Vol. 183, No. 1, 1992 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
February 28, 1992 Pages 107-113

SEQUENCES OF THE BOTULINAL NEUROTOXIN E DERIVED FROM *CLOSTRIDIUM BOTULINUM* TYPE E (STRAIN BELUGA) AND *CLOSTRIDIUM BUTYRICUM* (STRAINS ATCC 43181 AND ATCC 43755)

Sylvie Poulet[1], Daniel Hauser[1], Martin Quanz[2], Heiner Niemann[2], and Michel R. Popoff[1*]

[1]Unité des Antigènes Bactériens, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France

[2]Federal Research Center for Virus Diseases of Animals, W-7400 Tübingen, Federal Republic of Germany

Received November 14, 1991

SUMMARY:Recently, it has been shown that two *Clostridium butyricum* strains (ATCC 43181 and ATCC 43755), isolated from cases of infant botulism, produce a botulinal neurotoxin type E (BoNT/E). Here we have determined the nucleotide sequences of the BoNT/E genes of these two *C. butyricum* strains and from *C. botulinum* E strain Beluga. We show that the sequences of the BoNT/E genes from the two *C. butyricum* strains are identical and differ in only 64 positions resulting in 39 amino acid changes (97% identity at the amino acid level) from that derived from *C. botulinum*. Our data suggest a transfer of the BoNT/E gene from *C. botulinum* to the originally nontoxigenic *C. butyricum* strains.

The clostridial neurotoxins are highly potent protein toxins that inhibit neurotransmitter release at various synapses. These neurotoxins consist of tetanus toxin (TeTx) and seven serologically distinct botulinal neurotoxins designated BoNT/A, B, C1, D, E, F, and G, all of which are both structurally and functionally closely related. All of them are synthesized as single chain polypeptides of about 150 kDa which are proteolytically activated into di-chain derivatives constituted of a light (L) (Mr ap. 50 000), and a heavy (H) chain (Mr ap. 100 0000) linked by a single disulfide bridge (1).

*To whom correspondence should be addressed.

0006-291X/92 $1.50

A three step model has been proposed to explain the development of neurotoxicity (2). This model involves binding, internalization and intraneuronal sorting, and the actual poisoning of the nerve terminal. The heavy chains mediate the binding of toxins to cell receptors, while the light chains appear to enter target cells, and cause some internal changes (3,4). It is not known how the different heavy chains reach and affect different kinds of synapses, nor what is the molecular nature of the change caused by the light chains. However, these questions may be approached by sequence analyses and the identification of conserved domains. The genes encoding TeTx, BoNT/A, C1, and D have been sequenced previously (5, 6, 7, 8, 9). It has been shown that the BoNT/E gene is chromosomally located in *C. botulinum* E strain Beluga, and a partial sequence was established by cloning and sequencing of an EcoRI fragment (7).

Recently, it has been reported that *Clostridium* strains quite different from *C. botulinum* can produce BoNT. Thus, two *Clostridium* strains isolated from infant botulism, and identified as *C. butyricum* were shown to synthesize synthesize BoNT/E (10, 11). A comparison of purified BoNT/E derived from *C. botulinum* and *C. butyricum* showed that both toxins are very similar. Their molecular weights (145 000 Da) determined by polyacrylamide gel electrophoresis were in good agreement with that calculated from the deduced amino acid sequences. In addition, both toxins exhibited similar specific toxicity in mice (12).

Here, we have determined the complete nucleotide sequences of the BoNT/E genes from *C. botulinum* E strain Beluga, and from the toxigenic *C. butyricum* strains ATCC 43181 and ATCC 43755, and we present the deduced amino acid sequences.

MATERIALS AND METHODS

Bacterial DNA and Plasmids.*C. botulinum* type E strain Beluga, and toxigenic *C. butyricum* strains ATCC 43181 and ATCC 43755 received directely from ATCC, were grown in TGY broth (Trypticase, 30 g/l; yeast extract, 20 g/l; glucose, 5 g/l; HCl-cysteine, 0.5 g/l; pH7.2) in anaerobic conditions. Total DNA was extracted and purified as previously described (13).

PLasmid pUC 19 (Appligene, Strasbourg, France) was used for cloning in *Escherichia coli* strain TG1.

