

ARTICLE

Received 13 Jul 2016 | Accepted 2 Dec 2016 | Published 3 Aug 2017

DOI: 10.1038/ncomms14130

OPEN

# Identification and characterization of a novel botulinum neurotoxin

Sicai Zhang<sup>1</sup>, Geoffrey Masuyer<sup>2</sup>, Jie Zhang<sup>1</sup>, Yi Shen<sup>1</sup>, Daniel Lundin<sup>2</sup>, Linda Henriksson<sup>2</sup>, Shin-Ichiro Miyashita<sup>1</sup>, Markel Martínez-Carranza<sup>2</sup>, Min Dong<sup>1</sup> & Pål Stenmark<sup>2</sup>

Botulinum neurotoxins are known to have seven serotypes (BoNT/A–G). Here we report a new BoNT serotype, tentatively named BoNT/X, which has the lowest sequence identity with other BoNTs and is not recognized by antisera against known BoNTs. Similar to BoNT/B/D/F/G, BoNT/X cleaves vesicle-associated membrane proteins (VAMP) 1, 2 and 3, but at a novel site (Arg66-Ala67 in VAMP2). Remarkably, BoNT/X is the only toxin that also cleaves non-canonical substrates VAMP4, VAMP5 and Ykt6. To validate its activity, a small amount of full-length BoNT/X was assembled by linking two non-toxic fragments using a transpeptidase (sortase). Assembled BoNT/X cleaves VAMP2 and VAMP4 in cultured neurons and causes flaccid paralysis in mice. Thus, BoNT/X is a novel BoNT with a unique substrate profile. Its discovery posts a challenge to develop effective countermeasures, provides a novel tool for studying intracellular membrane trafficking, and presents a new potential therapeutic toxin for modulating secretions in cells.

<sup>1</sup>Department of Urology, Boston Children's Hospital, Department of Microbiology and Immunobiology and Department of Surgery, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>2</sup>Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden. Correspondence and requests for materials should be addressed to M.D. (email: min.dong@childrens.harvard.edu) or to P.S. (email: stenmark@dbb.su.se).

**B**oNTs are a family of bacterial toxins, classified as one of the six most dangerous potential bioterrorism agents (Category A and Tier 1 select agent in the United States)<sup>1</sup>. They are also widely used to treat a growing list of medical conditions<sup>2,3</sup>, including muscle spasms, chronic pain, overactive bladders, as well as having cosmetic applications. There are seven well-established serotypes of BoNTs (BoNT/A–G), traditionally defined based on a lack of cross-neutralization by different antisera raised against each toxin type. All BoNTs share the same structure and function<sup>4–6</sup>. They are composed of a light chain (LC, ~50 kDa) and a heavy chain (HC, ~100 kDa) connected by an inter-chain disulfide bond. The HC contains two sub-domains: the C-terminal H<sub>C</sub> that mediates binding to receptors, and the N-terminal H<sub>N</sub> that mediates translocation of the LC across endosomal membranes. The LC acts as a protease in neurons to cleave a set of proteins: BoNT/A, C and E cleave at distinct sites on a peripheral membrane protein known as SNAP-25; BoNT/B, D, F and G cleave at different sites on homologous vesicle proteins VAMP1, 2 and 3 (vesicle-associated membrane proteins); and BoNT/C also cleaves the plasma membrane protein syntaxin 1. These proteins are prototypes of the SNARE (soluble NSF attachment protein receptor) protein family, whose members mediate various membrane fusion events in eukaryotic cells<sup>7,8</sup>. Cleavage of any one of the three neuronal SNARE proteins blocks fusion of synaptic vesicles to plasma membranes, thus preventing neurotransmitter release from neurons.

Recognizing all distinct serotypes of BoNTs is essential for developing effective countermeasures against this family of toxins. BoNT/A and BoNT/B were first identified in 1919 by Georgina Burke<sup>9</sup>. The last of the seven serotypes, BoNT/G, was discovered in 1969 (ref. 10), and no new BoNT serotype has been found for the past four decades. Recent progress in genomic sequencing has revealed multiple subtypes (designated with Arabic numbers, for example, BoNT/A1), which can be recognized by the same antiserum but contain substantial sequence variations<sup>11–13</sup>. Furthermore, there are also multiple mosaic toxins. For instance, a ‘type H’ was reported in 2013 but was later designated as a mosaic toxin, as its LC shares ~80% identity with the LC of a BoNT/F subtype, BoNT/F5, and its H<sub>C</sub> shares ~84% identity with the H<sub>C</sub> of BoNT/A1 (refs 14–17). Consistently, this toxin can be neutralized by antibodies against BoNT/A (ref. 16).

The genes encoding BoNTs can be on a plasmid, a phage, or the chromosome, indicating that these genes are mobile and capable of horizontal gene transfer<sup>18–23</sup>. Some strains contain two or even three different BoNT genes<sup>15,24,25</sup>. These strains are usually designated with a capital letter for the toxin type that is expressed at higher levels than the other one, followed with a lower case letter for the second toxin type (for example, BoNT/Af strain). In addition, it has also been reported that some BoNT/A strains contain a complete BoNT/B gene, but only BoNT/A is expressed<sup>26–28</sup>. Thus, the BoNT/B gene is considered a silent gene and the strains are known as BoNT/A(B) strains. A recent survey of infant botulism cases reported that ~8% isolates are BoNT/A(B) strains<sup>29</sup>.

Here we searched published genomic sequences and identified a novel BoNT gene encoded on the chromosome of *Clostridium botulinum* strain 111. This strain was originally identified from an infant botulism patient in Japan in 1996 (ref. 30). The initial characterizations indicated that the toxicity of this strain is due to BoNT/B<sup>30</sup>. Later studies confirmed that this strain expresses a subtype of BoNT/B (BoNT/B2) encoded on a plasmid<sup>31,32</sup>. The sequence of this novel BoNT gene was deposited into the GenBank database in February 2015, as a part of the genomic sequence of *C. botulinum* 111. We characterized the protein

encoded by this gene at functional levels and established it as a new BoNT serotype with a unique substrate profile.

## Results

### Searching genomic databases revealed a novel BoNT gene.

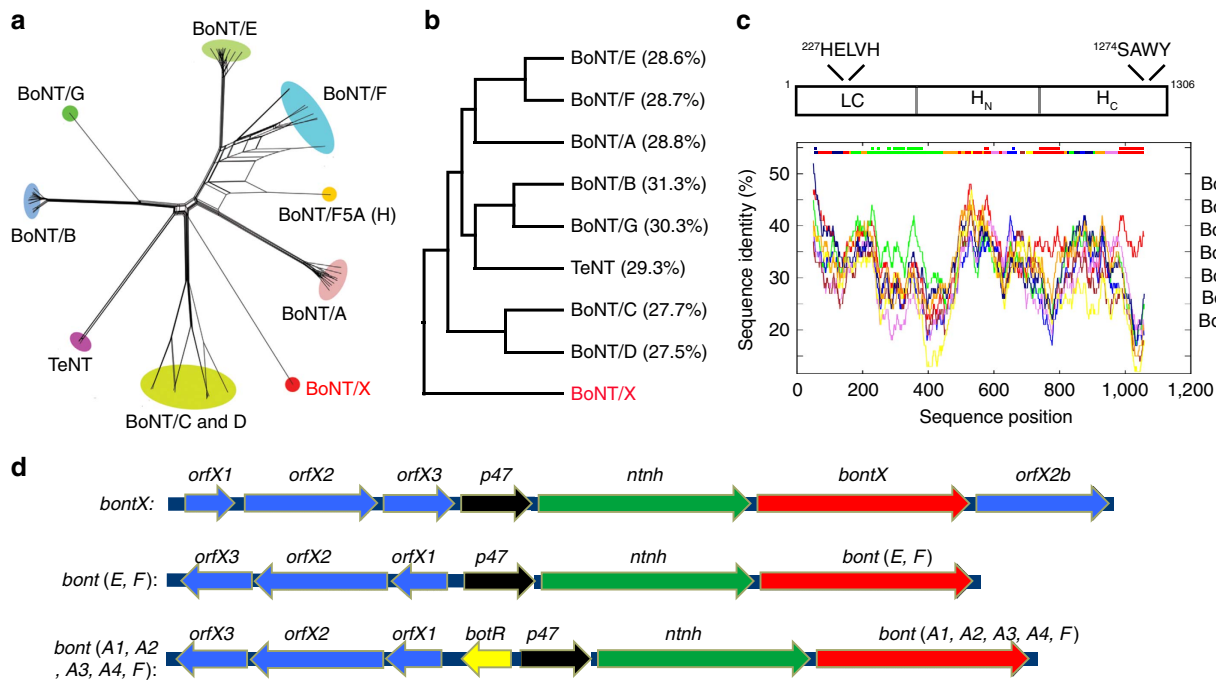
In an attempt to survey the evolutionary landscape of BoNTs, we performed iterative Hidden Markov model searches of the Uniprot sequence database. Our search identified all known BoNT subtypes and mosaic toxins, as well as the related tetanus neurotoxin (Fig. 1a; Supplementary Fig. 1). To our surprise, the search revealed a potentially new BoNT, tentatively designated BoNT/X (Fig. 1a, GenBank no.: BAQ12790.1), from the recently reported genomic sequence of *C. botulinum* strain 111. BoNT/X showed the least protein sequence identity with the other BoNTs in pairwise comparisons (Fig. 1b). Furthermore, the low sequence similarity is evenly distributed along the entire BoNT/X sequence (Fig. 1c), indicating that it is not a mosaic toxin. Despite this low sequence identity, the overall domain arrangement of BoNTs is conserved in BoNT/X (Fig. 1c), including a zinc-dependent protease motif HEXXH (residues 227–231, HELVH) in the LC (ref. 33), and a SXWY motif in the H<sub>C</sub> (residues 1,274–1,277, SAWY), which recognizes the lipid receptor gangliosides<sup>34</sup>.

