ELSEVIER

Journal of Chromatography A, 970 (2002) 95–115

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Characterisation of botulinum toxins type A and B, 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

Abstract

A method earlier developed for the mass spectrometric (MS) identification of tetanus toxin (TTx) was applied to botulinum toxins type A and B (BTxA and BTxB). Botulinum toxins are extremely neurotoxic bacterial toxins, likely to be used as biological warfare agent. Biologically active BTxA and BTxB are comprised of a protein complex of the respective neurotoxins with specific haemagglutinins (HAs) and non-toxic non-haemagglutinins (NTNHs). These protein complexes are also observed in mass spectrometric identification. The particular BTxA complex, from *Clostridium botulinum* strain 62A, almost completely matched database data derived from genetic sequences known for this strain. Although no such database information was available for BTxB, from *C. botulinum* strain okra, all protein sequences from the complex except that of HA-70 were found to match proteins known from other type B strains. It was found that matrix-assisted laser desorption ionisation MS provides provisional identification from trypsin digest peptide maps and that liquid chromatography electrospray (tandem) mass spectrometry affords unequivocal identification from amino acid sequence information of digest peptides obtained in trypsin or pepsin digestion.

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

1. Introduction

Biological warfare agent (BWA) use in military and terrorist attacks poses a potential threat to NATO forces, either on the battlefield or in the theatre of a peacekeeping deployment. Among these biological warfare agents, toxins form a specific class. The use of toxin BWA as a military weapon is subject to the 1972 Biological and Toxin Weapons Convention (BTWC; [1]), with the exception of ricin and saxitoxin, which two toxins are scheduled for control and monitoring under the 1997 Chemical Weapons Convention (CWC; [2]). However, control and monitoring of toxins and BWA are not regulated in the BTWC and discussions on a sound verification regime are on-going. Identification methods providing irrefutable proof, by the standards of the best technical means, will be required to strengthen the BTWC to include an effective arms control regime comparable to that of the CWC.

At the TNO Prins Maurits Laboratory (TNO-PML), the development of identification procedures for toxins takes place currently within the Programme on Passive NBC Defence. The work started several years ago, with a study on saxitoxin [3], and has been extended to investigations of protein toxins with increasing molecular size, i.e., staphylococcal enterotoxin B (SEB) [4], cholera toxin [5] and

*Corresponding author. Tel.: +31-15-284-3514; Fax: +31-15-284-3963.

E-mail address: baar@pml.tno.nl (B.L.M. van Baar).

0021-9673/02/$ – see front matter
PII: S0021-9673(02)00508-3

tetanus toxin [6]. The ultimate aim is a universal procedure, applicable to all protein toxins.

The present report concerns the characterisation of botulinum toxin (BTx), essentially employing the method developed in the characterisation of tetanus toxin (TTx) [6]. Both BTx and TTx belong to a family of potent bacterial neurotoxins, the clostridial toxins. The potency is reflected by an LD_{50} in the order of 1 ng/kg (mouse, intravenous; [7]) and a generally accepted human LD_{50} in the same range. In threat evaluations, BTx is most widely regarded as the most likely toxin to be applied on the battlefield or in a terrorist attack in a civil environment (see, for example, Refs. [8,9]). BTx for planned military use has actually been found in Iraq [10], whereas a failed terrorist dissemination of BTx has been reported from Japan [11]. As a consequence of this potential, BTx has been scheduled in the Protocol to strengthen the Biological and Toxin Weapons Convention [12] and has appeared on the export control list of the Australia Group (on the ''core list'' [13]).

The name ''botulinum toxin'' refers to a number of different compounds, all produced by different members of the bacteria family *Clostridium*, under anaerobic conditions: strains of *C. botulinum* produce BTx, but some strains of *C. barati* and *C. butyricum* produce the toxin as well [14,15]. Because of this diversity, use of the plural ''botulinum toxins'' (BTxs) is more appropriate. BTxs have, so far, been distinguished in seven serotypes: type A through type G. However, genetic analysis of some of the toxin producing bacteria has shown that there is considerable biovariation within each of these serotypes. Also, the mechanism of action, as far as it is known, is not the same for all serotypes (see, for example, Ref. [16]). Hence, a more detailed indication of the particular BTx at hand is required in characterisation.

This paper gives an account of the characterisation of BTx types A and B, 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. Type A (BTxA), was from *C. botulinum* strain 62A (also known as NCTC 3805), whereas type B (BTxB), was from *C. botulinum* strain okra. The amino acid sequence of strain 62A BTxA neurotoxin is completely known from nucleotide sequencing of the toxin genes [17]. In contrast, the amino acid sequence of strain okra BTxB was only known for less than 10%, from N-terminal protein sequencing experiments [18,19]. However, the amino acid sequence of BTxB neurotoxin from some other type B strains was completely known from nucleotide sequencing. In the Results section, characterisation of both toxins is discussed separately.

2. Experimental

2.1. Notice of caution

BTxs are extremely toxic and their handling requires extensive safety measures. All handling of μg 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. In case of emergency, horse antisera against BTx A, B and E were available from the ''Rijksinstituut voor Volksgezondheid en Milieuhygiëne'' (RIVM, Bilthoven, The Netherlands); in the course of the presently discussed research, no such emergency occurred.

2.2. Materials

BTx type A, from strain 62A, and BTx type B, from strain okra, and bis(2-mercaptoethyl)sulfone (BMS) were purchased from Calbiochem (La Jolla, CA, USA). Tris and Tris–HCl (Trizma), sodium acetate (analytical-reagent grade), and sodium chloride (analytical-reagent grade) were obtained from Sigma (Zwijndrecht, The Netherlands), whereas TPCK treated trypsin was obtained from Sigma–Aldrich (Steinheim, Germany). Pepsin was obtained from Roche (Mannheim, Germany), whereas 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 (Buchs, Switzerland). Bleach was available from laboratory stock as a solution of 25% technical quality sodium hypochlo-

rite in tap water. All water used in analytical procedures was drawn from a Milli-Q system (Millipore, Milford, MA, USA). Acetonitrile (LiChrosolv quality) and formic acid (analytical-reagent grade) were purchased from Merck (Darmstadt, Germany). Recrystallised α-cyano-4-hydroxycinnamic acid, used as a MALDI matrix, was purchased from Bruker Daltonik (Bremen, Germany). Molecular mass cut-off (MWCO) 30 000 and 50 000 filters were purchased from Amicon (Dronten, The Netherlands). ZipTip were obtained from Millipore.

2.3. Sample preparation for mass spectrometric analysis

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

2.3.1. Trypsin digestion

The sample treatment procedure previously developed with tetanus toxin [6] 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 m*M* EDTA at pH 8.4). Subsequently, disulfide 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 50 °C (45 min, in the dark). Free cysteine residues were then derivatised with an excess of sodium 2-iodoacetate (7 mg in 50 to 100 μl of the above guanidine buffer), at 40 °C (30 min, in the dark). 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 standing 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 needles and pipette tips used were rendered harmless by immersion in bleach.

2.3.2. Pepsin digestion

Pepsin digestion was used as an alternative to trypsin digestion. Handling of the sample was done in the same way as for trypsin digestion, but specific sample treatment was slightly different. Typically, 100 μg of the purchased toxin was reconstituted with 100 μl water (to give a solution buffered with 50 m*M* sodium acetate and 0.2 *M* sodium chloride). 200 μl of 1% (v/v) aqueous formic acid was added and the solution was distributed over two MWCO 50 000 filters (5 min, 11,000 rpm). The residue on each filter was then washed with 300 μl 1% (v/v) aqueous formic acid (15 min, 11 000 rpm). Subsequently, 400 μl of a 10 μg/ml solution of pepsin in 5% (v/v) aqueous formic acid was brought onto each MWCO filter, the filters were then kept at 37 °C for 1 h, and were centrifuged (15 min, 11 000 rpm). The filtrate was used in analysis, outside the containment, without further treatment. Any remaining washing solutions and residues and all disposable needles and pipette tips used were rendered harmless by immersion in bleach.

2.4. MALDI mass spectrometry

All MALDI-MS experiments were conducted on 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 or pepsin 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 flushed with 10 μl of a 0.1% trifluoroacetic acid (TFA) solution and eluted with 10 μl matrix solution (saturated α-cyano hydroxycinnamic acid with 0.1% TFA in acetonitrile–water, 1:2, v/v). Of the matrix–sample mixture 0.5 μl was brought on 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 on 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 six-port valve (Valco, Schenkon, Switzerland) with a 10 μl 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 (water with 0.2%, v/v, formic acid) and B (acetonitrile with 0.2%, v/v, 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, 0.5 ml/min flow) to 20% A and 80% B (at 60 min, 0.5 ml/min flow). The flow delivered by the liquid chromatograph was split pre-column to allow a flow of approximately 6 μl/min through the column and into the ES-MS interface.

The mass spectrometer was operated at a cone voltage of 25–35 V and employing nitrogen 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 and pepsin digest analyses.

3. Results and discussion

BTx type A and BTx type B were analysed following the sample preparation developed for tetanus toxin [6], which encompasses disulfide bridge reduction, cysteine–thiol derivatisation and trypsin digestion. After initial experiments with BTxA it turned out that the neurotoxin was still associated with other proteins through non-covalent bonds; such complexes have been described before (see, for example, Refs. [20–22]). Briefly, one molecule of 150 kDa neurotoxin can be associated with a ''non-toxic non-haemagglutinin'' (NTNH) protein, in a 300 kDa complex. The 300 kDa complex can be additionally associated with a variety of haemagglutinating proteins (HAs). The stoichiometry of these HAs in a complex is not known, but complexes of 500 kDa and of 900 kDa have been reported (see, for example, Ref. [21]). Hence, BTxs characterisation by MS is complicated, because of the presence of the other proteins, each with their own biovariations. This is in contrast to TTx [6], where the neurotoxin appears as a single neurotoxin molecule, without any additional protein factors.

The particular complexes studied here, type A from strain 62A and type B from strain okra, turned out to be of the 500 or 900 kDa type. Therefore, the initial characterisation of the neurotoxin was extended to NTNH and HA factors. Identification of these factors may point to the presence of BTx. For the sake of clarity of presentation, results on type A and type B toxin will be discussed separately, each after a brief introduction on available knowledge with respect to protein structure.

3.1. Botulinum toxin type A, from C. botulinum strain 62A

3.1.1. Available sequence information

Three gene nucleotide sequences are known for the precursor protein of BTx type A (known as bontoxylysin A). These are designated by NCBI database [23] accession codes: M30196 from strain 62A [17], X52066 from strain NCTC 2916 [24]; X73423 from strain Kyoto-F [25] (further referred to as BTxA1, BTxA2, and BTxA3, respectively). An additional protein sequence is available from a crystal structure of the neurotoxin chain of BTx type A (NCBI protein identity code gi|6137387 [26], strain not specified). The crystal structure amino acid sequence agrees completely with that of the BTxA2 nucleotide sequence minus the N-terminal methionine (M) residue. Removal of N-terminal M is a common post-translational modification in eukaryotic proteins [27]. Taking BTxA1 (M30196) as the wild-type sequence and using amino acid residue

numbering in reference to the precursor protein, the sequence difference between BTxA1 and BTxA2 is limited to two amino acids (proline to glutamine at position 2, or P2Q, and A27V), which differences map to the N-terminal first and fourth trypsin digest fragment (T4) of both protein sequences. The difference between BTxA1 and BTxA3 comprises 130 amino acid residues[1], but still leaves an 89% protein sequence identity. Because of this large difference BTxA3 is sometimes referred to as ''type A2'', where BTxA1 and BTxA2 are then ''type A1''. As a consequence of the extensive sequence difference, BTxA3 would give 146 trypsin digest peptides, whereas BTxA1 and BTxA2 would each give 144; of these peptides, 25 or 22, respectively, have a length with little or no sequence information (three or less residues).

[1]In reference to BTxA1: E171D (T18), G229E (T21), I293V (T29), V304I (T31) K356N (T39 and T40 in A1 vs. T39 in A3), L361I (T41 vs. T40), K375R (T44 vs. T43), K381D, V382E, Y387K, R393K (T45–46 vs. T44–45), N394G, T395A, A398S A399T, N410S, M411R, K415R (T47 vs. T46–47), T436P, S437F (T51), K444E (T53–54 vs. T53), N476D (T56 vs. T55), G478V, S483A, N507D, L530P (T57 vs. T56), K563D (T62–63 vs. T61) A566I, V571A, N572E, N578K, S580N, R581V, V582A, D589K (T64–65 vs. T62–63), V594I (T67 vs. T65), T598V, A601F, G605N, V607A, Q609E, S619N, S622T, T624M (T68 vs. T66), I634V, Y648S (T69 vs. T67), D650G, D651E, G654E, L656I, S659T, A661V, I663A, L664M, I671Y, I673L, L676F, L681I (T70 vs. T68), D696N (T71 vs. T69), I713T (T75 vs. T73), K730E (T77 vs. T75), E734K (T80 vs. T77), K779S (T83–84 vs. T81), N789D, G804A (T85 vs. T82), E809K, L815V, K816R (T87 vs. T84–85), A818V (T88 vs. T 86), I831V, G832L (T90 vs. T88), K840E, T847A (T92–93 vs. T90), Q860K, R861K (T94 vs. T92–93), I874V, N880S, L881I, R882V, E884K, S885K, N886D, H887D (T96–97 vs. T94–96), A895G, S896A (T98 vs. T97), S902D, K903R (T99 vs. T98), N905Y, F906Y, P908S (T110 vs. T99), Q915K, F917I, K923T (T101–102 vs. T100–101), R948K (T103 vs. T102), N954S, S955K, S957N, M968I (T105 vs. T104–105), T990N, Q991K, E992Q, I993N, K994I (T106–107 vs. T106–107), I1005V (T109), N1025T, N1026K (T111 vs. T111–112), N1052K (T113 vs. T114–115), T1063P, H1064R (T115 vs. T117–118), W1068M (T116 vs. T119), N1090S (T120 vs. T123), D1103N, Y1117F (T121 vs. T124), V1128I (T122 vs. T125), M1144G, S1153T, R1156E (T125–126 vs. T128), K1170E (T129–130 vs. T131), N1227D, T1232R (T138 vs. T 139–140), Q1254L, F1255Y, N1256D (T140 vs. T142), I1271V, E1272G, R1273K (T142 vs. T144), S1274A (T143 vs. T145), L1278F, R1294S, P1295S (T144 vs. T146).

