JOURNAL OF BACTERIOLOGY, Apr. 1993, p. 2465–2469
0021-9193/93/082465-05$02.00/0

Vol. 175, No. 8

Molecular Cloning, Sequencing, and Transcriptional Analysis of the *groESL* Operon from *Bacillus stearothermophilus*

ULRIKE SCHÖN AND WOLFGANG SCHUMANN*

Lehrstuhl für Genetik, Universität Bayreuth, W-8580 Bayreuth, Germany

Received 30 November 1992/Accepted 16 February 1993

Using a gene probe of the *Bacillus subtilis groEL* gene, a 7.3-kb *Hin*dIII fragment of chromosomal DNA of *Bacillus stearothermophilus* was cloned. Sequencing of 2,309 bp led to the detection of two open reading frames in the order *groES groEL*. Primer extension studies revealed one potential transcription start site preceding the *groESL* operon, which was activated upon temperature upshift. Northern blot (RNA) analysis resolved two mRNA species with lengths of 2.2 and 1.5 kb; RNA slot-blot experiments revealed an at least 10-fold increase in the amount of specific mRNA from 0 to 7 min postinduction followed by a decrease. The 9-bp inverted repeat characteristic of many gram-positive bacteria was found within the 5′ leader region of the mRNA. The *groESL* operon of *B. stearothermophilus* could complement *E. coli groES*(Ts) and *groEL*(Ts) mutants for growth at high temperature and for propagation of phage lambda.

Induction of heat shock protein synthesis in response to environmental stress is a universal phenomenon that occurs in most if not all organisms, and the primary sequences of some heat shock proteins are well conserved throughout evolution, from bacteria to humans (18). Major heat shock proteins GroES and GroEL of *Escherichia coli* promote assembly (6, 13, 15), disassembly (4, 17), or translocation (7, 8) of other proteins, presumably by preventing their misfolding and thereby sponsoring the formation of biologically functional structures. These kinds of proteins have been called "molecular chaperones" (9, 10, 15, 23).

GroEL (also called Cpn60) has been found in prokaryotes, mitochondria (Hsp60), and chloroplasts (ribulose-1,5-bisphosphate carboxylase subunit binding protein [13, 15]). GroES (Cpn10) has been found in prokaryotes (1, 2), and a functional equivalent was recently identified in mitochondria (20). *E. coli* GroES and GroEL are essential for morphogenesis of several coliphages (12, 29, 30, 38), for the assembly of ribulose bisphosphate carboxylase-oxygenase subunits from photosynthetic bacteria in *E. coli* (13), and for export of plasmid-encoded β-lactamase (16). Recently, the GroES and GroEL homologs from *Thermus thermophilus* have been copurified as a large complex (28). They can promote in vitro refolding of several guanidine hydrochloride-denatured enzymes from thermophilic bacteria at temperatures above 60°C. Furthermore, the crystallization of the GroES-GroEL complex from *T. thermophilus* has been reported (19).

To study the function of chaperones from a thermophilic organism in processes such as in vivo assembly, folding of proteins, and secretion of exoenzymes in comparison with the same chaperones from a related mesophilic bacterium, we have initiated a program to clone and to characterize genes coding for molecular chaperones from *Bacillus stearothermophilus*, in which temperature upshift is accompanied by the synthesis of at least four heat shock proteins (35). The biochemical properties of the chaperones at different temperatures will be compared with those of *Bacillus subtilis*, whose genes have been recently cloned by our group (27). In this paper, we report the nucleotide sequence of 2,309 kb of DNA which contains the complete *groESL* operon. Furthermore, we analyzed transcription of this operon and showed that both genes can complement temperature-sensitive mutants of *E. coli groES* and *groEL*, respectively, for growth at high temperature and for propagation of phage lambda.

