Structure and activity of Clostridium botulinum neurotoxin functional fragments

Masuyer, Geoffrey

Award date:
2012

Awarding institution:
University of Bath

Structure and activity of *Clostridium botulinum* neurotoxin functional fragments

Geoffrey Masuyer

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

September 2011

Acknowledgements

Thank you to my supervisors Ravi Acharya and John Chaddock, for giving me the opportunity to work on this exciting project and for your constant support and guidance throughout my thesis.

My special thanks to Ravi, for sharing your knowledge and enthusiasm for crystallography, and for your support in every trips to the synchrotron.

I would like to express my deepest appreciations to everyone at Syntaxin Ltd, for sharing their knowledge of the botulinum neurotoxin and more importantly always welcoming me. It has been a real pleasure to work with all of you. I am particularly thankful to Peter James for his help with protein purification, Matthew Beard for his knowledge of biochemistry and his kind help with the biochemical assays, and Patrick Stancombe for his essential contributions to the study on self-activating molecules.

I would like to extend my thanks to everyone from lab 0.34, past and present, for making the last four years extremely enjoyable. I am particularly grateful to Nethaji Thiyagarajan for his help with computing and crystallographic software.

Thank you to all my friends and family who have encouraged me through the years. Above all, thank you Sarah, for never doubting and always encouraging me through your love, patience and understanding.

Merci à mes parents, qui m'ont toujours soutenu et encouragé, dans mes études et au-delà, et sans qui je ne serais pas là aujourd'hui.

Abstract

Botulinum neurotoxins (BoNTs) cause flaccid paralysis by inhibiting neurotransmission at cholinergic nerve terminals. BoNTs consist of three essential domains for toxicity: the cell binding domain (Hc), the translocation domain (Hn) and the catalytic domain (LC). The binding function of the Hc domain is essential for BoNTs to bind the neuronal cell membrane, therefore removal of the Hc domain results in a product that retains the endopeptidase activity of the LC but is non-toxic. Functional derivatives (LHn) of the parent neurotoxin composed of Hn and LC domains have been recombinantly produced and characterised. The crystallographic structures of LHn from serotypes A and B are reported here and demonstrate the stability of the LHn fragment in comparison to the full length toxins. The activity of LHn has been assessed on recombinant substrates and on cultured neuronal cells. LHn retains the capacity to internalise and cleave its intracellular SNARE substrate when applied to the cells at high concentration. These activities demonstrate the utility of engineered botulinum neurotoxin fragments as analytical tools to study the mechanisms of action of BoNT neurotoxins and of SNARE proteins. Targeted secretion inhibitors (TSI) are a new class of engineered biopharmaceutical molecules derived from the botulinum neurotoxins. These functional derivatives are expressed as single-chain proteins and require post-translational activation into di-chain molecules for function. A range of BoNT derivatives are presented and demonstrate the successful use of engineered SNARE substrate peptides at the LC-Hn interface to give these molecules self-activating capabilities while retaining the functions of LHn. Several novel molecules with therapeutic potential have been produced and their crystallisation for structural investigation is reported. These results provide an understanding of the structural implications and challenges of engineering therapeutic molecules that combine functional properties of the LHn fragment from BoNTs with specific ligand partners to target different cell types.

Table of Contents

List of figures

List of tables

List of abbreviations

A_{280}	Absorbance at 280 nm
ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
BIS-TRIS	Bis(2-hydroxyethyl)-amino-tris(hydroxymethyl)-methane
BoNT(/X)	Botulinum neurotoxin (/serotype)
BSA	Bovine serum albumin
C. argentinense	*Clostridium argentinense*
C. baratii	*Clostridium baratii*
C. botulinum	*Clostridium botulinum*
C. butyricum	*Clostridium butyricum*
C. difficile	*Clostridium difficile*
C. perfringens	*Clostridium perfringens*
C. tetani	*Clostridium tetani*
CDT	Cytolethal distending toxin
DHFR	Dihydrofolate reductase
DLS	Diamond Light Source
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EC50	Half maximal effective concentration
ECL	Enhanced chemiluminescent
ECL	*Erythrina cristagalli* lectin
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EGFR (ECD)	Epidermal growth factor receptor (extracellular domain)
eSCN	Embryonic spinal cord neuron

GFP	Green fluorescent protein
GST	Glutathione S-transferase
HA	Hemagglutinin
Hc	Binding domain (of BoNT)
HC	Heavy chain (of BoNT)
Hc_C	C-terminal domain of Hc
Hc_N	N-terminal domain of Hc
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Hn	Translocation domain (of BoNT)
IgG	Immunoglobulin G
IL-8	Interleukin-8
IPTG	Isopryl-1-thio-β-galactopyranoside
Km	Michaelis constant
LB	Luria Bertani
LC	Light chain (catalytic domain of BoNT)
LC(0)	Inactive LC
LHn(/X)	Fragment LC+Hn of BoNT (/serotype)
MBP	Maltose binding protein
MOPS	3-(N-morpholino)propanesulfonic acid
MR	Molecular replacement
NGF	Nerve growth factor
NTNH	Non-toxic non-hemagglutinin
P(r)	Distance distribution function
P. pastoris	*Pichia pastoris*
PAGE	Polyacrylamide gel electrophoresis
PEG	Polyethylene glycol
PIP	Phosphatidylinositol phosphate
rDNA	Ribosomal DNA

Rg	Radius of gyration
Rmsd	Root means square deviation
RNA	Ribonucleic acid
S. aureus	*Staphylococcus aureus*
S. pyogenes	*Streptococcus pyogenes*
SAXS	Small-angle x-ray scattering
scLHn	Single-chain LHn
SDS	Sodium dodecyl sulphate
SNAP	Synaptosomal-associated membrane protein (of 23 or 25 kDa)
SNARE	N-ethyl-maleimide-sensitive fusion protein attachment receptor
SV2	Synaptic vesicle protein 2
Syt	Synaptotagmin
TB	Terrific Broth
TeNT	Tetanus neurotoxin (or tetanospasmin)
TRIS	Tris(hydroxymethyl)aminomethane
TSI	Targeted Secretion Inhibitors
t-SNARE	Target SNARE
VAMP	Vesicle-associated membrane protein
VIP	Vasoactive intestinal peptide
Vmax	Maximum reaction rate
VPAC1 (ECD)	Vasoactive intestinal peptide receptor 1 (extracellular domain)
v-SNARE	Vesicle SNARE
WGA	Wheat germ agglutinin
Z	modified IgG binding domain of protein A

Chapter 1. Introduction

1.1. Overview of bacterial protein toxins

The idea that bacteria could produce poisonous substances was developed early in the realisation that microbial agents were responsible for infectious diseases. Klebs, first in the late 19th century, suggested the presence of toxic molecules secreted by staphylococci, although experimental evidence only came later. Pioneering work by Koch and Loeffler anticipated that soluble molecules were involved in the cholera (*Vibrio cholerae*) and diphtheria (*Corynebacterium diphtheriae*) infections, respectively (Alouf, 2006). As the understanding of infectious diseases expanded, protein toxins were discovered and their role in microbial pathogenicity put in evidence. Because of the threat they represent for human health, either as natural infectious agents or as biological weapons, protein toxins have been extensively studied. Careful analyses of the toxins' mechanism of action have also led to their usage as therapeutic agents and also as important tools for understanding certain biological processes (Schiavo and Van der Goot, 2001).

Diversity of activities

The diphtheria toxin was the first to be isolated and described by Roux and Yersin (1889) as a toxic soluble product of the diphtheria bacillus and responsible for the illness. It took several decades however to extract the toxin and characterise it as a protein with potent enzymatic activity. The diphtheria toxin is responsible for the inhibition of protein synthesis in eukaryotic organisms by catalysing the ADP-ribosylation of elongation factor-2 inside the cell cytoplasm, thereby leading to cell death (Collier, 1975). This was to be the first of many virulence factors with ADP-ribosyltransferase activity. These toxins have been shown to be produced by a wide range of pathogenic bacteria of both Gram-positive and Gram-negative types. Modification of their target protein by the conserved ADP-ribosylation function disrupts essential processes of the eukaryotic cells and provokes cell death (Deng and Barbieri, 2008).

Another major type of toxin that was among the first discovered consists of the clostridial neurotoxins, an array of zinc-metalloproteases expressed by *Clostridia sp.* causing paralysis by inhibition of neurotransmission (Turton *et al.*, 2002). Among the most potent bacterial toxins, the lethal toxin component of anthrax toxin from *Bacillus anthracis* is also a metalloprotease (Agrawal and Pulendran, 2004). Other important bacterial protein toxin functions with role in human pathogens include:

- phospholipases: cytolytic *Clostridium perfringens* α-toxin (Sakurai *et al.*, 2004);

- adenylate cyclases: *Bordetella pertussis* toxin, anthrax oedema factor (Ahuja *et al.*, 2004);

- RNA N-glycosidases: *Shigella dysenteriae* shiga toxins (Johannes and Römer, 2010);

- glucosyltransferases*: Clostridium difficile* toxins A and B (Davies *et al.*, 2011);

- deamidase: Escherichia coli cytotoxic necrotising factors (Lemonnier *et al.*, 2007);

- protease: *Staphylococcus aureus* exfoliative toxins (Ladhani, 2003), *Streptococcus pyogenes* cysteine proteinases;

- deoxyribonuclease: cytolethal distending toxins: cytolethal distending toxin (CDT) produced by a number of bacterial pathogens (Smith and Bayles, 2006).

Another class of toxin, the superantigens, does not present any enzymatic activity but has important functions in bacterial pathogenicity. They bind to the major histocompatibility complex class II and T-cell receptors. They are mainly produced by the Gram-positive *S. aureus* and *S. pyogenes* and are the causative agents of the toxic shock syndrome (Fraser and Proft, 2008).

Structural aspects

Common structural features can be identified in bacterial protein toxins despite the diversity of functions they display. In particular, toxins may be classified according to their activity, and more generally the localisation of their action, with some acting intracellularly, and others that targets the surface of the cytoplasmic membrane.

For toxins with intracellular targets, crossing of the cell membrane is an essential step of the intoxication process. Advances in the understanding of the mode of action of these poisonous molecules led to the general principle that most of these toxins possess two functional parts. The first component A is responsible for the intracellular enzymatic activity while B allows for specific cell binding and transport of A inside the cell (Menetrey *et al.*, 2005). These complex mechanisms regroup cell recognition, endocytosis, translocation and trafficking into the cytoplasm, with high specificity and potency for their receptors and substrates (Aktories, 2002). Over the last twenty-five years, determination by x-ray crystallography of the three-dimensional structure of many of these toxins has greatly contributed to the characterisation of their mode of action and biochemical properties. More particularly, it has highlighted the various possible domain organisations seen in the toxins. The A and B components can be part of a single-chain polypeptide with two distinctive domains (e.g. clostridial neurotoxins, diphtheria toxin), or can be two separate proteins either in an oligomeric form (e.g. cholera, shiga toxins) or as binary toxins. In the latter case,

the two different components associate at the target cell surface (e.g. *C. difficile* ADP-ribosylating binary toxin, Davies *et al.*, 2011).

Biological and medical applications

Bacteria have evolved to produce virulence factors with high potency and specificity towards eukaryotic organisms. These essential properties make bacterial protein toxins efficient biological instruments to analyse the cellular processes they target (Aktories, 2006). Many of the toxins have greatly contributed the discovery of important pathways in cell biology which in return have shed light on the toxin mode of action itself. This is particularly true for toxins with intracellular targets that have developed unique mechanisms to enter the cell without damaging the membrane (Schiavo and van der Goot, 2001). An evident example is the use of the tetanus toxin, a specific protease acting on a presynaptic membrane protein involved in vesicle formation, which helped drawing the mechanism underlying neurosecretion, and eukaryotic cellular secretion in general (Schiavo *et al.*, 1992). Likewise, pertussis and cholera toxins have been important in understanding the mechanism of adenylate cyclase activation and the role of cyclic AMP as a second messenger in eukaryotic cells. The toxins are responsible for ADP-ribosylation of G proteins that control these signal transduction pathways (Harnett, 1994; Neer, 1995).

A direct medical application for bacterial toxins is the production of effective vaccines against corresponding bacterial pathogens. Derivatives of some of the toxins, cholera and heat-labile enterotoxin for examples, have been incorporated as adjuvants into human vaccines. Adjuvants are pharmacological or immunological agents that can stimulate the immune system and improve immunogenicity of co-administered antigens in order to improve the efficiency of vaccines. These novel toxin-based adjuvants have shown significant effects on vaccine potency in preclinical models (da Hora *et al.*, 2011).

An alternative approach to use bacterial protein toxins for medical purposes is to focus on some of their specific properties. Since many of the toxins present a structure divided into functional compartments, they represent ideal subjects for protein engineering. An early application for retargeting toxin consisted in harnessing the activities of cytotoxins as potential therapies for certain cancers. Such toxins can be engineered as components of immunotoxins where the potency of the toxin is combined with the specificity of an antibody partner to target and kill cancer cells (Pastan *et al.*, 2006). This has been carried out with the diphtheria toxin and *Pseudomonas aeruginosa* exotoxin A. Their specific binding domains were replaced by antibody Fv fragments, thus allowing the targeting to a particular antigen at the surface of cancer cells, while retaining the toxin's ability to reach its intracellular domain and cause cell death by inhibition of protein synthesis. Current

immunotoxins are being assessed for the treatment of leukaemia, adenocarcinomas, and T-cell lymphomas among other diseases (FitzGerald *et al.*, 2004).

One of the toxins with the most potential has been the botulinum neurotoxin. With its potency and specificity to inhibit neurotransmission at the neuromuscular junction, it has been clinically used for a wide range of clinical applications. These applications and the alternative uses for BoNT in protein engineering are reviewed in details in section 1.4 of this chapter.

1.2. *Clostridia*

1.2.1. Microbiology and pathological association

The genus *Clostridium* consists of Gram-positive, rod-shaped bacteria in the Phylum Firmicutes (Figure 1.1). It contains over 100 different species which have been characterised and classified according to their 16s rDNA sequences. All species are strictly anaerobic and form endospores. *Clostridia* have been one of the most studied bacterial classes since the early identification of its role in a number of human diseases (Melville, 2007).

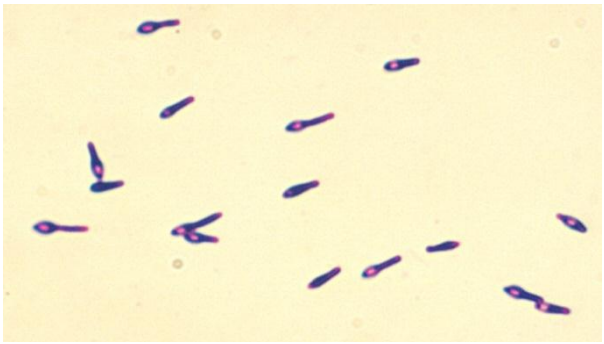

Figure 1.1. *Clostridium botulinum.* A photomicrograph of *C. botulinum* type A viewed using a Gram stain technique (from the Public Health Image Library, U.S). Gram-positive rod-shaped bacteria showing endospores.

Clostridia are ancient organisms that are widespread in the environment, more precisely in anaerobic habitats where organic compounds are present, including soils, and aquatic sediments. Many *Clostridia* are transient or permanent members of the normal flora of the gastrointestinal tracts of humans and animals (Melville, 2007). In most cases, *Clostridia* are opportunistic pathogens. All pathogenic clostridial species produce protein exotoxins that play an important role in their pathogenicity.

Several species are commonly listed as important human pathogens. *C. difficile* is a highly infectious bacterium that is increasingly being associated with hospital-acquired

infections. It is responsible for antibiotic-associated pseudomembranous colitis and involves several virulence factors that include glucosyltransferase toxins and the ADP-ribosylating binary toxin (Davies *et al.*, 2011).

C. perfringens is another human pathogen. It can produce up to 12 different toxins and is responsible for wound infections that may lead to gas gangrene. It produces a potent cytolytic α-toxin that hydrolyses phosphatidylcholine and sphingomyelin, leading to cell death (Sakurai *et al.*, 2004). Untreated infections are highly lethal. *C. perfringens* is also involved in cases of food poisoning and antibiotic-associated diarrhoea with strains producing a powerful enterotoxin. When released from the spores, the toxin binds to its receptor, claudin, at the epithelial cell junctions of the gut wall, and forms pores through the cell membrane (Briggs *et al.*, 2011).

Tetanus is an infectious disease caused by *C. tetani*, and is usually associated with wound contamination. The bacteria produce the tetanus neurotoxin (TeNT), or tetanospasmin. The toxin targets the central nervous system by inhibiting neurotransmission, thereby causing spastic paralysis and potentially leading to death. It binds to cholinergic motorneurons and is subjected to retrograde axonal transport to the spinal cord cells. It blocks neurotransmission by cleaving synaptobrevin, a protein involved in vesicular secretion (Poulain *et al.*, 2006). The structure of TeNT is closely related to that of the botulinum neurotoxin which is described in details in the following sections.

1.2.2. Botulism

Botulism

Botulism has been known for hundreds of years and was first described in the early nineteenth century by Justinus Kerner after a food poisoning outbreak that followed the ingestion of blood sausages. The illness was then named after the Latin term for sausage (*botulus*). It is a rare but serious paralytic illness caused by the botulinum neurotoxins (BoNTs), which are produced by *Clostridium botulinum* and other species under anaerobic conditions (Hatheway, 1990). The toxins are large proteins that act on cholinergic neuromuscular junctions and are responsible for inhibition of the neurotransmission. Botulism infections are characterised by progressive flaccid paralysis, firstly of the facial muscle. In most severe cases it can spread towards the limbs and cause respiratory failure, thereby leading to death.

There are three main clinical instances of botulism (Sobel, 2005). As it was first described in cases of food poisoning, one of the recurrent forms is foodborne botulism. This often results from contaminated food in which *C. botulinum* spores have been allowed to germinate in anaerobic conditions. This typically occurs in home-made preserves and fermented uncooked dishes where microbiological safety measures were not respected.

Clinical cases of wound botulism appear to be increasing (Yuan *et al.*, 2011). It follows infection of wounds caused by penetrating injuries where *C. botulinum* spores can germinate and produce the toxin in the anaerobic environment caused by abscesses. It is almost exclusively occurring among injection drug users and is the main form of botulism occurring in the UK (Brett *et al.*, 2004).

Intestinal (infant) botulism is the most common form of the disease and results from the absorption of toxin produced *in vivo* after colonisation of the intestines by *C. botulinum* in infants aged less than 1 year (Sobel, 2005). The lack of a fully competing intestinal gut flora allows for the colonisation of the *Clostridium* species from the ingestion of spores. Similar cases, although very rare, may happen in adults with a weakened intestinal microflora. The fatality rate for infant botulism is very low (< 1%) for hospitalised patients.

Other minor clinical cases of botulism include inhalation by laboratory worker handling the toxin, and a few cases of patients treated with botulinum neurotoxin injections for therapeutic purposes. Inhalation botulism is a cause of concern from the bioterrorism perspective since an aerosol form of the toxin could be a very efficient and lethal weapon. The lethal dose is predicted to be as little as 7 mg when ingested or 700 ng when inhaled for a 70 kg individual (Arnon *et al.*, 2001).

Different types of botulinum neurotoxins

Analysis of the strains responsible for the various outbreaks of botulism across the twentieth century led to the realisation that more than one toxin type was involved. Indeed, antitoxin serum prepared from the first described outbreak failed to react against some of the new cases (Hatheway, 1990). As new immunologically distinct strains were observed they were labelled from serotypes A to G in order of chronology. Epidemiologic data and later *in vitro* characterisation showed that only serotypes A, B, E, and F can cause human botulism (Coffield *et al.*, 1997), whereas serotypes C and D are prevalent in bird and cattle botulisms (Hedeland *et al.*, 2011).

Classification

C. botulinum strains are commonly divided in four different groups (groups I, II, III, and IV) based on physiologic characteristics (Collins and East, 1998). The toxins produced are ordered into seven serologically distinct groups (serotypes A to G), based on recognition by polyclonal serum (Hatheway, 1990). Further to *C. botulinum*, other *Clostridia* species have been shown to produce the botulinum neurotoxins, namely *C. butyricum* (BoNT/E), *C. baratii* (BoNT/F), and *C. argentinense* (BoNT/G) (Collins and East, 1998).

Each BoNT is encoded by an approximately 3.8 kb gene, which is preceded by a non-toxic non-hemagglutinin gene and several other genes that encode toxin-associated proteins (hemagglutinins HA-17, HA-33, HA-70, and p21, and/or p47) (Kubota *et al.*, 1998). The gene coding for BoNT serotypes A, B, E, and F are located on the bacterial chromosome. Strains producing serotypes C and D have the BoNT gene encoded on a phage genome. The gene for BoNT/G is present on a plasmid (Hill *et al.*, 2007).

C. botulinum strains producing cross-serotype chimera toxins, more particularly the C/D and D/C phage-encoded serotypes, have been described (Moriishi *et al.*, 1996). Moreover, several strains have been shown to produce multiple toxins. Multivalent strains, producing two toxins of serotypes A-B, B-A, A-F, and B-F, have been reported (Barash and Arnon, 2004).

Analysis of sequences from the toxin genes available from the databases has highlighted the significant sequence diversity within the BoNT subtypes in six of the seven serotypes (Smith *et al.*, 2005). The authors showed that the subtypes could differ from 2.5% to 30% at the amino acid level, and that these differences had effects on the binding and neutralisation by monoclonal and polyclonal antibodies. Further study into the genome of other *Clostridium* strains may reveal the presence more subtypes, although it is not known how the sequence variation may affect the toxins biochemical properties (Hill *et al.*, 2009).

Botulinum neurotoxin complexes

C. botulinum expresses the toxin with other non-toxic components and form large protein complexes called the progenitor toxin. The constitution of the progenitor toxin varies with serotype and can consist of a non-toxic non-hemagglutinin element (NTNH), and several hemagglutinin components (HA). The complex formed is not covalently linked but the protein interactions are stable at acidic pH. *C. botulinum* type A can produce several complexes known as LL (19S), L (16S) and M (12S) which range from 300 to 900 kDa resulting from the association of BoNT/A with several of the components mentioned above

(Sharma *et al.*, 2003). The most abundant component is the HA33-35 hemagglutinin which is involved in oligosaccharide binding.

The progenitor toxin was shown to increase oral toxicity of BoNT by protecting it from the acidic pH of the stomach and from intestinal proteases. Furthermore, it was suggested that the hemagglutinins may have a role in attachment to the intestine epithelial layer and thus supports internalisation of the toxin into the bloodstream (Inoue *et al.*, 2003). The progenitor toxin is therefore particularly important in foodborne botulism.

1.3. Botulinum neurotoxins

Botulinum neurotoxins (BoNTs) are the most powerful neurotoxins discovered to date and act by blocking the synaptic transmission at peripheral cholinergic nerve terminals. The seven serotypes of BoNT (A-G) are all expressed as single chain polypeptides of approximately 150 kDa. The toxins are later cleaved by clostridial or host proteases to its active form, a di-chain molecule that follows the A-B bacterial toxin motif with intracellular targets. The N-terminal catalytic light chain (LC), a zinc endopeptidase, is then linked by a single disulphide bridge to the heavy chain which is composed of the binding (Hc) and translocation (Hn) domains (Figure 1.2), both approximately 50 kDa (Turton *et al.*, 2002).

Figure 1.2. Domain organisation of botulinum neurotoxins. Each domain is approximately 50 kDa. LC (yellow) is an N-terminal zinc-endopeptidase domain linked to the heavy chain by a single disulphide bridge (marked). The translocation domain (Hn, green) is at the N-terminal of the heavy chain while the binding domain (Hc, blue) is at the C-terminus.

BoNTs bind specifically to the nerve terminals and are endocytosed into a vesicle where the acidic environment provokes a conformational change. This allows translocation of the BoNT catalytic component across the endosomal membrane and into the cytosol. The protease domain is then free to cleave one of the soluble N-ethyl-maleimide-sensitive fusion protein attachment receptor (SNARE) proteins (Figure 1.3). Impairing the function of the SNARE complex causes inhibition of neurotransmission, leading to flaccid paralysis and potentially death (Montal, 2010).

1.3.1. SNARE complex and exocytosis

The N-ethyl-maleimide-sensitive fusion protein attachment receptor (SNARE) proteins form a complex essential in the docking and fusion of synaptic vesicle (Südhof *et al.*, 2009). Each BoNT targets one of the SNARE proteins, with BoNT/A, /C and /E cleaving the synaptosomal-associated membrane protein of 25 kDa (SNAP-25); serotypes B, D, F and G are specific for the vesicle-associated membrane protein synaptobrevin (VAMP); syntaxin is cleaved only by BoNT/C (Schiavo *et al.*, 2000). It should also be added that BoNTs only cleave the SNARE proteins in their free form, before assembly of the complex. The SNARE complex was shown to be resistant to the proteolytic activity of the toxins as the recognition sites are made inaccessible (Hayashi *et al.*, 1994).

