APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Aug. 2005, p. 4478–4486 Vol. 71, No. 8
0099-2240/05/$08.00+0 doi:10.1128/AEM.71.8.4478–4486.2005

Characterization of Botulinum Progenitor Toxins by Mass Spectrometry†

Harry B. Hines,* Frank Lebeda, Martha Hale, and Ernst E. Brueggemann

Department of Cell Biology and Biochemistry, Toxinology Division, United States Army Medical Research Institute of Infectious Diseases, Frederick, Maryland 21702-5011

Received 22 September 2004/Accepted 24 February 2005

Botulinum toxin analysis has renewed importance. This study included the use of nanochromatography-nanoelectrospray-mass spectrometry/mass spectrometry to characterize the protein composition of botulinum progenitor toxins and to assign botulinum progenitor toxins to their proper serotype and strain by using currently available sequence information. *Clostridium botulinum* progenitor toxins from strains Hall, Okra, Stockholm, MDPH, Alaska, and Langeland and 89 representing serotypes A through G, respectively, were reduced, alkylated, digested with trypsin, and identified by matching the processed product ion spectra of the tryptic peptides to proteins in accessible databases. All proteins known to be present in progenitor toxins from each serotype were identified. Additional proteins, including flagellins, ORF-X1, and neurotoxin binding protein, not previously reported to be associated with progenitor toxins, were present also in samples from several serotypes. Protein identification was used to assign toxins to a serotype and strain. Serotype assignments were accurate, and strain assignments were best when either sufficient nucleotide or amino acid sequence data were available. Minor difficulties were encountered using neurotoxin-associated protein identification for assigning serotype and strain. This study found that combined nanoscale chromatographic and mass spectrometric techniques can characterize *C. botulinum* progenitor toxin protein composition and that serotype/strain assignments based upon these proteins can provide accurate serotype and, in most instances, strain assignments using currently available information. Assignment accuracy will continue to improve as more nucleotide/amino acid sequence information becomes available for different botulinum strains.

Toxigenic strains of the anaerobic bacterium *Clostridium botulinum* produce seven immunologically distinct protein neurotoxins (BoNTs) designated as serotypes A through G. Strains of two related species, *C. barati* and *C. butyricum*, also produce some of these toxic proteins. The neurotoxins block acetylcholine release at the neuromuscular junction, resulting in the flaccid paralysis in humans and animals commonly known as botulism (26). Whether growing in culture or foods, *Clostridia* release their neurotoxins as protein aggregates. These aggregates, designated progenitor toxins or toxin complexes, result from the noncovalent association of neurotoxin with up to seven other proteins known as neurotoxin-associated proteins (NAPs) (13, 21). NAPs include a nontoxic, nonhemagglutinin protein (NTNH) and several proteins possessing hemagglutination properties. Hemagglutinins are labeled as HA combined with their molecular mass, as determined by gel electrophoresis. For example, HA-70, HA-33, and HA-17 refer to three different hemagglutinins having masses of approximately 70, 33, and 17 kDa, respectively. Hemagglutinins are found in progenitor toxins from serotypes A, B, C1, and D, but not serotypes E and F (13). Other progenitor toxin hemagglutinins, HA-53 and HA-22, have also been described, but they are derived from the posttranslational cleavage of HA-70 (27). Table 1 lists the NAPs currently associated with botulinum progenitor toxins from each serotype.

The association of different NAPs with neurotoxin generates a range of progenitor toxin masses and sizes (13). Masses range from 300 to 900 kDa, and sizes, categorized by sucrose density gradient centrifugation, include 12S (300-kDa), 16S (500-kDa), and 19S (900-kDa) complexes, which have alternate designations of M, L, and LL, respectively (20, 25). Pure neurotoxins have molecular masses of approximately 150 kDa and size designations of either 7S or S (24). Progenitor toxin size distributions vary by serotype, with type A strains producing 19S, 16S, and 12S progenitor toxins, type B, C, D, and E strains producing 16S and 12S progenitor toxins, and type F and G strains producing only 12S and 16S progenitor toxins, respectively (9, 17, 22, 23, 24, 31, 34). Neurotoxin and NTNH comprise the 12S (M) progenitor toxin (30), the 16S (L) progenitor toxin is composed of neurotoxin, NTNH, and additional NAPs (10, 15, 19), and the 19S (LL) progenitor toxin is a dimer of the 16S (L) progenitor toxin. Genes encoding progenitor toxins are clustered, have a defined spatial order, and are expressed in a coordinated manner, and progenitor toxin assembly apparently proceeds in a concerted fashion (6, 7, 14, 16, 18). Functions ascribed to NAPs range from neurotoxin protection against acidic pH levels and proteases by NTNH to HA-33 assistance in binding to and translocating progenitor toxin(s) through alimentary epithelial barriers (11, 12, 24). However, specific functions concerning the toxic mechanisms of action are not fully characterized and have not been assigned to all NAPs. Progenitor toxins are, in general, thought to be more potent in vitro and in vivo against their respective serotype substrates than purified neurotoxins (3). This is important, as certain

* Corresponding author. Mailing address: Toxinology Division, USAMRIID, 1425 Porter St., Frederick, MD 21702-5011. Phone: (301) 619-2762. Fax: (301) 619-2348. E-mail: Harry.Hines@det.amedd.army.mil.

† Supplemental material for this article may be found at http://aem.asm.org.

TABLE 1. Proteins known to form progenitor toxins according to botulinum serotype

Protein of serotype A	Protein of serotype B	Protein of serotype C	Protein of serotype D	Protein of serotype E	Protein of serotype F	Protein of serotype G
BoNT	BoNT	BoNT	BoNT	BoNT	BoNT	BoNT
NTNH	NTNH	NTNH	NTNH	NTNH	NTNH	NTNH
HA-70	HA-70	HA-70	HA-70			HA-70
HA-52		HA-55				
HA-33/35	HA-33/35	HA-33	HA-33			
HA-19/20		HA-26/21				
HA-17/15	HA-17	HA-17	HA-17			HA-17

purified BoNT serotypes possess toxicities of 1 ng/kg in mice (50% lethal dose, intravenous administration). This value is generally accepted for human toxicity as well (32). If progenitor toxins possess similar toxicities, they pose an exposure danger comparable to that of purified neurotoxins.

