

Comparative sequence analyses of the neurotoxin complex genes in *Clostridium botulinum* serotypes A, B, E, and F

Ajay K. Singh, Eric W. Brown,
Shashi K. Sharma

Division of Microbiology, Office
of Regulatory Science, Center for Food
Safety and Applied Nutrition, Food and
Drug Administration, College Park, MD, USA

Abstract

Neurotoxin complex (NTC) genes are arranged in two known hemagglutinin (HA) and open reading frame X (ORFX) clusters. NTC genes have been analyzed in four serotypes A, B, E and F of *Clostridium botulinum* causing human botulism. Analysis of amino acid sequences of NT genes demonstrated significant differences among subtypes and four serotypes. Phylogram tree of NT genes reveals that serotypes A1 and B1 are much closer compared to serotype E1 and F1. However, non-toxic non-hemagglutinin (NTNH) gene is highly conserved among four serotypes. Analysis of phylogram tree of NTNH gene reveals that serotypes A and F are more closely related compared to serotype B and E. Additionally, sequences of HAs and ORFX genes are very divergent but these genes are specific in subtypes and serotypes of *Clostridium botulinum*. Information derived from sequence analyses of NTC has direct implication in development of detection tools and therapeutic countermeasures for botulism.

Introduction

Botulism is a neuroparalytic disease caused by seven antigenically distinct serotypes of neurotoxins (A-G) produced primarily by *Clostridium botulinum*. Human botulism is caused by toxins produced from serotypes A, B, E, and F with vast majority of cases due to serotypes A and B. Botulinum neurotoxins (BoNTs) are highly poisonous and have great potential as a biothreat agent.^{1,2} Thus, BoNTs are listed as a category A select agents by the Centers for Disease Control and Prevention (<http://emergency.cdc.gov/bioterrorism>).³ BoNTs are produced as a progenitor toxin complex consisting of neurotoxin (NT), and neurotoxin associated proteins (NAPs) including hemagglutinins (HAs), nontoxic non-hemagglutinin (NTNH), and other uncharacterized components such as P21, P47, bot-R, and

ORFX1-3. Neurotoxin complexes (NTCs) vary in sizes [viz. 900 kDa Large Large-Toxin Complex (LL-TC), 500 kDa Large-Toxin Complex (L-TC), and 300 kDa Medium-Toxin Complex (M-TC)] and that reflects to number, ratio, presence and absence of individual components. The production of botulinum NTC is known to vary with different serotypes, strains, medium composition, and culture conditions. LL-TC has only been produced by serotype A. L-TC and M-TC is produced by serotypes A-D, and G strains, while M-TC is produced by serotype E and F strains. The biological and structural roles of the NAPs are not fully characterized but it is believed that they serve the role of protecting BoNT from digestive enzymes and adverse conditions present in the gastrointestinal tract.⁴ The genes encoding NAPs are linked as a cluster and differ in composition and structure among different serotypes and strains.⁵⁻¹³ HA cluster and open reading frame X (ORFX) cluster are two known primary gene clusters vary in serotypes and subtypes, and located upstream of the toxin gene. The HA cluster encodes for the HA17, HA-33, HA-70, bot-R, and a NTNH. The ORFX cluster encodes for ORFX-3, ORFX-2, ORFX-1, P-47, P-21, and a NTNH. The function of ORFX genes and their role in the neurotoxin cluster are still unknown, but it is assumed that they perform a similar role like their counterparts HAS because the location and orientation of the ORFX and HA cluster genes relative to the NT gene in the neurotoxin cluster are analogous.⁸

Over the past decade, a significant amount of BoNT gene clusters sequence data was analyzed. Comparative sequence analysis of the NT gene clusters revealed interesting phylogenetic relationships both within and among serotypes.^{7,8,11} To our knowledge, no reports present a complete overview of sequence comparison of neurotoxin gene clusters among serotypes A, B, E, and F. In this communication, we have explored the sequence diversity of neurotoxin gene clusters of *C. botulinum* by comparing amino acid sequence of NT and NAPs genes in serotypes A, B, E, and F. Results indicated the existence of marked sequence diversity in neurotoxin gene clusters within subtypes and serotypes of *C. botulinum*.

