Volume 293, number 1,2, 25–28 FEBS 10374 November 1991

Nucleotide sequence of a novel δ-endotoxin gene *cryIg* of *Bacillus thuringiensis* ssp. *galleriae*

S.V. Smulevitch, A.L. Osterman, A.B. Shevelev, S.V. Kaluger, A.I. Karasin, R.M. Kadyrov, O.P. Zagnitko, G.G. Chestukhina and V.M. Stepanov

Institute of Microbial Genetics, Lab. of Protein Chemistry, Moscow, 113545, USSR, 1st Dorozhny, 1

Received 18 August 1991

A gene *cryIg* coding for entomocidal protein δ-endotoxin of *Bacillus thuringiensis* ssp. *galleriae* str. 11-67 named CryIG has been cloned and sequenced (EMBL accession number X58120). The deduced amino acid sequence that contains 1156 amino acid residues shows only 28% of identical residues, when compared with other δ-endotoxins of the CryI family. The extent of identity is substantially higher for some regions of the sequence ('conserved blocks'), that presumably bear important structural or functional properties. This implies that CryIG δ-endotoxin follows the same type of polypeptide chain folding as other CryI proteins, whereas peculiarities of primary structure help to explain its unique specificity.

CryIG; δ-Endotoxin; Primary structure; Genomic library; *Bacillus thuringiensis*

1. INTRODUCTION

Entomocidal protein δ-endotoxins are intensively studied in many laboratories. Their ability to selectively kill only certain species of insects creates a basis for their utilization in plant protection. Nevertheless the mode of the toxic action remains unknown.

Proteinaceous crystals of *B. thuringiensis* spp. *galleriae* str. 11-67 are composed of at least two δ-endotoxins of M_r ~130 kDa differing strongly in their immunological properties and entomocidal specificity, named 'positive' and 'negative' components according to the electrophoretic behavior of the respective true toxins [1]. Earlier we reported cloning of the gene coding for the 'positive' component, the endotoxin with unique specificity towards larvae of *Galleria mellonella* [2]. The partial amino acid sequence of this protein was determined by Edman's method. Here we present the complete sequence of this gene named *cryIg*, according to the principles of δ-endotoxin classification proposed in [3].

2. METHODS

2.1. cryIg *gene cloning*

A genomic library of *Bacillus thuringiensis* ssp. *galleriae* str. 11-67 was obtained on phasmid vector λ-pSL5, especially designed for procaryotic genomic library construction, using partial *Eco*RI digestion of total DNA [5]. Immunoscreening was performed with affinity-purified monospecific antiserum against pure CryIG protein. Restriction mapping of three immunopositive clones and preliminary localization of the toxin gene were reported earlier [1]. pOC10 phasmid, containing the full-size gene was used in further cloning procedures. The phage part of the pOC10 was removed by *Xba*I deletion, yielding pOK10 plasmid. Exact mapping and localization of the *cryIg* gene were performed using immunological tests (Fig. 1). *Kpn*I–*Xba*I insert, carrying the full-size gene, the 521 base-pair 5′-flank and the 1.3 kb 3′-flank was cloned into the pUC18 cloning vector, resulting in the