Botulinum neurotoxin identification, characterization, and assay have renewed importance because these toxins have been designated as potential biowarfare and terrorist threat agents due to attempts to deploy them in conventional military weapons (35) and to use them in aerosol form against a civilian population (2). Therefore, techniques capable of assaying all proteins in progenitor toxins to identify them at the strain level in a timely manner are required. Proteomic techniques based upon high-performance liquid chromatography (HPLC) and mass spectrometry (MS) can generate the amino acid sequence information essential for progenitor toxin/neurotoxin identification. When coupled with nanoscale HPLC (nLC) and electrospray ionization (nESI) techniques, mass spectrometry has the ability to identify proteins definitively at nanomole levels (1). This would be important for botulinum progenitor toxins because they are produced and active at low levels. However, factors such as protein variability, database composition, and search engine characteristics may influence identification, especially at the strain level.

This study examined the ability of MS combined with protein database searching to characterize botulinum progenitor toxins and to assign proteins to specific botulinum serotypes and strains based upon currently available sequence information. First, proteins were identified to establish progenitor toxin composition. Next, protein identification was used to assign progenitor toxins to serotypes and strains. Two major concerns regarding progenitor toxin strain assignment involve the availability of nucleotide/amino acid sequence information for individual *C. botulinum* strains and the level of amino acid sequence identity among NAPs from different strains. In recent reports, van Baar et al. (29, 30) addressed some of these issues when they characterized neurotoxin samples from serotypes A (strain 62A), B (strain Okra), C (003-9), D (CB-16), E (no designation; equivalent to NCTC 11219 by analysis), and F (Langeland) that also contained some progenitor toxin proteins. They analyzed these samples with matrix-assisted laser desorption ionization (MALDI)-MS and capillary LC-ESI-MS methods originally developed for tetanus toxin analysis (28). They showed that accurate strain assignments were possible when genetic sequences were available. Otherwise, toxin proteins matched the same class of toxin protein from other strains of the same serotype, except for HA-70 of serotype B. We have expanded upon these efforts and, in this study, examined the ability of nLC-nESI-MS/MS and protein database-searching techniques to characterize botulinum progenitor toxins by identifying the constituent proteins from serotypes A through G and to assign the proteins accurately to botulinum serotypes and strains.

MATERIALS AND METHODS

A word of caution. Due to their extreme toxicity, microgram quantities of botulinum progenitor toxins must be handled in approved laboratories only and strict safety and regulatory measures for toxin acquisition, storage, use, containment, and destruction must be observed. Individuals handling toxins in this study were vaccinated with an investigational anti-BoNT vaccine.

Progenitor toxins from the following strains of serotypes A through G were purchased from Metabiologics, Inc. (Madison, Wis.) and used without further purification: Hall (A), Okra (B), Stockholm (C1), MDPH (D), Alaska (E), Langeland (F), and 89 (G). Complexes were dialyzed extensively against and stored in 0.015 M Tris–0.1 M NaCl, pH 7.2, at −70°C until used. Porcine trypsin and proteinase K were acquired from Promega (Madison, Wis.) and Calbiochem (La Jolla, Calif.), respectively. Dithiothreitol, iodoacetamide, and ammonium bicarbonate were purchased from Sigma Chemical Co. (St. Louis, Mo.). Burdick & Jackson HPLC water was used to prepare all solutions and was acquired from VWR, Inc. (W. Chester, Pa.).

Each progenitor toxin sample was analyzed in triplicate. Before enzymatic digestion, 30 μg (total protein) of each complex was diluted with 50 μl of 8 M urea dissolved in 0.4 M ammonium bicarbonate and incubated at 23°C for 15 min. Samples were reduced with 5 μl of 0.045 M dithiothreitol and heating at 60°C for 15 min. After cooling to 23°C, samples had 5 μl of 0.1 M iodoacetamide added and then were incubated in the dark at 60°C for 15 min. Samples were cooled to 23°C again and diluted with 100 μl of HPLC water. Enzymatic digestions were performed at 37°C for 15 h after the addition of either 20 μg trypsin or 20 μg of proteinase K reconstituted separately in 50 μl of HPLC water. Enzymatic digestion was stopped with 30 μl of 0.1% formic acid. Digests were stored at −20°C until analyzed by MS.

Peptide data were acquired by using a Micromass Quadrupole-time of flight 2 mass spectrometer (Altrincham, United Kingdom) equipped with a New Objective (Woburn, Mass.) nanospray holder interfaced with a Water's CapLC high-performance liquid chromatograph (Milford, Mass.). Aliquots of 2.5 μl, containing approximately 375 ng of initial total protein, were injected onto a New Objective capillary LC column (10 cm by 75 μm) packed with Aquasil C_{18} (5-μm particle size, 100-Å pore size). Peptides were eluted with a flow rate of 500 nl/min and the following gradient: 0 to 20% B in 40 min (linear), 20 to 60% B in 40 min (linear), and 60 to 100% B in 20 min (linear). Solvents A and B consisted of 2% and 80% (vol/vol) acetonitrile in 0.1% (vol/vol) formic acid, respectively. The electrospray voltage was 2.0 kV, and the sample cone voltage was 35 V. Gas-assisted nebulization was not used. The mass spectrometer was operated in survey mode using the instrument's automatic switching feature to capture full-scan spectra (*m/z* 400 to 1,500 in 1.5 s) and product ion spectra (*m/z* 100 to 1,500 in 1.5 s). Product ion spectra were generated from multiply charged precursor ions with variable collision energies ranging from 10 to 60 eV based upon the mass and charge state of the eluting peptide. Argon was used as the collision gas at a nominal pressure of 1 bar.

Mascot (v1.0; Matrix Sciences, London, United Kingdom) database search engine was used to identify proteins. Several different search parameters were used for each variable in these search engines. The final search settings used in this study represented a compromise between including too many extraneous proteins and excluding known proteins. Mascot database MS/MS searches were

