ORIGINAL PAPER

# Conserved inserts in the Hsp60 (GroEL) and Hsp70 (DnaK) proteins are essential for cellular growth

**Bhag Singh · Radhey S. Gupta**

**Abstract** The Hsp60 and Hsp70 chaperones contain a number of conserved inserts that are restricted to particular phyla of bacteria. A one aa insert in the *E. coli* GroEL and a 21–23 insert in the DnaK proteins are specific for most Gram-negative bacteria. Two other inserts in DnaK are limited to certain groups of proteobacteria. The requirement of these inserts for cellular growth was examined by carrying out complementation studies with temperature-sensitive ($T_s$) mutants of *E. coli groEL* or *dnaK*. Our results demonstrate that deletion or most changes in these inserts completely abolished the complementation ability of the mutant proteins. Studies with GroEL and DnaK from some other species that either lacked or contained these inserts also indicated that these inserts are essential for growth of *E. coli*. The DnaK from some bacteria contains a two aa insert that is not found in *E. coli*. Introduction of this insert into the *E. coli* DnaK also led to its inactivation, indicating that these inserts are specific for different groups. We postulate that these conserved inserts that are localized in loop regions on protein surfaces, are involved in some ancillary functions that are essential for the groups of bacteria where they are found.

B. Singh · R. S. Gupta (✉)
Department of Biochemistry and Biomedical Sciences,
McMaster University, Hamilton L8N 3Z5, Canada
e-mail: gupta@mcmaster.ca

## Introduction

The Hsp60 and Hsp70 are two of the major and best-studied molecular chaperone proteins, whose homologs are present in virtually all organisms (Craig et al. 1993; Bukau and Horwich 1998; Gupta 1995; Gupta and Golding 1993). These proteins play essential cellular roles in the intracellular protein folding and translocation processes (Craig et al. 1993; Bukau and Horwich 1998; Ellis and Hartl 1999). Our earlier work has identified a number of conserved inserts in highly preserved regions of these proteins that are uniquely found in species from particular groups or phyla of bacteria (Gupta and Golding 1993; Gupta 1998, 2000). The species distribution pattern of these conserved indels (i.e., inserts or deletions), as well as numerous other conserved indels in widely distributed proteins that have been discovered in recent years, has provided important insights concerning the evolutionary relationships among different organisms (Rivera and Lake 1992; Gupta and Golding 1993; Baldauf and Palmer 1993; Rokas and Holland 2000; Gupta 1998). The discovery of these lineage-specific conserved indels in ubiquitous proteins raises the question concerning their cellular functions. Because the primary cellular functions of these proteins where such indels are present (e.g., Hsp60, Hsp70, EF-Tu, RpoB, RpoC, etc.) are expected to be the same in all species, it is important to determine whether these indels represent functionally important genetic events that are essential for the groups of organisms

where they are found. The answer to this question is of much interest and importance and it has not been carefully studied in the past.

In this work, we provide updated information regarding the species distribution patterns of several conserved indels in the Hsp60 and Hsp70 proteins and also describe some new conserved indels in these proteins. More importantly, this work examines the requirement for cellular growth of a number of conserved indels in these proteins in *E. coli* cells. The bacterial homologs of Hsp60 and Hsp70 are known as GroEL and DnaK, respectively and they are essential for the growth of *E. coli* cells. Of these, GroEL is essential at all temperatures, whereas mutation or deletion of DnaK permits limited growth below 34°C (Ang and Georgopoulos 1989; Fayet et al. 1989; Bukau and Walker 1989; Wild et al. 1992; Zeilstra-Ryalls et al. 1991). We describe here the results of experiments where the ability of different Hsp60 and Hsp70 constructs (i.e., + or – inserts or containing various modifications in the insert regions) to complement the $T_s$ phenotype of *E. coli* GroEL and DnaK mutants was examined (Ang and Georgopoulos 1989; Fayet et al. 1989). Results of our studies provide strong evidence that all of the conserved inserts in these proteins are essential for the growth of *E. coli* cells. The surface locations of these inserts in loop regions also suggest that they could be involved in interaction with other proteins and ligands.

## Experimental procedures

### Bacterial strains

The wild-type and mutant strains of *E. coli* that are temperature-sensitive ($T_s$) for growth at 42°C due to mutations in the *groEL* and *dnaK* genes have been previously characterized by Dr. Costa Georgopoulos and coworkers (Georgopoulos et al. 1973; Ang and Georgopoulos 1989; Klein and Georgopoulos 2001). The following wild-type and mutant strains were kindly made available to us by Drs. Debbie Ang and C. Georgopoulos (University of Geneva, Switzerland). The wild-type strains used were CG3014 (a derivative of B178) and CG799 (a derivative of C600) that grow normally at both 30° and 42°C. The strain CG3015 (B178 *groEL673* Tn*10* tet$^R$ nearby) is a mutant of CG3014, which contains two different point mutations in the *groEL* gene ($G_{173}D$ and $G_{337}D$) that makes it unable to grow at 42°C (Georgopoulos et al. 1973; Klein and Georgopoulos 2001). The complementation studies with *dnaK* (Hsp70) were carried out using the mutant strain CG800 (C600 *dnaK103 + thr*::Tn10), which is $T_s$ for growth at 42°C (Ang and Georgopoulos 1989).

### Construction of plasmids and creation of site-specific mutations

Full-length wild type *groEL* and *dnaK* genes were PCR amplified from *E. coli* DNA based on the published sequences for these genes/proteins (GroEL accession number AAS75782, DnaK accession number YP_851220). The amplified sequences were cloned in the pDrive vector and were sequenced to ensure that they contained no mutations. Drs. Lori Burrows and Dr. Turlough Finan of McMaster University generously provided the genomic DNA for *Pseudomonas aerguinosa* and *Sinorhizobium meliloti* to us. The cDNA for human and Chinese hamster Hsp60 have been cloned in our earlier work (Jindal et al. 1989; Picketts et al. 1989). For expression, the *groEL* and *dnaK* sequences were PCR amplified from the plasmid vector using a second set of primers containing the appropriate restriction sites and subcloned in the plasmid PKK-233 (Pharmacia). The list of different DNA primers that were used to amplify various genes or to create various site-specific changes is provided in the Supplemental Table 1. Specific changes in the *groEL* and *dnaK* genes were introduced by means of PCR using overlapping primers in opposite orientations carrying the appropriate changes employing the 'Quikchange' site-directed mutagenesis kit (Stratagene). The presence of the desired changes in the mutant plasmids (and absence of any other changes) was confirmed by DNA sequencing.

### Complementation studies with the temperature-sensitive mutants

All bacterial strains were verified for $T_s$ phenotypes by examining their growth at 30° and 42°C by streaking on the LB agar plates. The *E. coli* strains CG3014 and CG799 showed normal growth at both at both 30° and 42°C. In contrast, the mutant strains CG3015 ($T_s$ *groEL*) and CG800 ($T_s$ *dnaK*) showed no growth at 42°C, although they grew normally at 30°C. The competent cells for these *E. coli* strains were prepared by the CaCl$_2$ method. The CG3015 (or CG800) mutant cells were transformed with various PKK-233 plasmids harboring different *groEL* (or *dnaK*) gene constructs to test for complementation as measured by the ability of the cells to form colonies at 42°C. Our studies with the wild type *groEL* and *dnaK* constructs indicated that IPTG addition was not necessary in order to observe complementation of the $T_s$ phenotypes, hence IPTG was not included in such experiments. In a typical experiment, the CG3015 (or CG800) cells were transformed with about 50–200 ng of the plasmid DNA for either the wild type *groEL* (*dnaK*) or various other test construct(s). After incubating for 1 h at 30°C in a shaking water bath, different dilutions of the cultures were plated on LB agar plates

containing 50 µg/ml ampicillin. The cell dilutions were generally made to obtain between 100 and 10,000 colonies on the plates grown at 30°C with the wild type construct. All dilutions were plated in at least quadruplicate and two or more plates from each dilution were incubated at 30° and 42°C. The average number of colonies observed at 30° and 42°C upon transformation with different constructs was determined (based upon dilutions which yielded well-separated colonies) and the number of colonies observed at 42°C was normalized with respect to those seen at 30°C. All of these experiments were repeated at least three times to ensure reproducibility.

To examine the expression of the recombinant proteins in the mutant cells, CG3015 (or CG800) cells transformed with various plasmids were induced with 0.1 mmol IPTG for 2.5 h at 37°C and the cell extracts were analyzed by electrophoresis in 10% SDS-polyacrylamide gels. The cells transformed with PKK-233 plasmid containing no cloned gene was used as a control in these experiments. The band corresponding to the recombinant GroEL was readily detected in stained gels due to its high level of expression in comparison to the control cells. However, to detect the band corresponding to DnaK, it was necessary to carry out western blot analysis using an antibody to the Hsp70/DnaK protein. The rabbit polyclonal antibody to Hsp70/DnaK used in these studies (Ahmad et al. 1990) shows good reactivity against the *E. coli* DnaK protein and it was employed at a dilution of 1:2,000 in these experiments.

Blast searches

The presence or absence of various inserts in the Hsp60 and Hsp70 proteins in different groups was determined by Blast searches on the indicated sequence segments of these proteins against various groups of bacteria. Such searches were performed with both an insert containing (*E. coli*) as well as an insert lacking homolog (*Bacillus subtilis*) and the presence or absence of various inserts in species from different groups of bacteria was determined.

## Results

Species distribution and essential nature of a conserved insert in the Hsp60 protein

Figure 1 presents partial sequence alignment of the Hsp60 protein showing a one aa insert (boxed) in a conserved region that is specific for most Gram-negative bacteria, but not found in other phyla of bacteria including various Gram-positive bacteria (Gupta 1998, 2000). When this signature was originally described, sequence information for Hsp60 was available from only a limited number of bacteria. To examine the specificity of this insert, Blastp searches were carried out on the Hsp60 sequence segments shown in Fig. 1 against various taxonomic groups and the presence or absence of this insert in species from different bacterial groups was determined. The results of these analyses are also included in Fig. 1. For example, for the γ- and β-proteobacteria, 611 and 139 entries, respectively, were found in the NCBI database and all of them contained this insert. Similarly, for other phyla of Gram-negative bacteria (viz. α-, δ- and ε- proteobacteria, Aquificae, Chlamydiae–Verrucomicrobiae–Planctomyctes, Bacteroidetes–Chlorobi–Fibrobacter, Spirochaetes and Cyanobacteria), about 900 sequences were available and all of them contained this insert. In some bacteria, particularly the *Rhizobiaceae* and *Bradyrhizobiaceae*, multiple homologs of Hsp60 are found (Rodriguez-Quinones et al. 2005; Rusanganwa and Gupta 1993) and all of them contained this insert. In contrast to these bacterial groups, this insert was not present in nearly all of the sequences from Chloroflexi, Deinococcus–Thermus, Thermotogae, Fusobacteria, Actinobacteria and Firmicutes phyla. Of the >1,400 sequences from these groups, this insert was present in only ten entries (mostly from Firmicutes), which constituted <1% of the total sequences. These small number of exceptions are likely the results of lateral gene transfers or some other non-specific events (Gogarten et al. 2002; Doolittle 1999).

