MINIREVIEW

Phylogenetic Relationships of *Bacillus thuringiensis* δ-Endotoxin Family Proteins and Their Functional Domains

ALEJANDRA BRAVO*

*Department of Microbiology, Instituto de Biotecnología, Universidad Nacional
Autónoma de México, Cuernavaca, Morelos, México*

INTRODUCTION

Insecticidal crystal proteins (ICPs) from *Bacillus thuringiensis* have been used as biopesticides for the last 35 years. *B. thuringiensis* is a gram-positive bacterium which produces proteinaceous inclusions during sporulation; these inclusions can be distinguished as distinctively shaped crystals by phase-contrast microscopy. The inclusions are composed of proteins known as ICPs, Cry proteins, or δ-endotoxins, which are highly toxic to a wide variety of important agricultural and health-related insect pests as well as other invertebrates. Due to their high specificity and their safety for the environment, ICPs are a valuable alternative to chemical pesticides for control of insect pests in agriculture and forestry and in the home. It has been proposed that the rational use of *B. thuringiensis* toxins will provide a variety of alternatives for insect control and for coping with the problem of insect resistance to pesticides.

Intensive screening programs have identified strains of *B. thuringiensis* from soil samples, plant surfaces, dead insects, and stored grains from all over the world. The isolated strains show a wide range of specificity against different insect orders (Lepidoptera, Diptera, Coleoptera, Hymenoptera, Homoptera, Phthiraptera or Mallophaga, and Acari) and other invertebrates (Nemathelminthes, Platyhelminthes, and Sarcomastigorphora) (13). Currently 45 different serotypes have been catalogued, representing a total number of 58 serovars (28). Many of the ICP genes have been cloned, sequenced, and classified as *cry* and *cty* genes. The first classification was based on insecticidal activity (23), with the different Cry proteins denoting ICPs toxic to various insect and invertebrate groups as follows: CryI toxic to lepidopterans, CryII toxic to lepidopterans and dipterans, CryIII toxic to coleopterans, CryIV toxic to dipterans, and CryV and CryVI toxic to nematodes (14). Novel *cry* genes isolated recently have created some problems for this classification scheme, especially genes that were homologous to known genes but displayed different specificities and genes that had dual specificity. Recently, a novel nomenclature has been proposed based exclusively on amino acid identity (7). To date, over 50 *cry* gene sequences have been determined and classified into 15 families (7).

The *cry* genes code for proteins with a range of molecular masses from 50 to 140 kDa. Upon ingestion by the susceptible target, the protoxins are solubilized and proteolytically processed to release the toxic fragment (23). During proteolytic activation, peptides are removed from both amino- and car-

boxyl-terminal ends of the protoxin. For the 130- to 140-kDa protoxins, the carboxyl-terminal proteolytic activation removes half of the molecule, resulting in an active toxin fragment of 60 to 70 kDa.

A generally accepted model for Cry toxin action is that it is a multistage process. First, the activated toxin binds to receptors located on the apical microvillus membrane of epithelial midgut cells (6, 22, 49). After the toxin binds the receptor, it is thought that there is a change in the toxin conformation allowing toxin insertion into the membrane. Oligomerization of the toxin follows, and this oligomer then forms a pore that leads to osmotic cell lysis (26a, 30, 32, 40).

Receptor binding is a key factor in specificity. Two different insect proteins have been identified as receptors for Cry toxins, the 120-kDa aminopeptidase N Cry1Ac toxin-binding protein purified from brush border vesicles of *Manduca sexta*, *Heliothis virescens*, and *Lymantria dispar* (20, 25, 41, 48) and the 210-kDa cadherin-like glycoprotein Cry1Ab toxin-binding protein purified from *M. sexta* membranes (47). Specific binding involves two steps, one that is reversible and one that is irreversible. Recent data suggest that toxicity correlates with irreversible binding (31). Irreversible binding might be related to insertion of the toxin into the membrane but could also reflect a tighter interaction of the toxin with the receptor.

The crystal structures of Cry3A (coleopteran-specific) and Cry1Aa (lepidopteran-specific) toxins have been reported (21, 30). The Cry3A protoxin has a molecular mass of 70 kDa and does not contain the large carboxyl-terminal extension contained in the Cry1Aa toxin. The crystal structure of Cry1Aa toxin was determined from the activated toxin fragment. Both toxins share 36% amino acid identity, and the two structures show high overall similarity (21). Both are globular molecules containing three distinct domains connected by single linkers.

Domain I extends from residues 33 to 253 in Cry1Aa and from residues 58 to 290 in Cry3A; it is a seven α-helix bundle in which a central helix (helix α-5) is completely surrounded by six outer helices (Fig. 1A). This domain has been implicated in the channel formation in the membrane. The six α-helices are amphipathic and are long enough to span the 30-Å-thick hydrophobic region of a membrane bilayer. Point mutations in the region encoding the central α-5 helix of the Cry1Ac toxin (residues 163 to 170) drastically affect toxicity without affecting binding to larval midgut vesicles (52).

Residues 265 to 461 in Cry1Aa and 291 to 500 in Cry3A form domain II (Fig. 1B). Domain II consists of three antiparallel β-sheets with similar topologies packed around a hydrophobic core. This domain represents the most divergent part in structure between the two toxin molecules (21). This domain has been described as the specificity-determining domain, since reciprocal hybrid genes between closely related toxins

* Mailing address: Department of Microbiology, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Apdo. Postal 510-3, Cuernavaca, Morelos, México. Phone: (52) (5) 622-7635. Fax: (52) (73) 17-2388. E-mail: bravo@ibt.unam.mx.
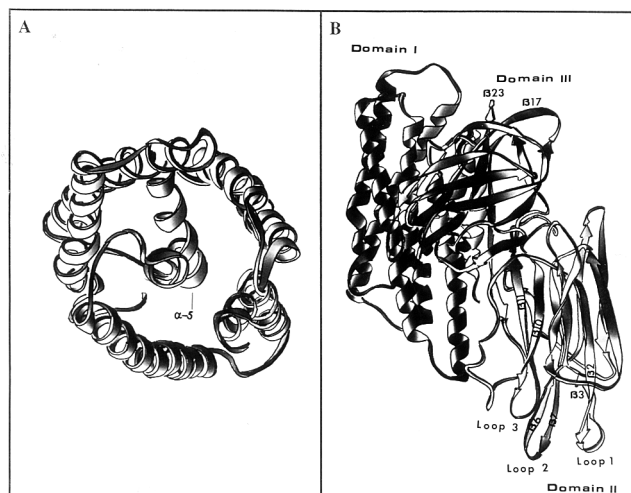
FIG. 1. Crystal structure of Cry3A toxin (crystallography data from Li et al. [30]). (A) A schematic ribbon representation of isolated domain I, showing an upper view of the α-helix bundle and the location of the central helix α-5. (B) The three-domain organization (domains I to III) of Cry3A toxin is shown, with the positions of the three surface-exposed domain II loops (loops 1, 2, and 3) forming the molecular apex of the toxin, and the internal β-sheets of domain III (β-17 and β-23) indicated.

