Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1035 (2004) 97–114
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Characterisation of botulinum toxins type C, D, E, and F by matrix-assisted laser desorption ionisation and electrospray mass spectrometry

Ben L.M. van Baar*, Albert G. Hulst, Ad L. de Jong, Eric R.J. Wils

TNO Prins Maurits Laboratory, Division Chemical and Biological Protection, P.O. Box 45, 2280 AA, Rijswijk, The Netherlands

Received 17 November 2003; received in revised form 16 February 2004; accepted 17 February 2004

Abstract

In a follow-up of the earlier characterisation of botulinum toxins type A and B (BTxA and BTxB) by mass spectrometry (MS), types C, D, E, and F (BTxC, BTxD, BTxE, BTxF) were now investigated. Botulinum toxins are extremely neurotoxic bacterial toxins, likely to be used as biological warfare agent. Biologically active BTxC, BTxD, BTxE, and BTxF are comprised of a protein complex of the respective neurotoxins with non-toxic non-haemagglutinin (NTNH) and, sometimes, specific haemagglutinins (HA). These protein complexes were observed in mass spectrometric identification. The BTxC complex, from *Clostridium botulinum* strain 003-9, consisted of a 'type C1 and D mosaic' toxin similar to that of type C strain 6813, a non-toxic non-hemagglutinating and a 33 kDa hemagglutinating (HA-33) component similar to those of strain C-Stockholm, and an exoenzyme C3 of which the sequence was in full agreement with the known genetic sequence of strain 003-9. The BTxD complex, from *C. botulinum* strain CB-16, consisted of a neurotoxin with the observed sequence identical with that of type D strain BVD/-3 and of an NTNH with the observed sequence identical with that of type C strain C-Yoichi. Remarkably, the observed protein sequence of CB-16 NTNH differed by one amino acid from the known gene sequence: L859 instead of F859. The BTxE complex, from a *C. botulinum* isolated from herring sprats, consisted of the neurotoxin with an observed sequence identical with that from strain NCTC 11219 and an NTNH similar to that from type E strain Mashike (1 amino acid difference with observed sequence). BTxF, from *C. botulinum* strain Langeland (NCTC 10281), consisted of the neurotoxin and an NTNH; observed sequences from both proteins were in agreement with the gene sequence known from strain Langeland. As with BTxA and BTxB, matrix-assisted laser desorption/ionisation (MALDI) MS provided provisional identification from trypsin digest peptide maps and liquid chromatography–electrospray (tandem) mass spectrometry (LC–ES MS) afforded unequivocal identification from amino acid sequence information of digest peptides obtained in trypsin digestion.

Keywords: Peptide mapping; Amino acid sequencing; Mass spectrometry; Toxins; Botulinum toxins

1. Introduction

The classes of chemical weapons (CW) and biological weapons (BW) are linked by the area of mid-spectrum agents (MSA). CW agents are typically highly toxic man-made, synthetic chemicals. BW agents are typically the weapons of nature: disease-causing organisms, like bacteria, viruses and rickettsiae, with the key ability of self-replication. MSA are highly toxic natural products, for example from bacteria or plants, that lack the ability of self-replication. In international treaties, for instance the Biological and Toxin Weapons Convention (BTWC, [1]), MSA are seen as belonging to the class of BW.

Theoretically, detection, identification and verification of MSA can be achieved by biochemical methods usually applied to BW and by analytical chemistry methods commonly applied to CW. Over the last few years, we have developed analytical chemistry methods for identification and verification of MSA, to find out whether the principle could be put to practice. Studies on mass spectrometric analysis of low molecular weight toxins (for example [2,3]), peptides [4], and protein toxins [5–7] have shown that neat MSA are amenable to chemical analysis, provided that typical bio-analytical methods are used and that attention is given to adequate sample handling. As a demonstration of the

* Corresponding author. Tel.: +31-15-2843514; fax: +31-15-2843963.
E-mail address: baar@pml.tno.nl (B.L.M. van Baar).

0021-9673/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2004.02.047

capabilities of mass spectrometry (MS) in identification and verification of MSA, we recently published a study on mass spectrometry of neat botulinum toxin (BTx) of serotypes A and B [8].

Presently, we report a follow-up study concerning neat BTx of serotypes C, D, E, and F. BTx types C and D are neurotoxins, but no human cases of botulism have been linked to these two serotypes. Therefore, it is not known if these types of BTx are toxic to humans. BTx types E and F are highly toxic to humans. BTx type E often occurs in vacuum packaged foodstuff, due to *Clostridium botulinum* type E contamination at the time of packaging (see, for example, [9]). BTx type F was the neurotoxin taken into production by Iraq, in its bioweapons program [10,11]. Characterisation of neat BTx was done by matrix assisted laser desorption/ionisation mass spectrometry (MALDI MS) and by coupled liquid chromatography and electrospray mass spectrometry (LC–ES MS) methods, in the laboratory. BTx type C (BTxC) was from *C. botulinum* strain 003-9, whereas type D (BTxD) was from strain CB-16, type E (BTxE) was from an unnamed strain isolated from herring sprats, and type F was from strain Langeland (NCTC 10281). Protein or nucleotide sequences from some of the proteins are available from independent sources. In Section 3, characterisation of the four toxins is discussed separately, with reference to relevant known sequence data.

2. Experimental

2.1. Notice of caution

BTxs are extremely toxic and their handling requires extensive safety measures. All handling of microgram quantities of intact BTx was performed in the containment of a glove box equipped with HEPA filters (biosafety level 3). No botulinum toxin vaccine has been accepted for use in The Netherlands.

2.2. Materials

BTx type C, from *C. botulinum* strain 003-9, type D, from strain CB-16, type E, from a strain isolated from herring sprats, and type F, from strain Langeland (NCTC 10281) were purchased from Calbiochem (La Jolla, CA, USA). Bis(2-mercaptoethyl)sulphone (BMS) was also obtained from Calbiochem. Tris and Tris–HCl (Trizma®), sodium acetate (p.A.), and sodium chloride (p.A.) were obtained from Sigma (Zwijndrecht, The Netherlands), whereas TPCK treated trypsin was obtained from Sigma–Aldrich (Steinheim, Germany). Guanidine–HCl and EDTA were purchased from Janssen (Geel, Belgium). Sodium 2-iodoacetate and ammonium hydrogencarbonate (NH_4HCO_3; >99.5%) were obtained from Fluka Chemie AG (Buchs, Switzerland). Bleach was available from laboratory stock as a solution of 25% technical quality sodium hypochlorite in tap water.

All water used in analytical procedures was drawn from a MilliQ system (Millipore, Milford, MA, US). Acetonitrile (LiChrosolv quality) and formic acid (p.A.) were purchased from Merck (Darmstadt, Germany). Recrystallised α-cyano-4-hydroxycinnamic acid, used as a MALDI matrix, was purchased from Bruker Daltonik (Bremen, Germany). Trifluoroacetic acid, biochemical grade, was purchased from Acros Organics (Geel, Belgium).

Molecular weight cut-off (MWCO) 30,000 and 100,000 filters and ZipTip™ were obtained from Millipore.

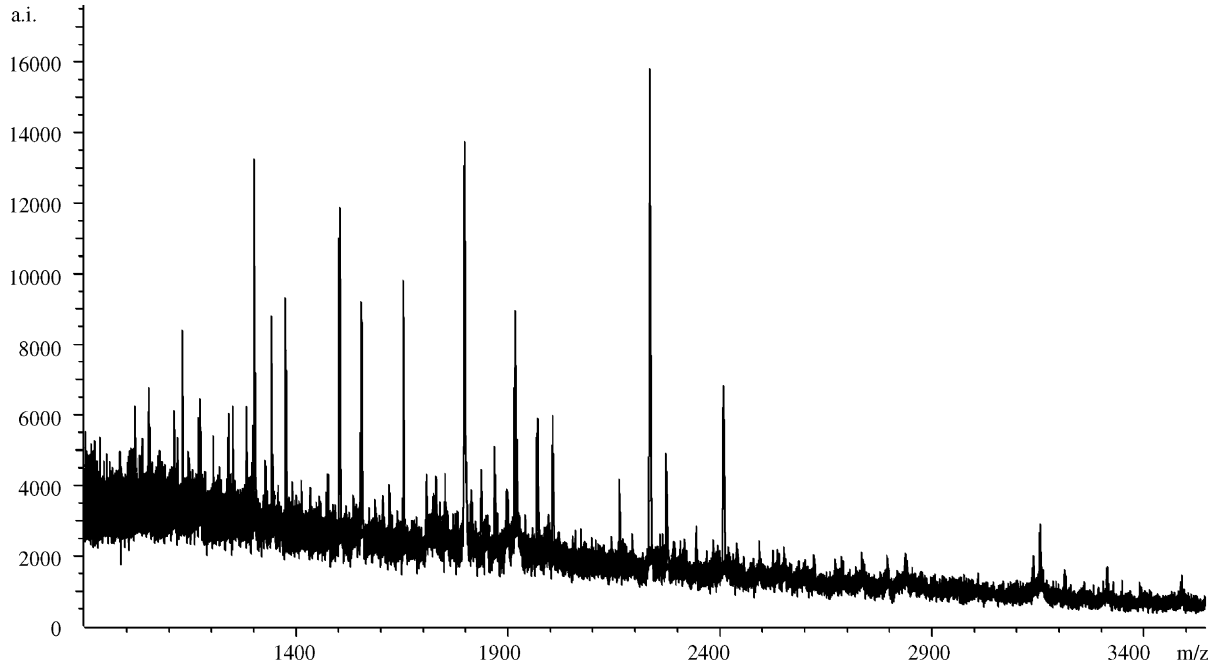

Fig. 1. MALDI mass spectrum of neat botulinum toxin C, from *C. botulinum* strain 003-9; see Table 1 for mass and peptide map assignment.

2.3. Sample preparation for mass spectrometric analysis

An Eppendorf centrifuge (type MC-13; Heraeus, Dijkstra Vereenigde B.V., Lelystad, The Netherlands) was used for all centrifugation in sample preparation.

The sample treatment procedure previously developed with tetanus toxin [7] was used, employing a common autopipette with disposable tips. This procedure allowed handling of any intact toxin inside adequate containment and subsequent mass spectrometry outside that containment. Briefly, 100 or 50 μg of neat toxin was denatured at room temperature, by adding a guanidine buffer (6 M guanidine–HCl in 0.1 M Tris/Tris–HCl with 2 mM EDTA at pH 8.4). Subsequently, disulphide bridges were reduced by the addition of 1 ml of a solution of BMS (2 mg/ml in the above guanidine buffer) and heating at 55 °C (45 min, in the dark). Free cysteine residues were then derivatised with an excess of sodium 2-iodoacetate (7 mg in 50–100 μl of the above guanidine buffer), at 37 °C (30 min, in the dark). Prior to use, MWCO were washed with 100 μl MilliQ water for preconditioning and leakage testing. The reactant solution was transferred to an MWCO 30,000 filter, centrifuged (14,000 × *g*, 25 min) and washed twice with 200 μl of a 0.1 M NH_4HCO_3 buffer. Trypsin in 0.1 M NH_4HCO_3 buffer was then added and the mixture was left overnight (approximately 14 h), at 37 °C. Finally, the mixture was centrifuged (14,000 × *g*, 20 min) and the supernatant was taken out of the containment for mass spectrometry. Any remaining washing solutions and residues and all disposable pipette tips used were rendered harmless by immersion in bleach.

2.4. MALDI mass spectrometry

All matrix-assisted laser desorption/ionisation MS experiments were conducted with a Biflex III™ reflectron time-of-flight instrument (Bruker, Bremen, Germany), equipped with delayed extraction and with a UV ionisation laser (N_2, 337 nm). To an aliquot of a crude trypsin digest of the toxin an equal volume of 5% aqueous formic acid was added. Of that mixture, 10 μl was flushed a few times over a single ZipTip™; the column material of that tip was then washed with three times 10 μl of a 0.1% TFA solution and eluted with 1 μl matrix solution (saturated α-cyano hydroxycinnamic acid with 0.1% TFA in acetonitrile:water, 1:2, v/v) onto the target. MALDI spectra were then recorded to obtain a peptide map for protein database searching. Post-source decay (PSD) was done with the same samples, using standard instrument procedures and no collision gas.

2.5. Liquid chromatography–electrospray (tandem) mass spectrometry

LC–ES MS(/MS) experiments were conducted with a Q-TOF™ hybrid instrument (Micromass, Altrincham, UK) equipped with a standard Z-spray™ ES interface (Micromass) and an Alliance, type 2690 liquid chromatograph (Waters, Milford, MA, USA). The chromatographic hardware consisted of a pre-column splitter (type Acurate; LC Packings, Amsterdam, The Netherlands), a sixport valve (Valco, Schenkon, Switzerland) with a 10 or a 50 μl injection loop mounted and a PepMap C_{18} column (15 cm × 300 μm i.d., 3 μm particles; LC Packings).

A gradient of eluents A (H_2O with 0.2 vol.% formic acid) and B (acetonitrile with 0.2 vol.% formic acid) was used to achieve separation, following: 100% A (at time 0 min, 0.1 ml/min flow) to 100% A (at 5 min, 0.6 ml/min flow) to 20% A and 80% B (at 60 min, 0.6 ml/min flow). The flow delivered by the liquid chromatograph was split pre-column

Table 1
Matching peptides from the trypsin digest peptide map of neat type *C. botulinum* toxin (from *C. botulinum* strain 003-9)

$[M+H]^+$ observed mass (Da)[a]	Peptide[b]	Calculated mass (Da)
BTxC		
800.5	T30	800.39
906.4	T43	906.48
908.4	T78–79	908.51
917.5[c]	T61	917.48
1121.6	T14	1121.58
1152.6	T125–126	1152.56
1205.7	T138–139	1205.60
1251.7[c]	T6	1251.64
1301.8	T81–82	1301.71
1390.8	T71–72	1390.70
1411.8	T18	1411.79
1571.8	T137–138	1571.86
2005.1	T117	2005.02
2343.2[c]	T8–9	2343.07
NTNH		
866.5[c]	T101	866.46
917.5[c]	T17	917.44
931.5	T106	931.51
1112.7	T3–5	1112.62
1220.7	T109	1220.59
1251.7[c]	T7	1251.68
1283.7	T100–101	1283.65
1327.8	T23	1327.66
1375.7	T46	1375.62
1754.9	T72	1754.84
1969.0	T52–54	1969.01
2343.2[c]	T71–72	2343.11
2440.2	T74–75	2440.18
2835.5	T19	2835.44
C3		
866.5[c]	T8	866.42
1018.6	T5	1018.39
1131.7	T18	1131.63
1362.8	T10	1362.73
1390.8[c]	T17–18	1390.77
1681.9	T14	1681.85
2582.4	T4	2582.45

[a] Monoisotopic mass.
[b] In reference to BTxC1 (strain 6813, NCBInr D49440), NTNH (strain C-Stockholm, X62389), and exoenzyme C3 (strain 003-9, M74038).
[c] NTNH in overlap with toxin or C3 peptide, or C3 in overlap with toxin or NTNH peptide.

