RESEARCH ARTICLE

ScienceAsia 36 (2010): **271–279**

doi: 10.2306/scienceasia1513-1874.2010.36.271

Analysis of the insecticidal crystal gene type 1 of *Bacillus thuringiensis* isolates affecting lepidopterans

Benjawan Lertwiriyawong[a], **Krit Pinthong**[a], **Jariya Chanpaisaeng**[b], **Panapa Saksoong**[a], **Pattana Srifah Huehne**[a,c,*]

[a] Department of Genetics, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand
[b] Department of Entomology, Faculty of Agriculture, Kasetsart University, Bangkok 10900, Thailand
[c] Laboratory of Biotechnology, Chulabhorn Research Institute, Bangkok 10210, Thailand

*Corresponding author, e-mail: fscipns@ku.ac.th

Received 2 May 2010
Accepted 30 Sep 2010

ABSTRACT: The novel δ-endotoxin from *Bacillus thuringiensis* (*Bt*) is one of the alternative measures for lepidopteran pest control. A survey of 31 *Bt* isolates, obtained from Thailand and previously screened for their toxicity to lepidopterans, was conducted to determine the presence of *cry1*-type genes, using polymerase chain reaction - restriction fragment length polymorphism analysis. Seven distinct types of *cry1* genes: *cry1Aa*, *cry1B*, *cry1C*, *cry1Cb*, *cry1D*, *cry1E*, and *cry1F*, were identified. The most common of the *cry1*-type genes was *cry1Cb*, followed by *cry1C* and *cry1D*, which covered 64.6, 48.4, and 25.8%, respectively. Besides *cry1C* and *cry1Cb*, two candidate *cry* genes, *cry1E* and *cry1D*, of isolate JC 190 (harbouring *cry1C/1E*), with cotton bollworm toxin specificity, showed 99% amino acid sequence identity to Cry1Ea of *B. thuringiensis* subsp. *kenyae*, while isolate JC 291 (containing *cry1C/1Cb/1D*), with Asian corn borer toxin specificity, harboured Cry1Dc, which exhibited only 84% amino acid sequence identity to Cry1Da of *B. thuringiensis* subsp. *aizawai*. The C-terminus of the JC291Cry1Dc possessed a unique sequence, 812-NVPIIPIISPW-822. Furthermore, the major differences between the 3D structure of Cry1Dc and Cry1Aa were confined to Domain II as three loop structures.

KEYWORDS: corn borer, *Bt*, Cry1D, delta endotoxin, PCR-RFLP

INTRODUCTION

Bacillus thuringiensis (*Bt*) is known to produce a variety of insecticidal crystal (*cry*) proteins, collectively referred to as δ-endotoxins, that for several decades have been used as biological pesticides against certain insect pests. To date, more than 495 *cry*-type genes have been identified with 59 holotype toxins, designated *cry1* to *cry*59. Of these, *cry1* is the most abundant and has diversified into more than 186 *cry*-type gene variants[1]. They are highly specific, in that they are active only against nematodes and a limited number of susceptible insects including lepidopterans, coleopterans, and dipterans[2,3]. A major group of Cry toxins is the three domain (3D)-Cry family, members of which share similarities in sequence and structure, but there are other groups of Cry proteins that are different from 3D-Cry[4,5]. The 3D-structure of the Cry1Aa protein (commonly used as a representative toxin produced by *Bacillus thuringiensis*) is an example of the three-domain family of Cry proteins[2]. Domain I is the pore-forming domain and Domains II and III are important for toxin-receptor interactions[4,6]. Two different hypotheses have been proposed for the mode of action of these toxins; one relies on pore formation and the other on signal transductions[3]. Proteolytic activation of Cry toxins takes place in the insect gut under high pH conditions (about pH 9.5), resulting in pathological effects that lead to larval death[4]. This highly specific toxin is considered environmentally friendly and is often used as an alternative approach for insect pest control[7]. Nevertheless, insects may develop resistance to the Cry toxin if only one type of Cry toxin is persistently used[8,9]. To overcome this problem, gene recombination and synergistic activity of different cry1-type toxins have been attempted to produce highly specific toxins that cause complete mortality of larval pests and thus prevent the development of resistance to the Cry proteins used in insect pest control[10,11].

In Thailand, many economically important crops, including corn, cotton, and tobacco, are destroyed by a variety of polyphagous lepidopteran pests. Major pests include the beet armyworm (*Spodoptera exigua* Hubner), the cotton bollworm (*Helicoverpa armigera* Hubner), the diamondback moth (*Plutella xylostella*), and the Asian corn borer (*Ostrinia furnacalis*). Pests are the cause of enormous economic loss to the Thai

agricultural industry. The use of δ-endotoxin is one attractive measure to solve this problem. However, development of effective crop treatment regimes is still in the preliminary stage[12]. In this report, various subtypes of the *cry1* gene in 31 *Bt* isolates known to be toxic to lepidopteran species were identified by PCR-RFLP analysis. The sequences and the predictable protein structures of two *cry1*-type genes cloned from efficient toxin producers were determined.

MATERIALS AND METHODS

Bacterial strains and insect toxicity bioassays

More than 260 bacterial samples were locally collected and single colonies isolated from soils in 18 Parks and National Parks all over Thailand. After cell culturing for 48 and 72 h, each isolate was characterized by staining for spore and crystal protein production with malachite green and safranin, respectively. The positive isolates were cultured in semi-solid nutrient and examined for H-serotyping with agglutinin titre 1:3200–1:25600 according to Ohba and Aizawa[13] and de Barjac and Frachon[14]. Then, 31 isolates were tested for their effectiveness as pest killers. Different logarithmic doses of sporulated *Bt* suspension (10^3, 10^5, 10^7, and 10^9 spores/ml), from each isolate were spread over the insect diet. Twenty of the second instars larvae of each of the lepidopteran pests, *H. armigera*, *O. furnacalis*, *P. xylostella*, and *S. exigua*, were incubated with each dose. Insect mortalities were scored at 6, 12, 24, 48, and 72 h post-inoculation. LC_{50} was calculated by probit analysis[15].