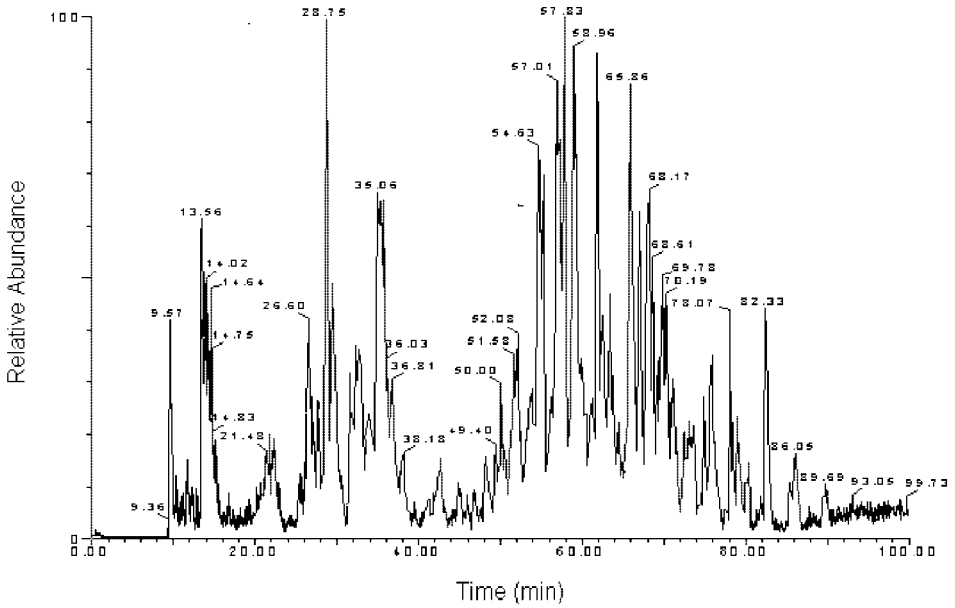

FIG. 1. Representative base peak chromatogram of a trypsin digestion of botulinum serotype C1 progenitor toxin. The following nLC conditions were used: Aquasil C18 column, 10 cm by 75 μm; flow rate of 500 nl/min; injection volume of 2.5 μl; gradient of linear segments of 0 to 20% B in 40 min, 20 to 60% B in 40 min, and 60 to 100% B in 20 min; for solvent A, 2% acetonitrile in 0.1% (vol/vol) formic acid; for solvent B, 80% acetonitrile in 0.1% (vol/vol) formic acid.

performed using peak list files generated by MassLynx (v3.5) and the following parameters: the National Center for Biotechnology Information nonredundant (NCBInr) database (v190304), eubacteria, trypsin, carbamidomethyl modification, methionine oxidation, 2+/3+ charge state, ±0.5-Da peptide tolerance, ±0.3-Da MS/MS mass tolerance, Micromass data format, and the ESI-Quad-time of flight instrument. Mowse scores greater than 38 indicated identity or extensive homology, but peptides with lower scores were used for identification if they matched peptides from the tentatively identified proteins exactly using BLAST searches (NCBInr database, exact matches for short sequences, PAM 30 matrix).

RESULTS

nLC-nESI-MS(/MS). Stringent tryptic and proteinase K digestion conditions were used to inactivate neurotoxins and to liberate a maximum number of peptides from each constituent protein. This was important, as HA-33 is reportedly resistant to trypsin digestion under nondenaturing conditions (8). Lower concentrations of dithiothreitol and iodoacetamide were used so that the enzymes could be added directly to the mixture to avoid potential protein losses that may occur when reagents are removed before digestion and to shorten analysis time. Using these lower reagent concentrations prompted the use of higher enzyme/protein mass ratios to promote protein cleavage. The success of this approach is shown in Fig. 1, which contains a representative base peak chromatogram for the tryptic digest of botulinum serotype C1 (strain Stockholm) progenitor toxin. Proteinase K digestion liberated many peptides also (data not shown), but random cleavages produced by this enzyme increased database search times to prohibitive levels. Consequently, proteinase K digestions were discontinued.

Figure 1 also indicated that tryptic peptides were present in a range of apparent relative abundances, which raised the question of how much peptide was needed to identify a protein. We determined that the product ion spectrum produced from 50 ng/ml of a standard peptide (Glu-fibrinogen) contained ions sufficient to identify correctly all amino acids in the standard peptide (Fig. 2).

Progenitor toxin protein identifications. Representative results obtained for progenitor toxin proteins from each botulinum serotype are listed in Table 2 (see the supplemental material for specific peptide information). In this report, identified proteins refer to precursor proteins typically contained in protein databases and HA-70 is used to refer to this hemagglutinin and its posttranslational fragments such as HA-52. A minimum of three peptides was required to match identical amino acid sequences in the protein database to identify a protein. Protein identifications were reproducible for each sample.

Serotype A progenitor toxin (Hall strain). (i) Protein composition. Comparing identified proteins contained in Table 2 with the list of known progenitor toxin proteins (Table 1) showed that all the proteins known to comprise serotype A progenitor toxin were identified. In addition, flagellins, not previously reported to be associated with progenitor toxins, were detected in this sample of serotype A progenitor toxin. *C. novyi* flagellin was identified with two different amino acid sequences. Although using two peptides instead of three peptides for protein identification was lower than the required number, the identification was included because Dineen et al. reported that putative flagellin genes are located directly upstream from the progenitor toxin gene cluster, adjacent to the HA-70 gene in type A strains 62A, Hall A-*hyper*, and NCTC 2916 (5). However, flagellar amino acid sequences obtained in this study for serotype A, strain Hall, did not match those reported by Dineen et al. for serotype A strains of Hall A-*hyper* plus 62A (gi:217025444), and NCTC 2916 (gi:21702565) (5).