Molecular cloning of the *groESL* locus. Southern blot experiments revealed an internal part of the *B. subtilis* (24) *groEL* gene hybridizing to a 7.3-kb *Hin*dIII fragment of *B. stearothermophilus* NUB36 (5) chromosomal DNA (data not shown). This fragment was ligated into pACYC177 (3) and transformed into *E. coli* RR1 (25). Transformants were analyzed by probing recombinant plasmids with a 1.0-kb *Hin*dIII-*Pst*I fragment of *B. subtilis groEL* gene recovered from pASG58 (27). DNA fragments were labeled with digoxigenin-[11]-dUTP (Boehringer GmbH, Mannheim, Germany) by the random-priming technique of Feinberg and Vogelstein (11, 37). One of these plasmids yielding a clear hybridization signal was chosen for further analyses (pUS01). In parallel, the *Hin*dIII fragment was inserted into pUC19 (21), resulting in pUS05, which could be stably maintained in different *E. coli* strains. Next, a restriction map of the 7.3-kb insert was established, and the *groEL* sequence could be located near one end of the *Hin*dIII fragment.

Sequencing and features of the *groE* locus. By using pUS01 and subclones in pUC18/19 the sequence of the *groE* locus was determined by the dideoxy-chain termination method (26) with [α-^{35}S]dATP (specific activity, 650 Ci/mmol; Amersham Corp.) and the Sequenase kit (version 2.0) of U.S. Biochemical Corp. Synthetic primers were used for double-stranded DNA sequencing (14) with plasmid DNA as the template. A total of 2,309 bp was sequenced, revealing the presence of two contiguous open reading frames (ORFs) (Fig. 1).

ORF1 is 282 nucleotides long, beginning with a GUG start codon at position 249 and ending with two stop codons at position 531. Alternative start codons GUG, UUG, and AUU are frequently used in prokaryotic genes; all are single-base changes from AUG (22). ORF1 encodes a polypeptide of 94 amino acid residues with a predicted molecular mass of 10,209 Da and a calculated pI of 4.64. This polypeptide exhibits significant sequence similarity to the GroES of *B. subtilis* (74% identical and 12% conservative replacements), *E. coli* (43% identical and 21% conservative

* Corresponding author.

