J. Fac. Decmber. Biol. Sci.,
Hiroshima Univ. (1990), 29 : 95～107

Nucleotide Sequence of the Lepidoptera-toxic Protein Gene of *Bacillus thuringiensis* subsp. *dendrolimus* T84A1

Masashi Ogo, Shoichi Yamada, Yasuo Kobayashi,
Jiro Shibata and Yasunori Nagamatsu

Faculty of Applied Biological Science, Hiroshima University,
Higashi-Hiroshima 724, Japan

Received September 26, 1990

Abstract A 3882 nucleotides sequence including the complete crystal protein gene of *Bacillus thuringiensis (B. t.) dendrolimus* T84A1 was determined by the dideoxy-chain termination method. It was revealed that the gene of *B. t. dendrolimus* is highly homologous to that of *B. t. sotto* (Shibano *et al.*, 1986). According to the suggested nomenclature, this gene is classified into *cryIA(a)* (Höfte and Whiteley, 1989). The open reading frame encoded a 133487.71 Da protein consisting of 1180 amino acid residues. The hydropathy of the predicted crystal protein was also analyzed. When the crystal protein gene was expressed in *E. coli* MV1184 using the vector pUC118, the lower cultivation temperature (25℃) yield larger accumulation of the crystal protein than that at ordinary temperature (37℃). Based on the sequence determined in this study, the functional structure of the crystal protein was discussed.

INTRODUCTION

Bacillus thuringiensis is a gram-positive bacterium, and is well known to produce an insecticidal protein crystal during sporulation (δ-endotoxin). Some δ-endotoxins have been widely used to control certain species of insects.

Crystal proteins from a number of *B. t.* strains can be classified into four Cry protein classes and one Cyt (cytolytic) protein, based on the host spectra of insecticidal activities and the primary structures. The four Cry classes are (I) Lepidoptera-specific, (II) Lepidoptera-and Diptera-specific, (III) Coleoptera-specific, (IV) Diptera-specific, and the genes coding these proteins are called *cry* genes. And the remaining cytolytic protein gene is designated *cytA*.

From its lepidoptera-specific spectrum, the crystal protein produced by *B. t. dendrolimus* may belong to CryI. CryI proteins dissolve in the larval lumen and release smaller toxic polypeptide fragments. A trypsin-resistant polypeptide (Fragment T, app. M. W. 60-kDa) from *B. t. dendrolimus* crystal protein is the minimum unit and its N-terminal sequence was determined previously (Nagamatsu *et al.*, 1984). Several recent studies demonstrated the correlation between the insecticidal specificity and the interaction of the toxic fragments with the receptors in the brush border membrane of insect midgut (Hofmann *et al.*, 1988, Rie *et al.*, 1989).

It is important to know the full primary structure of the crystal protein, because it may

help to understand which part of the crystal protein interacts with the insect larval midgut, and how their biochemical mechanisms are. To date, several *cryI* genes from different *B. t.* strains were cloned and sequenced, but not *B. t. dendrolimus*. In this study, we present its complete nucleotide sequence and the deduced amino acid sequence of the crystal protein. Temperature influence on expression of the crystal protein in *E. coli* MV1184 was also examined.

MATERIALS AND METHODS

Hosts and Plasmids.

Escherichia coli MV1184 [*ara, Δ(lac-pro), strA, thi, (φ80Δ1acZΔM15), Δ (srl-recA)306::Tn10(tet^r);F' :traD36, proAB, lacI^{q}Z ΔM15*] was used as a host with plasmids, pUC118 and pUC119. Plasmid pHP206 was used as the source of full length of the crystal protein gene (Shibata *et al.*, 1989).

Nucleotide sequencing.

Plasmid pHP206 containing the 5 kb *Pst*I fragment of *B. t. dendrolimus* was digested with some restriction enzymes and the generated fragments were inserted into pUC118 or pUC119 vectors to create subclones for sequencing. Some of them were further digested with exonuclease III and Mung bean nuclease to generate overlapping deletion mutants (Henikoff, 1984). The template single-stranded DNAs were prepared by infecting with helper phage M13K07 (Vieira and Messing, 1987), and sequenced by the dideoxy chain-termination method (Sanger *et al.*, 1977).

Analysis of the open reading frame and the transcriptional terminator of the crystal protein gene, and the hydropathy profile of the crystal protein.

Searching for the ORF (open reading frame) and the transcriptional terminator and calculation of the average hydropathy indexes (Kyte and Doolittle, 1982) were carried out by using the computer program (GENETYX: SDC software development inc.) at Center for Gene Science Hiroshima University.

Construction of pPP03.

The 5 kb *Pst*I fragment containing the crystal protein gene on pHP206 was ligated with pUC118 and introduced into *E. coli* MV1184.

Analysis of the crystal protein produced in E. coli.

E. coli MV1184 cells carrying the pPP03 were grown in 100 ml of L-broth at 37℃ or 25℃ for 48 hr. Cell concentrations of *E. coli* MV1184 in a medium were measured by the absorbance at 600 nm.

The cells were harvested by centrifugation at 7000 × g for 10 min at 4℃, resuspended with 20 ml PBS, and sonicated at 30 W for 10 min in an ice-bath. After centrifugation at 15000 × g and 4℃ for 90 min, the crystal protein in the precipitate was prurified by selective solubilization in sodium carbonate buffer (Na^+; 0.1 M), pH 10.2, including 10 mM dithiothreitol. Samples prepared as described above were subjected to SDS (sodium dodecyl sulfate)-7.5% PAGE (polyacrylamide gel electrophoresis) analysis (Laemmli *et al.*, 1970). Protein bands were visualized by Coomasie Brilliant Blue stainning. The crystal protein was immunodetected after electrotransfer to a polyvinylidene difluoride membrane as described previously (Shibata *et al.*, 1989).

Transformation, media and other techniques.

Transformation of *E. coli* MV1184 was performed as described by CHUNG *et al.*, (1989). Media and other techniques used in this study were described by SAMBROOK *et al.*, (1989).

