Nucleotide sequence of *Clostridium botulinum* C1 neurotoxin

Daniel Hauser, Melvin W.Eklund², Hisao Kurazono³, Thomas Binz³, Heiner Niemann³, D.Michael Gill⁴, Patrice Boquet¹ and Michel R.Popoff*

Unité des Anaérobies and ¹Unité des Antigènes Bactériens, Institut Pasteur, 28 rue du Dr.

Roux – 75724 Paris cedex 15, France, ²Northwest and Alaska Fisheries Center, 2725 Montlake

Boulevard East, Seattle, WA 98112, USA, ³Institut für Mikrobiologie an der Bundesforschungsanstalt

für Viruserkrankheiten der Tiere, Postfach 1149, 7400 Tübingen, FRG and ⁴Department of Molecular

Biology and Microbiology, Tufts University Schools of Medicine, Boston, MA 02111, USA

Submitted July 13, 1990

EMBL accession no. X53751

Clostridium botulinum neurotoxins consist of seven serologically distinct types: A, B, C1, D, E, F, and G. All of these are synthesized as single-chain polypeptides of about 150 kDa which are generally nicked after secretion to yield N-terminal light (≈ 50 kDa) and C-terminal heavy (≈ 100 kDa) chains connected by a single disulfide bond. *C. botulinum* C1 neurotoxin is encoded by a bacteriophage (1, 2).

C. botulinum C1 neurotoxin DNA sequence shown below was derived from cloned C phage DNA fragments the sizes of which were less than 2.0 kb. The initial clone containing a C1 neurotoxin coding DNA was identified using a mixed 29-mer probe [5'-GA(T/C)CCIGTIGA(T/C)AA(T/C)AA(A/G)AA(T/C)AT-ICTITA-3'] which corresponds to Asp¹² to Tyr²¹ of the partial amino acid sequence published previously (3). Potential biohazards associated with the experiments described had been approved by the French National Control Committee.

The complete C1 DNA sequence shows a single open reading frame which begins with an ATG at position 207 and encodes for 1291 amino acids (MR 148,698 Da). Before the ATG a Shine-Dalgarno sequence can be localized from position 191 to 196. Cys⁴³⁷ and Cys⁴⁵³ are probably involved in a disulfide bridge linking the light and heavy chains. A candidate transcription terminator ($\Delta G = -6.2$ kCal.mol⁻¹) is underlined. *C. botulinum* C1 neurotoxin exhibits no similarity with exoenzyme

C3 which is also encoded by *C. botulinum* C and D phages (4), in particular the C1 neurotoxin lacks a leader sequence which was found in exoenzyme C3.

ACKNOWLEDGEMENTS

We are very 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. This work was supported by grant Nie 175/5-2 from the Deutsche Forschungsgemeinschaft to H.N. (H.K. is a recipient of a fellowship from the Alexander von Humboldt Foundation) and by USAMRIID contract 84-PP-4861-(A-1) to M.E. and by NIH grant AI16928 to DMG.

REFERENCES

1. Eklund, M.W., Poysky, F.T., Reed, S.M. and Smith, C.A. (1971) *Science* **172**, 480–482.
2. Eklund, M.W. and Poysky, F.T. (1989) In: *Botulinum Neurotoxins and Tetanus Toxin* (Simpson, L.L. ed.), pp. 25–51, Academic Press, San Diego.
3. Tsuzuki, K., Yokosawa, N., Syuto, B., Ohishi, I., Fujii, N., Kimura, K. and Oguma, K. (1988) *Infect. Immun.* **56**, 898–902.
4. Popoff, M.R., Boquet, P., Gill, D.M. and Eklund, M.W. (1990) *Nucl. Acids Res.* **18**, 1291.