Similar to the other BoNTs, the BoNT/X gene is located in a gene cluster<sup>23</sup>. All seven established BoNTs are co-expressed with another 150 kDa protein known as NTNHA (non-toxic non-hemagglutinin protein), which forms a pH-dependent complex with BoNTs and protects them from proteases in the gastrointestinal tract<sup>35</sup>. The BoNT/X gene is also preceded by a potential NTNHA gene (Fig. 1d). Besides BoNT and NTNHA, a typical BoNT gene cluster contains genes encoding one of the two types of accessory proteins: (1) the HA cluster encoding three conserved proteins HA17, HA33 and HA70, which form a complex with BoNT/NTNHA and facilitate absorption of toxins across the intestinal epithelial barrier<sup>36–38</sup>; or (2) the OrfX cluster encoding conserved OrfX1, OrfX2, OrfX3 and P47 proteins with unknown function<sup>23</sup>. The BoNT/X gene is located in an OrfX gene cluster, as are BoNT/E, F and members of BoNT/A. Interestingly, the BoNT/X cluster has two unique features (Fig. 1d): (1) there is an additional OrfX2 gene that does not exist in any other BoNT clusters (we designated it OrfX2b); (2) the reading frame of OrfX genes is usually opposite to BoNT/NTNHA genes, but it has the same direction as the BoNT/X gene in the BoNT/X cluster (Fig. 1d). These findings suggest that BoNT/X is a unique branch of the BoNT family.

### The LC of BoNT/X cleaves VAMP2 at a novel site.

To characterize BoNT/X, we first focused on its LC (X-LC, residues 1–439) and produced it as a His6-tagged protein in *Escherichia coli*. LCs of BoNT/A (A-LC) and BoNT/B (B-LC) were produced and assayed in parallel as controls. Incubation of X-LC with rat brain detergent extracts (BDE) did not affect syntaxin 1 or SNAP-25, but abolished VAMP2 immunoblot signals (Fig. 2a). LCs of BoNTs are zinc-dependent proteases<sup>33</sup>. As expected, EDTA prevented cleavage of SNARE proteins by X-, A- and B-LCs (Fig. 2a). Furthermore, incubation of X-LC with the purified recombinant cytosolic domain of VAMP2 (residues 1–93) converted VAMP2 into two lower-molecular-weight bands (Fig. 2b), confirming that X-LC cleaves VAMP2.

To identify the cleavage site, we analysed the VAMP2 (1–93) protein, with or without pre-incubation with X-LC, by liquid chromatography–tandem mass spectrometry (LC–MS/MS, Fig. 2c–e). A single dominant peptide peak appeared after incubation with X-LC (Fig. 2c,e; Supplementary Fig. 2). Its molecular weight is 3,081.7, which fits only the peptide sequence of A67–L93 of VAMP2 (Fig. 2c,e). Consistently, another fragment



**Figure 1 | Identification of BoNT/X.** (a) A phylogenetic split network covering all BoNT serotypes, subtypes, mosaic toxins and related tetanus neurotoxin (TeNT) illustrates their potential evolutionary relationships, as well as conflicts arising from e.g. chimerisms, based on their protein sequences. BoNT/X is highlighted in red. An enlarged version of this panel is shown in Supplementary Fig. 1, with the sequence access number for each toxin gene noted. (b) A phylogenetic tree of the protein sequence alignment for BoNT/A-G, TeNT and BoNT/X, analysed by the ClustalW method. The percentages of sequence identity between each toxin and BoNT/X are noted. (c) Upper panel: a schematic drawing of the three domains of BoNT/X, with conserved protease motif in the LC and the ganglioside binding motif in the H<sub>C</sub> noted. Lower panel: analysis using a sliding sequence comparison window demonstrated that the low similarity between BoNT/X and other BoNTs/TeNT is evenly distributed along the entire BoNT/X sequence. The X axis represents the query sequence position at the center of a 100-amino-acid moving sequence-comparison window. The Y axis shows the percentage of identity between that sequence window and each of the aligned background sequences. The two bars at the top of the graph illustrate the best matching sequence (lower bar) and whether the best match is significantly separated from the second-best match (upper bar). (d) A schematic drawing of the *orf* gene cluster that hosts the BoNT/X gene (upper panel), which has two distinct features compared with other known *orfX* clusters (middle and lower panels): (1) there is an additional *orfX2* protein (designated *orfX2b*) located next to the BoNT/X gene; (2) the reading frame of *orfX* genes has the same direction as the BoNT/X gene.

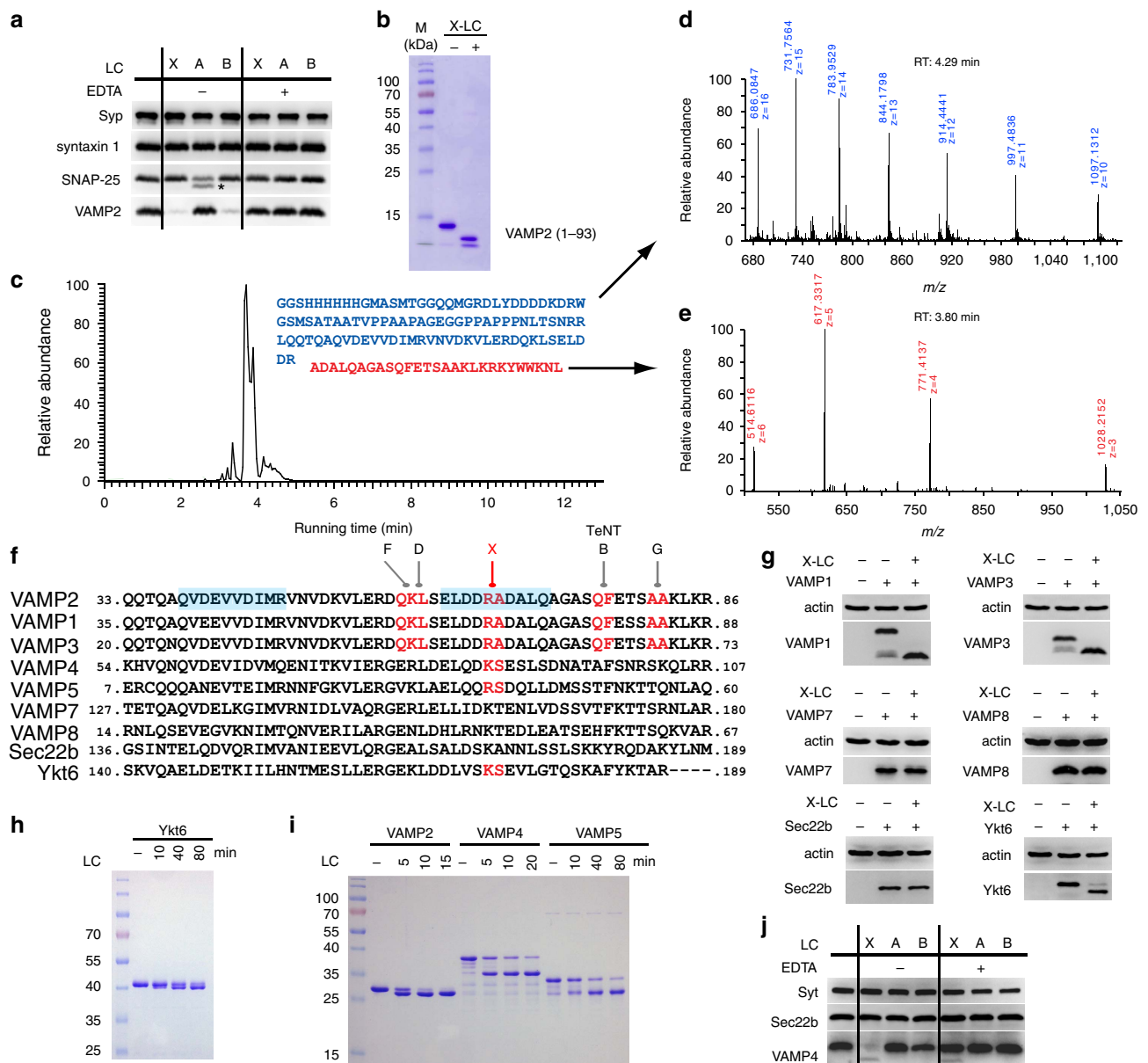
from the beginning of the His6-tag to residue R66 of VAMP2 was also detected (Fig. 2d). To further confirm this finding, we repeated the assay with a different VAMP2 fragment: glutathione S-transferase (GST) tagged VAMP2 (33–86) (Supplementary Fig. 3). Incubation with X-LC generated a single dominant peptide peak with a molecular weight of 2,063.1, which fits only A67–R86 of VAMP2 (Supplementary Fig. 3). Together, these results demonstrate that X-LC has a single cleavage site on VAMP2 between R66 and A67.

R66–A67 is a novel cleavage site on VAMP2, distinct from all established target sites of BoNTs (Fig. 2f). It is also the only BoNT cleavage site located within a region previously known as the SNARE motif (Fig. 2f, shaded regions)<sup>39</sup>. The VAMP protein family includes VAMP1, 2, 3, 4, 5, 7 and 8, as well as related Sec22b and Ykt6. R66–A67 is conserved in VAMP1 and 3, which are highly homologous to VAMP2. To validate the specificity of X-LC, we expressed HA-tagged VAMP1, 3, 7, 8 and Myc-tagged Sec22b and Ykt6 in HEK293 cells via transient transfection. Cell lysates were incubated with X-LC. Both VAMP1 and 3 were cleaved by X-LC, whereas VAMP7, VAMP8 and Sec22b were resistant to X-LC (Fig. 2g).