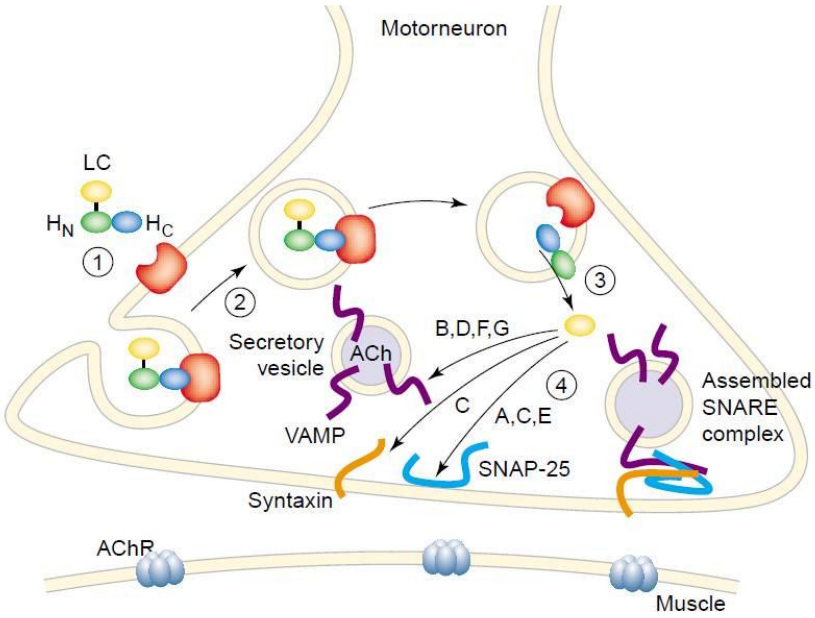

Figure 1.3. Mode of action of botulinum neurotoxins. Reproduced from Turton *et al.*, 2002. A four-step mechanism, (1) cell binding, (2) endocytosis, (3) translocation of LC in the cytosol, (4) LC cleavage of one of the SNARE proteins. (Ach, acetylcholine; AChR, acetylcholine receptor).

In neuronal exocytosis, syntaxin and synaptobrevin are anchored by their C-terminal domains in the cell and vesicular membranes, respectively. SNAP-25 is attached to the plasma membrane via several cysteine-linked palmitoyl chains. The SNARE proteins are usually divided into two broad categories, v-SNAREs and t-SNAREs after their cellular location, i.e. transport vesicles and target membranes respectively (Brünger, 2005).

The complex formed by the SNARE proteins folds into a four-α-helix bundle, where one α-helix is contributed by syntaxin-1 (residues 183-256), one α-helix by synaptobrevin

(residues 30-85) and SNAP-25 contributes two α-helices labelled sn1 (residues 7-83) and sn2 (residues 141-204) corresponding to the N- and C-terminal helices respectively (Sutton *et al.,* 1998) (Figure 1.4). These helical domains correspond to the SNARE motif, a structural feature common to all SNARE proteins. Furthermore, the SNARE complex is stabilised together by the hydrophobic core of the helices that expose either a central glutamine (Q) or an arginine (R) residue. This led to the reclassification of SNARE proteins into Q-SNAREs, with syntaxin and SNAP-25 providing Q-containing coils, and R-SNAREs, with synaptobrevin presenting the R-containing helix (Fasshauer *et al.,* 1998). The four helices are believed to interlock with a zippering action in the N- to C- terminal direction, thus bringing into juxtaposition the synaptic vesicle surface and the neuronal lipid bilayers. This action is mediated by the force applied by syntaxin and synaptobrevin on their transmembrane regions (Südhof *et al.,* 2009).

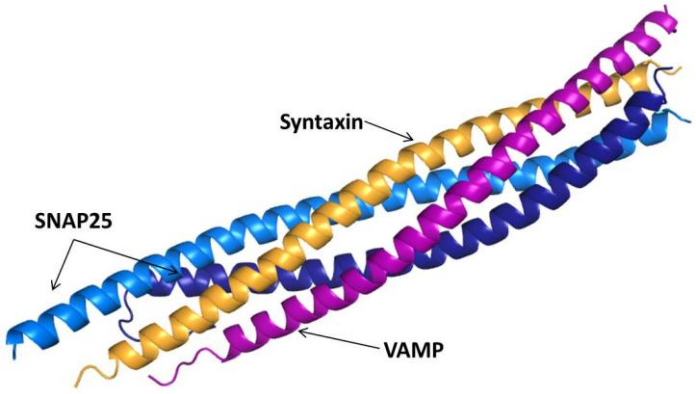

Figure 1.4. Crystal structure of a SNARE complex involved in synaptic exocytosis. (PDB 1SFC, Sutton *et al.,* 1998). Ribbon diagram representation of the four-helix bundle complex with VAMP in purple, syntaxin in orange and SNAP-25 (sn1 in light blue; sn2 in dark blue).

Noticeably, SNARE proteins are not only involved in all neurotransmission events, but it is also broadly accepted that they are essential in universally mediating membrane fusion processes. The SNARE proteins have ubiquitously expressed homologs in eukaryotic cells and tissues, which are implicated in various cellular processes, including cell growth, membrane repair, cytokinesis, and cellular secretion (Jahn and Scheller, 2006).

1.3.2. Crystal structure of botulinum neurotoxins

The first crystal structure of a full length botulinum neurotoxin to be determined what that of serotype A (Lacy et *al.,* 1998). It presented a linear arrangement of the three

functional domains (catalytic, translocation and binding) with the translocation domain holding a central location (Figure 1.5A). Swaminathan and Eswaramoorthy (2000) then described the high resolution structure (1.8Å) of BoNT/B which confirmed the modular arrangement of the toxin's structure. The two serotypes share 39 % identity (and 56 % similarity; Lacy and Stevens, 1999) and present common structural features. This is illustrated by the root mean square (rms) deviations between the binding, translocation, and catalytic domains of the superposed BoNT/A and BoNT/B which are 1.43, 1.56 and 1.43 Å, respectively (Swaminathan and Eswaramoorthy, 2000).

Structure of BoNT serotypes A and B

The catalytic domain (LC), a zinc protease of the thermolysin family, is a mixed bundle of α-helices and β-strands (Figure 1.5A). The translocation domain (Hn) is mostly helical with two long helices forming a coiled-coil fold. Part of Hn is also a large and unstructured loop region which had not been identified by primary sequence analysis. This region corresponds to approximately 100 residues on the N-terminus of Hn and include the cysteine bridge with LC. The crystal structure showed this loop is unique to BoNT in the way it surrounds and closely associates with LC, also preventing access to the catalytic cleft. It was termed "the belt" region (Lacy *et al.*, 1998). The binding domain (Hc) has limited interaction with Hn as it is angled away from it. Hc consists of two subdomains, Hc_N and Hc_C. Hc_N consists of 14 β-strands in a jellyroll motif, a fold usually associated with lectins, a class of sugar-binding proteins (Drickamer, 1988). The Hc_C domain is composed of loops and β-strands forming a β-trefoil fold which is the main region involved in the domain's binding function.

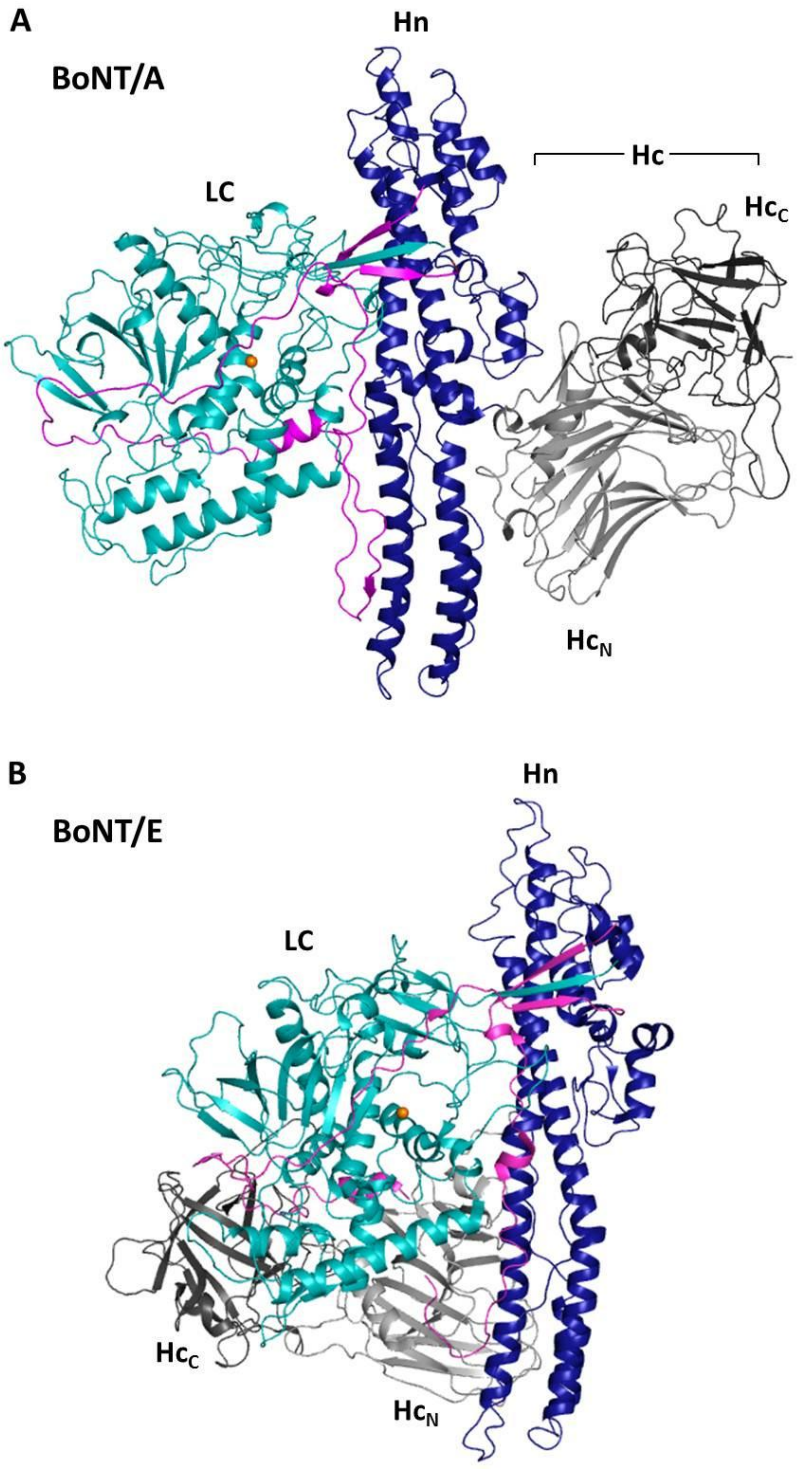

Figure 1.5. Crystal structure of botulinum neurotoxins. (A) Crystal structure of BoNT/A (PDB 3BTA, Lacy *et al.*, 1998). BoNT/A. Ribbon diagram representation with LC in cyan; Hn in blue; "belt" in pink, Hc in grey (Hc_N and Hc_C, light and dark respectively). Zinc ion shown as an orange sphere. **(B)** Crystal structure of BoNT/E (PDB 3FFZ, Kumaran *et al.*, 2009). Ribbon diagram representation with same colour as in (A).

Structure of BoNT/E

The level of sequence homology between the BoNTs (approximately 35%) would predict a conserved structure and domain organisation over the seven serotypes (Lacy and Stevens, 1999). However, structural analysis on BoNT/E led to the discovery of an alternative conformation. Fischer *et al.* (2008a) first observed the different molecular architecture of the holotoxin serotype E by single particle electron microscopy. The low resolution images showed a more globular fold and a likely rearrangement of the domains which was recently confirmed by the determination of the crystal structure by Kumaran *et al.* (2009) (Figure 1.5B). Interestingly, each individual domain has conserved its typical fold with the rmsd between the catalytic, translocation, and binding domains of the superposed E and B serotypes corresponding to 1.4 Å, 1.4 Å, and 1.9 Å, respectively. The main differences appear in unstructured loop regions and the belt which shows the highest sequence variation. In the BoNT/E conformation, Hc is located on the same side of the translocation domain as LC and thus this implies new inter-domain interactions, particularly between Hc and both Hn and LC.

Noticeably, although not as potent as BoNT/A, BoNT/E has been shown to act faster in the intoxication process with emphasis on the rate of translocation (Keller *et al.*, 2004; Wang *et al.*, 2008). Considering the unique arrangement of BoNT/E, Kumaran *et al.* (2009) suggested that LC was in a preferential position in relation to Hn which favoured the translocation process.

Overall, the crystal structure determination of the three holotoxins indicates that the general BoNT fold may be more diverse than expected. Furthermore, the modular aspect of the botulinum neurotoxins has allowed focusing on individual domains and their structural studies have provided essential information in the comprehension of the toxin's mode of action. These functions are described in details in the following section.

1.3.3. Mode of action

The botulinum neurotoxin is a single molecule capable of multiple functions to reach its aim, i.e. inhibition of neurosecretion. It first recognises and binds to cholinergic motor neurons with high affinity and specificity, and achieves entry into the cells where it can release a potent enzyme (Schiavo *et al.*, 2002). A step-by-step analysis allows a better understanding of the BoNT design.

Receptor-binding

The early findings on the binding mode of botulinum toxins to neuronal membranes led to the hypothesis of a double receptor formed by both a ganglioside and a protein component (Montecucco, 1986).

The neuron cell surface is particularly rich in gangliosides, a group of complex glycosphingolipids with poly-sialic acid moieties. Initial studies identified members of the G1b class as likely receptors, with GT1b and GD1b, tri- and di-sialo gangliosides respectively, presenting the highest affinities for BoNT binding (Montecucco, 1986). Kozaki *et al.* (1998) confirmed the importance of GT1b by inhibiting BoNT/A and /B activity on rat brain synaptosomes using a monoclonal antibody specific to this ganglioside. In an alternative approach, Yowler *et al.* (2002) demonstrated that gangliosides were essential for the binding of BoNT/A in ganglioside-deficient murine or human neuronal cells. The preferred ganglioside receptors for all of the serotypes have been identified. GT1b is recognised by all types but D, while GD1a and GD1b bind to most of the same serotypes with various affinities (Montal, 2010). Interestingly, BoNT/D presents unique neuron binding properties. While serotype D was first shown to lack the classical binding motif common to other BoNTs (Rummel *et al.,* 2004), it was recently demonstrated that it still has the ability to recognise b-series gangliosides, including GT1b and GD1b, via a different binding mode (Kroken *et al.,* 2011).

Although ganglioside binding is essential for the entry of BoNT in neurons, it is not sufficient. BoNTs exploit the synaptic vesicle recycling pathway further by also targeting receptor proteins at the surface of the presynaptic membrane (Montecucco, 1986). Synaptotagmin (Syt) was the first protein described to be involved in BoNT recognition, and more particularly of serotype B (Nishiki *et al.,* 1996). Synaptotagmins (Syt) are synaptic and secretory vesicle proteins that contain a single transmembrane region, and two C-terminal calcium-binding domains. Upon calcium binding, Syt triggers the SNARE fusion machinery in the events leading to exocytosis (Pang and Südhof, 2010). The protein receptor for BoNT/A was later identified as the synaptic vesicle protein 2 (SV2) by Dong *et al.* (2006). SV2 is a membrane glycoprotein that mediates calcium-stimulated transmitter release in neuronal and endocrine cells. Indeed, SV2 regulates the expression and trafficking of synaptotagmin in synapses (Nowack *et al.,* 2010). There seem to be a correlation between the sequence homology among the BoNT binding domains and their specific protein receptors, with serotypes A, E and F recognising SV2 (glycosylated in the case of E and F), while types B and G bind to synaptotagmins (I and II) (Binz and Rummel, 2009). The receptor for serotype C is not yet discovered while the role of SV2 in the binding of BoNT/D

remains unclear (Kroken *et al.,* 2011). Expectedly, the region within the protein receptors that is responsible for binding to BoNT is located in the vesicular lumen. For synaptotagmin, it was confirmed to be in the N-terminal ectodomain region for both serotypes B and G, in the segment which is closest to the membrane (Rummel *et al.*, 2007). Two crystal structures were simultaneously published of the BoNT/B binding domain in complex with a synaptotagmin II fragment (Chai *et al.,* 2006; Jin *et al.,* 2006). Both structures showed the SytII peptide adopting a α-helical conformation when bound to the Hc_C domain of the toxin (Figure 1.6).

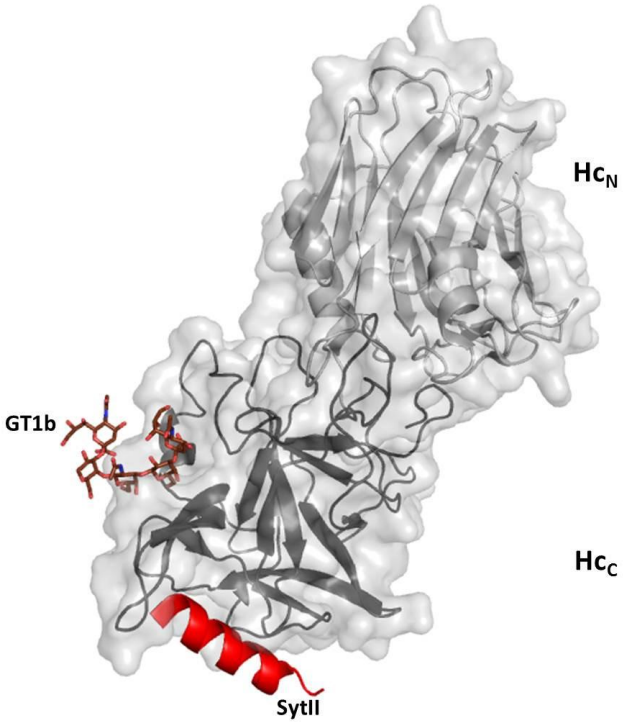

Figure 1.6. Model of the dual-receptor complex for botulinum neurotoxins binding to neuronal cells. Crystal structure of Hc/A in complex with GT1b (PDB 2VU9, Stenmark *et al.*, 2008) and crystal structure of Hc/B in complex with SytII (PDB 2NP0, Chai *et al.,* 2006) were superposed. Only Hc/B is shown as a ribbon diagram representation in grey (Hc_N and Hc_C, light and dark respectively) with SytII in red, and GT1b in brown stick representation.

The binding domain of BoNTs consists of two subdomains, N-terminal (Hc_N) and C-terminal (Hc_C) which are connected by a short helix (Figures 1.5, 1.6). The binding domains of all botulinum toxins share a similar fold (Swaminathan, 2011) and have been observed in the crystal structures of BoNT/A, /B and /E holotoxins as well as in the structures of Hc on its own for serotypes C (Karalewitz *et al.,* 2010), D (Strotmeier *et al.,* 2010; Kroken *et al.,* 2011), F (Fu *et al.,* 2009) and G (Schmitt *et al.,* 2010; Stenmark *et al.,* 2010). The structure of Hc/A in complex with the GT1b ganglioside (Stenmark *et al.,* 2008) complemented with

the structures of Hc/B bound to the SytII protein receptor (Chai *et al.,* 2006; Jin *et al.,* 2006) have provided essential information in understanding the detailed binding mechanism of BoNT (Figure 1.6). The two sites involved are located solely on Hc_C. The first site is a hydrophobic cavity constituted of a conserved lactose-binding motif (H…SXWY…G; Rummel *et al.,* 2004). In particular, residues Trp1266 and Tyr1267 of BoNT/A were shown to be directly in contact with GT1b. These two residues are highly conserved (except in BoNT/D, Kroken *et al.,* 2011) therefore all serotypes are likely to share the same ganglioside binding mechanism (Stenmark *et al.,* 2008). The protein receptor binding region is separated from the ganglioside site by a single loop. It consists of a hydrophobic pocket organised by two β-strands which are part of the β-trefoil fold. In both structures of the Hc/B – SytII complex, SytII residues 45 to 59 in one case (Chai *et al.,* 2006) and 47 to 58 in the second (Jin *et al.,* 2006) presented a α-helical conformation (Figure 1.6) while the rest of the peptide was not observed, and thus likely unstructured. With the difference in specificity observed among BoNTs, it is likely that this mode of protein interaction will vary from one serotype to another. However evidences suggest that the second binding site would correspond to the same, structurally conserved, hydrophobic pocket (Swaminathan, 2011). Interestingly, comparison of the structures of Hc with and without binding partners shows no visible conformational changes upon ligand binding. It demonstrates the rigidity of the β-trefoil motif and shows that the two binding sites, although in close proximity, do not interact with each other.

Despite the amount of structural information available, the role of Hc_N in the mechanism of BoNT intoxication remains unclear. Its jelly-roll fold is typical of lectins and other proteins involved in sugar-binding and protein interactions. A recent study showed Hc_N could bind to sphingomyelin-enriched membrane microdomains (Muraro *et al.,* 2009). More particularly, Hc_N could interact directly with phosphatidylinositol phosphates (PIP) which are molecules involved in vesicular traffic events. A potential binding site was identified from sequence comparison with PIP-binding protein. However this site would be located on the other side of Hc compared to the gangliosides and synaptotagmin binding region. Noticeably, PIP has been shown to promote the diphtheria toxin pore formation inside the membrane (Donovan *et al.,* 1982). Further work is necessary to understand the role PIP may have in the mechanism of BoNT either in terms of cell binding or mediation of membrane insertion (Muraro *et al.,* 2009).

Translocation

The only structural information available on the translocation domain comes from the crystal structures of the three holotoxins described previously (Swaminathan, 2011) (Figure

1.7). The most striking feature of Hn consists in a pair of long, anti-parallel and amphipathic, α-helices (approximately 100 Å in length) forming a coiled-coil fold. On both sides of these helices a shorter α-helix can be observed and packs in parallel to them. An unstructured region, traditionally associated with Hn from sequence analysis, forms an unusual loop which wraps around LC and was thus termed the "belt" (Lacy *et al.*, 1998). Hn shows a high sequence conservation among all the serotypes except for the belt region (Lacy and Stevens, 1999).

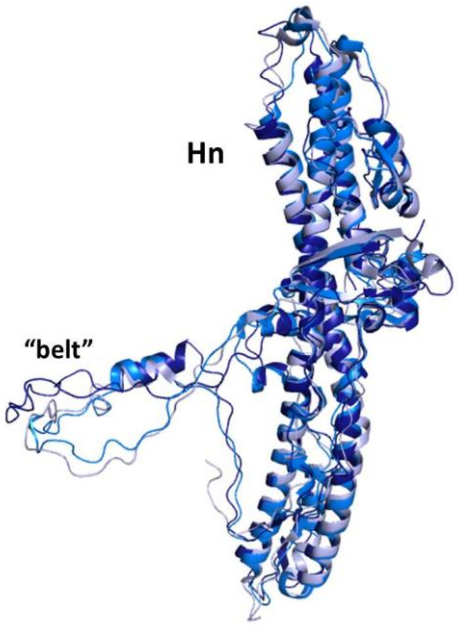

Figure 1.7. Structure of the botulinum neurotoxin translocation domain (Hn). Ribbon diagram representation of Hn in the crystal structures of BoNT/A (in dark blue, PDB 3BTA, Lacy *et al.*, 1998), BoNT/B (in marine blue, PDB 1EPW, Swaminathan and Eswaramoorthy, 2000) and BoNT/E (in light blue, PDB 3FFZ, Kumaran *et al.*, 2009).

The translocation domain of BoNT is structurally distinct from the other pore-forming toxins such as the diphtheria toxin and pseudomonas exotoxin that share a common structural motif (Parker and Pattus, 1993) which is not visible in the BoNT structures. The pore-formation mechanism necessary for translocation in BoNT intoxication remains elusive. The nature of the channel itself is not well understood and may involve a multimeric assembly of the toxin. A study by Schmid *et al.* (1993) showed a low-resolution electron micrograph of BoNT⁄B interaction with reconstituted vesicles in which it is suggested that the toxin forms a tetrameric channel. Further structural evidence has yet to confirm this analysis.

