

Article

The Light Chain Domain and Especially the C-Terminus of Receptor-Binding Domain of the Botulinum Neurotoxin (BoNT) Are the Hotspots for Amino Acid Variability and Toxin Type Diversity

Renmao Tian ¹, Melissa Widel ¹ and Behzad Imanian ^{1,2,*} ¹ Institute for Food Safety and Health, Illinois Institute of Technology, Bedford Park, IL 60501, USA² Food Science and Nutrition Department, Illinois Institute of Technology, Chicago, IL 60616, USA

* Correspondence: bimanian@iit.edu

Abstract: Botulinum neurotoxins (BoNT) are the most potent toxins in the world. They are produced by a few dozens of strains within several clostridial species. The toxin that they produce can cause botulism, a flaccid paralysis in humans and other animals. With seven established serologically different types and over 40 subtypes, BoNTs are among the most diverse known toxins. The toxin, its structure, its function and its physiological effects on the neural cell and animal hosts along with its diversity have been the subjects of numerous studies. However, many gaps remain in our knowledge about the BoNT toxin and the species that produce them. One of these gaps involves the distribution and extent of variability along the full length of the gene and the protein as well as its domains and subdomains. In this study, we performed an extensive analysis of all of the available 143 unique BoNT-encoding genes and their products, and we investigated their diversity and evolution. Our results indicate that while the nucleotide variability is almost uniformly distributed along the entire length of the gene, the amino acid variability is not. We found that most of the differences were concentrated along the protein's light chain (LC) domain and especially, the C-terminus of the receptor-binding domain (H_{CC}). These two regions of the protein are thus identified as the main source of the toxin type differentiation, and consequently, this toxin's versatility to bind different receptors and their isoforms and act upon different substrates, thus infecting different hosts.

Keywords: botulinum neurotoxins; *Clostridium botulinum*; gene diversity; receptor-binding domain



Citation: Tian, R.; Widel, M.; Imanian, B. The Light Chain Domain and Especially the C-Terminus of Receptor-Binding Domain of the Botulinum Neurotoxin (BoNT) Are the Hotspots for Amino Acid Variability and Toxin Type Diversity. *Genes* **2022**, *13*, 1915. <https://doi.org/10.3390/genes13101915>

Academic Editor: Silvia Turroni

Received: 26 September 2022

Accepted: 18 October 2022

Published: 21 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Toxins are organic or naturally occurring poisons, produced by a variety of organisms including bacteria in the form of peptides, single or conjugated proteins with adverse effects on the target cells and organisms. More than any other bacterial class, Clostridia, with over a dozen toxin-producing species especially within the Genus *Clostridium*, makes the highest and the most diverse number of toxins including alpha, beta, gamma, delta, epsilon, iota, zeta, eta, theta, TcdA and TcdB, tetanus and botulinum [1,2]. Each of these toxins is specialized in targeting a specific type of cell or tissue such as red blood cells, the liver or the nervous system in a wide ranging group of animals such as waterfowl, wild birds, cattle, horses, poultry, mink and humans [3].

A subset of the above-mentioned toxins are bacterial neurotoxins that can damage or destruct nerve cells/tissues. Neurotoxins often function through inhibiting the neuron control through disrupting either the ion concentrations across the cell membrane or the communication between the neurons across a synapse [4,5]. The common effects of neurotoxin exposure in humans include widespread central nervous system damage that results in epilepsy, dementia [6], intellectual disability [7] or persistent memory impairments [8]. Different neurotoxins have different potencies, and the botulinum neurotoxin (BoNT)

is the most potent/lethal known toxin, produced by several Gram-positive, anaerobic and spore-forming bacteria, mostly in *Clostridium* genus including *C. botulinum*, *C. baratii*, *C. butyricum* and *C. argentinense* [9–12].

In vivo, the toxic BoNT proteins, like other neurotoxins, form a toxin complex (TC) with a few non-toxic proteins [13], whose genes are in a cluster adjacent to the *bont* gene, located in the bacterial chromosome or extrachromosomal elements such as plasmids. These accessory genes invariably include a non-toxin/non-hemagglutinin (*ntnh*) gene along with either hemagglutinin (*ha*) genes (*ha70*, *ha17* and *ha33*) in some genomes or *orfX* genes (*orfX1*, *orfX2* and *orfX3*) in others. While little is known about the functions of ORFX1-3, a few studies suggest that the products of the *ha* genes are involved in providing help for docking of the BoNT protein to the receptors, found on the lumen of the small intestine [14] and also for the protection for the BoNT protein in its journey from the gastrointestinal tract to the circulatory system [15,16].

The BoNT protein is initially produced as a single soluble polypeptide chain, weighing 150 kDa. This precursor protein is not toxic to the neural tissues, until later, when it is cleaved by a protease that is either produced by the bacterium or made within the target cell/tissue itself [17]. The cleaving of the precursor protein generates two polypeptide chains that are linked by a functionally critical disulfide bond, and it results in the activation of the neurotoxin. One of the polypeptide chains, the light chain (LC, 50 kDa), has a zinc endopeptidase activity and it is, thus, the catalytic domain [18–20]. The function of the LC is to cleave a target protein in the neural cells such as the Soluble N-ethylmaleimide-Sensitive Factor Attachment Proteins (SNAP) Receptor (SNARE), with three SNARE proteins identified so far: Vesicle-Associated Membrane Protein (VAMP)/synaptobrevin, SNAP-25 and syntaxin [20]. The other polypeptide, the heavy chain (HC, 100 kDa), is composed of two functional domains: an N-terminus (H_N) translocation domain and a C-terminus (H_C) receptor-binding domain [19,21,22]. The H_C domain of the heavy chain can be further divided into two subdomains: the N-terminus (H_{CN}) and the C-terminus (H_{CC}). The three domains of BoNT work together to intoxicate the host. The specificity of a neurotoxin is ensured through its receptor-binding domain that recognizes specific receptor(s) in/on and delivers their effects to the specific neural cells. In the case of H_C , the specific recognition of neuronal cells occurs by a double anchorage to the receptors: first, to a polysialoganglioside (PSG) receptor, and then, this is followed by binding to a protein like synaptotagmin (Syt) (e.g., target of BoNT/B and BoNT/G) in two known isoforms (I–II) or in a Synaptic Vesicle protein (SV) receptor (e.g., target for BoNT/A and BoNT/E), with three identified isoforms (A–C) [23–27]. The special double receptor mechanism guarantees a specific binding to no other target except for the neural cell. Then, BoNT is internalized through its H_N translocation domain and the clathrin–dynamamin-mediated endocytosis of the recycling neurosecretion vesicles [28]. The endocytosis of BoNT is followed by the acidification of the lumen of the vesicle and the conformational change of BoNT. Finally, the LC zinc endopeptidase cleaves the SNARE proteins, which are involved in neurotransmitter release, thus, blocking the acetylcholine release, leading to a flaccid paralysis of the inflicted organism [20,29].