**BoNT/X cleaves VAMP4, VAMP5 and Ykt6.** Unexpectedly, Ykt6 was also cleaved by X-LC (Fig. 2g). This finding was confirmed using a purified GST-tagged Ykt6 fragment, which shifted to a lower-molecular-weight band after incubation with X-LC (Fig. 2h). The cleavage site was determined to be K173–S174 by

mass spectrometry analysis of the intact Ykt6 versus Ykt6 cleaved by X-LC (Supplementary Fig. 4). This site is homologous to the cleavage site of BoNT/X on VAMP2 (Fig. 2f). Among VAMP family of proteins, VAMP4 contains the same pair of residues (K87–S88) at this site as Ykt6. We found that X-LC cleaved both purified GST-tagged cytoplasmic domain of VAMP4 (Fig. 2i), as well as native VAMP4 in BDE (Fig. 2j). As a control, Sec22b was not cleaved by X-LC in BDE. In addition, the GST-tagged cytoplasmic domain of VAMP5 was also cleaved (Fig. 2i). The cleavage sites were determined by mass spectrometry analysis to be K87–S88 in VAMP4 and R40–S41 in VAMP5 (Supplementary Fig. 5). Both sites are homologous to the cleavage site of BoNT/X on VAMP2 (Fig. 2f), demonstrating that the location of the cleavage site is conserved across different VAMPs. The ability of X-LC to cleave VAMP4, VAMP5 and Ykt6 is highly unusual, as their sequences are substantially different from VAMP1/2/3. BoNT/X is the first and the only BoNT known that can cleave VAMPs beyond the canonical targets VAMP1, 2 and 3 (ref. 40).

**Proteolytic activation of BoNT/X.** We next examined the linker region between the LC and the HC, which must be cleaved by bacterial or host proteases to convert the toxin to an ‘active’ di-chain form. We produced a recombinant X-LC-H<sub>N</sub> fragment (residues 1–891) in *E. coli* and subjected it to limited proteolysis by endoproteinase Lys-C. Samples were analysed using Tandem Mass Tag (TMT) labelling and tandem mass spectrometry. TMT labels free N-termini (and lysines). Limited proteolysis by Lys-C

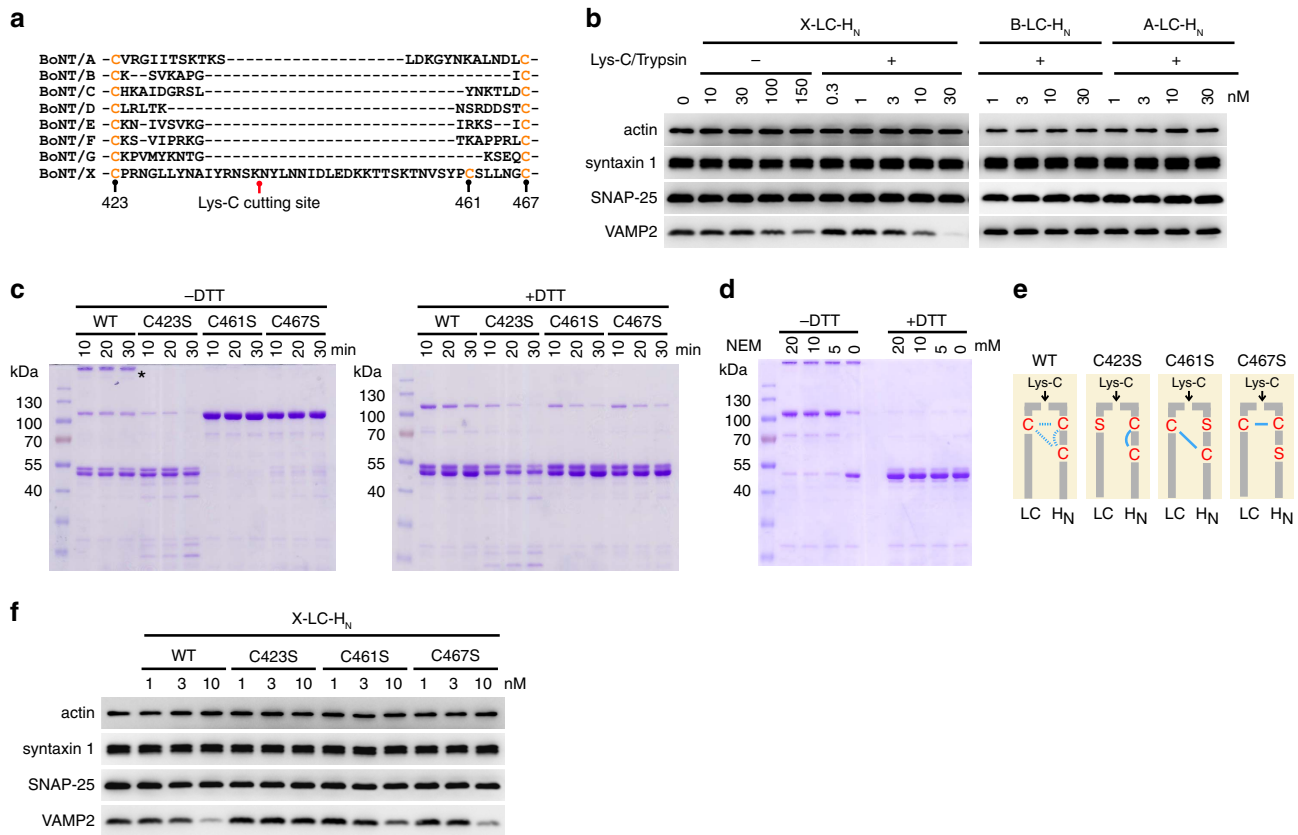


**Figure 2 | The LC of BoNT/X cleaves VAMPs at a unique site.** (a) X-LC was incubated with BDE. Immunoblot analysis was carried out to detect syntaxin 1, SNAP-25 and VAMP2. Synaptophysin (Syp) served as a loading control. A-LC and B-LC were analysed in parallel. Cleavage of VAMP2 by B-LC results in loss of immunoblot signals, while cleavage of SNAP-25 by A-LC generates a smaller fragment (marked with an asterisk). EDTA blocked the activity of X-, A- and B-LCs. (b) VAMP2 (1-93) was incubated with X-LC. Samples were analysed by SDS-PAGE and Coomassie Blue staining. X-LC converted VAMP2 (1-93) into two smaller fragments. (c-e) VAMP2 (1-93) was incubated with X-LC. Samples were analysed by mass spectrometry (LC-MS/MS) to determine the molecular weight of cleaved fragments. Eluted peptide peaks from the HPLC column are plotted over running time (RT, X axis). The mass spectrometry data for the two cleavage products are colour-coded, with mass-to-charge ratio ( $m/z$ ) noted. The molecular weight is deduced by multiplying  $m$  with  $z$ , followed by subtracting  $z$ . The protein sequences for the two cleavage products are colour-coded and listed in c. (f) Sequence alignment between VAMP family members, with the cleavage sites for BoNT/B, D, F, G and X marked in red, and the two SNARE motifs in blue shade. (g) HA-tagged VAMP1, 3, 7 and 8, and Myc-tagged Sec22b and Ykt6 were expressed in 293T cells via transient transfection. Cell lysates were incubated with X-LC and subjected to immunoblot analysis. Actin is a loading control. (h) GST-tagged Ykt6 was incubated with X-LC (100 nM). Samples were analysed by SDS-PAGE and Coomassie Blue staining. (i) GST-tagged VAMP2 (33-86), VAMP4 (1-115) and VAMP5 (1-70) were incubated with X-LC (100 nM). Samples were analysed by SDS-PAGE and Coomassie Blue staining. X-LC cleaved both VAMP4 and VAMP5. We note that VAMP5 protein contains a contaminant band that runs close to the cleavage product. (j) Experiments were carried out as described in a, except that VAMP4 and Sec22b were detected. Synaptotagmin I (Syt I) is a loading control. X-LC cleaved native VAMP4 in BDE. One of two (b,g,i) or three (a,h,i) independent experiments is shown.

produced one new free N terminus mapped to residue N439 in the linker region (Fig. 3a; Supplementary Data 1), confirming that the linker region is susceptible to proteases.

We then examined whether proteolytic activation enhances the potency of BoNT/X. It has been shown that incubation of high concentrations of LC- $H_N$  of BoNTs with cultured neurons results

in entry of LC- $H_N$ , likely through non-specific uptake into neurons<sup>41</sup>. Similarly, X-LC- $H_N$  entered cultured rat cortical neurons and cleaved VAMP2 in a concentration-dependent manner (Fig. 3b). Activation by Lys-C increased the potency of X-LC- $H_N$ : 10 nM activated X-LC- $H_N$  cleaved similar levels of VAMP2 as 150 nM intact X-LC- $H_N$  (Fig. 3b). Activated X-LC- $H_N$



