A Taste of Bioinformatics:

Exploring the Pathogenetics of *Clostridium botulinum*

Evgenia Trudova

Degree project/Independent project • 15 credits
Swedish University of Agricultural Sciences, SLU
Department of Energy and Technology
Food Inspection – Master`s Programme
Examensarbete 2024:07 • ISSN 1654-9392
Uppsala 2024

A Taste of Bioinformatics: Exploring the Pathogenetics of *Clostridium botulinum*

Evgenia Trudova

Supervisor: Christopher Malefors, SLU, Department of Energy and Technology
Examiner: Mattias Eriksson, SLU, Department of Energy and Technology

Credits: 15 hp
Level: Masters level (A1E)
Course title: Independent project in Food Science
Course code: EX1008
Programme/education: Food Inspection – Master´s Programme
Course coordinating dept: Department of Energy and Technology
Place of publication: Uppsala
Year of publication: 2024

Title of the Series: Examensarbete (Institutionen för energi och teknik, SLU)
Part number: 2024:07
ISSN: 1654-9392

Keywords: Clostridium botulinum, bioinformatics, BoNT , pathogenesis, botulism, toxicology

Abstract

This interdisciplinary thesis investigates *Clostridium botulinum* and botulism disease, spanning molecular biology, epidemiology, bioinformatics as well as clinical diagnosis and treatment. It delves into the morphology, genetics and environmental distribution of *Clostridium botulinum*, highlighting its adaptive strategies and spore formation and germination. This research aims to provide a comprehensive understanding of botulinum toxins, BoNT A-G by application of bioinformatic strategies to information gathering, information analysis and *in silica* methods, when researching protein interactions involved in intoxication. This study conducted a representative selection of BoNT strands, to visualise *in silica* relationship between toxins classified as BoNT A-G with a multiple sequence alignment of selected strands using Clustal O (1.2.3). Data was analysed for phylogenetic relationships among these strains. The results are organised, to highlight similarities and differences across the strains in relationship, providing figures for "Clustal highlight", "Positive", "Negative" and "Hydrophobicity" regions in sequences. Data collection included analysis of available data from different databases such as NCBI, PDB and UniProt. Protein structure of the BoNT/A Hall strain, protein database entry 2nyy, was visualized using PDB and BioPython.

Keywords: Clostridium botulinum, Bioinformatics, BoNT, Botulism, Toxin Analysis

Table of contents

List of tables

List of figures

Abbreviations

SLU	Swedish University of Agricultural Sciences
DNA	Deoxyribonucleic acid
BoNT	Botulinum neurotoxin
HC	Heavy Chain of BoNT
LC	Light Chain of BoNT
SNARE	Soluble N-Ethylmaleimide-Sensitive Factor Attachment Protein Receptor
SNAP-25	Synaptosomal-Associated Protein, 25kDa
VAMP-2	Synaptobrevin
SYX 1	Synataxin 1a
PCR	Polymerase Chain Reaction
ELISA	Enzyme-linked immunosorbent assay
PFGE	Pulsed Field Gel Electrophoresis
GMP	Good manufacturing practices
TPCY	Tryptone-Peptone-Glucose-Yeast Extract
EY	Egg Yolk Agar
Ct	Coat of proteins
Cx	Cortex of peptidoglycan
IM	Inner membrane
GCM	Germ cell wall
CaDPA	Calcium dipicolinic acid
CDC	Center for Disease Control and Prevention
BLAST	Basic Local Alignment Search Tool
FASTA	Fast- All text based format for nucleotide or protein sequences
NCBI	National Center for Biotechnology Information
UniProt	Universal Protein Resource
PDB	Protein Data Base

1. Introduction

Botulism is a severe disease affecting both humans and animals, with a high fatality rate. The sickness manifests in flaccid muscle paralysis, blindness and difficulties breathing. Botulism is difficult to diagnose, with early onset symptoms and once diagnosed, it requires prolonged recovery time which may take months or even years. In some cases, late diagnosis can develop into chronic damage (Harris et al., 2020).

Foodborne botulism is the most commonly clinically relevant case, caused by consumption of food contaminated by neurotoxins. Infant botulism arises from ecological infection of immature gut flora, and wound botulism is associated with drug use and contamination of open wounds with ecological flora containing *Clostridium botulism* bacteria (Harris et al., 2020).

The objective of this study is to investigate the neurotoxin BoNT, the agent of the disease botulism, through application of bioinformatic methodologies, as well as applying analysis of written material encompassing historical, toxicological, immunological, and pharmacological to closer study aspects related to the various serotypes of *Clostridium botulinum*. By enhanced comprehension of the pathogenic mechanism underlying botulism the aim is to give insights that might guide future research efforts in the field.

Additionally this study aims to provide examples with the goal of enhancing the understanding of this topic, it aims to do so by implementing information from interdisciplinary databases and computational tools on complex biological processes such as biochemistry of the intoxication process within BoNT toxin.

2. Theory

In 1919, Georgina Burke differentiated *Clostridium botulinum* into the serotypes BoNT/ A and BoNT/B. Since then, the differentiation of serotypes has evolved into

a tree of closely related strains, reflecting the nuanced variation in toxin production (Smith et al., 2018).

2.1 Serotypes

In addition to BoNT/ A and BoNT/B, C to G types have been identified. Furthermore, certain strains exhibit mixed serotype expression, complicating the classification by expression of toxicology. Serotype diversity expressed by strains, exhibits varying levels of neurotoxicity (Williamson et al., 2016).

As previously touched on, clinical botulism can be separated into three large groups: infant, wound, and foodborne botulism (Rawson et al., 2023). Infant botulism is not associated with a specific serotype, but instead with anatomical immaturity and low ecological protection of the child's gastrointestinal tract. This condition primarily affects infants under the age of one due to colonization of their gastrointestinal tract with *Clostridium botulinum* bacteria, which leads to local toxin production (Van Horn & Street, 2023).

Wound botulism occurs by contamination with *Clostridium botulinum* spores from soil and food, and is associated with BoNT/A and in rare cases BoNT/B and BoNT/F. Foodborne botulism is the most common variety, occurring by ingestion of toxin in contaminated food; this is associated with serotypes BoNT/ A, BoNT/B, BoNT/E, and BoNT/F (Rawson et al., 2023). Type A and B toxins are resistant to digestion due to the enzymes of the gastrointestinal system, making them more clinically relevant in cases of foodborne botulism (Corsalini et al., 2021).

This variety is further expanded with some strains of *Clostridium botulinum* which produce multiple serotypes of toxin simultaneously, classed as a hybrid subtype. This ability is likely linked to the horizontal transfer of genes within the species, or by the transfer of genetic information by plasmids. Plasmids are small, circular DNA molecules which exist independently of the bacterial chromosome and are used to directly transfer genetic information between the cells which in turn provides survival advantages for bacterial populations (Williamson et al., 2016).

The reclassification of *Clostridium botulinum* based on genetic and phenotypic characteristics provides more robust data compared to the reliance solely on clinical disease expression, particularly in a context where the expression of diseases is heavily influenced by serotype variability and the susceptibility to disruption by horizontal gene transfer among strains. Currently, there are four genetically distinct *Clostridium botulinum* groups (I -IV) with seven main serological types of synthesized neurotoxins (A - G) (Carter & Peck, 2015).

Group 1 (I) is associated with botulism in humans, is proteolytic, and produces toxins A, B, F, more than one toxin may be formed, a hybrid serology H is not yet verified (Carter & Peck, 2015).

Group 2(II) is associated with botulism in humans, is non - proteolytic, and produces toxins B, E, F (Carter & Peck, 2015).

Group 3(III)is associated with botulism in various animals, associated with C and D serotypes (Skarin et al., 2011).

Group 4 (IV) is not associated with botulism in humans or animals (Lindström & Korkeala, 2006).

The spores exhibit significant diversity among genetic groups of *Clostridium botulinum*, nested in different ecological niches. Group I is classified as a terrestrial type which produces highly heat resistant spores. Group II strains can be terrestrial or aquatic, preferring cooler temperatures and producing spores with moderate heat resistance. A different sporulation strategy suggests a potential role for exosporium, a protective layer, which enhances heat resistance. Two sporulation strategies have been identified in Group I, showing higher proportions of sporulating cells with slow release when compared to Group II and Group III strains. Spores of *Clostridium botulinum* have been identified as highly hydrophobic. The variation in heat resistance and temperature preference could be signs of environmental adaptation and genetic evolution of this bacterium (Portinha et al., 2022).

2.2 Mechanism of intoxication

The biomolecular mechanism of BoNT intoxication unfolds in sequential stages: upon entering the host cell, the mechanism progresses through a series of events. Initially, *Clostridium botulinum* synthesizes the non-activated BoNT toxin that exits the bacterial cell via a secretion system. BoNTs (e.g., BoNT-A) are activated in the digestive system by trypsin (a serine protease), activation occurs through cleavage in the scission region located between amino acids 449 and 448 between heavy and light chain (Lam & Jin, 2015).

Trypsin catalyzes the hydrolysis of peptide bonds, resulting in the separation of the BoNT molecule into two fragments bound by an internal disulfide bond. This makes the toxin more stable and bioactive.Having an intact polypeptide chain gives rigid structure and stability, but the activated state exhibits increased flexibility and geometric properties compared to the intact form, which contributes to its ability to interact with target cells and evade detection by the immune system (von Berg et al., 2019). Activated BoNT toxin spreads in the blood until contact with

neuromuscular and cholinergic junctions, penetrating the synaptic endings via endocytosis. Upon entering the neuronal cytoplasm, BoNT is engulfed by endocytosis into a vesicle within the neuron cytoplasm, endosome. Inside of the endosome acidification triggers the cleavage of the disulfide bond holding the heavy and light chain (LC) together, this results in the release of heavy chain (HC) outside of the endosome (Corsalini et al., 2021).

The architecture of the light chain (LC) of BoNT varies depending on which serotype is the origin of infection, with variation in amino acid sequences of the secondary structure. These differences influence stability, flexibility, substrate binding properties and pathogenetic properties. Zinc ions are structurally bound internally on the light chain of the BoNT molecule (Corsalini et al., 2021).

When the light chain of BoNT toxin encounters regulatory protein thioredoxin (Trx) in the cytoplasm of the cell, a catalytic process is initiated. Thioredoxin recognizes and catalyzes reduction of interchain disulfide bonds between light and heavy chains of BoNT molecule. This modification is essential for function of the BoNT toxin, thioredoxin reduction leads to the conversion of the disulfide bond in the BoNT into two thiol (sulfhydryl) groups. An activated light chain gains flexibility, and positively charged zink acts as cofactor for the protease activity, effectively giving the light chain enhanced properties, similar to a cleaving enzyme. The light chain, remaining inside of the endosome, identifies and cuts a substate Soluble N-Ethylmaleimide-Sensitive Factor Attachment Protein Receptor (SNARE), leading to disruption of the membrane fusion process and hindering the release of important neurotransmitter acetylcholine (ACh) into synaptic cleft (Pirazzini et al., 2014).

Each BoNT serotype cleaves one of three SNARE proteins; synaptosomal - associated protein 25 (SNAP-25) an attachment protein receptor found on the plasma membrane of neurons, vesicle associated membrane protein 2 (Synaptobrevin or VAMP- 2) is a protein on the membrane of synaptic vesicles or syntaxin 1a, protein found on the presynaptic membranes, facilitating the docking and fusion of SNAP-25 and VAMP-2 during release of neurotransmitter into synaptic cleft between neurons. BoNT toxin shows stereotypical and group differences between which of the SNARE proteins is cleaved, but it has been determined that BoNT/ A and BoNT E cleave SNAP-25. BoNT/C cleaves both SNAP-25 and syntaxin 1a (Syx 1). BoNT/ B, BoNT/ D, BoNT/ F and BoNT/ G cleave VAMP-2 at distinct sites (Chen, Hall, & Barbieri, 2008).

2.2.1 SNAP - 25

The key protein which is targeted by BoNT is synaptosomal-associated protein 25 (SNAP-25), which is essential for synaptic vesicle fusion and the release of neurotransmitters in the neuromuscular junction. It functions by forming a complex with other SNARE proteins, such as syntaxin and synaptobrevin, to facilitate the fusion of synaptic vesicles with the presynaptic membrane, leading to the release of neurotransmitters like acetylcholine. The light chain of BoNT acts as zinc-dependent endopeptidase, cleaving SNAP-25, thus effectively blocking the nerve signaling (Antonucci et al., 2016).

Human SNAP-25 is a protein found in nerve terminals containing 206 amino acid proteins with molecular weight of 25kDa, containing both positively and negatively charged amino acid residues giving it an overall neutral charge distribution. The positive charged area includes lysine and arginine where the negatively charged area includes aspartate and glutamate. SNAP-25 adopts a predominantly a-helical secondary structure, with two helical domains separated by a flexible linker region (Bajohrs et al., 2005). The N-terminal domain forms a four helix structure, while the C terminal domain has two helices. The domains are folded into a globular structure stabilized by hydrophobic interactions and hydrogen bonds. When interacting with partner proteins, syntaxin and synaprobrevin, the SNAP-25 undergoes conformational changes and folds to correct shape (Antonucci et al., 2016).

Developing understanding of how serotype BoNT/A cleaves SNAP-25 aids in developing insights into mechanical processes and understanding of intoxication processes.

The light chain of BoNT/A is classed as a zinc-dependent endopeptidase enzyme. The light chain of botulinum neurotoxin A (BoNT/A) was previously linked to the heavy chain of the toxin with a disulfide linkage. The translocation domain also contains 105Å - a long stretch of residues, known as a “belt” that crosses over and wraps around LC to shield the active site from solvent (Gul et al,. 2010). As a separate unit the protein is a 50ka composed of approximately 450 amino acids which adopts a predominantly a-helical and B-sheet secondary structure, folding into a compact globular structure, forming the catalytic domain of the enzyme, the zinc ion is positioned deep in the crevice of the pocket. The zinc ion on the active site of the LC acts as a cofactor, facilitating hydrolysis of the peptides. The zinc ion acts as a Lewis acid, performing a nucleophilic attack by a water molecule on the carbonyl carbon of the peptide bond within the SNAP-25.

This leads to cleavage of the bond. The SNAP-25 protein breaks into two pieces a larger N-terminal fragment and a smaller C-terminal fragment which leads to inhibition of function. In vivo LC/A cleaves C terminal residues but not N terminal residues from SNAP-25, which is an incomplete cleaving of the protein (Wang et al., 2011).

2.2.2 VAMP – 2

The principle mechanism of exocytosis and function of SNARE proteins are intricately linked to the understanding of the intoxication mechanism of the BoNT. Exocytosis refers to the fusion of synaptic vesicles with the plasma membrane, thus releasing neurotransmitters from the synaptic terminal into the synaptic cleft, which acts as a chemical signal transmission from one neuron cell to another. This process is triggered by an increase in calcium ions (Ca2+). SNARE proteins are important in orchestrating the exocytosis process, facilitating the process between neuronal membrane proteins. Calcium ions act as external stimuli which enters the neuronal cell through voltage-gated calcium channels upon electrical stimulation; as part of the neuron signaling process, the ions initiate the process of exocytosis.

Synaptobrevin, also known as VAMP-2 (vesicle- associated membrane protein 2) is located on the membrane of synaptic vesicles, as synaptotagmin changed form. Synaptotagmin is a calcium sensor located on the membrane of synaptic vesicles. When calcium ions enter the neuron, they bind to the synaptotagmin, triggering a conformational change of the protein. The VAMP-2 protein changes, becoming a functional side for interactions with SNARE complexes, consequently making the neuron vesicle containing neurotransmitter substances predisposed to interactions with membrane, thus enabling the possibility of facilitating the translocation of the substance across cellular boundaries (Somasundaram & Taraska, 2018).

Syntaxin and synaptotagmins are part of the SNARE complex, located on the plasma membrane of the neuronal cell that facilitate fusion of the VAMP-2 protein on the vesicle with the plasma membrane. This fusion results in the release of neurotransmitters stored within the vesicle into the synaptic cleft. BoNT disrupts this process by cleaving the SNARE proteins, targeting syntaxin and synaptobrevin proteins, making it impossible for VAMP-2 to fuse with them (Somasundaram & Taraska, 2018).

VAMP-2 protein is a protein involved in several types of paralysis intoxication. Both botulinum neurotoxin (BoNT) and closely related tetanus neurotoxin (TeNT) interfere in the synaptic vesicle fusion in neuron cells, by recognition of VAMP-2

protein. These toxins bind to different regions of the VAMP-2 protein, but cleave it in similar ways with a light chain zinc protease with enzymatic activity. Understanding this process could lead to better understanding of how toxins recognize protein residues (Chen et al., 2008).

2.2.3 SYX - 1

Syntaxin 1a (SYX-1) inhabits a mediating role in the exocytosis. BoNT/C degenerates Syx1 into fragments, which leads to neurodegeneration. Cleavage of Syx1 disrupts the coordination of vesicle fusion events, making neuronal communication difficult. A study has proven that even at low concentrations of BoNT/C nearly all Syx1 proteins are cleaved (Peng et al., 2013). Proteins in mouse cells can gradually recover after exposure to toxin as a dynamic response to intoxication. Neurodegeneration occurs when recovery of Syx1 proteins is overwhelmed, and nearly all Syx1 proteins are cleaved, which indicates the role of Syx1 in viability of neuron cells (Peng et al., 2013).

2.2.4 Proteomics of BoNT

The interaction between the BoNT toxin and SNARE proteins is not yet fully understood. The complexity is inherently nuanced as BoNT toxin is protected with neurotoxin-associated proteins (NAPs) that vary between serotypes, essential for shielding toxins against degradation and affecting pathogenicity of different strands. The BoNT toxin complex typically includes accessory proteins such as haemagglutinin(HA) and non-toxic-haemagglutinin (NTNH) of various sizes from 300kDa to 900kDa between different serotypes (Carter & Peck, 2015). These regions carry importance to the structure and function of the BoNT toxin (Kalb et al., 2017).

