Cloning and sequence analysis of the *dnaK* gene region of *Lactococcus lactis* subsp. *lactis*

TRACY EATON, CLAIRE SHEARMAN and MICHAEL GASSON*

AFRC Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, UK

(Received 15 April 1993; revised 13 July 1993; accepted 22 July 1993)

A 5.4 kb *Hind*III fragment of *Lactococcus lactis* subsp. *lactis* was identified using a homologous *dnaK* probe generated by PCR and cloned in *Escherichia coli*. Upstream sequences were generated by inverse PCR. The two cloned fragments partially overlapped, and sequencing of 5915 bp revealed the presence of four open reading frames in the order *orf1*-*grpE*-*dnaK*-*orf4*. *orf1* encodes a 39 kDa protein of unknown function which shows considerable sequence homology with the Orf39 and Orf4 proteins of *Bacillus subtilis* and *Clostridium acetobutylicum*, respectively. The downstream ORFs showed high homology to the *grpE* and *dnaK* genes of other prokaryotes. The DnaK protein has a characteristic 24-amino-acid deletion exhibited by all the known DnaK proteins of Gram-positive species. In many bacteria the *dnaK* and *dnaJ* genes are found as part of the same operon. The *L. lactis* *dnaK* operon is unusual in that the *dnaK* gene is followed by a putative transcription terminator and a fourth large ORF which shares no homology with the *dnaJ* genes of other bacteria but has a small degree of homology with various membrane proteins. Vegetative promoter sequences are found upstream of both *orf1* and *orf4*. A 12 bp inverted repeat is found upstream of the putative promoter of *orf1* and an 8 bp inverted repeat is found between this promoter and the *orf1* initiation codon. These repeats are thought to be involved in regulation of the heat-shock genes. The DnaK homologue is induced approximately 3-fold on heat shock at 42 °C.

Introduction

Lactic acid bacteria are important in many industrial applications and are used extensively in the dairy industry (Marshall, 1987). Analysis of inducible gene expression and characterization of regulatory elements are important areas of research involved in the genetic improvement of industrial strains. Strong controllable promoters facilitate the efficient and inducible expression of both heterologous and homologous genes. The principal aim of this work was to isolate and characterize potential inducible genes, their promoter regions and possible regulatory sequences in *Lactococcus lactis*.

The majority of organisms exhibit a heat-shock response when subjected to a sudden temperature increase above their normal growth temperature. Heat shock rapidly induces the synthesis of a group of specific heat-shock proteins (HSPs) (for a review see Morimoto

et al., 1990). Other inducers such as osmotic shock (Meury & Kohiyama, 1991; Völker *et al.*, 1992), pH shifts (Heyde & Portalier, 1990; Taglicht *et al.*, 1987) and unfolded and abnormal proteins (Parsell & Sauer, 1989; Goff & Goldberg, 1985) are all known to induce particular HSPs to various degrees. In *Escherichia coli*, mutations in the *dnaK* gene block bacteriophage lambda DNA replication at all temperatures (Georgopoulos & Herskowitz, 1971) and DnaK is essential for cell viability at high and low temperatures (Paek & Walker, 1987). At normal temperatures DnaK is involved in the synthesis of RNA and DNA and in cell division (Paek & Walker, 1987; Sakakibara, 1988).

The HSPs are highly conserved at the amino acid level and appear to have primary roles as molecular chaperones in mediating the correct folding and assembly of cellular proteins (Gething & Sambrook, 1992; Langer *et al.*, 1992). In *L. lactis*, 13 to 16 proteins are induced on heat shock at 42 °C (Whitaker & Batt, 1991; Boutibonnes *et al.*, 1992). Three of these appear to be homologues of GroEL, GrpE and DnaK based on their molecular masses and reactivity with the respective antisera. However, there is no information on the regulation of the heat-shock response or the function of HSPs in *L. lactis*. PCR has been used to generate a homologous

---

* Author for correspondence. Tel. +44 603 255248; fax +44 603 58414.

Abbreviation: HSPs, heat-shock proteins.

The EMBL accession number for the sequence reported in this paper is Z19148.

Table 1. *Bacterial strains and plasmids used in this study*

Strain or plasmid	Characteristics	Source or reference
Strain		
<i>E. coli</i> MC1022	<i>araD139</i> Δ (<i>ara leu</i>) (7697) Δ (<i>lacZ</i>)M15 <i>galU galK strA</i>	Casadaban & Cohen (1980)
<i>E. coli</i> SURE	Δ (<i>hsdRMS</i>) <i>mcrA mcrB mrr endA1 supE44 thi-1 gyrA96 relA1 lacrecB recJ sbcC umuC uvrC</i> [F' <i>proAB lacI^q</i> Δ (M15 Tn10) (Tetr)]	Stratagene
<i>E. coli</i> JM109	[<i>recA1 endA1 gyrA96 thi hsdR17</i> (rk ⁺ mk ⁺) <i>supE44 relA1</i> λ ⁻ Δ (<i>lac-proAB</i>) (F' <i>traD36 proAB⁺ lacI^q lacZ</i> Δ M15)]	Yanisch-Perron <i>et al.</i> (1985)
<i>L. lactis</i> subsp. <i>lactis</i> MG1363	Plasmid-free strain	Gasson (1983)
Plasmid		
pUC18	Ampicillin resistance	Yanisch-Perron <i>et al.</i> (1985)
pBR322	Ampicillin and tetracycline resistance	Bolivar <i>et al.</i> (1977)
pFI483	346 bp <i>dnaK</i> PCR fragment cloned into <i>SmaI</i> site of pUC18	This study
pFI573	pBR322 containing 5.4 kb chromosomal <i>dnaK</i> <i>HindIII</i> fragment	This study
pFI574 (pFI619)	1.6 kb <i>EcoRI</i> fragment of pFI573 cloned into <i>SmaI</i> site of pUC18 (pFI619, opposite orientation)	This study
pFI575 (pFI620)	1.8 kb <i>EcoRI</i> fragment of pFI573 cloned into <i>SmaI</i> site of pUC18 (pFI620, opposite orientation)	This study
pFI576	2.0 kb <i>EcoRI</i> fragment of pFI573 cloned into <i>SmaI</i> site of pUC18	This study
pFI672	2.3 kb inverse PCR fragment cloned into <i>EcoRV</i> site of pBR322	This study

Table 2. *Oligonucleotide primers used for PCR amplification*

Primer	Sequence	Description
P10	5'-GGTATCGATTAGGTACAAC-3'	Conserved region $\alpha\alpha$ 6-12* (6-12)†
P9	5'-TTAAARTANGCNGGACNGT-3'†	Conserved region $\alpha\alpha$ 115-121 (116-122)
P69	5'-GCAACTGCTGAAAGCTACCTTG-3'	Inverse PCR primer nucleotide no. 2225-2246 (Fig. 3)
P100	5'-GCATTCGCTCATCACT-3'	Inverse PCR primer nucleotide no. 1003-1021 (Fig. 3)

* Amino acid ($\alpha\alpha$) notation refers to *B. megaterium* DnaK (Sussman & Setlow, 1987).

† R = A or G, N = A, C, G or T.

‡ $\alpha\alpha$ notation refers to *L. lactis* subsp. *lactis* DnaK (this study).

probe for *L. lactis dnaK* based on multiple sequence alignments of DnaK homologues which reveal regions of particularly high homology. This report describes the cloning and nucleotide sequences of three heat-shock genes and their promoter region which constitute the *dnaK* locus, and a fourth potential ORF downstream of these genes.

Methods

Bacterial strains and media. The bacterial strains and plasmids used in this study are listed in Table 1. *E. coli* strains were grown aerobically at 37 °C in L Broth (Lennox, 1955) and *L. lactis* subsp. *lactis* MG1363 (hereafter *L. lactis*) was grown statically at 30 °C in M17 medium (Terzhagi & Sandine, 1975) supplemented with 0.5%

glucose. When necessary L-Broth was supplemented with ampicillin (100 µg ml⁻¹), isopropyl β -D-thiogalactopyranoside (IPTG, 100 µg ml⁻¹) and 5-bromo-4-chloro-3-indolyl β -D-galactopyranoside (X-Gal, 40 µg ml⁻¹).

PCR amplifications. Oligonucleotide primers for PCR were produced using an Applied Biosystems 392 DNA/RNA synthesizer (Table 2). Primers P9 and P10 were used to produce a 346 bp homologous DNA probe for the *dnaK* gene from *L. lactis* chromosomal DNA. PCR reactions were performed using either 1 µg of chromosomal DNA as template or by inoculation of colony cells in a 50 µl vol. of PCR mixture (10 mM-Tris/HCl, pH 8.3, 50 mM-KCl, 2 mM-MgCl₂, 0.1% gelatin, 200 µM of each dNTP and 0.1% Triton-X100) containing 100 ng of each primer and 0.5 U *Taq* polymerase (Promega). Samples were overlaid with 50 µl of mineral oil before being subjected to 25 cycles of 2 min denaturation at 92 °C, 2 min annealing at 47 °C and 2 min elongation at 72 °C. *L. lactis* chromosomal DNA was digested

completely with *EcoRI* and religated for use as the template for inverse PCR using the divergent primers P69 and P100. PCR-generated fragments were end-repaired using T4 DNA polymerase and T4 polynucleotide kinase before use in ligation reactions.

