

## Isolation and Characterization of a Novel Insecticidal Crystal Protein Gene from *Bacillus thuringiensis* subsp. *aizawai*

JUDITH A. CHAMBERS,<sup>†</sup> AMY JELEN, M. PEARCE GILBERT,<sup>‡</sup> CHRISTINE S. JANY,  
TIMOTHY B. JOHNSON, AND CYNTHIA GAWRON-BURKE\*

Ecogen Inc., 2005 Cabot Boulevard West, Langhorne, Pennsylvania 19047

Received 3 December 1990/Accepted 18 April 1991

*Bacillus thuringiensis* subsp. *aizawai* EG6346, a novel grain dust isolate, was analyzed by Southern blot hybridization for its insecticidal crystal protein (ICP) gene profile. Strain EG6346 lacks previously characterized *cryIA* ICP genes yet does possess novel *cryI*-related gene sequences. A recombinant genomic plasmid library was constructed for strain EG6346 in *Escherichia coli*. One recombinant plasmid, pEG640, isolated from the library contained a novel ICP gene on a 5.7-kb *Sau3A* insert. The sequence of this gene, designated *cryIF*, was related to, but distinct from, the published sequences for other *cryI* genes. A second novel *cryI*-related sequence was also located on pEG640, approximately 500 bp downstream from *cryIF*. Introduction of *cryIF* into a *Cry*<sup>-</sup> *B. thuringiensis* recipient strain via electroporation enabled sufficient production of CryIF protein for quantitative bioassay analyses of insecticidal specificity. The CryIF crystal protein was selectively toxic to a subset of lepidopteran insects tested, including the larvae of *Ostrinia nubilalis* and *Spodoptera exigua*.

Perhaps the most well-known and widely used bioinsecticides are those based on the insecticidal crystal proteins (ICPs) produced by the sporulating bacterium *Bacillus thuringiensis*. The ICPs (also termed delta endotoxins) can comprise up to 20 to 30% of the total dry weight of sporulated cells (28) and form crystalline inclusions which are toxic when ingested by susceptible insects. The crystalline inclusions may be of various morphologies which reflect the differences in the nature of the ICPs that comprise them. ICPs can exhibit a wide variety of insecticidal specificities, and crystal proteins toxic to lepidopteran, dipteran, and coleopteran insect species have been described (7, 13, 23).

Upon ingestion, ICPs are solubilized and, in some cases, proteolytically processed by insect gut proteases to yield an active truncated toxin moiety (28). This active toxin moiety disrupts the osmotic balance of midgut epithelial cells, eventually resulting in cell lysis. The insect stops feeding within minutes, followed by paralysis and death in 3 to 5 days.

The genes encoding ICPs have been localized to large (>30-MDa) plasmids (14, 15), and various ICP genes have been cloned and characterized (for a review, see reference 20). Generally, the sequences of genes encoding proteins active on different orders of insects are not well conserved. Rather, the gene sequences encoding a given crystal phenotype and proteins active against the same insect order are significantly more related. The sequence relatedness of ICPs as well as their insecticidal activity spectrum have been used to define an ordered classification of genes encoding *B. thuringiensis* ICPs (20). Four major classes of ICP genes have been identified; *cryI*, *cryII*, *cryIII*, and *cryIV* genes encode lepidoptera-specific (CryI), lepidoptera- and diptera-

specific (CryII), coleoptera-specific (CryIII), and diptera-specific (CryIV) proteins, respectively.

The *cryI* genes, encoding the 130- to 138-kDa lepidopteran-active ICPs that form bipyramidal crystalline inclusions, comprise the largest of these families. Within the *cryI* gene classification, a subranking has been established on the basis of further refinement of sequence relationship. The *cryIA* gene subfamily [*cryIA(a)*, *cryIA(b)*, and *cryIA(c)*] includes the previously designated 4.5, 5.3, and 6.6 Pl genes, originally differentiated according to the size (in kilobases) of a characteristic *HindIII* fragment associated with the presence of the gene (25). The amino acid sequences of CryIA proteins are highly homologous (>80%), with most of the sequence dissimilarity localized to a short internal variable region (40). It is believed that differences within this variable region account for the different insecticidal specificities exhibited by the CryIA(a), CryIA(b), and CryIA(c) proteins. Additional genes within the *cryI B. thuringiensis* family have been recently reported, such as the *cryIB* gene from *B. thuringiensis* subsp. *thuringiensis* (5), *cryIC* and *cryID* from *B. thuringiensis* subsp. *aizawai* (4, 21), and *cryIE* from *B. thuringiensis* subsp. *darmstadtensis* (4). Comparisons of the sequences for these genes reveal significant sequence dissimilarities throughout the N-terminal protein domain, in contrast to the more extensive N-terminal sequence homology among the CryIA subgroup.

In this report, we present data which establish the presence of at least one additional subgroup of *cryI* genes. The prototype of this class, designated *cryIF*, was isolated from a novel grain dust isolate of *B. thuringiensis* subsp. *aizawai*. The CryIF is distinctly different in protein sequence and insecticidal specificity from the other CryI proteins. We have also identified an open reading frame located downstream from the novel *cryIF* gene, which could possibly encode an additional novel toxin gene. Data are presented on the identification, cloning, sequencing, and expression of *cryIF*, as well as on the insecticidal activities of the CryIF protein.

\* Corresponding author.

<sup>†</sup> Present address: Office of Agriculture, Bureau for Science and Technology, Agency for International Development, Washington, DC 20523.

<sup>‡</sup> Present address: 4601 Pinedale Drive, Durham, NC 27705.

TABLE 1. Bacterial strains and plasmids

Strain or plasmid	Relevant characteristic(s)	Reference or source
<i>B. thuringiensis</i> subsp. <i>kurstaki</i>		
HD73-26	Cry <sup>-</sup> Cm <sup>s</sup>	11
EG1945	HD73-26(pEG642)	This study
<i>B. thuringiensis</i> subsp. <i>aizawai</i>		
EG6345	115- and 45-MDa plasmids	This study
EG6346	115-MDa plasmid, cured of 45-MDa plasmid	This study
<i>E. coli</i>		
DH5 $\alpha$	Amp <sup>s</sup>	Bethesda Research Laboratories
GM2163	Amp <sup>s</sup> Dam <sup>-</sup> Dcm <sup>-</sup>	30
EG1943	DH5 $\alpha$ (pEG640)	This study
Plasmids		
pEG434	Tc <sup>r</sup> <i>Bacillus</i> vector	31
pGEM <sup>TM</sup> -3Z	Amp <sup>r</sup> <i>E. coli</i> vector	Promega
pEG640	pGEM-3Z with 5.7-kb insert of EG6346 DNA ( <i>cryIF</i> ORF2)	This study
pEG642	Tc <sup>r</sup> , <i>E. coli</i> - <i>Bacillus</i> shuttle vector with pEG640 inserted into HindIII site of pEG434 ( <i>cryIF</i> ORF2)	This study

## MATERIALS AND METHODS

**Bacterial strains and plasmids.** The strains and plasmids discussed in this report are described in Table 1. *B. thuringiensis* subsp. *aizawai* EG6345 was isolated from a maize grain dust sample by using previously described procedures (11). Plasmid profiles were determined for each strain by electrophoresis through agarose gels (14). The crystal-negative (Cry<sup>-</sup>) strain *B. thuringiensis* HD73-26, which is a cured derivative of HD-73 containing a single 4.9-MDa plasmid, was used as a recipient for transformation of recombinant DNA constructs into *B. thuringiensis* (11, 13a). *B. thuringiensis* subsp. *kurstaki* HD-1 was obtained from the collection of H. T. Dulmage. Library efficiency *Escherichia coli* DH5 $\alpha$  competent cells, supplied by Bethesda Research Laboratories, were used in the construction of the recombinant plasmid genomic library. Plasmid pGEM-3Z (Promega Corp.) was the vector used to construct the genomic library. Plasmid pEG434 was used to facilitate expression of toxin genes in *B. thuringiensis* (31). Plasmid pEG434 contains the 3.1-kb *EcoRI* fragment from *Bacillus cereus* plasmid pBC16 modified by the insertion of a multiple cloning site at the *EcoRI* site. *E. coli* GM2163, obtained from New England BioLabs, was used to facilitate transfer of plasmids from the *E. coli* DH5 $\alpha$  background to *B. thuringiensis* HD73-26 (30).

**Nucleic acid hybridization.** Total DNA from strains EG6345 and EG6346 was prepared according to the procedure of Kronstad et al. (24). Restriction enzyme digests were performed as recommended by the manufacturer. Restricted DNAs were size separated by electrophoresis in horizontal 0.7% agarose slab gels and transferred to nitrocellulose by the procedure of Southern (37). All double-stranded DNA probes were radioactively labeled by nick translation (33). Nitrocellulose filters containing bound DNA were hybridized under either of two conditions to accommodate alterations in the stringency of the annealing reaction. Prehybridization and hybridization of filters were in a solution of 3 $\times$  SSC (1 $\times$  SSC is 0.15 M NaCl plus 0.015 M sodium citrate), 10 $\times$  Denhardt's solution (1 $\times$  Denhardt's solution is 0.02% bovine serum albumin, 0.02% Ficoll, and 0.02% polyvinylpyrrolidone), 200  $\mu$ g of heparin per ml and 0.1% sodium dodecyl sulfate (SDS). Standard (high-stringency) hybridization was conducted at 65°C; hybridization at lower stringency was performed at 50°C. Washes were in 3 $\times$  SSC-0.1% SDS at either temperature. Filters were dried and exposed to

Kodak X-Omat AR film, using DuPont Cronex intensifying screens.