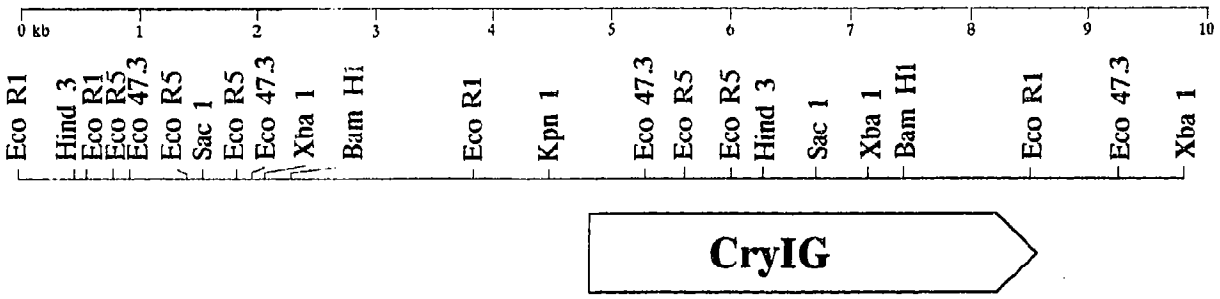
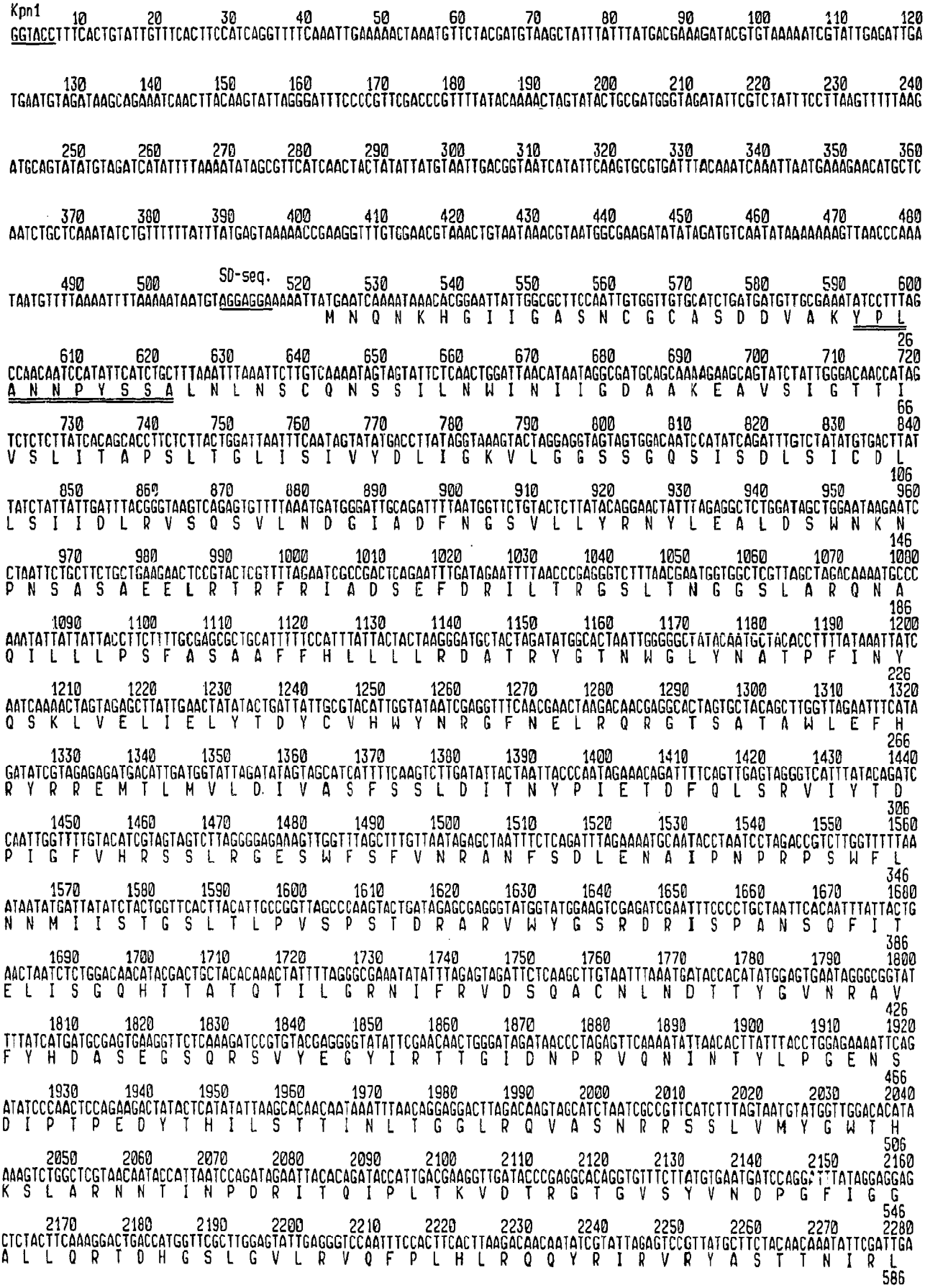


Fig. 1. Structural organization of the insert of pOK10 plasmid, containing full-size CryIG gene.