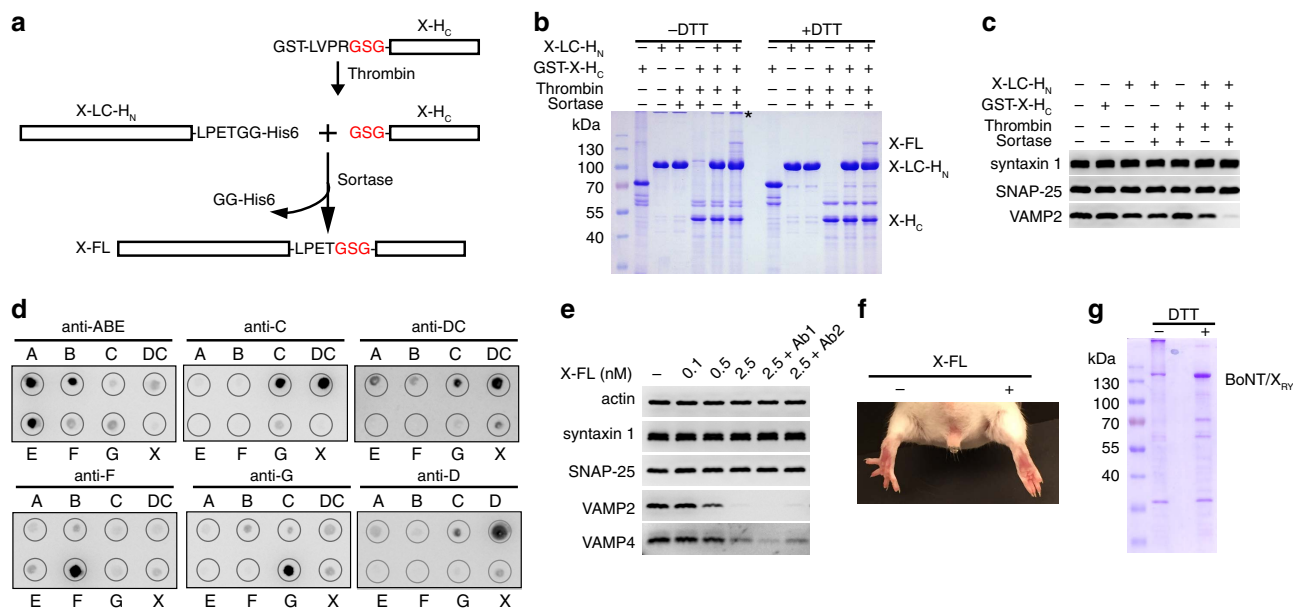
**Figure 3 | Proteolytic activation and inter-chain disulfide bond in BoNT/X.** (a) Sequence alignment of the linker between the LC and HC of the seven established BoNTs plus BoNT/X. The Lys-C cutting site was identified by mass spectrometry analysis (see Method and Supplementary Data 1). (b) Cultured rat cortical neurons were exposed to X-LC-H<sub>N</sub> for 12 h. Cell lysates were harvested and immunoblot analysis carried out to examine syntaxin 1, SNAP-25 and VAMP2. Actin is a loading control. Trypsin-activated A-LC-H<sub>N</sub> and B-LC-H<sub>N</sub> were analysed in parallel. X-LC-H<sub>N</sub> entered neurons and cleaved VAMP2. X-LC-H<sub>N</sub> activated by Lys-C showed a greater potency than non-activated X-LC-H<sub>N</sub>. X-LC-H<sub>N</sub> was more potent than B-LC-H<sub>N</sub> and A-LC-H<sub>N</sub>, neither of which cleaved their substrates. (c) WT and mutant X-LC-H<sub>N</sub> were activated by Lys-C and analysed by SDS-PAGE and Coomassie Blue staining, with or without DTT. C461S and C467S mutants showed as a single band at ~100 kDa without DTT, and separated into two ~50 kDa bands with DTT. A portion of WT X-LC-H<sub>N</sub> formed aggregates, marked by an asterisk, which disappeared with DTT. The majority of activated WT X-LC-H<sub>N</sub> separated into two ~50 kDa bands without DTT. This is due to disulfide bond shuffling as described in the following panel. (d) Lys-C-activated WT X-LC-H<sub>N</sub> was incubated with NEM to block disulfide bond shuffling. Samples were then analysed by SDS-PAGE and Coomassie Blue staining. A majority of WT X-LC-H<sub>N</sub> exists as a single band at ~100 kDa without DTT after NEM treatment, indicating that native WT X-LC-H<sub>N</sub> contains an inter-chain disulfide bond. (e) Schematic drawings of the disulfide bond in WT and three cysteine mutants of BoNT/X. (f) Experiments were carried out as described in b, except that neurons were exposed to WT or X-LC-H<sub>N</sub> mutants. C423S mutation abolished the activity of X-LC-H<sub>N</sub>, whereas mutating C461 or C467 did not affect the activity of X-LC-H<sub>N</sub>. These results confirmed that the inter-chain disulfide bond is essential for the activity of X-LC-H<sub>N</sub>, and this inter-chain disulfide bond can be formed via either C423-C461 or C423-C467. One of two (b) or three (b,c,f) independent experiments is shown.

appears to be more potent than activated LC-H<sub>N</sub> of BoNT/A (A-LC-H<sub>N</sub>) and BoNT/B (B-LC-H<sub>N</sub>), which did not show any detectable cleavage of their substrates under the same assay conditions (Fig. 3b).

**The inter-chain disulfide bond in BoNT/X.** Like other BoNTs, the linker region of BoNT/X contains two conserved cysteines, but there is also an additional cysteine (C461) unique to BoNT/X (Fig. 3a). To determine the cysteine residues that form the essential inter-chain disulfide bond, we generated three X-LC-H<sub>N</sub> mutants, each with one of the three cysteine residues mutated (C423S, C461S and C467S). These mutants, as well as wild-type (WT) X-LC-H<sub>N</sub>, were subjected to limited proteolysis with Lys-C and then analysed via SDS-PAGE and Coomassie Blue staining, with or without the reducing agent dithiothreitol (DTT; Fig. 3c). Mutating the only cysteine on the LC (C423S) is expected to abolish the inter-chain disulfide bond. Consistently, C423S mutant separated into two ~50 kDa bands without DTT.

In contrast, both C461S and C467S mutants showed as a single band at 100 kDa in the absence of DTT and separated into two ~50 kDa bands in the presence of DTT. These results suggested that C423 on the LC can form the inter-chain disulfide bond with either C461 or C467 on the HC. We also found that Lys-C treatment degraded a significant portion of C423S mutant as compared with C461S or C467S mutants (Fig. 3c, + DTT), suggesting that losing the inter-chain disulfide bond makes the molecule more susceptible to proteases. We noticed that a portion of WT X-LC-H<sub>N</sub> formed aggregates at the top of the SDS-PAGE gel (Fig. 3c, marked by an asterisk). These aggregates disappeared in the presence of DTT. C423/C461/C467 are the only three cysteines in the X-LC-H<sub>N</sub>; mutating any one of them abolished formation of aggregates (Fig. 3c, - DTT), suggesting that these aggregates are formed by inter-molecular disulfide bonds due to the existence of an extra cysteine in the linker region.

Interestingly, the majority of activated WT X-LC-H<sub>N</sub> separated into two ~50 kDa bands without DTT (Fig. 3c), which is similar to C423S mutant. On the other hand, WT X-LC-H<sub>N</sub> did not show



**Figure 4 | Full-length BoNT/X is active on cultured neurons and *in vivo* in mice.** (a) A schematic drawing of the sortase ligation method. (b) Sortase ligation reaction mixtures were analysed by SDS-PAGE and Coomassie Blue staining. The asterisk marks the proteins aggregates due to inter-molecular disulfide bonds. Full-length BoNT/X (X-FL) appeared only in the sortase ligation mixture. (c) Neurons exposed to the sortase ligation mixture (15  $\mu$ l) or control mixtures for 12 h in culture medium. Cell lysates were analysed by immunoblot. The mixture containing both X-LC-H<sub>N</sub> and X-H<sub>C</sub> (but not sortase) cleaved slightly more VAMP2 than X-LC-H<sub>N</sub> alone. Ligating X-LC-H<sub>N</sub> and X-H<sub>C</sub> by sortase further enhanced cleavage of VAMP2, demonstrating that ligated X-FL is functional on neurons. (d) BoNT/A-G, BoNT/DC and BoNT/X were subjected to the dot blot assay, using four horse antisera (trivalent anti-BoNT/A, B and E, anti-BoNT/C, anti-BoNT/DC and anti-BoNT/F), as well as two goat antisera (anti-BoNT/G and anti-BoNT/D). BoNT/X is composed of X-LC-H<sub>N</sub> and X-H<sub>C</sub> at 1:1 molar ratio. These antisera recognized their corresponding target toxins, yet none recognized BoNT/X. The antisera against BoNT/DC and BoNT/C cross-react, as these two toxins share a high degree of similarity within their H<sub>C</sub> domains. (e) Cultured rat cortical neurons were exposed to ligated X-FL in culture medium for 12 h, with or without two combinations of anti-sera. Ab1: trivalent anti-BoNT/A/B/E, anti-BoNT/C and anti-BoNT/F. Ab2: anti-BoNT/G and anti-BoNT/D. The trivalent anti-BoNT/A/B/E was used at 1:50 dilution. All other anti-sera were used at 1:100 dilution. None of the antisera affected the cleavage of VAMP2 and VAMP4 by X-FL. The specificity and potency of these antisera were validated for their ability to neutralize target serotypes in the same assay as described in Supplementary Fig. 7. (f) X-FL linked by sortase reaction (0.5  $\mu$ g) was injected into the gastrocnemius muscles of the right hind limb of mice ( $n=4$ ). The injected limb developed typical flaccid paralysis, and the toes failed to spread within 12 h. The left limb was not injected with toxins, serving as a control. (g) Full-length inactive form of BoNT/X (BoNT/X<sub>RY</sub>) was purified as a His6-tagged recombinant protein in *E. coli*. Further purified BoNT/X<sub>RY</sub> is shown in Supplementary Fig. 8b. One of two (e) or three (c,d) independent experiments is shown.

increased degradation by Lys-C compared with C423S mutant (Fig. 3c, + DTT). One possible explanation is that WT X-LC-H<sub>N</sub> contains an inter-chain disulfide bond under native conditions, but this bond can rearrange to intra-chain C461–C467 pair under denaturing conditions in the SDS buffer. This phenomenon is known as disulfide bond shuffling, which often occurs among adjacent cysteines. To test this hypothesis, we utilized an alkylating reagent, *N*-Ethylmaleimide (NEM), which permanently blocks free cysteines and prevents disulfide bond shuffling. As shown in Fig. 3d, WT X-LC-H<sub>N</sub> pretreated with NEM showed as a single band at 100 kDa in the absence of DTT, and separated into two  $\sim$ 50 kDa bands in the presence of DTT. These results confirm that WT X-LC-H<sub>N</sub> contains mainly an inter-chain disulfide bond, but it is susceptible to disulfide bond shuffling due to an extra cysteine in the linker region (Fig. 3e).

We further examined the activity of the three X-LC-H<sub>N</sub> cysteine mutants on cultured neurons. As expected, C423S mutant was inactive, whereas C461S and C467S mutants both showed similar levels of activity as WT X-LC-H<sub>N</sub> (Fig. 3f). These results confirm that the inter-chain disulfide bond is critical for the activity of BoNT/X.

