Volume 13 Number 22 1985 Nucleic Acids Research

Molecular cloning and the nucleotide sequence of the M_r 28 000 crystal protein gene of *Bacillus thuringiensis* subsp. *israelensis*

Cees Waalwijk*, Annette M.Dullemans, Mariëlle E.S.van Workum and Bert Visser

Research Institute ITAL, P.O. Box 48, 6700 AA Wageningen, The Netherlands

Received 27 August 1985; Revised and Accepted 21 October 1985

ABSTRACT

The M_r 28.000 crystal protein gene of Bacillus thuringiensis subspecies israelensis has been cloned into pBR322 as part of a 9.7 kb HindIII fragment. From hybridization experiments of recombinant p425 DNA with B.t. subspecies israelensis RNA from different stages of growth it was concluded that transcription of the gene is restricted to early sporulation stages. Nucleotide sequence analysis revealed the presence of a large open reading frame with a coding capacity of 249 amino acids (M_r 27.340). Nuclease S1 mapping demonstrated that transcription starts 44 nucleotides upstream of the initiation codon. A Shine-Dalgarno sequence (AAGGAG) was found 10 nucleotides upstream of the translation startpoint. At the 3'-end of the gene a complex secondary structure was found immediately after the stop-codon. Despite the presence of these regulation signals only limited expression in E. coli was detected. This can be explained by assuming that B.t. subsp. israelensis promotor sequences are poorly recognized by E. coli RNA polymerase.

INTRODUCTION

Environmental and health hazards of chemical insecticides and the appearance of resistance against these compounds have urged research for biological alternatives such as Bacillus thuringiensis (B.t.). The majority of B.t. subspecies produce parasporal crystals that are composed of a single protein (M_r 130.000 - 140.000) that is toxic for lepidopteran insect larvae (1).
In contrast, B.t. subsp. israelensis (B.t.i.) produces crystals that are toxic to the larvae of dipteran insects, including mosquitoes and blackflies that transmit devastating diseases like malaria, yellow fever and river-blindness. The protein crystals of B.t.i. are different from those of most other B.t. subspecies, since they are composed of several proteins ranging in size from M_r 130.000 to M_r 28.000 (2, 3). The assignment of toxic activities to particular proteins is a matter of current debate, mainly because of the difficulties encountered in the purification of the different proteins. Recently, the protein of M_r 28.000 has been purified successfully by gel filtration (4, 5). In our lab, proteins of M_r 130.000 and M_r 28.000

have been purified by sucrose gradient centrifugation (Visser et al., submitted for publication). The M_r 130.000 protein possesses larvicidal activity, whereas the M_r 28.000 protein is hemolytic in vitro. Two reports on the cloning of B.t.i. endotoxin genes have been published (6, 7). On the basis of the results presented here it must be concluded that two different genes are involved.

In this paper we report the molecular cloning of the M_r 28.000 crystal protein gene and the complete nucleotide sequence is presented. Moreover, data will be shown suggesting that the recognition signals for the regulation of transcription of the M_r 28.000 crystal protein gene are different from those described for other Bacillus genes, expressed during sporulation.

MATERIALS AND METHODS

Enzymes and reagents

All enzymes used in this study were obtained from Boehringer (Mannheim). The in vitro transcription-translation system was purchased from Amersham (England) and used according to the protocol of the supplier. Antiserum against B.t.i. crystal protein was obtained by subcutaneous priming of a New Zealand rabbit with purified solubilized crystals followed by a boost 4 weeks later (L. Visser et al., submitted for publication).

Cloning and screening of recombinants

Total plasmid DNA from B.t.i. IPS82 (strain 1884 obtained from H. de Barjac, Pasteur Institute, Paris) was isolated according to Lereclus et al. (8). The DNA was digested with appropriate restriction enzymes, ligated to linearized pBR322 and competent E. coli RRI cells were transformed with the ligation mixtures according to standard procedures (9). Recombinants were screened for crystal protein production by SDS-PAGE analysis of immunoprecipitated material from in vitro transcription-translation assays.

Identification of developmental regulation

RNA was isolated according to a procedure developed by Srinivasan et al. (10) from synchronized B.t.i. cells harvested at various times during growth and sporulation. These RNAs were separated in formaldehyde-agarose gels, blotted to nylon membrane (Gene Screen, NEN) and hybridized with the nick-translated recombinant DNAs according to standard protocols (9).

