

## Expression of a Cloned *Bacillus thuringiensis* Crystal Protein Gene in *Escherichia coli*

H. ERNEST SCHNEPF, HING C. WONG,<sup>†</sup> AND H. R. WHITELEY\*

Department of Microbiology, University of Washington, Seattle, Washington 98195

Received 16 April 1987/Accepted 11 June 1987

**The expression in *Escherichia coli* of a cloned crystal protein gene from *Bacillus thuringiensis* was investigated through the use of fusions of the crystal protein gene promoter to  $\beta$ -galactosidase and catechol oxidase genes. Analysis of deletion and insertion derivatives of the crystal protein gene promoter showed that a region of *B. thuringiensis* DNA located between 87 and 258 base pairs upstream from the transcription initiation site caused reduced transcription from this promoter. Insertion of Tn5 145 base pairs upstream from the transcription initiation site resulted in overproduction of the crystal protein. S1 nuclease mapping experiments failed to detect transcription from an outwardly directed promoter in Tn5, indicating that the overproduction resulted from the disruption or repositioning of the transcription-suppressing region.**

During sporulation, *Bacillus thuringiensis* produces insecticidal, bipyramidal crystalline inclusion bodies that are composed of one or more polypeptides called crystal proteins. The genes encoding several of these crystal proteins have been cloned and sequenced (reviewed in reference 28). Our previous analysis of the expression of a crystal protein gene cloned from *B. thuringiensis* subsp. *kurstaki* HD-1-Dipel (31) indicated that transcription of the gene in *Escherichia coli* is not dependent on the stage of growth and that it initiates at a distinct start site lying between the transcription start sites used during sporulation in *B. thuringiensis*. The cloned gene is expressed relatively poorly in the heterologous *E. coli* host, and crystals cannot be detected. However, we isolated a mutant in which Tn5 had inserted 145 base pairs (bp) upstream from the transcription start site (29). This mutant produces substantial amounts of crystal protein that accumulate in bipyramidal crystals resembling those found in *B. thuringiensis* (H. R. Whiteley, H. E. Schnepf, K. Tomczak, and J. C. Lara, in K. Maramorosch, ed., *Biotechnology Advances in Invertebrate Pathology and Cell Culture*, in press).

Since Tn5 insertions normally cause polar mutations in operons, it seemed unusual that the expression of a gene could be stimulated by the insertion of Tn5 some distance from its promoter. While there have been some reports showing that Tn5 is occasionally nonpolar or stimulatory towards the expression of downstream genes, the mechanism by which Tn5 enhances downstream gene expression is not understood. Therefore, an investigation of how an upstream Tn5 insertion mutation enhances the expression of the crystal protein gene could provide some insight into the sporadic anomalous behavior of this transposon in *E. coli*.

In the current study, we used 5'-end deletions and S1 nuclease mapping to analyze the effect of Tn5 insertions near the promoter and to localize the promoter for the crystal protein gene in *E. coli*. The results indicated that (i) no transcription could be detected from a promoter located in Tn5 or from a promoter created by Tn5 insertion; (ii) deletion of a segment of DNA 87 to 258 bp upstream from the transcription start site caused increased transcription from the crystal protein promoter; and (iii) deletions closer than

47 bp severely reduced transcription. A comparison of the effects of 5'-end deletions and Tn5 insertions suggested that the increased expression caused by Tn5 in the overproducer strain resulted from disruption or displacement of the -87 to -258 DNA segment.

### MATERIALS AND METHODS

**Bacterial and bacteriophage strains and plasmids.** *E. coli* HB101 bearing the pBR322-derived recombinant plasmid pES1 was the source of the DNA coding for the crystal protein gene from *B. thuringiensis* subsp. *kurstaki* HD-1-Dipel (23). Bacteriophages M13 mp8 and mp9, *E. coli* JM103, and plasmids pUC8, pUC9, and pUC13 (13, 18) have been described previously. Plasmid pMC1403 and *E. coli* MC1000 were kindly provided by Malcolm J. Casadaban, University of Chicago. Plasmids and double-stranded DNA of M13 phages were prepared by the method of Birnboim and Doly (3). Insertions of the kanamycin resistance transposon Tn5 into plasmids pES1 and pPLZ-1600 were made as described previously (31). The position of Tn5 in the resulting mutant plasmids was determined by restriction enzyme mapping and nucleotide sequence analysis.

**Plasmid stability and culture maintenance.** All experiments were performed with freshly transformed cells or cells taken from stock cultures made from freshly transformed cells. Plasmid preparations from all strains used for assays of gene expression were subjected to restriction enzyme analysis to confirm the plasmid content. The stock cultures were stored in 50% glycerol-50% LB medium (14) at -70°C and were used three or fewer times before disposal. This procedure was adopted because several of the *E. coli* strains containing *lacZ* fusion plasmids that overproduced  $\beta$ -galactosidase were found to revert to lower levels of production with successive passages on agar plates. Despite this culture instability, gross rearrangement of the plasmids was not observed except as noted. In addition, restriction digests of some plasmid preparations of pPLZ-377::Tn5 and pPLZ-12::Tn5 (two Tn5 insertion derivatives of the *lacZ* fusion plasmid pPLZ-1600 described below) contained not only the expected restriction fragments but also a 1,600-bp fragment (i.e., a fragment of the same size as the fragment into which Tn5 had inserted). Possibly, Tn5 was precisely deleted from part of the total plasmid population. This may explain the difference between our current results for strains containing

\* Corresponding author.

<sup>†</sup> Present address: Cetus Corporation, Emeryville, CA 94608.

pPLZ-12::Tn5 and those obtained earlier (29). Expression assay data for these strains were from bacteria containing plasmids having no detectable deletion of Tn5 (pPLZ-12::Tn5) or at most a 10% deletion of Tn5 (pPLZ-377::Tn5).

**Enzymes and radiolabeling procedures.** Restriction enzymes and T4 DNA ligase were obtained from New England BioLabs, Inc., and S1 nuclease and bacterial alkaline phosphatase were obtained from Bethesda Research Laboratories, Inc.; all enzymes were used as recommended by the suppliers. Previously described methods (31) were used for the 5'-end labeling of DNA fragments with [ $\gamma$ -<sup>32</sup>P]ATP (3,000 Ci/mmol; New England Nuclear Corp.) and the 3'-end labeling of DNA fragments with the DNA polymerase I large fragment (New England Nuclear)-catalyzed fill-in reaction.