DNA isolation and cloning of the *dnaK* locus. Total DNA was extracted from overnight cultures of *E. coli* and *L. lactis* by the method of Lewington *et al.* (1987). Plasmid DNA was isolated from *E. coli* by the alkaline lysis method (Birnboim & Doly, 1979) and all DNA manipulations were performed as described by Maniatis *et al.* (1982). For construction of a genomic library, 100 µg of completely digested DNA was centrifuged in a sucrose density-gradient from 10–40% (w/v). The fractions were ethanol-precipitated, ligated to *HindIII* digested dephosphorylated pBR322 and transformed into either *E. coli* SURE (Stratagene) or *E. coli* JM109 (Table 1). All the resulting transformants were screened by colony hybridization with the 346 bp PCR probe.

Hybridizations. For Southern hybridizations, chromosomal DNA was digested to completion with the desired restriction enzymes and separated on 1% (w/v) agarose gels. DNA was transferred to a nylon membrane (Hybond-N, Amersham) as recommended by the manufacturer and hybridized with the 346 bp *L. lactis* PCR-generated fragment described earlier (P9/P10). The DNA probe was labelled with [α -³²P]ATP using the Multiprime DNA labelling kit (Amersham), in low melting point agarose. For colony hybridization, colonies were picked onto duplicate plates. Hybond-N nylon membranes were used for colony lifts and hybridization was performed as described by Grunstein & Wallis (1979).

Protein extraction. *E. coli* and *L. lactis* cultures were grown to an OD₅₅₀ of between 0.9 and 1.1, or were grown to an OD₅₅₀ of 0.6 before being subjected to 30 min heat shocks at 46 °C and 42 °C, respectively. *E. coli* protein was extracted exactly as described by Dean & James (1991). *L. lactis* cultures were harvested at 7000 r.p.m. for 10 min, resuspended in 1/4 strength Ringer's solution (37 mM-NaCl, 1 mM-KCl, 1 mM-CaCl₂, 2 mM-Na₂S₂O₃, pH 6.6) and the suspension added to an equivalent volume of 0.1 mm glass beads in a 2 ml screw-cap Eppendorf tube. Cells were disrupted by high-speed vortexing for 3 min using a Biospec BeadBeater and subsequently placed on ice for 5 min. The supernatant was cleared by centrifugation at 14000 r.p.m. for 5 min. Protein determinations were carried out by the method of Bradford (1976) using the Bio-Rad Protein Assay reagent and proteins were separated on a 12% (w/v) acrylamide gel (Bio-Rad Miniprotein II system).

Western immunoblot analysis. The SDS-PAGE gels were electroblotted for 22 min at 220 V onto Immobilon-P membranes (Millipore). The membrane filters were washed in TBS-T (50 mM-Tris/HCl, pH 7.4, 200 mM-NaCl, 0.1% Tween 20) to remove excess SDS and then blocked with 5% (w/v) skimmed milk (Marvel) in TBS-T for 30 min at 37 °C. The filters were subsequently incubated for 1 h at room temperature with 1% (w/v) Marvel in TBS-T containing a 1/10000 dilution of a polyclonal DnaK antibody (kindly provided by Richard James, University of East Anglia, Norwich, UK) and incubated for 1 h in TBS-T containing a 1/3000 dilution of goat anti-rabbit antibody labelled with horseradish peroxidase (Bio-Rad). Excess antibody in each case was removed by washing for 2 × 10 min in TBS-T. The immunoreactive bands were visualized on X-ray film by using ECL Western blot reagents (Amersham).

DNA sequencing. DNA was sequenced by the dideoxy chain-termination method (Sanger *et al.*, 1977) using deoxyadenosine 5'- α -[³⁵S]thiotriphosphate and a Sequenase T7 sequencing kit (USB). Sequencing of the *dnaK* locus was achieved using plasmids pFI574, pFI575, pFI576, pFI619 and pFI620. Several series of deletion derivatives were generated from each plasmid using the Erase-a-Base

system (Promega). Sequence determination across the *EcoRI* junctions was obtained using pFI573. Further upstream sequences were obtained from several different clones containing the 2.3 kb inverse PCR fragment (equivalent to pFI672).

Computer programs. The Wisconsin Genetics Computer Group sequence analysis software package, version 7 (University of Wisconsin) was used for sequence comparisons, analysis of non-coding regions and to deduce protein sequences. Analysis of the DNA coding regions was done using the DNA Strider program (Marck, 1988) on a Macintosh LCII computer.

Results

Detection of a 70 kDa dnaK homologue in L. lactis subsp. lactis

Western blot analysis of total protein extracts of *L. lactis* and *E. coli* under both normal growth and heat-shock conditions revealed a 70 kDa protein which cross-reacted with the DnaK antibody (Fig. 1). Based on densitometric studies the *L. lactis* protein appears to be induced approximately 3-fold on heat shock at 42 °C. Other cross-reacting proteins with molecular masses of approximately 50 kDa (*L. lactis*) and 13 kDa (*E. coli*) were apparent.

PCR amplification of the dnaK region and hybridization of L. lactis chromosomal DNA

The computer program Ialien (A. J. Bleasby, Daresbury Laboratory, Warrington, UK) was used to generate the maximum multiple alignment of *dnaK* genes from several bacterial species available in the EMBL and GenBank databases (releases 26.0 and 67.0, respectively), including *E. coli* and *Bacillus megaterium*. Within the N-termini of the proteins, two regions are particularly highly conserved. Primers P9 and P10 were designed from these regions based on the DNA sequence of *B. megaterium* (Sussman & Setlow, 1987), taking into account trends in lactococcal codon usage (Van de Guchte *et al.*, 1992). Of the resulting primers, P9 is degenerate, having 1024 possible variants. PCR amplification of the *L. lactis* and *E. coli* chromosomal DNA resulted in the production of 346 bp and 424 bp fragments, respectively. The *L. lactis* 346 bp fragment was cloned into pUC18 (pFI483) and its sequence determined. Sequence comparisons revealed high homology of the sequenced region with *dnaK* homologues, sharing most identity with the *dnaK* gene of *B. megaterium* (70%) at the DNA level. This PCR fragment was therefore used as a homologous probe in Southern hybridizations of *L. lactis* chromosomal DNA in order to isolate the complete *dnaK* gene.

Southern blots of *L. lactis* subsp. *lactis* chromosomal DNA cleaved with *HindIII* and *EcoRI* produced hybrid-

Fig. 1. Western blot analysis of extracts separated by SDS-PAGE (12% gels) from *L. lactis* subsp. *lactis* MG1363 and *E. coli* MC1022 with goat anti-rabbit antibody raised against polyclonal DnaK antibody. Lanes: A, *E. coli* heat shocked at 46 °C for 30 min; B, *E. coli* grown at 37 °C; C, *L. lactis* subsp. *lactis* heat shocked at 42 °C; D, *L. lactis* subsp. *lactis* grown at 30 °C. Molecular mass markers (kDa) are on the left.

izing bands of 5.4 kb and 3.5 kb. The larger 5.4 kb *Hind*III fragment was selected for cloning.

Cloning and sequencing of the *dnaK* gene region

A partial genomic library of *Hind*III chromosomal fragments ranging in size from 4.5–6 kb was constructed in pBR322 as described in Methods. Several attempts were made to clone the *Hind*III fragments into pUC plasmids. However, the resulting plasmids all appeared to be unstable and possessed large deletions (data not shown). Stable clones could only be obtained and maintained in the lower-copy-number vector pBR322 and by using *rec*⁻ *E. coli* hosts. Screening of 950 individual transformants by colony hybridization revealed one clone that gave a clear hybridization signal (pFI573). Sequencing of the 5.4 kb insert showed the presence of both a *dnaK* and a *grpE* homologue (see below). The sequence showed a considerable degree of homology to the corresponding upstream sequences found in *Bacillus subtilis* (Wetzstein *et al.*, 1992) and *Clostridium acetobutylicum* (Narberhaus *et al.*, 1992) and no potential promoter sequences were found in the cloned insert upstream of the *grpE* gene. A 5'-truncated ORF was identified upstream of the *GrpE* homologue. Cloning of sequences further upstream was achieved by inverse PCR as described in Methods. A 2.3 kb PCR fragment was generated from the *Eco*RI-digested

chromosomal DNA which corresponds to a template size of 3.5 kb. The 2.3 kb fragment was cloned into the *Eco*RV site of pBR322 to produce pFI672 and its sequence determined.