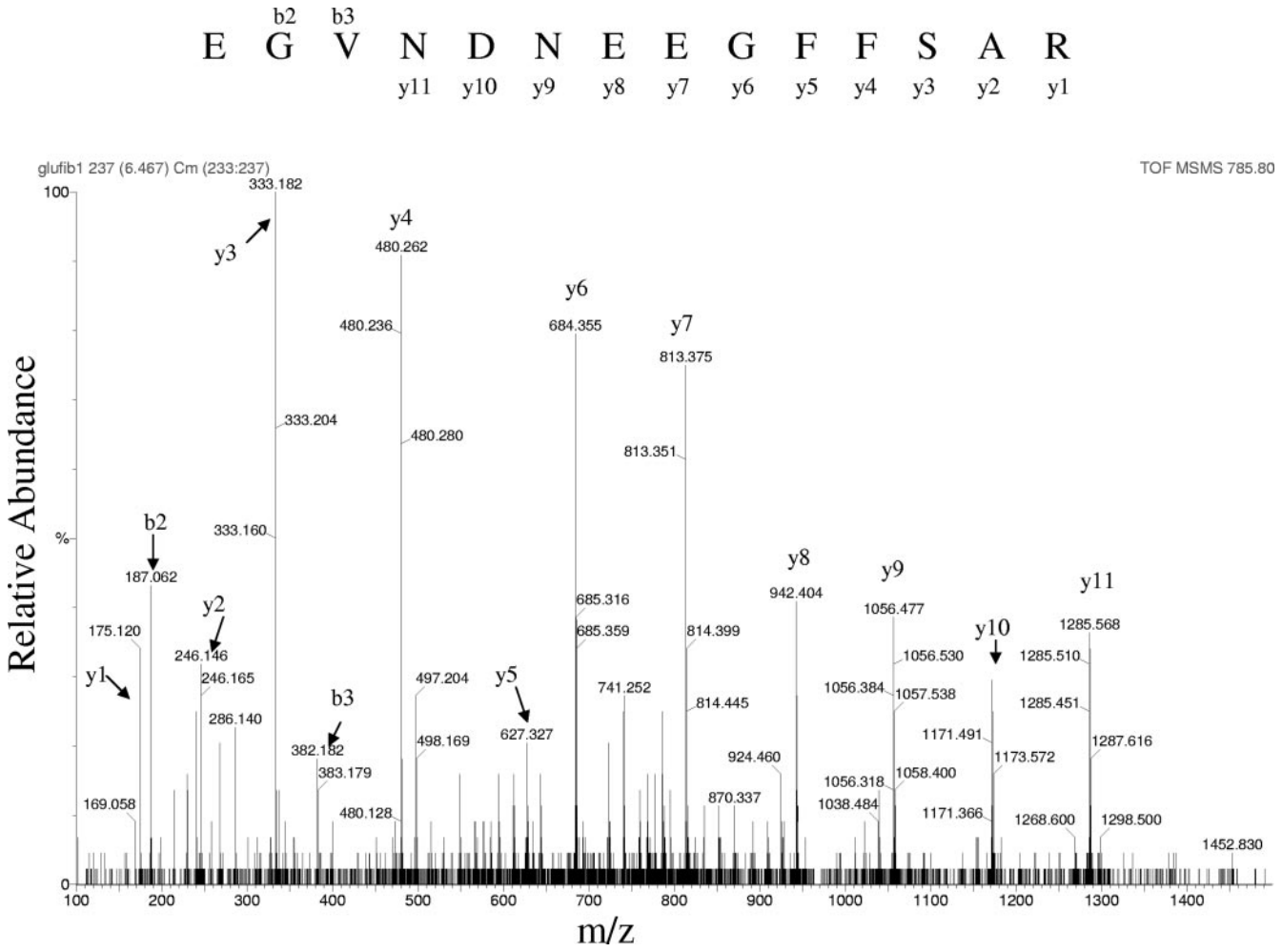

FIG. 2. Product ion spectrum of standard Glu-fibrinogen peptide (50 ng/ml) demonstrating that all ions necessary for identification were present for this peptide. The amino acid sequence and detected product ions are depicted in this figure also. The nLC conditions were the same as those described in the legend to Fig. 1.

BLASTp searches for several tryptic peptides selected randomly from the flagellin sequences submitted by Dineen et al. (5) matched only their submissions, possibly because only partial nucleotide sequences were submitted. BLASTp comparisons for the *C. novyi* flagellar sequences revealed that these amino acid sequences occur in several clostridial, as well as other bacterial, species, including *C. acetobutylicum*, *C. tyrobutyricum*, *C. novyi*, *C. tetani*, *C. chauvoei*, and *C. haemolyticum*. Furthermore, two additional, single-peptide flagellins were identified that are not included in Table 2. One peptide sequence, (224)MEY TVVGLDIAAENLQAAESR(244) (gi:15639853), matched only *Treponema pallidum* flagellin, while another sequence, (65)NAQDSISLIQTAEGALNETHSILQR(89) (gi:2829686), matched *C. tyrobutyricum* flagellin.

(ii) Serotype and strain assignment. Tryptic neurotoxin peptides from this sample of progenitor toxin matched type A1 toxin from strain 62A (gi:279630), with 21 peptides providing 18.0% coverage of the amino acid sequence. Generally, serotype A NAPs contained in this sample matched serotype A progenitor toxin NAP sequences. One apparent exception occurred for HA-17, where serotype Bnp (np, nonproteolytic) protein from Eklund 17B was identified. Two peptides were used to identify serotype Bnp HA-17 in the NCBInr database. This low number of peptides was inadequate to conclusively identify this protein, despite the 12.0% amino acid coverage provided by these peptides. Closer examination of these two peptides using BLASTp for exact matches of short peptide sequences showed that the first peptide, (58)ISNVAEPKN (66), is common to HA-17 in most botulinum serotypes, including A and B. The second peptide, (18)TFLPDGNYNIK (28), matched only Eklund 17B, as indicated in Table 2. However, D22 in serotype B HA-17 corresponds to N22 in serotype A HA-17, raising the possibility of a misleading strain assignment due to deamidation.

Serotype B progenitor toxin (Okra strain). (i) Protein composition. All the proteins known to comprise serotype B progenitor toxin (Table 1) were identified in this sample or serotype B progenitor toxin (Table 2). In addition, flagellins, not previously reported to be associated with progenitor toxins, were detected in this sample of serotype B progenitor toxin, supporting the results obtained for serotype A. The *C. novyi* match was made with the experimental sequence (65)NAQD

TABLE 2. Strongest progenitor toxin protein identities for serotypes A through G based upon the greatest number of peptides used for identification

Botulinum serotype	ID[b]	Accession no.	Strain(s)/organism	No. of peptides[c]	% Coverage
A[a] (Hall)					
BoNT	A1	279630	62A	21	18.0
NTNH	A1	2127324	NIH	11	11.2
HA-70	A1	21702546	62A	8	12.5
	A1	840638	62A	5	9.7
HA-33	A1	2127323	Hall	7	28.0
HA-17	Bnp	2104808	Eklund 17B	2	12.7
Other	Flagellin	19910967	*C. novyi*	2	12.6
B (Okra)					
BoNT	B	399134	NCTC 7273	26	23.5
NTNH	B	1619270	NCTC 7273	17	14.7
HA-70	A1	21702546	62A	13	24.3
HA-33	B	2145609	NCTC 7273	11	11.6
HA-17	A/B	840638	NCTC 2916	3	22.6
Other	Flagellin	19910969	*C. novyi*	1	9.3
	Flagellin	20807004	*Thermoanaerobacter tencongensis*	1	3.6
C (Stockholm)					
BoNT	C1	538636	Stockholm	36	33.3
NTNH	C1	1085641	C-468 (phage)	34	22.3
HA-70	C1	1170162	Stockholm	28	51.0
HA-33	C1	1085643	Stockholm	13	54.5
HA-17	C1	1170160	Stockholm	3	33.1
D (MDPH)					
BoNT	D	6939795	D-4947	18	18.6
NTNH	D	6939794	D-4947	19	18.9
HA-70	D	6939791	D-4947	17	35.3
HA-33	D	6939793	D-4947	10	22.3
HA-17	D	1075951	NI[d]	2	15.9
Other	Flagellin	19910967	*C. novyi*	5	27.4
	Flagellin	19910971	*C. novyi*	3	17.0
E (Alaska)					
BoNT	E	98569	Beluga, NCTC 11219	39	36.5
NTNH	E	1168713	Mashike	46	59.0
Other	ORF-X1	2897696	Iwana, *C. butyricum*	3	22.9
	NBP	4097887	Alaska (E)	2	48.8
F (Langeland)					
BoNT	F	529984	Langeland	25	21.9
NTNH	F	3757741	Langeland, *C. barati*	22	23.0
G (89)					
BoNT	G	2499920	113/30	14	16.1
NTNH	G	2104804	ATCC 27322	9	9.8
HA-70	Bnp/G	2104801	ATCC 27322	15	37.9
HA-17	G	2104802	ATCC 27322	5	49.0

[a] Serotype (strain used in this study).
[b] ID, serotype or protein identified.
[c] Number of peptides used for identification.
[d] NI, not indicated.