In eukaryotic organisms, Hsp60 homologs are only found in organelles such as mitochondria and plastids (Gupta 1995, 2000), which have originated from bacterial ancestors belonging to the α-proteobacteria and Cyanobacteria groups, respectively (Margulis 1993; Gray 1999; Palmer and Delwiche 1998). Similar to various bacteria from these groups, all of the mitochondrial and plastids homologs of Hsp60 also contained this insert in the same position. Further, similar to the proteobacteria, the insert in various mitochondrial homologs consisted of an N (Asn) residue, whereas in all of the plastid homologs a G (Gly) was present in this position. These results provide further evidence for the origin of plastids from a cyanobacterial ancestor (Palmer and Delwiche 1998; Gupta et al. 2003).

The species distribution pattern of the Hsp60 insert indicates that it is a stable and highly reliable characteristic of the above groups of Gram-negative bacteria. Most likely explanation for this result is that the rare genetic change (RGC) responsible for this insert was introduced in a common ancestor of the above groups of bacteria, as indicated in Fig. 2. The retention of this insert by all species from the indicated groups of bacteria without any loss (e.g., all 1,385 proteobacterial homologs contain this insert) provides persuasive evidence that there is a strong selection pressure for the retention of this insert in these bacteria. In the Hsp60 structure (Braig et al. 1994), the $N_{153}$ residue, corresponding

**Fig. 1** Partial sequence alignment of the Hsp60 protein showing a one aa insert (*boxed*) in a conserved region that is specific for certain major phyla of bacteria and the eukaryotic homologs, but not found in other bacteria. The presence or absence of this insert in all available sequences from different bacterial groups is indicated along with their names. For example, for $\gamma$-proteobacteria a total of 611 hits corresponding to Hsp60 were observed and all of them contained this insert (611 with insert, 0 without insert). Similarly, for the Firmicutes phylum, a total of 778 hits were observed and of these only seven contained the insert (7/771). Only representative sequences from different major groups are shown here. The *dashes* in the alignment indicate that the same amino acid as that found on the top line (i.e., *E. coli* protein) is present in that position. The accession numbers of sequences are given in the second column. The *numbers* on the *top* indicate the position of this sequence in *E. coli* protein

```
                                                                144                                178
                    Escherichia coli              P06139         IAQVGTISA N SDETVGKLIAEAMDKVGKEGVITVE
                    Pasteurella multocida         Q59687         -E------- - --SI--QI--Q--------------
Gamma-Proteo (611/0) Pseudomonas aeruginosa        P30718         --------- - ---SI-QI-----E-----------
                    Legionella pneumophila        YP_094724      --------- - ---AI-AI-----E-----------
                    Xylella fastidiosa            AE003849       ----A---- - ---SI-NI-----K---------I-
Beta-Proteo (139/0)  Neisseria meningitidis        P42385         -----S--- - ---Q--AI-----E-----------
                    Ralstonia solanacearum        17545361       -----A--- - ---SI-AR-----------------
                    Rickettsia prowazekii         3861167        --------S G-KEI-EK--K--EE----------
Alpha-Proteo (414/0) Caulobacter crescentus        P48211         --------- - G-KE--EM--K------N-------
                    Brucella abortus              P25967         V-------- - AEAEI--M-----Q---N------
Delta-Epsilon       Geobacter metallireducens     ZP_00082424    --------- - N-K-I-DI--Q--E----------
Proteo (215/0)      Desulfovibrio vulgaris        ZP_00129431    --------- - --S-I-NI-----S-----------
                    Campylobacter jejuni          e1287421       ----A---- - ---KI-N---D--E----D------
                    Helicobacter pylori           P42383         -T--A---- - --HNI-----D----E---D------
Aquificae (6/0)      Aquifex aeolicus              2984379        -E--A---- - N-PEI--I--D--EE---D------
                    Hydrogenivirga sp.128-5       ZP_02176593    -E--AA--- - N-PEI--I-D--EA---D------
Chlamydiae-         Chlamydia muridarum           AAF39243       ----A---- - N-ABI-N-------E----N-S---
Verrucomicrobiae-   Protochlamydia amoebophila    YP_0081791     ----A---- - N-AEI-EM--Q-IE---RD-T----
Planctomycetes (36/0) Lentisphaera araneosa        ZP_01875439    VK-IA---- - G--EI-SI--D-------D-T-A--
                    Verruco. spinosum             ZP_02926970    ----A-V-- - W-TBI-NI--D-------D-T----
                    Planctomyces maris            ZP_01855423    -----A--- - N-SVI-N---D-VE---RD------
                    Blastopirellula marina        ZP_01094293    V-H--A--- - N-HAI-E-L---LYR---D------
Bacteroides-        Chlorobium tepidum            AAM71772       --------- - N-PEI-E-----------D------
Chlorobi-           Fibrobacter succinogenes      AY017380       ----A---- - N-PEI-E-L-N--E---ND----I-
Fibrobacter (107/0)  Cytophaga aquatilis           AF335327       -K-IAS--- - N--VI-E---T-FA-----------
                    Bacteroides forsythus         CAB43992       -EH-AK--- - G--GI-------Q--K---------
Spirochaetes(16/0)   Borrelia burgdorferi          P27575         ----AS--- - N-SYI-EK---------D----
                    Treponema pallidum            AAC65026       V-H-ASV-- - N-KEI-RIL-S-IE---ND---D-D
                    Leptospira interrogans        L14682         --N-AS--- - N-N-I-N---D-------D------
Cyanobacteria (105/0) Synechocystis sp. PCC6803    2506274        -----A--- G N-E--QM--N-----Q----SL-
                    Nostoc punctiforme            17132795       -----A--- G N-E--QM--Q----------SL-
                    Thermosyn. elongatus          11131866       ----AA--- G N--E--RM--D---------SL-
Chloroflexi (0/10)   Chloroflexus aurantiacus      AAD33468       --H-A-N-- - A-SEI-E---V-A----D------
                    Roseiflexus castenholzii      YP_001431326   --H-AA--- - A-SEI-D---V-E----D------
                    Herpetosiphon aurantiacus     YP_001546443   -SS-A---- - A-SSI-D---V------D------
                    Deinococcus radiodurans       G75499         -KK-AG--- - N-----QE--S-----------I-
Deinococcus-        Deinococcus grandis           AAS10491       -KK-AG--- - N--Q--QE--S------------I-
Thermus (1/9)       Thermus aquaticus             P45746         -EE-A---- - N-PE------D--E------I----
                    Thermus igniterrae            AAX24029       -EE-A---- - N-PD------D--E------I----
Thermotogae (0/7)    Thermotoga maritima           H72367         --HA-A--- - NSABI-E---------ED------
                    Fervidobacterium nodosum      YP_001411043   --H-AS--- - NS-EI--------E---ED------
                    Petrotoga mobilis             YP_001568199   --H-AS--- - NN-EI-NI---------ED------
Fusobacteria (0/4)   Fusobacterium nucleatum       ZP_00273152    -S--AS--- - G--EI-----Q--A---ET------
                    Fusobacterium perfoetens      AAT96607       -E--AS--- - G-KEI--------A---ET------
                    Mycobacterium leprae          P09239         --ATAA--- - G-QSI-D---------N--------
                    Corynebacterium glutamicum    19553910       --N-A-V-S R--V--EIV-A--E----D--V---
Actinobacteria      Bifidobacterium longum        24850323       --A-A---- - A-PE--EK----L----QD-V---
(2/639)             Streptomyces coelicolor       8248795        --STAS--- - A-TQI-E------------------
                    Tropheryma whipplei           CAD830831      --ACAS--- - G-PQI-DI--Q-LE-------V---
                    Arthrobacter ureafaciens      AAY233611      --ATAS--- - G-QBI-A-----L------------
                    Frankia sp. EAN1pec           YP_001507661   --SAAS--- - G-PAI-AM-----------------
                    Clostridium perfringens       P26821         --R-AA--- - A--KI-----D--E---N-------
                    Lactococcus lactis            X71132         ----A-V-S RS-K--EY-SD--ER--SD----I-
Firmicutes          Staphylococcus aureus         15927604       -----A--- - A--EI-RY-S---E---ND----I-
(7/771)             Bacillus subtilis             16077670       ----AA--- - A--E--S-------ER--ND----I-
                    Streptococcus pneumoniae      2545870        ----AAV-S RS-K--EY-S---E---D----I-
                    Mycoplasma genitalium         12045254       -E--AA--S GSKEI-----Q--AL---N----TD
                    Lactobacillus plantarum       NP_7844831     --IASV-S AS-ET--------E---HD----I-
                    Saccharomyces cerevisiae      NP_0133601     ----A---- - G-SH----L-S--E--------IR
                    Arabidopsis thaliana          CAA776461      --------- - GEREI-E---K--E---------IQ
Eukaryotes          Oryza sativa                  EAZ163411      --------- - GEREI-E---K--E---------I
(mitochondria)      Chlamydomonas reinhardtii     XP_001691353   --------- - GEREI-E---R--E-----------
(250/0)             Drosophila melanogaster       CAA677201      ----A---- - G-QAI-N--S---K---RD-----K
                    Trypanosoma cruzi             Q95046         --V--A---- - G--EL-R--GQ--E----D-----TQ
                    Homo sapiens                  AAX931451      ----A---- - G-KEI-NI-SD--K---RK-----K
Eukaryotes          Cyanidium caldarium           NP_0451261     -I--SS--- G N--EI-SM--N-IQ----D---SL-
(Plastids)(86/0)    Oryza sativa                  EAZ201561      -KA-AS--- G N--LI-SM--D-I----PD--LSI-
                    Arabidopsis thaliana          NP_1879561     L-D-AAV-- G NNHE--SM-----S---RK--V-L-
                    Chlamydomonas reinhardtii     XP_001701113   L-N-ACV-- G GNTDI-S--SD--A---RT--V-M-
```

to the insert, is present on the surface of this protein in a loop connecting two alpha helices (Supplemental Figure 1). In this figure, only three of the seven subunits of the GroEL ring structure are shown. The insert is shown in purple with its side chain protruding. Earlier studies indicate that mutations in a number of neighboring conserved residues viz. $I_{150} \to E$, $S_{151} \to V$ and $A_{152} \to E$, affect both ATPase activity and binding of the co-chaperone GroES by the mutant proteins (Fenton et al. 1994).