RESULTS

SHIBATA *et al.*, (1989) isolated a 9.91 kb *Pst*I fragment including the crystal protein gene from the 51.1 kb plasmid DNA of *Bacillus thuringiensis dendrolimus* T84A1. The entire *cry* gene resided on the 5.0 kb *Hpa*I -*Pst*I segment within the DNA fragment and it was the only *cry* gene in this *B. t.* strain. The 5.0 kb DNA fragment was attached with a *Pst*I linker and inserted into vector pBR322 to generate plasmid pHP206.

Figure 1 shows the restriction enzyme map of the DNA fragment and the sequencing strategy. By analyses of 48 partial sequences, a 3882-base continuous sequence from the *Hpa*I site to the second *Kpn*I site was determined.

In this sequence, two open reading frames (ORF) were found together with the two identical potential ribosome binding regions (Fig. 2). The shorter one (ORF2, 6 bases) is of course too short to encode the crystal protein. The calculated molecular weight (133,487

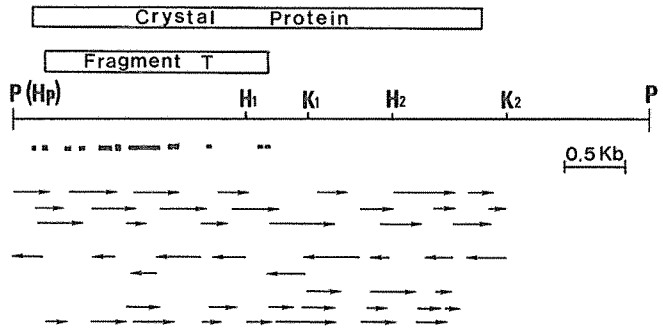

Fig. 1. Restriction enzyme map of the crystal protein gene and sequencing strategy.
Arrows represent the areas sequenced and their direction. Bold lines show the portion confirmed by amino acid sequencing. Abbreviations for restriction enzymes: P, *Pst*I ; H, *Hin*dIII ; K, *Kpn*I ; Hp, *Hpa*I.

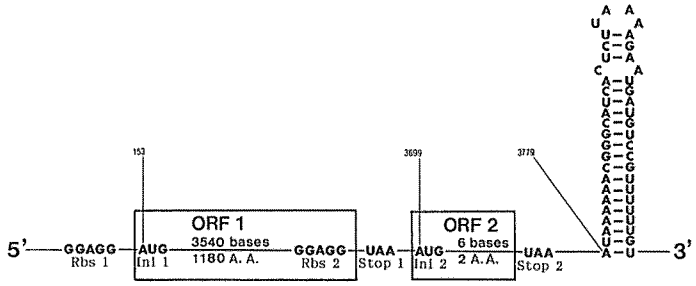

Fig. 2. Two ribosome binding sites and the structure of the transcriptional terminator in the crystal protein mRNA.
Ribosome binding sites are indicated as Rbs. Translational initiation sites and stop sites are indicated as Ini and Stop, respectively.

```
        10         20         30         40         50         60
AACACCCTGGGTCAAAAATTGATATTTAGTAAAATTAGTTGCACTTTGTGCATTTTTTCA

        70         80         90        100        110        120
TAAGATGAGTCATATGTTTTAAATTGTAGTAATGAAAAACAGTATTATATCATAATGAAT

       130        140        150        160        170        180
TGGTATCTTAATAAAAGAGATGGAGGTAACTTATGGATAACAATCCGAACATCAATGAAT
                                    MetAspAsnAsnProAsnIleAsnGluCys

       190        200        210        220        230        240
GCATTCCTTATAATTGTTTAAGTAACCCTGAAGTAGAAGTATTAGGTGGAGAAAGAATAG
  IleProTyrAsnCysLeuSerAsnProGluValGluValLeuGlyGlyGluArgIleGlu

       250        260        270        280        290        300
AAACTGGTTACACCCCAATCGATATTTCCTTGTCGCTAACGCAATTTCTTTTGAGTGAAT
  ThrGlyTyrThrProIleAspIleSerLeuSerLeuThrGlnPheLeuLeuSerGluPhe

       310        320        330        340        350        360
TTGTTCCCGGTGCTGGATTTGTGTTAGGACTAGTTGATATAATATGGGGAATTTTTGGTC
  ValProGlyAlaGlyPheValLeuGlyLeuValAspIleIleTrpGlyIlePheGlyPro

       370        380        390        400        410        420
CCTCTCAATGGGACGCATTTCTTGTACAAATTGAACAGTTAATTAACCAAAGAATAGAAG
  SerGlnTrpAspAlaPheLeuValGlnIleGluGlnLeuIleAsnGlnArgIleGluGlu

       430        440        450        460        470        480
AATTCGCTAGGAACCAAGCCATTTCTAGATTAGAAGGACTAAGCAATCTTTATCAAATTT
  PheAlaArgAsnGlnAlaIleSerArgLeuGluGlyLeuSerAsnLeuTyrGlnIleTyr

       490        500        510        520        530        540
ACGCAGAATCTTTTAGAGAGTGGGAAGCAGATCCTACTAATCCAGCATTAAGAGAAGAGA
  AlaGluSerPheArgGluTrpGluAlaAspProThrAsnProAlaLeuArgGluGluMet

       550        560        570        580        590        600
TGCGTATTCAATTCAATGACATGAACAGTGCCCTTACAACCGCTATTCCTCTTTTTGCAG
  ArgIleGlnPheAsnAspMetAsnSerAlaLeuThrThrAlaIleProLeuPheAlaVal

       610        620        630        640        650        660
TTCAAAATTATCAAGTTCCTCTTTTATCAGTATATGTTCAAGCTGCAAATTTACATTTAT
  GlnAsnTyrGlnValProLeuLeuSerValTyrValGlnAlaAlaAsnLeuHisLeuSer

       670        680        690        700        710        720
CAGTTTTGAGAGATGTTTCAGTGTTTGGACAAAGGTGGGGATTTGATGCCGCGACTATCA
  ValLeuArgAspValSerValPheGlyGlnArgTrpGlyPheAspAlaAlaThrIleAsn

       730        740        750        760        770        780
ATAGTCGTTATAATGATTTAACTAGGCTTATTGGCAACTATACAGATTATGCTGTGCGCT
  SerArgTyrAsnAspLeuThrArgLeuIleGlyAsnTyrThrAspTyrAlaValArgTrp

       790        800        810        820        830        840
GGTACAATACGGGATTAGAGCGTGTATGGGGACCGGATTCTAGAGATTGGGTAAGGTATA
  TyrAsnThrGlyLeuGluArgValTrpGlyProAspSerArgAspTrpValArgTyrAsn

       850        860        870        880        890        900
ATCAATTTAGAAGAGAGCTAACACTTACTGTATTAGATATCGTTGCTCTATTCTCAAATT
  GlnPheArgArgGluLeuThrLeuThrValLeuAspIleValAlaLeuPheSerAsnTyr

       910        920        930        940        950        960
ATGATAGTCGAAGGTATCCAATTCGAACAGTTTCCCAATTAACAAGAGAAATTTATACGA
  AspSerArgArgTyrProIleArgThrValSerGlnLeuThrArgGluIleTyrThrAsn
```

