Molecular Cloning and Characterization of the Genes Encoding the L₁ and L₂ Components of Hemolysin BL from *Bacillus cereus*†

PATRICIA A. RYAN,1 JAMES D. MACMILLAN,1,* AND BARBARA A. ZILINSKAS2

Department of Biochemistry and Microbiology1 and Plant Science Department,2 Cook College, Rutgers University, New Brunswick, New Jersey 08803-0231

Received 9 September 1996/Accepted 3 February 1997

*B. cereus*, a widely distributed bacterium commonly isolated from soil and food, is the causative agent of diarrheal and emetic types of food poisoning (23) as well as a number of opportunistic infections, including endophthalmitis (24). This paper deals with the enterotoxin responsible for the diarrheal food poisoning syndrome caused by strains of *B. cereus*. To further characterize the toxin, we sought to clone and sequence the genes encoding the L₁ and L₂ proteins. A genomic library was screened with polyclonal antibody to the L₁ and L₂ proteins to identify recombinant clones containing the genes. Five clones reacted with the antibody to L₂, but none reacted with the antibody to L₁. Southern hybridization analysis with oligonucleotide probes designed from the N-terminal amino acid sequences of the L₁ and L₂ proteins, in conjunction with immunoblot and nucleotide sequence analysis, revealed that the recombinant plasmid from one of the clones contained two genes, *hblC* and *hblD*, which encode L₂ and L₁, respectively. The two genes are arranged in tandem and are separated by only 37 bases. The gene which encodes the B component of hemolysin BL (*hblA*) is located immediately downstream from the gene encoding the L₁ protein. Northern blot analysis of RNA isolated from cells grown to mid-log (2.5 h) in LB broth revealed a 5.5-kb transcript which was hybridized with DNA fragments internal to, or including, the coding sequences of the B, L₁, and L₂ genes, suggesting that the clustered genes which encode the components of hemolysin BL are cotranscribed and constitute an operon.

---

**Hemolysin BL**, which is composed of a binding component, B, and two lytic components, L₁ and L₂, is the enterotoxin responsible for the diarrheal food poisoning syndrome caused by strains of *B. cereus*. To further characterize the toxin, we sought to clone and sequence the genes encoding the L₁ and L₂ proteins. A genomic library was screened with polyclonal antibody to the L₁ and L₂ proteins to identify recombinant clones containing the genes. Five clones reacted with the antibody to L₂, but none reacted with the antibody to L₁. Southern hybridization analysis with oligonucleotide probes designed from the N-terminal amino acid sequences of the L₁ and L₂ proteins, in conjunction with immunoblot and nucleotide sequence analysis, revealed that the recombinant plasmid from one of the clones contained two genes, *hblC* and *hblD*, which encode L₂ and L₁, respectively. The two genes are arranged in tandem and are separated by only 37 bases. The gene which encodes the B component of hemolysin BL (*hblA*) is located immediately downstream from the gene encoding the L₁ protein. Northern blot analysis of RNA isolated from cells grown to mid-log (2.5 h) in LB broth revealed a 5.5-kb transcript which was hybridized with DNA fragments internal to, or including, the coding sequences of the B, L₁, and L₂ genes, suggesting that the clustered genes which encode the components of hemolysin BL are cotranscribed and constitute an operon.

---

**B. cereus**, an isolate from a surgical wound infection and a high-level producer of enterotoxin, was obtained from J. M. Kramer, Public Health Laboratory Service, London, England. For DNA isolation, cells were grown in Luria-Bertani (LB) broth with shaking at 30°C for 4 h. RNA was obtained from cells grown to early log (2.5 h) in LB broth at 37°C.

**Screening of genomic library with polyclonal antisera.** A genomic library of *B. cereus* F837/76 was obtained as a gift from A. B. Kolsto (University of Oslo, Oslo, Norway). The library had been constructed by Toril Lindback (University of Oslo) in pUC19, using an endonuclease-deficient host strain, *E. coli* BK2116, a derivative of one described by Evensen (9, 14a). The library was plated on LB agar containing ampicillin (100 μg/ml), isopropyl-β-D-thiogalactopyranoside (IPTG; 1 mM), and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal; 40 μg/ml). The colonies were lifted in duplicate onto nitrocellulose membranes (Millipore Corp., Bedford, Mass.) and screened with a colony immunoblot assay (19) with polyclonal antibodies specific for L₁ and L₂, obtained from D. Beecher (University of Wisconsin, Madison) and prepared as described by Beecher and Wong (2).