3.1.2. MALDI-MS of trypsin digested material

A MALDI-MS spectrum of the trypsin digested sample type A, from strain 62A was obtained; an example of a typical spectrum is given in Fig. 1.

The spectrum provides a typical ''digest peptide map'', consisting of 65 average peptide masses (transformed from observed $[M+H]^+$), which can be used for database searching against known protein sequences. In this case, a database search was performed using ProFound [28]; search limitations were: against the non-redundant NCBI (NCBInr) protein sequence database [23], with a mass window of ± 0.5 Da, the taxonomic category ''bacteria'' and the relevant cysteine modification. This gave a best match for BTxA, with over 20% sequence coverage for the match with BTxA1 and BTxA2 sequence and 15% for BTxA3. BTxA clearly lead the score, with a probability of 1.0, versus 0.0003 for the second in rank. In reference to BTxA1 and BTxA2, the spectrum fit is summarised in Table 1.

A further search with the 49 remaining unmatched masses still gave a reasonable fit (probability 0.27, rank 2) for the *C. botulinum* 70 kDa haemagglutinin, HA-70 (specifically, with gene sequences NCBI nucleotide access code L42537 [29] and Y13630 [30]); the fitting peptides are summarised in Table 1. This data effectively pinned down two proteins, BTxA and HA-70, for a more detailed confirmation. Although such confirmation could have been obtained from MALDI-PSD experiments, we chose to use electrospray MS and MS–MS.

3.1.3. LC–ES-MS(–MS) of trypsin digested material

First, an LC–ES-MS run was performed with the crude trypsin digest mixture. The data was manually checked for m/z with abundant signals and ion chromatograms were then constructed for these m/z. The m/z observed were converted to the corresponding $[M+H]^+$ ion masses, by simple calculation. The list of peptide masses, thus obtained, is much larger than that obtained from a direct MALDI-MS analysis. Although MALDI and ES do differ in their ionisation efficiency, the more abundant information from LC–ES-MS is also due to the fact that separation of peptides prevents suppression

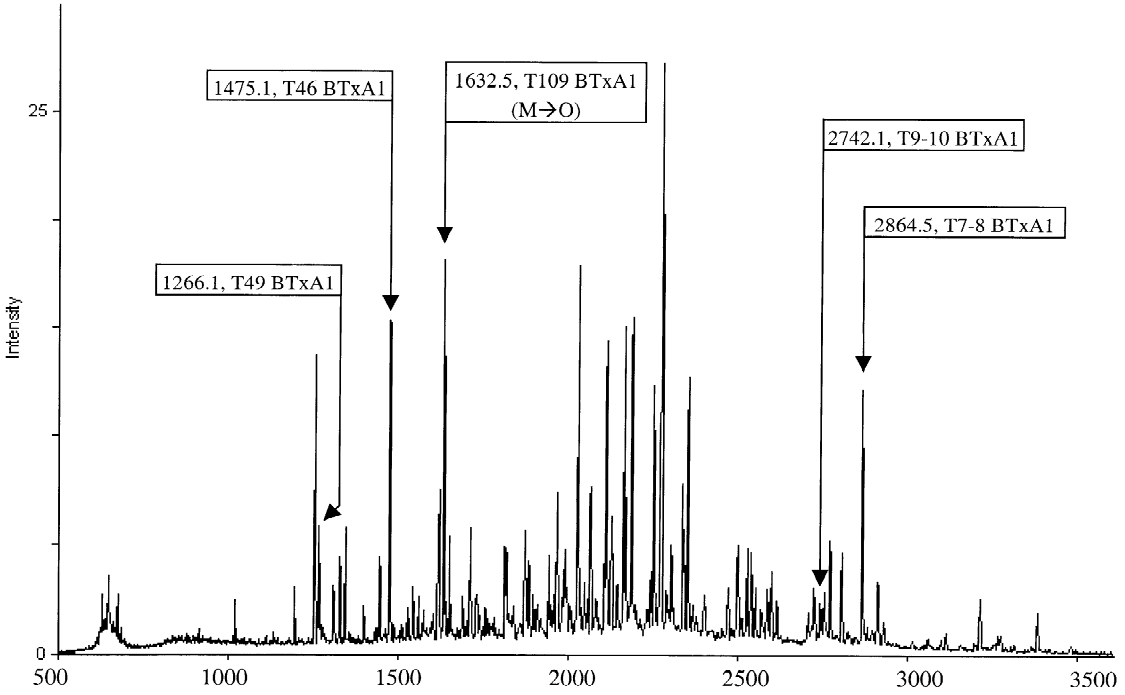

Fig. 1. MALDI-MS mass spectrum of the trypsin digest peptides from a botulinum toxin type A preparation, annotated with selected peptides (''M→O'' indicates a match for possible methionine oxidation).

Table 1
Summary of matching masses in the MALDI spectrum (Fig. 1) of a BTxA trypsin digest

Observed mass (Da)[a]	Calculated mass (Da)[a]	Assigned peptide	Observed mass (Da)[a]	Calculated mass (Da)[a]	Assigned peptide
2864.5	2864.4	T7–8, BTxA1/2	2027.1	2027.0	T4, HA-70
2742.1	2742.3	T9–10, BTxA1/2	2537.1	2537.2	T8, HA-70
1965.1	1964.9	T16, BTxA1/2	1309.1	1308.8	T14, HA-70
2351.1	2351.1	T24, BTxA1/2	2498.4	2498.4	T25–26, HA-70
1398.0	1397.7	T26, BTxA1/2	1021.1	1021.5	T32, HA-70
1603.0	1602.8	T39, BTxA1/2	1651.0	1650.8	T32–33, HA-70
1475.1	1474.7	T46, BTxA1/2	2550.2	2550.4	T41–42, HA-70
2470.0	2470.1	T47, BTxA1/2	2597.2	2597.2	T49, HA-70
1266.1	1265.6	T49, BTxA1/2	2803.2	2803.3	T36, HA-70
1712.2	1711.9	T64, BTxA1/2			
3380.4	3380.6	T81–82, BTxA1/2			
1271.2	1270.8	T96, BTxA1/2			
1560.1	1559.8 (both)	T99–100 or T117–118, BTxA1/2			
1264.1	1263.7	T110, BTxA1/2			
2125.2	2125.2	T136, BTxA1/2			
2335.2	2335.1	T144, BTxA1/2			

[a] Monoisotopic mass of $[M+H]^+$.

effects in ionisation. In a further comparison to MALDI-MS, a list of 81 peptide masses from LC–ES-MS was subjected to a similar database search. ProFound [28] search with the crude LC–ES-MS derived masses (81 ''peptide masses'') in the non-redundant NCBI protein sequence database [23], using a mass window of ±0.2 Da, the taxonomic category ''bacteria'' and the relevant cysteine modification, gave a best fit for BTxA, with over 35% sequence coverage for the match with BTxA1 and BTxA2 sequence and 20% for BTxA3. In analogy to the MALDI results, BTxA lead the score, but with sharper distinction to any alternative: a probability of 1.0 versus $2.0 \cdot 10^{-51}$ for the second in rank. A further search with the 36 remaining unmatched masses only gave a poor fit (probability 0.0003, rank 5) for the *C. botulinum* NTNH. Again, this data effectively pinned down BTxA for a more detailed confirmation.

Second, the m/z observed and used in the database search were selected for MS–MS experiments, in order to obtain amino acid sequence information. After the initial database match of the LC–ES-MS peptide map to BTxA, the data was also checked with special attention for all m/z of $[M+nH]^{n+}$, with n between 1 and 4, of digest peptides possibly resulting from trypsin digestion of BTxA (and with a length of three or more amino acid residues). This did not result, however, in additional precursors selection for MS–MS. Apparently the other expected fragments were not present in appreciable quantity, due to losses in sample treatment, or were suppressed in the ionisation process. All peptides thus selected were used to direct time-scheduled LC–ES-MS–MS analysis of precursor ions. There is a software option for automatic selection of precursor ions for MS–MS and immediate switching between MS and MS–MS in a single LC run. However, the peptide mixture turned out to be too complex for successful automated operation. Instead, manual operation and repetitive LC–ES-MS–MS runs, with optimisation of collision energy when necessary, produced a vast quantity of detailed sequence data. A typical product ion MS–MS spectrum is given in Fig. 2, whereas the sequence information obtained is summarised in Table 2.

From these MS–MS results, 40 peptides were unambiguously attributed to either of the BTxA1 or BTxA2 sequences. The observation of the T4 peptide allows a specific limitation to BTxA1, because this peptide contains the A27 residue specific to BTxA1 (V27 in BTxA2). Thus, the peptide sequences observed are in full agreement with BTxA1 from *C. botulinum*, strain 62A.

A further 22 peptide sequences were attributed to four different proteins known from various strains of *C. botulinum*. This was achieved by subjecting each non-BTxA sequence to a comparative search in the NCBInr nucleotide database [23] using the translated Basic Local Alignment Search Tool (BLAST, in ''tblastn'' mode [32]). The BLAST search results were further restricted by a manual check of the compatibility of the protein amino acid sequences with trypsin cleavage sites for the matched peptide sequence. Where individual peptides fit many proteins from different *C. botulinum* types and strains, even with the trypsin cleavage restriction, the combined sequence coverage of several peptides from one protein was distinctive in most cases.

Eight of the 22 peptides belonged to NTNH. Seven amino acid sequences out of these eight (T7, T26, T51, T62, T100, T101, and T102) agree with the NTNH sequences known from the corresponding genes of *C. botulinum* type A, strains 62A (NCBInr access X92973 [33]) and NIH (D67030 [21]). The eighth peptide, T35, was unique to the strain 62A NTNH: it contained the only different amino acid residue (G333 in 62A and E333 in NIH). Thus, the NTNH sequence data is in full agreement with the known strain 62A NTNH sequence.

No nucleotide or amino acid sequences are available for the strain 62A HA proteins. However, BLAST searching provided good matches for many of the MS–MS sequenced peptides to either of three HAs, HA-17, HA-33, and HA-70 (the numbers denote kDa and derive from gel electrophoresis). The smaller, HA-17, is represented by three peptides, which together match two identical HA-17 sequences from *C. botulinum* type A, strain NCTC 2916 (from the A gene cluster of this species; NCBI access L42537 [29]) and from type B, strain ATCC 43757 (from the B gene cluster of this species; Y13630 [30]). These three peptides, T3, T7, and T9, correspond with a 20% coverage of the 146 amino acid

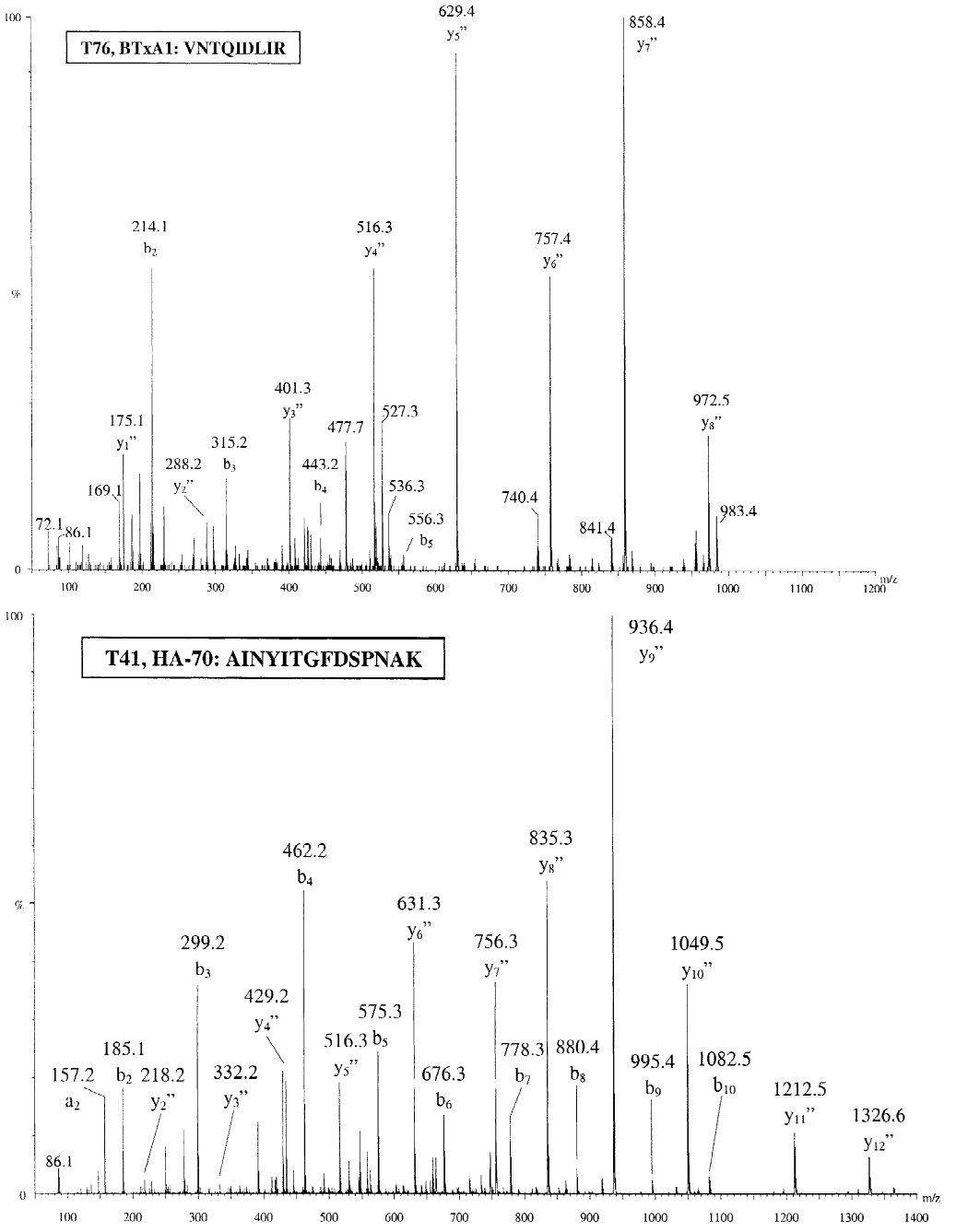

Fig. 2. Product ion MS–MS spectra of $[T76+2H]^{2+}$ from BTxA (top) and of $[T41+2H]^{2+}$ from the accompanying HA-70 (bottom); annotation only covers singly charged b and y″ type fragment ions (following Roepstorff–Fohlman nomenclature [31]).