**Construction of *lacZ* fusion plasmids.** The *EcoRI* D fragment from plasmid pES1 (see Fig. 1B) and the *EcoRI* C fragment from plasmid pES1-B8-1 (see Fig. 1B) were digested with *XmnI*. The *EcoRI-XmnI* fragments were then cloned into M13 mp8 which had been digested with *EcoRI* and *HincII*, and the recombinant phage DNAs were digested with *HindIII*. For the ca. 2-kilobase *EcoRI-XmnI* fragment derived from pES1, the *HindIII* end was filled in, digestion was completed with *EcoRI*, and the fragment was subcloned into the *EcoRI-SmaI* site of plasmid pMC1403 to construct plasmid pPLZ-1600. For the ca. 2-kilobase *EcoRI-XmnI* fragment obtained from pES1-B8-1, the ends of the fragment were filled in, and the fragment was subcloned into the *SmaI* site of plasmid pMC1403 to construct plasmid pPLZ-145::Tn5. In both cases, the transcriptional and translational control signals of the crystal protein gene were used to synthesize hybrid  $\beta$ -galactosidases which contained 15 amino acids coming from the crystal protein gene and the phage DNA.

**Construction of 5'-end deletions of the crystal protein promoter region.** Plasmid pPLZ-422 was made by digesting the promoter fragment of pPLZ-1600 with *AccI*, filling in the restriction site by using DNA polymerase I and deoxynucleotide triphosphates, completing the excision of the promoter fragment with *BamHI*, and ligating the resulting fragment into pMC1403 which had been previously digested with *SmaI* and *BamHI*. Deletions in this plasmid from -422 toward the promoter were carried out by digestion with *EcoRI* and sequential treatment with exonuclease III and S1 nuclease essentially as described by Henikoff (8), except that the digestions were performed at 20°C to slow the rate of digestion by exonuclease III to obtain deletions of 50 to 450 bp. Following phenol extraction, ether extraction, and ethanol precipitation, the nuclease-treated plasmids were digested with *BamHI* and the digestion products were applied to a 0.7% agarose gel. Sections of the gel corresponding to shortened promoter fragments were purified from the gel and ligated into pMC1403 which had been digested with *SmaI* and *BamHI*. Fusion clones containing plasmids with less extensive deletions were detected on lactose-MacConkey agar (14). Clones with plasmids containing more extensive deletions and which were later found to be missing portions of the crystal protein promoter were identified by performing restriction enzyme digestions on small-scale plasmid preparations. The exact 5' deletion endpoints of all of the deleted constructions were determined by subcloning into M13 mp9 and DNA sequencing.

pPLZ-422A was constructed by using a three-way ligation reaction containing the *EcoRI-BamHI* vector fragment from pMC1403, the *PstI-BamHI* digestion products of a chloramphenicol-resistant derivative of pUC13 (provided by Martin

Yanofsky, Department of Microbiology, University of Washington), and the *EcoRI-PstI* promoter fragment from pPLZ-422. The products of this ligation reaction were transformed into *E. coli* MC1000, and colonies resistant to ampicillin but sensitive to chloramphenicol were retained for further analysis. The same strategy was used to attempt the construction of a similar derivative of pPLZ-145::Tn5, except that the *EcoRI-PstI* promoter fragment was from pPXE-145::Tn5 (described below) rather than from pPLZ-422.

**Construction of fusions of the crystal protein promoter to the *xylE* gene.** Since the promoter-cloning site of the *xylE*-containing vector pTG402 (32) is incompatible with the current set of restriction fragments containing the crystal protein promoter (which require an *EcoRI* site at the 5' end and either a *PstI* site or a *BamHI* site at the 3' end), the fusion vector was reconstructed. The *SalI* partial digest fragment of pTG402 containing the *xylE* gene was purified and cloned into pUC9. Clones bearing the *xylE* gene in the appropriate orientation were detected by reaction with 0.5 M catechol sprayed on agar plates. The *xylE* gene of the resultant pUC9 derivative was obtained by digestion with *SmaI* and *PstI* and was fused to the crystal protein promoter in plasmid pPL703E-422. The promoter in the latter construction came from pPLZ-422 which had been subcloned into pPL703E, a transcription fusion vector (7) which we modified to replicate in *E. coli*. The crystal protein promoter-*xylE* fusion was achieved by digesting pPL703E-422 with *XbaI*, filling in the site with DNA polymerase I and deoxynucleotide triphosphates, digesting with *PstI*, and purifying the largest restriction fragment. This fragment was then ligated to the *xylE*-containing fragment with *PstI* and *SmaI* ends, described above. The resulting plasmid could be used to reclone the various promoter deletions as *PstI-EcoRI* fragments. Since transcription of the segment of pPL703E upstream from the promoter-cloning site is not well characterized in *E. coli*, the crystal protein promoter-*xylE* fusions were recloned as *EcoRI-BamHI* fragments into pBR322 (which had also been cut with *EcoRI* and *BamHI*) to put them in the same transcriptional context as the *lacZ* fusions in pMC1403. The resulting plasmids were designated pPXE (for promoter *xylE*) with a number indicating the position, in base pairs, of either a Tn5 insertion or the 5' end of the promoter fragment.

The construction of pPXE-145::Tn5 required modification of the crystal protein promoter-containing fragment from pPLZ-145::Tn5 to remove the upstream *PstI* site. The promoter fragment from the *XhoI* site of Tn5 (see Fig. 1C) to the *PstI* site downstream from the promoter (see Fig. 2A) was purified and ligated into the *SalI* and *PstI* sites of pPL703E that were derived from the polycloning site of pUC8 (18). The resulting construction contained an *EcoRI* site 18 bp upstream from the *SalI-XhoI* fusion site and a *PstI* site downstream from the promoter. This promoter fragment could be used to construct *xylE* fusion plasmids as described above.

**DNA sequencing analysis.** The nucleotide sequence of the promoter region of the crystal protein gene was reported earlier (24, 31). The insertion site of Tn5 in plasmid pPLZ-377::Tn5 was determined by *HinfI* digestion of the *XhoI-BamHI* restriction fragment of the recombinant plasmid containing the crystal protein gene promoter and the Tn5 end sequence (*HinfI* cuts at bp 4 of Tn5[1]). The resulting *HinfI* fragments were labeled at the 3' termini by using the Klenow fragment of DNA polymerase; they were digested with *HpaI* and sequenced by the method of Maxam and Gilbert (12). The insertion site of Tn5 in plasmid pPLZ-12::Tn5 was

