. The number of allowed substitutions was between 0 and 2, implemented in a Perl script. Genome sequences were obtained from GenBank (Wheeler et al., 2003).

3. Results and discussion

3.1. Distribution of chaperones and proteases across genomes

Because completely sequenced genomes of a large number of bacterial and archaeal organisms are available, the distribution of chaperone and protease families across these organisms can be assessed. Clusters of orthologous groups (COGs) (Tatusov et al., 1997) can be used to rapidly search for these families across genomes since proteins within a COG are predicted to function

similarly. Fig. 1 shows the distribution of 14 chaperone/protease COGs across 67 bacterial (numbered 1–67) and 16 archaeal (numbered 68–83) genomes.

The figure clearly shows that none of the major chaperone and protease families considered are present in all genomes. The COGs for the chaperones Tig and DnaK/DnaJ/GrpE, and the protease FtsH, are the only absolutely conserved COGs in the listed bacteria. FtsH COG is absent only in listed archaea. However, for the 83 genomes, GroEL is the most conserved chaperone, and it is absent in only two mycoplasmas (*Ureaplasma urealyticum* and *Mycoplasma pulmonis*). Note that in archaea, it is the archaeal thermosome and proteasome that function like *E. coli* GroEL and HslV, respectively. These archaeal proteins are closer to the eukaryotic CCT and proteasome than to the bacterial GroEL and HslV, respectively.

3.2. Building the chaperone network

To reveal possible functional associations between the different chaperones and proteases, we decided to build a chaperone network. In the past decade, tools to cluster functionally related proteins have been developed based on: (1) phylogenetic profiling, (2) gene fusion, and (3) gene neighborhood searches. These methods provide a good prediction of possible functional association between proteins (von Mering et al., 2003), and we apply them in combination using the STRING tool (von Mering et al., 2003) to determine whether we can connect the major chaperone and protease families using a seed protein COG. The network is built by adding new COGs to these seed COGs if a certain threshold of association between the COGs is reached. The process is repeated until no more COGs can be added. Results with a threshold of 75% confidence show that major components of the protein homeostatic network can be connected (connected by arrows in Fig. 2).

The advantage of such a graph is that it allows predictions of functional relationships between proteins by means of transitivity: when one knows that protein A is predicted to be functionally associated with protein B, and that protein B is predicted to be functionally associated with protein C, one may predict that A and C are functionally associated. For instance, a functional relationship between GroEL and trigger factor is predicted based on this network, a prediction which agrees with published experimental results. For example, the ability of GroEL to form complexes with a model substrate protein can be enhanced upon Tig overproduction (Kandror et al., 1997). Note that, from a purely bioinformatics perspective, the relationship between GroEL and Tig cannot be established based solely on phylogenetic profiling, gene fusion, or gene neighborhood search information; rather, a combination of the methods is required—at the employed threshold.

Fig. 2 shows that the four proteins ClpX, ClpP, Lon, and Tig form a 'complete graph' whereby each protein is directly connected to the other three proteins. When multiple cycles of transitive relationships exist in complete graphs, this reinforces the predicted associations. Cyclic reinforcement has been exploited successfully with respect to gene neighborhood relationships to confirm gene associations (Kolesov et al., 2001). The complete association graph obtained for ClpX, ClpP, Lon, and Tig, suggests that either a substrate or one or more biochemical processes common to many organisms require all four proteins.

It should be noted that HtpG could not be connected to any of the chaperones or proteases in Fig. 2; nor could a connection be made between the DnaK chaperone system and other systems. The latter observation is contrary to experimental data showing, for example, that *E. coli* DnaK and Tig have overlapping functions as well as overlapping substrate-binding specificities (Deuerling et al., 1999; Deuerling et al., 2003; Teter et al., 1999). Furthermore, no connections are observed between ClpB and the DnaK system in Fig. 2, although it has been proposed that these two *E. coli* chaperones form a bichaperone network involved in the solubilization and refolding of stable protein aggregates (Goloubinoff et al., 1999). Since Fig. 2 incorporates all sequenced bacterial genomes, it implies that such a bichaperone system is not conserved across genomes. However, if the threshold of the COG associations is lowered to 20%, ClpB can be connected to DnaK. Hence, by lowering the confidence threshold, weaker associations can be revealed.

Since genes that function similarly do not necessarily occupy the same COG, this network can also be extended by searching for protein sequences similar to those already connected (connected by lines in Fig. 2). For example, connections by this approach can be made between ClpX and FtsH, DnaJ and its mutation suppressing paralogue CbpA (Ueguchi et al., 1995), Lon and the DNA repair protein RadA (Beam et al., 2002), and DnaK and MreB/YegD/HscA/HscC. YegD, HscA, and HscB are members of the Hsp70 family, while MreB is a rod-forming protein; therefore, Fig. 2 can be expanded using the proposed approach.

The importance of the network shown in Fig. 2 is that it reinforces the concept of functional associations or functional overlap between the different chaperones and proteases in the cell.

3.3. Missing chaperones and proteases in mycoplasmas

As seen in Fig. 1, the distribution of chaperones and proteases can readily distinguish bacteria from archaea. Archaea appear to be missing many of the COGs listed