DNA sequence analysis

Appropriate restriction fragments were labelled at the 5'-end with [γ-^{32}P]-ATP using T4 polynucleotide kinase or at the 3'-end with [α-^{32}P]ddATP using deoxynucleotidyl terminal transferase (9). The nucleotide sequence was elu-

cidated with the chemical degradation method (11). Alternatively, fragments were cloned into M13mp10 or M13mp11. The nucleotide sequences of these inserts were determined by the chain termination method (12) using [α-^{32}P]dATP and ddNTPs.

Nuclease S1 mapping

Mapping experiments with nuclease S1 were performed essentially according to standard protocols (9, 13). ^{32}P-end labelled DNA probes were mixed with 100 µg of RNA from different growth stages in 30 µl of hybridization buffer (9) and incubated for 15 min at 74 °C. Hybridization was performed for 16 h at 49 °C. The samples were diluted 30 times in S1 nuclease buffer (9) and incubated with 100 or 1000 units/ml S1 nuclease for 45 min at 37 °C. Nuclease S1 resistant hybrids were analysed on 8% polyacrylamide gels in 7M urea along with the chemical degradation products of the ^{32}P-end labelled DNA probe used.

RESULTS AND DISCUSSION

Cloning of the M_r 28.000 crystal protein gene

From plasmid-curing and plasmid-transfer experiments it has been shown that crystal production is associated with a large plasmid in *B.t.i.* (14). Therefore plasmid DNA from *B.t.i.* IPS82 was digested with several restriction enzymes and ligated to linearized pBR322. *HindIII* was included, because it was indicated that the enzyme does not recognize sequences within the toxin gene (6). Competent *E. coli* RRI cells were subsequently transformed with the ligation mixture. For the identification of the coding capacity of the recombinants we used an *in vitro* transcription-translation system, because expression levels of other *B.t.* toxin genes cloned in *E. coli* have been reported to be rather low (15, 16). Moreover the feeding behaviour of the larvae, that filter out particles from their environment might interfere with the detection of toxicity. The [^{35}S]-methionine labelled polypeptides were subsequently immunoprecipitated with antiserum raised against all proteins solubilized from the crystals (Visser *et al.*, submitted for publication).

When plasmid DNA from one of these recombinants (designated p425) was used as a template in the *in vitro* transcription-translation assay a considerable amount of radioactivity could be precipitated. This recombinant originated from a cloning experiment with *HindIII* and contained a 9.7 kb insert. SDS-PAGE analysis of the translation products encoded by p425 DNA is shown in Fig. 1. This DNA gives rise to a protein of M_r 28.000 and a somewhat