Probes and hybridization conditions.Oligonucleotides were synthesized by the phosphoramidite method using a Cyclone Miligen automated DNA synthesizer. Gene Screen Plus filters (New England Nuclear Research Products, Du Pont Nemours, Boston, USA) were pretreated with 200 µg/ml heat-denaturated salmon sperm DNA in 1 M NaCl, 10% dextran sulfate, 0.5% SDS, 50 mM Tris HCl, pH 7.5 at 40°C and then with a 5'(^{32}P) labeled oligonucleotide (10^6 cpm/ml) in the same mixture overnight at 40°C. Filters were washed in 6 X SSC, 0.1% SDS at 40°C for two hours and exposed overnight to Fuji RX films.

Polymerase chain reaction amplification. *C. butyricum* DNA (100ng) was amplified using the polymerase chain reaction (PCR) and primers deduced from the *C. botulinum* E strain Beluga BoNT/E gene. The reactions were done in a total volume of 100 µl containing 10 mM Tris HCl, pH 8.3, 50 mM KCl, 4 mM $MgCl_2$, 0.1% BSA, 100 µM dNTP, 10 mM beta-mercaptoethanol, 25 pmol of each primer, and 2.5 U of Taq polymerase (Beckman, Paris, France). Reaction mixtures were denaturated at 95°C for 2 min and then submitted to 30 subsequent cycles consisting of denaturation (20 s at 94°C), annealing (20s at hybridization temperature which was 5°C below the theoretical melting temperature of the primers), and extension (20 s at 72°C) in a DNA Thermal Cycler version 2.2 (Perking Elemer Cetus, Emeryville, USA). Amplification products were purified by Geneclean (Bio 101 Inc., La Jolla, USA) and sequenced.

Other molecular biological techniques. Ligation reactions and preparation of plasmid DNA from *E. coli* were conducted as described by Maniatis *et al.* (14). Bacteria were transformed by electroporation. T4 polynucleotide kinase and calf intestinal phosphatase were from Boehringer-Mannheim France, and other enzymes from Pharmacia (Paris, France). DNA was sequenced by the dideoxy-chain-termination procedure (15) using the Sequenase Kit (United States Biochemical Corporation, Cleveland, USA).

RESULTS AND DISCUSSION

We used a synthetic 42mer oligonucleotide representing the nucleotide sequence from positions 716 to 757 of the previously published EcoRI fragment of the BoNT/E gene to identify and clone an overlapping 1373 bp NsiI fragment. This fragment was then used to clone a ScaI fragment (1639 bp) to yield pMRP43, and with the help of the latter clone, pMRP46 containing an HindIII fragment of 2500 bp was obtained. Together, these four clones encompass the entire coding region of the BoNT/E gene from *C. botulinum* E strain Beluga. The nucleotide sequence (4017 bp) has been entered in the EMBL database under the accession number X62089. It contains a single open reading frame of 3753 nucleotides. The deduced amino acid sequence (1251 residues) is shown in Fig. 1.

Overlapping DNA fragments of 400 to 500 bp were amplified from *C. butyricum* strains ATCC 43181 and ATCC 43755 DNA by PCR using oligonucleotides the sequences of which were deduced from the *C. botulinum* E BoNT/E gene. The nucleotide sequences of the BoNT/E genes from the two *C. butyricum* strains were determined to be identical and have been entered in the EMBL database under the accession number X62088.

A comparison of the DNA sequences established for the BoNT/E genes from the *C. botulinum* and the *C. butyricum* strains revealed differences in 69 positions, 64 of them in the coding region and 5 in the 5' noncoding region. Thirty nine amino-acid changes were detected, 19 of them (48%) being located within the 198 N-terminal residues which represent 16% of the entire sequence. Amino-acid identities between *C. botulinum* E and *C. butyricum* BoNT/E sequences are 95% for the L