The importance of the Hsp60 insert for cellular growth was examined by employing a $T_s$ mutant of *E. coli* (CG3015, *groEL673*) (Georgopoulos et al. 1973; Klein and Georgopoulos 2001). Transformation of this mutant with an expression plasmid containing the wild-type *E. coli groEL* gene led to restoration of normal growth at 42°C (Table 1).

When the one aa insert ($N_{153}$) from the *E. coli* GroEL was deleted, cells transformed with the (–) insert plasmid showed no growth at 42°C (Table 1) indicating that the protein lacking this insert was unable to meet the GroEL requirement for cellular growth. To understand the significance of the amino acid found in this position, a number of mutants were made in which the $N_{153}$ was replaced with other amino acids [viz. Glycine (G), Alanine (A), Aspartic acid (D), Glutamine (Q) and Valine (V)]. The results of studies with these mutants are also presented in Table 1. Of these mutants, very weak complementation (0.3–4.0% plating efficiency) was observed with the $N_{153} \to D$, $N_{153} \to Q$ and $N_{153} \to A$ mutants, while the other two mutations, $N_{153} \to G$ and $N_{153} \to V$, were totally inactive in complementing the $T_s$ defect. In the first three of these mutants
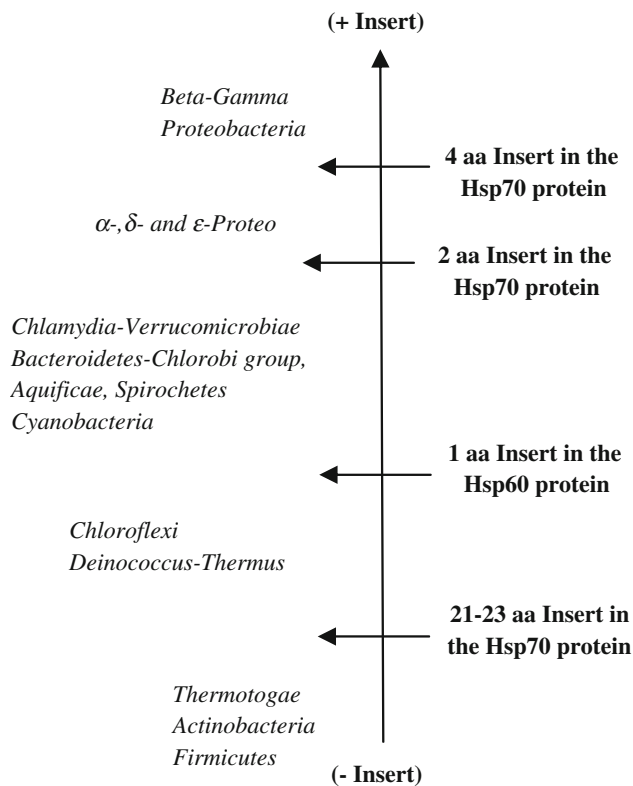
**Fig. 2** A diagram showing the species distribution patterns of various conserved inserts in the GroEL and DnaK proteins (shown in Figs. 1, 4 and 6) and the evolutionary stages where genetic changes responsible for these inserts likely occurred (Gupta 2000; Griffiths and Gupta 2004). The different inserts shown in this diagram are generally present in all of the species from various bacterial groups that lie above the *horizontal lines* marking the positions of these inserts, but they are absent from most other groups that fall below these lines

**Table 1** Complementation of the *E. coli* GroEL $T_s$ mutant with different constructs

| Plasmid construct | Growth at 30°C | Growth at 42°C | Comments |
|---|---|---|---|
| Control plasmid (no insert) | 100 | 0 | Negative control |
| Wild-type *E. coli* GroEL | 100 | 100 | Contains insert |
| *E. coli* GroEL (ΔN153) | 100 | 0 | Insert deleted |
| *E. coli* GroEL $N_{153}D$ | 100 | <3–4% | Modified insert |
| *E. coli* GroEL $N_{153}G$ | 100 | 0 | Modified insert |
| *E. coli* GroEL $N_{153}V$ | 100 | 0 | Modified insert |
| *E. coli* GroEL $N_{153}Q$ | 100 | ~1% | Modified insert |
| *E. coli* GroEL $N_{153}A$ | 100 | ~0.3–0.4% | Modified insert |
| GroEL, *Bacillus subtilis* | 100 | 0 | Lacks insert |
| GroEL, *P. aeruginosa* | 100 | 100 | Contains insert |
| GroEL, *S. meliloti* | 100 | 100 | Contains insert |
| Hsp60, Chinese hamster | 100 | 0 | Contains insert |
| Hsp60, Human | 100 | 0 | Contains insert |

Complementation of the $T_s$ phenotype of *E. coli* CG3015 (B178 *groEL673*) cells upon transformation with different *groEL* gene constructs expressing either the wild type or mutant GroEL protein from *E. coli* or other prokaryotic/eukaryotic organisms. The experiment was carried out as described in "Experimental procedures". Assuming the number of colonies observed at 30°C to be 100%, the relative plating efficiency of the cells at 42°C was determined

showing weak complementation, the asparagine residue in the insert position was replaced with either a conservative (D or Q) or similar size amino acids, whereas in the other two cases the replacements were more drastic. These results indicated that not only that the insert in this position, but the nature of the amino acid present at this position is also crucial for the proper function of the protein. As noted earlier, in contrast to other bacteria, all cyanobacteria have a glycine (G) residue in the insert position. However, a N → G substitution is not compatible with the growth of *E. coli*, indicating that this change to Gly is specific for cyanobacteria and plastids.

We have also examined the expression of the mutant proteins in the CG3015 (or CG800) to determine whether the lack of complementation in some cases could be caused by this factor. However, there was good and comparable level of expression of the mutant proteins (GroEL or DnaK) in all cases (Fig. 3), indicating that the observed lack of complementation for many mutants, or proteins from other bacteria, that were studied in this work was not due to their lack of expression.

We have also examined whether Hsp60 from other organisms can complement the $T_s$ defect of the *E. coli* *groEL* mutant. To test this, full-length *groEL* genes from a number of bacteria viz. *P. aeruginosa* (a γ-proteobacteria), *S. meliloti* (an α-proteobacteria) and *B. subtilis* (a low G + C Gram-positive bacterium) were cloned in the expression vector and their ability to complement the $T_s$ phenotype of *E. coli* CG3015 was examined. All of these proteins showed comparable level of expression in *E. coli* cells (Fig. 3a). However, of these, only the GroEL homologs from *P. aeruginosa* and *S. meliloti*, which contained the Hsp60 insert and exhibited 79 and 68% sequence identity, respectively, to the *E. coli* protein, were able to complement the $T_s$ defect of the mutant (about 50% colony forming efficiency at 42°C; Table 1). In contrast, the GroEL homolog from *B. subtilis*, which lacked the Hsp60 insert was unable to complement the $T_s$ defect. This result is consistent with the requirement of the Hsp60 insert for its proper functioning in *E. coli*. However, the inability of this latter protein to function in *E. coli* can also be due to other changes in the molecule as the *B. subtilis* Hsp60 exhibits only 61% sequence identity to the *E. coli* protein.

We also studied whether the human and Chinese hamster Hsp60 (i.e., mitochondrial homologs) can complement the $T_s$ defect of the mutant. These mitochondrial proteins are derived from an α-proteobacterial ancestor and they contain the conserved insert in the Hsp60 protein. However, these proteins exhibit only 50% sequence identity to the *E. coli*
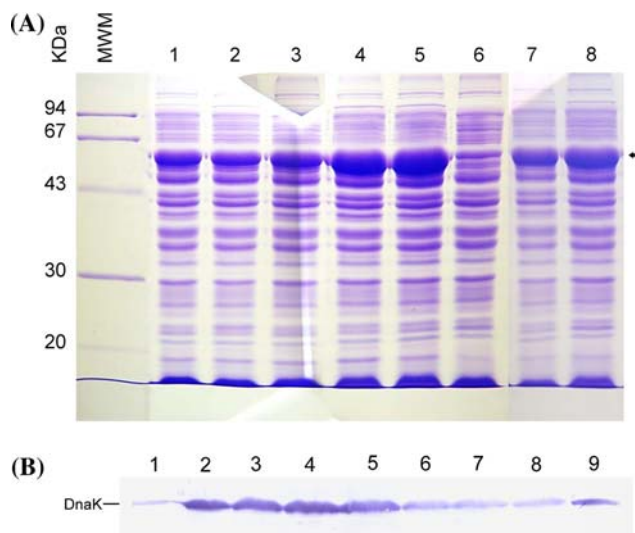
**Fig. 3** Expression of the recombinant GroEL/Hsp60 or DnaK proteins in the mutant cells. **a** SDS-PAGE analysis of expression of the recombinant GroEL proteins in *E. coli* CG3015 cells grown at 37°C. The cells transformed with various plasmids (see Table 1) were induced with 0.1 mmol IPTG for 2.5 h and then analyzed. *Left lane MWM* molecular weight markers, *lane 1* cells transformed with a plasmid containing the wild-type *E. coli* GroEL, *lane 2 E. coli* GroEL with deletion of $N_{153}$, *lane 3 E. coli* GroEL with $N_{153}D$ mutation, *lane 4 E. coli* GroEL with $N_{153}V$ mutation, *lane 5 E. coli* GroEL with $N_{153}Q$ mutation, *lane 6* control *E. coli* cells transformed with empty plasmid, *lane 7* plasmid with *B. subtilis* GroEL, *lane 8* plasmid containing full-length Hsp60 cDNA from human cells. Cells transformed with other plasmids not shown here also showed comparable level of protein expression. **b** Western blot analysis showing the expression at 37°C of DnaK in *E. coli* CG800 cells transformed with various plasmids (Table 2). *Lane 1* cells transformed with empty vector, *lane 2* plasmid with wild-type DNA from *E. coli*, *lane 3 E. coli* DnaK with Δ75–97, *lane 4 E. coli* DnaK with ΔEE deletion from positions 80–81, *lane 5 E. coli* DnaK with the four aa deletion of DEVD from positions 208–211, *lane 6 E. coli* DnaK with $\Delta E_{209}$, *lane 7 E. coli* DnaK with two aa addition (SG) after position 385, *lane 8* DnaK from *P. aeruginosa*, *lane 9* DnaK from *S. meliloti*

homolog (Jindal et al. 1989; Picketts et al. 1989), indicating that they have diverged significantly from their bacterial ancestor. The results of our studies revealed that these proteins, despite containing the conserved insert and showing good expression (Fig. 3a), were unable to complement the $T_s$ defect of the mutant (Table 1). Richardson et al. (2001) have previously shown that the mammalian Hsp60 functions specifically with its own co-chaperonin and it does not function with the bacterial GroES protein. The inability of the mammalian Hsp60 to complement the $T_s$ defect of *E. coli* mutant is thus very likely due to this reason.