In bacterial genetics, the term *locus* refers to the precise physical position of a gene within a chromosome. Conversely an operon presents a functional unit of genetic material, a cluster of genes with interconnected functional roles. *Clostridium botulinum* genes are organised as a singular circular chromosome, with a variance in genome size between serotypes, providing genetic diversity across strains. Genes involved in the synthesis and regulation of BoNT are associated with non-toxic proteins (ANTPs) in two operon zones. Regulation of these genes is done by a sigma factor located on the botR/A region, between the two operons. First the operon synthesis the toxin, the genes are located at ntnh/A - bont/A genes situated

at the 3` terminus of the botulinum locus, followed closely with non hemagglutinin (NTNH) genes. Secondly, the operon has genes which differ between serotypes, the subtype A1 carries genes for hemagglutinin proteins (ha), ha34 (size 34kDa), ha17(17kDa) and ha70(79kDa). ANTP hemagglutinin proteins form complexes with BoNT at low pH and dissociate at high pH, protecting BoNT against degradation. Region at ha33/35 and ha50 has been linked to facilitation of BoNT transport into the cell (Connan et al., 2012).

The BoNT toxin complex demonstrates intricate internal and external adaptability, characterized by the presence of functional and protective proteins, regulatory units as well as gangliosides and synaptic vesicle proteins. This complex is involved in the formation of receptor complexes and facilitated by neurotoxin-associated proteins (ANTPs and NAPs), while showing serotype variation in synthesis and expression of genes involved in intoxication mechanics (Connan et al., 2012).

2.3 Clinical

Infections by Clostridia are divided into endogenous, one with presence in the patient microbiota and exogenous, origin of bacteria is from the external environment. Exact identification of species is rarely clinically relevant, because symptom offset is early and acute. Dysfunction of motor signaling from the upper esophagus is known as dysphagia. Individuals succumbing to symptoms of botulism which exhibit fatal outcomes show clinical asphyxia characterized by loss of muscle control in the abdomen following other neurological intoxication symptoms and profound dysphagia (Goudarzi et al., 2014).

The toxicity of a toxin is measured in a lethality dose of an expectancy of 50% of the population to die from a toxin. Standard evaluation of toxin is done by intravenous examination of death rate in mice referenced to as LD50. The value of LD50 determines that *Clostridium botulinum* toxin has 0.0003 in micrograms (μg) per one kilogram (kg). In comparison, the most venomous species of snake *Crotalus viridis helleri* has LD50 1.3, which is 4333 times less lethally toxic than *Clostridium botulinum* toxin (Casarett et al,. 2019).

2.3.1 Foodborne botulism

Botulism is classified into several types; the foodborne type is linked to ingestion of substances contaminated with toxins. The failure to effectively apply the botulinum cook (121^{o}C/ 3 min) to home canned or bottled foods has led to outbreaks of foodborne botulism associated with *Clostridium botulinum.* Botulism is an acute infectious disease which is characterized by paralysis of both smooth and skeletal muscles, as well as dysfunction of the intestines, disturbances in vision, and struggles with swallowing.

This danger serves to exemplify the importance of rigorous food inspection practices, both at smaller local scales and large industrial ones, as pivotal preventative measures against outbreaks associated with failure to uphold appropriate cooking and sterilizing temperatures throughout the food productions and distribution chain are needed to mitigate the risk of contamination and subsequent public health crisis. The results of failure to uphold these standards are exemplified by the following cases of: commercial potato soup storage (CDC., 2011), very small scale of prison-made alcohol (CDC., 2012) and Thai outbreak of 2007 (Ungchusak et al., 2007).

Initial stages show nausea, vomiting, cramping, abdominal pains and diarrhea preceding neurological symptoms. Clinical development of intoxication from BoNT does not include fevers in the subject (World Health Organization, 2023). Within hours to days, the neurological symptoms show with initial anticholinergic symptoms such as a dry month, constipation, urinary retention, diplopia, blepharoptosis, loss of accommodation and decrease or total loss of the pupillary reflex to light. Neurological symptoms progress to bilateral and symmetrical effects of the cranial nerves and continue with descending weakness, paralysis and bulbar paralysis development of dysarthria and respiratory failure (Busl & Bleck, 2012).

Cases of foodborne botulism should be differentiated from the Guillain-Barre syndrome, Miller Fisher variant, Myasthenia gravis and Lambert- Eaton syndrome. It is also important to exclude toxic causes of neuromuscular junction disorders caused by venoms, drugs and poisoning (Costa et al., 2021).

Adult cases of botulism will show absence of sensory areflexia, which is present in Guillain-Barre syndrome. While botulism does have an impact on oculomotor function, the symptoms include disturbance of vision with cloudiness, although patience with the Miller Fisher variant display distinct oculomotor dysfunctions and cranial neuropathies. Botulism is not an autoimmune disorder like Myasthenia gravis which can be diagnosed with anti-acetylcholine receptor antibody test

AChR- Ab (Testing.com, 2021). Autonomic nervous system symptoms, decline in ability to control muscles is associated with botulism, but botulism is not associated with small cell lung carcinoma like Lambert Eaton syndrome (Costa et al., 2021).

The botulism diagnosis is time sensitive and has some shared characteristics with electromyography patterns. In botulism, compound muscle action potential (CMAP) amplitude decreases due to presynaptic block. In severe cases, CMAP may be too low to observe, followed by repetitive nerve stimulation at high rates (20-50Hz) which increases CMAP amplitude to register decline in botulism with the Lambert- Eaton variant but not in the Myasthenia gravis variant (Costa et al., 2021).

2.3.2 Infant botulism

Infant botulism is the most common occurring form of botulism in the United States, with 70% of all cases, 100 cases annually. In Canada, 38 cases were reported between 1979 and 2010. Such variations in case numbers suggest risk of misdiagnosis or potential disparities in the pathogenicity of local serotypes between the two countries (Schwartz et al., 2012). The practice guidelines of the American Academy of Pediatrics recognizes that infantile botulism is caused by ingestion of spores from contaminated soil, milk and other foods, which had led them to adopt a cautionary stance of raw honey and honey products (Korioth, 2018).

In 2012, a 5-month- old patient in Canada displayed a combination of constipation together with severe neurologic findings of peripheral hypotonia even known as limb weakness or “floppy baby” syndrome (Schwartz et al., 2012). Symptoms included bulbar involvement, severe motor neuron impairment with dysfunction of speech and swallowing and effect on the brain stem (Tena et al., 2021). The patient had feeding difficulties and was observed to be drooling excessively, they were fed through a nasogastric tube and showed no signs of breathing difficulties. The case was classed as a mild expression of infant botulism, a (non- frozen) fecal sample was analyzed with a PCR test for *C. Botulinum* type A. The patient was administered antitoxin and discharged home after seven days to make a full recovery (Schwartz et al., 2012).

2.3.3 Wound botulism

In the context of this form of botulism, toxin is absorbed through the wound tissue and into the bloodstream, this is referred to as hematogenous circulation. The toxin travels in the bloodstream until it reaches nerve endings, undergoing retrograde

axonal transport from thin nerve endings to the axon bodies and the nerve bundles of the central nervous system (Notermans, 1999). Clinical cases of adult wound botulism have sudden onset of neurological symptoms, from vision impairment to swallowing and speaking difficulties. The progression of botulism shows bilateral facial weakness and paralysis of eye muscles, ophthalmoplegia. The diagnosis of wound botulism was given to the patient, based on medical history and intravenous drug use by them. The patient was admitted to the intensive case unit for supportive cases and treated with antitoxin, showing improvement after a few days of treatment (Montes-Velez et al., 2021).

2.3.4 Other cases

Furthermore, botulism is not solely restricted to humans but can affect various animal species, arising from ingestion of contaminated feed, water or carcasses containing vegetative forms of bacteria or its toxins. Clinical presentation in animals mirrors that seen in humans, with neurological symptoms ranging from muscle weakness to respiratory paralysis. Expression of botulism toxin exhibits variability which influences the severity and clinical manifestation of botulism cases (Pinna et al., 2023). A to F cases lead to clinically relevant botulism toxin expression but G serotype has historically been considered non-pathogenic to humans and mainly a source of outbreaks among livestock and wildlife. The exploration of variances in pathogenicity and methods of toxin expression between serotypes may yield scientific insight of clinical relevance (Rawson et al., 2023).

2.3.5 Clinical treatment

The treatment of BoNT intoxication in clinical settings, in abstinence of the antitoxin, is limited to breathing assistance and awaiting the elimination of the toxin from the body. Natural recovery is possible, but carries high risks of various complications, such as muscle weakness and muscle paralysis, besides the suffering and the psychological effects of long term intensive care of that variety (Casarett et al., 2019).

The method for handling clinical toxicology patients focuses on clinical stabilization, where the primary goal is to ensure that the patient's airway, breathing and circulation are maintained. Rapid decline in function can lead to toxin induced seizures. For cases of botulism without early offset of neurological symptoms there

is a risk that a patient with high care needs is placed to lower treatment priority (Casarett at al., 2019).

Early gathering of medical history makes clinical evaluation easier, and in case of botulism, can be gathered while the patient still has ability to communicate with healthcare providers. Ideally, a toxin treatment includes a prevention of further absorption of the toxin. Common tools are administration of porous activated charcoal with high absorption capability, and gastric emptying using ipecac syrup. Toxin elimination can be enhanced with multiple dose therapy of activated charcoal (MDAC), however that is effective under certain specific conditions which need to be met. Examples such as significant liver and renal circulation of toxin, formation of active recirculating metabolites, prolonged plasma half-life after an overdose, small (<1,0L/kg) volume of distribution, limited (<60%) plasma protein binding, pKa that maximizes transport of the drug across cell membrane and onset of organ failure (kidneys) which results in reduced capacity of the major route of elimination of the toxicant. The final and perhaps most desirable step in toxin treatment, is the identification of the toxin and administration of vaccine or antitoxin to treat a clinical case of toxin poisoning (Casarett et al., 2019).

2.4 Medical potential

In the realm of aesthetic medicine, BoNT injections are commonly marketed under the trade name Botox, they are dermatology injections of low concentration of toxins which reduce the appearance of wrinkles and fine lines (Hong, 2023). One of the clinical properties of Botox is temporary relaxation of muscles, making Botox effective treatment of some forms of tension migraines and chronic pain syndromes. Additionally clinical treatments have shown a capacity to alleviate pain sensitivity (Dekhne et al., 2023).

In the field of pharmacology, the light chain of botulinum toxin possesses therapeutic potential, enzymatic ability and precise target specificity. In the context of drug development, toxins exhibiting analogous properties to the light chain of botulinum toxin, are targeted during development of pharmacotherapies aiming for modification of the toxin. The toxin targets are chosen based on mechanism of action, targeting specificity, well-defined and understood molecular structures with distinct biological activity. Once the mechanics of intoxication are understood, the toxin undergoes modification to understand the pharmacokinetic properties. Such modification may involve structural alteration with loss and gain of function. Involvement of botulinum toxin in initiation of neurodegenerative processes holds

promise for development of intracellular treatments of neurodegenerative disorders (Walker et al., 2018).

This research is, at the time of writing, still ongoing. When a 3D human neuron glial model was treated with BoNT/A, the cell revealed an expected drastic decrease of acetylcholine. In addition to that, the glial model showed a release of transforming growth factor beta leading to microglial proinflammation and phosphorylation of tau (pTau) protein. The goal of the study is to gather information on the functionality of: astrocyte cells, a neuronal cell reacting to inflammation, intoxication and oxidative stress (Ambrin et al., 2024).

2.5 BoNT vaccination and antitoxins

The production of methods to treat toxin isolated intoxication presents unique challenges. One such challenge is development of therapeutic uses of products derived from the BoNT toxin itself. Vaccination leading to lifelong immunity to the active substance risks limiting the market for novel pharmacological advancements. Additionally, safety and cost analyses are imperative.

Immunogenicity associated with botulinum toxin is a topic which is not fully researched. Some clinical cases show varied response to BoNT treatment, this reaction might be sign of existence of a genetic predisposition to BoNT antibody formation due to naturally occurring differences in histocompatibility complexes (MHC) which is involved in immune response reaction on intoxication with botulism toxin, 19 out of 117 patients (16,3%) had higher level of cytokine IP10 (Baizabal-Carvallo et al., 2013).

Presently, commercial vaccines targeting BoNT rely on formulation centered on inactivated toxins (toxoids), with the earliest iterations dating back to the 1930s, employing formalin- inactivated penta- serotype- BoNT/A- E toxoid variants (Graham & Thorp, 1929).

The toxoid vaccine formulation is characterized by variable antigen yield, and many toxoid vaccines require additional booster shots to develop longer-term immunity. When a toxoid vaccine is administered, the antigen- presenting cells (APCs) such as dendritic cells process the toxoid and present peptide antigen on their cell surface using major histocompatibility complex class II (MHC-II) molecules as signal for CD4+ T cells to trigger a immunological cascade reaction, leading to activation of B cells and cytokine secretion (Gupta & Pellett, 2023).

In 1965 a pentavalent botulinum toxoid (PBT) obtained Investigation New Drug (IND) status from the Center for Disease Control IND161. Additionally in 1991,

the formula was developed under endorsement from the United States Army's Office under IND3723. This milestone vaccine using bacteria instead of animals, was reached in 2004 when first recombinant vaccine rBV A/B (*Pichia pastoris*) underwent phase 1 in clinical trials, demonstrating efficacy against a BoNT serotypes A and B (Smith & Rusnak, 2007).

Clinical cases of botulism in adults are typically treated with antitoxins instead of vaccines. Antitoxins are designed to neutralize present toxins and prevent further harm. Antitoxins are developed through immunization of animals. Two types of antitoxin are used clinically; adult equine-derived heptavalent antitoxin derived from horse blood, and human-derived immunoglobulin (BabyBIG®), derived from donors vaccinated with a pentavalent (ABCDE) toxoid vaccine, used exclusively in treatment of infant botulism. Usage of animals is a practice which has increased the risk of hypersensitivity, low specific activity has a batch to batch variability and disease risk due to pathogen transmission. The main problem with BabyBig® antitoxin is its low yield and dependence on donated blood. These issues have been addressed in Thailand, by a study that genetically engineered the *Nucotiana benthamiana* plant to produce anti- BoNT monoclonal antibodies (mAbs) by combining SDS-PAGE, western blotting and size-exclusion chromatography. Assembly of anti- BoNT/A mAbs and anti-BoNT/B mAbs have been successfully proven, and the product been purified as monomeric proteins with an estimated reduction of costs from animal cell platforms at 232USD per gram to 99USD per gram using a plant expression system, a reduction by more then 50% (Sangprasat et al., 2024).

A Chinese study addressed the potential of alternatives to non-toxic recombinant BoNT development of antitoxin in different organisms E.coli (Baldwin et al. 2008; Ben David et al. 2022; Gao et al. 2010; Moreira et al. 2014; Shi et al. 2019; Webb et al. 2017) , yeast (Liu et al. 2015; Sinha et al. 2007; Smith 2009)or 293R cell (Yu et al. 2015). Instead of targeting each serotype individually, the study developed a novel tetravalent botulism antitoxin (T-BAT) which targets a shared region between BoNT serotypes, a Hc domain with anti-Hc horse sera antibodies which could bind to all of domains derived from AL-HN of BoNT/ A with strong binding affinity. This approach represents a promising strategy for development of antitoxins (Shi et al., 2023).

2.6 BoNT, bioweapon

Botulinum toxin, produced by *Clostridium botulinum* is classified as a Category A bioterrorism agent by the Centers for Disease Control and Prevention (CDC), which is defined as agent which has the highest priority, positing a risk for national security, due to high risk of mortality rates with the potential for major public health impact, risk to cause public panic and social disruption. Despite its designation as a high-priority threat, laboratories handling this toxin operate under Biosafety Level 2 (BSL-2) conditions, signifying a moderate level of risk to both individuals and the environment (Casarett et al., 2019).

The first publicly funded biological weapons program was initiated in 1942 - 1943 in the United States, its aim was to research and develop vaccines for the Army in the Second World War (The Department of the Army, United States of America, 1977). In 1944, the responsibility for implementation was entrusted to the Chemical Warfare Service of the US Army. As part of the program, a Center for Biomedical Weapons "Fort Detrick" program was initiated from 1943 to 1969. The lab was located in the United States, Frederick County, Maryland. In unclassified documentation from 1977, the scope of the biological weapons including BoNT/A toxin research was confirmed to be primarily confined to Camp Dentrick, with a small number of contracts in universities and industries, which received agents from Camp Detrick during WW2 (The Department of the Army, United States of America, 1977).

The cultivation process of *Clostridium botulinum* was initially explained by Sommer and Schantz in 1928, reviewing methods of acid treating bacteria growing in culture to retrieve toxins by centrification. Type BoNT/A toxin of *Clostridium botulinum* was isolated and crystallized in 1946 at Camp Detrick, this was achieved using chloroform, sodium chloride and sodium acetate at controlled pH values (Lamanna et al., 1946).

Operation "Paperclip", which was conducted in 1945 to identify and relocate scientists from the Third Reich to U.S. Key relocation recruits to Fort Blizz in Texas and Fort Detrick in Maryland (Jacobsen., 2014). Notably, the Paperclip group included war criminals Kurt Blome, Erich Traub and Walter Schreiber. Kurt Blome (Yale Law School n.d) was a German microbiologist and Nazi official who was widely associated with establishment of unethical human experiments with bacterial agents, Erich Traub was a virologist developing a FMDV agent as bioweapon (Central Intelligence Agency, 1949). Walter Schreiber (Yale Law School n.d) was a German professor and bacteriologist.

