Nucleotide sequence of a *Bacillus megaterium* gene homologous to the *dnaK* gene of *Escherichia coli*

Michael D.Sussman and Peter Setlow

Department of Biochemistry, University of Connecticut Health Center, Farmington, CT 06032, USA
Submitted April 7, 1987 Accession no. Y00154

The dnaK gene of Escherichia coli codes for a major heat shock protein whose primary sequence is extremely homologous to those of 70 kd heat shock proteins from higher organisms (1). Bacillus subtilis also contains an analogous heat shock protein (2). We have cloned and sequenced (see figure) a B. megaterium gene which codes for a protein with 59% sequence identity with the E. coli dnaK protein. While this B. megaterium gene did not complement an E. coli dnaK mutant, the B. megaterium protein cross-reacted strongly with antiserum against purified E. coli dnaK protein (a gift of Dr. Costa Georgopoulous, Univ. of Utah).

```
                                                                                                                    120
TTCCGGCTTCAAAAAGGTTACAAATTAAAAGATAAAATCATCCGCCCTGCGATGGTTAAAGTGAATCAATAACACTACATATTATGAAGGAGGTCATTGTCAATGAGTAAGATCATTGGT
                                                                                                  RBS           M  S  K  I  I  G
                                                                                                                    240
ATCGATTTAGGTACAACTAACTCTTGTGTCGCTGTATTAGAAGGCGGCGAACCAAAAGTAATTCCAAATCCAGAAGGAAACCGTACAACGCCATCAGTTGTGGCATTCAAAAACGGTGAG
I  D  L  G  T  T  N  S  C  V  A  V  L  E  G  G  E  P  K  V  I  P  N  P  E  G  N  R  T  T  P  S  V  V  A  F  K  N  G  E
                                                                                                                    360
CGTCAAGTTGGGGAAGTAGCGAAACGTCAAGCTATTACAAACCCTAACACAATTATTTCAGTTAAACGTCATATGGGTACAGATCATAAGGTTGAAGCTGAAGGCAAGCAATACACGCCT
R  Q  V  G  E  V  A  K  R  Q  A  I  T  N  P  N  T  I  I  S  V  K  R  H  M  G  T  D  H  K  V  E  A  E  G  K  Q  Y  T  P
                                                                                                                    480
CAAGAAATGTCAGCTATCATTCTTCAACATTTAAAAGGTTATGCTGAAGAGTATTTAGGTGAGCCTGTAACAAAAGCTGTTATCACAGTTCCTGCTTACTTTAATGATGCTGAGCGTCAA
Q  E  M  S  A  I  I  L  Q  H  L  K  G  Y  A  E  E  Y  L  G  E  P  V  T  K  A  V  I  T  V  P  A  Y  F  N  D  A  E  R  Q
                                                                                                                    600
GCAACAAAAGATGCTGGTAAAATTGCTGGTTTAGAAGTAGAGCGTATTATTAACGAGCCTACTGCAGCAGCACTTGCATACGGGTTAGAAAAAACAGATGAAGATCAAACAGTTTTAGTT
A  T  K  D  A  G  K  I  A  G  L  E  V  E  R  I  I  N  E  P  T  A  A  A  L  A  Y  G  L  E  K  T  D  E  D  Q  T  V  L  V
                                                                                                                    720
TATGACCTTGGTGGCGGTACGTTTGACGTATCTATTCTAGAACTTGGCGACGGCGTATTTGAAGTTCGCGCAACTGCAGGTGACAACCGCCTTGGTGGTGACGACTTTGACCAAGTAATC
Y  D  L  G  G  G  T  F  D  V  S  I  L  E  L  G  D  G  V  F  E  V  R  A  T  A  G  D  N  R  L  G  G  D  D  F  D  Q  V  I
                                                                                                                    840
ATCGACTATTTAGTCGCTGAATTCAAAAAAGAAAACGGCGTGGATTTAAGCAAAGATAAAATGGCGCTTCAACGTTTAAAAGATGCGGCTGAAAAAGCGAAAAAAGATTTATCAGGCGTA
I  D  Y  L  V  A  E  F  K  K  E  N  G  V  D  L  S  K  D  K  M  A  L  Q  R  L  K  D  A  A  E  K  A  K  K  D  L  S  G  V
                                                                                                                    960
ACATCTACACAAATTTCTTTACCATTTATCACTGCTGGAGAAGCTGGTCCTCTTCACTTAGAGGTATCTTTATCACGTGCTAAATTTGACGAGTTATCAGCAGGCCTTGTAGAGCGTACA