*Description and studies on the conserved inserts in the Hsp70 (DnaK) protein*

The Hsp70 family of proteins comprise one of the most conserved proteins known that are ubiquitously found in

species from all three domains, with the exception of a few archaebacteria (Gupta and Golding 1993; Gupta 1998; Macario et al. 1999). A number of conserved indels have been described in this protein that are of considerable evolutionary significance. Figure 4 shows partial sequence alignment of the Hsp70 protein from some representative bacteria highlighting a 21–23 aa insert (boxed) that is a distinctive characteristic of various Gram-negative bacteria that are bounded with two membranes (i.e., diderm prokaryotes) (Gupta and Golding 1993; Gupta 1998; Lake et al. 2007). When blast searches were carried out with the segment of the DnaK protein shown in Fig. 4, all of the blast hits (>800) from various phyla of diderm bacteria contained this insert. In some proteobacteria, as a result of gene duplications, up to three homologs of Hsp70 are found (i.e., DnaK, HscA and HscC) (Vickery and Cupp-Vickery 2007; Arifuzzaman et al. 2004) and this insert was present in all of them (results not shown). In contrast, this insert was not found in any of the blast hits from the Firmicutes and Thermotogae phyla and most of the Actinobacteria. However, in a small number of Actinobacteria (e.g., *Frankia*, *Arthrobacter*, *Streptomyces avermitilis*, *Rhodococcus*, *Actinomadura*) two different Hsp70 homologs are present and of these only one contained this insert. Because most other Actinobacteria have only a single copy of the Hsp70 gene, it is likely that these soil Actinobacteria have acquired the additional Hsp70 gene from some soil-resident Gram-negative bacteria by lateral gene transfer (Gogarten et al. 2002; Doolittle 1999). Evidence has been presented that this large indel, which is also lacking in various archaebacteria, is an insert in the Gram-negative bacteria (see Fig. 2), thereby indicating that the prokaryotic phyla lacking this indel are ancestral (Gupta and Golding 1993; Gupta 1998; Gupta and Singh 1994; Lake et al. 2007). The presence of this large insert in all of the eukaryotic Hsp70 homologs also provides strong evidence for their origin from an insert-containing Gram-negative bacterium (Gupta et al. 1994; Gupta 1998). Within this large insert, a two aa insert that is specific for various proteobacteria has also been identified (smaller box in Fig. 4) (Gupta 1998).

The structure of the Hsp70/DnaK protein has been solved from a number of sources including bovine (Flaherty et al. 1990), *E. coli* (Harrison et al. 1997; Pellecchia et al. 2000) as well as a low G + C Gram-positive bacterium *Geobacillus kaustophilus* (Chang, Y.-W., Protein Data Bank I.D. No. 2v7y). This insert is located in the N-terminal ATPase domain of the protein (aa 1–383) and the structures of this domain for the *E. coli* and *G. kaustophilus* proteins are shown in Fig. 5. This large insert is present on the surface of the *E. coli* protein as an additional alpha helix inserted within a loop. The importance of this indel was examined by complementation studies using the DnaK $T_s$ mutant CG800, which shows no growth at 42°C (Ang and Georgopoulos

```
                                                      61                                                                       123
       ┌ Escherichia coli       SP/P04475    NPQNTLFAIKRLIG RRF QD EEVQRDVSIMPFKIIAAD NGDAWVEVKGQK MAPPQISAEVLKKM
       │  Xylella fastidiosa     AAF85139     --K--FY-V----- -K- G- A---K-LDLV-Y--TQH- ------ATADAK L--QE---K--E--
       │  Pse. aeruginosa        AAG08147     ------Y-V----- --- EE NV--K-IQMV-YS-VK-- ------------ --------------
       │  Vibrio cholerae        2623605      -------------- --- E- ------IK---Y--VK-- -------A---- --A--V-----L--
       │  Ral. solanacearum      16429657     --R---Y-V----- -K- EE K---K-IGL--YT-SK-- --------RDK- ------------R--
       │  Nei. meningitidis      H81185       -AK--IY-A----- HK- E- K-----IES---E--K-N ------KAQ-KE LS---------R--
Proteo-│  Mesorhizobium loti     13473981     --E--I--V----- --Y D- PVTEK-KKLV-Y--VKG- --------AG-K- QS-S----MI-Q--
Bacteria│  Rickettsia conorii    15892156     --R--IY-V----- -N- I- PM-RK-QG-V-YN-VK-- -------ADNN- YS-S----FI-Q--
(596/0)│  Caulo. crescentus      GB/M95799    --T----------- -TA S- PV-EK-KGMV-YRSSR-R A-----KAH-KD YS-QEV--FI-Q--
       │  Myxococcus xanthus     1805284      --E--V--A----- -K- DS P-GKKAIGVS---VASSP -------IR-KG YS--EV--I--M--
       │  Geo. metallireducens   ZP_00080430  --E----------- -KY DT ---RK-I--S----VK-- -------AR-KM YSA-E---M--Q--
       │  Des. vulgaris          ZP_00130430  --ER-V--V---M- --G DA P--G-WKEHS-YR-V-GA ----A---Q-RP YSA-E---MI-G-L
       │  Campylobacter jejuni   CAB73024     --EK-IYS---IM- LMI NE DAAKEAKNRL-YH -TER --ACAI-IA-KI YT-QE---K--M-L
       └ Helicobacter pylori     2072520      --EK-IYS---IM- LM- NE DKAKEAEKRL-Y--VDRN GA-CAI-IS-KI YT-QE---KI-M-L
       ┌ Chl. trachomatis        SP/P17821    --EK--AST--F-- -K-    S--ESEIKTV-Y-VAPNS K----VFD-EQKL YT-EE-G-QI-M--
       │  Chlam. pneumoniae      AAF38114     --EK--GS---F-- -KY    S--ASEIQTV-YTVTSGS K---VF--D-KQ  YT-EE-G-QI-N--
       │  Prot. amoebophila      YP_008498    --E--ITSS--F-- -KY    Q--LSEIKTV-Y-VTNNN ----VF-IQ-KI  VT-EE-A-QI-I--
       │  Lentisphaera araneosa  ZP_01875441  --T--I-SA--FM- -K-    A--KHEIDLV-YEVVE-K ----HIK-GDKV  YS--E---KIIA-L
       │  Cyt. aquatilis         AF013110     --TK-IAS---FM- HT-    A-TTNESKRVSY-VVKGV QQYST-DID-RL  YTAQEL--MT-Q--
       │  Fib. succinogenes      AY017382     --EK-IYS---FM- -TA    G-CSAAEKN--Y-LVGTG SDPVR-QIDDKQ  F---E---A--QT-
       │  Por. gingivalis        Q9ZAD3       --TK-IYS---FM- ETY    DQ-S-E-ERV---VVRG- -NTPR-DID-RL  YT-QE---MI-Q-L
Other Gram-│ Chlrobium tepidum   AF130447     --K--I-S---FM- -KY    D--PNEKKLASYDVVN-E G-Y-K-KIGDKT  YS-QE---MI-Q--
Negative│  Hydro. marinus        AY188440     D----IYES--F-- -K-    D--KEEIKYV-Y-VV-D- K---AFD-PNAGKIVR-EEVG-Q----L
Bacteria│  Aquifex aeolicus      2983493      D-E--VYES--F-- -K-    N---KEAKRVSY-VVPDE K---AFDIP-KL  VR-EEVG-H--R-L
(230/0)│  Treponema pallidum     3322484      --EH-IYS---F-- S--    N-LTGEAKKV-Y--VPQG -D-VR---E-KL  YSTQE---FI-Q--
       │  Bor. burgdorferi       SP/P28608    ---            ---ASEIKMV-Y--EKGL ----R-NISNI-KQ-S--E---AT-T--
       │  Lep. interrogans       AAN50903     -AV--IRSA--F-- --L    N-CESEMKHVSY-V-RSG -EGVKF-TSAGE  FT-QE---R--M--
       │  Synechococcus PCC7942  GB/D29668    --R--FAN---F-- --Y    D-LTDESKRV-YTVRRDP E-NVRIVCPQLSREF--EEVA-MI-R-L
       │  Synechocystis PCC6803  GB/M57518    --G--FYSV--F-- -K-    D-ITNEATEVAYSVVKDG --NVKLDCPA-GKQF--EE---Q--R-L
       │  Anabaena variabilis    2073390      -----F--V--Y-- --Y    N-LSPESKRV-YT-RKD- V-NIK-ARLNKE  F-AEE---M----L
       │  Cfx. aurantiacus       AF130446     --E---VSV--F-- -S-    D--TEERDV----VVKGP RN-VRIY-PQTNKEY--QE---M--Q-L
       │  Herpeto. aurantiacus   YP_001543093 --E--I-S---FM- LE-    DKATQER-TV-YE-VNGP R---R-R-PRTGKDYT-QE---MT-Q-L
       │  D. radiodurans         A75557       --AA---EV--F-- --W    D--KEEAARS--TVKEGP S-SVRI--N-KD  L--E-V-----R-L
       └ Thermus aquaticus       GB/Y07826    --EG-I-E---F-- ---    ----EEAKRV-Y-VVPGP D-GVR-----KL  YT-EE---MI-R-L
Thermotogae ┌ Thermotoga maritima 3882431                                                T-YK-RIDDKE   YT-QE---FI---L
(0/8)      └ Fervido. nodosum    YP_001410713 -SDR-IKS---KM-                             S-YK-RIDDKE   YT-QE---YI---L
Actinobacteria ┌ Mycobacterium leprae SP/P19993  -VDR-IRSV--HM-                          S-WSI-ID-K-   YTAQE--R--M-L
(12/71)    │  Str. coelicolor    GB/O14499     -VDR-IRSV--HM-                             T-WK-NLD-KD   FN-Q----F--Q-L
           │  Bifido. longum     AAN24348      -VDR-ISSV--HM-                             S-WT-DID-K-   WT-QE--AQILMKL
           └ Cor. glutamicum     Q8NLY6        -VDR-IRSV--H--                             T-WS-AIDDKN   YTSQE---RT-M-L
Firmicutes ┌ Clo. acetobutylicum 118726        --DK-IIS---KM-                             TAEK-AIDDKN   YT-QE---MI-Q-L
(0/262)    │  Bacillus subtilis  GB/U39711     -- --IMS----M-                             T-YK--IE-KD   YT-QEV--II-QHL
           │  Mycoplasma genitalium 1361531    -- --IVS----M-                             TSNK-K-TTKE   LS-E-V--QI-SYL
           │  Strep. mutans      2145133       -- E-ILS--SKM-                             TSEK-SANAKE   YT-QE---MI-QYL
           │  Lactococcus lactis GB/M98865     -- E-IIS---KM-                             TSEK-SAN-KE   YT-QE---MI-QNL
           └ Sta. aureus         GB/X75428     -- --VQS---HM-                             T-YK-DIE-KS   YT-QE---MI-QNL
```