(Cry1Aa and Cry1Ac) resulted in chimeric toxins with altered specificity (19, 42). The two protruding loops oriented parallel with the helical bundle of domain I (loop 1 [between β-2 and β-3] and loop 2 [between β-6 and β-7]) (Fig. 1B) were suggested to be involved in receptor binding in Cry1-type toxins. Mutations located in loop 1 of Cry1Aa toxin demonstrated that these residues are essential for binding to the brush border membrane of *Bombyx mori* midgut cells (33). Also, mutations in this region in the Cry1Ab toxin affect binding to *M. sexta* and *H. virescens* midgut membranes (39). Smith and Ellar (44) showed that mutations in loops 1 and 2 of Cry1C toxin were able to modulate toxicity and specificity. Site-directed mutagenesis analysis of Cry3A toxin showed that in addition to loop 1, loop 3 (between β-10 and β-11) is involved in irreversible binding to *Tenebrio molitor* midgut membranes (53). It is interesting to note that loops 1, 2, and 3 correspond to the regions that showed the largest structural differences between Cry1Aa and Cry3A toxins (21).

Finally, domain III is a β-sandwich of two antiparallel β-sheets (Fig. 1B). This domain comprises residues 463 to 609 in the Cry1Aa toxin and residues 501 to 644 in the Cry3A toxin (21). The function of domain III is still under discussion. It has been proposed that it stabilizes the toxin by protection from proteolysis (30). However, recent reports suggest that it may be involved in channel function as a voltage sensor, since conservative mutations in R521K and R527K reduced toxicity without reducing binding (9). The location of the arginine residues in the three-dimensional structure (β-17 [Fig. 1B]) indicates that they have an important role in stabilizing the structure by forming salt bridges and hydrogen bonds with other residues in the vicinity (21). Consequently, conversion of all arginine residues from β-17 to glutamic acid or glycine resulted in protein instability or poor expression (9). Mutations in β-23 of Cry4A toxin (37) indicated that D670K and E673K caused an unstable conformation, as judged by digestion with trypsin and thermolysin. These two acidic residues form hydrogen bonds with the arginine residues of β-17, supporting the idea that both β-

sheets (β-17 and β-23 [Fig. 1B]) are important determinants for the proper folding of the toxin (21).

Several lines of evidence indicate that domain III may be involved in receptor binding. The construction of chimeric proteins between Cry1Ea and Cry1C toxins has shown that domain III of Cry1C is a significant determinant of specificity to *Spodoptera exigua* and *Mamestra brassicae* (4). Also, it has been demonstrated that domain III exchanges between Cry1Ac and Cry1Aa toxins affect binding to different *L. dispar* midgut receptors (29). Finally, mutations S503I and S504I in Cry1Ac toxin resulted in proteins that bind poorly to the *M. sexta* and *H. virescens* toxin-binding proteins and showed extensive loss of toxicity for both insects (1).

Höfte and Whiteley (23) have identified five highly conserved regions among the sequences of Cry toxins. The locations of these regions in the three-dimensional structures of Cry1Aa and Cry3A are the same: they are found at the central positions of each domain or are involved in interdomain contacts. Li et al. (30) proposed that the high degree of conservation of these blocks and their important structural location would imply that the Cry toxins which possess these blocks would share a similar structure, that of globular toxins composed of three structural domains.

Experimental data from several laboratories have shown that domains from Cry proteins are structurally independent. Domain I (50, 51) and helix α-5 peptides (8, 18), expressed independently, retain their ability to form cation channels in planar lipid bilayers. There have been no reports of domain II or III isolation and expression, but the exchange of sequence segments within domains II and III resulted in specificity changes (19, 42). These observations support the hypothesis that δ-endotoxins have a modular structure and suggest that their different domains could have evolved independently.

Protein sequence analysis of ICPs has been previously done (46, 54). These alignments showed the amino acid identity among the Cry sequences. Nevertheless, the percentage of amino acid identity does not necessarily reflect evolutionary relationships. In order to draw an evolutionary tree for different proteins, it is necessary to calculate the approximate constancy of amino acid substitutions by a distance matrix method or by a maximum parsimony method (36). In this article, an amino acid sequence alignment of some ICPs has been displayed and the phylogenetic relationships of these proteins and their different functional domains were determined. It is demonstrated that in fact the three structural domains show different evolutionary relationships, suggesting that natural selection occurred at the domain level in these proteins.

AMINO ACID SEQUENCE ALIGNMENT

To date, more than 50 sequences of different δ-endotoxins from 21 *B. thuringiensis* subspecies have been reported (7). A multiple sequence alignment of the δ-endotoxins shown in Table 1 was generated by using the Genetics Computer Group (GCG) sequence analysis program PILEUP (11) (Fig. 2). This program uses a simplified version of the progressive alignment method of Feng and Doolittle (16). The procedure begins with the determination of all possible pairwise similarity scores. The two most similar sequences are aligned by using the Needleman and Wunsch algorithm (35) forming the first cluster, and then the next most related sequences are progressively aligned to this cluster. We have analyzed one toxin from each subgroup from Cry1 to Cry14 without including the different alleles (Table 1), and we have withdrawn from our analysis the Cry6 and Cry15 proteins, since they do not share any sequence similarities with the rest of the Cry protein family, not even within the

TABLE 1. δ-Endotoxin sequences used for determination of phylogenetic relationships

ICP[a]	Accession no.	Specificity range[b]	ICP	Accession no.	Specificity range	ICP	Accession no.	Specificity range	ICP	Accession no.	Specificity range
Cry1Aa	M11250	L	Cry1Eb	M73253	L	Cry3Ba	X17123	C	Cry8Ca	U04366	C
Cry1Ab	M13898	L	Cry1Fa	M63897	L	Cry3Bb	M89794	C	Cry9Aa	X58120	L
Cry1Ac	M11068	L	Cry1Ga	Z22510	L	Cry3Ca	X59797	C	Cry9Ba	X75019	L
Cry1Ad	M73250	L	Cry1Ha	Z22513	L	Cry4Aa	Y00423	D	Cry9Ca	Z37527	L
Cry1Ae	M65252	L	Cry1Ia	X62821	L, C	Cry4Ba	X07423	D	Cry10Aa	M12662	D
Cry1Ba	X06711	L, C	Cry1Ib	U07642	L, C	Cry5Aa	L07025	N	Cry11Aa	M31737	D
Cry1Ca	X07518	L, D	Cry1Ja	L32019	L	Cry5Ab	L07026	N	Cry12Aa	L07027	N
Cry1Cb	M97880	L	Cry2Aa	M23723	L, D	Cry7Aa	M64478	C	Cry13Aa	L07023	N
Cry1Da	X54160	L	Cry2Ab	M23724	L	Cry7Ab	U04367	C	Cry14Aa	U13955	C
Cry1Db	Z22511	L	Cry2Ac	X57252	L	Cry8Aa	U04364	C			
Cry1Ea	X53985	L	Cry3Aa	M22472	C	Cry8Ba	U04365	C			

[a] Crystal proteins are named according to the revised nomenclature (7).
[b] Abbreviations: L, lepidopteran; C, coleopteran; D, dipteran; N, nematode.

conserved regions described by Höfte and Whiteley (23). The analyzed proteins represent toxins active against different insect orders (Lepidoptera, Coleoptera, and Diptera) and nematodes. The alignment presented here is accurate since it was done by modulating the gap creation and gap extension penalties in order to have an optimum alignment of the structural motifs (α-helices and β-sheets) delimited in the three-dimensional structure analysis of Cry1Aa and Cry3A toxins (21, 30). This alignment was used to establish the limits of each domain, as well as the carboxyl-terminal region of the protoxin. To optimize the individual alignments, each domain was realigned first with the GCG PILEUP program and then improved manually.

As has been previously described, ICPs have substantially higher sequence similarity in the protoxin segment from the end of β-23 through the carboxyl terminus (Fig. 2). The toxic fragment is less conserved than the protease-susceptible fragment. Among the three domains of the toxin, domain I has the highest similarities, specially within helix α-5 and α-7; domain II is the least-conserved domain among all toxins. Cry2 and Cry11 toxins share significant homology only within domain I, but almost no similarity is found within domains II and III.