GISLIQTAEGALNETHAILQR(89), which is also found in flagellin from *C. haemolyticum*, *C. chauvoei*, *C. thermocellum*, *C. novyi*, and *C. tetani*, as well as a number of other bacterial species. This sequence from the *C. tyrobutyricum* sequence differed from this sequence with an S5G substitution. The second sequence, (124)IASTTQFNTR(133), matched only *Thermoanaerobacter* entries exactly. *Thermoanaerobacter* belongs to the class *Clostridia*.

(ii) Serotype and strain assignment. Neurotoxin, NTNH, HA-33, and HA-17 proteins in this sample of serotype B, strain Okra, progenitor toxin matched entries from other serotype B strains or serotypes containing silent genes for type B progenitor toxin (Table 2), as no Okra strain sequence data are available in the NCBInr database. HA-70 was an exception to serotype matching. Thirteen peptides (24.3% coverage) apparently matched type A serotype (strain 62A) HA-70. Closer examination of the peptides used for the match showed that seven peptides were common to both serotypes A and B. The remaining six peptides were used to identify serotype A. Based upon BLASTp results for 100% matching, five of the six remaining peptides were found exclusively in serotype A strains such as 62A and Hall A-*hyper*. One of these five peptides was also found in the type A/B strain NCTC 2916. The sixth peptide, (154)SIEFNPGEK(162), matched strains 62A, Hall A-

hyper, and NCTC 2916 and a B/F strain of serotype B, CDC 3281, indicating it is common to strains from serotypes A and B.

Serotype C progenitor toxin (Stockholm strain). (i) Protein composition. Table 2 lists proteins identified in this sample of serotype C progenitor toxin. Comparing Table 2 to Table 1 shows that all the proteins known to comprise serotype C progenitor toxin were identified. No additional proteins were detected in this sample.

(ii) Serotype and strain assignments. Proteins from this sample of serotype C1, strain Stockholm, progenitor toxin matched known serotype C1 proteins (Table 2). Strain assignments were consistent for this serotype because nucleotide information is available for the following progenitor toxin proteins: for BoNT, X62389; for NTNH, X62389; for HA-70, D38562; for HA-33, X5301; and for HA-17, X62389.

Serotype D progenitor toxin (MDPH strain). (i) Protein composition. All proteins known to comprise serotype D progenitor toxin were identified in this sample of serotype D toxin (Tables 1 and 2). Flagellins were also associated with this isolate of strain MDPH progenitor toxin. Two forms of flagellin, FliA(A) and FliA(B), matched *C. novyi* flagellin database entries by using five and three unique peptides, respectively. As before, these proteins are common to several bacterial species, but the larger number of identified peptides substantiates the flagellin identified for samples from serotypes A and B.

(ii) Serotype and strain assignment. Serotype D, strain MDPH, neurotoxin matched neurotoxin entries from other D serotype strains (Table 2) because no nucleotide data are available for this strain. The smallest protein in this progenitor toxin, HA-17, again proved difficult to identify because only two peptides were used for the assignment.

Serotypes E and F progenitor toxins (Alaska and Langeland strains, respectively). (i) Protein composition. Each protein known to comprise serotype E and F progenitor toxins, neurotoxin and NTNH, were identified in these samples. However, additional proteins were identified in serotype E progenitor toxin samples but not in serotype F samples. Three peptides matched ORF-X1 from *C. botulinum* serotype E and provided 23.2% coverage of the protein's 142 amino acid residues, which was reported for strain Iwanai and *C. butyricum*. BLASTp searches for each of the three detected peptides matched only ORF-X1, although entries for ORF-X1 from serotypes A2 (Kyoto-F) and F (Langeland) are also present in the NCBInr database. The gene that produces serotype E ORF-X1 is found upstream from the NTNH gene (34). However, the protein was not detected in serotype F and serotype A2 was not used in this study. Two other proteins, ORF-X2 and p47, are also produced by E and F serotypes, but neither protein was detected. A second protein designated neurotoxin-binding protein (NBP; gi:4097887), having a possible transcriptional regulatory function, was also identified in the serotype E complex. This protein consists of 43 amino acids, is an unpublished, direct submission to the NCBInr database, and was reported for the Alaska strain of *C. botulinum* serotype E. Two peptides comprising 48.8% (21/43 residues) of the submitted amino acid sequence were identified for the Alaska strain used in our study. Although only two peptides were used for identification, strain specificity and high amino acid coverage justified its inclusion in Table 2.

(ii) Serotype and strain assignment. Serotype E neurotoxin and NTNH were assigned to various strains because no sequence data were available in the NCBI database for this strain. NBP was isolated from the Alaska strain and was assigned properly in this study. On the other hand, serotype F neurotoxin and NTNH were assigned to the Langeland strain because each protein in the progenitor toxin has a database entry. The entries are S76749 for the BoNT and X99064 for NTNH.

Serotype G progenitor toxin (89 strain). Strain 89 progenitor toxin proteins matched BoNT, NTNH, HA-70, and HA-17 produced by other serotype G strains (Table 2). HA-33 is not produced in this serotype. While some sequence information is available for most serotypes, very little information is available for serotype G progenitor toxin in the NCBI database. One entry (X87972) covers the nucleotide sequences of NTNH, HA-70 (partial), and HA-17 present in the *C. argentinense* progenitor toxin gene cluster, and another entry contains the neurotoxin nucleotide sequence from *C. botulinum* strain 1113/30 or NCBF 3012. These entries were used to identify strain 89 used in our study. Therefore, substantial amino acid identities must exist among the different progenitor toxin proteins of this strain and those contained in the database to make these matches. The paucity of sequence data precludes accurate strain assignment, although the serotype assignment was accurate.

DISCUSSION

van Baar et al. proposed that protein toxins can be unambiguously identified with mass spectrometry (29), and they applied this premise to the analyses of tetanus (28) and botulinum toxins (29, 30). Their MALDI-MS and LC-ESI-MS analyses of botulinum serotypes A through F identified neurotoxin and each protein present in their toxin samples. They also showed that ESI-MS with CID (ESI-MS/MS) or MALDI-MS with postsource decay identified protein unambiguously (29, 30). Several neurotoxin samples they purchased included some known progenitor proteins, but they did not always include each progenitor toxin protein in each sample. They were reportedly unable to purchase serotype G toxin. So, complete progenitor toxin protein data were not provided for serotypes B, C, D, and G.