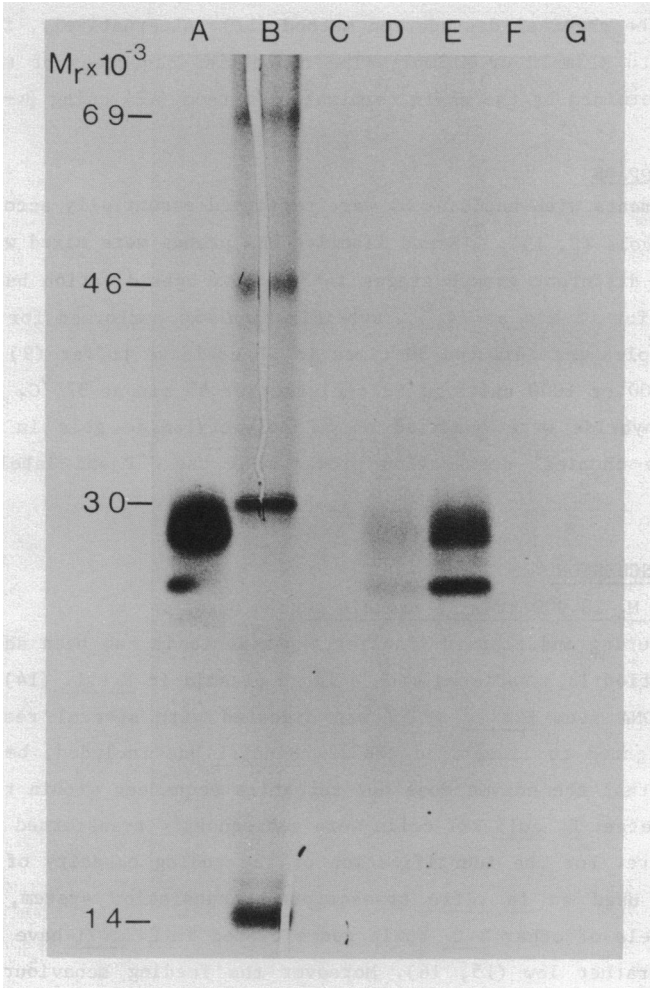

Fig. 1. SDS-PAGE analysis of immunoprecipitated polypeptides. Either intact or restricted p425 DNA was used as a template in the *in vitro* transcription-translation assays.
A, intact p425 DNA; C-G, digested p425 DNA; C, *BamHI*; D, *BglII*; E, *EcoRI*; F, *HpaI* and G, *PstI*. ^{14}C-labelled protein mixture (Amersham, England) is shown in lane B.

smaller polypeptide (M_r 25.000) that may be a breakdown product that specifically precipitates with the antiserum. These data suggest that p425 DNA encodes the M_r 28.000 crystal protein of *B.t.i.*

Regulation of the expression of the M_r 28.000 protein gene

In *B.t.i.* crystals are produced during sporulation and expression of the crystal protein genes is therefore most likely restricted to this phase of

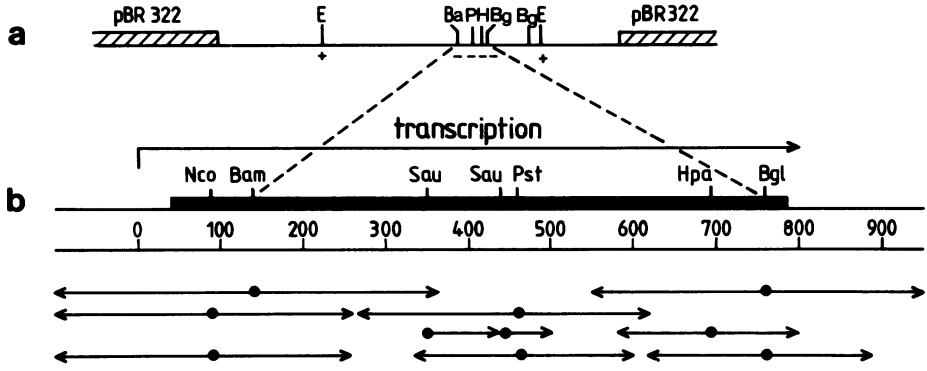

Fig. 2. Restriction enzyme cleavage map of p425. In a, the presence or absence of immunoprecipitable polypeptides after digestion of p425 DNA with the indicated enzymes is marked by + and -, respectively. In b, the location of the gene (marked by a black box) and the direction of transcription is shown by the arrow. The arrows in the bottom part indicate the direction and extent of the nucleotide sequence elucidation.

growth. Therefore, total RNA extracted from synchronized cells harvested at different stages of growth and sporulation was used for Northern hybridization analysis. Using nick-translated p425 DNA as a probe a very weak signal from a 800-900 nucleotides long mRNA was found with RNA from vegetative cells. RNA isolated from cells harvested at 4 h after the onset of sporulation gave rise to a dramatic increase of the hybridization signal and this signal decreased again during later stages of sporulation. The results from this experiment are not shown, since they can also be inferred from the results presented in Fig. 4. This result demonstrates that the M_r 28.000 crystal gene encoded by recombinant plasmid p425 is a developmentally regulated gene, most abundantly transcribed during the early stages of sporulation.

Restriction map and sequencing strategy

To localize the gene within the insert in vitro transcription-translation was performed on p425 DNA digested with one of several restriction enzymes. As can be seen in Fig. 1 digestion of the DNA with Bam HI, HpaI or PstI abolished the appearance of immunoprecipitable proteins. In contrast, no inhibition of transcription and translation was observed when the DNA was digested with EcoRI. Along with the results of single and double digestion on p425 DNA this allowed the construction of a restriction map of the gene and its flanking regions (Fig. 2).

