Volume 336, number 1, 79–82 FEBS 13375 December 1993

Primary structure of *cryX***, the novel δ-endotoxin-related gene from *Bacillus thuringiensis* spp. *galleriae*

A.B. Shevelev*, M.A. Svarinsky, A.I. Karasin, Ya.N. Kogan, G.G. Chestukhina, V.M. Stepanov

Institute of Microbial Genetics (VNIIGenetika), Laboratory of Protein Chemistry, 1st Dorozhny 1, Moscow, 113545, Russian Federation

Received 11 October 1993; revised version received 26 October 1993

A cry-related sequence, designated *cryX* (EMBL X75019), was localized upstream of *cry*IG, the δ-endotoxin gene cloned from spp. *galleriae* of *Bacillus thuringiensis* and sequenced earlier [(1991) FEBS Lett. 293, 25–28]. Analysis of the *cryX* complete nucleotide sequence enabled us to explain its virtual crypticity and to reveal the chimeric structure of the genes, *cryX* and *cry*IG. The amino acid sequence of 1,151 residues encoded by the continuous reading frame of *cryX* is similar to the other δ-endotoxins but differs essentially from them.

cryX; δ-Endotoxin; Primary structure; Cryptic gene; *Bacillus thuringiensis*

1. INTRODUCTION

Various subspecies of *Bacillus thuringiensis* (BT) produce a great variety of 70–130 kDa crystal-forming proteins, known as δ-endotoxins (dET), displaying highly specific entomocidal activity towards larvae of different insect Orders.

The spp. *galleriae* of BT is characterized by its ability to procedure a wide spectrum of dETs [2]. Earlier we reported cloning [3] and sequencing [4] the *cry*IG gene coding for the major crystal protein component of this spp. with unique toxicity toward *Galleria mellonella* larvae. Another *cry*-related sequence, designated *cryX*, was localized upstream of *cry*IG. Analysis of the *cryX* complete nucleotide sequence presented here explains its virtual crypticity and also reveals the chimeric structure of the genes, *cryX* and *cry*IG.

2. EXPERIMENTAL

2.1. cryX *cloning and sequence analysis*

The nucleotide sequence of subclones overlapping 4.0 kb the *Eco*RI-*Hin*dIII segment of interest (see Fig. 1) was determined by the double-stranded DNA sequenase version [5] of Sanger's dideoxynucleotide method.

3. RESULTS AND DISCUSSION

The recombinant phasmid pOC10 (with a 10 kb insertion fragment bearing the *cryX-cry*IG locus) was isolated from *Bacillus thuringiensis* spp. *galleriae* genomic library as described [3,4]. An additional *cry*-related sequence was detected by Southern-blot hybridization upstream to the 4.5 kb *cry*IG-containing segment (downstream of the unique *Kpn*I site, see Fig. 1) sequenced earlier [4].

The nucleotide sequence of the 4 kb segment covering the 5′-flank of the insert of pOC10 (from the first *Eco*RI down to the second *Hin*dIII site), contains the continuous reading frame (CRF) of 3,453 bp encoding a polypeptide of 1,151 amino acids (see Fig. 2). The length of this CRF corresponds to the normal size of the dETs, but there were no translation initiation elements detected in its 5′-flank. The new gene might therefore be considered the cryptic gene that cannot be normally expressed in vivo.

According to contemporary structure–function nomenclature dETs are classified into four major classes, I–IV [1]. The hypothetical CRF product, like dETs of classes CryI or CryIV, is a 130 kDa-type protein. The most striking feature of the C-terminal half (amino acids 653–1,151) of CryX is that it appears to be nearly identical (> 98%) to the corresponding region of CryIG. At the same time, its N-terminal half has only marginal similarity with the known dETs. The highest percentage of the coincident residues is observed in a comparison of the N-terminal half of CryX with CryIB [6] (23%), and to a lesser extent with dETs of other groups (CryI and CryIII-12–20%). The majority of the matching residues is located in five conserved regions ('blocks') common for all known dETs, but the percentage of identity does not exeed 40% even there. The presence and the mode of distribution of the blocks indicates that *cryX* can be considered a *cry*-related sequence.

The observed type of similarity between *cryX* and

*Corresponding author. Fax: (7) (095) 315 0501.

**The index *X* indicates the cryptic character of the new sequence and the unidentified nature of the biospecificity of its potential product.