Fig. 3. Nucleotide sequence and the deduced amio acid sequence of the crystal protein.

```
      970        980        990       1000       1010       1020
ACCCAGTATTAGAAAATTTTGATGGTAGTTTTCGTGGAATGGCTCAGAGAATAGAACAGA
  ProValLeuGluAsnPheAspGlySerPheArgGlyMetAlaGlnArgIleGluGlnAsn

      1030       1040       1050       1060       1070       1080
ATATTAGGCAACCACATCTTATGGATATCCTTAATAGGATAACCATTTATACTGATGTGC
  IleArgGlnProHisLeuMetAspIleLeuAsnArgIleThrIleTyrThrAspValHis

      1090       1100       1110       1120       1130       1140
ATAGAGGCTTTAATTATTGGTCAGGGCATCAAATAACAGCTTCTCCTGTAGGGTTTTCAG
  ArgGlyPheAsnTyrTrpSerGlyHisGlnIleThrAlaSerProValGlyPheSerGly

      1150       1160       1170       1180       1190       1200
GACCAGAATTCGCATTCCCTTTATTTGGGAATGCGGGGAATGCAGCTCCACCCGTACTTG
  ProGluPheAlaPheProLeuPheGlyAsnAlaGlyAsnAlaAlaProProValLeuVal

      1210       1220       1230       1240       1250       1260
TCTCATTAACTGGTTTGGGGATTTTTAGAACATTATCTTCACCTTTATATAGAAGAATTA
  SerLeuThrGlyLeuGlyIlePheArgThrLeuSerSerProLeuTyrArgArgIleIle

      1270       1280       1290       1300       1310       1320
TACTTGGTTCAGGCCCAAATAATCAGGAACTGTTTGTCCTTGATGGAACGGAGTTTTCTT
  LeuGlySerGlyProAsnAsnGlnGluLeuPheValLeuAspGlyThrGluPheSerPhe

      1330       1340       1350       1360       1370       1380
TTGCCTCCCTAACGACCAACTTGCCTTCCACTATATATAGACAAAGGGGTACAGTCGATT
  AlaSerLeuThrThrAsnLeuProSerThrIleTyrArgGlnArgGlyThrValAspSer

      1390       1400       1410       1420       1430       1440
CACTAGATGTAATACCGCCACAGGATAATAGTGTACCACCTCGTGCGGGATTTAGCCATC
  LeuAspValIleProProGlnAspAsnSerValProProArgAlaGlyPheSerHisArg

      1450       1460       1470       1480       1490       1500
GATTGAGTCATGTTACAATGCTGAGCCAAGCAGCTGGAGCAGTTTACACCTTGAGAGCTC
  LeuSerHisValThrMetLeuSerGlnAlaAlaGlyAlaValTyrThrLeuArgAlaPro

      1510       1520       1530       1540       1550       1560
CAACGTTTTCTTGGCAGCATCGCAGTGCTGAATTTAATAATATAATTCCTTCATCACAAA
  ThrPheSerTrpGlnHisArgSerAlaGluPheAsnAsnIleIleProSerSerGlnIle

      1570       1580       1590       1600       1610       1620
TTACACAAATACCTTTAACAAAATCTACTAATCTTGGCTCTGGAACTTCTGTCGTTAAAG
  ThrGlnIleProLeuThrLysSerThrAsnLeuGlySerGlyThrSerValValLysGly

      1630       1640       1650       1660       1670       1680
GACCAGGATTTACAGGAGGAGATATTCTTCGAAGAACTTCACCTGGCCAGATTTCAACCT
  ProGlyPheThrGlyGlyAspIleLeuArgArgThrSerProGlyGlnIleSerThrLeu

      1690       1700       1710       1720       1730       1740
TAAGAGTAAATATTACTGCACCATTATCACAAAGATATCGGGTAAGAATTCGCTACGCTT
  ArgValAsnIleThrAlaProLeuSerGlnArgTyrArgValArgIleArgTyrAlaSer

      1750       1760       1770       1780       1790       1800
CTACTACAAATTTACAATTCCATACATCAATTGACGGAAGACCTATTAATCAGGGTAATT
  ThrThrAsnLeuGlnPheHisThrSerIleAspGlyArgProIleAsnGlnGlyAsnPhe

      1810       1820       1830       1840       1850       1860
TTTCAGCAACTATGAGTAGTGGGAGTAATTTACAGTCCGGAAGCTTTAGGACTGTAGGTT
  SerAlaThrMetSerSerGlySerAsnLeuGlnSerGlySerPheArgThrValGlyPhe

      1870       1880       1890       1900       1910       1920
TTACTACTCCGTTTAACTTTTCAAATGGATCAAGTGTATTTACGTTAAGTGCTCATGTCT
  ThrThrProPheAsnPheSerAsnGlySerSerValPheThrLeuSerAlaHisValPhe
```