Materials and Methods

The gene bank accession numbers of the National Center for Biotechnology Information were used for amino acid sequences from fourteen *C. botulinum* strains including subtypes and serotypes A, B, E, and F (Table 1). Amino acid sequences of BoNT, and NAPs genes were aligned individually by a Clustal W2.1 program of European Bioinformatics Institute. Four strains (Hall-H, Hall-A, ATCC-

Correspondence: Shashi K. Sharma, Division of Microbiology, Office of Regulatory Science, Center for Food Safety and Applied Nutrition, Food and Drug Administration, College Park, MD 20740, USA.
Tel. +1.240.4021570 - Fax: +1.301.4322632
E-mail: ajay.singh@fda.hhs.gov;
shashi.sharma@fda.hhs.gov

Key words: botulism, neurotoxin, neurotoxin complex gene, amino acid sequence.

Conflict of interests: the authors report no potential conflict of interests.

Funding: this work was supported by funding from the US Department of Homeland Security to SKS. The views or opinions presented in this review are those of the authors and should not be construed as an official view of the U.S. Food and Drug Administration.

Received for publication: 24 April 2012.

Revision received: 14 June 2012.

Accepted for publication: 9 July 2012.

This work is licensed under a Creative Commons Attribution NonCommercial 3.0 License (CC BY-NC 3.0).

©Copyright S.K. Sharma et al., 2012
Licensee PAGEPress, Italy
Microbiology Research 2012; 3:e21
doi:10.4081/mr.2012.e21

3502, and 62A) were analyzed for NT, NTNH, and HAs genes within subtype A1. Comparative sequence analysis of NT and NAP genes was performed within serotypes A, B, E, and F. Phylogram trees of BoNT and NTNH were also generated within four serotypes by Clustal W2.1 program.

Results and Discussion

The homology analysis of BoNT, NTNH, and three HA genes (HA-17, HA-34, and HA-70) within and between five subtype A is shown in Table 2. Among the four strain of subtype A1, three strains show 100% sequence identity for NT and NTNH genes except strain 62A where 1% difference was observed in NT and NTNH gene because of a single mutation in both gene (Table 2). Analysis of three HA genes revealed 100% sequence identity in all four strains. Next, the comparison of NT and NTNH gene was performed in five subtype of serotype A (A1-A5). Kyoto-F (subtype A2) and 657 Ba4 (subtype A4) strains showed 89%, and <78% sequence identity in NT and NTNH gene respectively compared to the Hall-H strain of subtype A1. NT and NTNH gene from Loch

Table 1. Gene bank accession numbers of *Clostridium botulinum* strains used in sequence analyses. Gene bank accession numbers are from National Center for Biotechnology Information databases.

Strain	Serotype/Subtype	NT	NTNH	HA-17	HA-33	HA-70	ORFX 1	ORFX 2	ORFX 3
Hall-Hyper	A1	AAM75961	AAM75960	AAM75957	AAM75958	AAM75956	-	-	-
Hall-Allergan	A1	AF488749	AF488748	AF488745	AF488746	AF488747	-	-	-
62A	A1	ACS52162	CAA63550	AAM75950	AAM75951	AAM75949	-	-	-
ATCC3502	A1	YP_001253342	YP_001253341	YP_001253333	CAL82357	CAL82355	-	-	-
Kyoto-F	A2	CAA51824	X87974	-	-	-	-	-	-
NCTC-2916	A2	X52066	CAA74634	-	-	-	AY497357	AY497357	AY497357
Loch Maree	A3	NC_010418	NC_010418	-	-	-	NC_010418	NC_010418	NC_010418
657Ba4	A4	EU341307	EU341307	-	-	-	EU341307	EU341307	EU341307
IBCA94-0216	A5	FJ959094	FJ959094	FJ959094	FJ959094	FJ959094	-	-	-
657Ba4	B	EU341304	EU341304	EU341304	EU341304	EU341304	-	-	-
Okra	B1	AB23297	AB23297	AB23297	AB23297	AB23297	-	-	-
K-81	E	AM695753	AM695753	-	-	-	AM695753	AM695753	AM695753
Beluga	E1	ZP_04822342	ZP_04822317	-	-	-	ZP_04820916	ZP_04821956	ZP_04822382
Langeland	F1	NC_009699	NC_009699	-	-	-	NC_009699	NC_009699	NC_009699
IBCA66-5436	F6	HQ441176	HQ441176	-	-	-	HQ441176	HQ441176	HQ441176