**Construction of a *B. thuringiensis* EG6346 genomic library.** High-molecular-weight DNA, obtained from *B. thuringiensis* EG6346, was partially digested with *Sau3A* and size fractionated on a 10 to 40% sucrose gradient in 100 mM NaCl-10 mM Tris hydrochloride (pH 7.4)-1 mM EDTA (29). Gradient fractions, containing DNA ranging in size from 5 to 10 kb, were pooled, dialyzed against 10 mM Tris-1 mM EDTA (pH 7.4), extracted with 2-butanol to reduce the volume, and ethanol precipitated (29). The purified insert DNA was ligated to pGEM-3Z *Bam*HI-digested vector DNA at a 1:2 molar ratio of vector to insert and at a final DNA concentration of 20  $\mu$ g/ml, using T4 DNA ligase (Promega). Transformation into *E. coli* DH5 $\alpha$  cells was done according to the manufacturer's directions. Transformed colonies were plated on LB medium containing 100  $\mu$ g of ampicillin and 50  $\mu$ g of X-Gal (5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactoside) per ml. Approximately 3.3  $\times$  10<sup>6</sup> colonies were screened for the presence of *cryI*-related toxin gene sequences under low-stringency conditions, using a probe comprised of a 2.2-kb *Pvu*II intragenic fragment obtained from a *cryIA(a)* gene present within *B. thuringiensis* HD-1. Rapid, small-scale isolation of plasmid DNA from recombinant colonies was performed by the procedure of Birnboim and Doly (3).

**Transformation procedures.** Transformation of *E. coli* DH5 $\alpha$  was performed according to the manufacturer's protocol (Bethesda Research Laboratories' recommended protocol, which was adapted from the procedure of Hanahan [17]). Transformation into the Dam<sup>-</sup> Dcm<sup>-</sup> *E. coli* GM2163 (30) was facilitated by preparation of frozen competent cells according to the procedure of Maniatis et al. (29). Transformation of *B. thuringiensis* HD73-26 was accomplished, as previously described, by electroporation in a Bio-Rad GenePulser (31).

**DNA sequencing.** Standard dideoxy sequencing procedures (34) using Sequenase (U.S. Biochemical) were employed to sequence the 5.7-kb pEG640 insert. Sequencing of the insert was initiated in both directions and on both strands from the SP6 and T7 promoters present on vector pGEM-3Z and was done with use of the specific primers supplied by the manufacturer (Promega). Preparation of and denaturation of the double-stranded template were also done as instructed by the manufacturer (Promega or U.S. Biochemical). Sub-

sequent 17-mer oligonucleotide primers were synthesized on a model 380B Applied Biosystems DNA synthesizer. The sequence analysis program of Queen and Korn was used to compare the sequences of *cryIF* and ORF2 with the published sequences of other *B. thuringiensis* ICP genes (32).

**Protein analysis.** *B. thuringiensis* EG6345, EG6346, and EG1945 were grown for 72 h at 30°C in M55 medium [29 mM  $K_2HPO_4$ , 37 mM  $KH_2PO_4$ , 1 mM citric acid  $\cdot H_2O$ , 5 mM  $(NH_4)_2SO_4$ , 150 mM NaCl, 1 mM  $CuCl_2 \cdot 2H_2O$ , 1 mM  $ZnCl_2$ -trisodium citrate, 1 mM  $Na_2MoO_4$ , 0.3 mM  $MgCl_2 \cdot 6H_2O$ , 5  $\mu M$   $MnCl_2 \cdot 4H_2O$ , 0.5 mM  $CaCl_2 \cdot 2H_2O$ , 0.15% potato dextrose broth (Difco Laboratories), 0.265% nutrient broth (Difco), 0.67 mM L-methionine (Sigma Chemical Co.)] until cultures were fully sporulated. Cultures were harvested, resuspended in TNT (50 mM Tris [pH 7.5], 100 mM NaCl, 0.05% Triton X-100) with a final lysozyme (Sigma) concentration of 0.5 mg/ml, and lysed at 37°C for 2 h. Lysed cultures were pelleted, resuspended in TNT, and loaded onto linear 78 to 55% Renografin-76 (Squibb Diagnostics) gradients containing 0.05% Triton X-100. Gradients were centrifuged at 18,000 rpm in an SW28 rotor, using a Beckman model L8 ultracentrifuge, for at least 2 h. Crystal bands were collected from gradients and fractionated over an additional Renografin gradient for further purification of crystals. The crystals were washed and resuspended in 0.005% Triton X-100 and stored at 4°C. Purified crystal preparations used in bioassay analyses were also examined by using a discontinuous SDS-polyacrylamide gel electrophoresis (PAGE) system consisting of a 3% acrylamide stacking gel (pH 6.8), with a linear gradient gel from 5 to 20% acrylamide (pH 8.8) used for resolution of protein bands (6, 26). Crystal protein concentration was determined by the Bio-Rad protein assay (Bio-Rad Laboratories, Richmond, Calif.).

**Insect toxicity assays.** The insecticidal activity of the CryIF protein was tested against four lepidopteran larvae, one member of the Pyralidae (*Ostrinia nubilalis*), and three Noctuidae (*Spodoptera exigua*, *Heliothis virescens*, and *Helicoverpa zea*), using Renografin-purified crystal protein from *B. thuringiensis* EG1945 harboring the *cryIF* gene on plasmid pEG642. For comparison, Renografin-purified crystal proteins from recombinant *B. thuringiensis* EG7077, containing the *cryIA(b)* gene, and EG1861, containing the *cryIA(c)* gene, were included (39a). Insecticidal activity was measured by using an overlay technique in which the surface of an agar-based artificial diet (22) was covered with an aliquot suspension containing CryIF protein crystals. Each bioassay consisted of eight serial dilutions in 0.005% Triton X-100. Fifty-microliter aliquots were delivered to each of 32 2-ml wells containing 1 ml of diet (surface area, 175 mm<sup>2</sup>). The diluent only served as a control treatment. After the diluent was allowed to dry, one neonate larva of the test species was placed in each well, for a total of 256 larvae per bioassay. After covering, bioassays were held at 28°C for 7 days, at which time mortality was scored. If insecticidal activity was sufficient to determine 50% lethal concentrations, bioassays were repeated. Bioassay data were adjusted for control mortality with Abbott's formula (1), with replications combined for composite probit analysis (12) using the program of Daum (8).

**Nucleotide sequence accession number.** The nucleotide sequence of the 5.7-kb insert of *B. thuringiensis* DNA in pEG640, including the coding region of CryIF and its deduced amino acid sequence, have been filed with GenBank, Los Alamos National Laboratory, under accession number M63897.

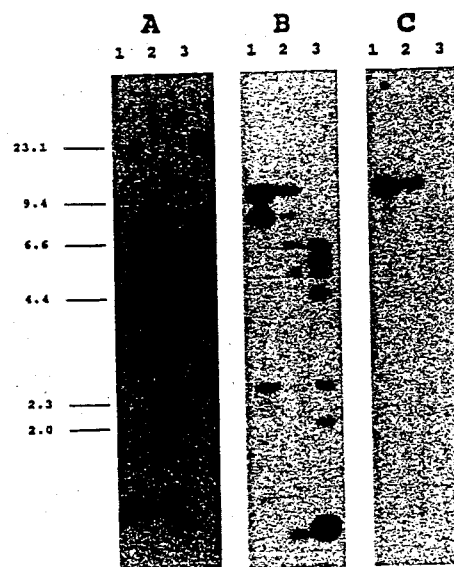


FIG. 1. Hybridization analyses of *B. thuringiensis* subsp. *aizawai* isolates. (A) Southern blot of an agarose gel containing *Hind*III-digested total DNA from *B. thuringiensis* strains EG6346 (lane 1), EG6345 (lane 2), and HD-1 (lane 3) following hybridization to a <sup>32</sup>P-labeled, N-terminal, 726-bp *Eco*RI probe from the *B. thuringiensis* subsp. *kurstaki* HD-1 *cryIA(a)* gene at low-stringency conditions. (B) Southern blot of an agarose gel containing *Hind*III-digested total DNA from *B. thuringiensis* strains (lanes as in panel A) following hybridization to the <sup>32</sup>P-labeled, 2.2-kb intragenic *Pvu*II probe from the *cryIA(a)* gene of *B. thuringiensis* HD-1 at low-stringency conditions. (C) Southern blot analysis of *Hind*III-digested total DNA from *B. thuringiensis* strains (lanes as in panel A) following hybridization to a <sup>32</sup>P-labeled, N-terminal, 0.4-kb *Pst*I-*Sac*I intragenic, *cryIF*-specific probe from plasmid pEG640. *Hind*III-digested phage lambda DNA was used for size standards.

## RESULTS

**Identification and isolation of the *cryIF* gene.** *B. thuringiensis* subsp. *aizawai* EG6345 was isolated from a maize grain dust sample and selected for further study on the basis of its insecticidal activity against a variety of lepidopteran larvae (data not shown). Strain EG6346 was identified by its distinct colony morphology (i.e., shinier) in comparison with strain EG6345 on a nutrient salts agar plate and subsequently shown by plasmid profile agarose gel electrophoresis analyses to be a spontaneously cured derivative of strain EG6345, which lacked the 45-MDa plasmid. Both strains, EG6345 and the cured derivative, EG6346, produced large bipyramidal inclusions during sporulation.

