# Amino acid sequences from the N-terminal domain of *Bacillus thuringiensis*, subspecies *alesti*, δ-endotoxin

## Hypervariable regions of *Bacillus thuringiensis* δ-endotoxins

G.G. Chestukhina, S.A. Tyurin, A.L. Osterman, O.P. Khodova, V.M. Stepanov

*Institute of Genetics and Selection of Industrial Microorganisms, 113545 Moscow, USSR*

Received 20 January 1986

The tryptic peptides derived from the N-terminal domain of δ-endotoxin produced by *Bacillus thuringiensis* ssp. *alesti* have been sequenced and aligned by homology with the primary stuctures of δ-endotoxins formed by *B.t. kurstaki* subspecies (K-1 and K-73). The N-terminal regions of the domains (residues 1–346) show only marginal variations: 9 amino acid substitutions were found among 158 aligned residues of *B.t.* ssp. *alesti* δ-endotoxin. In contrast, the following stretch of the sequence starting from residue 347 reveals ~ 50% of amino acid substitutions when compared with the structures of K-1 or K-73. As a rule, the fragments of *B.t.* ssp. *alesti* δ-endotoxin that differ from K-1 in structure precisely reproduce the corresponding sequences of K-73 δ-endotoxin and, vice versa, those different from the K-73 sequences coincide with the respective structures of K-1 δ-endotoxin. A few fragments present in *B.t.* ssp. *alesti* were also found but not in both *B.t.* ssp. *kurstaki* δ-endotoxins. It appears that the structure of this region of *B.thuringiensis* δ-endotoxins (347–625 residues) contains a pattern of relatively conservative and highly variable fragments. These variations might be responsible for the variations in specificity of the entomocidal action of δ-endotoxins.

(Bacillus thuringiensis) *Entomocidal crystal* *Insecticide* *δ-Endotoxin* *Primary structure*

## 1. INTRODUCTION

*Bacillus thuringiensis* produces crystals of protein – δ-endotoxin – that possesses insecticidal activity. Among 25 subspecies of *B.thuringiensis*, several are known to produce δ-endotoxins differing in specificity of action on various Lepidoptera or Diptera. At the same time, the results of limited proteolysis indicate that all these proteins share the same type of domain organization of their molecules [1]. Hence, one might presume that rather slight differences in their structures, probably those in the primary structures, are responsible for these variations in biological specificity of δ-endotoxins.

This paper describes the structural features of the N-terminal domain of δ-endotoxin produced by subspecies *alesti* of *B.thuringiensis*. Extensive hydrolysis of the δ-endotoxin with subtilisin BPN′ yielded fragments, rather stable to further degradation, derived from the N-terminal half of the molecule corresponding to the 'true toxin'. These fragments with $M_r$ close to 10000 were subjected to tryptic hydrolysis without preliminary separation. The tryptic peptides were isolated and sequenced, and then their structures compared with known sequences of two *B.t.* ssp. *kurstaki* δ-endotoxins, K-1 and K-73, established by sequencing the respective structural genes [2,3]. This allowed us to reveal the characteristic distribution of conservative and highly variable stretches along the sequences of these δ-endotoxins. These observations appear to be of general importance and might serve as a basis for explaining variations in the biological specificity of δ-endotoxins produced by various subspecies of *B.thuringiensis*.

*Published by Elsevier Science Publishers B.V. (Biomedical Division)*
00145793/86/$3.50 