A channel is nevertheless formed and is expected to be approximately 15 Å wide (Koriazova and Montal, 2003) and thus not large enough to accommodate a translocating

light chain in its integral form. In order to explain the translocation process, Montal and collaborators have developed a single-molecule detection assay (Fischer and Montal 2007b). Their analysis with BoNT/A and a derivative molecule composed of LC and Hn have allowed determination of the minimum conditions required for translocation and deducted a step-by-step mechanism (Fischer *et al.,* 2008b). The conditions offered by the vesicle are as follow: (i) a pH gradient, going from a low endosomal pH (5.0) to a neutral cytoplasmic one; (ii) a redox gradient from an oxidising environment inside the endosome into a reducing cytosol; (iii) a transmembrane potential. The low pH in the endosome incites a conformational change which is likely to involve membrane insertion of Hn and the unfolding of LC (Koriazova and Montal, 2003). At this stage the single disulphide bond linking Hn to LC plays an essential role (Fischer and Montal 2007a). It stays intact up to the complete translocation of LC across to the intracellular milieu. Only then does the reducing environment allow release of LC in the cytosol. Interestingly the cysteine bridge is located at the C-terminal end of LC in all the serotypes. The unfolding and subsequent refolding of LC inside the cell is not fully understood but is essential to the intoxication process. A similar phenomenon has been observed in other bacterial protein toxins with translocation activity. The recent crystal structure of the anthrax protective antigen oligomers in complex with an unfolded lethal factor intermediate provided a unique insight into the mechanism of translocation-coupled protein unfolding (Feld *et al.,* 2010). There is no direct evidence available for the refolding mechanism of LC apart that it is active on its SNARE protein substrate once released and thus must be in its original conformation. The refolding is likely to involve intracellular chaperone proteins such that were reported for the diphtheria (Ratts *et al.,* 2003) and the ADP-ribosylating *C. botulinum* C2 toxins (Haug *et al.,* 2003).

While each domain of the botulinum neurotoxin has its own specific function, the intoxication process as a whole relies on complex interactions between the three functional elements so that the potent catalytic LC can reach its final intracellular destination.

Proteolysis

Once inside the cell the catalytic light chain is free to find its SNARE substrate. The location of the SNARE proteins differs, with VAMP attached to the vesicular membrane while SNAP-25 and Syntaxin are on the plasma membrane (Südhof *et al.,* 2009). Accordingly the various BoNT serotypes have adapted different means to access its target. A study on the cellular localisation of LC was performed with recombinant GFP-fusion molecules. This demonstrated the ability of LC/A to reside along SNAP-25 at the plasma membrane whereas LC/E was distributed in the cytoplasm (Fernandez-Salas *et al.,* 2004).

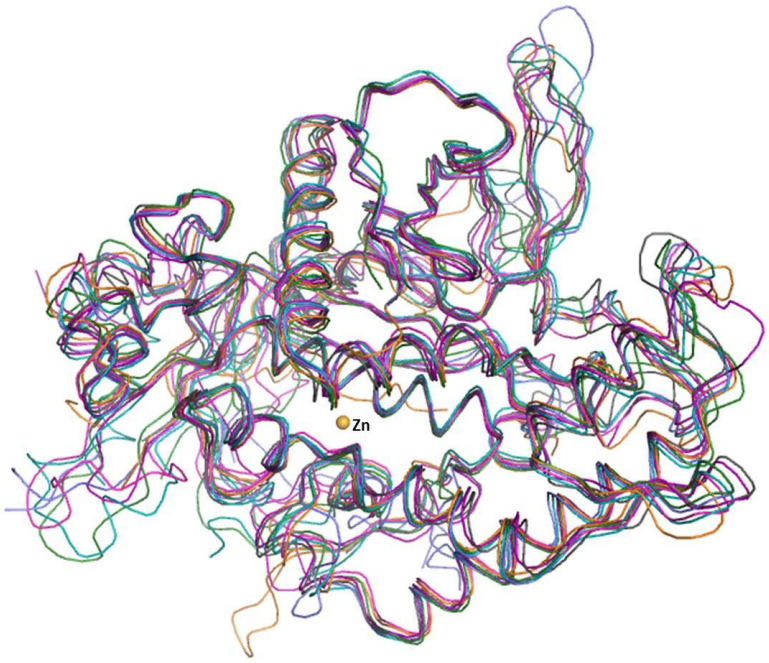

Figure 1.8. Structure of the botulinum neurotoxin catalytic domain (LC). Ribbon representation of LC of all serotypes; LC/A (in dark green, PDB 1XTF, Breidenbach and Brünger, 2004), /B (in light blue, PDB 2ETF, Swaminathan and Eswaramoorthy, unpublished), LC/C (in pink, PDB 2QN0, Jin et al., 2007), /D (in orange, PDB 2FPQ, Arndt *et al.*, 2006), /E (in purple, Agarwal *et al.*, 2004), /F (in grey, Agarwal *et al.*, 2005) and /G (in slate blue, PDB 3FFZ, Arndt *et al.*, 2005).

The catalytic light chain is a zinc protease with similarities to thermolysin (Lacy *et al.*, 1998). This domain has been particularly well characterised biochemically and structurally. Indeed the crystal structures of LC from all serotypes have been determined and are shown in figure 1.8 (Agarwal *et al.*, 2004, 2005b; Segelke *et al.*, 2004; Arndt *et al.*, 2005, 2006; Jin *et al.*, 2007). LC is constituted of a mixture of α-helices and β-strands and its individual fold is similar across all serotypes (which share approximately 35 % sequence identity, Lacy and Stevens, 1999), and also has the same overall conformation as the full length toxin. The active site is buried approximately 20 Å deep inside an open pocket with negative electrostatic potential. The zinc ion is coordinated by two histidines and one glutamate. The tetrahedral coordination, which is classical of a HEXXH zinc-binding motif seen in other proteases of the same family, is completed by a water molecule linked to the conserved glutamate of the motif and acting as the nucleophilic base. The residues involved in the catalytic pocket are particularly well conserved (Lacy and Stevens, 1999) and several mutation studies have highlighted their importance for the catalytic activity (Binz *et al.*, 2002; Agarwal *et al.*, 2005a). Furthermore, the structural studies demonstrated that the zinc ion was not involved in the site's architecture and thus only had a functional role (Breidenbach and Brünger, 2004).

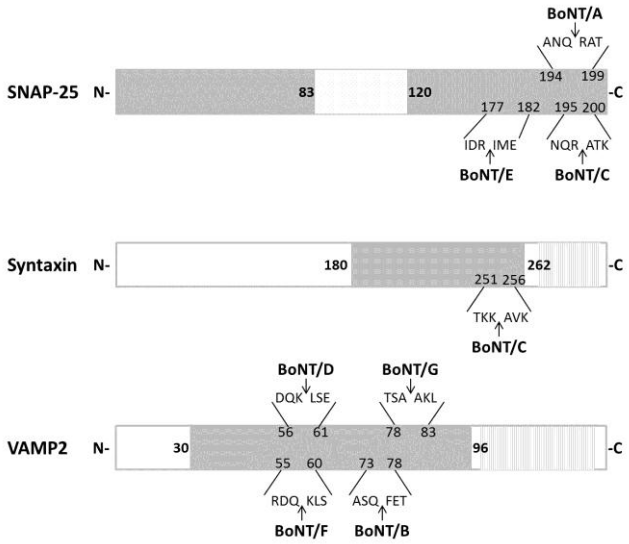

Figure 1.9. Schematic structure of SNARE proteins with BoNT cleavage sites. Schematic structures of SNAP-25, syntaxin, and VAMP2 and their respective cleavage sites by BoNTs (Schiavo *et al.,* 2000). The α-helical SNARE motif is shown in grey, the transmembrane domains in vertical stripes, region of SNAP-25 with palmitoylated cysteines in dots (responsible for membrane anchorage).

The botulinum neurotoxins are remarkably specific proteases. Except for serotype C which can cleave both SNAP-25 and syntaxin, each of the BoNTs only targets one of the SNARE proteins. Interestingly, each serotype cleaves its substrate at a different site (Figure 1.9). Early experimental evidence suggested the BoNT-SNARE interaction was unusual in that a long substrate was required for efficient proteolysis. Rossetto *et al.* (1994) put this hypothesis in evidence by identifying two sites essential for the enzymatic activity in the SNARE proteins. The first site was the target peptide bond. The second site was a region located further on the N-terminal side which was part of the α-helical SNARE motif and was involved in substrate recognition. The information provided by the crystal structure of enzyme-substrate complexes was essential in identifying the full extent of this interaction. Breidenbach and Brünger (2004) could solve the structure of the complex my mutating two key residues of the catalytic site. Thus, although it could not provide detailed information on the catalytic site, it demonstrated a complex binding mechanism with several exosites located around LC and away from the catalytic sites (Figure 1.10A). While SNAP-25 is known to be unstructured in its free state (Fasshauer *et al.,* 1997) and have a α-helical fold when in complex with its SNARE partners (Figure 1.4), it is mainly unstructured when bound to LC/A, with the exception of its N-terminus (i.e. residues 147-167) which binds at the α-exosite. SNAP-25 wraps around a crevice on LC/A's surface which is stabilised on its

C-terminus by a short β-strand forming the β-exosite. Binding of SNAP-25 causes the movement of flexible loops in proximity to the catalytic site.

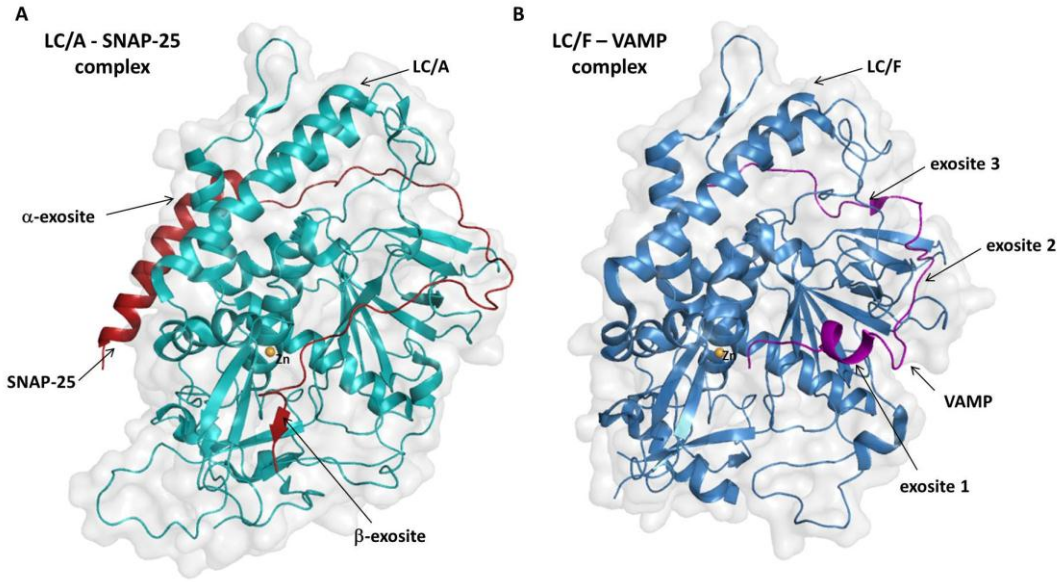

Figure 1.10. SNARE interaction with LC. (A) Crystal structure of LC/A in complex with SNAP-25 (sn2 segment, residues 146-204; PDB 1XTG, Breidenbach and Brünger, 2004). Cartoon diagram representation with LC/A in cyan and SNAP-25 in red. **(B)** Crystal structure of LC/F in complex with a VAMP peptide inhibitor (residues 22-58/Q58D-Cys 146-204; PDB 3FIE, Agarwal *et al.*, 2009). Ribbon diagram representation with LC/F in blue and VAMP in purple. Sites important for substrate binding are indicated.

The crystal structure of LC/A in complex with a short peptide inhibitor (QRATKM) allowed for a clear analysis of substrate binding at the catalytic site (Kumaran *et al.*, 2008). Furthermore, it confirmed the importance of several residues of the catalytic pocket that are not involved in the zinc coordination but are conserved among all BoNT serotypes (Binz *et al.*, 2002). Indeed, residues P1 (Gln197) and P1' (Arg198) present strong hydrogen bonds with Tyr366 and Arg363, respectively (Figure 1.11). A catalytic mechanism was deduced from the crystal structure of the complexes and mutational studies available. Residues Tyr and Arg help position and stabilise the substrate in the active site. The nucleophilic water attacks the carbonyl carbon of the scissile bond. A transition intermediate is formed that may be stabilised by Zn^{2+} and the tyrosine. Residue Glu 224 (LC/A) absorbs the transferred proton and mediates the formation of the amino group (Swaminathan, 2011).

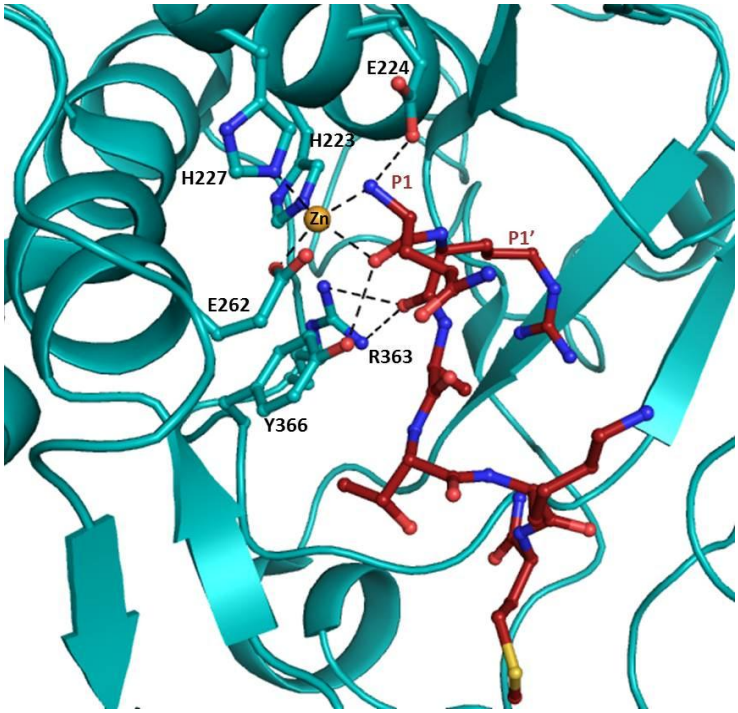

Figure 1.11. LC/A interaction with a SNAP-25 peptide inhibitor. Crystal structure of LC/A in complex with the SNAP-25 peptide QRATKM (PDB 3DDA, Kumaran *et al.,* 2008). Cartoon diagram representation of LC/A in cyan with residues involved in the catalytic site shown in stick and ball representation, QRATKM in red. Zinc ion shown as an orange sphere.

The mechanism of substrate binding and catalysis are expected to be the same for all serotypes. Extensive mutagenesis work on synaptobrevin allowed the identification of the regions involved in substrate binding for cleavage by BoNTs located on both sides of the scissile bond (Chen *et al.,* 2008; Sikorra *et al.,* 2008). The crystal structure of LC/F in complex with a VAMP-based peptide inhibitor (Agarwal *et al.,* 2009) confirmed this common mode of an extended substrate binding (Figure 5.10B). Although only a limited part of the peptide was visible, it clearly put in evidence several exosites that were different from the ones observed in the LC/A – SNAP-25 complex. Therefore the diversity of the exosites dictates the specificity of each serotype for its substrate.

Interestingly, the period of activity of BoNT is different for each serotype and has been linked to the intracellular longevity of the protease domain (Forlan *et al.,* 2003). The study showed that while BoNT/A can be active for several months, BoNT/E only lasts a few days. The mechanism underlying these differences and the spectacular longevity of BoNT/A is not well understood. It likely involves several factors that allow LC/A to be resistant to intracellular degradation pathways (Montal, 2010). The intracellular localisation of LC/A on the plasma membrane was proposed as a potential explanation for its longevity (Fernandez-

Salas *et al.*, 2004). Two sites involved in the LC/A interaction with the membrane were identified. A recent study confirmed one of these sites, a di-leucine on the C-terminus of LC/A, to be essential in the resistance of LC/A to degradation (Wang *et al.*, 2011). Understanding the longevity of LC/A offers interesting prospect for protein engineering on BoNT in order to optimise its therapeutic properties.

Functional role of the belt

One of the most interesting observations from the crystal structures of the holotoxins is the belt, an extended loop of Hn that wraps around LC and is common to all serotypes (Montal, 2010). The belt is the primary interface between LC and Hn and thus is expected to play a key role in the intoxication process. This role is however not well understood. The most interesting observation was made after determination of the LC/A – SNAP-25 crystal structure (Breidenbach and Brünger, 2004), which is that the belt in BoNT/A actually occupies the same crevice as SNAP-25 at the surface of LC (Figure 1.12).

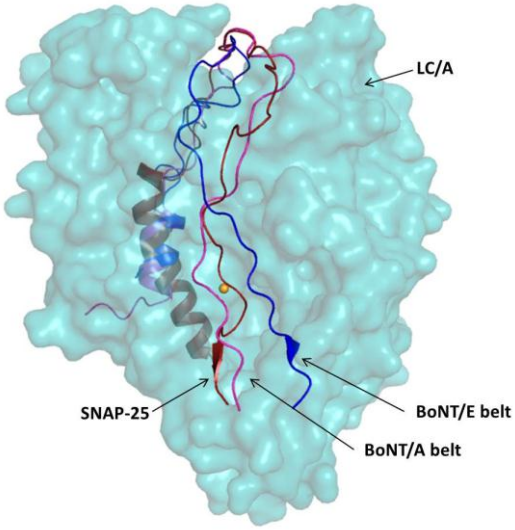

Figure 1.12. Superposition of the belt region with SNAP25. Superposition of the crystal structures of LC/A in complex with SNAP-25 (PDB 1XTG, Breidenbach and Brünger, 2004), BoNT/A (PDB 3BTA, Lacy *et al.*, 1998) and BoNT/E (PDB 3FFZ, Kumaran *et al.*, 2009). Surface diagram representation of LC/A in cyan, ribbon diagram representation of SNAP-25 in red, .BoNT/A belt (pink) and BoNT/E belt (blue).

The belt thus seems to play a role of pseudo-substrate and proteolysis inhibitor by preventing binding of the SNARE substrate when the toxin is in its di-chain form. The belt lacks the scissile bond and thus is not cleaved by LC/A (Brünger *et al.*, 2007). Information provided by the crystal structures of BoNT/B and LC/F-VAMP show that a similar relationship exist for the other serotypes (Swaminathan, 2011). Further functional roles for

the belt have also been investigated which particularly look at the translocation process. Brünger *et al.* (2007) proposed a role as a chaperone for protein-assisted unfolding of the LC protease. In another study, it was shown that lowering the pH neutralised acidic residues of the belt region and thus eliminated repulsive electrostatic interactions between Hn and the membrane, consequently allowing its penetration into the membrane (Galloux *et al.,* 2008). Further work is necessary to fully determine the role of the belt in the translocation and whether it helps initiates or control this process.

1.3.4. Inhibition of the botulinum neurotoxins

The botulinum neurotoxins are among the most poisonous protein toxins and as such have been feared for their use in biologic warfare and terrorism. Despite this threat there is still no efficient drug available for treating or preventing the potentially lethal botulism.

So far, a pentavalent toxoid vaccine composed of formalin inactivated botulinum neurotoxin has been used as the botulinum neurotoxin vaccine (Smith, 2009). The decreasing immunogenicity of the toxoid vaccine and the recent production of non-toxic subunit immunogens have led to the development of novel botulinum vaccines that are based on recombinant fragments of the heavy chain. More particularly a recombinant botulinum vaccine based on the cell binding domain (Hc) is being tested in human clinical trials (Staats *et al.,* 2011).

An alternative approach is to find small molecule inhibitors able to prevent or stop intoxication. The recent advances in understanding the BoNT mechanism of action, and the determination of the crystal structures described previously, have allowed the development of molecules targeting specifically one of the BoNT functions: cell binding, translocation or proteolysis. While Hc has been the main target for antibody recognition, it also represents a potential target for small molecule inhibition. For example, Eswaramoorthy *et al.* (2001) presented crystallographic evidence for the binding of doxorubicin (a DNA intercalator) at the ganglioside binding site. More recently, toosendanin, a natural molecule found in the bark of the tree *Melia toosendan* has been reported to block the translocation channel in a single-current assay (Fisher *et al.,* 2009). Finally, inhibitors of the catalytic domain have been investigated. Early studies focused on heavy-metal chelators with cell permeability ability, including hydroxamate zinc-binding molecules (Thompson *et al.,* 2011). However such molecule may interfere with other host proteases (Dickerson *et al.*, 2006). Peptide and peptidomimetics based on the SNARE substrates were showed to bind efficiently to LC, particularly the heptapeptide QRATKML. Thus smaller peptides and derivatives are being investigated (Zuniga *et al.*, 2010). Although they are potent ligands, these molecules are likely to be limited by their serotype specificity. The crystal structures of all LCs should help

the development on future generations of potent inhibitors that are active on all BoNT serotypes.

1.4. Biological and pharmaceutical applications of the botulinum neurotoxins

1.4.1. Current clinical applications

Despite their high toxicity, various preparations of BoNTs are available commercially for the treatment of diseases and conditions involving neuromuscular or autonomic neuronal transmission where at safe doses they can be used as therapeutic and cosmetic agents.

The first clinical use of a botulinum neurotoxin was for the treatment of strabismus in humans, where purified BoNT/A was injected (Scott, 1980). Following this first step, BoNT/A was used for the treatment of dystonia, a neurological movement disorder. The Food and Drug Administration (US) approved BoNT/A as a drug for these specific disorders in 1989. Since then, it has received approval for a growing number of diseases, including cervical dystonia, hyperhidrosis and more famously, for glabellar wrinkles. Its clinical use has been extended to over 100 conditions (Barnes *et al.,* 2007). Beyond its effect in neuromuscular pathologies, the toxin has also shown interesting analgesic activities. Although the underlying mechanism is not understood, it has led to the utilisation of the toxin for various pain-related conditions such as myofascial pain syndrome and migraine (Moore *et al.,* 2007). The mode of delivery consists simply in direct and localised intramuscular injection of a saline solution containing the diluted toxin. This allows for very little side effects, although some individuals may develop an immune response after repeated injections. A single injection of BoNT/A is usually effective for 3-6 months (Barnes *et al.,* 2007).

BoNT/A and /B are the most potent serotypes and have been the only ones approved for clinical use until now. Several formulations are commercially available in various parts of the world, with BOTOX® (Allergan Inc., US) and DYSPORT® (Ipsen Ltd, UK) corresponding to serotype A, while BoNT/B is distributed as MYOBLOC® and NEUROBLOC® in the US and Europe respectively (Solstice Neuroscience Inc., Ireland). Serotypes C and F have also been investigated for clinical applications (Eleopra *et al.,* 1997; Mezaki, 1995) but have not been commercialised.

1.4.2. Engineering botulinum neurotoxins

The botulinum neurotoxins are powerful molecule with already a wide range of therapeutic applications. This is the result of a single molecule with three domains responsible for specific functions i.e. cell binding, translocation inside the cells, and cleavage of intracellular substrate. The modular arrangement of BoNT and the natural range of biochemical properties provided by the seven serotypes have made it a molecule of choice for protein engineering. Several approaches have been considered which consist in enhancing the toxin's capabilities or retargeting one of its functions, thus leading to novel BoNT-derived applications.

Enhancing the botulinum neurotoxin activity

BoNT targets cholinergic motor neurons and its substrate is specifically involved in neuronal transmission, which limits the effect of the toxin to neuronal targets. SNARE-mediated secretion is a universal mechanism in eukaryotic cells (Südhof *et al.*, 2009), it however involves various SNARE protein homologs which are not all susceptible to BoNT cleavage (Schiavo *et al.,* 2000). Therefore, attempts to retarget BoNT to non-neuronal targets may require a modification of its enzymatic specificity. Chen and Barbieri (2006) designed a BoNT derivative that could cleave both SNAP-25 and SNAP-23. SNAP-23 is a non-neuronal isoform that is involved in the secretion process of various cell types, thus mediating the release of mucus, antibodies and many hormones (Jahn and Scheller, 2006). The novel design consisted in a single mutation of the BoNT/E light chain (K224D) (Figure 1.13A). This engineered molecule was able to cleave intracellular SNAP-23 when delivered to cultured human epithelial cells, consequently inhibiting the secretion of mucin and IL-8. This study illustrated the potential for extending the therapeutic application of BoNT to a wider range of hypersecretory disorders.

One of the most striking differences between the serotypes is in the speed and duration of intoxication (Forlan *et al.*, 2003). For example, although both BoNT/A and /E cleave SNAP-25, serotype E was shown to block neurotransmission faster and more potently. Wang *et al.* (2008) first identified the domains of BoNT responsible for these properties by engineering chimera proteins consisting of mixed fragments from serotypes A and E. They demonstrated that the speed of translocation was dependent on Hn while the potency and duration of toxicity came from LC. The same group later managed to design a novel molecule that combined the resistance of LC/A with the potency of LC/E (Wang *et al.*, 2011). The engineered protein consisted of LC/E fused to the amino terminus of a modified enzymatically inactive BoNT/A (Figure 1.13B). The novel molecule was able to target

neurons and demonstrated a potent and persistent cleavage of SNAP-25, thus providing an optimised BoNT derivative.