ESTIMATION OF PHYLOGENETIC RELATIONSHIPS

Many proteins from different organisms are composed of several structural domains, some of which have shown independent domain evolution (2, 34). As discussed above, ICPs display a modular structure; however, it is not well understood if the different domains could function independently and how they have evolved. In order to estimate the phylogenetic relationships of ICPs and of each of their functional domains, the genetic distances among the Cry sequences were calculated with the PROTDIST program of J. Felsenstein's PHYLIP 3.5 phylogeny inference package with the Dayhoff PAM matrix (15), using the alignment previously obtained (Fig. 2). This program computes a distance measure for protein sequences. The distance that is computed is scaled in units of expected fraction of amino acids changed. The FITCH program (15) was then used to estimate phylogenies from the distance matrix data under the additive tree model, in which the distances are expected to equal the sums of branch lengths between the species. This program uses the method of Fitch and Margoliash (17) and the least-squares criterion. Phylogenetic analyses were also done by the parsimony method using the PROT-PARS program (15). The principle of this method is to infer the amino acid sequences of the ancestral species and choose a tree that requires the minimum number of mutational changes. Both types of phylogenetic analyses were carried out 100 times in order to get a strict consensus tree by using the bootstrapping tool which generates multiple data sets that are resampled versions of the input data set. The consensus phylogenetic trees were computed by the CONSENSE program (15).

The phylogeny of the entire Cry protoxin sequences is shown in Fig. 3A. The circled branches are branches with accurate topology, since they were found in more than 90% of the trees. This analysis demonstrated that some of the different protoxin classes (Cry1, Cry3, and Cry7) form independent clusters. The Cry1 protoxins are arranged in one main group. The branch between Cry1 protoxins and the rest of the Cry family is present in 100% of the trees, indicating a clear subdivision between Cry1 and the rest of the protoxins. In contrast, the nematode-specific protoxins (Cry5, Cry12, and Cry13) and the Cry14 toxin cluster together, indicating their close relationship in the evolutionary process. Also, Cry8 and Cry9 protoxins are arranged in the same branch as the Cry2 and the Cry11 protoxins, which also suggests a common origin for these protoxins.

The phylogenetic relationships of the presumed toxic portion were analyzed. The topology of the obtained phylogenetic tree (Fig. 3B) is different from that of the protoxin tree. There are three main groups, the first one containing some of the lepidopteran-specific toxins. One of the main differences between both phylogenetic trees (toxin versus protoxin) is that Cry1B, Cry1Ia, and Cry1Ib toxins are not contained in the same group of the Cry1 toxins. These three Cry1 toxins are arranged in the second group consisting of the toxins that have shown activity against coleopteran insects (Cry3, Cry7, Cry8, Cry1I, and Cry1B toxins) (5, 45). The second group did not consist solely of coleopteran-active toxins, since two lepidopteran-specific toxins (Cry9Ba and Cry9Ca) were also in this group. The Cry9Aa toxin is located far away from the other two Cry9Ba and Cry9Ca toxins, suggesting that Cry9Aa toxin evolved independently from the other Cry9 toxins. In fact, the Cry9Aa protoxin is homologous to the Cry9Ba and Cry9Ca protoxins only at the carboxyl-terminal end. Finally, in the third group, very different toxins are localized. This group is composed of the dipteran-specific toxins (Cry4, Cry10, Cry11, and Cry2Aa), the nematode-specific toxins (Cry5, Cry12, and Cry13), the Cry14A toxin, and the lepidopteran-specific Cry2 toxins which are smaller and very different from the rest of the lepidopteran-specific toxins.

The main difference between the protoxin and the toxin sequences is the large carboxyl-terminal end contained in the
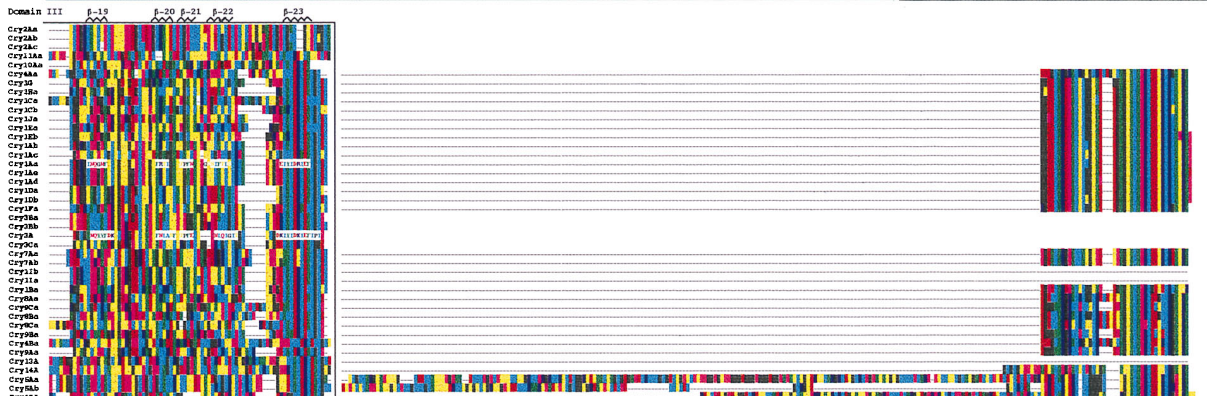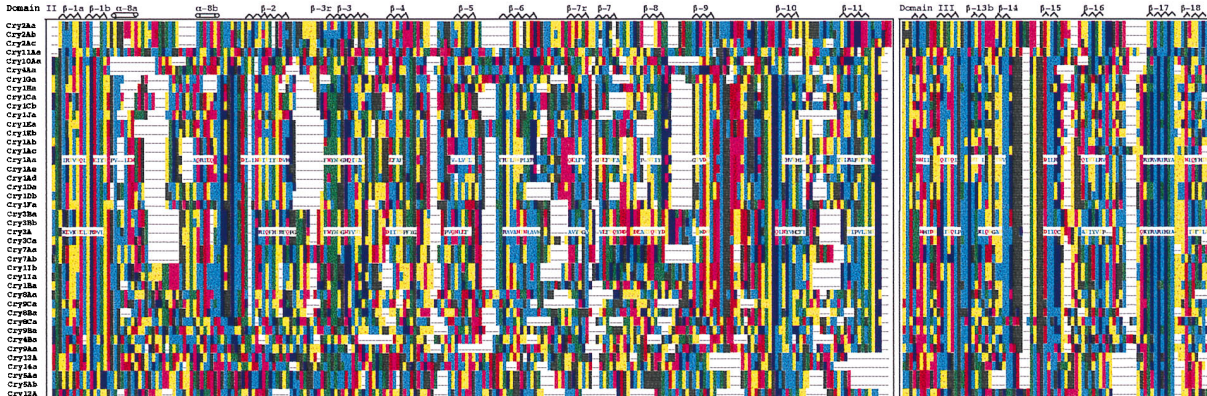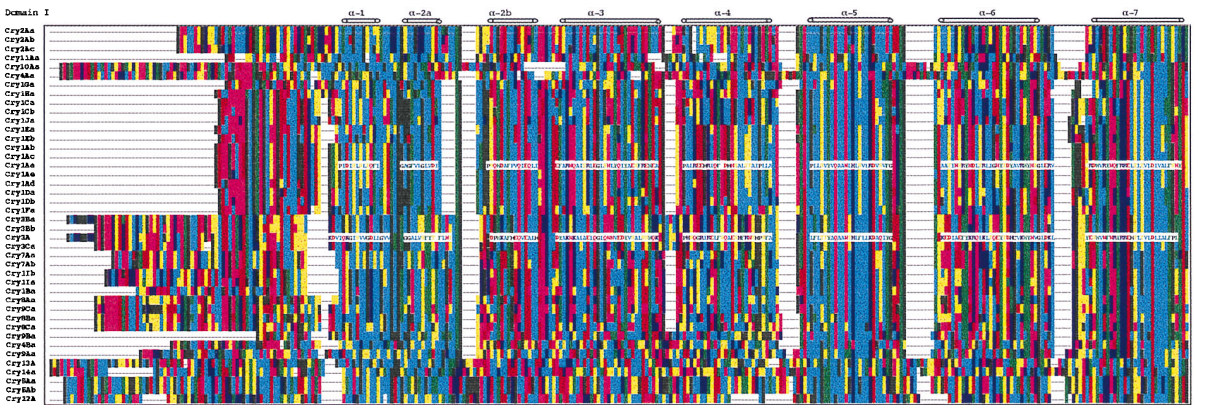
FIG. 2. Amino acid sequence alignment of the δ-endotoxin (Cry) protein family. The sequences of 42 Cry proteins were aligned by using the GCG program PILEUP (11). Amino acid sequences were translated from the nucleotide sequences indicated by the GenBank database accession numbers (Table 1). The code for colors is as follows: red, acidic residues (Glu and Asp); rose, amide residues (Asn and Gln); dark blue, basic residues (Lys, Arg, and His); light blue, aliphatic residues (Ala, Val, Leu, and Ile); green, aromatic residues (Phe, Tyr, and Trp); yellow, hydroxyl residues (Ser and Thr); black, the remaining residues (Gly, Pro, Met, and Cys). The boxes indicate the limits of each domain. The structural motifs (α-helices and β-sheets) delimited in the three-dimensional structures of Cry1Aa and Cry3A toxins (20, 29) are presented in color characters on a white background and indicated at the top of the boxes by a cylinder (α-helices) or a broken line (β-sheets).