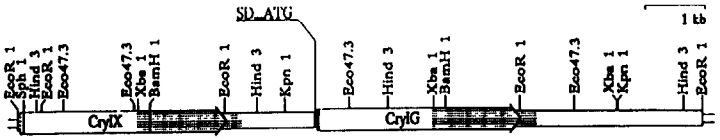

Fig. 1. Structural organization of the *cryX–cry*IG-containing segment of *Bacillus thuringiensis* spp. *galleriae:* the map and gene localization in the 11.7 kb insertion segment of the pOC10 recombinant phasmid. The regions of *cryX* vs. *cry*IG nucleotide sequence identity are shaded. The presence of the normal SD-ATG initiator motif in *cry*IG (absent in *cryX*) is indicated.

```
EcoRI                                31                              61                              91
GAATTCCTCTATTTGATATAGATGATTTCTCTTATTGGAGCATTTTCTTTTTCATGTTTTCGAATTGAAAACATGTATTTGGCATGCTTTT
                          *  F  L  L  L  E  H  F  L  F  H  V  F  E  L  K  T  C  I  W  H  A  F
                        121                             151                             181
TTCCTAACTAAGCTGAGTAGTTACAAAGATTATTTGAAAATGTCCGAAGGAGACTATATAGACTCGTATATAAATCCTGGAAATGTTAGA
 F  L  T  K  L  S  S  Y  K  D  Y  L  K  M  S  E  G  D  Y  I  D  S  Y  I  N  P  G  N  V  R
                        211                             241                             271
ACTGGACTACAAACTGGAATTGATATTGTTGCAGTAGTAGTAGGTGCTTTAGGTGGACCAGTTGGTGGCATACTCACTGGTTTTCTTTCT
 T  G  L  Q  T  G  I  D  I  V  A  V  V  V  G  A  L  G  G  P  V  G  G  I  L  T  G  F  L  S
                        301                             331                             361
ACTCTTTTTGGTTTTCTTTGGCCATCTAATGATCAAGCAGTATGGGAAGCTTTTATAGAACAAATGGAAGAACTGATTGAACAAAGGATA
 T  L  F  G  F  L  W  P  S  N  D  Q  A  V  W  E  A  F  I  E  Q  M  E  E  L  I  E  Q  R  I
                        391                             421                             451
TCAGATCAAGTAGTAAGGACTGCACTCGATGACTTAACTGGAATTCAAAATTATTATAATCAATATCTAATAGCATTAAAGGAATGGGAG
 S  D  Q  V  V  R  T  A  L  D  D  L  T  G  I  Q  N  Y  Y  N  Q  Y  L  I  A  L  K  E  W  E
                        481                             511                             541
GAAAGACCAAACGGCGTAAGAGCAAACTTAGTTTTGCAAAGATTTGAAATCTTGCACGCGCTATTTGTAAGTAGTATGCCAAGTTTTGGT
 E  R  P  N  G  V  R  A  N  L  V  L  Q  R  F  E  I  L  H  A  L  F  V  S  S  M  P  S  F  G
                        571          block 1            601                             631
AGTGGCCCTGGAAGTCAAAGGTTTCAGGCACAATTGTTGGTTGTTTATGCGCAAGCAGCAAATCTTCATTTACTATTATTAGCTGATGCT
 S  G  P  G  S  Q  R  F  Q  A  Q  L  L  V  V  Y  A  Q  A  A  N  L  H  L  L  L  L  A  D  A
      block 1                   661                             691                             721
GAAAAGTATGGGGCAAGATGGGGACTCCGTGAATCCCAGATAGGAAATTTATATTTTAATGAACTACAAACTCGTACTCGAGATTACACC
 E  K  Y  G  A  R  W  G  L  R  E  S  Q  I  G  N  L  Y  F  N  E  L  Q  T  R  T  R  D  Y  T
                        751                             781                             811
AACCATTGTGTAAACGCGTATAATAACGGGTTAGCCGGGTTACGAGGAACGAGCGCTGAAAGTTGGTTAAAGTACCATCAATTCCGCAGA
 N  H  C  V  N  A  Y  N  N  G  L  A  G  L  R  G  T  S  A  E  S  W  L  K  Y  H  Q  F  R  R
                        841                             871                             901
GAAGCAACCTTAATGGCAATGGATTTGATAGCTTTATTTCCATATTATAACACCCGGCGATATCCAATCGCAGTAAATCCTCAGCTTACA