**Fig. 4** Partial sequence alignment of the Hsp70/DnaK protein show-ing a large insert (*boxed*) that is specific for various Gram-negative bacteria (i.e., diderm bacteria bounded by two membranes) and the eukaryotic homologs, but not found in other bacteria. The numbers with the group's name indicate the presence or absence of this insert in various blast hits from each group (see Fig. 1 and text for details). The distribution (presence/absence) of this insert in various subgroups of Gram-negative bacteria was as follows: γ proteobacteria 213/0; β pro-teobacteria, 120/0; α proteobacteria, 147/0; δ–ε proteobacteria, 116/0; Chlamydiae–Verrucomicrobiae, 10/0; Bacteriodetes–Chlorobi, 54/0; Cyanobacteria, 121/0; Aquificae, Spirochetes, Chloroflexi, Deinococ-cus–Thermus, 45/0; All of the sequences shown here are for the DnaK homologs, however, this insert is also present in other Hsp70 homologs (viz. HscA or HscC) that are present in some bacteria (mainly proteo-bacteria). This insert is also not found in archaeal homologs (0/40), whose sequences are not shown (Gupta and Golding 1993; Gupta 1998; Lake et al. 2007). The *smaller box* inside the *larger box* indicates a two aa insert that is specific for proteobacteria and not found in other Gram-negative bacteria (Gupta 1998, 2000)

1989). The transformation of this mutant with the WT copy of the *dnaK* gene led to normal growth at 42°C (Table 2). However, when the large insert that is characteristics of the Gram-negative bacteria was deleted from the DnaK protein (Δ75–97) the resulting protein despite showing good expres-sion (Fig. 3b) showed no complementation of the *dnaK* $T_s$ mutant (Table 2). Previously, another deletion from this region (Δ74–96) of the DnaK protein was also found to lead to loss of its biological activity (Sugimoto et al. 2007). We also made a two aa deletion ($\Delta EE_{80-81}$) within this large insert and this plasmid was also completely inactive in com-plementing the $T_s$ phenotype of the mutant (Table 2).

The Hsp70 protein contains another highly conserved insert that is specific for the β–γ proteobacteria (Fig. 6). This four aa insert is present in all of the DnaK sequences (>300) from β–γ proteobacteria, but it is not found in any other Gram-negative bacteria. This insert is also not present in the HscA or HscC homologs from β–γ proteobacteria

(results not shown), indicating that it is a distinctive charac-teristic of the DnaK homologs from these bacteria (see Fig. 2). Within Gram-positive bacteria this insert is absent in virtually all Firmcutes (only one exception in 209 entries) and all Actinobacteria except those belonging to the orders *Micrococcineae* and *Bifidobacteria*, which form a distinct clade within Actinobacteria (Gao and Gupta 2005; Embley and Stackebrandt 1993; Ventura et al. 2007). All of the DnaK homologs from these two orders of Actino-bacteria contain either a four or a five aa insert (mostly five aa) in the same position (Fig. 6), where the four aa insert is found in β–γ proteobacteria. Because the sequences of the inserts in Actinobacteria and β–γ proteobacteria differ from each other and these two groups of bacteria branch very distantly in Hsp70 phylogenetic trees (Gupta 1998) (unpub-lished results), the inserts in these two cases have origi-nated independently. In the DnaK structure, these inserts are located in a surface loop (Fig. 5), but the structure of the
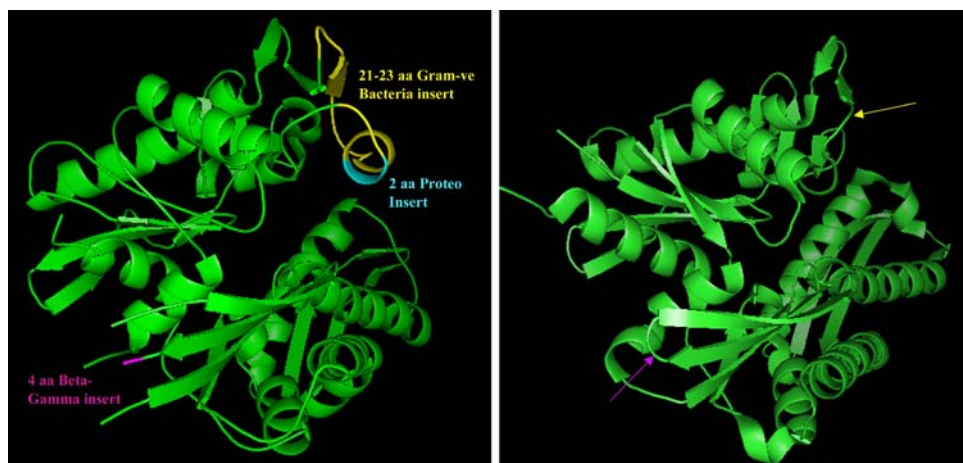
**Fig. 5** Structure of the ATPase domain (aa 1–383) of DnaK from *E. coli* (*left panel*) and *Geobacillus kaustophilus* (*right panel*) showing the positions of different conserved inserts (described in Figs. 4, 6) in the *E. coli* protein. The images were constructed by means of PyMol (version 0.93) using the PDB file 1DKG for the *E. coli* protein and 2V7Y for *G. kaustophilus* homolog. The large insert specific for the Gram-negative bacteria is shown in yellow and it is present as an extra alpha helix. The two aa insert within the large insert that is specific for the proteobacteria is shown in *blue*. The position of the four aa *β–γ* proteobacteria insert in the *E. coli* protein is shown in *pink*. However, the structure of this region, which forms part of a loop is not resolved in the *E. coli* protein structure (Harrison et al. 1997). The *colored arrows* mark the positions of these inserts in the *G. kaustophilus* protein, which lacks these inserts. Note: the colors can be seen only on the online version of the article

**Table 2** Complementation of the *E. coli* DnaK $T_s$ mutant with different constructs

| Plasmid construct | Growth at 30°C | Growth at 42°C | Comments |
|---|---|---|---|
| Control plasmid (no insert) | 100 | 0 | Control |
| Wild-type *E. coli* DnaK | 100 | 100 | Contains inserts |
| *E. coli* DnaK (Δ amino acids 75–97) | 100 | 0 | Deletion of diderm insert |
| *E. coli* DnaK (ΔEE from positions 80–81 within the large insert) | 100 | 0 | Part of the diderm insert deleted |
| *E. coli* DnaK (Δ four amino acids DEVD, from positions 208–211) | 100 | 0 | Deletion of βγ-proteo insert |
| *E. coli* DnaK (ΔE from position 209) | 100 | ~0.3% | Part deletion of the βγ-insert |
| *E. coli* DnaK ($E_{209}K$) | 100 | ~50% colonies | Modification of the βγ-insert |
| *E. coli* DnaK ($D_{208}G$ and $E_{209}K$) | 100 | ~50% colonies | Modification of the βγ-insert |
| *E. coli* DnaK with two aa (SG), corresponding to the insert found in some other bacteria, inserted after position 385 | 100 | 0 | This insert is not found in βγ-proteobacteria |
| DnaK, *S. meliloti* | 100 | 0 | Lacks βγ-insert |
| DnaK, *P. aeruginosa* | 100 | 0 | Contains insert |

Complementation of the $T_s$ phenotype of *E. coli* CG800 (C600 *dnaK103*) mutant cells upon transformation with different *dnaK* gene constructs from *E. coli* or other bacteria. The experiment was carried out as described in "Experimental procedures". Assuming the number of colonies observed at 30°C to be 100%, the relative plating efficiency of the cells at 42°C was determined

insert region in the *E. coli* DnaK protein was not resolved (Harrison et al. 1997).