Identification of the genes of the *dnaK* gene region of *L. lactis*

The complete nucleotide sequence of the *dnaK* gene region of *L. lactis* is shown in Fig. 2. A total of 5915 bp was sequenced and analysed using the computer program Codonpreference with lactococcal codon usage tables (Van De Guchte *et al.*, 1992) revealing the presence of four potential ORFs. Low values of third-position GC bias were obtained for the four ORFs. *L. lactis* has a very low GC content ranging from 35.5–38.6% (Hardie, 1986) and the predicted ORFs are therefore likely to be expressed efficiently in *Lactococcus*. Three of the ORFs appear to be contiguous and are followed by a potential transcription terminator, a vegetative promoter sequence and a fourth large ORF (Fig. 3).

ORF 1 is 1044 bp long (Fig. 2, nucleotides 247–1290) encoding a putative polypeptide of 347 amino acids with a predicted molecular mass of 38.9 kDa. Homology searches identified sequence homology with the *orfA* and *orf39* genes found in *C. acetobutylicum* and *B. subtilis*, respectively, and has been designated *orf1*.

ORF 2 is 540 bp long (nucleotides 1320–1859). The start codon is 32 nucleotides from the last stop codon of ORF 1. ORF 2 encodes a putative polypeptide of 179 amino acids with a predicted molecular mass of 20.6 kDa. This value is comparable with the 94 kDa cross-reacting protein (tetramer) detected with anti-GrpE antiserum by Auffray *et al.* (1992).

ORF 3 is 1824 nucleotides long (nucleotides 1931–3754) and is 74 nucleotides from the last stop codon of ORF 2. ORF 3 encodes a putative protein of 607 amino acids with a predicted molecular mass of 64.9 kDa, which is slightly smaller than the 70 kDa cross-reacting DnaK protein observed in this study and by Boutibonnes *et al.* (1992). ORF 2 and ORF 3 show high homology (see Table 3 and below) to several other *grpE* and *dnaK* homologues and are named accordingly.

ORF 4 is 1845 bp long (nucleotides 3918–5762) encoding a putative polypeptide of 614 amino acids with a predicted molecular mass of 69.8 kDa. The deduced sequence shows no extensive homology with any genes found in the GenBank or EMBL databases (release 75.0 and 34.0, respectively) although there is a low degree of homology with various membrane-associated proteins. As yet, no specific function has been assigned to this gene and it has been designated *orf4*.

The ORFs described follow the pattern of codon usage described for *Lactococcus* (Van de Guchte, 1992). All

1 CAACATCAAGTTGGATTGAAAACTTTTATTATGCTGTAATGGTAGTTACACTAATGAATATCTGGCAGACATCCAAAGATTCTAATCTTTTATAAC
 -----> <----- -35 -10 ----->
 101 AGAAATTAGCACTCTAATAAAAGAGTGTCTAATTTTATGTCTTTTCTTTGACAAGGTATAAACAAAGGGTTATAATAAACTTGTAGCACTTAGAGGT
 <-----
 • *****
 201 AAAGAGTGTCTAAAGAAAAAAGCTTAGAGATAAAAAAGGAGCAGCTATGATTACAGAACGTCAAAGACAAATTTAAATTTGATTGTTTCACTTTATGCT
 M I T E R Q R Q I L N L I V S L Y A
 orf1
 301 AAGGACCATACACCAATCGGTTCAAAATCGCTCCTTGATAGTATCCAAGCCTCTAGCGCAACCATTCGAAATGATATGAAAGCTTTGGAAAGGCTGGGTT
 K D H T P I G S K S L L D S I Q A S S A T I R N D M K A L E R L G L
 401 TGATTCAAAGGAACATACCTCAAGTGGACGATTCCGAGCGTTTCGGGCTATAAGTATTTTGTGGAAATGTCATCCAGTTAGAAGAGTTTCAGTCAAAA
 I Q K E H T S S G R I P S V S G Y K Y F V E N V I Q L E E F S Q N
 501 TGACTTGTTCAAAGTGATGAAAGCTTTTGATGGCGATTTTATCGCTTGTCGGATTTGTTTAAACTGCGGCTAAAAGCTTATCAGAATAACAGGATTG
 D L F K V M K A F D G D F Y R L S D L F K T A A K S L S E L T G L
 601 ACGAGTTTGTCTTAATGCTCCGCAACGTGACCAACAATTAGTCTCATTGAAATGGTTATGCTTGATAATCACTCGGTGCTCTCGGTCATTACTTTGG
 T S F V L N A P Q R D Q Q L V S F E M V M L D N H S V L S V I T L G
 701 GAACTGGTGAGGTAAAGACAAATCAATTTATTTTACCAAAATCAATGACAGAAGCAGATTTAGCTGTTTTAGCAACTTGGTTAAGGAACGATTAGTCGG
 T G E V R T N Q F I L P K S M T E A D L A V F S N L V K E R L V G
 801 TAAAAAGTGATTGATATCACTATACGTTACGCACTGAAATCCACAATTTGTTCAACGCTATTTTAAGGTTACGAGTGAAGTTTACAACCTTTTGAA
 K K V I D I H Y T L R T E I P Q I V Q R Y F K V T S E V L Q L F E
 901 TCAATCTTTGATGATTATTCAAAGAATCTGACAGTCGCTGGTCATAAAAAATTTTTTGACTATGCGACGGATAATCTTGCAGAGCTCTATAAACTCT
 S I F D D L F K E H L T V A G H K N I F D Y A T D N L A E L Y K L F
 1001 TTAGTGATGATGAGCGAATGCTTCATGAAATTCGCGAGATTACAAATAATGATGAAATGAGAGCTGTTAAATTTGATAATGATGAAAAATTTATGAAAA
 S D D E R M L H E I R E I T N N D E M R A V K F D N D E K F M K N
 1101 TTAAACAATTATCTCGCAAAAATTTGTCATTCCTTATCGTGGATTGGTGACATTGACAGTTGTAGGACCTGTAGAAATGGATTATCAAGAACCTTTATCA
 L T I I S Q K F V I P Y R G F G T L T V V G P V E M D Y Q R T L S
 1201 GTTCTTGATTAGTTGCTAAAGTTCTAACAATGAAATTAAGTGATTATTATCGTTACCTTGATGGTAACCATTATGAAATTAGTAAATGATTAAAAAGTT
 V L D L V A K V L T M K L S D Y Y R Y L D G N H Y E I S K *
 * *****
 1301 ATTTAAAAAGGAGGGCCTATGTCCGAAGAACTAAAGAAGAAATCAAAAACGAAAAAGTTGATGAAGAAGTGACTGAGGAGTTAACAGAAGAAGCGCTT
 M S E E T K E E I K N E K V D E E V T E E L T E E A L
 grpE
 1401 GAAGATATCGTTGAAGAAGAAATCAATGAACCTTGATGAAGCGCAAAAGTTAGCAACTGAGTGGGAAAAATAAATTCCTTCGTGTTTCTGCTGAAATGCAAA
 E D I V E E E I N E L D E A Q K L A T E W E N K F L R V S A E M Q N
 1501 ATGTGCAACGTCGCGAAATGAAGAACGTTACAACCTGTCAAGTATCGCTCGCAAGATTTAGCAAAAAAATCTTGTCTTCATTAGATAATCTTGAACG
 V Q R R G N E E R L Q L V K Y R S Q D L A K K I L S S L D N L E R
 1601 TGCCTTGTCTGTTGAAGTTTGACAGATGATGTTAAAAAGGGCTTGAATGGTACAAGAAAGTTTAATTTAGCCCTTAAAGAAGAAGGGGTTGAAGAA
 A L A V E G L T D D V K K G L E M V Q E S L I S A L K E E G V E E
 1701 GTTCTTATGAAAGTTTCGACCATAACCTTCACATGGCTGTTCAAATGTTCCAGCTGATGACGAACATCCAGCTGACAGCATTGTGCAAGTCTTCCAAA
 V S Y E S F D H N L H M A V Q T V P A D D E H P A D S I V Q V F Q K
 1801 AaGGCTACCAACTTCACGAACGTTTACTACGTCTGCAATGGTGTGTAGTACACAATAATTAAATAAAAAATATTGACCGCAAAGTCTTTAACTATTCA
 G Y Q L H E R L L R P A M V V V A Q *
 • ***** *
 1901 TAAGTAAAGAAATAAAGGAGATTAAAAATATGCTAAAATATCGGTATTGACCTTGGTACAACCTAAGTACAGCAGTCGAGTCTTGAAGGAACTGAAT
 M S K I I G I D L G T T N S A V A V L E G T E S
 dnaK
 2001 CAAAAATTTATCCAAATCCAGAAGGAAATCGCACAACTCCATCAGTAGTGCATTAAAAATGGTGAAATCATTGTTGGTGATGCTGCAAAACGTCAAGC
 K I I P N P E G N R T T P S V V A F K N G E I I V G D A A K R Q A
 2101 AGTAACAAATCCAGAAACAATCATTTCTATCAAAATCAAAATGGGAACATCTGAAAAAGTTTCTGCAAAATGGTAAAGAATATACACCACAAGAAATTTCA
 V T N P E T I I S I K S K M G T S E K V S A N G K E Y T P Q E I S
 2201 GCAATGATCTTCAAACCTGAAAGCAACTGCTGAAAGCTACCTTGGCGAAAAAGTTGAAAAAGCAGTTATCACTGTTCTGCTTACTTCAATGACGCTC
 A M I L Q N L K A T A E S Y L G E K V E K A V I T V P A Y F N D A Q
 2301 AACGCTAAGCGACTAAAGATGCTGGTAAAATGCTGGTCTTGAAGTAGAACGTATCGTTAACGAACCAACTGCCGAGCTCTGCTTATGGTCTTGATAA
 R Q A T K D A G K I A G L E V E R I V N E P T A A A L A Y G L D K