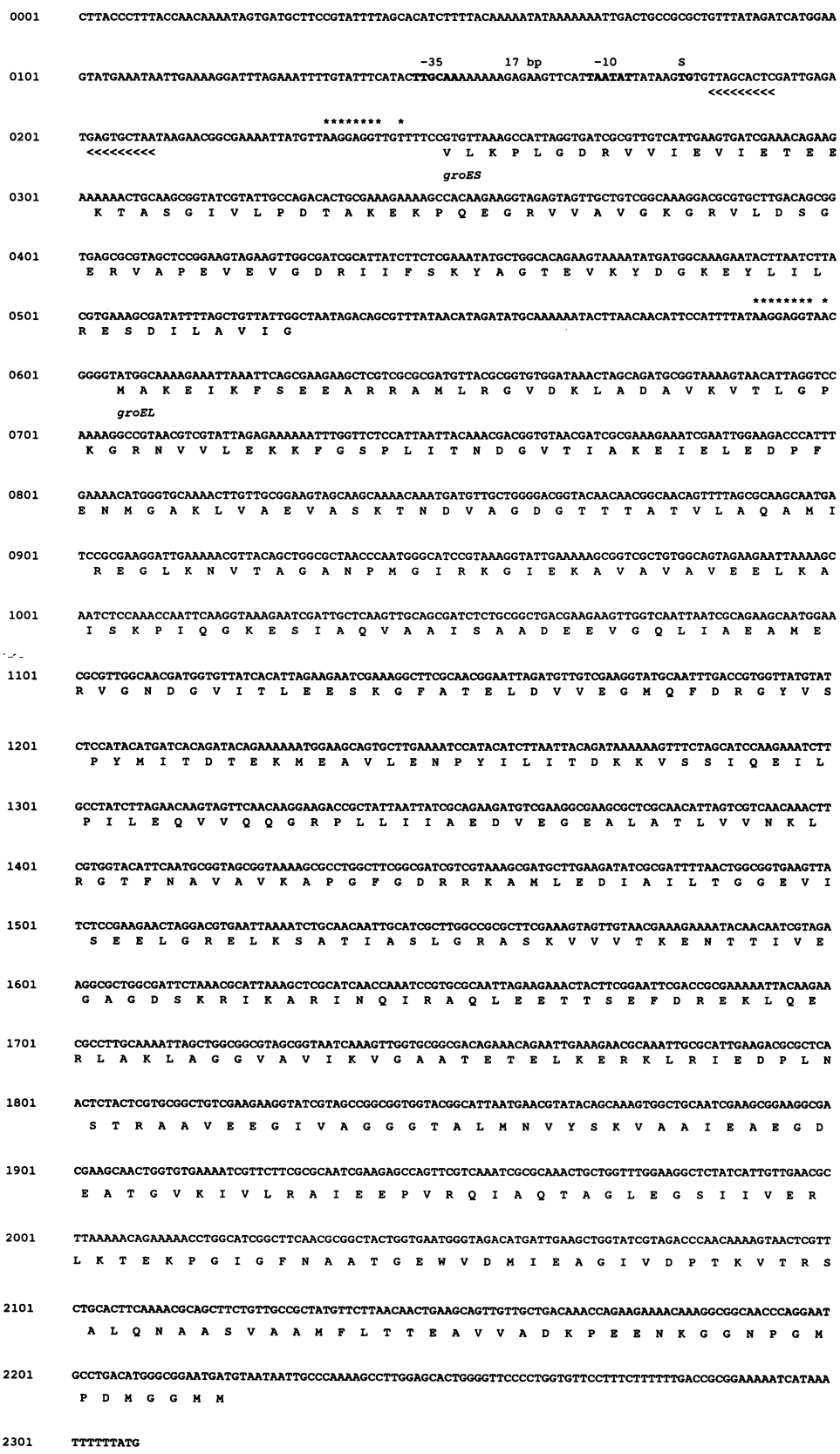

FIG. 1. Nucleotide sequence of the *groESL* operon. Nucleotides are numbered from the 5′ end of the nontranscribed strand, and the predicted amino acid sequences for the *groES* and *groEL* genes are given in the single-letter code below the DNA sequence. Indicated are the putative Shine-Dalgarno sequences (asterisks), one potential vegetative promoter (boldface letters), and one inverted repeat (arrowheads). The 5′ end of mRNA (S), as identified by primer extension, is marked by a boldface letter.

replacements), and other bacterial species (data not shown). Therefore, we designated ORF1 as *groES*.

ORF2 is 1,617 nucleotides long, beginning with an AUG start codon at position 606 and ending with two UAA stop codons at position 2223. The start codon of ORF2 is 70 nucleotides downstream from the stop codons of ORF1. ORF2 encodes a polypeptide of 539 amino acid residues with a predicted molecular mass of 57,244 Da, a calculated pI of 4.75, and one Gly-Gly-Met (GGM) motif at its C-terminus. This motif has been found in GroEL proteins of various species, but its biological function remains unclear. Recently, the DNA sequence of the *groESL* genes from the thermophilic bacterium PS3 was reported (30). Most interestingly, the GroEL protein encoded by PS3 also ends with only one GGM motif. It is tempting to speculate that the number of GGM motifs at the C-terminus of GroEL will prove to be correlated with the thermostability of the protein. Because the polypeptide encoded by ORF2 exhibits significant sequence similarity with GroEL of *B. subtilis* (86% identical and 7% conservative replacements), of *E. coli* (61% identical and 18% conservative replacements), and other bacterial species (data not shown), ORF2 was designated *groEL*.

Putative ribosome-binding sites, marked with asterisks in Fig. 1, are 6 and 7 nucleotides upstream from the start codons of ORF1 and ORF2, respectively. These sites consist of 9 nucleotides each complementary to the 3′ end of *B. stearothermophilus* 16S rRNA (32).

Transcriptional analysis. The start of transcription of the *groESL* operon was determined by primer extension analysis of RNA isolated from cells grown at 55°C or after heat shock at 70°C. Synthetic oligonucleotides ON1 (5′-CTTGC TACTTCCGCAA-3′) and ON2 (5′-CTCTACCTTCTTGTG GC-3′) were used as primers. These oligonucleotides are complementary to the non-coding strand at the beginning of *groES* (ON2) and *groEL* (ON1). Primer extension experiments revealed one potential transcription start point located 69 nucleotides upstream of the translation initiation codon of *groES* (designated S; Fig. 2A and Fig. 1). This position corresponds to a T residue. Upon longer exposure, we observed an additional weak band immediately below the strong band corresponding to a G residue (Fig. 2A). The same results have been obtained with the *groESL* operon of *B. subtilis* (27), underlining the close transcriptional organization of these two operons. The number of transcripts initiated at S increased with the temperature shift from 55 to 70°C (Fig. 2A).