DNA isolation

Thirty-one *Bt* isolates that had been tested for endotoxin toxicity were used in *cry1* gene typing studies. DNA extraction was performed according to the method described in Ref. 16. Cells were washed once with GTE (50 mM glucose, 25 mM Tris-HCl (pH 8.0), 10 mM EDTA). The pellets were lysed with 500 µl of lysis buffer (GTE containing 1 mg/ml lysozyme), incubated at 37 °C for 15 min, then 60 µl of 10% SDS was added and incubated at 50 °C for 15 min. The DNA in each sample was separated on 0.8% agarose gel.

PCR-RFLP analysis

For the *cry1* gene typing of individual isolates, the DNA sequence of the *cry1* gene was partially amplified using two pairs of universal oligonucleotide primers (K5un3 and K3un3, and K5un2 and K3un2) designed by Kuo and Chak[17]. The first pair amplifies the 5′ end of the gene spanning 1.4 kb, while the second pair yields a 1.6 kb PCR product at the 3′ end of the *cry1* gene using *i-Taq* DNA Polymerase (iNtRON Biotechnology Inc., Korea). Amplification was accomplished using the following conditions: denaturizing for 4 min at 94 °C followed by 35 cycles of amplification with 1 min denaturizing at 94 °C, 1 min of annealing at 60 °C, 2 min of extension at 72 °C, and a final extra extension step of 15 min at 72 °C at cycle completion. The products were size-analysed by 1% agarose gel electrophoresis with a 1 kb DNA marker ladder. The RFLP was performed and analysed according to the procedure described in Ref. 17. Individual derived DNA profiles were then assigned to the known gene types.

Full-length gene cloning and sequence analysis

Isolates JC 190 and 291 were arbitrarily chosen for DNA sequencing of the full-length *cry1* gene. Partially amplified fragments (1.4 kb at the 5′ and 1.6 kb of the 3′ end fragments) of the *cry1*-type genes were cloned into pGEM-T easy vector (Promega, USA), walking sequenced, and BLAST searched against the GenBank database to identify the *cry1* gene type. Amino acid sequence alignment of all *cry1* gene types was carried out using CLUSTALW[18]. Phylogenetic trees were inferred via neighbour-joining analysis using MEGA Version 3.1[19] based on a Dayhoff matrix[20] with 1000 bootstrap replicates.

Protein structure homology modelling

Identification of regions of homology between the amino acid sequences of JC 291 Cry1 and JC 190 Cry1 with Cry1Aa (PDB entries 1CIY) involved using the structural alignment tool of the SWISS-MODEL database[21]. The percentage sequence identity and three-dimensional structures were determined using CPHMODELS-2.0[22] employing the most homologous amino acid sequence resulting from the 1CIY sequences as standards. Figure construction and Ca RMSD calculations were performed with SWISSPDB VIEWER 4.01[23].

RESULTS

Bt serotyping and toxicity

Ninety-four out of 260 collected samples were selected as *Bt* positive isolates for spore and crystal protein production and further examined for H-serotyping. Thirty-one of the *Bt* positive clones comprised five known subspecies: *Bt. kurstaki* (3abc), *Bt. kenyae* (4ac), *Bt. galleriae* (5ab), *Bt. neoleonensi* (24), and *Bt. thailandensis* (68), and three unknown subspecies, based on H-serotype bioassay typing

(Table 1). A wide variation in virulence was expressed by 31 isolates, with 17 out of 31 virulent isolates being the most potent and having a killing dose LC_{50} value ranging from 10^3–10^7 spores/ml. Isolates JC 81, JC 82, JC 189, JC 190, JC 389 affected *H. armigera*, isolates JC 150, JC 173, JC 316, JC 384 affected *P. xylostella*, isolates JC 81, JC 82, JC 189, JC 353, JC 398, JC 595 affected *S. exigua*, and isolate JC 291 affected *O. furnacalis*. The *cry1*-type gene profiles in Table 1 show clearly that isolates containing many inconsistent variations of *cry1*-type genes feature in the group with high toxicity activity.

Cry1 gene type determination

Analysis of the *cry1* PCR-RFLP patterns of the 31 *Bt* isolates, generated using two universal primer pairs designed to amplify the 1.6 kb 3′ end fragments (Fig. 1a), and the 1.4 kb 3′ ends of *cry1* (Fig. 1c), revealed the presence of seven different *cry1*-type gene subfamilies (*cry1Aa*, *cry1B*, *cry1C*, *cry1Cb*, *cry1D*, *cry1E*, and *cry1F*) that were identical to those previously identified in Ref. 17, as shown in Fig. 1b and 1e for the 3′ end RFLP pattern, and in Fig. 1d and 1f for the 5′ end RFLP pattern. Of the 31 *Bt* isolates, *cry1Cb*, followed by *cry1C*, had the highest two frequencies with values of 64.6 and 48.4%, respectively. Two genes, *cry1D* and *cry1F*, were detected with an almost similar incidence of around 22–25%, while *cry1B* at 9.6% had the lowest rate of discovery (Fig. 2a) in the *Bt* isolates. The restriction mapping profiles of the *cry1*-type genes among the 31 lepidopteran-effective killers fell into eleven groups (Table 1) according to their RFLP-PCR patterns of the 5′ end and the 3′ end of *cry1* genes. Some of the 31 *Bt* isolates demonstrated toxicity towards one or two of the species: *H. armigera*, *O. furnacalis*, *P. xylostella*, and *S. exigua* in insect bioassays, when their LC_{50} was determined by probit analysis. The restriction mapping profiles also indicated that most individual isolates contained two *cry1* gene types. The combined results of 5′ and 3′ PCR-RFLP analysis (Fig. 2b, Table 1) revealed that 15 of the isolates contained the two-gene combination of *cry1Cb*/*cry1D* or *cry1Cb*/*cry1F*. Eight isolates contained two-gene combinations, consisting of either *cry1Aa*/*cry1C* (JC 81) or *cry1C*/*cry1E* (JC 189 and JC 190) or *cry1Cb*/*cry1D* (JC 82, JC 150, JC 173, and JC 316) or *cry1Cb*/*cry1F* (JC 384) and five isolates of a three-gene combination of *cry1C*/*cry1Cb*/*cry1D* (JC 291 and JC 389) and *cry1C*/*cry1Cb*/*cry1E* (JC 353, JC 590 and JC 595) correlated with high lepidopteran toxicity LC_{50} at 10^3–10^7 spores/ml (Table 1). However, the PCR-RFLP pattern of *cry1* genes from the JC 190 and