Table 2
Summary of trypsin digest peptide sequence information obtained for botulinum toxin type A, from strain 62A

Digest fragment	m/z_{obsd} (Da)[a]	Peptide sequence[b,c]	Sequence ions[d]	Digest fragment	m/z_{obsd} (Da)[a]	Peptide sequence[b,c]	Sequence ions[d]
BTxA1[e]				T105	971.4^{3+}	*YFNSI*SLNNEY*TIINZMENNSGWK*	a_1, a_2, b_3–b_5, b_{12}–b_{14}, y_1''–y_{13}''
T2	350.2^{2+}	QFNYK	Not determined	T106	750.8^{3+}	*VSLNYGEIIWTLQDTQEIK*	a_2, b_2–b_9, y_1''–y_{10}''
T4	591.8^{2+}	*IPNAGQMQPVK*	a_2, b_2–b_5, b_7, b_8, y_1'', y_3''–y_{10}'', $[y_{10}+H]^{2+}$	T109	808.9^{2+}	YSQMINISDYINR	Not determined
T7–8	955.8^{3+}	*IWVIP*ERDTFTN*PEEGDLNPPPEAK*	b_2–b_4, y_1''–y_{11}'', y_{13}'', $[y_{21}+H]^{2+}$–$[y_{23}+H]^{2+}$	T116	361.7^{2+}	*YIWIK*	a_1, a_2, y_1''–y_4''
T9–10	914.8^{3+}	*QVPVSYYD*STYLSTDNEKDNYLK	a_2, b_2, y_{11}'', $[y_{16}''+H]^{2+}$–$[y_{21}''+H]^{2+}$, $[y_{21}+2H]^{3+}$	T120	733.9^{2+}	DL*YDNQSNSGILK*	a_2, b_2–b_4, b_6, b_{11}, b_{12}, y_1''–y_{11}'', y_9''–NH_3, y_{10}''–NH_3, $[y_{11}''+H]^{2+}$
T10	326.7^{2+}	DNYLK	Not determined	T122	574.8^{2+}	*YVDVNNVGIR*	a_1, a_2, a_4, b_2–b_4, b_9, y_1''–y_9''
T13	462.7^{2+}	*IYSTDLGR*	a_1, a_2, b_2, y_1''–y_7'', y_6''–H_2O, y_7''–H_2O	T135	594.8^{2+}	*LATNASQAGVEK*	a_1, a_2, b_2–b_5, y_1''–y_{11}''
T14	466.8^{2+}	*MLLTSIVR*	a_1, a_2, b_2, b_3, y_1''–y_7''	T136	709.1^{3+}	*ILSALEIPDVGNLSQVVVMK*	a_1, a_2, b_2–b_{12}, y_1''–y_{13}''
T15	810.9^{2+}	*GIPFWGGSTIDTELK*	a_2, b_2–b_7, y_2''–y_{13}'', $[y_{13}+H]^{2+}$	T140	907.8^{3+}	*MNLQDNN*GNDIGFIGFHQFNNI*AK*	a_1, b_2–b_6, y_1'', $[y_{17}''+H]^{2+}$–$[y_{23}''+H]^{2+}$
T16	655.7^{3+}	*VIDTNZINVIQPDGS*YR	a_2, b_2, b_3, b_7–b_9, b_8–NH_3, b_9–NH_3, y_4''–y_8'', y_{10}'', y_7''–NH_3	T141	561.8^{2+}	*LVASNWYNR*	a_1, a_2, b_2, b_3, y_3''–y_8''
T18	636.8^{2+}	SF*GHEVLNLTR*	a_2, b_4, y_1'', y_2'', y_4''–y_9'', GHE	T144	779.0^{3+}	TL*GZSWEFIPVDDGWGERPL*	b_2, b_7–b_9, b_{14}, b_8–H_2O, b_9–H_2O, y_3''–y_{11}'', $[y_{11}''+H]^{2+}$
T20	906.4^{3+}	*FSPDFTFGFEESLEVDTNPLLGAGK*	b_4–b_{15}, $[b_{15}+H]^{2+}$, $[b_{17}+H]^{2+}$, y_2''–y_{12}''	NTNH[e]			
T22	565.8^{2+}	*LYGIAINPNR*	a_2, b_2–b_4, y_1''–y_9''	T7	626.3^{2+}	*VAPNIWVAPER*	a_2, b_2, b_5–b_8, y_1'', y_3''–y_{10}'', $[y_9''+H]^{2+}$
T24	784.4^{3+}	*VNTNA*YYEM*SGLEVSFEELR*	b_5, b_{11}–b_{13}, y_1''–y_{11}''	T26	915.8^{3+}	*SELENIEYSQLNIVDLLVSGGIDPK*	b_4, b_6–b_8, b_{10}–b_{13}, $[b_{16}+H]^{2+}$, $[b_{17}+H]^{2+}$, $[b_{18}+H]^{2+}$, $[b_{19}+H]^{2+}$, $[b_{21}+H]^{2+}$,$[b_{22}+H]^{2+}$,y_1''–y_{12}'', y_{14}''
T25	416.7^{2+}	*TFGGHDAK*	a_2, b_2, y_1''–y_7'', GGHD, GGHDA	T35	666.0^{3+}	*ININDIWELNLNYFSK*	b_2 –b_5, b_7, b_8/y_8'', b_9, b_{10}, b_7–NH_3, b_8/y_8''–NH_3, y_1''–y_7'', y_9'', y_{10}''
T26	699.3^{2+}	*FIDSLQENEFR*	a_1, a_2, b_2–b_4, y_1''–y_{10}''	T51	765.9^{2+}	*FSLSSDFVEVVSSK*	a_1, b_2–b_4, b_6, b_8, b_{10}, b_{11}, y_1''–y_{13}''
T27	432.2^{2+}	LYYYNK	Not determined	T62	1068.2^{3+}	*ENLSMPIIESYEI*PND*MLGLPLNDLNEK*	b_2–b_5, b_7–b_{13}, b_{18}–b_{20}, b_{23}, b_{25}, $[b_{10}+H]^{2+}$, $[b_{11}+2H]^{3+}$, y_1'', y_3''–y_{12}'', y_{15}''–y_{21}'', y_{23}'', y_{24}'', $[y_{11}''+H]^{2+}$, $[y_{12}''+2H]^{3+}$, $[y_{14}''+2H]^{3+}$
T31	699.9^{2+}	*SIVGTTASLQYMK*	a_2, b_2–b_5, y_1''–y_{11}''	T100	809.4^{2+}	*LDEVIISVLDNMEK*	a_1, a_2, b_2–b_6, y_1''–y_{13}''
T34	556.8^{2+}	*YLLSEDTSGK*	a_1, a_2, b_2, y_1''–y_9''	T101	562.8^{2+}	*YIDISEDNR*	a_1, a_2, b_2–b_4, y_1''–y_8''
T39	801.9^{2+}	*MLTEIYTEDNFVK*	a_1, a_2, b_2–b_5, b_4–H_2O, y_1''–y_{12}'' (y_6'' /b_6 and y_7'' /b_7 isobaric pairs)	T102	422.0^{2+}	*LQLIDNK*	a_1, b_2, b_3, y_1''–y_6'', y_6''–NH_3
T43	450.7^{2+}	*TYLNFDK*	a_2, b_2, y_1''–y_6''	HA-17[e]			
T45	342.2^{2+}	*INIVPK*	a_1, a_2, b_2, y_1''–y_5'', y_3''–H_2O, y_5''–H_2O	T3	756.2^{2+}	SI*FSGSLYLNPVSK*	a_2, b_2–b_5, y_2'', y_4''–y_{12}''
T46	737.9^{2+}	*VNYTIYDGFNLR*	b_2–b_5, y_1''–y_{11}''	T7	486.0^{2+}	*ISNVAEPNK*	a_1, b_2–b_4, y_1''–y_8''
T49	633.3^{2+}	NF*TGLFEFYK*	a_2, b_2, b_3, y_1''–y_9''	T9	507.7^{2+}	*ZYWFPIK*	a_2, a_2–H_2O, b_2, y_1'', y_3''–y_6''
T64	856.5^{2+}	*IALTNSVNEALLNPSR*	a_1, b_2–b_{13}, y_1''–y_{12}''	HA-33[e]			
T65	678.3^{2+}	*VYTFFSSDYVK*	a_2, b_2, b_3, y_1''–y_{10}''	T2	411.0^{2+}	*IVTISZK*	a_1, a_2, b_2, y_1''–y_6''
T68	1128.5^{3+}	*ATEAAMFLGW*VEQLVYDF*TDETSEVSTTDK*	b_4–b_{10}, $[b_{15}+H]^{2+}$, $[b_{16}+H]^{2+}$, $[b_{25}+H]^{2+}$, $[b_{26}+H]^{2+}$, y_1''–y_{12}''	T6	427.0^{2+}	*LIYDSNK*	a_1, a_2, b_2, y_1''–y_6'', y_5''–H_2O
T71	701.4^{2+}	VLTVQTIDNALSK	Not determined	T13	633.9^{2+}	*LSTLNNSNYIK*	a_1, a_2, b_2–b_5, y_1''–y_{10}''
T75	554.3^{2+}	YIVTNWLAK	Not determined	T14	1054.1^{2+}	FI*IEDYIISDLNNFTZK*	b_2–b_7, y_1''–$y_{14''}$
T76	536.3^{2+}	*VNTQIDLIR*	a_1, a_2, b_2–b_5, y_1''–y_8'', y_6''–NH_3, y_7''–NH_3, y_8''–NH_3	T15	506.8^{2+}	*ISPILDLNK*	b_2, y_2''–y_8'', $[y_7''+H]^{2+}$
T80	602.3^{2+}	*EALENQAEATK*	a_1, a_2, b_2–b_5, y_1''–y_{10}'', b_3–H_2O, b_4–H_2O, y_7''–NH_3, y_8''–H_2O	HA-70[e]			
T83	409.2^{2+}	*LNESINK*	a_1, b_2, y_1''–y_6'', y_5''–H_2O, y_6''–H_2O	T9	920.5^{3+}	NN*IQTVFTNFTEANQIPIGFEFSK*	b_2–b_8, y_8''–y_{13}'', $[y_8''+H]^{2+}$
T87	519.3^{2+}	DNYLKGVTK	Not determined	T13	654.9^{2+}	*AVLYVPSLGYVK*	a_2, b_2–b_5, y_8''–y_{11}''
T90	479.8^{2+}	*GTLIGQVDR*	b_2, b_3, b_3–H_2O, y_1''–y_7''	T18	665.2^{2+}	*ZILNEQFLYK*	a_2, b_2, b_3, y_1''–y_9'', $[y_8''+H]^{2+}$
T93	839.0^{2+}	VN*NTLSTDIPFQLSK*	b_2, y_1''–y_3'', y_5''–y_{13}''	T19–20	416.8^{2+}	*KILETTK*	b_1, y_1''–y_6'', y_5''–H_2O, y_6''–H_2O
T95	607.8^{2+}	LLSTFTEYIK	Not determined	T32	511.3^{2+}	*FVEEAPSDK*	a_1, a_2, b_2, y_1''–y_8''
T96	635.9^{2+}	*NIINTSILNLR*	a_2, b_2, b_3, y_1''–y_{10}''	T41	755.9^{2+}	*AINYITGFDSPNAK*	a_2, a_4, a_5, b_2–b_{10}, y_1'' –y_{12}''
T102	357.7^{2+}	*IEVILK*	a_1, a_2, b_2, b_3, y_1''–y_5'', y_4''–NH_3, y_5''–NH_3				

[a] Observed m/z ratio, with apparent monoisotopic mass in Da.

[b] Amino acid character codes: A=alanine, C=cysteine, D=aspartic acid, E=glutamic acid, F=phenylalanine, G=glycine, H=histidine, I=isoleucine, K=lysine, L=leucine, M=methionine, N=asparagine, P=proline, Q=glutamine, R=arginine, S=serine, T=threonine, V=valine, W=tryptophan, Y=tyrosine, Z=*S*-carboxymethylcysteine.

[c] Italic characters indicate residues and position confirmed by sequence peaks.

[d] According to common Roepstorff–Fohlman nomenclature [31].

[e] Sequence alignments of digest peptides refer to BTxA1 neurotoxin sequence from *C. botulinum* strain 62A (NCBInr access code M30196), NTNH sequence from strain 62A (NCBInr access code X92973), HA-17 and HA-70 sequences from strain NCTC 2916 (access L42537), and HA-33 sequence of strain NCTC 7272 (access X79104).

residues of the HA-17 sequence. The middle size HA, HA-33, is represented by five peptides, which together match one single HA-33: of *C. botulinum* type A, strain NCTC 7272 (X79104 [34]), with a 15% coverage of the 293 amino acid sequence. The larger size HA, HA-70, is represented by six peptides, of which five match one single HA-70: of *C. botulinum* type A, strain NCTC 2916 (from the A gene cluster of this species; NCBI access L42537 [29]). However, the observed sequence of the sixth peptide, AINYITGFDSPNAK, does not match any HA-70 protein sequence exactly (see Fig. 1). Since it largely matches T41 of the NCTC 2916 sequence (L42537), AINYITGVDDPNTK, the particular peptide must be a T41 fragment of HA-70 specific to strain 62A. Despite the fact that no sequences of HA proteins were available for strain 62A, the proteins were identified as HA factors of *C. botulinum* type A.