2401 AACAGATAAAGATGAAAAATCCTTGTCTTTGACCTTGGTGGTGGGACATTCGACGTTTCAATCCTTGAACCTGGCGACGGTGTCTTTGATGTTCTTGCA
 T D K D E K I L V F D L G G G T F D V S I L E L G D G V F D V L A
 2501 ACAGCAGGTAACAACAACTTGGTGGGACGATTTTGACCAAAAAATCATTGACTGGATGGTTGCCGAATCAAGAAAGAAAACGGAATCGACCTTGGTC
 T A G N N K L G G D D F D Q K I I D W M V A E F K K E N G I D L G Q
 2601 AAGACAAAATGGCATTGCAACGTTTGAAGATGCCGCTGAAAAAGCGAAAAAGACCTTTGAGGTGTAACAACAACCTCAAAATCAGCTTACCATTTCATCAC
 D K M A L Q R L K D A A E K A K K D L S G V T T T Q I S L P F I T
 2701 AGCAGGAGCTGCCGGACCTCTTCACTTGGAAATGGCTTTGACTCGTGCCAAATTTGACGAATTGACACATGATCTTGTGTAAGCGACACGTCAACCAAGTT
 A G A A G P L H L E M A L T R A K F D E L T H D L V E A T R Q P V
 2801 CGTCAAGCCCTTTCTGATGTGGTTTGTCAACTTCTGATATTGATGAAGTCTCTTAGTTGGTGGTTCAACTCGTATTCTGTGTGTGTTGAACCTTGTTC
 R Q A L S D A G L S T S D I D E V L L V G G S T R I P A V V E L V R
 2901 GTCATGAAACAAACAAAGAACCAATAAATCAGTAAACCTGATGAAGTTGTTGCTATGGGTGCCGAATTCAAGGTGGTGTCACTGGTGACGTTAA
 H E T N K E P N K S V N P D E V V A M G A A I Q G G V I T G D V K
 3001 AGACGTTGTCCTTCTGATGTTACTCCATTGTCACTTGGTATTGAAACAATGGGTGGAGTCTTCACAAAATTGATTGACCGTAACACAACATATCCCAACT
 D V V L L D V T P L S L G I E T M G G V F T K L I D R N T T I P T
 3101 TCTAAATCACAAGTCTTCTCAACTGCCGCTGATAATCAACAGCTGTTGACATTCACGTCCTCCAAGGAGAACGTCCAATGGCCGCTGATAACAAAACCTT
 S K S Q V F S T A A D N Q P A V D I H V L Q G E R P M A A D N K T L
 3201 TGGGTCGTTTCCAATTGACAGATATCCCTGCAGCACCAGTGGTATCCCAACAATCGAAGTAACCTTTGATATTGATAAAAATGGTATTGTTCTGTGTTAA
 G R F Q L T D I P A A P R G I P Q I E V T F D I D K N G I V S V K
 3301 AGCTAAAGACCTTGGCACACAAAAGAACAACAATGTTATCAAAATCAAACTCAGGTCTTCTGATGAAGAGATTGATAAAATGATGAAAGACGCCGAA
 A K D L G T Q K E Q T I V I K S N S G L S D E E I D K M M K D A E
 3401 GCAAAATGCTGACGCTGACGCAAAACGTAAGAAGAAGTTGATACACGTAACGAAGCTGATGCTTTAGTTTTCCAAACTGAAAAAATTTGAAAGACCTTG
 A N A D A D A K R K E E V D T R N E A D A L V F Q T E K T L K D L E
 3501 AAGGAAAAGTTGAAGAAGCTGAAGTTAAAAAGCCGAAGACGCTAAAGAAGAACTCAAAAAAGCGCTTGAAGGTGAAGATATTGATGATATCAAGCTAA
 G K V E E A E V K K A E D A K E E L K K A L E G E D I D D I K A K
 3601 ATCAGAAGCACTTTCTGAAATCGCTCAAAACCTTGCTGTAACCTTTATGAACAAGCAACGCAGCTCAAGGAGAAGCGAGTGAAGCAACAGATGCTCAA
 S E A L S E I A Q N L A V K L Y E Q A N A A Q G E A S E A T D A Q
 3701 GAAGGACCAAAAGATGCTAATACATTTGATGGCGACTTTGAAGAAAGCAAAATATTTTAAAAAGTACTGATCGTTTCGTCAGTACTTTTTTTATGTT
 E G P K D A N T F D G D F E E S K *
 -35 -10
 3801 TAAAAGGATTCACTGACAAGCTGTCAAGTGGATTTTTGTTTATTTGACATAGGGACGAATTGAGTAATATGATATAATTGGACAAGATTAGGATTAAAT
 • *****
 3901 TTTAGGAGAAATAATAGATGGAAAGATTAGAAAAAGAGGGCAAGGATTAGAGAATTTCAAAAAATACAATTTGAATTTTATATCATTATTCTTTTAT
 M E R L E K R G Q G L E N F K K Y N F E F Y I I I L F I
 orf4
 4001 TACTATATTGGCAACATTCCTTTTATAGAGGTAATTTTCATGCTGGAATGATTTTGCAATTTAATATGCAAGGGTGTGTCACAATTAGTGCAATTA
 T I L A T F P F Y R G N F H A G N D F A F N Y A R V M S T I S A L
 4101 AAAGATGGACAAGTAATCCGCAATTTGATCCAAATGCTTTAAGTGGTTTGGATATGCCTGGAATGAATTTTATGGTCCACTTCCAACCTATTTTATTT
 K D G Q V I P Q F D P N A L S G F G Y A W N E F Y G P L P T Y F I S
 4201 CTGTCATTAATTTATGTCAGTCATGGTCACTATCCTTTTCACTTTTATAGTTTATGCTTTTATATCAGGTATTTTATTTAATTTTTCGAG
 V I K F I V K S W S L S F S L F Y S L C L F I S G I F I F N F S S
 4301 TTTTCTATTAAAGACCATACTAATAGTAAATTTTGGCTTATGGCAGTTGCCCTTTTACATTTTCGAATTCACATATATCAATCTTTATTTATTT
 F L L K D H T N S K L F G L L A V A L F T F S N S T Y I N L Y Y Y
 4401 GCCAACCTTAGCCAACTCTTGCCTTTTATTTGTAATCCTCTGTTTGGGGAATGAACAAAATGTTTAATAAACGGAGCTTTGCTGCCTTTTAAATGG
 A N P S Q P L A L L F V I L L F W G M N K M F N K R S F A A F L M V
 4501 TAGCTTTTGGGGCAGCAGGCTTACCTCTTTCACATACAGTGACCACGATTTGACTCTTCCCTTGTATTATTATATCTCCTCTTTTGGATAAAAAA
 A F G A A G L P L S H T V T T I C T L P F V L L Y L L F L I I K K
 4601 AGGGAATTTAAAGAAAATATAAAAAATCATTGGTTTGGGATTCTGTCACTGACTTCGGCAATTGGGCTGTCAGCATTCTTCTTATTTCTTTTATTGGAG
 G N L K E N I K I I G L G F L S V T S A I G L S A F F L F P L L E
 4701 AATTTGAAATCTGAATCTATAACGTTTCTAATTCAGATTTTCTCGTAGTTTGGTTGGAATAATATTGCTTATTTTCAAGGGAATGGGAGCCACTTT
 N L K S G I Y N V S N S D F S R S F G W N N I A Y F Q G K W E P L Y

4801 ATAAATGAATTTTCATATTATAAATTCCTTCTCTCTATTGTTTTATTGTACTATTATTTTATTTTCTCTGATAAATTTCAAGAAACTAA
 K I E F S Y Y K F P S L L F V F I V L F I F I F S L I N F K K T N

4901 TGCCAAGTAGCTGATTTTCTCCTGTTTTTCATTAGTCTGGTTTTGATGCAACTTCCGATATTTCTTGGAACTATTTTCTATTTTACGATTGTT
 A K Y S L I F S C F S L V L V L M Q L P I F P W K L F S I F T I V

5001 CAAGATCCAGCAAGATTTTCAACTCTTTTGGCCTATTTCTGCTCTCAGCCTCGTATTAATATTGCCATCTTATTGGACAAAATTTCAAGAAAACTT
 Q D P A R F S T L F G L F S A L S L V L I L P I L L D K I S G K T S

5101 CTTATTATCTGACAATCGGTTTATTAGTCATTTTCAATTCTAGGATTGTCTGAATTTAGAAATCGAATCCAAAAGGTTCTCAACCATTATTGCTTC
 Y Y L T I G L L V I F S I L G F A E F R N R I Q K G S Q P L F A S