BoNTs are extraordinarily diverse and so are the species that produce them. Many studies have investigated the diversity of the BoNT-producing bacteria, employing serological, physiological, biochemical, structural and genetic and genomic methods and analyses. For a long time, producing the botulinum neurotoxin was the only criterion to classify a species as *C. botulinum*. Based on several phenotypic and biochemical differences, four different groups (I–IV) were recognized within the BoNT-producing '*C. botulinum*' that are now, in the light of recent molecular interrogations, acknowledged to be at least four different taxonomic species [10]. As for the BoNT diversity, there are currently seven or possibly eight serologically different established types of proteins/toxins, designated with letters A–G (amino acid differences ranging from 37.2% to 69.6%) [30], with over 40 BoNT subtypes (e.g., A1, B1, E12), and also chimeric forms with a BoNT type appearing in combinatory forms with another BoNT type or types, usually leading to the production

of more than one toxin type and often in different ratios (e.g., BoNT/CD, /DC, /FA, /A2F4, /A2F5, /A2B6, /B5A4) [10,15,31,32]. Recent genetic studies have uncovered new types of BoNT, for example, BoNT/H [33], later characterized as a chimeric BoNT/FA [20] and as BoNT/HA [34], and BoNT-like proteins, BoNT/Wo, BoNT/J [35] and BoNT/X [36]. Types A, B, E and F can cause botulism in humans [37], while BoNTs C and D intoxicate animals including birds and mammals. [3,38]. Not much is known about the effects of BoNT/G or the newly discovered BoNT types (H, J and X) on humans or animals. Interestingly, BoNTs are found to have also valuable cosmetic and medical uses [39,40].

Beyond satisfying the academic curiosity and interest, studying the BoNT diversity provides real-life, practical, medical and health benefits since the big or small differences at the nucleotide/amino acid levels are not trivial. These differences determine the toxin types and subtypes, influencing the final folding of the proteins, and thus, their functions and efficiencies at every step from the formation of TC and receptor binding to the BoNT internalization within the neural cell and the endopeptidase activity of LC domain. They can, therefore, affect intoxication and the kind and intensity of botulism symptoms in vivo among other things [20].

A mplified fragment length polymorphism (AFLP) analyses, gene (e.g., 16S rRNA, *bont*) and whole genome sequencing in conjunction with various phylogenetic/phylogenomic and other molecular techniques and analyses have been employed to tackle the important question of the diversity of BoNT-producing species and strains, *bont* genes and BoNT proteins. These approaches, strengthened by ecological and biogeographical studies, have begun to (1) help researchers to revise and improve the confusing historical taxonomy and evolutionary relationships between the BoNT-producing species/strains; (2) identify the location of the *bont* gene and the *bont* gene clusters (chromosomal, extrachromosomal); (3) unravel the *bont* gene and protein evolution; (4) discover new BoNT subtypes and occasionally new types; (5) provide new insights into BoNT pathogenesis [41–45]. Despite these excellent studies, many gaps remain in our knowledge about the diversity and evolution of *bont* genes and proteins. In order to address one of these gaps and to provide a more granular insight into the BoNT diversity, we investigated the distribution of this diversity among different domains and subdomains of all the available BoNT proteins, including the within types and between types and spatially along the 3D structure of BoNT, with the null assumption that the amino acid variations along different domains/subdomains have a uniform distribution. The results of our study clearly show that it is not the case, and the LC domain, and especially, the H_{CC} subdomain of BoNT are the hotspots and the sites for the highest number of amino acid variations.

2. Materials and Methods

2.1. Collection of Bont Gene Sequences

Sequences of the genes for botulinum neurotoxin (*bont*) were collected through a custom pipeline. Seed sequences of the genes were obtained (81 sequences in total) from NCBI RefSeq database <https://www.ncbi.nlm.nih.gov/protein> (accessed on 15 November 2021) by searching the protein names. The sequences were further processed to manually remove the incorrectly annotated sequences (e.g., based on the sequence description). The seed sequences were then clustered at 95% identity to acquire representative sequences (28 in total) using CD-HIT (version 4.8.1) [46], which were then used as queries for tBLASTn against NCBI NT database. The output was parsed using the Python package Bio.Blast.NCBIXML to screen for qualified hits with identity > 30% and alignment coverage > 80%. The accession IDs of qualified hits were used to retrieve the nucleotide coding sequences using the Python package Bio.efetch, and the aligned regions were extracted according to the alignment coordinates. The downloaded sequences were then clustered at 100% identity using CD-HIT to remove the duplicates. Because botulinum and tetanus neurotoxin genes share ~30% sequence similarity, the sequences of *bont* with taxonomic assignment of *Clostridium tenani* in the sequence description were removed (and so, 143 sequences remained).

2.2. Assignment of Toxin Types

The above seed BoNT protein sequences (81) were entered into a custom database, against which the collected nucleotide coding sequences (143) were searched using BLASTx (version 2.12.0) [47] with a query coverage cutoff of 90% and an E value cutoff of 10^{-5} . The BLASTx output was then parsed, and the best hits with an identity > 90% were selected for the toxin type assignment. For the subtype analysis, the protein sequences were clustered using CD-HIT (version 4.8.1) at 97% identity, and each cluster was assigned a unique cluster ID of its own type. We then investigated the sequence variation within each cluster. Note that the sequences within a type do not exactly correspond to the subtypes defined in other studies, which may include sequences with as little as 1.5% dissimilarity to as high as 36.2% sequence variation [10,32,48,49].

2.3. Alignment and Sequence Diversity Analysis

All of the nucleotide coding sequences were first translated to protein sequences using the Python package Bio.SeqIO with the genetic codon table 11. Any pseudogenes with premature stop codon were removed. The protein sequences were then aligned using MUSCLE (version 3.8.1551) [50]. The nucleotide coding sequences were also aligned using PAL2NAL (version 14) [51] with the protein sequence alignment being used as a template. The protein or nucleotide sequence multiple alignment output was imported with the Python package Bio.AlignIO, and it was converted into a 2D array with the Python package Numpy. In order to construct a sequence diversity index, we counted the number of unique amino acids or nucleotides of each column in the protein and gene multiple alignment files, respectively. To demonstrate the sequence variation in each domain, the sequence diversity indexes of all of the positions were calculated and mapped onto the nucleotide sequence of BoNT/A of *C. botulinum* 62A; a commonly used reference in *C. botulinum* studies. The same was performed for each type of BoNT gene to investigate the sequence variation profiles.

2.4. Visualizing BoNT Sequence Variation on a BoNT 3D Structure

In order to visualize the sequence variation on a 3D structure of a BoNT protein, we mapped the diversity index at each position onto PDB data. A reference protein crystal structure PDB data (3BTA, serotype A) which was highly similar (>99.9%) to the reference protein sequences in the diversity analysis was downloaded from the Protein Databank (PDB) database. The amino acids and corresponding coordinates were retrieved from the PDB file using the Python package Bio.PDB. The retrieved protein sequences were aligned to the reference protein sequence in the diversity analysis, and then, the diversity indexes were mapped onto each amino acid position in the PDB data by replacing the B factor value with the corresponding diversity index. The sequence diversity was then visualized by importing the revised PDB file into PyMOL (version 2.5.2) (Schrödinger, LLC, New York, NY, USA) and executing 'spectrum b, blue_white_red' command. The same procedure was performed for each toxin type separately to demonstrate within-type sequence variation on the 3D structure for each type.