The transcriptional start point S is preceded by a potential *B. stearothermophilus* vegetative cell promoter (34) with two mismatches each in the −35 and −10 regions which are separated by a 17-bp spacer (Fig. 1, bp 145 to 173). Therefore, heat-sensitive transcription from S might also be initiated by the vegetative sigma factor. Furthermore, we found a palindromic structure downstream of S consisting of a 9-bp inverted repeat (IR) separated by a 9-bp spacer (Fig. 1, nucleotides 184 to 210). This palindromic structure has already been found preceding the *groESL* and *dnaK* operons of *B. subtilis, Clostridium acetobutylicum*, and several other bacterial species (see Table 2 in reference 33). Up to now, this IR was only found in front of the genes of the *dnaK* or *groE* operons. Will this IR be restricted only to genes encoding these two molecular chaperones? Recently, Wu and Welker published results on the cloning of two genes from *B. stearothermophilus* exhibiting homology with the *glnQH* genes of *E. coli* (34). Expression of these genes was activated by temperature upshift, but they are not preceded

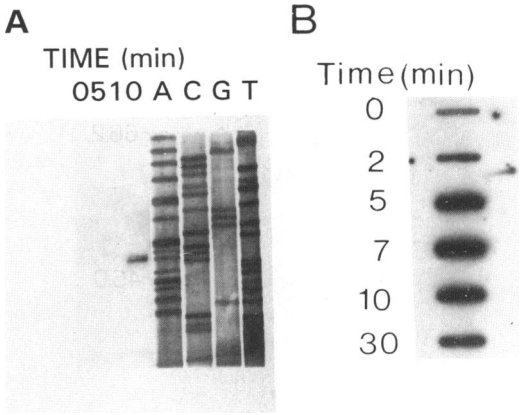

FIG. 2. Transcriptional analyses of the *groESL* operon. (A) Mapping the 5′ end of the *groESL* operon message by primer extension. The autoradiogram of a 6% polyacrylamide sequencing gel used to analyze the primer extension products is shown. Equal amounts of RNA isolated before (0 min) or at 5 and 10 min after heat shock (70°C) were hybridized with primer ON2 complementary to the 5′ region of *groES* and extended with avian myeloblastosis virus reverse transcriptase. A, C, G, and T indicate which dideoxynucleotide was used to terminate the reaction. (B) Concentration of *groE* mRNA. Slot-blot analyses of RNA isolated before (0 min) and at different times after heat shock to 70°C (2, 5, 7, 10, and 30 min) are shown. The digoxigenin-[11]-dUTP-labeled 582-bp internal *groEL* *Cla*I-*Eco*RI fragment was used as probe.

by the IR. Recent experiments revealed that introduction of point mutations within the IR preceding the *dnaK* operon of *B. subtilis* leads to a constitutively high level of expression of that operon (36), suggesting that a repressor is involved in the regulation of the heat shock response.

The in vivo transcripts of the *groE* locus were detected by slot-blot and Northern (RNA) blot analyses with total RNA prepared either from exponentially growing *B. stearothermophilus* cells at 55°C or at different times after the bacteria had been transferred to 70°C. The slot-blot experiments revealed a rapid increase in the mRNA level within 5 min after temperature upshift followed by a decline between 7 and 10 min postinduction (Fig. 2B). These results show that induction of the *groESL* operon is regulated at the level of transcription. This can be accomplished either by increased synthesis of mRNA, by a change in its half-life, or by a mixture of both.