Total DNA, prepared from both EG6345 and EG6346, was digested with *Hind*III, electrophoresed through a 0.7% agarose gel, transferred to nitrocellulose, and hybridized at 50°C either to a 726-bp *Eco*RI N-terminal probe (Fig. 1A) or to the 2.2-kb intragenic *Pvu*II probe isolated from the *cryIA(a)* gene of *B. thuringiensis* HD-1 (Fig. 1B) (24). Digested DNA from *B. thuringiensis* subsp. *kurstaki* HD-1, which harbors the *cryIA(a)*, *cryIA(b)*, and *cryIA(c)* genes, was included as a control.

As shown in Fig. 1A, the 726-bp *Eco*RI probe detected the expected 4.5-, 5.3-, and 6.6-kb fragments in HD-1 DNA (lane 3) corresponding to the previously described characteristic *Hind*III fragments for the *cryIA(a)*, *cryIA(b)*, and *cryIA(c)* genes, respectively (24). This probe also detected a promi-

nent 5.3-kb band in EG6345 (lane 2) which was absent in the cured derivative EG6346 (lane 1). This result indicated that the 45-MDa plasmid of EG6345 harbored at least one *cryIA(b)* gene. Subsequent conjugal transfer experiments confirmed these data (13b). The N-terminal 726-bp *EcoRI* probe also hybridized to a 1.4-kb *HindIII* fragment of unknown origin in both EG6345 (lane 2) and EG6346 (lane 1). Independent experiments confirmed the presence of the 1.4-kb hybridizing fragment in strains EG6345 and EG6346 (data not shown).

The hybridization pattern obtained with the intragenic *PvuII* probe from *cryIA(a)* was more complex (Fig. 1B). This probe, as expected, also hybridized to the 4.5- and 6.6-kb N-terminal flanking *HindIII* fragments in HD-1 (lane 3), confirming the presence of the *cryIA(a)* and *cryIA(c)* genes resident in this strain. In addition, a C-terminal 2.2-kb flanking *HindIII* fragment from *cryIA(c)*, as well as an internal 1.1-kb fragment corresponding to the presence of the *cryIA(a)* gene, was detected in HD-1. As expected, the 4.5- and 1.1-kb fragments, corresponding to the *cryIA(a)* gene, were absent in both EG6345 and EG6346, consistent with the lack of a *cryIA(a)* gene in both strains. Although a band of approximately 6.6 kb was observed in EG6345 (lane 2), which appears to comigrate with the *cryIA(c)*-specific 6.6-kb band from HD-1 (lane 3), the appearance of this band is coincidental and does not suggest the presence of a *cryIA(c)* gene in EG6346, as determined by independent confirmation with the *EcoRI* probe (Fig. 1A, lane 2).

A 5.3-kb *HindIII* fragment was detected with the *PvuII* probe (Fig. 1B) in both HD-1 (lane 3) and EG6345 (lane 2) but was not detected in EG6346 (lane 1). Similarly, an internal 0.9-kb fragment derived from *cryIA(b)* was detected by the *PvuII* probe in both HD-1 and EG6345 but was absent in the EG6346 digest. Lastly, a 6.0-kb fragment corresponding to the 3'-terminal and flanking sequences of the *cryIA(b)* gene was also detected with the *PvuII* probe in strain HD-1. These data confirmed the presence of a *cryIA(b)* gene in strains HD-1 and EG6345 and its absence in EG6346.

A 1.4-kb *HindIII* fragment, detected in strains EG6345 and EG6346 by the *EcoRI* probe, was also faintly detected in these strains by the *PvuII* probe. A 2.5-kb fragment was also detected with the *PvuII* probe in all three strains. For strains EG6345 and EG6346, this band may correspond to the presence of a characteristic *HindIII* fragment from a *cryIC* gene, which has been detected in other *B. thuringiensis* subsp. *aizawai* strains of the Ecogen collection (7a). The appearance of a similarly sized fragment in HD-1 is most likely coincident, since HD-1 does not harbor a *cryIC* gene. This fragment could correspond to a C-terminal flanking *HindIII* fragment from the *cryIA(a)* gene resident in the strain. The *PvuII* probe also hybridized to two large *HindIII* fragments present in both EG6345 and EG6346. These fragments, approximately 8.2 and 10.4 kb in length, were not detected by the *EcoRI* probe in EG6345 and EG6346, nor were they observed with either probe in HD-1 DNA. These data suggested the presence of at least one or more novel toxin genes in strains EG6345 and EG6346.

A partial *Sau3A* genomic library was constructed for EG6346 and was screened at low-stringency conditions with the intragenic *PvuII* probe. EG6346 DNA was chosen as the substrate DNA due to its apparent lack of *cryIA* type toxin genes, whose presence could potentially increase the difficulty in screening the library at low stringency with the *PvuII* probe. The probe hybridized strongly to one *E. coli* recombinant colony (EG1943) which contained a recombi-

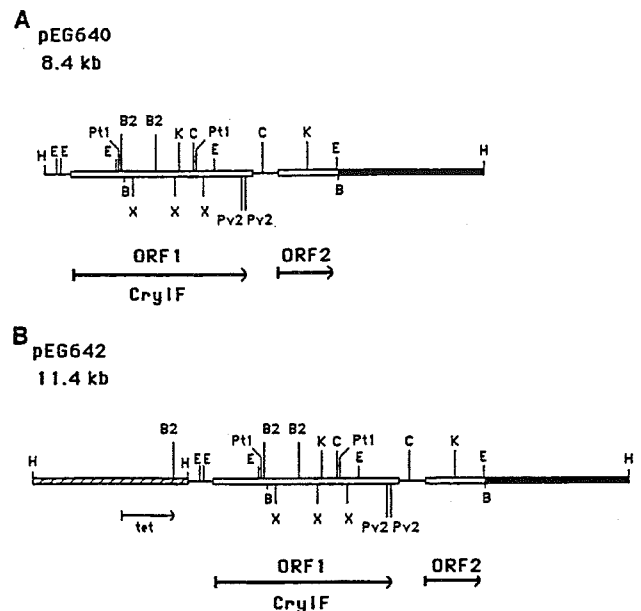


FIG. 2. Restriction maps of plasmids pEG640 (*E. coli*) and pEG642 (*E. coli*-*Bacillus*) containing *cryIF*. Restriction enzymes: H, *HindIII*; E, *EcoRI*; Pt1, *PstI*; B, *BamHI*; B2, *BstEII*; X, *XbaI*; K, *KpnI*; C, *Clal*; Pv2, *PvuII*. Open boxes denote toxin gene sequences; closed box is *E. coli* cloning vector pGEM-3Z (Promega); hatched box is the *Bacillus* vector pEG434. Arrows indicate direction of transcription. *tet*, the tetracycline resistance gene encoded on plasmid pEG434.

nant plasmid, pEG640, that consisted of pGEM-3Z ligated to a 5.7-kb *Sau3A* insert.

Sequence analyses of *cryIF*. A restriction map for the pEG640 insert was generated (Fig. 2A). Relative positions of restriction sites and localization of toxin gene sequences within the map were initially determined by low-stringency hybridization of Southern blots containing digested pEG640 DNA to the *EcoRI* and *PvuII* toxin gene probes (data not shown). Initial mapping data identified two regions on the pEG640 insert which reacted with varying intensity to the toxin gene probes. The larger of these, spanning over 3 kb in length, hybridized strongly to the *PvuII* probe at low- and high-stringency conditions. Significant hybridization was also observed with the *EcoRI* probe at low-stringency conditions. A smaller region, positioned in close proximity to the vector, weakly hybridized to the *EcoRI* probe at low-stringency conditions only. These data suggested the presence of two distinct *cryIA*-related sequences on the pEG640 insert.

The presence of at least one complete toxin gene and a possible truncated toxin gene was subsequently verified by DNA sequence analysis of the entire 5.7-kb insert. The DNA sequence, which is flanked by *Sau3A* cloning sites (GATC), extends 5,649 nucleotides in length (Fig. 3). Translation of the sequence revealed the presence of two open reading frames which were separated by approximately 500 bases of noncoding DNA sequence and are out of frame with respect to one another. The gene potentially encoded by the upstream open reading frame (ORF1) has been designated *cryIF*. Justification for this designation derives from sequence comparisons with other toxin genes and is discussed below.