2.5. Visualizing the BoNT Amino Acids Interacting with Ganglioside, SV2 and Monoclonal Antibody (mAb) CR1 on a Reference 3D Structure

The PDB data of the reference BoNT protein crystal structure combined with ganglioside, SV2 and monoclonal antibody (mAb) were searched and downloaded from the PDB database. Amino acids from the BoNT interacting with the ligands were identified using iCn3D [52] with the function icn3dnode. The protein sequence was aligned with the one from the analysis of sequence variation in 3D structure and the corresponding interacting amino acids were labelled in the visualization.

3. Results

3.1. Gene Sequence Collection

In total, nucleotide coding sequences of 143 unique *bont* genes were collected from the NCBI NT database. One of these sequences was identified as a pseudogene due to it having a premature stop codon, and it was removed from the analysis. We were able to assign a toxin type to 131 of the collected sequences (Figure S1). There were four dominant types: A (29 genes), B (34 genes), E (35 genes) and F (17 genes). Due to the low number of retrieved sequences (<7), the types C, D, G, H and X were not considered for further analysis.

3.2. The LC Domain and Especially the C-Terminus of the Receptor-Binding Domain (H_{CC}) of BoNT Contain More Amino Acid Diversity Than Other Domains Do

Our nucleotide and amino acid diversity index analyses of the *bont* genes and the BoNT proteins revealed that the distribution of amino acid variations is not uniform along the protein and its various domains and subdomains. We found that the light chain domain (LC, 4–409), and especially, the C-terminus of the receptor-binding domain (H_{CC} , 1088–1293 aa) contained more amino acid diversity, 24% and 45% more than that in the HN domain, respectively, with the peak occurring along the end of the H_{CC} subdomain (Figure 1).

3.3. Inter-Type Amino Acid Variation Was Much Higher Than Intra-Type Variation in BoNT Proteins

We further compared the protein sequences within each BoNT type, for which a sufficient number of sequences were available (A, B, E and F). A close inspection of the diversity index within each BoNT type revealed that in each position in the alignment file for types A, B and E, there were on average only one to two amino acid types or gaps. The highest variability was found within type F (2.5 on average for some positions, Figure 2). We also conducted similar calculations at the subtype level for the clusters within each type, and we found that the protein sequences within each cluster were even more conserved, with the sequence diversity index being < 1.3 in all of the subtypes (Figure S2). By contrast and not surprisingly, the between-types comparisons showed more amino acid variability for the BoNT proteins (with the diversity index value > 3.9 on moving average), reaching a maximum of eight different amino acids per position along the H_{CC} domain (Figure 1A). To recap, the protein sequences were found to be highly conserved within each BoNT type, and they were even more conserved within the subtypes, and highly variable between the types, with the highest amino acid variability being observed along the H_{CC} subdomain.

3.4. The Sequence Variations Are Mainly Concentrated at the Surface of the Protein's 3D Structure Where the LC Domain and H_{CC} Subdomain Are Located

In order to examine the spatial location of the highly variable amino acids, we mapped the sequence diversity profile (the actual diversity index, rather than the moving average values) on the 3D structure of the BoNT protein (3BTA, serotype A). We found that the light chain (LC, peptidase domain) and the H_{CC} region of the receptor-binding domain had more amino acid diversity than the other domains and regions of the protein did. These two domains are located near the ends of the protein (N-terminus and C-terminus, respectively). In contrast, the N-terminus of the receptor-binding domain (H_{CN}) and the translocation domain (H_N) that harbored less amino acid diversity were located near the core of the folded protein, away from the BoNT protein's termini. In addition, within the H_{CC} sub-domain, the highly variable amino acids were found to be located on the exterior or surface of the folded protein, while the more conserved positions were embedded in the interior or core of the domain (Figure 3). In summary, the amino acids on the surface or exterior of the folded BoNT protein showed more diversity than those that are positioned interiorly with the maximum diversity values being observed along the surface amino acids of the LC domain, and especially, on the H_{CC} sub-domain.

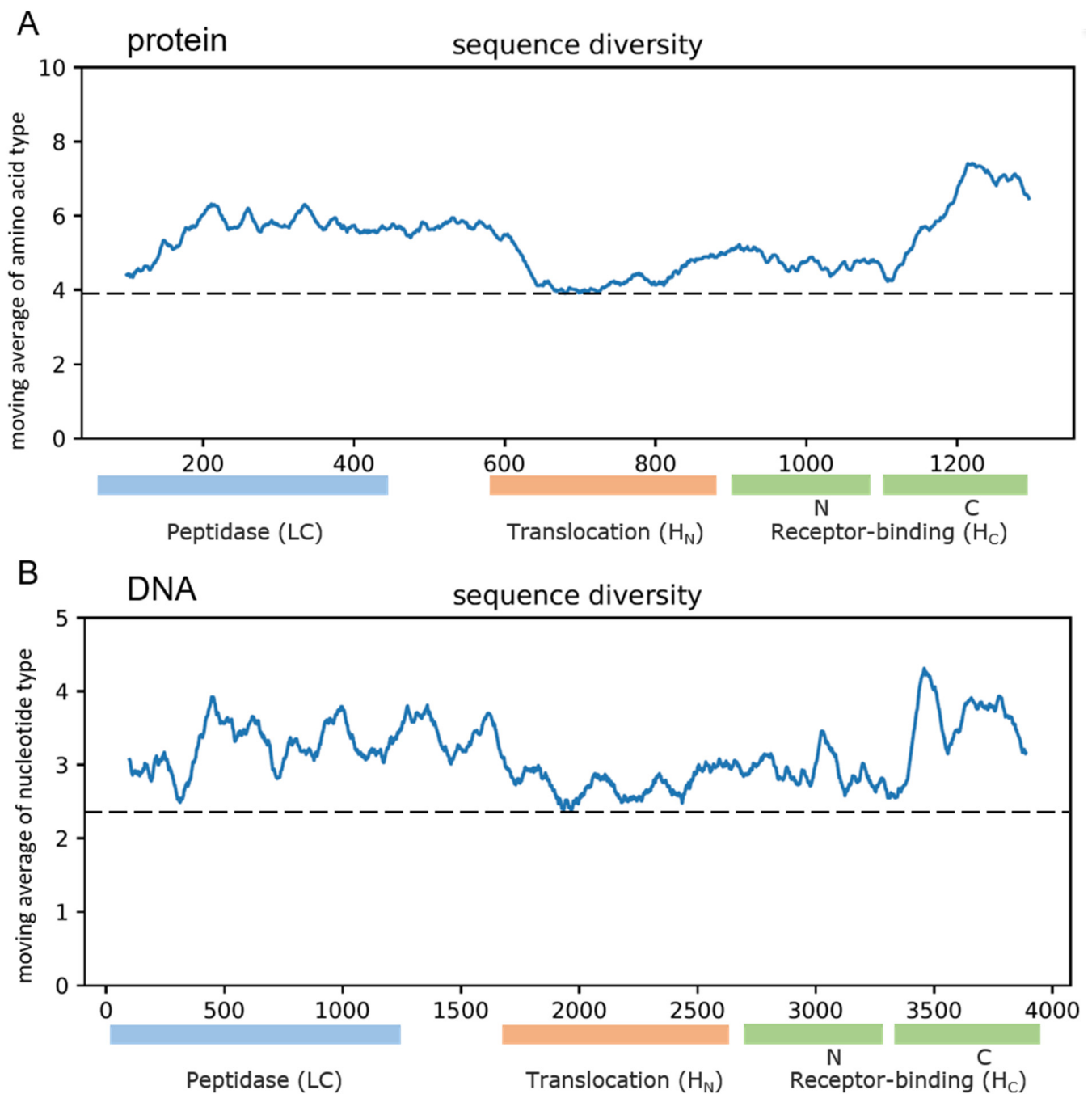


Figure 1. The amino acid and nucleotide diversity distribution of different regions of botulinum neurotoxin (A) protein and (B) gene, respectively. The diversity index (produced by counting the number of unique amino acids or nucleotides in each column in the multiple alignment files) of all of the positions were calculated using *C. botulinum* 62A BoNT gene as a reference. The moving average ($n = 100$) of the diversity index was used to generate an overview of the sequence variation. The dashed black horizontal lines demarcate the lowest diversity value in the graph and are used only as visual aid.