Northern analysis was performed with an internal fragment of *groESL* as probe which hybridized to two mRNA species with lengths of 2.2 and 1.5 kb (data not shown). By length, the 2.2-kb RNA could comprise the whole operon. The smaller mRNA species could arise either by premature termination of transcription or by processing of a larger species. To prove whether *groES* and *groEL* form one operon, hybridization was repeated with ON1 and ON2. Both oligonucleotides hybridized to the 2.2- and 1.5-kb mRNAs, respectively (data not shown), thereby proving that both genes constitute one operon.

Identification of the GroEL protein. In an attempt to identify the GroEL protein of *B. stearothermophilus*, extracts were analyzed from *B. stearothermophilus* cells grown at 55°C and prepared 30 min after temperature upshift to 70°C by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (4.5% polyacrylamide). In parallel, extracts from *E. coli* RR1 strains carrying either pACYC177 or pUS01 were analyzed. With *B. stearothermophilus*, a

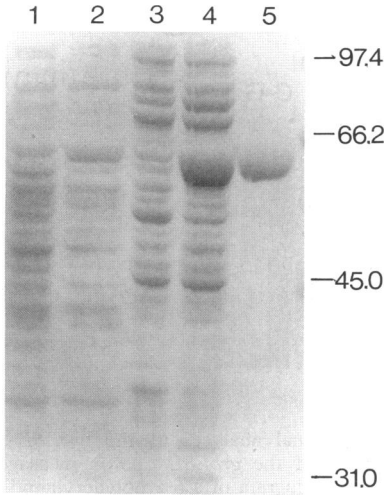

FIG. 3. SDS-polyacrylamide (4.5%) gel analysis of proteins from different bacterial strains stained with Coomassie blue. Each well contained 25 μg of protein. Lanes are as follows: 1, *B. stearothermophilus* NUB36 at 55°C; 2, NUB36 treated for 30 min at 70°C; 3, *E. coli* RR1, containing pACYC177 at 37°C; 4, *E. coli* RR1, containing pUS01 at 37°C; 5, extract of *E. coli* RR1 carrying pUS01 treated for 30 min at 65°C prior to analysis. Positions of molecular size markers are indicated in kilodaltons.

protein band with a molecular mass of about 60 kDa increased in amount after temperature upshift (Fig. 3, compare lanes 1 and 2). The same band can be seen in extracts prepared from *E. coli* containing pUS01, but not from the strain carrying the vector alone (compare lanes 3 and 4), demonstrating that the *groEL* gene of *B. stearothermophilus* is efficiently expressed in *E. coli*. If the *E. coli* extract containing GroEL of *B. stearothermophilus* was heated to 65°C for 30 min, most proteins aggregated and could be removed by centrifugation. Analysis of the clear supernatant by SDS-PAGE revealed that most of GroEL protein was not precipitated by the heating step (lane 5). This behavior of the GroEL protein might be useful during purification of this protein.

The *groESL* operon of *B. stearothermophilus* can complement defects in *E. coli groES* and *groEL*. Because at least *groEL* of *B. stearothermophilus* was expressed in *E. coli*, we asked whether the *groESL* operon can complement the *groE* mutants of *E. coli*. *E. coli* SKB178S (*groES42*) and SKB178L (*groEL44*) carry temperature-sensitive mutations in *groES* and *groEL*, respectively (31). They both cannot grow at 42°C and fail to plate phage lambda. Plasmid pUS01 was introduced into both strains, and transformants were analyzed for complementation of both phenotypes. Both recombinant strains were able to form colonies at high temperatures and to propagate phage lambda at all temperatures, indicating efficient complementation of the defects conferred by the mutations (data not shown).

We conclude from this result that (i) the *groES* and *groEL* genes of *B. stearothermophilus* are expressed in *E. coli*, and (ii) these proteins are active in this host, demonstrating functional homology to known GroE proteins. Therefore, besides the structural similarity, the GroE proteins of *B. stearothermophilus* and of *E. coli* seem to be functionally interchangeable. We envisage three different mechanisms to obtain complementation: (i) formation by the GroES and GroEL proteins of *B. stearothermophilus* of an active complex to substitute for the *E. coli* nonfunctional GroES-GroEL complex; (ii) formation of mixed complexes, e.g., active GroES of *E. coli* (in the *groELts* mutant) and GroEL of *B. stearothermophilus*; and (iii) change by the active GroES-GroEL complex of *B. stearothermophilus* of the inactive GroES-GroEL complex of *E. coli* into an active one. Experiments are in progress to distinguish between these possibilities.