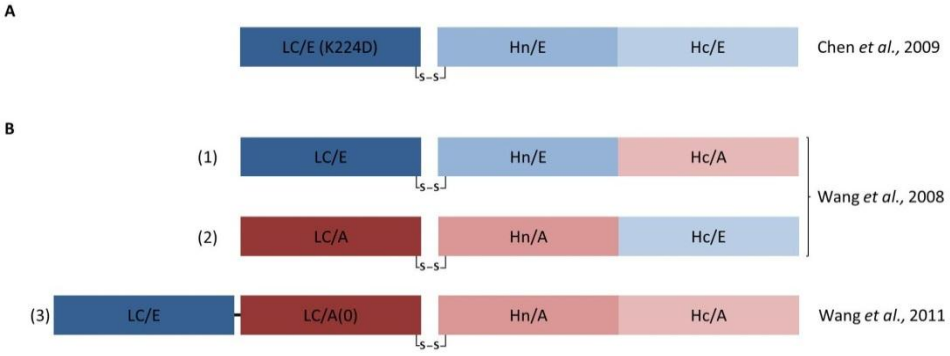

Figure 1.13. Engineering of BoNT for enhanced activity. (A) Domain organisation of modified BoNT/E(K224D) able to cleave both SNAP-25 and SNAP-23 (Chen and Barbieri 2006). **(B)** Domain organisation of engineered (1, 2) BoNT/A-/E chimeras (Wang *et al.*, 2008); and (3) a novel construct combining an inactive BoNT/A (designated here as "LC/A(0)") with LC/E and that have the potency of LC/E and the longevity of LC/A (Wang *et al.*, 2011).

The botulinum neurotoxin as a protein transporter

The first investigation in the use of BoNT to transport a cargo protein different from LC into the cytosol came from the need for intracellular BoNT inhibitors. Goodnough *et al.* (2002) designed a delivery machinery which consisted in the heavy chain of BoNT/A (HC/A) covalently attached to a chimeric glycoprotein with a 10-kDa amino dextran (Figure 1.14A). Thus HC/A served for specific neuronal cell targeting and internalisation of the complex while the dextran represented an inert carrier of BoNT/A antagonist. The cellular uptake of this molecule was successfully demonstrated using fluorescent markers and confirmed BoNT could be used for the intracellular transport of proteins.

Further attempts were made to use BoNT as a protein transporter and helped determine the conditions required for successful intracellular delivery. Bade *et al.* (2004) fused various proteins to the amino terminus of the full-length BoNT/D (Figure 1.14B) and assessed the delivery efficiency of this system using the neurotoxin's enzymatic activity. Cargo proteins consisting of LC/A, a dihydrofolate reductase, and luciferase were successfully translocated inside neurons. A less efficient translocation rate was however observed for the luciferase- and a GFP-fusion protein. This was linked to the structural constraints associated with the translocation process in which the cargo protein needs the ability to unfold in order to cross the narrow pore formed by BoNT. This study demonstrated

the potential of BoNT for the delivery of active enzymes inside neurons, and thus as a biological tool to investigate intracellular processes.

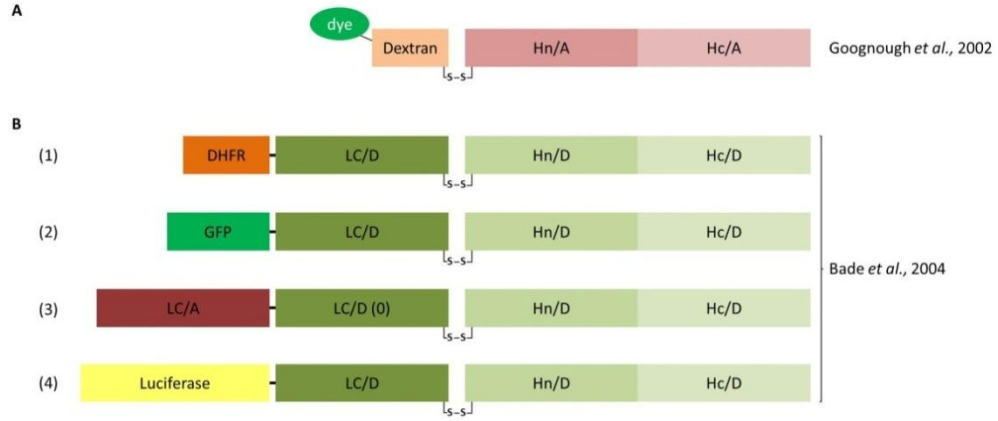

Figure 1.14. Engineering of BoNT as a cargo protein. (A) Domain organisation of a BoNT/A derived molecule for transport of a BoNT inhibitor (dextran, attached with a reporter dye; Goodnough *et al.*, 2002). **(B)** Domain organisation of engineered BoNT/D derivatives for intracellular transport of (1) dihydrofolate reductase (DHFR); (2) Green fluorescent protein (GFP); (3) LC/A; (4) firefly luciferase (Bade et *al.*, 2004).

The LHn fragment

Trypsinisation of BoNT/A led to the discovery of a fragment composed of the LC and Hn domains only, termed LHn/A (Shone *et al.*, 1985). This fragment was resistant to light trypsin treatment and could be recovered from the holotoxin. Additionally it was cleaved into a di-chain by the exoprotease thus mimicking host protease activation necessary for BoNT activity. The LHn fragment retains the catalytic activity of BoNT/A and the ability to form pores through membranes under acidic conditions (Shone *et al.*, 1987). The fragment was not toxic due to the loss of Hc and thus the lack of a cell targeting domain (Figure 1.15A). This fragment was later produced by recombinant expression as a GST-fusion protein and showed catalytic properties similar to the previously prepared fragment. LHn/A was active on neuronal cells when added at high concentration (Chaddock *et al.*, 2002). Further work on the recombinant production of LHn led to the development of optimised construct for serotypes A, B and C. These novel constructs were codon optimised for expression in *E. coli* and included a MBP-fusion tag as well as a protease cleavage site between the LC and Hn domains. These molecules were shown to be suitable for vaccine preparation and potential tools to investigate SNARE-mediated events (Sutton *et al.*, 2005). The LHn fragment represented a first step in the development of retargeted toxins for therapeutic use.

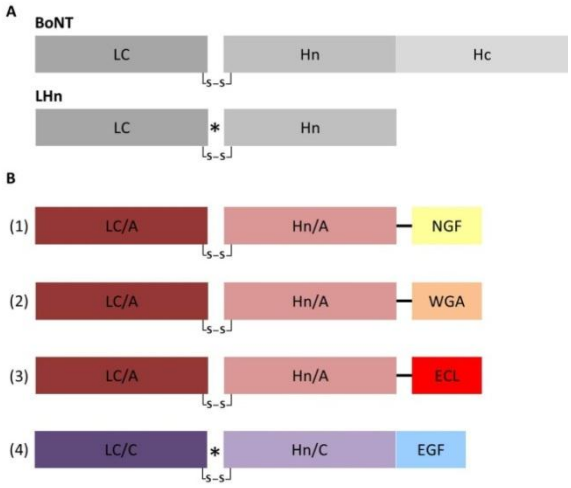

Figure 1.15. Engineering of BoNT for targeted secretion inhibition. (A) Domain organisation of BoNT and the LHn fragment, star marks site of engineered exoprotease cleavage site produced by recombinant expression (Chaddock *et al.*, 2002). **(B)** Domain organisation of engineered LHn derivatives as targeted secretion inhibitors (1) NGF-LHn/A (nerve growth factor, Chaddock *et al.*, 2000a); (2) WGA-LHn/A (wheat germ agglutinin, Chaddock *et al.*, 2000b); (3) ECL-LHn/A (*Erythrina cristagalli* lectin, Duggan *et al.*, 2002); (4) Recombinant EGF-LHn/C (epidermal growth factor, Foster *et al.*, 2006).

Engineering targeted secretion inhibitors

LHn is a fragment that combines the translocation and catalytic functions of BoNT. By associating LHn with various ligands, it was proven possible to inhibit cellular secretion of several cell types, neuronal and non-neuronal, by cleavage of their intracellular SNARE proteins (Foster, 2009).

The principle for retargeting BoNT derivatives was first demonstrated in two studies in which LHn/A, purified from BoNT/A, was chemically conjugated to the nerve growth factor (NGF) (Chaddock *et al.*, 2000a) and the wheat germ agglutinin (WGA) (Chaddock *et al.*, 2000b) (Figure 1.15B). In the first case, NGF-LHn/A was shown to inhibit the release of noradrenaline in neuronal cells, in correlation with intracellular SNAP-25 cleavage. The second investigation added incremental interest in that WGA-LHn/A not only inhibited neurotransmission in several neuronal cell types but could also inhibit insulin secretion from a pancreatic cell line. The potential of this technology was therefore demonstrated in neuronal and non-neuronal cells.

Further work looking into possible therapeutic application for retargeted LHn derivatives led to the design of several novel molecules. With the proven length of BoNT/A activity in clinical practices, LHn derivatives could prove a great relief in the treatment of chronic hypersecretory disorders. LHn/A was chemically conjugated to a lectin from a tree,

Erythrina cristagalli (ECL) (Figure 1.15B), which was shown to specifically bind nociceptive afferent neurons (Duggan *et al.*, 2002). This molecule was tested both *in vitro* and *in vivo*, demonstrating its ability to successfully inhibit substance P release in several models of pain. This was the first example of specific retargeting of LHn to a particular cell type, albeit a neuronal one.

To extend the principle of specific targeting to other cell types, it is essential to consider the LHn fragment from alternative serotypes. Indeed the specificity of BoNT/A for SNAP-25 limits its use to neuronal cells and cells of a neuroendocrine origin. Using the endopeptidase domain from other BoNTs should allow intracellular cleavage of SNARE isoforms involved in non-neuronal cell secretion (Jahn and Scheller, 2006). The production of a recombinant molecule associating LHn/C to the epidermal growth factor (EGF) represented another step in the development of targeted secretion inhibitors (Foster *et al.*, 2006) (Figure 1.15B). This molecule was able to inhibit *in vitro* the release of mucin in a human pulmonary epithelial cell line, thus showing a therapeutic potential in the treatment of chronic obstructive pulmonary disease. Furthermore it provided the proof of principle that the BoNT retargeting technology was efficient on non-neuronal cell lines that do not rely on SNAP-25 in its secretory pathway.

1.5. Aims

The botulinum neurotoxin is a powerful molecule that inhibits neurotransmission at the cholinergic nerve terminal. The modular architecture of BoNT has been highlighted, with each domain responsible for a specific function (Montal, 2010). It was demonstrated that a fragment of BoNT consisting of the catalytic and translocation domains (LHn) could be associated with novel ligands to bring the toxin's activity to newly targeted cells, leading to the development of a new class of therapeutic products (Foster, 2009).

The LHn fragment represents the essential framework on which these novel engineered molecules are based. The structure and function of the LHn fragment from several serotypes was therefore analysed. This was first carried out by the optimised production of various recombinant LHn constructs. The purified proteins were prepared for investigation by x-ray crystallography as a mean to solve their three-dimensional structure. Additionally, analysis of LHn's structure by alternative methods such as small angle x-ray scattering also provided useful information. Furthermore, the functionality of the LHn fragment was analysed by considering its proteolytic activity on recombinant SNARE substrates and in neuronal cell cultures. Such information should prove essential in understanding the behaviour of LHn as a basis for future optimisation of therapeutic molecules.

In the context of biopharmaceutical production, the LHn fragment is particularly challenging, owing to the necessity of obtaining an active di-chain form (Wey *et al.*, 2006). Thus proteolytic activity of BoNT towards SNARE peptides was investigated as a tool for engineering both self-activating and LC-activating molecules. A SNARE peptide region was inserted between the LC and Hn domains in several LHn constructs. The activity and structure of these novel molecules were investigated, providing valuable information to the design of self-processing pharmaceutical products.

Finally, several molecules representing LHn-based potential targeted secretion inhibitors were purified in order to be studied by x-ray crystallography. This led to preliminary crystallisation successes which may offer a basis for future structural investigations.

Chapter 2. Materials and methods

2.1. Chemical and reagents

All chemicals and reagents were purchased from Sigma-Aldrich unless otherwise stated.

2.2. Bacterial strains and plasmids

Competent cells

The *E. coli* One Shot TOP 10 Chemically Competent cells (Invitrogen) were used for cloning and plasmid propagation.

Several strains of *E. coli* were used for protein expression, including One Shot BL21 (DE3) Chemically Competent cells (Invitrogen), and Origami 2 (DE3) Competent cells (Novagen).

Expression plasmids

All the plasmids used for cloning were kindly provided by Syntaxin Ltd.

The plasmids used for expression of *C. botulinum* neurotoxin derivatives consisted in modified pET vectors (Novagen). These vectors carry an N-terminal pelB signal sequence for potential periplasmic localisation, plus optional C- or N-terminal His-Tag sequences. The plasmids contain an antibiotic resistance gene (either ampicillin or kanamycin) and a T7 promoter region for protein expression.

Other plasmids used for expression included modified pMal from New England BioLabs (NEB) and pGEX vectors (GE LifeSciences). The pMal vector encodes for maltose-binding protein (MBP) as a cleavable N-terminal tag. The pGEX vector encodes for glutathione S-transferase (GST) as a cleavable N-terminal tag. Both these vectors have an ampicillin resistance gene and rely on the P_{tac} promoter for protein expression.

The epidermal growth factor receptor extracellular domain (EGFR ECD) was cloned into a pPICZα vector from the EasySelect *Pichia* Expression Kit (Invitrogen). This vector encodes for a zeocin resistance gene, and a signal peptide for extracellular secretion of proteins expressed in *P. pastoris* and promoted by the AOX1 promoter.

2.3. Media and supplements

Competent cells for subcloning were grown in Luria Bertani (LB) medium which contained 10 g tryptone, 5 g yeast extract and 10 g sodium chloride per litre of distilled water. Medium was sterilised by autoclaving prior to use.

LB agar plates were prepared as per LB medium with the addition of 12 g agar per litre. After sterilisation by autoclaving, the medium was allowed to cool down before being supplemented with the appropriate antibiotic, and poured in Petri plates. Plates were stored at 4°C until use.

Expression cells were grown in Terrific Broth (TB) medium which contained 12 g tryptone, 24 g yeast extract, 4 ml glycerol, 9.4 g dipotassium hydrogen phosphate and 2.2 g potassium dihydrogen phosphate per litre of distilled water. Medium was sterilised by autoclaving prior to use.

During bacterial growth, cultures were supplemented with the appropriate antibiotic. Stock solutions of ampicillin (100 mg/ml) and kanamycin (30 mg/ml) were prepared and used to supplement culture medium at a 1/1000 dilution.

2.4. Molecular biology

2.4.1. Agarose gel electrophoresis

Agarose gel electrophoresis was utilised to analyse DNA samples. Gels of 1 % agarose were prepared by dissolving agarose in TRIS acetate EDTA (TAE) (40 mM TRIS acetate, 1 mM EDTA). Samples were prepared with SYBR Safe DNA gel stain (Invitrogen) before being loaded onto the agarose gel. NEB DNA ladders were used as markers (1 kb or 100 bp as appropriate). Electrophoresis was carried out using Bio-Rad apparatus. Visualisation was performed under UV light on an InGenius system (Syngene).

2.4.1. Cloning

Synthetic genes

All the designed synthetic genes were purchased from GeneArt and kindly provided by Syntaxin Ltd. The sequences for all proteins had been obtained from the UniProt Knowledgebase (UniProtKB). Sequences were back-translated and the corresponding genes optimised for the expression host codon usage (Entelechon tool). Restrictions sites were added as appropriate for sub-cloning.

Plasmid isolation

For plasmid DNA propagation, 10 ml cultures of LB supplemented with the appropriate antibiotic were inoculated from a glycerol stock and grown overnight at 37°C in a shaking incubator (200 rpm). Cells were harvested and plasmid DNA was then isolated from the cultures using a PureYield Plasmid Miniprep System (Promega) following the manufacturer's instructions.

Restriction digestion

DNA vectors and fragments were digested using the appropriate restriction endonucleases. A general overview of the cloning cassettes is presented in figure 2.1. All restriction enzymes were purchased from NEB and used with the supplied digestion buffers according to the manufacturer's instructions. When the restriction endonucleases could be optimally used together, double digestions were set up. All digestions were carried out at 37°C for 2 hours.

Double digestions were set up simultaneously on the plasmid containing the fragment to insert and the one carrying the vector. Each digestion reaction contained 30 µl DNA, 2.5 µl of each restriction enzymes, 5 µl buffer (10X), 5 µl Bovine Serum Albumin (BSA, 10X), 5 µl water (all reagents from NEB).

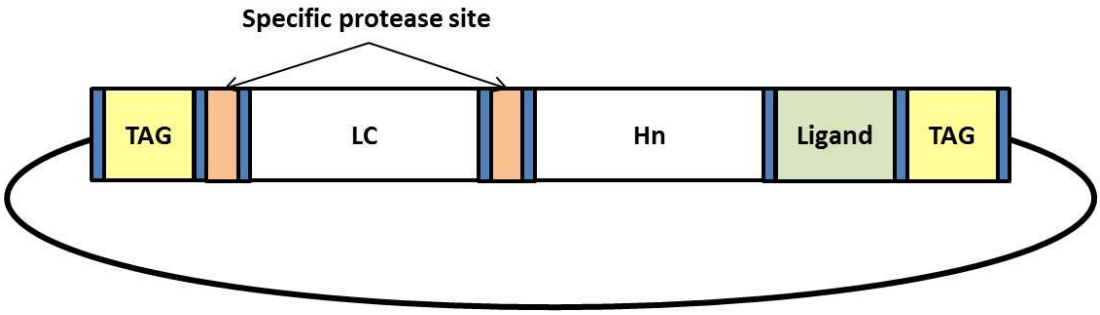

Figure 2.1. Schematic of cloning cassette used for BoNT derivatives. Regions coding for protein domains and endonuclease restriction sites are represented. LC and Hn of BoNT are white rectangles, N- and/or C-terminal tag in yellow, specific protease sites in orange (Factor Xa/Enterokinase), C-terminal ligand in green, all endonuclease restriction sites in blue. Details of the constructs used are given in the following chapters.

For VIP-LHn/A, the pET vector carrying LHn/A and the vector with the VIP gene fragment were double digested with *Nde*I and *Xba*I.

The plasmid carrying the VPAC1 extracellular domain and the modified pMal and pGEX vectors were digested with *EcoR*I and *Hind*III.

The plasmid with the EGFR gene and the pPIZa vector were digested with *Eco*RI and *Not*I.

The reaction samples were loaded onto 1 % agarose gel for analysis. The digested DNA fragment and vector required were cut from the gel and extracted with a DNA gel extraction kit (QIAGEN).

Ligation

Restricted DNA fragment and vector were ligated using T4 DNA ligase (NEB). Each ligation reaction was carried out in a total volume of 20 µl and contained 3 µl vector DNA, 13 µl fragment DNA, 2 µl T4 ligase buffer (10X) and 2 µl T4 ligase. The reaction was incubated for 2 hours at room temperature before being used for transformation in cloning competent cells.

Transformation

Ligated DNA was used to transform TOP10 competent cells (Invitrogen) according to the manufacturer's instructions. The competent cells (50 µl) were thawed on ice from their -80°C storage. For each experiment 5 µl of ligation mixture was added to the cells and incubated for 10 minutes on ice. Samples were then heat-shocked at 42°C for 90 seconds and incubated back on ice for 10 minutes. Then 150 µl of pre-warmed SOC medium was added to the reaction and incubated for 1 hour at 37°C in a shaking incubator (200 rpm). Reactions (50 – 150 µl) were plated on LB agar (with antibiotic) and incubated overnight at 37°C.

Purified DNA plasmids were used to transform expression competent cells (BL21 (DE3), Invitrogen; Origami 2 (DE3), Novagen) according to the manufacturer's instructions. The competent cells (50 µl) were thawed on ice from their -80°C storage. For each experiment 5 µl of plasmid DNA was added to the cells and incubated for 30 minutes on ice. Samples were then heat-shocked at 42°C for 30 seconds and incubated back on ice for 5 minutes. Then 250 µl of pre-warmed SOC medium was added to the reaction and incubated for 1 hour at 37°C in a shaking incubator (200 rpm). Reactions (50 – 200 µl) were plated on LB agar (with antibiotic) and incubated overnight at 37°C.

Screening for positive clones

After ligation and transformation into cloning competent cells, resulting transformants were tested for the presence of the desired inserts. Individual colonies of transformants were used to inoculate 10 ml LB cultures (with antibiotic) and their plasmid DNA extracted as

described previously. A test digest was performed on the resulting DNA. Each test reaction contained 6 µl DNA, 1 µl buffer (10X), 1 µl BSA (10X), and 1 µl of each restriction enzymes (same as previously). After 1 hour incubation at 37°C, reactions were analysed by agarose gel electrophoresis. Samples showing DNA fragments at the expected size were sent for sequencing (Geneservice, Oxford, UK).

Glycerol stock preparation

After being transformed with the sequence-verified plasmid DNA, cells were plated on LB agar with the appropriate antibiotic and incubated overnight at 37°C. A colony was picked and used to inoculate 5 ml LB broth (with antibiotic) and the culture was grown at 37°C (shaking) until the absorbance at 600 nm (OD_{600}) reached between 0.6 and 0.9. For each glycerol stock, 500 µl of culture were added with 500 µl of 50 % glycerol in a cryovial containing ceramic beads (Mast Group) and frozen at -80°C for storage. Primary cultures were inoculated using beads from these stocks.

2.5. Protein methods

2.5.1. SDS PAGE electrophoresis

At every steps of the protein purification process, samples were analysed by SDS-PAGE. Fractions of interests (column loads and wash, elution) were prepared by dilution with NuPAGE LDS Sample Buffer (4X, Invitrogen) with or without 1 mM DTT for oxidised and reduced samples respectively. Samples were denatured by heating for 5 minutes at 95°C before loading onto a 4-12 % BIS-TRIS gel (Invitrogen). Depending on samples, 5 to 15 µl were loaded. BenchMark molecular weight markers (Invitrogen) were used as standards. Gels were run for 45 minutes at 200 Volts in MOPS buffer (Invitrogen) using a Novex system (Invitrogen).

Gels were washed 3 times 5 minutes in distilled water. Staining was carried out using Simply Blue Safestain (Invitrogen) according to the manufacturer's instructions. After staining gels were washed with distilled water. Gel scanning and densitometry (when performed), were done using a Dyversity system and the GeneTools software (Syngene).

2.5.2. Protein expression

Typically, 100 ml primary cultures of TB (in 500 ml conical flasks) supplemented with the appropriate antibiotic were inoculated from a bead of the glycerol stock and grown overnight at 37°C in a shaking incubator (200 rpm). The following day, 10 ml of these cultures were used to inoculate 1 L secondary cultures of TB (with antibiotic) and grown at

37°C in a shaking incubator (200 rpm) until OD_{600} reached between 0.6-0.9. Protein expression was then induced by addition of isopryl-1-thio-β-galactopyranoside (IPTG) to a final concentration of 1 mM after the incubation temperature was lowered to 16°C. Cells were grown overnight and harvested by centrifugation. For storage, cell pastes were resuspended in 15 ml 50 mM HEPES pH7.2, 200 mM sodium chloride and frozen at -80°C until day of use.

2.5.3. Protein purification

Details of the purification methods are given in each chapter for the various proteins purified. A summary of the common experimental procedures used is given below.

Cell preparation

Stored cell pastes (corresponding typically to 10-15 g) were thawed and resuspended in approximately 100-150 ml of buffer used for the first purification stage. Cell lysis was carried out using a cell disruptor (TS Series, Constant Systems) at 20,000 Psi at 4°C. Lysate was centrifuged at 18,000 g for 30 minutes at 4°C to separate the clear soluble fraction. The resulting supernatant was used for further processing by loading onto an appropriate chromatography medium.

Protein purification

Each protein was purified according to its biochemical properties and the presence of affinity tags. Typically, the process started by an affinity chromatography stage, followed by treatment with a selective exoprotease for tag removal or protein activation. Details of the chromatography media used are detailed in the following chapters. In all cases, purifications were carried out using the manufacturer's instructions for each column and media. Table 2.1 summarises the proteins purified for this study and the techniques that were used. All proteins were purified using AKTA systems (GE Healthcare).