5201 TGCTCAAAGTTTATTAAATAAACTCCCTTTAATTATATGGAACCCGGATTCTATTGCCATTGGTGAATATTACCCCAAGTTATTGGAAGCCATAAT
 A Q S L L N K T P F N Y M E N P D S I A I G E Y L P Q V I G S H N

5301 CAACCTTATGAAAAGACTATTCAACAATTTTATAAGATAAGAATGTCTATGGAATGCGCAATCAAGCGATGACATATTGAGTCAAGAGGAAAACTCC
 Q P Y E K T I Q Q F Y K D K N V Y G M R N Q A M T Y L S Q R G K L P

5401 CAGAGGGGCTTGCAAAAAGCATTGAGATATCAGACTATAGTAAAAAGGCTCACATGTCACTTTTACAGCGACTACTGGTTCAAAAAGTCTAGTGTGA
 E G L A K S I Q I S D Y S K K G S H V T F T A T T G S K S A S V E

5501 AATACCAGAAATTTATTACAAAGGATTTATCGCCTTTACGAAAGAGGAAATAAAAAATAAGCTCACAAGCCAAATTTCTAAAAATGGATTCTAGAA
 I P E I Y Y K G F I A F T K E G N K K I K L T S Q I S K N G F L E

5601 ATAAAGTCCCCGAAAGTTTTCGGGAAATTTATTCTACTTTCGCAATGAGTTCAGCAACAAATATGGCGTACTTTTATCACTTTTGACATTTGTGA
 I K V P A K F S G K I Y S Y F A M S S A T K Y G V L L S L L T F V S

5701 GCCTGCTGATAATTTCTTATCACTAAAACCTTCATAAAAAGGAAATATCCAAAGATAAATAAAGAAAATGGACAAAATGGACAAGGTGTCCAAGAAAC
 L L I I L I T K T F I K R K L S K D K *

5801 GTTTCGGATTCTTGCAATCGTTTACTCTGTACTATGATGAAATTAGTAAGAGATTGGAGGAGAATTAAAAAGAAAGGCTAAAATTTTAGCCGAAATA

5901 TGTGTTTAGAATTC 5915

Fig. 2. DNA sequence of *orf1*, *grpE*, *dnaK* and *orf4* of the *dnaK* operon and the deduced amino acid sequences. Indicated are the putative Shine–Dalgarno sequences (asterisks above sequence), -35 and -10 promoter sequences (underlined), a potential stem-loop (nucleotides 103–135 and 186–212; converging dashed arrows) and a possible transcription terminator (nucleotides 3761–3794; solid converging arrows).

begin with an ATG start codon and end with a TAA stop codon (except *orf1* which has a TGA stop codon) which is the most commonly used lactococcal stop codon. The number of rare codons present in the *grpE* and *dnaK* genes is very low (two and three, respectively). The *grpE* gene contains one TCC and one TCG serine codon (amino acids 2 and 78 respectively) which occur at a frequency of only 6% in *L. lactis*. Three CTC leucine codons are present in the *dnaK* gene (amino acids 308, 411 and 542) which occur at a frequency of 8%. *orf1* and *orf4* contain considerably more rare codons; 11 and 29, respectively. This presumably reflects the fact that *grpE* and *dnaK* heat-shock genes are among the most highly conserved proteins in nature, common to all organisms.

Features of the non-coding regions

Putative ribosome-binding sites were identified in front of all four genes (Fig. 2) which show varying degrees of complementarity to the 16S rRNA sequence of *L. lactis* (3'-UCUUUCCUCCA-5'; Ludwig *et al.*, 1985). A promoter region identical to the consensus Gram-positive promoter sequence (Graves & Rabinowitz, 1986) is found upstream of *orf1* (nucleotides 150–178) with a -35 sequence, 5'-TTGACA-3', and -10 sequence, 5'-

TATAAT-3', separated by 17 bp. Two stem-loop structures flank this putative promoter. An inverted repeat situated 5' to the promoter has the sequence AAATTAGCACTC-N₉-GAGTGCTAATTT and a free energy of -64.26 kJ mol⁻¹ [calculated using the Fold computer program (Zuker & Stiegler, 1981) as determined by Freier *et al.*, 1986].

Between the promoter and the ATG start codon is a second inverted repeat with a free energy of -37.8 kJ mol⁻¹ which has the arrangement TTAGCACT-N₁₁-AGTGCTAA. A putative rho-independent transcription terminator (Platt, 1986) is found in the 163 bp region separating ORF 3 and ORF 4 (Fig. 2, nucleotides 3761–3794). This inverted repeat has a free energy of -58.8 kJ mol⁻¹ and is followed by a stretch of several Ts and As. *orf4* is preceded by a vegetative promoter sequence with a -35 sequence, 5'-TTGACA-3', and three possible -10 regions with the sequences 5'-TAATAT, TATGAT and TATAAT-3' which are separated from the -35 region by 15, 18 and 23 bp, respectively. Lactococcal promoter sequences are commonly separated by slightly more than the 17 bp found in *E. coli* and it is likely that either of the latter two -10 regions are the genuine promoter sequences (shown in Fig. 2).

Fig. 3. Restriction map of the *L. lactis* subsp. *lactis* *dnaK* operon, plasmids, location of the PCR primers and putative transcription terminator (T). Inserts of various plasmids described in the text are shown. P9 and P10 were used to produce the homologous probe used to screen the pBR322 library. P69 and P100 were used to isolate upstream promoter sequences (P) by inverse PCR.

Table 3. Homology of the *DnaK* and *GrpE* proteins of *L. lactis* subsp. *lactis* to the corresponding proteins of other bacteria

Organism	Percentage identity (percentage similarity) of the following proteins:		
	Orf1	GrpE	DnaK
<i>E. coli</i>	—	29.0% (53.6%)	57.0% (74.2%)
<i>B. subtilis</i>	27.2% (52.5%)	43.3% (63.5%)	77.0% (87.3%)
<i>B. megaterium</i>	—	—	77.4% (87.9%)
<i>C. acetobutylicum</i>	24.6% (56.2%)	38.6% (56.6%)	66.1% (78.9%)

Sequence similarities between *DnaK*, *GrpE* and the *Orf1* protein with other bacterial species

The deduced amino acid sequences of the four ORFs were aligned to sequences available in the EMBL and GenBank databases. The amino acid sequence of the *L. lactis* *DnaK* protein shows high sequence similarity to the *DnaK* proteins of other bacterial species (Table 3). Although the actual identity with the *E. coli* amino acid sequence is 57.0%, conservative amino acid changes increase the similarity value to 74.2%. Very high identity (77.0%; 87.9% similarity) is observed with the *B. megaterium* protein.

Considerably lower homology is observed between the *L. lactis* *grpE* amino acid sequence and other bacterial species. Although identity with other proteins ranges from 29.0% (*E. coli*) to 43.3% (*B. subtilis*), inclusion of conservative amino acid changes increases the range of similarity to 53.6% (*E. coli*) and 63.5% (*B. subtilis*), respectively. Another start point for transcription of the *L. lactis* *grpE* gene is possible (Fig. 2, nucleotide 1287). Although this would increase the deduced protein size

from 179 to 190 amino acids, which is more similar to the other bacterial *GrpE* proteins, a suitable ribosome-binding site could not be identified in the immediate upstream region.

The amino acid sequence deduced from *orf1* is the third example in Gram-positive bacteria of a heat-inducible gene found upstream of the *grpE* and *dnaK* genes as part of the same operon. Similarly sized ORFs are found in both *C. acetobutylicum* (*orfa*; Narberhaus *et al.*, 1992) and *B. subtilis* (*orf39*; Wetzstein *et al.*, 1992) which show 31.6% identity and 59.5% similarity with conserved amino acid changes. The computer programs Pileup and Prettybox were used to produce a boxed sequence alignment (Fig. 4) of these three proteins. The degree of similarity is consistently higher at the N-terminus and declines towards the C-terminus. Within the N-terminus two consensus sequences are evident; SSATIRN and SSGRXPS.