1 GAATCAATG TAAGTCAAG ATAATGCAA AGGATTACG TTAATCTGA TTATTAGAC
61 CTGATTGTAT AACATTTTCA TATAATGATA AATATTTCT CTATCACTTA GAGATGAGA
121 TTATTAATGG ATGATATGTA ATGCAATATA CAGGTGCTA AAGGTGCACA TTGTGGBATA
181 TTAGAAGTGT AGGAGATGTT AGTATTATGC CATATACCAAT TAACAACTTT AATTATTCAG
241 ATCCCTGTGA TAATAAAAT ATTTTATATT TAGATACTCA TTAAATACA CTAGCTAATG
301 AGCCTGAAAG AGCCTTTCGC ATACAGGAA ATATATGSGT AATACCTGAT AGATTTCGA
361 GAAATTTCTAA TCCAAATTTA AATAAAGCTT CTGAGATTAC AAGCCTTAA AGTGTATTAT
421 ATGATCTTAA TTATTGATGT AGTATCTGCT ACAGAGATCC ATTTTAAAA GAATATTAA
481 AGTATTTTAA AAGATTAAT TCTAGAGAA TAGAGAGAGA ATTAATATAT AGACTTTCA
541 CAGATATACC CTTTCTGGG AATAACATA CTTCAATTAA TACTTTTGT TTGATGTAG
601 ATTTTAAACG TGTGATGTT AAACTAGAG AAGTAAACAA CTGGTTTAAA ACTGTAGCA
661 TAATCTTAG TGTATAATA ACTGAGCTTA GAGAAACAT TATAGATCCA GAACCTTCTA
721 CBTTTAAAT AACTAAATAT ACTTTTCCG CACAAAGAG ATTGTGCTCT TTATCAATAA
781 TTTCAATATC ACCTAGATTT ATGCTAACAT ATAGTATGTC AACTAATGAT GTAGAGAGG
841 GTAGATTTTC TAAGTCTGAA TTGTGATG ATCCAACTAT AATTTTAATG CATGAACCTA
901 ATGATGCAAT GCATTAATTA TAGGAATAG CTATACCAAA TGATCAACCA ATTTCTCTG
961 TACTAGTAA TATTTTTTAT TCTCAATATA ATGTGAATAT AGATATGCA GAATATATG
1021 CATTTGAGG TCCAACTATA GACTTTATCT CTAAGAGTGC AAGSAAATAT TTGAGAGAA
1081 AGGCTTGTGA TTATTATAGA TCTATAGCTA AAGAGCTTAA TAGTATACCT ACTGCAATC
1141 CTTCAAGCTT TAATAAATAT ATAGGGAAT ATAAACAGAA ACTTATTAGA AAGTATAGAT
1201 TGTATAGAA ATCTTCAGGT BAAGTTACAG TAAGTCTGAA TAAGTTTGTG BASTATATA
1261 ATGACATTAC ACAATATATT ACAGATTTA ACTAGCTGAA AATATATAAT GTACAAATA
1321 GBAATAATATA TCTTTCAAT GTATATCTC CGTATACGAC GAATATATTA GACATAGT
1381 TTATGATAT ACAAAATGGA TTAAATATAC CTAAGAGTAA TTAAATGTA CTATTATAG
1441 GTCAAAATTT ATCTCBAAT CCAGCATTA AAGAGTCAA TCTTGAATAT ATGCTTTATT

1501 TATTTCACAA ATTTTGTCT AAAGCAATG ATGATGATC ATTATATAA AAAACATAG
1561 ATTTGAGAGA CTTTATAGT AAAAGTACT ACTTACCTT TATAGTATAT ATTAGTATG
1621 TTAAGACTGA TATATTTTGA AGAAGAGATA TAAATGAGA AACTGAAGTT ATATACTATC
1681 CAGCAATGTT TTAGATAGAT CAGGTATATC TCAATAGAA TACCTCAGAA CATGCAAC
1741 TAGATTTATT ATACCTTAT ATTAGAGTGT AGATGAGAT ATTACAGGG GAGATCAAG
1801 TCTTTATGA TAATGAGACT CAAGATGTTG ATTTATTTGA TCTTATTTAT TACCTAGAT
1861 CTCAAAACCT AAGTATATAT GTTAGAGAT TTACTTTTAC GAGATCAATT GAGGAGCTT
1921 TGAATATAG TCAAAAGATA TATCTTACT TCTTACACT AGCTAATAAA GTAAATGCG
1981 GTGTTCAAG TGGTTTATT TTAAATGAG CAATGATGT ATTTGAGAT TTACTACAA
2041 ATATTCTAG AAAAGATACA TTAGATAAAA TACAGATGT ATCAGCTATT ATTCCTATA
2101 TAGGACCCGC ATTAATATA AGTATCTCT TAAGAGAGG AATTTTACT GAAGCATTTG
2161 CAGTTACTGG TGTACTATT TTATTAGAG CATTTCTGTA ATTTCAATA CTGCACTTG
2221 GTGCTATTGT GATTATAGT AAGGTTCAG AAAGAACGA GATTATGAA ACTATAGATA
2281 ATTTTATGA ACAAGGAT ATAGATGAGA AAGATTCATA TGAATGATG ATGGAGCT
2341 GGTATCTCAG GATTATCT CAATTTAATA ATATAGTTA TCAATGTAT GATTCTTAA
2401 ATATCAAGC AGGTGCACT AAAGTAAAG TAGATTAGA ATATAAAGA TATTCAGAA
2461 GTATTAAGA AATATATAA AGTCAGTGT AAAATTTAAA AATAGTTTAA BATGTAAAA
2521 TTTGCGAGC AATGATTAAT ATAAATTAAT TTATACAGAA ATTTCCGTA ACATATTTAT
2581 TTAAGAAAT GTTACCTAAA GTAAATGATG AATTAAATGA GTTTGATGAA AATCAAGG
2641 CAAATTAAT TAATCTTATA GATAGTATA ATATTATCT AGTTGATGAA GTAGATAAT
2701 TAAAGCAAA AGTAAATAT AGCTTTCAAA ATCAATACC CTTAATATT TTTCATATA
2761 CTAATTAATC TTATTTAAAA GATATATATA ATGAATATT CAATATATT AATGATCAA
2821 AATTTTGGG CCTACAAAC AGAAAAATA CTTAGTGA TACATAGGA TAAATGCG
2881 AAGTATGTA AGAAGGCAAT GTTCACTTAT ATCCATATT TCCATTTAG TTAAATAG
2941 GTAGTTCAG GAGAGATGA GATAAGTTA TAGTAACCA GAATGAAAT ATTTGATATA