In our study, we analyzed complete progenitor toxins from serotypes A through G. Five of the strains we used in this study differed from those analyzed by van Baar et al. (29, 30), who used strains 62A, Okra, 003–9, CB-16, no designation (equivalent to NCTC 11219 by analysis), and Langeland from serotypes A through F, respectively (30). Okra and Langeland strains from serotypes B and F, respectively, overlapped those used in this study.

We used LC-ESI-MS/MS to analyze progenitor toxins from each botulinum serotype to provide the best possible protein identifications. Using nLC-nESI in our study represented a refinement of the capillary LC and micro-ESI techniques employed by van Baar et al. (28–30). Because sample requirements are lower for nLC-nESI, total protein quantities used in this study were approximately 40% lower than quantities used by van Baar et al. (29, 30).

Progenitor toxin protein composition. The first goal of this study was to characterize progenitor toxins from one strain of each serotype by identifying proteins in the toxin. We made no effort to further purify or separate different sizes of progenitor toxin to determine the total protein composition of each sample. To date, most progenitor toxin studies have concentrated upon gene clusters or purifying progenitor toxin for further study, but not upon the toxin's final protein composition. The ability to characterize the final protein composition is needed to study progenitor toxin formation and progenitor toxin aging. It is especially needed for detection and assignment purposes because these will be affected by extraction conditions that may alter protein composition, by processing/treatment of the extract that may alter the physical condition of the constituent proteins, and by analytical instrumentation characteristics and parameters that may be required to differentiate nearly identical amino acid sequences. We found that mass spectrometry identified accurately all known proteins in botulinum progenitor toxins from each serotype. During the identification process, it became apparent that additional information, such as BLAST searches, was needed occasionally to differentiate two possible identifications because some proteins from different strains within a serotype possessed nearly identical amino acid sequences.

In addition to the known protein components of botulinum progenitor toxins, we found that proteins produced by genes near or within the progenitor toxin gene complex can coisolate with progenitor toxins (5). One example was flagellin, which was identified in extracts from several serotypes. Whether or not flagellin was bound to progenitor toxins was not determined, but its presence in extracts from three different serotypes indicates that association may have occurred. Flagellin's presence is interesting because Dineen et al. postulated that the proximity of flagellin genes to the progenitor toxin gene cluster in several serotype A strains may be important because of the interaction of flagellin with sigma factors that affect gene regulation and the possible involvement of flagellin in protein secretion (5). Additional studies will be needed to determine if flagellin coextracts only or is associated specifically or nonspecifically with the progenitor toxin and to determine if a specific association involves one or more of the constituent proteins.

In one of their studies, van Baar et al. also detected exoenzyme C3 in samples from serotype C1, strain 003-9 (30). Although we failed to detect this enzyme, we did detect ORF-X1 and NBP in serotype C1, strain Stockholm, samples. The presence of these proteins and the flagellins in samples from separate suppliers indicates that different isolation/purification procedures can affect the final protein composition of the progenitor toxin.

Serotype and strain assignments. The second goal of this study was to determine if protein identities could be used to assign accurate serotypes and strains. Examination of the assignments showed that neurotoxin tryptic peptides matched the correct serotype for each sample and represented a reliable marker for serotype identification. Strain assignments were also accurate when sequence data were available for the strain, as indicated by van Baar et al. (29, 30). For example, serotype A, C, and F strain assignments matched database entries because neurotoxin sequences are available. Conversely, serotype B neurotoxin peptides matched the NCTC 7273 entry because no sequence data are available for strain Okra in the NCBI database. Strains Okra and NCTC 7273 must share a high level of neurotoxin amino acid identity because 26 Okra strain peptides matched NCTC 7273 peptides in this study.

Strain assignments can be difficult, especially when proteins from different strains share high amino acid sequence identities. For example, multiple database entries are available for neurotoxins produced by strains of serotype A and several of these neurotoxins share high amino acid sequence identities. Zhang et al. compared available botulinum progenitor toxin amino acid identities for several strains from each botulinum serotype (33). When serotype A, strain Hall/Allergan, neurotoxin was used as the reference amino acid sequence, it differed from 62A, Hall A/ATCC 3502, NCTC 2916, and Kyoto-F (A2 toxin) neurotoxin amino acid sequences by 0, 1, 0.2, and 10%, respectively. This indicates that it will not be possible to distinguish strains Hall/Allergan and 62A based solely upon neurotoxin peptides and that differentiating between strains Hall/Allergan, Hall A/ATCC 3502, and NCTC 2916 (A toxin) may be challenging but possible with mass spectrometry. This was supported when neurotoxin peptides used in this study to match strain 62A neurotoxin A1 also matched strain Hall/Allergan peptides. This situation exists for neurotoxins from all serotypes with multiple strains.

Using NAPs in addition to neurotoxins should provide additional information for definitive serotype/strain assignments, but similar complexities also exist for these proteins. For example, the serotype A strain assignment for Hall strain NTNH used in this study was the National Institutes of Health (NIH) strains. Zhang et al. reported 100% amino acid identity between the Hall A/Allergan and NIH strains (33). Dineen et al. (5) reported that NTNHs from the Hall A-*hyper* and NIH strains are identical and that NTNHs from these strains differ from strain 62A by only 1 amino acid substitution (G333E), which also corresponds to the 99.9% identity presented by Zhang et al. (33).

One anomalous serotype assignment occurred when HA-70 protein from serotype B, strain Okra, was assigned to serotype A, strain 62A (Table 2). As no sequence information is available for serotype B, strain Okra, it was not possible to know how this strain relates to other strains within the serotype. However, comparing HA-70 amino acid sequence identities for other strains indicated that only partial sequences are available for most serotypes. Dineen et al. used residues 1 through 442 for identity comparisons because only partial sequence data were available for the strains compared (5). In their analyses, van Baar et al. (29) indicated that, although HA-70 from the Okra strain of serotype B appeared to be similar to HA-70 contained in the silent gene of strain NCTC 2916 (serotype A), it appeared to differ from other known serotype B HA-70 proteins, especially in the posttranslational HA-52 N-terminal region. Dineen et al. reported that 62A and Hall A-*hyper* HA-70 shares 99% identity with serotype A strain NCTC 2916 and serotype BF strain 3281 and 96% identity with serotype B strain Eklund 17B for available residues (5). These close similarities illustrate how an anomalous assignment may occur.