The *cryIF* open reading frame, which is the larger of the

30 60  
GATCTTCAAAATGAGAAAAAAGGGTATTCGGTATGGGATGCCTTATTTGGTTGGGAAG  
150 180  
AAGGATTAATAATCAAAATGTAATCAGATATAGTCCAGATAATTTTAAAGAGTGTA  
210 240  
GTATATTAATAATGTTCTTATAACATATATGTTGATTTTAAAGAAATATTTGTTTA  
270 300  
AGAATTCAATCCATATGAGTATAAAAGTTAAAGGCCAAAAATAAGTTAAGGGAAATC  
330 360  
AAGCTTTAATACAAAGTTTATCTCAGGAATCTCAACTATGGATAGCAGGAAGAGAAG  
390 420  
TAAGCACAATTATTAACATATTAGGTCATTTAAATTAAGGGCATATAGTGATATTTTATA  
450 480  
AGATTGGTGCACCTTTGTGCAATTTTCATAAGATGAGTCAATATTTACATTGTAATA  
510 540  
CAGTAAGAGGTTTTAGTTTAAAGAACTATTATGATGAAATGTGGAGGAACTATG  
570 600  
GAGAATAATATTCAAAATCAATGCGTACCTTCAATTTTAAATAATCCTGAAGTAGAA  
630 660  
GluAsnAsnIleGlnAsnGlnCysValProTyrAsnCysLeuAsnAsnProGluValGlu  
690 720  
ATATTAATGAAGAAAGTACTGGCAGATTACCGTTAGATATATCCTTATCGCTTACA  
750 780  
IleLeuAsnGluGluArgSerThrGlyArgLeuProLeuAspIleSerLeuSerLeuThr  
810 840  
CGTTTCCTTTTGGAGTAAATTTGTTCCAGGTGTGGGAGTTGCGTTTGGATTATTTGATT  
870 900  
ArgPheLeuLeuSerGluPheValProGlyValGlyValAlaPheGlyLeuPheAspLeu  
930 960  
ATATGGGTTTTATACTCCTTCTGATTTGGAGCTTATTTCTTTTACAGATTGAACAATTG  
990 1020  
IleTrpGlyPheIleThrProSerAspTrpSerLeuPheLeuLeuGlnIleGluGlnLeu  
1050 1080  
ATTGAGCAAGAATAGAACTTGGAAAGGACCGGCAATTTACTACATTACGAGGGTTA  
1110 1140  
IleGluGlnArgIleGluThrLeuGluArgAsnArgAlaIleThrThrLeuArgGlyLeu  
1170 1200  
GCAGATAGCTATGAATTTATATGAAGCACTAAGAGAGTGGGAAGCAATCTAATAAT  
1230 1260  
AlaAspSerTyrGluIleTyrIleGluAlaLeuArgGluTrpGluAlaAsnProAsnAsn  
1290 1320  
GCACAATTAAGGGAAGATGTCGTTTCCGATTTGCTAATACAGACGACCTTTAATAACA  
1350 1380  
AlaGlnLeuArgGluAspValArgIleAspPheAlaAsnThrAspAspAlaLeuIleThr  
1410 1440  
GCAATAAATATTTTACACTTACAAGTTTGAATCCCTCTTTATCGGTCTATGTTCAA  
1470 1500  
AlaIleAsnAsnPheThrLeuThrSerPheGluIleProLeuLeuSerValTyrValGln  
1530 1560  
GCGGCAATTTACATTATCACTATTAGAGACGCTGTATCGTTGGGACAGGTTGGGA  
1590 1620  
AlaAlaAsnLeuHisLeuSerLeuLeuArgAspAlaValSerPheGlyGlnIleTrpGly  
1650 1680  
CTGGATATAGCTACTGTTAATAATCATTATAATAGATTAAATAATCTTATCATAGAT  
1710 1740  
LeuAspIleAlaThrValAsnAsnHisTyrAsnArgLeuIleAsnLeuIleHisArgTyr  
1770 1800  
ACGAAACATTTGGGACACATCAATCAAGGATTAGAAAACCTTAAGAGGTACTAATCT  
1830 1860  
ThrLysHisCysLeuAspThrTyrAsnGlnGlyLeuGluAsnLeuArgGlyThrAsnThr  
1890 1920  
CGACAATGGGCAAGATTCAATCAGTTTAGAGAGATTTAACACTACTGTATTAGATATC  
1950 1980  
ArgGlnTrpAlaArgPheAsnGlnPheArgArgAspLeuThrLeuThrValLeuAspIle  
2010 2040  
GTTGCTCTTTTCCGAACATGATGTTAGACATATCCAATCAACGTCATCCCAATTA  
2070 2100  
ValAlaLeuPheProAsnTyrAspValArgThrTyrProIleGlnThrSerSerGlnLeu  
2130 2160  
ACAAGGGAATTTATACAAGTTCAAGTATTGAGGATTTCCAGTTTCTGCTAATATACCT  
2190 2220  
ThrArgGluIleTyrThrSerSerValIleGluAspSerProValSerAlaAsnIlePro  
2250 2280  
AATGGTTTTAATAGCGGAATTTGGAGTTAGACCGCCCATCTTATGCACTTTATGAAT  
2310 2340  
AsnGlyPheAsnArgAlaGluPheGlyValArgProProHisLeuMetAspPheMetAsn  
2370 2400  
TCTTTGTTTGAAGTCCAGAGACTGTTAGAACTCAAACTGTGTGGGAGGACACTAGTT  
2430 2460  
SerLeuPheValThrAlaGluThrValArgSerGlnThrValTrpGlyGlyHisLeuVal  
2490 2520  
AGTTACGAAATACGGCTGGTAACCGTATAAATTTCCCTAGTTACGGGCTCTCAATCT  
2550 2580  
SerSerArgAsnThrAlaGlyAsnArgIleAsnPheProSerTyrGlyValPheAsnPro  
2610 2640  
GGTGGCGCAATTTGGATTGAGATGAGGATCCACGCTCTTTTATCGGACATTATCAGAT  
2670 2700  
GlyGlyAlaIleTrpIleAlaAspGluAspProArgProPheTyrArgAsp  
2730 2760  
TCTTGTGTTGTAAGTCCAGAGACTGTTAGAACTCAAACTGTGTGGGAGGACACTAGTT  
2790 2820  
SerLeuPheValThrAlaGluThrValArgSerGlnThrValTrpGlyGlyHisLeuVal  
2850 2880  
AGTTACGAAATACGGCTGGTAACCGTATAAATTTCCCTAGTTACGGGCTCTCAATCT  
2910 2940  
SerSerArgAsnThrAlaGlyAsnArgIleAsnPheProSerTyrGlyValPheAsnPro  
2970 3000  
GGTGGCGCAATTTGGATTGAGATGAGGATCCACGCTCTTTTATCGGACATTATCAGAT  
3030 3060  
GlyGlyAlaIleTrpIleAlaAspGluAspProArgProPheTyrArgAsp  
3090 3120  
TCTTGTGTTGTAAGTCCAGAGACTGTTAGAACTCAAACTGTGTGGGAGGACACTAGTT  
3150 3180  
SerLeuPheValThrAlaGluThrValArgSerGlnThrValTrpGlyGlyHisLeuVal  
3190 3220  
AGTTACGAAATACGGCTGGTAACCGTATAAATTTCCCTAGTTACGGGCTCTCAATCT  
3250 3280  
SerSerArgAsnThrAlaGlyAsnArgIleAsnPheProSerTyrGlyValPheAsnPro  
3290 3320  
GGTGGCGCAATTTGGATTGAGATGAGGATCCACGCTCTTTTATCGGACATTATCAGAT  
3350 3380  
GlyGlyAlaIleTrpIleAlaAspGluAspProArgProPheTyrArgAsp

FIG. 3. DNA sequence of the 5.7-kb insert of *B. thuringiensis* EG6346 DNA on pEG640. The sequence is flanked by the *Sau*3A cloning sites and extends 5,649 nucleotides in length. Two open reading frames are indicated as diagrammed in Fig. 2. The larger of these is *cryIF*, and the smaller is designated ORF2. Putative ribosome binding sites (RBS) are shown. A putative promoter region upstream of *cryIF* is numerically designated as the -35 and -10 region. Amino acid sequence for *CryIF* and ORF2 are shown below the nucleotide sequence.