3.2. Botulinum toxin type B, from C. botulinum strain okra

3.2.1. Available sequence information

Three complete gene nucleotide sequences are known for the precursor protein of BTxB, bontoxylysin B, from different BTxB producing strains: M81186 from strain Danish (BTxB1 [35]), X71343 from strain Eklund 17B (ATCC 25765; BTxB2 [36]), and Y13630 from strain ATCC 43757 (CDC 3281; BTxB3 [30]). The latter, strain ATCC 43757, produces type B and F toxin and the BTxB3 sequence derives from the B-gene cluster. These BTxB sequences produce 154 or 155 trypsin digest fragments, of which 34 or 36 span less than three amino acid residues. In addition, some partial nucleotide sequences of BTxB are known, among others from BTxB silent genes, for example of strains 667Ab (X87849 [37]) and NCTC 2916

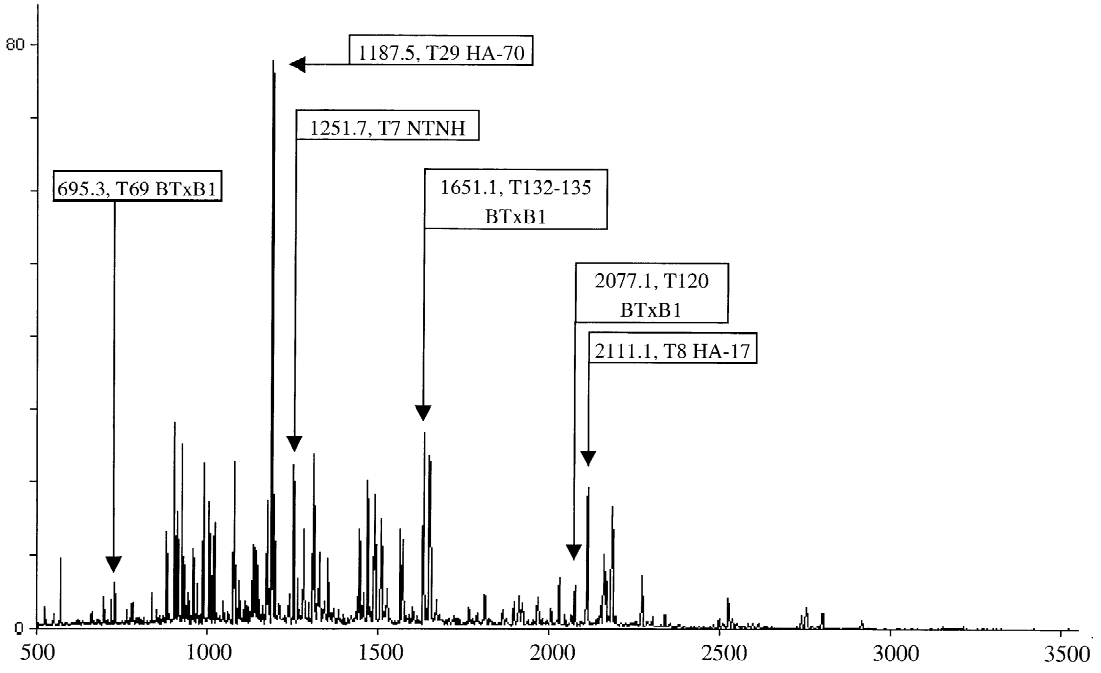

Fig. 3. MALDI-MS mass spectrum of the trypsin digest peptides from a botulinum toxin type B preparation; the annotated signals of non-BTxB peptides were identified by MALDI-PSD; attributed BTxB1 signals give an impression of relative intensities of the digest peptides.

(Y14239 [38]). The differences between the three complete neurotoxin amino acid sequences comprise 91 amino acid residues for BTxB1 and BTxB2, and 50 residues for BTxB1 and BTxB3[2] . For further reference throughout this paper, BTxB1 is taken as the wild-type sequence.

Little sequence information is available for the *C. botulinum* type B strain okra reported here. In particular, the N-terminal sequences of the L and H chains of the toxin are known from amino acid sequencing [18,19]. The 44 residues of the N-terminal sequence of the L-chain[3] agree with the sequence of BTxB2, of course with the exception of the N-terminal methionine involved in post-translational modification. The only difference with the L-chain N-termini from BTxB1 and BTxB3 is a T for an *M* in the strain okra sequence, at position 30 (M30T). The 26 residues of the N-terminal of the H-chain[4] agree for the larger part with either of the BTxB genetic sequences, when aligned onwards from residue 442 of the bontoxylysin B sequence; one unidentified amino acid fits C446 (all BTxB), whereas S464 in BTxB1 and BTxB3 aligns with R in the strain okra H-chain N-terminal sequence. The H-chain N-terminal sequence gives some information about the post-translational modification that is involved in the formation of the L-chain and H-chain structure from the entire bontoxylysin B protein (''nicking''): an enzymatic cleavage must have occurred between amino acid residues 441 and 442.

Interestingly, N-terminal sequences of the *C. botulinum* type B strain okra of three HA factors are also known [20]. They comprise the HA-52 N-terminal VFVTQRVLPY . . . and the HA-22 N-terminal VINYSDTIDL . . . . These two HA factors result from post-translational modification of the type B, strain okra, HA-70. The N-terminal reported for the type B, strain okra, HA-22 implies that the 15 N-terminal residues of HA-70 have been cut off. The HA-52 N-terminal is at variance with all known sequences of type B HA-70 and matches best with VSSTQRVLPY . . . observed in strain NCTC 2916 [29]. With respect to the known nucleotide sequences of type B HA factors, the HA-17 N-terminal misses the methionine residue and has 31N instead of the common 31S (for example from strain Lamanna [39]). The HA-33 from type B, strain okra, seems truncated, with the amino acid sequence missing the five N-terminal residues in comparison to the known nucleotide sequence from other type B HA-33 (for example from strain Lamanna).

The fact that there is only little sequence information available for the type B, strain okra, toxin and associated NTNH and HA factors means that peptide mapping may be hampered: deviant amino

[2]BTxB2 in reference to BTxB1: N17D (T1), L89F, M92L (T11), A263T (T29), D346N (T40), F354L, T359I (T42), D404N, E406G (T48), A442V, D453N (T56), I473V, S479N, E483G (T59), I530V, D549N (T63), F565V, N567S (T64), N599D (T68), N607S (T69), N632D (T70), N641S, A648S, A665V (T71), N690V (T76), S695I (T77), R737K (T80), K744E (T81–82 of BTxB1 vs. T81 of BTxB2), D752N (T83 vs. T82), E762D, I768M, N773D, G777E (T84 vs. T83), E793K (T86 vs. T85–86), A821V, Y823D (T90), N829D (T92), M836I, I842T, T844S, D846I, T847E, E851K, M852I (T94 vs. T94–95), K871R (T96 vs. T106), E892K (T98 vs. T99–100), N906D (T100 vs. T102), V922M (T102 vs. T104), K938R (T104 vs. T106), G941D (T105 vs. T107), N1012D (T113 vs. T115), K1021T, T1026M (T114–115 vs. T116), R1032G, A1036V, I1041T (T116–117 vs. T117), I1048V (T118), E1065Q, S1067N, E1072K, R1074I (T120–121), K1113V (T126), P1117S, T1123I (T128 vs. T127), K1132N (T130–131 vs. T129), D1138N (T132 vs. T130), K1150E (T135–136 vs. T133), Y1168H, F1171L, F1172V, N1173L, L1174H, N1175H, Q1176E (T138 vs. T135), T1182A (T139 vs. T136), K1188E, E1189Q (T141–142 vs. T138), A1196S, P1197I, D1202N, N1206K, Q1209E (T143 vs. T139–140), T1218S (T144 vs. T141), E1234D (T146 vs. T143), I1248V, V1249L, F1250R, E1251K, E1252K (T147 vs. T144–147), N1273K, L1274S, K1275N (T152–153); BTxB3 in reference to BTxB1: T30M (T2), D245N (T27), T281S (T29), S372A (T45), D404N (T48), E474A, S479N, P487S (T59), D517Y, E522K (T60 of BTxB1 vs. T60–61 of BTxB3), N599D (T68 vs. T69), N609S (T69 vs. T70), K680E, D683N, N684S (T73–74 vs. T74), N690D (T76), S694I (T77), R737K (T80), K746R (T82), I756V (T83), G777E (T84), K803R (T87), N829D (T92), I835S, M836I, I842T, D846N, M852I (T94), E859D (T95), K871R (T96), N874K (T97 vs. T97–98), E892K (T98 vs. T99–100), T912I, V922M (T102 vs. T104), R970M (T108–109 vs. T110), T982I (T110 vs. T111), N990S, R992K (T111 vs. T112), L1011S, N1012D (T113 vs. T114), N1025H, T1026I, K1029R (T115 vs. T116), G1038D (T117 vs. T118), D1047N (T118 vs. T119), R1074I (T120–121 vs. T121), P1117S (T128), T1182M (T139), E1251K (T147 vs. T147–148), L1274S (T152 vs. T153).

[3]The strain okra L-chain N-terminal sequence is: PVTIN NFNYN DPIDN NNIIM MEPPF ARGMG RYYKA FKITD RIWI [18,19].

[4]The strain okra H-chain N-terminal sequence is: APGIXIDVDNEDLFFIADKNSFRDDL, where ''X'' was unidentified [18,19].

acid sequences will not give adequate matches to database derived digest peptide masses. However, this lack of information did not pose a major problem in practice.

3.2.2. MALDI-MS of trypsin digested material

A MALDI mass spectrum of trypsin digested BTxB material from strain okra is given in Fig. 3.

A search with the trypsin digest peptide map, following the approach used for the BTxA MALDI mass spectrum gave a BTxB ranking of 0.71 versus 0.14 for the second and third in rank (both not even *C. botulinum* proteins). The ranking difference is relatively small and typical of poor distinction in protein identification. Apparently, the fact that the BTxB starting material was not a single pure protein prevents adequate matching of the peptide map data, much in contrast to the observations for BTxA. An obvious problem is then that the toxin might not be recognised from a MALDI mass spectrum of a peptide map of an unknown sample.

A manual check of the peptide map against BTxB1 was performed; the assigned peptide masses are summarised in Table 3.

Of the peptides typically observed, 15 to 20 matched either BTxB1, BTxB2 or BTxB3. Therefore, distinction of these neurotoxin proteins was not possible. Another 20 peptides in the peptide map were attributed to NTNH or one of the HA. This corroborates the observation that the mixture of proteins obscures the presence of the neurotoxin in a peptide map.

Given the poor database match of peptide maps, PSD mass spectra were obtained from a few selected $[M+H]^+$ precursor ions. The spectra obtained (not shown) were used in a Mascot database search [40], and produced some sequence information that allowed assignment of the protein material to *C. botulinum* NTNH and HA-70, particularly FSQQYTEER (1187.5 Da), VAPNIWVAPER (1251.7 Da), and YLSYDNFGFISLDSLSNR (2111.1 Da), of which the observed $[M+H]^+$ ions are identified in Fig. 3. Identification of such peptides allows a limitation of a subsequent search to *C. botulinum* proteins only, with a better support for positive identification of the neurotoxin. As a consequence, an initial identification of botulinum toxin in an unknown sample can only be made by a combination of MALDI peptide mapping and selected peptide sequencing by MALDI-PSD.

3.2.3. LC–ES-MS(–MS) of trypsin digested material

Masses corresponding with peaks from LC–ES-MS ion chromatograms were selected as precursor ions for LC–ES-MS–MS. In contrast to BTxA, the LC–ES-MS trypsin digest peptide map was not used to perform database searching. Instead, the ES-MS–MS sequence data from trypsin and pepsin digest peptides was directly used for identification. Table 4 gives a summary of all results, in reference to the known sequence of BTxB1 [35]. Fig. 4 gives a typical LC–ES-MS spectrum and corresponding MS–MS spectrum from the toxin material.

From all MS–MS results on this toxin, 31 peptide amino acid sequences were in complete agreement

Table 3
Summary of matching neurotoxin digest peptide masses in the MALDI spectrum (Fig. 3) of a BTxB trypsin digest

Observed mass (Da)[a]	Calculated mass (Da)[a]	Assigned peptide	Observed mass (Da)[a]	Calculated mass (Da)[a]	Assigned peptide
1412.2	1411.8	T5–6, BTxB1	1324.2	1324.7	T99–100, BTxB1
1508.2	1508.7	T7, BTxB1	2077.0	2077.0	T120, BTxB1
2152.9	2152.9	T9–10, BTxB1	1130.3	1130.6	T122, BTxB1
2031.2	2031.1	T16–17, BTxB1	1384.2	1384.6	T123, BTxB1
914.4	914.5	T32–33, BTxB1	1671.2	1671.0	T126–129, BTxB1
1400.2	1402.8	T65, BTxB1	837.4	837.4	T132, BTxB1
1813.1	1813.9	T65–66, BTxB1	1651.1	1651.0	T132–135, BTxB1
695.4	695.3	T69, BTxB1	1511.2	1510.7	T147, BTxB1
1604.3	1603.9	T79, BTxB1	1865.1	1864.9	T149–152, BTxB1
1922.0	1921.9	T79–80, BTxB1			

[a] Monoisotopic mass of $[M+H]^+$.

