INFECTION AND IMMUNITY, Oct. 1991, p. 3673–3679
0019-9567/91/103673-07$02.00/0

Vol. 59, No. 10

Characterization of the C3 Gene of *Clostridium botulinum* Types C and D and Its Expression in *Escherichia coli*

MICHEL R. POPOFF,[1]* DANIEL HAUSER,[1] PATRICE BOQUET,[1] MELVIN W. EKLUND,[2] AND D. MICHAEL GILL[3]†

Unité des Antigènes Bactériens, Institut Pasteur, 28 rue du Docteur Roux, F-75724 Paris Cedex 15, France[1]; Utilization Research Division, Northwest Fisheries Center, Seattle, Washington 98112[2]; and Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, Massachusetts 02111[3]

Received 3 May 1991/Accepted 20 July 1991

Clostridium botulinum **type C and D strains produce exoenzyme C3, which ADP-ribosylates the Rho protein, a 21-kDa regulatory GTP-binding protein. In a previous work, we demonstrated that the C3 gene is encoded by bacteriophages C and D of *C. botulinum* by using DNA-DNA hybridizations with oligonucleotides deduced from the C3 protein N-terminal sequence. The C3 coding gene was cloned and sequenced, but its upstream DNA region could not be studied because of its instability in *Escherichia coli*. In this work, the upstream DNA region of the C3 gene was directly amplified by the polymerase chain reaction and sequenced. The C3 gene encodes a polypeptide of 251 amino acids (27,823 Da) consisting of a 40-amino-acid signal peptide and a mature protein of 211 amino acids (23,546 Da). The C3 mature protein was expressed in *E. coli* under the control of the *trc* promoter. The recombinant polypeptide obtained was recognized by C3 antibodies and ADP-ribosylated the Rho protein. The C3 gene nucleotide sequence is identical on C and D phage DNAs. At the amino acid sequence level, no similarity was found among C3, other ADP-ribosylating toxins, or tetanus or botulinal A, C1, and D neurotoxins.**

Clostridium botulinum type C and D strains produce the botulinal neurotoxins C1 and D, respectively, a cytotoxin called C2, and an ADP ribosyltransferase termed exoenzyme C3 (30). C2 toxin is enterotoxic, cytopathic, and lethal. This toxin is composed of two polypeptide chains; one is the binding component, whereas the other is the enzymatic component, which catalyzes the ADP-ribosylation of nonmuscular G-actin (1). Exoenzyme C3 is produced as a single polypeptide of 26 kDa and ADP-ribosylates a eukaryotic protein of 21 kDa (4, 30). The substrate of exoenzyme C3 has been identified as the Rho protein (2, 24), a member of the *ras* superfamily of proteins (9). The ADP-ribosylation site has been determined as Asn-41 in the RhoA protein (32). Exoenzyme C3 has been shown to be a useful tool for analyzing the function of the Rho protein (3). C3 treatment of Vero cells induces rounding up and disassembly of actin microfilaments (10). Microinjection of Rho ADP-ribosylated by C3 into cells produces similar effects (25). It seems, therefore, that one function of the Rho protein is to control the formation or stability of actin microfilaments. Very recently, exoenzyme C3 was used in the yeast *Saccharomyces cerevisiae* to localize $Rho1_p$ at the level of the Golgi apparatus and post-Golgi vesicles (23).

It has been shown that botulinal neurotoxins C1 and D are encoded by bacteriophages C and D, respectively (13, 14). The botulinal neurotoxin C1 and D genes have been cloned and sequenced (6, 19). Since *C. botulinum* type C and D strains that had lost their bacteriophages did not produce either the neurotoxin or the C3 exoenzyme, it has been suggested that C3 is also bacteriophage encoded (30). The localization on bacteriophage DNA of the C3 gene has been confirmed by DNA-DNA hybridization with an oligonucleotide deduced from the C3 N-terminal sequence, and the bacteriophage coding region of C3 has been cloned and sequenced (26). We report in the present study the characterization of the exoenzyme C3 gene from *C. botulinum* type C and D bacteriophages and its expression in *Escherichia coli*.

MATERIALS AND METHODS

Exoenzyme C3 protein sequencing. Exoenzyme C3 was purified from *C. botulinum* type D strain 1873 as previously described (30). Purified exoenzyme C3 was run on a 12% polyacrylamide gel containing 0.1% sodium dodecyl sulfate (SDS) and transferred to a polyvinylidene difluoride membrane (Immobilon; Millipore, Paris, France). After Coomassie blue staining and destaining, the C3 band was cut out and peptide sequencing was done directly from the Immobilon membrane with a gas-phase protein sequencer (Applied Biosystems).

Bacterial DNA and plasmids. *C. botulinum* type C strain 468 and *C. botulinum* type D strain 1873 were grown in broth containing Trypticase (30 g/liter), yeast extract (20 g/liter), glucose (5 g/liter), and cysteine HCl (0.5 g/liter) (pH 7.2) under anaerobic conditions. Genomic DNA was extracted and purified as previously described (27). Bacteriophage DNAs from *C. botulinum* type C strain 468 and *C. botulinum* type D strain 1873 were purified as previously described (14). Plasmid pUC19 (Appligène, Strasbourg, France) was used for cloning in *E. coli* TG1 [Δ(*lac-pro*) *thi supE hsdD5* F′ (*traD36 proAB*$^+$ Δ*lacZ* M15)], and plasmids pRX1 (29) and pKK233-2 (LKB-Pharmacia, Paris, France) were used for expression in *E. coli* RR1 and TG1, respectively.

Probes and hybridization conditions. Oligonucleotides were synthesized by the phosphoramidite method with a Cyclone Milligen automated DNA synthesizer. GeneScreen Plus filters (New England Nuclear Research Products, du

* Corresponding author.
† Deceased 30 June 1990.

