Volume 17 Number 21 1989 Nucleic Acids Research

Nucleotide sequence of the *Thermus thermophilus HB8* gene coding for elongation factor G

A.V.Yakhnin, D.P.Vorozheykina and N.I.Matvienko

Institute of Protein Research, USSR Academy of Sciences, Pushchino, Moscow region, USSR
Submitted September 29, 1989 EMBL accession no. X16278

A clone with the Thermus thermophilus HB8 gene coding for elongation factor G was identified in a genomic library in plasmid pBR 322 by hybridization with labeled oligonucleotide 19 bases in length complementary to the 3'- end of the T.thermophilus fus gene (1). Fragments of the fus gene were subcloned into pTZ 18R, M13 mp18 and mp 19 vectors. Their nucleotide sequence were determined by the dideoxy chain termination method (2). Computer manipulation with sequence data was made with the MicroGenie sequence analysis program (Beckman,USA). The open reading frame starting with ATG (position 1) and ending with TAG (position 2074) codes the elongation factor G 690 amino acids in length (Mr 76756 Da). The termination codon TGA (position -5) of the preceeding rps7 gene coding for the ribosomal protein S7 is underlined. The G+C content in the thermophilic fus gene is 66.0 %, while it is 96.5 % in the third letter of the codons. C is an especially prefered nucleotide in this position (55.7 %). The amino acid sequence of EF-G from T.thermophilus has a 59.7 % homology with that of E.coli. There are some highly conservative amino acid sequences in the N-terminal part of the thermophilic EF-G (underlined) which are found in different GTP-binding proteins (3).

```
  -6 GtgagtATGGCGGTCAAGGTAGAGTACGACCTCAAGCGGCTCCGGAACATCGGCATCGCCGCCCACATTGACGCGGGCAAGACCACCACCACCGAGCGCATCCTCTACTACACGGGCCGG
              A  V  K  V  E  Y  D  L  K  R  L  R  N  I  G  I  A  A_H_I_D_A_G_K  T  T  T  T  E  R  I  L  Y  Y  T  G  R
 115 ATCCACAAGATCGGCCAGGTCCACCAGGGCGCCGCCACCATGGACTTCATGGAGCAGGAGCGGGAGCGGGGCATCACCATCACCGCCGCCGTGACCACCTGCTTCTGGAAGGACCACCGC
      I  H  K  I  G  E  V  H  E  G  A  A  T  M  D  F  M  E  Q  E  R  E  R_G_I_T_I_T  A  A  V  T  T  C  F  W  K  D  H  R
 235 ATCAACATCATCGACGCCCGGCCAGGTGGACTTCACCATTGAGGTGGAGCGCTCCATGCGGGTCCTGGACGGGGCCATCGTGGTCTTTGACTCCAGCCAGGGGGTGGAGCCCCAGTCG
      I  N  I  I  D_T_P_G_H  V  D  F  T  I  E  V  E  R  S  M  R  V  L  D  G  A  I  V  V  F  D  S  S  Q  G  V  E  P  Q  S
 355 GAGACCGTCTGGCGCCAGCGGAGAAGTACAAGGTCCCCCGCATCGCCTTCGCCAACAAGATGGACAAGACCGGGGCCGAACCTCTGGCTCGTGATCCGCACCATGCAGGAGCGCCTGGGG
      E  T  V  W  R  Q  A  E  K  Y  K  V  P  R  I  A  F  A_N_K_M_D_K  T  G  A  D  L  W  L  V  I  R  T  M  Q  E  R  L  G
 475 GCGCGCCCGGTGGTGATGCAGCTCCCCATCGGCCGGGAGGACACCTTCTCGGGGATCATTGACGTCCTCCGGATGAAGGCCTACACCTACGGCAACGACCTGGGCACGGACATCCGCGAG
      A  R  P  V  V  M  Q  L  P  I  G  R  E  D  T  F  S  G  I  I  D  V  L  R  M  K  A  Y  T  Y  G  N  D  L  G  T  D  I  R  E
 595 ATCCCCATCCCCGAGGATACCTGGACCAGGCCCGGGAGTACCACGAGAAGCTCGTGGAGGTGGCCGCCGACTTTGACGAGAACATCATGCTGAAGTACCTCGAGGCCGAGGAGCCCACC
      I  P  I  P  E  E  Y  L  D  Q  A  R  E  Y  H  E  K  L  V  E  V  A  A  D  F  D  E  N  I  M  L  K  Y  L  E  G  E  E  P  T