When dialysis was performed between protein purification stages, the pooled eluate fractions were set to dialyse in a SnakeSkin Dialysis Tubing (Thermo Scientific) with an appropriate molecular weight cut-off) against 10 L of buffer used for the following purification step and left overnight at 4°C.

Specific protease treatment

Before addition of the specific exoprotease for treatment, protein concentration was assessed by measuring absorbance at 280 nm (A_{280}). All LHn-derived molecules have an extinction coefficient such that an absorbance of 1 (A_{280}) corresponds to approximately 1

mg/ml. Enterokinase (NEB) was added at 6.4 ng and Factor Xa (NEB) at 10 μg units per mg of protein. PreScission protease (GE Healthcare) was used according to the manufacturer's instructions. Reactions were incubated overnight at 25°C unless otherwise stated.

Protein concentration determination

Final concentration of the pure proteins was determined by A_{280} measurements using a NanoDrop 2000 instrument (Thermo Scientific). For each samples, absorbance was averaged from three readings at an appropriate dilutions (so that A_{280} is between 0.1 and 1.0). The specific extinction coefficient derived from the protein sequence was used in the calculations.

Table 2.1. Summary of proteins purified.

	Proteins	Purification stages				
		I	II	III	IV	V
Proteins for crystallisation	LHn/A	AC(His)	F. Xa	AC(His)	HIC	
	LHn/B	AC(His)	F. Xa	HIC		
	LHn/C	AC(His)	F. Xa	AC(His)	IEX	
	LHn/D	AC(His)	Ek	HIC		
	scLHn/D	AC(His)	HIC			
	LHn/E	AC(His)	HIC			
	LC/A-SNAP23-Hn/A	AC(His)	HIC			
	LC/A-SNAP25(0)-Hn/A	AC(His)	GF			
	LC/B-GS-Hn/B	AC(His)	F. Xa	HIC		
	ZZ-LHn/A	AC(His)	Ek	AC(His)		
	VIP-LHn/A	AC(His)	F. Xa	HIC		
	VPAC1 ECD	AC(MBP)	F. Xa	AC(MBP)	AC(His)	
	EGF-LHn/A	AC(His)	Ek	AC(His)	HIC	GF
	EGF-LHn/C	AC(His)	Ek	AC(His)	HIC	
	EGF-LHn/D	AC(His)	Ek	AC(His)	HIC	
Other proteins	VAMP1-GFP	AC (GST)	PreSc.	AC(GST)		
	VAMP2-GFP	AC (GST)	PreSc	AC(GST)		
	VAMP3-GFP	AC (GST)	PreSc	AC(GST)		
	SNAP25-GFP	AC (GST)	PreSc	AC(GST)		
	scLHn/A	AC(His)				
	LC/A-SNAP25-Hn/A	AC(His)				
	LC/A-VAMP2-Hn/A	AC(His)				
	LC/A-VAMP2	AC(His)				
	LC/B-SNAP25					

AC, affinity chromatography using His-, GST- or MBP- tags; Exoprotease treatment with F. Xa, Factor Xa, Ek, enterokinase, PreSc., PreScission; HIC, hydrophobic interaction chromatography; IEX, ion exchange chromatography; GF, gel filtration chromatography.

2.6. Biochemical assays

2.6.1. Western blotting

Samples were prepared as per SDS PAGE analysis, with the addition of a MagicMark protein ladder (Invitrogen) for molecular weight standards. After electrophoresis, gels were set up in a blotting cassette with filter papers and either nitrocellulose of PVDF membranes (Millipore). Blotting was carried out in SDS running buffer with 20 % methanol for 1 hour at 40V, 100 mA. The blots were then blocked for 1 hour in 3 % Marvel milk powder in PBS with 0.001% Tween (PBS Tween). Primary antibodies were diluted appropriately in 3 % Marvel milk powder in PBS Tween and added to the blots. After incubation for 1 hour at room temperature (shaking), blots were washed three times 10 minutes in PBS Tween. Secondary (peroxidase conjugated) antibodies were diluted appropriately in 3 % Marvel milk powder in PBS Tween and added to the blots with a 1 hour incubation period. After a washing step, blots were revealed using an enhanced chemiluminescent (ECL) detection system (Thermo Scientific).

Details of the antibodies used for specific detection are given in the following chapters.

2.6.2. Protein N-terminal sequencing

Samples prepared for N-terminal sequencing were run on SDS-PAGE and blotted on an ImmobilonP membrane (Millipore) as described previously. Blots were stained by Ponceau S (Sigma) for 5 minutes and rinsed in water. The bands corresponding to the proteins to be analysed were cut and sent to AltaBioscience (UK) for sequencing by Edman degradation.

2.6.3. Substrate cleavage assays

Purification of recombinant SNARE substrates

Clones corresponding to recombinant SNARE-GFP fusion proteins were kindly provided by Syntaxin Ltd. The soluble region of SNAP25 (2-206), VAMP-1 (2-96), -2 (1-94) and -3 (2-77) were cloned to include an N-terminal cleavable GST tag and a C-terminal GFP tag. Protein expression and cell preparation were carried out as described previously. Cell pastes were resuspended and lysed in 50 mM HEPES (pH 7.2), 0.2 M sodium chloride (buffer A) with 5 mM DTT added freshly before use. Cell lysate supernatant was loaded onto a Glutathione Sepharose 4B column (GE Healthcare) and washed with buffer A. Elution was carried out with 100 mM reduced glutathione (in buffer A). The pooled eluate

was treated with PreScission protease (GE Healthcare) for tag removal and dialysed against buffer A (three times 1 L buffer A for 3 hours at 4°C). The dialysate was loaded onto the same affinity chromatography media and the unbound fraction collected. Sample was then concentrated using Vivaspin 5000 MWCO concentrator (Millipore). Figure 2.2 presents a summary of VAMP2-GFP purification by SDS PAGE analysis. Final protein concentrations were determined by densitometry analysis on SDS PAGE against a BSA standard curve. Protein was stored at -20°C until further use.

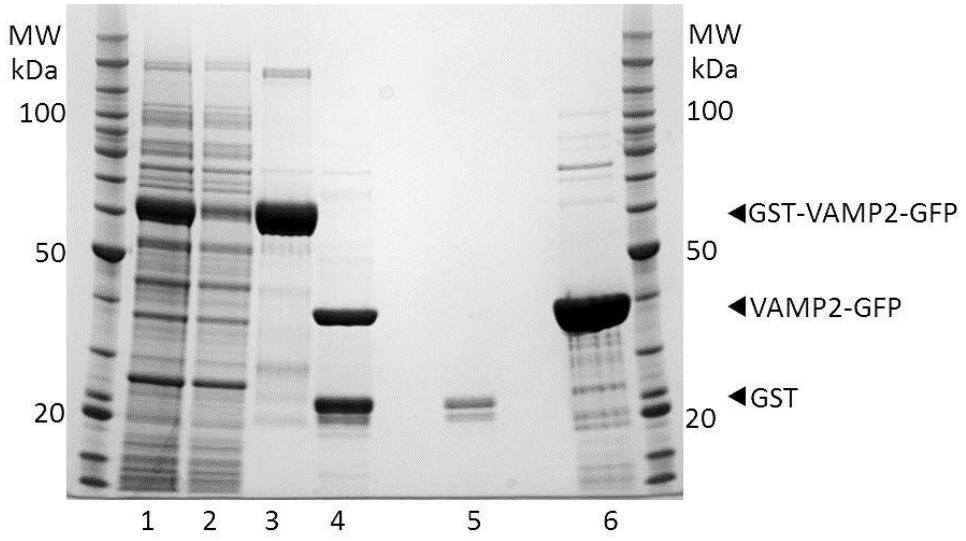

Figure 2.2. Purification of VAMP2-GFP. SDS-PAGE summary of VAMP2-GFP purification. Lane 1, cell lysate supernatant; 2, unbound fraction from first purification step; 3, pooled eluate from first purification step; Tag removal by PreScission treatment; 5, Eluate from final purification step; 6, Final and concentrated VAMP2-GFP fraction.

Cleavage of recombinant SNARE substrates by BoNT derivatives

All the assays were optimised by performing preliminary experiments to determine the most appropriate concentration range for substrate and test proteins. Details of the experiments are given in the following chapters. As a summary, test proteins were diluted in buffer containing 50 mM HEPES pH 7.2, 20 μM zinc chloride, 1 μg/μl BSA, 10 mM DTT, and incubated at 37°C for 30 minutes. Recombinant substrates were diluted and incubated with the test enzyme at 37°C for 1 hour. Reactions were stopped by adding 2x reducing sample buffer (NuPAGE, Invitrogen). Samples were then analysed by SDS PAGE along with BSA standards. Assay results were quantified by densitometry (Syngene).

2.6.4. Embryonic spinal cord neuron (eSCN) assays

Details of the BoNT serotype specific experimental conditions are given in the following chapters. A general summary for the eSCN assays is given below.

Spinal cords dissected from 14–15 day old foetal Sprague Dawley rats were cultured for 21 days using a modification of previously described method (Chaddock *et al.,* 2002) and provided by Syntaxin Ltd. eSCN were treated with serial dilutions of the test enzymes, and incubated at 37 °C with 10% CO_2 for 24 hours. Cells were lysed by removing all media and adding sample buffer (25% NuPAGE buffer (Invitrogen), 10 mM DTT). After 20 minutes, samples were transferred into microcentrifuge tubes and heated at 95°C for 5 minutes. All samples were run on 12 % BIS-TRIS gels (Invitrogen) and proteins were transferred onto nitrocellulose membranes using an iBlot Dry Blotting System (Invitrogen) following the manufacturer's instructions. Substrate cleavage was monitored by measuring the disappearance of the specific substrate immunoreactive bands (and for VAMPs compared to an internal control protein). Specific binding of substrate primary antibodies was visualised using peroxidase-conjugated secondary antibodies and an enhanced chemiluminescent (ECL) detection system (Thermo Scientific), similarly to the Western blotting protocol described previously. Analysis was performed by densitometry (Syngene). Assay results (calculations and curve fittings) were processed with Prism5 (GraphPad Software).

2.7. X-ray crystallography

2.7.1. Crystallisation

Vapour diffusion method

Proteins may form crystals when set in the right environment. In the crystallisation process, the protein in a soluble state is slowly transformed towards precipitation. When conditions for crystallisation are met, the protein molecules arrange themselves in regular orientations to form repeating unit cells. The saturation curve for protein crystals illustrates the conditions required for this mechanism (Figure 2.3).

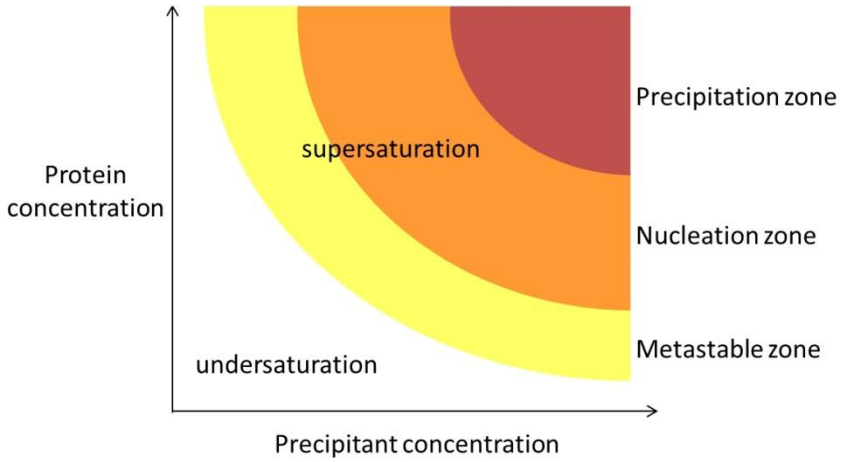

Figure 2.3. Protein crystallisation saturation curve. Crystallisation can happen under supersaturated conditions after nucleation occurs. In the precipitation zone, the level of saturation is too high and makes the protein precipitates out of solution. Crystals may be grown in the metastable zone from previous nucleation.

Vapour diffusion is the most commonly used methods for protein crystallisation and regroups two techniques known as hanging drop and sitting drop. Both rely on the same principle whereby a drop, containing the protein to crystallise along with buffer and precipitant, is set to equilibrate against a reservoir of solution composed of the same buffer and precipitant but at higher concentration. This allows the water to vaporise from the drop and transfer to the reservoir. Consequently the precipitant concentration within the drop increases until equilibrium is reached and permits crystallisation (McPherson, 1985). The difference between the two techniques is in the position of the drop in regard to the well, with the drop being set on a shelf next to the reservoir in a sitting drop arrangement, while the drop is suspended above the reservoir in the hanging drop situation.

Preliminary automated crystallisation screens

A common method was used to determine the crystallisation conditions for all of the purified proteins. Preliminary crystallisation screens were carried out using an automated nano-dispenser (Crystal Phoenix, Art Robbins Instruments). 96-well INTELLI-plates (Art Robbins Instruments) were set up following a sitting-drop vapour diffusion method with 40 µl mother liquor in the reservoir, and 200 nl drop size with 2:1 and 1:1 protein to mother liquor ratios. Available commercial crystallisation screens were used (all from Molecular Dimensions): Structure screen I + II HT-96; Clear Strategy Screen I HT-96; Clear Strategy Screen II HT-96; - Heavy + Light twin pack HT-96; PACT *premier* HT-96; JCSG-*plus* HT-96; Morpheus HT-96; PGA Screen HT-96.

All plates were incubated at 16°C. Typically, crystallisation plates were checked after 2 days, 1 week, 2 weeks and then regularly until drops had dried. When potential crystallisation hits were observed, further optimisation trials were performed, either manually or by using the robot and commercially available additive screens (Additive Screen HT, Silver Bullets HT, from Hampton Research).

Crystallisation manual optimisation trials

When crystals were observed in the preliminary screens, the corresponding conditions were further tested. Trials by the hanging-drop method using 24-well plates (XRL plate, Molecular Dimensions) were set up. Typically, the precipitants were tested at a broad concentration range (5-200 % of initial concentration), and with different buffer systems for pH optimisation. Other parameters tested included the protein to mother liquor ratio (1:1; 2:1; 1:2), drop size (1 to 4 µl) and reservoir volume (200 to 1000 µl). Alternatives to the initial precipitants were also trialled such as, for examples, PEG of different molecular weights, and various salts with a common element (i.e. sodium, potassium salts…).

Several seeding methods were also used to optimise the crystals, particularly for transferring crystals from the preliminary 96-well plate to the 24-well plate format. Firstly streak-seeding was performed for smaller crystals using a thin needle. Secondly, when crystals of a suitable size were obtained, they could be isolated and inoculated in a new drop with lower precipitant concentration, thus favouring crystal development in the metastable zone (Figure 2.3) after nucleation.

Details of the crystallisation conditions obtained for each protein is given in the following chapters.

2.7.2. Data collection and processing

Data collection at the synchrotron source

All the data presented in this study were collected at the Diamond Light Source (UK), macromolecular crystallography stations IO2, IO3, IO4 or I24 (microfocus beamline). Crystals were mounted in cryoloops on site on the day of experiment. When possible, various cryoprotectants were tested depending on the crystallisation conditions. Crystals were centred and tested for x-ray diffraction. When diffraction spots were observed, appropriate settings were applied for data collection (resolution, oscillation range, beam transmission) and images were collected. Depending on the radiation damages observed, and unit cell parameters, 100 to 200 images with a 1° oscillation range would usually be collected from a single crystal when using a regular beamline.

At the time of experiments, beamlines IO2, IO3, and IO4 were equipped with a Quantum-4 CCD detector (ADSC Systems, CA). The I24 microfocus beamline was equipped with a PILATUS-6M (Dectris) detector.

Data processing

The data were processed and scaled using MOSFLM and SCALA of the Collaborative Computational Project No. 4 softwares for macromolecular x-ray crystallography (CCP4, 1994; Leslie, 2006). Indexing in the appropriate space group was supported by analysis with POINTLESS (CCP4, 1994).

Details on data collection and processing are given in the following chapters.

2.7.3. Structure determination and validation

Molecular replacement

Molecular-replacement is one of the main methods for determining the crystal structures of macromolecules. It relies on the existence of known structural models for molecules that share a degree of homology with the macromolecular structure to be solved. The different botulinum neurotoxin serotypes share 30 to 60 % homology (Lacy and Stevens, 1999). With the crystal structures of three full-length toxins available, along with the catalytic domain from each serotype (Swaminathan, 2011), the BoNT fragments and derivatives described in this study presented an ideal case for molecular replacement.

The molecular replacement method is based on the Patterson function properties. Phase information is not necessary to calculate the Patterson function which can be extracted from a potential structural model, and compared to the experimental, measured Patterson. There are six dimensions in the molecular replacement problem which may be simplified in a subset of two three-dimensional searches. The program MolRep (Vagin and Isupov, 2001) is based on the Patterson method. A molecular replacement solution is found when the search model is correctly located in the crystallographic unit cell. Its Patterson function should then be similar to the experimental one. The first three-dimensional search corresponds to a cross-rotation function. The spatial points with a high rotation function values are kept as the most likely orientations of the search model. Then a search with the translation function, the second three-dimensional function, is applied on these orientations.

A more recent technique to molecular replacement has been including the maximum likelihood method to these functions. The program Phaser (McCoy *et al.*, 2007) allows for calculation of the probability that the data would be measured with the given model. Hence

this provides an indication of the consistency between the model and the experimental data, and offers a more sensitive approach to molecular replacement.

Details on molecular replacement are given for each of the structures solved with this method in the following chapters.

Refinement and validation

Refinement of the crystal structures was carried out with REFMAC (Vagin *et al.*, 2004) and manually with Coot (Emsley and Cowtan, 2004). The REFMAC program can perform rigid body, TLS (Translation-Libration-Screw motion), and restrained or unrestrained refinement against x-ray data. It minimises the coordinate parameters to satisfy either a maximum likelihood or least squares residual. Typically a first round of rigid body refinement was performed on the molecular replacement solution, and then several rounds of restrained refinement. Weight of geometrical parameters and the presence of non-crystallographic symmetry were adjusted according to each structure. Between REFMAC refinement rounds, manual adjustments and model fitting were done with Coot using real-space refinement and regularisation options. Coot was also used to add solvent and ligand molecules when appropriate. TLS refinement was usually performed in the final stages of refinement after validation of the structure's geometry.

Quality of the final structures was assessed by the figure of merit, R factors and geometric statistics. All structures were checked with the MolProbity program (Davis *et al.*, 2007) and protein geometry optimised for acceptable Ramachandran statistics, Cβ deviation, bond length and bond angles (CCP4, 1994).

Details of the final crystal structures and their refinement statistics are given in the following chapters. Coordinates and structure factors of new structures were deposited to the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/), using the AutoDep server (EBI).

2.8. Small-angle x-ray scattering

Principles of small-angle x-ray scattering

Small-angle x-ray scattering (SAXS) is a technique for studying the solution structure of biological molecules and macromolecular complexes. SAXS is particularly useful for analysing large conformational changes such as transitions in folding states due to ligand binding or to modifications of the physicochemical environment. These observations are

allowed by the non-intrusive sample preparation where the proteins are left in a simple soluble environment.

X-rays are a form of radiation characterised by their wavelength and amplitude. Their interaction with individual atoms of a macromolecule causes the formation of secondary wavelets which results in a coherent scattering. In SAXS, the protein sample is exposed to x-rays and the scattered intensity "I(s)" is measured as a function of the momentum transfer "s" against the solvent scattering intensity which is subtracted (Jacques and Trewhella, 2010). It is important that the macromolecule sample be pure and monodisperse in solution. Indeed the intensity measured is an average of the isotropic intensity distribution which is in relation with all the random orientations of the molecules in solution. The net particle scattering is in correlation with the contrast which corresponds to the squared difference in electron density between particle and solvent. The signal measured is dependent on the size and shape of the particles being analysed and may thus provide information on the tertiary structure and molecular assembly of proteins. The resolution of the structural information obtained is limited by the smallest angles at which the data are measured. In the real space dimension, this corresponds to the longest distance that can be observed from the data.

Analysis of SAXS data may be performed following the Guinier approximation (Guinier, 1938). This analysis at small scattering angles gives a direct estimation of two important parameters: the radius of gyration, Rg, and the derived intensity at zero scattering angle, I(0). Rg is defined as the mass distribution of the molecule around its centre of gravity. Variation in Rg may be an indication of conformational change in the macromolecule structure. Furthermore, an important characteristic and another tool for determining the shape of a molecule is the distance distribution function, P(r) (Putnam *et al.*, 2007). This function describes the paired-set of distances between all of the electrons within the molecule.

Ab initio modelling

Recent advances in SAXS data analysis have led to the development of *ab initio* algorithms that allow the modelling of a molecular shape from the data alone. The DAMMIF (Franke and Svergun, 2009) program allows representation of the macromolecule by finite volume elements (or "dummy atoms") and fits the experimental data using simulated annealing. The computed model provides a compact shell corresponding to the surface of the molecule. This ab initio approach usually produces several models that need to be averaged to provide for a more accurate representation. The DAMAVER (Volkov and Svergun, 2003) program can be utilised to classify structurally similar models and generate an average structure.

Details of SAXS experimental procedures for sample preparation and data analysis are given in chapter 3.4.

Chapter 3. Structure and activity of the LHn fragments from *Clostridium botulinum* neurotoxins

3.1. Structure and activity of LHn/A

3.1.1. Introduction

Toxins from *Clostridium botulinum* species are the causative agent of the rare neuroparalytic illness botulism. There exist seven distinct serotypes (A–G) of botulinum neurotoxins (BoNTs) that affect humans and other species to varying degrees. Once inside the neuronal cell, BoNTs block the release of neurotransmitters leading to paralysis. Despite their potent toxicity, various preparations of BoNT/A are available commercially as therapeutic and cosmetic agents.

BoNT/A is synthesised as a single polypeptide chain (150 kDa), which is post-translationally cleaved into a di-chain molecule composed of light chain (LC/A) (50 kDa) and heavy chain (HC/A) (100 kDa). LC/A is the catalytic domain and a zinc-endopeptidase, while HC/A is further divided into two sub-domains of equal molecular mass called the translocation domain (Hn/A) and the membrane binding domain (Hc/A). The crystal structure of serotype A was solved and presented a linear arrangement of these three functional domains (Lacy *et al.,* 1998). On binding to the nerve terminals, BoNT/A is endocytosed into a vesicle, where the acidic environment causes some conformational changes and LC/A is translocated into the cytosol (Montal, 2010). Inhibition of neurotransmission takes place by proteolysis of the pre-synaptic membrane protein SNAP-25 (synaptosomal associated protein of 25 kDa; Schiavo *et al.*, 2000). SNAP-25 is one of the soluble N-ethyl-maleimide-sensitive fusion protein attachment receptor (SNARE) proteins which mediate cell secretion (Südhof *et al.*, 2009).

LHn/A is a non-toxic fragment of BoNT/A. It consists of only the LC and Hn of the whole type A toxin, hence it is incapable of binding extracellularly to cell membranes, but retains its full catalytic activity (Chaddock *et al.*, 2002). LHn/A was originally prepared by trypsinisation of whole BoNT/A (Shone *et al*., 1985). However, preparation of recombinant LHn/A has subsequently been achieved by Chaddock *et al.* (2002). LHn/A was proposed as a component in the design of a range of therapeutic molecules for the treatment of pain, control of secretion from multiple cell types and other muscular disorders. By coupling the LHn/A molecule to a ligand that selectively binds to a specific target cell that plays a role in the maintenance of a chronic disease state, it is proposed that the retargeted LC/A activity

will result in inhibition of secretion from said cell, thereby leading to alleviation of symptoms and improvement in quality of life. As proof-of- principle, retargeting of LHn/A in fusion with other cell binding proteins such as *Erythrina cristagalli* lectin (Chaddock *et al.*, 2004), wheat germ agglutinin (Chaddock *et al.*, 2000a), and nerve growth factor (Chaddock *et al.*, 2000b) have been successfully tested in vitro and in vivo models.

Elucidation of the tertiary structure of LHn/A to confirm its native state and to aid in designing suitable ligands as delivery vehicle has become important. Details of the first crystal structure of the LHn non-toxic fragment from BoNT/A at 2.6 Å resolution are presented here. The catalytic activity of LHn/A was confirmed by SNAP-25 cleavage assay, and the stability of the fragment assessed. These results have provided the structural basis for the LHn/A protein and should lead the way forward for the design of a new class of therapeutic molecules.

