JOURNAL OF BACTERIOLOGY, May 1992, p. 3290–3299
0021-9193/92/103290-10$02.00/0

Vol. 174, No. 10

Molecular Characterization of the *dnaK* Gene Region of *Clostridium acetobutylicum*, Including *grpE*, *dnaJ*, and a New Heat Shock Gene

FRANZ NARBERHAUS, KATHARINA GIEBELER,† AND HUBERT BAHL*

Institut für Mikrobiologie, Georg-August-Universität Göttingen, Grisebachstrasse 8, W-3400 Göttingen, Germany

Received 2 December 1991/Accepted 10 February 1992

The *dnaK* gene region of *Clostridium acetobutylicum* was cloned in *Escherichia coli* by using the pBluescript SK+ and pUC18 vectors. By using the *E. coli dnaK* gene as a probe and by in vivo chromosome walking, three positive clones harboring the recombinant plasmids pKG1, pKG2, and pKG3 containing 1.2-kbp *Hin*dIII, 3.55-kbp *Eco*RV, and 1.2-kbp *Pst*I fragments of the chromosome of *C. acetobutylicum*, respectively, were isolated. The cloned fragments partially overlapped, and together they spanned 4,083 bp of the clostridial genome that were completely sequenced. On one strand, four open reading frames of which the last was obviously truncated were identified. The last three genes showed high homology to the *grpE*, *dnaK*, and *dnaJ* heat shock genes of *E. coli*, respectively. They were preceded by an open reading frame (*orfA*) without any homology to sequences available in the EMBL or GenBank data bases. Typical translational start sites could be found in front of all four genes. Northern (RNA) blot analysis revealed several transcripts of this region with a maximum length of 5.0 kb. Thus, these genes are probably organized in an operon. A transcription terminator could be found between the *dnaK* and *dnaJ* genes. By primer extension analysis, a major heat-inducible transcription start site was identified 49 bases upstream of *orfA*. This site was preceded by a region (5′-TTGACA[17 bp]TATTTT) that exhibited high homology to the consensus promoter sequences of gram-positive bacteria as well as σ^{70}-dependent *E. coli*. Between this promoter and the initiation codon of *orfA*, a hairpin-loop structure with a possible regulatory role in the expression of these genes was found. Additional heat-inducible transcription start sites were located 69 bases upstream of *orfA* and 87 bases upstream of *grpE*; the corresponding promoter regions showed less similarity to other known promoter sequences. Maximum mRNA levels of this heat shock operon were found about 15 min after a heat shock from 30 to 42°C. Our results indicate that *orfA* codes for an unknown heat shock protein.

Clostridium acetobutylicum, a gram-positive, strictly anaerobic sporeformer, has been used for the production of acetone and butanol for a long time. This organism is able to change its metabolism according to growth conditions. During growth in a batch culture, *C. acetobutylicum* ferments sugars to acetate, butyrate, H_2, and CO_2. Shortly before entering the stationary phase, the residual sugars and most of the performed acids are converted to butanol and acetone (for reviews, see references 3 and 23). The parameters effecting this shift (e.g., low pH, threshold concentrations of butyrate, and suitable growth-limiting factors such as phosphate) have been determined, mostly by continuous culture experiments (1, 2). Furthermore, the switch from acid to solvent formation seems to be associated with cell differentiation (alterations in cell shape, storage of granulose, and sporulation) (25) and to be connected with the heat shock response (37, 48). Specific heat shock proteins including DnaK and GroEL are induced during the onset of solvent formation (37). The induction of these stress proteins was observed before acetone and butanol were detectable in the medium and was therefore not a response to the presence of these solvents in the medium. It is likely that the above-mentioned stressful conditions, responsible for the metabolic switch, are also able to trigger the heat shock response.

To elucidate a possible connection between the heat shock response and the initiation of solvent formation at the molecular level, an analysis of the corresponding genes is necessary. Recently, much progress in the characterization of genes encoding enzymes necessary for solvent formation has been made. Thus, the NADPH-dependent ethanol dehydrogenase, an NADP-dependent butanol dehydrogenase, the acetoacetate decarboxylase, and the acetyl coenzyme A acyltransferase (thiolase) have been cloned and in some cases sequenced (18, 35, 36, 55, 56).

The aim of this study was to clone the *dnaK* locus of *C. acetobutylicum* to determine the nucleotide sequence and to identify promoter regions and possible regulatory sequences in order to better understand heat shock gene expression in *C. acetobutylicum*.

MATERIALS AND METHODS

Bacterial strains, plasmids, and growth conditions. *C. acetobutylicum* DSM 1731 was used as a source of genomic DNA and total RNA. *Escherichia coli* XL1-Blue [*recA1 endA1 gyrA96 thi relA1 lacZ*ΔM15 *lacQ* F′ *proAB* Tn*10*] (11) and *E. coli* JM109 [*recA1 endA1 gyrA96 thi hsdR17* (r_K^+ m_K^+) *supE44 relA1* λ⁻ Δ(*lac-proAB*)(F′ *traD36 proAB*$^+$ *lacI*^{q} *lacZ*ΔM15)] (54) were used as hosts for the cloning experiments. The plasmids pBluescript SK+ (Stratagene, San Diego, Calif.) and pUC18 (54) served as vectors for the construction of genomic libraries. Plasmid pMK341, a derivative of pMOB45 (10), contained the *dnaK* gene of *E. coli* on a 5.3-kbp *Hin*dIII fragment, which was used as a probe in

* Corresponding author.
† Present address: Europäisches Patentamt, W-8000 München 2, Germany.

hybridization experiments. *C. acetobutylicum* was grown anaerobically in CBM medium (33) at 37°C. *E. coli* was routinely grown at 37°C in Luria-Bertani medium (40) supplemented with ampicillin (50 μg/ml), isopropyl-β-thiogalactopyranoside (IPTG, 50 μg/ml), or 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal, 40 μg/ml) if required.

DNA isolation and manipulation. Chromosomal DNA of *C. acetobutylicum* was isolated by the method of Marmur (27) as modified by Bertram and Dürre (8). Plasmid isolation from *E. coli* was performed by the method of Birnboim and Doly (9) or with the Quiagen Midi Kit (Diagen GmbH, Düsseldorf, Germany).

DNA was manipulated by standard methods (40). Restriction enzymes were obtained from GIBCO/BRL GmbH (Eggenstein, Germany) or Pharmacia LKB GmbH (Freiburg, Germany), and calf intestinal phosphatase was from Boehringer GmbH (Mannheim, Germany). The enzymes were used according to the instructions of the manufacturers. DNA restriction fragments were isolated from agarose gels by electroelution in a Biotrap BT1000 (Schleicher & Schuell, Dassel, Germany).

Construction and screening of gene banks. Gene banks were constructed with completely digested chromosomal DNA of *C. acetobutylicum*. For partial gene banks, DNA fragments in the respective size ranges were electroeluted from an agarose gel; for total gene banks, digested chromosomal DNA was used after extraction with phenol-chloroform and chloroform and precipitation with ethanol. The DNA fragments were ligated to the appropriately digested and dephosphorylated vectors. *E. coli* was transformed with the ligation mixture by electroporation in a Gene Pulser (Bio-Rad Laboratories GmbH, Munich, Germany). White colonies containing inserts were screened by colony hybridization with the respective probes. Positive clones were tested by restriction endonuclease digestion and were sequenced.

Preparation of RNA. Total RNA was isolated from *C. acetobutylicum* by the hot phenol-chloroform procedure described by Oelmüller et al. (34). To obtain higher yields of RNA, cells were not washed in acetate-EDTA buffer and the extraction in 60°C hot phenol-chloroform was extended to 10 min.