protoxin sequence. This fragment is highly conserved among some of the protoxin sequences; the putative function of this long carboxyl-terminal segment is to aid in the formation of an ordered crystalline array. Since most of the cysteine residues are located in this fraction of the protoxin, it has been suggested that the alkaline and reducing conditions required for the solubility of these proteins are related to disulfide bridge formation within the protoxin fragment (10). However, this fragment is not found in some toxins (Cry3A, Cry3Ba, Cry3Bb, Cry3Ca, Cry2Aa, Cry2Ab, Cry2Ac, and Cry11Aa) or is very small in some other protoxins, like Cry1Ia and Cry1Ib (75 residues) and Cry13A (111 residues). Figure 3C shows the consensus phylogenetic tree obtained with the carboxyl-terminal end of the protoxin. The topology of this tree is rather similar to the topology of the tree obtained with the complete protoxin sequence (Cry1 toxins far away from the rest of the toxins; the nematode-specific toxins and Cry14 clustered in the same group and Cry8 and Cry9 toxins arranged in the same branch), implying that the difference in the obtained protoxin and toxin phylogenetic trees is due principally to the presence of the carboxyl-terminal end.

PHYLOGENETIC RELATIONSHIP ESTIMATIONS FOR THE TOXIN DOMAINS

Domain I. In the current δ-endotoxin mode of action model, domain I has been thought to be responsible for the toxic activity in the membrane. It has been proposed that after the toxin binds to the receptor, there is a change in the conformation of this domain allowing the hydrophobic surfaces of the helices to face the exterior of the bundle, leading to insertion into the membrane and the formation of ion channels (26a, 30).

The obtained phylogenetic tree is composed of three main groups (Fig. 4A). The first group contains domain I from lepidopteran-specific toxins (most of the Cry1 toxins excluding Cry1B, Cry1Ia, and Cry1Ib). This group is very reliable since all tree branches were found in more than 90% of the analyzed trees. These data suggest that all the domain I sequences from this group of lepidopteran-specific toxins have evolved from a common ancestor. The second group is comprised of domains I from coleopteran (Cry3, Cry7, and Cry8) and lepidopteran-coleopteran (Cry1I and Cry1B)-specific toxins. Within this group, domains I from Cry9Ba and Cry9Ca are also localized. Finally, the third main group is
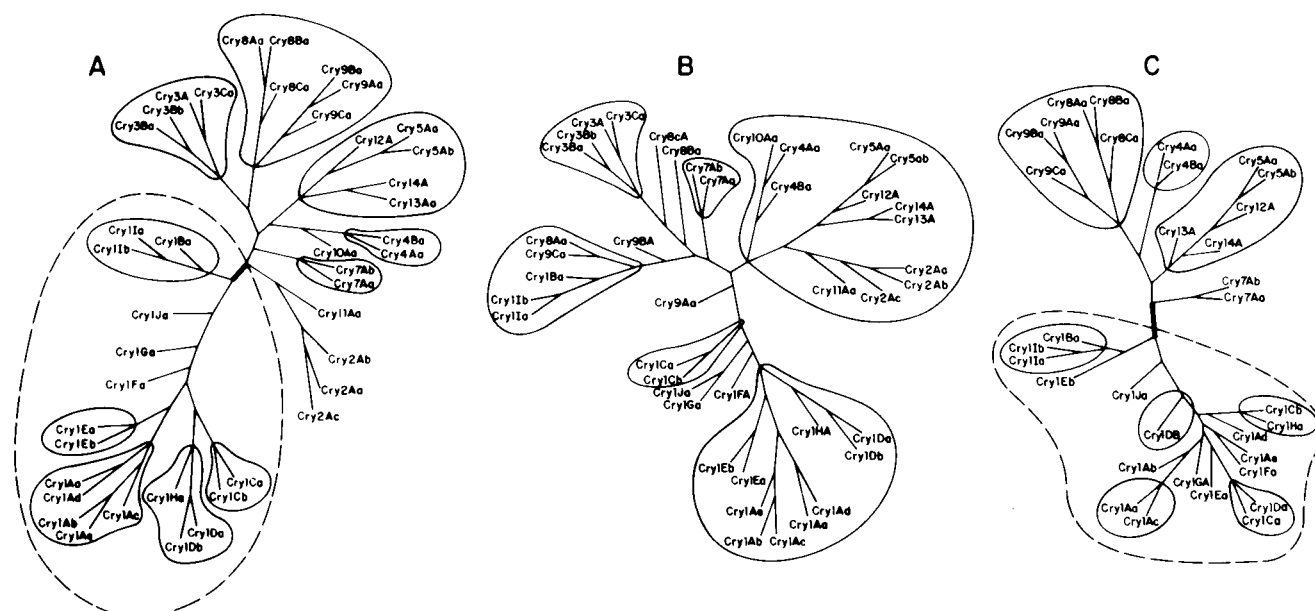