The conclusive evidence regarding the active development of botulinum toxin as a weapon of mass destruction by the Third Reich remains elusive. In March 1943, Marshal Wilhelm Keutel and Adolf Hitler underscored the potential of combining chemical and biological warfare, leading to the establishment of the Blitzableiter group. However, this topic underwent thorough examination (Tate & Freugeas, 2021) and was ultimately dismissed as a myth. What is unequivocal, though, is that surge in research related to *Clostridium botulism* growth and optimization followed arrival of Third Reich microbiologists under Operation Paperclip to Fort Detrick, Maryland (Lebeda et al., 2018).

3. Method

Bioinformatics is a relatively young field of science that dates to the 1970s, when computers and programming started to become tools within science and data processing. The integration of technology in the biology field has catalyzed innovations. With exponentially increasing volumes of data produced by various instruments, bioinformatics emerged as a tool driving the progress. Bioinformatics is an interdisciplinary science focused on foundational understanding of cellular biology and computer science, mathematics and statistics. By applying bioinformatics one can extract valuable insights from diverse biological data types and grasp better understanding of biological subjects or processes (Agostino, 2012). Data processing in itself starts with fundamental tools, namely the ability to identify and align one's data (Sarkans et al., 2018).

3.1 Data collection

For this study, *Clostridium botulinum* strains of types A to G have been chosen as they are linked to the production of the neurotoxin botulinum (BoNT). Data have been collected on sequence length, mass and genetic makeup, the FASTA format for the genetic code for serotypes of interest have been retrieved from the bioinformatic database The Universal Protein Resource (UniProt). UniProt is a global attempt to organize known knowledge about proteomics and structural biology and function of proteins. The UniProt entry is a recognized identification marker, found in the FASTA format of the protein identifiers. UniProt searches provided numerous genetic relatives to the *Clostridium botulinum* as potential comparison organisms; the chosen serotypes showed variability in protein

information available in the database and potential search discoveries. It shall be noted that the data obtained at UniProt may not always contain complete information since it is continuously updated as new data becomes available. The three main data points have been isolated and presented in this study: the FASTA format for the genetic code, sequence length, and mass of the protein, this gives insight into the complexity of each of the BoNTs. Initial search followed a recommended selection of organisms, from parent (*Clostridium botulinum*) to closest genetic children by "browse all direct children" research. UniProt provides overview of genetics and an annotation score and organises browse results by best genetic match and best algorithmic find in relevance to the search question. Strains that have been chosen for this study have all had 5/5 best value, in the related UniProtKD entry results among 2,530 – 8,670 possible entries.

3.2 Data processing

Data was controlled to ensure correct FASTA format, containing headers and sequences in preparation for alignment.FASTA is a widely used format of data identification in bioinformatics. Genetic information is delineated by a descriptive header line preceded by a ">" symbol, followed by structure description and sequenced data. Understanding the structure of FASTA allows insight in data quality and organisation of genetic information. FASTA containins "sp" accession code followed by the identifier number and database linked protein name. The following sequences contain proteins, common name and description of organism species, strain identification number and subtypes identifier. Fasta format ends with organism taxonomy id (OX), the gene name associated with the protein (GN) and the protein existence level, indicating whether the protein has been experimentally verified (PE). This ranges from confirmed existence at PE = 1 to suggested theoretical existence PE = 4 (Agostino, 2012).

FASTA is structured as a searchable exact identification of the genome through multiple databases and as result genetically similar organisms have different FASTA codes, and genetic differences between nucleotide sequences can be found by comparing FASTA to each other. Alignment of FASTA is the most common of the *in silico* techniques to analyse genetic data of biological processes (Barh et al., 2014).

3.3 Sequence alignment

Multiple sequence alignment was performed using FASTA with UniProt, tested against the UniProtKB/Swiss-Prot database. The resulting data was analysed using CLUSTAL O, which provides a detailed cluster analysis. In the CLUSTAL alignment, sequences from the input data are aligned, with each sequence represented by a row. This format allows for simultaneous comparison of genetic codes, yielding reliable results. The CLUSTAL alignment highlights both regions of similarity and differences between sequences. Identical residues are aligned vertically, while gaps are indicated by dashes.

The alignment of FASTA sequences produce a numerical representation of representation of sequence similarity known as percent matrix, where higher values indicate greater similarity. Additionally the method includes generating a phylogenetic tree, which illustrates common ancestry and divergence among the sequences (Sievers & Higgins, 2018).

3.4 Visualisation

This study explored the visualisation capabilities of working with PDB-formatted data and PyMol software. PDB files for serotypes were identified and downloaded from the Protein Data Bank. Links to available visualisation of BoNT toxins selected in this study are provided in the result section for each serotype. In PyMol, these files can be opened using the open file function, which allows direct interaction with the molecular structures. Users can rotate the molecules and examine their structure from various angles.

Bioinformatics provides a methodological approach in molecular sciences, enabling the formation of hypotheses through visualization of experimentally obtained data (Bunyat-zada & Ross, 2023).

PyMol is one of the molecular visualisation tools that utilizes structural predictions and genome for 3D structure development, that allows to conduct in depth in silico analysis and provides opportunity to render molecular structures to simulate chemical reactions or environmental influences, visualising protein-ligand interaction and generating high quality images of molecules (UniProt Consortium, 2008).

3.5 Literature study

As selection of the strains was done, a literature study started with initial search and development of a search strategy working with UniProt database. Initial data was gathered from experimental results in the database, under “Function”, “Feature viewer” and “Publications”. Links at publication mention, studies done of proteins of interest. Selection of articles was a limitation to the screening, databases such as PubMed and NCBI provided a wider comprehensive search across biochemistry and medicine fields to add to the interdisciplinary nature of this study. Search strategy developed using keywords and boolean operators to find structural, functional and clinical information on selected toxins.

Literature study was done by screening of abstracts and titles, to determine relevance, followed by a selection of a hundred articles. Out of hundred articles, 85 articles were deemed relevant, and 15 have been sorted based on objective of the study not being focused on the BoNT toxin. Important part of this study was an attempt to widen the reference database to include global perspective, with articles from China, India, South Asia, and Russia to lift different directions of research that is done on the topic across the globe to avoid Western- centric bias. Fifty of the articles have been read in full text review to confirm their suitability for inclusion in the study. A third of the articles (35.29%) included high percentage of similarity, proving overview of the BoNT toxicity from different interdisciplinary perspectives. Articles with focus on biochemical data and pharmacological target identification was used to describe proteomics, intoxication mechanisms and potential of the protein in the pharmacology. Articles with focus on analysis of clinical cases was used for clinical summary of foodborne, infant and wound botulism and vaccination chapters. Inclusion of the historical data on relevance of BoNT as a biological weapon was done due to recent in-depth publications on the topic (Tate & Freugeas, 2021). These studies referenced existence of Camp Detrick documentation, without providing direct access to the resource, which this study did by analysing available declassified documentation mentioning BoNT toxin and *Clostridium botulinum* by going to the earliest publication about BoNT/A toxin in the NCBI database.

4. Results

The results section of this study aims to present a comprehensive overview of various botulinum neurotoxin subtypes providing FASTA format details on the primary amino acid structures of BoNT/A, BoNT/B, BoNT/C, BoNT/D, BoNT/E, BoNT/F from different strains of *Clostridium botulinum* strains, providing additional links to the documentation in the data banks UniProt, Protein Data Bank, Alpha Fold and Prosite. Information on the BoNT/G is limited due to lack of genetic information on chromosomal DNA.

The data provided is outlined for alignment comparison of the BoNT A-F strands, visualising variance in size of genome and volume of available information on each serotype. It shall be noted that in this study, BoNT/A is the most in depth researched serotype.

It should be noted that in this section of the study, the following amino acid sequences have all been provided in FASTA format.

4.1 BoNT/A

The following data describes the primary structures of the botulinum neurotoxin type A (BXA1_CLOBH) from the *Clostridium botulinum* strain Hall (ATCC2301/NCTC 13319/Type A). The sequence consists of 1,296 amino acids and a molecular mass of 149,426 Daltons. Additional data bank lists to the Protein Data Bank and UniProt Taxonomy have been provided for further reference.

P0DPI1 · BXA1_CLOBH

Organism: Clostridium botulinum (strain Hall / ATCC 3502 / NCTC 13319 / Type A)

Seq Length: 1,296 amino acids

Mass (Da): 149,426

>sp|P0DPI1|BXA1_CLOBH Botulinum neurotoxin type A OS=Clostridium botulinum (strain Hall / ATCC 3502 / NCTC 13319 / Type A) OX=441771 GN=botA PE=1 SV=1

MPFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLN
PPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG
STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLNLTRNGY
GSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLYGIAINPN

RVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKFKDIASTLNKA
KSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYTEDNFVKFFKV
LNRKTYLNFDKAVFKINIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTKLKNFT
GLFEFYKLLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTNDLNKGEE
ITSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISIENLSSDIIGQLELMPNIERFPNG
KKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNPSRVYTFFSSDYVKKVNKATEA
AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGPALNIGNMLYKDDFVGALIFSG
AVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTIDNALSKRNEKWDEVYKYIVTNWLAK
VNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEKNNINFNIDDLSSKLNESINKA
MININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDALLKYIYDNRGTLIGQVDRLKDK
VNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINI
GSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNN
EYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTIT
NNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELN
EKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPR
GSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRLATNASQA
GVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAK
LVASNWYNRQIERSSRTLGCSWEFIPVDDGWGERPL

Data bank links:
https://www.ebi.ac.uk/pdbe/entry/pdb/2NYY
https://www.uniprot.org/taxonomy/36826

4.2 BoNT/B

The following data describes the primary structures of the botulinum neurotoxin type B (BXB_CLOBK) from the *Clostridium botulinum* strain Okra (Type B). The sequence consists of 1,291 amino acids and a molecular mass of 150,803 Daltons. Additional data bank lists to the AlfaFold database and UniProt Taxonomy have been provided for further reference.

B1INP5 · BXB_CLOBK
Organism: Clostridium botulinum (strain Okra / Type B1)
Seq Length: 1291 amino acids
Mass (Da): 150,803

>sp|B1INP5|BXB_CLOBK Botulinum neurotoxin type B OS=Clostridium botulinum (strain Okra / Type B1) OX=498213 GN=botB PE=1 SV=1

MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN

KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG
DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH
FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLY
GIKVDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRGIV
DRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETN
IAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINKQA
YEEISKEHLAVYKIQMCKSVKAPGICIDVDNEDLFFIADKNSFSDDLSKNERIEYNTQSN
YIENDFPINELILDTDLISKIELPSENTESLTDFNVDVPVYEKQPAIKKIFTDENTIFQY
LYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAGWVKQIVND
FVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGASILLEFIPELLI
PVVGAFLLESYIDNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLSTVNTQFYTIKEGMY
KALNYQAQALEEIIKYRYNIYSEKEKSNINIDFNDINSKLNEGINQAIDNINNFINGCSV
SYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYEKSKVNKYLKTIMPFDL
SIYTNDTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFK
LTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNS
GWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYING
KLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSY
SEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLY
IGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISD
SDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCIS
KWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE

Data bank links:
https://alphafold.ebi.ac.uk/entry/B1INP5
https://www.uniprot.org/taxonomy/36827

4.3 BoNT/C

The following data describes the primary structures of the toxicity linked to protein AOA916LMJ4 (A0A916LMJ4_CLOBO) from *Clostridium botulinum* C str. Eklund. This protein consists of 330 amino acids with a molecular mass of 36, 705 Daltons. On May 1 2024, no structural information was available for this protein; the UniProt taxonomy link has thus been provided for reference.

A0A916LMJ4 · A0A916LMJ4_CLOBO
Organism: Clostridium botulinum C str. Eklund
Seq Length: 330 amino acids

Mass (Da): 36,705

>tr|A0A916LMJ4|A0A916LMJ4_CLOBO Aspartate-semialdehyde dehydrogenase OS=Clostridium botulinum C str. Eklund OX=445337 GN=asd PE=3 SV=1

MNYNVAVVGATGMVGNKFIEVLAERDFPINNLYFFASKKSAGKVLKFKDKDIVVEELKED
NIKTKKIDFALFSAGGNISLGYAPVFAKYNAIVIDNSSAWRMNPEVPLVVPEVNPQDIKL
NKGIIANPNCSTIQAVVALKPLYDKYGIKRIIYSTYQAVSGAGVGGFNDLKNGYNGEAPT
KFPYPIAGNILPHIDDFLDNGYTKEEMKMINETQKIFHDNNIKITATTARVPVFYGHSES
INVELEKPFEIEDIFNLYKNAEGILLKDDVKNLVYPLPLDAEGHDEVYVGRIRRDFSLDN
GLNIWVVADNIRKGAASNAIQIAEKIISMK

Data bank links:

No structure information available (2024/05/01)

https://www.uniprot.org/taxonomy/36828

4.4 BoNT/D

The following data describes the primary structure of the protein A0A9P2G6W8 · (A0A9P2G6W8_CLOBO) from the *Clostridium botulinum* D str. 1873. This protein consists of 501 amino acids with a molecular mass of 55,018 Daltons. The data bank links provided include Prosite documentation and UniProt taxonomy.

A0A9P2G6W8 · A0A9P2G6W8_CLOBO

Organism: Clostridium botulinum D str. 1873

Seq Length: 502 amino acids

Mass (Da): 55,018

>tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO Bifunctional NAD(P)H-hydrate repair enzyme OS=Clostridium botulinum D str. 1873 OX=592027 GN=nnrE PE=3 SV=1

MEIITAQKMRDIDRFSIENIGIPSMVLMENAALKVLKNIDINVVNSFTIICGNGNNGGDG
LALARHLIVLKKQVDIFIIGSEYTLSKDCENNYKILCNMNVKVLFIKTVEDIEFIRKSME
RKDIVIDAIFGTGLSRDIKGIHKEVIFSINKINSYIISIDIPSGLNSDTGEILGCCVKAN
KTISFQFYKKGFLNYESFKYIGELIVENIGIPQSVTKNFFINDYLVEKNSVKSIIPIRTN
YLHKGNFGRTSIVAGSVGFTGAAYIATQAAVRSGAGLVTLCCPEKIQDILSNKLVEAMTL
SFKDKEKLNDILKKSDSIAVGPGMGNNEGTLKILSNIIRYTNSPIVIDADAINVLKDNLE
MLKEKNNKIVLTPHVGEMSKITGISIDTINKNRIDIARQFAKEYDIIVLLKGYNTIITDG
ITTMVNTTGNSAMASGGMGDCLTGIIAALMSQGLDAFEAAYVGAYIHGYSGDKLSKNKFCVNA
TDILEEIPFTLKELQGITS

Data bank links:
https://prosite.expasy.org/doc/PS51385
https://www.uniprot.org/taxonomy/36829

4.5 BoNT/E

The following data describes the primary structure of the protein B2V4B5 (RLMN_CLOBA) from the *Clostridium botulinum* (strain Alaska E43 / Type E3). This protein consists of 347 amino acids with a molecular mass of 39,772. Data bank links to Alphafold and UniProt have been provided for further reference.

B2V4B5 · RLMN_CLOBA
Organism: Clostridium botulinum (strain Alaska E43 / Type E3)
Seq Length: 347 amino acids
Mass (Da): 39,772
>sp|B2V4B5|RLMN_CLOBA Probable dual-specificity RNA methyltransferase RlmN OS=Clostridium botulinum (strain Alaska E43 / Type E3) OX=508767 GN=rlmN PE=3 SV=1
MKNILDYTLEELTLWMKENNESSFRAKQIMSWIYKDVRNFSDMRNMPKSLIAKLEENFEI
SLPEIEEIYKSELDGTEKFLFKFSDGNLIESVLMRYKHGNSICISTQIGCRMGCKFCAST
IDGRIRNLTTGEILSQILVVQNYIGERISNVVLMGSGEPLDNYENVMKFLEVVSAEYGLN
IGQRHITLSTCGIVPKIYELADKELSITLAISLHAFSDEKRKEIMPIANKYSIDEILNAC
KYFINKTKRRITFEYSLVKDVNDSKEDARALGKLLKGMLCHVNLIPVNEIKERTFKRSSK
ETIQDFANILSNLGIEVTVRREMGSDINAACGQLRRSYIKTQETRGE
Data bank links:
https://alphafold.ebi.ac.uk/entry/B2V4B5
https://www.uniprot.org/taxonomy/36830

4.6 BoNT/F

The following data describes the primary structure of the protein A7GBG3 (BXF_CLOBL) from the Clostridium botulinum strain Langeland (NCTC 10281 / Type F). The protein consists of 1278 amino acids with a molecular mass of 147,075 Daltons. Data bank links to Protein Data Bank and UniProt have been provided for further reference.