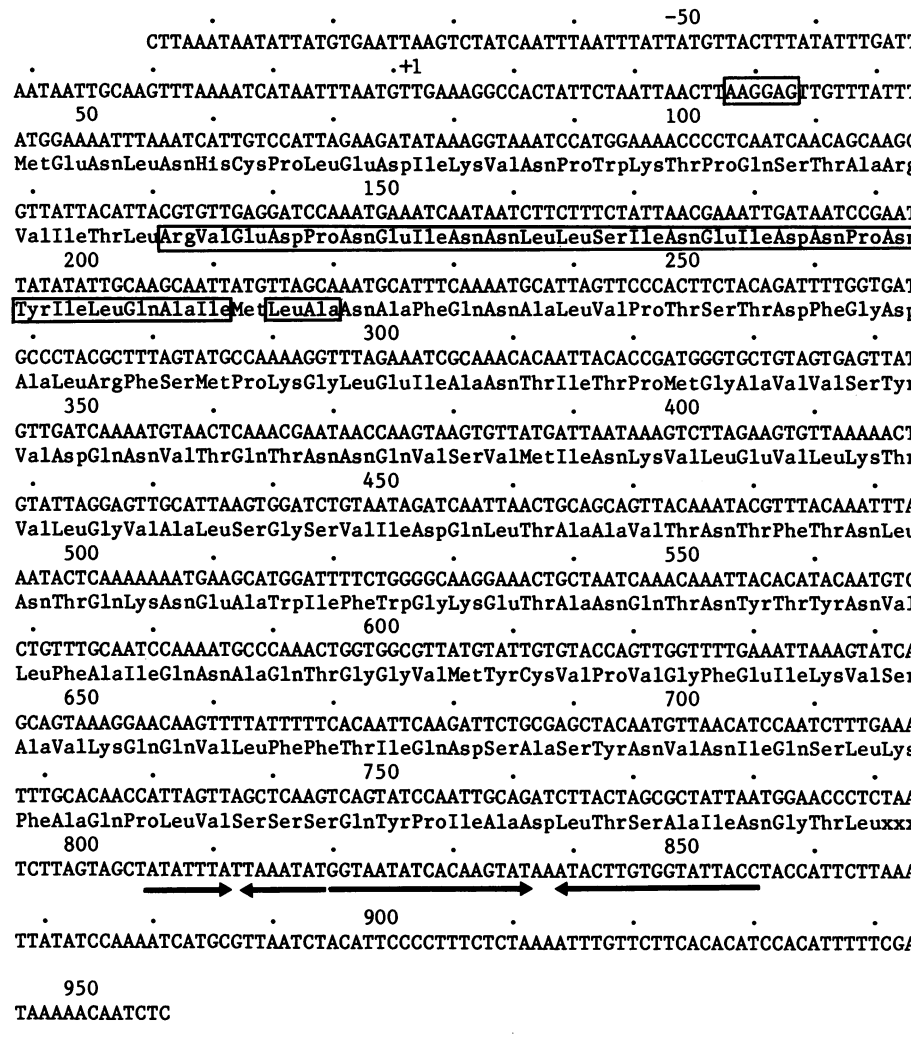

Fig. 3. The nucleotide sequence of the M_r 28.000 crystal protein gene. The amino acid sequence of the protein, deduced from the nucleotide sequence is indicated. The T-residue at position +1 marks the major transcription initiation site. The putative Shine-Dalgarno sequence (AAGGAG) is indicated by a box at positions 28 to 33. Arrows between nt805 and nt855 mark the regions of dyad symmetry at the 3'-end of the gene. Boxed amino acids indicate the region identical to the amino acid sequence determined by Armstrong et al. (22).

Nucleotide sequence analysis

The nucleotide sequence of the M_r 28.000 crystal protein gene as shown in Fig. 3 was elucidated according to the strategy depicted in Fig. 2. Inspection of the nucleotide sequence for the presence of open reading frames

(ORFs) revealed only one frame with sufficient length to specify a protein of M_r 28.000. This frame has a coding capacity for a polypeptide of 249 amino acids. The coding region starts with the initiation codon ATG (nt44-46 in Fig. 3) and ends with a TAA stopcodon (nt791-793). In the same reading frame two other stopcodons are present (nt797-799 and nt800-802, respectively).

Ten nucleotides upstream of the initiation codon a Shine-Dalgarno sequence is found (nt28-33) that has a calculated free energy of double-helix formation (ΔG_f) with 16 S rRNA from E. coli and B. subtilis of -12.8 kcal/mol (17).

This value is well within the range of values found for ribosome binding sites in E. coli but low when compared to those in B. subtilis (18).