Pont de Nemours, Boston, Mass.) were pretreated with 200 μg of heat-denatured salmon sperm DNA per ml in 1 M NaCl–10% dextran sulfate–0.5% SDS–50 mM Tris HCl (pH 7.5) at 40°C and then treated with a 5′-^{32}P-labeled oligonucleotide (10^6 cpm/ml) in the same mixture overnight at 40°C. Filters were washed in 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% SDS at 40°C for 2 h and exposed overnight to Fuji RX films.

Vectorette system and PCR amplification. The vectorette system, which was developed to amplify and sequence DNA fragments adjacent to a region for which the sequence is known, was used as previously described (28). In brief, the vectorette system consists of two synthetic 86-mer oligonucleotides (upper strand and lower strand), made according to the sequence published by Riley et al. (28) and containing the following restriction enzyme sites: *Sma*I, *Bam*HI, *Acc*I, *Pst*I, and *Hin*dIII. These oligonucleotides constitute a duplex with a noncomplementary region. The vectorette primer (VP) is a 30-mer oligonucleotide which is identical to a noncomplementary portion of the lower strand (28). Both the vectorette and the DNA are digested with an appropriate restriction enzyme and ligated. Amplification is performed between a primer derived from the known DNA sequence and VP. In the first cycle, amplification products are derived only from the primer hybridized to the known DNA sequence and encompass the complementary strand of the vectorette lower strand. During the subsequent cycles, VP can hybridize to the complementary vectorette lower strand and exponential DNA amplification takes place between VP and the primer derived from the known DNA sequence.

Phage C and D DNAs (300 ng) cut by *Mbo*I were ligated to 10 ng of *Bam*HI-digested vectorette in a final volume of 20 μl for 2 h at room temperature. The *Mbo*I-cut vectorette concentration was determined by checking several vectorette concentrations in the ligation reaction. Two microliters of the ligation mixture was amplified by the polymerase chain reaction (PCR) in a total volume of 100 μl of 10 mM Tris HCl (pH 8.3)–50 mM KCl–4 mM $MgCl_2$–0.1% bovine serum albumin–100 μM deoxynucleoside triphosphate–10 mM β-mercaptoethanol–25 pmol of each primer–2.5 U of *Taq* polymerase (PCR buffer) (Beckman, Paris, France). Reaction mixtures were denatured at 95°C for 2 min and then subjected to 30 cycles of denaturation (20 s at 94°C), annealling (20 s at the hybridization temperature; see Results), and extension (20 s at 72°C) in a version 2.2 DNA thermal cycler (Perkin-Elmer Cetus, Emeryville, Calif.). Amplification products were purified with Geneclean (Bio 101, Inc., La Jolla, Calif.) and sequenced.

For PCR amplification directly with phage DNA, 100 ng of phage DNA and 25 pmol of each primer were used in the same PCR buffer.

Other molecular biology techniques. Ligation and preparation of plasmid DNA from *E. coli* were conducted as described by Maniatis et al. (22). Bacteria were transformed by electroporation (11). T4 polynucleotide kinase and calf intestinal phosphatase were from Boehringer-Mannheim France, and other enzymes were from Pharmacia (Paris, France). DNA was sequenced by the dideoxy chain termination procedure (31) with a Sequenase kit (United States Biochemical Corp., Cleveland, Ohio).

Potential biohazards associated with the experiments described had been approved by the French National Control Committee.

Expression of recombinant proteins. *E. coli* TG1 transformed with recombinant plasmid pKK233-2 was grown in M9 minimal medium with 5% Casamino Acids (22) and 50 μg of ampicillin per ml to an optical density at 600 nm of 0.5. Isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 100 μg/ml, and the cells were grown for an additional 2 h. One milliliter of the cell suspension was pelleted in a Microfuge, resuspended in 100 μl of Laemmli sample buffer (21), and boiled for 3 min. Ten microliters was loaded on a 0.1% SDS–polyacrylamide gel (21).

E. coli RR1 harboring plasmid pRX1 was grown in M9 minimal medium with Casamino Acids and ampicillin as described above. Induction was performed with indolyl acrylic acid (Sigma, Paris, France) to a final concentration of 10 μg/ml, and the cells were grown for an additional 4 h (29).

For immunoblotting analysis, proteins were transferred electrophoretically to nitrocellulose (Hybond C; Amersham, Paris, France). The nitrocellulose was first incubated for 1 h in phosphate-buffered saline containing 5% milk and then incubated overnight at room temperature with a 1:400 dilution of C3 antibodies (30). Bound antibodies were detected with ^{125}I-labeled protein A and autoradiography.

Gel assay for ADP-ribosylation. *E. coli* TG1 carrying recombinant plasmid pKK233-2 was grown and induced by IPTG as described above. The bacteria were resuspended in 20 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES) buffer (pH 8) and opened by freezing-thawing (three cycles). After centrifugation in a Microfuge, the supernatant was used as a source of C3 exoenzyme. Samples (1 to 10 μl) to be assayed were incubated in a total volume of 15 μl of 1 μg of partially purified recombinant RhoC–1 mM AMP–30 μM [^{32}P]NAD (10,000 cpm/pmol). The RhoC gene was cloned from B lymphocytes and expressed in baculovirus (unpublished data). After 60 min at 37°C, gel application buffer was added and the sample was fractionated by SDS-polyacrylamide gel electrophoresis (PAGE) (15%). After being stained, destained, and dried, the gel was exposed to X-ray film with intensifying screens. Quantitation of ADP-ribosylation was done by cutting out the radioactive spots from the gel and subsequently counting.

Nucleotide sequence accession number. The nucleotide sequences reported in this paper have been submitted to the EMBL Data Library with accession numbers X59039 and X59040 for the C3 gene sequences from *C. botulinum* type C (1,977 bp) and D (1,979 bp) phage DNAs, respectively.