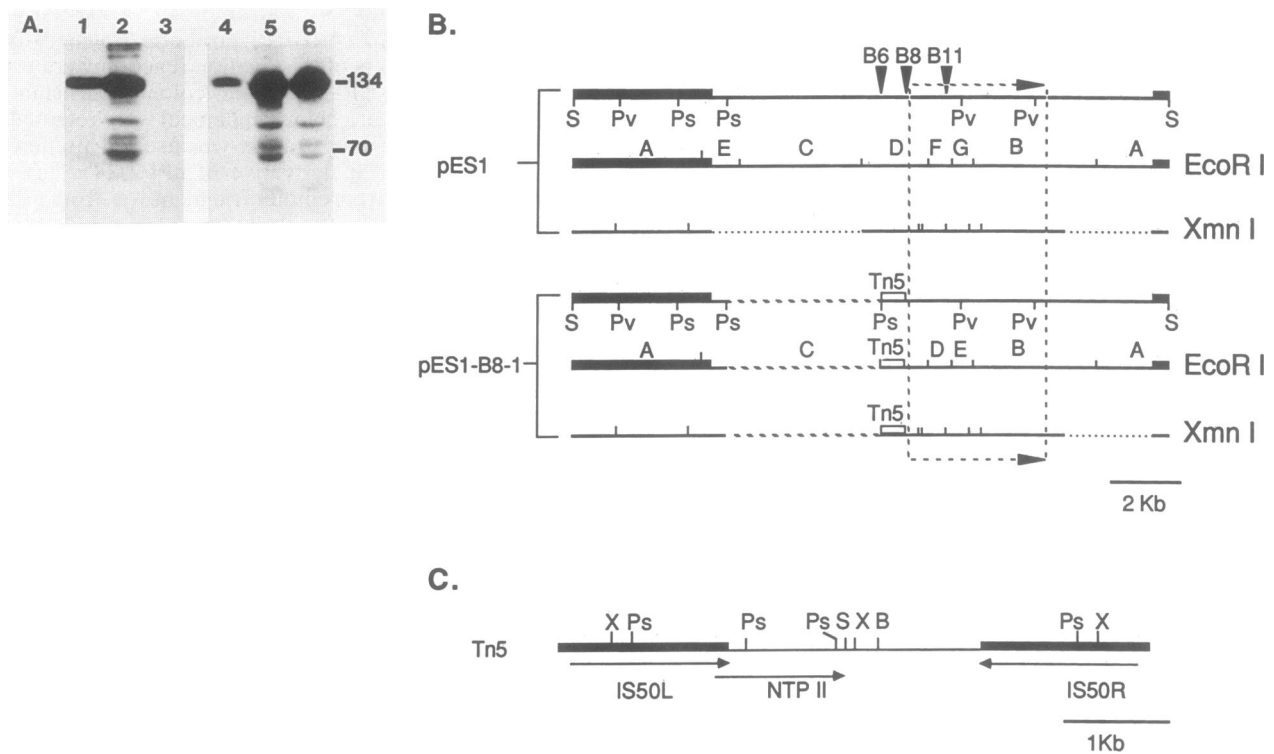


FIG. 1. Effect of Tn5 insertions on crystal protein gene expression and restriction enzyme maps. (A) Immunoblot of extracts of *E. coli* containing the following plasmids: lane 1, pES1-B6; lanes 2 and 5, pES1-B8; lane 3, pES1-B11; lane 4, pES1; and lane 6, pES1-B8-1. As described in the text, all of the above-listed plasmids except pES1 contain insertions of all or part of Tn5; the positions of these insertions are indicated in panel B. Numbers at the right are kilodaltons. (B) Restriction enzyme maps of pES1 and pES1-B8-1. Vertical arrowheads indicate sites of Tn5 insertions, the dotted box indicates the position of the crystal protein gene mRNA, and the horizontal arrowheads indicate the direction of transcription. The thick horizontal lines denote pBR322 vector sequences. For pES1-B8-1, the dashed portion of the map represents deleted DNA, and the boxed portion indicates residual Tn5 sequences. The *XmnI* maps of pES1 and pES1-B8-1 are incomplete; the dotted portions of the *XmnI* maps represent the unmapped segments. (C) Physical map of Tn5. The thick horizontal lines represent the IS50 inverted repeats; the arrows indicate the transcription units. NTP II indicates the location of the gene for neomycin phosphotransferase type II. Abbreviations: S, *Sall*; Pv, *PvuII*; Ps, *PstI*; X, *XhoI*; B, *BamHI*; Kb, kilobases.

determined in a similar way by sequencing from the *BamHI* site at the  $\beta$ -galactosidase junction into the crystal protein promoter region. The location of the Tn5 insertion in pPLZ-145::Tn5 was determined by the chain termination procedure (21) starting from the M13 DNA at the  $\beta$ -galactosidase-crystal protein junction.

**RNA purification and S1 nuclease mapping.** Previously described methods (31) were used to purify RNAs from *E. coli* strains carrying plasmids pPLZ-1600 and pPLZ-145::Tn5 and to perform S1 nuclease mapping. The 2-kilobase *EcoRI*-*BamHI* fragment from pPLZ-145::Tn5 which had been previously 5'-end labeled with [ $\gamma$ - $^{32}$ P]ATP and polynucleotide kinase at the *BamHI* site was used as the DNA probe.

**$\beta$ -Galactosidase assays.** *E. coli* strains carrying the fusion plasmids were grown at 37°C with vigorous shaking in 5 ml of L broth containing 30  $\mu$ g of ampicillin per ml to an  $A_{600}$  of  $0.7 \pm 0.1$  and were harvested by centrifugation. Cells disrupted by sonification (14) were used for assays of  $\beta$ -galactosidase. Enzyme activities are expressed in terms of nanomoles of *o*-nitrophenol formed per minute per milligram of protein. Protein was determined by the procedure of Bradford (4).

**Miscellaneous techniques.** Previously described methods were used for the sodium dodecyl sulfate-polyacrylamide gel immunoblot assay (23), transformation (11), and the deter-

mination of plasmid copy number (15, 26). Catechol oxidase was assayed as described by Ray et al. (19).

## RESULTS

**Isolation of a Tn5-overproducing strain and fusion of the crystal protein gene promoter region to *lacZ*.** Our interest in a more detailed analysis of the expression of a crystal protein gene cloned in *E. coli* was stimulated by the isolation of a Tn5 insertion mutant, B8, which synthesized 5 to 10 times more 134-kilodalton crystal protein than did our original clone (Fig. 1A). When the plasmid bearing this insertion mutation was introduced into some strains of *E. coli*, sufficient crystal protein was synthesized to observe crystals resembling those produced in *B. thuringiensis* (Whiteley et al., in press).

As reported previously (31), the position of the crystal protein gene in recombinant plasmid pES1 (23) was determined by insertion mutagenesis with transposon Tn5. The positions of three of the insertions near the 5' end of the crystal protein gene are shown in Fig. 1B. Data presented in the immunoblot in Fig. 1A indicate that when Tn5 was inserted at the position labeled B8 in Fig. 1B, more crystal protein, the 134-kilodalton polypeptide, was synthesized than in the original recombinant strain (compare lanes 2, 4,