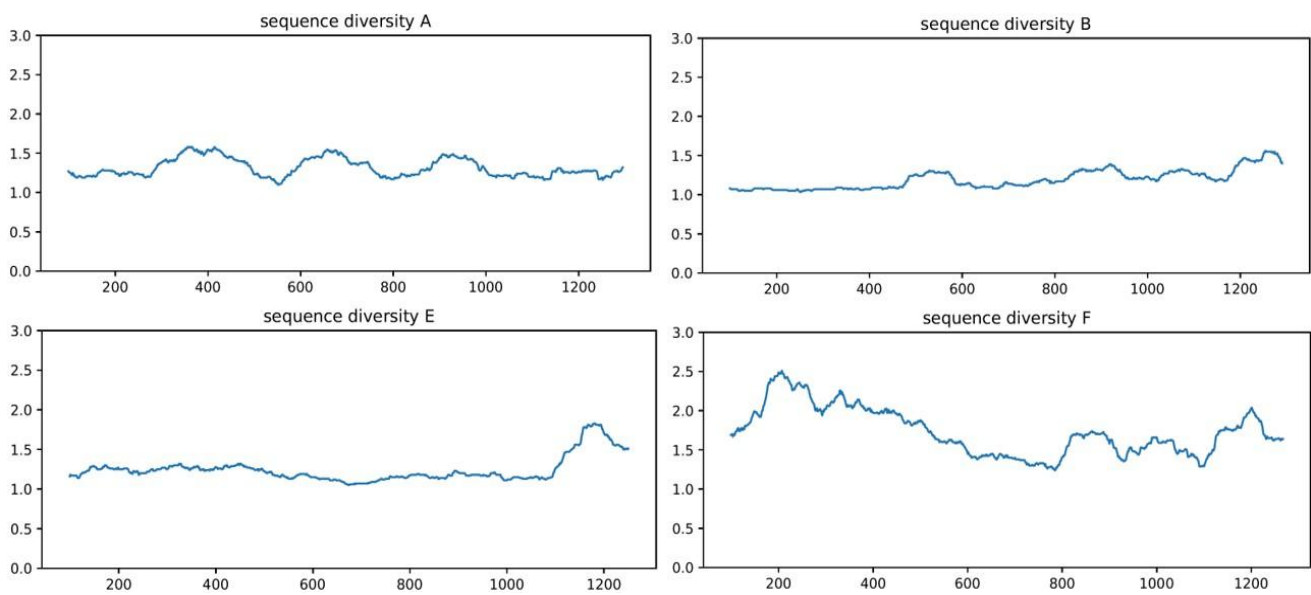


Figure 2. Protein sequence diversity within each type of BoNT. Protein sequences of each type were aligned and the diversity index (produced by counting the number of unique amino acids or nucleotides in each column in the multiple alignment files) of all of the positions were calculated and mapped onto a representative BoNT gene of each type. The moving average ($n = 100$) of the diversity index was shown for an overview of the sequence variation.

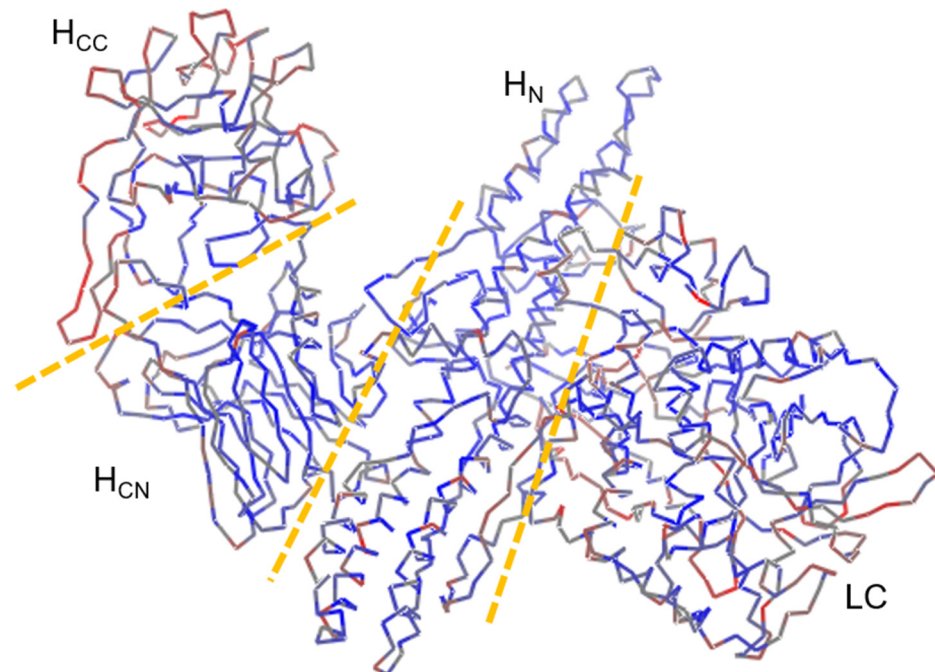


Figure 3. Amino acid sequence diversity of BoNT proteins mapped on its 3D structure (3BTA, serotype A). The diversity index, the actual counts of the number of unique amino acids or nucleotides of each column in the multiple alignment files, of all the positions were mapped on the 3D structure of a reference gene, 3BTA, downloaded from the PDB database. The diversity index is indicated by different shades of a color from blue (low) to medium (grey) to red (high). LC: Light chain. H_N: Heavy chain N-terminus. H_{CN}: N-terminus of heavy chain C-terminus. H_{CC}: C-terminus of heavy chain C-terminus.

In order to further explore the function of the highly diverse amino acids of the H_{CC} sub-domain of the receptor-binding domain of the BoNT protein, we identified the sites

that interacted with the receptors, a ganglioside and a synaptic vesicle (SV2), and also with a SNAP-25 substrate and a mAb CR1 against the BoNT toxin. Our examinations indicated that the highly variable sites of H_{CC} were mostly made of the same residues that interacted with the two receptors (Figure 4A,B). The different isoforms of the ganglioside had highly consistent interacting sites in BoNT (Figure S3). Interestingly, the interacting sites with the mAb CR1 against the toxin and those interacting with the SV2 receptor clearly showed a large overlap (Figure 4B,D), hinting at the possible modes of function for the antitoxin quality of mAb CR1 by competing with or blocking the sites that bind the SV2 receptor.