Nucleotide sequence accession number. The DNA sequence of the *B. stearothermophilus groESL* operon was submitted to GenBank and assigned the accession number L10132.

We thank Debbie Ang and Neil Welker for providing strains.

This work was supported by grants from the BMFT to W.S. and by a predoctoral fellowship from the DFG to U.S.

REFERENCES

1. **Baird, P. N., L. M. C. Hall, and A. R. M. Coates.** 1988. A major antigen from *Mycobacterium tuberculosis* which is homologous to the heat shock proteins *groES* from *E. coli* and the *htpA* gene product of *Coxiella burnetii*. Nucleic Acids Res. **16:**9047.
2. **Chandrasekhar, G. N., K. Tilly, C. Woolford, R. Hendrix, and C. Georgopoulos.** 1986. Purification and properties of the *groES* morphogenetic protein of *Escherichia coli*. J. Biol. Chem. **261:**12414–12419.
3. **Chang, A. C. Y., and S. N. Cohen.** 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J. Bacteriol. **134:** 1141–1156.
4. **Chappell, T. G., W. J. Welch, D. M. Schlossman, K. B. Palter, M. J. Schlesinger, and J. E. Rothman.** 1986. Uncoating ATPase is a member of the 70 kilodalton family of stress proteins. Cell **45:**3–13.
5. **Chen, Z., S. F. Wojcik, and N. E. Welker.** 1986. Genetic analysis of *Bacillus stearothermophilus* by protoplast fusion. J. Bacteriol. **165:**994–1001.
6. **Cheng, M. Y., F.-U. Hartl, J. Martin, R. A. Pollock, F. Kalousek, W. Neupert, E. M. Hallberg, R. L. Hallberg, and A. L. Horwich.** 1989. Mitochondrial heat-shock protein hsp60 is essential for assembly of proteins imported into yeast mitochondria. Nature (London) **337:**620–625.
7. **Chirico, W. J., M. G. Waters, and G. Blobel.** 1988. 70K heat shock related proteins stimulate protein translocation into microsomes. Nature (London) **332:**805–810.
8. **Deshaies, R. J., B. D. Koch, M. Werner-Washburne, E. A. Craig, and R. Schekman.** 1988. A subfamily of stress proteins facilitates translocation of secretory and mitochondrial precursor polypeptides. Nature (London) **332:**800–805.
9. **Ellis, J.** 1987. Proteins as molecular chaperones. Nature (London) **328:**378–379.
10. **Ellis, R. J., and S. M. Hemmingsen.** 1989. Molecular chaperones: proteins essential for the biogenesis of some macromolecular structures. Trends Biochem. Sci. **14:**339–342.
11. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:**6–13.
12. **Georgopoulos, C. P., R. W. Hendrix, A. D. Kaiser, and W. B. Wood.** 1972. Role of the host cell in bacteriophage morphogenesis: effects of a bacterial mutation on T4 head assembly. Nature (London) New Biol. **239:**38–42.
13. **Goloubinoff, P., A. A. Gatenby, and G. H. Lorimer.** 1989. GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in *Escherichia coli*. Nature (London) **337:**44–47.
14. **Heinrich, P.** 1986. Guidelines for quick and simple plasmid sequencing. Boehringer GmbH, Mannheim, Germany.
15. **Hemmingsen, S. M., C. Woolford, S. M. van der Vies, K. Tilly, D. T. Dennis, C. P. Georgopoulos, R. W. Hendrix, and R. J. Ellis.** 1988. Homologous plant and bacterial proteins chaperone oligomeric protein assembly. Nature (London) **333:**330–334.
16. **Kusukawa, N., T. Yura, C. Ueguchi, Y. Akiyama, and K. Ito.** 1989. Effects of mutations in heat-shock genes *groES* and *groEL*