Fig. 3. Continued.

```
      1930       1940       1950       1960       1970       1980
TCAATTCAGGCAATGAAGTTTATATAGATCGAATTGAATTTGTTCCGGCAGAAGTAACCT
  AsnSerGlyAsnGluValTyrIleAspArgIleGluPheValProAlaGluValThrPhe

      1990       2000       2010       2020       2030       2040
TTGAGGCAGAATATGATTTAGAAAGAGCACAAAAGGCGGTGAATGAGCTGTTTACTTCTT
  GluAlaGluTyrAspLeuGluArgAlaGlnLysAlaValAsnGluLeuPheThrSerSer

      2050       2060       2070       2080       2090       2100
CCAATCAAATCGGGTTAAAAACAGATGTGACGGATTATCATATTGATCAAGTATCCAATT
  AsnGlnIleGlyLeuLysThrAspValThrAspTyrHisIleAspGlnValSerAsnLeu

      2110       2120       2130       2140       2150       2160
TAGTTGAGTGTTTATCAGATGAATTTTGTCTGGATGAAAAACAAGAATTGTCCGAGAAAG
  ValGluCysLeuSerAspGluPheCysLeuAspGluLysGlnGluLeuSerGluLysVal

      2170       2180       2190       2200       2210       2220
TCAAACATGCGAAGCGACTTAGTGATGAGCGGAATTTACTTCAAGATCCAAACTTCAGAG
  LysHisAlaLysArgLeuSerAspGluArgAsnLeuLeuGlnAspProAsnPheArgGly

      2230       2240       2250       2260       2270       2280
GGATCAATAGACAACTAGACCGTGGCTGGAGAGGAAGTACGGATATTACCATCCAAGGAG
  IleAsnArgGlnLeuAspArgGlyTrpArgGlySerThrAspIleThrIleGlnGlyGly

      2290       2300       2310       2320       2330       2340
GCGATGACGTATTCAAAGAGAATTACGTTACGCTATTGGGTACCTTTGATGAGTGCTATC
  AspAspValPheLysGluAsnTyrValThrLeuLeuGlyThrPheAspGluCysTyrPro

      2350       2360       2370       2380       2390       2400
CAACGTATTTATATCAAAAAATAGATGAGTCGAAATTAAAAGCCTATACCCGTTATCAAT
  ThrTyrLeuTyrGlnLysIleAspGluSerLysLeuLysAlaTyrThrArgTyrGlnLeu

      2410       2420       2430       2440       2450       2460
TAAGAGGGTATATCGAAGATAGTCAAGACTTAGAAATCTATTTAATTCGCTACAATGCAA
  ArgGlyTyrIleGluAspSerGlnAspLeuGluIleTyrLeuIleArgTyrAsnAlaLys

      2470       2480       2490       2500       2510       2520
AACATGAAACAGTAAATGTGCCAGGTACGGGTTCCTTATGGCCGCTTTCAGCCCAAAGTC
  HisGluThrValAsnValProGlyThrGlySerLeuTrpProLeuSerAlaGlnSerPro

      2530       2540       2550       2560       2570       2580
CAATCGGAAAGTGTGGAGAGCCGAATCGATGCGCGCCACACCTTGAATGGAATCCTGACT
  IleGlyLysCysGlyGluProAsnArgCysAlaProHisLeuGluTrpAsnProAspLeu

      2590       2600       2610       2620       2630       2640
TAGATTGTTCGTGTAGGGATGGAGAAAAATGTGCCCATCATTCCCATCATTTCTCCTTGG
  AspCysSerCysArgAspGlyGluLysCysAlaHisHisSerHisHisPheSerLeuAsp

      2650       2660       2670       2680       2690       2700
ACATTGATGTTGGATGTACAGACTTAAATGAGGACTTAGGTGTATGGGTGATATTCAAGA
  IleAspValGlyCysThrAspLeuAsnGluAspLeuGlyValTrpValIlePheLysIle

      2710       2720       2730       2740       2750       2760
TTAAGACGCAAGATGGCCATGCAAGACTAGGAAATCTAGAATTTCTCGAAGAGAAACCAT
  LysThrGlnAspGlyHisAlaArgLeuGlyAsnLeuGluPheLeuGluGluLysProLeu

      2770       2780       2790       2800       2810       2820
TAGTAGGAGAAGCACTAGCTCGTGTGAAAAGAGCGGAGAAAAAATGGAGAGACAAACGTG
  ValGlyGluAlaLeuAlaArgValLysArgAlaGluLysLysTrpArgAspLysArgGlu

      2830       2840       2850       2860       2870       2880
AAAAATTGGAATGGGAAACAAATATTGTTTATAAAGAGGCAAAAGAATCTGTAGATGCTT
  LysLeuGluTrpGluThrAsnIleValTyrLysGluAlaLysGluSerValAspAlaLeu
```