AAACGTGAAATTTGGAATTTGGAACAAATATCTTTATAAGAGGCAAAAGAAATCTGTA LysArgGluLysLeuGluLeuGluThrAsnIleValTyrLysGluAlaLysGluSerVal	3150	3180	TAAGTATAATTTGTATGAATAAAATTATATCTGAAAAATAAATAATTACAGTGGAGG 238	4500
GATGCTTTATTTGTAACCTCTCAATATGATCAATTACAAGCGGATACGAATATTGCCATG AspAlaLeuPheValAsnSerGlnTyrAspGlnLeuGlnAlaAspThrAsnIleAlaMet	3210	3240	GATTAATATGAACTAAAGATCCAGATAAGCATCAAAATTTTCTAGCAATGCGAAAGT MetLysLeuLysAsnProAspLysHisGlnSerPheSerSerAsnAlaLysVa	4560
ATTGATGCGGCGAGATAAAGCTGTTTCATAGAAATTCGGGAAGCGTATCTTCCAGAGTTATCT IleHisAlaAlaAspLysArgValHisArgIleArgGluAlaTyrLeuProGluLeuSer	3270	3300	AGATAAAATCTCTACGGATTCACTAAAAATGAAACAGATATAGAATTACAAACATTAA LAspLysIleSerThrAspSerLeuLysAsnGluThrAspIleGluLeuGlnAsnIleAs	4620
GTGATTCCGGGTGTAATGTAGACATTTTCGAAGAATTAAGGGCGTATTTTCACTGCA ValIleProGlyValAsnValAspIlePheGluGluLeuLysGlyArgIlePheThrAla	3330	3360	TCATGAAGATTGTTTGAATAATCTGAGTATGAAAAATGAGAGCCGTTTGTAGTGCATC nHisGluAspCysLeuLysIleSerGluTyrGluAsnValGluProPheValSerAlaSe	4680
TTCTCTCTATGATCGGAGAAATGCTATTAAGAAACCGGTATTTCAATAATGGCTTATCA PhePheLeuTyrAspAlaArgAsnValIleLysAsnGlyAspPheAsnAsnGlyLeuSer	3390	3420	AACAATTCMAACAGGTATTAGTATTGGGGTAAAACTAGTGGCACCCCTAGGCGTTCTTT rThrIleGlnThrGlyIleSerIleAlaGlyLysIleLeuGlyThrLeuGlyValProPh	4740
TGCTGGAACGTGAAGGGCATGTAGATAGAGAACAACACCCCGTTCGGTCCCT CysTrpAsnValLysGlyHisValAspValGluGluGlnAsnAsnHisArgSerValLeu	3450	3480	TGCAGGACAGTAGCTAGTCTTTATAGTTTATCTTAGGTAGAGCTATGGCCTAAGGGGAA eAlaGlyGlnValAlaSerLeuTyrSerPheIleLeuGlyGluLeuTrpProLysGlyLy	4800
GTGTTCCGGGAATGGGAAGCAGAGTGTCAAGAAGTTCGTGTCGCGGGTCCGTGGC ValValProGluTrpGluAlaGluValSerGlnGluValArgValCysProGlyArgGly	3510	3540	AAATCAATGGGAAATCTTTATGGAACATGTAGAAGATTATTAATCAAAAATATCAAC sAsnGlnTrpGluIlePheMetGluHisValGluGluIleIleAsnGlnLysIleSerTh	4860
TATATCCTCTGTCACAGCGTACAGAGGGGATATGGAGAAGGTTGCGTAACCATTCAT TyrIleLeuArgValThrAlaTyrLysGlyGlyTyrGlyGlyCysValThrIleHis	3570	3600	TTATCGAAGAAATAAGCACTACAGACTGAAAGGATTAGGAGATGCCCTTAGCTGTCTA rTyrAlaArgAsnLysAlaLeuThrAspLeuLysGlyLeuGlyAspAlaLeuAlaValTy	4920
GAGATCGAGAACATACAGACGAAGTGAAGTTTAGCAACTGCGTGAAGAGGCAAGCTAT GluIleGluAsnAsnThrAspGluLeuLysPheSerAsnCysValGluGluGluValTyr	3630	3660	CCATGAATCCCTGAAAGTGGGTGGAAATCGTAAGAACACAAGGGCTAGGAGTGTGT rHisGluSerLeuGluSerTrpValGlyAsnArgLysAsnThrArgAlaArgSerValVa	4980
CCAAACAACCGGTAACTGTAATGATTACTGCAAAATCAAGAAGAAATACGGGGTCCG ProAsnAsnThrValThrCysAsnAspTyrThrAlaAsnGlnGluGluTyrGlyGlyAla	3690	3720	CAAGGCCAATATATCGCATTAGAATTGATGTTCTGTCAGAACTACCTTCTTTTGGT lLysSerGlnTyrIleAlaLeuGluLeuMetPheValGlnLysLeuProSerPheAlaVa	5040
TACACTCCCGTAACTCGTGATATGAGAACTTATGGAAGCAATCTCTGTACAGCT TyrThrSerArgAsnArgGlyTyrAspGluThrTyrGlySerAsnSerSerValProAla	3750	3780	GTCTGGAGAGGAGGTACCAATTATACCAATATATGCCAAGCTGCAAAATTACATTGTT lSerGlyGluGluValProLeuLeuProIleTyrAlaGlnAlaAlaAsnLeuHisLeuLe	5100
GATTATCGCTCAGTCTATGAAGAAATCGTATACAGATGGACGAAGAGACAATCTTGT AspTyrAlaSerValTyrGluGluLysSerTyrThrAspGlyArgArgAspAsnProCys	3810	3840	GCTATTAAGAGATGCATCTATTTTGGAAAGATGGGGATTATCATCTTCAGAAATTTC uLeuLeuArgAspAlaSerIlePheGlyLysGluTrpGlyLeuSerSerSerGluIleSe	5160
GAATCTAACAGAGGATATGGGATACACACCACTACAGCTGGCTATGTGCAAAAGAA GluSerAsnArgGlyTyrGlyAspTyrThrProLeuProAlaGlyTyrValThrLysGlu	3870	3900	AACATTTTATAACCGTCAAGTCAAGCAGCAGGAGATTATCCGACCATTTGTGTGAATG rThrPheTyrAsnArgGlnValGluArgAlaGlyAspTyrSerAspHisCysValLysTr	5220
TTAGAGTACTTCCAGAAACCGATAAGGTATGGATTGAGATCGGAGAAACGGAAGACA LeuGluTyrPheProGluThrAspLysValTrpIleGluIleGlyGluThrGluGlyThr	3930	3960	GTATAGTACAGGTCTAAATAACTTACGGGTACAAATGCCGAAAGCTGGGTTCGTTATAA pTyrSerThrGlyLeuAsnAsnLeuArgGlyThrAsnAlaGluSerTrpValArgTyrAs	5280
TTCATCGTGACAGCGTGAATTAAGTCTTATGGAGAAATAGTCTCATACAAATTAGTT PheIleValAspSerValGluLeuLeuLeuMetGluGluEnd	3990	4020	TCAATTTCTGTAAGATATGACATTAATGTTACTTGTATTAGTGGCACTATTCCCAAGCTA nGlnPheArgLysAspMetThrLeuMetValLeuAspLeuValAlaLeuPheProSerTy	5340
TTAAATATCGTTTCAAATCAATTCGCTAAGAGCATCATTACAAATAGATAAGTAATTG 4050	4080	5370	TGATACACTTGTATATCCAAATTAAGTACTTCTCAACTTACAAGAGAGTATATACAGA rAspThrLeuValTyrProIleLysThrThrSerGlnLeuThrArgGluValTyrThrAs	5400
TTGTAATGAAAAACGGACATCACCTCCATTGAAACGGTGAGATGTCGGTTTACTATGTT 4110	4140	5430	CGCAATTGGGACAGTACATCCGAATGCAAGTTTTCGAAGTACGACTTGGTATATAATAA pAlaIleGlyThrValHisProAsnAlaSerPheAlaSerThrThrTrpTyrAsnAsnAs	5460
ATTTTCTAGTAATACATATGTACAGAGCAACTTAATTAAGCAGAGATATTTCCCTATC 4170	4200	5490	TGCCCCCTGTTCTCTACCATAGAGTCTGCTGTTGTTGCGAAACCCGATCTACTCGATT nAlaProSerPheSerThrIleGluSerAlaValValArgAsnProHisLeuLeuAspPh	5520
GATGAAATATCTCTGCTTTTCTTTCTTTTATTCGGTATATGCTTACTTGTAAATGAAA 4230	4260	5550	TCTAGAACAAGTTACAATTACAGCTTATTAGTAGGTGGAGTAACTCAGTATATGAA eLeuGluGlnValThrIleTyrSerLeuLeuSerArgTrpSerAsnThrGlnTyrMetAs	5580
ATAAGCACTAATAAGAGTATTTATAGGTGTTTGAAGTTATTTTCACTTTATTTTAAAGG 4290	4320	5610	TATGTGGGAGGACATAGACTTGAATTCGAACCAATCGAGGAATGTTAAATACCTCAAC nMetTrpGlyGlyHisArgLeuGluPheArgThrIleGlyGlyMetLeuAsnThrSerTh	5640
AGGTTTAAAAACGTTAGAAAGTTATTAAAGGAATAACTTATTAGTAAATCCACATATA 4350	4380	ACAAGGATG rGlnGly		
TTTTATAATTAATTATGAAATATATGTATAAATGAAATGCTTATTGACATTACAGC 4410	4440			

FIG. 3—Continued.

two, is 3,522 nucleotides in length, encoding a putative peptide consisting of 1,174 amino acids. Its position within pEG640 and its relationship to the location of the downstream open reading frame (ORF2) are schematically represented in Fig. 2A. As shown in Fig. 3, the coding region of the *cryIF* gene extends from nucleotide positions 478 to 3999. An NH<sub>2</sub>-terminal methionine translational start site was identified for *cryIF* at nucleotide position 478 of the

sequence. It was immediately preceded by a putative ribosome binding site. A putative promoter sequence for the *cryIF* gene was located via sequence inspection upstream of the ribosome binding site, beginning at nucleotide position 389 (Fig. 3). The sequences of the *cryIF* presumed -10 and -35 regions were homologous to those identified for the HD-1-Dipel *cryIA(a)* gene Btl promoter (41).

A methionine codon, followed by an open reading frame of



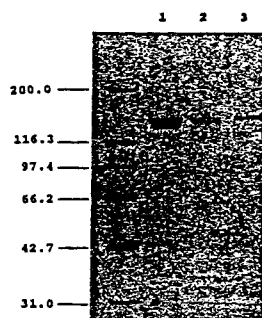


FIG. 4. SDS-PAGE analysis of crystal protein from *B. thuringiensis* native and recombinant strains. Lanes: 1 and 2, Renografin gradient-purified crystal protein from *B. thuringiensis* EG6345 and EG6346, 2.8 and 0.75  $\mu$ g, respectively; 3, 0.70  $\mu$ g of Renografin gradient-purified crystal protein from *B. thuringiensis* EG1945 (recombinant strain harboring *cryIF*). Numbers at the left indicate the positions of protein size standards in kilodaltons.

The smaller protein present in EG6346 and also evident in EG6345 most likely represents the protein encoded by the *cryIC* gene, which has been identified in each of these strains by DNA hybridization analysis with a *cryIC*-specific oligonucleotide probe (data not shown). However, because of the sometimes spurious nature of protein migration in gradient SDS-polyacrylamide gels, specific confirmation of protein size for the CryIC protein compared with the CryIF protein will depend on analysis of the full-size, cloned *cryIC* gene product.