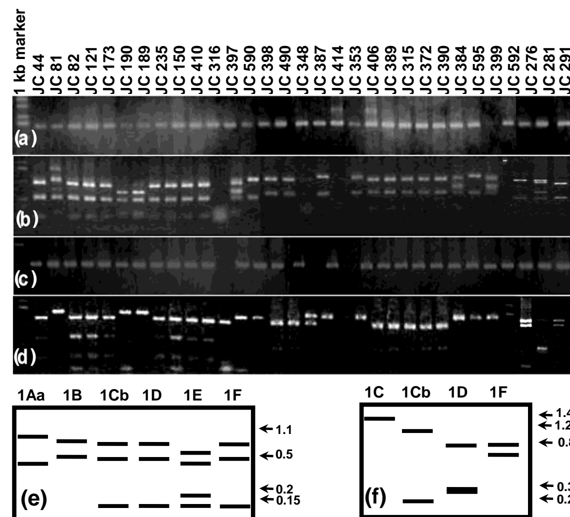

Fig. 1 PCR-RFLP pattern identifying the *cry*1 genes of 31 *Bt* isolates: (a) 1.6 kb 3′ *cry*1 PCR products; (b) after digestion with the *Pst*I and *Xba*I generated 1.1, 1, 0.95, 0.75, 0.6, 0.58,0.5, 0.2, and 0.14 kb fragments; (c) 1.4 kb 5′ *cry1* PCR products; (d) after digestion with the *Pst*I and *Eco*RI generated 1.4, 1.2, 0.8, 0.6, 0.3, and 0.2 kb fragments; (e) Kuo and Chak [13] pattern model of 3′ *cry1* PCR-RFLP; (f) the 5′ PCR-RFLP model illustrates the predicted sizes of *cry*-type genes, *cry1Aa*, *1B*, *1Cb*, *1D*, *1E*, and *1F* for 3′ PCR-RFLP patterns and *cry1C*, *1Cb*, *1D*, and *1F* for 5′ patterns.

JC 291 isolates that killed *H. armigera* and *O. furnacalis*, respectively, were the most distinctive among the 31 *Bt* isolates. Therefore, full-length PCR-fragments containing the *cry1*-type genes of *Bt* JC 190 and *Bt* JC 291 were cloned for further study.

Full-length cry1 sequence analysis

The DNA sequence of JC 291 *cry1* was 3519 bp in length and encoded a protein of 1173 amino acids. Comparison of the JC 291 *cry1*-type clone deduced amino acid sequence with 13 other Cry1 proteins available in the GenBank database indicated that the sequence was new with 84% identity to Cry1Da (accession no. X54610) and 75% identity to Cry1Aa and Cry1Ha (Fig. 3, Table 2). This new nucleotide sequence, denoted Cry1Dc (www.lifesci.sussex.ac.uk/home/Neil_Crickmore/Bt/), the insecticide affecting *O. furnacalis* isolated from the JC 291 isolate (carrying *cry1C/1Cb/1D*), was deposited in the GenBank database under accession no. EF059913.

Comparisons of the Cry1Dc insecticidal toxic Domain I region of Cry1Dc with those of Cry1Da and Cry1Db (accession no. Z22511) revealed 71%

Table 1 The RFLP-PCR *cry1* profiles, H-serotyping, and toxicity bioassay of 31 *Bt* isolates affecting four lepidopterans.

Cry1 profile		5′ end	3′ end	[a]Subspecies of *Bt*	[b]#JC isolate affecting			
					HA	SE	OF	PX
1.	*cry1Aa*, *1B*, *1C*	*cry1C*	*cry1Aa*, *1B*	unidentified subspecies: JC 281	-	-	-	-
2.	*cry1Aa*, *1C*	*cry1C*	*cry1Aa*	*kurstaki*: JC 372, JC 592	81	81	-	-
				kenyae: JC 81				
3.	*cry1Aa*, *1Cb*, *1C*	*cry1C*, *1Cb*	*cry1Aa*	unidentified subspecies: JC 276	-	-	-	-
4.	*cry1B*, *1C*	*cry1C*	*cry1B*	*kurstaki*: JC 390, JC 398	-	398	-	-
5.	*cry1C*, *1E*	*cry1C*	*cry1E*	*kenyae*: JC 189, JC 190	189,190	189	-	-
6.	*cry1C*, *1Cb*, *1E*	*cry1C*	*cry1E*,	*thailandensis*: JC 353	-	353	-	-
			cry1Cb/*1D*/*1F*	*kurstaki*:JC 590, JC 595	-	595	-	590
7.	*cry1C*, *1Cb*, *1D*	*cry1C*, *1Cb*	*cry1D*	*galleriae*: JC 291	389	-	291	-
				kurstaki: JC 389				
8.	*cry1Cb*, *1C*	*cry1C*, *1Cb*	-	*kurstaki*: JC 397	-	-	-	-
9.	*cry1Cb*, *1D*	*cry1Cb*, *1D*	*cry1Cb*/*1D*/*1F*	*galleriae*: JC 44, JC 121,	82	82	-	150,173
				JC 150, JC 173, JC 235				
				kurstaki: JC 82, JC 315, JC 316	316			
10.	*cry1Cb*, *1F*	*cry1Cb*, *1F*	*cry1Cb*	*kurstaki*: JC 384, JC 399,	-	-	-	384
				JC 406, JC 410, JC 414, JC 490				
				neoleonensis: JC 387				
11.	*cry1Cb*	*cry1Cb*	-	*kurstaki*: JC 348	-	-	-	-