**Generating full-length BoNT/X via sortase-mediated ligation.** We then sought to determine whether full-length BoNT/X is a functional toxin. As no antisera against BoNT/X are available,

we decided to avoid generating the full-length active toxin gene. Instead, we developed an approach to generate a limited amount of full-length BoNTs in test tubes by enzymatic ligation of two non-toxic fragments of BoNTs. This method utilizes a transpeptidase known as sortase<sup>42,43</sup>, which recognizes the peptide motif LPXTG, cleaves between T-G, and concurrently forms a new peptide bond with other proteins/peptides containing N-terminal glycine (Fig. 4a). We produced two non-toxic fragments of BoNT/X: (1) LC-H<sub>N</sub> with a LPETGG motif and a His6-tag fused to the C terminus; and (2) the H<sub>C</sub> of BoNT/X (X-H<sub>C</sub>) with a GST tag, thrombin cleavage site, and an additional glycine residue at its N terminus. Cutting by thrombin releases X-H<sub>C</sub> with a free glycine at its N terminus. Incubation of these two fragments with sortase generated a small amount of  $\sim$ 150 kD full-length BoNT/X (X-FL, Fig. 4a,b). We note that X-H<sub>C</sub> showed poor solubility and a strong tendency towards aggregation, which might be the reason for the low ligation efficiency (Fig. 4b). In contrast, ligation of X-LC-H<sub>N</sub> with the H<sub>C</sub> of BoNT/A (A-H<sub>C</sub>) achieved a better efficiency, with the majority of X-LC-H<sub>N</sub> ligated into a XA chimeric toxin (Supplementary Fig. 6a). To ensure biosafety, the amount of precursor fragments in the reaction is strictly limited to generate the minimum amount of ligated toxin necessary for functional assays.

We first analysed the activity of ligated BoNT/X using cultured rat cortical neurons. Neurons were exposed to the sortase ligation mixture and control mixtures in culture medium. As shown in

Fig. 4c, X-LC-H<sub>N</sub> alone cleaved some VAMP2 due to its high concentration in the reaction mixture. Mixing X-H<sub>C</sub> with X-LC-H<sub>N</sub> without sortase slightly enhanced cleavage of VAMP2 compared with X-LC-H<sub>N</sub> alone, suggesting that X-H<sub>C</sub> might be associated with X-LC-H<sub>N</sub> via non-covalent interactions. This interaction appears to be specific, as mixing A-H<sub>C</sub> with X-LC-H<sub>N</sub> did not enhance cleavage of VAMP2 in neurons (Supplementary Fig. 6b). Ligating X-LC-H<sub>N</sub> with X-H<sub>C</sub> by sortase clearly enhanced cleavage of VAMP2 compared with the mixture of X-LC-H<sub>N</sub> and X-H<sub>C</sub> without sortase (Fig. 4c). These results demonstrated that the X-H<sub>C</sub> is functional for targeting cells and that ligated full-length BoNT/X entered neurons and cleaved VAMP2. Similarly, ligated XA also entered neurons and cleaved VAMP2 (Supplementary Fig. 6b).

#### BoNT/X was not recognized by antisera against known BoNTs.

We next carried out dot blot assays using antisera raised against known BoNTs, including all seven serotypes as well as one mosaic toxin (BoNT/DC), to confirm that BoNT/X is serologically unique. Four horse antisera were utilized (trivalent anti-BoNT/A, B and E, anti-BoNT/C, anti-BoNT/DC, and anti-BoNT/F) as well as two goat antisera (anti-BoNT/G and anti-BoNT/D). The specificity and potency of these antisera were first validated by analysing their ability to neutralize BoNTs on cultured neurons. As expected, all antisera neutralized their target BoNTs, without affecting the activity of a different serotype (Supplementary Fig. 7). We found that these antisera recognized their corresponding BoNTs in the dot blot assay, yet none recognized BoNT/X (Fig. 4d).

We further analysed whether the toxicity of BoNT/X on neurons can be neutralized by these antisera. X-FL generated by sortase-mediated ligation was first activated with limited proteolysis using trypsin. We used trypsin to activate X-FL instead of Lys-C for functional assays, as trypsin allows us to stop proteolysis using trypsin inhibitors. Activated X-FL entered cultured rat cortical neurons and cleaved both VAMP2 and VAMP4 in a concentration-dependent manner (Fig. 4e). Combinations of antisera against known BoNTs (Ab1 (horse antisera): trivalent anti-BoNT/A, B and E, anti-BoNT/C, and anti-BoNT/F; Ab2 (goat antisera): anti-BoNT/G and anti-BoNT/D) did not affect the activity of ligated X-FL, as evidenced by similar degrees of VAMP2 and VAMP4 cleavage in the presence of these antisera (Fig. 4e). These results confirmed that BoNT/X is a new BoNT serotype.

**BoNT/X induced flaccid paralysis *in vivo* in mice.** We next sought to determine whether BoNT/X is active *in vivo* using a well-established non-lethal assay in mice, known as the Digit Abduction Score (DAS) assay, which measures local muscle paralysis following injection of BoNTs into mouse hind limb muscles<sup>44</sup>. BoNTs cause flaccid paralysis of limb muscles, which is manifested as the failure to spread the toes in response to a startle stimulus. We injected ligated X-FL (0.5 µg, activated by trypsin treatment) into the gastrocnemius muscles of the right hind limb in mice, which induced typical flaccid paralysis and the failure of toes to spread (Fig. 4f), indicating that BoNT/X is capable of causing flaccid paralysis *in vivo*. We note that the potency of ligated X-FL appears to be much lower than other BoNTs in this assay. To further confirm the low toxicity of ligated X-FL, we injected mice with 1 µg of ligated X-FL intraperitoneally ( $n = 3$ ). No mice showed any systemic effects and all survived at this dose. Thus, ligated X-FL has a rather low toxicity *in vivo* in mice compared with other native BoNTs, which usually have lethal doses at low picogram levels per mouse.

**Full-length inactive BoNT/X.** Finally, we developed an inactive mutant of BoNT/X as a potential reagent for generating neutralizing antibodies. Mutations at two residues (R362A/Y365F) in BoNT/A inactivate the protease activity of the LC and abolish the toxicity of BoNT/A *in vivo*<sup>45,46</sup>. These two residues are conserved in BoNT/X. We introduced the corresponding mutations (R360A/Y363F) in BoNT/X and generated a full-length inactive form, designated as BoNT/X<sub>RY</sub>. As shown in Fig. 4g, BoNT/X<sub>RY</sub> was purified as a His6-tagged protein in *E. coli*, and it had no activity on cultured neurons (Supplementary Fig. 8a). Furthermore, intraperitoneal injection of mice with 30 µg BoNT/X<sub>RY</sub> (activated by trypsin treatment) did not cause any adverse effects ( $n = 5$ ), demonstrating that it is not toxic *in vivo*. A substantial portion of BoNT/X<sub>RY</sub> formed aggregates at the top of the SDS-PAGE gel (Fig. 4g). Adding DTT reduced these aggregates to monomeric BoNT/X<sub>RY</sub> (Fig. 4g). Thus, full-length BoNT/X is susceptible to forming intermolecular disulfide bonds. Nevertheless, the monomeric form of BoNT/X can be purified and is stable in solution (Fig. 4g). Furthermore, we developed a scale-up purification protocol, which generated BoNT/X<sub>RY</sub> with a yield of ~3 mg per liter of culture and ~90% purity (Supplementary Fig. 8b). Highly purified BoNT/X<sub>RY</sub> remained stable in solution up to 10 mg ml<sup>-1</sup> in the presence of reducing agent. This atoxic BoNT/X<sub>RY</sub> will be a valuable reagent for generating neutralizing antibodies.

#### Discussion

BoNT/X is the first serotype of BoNTs identified by genomic sequencing and bioinformatics approaches. It remains unknown whether BoNT/X is ever produced in *C. botulinum* strain 111. BoNT/X could be a silent gene, or it may not be expressed at detectable toxicity levels under culture conditions in the lab. Thus, this toxin was revealed only by sequencing *C. botulinum* 111. This illustrates the importance of genomic sequencing and bioinformatics approaches for understanding microbial virulence factors. Whether BoNT/X could be expressed and exhibit toxicity under certain environmental conditions remains an intriguing question.

A remarkable feature of BoNT/X is its unique ability to cleave VAMP4 and Ykt6. VAMP4 is widely expressed and is known to mediate vesicle fusion between the trans-Golgi network (TGN) and endosomes, as well as homotypic fusion of endosomes<sup>47,48</sup>. Ykt6 is an atypical SNARE without a transmembrane domain<sup>49</sup>. It is anchored to membranes via lipidation, which allows dynamic regulation of its membrane association. Ykt6 is an essential protein in yeast and implicated in multiple membrane fusion events including ER-Golgi, intra-Golgi, endosome-Golgi-vacuolar, and autophagosome formation. Its function in mammalian cells remains to be established. BoNT/X is the first and only BoNT to cleave these SNAREs that mediate various intracellular membrane-trafficking events.

Interestingly, both VAMP4 and Ykt6 are enriched in neurons. Recent studies suggested that VAMP4 contributes to asynchronous synaptic vesicle exocytosis, enlarges exocytosis and activity-dependent bulk endocytosis (ADBE) in neurons<sup>50-52</sup>. The role of Ykt6 in neurons remains to be established, but it has been shown to suppress the toxicity of  $\alpha$ -synuclein in Parkinson's disease models<sup>53,54</sup>. The other substrate of BoNT/X, VAMP5, is mainly expressed in muscle cells and its function remains to be established<sup>55</sup>. BoNT/X will be a useful tool to investigate the function of VAMP4, Ykt6 and VAMP5, as well as related membrane trafficking events. In addition, because VAMP4 has been implicated in granule release in immune cells<sup>56</sup>, BoNT/X may have the potential to modulate inflammatory secretion in immune cells.

The X-LC-H<sub>N</sub> fragment showed a higher level of activity in neurons than either A-LC-H<sub>N</sub> or B-LC-H<sub>N</sub>, suggesting that its membrane translocation and/or protease activity might be more efficient than the corresponding fragments in BoNT/A and BoNT/B. X-H<sub>C</sub> is functional for targeting cells, as its presence enhanced cleavage of VAMP2 in neurons over LC-H<sub>N</sub> alone (Fig. 4c). When present without the translocation and LC domains, X-H<sub>C</sub> is prone to aggregation. This solubility issue is likely due to separation of X-H<sub>C</sub> from X-LC-H<sub>N</sub>, as full-length BoNT/X<sub>RY</sub> remains stable at high concentrations. The X-FL generated by sortase-mediated ligation has a rather low toxicity *in vivo* in mice. It remains unknown whether this low *in vivo* toxicity was intrinsic to BoNT/X. It is also possible that the sortase linking method resulted in an attenuated toxin, as the H<sub>C</sub> and LC-H<sub>N</sub> folded separately and there is an additional linker between the H<sub>N</sub> and H<sub>C</sub> in ligated toxins. Nevertheless, X-FL produced by sortase-mediated ligation is active on neurons and induced typical flaccid paralysis when injected locally in mice, demonstrating that BoNT/X is a functional toxin (Fig. 4f). It will be necessary to produce native BoNT/X to characterize its *in vivo* potency and determine its biosafety risk. It will be important to generate neutralizing antisera against BoNT/X before producing any native toxin.