T  S  T  Q  I  S  L  P  F  I  T  A  G  E  A  G  P  L  H  L  E  V  S  L  S  R  A  K  F  D  E  L  S  A  G  L  V  E  R  T
                                                                                                                   1080
ATGGCTCCTGTGCGTCAAGCTTTAAAAGATGCAGGCCTTTCTGCAAGCGAACTTGATAAAGTAATCTTAGTTGGTGGTTCAACTCGTATCCCAGCGGTACAAGATGCAATCAAAAAAGAA
M  A  P  V  R  Q  A  L  K  D  A  G  L  S  A  S  E  L  D  K  V  I  L  V  G  G  S  T  R  I  P  A  V  Q  D  A  I  K  K  E
                                                                                                                   1200
ACTGGTCAAGATCCTCACAAAGGTGTAAACCCTGATGAAGTAGTTGCACTTGGTGCAGCAATTCAAGGTGGCGTATTAACTGGTGATGTAAAAGACGTTGTATTACTAGACGTAACGCCT
T  G  Q  D  P  H  K  G  V  N  P  D  E  V  V  A  L  G  A  A  I  Q  G  G  V  L  T  G  D  V  K  D  V  V  L  L  D  V  T  P
                                                                                                                   1320
TTATCACTAGGTATCGAAACAATGGGTGGCGTATTTACAAAGCTAATTGAGCGTAATACGACGATTCCAACAAGTAAATCACAAGTATTCTCAACGGCTGCAGATAGCCAAACAGCTGTA
L  S  L  G  I  E  T  M  G  G  V  F  T  K  L  I  E  R  N  T  T  I  P  T  S  K  S  Q  V  F  S  T  A  A  D  S  Q  T  A  V
                                                                                                                   1440
GATATTCATGTTCTTCAAGGTGAGCGTCCAATGTCTGCAGACAACAAAACGCTAGGTCGTTTCCAGTTAACTGATATTCCACCTGCACCACGCGGAGTACCTCAAATCGAAGTGTCATTC
D  I  H  V  L  Q  G  E  R  P  M  S  A  D  N  K  T  L  G  R  F  Q  L  T  D  I  P  P  A  P  R  G  V  P  Q  I  E  V  S  F
                                                                                                                   1560
GATATTGACAAAAATGGTATCGTAAACGTTCGTGCAAAAGATTTAGGTACAAACAAAGAGCAAGCTATTACAATTAAATCTTCAACAGGTTTATCAGATGATGAAATCGATCGCATGGTA
D  I  D  K  N  G  I  V  N  V  R  A  K  D  L  G  T  N  K  E  Q  A  I  T  I  K  S  S  T  G  L  S  D  D  E  I  D  R  M  V
                                                                                                                   1680
AAAGAAGCGGAAGAAAACGCAGATGCTGATAAGCAACGTAAAGAAGAAGTGGAACTACGCAATGAAGCAGATCAATTAGTGTTTACAACTGAAAAAACATTAAAAGATCTTGAAGGAAAA
K  E  A  E  E  N  A  D  A  D  K  Q  R  K  E  E  V  E  L  R  N  E  A  D  Q  L  V  F  T  T  E  K  T  L  K  D  L  E  G  K
                                                                                                                   1800
GTAGAAGAAGCTGAAGTAACAAAAGCTAACGAAGCAAAAGATGCTTTAAAAGCAGCGATTGAAAAGAATGACCTTGAAGAAATCAAAGCGAAAAAAGATGAACTTCAAGAAATCGTTCAA
V  E  E  A  E  V  T  K  A  N  E  A  K  D  A  L  K  A  A  I  E  K  N  D  L  E  E  I  K  A  K  K  D  E  L  Q  E  I  V  Q
                                                                                                                   1920
GCGTTAACTGTAAAATTGTATGAGCAAGCTCAACAAGCTCAGCAAGCAGGTGAACAAGGCGCTCAAAATGATGATGTTGTAGATGCAGAGTTTGAAGAAGTAAACGACGACAAAAAATAA
A  L  T  V  K  L  Y  E  Q  A  Q  Q  A  Q  Q  A  G  E  Q  G  A  Q  N  D  D  V  V  D  A  E  F  E  E  V  N  D  D  K  K  *
                                                                                                                   1973
TGCATCTTGACGACTAACGTGAATGACGCCATGGTGCGATTCACATCTTTTCA
```

LEGEND. The complete B. megaterium homolog of the E. coli dnaK gene is located on two adjacent Hind III fragments of 3.5 and 1kb. The nucleotide sequence was determined by the method of Maxam and Gilbert(3). The coding sequence is preceded by a strong ribosome binding site (RBS). Underlined amino acids are identical to those in the E. coli dnaK protein (1).

References
1. Bardwell, J. C. and Craig, E. A. (1984) Proc. Nat. Acad. Sci. USA 81:848-85.
2. Amosti, D. N., Singer, V. L., and Chamberlain, M. J. (1986) J. Bacteriol. 168:1243-1249.
3. Maxam, A. M. and Gilbert, W. (1980) Meth. in Enzymol. 65:499-560.