A7GBG3 · BXF_CLOBL

Organism: Clostridium botulinum (strain Langeland / NCTC 10281 / Type F)

Seq Length: 1278 amino acids

Mass (Da): 147,075

>sp|A7GBG3|BXF_CLOBL Botulinum neurotoxin type F OS=Clostridium botulinum (strain Langeland / NCTC 10281 / Type F) OX=441772 GN=F PE=1 SV=1

MPVVINSFNYNDPVNDDTILYMQIPYEEKSKKYYKAFEIMRNVWIIPERNTIGTDPSDFD
PPASLENGSSAYYDPNYLTTDAEKDRYLKTTIKLFKRINSNPAGEVLLQEISYAKPYLGN
EHTPINEFHPVTRTTSVNIKSSTNVKSSIILNLLVLGAGPDIFENSSYPVRKLMDSGGVY
DPSNDGFGSINIVTFSPEYEYTFNDISGGYNSSTESFIADPAISLAHELIHALHGLYGAR
GVTYKETIKVKQAPLMIAEKPIRLEEFLTFGGQDLNIITSAMKEKIYNNLLANYEKIATR
LSRVNSAPPEYDINEYKDYFQWKYGLDKNADGSYTVNENKFNEIYKKLYSFTEIDLANKF
KVKCRNTYFIKYGFLKVPNLLDDDIYTVSEGFNIGNLAVNNRGQNIKLNPKIIDSIPDKG
LVEKIVKFCKSVIPRKGTKAPPRLCIRVNNRELFFVASESSYNENDINTPKEIDDTTNLN
NNYRNNLDEVILDYNSETIPQISNQTLNTLVQDDSYVPRYDSNGTSEIEEHNVVDLNVFF
YLHAQKVPEGETNISLTSSIDTALSEESQVYTFFSSEFINTINKPVHAALFISWINQVIR
DFTTEATQKSTFDKIADISLVVPYVGLALNIGNEVQKENFKEAFELLGAGILLEFVPELL
IPTILVFTIKSFIGSSENKNKIIKAINNSLMERETKWKEIYSWIVSNWLTRINTQFNKRK
EQMYQALQNQVDAIKTVIEYKYNNYTSDERNRLESEYNINNIREELNKKVSLAMENIERF
ITESSIFYLMKLINEAKVSKLREYDEGVKEYLLDYISEHRSILGNSVQELNDLVTSTLNN
SIPFELSSYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIY
STNRNQFGIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDC
IRNNNSGWKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGN
SRIYINGNLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETL
YSDEPDPSILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSN
TRLYTGVEVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKL
IRTSNSNNSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCF
WSFISKEHGWQEN

Data bank links:

https://www.ebi.ac.uk/pdbe/entry/pdb/3FIE

https://www.uniprot.org/taxonomy/36831

4.7 BoNT/G

The following data describes the admittedly limited information which can be found regarding botulinum neurotoxin type G (BoNT/G) in the bioinformatic databases. This is primarily due to the fact that the genes transfer by plasmids rather than on chromosomal DNA. The size of the plasmid is calculated to be 114kb(76MDa). In this case, no genetic information could be found in FASTA format. Links to the NCBI and UniProt have been provided for further reference.

In contrast to the genes of other C. botulinum serotypes, the genes encoding type G toxin are found on plasmids. The three genes encoding these proteins were closely linked on a plasmid of about 114 kb (76 MDa) but not on chromosomal DNA.

No structure information available (2024/05/01)
https://www.uniprot.org/taxonomy/447213
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC173270/

4.8 Sequence alignment

The alignment results provided by UniProt`s alignment tool are visualised with colorful highlights, providing an aid in the interpretation of sequence similarities and differences.

Green indicates high similarity regions and functional importance. Yellow indicates moderate similarity sequences and some degree of functional significance. Red indicates a highly variable region between the FASTA amino acids, these regions might show higher variance and rapid diversification or possible mutations. Blue represents gaps in the alignment, indicating that the sequences are of different lengths. Additionally, signs have been positioned at the gaps in one or more sequences compared to the reference sequence, set in this alignment to the AOA916LMJ4_CLOBO from BoNT/C strain see Figure 1.

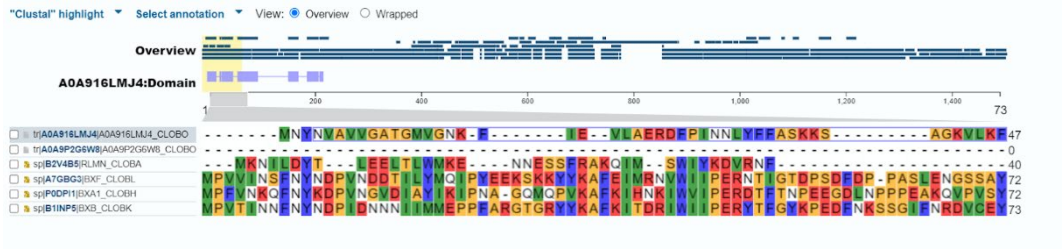

Figure 1. Image with sequence alignment of FASTA BoNT A-G in UniProt at "Clustal highlight" setting at AOA916LMJ4_CLOBO from BoNT/C str. (Eklund sample.)

The alignment results provided by UniProt`s alignment tool displays positively changed zones in the aligned sequences, marked with blue see Figure 2.

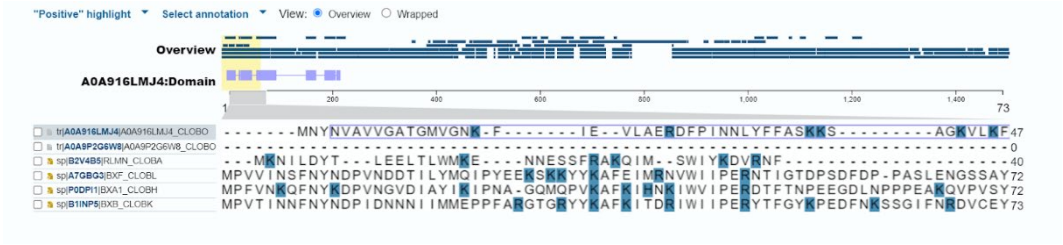

Figure 2. Image with sequence alignment of FASTA BoNT A-G in UniProt at "Positive" highlight setting at AOA916LMJ4_CLOBO from BoNT/C str. (Eklund sample.)

The alignment results provided by UniProt`s alignment tool displays negatively changed zones in the aligned sequences, marked with blue see Figure 3.

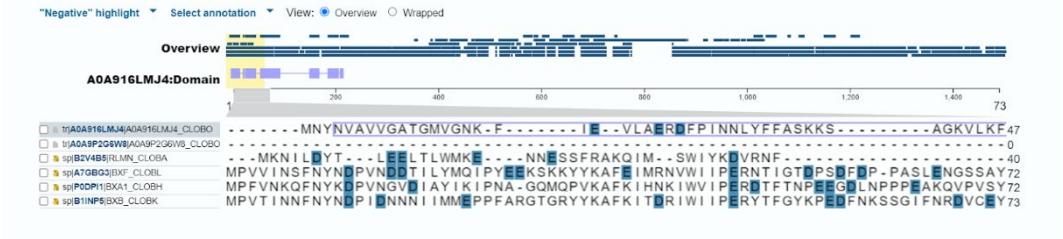

Figure 3. Image with sequence alignment of FASTA BoNT A-G in UniProt at "Negative" highlight setting at AOA916LMJ4_CLOBO from BoNT/C str. (Eklund sample.)

The alignment results provided by UniProt`s alignment tool displays hydrophobicity zones in the aligned sequences, marked with blue for hydrophobic region, purple for hydrophilic regions and charged regions in red, see Figure 4.

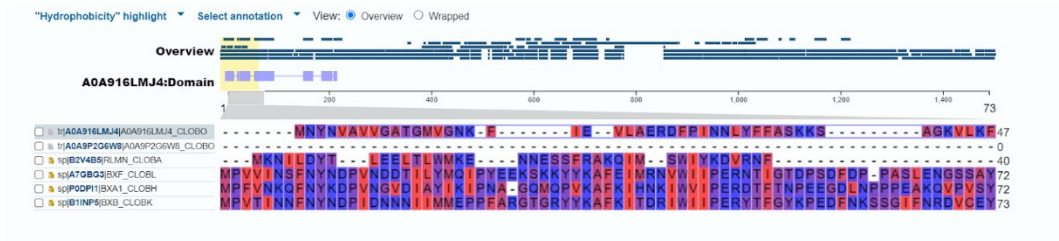

Figure 4. Image with sequence alignment of FASTA BoNT A-G in UniProt at "Hydrophobicity" highlight setting at AOA916LMJ4_CLOBO from BoNT/C str. (Eklund sample.)

4.9 Multiple sequence alignment CLUSTAL 0 (1.2.4).

The results, see Table 1, describes a multiple sequence alignment using CLUSTAL O (1.2.4) of protein sequences for BoNT, organised by the number of UniProt ID in vertical order of first value being FASTA format of BonT/C. The Eklund stand (tr|A0A916LMJ4|A0A916LMJ4_CLOBO) is followed by BoNT/D strand 1873 (tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO), BoNT/E strand Alaska E43 (sp|B2V4B5|RLMN_CLOBA), BoNT/F strand Lengeland (sp|A7GBG3|BXF_CLOBL), BoNT/A strand Hall (sp|P0DPI1|BXA1_CLOBH) and BoNT/B strand Okra(sp|B1INP5|BXB_CLOBK). This analysis has been performed to visualize the similarities and differences between the genomes of the selected strands of BoNT.

Table 1: Clustal O (1.2.3) multiple sequence alignment raw data in clustal o format for alignment in UniProt for multiple alignment organised by UniProt ID number in vertical order of first value being FASTA format of BonT/C, Eklund stand (tr|A0A916LMJ4|A0A916LMJ4_CLOBO) followed by BoNT/D strand 1873 (tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO), BoNT/E strand Alaska E43 (sp|B2V4B5|RLMN_CLOBA), BoNT/F strand Lengeland (sp|A7GBG3|BXF_CLOBL),BoNT/A strand Hall (sp|P0DPI1|BXA1_CLOBH)and BoNT/B strand Okra(sp|B1INP5|BXB_CLOBK).