[a] *cry1* gene profiles were determined by RFLP-PCR patterns.
[b] Subspecies were determined by H-serotyping: *Bt. kurstaki* (3abc); *Bt. kenyae* (4ac); *Bt. galleriae* (5ab); *Bt. neoleonensi* (24); *Bt. thailandensis* (68).
[c] LC_{50} at 10^3–10^7 spores/ml was determined by probit analysis.
HA = *Helicoverpa armigera* Hubner; SE = *Spodoptera exigua* Hubner ; OF = *Ostrinia furnacalis*; PX = *Plutella xylostella*.

Table 2 Comparison of amino acid sequence identity of Cry1Dc and Cry1Ea7 with published Cry1.

Cry toxin	GenBank Accession #	aa size	% aa sequence identity		Endotoxin Region		
			Cry1Ea7	Cry1Dc	Domain I	Domain II	Domain III
Cry1Aa	M11250	1176	72	75	36–254	259–460	470–607
Cry1Ba	X06711	1228	58	55	51–273	278–489	499–636
Cry1Ca	X07518	1189	74	69	35–253	261–457	467–616
Cry1Da	X54160	1165	72	84	35–253	258–450	460–592
Cry1Db	Z22511	1160	70	78	35–253	258–450	460–592
Cry1Dc	EF059913	1173	69	100	38–256	261–453	463–600
Cry1Ea	X53985	1171	99	69	36–253	258–453	468–601
Cry1Ea7	AY894137	1171	100	69	36–253	258–453	468–601
Cry1Fa	M63897	1174	69	70	34–252	257–454	464–601
Cry1Ga	Z22510	1166	69	71	35–248	253–446	456–593
Cry1Ha	Z22513	1172	68	75	38–252	257–455	465–599
Cry1Ia	X62821	719	36	40	60–280	287–497	507–644
Cry1Ja	L32019	1167	64	66	51–253	258–447	464–595
Cry1Ka	U28801	1215	56	55	56–279	284–490	500–637
Cry1La	AY554171	1170	64	66	38–251	256–452	462–598
Cry1Ma	Y09326	1173	66	68	35–249	254–447	457–595

and 63% identity, respectively. The sequence identity within Domain II was the most conserved at 96% and 91%, respectively, while Domain III shared the lowest sequence identity, namely 52% identity with Cry1Da and 51% identity with Cry1Db. The amino acid substitutions acquired during the evolution of the Cry1D gene types were not evenly spread over the length of the molecule. Comparison of the amino acid

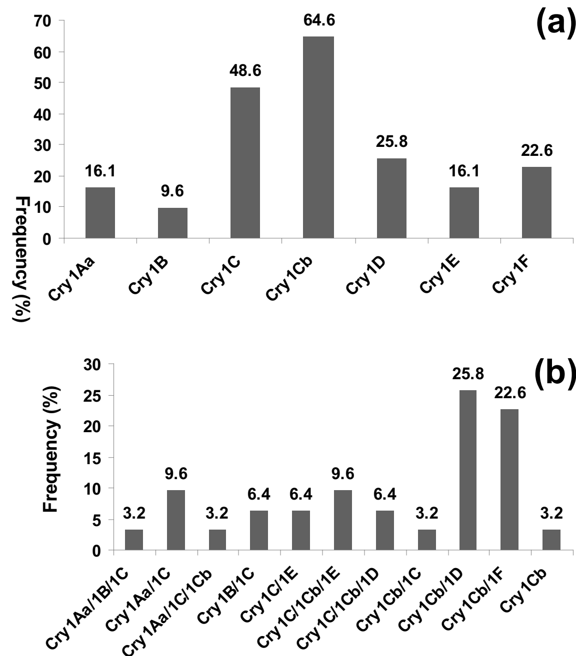

Fig. 2 (a) Frequency of individual *cry*1 type genes of 31 *Bt* isolates based on the PCR-RFLP patterns; (b) their combination profiles.

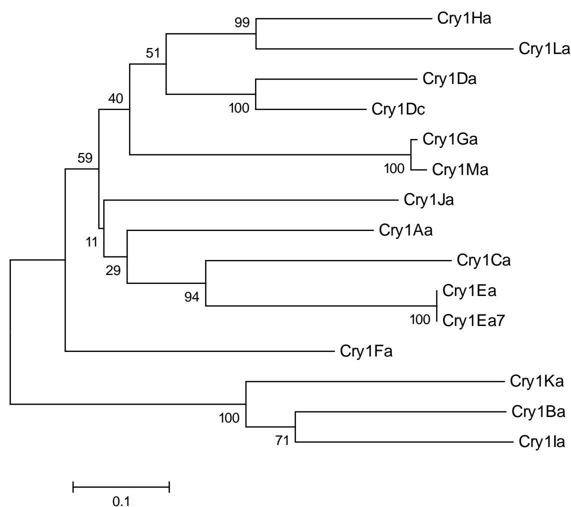

Fig. 3 Phylogenetic tree of Cry1A-1M on the basis of full-length amino acid sequences. The relatedness cladogram, inferred from the known 13 Cry1 proteins in the Genbank database and the JC291Cry1Dc and JC190Cry1Ea7, was constructed with the neighbour-joining method, using MEGA 3.1, based on a Dayhoff matrix. The numbers at the nodes represent the percentage of 1000 bootstrap resamplings. The scale on the bottom indicates the number of substitutions per amino acid site.