RESULTS

Cloning of the C3 gene. By microsequencing of the N-terminal part of purified C3, the 26 first amino acids were determined (26). Two oligonucleotides were synthesized: one (P4) corresponds to the DNA sequence deduced from the amino acid sequence from Ala-1 to Ala-17, and the other (P5) corresponds to the sequence from Trp-18 to Tyr-26, according to *Clostridium* codon usage (39) and with inosine (I) at the most degenerated positions (Fig. 1). These probes hybridized with DNA fragments from phages C and D (Fig. 2) but not with chromosomal DNA from *C. botulinum* type C strain 468 or *C. botulinum* type D strain 1873 (data not shown). In particular, P4 recognized a 0.9-kbp *Mbo*I fragment on phage C and D DNAs, and P5 recognized a 2-kbp fragment on phage D DNA and a 1.7-kbp fragment on phage C DNA digested with *Hin*dIII (Fig. 2).

We observed that restriction enzymes *Sau*3A, *Bam*HI and *Bgl*II were able to cut phage D DNA but not phage C DNA. This difference in susceptibility to *Sau*3A between phage C and D DNAs has also been reported by Fujii et al. (16). This result suggests that *C. botulinum* type C strain 468 possesses a restriction-modification system acting on the GATC se-

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
aa	A	Y	S	N	T	Y	Q	E	F	T	N	I	D	Q	A	K	A
Probe P4 5'GCI		TAT	TCI	AAT	ACI	TAT	CAA	GAA	TTT	ACI	AAT	ATI	GAT	CAA	GCI	AAA	GC 3'
b	GCT	TAT	TCA	AAT	ACT	TAC	CAG	GAG	TTT	ACT	AAT	ATT	GAT	CAA	GCA	AAA	GCT

	18	19	20	21	22	23	24	25	26
aa	W	G	N	A	IorQ	Y	K	K	Y
Probe P5 5'	TGG	GGI	AAT	GCI	III	TAT	AAG	AAG	TAT 3'
b	TGG	GGT	AAR	GCT	CAG	TAT	AAA	AAG	TAT

FIG. 1. Amino acid (aa) sequences of the N-terminal part of exoenzyme C3 and oligonucleotide probes P4 (50-mer), and P5 (27-mer) complementary to this sequence, according to *Clostridium* codon usage (see the text). The nucleotide sequence determined by DNA sequencing is shown in line b.

quence and presumably methylating the cytosine, as was previously found for *C. perfringens* (17).

Phage C and D DNA fragments cut with *Hin*dIII were ligated to *Hin*dIII-digested pUC19, and recombinant transformants were screened by colony hybridization with radiolabeled P5. Two clones recognized by P5 were selected: pMRP11 contains 2 kbp of phage D DNA and pMRP12 contains 1.7 kbp of phage C DNA (Fig. 3). pMRP11 and pMRP12 were sequenced on both DNA strands until about 0.5 kbp downstream of the two stop codons of the C3 gene (26).

Sequencing of the upstream C3 gene region. Several attempts were made to clone the upstream C3 gene region, which corresponds to the 0.9-kbp fragment recognized by oligonucleotide P4 on phage C and D DNAs cut by *Mbo*I. However, no clone could be obtained, despite the use of several *E. coli* host strains, such as TG1, TG2 (*rec*), ER1451 [*hsdR* (r_K^- m_K^+) *mcrB1*] (New England Biolabs), or Sure (Stratagene), designed specially to clone DNAs with secondary or tertiary structures. Moreover, several plasmid vectors, such as pUC19, pBR322, and pKK232-8 (Pharmacia), which is used in promoter selection, did not allow us to clone the 5′ DNA region of the C3 gene. In addition, no other restriction enzyme could be found to determine a smaller DNA fragment corresponding to the 5′ DNA region of the C3 gene.

By screening a *Hin*dIII library of phage D DNA in pUC19 and using host strain TG1, we found that 1 clone (pMRP6) of 1,500 was recognized by P4. This clone contains a deleted insert comparable in size to the *Hin*dIII phage D fragment (10 to 12 kbp) recognized by P4 in Southern blots (Fig. 2). This result indicated that the upstream DNA region of the C3 gene has a special feature which does not allow it to be cloned.

Sequencing of the pMRP6 insert showed that this fragment corresponds to the coding region of C3 from amino acids 1 to 17 (determined by protein sequencing) (Fig. 1) with 7 extra N-terminal amino acid residues.

Sequencing of the *Mbo*I fragment from phage C and D DNAs recognized by P4 was finally carried out with the vectorette system as described in Materials and Methods. The vectorette library was amplified with VP and P2 (corresponding to P4 but on the complementary DNA strand) at a hybridization temperature of 57°C, and the amplification product (P2-VP) was directly sequenced (Fig. 4) with primers P2 and allowed to synthesize a new primer (P97) located

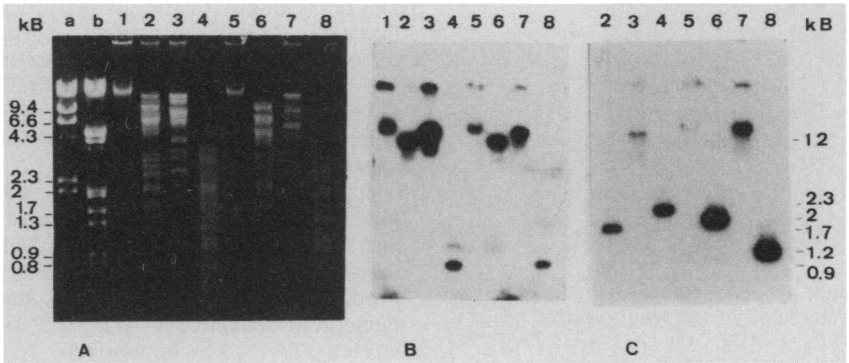

FIG. 2. Identification of the C3 gene in phage C and D DNAs. (A) Ethidium bromide-stained agarose gel of phage C DNA from *C. botulinum* type C strain 468 undigested (lane 1) and digested with *Hin*dIII (lane 2), *Eco*RI (lane 3), and *Mbo*I (lane 4) and of phage D DNA from *C. botulinum* type D strain 1873 undigested (lane 5) and digested with *Hin*dIII (lane 6), *Eco*RI (lane 7), and *Mbo*I (lane 8). Lanes a and b, lambda DNA ladder. (B and C) Southern blot analysis of the agarose gel by hybridization with 5′-labeled oligonucleotides P4 (B) and P5 (C).