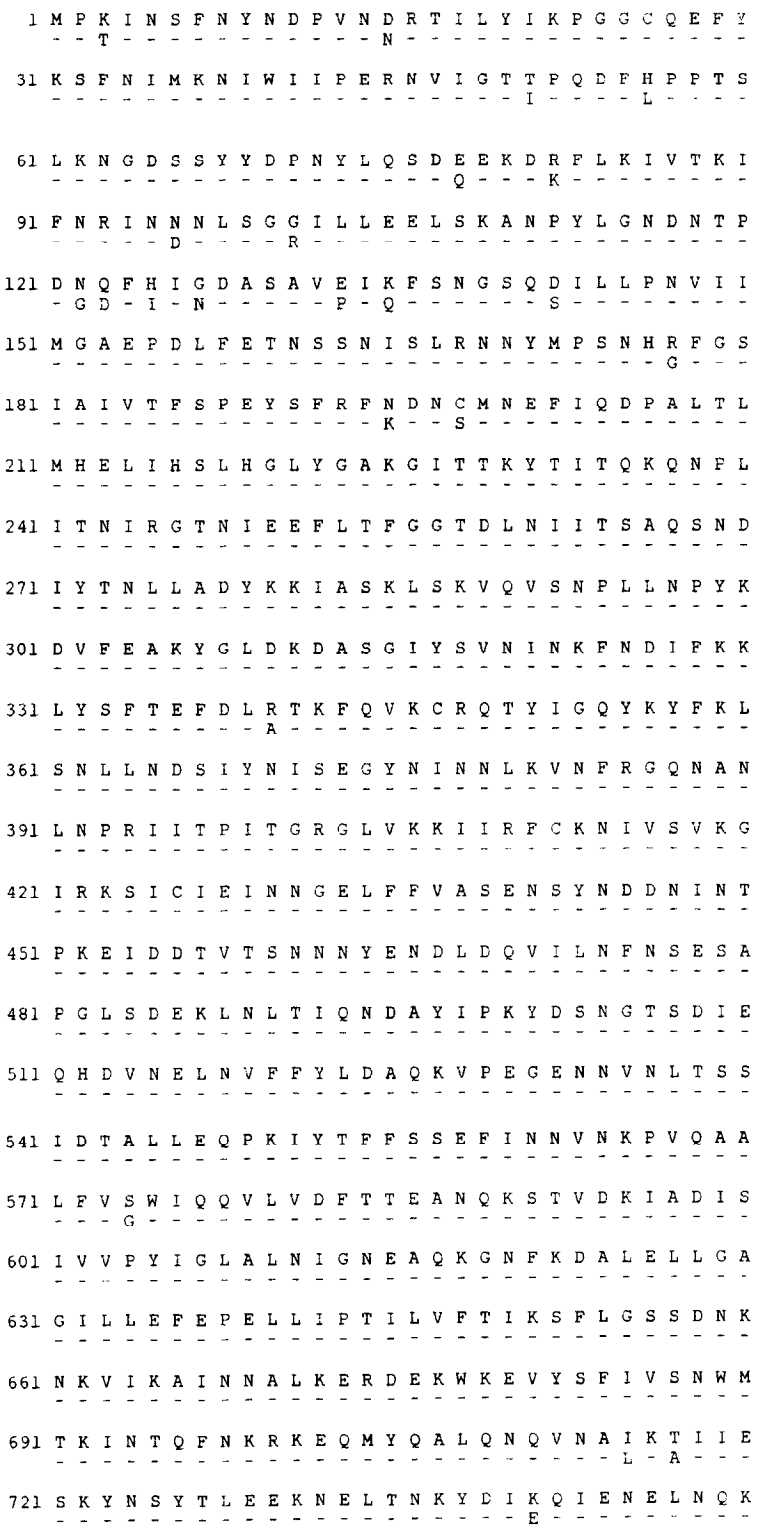

Fig. 1 . Alignment of the deduced amino-acid sequences of BoNT/E derived from *C. botulinum* E strain Beluga (upper line), and *C. butyricum* strains ATCC 43181 or ATCC 43755 (lower line). Dashes represent identical amino acids.

```
751 V S I A M N N I D R F L T E S S I S Y L M K I I N E V K I N
    - - - - - - - - - - - - - - - - - - - - - L - - - - - - - -

781 K L R E Y D E N V K T Y L L N Y I I Q H G S I L G E S Q Q E
    - - - - - - - - - - - - - - D - - - K - - - - - - - - - - -

811 L N S M V T D T L N N S I P F K L S S Y T D D K I L I S Y F
    - - - - - I - - - - - - - - - - - - - - - - - - - - - - -

841 N K F F K R I K S S S V L N M R Y K N D K Y V D T S G Y D S
    - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

871 N I N I N G D V Y K Y P T N K N Q F G I Y N D K L S E V N I
    - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

901 S Q N D Y I I Y D N K Y K N F S I S F W V R I P N Y D N K I
    - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

931 V N V N N E Y T I I N C M R D N N S G W K V S L N H N E I I
    - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

961 W T F E D N R G I N Q K L A F N Y G N A N G I S D Y I N K W
    - - L Q - - S - - - - - - - - - - - - - - - - - - - - - -

991 I F V T I T N D R L G D S K L Y I N G N L I D Q K S I L N L
    - - - - - - - - - - - - - - - - - - - - - - K - - - - - -

1021 G N I H V S D N I L F K I V N C S Y T R Y I G I R Y F N I F
     - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

1051 D K E L D E T E I Q T L Y S N E P N T N I L K D F W G N Y L
     - - - - - - - - - - - - N - - - - A - - - - - - - - - - -

1081 L Y D K E Y Y L L N V L K P N N F I D R R K D S T L S I N N
     - - - - - - - - - - - - - - - - - N - - T - - - - - - - -

1111 I R S T I L L A N R L Y S G I K V K I Q R V N N S S T N D N
     - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

1141 L V R K N D Q V Y I N F V A S K T H L F P L Y A D T A T T N
     - - - - - - - - - - - - - - - - - - - - L - - - - - - - -

1171 K E K T I K I S S S G N R F N Q V V V M N S V G N C T M N F
     - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

1201 K N N N G N N I G L L G F K A D T V V A S T W Y Y T H M R D
     - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

1231 H T N S N G C F W N F I S E E H G W Q E K 1251
     N - - - - - F - - - - - - - - - - - - - -
```

Fig. 1 – continued

chain, and 98% for the H chain. The molecular weights of the predicted polypeptides are 143 836.70 Da for BoNT/E from *C. botulinum* E, and 143 389.33 Da for BoNT/E from the *C. butyricum* strains, respectively.