Hybridization. Total chromosomal DNA of *C. acetobutylicum* was digested to completion with the corresponding restriction enzymes and separated on agarose gels. Southern blots on nylon membranes (GeneScreen Plus; Dupont, NEN Research Products, Dreieich, Germany) were performed according to the manufacturer's instructions, and prehybridization was done in 0.2% (wt/vol) polyvinylpyrrolidone–0.2% (wt/vol) Ficoll–0.2% (wt/vol) bovine serum albumin–50 mM Tris hydrochloride (pH 7.5)–1 M NaCl–0.1% (wt/vol) sodium pyrophosphate–1% (wt/vol) sodium dodecyl sulfate (SDS)–10% (wt/vol) dextran sulfate at 60°C for 2 to 6 h. The DNA probes were radiolabeled with [α-^{32}P]dCTP (Amersham Buchler GmbH, Braunschweig, Germany) by using a nick translation kit (GIBCO/BRL). The labeled fragments were purified by column chromatography on Sephadex G-25, and 0.1 to 0.2 μCi/ml was added to the prehybridization solution. After hybridization at 60°C for 15 to 20 h, the membranes were washed twice in 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) for 5 min at room temperature and subjected to autoradiography. If necessary, the blot was washed again twice in 2× SSC–1% (wt/vol) SDS for 30 min at 60°C and twice in 0.1% (wt/vol) SDS for 30 min at room temperature.

Colony hybridization was done with BA85 nitrocellulose filters (Schleicher & Schuell). Colonies were picked directly on a master plate and on a plate containing the filter. Filters were prepared for hybridization as described by Grunstein and Wallis (21). Hybridization and washing were done as described above for Southern blots.

RNA for Northern blots was separated in formaldehyde gels and transferred to nylon membranes (GeneScreen Plus). Size determination was done by using an RNA ladder (0.24, 1.4, 2.4, 4.4, 7.5, and 9.5 kb; GIBCO/BRL) as the standard. RNA for dot blot experiments was denatured in 50% formamide–6% formaldehyde and spotted onto nylon filters by using a dot blot apparatus. Hybridization and washing were done as described above for the DNA hybridization procedure.

DNA sequencing. DNA was sequenced by the dideoxy-chain termination method of Sanger et al. (41), by using [^{35}S]dATP and a T7 sequencing kit (Pharmacia LKB). Single-stranded templates were prepared from both strands of the plasmids. Sequencing was started with commercially available M13/pUC universal sequencing forward primer and reverse primer. Synthetic oligonucleotides (17-mers) complementary to the ends of already sequenced templates were prepared by a Gene Assembler Plus (Pharmacia LKB) according to the instructions of the manufacturer and used as primers for continued sequencing. The dideoxy-terminated fragments were separated on 55-cm wedge-shaped thickness gradient gels (0.2 to 0.4 mm, 7% [wt/vol] polyacrylamide) with a Macrophor sequencing unit (Pharmacia LKB) as recommended by the manufacturer.

Determination of the transcription start site. For the determination of the transcription start site, primer extension analysis was used. Two picomoles of oligonucleotide (17-mer) was labeled with 10 μCi of [γ-^{32}P]ATP (Amersham Buchler) and 7 U of polynucleotide kinase (GIBCO/BRL) in PNK buffer (100 mM Tris [pH 7.6], 20 mM $MgCl_2$, 30 mM dithiothreitol) for 3 h at 37°C (10-μl reaction volume). The reaction was stopped by heating for 2 min at 95°C. Ten micrograms of total RNA was incubated together with 12.5 U of RNasin (GIBCO/BRL) and 0.2 pmol of kinased oligonucleotide in 500 mM KCl–20 mM Tris (pH 7.9) buffer for 3 h at 30°C (final volume, 10 μl).

For the primer extension reaction, 6 μl of 5× PE buffer (250 mM Tris [pH 8.3], 125 mM KCl, 15 mM $MgCl_2$, 50 mM dithiothreitol), 500 μM dNTPs of each nucleotide, 2.5 μg of actinomycin D and 200 U of Moloney murine leukemia virus or SuperScript RNase H$^-$ reverse transcriptase (GIBCO/BRL) per 10 μg of RNA and 30 μl of H_2O was combined in the annealing assay and incubated at 37°C for 1 h. The RNA primer-extended hybrid was phenol-chloroform extracted, ethanol precipitated, and analyzed on 7% polyacrylamide–urea sequencing gels. The lengths of the primer extension products were calculated by running a sequencing reaction with the same primer on the same gel.

Computer programs. The DNA region sequenced and the deduced proteins were analyzed with the DNA Strider program (26) on a Macintosh SE computer (Apple Computer, Inc., Cupertino, Calif.). Sequence comparisons were done by using the Wisconsin Genetics Computer Group sequence analysis software package, version 6.0 (University of Wisconsin Biotechnology Center, Madison) (15).

Nucleotide sequence accession number. The sequence data reported here (see Fig. 3) have been submitted to GenBank and assigned accession no. M74569.

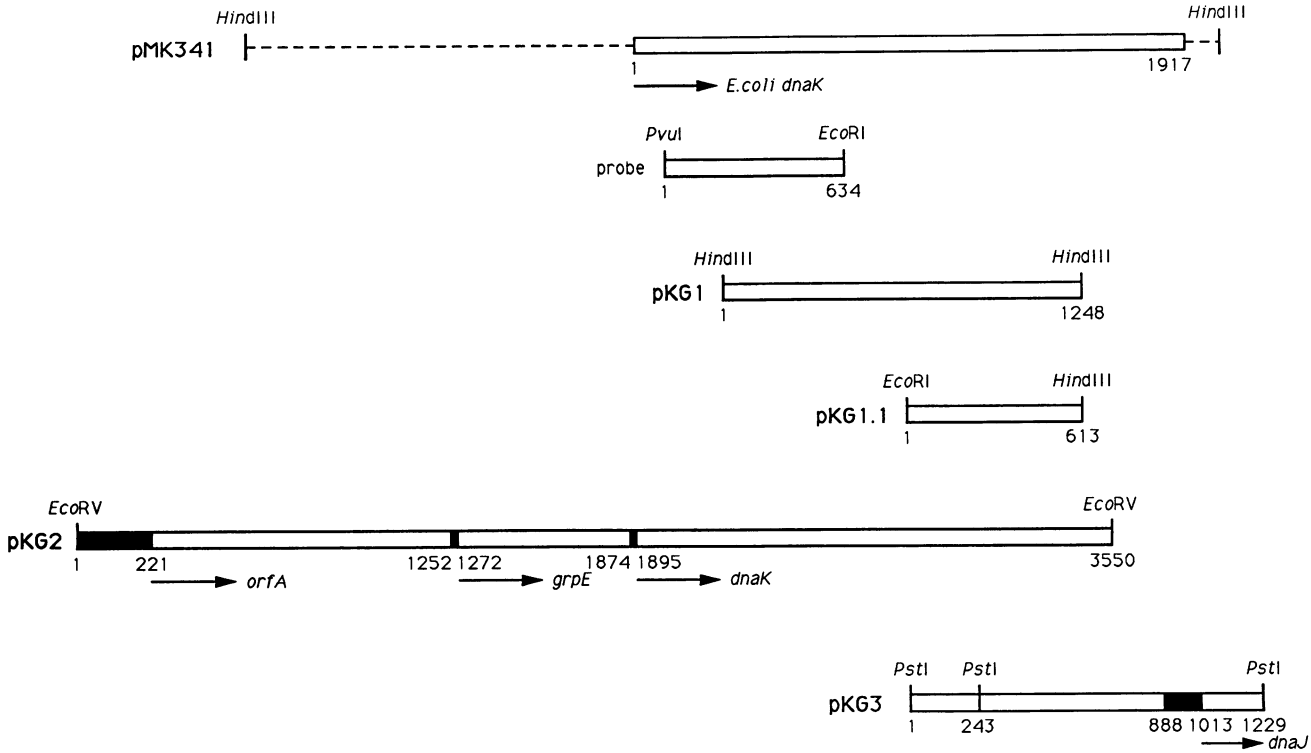


FIG. 1. Schematic representation of the inserts of pMK341, pKG1, pKG1.1, pKG2, and pKG3. The positions and lengths of ORFs (open segments) and the direction of transcription (arrows) have been determined by DNA sequencing; numbers represent nucleotide positions, starting from the left. According to identical sequences, the relationship between the clostridial fragments of pKG1, pKG1.1, pKG2, and pKG3 are shown in respect to each other. The *grpE*, *dnaK*, and *dnaJ* genes were identified because of their homology to corresponding genes of *E. coli* and *Bacillus* species.