Table 2
Summary of sequence information obtained for botulinum toxin type C, by trypsin digest

Digest fragment[a]	m/z_{obsd} (Da)[b]	Peptide sequence[c]	Sequence ions[d]
BTxC[e]			
T18	706.4^{2+}	TG*SINPSVIITGPR*	a_2, b_2–b_6, b_2-H_2O, b_3-H_2O, b_4-H_2O, b_5-H_2O, b_6-H_2O, y''_3–y''_{12}
T19	697.6^{2+}	EN*IIDPETSTFK*	a_2, b_2–b_5, b_2-H_2O, b_3-H_2O, b_4-H_2O, b_5-H_2O, y''_1–y''_{10}
T30	400.7^{2+}	*ALDYYR*	a_2, b_2, b_3, y''_1–y''_5, y''_2-NH_3
T39	697.3^{2+}	*FVVESSGEVAVDR*	a_2, b_2–b_4, y''_1–y''_{12}, $[y''_{11} + H]^{2+}$, $[y''_{11} + H\text{-}H_2O]^{2+}$
T41	385.7^{2+}	*FAELYK*	a_1, a_2, b_2, b_3, y''_1–y''_5, y''_4-H_2O, AEL
T55	774.4^{2+}	NT*DLPFIGDISDIK*	a_2–a_4, b_2–b_7, y''_1–y''_{12}, $[y''_{10} + H]^{2+}$, $[y''_{11} + H]^{2+}$
T56	412.2^{2+}	*TDIFLSK*	a_2, b_2–b_4, y''_1–y''_6, DIF
Y71/T66	620.8^{2+}	*IGPALNISNSVR*	a_4, a_5, b_4, b_5, y''_1–y''_{11}
T71	575.3^{2+}	*TIDNZLEQR*	a_2, b_2, b_3, y''_1–y''_8, y''_7-NH_3
T82	409.2^{2+}	*SQVENLK*	b_2, y''_1–y''_6, y_4-H_2O, y''_6-NH_3, VEN
T95	836.4^{2+}	DI*INEYFNSINDSK*	a_2, b_2–b_{11}, y''_1–y''_{12}, $[y''_{12} + H]^{2+}$, NEY
T98	754.9^{2+}	NA*LVDTSGYNAEVR*	a_2–a_4, b_2–b_7, y''_1–y''_{12}
T137	534.8^{2+}	*LLSTSSF*WK	a_1, a_2, b_2, b_3, y''_3–y''_8
NTNH[e]			
T97	517.3^{2+}	*TVTSEEVIR*	a_2, b_2, b_3, y''_1–y''_8
T106	932.6^{2+}	*LINIDESK*	a_1, a_2, b_2, y''_1–y''_7, y''_6-NH_3
T110	387.7^{2+}	*IQLVSSK*	a_1, a_2, b_2, b_3, b_3-H_2O, y''_1–y''_6, y''_5-H_2O, y''_6-NH_3
HA-33[e]			
T8	433.8^{2+}	*LIYDTNK*	a_1, a_2, b_2, y''_1–y''_6, YD, YDT
T17	537.8^{2+}	*LQTQLNSDR*	a_1, b_2, b_3, b_2-NH_3, y''_1–y''_8, y''_6-NH_3, y_8-NH_3, $[y''_8 + H\text{-}NH_3]^{2+}$
C3[e]			
T14	841.9^{2+}	*ANQGNENGLPADILQK*	b_2–b_9, y''_1–y''_{13}, y''_{12}-NH_3
T18	566.4^{2+}	*MPQNIILFR*	a_1, a_2, b_2–b_7, b_3-NH_3, b_4-NH_3, y''_1–y''_8, $[y''_7 + H]^{2+}$, $[y''_8 + H]^{2+}$
T22	425.8^{2+}	*TVFEQVK*	a_2, b_2, y''_1–y''_6, y''_3-NH_3, y''_3-NH_3, y''_3-H_2O, FEQ, FE
T30	745.5^{3+}	GG*YIDPISYFPGQLEVLLPR*	b_2–b_{10}, b_{13}–b_{16}, y''_1–y''_{11}, $[y''_{10} + H]^{2+}$, $[y''_{11} + H]^{2+}$, $[y''_{15} + H]^{2+}$, $[y''_{16} + H]^{2+}$

[a] T for trypsin cleavage site, Y for chymotrypsin cleavage site.
[b] Observed m/z ratio, with apparent monoisotopic mass in Da.
[c] Italic characters indicate residues and position confirmed by sequence peaks; 'Z': carboxymethylcysteine.
[d] According to common Roepstorff–Fohlmann nomenclature [24].
[e] Sequence alignments of digest peptides refer to BTxC1 neurotoxin, *C. botulinum* strain C-Stockholm NTNH (NCBInr access code X62389), strain 6814 HA-33 (AB037166), and strain 003-9 exoenzyme C3 (M74038).

to allow a flow of approximately 7 μl/min through the column and into the ES MS interface.

The mass spectrometer was operated at a cone voltage of 25–35 V. Nitrogen was used as the nebuliser and desolvation gas (at a flow of 20 and 400 l/h, respectively). MS/MS product ion spectra were recorded using a collision energy between 14 and 30 V, with argon as the collision gas (at an indicated pressure of 10^{-4} mbar). LC–ES MS(/MS) was used for trypsin digest analyses.

3. Results and discussion

3.1. Botulinum toxin, type C

3.1.1. Available sequence information

From information supplied by the manufacturer, the neat BTxC purchased was from strain 003-9. No details of the particular strain are known, neither from the manufacturer, nor in any sequence database. By the present state of knowledge on BTx, BTxC is the most complicated of all botulinum toxins: the type C strains can often produce two different toxins, generally denoted C1 and C2, and the capability of production of these toxins can be transmitted by phages. However, the mechanism of action of the C2 toxin is quite different from that of the actual neurotoxin, C1. C2 is known to be cytotoxic, by ADP-ribosylation of monomeric actin (see, for example [12]). For the purpose of the present neurotoxin investigations the scope is limited to the type C1 botulinum neurotoxin and its phage-borne variants.

So far, five *C. botulinum* type C strains have had their C1 toxin genes sequenced: strains 6813 (NCBInr access code D49440, [13]), 6814 (AB037166, [14]), 468C (X72793, [15]), C-Stockholm (X62389 [16]), and C-Yoichi (AB061780, [17]). The strain 6813 sequence (BTxC1) differs slightly from that of strain 6814 (BTxC2;).[1] The strain C-Stockholm sequence (BTxC3) is identical with that of strains C-Yoichi and 468C. In addition to these bacterial sequences, neurotoxin C1 gene sequences from a number of phage genomes have been mapped: 1C phage (X53751, [18]; in unnamed host: X66433, [19]), A028-CN phage (X71126, [20]), and C-ST phage (D90210 [16]). The C-ST

[1] Differences between strain 6813 and 6814 sequences, in reference to 6813 sequence ([6813] position [6814]): T234A, G285A, V484E, and E645G.

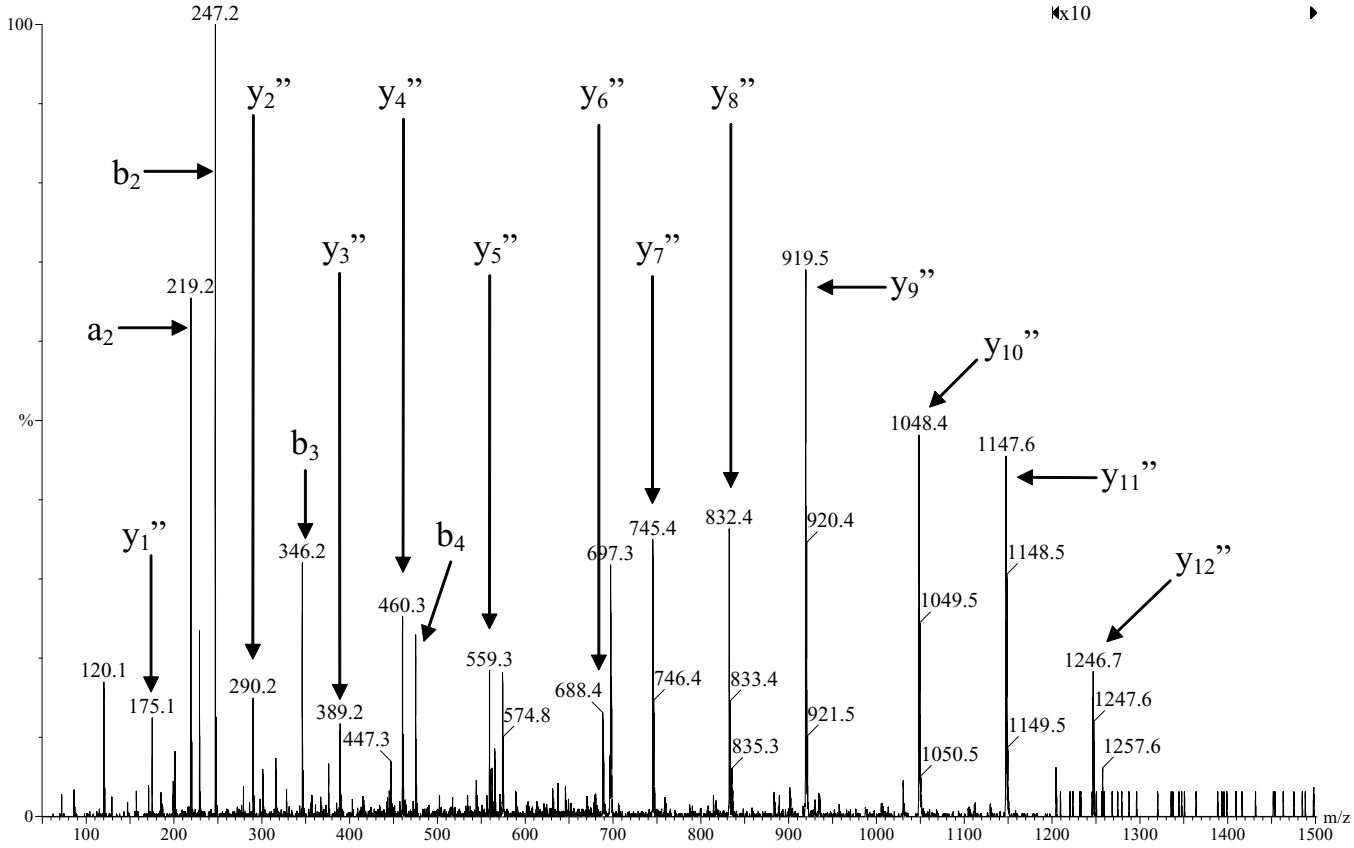

Fig. 2. Product ion MS/MS spectrum (collision energy 25 eV) of the T39 trypsin digest peptide, FVVESSGEVAVDR of BTxC; main series singly charged ions observed are indicated by b- and y″-labels.

phage sequence is identical with BTxC3, in line with the fact that this phage is the infectious agent isolated from the *C. botulinum* C-Stockholm strain. Remarkably, the 1C phage sequence (BTxC4) differs by one amino acid from BTxC3,[2] where the BTxC3 strain 468C was used as a laboratory host for the 1C phage [15]. The A028-CN phage DNA-sequence has an intragenic stop codon and, therefore, does not give rise to neurotoxin production. This condenses the number of known different C1 neurotoxin sequences to four.

The differences between the sequences of BTxC1 and BTxC2 are small, as are those between BTxC3 and BTxC4. However, the differences between the BTxC1/C2 two pair on the one hand and the BTxC3/C4 pair, on the other covers more than 30% of the sequence. Nevertheless, serotyping of, for example, BTxC1 and BTxC3 indicates that they are identical. This sequence heterogeneity reflects the instability of the C1 neurotoxin gene, as compared to stability of the A, B, E, and F neurotoxin genes. This genetic instability is promoted by the possibility of phage transmission of the neurotoxin gene. Since phage transmission is also possible in *C. botulinum* type D, ‘genetic mixes’ of C1 and D type toxin do occur; these mixes are known as ‘type C1 and D mosaic genes’ [13,21]. In the case of neurotoxin C1, the BTxC1 and BTxC2 sequences have a high similarity to a botulinum toxin D sequence (X54254, [22]), particularly in the C-terminal 800 amino acids of the toxin H-chain. Given the detail of distinction attainable by chemical analysis, serological distinction of C1 and D type neurotoxins obscures, rather than clarifies the differences.

3.1.2. MALDI mass spectrometric peptide mapping

The MALDI mass spectrum of neat BTxC from strain 003-9 (Fig. 1) produced a peptide map that was subjected to a ProFound [23] database search (limits: bacteria, masses within ±100 ppm). The search did come up with a poor match for a non-toxic non-hemagglutinin (NTNH) protein (rank 3, estimated Z 0.03 [23][3]), a protein commonly associated with BTx in an non-covalent complex, and an even poorer match for BTxC1 (rank 7); ProFound rather attributed the spectrum to a *Mycobacterium tuberculosis* protein (rank 1, estimated *Z* 0.17) or a *Bacillus subtilis* protein (rank 2, estimated *Z* 0.05). The NTNH and BTxC matched by 12 and 11 peptides, respectively.

This matching result is poor as compared to that for the other botulinum toxins. Therefore, a closer inspection was

[2] In reference t BTxC3: T85 is P85 in 1C phage.

[3] For reference: an ’estimated *Z* value’ of approximately 2.0 for a rank 1 match reflects a typical value for a high probability correct match.

made, using the information gained from the LC–ES MS experiments discussed below. These experiments gave evidence for the presence of BTxC1, an NTNH component, and exoenzyme C3. The peptide attribution is summarised in Table 1. The 33 kDa hemagglutinin (HA-33) is not included in that table, because the only signal attributable to a digest peptide mass from that protein (T8, 866.48 Da) might also result from NTNH or exoenzyme C3. The relatively poor protein library match of the MALDI peptide map seems to originate from the fact that the BTxC material used is a complex mixture of three or more proteins. This lack of differentiation is a general weakness of peptide mapping, which approach would typically be used to identify relatively pure proteins, for instance from two dimensional polyacrylamide gel electrophoresis (2D-PAGE). However, this kind of elaborate separation step was outside the scope of the present study.