Introducing a full-length active toxin gene into any expression system/organism is always a significant biosafety concern. Sortase-mediated ligation assembles a small quantity of full-length toxin from two complementary and non-toxic fragments expressed and purified individually. The amount of the precursor fragments in the reaction can be strictly controlled, so the amount of ligated toxin is precisely limited to ensure biosafety. The ligated toxin should possess the same mode of action as native toxins, but may exhibit lower toxicity *in vivo*, possibly due to the addition of the sortase linker and/or reduced compactness of the molecule. Thus, the ligated toxin could be used for functional studies, but may not be suitable for estimating the lethal dose of native toxin. On the other hand, this reduced toxicity from the sortase linking method could be an advantage to create attenuated toxins for research use. It might be also possible to deliberately attenuate the toxicity of ligated toxins *in vivo* by including additional linkers, such as peptide sequences that are sensitive to serum proteases, to further mitigate biosafety concerns.

## Methods

**Materials.** Mouse monoclonal antibodies for syntaxin 1 (HPC-1, dilution 1:5,000), SNAP-25 (C171.2, 1:5,000), VAMP2 (C169.1, 1:2,000), Syp (C17.2, 1:2,000) and Syt I (mAB48, 1:1,000) were generously provided by E. Chapman (Madison, WI, USA). Most are available from Synaptic Systems (Goettingen, Germany), with the exception of Syt I (mAB48), which is available from the Developmental Studies Hybridoma Bank. Rabbit polyclonal antibodies against VAMP4 (Cat. no. 136002, 1:1,000) and Sec22b (Cat. no. 186003, 1:1,000) were purchased from Synaptic Systems. The following mouse monoclonal antibodies were purchased from the indicated vendors: actin (Sigma, AC-15, 1:2,000); anti-HA (Covance, 16B12, 1:2,000); anti-Myc (Millipore, 9E10, 1:1,000). Equine polyclonal antisera against BoNT/A/B/E, BoNT/C, BoNT/DC, BoNT/F and goat polyclonal antisera against BoNT/G were generously provided by S. Sharma (FDA). Goat polyclonal antibody against BoNT/D was purchased from Fisher Scientific (NB10062469). BoNTs were purchased from Metabolics (Madison, WI, USA). Antibody validation is available on the manufacturers' websites. 293T (#CRL-3216) cells were originally obtained from ATCC, which were negative for mycoplasma contamination but have not been authenticated.

**cDNA and constructs.** The cDNAs encoding X-LC (residues 1–439), X-H<sub>C</sub> (residues 893–1,306), A-LC-H<sub>N</sub> (residues 1–874, GenBank no. M30196), and B-LC-H<sub>N</sub> (residues 1–860, GenBank no. AB232927) were synthesized by GenScript (New Brunswick, NJ, USA). The cDNA encoding X-H<sub>N</sub> was generated in-house using the Gibson assembly method. X-LC, A-LC (residues 1–425) and B-LC (residues 1–439) were cloned into pET28 vectors with His6-tag on their N-termini. X-H<sub>C</sub> and A-H<sub>C</sub> (residues 875–1,297, GenBank No. AF488749) were cloned into pGEX4T to express as GST-tagged proteins. One extra glycine was introduced into the N terminus of X-H<sub>C</sub> to increase the sortase ligation efficiency. X-LC-H<sub>N</sub>,

A-LC-H<sub>N</sub> and B-LC-H<sub>N</sub> were cloned into pET22b vector, with the peptide sequence LPETGG fused to their C-termini, followed by a His6-tag, and purified as His6-tagged proteins. The cDNA encoding rat VAMP2 was generously provided by E. Chapman (Madison, WI, USA). VAMP2 (1–93) was cloned into pET28 vector with a His6-tag on the N terminus. VAMP2 (33–86) was cloned into pGEX4T vector and expressed as a GST-tagged protein. The cDNAs encoding mouse VAMP1, VAMP3, VAMP4, human VAMP5, rat VAMP7 and VAMP8 were generously provided by C. Hu (Louisville, KY, USA). Full-length VAMP1, 3, 7 and 8 were cloned into modified pcDNA3.1 vectors, with an HA tag fused to their C termini. Constructs expressing full-length rat Ykt6 and mouse Sec22b, both in pcDNA3.1 vector with a Myc tag fused to the N terminus of the protein, were generously provided by J. Hay (Missoula, MT, USA). The cytoplasmic domains of VAMP4 (1–115) and VAMP5 (1–70) were cloned between BamHI/XhoI sites in pGEX4T and expressed as GST-tagged proteins. We note that there are seven extra residues from the pGEX4T vector fused to the C-termini of VAMP4 and VAMP5 fragments. Ykt6 fragment (residues 1–192) was also cloned into pGEX4T and expressed as GST-tagged proteins. The construct encoding His6-tagged sortase (SrtA\*) was generously provided by B. Pentelute (Boston, MA, USA)<sup>43</sup>.

**Bioinformatics.** The Uniprot database was searched with JackHMMER on the HMMER web server, using a BoNT/A1 sequence as the seed (Uniprot accession number A5HZZ9) until convergence. Returned sequences were aligned with Clustal Omega and a NeighborNet phylogenetic network estimated with SplitsTree.

**Protein purification.** *E. coli* BL21 (DE3) was utilized for protein expression. In general, induction of expression was carried out with 0.1 mM IPTG at 22 °C overnight. Bacterial pellets were disrupted in lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl) by sonication, and supernatants were collected after centrifugation at 20,000 g for 30 min at 4 °C. Protein purification was carried out using AKTA Prime FPLC system (GE), and purified proteins were further desalted with a PD-10 column (GE, 17-0851-01).

**Large-scale production and purification of BoNT/X<sub>RY</sub>.** cDNA encoding BoNT/X<sub>RY</sub> was assembled in-house from mutated X-LC (R360A/Y363F), X-H<sub>N</sub> and X-H<sub>C</sub>. It was cloned into a pET22b vector, with the His6-tag on its C terminus. The corresponding plasmid was transformed into *E. coli* BL21 (DE3). Cultures for expression were first grown using a LEX Bioreactor (Epiphyte3, Ontario, Canada) at 37 °C in 1.5 l of medium until OD<sub>600</sub> reached 0.8. The temperature was then reduced to 18 °C for induction of expression with 1 mM IPTG, and grown for 16–17 h. Bacteria were harvested, re-suspended in HEPES buffer (50 mM HEPES pH 7.2, 500 mM NaCl, 25 mM imidazole, 5% glycerol, 2 mM TCEP), and lysed with an Emulsiflex-C3 (Avestin, Mannheim, Germany) at 20,000 p.s.i. Lysates were ultra-centrifuged at 200,000 g for 45 min. Supernatant was loaded onto a 15 ml Ni-NTA agarose column and washed with wash buffer (50 mM HEPES pH7.2, 500 mM NaCl, 100 mM imidazole, 5% glycerol, 1 mM TCEP). Proteins were eluted with elution buffer (50 mM HEPES pH 7.2, 500 mM NaCl, 250 mM imidazole, 5% glycerol, 1 mM TCEP) and then dialyzed overnight in 50 mM HEPES, 500 mM NaCl, 5% glycerol, and 0.5 mM TCEP. Dialysate was concentrated using a Vivaspin concentrator before being loaded on a Superdex200-16/60 column pre-equilibrated in the same buffer used for dialysis. Elution peak corresponding to BoNT/X was collected and concentrated to ~10 mg ml<sup>-1</sup>. Sample was aliquoted and flash-frozen in liquid nitrogen for storage at -80 °C.

**Cleavage of SNARE proteins in rat BDE.** Rat brain was homogenized in 15 ml 320 mM sucrose buffer, followed by centrifugation at 5,000 r.p.m. for 2 min at 4 °C. Supernatants were collected and centrifuged at 11,000 r.p.m. for 12 min. The pellet was collected and solubilized for 30 min in 15 ml Tris-buffered saline (TBS: 20 mM Tris, 150 mM NaCl) plus 2% of Triton X-100 and a cocktail of protease inhibitors (Roche, CA). Samples were subsequently centrifuged at 17,000 r.p.m. for 20 min to remove insoluble materials. The final BDE concentration was ~2 mg ml<sup>-1</sup>. BDE (60 µl) were incubated with X-LC (0.5 µM), A-LC (1 µM), or B-LC (1 µM), for 1 h at 37 °C, and then analysed by immunoblot using the enhanced chemiluminescence (ECL) method (Pierce). As controls, LCs were pre-incubated with 20 mM EDTA for 20 min at room temperature before adding to BDE. Full-blot scans are shown in Supplementary Fig. 9.

**Cleavage of recombinant VAMPs by X-LC.** VAMP2 (1–93) was expressed and purified as a His6-tagged protein. VAMP2 (33–86), VAMP4 (1–115), VAMP5 (1–70) and Ykt6 (1–192) were expressed and purified as GST-tagged proteins. These proteins (0.3 mg ml<sup>-1</sup>) were incubated with 0.1 µM X-LC in TBS buffer at 37 °C. Samples were either analysed by SDS-PAGE gels and Coomassie Blue staining, or subjected to mass spectrometry analysis.

**Cleavage of VAMPs in cell lysates:** Full-length HA-tagged VAMP1, 3, 7, 8 and Myc-tagged Sec22b, and Ykt6 were transfected into 293T cells using PolyJet transfection reagents (SigmaGen, MD) following the manufacturer's instructions. Cell lysates were harvested 48 h later in RIPA buffer (50 mM Tris, 1% NP40, 150 mM NaCl, 0.5% sodium deoxycholate, 0.1% SDS, 400 µl per 10-cm dish) plus a



protease inhibitor cocktail (Sigma-Aldrich). Cell lysates (250  $\mu$ l) were incubated with X-LC (0.5  $\mu$ M) for 1 h at 37 °C. Samples were then analysed by immunoblot.