The deduced amino-acid sequences are in close agreement with partial sequences determined for the N-termini of the L and the H chains of BoNT/E from *C. botulinum* and *C. butyricum* (1). As shown by Das Gupta (1), the H chain of BoNT/E from *C. botulinum* E and *C. butyricum* begins at Lys in position 423. Thus, the Cys residues located at positions 412 and 426 are probably involved in the disulfide bridge between the L and H chains.

Fujii *et al.* have previously reported the nucleotide sequences of the EcoRI fragment encoding the 5'-terminus of the BoTN/E gene from *C. botulinum* E strains

Mashike, Iwanai, and Otaru, and from *C. butyricum* strain BL6340 (16, 17). Several differences were found between these nucleotide sequences and the corresponding sequences of *C. botulinum* E strain Beluga and *C. butyricum* strains ATCC 43181 and ATCC 43755. As reported in this study, the Beluga strain produces a BoNT/E that contains an Arg at position 177, Cys at 198, and Lys at 230. The Japanese BoNT/E reference strains contained a neurotoxin with Gly at position 177, Ser at 198, and Met at position 230. The presence of Met in this latter position was also detected in BoNT/E from the *C. butyricum* strain BL6340 (17). Taken together, these findings indicate that botulinal neurotoxins, although belonging to the same toxinotype, may exhibit minor differences in their amino acid sequences.

BoNT/E (1251 residues) is significantly shorter than TeTx (1315 residues), BoNT/A (1296 residues), BoNT/C1 (1291 residues) and BoNT/D (1276 residues). At the amino-acid level, BoNT/E from *C. botulinum* has an overall identity of 38.3% with TeTx, 44.0% with BoNT/A, 38.3% with BoNT/C1, and 32.0% with BoNT/D.

Functional domains which constitute the receptor binding sites on the H chains or which are involved in the as yet unspecified toxification process mediated by the L chains are expected to be conserved. An alignment of the amino acid sequences of the L chains of TeTx, BoNT/A, B, C1, D, and E revealed that indeed highly conserved domains are separated from each other by short variable regions composed of 10 to 30 unrelated residues (18). In this respect, it is noteworthy that only 1 out of the 19 amino acid exchanges found in the N-terminal portions of the two BoNT/E L chain sequences involves a residue that is conserved in the other neurotoxins, while 10 of the replacements reside in such variable regions. It is possible that these regions define the immunologic differences of the individual toxinotypes. This would explain why only five out of nine monoclonal antibodies raised against BoNT/E from *C. botulinum* E reacted with the neurotoxin from the *C. butyricum* strains (19).

