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 HindIII 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 Orfa 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
Table 1. Bacterial strains and plasmids used in this study

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Characteristics</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strain</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. coli SURE</td>
<td>Δ(hsdRMS) merC mcrB mrr endA1 supE44 thi-1 gyrA96 relA1 lacC2 recF1 sbeC urac1 uvrC [F' proAB lacFΔM15 Tn10 (Tet)]</td>
<td>Stratagene</td>
</tr>
<tr>
<td>E. coli JM109</td>
<td>[recA1 endA1 gyrA96 thi hsdR17 (rk-) mcrA mrr Δ(lacZ)M15]</td>
<td>Yanisch-Perron et al. (1985)</td>
</tr>
<tr>
<td>L. lactis subsp. lactis MG1363</td>
<td>Plasmid-free strain</td>
<td>Gasson (1983)</td>
</tr>
<tr>
<td><strong>Plasmid</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pUC18</td>
<td>Ampicillin resistance</td>
<td>Yanisch-Perron et al. (1985)</td>
</tr>
<tr>
<td>pBR322</td>
<td>Ampicillin and tetracycline resistance</td>
<td>Bolivar et al. (1977)</td>
</tr>
<tr>
<td>pFI483</td>
<td>346 bp dnaK PCR fragment cloned into SmaI site of pUC18</td>
<td>This study</td>
</tr>
<tr>
<td>pF1573</td>
<td>pBR322 containing 54 kb chromosomal dnaK HindIII fragment</td>
<td>This study</td>
</tr>
<tr>
<td>pF1574 (pF1619)</td>
<td>1.6 kb EcoRI fragment of pF1573 cloned into SmaI site of pUC18 (pF1619, opposite orientation)</td>
<td>This study</td>
</tr>
<tr>
<td>pF1575 (pF1620)</td>
<td>1.8 kb EcoRI fragment of pF1573 cloned into SmaI site of pUC18 (pF1620, opposite orientation)</td>
<td>This study</td>
</tr>
<tr>
<td>pF1576</td>
<td>2.0 kb EcoRI fragment of pF1573 cloned into SmaI site of pUC18</td>
<td>This study</td>
</tr>
<tr>
<td>pF1672</td>
<td>2.3 kb inverse PCR fragment cloned into EcoRV site of pBR322</td>
<td>This study</td>
</tr>
</tbody>
</table>

Table 2. Oligonucleotide primers used for PCR amplification

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>P10</td>
<td>5'-GGTATCGATTTAGGTACAACT-3'</td>
<td>Conserved region aa 6-12* (6-12)†</td>
</tr>
<tr>
<td>P9</td>
<td>5'-TTRAARTANGCNGGNACNGT-3'†</td>
<td>Conserved region aa 115-121 (116-122)</td>
</tr>
<tr>
<td>P69</td>
<td>5'-GCAACTGTGAAAGCTACCTTG-3'</td>
<td>Inverse PCR primer nucleotide no. 2225-2246 (Fig. 3)</td>
</tr>
<tr>
<td>P100</td>
<td>5'-GCATTCGCTCATCATCACT-3'</td>
<td>Inverse PCR primer nucleotide no. 1003-1021 (Fig. 3)</td>
</tr>
</tbody>
</table>

* Amino acid (aa) notation refers to *B. megaterium* DnaK (Sussman & Setlow, 1987).
† R = A or G, N = A, C, G or T.
‡ aa 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 (Terzagi & 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 (P9/10). The DNA probe was labelled with [α-32P]ATP using the Multiprime DNA labelling kit (Amersham), in 1% low melting point agarose. For colony hybridization, colonies were resuspended in a Biospec BeadBeater and subsequently placed on ice for 5 min.

Protein extraction. E. coli and L. lactis cultures were grown to an OD of between 0.9 and 1.1, or were grown to an OD 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-Na2HPO4·12H2O, pH 6.6) and the suspension added to 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% (v/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'- triphosphate 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 23 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 Gucht 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 was 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-
Cloning and sequencing of the dnaK gene region

A partial genomic library of HindIII 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 HindIII 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 EcoRI-digested chromosomal DNA which corresponds to a template size of 3.5 kb. The 2.3 kb fragment was cloned into the EcoRV 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 65.4 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
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%.
orf4 contains 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-N₆-GAGTGCTAAATTTT 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 TTAGCACTC-N₆-GAGTGCTAAATTTT. 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).
Although the actual identity with the other proteins ranges from 29.0% to 63.5% (E. coli), 52.5% to 66.1% (B. subtilis), 53.6% to 77.4% (B. megaterium) and 56.6% to 78.9% (C. acetobutylicum), 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.