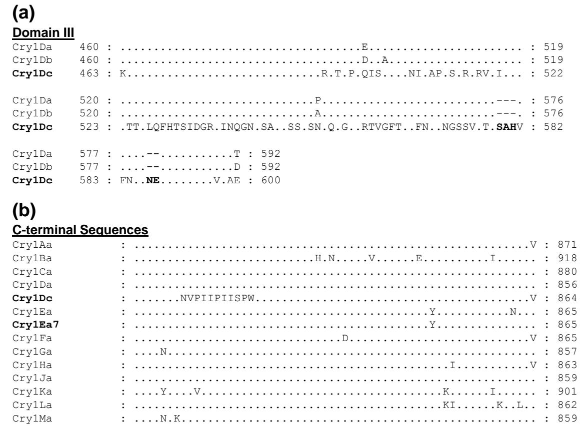

Fig. 4 Multiple alignment of cry1D domains and C terminal of cry1A–M. (a) Extra and different amino acid compositions presented on three domains of Cry1Da, Cry1Db, and JC291Cry1Dc; (b) different amino acid residues occurring on the C terminal of the Cry1Dc protein as compared with the other Cry1 sequences available in GenBank. The extra residues are indicated by bold letters; conserved residues are indicated by dots (.) and the gaps are indicated by dashes (-).

sequences of Cry1Dc, Cry1Db, and Cry1Dc revealed that 85 of the 219 amino acid residues of Domain I, or 38.8%, had undergone substitution. In contrast, the substitution frequency was only 8.3% (16/193) in Domain II and rose to 46.6% (62/133) in Domain III, excluding a 5-amino-acid insertion. Sporadic consecutive amino acid substitutions were observed in all three domains, with the longest stretch of 10 amino acids, 527-LQFHTSIDGR-536, in Domain III. In addition, two apparent amino acid insertion sites were observed, one with an extra 3 amino acids at positions 579-SAH-581 and another, with two extra residues, at positions 587-NE-588 (Fig. 4a). When the C-terminal sequence of the protein was aligned with 12 other Cry1 sequences (A–M except I), a high level of sequence conservation was noted, except for an obvious stretch of amino acid substitutions at positions 812-NVPIIPIISPW-822 (Fig. 4b). However, the *cry1E*-type gene, isolated from JC 190 (carrying *cry1C/1E*) affecting *H. armigera*, was identical in amino acid sequence to Cry1Ea1 (Table 2 and Fig. 3) and its nucleotide sequence was 99% identical. This new Cry1Ea isolate was designated Cry1Ea7 and deposited under accession no. AY894137 in the database.

Active toxin homology modelling

The Cry1Dc amino acid sequence shared the highest level of sequence identity (59%) with Cry1Aa (1CIY), while Cry1Ea7 was slightly less (54%). Cry1Dc and

```
                         Block 2
Cry1Dc    ISVLDIVAFFPNYDIRTYPIHTATQLTREIYLDLPFINENLSPAASYPTFSAAESAIIRS   299
Cry1Aa     LTVLDIVALFSNYDSRRYPIRTVSQLTREIYTNP--VLENFD-GSFRGMAQRIEQNIRQ   293
Cry1Ea7   ISVLDIISFFRNYDSRLYPIPTSSQLTREVYTDP---VINITDYRVGPSFENIENSAIRS   293
          ::****:::* *** * *** * :*****:* :       *:        :    .  **.
           Endotoxin N                  Endotoxin M
Cry1Dc    PHLVDFLNSFTIYTDSLARYAYWGGHLVNCFRTGTTTNLIRSPLYGREGNTERPVTITAS   359
Cry1Aa    PHLMDILNSITIYTDVHRGFNYWSGHQITASPVGFSGPEFAFPLFGNAGNAAPPVLVSLT   353
Cry1Ea7   PHLMDFLNNLTIDTDLIRGVHYWAGHRVTSHFTGSS-QVITTPQYGITANAEPRRTIAPS   352
          ***:*:**.:** **      **.** :..  .* :   :  * :*  .*:      :: :
Cry1Dc    PSVPI---FRTLS------YITGLD-NSNPVAGIEGVEF-----QNTISRSIYRKSGPID   404
Cry1Aa    G-LGI---FRTLSSPLYRRIILGSGPNNQELFVLDGTEFSFASLTTNLPSTIYRQRGTVD   409
Cry1Ea7   TFPGLNLFYRTLSN-----PFFRRSENITPTLGINVVQG-VGFIQPNNAEVLYRSRGTVD   406
          :   :****         :    . *       :: .:           . .  :**. *.:*
Cry1Dc    SFNELPPQDASVSPSIGYSHRLCHATFLERISGPR--IAGVVFPWTHRSASPTNEVSSSR   462
Cry1Aa    SLDVIPPQDNSVPPRAGFSHRLSHVTMLSQAAGAVYTLRAPTFSWQHRSAEFNNIIPSSQ   470
Cry1Ea7   SLNELPIDGE--NSLVGYSHRLSHVTLTRSLYNTN-ITSLPTFVWTHHSATNTNTINPDI   463
          *:: :* :.     .  *:****.*.*:      ..        .* * *:**  .* : ..
                                                        Block 3
```

Fig. 5 Amino acid sequence alignment of the Domain II region of the Cry1Aa, Cry1Dc and Cry1Ea7. * = identical residues represented; : = strongly homologous residues ; . = weakly homologous residues; and - = no consensus residue. The major differences in conferred loop structures are underlined and conserved sequence are boxed. The amino acid residues in β-sheet (S) are in bold. S1=IRTVS, S2=EIYTN, S3=GSFRG, S4=QRIEQN, S5=MDILN, S6=TIYTD, S7=GFNYWSGHQITASPV, S8=EFAFP, S9=VLVSL, S10=GLGIF, S11=TLSSPLYRRIILG, S12=ELFVL, S13=STIYR, S14=TVDSL, S15=SHRLSHV, S16=MLSQAAGA, S17=TLRAPTFSWQHR.

