Volume 16 Number 15 1988 Nucleic Acids Research

Nucleotide and deduced amino acid sequence of the *Bacillus sphaericus* 1593M gene encoding a 51.4 kD polypeptide which acts synergistically with the 42 kD protein for expression of the larvicidal toxin

Calliope Arapinis, Françoise de la Torre and Jekisiel Szulmajster

Laboratoire d'Enzymologie du CNRS, 91190 Gif-sur-Yvette, France

Submitted June 28, 1988 Accession no. X07992

The HindIII DNA fragment of Bacillus sphaericus 1593M coding for the larvicidal toxin (1) bears two genes, separated by 171 nt coding for two polypeptides of the Mr of 42 and 51 kD. The nucleotide sequence analysis of the gene coding for the 42 kD protein has been reported by Hindley and Berry (2). Here we describe the nucleotide sequence of the gene coding for the second protein. This sequence contains an ORF of 1344 nucleotides which stops by an amber codon and corresponds to a 448 aminoacid polypeptide with a Mr of 51.37. Ten nt upstream the initiation codon there is a putative RBS sequence (underlined) complementary to the 3' terminus of the 16S rRNA characteristic for a number of Bacilli genes (3).

```
                                                    gcttgtcaacatgtgaagattaaaggtaactttcagtttctttctgtg -449
taacaatacacgaagtaatatatgtatttctatagaaattaatcaaaaaaagacctagtgaactgcacoctgtcaagtagatagtggaaataataaaaatgtattaagcggc -337
ttgtgctcttaattctaagaaggtgatgaaaaacttgcactctagggcgtaccaagcaatcaaggagaaacgttgtatctccggattaattcgaactacaacgtctattgt -225
agattaaaccgtaagagattatataattattataacaatactttaattaattcaattgaaattaaacaaacaacttttttcggtatatactatctaacatatcggtttcagtc -113
ccatcgttctaaaggtaccttcttttggttaacgttatttaatgaacttttttaggttttaaataatataatgagaagtatttttttatcaatgataaggagatgaagaaagc   -1
 M   C   D   S   K   D   N   S   G   V   S   E   K   C   G   K   K   F   T   N   Y   P   L   N   T   T   P   T    28
ATG TGC GAT TCA AAA GAC AAT TCT GGC GTT TCA GAA AAA TGC GGA AAG AAA TTT ACT AAT TAC CCG CTA AAT ACT ACT CCT ACA   84
 S   L   N   Y   N   L   P   E   I   S   K   K   F   Y   N   L   K   N   K   Y   S   R   N   G   Y   G   L   S    56
AGC CTA AAT TAT AAC CTT CCA GAA ATA TCA AAA AAA TTT TAT AAC CTT AAG AAT AAA TAT TCA CGG AAT GGT TAT GGT TTA TCA  168
 K   T   E   F   P   S   S   I   E   N   C   P   S   N   E   Y   S   I   M   Y   D   N   K   D   P   R   F   L    84
AAA ACC GAA TTT CCT TCA AGT ATC GAA AAT TGC CCA TCT AAC GAA TAT TCA ATA ATG TAT GAT AAT AAA GAT CCT CGA TTC TTG  252
 I   R   F   L   L   D   D   G   R   Y   I   I   A   D   R   D   D   G   E   V   F   D   E   A   P   T   Y   L   112
ATT CGG TTT TTA TTA GAT GAT GGT AGA TAT ATT ATT GCA GAT AGA GAC GAT GGA GAA GTT TTT GAT GAA GCA CCT ACT TAT TTG  336
 D   N   N   N   H   P   I   I   S   R   H   Y   T   G   E   E   R   Q   K   F   E   Q   V   G   S   G   D   Y   140
GAT AAT AAC AAT CAC CCT ATC ATA AGT AGA CAT TAT ACC GGA GAA GAG AGA CAA AAG TTT GAG CAG GTA GGT AGT GGA GAT TAT  420
 I   T   G   E   Q   F   F   Q   F   Y   T   Q   N   K   T   R   V   L   S   N   C   R   A   L   D   S   R   T   168