## 2. MATERIALS AND METHODS

Crystal-forming protein produced by *B.t.* ssp. *alesti* and its N-terminal domain were isolated as described in [1,4]. To obtain fragments stable to further degradation, 600 mg *B.t.* ssp. *alesti* δ-endotoxin were dissolved in 400 ml of 0.05 M NaOH and the solution (pH 12.5) incubated at 37°C for 1 h. The pH was then lowered to 10.5 with concentrated HCl; sodium carbonate-bicarbonate buffer (pH 10.5) was added to 0.01 M followed by 2.7 mg subtilisin BPN′. The mixture was incubated at 37°C for 30 min, the pH lowered to 7.5 and the enzyme inactivated by 0.01% (v/v) diisopropyl fluorophosphate. The mixture was concentrated on an Amicon PM-30 membrane to 10 ml and applied to the top of a Sephadex G-75 column (5 × 180 cm) equilibrated with 4 M urea containing 0.01 M sodium carbonate buffer (pH 8.6), 0.01 M dithiothreitol, 1 mM $Na_2EDTA$ and 1% SDS. A fraction (840–960 ml) of the eluate was collected, rechromatographed twice under the same conditions, dialyzed against water and lyophilized. As shown by SDS electrophoresis in polyacrylamide gels, this fraction contained peptide material corresponding to $M_r$ 10000. This material was treated with 90% phenol for 12 h at 37°C to ensure complete unfolding, then hydrolysed with trypsin (1:20, v/v) for 36 h at pH 9.5 and 37°C. The course of hydrolysis was checked by SDS electrophoresis and carboxypeptidase B digestion [5,6]. The hydrolysate was fractionated on a Chromobeads (Technicon, Ireland) cation exchanger column [4]. The fractions obtained were further purified by HPLC on a reversed-phase column (ODS-Hypersil, Shandon) in an acetonitrile concentration gradient at pH 5.5. Amino acid analysis and sequencing of the peptides were performed as in [4,7].

## 3. RESULTS AND DISCUSSION

The N-terminal domain that contains ~700 amino acid residues corresponds to the true toxin, which appears after entomocidal protein ('protoxin') hydrolysis with various proteinases including those present in larva gut. Automated sequencing of the true toxin formed by the action of trypsin on the δ-endotoxin produced by *B.t. alesti* subspecies showed that simultaneously with limited proteolysis of the hinge region, joining this domain with the C-terminal half of δ-endotoxin, the N-terminal stretch of 28 amino acid residues was split off. Apparently, the loss of this N-terminal fragment has a negligible effect on the entomocidal activity of the N-terminal domain. The following residues, 29–57, were found to be clearly homologous to the respective sequences of both δ-endotoxins formed by *B.t.* ssp. *kurstaki* (K-1, K-73) (fig.1).

Fig.1. Amino acid sequence of the N-terminal halves of *B.thuringiensis* δ-endotoxins from subspecies *kurstaki* K-1 and K-73 and aligned peptides of *B.t.* ssp. *alesti* δ-endotoxin (Ale).

Whereas the δ-endotoxins produced by *B.t.* ssp. *kurstaki* (K-1) and *sotto* [8] have been found to be almost identical – only 4 amino acid substitutions observed among 900 residues compared – the difference between the primary structures of *B.t.* ssp. *kurstaki* K-1 and K-73 δ-endotoxins was substantially greater. These proteins, which also differ according to immunochemical criteria, differ by 171 substitutions out of 1176 amino acid residues in each of their polypeptide chains (15%). Remarkably, all these substitutions are localized in the N-terminal halves of the molecules, i.e. within the true toxin domains, whereas their C-terminal halves containing ~600 amino acid residues are strictly identical.

The distribution of the amino acid substitutions within the N-terminal domains is also strikingly uneven. The most variable is the C-terminal part of this domain embracing residues 347–625 and containing ~90% of all substitutions. It appears to be particularly important that the amino acid substitutions are not randomly dispersed along this region but clustered within several highly variable stretches of 7–28 residues in length. Other stretches of both sequences are nearly identical.

Comparison of the tryptic peptides isolated from *B.t.* ssp. *alesti* δ-endotoxin with the sequences of both *B.t.* ssp. *kurstaki* δ-endotoxins led us to the following conclusions. Almost all peptides might be localized within the N-terminal domain according to homology with either sequence of *B.t.* ssp. *kurstaki* δ-endotoxins. The sequences of the peptides belonging to the N-terminal 346-residue stretch of this domain are very similar to those of both *B.t.* ssp. *kurstaki* δ-endotoxins (K-1, K-73) – only 9 substitutions found for 158 residues that might be aligned. Hence, this part of the molecule appears to be rather conservative. The most variable is the sequence starting from residue 347 but, as a rule, these variations in the structure of *B.t.* ssp. *alesti* δ-endotoxin reproduce the sequence of either *B.t.* ssp. *kurstaki* δ-endotoxin. Thus, the sequence of *B.t.* ssp. *alesti* δ-endotoxin coincides precisely with that of *B.t.* ssp. *kurstaki* K-73 on at least two long stretches where the K-1 and K-73 sequences differ markedly (residues 347–359 and 436–450). Even the insertion of one residue (351) in the first sequence and two (448, 449) in the second is copied in *B.t.* ssp. *alesti* protein. In contrast, sequences 502–529 of K-1 and *B.t.* ssp. *alesti* toxin are identical, both differing strongly from that of K-73 δ-endotoxin.