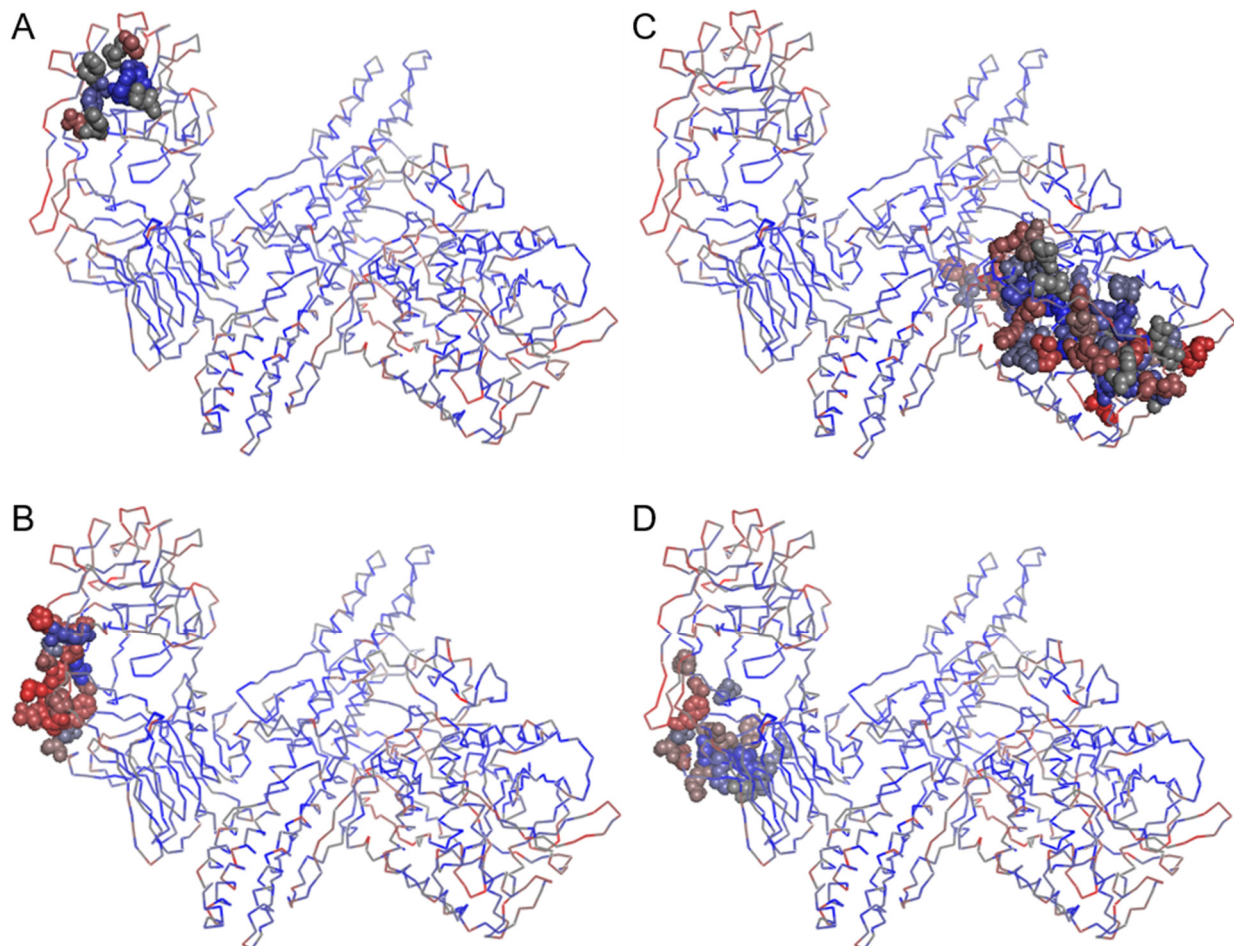


Figure 4. Three-dimensional structure of BoNT protein and the interacting sites with (A) ganglioside, (B) SV2, (C) SNAP-25 substrate and (D) monoclonal antibody CR1 against BoNT toxin. The diversity index of each position were mapped onto the 3D structure indicated by the color (from blue to grey to red, low to medium to high). The interacting amino acids were highlighted as spheres with color indicating the diversity index.

4. Discussion

With a few exceptions, all of the BoNT-producing clostridia were traditionally classified as *C. botulinum*, whose members then were categorized into several groups based on a few physiological characteristics (e.g., proteolytic differences) [53,54], thus leading to a great deal of confusion about the taxonomy, phylogeny and diversity of this class of bacteria. Recent molecular studies have begun to address and resolve this confusion. Characterizing the diversity of the neurotoxin genes (*bont*) and their products (BoNT) is particularly important because the differences between the BoNT types and the subtypes determine, or at least influence, the functional speed and efficiency of the toxin, the class or type of receptors and the hosts, intoxication, botulism symptoms, and so on [20]. Many of the previous studies have focused on characterizing and enumerating the toxin types/subtypes

and cataloguing and quantifying the sequence similarities and differences. The diversity of the *bont* genes and their products have also been explored by employing a phylogenetic analysis of the *bont* genes, the BoNT proteins with or without the toxin accessory genes and in the *bont*-encoding clostridial species/strains [12,55]. However, few studies have focused on the distribution of the nucleotide and amino acid diversity along the entire *bont* gene, its product and its domains/subdomains [10] across the primary sequence as well as spatially along the 3D structure of the folded protein in the clostridial species/strains.

Here, we examined this distribution, and we found that the amino acid variations generally increased from the within-subtypes (Figure S2) to the within-types (Figure 2), and from there to the between-types of toxins (Figure 1), where the variations peaked. These variations at different levels might provide clues as to why different types or subtypes of the toxin function differently with different degrees of efficiencies. As an example, the LC domain of the BoNT/F contains more variations than the same domain does in the other BoNT types (Figure 2), and this could provide one possible explanation for the differential catalytic activities of the LC domain that has been reported in different BoNT/F subtypes [32].

We also found that the nucleotide variations had nearly uniform distribution along the entire length of all of the available *bont* genes. However, the amino acid variations were not uniform along the length of the protein (BoNT) and its various domains and subdomains. This implies that while all of the types of mutations might have occurred uniformly along the gene sequence, the non-synonymous or missense mutations that change the encoded amino acid had not. In fact, the LC domain and especially the C-terminus of the receptor-binding subdomain (H_{CC}) in the BoNT proteins harbored more diversity (contained more amino acid types per position) than the other domains/subdomains did. In other words, the LC and particularly H_{CC} were the main contributors to the total diversity found in the BoNT proteins, and thus, they are mainly responsible for BoNT type differentiation. Further studies are needed to investigate whether these two domains are under selective pressure, and diversifying these two domains is accompanied by any improved fitness, and perhaps, this may contribute to the speciation of the clostridial species.