ATT ACG GGA GAG CAA TTT TTT CAA TTC TAT ACA CAA AAC AAA ACA CGT GTA TTG TCA AAT TGT AGG GCG CTT GAC AGT AGG ACA  504
 I   L   L   S   T   A   K   I   F   P   I   Y   P   P   A   S   E   T   Q   L   T   A   F   V   N   S   S   F   196
ATA TTA CTA TCT ACT GCA AAA ATC TTC CCA ATT TAC CCT CCA GCT TCT GAA ACT CAA CTA ACA GCT TTC GTT AAT AGT TCA TTT  588
 Y   A   A   A   I   P   Q   L   P   Q   T   S   L   L   E   N   I   P   E   P   T   S   L   D   D   S   G   V   224
TAT GCT GCG GCA ATT CCT CAA TTA CCC CAA ACA TCC TTA CTT GAG AAT ATT CCT GAG CCT ACT AGT CTC GAT GAT TCT GGA GTA  672
 L   P   K   D   A   V   R   A   V   K   G   S   A   L   L   P   C   I   I   V   H   D   P   N   L   N   N   S   252
TTA CCA AAA GAT GCA GTA AGA GCA GTT AAA GGA AGT GCG CTA TTA CCT TGT ATA ATA GTA CAT GAT CCT AAT TTA AAC AAT TCC  756
 D   K   M   K   F   N   T   Y   Y   L   L   E   Y   K   E   Y   W   H   Q   L   W   S   Q   I   I   P   A   H   280
GAT AAA ATG AAA TTT AAT ACC TAC TAT CTT TTA GAA TAT AAA GAA TAC TGG CAT CAA TTA TGG TCA CAA ATT ATA CCT GCT CAT  840
 Q   T   V   K   I   Q   E   R   T   G   I   S   E   V   V   Q   N   S   M   I   E   D   L   N   M   Y   I   G   308
CAA ACT GTA AAA ATA CAG GAA CGA ACA GGA ATA TCT GAA GTT GTA CAA AAT AGC ATG ATT GAA GAT TTA AAT ATG TAT ATT GGA  924
 A   D   F   G   M   L   F   Y   F   R   S   S   G   F   K   E   Q   I   T   R   G   L   N   R   P   L   S   Q   336
GCA GAT TTT GGC ATG CTT TTT TAT TTT AGA TCT AGT GGA TTT AAG GAA CAA ATA ACA AGG GGG CTA AAT AGG CCT TTA TCC CAA 1008
 T   T   T   Q   L   G   E   R   V   E   E   M   E   Y   Y   N   S   N   D   L   D   V   R   Y   V   K   Y   A   364
ACG ACC ACT CAG TTA GGA GAA AGA GTA GAA GAA ATG GAG TAT TAT AAT TCT AAT GAT TTG GAT GTT AGA TAT GTG AAA TAC GCA 1092
 L   A   R   E   F   T   L   K   R   V   N   G   E   I   V   K   N   W   V   A   V   D   Y   R   L   A   G   I   392
TTG GCT AGA GAA TTC ACA CTA AAA CGC GTT AAT GGT GAA ATT GTA AAA AAT TGG GTT GCT GTA GAT TAT CGA TTG GCA GGT ATA 1176
 Q   S   Y   P   N   A   P   I   T   N   P   L   T   L   T   K   H   T   I   I   R   C   E   N   S   Y   D   G   420
CAA TCG TAT CCT AAT GCA CCT ATA ACT AAT CCA CTT ACG CTA ACA AAA CAT ACA ATT ATT CGA TGT GAA AAT AGT TAC GAT GGA 1260
 H   I   F   K   T   P   L   I   F   K   N   G   E   V   I   V   K   T   N   E   E   L   I   P   K   I   N   Q   448
CAC ATA TTT AAA ACA CCT TTA ATC TTT AAA AAT GGT GAA GTT ATT GTA AAA ACG AAT GAA GAA TTA ATA CCT AAA ATT AAC CAG 1344
TGA tactttaacttcaaatattcattaccatgttatttaaaatagtagatagatgaaataaatagtatatattaagacaacaacttaattttgacacataagaataatttt 1456
taaatgtataaatagtatttagagtgttattgcaatatattttttgaaagggagctaaaagacatgagaaatttggattttattgattcttttatacccacagaaggaaagt 1568
acattcgcgttatggatttt                                                                                             1587
```

References
1) Tandeau de Marsac, N., de la Torre, F. and Szulmajster, J. (1987) Mol. Gen. Genet. 209:396-398.
2) Hindley, J. and Berry, C. (1987) Mol. Microbiol. 1(2):187-194.
3) McLaughlin, J.R., Murray, C.L. and Rabinowitz, J. (1981) J. Biol. Chem. 256:11283-11291.