 E  A  T  L  M  A  M  D  L  I  A  L  F  P  Y  Y  N  T  R  R  Y  P  I  A  V  N  P  Q  L  T
                  block 2       931                             961                             991
CGTGAGGTATATACAGATCCATTAGGCGTTCCTTCTGAAGAATCAAGTTTATTTCCAGAATTGAGATGCTTAAGATGGCAAGAGACTTCT
 R  E  V  Y  T  D  P  L  G  V  P  S  E  E  S  S  L  F  P  E  L  R  C  L  R  W  Q  E  T  S
                        1021                            1051                 block2     1081
GCCATGACTTTTTCAAATTTGGAAAATGCAATAATTTCGTCACCACATCTATTTGACACAATAAACAATTTAATGATTTATACCGGTTCC
 A  M  T  F  S  N  L  E  N  A  I  I  S  S  P  H  L  F  D  T  I  N  N  L  M  I  Y  T  G  S
                        1111                            1141                            1171
TTTTCCGTTCACCTAACCAATCAATTAATTGAAGGGTGGATTGGACATTCTGTAACTAGTAGTTTGTTGGCCAGTGGACCAACAACAGTA
 F  S  V  H  L  T  N  Q  L  I  E  G  W  I  G  H  S  V  T  S  S  L  L  A  S  G  P  T  T  V
                        1201                            1231                            1261
CTGAGAAGAAATTACGGTAGCACGACATCTATTGTAAACTATTTTAGTTTTAATGATCGTGATGTTTATCAGATTAATACGAGATCACAT
 L  R  R  N  Y  G  S  T  T  S  I  V  N  Y  F  S  F  N  D  R  D  V  Y  Q  I  N  T  R  S  H
                        1291                            1321                            1351
ACTGGGTTGGGATTCCAGAACGCACCTTTATTTGGAATCACTAGAGCTCAATTTTACCCAGGTGGGACTTATTCAGTAACTCAACGAAAT
 T  G  L  G  F  Q  N  A  P  L  F  G  I  T  R  A  Q  F  Y  P  G  G  T  Y  S  V  T  Q  R  N
                        1381                            1411                            1441
GCATTAACATGTGAACAAAATTATAATTCAATTGATGAGTTACCGAGCCTAGACCCAAATGAACCTATCAGTAGAAGTTATAGTCATAGA
 A  L  T  C  E  Q  N  Y  N  S  I  D  E  L  P  S  L  D  P  N  E  P  I  S  R  S  Y  S  H  R
                        1471                            1501                 block3     1531
TTATCTCATATTACCTCCTATTTGCATCGTGTATTGACTATTGATGGTATTAATATATATTCAGGAAATCTCCCTACTTATGTATGGACC
 L  S  H  I  T  S  Y  L  H  R  V  L  T  I  D  G  I  N  I  Y  S  G  N  L  P  T  Y  V  W  T
          block3                1561                            1591                            1621
CATCGCGATGTGGACCTTACAAACACGATTACCGCAGATAGAATTACACAACTACCATTGGTAAAGTCATTTGAAATACCTGCGGGTACT
 H  R  D  V  D  L  T  N  T  I  T  A  D  R  I  T  Q  L  P  L  V  K  S  F  E  I  P  A  G  T
                        1651          block3            1681                            1711
ACTGTCGTAAGAGGACCAGGTTTTACAGGAGGGGATATACTCCGAAGAACAGGGGTTGGTACATTTGGAACAATAAGGGTAAGGACTACT
 T  V  V  R  G  P  G  F  T  G  G  D  I  L  R  R  T  G  V  G  T  F  G  T  I  R  V  R  T  T
                  block4      1741                            1771                            1801
GCCCCCTTAACACAAAGATATCGCATAAGATTCCGTTTCGCTTCTACCACAAATTTGTTCATTGGTATAAGAGTTGGTGATAGACAAGTA
 A  P  L  T  Q  R  Y  R  I  R  F  R  F  A  S  T  T  N  L  F  I  G  I  R  V  G  D  R  Q  V
                        1831                            1861                            1891
AATTATTTTGACTTCGGAAGAACAATGAACAGAGGAGATGAATTAAGGTACGAATCTTTTGCTACAAGGGAGTTTACTACTGATTTTAAT
 N  Y  F  D  F  G  R  T  M  N  R  G  D  E  L  R  Y  E  S  F  A  T  R  E  F  T  T  D  F  N
```