**Plasmid localization of the *cryIF* gene.** To determine the location of the *cryIF* gene in *B. thuringiensis* EG6345 and EG6346 and to compare its location with that of the *cryIA(b)* gene present within strain EG6345, plasmid DNAs from EG6345 and EG6346 were resolved by agarose gel electrophoresis according to the method of González et al. (14) (Fig. 5A). Plasmid DNAs were then transferred to nitrocellulose and hybridized either to the intragenic 2.2-kb *PvuII* probe or to a *cryIF* gene-specific probe consisting of a gel-purified 0.4-kb *PstI-SacI* fragment isolated from the 5'-terminal region of the *cryIF* gene on pEG640. As shown in Fig. 5B, the *PvuII* intragenic *cryIA(a)* probe hybridized strongly to the 44-MDa plasmid present within HD-1 (lane 1), which harbors a *cryIA(b)* gene (25). Hybridization of the *PvuII* probe to this plasmid was expected, since the nucleotide sequence of the probe is highly conserved among all three *cryIA* genes. Similarly, the *PvuII* probe also hybridized to the large 110-MDa plasmid in strain HD-1 containing the *cryIA(a)* and *cryIA(c)* toxin genes (25).

The *PvuII* probe also hybridized to the 45-MDa plasmid encoding the *cryIA(b)* gene in strain EG6345. Differences in the hybridization signal intensity of the *PvuII* probe in detecting the *cryIA(b)* gene in strains HD-1 and EG6345 may be attributed to inconsistent amounts of DNA loaded onto the gel (Fig. 5A). Lack of hybridization by the *PvuII* probe to a 45-MDa plasmid in strain EG6346 (Fig. 5B, lane 3) was consistent with the absence of the 45-MDa plasmid in this cured derivative of EG6345. The 115-MDa plasmid present within strains EG6345 and EG6346 was weakly detected by the *PvuII* probe (Fig. 5B, lanes 2 and 3).

An autoradiogram showing hybridization of the *cryIF* *PstI-SacI* probe to plasmid DNAs from strains HD-1 (lane 1), EG6345 (lane 2), and EG6346 (lane 3) is shown in Fig. 5C. The *cryIF* probe failed to hybridize to plasmids harboring

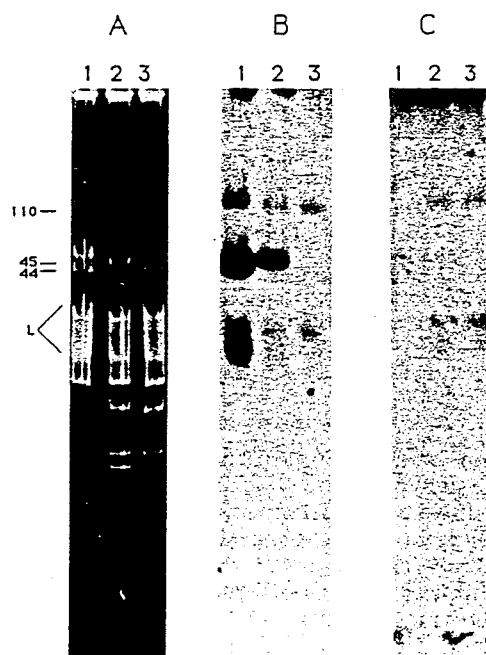


FIG. 5. Plasmid location of *cryIF*. (A) Ethidium bromide-stained agarose gel resolving plasmid DNA prepared from *B. thuringiensis* HD-1 (lane 1); EG6345 (lane 2), and EG6346 (lane 3). (B) Southern blot of the gel in panel A probed with the  $^{32}$ P-labeled 2.2-kb *PvuII* intragenic fragment obtained from the *cryIA(a)* gene of HD-1 (lanes as in panel A). (C) Southern blot of the gel in panel A probed with the  $^{32}$ P-labeled 0.4-kb *PstI-SacI* intragenic N-terminal fragment of *cryIF* (lanes as in panel A). Numbers at the left indicate approximate plasmid mass in megadaltons. L, low-molecular-weight smear. Hybridizations were conducted at high stringency as described in Materials and Methods.

*cryIA* genes in strain HD-1 or EG6345 but did hybridize to the 115-MDa plasmid present in strains EG6345 and EG6346, indicating that the 115-MDa plasmid contains the *cryIF* gene. The comparative intensity of the hybridization signal obtained is weak. This is particularly evident in comparison with the signal that results from hybridization of the *PvuII* probe to the *cryIA* gene(s) present in HD-1 and EG6345. Several factors may be responsible for this weak hybridization signal. First, the *PvuII* probe hybridizes strongly to all *cryIA* genes located on the 110- and 45-MDa toxin plasmids of HD-1. In contrast, the *cryIF*-specific *PstI-SacI* probe, which is derived from an N-terminal region of *cryIF*, which has limited sequence homology to the *cryIA* genes, is specific for the detection of *cryIF* only. In addition, high-molecular-weight, closed covalent circular DNA, such as that represented by the 115-MDa plasmid, may be less efficiently transferred by the Southern blot procedure than are smaller plasmids (such as the 45-MDa plasmid shown in Fig. 5B). Finally, the relative sizes of the probes used in Fig. 5B and C (2.2 kb for the *PvuII* probe in Fig. 5B and 0.4 kb for the *cryIF* probe in Fig. 5C) may also contribute to the comparatively weaker signal in Fig. 5C. Both the *PvuII* and the *PstI-SacI* probes hybridized to a low-molecular-weight smear (L in Fig. 5A), which likely represents shearing of the larger toxin plasmids.

As shown in Fig. 1C, the 0.4-kb *PstI-SacI* intragenic *cryIF*-specific probe detected a 10.4-kb *HindIII* fragment

TABLE 2. Sequence homology of *cryIF* and CryIF to other ICP genes and proteins<sup>a</sup>

Gene	Homology with <i>cryIF</i>		
	DNA	Amino acid	
		Total	N-terminal region <sup>b</sup>
<i>cryIA(a)</i>	77.6	71.7	51.0 (1-608)
<i>cryIA(b)</i>	75.8	70.4	52.0 (1-609)
<i>cryIA(c)</i>	75.8	69.9	49.0 (1-610)
<i>cryIB</i>	66.6	58.3	40.1 (1-637)
<i>cryIC</i>	75.3	70.0	48.8 (1-617)
<i>cryID</i>	75.6	71.5	52.0 (1-593)
<i>cryIE</i>	77.2	69.8	48.1 (1-602)
<i>cryIIA</i>	43.9	24.6	
<i>cryIIIA</i>	53.0	35.6	
<i>cryIVD</i>	44.5	20.8	

<sup>a</sup> The *cryIF* and CryIF sequences were compared with the sequences of *cryIA(a)* (36), *cryIA(b)* (19), *cryIA(c)* (2), *cryIB* (5), *cryIC* (21), *cryID* (4), *cryIE* (4), *cryIIA* (10), *cryIIIA* (11), and *cryIVD* (9). Values denote percent positional identity as determined by Queen and Korn (32).

<sup>b</sup> Amino acids 1 to 602 of the CryIF protein were compared with the N-terminal regions of other CryI proteins as indicated above.

379 codons (ORF2), was identified at nucleotide position 4508. ORF2 terminates with the GATC *Sau3A* cloning site delimiting the insert DNA. The sequence described for ORF2 may represent an artificially truncated version of the native gene present within *B. thuringiensis* EG6346. Although a ribosome binding site has been identified upstream of ORF2, we were unable to identify -10 and -35 promoter sequences homologous to those already described for other *cryI* genes within the intervening DNA sequence between *cryIF* and ORF2.

The sequence analysis program of Queen and Korn (32) was used to compare the *cryIF* and CryIF sequences with the published sequences of other *B. thuringiensis* insecticidal crystal genes and proteins (Table 2). For comparisons between genes of widely differing lengths, such as *cryIF* (3.5 kb) and *cryIVD* (1.9 kb), alignments were performed as follows. Alignment of sequences was first determined by using full-length sequences for both genes. The sequence of the larger of the two genes was then truncated at the last nucleotide shared between the two genes, and the two sequences were realigned to determine the percentage of matched nucleotides within the general area of homology defined by the first alignment. Full-length amino acid sequences were similarly compared.

As deduced from Table 2, the nucleotide sequence of the *cryIF* gene is only about 67 to 78% homologous (positionally identical) to those of the *cryIA* subgroup, *cryIB*, *cryIC*, *cryID*, and *cryIE* genes. Among these crystal protein gene sequences, the DNA sequence of *cryIF* was most homologous to the *cryIA(a)* nucleotide sequence from *B. thuringiensis* HD-1, with 77.6% of the nucleotides conserved between the two genes. Nucleotide sequence comparisons between the *cryIF* and *cryII*, *cryIII*, and *cryIV* genes revealed, as expected, significantly less homology. The sequence of the *cryIIA* gene was most divergent, with only 43.9% of the nucleotides conserved between the two genes.

Comparisons of the amino acid sequences generally reflect results obtained between nucleotide sequence comparisons. Again, the CryIF protein sequence was distinct from but significantly homologous to the other CryI proteins, with 58 to 72% of the amino acids conserved. Although the nucleotide sequence of *cryIIA* was least related to that of *cryIF*, the

CryIIA protein was slightly more related to the CryIF protein than was CryIVD (24.6% versus 20.8% shared amino acids).

Particular attention was focused on the 5'-terminal region of the *cryIF* gene, since this region has been shown to encode the active toxin moiety of other CryI ICPs (2, 19, 35). Amino acid homologies ranged between 49 and 52% for the N-terminal region of the CryIF protein and similar domains in the CryIA, CryIC, and CryIE proteins. The CryIF N-terminal amino acid sequence was less related to CryIB (40%).