---

**MATERIALS AND METHODS**

**Production of cells.** *B. cereus* F837/76, an isolate from a surgical wound infection and a high-level producer of enterotoxin, was obtained from J. M. Kramer, Public Health Laboratory Service, London, England. For DNA isolation, cells were grown in Luria-Bertani (LB) broth with shaking at 30°C for 4 h. RNA was obtained from cells grown to early log (2.5 h) in LB broth at 37°C.

**Screening of genomic library with polyclonal antisera.** A genomic library of *B. cereus* F837/76 was obtained as a gift from A. B. Kolsto (University of Oslo, Oslo, Norway). The library had been constructed by Toril Lindback (University of Oslo) in pUC19, using an endonuclease-deficient host strain, *E. coli* BK2116, a derivative of one described by Evensen (9, 14a). The library was plated on LB agar containing ampicillin (100 μg/ml), isopropyl-β-D-thiogalactopyranoside (IPTG; 1 mM), and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal; 40 μg/ml). The colonies were lifted in duplicate onto nitrocellulose membranes (Millipore Corp., Bedford, Mass.) and screened with a colony immunoblot assay (19) with polyclonal antibodies specific for L₁ and L₂, obtained from D. Beecher (University of Wisconsin, Madison) and prepared as described by Beecher and Wong (2).
Isolation of genomic DNA. A method employing a cesium chloride gradient, provided to us by A. B. Kolsto (14a), was used to isolate DNA from B. cereus.

DNA sequencing. Plasmid DNA was sequenced with the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction kit and the ABI 373 DNA sequencer (Perkin-Elmer Cetus, Norwalk, Conn.). The sequence was derived by progressive sequencing, using oligonucleotide primers (17 to 18 bases) which were designed based on the preceding section of DNA. Both strands were sequenced in their entirety, without ambiguity, and the sequence was compiled with the GeneWorks (IntelliGenetics, Mountain View, Calif.) computer package. The composite sequence was then compared for homology to other sequences contained in the GenBank EMBL database.

Preparation of recombinant proteins for immunoblot analysis. Overnight cultures of the recombinant E. coli strains suspected to contain genes encoding the L1 and L2 proteins were inoculated (0.5% [vol/vol]) into LB broth supplemented with ampicillin (100 µg/ml). The cells were grown at 37°C with shaking to an optical density of approximately 1 at 600 nm. Cells were harvested by centrifugation, washed twice in 100 ml of 10 mM Tris buffer, pH 7.4, and resuspended in 5 ml of the same buffer. The cells were lysed by sonication at 0°C for 5 min (10 30-s pulses, each followed by a 30-s rest interval), and the cellular debris was removed by centrifugation. The protein concentration of the lysate was determined with the Bio-Rad (Hercules, Calif.) protein assay reagent, using deoxycholate (0.1% [vol/vol]) which starts 221 bases downstream of the hblA site at the 3' end of hblD and a 3' primer (5'-AGTTCGTTCACTTCG-3') which starts 221 bases downstream from the EcoRI site at the 5' end of hblD and a 3' primer (5'-CAGACTGTAGATCAAGC-3') which starts 183 bases upstream from the Bio-Rad (Hercules, Calif.) protein assay reagent, using deoxycholate (0.1% [vol/vol]) which starts 221 bases downstream of the hblA site at the 3' end of hblD and a 3' primer (5'-AGTTCGTTCACTTCG-3') which starts 221 bases downstream from the EcoRI site at the 5' end of hblD. The primers were added to tubes containing 0.1 µg of genomic DNA, 50 µM concentrations of each of the four deoxynucleoside triphosphates (dATP, dGTP, dCTP, and dTTP), 1.5 mM MgCl₂, Taq polymerase buffer (10 mM Tris-HCl, pH 9.0, containing 50 mM KCl and 0.1% Triton X-100), and 5 U of Taq polymerase (Promega). The PCR was performed in a thermocycler (Perkin-Elmer Cetus). The template was initially denatured at 95°C for 5 min and was then amplified by 33 cycles as follows: denaturation at 95°C for 1 min, annealing at 56°C for 1 min, and extension at 72°C for 2 min; in the final cycle, the extension time was increased to 10 min. The DNA obtained from the amplification reaction was electrophoresed on preparative-scale agarose gels, and the DNA corresponding to the predicted size (approximately 400 bases) was purified with the QIAEX II gel extraction kit. Both strands of the PCR product were sequenced in their entirety as described above, and the primers used for sequencing were the same two primers used in the amplification process.