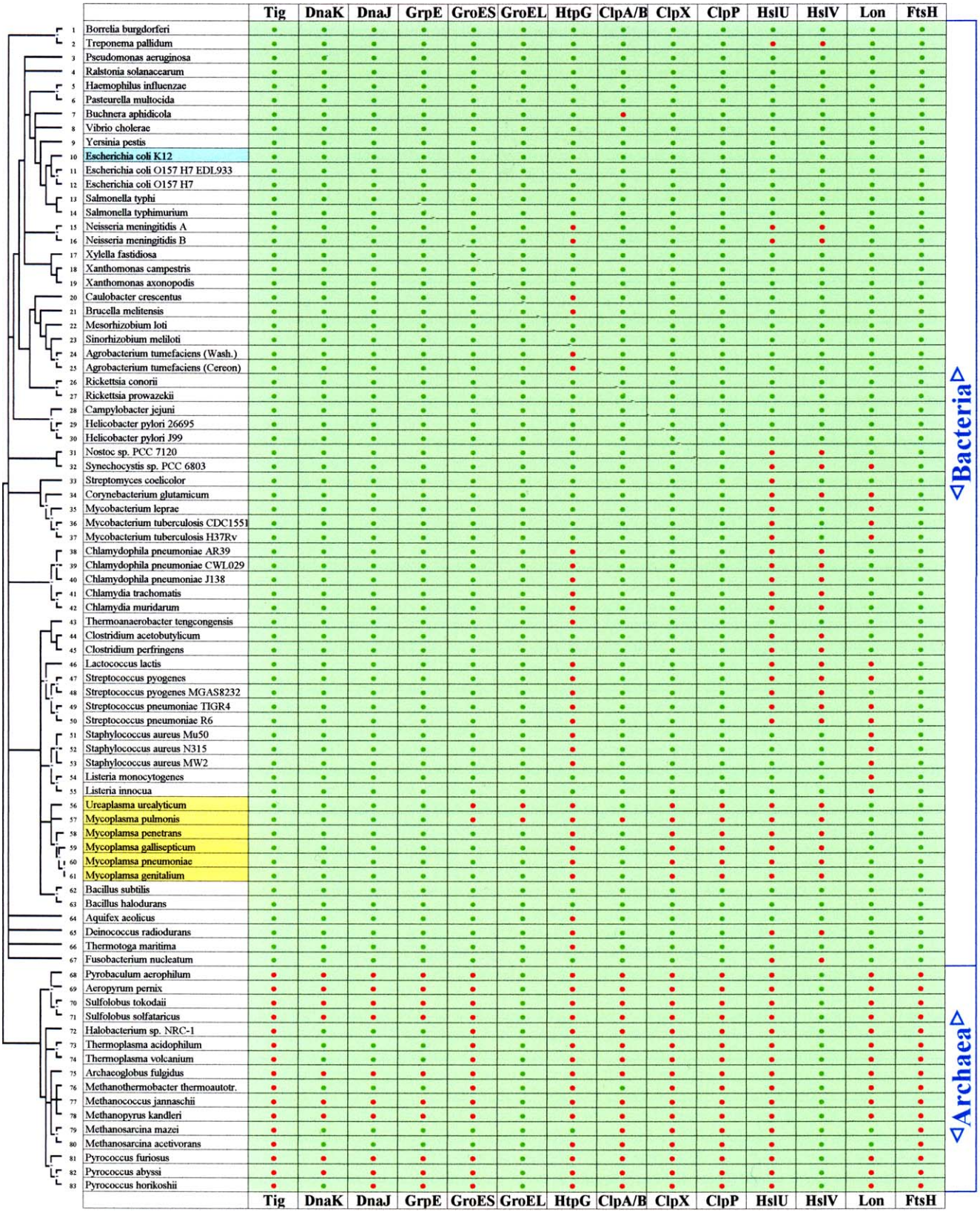

Fig. 1. Distribution of major chaperones and proteases. Green and red dots represent existence or absence, respectively, of a protein COG for each of the 83 organisms listed on the left panel. Representative members of each COG from *E. coli* are listed at the bottom. The diagram was modified from output obtained using the STRING tool (von Mering et al., 2003). The phylogenetic tree is according to STRING except that the Mycoplasma branch was rearranged based on EF-Tu phylogeny (see Fig. 5).

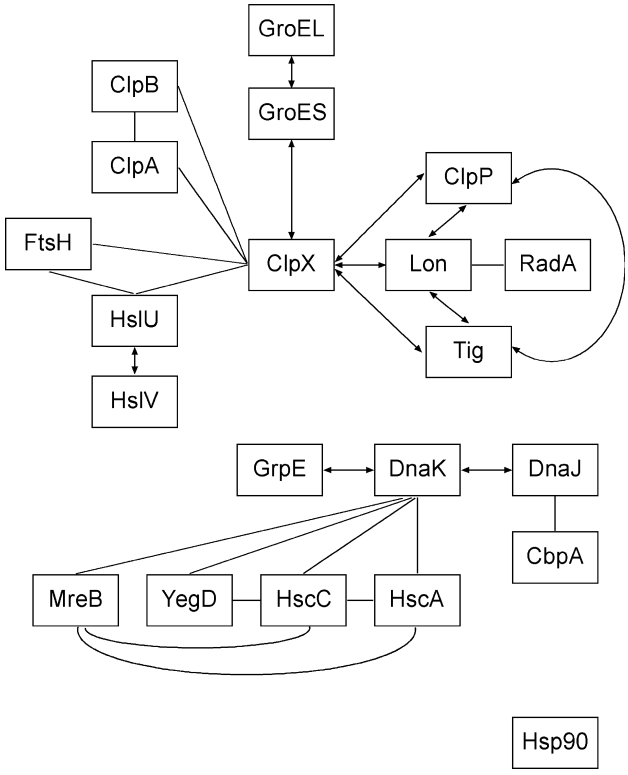

Fig. 2. The chaperone and protease network. The predicted functional linkage between major cytoplasmic *E. coli* chaperones and proteases. A ↔ B indicates that A and B are linked by STRING with a confidence of 75% or greater. A — B indicates that a bidirectional BlastP (Altschul et al., 1997) *E*-value between A and B is less than $1e^{-5}$.

in Fig. 1. Furthermore, the mycoplasmas (highlighted in yellow in Fig. 1) can also be readily separated from the other bacteria, based on the distribution of their chaperones and proteases. Among listed bacteria, the mycoplasmas contain the smallest number of chaperones and proteases.

One distinguishing characteristic of the Mycoplasma genomes is that the genes for ClpX, ClpP, HslU, and HslV are absent in all mycoplasmas sequenced so far (Fig. 1). Furthermore, ClpA/B is also missing in *M. pulmonis*. Also absent from all sequenced Mycoplasma genomes are the following: CtsR, a negative regulator of Clp proteins in many gram positive bacteria (Derre et al., 1999); ClpS, a modulator of ClpAP substrate recognition (Dougan et al., 2002c); and SspB, a stimulator of ClpXP degradative activity (Wah et al., 2002).