3.1.3. LC–ES MS(/MS) sequencing

Sequencing of the neat *C. botulinum* strain 003-9 BTxC revealed the presence of BTxC1 along an NTNH component, a HA-33 and an exoenzyme C3. Results of sequencing experiments are summarised in Table 2, using Roepstorff–Fohlman nomenclature [24]. A sequence coverage of 11% was achieved for the neurotoxin. The sequence covered also spans parts of the toxin that particularly match BTxC1 and BTxC2; however, the similarity between these two variants is so high (four residues out of 1280 differ) that the strain 003-9 neurotoxin sequence cannot exclusively be assigned to either of the two. Since BTxC1 and BTxC2 come from a 'type C1 and D mosaic gene' [13,21], BTxC from strain 003-9 also has this type of mosaic gene. As an illustrative example, the product ion MS/MS spectrum of trypsin digest fragment T39 of neat BTxC is given in Fig. 2.

Three peptides with a typical NTNH sequence were found. The three peptide sequences collectively match three known *C. botulinum* NTNH sequences: (1) from strain C-Stockholm (X62389, [16]) which is identical with that of an unnamed strain (X66433 [19]); (2) from type D strain CB-16 (S80809, [25]), which is identical with that of strain C-Yoichi (AB061780 [17]), and (3) from strain C-6814 (AB037166, [14]). The C-Stockholm and strain CB-16 NTNH amino acid sequences differ by a single residue (I170 versus V170, respectively), whereas the C-6814 sequence differs by 49 out of 1196 residues from that of C-Stockholm. The three different NTNH sequences cannot be distinguished with the observed peptide sequences.

Two peptides with a sequence belonging to an HA-33 protein were found. The combined sequences of these peptides match either of five known HA-33 sequences: those of 1C phage (X66433, [19]) and C-ST phage (X62389, [16]), those of *C. botulinum* type C strains 6814 (AB037166, [14]) and C-Stockholm (X53041, [26]), and that of type D strain D-4947 (AB037920, [27]). Although these HA-33 sequences differ by eight or more amino acid residues, the peptides observed do not allow further distinction. However, it is certain that an HA-33 component was present along the neurotoxin.

The sequence of the exoenzyme C3 of the type C strain 003-9 is known at the DNA level (M74038, [28]). In the present experiments, sequences were obtained from four peptides, T14, T18, T22, and T30, which exactly match the known sequence. Extracted ion chromatograms of possible other trypsin digest fragments of exoenzyme C3

Table 3
Matching peptides from the trypsin digest peptide map of neat BTxD (from strain CB-16)

$[M+H]^+$ observed mass (Da)[a]	Peptide[b]	Calculated mass (Da)
BTxD		
1362.0	T81–82	1361.74
1564.0	T78–79	1563.79
1572.0	T129–130	1571.86
1718.1	T5	1717.91
1854.1	T42–43	1853.97
1886.1	T118	1886.00
1954.0	T112	1953.89
1976.2[c]	T41–42	1976.07
	T113	1976.05
1986.0	T67–68	1985.93
2005.1	T109	2005.02
2096.1	T49–50	2096.06
2177.2	T38	2177.14
2191.2	T113–114	2191.18
2263.4	T84	2263.23
2306.2	T107	2306.15
2382.5	T60	2382.34
2410.2	T19	2410.25
2492.5	T83–84	2492.37
2504.5	T84–85	2504.41
2591.2	T7–8	2591.18
2684.4	(T1–2)-M	2684.29
2733.7	T83–85	2733.55
2908.5	T15	2908.37
3266.8	T102–104	3266.67
3381.9	T110–112	3381.64
NTNH		
1541.8	T74	1541.76
1554.9	T99	1554.72
1586.9	T43	1586.86
1755.0	T72	1754.84
1839.0	T103–104	1838.99
1869.1	T17	1868.99
1976.2[c]	T109–110	1976.05
2244.3	T25	2244.21
2420.2	T113	2420.14
2548.3	T112–113	2548.24
2585.4	T72–73	2585.27
2603.4	T59–60	2603.32
2633.4	T98	2633.33
2668.3	T45–46	2668.17
2727.3	T102	2727.26
2813.4	T57	2813.42

[a] Monoisotopic mass.
[b] In reference to BTxD1 (strain BVD/-3, X54254) and NTNH (strain C-Yoichi; AB061780).
[c] Overlap of toxin and NTNH.

Table 4
Summary of sequence information obtained for botulinum toxin type D, by trypsin digest

Digest fragment[a]	m/z_{obsd}(Da)[b]	Peptide sequence[c]	Sequence ions[d]
BTxD[e]			
T1–2(-M)	895.4^{3+}	TWPV*KDFNYSDPVNDNDILYLR*	a_2, b_2–b_5, b_9–b_{11}, $[b_{13}+H]^{2+}$–$[b_{15}+H]^{2+}$, y''_1–y''_{13}, $[y''_9+H]^{2+}$, $[y''_{11}+H]^{2+}$–$[y''_{14}+H]^{2+}$, $[y''_{20}+H]^{2+}$, $[y''_{20}+2H]^{3+}$
T2	1037.0^{2+}	DF*NYSDPVNDN*DI*LYLR*	a_2, a_3, b_2–b_6, b_8, b_9, y''_1–y''_4, y''_6–y''_9, y''_{11}–y''_{14}
T4	386.3^{2+}	*LITTPVK*	a_1, a_2, b_2, y''_1–y''_6
T5	859.4^{2+}	AF*MITQNIWVIPER*	a_2, b_2–b_{11}, y''_1–y''_{11}
T6	649.4^{3+}	*FSSDTNP*SL*SKPP*RPTSK	a_2, b_2, b_4, b_3-H_2O, y''_7–y''_9, y''_{12}, $[y''_{12}+H]^{2+}$–$[y''_{17}+H]^{2+}$
T7–8	864.4^{3+}	YQ*SYYDPSYLSTDEQKDTFLK*	a_1, a_2, b_2–b_6, b_2-NH_3, y''_1, y''_{10}–y''_{12}, y''_{15}, $[y''_{15}+H]^{2+}$–$[y''_{19}+H]^{2+}$
T10–11	282.2^{2+}	*LFKR*	a_1, a_2, y''_1–y''_3, y''_3-NH_3
T15	970.0^{3+}	LI*NYLVVGSPF**MGDSSTPEDTFDFTR*	a_2, b_2–b_9, b_{11}–b_{17}, y''_1–y''_{14}, $[y''_9+H]^{2+}$–$[y''_{12}+H]^{2+}$, $[y''_{18}+H]^{2+}$–$[y''_{20}+H]^{2+}$
T16	506.7^{2+}	*HTTNIAVEK*	a_1, b_2, b_4–b_8, y''_1, y''_2–y''_4, y''_7, y''_8, y''_2-H_2O
T19	804.2^{3+}	VAPE*FLLTFSDVTSNQSSAVL*GK	a_2, b_2, b_4–b_{12}, y''_2–y''_{11}, $[y''_{21}+H]^{2+}$
T22	1090.0^{4+}	IRPQVSEGFFSQDGPNVQFEELY TFGGL*DVEIIPQIER*	$[b_{29}+H]^{2+}$–$[b_{33}+H]^{2+}$, $[b_{30}+H]^{2+}$–$[b_{34}+H]^{2+}$, y''_1–y''_8
T33–35	1193.5^{4+}	YNFDKDNTGNFVVNIDKFNSLY SDLTNVMS*EVVYSSQYNVK*	y''_1–y''_{11}
T38	726.4^{2+}	HY*LPVFANILDDNIYTIR*	b_2–b_{13}, $[b_8+H]^{2+}$–$[b_{11}+H]^{2+}$, $[b_{14}+H]^{2+}$, y''_1–y''_{11}
T39	454.8^{2+}	*DGFNLTNK*	a_2, a_3, a_3-H_2O, b_2–b_4, y''_1–y''_7, y''_5-NH_3
T40	739.4^{2+}	GF*NIENSGQNIER*	a_2–a_4, b_2–b_5, y''_1–y''_{10}
T42	662.7^{2+}	*LSSESVVDLFTK*	a_2, b_2, b_4–b_6, b_3-H_2O, y''_1–y''_{11}
T43	274.7^{2+}	*VZLR*	a_1, y''_1–y''_3, y''_1-NH_3, a_1 and a_2 of $[ZLR+H]^+$ (y''_3)
T43	548.3^{+}	*VZLR*	a_2, b_2, b_3, x_3-NH_3, y''_1–y''_3, y_1-NH_3, y''_2-NH_3, ZL
T45–46	598.7^{2+}	NSR*DDSTZIK*	a_7, a_8, b_4–b_9, y''_1–y''_5
T46	420.2^{2+}	*DDSTZIK*	a_1, a_2, a_2-H_2O, b_2, y''_1–y''_6, STZ
T53	974.4^{2+}	*YVDYLNSYYYLESQK*	a_1, a_2, a_4, b_2–b_{12}, y''_1–y''_{14}
T55	641.4^{2+}	*IYTFLPSLAEK*	a_1, a_2, b_2–b_4, b_3-H_2O, b_4-H_2O, y''_1–y''_{10}
T57	899.4^{3+}	GV*QAGLFLNWANEVVEDFTTNIMK*	a_2, b_2–b_{15}, $[b_{13}+H]^{2+}$–$[b_{19}+H]^{2+}$, y''_1–y''_{11}, $[y''_9+H]^{2+}$–$[y''_{13}+H]^{2+}$
Y69/T60	648.4^{2+}	*IGPALNIGNSALR*	a_1, a_4, a_5, b_4, b_5, y''_1–y''_{12}, $[y''_{11}+H]^{2+}$, $[y''_{12}+H]^{2+}$
T64	582.3^{2+}	*TIENZLEQR*	a_2, b_2, b_3, y''_1–y''_8, y''_7-H_2O, $[y''_7+H]^{2+}$
T69	955.4^{3+}	ITTQFNHI*NYQMYDSLSYQADAIK*	b_9, b_{10}, $[b_{10}+H]^{2+}$–$[b_{23}+H]^{2+}$, y''_1–y''_{13}
T71	390.7^{2+}	*IDLEYK*	a_1, a_2, b_2, y''_1–y''_5, y''_3-H_2O, DLE
T75	408.3^{2+}	*SQVENLK*	b_2, b_3, y''_1–y''_5, y''_3-H_2O, y''_4-H_2O, y''_6-NH_3
T77	567.2^{2+}	*ISEAMNNINK*	a_1, a_2, b_2–b_4, b_3-H_2O, y''_1–y''_9
T79	574.3^{2+}	*EZSVTYLFK*	a_2, b_2, b_3, b_3-H_2O, y''_1–y''_8
T82	275.7^{2+}	*FDLR*	a_1, y''_1–y''_3, y''_1-NH_3, y''_2-NH_3, a_2 and b_2 of $[LDR+H]^+$ (y''_3)
T84	755.2^{3+}	TE*LINLIDSHNIILVGE*VD*R*	a_2, b_2–b_5, $[b_{10}+H]^{2+}$–$[b_{14}+H]^{2+}$, y''_1, y''_3–y''_{13}, $[y''_{13}+H]^{2+}$–$[y''_{18}+H]^{2+}$
T88	836.3^{2+}	DI*INEYFNSINDSK*	a_2, b_2–b_9, y''_1–y''_{11}
T89	408.3^{2+}	*ILSLQNK*	a_1, a_2, b_2, y''_1–y''_6, y''_2-NH_3, y''_3–NH_3
T91	754.8^{2+}	NA*LVDTSGYNAEVR*	a_2, a_3, b_2–b_{12}, y''_1–y''_{12}
T92	921.0^{2+}	VG*DNVQLNTIYTNDFK*	a_4, a_5, a_4-NH_3, a_5-NH_3, b_4–b_7, b_{10}, b_3-NH_3, b_4-NH_3, b_5-NH_3, b_6-NH_3, b_7-H_3, y''_1–y''_{12}, y''_{10}-NH_3, y''_{11}-NH_3
T96	855.1^{3+}	DLTNSHN*EYTIINSIEQNSGWK*	a_2, b_2, b_8–b_{13}, $[b_{10}+H]^{2+}$–$[b_{19}+H]^{2+}$, y''_1–y''_{11}
T97	281.7^{2+}	*LZIR*	a_1, a_2, b_2, y''_1–y''_3
T103	475.3^{2+}	*LYINGELK*	a_1, a_2, b_2, y''_1, y''_2, y''_4–y''_7, NGE, NGEL, INGEL
T105	480.8^{2+}	*IEDL*DEV*K*	a_1, a_2, b_2, b_3, y''_1–y''_7, y''_4-NH_3, y''_7-H_2O
T105–106	658.8^{2+}	*IEDLDEVKLDK*	a_2, b_2–b_{10}, y''_1–y''_{10}
T108	435.7^{2+}	*DFNIFSK*	a_1, a_2, b_2, y''_1–y''_6, y''_5-NH_3
T109	1003.1^{2+}	EL*SNEDINIVYEGQILR*	b_5–b_8, y''_1–y''_{15}
Y124/T112	648.4^{2+}	*YIINDNYIDR*	a_1, a_2, b_2–b_4, y''_1–y''_7, $[y''_8+H]^{2+}$, $[y''_9+H]^{2+}$
T115	560.3^{2+}	*LYTGNPITIK*	a_1, a_2, b_2, b_3, y''_1–y''_3, y''_5–y''_9, y''_9-H_2O
T118	629.3^{3+}	*ILNGDNIILHMLYNSR*	a_2, b_2, b_3, y''_1–y''_{11}, $[y''_{10}+H]^{2+}$–$[y''_{13}+H]^{2+}$, $[y''_{15}+H]^{2+}$
T128	786.0^{2+}	*NAYTPVAVTNYETK*	a_2, b_2–b_4, y''_1–y''_{12}, $[y''_{10}+H]^{2+}$, YT
T129	534.8^{2+}	*LLSTSSFWK*	a_1, a_2, b_2, y''_1–y''_8
T130–131	603.3^{2+}	*FISRDPGWVE*	a_1, a_2, a_7–a_9, b_2, b_5–b_9, $[b_8+H]^{2+}$, $[b_9+H]^{2+}$, y''_1–y''_3, y''_5, y''_7–y''_9, SRD, PGW