Discussion

This report describes the first complete heat-shock operon within lactococcal species. The *dnaK* gene region of *L. lactis* subsp. *lactis* contains at least three heat-shock genes with the arrangement *orf1*-*grpE*-*dnaK*, followed by a fourth large ORF. This chromosomal organization is identical to that found in two other Gram-positive bacteria *B. subtilis* and *C. acetobutylicum*, except that in these cases the *dnaJ* gene is found downstream of *dnaK*. The *Orf1* protein has homology with the *Orfa* (*C. acetobutylicum*) and *Orf39* (*B. subtilis*) proteins, particularly within the N-terminus. The amino acid sequences SSATIRN and SSGRXPS are highly conserved in the N-termini of these proteins suggesting some functional role of these regions. As *orf1* is under the same regulatory signals as the other *dnaK* region genes, its gene product is assumed to be part of the heat-shock response. No homologous gene could be detected in *E.*

Orfa	M E M E E R K L K I	L Q A I I N D Y I N	N G E P V G S R T I	A K K Y N I G I S S	A T I R N	45
Orf39	. M I T N R Q L L I	L Q V I I N D I K	S A Q P V G S R T L	S K K D E I T F S S	A T I R N	44
Orf1	. M I T E R Q R Q I	L N L I V S L Y A K	D H T P F G S K S L	L . . D S I Q A S S	A T I R N	42
Orfa	E M A D L E E M G Y	I E Q L H T S S G R	K P S D K G Y R L Y	V D R L M E I P S M	S V E E E	90
Orf39	E M A D L E E L G F	I E K T H S S S G R	V P S E K G Y R Y Y	V D H L L S P V K L	T K S D L	89
Orf1	D M K A L E R L G L	I Q K E H T S S G R	I P S V S G Y K Y F	V E N V I Q L E E F	S Q N D L	87
Orfa	M L I K A K I I D S	A L Y E I D K L V K	Q A M S L V S E M T	K L T C V V K S L S	A R K S Y	135
Orf39	D Q I H S . I F K E	K I E E L E K T V Q	K S A Q I L S D L T	N Y T S I V L G P K	L S E N Y	133
Orf1	F K V M . K A F D G	D F Y R L S D L F K	T A A K S L S E L T	G L T S F V L N A P	Q R D Q Q	131
Orfa	K S I S L I N I E	P N M I L C V F I T	D S G M K N S I S	R V K S N I E N S S	L E R I A	180
Orf39	L K Q I Q I I P I Q	P D M A V A L V T	N T G H V E N K T I	N F P T K M D L S D	L E K L V	178
Orf1	L V S F E M V M L D	N H S V L S V I T L	G T G E V R T N Q F	I L P K S M T E A D	L A V F S	176
Orfa	N I L N S K L K G L	T I E Q I N L E V I	N I K K D L R E Y	G H I F D C I M P N	L Y D I F	225
Orf39	N I L N D R L S G V	P M D E L N E R I F	K E V V M Y L R Q H	I K N Y D N I L D A	L R S T F	223
Orf1	N L V K E R L V G K	K V I D I H Y T L R	T E I P Q I V Q R Y	F K V T S E V L Q L	E S I F	221
Orfa	R E A D S T E . V Y	K E G T M N I F N Y	P E F K D I E K A K	E F L S V I D . D R	R I L D T	268
Orf39	H S T N H V E K L F	F G G K I N M L N Q	P E F H D I T R V R	S L L S L I E K E Q	D V L K L	268
Orf1	D D L . F K E H L T	V A G H K N I F D Y	A T . D N L A E L Y	K L E S D D E R M L	H E I R E	264
Orfa	L F N A S G G V T V	N I G N E N S I K E	A R D F R V V S S V	Y K Y N G R P L G T	I G I G	313
Orf39	V Q S P H T G I S T	K I G K E N D Y E E	M E N C S L I T A S	Y S V D Q K Q I G S	I A I I G	313
Orf1	I T N N D E M R A V	K F D N D E K F . .	M K N L T I I S Q K	V I P Y R G F G T	L T V V G	307
Orfa	P T R I P Y S K V I	K V I M E V V D Q E	N N N L D K M N N S	*	.343	
Orf39	P T R M N Y S R V V	S L L Q H V T S D L	S K A L T S L Y D E	*	.343	
Orf1	P V E M D Y Q R T L	S V L D L V A K V L	T M K L S D Y Y R Y	L D G N H Y E I S K	*347	

Fig. 4. Comparison of the deduced amino acid sequences encoded by *orf1*, *orfa* and *orf39* from *L. lactis*, *C. acetobutylicum* and *B. subtilis*, respectively. Amino acids are given in the standard one-letter code. Identical amino acids in two or three of the aligned sequences are boxed in black, conserved amino acids in grey (darkest grey indicating highest similarity). Gaps indicated by dots are introduced in order to obtain maximum alignment.

coli (Wetzstein *et al.*, 1992) and sequence data is not yet available to determine whether a homologous gene is present in other Gram-positive species.

The GrpE protein of *L. lactis* shows lower overall homology than the DnaK proteins (Table 3). The identity is highest with the Gram-positive bacteria. The alignment of the three proteins reveals two regions of increased homology (amino acids 50–96 and 137–178) as described previously for the GrpE protein of *B. subtilis* (Wetzstein *et al.*, 1992). It has been suggested that these regions could be involved in the functional interactions that occur (Johnson *et al.*, 1989) between GrpE and the DnaK and DnaJ proteins (Wetzstein *et al.*, 1992). As stated previously, the *L. lactis* *grpE* gene has two possible translational start points. The upstream translational start point overlaps with the translational stop codon of *orf1*. Despite the fact that this ORF lacks a credible ribosome-binding site, translational coupling could enhance expression from this potential initiation site. Although the indicated start point for translation (Fig. 2, nucleotide 1320) results in a smaller protein, it has a suitable ribosome-binding site and the N-terminus of the protein (amino acids 1–4; MSEE) is identical to that of the *B. subtilis* GrpE protein.

The DnaK protein shows very high homology with the DnaK proteins of other bacterial species which declines towards the C-terminus. There is a characteristic 24-amino-acid deletion near the N-terminus which appears to be a distinguishing feature of the Gram-positive DnaK proteins (Wetzstein *et al.*, 1992). This accounts for the smaller predicted molecular mass of the *L. lactis* DnaK protein and the difference in size of the PCR fragments amplified from lactococcal and *E. coli* chromosomal DNA by primers P9 and P10.

In many bacteria the *grpE*, *dnaK* and *dnaJ* genes are found as part of the same operon (Table 4), with the exception of *E. coli* and *Brucella ovis* where the *grpE* gene is found at a different chromosomal location. The *L. lactis* *dnaK* operon is unusual in that the *dnaJ* gene is not part of it and forms a completely separate transcriptional unit having its own promoter (Van Asseldonk *et al.*, 1993) immediately upstream. In *L. lactis* the ORF downstream of *dnaK* (*orf4*) appears to form a separate transcriptional unit, being separated from the *dnaK* gene by a potential transcription terminator and its own vegetative promoter. In *Bacillus* and *Clostridium* a transcription terminator is present immediately downstream of the *dnaK* gene and transcripts are initiated that

Table 4. *Regulatory sequence and gene organization of different bacterial species*

+, presence of; -, absence of; ND, not determined; v, vegetative promoter sequence; hs, heat-shock promoter sequence; T, transcriptional terminator; ORF, single ORF encoding the described protein; ?, no extensive homology to vegetative or heat-shock promoter consensus.

Organism	Gene or protein	Inverted repeat	Promoter		Gene organization	Source/reference
			- 35	- 10		
Gram-positive						
<i>L. lactis</i>	<i>orf1</i>	+	v	v	<i>orf1-grpE-dnaK-T-orf4</i>	This study
<i>L. lactis</i>	<i>dnaJ</i>	+	v	v	<i>-dnaJ-</i>	Van Asseldonk <i>et al.</i> (1993)
<i>B. subtilis</i>	<i>orf39</i>	+	v	v	<i>orf39-grpE-dnaK-T-dnaJ</i>	Wetzstein <i>et al.</i> (1992); Hearne & Ellar (1989)
<i>B. subtilis</i>	<i>groESL</i>	+	v	v	<i>groES-groEL</i>	Li & Wong (1992); Schmidt <i>et al.</i> (1992)
<i>C. acetobutylicum</i>	<i>orfa</i>	+	v	v	<i>orfa-grpE-dnaK-T-dnaK</i>	Narberhaus <i>et al.</i> (1992)
<i>C. acetobutylicum</i>	<i>groESL</i>	+	v	v	<i>groES-groEL</i>	Narberhaus & Bahl (1992)
<i>M. tuberculosis</i>	10 kDa antigen	+	v/hs	v/hs	ORF	Baird <i>et al.</i> (1989)
<i>M. tuberculosis</i>	65 kDa antigen	+	v/hs	v/hs	ORF	Shinnick (1987)
<i>M. tuberculosis</i>	<i>dnaK</i>	—	hs	hs	<i>dnaK-grpT-T-dnaJ</i>	EMBL database
<i>Streptomyces coelicolor</i>	<i>dnaK</i>	ND	ND	ND	<i>dnaK-grpE</i>	EMBL database
<i>B. megaterium</i>	<i>dnaK</i>	ND	ND	ND	<i>-dnaK-</i>	Sussman & Setlow (1987)
<i>Erysipelothrix rhusiopathiae</i>	<i>dnaK</i>	ND	ND	ND	<i>-dnaK-</i>	EMBL database
Gram-negative						
<i>E. coli</i>	<i>dnaK</i>	—	hs	hs	<i>dnaK-T-dnaJ-T</i>	Bardwell & Craig (1984); Bardwell <i>et al.</i> (1986)
<i>Brucella ovis</i>	<i>dnaK</i>	—	hs	hs	<i>dnaK-T-dnaJ-T</i>	Cellier <i>et al.</i> (1992)
<i>Borrelia burgdorferi</i>	<i>grpE</i>	ND	ND	ND	<i>grpE-dnaK-dnaJ-T</i>	Tilly <i>et al.</i> (1993)
<i>Synechococcus</i> sp.	<i>groESL</i>	+	hs	hs	<i>groES-groEL</i>	Webb <i>et al.</i> (1990)
<i>Synechococcus</i> sp.	<i>dnaK</i>	ND	?	hs	<i>orf70.5-dnaK-T-orf70.3</i>	Chitnis & Nelson (1991)
<i>Synechocystis</i> sp.	<i>cps-60</i>	+	?	hs	<i>orf60.5a-orf60.5-cpn60-orf60.3</i>	Chitnis & Nelson (1991)