RESULTS

Cloning and sequencing of the *dnaK* gene region of *C. acetobutylicum*. A 634-bp fragment of the *E. coli dnaK* gene (Fig. 1) was used as a probe to detect homologous sequences in the DNA of *C. acetobutylicum*. Hybridization of this fragment to chromosomal DNA of *C. acetobutylicum* digested with different restriction enzymes revealed several specific signals (Fig. 2). To isolate positive clones in a gene bank, it was necessary to separate the cloned fragments from host and vector DNAs before hybridization to avoid unspecific signals due to the required nonstringent washing conditions. Therefore, at the beginning of our experiments, a partial gene bank was constructed with the pBluescript SK+ vector and 1.2-kbp *Hin*dIII fragments of chromosomal DNA of *C. acetobutylicum*, which could be well separated from the vector band in agarose gels. Pools of 50 clones were grown in Luria-Bertani medium, and the recombinant plasmids were isolated, digested with *Hin*dIII, transferred to a nylon membrane, and hybridized with the radioactively labeled probe. One positive pool was detected, and the clone containing the respective 1.2-kbp *Hin*dIII fragment was designated pKG1 (Fig. 1). DNA sequencing revealed an open reading frame (ORF) truncated at both ends. The deduced amino acid sequence showed high homology to the amino acid sequence of the *E. coli* DnaK protein. The clostridial insert of pKG1 was then used as a homologous probe (stringent hybridization and washing conditions were possible) to clone the 3.5-kbp *Eco*RV fragment (Fig. 2) of *C. acetobutylicum*. *Eco*RV-digested chromosomal DNA of *C. acetobutylicum* and pUC18 were used to construct a gene bank, which was screened by colony hybridization. DNA sequencing of a positive clone (pKG2; Fig. 1) revealed two ORFs upstream of the *dnaK* gene. Since the *dnaK* gene was still truncated at the 3′ end, another gene bank containing *Pst*I fragments of *C. acetobutylicum* DNA in pUC18 was

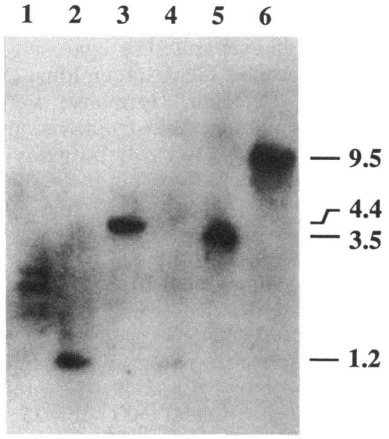


FIG. 2. Southern hybridization of restriction endonuclease-digested *C. acetobutylicum* DNA with the ^{32}P-labeled 0.63-kbp *Pvu*I-*Eco*RI fragment of *E. coli dnaK*. Lane 1, λ DNA digested with *Pst*I; lanes 2 to 6, 10 μg of chromosomal DNA digested with *Hin*dIII (lane 2), *Acc*I (lane 3), *Eco*RI (lane 4), *Eco*RV (lane 5), and *Msp*I (lane 6). Fragment sizes (kilobase pairs) of positive bands are indicated to the right.

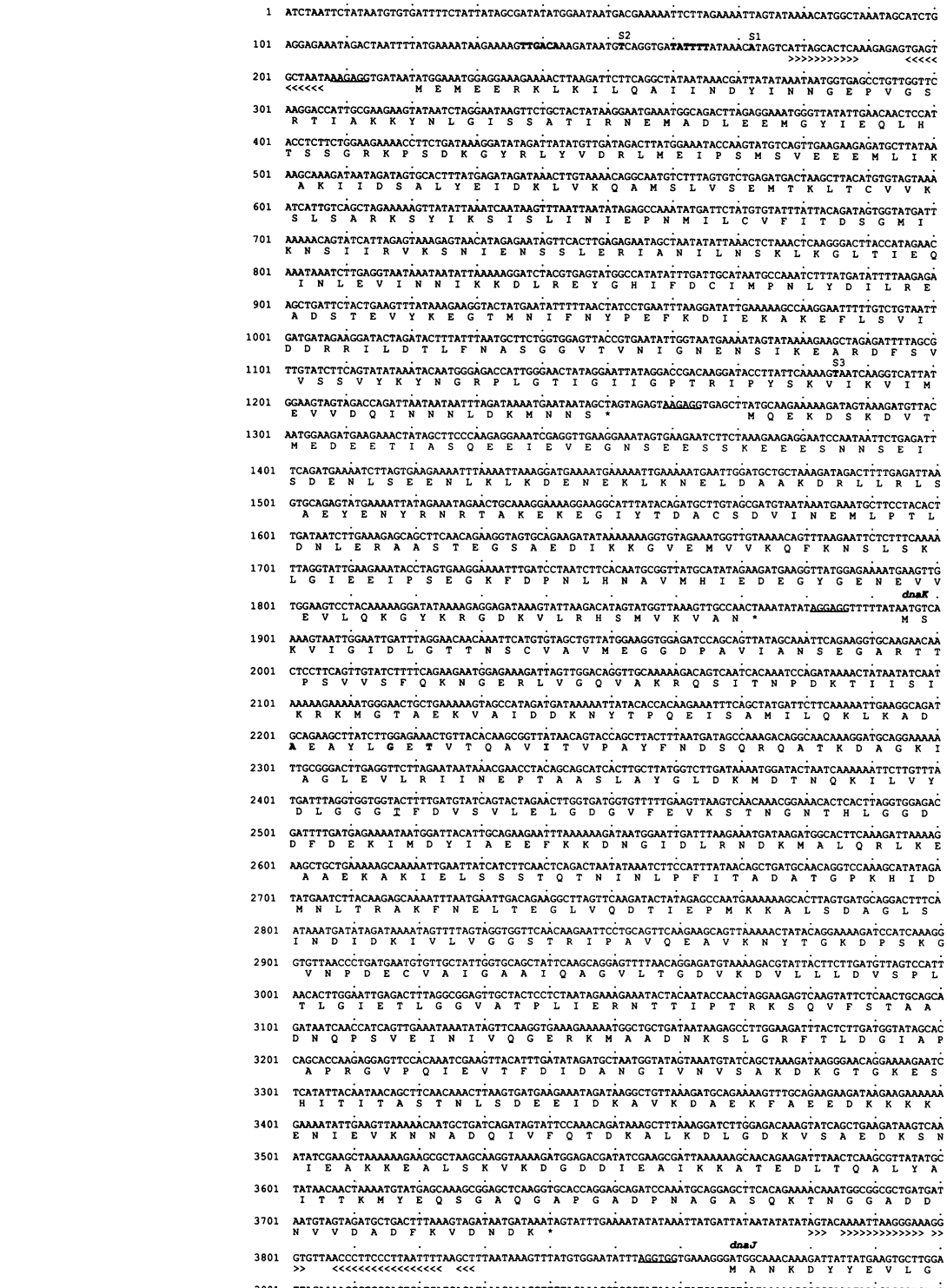


FIG. 3. Nucleotide sequence of the 4,083-bp *dnaK* gene region of *C. acetobutylicum*. Only the antisense strand is shown. The deduced amino acid sequences of *orfA*, *grpE*, *dnaK*, and *dnaJ* are shown below the DNA sequence (single-letter code). Translation stop signals are marked by asterisks below the respective codons. The putative ribosome-binding sites are underlined. Arrow heads below the DNA sequence show regions of dyad symmetry. The main transcription start site (S1) identified by primer extension analysis and the corresponding promoter are marked by bold letters in the −10 and −35 regions. Further putative transcription start sites are labeled S2 and S3. Amino acids of the possible ATP binding domain of the DnaK protein are marked with bold letters, and the threonine residue of the putative autophosphorylation site is underlined.