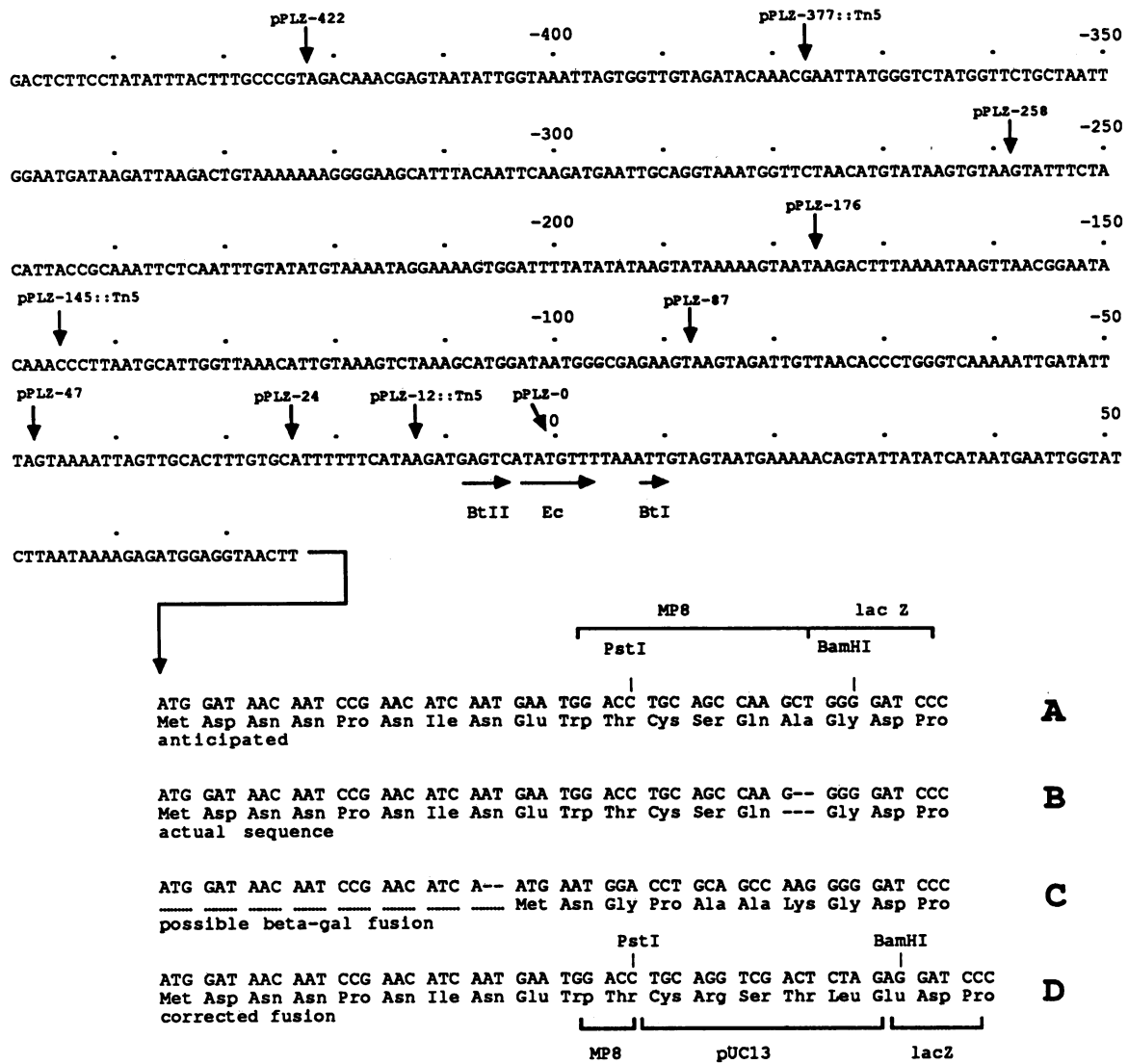


FIG. 2. DNA sequence of the 5'-flanking region of the crystal protein and fusions to *lacZ*. Vertical arrows denote sites of Tn5 insertions or the 5' endpoint of a deletion in the indicated plasmid. Horizontal arrows indicate the transcription start sites for *B. thuringiensis* (BtI and BtII) and *E. coli* (Ec). Sequences A to D show the beginning of the crystal protein-*lacZ*-coding region: (A) anticipated DNA and amino acid sequences; (B and C) determined DNA sequences showing the position of the frameshift in the corresponding amino acid sequence (B) or a potential alternative translation initiation point for the fusion protein (C); (D) corrected reading frame of pPLZ-422A. Portions of the sequences derived from the synthetic linker sequences of phages M13 mp8 and pUC13 and the *lacZ* portion of plasmid pMC1403 are indicated with brackets; the first 555 bases are from the amino-terminal-coding and 5'-flanking sequences of the *B. thuringiensis* crystal protein gene. beta-gal,  $\beta$ -Galactosidase.

and 5 of Fig. 1A) or if the insertion was farther upstream at the position labeled B6 (lane 1 of Fig. 1A). A Tn5 insertion at position B11 (lane 3 of Fig. 1A), within the crystal protein gene, resulted in no detectable crystal protein synthesis. A deletion derivative of pES1-B8 designated pES1-B8-1, which retained only 680 bp of Tn5, was still capable of producing the same high level of the crystal protein (lane 6 of Fig. 1A; Fig. 1B) as a strain containing pES1-B8; this result indicated either that some feature of the end of the transposon was required for overexpression of the crystal protein gene or that the transposon at position B8 had interrupted a sequence responsible for reducing its expression.

Comparisons were made of the copy numbers of plasmids pES1 and pES1-B8 as well as of the copy numbers of fused versions of these plasmids (plasmids pPLZ-1600 and pPLZ-

145::Tn5, in which the fusion is to *lacZ*, and plasmids pPX-422 and pPX-145::Tn5, in which the fusion is to *xylE*; these plasmids are described below). The copy number of each pair of plasmids was found to be approximately the same, thus ruling out gene dosage differences as a cause of overproduction.

**Fusion of the crystal protein gene promoter to *lacZ* and isolation of additional Tn5 mutants.** To analyze sequences regulating crystal protein transcription, we fused wild-type and mutant crystal protein promoters to the promoterless *lacZ* gene in pMC1403 (5). The previously determined DNA sequence of the promoter region of the crystal protein gene (24, 31) is shown in Fig. 2, and the positions of the transcription start sites in *B. thuringiensis* (BtI and BtII) and in *E. coli* (Ec) are indicated. Fusions were made by using the *EcoRI* D