The serotype A HA-17 match with the nonproteolytic B strain Eklund 17B represented in Table 2 was anomalous due to the low number of peptides used to identify it. Because few peptides were used and close similarities exist among the A/B

and C/D serotypes, changing 1 amino acid affected an assignment when Eklund 17B matched serotype A HA-17. One peptide used for the match, ISNVAEPNK, is also part of HA-17 found in strains NCTC 2916 (gi:840638), Hall A-*hyper* (gi:21702556), and 62A (gi:21702547), as well as a B/F strain (gi:380578). The distinguishing peptide had the sequence TF LPDGNYNIK. The aspartic acid at position 5 (underlined) caused the match with strain Eklund 17B HA-17, as most serotype A HA-17 proteins have an asparagine at that position. BLASTp searches against serotype B HA-17 protein (gi:2104808) found five database entries with identities equal to or greater than 94%. Strains Hall A-*hyper*, 62A, NCTC 2916, and CDC 3281 share 95% amino acid identity with the Eklund 17B strain, and the Lamanna strain of the B serotype shares 94% identity. Generally, strains possessing A serotype HA-17 matched best with the Hall A serotype strain used in this study, although some B serotype proteins are very similar. For example, the amino acid sequences for HA-17 from type B/F strain CDC 3281 and the type A protein are identical (5). Comparing the 146 residues of gi:2104808 to other entries with 95% amino acid identity revealed seven residue/alignment differences, whereas 94% matches had eight differences (4). Amino acid identity results among different serotypes and strains included in this study corresponded to those reported by Dineen et al. for HA-17 (5).

One difficulty encountered during these analyses involved detecting HA-17 in samples from each serotype known to produce this protein. Quantitative data concerning NAPs in 12S and 16S progenitor toxins are available for serotype D, strain 4947 (16). For this serotype, four heterodimers of HA-33/HA-17 are present in both 12S and 16S progenitor toxins. Similar protein stoichiometries seem to occur in serotype C (15), but this has not been documented for the other serotypes. Therefore, an adequate amount of HA-17 should be available for detection, assuming that progenitor toxins containing HA-17 are assembled in a similar manner by other serotypes. Closer examination of known amino acid sequences for serotypes producing HA-17 showed that trypsin cleavage of serotypes A, B, and G would produce 9, 8, and 9 peptides, respectively, which would be most useful for identification by mass spectrometry (optimum lengths between 5 and 20 amino acid residues). On the other hand, only three optimal tryptic peptides could be generated from serotypes C and D HA-17. Three other peptides comprised of 25, 28, and 51 residues could be generated with tryptic digestion (data not shown). Therefore, using a second proteolytic enzyme would be helpful to identify HA-17 in serotypes C and D. Pepsin was used by van Baar et al. (29, 30) and provided useful peptides, which seems to be a better compromise than the proteinase K we used in our study. Changes in reaction conditions could limit the proteinase K cleavages to reduce the number of peptides that must be considered during protein searches.

Overall, these results demonstrate that analyses based upon nanoscale chromatographic and ionization techniques benefit *C. botulinum* progenitor toxin analysis by lowering the quantities required for analysis. Proteins from the seven serotypes of *C. botulinum* progenitor toxins were identified successfully using this approach. Neurotoxin database identifications were generally accurate and conclusive for a given serotype, but careful analysis was needed for certain proteins to ensure accuracy. Strain assignment can be challenging due to the nearly identical amino acid sequences shared by NAPs from different serotypes and strains. While nucleotide sequence data are the best references for matching experimental data to strains, it is not essential to assign proteins to a serotype accurately. Furthermore, despite high levels of amino acid similarities, the mass spectrometer is able to differentiate sequences and serotypes that differ by only one amino acid, within the known constraints of an instrument. Future efforts based upon these results will include assessing detection limits for progenitor toxins, establishing a proteolytic peptide library for different strains, establishing identification limits for progenitor toxins, and determining stoichiometric relationships among proteins comprising progenitor toxins with mass spectrometry.

ACKNOWLEDGMENTS

Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the U.S. Army.

The research described herein was sponsored by the U.S. Army Medical Research Institute of Infectious Diseases, U.S. Army Medical Research and Material Command, project number 02-4-3U-058.

We thank Tony Garza and Kari Holman for their technical assistance.

REFERENCES

1. **Aebersold, R.** 2003. A mass spectrometric journey into protein and proteome research. J. Am. Soc. Mass Spectrom. **14:**685–695.
2. **Arnon, S. S., R. Schechter, T. V. Inglesby, D. A. Henderson, J. G. Bartlett, M. G. Ascher, E. Eitzen, A. D. Fine, J. Hauer, M. Lyton, S. Lillibridge, M. T. Osterholm, T. O'Toole, G. Parker, T. M. Perl, P. K. Russel, D. L. Swerdlow, and K. Tonat.** 2001. Botulinum toxin as a biological weapon: medical and public health management. JAMA **285:**1059–1070.
3. **Cai, S., H. K. Sarkar, and B. R. Singh.** 1999. Enhancement of the endopeptidase activity of botulinum neurotoxin by its associated proteins and dithiothreitol. Biochemistry **38:**6903–6910.
4. **ClustalW.** http://www.ebi.ac.uk/clustalw/.
5. **Dineen, S. S., M. Bradshaw, and E. A. Johnson.** 2003. Neurotoxin gene clusters in *Clostridium botulinum* type A strains: sequence comparision and evolutionary implications. Curr. Microbiol. **46:**345–352.
6. **East, A. K., J. M. Stacey, and M. D. Collins.** 1994. Cloning and sequencing of a hemagglutinin component of the botulinum neurotoxin complex encoded by *Clostridium botulinum* types A and B. Syst. Appl. Microbiol. **17:**306–312.
7. **East, A. K., M. Bhandari, J. M. Stacey, K. D. Campbell, and M. D. Collins.** 1996. Organization and phylogenetic interrelationships of genes encoding components of the botulinum toxincomplex in proteolytic *Clostridium botulinum* type A, B, and F: evidence of chimeric sequences in the gene encoding the nontoxic nonhemagglutinin component. Int. J. Syst. Bacteriol. **46:**1105–1112.
8. **Fu, F.-N., S. K. Sharma, and B. R. Singh.** 1998. A protease-resistant novel hemagglutinin purified from type A *Clostridium botulinum.* J. Protein Chem. **17:**53–60.
9. **Fugita, R., Y. Fujinaga, K. Inoue, H. Nakajima, H. Kumon, and K. Oguma.** 1995. Molecular characterization of two forms of nontoxic-nonhemagglutinin components of *Clostridium botulinum* type A progenitor toxins. FEBS Lett. **376:**41–44.
10. **Fujinaga, Y., K. Inoue, S. Shimazaki, K. Tomochika, K. Tsuzuki, N. Fujii, T. Watanabe, T. Ohyama, K. Takeshi, K. Inoue, and K. Oguma.** 1994. Molecular construction of *Clostridium botulinum* type C progenitor toxin and its gene organization. Biochem. Biophys. Res. Commun. **205:**1291–1298.
11. **Fujinaga, Y., K. Inoue, S. Watanabe, K. Yokota, Y. Hirai, I. Nagamachi, and K. Oguma.** 1997. The haemagglutinin of *Clostridium botulinum* type C progenitor toxin plays an essential role in binding of toxin to the epithelial cells of guinea pig small intestine, leading to the efficient absorption of the toxin. Microbiology **143:**3841–3847.
12. **Fujinaga, Y., K. Inoue, T. Nomura, J. Sasaki, J. C. Marvaud, M. R. Popoff, S. Kozaki, and K. Oguma.** 2000. Identification and characterization of functional subunits of *Clostridium botulinum* type A progenitor toxin involved in binding to intestinal microvilli and erythrocytes. FEBS Lett. **467:**179–183.
13. **Inoue, K., Y. Fujinaga, T. Watanabe, T. Ohyama, K. Takeshi, K. Moriishi, H. Nakajima, K. Inoue, and K. Oguma.** 1996. Molecular composition of *Clostridium botulinum* type A progenitor toxins. Infect. Immun. **64:**1589–1594.
14. **Johnson, E. A., and M. Bradshaw.** 2001. *Clostridium botulinum* and its neurotoxins: a metabolic and cellular perspective. Toxicon **39:**1703–1722.