The transcription startpoint was determined by nuclease S1 mapping with RNA isolated from cells harvested at different stages of growth and development. The results of these experiments (Fig. 4) clearly demonstrate that transcription starts 28 nucleotides upstream of the putative ribosome binding site (designated +1) with a minor transcription start at +2. Fig. 4 also illustrates the forementioned developmental regulation of the transcription of the gene. Only RNA isolated from cells harvested at 4 h or at 7 h after the onset of sporulation gave protection of the DNA probe (Fig. 4, lanes 2 and 3). In B. subtilis at least five different promotor sequences have been identified that are recognized by different σ-factors of the RNA polymerase (19). Inspection of the nucleotide sequence of the 5' proximal part of the gene revealed several nucleotide stretches that show homology with these -35 and -10 promotor consensus sequences. However, no combinations could be made that showed a correct spacing. Although low levels of transcription may occur when the -35 box and the -10 box are separated by an incorrect number of nucleotides or even in the absence of the -35 sequence (20) this is not likely to happen in the case of the M_r 28.000 crystal protein gene. The high amounts of protein found during sporulation in B.t.i. are not caused by gene dosage effects (data not shown) so either the M_r 28.000 mRNA is extremely stable or the gene is preceded by a strong promotor, recognized by a still unknown σ-factor.

Recently, a promotor sequence upstream of the genes coding for the small acid-soluble spore proteins of B. megaterium has been reported (21). This sequence does not resemble either of the hitherto identified consensus sequences.

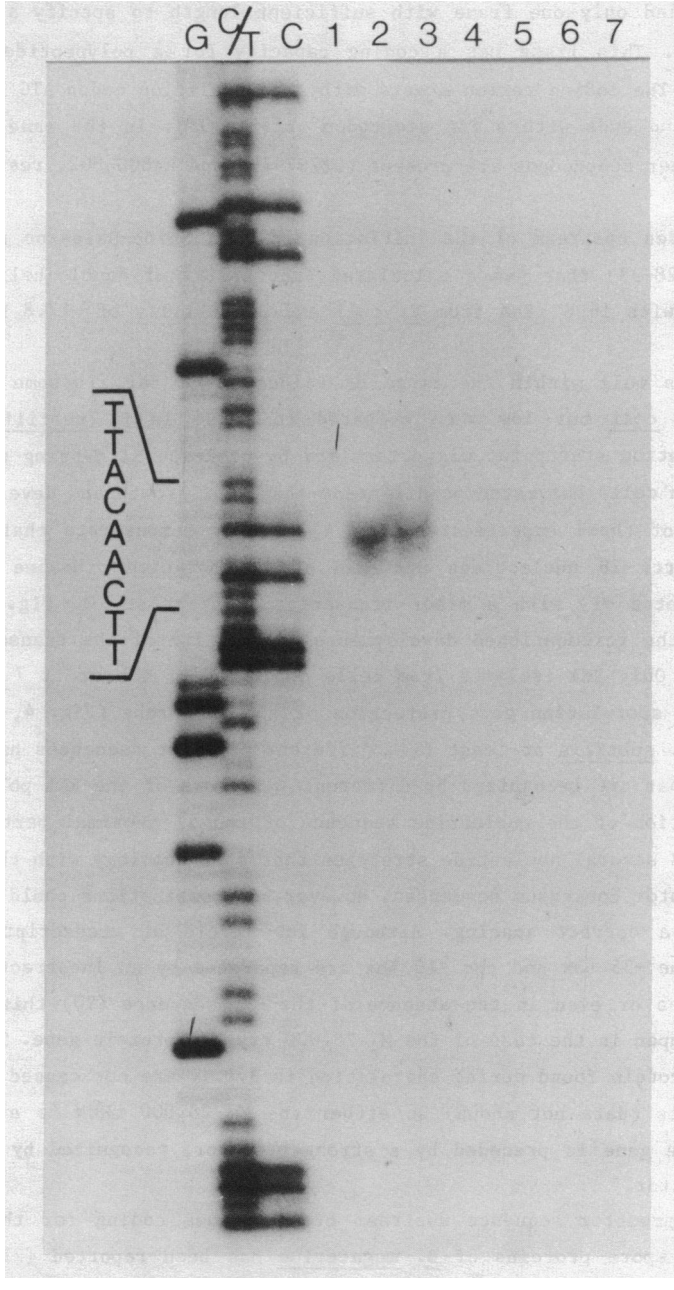

Fig. 4. Nuclease S1 mapping of the 5'-end of the M_r 28.000 mRNA. A 320-bp HinfI-NcoI fragment (nt-230 to nt90) 5'-labelled at the NcoI end was used as a probe.
Lanes 1-7, RNAs isolated from B.t.i. cells harvested at log phase