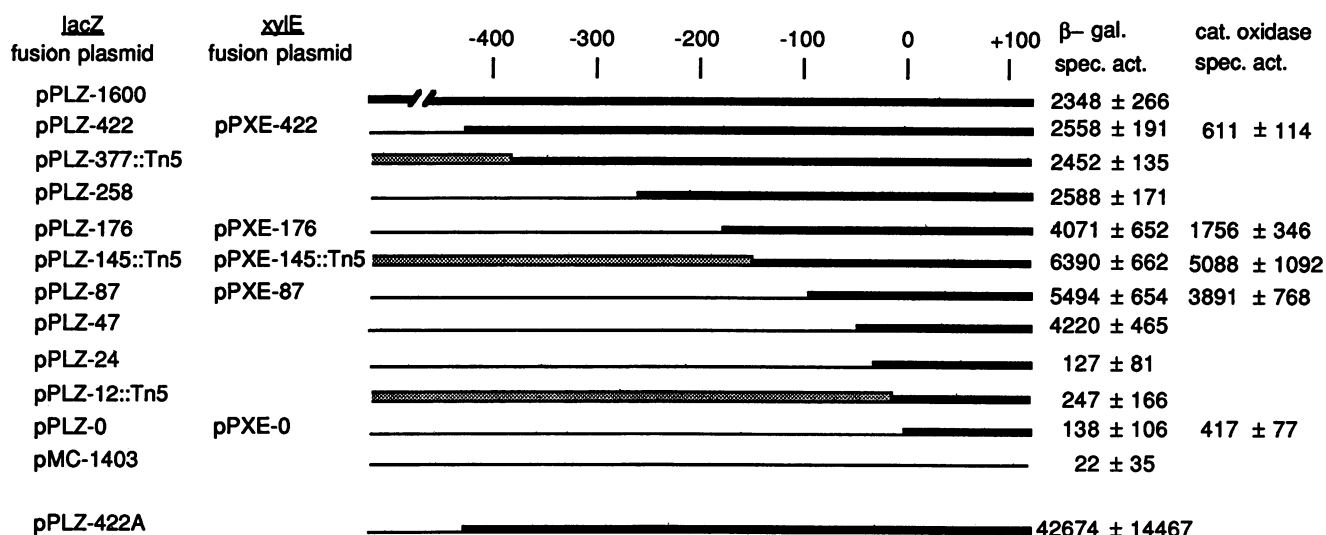


FIG. 3. Effect of Tn5 insertions and 5'-end deletions on the  $\beta$ -galactosidase ( $\beta$ -gal.) and catechol oxidase (cat. oxidase) activities of strains containing a crystal protein gene promoter fused to *lacZ* or *xylE*. Numbers over vertical lines indicate the sequence position relative to the *Ec* transcription start site. Thick black lines indicate crystal protein gene sequences fused to *lacZ* or *xylE*; thin black lines denote a segment of pBR322 DNA present in the *lacZ* and *xylE* fusions (the same pBR322 sequences are present in both); shaded bars represent Tn5 sequences. The junction is always to the *EcoRI* or adjacent *SmaI* site of pMC1403 (present in both *xylE* and *lacZ* fusion plasmids). Specific activities (spec. act.) are given in nanomoles of *o*-nitrophenol formed per minute per milligram of protein for  $\beta$ -galactosidase and in nanomoles of hydroxymuconic semialdehyde formed per minute per milligram of protein for catechol oxidase.

fragment of pES1 to construct pPLZ-1600 and the *EcoRI* C fragment of pES1-B8-1 to construct pPLZ-145::Tn5 (see Fig. 1B for the location of the restriction fragments). The fragments used in the fusions included sequences beginning several hundred base pairs upstream from the promoters, the promoters themselves, and the first 10 codons of the crystal protein gene. (The fusion point is located at an *XmnI* site which is 243 bp from the downstream *EcoRI* site of the *EcoRI* D and *EcoRI* C fragments noted above; see Materials and Methods for details of the constructions). A short segment of the M13 mp8 linker sequence was used to connect the crystal protein and  $\beta$ -galactosidase reading frames (Fig. 2A). Fusion plasmids were labeled pPLZ (for promoter *lacZ*) with a number indicating either the position in base pairs of a Tn5 insertion or the 5' end of the promoter fragment in base pairs relative to the *E. coli* transcription start site.

Recombinant strains containing the fusion plasmids produced reasonable amounts of  $\beta$ -galactosidase (Fig. 3), and the strain with the Tn5 mutant promoter (pPLZ-145::Tn5) showed a 2.6-fold increase in synthesis as compared with the strain with the wild-type promoter (pPLZ-1600). The differences in the effect of Tn5 on the synthesis of the crystal protein (Fig. 1) versus the effect on  $\beta$ -galactosidase synthesis (Fig. 3) will be addressed below. Two additional insertion mutants with Tn5 near the promoter in pPLZ-1600 were obtained by transducing *E. coli* containing pPLZ-1600 with  $\lambda$ ::Tn5. The location of the insertions (at positions -377 and -12) and the position of the Tn5 insertion in strain B8 (at position -145) are shown in Fig. 2 and were determined by DNA sequencing. DNA sequencing (data not shown) also provided evidence that there were no additional DNA rearrangements or other mutations in the sequence upstream from the transcription start site which might have an effect on transcription. Extracts of the strain containing pPLZ-377::Tn5 had the same  $\beta$ -galactosidase activity as did ex-

tracts of the strain containing pPLZ-1600, while extracts of the strain containing pPLZ-12::Tn5, which had Tn5 inserted in the promoter region, synthesized only ca. 10% the amount of  $\beta$ -galactosidase as did extracts of *E. coli* containing pPLZ-1600 (Fig. 3).

**S1 nuclease mapping of transcripts.** One explanation for the high level of expression observed in strains containing the Tn5 insertion pPLZ-145::Tn5 is that an outward promoter from Tn5 or a promoter created by Tn5 insertion could contribute to the expression. We therefore used S1 nuclease mapping to determine if extracts of this mutant contained RNAs originating in or near Tn5. RNAs from cells containing pPLZ-1600 (Fig. 4, lane 1) and pPLZ-145::Tn5 (Fig. 4, lane 2) had the same set of 5' RNA ends and mapped to the *E. coli* start site identified previously (31). However, the current level of resolution indicated that the first two bands, bases -3 and -2, were the most intense; they are likely to be the most frequently used transcription initiation sites. The lack of additional, higher-molecular-weight bands indicated that no additional 5' RNA ends originated in Tn5; additional S1 nuclease mapping experiments with continuously labeled single-stranded probes (data not shown) have confirmed this result.

**Deletion analysis of the crystal protein gene promoter.** A second way in which Tn5 could activate crystal protein gene expression is by inactivating a negative regulatory region. To determine the effect of *B. thuringiensis* sequences upstream from the crystal protein gene on transcription in *E. coli* and to further localize the promoter itself, we constructed derivatives of pPLZ-1600 bearing 5'-end deletions to endpoints from -422 through the *E. coli* transcription initiation site. Extracts of *E. coli* carrying plasmids with deletions to bases -422 and -258 displayed the same  $\beta$ -galactosidase activity as did extracts of a strain containing pPLZ-1600 (Fig. 3). However, deletions to bases -176, -87, and -47 resulted in levels of  $\beta$ -galactosidase expression that were, respectively,

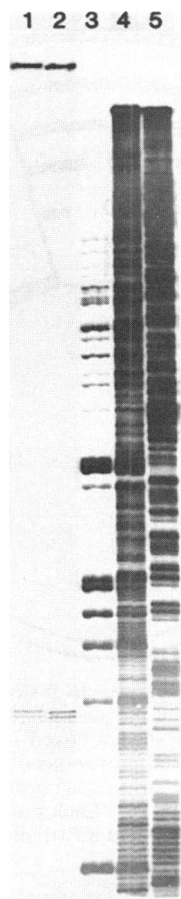


FIG. 4. S1 nuclease analysis of hybrid crystal protein gene- $\beta$ -galactosidase transcription. The 2-kilobase *EcoRI*-*Bam*HI fragment from pPLZ-145::Tn5 which had been 5'-end labeled at the *Bam*HI site (position 128 of Fig. 2) was hybridized with the indicated RNA samples, treated with S1 nuclease, and analyzed on a DNA sequencing gel. RNA was from *E. coli* containing pPLZ-1600 (lane 1, 200  $\mu$ g of RNA) or pPLZ-145::Tn5 (lane 2, 45  $\mu$ g of RNA). Base-specific chemical cleavages of the *Bam*HI-*Acc*I fragment from pPLZ-1600 (positions 128 to -422 of Fig. 2) which had also been 5'-end labeled at the *Bam*HI site are shown in lanes 3 (guanine reaction), 4 (guanine-plus-adenine reaction), and 5 (cytosine-plus-thymine reaction).