3001 ATTTCTATGA TGAAGATTTT AGCATATGTT TTGTGATTAG AATAAATAA TGGTATAGA
3061 ATTTACCTGG ATATACATA ATGTATAGTG TTAAGAAATA CTCAGGTGG ABATAGSTA
3121 TTATTAGTAA TTTTATAGTA TTACTTTTAA AACAAATGA AGATAGTGA CAAATATGA
3181 ATTTTATGTA TGAATATACA AATAAGTCTC CTGATACAA TAAATGTTT TTGTAGCTG
3241 TTACTACAA TATGATGAGA AATATGAGA TTATATATA TGAAGATTA ATAGATACTA
3301 TAAAGTTAA ABACATATCT GBAATTAATT TTAGCAAAAC TATACTATT GAATTAATA
3361 AATTTCCGA TACCGGTTTG ATTTCTCAG ATTTCTGATA CATCAATATG TGAATAGAG
3421 ATTTTATAT ATTTGCTAAA GAATGATAGT GTAAAGATAT TAATATATTA TTTAATAGCT
3481 TGAATATAC TAATGTTGA AAGATATTAT GGGGAATGA TTTAGATAT AATAAGATG
3541 ATTTATGAT TAATATAGAT TATTAAATA GATATATGTA TGCBACTCA CBAAGATTG
3601 TTTTATATC ACTGAGAAAT AATAAGTACT CAAATGAGG ATATAAATAT ATAAATGAA
3661 GAATCAGAG AATACAAAT GATACATGAG TACAGAGAG AGATATTTTA TATTTGATA
3721 TGCAATTA TAACAAAGA TATAATTTG TTATGAGAA TGAACATGAT TATGAGATA
3781 ATCATAGT TAAGATATA TATGCTATG GTTTAAGAGA ACAACCAAG GATATAAGT
3841 ATATATAT ATTTCAATA CAACTAGTA ATAACTACTA TTATTACCA TCTCAATAT
3901 TTAAATCAA TTTTATGGA GAATATTTT CTGAGATGT TCAATAGT ACTATGCT
3961 TTAGCTTGG AGGTGATGG TATAGACAA ATATTGTTGT GCTTACTGTC AAGCAAGGA
4021 ATTTGCTTC ATTTATGAA TCAACATCAA CTCATGGGG TTTGTACTCT GTAGTGAAT
4081 AATATGAT TAATAATATA AATTATGTTA AATATTTTAA TATTAGSTT AATATGTTT
4141 TGATTTGTC AACTATATA AATACATGA ATAAATTTA TTTAAGCTG TTAAGATAGT
4201 ATTTTATTA TCAATATGAG TAGTAATAA CTGATGCTC TGTCTATAT CTTCAGAT
4261 TAAAGAGAA GATGTTAAT TTTAGAGAT TATTAGTAA AAGAGATACA GTTACGAC
4321 TAAAGATTA AATTAATTA TTTTAAAGG TATTGACAT TGAAGATA TGTATATA
4381 CTTAATAGT TGAAGAGAG ATAAAGCAA TAACTTAG TACAAATA AATTAGAAA
4441 TATTAAAGA GBAATTAATA TGAACACCA TC

* To whom correspondence should be addressed