3.1.2. Materials and methods

LHn/A cloning and expression

The synthetic gene encoding 877 amino acids of LHn/A was cloned into modified pMAL-c2x vector (NEB, UK) with a Factor Xa cleavable N-terminal poly-His(6x) tag and transformed into *E. coli* BL21 expression cells. The LHn/A gene was engineered to encode for Factor Xa cleavage site (IEGR) between the LC and Hn domain between positions 443-446. The clone was provided by Syntaxin Ltd. Expression of LHn/A was carried out by inoculating terrific broth complex medium (100 μg/ml Ampicillin) with an overnight culture grown at 37^{o}C. The cultures were grown in 2 L conical flasks at 200 rpm and 37^{o}C until the OD_{600} reached 0.5. Cultures were then set at 16°C for induction with 1 mM IPTG and left growing overnight (200 rpm). The cells were then harvested and stored at -80^{o}C until further use.

LHn/A purification

Cells were resuspended in 50 mM M HEPES, pH 7.2, 0.2 M NaCl and lysed using a homogeniser (Constant Systems Ltd). Lysate was centrifuged for 45 minutes at 18,000 g. Soluble fraction was loaded onto a Ni^{2+}-charged chelating sepharose column (GE Healthcare). Bound proteins were then eluted in steps at 10, 40, 100 and 250 mM imidazole (dissolved in 50 mM HEPES, pH 7.2, 0.2 M NaCl) (Figure 3.1A). The fractions of interest were pooled together and dialysed overnight at 4 °C against 50 mM HEPES, pH 7.2, 0.2 M NaCl.

Factor Xa (BioLabs) treatment was carried out to remove the N-terminal 6xHis-tag. This step was also necessary to mimic the native protein endoprotease activation by *C. botulinum*, allowing formation of the two-chain (LC and Hn) protein. Digestion was carried out by adding 1 unit of Factor Xa per 0.1 mg fusion protein and incubated at 25°C overnight.

The cleaved fusion protein was loaded onto a Ni^{2+}-charged chelating sepharose column equilibrated with 50 mM HEPES, pH 7.2, 0.2 M NaCl. The protein of interest did not bind to the column as its tag was cut-off. The flow-through was therefore collected and further purified. Ammonium sulphate was added to the fraction to make up a 1 M concentration, and the sample was loaded onto a phenyl sepharose column (GE Healthcare) previously equilibrated with 1 M ammonium sulphate in 50 mM HEPES, pH 7.2. Elution was performed in steps at 0.7, 0.6, 0.4, 0.3 and 0 M salt in 50 mM HEPES, pH 7.2. LHn/A eluted at 0.7 and 0.6 M (Figure 3.1B) and corresponding fractions were pooled together. Pooled sample was dialysed overnight against 50 mM HEPES, pH 7.2, 0.2 M NaCl at 4°C. The sample was finally concentrated using Vivaspin 50000 MWCO concentrator (Millipore), spinning at 2500 g until concentration reached approximately 5 mg/ml. All concentrations were determined by absorbance measurement at 280 nm. Protein was stored at -20°C until further use.

Enzymatic assay

The SNAP25 cleavage assay allows the monitoring of botulinum toxin serotype A catalytic activity by SDS PAGE and is described below.

LHn/A was diluted down to 1.5 mg/ml in buffer containing 50 mM HEPES pH 7.2, 20 μM $ZnCl_2$, 1 μg/μl BSA, 10 mM DTT, and incubated at 37 °C for 30 minutes. SNAP25-GFP recombinant substrate (Syntaxin Ltd, UK) was diluted and used at a final concentration of 4 μM and incubated with the test enzyme at appropriate concentrations, at 37 °C for 4 hours. Reaction was stopped by adding 2x reducing sample buffer (Invitrogen). Sample was loaded on a NuPAGE 4-12 % BIS-TRIS gel (Invitrogen) and visualised by staining with Simply Blue Safestain (Invitrogen).

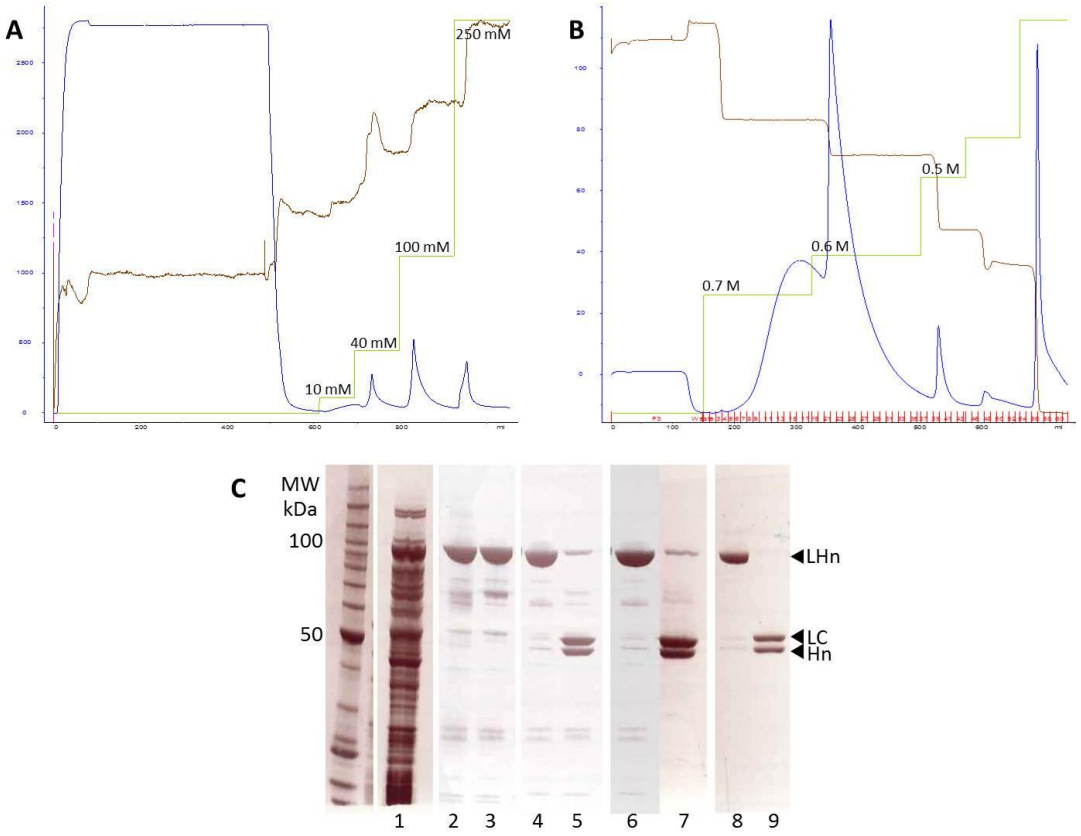

Figure 3.1. Purification of LHn/A. (A) Chromatograph from first purification step using a Ni-charged chelating sepharose column (imidazole concentration is indicated). **(B)** Elution profile from final purification step using a phenyl sepharose column (ammonium sulphate concentration is indicated). Absorbance at 280 nm, blue; gradient concentration, green; conductivity, brown. **(C)** SDS-PAGE summary of purification. Lane 1, cell lysate supernatant; 2, 3, pooled eluate from first purification step in oxidised (O) and reduced (R) conditions (respectively); 4, 5, Factor Xa-activated O and R sample; 6, 7, Eluate from second affinity chromatography (O and R); 8,9, Eluate from final purification step (O and R).

Stability assay

Purified LHn/A sample was dialysed in either 50 mM HEPES pH 7.2, 200 mM NaCl or 100 mM TRIS Acetate pH 8.5, 1.5 M Ammonium Sulphate, 15 % glycerol. Samples were then diluted down to 1.5 mg/ml in corresponding buffer, aliquoted and stored at -20˚C, +4˚C and +25˚C until day of testing. All samples were loaded in duplicate (2 μg) on non-reduced and reduced gels (4-12 % BIS-TRIS, Invitrogen). The percentage purity of LHn/A was determined by densitometry (GeneTools, Synoptics Ltd).

Crystallisation

Final purified LHn/A (residues 1-877) was stored as a frozen stock (-20°C) at 4.8 mg/ml in solution with 50 mM HEPES (pH 7.2) and 0.2 M sodium chloride. Initial crystallisation condition was identified from automated preliminary trials (Phenix, Art Robbins instruments) with Structure Screen I and II (Molecular Dimensions). The condition was repeated by the hanging-drop vapour diffusion method, with 3 µl drops (2 µl protein and 1 µl mother liquor) against a reservoir (500 µl) at 16°C. Microcrystals were observed with 12% glycerol, 0.1 M TRIS pH 8.5, 1.5 M ammonium sulphate. Optimisation gave small reproducible crystals after 14 days. Macro-seeding was necessary to obtain crystals of suitable size and quality, presenting a single rectangular prism, in 15% glycerol, 0.1 M TRIS Acetate pH 8.5, 1.5 M ammonium sulphate. LHn/A crystals were also obtained by using 15 % sucrose for glycerol and diffracted at 2.6 Å resolution (Figure 3.2).

Data collection and structure solution

X-ray diffraction data for LHn/A were collected at Diamond Light Source (DLS, UK), beamline I03. Prior to data collection, LHn/A crystals were transferred to a cryoprotectant solution with 25% (v/v) glycerol for 1 min, and then flash-frozen under a nitrogen stream. The crystals grown with glycerol diffracted to 2.75 Å in space group $P2_12_12_1$ (a= 78.4; b= 157.0; and c= 211.7 Å) with two molecules per asymmetric unit. Crystals grown in sucrose required annealing for 30 seconds to show diffraction up to 2.6 Å in a similar cell. The data were processed in primitive orthorhombic space group $P2_12_12_1$ using MOSFLM version 7.0.3, and scaled up to 2.6 Å using SCALA from CCP4 suite (CCP4, 1994; Leslie, 2006). Initial phases were obtained by the molecular replacement (MR) method using MolRep (Vagin and Isupov, 2001) with the coordinates of a homology based model of the LHn/A fragment from the holotoxin model (PDB 3BTA, Lacy *et al.,* 1998). Crystallographic refinement were carried out using CNS suite version 1.2 *(*Brünger *et al., 1998)* and model fitting were done using Coot version 0.4.1 *(*Emsley *et al., 2004)*. PROCHECK *(*Laskowski *et al., 1993),* and MolProbity *(*Davis *et al., 2007)* were used to check the geometry of the structure. Data collection and processing statistics are provided in Table 3.1.

The atomic coordinates and structure factors (codes 2W2D and r2W2Dsf, Masuyer *et al.,* 2009) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

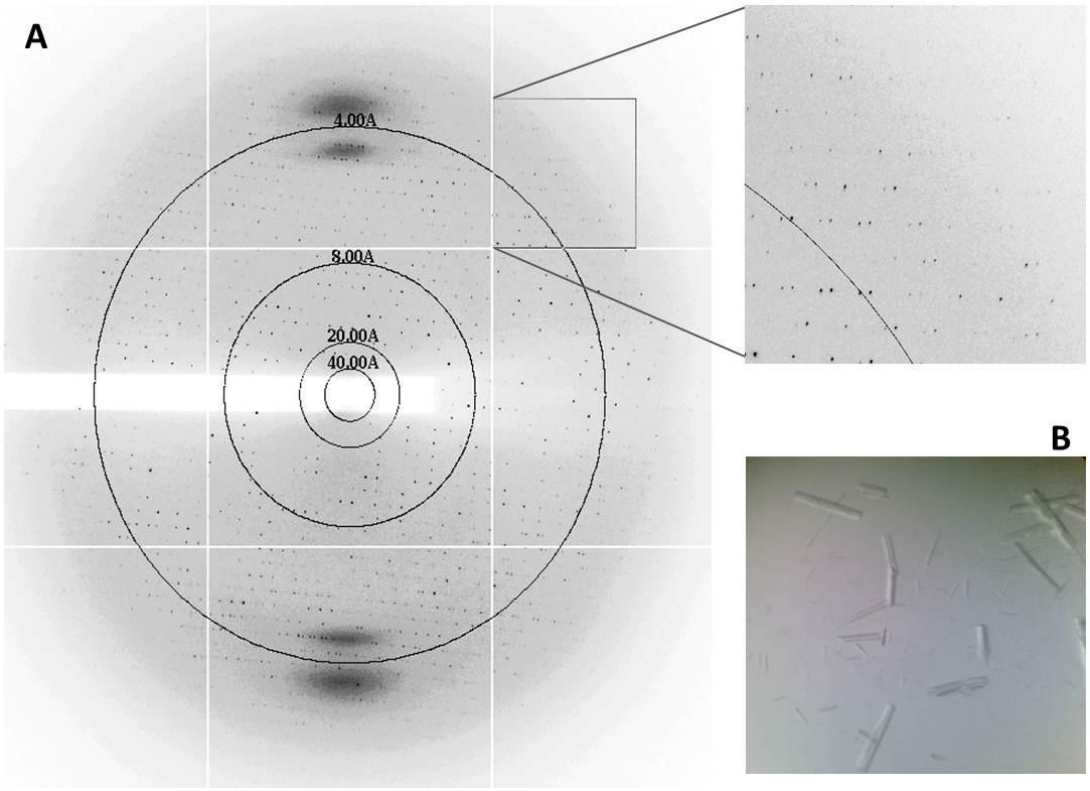

Figure 3.2. Crystals and x-ray diffraction of LHn/A. (A) Diffraction image collected at DLS IO3 where LHn/A crystals diffracted at 2.6 Å. **(B)** Crystals of LHn/A after optimisation in 15% sucrose, 0.1 M TRIS Acetate pH 8.5, 1.5 M ammonium sulphate.

Table 3.1. Data collection and refinement statistics of LHn/A

A. Data Collection Statistics	
Space Group	$P2_12_12_1$
Cell	a = 78.4, b =157.0, c =211.7 Å ; α, β, γ = 90°
Resolution range (Å)	50-2.6
R_{sym}[1] (%)	10.9 (55.3)
I/σI (outer shell[2])	11.5 (2.0)
Completeness %	97.1 (98.4)
Redundancy	1.9 (1.8)
B. Refinement Statistics	
Resolution range (Å)	50-2.6
No. of reflections	528,420
Unique no. of reflections	82,260
R_{cryst}[3] (%)	21.2
R_{free}[4] (%)	25.3
Number of non-H atoms	
Protein	13,740
Ligand (ions)	2 Zinc, 3 Chloride, 1 Acetate, 4 Sulphate, 5 Glycerol
Water	309
Average temperature factor (Å^2)	41.0
RMSD for bond lengths (Å)	0.009
RMSD for bond angles (°)	1.3

[1] $R_{sym} = \Sigma_h \Sigma_i |I(h) - I_i(h)| / \Sigma_h \Sigma_i I(h)$, where $I_i(h)$ and $I(h)$ are the *i*th and the mean measurements of the intensity of reflection *h*, respectively.
[2] Outer shell is 2.74Å – 2.60Å.
[3] $R_{cryst} = \Sigma_h |F_o - F_c| / \Sigma_h F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection *h*, respectively.
[4] R_{free} is equal to R_{cryst} for a randomly selected 814 (1.0 %) reflections not used in the refinement.

3.1.3. Results and discussion

3.1.3.1. Structure

First crystal structure of LHn/A

The structure of LHn/A using X-ray diffraction studies has been determined at 2.6 Å resolution. The structure was refined to a final R_{free} = 25.3%, and R_{cryst} = 21.2% (Table 3.1, Figure 3.3), with 98.6% of amino acids in allowed region and 1.4% in the Ramachandran plot generously allowed region. The two molecules in the crystallographic asymmetric unit are tightly packed against each other in a dimeric arrangement. The interactions are stabilised by 15 potential H-bonds and 152 van der Waals contacts. However, the interactions between the two molecules could be due to crystal packing and may not be an inherent property of LHn/A. The buried surface area between the two molecules is 377 $Å^2$ compared to the total surface area of 21802 $Å^2$. Engineered Factor Xa site for LC activation was confirmed both by SDS-PAGE analysis and in the crystal structure where there is a clear evidence of break in electron density between LC and Hn (Figure 3.3C). Furthermore, ordered electron density allowed model building of the full-length LC C-terminus and the disulphide bridge (Cys 430–Cys 454) between LC and Hn which are stabilised through an anti-parallel β-sheet arrangement.

The overall structure of LHn/A (Figure 3.3) resembles BoNT/A without the binding domain, and superposition of LHn/A on BoNT/A gives an overall root mean square deviation of 0.85 Å for 830 Cα-atoms (Figure 3.3B). This proves that the absence of binding domain in LHn/A did not produce any major conformational change compared with the structure of full length BoNT/A. LC is characterised by a mixture of α-helix and β-strands and has the conserved zinc binding motif HEXXH. Hn consists of three long anti-parallel α-helices of length 110 Å. A small region of Hn, called the 'belt', wraps LC and prevents access to the catalytic site. At least two roles are proposed for the belt, one as a chaperone for the catalytic domain (Brünger *et al.*, 2007) and the other as a regulatory element in membrane interaction (Galloux *et al.*, 2008).

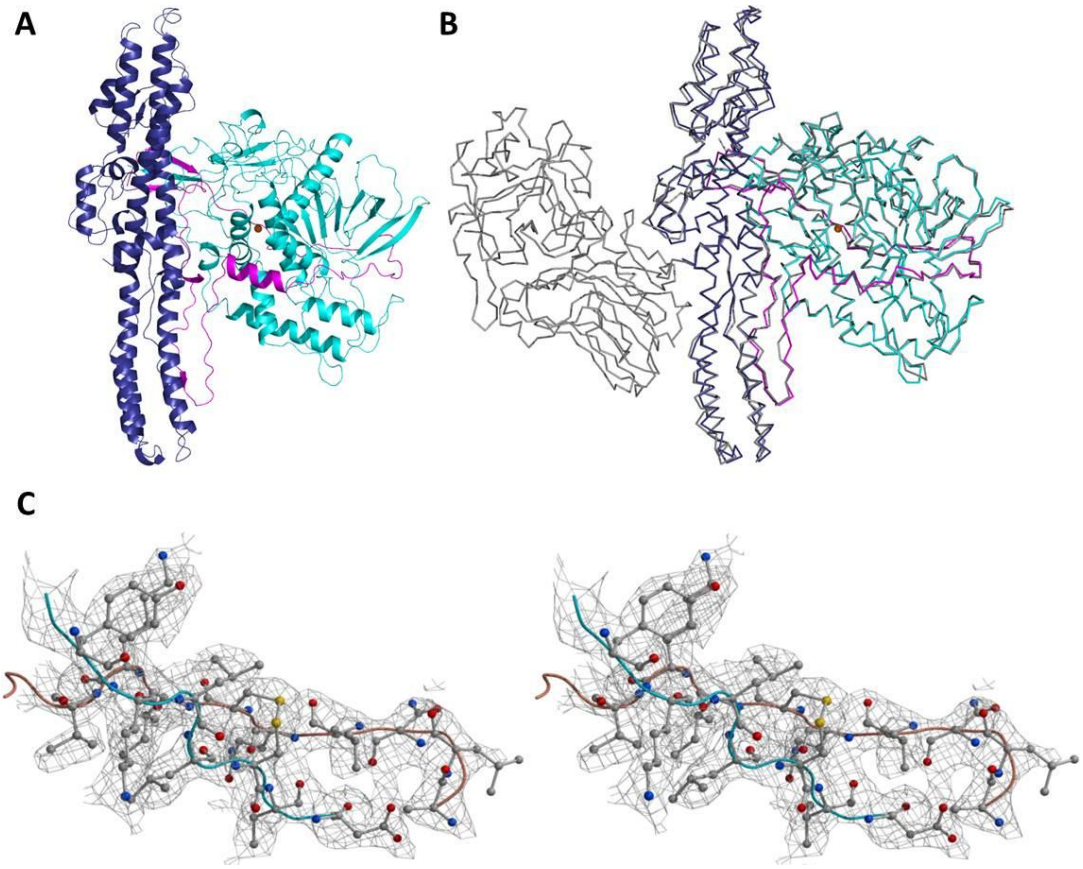

Figure 3.3. Crystal structure of LHn/A. (A) Ribbon diagram representation of LHn/A structure. LC in cyan; Hn in blue; "belt" in pink. Zinc ion shown as orange spheres. **(B)** Superposition of LHn/A on BoNT/A (grey) (PDB code 3BTA). **(C)** Electron density showing the engineered peptide break between LC and Hn. 2|Fo|-|Fc| map contoured at 1σ. All the residues are shown in grey and ball-and-stick model. Disulphide link between LC and Hn is shown and sulphur atoms are coloured yellow.

Domain components of engineered LHn/A

Requirement of optimal length for a functional LHn/A is critical in LHn/A engineering. Deletion analysis carried out by Kurazono *et al.* (1992), showed that removal of 10 amino acids from the amino terminal and 57 amino acids from the carboxy terminal of LC abolished the toxicity even in the presence of Hn. Hence, a full length catalytic domain is a requisite. A construct up to Ser 877 was selected to be an optimal length for Hn domain ending at the linker between Hn and Hc.

The light chain (LC)

The light chains from all serotypes exhibit high degree of structural similarity. The catalytic zinc site is conserved in all serotypes and zinc is essential for LC's enzymatic activity. In the engineered LHn/A molecule, the zinc ion is tetrahedrally coordinated to His

223, His 227, Glu 262 and a water molecule involving Glu 224, as observed in LC structures of BoNT/A (Figure 3.4).

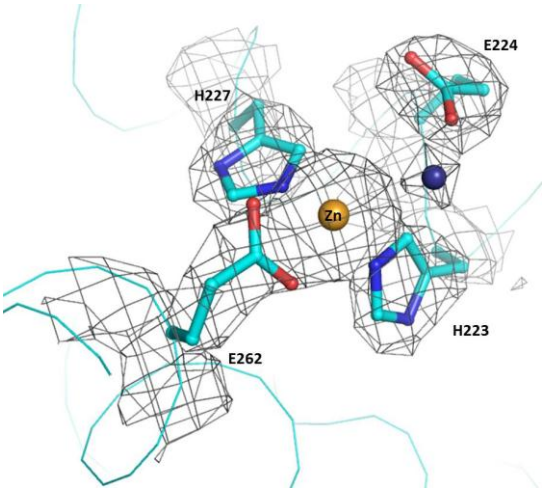

Figure 3.4. Zinc coordination in LHn/A. Ribbon diagram representation of LC/D (cyan). Zinc ion is shown as an orange sphere, water molecule in dark blue. Residues involved in the zinc coordination are shown in stick and ball representation. 2|Fo|-|Fc| map contoured at 1σ.

The translocation domain (Hn)

The non-toxic Hn domain is required after endocytosis of the toxin for translocation of LC within the cytosol (Koriazova and Montal, 2003) where it is released and cleaves its substrate, thereby blocking neurotransmitter release. Hn domain's accessible surface area is 6038 $Å^2$ compared to its total area of 19992 $Å^2$ excluding the belt. The absence of the binding domain did not cause any major conformational change in the translocation domain. Hn normally interacts with Hc through eight potential hydrogen bonds and some 160 van der Waals' contacts (calculated using holotoxin structure, PDB code 3BTA) around the 650 loop and residues Pro 801, Lys 805, Glu 808 and Ser 845. Superposition with the holotoxin model shows a change in the position of loop 642–650 and in particular with Tyr 648 which would form a hydrogen bond with Hc at His 886. Hn stability could be explained by strong inter-helical interactions not influenced by the new solvent accessibility.

3.1.3.1. Endopeptidase activity

SDS-PAGE of the protein sample was carried out at various temperatures to assess its stability during storage and under crystallisation conditions. Analysis by densitometry revealed that LHn/A was more than 95% pure after the final purification step (Figure 3.5A). Reduced conditions put in evidence some residual non-activated (single chain) peptide at 100 kDa, representing less than 4% of the sample. LHn/A's stability is also not affected by

high concentration of ammonium sulphate (1.5 M) from the crystallisation condition. There was approximately 10% degradation observed after 36 days of incubation at 25 °C.