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

<table>
<thead>
<tr>
<th>Organism</th>
<th>Orf1</th>
<th>GrpE</th>
<th>DnaK</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli</td>
<td>–</td>
<td>29.0%</td>
<td>57.0%</td>
</tr>
<tr>
<td>B. subtilis</td>
<td>27.2%</td>
<td>43.3%</td>
<td>77.0%</td>
</tr>
<tr>
<td>(52.5%)</td>
<td>(63.5%)</td>
<td>(87.3%)</td>
<td></td>
</tr>
<tr>
<td>B. megaterium</td>
<td>–</td>
<td>–</td>
<td>77.4%</td>
</tr>
<tr>
<td>(56.2%)</td>
<td>(56.6%)</td>
<td>(87.9%)</td>
<td></td>
</tr>
<tr>
<td>C. acetobutylicum</td>
<td>24.6%</td>
<td>38.6%</td>
<td>66.1%</td>
</tr>
<tr>
<td>(53.6%)</td>
<td>(56.5%)</td>
<td>(87.9%)</td>
<td></td>
</tr>
</tbody>
</table>

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.
Analysis of the Lactococcus lactis dnaK operon

Orf a
Orf 39
Orf 1

Fig. 4. Comparison of the deduced amino acid sequences encoded by orf1, orfa and orf39 from L. lactis, C. acetobutyllicum 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
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-N,-NAGTGCTAA; 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) 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, whereas 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.

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

<table>
<thead>
<tr>
<th>Organism</th>
<th>Gene or protein</th>
<th>Inverted repeat</th>
<th>Promoter</th>
<th>Gene organization</th>
<th>Source/reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gram-positive</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L. lactis</td>
<td>orf1</td>
<td>+</td>
<td>v</td>
<td>v</td>
<td>orf1-grpE-dnaK-T-orf4</td>
</tr>
<tr>
<td>L. lactis</td>
<td>dnaJ</td>
<td>+</td>
<td>v</td>
<td>v</td>
<td>dnaJ</td>
</tr>
<tr>
<td>B. subtilis</td>
<td>orf39</td>
<td>+</td>
<td>v</td>
<td>v</td>
<td>orf39-grpE-dnaK-T-dnaJ</td>
</tr>
<tr>
<td>B. subtilis</td>
<td>groESL</td>
<td>+</td>
<td>v</td>
<td>v</td>
<td>groES-groEL</td>
</tr>
<tr>
<td>C. acetobutylicum</td>
<td>orfa</td>
<td>+</td>
<td>v</td>
<td>v</td>
<td>orfa-grpE-dnaK-T-dnaK</td>
</tr>
<tr>
<td>C. acetobutylicum</td>
<td>groESL</td>
<td>+</td>
<td>v</td>
<td>v</td>
<td>dnaK-grpE-groEL</td>
</tr>
<tr>
<td>M. tuberculosis</td>
<td>10kDa antigen</td>
<td>+</td>
<td>v</td>
<td>v</td>
<td>ORF</td>
</tr>
<tr>
<td>M. tuberculosis</td>
<td>65kDa antigen</td>
<td>+</td>
<td>v</td>
<td>v</td>
<td>ORF</td>
</tr>
<tr>
<td>M. tuberculosis</td>
<td>dnaK</td>
<td>-</td>
<td>hs</td>
<td>hs</td>
<td>dnaK-grpT-T-dnaJ</td>
</tr>
<tr>
<td>Streptomyces coelicolor</td>
<td>dnaK</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>dnaK-grpE</td>
</tr>
<tr>
<td>B. megaterium</td>
<td>dnaK</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>dnaK</td>
</tr>
<tr>
<td><em>Eryxiploforix rhisiopathiae</em></td>
<td>dnaK</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>dnaK</td>
</tr>
<tr>
<td><strong>Gram-negative</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. coli</td>
<td>dnaK</td>
<td>-</td>
<td>hs</td>
<td>hs</td>
<td>dnaK-T-dnaJ-T</td>
</tr>
<tr>
<td>Brucella ovis</td>
<td>dnaK</td>
<td>-</td>
<td>hs</td>
<td>hs</td>
<td>dnaK-T-dnaJ-T</td>
</tr>
<tr>
<td>Borrelia burgdorferi</td>
<td>grpE</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>grpE-dnaK-dnaJ-T</td>
</tr>
<tr>
<td>Synechococcus sp.</td>
<td>groESL</td>
<td>+</td>
<td>v</td>
<td>v</td>
<td>groES-groEL</td>
</tr>
<tr>
<td>Synechococcus sp.</td>
<td>dnaK</td>
<td>ND</td>
<td>?</td>
<td>hs</td>
<td>orf70.5-dnaK-T-orf70.3</td>
</tr>
<tr>
<td>Synechocystis sp.</td>
<td>cps-60</td>
<td>+</td>
<td>?</td>
<td>hs</td>
<td>orf60.5a-orf60.5-cpn60-orf60.3</td>
</tr>
</tbody>
</table>

+ = 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.
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 groEESL 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


James, R., Dean, D. O. & Debbage, J. (1993). Five open reading frames