15. **Kouguchi, H., T. Watanabe, Y. Sagane, and T. Ohyama.** 2001. Characterization and reconstitution of functional hemagglutinin of the *Clostridium botulinum* type C progenitor toxin. Eur. J. Biochem. **268:**4019–4026.
16. **Kouguchi, H., T. Watanabe, Y. Sagane, H. Sunagawa, T. Ohyama.** 2002. *In vitro* reconstitution of the *Clostridiium botulinum* type D progenitor toxin. J. Biol. Chem. **277:**2650–2656.
17. **Li, L., B. Li, S. N. Parikh, R. B. Lomenth, and B. R. Singh.** 1997. A novel type E *Clostridium botulinum* neurotoxin progenitor complex. Protein Sci. **6**(Suppl. 2)**:**139T.
18. **Marvaud, J. C., M. Gilbert, K. Inoue, V. Fujinaga, K. Oguma, and M. R. Popoff.** 1998. *botR* is a positive regulator of botulinum neurotoxin and associated nontoxic protein genes in *Clostridium botulinum* A. Mol. Microbiol. **29:**1009–1018.
19. **Nakajima, H., K. Inoue, T. Ikeda, Y. Fujinaga, H. Sunagawa, K. Yakesi, T. Ohyama, T. Watanabe, K. Inoue, and K. Oguma.** 1998. Molecular composition of the 16S toxin produced by a *Clostridium botulinum* type D strain, 1873. Microbiol. Immunol. **42:**599–605.
20. **Oguma, K., K. Inoue, Y. Fujinaga, K. Yokota, T. Watababe, T. Ohyama, K. Takeshi, and K. Inoue.** 1999. Structure and function of *Clostridium botulinum* progenitor toxin. J. Toxicol. Tox. Rev. **18:**17–34.
21. **Oguma, K., Y. Fujinaga, and K. Inoue.** 1995. Structure and function of *Clostridium botulinum* toxins. Microbiol. Immunol. **39:**161–168.
22. **Ohishi, I., and G. Sakaguchi.** 1980. Oral toxicities of *Clostridium botulinum* type C and D toxins of different molecular sizes. Infect. Immun. **28:**303–309.
23. **Sagane, Y., T. Watanabe, H. Kouguchi, H. Sunagawa, S. Obata, K. Oguma, and T. Ohyama.** 2002. Spontaneous nicking in the nontoxic-nonhemagglutinin component of the *Clostridium botulinum* toxin complex. Biochem. Biophys. Res. Commun. **292:**434–440.
24. **Sakaguchi, G.** 1983. *Clostridium botulinum* toxins. Pharmacol. Ther. **19:**165–194.
25. **Shunji, S., I. Ohishi, and G. Sakaguchi.** 1977. Intestinal absorption of botulinum toxins of different molecular sizes in rats. Infect. Immun. **17:**491–496.
26. **Singh, B. R.** 2000. Intimate details of the most poisonous poison. Nat. Struct. Biol. **7:**617–619.
27. **Somers, E., and B. R. DasGupta.** 1991. *Clostridium botulinum* types A, B, C_1, and E produce proteins with or without hemagglutinating activity: do they share common amino acid sequences and genes? J. Protein Chem. **10:**415–425.
28. **van Baar, B. L. M., A. G. Hulst, and E. R. J. Wils.** 2002. Characterisation of tetanus toxin, neat and in culture supernatant, by electrospray mass spectrometry. Anal. Biochem. **301:**278–289.
29. **van Baar, B. L. M., A. G. Hulst, A. L. de Jong, and E. R. J. Wils.** 2002. Characterisation of botulinum toxins type A and B, by matrix-assisted laser desorption ionisation and electrospray mass spectrometry. J. Chromatogr. A **970:**95–115.
30. **van Baar, B. L. M., A. G. Hulst, A. L. de Jong, and E. R. J. Wils.** 2004. Characterisation of botulinum toxins type C, D, E, and F by matrix-assisted laser desorption ionisation and electrospray mass spectrtometry. J. Chromatogr. A **1035:**97–114.
31. **Watanabe, T., Y. Sagane, H. Kouguchi, H. Sunagawa, K. Inoue, Y. Fujinaga, K. Oguma, and T. Ohyama.** 1999. Molecular composition of progenitor toxin produced by *Clostridium botulinum* type C strain 6813. J. Protein Chem. **18:**753–760.
32. **Weller, U., M.-E. Dauzenrothe, D. Meyer zu Heringdorf, and E. Habermann.** 1989. Chains and fragments of tetanus toxin. Separation, reassociation, and pharmacological properties. Eur. J. Biochem. **182:**649–656.
33. **Zhang, L., W.-J. Lin, S. Li, and R. Aoki.** 2003. Complete DNA sequences of the botulinum neurotoxin complex of *Clostridium botulinum* type A-Hall (Allergan) strain. Gene **315:**21–32.
34. **Zhang, Z., and B. R. Singh.** 1995. A novel complex of type E *Clostridium botulinum.* Protein Sci. **4**(Suppl. 2)**:**110.
35. **Zilinskas, R. A.** 1997. Iraq's biological weapons. The past as future? JAMA **278:**418–424.