In bacterial cells, incomplete protein fragments generated due to ribosomal-stalling are C-terminally tagged with a hydrophobic SsrA sequence (AANDENYALAA in *E. coli*) by a tmRNA (Keiler et al., 1996). In *E. coli*, the translated tag targets the protein fragments for degradation by FtsH, ClpXP, or ClpAP (Gottesman et al., 1998; Herman et al., 1998). A similar system likely exists in mycoplasmas since the tmRNA is present in those genomes (Williams, 2002). Furthermore, proteins thought to associate or function with the tmRNA—such as SmpB, phosphoribosyl pyrophosphate synthase, RNase R, and YfbG-like methionine tRNA formyltransferase (Karzai et al., 2000; Karzai and Sauer, 2001)—are also encoded in all sequenced mycoplasma genomes. Since ClpXP and ClpAP systems are missing from mycoplasmas, it is likely that FtsH could replace the function of these systems in degrading at least some of the SsrA-tagged proteins (Karzai et al., 2000).

The loss of the HslUV system in mycoplasmas indicates, perhaps, that Lon or FtsH carry a function similar to that of HslUV in these organisms. In *E. coli*, it has been observed that *lon* mutants can be suppressed by the overproduction of HslUV protease; furthermore, there is an overlap of substrate specificity between Lon and HslUV (Wu et al., 1999). This suggests that, in mycoplasmas, Lon is the likely candidate to replace the function of the HslUV system.

Another distinguishing feature of the Mycoplasma genomes is the striking observation that *U. urealyticum* and *M. pulmonis* are the only two organisms among the 83 listed in Fig. 1 that do not have GroEL/GroES. We further verified that a protein with a GroEL-like structure is not present in these two organisms, using CATH-assigned domains from Gene3D (Buchan et al., 2003) as well as the SCOP database (Lo Conte et al., 2002) as implemented in PEDANT (Frishman et al., 2003). Transposon mutagenesis of GroEL in *Mycoplasma pneumoniae* and *Mycoplasma genitalium* suggests that GroEL is not essential for the viability of these two organisms (Hutchison et al., 1999). In *Mycoplasma capricolum*, which is not yet completely sequenced, results of 1D-PAGE analysis of pulse-labeled proteins from heat-shocked cells suggest that the chaperonin is missing from this organism as well (Table 1).

The loss of GroEL in mycoplasmas is an intriguing finding since GroEL was thought to be essential for viability. For example, Fayet has shown that in *E. coli*, the GroEL/GroES system is the only essential chaperone system under all growth temperatures (Fayet et al.,

Table 1
GroEL is missing or not essential in mycoplasmas

Mycoplasma species	Presence of GroEL
M. genitalium	Present but not essential (Hutchison et al., 1999)[a]
M. pneumoniae	Present but not essential (Hutchison et al., 1999)[a]
M. gallisepticum	Present (Scherm et al., 2002)[b]
M. penetrans	Present (Sasaki et al., 2002)
M. hyopneumoniae	Present (Scherm et al., 2002)[b]
M. capricolum	Missing (Dascher and Maniloff, 1992)[c]
M. pulmonis	Missing (Chambaud et al., 2001)
U. urealyticum	Missing (Glass et al., 2000)

[a] Not essential as shown by transposon mutagenesis.
[b] As shown by immunoblotting.
[c] 1D-PAGE of pulse-labeled proteins after heat shock did not reveal the presence of GroEL.

Table 2
Chaperones downstream of CIRCE-like sequences in mycoplasmas

Species	CIRCE-like sequences (consensus: TTAGCACTC-N9-GAGTGCTAA)			ORFs	Location of ORFs (strand)	Distance between CIRCE and ORFs (bp)
M. gallisepticum	651401	**TTAGCTAGAAGTGTGGCTAATGAATTAA**	651429	GroES	649868–650161(−)	1240[a]
				GroEL	648264–649871(−)	
M. genitalium	496987	**TTAGCGCTAATATCAGATAAATTTAA**	497013	GroES	496628–496960(−)	27
				GroEL	494984–496615(−)	
M. pneumoniae	697678	**TTTGCACTTTAAGCATTTAACTGCTAA**	697705	GroES	697326–697676(−)	11
				GroEL	695692–697323(−)	
M. penetrans	124473	**TTAGCAGTTATAATAGTTAAGTGATAA**	124500	GroES	124534–124806(+)	34
				GroEL	124836–126473(+)	
M. gallisepticum	739626	**TTAGCAATCTACTTGCAAAAGTGCTAA**	739653	DnaK	737780–739567(−)	59
M. genitalium	376726	**TTAGCACTTTTAGTGTTTGAGTGCTAA**	376753	DnaK	374918–376705(−)	26
M. penetrans	1271763	**TTAGCATAAAAGAGTTGAGAGTGCTAA**	1271790	DnaK	1269938–1271803(−)	40
M. pneumoniae	523629	**TTAGCACTCAAACGCTAAAAGTGCTAA**	523656	DnaK	521837–523624(−)	5
M. pulmonis	255797	**TTAGCACTTTATTCATTTGAGTGCTAA**	255824	DnaK	255963–257762(+)	139
U. urealyticum	385711	**TTAGCACTCAATATTGACAAGTGCCAA**	385738	DnaK	385804–387615(+)	66
M. gallisepticum	668480	**TTAGCACTCTAATAGCTCAAGTGCTAA**	668507	ClpB	666315–668468(−)	12
M. genitalium	453770	**TTAGCACTCGAATGCCTTGAGTGCTAA**	453797	ClpB	451613–453757(−)	13
M. penetrans	490509	**TTAGCACTCCTATATTTAATCTGCTAA**	490536	ClpB	488365–490482(−)	27
M. pneumoniae	653965	**TTAGCACTCAAGCCATTCGAGTGCTAA**	653992	ClpB	651802–653949(−)	16
U. urealyticum	451467	**TTAGCACTCATTATTTTAAAGTGCTAA**	451494	ClpB	451524–453638(+)	30
M. gallisepticum	516411	**TTAGCACTTGAATTCTTAGAGTGCTAA**	516438	Lon	516459–518939(+)	21
M. genitalium	283322	**TTAGCACTCAAAGCTTGTGAGTGCTAA**	283349	Lon	283395–285782(+)	46
M. penetrans	890414	**TTATCACTTAAATATTTAAAGTGCTAA**	890441	Lon	888037–890382(−)	32
M. pneumoniae	390253	**TTAGCACTCAAAGTATTAGAGTGCTAA**	390280	Lon	390328–392715(+)	48
M. pulmonis	636422	**ATAGCACTTTATTCAAAAGAGTGCCAA**	636449	Lon	633790–636318(−)	104
U. urealyticum	394936	**TTAGCACTCGTAATAATAGTTTGCTAA**	394963	Lon	395023–397398(+)	60