Fig. 3. Continued.

```
      2890       2900       2910       2920       2930       2940
TATTTGTAAACTCTCAATATGATAGATTACAAGCGGATACCAACATCGCGATGATTCATG
  PheValAsnSerGlnTyrAspArgLeuGlnAlaAspThrAsnIleAlaMetIleHisAla

      2950       2960       2970       2980       2990       3000
CGGCAGATAAACGCGTTCATAGCATTCGAGAAGCTTATCTGCCTGAGCTGTCTGTGATTC
  AlaAspLysArgValHisSerIleArgGluAlaTyrLeuProGluLeuSerValIlePro

      3010       3020       3030       3040       3050       3060
CGGGTGTCAATGCGGCTATTTTTGAAGAATTAGAAGGGCGTATTTTCACTGCATTCTCCC
  GlyValAsnAlaAlaIlePheGluGluLeuGluGlyArgIlePheThrAlaPheSerLeu

      3070       3080       3090       3100       3110       3120
TATATGATGCGAGAAATGTCATTAAAAATGGTGATTTTAATAATGGCTTATCCTGCTGGA
  TyrAspAlaArgAsnValIleLysAsnGlyAspPheAsnAsnGlyLeuSerCysTrpAsn

      3130       3140       3150       3160       3170       3180
ACGTGAAAGGGCATGTAGATGTAGAAGAACAAAACAACCACCGTTCGGTCCTTGTTGTTC
  ValLysGlyHisValAspValGluGluGlnAsnAsnHisArgSerValLeuValValPro

      3190       3200       3210       3220       3230       3240
CGGAATGGGAAGCAGAAGTGTCACAAGAAGTTCGTGTCTGTCCGGGTCGTGGCTATATCC
  GluTrpGluAlaGluValSerGlnGluValArgValCysProGlyArgGlyTyrIleLeu

      3250       3260       3270       3280       3290       3300
TTCGTGTCACAGCGTACAAGGAGGGATATGGAGAAGGTTGCGTAACCATTCATGAGATCG
  ArgValThrAlaTyrLysGluGlyTyrGlyGluGlyCysValThrIleHisGluIleGlu

      3310       3320       3330       3340       3350       3360
AGAACAATACAGACGAACTGAAGTTTAGCAACTGTGTAGAAGAGGAAGTATATCCAAACA
  AsnAsnThrAspGluLeuLysPheSerAsnCysValGluGluGluValTyrProAsnAsn

      3370       3380       3390       3400       3410       3420
ACACGGTAACGTGTAATGATTATACTGCGACTCAAGAAGAATATGAGGGTACGTACACTT
  ThrValThrCysAsnAspTyrThrAlaThrGlnGluGluTyrGluGlyThrTyrThrSer

      3430       3440       3450       3460       3470       3480
CTCGTAATCGAGGATATGACGGAGCCTATGAAAGCAATTCTTCTGTACCAGCTGATTATG
  ArgAsnArgGlyTyrAspGlyAlaTyrGluSerAsnSerSerValProAlaAspTyrAla

      3490       3500       3510       3520       3530       3540
CATCAGCCTATGAAGAAAAAAGCATATACAGATGGACGAAGAGACAATCCTTGTGAATCTA
  SerAlaTyrGluGluLysAlaTyrThrAspGlyArgArgAspAsnProCysGluSerAsn

      3550       3560       3570       3580       3590       3600
ACAGAGGATATGGGGATTACACACCACTACCAGCTGGCTATGTGACAAAAGAATTAGAGT
  ArgGlyTyrGlyAspTyrThrProLeuProAlaGlyTyrValThrLysGluLeuGluTyr

      3610       3620       3630       3640       3650       3660
ACTTCCCAGAAACCGATAAGGTATGGATTGAGATCGGAGAAACGGAAGGAACATTCATCG
  PheProGluThrAspLysValTrpIleGluIleGlyGluThrGluGlyThrPheIleVal

      3670       3680       3690       3700       3710       3720
TGGACAGCGTGGAATTACTTCTTATGGAGGAATAATATATGCTTTAAAATGTAAGGTGTG
  AspSerValGluLeuLeuLeuMetGluGlu

      3730       3740       3750       3760       3770       3780
CAAATAAAGAATGATTACTGACTTGTATTGACAGATAAATAAGGAAATTTTTATATGAAT

      3790       3800       3810       3820       3830       3840
AAAAAACGGGCATCACTCTTAAAAGAATGATGTCCGTTTTTTGTATGATTTAACGAGTGA

      3850       3860       3870       3880
TATTTAAATGTTTTTTTGCGAAGGCTTTACTTAACGGGGTAC
```

Fig. 3. Continued.

Da) of the ORF1 protein is well consistent with the apparent molecular weight 140,000 estimated on SDS-PAGE of the crystal protein. Among the several stem loop structures (data not shown), the largest (20 bases) one locating behind the two ORFs was revealed to be a potential transcriptional terminator because of the highest stability and its location. The minimum free energy (ΔG) for the structure formation was calculated as -25.1 kcal/mol.

Detailed nucleotide sequence and the deduced amino acid sequence were shown in Fig. 3. ORF1 (#153-3692) encodes 1180 amino acid residues. The ribosome binding sequence is located six bases upstream from the initiation codon. The N-terminal amino acid sequence of the crystal protein and the several partial sequences within the toxic trypsin-resistant fragment (Fragment) were also determined (NAGAMATSU *et al.*, 1984, and our unpublished data). The sequences underlined were identical with those deduced from the nucleotide sequences, giving the evidence that ORF1 is the crystal protein gene. The N-terminal and the C-terminal residues of Fragment T correspond to #29 and #618 amino acid residues of the crystal protein, respectively. The minimum toxic unit resides on the N-terminal half of the crystal protein. Compared with the N-terminal half and the remaining C-terminal half, most of cysteine and lysine residues distributed in the latter region.

From the modes of action of several well-studied protein toxins, it was expected that Fragment T could be consisted of several functional domains. A hydropathy profile of the crystal protein was created based on the deduced amino acid sequence (Fig. 4). The N-ter-