FIG. 3. Unrooted phylogenetic trees of the entire protein sequence (A), the toxin fragment (B), and the carboxyl-terminal fragment (C) of the δ-endotoxin protein family. Phylogenetic analysis of 42 Cry sequences was performed. Initially, a multiple sequence alignment of all the members of this protein family was generated by using the GCG sequence analysis package program PILEUP (11). The alignment obtained was further refined with the manual multiple alignment program LINEUP. For each fragment, the alignments were generated independently of the alignment of the entire sequences. The genetic distances were calculated by using the Dayhoff PAM matrix with the program PROTDIST of J. Felsenstein's PHYLIP 3.5 phylogeny inference package (15). Subsequently, the phylogenetic relationships of these sequences were determined by the method of Fitch and Margoliash (17) and by using the least-squares criterion and the FITCH program (15). Finally, the phylogenetic analyses were carried out 100 times in order to get a strict consensus tree by using the bootstrapping tool and the CONSENSE program (15). Circled branches are branches that were found in more than 90% of the trees. The branch between Cry1 sequences and the rest of the Cry family is present in 100% of the trees, and the broken-line circle indicates a subdivision between Cry1 and the rest of the sequences.
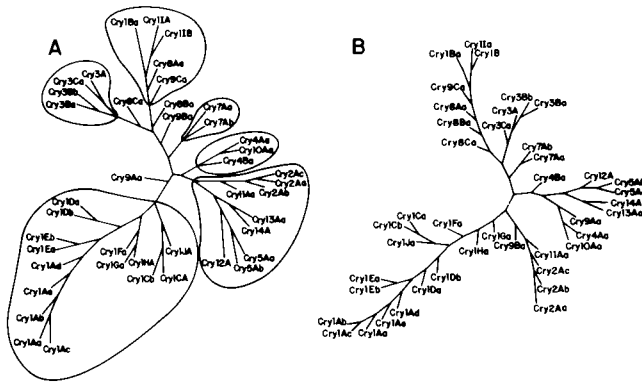
FIG. 4. Unrooted phylogenetic trees of the domain I sequences of the δ-endotoxin protein family. The phylogenetic analysis was performed by Fitch and Margoliash's method (17) and by using the least-squares criterion and the FITCH program (15) as described in the legend to Fig. 3 (A) and by the parsimony method with the PROTPARS program (14) (B). Circled branches are branches that were found in more than 90% of the trees.

formed by domains I from very different toxins (nematode, dipteran, and the Cry2 lepidopteran-specific toxins). However, each specificity group is clustered in separate small branches; domains I from the nematode toxins (Cry5, Cry12, and Cry13) are arranged together in a single branch, as are the dipteran-specific toxins (Cry4 and Cry10) as well as the small Cry2 and Cry11 toxins. The locations of domain I sequences from Cry8Ca, Cry8Ba, Cry9Ba, and Cry9Aa within the tree topology are not very reliable since their locations within the consensus tree were found in 52, 40, 70 and 34% of the analyzed trees, respectively. However, the phylogenetic relationships obtained by parsimony analysis of domain I (Fig. 4B) confirmed the distribution of domains I in three main groups, one composed of domains I from lepidopteran-specific toxins, a second composed of domains I from all the Cry toxins that have showed activity against coleopteran insects, including domain I from Cry9Ca toxin (lepidopteran specific), and a third group composed of domains I from the dipteran- and nematode-specific toxins and the Cry2 lepidopteran-specific toxins. Taking together the phylogenetic data obtained by the two different methodologies (maximum parsimony and the FITCH program), we can conclude that there is a correlation between the degree of relatedness among domain I segments and the specificity of the toxin proteins with which they are associated. These data may suggest that different characteristics of domain I are necessary to achieve successful domain I integration in the distinct target membranes. Lepidopteran and coleopteran insects have very different pH midgut conditions (alkali versus acidic) (12, 27), and there may also be differences in the protein and phospholipid compositions of the membranes, suggesting that special types of ion channels have been selected in the different targets. The single example of changes in specificity by mutations in domain I is the point mutation A92D of Cry1Ac toxin, which showed a selective loss of activity against different lepidopteran insects (52).

From the distribution of domain I sequences found in both Fitch and parsimony trees (Fig. 4), it is evident that lepidopteran membranes are sensitive to at least three types of domain I, while coleopteran insects are sensitive just to one type of domain I. It would be interesting to compare at the single-channel level the ionic and kinetic properties of the three different types of domain I.

Since the phylogenetic analyses suggest that Cry9Ca domain I (lepidopteran specific) has evolved from the same origin as

that of the coleopteran-specific domain I, it would be worth determining whether trypsin-activated Cry9Ca toxin has any activity against coleopteran species, as is the case of Cry1B and Cry1I toxins.

Domain II. The second domain of δ-endotoxins has been thought to be the receptor binding domain. Recently, it has been reported that the vitelline membrane outer layer protein I (VMO-I) and domain II from ICPs have similar three-dimensional structures. It has been proposed that the β-prism-fold may be a structural domain associated with carbohydrate binding functionality (43). Both proteins may have a carbohydrate binding site, since binding of Cry1Ac toxin to its glycoprotein receptor is inhibited by N-acetylgalactosamine (26) and VMO-I binds hexasaccharides of N-acetylglucosamine (24).

The estimation of phylogenetic relationships of domain II sequences suggests that this domain is probably derived from different evolutionary roots, because some sequences showed infinite distances and could not be grouped in the same tree. Those unrelated sequences have been grouped and analyzed independently. We have found that the domain II sequences can be distributed into three different trees. The first phylogenetic tree is composed of domains II from Cry2 and Cry11 toxins (Fig. 5A). The second tree grouped domains II from nematode-specific (Cry5, Cry12, and Cry13) and Cry14A toxins (Fig. 5B). Cry5Aa and Cry5Ab toxins are so homologous along the domain II sequence that they can be considered variants of the same molecule. The third tree is constituted of the rest of the domain II sequences (Fig. 5C). In this tree there are two main branches, one composed exclusively by domains II from Cry1 lepidopteran-specific toxins (excluding Cry1B, Cry1Ia, and Cry1Ib toxins). It is clear that some toxins have significant similarity in this region, suggesting that they have evolved from a common protein (examples are Cry1Ab and Cry1Ac, Cry1Aa and Cry1Ad, and both Cry1E toxins, both Cry1C toxins, and both Cry1D toxins). The second main branch is composed of four smaller branches. One small branch grouped the dipteran-specific toxins. This branch included the domain II from Cry9Aa toxin, which is a lepidopteran-specific toxin. The second small branch included domains II from Cry3 and Cry7 coleopteran-specific toxins. Again, there are pairs of domain II
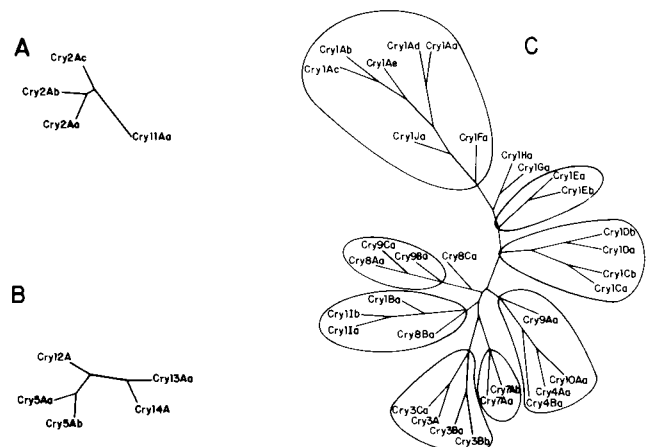

FIG. 5. Unrooted phylogenetic trees of the domain II sequences of the δ-endotoxin protein family. The phylogenetic analysis was performed with different groups of domain II sequences (A, B, and C) by Fitch and Margoliash's method (17) and by using the least-squares criterion and the FITCH program (15) as described in the legend to Fig. 3. The consensus phylogenetic trees obtained by the CONSENSE program (15) are presented. Circled branches are branches that were found in more than 90% of the trees.