**Expression of *cryIF* in *B. thuringiensis*.** Previous reports from this laboratory (11) have indicated that *E. coli* cells harboring *B. thuringiensis* ICP genes fail to produce significant amounts of toxin protein when ICP genes are expressed from their native promoters. Returning the cloned *B. thuringiensis* ICP gene to a *Bacillus* species, and ideally to a *B. thuringiensis* host, maximizes ICP gene expression from its native promoter and enables sufficient crystal protein yields required for critical evaluations of insecticidal activity. To this end, the cloned *cryIF* gene was introduced into an acrySTALLIFEROUS recipient *B. thuringiensis* strain, HD73-26, as described below.

The pEG640 plasmid construct was ligated to the modified pBC16 vector pEG434 (31) at the unique *HindIII* site present on both pEG640 and pEG434. The resulting recombinant plasmid, designated pEG642 (Fig. 2B), possessed both *E. coli* and *Bacillus* replication origins and a selectable marker (*tet*) that encoded tetracycline resistance in a *B. thuringiensis* host. A previous report documented increased transformation efficiency of *B. thuringiensis* strains with DNA isolated from GM2163, an *E. coli* mutant strain defective for both adenine and cytosine methylation (31). Therefore, pEG642 plasmid DNA was first used to transform *E. coli* GM2163. Plasmid DNA prepared from this recombinant strain (GM2163 containing plasmid pEG642) was used to transform the *B. thuringiensis* Cry<sup>-</sup> recipient strain HD73-26 by electroporation. A single tetracycline-resistant *B. thuringiensis* HD73-26 transformant, strain EG1945, contained pEG642, as verified by restriction enzyme and hybridization analyses (data not shown), and was chosen for further study. Microscopic examination of sporulated *B. thuringiensis* EG1945 cultures revealed the presence of crystalline inclusions (large, irregularly shaped rods and bipyrAmidals).

Renografin gradient-purified crystal protein from strain EG1945 was used for SDS-PAGE analyses of the *cryIF* gene product. The Renografin-purified CryIF protein from the *B. thuringiensis* EG1945 recombinant strain was compared with similarly purified proteins obtained from the native *B. thuringiensis* isolates, EG6345 and EG6346 harboring the *cryIF* gene. A single large protein of approximately 135 kDa was observed in strain EG1945 (Fig. 4, lane 3), consistent with expression of *cryIF* in this background. The size of the observed protein correlates well with the predicted molecular mass of 134 kDa deduced from the amino acid sequence. At least three distinct protein species were observed in EG6345 (lane 1), which confirms the DNA hybridization analyses (Fig. 1) and verifies the presence of the *cryIA(b)*, *cryIC*, and *cryIF* genes in this strain. It is possible, however, that other proteins of similar size encoded by additional toxin genes are present in EG6345 which are not resolved under these electrophoretic conditions. *B. thuringiensis* EG6346 (lane 2), which was used to construct the library from which *cryIF* was cloned, produces at least two ICPs, the largest of which appears to comigrate with the 135-kDa CryIF protein produced by the recombinant strain EG1945.



TABLE 3. Insecticidal activity of CryIF protein against several neonate lepidopteran larvae

Crystal protein	50% Lethal concn (ng of ICP/mm <sup>2</sup> of diet surface) <sup>a</sup>			
	<i>H. virescens</i>	<i>H. zea</i>	<i>O. nubilalis</i>	<i>S. exigua</i>
CryIA(b)	0.68 [2] (0.57–0.82)	16.9 [2] (13.0–24.3)	0.27 [2] (0.22–0.35)	38.8 [8] (34.0–44.8)
CryIA(c)	0.04 [4] (0.03–0.05)	1.2 [4] (1.1–1.4)	0.11 [4] (0.10–0.13)	>57.0 [1]
CryIF	0.31 [2] (0.26–0.38)	>57.0 [2]	0.27 [4] (0.22–0.33)	25.6 [4] (20.9–31.9)

<sup>a</sup> Numbers of bioassays performed are in brackets; 95% confidence intervals are in parentheses.

present in strains EG6346 (lane 1) and EG6345 (lane 2) but failed to hybridize to *Hind*III-digested HD-1 DNA (lane 3). The 10.4-kb *Hind*III fragment had also hybridized to the *Pvu*II probe at low stringency as described above (Fig. 1B).

**Insect toxicity of the CryIF protein.** The CryIF protein was tested for its insecticidal activity in a bioassay against four lepidoptera; CryIA(b) and CryIA(c) proteins were included for comparison (Table 3). The insecticidal activity profile of the CryIF protein was different from that of CryIA(b) or CryIA(c). The CryIF protein was highly active against *O. nubilalis* and *H. virescens*, was moderately active against *S. exigua*, and demonstrated little activity against *H. zea* at the highest dose tested. CryIF crystals were more active than CryIA(b) crystals against *H. virescens* and *S. exigua*, equivalent in activity against *O. nubilalis*, and significantly less active against *H. zea*. CryIF crystals were significantly more active than CryIA(c) crystals against *S. exigua* and were significantly less active against *H. virescens*, *H. zea*, and *O. nubilalis*.

## DISCUSSION

Using the technique of low-stringency DNA-DNA hybridization and *cryI*-specific DNA probes, we have analyzed a novel *B. thuringiensis* subsp. *aizawai* strain, EG6345, and a cured derivative of strain EG6345, designated EG6346. Both strains produce large bipyramidal crystals during sporulation. Characteristic *Hind*III restriction fragments, associated with the presence of previously described, lepidopteran-active *cryI* genes which could give rise to the observed crystal phenotype, were not detected in EG6346 by Southern blot hybridization analysis. Rather, low-stringency hybridization to the intragenic *Pvu*II probe from *cryIA(a)* detected several atypically sized fragments which suggested the presence of one or more novel toxin genes within *B. thuringiensis* EG6346. Subsequent screening of an EG6346 *E. coli* genomic library identified one recombinant plasmid, pEG640, which contained a 5.7-kb insert.

Sequencing of the entire pEG640 insert, and subsequent comparison with the published sequences of other toxin genes, resulted in the identification of one intact novel toxin gene, which we have designated *cryIF*, and an additional novel *CryI*-related sequence, designated ORF2. Comparison of the CryIF amino acid sequence with that of ORF2 indicates that the N-terminal regions of these two sequences, compared over the length delimited by the truncation of

ORF2, are quite distinct, with only 35.8% of the amino acids positionally aligned.

Justification for the *cryIF* designation derives from a consideration of DNA and amino acid sequence comparisons with other *cryI* genes and ICPs, as well as the *cryIF* gene product's potent activity against several lepidopteran insects. Amino acid sequence comparisons of CryIF with CryIA, CryIB, CryIC, CryID, and CryIE proteins show, at best, only a 72% conservation of amino acid sequence. In contrast, CryIA subgroup crystal proteins are greater than 80% homologous. More important is the unique N-terminal amino acid sequence of the CryIF protein, which is at most 52% homologous to that of the other *CryI* proteins. The insecticidal activity spectrum of the CryIF protein was likewise distinct from those of the other *CryI* crystal proteins tested. Significant larvicidal activity was observed for a number of important lepidopteran pests, including *H. virescens* (tobacco budworm), *S. exigua* (beet armyworm), and *O. nubilalis* (European corn borer).

The CryIF amino acid sequence was analyzed for the presence of the five conserved domains or homology boxes which have been previously identified for the *CryI*, *CryIII*, and *CryIV* ICPs (20, 27). Not surprisingly, all five conserved domains are present in CryIF. The box 1 and 2 conserved domains are highly hydrophobic and have been hypothesized to comprise a toxicity domain capable of membrane insertion (16). Interestingly, greater homology was evident between CryIF and CryIA(b) at homology boxes 1 and 2 than was present at homology box 3. At box 3, homologies between CryIF and CryIA(b) and between CryIF and CryIC were 63 and 76%, respectively.

Experiments are currently under way to clone and characterize the potential gene suggested by the presence of ORF2 from the EG6346 genomic library. Of interest was the close proximity of *cryIF* to the ORF2 sequence, with an intervening sequence of only 500 bp separating the open reading frames. Although sequence inspection located a potentially functional promoter sequence upstream of the *cryIF* open reading frame, a similar sequence was not observed for ORF2. Sequence inspection has identified, however, a putative termination structure within the 500-bp intervening sequence at nucleotide positions 4090 to 4132 that is nearly identical to the termination structure described for the *cryIA(a)* gene of the HD-1-Dipel strain (41).

In conclusion, a novel ICP gene, *cryIF*, has been identified that directs the synthesis in *B. thuringiensis* of a 133.6-kDa protein, CryIF, with significant insecticidal activity against *H. virescens*, *S. exigua*, and *O. nubilalis* larvae. The characterization of novel ICP genes, such as *cryIF*, furthers our understanding of the molecular genetic basis and diversity of *B. thuringiensis* ICP specificity. This diversity of ICPs is of paramount importance in the creation of new *B. thuringiensis*-based bioinsecticides by using both microbial genetic and recombinant DNA techniques. It will be of interest to determine whether the CryIF active toxin moiety binds to a midgut epithelium receptor population that is distinct from those identified for other *CryI* proteins (18, 38). Lepidopteran-active ICPs with distinct receptor binding characteristics are of particular interest, since recent evidence indicates that laboratory-selected insect resistance to a specific ICP can be correlated with a reduced affinity of the membrane receptor for that protein (39). *B. thuringiensis* bioinsecticide products composed of multiple ICPs that interact with distinct membrane receptors may therefore be less likely to lead to resistant insect populations (39).