To determine the functional significance of this insert, complementation studies were carried out with *E. coli dnaK* gene where this four aa (12 nt) insert was deleted (Table 2). The plasmid lacking this insert was unable to complement the $T_s$ growth defect of the *dnaK* mutant, indicating that it is essential for the proper function of the protein in *E. coli*. We have also examined the effect of a one aa deletion (deletion of $E_{209}$) within this insert and the plasmid with this deletion showed very poor plating efficiency at 42°C (<0.5% in comparison to the + insert plasmid) provid-

ing further confirmation that this insert is essential for the proper function of the protein in *E. coli*. These mutant proteins were expressed at a similar level as the WT DnaK protein (Fig. 3b), indicating that their inability to support the growth of mutant *E. coli* was not due to their lack of expression. We have also replaced some amino acids in this proteobacterial insert with those found in the *Micrococcineae* and *Bifidobacteria*. Interestingly, the mutant plasmids containing these changes (viz. $E_{209}K$ or a double mutant $D_{208}G$ and $E_{209}K$) in the *E. coli* DnaK partially supported (about 50% plating efficiency) the growth of the DnaK mutant cell at 42°C. We also studied the ability of DnaK

**Fig. 6** Excerpt from DnaK sequence alignment showing a four aa conserved insert (*boxed*) that is specific for the β and γ proteobacteria, except as noted below. A 4–5 aa insert of independent origin is also present in this position in various *Micro-coccineae* and *Bifidobacteria*. Other details are the same as in Figs. 1 and 4

```
                                                     190                                          233
Gamma              Escherichia coli            ZP_00719907   IAVYDLGGGTFDISIIEI DEVD GEKTFEVLATNGDTHLGGEDFD
Proteobacteria     Haemophilus influenzae      ZP_01790143   ------------------ -NF- --Q-------G-N---------
(223/0)            Vibrio cholerae             ZP_01681355   ------------------ ---E --------S------------
                   Shewanella baltica          YP_001051666  V----------------- -SN- --DQ-----------------
                   Xylella fastidiosa          NP_779568     ------------V----- A--- ---Q---------F-------
                   Coxiella burnetii           NP_820282     ------------------ A--- --HQ---------F-------
                   Legionella pneumophila      YP_001250804  ------------------ A--- --HQ---------F-------
                   Pseudomonas aeruginosa      YP_001350803  VI----------V-V--- A--- --HQ---------F-------
                   Thiomicrospira crunogena    YP_391140     -----------------V ADL- ---QV---S-----F-----
                   Francisella tularensis      ACD30966      V----------------- AD-- -DNQI---S-----F-----
Beta               Ralstonia solanacearum      NP_520756     ------------------ AD-- ---Q-----S-----D----
Proteobacteria     Nitrosomonas europaea       NP_841967     ------------------ A--E --HQ----------F-----
(79/0)             Neisseria meningitidis      YP_974594     V----------------- ANL- -D-Q----------F-----
                   Dechloromonas aromatica     YP_284149     ------------------ AN-- --TQ----------F-----
                   Thiobacillus denitrificans  YP_315296     ------------------ A-M- --HQ-----S-----F----
                   Chromobacterium violaceum   NP_901313     ------------V----- AD-- --HQ-----S-----F----
                   Burkholderia mallei         AAY63995      ------------V----- AD-- --MQ-----S-----F----
                   Bordetella pertussis        NP_881126     -V----------V----- AD-- --MQ-----S-----F----
Other              Caulobacter crescentus      AAA62723      ------------V--L-- -DGV---KS-----F------
Proteobacteria     Rickettsia prowazekii       NP_220574     ------------V--L-- SDGV---KS-----F------
(1/295)            Brucella suis               NP_699100     ------------V-VL-- -DGV---KS-----F------
                   Rhodobacter sphaeroides     YP_354258     -------------T-L-- DDGL---KS------F-----
                   Myxococcus xanthus          AAC64205      --------------L-L NAGV---KS-----F------
                   Geobacter metallireducens   YP_386462     ---F----------L-L -DGV---KS-----F------
                   Desulfovibrio vulgaris      YP_010032     ---F----------L-V -DNVV--R------F-----
                   Campylobacter jejuni        YP_001000443  -V----------VTVL-T -DNVV-----G-NAF---D--
                   Helicobacter pylori         YP_626850     -M----------VTVL-T -DNVV-----G--AF---D--
Other              Bacteroides fragilis        YP_098509     ---F----------L-F -GGV---S--------D---
Gram-negative      Salinibacter ruber          YP_445424     V----------V--L-L -DGV---N--Y-------DN--
Bacteria           Chlamydia trachomatis       YP_328213     ---F----------L-- -DGV---S---------D---
(0/370)            Protochlamydia amoebophila  YP_008498     ---F----------VL- -DGV----S-S------D---
                   Aquifex aeolicus            NP_21368      -L---F------V-L-G --GVI--KV-A-------ANI-
                   Chlorobium tepidum          NP_661540     V--F----------L-L -GGV---KS-D------D---
                   Planctomyces maris          CAA75091      ---F----------LDV ADGV----S-S------D-W-
                   Borrelia burgdorferi        NP_212652     V-------------L-L -DGV---KS--------DN--
                   Treponema pallidum          NP_218656     --------------L-L -DGV---KS---------D---
                   Synechocystis sp. PCC 6803  NP_441989     -L-F--------V--L-V --GV------S-------D---
                   Thermosyn. elongatus        NP_682523     -L-F--------V--L-V -DGV------S-------D---
                   Nostoc punctiforme          ZP_00111247   -L-F--------V--L-V --GV---K--S-------D---
                   Thermus thermophilus        YP_005096     VL-F--------VTVL-- --GV---K--S-------S-M-
                   Deinococcus radiodurans     NP_293855     VL-F--------VT-L-L -DGV---KS-A--------A---
                   Thermotoga maritima         AAC77523      VL----------V--L-- --GVT--I--R-NN----D---
Other              Streptomyces coelicolor     NP_733614     -L-F--------V-LL-- -DGVV--K-----NN---D-W-
Actinobacteria     Mycobacterium leprae        NP_302613     -L-F--------V-LL-- --GVV--R--S--N----D-W-
(0/107)            Propioni. freudenreichii    CAD59395      VL-F--------V-LLD- SDGV---K-----NR---D-W-
                   Thermobifida fusca          YP_288257     -L-F--------V-LLDV -DGVV--K--H--N----D-W-
Micrococcineae +   Tropheryma whipplei         NP_7878781    -L-F--------V-LL-V GKD- DFS-IQ-R--S--NR---D-W-
Bifidobacteria     Leifsonia xyli              YP_0631211    -L-F--------V-LL-V GKD- DFS-IQ-R--A--NR---D-W-
(31/31)            Clavibacter michiganensis   YP_001220890  -L-F--------V-LL-V GKD- DFS-IQ-RS-A--NR---D-W-
Firmicutes         Mycoplasma genitalium       NP_072972     VL----------V-LLD- A-G------A--NR---D-W-
(1/208)            Clostridium perfringens     ZP_02864700   -L----------V--LDL -DGV---VS----AR---D---
                   Staphylococcus aureus       NP_372104     VL-F--------V--L-L -DGV----S-A--NK---D---
                   Streptococcus pneumoniae    AAC15892      -L-F--------V--L-L -DGV-D--S-A--NK---D---
                   Lactococcus lactis          YP_001032859  -L-F--------V--L-L -DGV-D----A-NNK---D---
                   Bacillus subtilis           CAA34429      -LL---------V--L-L -DGV---RS-A--NR---D---
```

homologs from *P. aeruginosa* and *S. meliloti* to complement the $T_s$ defect of the *E. coli dnaK* mutant. However, both these homologs despite their good expression (Fig. 3b) were inactive in the complementation test. The failure of the *P. aeruginosa* DnaK homolog, which contained various conserved inserts found in the *E. coli* protein, was somewhat surprising and it suggests that other changes in the molecule besides the inserts are responsible for its lack of complementation. In earlier work, DnaK homologs from *B. megaterium* (Sussman and Setlow 1987), *Borrelia burgdorferi* (Tilly et al. 1993), and rat Hsc70 (Suppini et al. 2004), were also found to be ineffective in complementing the $T_s$ defect of the same *E. coli* $T_s$ mutant.

We have identified yet another conserved insert in the Hsp70 protein that is present in a number of clostridia as well as few *Rickettsiae* and some other bacteria (Fig. 7). Unlike the other inserts described above, the groups of bacteria containing this insert are phylogenetically unrelated and they branch in very different positions in various phylogenetic trees (Gupta 1998; Ciccarelli et al. 2006; Viale et al. 1994). In the Hsp70 trees, these groups do not cluster together (Gupta 1998) indicating that the shared presence of this insert in these species is not due to lateral gene transfers. The most likely explanation for this two aa indel is

that it has occurred independently in these groups of bacteria. This insert in the Hsp70 is present in the unstructured part of the molecule (between residues 385 and 386 in the *E. coli* DnaK protein) that links the ATPase or nucleotide-binding domain of the molecule to the substrate-binding domain of the protein (Harrison et al. 1997; Pellecchia et al. 2000; Suppini et al. 2004). It is of interest that this two aa insert, although present in a number of unrelated bacteria, was not found in any of the β–γ proteobacteria. Thus, it was of interest to determine whether *E. coli* DnaK can tolerate this insert. To examine this, a two aa insert (SG) corresponding to that found in some of these bacteria was introduced into the *E. coli* DnaK in the same position where it is found in other bacteria. The resulting mutant protein despite good expression (Fig. 3b) showed no complementation of the *dnaK* mutant (Table 2); this observation indicates that this insert is incompatible with the proper function of the DnaK in *E. coli* cells.

## Discussion

In this work, we have examined the biological significance of several conserved inserts in the Hsp60 and Hsp70