Fig. 2 (*For caption, see next page*).

```
                                1921                                1951 block5                                1981
TTTAGACAACCTCAAGAATTAATCTCAGTGTTTGCAAATGCATTTAGCGCTGGTCAAGAAGTTTATTTTGATAGAATTGAGATTATCCCC
 F  R  Q  P  Q  E  L  I  S  V  F  A  N  A  F  S  A  G  Q  E  V  Y  F  D  R  I  E  I  I  P
                                2011                                2041                                2071
GTTAATCCCGCACGAGAGGCGAAAGAGGATCTAGAAGCAGCAAAGAAAGCGGTGGCGAGCTTGTTTACACGCACAAGGGACGGATTACAA
 V  M  P  A  R  E  A  K  E  D  L  E  A  A  K  K  A  V  A  S  L  F  T  R  T  R  D  G  L  Q
                                2101                                2131                                2161
GTAAATGTGAAAGATTATCAAGTCGATCAAGCGGCAAATTTAGTGTCATGCTTATCAGATGAACAATATGGGTATGACAAAAAGATGTTA
 V  N  V  K  D  Y  Q  V  D  Q  A  A  N  L  V  S  C  L  S  D  E  Q  Y  G  Y  D  K  K  M  L
                                2191                                2221                                2251
TTGGAAGCGGTACGCGCGGCAAAACGCCTCAGCCGAGAACGTAACTTACTTCAGGATCCAGATTTTAATACAATCAATAGTACAGAAGAA
 L  E  A  V  R  A  A  K  R  L  S  R  E  R  N  L  L  Q  D  P  D  F  N  T  I  N  S  T  E  E
                                2281                                2311                                2341
AATGGATGGAAAGCAAGTAACGGCGTTACTATTAGTGAGGGCGGTCCATTCTATAAAGGCCGTGCACTTCAGCTAGCAAGTGCACGAGAA
 N  G  W  K  A  S  N  G  V  T  I  S  E  G  G  P  F  Y  K  G  R  A  L  Q  L  A  S  A  R  E
                                2371                                2401                                2431
AATTATCCAACATACATTTATCAAAAAGTAGATGCATCGGAGTTAAAACCTTATACACGTTATAGATCAGATGGGTTCGTGAAGAGTAGT
 N  Y  P  T  Y  I  Y  Q  K  V  D  A  S  E  L  K  P  Y  T  R  Y  R  S  D  G  F  V  K  S  S
                                2461                                2491                                2521
CAAGATTTAGAAATTGATCTCATTCACCATCATAAAGTCCATCTTGTGAAAAATGTACCAGATAATTTAGTATCTGATACTTACCCAGAT
 Q  D  L  E  I  D  L  I  H  H  H  K  V  H  L  V  K  N  V  P  D  N  L  V  S  D  T  Y  P  D
                                2551                                2581                                2611
GATTCTTGTAGTGGAATCAATCGATGTCAGGAACAACAGATGGTAAATGCGCAACTGGAAACAGAGCATCATCATCCGATGGATTGCTGT
 D  S  C  S  G  I  N  R  C  Q  E  Q  Q  M  V  N  A  Q  L  E  T  E  H  H  H  P  M  D  C  C
                                2641                                2671                                2701
GAAGCAGCTCAAACACATGAGTTTTCTTCCTATATTGATACAGGGGATTTAAATTCGAGTGTAGACCAGGGAATCTGGGCGATCTTTAAA
 E  A  A  Q  T  H  E  F  S  S  Y  I  D  T  G  D  L  N  S  S  V  D  Q  G  I  W  A  I  F  K
                                2731                                2761                                2791
GTTCGAACAACCGATGGTTATGCGACGTTAGGAAATCTTGAATTGGTAGAGGTCGGACCGTTATCGGGTGAATCTTTAGAACGTGAACAA
 V  R  T  T  D  G  Y  A  T  L  G  N  L  E  L  V  E  V  G  P  L  S  G  E  S  L  E  R  E  Q
                                2821                                2851                                2881
AGGGATAATACAAAATGGAGTGCAGACTAGGAAGAAAGCGTGCAGAAACAGATCGCGTGTATCAAGATGCCAAACAATCCATCAAT CAT
 R  D  N  T  K  W  S  A  E  L  G  R  K  R  A  E  T  D  R  V  Y  Q  D  A  K  Q  S  I  N  H
                                2911                                2941                                2971
TTATTTGTGGATTATCAAGATCAACAATTAAATCCAGAAATAGGGATGGCAGATATTATGGACGCTCAAAATCTTGTCGCATCAATTTCA
 L  F  V  D  Y  Q  D  Q  Q  L  N  P  E  I  G  M  A  D  I  M  D  A  Q  N  L  V  A  S  I  S
                                3001                                3031                                3061
GATGTATATAGCGATGCCGTACTGCAAATCCCTGGAATTAACTATGAGATTTACACAGAGCTGTCCAATCGCTTACAACAAGCATCGTAT
 D  V  Y  S  D  A  V  L  Q  I  P  G  I  N  Y  E  I  Y  T  E  L  S  N  R  L  Q  Q  A  S  Y
                                3091                                3121                                3151
CTGTATACGTCTCGAAATGCGGTGCAAAATGGGGACTTTAACAACGGGCTAGATAGCTGGAATGCAACAGCGGGTGCATCGGTACAACAG
 L  Y  T  S  R  N  A  V  Q  N  G  D  F  N  N  G  L  D  S  W  N  A  T  A  G  A  S  V  Q  Q
                                3181                                3211                                3241
GATGGCAATACGCATTTCTTAGTTCTTTCTCATTGGGATGCACAAGTTTCTCAACAATTTAGAGTGCAGCCGAATTGTAAATATGTATTA
 D  G  N  T  H  F  L  V  L  S  H  W  D  A  Q  V  S  Q  Q  F  R  V  Q  P  N  C  K  Y  V  L
                                3271                                3301                                3331
CGTGTAACAGCAGAGAAAGTAGGCGGCGGAGACGGATACGTGACTATCCGGGATGGTGCTCATCATACAGAAACGCTTACATTTAATGCA
 R  V  T  A  E  K  V  G  G  G  D  G  Y  V  T  I  R  D  G  A  H  H  T  E  T  L  T  F  N  A
                                3361                                3391                                3421
TGTGATTATGATATAAATGGCACGTACGTGACTGATAATACGTATCTAACAAAAGAAGTGATATTCTATTCACATACAGAACACATGTGG
 C  D  Y  D  I  N  G  T  Y  V  T  D  N  T  Y  L  T  K  E  V  I  F  Y  S  H  T  E  H  M  W
                                3451                                3481                                3511
GTAGAGGTAAATGAAACAGAAGGTGCATTTCATATAGATAGTATTGAATTCGTTGAAACAGAAAAGTAACGGGATGATGTTCCGAACATA
 V  E  V  N  E  T  E  G  A  F  H  I  D  S  I  E  F  V  E  T  E  K  *
                                3541                                3571                                3601
TAAGGTATAAGGACGATACGCCGTATAAAAGATTCTCAAACAGAATGTGAAATAAATGAGGACCCCTCCGGGTAGTCGTACATGGAAAAT
                                3631                                3661                                3691
ACACAGACTACCCGGAGGTATTTTTTATATAAAAAATGTGGTTTTTCACTACGGTATATCGAAAGAATTACAGTTATGAGCGAATCAAAA
                                3721                                3751                                3781
TGAATAGAAAGCAGGATTCTGTACCTAAGACATATAAAATAAGAAGAGGGCGCATTGAAATAAGGATAGGAATAAATGTATTCTCATTTG
                                3811                                3841                                3871
AAAGTTTAGTCTGCAAGGGAAATATCTAAATTGCCTATTGGAGGGAATGGGATACAAAAGAAGTGTGCTACCTATAGACAAGAATTTCAG
                                3901                                3931                                3961
TAGGATAAATCTATGGATTTGAGGTAATAATAAAATGGTGGGACAAAAATGGTTATTTCATCCCACCGTTATATACAGGAATAAGGAATT
                                3991  HindIII
ATATTCGCAAATGATTTGTTAATATCTGTTCCTAAGCTT
```