Table 4 (*Continued*)

Digest fragment[a]	m/z_{obsd}(Da)[b]	Peptide sequence[c]	Sequence ions[d]
NTNH[e]			
T7	625.8^{2+}	VA*PNIWVAPER*	a_2, b_2, b_5–b_7, y''_1, y''_3–y''_9, y''_2-NH_3, $[y''_9 + H]^{2+}$
T18	670.3^{2+}	*ETNYIESQNNK*	a_2, b_2–b_6, b_3-H_2O, b_4-H_2O, b_5-H_2O, b_6-H_2O, y''_1–y''_{10}
T23	664.3^{2+}	*FYIDPAMELTK*	a_2, b_2–b_4, y''_1–y''_{10}
T41	336.2^{2+}	*LPLSNK*	a_1–a_3, a_3-NH_3, b_2, y''_1–y''_5, y''_4-NH_3, y''_5-NH_3
T42	572.4^{2+}	*NTNIISKPEK*	a_2–a_4, b_2–b_7, b_9, y''_1, y''_3–y''_8
T66	608.3^{2+}	*QSILAQESLVK*	a_1, a_2, b_2–b_4, b_4-NH_3, b_3-NH_3, b_3-H_2O, y''_1–y''_{10}
T74	771.4^{2+}	A*SIZVFVEDIYPK*	b_2–b_{11}, b_3-H_2O, y''_1–y''_{11}
T76	496.7^{2+}	*YINNINIK*	a_1, a_2, b_2, b_3, y''_1–y''_7, y''_6-NH_3
T83	735.9^{2+}	N*FFNSQVEQVMK*	a_2, b_2–b_5, b_7–b_{11}, y''_1–y''_{10}, $[y''_{10} + H]^{2+}$
T85	672.3^{2+}	G*PNSNIIEDISGK*	a_2, b_2–b_{11}, y''_1–y''_{11}
T86	916.5^{3+}	NT*LIQYTESIELVYGVNGESLYLK*	a_2, b_2–b_{13}, $[b_{10} + H]^{2+}$–$[b_{13} + H]^{2+}$, y''_1–y''_5, y''_7–y''_{12}
T99	777.9^{2+}	*NYFSYLDNSYIR*	a_2, b_2–b_9, y''_1–y''_{10}, y''_1 $[y''_{10} + H]^{2+}$
T104	574.3^{2+}	*NTDGINISSVK*	a_2, b_2–b_6, b_3-H_2O, b_4-H_2O, b_5-H_2O, y''_1–y''_{10}
T109	610.8^{2+}	*YLDISPENNR*	a_1, a_2, b_2–b_4, y''_1–y''_{10}
T110	387.8^{2+}	*IQLVSSK*	a_1, a_2, b_2, b_3, y''_1–y''_6, y''_6-NH_3

[a] T for trypsin cleavage site, Y for chymotrypsin cleavage site.
[b] Observed *m*/*z* ratio, with apparent monoisotopic mass in Da.
[c] Italic characters indicate residues and position confirmed by sequence peaks; 'z' for carboxymethylcysteine.
[d] According to common Roepstorff–Fohlmann nomenclature [24].
[e] Sequence alignments of digest peptides refer to BTxD1 neurotoxin, and *C. botulinum* strain C-Yoichi NTNH (NCBInr access code AB061780).

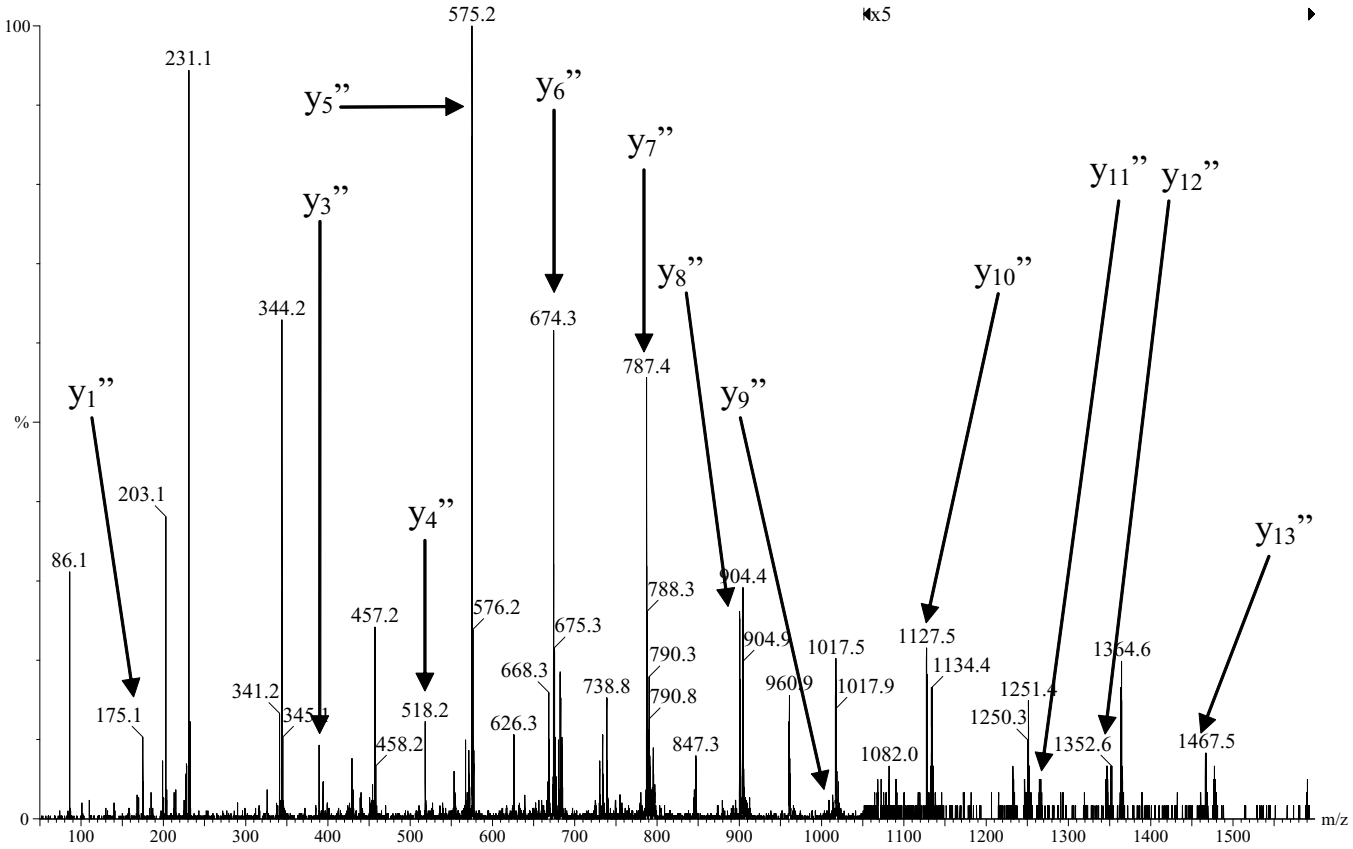

Fig. 3. Product ion MS/MS spectrum of $[T84 + 3H]^{3+}$, 755.2 Da, precursor ions (collision energy 25 eV); for clarity, only y″ series ions were labelled (note that y''_2 was absent).

and subsequent inspection of the MS spectra showed that T8, T16, T19, and T32 were also present with their digest peptide mass corresponding to those of the known strain 003-9 sequence. Since the exoenzyme C3 has mono-ADP–ribosyltransferase activity, contamination with this enzyme implies that the BTxC preparation from the commercially obtained sample will have ribosylating activity.

3.2. Botulinum toxin, type D

3.2.1. Available sequence information

The neat BTxD purchased was from strain CB-16, according to the manufacturer. No details were given with regard to properties of the particular strain. The gene sequence of the toxin gene of the CB-16 strain is not available, but the non-toxin component gene sequences are known. This prompted for a brief review of the information available on BTxD.

The current state of knowledge on type D neurotoxin sequences gives a quite complicated picture. The first known BTxD sequence was that of *C. botulinum* strain BVD/-3 (BTxD1; NCBInr access code X54254 [22]). The BTxD1 sequence differs from that obtained from the genome of the type D bacteriophage, D-16 phi (BTxD2; S49407 [29]), by only two amino acids: the mutations K489N and Q1122R. The possibility of phage transmission of BTxD production capability implies that the type D neurotoxin gene is less stable than that of type A, B, E or F neurotoxins. That is reflected by later reports on type D toxins from strains D-4947 (BTxD3; AB037920 [27]) and South Africa (BTxD4; D38442, [21]). The toxins from these two strains have a high sequence similarity with BTxD1 and BTxD2, but only in the N-terminal 912 amino acid residues (BTxD3 has 33 and BTxD4 has 34 residues in difference with BTxD1).[4] The 900 BTxD3 and BTxD4 N-terminal residues differ in only three amino acid residues (V341L, V362I, and R715Q). The over 300 residues of BTxD3 and BTxD4 C-terminal sequence are completely different from those of BTxD1 and BTxD2, whereas there is only a single residue difference between BTxD3 and BTxD4 (V917I). Actually, the C-terminal part of BTxD3 and BTxD4 resembles that of the other phage transmissible BTx, the C1 botulinum neurotoxin. As a consequence, differentiation between BTxC and BTxD requires a distribution of sequence information over N- and C-terminal parts of the toxin molecule.

The situation seems less complicated with respect to the proteins that might accompany the type D neurotoxin. Two NTNH sequences are known from *C. botulinum* type D. One

[4] BTxD3 in reference to BTxD1: L276S, S287L, I316S, R368K, R396I, D397N, N403T, K442R, K452Q, R457T, N474S, K475Q, Q484E, K489N, G499A, Q500K, I503T, I507A, L522V, I527E, Y536D, S548A, V623A, Q715R, N720S, R805K, M840I, I856M, A878T, V880M, G892E, D893G, T898P; BTxD4 in reference to BTxD1: L276S, S287L, I316S, V341L, V362I, R368K, R396I, D397N, N403T, K442R, K452Q, R457T, N474S, K475Q, Q484E, K489N, G499A, Q500K, I503T, I507A, L522V, I527E, Y536D, S548A, V623A, N720S, R805K, M840I, I856M, A878T, V880M, G892E, D893G, T898P.

Table 5
Matching peptides from the trypsin digest peptide map of neat type E botulinum toxin

$[M+H]^+$ observed mass (Da)[a]	Peptide[b]	Calculated mass (Da)
BTxE		
621.2	T104	621.37
659.2	T43	659.41
870.2	T38	870.54
983.2	T37	983.50
1020.2	T64–65	1020.56
1040.3	T5	1040.59
1068.3[c]	T19	1068.62
1098.2	T90	1098.52
1132.2	T73	1132.58
	T111	1132.60
1165.3[c]	T77–78	1165.59
1196.3[c]	T56–57	1196.59
1264.4	T99	1264.67
1291.3[c]	T101	1290.71
1434.5	T30	1434.72
1503.5	T48	1502.82
1567.4	T2	1567.71
1604.5[c]	T63	1603.78
1692.7	T119	1692.86
1699.6	T23–24	1699.97
1791.6	T100–101	1790.97
1817.7	T68–69	1817.92
	T3	1817.91
1851.8	T6	1851.96
1874.7	T98	1873.91
1949.8	T119–120	1950.00
2060.9	T24–25	2061.10
2801.4	T13	2800.32
NTNH		
854.2	T60	854.53
926.2	T68–69	926.51
958.3	T42–43	958.57
1051.3	T112–113	1050.64
	T10	1050.55
1068.3[c]	T31–32	1068.61
	T62	1068.64
1072.3	T27–28	1072.53
1165.3[c]	T95	1165.62
1196.3[c]	T108	1196.59
	T62–63	1196.74
1291.3[c]	T79	1291.57
1327.3	T36	1327.62
1440.4	T101	1440.68
1604.5[c]	T47	1604.80
1639.7	T6	1639.90
1869.7	T65–66	1868.98
2273.1	T111–112	2273.07

[a] Monoisotopic mass.
[b] In reference to BTxE1 (strain NCTC 11219; X62683), NTNH (strain Mashike; D12697).
[c] Overlap of NTNH and toxin peptide.

Table 6
Summary of sequence information obtained for botulinum toxin type E, by trypsin digest