**Fig. 7** Partial sequence alignment of DnaK showing a two aa insert (*boxed*) that is present in a number of bacteria belonging to different groups. This insert is not present in other bacteria and it is also not found in any of the β–γ proteobacteria, which includes *E. coli*. The *numbers* on the *top* indicate the position of this sequence in the *Clostridium bolteae* and *E. coli* (in parenthesis) proteins. In the DnaK structure, this insert is located in the linker region that joins the ATPase domain to the substrate-binding domain

```
                                                                 338(367)                 376(403)
           ┌ Clostridium bolteae           ZP_02084488   PDECVAIGAAIQGGKLAGD AG AGDILLLDVTPLSLSIET
           │ Clostridium sp. SS2/1          ZP_02440147   ---------SV-------E  -- ------------------
           │ Clostridium sp. L2-50          ZP_02076109   ---------S---------  -- --E--------T-----
           │ Clostridium phytofermentans    YP_00155941   ---------SV--------  T- ------------------
           │ Anaerostipes caccae            ZP_02419215   ---------S---------  -- ------------------
           │ Epulopiscium sp. 'N.t'         ZP_0269286    ------------------- -- -----------T-G---
Species with┤ Prevotella loescheii          Q93GF1        ---V--V------AI-NKE S- V---V--------T-G---
  insert   │ Prevotella albensis            CAB46748      ---V--V------AV-NKE E- V-N-V--------T-G---
           │ Thermosyn. elongatus           NP_681580     ---A----------I-GKE TT VK-L----------GL--
           │ Fibrobacter succinogenes       AAK130221     ---V------V-AAV-S-- SA VK-VA--------G---
           │ Neorickettsia sennetsu         YP_505922     ---V------ANI---G  SD -Q--V--------G---
           └ Ehrlichia sennetsu             AAC27487      ---V------ANI---G  SD -Q--V--------G---
           ┌ Escherichia coli               ZP_00719907   ---A------V---V-T-- VK-V----------G---
           │ Haemophilus influenzae         ZP_01792567   ---A------V--V-K-- VK-V----------G---
           │ Yersinia pestis                NP_404110     ---A------V--V-S-E VK-V----------G---
           │ Vibrio cholerae                ZP_01980924   ---A--V---V---V---E VK-V----------G---
           │ Shewanella putrefaciens        ZP_01705390   ---A--V-----A-V-S-- VK-V----------G---
           │ Thiobacillus denitrificans     YP_315296     ---A--V--------Q-E VK-V----------G---
           │ Xanthomonas campestris         YP_363295     ---A--V-------V---- VK-V----------G---
           │ Francisella tularensis         YP_001121600  ---A--V-------V-S-- VK------------G---
           │ Nitrosomonas europaea          NP_841967     ---A----------V-K-- VK-V----------G---
Proteobacteria┤ Dechloromonas aromatica     YP_284149     ---A--V-------V-Q-E VK-V----------G---
           │ Rhodospirillum rubrum          YP_428636     ---V--M-------V-K-- VK-V----------G---
           │ Ehrlichia ruminantium          YP_180413     ---V---------I---- VR-V----------G---
           │ Bartonella henselae            YP_032930     ---V--M-------V-Q-- VK-V----------G---
           │ Bdellovibrio bacteriovorus     NP_96820      ---V------V---V--- VK-V----------G---
           │ Geobacter sulfurreducens       NP_951095     ---V----------V-R-- VK-V----------G---
           └ Desulfuromonas acetoxidans     ZP_01313919   ---V-S--------V-K-- VK-V----------G---
            ┌ Helicobacter pylori           YP_626850     ---V--V--S----V-K-- VK-V----------G---
            │ Bacteroides fragilis          YP_098509     ---V--V--V--V-AV-TDE IKGVV--------MG---
            │ Porphyromonas gingivalis      NP_905399     ---V--V-------V-T-E VK-V----------G---
            │ Flavobacteria bacterium       ZP_01203002   ---V----------V-T-- VK-V----------G---
            │ Cytophaga hutchinsonii        YP_679683     ---V--L-------V-T-E VT-V-----I----G---
Other Bacteria┤ Salinibacter ruber          YP_445424     ---V-SL---V---V-S-- VD-V--------N-G---
            │ Chlorobium tepidum            NP_661540     ---V----------V-Q-- VT-V----------G---
            │ Pelodictyon luteolum          YP_374547     ---V----------V-K-- VT-V----------G---
            │ Treponema denticola           NP_971242     ---A--M-------I-G-- VK-V----------G---
            │ Borrelia hermsii              YP_001883946  ---A----------I-T-E TK-MV--------G---
            │ Lentisphaera araneosa         ZP_0187544    ------L-------V-T-- VN-V----------G---
            │ Lactobacillus plantarum       NP_785552     ---A--L---V---VIT-- VK-VV--------G---
            └ Bacillus sphaericus           O69268        ---V--M---V---V-T-- VK-VV--------G---
```

proteins. These inserts (or indels) are located in highly conserved regions of the proteins and they are uniquely shared by various species from particular groups or phyla of bacteria. The species distribution patterns of these conserved indels strongly suggest that the genetic changes responsible for them occurred at important evolutionary branch points in the course of bacterial evolution (Fig. 2). After their introduction into these genes, these indels have been retained, apparently without any loss, by all descendant species diverging from these evolutionary branch points. The fact that similar indels are either completely absent or rarely observed in other groups of bacteria strongly indicates that these indels constitute rare genetic events and they do not occur frequently or randomly in these genes. In accordance with their highly conserved nature, results of our studies presented here show that deletion or any substantial changes in these conserved inserts lead to either complete loss or nearly complete loss (>95%) of the ability of the mutant Hsp60 or Hsp70 genes/proteins to complement the $T_s$ growth defects of the *E. coli* mutants affected in these proteins. Similar results were obtained for all four conserved inserts in these proteins that were studied in this work. These results provide strong evidence that evolutionary conserved inserts in protein sequences are essential for the proper functions of these proteins in *E. coli* and possibly all other bacteria that commonly share these characteristics. It should be acknowledged that in the present work cellular function of various Hsp60 or Hsp70 proteins affected in these insert(s) was studied only by their ability to support the growth of the $T_s$ mutants at non-permissive temperature. Although the mutant proteins affected in these inserts were inactive in these regards, it is possible that they may retain limited functions with regard to some of the specific steps in the functioning of these proteins (e.g., ATPase activity, binding to their substrates or co-chaperone, etc.) (Fenton et al. 1994; Sugimoto et al. 2007).

Of the inserts whose functions were studied, the indel in the GroEL protein consisted of only a one aa insert. Smaller indels in protein sequences are generally considered to be not as important as larger indels (Artsimovitch et al. 2003; Rokas and Holland 2000; Gupta 1998). However, deletion of this one aa insert from the GroEL protein, or most amino acid substitutions in this position, led to either complete inactivation or nearly complete loss of the activity of this protein in supporting the growth of mutant *E. coli* cells. These results demonstrate that not only this one aa insert is essential for the proper functioning of the GroEL protein in *E. coli*, but that the specific amino acids found in this position are also of critical importance and they are under strong selection pressure. Except for cyanobacteria, Hsp60 homologs from all other phyla of bacteria as well as mitochondria (where this insert is found) contain an asparagine (N) residue in this position. However, both in cyanobacteria as well as various plastids, which are derived from cyanobacteria (Palmer and Delwiche 1998; Gupta et al. 2003), a change from N → G has occurred in this position. The inability of the $N_{153}$ → G mutant to function in *E. coli* provides evidence that this change is tolerated, and very likely

required, in a specific genetic background, as in cyanobacteria and plastids.

We have also examined the effect of a conserved insert in the Hsp70 protein that is specific for other groups of bacteria, but not found in *E. coli*, on the function of the *E. coli* protein. We have described a two aa insert in DnaK in this work that is commonly shared by several *Clostridia*, a few *Rickettsiae* and some other bacteria, but which is not found in *E. coli* or various other β or γ- proteobacteria. When this two aa insert was introduced into the *E. coli* DnaK in the same position as found in the other bacteria, the resulting protein was completely inactive in complementing the growth defect of the *E. coli* DnaK $T_s$ mutant. This result provides strong evidence that these inserts are highly specific in terms of their functional requirement by particular groups of bacteria where they are found and they do not constitute random genetic events in functionally less constrained parts of the proteins. This two aa insert in the Hsp70 protein is located within the linker region that joins the nucleotide-binding domain of the protein with the substrate-binding domain (Pellecchia et al. 2000; Suppini et al. 2004). Based upon the high degree of flexibility that is shown by these two domains (Rist et al. 2006; Popp et al. 2005), one may expect that the length or the sequence of this linker region may not be important for the function of this protein. However, our studies indicate that this is not the case; certain changes in this region such as this insert that are specific (and presumably required) for particular groups of bacteria are not compatible with the function of the Hsp70 protein in other bacteria (*E. coli*). Recent studies by others have also provided evidence that the linker region of DnaK/Hsp70 plays an important role in the function of this protein (Popp et al. 2005; Rist et al. 2006). The work by Rist et al. (2006) indicates that the linker region undergoes a conformational change from a solvent-exposed to a solvent-protected state upon addition of ATP and this change is reversed upon binding of the substrate peptide to the DnaK–ATP complex.

All of the conserved inserts in the GroEL and DnaK proteins that we have studied are located on the surface of these proteins and they are generally found in loop regions connecting other secondary structural elements. This is also true of conserved inserts in a number of other proteins that have been studied (Artsimovitch et al. 2003; Opalka et al. 2000) (Gupta RS, unpublished results). Although our studies provide convincing evidence that these conserved inserts (indels) are essential for the species in which they are found, there is no information available at present regarding their cellular functions. Most of the proteins in which such conserved indels have been identified (e.g., RpoA, RpoB, RpoC, SecA, PolA, EF-Tu, GyrA, GyrB, several ribosomal proteins and aminoacyl-tRNA synthetases, etc.) (Griffiths et al. 2005; Gao and Gupta 2005; Griffiths and Gupta 2004; Gupta 2004; Gupta et al. 2003) carry out important housekeeping functions. It is expected that these functions are still carried out by these inserts-containing proteins. In view of this and the location of these inserts on the protein surfaces, it is possible that these conserved inserts are imparting some new functional capability (i.e., an ancillary function) on these proteins, which is essential for the groups of organisms where they are found. This ancillary function could include altered biochemical regulation of the protein due to its interaction with some other ligands or proteins in the cells (with the insert serving an important role in enabling this interaction), or some other novel function.

It is of much interest that recent comparative genomic studies have identified numerous proteins that are also specific for the same taxonomic groups of bacteria as these and other conserved indels (Galperin and Koonin 2004; Daubin and Ochman 2004; Gupta and Griffiths 2006; Gupta and Mok 2007; Bork and Koonin 1998). The cellular functions of most of these lineage-specific proteins are not known (Bork and Koonin 1998; Danchin 1999). However, the presence of certain conserved indels as well as lineage-specific proteins in the same groups of bacteria indicates that these genetic characteristics have coevolved. Therefore, it is possible that some of these conserved indels in the essential proteins could be interacting with one or more of the lineage-specific proteins, thereby leading to some new functional or regulatory capabilities that are specific and essential characteristics of particular groups of bacteria. Thus, further studies on understanding the cellular functions of these conserved indels are of much interest. In addition to providing novel information regarding the cellular functions of these important house keeping proteins, these studies may also lead to identification of novel biochemical and/or physiological properties that are unique and distinctive characteristics of different groups of bacteria.