[a] GroES/GroEL may be part of an upstream operon containing pykF and pykA.

1989); it is estimated to be involved in folding approximately 10% of newly translated proteins (Ewalt et al., 1997). GroEL homologues are also known to be essential for *Saccharomyces cerevisiae* (Esser et al., 1999) and *Bacillus subtilis* (Kobayashi et al., 2003) viability.

It is possible that Tig or the DnaK/DnaJ/GrpE chaperones could replace some of the missing function of GroEL. However, as has been clearly demonstrated by extensive biophysical, biochemical, and genetic studies (Hartl and Hayer-Hartl, 2002; Saibil et al., 2001), the functions and substrate specificities of Tig and DnaK are fundamentally different from those of GroEL. Tig and DnaK bind proteins in an extended conformation at early stages of folding, while GroEL recognizes a compact folding intermediate. Furthermore, GroEL/GroES is the only chaperone system that can encapsulate its substrates in an isolated cavity to promote proper folding. Therefore, it is unlikely that other chaperones in Mycoplasma can 'absolutely' replace GroEL function.

The chaperones and proteases lost in Mycoplasma play an important role in protein homeostasis in other bacteria such as *E. coli*, indicating, perhaps, that the mechanisms by which mycoplasmas achieve their protein homeostasis is quite different from that found in *E. coli*. We propose that there has been a fundamental shift in mycoplasmas towards favoring processes that promote protein degradation rather than protein folding. The proposal is based on two different premises: (1) the regulation of stress response in Mycoplasma and (2) the characteristics of the Mycoplasma proteome.

3.4. Possible evidence for the shift towards proteolysis in *Mycoplasma*

3.4.1. Regulation of stress response in *Mycoplasma*

If the functions of missing chaperones and proteases in Mycoplasma are to be replaced by other mechanisms in the protein homeostatic network as highlighted by

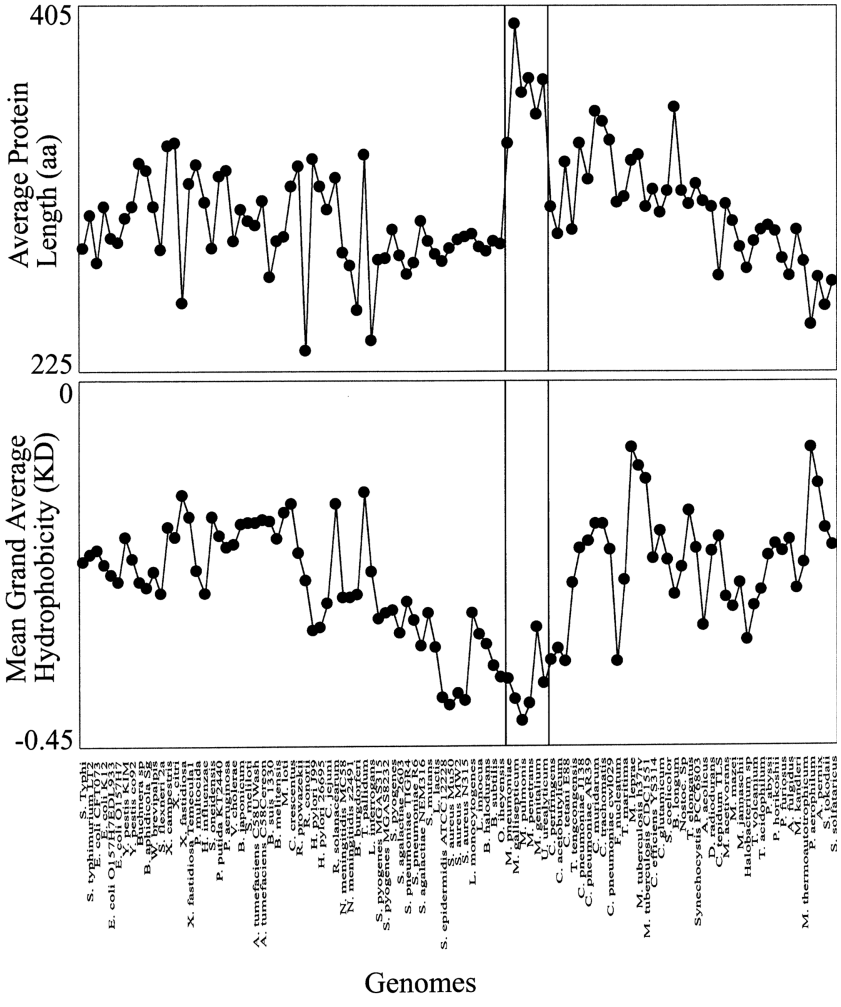

Fig. 3. Length and hydrophobicity of cytoplasmic proteins in prokaryotes. The distributions are those of cytoplasmic proteins. Proteins predicted to contain two or more transmembrane helices by TMHMM2 (Krogh et al., 2001) were excluded. Top panel shows the average protein length across genomes; bottom panel shows the Mean Grand Average Hydrophobicity (GRAVY) across genomes. GRAVY was calculated using the Kyte and Doolittle scale (KD) (Kyte and Doolittle, 1982) for each protein, and the mean was obtained.