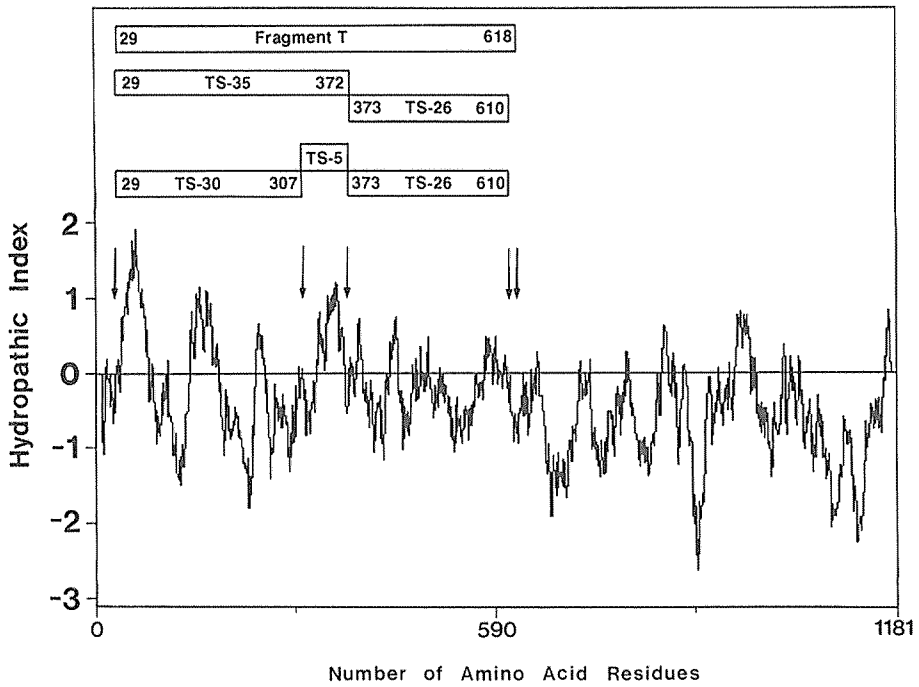

Fig. 4. Hydropathy profile of the crystal protein.
The location of the minimum toxic region (Fragment T) and the polypeptides generated by subtilisin cleavage were shown by the boxes. Trypsin and subtilisin cleaving sites are given by arrows.

minal half of the crystal protein is more hydrophobic as than the C-terminal half. Trypsin cleavage sites, the N- and C-terminals of Fragment T, are located in boundary regions between the hydrophilic regions and the hydrophobic regions. Further digestion of the low specific protease, subtilisin gave two (TS-35, TS-26) or three (TS-30, TS-5, TS-26) polypeptides. The results from their N-terminal and C-terminal sequence analysis and the sequence determined in this study, these polypeptides was positioned as illustrated in Fig. 4. These subtilisin cleavage sites were also positioned in the boundary regions between the hydrophilic regions and the hydrophobic regions. These generated polypeptides associate each other, making a nicked protein missing the insecticidal activity. Highly hydrophobic region can be seen around #50 amino acid residue.

The expression of the crystal protein gene in *E. coli* was demonstrated, but the productivity was quite low (SHIBATA, *et al.*, 1989). Using the recombinant plasmid pPP03 (vector pUC118-host *E. coli* MV1184 system), influence of the cultivated temperature on expression of the crystal protein gene was examined. Figure 5 shows the SDS-PAGE patterns of proteins in the soluble and insoluble fractions. In the insoluble fraction (lane 5) from *E. coli*

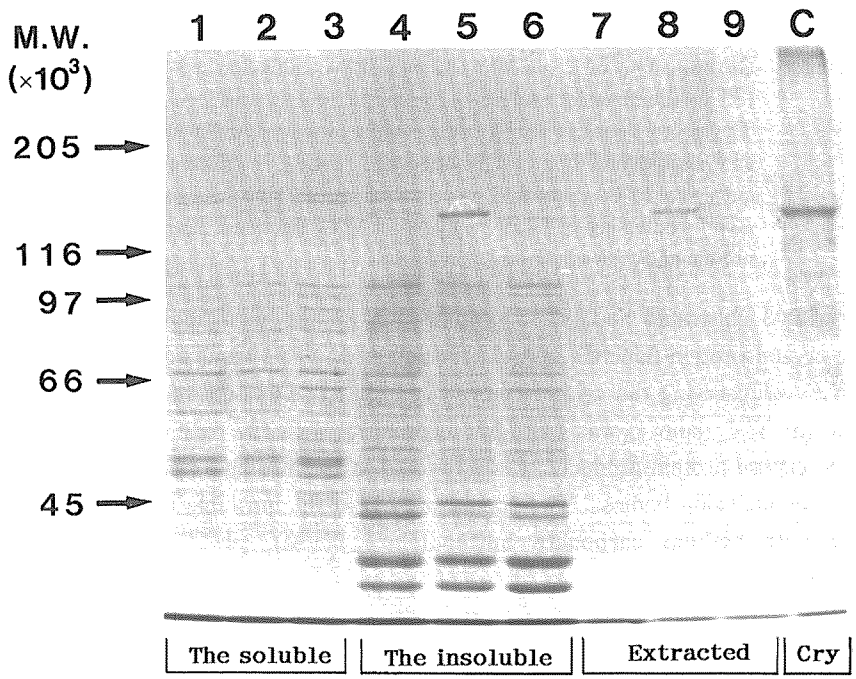

Fig. 5. SDS-PAGE analysis of the crystal protein produced in *E. coli* cultivated at 25℃ and 37℃.
E. coli carrying the recombinant plasmid, pPP03 (pUC118 containing the crystal protein gene), was cultivated at 37℃ (lanes 1, 4 and 7) and 25℃ (lanes 2, 5 and 8) for 48 hr. As a control, *E. coli* carrying vector pUC118, was also cultivated (lanes 3, 6 and 9). The cell suspension was disrupted by sonication and fractionated into the soluble (lanes 1, 2, 3) and the insoluble (lanes 4, 5, 6). Each of the insoluble fractions were extracted with the buffer, pH 10.2, containing 10 mM dithiothreitol (lanes 7, 8, 9).