Table 4
Summary of sequence information obtained for botulinum toxin type B strain okra, by trypsin digest[c]

Digest fragment[a]	m/z_{obsd} (Da)[b]	Peptide sequence[c,d]	Sequence ions[e]	Digest fragment[a]	m/z_{obsd} (Da)[b]	Peptide sequence[c,d]	Sequence ions[e]
BTxB1[f]				NTNH[f]			
T6	463.7^{2+}	*IWIIPER*	a_2, b_2, y''_1, y''_3–y''_6	T7	626.3^{2+}	*VAPNIWVAPER*	a_2, b_2, b_5–b_8, y''_1, y''_3–y''_{10}, $[y''_9+H]^{2+}$
T8	390.7^{2+}	*SSGIFNR*	a_2, b_2, b_3, b_3–H_2O, y''_1–y''_6	T18	650.5^{2+}	*ETNYLSSEDNK*	a_1, b_2, b_3, b_4, b_2–H_2O, b_3–H_2O, b_4–H_2O, b_5–H_2O, y''_1–y''_9
T11	554.3^{2+}	NI*FLQTMIK*	a_2, b_2, b_3, y''_1, y''_2, y''_4–y''_7	T22	864.9^{2+}	*YDEFYIDPAIELIK*	a_1, b_2–b_7, y''_1–y''_{13}
T14	379.8^{2+}	*SKPLGEK*	a_2–H_2O, b_2, b_2–H_2O, y''_1–y''_5, $[y''_5+H]^{2+}$, KPLGE, PLGE	T43	483.7^{2+}	*SNIYGDGLK*	b_2–b_4, y''_1, y''_3–y''_8
T18	557.3^{2+}	*LISNPGEVER*	a_2, b_2–b_4, b_4–H_2O, y''_1–y''_9	T44	670.4^{2+}	STV*DDFYSNYK*	b_2–b_4, b_3–H_2O, y''_1–y''_9
T22	549.2^{2+}	*EGFGGIMQMK*	b_3, b_3–H_2O, a_5–H_2O, b_6–H_2O, b_9–H_2O, y''_1–y''_9, $[y''_8+H]^{2+}$	T51	765.9^{2+}	*FSLSSDFVEVVSSK*	a_2, b_2–b_5, b_3–H_2O, y''_1–y''_{13}
T27	619.8^{2+}	*VDDLPIV*PNEK	a_4, b_2–b_4, b_7, y''_1, y''_4–y''_9	T56	457.2^{2+}	*YYLWLR*	a_1, a_2, b_2, y''_1–y''_5, y''_3–NH_3
T31	388.7^{2+}	*VLQNFR*	a_2, b_2, y''_1–y''_5	T67	458.8^{2+}	*QSILAQEK*	a_1, a_2, y''_1–y''_7, $[M+2H–H_2O]^{2+}$
T39	455.7^{2+}	*FVEDSEGK*	a_1, a_2, b_2, y''_1–y''_7	T70	382.2^{2+}	*LQDLFK*	a_1, b_2, b_3, y''_1–y''_5, y''_5–NH_3
T40	601.8^{2+}	*YSIDVESFDK*	a_1, a_2, b_2, b_4, b_3–H_2O, b_5–H_2O, y''_2–y''_9	T73	841.7^{3+}	TFIDLSNESQIAINNINDFLNK	b_2–b_8, b_{10}–b_{13}, $[b_{11}+H]^{2+}$, $[b_{12}+H]^{2+}$, $[b_{13}+H]^{2+}$, y''_1–y''_{12}
T42	932.9^{2+}	SL*MFGFTETNIAENYK*	a_2, b_2–b_7, b_5–NH_3, b_6–NH_3, b_7–NH_3, y''_1–y''_{13}	T87	545.2^{2+}	*NLGEDIITSK*	a_2, b_2, y''_1–y''_8
T45	655.8^{2+}	AS*YFSDSLPPVK*	a_2, b_2, y''_1, y''_3, y''_5–y''_{10}	Y131/T94	565.8^{2+}	*SYLNNSYIR*	a_2, b_2, b_3, y''_1–y''_7, $[y''_7+H]^{2+}$
T54	340.7^{2+}	*IQMZK*	a_1, b_2, b_4, y''_1–y''_4, y''_4–NH_3	T100	450.7^{2+}	*LINIDANK*	a_1, a_2, b_2, y''_1–y''_7
T65/Y68	662.2^{+}	*VYSFF*	a_2, a_4, a_4–H_2O, b_2–b_4, b_3–H_2O, y''_1–y''_4, $[M+H–H_2O]^{+}$	T111	455.2^{2+}	AA*HLWALK*	b_3, b_4, y''_1, y''_3–y''_6, $[y''_6+H]^{2+}$
T67	638.5^{2+}	*VVEAGLFAGWVK*	a_2, b_2–b_4, b_8, b_9, y''_1, y''_2, y''_4–y''_{10}	HA-33[f]			
T68	695.4^{2+}	Q*IVNDFVIEANK*	a_2, a_2–NH_3, b_2, b_3, b_5–NH_3, y''_1–y''_{10}, $[M+2H–H_2O]^{2+}$	T2	411.2^{2+}	*IVTISZK*	a_2, b_2, y''_1–y''_6
Y75/T70	650.4^{2+}	*IGLALNVGNETAK*	a_4, a_5, b_2–b_5, y''_1–y''_{12}	T6	426.7^{2+}	*IIYDSNK*	a_2, b_2, y''_1–y''_6
T74	438.2^{2+}	*TIDNALTK*	a_2, b_2, y''_1–y''_7	Tx[g]	455.2^{2+}	*LYADTVAR*	a_2, b_2, b_3, y''_1–y''_7
T79	802.4^{2+}	AL*NYQAQALEEIIK*	a_2, a_4, b_2–b_6, b_6–NH_3, y''_1–y''_{12}	T13	620.3^{2+}	*LSTLNNSSYIK*	a_1, a_2, b_2, y''_1–y''_{10}
T83	747.3^{2+}	SN*INIDFNDINSK*	b_2–b_5, y''_2–y''_{11}	T14	744.8^{2+}	*FIIEDYVISDFK*	a_1, a_2, b_2, y''_1–y''_{10}
T87	539.8^{2+}	*LLDFDNTLK*	a_1, a_2, b_2, y''_1–y''_8	T16	428.2^{2+}	*ISPILAGGK*	a_1, a_2, b_2, y''_1, y''_3–y''_8, $[y''_7+H]^{2+}$
T89	618.3^{2+}	NL*LNYIDENK*	a_2, b_2, b_3, y''_1–y''_8	T18	598.2^{2+}	*WTIIYNEEK*	a_1, a_2, b_2, b_3, y''_1–y''_8
T90	643.3^{2+}	*LYLIGSAEYEK*	a_1, a_2, a_4, b_2, b_3, b_4, y''_1–y''_{10}, GSAE	T19	494.7^{2+}	*AAYQFFNK*	a_2, b_2, b_3, y''_1–y''_7, y''_5–NH_3
T95	844.9^{2+}	*YNSEILNNIILNLR*	a_1, b_2–b_5, y''_1–y''_{12}	/T20[h]	504.7^{2+}	*IFSDGNTVR*	a_1, a_2, b_2, y''_1, y''_2, y''_5–y''_8
T111	587.4^{2+}	*SVFFEYNIR*	a_2, b_2–b_4, y''_1–y''_7, $[y''_7+H]^{2+}$	Y33/T21	653.3^{2+}	*LINPVSDNYDR*	a_1, a_2, b_2, b_3, y''_1, y''_3–y''_{10}, $[y''_8+H]^{2+}$, $[y''_9+H]^{2+}$
T115	460.2^{2+}	*LESNTDIK*	a_2, b_2, y''_1–y''_7, y''_6–NH_3, y''_7–H_2O	Y32/T21	746.2^{2+}	WL*INPVSDNY*DR	a_2, a_2–NH_3, b_3–b_8, b_4–NH_3, b_5–NH_3, b_6–NH_3, y''_3–y''_{11}
T128	543.8^{2+}	*DSPVGEILTR*	a_2, a_3, a_3–H_2O, b_2, b_3, y''_1, y''_3–y''_8, $[y''_8+H]^{2+}$	T22	440.7^{2+}	*YTITNLR*	a_1, a_2, b_2, y''_1–y''_6
T131	364.7^{2+}	*YINYR*	a_1, a_2, y''_1–y''_4, y_3–NH_3	Y33–34	734.8^{2+}	*LINPVSDNYDRY*	a_1, a_2, b_2, b_3, b_5, y''_2–y''_{11}, $[y''_9+H]^{2+}$, $[y''_{10}+H]^{2+}$
T132	419.2^{2+}	D*LYIGEK*	a_2, b_2, b_3, y''_1–y''_5				
T143	738.6^{3+}	*LFLAPI*SDSDE*FYNTIQIK*	b_2–b_4, $[b_{12}+H]^{2+}$, $[b_{15}+H]^{2+}$–$[b_{18}+H]^{2+}$, y''_1–y''_8, $[y''_{15}+H]^{2+}$, $[M+3H–H_2O]^{3+}$				
T153	632.3^{2+}	*LGZNWQFIPK*	b_2, b_7, b_8, b_7–NH_3, b_8–NH_3, y''_1–y''_9, y''_5–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] Amino acid character codes as in Table 1.

[d] Italic characters indicate residues and position confirmed by sequence peaks.

[e] According to common Roepstorff–Fohlman nomenclature [31].

[f] Sequence alignments of digest peptides refer to BTxB1 neurotoxin sequence from *C. botulinum* strain Danish (NCBInr access code M81186), NTNH sequence from strain ATCC 43757 (NCBInr access code Y13630), and HA-33 sequence of strain NCTC 7273 (access X79102), whereas HA-17 and HA-70 are represented in Figs. 5 and 6.

[g] See text.

[h] N-terminal of peptide not conform trypsin or chymotrypsin cleavage of any known HA-33 sequence.

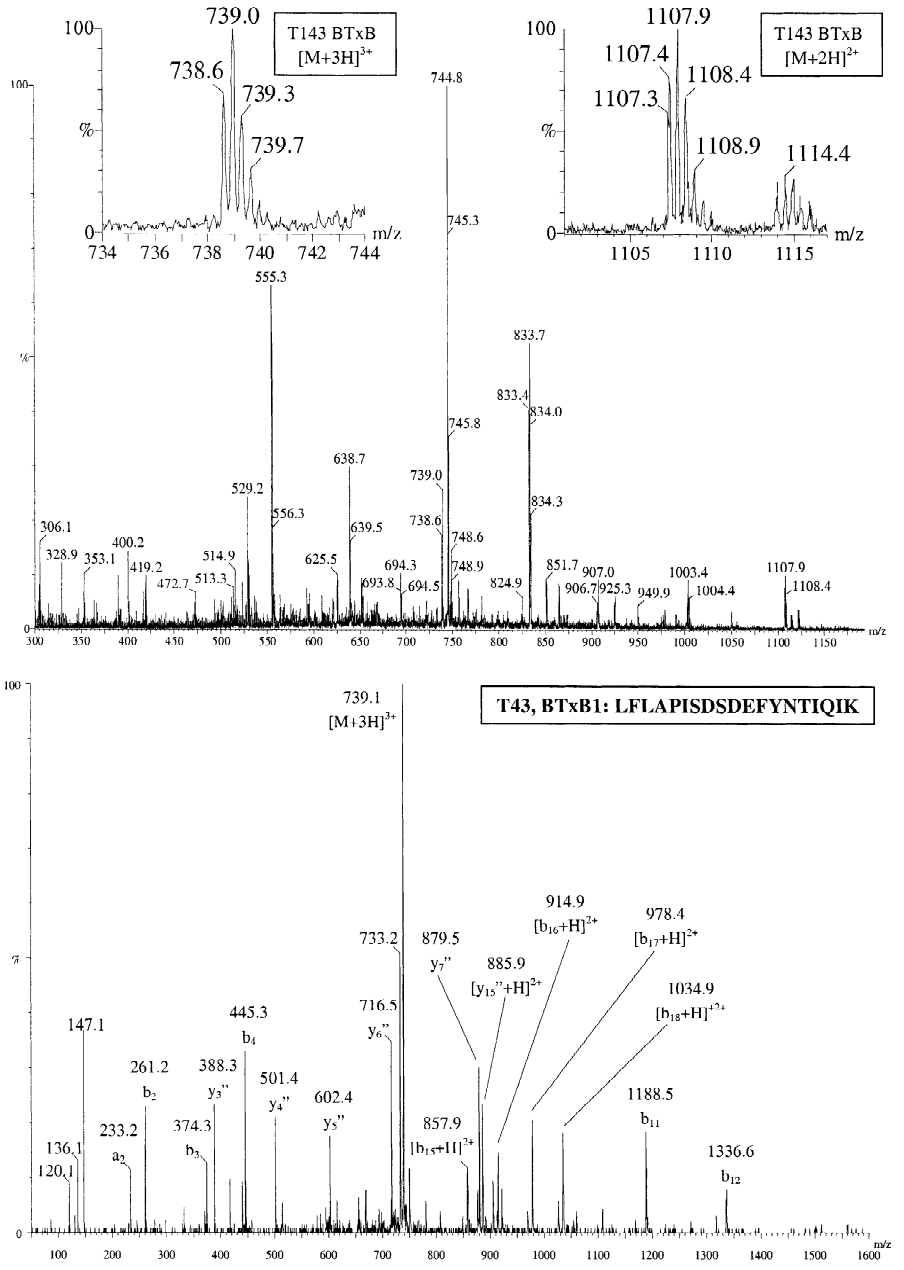

Fig. 4. ES-MS mass spectrum of components eluting at 34.61 min (top), with relevant ions of T143 of BTxB enlarged (inserts) and MS–MS product ion spectrum of the 738.6^{3+} precursor ions (bottom); the MS–MS spectrum is only partly annotated, for clarity.

with the BTxB1 sequence (from strain Danish). It is noted that chymotryptic cleavage of trypsin digest peptides must have occurred to explain the occurrence of two of the observed BTxB peptides. Since other toxin related peptides might have sequences deviating from the known BTxB sequences, all other observed sequences were subjected to BLAST search, as described above for BTxA. However, none