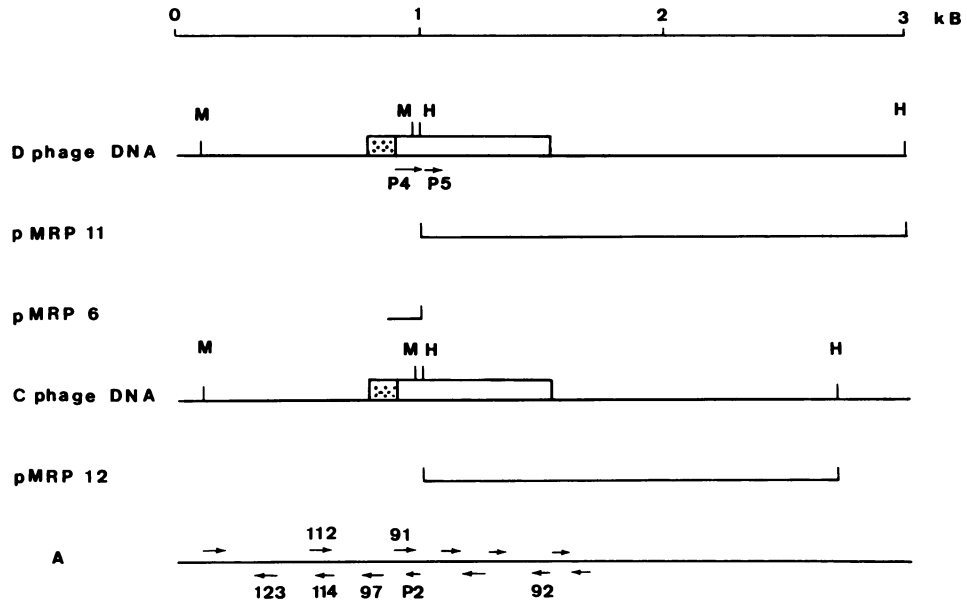

FIG. 3. Cloning of the C3 gene from phage C and D DNAs. The open box corresponds to the coding region for the C3 mature protein, and the dotted box corresponds to the leader sequence. pMRP6, pMRP11, and pMRP12 are pUC19 recombinant plasmids. The 0.9-kbp (kB) *Mbo*I fragments from phages C and D were amplified and sequenced by PCR with the vectorette system. M, *Mbo*I; H, *Hin*dIII. P4 and P5 are the oligonucleotides (Fig. 1) used for screening the recombinant plasmids. The strategy used to determine the nucleotide sequence with synthetic oligonucleotides is shown in line A.

194 bp upstream of P2. A 10-μl aliquot of the P2-VP amplification product was amplified a second time with VP and P97 at a hybridization temperature of 60°C and sequenced. Thereafter, new primers were synthesized (P114 and then P123) and used in PCR amplification and sequencing (Fig. 3 and 4). The complementary DNA strand was sequenced with synthetic oligonucleotides (Fig. 3).

The phage C DNA region corresponding to the phage D DNA fragment cloned in pMRP6 was sequenced after PCR amplification with primers P112 and P92. Similar experiments were done on phage D with the same primers, and an identical nucleotide sequence was obtained. The fact that this sequencing strategy was successful supports the contention that fragments cloned in pMRP6 and pMRP11 from phage D DNA are contiguous (Fig. 3). Figure 5 shows the alignment of the nucleotide sequences derived from phage C and D DNAs which encompass the C3 gene.

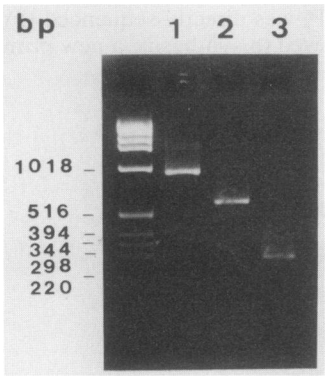

FIG. 4. Ethidium bromide-stained agarose gel demonstrating amplification of the vectorette library from *Mbo*I-digested phage C DNA with VP and (lane 1) P2 (5′-ACT CCT GGT AAG TAT TTG AAT-3′) at a hybridization temperature of 57°C. After sequencing of the amplification product, new primers were synthesized and used in PCR amplification with VP: lane 2, P114 (5′-CAT AGG TTA CTA CTT CTA TAA CCC-3′) at a hybridization temperature of 64°C; lane 3, P123 (5′-CAA GTT ATT TAC CTG AAC CAA AC-3′) at a hybridization temperature of 60°C.

Features of C3 genes. An open reading frame (ORF) covering 759 bp beginning with an AUG start codon and terminating with two stop codons (UAA and UAG) was identified. This ORF nucleotide sequence is identical in phage C and D DNAs. The deduced polypeptide is composed of 251 amino acid residues and has a predicted molecular mass of 27,823 Da. The amino acid sequence from Ala-41 to Tyr-66 is in close agreement with the sequence determined by protein sequencing of the N-terminal part of C3 (Fig. 1). These data suggest that the amino acid sequence from Ala-41 to Lys-211 (predicted molecular mass, 23,546 Da) corresponds to the C3 mature protein and that the 40 first amino acids correspond to a signal peptide. The sequence from Lys-40 to Ala-41 represents a common cleavage site for proteases, such as trypsin. The low G+C content (30.6%) and codon usage of the C3 gene conform to the pattern generally seen for genes isolated from *Clostridium* species (39).