These findings allow us to discuss the following mechanism of variations in δ-endotoxin structures and, possibly, in their biological activity. We presume that the patterns of molecular organization of the δ-endotoxins are very similar. This is especially true for the C-terminal half of the molecule as well as for its N-terminal fragment embracing the first 350 amino acid residues. The remaining variable part of the structure, residues 350–625, showed 50% homology when the sequences of 3 δ-endotoxins were compared. This degree of homology is large enough to presume that the folding pattern is preserved in this part of the molecule, although large differences in the sequences of certain hypervariable stretches suggest some specific modification of the secondary or tertiary structure acquired by this segment. It is significant that the several stretches localized within this highly variable region of δ-endotoxin structures are very different in their sequences showing no homology, or only a marginal degree thereof.

One might suggest that these hypervariable sequences, e.g. 347–359, 438–450 or 502–529, could be of crucial importance for the variations in biological specificity, especially in the host range characteristic for the different δ-endotoxins. Comparison of the K-1 and K-73 sequences shows that these toxins have 8 such sequences. Three hypervariable sequences, one identical with K-1 (502–529) and two identical with K-73 (347–359 and 438–450), were found among the tryptic peptides of *B.t.* ssp. *alesti* δ-endotoxin. In addition, it should to be noted that certain peptides of *B.t.* ssp. *alesti* δ-endotoxin tryptic hydrolysate (GLDPATTR, LWLGL, LWGTYTDYAFGR) could not be aligned with *B.t.* ssp. *kurstaki* sequences. Apparently, *B.t.* ssp. *alesti* δ-endotoxin, in addition to the hypervariable sequences that copy the respective stretches of K-1 and K-73 toxins, also possesses fragments that might correspond to the hypervariable sequences of other δ-endotoxins, which have not yet been studied structurally.

The above data indicate that these hypervariable sequences might recombine independently within different δ-endotoxin molecules. If this assumption were valid, it would explain the adaptive

variations in biological specificity of *B. thuringiensis* δ-endotoxins.

## REFERENCES

[1] Chestukhina, G.G., Kostina, L.I., Mikhailova, A.L., Tyurin, S.A., Klepikova, F.S. and Stepanov, V.M. (1982) Arch. Microbiol. 132, 159–162.
[2] Schnepf, H.E., Wong, H.C. and Whiteley, H.R. (1985) J. Biol. Chem. 258, 6264–6273.
[3] Adang, M.J., Staver, M.J., Rocheleau, T.A., Leighton, J., Bareker, R.F. and Thompson, D.V. (1985) Gene 36, 289–300.
[4] Zalunin, I.A., Levin, E.D. Timokhina, E.A., Chestukhina, G.G. and Stepanov, V.M. (1982) Biokhimia 47, 1570–1575.
[5] Weber, K. and Osborn, M. (1969) J. Biol. Chem. 224, 4406–4412.
[6] Kotlova, E.K., Ostoslavskaya, V.I., Revina, L.P., Kovaleva, G.G., Surova, I.A., Levin, E.D., Timokhina, E.A. and Stepanov, V.M. (1981) Bioorg. Khim. 7, 58–67.
[7] Chang, J.-Y. (1983) Methods Enzymol. 91, 79–83.
[8] Shibano, Y., Yamagata, A., Nakamura, N, Iizuka, T., Sugisaki, H. and Takanami, M. (1985) Gene 34, 243–251.