Coat protein TP4 of the virus TTV1: primary structure of the gene and the protein

Horst Neumann and Wolfram Zillig

Max-Planck-Institut für Biochemie, Am Klopferspitz 18, D-8033 Martinsried, FRG
Submitted October 17, 1989 EMBL accession no. X14855

The coat protein TP4 of the virus TTV1 of the extremely thermophilic archaebacterium *Thermoproteus tenax* strain Kral was mapped within the viral genome by locating the N-terminal amino acid sequence of the protein in the TTV1 DNA sequence. TP4 is a slightly acidic protein, the predicted size is 28.5 kDa. The start codon is GTG whereas the N-terminal residue is proline. In agreement with the GC-content of the TTV1 DNA which is only 37% (1), a preference for A or U as third base in the codons was found and C was even less common than G as third base. A Shine-Dalgarno sequence was not recognizable upstream or immediately downstream of the initiation codon of the gene.

```
  1  CCTAAAATACTCTATATTGTGCCATCATTTAGTACACAGTCAATAACTAGACCG
                           P  S  F  S  T  Q  S  I  T  R  P
 55  TCTGTAGGTGTAGCAATTGGACCCGATGGAATATATAGAGCTCTTTATGGATGG
     S  V  G  V  A  I  G  P  D  G  I  Y  R  A  L  Y  G  W
109  TATGGAAATGAACAGCCGCTGGAAATAGATAAATGTTCATATGATTTATCGACT
     Y  G  N  E  Q  P  L  E  I  D  K  C  S  Y  D  L  S  T
163  TGTTCTCCTGTAGTATCGAGTGGACCAACAATAGGTGATGGATATGGAGGATTT
     C  S  P  V  V  S  S  G  P  T  I  G  D  G  Y  G  G  F
217  TATGATGGAACTAATATATGGTTCAGTGGTTCAGATGAGGCGAATGATCAAGGA
     Y  D  G  T  N  I  W  F  S  G  S  D  E  A  N  D  Q  G
271  GTGGTAGCAAGCTATAATCCATCTACAGGAGCATTCAATTATGCATATTATCCT
     V  V  A  S  Y  N  P  S  T  G  A  F  N  Y  A  Y  Y  P
325  GGTAATTCTACAAAATACGTAATTAAGATATTCTACTATAATGGATACTACTAT
     G  N  S  T  K  Y  V  I  K  I  F  Y  Y  N  G  Y  Y  Y
379  TTGATAACATGCTGTGACCCAGGAACACTACTTAAGTGCACAAATCCTCTAAAT
     L  I  T  C  C  D  P  G  T  L  L  K  C  T  N  P  L  N
433  CCGTCTACATGTACCCAGCTTAATATAAATGCACCTACTTCATTTACAGTTCAA
     P  S  T  C  T  Q  L  N  I  N  A  P  T  S  F  T  V  Q
487  GAATACTATATGTATACACAAGGCTCATATGCACTACTATCATACTTCCAGAAA
     E  Y  Y  M  Y  T  Q  G  S  Y  A  L  L  S  Y  F  Q  K
541  AATATGAATAATAAATTGTTCAGTTTATGTAATTTCGATGGAACAAATCTATAT
     N  M  N  N  K  L  F  S  L  C  N  F  D  G  T  N  L  Y
595  AACTGTATAAATATATACTCTACGACTGGAATACTACATTATCTCCATCACAGT
     N  C  I  N  I  Y  S  T  T  G  I  L  H  Y  L  H  H  S
649  TTACTAGAGGAGCAATATGTGGAGATAATATAGTATTCCCAGATATTATAAACA
     L  L  E  E  Q  Y  V  E  I  I  *
```

References

(1) Neumann H, Schwass V, Eckerskorn C, Zillig W (1989) Mol Gen Genet 217, 105-110