Digest fragment[a]	m/z_{obsd} (Da)[b]	Peptide sequence[c]	Sequence ions[d]
BTxE[e]			
T2	784.4^{2+}	IN*SFNYNDPVNDR*	b_2–b_8, y''_1–y''_{11}
T3	606.7^{3+}	T*ILYIK*PGGZQEFYK	a_2, b_2, b_3, y''_9, $[y''_{11} + H]^{2+}$–$[y''_{13} + H]^{2+}$
T5	520.8^{2+}	NI*WIIPER*	a_2, b_2–b_4, y''_1, y''_3–y''_6
T6	926.5^{2+}	NV*IGTTPQDFHPP*TSLK	a_2, b_2, b_3, y''_5–y''_{15}
T12	857.5^{2+}	IN*NNLSGGILLEELSK*	b_2–b_6, y''_1–y''_{13}
T13	934.2^{3+}	AN*PYLGNDNT*PDNQFHIGD*ASAVEIK*	a_2, b_2–b_6, y''_1–y''_5, y''_7, y''_9–y''_{12}, $[y''_{17} + H]^{2+}$–$[y''_{23} + H]^{2+}$
T19	534.8^{2+}	*QNPLITNIR*	b_2–b_5, b_2-NH_3, b_3-NH_3, b_4-NH_3, b_5-NH_3, y''_1–y''_7, $[y''_7 + H]^{2+}$
T30	717.9^{2+}	*LYSFTEFDLATK*	a_2, b_2–b_6, y''_1–y''_{11}
T33	500.8^{2+}	*QTYIGQYK*	a_2, a_4-NH_3, b_2, b_3, b_6-H_2O, y''_1, y''_2, y''_4–y''_7
T37	492.2^{2+}	GQ*NANLNPR*	b_2–b_7, b_2-NH_3, y''_1–y''_7
T38	435.8^{2+}	IITPITGR	a_1, a_2, b_2, b_3, y''_3–y''_7, ITG
T48	751.9^{2+}	LN*LTIQNDAYIPK*	a_1, a_2, b_2–b_5, b_4-H_2O, b_5-H_2O, y''_1–y''_{11}
T50	823.4^{3+}	VPEGENNVNLT*SSIDTALLEQPK*	$[b_{16} + H]^{2+}$–$[b_{21} + H]^{2+}$, y''_1–y''_{12}
T56	477.7^{2+}	*SFLGSSDNK*	a_2, b_2, b_3, y''_5–y''_7
T64	432.7^{2+}	*INTQFNK*	a_1, a_2, b_2, b_3, y''_1–y''_6, y''_6-NH_3
T67	889.5^{2+}	EQ*MYQALQNQVNAIK*	b_2–b_6, b_2-H_2O, b_3-H_2O, b_4-H_2O, b_5-H_2O, b_6-H_2O, y''_1–y''_{13}
T69	573.8^{2+}	*YNSYTLEEK*	a_1, a_2, b_2, b_3, y''_1–y''_8
T72	558.3^{2+}	*QIENELNQK*	a_2, a_2-NH_3, b_2–b_6, b_2-NH_3, b_3-NH_3, b_4-NH_3, b_5-NH_3, b_6-NH_3, y''_1–y''_7
T73	566.8^{2+}	*VSIAMNNIDR*	a_2, b_2–b_6, b_3-H_2O, y''_1–y''_9, $[y''_8 + H]^{2+}$, $[y''_9 + H]^{2+}$
T90	549.8^{2+}	*NQFGIYNDK*	b_2–b_5, b_2-NH_3, y''_1–y''_7
T94	432.3^{2+}	*IPN*YD*NK*	a_1, a_2, b_2, y''_1, y''_2, y''_4–y''_6, $[y''_6 + H]^{2+}$, PNYD
T95	927.4^{2+}	IV*NVNNEYTIINCMR*	a_1, a_2, b_2–b_8, y''_1–y''_{13}
T97	803.0^{2+}	VSLNHN*EIIWTLQDNAGINQK*	b_7–b_{10}, $[b_9 + H]^{2+}$–$[b_{20} + H]^{2+}$, y''_1–y''_{11}
T99	632.8^{2+}	*WIFVTITNDR*	a_1, a_2, b_2–b_6, y''_1–y''_9
T101	645.8^{2+}	*LYINGNLIDQK*	a_1, a_2, b_2–b_4, y''_1–y''_9, y''_7-NH_3, y''_8-NH_3
T102	633.0^{3+}	S*ILNL*GNIHVSDNILFK	a_2, b_2–b_5, $[y''_{12} + H]^{2+}$–$[y''_{15} + H]^{2+}$, $[y''_{15} + 2H]^{3+}$
T103	507.3^{2+}	*IVNZSYTR*	a_2, b_2, y''_1–y''_7
T105	473.3^{2+}	*YFNIFDK*	a_1, a_2, b_2, y''_1–y''_6
T111	566.8^{2+}	DS*TLSINNIR*	b_3–b_5, b_3-H_2O, b_4-H_2O, y''_1–y''_8
T116	666.8^{2+}	VN*NSST*ND*NLVR*	b_2–b_4, y''_1–y''_4, y''_6–y''_{10}
T123	1102.5^{2+}	FN*QVVVMNSVG*NNZTMNFK	b_3–b_6, y''_2, y''_6, y''_9–y''_{16}
NTNH[e]			
T6	820.5^{2+}	AF*QVAPNIWIVPER*	b_2–b_5, b_8, y''_1, y''_3–y''_{11}
T7	430.2^{2+}	*YYG*ESL*K*	a_2, b_2, y''_1, y''_5, y''_6
T12	464.7^{2+}	*INNNVVGAK*	b_2, b_4–b_6, y''_1–y''_8
T14	470.7^{2+}	*QTNYLSSK*	y''_1–y''_7
T19	858.0^{2+}	*YDEFYVDPALELIK*	a_1, a_2, b_2–b_7, y''_1, y''_3–y''_{12}
T27–28	536.8^{2+}	*YKND*YE*IK*	b_2, b_6, b_7, y''_1, y''_2, y''_5–y''_7
T41–42	393.8^{2+}	*TKL*PL*SK*	b_2, b_3, y''_1, y''_2, y''_4–y''_6
T61	927.0^{2+}	AL*NILNTNNSFVEEFK*	a_2, b_2–b_{10}, y''_1–y''_{13}
T62	534.8^{2+}	NL*GPISLI*NK	a_2, b_2, b_3, y''_2–y''_8
/T76	686.4^{2+}	*ILINLTTTNTLR*	a_2, b_2–b_5, y''_1–y''_{11}
T79	646.4^{2+}	*FTSFMEQZIK*	a_1, a_2, b_2, b_3, b_3-H_2O, y''_1–y''_9
T83	555.3^{2+}	ZT*NINETEK*	a_2, b_2–b_5, y''_1–y''_7
T85[f]	728.0^{2+}	NL*FNSYTTELLIK*	a_1, a_2, b_2–b_4, y''_1–y''_{10}
/T86	861.8^{2+}	AF*QEQDNNVIGDTSGK*	a_2, b_2–b_5, y''_1–y''_{11}
T87	482.3^{2+}	*NTLVEYPK*	a_2, b_2, b_3, b_3-H_2O, y''_1–y''_7
T101	720.8^{2+}	*NFDEEILQYNR*	a_2, b_2–b_7, y''_1–y''_{10}
T105	457.3^{2+}	*LLNTNPNK*	a_1, a_2, b_2, b_3, y''_1–y''_7
T107	819.4^{2+}	*WDEVIFSVLDG*TEK	a_1, b_2–b_5, y''_4–y''_{13}
T108	598.8^{2+}	*YLDISTTNNR*	a_1, a_2, a_4, b_2–b_5, y''_1–y''_9

[a] T for trypsin cleavage site, '/' for non-trypsin cleavage site with respect to reference sequence (see foot note f).
[b] Observed *m/z* ratio, with apparent monoisotopic mass in Da.
[c] Italic characters indicate residues and position confirmed by sequence peaks; 'z' for carboxymethylcysteine.
[d] According to common Roepstorff–Fohlmann nomenclature [24].
[e] Sequence alignments of digest peptides refer to BTxE1 neurotoxin, and *C. botulinum* strain Mashike NTNH (NCBInr access code D12697).
[f] Deviates from strain Mashike NTNH (NLFNLYTELLIK); observed NLFNSYTTELLIK occurs in a few other NTNH (for example that of type F strain 202F, NCBInr access code X71086).

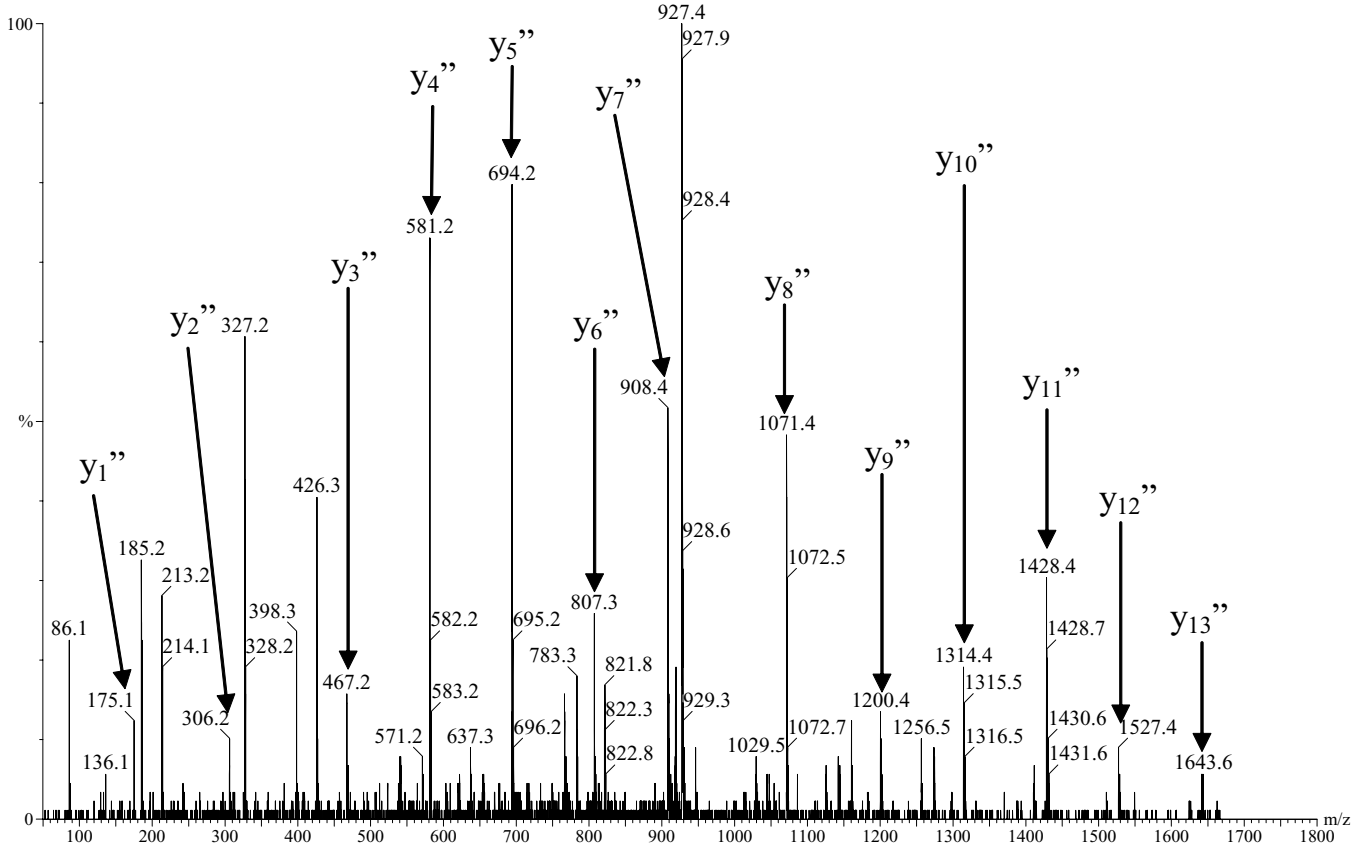

Fig. 4. Product ion MS/MS spectrum of $[T95 + H]^{2+}$, 927.4 Da, precursor ions (collision energy 28 eV); for clarity, only y″ series ions were labelled.

is from the CB-16 strain presently reported on (S80809 and AB012111 [25]), whereas the other is from strain D-4947 (AB037920 [27]). Specifically for the CB-16 strain, the NTNH is known to occur as the full 130 kDa protein, in an “L complex” with the neurotoxin, and as an N-terminally truncated 115 kDa protein, in an “M complex” with the neurotoxin [25]. Both strain CB-16 [25] and strain BVD/-3 [22] gene sequences cover hemagglutinin factors as well. Therefore, the strain CB-16 BTxD investigated may occur in the company of NTNH and HA.

3.2.2. MALDI mass spectrometric peptide mapping

The MALDI mass spectrum of neat BTxD (not shown) produced a peptide map that was subjected to a ProFound [23] database search (limits: bacteria, masses within 100 ppm). The search invariably came up with a best match for BTxD, although the number of matched peptides varied from 10 to 16 between repeated experiments. A closer inspection of the spectrum by hand, summarised in Table 3, revealed 28 matching peptide signals, 12 more than by automated database match. One of the extra matches is the (T1–2)-M peptide, for which the database search did not account because of the post-translational modification involved.

As a confirmation of the MALDI spectrum match, a PSD experiment was done with the *m/z* 1976.2 precursor ions; the PSD spectrum was initially matched with protein database information by the Mascot program [30]. The PSD spectrum of 1976.2^{+} ions yielded signals in line with b and y″ calculated for the T113 sequence (a2, b2, b3, b8–b10, b3-H2O, b4-H2O, y''_1, y''_3, y''_4, y''_6, y''_7, y''_{12}, y''_3-NH3, y''_7-NH3, y''_9-NH3). This finding is in agreement with the initial match of the common MALDI mass spectrum to BTxD1.

3.2.3. LC–ES MS(/MS) sequencing

Sequencing of trypsin digested neat BTxD revealed the presence of the neurotoxin and an NTNH component. No evidence was found for the presence of haemagglutinins. Results of sequencing experiments are summarised in Table 4. A sequence coverage of 49% was achieved for the neurotoxin. As an example, the product ion MS/MS spectrum of $[T84 + 3H]^{3+}$ is given in Fig. 3.

The N-terminal sequence showed up as a T1–2 fragment that is missing the N-terminal methionine residue ([T1–2]-M), in agreement with the MALDI results. The two amino acid residues that distinguish BTxD1 from BTxD2, K/N^{489} and Q/R^{1122}, are not covered by any of the digest peptides observed. In the sequence overlap area of BTxD1 and BTxD2 with BTxD3 and BTxD4, the 912 N-terminal residues, ten of the observed digest peptides cannot belong to BTxD3 or BTxD4. The C-terminal remainder is covered

by 13 peptides that cannot belong to BTxD3 or BTxD4. Therefore, all neurotoxin peptide sequences observed correspond exactly with the sequence of BTxD1 (strain BVD/-3) and BTxD2 (phage D-16 phi).

The observed NTNH related peptides match a number of NTNH sequences in the NCBInr database. There was a complete match of all 15 peptides with the sequences known from the type C strain C-Yoichi (AB061780, [17]), from an unnamed C-strain (X66433, [19]), and from the C-ST phage (X62389, [16]). Actually, the C-Yoichi and C-ST phage NTNH sequences are identical and differ by only one amino acid from that of the 1C phage (I^{169} is V^{169} in 1C phage). This single residue difference was not covered by the peptides observed. The C-Yoichi NTNH differs by two residues from that of the strain CB-16 (L^{859} is F^{859} and V^{1103} is G^{1103}, in CB-16); these differences both correspond to a single nucleotide difference. Notably, one of the 15 peptides observed, T86, spans one of these differences and shows a leucine residue at position 859. Hence, the observed NTNH sequence is not completely in line with the CB-16 NTNH gene sequence known from literature [25]. This minor discrepancy may be explained by an error in the reported CB-16 sequence (where the database sequence, NCBInr access code S80809, already differs from that reported in the corresponding paper [25] by mismatches H375N, K413L, K619S, and K624L), or by a C-type phage infection of the CB-16 strain used in the production of the toxin. Either way, the difference would lead to a correct identification of the sample material as BTxD.

3.3. Botulinum toxin, type E

3.3.1. Available sequence information

According to the information obtained from the manufacturer, the neat BTxE was from a *C. botulinum* strain originally isolated from herring sprats. No details were available with regard to a strain collection number or to properties of the particular strain. Therefore, any database information on BTxE might apply.