Fig. 2, one would expect some commonality in their regulation. Heat regulation at CIRCE (HrcA) is a negative regulator of GroEL in gram positive bacteria. HrcA binds to the control inverted repeat chaperone expression (CIRCE) sequence that has the consensus TTAGCACTC-N9-GAGTGCTAA, where N is any base (Narberhaus, 1999). The CIRCE sequence is upstream of the *groESL* operon in many gram positive bacteria (Narberhaus, 1999) but can be found upstream of other genes as well, and it is also present in other bacterial lineages (Chastanet et al., 2003; Narberhaus, 1999; Servant and Mazodier, 2001).

Regular expression searches for CIRCE-like sequences in mycoplasmas were conducted (see Section 2), and a large number of heat shock-related genes were found downstream of these sequences (Table 2). A full list of potential CIRCE-regulated proteins in mycoplasmas can be downloaded at http://biochemistry.utoronto.ca/houry/Labpage/Bioinfopage.html. Among these genes are those encoding for DnaK, ClpB, and Lon protease (Table 2). Notably, no CIRCE consensus sequence was found upstream of FtsH. Common regulation of ClpB, DnaK, Lon, and GroEL could support overlapping functionality in mycoplasmas. Thus, the

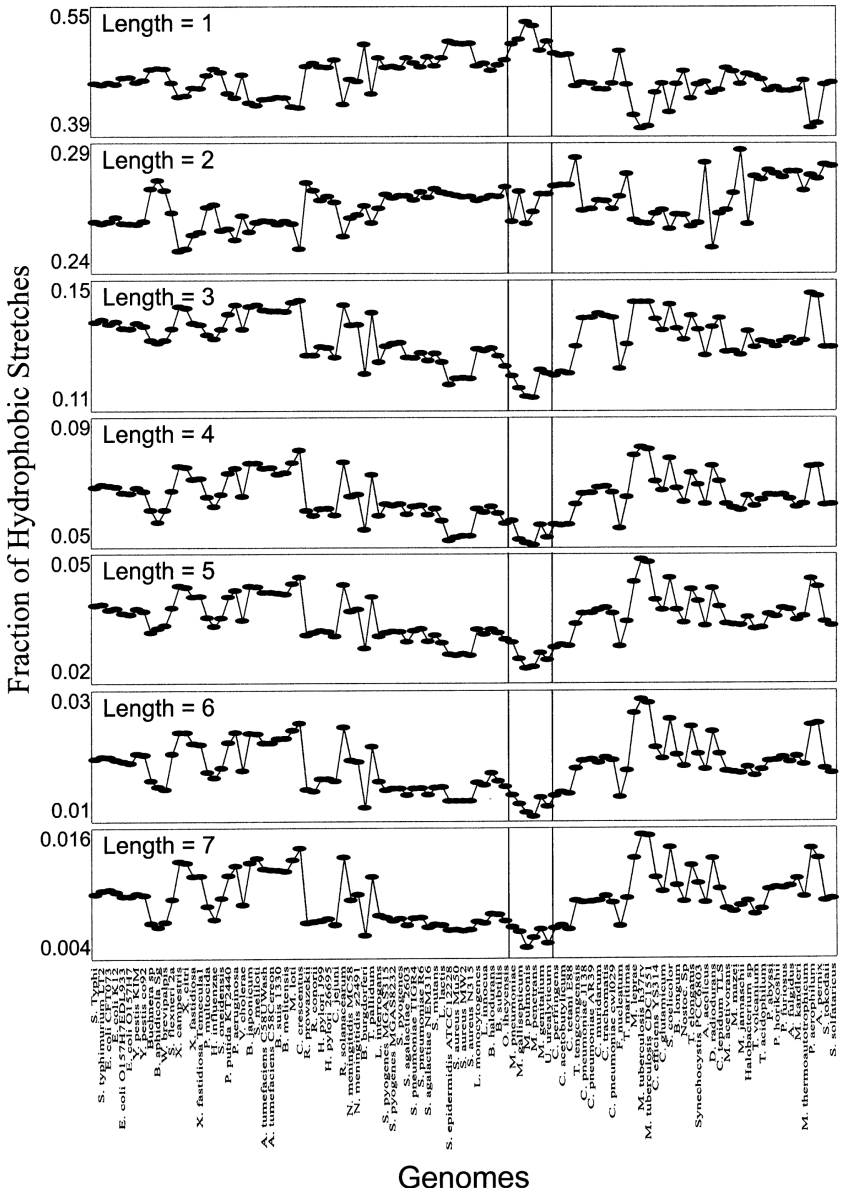

Fig. 4. Distribution of hydrophobic stretches in prokaryotes. Amino acids A, C, F, G, I, L, M, P, V, W, and Y were considered hydrophobic residues. A hydrophobic stretch is a contiguous sequence of such residues. The fraction of hydrophobic stretches of length '*i*' is defined as the number of stretches of length '*i*' encoded in the genome of the given organism divided by the total number of all hydrophobic stretches of length one or more residues in that organism. Only those cytoplasmic proteins predicted to contain less than two transmembrane helices by TMHMM2 (Krogh et al., 2001) were included.

function of GroEL that is absent in some mycoplasmas may be partially replaced by the action of ClpB and DnaK chaperones, and the Lon protease. Consequently, this may result in a net increase in degradative—rather than folding—activities in these organisms.

3.4.2. Characteristics of the Mycoplasma proteome

A high mutation rate is a plausible strategy by parasites to avoid detection by host defenses. Comparison of bacterial 16S RNA sequences indicates that mycoplasmas evolve 50% faster than related organisms in the *Lactobacillus* group (Maniloff, 1992). With mutation of the genome, it is expected that the proteome mutates as well. A likely consequence of high mutation rates is the destabilization of proteins. Indeed, a recent study in which protein stability was estimated from the thermodynamic parameters of 11 protein families derived from 40 bacterial species suggests that obligate pathogens, including the mycoplasmas, have more unstable proteins than have other bacteria (van Ham et al., 2003).