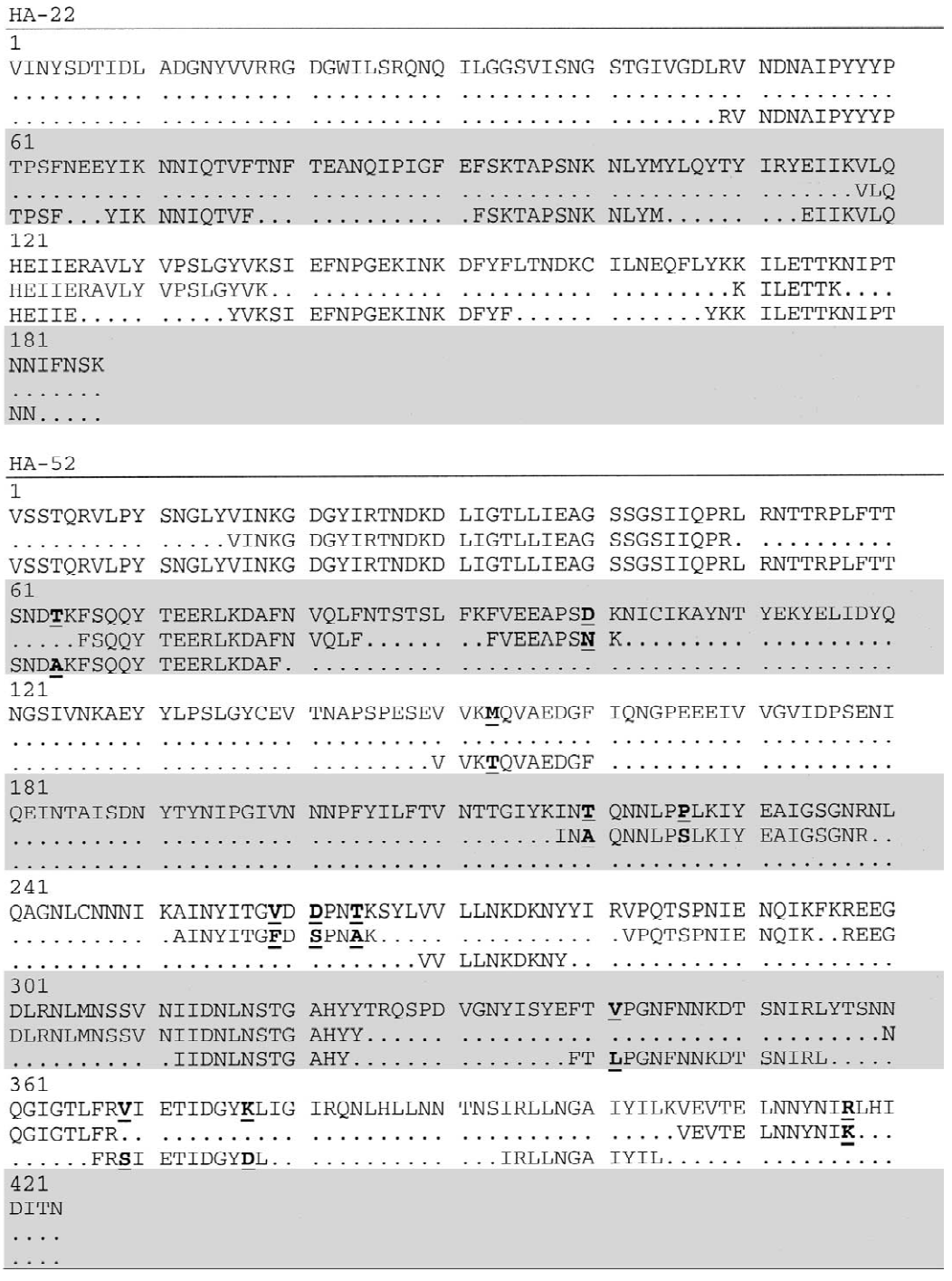

Fig. 5. Sequence alignment of HA-70 sequence with the known NCTC 2916 sequence [29] adjusted to strain okra by inclusion of known post-translational modification and N-terminal sequence differences [20] (top rows) and observed trypsin (middle rows) and pepsin (bottom rows) digest peptide sequences; sequence differences in bold underlined.

of these sequences were attributed to any BTxB. Rather, all other peptides partly or entirely matched either NTNH or HA type proteins associated with known BTxB gene cluster sequences.

Of the remainder of peptides, 14 belonged to NTNH. All 14 sequences agreed completely with the NTNH sequences known from the corresponding genes of *C. botulinum* type B strain ATCC 43757 (CDC 3281; Y13630 [30]) and strain Lamanna (U63808 [41]), assuming that one of the peptides is a chymotrypsin digest fragment of a trypsin digest fragment. The ATCC 43757 and Lamanna NTNH protein sequences differ by only five amino acids (R361K, C761Y, M783I, Q1089E, R1152M) and cannot be distinguished by the peptides observed.

Another 17 peptides match the HA-70 sequence and belong to either of the HA-22 or HA-52 parts that result from post-translational modification of HA-70. Some of the observed HA-70 peptides display mutations with respect to the best matching known sequence, that of *C. botulinum* strain NCTC 2916 (L42537 [34]). Strain NCTC 2916 is a type A strain with a silent B gene cluster [37]; there has been some controversy about the typing of the particular HA-70 [34,37], but it is now clear that the particular HA-70 (L42537) sequence comes from the type B gene cluster [37]. A comparison of the observed sequence and the NCTC 2916 sequence is given in Fig. 5; in this alignment, the above mentioned strain okra post-translational cleavage of HA-70 into HA-22 and HA-52 has been accounted for. Some details are discussed in Section 3.2.4.

A further 13 peptides belong to HA-33. From these peptides, 12 observed sequences agreed completely with that of HA-33 sequences known from HA-33 of *C. botulinum* type B strain NCTC 7273 (X79102 [34]) and strain Lamanna (U24431 [41]), provided that four of these peptides result from chymotrypsin cleavage of trypsin digest fragments. Three of these mixed digest peptides, LINPVSDNYDR, WLINPVSDNYDR, and LINPVSDNYDRY have a pairwise sequence overlap and the C- or N-termini demonstrate that chymotrypsin activity is indeed involved. The thirteenth peptide, LYADTVAR, did not match a trypsin or chymotrypsin digest fragment, but it does match part of the sequence (of T11) of the above mentioned two HA-33 proteins and of six other type B or A related HA-33. The peptide mismatch might be the result of an amino acid residue difference to either of the two known HA-33 sequences, for instance V119K or V119R and trypsin cleavage, but fitting anomalous peptides reflecting such mutations were not found. It is noted that a mutation from V to another amino acid residue would require an unlikely selective two-nucleotide mutation at the gene level. Therefore, the appearance of the LYADTVAR peptide cannot be adequately explained. The strain NCTC 7273 and Lamanna HA-33 sequences differ only slightly (Q87L, deletion of N234, 5 or 3 C-terminal residues) and cannot be distinguished by the trypsin digest peptides observed.

Another five peptides belong to HA-17, which all match three known HA-17 sequences from *C. botulinum* type B, strain ATCC 43757 (CDC 3281; Y13630 [30]), strain Lamanna (U24431 [41]), and strain NCTC 2916 (L42537 [34]). The NCTC 2916 and ATCC 43757 HA-17 amino acid sequences are identical, whereas the difference of these two with the Lamanna sequence comprises only two amino acids (V3A and K30G). A sequence alignment of the observed peptides with the NCTC 2916 sequence is given in Fig. 6.

The observed peptides do not allow a distinction between HA-17 from strain Lamanna and that of NCTC 2916 (identical sequence to that of ATCC 43757). The T2 sequence is in agreement with the N-terminal sequence reported for strain okra[5] . An MS–MS spectrum showed the presence of TFLPDGNYNIK, where a single asparagine was deamidated. In this case, the observation of the common peptide, TFLPNGNYNIK, proved that the peptide sequence is in line with the known HA-17 sequences and that the N/D change was due to deamidation during sample storage or treatment.

The observation of trypsin digest peptides and their sequences gives only a limited coverage. Because toxin, NTNH and HA factors from *C. botulinum* strain okra have not been fully covered by other sequencing methods, attempts were made to gain more information from pepsin digest peptides. The alignment of sequences from both peptide digests and known protein sequences provides an

[5]The strain okra HA-17 N-terminal protein sequence is: SVERTFLPNGNYNIKSIFS [18,19].

```
HA-17
1
MSVERTFLPN GNYNIKSIFS GSLYLNPVSK SLTFSNESSA NNQKWNVEYM AENRCFKISN
.....TFLPD GNYNIK.... .......... .......... ....WNVEYM AENR...ISN
.......... .......... .......... .......... .......... .....FKISN
61
VAEPNKYLSY DNFGFISLDS LSNRCYWFPI KIAVNTYIML SLNKVNELDY AWDIYDTNEN
VAEPNKYLSY DNFGFISLDS LSNRCYWFPI K......... .......... ..........
VAEPNKYLSY DNFGF..... .....YWFPI KIAVNT...L SLNKVNELD. .........N
121
ILSQPLLLLP NFDIYNSNQM FKLEKI
.......... .......... ......
ILSQPLLLLP NF........ FKLEKI
```

Fig. 6. Sequence alignment of HA-17 sequence from known NCTC 2916 [29], (top rows) and observed trypsin digest (middle rows) and pepsin digest peptide sequences (bottom rows).

approximation to de novo sequencing of the proteins from strain okra.

3.2.4. LC–ES-MS(–MS) of pepsin digested material

Pepsin digestion has the advantages that reduction and derivatisation are not required for adequate enzyme digest action and that digestion is sufficiently progressive within one h. However, pepsin has the disadvantages that it has many cleavage sites and does generally not give basic amino acids at the C-terminal. Of the observed cleavages, approximately 60% occur C-terminal to F, L, M, or W, whereas the remainder occurs mainly C-terminal to E, D and Y. This diversity of cleavage sites does not allow easy prediction of cleavage peptides from known proteins. Where terminal basic amino acids generally force peptide $[M+H]^+$ ion MS–MS fragmentation along b_n or y''_n series ions, the lack of such terminal residues complicates MS–MS product ion spectrum interpretation. In addition, interpretation cannot be guided by the typical y''_1 signal of K or R of a trypsin digest peptide. Notwithstanding this balance between advantages and disadvantages, we attempted pepsin digestion as an additional or alternative route to sequence information.

Fig. 7 gives a typical pepsin digest peptide product ion spectrum, whereas Table 5 gives a summary of results, in reference to known protein sequences. Many pepsin digest cleavages did not occur at the C-terminal side of F, L, M or W (the known main cleavage sites of pepsin); these cleavages have been indicated by ''/''. In addition, not a single peptide of HA-33 was found, although this protein was readily observed in trypsin digests of the BTxB preparation. It has been postulated that the non-neurotoxin proteins in a BTxB preparation play a role in neurotoxin transport or neurotoxin protection in the digestive tract, although the precise functions of these proteins remain elusive (see, for example, Refs. [42,43]). Therefore, it may be speculated that HA-33 is not observed because it interferes somehow with pepsin digestion. However, confirmatory experiments with purified HA-33 and pepsin digestion were outside the scope of the present investigation. In general, the information obtained from pepsin digest peptides was as good as that obtained from trypsin digest peptides.

Pepsin digest results confirm the trypsin digest data with respect to similarity of the type B strain okra neurotoxin to that from type B, strain Danish (BTxB1). This confirmation also holds for HA-17 (see Fig. 6) and NTNH. In contrast, pepsin digest peptides show an even more extensive deviation of strain okra HA-70 from any known HA-70 sequences than trypsin digest results (Tables 4 and 5). Firstly, the N-terminal found for the HA-52 post-translationally formed from HA-70, VSSTQ . . . , is at variance with literature data [20] for an amino acid sequenced HA-52 from strain okra, VFVTQ . . . . However, the presently observed N-terminal sequence matches that of other known type B HA-70 (HA-52), for example of strain NCTC 2916 [29,38]. The observed trypsin digest peptides FVEEAPSNK

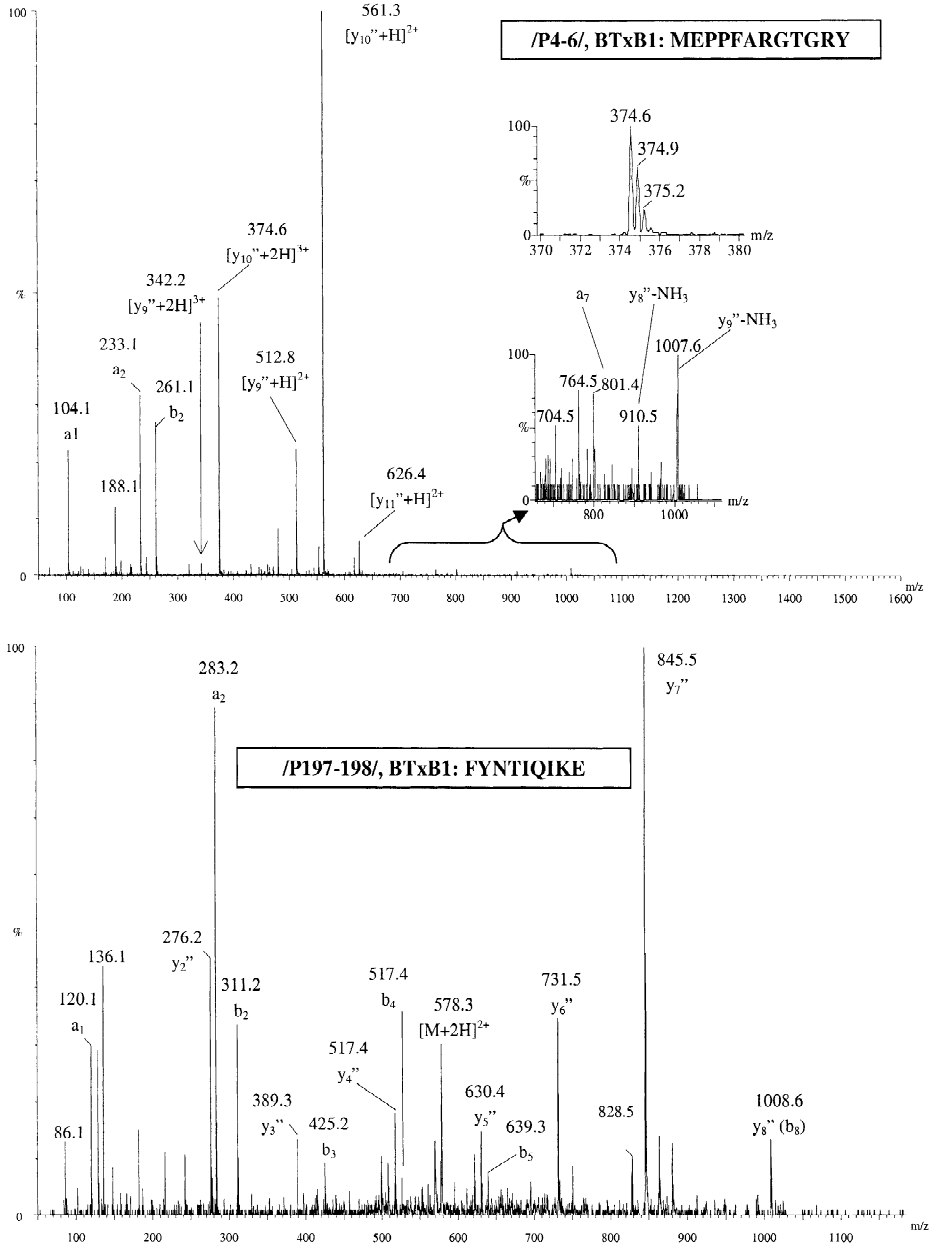

Fig. 7. Annotated product ion mass spectra of two pepsin digest fragments: triply protonated MEPPFARGTGRY with insert showing triply charged fragments (top) and doubly protonated FYNTIQIKE (bottom).

and INAQNNLPSLK do occur in the partially known HA-70 sequence from type B strains ATCC 25765 (Eklund 17B [44]) and ATCC 43757 [30], but other amino acid residue differences either do not match known HA-70 sequences or lie outside known (partial) sequences.