on protein export in *Escherichia coli*. EMBO J. **8**:3517–3521.
17. **Liberek, K., C. Georgopoulos, and M. Zylicz.** 1988. Role of the *Escherichia coli* DnaK and DnaJ heat shock proteins in the initiation of bacteriophage lambda DNA replication. Proc. Natl. Acad. Sci. USA **85**:6632–6636.
18. **Lindquist, S., and E. A. Craig.** 1988. The heat-shock proteins. Annu. Rev. Genet. **22**:631–677.
19. **Lissin, N. M., S. E. Sedelnikova, and S. N. Ryazantsev.** 1992. Crystallization of the cpn60/cpn10 complex ('holo-chaperonin') from *Thermus thermophilus*. FEBS Lett. **311**:22–24.
20. **Lubben, T. H., A. A. Gatenby, G. K. Donaldson, G. H. Lorimer, and P. V. Viitanen.** 1990. Identification of a *groES*-like chaperonin in mitochondria that facilitates protein folding. Proc. Natl. Acad. Sci. USA **87**:7683–7687.
21. **Norrander, J., T. Kempe, and J. Messing.** 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene **26**:101–106.
22. **Parker, J.** 1989. Errors and alternatives in reading the universal genetic code. Microbiol. Rev. **55**:273–298.
23. **Pelham, H. R. B.** 1986. Speculations on the functions of the major heat shock and glucose-regulated proteins. Cell **46**:959–961.
24. **Puyet, A., H. Sandoval, P. López, A. Aguilar, J. F. Martin, and M. Espinosa.** 1987. A simple medium for rapid regeneration of *Bacillus subtilis* protoplasts transformed with plasmid DNA. FEMS Microbiol. Lett. **40**:1–5.
25. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
27. **Schmidt, A., M. Schiesswohl, U. Völker, M. Hecker, and W. Schumann.** 1992. Cloning, sequencing, mapping, and transcriptional analysis of the *groESL* operon from *Bacillus subtilis*. J. Bacteriol. **174**:3993–3999.
28. **Taguchi, H., J. Konishi, N. Ishii, and M. Yoshida.** 1991. A chaperonin from a thermophilic bacterium, *Thermus thermophilus*, that controls refoldings of several thermophilic enzymes. J. Biol. Chem. **266**:22411–22418.
29. **Takano, T., and T. Kakefuda.** 1972. Involvement of a bacterial factor in morphogenesis of bacteriophage capsid. Nature (London) New Biol. **239**:34–37.
30. **Tamada, H., T. Ohta, T. Hamamoto, Y. Otawara-Hamamoto, M. Yanagi, H. Hiraiwa, H. Hirata, and Y. Kagawa.** 1991. Gene structure of heat shock proteins 61 KDa and 12 KDa (thermophilic chaperonins) of thermophilic bacterium PS3. Biochem. Biophys. Res. Commun. **179**:565–571.
31. **Tilly, K., and C. Georgopoulos.** 1982. Evidence that the two *Escherichia coli groE* morphogenetic gene products interact in vivo. J. Bacteriol. **149**:1082–1088.
32. **Van Charldorp, R., A. M. A. van Kimmenade, and P. H. van Knippenberg.** 1981. Sequence and secondary structure of the colicin fragment of *Bacillus stearothermophilus* 16S ribosomal RNA. Nucleic Acids Res. **9**:4909–4917.
33. **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.
34. **Wu, L., and N. E. Welker.** 1991. Cloning and characterization of a glutamine transport operon of *Bacillus stearothermophilus* NUB36: effect of temperature on regulation of transcription. J. Bacteriol. **173**:4877–4888.
35. **Wu, L., and N. E. Welker.** 1991. Temperature-induced protein synthesis in *Bacillus stearothermophilus* NUB36. J. Bacteriol. **173**:4889–4892.
36. **Zuber, U., and W. Schumann.** Unpublished data.
37. **Zuber, U., and W. Schumann.** 1991. Tn*5cos*: a useful transposon for restriction mapping of large plasmids using phage lambda terminase. Gene **103**:69–72.
38. **Zweig, M., and D. J. Cummings.** 1973. Cleavage of head and tail proteins during bacteriophage T5 assembly: selective host involvement in the cleavage of a tail protein. J. Mol. Biol. **80**:505–518.