

Complete nucleotide sequence of the *Bacillus subtilis* *dnaK* gene

Michael Wetzstein, Jürgen Dedio and Wolfgang Schumann
Lehrstuhl für Genetik, Universität Bayreuth, D-8580 Bayreuth, FRG

Submitted March 23, 1990

EMBL accession no. X52064

Recently, we published the isolation and sequencing of a chromosomal fragment of *B. subtilis* DNA containing an open reading frame with significant homology to the *E. coli* *grpE* protein (1). Another open reading frame encoding the DnaK protein is located 23 bp downstream of the *grpE* gene. Using the method of *in vivo* chromosome walking, flanking sequences were isolated, and the complete sequence of the *dnaK* gene has been established. The deduced amino acid sequence showed 94% homology to the DnaK protein of *B. megaterium* (2) and 71% to that of *E. coli* (3). The *dnaK* gene is followed by a 90 bp region containing a putative terminator ($\Delta G = -13$ kcal), a sigma-A dependent promoter and the beginning of another open reading frame. The deduced amino acid sequence of this second gene showed high homology to the DnaJ protein of *E. coli* (4). Start and stop codons of the open reading frames are underlined,

a potential sigma-A promoter is boxed and a region of dyad symmetry indicated by opposing arrows.

ACKNOWLEDGEMENT

This research project was supported by the BMFT through Grant No. PBE/II/0318842A.

REFERENCES

1. Wetzstein, M. and Schumann, W. (1990) *Nucl. Acids Res.* **18**, 1289.
2. Sussman, M.D. and Setlow, P. (1987) *Nucl. Acids Res.* **15**, 3923.
3. Bardwell, J.C.A. and Craig, E. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 848–852.
4. Bardwell, J.C.A., Tilly, K., Craig, E., King, J., Zyllicz, M. and Georgopoulos, C. (1986) *J. Biol. Chem.* **261**, 1782–1785.

```

1  AATCAATAAT TACATAGCAG GAGGTTATTC AAGTGAGTAA AGTTATCGGA ATCGACTTAG
61 GAACAACAAA CTCATGTGTG GCAGTGCTTG AAGCGGCGA GCCTAAAGTT ATTGCTAACG
121 CTGAAGGAAA CCGCACAACG CCATCAGTTG TTGCATTTAA AAACGGCGAA CGTCAAGTAG
181 GGGAAGTGGC TAAACGCCAA TCTATTACAA ACCCTAACAC AATTATGTCT ATCAAACGTC
241 ATATGGGTAC TGATTATAAA GTTGA AATG AAGGAAAGGA TTACTCTCCA CAAGAAGTGT
301 CTGCTATCAT CCTTCAACAC CTFAATCAT ACGCTGAAAG CTATCTGGC GAAACAGTAT
361 CAAAAGCAGT TATCACAGTT CCTGCATACT TTAACGATGC TGAGCGTCAA GCAACAAAAG
421 ACGCTGGTAA AATTGCAGGT CTGGAAGTAG AACGTATCAT CAACGAGCCG ACTGCAGCAG
481 CGCTTGCATA CGGACTTGAT AAAACAGATG AAGATCAAAC GATCCTAGTA TACGACCTTG
541 GCGGCGGTAC ATTTCGACGTT TCCATCCTTG AGCTTGGCGA CCGTGTATTC GAAGTTCGTT
601 CAACTGCCGG CGACAACCGT CTGGGTGGGG ACGATTTTGA CCAAGTTATC ATCGATCATC
661 TGGTGTCTGA ATTCAAAAAA GAAAACGGCA TTGATTTGTC AAAAGACAAA ATGGCGCTTC
721 AGCGTTTGA AGACGCAGCT GAAAAGCGA AAAAAGATCT TTCCGGCGTA TCTTCTACGC
781 AAATTTCTTT ACCGTTTATC ACAGCTGGAG AAGCAGGACC GCTTCACCTT GAACCTACAT
841 TAACTCGCGC TAAATTCGAA GAGCTTTCTT CTCATTTAGT AGAGCGCAC ATGGGTCTTG
901 TCCGTCAAGC GCTTCAAGAT GCAGGACTTT CTGCAAGCGA AATCGACAAA GTCATCCTTG
961 TCGGCGGATC AACTCGTATC CCTGCCGTAC AAGAAGCAAT CAAAAAAGAA ACTGGAAAAG
1021 AAGCGCATAA AGGCGTAAAC CCGGATGAAG TTGTAGCGCT TGGTGTGCG ATTCAGGGCG
1081 GCGTTATCAC AGGTGACGTA AAAGATGTTG TTCTTCTTGA CGTTACACCG CTTTCTCTCG
1141 GTATCGAAAC AATGGGCGGT GTGTTTACAA AACTGATCGA CCGCAACACG ACGATCCCAA
1201 CAAGCAAATC TCAAGTGTTT TCAACTGCTG CTGATAACCA AACAGCTGTT GATATCCATG
1261 TTCTTCAAGG TGAGCGCCCA ATGTCTGCCG ACAACAAAAC ACTCGGCGCG TTCCAGCTTA
1321 CTGATATCCC GCCAGCACCG CGCGGCGTGC CTCAAATCGA AGTTTCTTTC GATATTGACA
1381 AAAACGGTAT CGTAAACGTA AGAGCAAAAAG ACTTAGGCAC AGGAAAAGAA CAAAACATTA
1441 CAATCAAATC TTCTTCAGGT CTCTCAGATG AAGAGATCGA ACGCATGGTA AAAGAAGCGG
1501 AAGAAAATGC TGACGCTGAT GCGAAGAAAA AAGAAGAAAT CGAAGTCCGC AACGAAGCAG
1561 ATCAGCTTGT TTTCCAAACT GAGAAAACAT TAAAAGATCT TGAAGGCAA GTGGACGAAG
1621 AACAAAGTAA AAAAGCCAAC GATGCCAAAG ACGTTTAAA AGCAGCGATT GAGAAAACG
1681 AATTTGAAGA GATCAAAGCG AAAAAAGATG AGCTTCAAAC AATCGTCAA GAGCTTCTTA
1741 TGAAGCTTTA TGAAGAAGCT GCTAAAGCAC AGCAAGCTCA AGGCGGAGCA AACGCTGAAG
1801 GCAAAGCGGA TGACAACGTT GTCGACGCTG AATACGAAGA AGTAAACGAC GACCAAAACA
1861 AAAAAATAAGT TCTTTTTAGT GTCAGCCCCG CTTCCGGAGC TGACCGAAAA GAACACATTT
1921 CATAATCTGA TTCAATGATT AGAAAAGTCAA AGTCAGGCAT CTTTGGCTT TGACTTTTTT
1981 TCTTGCCCCG GATAAAAGGA AATGAAAAT CATAATCAA AATGATACAA TCTAATTTAT
2041 GCGGAGAGTG AAGCGAGATG AGTAAAGCTG ATTACTATGA AGTGCTGGGA
2101 GTAAGTAAGA GCGCTTCAA GGATGAAATT AAAAAAGCTT A

```