terminate after *dnaK* under non-heat-shock conditions. Transcripts extending through *dnaJ* are only clearly detected upon heat shock (Wetzstein *et al.*, 1992; Narberhaus *et al.*, 1992). Considering this arrangement it cannot be completely ruled out that the *orf4* gene is part of the same operon which has its own promoter allowing constitutive expression under normal conditions. Analysis of transcriptional start sites will help to determine whether *orf4* is part of the *dnaK* operon or forms an exclusively separate transcriptional unit. Alignment of the *orf4* amino acid sequences with all sequences available in the EMBL and GenBank databases indicated homology with several membrane-associated proteins, particularly at the C-terminus (Fischetti *et al.*, 1991). Membrane-associated proteins have been found upstream of heat-shock genes in *Synechocystis* sp. (Chitnis & Nelson, 1991) and *E. coli* (James *et al.*, 1993). As well as acting as molecular chaperones in the correct folding and assembly of proteins, heat-shock genes have been implicated in protein translocation processes (Parsell & Sauer, 1989; Pelham, 1986) and therefore may be expected to be associated with the membrane in some way.

In *E. coli* the heat-shock response is regulated at the transcriptional level by a unique sigma factor (σ^{32}) which recognizes specific heat-shock promoters (Cowing *et al.*, 1985). In the *L. lactis* *dnaK* operon, only vegetative

promoter sequences could be identified which showed no homology with the consensus heat-shock promoter sequences. Stem-loop structures are found flanking the putative *orf1* promoter. Several heat-shock genes possess vegetative promoter sequences upstream of inverted repeat sequences (Table 4) and a consensus inverted repeat sequence has been described (TTAGCACTC-N₉-GAGTGCTAA; Wetzstein *et al.*, 1992), suggesting an involvement in the regulation of the heat-shock response. The precise role of the inverted repeat has not been determined but the formation of mRNA secondary structures which may protect against mRNA degradation and/or act as an activator-binding site for a DNA- or RNA-binding protein (Wetzstein *et al.*, 1992; Li & Wong, 1992; Narberhaus & Bahl, 1992; Schmidt *et al.*, 1992) has been proposed. In the case of the recently described *dnaJ* heat-shock gene of *L. lactis*, the stem-loop is found upstream of the promoter sequence and is thought to act as a repressor-binding site. Expression levels of a reporter gene (α -amylase) fused to the promoter region both with and without the inverted repeat region were compared. Constitutively high levels of α -amylase were produced from the promoter lacking an inverted repeat comparable with levels obtained from the promoter with an inverted repeat after heat shock. The unique arrangement of the *dnaK* regulatory region may reflect a more complex mechanism of control.

The inverted repeat is not always preceded by a vegetative promoter (Table 4). For example, the genes encoding the 10 kDa and 65 kDa antigens of *Mycobacterium tuberculosis* possess both vegetative and heat-shock-type promoter sequences. The *dnaK* heat-shock operon of *M. tuberculosis* is transcribed from heat-shock promoters but an inverted repeat has not been identified. In the Gram-negative species *Synechococcus* the *groESL* operon possesses both an inverted repeat and heat-shock promoter sequences and both systems may regulate the heat-shock response in these organisms. It is evident that a unique sigma factor is not always required for induction of the heat-shock response in Gram-positive bacteria.

The complete *dnaK* gene could not be cloned on the high-copy-number pUC vector suggesting that high levels of *L. lactis* DnaK expression may have an adverse effect on the host. In *E. coli* DnaK overproduction has several detrimental effects associated with plasmid instability and growth (Blum *et al.*, 1992). It may be that the *L. lactis* DnaK protein has some functional activity in *E. coli*.

The induction of the *L. lactis* DnaK protein appears to be approximately 3-fold on heat shock at 42 °C based on the Western blot data. This value corresponds well with the 2 to 3-fold induction of the *dnaJ* gene on heat shock (Van Asseldonk *et al.*, 1993). Another immunoreactive protein of 50 kDa can be seen in *L. lactis* and although HSPs of 51 and 49 kDa have been observed in *L. lactis* (Whitaker & Batt, 1991) the 50 kDa cross-reacting protein is not induced on heat shock. The presence of other immunoreactive *E. coli* and *L. lactis* proteins is probably due to the polyclonal DnaK antibody used. Work is currently in progress to study regulation of the *L. lactis* *dnaK* promoter under different environmental conditions by chromosomal integration of bacterial luciferase reporter genes (Stewart & Williams, 1992) into the *L. lactis* *dnaK* operon.

We wish to thank Dr R. James for helpful discussions and for supplying the DnaK antibody. This work was supported by contract BIOTC CT-910363 from the BRIDGE programme of the European Community.

References

- AUFFRAY, Y., GANSEL, X., THAMMAVONGS, B. & BOUTIBONNES, P. (1992). Heat shock induced protein synthesis in *L. lactis* subsp. *lactis*. *Current Microbiology* **24**, 281–284.
- BAIRD, P. N., HALL, L. M. C. & COATES, A. R. M. (1989). Cloning and sequence analysis of the 10 kDa antigen gene of *Mycobacterium tuberculosis*. *Journal of General Microbiology* **135**, 931–939.
- BARDWELL, J. C. A. & CRAIG, E. A. (1984). Major heat shock genes of *Drosophila* and the *Escherichia coli* heat-inducible *dnaK* gene are homologous. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 848–852.
- BARDWELL, J. C. A., TILLY, K., CRAIG, E., KING, J., ZYLICZ, M. & GEORGIOPOULOS, C. (1986). The nucleotide sequence of the *Escherichia coli* K12 *dnaJ*⁺ gene. *Journal of Biological Chemistry* **261**, 1782–1785.
- BIRNBOIM, H. C. & DOLY, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research* **7**, 1513–1519.
- BLUM, P., VELLIGAN, M., LIN, N. & ABDUL, M. (1992). DnaK mediated alterations in human growth hormone protein inclusion bodies. *Biotechnology* **10**, 301–304.
- BOLIVAR, F., RODRIGUES, R. L., GREENE, P. J., BETLACH, M. C., HEYNEKER, H. L., BOYER, H. W., CROSA, J. H. & FALKOW, S. (1977). Construction and characterization of new cloning vehicles: a multi purpose cloning system. *Gene* **2**, 95–114.
- BOUTIBONNES, P., TRANCHARD, C., HARTKE, A., THAMMAVONGS, B. & AUFRAY, Y. (1992). Is thermotolerance correlated to heat-shock protein synthesis in *L. lactis* subsp. *lactis*? *International Journal of Food Microbiology* **16**, 227–236.
- BRADFORD, M. M. (1976). A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* **72**, 248–254.
- CASADABAN, M. J. & COHEN, S. N. (1980). Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *Journal of Molecular Biology* **138**, 179–207.
- CELLIER, M. F. M., TEYSSIER, J., LAUTARD, J. P., MARTI, J. & SRIWIDADA, J. (1992). Cloning and characterization of the *Brucella ovis* heat shock protein DnaK functionally expressed in *Escherichia coli*. *Journal of Bacteriology* **174**, 8036–8042.
- CHITNIS, P. R. & NELSON, N. (1991). Molecular cloning of genes encoding two chaperone proteins from the cyanobacteria *Synechocystis* sp. PCC6803. *Journal of Biological Chemistry* **266**, 58–65.
- COWING, D. W., BARDWELL, J. C. A., CRAIG, E. A., WOOLFORD, C., HENDRIX, R. W. & GROSS, C. A. (1985). Consensus sequence for *Escherichia coli* heat shock gene promoters. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 2679–2683.
- DEAN, D. O. & JAMES, R. (1991). Identification of a gene, closely linked to *dnaK*, which is required for high-temperature growth of *Escherichia coli*. *Journal of General Microbiology* **137**, 1271–1277.
- FISCHETTI, V. A., PANCHOLI, V. & SCHNEEWIND, O. (1991). Common characteristics of the surface proteins from Gram positive cocci. In *Genetics and Molecular Biology of Streptococci, Lactococci and Enterococci*, pp. 290–294. Edited by G. M. Dunne, P. P. Cleary & L. L. McKay. Washington DC: American Society for Microbiology.
- FREIER, S. M., KIEREK, R., JAEGER, J. A., SUGIMOTO, N., CARUTHERS, M. H., NEILSON, T. & TURNER, D. H. (1986). Improved free-energy parameters for predictions of RNA duplex stability. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 9373–9377.
- GASSON, M. J. (1983). Plasmid complements of *Streptococcus lactis* NCDO712 and other lactic streptococci after protoplast induced curing. *Journal of Bacteriology* **154**, 1–9.
- GEORGIOPOULOS, C. P. & HERSKOWITZ, I. (1971). *Escherichia coli* mutants blocked in lambda synthesis. In *The Bacteriophage Lambda*, pp. 553–564. Edited by A. D. Hershey. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- GETTING, M. J. & SAMBROOK, J. (1992). Protein folding in the cell. *Nature, London* **355**, 33–45.
- GOFF, S. A. & GOLDBERG, A. L. (1985). Production of abnormal proteins in *Escherichia coli* stimulates transcription of *lon* and other heat shock genes. *Cell* **41**, 587–595.
- GRAVES, M. C. & RABINOWITZ, J. C. (1986). *In vivo* and *in vitro* transcription of the *Clostridium pasteurianum* ferredoxin gene. *Journal of Biological Chemistry* **261**, 11409–11415.
- GRUNSTEIN, M. & WALLIS, J. (1979). Colony hybridization. *Methods in Enzymology* **68**, 379–389.
- HARDIE, J. M. (1986). In *Bergey's Manual of Systematic Bacteriology*, vol. 2, pp. 1065–1068. Edited by J. G. Holt, P. H. A. Sneath, N. S. Mair & M. E. Sharpe. Baltimore: Williams & Wilkins.
- HEARNE, C. M. & ELLAR, D. J. (1989). Nucleotide sequence of a *Bacillus subtilis* gene homologous to the *dnaK* gene of *Escherichia coli*. *Nucleic Acids Research* **17**, 8873.
- HEYDE, M. & PORTALIER, P. (1990). Acid shock proteins of *Escherichia coli*. *FEMS Microbiology Letters* **69**, 19–26.
- JAMES, R., DEAN, D. O. & DEBBAGE, J. (1993). Five open reading frames