At least 17 genetic sequences are known of complete neurotoxin genes from BTxE producing organisms. However, most of these are from a comparative investigation of BTxE from strains of *Clostridium butyricum*, rather than from *C. botulinum* [31–33]. Three complete neurotoxin type E sequences from *C. botulinum* are known. That from strain NCTC 11219 (BTxE1; NCBInr access code X62683 [34]) is identical with that from strain 35396 (AB082519) [35], whereas that from strain Beluga (BTxE2; X62089, [31]) differs by only seven out of 1252 amino acid residues.[5] In addition, eight partial toxin gene sequences derive from a PCR application to food contamination with *C. botulinum* capable of BTxE production [9,36]. These partial sequences, corresponding to about 600 amino acid residues, differ at most by two amino acid residues from the BTxE1 sequence. The amino acid sequence difference between *C. botulinum* and *C. butyricum* type E neurotoxins is between 5 and 30 residues, out of 1252, with intraspecies differences in the *C. butyricum* toxins between 5 and 30 residues as well. Hence, species distinction of botulinum toxin type E producing *C. butyricum* and *C. botulinum* can generally not be made at the neurotoxin amino acid level.

A complete NTNH nucleotide sequence from a *C. botulinum* type E strain is known from strain Mashike (D12697

[5] Relative to BTxE1 (first residue): G177R, S198C, A340R, L774I, L963F, Q964E, deletion of N1215.

Table 7
Matching peptides from the trypsin digest peptide map of neat type F botulinum toxin

$[M+H]^+$ observed mass (Da)[a]	Peptide[b]	Calculated mass (Da)
BTxF		
885.5[c]	T44–45	885.58
1026.6	T6	1026.57
1067.7	T115	1067.63
1354.7	T25	1354.70
1366.8	T22	1366.79
1413.7	T34	1413.73
1426.7[c]	T43–44	1426.82
1437.7	T86	1437.70
1523.8	T118	1523.80
1581.7	T54	1581.71
1601.8	T102	1601.83
1688.8	T34–35	1688.89
1737.9[c]	T99	1737.85
1776.9	T117	1776.90
1818.8	T105–106	1818.97
1872.0	T107	1871.99
1887.9[c]	T30–31	1887.87
1978.0	T77–78	1977.99
2106.0	T31–32	2105.98
2121.0	T116–117	2121.07
2258.1[c]	T85–86	2258.06
2304.1	T53	2304.07
2506.2	T28–29	2506.16
2943.5	T60–61	2943.62
NTNH		
843.3	T98	843.49
885.5[c]	T3–4	885.53
993.5	T88	993.51
1124.5	T97	1124.52
1239.6	T34	1239.59
1396.7	T89–90	1396.63
1426.7[c]	T59	1426.72
1572.7	T8	1572.72
1625.9	T7	1625.88
1713.8	T18	1713.87
1737.9[c]	T43	1737.87
1809.9	T54	1809.94
1887.9[c]	T9	1887.86
1949.0	T97–98	1949.00
2258.1[c]	T24	2258.08

[a] Monoisotopic mass.
[b] In reference to BTxF1 (X81714) and NTNH (X99064), both *C. botulinum* strain Langeland.
[c] Overlap of NTNH and toxin peptide.

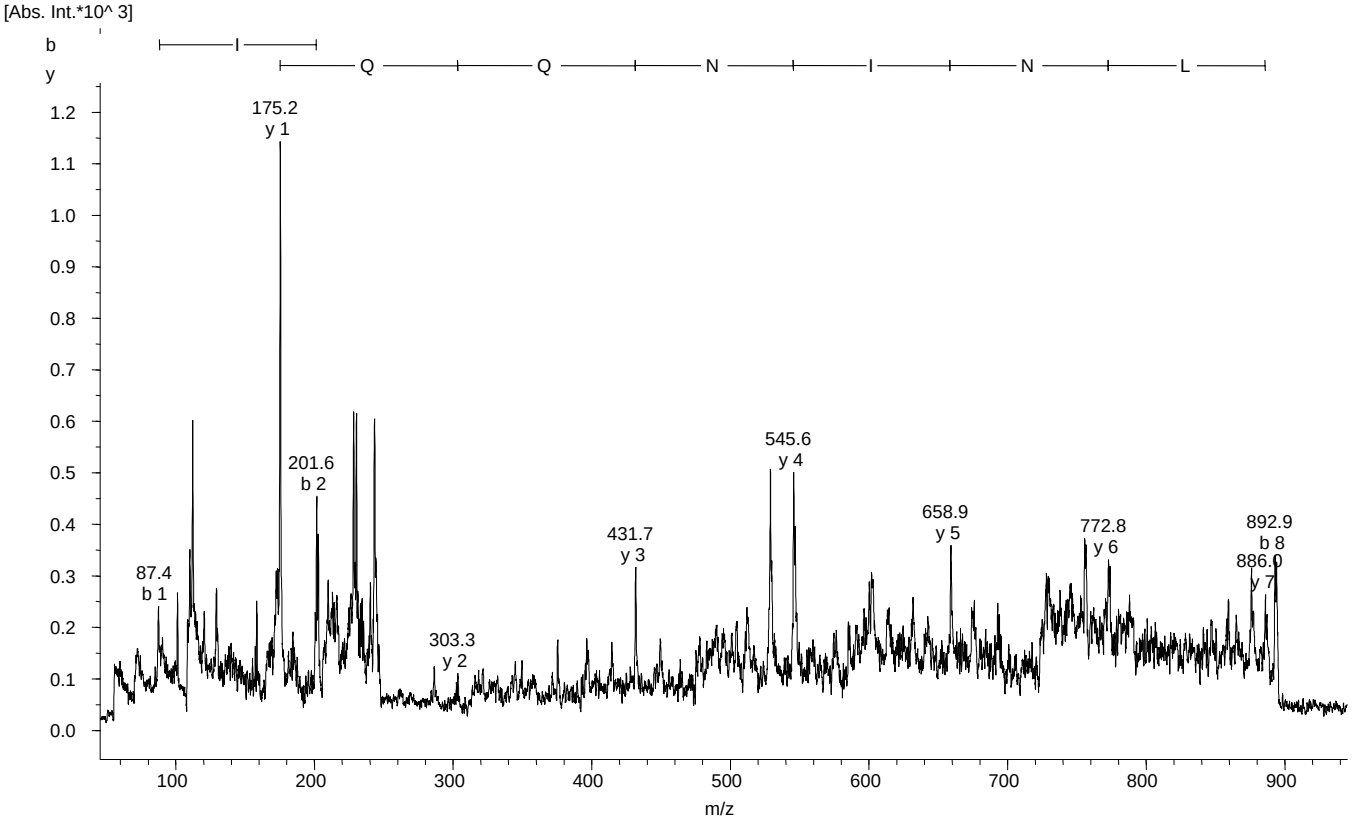

Fig. 5. Annotated partial PSD spectrum of $[T117 + H]^+$ of BTxF, 1776.9 Da precursor ions.

[37]), along a partial sequence from strain Alaska (U70780 [38]). In addition, the complete NTNH gene of a *C. butyricum* strain (ATCC 43755, D12739 [39]) was reported. As with the type E neurotoxin, *C. botulinum* and *C. butyricum* NTNH sequences are closely similar, with 25 amino acid residues difference out of 1163, and over 10% distinct from the NTNH of other *Clostridia*.

The available genetic information does not show evidence of the presence of HA components in toxin genes from any BTxE producing *Clostridia*. That implies that BTxE appears mainly as a complex of an NTNH and the neurotoxin.

3.3.2. MALDI mass spectrometric peptide mapping

The MALDI mass spectrum of the BTxE trypsin digest (not shown) provided a good peptide map: BTxE was the rank 1 candidate in a ProFound [23] search (limits: bacteria, masses within 100 ppm), where 27 peptides matched. The type E related NTNH ranked 9, with 18 matched peptides, of which 5 indistinguishable in mass from BTxE digest peptides. The matching peptides are summarised in Table 5.

3.3.3. LC–ES MS(/MS) sequencing

Sequencing of trypsin digested neat BTxE revealed the presence of the neurotoxin and an NTNH component. No evidence was found for the presence of haemagglutinins. Results of sequencing experiments are summarised in Table 6. Sequence coverage of 30 and 17% was achieved for the neurotoxin and the NTNH, respectively. All neurotoxin peptide sequences observed correspond exactly with the sequence of BTxE1 (strain NCTC 11219). As an illustration of the findings, the product ion spectrum of $[T95 + 2H]^{2+}$ of BTxE is given in Fig. 4.

Most NTNH sequences observed correspond to that of the known sequence of the *C. botulinum* type E strain Mashike [37]. One peptide, NLFNSYTELLIK, aligns with T85 (NLFNLYTELLIK in strain Mashike NTNH) but deviates in one amino acid. The observed trypsin digest fragment occurs in some of the known *C. botulinum* type F NTNH, for example that of strain 202F (NCBInr X71086, [40]). The other two non-corresponding peptides are only partial trypsin digest fragments (of T76 and T86), apparently missing amino acid residues from the peptide N-terminal. The amino acids adjacent to the observed N-terminal do not constitute a chymotrypsin cleavage site in strain Mashike NTNH. There were also no signs that the T76 and T86 truncation was due to in-source fragmentation. It can only be guessed at that the N-terminus of these two peptides is adjacent to a trypsin cleavage site (K or R residue). Thus, the two apparent miscleavages do not give direct evidence for a sequence difference with strain Mashike NTNH, but the odd T85 peptide does. Hence, there is at least a small

Table 8
Summary of sequence information obtained for botulinum toxin type F, by trypsin digest