A presumptive Shine-Dalgarno sequence (GGAGGG) was identified 5 nucleotides upstream of the AUG start codon and is closely related to the Shine-Dalgarno sequences found in other *Clostridium* species genes (39). Upstream of the proposed Shine-Dalgarno region, starting at base 652 of the nucleotide sequence of phage C DNA, a hexanucleotide stretch (TATAAT) homologous to the *E. coli* T^{55} RNA polymerase −10 consensus recognition sequence (TATAAT) (39) was found. At 21 bases upstream of the proposed −10 recognition sequence, a heptanucleotide stretch (TTGAACA on phage C DNA and TTGAATA on phage D DNA) homologous to the reported −35 consensus sequence (TTG

```
603  TTACTCATTGATATTTAATATTATTGAATAATAAATTTTTATTATAAAAATTATAATAAAGTATTTACAAAATTACAAAATATGGTAAAATAGTATATAT
1                                  M  K  G  L  R  K  S  I  L  C  L  V  L  S  A  G  V  I  A  P  V  T  S  G  M  I
703  AAATTTAGTTTAGGAGGGATTTTATGAAAGGTTTAAGAAAATCAATTTTATGTTTAGTTTTGTCAGCAGGAGTAATAGCTCCAGTAACATCTGGGATGAT
27   Q  S  P  Q  K  C  Y  A  Y  S  I  N  Q  K  A  Y  S  N  T  Y  Q  E  F  T  N  I  D  Q  A  K  A  W  G
803  TCAAAGTCCTCAAAAATGTTATGCTTATTCCATTAATCAAAAGGCTTATTCAAATACTTACCAGGAGTTTACTAATATTGATCAAGCAAAAGCTTGGGGT
```

FIG. 5. Nucleotide sequence of phage D DNA and predicted amino acid sequence of the C3 gene N-terminal part. The signal peptide amino acids are in italic type. The putative Shine-Dalgarno sequence is boxed. The presumptive −10 and −35 promoter consensus sequences are underlined.

ACA) was found (Fig. 5). The 21-nucleotide spacing between the putative −10 and −35 recognition sequences of the C3 gene is similar to the reported spacer lengths in other clostridial genes (39).

A candidate transcription terminator (ΔG, −16.5 kcal [ca. −69.0 kJ] for the C3 gene in phage D DNA; ΔG, −14.4 kcal [ca. −60.2 kJ] for the C3 gene in phage C DNA) was found 33 nucleotides downstream of the two stop codons. The inverted repeat of the putative transcription terminator encompasses 17 and 16 nucleotides on phage D and C DNAs, respectively, with one impairment of bases. The 3′-untranslated region encompassing the 223 nucleotides downstream of the two stop codons is identical in phage C and D DNAs, except for two nucleotide differences: A at 1484 and T at 1558 in phage C instead of G at 1486 and C at 1560 in phage D. However, nucleotide sequences 3′ to the G at position 1705 on phage C DNA are unrelated to those on phage D DNA. The phage C and D DNA sequences for the 725-nucleotide region upstream of the AUG start codon are very similar. Only three nucleotide differences (T at 459, G at 484, and C at 629 on phage C DNA instead of A at 459, A at 484, and T at 631 on phage D DNA), and two additional nucleotides (G at 551 and T at 552) on D phage DNA were observed on this upstream nucleotide sequence of the C3 gene.

Expression of the C3 gene in *E. coli*. To demonstrate that the ORF from phage C and D DNAs and identified with oligonucleotides P4 and P5 is the structural C3 gene, we expressed this ORF in *E. coli*. A fusion protein expression plasmid (pMRP34) derived from the tryptophan synthetase expression system pRX1 (29) and encoding TrpE amino acids 1 to 18, 18 amino acids of the pUC19 polylinker, the last 7 amino acids of the C3 signal peptide, and the coding region of the C3 mature protein under the control of the *trpE* promoter was constructed (Fig. 6). This plasmid was transferred in host strain *E. coli* RR1 by electrotransformation. PAGE showed that strain RR1 harboring pMRP34 and induced with indolyl acrylic acid expressed a peptide of 29 kDa; the same strain grown without induction did not (Fig. 7). Approximately 4 μg of C3 fusion protein per 100 μg of bacterial protein was produced. This polypeptide was recognized by antibodies raised against purified C3 from *C. botulinum*, and a cellular extract from RR1(pMRP34) induced with indolyl acrylic acid exhibited an ADP ribosyltransferase activity.

The DNA region coding for the C3 mature protein, from Ala-41 to the two stop codons, was amplified by PCR with two primers adding an *Nco*I restriction site on the 5′-terminal part and a *Pst*I restriction site on the 3′-terminal part of this DNA fragment (Fig. 6). The amplified product was cloned in expression vector pKK233-2 digested with *Nco*I and *Pst*I. The resulting recombinant plasmid was called pMRP36 (Fig. 6). The fragment cloned in pMRP36