The two amino acid differences matching other known HA-70 each map to a single nucleotide in the HA-70 gene sequence (codon 302 cta→tta and codon 422 tga→cga, respectively). Of the other 11 differences, eight map to a single nucleotide difference, provided that the I/L indistinguishable to present

Table 5
Summary of sequence information obtained from botulinum toxin type B strain okra, by pepsin digest[c]

Digest fragment[a]	m/z_{obsd} (Da)[b]	Peptide sequence[c,d]	Sequence ions[e]	Digest fragment[a]	m/z_{obsd} (Da)[b]	Peptide sequence[c,d]	Sequence ions[e]
BTxB1[f]				NTNH[f]			
P4–6/	461.3^{3+}	*MEPP*FARGTGRY	a_1, a_2, a_7, b_2, $y_8''-NH_3$, $y_9''-NH_3$, $[y_9''+H]^{2+}-[y_{11}''+H]^{2+}$, $[y_9''+2H]^{3+}$, $[y_{10}''+2H]^{3+}$	P3	640.3^{4+}	*SINSPVDNK*NVVVVRARKTDTV*F*	a_2, b_2, y_1'', y_2'', $[y_{15}''+H]^{2+}-[y_{19}''+H]^{2+}$, $[y_{19}''+2H]^{3+}-[y_{21}''+2H]^{3+}$
/P11–12	675.1^{3+}	*YYDPDYLNTNDKKNIF*	a_1, a_2, b_2, b_3, $y_2''-y_5''$, $y_7''-y_{10}''$, $y_8''-NH_3$, $y_8''-NH_3$, $[y_{10}''+H]^{2+}-[y_{15}''+H]^{2+}$	P6/	585.9^{2+}	*VAPERYYGES*	a_2, b_2, $y_6''-y_8''$, $[y_8''+H]^{2+}$, PERYYGE, PERYYG, PERYY
/P20–23/	695.8^{3+}	*MIINGIP*YLGDRRVPLEE	a_1, a_2, b_2-b_6, $[y_{12}''+H]^{2+}-[y_{17}''+H]^{2+}$, $[y_{16}''+2H]^{3+}$	P8/	503.7^{3+}	*LSQDS*EKDKF*LQA*	a_1, a_2, b_2, y_2'', y_3'', $[y_9+H]^{2+}-[y_{12}+H]^{2+}$
/P20–23	744.7^{3+}	*MIINGIPYLGDRRV*PLEEF	a_2, b_2-b_6, y_8'', $[y_{13}''+H]^{2+}-[y_{18}''+H]^{2+}$, $[y_{17}''+2H]^{3+}$, PYLGD, RVP	P12–13	481.9^{3+}	LKRINS*TNAGEKL*	b_2, b_5, b_3-H_2O, $[b_7+H]^{2+}-[b_{12}+H]^{2+}$, $y_1''-y_8''$
P21–23	701.0^{3+}	*IINGIPYLGD*RRVPLEEF	a_1, a_2, b_2-b_5, y_5'', y_8'', $[y_{13}''+H]^{2+}-[y_{17}''+H]^{2+}$, PYLGD	/P22–23	513.8^{2+}	*LSSEDNKSF*	a_2, b_2, b_3-H_2O, b_8, y_1'', y_2'', $y_5''-y_8''$, SEDNKS/SSEDNK
P25	525.2^{3+}	*ISNP*GEVERKKGIF	a_1, a_2, b_2, b_3, b_3-H_2O, $[y_{11}''+H]^{2+}-[y_{13}''+H]^{2+}$, $[y_{11}''+2H]^{3+}$, $[y_{12}''+2H]^{3+}$	/P36–P38/	577.9^{2+}	*FLYGIKPSDD*	a_1, a_2, b_2-b_6, b_9, y_1'', y_2'', $y_4''-y_9''$, YGI, LYGI
P27–29/	578.4^{2+}	*IIFGPGPVLNE*	a_1, a_2, b_2, b_6, b_8-b_{10}, $y_7''-y_9''$, FGPGPV, FGPGPVL	P38/	447.7^{2+}	*YGIKPSDD*	a_1, b_2, b_3, b_6, b_7, $y_1''-y_7''$
/P29–32	698.3^{3+}	*IGIQNHFASREGFGGIMQM*	$[b_{13}+H]^{2+}-[b_{18}+H]^{2+}$, $[b_{18}+2H]^{3+}$, $y_1''-y_3''$, $y_2''-H_2O$, $y_3''-NH_3$	/P38–40	453.7^{3+}	*LVIP*YRLRSEL	a_2, b_2, b_3, $[y_8''+H]^{2+}-[y_{10}+H]^{2+}$
/P34–37	758.2^{4+}	*VSVFNNV*QENKGASIFNRRGYFSDPAL	a_1, a_2, a_4, b_2-b_4, b_3-H_2O, $[y_{21}''+2H]^{3+}-[y_{26}''+2H]^{3+}$, $[y_{25}''+3H]^{4+}$	P42	573.4^{+}	*NIVDL*	a_2, b_2-b_4, $y_1''-y_3''$
P38–43	607.8^{4+}	*ILMHELIHVLHGLYGIKVDDL*	a_1, a_2, b_2, b_8, b_{10}, $y_5''-y_{10}''$	P43–44	516.9^{2+}	*LVSGGIDPKF*	a_1, a_2, b_2, b_3, b_7, b_7-H_2O, $y_1''-y_4''$, $y_6''-y_9''$, SGGID
/P47	461.3^{+}	*AEEL*	a_2, b_2, b_3, y_2'', y_3''	/P42–44	573.7^{2+}	*LLVSGGIDPKF*	a_1, a_2, b_2, b_3, $y_1''-y_{10}''$, $[y_9''+H]^{2+}$
P48–50	1012.2^{3+}	*YTFGGQDPSIITP*STDKSIYDKVLQNF	a_2, b_2, b_7, b_3-H_2O, b_6-H_2O, $[y_{15}''+H]^{2+}$, $[y_{16}''+H]^{2+}$, $[y_{19}''+H]^{2+}-[y_{25}''+H]^{2+}$, $[y_{25}''+2H]^{3+}$	/P47–49/	540.7^{4+}	*YFS*NAKKV*FEDHR*NIYE	a_1, a_2, b_2, y_5'', y_6'', $[y_6''+H]^{2+}$, $[y_7''+H]^{2+}$, $[y_9''+H]^{2+}$, $[y_{15}''+2H]^{3+}$, $[y_{16}''+2H]^{3+}$
P48–51/	812.4^{4+}	*YTFGGQDPSIITP*STDKSIYDKVLQNFRG	a_1, a_2, b_2, $[y_{17}''+H]^{2+}-[y_{19}''+H]^{2+}$, $[y_{22}''+2H]^{3+}-[y_{27}''+2H]^{3+}$, $[y_{27}''+3H]^{4+}$	/P63–65/	624.3^{3+}	*LSL*SDRN*QDIINKPEE*	a_1, a_2, b_2, b_3, b_3-H_2O, $[b_8+H]^{2+}-[b_{11}+H]^{2+}$, $y_1''-y_8''$, $[y_{13}''+H]^{2+}-[y_{15}''+H]^{2+}$, SLSNQDRI, SLSDRNQDII, SDRNQDIINK
/P60–61/	537.9^{4+}	*FTETNIAENYKIK*TRASY	a_1, a_2, b_2, b_3-H_2O, b_5, b_6, $y_5''-y_8''$, $[y_{10}''+H]^{2+}-[y_{12}''+H]^{2+}$, $[y_{16}''+2H]^{3+}$, $[y_{17}''+2H]^{3+}$	P64–65/	586.6^{3+}	*SL*SDRN*QDIINKPEE*	a_2, b_2, b_8, b_9, $[b_7+H]^{2+}-[b_{10}+H]^{2+}$, $y_1''-y_8''$, $[y_{12}''+H]^{2+}$, $[y_{13}''+H]^{2+}$
/P61–P65/	643.7^{3+}	*FSDSLPPVKIKNLLDNE*	a_1, a_2, b_2-b_5, y_{11}'', y_{12}'', $[y_{12}''+H]^{2+}-[y_{16}''+H]^{2+}$, LPPVKIKNLLD, PPVKIKNLLDN	P71/	463.6^{3+}	*YSNYKIPYN*RA	a_1, a_2, b_2-b_6, $y_3''-y_7''$, $[y_5''+H]^{2+}$, $[y_7''+H]^{2+}-[y_{10}''+H]^{2+}$
P78–80	961.9^{2+}	*ISKIELPSENTESLTDF*	a_1, a_2, b_2-b_6, $b_{12}-b_{14}$, $[b_{15}+H]^{2+}$, $[b_{16}+H]^{2+}$, $y_1''-y_3''$, $y_{11}''-y_{13}''$, SKIELPSEN, PSENTESLTD	/P78–P80	612.2^{3+}	*FVEVVSSKDKSLVYSF*	a_2, b_2-b_5, $[b_{15}+H]^{2+}$, $[y_{13}''+H]^{2+}-[y_{15}''+H]^{2+}$, SKDKSLVYS/SSKDKSLVY, SSKDKSLVYS, VSSKDKSLVY, EVVSSKDKSLV/VEVSSKDKSL, VSSKDKSLVYS
P81–82/	584.0^{4+}	*NVDVPVYEKQPAIK*KIFT*DE*	a_2, a_2-NH_3, b_2, y_1'', y_2'', $y_7''-y_{10}''$, $[y_{14}''+H]^{2+}-[y_{16}''+H]^{2+}$, $[y_{16}''+2H]^{3+}-[y_{19}''+2H]^{3+}$	P98–99	538.7^{2+}	*INKKENLSM*	a_1, b_2, b_5-b_8, b_3-NH_3, $y_1''-y_4''$, $y_6''-y_8''$, $[y_8''+H]^{2+}$, NKKENLS, KKENLS, KENLS
P81–82	703.0^{4+}	N*VDVP*VYEKQPAIKKIFT*DENTIF*	a_2, a_4, b_2-b_5, $[b_{20}+2H]^{3+}-[b_{23}+2H]^{3+}$, $[b_{22}+3H]^{4+}$, y_1'', y_2'', $[y_{22}''+2H]^{3+}$	P106–108	569.3^{2+}	*FNIYLKNIL*	a_1, a_2, b_2-b_8, $y_1''-y_8''$
/P90–91	496.3^{2+}	*FSNKVYSF*	a_2, b_2, b_4-b_7, $y_1''-y_3''$, $y_5''-y_7''$	/P117–118	395.7^{2+}	*MAKQSIL*	a_1, a_2, a_5-H_2O, $[a_6+H]^{2+}$, $[a_6+H-H_2O]^{2+}$, b_2, b_5, b_6, b_5-H_2O, b_6-NH_3, y_1'', y_5'', y_6'', $y_6''-NH_3$
/P94	469.2^{3+}	*YIKTANKVVEAGL*	a_1, a_2, b_2, b_3, $[b_8+H]^{2+}-[b_{12}+H]^{2+}$, $y_1''-y_4''$, $[y_{11}''+H]^{2+}$, TANKVVEA, TANKVVEAG, KTANKVVEAG	P137/	525.2^{2+}	*IQKCTNITE*	a_1, b_2, b_6-b_8, $[b_8+H]^{2+}$, y_1'', y_2'', y_4'', y_7'', y_8'', KCTNI, KCTNIT
/P96–97/	501.2^{2+}	*WVKQIVND*	a_2, b_2, b_3, b_4-b_7, b_5-NH_3, $y_1''-y_3''$, y_6'', y_7'', VKQI/KQIV	P144–145/	483.2^{2+}	*FDIQSIKD*	a_2, b_2-b_6, $b_5''-H_2O$, $b_6''-H_3O$, $y_1''-y_7''$
/P97–99/	599.3^{3+}	*FVIEANKSNTMDKIAD*	a_1, a_2, b_2-b_4, y_1'', y_2'', $y_4''-y_9''$, $[y_{11}''+H]^{2+}-[y_{15}''+H]^{2+}$, $[y_{14}''+2H]^{3+}$	P144–146/	646.9^{2+}	*FDIQSIKDLIT*	a_1, a_2, b_2-b_{10}, $y_5''-y_8''$, IKDL/KDLI, IKD/KDL
/P122	663.3^{4+}	IIKYRYNIYSEKEKSN*INIDF*	a_2, b_2, $[b_{17}+H]^{2+}$, $[b_{18}+2H]^{3+}-[b_{20}+2H]^{3+}$, $[b_{20}+3H]^{4+}$, $y_1''-y_3''$	P144–146/	646.9^{2+}	*FDIQSIKDLITSE*	$a_1(y_1'')$, $a_2(y_2'')$, b_2-b_{11}, y_3'', y_4'', $y_7''-y_{10}''$, IKDL/KDLI, IKD/KDL
P130–135/	570.9^{3+}	*LDFDN*TLKKNLL*NY*	a_1, a_2, b_2, b_4, y_1'', y_2'', y_7'', $[y_{10}''+H]^{2+}-[y_{13}''+H]^{2+}$	P150–152	759.9^{3+}	*FTLQEDNNKVIEDISGKNTL*	a_1, a_2, b_2-b_5, b_9-b_{13}, b_3-H_2O, b_6-H_2O, y_1'', y_2'', $y_4''-y_{11}''$, $[y_{14}''+H]^{2+}-[y_{18}''+H]^{2+}$
/P135	366.3^{2+}	*IDENKL*	a_1, a_2, b_2, $y_1''-y_5''$	P155/	490.8^{2+}	*YLKEPDES*	a_1, a_2, b_2-b_3, b_6, b_7, y_1'', $y_2''-H_2O$, y_4'', y_6'', $y_6''-H_2O$, y_7'', PDE, LKEPD, KEPDE
P143	678.4^{2+}	*FNKYNSEILNN*	a_1, b_2-b_{10}, b_7-NH_3, b_8-NH_3, b_9-NH_3, $y_1''-y_4''$, y_3-NH_3, KYNSEI	P155–156	657.3^{2+}	*YLKEPDESVSF*	a_2, b_2-b_4, b_6-b_{10}, $y_1''-y_5''$, $y_7''-y_9''$
P147	461.7^{2+}	*RYKDNNL*	b_2-b_6, b_4-H_2O, b_5-NH_3, b_6-H_2O, y_1'', y_2'', $y_2''-H_2O$	P161–164/	501.7^{2+}	*WLRNGED*	a_1, a_2, b_1-b_7, b_1-NH_3, $[b_7+H]^{2+}$, $y_1''-y_3''$, y_6'', y_7'', $y_2''-H_2O$, $y_3''-H_2O$
/P162	554.7^{2+}	*DINGKTKSVF*	a_1, a_2, b_2, b_6-b_9, y_1'', y_2'', y_5'', y_7'', y_8'', $[y_8''+H]^{2+}$, $[y_9''+H]^{2+}$	P192–195/	788.6^{5+}	*LGDEEKYVDISSE*NNRIQLVN-NSKDTAKRIIFNND	a_1, a_2, b_2-b_{11}, $[b_9+H]^{2+}$, $[b_{10}+H]^{2+}$, $[y_{22}''+2H]^{3+}-[y_{26}''+2H]^{3+}$, $[y_{33}''+3H]^{4+}$
P163–164/	485.8^{2+}	*FEYNIRE*	a_1, a_2, b_2-b_4, $y_1''-y_6''$				
/P164–166	523.4^{2+}	*YINRWFF*	a_1, a_2, a_5, a_6, $[a_5+H]^{2+}$, $[a_6+H]^{2+}$, b_2, b_4-b_6, y_1'', y_2'', y_5'', y_6''				
P167–169/	727.1^{4+}	*VTITNNL*NNAKIYINGKL*ESNTDI*KD	a_2, b_2-b_4, b_3-H_2O, $y_2''-y_4''$, y_9'', y_{12}'', $[y_{21}''+2H]^{3+}-[y_{25}''+2H]^{3+}$, $[y_{24}''+3H]^{4+}$				
/P177–181	692.5^{2+}	*YLKDFWGNPLM*	a_1, a_2, $[a_{10}+H]^{2+}$, b_2, b_5-b_9, $y_1''-y_6''$, y_9'', y_{10}'', KDFWGN				
/P197–198/	578.3^{2+}	*FYNTIQIKE*	a_1, a_2, b_2-b_5, $y_1''-y_8''$				