NT, neurotoxin; NTNH, non-toxic non-hemagglutinin; HA, hemagglutinin; ORFX, open reading frame X.

Maree strain (subtype A3) showed 84% and 79% sequence identity respectively to the Hall-H strain. The sequence identity of NT, NTNH, and HA genes for IBCA94-0216 strain (subtype A5) was 90%, 97%, and 91-97% respectively to the Hall-H strain of subtype A1 (Table 2).

The amino acid sequence of NT gene in serotype A was 37-39%, identical to serotypes B, E, and F respectively. However, sequence of NT gene in serotype E was 64% identical to serotype F (Table 3). Analysis of phylogram tree of NT genes revealed that serotype A1 is much closer to serotype B1 than serotype E1 and F1 (*data not shown*). A report described several noteworthy observations between the four subtypes (A1-A4) for type A including highly conserved catalytic domain residues of light chain (LC), divergent regions in protease nicking site, amino terminal of a translocation domain (H_N), and amino terminal of a receptor binding domain (H_{CN}).¹⁴ Therefore, analysis of amino acid sequences of toxin gene was performed in detail including a catalytic domain of LC, protease nicking sites, H_N and H_{CN} domains for four serotypes. The catalytic domain of LC (residues His 223, Glu 224, His 227, Glu 261, Glu 262, Glu 351, Arg 362, and Tyr 366) was completely conserved among four serotypes. The disulfides that are in between LC and HC are also highly conserved for four subtypes. Protease nicking site (residues 436-444) in serotype A1 was compared to serotypes B, E, and F. We noticed that residues 436-439 are divergent. Residues 440-444 are only present in serotype A but absent in serotypes B, E, and F. The variation in sequences of protease nicking site may be accounted for the differences in formation of double chain. We have also

Table 2. Comparison of percentage amino acid identity of neurotoxin and neurotoxin associated proteins genes in serotype A of *C. botulinum*.

Subtype	NT	NTNH	HA-17	HA-33	HA-70
A1					
Hall-H	100	100	100	100	100
Hall-A	100	100	100	100	100
62A	99	99	100	100	100
ATCC 3502	100	100	100	100	100
A2					
Kyoto-F	89	78	-	-	-
A3					
Loch Maree	84	79	-	-	-
A4					
657 Ba4	89	77	-	-	-
A5					
BCA94-0216	90	97	97	91	97

Table 3. Comparison of percentage amino acid identity of neurotoxin gene cluster, and antibody neutralization specific polypeptides (ANSPs) of neurotoxin in serotypes A, B, E, and F of *C. botulinum*.