The His-rich motif in the center of the BoNT/E L chain involving His at positions 212, 216, and 219, is conserved in the sequences of all clostridial neurotoxins analysed so far (7). It remains to be shown, wether this motif is directely involved in the toxification reaction or in the translocation process of the L chain into the cytosole. Alternatively, this motif could merely serve as an element that stabilizes the tertiary structures of the individual L chains (18).

The high level of identity between BoNT/E from *C. botulinum* and *C. butyricum* suggests a transfer of the BoNT/E gene from *C. botulinum* to originally nontoxigenic *C. butyricum* strains, and an independent evolution from each other by point mutations. A similar unexpected synthesis of BoNT/F in *C. baratii* has been reported previously (11, 20). These findings underscore the general observation that classification of *C. botulinum* strains on the basis of their neurotoxins bears little

significance with respect to the physiologic properties of the corresponding host strains.

ACKNOWLEDGMENT

This work was supported by grant Nie 175/5-2 from the Deutsche Forschungsgemeinschaft .

REFERENCES

1. DasGupta, B.R. (1989). In: Botulinum Neurotoxin and Tetanus Toxin (Simpson L.L. Ed.) pp 53-67 Academic Press, San Diego.
2. Simpson L.L. (1989). In: Botulinum Neurotoxin and Tetanus Toxin (Simpson L.L. Ed.) pp 153-178 Academic Press, San Diego.
3. Ahnert-Hilger, G., Weller, U., Dauzenroth, M.E., Habermann, E., and Gratzl, M. (1989). FEBS Let. 242,245-248.
4. Bittner, M.A., Habig, W.H., and Holz, R.W. (1989). J. Neurochem. 53, 966-968.
5. Fairweather, N.F., Lyness, V.A., Pickard, D.J., Allen, G., and Thomson, R.O. (1986). J. Bacteriol. 165, 21-27.
6. Eisel, U., Jarausch, W., Goretzki, K., Henschen, A., Engels, J., Weller, U., Hudel, M., Habermann, E., and Niemann, H. (1986). EMBO J. 5,2495-2502.
7. Binz, T., Kurazono, H., Wille, M., Frevert, J., Wernars, U., and Niemann, H. (1990). J. Biol. Chem. 265,9153-9158.
8. Hauser, D., Eklund, M.W., Kurazono, H., Binz, T., Niemann, H., Gill, D.M., Boquet, P., and Popoff, M.R. (1990). Nucl. Acids Res. 18,4924.
9. Binz, T., Kurazono, H., Popoff, M.R., Eklund, M.W., Sakaguchi, G.S., Kozaki, S., Krieglstein, K., Henschen, A., Gill, D.M., and Niemann, H. (1990) Nucl. Acids Res. 18, 5556.
10. Aureli, P., Fenici, L., Pasolini, B., Gianfranceshi, M., McCroskey, L.M., and Hatheway, C.L. (1986). J. Infect. Dis. 154,207-211.
11. Suen, J.C., Hatheway, C.L., Steigerwalt, A.G., and Brenner, D.J. (1988). J. Clin. Microbiol. 26,2191-2192.
12. Gimenez, J.A., and Sugiyama, H. (1988). Infect. Immun. 56,926-929.
13. Popoff, M.R., Guillou, J.P., and Carlier, J.P. (1985). J. Gen. Microbiol. 131,1697-1703.
14. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
15. Sanger, F., Nicklen, S., and Coulson, A.R. (1977). Proc. Natl. Acad. Sci. USA 74, 5463-5467.
16. Fujii, N., Kimura, K., Indoh, T., Murakami, T., Yashiki, T., Tsuzuki, K., Yokosawa, N., and Oguma, K. (1990). Microbiol. Immunol. 34,1041-1047.
17. Fujii, N., Kimura, K., Yashiki, T., Indoh, T., Murakami, T., Tsuzuki, K., Yokosawa, N., and Oguma, K. (1991). J. Gen. Microbiol. 137,519-525.
18. Niemann, H. (1991). In: Sourcebook of Bacterial Protein Toxins (Alouf J., and Freer J., Eds). Academic Press. New York.
19. Kozaki, S., Onimaru, J., Kamata, Y., and Sakaguchi, G. (1991). Infect. Immun. 59,457-459.
20. Hall, J.D., McCroskey, L.M., Pincomb, B.J., and Hatheway, C.L. (1985). J. Clin. Microbiol. 21,654-655.