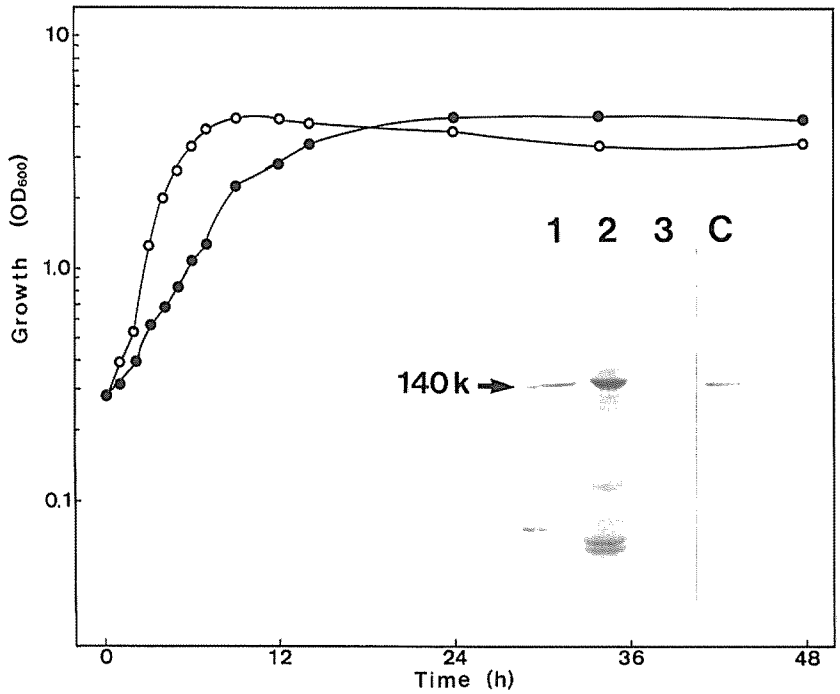

Fig. 6. Growth curves of *E. coli* MV1184 with pPP03 cultivated at 37℃ (○) and 25℃ (●) and immunoblot analysis of the crystal protein produced (the insert). The optical densities of cell cultures were measured at 600 nm. Immunoblot analysis with anti-fragment T antibodies. Lane 1, *E. coli* containing pPP03 incubated at 37℃. Lane 2, *E. coli* containing pPP03 incubated at 25℃. Lane 3, *E. coli* containing pUC118 incubated at 37℃. fractions. Lane C, the crystal protein of *B.t. dendrolimus*.

carrying pPP03 cultivated at 25℃, a protein band of the same molecular size as that of authentic crystal protein (lane C), was clearly detected. This band was not seen in the fraction from *E. coli* carrying the vector (control, lane 6). The amount of the protein band from the culture at 37℃ (lane 4) was quite low comparing with the case of 25℃ (lane 5). The bulk of the crystal protein produced was recovered in the insoluble fraction, indicating its accumulation as inclusion bodies. The crystal protein in the insoluble fraction was selectively extracted with sodium carbonate buffer (Na^+; 0.1 M), pH 10.2, containing 10 mM dithiothreitol (lane 8).

Figure 6 shows growth curves of *E. coli* cultivated at 25℃ and 37℃. The growth rate at 25℃ was lower during the log phase than that at 37℃, but cell concentrations in the stationary phase were on a similar level at those two temperatures. Influence of cultivation temperature on the crystal protein production was again examined by immunoblot analysis with anti-Fragment T antibody. Expression of the crystal protein gene in this *E. coli* system was confirmed, and accumulation of the crystal protein was higher at the lower temperature.

DISCUSSION

In this study, the crystal protein structural gene of *Bacillus thuringiensis dendrolimus*

was found to be consisted of 3540 nucleotides encoding 1180 amino acid residues.

Comparison of the sequence for the *dendrolimus* strain with other sequences for several *B. t.* strains revealed its high homology with *cryIA(a)* genes. Between the *kurstaki* HD-1 Dipel gene (SCHNEPF *et al.*, 1985) and the *dendrolimus* gene, there are 57 nucleotides differences; 19 amino acid residues are substituted and 4 residues are deleted in the *kurstaki* protein. Within the toxin region (the N-terminal half of the crystal protein), only 3 amino acids changes were seen. Compairng with the *sotto* gene (SHIBANO *et al.*, 1986), only one nucleotide difference at position 2795, C for *sotto*, G for *dendrolimus* was found. Structural differences were revealed, in 5'-region (SHIBATA *et al.*, 1989). Subspecies *sotto* is categorized into the same serotype as subsp. *dendrolimus*.

YAMADA (1990) showed two transcriptional start sites of the *B. t. dendrolimus* crystal protein gene, Bt I (#83 base in early sporulation), and Bt II (#69 base in mid sporulation). Results for the *kurstaki* HD-1 Dipel gene (WONG *et al.*, 1983) were coincident with the case of *dendrolimus*, and the promotor regions of both strains were well conserved.

A very short open reading frame (ORF2) with the identical ribosome binding sequence was found in the 3'-terminal regions of the crystal protein gene (Fig. 2). Influence of the second ribosome binding sequence on the crystal protein production remains to be studied.

The hydropathy and the distribution of certain amino acid revealed the differences in protein structural features between the N- and C-terminal halves. The C-terminal half containing most of cysteine residues may be involved in crystal formation. BIETLOT *et al.*, (1990) suggested that the thiol groups of the crystal protein are exposed on the surface of the molecule, and form symmetrical interchain disulfide bridges. Comparing with the C-terminal half, several hydrophobic peaks were clearly seen in the toxic N-terminal half, suggesting that the existence of several structural domains. Subtilisin digestion of the toxic fragment gave three polypeptide fragments, generating no toxic protein. One of the hydrophobic peaks locates between the two subtilisin cleavage sites, and this part corresponding to TS-5 domain (Fig. 4). Comparison of many crystal protein sequences revealed two hypervariable regions in the minimum toxic unit (GEISER *et al.*, 1986). One of them located from #286-457 corresponds to the boundary regions of three domains with TS-5 as the central figure. The other one #466-618 corresponds to the C-terminal remaining part of TS-26 region, and most of TS-30 region are well conserved. The combination of the primary sequence analysis and the domain searching with limited proteolysis would be a great help to study the relation between the function and the structure. Subtilisin-cleaved nicked protein missed the insecticidal activity, suggesting some synergestic action of these domains. Binding of this toxin to the receptor, a specificity-determining step, might be exerted by the interface region or the C-terminal part of TS-26 polypeptide. In addition, highly hydrophobic region is observed around #50 amino acid residue. Role of the potential membrane-spanning sequence is very interested from the point of view that this protein is a membrane-acting toxin.