Mapping the amino acid type diversity of the BoNT proteins on a 3D reference structure provided a quick visual reference for the diversity of the BoNT proteins, demonstrating the same pattern, but here spatially: there was more amino acid diversity within the LC domain, and especially, in the H_{CC} subdomain of the protein (Figure 3). The expected consequences of this diversity especially in these two regions of the protein have already been documented. For example, it has been shown that the LC of different BoNTs in different species/strains cleave different SNARE proteins with many cleavage sites having been already identified. For example, BoNT/B, BoNT/D, BoNT/F and BoNT/G cut VAMP, while BoNT/A and BoNT/E cleave SNAP-25, and BoNT/C cleaves SNAP-25 as well as syntaxin (Stx) [20]. More importantly, we also found, and here we show, that many of the amino acids within the receptor-binding domain that interacted with the ganglioside, the SV2 receptors and the SNAP-25 substrate on the target cells were among the highly variable amino acids (Figure 4A–C). It has been demonstrated that BoNT acts on different hosts and its targets have usually more than one isoform: the ganglioside receptors have several (e.g., GM1, GD1a, GD1b and GT1b) [31,56]; the synaptotagmin (SytII-II) has two known isoforms; and the SV2 receptors have at least three (SV2A–C) [31,57]. Our results once more show and underline that the functional diversity of the BoNT proteins can be linked, at least in part, to the observed nucleotide variability of the *bont* genes and the amino acid type diversity of their products in different clostridial species/strains. This underlying diversity is at the heart of BoNT's versatility in infecting different hosts, targeting different receptors, acting on different substrates with different speed and efficiencies and causing different symptoms with different intensities. Studies in avian and mammalian botulism, for example, have confirmed that BoNT/C and BoNT/D are specific to animal rather than human hosts [3,38], whereas BoNT/A, BoNT/B, BoNT/E and BoNT/F cause botulism in human rather than animal hosts. The sequence and structure of SV2 and its various

isoforms in animals are different enough from those in humans so that the BoNT/C and BoNT/D could bind the receptor(s) in the target cell and/or cleave the substrate in animals, but rarely in humans.

Some of the recent studies have shed light on the possible underlying mechanisms behind this extraordinary diversity. In addition to the slow random mutations, researchers have discovered extensive evidence of multiple horizontal gene transfer (HGT) events of the *bont* gene or the *bont* gene cluster between the closely or distantly related strains and species and the footprints of homologous recombination [42,54,57]. While it is clear that BoNT diversity plays an important role in the collective versatility of the toxin in binding a variety of receptors that are found in a number of different hosts and in the success of the BoNT-producing species, it is not clear whether there is a selective pressure driving the diversification of BoNT and its domains and subdomains. It is conceivable that host specialization by the Clostridial species is one of the possible consequences of this diversification, a specialization that could lessen the costly competitions between the closely related bacterial species in the same environment. However, further studies are needed to validate this speculation.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13101915/s1>. Figure S1: Number of *bont* genes collected for each type. The types C, D, G, H and X had <7 sequences and were thus not considered for further comparison analysis; Figure S2: The diversity index at subtype level. Subtypes were generated by protein sequence clustering at 97%. (A) shows the number of protein sequences of each subtype. (B) shows the diversity index, counting the number of unique amino acids of each column in the multiple alignment, of each subtype. The moving average ($n = 100$) of the diversity index were shown for an overview of the sequence variation; Figure S3: Interacting sites of BoNT with ganglioside receptors. The highlighted spheres corresponds to the interacting sites with ganglioside (A) GD1a (in 5PTC), (B) GT1B (in 2VU9), (C) GD1a (in 7QPT). The interacting sites were mapped onto the BoNT structure from PDB database (3BTA). The amino acid residue numbers were displayed.

Author Contributions: B.I. and R.T. conceptualized and designed the project; R.T. and B.I. performed experiments; B.I., R.T. and M.W. provided comparative analyses of genes/proteins. B.I. and R.T. wrote the first draft. All authors have read and agreed to the published version of the manuscript.

Funding: This publication is supported by the Food and Drug Administration (FDA) of the U.S. Department of Health and Human Services (HHS) (Grant No. 5U19FD005322) as part of an award totaling \$4,148,332 with 0% financed with nongovernmental sources. The funding body did not play a role in the design of the study and collection, analysis and interpretation of data and in writing the manuscript. The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the official views of, nor endorsement by, the FDA, HHS, U.S. Government or Illinois Institute of Technology. For more information, please visit <https://www.fda.gov/>.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors acknowledge the Food and Drug Administration (FDA) of the U.S. Department of Health and Human Services (HHS) for the support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Baldassi, L. Clostridial Toxins: Potent Poisons, Potent Medicines. *J. Venom. Anim. Toxins Incl. Trop. Dis.* **2005**, *11*, 391–411. [[CrossRef](#)]
2. Popoff, M.R.; Bouvet, P. Clostridial Toxins. *Future Microbiol.* **2009**, *4*, 1021–1064. [[CrossRef](#)] [[PubMed](#)]
3. Anniballi, F.; Fiore, A.; Löfström, C.; Skarin, H.; Auricchio, B.; Woudstra, C.; Bano, L.; Segerman, B.; Koene, M.; Båverud, V.; et al. Management of Animal Botulism Outbreaks: From Clinical Suspicion to Practical Countermeasures to Prevent or Minimize Outbreaks. *Biosecurity Bioterrorism Biodefense Strategy Pract. Sci.* **2013**, *11*, S191–S199. [[CrossRef](#)] [[PubMed](#)]