Fig. 2 (*continued*). The nucleotide sequence of the 4,000 bp *Eco*RI–*Hin*dIII segment of the insert in the pOC10 recombinant phasmid from the genomic library of *Bacillus thuringiensis* spp. *galleriae*. The position of relevant restriction enzyme sites, the same as in Fig. 1, are indicated. In the deduced amino acid sequence of *cryX* the 5 highly conserved regions (blocks) are delineated.

*cry*IG indicates clearly the participation of recombination events in the origination of these genes.

Acknowledgements: This work was partly supported by a grant from the Russian State Program 'Frontiers in Genetics'.

REFERENCES

[1] Hofte, H. and Whiteley, H.R. (1989) Microbiol. Rev. 53, 242–255.
[2] Chestukhina, G.G., Kostina, L.I., Zalunin, I.A., Khodova, O.M. and Stepanov, V.M. (1988) FEBS Lett. 232, 249–251.
[3] Osterman, A.L., Karasin, A.I., Zagnitko, O.P., Kaluger, S.V., Chestukhina, G.G., Fonstein, M.Yu., Yankovsky, N.K. and Stepanov, V.M. (1989) Mol. Biol. (USSR) 23, 463–472.
[4] Smulevitch, S.V., Osterman, A.L., Shevelev, A.B., Kaluger, S.V., Karasin, A.I., Kadyrov, R.M., Zagnitko, O.P., Chestukhina, G.G. and Stepanov, V.M. (1991) FEBS Lett. 293, 25–28.
[5] Tabor, S. and Richardson, C.C. (1987) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[6] Brizzard, B.L. and Whiteley, H.R. (1988) Nucleic Acids Res. 16, 4168–4169.