Northern blot analysis. Total RNA from B. cereus was isolated by the method of Chomczynski and Sacchi (8) as modified by Philbrick et al. (16). Denatured RNA was subjected to electrophoresis on a 1.2% agarose gel which contained 2.2 M deionized formaldehyde and then transferred to Duralon-UV membranes (Stratagene, La Jolla, Calif.). For determination of molecular sizes, an RNA ladder (Gibco-BRL, Grand Island, N.Y.) was also transferred to the same membrane and stained in 0.1% methylene blue in 0.3 M sodium acetate (pH 5.5). Membranes containing RNA from B. cereus were immersed first in prehybridization solution for 12 h at 60°C and then in hybridization solution for 12 h at 60°C. Prehybridization solution consisted of 5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), SDS (0.1%), bovine serum albumin (0.2 mg/ml), Ficoll (0.2 mg/ml), and denatured calf thymus DNA (100 ng/ml; Sigma). The hybridization solution was prehybridization solution containing 2× 10⁷ cpm of one of three radiolabeled DNA probes, each one corresponding to a sequence either internal to or including at least part of the coding region of hblA, hblC, or hblD. All probes were radiolabeled by random primer labeling using [³²P]dATP with a kit from U.S. Biochemicals Cleveland, Ohio. Following hybridization, the membranes were washed four times (15 min each) at 65°C in 0.1× SSC containing 0.2% SDS and exposed to X-ray film with intensifier screens at −70°C.

Nucleotide sequence accession number. The GenBank accession number for the DNA sequence reported is U63928.

RESULTS

Immunoblotting. Screening of 20,000 colonies from a genomic library of B. cereus F837/76 DNA with polyclonal antibody to L2 resulted in the identification of 5 colonies which clearly reacted with this antibody. Immunoblot analysis of the cell lysate from one of the five clones, designated E. coli BK23, showed a protein of approximately 43.5 kDa which was reactive with the antibody to the L2 component and, surprisingly, proteins of approximately 41, 38, and 31 kDa which were reactive with the antibody to the L₁ component (Fig. 1).

Southern hybridization analysis of the recombinant plasmid from E. coli BK3R identified two restriction fragments (a...
5.5-kb EcoRI fragment and a 4-kb fragment produced by digestion of the plasmid with EcoRI and HindIII, each able to hybridize with the oligonucleotide probes for both L₂ and L₁ (data not shown). Southern blot analysis of B. cereus genomic DNA, which was digested with EcoRI and HindIII and probed with the same oligonucleotide probes, confirmed these results (data not shown). The 4-kb restriction fragment from the recombinant plasmid was selected to simplify nucleotide sequencing. This fragment was subcloned into the vector pBlue script KS+(Stratagene) and used to transform the host strain, E. coli BK2118 (11). The colonies were rescreened with the oligonucleotide probes to L₁ and L₂. Lysates from the clones that tested positively with both DNA probes were immuno-blotted and screened with the polyclonal antibodies to ensure that full-length proteins were being produced (data not shown). The plasmid from one of the subclones, E. coli RC7, detected in this manner was isolated, and both strands of the DNA were sequenced, without ambiguity, in their entirety.

**DNA sequence.** The nucleotide sequence of the cloned fragment (Fig. 2) revealed two open reading frames, one of 1,343 bases, hblC, encoding a protein with a predicted molecular mass of 50 kDa, and the other of 1,154 bases, hblD, encoding a protein with a predicted size of 41 kDa. Both sequences are preceded by probable ribosome binding sites. The deduced amino acid sequences of hblC and hblD include 32- and 30-amino-acid peptides, respectively, prior to the first residues of the N termini of L₁ and L₂, previously determined by Edman degradation (2). Both the 32- and the 30-amino-acid sequences exhibit features of the classical prokaryotic signal sequence (20), and the signal peptidase cleavage sites follow the consensus sequences as described by von Heijne (25). The presence of the signal peptides explains the difference between the predicted sizes of the recombinant L₁ and L₂ proteins as determined from the deduced amino acid sequences and those reported for the mature L₁ and L₂ proteins (2). In addition, the 41-kDa protein seen on immunoblots of cell lysate probed with the L₁ antibody has the correct size for the recombinant protein with an unprocessed signal peptide.