Correspondence address: A.L. Osterman, Institute of Microbial Genetics, Lab. of Protein Chemistry, Moscow, 113545, USSR, 1st Dorozhny, 1. Fax: (7) (095) 315 05 01.


Kpn1
SD-seq.

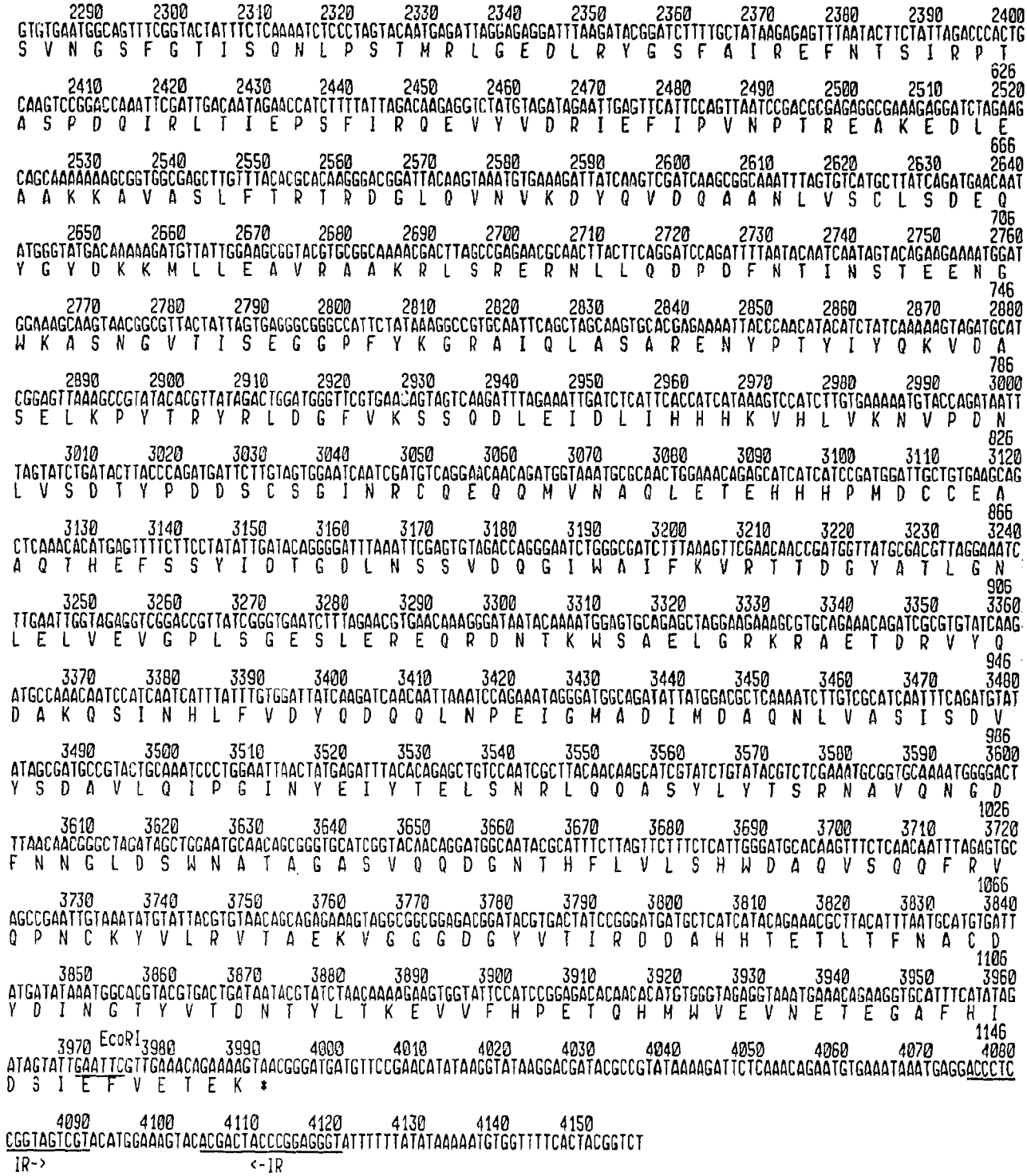


Fig. 2. The nucleotide sequence of a 4156 bp fragment of *Bacillus thuringiensis* ssp. *galleriae* str. 11-67 DNA and the deduced amino acid sequence of CryIG δ-endotoxin. The N-terminal amino acid sequence determined with the Edman method is doubly underlined. Inverted repeat in 3'-flanking region is underlined. Also shown are the SD-sequence, *Kpn*I and *Eco*RI sites.

pKP7 plasmid, which allowed expression of *cry1g* gene in *E. coli* using the Lac-promoter. Immunological and toxicity assays of the expression product confirmed its identity with CryIG toxin.

2.2. *Sequence analysis*

Sequencing of pKP7 was carried out by the SEQUENASE version [4] of Sanger dideoxynucleotide method using a number of subclones in single- or double-strand form obtained in pUC or M13 vectors. The

sequence protocol included use of standard and custom synthetic primers. Sequence data were submitted to EMBL Data Library – accession number X58120.

3. RESULTS AND DISCUSSION

The phasmid library on the λ-pSL5 vector has shown good expression properties, producing a quantity of protein, sufficient for immunoscreening in the phage form, although *Eco*RI site of λ-pSL5 is not situated in the region transcribed from the Lac-promoter. The strategy based on primary cloning of long inserts with consequent localization and subcloning of the full-size gene appeared to be productive.

Analysis of the DNA-derived amino-acid sequence shows its identity with the N-terminal fragment determined with Edman's method (Fig. 2). The sequence contains an open reading frame of 1156 codons with the AUG initiator codon and TAA terminator ochre-codon. The calculated M_r of the protein product is 129 740 kDa; experimental measurement of the M_r of CryIG has given essentially the same value [1].