1.6-, 2.2-, and 1.7-fold the level observed in the strain containing pPLZ-1600. Additional deletions to bases -24 and 0 resulted in drastically reduced  $\beta$ -galactosidase synthesis, as might be expected if the promoter had been interrupted.

S1 nuclease mapping had been used previously (31) to localize the 5' end of the crystal protein mRNA in *E. coli*. This method did not formally rule out the possibility that the detected 5' RNA ends were generated by the processing of RNA from an upstream promoter, as opposed to transcription initiation. The results shown in Fig. 3 demonstrate that efficient  $\beta$ -galactosidase synthesis, and hence efficient transcription initiation, required the sequence between -47 and the transcription initiation site identified earlier by S1 nuclease mapping.

**Fusion of the crystal protein gene promoter to *xylE*.** A comparison of the intensity of the bands in Fig. 1A suggested

that *E. coli* containing pES1-B8 or pES1-B8-1 (lanes 2, 5, and 6) synthesized approximately 10 times more 134-kilodalton polypeptide (the crystal protein) than did the original strain containing pES1 (lane 4). This conclusion was confirmed by comparison of the relative amounts of crystal protein mRNA in the parent and overproducer strains by dot hybridization (data not shown). However, when the promoters of these two strains were fused to *lacZ* (pPLZ-422 and pPLZ-145::Tn5; Fig. 3), approximately a 2.6-fold difference was found in the  $\beta$ -galactosidase levels. Sequencing of the crystal protein-*lacZ* junctions of pPLZ-422 and pPLZ-145::Tn5 showed that although there were no differences in the promoters or upstream sequences, both of these promoter fragments, after fusion to *lacZ*, had a 2-base deletion in the coding region (Fig. 2B). This deletion would lead to translation termination at a UGA codon beginning at the third position of codon 15 of *lacZ* (9). Examination of the first 10 codons of the crystal protein gene indicated that there was an AUG codon (Fig. 2C) in-phase with  $\beta$ -galactosidase. However, translation from this site would be expected to be inefficient, since initiation from the out-of-phase crystal protein start codon might interfere with productive translation initiation, which leads to the synthesis of the fusion protein. Alternatively, "shifty" codons (27) in the sequence before the stop codon may allow for a low-frequency correction of the *lacZ* reading frame initiated from the crystal protein start codon.

Attempts were made to create in-phase fusions in pPLZ-422 and pPLZ-145::Tn5 by modifying the M13-derived connecting sequence as shown in Fig. 2D (see Materials and Methods). The in-phase fusion variant of pPLZ-422, pPLZ-422A, was obtained readily; however, repeated attempts to obtain the in-phase fusion variant of pPLZ-145::Tn5 were unsuccessful.  $\beta$ -Galactosidase constituted a large percentage of the total cell protein (14%, assuming a specific activity of 300,000 for pure  $\beta$ -galactosidase [14]) when the wild-type crystal protein promoter was fused in-phase to *lacZ* (Fig. 3). Thus, the crystal protein promoter can be used more efficiently in *E. coli* than was originally inferred by the level of crystal protein expression (23). The difference between the amounts of  $\beta$ -galactosidase and crystal protein produced in *E. coli* from this promoter may reflect differences in transcription termination, mRNA stability, or translation elongation.

Since  $\beta$ -galactosidase was produced so efficiently in cells containing pPLZ-422A, it is possible that *E. coli* is unable to synthesize still more of the fusion protein or the *lacY* gene product, from the pPLZ-145::Tn5 crystal protein promoter; indeed, such continued synthesis could be lethal to the cells. Restriction enzyme analysis of preparations of pPLZ-422A revealed that 5 to 10% of the digest consisted of anomalously sized fragments (data not shown), indicating that even this construction was not completely stable. This lack of stability may also account, in part, for the variability in the amount of  $\beta$ -galactosidase activity obtained from *E. coli* containing pPLZ-422A, as suggested by the high standard deviation in Fig. 3.

The correspondence between  $\beta$ -galactosidase synthesis and transcription may not be directly proportional in the out-of-phase translation initiation constructions; however, remaking of all of the fusions in-phase to *lacZ* would be impractical, since they would be mostly unstable or lethal. Since fusions could be obtained between the crystal protein promoter from pPLZ-145::Tn5 and a promoterless catechol oxidase gene (*xylE* [32]) (Fig. 5) containing its own ribosome-binding site, several representative mutant promoters were

fused to the *xylE* gene to supplement the out-of-phase  $\beta$ -galactosidase data. The strain containing pPXE-145::Tn5 synthesized ca. 8.3 times more catechol oxidase than did *E. coli* containing pPXE-422, while strains bearing pPXE-176 and pPXE-87 synthesized 2.9- and 6.4-fold more, respectively (Fig. 3). *E. coli* containing pPXE-0 produced about 0.7-fold the amount of catechol oxidase produced by a strain containing pPXE-422. Since the latter level of expression is higher than would be expected on the basis of the  $\beta$ -galactosidase data presented above, it is possible that the *xylE* fragment itself contains a promoter or that the fusion between the crystal protein gene and the *xylE* gene generates a promoter. Overall, the *xylE* fusion data more closely parallel estimates of the differences in crystal protein production by strains containing the original plasmid (pES1) or an overproducer plasmid (pES1-B8 or pES1-B8-1), on the basis of immunologic methods (Fig. 1), than do the  $\beta$ -galactosidase fusion data presented above.