CLUSTAL 0 (1.2.4)

```
tr|A0A916LMJ4|A0A916LMJ4_CLOBO      -------MNYNVAVVGATGMVGNK-F-------IE--VLAERDFPINNLYFFASKKS---      40
tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO      ------------------------------------------------------------      0
sp|B2V4B5|RLMN_CLOBA                ---MKNILDYT---LEELTLWMKE----NNESSFRAKQIM--SWIYKDVRNF--------      40
sp|A7GBG3|BXF_CLOBL                 MPVVINSFNYNDPVNDDTILYMQIPYEEKSKKYYKAFEIMRNVWIIPERNTIGTDPSDFD      60
sp|P0DPI1|BXA1_CLOBH                MPFVNKQFNYKDPVNGVDIAYIKIPNA-GQMQPVKAFKIHNKIWVIPERDTFTNPEEGDL      59
sp|B1INP5|BXB_CLOBK                 MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN      60

tr|A0A916LMJ4|A0A916LMJ4_CLOBO      ------AGKVLKFKDKDIVVEELKEDNIKTKKIDFALFSAGGN----------------      77
tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO      ------------------------------------------------------------      0
sp|B2V4B5|RLMN_CLOBA                ------------------------------------------------------------      40
sp|A7GBG3|BXF_CLOBL                 P-PASLENGSSAYYDPNYLTTDAEKDRYLKT--TIKLFKRINSNPAGEVLLQEISYAKPY      117
sp|P0DPI1|BXA1_CLOBH                NPPPEAKQVPVSYYDSTYLSTDNEKDNYLKG--VTKLFERIYSTDLGRMLLTSIVRGIPF      117
sp|B1INP5|BXB_CLOBK                 KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQT--MIKLFNRIKSKPLGEKLLEMIINGIPY      118

tr|A0A916LMJ4|A0A916LMJ4_CLOBO      ----------------------------------------ISLGYAPVFAKYNAIVID--      95
tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO      ------------------------------------------------------------      0
sp|B2V4B5|RLMN_CLOBA                ------------------------------------------------------------      40
sp|A7GBG3|BXF_CLOBL                 LGNEHTPINEFHPVTRTTSVN----IKSSTNVKSSIILNLLVLGAGPDIFENSSYPVRKL      173
sp|P0DPI1|BXA1_CLOBH                WGGSTIDTELKVIDTNCI---NVIQPDGS---YRSEELNLVIIGPSADIIQFECKSFG--      169
sp|B1INP5|BXB_CLOBK                 LGDRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIG--      176

tr|A0A916LMJ4|A0A916LMJ4_CLOBO      ---------------NSSAWRMNPEVPLVVPEVNP---QDIKLNKGIIANPNCS------      131
tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO      ------------------------------------------------------------      0
sp|B2V4B5|RLMN_CLOBA                ------------------------------------------------------------      40
sp|A7GBG3|BXF_CLOBL                 MDSGGVYDPSNDGFGSINIVTFSPEYEYTFNDISGG---YNSSTESFIADPAISLAHELI      230
sp|P0DPI1|BXA1_CLOBH                ---HEVLNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELI      226
sp|B1INP5|BXB_CLOBK                 ---IQNHFASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELI      233

tr|A0A916LMJ4|A0A916LMJ4_CLOBO      ------------------------------------------------------------      131
tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO      ------------------------------------------------------------      0
sp|B2V4B5|RLMN_CLOBA                ------------------------------------------------------------      40
sp|A7GBG3|BXF_CLOBL                 HALHGLYGARGVTYKETIKVKQAPLMIAEKPIRLEEFLTFGGQDLNIITSAMKEKIYNNL      290
sp|P0DPI1|BXA1_CLOBH                HAGHRLYGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYY      286
sp|B1INP5|BXB_CLOBK                 HVLHGLYGIKVDDLPIVP-NEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKV      292
```

```
tr|A0A916LMJ4|A0A916LMJ4_CLOBO     ------------------------------------------------------------     131

tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO     ------------------------------------------------------------     0

sp|B2V4B5|RLMN_CLOBA               ------------------------------------------------------------     40

sp|A7GBG3|BXF_CLOBL                LANYEKIATRLSRVNS--APPEYDINEYKDYFQWKYGLDKNADGSYTVNENKFNEIYKKL     348

sp|P0DPI1|BXA1_CLOBH               YNKFKDIASTLNKAKS-IVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKML     345

sp|B1INP5|BXB_CLOBK                LQNFRGIVDRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSL     352


tr|A0A916LMJ4|A0A916LMJ4_CLOBO     ------------------------------------------------------------     131

tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO     MEIITA-------------------------QKM------RDIDRFSIENIGIPSMVLME     29

sp|B2V4B5|RLMN_CLOBA               ----------------------------SDMRNMPKSLIAKLEENFEISLPEIEEIYKSE     72

sp|A7GBG3|BXF_CLOBL                -YSFTEIDLANKFKVKCRNTYFIK-YGFLKVPNLLDDDIYTVSEGFNIGN--LAVNNRGQ     404

sp|P0DPI1|BXA1_CLOBH               TEIYTEDNFVKFFKVLNRKTYLNFDKAVFK-INIVPKVNYTIYDGFNLRNTNLAANFNGQ     404

sp|B1INP5|BXB_CLOBK                MFGFTETNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQ     412


tr|A0A916LMJ4|A0A916LMJ4_CLOBO     ------------------------------------------------------------     131

tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO     NAA-----LKVLKNIDINVVNSFTIICGNGNNGGDGLALAR--------HLIVL--KKQV     74

sp|B2V4B5|RLMN_CLOBA               LDGTEKFLFK-FSD-G-NLIESVLM------------------RYKHGNSICISTQIGCRM     112

sp|A7GBG3|BXF_CLOBL                NIKLNPKIIDSIPD-K-GLVEKIVKFC--------KSVIPRKGTKAPPRLCIRVNNRE--     452

sp|P0DPI1|BXA1_CLOBH               NTEINNMNFTKLKN-FTGLFEFYKLLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWD--     461

sp|B1INP5|BXB_CLOBK                NKAINKQAYEEISK-EH--LAVYKIQM--------------CKSVKAPGICIDVDNED--     453


tr|A0A916LMJ4|A0A916LMJ4_CLOBO     ------------------------------------------------------------     131

tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO     DIFIIGSEYTLSKDCENNYKILCNMNVKVLF----------IKTVEDIEFIRKSMERKDI     124

sp|B2V4B5|RLMN_CLOBA               GCKFCASTIDGRIRNLTTGEILSQILVVQNYIGERISNVVLMGSGEPLDNYENVMKFLEV     172

sp|A7GBG3|BXF_CLOBL                -LFFVASESSYNENDINTPKE--------------ID------DTTNLNNNYRN-NLDEV     490

sp|P0DPI1|BXA1_CLOBH               -LFFSPSEDNFTN-DLNKGEE--------------IT------SDTNIEAAEENISLDLI     499

sp|B1INP5|BXB_CLOBK                -LFFIADKNSFSD-DLSKNER--------------IE------YNTQSNYIENDFPINEL     491


tr|A0A916LMJ4|A0A916LMJ4_CLOBO     ----------------------------TIQAVVALKPLYDKYGI---KRIIYSTY----     156

tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO     VIDAIFGTGLSRDIKGIH------------------------------KEVIFSINKINS     154

sp|B2V4B5|RLMN_CLOBA               -VS--AEYGL-----NIGQRHIT-------LSTCGIVPKIYELA--------DKELSITL     209

sp|A7GBG3|BXF_CLOBL                -I---LDYNSETIPQ------ISNQTLNTLVQDDSYVPRYDSNGTSEIEEHNVVDLNVFF     540

sp|P0DPI1|BXA1_CLOBH               -QQYYLTFNFDNEPENISIENLSS----DIIGQLELMPNIERFP--NGKKYELDKYTMFH     552

sp|B1INP5|BXB_CLOBK                -I---LD------TDLISKIELPSENTESLTDFNVDVPVYEKQP--AIKKIFTDENTIFQ     539


tr|A0A916LMJ4|A0A916LMJ4_CLOBO     ----QAVSGAGVGGFND--------------------------LKNGYNGEA-PTKFPYP     185

tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO     YIISIDIPS---------GLNSDTGE-ILGCCVKANKTISFQFYKKGFLNYE-----SFK     199
```

```
sp|B2V4B5|RLMN_CLOBA            AISLHAFSDEKRKEIMPIANKYSIDEILNACKYFINKTKRRITFEYSLVKDVNDSKEDAR                269
sp|A7GBG3|BXF_CLOBL             YLHAQKVPE----GETNISLTSSIDTALS------EESQVYTFFSSEFINTINKPVHAAL                590
sp|P0DPI1|BXA1_CLOBH            YLRAQEFEH----GKSRIALTNSVNEALL------NPSRVYTFFSSDYVKKVNKATEAAM                602
sp|B1INP5|BXB_CLOBK             YLYSQTFPL----DIRDISLTSSFDDALL------FSNKVYSFFSMDYIKTANKVVEAGL                589
.                          .

tr|A0A916LMJ4|A0A916LMJ4_CLOBO  IAGNI-------------LPHIDDFLDNGYTKEEMKMINETQKIFHDNNIKITAT-TARV                231
tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO  YIGELIVENIGIPQSVTKNFFINDYLVEKNSVK---------SIIPIRTNYLHKGNFGRT                250
sp|B2V4B5|RLMN_CLOBA            ALGKLLKGML----CHVNLIPVNEIKERTFKRSSKETIQDFANILSNLGIEVTVRREMGS                325
sp|A7GBG3|BXF_CLOBL             FISWI-------------NQVIRDFTTEATQKSTFDKIADISLVVPYVGLALNIGNEVQK 637
sp|P0DPI1|BXA1_CLOBH            FLGWV-------------EQLVYDFTDETSEVSTTDKIADITIIIPYIGPALNIGNMLYK 649
sp|B1INP5|BXB_CLOBK             FAGWV-------------KQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAK                636
. :           : :  .  .      :.   :

tr|A0A916LMJ4|A0A916LMJ4_CLOBO  PVF-----------------YGHSE----------------------------------  239
tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO  SIVAGSVGFTGAAYIATQAAVRSGAGLVTLCCPEKIQDILSNKLVEAMTLSFKDKEKLND  310
sp|B2V4B5|RLMN_CLOBA            DIN-----------------------------------AACGQLRRSYIKTQETR-----  345
sp|A7GBG3|BXF_CLOBL             ENFKEA-------------FELLGAGILLEFVPELLIPTILVFTIKSFIGSSENK-----  679
sp|P0DPI1|BXA1_CLOBH            DDFVGA-------------LIFSGAVILLEFIPEIAIPVLGTFALVSYIA---NK-----  688
sp|B1INP5|BXB_CLOBK             GNFENA-------------FEIAGASILLEFIPELLIPVVGAFLLESYID---NK-----  675


tr|A0A916LMJ4|A0A916LMJ4_CLOBO  ----------------------------------------------------------  239
tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO  ILKKSDSIAVGPGMGNNEGTLKILSNIIRYTNSPIVIDADAINVLKDNLEMLKEKNNKIV  370
sp|B2V4B5|RLMN_CLOBA            ----------------------------------------------------------  345
sp|A7GBG3|BXF_CLOBL             ----------------------------------------------------------  679
sp|P0DPI1|BXA1_CLOBH            ----------------------------------------------------------  688
sp|B1INP5|BXB_CLOBK             ----------------------------------------------------------  675


tr|A0A916LMJ4|A0A916LMJ4_CLOBO  ----------------------------------------------------------  239
tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO  LTPHVGEMSKITGISIDTINKNRIDIARQFAKEYDIIVL--L----------------  410
sp|B2V4B5|RLMN_CLOBA            ------------GE----------------------------------------------347
sp|A7GBG3|BXF_CLOBL             ------------NKIIKAINNSLMERETKWKEIYSWIVSNWLTRINTQFNKRKEQMYQAL  727
sp|P0DPI1|BXA1_CLOBH            ------------VLTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEAL  736
sp|B1INP5|BXB_CLOBK             ------------NKIIKTIDNALTKRNEKWSDMYGLIVAQWLSTVNTQFYTIKEGMYKAL  723


tr|A0A916LMJ4|A0A916LMJ4_CLOBO  ----------------------------------------------------------  239
tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO  ------------KGYNTIIT------------DGITTMVNTTGNSAMAS--GGMGDCLTG  444
sp|B2V4B5|RLMN_CLOBA            ----------------------------------------------------------  347
sp|A7GBG3|BXF_CLOBL             QNQVDAIKTVIEYKYNNYTSDERNRLESEYNINNIREELNKKVSLAMENIERFITESSIF  787
sp|P0DPI1|BXA1_CLOBH            ENQAEATKAIINYQYNQYTEEEKNNI--NFNIDDLSSKLNESINKAMININKFLNQCSVS  794
```

```
sp|B1INP5|BXB_CLOBK              NYQAQALEEIIKYRYNIYSEKEKSNI--NIDFNDINSKLNEGINQAIDNINNFINGCSVS                781

tr|A0A916LMJ4|A0A916LMJ4_CLOBO   ----------------------------------------------------SINVELEKPFEI          251
tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO   ----IIAALMSQGLDAFEAAY---VGAYIHGY---SGDKLSKNKFCVNATDILEEIPFTL              494
sp|B2V4B5|RLMN_CLOBA             ------------------------------------------------------------  347
sp|A7GBG3|BXF_CLOBL              YLMKLINEAKVSKLREYDEGVKEYLLDYISEHRSILGNSVQELNDLV-TSTLNNSIPFEL              846
sp|P0DPI1|BXA1_CLOBH             YLMNSMIPYGVKRLEDFDASLKDALLKYIYDNRGTLIGQVDRLKDKV-NNTLSTDIPFQL              853
sp|B1INP5|BXB_CLOBK              YLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYEKSKV-NKYLKTIMPFDL              840

tr|A0A916LMJ4|A0A916LMJ4_CLOBO   EDIFN------LYKNAEGILLKDDVKNLVYP----------------------LPLDA--          281
tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO   KELQGITS-------------------------------------------------            502
sp|B2V4B5|RLMN_CLOBA             ------------------------------------------------------------  347
sp|A7GBG3|BXF_CLOBL              SSYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQ              906
sp|P0DPI1|BXA1_CLOBH             SKYVDNQRLLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQ              913
sp|B1INP5|BXB_CLOBK              SIYTNDTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVEL--NDKNQ              898

tr|A0A916LMJ4|A0A916LMJ4_CLOBO   --------------EGHDEVYVGR-----------------------------------          291
tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO   ------------------------------------------------------------          502
sp|B2V4B5|RLMN_CLOBA             ------------------------------------------------------------  347
sp|A7GBG3|BXF_CLOBL              FGIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKV---NLNNEYTIIDCIRN              963
sp|P0DPI1|BXA1_CLOBH             IQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSI---SLNNEYTIINCMEN              970
sp|B1INP5|BXB_CLOBK              FKLTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKN              958

tr|A0A916LMJ4|A0A916LMJ4_CLOBO   --IRRDFSLDNGLNIWVVADNIRK----------------------------GAASNAI          320
tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO   ------------------------------------------------------------          502
sp|B2V4B5|RLMN_CLOBA             ------------------------------------------------------------  347
sp|A7GBG3|BXF_CLOBL              NNSGWKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRI              1023
sp|P0DPI1|BXA1_CLOBH             -NSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKI              1029
sp|B1INP5|BXB_CLOBK              -NSGWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITN-NLNNAKI              1016

tr|A0A916LMJ4|A0A916LMJ4_CLOBO   QIAEKIISMK-------------------------------------------------          330
tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO   ------------------------------------------------------------          502
sp|B2V4B5|RLMN_CLOBA             ------------------------------------------------------------  347
sp|A7GBG3|BXF_CLOBL              YINGNLIDEKSISNLGDIHVSDNILFKIVGCND-TRYVGIRYFKVFDTELGKTEIETLYS              1082
sp|P0DPI1|BXA1_CLOBH             YINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYD              1089
sp|B1INP5|BXB_CLOBK              YINGKLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYK              1076

tr|A0A916LMJ4|A0A916LMJ4_CLOBO   -------------------------------------------------          330
```

```
tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO  ------------------------------------------------------------  502

sp|B2V4B5|RLMN_CLOBA  ------------------------------------------------------------  347

sp|A7GBG3|BXF_CLOBL  DEPDPSILKDFWGNYLLYNKRYYLLNLLRTDKSITQNS----NFLNINQQRGVYQKPN--  1136

sp|P0DPI1|BXA1_CLOBH  NQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIY  1149

sp|B1INP5|BXB_CLOBK  IQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSP---------VGEILTRSKY  1127


tr|A0A916LMJ4|A0A916LMJ4_CLOBO  ------------------------------------------------------------  330

tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO  ------------------------------------------------------------  502

sp|B2V4B5|RLMN_CLOBA  ------------------------------------------------------------  347

sp|A7GBG3|BXF_CLOBL  -----IFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISI  1191

sp|P0DPI1|BXA1_CLOBH  L-------NSSLYRGTKFIIKKYAS---GNKDNIVRNNDRVYINVVVKNKEYRLATNASQ  1199

sp|B1INP5|BXB_CLOBK  NQNSKYINYRDLYIGEKFIIRRKSN-SQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYF  1186


tr|A0A916LMJ4|A0A916LMJ4_CLOBO  ------------------------------------------------------------  330

tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO  ------------------------------------------------------------  502

sp|B2V4B5|RLMN_CLOBA  ------------------------------------------------------------  347

sp|A7GBG3|BXF_CLOBL  AKPEKIIKLIRTSNSNNSLGQIIVMD-----SIGNNCTMNF--QNNNGGNIGLLGFHSN-  1243

sp|P0DPI1|BXA1_CLOBH  AGVEKILSALEIPDVGNLSQ-VVVMKSKNDQGITNKCKMNL--QDNNGNDIGFIGFHQFN  1256

sp|B1INP5|BXB_CLOBK  KKEEEKLFLAPISDSDEFYN-TIQIKE-YDEQPTYSCQLLFKKDEESTDEIGLIGIHRFY  1244


tr|A0A916LMJ4|A0A916LMJ4_CLOBO  ----------  330

tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO  ----------  502

sp|B2V4B5|RLMN_CLOBA  ----------  347

sp|A7GBG3|BXF_CLOBL  ----------NLVASSWYYNNIRKNT--SSNGCFWSFISKEHGWQEN--  1278

sp|P0DPI1|BXA1_CLOBH  NIA--------KLVASNWYNRQIERSS--RTLGCSWEFIPVDDGWGERPL  1296

sp|B1INP5|BXB_CLOBK  ESGIVFEEYKDYFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE---  1291
```

4.10 Percent Identity Matrix

The results describe a percent Identity Matrix for performing multiple alignments of the selected serotypes. The main objective of the percent identity Matrix is to perform multiple alignment experiments; it allows a quickly assessable degree of similarity between sequences, used to sort and select suitable targets of interest between closely related organisms. The identity Matrix also helps to identify regions of conservation or divergence among the selected sequences and shows insight into evolutionary relationships. The Matrix organizes the data in a visual format table, with darker shades of blue representing higher statistical resemblance

or greater genetic similarity, while lighter shades of blue indicate more distant relations, see Figure 5. This data can also be formatted in different types of a table, see Figure 6.

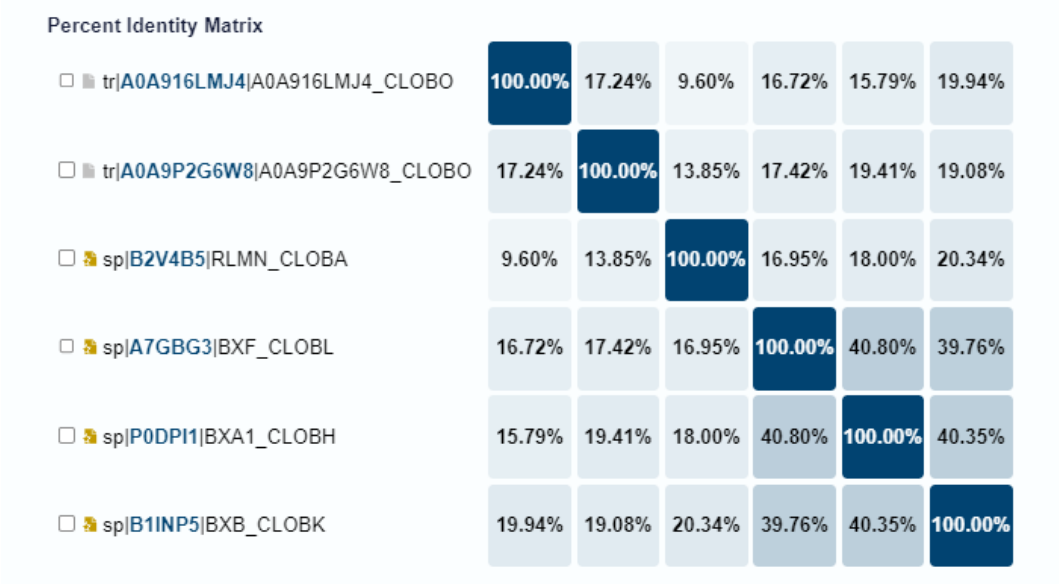

Figure 5. Image with percent identity matrix in UniProt for multiple alignment organised by UniProt ID number in vertical order of first value being. FASTA format of BonT/C, Eklund stand (tr|A0A916LMJ4|A0A916LMJ4_CLOBO) followed by BoNT/D strand 1873 (tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO), BoNT/E strand Alaska E43 (sp|B2V4B5|RLMN_CLOBA), BoNT/F strand Lengeland(sp|A7GBG3|BXF_CLOBL),BoNT/A strand Hall (sp|P0DPI1|BXA1_CLOBH)and BoNT/B strand Okra(sp|B1INP5|BXB_CLOBK).

```
#
#
#  Percent Identity  Matrix - created by Clustal2.1
#
#

     1: tr|A0A916LMJ4|A0A916LMJ4_CLOBO  100.00   17.24    9.60   16.72   15.79   19.94
     2: tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO   17.24  100.00   13.85   17.42   19.41   19.08
     3: sp|B2V4B5|RLMN_CLOBA              9.60   13.85  100.00   16.95   18.00   20.34
     4: sp|A7GBG3|BXF_CLOBL              16.72   17.42   16.95  100.00   40.80   39.76
     5: sp|P0DPI1|BXA1_CLOBH             15.79   19.41   18.00   40.80  100.00   40.35
     6: sp|B1INP5|BXB_CLOBK              19.94   19.08   20.34   39.76   40.35  100.00
```

Figure 6. Image with percent identity matrix in UniProt organised by UniProt ID number in vertical order of first value being FASTA format of BonT/C, Eklund stand (tr|A0A916LMJ4|A0A916LMJ4_CLOBO) followed by BoNT/D strand 1873 (tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO), BoNT/E strand Alaska E43 (sp|B2V4B5|RLMN_CLOBA), BoNT/F strand Lengeland (sp|A7GBG3|BXF_CLOBL),

BoNT/A strand Hall (sp|P0DPI1|BXA1_CLOBH) and BoNT/B strand Okra(sp|B1INP5|BXB_CLOBK).

4.11 Phylogenetic tree

The result displayed in Table 2 shows evolutionary relations between selected BoNT samples, identifying close relationships between BonT/C, str. Eklund and BoNT/D str.1873, similarly BoNT/E str. Alaska E43 and BoNT/B str. Okra appear to have a recent common ancestor. The branch length provided by the multiple alignment phylogenetic tree analysis is something linked to relation proximity and has thus been visualised in Figure 7.

Table 2: Raw data from multiple alignment phylogenetic tree analysis of organised by UniProt, ID numbers for tr|A0A916LMJ4|A0A916LMJ4_CLOBO, tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO,sp|B2V4B5|RLMN_CLOBA, sp|A7GBG3|BXF_CLOBL, sp|P0DPI1|BXA1_CLOBH and sp|B1INP5|BXB_CLOBK.

```
tr|A0A916LMJ4|A0A916LMJ4_CLOBO:0.42342,
tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO:0.40416)
:0.02944,
sp|B2V4B5|RLMN_CLOBA:0.43951)
:0.07227,
sp|B1INP5|BXB_CLOBK:0.28713)
:0.01634,
sp|A7GBG3|BXF_CLOBL:0.29945,
sp|P0DPI1|BXA1_CLOBH:0.29252);
```

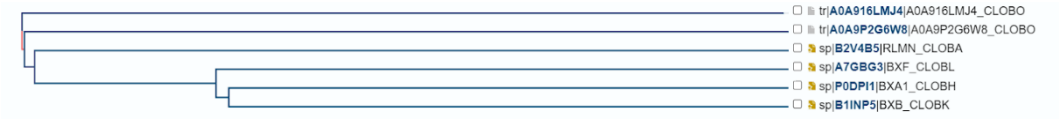

Figure 7. Image is a phylogenetic tree of organisms organised by UniProt ID numbers in horizontal order.

BoNT/A str. Hall (sp|P0DPI1|BXA1_CLOBH)

BoNT/B str. Okra(sp|B1INP5|BXB_CLOBK).

BonT/C, str. Eklund (tr|A0A916LMJ4|A0A916LMJ4_CLOBO)
BoNT/D str.1873 (tr|A0A9P2G6W8|A0A9P2G6W8_CLOBO)
BoNT/E str. Alaska E43 (sp|B2V4B5|RLMN_CLOBA)
BoNT/F str. Lengeland (sp|A7GBG3|BXF_CLOBL)

4.12 Visualisation

Result of analysis of Protein Data Base revealed limitations within the database. Entry showed in Figure 8 is for BoNT/A of this strain was available for analysis under id: 2nyy protein but entries were missing for other toxins chosen for this study. Results can be open and further analysed with Biopython and PyMol software (Cock et al., 2009).

.