Computer assay of 5'-non-coding region of the CryIG gene failed to reveal sequences similar to known bacillar promoters, which might imply that expression of the *cryIg* gene proceeds via polycystronic mRNA. The Shine–Delgarno sequence GGAGGA was found 7 base pairs upstream from the initiator codon (Fig. 2). In the 3'-non-coding region 85 bp downstream from terminator codon,a 17-bp inverted repeat was found. The hairpin structure that might be presumed for this repeat cannot be considered a good terminator because the loop part of it is too long, i.e. 13 base pairs.

CryIG differs markedly from all other members [3] of the CryI family. Its primary structure reveals only a marginal extent of identity when compared with other known members of the CryI family; it does not exceed 21% for the N-terminal half (1–670 bp) and 35% for the C-terminal half (671–1156 bp). The identical residues are not evenly distributed along the sequence forming pronounced stretches. In particular, all five conserved blocks of amino-acid residues, common for all CryI δ-endotoxins, have also been found in the CryIG sequence. Characteristically, the identity extent for these blocks is substantially higher, i.e. 60–80%.

A more detailed discussion of these data, especially of the alignment of sequences of CryI proteins will appear elsewhere.

REFERENCES

[1] Chestukhina, G.G., Kostina, L.I., Zalunin, I.A., Khodova, O.M. and Stepanov, V.M. (1988) FEBS Lett. 232, 249–251.
[2] Osterman, A.L. et al. (1989) Mol. Biol. (USSR) 23, 463–472.
[3] Hofte, H. and Whiteley, H.R. (1989) Microbiol. Revs. 53, 242–255.
[4] Tabor, S. and Richardson, C.C. (1987) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[5] Yankovsky, N.K., Fonstein, M.Yu. and Lashina, S.Y. (1989) Gene 81, 203–207.