Fig. 6 Superimposition of the 3D structure of Cry1Dc or Cry1Ea7 with Cry1Aa (1CIY). (a) and (b): The schematic ribbon diagram using superimposition of the α C carbon traces of the Cry1Dc, Cry1Ea7, and Cry1Aa. Three major differences between the toxin proteins are confined to Domain II of the β-sheet structure, as indicated by arrows.

Cry1Ea7 were co-modelled with Cry1Aa and their 3D structures are displayed as superimposed backbone traces of 2.17 Å RMSD for Cα in Fig. 3a and 3b. In each case, the general structural features of the three domains of the Cry-type toxins were conserved. The Cry1Dc Domain I, residues 38–256, form seven highly conserved α helices, while Domain II, residues 261–453, comprise three antiparallel β-sheets joined to form a Greek key topology[6]. The major differences between the two modelled structures centre on three major loops corresponding to regions containing a large number of amino acid sequence differences located in the endotoxin M region between conserved amino acid block 2 and block 3 (Fig. 5). Domain III is composed of two antiparallel β-sheets forming a β-sandwich spanning residues 463–600. As shown in Fig. 6b, five extra amino acids in Cry1Dc seemed to cause no apparent effects on the 3D model. Nevertheless, diversity prevailed when comparisons were made among the deduced model structures of Cry1Dc and Cry1Ea7, using Cry1Aa as a standard. An extended loop was noted at the insertion/deletion sites in Domain III (Fig. 6a). For the amino acid substitution along the domains, no apparent change of the general feature was observed, except for several minor deviations, especially at the point of ten consecutive amino acid substitutions (Fig. 5).

For the Cry1Ea7 structural analysis, the three domains ranged from residue 36–601 (Fig. 6b). The superimposed backbone traces of Cry1Ea7 and Cry1Aa (1CIY) displayed 1.53 Å RMSD for Cα. The major differences between the two proteins were also noted by the presence of three loops in Domain II that spanned residues 258–453 and formed three antiparallel β-sheets. On the other hand, Domain I was highly conserved and consisted of residues 36–253 that formed seven α helices, while Domain III formed two antiparallel β-sheets between residues 453 and 601.

DISCUSSION

The present study reports that the RFLP-PCR technique according to Kuo and Chak[17] is a simple rapid method not only for Bt*Cry1*gene identification but also for full-length gene cloning, unlike denaturing gradient gel electrophoresis[24] (DGGE) and multiplex-PCR[25] techniques which are used for distinguishing between *cry1* and the other particular genes in large-scale screening. The RFLP-PCR profile patterns of 3′ and 5′ end amplified *cry1* of 31 *Bt* isolates affecting

lepidopteran insects provided sufficient information to classify *Cry1* genes into seven subfamilies as shown in Table 1.

The prevalence of *cry1* gene type 1C and 1D found in the present study was not surprising, since they have been noted earlier in Asian *Bt* collections[26–30]. Similarly, the low occurrence of *cry1B* (one isolate) among the 31 isolates was consistent with similar findings in Korean and Iranian collections[27,30]. The results reported here differ from the results of a recent survey of *Bt* isolates obtained in Thailand by Thammasittirong and Attathom[12] in which *cry1A*, found at a relatively low frequency in the current work, was the most prevalent type. Moreover, *cry1F* was not found in the previous study[12], but occurred at a relatively high frequency (22.6%) in the present study. Such variability in *cry1* gene types and their frequencies is thought to be due to differences in the geography and ecology of the collection sites and/or artefacts arising from the methods used for gene type identification[12,27–29].

The presence of more than one *cry*-type gene in individual isolates was not surprising since, in addition to their own chromosomal *cry* gene, strains can carry plasmids that encode additional *cry* genes[31,32]. Moreover, isolates can harbour more than one type of plasmid. Thus it was not surprising that at least two cry*1* gene types were observed in the present work in almost all isolates, with the highest number being three.

Although possessing several *cry*-type genes was not considered to directly reflect the toxic level of *Bt* strains[28,30], harbouring different *cry* gene types can give the isolates a broader spectrum of insect specificity. The toxicity level depends more on the level of gene expression and the synergism between different *cry*-type genes[33,34], as well as on the copy number of plasmid templates for toxin production. For these reasons, the current isolates, which had previously been screened for their toxicity to some important lepidopteran pests, produced different levels of toxicity, despite having the same *cry1* gene types. Although the *cry1*-type gene features had inconsistent variations, the concurrence of Cry1C or Cry1Cb (100%) was found in all *Bt* isolates JC 81, JC 82, JC 189, JC 190, JC 389 that were highly toxic to *H. armigera*. Isolates JC 150, JC 173, JC 316, JC 384 affected *P. xylostella*, isolates JC 81, JC 82, JC 189, JC 353, JC 398, JC 595 affected *S. exigua*, and isolate JC 291 (carrying *cry1C/1Cb/1D*) affected *O. furnacalis* (Table 1). There were inconsistent variations of the *cry1*-type genes feature. This suggests some genetic linkage between *cry1Cb* and *1D* (Fig. 2b) as has been previously reported[30,35].

The JC 291 isolate showed high toxicity to *O. furnacalis*, but differed in its *cry1* PCR-RFLP pattern in the collection from the current study, as it contained three different *cry1* type genes: *cry1C*, *cry1Cb* and *cry1D*. The 84% and 78% homology of the deduced amino sequence of the Cry1Dc to that of Cry1Da and Cry1Db, respectively, makes it a novel subtype. The high degree of amino acid substitutions in Domain I (38.8%) was surprising since it had previously been regarded as the most conserved of the three domains[2,3,6]. Domain I is responsible for ion channel formation[36] at the target site, so any alteration of amino acid types might be expected to affect its function, especially when the changes are open in a consecutive order, like those occurring at locations 45–65, and 106–141. However, toxicity tests confirmed the retention of toxicity (Table 1).