FIG. 6. Unrooted phylogenetic trees of the domain III sequences of the δ-endotoxin protein family. Phylogenetic analysis was performed with different groups of domain III sequences (A and B) by Fitch and Margoliash's method (17) and by using the least-squares criterion and the FITCH program (15) as described in the legend to Fig. 2. The consensus phylogenetic trees obtained by the CONSENSE program (15) are presented. Circled branches are branches that were found in more than 90% of the trees.
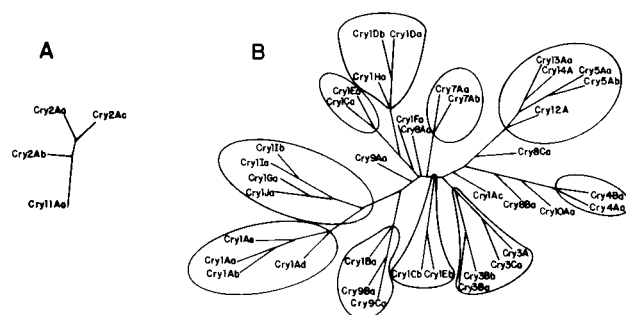
sequences which showed high similarity, like Cry3A and Cry3Ca, both Cry7 toxins, and both Cry3B toxins, implying that they have evolved from common ancestors. Domains II from the lepidopteran-coleopteran toxins (Cry1B and Cry1I toxins) and the coleopteran-specific Cry8Ba toxin are grouped in a single branch, suggesting a common origin of these domains. Finally, the fourth small branch is composed of domains II from the lepidopteran-specific Cry9Ba and Cry9Ca toxins together with the coleopteran-specific Cry8Aa and Cry8Ca toxins. It would be interesting to ask whether the proteins Cry9Ba and Cry9Ca that share similarity in domain II with other coleopteran-specific toxins (Cry8) have any activity against coleopteran insects. Also, the activity of Cry9Aa against dipteran insects could be tested.

The analysis of the topology of the third tree (Fig. 5C) which groups lepidopteran, coleopteran, and dipteran toxins, showed some correlation between the origin of domain II and specificity. However, this does not imply that domains II grouped in the same branch (probably same origin) will bind to the same type of protein receptors, since domains II from Cry1Ac and Cry1Ab are highly related and both toxins bind to different receptors (an aminopeptidase N versus a cadherin glycoprotein) (20, 25, 41, 47, 48). These data suggest that small differences within domain II contribute to the binding to different receptors or that some other parts of the protein besides domain II are involved in the recognition of the binding site. As receptors are known to be glycoproteins, a third possibility is that binding specificity involves interaction with a similar carbohydrate moiety on different receptor polypeptides.

Domain III. It has been proposed that domain III stabilizes the toxin by protecting against proteolysis (30) and that this domain may be implicated in receptor binding (1, 4, 29). During the sequence distance determination, we found that domains III from Cry2 and Cry11 toxins have infinite distances from the rest of the Cry family, suggesting a very different origin, and therefore the domain III sequences from those toxins were analyzed independently. Figure 6A shows the resulting phylogenetic tree with domain III sequences from Cry2 and Cry11 toxins, and Fig. 6B presents the results of analysis of the rest of the domain III sequences. The topology of the phylogenetic tree of Fig. 6B is very different from the topology of the trees obtained with the domain I or II sequences. The domain III sequences from the coleopteran-specific toxins are distributed along the tree in different branches, and only domains III from Cry3 toxins are arranged in a single branch.

These data suggest that many different types of domains III are compatible with coleopteran specificity. In contrast, domains III from all the nematode-specific toxins are arranged in a single branch; specifically, domains III from Cry5Aa and Cry5Ab toxins are very similar. The phylogenetic analysis of the three domains of this group of protoxins indicate that they have coevolved as a separate group that is relatively far from the rest of the Cry protein family. It will be worthwhile to determine if Cry14A toxin has any activity against nematodes, since this toxin has been described as toxic to *Diabrotica* sp. (38).

Regarding domain III sequences from the lepidopteran-specific toxins, most are closely arranged in the tree, suggesting that they have coevolved from a common ancestor. Domains III from Cry1A toxins are arranged in a single branch, with the exception of Cry1Ac that clearly evolved from a different origin. Domain III from the Cry1Ac toxin is not highly related to any other toxin. The Cry proteins with dual specificity (lepidopteran and coleopteran) with a domain I and domain II which share high similarity with the coleopteran-specific toxins have a domain III more related to the lepidopteran toxins than to the coleopteran toxins. These data may suggest that shuffling of domain III between lepidopteran and coleopteran toxins may be the origin of proteins with the capacity to affect both types of insects. However, the lepidopteran-specific Cry9Ca toxin also has a domain I and domain II that are more related to the coleopteran Cry8 toxins and a domain III that shares more similarity with the lepidopteran-specific toxins, implying that not all domain III shuffling between coleopteran and lepidopteran toxin could result in double specificity.

There are some clear examples of domain III shuffling among Cry1 toxins. Toxins Cry1Ca and Cry1Cb have domain I and II sequences that are so similar that they can be considered variants of the same protein. The same is true for the Cry1Ea and Cry1Eb toxins, but domains III from these four toxins have a different distribution in the tree. It is clear that domains III from Cry1Ca and Cry1Ea have a common origin, while domains III from Cry1Cb and Cry1Eb are variants of the same molecule. Both types of domain III (from subspecies a and b) group far away from the other. These data are in agreement with the proposition of Thompson et al. (46) that *cry1Cb* and *cry1Ea* genes could have arisen from ancestral crossovers between *cry1Eb* and *cry1Ca* genes.

Finally, the last example of a probable domain III shuffling is between Cry1Ga and Cry1Ha toxins. Both toxins have similar domain I and II sequences, but they have different domain III sequences; domain III from Cry1Ha clusters together with Cry1D toxins, while domain III from Cry1Ga resembles domain III from Cry1J toxin.

CONCLUSIONS

Several proteins are organized as discrete modules which may have different functions. It has been proposed that domain swapping may contribute to the versatility of protein function and therefore be an important molecular mechanism for their evolution (2, 3). The ICPs are a family of proteins that have biocidal activities against very different targets. These proteins are modular in structure, consisting of three different functional domains. In this work, the evolutionary relationships of the Cry protein family are presented. The phylogenies were estimated by two different methods (Fitch and Margoliash's method and maximum parsimony), and both types of analysis gave phylogenetic trees with similar topologies (data not shown), implying that there is a high probability that the phylogenetic trees presented here represent the correct topology

through which the Cry proteins evolved. Only the phylogenetic trees prepared by the method of Fitch and Margoliash are presented here.

The results of phylogenetic analysis of the whole Cry protein sequences do not reflect the complex evolutionary relationships found in the analysis of the independent functional domains. The results of phylogenetic analysis of domain I sequences suggest that domain I sequences have a common origin for the whole protein family (Fig. 4), while domain II and III sequences seem to be common only for a subgroup of proteins (Fig. 5 and 6). Unexpectedly, domain I, which is involved in the pore formation activity of the toxin, showed a topology clearly related to the specificity of the toxin proteins with which they are associated, suggesting that different types of domain I have been selected for acting in particular membrane conditions from the distinct target types (Fig. 4).

The data presented here suggest that there are three independent origins of domain II. The low degree of similarity among the three domain II groups (Fig. 5) could suggest that each type of domain II interacts with very different receptors, although there is no experimental evidence that support this hypothesis. The analysis showed that domains II from Cry1Ac and Cry1Ab toxins were derived from a common origin; nevertheless, both toxins bind to different receptors (an aminopeptidase N versus a cadherin glycoprotein) (20, 25, 41, 47, 48). It is proposed that both proteins interact with similar carbohydrate moieties on different receptor polypeptides or that small differences within domain II contribute in the binding to different receptors. Alternatively, additional regions of the protein besides domain II could be involved in the recognition of the binding site.

