

A gene from *Mycobacterium tuberculosis* which is homologous to the DnaJ heat shock protein of *E. coli*

Raju B.Lathigra, Douglas B.Young, Doug Sweetser¹ and Richard A.Young¹

MRC Tuberculosis and Related Infections Unit, Hammersmith Hospital, London W12 0HS, UK and ¹Whitehead Institute for Biomedical Research, Cambridge, MA 02404, USA

Submitted December 17, 1987

Accession no.X06422

The *dnaK* and *dnaJ* genes code for heat shock proteins and are located on a single operon in *E. coli* (1,2). The DnaK protein is a member of a highly conserved family of proteins (hsp 70) found in all living cells (3). The derived sequence of a 71kD protein antigen from *Mycobacterium tuberculosis* shows it to be homologous to the *E. coli* DnaK protein (4). Sequence analysis of an open reading frame located 788 bp downstream from the *dnaK* gene of *M. tuberculosis* indicates the occurrence of a 356 amino acid protein (fig.1) sharing considerable sequence homology with the *E. coli* DnaJ protein (fig.2) including conservation of four tandem repeats of a motif consisting of Cys-X-X-Cys-X-Gly-X-Gly (marked in fig.1)(2). The DnaJ protein may resemble DnaK in being highly conserved amongst widely differing organisms.

```

(885)          R A Q P E M V E
ACAACAAGAGGAGAGCGGGGTCACCGGCACATGGCCCAAGGGAATGGCTCGAA 1429
  N D F Y Q E L C V S S D A S P E I N P
AAAGACTTCTACAGCGCTGGCCCTCTCTCTGATCCAGTCTCAAGAGATCAACCT 1480
  A Y R K L A R D L N P D A N P C W P A
GCCTATCGGAAGTTGGCCCGGACCTGCAATCCGGACCGCAACCCGGGCAACCCGGCC 1545
  C E R F K A V S E A H V L S D P A K E
GGCAACCGGTTCAAGCCGGTTTCGGAGCGCATAACCTGCTCTCGGATCCGCCAAGCC 1600
  K E Y D E T P I I A C G C G S G R E T
AAGGACTACGCAAAACCCGCCCTTCTTCCCGGGGGCGGCTTCGGCGCCCTCGGTT 1665
  D S G F G G C F G C F C V G C D G A E F
GACAGCGGCTTTGGGGCGGGTTTCGGCGGTTTCGGGCGCGTGGAGACGGCGCCGACT 1729
  N L N D L F D A A S B T G G T T I G D L
AACTCAACGACTTTTCCAGCCCGCCAGCCGAAACCGCGGCTACCACTCGGCTGACT 1789
  T G C G L F C G C S A P F S P E E G
TTGGTGCTTCTTCCGACCGCTGGCAGCCGCCCTCCAGCCCGCCCGGACCGCCCA 1849
  D L E Y E T E L D F V E A A K G V A N P
GACCTGGAGACCGAGACCGAATGGATTCCTGGAGGCCCAAGGGCTGGCGATGCC 1909
  L B L Y S P A P C T T N C H C S C A R P C
CTGGATTAACCAACCCCGCGCGCTGCACCAATGCCATGGCAGCCGGCCCGCCAGCC 1969
  T S P K V C F T T C H E G S G V I N R N Q
ACCACCCAAAGGTGTCTCCCACTTGCACCGGCTCGGGCGTGCATCAACCCCAATCAG 2029
  A P C F S E P C T T D E S T S G S I E H
CGCTTCGCTTCTCCGACCGCTGCACCGACTCCCACTAGCGGCTCGATATCGACCAC 2089
  P C E E C K G C T T G V T T B T B T I N V P
CCTTCGAGGAGTCCAAAGCCACCGCGTACCTCCACCCCAACCATCAAGCTGCGG 2145
  I P F C V T D G G R I B L A G C G E A C
ATCCCGCCCGTCTCGAGGATGGCCAGCCATCCCGTACCGGTCAGGGCCGAGCCCG 2209
  L B G A P S C D L V V T V H V B P D K I
TTGCGCGCGCTCCCTCGGGGATCTCTACTGACCGGTGATGTCGGCCCGCAAGATC 2269
  P E B D G D D L T V T V F S F T L L A
TTGCGCCGCAACCGCGACACTCAGCGTCCCGTCCCGTCACTCAGCGAATGGCT 2329
  L G S T L S V P T L D G T V G V B V P K
TTGGCTCAAGCTCTCGGTACTACCTACCGAGCCAGCGGTCGGGCTCCGGGTGCCAAA 2385
  C T A D G D I L E V E G C G V L F S A V G
GGCAGCTGACCGCCGCAATTCTCGCTGTGGCCGACCGGTTCTCCCAACCGCAATGG 2449
  V A A T Y L S P S
GTAGCCGCACTACTTCTCACCGTCAAGCTGGCCG 2479
    
```

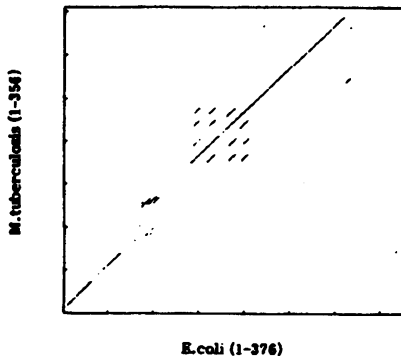


Figure 2. Comparative analysis of *E. coli* and *M. tuberculosis* genes.

Amino acid sequences were compared using dot matrix program (Staden-Plus Software, Amersham, UK) with span length 15, proportional score 180, identities score 8. The repeated sequences between residues 150 and 250 are clearly seen in the dot matrix analysis.

Figure 1. Nucleotide sequence for proposed *dnaJ* gene from *M. tuberculosis*. Insert DNA from lambda gtl1 clone Y3273 (5) was sequenced by Sanger dideoxy procedure (6). A potential ribosome binding site is indicated along with the derived amino acid sequence of an open reading frame downstream from the *dnaK* gene.

References

1. Bardwell, J.C.A., Tilly, K., Craig, E., King, J., Zylicz, M. and Georgopoulos, C. (1986) *J. Biol. Chem.* 261:1782-1785.
2. Okhi, M., Tamura, F., Nishisura, S. and Uchida, H. (1986) *J. Biol. Chem.* 261:1778-1781.
3. Bardwell, J.C. and Craig, E.A. (1984) *Proc. Natl. Acad. Sci. USA* 81:848-852.
4. Young, D.B., Lathigra, R.B., Hendrix, R., Sweetser, D. and Young, R.A. (1988) submitted for publication.
5. Hussion, R.N. and Young, R.A. (1987) *Proc. Natl. Acad. Sci. USA* 84:1679-1683.
6. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74:5463-5467.