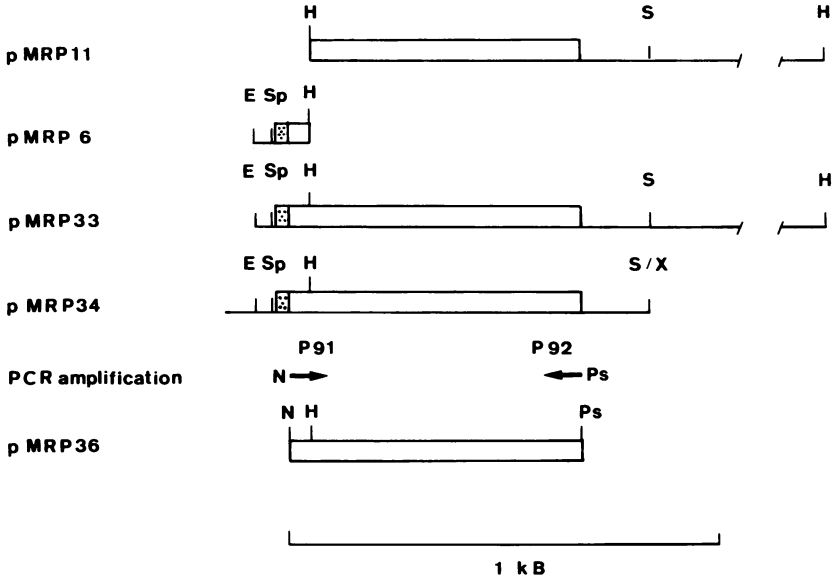

FIG. 6. Construction of C3 expression vectors. pMRP33 resulted from the insertion of the *Hin*dIII fragment from pMRP11 into pMRP6. The E-Sp region corresponds to the pUC19 polylinker. pMRP34 resulted from the insertion of the *Eco*RI-*Spe*I fragment from pMRP33 into the *Eco*RI-*Xba*I sites of pRX1. Positions 1 to 18 correspond to the 18 TrpE N-terminal amino acids. The coding region of the C3 mature protein (open box) was amplified by PCR with primers P91 (5′-CCA TGG CTT ATT CAA ATA CTT ACC AGG AG-3′) and P92 (5′-CTG CAG CTC TAT TAT TTA GGA TTG ATA GC-3′). The amplified fragment was cloned in the *Nco*I-*Pst*I sites of pKK233-2. The dotted box corresponds to the leader sequence. E, *Eco*RI; H, *Hin*dIII; S, *Spe*I; Sp, *Sph*I; X, *Xba*I; N, *Nco*I; Ps, *Pst*I.

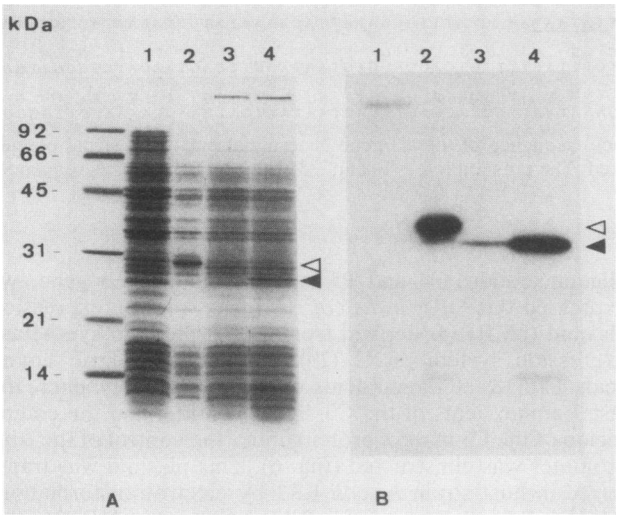

FIG. 7. Analysis of C3 expression in *E. coli*. (A) SDS-PAGE (12% polyacrylamide) of protein extracted from *E. coli* RR1 (pMRP34) without induction (lane 1) and with indolyl acrylic acid induction (lane 2) and from *E. coli* TG1(pMRP36) without induction (lane 3) and with IPTG induction (lane 4). (B) Western blot with C3 antibodies. Symbols: ◀, C3 expressed from pMRP36; ◁, truncated TrpE-C3 fusion protein.

was sequenced and had the same nucleotide sequence as that previously found for C3, except that an ATG initiation codon was located just before the codon for Ala-41 by the introduction of an *Nco*I restriction site. This ORF in pKK233-2 is under the control of the *trc* promoter. A peptide of 27 kDa was expressed at a higher level after IPTG induction in host strain *E. coli* TG1 harboring pMRP36 than in the same strain without induction (Fig. 7). This peptide was also recognized by C3 antibodies, and a cellular extract of this *E. coli* strain ADP-ribosylated the Rho protein. This expression vector construction yields approximately 2 μg of recombinant C3 per 100 μg of bacterial protein.

DISCUSSION

In *C. botulinum* type C strain 468 and *C. botulinum* type D strain 1873, the C3 gene was located on phage C and D DNAs, respectively, by hybridization with oligonucleotides (P4 and P5) deduced from the C3 N-terminal protein sequence. The C3 gene from both phages C and D was cloned. The *Mbo*I restriction fragment encompassing the promoter and the beginning of the C3 gene leader sequence is unstable in *E. coli*. The reason for this instability is not clear, since neither a complete ORF encoding a putative poisonous peptide nor significant secondary DNA structures are present in this fragment. A possibility is that the C3 promoter has strong functional activity in *E. coli*. However, if this were the case, its cloning should have been successful in pKK232-8, which contains two efficient rRNA terminators downstream of the multiple cloning site, and this was not the case. The vectorette system described by Riley et al. (28) allowed us finally to amplify and sequence the upstream part of the C3 gene without cloning.

The ORF that we have identified encodes a native peptide of 251 amino acids with a predicted molecular mass of 27,823 Da. A smaller molecular mass (26 kDa) was determined for C3 secreted by *C. botulinum* (4, 30). Moreover, the C3 N-terminal sequence showed that the mature protein starts at Ala-41. Expression in *E. coli* of the C3 gene region encoding the 211 amino acids from Ala-41 to the two stop codons in the pKK233-2 vector yielded a 27-kDa peptide in SDS-PAGE that was recognized by C3 antibodies and exhibited an ADP-ribosylating activity. These data indicate that the ORF that we have identified corresponds to the C3 structural gene.

The fusion protein construction with 18 amino acids of TrpE and C3 showed approximately twice the level of expression as the C3 mature protein only. The presence of an *E. coli* homologous peptide at the beginning of C3 probably protected the foreign protein against protease degradation. The level of recombinant C3 expression in *E. coli* is at least 10 to 20 times greater than that in *C. botulinum* type D. However, the ADP-ribosylation activity of the recombinant C3 protein is lower than that of native C3 from *C. botulinum*. When present in identical concentrations, recombinant C3 fusion protein (expressed from pMRP34) and recombinant C3 protein (expressed from pMRP36) had, respectively, 50 and 55% of the activity of C3 purified from *C. botulinum*.