TABLE 1. Homology of the GrpE, DnaK, and DnaJ proteins of *C. acetobutylicum* to the corresponding proteins of other bacteria

Organism	% Identity and % similarity of the following proteins:			
	GrpE	DnaK	DnaJ[a]	Reference(s)
E. coli	30.7, 55.0	61.1, 76.2	61.1, 80.6	6, 7, 24
B. subtilis	43.2, 63.9	67.5, 81.1		22, 51
Bacillus megaterium		66.8, 80.8		47

[a] Based on the 5′ end of the gene.

screened with an *Eco*RI-*Hin*dIII fragment of pKG1 (subcloned in pKG1.1; Fig. 1) to obtain the entire sequence of the *dnaK* gene of *C. acetobutylicum*. A positive clone (pKG3) containing a 1.2-kbp *Pst*I fragment, which overlapped with the *Eco*RV fragment of pKG2 and contained the 3′ end of the *dnaK* gene and the start of another ORF (Fig. 1), was isolated. The complete nucleotide sequence of the *dnaK* gene region of *C. acetobutylicum* is shown in Fig. 3.

In front of all four genes, putative ribosome-binding sites were identified (Fig. 3). The sequences and distances showed reasonable homology to the *E. coli* prototype (43), although the site upstream of the fourth gene was somewhat unusual (5′-AGGTGG). A region with reasonable homology to the consensus promoter sequence of gram-positive bacteria (19) as well as *E. coli* (σ^{70} dependent [28]) was observed upstream of *orfA* at bases 137 to 165. The −35 sequence was 5′-TTGACA, and the −10 sequence was 5′-TATTTT, with 17 bp in between. Interestingly, between this region and the ATG start codon of the *orfA* gene, a hairpin-loop structure with a free energy of −57 kJ/mol (calculated with the Fold computer program [57] according to the determination of Freier et al. [17]) was identified with an 11-base palindromic sequence (5′-ATTAGCACTCA) forming a stem and 7 bases forming a loop, not followed by a sequence of several T's. A nucleotide sequence resembling a procaryotic rho-independent transcription terminator (38) consisting of a stem-loop structure with a free energy of −82.5 kJ/mol followed by sequence of 8 U's and A's is located downstream of the *dnaK* gene.

Identification of the genes of the *dnaK* gene region of *C. acetobutylicum*. The deduced amino acid sequences of the four ORFs of the cloned clostridial fragments in pKG2 and pKG3 were aligned to sequences available in the EMBL or GenBank data base. The two ORFs immediately upstream and downstream of the *dnaK* gene were identified as the *grpE*- and *dnaJ*-homologous genes of *C. acetobutylicum*, respectively, and were named accordingly. The GrpE, DnaK, and DnaJ heat shock proteins of *C. acetobutylicum* were most homologous to the corresponding proteins of *Bacillus* species (Table 1). Interestingly, the DnaK proteins of both organisms have an identical gap of 24 amino acids near the N terminus (Fig. 4) compared with the *E. coli* protein (69,072 Da), and thus, the molecular mass of *C. acetobutylicum* DnaK calculated from the deduced amino acid sequence is lower (65,608 Da). The proposed ATP binding (12) and autophosphorylation (58) sites of the *E. coli* DnaK protein were identified in the enzyme of *C. acetobutylicum* and are shown in Fig. 3.

The deduced protein encoded by the ORF upstream of *grpE* (*orfA*) showed no significant homology to other sequences of the data bases. However, the computer programs CodonPreference and TestCode (16) gave strong evidence for a protein coding function of *orfA*. TestCode identifies protein coding sequences independently of the reading frame by plotting a measure of nonrandomness of the composition at every third base. The entire *dnaK* gene region was subjected to this analysis (Fig. 5). The plot is divided into three regions for which the statistic makes prediction. The top region predicts coding sequences with 95% confidence. The bottom region predicts noncoding sequences with the same probability. The middle region does not allow statistic prediction. The data clearly confirmed coding regions for the *grpE*, *dnaK*, and *dnaJ* genes, and a coding sequence for *orfA* seems to be very likely. The calculated molecular mass of the deduced protein is 38,851 Da, with a pI of 6.07.

mRNA analysis of the *dnaK* gene region. Total RNA was isolated from *C. acetobutylicum* cells grown at 30°C and subjected to a heat shock from 30 to 42°C between 5 and 60 min. First, a dot blot was prepared and hybridized with a radioactively labeled fragment of the *dnaK* gene of *C. acetobutylicum* (Fig. 6A). It is obvious that transcription was transiently induced by a heat shock with maximum mRNA levels around 15 min after the temperature upshift. On Northern blots hybridized with *orfA*- and *dnaK*-specific probes, several bands were visible (Fig. 6B and C). With both probes, two identical RNA species of 5.0 and 3.8 kb each were detected; they represent transcripts of the total operon (*orfA*, 1,032 bp; *grpE*, 603 bp; *dnaK*, 1,848 bp) with and without the *dnaJ* gene, respectively, provided that the *dnaJ* gene in *C. acetobutylicum* has a similar size as in *E. coli* (1,128 bp) (7).

In addition, a transcript of 2.6 kb was found with the *dnaK*-specific probe, which might initiate in front of the *grpE* gene and then extend to the putative transcription terminator between *dnaK* and *dnaJ*, since the calculated size of such a transcript is in the same size range (around 2,570 bases). Other smaller transcripts evident on both blots are most likely specific degradation products, although additional defined transcription start sites cannot be excluded. The Northern blot experiments again proved that the mRNAs of these heat shock genes reached maximum levels around 15 min after the shock.

Determination of transcription start sites. Sequence analysis and Northern blot experiments of the *dnaK* gene region

```
C. acetobutylicum DnaK      51  GQVAKRQSITNPDKTIISIKRKMG**********************TAEKVA   80
                                :.:::::::::::  ::.:::::.::                      :. ::.
B. subtilis DnaK (22)       50  GEVAKRQSITNP.NTIMSIKRHMG**********************TDYKVE   78
                                :::::::.:::: :::.:.:::::                        ::.:::
B. megaterium DnaK (47)     50  GEVAKRQAITNP.NTIISVKRHMG**********************TDHKVE   78
                                :..:::::.::: ::.....:: .:                       .: .::
E. coli DnaK (6)            51  GQPAKRQAVTNPQNTLFAIKRLIGRRFQDEEVQRDVSIMPFKIIAADNGDAWVE  104
```

FIG. 4. Comparison of a part of the N-terminal region of the DnaK proteins from different bacteria. Numbers indicate the positions of the first and last amino acids shown. Missing, identical, and similar amino acids are represented by asterisks, double points, and single points, respectively.