(1) 4 h (2), 7 h (3), 24 h (4), 27 h (5), 31 h (6) and 48 h (7) after the onset of sporulation, were used to protect the DNA probe. Chemical degradation products from the DNA probe specific for G, pyrimidines or C alone are shown in the lanes marked G, C/T and C, respectively. The nucleotide stretch of the anti-sense strand in the viscinity of the transcription startpoint is depicted in the figure.

At the 3' end of the gene the three stopcodons are immediately followed by a double hairpin structure (Fig. 5) which is indicative for transcription termination (22). In the 3' proximal part of the gene 4 nucleotide stretches are found that resemble the spacer consensus sequence deduced from regions between genes in E. coli as well as in B. subtilis (20). The size of the mRNA (855nt) calculated from the transcription startpoint at nucleotide +1 and assignment of the termination point at the end of the region of dyad symmetry agrees well with the size measured in the Northern hybridization analysis (800- 900 nt).

Size and composition of the crystal protein

The primary amino acid sequence deduced from the nucleotide sequence suggests a polypeptide of M_r 27.340, which is close to the size estimated by SDS-PAGE analysis. The amino acid composition deduced from the sequence agrees very well with the empirically determined amino acid composition of the M_r 28.000 protein of another B.t.i. strain (5). Moreover, the amino acid sequence from residue 30 to 59 as predicted by the nucleotide sequence

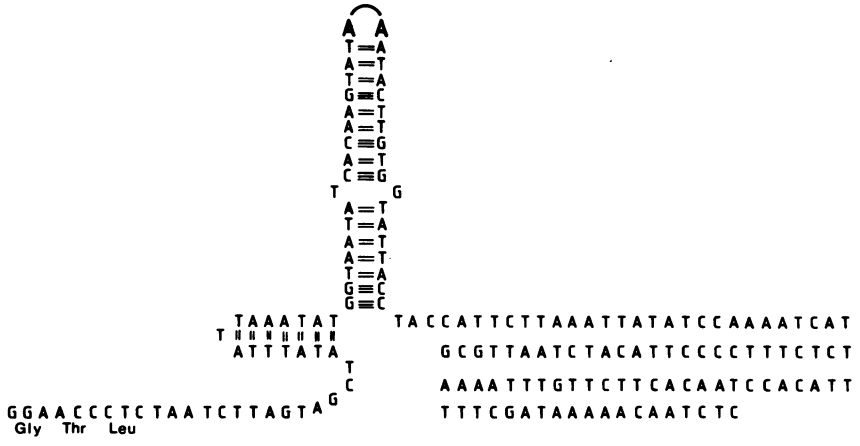

Fig. 5. Predicted secondary structure of the 3' end of the mRNA. The COOH-terminal amino acids are also shown.
The calculated free energies of the 7bp-stem and the 17 bp-stem are -9.6 kcal/mol and -29.6 kcal/mol, respectively.

appeared nearly identical to the sequence determined for the NH_2-terminal end of an M_r 25.000 protein which is an enzymatically truncated product of the M_r 28.000 crystal protein (5). The only difference occurs at residue 57, where a methionine is predicted by the nucleotide sequence and a serine is found in the truncated protein. Other differences observed are the presence of a single cysteine as deduced from the nucleotide sequence, that is not reported for the M_r 25.000 protein and the lower number of arginine residues predicted by the nucleotide sequence. These differences could be explained by assuming that a point mutation has occurred at nucleotide 611 where at T to C transition would result in the replacement of a cysteine by an arginine residue. The difference of at least two amino acids might reflect a recent divergence of the two toxin genes.

Codon usage

A striking feature of the codon usage of the M_r 28.000 crystal protein gene is the strong preference of U and A over C and G at the third position of the codon, which is most pronounced for the glutamine and glutamic acid codons of which only 1 out of 16 and 1 out of 11, respectively, contained a G in the third position. As a consequence of this preference the C-G content of the coding region (33.6%) is rather low, when compared to 39.4% for the B.t. subsp. kurstaki toxin gene (23), 43.8% for the penicillinase gene of B. licheniformis (24) and 46.5% for the α-amylase gene of B. amyloliquefaciens (25).