Domain II of Cry1Dc, the most variable domain according to previous reports[37], was found to be relatively conserved (8.3% amino acid substitutions) between the Cry1D-type protein examined in the present study. The observed conservation of Domain II between the Cry1D-type proteins emphasizes its importance in specific receptor binding[38,39]. In contrast, the high degree of amino acid sequence divergence (46.6%) in Domain III at locations 493–519 and 527–577 of the sequence residues would again cause some changes of the domain 3D structure, as well as its receptor binding function and the potency of toxicity[40]. Nonetheless, five extra amino acids in Domain III may cause just two small loops, which may not harm its activities, unless the locations of insertion are very crucial for the domain function.

Unfortunately, the only available 3D structure of a Cry1-type protein is that of Cry1Aa[3,6]. Therefore, any comparison of the Cry1Dc 3D structure has been carried out through the Cry1Aa model. Apparently, the 5-amino-acid insertion in Domain III of Cry1Dc is, in fact, the original amino acid sequence, as it superimposed well with that of the Cry1Aa, but formed a larger loop when compared with the modelled structure of Cry1Da and Cry1Db (Fig. 4a). It is obvious that deletion of these 5-amino-acids in Cry1Da and Cry1Db has occurred in the course of their evolution. It remains to be seen if these extra 5 amino acids, as well as other amino acid substitutions, alter the domain function.

In conclusion, the PCR-RFLP method was effective for preliminary screening for the *Bt cry1* gene types, but it requires confirmation by DNA sequencing, as illustrated in the present study by the two *cry1* type genes, *cry1Ea7* and *cry1Dc*, cloned from

isolates *Bt*JC190 (harbouring *cry1C/1E*) and *Bt*JC291 (harbouring *cry1C/1Cb/1D*), respectively. The present work identified a novel tertiary rank Cry1Dc and reported its DNA sequence, and deduced the amino acid sequence and the most likely 3D structure. This finding provides the 3D-structure of the receptor binding Domain II and Domain III in insecticidal active domains of the Cry1Dc (JC 291) and Cry1Ea7 (JC 190) toxin. The fact that these domain structures may correlate with the toxin-receptors of *H. armigera* and *O. furnacalis* is potentially of importance in the control of crop pests affecting cotton and corn.

Acknowledgements: This study was supported by the Golden Jubilee Program, Thailand Research Fund, and Kasetsart University Research and Development Institute, Thailand. The authors would like to thank Dr Anchanee Kubera for 3D structure analysis and Dr James M. Dubbs for manuscript editing.

REFERENCES

1. Crickmore N, Zeigler DR, Schnepf E, Van Rie J, Lereclus D, Baum J, Bravo A, Dean DH (2009) *Bacillus thuringiensis* toxin nomenclature. www.lifesci.sussex.ac.uk/Home/Neil_Crickmore/Bt
2. Crickmore N, Zeigler DR, Feiteison J, Schnepf E, Van Rie J, Lereclus D, Baum J, Dean DH (1998) Revision of the nomenclature for the *Bacillus thuringiensis* pesticidal crystal proteins. *Microbiol Mol Biol Rev* **62**, 807–13.
3. Bravo A, Gill SS, Soberón M (2007) Mode of action of *Bacillus thuringiensis* Cry and Cyt toxins and their potential for insect control. *Toxicon* **49**, 423–35.
4. Schnepf E, Crickmore N, Van Rie J, Lereclus D, Buam J, Feitelson J, Zeigler DR, Dean DH (1998) *Bacillus thuringiensis* and Its Pesticidal Crystal proteins. *Microbiol Mol Biol Rev* **62**, 775–806.
5. Bravo A, Soberón M (2008) How to cope with insect resistance to Bt toxins? *Trends Biotechnol* **26**, 573–9.
6. Li J, Carroll J, Ellar DJ (1991) Crystal structure of insecticidal δ-endotoxin from *Bacillus thuringiensis* at 2.5 Å resolution. *Nature* **353**, 815–21.
7. Siegel JP (2001) The mammalian safety of *Bacillus thuringiensis* based insecticides. *J Invertebr Pathol* **77**, 13–21.
8. Ferre J, McGaughey WH, Johnson DE, Barnet BD, Van Mellaert H (1991) Resistance to the *Bacillus thuringiensis* bioinsecticide in a field population of *Plutella xylostella* is due to a change in mid gut membrane receptor. *Proc Natl Acad Sci USA* **88**, 5119–23.
9. Jansens S, Cornelissen M, de Clercg R, Reynaerts A, Peferoen M (1995) *Pthorimea operculella* (Lepidoptera: Gelechiidae) resistance in potato by expression of the *Bacillus thuringiensis* CryIA(b) insecticidal crystal protein. *J Econ Entomol* **88**, 1469–76.
10. Mandaokar A, Chakrabarti SK, Rao NGV, Kumar PA, Shama RP (1998) Short communication: A fusion gene coding for two different δ-endotoxins of *Bacillus thuringiensis* toxic to *Plutella xylostella* and useful for resistance management. *World J Microbiol Biotechnol* **14**, 599–601.
11. Coa J, Shelton AM, Earle ED (2005) Development of transgenic collards (*Brassica oleracea* L., var. *acephala*) expressing a *cry1Ac* or *cry1C* Bt gene for control of the diamondback moth. *Crop Protect* **24**, 804–13.
12. Thammasittirong A, Attathom T (2008) PCR-based method for the detection of *cry* genes in local isolates of *Bacillus thuringiensis* from Thailand. *J Invertebr Pathol* **98**, 121–6.
13. Ohba M, Aizawa K (1978) Serological identification of *Bacillus thuringiensis* and related bacteria isolated in Japan. *J Invertebr Pathol* **33**, 303–9.
14. de Barjac H, Franchon E (1990) Classification of *Bacillus thuringiensis* strains. *Entomophaga* **35**, 233–40.
15. Finney DJ (1971) *Probit Analysis*, Cambridge Univ Press, Cambridge.
16. Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn, Cold Spring Harbor Laboratory Press, New York.
17. Kuo WS, Chak KF (1996) Identification of novel *cry*-type genes from *Bacillus thuringiensis* strains on the basis of restriction fragment length polymorphism of the PCR-Amplified DNA. *Appl Environ Microbiol* **62**, 1369–77.
18. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **2**, 4673–80.
19. Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* **5**, 150–63.
20. Dayhoff MO, Schwartz RM, Orcutt BC (1978) Composition of proteins. In: Dayhoff MO (ed) *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, Washington, DC, pp 345–52.
21. Kopp J, Schwede T (2004) The SWISS-MODEL repository of annotated three-dimensional protein structure homology models. *Nucleic Acids Res* **32**, 230–4.
22. Lund O, Nielsen M, Lundegaard C, Worning P (2002) CPH models 2.0: X3M a computer program to extract 3D Models. In: CASP5 Conference, p A102.
23. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**, 2714–23.
24. Jung S, Park S, Kim D, Kim SB (2008) Denaturing gradient gel electrophoresis analysis of bacterial community profiles in the rhizosphere of cry1AC-carrying *Brassica rapa* subsp. pekinensis. *J Microbiol* **46**, 12–5.
25. Vidal-Quist JC, Castañera P, González-Cabrera J (2009) Simple and rapid method for PCR characteri-