**Identification of cleavage sites in VAMPs by LC-MS/MS.** Samples were analysed at Taplin Biological Mass Spectrometry Core Facility at Harvard Medical School. For VAMP2, whole-protein samples were loaded onto a 100  $\mu$ m internal diameter C18 reverse-phase HPLC column packed with 3 cm of beads off-line using a pressure cell. The column was re-attached to an Accela 600 Pump (Thermo Fisher Scientific). A rapid gradient of increasing acetonitrile was used to elute the protein/peptide from the HPLC column. As peptides eluted, they were subjected to electrospray ionization and then placed into an LTQ Orbitrap Velos Pro ion-trap mass spectrometer to acquire a high-resolution FTMS scan at 60,000 resolution, a second scan at low resolution in the ion trap, and a final scan to perform data-dependent MS/MS. The charge state envelopes were de-convoluted manually to obtain mono-isotopic masses when possible or average masses for the proteins. Peptide and protein identity were determined by matching protein databases with the acquired fragmentation pattern using the software program Sequest (Thermo Fisher Scientific). All databases include a reversed version of all the sequences, and the data were filtered to 1–2% peptide false-discovery rate.

For Ykt6, VAMP4 and VAMP5, samples were first separated on SDS-PAGE. Protein bands were excised and cut into  $\sim$ 1 mm<sup>2</sup> pieces. Gel pieces were incubated with 50 mM ammonium bicarbonate solution containing 12.5 ng  $\mu$ l<sup>-1</sup> modified sequencing-grade chymotrypsin (Roche Diagnostics). Samples were digested overnight at room temperature. Peptides were then extracted and separated with reverse-phase HPLC. As peptides were eluted, they were subjected to electrospray ionization and transferred into an LTQ Orbitrap Velos Pro ion-trap mass spectrometer (Thermo Fisher Scientific). Eluted peptides were detected, isolated and fragmented to produce a tandem mass spectrum of specific fragment ions for each peptide.

**Identification of the protease cleavage site between LC and H<sub>N</sub>.** His6-tagged recombinant X-LC-H<sub>N</sub> fragment (residues 1–891) was purified in *E. coli* and subjected to limited proteolysis by endoprotease Lys-C (Sigma P2289, 100:1 (toxin:Lys-C) molar ratio, 25 min at room temperature). The cleavage site was determined by TMT labelling and tandem mass spectrometry. Briefly, intact X-LC-H<sub>N</sub> samples were labelled with the light TMT, and equal amounts of X-LC-H<sub>N</sub> samples treated with Lys-C were labelled with the heavy TMT. Both samples were then digested with chymotrypsin, combined and subjected to quantitative mass spectrometry analysis.

**Cysteine alkylation by NEM.** Lys-C-activated X-LC-H<sub>N</sub> fragment was diluted into sodium phosphate buffer (10 mM, pH 6.5) at a final concentration of 0.3 mg ml<sup>-1</sup>, with or without NEM at indicated concentrations (20, 10 and 5 mM) and incubated for 10 min at room temperature. NEM was freshly prepared in sodium phosphate buffer. Samples were mixed with 3  $\times$  neutral loading dye (200 mM Tris pH 6.8, 30% glycerol, 6% Lithium Dodecyl sulfate, 10 mM NEM, and 0.06% BPB) at room temperature for 10 min, heated for 10 min at 55 °C, and then analysed by SDS-PAGE and Coomassie Blue staining.

**Neuron culture and immunoblot analysis.** Primary rat cortical neurons were prepared from E18–19 embryos using a papain dissociation kit (Worthington Biochemical) following the manufacturer's instruction<sup>57</sup>. Neurons were exposed to either BoNT/X fragments or sortase ligation mixtures in culture medium for 12 h. Cells were then lysed with RIPA buffer plus a protease inhibitor cocktail (Sigma-Aldrich). Lysates were centrifuged for 10 min at maximum speed using a microcentrifuge at 4 °C. Supernatants were subjected to SDS-PAGE and immunoblot analysis.

**Dot blot.** BoNTs (0.2  $\mu$ g in 1  $\mu$ l) were spotted onto nitrocellulose membranes and dried (10 min at RT). The membranes were blocked with 5% milk in TBST (TBS plus 0.05% Tween20) for 30 min and then incubated with appropriate antisera (1:500 dilution) for 30 min. The membranes were then washed three times with TBST and incubated with HRP (horseradish peroxidase)-conjugated secondary antibodies for 30 min, washed three more times with TBST, and analysed by the ECL method. The BoNT/X sample was composed of X-LC-H<sub>N</sub> and GST-X-H<sub>C</sub> at 1:1 molar ratio.

**Sortase-mediated ligation.** GST-X-H<sub>C</sub> or GST-A-H<sub>C</sub> was cleaved overnight at 4 °C by thrombin before being added into the ligation reaction mixture. Ligation reaction was set up in 50  $\mu$ l TBS buffer with X-LC-H<sub>N</sub> (8  $\mu$ M), X-H<sub>C</sub> (4  $\mu$ M) or A-H<sub>C</sub> (25  $\mu$ M), Ca<sup>2+</sup> (10 mM) and sortase (10  $\mu$ M), for 40 min at room temperature.

**DAS assay.** All procedures were conducted in accordance with the guidelines approved by the Institute Animal Care and Use Committee (IACUC) at Boston Children's Hospital (#3030). Briefly, X-FL generated by sortase-mediated ligation was first activated with limited proteolysis using trypsin (60:1 (toxin:trypsin) molar

ratio, 30 min at room temperature). We chose trypsin instead of Lys-C here, as it allows us to stop proteolysis by adding trypsin inhibitors (Soybean trypsin inhibitor, 1:10 ratio (trypsin:trypsin inhibitor)). Mice (CD-1 strain, male, purchased from Charles River, 5–6 weeks old, 21–25 g,  $n=4$ ) were anesthetized with isoflurane (3–4%) and injected with X-FL (0.5  $\mu$ g) using a 30-gauge needle attached to a sterile Hamilton syringe, into the gastrocnemius muscles of the right hind limb. Muscle paralysis and the spread of hind paw in response to a startle stimulus were observed 12 h after injection as previously described<sup>44</sup>.

**Biosafety and biosecurity.** All procedures were approved by the Institute of Biosafety Committees at Boston Children's Hospital and at Stockholm University. To ensure biosafety and biosecurity, no active full-length toxin gene was produced in any form. The amount of sortase linking reaction is strictly controlled to ensure that only a minimal amount of ligated toxins was produced, which was immediately utilized for functional studies. The loss of toxicity of BoNT/X<sub>RY</sub> and reduced toxicity of sortase-linked X-FL were confirmed using the mouse lethality assay. Inactive BoNT/X<sub>RY</sub> was prepared and is available for developing neutralizing antibodies.

**Data availability.** The data and materials that support the findings of this study are available from the corresponding authors upon request.

## References

- Arnon, S. S. *et al.* Botulinum toxin as a biological weapon: medical and public health management. *JAMA* **285**, 1059–1070 (2001).
- Johnson, E. A. Clostridial toxins as therapeutic agents: benefits of nature's most toxic proteins. *Annu. Rev. Microbiol.* **53**, 551–575 (1999).
- Montecucco, C. & Molgo, J. Botulinum neurotoxins: revival of an old killer. *Curr. Opin. Pharmacol.* **5**, 274–279 (2005).
- Schiavo, G., Matteoli, M. & Montecucco, C. Neurotoxins affecting neuroexocytosis. *Physiol. Rev.* **80**, 717–766 (2000).
- Montal, M. Botulinum neurotoxin: a marvel of protein design. *Annu. Rev. Biochem.* **79**, 591–617 (2010).
- Rossetto, O., Pirazzini, M. & Montecucco, C. Botulinum neurotoxins: genetic, structural and mechanistic insights. *Nat. Rev. Microbiol.* **12**, 535–549 (2014).
- Jahn, R. & Scheller, R. H. SNAREs—engines for membrane fusion. *Nat. Rev. Mol. Cell Biol.* **7**, 631–643 (2006).
- Sudhof, T. C. & Rothman, J. E. Membrane fusion: grappling with SNARE and SM proteins. *Science* **323**, 474–477 (2009).
- Burke, G. S. Notes on *Bacillus botulinus*. *J. Bacteriol.* **4**, 555–570 (1919).
- Gimenez, D. F. & Ciccarelli, A. S. Another type of *Clostridium botulinum*. *Zentralbl. Bakteriolog. Orig.* **215**, 221–224 (1970).
- Smith, T. J. *et al.* Sequence variation within botulinum neurotoxin serotypes impacts antibody binding and neutralization. *Infect. Immun.* **73**, 5450–5457 (2005).
- Hill, K. K. *et al.* Genetic diversity among Botulinum Neurotoxin-producing clostridial strains. *J. Bacteriol.* **189**, 818–832 (2007).
- Montecucco, C. & Rasotto, M. B. On botulinum neurotoxin variability. *MBio.* **6**, e02131–e02134 (2015).
- Dover, N., Barash, J. R., Hill, K. K., Xie, G. & Arnon, S. S. Molecular characterization of a novel botulinum neurotoxin type H gene. *J. Infect. Dis.* **209**, 192–202 (2014).
- Barash, J. R. & Arnon, S. S. A novel strain of *Clostridium botulinum* that produces type B and type H botulinum toxins. *J. Infect. Dis.* **209**, 183–191 (2014).
- Maslanka, S. E. *et al.* A novel botulinum neurotoxin, previously reported as serotype H, has a hybrid-like structure with regions of similarity to the structures of serotypes A and F and is neutralized with serotype A antitoxin. *J. Infect. Dis.* **213**, 379–385 (2015).
- Kalb, S. R. *et al.* Functional characterization of botulinum neurotoxin serotype H as a hybrid of known serotypes F and A (BoNT F/A). *Anal. Chem.* **87**, 3911–3917 (2015).
- Zhou, Y., Sugiyama, H., Nakano, H. & Johnson, E. A. The genes for the *Clostridium botulinum* type G toxin complex are on a plasmid. *Infect. Immun.* **63**, 2087–2091 (1995).
- Marshall, K. M., Bradshaw, M., Pellett, S. & Johnson, E. A. Plasmid encoded neurotoxin genes in *Clostridium botulinum* serotype A subtypes. *Biochem. Biophys. Res. Commun.* **361**, 49–54 (2007).
- Jacobson, M. J., Lin, G., Raphael, B., Andreadis, J. & Johnson, E. A. Analysis of neurotoxin cluster genes in *Clostridium botulinum* strains producing botulinum neurotoxin serotype A subtypes. *Appl. Environ. Microbiol.* **74**, 2778–2786 (2008).
- Eklund, M. W., Poysky, F. T., Reed, S. M. & Smith, C. A. Bacteriophage and the toxigenicity of *Clostridium botulinum* type C. *Science* **172**, 480–482 (1971).
- Smith, T. J. *et al.* Analysis of the neurotoxin complex genes in *Clostridium botulinum* A1–A4 and B1 strains: BoNT/A3, /Ba4 and /B1 clusters are located within plasmids. *PLoS ONE* **2**, e1271 (2007).