- upstream of the *dnaK* gene of *Escherichia coli* DNA sequence. *Journal of DNA Sequencing & Mapping* **3**, 327–332.
- JOHNSON, C., CHANDRASEKHAR, G. N. & GEORGIOPOULOS, C. (1989). *Escherichia coli* DnaK and GrpE heat shock proteins interact both *in vivo* and *in vitro*. *Journal of Bacteriology* **171**, 1590–1596.
- LANGER, T., LU, C., ECHOLS, H., FLANAGAN, J., HAYER, M. K. & HARTL, F. U. (1992). Successive action of DnaK, DnaJ and GroEL along the pathway of protein mediated folding. *Nature, London* **356**, 683–689.
- LENNOX, E. S. (1955). Transduction of linked genetic characters of the host bacteriophage P1. *Virology* **1**, 190–200.
- LEWINGTON, J., GREENAWAY, S. D. & SPILLANE, B. J. (1987). Rapid small scale preparation of bacterial genomic DNA suitable for cloning and hybridization analysis. *Letters in Applied Microbiology* **5**, 51–53.
- LI, M. & WONG, S. L. (1992). Cloning and characterization of the *groESL* operon from *Bacillus subtilis*. *Journal of Bacteriology* **174**, 3981–3992.
- LUDWIG, W., SEEWALDT, E., KILPPER-BALZ, R., SCHEIFER, K. H., MAGNUM, L., WOESE, C. R., FOX, G. E. & STACKEBRANDT, E. (1985). The phylogenetic position of *Streptococcus* and *Enterococcus*. *Journal of General Microbiology* **131**, 543–551.
- MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- MARCK, C. (1988). 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acids Research* **16**, 1829–1836.
- MARSHALL, V. M. (1987). Fermented milks and their future trends. I. Microbiological aspects. *Journal of Dairy Research* **54**, 559–574.
- MEURY, J. & KOHIYAMA, M. (1991). Role of heat shock protein DnaK in osmotic adaptation of *Escherichia coli*. *Journal of Bacteriology* **173**, 4404–4410.
- MORIMOTO, R. I., TISSIERES, A. & GEORGIOPOULOS, C. (1990). The stress response, function of the proteins, and perspectives, 1. In *Stress Proteins in Biology and Medicine*, pp. 1–36. Edited by R. I. Morimoto, A. Tissieres & C. Georgopoulos. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- NARBERHAUS, F. & BAHL, H. (1992). Cloning, sequencing and molecular analysis of the *groESL* operon of *Clostridium acetobutylicum*. *Journal of Bacteriology* **174**, 3282–3289.
- NARBERHAUS, F., GIEBELER, K. & BAHL, H. (1992). Molecular characterization of the *dnaK* gene region of *Clostridium acetobutylicum*, including *grpE*, *dnaJ* and a new heat shock gene. *Journal of Bacteriology* **174**, 3290–3299.
- PAEK, K. H. & WALKER, G. C. (1987). *Escherichia coli* *dnaK* null mutants are inviable at high temperatures. *Journal of Bacteriology* **169**, 283–290.
- PARSELL, D. A. & SAUER, R. T. (1989). Induction of a heat shock-like response by unfolded protein in *Escherichia coli*: dependence on protein level not protein degradation. *Genes and Development* **3**, 1226–1232.
- PELHAM, H. R. B. (1986). Speculations on the functions of the major heat shock and glucose related proteins. *Cell* **46**, 959–961.
- PLATT, T. (1986). Transcription termination and the regulation of gene expression. *Annual Review of Biochemistry* **55**, 339–372.
- SAKAKIBARA, Y. (1988). The *dnaK* gene of *Escherichia coli* functions in initiation of chromosome replication. *Journal of Bacteriology* **170**, 972–979.
- SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5463–5467.
- SHINNICK, T. M. (1987). The 65 kilodalton antigen of *Mycobacterium tuberculosis*. *Journal of Bacteriology* **169**, 1080–1088.
- SCHMIDT, A., SCHIESSWOHL, M., VOLKER-HECKER, M. & SCHUMANN, W. (1992). Cloning, sequencing, mapping, and transcriptional analysis of the *groESL* operon from *Bacillus subtilis*. *Journal of Bacteriology* **174**, 3993–3999.
- STEWART, G. S. A. B. & WILLIAMS, P. (1992). *lux* genes and the applications of bacterial bioluminescence. *Journal of General Microbiology* **138**, 1289–1300.
- SUSSMAN, M. D. & SETLOW, P. (1987). Nucleotide sequence of a *Bacillus megaterium* gene homologous to the *dnaK* gene of *Escherichia coli*. *Nucleic Acids Research* **15**, 3923–3928.
- TAGLICH, D., PADAN, E., OPPENHEIM, A. B. & SCHULDINER, S. (1987). An alkaline shift induces the heat shock response in *Escherichia coli*. *Journal of Bacteriology* **169**, 885–887.
- TERZHAGI, B. E. & SANDINE, W. E. (1975). Improved medium for lactic *Streptococci* and their bacteriophages. *Applied Microbiology* **29**, 807–813.
- TILLY, H., HANSER, R., CAMPBELL, J. & OSTHEIMER, G. J. (1993). Isolation of *dnaJ*, *dnaK* and *grpE* homologues from *Borrelia burgdorferi* and complementation of *Escherichia coli* mutants. *Molecular Microbiology* **1**, 359–369.
- VAN ASSELDONK, M., SIMONS, A., VISSER, H., DE VOS, W. M. & SIMONS, G. (1993). Cloning, nucleotide sequence and regulatory analysis of the *Lactococcus lactis* *dnaJ* gene. *Journal of Bacteriology* **175**, 1637–1644.
- VAN DE GUCHTE, M., KOK, J. & VENEMA, G. (1992). Gene expression in *Lactococcus lactis*. *FEMS Microbiology Reviews* **88**, 73–92.
- VÖLKER, U., MACH, H., SCHMIDT, R. & HECKER, M. (1992). Stress proteins and cross-protection by heat shock and salt stress in *Escherichia coli*. *Journal of General Microbiology* **138**, 2125–2135.
- WEBB, R., REDDY, K. J. & SHERMAN, L. A. (1990). Regulation and sequence of the *Synechococcus* sp. strain PCC7942 *groESL* operon, encoding a cyanobacterial chaperonin. *Journal of Bacteriology* **172**, 5079–5088.
- WETZSTEIN, M., VOLKER, U., DEDIO, J., LOBAU, S., ZUBER, U., SCHIESSWOHL, M., HERGET, C., HECKER, M. & SCHUMANN, W. (1992). Cloning, sequencing and molecular analysis of the *dnaK* locus from *Bacillus subtilis*. *Journal of Bacteriology* **174**, 3300–3310.
- WHITAKER, R. D. & BATT, C. A. (1991). Characterization of the heat shock response in *Lactococcus lactis* subsp. *lactis*. *Applied and Environmental Biology* **57**, 1408–1412.
- YANISCH-PERRON, C., VIEIRA, J. & MESSING, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119.
- ZUKER, M. & STIEGLER, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research* **9**, 133–148.