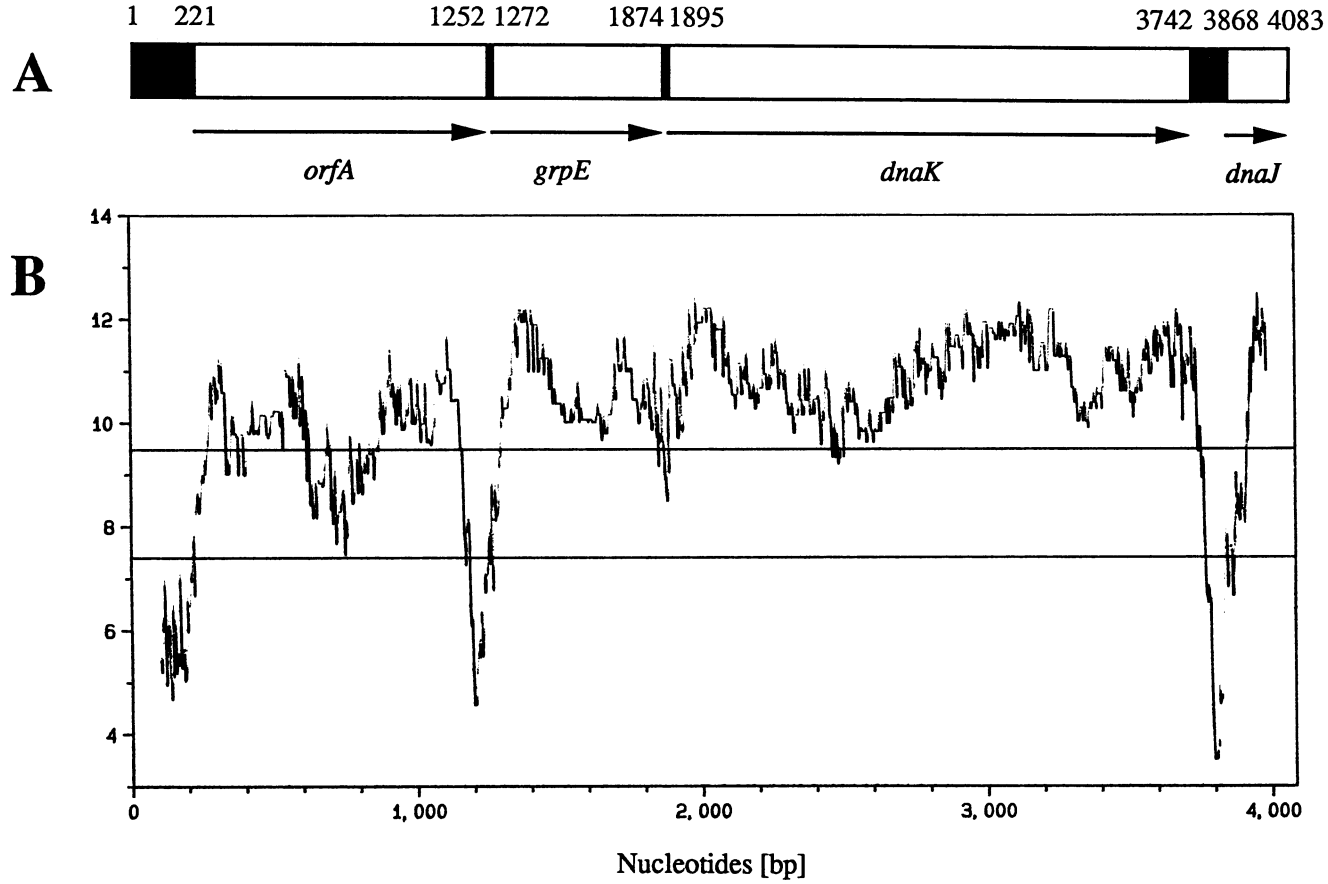
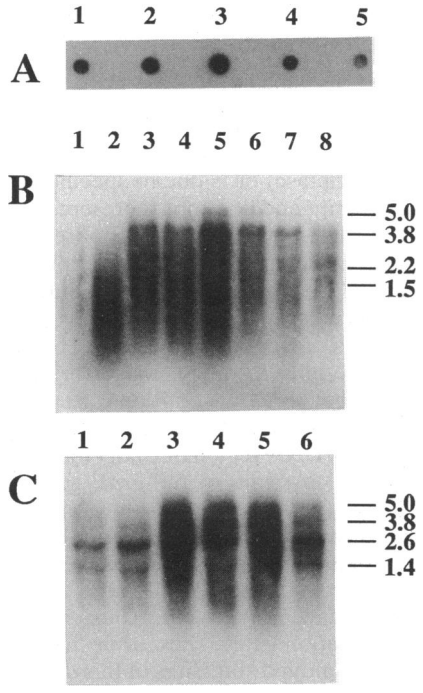


FIG. 5. Computer analysis for protein coding functions of ORFs in the *dnaK* gene region of *C. acetobutylicum*. (A) Schematic presentation of the *dnaK* gene region (compare with Fig. 1). (B) Results of a computer search for coding regions by using the program TestCode. The direction of screening was from left to right. The window was set at 200 bp to detect only longer coding regions. For further explanation, see the text.



of *C. acetobutylicum* suggested that transcription is initiated at least at two sites, upstream of *orfA* and of *grpE*. The transcription start points of the heat shock gene operon were determined by primer extension analysis using 17-mer synthetic oligonucleotide primers, complementary to the 5′ end of *orfA*, *grpE*, *dnaK*, and *dnaJ*. Total RNA used in these experiments was isolated from cells before and 7.5 min after a heat shock from 30 to 42°C. The results are shown in Fig. 7. Strong signals were obtained with RNA preparations from heat-shocked cells. A transcription start point (S1) was located 49 bases upstream of the ATG start codon of *orfA* with an A as the 5′ end of the mRNA. The deduced promoter

FIG. 6. Hybridization of total RNA of *C. acetobutylicum*. (A) Dot blot hybridization with the radioactively labeled insert of pKG1. Each spot contains 3 μg of RNA isolated before (lane 1) and 5 (lane 2), 15 (lane 3), 30 (lane 4), and 60 (lane 5) min after a heat shock from 30 to 42°C. (B) Northern hybridization with a radioactively labeled 0.85-kbp *Ava*II fragment of *orfA* as a probe. Each lane contains 5 μg of RNA before (lane 1) and 5 (lane 2), 7.5 (lane 3), 10 (lane 4), 15 (lane 5), 20 (lane 6), 30 (lane 7), and 60 (lane 8) min after a heat shock from 30 to 42°C. (C) Northern hybridization with the radioactively labeled insert of pKG1 as a probe. Each lane contains 8 μg of RNA before (lane 1) and 5 (lane 2), 10 (lane 3), 15 (lane 4), 30 (lane 5), and 60 (lane 6) min after a heat shock from 30 to 42°C. Fragment sizes (kilobase pairs) are indicated to the right of the gels.

