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

JUDITH A. CHAMBERS,† AMY JELEN, M. PEARCE GILBERT,‡ 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 ICPs 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- 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% 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 subbranching 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 P1 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. darmstadensis (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.
† Present address: Office of Agriculture, Bureau for Science and Technology, Agency for International Development, Washington, DC 20523.
‡ Present address: 4601 Pinedale Drive, Durham, NC 27705.
### MATERIALS AND METHODS

**Bacterial strains and plasmids.** The strains and plasmids discussed in this report are described in Table 1. *B. thuringiensis* EG6346 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⁻) 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* DH5α 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* DH5α 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× SSC (1× SSC is 1.0 M NaCl plus 0.015 M sodium citrate), 10× Denhardt’s solution (1× Denhardt’s solution is 0.02% bovine serum albumin, 0.02% Ficoll, and 0.02% polyvinylpyrrolidone), 200 μg 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× 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 *SacI* 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 *BamHI*-digested vector DNA at a 1:2 molar ratio of vector to insert and at a final DNA concentration of 20 μg/ml, using T4 DNA ligase (Promega). Transformation into *E. coli* DH5α cells was done according to the manufacturer’s directions. Transformed colonies were plated on LB medium containing 100 μg of ampicillin and 50 μg of X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) per ml. Approximately 3.3 × 10⁶ 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 *PvuII* 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* DH5α 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− Dcm− 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-

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Relevant characteristics</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>B. thuringiensis subsp. kurstaki</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HD73-26</td>
<td>Cry⁻ Cm¹</td>
<td>11</td>
</tr>
<tr>
<td>EG1945</td>
<td>HD73-26(pEG642)</td>
<td>This study</td>
</tr>
<tr>
<td><strong>B. thuringiensis subsp. aizawai</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EG6345</td>
<td>115- and 45-MDa plasmids</td>
<td>This study</td>
</tr>
<tr>
<td>EG6346</td>
<td>115-MDa plasmid, cured of 45-MDa plasmid</td>
<td>This study</td>
</tr>
<tr>
<td><strong>E. coli</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DH5α</td>
<td>Amp⁴</td>
<td>Bethesda Research Laboratories</td>
</tr>
<tr>
<td>GM2163</td>
<td>Amp⁴ Dam− Dcm−</td>
<td>30</td>
</tr>
<tr>
<td>EG1943</td>
<td>DH5α(pEG640)</td>
<td>This study</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pEG434</td>
<td>Te⁺ <em>Bacillus</em> vector</td>
<td>31</td>
</tr>
<tr>
<td>pGEM-3-Z</td>
<td>Amp⁴ <em>E. coli</em> vector</td>
<td>Promega</td>
</tr>
<tr>
<td>pEG640</td>
<td>pGEM-3Z with 5.7-kb insert of EG6346 DNA (cryI progenitor) DNA (cryI ORF2)</td>
<td>This study</td>
</tr>
<tr>
<td>pEG642</td>
<td>Te⁺, <em>E. coli-Bacillus</em> shuttle vector with pEG640 inserted into HindIII site of pEG434 (cryI ORF2)</td>
<td>This study</td>
</tr>
</tbody>
</table>

---

**TABLE 1. Bacterial strains and plasmids**

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 *SacI* 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 *BamHI*-digested vector DNA at a 1:2 molar ratio of vector to insert and at a final DNA concentration of 20 μg/ml, using T4 DNA ligase (Promega). Transformation into *E. coli* DH5α cells was done according to the manufacturer’s directions. Transformed colonies were plated on LB medium containing 100 μg of ampicillin and 50 μg of X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) per ml. Approximately 3.3 × 10⁶ 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 *PvuII* 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* DH5α 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− Dcm− 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 crylF 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 K2HPO4, 37 mM KH2PO4, 1 mM citric acid · H2O, 5 mM (NH4)2SO4, 150 mM NaCl, 1 mM CuCl2 · 2H2O, 1 mM ZnCl2, trisodium citrate, 1 mM Na3MoO4, 0.3 mM MgCl2 · 6H2O, 5 μM MnCl2 · 4H2O, 0.5 mM CaCl2 · 2H2O, 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 CrylF 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 crylF gene on plasmid pEG642. For comparison, Renografin-purified crystal proteins from recombinant *B. thuringiensis* EG7077, containing the crylA(b) gene, and EG1861, containing the crylA(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 CrylF 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 mm²). 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 replicates 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 CrylF and its deduced amino acid sequence, have been filed with GenBank, Los Alamos National Laboratory, under accession number M63897.