A search of the GenBank-EMBL database showed that the genes encoding the L₁ and L₂ proteins have no significant homology with any known genes. However, at the amino acid level, the L₁ protein shows 22% identity with the B component of hemolysin BL. When conservative substitutions are allowed, there is 43% similarity between the deduced amino acid sequences of these two proteins. The amino acids which are identical or similar appear to be randomly distributed throughout the proteins and are not conserved within any particular domains.

**Amplification of genomic DNA fragment by PCR.** With primers designed from the 3’ end of the coding sequence of hblD and the 5’ end of hblA, a 417-bp DNA fragment was amplified from B. cereus genomic DNA. This is the size of the product expected if hblA and hblD were arranged in tandem. The first 180 bases of the fragment, immediately followed by an EcoRI restriction site, are identical to the last 180 bases of the sequence, shown in Fig. 2, which include part of the hblD gene. The 221 nucleotides, which start 10 bases downstream from the EcoRI restriction site, are identical to the first 221 nucleotides at the 5’ flanking end of the hblA gene (12, 13).

**Map of hemolysin BL gene cluster.** The Southern blot data (not shown), the restriction maps generated from the nucleotide sequences of hblA (13), hblC, and hblD, and the nucleotide sequence of the PCR product that extends from the 3’ end of hblD through the 5’ end of hblA were used to construct a map depicting the arrangement of these genes (Fig. 3).

**RNA analysis.** To determine if the genes encoding the B, L₁, and L₂ proteins are transcribed as part of a polycistronic message, total RNA was isolated from an early-log-phase culture of B. cereus. Northern blots were probed with three DNA fragments internal to, or including a portion of, the coding sequences for all three proteins. This analysis revealed a single 5.5-kb transcript which hybridized with all three probes (Fig. 4). This size agrees well with that reported by Heinrichs et al. (13) for the RNA transcript encoding the B component in early-log-phase cultures of B. cereus.

**DISCUSSION**

The identity of the diarrheal enterotoxin had been unclear until Beecher and coworkers (5) proved that hemolysin BL and the enterotoxin are the same entity. All three components of this hemolysin are required for maximal fluid accumulation in ligated rabbit ileal loops; however, individual and binary combinations of components also caused some degree of fluid accumulation. The testing of recombinant proteins in this and other biological assays will establish if this accumulation is due to residual contamination with other components or if individual components are indeed enterotoxigenic. To this end, we have cloned and characterized the hblC and hblD genes, which encode the L₂ and L₁ proteins, respectively.

A single recombinant clone, expressing both the L₁ and the L₂ components, was identified. In addition to producing proteins of 41 and 38 kDa which react with the L₁ antibody, the clone also produces a reactive protein of 31 kDa which may represent a truncated gene product or may have resulted from proteolytic degradation of the L₁ component. Analysis of the nucleotide sequences of hblC and hblD shows that the genes are arranged in tandem in the genome of B. cereus. The presumptive translation initiation codon for hblD is only 37 bases downstream from the translation termination codon of hblC. Although it is difficult to assign promoter sequences to these two genes due to their high A+T content, there is not sufficient space between them to contain a promoter sequence for hblD, indicating that the two genes are cotranscribed. Additionally, there is no apparent transcriptional terminator following hblC, providing more evidence that the genes are transcribed as part of an operon.

Southern hybridization analysis of B. cereus genomic DNA (not presented) shows that the gene encoding the L₁ component (hblD) lies directly upstream from the gene encoding the B component (hblA) of hemolysin BL. To prove that all three genes are clustered, a single DNA fragment was amplified from B. cereus genomic DNA with primers designed from sequences internal to the coding regions of hblD and hblA. This fragment encodes the last 177 bases of the 3’ end of hblD and the first 222 bases of the 5’ end of the previously cloned EcoRI fragment (13) which includes the hblA gene, proving definitively that hblA is directly downstream from hblD and that all four genes (hblC, hblD, hblA, and hblB, in that order) are arranged in tandem in the genome. A single 5.5-kb RNA transcript hybridizes with DNA fragments internal to, or including the coding regions of, hblA, hblC, and hblD, suggesting that these clustered genes are cotranscribed and constitute an operon. It is likely that hblB is also transcribed as part of this polycistronic message.