23. Hill, K. K., Xie, G., Foley, B. T. & Smith, T. J. Genetic diversity within the botulinum neurotoxin-producing bacteria and their neurotoxins. *Toxicon* **107**, 2–8 (2015).
24. Luquez, C., Raphael, B. H. & Maslanka, S. E. Neurotoxin gene clusters in *Clostridium botulinum* type Ab strains. *Appl. Environ. Microbiol.* **75**, 6094–6101 (2009).
25. Dover, N. *et al.* *Clostridium botulinum* strain Af84 contains three neurotoxin gene clusters: bont/A2, bont/F4 and bont/F5. *PLoS ONE* **8**, e61205 (2013).
26. Franciosa, G., Ferreira, J. L. & Hatheway, C. L. Detection of type A, B, and E botulinum neurotoxin genes in *Clostridium botulinum* and other *Clostridium* species by PCR: evidence of unexpressed type B toxin genes in type A toxigenic organisms. *J. Clin. Microbiol.* **32**, 1911–1917 (1994).
27. Hutson, R. A. *et al.* Genetic characterization of *Clostridium botulinum* type A containing silent type B neurotoxin gene sequences. *J. Biol. Chem.* **271**, 10786–10792 (1996).
28. Dover, N., Barash, J. R. & Arnon, S. S. Identical novel A5(B3') botulinum neurotoxin gene arrangements isolated from widely disparate geographical and patient sources suggest their independent origins. *J. Clin. Microbiol.* **48**, 1989 (2010).
29. Dabritz, H. A. *et al.* Molecular epidemiology of infant botulism in California and elsewhere, 1976–2010. *J. Infect. Dis.* **210**, 1711–1722 (2014).
30. Kakinuma, H., Maruyama, H., Takahashi, H., Yamakawa, K. & Nakamura, S. The first case of type B infant botulism in Japan. *Acta Paediatr. Jpn* **38**, 541–543 (1996).
31. Kozaki, S. *et al.* Characterization of *Clostridium botulinum* type B neurotoxin associated with infant botulism in Japan. *Infect. Immun.* **66**, 4811–4816 (1998).
32. Ihara, H. *et al.* Sequence of the gene for *Clostridium botulinum* type B neurotoxin associated with infant botulism, expression of the C-terminal half of heavy chain and its binding activity. *Biochim. Biophys. Acta* **1625**, 19–26 (2003).
33. Schiavo, G. *et al.* Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin. *Nature* **359**, 832–835 (1992).
34. Rummel, A., Mahrhold, S., Bigalke, H. & Binz, T. The HCC-domain of botulinum neurotoxins A and B exhibits a singular ganglioside binding site displaying serotype specific carbohydrate interaction. *Mol. Microbiol.* **51**, 631–643 (2004).
35. Gu, S. *et al.* Botulinum neurotoxin is shielded by NTNHA in an interlocked complex. *Science* **335**, 977–981 (2012).
36. Lee, K. *et al.* Molecular basis for disruption of E-cadherin adhesion by botulinum neurotoxin A complex. *Science* **344**, 1405–1410 (2014).
37. Lee, K. *et al.* Structure of a bimodular botulinum neurotoxin complex provides insights into its oral toxicity. *PLoS Pathog.* **9**, e1003690 (2013).
38. Sugawara, Y. *et al.* Botulinum hemagglutinin disrupts the intercellular epithelial barrier by directly binding E-cadherin. *J. Cell Biol.* **189**, 691–700 (2010).
39. Rossetto, O. *et al.* SNARE motif and neurotoxins. *Nature* **372**, 415–416 (1994).
40. Yamamoto, H. *et al.* Specificity of botulinum protease for human VAMP family proteins. *Microbiol. Immunol.* **56**, 245–253 (2012).
41. Chaddock, J. A. *et al.* Expression and purification of catalytically active, non-toxic endopeptidase derivatives of *Clostridium botulinum* toxin type A. *Protein Expr. Purif.* **25**, 219–228 (2002).
42. Popp, M. W., Antos, J. M., Grotenbreg, G. M., Spooner, E. & Ploegh, H. L. Sortagging: a versatile method for protein labeling. *Nat. Chem. Biol.* **3**, 707–708 (2007).
43. McCluskey, A. J. & Collier, R. J. Receptor-directed chimeric toxins created by sortase-mediated protein fusion. *Mol. Cancer Ther.* **12**, 2273–2281 (2013).
44. Aoki, K. R. A comparison of the safety margins of botulinum neurotoxin serotypes A, B, and F in mice. *Toxicon* **39**, 1815–1820 (2001).
45. Binz, T., Bade, S., Rummel, A., Kollwe, A. & Alves, J. Arg(362) and Tyr(365) of the botulinum neurotoxin type A light chain are involved in transition state stabilization. *Biochemistry* **41**, 1717–1723 (2002).
46. Pier, C. L. *et al.* Recombinant holotoxoid vaccine against botulism. *Infect. Immun.* **76**, 437–442 (2008).
47. Steegmaier, M., Klumperman, J., Foletti, D. L., Yoo, J. S. & Scheller, R. H. Vesicle-associated membrane protein 4 is implicated in trans-Golgi network vesicle trafficking. *Mol. Biol. Cell* **10**, 1957–1972 (1999).
48. Brandhorst, D. *et al.* Homotypic fusion of early endosomes: SNAREs do not determine fusion specificity. *Proc. Natl Acad. Sci. USA* **103**, 2701–2706 (2006).
49. Daste, F., Galli, T. & Tareste, D. Structure and function of longin SNAREs. *J. Cell Sci.* **128**, 4263–4272 (2015).
50. Raingo, J. *et al.* VAMP4 directs synaptic vesicles to a pool that selectively maintains asynchronous neurotransmission. *Nat. Neurosci.* **15**, 738–745 (2012).
51. Cocucci, E., Racchetti, G., Rupnik, M. & Meldolesi, J. The regulated exocytosis of enlargosomes is mediated by a SNARE machinery that includes VAMP4. *J. Cell Sci.* **121**, 2983–2991 (2008).
52. Nicholson-Fish, J. C., Kokotos, A. C., Gillingwater, T. H., Smillie, K. J. & Cousin, M. A. VAMP4 is an essential cargo molecule for activity-dependent bulk endocytosis. *Neuron* **88**, 973–984 (2015).
53. Cooper, A. A. *et al.* Alpha-synuclein blocks ER-Golgi traffic and Rab1 rescues neuron loss in Parkinson's models. *Science* **313**, 324–328 (2006).
54. Thayanidhi, N. *et al.* Alpha-synuclein delays endoplasmic reticulum (ER)-to-Golgi transport in mammalian cells by antagonizing ER/Golgi SNAREs. *Mol. Biol. Cell* **21**, 1850–1863 (2010).
55. Zeng, Q. *et al.* A novel synaptobrevin/VAMP homologous protein (VAMP5) is increased during *in vitro* myogenesis and present in the plasma membrane. *Mol. Biol. Cell* **9**, 2423–2437 (1998).
56. Krzewski, K., Gil-Krzewska, A., Watts, J., Stern, J. N. & Strominger, J. L. VAMP4- and VAMP7-expressing vesicles are both required for cytotoxic granule exocytosis in NK cells. *Eur. J. Immunol.* **41**, 3323–3329 (2011).
57. Peng, L., Tepp, W. H., Johnson, E. A. & Dong, M. Botulinum neurotoxin D uses synaptic vesicle protein SV2 and gangliosides as receptors. *PLoS Pathog.* **7**, e1002008 (2011).

### Acknowledgements

We thank members of the Dong and Stenmark labs for discussion, suggestions and comments. We would like to acknowledge Daniel Nilsson for his technical assistance. We thank Edwin R. Chapman (Univ. of Wisconsin—Madison), Shashi Sharma (FDA), Chuan Hu (University of Louisville), Bradley L. Pentelute (MIT) and Jesse C. Hay (University of Montana) for providing antibodies, cDNAs and other reagents. This study was supported in part by NIH 1R01NS080833 (to M.D.), the Swedish Research Council (2014-5667), the Wenner-Gren Foundation, and the Swedish Cancer Society (to P.S.). M.D. also holds the Investigator in the Pathogenesis of Infectious Disease award from the Burroughs Wellcome Fund.

### Author contributions

P.S. identified BoNT/X and analysed its gene cluster. P.S. and D.L. carried out the bioinformatics analysis. P.S., G.M., M.M.-C. and L.H. purified the full-length BoNT/X<sub>RY</sub>. P.S., S.Z. and M.D. designed and performed all other experiments and analysed data. J.Z. performed the DAS assay. Y.S. and S.-I.M. synthesized the cDNA for X-LC-H<sub>N</sub> of BoNT/X and helped with protein purification. M.D. and P.S. wrote the manuscript, with contributions from all other authors.

### Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** Boston Children's Hospital has filed a provisional patent application (U.S. 62/360,239) for medical use of BoNT/X, with P.S., S.Z. and M.D. as inventors.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Zhang, S. *et al.* Identification and characterization of a novel botulinum neurotoxin. *Nat. Commun.* **8**, 14130 doi: 10.1038/ncomms14130 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017