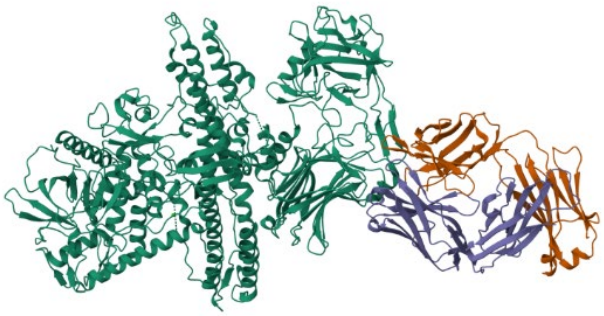

Figure 8: Protein Data Base entry for 2nyy protein of BoNT/A coloured by chain, front view.

5. Discussion

The main methodology of this study primarily took it´s form as online literature searches, performed with Google Scholar, UniProt, PubMed and NCBI databases. The search terms deepened from the initial foundational structure of botulinum toxin to intricate interactions with other proteins and mechanics of intoxication, as well as medical relevance at a molecular level. The bioinformatics results provided search terms relevant to the structure of protein enabling the possibility to investigate the pharmacological potential of the BoNT toxin. Early, when working with literature review process, an ethical discourse emerged, revealing a literature review deficiency, particularly evident in the cursory overview of lifecycle of *Clostridium botulinum* in context to the botulinum neurotoxin production. Consequently, this gap has lead to a superficial understanding of the bacterium`s lifecycle and the germination process within the intoxication pathway.Numerous studies have concentrated on the reclassification of *Clostridium botulinum* serotypes from the alphabetic categorization (A-G) to a grouping system denoted as Group 1 to 4 for easier organisation and communication, offering more flexibility for future modifications. Studies that focused on the intoxication process often lacked a comprehensive understanding of the bacterium`s lifecycle. Few articles focused on the genetic component of involved in proteomics of BoNT or immunological response to intoxication in human or animal body. Furthermore, few studies investigated possibility of clinical application of theoretical knowledge available about the toxin, such as size of the toxin, outside of methodology development for affordable antitoxin production in different organisms.

Main conclusion of this study is that serotypes BoNT/A-F show significant similarities in pathogenetic properties, germination and toxin producing behaviour while having a significant variance in structure, size and weight. Interesting research have been done identifying strains by geographical location, that is noted as one of potential reasons to this wide variety of BoNT toxin structure. This research can aid in understanding temperature preferences and adaptations between the serotypes. More regions and analysis of groups 2 to 4 can add to comprehensive grasp on geographical locations of on *Clostridium botulinum* in nature, se figure 9.

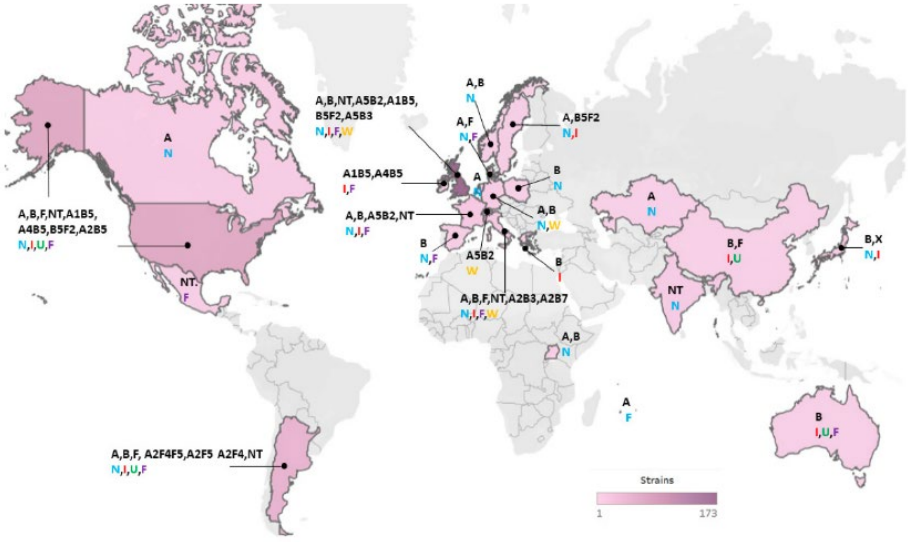

Figure 9: Geographical locations and heat map of C. botulinum Group I and C. sporogenes strains included in the present study. A total of 359 strains were attributed to a geographical location. For each specified country, the detected botulinum neurotoxin genes for isolates are indicated (in black; with NT = no toxin gene), and the types of botulism are shown in colour (N = non-clinical; I = infant botulism; U = unknown; F = foodborne botulism; W = wound botulism) (brunt et al.,2020).

These gaps in information arise from distance between experimental and practical analysis and complexities of interdisciplinary research and can be solved with bioinformatic approach to science, where available data is gathered and sorted *in silico* before theoretical conclusions are made, with focus on identification of gaps in knowledge.

Diagnostic of botulism is a difficult task which includes laboratory tests to detect the presence of BoNT or *Clostridium botulinum* bacteria in clinical stool or blood samples. The treatment involves administration of antitoxin to neutralize the effects of the toxin, along with supportive care to manage symptoms and complications. A development of a cost effective rapid diagnostic assay for presence of *Clostridium botulinum* in stool sample, can decrease laboratory costs in healthcare and industrial sector. This potential rapid test needs to differentiate between presence of *Clostridium* bacteria and presence of *Clostridium botulinum*, as *Clostridium difficile* is found in the human gut microbiota. Further research should focus on identification of targets on morphology of the rod shaped, anaerobe, gram positive *Clostridium botulinum* bacteria for potential attachment of an identification markers on cell membrane of bacterium.

Necessity of development of BoNT antidote have been communicated by several studies, but so far none have arrived at a clinically approved solution even through

3,4-diaminopyridine (3,4-DAP) shows promise as a potential in BoNT/A mediated inhibition of muscle paralysis in mouse models (Adler et al. 2000).

Difficulties in development of antidote is understandable with large morphological variables in the structure of the BoNT toxins, antidote development would need to target domain of the BoNT toxins that is common on surface of the toxin or is present across all variants of the toxin. With variation is both heavy and light chain of the toxin, this is a challenging task.

Availability of cost-effective rapid diagnostics would give better data on number of people who potentially died from botulism, without access to healthcare in rural areas. Another potential benefit is analysis of number of farm animals that die from botulism without full diagnostics. Diagnostic of farm animals can also aid in localisation of ecological areas that caused a disturbance of *Clostridium botulinum* lifecycle and potential repression of silent epidemics with science aware methods, like treatment of carcases with high heat to destroy active toxin in the gut of the animal and lower secondary infection risk.

Bioinformatics techniques such as genome alignment and phylogenetic analysis provide insight into the genetic diversity of the *Clostridium botulinum* strains and links that to diversity in structure of BoNT toxins. By examination of the structural characteristics, bioinformatics aid in understanding of pathogenesis and provide insight into target diagnostics and treatment strategies. Structural similarities and differences should be studied further to gain an improved understanding of protein interaction through transformation, toxin activation, passage into neuron cytoplasm and toxicological effect. More research is needed to understand how BoNT toxin conceals itself from detection by the immune system.

In clinical practice botulism treatment is time sensitive and carries risks of permanent damage on the body of the patient. Full recovery is not guaranteed, and intensive care is needed even after admission of antitoxin. The toxin function leads to inhibition of neurotransmitter release and muscle paralysis, often leading to loss of sight, blurry vision and loss of neurological function, mobility and breathing difficulties.

One of the possible, but perhaps considered more extreme treatments of botulism is physical removal of toxin from the blood by replacement of a patients' blood volume with blood from a donor. This solution might be an option when treating botulism in neonate infant care, if the child is not responding to vaccine and antitoxins and has a relatively low blood volume (Casarett et al., 2019).

The pathogenicity of this toxin indicates that the toxin is traveling in the bloodstream of the patient, thus, a clinical treatment by blood filtration (haemodialysis), a method of blood passing through a filter adjusted to size of the toxin, is a suitable treatment method against toxins with low volume of distribution

(0.5-0,7 L/kg), low protein binding (30-50%), relatively high degree of water solubility and low molecular weight. If the volume of distribution is high, the toxin will no longer be present in the blood, having already passed through the membrane barrier into the cells (Casarett et al., 2019).

A similar tactic of manual removal of toxin from circulation in the body might be clinically beneficial if BoNT toxin is present in the liquor, a method of filtration of liqvourpheresis (CSF-filtration), principle of particle filtration of cerebrospinal fluid (Wollinshy et al., 2001). No clinical data is available on laboratory examination of liquor in cases of acute botulism. No clinical data is available on potential accumulation of toxin in the body, structural similarities with *Clostridium tetani toxin ToNT* might indicate that the BoNT toxin could also be accumulated and transported via nerve cells to the spinal cord and is present in the brainstem. This physical proximity of liquor suggests that further studies can be done to explore the presence of BoNT toxin in liquor, thus providing potential clinical benefits. However, the filtration method is only beneficial in the absence of vaccines and antitoxin, and a significantly higher concentration of BoNT in liquor compared to blood, as CSF-filtration is a complicated treatment with higher risks than haemodialysis.

This study analyses a FASTA of *Clostridium botulinum* BoNT/C strand Eklund that has a mass 36,7 kDa and sequence length of 330 amino acids, this is on the smaller size of biological proteins and blunt filtration of all molecules of this size through a filter, would lead to complications such as removal of albumin that has a 66.5 kDa mass and 585 amino acids. Albumin is important transport protein in the blood. To achieve filtration of toxin from the blood one would need to be mindful of the size of proteins and biochemical balance in the body. Further studies are required on potential benefit of plasmapheresis filtration through absorbents such as activated charcoal during acute intoxication with BoNT toxins.

Botulinum neurotoxin is clinically significant as a potent neurotoxin that inhibits the release of acetylcholine (ACh) leading to muscle paralysis by targeting one or more of three SNARE proteins: synaprosomal-associated protein 25 (SNAP-25), synaptobrevin (VAMP-2) or synataxin (Syx1). This specificity for SNARE proteins but a variation on which of the proteins are impacted can be studied further, as possible indicator of relations between serotypes. In this study, selected serotype strains covered this variation in function, SNAP-25 is targeted by BoNT/A and BoNT/E, VAMP-2 is targeted by BoNT/B, BoNT/D, BoNT/F and BoNT/G. The only serotype that targeted two proteins is BoNT/C that targets both SNAP-25 and Syx1.

Understanding the intoxication mechanisms is essential for developing effective diagnostics and treatment strategies. Further research on this topic, can aid in identifying gaps in data available in databases such as X-ray crystallography data on different serotypes and their interactions with SNARE proteins *in silico*. SNARE proteins are interesting in other clinical cases, such as neurological diseases, synaptic dysfunction and vesicle trafficking abnormalities. For instance, a study of the degeneration of SNARE proteins with botulinum toxin can aid in understanding Alzheimer`s disease where SNARE proteins interact with amyloid-beta peptides in an ongoing process of neurodegeneration (Costa et al., 2019).

An interesting aspect is BoNT toxin`s ability to find neurological cells of the body and bind to SNARE toxins selectively, by researching protein modification and architecture and detachment of the light chain of BoNT from the heavy chain, one might develop non-toxic fragments that still have desired functionality to target neuron cells or to target a SNARE protein with extreme specificity that BoNT toxins have. A deeper study of the process of neuron cell degeneration can provide insight on the process of degeneration mechanisms triggered by different BoNT toxins, and how these mechanisms compare when different SNARE proteins are targeted into inhibition. Many toxins show pharmaceutical promise but are not developed further to clinical trials. BoNT toxins, already have a clinically approved example of toxin application, active ingredient in cosmetic Botox injections is BoNT/A toxin. This is promising, as more clinical treatments can be developed in the future.

The bioinformatic part of this study demonstrates a methodology for how to work with databases to find FASTA formatted data and do multiple sequence alignments to examine the evolutionary relationship between the strains. Based on phylogenetic tree data, *in silico* study determines that there is an evolutionary relationship between selected BoNT strains in Figure 7 and Table 2. The closest related strains are BoNT/C and BoNT/D, these strains have a very short branch length between them (0.02944), indicating recent mutation between the strains. BoNT/A strain has a variance in branch length to other strains, BoNT /B (0.28713), BoNT/ C (0.42342), BoNT/D (0.40416), BoNT/E (0.43951), BoNT/F (0.29945) indicating closest relationship to BoNT/B and next closest relationship to BoNT/F toxin. These distances illustrate evolutionary relationships, which cannot directly suggest their relative age. But this examination can suggest that BoNT/A, BoNT/B, and BoNT/F are likely close relatives, while BoNT/C and BoNT/D are on a different branch. This theory is strengthened by data from percent identity matrix results for these strains, where BoNT/C and BoNT/D have an evolutionary relationship of 40.80% shared amino acid sequences, suggesting development from a common ancestor with similar structure and conserved regions.

Visualisation of the proteins with protein database revealed notable limitations within the database and gaps in available PDB formatting for BoNT toxins. This gives impression that all BoNT toxins look and function as BoNT/ A toxin, which is visual manipulation of the data. The size of genome indicates that other toxins are much smaller in size and look different from the BoNT/A toxin (id:2nyy). Visual representation of protein provides availability for structural analysis. The PDB formatted data can be directly applied in the mechanism of action simulations with tools such as Biopython and PyMol. These tools could not be used in scope of this study, but in further research interactions between BoNT toxins and SNARE proteins can be developed. Visualisation of interactions between proteins can aid in identification of domains relevant for functionality of a protein, highlighting specific chains, residues and domains within proteins internal structure or external areas. Folding of protein, spatial arrangement and movement within the protein is important in analysis of binding sites for identification markers.

Gaps in information limits our ability to draw comprehensive conclusions about the structural and functional similarities and differences among these toxins, as physical mobility within a protein and folding form are important parameters for which part of the protein interacts with the target. This hinders our ability to validate findings across multiple toxins.

Identification of targets for multiple alignments involves understanding of characteristics of the database and bacterium and its genetic makeup. The alignment of FASTA can be realized using a sequence similarity search tool, BLAST with NCBI database, or by alignment tool provided by the UniProt database. These tools use different algorithms while providing similar services, with some differences in speed and efficiency (Boratyn et al., 2013).

This study uses UniProt alignment tool. This *in silica* analysis is dependent on correct parameter setting, sequence quality, and algorithm, to produce accurate results. Different algorithms produce different results. As the volume of data increases, the selection of alignment methods becomes more complicated. Analysis itself is computationally intensive when working with large data volumes, this is a time-requiring process. A weakness of multiple alignment is that it can be difficult to see the biological significance of findings, this method often requires that it is used in combination together with *in silica* visualisation methods such as PyMol or experimental validation by chemical reaction, to confirm functional relevance of discovered domains or regions. For the serotype BoNT/G database information was so limited, that the toxin was excluded from multiple alignment analysis of the proteins, which cannot be performed without a FASTA sequence.

Besides gaps in knowledge about BoNT/G primary limitation of this study is gaps in data for other BoNT toxins across serotypes in the PDB format. Without entries for toxins, one is unable to perform a visual, complete comparative structural analysis. A good approach to describing a protein often includes a robust strategy for selecting samples, followed by sequence alignment to validate the selected samples, then a variety of structural predictions, and finally a functional annotation tool. This study showed high sequence diversity and structural variability, which was a designed outcome to represent the complexity of BoNT serotypes. To improve accuracy, one might repeat multiple alignments on evolutionary closely related serotypes such as BoNT/C and BoNT/D to study similarities and differences between these proteins.

6. Summary

This literature review synthesizes finding from multiple disciplines to provide a comprehensive overview of the topic. This study examines botulism as a clinical disease leaning into significant finds on function of BoNT toxins and function of neurotransmission in human cells. This theoretical framework was deepened by exploration of available data in bioinformatic databases and multiple sequence alignment, phylogenetic tree, and potential of visual analysis of found data with Biopython and PyMol software (Cock et al., 2009). Previous research has highlighted the significant role of various factors in the process of botulism infection. Utilizing an interdisciplinary approach, supported by bioinformatic databases, enables a comprehensive understanding of botulism without prior insights into the research on this toxin group.

Bioinformatic databases are crafted to accommodate, organize and administer vast quantities of raw biological data with high accuracy and integrity. These databases serve as a global centralized hubs where data from experimental studies are collected and curated, into universal formatting such as FASTA, PDB and CLUSTAL O. Which creates a collaborative platform, where interdisciplinary scientists can contribute to the enrichment and refinement of available data, fostering a collective effort to expand our understanding of nature. This approach to research, from the roots up. is a methodology that brings analysis of literature to analysis of available bioinformatic data, in context to microbiology or microbiome, genetics, clinical or practical application of found data and finds of relationships

within selected data group. This method might be useful in shaping understanding and brings forward bioinformatic as a useful tool of analysis across scientific disciplines.