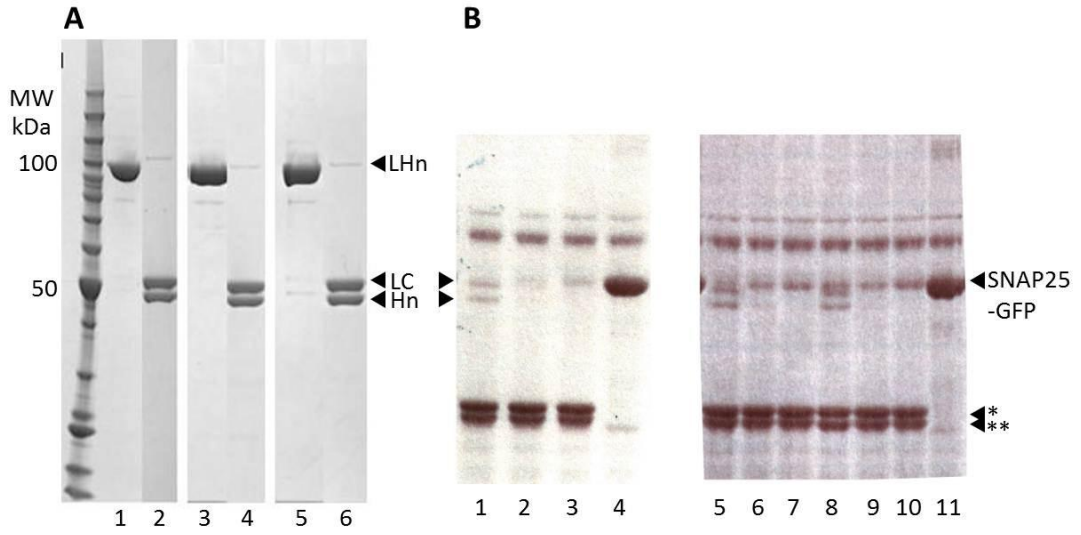

Figure 3.5. Stability and activity of LHn/A. (A) SDS-PAGE analysis of LHn/A stability. Lanes 1, 3 and 5, oxidised samples; 2, 4 and 6, reduced samples. Lanes 1, 2 on day d1; 3, 4 on day d36 in buffer A; 5, 6 on day d36 in crystallisation buffer. **(B)** SDS-PAGE analysis of SNAP-25 cleavage assay by LHn/A. GFP-SNAP-25 is 51 kDa. LC cleaves GFP-SNAP-25 in two peptides of 28 (*) and 23 (**) kDa, respectively. Lanes 1, 5, 8, LHn/A (100 μg/ml); 2, 6, 9, LHn/A (20 μg/ml); 3, 7, 10, LHn/A (4.0 μg/ml). Lanes 4, 11, negative control. BSA visible at 70 kDa in all lanes. On lanes 1, 5, 8 the two bands below the non-cleaved substrate correspond to reduced LHn/A domains.

The catalytic property of LHn/A was confirmed by the SNAP-25 cleavage assay. (Figure 3.5B). A SNAP-25-GFP substrate was engineered to monitor LHn/A proteolytic action resulting in two peptides of approximately 28 and 23 kDa respectively. The results were in correlation with the stability study since less than a 10% loss of activity was noticeable after 36 days of storage under crystallisation conditions at 25˚C. LHn/A tested at 0.1 mg/ml showed complete substrate cleavage and uncleaved substrate was only just detected by densitometry with LHn/A at 0.02 mg/ml. Overall, LHn/A presented strong stability and kept its catalytic activity over a period of 36 days in the crystallisation storage conditions. This corresponds to the length required for crystal formation, implying that the structure described here correspond to a functional toxin with its native property.

The data presented here have led to the solution of the first crystal structure for an LHn fragment of botulinum neurotoxin and does not show any noticeable structural difference between the parent neurotoxin and the recombinant fragment lacking the Hc domain. LHn/A was shown to be catalytically active, and a recent study demonstrated its ability to translocate the active protease within the cytosol of target cells (Fischer *et al.*, 2008b). LHn/A thus represents a fully functional BoNT fragment. With a growing interest in the use of the LHn fragment for therapeutic and vaccine purposes (Acharya and Chaddock, 2011), the structure presented here will facilitate the development of optimised proteins that will have greater applicability. LHn/A's structural integrity is unlikely to alter the property of any fusion proteins and provides additional degree of freedom for the design of future novel ligands.

3.2. Structure and activity of LHn/B

3.2.1. Introduction

BoNT/B is synthesised as a single polypeptide cleaved by clostridial or host proteases to its active form. The C-terminal heavy chain composes the binding (Hc/B) and translocation (Hn/B) domains of 50 kDa each, and is linked by a single disulphide bridge to the catalytic light chain (LC/B), a zinc endopeptidase. The crystal structure of BoNT/B was reported in 2001 (Swaminathan and Eswaramoorthy, 2000) and confirmed the individual fold of each domain, as well as their three-dimensional linear arrangement.

BoNT/B binds specifically to the nerve terminals by a dual interaction with a protein receptor, synaptotagmin, and a ganglioside (GT1b) (Baldwin and Barbieri, 2007). The complex is endocytosed into a vesicle where the acidic environment provokes conformational changes and mediates translocation of LC/B into the cytosol (Montal, 2010). The protease domain is then free to cleave the vesicle-associated membrane protein (VAMP, also known as synaptobrevin and cellubrevin.), one of the SNARE proteins. Impairing the function of the SNARE complex causes inhibition of neurotransmission and hence paralysis.

BoNT/B substrate recognition relies on an extended set of exosites located downstream and upstream of the scissile bond. A number of mutation and kinetic studies with VAMP-2 identified some of the regions important for proteolysis by VAMP-specific Clostridial toxins (Chen *et al.,* 2008; Sikorra *et al.,* 2008). This suggests a common mode of an extended substrate binding, each presenting with a different set of enzyme-substrate interactions.

The LHn fragment of BoNT/A was shown to conserve the proteolytic activity of its parent neurotoxin and was active on cells when added at high concentration (Chaddock *et al.,* 2002; Fischer *et al.,* 2008b). LHn are therefore attractive candidates as scaffolds for the design of novel biological therapeutics (Chaddock and Marks, 2006; Foster, 2009). BoNT/A and /B are the only serotypes currently approved for pharmaceutical applications. Characterisation of the LHn/B fragment is therefore necessary if it is to be used in novel therapeutic molecules.

The crystal structure of the LHn fragment from serotype B botulinum neurotoxin described here was determined at 2.8 Å. It demonstrated the stability of the translocation domain in association with the catalytic light chain. The metalloprotease activity of the parent toxin was conserved and presented differences in VAMP substrates specificity. If applied at suitably high concentrations, this active protease could be translocated inside

neuronal cells despite the lack of a specific binding domain. It confirmed the relevance of LHn/B as a functional molecule and highlights the use of LHn as a tool for deciphering the mechanism of botulinum neurotoxin's action.

3.2.2. Materials and methods

LHn/B cloning, expression and purification

The synthetic gene encoding 880 amino acids of LHn/B was cloned into modified pET vector (Novagen, UK) with a C-terminal 6 x His-tag and transformed into *E. coli* BL21 expression cells. The LHn/B gene was engineered to encode for Factor Xa cleavage site (IEGR) between the LC and Hn domain between positions 449-452. The clone was provided by Syntaxin Ltd. Expression cultures of LHn/B were grown in 1L terrific broth at 200 rpm, 37°C until OD_{600} read 0.5-0.6. Then cultures were incubated at 16°C and induced with IPTG (1mM). Cells were harvested after 18 h, and stored at -80°C until further use.

Cells were resuspended in 50 mM HEPES, pH 7.2, 0.2 M NaCl (buffer A) and lysed at 20,000 psi with a homogeniser (Constant Systems Ltd). Lysate was centrifuged for 45 minutes at 18,000 g. Soluble fraction was loaded onto a Ni^{2+}-charged chelating sepharose column (GE Healthcare). LHn/B eluted at 100 mM imidazole (dissolved in buffer A) and dialysed overnight at 4°C against buffer A. Activation of purified LHn/B is achieved by Factor Xa (New England BioLabs) treatment. The cleaved fusion protein was supplemented with ammonium sulphate to 1 M and loaded onto a Toyopearl Phenyl-650M column, equilibrated with 50 mM HEPES, pH 7.2, 1.0 M ammonium sulphate. LHn/B eluted at 0.7 M ammonium sulphate (in buffer) and dialysed overnight against buffer A at 4°C. A SDS-PAGE summary of LHn/B purification is presented in figure 3.6A. The sample was finally concentrated using Vivaspin 50000 MWCO concentrator (Millipore) to 9 mg/ml and stored at -20°C. All concentrations were determined by A_{280} measurement.

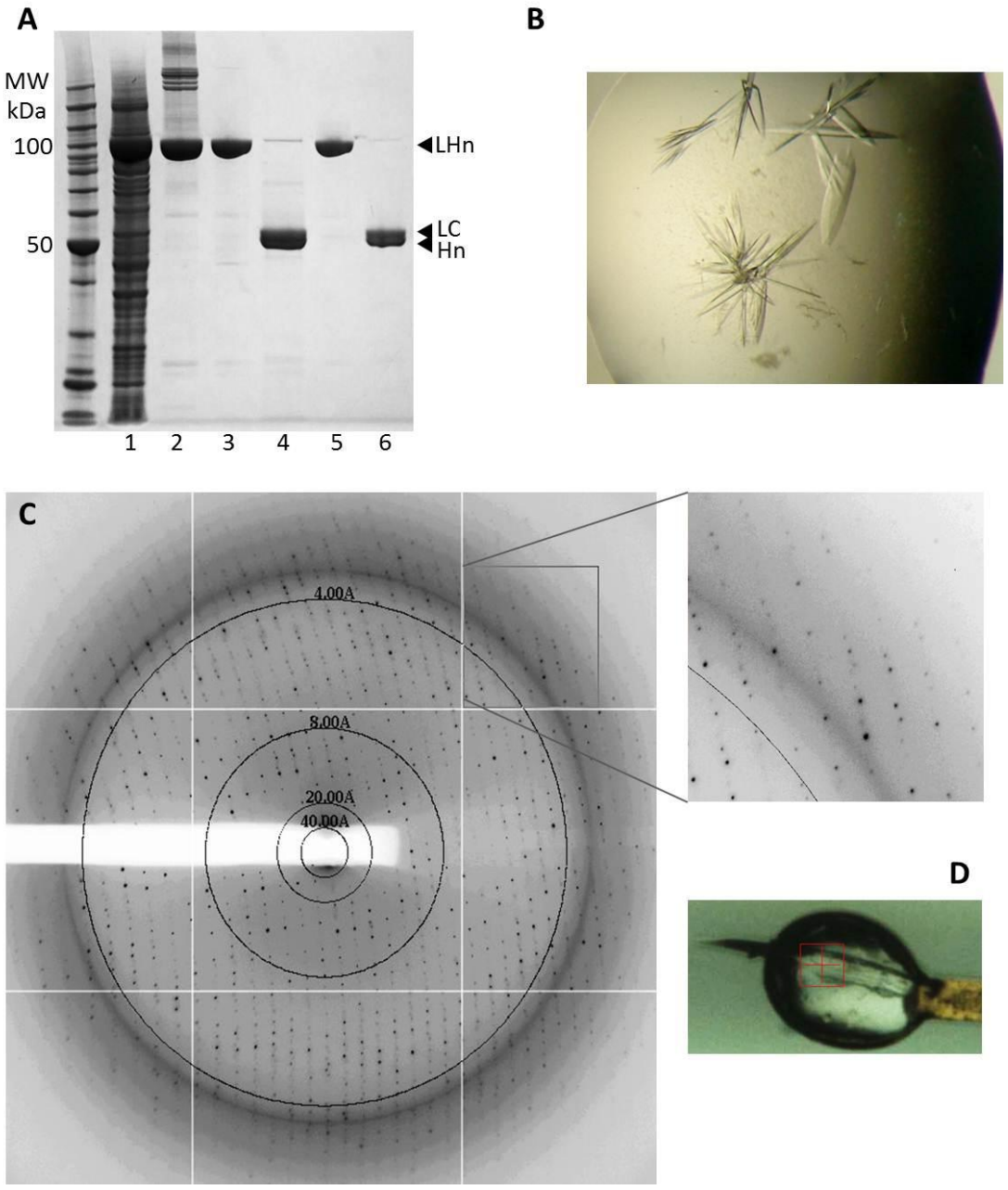

Figure 3.6. Purification and crystallisation of LHn/B. (A) SDS-PAGE summary of LHn/B purification. Lane 1, cell lysate supernatant; 2, pooled eluate from first purification step; 3, 4, Factor Xa-activated sample in oxidised (O) and reduced (R) conditions (respectively); 5, 6, Eluate from final purification step (O and R). **(B)** Crystals of LHn/B after optimisation in 15% PEG3350, 0.1 M BIS-TRIS-propane pH 6.5, 0.2 M sodium sulphate. **(C)** X-ray diffraction image collected at DLS, IO3 where LHn/B crystals diffracted at 2.8 Å. **(D)** LHn/B crystal mounted for data collection.

Crystallisation and structure determination

Several crystallisation conditions were identified using PACT premier (Molecular Dimensions). The conditions which were repeatable by hanging-drop with 3 µl drops (1.5 µl protein and 1.5 µl mother liquor) against a reservoir (500 µl) at 16°C were used in several round of optimisation. Crystals could be readily grown in 15% PEG3350, 0.1 M BIS-TRIS-propane pH 6.5, 0.2 M sodium sulphate. Crystals had a tendency to grow as stacks of blade-shaped forms. Macro-seeding was performed to obtain single crystals of suitable size and quality (Figure 3.6).

X-ray diffraction data were collected at the Diamond Light Source, UK, beamline IO3. A complete dataset to 2.8 Å was collected from a single crystal at 100K (25% glycerol as cryoprotectant) using a Quantum-4 CCD detector (ADSC Systems, CA) (Figure 3.6). The data were processed and scaled in orthorhombic space group $P2_12_12$ using MOSFLM and SCALA (CCP4, 1994; Leslie, 2006) (Table 3.2). Initial phases were obtained by molecular replacement using Phaser (McCoy *et al.*, 2007) with the coordinates of a model based on the LHn fragment of the BoNT/B structure (PDB code 1EPW, Swaminathan and Eswaramoorthy, 2000). Crystallographic refinement was carried out using REFMAC5 (version 5.5; Vagin *et al.*, 2004). Manual adjustments and model fitting were done using Coot (version 0.6.1) (Emsley and Cowtan, 2004). Water molecules were added at positions where F_o–F_c electron density peaks exceeded 3σ and potential H-bonds could be made. The structure was validated using MolProbity (Davis *et al.*, 2007). Structure figures were drawn with PyMOL (DeLano Scientific LLC).

The atomic coordinates and structure factors (codes 2XHL and r2XHLsf, Masuyer *et al.*, 2011) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

Table 3.2. Data collection and refinement statistics of LHn/B

A. Data collection statistics	
Space group	$P2_12_12$
Number of molecules/asymmetric unit	1
Cell dimensions	a = 66.9, b =113.5, c =149.1Å ; α, β, γ = 90°
Resolution range (Å)	50-2.8
R_{sym}[1] (%)	12.1 (51.4)
I/σI (outer shell)	10.3 (3.5)
Completeness (outer shell) %	99.8 (100.0)
Total no. of reflections	165,586
Unique no. of reflections	28,644
Redundancy	5.8 (5.8)
Wilson B-factor (Å^2)	57.6
B. Refinement statistics	
Resolution range (Å)	50-2.8
R_{cryst}[2] (%)	24.0
R_{free}[3] (%)	28.2
Number of non-H atoms	
Protein	6,828
Ligand	1 zinc ion
Water	31
Average temperature factor (Å^2)	Protein atoms- LC/B = 32.6, Hn/ B = 40.8 (water molecules = 22.1)
RMSD in bond lengths (Å)	0.006
RMSD in bond angles (°)	0.86

[1] $R_{sym} = \Sigma_h \Sigma_i |I(h) - I_i(h)| / \Sigma_h \Sigma_i I_i(h)$, where $I_i(h)$ and $I(h)$ are the i^{th} and the mean measurements of the intensity of reflection *h*, respectively.
[2] $R_{cryst} = \Sigma_h ||F_o| - |F_c|| / \Sigma_h F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection *h*, respectively.
[3] R_{free} is equal to R_{cryst} for a randomly selected 5.0% subset of reflections not used in the refinement.

Enzymatic assay

LHn/B and BoNT/B (Metabiologics, US) were diluted to 0.1 μg/ml in buffer containing 50 mM HEPES pH 7.2, 20 μM $ZnCl_2$, 1 μg/μl BSA, 10 mM DTT, and incubated at 37°C for 30 minutes. Recombinant VAMP-1 (2-96), -2 (1-94) and -3 (2-77) substrates were expressed with a C-terminal GFP tag and purified. VAMP-GFP substrates were serial diluted and incubated with the test enzyme at 37°C for 1 hour. Reactions were stopped by adding 2x reducing sample buffer (Invitrogen). Samples were then loaded on to a 4-12 % BIS-TRIS gel (Invitrogen) along with BSA standards and visualised by staining with Simply Blue Safestain (Invitrogen). Assay results were quantified by densitometry (Syngene Bioimaging).

Embryonic spinal cord neuron (eSCN) assay

Spinal cords dissected from 14–15 day old foetal Sprague Dawley rats were cultured for 21 days using a modification of previously described method (Chaddock *et al.*, 2002) and provided for the assay. eSCN were treated with serial dilutions of LHn/B and BoNT/B, and incubated at 37 °C with 10% CO_2 for 24 hours. Cells were lysed by removing all media and adding sample buffer (25% NuPAGE buffer, 10 mM DTT). After 20 minutes, samples were transferred into microcentrifuge tubes and heated at 95°C for 5 minutes. All samples were run on 12 % BIS-TRIS gels (Invitrogen) and proteins were transferred onto nitrocellulose membranes (Invitrogen) using standard protocols. VAMP cleavage was monitored by measuring the disappearance of the specific VAMP immunoreactive bands compared to an internal control protein (GAPDH). Specific binding of VAMP-1, -2 and -3 primary antibodies (Abcam #ab3346, #ab3347 and #ab43080) was visualised using peroxidase-conjugated secondary antibodies and an enhanced chemiluminescent (ECL) detection system (Thermo Scientific), analysis was performed by densitometry (Syngene Bioimaging).

3.2.3. Results and discussion

3.2.3.1. Structure

The crystal structure of LHn/B has been determined at 2.8 Å resolution. The structure was refined to a final R_{free} = 28.2%, and R_{cryst} = 24.1% (Table 3.2, Figure.3.7), with 97.8% of amino acids in the Ramachandran plot favoured region.

The structure presents the two domains interacting tightly in a di-chain complex. The successful activation of LHn/B by Factor Xa at the engineered site was confirmed by SDS–

PAGE analysis (Figure 3.6A) and the crystal structure, which shows a clear evidence of the cleavage by a break in electron density between LC and Hn (Figure 3.7C). This region is likely to be particularly important for BoNT activity and Hn-mediated delivery of LC in the cytosol. The resulting LC C-terminus and the disulphide bridge (A Cys 437– B Cys 461) between LC and Hn are stabilised through an anti-parallel β-sheet arrangement.

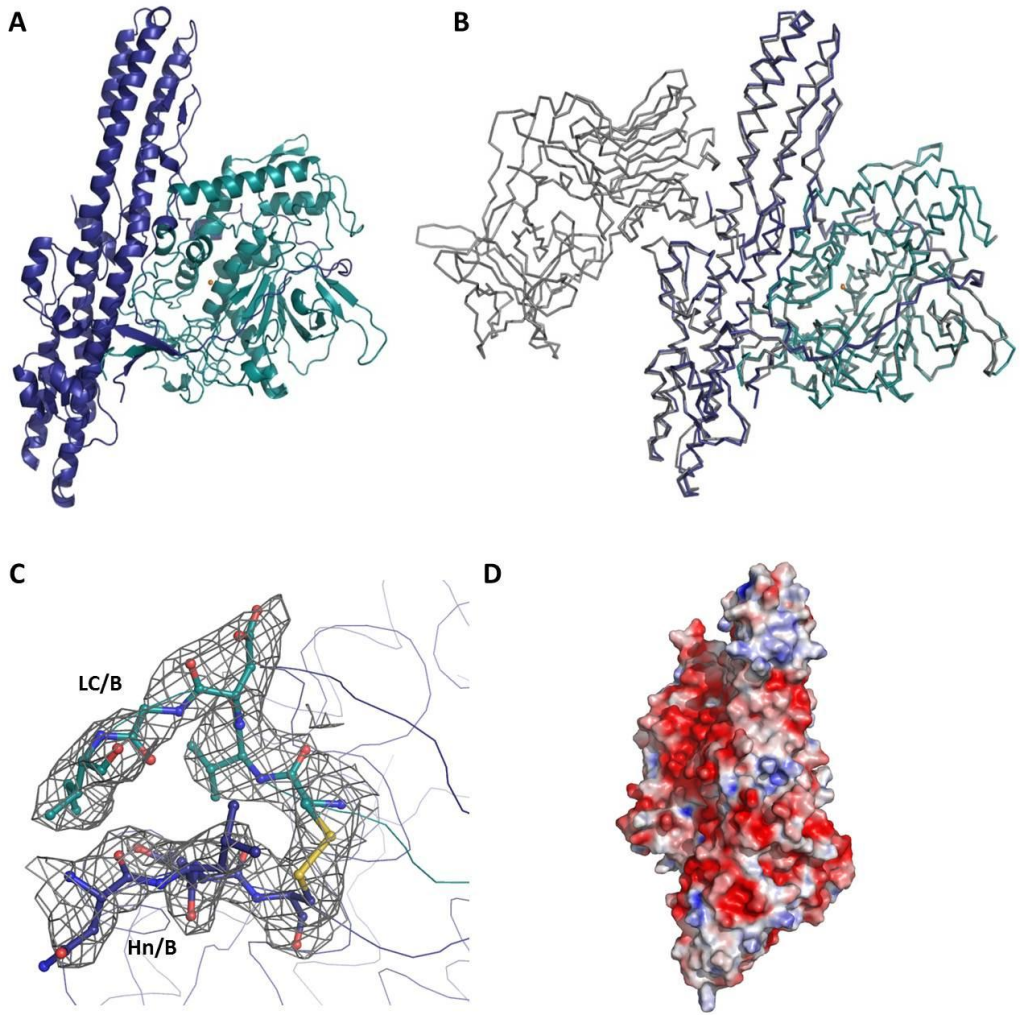

Figure 3.7. Crystal structure of LHn/B. (A) Ribbon diagram representation of LHn/B structure, Hn in blue, LC in cyan. Zinc ion is shown as orange spheres. **(B)** Superposition of LHn/B with BoNT/B (in grey, PDB 1EPW, Swaminathan and Eswaramoorthy, 2000). Overall root mean square deviation was calculated with Swiss PDB viewer and was 0.95 Å for 837 Cα-atoms. **(C)** LC-Hn interaction at the disulphide bridge (sulphur atoms in yellow) – Factor Xa cleavage site for activation. 2|Fo|-|Fc| map contoured at 1σ. **(D)** LHn/B molecule rotated 90° to highlight the C-terminal accessible solvent surface. Electrostatic potential calculated using APBS in PyMOL (Negative potential in red, Positive in blue).

The overall structure of LHn/B (Figure 3.7) resembles BoNT/B without the binding domain, and superposition of LHn/B with its parent toxin gives an overall root mean square

deviation of 0.95 Å for 837 C_α-atoms (Figure 3.7B). The absence of the 50 kDa binding domain in LHn/B did not result in any major conformational change compared with the structure of full length BoNT/B.

LC is a zinc protease with the catalytic ion coordinated by the conserved tetrahedral arrangement of His 229, His 233, Glu 267 and a water molecule bonded to Glu 230. While LC primary sequence is fairly conserved among BoNTs (Lacy and Stevens, 1999), its activity is regulated by a complex substrate binding mechanism relying on dispersed exosites, away from the catalytic site. In the structure of the LC/A-SNAP-25 complex, these sites consisted of dispersed flexible loop regions (Breidenbach and Brünger, 2004). The unstructured regions seen in LHn/B follow the same arrangements as seen in the full length toxin (Figure 3.7). Residues 208-218, downstream of the active site, could not be modelled due to the lack of electron density. This may be the result of the loop's flexibility, even though the holotoxin structure shows it to be stabilised by interactions with helices α17 and α19 of Hn (Swaminathan and Eswaramoorthy, 2000).

Two long anti-parallel α–helices structure the Hn domain. The newly accessible solvent area left open by the lack of binding domain represents 1880 $Å^2$ and shows weak electrostatic potential (Figure 3.7D). The interface between Hn and Hc of the holotoxin is based on 12 potential hydrogen bonds and some weaker van der Waals' contacts between loop 916-920 of Hc and helix α20 of Hn, as well as between the short linker helix of Hc subdomains and loop 610-615 (Hn) (Swaminathan and Eswaramoorthy, 2000). The loop between residues 626-630 and the last 10 C-terminal residues (including the poly-His tag) of LHn/B could not be modelled due to lack of electron density. Strong inter-helical interactions allow Hn to keep its long helical conformation as illustrated by the perfect superposition of LHn/B with BoNT/B (Figure 3.7B). The belt of Hn surrounding LC appears to be important for aspects of BoNT activity. The LHn/B structure demonstrates that the stability of this fold is based on strong interactions with LC and is similar to that seen in the holotoxin structure.