[a] P for pepsin cleavage site, a slash indicates a C- or N-terminal cleavage at a less common pepsin cleavage site (that is: not directly at the C-terminal side of an F, L, M or W residue).
[b] Observed m/z ratio, with apparent monoisotopic mass in Da.
[c] Amino acid character codes as in Table 1.
[d] Italic characters indicate residues and position confirmed by sequence peaks.
[e] According to common Roepstorff–Fohlman nomenclature [31].
[f] Sequence alignments of digest peptides refer to BTxB1 neurotoxin sequence from *C. botulinum* strain Danish (NCBInr access code M81186), and NTNH sequence from strain ATCC 43757 (NCBInr access code Y13630), whereas HA-17 and HA-70 are represented in Figs. 5 and 6.

mass spectrometry experiments in/P63–65 (Table 5) is L. In any case, most differences involve unlike amino acids (T→ A, M→T, P→S, V→F, V→S, and K→D). The detail of information thus obtained shows that the HA-70 protein, particularly the HA-52 mature protein, seems to be the only protein in the *C. botulinum* type B strain okra that is widely differing from other known type B HA-52 (HA-70) proteins. Therefore, this protein would be a candidate for more elaborate strain typing, for example by gel electrophoresis and subsequent enzymatic digestion and MALDI-MS.

4. Conclusions

It was shown that an identification method developed earlier for tetanus toxin can straightforwardly be applied to botulinum toxin. This finding underpins the thesis that all toxins can be identified unequivocally y mass spectrometry and is a step forward in a generic procedure applicable to all protein toxins. In addition, pepsin digestion was shown to be an alternative to trypsin digestion when LC–ES-MS–MS amino acid sequence analysis is performed. Pepsin digestion is not a good alternative in case of MALDI-MS peptide mapping, because enzymatic cleavages with pepsin are much less predictable than with trypsin.

The trypsin digest peptide map obtained from MALDI MS experiments allowed straightforward identification of a botulinum neurotoxin component from the known *C. botulinum* strain 62 type A toxin. In contrast, the unknown strain okra type B neurotoxin did not allow identification, because the peptide map was more complex, due to the presence of neurotoxin associated proteins (NTNH and HA). In this case, MALDI-PSD sequence elucidation of some peptides was required to attribute the sample to *C. botulinum*. From the point of view of identification of an unknown toxin, the MALDI peptide maps with limited MALDI-PSD provide a means of ''provisional identification'', down to the type A–G level. The amino acid sequence information obtained from subsequent LC–ES-MS–MS experiments does allow unambiguous identification of the neurotoxin. This was demonstrated for a type A neurotoxin sequence for which DNA sequences are available. Even with a type B sequence for which no DNA sequence is available, identification of the neurotoxin did not pose a problem, given the fact that DNA sequences of other type B neurotoxins are known.

The presence of NTNH and HA proteins along the botulinum neurotoxins complicates identification, although in the end evidence from NTNH and HA sequences corroborates that identification. Actually, the full set of proteins gives information on the *C. botulinum* source, because it allows a certain degree of strain typing. In the case of the hitherto unknown sequences of type B, strain okra, the biovariation in HA-70, particularly in the post-translationally modified HA-52 part, can be used as a strain indicator.

Acknowledgements

The authors thank the Netherlands MOD 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] Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on Their Destruction, signed at Paris, 13 January 1993; entered into force on 29 April 1997.
[3] E.R.J. Wils, A.G. Hulst, Rapid Commun. Mass Spectrom. 7 (1993) 413.
[4] C.E. Kientz, A.G. Hulst, E.R.J. Wils, J. Chromatogr. A 757 (1997) 51.
[5] B.L.M. van Baar, A.G. Hulst, E.R.J. Wils, Toxicon 37 (1998) 85.
[6] B.L.M. van Baar, A.G. Hulst, B. Roberts, E.R.J. Wils, Anal. Biochem. 301 (2002) 278.
[7] U. Weller, M.-E. Dauzenroth, D. Meyer zu Heringdorf, E. Habermann, Eur. J. Biochem. 182 (1989) 649.
[8] S.S. Arnon, R. Schechter, T.V. Inglesby, D.A. Henderson, J.G. Bartlett, M.G. Ascher, E. Eitzen, A.D. Fine, J. Hauer, M. Layton, S. Lillibridge, M.T. Osterholm, T. O'Toole, G. Parker, T.M. Perl, P.K. Russel, D.L. Swerdlow, K. Tonat, J. Am. Med. Assoc. 285 (2001) 1059.
[9] R.H. Hill Jr., M.L. Hemphill, J.Y. Richmond, Chem. Health Safety 8 (2001) 30.
[10] R.A. Zilinskas, J. Am. Med. Assoc. 278 (1997) 418.
[11] S.S. Arnon, R. Schechter, T.V. Inglesby, D.A. Henderson, J.G. Bartlett, M.G. Ascher, E. Eitzen, A.D. Fine, J. Hauer,

M. Layton, S. Lillibridge, M.T. Osterholm, T. O'Toole, G. Parker, T.M. Perl, P.K. Russel, D.L. Swerdlow, K. Tonat, J. Am. Med. Assoc. 285 (2001) 1060.
[12] A latest version (September 2001) of the ''rolling text'' is accessible through the internet: http://www.bradford.ac.uk/acad/sbtwc/evaluation/evalu19.pdf.
[13] Australia Group document AG/Jun92/BW/Chair/12, 1993, accessible through the internet: http://www.australiagroup.net/control_list/bio_agents.htm.
[14] D.E. Thompson, R.A. Hutson, A.K. East, D. Allaway, M.D. Collins, P.T. Richardson, FEMS Microbiol. Lett. 108 (1993) 175.
[15] X. Meng, T. Karasawa, K. Zou, X. Kuang, X. Wang, C. Lu, C. Wang, K. Yamakawa, S. Nakamura, J. Clin. Microbiol. 35 (1997) 2160.
[16] H. Niemann, in: J.E. Alouf, J. Freer (Eds.), Sourcebook of Bacterial Protein Toxins, Academic Press, London, 1991, p. 303.
[17] T. Binz, H. Kurazono, M. Wille, J. Frevert, K. Wernars, H. Niemann, J. Biol. Chem. 265 (1990) 9153.
[18] J.J. Schmidt, V. Sathyamoorthy, B.R. DasGupta, Arch. Biochem. Biophys. 238 (1985) 544.
[19] B.R. DasGupta, A. Datta, Biochimie 70 (1988) 811.
[20] E. Somers, B.R. Dasgupta, J. Prot. Chem. 10 (1991) 375.
[21] R. Fujita, Y. Fujinaga, K. Inoue, H. Nakajima, H. Kumon, K. Oguma, FEBS Microbiol. Lett. 376 (1995) 41.
[22] A.B. Maksymowych, M. Reinhard, C.J. Malizio, M.C. Goodnough, E.A. Johnson, L.L. Simpson, Infect. Immun. 67 (1999) 4708.
[23] Accessible through the internet, at http://www.ncbi.nlm.nih.gov/.
[24] D.E. Thompson, J.K. Brehm, J.D. Oultram, T.J. Swinfield, C.C. Shone, T. Atkinson, J. Melling, N.P. Minton, Eur. J. Biochem. 189 (1990) 73.
[25] A. Willems, A.K. East, P.A. Lawson, M.D. Collins, Res. Microbiol. 144 (1993) 547.
[26] D.B. Lacy, W. Tepp, A.C. Cohen, B.R. DasGupta, R.C. Stevens, Nat. Struct. Biol. 5 (1998) 898.
[27] N.N. Nalivaeva, A.J. Turner, Proteomics 1 (2001) 735.
[28] W. Zhang, B.T. Chait, Anal. Chem. 72 (2000) 2482, (Search facility accessible at http://prowl.rockefeller.edu/ and at http://www.proteometrics.com/).
[29] I. Henderson, S.M. Whelan, T.O. Davis, N.P. Minton, FEMS Microbiol. Lett. 140 (1996) 151.
[30] J.A. Santos-Buelga, M.D. Collins, A.K. East, Curr. Microbiol. 37 (1998) 312.
[31] P. Roepstorff, J. Fohlman, Biomed. Mass Spectrom. 11 (1984) 601.
[32] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Nucleic Acids Res. 25 (1997) 3389.
[33] A.K. East, M. Bhandari, J.M. Stacey, K.D. Campbell, M.D. Collins, Int. J. Syst. Bacteriol. 46 (1996) 1105.
[34] A.K. East, J.M. Stacey, M.D. Collins, Syst. Appl. Microbiol. 17 (1994) 306.
[35] S.M. Whelan, M.J. Elmore, N.J. Bodsworth, J.K. Brehm, T. Atkinson, N.P. Minton, Appl. Environ. Microbiol. 58 (1992) 2345.
[36] R.A. Hutson, M.D. Collins, A.K. East, D.E. Thompson, Curr. Microbiol. 28 (1994) 101.
[37] R.A. Hutson, Y. Zhou, M.D. Collins, E.A. Johnson, C.L. Hatheway, H. Sugiyama, J. Biol. Chem. 271 (1996) 10786.
[38] M. Rodriguez Jovita, M.D. Collins, A.K. East, Curr. Microbiol. 36 (1998) 226.
[39] G.H. Yang, S.D. Rhee, H.H. Jung, K.H. Yang, Biochem. Mol. Biol. Int. 39 (1996) 1141.
[40] D.N. Perkins, D.J. Pappin, D.M. Creasy, J.S. Cottrell, Electrophoresis 20 (1999) 3551.
[41] G.H. Yang, unpublished (NCBI access number: U63808).
[42] F. Chen, G.M. Kuziemko, R.C. Stevens, Infect. Immun. 66 (1998) 2420.
[43] Y. Fujinaga, K. Inoue, T. Nomura, J. Sasaki, J.C. Marvaud, M.R. Popoff, S. Kozaki, K. Oguma, FEBS Lett. 467 (2000) 179.
[44] M. Bhandari, K.D. Campbell, M.D. Collins, A.K. East, Curr. Microbiol. 35 (1997) 207.