4. Arnon, S.S.; Schechter, R.; Inglesby, T.V.; Henderson, D.A.; Bartlett, J.G.; Ascher, M.S.; Eitzen, E.; Fine, A.D.; Hauer, J.; Layton, M.; et al. Botulinum Toxin as a Biological Weapon: Medical and Public Health Management. *JAMA* **2001**, *285*, 1059–1070. [[CrossRef](#)]
5. Popoff, M.R.; Poulain, B. Bacterial Toxins and the Nervous System: Neurotoxins and Multipotential Toxins Interacting with Neuronal Cells. *Toxins* **2010**, *2*, 683–737. [[CrossRef](#)] [[PubMed](#)]
6. Nadler, J.V.; Perry, B.W.; Cotman, C.W. Intraventricular Kainic Acid Preferentially Destroys Hippocampal Pyramidal Cells. *Nature* **1978**, *271*, 676–677. [[CrossRef](#)]
7. Olney, J.W. New Insights and New Issues in Developmental Neurotoxicology. *Neurotoxicology* **2002**, *23*, 659–668. [[CrossRef](#)]
8. Jevtovic-Todorovic, V.; Hartman, R.E.; Izumi, Y.; Benshoff, N.D.; Dikranian, K.; Zorumski, C.F.; Olney, J.W.; Wozniak, D.F. Early Exposure to Common Anesthetic Agents Causes Widespread Neurodegeneration in the Developing Rat Brain and Persistent Learning Deficits. *J. Neurosci.* **2003**, *23*, 876–882. [[CrossRef](#)]
9. Horowitz, B.Z. Botulinum Toxin. *Crit. Care Clin.* **2005**, *21*, 825–839. [[CrossRef](#)]
10. Smith, T.J.; Hill, K.K.; Raphael, B.H. Historical and Current Perspectives on Clostridium Botulinum Diversity. *Res. Microbiol.* **2015**, *166*, 290–302. [[CrossRef](#)]
11. Poulain, B.; Popoff, M.R. Why Are Botulinum Neurotoxin-Producing Bacteria So Diverse and Botulinum Neurotoxins So Toxic? *Toxins* **2019**, *11*, 34. [[CrossRef](#)] [[PubMed](#)]
12. Williamson, C.H.D.; Sahl, J.W.; Smith, T.J.; Xie, G.; Foley, B.T.; Smith, L.A.; Fernández, R.A.; Lindström, M.; Korkeala, H.; Keim, P.; et al. Comparative Genomic Analyses Reveal Broad Diversity in Botulinum-Toxin-Producing Clostridia. *BMC Genom.* **2016**, *17*, 180. [[CrossRef](#)] [[PubMed](#)]
13. Fredrick, C.M.; Lin, G.; Johnson, E.A. Regulation of Botulinum Neurotoxin Synthesis and Toxin Complex Formation by Arginine and Glucose in Clostridium Botulinum ATCC 3502. *Appl. Environ. Microbiol.* **2017**, *83*, e00642-17. [[CrossRef](#)] [[PubMed](#)]
14. Hasegawa, K.; Watanabe, T.; Suzuki, T.; Yamano, A.; Oikawa, T.; Sato, Y.; Kouguchi, H.; Yoneyama, T.; Niwa, K.; Ikeda, T.; et al. A Novel Subunit Structure of Clostridium Botulinum Serotype D Toxin Complex with Three Extended Arms. *J. Biol. Chem.* **2007**, *282*, 24777–24783. [[CrossRef](#)] [[PubMed](#)]
15. Carter, A.T.; Peck, M.W. Genomes, Neurotoxins and Biology of Clostridium Botulinum Group I and Group II. *Res. Microbiol.* **2015**, *166*, 303–317. [[CrossRef](#)]
16. Lam, K.-H.; Jin, R. Architecture of the Botulinum Neurotoxin Complex: A Molecular Machine for Protection and Delivery. *Curr. Opin. Struct. Biol.* **2015**, *31*, 89–95. [[CrossRef](#)]
17. Pirazzini, M.; Rossetto, O.; Eleopra, R.; Montecucco, C. Botulinum Neurotoxins: Biology, Pharmacology, and Toxicology. *Pharmacol. Rev.* **2017**, *69*, 200–235. [[CrossRef](#)]
18. Schiavo, G.; Matteoli, M.; Montecucco, C. Neurotoxins Affecting Neuroexocytosis. *Physiol. Rev.* **2000**, *80*, 717–766. [[CrossRef](#)]
19. Rossetto, O.; Montecucco, C. Presynaptic Neurotoxins with Enzymatic Activities. *Handb. Exp. Pharmacol.* **2008**, 129–170. [[CrossRef](#)]
20. Lacy, D.B.; Tepp, W.; Cohen, A.C.; DasGupta, B.R.; Stevens, R.C. Crystal Structure of Botulinum Neurotoxin Type A and Implications for Toxicity. *Nat. Struct. Biol.* **1998**, *5*, 898–902. [[CrossRef](#)]
21. Kumaran, D.; Eswaramoorthy, S.; Furey, W.; Navaza, J.; Sax, M.; Swaminathan, S. Domain Organization in Clostridium Botulinum Neurotoxin Type E Is Unique: Its Implication in Faster Translocation. *J. Mol. Biol.* **2009**, *386*, 233–245. [[CrossRef](#)]
22. Swaminathan, S.; Eswaramoorthy, S. Structural Analysis of the Catalytic and Binding Sites of Clostridium Botulinum Neurotoxin B. *Nat. Struct. Mol. Biol.* **2000**, *7*, 693–699. [[CrossRef](#)]
23. Montecucco, C. How Do Tetanus and Botulinum Toxins Bind to Neuronal Membranes? *Trends Biochem. Sci.* **1986**, *11*, 314–317. [[CrossRef](#)]
24. Rummel, A.; Eichner, T.; Weil, T.; Karnath, T.; Gutcaits, A.; Mahrhold, S.; Sandhoff, K.; Proia, R.L.; Acharya, K.R.; Bigalke, H.; et al. Identification of the Protein Receptor Binding Site of Botulinum Neurotoxins B and G Proves the Double-Receptor Concept. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 359–364. [[CrossRef](#)]
25. Rummel, A. Double Receptor Anchorage of Botulinum Neurotoxins Accounts for Their Exquisite Neurospecificity. *Curr. Top. Microbiol. Immunol.* **2013**, *364*, 61–90. [[CrossRef](#)]
26. Binz, T.; Rummel, A. Cell Entry Strategy of Clostridial Neurotoxins. *J. Neurochem.* **2009**, *109*, 1584–1595. [[CrossRef](#)] [[PubMed](#)]
27. Weisemann, J.; Stern, D.; Mahrhold, S.; Dorner, B.G.; Rummel, A. Botulinum Neurotoxin Serotype A Recognizes Its Protein Receptor SV2 by a Different Mechanism than Botulinum Neurotoxin B Synaptotagmin. *Toxins* **2016**, *8*, 154. [[CrossRef](#)]
28. Harper, C.B.; Martin, S.; Nguyen, T.H.; Daniels, S.J.; Lavidis, N.A.; Popoff, M.R.; Hadzic, G.; Mariana, A.; Chau, N.; McCluskey, A.; et al. Dynamin Inhibition Blocks Botulinum Neurotoxin Type A Endocytosis in Neurons and Delays Botulism. *J. Biol. Chem.* **2011**, *286*, 35966–35976. [[CrossRef](#)]
29. Humeau, Y.; Doussau, F.; Grant, N.J.; Poulain, B. How Botulinum and Tetanus Neurotoxins Block Neurotransmitter Release. *Biochimie* **2000**, *82*, 427–446. [[CrossRef](#)]
30. Hill, K.K.; Smith, T.J. Genetic Diversity within Clostridium Botulinum Serotypes, Botulinum Neurotoxin Gene Clusters and Toxin Subtypes. *Curr. Top. Microbiol. Immunol.* **2013**, *364*, 1–20. [[CrossRef](#)]
31. Davies, J.R.; Liu, S.M.; Acharya, K.R. Variations in the Botulinum Neurotoxin Binding Domain and the Potential for Novel Therapeutics. *Toxins* **2018**, *10*, 421. [[CrossRef](#)] [[PubMed](#)]