Serotypes	A1:B1	A1:E1	A1:F1	B1:E1	B1:F1	E1:F1
NT	39	38	39	37	38	64
NTNH	83	68	77	62	65	75
HA-17	97	-	-	-	-	-
HA-33	84	-	-	-	-	-
HA-70	98	-	-	-	-	-
ANSPs (amino acid position)						
925-957	48	45	56	42	50	71
967-1013	46	55	65	48	51	66
1051-1069	47	50	55	22	27	77
1257-1296	57	40	40	42	47	80

NT, neurotoxin; NTNH, non-toxic non-hemagglutinin; HA, hemagglutinin.

analyzed four variable specific polypeptides regions including H_N domain (residues 276-357, 572-594), H_{CN} domain (residues 647-673), and H_{CN}/H_N interface (residues 880-908) as reported by Arndt *et al.* (2006) for four subtype (A1-A4).¹⁴ Sequences of four variable regions of serotype A showed significant sequence diversity compared to serotypes B, E, and F. Recently, a paper described comparison of four specific peptides (residues 925-957, 967-1013, 1051-1069, and 1275-1296) of BoNT/A1, /A2, and /A5 that are known to be a targets for antibody neutralization.¹⁵ Analysis of these four specific polypeptides was performed in serotypes B, E, and F, and compared to serotype A. Results of sequence comparisons of these four polypeptides showed 66-80% sequence identity between serotypes E1 and F1. 40-65% sequence identity was observed for serotypes B1, E1, and F1 compared to serotype A1 with the exception of polypeptide sequence 1051-1069 that shows only 22-27% sequence identity for serotype B compared to serotype E, and F (Table 3). Variability in sequences of LC, H_N, and H_{CN} regions were observed and this may be responsible for differences in structural configuration of NT because these residues are located in the belt region, at the H_{CN}/H_N interface, and the N-terminus of the H_{CN}. The diversity in sequence of NT gene within serotypes and subtypes can account for significant differences in protection with vaccines and therapeutic agents.^{12,16} Due to universal presence of NTN gene of *C. botulinum*, analysis of this gene was performed in four serotypes.¹⁰ Comparison of amino acid sequences of NTN gene for serotype A showed 68-83% identity to serotypes B, E, and F. The sequence identity for serotype B was 62-65% compared to serotype E and F, and even higher up to 75% in between serotype E and F (Table 3). We also noticed that the ORF of NTN gene for serotypes A and B consist of 1193 and 1197 amino acid residues, respectively. However, the ORF of NTN gene for serotypes E and F consist of 1163 residues and are lacking 30-34 amino acid residues compared to serotypes A and B. Analysis of phylogram tree of NTN genes from four serotypes reveal that serotype A and F are more closely related compared to serotype B and E (*data not shown*). Based on sequence homology data, it appears that NTN gene has higher sequence homology than NT gene among four serotypes (Table 3). NTN gene has been explored as useful tool to rule out the presence of *C. botulinum* because of its universal nature in NT gene clusters.^{17,18}

Furthermore, ORFX genes were analyzed in serotypes A, E, and F. Comparison of ORFX genes revealed 78-95% sequence identity in subtypes A2-A4. Among three ORFX genes, the highest sequence identity was observed for

ORFX-3 (92-95%) between subtypes A2-A4 (*data not shown*). We also observed that serotypes A (subtypes A2-A4) show 78-97% sequence identity for ORFX genes to serotype F, and 48-77% to serotype E. Analysis of ORFX genes in serotype E and F showed 50-77% sequence identity. Based on ORFX sequence data analysis, serotype A and F has more resemblance than serotype E. We are also interested in comparing sequences of HA-17, HA-33, and HA-70 genes to ORFX-1, ORFX-2, and ORFX-3 genes because they are counterpart in NT gene clusters. Analysis of sequence comparison of HA genes show <15% identity to ORFX genes within serotypes (*data not shown*). This clearly indicates diverse nature of sequences of HA and ORFX genes even though they are counterpart in NT gene clusters.

Conclusions

In conclusion, we have analyzed amino acid sequences of NT gene clusters in serotypes A, B, E, and F of *C. botulinum*. Sequence of NT gene is highly divergent in subtypes and serotypes. However, the sequence of NTN gene is highly conserved compared to NT among serotypes. Analysis of HA and ORFX genes revealed sequence diversity among subtypes and serotypes. Information derived from sequence analyses of neurotoxin gene clusters has direct implication in development of therapeutic countermeasures and detection tools for botulism.