## DISCUSSION

**Defining the promoter region.** As shown earlier (31), transcription of the *B. thuringiensis* crystal protein gene in *E. coli* is initiated from a start site (Ec) located between the two start sites (BtI and BtII) used in *B. thuringiensis*. The present deletion analysis demonstrated that transcription from the crystal protein promoter used in *E. coli* required the presence of sequences located closer than 47 bp to the transcription start site determined by S1 nuclease mapping. Inspection of the sequence between bases 0 and -47 revealed sequences moderately homologous to the -10 and -35 regions recognized by  $\sigma^{70}$ -containing *E. coli* RNA polymerase centered at positions -13 (four of six bases) and -35 (three of six bases). The putative promoter sequence defined by these homologies was analyzed by using the scoring algorithm of Mulligan et al. (16), which gives an approximate prediction of in vitro promoter strength. The score obtained from this analysis (58.05) was similar to that of the  $\lambda p_L$  or Tn5 *neo* promoters (58.0) but lower than that of the strong promoters A1, A2, and A3 used early during bacteriophage T7 infection (74 to 72.8). This result indicates that the promoter could be moderately strong in vitro and might function well in vivo. The results of in vitro transcription (H. R. Whiteley and K. L. Brown, unpublished data) with an *E. coli* RNA polymerase preparation containing  $\sigma^{70}$  (and perhaps other  $\sigma$  factors) indicated that the enzyme can initiate transcription at the Ec start site. However, we have not yet rigorously demonstrated that it is  $\sigma^{70}$ -containing RNA polymerase that initiates transcription at this start site.

**Effect of upstream deletions.** Deletions and insertions in sequences upstream from the crystal protein gene promoter resulted in a significant increase in transcription from the promoter. Thorne et al. (25) also reported that deletion of sequences upstream from a cloned crystal protein gene led to an increase in its expression in *E. coli*. Our current results indicated that the deletion of sequences between -258 and -87 relative to the *E. coli* start site increased transcription from the *E. coli* promoter by over twofold, as assayed with  $\beta$ -galactosidase fusions, and by ca. six- to eightfold, as assayed with *xylE* gene fusions or RNA dot hybridizations. An unusual feature of this region, in addition to its distance from the promoter, is that the increase in transcription was incremental: a deletion to -176 resulted in a 1.5-fold stimulation of  $\beta$ -galactosidase synthesis and a 3-fold stimulation of *xylE*, while a deletion to -87 resulted in maximal stimulation of the expression of the fusion proteins.

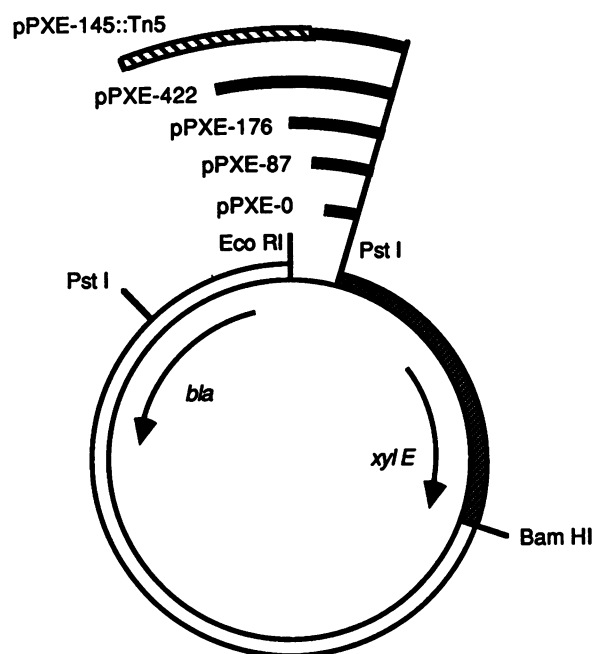


FIG. 5. Restriction maps of *xylE* fusion plasmids. The circular portion is common to all of the plasmids, the white boxed segment is from pBR322, and the shaded boxed segment is from pTG402. Black arcs show the crystal protein gene promoter portions of the indicated plasmids. The striped segment of the pPXE-145::Tn5 promoter fragment is from Tn5. Each promoter fragment is bound by an *Eco*RI site on the left and a *Pst*I site on the right.

The upstream regulatory segment contains a region of dyad symmetry between bases -258 and -176 (24), and the 50-base sequence preceding base -176 has an adenine-plus-thymine (A+T) content of 82%, as opposed to the ca. 63% A+T content for the rest of the crystal protein gene. However, *B. thuringiensis* DNA in general is significantly more A+T-rich than is *E. coli* DNA (37% guanine plus cytosine [G+C] versus 50% G+C, respectively). The role that these features play in the observed regulation of transcription is not known. Also, there are no significant open reading frames in this region large enough to encode, for example, a repressor. We do not yet know whether the transcription-suppressing region described in the current report can function at various distances from the promoter or in the opposite orientation; however, our ability to accurately transcribe the promoter in vitro may allow us to determine if the transcription-suppressing effect can be accounted for solely by interactions between RNA polymerase and DNA.

Reznikoff et al. (20) have summarized several recent reports indicating that the deletion of sequences 15 to 145 bases upstream from the -35 region of certain promoters decreases transcription from those promoters—the opposite effect of that observed in the current study. However, we did notice a decrease in expression caused by the deletion of sequences between -87 and -47 relative to the Ec start site of the crystal protein promoter, and this effect may be similar to the phenomenon noted by Reznikoff et al. (20).

Preliminary results obtained by fusing several of the same deleted crystal protein gene promoters to a chloramphenicol acetyltransferase gene indicate that the upstream region does not significantly affect transcription during vegetative growth or sporulation in *B. subtilis* (unpublished observa-

tions). While we do not understand how this transcription-suppressing region functions in *E. coli*, it seems likely that such regions could act to minimize potentially disruptive transcription from promoterlike sequences located within coding regions or in intergenic regions in operons. Well-documented examples of such latent promoter sequences have not, to our knowledge, been reported. However, several promoterlike sites with an unknown potential for promoting transcription have been identified (e.g., in the *E. coli lac* operon [17]). It is possible that the effects observed in some nonpolar Tn5 insertions (see below for examples) could have resulted from the separation of a latent promoter sequence from a transcription-suppressing sequence.

**Effect of Tn5.** The insertion of Tn5 at -145 relative to the *E. coli* transcription start site led to increased expression of the crystal protein and, in gene fusion strains, to increased expression of  $\beta$ -galactosidase and catechol oxidase. The increase in  $\beta$ -galactosidase expression was consistently higher in strains carrying pPLZ-145::Tn5 than in strains carrying plasmids containing deletions to -176; however, it was only marginally higher than in strains carrying plasmids containing deletions to -87. It is possible that a deletion to -145 would have had the same effect as a Tn5 insertion. Analysis of the transcription initiation site in strains carrying pPLZ-145::Tn5, estimated by S1 nuclease mapping, indicated that the stimulated transcription was from the same site used by the crystal protein promoter in *E. coli* and not from an outward-facing promoter of Tn5 or a promoter created by Tn5 insertion. Furthermore, if enhanced expression were due to another promoter, Tn5 insertions at -377 or -12 would be expected to yield higher levels of  $\beta$ -galactosidase, but this result was not observed.