32. Kalb, S.R.; Baudys, J.; Rees, J.C.; Smith, T.J.; Smith, L.A.; Helma, C.H.; Hill, K.; Kull, S.; Kirchner, S.; Dorner, M.B.; et al. De Novo Subtype and Strain Identification of Botulinum Neurotoxin Type B through Toxin Proteomics. *Anal. Bioanal. Chem.* **2012**, *403*, 215–226. [[CrossRef](#)]
33. 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.* **2014**, *209*, 192–202. [[CrossRef](#)]
34. Fan, Y.; Barash, J.R.; Conrad, F.; Lou, J.; Tam, C.; Cheng, L.W.; Arnon, S.S.; Marks, J.D. The Novel Clostridial Neurotoxin Produced by Strain IBCA10-7060 Is Immunologically Equivalent to BoNT/HA. *Toxins* **2019**, *12*, 9. [[CrossRef](#)] [[PubMed](#)]
35. Brunt, J.; Carter, A.T.; Stringer, S.C.; Peck, M.W. Identification of a Novel Botulinum Neurotoxin Gene Cluster in Enterococcus. *FEBS Lett.* **2018**, *592*, 310–317. [[CrossRef](#)] [[PubMed](#)]
36. Zhang, S.; Masuyer, G.; Zhang, J.; Shen, Y.; Lundin, D.; Henriksson, L.; Miyashita, S.-I.; Martínez-Carranza, M.; Dong, M.; Stenmark, P. Identification and Characterization of a Novel Botulinum Neurotoxin. *Nat. Commun.* **2017**, *8*, 14130. [[CrossRef](#)] [[PubMed](#)]
37. De Medici, D.; Anniballi, F.; Wyatt, G.M.; Lindström, M.; Messelhäuser, U.; Aldus, C.F.; Delibato, E.; Korkeala, H.; Peck, M.W.; Fenicia, L. Multiplex PCR for Detection of Botulinum Neurotoxin-Producing Clostridia in Clinical, Food, and Environmental Samples. *Appl. Environ. Microbiol.* **2009**, *75*, 6457–6461. [[CrossRef](#)]
38. Woudstra, C.; Skarin, H.; Anniballi, F.; Fenicia, L.; Bano, L.; Drigo, I.; Koene, M.; Bâyon-Auboyer, M.-H.; Buffereau, J.-P.; De Medici, D.; et al. Neurotoxin Gene Profiling of Clostridium Botulinum Types C and D Native to Different Countries within Europe. *Appl. Environ. Microbiol.* **2012**, *78*, 3120–3127. [[CrossRef](#)]
39. Farag, S.M.; Mohammed, M.O.; El-Sobky, T.A.; ElKadery, N.A.; ElZohiery, A.K. Botulinum Toxin A Injection in Treatment of Upper Limb Spasticity in Children with Cerebral Palsy: A Systematic Review of Randomized Controlled Trials. *JBJS Rev.* **2020**, *8*, e0119. [[CrossRef](#)]
40. Blumetti, F.C.; Belloti, J.C.; Tamaoki, M.J.; Pinto, J.A. Botulinum Toxin Type A in the Treatment of Lower Limb Spasticity in Children with Cerebral Palsy. *Cochrane Database Syst. Rev.* **2019**, *2019*, CD001408. [[CrossRef](#)]
41. Bintsis, T. Foodborne Pathogens. *AIMS Microbiol.* **2017**, *3*, 529–563. [[CrossRef](#)]
42. Brüggemann, H. Genomics of Clostridial Pathogens: Implication of Extrachromosomal Elements in Pathogenicity. *Curr. Opin. Microbiol.* **2005**, *8*, 601–605. [[CrossRef](#)] [[PubMed](#)]
43. Carter, A.T.; Austin, J.W.; Weedmark, K.A.; Peck, M.W. Evolution of Chromosomal Clostridium Botulinum Type E Neurotoxin Gene Clusters: Evidence Provided by Their Rare Plasmid-Borne Counterparts. *Genome Biol. Evol.* **2016**, *8*, 540–555. [[CrossRef](#)]
44. Cruz-Morales, P.; Orellana, C.A.; Moutafis, G.; Moonen, G.; Rincon, G.; Nielsen, L.K.; Marcellin, E. Revisiting the Evolution and Taxonomy of Clostridia, a Phylogenomic Update. *Genome Biol. Evol.* **2019**, *11*, 2035–2044. [[CrossRef](#)]
45. Brunt, J.; van Vliet, A.H.M.; Stringer, S.C.; Carter, A.T.; Lindström, M.; Peck, M.W. Pan-Genomic Analysis of Clostridium Botulinum Group II (Non-Proteolytic C. Botulinum) Associated with Foodborne Botulism and Isolated from the Environment. *Toxins* **2020**, *12*, 306. [[CrossRef](#)]
46. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data. *Bioinformatics* **2012**, *28*, 3150–3152. [[CrossRef](#)]
47. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and Applications. *BMC Bioinform.* **2009**, *10*, 421. [[CrossRef](#)]
48. Smith, T.J.; Hill, K.K.; Foley, B.T.; Detter, J.C.; Munk, A.C.; Bruce, D.C.; Doggett, N.A.; Smith, L.A.; Marks, J.D.; Xie, G.; et al. Analysis of the Neurotoxin Complex Genes in Clostridium Botulinum A1-A4 and B1 Strains: BoNT/A3, /Ba4 and /B1 Clusters Are Located within Plasmids. *PLoS ONE* **2007**, *2*, e1271. [[CrossRef](#)]
49. Smith, T.J. Clostridium Botulinum Genomes and Genetic Diversity. In *Molecular Aspects of Botulinum Neurotoxin*; Foster, K.A., Ed.; Current Topics in Neurotoxicity; Springer: New York, NY, USA, 2014; pp. 207–228. ISBN 978-1-4614-9454-6.
50. Edgar, R.C. MUSCLE: A Multiple Sequence Alignment Method with Reduced Time and Space Complexity. *BMC Bioinform.* **2004**, *5*, 113. [[CrossRef](#)] [[PubMed](#)]
51. Suyama, M.; Torrents, D.; Bork, P. PAL2NAL: Robust Conversion of Protein Sequence Alignments into the Corresponding Codon Alignments. *Nucleic Acids Res.* **2006**, *34*, W609–W612. [[CrossRef](#)] [[PubMed](#)]
52. Wang, J.; Youkharibache, P.; Zhang, D.; Lanczycki, C.J.; Geer, R.C.; Madej, T.; Phan, L.; Ward, M.; Lu, S.; Marchler, G.H.; et al. ICn3D, a Web-Based 3D Viewer for Sharing 1D/2D/3D Representations of Biomolecular Structures. *Bioinformatics* **2020**, *36*, 131–135. [[CrossRef](#)] [[PubMed](#)]
53. Collins, M.D.; East, A.K. Phylogeny and Taxonomy of the Food-Borne Pathogen Clostridium Botulinum and Its Neurotoxins. *J. Appl. Microbiol.* **1998**, *84*, 5–17. [[CrossRef](#)] [[PubMed](#)]
54. Dahlsten, E.; Lindström, M.; Korkeala, H. Mechanisms of Food Processing and Storage-Related Stress Tolerance in Clostridium Botulinum. *Res. Microbiol.* **2015**, *166*, 344–352. [[CrossRef](#)] [[PubMed](#)]
55. Benoit, R.M. Botulinum Neurotoxin Diversity from a Gene-Centered View. *Toxins* **2018**, *10*, 310. [[CrossRef](#)] [[PubMed](#)]

56. Flores, A.; Ramirez-Franco, J.; Desplantes, R.; Debreux, K.; Ferracci, G.; Wernert, F.; Blanchard, M.-P.; Maulet, Y.; Youssouf, F.; Sangiardi, M.; et al. Gangliosides Interact with Synaptotagmin to Form the High-Affinity Receptor Complex for Botulinum Neurotoxin B. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 18098–18108. [[CrossRef](#)]
57. Rummel, A.; Häfner, K.; Mahrhold, S.; Darashchonak, N.; Holt, M.; Jahn, R.; Beermann, S.; Karnath, T.; Bigalke, H.; Binz, T. Botulinum Neurotoxins C, E and F Bind Gangliosides via a Conserved Binding Site Prior to Stimulation-Dependent Uptake with Botulinum Neurotoxin F Utilising the Three Isoforms of SV2 as Second Receptor. *J. Neurochem.* **2009**, *110*, 1942–1954. [[CrossRef](#)]