## ACKNOWLEDGMENTS

We thank J. Baum, B. C. Carlton, and W. Donovan for critical reading of the manuscript; M. Von Tersch and Helen Loidl Robbins for providing strains; J. M. González, Jr., for providing strains and the recipe for M55 medium; D. Coyle for assistance with insect toxicity assays; and S. Brussock for assistance with qualitative analyses of ICP. We also thank B. Stakes for excellent secretarial assistance.

This research was supported in part by a Phase II SBIR National Science Foundation grant to Ecogen Inc.

## REFERENCES

- Abbott, W. S. 1925. A method of computing the effectiveness of an insecticide. *J. Econ. Entomol.* 18:265-267.
- Adang, M. J., M. J. Staver, T. A. Rocheleau, J. Leighton, R. F. Barker, and D. V. Thompson. 1985. Characterized full-length and truncated plasmid clones of the crystal protein of *Bacillus thuringiensis* subsp. *kurstaki* HD-73 and their toxicity to *Manduca sexta*. *Gene* 36:289-300.
- Birnboim, H. C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* 7:1513.
- Botterman, J., M. Peferoen, H. Höfte, and H. Joos. 1990. European patent application publication no. 358 557.
- Brizzard, B. L., and H. R. Whiteley. 1988. Nucleotide sequence of an additional crystal protein gene cloned from *Bacillus thuringiensis* subsp. *thuringiensis*. *Nucleic Acids Res.* 16:4168-4169.
- Brussock, S. M., and T. C. Currier. 1990. Use of sodium dodecyl sulfate-polyacrylamide gel electrophoresis to quantify *Bacillus thuringiensis* delta-endotoxins, p. 78-87. In L. A. Hickie and W. L. Fitch (ed.), *Analytical chemistry of Bacillus thuringiensis*. American Chemical Society, Washington, D.C.
- Burges, H. D. (ed.). 1981. *Microbial control of pests and plant diseases 1970-1980*. Academic Press, Inc., (London), Ltd., London.
- Chambers, J. Unpublished data.
- Daum, R. J. 1970. Revision of two computer programs for probit analysis. *Bull. Entomol. Soc. Am.* 16:10-15.
- Donovan, W. P., C. C. Dankocsik, and M. P. Gilbert. 1988. Molecular characterization of a gene encoding a 72-kilodalton mosquito-toxic crystal protein from *Bacillus thuringiensis* subsp. *israelensis*. *J. Bacteriol.* 170:4732-4738.
- Donovan, W. P., C. C. Dankocsik, M. P. Gilbert, M. C. Gawron-Burke, R. G. Groat, and B. C. Carlton. 1988. Amino acid sequence and entomocidal activity of the P2 crystal protein. *J. Biol. Chem.* 263:561-567.
- Donovan, W. P., J. M. González, Jr., M. P. Gilbert, and C. C. Dankocsik. 1988. Isolation and characterization of EG2158, a new strain of *Bacillus thuringiensis* toxic to coleopteran larvae, and nucleotide sequence of the toxin gene. *Mol. Gen. Genet.* 214:365-372.
- Finney, D. J. 1971. *Probit analysis*. Cambridge University Press, London.
- Goldberg, L. J., and J. Margalit. 1977. A bacterial spore demonstrating rapid larvicidal activity against *Anopheles sergenti*, *Uranotaenia unguiculata*, *Culex univittatus*, *Aedes aegypti*, and *Culex pipiens*. *Mosquito News* 37:355-358.
- González, J. M. 1984. Ph.D. dissertation. University of Georgia, Athens.
- González, J. M. Unpublished data.
- González, J. M., Jr., B. J. Brown, and B. C. Carlton. 1982. Transfer of *Bacillus thuringiensis* plasmids coding for delta-endotoxin among strains of *B. thuringiensis* and *B. cereus*. *Proc. Natl. Acad. Sci. USA* 79:6951-6955.
- González, J. M., Jr., H. T. Dulmage, and B. C. Carlton. 1981. Correlation between specific plasmids and delta-endotoxin production in *Bacillus thuringiensis*. *Plasmid* 5:351-365.
- Haider, M. Z., and D. J. Ellar. 1989. Mechanism of action of *Bacillus thuringiensis* insecticidal delta-endotoxin: interaction with phospholipid vesicles. *Biochim. Biophys. Acta* 978:216-222.
- Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166:557-580.
- Hofmann, C., H. Vanderbruggen, H. Höfte, J. Van Rie, S. Jansens, and H. Van Mellaert. 1988. Specificity of *Bacillus thuringiensis* delta-endotoxins is correlated with the presence of high affinity binding sites in the brush border membrane of target insect midguts. *Proc. Natl. Acad. Sci. USA* 85:7844-7848.
- Höfte, H., H. de Greve, J. Seurinck, S. Jansens, J. Mahillon, C. Ampe, J. Vandekerckhove, H. Vanderbruggen, M. van Montagu, M. Zabeau, and M. Vaeck. 1986. Structural and functional analysis of a cloned delta-endotoxin of *Bacillus thuringiensis* berliner 1715. *Eur. J. Biochem.* 161:273-280.
- Höfte, H., and H. R. Whiteley. 1989. Insecticidal crystal proteins of *Bacillus thuringiensis*. *Microbiol. Rev.* 53:242-255.
- Honée, G., T. van der Salm, and B. Visser. 1988. Nucleotide sequence of crystal protein gene isolated from *Bacillus thuringiensis* subspecies *entomocidus* 60.5 coding for a toxin highly active against *Spodoptera* species. *Nucleic Acids Res.* 16:6240.
- King, E. G., and G. G. Hartley. 1985. *Heliothis virescens*, p. 323-328. In P. Singh and R. F. Moore (ed.), *Handbook of insect rearing*, vol. 2. Elsevier Science Publishers, B.V., Amsterdam.
- Krieg, A., A. Huger, G. Langenbruch, and W. Schnetter. 1983. *Bacillus thuringiensis* var. *tenebrionis*: a new pathotype effective against larvae of Coleoptera. *J. Appl. Entomol.* 96:500-508.
- Kronstad, J. W., H. E. Schnepf, and H. R. Whiteley. 1983. Diversity of locations for *Bacillus thuringiensis* crystal protein genes. *J. Bacteriol.* 154:419-428.
- Kronstad, J. W., and H. R. Whiteley. 1986. Three classes of homologous *Bacillus thuringiensis* crystal protein genes. *Gene* 43:29-40.
- Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* 227:680-685.
- Lereclus, D., C. Bourgouin, M.-M. Lecadet, A. Klier, and G. Rapoport. 1989. Role, structure and molecular organization of the genes coding for the parasporal delta-endotoxins of *Bacillus thuringiensis*, p. 255-276. In I. Smith, R. A. Slepecky, and P. Setlow (ed.), *Regulation of prokaryotic development*. American Society for Microbiology, Washington, D.C.
- Lilley, M., R. N. Ruttell, and H. J. Somerville. 1980. Purification of the insecticidal toxin in crystals of *Bacillus thuringiensis*. *J. Gen. Microbiol.* 118:1-11.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Marinus, M. G., M. Carraway, A. Z. Frey, L. Brown, and J. A. Arraj. 1983. Insertion mutations in the *dam* gene of *Escherichia coli* K-12. *Mol. Gen. Genet.* 192:288-289.
- Mettus, A.-M., and A. Macaluso. 1990. Expression of *Bacillus thuringiensis* delta-endotoxin genes during vegetative growth. *Appl. Environ. Microbiol.* 56:1128-1134.
- Queen, C., and L. J. Korn. 1984. A comprehensive sequence analysis program for the IBM personal computer. *Nucleic Acids Res.* 12:581-599.
- Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg. 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. *J. Mol. Biol.* 113:237-251.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467.
- Schnepf, H. E., and H. R. Whiteley. 1985. Delineation of a toxin-encoding segment of a *Bacillus thuringiensis* crystal protein gene. *J. Biol. Chem.* 260:6273-6280.
- Schnepf, H. E., H. C. Wong, and H. R. Whiteley. 1985. The amino acid sequence of a crystal protein from *Bacillus thuringiensis* deduced from the DNA base sequence. *J. Biol. Chem.* 260:6264-6272.

37. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503-517.
38. Van Rie, J., S. Jansens, H. Höfte, D. Degheele, and H. Van Mellaert. 1989. Specificity of delta-endotoxins: importance of specific receptors on the brush border membrane of the midgut of target insects. *Eur. J. Biochem.* 186:239-247.
39. Van Rie, J., W. H. McGaughey, D. E. Johnson, B. D. Barnett, and H. Van Mellaert. 1990. Mechanism of insect resistance to the microbial insecticide *Bacillus thuringiensis*. *Science* 247:72-74.
- 39a. Von Tersch, M. Unpublished data.
40. Whiteley, H. R., and H. E. Schnepf. 1986. The molecular biology of parasporal crystal body formation in *Bacillus thuringiensis* subsp. *kurstaki*. *Annu. Rev. Microbiol.* 40:549-576.
41. Wong, H. C., H. E. Schnepf, and H. R. Whiteley. 1983. Transcriptional and translational start sites for the *Bacillus thuringiensis* crystal protein gene. *J. Biol. Chem.* 258:1906-1967.