 715 GAGGAAGAGCTGGTGCCCCATCCGCAAGGGCACCATTGACCTCAAGATCACCCCCGTCTTCTGGGCTCCGCCCTGAAGAACAAGGGCTCCAGCTCCTCCTGGACGCGGTGGTGGAC
      E  E  E  L  V  P  A  I  R  K  G  T  I  D  L  K  I  T  P  V  F  L  G  S  A  L  K  N  K  G  V  Q  L  L  L  D  A  V  V  D
 835 TACCTGCCCTCCCCCTTGGACATCCCCCCCATCAAGGGCACCACCCCTGAGGGCG-GGTGGTGGAGATCCACCCCGACCCCAATGGCCCTTGGCGGCTTGGCCTTCAAGATCATGGCC
      Y  L  P  S  P  L  D  I  P  P  I  K  G  T  T  P  E  G  E  V  V  E  I  H  P  D  P  N  G  P  L  A  A  L  A  F  K  I  M  A
 955 GACCCCTACGTGGGCCGCCTCACCTTCATCCGCGTCTACTCGGGGACCCTCACCTCGGGCTCCTACGTCTACAACACCACCAAGGGCCGCAAGGAACGGGTGGCCCGGCTCCTCCGCATG
      D  P  Y  V  G  R  L  T  F  I  R  V  Y  S  G  T  L  T  S  G  S  Y  V  Y  N  T  T  K  G  R  K  E  R  V  A  R  L  L  R  M
1075 CACGCCAACCACCGGAGGAGGTGGAGGAGCTCAAGGCGGGCGACCTGGGCGCCGTGGTGGGCCTCAAGGAGACGATCACCGGGGACACCCTGGTGGGCGAGGACGCGCCCCGGGTGATC
      H  A  N  H  R  E  E  V  E  E  L  K  A  G  D  L  G  A  V  V  G  L  K  E  T  I  T  G  D  T  L  V  G  E  D  A  P  R  V  I
1195 CTGGAGTCCATTGAGGTGCCGGACCCCTGATTGACGTGCGCCATTGAGCCCAAGACCAAGGCCGACCAGGATGAAGCTTTCCCAGCCCTGCCCCGCCTGGCGGAGGAGGACCCCACCTTC
      L  E  S  I  E  V  P  E  P  V  I  D  V  A  I  E  P  K  T  K  A  D  Q  E  K  L  S  Q  A  L  A  R  L  A  E  E  D  P  T  F
1315 CGCGTCTCCACCCACCCCGAGACCGCGCAGACCATCATCAGTGGGATGGCGACCTCCACCTGGATCATCGTGGACCGGCTCAACCCGGAGTTCAAGGTGGACGCCAACGTGGGCAAG
      R  V  S  T  H  P  E  T  G  Q  T  I  I  S  G  M  G  E  L  H  L  E  I  I  V  D  R  L  K  S  E  F  K  V  D  A  N  V  G  K
1435 CCCCAGGTGGCGTACCGGGAGACCATCACCAAGCCGGTGGACGTGGAGGTCAAGTTCATCCCCAGACCGGCGCCGCGGCCAGTACGGCCACGTGAAGATCAAGGTGGAGCCCCTCCCC
      P  Q  V  A  Y  R  E  T  I  T  K  P  V  D  V  E  G  K  F  I  R  Q  T  G  G  R  G  Q  Y  G  H  V  K  I  K  V  E  P  L  P
1555 CGGGGCTCGGGCTTTGAGTTCGTCAACGCCATCGTGGGCGGGTGATCCCCAAGGAGTACATCCCCGCCGTGCAGAAGGGGATCGAGGAGGCCATGCAGTCCGGGCCCCTTATCGGCTTC
      R  G  S  G  F  E  F  V  N  A  I  V  G  G  V  I  P  K  E  Y  I  P  A  V  Q  K  G  I  E  E  A  M  Q  S  G  P  L  I  G  F
1675 CCCGTGGTGGACATCAAGGTCACCCTTTACGACGGCTCCTACCACGAGGTGGACTCCTCCGAGATGGCCTTTAAGATCGCCGGCTCCATGGCCATCAAGGAGGCGGTGCAGAAGGGGGAC
      P  V  V  D  I  K  V  T  L  Y  D  G  S  Y  H  E  V  D  S  S  E  M  A  F  K  I  A  G  S  M  A  I  K  E  A  V  Q  K  G  D
1795 CCGGTGATCCTCGAGCCCATCATGCGGGTGGAGGTCACCACCCCCGAGGAGTACATGGGGGACGTCATCGGCGACCTGAACGCCCGGCGGGGGCAGATCCTGGGGATGGAGCCCCGGGGC
      P  V  I  L  E  P  I  M  R  V  E  V  T  T  P  E  E  Y  M  G  D  V  I  G  D  L  N  A  R  R  G  Q  I  L  G  M  E  P  R  G
1915 AACGCCCAGGTGATCCGGGCCTTCGTGCCCTTGGCGGAGATGTTCGGCTACGCCACGGACCTGCGCTCCAAGACCCAGGGGCGAGGCTCCTTCGTCATGTTCTTTGACCACTACCAGGAG
      N  A  Q  V  I  R  A  F  V  P  L  A  E  M  F  G  Y  A  T  D  L  R  S  K  T  Q  G  R  G  S  F  V  M  F  F  D  H  Y  Q  E
2035 GTGCCCAAGCAGGTCCAGGAGAAGCTCATCAAGGGTCAAtag
      V  P  K  Q  V  Q  E  K  L  I  K  G  Q  *
```

REFERENCES

1.Seidler L.,Peter M.,Meissner F.,Sprinzl M.(1987) Nucl.Acids Res.15, 9263-9277.
2.Sanger F.,Nicklen S., Coulson A.(1977) Proc.Natl.Acad.Sci.USA 74, 5463-5467.
3.Kushro A.,Shimizu M.,Tomita K.(1987) Eur.J.Biochem.170, 93-98.