zation of large *Bacillus thuringiensis* strain collections. *Curr Microbiol* **58**, 421–5.
26. Chak KF, Chao DC, Tseng MY, Kao SS, Tuan SJ, Feng TY (1994) Determination and distribution of *cry*-type genes of *Bacillus thuringiensis* isolates from Taiwan. *Appl Environ Microbiol* **60**, 2415–20.
27. Kim HS (2000) Comparative study of the frequency, flagellar serotype, crystal shape, toxicity and *cry* gene contents of *Bacillus thuringiensis* from three environments. *Curr Microbiol* **41**, 250–6.
28. Hongyu Z, Ziniu Y, Wangxi D (2000) Composition and ecological distribution of Cry proteins and their genotypes of *Bacillus thuringiensis* isolates from warehouse in China. *J Invertebr Pathol* **76**, 191–7.
29. Wang J, Boets A, Van Rie J, Ren G (2003) Characterization of *cry*1, *cry*2 and *cry*9 genes in *Bacillus thuringiensis* isolates from China. *J Invertebr Pathol* **82**, 63–71.
30. Seifinejad A, Salehi Jouzani GR, Hosseinzadeh A, Abdmishani C (2008) Characterization of Lepidoptera-active *cry* and *vip* genes in Iranian *Bacillus thuringiensis* strain collection. *Biol Contr* **44**, 216–26.
31. Carlton BC, Gonzalez JM Jr (1985) Plasmids and delta-endotoxin production in different subspecies of *Bacillus thuringiensis*. In: Hoch JA, Setow P (eds) *Molecular Biology of Microbial Differentiation*, American Society of Microbiology, Washington, DC, pp 246–52.
32. Carlson CR, Kolsto AB (1993) A complete physical map of a *Bacillus thuringiensis* chromosome. *J Bacteriol* **175**, 1053–60.
33. Martinez C, Ibarra JE, Caballero P (2005) Association analysis between serotype, *cry* gene content, and toxicity to *Helicoverpa armigera* larvae among *Bacillus thuringiensis* isolates native to Spain. *J Invertebr Pathol* **90**, 91–7.
34. Xue JL, Cai QX, Zheng DS, Yuan ZM (2005) The synergistic activity between Cry1Aa and Cry1C from *Bacillus thuringiensis* against *Spodoptera exigua* and *Helicoverpa armigera*. *Lett Appl Microbiol* **40**, 460–5.
35. Uribe D, Martinez W, Ceron J (2003) Distribution and diversity of *cry* genes in native strains of *Bacillus thuringiensis* obtained from different ecosystems from Columbia. *J Invertebr Pathol* **82**, 119–27.
36. Grochulski P, Masson L, Borisova S, Pusztai-Carey M, Schwartz JL, Brousseu R, Cygler M (1995) *Bacillus thuringiensis* CryIA(a) insecticidal toxin: crystal structure and channel formation. *J Mol Biol* **254**, 447–64.
37. Bravo A (1997) Phylogenetic relationships of the *Bacillus thuringiensis* δ-endotoxin family proteins and their functional domains. *J Bacteriol* **179**, 2793–801.
38. Sankaranarayanan R, Sekar K, Banerjee R, Sharma V, Surolia A, Vijayan M (1996) A novel mode of carbohydrate recognition in jacalin, a *Moraceae* plant lectin with a β-prism fold. *Nat Struct Mol Biol* **3**, 596–603.
39. Schwartz JL, Lu YJ, Sohnlein P, Brousseau R, Laprade R, Masson L, Adang MJ (1997) Ion channels formed in planar lipid bilayers by *Bacillus thuringiensis* toxins in the presence of *Manduca sexta* midgut receptors. *FEBS Lett* **412**, 270–6.
40. de Maagd RA, Kwa MSG, van der Klei H, Yamamoto T, Schipper B, Vlak JM, Stiekema WJ, Bosch D (1996) Domain III substitution in *Bacillus thuringiensis* delta-Endotoxin CryIA(b) results in superior toxicity for *Spodoptera exigua* and altered membrane protein recognition. *Appl Environ Microbiol* **62**, 1537–43.