References

Harris, R. A., Anniballi, F., & Austin, J. W. (2020). *Adult Intestinal Toxemia Botulism. Toxins (Basel), 12(2), 81*. https://doi.org/10.3390/toxins12020081

Monheit, G. D., & Pickett, A. (2017). *Abobotulinumtoxin A: A 25-Year History. Aesthetic Surgery Journal, 37(Suppl 1), S4–S11.* https://doi.org/10.1093/asj/sjw284

Erkmen, O. (2022). *Practice 22 - Isolation of Clostridium botulinum.* In O. Erkmen (Ed.), *Microbiological Analysis of Foods and Food Processing Environments* (pp. 243–251). Academic Press. https://doi.org/10.1016/B978-0-323-91651-6.00054-9

Portinha, I. M., Douillard, F. P., Korkeala, H., & Lindström, M. (2022). *Sporulation Strategies and Potential Role of the Exosporium in Survival and Persistence of Clostridium botulinum. International Journal of Molecular Sciences, 23(2), 754.* https://doi.org/10.3390/ijms23020754

da Silva, T. L., Reis, A., Kent, C. A., Kosseva, M., Roseiro, J. C., & Hewitt, C. J. (2005). *Stress-induced physiological responses to starvation periods as well as glucose and lactose pulses in Bacillus licheniformis CCMI 1034 continuous aerobic fermentation processes as measured by multi-parameter flow cytometry. Biochemical Engineering Journal, 24(1), 31-41.*

Makarova, K. S., Anantharaman, V., Aravind, L., & Koonin, E. V. (2012). *Live virus-free or die: coupling of antivirus immunity and programmed suicide or dormancy in prokaryotes. Biology direct, 7, 1-10.*

Mearls, E. B., Izquierdo, J. A., & Lynd, L. R. (2012). *Formation and characterization of non-growth states in Clostridium thermocellum: spores and L-forms. BMC microbiology, 12, 1-9.*

Romero-Rodríguez, A., Ruiz-Villafán, B., Martínez-de la Peña, C. F., & Sánchez, S. (2023). *Targeting the Impossible: A Review of New Strategies against Endospores. Antibiotics, 12(2), 248.* https://doi.org/10.3390/antibiotics12020248

Centers for Disease Control and Prevention. (2008). *Steam Sterilization Guideline for Disinfection and Sterilization in Healthcare Facilities.* Retrieved from https://www.cdc.gov/infectioncontrol/guidelines/disinfection/sterilization/steam.html

Glass, K., & Marshall, K. (2013). *Clostridium botulinum.* In J. G. Morris & M. E. Potter (Eds.), *Foodborne Infections and Intoxications* (Fourth Edition) (pp. 371-387). Academic Press. https://doi.org/10.1016/B978-0-12-416041-5.00027-5.

Rawson, A. M., Dempster, A. W., Humphreys, C. M., & Minton, N. P. (2023). *Pathogenicity and virulence of Clostridium botulinum. Virulence, 14(1),* 2205251. https://doi.org/10.1080/21505594.2023.2205251

Proydakov, M. A. (2001). *Lecture on Botulism by Proydakov Michail Andreevich* [Пройдаков Михаил Андреевич - ботулизм (2001)]. Retrieved April 20, 2024, from https://www.youtube.com/watch?v=r6C4GfDIuWg&t=2808s. Rostov State Medical University, Infectious disease depatment [Кафедра инфекционных болезней РостГМУ]

Lindström, M., & Korkeala, H. (2006). *Laboratory diagnostics of botulism. Clinical microbiology reviews, 19(2), 298–314.* https://doi.org/10.1128/CMR.19.2.298-314.2006

Corsalini, M., Inchingolo, F., Dipalma, G., Wegierska, A. E., Charitos, I. A., Potenza, M. A., Scarano, A., Lorusso, F., Inchingolo, A. D., Montagnani, M., & Santacroce, L. (2021). *Botulinum Neurotoxins (BoNTs) and Their Biological, Pharmacological, and Toxicological Issues: A Scoping Review. Applied Sciences, 11(19),* 8849. https://doi.org/10.3390/app11198849

Notermans, S. H. W. (1999). *CLOSTRIDIUM | Detection of Neurotoxins of Clostridium Botulinum.* In R. K. Robinson (Ed.), *Encyclopaedia of Food Microbiology (pp. 463-466).* Elsevier. https://doi.org/10.1006/rwfm.1999.0400.

Dembek, Z. F., Smith, L. A., & Rusnak, J. M. (2007). *Botulism: cause, effects, diagnosis, clinical and laboratory identification, and treatment modalities. Disaster medicine and public health preparedness, 1(2), 122–134.* https://doi.org/10.1097/DMP.0b013e318158c5fd

Wilkins, P. A. (2014). *Botulism.* In D. C. Sellon & M. T. Long (Eds.), *Equine Infectious Diseases* (Second Edition) (pp. 364-367.e2). W.B. Saunders. https://doi.org/10.1016/B978-1-4557-0891-8.00043-9.

Lebeda, F. J., Cer, R. Z., Mudunuri, U., Stephens, R., Singh, B. R., & Adler, M. (2010). *The zinc-dependent protease activity of the botulinum neurotoxins. Toxins, 2(5), 978–997.* https://doi.org/10.3390/toxins2050978

Ramakrishnan, N. A., Drescher, M. J., & Drescher, D. G. (2012). *The SNARE complex in neuronal and sensory cells. Molecular and cellular neurosciences, 50(1), 58–69.* https://doi.org/10.1016/j.mcn.2012.03.009

Smith, T., Williamson, C. H. D., Hill, K., Sahl, J., & Keim, P. (2018). *Botulinum Neurotoxin-Producing Bacteria. Isn't It Time that We Called a Species a Species?* mBio, 9(5), e01469-18. https://doi.org/10.1128/mBio.01469-18

Williamson, C.H.D., Sahl, J.W., Smith, T.J. (2016) *Comparative genomic analyses reveal broad diversity in botulinum-toxin-producing Clostridia* . BMC Genomics 17, 180 https://doi.org/10.1186/s12864-016-2502-z

Rawson, A. M., Dempster, A. W., Humphreys, C. M., & Minton, N. P. (2023). *Pathogenicity and virulence of Clostridium botulinum. Virulence, 14(1), 2205251*. https://doi.org/10.1080/21505594.2023.2205251

Van Horn NL, Street M.(2023). *Infantile Botulism*. [Updated 2023 Jun 12]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK493178/

Rawson, A. M., Dempster, A. W., Humphreys, C. M., & Minton, N. P. (2023). *Pathogenicity and virulence of Clostridium botulinum. Virulence, 14(1), 2205251.* https://doi.org/10.1080/21505594.2023.2205251

Corsalini, M., Inchingolo, F., Dipalma, G., Wegierska, A. E., Charitos, I. A., Potenza, M. A., Scarano, A., Lorusso, F., Inchingolo, A. D., Montagnani, M., &

Santacroce, L. (2021). *Botulinum Neurotoxins (BoNTs) and Their Biological, Pharmacological, and Toxicological Issues: A Scoping Review. Applied Sciences, 11(19), 8849.* https://doi.org/10.3390/app11198849

Williamson, C.H.D., Sahl, J.W., Smith, T.J. (2016) *Comparative genomic analyses reveal broad diversity in botulinum-toxin-producing Clostridia.* BMC Genomics 17, 180 https://doi.org/10.1186/s12864-016-2502-z

Carter, A. T., & Peck, M. W. (2015). *Genomes, neurotoxins and biology of Clostridium botulinum Group I and Group II. Research in Microbiology, 166(4), 303-317.* https://doi.org/10.1016/j.resmic.2014.10.010

Skarin, H., Håfström, T., Westerberg, J., & Segerman, B. (2011*). Clostridium botulinum group III: a group with dual identity shaped by plasmids, phages and mobile elements.* BMC genomics, 12, 185. https://doi.org/10.1186/1471-2164-12-185

Lindström, M., & Korkeala, H. (2006). *Laboratory diagnostics of botulism. Clinical microbiology reviews, 19(2), 298–314.* https://doi.org/10.1128/CMR.19.2.298-314.2006

Portinha, I. M., Douillard, F. P., Korkeala, H., & Lindström, M. (2022). *Sporulation Strategies and Potential Role of the Exosporium in Survival and Persistence of Clostridium botulinum. International Journal of Molecular Sciences, 23(2), 754.* https://doi.org/10.3390/ijms23020754

Lam, K.-H., & Jin, R. (2015). *Architecture of the botulinum neurotoxin complex: a molecular machine for protection and delivery. Current Opinion in Structural Biology, 31, 89–95.* https://doi.org/10.1016/j.sbi.2015.03.013

von Berg, L., Stern, D., Pauly, D. (2019) *Functional detection of botulinum neurotoxin serotypes A to F by monoclonal neoepitope-specific antibodies and suspension array technology. Sci Rep 9, 5531* https://doi.org/10.1038/s41598-019-41722-z

Corsalini M, Inchingolo F, Dipalma G, Wegierska AE, Charitos IA, Potenza MA, Scarano A, Lorusso F, Inchingolo AD, Montagnani M, (2021) *Botulinum*

Neurotoxins (BoNTs) and Their Biological, Pharmacological, and Toxicological Issues: A Scoping Review. Applied Sciences. 2021; 11(19):8849. https://doi.org/10.3390/app11198849

Pirazzini, M., Azarnia Tehran, D., Zanetti, G., Megighian, A., Scorzeto, M., Fillo, S., Shone, C. C., Binz, T., Rossetto, O., Lista, F., & Montecucco, C. (2014). *Thioredoxin and its reductase are present on synaptic vesicles, and their inhibition prevents the paralysis induced by botulinum neurotoxins. Cell Reports, 8(6), 1870–1878.* https://doi.org/10.1016/j.celrep.2014.08.017

Chen, S., Hall, C., & Barbieri, J. T. (2008). *Substrate recognition of VAMP-2 by botulinum neurotoxin B and tetanus neurotoxin. The Journal of biological chemistry, 283(30), 21153–21159.* https://doi.org/10.1074/jbc.M800611200

Antonucci, F., Corradini, I., Fossati, G., Tomasoni, R., Menna, E., & Matteoli, M. (2016). *SNAP-25, a Known Presynaptic Protein with Emerging Postsynaptic Functions. Frontiers in synaptic neuroscience, 8, 7.* https://doi.org/10.3389/fnsyn.2016.00007

Wang, J., Zurawski, T. H., Meng, J., Finn, D. P., Wheeler, L., & Dolly, J. O. (2011). *A Dileucine in the Protease of Botulinum Toxin A Underlies Its Long-lived Neuroparalysis: Transfer of Longevity to a Novel Potential Therapeutic. Neurobiology, 286(8), 6375-6385*. https://doi.org/10.1074/jbc.M110.181784

Somasundaram, A., & Taraska, J. W. (2018). *Local protein dynamics during microvesicle exocytosis in neuroendocrine cells. Molecular biology of the cell, 29(15), 1891–1903.* https://doi.org/10.1091/mbc.E17-12-0716

Chen, S., Hall, C., & Barbieri, J. T. (2008). *Substrate recognition of VAMP-2 by botulinum neurotoxin B and tetanus neurotoxin. The Journal of biological chemistry, 283(30), 21153–21159.* https://doi.org/10.1074/jbc.M800611200

Peng, L., Liu, H., Ruan, H., Tepp, W. H., Stoothoff, W. H., Brown, R. H., Johnson, E. A., Yao, W. D., Zhang, S. C., & Dong, M. (2013). *Cytotoxicity of botulinum neurotoxins reveals a direct role of syntaxin 1 and SNAP-25 in neuron survival. Nature communications, 4, 1472.* https://doi.org/10.1038/ncomms2462

Kalb, S. R., Baudys, J., Smith, T. J., Smith, L. A., & Barr, J. R. (2017). *Characterization of Hemagglutinin Negative Botulinum Progenitor Toxins. Toxins, 9(6), 193.* https://doi.org/10.3390/toxins9060193

Connan, C., Brüggemann, H., Mazuet, C., Raffestin, S., Cayet, N., & Popoff, M. R. (2012). *Two-component systems are involved in the regulation of botulinum neurotoxin synthesis in Clostridium botulinum type A strain Hall.* PloS one, 7(7), e41848. https://doi.org/10.1371/journal.pone.0041848

Goudarzi, M., Seyedjavadi, S. S., Goudarzi, H., Mehdizadeh Aghdam, E., & Nazeri, S. (2014). *Clostridium difficile Infection: Epidemiology, Pathogenesis, Risk Factors, and Therapeutic Options.* Scientifica, 2014, 916826. https://doi.org/10.1155/2014/916826

Casarett, L. J., Doull, J., & Klaassen, C. D.(Eds.) (2019). *Casarett & Doull's Essentials of Toxicology: The Basic Science of Poisons* (9th ed.)(pp. 65-122). McGraw-Hill Education. [Note: Mary Amdur, who held a PhD in biochemistry, contributed to all chapters, although her contribution was not acknowledged due to Harvard´s "Donora smog scandal" 1948.]

Centers for Disease Control and Prevention (CDC. (2011). *Notes from the field: botulism caused by consumption of commercially produced potato soups stored improperly---ohio and georgia, 2011.* MMWR: Morbidity & Mortality Weekly Report, 60(26).

Centers for Disease Control and Prevention (CDC. (2012). *Botulism from drinking prison-made illicit alcohol-utah 2011.* MMWR: Morbidity & Mortality Weekly Report, 61(39).

Ungchusak, K., Chunsuttiwat, S., Braden, C. R., Aldis, W., Ueno, K., Olsen, S. J., & Wiboolpolprasert, S. (2007). *The need for global planned mobilization of essential medicine: lessons from a massive Thai botulism outbreak. Bulletin of the World Health Organization, 85(3), 238-240.*

World Health Organization. (2023, September 25). *Botulism.* Retrieved from https://www.who.int/news-room/fact-sheets/detail/botulism

Busl, K. M., & Bleck, T. P. (2012). *Treatment of neuroterrorism. Neurotherapeutics : the journal of the American Society for Experimental NeuroTherapeutics, 9(1), 139–157.* https://doi.org/10.1007/s13311-011-0097-2

Costa, A. M., Silva, J. M., Belém, F., Silva, L. P., Ascensão, M., & Evangelista, C. (2021). *Foodborne botulism: a case report. Porto biomedical journal, 6(1), e115.* https://doi.org/10.1097/j.pbj.0000000000000115

Schwartz, K. L., Austin, J. W., & Science, M. (2012*). Constipation and poor feeding in an infant with botulism.* CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne, 184(17), 1919–1922. https://doi.org/10.1503/cmaj.120340

Korioth, T. (2018, November 19). *Remind families: honey can cause infant botulism. American Academy of Pediatrics.* Retrieved from: https://publications.aap.org/aapnews/news/13225/Remind-families-honey-can-cause-infant-botulism?autologincheck=redirected

Schwartz, K. L., Austin, J. W., & Science, M. (2012). *Constipation and poor feeding in an infant with botulism.* CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne, 184(17), 1919–1922. https://doi.org/10.1503/cmaj.120340

Tena, A., Claria, F., Solsona, F., Meister, E., & Povedano, M. (2021). *Detection of Bulbar Involvement in Patients With Amyotrophic Lateral Sclerosis by Machine Learning Voice Analysis: Diagnostic Decision Support Development Study.* JMIR medical informatics, 9(3), e21331. https://doi.org/10.2196/21331

Schwartz, K. L., Austin, J. W., & Science, M. (2012). *Constipation and poor feeding in an infant with botulism.* CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne, 184(17), 1919–1922. https://doi.org/10.1503/cmaj.120340

Notermans, S. H. W. (1999). *Clostridium: Detection of Neurotoxins of Clostridium Botulinum.* In R. K. Robinson (Ed.), Encyclopedia of Food Microbiology (pp. 463-466). Elsevier. https://doi.org/10.1006/rwfm.1999.0400

Montes-Velez, D., Bautista, W., Brophy, S., Chatten-Brown, J., & Ranasinghe, L. (2021). *A Rare Case of Botulism in an Adult Patient. Iberoamerican Journal of Medicine.* Retrieved from http://www.iberoamericanjm.tk

Pinna, L., Coccollone, A., Maxia, M., Bano, L., Scalfaro, C., Mandas, D., & Liciardi, M. (2023). *Botulism in Cattle: A Case Report of an Outbreak in Sardinia (Italy). Animals : an open access journal from MDPI, 13(15), 2435.* https://doi.org/10.3390/ani13152435

Rawson, A. M., Dempster, A. W., Humphreys, C. M., & Minton, N. P. (2023). *Pathogenicity and virulence of Clostridium botulinum. Virulence, 14(1), 2205251.* https://doi.org/10.1080/21505594.2023.2205251

Hong S. O. (2023). *Cosmetic Treatment Using Botulinum Toxin in the Oral and Maxillofacial Area: A Narrative Review of Esthetic Techniques. Toxins, 15(2), 82.* https://doi.org/10.3390/toxins15020082

Dekhne, A., Goklani, H. D., Doshi, N., Baskara Salian, R., Gandhi, S. K., & Patel, P. (2023). *Effectiveness of Botulinum Toxin in the Treatment of Neuropathic Pain: A Literature Review. Cureus, 15(10), e46848.* https://doi.org/10.7759/cureus.46848

Walker, A. L., Imam, S. Z., & Roberts, R. A. (2018). *Drug discovery and development: Biomarkers of neurotoxicity and neurodegeneration. Experimental biology and medicine (Maywood, N.J.), 243(13), 1037–1045.* https://doi.org/10.1177/1535370218801309

Ambrin, G., Kang, Y. J., Van Do, K., Lee, C., Singh, B. R., & Cho, H. (2024). *Botulinum Neurotoxin Induces Neurotoxic Microglia Mediated by Exogenous Inflammatory Responses. Advanced science (Weinheim, Baden-Wurttemberg, Germany), 11(15), e2305326.* https://doi.org/10.1002/advs.202305326

Baizabal-Carvallo, J. F., Jankovic, J., & Feld, J. (2013). *Flu-like symptoms and associated immunological response following therapy with botulinum toxins. Neurotoxicity research, 24(2), 298–306.* https://doi.org/10.1007/s12640-013-9400-9

Graham, R., Thorp, F.(1929) *The Effect of Formalin on Botulinum Toxins A, B and C. J. Immunol. 1929, 16, 391–401.*

Gupta S, Pellett S. (2023) *Recent Developments in Vaccine Design: From Live Vaccines to Recombinant Toxin Vaccines. Toxins. 2023; 15(9):563.* https://doi.org/10.3390/toxins15090563

Smith, L. A., & Rusnak, J. M. (2007). *Botulinum Neurotoxin Vaccines: Past, Present, and Future. Critical Reviews in Immunology, 27(4), 303-318.* https://doi.org/10.1615/CritRevImmunol.v27.i4.20

Sangprasat, K., Bulaon, C. J. I., Rattanapisit, K., Srisangsung, T., Jirarojwattana, P., Wongwatanasin, A., & Phoolcharoen, W. (2024). *Production of monoclonal antibodies against botulinum neurotoxin in Nicotiana benthamiana. Human vaccines & immunotherapeutics, 20(1), 2329446.*

Sangprasat, K., Bulaon, C. J. I., Rattanapisit, K., Srisangsung, T., Jirarojwattana, P., Wongwatanasin, A., & Phoolcharoen, W. (2024). *Production of monoclonal antibodies against botulinum neurotoxin in Nicotiana benthamiana. Human vaccines & immunotherapeutics, 20(1), 2329446.*

Shi, DY., Lu, JS., Mao, YY. (2023) *Characterization of a novel tetravalent botulism antitoxin based on receptor-binding domain of BoNTs. Appl Microbiol Biotechnol 107, 3205–3216.* https://doi.org/10.1007/s00253-023-12515-2

Agostino, M. (2012). *Practical Bioinformatics. CRC Press Inc. (pp. 23-71).*

Sarkans U, Gostev M, Athar A (2018) *The BioStudies database-one stop shop for all data supporting a life sciences study. Nucleic Acids Research. 2018 Jan;46(D1):D1266-D1270.* https://doi.org/10.1093/nar/gkx965. PMID: 29069414; PMCID: PMC5753238.