Since instability and folding difficulty are more common in long proteins (Chakravarty and Varadarajan, 2000), we compared the Mycoplasma cytoplasmic protein lengths with those of other prokaryotes. Results shown in Fig. 3, upper panel, indicate that Mycoplasma protein lengths are amongst the highest for prokaryotes sequenced so far—a mean of 364 amino acids for six sequenced mycoplasmas compared to a mean of 302 amino acids for 104 sequenced prokaryotes, further supporting the proposal that mycoplasmas have more unstable proteins than have other bacteria.

To tolerate high instability in their proteins, organisms require a high protein homeostatic capacity. For example, the abundant expression of GroEL, which was shown to be a mutational buffer (Fares et al., 2002), has been suggested as aiding the symbiont Buchneras in tolerating their high protein instability (van Ham et al., 2003). Mycoplasmas' apparent loss—or impending loss—of GroEL may indicate that mycoplasmas employ proteolysis rather than folding to buffer for protein instability. However, for such a strategy to be successful, protein aggregation levels must be kept low since, once large protein aggregates are formed, these aggregates cannot be effectively degraded. Indeed, the hydrophobicity of Mycoplasma cytoplasmic proteins is amongst the lowest of sequenced species (Fig. 3, lower panel). Furthermore, the fraction of hydrophobic stretches having residue length greater than 2 in mycoplasmas is also amongst the lowest compared with those in other species (Fig. 4). Therefore, although Mycoplasma proteins may be unstable, it is possible that, with less hydrophobic amino acids and with reduced numbers of long hydrophobic stretches, there is a reduced chance that unfolding or misfolding of Mycoplasma proteins will lead to aggregation. Hence, proteolysis would be an effective way to deal with these proteins.

The concept that increased degradative activities might reduce the requirement for folding activities has recently been demonstrated in yeast (Friant et al., 2003). It was shown that overexpression of ubiquitin, which targets proteins for degradation by the proteasome, can allow the cells to tolerate heat shock without the induction of heat shock genes.

3.5. Consequences of the loss of GroEL in mycoplasmas

What is the advantage of losing a chaperone such as GroEL? For parasites, host recognition of stable phenotypes may be a significant driving force for loss of genes that express or maintain those phenotypes. This reasoning may explain why certain mycoplasmas have lost the chaperonin—it is immunogenic (Dascher and Maniloff, 1992). Hence, mutations leading to the loss of GroEL can be viewed as an adaptation to host defenses.

Besides allowing parasites to exploit new niches, the loss of GroEL may also affect the evolution of proteins. EF-Tu is a possible candidate for such a protein since the absence or presence of GroEL in mycoplasmas can largely be distinguished by EF-Tu-based phylogeny (Kamla et al., 1996) (Fig. 5), an alternative classification to the classical 16S RNA-based phylogeny (Maniloff, 1992). Based on the position of two mycoplasmas that lack GroEL, *M. pulmonis* and *U. urealyticum*, we predict that *Mycoplasma hominis* and *Mycoplasma mycoides* subsp. *mycoides*, whose complete genome sequences are not yet available, have also lost the chaperonin. Our prediction can be verified when the complete genome sequences of these organisms become available.

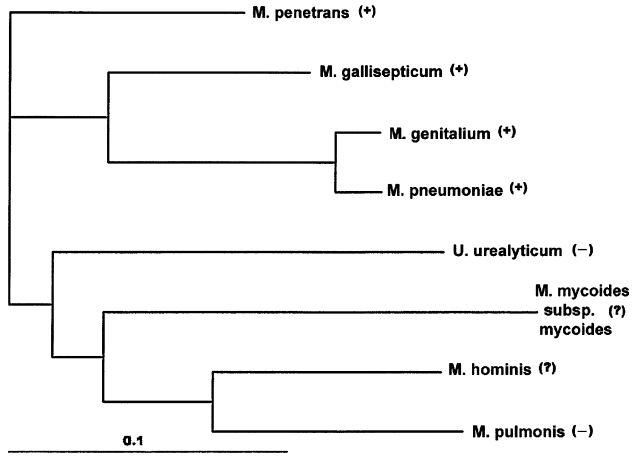

Fig. 5. Phylogram of mycoplasmas based on EF-Tu amino acid sequences. Evolutionary distance is based on the neighbor joining method (Saitou and Nei, 1987) as implemented in CLUSTALW (Chenna et al., 2003). The black bar represents 10% sequence divergence. (+) indicates that a strain of the species has been identified as containing GroEL. Strains of *M. pulmonis* and *U. urealyticum* lack GroEL and are labeled (−). No published data are available on whether there is GroEL in *M. mycoids* subsp. *mycoides* or *M. hominis*; these strains are labeled (?).

4. Concluding remarks

The birth of the Mycoplasma genus is thought to have resulted from massive gene loss; however, the driving force remains unclear (Moran, 2002). In several mycoplasmas, this gene loss included the loss of the essential chaperone system GroEL/GroES, among other chaperones and proteases (Fig. 1). In this paper, we have argued from different perspectives that there has been a fundamental shift towards proteolysis rather than protein folding as a means to maintain protein homeostasis in Mycoplasma.

Acknowledgment

This work was supported by a Canadian Institutes of Health Research Grant to W.A.H.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Bardwell, J.C., Craig, E.A., 1988. Ancient heat shock gene is dispensable. J. Bacteriol. 170, 2977–2983.

Beam, C.E., Saveson, C.J., Lovett, S.T., 2002. Role for RadA/Sms in recombination intermediate processing in *Escherichia coli*. J. Bacteriol. 184, 6836–6844.

Blaha, G., Wilson, D.N., Stoller, G., Fischer, G., Willumeit, R., Nierhaus, K.H., 2003. Localization of the trigger factor binding site on the ribosomal 50S subunit. J. Mol. Biol. 326, 887–897.

Buchan, D.W., Rison, S.C., Bray, J.E., Lee, D., Pearl, F., Thornton, J.M., Orengo, C.A., 2003. Gene3D: structural assignments for the biologist and bioinformaticist alike. Nucleic Acids Res. 31, 469–473.