References

1. Arnon SS, Schechter R, Inglesby TV, et al. Botulinum toxin as a biological weapon: medical and public health management. *JAMA* 2001;285:1059-70.
2. Atlas RM. Bioterrorism: from threat to reality. *Ann Rev Microbiol* 2002;56:167-85.
3. CDC Bioterrorism Agents/Diseases. Available from: <http://emergency.cdc.gov/bioterrorism>. Accessed on January 17, 2012.
4. Oguma K, Fujinaga Y, Inoue K. Structure and function of Clostridium botulinum toxins. *Micobiol Immunol* 1995;39:161-8.
5. Bradshaw M, Dineen SS, Maks ND, Johnson EA. Regulation of neurotoxin complex expression in Clostridium botulinum strain 62A, Hall A-Hyper, and NCTC 2916. *Anaerobe* 2004;10:321-33.
6. Chen Y, Korkeala H, Aarnikunnas J, Lindstrom M. Sequencing the botulinum neurotoxin gene and related genes in Clostridium botulinum type E strains reveals orfx and a novel type E neurotoxin

- subtype. *J of Bacteriol* 2007;189:8643-50.
7. Collins MD, East AK. Phylogeny and taxonomy of the food borne pathogen Clostridium botulinum and its neurotoxins. *J Appl Microbiol* 1998;84:5-17.
8. Dineen SS, Bradshaw M, Johnson EA. Neurotoxin gene clusters in Clostridium botulinum type A strains: sequence comparison and evolutionary implications. *Curr Microbiol* 2003;46:345-52.
9. Dover N, Barash JR, Hill KK, et al. Novel structural elements within the non-proteolytic Clostridium botulinum type F toxin gene cluster. *App Environ Microbiol* 2011;77:1904-6.
10. Franciosa G, Floridi F, Maugliani A, Aureli P. Differentiation of the gene clusters encoding botulinum neurotoxin type A complexes in Clostridium botulinum type A, Ab, and A(B) strains. *App Environ Microbiol* 2004;70:7192-9.
11. Jacobson MJ, Lin G, Raphael B, et al. Analysis of neurotoxin cluster genes in Clostridium botulinum strain producing botulinum neurotoxin serotype A subtypes. *App Environ Microbiol* 2008;74:2778-86.
12. Smith TJ, Lou J, Geren IN, et al. Sequence variation within botulinum neurotoxin serotypes impacts antibody binding and neutralization. *Infect Immun* 2005;73:5450-7.
13. Smith TJ, Hill KK, Foley BT, et al. Analysis of the neurotoxin complex genes in Clostridium botulinum A1-A4 and B1 strains: BoNT/A3, /Ba4 and B1 clusters are located within plasmids. *PLoS One* 2007;12:e1271.
14. Arndt JW, Jacobson MJ, Abola EE, et al. A structural perspective of the sequence variability within botulinum neurotoxin subtype A1-A4. *J Mol Biol* 2006;362:733-42.
15. Jacobson MJ, Lin G, Tepp WH, et al. Purification, modeling, and analysis of botulinum neurotoxin subtype A5 (BoNT/A5) from Clostridium botulinum strain A661222. *App Environ Microbiol* 2011;77:4217-22.
16. Webb RP, Smith TJ, Wright PM, et al. Protection with recombinant Clostridium botulinum C1 and D binding domain subunit (Hc) vaccines against C and D neurotoxins. *Vaccine* 2007;25:4273-82.
17. Raphael BH, Andreadis JD. Real-time PCR detection of the nontoxic nonhemagglutinin gene as a rapid screening method for bacterial isolates harboring the botulinum neurotoxin (A-G) gene complex. *J Microbiol Methods* 2007;71:343-6.
18. Hill BJ, Skerry JC, Smith TJ, et al. Universal and specific quantitative detection of botulinum neurotoxin genes. *BMC Microbiol* 2010;10:267.