Digest fragment[a]	m/z_{obsd} (Da)[b]	Peptide sequence[c]	Sequence ions[d]
BTxF[e]			
T1-M	833.4^{4+}	PVVIN*SFN*YNDPVNDDTI*LYMQIPYE*EK	b_2, b_6, b_8, b_9, y''_2, y''_3, y''_5–y''_{10}
T5	383.7^{2+}	*AFEIMR*	a_2, b_2, b_3, y''_1–y''_5, y''_4-H_2O
T6	513.8^{2+}	NV*WIIPER*	a_2, b_2, b_3, y''_1–y''_6
T10	462.2^{+}	*TTIK*	a_2, b_2, b_3, b_2-H_2O, b_3-H_2O, y''_1–y''_3, y''_1-H_2O, y''_2-H_2O, y''_3-H_2O
T11	407.3^{+}	*LFK*	a_2, b_2, y''_1
T13	810.8^{5+}	I*NSNPAG*EVLLQEISYAKPYLGNEHTPINEFHPVTR	a_2, b_2–b_6, y''_4, y''_5, $[y''_{30}+2H]^{3+}$, $[y''_{32}+2H]^{3+}$, $[y''_{32}+3H]^{4+}$, $[y''_{33}+3H]^{4+}$, $[y''_{34}+3H]^{4+}$
T14	381.7^{2+}	*TTSVNIK*	a_2, b_2, y''_1, y''_3–y''_6, SVN
T14	762.4^{+}	*TTSVNIK*	a_2, b_2–b_6, b_3-H_2O, b_4-H_2O, b_5-H_2O, b_6-H_2O, y''_1–y''_5, y''_3-NH_3, y''_4-H_2O
T15	635.2^{+}	*SSTNVK*	b_2, b_4, b_5, y''_1–y''_4, TN
T16	892.0^{3+}	SSIILNL*LVLGAGP*DIFEN*SSYPVR*	b_9–b_{13}, y''_1, y''_3–y''_6, $[y''_{12}+H]^{2+}$–$[y''_{18}+H]^{2+}$
T19	567.3^{+}	GV*TYK*	a_2, b_2, b_3, b_3-H_2O, y''_1–y''_3
T22	683.8^{2+}	QAP*LMIAEKPIR*	b_5, y''_1–y''_7, LMI
T23	743.0^{3+}	LEEFLTFGGQDLNIITSAMK	a_1, a_2, b_2–b_8, b_{10}–b_{13}, y''_1–y''_{10}
T25	677.8^{2+}	*IYNNLLANYEK*	a_1, a_2, b_2–b_7, y''_1–y''_{10}
T26	460.3^{+}	*IATR*	a_2, b_2, y''_1–y''_3, y''_1-NH_3, y''_2-NH_3, y''_3-NH_3, $[M+H\text{-}NH_3]^{+}$
T28	819.8^{2+}	VN*SAPPEYDINEYK*	b_2, b_3, b_4, b_4-H_2O, y''_1–y''_7, y''_9–y''_{12}, PPEYDINE
T29	443.7^{2+}	DY*FQWK*	a_2, b_2, b_3, y''_1–y''_4, y''_3-NH_3
T30	595.3^{+}	*YGLDK*	b_2–b_4, y''_1–y''_4
T31	656.2^{2+}	NA*DGSYTVNENK*	a_2, b_2–b_{11}, y''_1, y''_2, y''_4–y''_{10}
T32	407.2^{2+}	*FNEIYK*	b_2–b_4, y''_1–y''_5, y''_5-NH_3, y''_4-H_2O
T34	707.3^{2+}	*LYSFTEIDLANK*	a_1, a_2, b_2, b_3, y''_1–y''_{11}
T38	393.2^{2+}	*NTYFIK*	a_1, a_2, b_2, b_4, b_4-H_2O, y''_1–y''_5
T39	314.2^{2+}	*YGFLK*	a_1, a_2, b_2, b_3, y''_1–y''_4
T40	954.6^{3+}	VPNLLDDDIYTVSEGF*NIGNLAVNNR*	b_{12}, $[b_{18}+H]^{2+}$–$[b_{25}+H]^{2+}$, y''_1–y''_{10}
T42	471.2^{+}	*LNPK*	b_2, y''_1, y''_2, y''_3, y''_1-H_2O, y''_2-H_2O, y''_3-H_2O
T47	571.4^{+}	*SVIPR*	a_2, b_2–b_4, y''_1–y''_3, y''_2-NH_3, y''_3-NH_3^{+}
T51	562.3^{+}	*LZIR*	a_1, a_2, b_2, y''_1–y''_3, y''_1-NH_3, y''_2-NH_3, y''_3-NH_3
T53	768.8^{3+}	EL*FFVASESSYNENDINTPK*	a_2, b_2–b_{11}, b_3-H_2O, b_4-H_2O, b_5-H_2O, b_6-H_2O, b_{11}-H_2O, y''_1–y''_{11}
T54	791.4^{2+}	EI*DDTTNLNNNYR*	a_2, b_2–b_6, b_8–b_{12}, b_5-H_2O, y_1-y''_{11}
T55	1006.5^{4+}	NNLDEV*IL*DYNSETIPQISNQTLNTL*VQDDSYVPR*	b_6–b_8, b_{15}, y''_1–y''_9, $[y''_{20}+H]^{2+}$, $[y''_{22}+H]^{2+}$
T58	520.5^{2+}	*DFTTEATQK*	a_2, b_2, b_3, y''_1–y''_7
T59	597.2^{+}	*STFDK*	a_2, b_2, b_4, b_3-H_2O, y''_1–y''_4
T61	537.2^{+}	*ENFK*	a_1, b_3, y''_1, y''_2, y''_1-H_2O, y''_2-H_2O, y''_2-$2H_2O$, $[M+H\text{-}NH_3]^{+}$
T63	484.7^{2+}	SF*IGSSENK*	a_2, b_2, b_3, y''_1–y''_7, $[y''_7+H]^{2+}$
T66	524.1^{2+}	*AINNSLMER*	a_2, b_2, b_3, y''_1–y''_7, y''_7-NH_3
T66(ox)	532.3^{2+}	AINNSLM(→ O)ER	a_2, b_2–b_8, y''_1–y''_7
T69	833.9^{2+}	EI*YSWIVSNWLTR*	a_2, b_2–b_7, b_4-H_2O, b_5-H_2O, b_6-H_2O, y''_1–y''_{11}
T70	432.7^{2+}	*INTQFNK*	a_1, a_2, b_2, y''_1–y''_6, y''_5-H_2O, y''_6-H_2O
T73	889.9^{2+}	EQ*MYQALQNQVDAIK*	b_2–b_{13}, b_2-H_2O, b_3-H_2O, b_4-H_2O, b_5-H_2O, b_6-H_2O, b_7-H_2O, y''_1–y''_6, y''_8–y''_{13}
T74	376.5^{2+}	*TVIEYK*	a_2, b_2, y''_1–y''_5, y''_3-H_2O
T75	581.1^{2+}	*YNNYTSDER*	a_1, b_2–b_5, y''_1, y''_2, y''_4–y''_8
T80	581.2^{2+}	*VSLAMENIER*	a_2, b_2–b_5, b_3-H_2O, y''_1–y''_9
T82	344.1^{2+}	*LINEAK*	a_2, b_2, y''_1–y''_5, y''_4-NH_3
T85–86	753.3^{3+}	EY*DEGVKEYLLDYISEHR*	a_2, b_2, b_3, b_{12}, y''_1–y''_{11}, $[y''_{14}+H]^{2+}$, $[y''_{15}+H]^{2+}$, $[y''_{16}+H]^{2+}$
T86	719.2^{2+}	EY*LLDYISEHR*	a_2, b_2, b_3, b_3-H_2O, y''_1–y''_9, $[y''_9+H]^{2+}$, LLD
T87	904.3^{4+}	S*ILGN*SVQELNDLVTSTLNNSIPF*ELSSYTNDK*	b_2–b_5, b_{14}, y''_1–y''_9, y''_{11}, y''_2-H_2O, $[y''_{11}+H]^{2+}$
T88	512.2^{2+}	*ILILYFNK*	a_1, a_2, b_2, b_3, y''_1–y''_7
T92	482.1^{2+}	*DNSILDMR*	a_1, a_2, b_2–b_6, b_3-H_2O, b_4-H_2O, y''_1–y''_6
T98	571.3^{1+}	*YFNK*	a_1, a_2, b_2, b_3, b_3-H_2O, y''_1–y''_3
T99	869.4^{2+}	*VNLNNEYTIIDZIR*	a_1, b_2–b_{11}, b_4-NH_3, b_5-NH_3, b_6-NH_3, y''_1–y''_{12}
T100	410.2^{2+}	*NNNSGWK*	b_2, b_2-NH_3, y''_1, y''_3–y''_6, y''_5-NH_3
T101	426.2^{2+}	*ISLNYNK*	a_1, a_2, b_2, b_2-H_2O, y''_1–y''_6
T102	801.4^{2+}	*KIIWTLQDTAGNNQK*	a_1, a_2, b_2–b_9, y''_1–y''_{12}, y''_6-H_2O, y''_7-H_2O, TLQDTAGN
T106	646.3^{2+}	*IYINGNLIDEK*	a_1, a_2, b_2, y''_1–y''_{10}, y''_7-NH_3, y''_8-NH_3, y''_9-NH_3
T107	936.8^{2+}	SI*SNLGDIHVSDNILFK*	a_2, b_2–b_5, b_7, b_8, b_{10}–b_{15}, y''_1–y''_3, y''_5–y''_{14}
T108	468.1^{2+}	*IVGZNDTR*	a_1, a_2, b_2, y''_2–y''_7
T111	454.6^{2+}	*VFDTELGK*	a_1, a_2, b_2, b_3, y''_1–y''_7
T115	534.3^{2+}	*YYLLNLLR*	a_1, a_2, b_2, b_3, y''_1–y''_7

Table 8 (*Continued*)

Digest fragment[a]	m/z_{obsd} (Da)[b]	Peptide sequence[c]	Sequence ions[d]
T117	888.8^{2+}	SI*TQNSNFLNINQQR*	a_2, b_2–b_5, y''_1–y''_{12}
T118	762.4^{2+}	GV*YQKPNIFSNTR*	a_2, a_3, b_2–b_4, y''_1, y''_2, y''_4–y''_{10}
T119	581.8^{2+}	*LYTGVEVIIR*	a_1, a_2, b_2–b_6, b_9–b_{11}, y''_1–y''_9
T121	770.3^{2+}	NG*STDISNTDNFVR*	b_2, b_3, b_4-H_2O, y''_1, y''_3–y''_{11}
T125	674.2^{2+}	*LYADISIAKPEK*	a_1, a_2, b_2–b_3, y''_1–y''_{10}, $[y''_{11} + H]^{2+}$
T131	450.2^{2+}	EH*GWQEN*	a_2, b_2–b_6, b_4-H_2O, b_5-H_2O, b_6-H_2O, b_6-NH_3, y''_1, y''_2, y''_2-NH_3
NTNH[e]			
T5–6	432.7^{2+}	*KTDIFLK*	b_3–b_6, y''_1–y''_6
T7	813.4^{2+}	VF*QVAPNIWVAP*ER	a_2, b_2, b_3–b_5, y''_3, y''_4, y''_6–y''_{12}
T12	471.9^{2+}	*INNNVIGAK*	a_1, a_2, b_2–b_5, y''_1–y''_8
T13	815.0^{3+}	LL*SLISTAIP*FPYEYKPGDYR	a_2, b_2–b_{10}, $[y''_{12} + H]^{2+}$–$[y''_{19} + H]^{2+}$
T18	857.4^{2+}	*YDQFYVDPALELIK*	a_1, a_4, b_2–b_7, y''_1–y''_3, y''_5–y''_{12}
T23	869.3^{3+}	YSELDM*VDFLISGGTDYK*	b_6–b_{11}, y''_1–y''_{11}
T23	1027.3^{2+}	YS*ELDMVDFLISGGTDYK*	b_2–b_5, y''_1–y''_{14}
T24	753.3^{3+}	*LLNTNPYWVTDNYFSDAPK*	a_1, a_2, b_2–b_{11}, y''_1–y''_{12}
T34	620.3^{2+}	*EYYVIDYFK*	a_2, b_2–b_5, b_3-H_2O, b_4-H_2O, b_5-H_2O, y''_1–y''_8
T43	685.0^{2+}	NS*DPFIPVYNITE*TK	b_2–b_6, y''_1–y''_{12}
T45	559.4^{+}	*VISLK*	a_2, b_2, b_3, b_3-H_2O, b_4-H_2O, y''_1–y''_4
T47	702.7^{3+}	SL*VYS*FLDNT*IDYL*DSIK	a_2, b_2–b_5, b_{10}, y''_5–y''_8
Y68/T47	822.4^{2+}	SF*LDNTIDYLDSIK*	a_2, b_2, b_3, b_6–b_9, b_6-NH_3, y''_1–y''_{12}
T50	443.2^{2+}	*YYLWLK*	a_1, a_2, b_2, y''_1–y''_5
T54	604.0^{3+}	*ALNILNTGNSFIEEFK*	a_1, a_2, b_2–b_7, y''_1–y''_{10}
T54	905.5^{2+}	AL*NILNTGNSFIEEFK*	a_2, a_4, a_5, b_2–b_9, y''_1–y''_{14}
T58	932.0^{2+}	*IEIDEIPNSMLNLSFK*	a_1, a_2, b_2–b_6, y''_1–y''_{13}, y''_{10}-NH_3, EIDEIPN
T59	713.9^{2+}	*DLSENLFNIFSK*	a_1, a_2, b_2–b_{10}, b_3-H_2O, b_4-H_2O, b_5-H_2O, b_6-H_2O, b_7-H_2O, b_8-H_2O, b_9-H_2O, y''_1–y''_{10}
T63	544.2^{2+}	SV*LAQESLIK*	a_2, b_2–b_9, y''_1–y''_8, $[y''_8 + H]^{2+}$
T68	1188.2^{3+}	DISNESQ*IAMN*NVDSFLNSA*AIZVFEGNIY*PK	b_7–b_{10}, y''_2–y''_{10}, $[b_{29} + H]^{2+}$
T69	879.9^{2+}	*FISFMEQZINNINK*	a_1, a_2, b_2–b_{11}, y''_1–y''_{12}
T74	677.0^{3+}	*NIFSSLDFDFLNIENLK*	a_1, a_2, b_2–b_{10}, b_{12}, y''_1–y''_{10}, $[y_{15} + H]^{2+}$
T75	655.0^{2+}	SL*FSSETALLIK*	a_2, b_2–b_4, b_9–b_{11}, y''_1–y''_{10}
Y105	635.8^{+}	*ELVLY*	a_2, b_2–b_4, b_3-H_2O, b_4-H_2O, y''_2, y''_3
T77	471.0^{2+}	NT*SIEYSK*	a_2, b_2, b_5, b_6, y''_1–y''_6
T88	497.3^{2+}	NL*NNSYIR*	a_2, b_2, y''_1–y''_6
T89–90	698.8^{2+}	DS*NEERLEYNK*	a_2, b_2, b_4, b_6–b_{10}, y''_1–y''_8, $[y''_{10} + H]^{2+}$
T96	833.8^{2+}	*FDEVIISILDNMEK*	b_2–b_9, y''_1–y''_{13}
T97	562.7^{2+}	*YIDISEDNR*	a_1, a_2, b_2–b_4, y''_1–y''_8
T98	422.2^{2+}	*LQLIDNK*	a_1, b_2, b_3, y''_1–y''_6, y''_6-NH_3

[a] T for trypsin cleavage site, Y for chymotrypsin cleavage site, (ox) for oxidised (M–M sulphoxide).
[b] Observed *m/z* ratio, with apparent monoisotopic mass in Da.
[c] Italic characters indicate residues and position confirmed by sequence peaks; 'z' for carboxymethylcysteine.
[d] According to common Roepstorff-Fohlmann nomenclature [24].
[e] Sequence alignments of digest peptides refer to BTxF1 neurotoxin, *C. botulinum* strain Langeland NTNH (NCBInr access code X99064).

difference between the NTNH with this type E neurotoxin and other known type E NTNH sequences.

3.4. Botulinum toxin, type F

3.4.1. Available sequence information

Three complete botulinum neurotoxin type F gene sequences are known from as many *C. botulinum* strains. Information from the manufacturer stated that the purchased type F toxin was from *C. botulinum* strain Langeland. The toxin gene sequence from this strain, also known as strain NCTC 10281 (NCBInr X81714 [41]), is identical with that of an unnamed strain (L35496 [42]). Type F strains 202F (is ATCC 23387; M92906 [40]) and CDC 3281 (is ATCC 43757; Y13631 [43]) have quite similar neurotoxin sequences, different by about 15% from that of strain Langeland.[6] In addition, a type F botulinum toxin is known from *Clostridium baratii* strain ATCC 43756 (NCBInr X68262 [44]); that sequence differs by well over 20% from that of the *C. botulinum* strains. The amino acid sequences corresponding to these gene sequences are further referred to as BTxF1 (Langeland), BTxF2 (202F), BTxF3 (CDC 3281), and BTxF4 (*C. barati* ATCC 43756), respectively. For the *C. botulinum* type F strain investigated here, strain Langeland, the neurotoxin gene sequence was partly confirmed before, in two independent studies [45,46].

[6] Sequence alignment for a direct comparison cannot be done in an unequivocal way, due to apparent frameshift and deletion mutations.

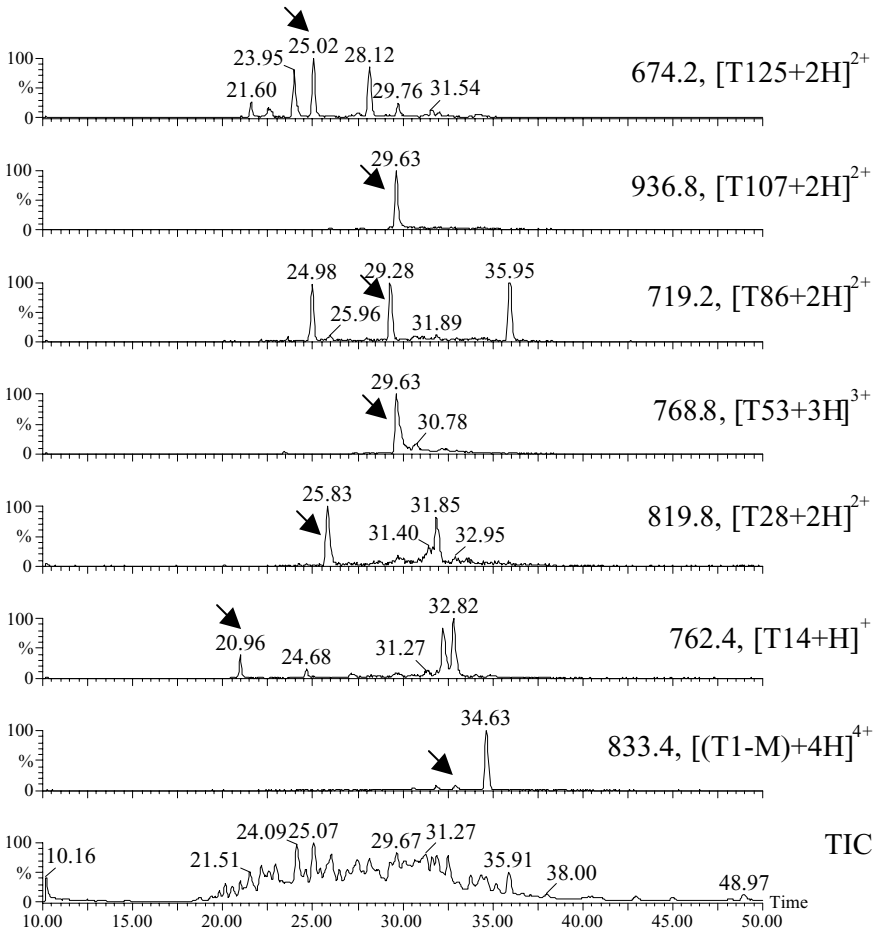

Fig. 6. TIC (bottom) and selected ion chromatograms from an LC–ES MS analysis of a BTxF trypsin digest; peptides eluted between 10 and 50 min and arrows indicate peaks with identity confirmed by MS and MS/MS spectra; masses (in Da) and digest peptides given with chromatogram.