Since it has been suggested that the M_r 28.000 crystal protein causes a detergent-like rearrangement of membrane lipids, leading to disruption of membrane integrity and eventual cytolysis (26) the protein is likely to possess hydrophobic domains capable of interacting with membranes.

When the hydrophilicity profile of the predicted amino acid sequence was determined using the program of Hopp and Woods (27) the general hydrophobic nature of the protein could be observed (data not shown).

The NH_2-terminal part of the protein (residues 1 to 45) appears to be hydrophilic whereas the remainder of the protein is hydrophobic. It was also noted that proline residues are absent in the central portion of the protein.

Site-directed mutagenesis should provide information on the significance of these findings for the biological activity of the M_r 28.000 crystal protein.

*To whom correspondence should be addressed

REFERENCES

1. Dulmage, H.T. (1981) in Microbial control of Pests and Plant Diseases 1970-1980, Burges, H.D. ed. pp. 193-222, Academic Press, London.
2. Tyrell, D.J., Bulla, L.A. Jr., Andrews, R.E. Jr., Kramer, K.J., Davidson, L.I. and Nordin, P. (1981) J. Bacteriol. 145, 1052-1062.
3. Huber, H.E., Lüthy, P., Ebersold, H.-R. and Cordier, J.-L. (1981) Arch. Microbiol. 129, 14-18.
4. Hurley, J.M., Lee, S.G., Andrews, R.E. Jr., Klowden, M.J. and Bulla, L.A. Jr. (1985) Biochem. Biophys. Res. Commun. 126, 961-965.
5. Armstrong, J.L., Rohrmann, G.F. and Beaudreau, G.S. (1984) J. Bacteriol. 161, 39-54.
6. Ward, E.S., Ellar, D.J. and Todd, J.A. (1984) FEBS Letters 175, 377-382.
7. Sekar, V. and Carlton, B.C. (1985) Gene 33, 151-158.
8. Lereclus, D., Lecadet, M.M., Ribier, J. and Dedonder, R. (1982) Mol. Gen. Genet. 186, 391-398.
9. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
10. Srinivasan, P.R., Ramanarayanan, M. and Rabbani, E. (1975). Proc. Natl. Acad. Sci. USA 72, 2910-2914.
11. Maxam, A.M. and Gilbert, W. (1980) in: Methods in Enzymology, Colowick, S.P. and Kaplan, N.O. eds. vol. 69, pp. 499-559. Academic Press, New York.
12. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467.
13. Berk, A.J. and Sharp, P.A. (1977) Cell 12, 721-732.
14. Gonzalez, J.M. Jr., and Carlton, B.C. (1984) Plasmid 11, 28-38.
15. Schnepf, H.E. and Whiteley, H.R. (1981) Proc. Natl. Aca. Sci. USA 78, 2893-2897.
16. Klier, A., Fargette, F., Ribier, J. and Rapoport, G. (1982) EMBO J. 1, 791-799.
17. Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature New Biology 246, 40-41.
18. McLaughlin, J.R., Murray, C.L. and Rabinowitz, J.C. (1981) J. Biol. Chem. 256, 11283-11291.
19. Johnson, W.C., Moran, C.P. Jr. and Losick, R. (1983) Nature 302, 800-804.
20. Moriya, S., Ogasawara, N. and Yoshikawa, H. (1985) Nucl. Acids Res. 13, 2251-2265.
21. Fliss, E.R. and Setlow, P. (1985) Gene 35, 151-157.
22. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353.
23. Schnepf, H.E., Wong, H.C. and Whiteley, H.R. (1985) J. Biol. Chem. 260, 6264-6272.
24. Neugebauer, K., Sprengel, R. and Schaller, H. (1981) Nucl. Acids Res. 9, 2577-2588.
25. Takkinen, K., Pettersson, R.F., Kalkkinen, N., Palva, I., Soderlund, H. and Kääriäinen, L. (1983) J. Biol. Chem. 258, 1007-1013.
26. Thomas, W.E. and Ellar, D.J. (1983) FEBS Letters 154, 362-368.
27. Hopp, T.P. and Woods, K.R. (1981) Proc. Natl. Acad. Sci. USA 78, 3824-3828.