In an earlier study (18), workers from our laboratory found that B. cereus culture broth contained a large extracellular protein (approximately 100 kDa, as determined by SDS-PAGE), which reacts with a monoclonal antibody specific to the B component (3). This protein, but not a protein the size of the B component (38.2 kDa), is detectable in early-log-phase...
cultures (2.5 h) of B. cereus. In later growth stages, some of this larger protein becomes degraded, and concomitantly the B component becomes detectable with monoclonal antibody (3), suggesting that this large protein might be a precursor to the B component. The Northern blot data presented here show a large transcript, produced during early log phase, which hybridizes with probes for \( hblA \), \( hblC \), and \( hblD \). Although we have evidence only that the 100-kD protein is a precursor to the B component, an RNA transcript, which is sufficiently large to encode all three proteins and which hybridizes to probes for all three genes, is produced at the same time in the growth phase as when the large protein appears, suggesting that this protein might be a precursor to more than one, or possibly all three, of the components of hemolysin BL. If this were true, one would expect that the protein would need to have a mass of 119 kDa (38.2 kDa [B component] + 37 kDa [L1] + 43.5 kDa) rather than of 100 kDa. However, the molecular weight of the large protein was calculated by SDS-PAGE, which predicts only apparent molecular weights. We also used 12.5% acrylamide gels in these studies, whereas 8.5% gels would be better for a protein of this size. Recent experiments, however,
show that the 100-kDa protein is not reactive with the polyclonal antibodies to the L1 or L2 component on immunoblots. The large protein may not include the L1 or L2 component but may instead include the B component and the hblB gene product. The production of protoxins is not uncommon, as both the tetanus toxin (17) and the botulinum toxin (21) are formed first as 150-kDa preproteptides which are then cleaved into a heavy chain (100 kDa) and a light chain (50 kDa), which are linked together by a disulfide bond.

Hemolysin BL is not the only multicomponent hemolysin of B. cereus that is encoded by genes arranged in tandem. Cereolysin AB, composed of phospholipase C and sphingomyelinase, which act synergistically to lyse erythrocytes, is encoded in this manner (10). The genes encoding these components, however, are not transcribed as part of a dicistronic message, as indicated by a transcriptional terminator present at the 3' end of the gene encoding phospholipase C, which lies upstream from the gene encoding sphingomyelinase (14).

Cloning of the hblC and hblD genes will provide a source of the L2 and L1 components that is completely free of other contaminating B. cereus proteins. Our studies of the activities of the L1 and L2 components have been hindered due to the high-level toxicity of the individual proteins to E. coli, which became apparent when we attempted to clone the genes encoding these proteins separately into this bacterium by using various high-level-expression vectors. We are now working on formulating a tightly regulated T7 polymerase-based expression system to separately clone the hblC and hblD genes without the signal sequences so that increased levels of protein expression can be achieved and problems of toxicity can be overcome. Recombinant strains for overexpression of the L1 and L2 proteins will then be constructed, and the proteins will be used in biological assays to determine definitively if the individual or binary combinations of components show enterotoxigenic activity. As the biological activities of the components of hemolysin BL are unknown, the recombinant proteins will also provide a convenient source of these proteins to be used in determining the nature of their physiological effects.

ACKNOWLEDGMENTS

We gratefully acknowledge Douglas Beecher for providing us with the polyclonal antibodies, Anne-Brit Kolsto and Toril Lindback for providing the B. cereus genomic library, and Gerben Zylstra for use of the automated DNA sequencer. We also thank Jon Heinrichs, John Cigolini, Siva Murthy, and Patrick O’Mullen for technical help and insightful discussions.

The research was supported by grant 91-03816 under the National Research Initiative Competitive Grants Program of the United States Department of Agriculture, by New Jersey state funds, and by the United States Hatch Act.

REFERENCES


FIG. 4. Northern blot analysis of B. cereus total RNA. Cells were grown to early log phase and total RNA was isolated, separated on an agarose gel, and transferred to a nylon membrane. Membrane strips were probed with a 1.8-kb Eco RI/Xbal fragment which includes the hblD gene (encoding L2), a 1.3-kb Acc I fragment which includes a portion of the hblC gene (encoding L1), and a 506-bp PvuI/Acc I fragment internal to the hbl4 coding sequence (encoding component B) (refer to Fig. 3 for positions of restriction sites). All probes were radiolabeled with 32P by random primer labeling. Molecular size (in kilobases) was determined by comparison with an RNA ladder (not shown; Gibco-BRL). The autoradiograph was scanned by the Hewlett-Packard ScanJet 4c, using ScanJet II software.
14a. Kolsto, A. B. Personal communication.