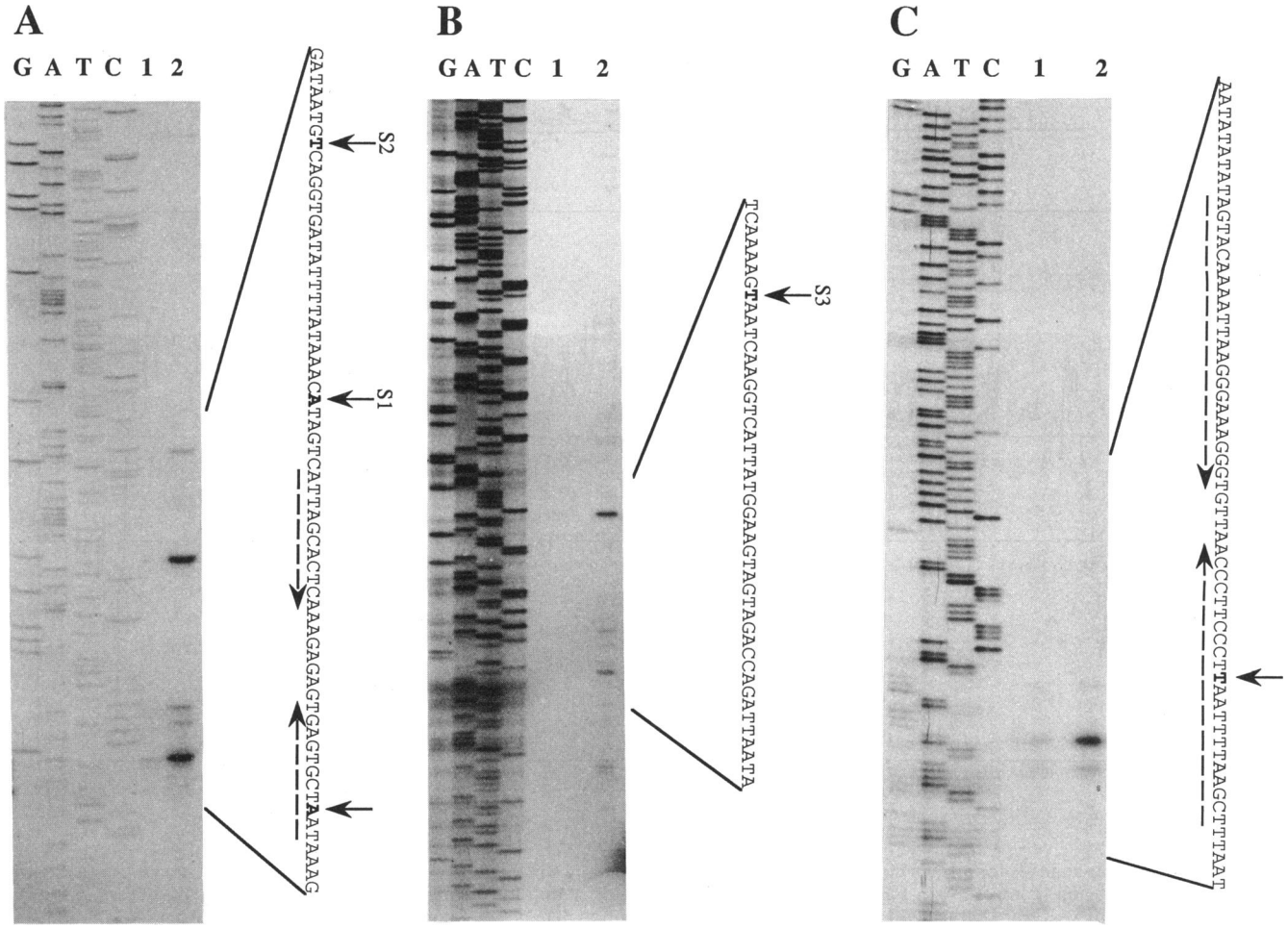


FIG. 7. Mapping of the 5′ end of mRNA from the *dnaK* gene region by primer extension analysis. ^{32}P-radiolabeled primers complementary to the 5′ end of *orfA* (A), *grpE* (B), and *dnaJ* (C) were hybridized to 10 μg each of total RNA from *C. acetobutylicum* isolated before (lanes 1) and 7.5 min (lanes 2) after a heat shock from 30 to 42°C. The primer extension products were analyzed on a sequence gel. G, A, T, and C are products of sequencing reactions with the same oligonucleotide as the primer. The sequences shown are antisense strands. Broken arrows mark regions of dyad symmetry. Arrows marked with S1, S2, and S3 point to bases representing the 5′ end of the respective mRNA.

sequence (P1) was identical with that identified above by its homology to the consensus for gram-positive bacteria. Twenty bases upstream of this major heat-inducible promoter, a minor band (S2) indicated a less frequently used initiation site. The corresponding promoter (P2) sequence showed less similarity to known promoter sequences. Another strong signal, located in the hairpin-loop structure immediately upstream of the ribosome-binding site, is probably the result of this secondary structure (interference with the reverse transcriptase) or is due to specific processing of the primary transcript and does not represent a transcription start point. The same phenomenon was observed upstream of the *dnaJ* gene, where the putative transcription terminator is located (Fig. 7C). Neither by the use of different reaction conditions nor special reverse transcriptases (Moloney murine leukemia virus or SuperScript RNase H$^-$ reverse transcriptase) were we able to obtain longer primer extension products. Thus, whether a transcription start point is located in front of *dnaJ* is not clear. However, a third transcription start site (S3) in this DNA region of *C. acetobutylicum* could be found 87 bases upstream of the *grpE* gene. Again, the deduced promoter (P3) showed only low similarity to promoters described so far. No transcription start site in front of *dnaK* could be found. Identical signals were also obtained in non-heat-shocked cells, however only after a long exposure time for the gels (data not shown). Thus, no evidence for additional transcription start sites after heat shock was found. The three identified start sites of this *dnaK* heat shock

	−35	−10	+1
Consensus	Ta AAAAA TTGACA a A a T TG TATAAT AAtAt		
P1	AAATAAGAAAAGTTGACAAAGATAATGTCAGGTGATATTTTATAAACA		
P2	GAAATAGACTAATTTTATGAAAATAAGAAAAGTTGACAAAGATAATGT		
P3	GGGAACTATAGGAATTATAGGACCGACAAGGATACCTTATTCAAAAGT		

FIG. 8. Comparison of the promoter regions P1, P2, and P3 deduced from the transcription start sites S1, S2, and S3, respectively (Fig. 7), with the consensus promoter of gram-positive bacteria (19). Putative −10 and −35 regions are boxed.

operon of *C. acetobutylicum* and the corresponding promoter sequences are shown in Fig. 8, as well as the consensus sequence of promoters in gram-positive bacteria.

DISCUSSION

This paper describes for the first time the DNA sequence of a heat shock operon of *C. acetobutylicum*, and the data presented provide one of the first examples of inducible gene expression in this organism. The *dnaK* locus of *C. acetobutylicum* contains four heat shock genes in the order *orfA-grpE-dnaK-dnaJ*. Our results make it evident that the chromosomal organization as well as the regulatory mechanism for the expression of these genes in *C. acetobutylicum* is different from that of the corresponding genes in *E. coli*.

In *E. coli*, *dnaK* and *dnaJ* are organized in a bicistronic operon (39), and the *grpE* gene is located in a different region of the chromosome (24). In addition, an *orfA*-homologous gene in *E. coli* has not been described so far, and we were unable to detect homologous sequences in the *E. coli* chromosome by Southern hybridization (31). Interestingly, in another gram-positive bacterium, *Bacillus subtilis*, an identical chromosomal organization, including the *orfA* homolog *orf39* (52), has been established. The *orfA*- and *orf39*-encoded proteins show 31.6% identity and 59.5% similarity, respectively, at the amino acid level (32). A heat shock protein with an *orfA*-corresponding molecular mass of 38 kDa has not been identified in *C. acetobutylicum* (37, 48). However, the clearly heat-inducible expression of the gene initiated at a major promoter of the *dnaK* locus indicates that the increased synthesis of the encoded protein is part of the heat shock response of *C. acetobutylicum*. The biological function of that protein is unknown, and it will be of interest to determine in which other bacteria a similar protein can be found.