Chakravarty, S., Varadarajan, R., 2000. Elucidation of determinants of protein stability through genome sequence analysis. FEBS Lett. 470, 65–69.

Chambaud, I., Heilig, R., Ferris, S., Barbe, V., Samson, D., Galisson, F., Moszer, I., Dybvig, K., Wroblewski, H., Viari, A., Rocha, E.P., Blanchard, A., 2001. The complete genome sequence of the murine respiratory pathogen *Mycoplasma pulmonis*. Nucleic Acids Res. 29, 2145–2153.

Chastanet, A., Fert, J., Msadek, T., 2003. Comparative genomics reveal novel heat shock regulatory mechanisms in *Staphylococcus aureus* and other Gram-positive bacteria. Mol. Microbiol. 47, 1061–1073.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., Thompson, J.D., 2003. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 31, 3497–3500.

Dascher, C.C., Maniloff, J., 1992. Heat shock response. In: Maniloff, J., McElhaney, R.N., Finch, L.R. (Eds.), Mycoplasmas: Molecular Biology and Pathogenesis. American Society for Microbiology, Washington, DC, p. 353.

Derre, I., Rapoport, G., Msadek, T., 1999. CtsR, a novel regulator of stress and heat shock response, controls Clp and molecular chaperone gene expression in gram-positive bacteria. Mol. Microbiol. 31, 117–131.

Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A., Bukau, B., 1999. Trigger factor and DnaK cooperate in folding of newly synthesized proteins. Nature 400, 693–696.

Deuerling, E., Patzelt, H., Vorderwulbecke, S., Rauch, T., Kramer, G., Schaffitzel, E., Mogk, A., Schulze-Specking, A., Langen, H., Bukau, B., 2003. Trigger factor and DnaK possess overlapping substrate pools and binding specificities. Mol. Microbiol. 47, 1317–1328.

Dougan, D.A., Mogk, A., Bukau, B., 2002a. Protein folding and degradation in bacteria: to degrade or not to degrade? That is the question. Cell. Mol. Life Sci. 59, 1607–1616.

Dougan, D.A., Mogk, A., Zeth, K., Turgay, K., Bukau, B., 2002b. AAA+ proteins and substrate recognition, it all depends on their partner in crime. FEBS Lett. 529, 6–10.

Dougan, D.A., Reid, B.G., Horwich, A.L., Bukau, B., 2002c. ClpS, a substrate modulator of the ClpAP machine. Mol. Cell 9, 673–683.

Esser, K., Scholle, B., Michaelis, G., 1999. Disruption of six open reading frames on chromosome X of *Saccharomyces cerevisiae* reveals a cluster of four essential genes. Yeast 15, 921–933.

Ewalt, K.L., Hendrick, J.P., Houry, W.A., Hartl, F.U., 1997. In vivo observation of polypeptide flux through the bacterial chaperonin system. Cell 90, 491–500.

Fares, M.A., Ruiz-Gonzalez, M.X., Moya, A., Elena, S.F., Barrio, E., 2002. Endosymbiotic bacteria: groEL buffers against deleterious mutations. Nature 417, 398.

Fayet, O., Ziegelhoffer, T., Georgopoulos, C., 1989. The groES and groEL heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. J. Bacteriol. 171, 1379–1385.

Friant, S., Meier, K.D., Riezman, H., 2003. Increased ubiquitin-dependent degradation can replace the essential requirement for heat shock protein induction. EMBO J. 22, 3783–3791.

Frishman, D., Mokrejs, M., Kosykh, D., Kastenmuller, G., Kolesov, G., Zubrzycki, I., Gruber, C., Geier, B., Kaps, A., Albermann, K., Volz, A., Wagner, C., Fellenberg, M., Heumann, K., Mewes, H.W., 2003. The PEDANT genome database. Nucleic Acids Res. 31, 207–211.

Fu, G.K., Smith, M.J., Markovitz, D.M., 1997. Bacterial protease Lon is a site-specific DNA-binding protein. J. Biol. Chem. 272, 534–538.

Glass, J.I., Lefkowitz, E.J., Glass, J.S., Heiner, C.R., Chen, E.Y., Cassell, G.H., 2000. The complete sequence of the mucosal pathogen *Ureaplasma urealyticum*. Nature 407, 757–762.

Goloubinoff, P., Mogk, A., Zvi, A.P., Tomoyasu, T., Bukau, B., 1999. Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network. Proc. Natl. Acad. Sci. USA 96, 13732–13737.

Gottesman, S., Roche, E., Zhou, Y., Sauer, R.T., 1998. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. Genes Dev. 12, 1338–1347.

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

Herman, C., Prakash, S., Lu, C.Z., Matouschek, A., Gross, C.A., 2003. Lack of a robust unfoldase activity confers a unique level of substrate specificity to the universal AAA protease FtsH. Mol. Cell 11, 659–669.

Herman, C., Thevenet, D., Bouloc, P., Walker, G.C., D'Ari, R., 1998. Degradation of carboxy-terminal-tagged cytoplasmic proteins by the *Escherichia coli* protease HflB (FtsH). Genes Dev. 12, 1348–1355.

Hutchison, C.A., Peterson, S.N., Gill, S.R., Cline, R.T., White, O., Fraser, C.M., Smith, H.O., Venter, J.C., 1999. Global transposon mutagenesis and a minimal Mycoplasma genome. Science 286, 2165–2169.

Jordan, I.K., Rogozin, I.B., Wolf, Y.I., Koonin, E.V., 2002. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. Genome Res. 12, 962–968.

Kamla, V., Henrich, B., Hadding, U., 1996. Phylogeny based on elongation factor Tu reflects the phenotypic features of mycoplasmas better than that based on 16S rRNA. Gene 171, 83–87.

Kandror, O., Sherman, M., Moerschell, R., Goldberg, A.L., 1997. Trigger factor associates with GroEL in vivo and promotes its binding to certain polypeptides. J. Biol. Chem. 272, 1730–1734.

Karzai, A.W., Sauer, R.T., 2001. Protein factors associated with the SsrA.SmpB tagging and ribosome rescue complex. Proc. Natl. Acad. Sci. USA 98, 3040–3044.