Barh, D., Chaitankar, V., Yiannakopoulou, E. C., Salawu, E. O., Chowbina, S., Ghosh, P., & Azevedo, V. (2014). *In Silico Models: From Simple Networks to Complex Diseases. Animal Biotechnology, 385–404.* https://doi.org/10.1016/B978-0-12-416002-6.00021-3

Boratyn, G. M., Camacho, C., Cooper, P. S., Coulouris, G., Fong, A., Ma, N., Madden, T. L., Matten, W. T., McGinnis, S. D., Merezhuk, Y., Raytselis, Y., Sayers, E. W., Tao, T., Ye, J., & Zaretskaya, I. (2013). *BLAST: a more efficient report with usability improvements. Nucleic acids research, 41(Web Server issue), W29–W33.* https://doi.org/10.1093/nar/gkt282

Sievers, F., & Higgins, D. G. (2018). *Clustal Omega for making accurate alignments of many protein sequences. Protein science : a publication of the Protein Society, 27(1), 135–145.* https://doi.org/10.1002/pro.3290

UniProt Consortium (2008). *The universal protein resource (UniProt). Nucleic acids research, 36(Database issue), D190–D195.* https://doi.org/10.1093/nar/gkm895

Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., & de Hoon, M. J. (2009). *Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics (Oxford, England), 25(11), 1422–1423.* https://doi.org/10.1093/bioinformatics/btp163

Bunyat-zada, A. R., & Ross, A. C. (2023). *Highlights of bioinformatic tools and methods for validating bioinformatics derived hypotheses for microbial natural products research. Current Opinion in Chemical Biology, 76, 102367.* https://doi.org/10.1016/j.cbpa.2023.102367

Department of the Army, United States of America. (1977). *U.S. Army activity in the U.S. biological warfare programs: Volume 1 [Unclassified document]*. Retrieved from National Security Archive website:https://nsarchive2.gwu.edu/NSAEBB/NSAEBB58/RNCBW_USABWP.pdf

Lamanna, C., Eklund, H. W., & McElroy, O. E. (1946). *Botulinum toxin (Type A); Including a study of shaking with chloroform as a step in the isolation procedure. Journal of Bacteriology, 52(1), 1-13.* https://doi.org/10.1128/jb.52.1.1-13.1946

Jacobsen, A. (2014). *Operation Paperclip: The Secret Intelligence Program That Brought Nazi Scientists to America. Opensource.* Retrieved from https://archive.org/details/annie-jacobsen-operation-paperclip.-the-secret-

intelligence-program-that-brought/page/n207/mode/2up and https://www.archives.gov/files/iwg/declassified-records/rg-330-defense-secretary/foreign-scientist-case-files.pdf

Yale Law School. (n.d.). *Nuremberg Trial Proceedings Volume 21: Transcript of Two Hundred and Eleventh Day, Morning Session. Avalon Project.* Retrieved from https://avalon.law.yale.edu/imt/08-26-46.asp

Central Intelligence Agency. (1949). *The Bacteriological Research Institute on the Island of Riems.* Retrieved from https://www.cia.gov/readingroom/docs/THE%20BACTERIOLOGICAL%20RESEA%5B16208214%5D.pdf

Tatu, L., & Feugeas, J.-P. (2021). *Botulinum Toxin in WW2 German and Allied Armies: Failures and Myths of Weaponization. Eur Neurol, 84(1), 53–60.* https://doi.org/10.1159/000512812

Lebeda, F. J., Adler, M., & Dembek, Z. F. (2018). *Yesterday and today: The impact of research conducted at Camp Detrick on botulinum toxin. Military Medicine, 183(5-6), 85–95.* https://doi.org/10.1093/milmed/usx047

Wollinsky, K. H., Hülser, P. J., Brinkmeier, H., Aulkemeyer, P., Bössenecker, W., Huber-Hartmann, K. H., Rohrbach, P., Schreiber, H., Weber, F., Kron, M., Büchele, G., Mehrkens, H. H., Ludolph, A. C., & Rüdel, R. (2001). *CSF filtration is an effective treatment of Guillain-Barré syndrome: a randomized clinical trial. Neurology, 57(5), 774–780.* https://doi.org/10.1212/wnl.57.5.774

Brunt J, van Vliet AHM, Carter AT, Stringer SC, Amar C, Grant KA, Godbole G, Peck MW. Diversity of the Genomes and Neurotoxins of Strains of *Clostridium botulinum* Group I and *Clostridium sporogenes* Associated with Foodborne, Infant and Wound Botulism. *Toxins*. 2020; 12(9):586. https://doi.org/10.3390/toxins12090586

Adler, M., Capacio, B., & Deshpande, S. S. (2000). Antagonism of botulinum toxin A-mediated muscle paralysis by 3, 4-diaminopyridine delivered via osmotic minipumps. *Toxicon : official journal of the International Society on Toxinology*, *38*(10), 1381–1388. https://doi.org/10.1016/s0041-0101(99)00231-7

Costa, A. S., Guerini, F. R., Arosio, B., Galimberti, D., Zanzottera, M., Bianchi, A., Nemni, R., & Clerici, M. (2019). SNARE Complex Polymorphisms Associate with Alterations of Visual Selective Attention in Alzheimer's Disease. *Journal of Alzheimer's disease : JAD*, *69*(1), 179–188. https://doi.org/10.3233/JAD-190147

Popular science summary

The connection between botulism outbreaks (foodborne, wound, infant, iatrogenic, and adult intestinal toxemia) and the botulinum toxin of *Clostridium botulinum* is well-documented (Harris et al., 2020). In 1820, Justinus Kerner published his findings, connecting blood sausage as the origin of a paralytic illness with clearly expressed motor control symptoms. In December of 1895, Emilie van Ermengem investigated an outbreak of paralytic illness with origins of smoked and pickled ham, by isolation of bacteria, identifying bacteria as anaerobic, and proving the presence of it in both the origins and body of the victims. The bacteria was named *Bacillus botulinus* and recorded as a toxin-producing species. In 1917, the *anaerobic bacilli* were reclassified under the genus *Clostridium* and gained the name of *Clostridium botulinum* (Monheit & Pickett, 2017).

Clostridium botulinum is characterized by a Gram-positive, obligate anaerobic oval endospore-forming ability, and exhibits a slightly curved rod-shaped morphology. Typically, it is found as individual cells or in small chains, with flagella facilitating motility. This bacterium thrives in environments with limited oxygen, such as soil and aquatic ecosystems. It further demonstrates a preference for slightly alkaline conditions, reflecting its adaptation as a saprophytic organism engaged in decomposing organic matter. Isolation of bacteria is difficult and dangerous. It cannot be done in standard equipped laboratories by streak planting from enriched broth into liver egg yolk agar, or identification of strain by phenotypic observation of traits of organisms, growth rate, morphology, metabolic activity and presence of toxins (Erkmen, 2022).

Endospores are a result of bacteria sporulation. The formation of dormant spores is characterized by a spore coat; a thick peptidoglycan layer on top of the inner membrane protecting the genome. The spore is dehydrated and has extremely low metabolic activity, this allows the spore to survive without requiring nutrients for long periods of time (Portinha et al., 2022). The complete set of environmental

stimuli which induce sporulation is currently unclear, but some signals include starvation (da Silva et al., 2005), a vital attack (Makarova, et al., 2012) or abrupt oxygen change (Mearls et al., 2012). Once a stimuli event occurs, sporulation is irreversible resulting in a coordination of hundreds of genes. The onset of sporulation is governed by transcription factors in cooperation with sigma factors. The dormant spore is a circular structure of concentric layers with exteriors of proteins/glycoproteins (exosporium), a coat of proteins(ct) which shows resistance to lytic enzymes and chemicals, outer membrane of lipids and proteins, a thick cortex of peptidoglycan (Cx), inner membrane (IM), a germ cell wall (GCM) and a core containing chromosomes, enzymes, ribosomes, Ca2+ and calcium dipicolinic acid (CaDPA) which is used to stabilize DNA and proteins from any heat, chemicals, or radiation damage (Romero-Rodríguez et al., 2023). In laboratory conditions, spores are removed by sterilization with autoclaving at 121°C (250°F) for 20 minutes (Centers for Disease Control and Prevention, 2008).

The formation of spores ensures survival advantages for bacteria. Typically, the spores remain intact and transit through the bodies without causing harm, and are ultimately expelled through excretion. The spores of *Clostridium botulinum* are ubiquitously distributed in the environment and are frequently encountered in soil, sediments, and the intestinal tracts of various animals. Consequently, these spores can infiltrate the food chain through direct contamination of animals or plants during production processes, as well as through cross-contamination, during harvesting, handling, or processing stages. The distribution and prevalence of spore types exhibit geographical variability, with regions such as Central and South America, Europe, and Asia exhibiting contamination, particularly in proximity to lakes, rivers, coastal areas, and flood-prone regions. Sediment samples collected from various locations, including Alaska, Washington, the Baltic regions, China, and Taiwan have revealed that 30% to 95% of samples harbor *Clostridium botulinum* spores, predominantly strains of serotype E. Consequently, marine and freshwater foods are frequently contaminated with these spores. A survey in France identified that 7.8% of fish and shellfish utilized as refrigerated ingredients were positive for *Clostridium botulinum* type A and B, with a population of 2-3 spores/ kilogram (Glass & Marshall, 2013).

Clostridium botulinum serotypes are commonly found in the intestines of clinically healthy animals. Spores are anticipated find in biowaste, manure, slaughterhouse waste, and raw fertilizers. Furthermore, spores are even found on vegetables and fruit grown in proximity to the soil, as well as contamination agents in meat and dairy products.

Upon germination, the spores transition into dangerous metabolically active, vegetative cells, initiating growth processes and intracellular synthesis of botulinum toxin (BoNT), into a dangerous neurotoxin. The release of toxins serves as both a defensive strategy and a means to acquire nutrients. Notably, contamination with the toxin differs from contamination with the bacteria (Rawson et al., 2023).

The detection of bacterial contamination in food is typically indicated by alterations in texture, colour, scent, and flavour; however, these safety measures cannot be applied to contamination with toxins. Botulinum neurotoxin (BoNT) contamination presents a unique challenge as it lacks visually identifiable signs. BoNT toxin is known to form infected clusters; a product may exhibit contamination on one side while remaining safe for consumption on the other. However, in a liquid solution, the entire product is affected by the toxin. A prevalent misconception and concern of the public revolves around the belief that bulging and dented canned goods pose a risk of botulism infection. Incomplete cooking of the food may allow some bacteria to survive and proliferate within the cans. Numerous bacteria generate gas, thereby elevating pressure within the container. While abstaining from such products is sensible to ensure food safety in general, the presence of gas does not necessarily imply the existence of metabolically active vegetative cells of *Clostridium botulinum* or BoNT toxins in the product (Proydakov, 2001).

Exact identification of *Clostridium botulinum* species and presence of BoNT toxin is a laborious and resource intensive task. Primary problem lies in the rapid offset of symptoms from moment of intoxication and low volume of identifiable sample. Diagnostics of sample, takes longer time then available safe timeframe of progression of the symptomatic presentation of botulism in patient, coupled with high risk of inadequacy when working with low volume of toxin or bacterial DNA in the stool or serum tests. New rapid methods are needed for robust and clinically helpful diagnostics of botulism before patient shows symptoms that carry risk of damage to nervous system. Traditional laboratory methods for neurotoxin detection, such as mouse lethality assay, remain usable, although newer in vitro assays have been developed, yet low volume of pathogenic agents leads to a need to enrich the medium. Enrichment of medium is a time-consuming process. The culture growing methods for botulinum are underdeveloped, for the purpose of clinical diagnostics. Molecular techniques targeting neurotoxin genes are useful for detection but do not identify biologically active toxins (Lindström & Korkeala, 2006).

Direct toxic detection is achieved through either mouse assay or simultaneous screening in vitro assays such as ELISA, endopeptidase assays or lateral flow tests.

Bacterial presence can be identified through DNA extraction via PCR, incorporating an enrichment step for samples such as gastric fluid, intestinal content and environmental samples. Cultivation methods entail growth in either liquid medium (e.g., Tryptone-Peptone-Glucose-Yeast Extract) or solid agar (e.g., Egg Yolk Agar). Following cultivation, toxin presence is validated using Polymerase Chain Reaction (PCR) or mouse bioassay. Moreover, pure cultures may undergo additional testing, including toxin typing via PCR or genetic characterization employing Pulsed Field Gel Electrophoresis (PFGE) methodology (Lindström & Korkeala, 2006).

Differentiation diagnoses include Guillain - Barre syndrome, Miller-Fisher syndrome, chemical intoxication, stroke and staphylococcal food poisoning. Foodborne botulism analysis can be performed on serums, feces, gastric fluids, and foods suspected of contamination. Infant botulism can additionally be detected in intestinal contents and environmental samples. Wound botulism can be detected in serums, tissues, wound swabs and pus (Lindström & Korkeala, 2006).

The industrial processes of prevention of contamination of food with spores of *Clostridium botulinum* involve stringent hygiene practices during ingredient handling, industrial sterilization of canning containers (boiled for 7 - 10 hours before use), and employing sterilization (121^{o}C for 15 minutes) for additives and preservatives (Corsalini et al., 2021). The industry-preferred method for detecting the presence of metabolically active vegetative cells of *Clostridium botulinum* is through PCR which amplifies the number of DNA copies in the sample, combined with enzyme-linked immunosorbent assay (ELISA) for diagnostic purposes, and is widely used for good manufacturing practices (GMP) in the food industry (Notermans, 1999).

Because botulism is early offset in the clinical setting, treatment of symptoms starts before the origins of intoxication is determined while, in industrial settings, the laboratory might choose to use a simplified and fast *Clostridium* genus assay for quality control to have substantial basis for action to limit food borne outbreak in the industrial processes (Dembek et al., 2007).

BoNT is a protein complex with high specificity. All serotypes of the botulinum toxin have a heavy chain (HC) with a light chain (LC), bound with a single disulfide bond. The toxin chains have different functions, the heavy chain is responsible for binding to receptors on the surface of neuronal cells with a receptor binding domain and translocation domain. The light chain is the core of catalytic activity which is transferred into the cytoplasm of the host cell as a result of endocytosis (Wilkins, 2014).

BoNT toxin has regions which play a role in maintaining the structural integrity of the toxin and the proteins which interact with other proteins. Botulinum neurotoxin (BoNT) is a zinc dependent metalloprotease which acts by cleaving Soluble N-Ethylmaleimide-Sensitive Factor Attachment Protein Receptor (SNARE) proteins within postsynaptic nerve terminals (Lebeda et al., 2010). The function of SNARE proteins is to mediate the fusion of synaptic vesicles with the presynaptic membrane during the neurotransmitter release, playing a crucial role in communication between neurons. Consequently, intoxication with BoNT leads to neurotransmitter release inhibition, disrupting neural transmission to effector muscles (Ramakrishnan et al., 2012).

Bioinformatics plays important role in interdisciplinary *in silica* research, offering computational tools when handling large data volumes. This study highlights some of the bioinformatics methods and applies them for development of deeper insight into complex biological research on BoNT serotypes and BoNT toxin, by showing efficient storing tactics for genetic sequencing data (FASTA format) followed by a sequence comparison between strains BoNTA-G and exploration of protein structure.

Acknowledgements

My family,

Nathaniel and Liam.

Publishing and archiving

Approved students' theses at SLU are published electronically. As a student, you have the copyright to your own work and need to approve the electronic publishing. If you check the box for **YES**, the full text (pdf file) and metadata will be visible and searchable online. If you check the box for **NO**, only the metadata and the abstract will be visible and searchable online. Nevertheless, when the document is uploaded it will still be archived as a digital file. If you are more than one author, the checked box will be applied to all authors. You will find a link to SLU's publishing agreement here:

- https://libanswers.slu.se/en/faq/228318.

☒ YES, I/we hereby give permission to publish the present thesis in accordance with the SLU agreement regarding the transfer of the right to publish a work.

☐ NO, I/we do not give permission to publish the present work. The work will still be archived and its metadata and abstract will be visible and searchable.