An analysis of mRNA by Northern and dot blot experiments proved that the synthesis of the heat shock proteins OrfA, GrpE, DnaK, and DnaJ is induced at the level of transcription and that the genes are organized in an operon. The fact that the largest transcript (5 kb) is visible only as a faint band is in agreement with the putative transcription terminator described above between *dnaK* and *dnaJ*, which probably allows a certain level of readthrough. A similar structure which was held responsible for low levels of *dnaJ*-specific transcripts (7) has been found in the intergenic region between *dnaK* and *dnaJ* of *E. coli*. Furthermore, a Shine-Dalgarno sequence located unusually far from the AUG initiation codon might cause low translation of the *dnaJ* mRNA in *E. coli*. The unusual ribosome-binding site of the *C. acetobutylicum* gene (5′-AGGTGG) may have a similar effect.

Three transcription start sites in the *dnaK* gene region of *C. acetobutylicum* were identified by primer extension experiments. The similarity of the three deduced promoters, P1, P2, and P3, to consensus sequences for vegetative sigma factors in gram-positive bacteria and σ^{70} of *E. coli* and the absence of additional transcription start sites after heat shock suggest that an alternative sigma factor is not involved in the regulation of the heat shock response in *C. acetobutylicum* compared with *E. coli* (14, 20). Instead, an 11-bp inverted repeat between the transcription and translation start sites may be important for the expression of these genes, as has been speculated for a similar hairpin-loop in front of the 10-kDa antigen gene of *Mycobacterium tuberculosis* (4). The significance of this site is strengthened by the fact that similar structures with an identical sequence (5′-GCACTC) in the core of the stem have been found in front of the *groESL* operon of *C. acetobutylicum* (30) and *B. subtilis* (42); the *dnaK* locus of *B. subtilis* (52); and several (mostly heat shock) genes of *M. tuberculosis* (4, 44, 45), *M. bovis* (53), *M. leprae* (29), *Synechococcus* sp. (50), and *Synechocystis* sp. (13). However, this inverted repeat is not present upstream of the acetoacetate decarboxylase-encoding *adc* gene of *C. acetobutylicum*, the final enzyme in the acetone-forming pathway. Therefore, the proposed common regulatory unit for the heat shock response and solventogenesis in *C. acetobutylicum* (37) seems to be at another level.

The DnaK, DnaJ, and GrpE heat shock proteins of *E. coli* are negative regulators of heat shock gene expression by controlling the synthesis and stability of the heat shock sigma factor, σ^{32} (46, 49). Whether the corresponding proteins in *C. acetobutylicum* play a similar central role in the regulation of the heat shock response will be of interest. The DnaK protein of *C. acetobutylicum* has been purified and exhibits ATPase activity (5), which is in good agreement with the proposed ATP binding site near the N terminus (Fig. 3). However, in contrast to the enzyme of *E. coli* (58), autophosphorylation activity is absent, although a corresponding site is also present (Fig. 3). Whether the gap of 24 amino acids present at 74 amino acids from the N terminus in the DnaK protein of *C. acetobutylicum*, compared with the protein of *E. coli*, is involved in this phenomenon or whether it influences the activity of this protein in other aspects (e.g., regulation of the heat shock response) remains an interesting question. Again, the identical gap in DnaK proteins from other gram-positive bacteria points to a biological function of this structural feature.

ACKNOWLEDGMENTS

We thank J. Bardwell for providing plasmid pMK341, P. Dürre for stimulating discussions, and G. Gottschalk for continuous support. We are obliged to W. Schumann and M. Hecker for sharing their results on the *dnaK* locus of *B. subtilis* with us prior to publication.

This work was supported by a grant from the Deutsche Forschungs-gemeinschaft to H.B. and a fellowship of the Graduiertenförderung of the University of Göttingen to F.N.

REFERENCES

1. **Bahl, H., W. Andersch, K. Braun, and G. Gottschalk.** 1982. Effect of pH and butyrate concentration on the production of acetone and butanol by *Clostridium acetobutylicum* in continuous culture. Eur. J. Appl. Microbiol. Biotechnol. **14**:17–20.
2. **Bahl, H., and G. Gottschalk.** 1984. Parameters effecting solvent production by *Clostridium acetobutylicum* in continuous culture. Biotechnol. Bioeng. Symp. **11**:215–223.
3. **Bahl, H., and G. Gottschalk.** 1988. Microbial production of butanol/acetone, p. 1–30. *In* H. J. Rehm and G. Reed (ed.), Biotechnology, vol. 6b. VCH Verlagsgesellschaft mbH, Weinheim, Germany.
4. **Baird, P. N., L. M. C. Hall, and A. R. M. Coates.** 1989. Cloning and sequence analysis of the 10 kDa antigen gene of *Mycobacterium tuberculosis*. J. Gen. Microbiol. **135**:931–939.
5. **Balodimos, I. A., E. Rapaport, E. R. Kashket.** 1990. Protein phosphorylation in response to stress in *Clostridium acetobutylicum*. Appl. Environ. Microbiol. **56**:2170–2173.
6. **Bardwell, J. C. A., and E. A. Craig.** 1984. Major heat shock gene of *Drosophila* and the *Escherichia coli* heat-inducible *dnaK* gene are homologous. Proc. Natl. Acad. Sci. USA **81**:848–852.
7. **Bardwell, J. C. A., K. Tilly, E. Craig, J. King, M. Zylicz, and C. Georgopoulos.** 1986. The nucleotide sequence of the *Escherichia coli* K12 *dnaJ*$^+$ gene. J. Biol. Chem. **261**:1782–1785.
8. **Bertram, J., and P. Dürre.** 1989. Conjugal transfer and expression of streptococcal transposons in *Clostridium acetobutylicum*. Arch. Microbiol. **151**:551–557.

9. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
10. **Bittner, M., and D. Vapnek.** 1981. Versatile cloning vectors derived from the runaway-replication plasmid pKN402. Gene **15**:319–329.
11. **Bullock, W. O., J. M. Fernandez, and J. M. Stuart.** 1987. XL1-Blue: a high efficiency plasmid transforming *recA Escherichia coli* strain with beta-galactosidase selection. BioTechniques **5**:376–379.
12. **Cegielska, A., and C. Georgopoulos.** 1989. Functional domains of the *Escherichia coli* dnaK heat shock protein as revealed by mutational analysis. J. Biol. Chem. **264**:21122–21130.
13. **Chitnis, P. R., and N. Nelson.** 1991. Molecular cloning of the genes encoding two chaperone proteins of the cyanobacterium *Synechocystis* sp. PCC 6803. J. Biol. Chem. **266**:58–65.
14. **Cowing, D. W., J. C. A. Bardwell, E. A. Craig, C. Woolford, R. W. Hendrix, and C. A. Gross.** 1985. Consensus sequence for *Escherichia coli* heat shock gene promoters. Proc. Natl. Acad. Sci. USA **80**:2679–2683.
15. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
16. **Fickett, J. W.** 1982. Recognition of protein coding regions in DNA sequences. Nucleic Acids Res. **10**:5303–5318.
17. **Freier, S. M., R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Neilson, and D. H. Turner.** 1986. Improved free-energy parameters for predictions of RNA duplex stability. Proc. Natl. Acad. Sci. USA **83**:9373–9377.
18. **Gerischer, U., and P. Dürre.** 1990. Cloning, sequencing and molecular analysis of the acetoacetate decarboxylase gene region from *Clostridium acetobutylicum*. J. Bacteriol. **172**:6907–6918.
19. **Graves, M. C., and J. C. Rabinowitz.** 1986. *In vivo* and *in vitro* transcription of the *Clostridium pasteurianum* ferredoxin gene. J. Biol. Chem. **261**:11409–11415.
20. **Grossman, A. D., J. W. Erickson, and C. A. Gross.** 1984. The *htpR* gene product of *E. coli* is a sigma factor for heat-shock promoters. Cell **38**:383–390.
21. **Grunstein, M., and J. Wallis.** 1979. Colony hybridization. Methods Enzymol. **68**:379–389.
22. **Hearne, C. M., and P. J. Ellar.** 1989. Nucleotide sequence of a *Bacillus subtilis* gene homologous to the *dnaK* gene of *Escherichia coli*. Nucleic Acids Res. **17**:8373.
23. **Jones, D. T., and D. R. Woods.** 1986. Acetone-butanol fermentation revisited. Microbiol. Rev. **50**:484–524.
24. **Lipinska, B., J. King, D. Ang, and C. Georgopoulos.** 1988. Sequence analysis and transcriptional regulation of the *Escherichia coli grpE* gene, encoding a heat shock protein. Nucleic Acids Res. **16**:7545–7562.
25. **Long, S., D. T. Jones, and D. R. Woods.** 1984. Initiation of solvent production, clostridial stage, and endospore formation in *Clostridium acetobutylicum* P262. Appl. Environ. Microbiol. **20**:493–498.
26. **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 Res. **16**:1829–1836.
27. **Marmur, J.** 1960. A procedure for the isolation of deoxyribonucleic acid from microorganisms. J. Mol. Biol. **3**:208–218.
28. **McClure, W. R.** 1985. Mechanism and control of transcription initiation in prokaryotes. Annu. Rev. Biochem. **54**:171–204.
29. **Mehra, V., D. Sweetser, and R. Young.** 1986. Efficient mapping of protein antigenic determinants. Proc. Natl. Acad. Sci. USA **83**:7013–7017.
30. **Narberhaus, F., and H. Bahl.** 1992. Cloning, sequencing, and molecular analysis of the *groESL* operon of *Clostridium acetobutylicum*. J. Bacteriol. **174**:3282–3289.
31. **Narberhaus, F., and H. Bahl.** Unpublished results.
32. **Narberhaus, F., M. Wetzstein, W. Schumann, and H. Bahl.** Unpublished results.
33. **O'Brien, R. W., and J. G. Morris.** 1971. Oxygen and growth and metabolism of *Clostridium acetobutylicum*. J. Gen. Microbiol. **68**:307–318.
34. **Oelmüller, U., N. Krüger, A. Steinbüchel, and C. Friedrich.** 1990. Isolation of prokaryotic RNA and detection of specific mRNA with biotinylated probes. J. Microbiol. Methods **11**:73–81.
35. **Petersen, D. J., and G. N. Bennett.** 1990. Purification of acetoacetate decarboxylase from *Clostridium acetobutylicum* ATCC 824 and cloning of the acetoacetate decarboxylase gene in *Escherichia coli*. Appl. Environ. Microbiol. **56**:3491–3498.
36. **Petersen, D. J., R. W. Welch, F. B. Rudolph, and G. N. Bennett.** 1991. Molecular cloning of an alcohol (butanol) dehydrogenase gene cluster from *Clostridium acetobutylicum*. J. Bacteriol. **173**:1831–1834.
37. **Pich, A., F. Narberhaus, and H. Bahl.** 1990. Induction of heat shock proteins during solvent formation in *Clostridium acetobutylicum*. Appl. Microbiol. Biotechnol. **33**:697–704.
38. **Platt, T.** 1986. Transcription termination and the regulation of gene expression. Annu. Rev. Biochem. **55**:339–372.
39. **Saito, H., and H. Uchida.** 1978. Organization and expression of the *dnaJ* and *dnaK* genes of *Escherichia coli* K12. Mol. Gen. Genet. **164**:1–8.
40. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
41. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
42. **Schumann, W.** Personal communication.
43. **Shine, J., and L. Dalgarno.** 1974. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementary to nonsense triplets and ribosome-binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
44. **Shinnick, T. M.** 1987. The 65-kilodalton antigen of *Mycobacterium tuberculosis*. J. Bacteriol. **169**:1080–1088.
45. **Shinnick, T. M., B. B. Plikaytis, A. D. Hyche, R. M. van Landingham, and L. L. Walker.** 1989. The *Mycobacterium tuberculosis* BCG—a protein has homology with the *Escherichia coli* GroES protein. Nucleic Acids Res. **17**:1254.
46. **Straus, D., W. Walker, and C. A. Gross.** 1990. DnaK, DnaJ, and GrpE heat shock proteins negatively regulate heat shock gene expression by controlling the synthesis and stability of σ^{32}. Genes Dev. **4**:2202–2209.
47. **Sussman, M. D., and P. Setlow.** 1987. Nucleotide sequence of a *Bacillus megaterium* gene homologous to the *dnaK* gene of *Escherichia coli*. Nucleic Acids Res. **15**:3923.
48. **Terracciano, J. S., E. Rapaport, and E. R. Kashket.** 1988. Stress and growth-phase associated proteins of *Clostridium acetobutylicum*. Appl. Environ. Microbiol. **54**:1989–1995.
49. **Tilly, K., N. McKittrick, M. Zylicz, and C. Georgopoulos.** 1983. The *dnaK* protein modulates the heat shock response of *Escherichia coli*. Cell **34**:641–646.
50. **Webb, R., K. J. Reddy, and L. A. Sherman.** 1990. Regulation and sequence of the *Synechocystis* sp. strain PCC 7942 *groESL* operon, encoding a cyanobacterial chaperonin. J. Bacteriol. **172**:5079–5088.
51. **Wetzstein, M., and W. Schumann.** 1990. Nucleotide sequence of a *Bacillus subtilis* gene homologous to the *grpE* gene of *E. coli* located immediately upstream of the *dnaK* gene. Nucleic Acids Res. **18**:1289.
52. **Wetzstein, M., U. Völker, J. Dedio, S. Löbau, U. Zuber, M. Schiesswohl, C. Herget, M. Hecker, and W. Schumann.** 1992. Cloning, sequencing and molecular analysis of the *dnaK* locus from *Bacillus subtilis*. J. Bacteriol. **174**:3300–3310.
53. **Yamaguchi, R., K. Matsuo, A. Yamazaki, S. Nagai, K. Terasaka, and T. Yamada.** 1988. Immunogenic protein MPB57 from *Mycobacterium bovis* BCG: molecular cloning, nucleotide sequence and expression. FEBS Lett. **240**:115–117.
54. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. Gene **33**:103–119.
55. **Youngleson, J. S., D. T. Jones, and D. R. Woods.** 1989. Homology between hydroxybutyryl and hydroxyacyl coenzyme A

dehydrogenase enzymes from *Clostridium acetobutylicum* fermentation and vertebrate fatty acid β-oxidation pathways. J. Bacteriol. **171**:6800–6807.

56. **Youngleson, J. S., W. A. Jones, D. T. Jones, and D. R. Woods.** 1989. Molecular analysis and nucleotide sequence of the *adh1* gene encoding an NADPH-dependent butanol dehydrogenase in the gram-positive anaerobe *Clostridium acetobutylicum*. Gene **78**:355–364.
57. **Zuker, M., and P. Stiegler.** 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. **9**:133–148.
58. **Zylicz, M. J., H. LeBowitz, R. McMacken, and C. Georgopoulos.** 1983. The dnaK protein of *Escherichia coli* posesses an ATPase and autophosphorylating activity and is essential in an *in vitro* DNA replication system. Proc. Natl. Acad. Sci. USA **80**:6431–6435.