Karzai, A.W., Roche, E.D., Sauer, R.T., 2000. The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. Nat. Struct. Biol. 7, 449–455.

Keiler, K.C., Waller, P.R., Sauer, R.T., 1996. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. Science 271, 990–993.

Keskin, O., Bahar, I., Flatow, D., Covell, D.G., Jernigan, R.L., 2002. Molecular mechanisms of chaperonin GroEL-GroES function. Biochemistry 41, 491–501.

Kobayashi, K. et al., 2003. Essential *Bacillus subtilis* genes. Proc. Natl. Acad. Sci. USA 100, 4678–4683.

Kolesov, G., Mewes, H.W., Frishman, D., 2001. SNAPping up functionally related genes based on context information: a colinearity-free approach. J. Mol. Biol. 311, 639–656.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580.

Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105–132.

Lo Conte, L., Brenner, S.E., Hubbard, T.J., Chothia, C., Murzin, A.G., 2002. SCOP database in 2002: refinements accommodate structural genomics. Nucleic Acids Res. 30, 264–267.

Maier, R., Eckert, B., Scholz, C., Lilie, H., Schmid, F.X., 2003. Interaction of trigger factor with the ribosome. J. Mol. Biol. 326, 585–592.

Maniloff, J., 1992. Mycoplasma phylogeny. In: Maniloff, J., McElhaney, R.N., Finch, L.R. (Eds.), Mycoplasmas: Molecular Biology and Pathogenesis. American Society for Microbiology, Washington, DC, p. 557.

Mogk, A.B.B., Bukau, B., Deuerling, E., 2001. Cellular functions of cytosolic *E. coli* chaperones. In: Lund, P.A. (Ed.), Molecular Chaperones in the Cell. Oxford University Press, Oxford; New York, p. 3.

Moran, N.A., 2002. Microbial minimalism: genome reduction in bacterial pathogens. Cell 108, 583–586.

Narberhaus, F., 1999. Negative regulation of bacterial heat shock genes. Mol. Microbiol. 31, 1–8.

Peterson, S.N., Fraser, C.M., 2001. The complexity of simplicity. Genome Biol. 2, 2002.

Razin, S., 1995. Molecular properties of mollicutes: a synopsis. In: Rasin, S., Tully, J.G. (Eds.), Molecular and Diagnostic Procedures in Mycoplasmology. Academic Press, San Diego, pp. 1–19.

Saibil, H.R., Horwich, A.L., Fenton, W.A., 2001. Allostery and protein substrate conformational change during GroEL/GroES-mediated protein folding. Adv. Protein Chem. 59, 45–72.

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

Sasaki, Y., Ishikawa, J., Yamashita, A., Oshima, K., Kenri, T., Furuya, K., Yoshino, C., Horino, A., Shiba, T., Sasaki, T., Hattori, M., 2002. The complete genomic sequence of *Mycoplasma penetrans*, an intracellular bacterial pathogen in humans. Nucleic Acids Res. 30, 5293–5300.

Scherm, B., Gerlach, G.F., Runge, M., 2002. Analysis of heat shock protein 60 encoding genes of mycoplasmas and investigations concerning their role in immunity and infection. Vet. Microbiol. 89, 141–150.

Servant, P., Mazodier, P., 2001. Negative regulation of the heat shock response in Streptomyces. Arch. Microbiol. 176, 237–242.

Takaya, A., Tomoyasu, T., Tokumitsu, A., Morioka, M., Yamamoto, T., 2002. The ATP-dependent lon protease of *Salmonella enterica* serovar *Typhimurium* regulates invasion and expression of genes carried on *Salmonella* pathogenicity island 1. J. Bacteriol. 184, 224–232.

Tatusov, R.L., Koonin, E.V., Lipman, D.J., 1997. A genomic perspective on protein families. Science 278, 631–637.

Teter, S.A., Houry, W.A., Ang, D., Tradler, T., Rockabrand, D., Fischer, G., Blum, P., Georgopoulos, C., Hartl, F.U., 1999. Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. Cell 97, 755–765.

Ueguchi, C., Shiozawa, T., Kakeda, M., Yamada, H., Mizuno, T., 1995. A study of the double mutation of dnaJ and cbpA, whose gene products function as molecular chaperones in *Escherichia coli*. J. Bacteriol. 177, 3894–3896.

van Ham, R.C., Kamerbeek, J., Palacios, C., Rausell, C., Abascal, F., Bastolla, U., Fernandez, J.M., Jimenez, L., Postigo, M., Silva, F.J., Tamames, J., Viguera, E., Latorre, A., Valencia, A., Moran, F., Moya, A., 2003. Reductive genome evolution in *Buchnera aphidicola*. Proc. Natl. Acad. Sci. USA 100, 581–586.

von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., Snel, B., 2003. STRING: a database of predicted functional associations between proteins. Nucleic Acids Res. 31, 258–261.

Wah, D.A., Levchenko, I., Baker, T.A., Sauer, R.T., 2002. Characterization of a specificity factor for an AAA+ ATPase. Assembly of SspB dimers with ssrA-tagged proteins and the ClpX hexamer. Chem. Biol. 9, 1237–1245.

Wheeler, D.L., Church, D.M., Federhen, S., Lash, A.E., Madden, T.L., Pontius, J.U., Schuler, G.D., Schriml, L.M., Sequeira, E., Tatusova, T.A., Wagner, L., 2003. Database resources of the National Center for Biotechnology. Nucleic Acids Res. 31, 28–33.

Wickner, S., Maurizi, M.R., Gottesman, S., 1999. Posttranslational quality control: folding, refolding, and degrading proteins. Science 286, 1888–1893.

Williams, K.P., 2002. The tmRNA website: invasion by an intron. Nucleic Acids Res. 30, 179–182.

Wu, W.F., Zhou, Y., Gottesman, S., 1999. Redundant in vivo proteolytic activities of *Escherichia coli* Lon and the ClpYQ (HslUV) protease. J. Bacteriol. 181, 3681–3687.