The conserved topology of domain I and II phylogenetic trees from some toxins suggest that these domains have co-evolved.

Finally, it is likely that the domain III sequences have evolved from two independent origins (Fig. 6). Shuffling of the functional domains was observed only for domains III of some toxins. Besides domain II, domain III is believed to be involved in receptor recognition. Toxins with dual specificity (lepidopteran and coleopteran) are examples of domain III shuffling among coleopteran- and lepidopteran-specific toxins. The phylogenetic relationships presented here suggest that the in vitro swapping of domains III from different toxins result in some cases in the production of novel chimeric toxins with altered specificity.

The phylogenetic analysis of the Cry toxin family shows that the great variability in the biocidal activity of this family has resulted from two fundamental evolutionary process: (i) independent evolution of the three functional domains and (ii) domain swapping among different toxins. These two processes have generated proteins with similar modes of action but with very different specificities. Also, this analysis suggests that some proteins (like Cry9A, Cry9B, Cry9C, and Cry14A) have additional biocidal activities against different insect orders. Finally, knowledge of the evolutionary relationships of the different Cry toxin domains may set the basis for a more rational and directed strategy to create novel chimeric toxins that have different specificities.

REFERENCES

1. **Aronson, A. I., D. Wu, and C. Zhang.** 1995. Mutagenesis of specificity and toxicity regions of a *Bacillus thuringiensis* protoxin gene. J. Bacteriol. **177:** 4059–4065.
2. **Baron, M., D. G. Norman, and L. D. Campbell.** 1991. Protein modules. Trends Biochem. Sci. **16:**13–17.
3. **Bennet, M. J., S. Choe, and D. Eisenberg.** 1994. Domain swapping: entangling alliances between proteins. Proc. Natl. Acad. Sci. USA **91:**3127–3131.
4. **Bosch, D., B. Schipper, H. van der Kleij, R. A. de Maagd, and J. Stiekema.** 1994. Recombinant *Bacillus thuringiensis* crystal proteins with new properties: possibilities for resistance management. Bio/Technology **12:**915–918.
5. **Bradley, D., M. A. Harkey, M. K. Kim, D. Biever, and L. S. Bauer.** 1995. The insecticidal CryIB protein of *Bacillus thuringiensis* has dual specificity to coleopteran and lepidopteran larvae. J. Invertebr. Pathol. **65:**162–173.
6. **Bravo, A., S. Jansens, and M. Peferoen.** 1992. Immunocytochemical localization of *Bacillus thuringiensis* insecticidal crystal proteins in intoxicated insects. J. Invertebr. Pathol. **60:**237–246.
7. **Crickmore, N., D. R. Zeigler, J. Feitelson, E. Shnepf, B. Lambert, D. Lereclus, C. Gawron-Burke, and D. H. Dean.** 1995. Revision of the nomenclature for *Bacillus thuringiensis cry* genes, p. 14. *In* Program and Abstracts of the 28th Annual Meeting of the Society for Invertebrate Pathology. Society for Invertebrate Pathology, Bethesda, Md.
8. **Cummings, C. E., G. Armstrong, T. C. Hodgman, and D. J. Ellar.** 1994. Structural and functional studies of a synthetic peptide mimicking a proposed membrane inserting region of a *Bacillus thuringiensis* delta-endotoxin. Mol. Membr. Biol. **11:**87–92.
9. **Chen, X. J., M. K. Lee, and D. H. Dean.** 1993. Site-directed mutations in a highly conserved region of *Bacillus thuringiensis* δ-endotoxin affect inhibition of short circuit current across *Bombyx mori* midguts. Proc. Natl. Acad. Sci. USA **90:**9041–9045.
10. **Choma, C. T., and H. Kaplan.** 1992. *Bacillus thuringiensis* crystal protein: effect of chemical modification of the cysteine and lysine residues. J. Invertebr. Pathol. **59:**75–80.
11. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
12. **Dow, J. A. T.** 1986. Insect midgut function. Adv. Insect Physiol. **19:**187–238.
13. **Feitelson, J. S.** 1993. The *Bacillus thuringiensis* family tree, p. 63–72. *In* L. Kim (ed.), Advanced engineered pesticides. Marcel Dekker, Inc., New York, N.Y.
14. **Feitelson, J. S., J. Payne, and L. Kim.** 1992. *Bacillus thuringiensis*: insects and beyond. Bio/Technology **10:**271–275.
15. **Felsenstein, J.** 1993. Phylip-Phylogeny interference package, version 3.5c (Distributed by J. Felsenstein, Department of Genetics, University of Washington, Seattle.)
16. **Feng, D. F., and R. F. Doolittle.** 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol. **25:**351–360.
17. **Fitch, W. M., and E. Margoliash.** 1967. Construction of phylogenetic trees. Science **155:**279–284.
18. **Gazit, E., D. Bach, I. D. K. M. S. P. Sansom, N. Chejanovsky, and Y. Shai.** 1994. The α-5 segment of *Bacillus thuringiensis* δ-endotoxin: *in vitro* activity, ion channel formation and molecular modeling. Biochem. J. **304:**895–902.
19. **Ge, A. Z., D. Rivers, R. Milne, and D. H. Dean.** 1991. Functional domains of *Bacillus thuringiensis* insecticidal crystal proteins. J. Biol. Chem. **266:**17954–17958.
20. **Gill, S. S., E. A. Cowlest, and V. Francis.** 1995. Identification, isolation and cloning of a *Bacillus thuringiensis* CryIAc toxin-binding protein from midgut of the lepidopteran insect *Heliothis virescens*. J. Biol. Chem. **270:**27277–27282.
21. **Grochulski, P., L. Masson, S. Borisova, M. Pusztai-Carey, J. L. Schwartz, R. Brousseau, and M. Cygler.** 1995. *Bacillus thuringiensis* CryIA(a) insecticidal toxin: crystal structure and channel formation. J. Mol. Biol. **254:**447–464.
22. **Hofmann, C., P. Lüthy, R. Hütter, and V. Pliska.** 1988. Binding of the δ-endotoxin from *Bacillus thuringiensis* to brush-border membrane vesicles of the cabbage butterfly (*Pieris brassicae*). Eur. J. Biochem. **173:**85–91.
23. **Höfte, H., and H. R. Whiteley.** 1989. Insecticidal crystal proteins of *Bacillus thuringiensis*. Microbiol. Rev. **53:**242–255.
24. **Kido, S., Y. Doi, F. Kim, E. Morishita, H. Narita, S. Kanaya, T. Ohkubo, K. Nishikawa, T. Yao, and T. Ooi.** 1995. Characterization of vitelline membrane outer layer protein I, VMO-I: amino acid sequence and structural stability. J. Biochem. (Tokyo) **117:**1183–1191.
25. **Knight, P., N. Crickmore, and D. Ellar.** 1994. The receptor for *Bacillus thuringiensis* CryIA(c) δ-endotoxin in the brush border membrane of the lepidopteran *Manduca sexta* is aminopeptidase N. Mol. Microbiol. **11:**429–436.
26. **Knowles, B. H., P. J. K. Knight, and D. J. Ellar.** 1991. N-acetyl galactosamine is part of the receptor in insect gut epithelia that recognizes an insecticidal protein from *Bacillus thuringiensis*. Proc. R. Soc. Lond. B **245:**31–35.
26a. **Knowles, B. H.** 1994. Mechanism of action of *Bacillus thuringiensis* insecticidal δ-endotoxins. Adv. Insect Physiol. **24:**275–308.