---

**RESULTS**

**Identification and isolation of the crylF 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 EG6345 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 *EcoRI* N-terminal probe (Fig. 1A) or to the 2.2-kb intragenic *PvuII* probe isolated from the *crylA(a)* gene of *B. thuringiensis* HD-1 (Fig. 1B) (24). Digested DNA from *B. thuringiensis* subsp. *kurstaki* HD-1, which harbors the *crylA(a), crylA(b), and crylA(c)* genes, was included as a control.

As shown in Fig. 1A, the 726-bp *EcoRI* 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 *crylA(a)*, *crylA(b)*, and *crylA(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 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 absence of this band is coincidental and does not support the presence of a cryIA(c) gene in EG6346, as confirmed 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 strain 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 recombinant 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
FIG. 3. DNA sequence of the 5.7-kb insert of *B. thuringiensis* EG6346 DNA on pEG640. The sequence is flanked by the Sau3A cloning sites and extends 5,649 nucleotides in length. Two open reading frames are indicated as diagramed 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 sequences for *cryIF* and *ORF2* are shown below the nucleotide sequence.
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 NH2-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 presumptively 30 and 35 regions were homologous to those identified for the HD-1-Dipel cryIA(a) gene BH promoter (41).

A methionine codon, followed by an open reading frame of
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 *cry* 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 *CryI* protein was slightly more related to the *CryIF* protein than was *CryIVD* (24.6% versus 20.8% shared amino acids).

Table 2. Sequence homology of *cryIF* and *CryIF* to other ICP genes and proteins

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

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

^b Amino acids 1 to 602 of the *CryIF* protein were compared with the N-terminal regions of other *Cry* proteins as indicated above.

Table 2 shows the sequence homology of *cryIF* and *CryIF* to other ICP genes and proteins. The *cryIA* gene was used as a reference for comparison. The sequence of *cryIA* was most homologous to the sequence of *cryIF* in terms of both DNA and amino acid sequences. The percentage of homology between the two genes was 77.6% for DNA and 51.0% for amino acids. The *cryII* gene had the least homology with *cryIF*, with only 43.9% of the nucleotides conserved between the two genes. The *cryII* gene had the least homology with *cryIF*, with only 43.9% of the nucleotides conserved between the two genes.
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 EG6345 (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 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 3. Insecticidal activity of CryIF protein against several neonic neopapun larvae

<table>
<thead>
<tr>
<th>Crytal protein</th>
<th>50% Lethal concn (ng of ICP/mg2 of diet surface)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H. virescens</td>
</tr>
<tr>
<td>CryIA(b)</td>
<td>0.68 [2]</td>
</tr>
<tr>
<td></td>
<td>(0.57-0.82)</td>
</tr>
<tr>
<td>CryIA(c)</td>
<td>0.04 [4]</td>
</tr>
<tr>
<td></td>
<td>(0.03-0.05)</td>
</tr>
<tr>
<td></td>
<td>(0.26-0.38)</td>
</tr>
</tbody>
</table>

* Numbers of bioassays performed in brackets; 95% confidence intervals are in parentheses.

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 cryl-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 HindIII restriction fragments, associated with the presence of previously described, lepidopteran-active cryl 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 PvuII 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 Cryl-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 cryl 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 Cryl proteins. The insecticidal activity spectrum of the CryIF protein is likewise distinct from those of the other Cryl 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 Cryl, 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 cryl(a) gene of the HD-1-Dipel strain (41).

In conclusion, a novel ICP gene, cryIF, has been identified that directs the synthesis of 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 Cryl 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


39a. Von Tersch, M. Unpublished data.