The production of the crystal protein in *E. coli* was higher at 25℃ than that at 37℃ in contrast to the growth rates at these temperatures. As the explanation of the preferential accumulation at the lower temperature, the following three factors could be considered; ① the gene may has a temperature-dependent regulation system, ② the temperature difference

would affect the protein synthesis rate, ③ the degradation rate of the crystal protein might change at different temperatures. *B. t. dendrolimus* incubated at temperatures above 30℃ decrease the crystal productivity. So temperature-dependent expression of the crystal protein may be regulated under a similar mechanism in both organisms.

ACKNOWLEDGEMENT

We would like to thank all staff members of Center for Gene Science Hiroshima University for their kind offer of the facilities.

REFERENCES

Bietlot, H. P. L., Vishnubhatla, I., Carey, D. R., and Pozsgay, M., 1990, Characterization of the cysteine residues and disulfide linkages in the protein crystal of *Bacillus thuringiensis. Biochem. J.*, 267: 309-315.

Chung, C., Niemela, S. L. and Miller, R. H., 1989, One-step preparation of competent *Escherichia coli*: Transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. USA*, 86: 2172-2175.

Geister, M., Schweitzer, S., and Grimm, C., 1986, The hypervariable region in the genes coding for entomopathogenic crystal proteins of Bacillus thuringiensis: nucleotide sequence of the *kurhdl* gene of subsp. *kurstaki* HD1. *Gene*, 48: 109-118.

Henikoff, S., 1984, Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene*, 28: 351-359.

Hofmann, C., Vanderbruggen, H., Höfte, H., Van Rie, J., Jansens, S. and Van Mellaert, H., 1988, Sepecificity of *Bacillus thuringiensis* δ-endotoxins is correlated with the presence of high-affinity binding sites in the brush border membrane of target insect midguts. *Proc. Natl. Acad. Sci. USA*, 85: 7844-7848.

Höfte, H. and Whiteley, H.R., 1989, Insecticidal crystal proteins of *Bacillus thuringiensis. Microbiological Reviews*, 53: 242-255.

Kyte, J. and Doolittle, R. F., 1982, A simple method for displaying the hydropathic Character of a protein. *J.Mol.Biol.*, 157: 105-132.

Laemmli, U. K., 1970, Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227: 680.

Nagamatsu, Y., Itai, Y., Hatanaka, C., Funatsu, G. and Hayashi, K., 1984, A toxic fragment from the entomocidal crystal protein of *Bacillus thuringiensis. Agric. Biol. Chem.*, 43: 661-619.

Sambrook, J., Fritsch, E. F. and Maniatis, T., 1989, "Molecular Cloning 2nd Edition", Cold Spring Harbor Laboratory, New York.

Sanger, F., Nicklen, S. and Coulson, A. R., 1977, DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, 74: 5463-5467.

Schnepf, H. E., Wong, H. C. and Whiteley, H. R., 1985, Amino acid sequence of a crystal protein from *Bacillus thuringiensis* deduced from the DNA base sequence. *J. Biol. Chem.*, 260: 6264-6272.

Shibano, Y., Yamagata, A., Amachi, T., and Takanami, M., 1986, In "Bacillus Molecular Genetics and Biotechology Applications" (Ganesan, A. T. and Hoch J. H. *ed.*), pp. 307-320, Academic Press, New York.

Shibata, J., Tamiya, M., and Nagamatsu, Y., 1989, Cloning and Partial Characterization of the *Bacillus thuringiensis dendrolimus* T84A1 Insecticidal Crystal Protein Gene. *J. Fac. Appl. Biol. Sci., Hiroshima Uviv.* 28: 63-69.

VAN RIE, J., MCGAUGHEY, W. H., JOHNSON, D. E., BARNETT, B. D. and VAN MELLAERT, H., 1990, Mechanism of insect resistance to the microbial insecticide *Bacillus thuringiensis*. *Science*, **247**: 72-74.

VIEIRA, J. and MESSING, J., 1987, In "Methods in Enzymology" (WU, R., GROSSMAN, L. and MOLDAVA, K. ed.), Vol. 153, pp 3-11, Academic Press, New York.

WONG, H. C., SCHNEPF, H. E. and WHITELEY, H. R., 1983, Transcriptional and translational start sites for the *Bacillus thuringiensis* crystal protein gene. *J. Biol. Chem.*, **258**: 1960-1967.

YAMADA, S., 1990 Master thesis, Dept. Appl. Biol. Sci., Hiroshima Univ.

Bacillus thuringiensis dendrolimus T84A1 の殺虫性蛋白質遺伝子の塩基配列の決定

尾郷　正志・山田　彰一・小林　泰夫
柴田　治郎・永松　康徳

広島大学生物生産学部，東広島市 724

Bacillus thuringiensis dendrolimus T84A1 の殺虫性蛋白質遺伝子を含む 3882 塩基の配列を決定した。本菌の結晶蛋白質は 1180 個のアミノ酸（分子量133, 487 Da）から成ることが推定された。

蛋白質の研究と今回決定した配列の結果から，最小の殺虫活性断片である Fragment T が結晶蛋白質のN-末端側の半分に位置することがわかった。結晶蛋白質のハイドロパシー分析からN-末端側の半分の領域は疎水的でC-末端側は親水的であることが明らかになった。そのC-末端側に大部分のシステイン，リジン残基が偏って存在していることから，C-末端側は結晶の形成に関与していることが予測された。

サテライシンによる Fragment T の限定水解とハイドロパシーの解析によって３つのドメインから成る毒性発現領域の構造の構築が明確になった。

また，大腸菌における殺虫性結晶蛋白質の生産に与える培養温度の影響を検討したところ，通常の培養温度である37℃に比べてより低い25℃の方が顕著な蓄積を示すことが明らかになった。