27. **Koller, C. N., L. S. Bauer, and R. M. Hollingworth.** 1992. Characterization of the pH-mediated solubility of *Bacillus thuringiensis* var *san diego* native δ-endotoxin crystals. Biochem. Biophys. Res. Commun. **184:**692–699.

28. **Lecadet, M. M., E. Frachon, V. C. Dumanoir, and H. de Barjac.** 1994. An update version of the *Bacillus thuringiensis* strains classification according to H-serotypes, p. 345. *In* Abstracts of the IInd International Conference on *Bacillus thuringiensis* 1994. Society for Invertebrate Pathology, Montpellier, France.

29. **Lee, M. K., B. A. Young, and D. H. Dean.** 1995. Domain III exchanges of *Bacillus thuringiensis* CryIA toxins affect binding to different gypsy moth midgut receptors. Biochem. Biophys. Res. Commun. **216:**306–312.

30. **Li, J., J. Carroll, and D. J. Ellar.** 1991. Crystal structure of insecticidal δ-endotoxin from *Bacillus thuringiensis* at 2.5 Å resolution. Nature **353:**815–821.

31. **Liang, Y., S. S. Patel, and D. H. Dean.** 1995. Irreversible binding kinetics of *Bacillus thuringiensis* CryIA δ-endotoxins to gypsy moth brush border membranes vesicles is directly correlated to toxicity. J. Biol. Chem. **270:**24719–24724.

32. **Lorence, A., A. Darszon, C. Díaz, A. Liévano, R. Quintero, and A. Bravo.** 1995. δ-Endotoxins induce cation channels in *Spodoptera frugiperda* brush border membranes in suspension and in planar lipid bilayers. FEBS Lett. **360:**217–222.

33. **Lu, H. L., F. Rajamohan, and D. H. Dean.** 1994. Identification of amino acid residues of *Bacillus thuringiensis* δ-endotoxin CryIAa associated with membrane binding and toxicity to *Bombyx mori*. J. Bacteriol. **176:**5554–5559.

34. **Morett, E., and L. Segovia.** 1993. The σ^{54} bacterial enhancer-binding protein family: mechanism of action and phylogenetic relationship of their functional domains. J. Bacteriol. **175:**6067–6074.

35. **Needleman, S. B., and C. D. Wunsch.** 1970. A general method applicable to the search for similarities in the amino acid sequences of two proteins. J. Mol. Biol. **48:**443–453.

36. **Nei, M.** 1987. Phylogenetic trees, p. 287–327. *In* M. Nei (ed.), Molecular evolutionary genetics. Columbia University Press, New York, N.Y.

37. **Nishimoto, T., H. Yoshisue, K. Ihara, H. Sakai, and T. Komano.** 1994. Functional analysis of block 5, one of the highly conserved amino acid sequences in the 130-kDa CryIVA protein produced by *Bacillus thuringiensis* subsp. *israelensis*. FEBS Lett. **348:**249–254.

38. **Payne, J. M., and K. E. Narva (Mycogene Corporation).** July 1994. Process for controlling corn rootworm larvae. Patent WO 94/16079.

39. **Rajamohan, F., J. A. Cotrill, F. Gould, and D. H. Dean.** 1996. Role of domain II, loop 2 residues of *Bacillus thuringiensis* CryIAb δ-endotoxin in reversible and irreversible binding to *Manduca sexta* and *Heliothis virescens*. J. Biol. Chem. **271:**2390–2396.

40. **Sacchi, V. F., P. Parenti, G. M. Hanozet, B. Giordana, P. Lüthy, and M. Wolfersberger.** 1986. *Bacillus thuringiensis* toxin inhibits K$^+$-gradient-dependent amino acid transport across the brush border membrane of *Pieris brassicae* midgut cells. FEBS Lett. **204:**213–218.

41. **Sangadala, S., F. W. Walters, L. H. English, and M. J. Adang.** 1994. A mixture of *Manduca sexta* aminopeptidase and phosphatase enhances *Bacillus thuringiensis* insecticidal CryIA(c) toxin binding and ^{86}Rb$^+$-K$^+$ efflux *in vitro*. J. Biol. Chem. **269:**10088–10092.

42. **Schnepf, H. E., K. Tomczak, J. P. Ortega, and H. R. Whiteley.** 1990. Specificity-determining regions of a lepidopteran-specific insecticidal protein produced by *Bacillus thuringiensis*. J. Biol. Chem. **265:**20923–20930.

43. **Shimizu, T., and K. Morikawa.** 1996. The β-prism: a new folding motif. Trends Biochem. Sci. **21:**3–6.

44. **Smith, G. P., and D. J. Ellar.** 1994. Mutagenesis of two surface-exposed loops of the *Bacillus thuringiensis* CryIC δ-endotoxin affects insecticidal specificity. Biochem. J. **302:**611–616.

45. **Tailor, R., J. Tippett, G. Gibb, S. Pells, D. Pike, L. Jordan, and S. Ely.** 1992. Identification and characterization of a novel *Bacillus thuringiensis* δ-endotoxin entomocidal to coleopteran and lepidopteran larvae. Mol. Microbiol. **6:**1211–1217.

46. **Thompson, M. A., H. E. Schnepf, and J. S. Feitelson.** 1995. Structure, function and engineering of *Bacillus thuringiensis* toxins. Genetic Eng. **17:**99–117.

47. **Vadlamudi, R. K., E. Weber, I. Ji, T. H. Ji, and L. A. Bulla.** 1995. Cloning and expression of a receptor for an insecticidal toxin of *Bacillus thuringiensis*. J. Biol. Chem. **270:**5490–5494.

48. **Valaitis, A. P., M. K. Lee, F. Rajamohan, and D. H. Dean.** 1995. Brush border membrane aminopeptidase-*N* in the midgut of the gypsy moth serves as the receptor for the CryIA(c) δ-endotoxin of *Bacillus thuringiensis*. Biochem. Mol. Biol. Int. **25:**1143–1151.

49. **VanRie, J., S. Jansens, H. Höfte, D. Degheele, and H. VanMellaert.** 1990. Receptors on the brush border membrane of the insect midgut as determinants of the specificity of *Bacillus thuringiensis* δ-endotoxins. Appl. Environ. Microbiol. **56:**1378–1385.

50. **Von-Tersch, M. A., S. L. Slatin, C. A. Kulesza, and L. H. English.** 1994. Membrane-permeabilizing activities of *Bacillus thuringiensis* coleopteran-active toxin CryIIIB2 and CryIIIB2 domain I peptide. Appl. Environ. Microbiol. **60:**3711–3717.

51. **Walters, F. S., S. L. Slating, C. A. Kulesza, and L. H. English.** 1993. Ion channel activity of N-terminal fragments from CryIA(c) δ-endotoxin. Biochem. Biophys. Res. Commun. **196:**921–926.

52. **Wu, D., and A. I. Aronson.** 1992. Localized mutagenesis defines regions of the *Bacillus thuringiensis* δ-endotoxin involved in toxicity and specificity. J. Biol. Chem. **267:**2311–2317.

53. **Wu, S. J., and D. H. Dean.** 1996. Functional significance of loops in the receptor binding domain of *Bacillus thuringiensis* CryIIIA δ-endotoxin. J. Mol. Biol. **255:**628–640.

54. **Yamamoto, T., and G. K. Powel.** 1993. *Bacillus thuringiensis* crystal proteins: recent advances in understanding its insecticidal activity, p. 3–42. *In* L. Kim (ed.), Advanced engineered pesticides, Marcel Dekker, Inc., New York, N.Y.