The putative signal peptide of C3 encompasses 40 amino acids and shows the general features of secretory signal peptides (38) with positively charged amino acids (Lys-2, Arg-5, and Lys-6) at the extreme N terminus, a central hydrophobic core region (Ile-8 to Ile-26), a polar C-terminal region (Pro-29 to Gln-39), and a proteolytic cleavage site (Lys-40 to Ala-41) (Fig. 5). The known signal peptides in *Clostridium* species are 28 to 41 amino acids long (Fig. 5). No identity in amino acid sequences was shown between the C3 signal peptide and other clostridial signal peptides, such as those of *C. thermocellum* cellulases A, B, and D, *C. perfringens* phospholipase C and perfringolysin O, and *C. bifermentans* phospholipase C (5, 18, 20, 34, 35, 37).

The C3 nucleotide sequence is exactly the same in phage C and D DNAs. Moreover, the 223-nucleotide sequence downstream and at least the 723-nucleotide sequence upstream of the C3 gene are closely related in phages C and D, suggesting that the C3 gene could be part of a larger DNA fragment that could represent a genetic mobile element.

At the amino acid sequence level, no similarity was found between C3 and other ADP-ribosylating toxins or tetanus or botulinal A, C1, and D neurotoxins (6, 7, 12, 15, 19, 33). Up to now, it has been shown that phages in *C. botulinum* types C and D encode C1 and D neurotoxins, respectively (6, 19), the main hemagglutinin component (36), and ADP ribosyltransferase C3. The advantages of these components for *C. botulinum* are still speculative.

Molecular cloning and expression in *E. coli* will allow us to do experiments such as in vitro mutagenesis to study the structure-function relationships in ADP ribosyltransferase C3. Furthermore, the recombinant protein obtained is totally devoid of other clostridial ADP-ribosylating toxins, such as C2, and thus provides an excellent tool for studying the role of the p21 Rho protein in microfilament assembly and/or stability.

ACKNOWLEDGMENTS

We are indebted to P. Cossart and J. Mengaud for providing us with synthetic oligonucleotides. We thank F. Poysky and L. Jordan for help in preparing clostridial phage DNA and F. Godeau for providing plasmid pRX1 and *E. coli* RR1.

This work was supported by USAMRIID contract 84-PP-4861-(A-1) to M.W.E. and by NIH grant AI16928 to D.M.G.