In addition, the gene sequence of the NTNH of strain Langeland is known (NCBInr X99064; [47]). The available genetic information does not show evidence of the presence of HA components in toxin genes from any BTxF producing *Clostridia*. That implies that BTxF, at least from strain Langeland, appears mainly as a complex of an NTNH and the neurotoxin.

3.4.2. MALDI mass spectrometric peptide mapping

The MALDI mass spectrum of the BTxF trypsin digest (not shown) provided a good peptide map: BTxF was the rank 1 candidate in a ProFound [23] search (limits: bacteria, masses within 100 ppm), where 24 peptides matched. Of the remainder of peptides, 10 matched a corresponding NTNH; 5 NTNH digest peptides have masses indistinguishable from BTxF digest peptides. The matching peptides are summarised in Table 7.

This primary MALDI identification was confirmed by a PSD experiment on one of the attributed trypsin digest peptides of BTxF, T117, and one of NTNH, T7. Fig. 5 shows the partial PSD spectrum of BTxF T117, with clear evidence of the C-terminal sequence, LNINQQR, of the attributed T117.

Similarly, the PSD spectrum of the attributed T7 of NTNH displayed sequence ion signals of the expected C-terminal sequence WVAPER. Hence, MALDI PSD confirmed the provisional identification made from the trypsin digest peptide map.

3.4.3. LC–ES MS(/MS) sequencing

In agreement with the genetic information available on strain Langeland, only peptides of type F neurotoxin and of an NTNH component were observed. Results of the LC–ES MS/MS experiments are summarised in Table 8. The observed sequences cover 53 and 31% of the neurotoxin and NTNH sequence, respectively. All peptides exactly matched those predicted from the available gene sequences. A TIC and selected ion chromatograms are given in Fig. 6.

Only a few peculiarities were observed, all due to the sample treatment. The T66 peptide of the neurotoxin, AINNSLMER, was found with a methionine sulphoxide along the non-oxidised form. After this observation, all other methionine containing peptides were also checked for possible oxidation. However, no other signs of this typical ageing reaction were found. Signs of low chymotrypsin activity were only observed with peptides from

the NTNH component. The NTNH T47 peptide, SLVYSFLDNTIDYLDSIK, occurred as such, but also as the N-terminally truncated SFLDNTIDYLDSIK. In addition, a small peptide ELVLY was found that represents a complete Y105 chymotryptic peptide (from ... YELVLYA ... that falls within T76). These minor procedure artefacts did not give serious problems in the characterisation of the neurotoxin or the NTNH. All protein sequence information obtained was in line with that known from *C. botulinum* strain Langeland DNA sequences.

4. Conclusions

The presently reported mass spectrometric experiments with neat botulinum toxin types C, D, E, and F complete the characterisation of botulinum toxins started with experiments on types A and B [8]. The type G neurotoxin is not commercially available for characterisation, but C. botulinum type G is rare and the neurotoxin toxicity has not been fully established (only two reports have appeared, one in 1993 [48], and one in 1997 [49]). In addition, there is no reason why the method should not work with the type G neurotoxin, because it is chemically of the same class as the other botulinum toxins.

The characterisation has shown that the various types of botulinum toxin can be distinguished, sometimes to the level of the *C. botulinum* strain the material originates from. The characterisation has also shown that the information obtained is excessive with regard to identification of the toxin: identification only requires sequencing of a few peptides, with subsequent positive attribution of the sequences to one of the botulinum toxin types. It was noted that for the type C and D neurotoxins, the peptides used in identification should cover amino acid sequence across the molecule to allow distinction of 'pure' types C and D from C and D mosaic type neurotoxins. In general, distribution of identified peptides across the entire neurotoxin protein sequence is to be preferred, where this is a restrictive requirement in type C and D identification. Given the detail of distinction attainable by chemical analysis, serological distinction of C1 and D type neurotoxins obscures, rather than clarifies the differences.

Further work is on-going to identify botulinum toxins in a 'real-world' sample of weaponised material.

Acknowledgements

The authors thank The Netherlands Ministry of Defence for continuous support.

References

[1] The Biological and Toxin Weapons Convention, signed at London, Moscow and Washington on 10 April 1972 entered into force on 26 March 1975.
[2] E.R.J. Wils, A.G. Hulst, Rapid Commun. Mass Spectrom. 7 (1993) 413.
[3] R. Draisci, L. Gianetti, L. Lucentini, C. Marchiafava, K.J. James, A.G. Bishop, B.M. Healy, S.S. Kelly, J. Chromatogr. A 798 (1998) 137.
[4] T. Krishnamurthy, C.R. Hauer, M. Prabhakaran, J.G. Freedy, K. Hauashi, Biol. Mass Spectrom. 23 (1994) 719.
[5] C.E. Kientz, A.G. Hulst, E.R.J. Wils, J. Chromatogr. A 757 (1997) 51.
[6] B.L.M. van Baar, A.G. Hulst, E.R.J. Wils, Toxicon 37 (1998) 85.
[7] B.L.M. van Baar, A.G. Hulst, B. Roberts, E.R.J. Wils, Anal. Biochem. 301 (2002) 278.
[8] B.L.M. van Baar, A.G. Hulst, A.L. de Jong, E.R.J. Wils, J. Chromatogr. A 970 (2002) 95.
[9] B. Kimura, S. Kawasaki, H. Nakano, T. Fujii, Appl. Environ. Microbiol. 67 (2001) 206. AB040123 (strain Iwanai), AB040124 (Tenno2), AB040125 (Biwako), AB040126 (5545), AB040127 (35396), AB040128 (164-1).
[10] R. Ekeus, Report of the Secretary General on the status of the implementation of the Special Commission's plan for the ongoing monitoring and verification of Iraq's compliance with relevant parts of Sector C of Security Council Resolution 687, United Nations Special Commission (UNSCOM), New York, 1991.
[11] D.W. Riegle Jr., A.M. D'Amato, US Chemical and Biological Warfare-Related Dual Use Exports to Iraq and their Possible Impact on the Health Consequences of the Gulf War, United States Senate, 103d Congress, 2d Session, 25 May 1994 (Chapter 1, part 2, 'US Exports of Biological Materials to Iraq').
[12] H. Barth, M. Klingler, K. Aktories, V. Kinzel, Infect. Immun. 67 (1999) 5083.
[13] K. Moriishi, M. Koura, N. Fujii, Y. Fujinaga, K. Inoue, B. Syuto, K. Oguma, Appl. Environ. Microbiol. 62 (1996) 662.
[14] Y. Sagane, T. Watanabe, H. Kouguchi, T. Yamamoto, T. Kawabe, F. Murakami, M. Nakatsuka, T. Ohyama, Published Only in DataBase (2000) NCBI access AB037166.
[15] D. Hauser, M.W. Eklund, P. Boquet, M.R. Popoff, Mol. Gen. Genet. 243 (1994) 631.
[16] K. Kimura, N. Fujii, K. Tsuzuki, T. Murakami, T. Indoh, N. Yokosawa, K. Takeshi, B. Syuto, K. Oguma, Biochem. Biophys. Res. Commun. 171 (1990) 1304.
[17] Y. Sagane, H. Kouguchi, T. Watanabe, H. Sunagawa, K. Inoue, Y. Fujinaga, K. Oguma, T. Ohyama, Biochem. Biophys. Res. Commun. 288 (2001) 650.
[18] D. Hauser, M.W. Eklund, H. Kurazono, T. Binz, H. Niemann, D.M. Gill, P. Boquet, M.R. Popoff, Nucleic Acids Res. 18 (1990) 4924.
[19] D.F. Hauser, M.W. Eklund, M.R. Popoff, 1992, NCBI access X66433, unpublished data.
[20] D. Hauser, M. Gibert, M.W. Eklund, P. Boquet, M.R. Popoff, J. Bacteriol. 175 (1993) 7260.
[21] K. Moriishi, M. Koura, N. Abe, N. Fujii, Y. Fujinaga, K. Inoue, K. Oguma, Biochim. Biophys. Acta 1307 (1996) 123.
[22] T. Binz, H. Kurazono, M.R. Popoff, M.W. Eklund, G. Sakaguchi, S. Kozaki, K. Krieglstein, A. Henschen, D.M. Gill, H. Niemann, Nucleic Acids Res. 18 (1990) 5556.
[23] W. Zhang, B.T. Chait, Anal. Chem. 72 (2000) 2482–2489 (Search facility accessible at http://prowl.rockefeller.edu and at http://proteometrics.com/).
[24] P. Roepstorff, J. Fohlman, Biomed. Mass Spectrom. 11 (1984) 601.
[25] T. Ohyama, T. Watanabe, Y. Fujinaga, K. Inoue, H. Sunagawa, N. Fujii, K. Inoue, K. Oguma, Microbiol. Immunol. 39 (1995) 457.
[26] K. Tsuzuki, K. Kimura, N. Fujii, N. Yokosawa, T. Indoh, T. Murakami, K. Oguma, Infect. Immun. 58 (1990) 3173.
[27] H. Kouguchi, T. Watanabe, Y. Sagane, H. Sunagawa, T. Ohyama, J. Biol. Chem. 277 (2002) 2650.
[28] Y. Nemoto, T. Namba, S. Kozaki, S. Narumiya, J. Biol. Chem. 266 (1991) 19312.

[29] H. Sunagawa, T. Ohyama, T. Watanabe, K. Inoue, J. Vet. Med. Sci. 54 (1992) 905.
[30] D.N. Perkins, D.J. Pappin, D.M. Creasy, J.S. Cottrell, Electrophoresis 20 (1999) 3551.
[31] S. Poulet, D. Hauser, M. Quanz, H. Niemann, M.R. Popoff, Biochem. Biophys. Res. Commun. 183 (1992) 107. NCBI access X62088.
[32] X. Wang, T. Maegawa, T. Karasawa, S. Kozaki, K. Tsukamoto, Y. Gyobu, K. Yamakawa, K. Oguma, Y. Sakaguchi, S. Nakamura, Appl. Environ. Microbiol. 66 (2000) 4992. NCBInr access AB037704 (strain KZ 1885), AB037705 (KZ 1899), AB037706 (KZ 1897), AB037707 (KZ 1898), AB037708 (KZ 1886), AB037709 (KZ 1887), AB037710 (KZ 1889), AB037711 (KZ 1890), AB037712 (KZ 1891), AB037713 (LCL 063), and AB037714 (LCL 095).
[33] K. Tsukamoto, M. Mukamoto, T. Kohda, H. Ihara, X. Wang, T. Maegawa, S. Nakamura, T. Karasawa, S. Kozaki, Published Only in Database (2002) NCBInr access AB088207 (Strain BL 5262).
[34] S.M. Whelan, M.J. Elmore, N.J. Bodsworth, T. Atkinson, N.P. Minton, Eur. J. Biochem. 204 (1992) 657.
[35] K. Tsukamoto, M. Mukamoto, T. Kohda, H. Ihara, X. Wang, T. Maegawa, S. Nakamura, T. Karasawa, S. Kozaki, Published Only in Database (2002), NCBInr access code AB082519.
[36] K.D. Campbell, M.D. Collins, A.K. East, J. Clin. Microbiol. 31 (1993) 2255. NCBInr access X70815 (strain ATCC 9564) and X70818 (VH Dolman).
[37] N. Fujii, K. Kimura, N. Yokosawa, T. Yashiki, K. Tsuzuki, K. Oguma, J. Gen. Microbiol. 139 (1993) 79. NCBInr access D12697.
[38] B. Li, X. Qian, H.K. Sarkar, B.R. Singh, 1996, NCBInr access U70780, unpublished data.
[39] N. Fujii, K. Kimura, N. Yokosawa, K. Oguma, T. Yashiki, K. Takeshi, T. Ohyama, E. Isogai, H. Isogai, Microbiol. Immunol. 37 (1993) 395.
[40] A.K. East, P.T. Richardson, D. Allaway, M.D. Collins, T.A. Roberts, D.E. Thompson, FEMS Microbiol. Lett. 96 (1992) 225.
[41] R.A. Hutson, M.D. Collins, 1994, NCBInr access code: X81714 (gi|971349), unpublished data.
[42] M.J. Elmore, N.J. Bodsworth, S.M. Whelan, N.P. Minton, 1994, NBBInr access code L35496, unpublished data.
[43] J.A. Santos-Buelga, M.D. Collins, A.K. East, Curr. Microbiol. 37 (1998) 312.
[44] D.E. Thompson, R.A. Hutson, A.K. East, D. Allaway, M.D. Collins, P.T. Richardson, FEMS Microbiol. Lett. 108 (1993) 175.
[45] A.K. East, M. Bhandari, J.M. Stacey, K.D. Campbell, M.D. Collins, Int. J. Syst. Bacteriol. 46 (1996) 1105. NCBInr X87974.
[46] J.L. Ferreira, M.K. Hamdy, S.G. McCay, M. Hemphill, N. Kirma, B.R. Baumstark, Mol. Cell. Probes 8 (1994) 365. NCBInr S76749.
[47] A.K. East, M. Bhandari, S. Hielm, M.D. Collins, Curr. Microbiol. 37 (1998) 262.
[48] K. Campbell, M.D. Collins, A.K. East, Biochim. Biophys. Acta 1216 (1993) 487.
[49] M. Bhandari, K.D. Campbell, M.D. Collins, A.K. East, Curr. Microbiol. 35 (1997) 207.