Several other examples of nonpolar effects or activation due to the insertion of Tn5 have been reported: in the *lac* operon of *E. coli* (2), in the *nif* operon of *Rhizobium meliloti* (6), near the *kilA* gene of plasmid pKM101 (30), and in the *trfA* operon of plasmid RK2 (22). Lupski et al. (10) described a Tn5 insertion mutation in pBR322 which led to increased copy number. Their analysis of transcription in two deletion mutants indicated that a new transcript was directed out of Tn5, and although their results suggested that a promoter in Tn5 was responsible for this transcription, the transcription start site has not yet been reported. Only in the last two of these examples, however, is the precise location of Tn5 known, and none of the transcription start sites have been reported. It remains to be determined whether any of these nonpolar effects are mediated by mechanisms similar to that found in the Tn5 mutant that overproduced crystal protein.

#### ACKNOWLEDGMENTS

This research was supported by Public Health Service grant GM-20784 from the National Institute of General Medical Sciences and by grant PCM-8315859 from the National Science Foundation. H. R. Whiteley is the recipient of Research Career Award K6-GM-442 from the National Institute of General Medical Sciences.

#### LITERATURE CITED

1. Auerswald, E. -A., G. Ludwig, and H. Schaller. 1981. Structural analysis of Tn5. Cold Spring Harbor Symp. Quant. Biol. **45**: 107-114.
2. Berg, D. E., A. Weiss, and L. Crossland. 1980. Polarity of Tn5 insertion mutations in *Escherichia coli*. J. Bacteriol. **142**:439-446.
3. Birnboim, H. C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513-1523.
4. Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**:248-254.
5. Casadaban, M. J., J. Chou, and S. N. Cohen. 1980. In vitro gene fusions that join an enzymatically active  $\beta$ -galactosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for the detection and cloning of translational initiation signals. J. Bacteriol. **143**:971-980.
6. Corbin, D., L. Barran, and G. Ditta. 1983. Organization and expression of *Rhizobium meliloti* nitrogen fixation genes. Proc. Natl. Acad. Sci. USA **80**:3005-3009.
7. Duvall, E. J., D. M. Williams, S. Mongkolsuk, and P. S. Lovett. 1984. Regulatory regions that control expression of two chloramphenicol-inducible *cat* genes cloned in *Bacillus subtilis*. J. Bacteriol. **158**:784-790.
8. Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**: 351-359.
9. Kalnins, A., K. Otto, and B. Muller-Hill. 1983. Sequence of the *lacZ* gene of *Escherichia coli*. EMBO J. **2**:593-597.
10. Lupski, J. R., S. J. Projan, L. S. Ozaki, and G. N. Godson. 1986. A temperature-dependent pBR322 copy number mutant resulting from a Tn5 position effect. Proc. Natl. Acad. Sci. USA **83**: 7381-7385.
11. Mandel, M., and A. Higa. 1970. Calcium-dependent bacteriophage DNA infection. J. Mol. Biol. **53**:159-162.
12. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499-560.
13. Messing, J., and J. Vieira. 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19**:269-276.
14. Miller, J. H. 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
15. Moser, D. R., and J. L. Campbell. 1983. Characterization and complementation of pMB1 copy number mutants: effect of RNA I gene dosage on plasmid copy number and incompatibility. J. Bacteriol. **154**:809-818.
16. Mulligan, M. E., D. K. Hawley, and W. R. McClure. 1984. *Escherichia coli* promoter sequences predict in vitro RNA polymerase selectivity. Nucleic Acids Res. **12**:789-800.
17. Mulligan, M. E., and W. R. McClure. 1986. Analysis of the occurrence of promoter-sites in DNA. Nucleic Acids Res. **14**: 109-126.
18. Norrander, J., T. Kempe, and J. Messing. 1983. Construction of improved M13 vectors using oligonucleotide-directed mutagenesis. Gene **26**:101-106.
19. Ray, C., R. E. Hay, L. Carter, and C. P. Moran, Jr. 1985. Mutations that affect utilization of a promoter in stationary-phase *Bacillus subtilis*. J. Bacteriol. **163**:610-614.
20. Reznikoff, W. S., D. A. Siegel, D. W. Cowing, and C. A. Gross. 1985. The regulation of transcription initiation in bacteria. Annu. Rev. Genet. **19**:355-387.
21. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463-5467.
22. Schmidhauser, T. J., and D. R. Helinski. 1985. Regions of the broad-host-range plasmid RK2 involved in replication and stable maintenance in nine species of gram-negative bacteria. J. Bacteriol. **164**:446-455.
23. Schnepf, H. E., and H. R. Whiteley. 1981. Cloning and expression of the *Bacillus thuringiensis* crystal protein gene in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **78**:2893-2897.
24. Schnepf, H. E., H. C. Wong, and H. R. Whiteley. 1985. The amino acid sequence of a crystal protein from *Bacillus thuringiensis* deduced from the DNA base sequence. J. Biol. Chem. **260**:6264-6272.
25. Thorne, L., F. Garduno, T. Thompson, D. Decker, M. Zounes, M. Wild, A. M. Walfield, and T. J. Pollock. 1986. Structural similarity between the Lepidoptera- and Diptera-specific insecticidal endotoxin genes of *Bacillus thuringiensis* subsp. "kurstaki" and "israelensis." J. Bacteriol. **166**:801-811.



26. Weisblum, B., M. Y. Graham, T. Gryczan, and D. Dubnau. 1979. Plasmid copy number control: isolation and characterization of high-copy-number mutants of plasmid pE194. *J. Bacteriol.* **137**:635-643.
27. Weiss, R. B. 1984. Molecular model of ribosome frameshifting. *Proc. Natl. Acad. Sci. USA* **81**:5797-5801.
28. Whiteley, H. R., and H. E. Schnepf. 1986. The molecular biology of parasporal crystal body formation in *Bacillus thuringiensis*. *Annu. Rev. Microbiol.* **40**:549-576.
29. Whiteley, H. R., H. E. Schnepf, H. C. Wong, and W. R. Widner. 1986. Regulation of a cloned *Bacillus thuringiensis* crystal protein gene in *Escherichia coli* and *Bacillus subtilis*, p. 377-386. In A. T. Ganesan and J. A. Hoch (ed.), *Bacillus* molecular genetics and biotechnology applications. Academic Press, Inc., New York.
30. Winans, S. C., and G. C. Walker. 1985. Identification of pKM101-encoded loci specifying potentially lethal gene products. *J. Bacteriol.* **161**:417-424.
31. Wong, H. C., H. E. Schnepf, and H. R. Whiteley. 1983. Transcriptional and translational start sites for the *Bacillus thuringiensis* crystal protein gene. *J. Biol. Chem.* **258**:1960-1967.
32. Zukowski, M. M., D. F. Gaffney, D. Speck, M. Kauffmann, A. Findeli, A. Wisecup, and J.-P. Lecocq. 1983. Chromogenic identification of genetic regulatory signals in *Bacillus subtilis* based on expression of a cloned *Pseudomonas* gene. *Proc. Natl. Acad. Sci. USA* **80**:1101-1105.