REFERENCES

1. **Aktories, K., M. Barmann, I. Ohishi, S. Tsuyama, K. H. Jakobs, and E. Habermann.** 1986. Botulinal C2 toxin ADP-ribosylates actin. Nature (London) **322**:390–392.
2. **Aktories, K., U. Braun, S. Rösener, I. Just, and A. Hall.** 1989. The *rho* gene product expressed in *E. coli* is a substrate of botulinum ADP-ribosyltransferase C3. Biochem. Biophys. Res. Commun. **158**:209–213.
3. **Aktories, K., and A. Hall.** 1989. Botulinum ADP-ribosyltransferase C3: a new tool to study low molecular weight GTP-binding proteins. Trends Pharmacol. Sci. **10**:415–418.
4. **Aktories, K., S. Rösener, U. Blaschke, and S. Chhatwal.** 1988. Botulinum ADP-ribosyltransferase C3. Purification of the enzyme and characterization of the ADP-ribosylation reaction in platelet membranes. Eur. J. Biochem. **172**:445–450.
5. **Beguin, P., P. Cornet, and J.-P. Aubert.** 1985. Sequence of a cellulase gene of the thermophilic bacterium *Clostridium thermocellum*. J. Bacteriol. **162**:102–105.
6. **Binz, T., H. Kurazono, M. R. Popoff, M. W. Eklund, G. S. Sakaguchi, S. Kozaki, K. Kriedlstein, A. Henschen, D. M. Gill, and H. Niemann.** 1990. Nucleotide sequence of the gene encoding *Clostridium botulinum* neurotoxin type D. Nucleic Acids Res. **18**:5556.
7. **Binz, T., H. Kurazono, M. Wille, J. Frevert, K. Wernars, and H. Niemann.** 1990. The complete sequence of botulinum neurotoxin type A and comparison with other clostridial neurotoxins. J. Biol. Chem. **265**:9153–9158.
8. **Bourne, H. R., D. A. Sanders, and F. McCormick.** 1990. The GTP-superfamily: a conserved switch for diverse cell functions. Nature (London) **349**:117–126.
9. **Chardin, P.** 1988. The *ras* superfamily of proteins. Biochimie **70**:865–868.
10. **Chardin, P., P. Boquet, P. Madaule, M. R. Popoff, E. J. Rubin, and D. M. Gill.** 1989. The mammalian G protein *rho* C is ADP-ribosylated by *Clostridium botulinum* exoenzyme C3 and affects actin microfilaments in Vero cells. EMBO J. **8**:1087–1092.
11. **Dover, W. J., J. F. Miller, and C. W. Ragsdale.** 1988. High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res. **16**:6127–6145.
12. **Eisel, U., W. Jarausch, K. Goretzki, A. Henschen, J. Engels, U. Weller, M. Hudel, E. Habermann, and H. Niemann.** 1986. Tetanus toxin: primary structure, expression in *E. coli* and homology with botulinum toxins. EMBO J. **5**:2495–2502.
13. **Eklund, M. W., F. T. Poysky, and W. H. Habig.** 1989. Bacteriophages and plasmids in *Clostridium botulinum* and *Clostridium tetani* and their relationship to production of toxins, p. 25–51. *In* L. L. Simpson (ed.), Botulinum neurotoxin and tetanus toxin. Academic Press, San Diego.
14. **Eklund, M. W., F. T. Poysky, S. M. Reed, and C. A. Smith.** 1971. Bacteriophages and toxigenicity of *Clostridium botulinum* type C. Science **172**:480–482.
15. **Fairweather, N. F., V. A. Lyness, D. J. Pickard, G. Allen, and R. O. Thomson.** 1986. Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in *Escherichia coli*. J. Bacteriol. **165**:21–27.
16. **Fujii, N., K. Oguma, N. Yokosawa, K. Kimura, and K. Tsuzuki.** 1988. Characterization of bacteriophage nucleic acids obtained from *Clostridium botulinum* types C and D. Appl. Environ. Microbiol. **54**:69–73.
17. **Garnier, T., and S. T. Cole.** 1986. Characterization of a bacteriocinogenic plasmid from *Clostridium perfringens* and molecular genetic analysis of the bacteriocin-encoding gene. J. Bacteriol. **168**:1189–1196.
18. **Grepinet, O., and P. Beguin.** 1986. Sequence of the cellulase gene of *Clostridium thermocellum* coding for endoglucanase B. Nucleic Acids Res. **14**:1791–1798.
19. **Hauser, D., M. W. Eklund, H. Kurazono, T. Binz, H. Niemann, D. M. Gill, P. Boquet, and M. R. Popoff.** 1990. Nucleotide sequence of *Clostridium botulinum* C1 neurotoxin. Nucleic Acids Res. **18**:4924.
20. **Joliff, G., P. Beguin, and J.-P. Aubert.** 1986. Nucleotide sequence of the cellulase gene *cel D* encoding endoglucanase D of *Clostridium thermocellum*. Nucleic Acids Res. **14**:8605–8613.
21. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
22. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **McCaffrey, M., J. S. Johnson, B. Goud, A. Myers, J. Rossier, M. R. Popoff, P. Madaule, and P. Boquet.** The small GTP-binding protein $Rho1_p$ is localized in the Golgi apparatus and post Golgi vesicles in *S. cerevisiae*. J. Cell Biol., in press.
24. **Narumiya, S., A. Sekine, and M. Fujiwara.** 1988. Substrate for botulinum ADP-ribosyltransferase, Gb, has an amino acid sequence homologous to a putative *rho* gene product. J. Biol. Chem. **263**:17255–17257.
25. **Paterson, H. F., A. J. Self, M. D. Garrett, I. Just, K. Aktories, and A. Hall.** 1990. Microinjection of recombinant p^{21} *rho* induces rapid changes in cell morphology. J. Cell Biol. **111**:1001–1007.
26. **Popoff, M. R., P. Boquet, D. M. Gill, and M. W. Eklund.** 1990. DNA sequence of exoenzyme C3, an ADP-ribosyltransferase encoded by *Clostridium botulinum* C and D phages. Nucleic Acids Res. **18**:1291.
27. **Popoff, M. R., J.-P. Guillou, and J.-P. Carlier.** 1985. Taxonomic position of lecithinase-negative strains of *Clostridium sordellii*. J. Gen. Microbiol. **131**:1697–1703.
28. **Riley, J., R. Butler, D. Ogilvie, R. Finnieur, D. Jenner, S. Powell, R. Anand, J. C. Smith, and A. F. Markham.** 1990. A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. Nucleic Acids Res. **18**:2887–2890.
29. **Rimm, D. L., and T. D. Pollard.** 1989. New plasmid vectors for high level synthesis of eukaryotic fusion proteins in *Escherichia coli*. Gene **75**:323–327.
30. **Rubin, E. J., D. M. Gill, P. Boquet, and M. R. Popoff.** 1988. Functional modification of a 21-kilodalton G protein when ADP-ribosylated by exoenzyme C3 of *Clostridium botulinum*. Mol. Cell. Biol. **8**:418–426.
31. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
32. **Sekine, A., M. Fujiwara, and S. Narumiya.** 1989. Asparagine residue in the *rho* gene product is the modification site for botulinum ADP-ribosyltransferase. J. Biol. Chem. **264**:8602–8605.
33. **Thompson, D. E., J. K. Brehm, J. D. Oultram, T.-J. Swinfield, C. S. Shone, T. Atkinson, J. Melling, and N. P. Minton.** 1990. The complete amino acid sequence of the *Clostridium botulinum* type A neurotoxin, deduced by nucleotide sequence analysis of the encoding gene. Eur. J. Biochem. **189**:73–81.
34. **Titball, R. W., S. E. C. Hunter, K. L. Martin, B. C. Morris, A. D. Shuttleworth, R. Rubidge, D. W. Anderson, and D. C. Kelly.** 1989. Molecular cloning and nucleotide sequence of the alpha-toxin (phospholipase C) of *Clostridium perfringens*. Infect. Immun. **57**:367–376.
35. **Tso, J. Y., and C. Siebel.** 1989. Cloning and expression of the phospholipase C gene from *Clostridium perfringens* and *Clostridium bifermentans*. Infect. Immun. **57**:468–476.
36. **Tsuzuki, K., K. Kimura, N. Fujii, N. Yokosawa, T. Indoh, T. Murakami, and K. Oguma.** 1990. Cloning and complete nucleotide sequence of the gene for the main component of hemagglutinin produced by *Clostridium botulinum* type C. Infect. Immun. **58**:3173–3177.
37. **Tweten, R. K.** 1988. Cloning and expression in *Escherichia coli* of the perfringolysin O (theta-toxin) gene from *Clostridium perfringens* and characterization of the gene product. Infect. Immun. **56**:3228–3234.
38. **Von Heijne, G.** 1985. Signal sequences. The limits of variation. J. Mol. Biol. **184**:99–105.
39. **Young, M., N. P. Minton, and W. L. Staudenbauer.** 1989. Recent advances in the genetics of the clostridia. FEMS Microbiol. Rev. **63**:301–326.