## References

Ahmad S, Ahuja R, Venner TJ, Gupta RS (1990) Identification of a protein altered in mutants resistant to microtubule inhibitors as a member of the major heat shock protein (hsp70) family. Mol Cell Biol 10:5160–5165

Ang D, Georgopoulos C (1989) The heat-shock-regulated grpE gene of *Escherichia coli* is required for bacterial growth at all temperatures but is dispensable in certain mutant backgrounds. J Bacteriol 171:2748–2755

Arifuzzaman M, Oshima T, Mori H (2004) The ATPase domain of HscC (DnaK homolog) is essential for interfering sigma70 activity in *E. coli*. FEMS Microbiol Lett 230:99–104

Artsimovitch I, Svetlov V, Murakami KS, Landick R (2003) Co-overexpression of *Escherichia coli* RNA polymerase subunits allows isolation and analysis of mutant enzymes lacking lineage-specific sequence insertions. J Biol Chem 278:12344–12355

Baldauf SL, Palmer JD (1993) Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. Proc Natl Acad Sci USA 90:11558–11562

Bork P, Koonin EV (1998) Predicting functions from protein sequences—where are the bottlenecks. Nat Genet 18:313–318

Braig K, Otwinowski Z, Hegde R, Boisvert DC, Joachimiak A, Horwich AL, Sigler PB (1994) The crystal structure of the bacterial chaperonin GroEL at 2.8 A. Nature 371:578–586

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

Bukau B, Walker GC (1989) Cellular defects caused by deletion of the *Escherichia coli* dnaK gene indicate roles for heat shock protein in normal metabolism. J Bacteriol 171:2337–2346

Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P (2006) Toward automatic reconstruction of a highly resolved tree of life. Science 311:1283–1287

Craig EA, Gambill BD, Nelson RJ (1993) Heat shock proteins: molecular chaperones of protein biogenesis. Microbiol Rev 57:402–414

Danchin A (1999) From protein sequence to function. Curr Opin Struct Biol 9:363–367

Daubin V, Ochman H (2004) Bacterial genomes as new gene homes: the genealogy of ORFans in *E. coli*. Genome Res 14:1036–1042

Doolittle WF (1999) Phylogenetic classification and the universal tree. Science 284:2124–2128

Ellis RJ, Hartl FU (1999) Principles of protein folding in the cellular environment. Curr Opin Struct Biol 9:102–110

Embley TM, Stackebrandt E (1993) The molecular phylogeny and systematics of the *actinomycetes*. Annu Rev Microbiol 48:257–289

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

Fenton WA, Kashi Y, Furtak K, Horwich AL (1994) Residues in chaperonin GroEL required for polypeptide binding and release. Nature 371:614–619

Flaherty KM, DeLuca-Flaherty C, McKay DB (1990) Three-dimensional structure of the ATPase fragment of a 70 K heat-shock cognate protein. Nature 346:623–628

Galperin MY, Koonin EV (2004) 'Conserved hypothetical' proteins: prioritization of targets for experimental study. Nucleic Acids Res 32:5452–5463

Gao B, Gupta RS (2005) Conserved indels in protein sequences that are characteristic of the phylum *Actinobacteria*. Int J Syst Evol Microbiol 55:2401–2412

Georgopoulos CP, Hendrix RW, Casjens SR, Kaiser AD (1973) Host participation in bacteriophage lambda head assembly. J Mol Biol 76:45–60

Gogarten JP, Doolittle WF, Lawrence JG (2002) Prokaryotic evolution in light of gene transfer. Mol Biol Evol 19:2226–2238

Gray MW (1999) Evolution of organellar genomes. Curr Opin Genet Dev 9:678–687

Griffiths E, Gupta RS (2004) Signature sequences in diverse proteins provide evidence for the late divergence of the order *Aquificales*. Int Microbiol 7:41–52

Griffiths E, Petrich A, Gupta RS (2005) Conserved indels in essential proteins that are distinctive characteristics of *Chlamydiales* and provide novel means for their identification. Microbiology 151:2647–2657

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

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

Gupta RS (2000) The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. FEMS Microbiol Rev 24:367–402

Gupta RS (2004) The Phylogeny and signature sequences characteristics of *Fibrobacters*, *Chlorobi* and *Bacteroidetes*. Crit Rev Microbiol 30:123–143

Gupta RS, Golding GB (1993) Evolution of HSP70 gene and its implications regarding relationships between archaebacteria, eubacteria, and eukaryotes. J Mol Evol 37:573–582

Gupta RS, Griffiths E (2006) Chlamydiae-specific proteins and indels: novel tools for studies. Trends Microbiol 14:527–535

Gupta RS, Mok A (2007) Phylogenomics and signature proteins for the alpha proteobacteria and its main groups. BMC Microbiol 7:106

Gupta RS, Singh B (1994) Phylogenetic analysis of 70 kD heat shock protein sequences suggests a chimeric origin for the eukaryotic cell nucleus. Curr Biol 4:1104–1114

Gupta RS, Aitken K, Falah M, Singh B (1994) Cloning of *Giardia lamblia* heat shock protein HSP70 homologs: implications regarding origin of eukaryotic cells and of endoplasmic reticulum. Proc Natl Acad Sci USA 91:2895–2899

Gupta RS, Pereira M, Chandrasekera C, Johari V (2003) Molecular signatures in protein sequences that are characteristic of cyanobacteria and plastid homologues. Int J Syst Evol Microbiol 53:1833–1842

Harrison CJ, Hayer-Hartl M, Di Liberto M, Hartl F, Kuriyan J (1997) Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK. Science 276:431–435

Jindal S, Dudani AK, Singh B, Harley CB, Gupta RS (1989) Primary structure of a human mitochondrial protein homologous to the bacterial and plant chaperonins and to the 65-kilodalton mycobacterial antigen. Mol Cell Biol 9:2279–2283

Klein G, Georgopoulos C (2001) Identification of important amino acid residues that modulate binding of *Escherichia coli* GroEL to its various cochaperones. Genetics 158:507–517

Lake JA, Herbold CW, Rivera MC, Servin JA, Skophammer RG (2007) Rooting the tree of life using nonubiquitous genes. Mol Biol Evol 24:130–136

Macario AJL, Lange M, Ahring BK, De Macario EC (1999) Stress genes and proteins in the archaea. MMBR 63:923–967

Margulis L (1993) Symbiosis in cell evolution. W.H. Freeman, New York, p 452

Opalka N, Mooney RA, Richter C, Severinov K, Landick R, Darst SA (2000) Direct localization of a beta-subunit domain on the three-dimensional structure of *Escherichia coli* RNA polymerase. Proc Natl Acad Sci USA 97:617–622

Palmer JD, Delwiche CF (1998) The origin and evolution of plastids and their genomes. In: Sotis DE, Soltis PE, Doyle JJ (eds) Molecular systematics of plants II DNA sequencing. Kluwer, Norwell, pp 375–409

Pellecchia M, Montgomery DL, Stevens SY, Vander Kooi CW, Feng HP, Gierasch LM, Zuiderweg ER (2000) Structural insights into substrate binding by the molecular chaperone DnaK. Nat Struct Biol 7:298–303

Picketts DJ, Mayanil CS, Gupta RS (1989) Molecular cloning of a Chinese hamster mitochondrial protein related to the "chaperonin" family of bacterial and plant proteins. J Biol Chem 264:12001–12008

Popp S, Packschies L, Radzwill N, Vogel KP, Steinhoff HJ, Reinstein J (2005) Structural dynamics of the DnaK–peptide complex. J Mol Biol 347:1039–1052

Richardson A, Schwager F, Landry SJ, Georgopoulos C (2001) The importance of a mobile loop in regulating chaperonin/co-chaperonin

interaction: humans versus *Escherichia coli*. J Biol Chem 276:4981–4987

Rist W, Graf C, Bukau B, Mayer MP (2006) Amide hydrogen exchange reveals conformational changes in hsp70 chaperones important for allosteric regulation. J Biol Chem 281:16493–16501

Rivera MC, Lake JA (1992) Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. Science 257:74–76

Rodriguez-Quinones F, Maguire M, Wallington EJ, Gould PS, Yerko V, Downie JA, Lund PA (2005) Two of the three groEL homologues in Rhizobium leguminosarum are dispensable for normal growth. Arch Microbiol 183:253–265

Rokas A, Holland PW (2000) Rare genomic changes as a tool for phylogenetics. Trends Ecol Evol 15:454–459

Rusanganwa E, Gupta RS (1993) Cloning and characterization of multiple groEL chaperonin-encoding genes in *Rhizobium meliloti*. Gene 126:67–75

Sugimoto S, Higashi C, Saruwatari K, Nakayama J, Sonomoto K (2007) A gram-negative characteristic segment in *Escherichia coli* DnaK is essential for the ATP-dependent cooperative function with the co-chaperones DnaJ and GrpE. FEBS Lett 581:2993–2999

Suppini JP, Amor M, Alix JH, Ladjimi MM (2004) Complementation of an *Escherichia coli* DnaK defect by Hsc70-DnaK chimeric proteins. J Bacteriol 186:6248–6253

Sussman MD, Setlow P (1987) Nucleotide sequence of a *Bacillus megaterium* gene homologous to the dnaK gene of *Escherichia coli*. Nucleic Acids Res 15:3923

Tilly K, Hauser R, Campbell J, Ostheimer GJ (1993) Isolation of dnaJ, dnaK, and grpE homologues from *Borrelia burgdorferi* and complementation of *Escherichia coli* mutants. Mol Microbiol 7:359–369

Ventura M, Canchaya C, Tauch A, Chandra G, Fitzgerald GF, Chater KF, van Sinderen D (2007) Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. Microbiol Mol Biol Rev 71:495–548

Viale AM, Arakaki AK, Soncini FC, Ferreyra RG (1994) Evolutionary relationships among eubacterial groups as inferred from GroEL (chaperonin) sequence comparisons. Int J Syst Bacteriol 44:527–533

Vickery LE, Cupp-Vickery JR (2007) Molecular chaperones HscA/Ssq1 and HscB/Jac1 and their roles in iron–sulfur protein maturation. Crit Rev Biochem Mol Biol 42:95–111

Wild J, Kamath-Loeb A, Ziegelhoffer E, Lonetto M, Kawasaki Y, Gross CA (1992) Partial loss of function mutations in DnaK, the *Escherichia coli* homologue of the 70-kDa heat shock proteins, affect highly conserved amino acids implicated in ATP binding and hydrolysis. Proc Natl Acad Sci USA 89:7139–7143

Zeilstra-Ryalls J, Fayet O, Georgopoulos C (1991) The universally conserved GroE (Hsp60) chaperonins. Annu Rev Microbiol 45:301–325