The crystallographic studies of LHn, serotypes A and B, highlight the exceptional structural stability of these fragments when compared to their parent holotoxins. LHn/A has been shown to conserve not only the structure, but also the catalytic property of BoNT/A, as well as a potential for intracellular activity (Chaddock *et al.,* 2002; Fischer *et al.,* 2008b; Masuyer *et al.,* 2009). Analysing the functionality of these fragments can therefore help understand the relationship between BoNT domains.

3.2.3.2. Endopeptidase activity

Proteolytic activity of LHn/B

The ability of LHn/B to retain its proteolytic activity was tested by looking at the cleavage of VAMP-1, -2 and -3. The three substrates were recombinantly produced as GFP-tagged constructs and were successfully hydrolysed by LHn/B (Figure 3.8A). The cleavage site was checked for each substrate by N-terminal sequencing (Alta Bioscience, UK) and confirmed to be at the expected position, corresponding to the site of BoNT/B action (Figure 3.8A).

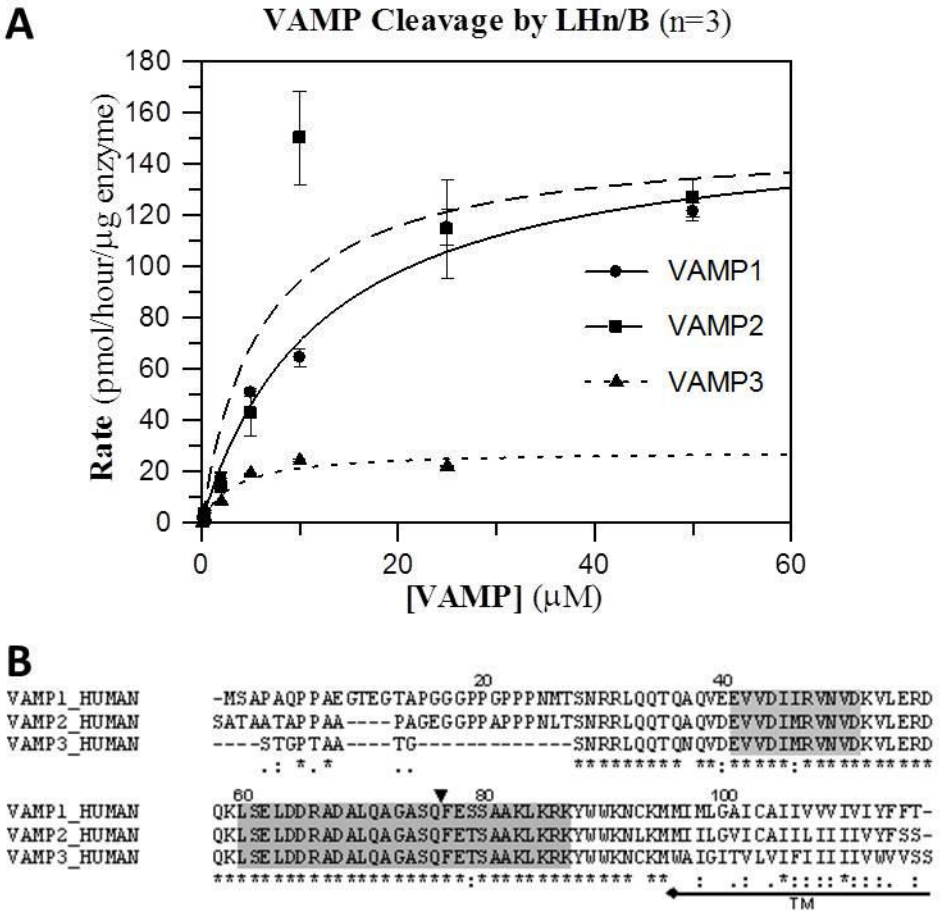

Figure 3.8. VAMP cleavage assay. (A) Enzymatic cleavage of VAMPs by LHn/B. Results were obtained by densitometry analysis and plotted. Non-linear fit was done with GraFit following Michaelis-Menten equation. Results for VAMP-1: *Km*= 12±3, *Vmax*= 157±12; VAMP-2: *Km*= 6.0±4.3, *Vmax*= 150±27; VAMP-3: *Km*= 3.2 ±1.4, *Vmax*= 27.9±3.5 (*Km* in µM, *Vmax* in pmol/hour/µg enzyme). **(B)** CLUSTALW alignment (EBI) of human VAMP-1,-2 and -3. Regions important for VAMP-2 cleavage by BoNTs are coloured in grey (Chen *et al.*, 2008; Sikorra *et al.*, 2008), numbers correspond to VAMP-2 residues. The C-terminal transmembrane domain is indicated.

A quantitative analysis of VAMP cleavage was performed by testing a range of different substrate concentrations at a fixed concentration of LHn/B (0.01 μg/ml). The plotted results were fitted according to a Michaelis-Menten equation. Substrate cleavage showed a higher maximal velocity (*Vmax*) for VAMP-1 and -2, (157±12 and 150±27 pmol/hour/μg enzyme respectively) than it did for VAMP-3 (27.9±3.5 pmol/hour/μg enzyme). Similar results were obtained with the BoNT/B control (unpublished results). A precise calculation of *Km* is difficult due to the nature of the assay, making the visualisation of low product levels hard to quantify. However determinations of *Km* only varied between 3.2 and 12 μM. It is likely that these apparent differences are within the errors of the estimations. Extensive work on the requirements for VAMP cleavage by Clostridial toxins have demonstrated that residues 60-87 of VAMP-2 were sufficient for efficient cleavage by BoNT, with residues within close distance of the scissile bond influencing the catalytic rate of reaction whereas several exosites on both sides of the bond were involved in binding (Chen *et al.*, 2008). The primary sequences of the three VAMP tested (Figure 3.8B) show that the scissile bond area is generally very conserved, apart for a minor difference at position P2 in VAMP-1 (VAMP-1/S – VAMP-2-3/T). This high degree of sequence conservation may explain the low variation in *Km* observed between the three substrates. The sequence alignment presents only one noticeable difference between VAMP-1 and -2 compared to VAMP-3 that lies near to the previously identified BoNT binding regions (VAMP-1-2/ Ala 39, 37 respectively compared to VAMP-3/ Asn 20). Further work is necessary to determine if this change accounts for the *Vmax* difference observed for these substrates. Another obvious difference in VAMP-3 compared to VAMP-1 and -2 is the N-terminal region which is considerably shorter and lacking the proline-rich region seen in VAMP-1 and -2. However, LC/B was proved to efficiently cleave a construct from VAMP-2 lacking residues 1-59 (Chen *et al.*, 2008). The C-terminal transmembrane regions of the VAMP proteins were not included in these recombinant substrates. These regions were replaced by a GFP-tag and are unlikely to affect substrate binding.

Activity of LHn/B on embryonic spinal cord neurons

Studies with BoNT/A have shown that Hn is necessary for the transport of LC into cells (Koriazova and Montal, 2003). LHn/B is similarly composed of the catalytic light chain with its translocating partner domain. The function of Hn in absence of the BoNT binding domain was investigated on embryonic spinal cord neurons (eSCN) and compared to the holotoxin's effect. VAMP cleavage was monitored by Western blotting eSCN lysates after incubation with each protein. The blots were quantified by densitometry and the proportion of uncleaved VAMP remaining after treatment was plotted (Figure 3.9).

The response followed dose-dependent sigmoidal curves for BoNT/B on VAMP-1 and -2, with sub-picomolar EC50. High concentrations of LHn/B also showed a sigmoidal dose-dependent response with EC50 of 15 and 170 nM on VAMP-2 and -1, respectively. This suggests that the binding domain of BoNT/B confers the toxin with more than 10^5-fold better efficiency in reaching its intracellular substrate. This is analogous to the phenomenon seen with BoNT/A and LHn/A where a 10^5 fold difference in concentration was observed on the inhibition of neurosecretion in similar cells (Chaddock *et al.*, 2002). It is expected that intracellular VAMP cleavage observed here would also result in a similar inhibition of neurotransmission, thus representing a reliable model to assay VAMP-specific BoNTs.

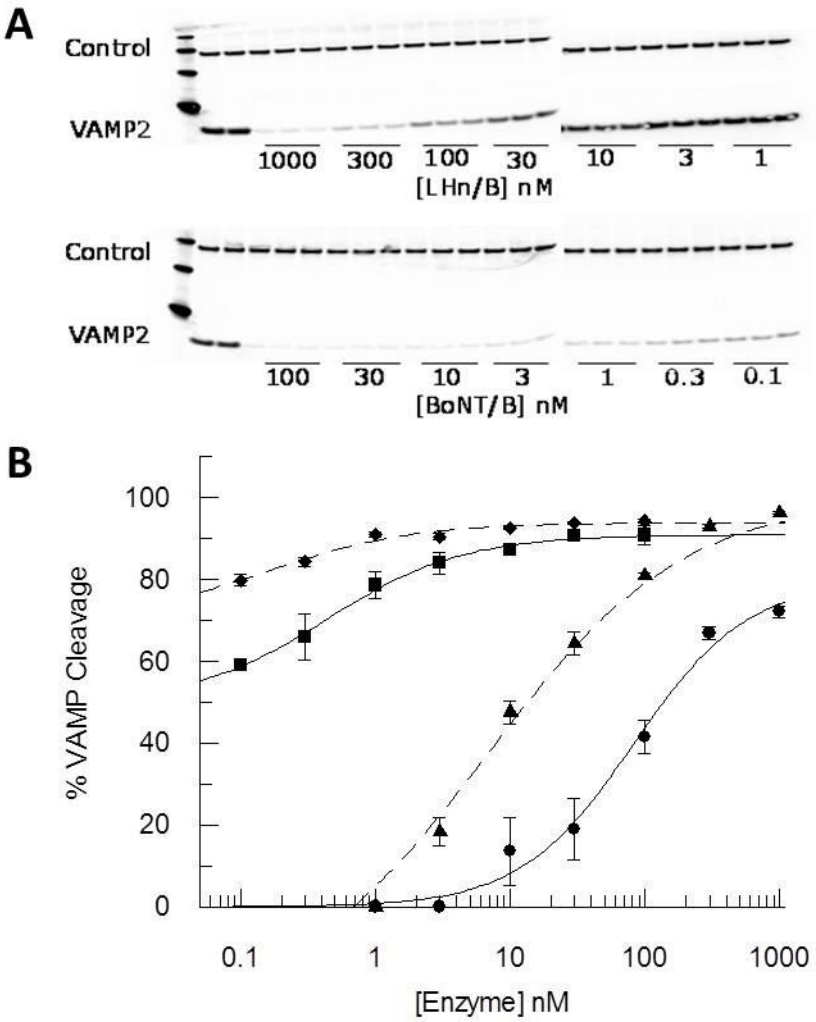

Figure 3.9. Spinal cord neuron assay. (A) Western blot analysis of VAMP-2 from cell lysates after treatment with decreasing concentrations of LHn/B and BoNT/B. **(B)** Analysis of intracellular cleavage of VAMP-1 (circles and squares) and -2 (triangles and diamonds) after treatment with LHn/B and BoNT/B respectively. Results were obtained by densitometry analysis and plotted. Non-linear fit was done with GraFit.

The mechanism of entry of LHn/B into eSCN is unknown but is likely to rely on a non-specific process. Fisher *et al.* (2008) reported LHn/A's ability to translocate LC over a wide pH range compared to the native toxin, therefore increasing the likelihood of translocation. On the other hand, the possibility of a low affinity receptor for LHn/B in this cell type should not be excluded. The effect observed here is enhanced by the high LHn/B concentration and the sensitivity of the system used for testing. However, further experimental evidence is required to fully understand the mechanism of internalisation by the LHn molecule.

BoNT/B and LHn/B both showed higher apparent efficiency in cleaving VAMP-2 compared to VAMP-1 in eSCN (Figure 3.9). The significance of this observation is not fully understood. Differences in the immunoreactivity, abundance and sub-cellular localisation of these two SNAREs may play roles in accounting for this observation. VAMP-2 is known to be the most abundant protein in synaptic vesicles (Takamori *et al.,* 2006). One possibility is that VAMP-2 might compete with VAMP-1 for binding to the LC. VAMP-3 was also monitored in treated eSCN lysates and did not show any evidence for VAMP-3 cleavage (results not shown). The role of VAMP-3 (also called cellubrevin) has been previously investigated in non-neuronal cells using Clostridial tetanus toxin (McMahon *et al.,* 1996). It highlighted its importance in the recycling of the plasma membrane, early endosome pathways, epithelial cell migration and adhesion (Proux-Gillardeaux *et al.,* 2005). Despite VAMP-3 homology with VAMP-1 and -2 (Figure 3.8B), there is no strong evidence for a direct role of VAMP-3 in neurosecretion. The lower specificity of LHn/B for this substrate demonstrated in the cell-free assay, along with the lack of intracellular proteolysis in eSCN, suggests that the toxin may have evolved to specifically target SNAREs directly mediating neurotransmission.

In the present study, the LHn/B fragment consisting of the catalytic and translocation domains of BoNT/B, is shown to retain its native structure after deletion of the Hc domain. The functionality of the fragment was also demonstrated *in vitro* with its ability to cleave several VAMP substrates. Similar to wild type BoNT/B, LHn/B showed higher specificity for VAMP-1, and -2 compared to VAMP-3. Furthermore, LHn/B retained an ability to cleave intracellular VAMP in spinal cord neurons, indicating an intrinsic capacity of Hn to transport its catalytic partner within the cytosol. A detailed understanding of the processes involved in internalisation and translocation into the cytoplasm is not yet available and requires further analysis, which may in turn give some indication on the translocation mechanism in the corresponding holotoxins.

LHn/B does not have the BoNT cell binding domain. The extent to which it can interact with and become internalised by other cell types represents an area for further study. This may reveal LHn/B as a useful pharmacological tool in the study of VAMP-mediated secretion events.

LHn fragments studied so far constitute stable soluble proteins that conserve the functionality of their parent holotoxin but lack the neuron specific targeting conferred by the binding domains of full-length BoNT. These fragments provide additional tools for understanding the structure-function relationship of BoNT domains in the intoxication process, and offer a new strategy for therapeutic and vaccine development (Chaddock and Acharya, 2011). Determining the structure of other LHn serotypes will help refine our knowledge and explain the variations in activities seen between the botulinum neurotoxin serotypes.

3.3. Structure of LHn/D

3.3.1. Introduction

Of the seven types of botulism, A, B, E and F are known to cause the disease in human while C and D have only been observed in animal cases. More particularly, D has been responsible for several recent outbreaks of botulism in cattle (Steinman *et al.*, 2007). This has raised some interest in this serotype and its precise mechanism of action.

No cases of type D human botulism have ever been recorded. Coffield *et al.* (1997) investigated the impact of serotype D on human tissues and demonstrated its inability to block neuromuscular transmission when tested at level 10 times higher than that of serotype A. This result implies that BoNT/D lacks a specific receptor to target human neurons. BoNT/D however acts similarly to the other botulinum neurotoxins by targeting one of the intracellular SNARE proteins. Synaptobrevin (or VAMP) was identified as BoNT/D specific substrate (Schiavo *et al*, 1993). It is cleaved at a unique position, the Lys-59- Leu-60 peptide bond.

Efforts to understand the mechanisms of BoNT/D have led to the determination of the crystal structures corresponding to the binding (Hc) and (LC) catalytic domains. Interestingly, several recent studies have focused on the Hc structure to explain the unusual entry of BoNT/D into neurons. As a summary, Hc/D was shown to lack a ganglioside binding motif common to other BoNT serotypes (H…SXWY) despite the domain conserving its typical β-trefoil structure (Karalewitz *et al*., 2010). While one of the crystal structures highlighted the presence of two carbohydrate binding pockets (Strotmeier *et al*., 2010) that could be involved in neuron recognition, a more recent structure along with binding and immunofluorescent assays seemed to favour a ganglioside-mediated entry (Kroken *et al*., 2011). Key residues unique to BoNT/D form a hydrophobic loop corresponding to the ganglioside binding region and are responsible for this serotype selectivity towards particular animal species.

The high resolution crystal structure of LC/D (Arndt *et al.*, 2006) showed features similar to the other BoNTs' LC and gave some clues regarding its mode of substrate binding. LC/D is a zinc metalloprotease that possess the classical zinc binding motif (HEXXH) at the active site. It is of a mixed composition with nine α-helices and twelve β-strands. The main differences with other LC are in the loop regions which share the lower sequence identity, and are likely to be involved in substrate binding (Breidenbach and Brünger, 2004). Indeed, the long substrate requirement for all LC is a feature shared by LC/D since the minimum substrate length corresponds to segment Thr27-Gln 76 of VAMP2

(Yamasaki *et al.*, 1994). This was highlighted by modelling of the LC/D-VAMP2 interaction based on the LC/A-SNAP25 complex (Arndt *et al.*, 2006), which showed a putative interface largely based on LC/D hydrophobic surface potential.

Despite its lack of human toxicity, BoNT/D has been proposed as an alternative to other serotypes for engineering proteins able to transport active enzymes within neuronal cells. Bade *et al.* (2004) described the building of full length recombinant BoNT/D associated with various reporter enzymes. Intracellular function of BoNT/D was conserved and added activity observed when associated with an amino-terminal LC/A domain. Intracellular localisation was also confirmed by fluorescence microscopy with a GFP-tag construct, as well as with the successful transport of a luciferase-BoNT/D molecule. Overall it was observed that successful transportation of an enzymatic cargo depended not only on the size of the attached protein, but also on its structural ability to form unfolded intermediates.

BoNT/A and /B are currently the only serotypes approved for medical uses. With the emergence of immuno-reactivity among patients, other serotypes could provide a useful alternative (Turton *et al.*, 2002). BoNT/D efficiency at targeting intracellular VAMP (Schiavo *et al*, 1993) makes it an interesting candidate for protein engineering. Especially, the successful retargeting of BoNT activity for therapeutic purposes, by association of the LHn fragment with various ligand (Foster *et al.*, 2009), highlights the relevance of using serotype D. In this context, BoNT/D inability to bind human neurons would not be an issue as Hc would be replaced by a specific targeting polypeptide.

Structural and biochemical characterisation of the LHn fragments from serotypes A (Section 3.1; Masuyer *et al.*, 2009) and B (Section 3.2; Masuyer *et al.*, 2011) have demonstrated their stability and functionality, representing an important step forward for the design of novel molecules based on these frameworks. In order to assess its applicability for further pharmaceutical development, LHn/D was cloned, expressed and purified. Crystallisation trials were set up and yielded protein crystals that diffracted to 2.2 Å. Careful processing of the data showed that only the LC domain was present in the crystals. The structure of LC/D was therefore refined and compared to the published structure (PDB 2FPQ, Arndt *et al.*, 2006). An alternative LHn/D construct was also crystallised and showed a similar truncated LC/D. However, several noticeable differences were observed between these structures and are discussed below.

3.3.2. Materials and methods

LHn/D cloning, expression and purification

The synthetic gene encoding 872 amino acids of LHn/D was cloned into modified pET vector (Novagen, UK) with a C-terminal 6 x His-tag and transformed into *E. coli* BL21 expression cells. The LHn/D gene was engineered to encode for an enterokinase cleavage site (DDDDK) between the LC and Hn domain between positions 448-452. The clone was provided by Syntaxin Ltd. Expression cultures of LHn/D were grown in 1L terrific broth at 200 rpm, 37°C until OD_{600} reads 0.5-0.6. Then cultures were incubated at 16°C and induced with IPTG (1mM). Cells were harvested after 18 h, and stored at -80°C until further use.

Cells were resuspended in 50 mM HEPES, pH 7.2, 0.2 M NaCl (buffer A) and lysed at 20,000 psi with a homogeniser (Constant Systems Ltd). Lysate was centrifuged for 45 minutes at 18,000 g. Soluble fraction was loaded onto a Ni^{2+}-charged chelating sepharose column (GE Healthcare). LHn/D eluted at 100 mM imidazole (dissolved in buffer A) and dialysed overnight at 4°C against buffer A. Activation of purified LHn/D is achieved by enterokinase (New England BioLabs) treatment. The cleaved fusion protein was supplemented with ammonium sulphate to 1 M and loaded onto a Toyopearl Phenyl-650M column, equilibrated with 50 mM HEPES, pH 7.2, 1.0 M ammonium sulphate. LHn/D eluted at 0.4 M ammonium sulphate (in buffer) and was dialysed overnight against buffer A at 4°C. The sample was finally concentrated using Vivaspin 50000 MWCO concentrator (Millipore) to 5.2 mg/ml and stored at -20°C. All concentrations were determined by A_{280} measurement. A SDS-PAGE summary of the purification is shown in figure 3.10A.

A non-activated, single-chain, form of LHn/D was also purified (scLHn/D). Briefly, initial preparation was identical to its activated counterpart and then followed by a two-step purification process. After a first affinity chromatography (His Trap HP, GE Healthcare), the fractions of scLHn/D were pooled and supplemented with ammonium sulphate to 1 M for loading onto a Phenyl-650M column (Toyopearl). The sample was dialysed and concentrated to 5.2 mg/ml (storage at -20°C). SDS-PAGE of the purified material is shown in figure 3.10B.

A protein based on LHn/D with a Factor Xa activation site (LHn/D/Xa) was provided by Syntaxin Ltd for crystallographic analysis. SDS-PAGE of the purified material is shown in figure 3.10C, mass spectrometry confirmed the construct was of the expected size (results not shown).

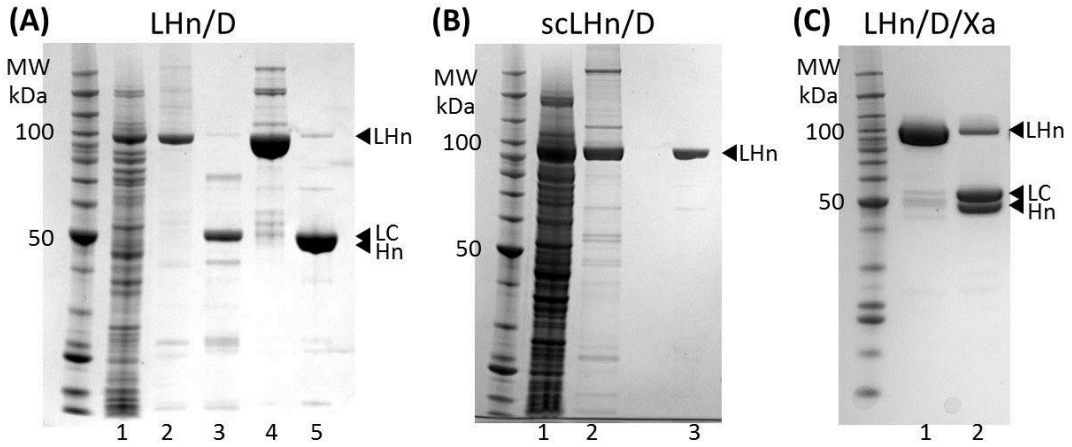

Figure 3.10. SDS-PAGE analysis of LHn/D, scLHn/D, and LHn/D/Xa purifications. (A) Summary of LHn/D purification. Lane 1, cell lysate supernatant; 2, 3, enterokinase-activated eluate from first purification step in oxidised (O) and reduced (R) conditions (respectively); 4, 5, eluate from final purification step (O and R). **(B)** Summary of scLHn/D purification. Lane 1, cell lysate supernatant; 2, eluate from first purification step; 3, eluate from final purification step. **(C)** Lanes 1, 2, purified LHn/D/Xa sample provided (O and R).

Characterisation of LHn/D by Western-blot.

Final sample of purified LHn/D was run on a SDS-PAGE gel and blotted on a PVDF membrane (Millipore). Primary detection was performed using anti-LC/D and anti-Hn/D antibodies raised in rabbit (obtained from Eurogentec), a Tetra-His (QIAGEN) antibody was used for the C-terminal tag detection. For secondary recognition, horseradish peroxidase-conjugated anti-rabbit and anti-mouse antibodies (respectively) were applied (Sigma). Enhanced chemiluminescent (ECL) substrate reagent (Thermo Scientific) was utilised for detection.

Crystallisation

Different conditions gave crystallisation hits for LHn/D, all in the shape of thin needle-like crystals. A single condition was successfully repeated by hanging-drop with manual trials: 24% PEG1500, 20% glycerol (with 3 µl drops - 2 µl protein and 1 µl mother liquor - against a reservoir 500 µl). After several rounds of optimisation and seeding, single larger needles were grown in 10% PEG1500, 20% glycerol and x-ray diffraction data could be collected (Figure 3.11).

Crystals for scLHn/D were also obtained. Crystals grew in stack of thin rod-shape forms in 0.1 M BIS-TRIS-propane pH 6.5, 0.3 M potassium thiocyanate, 18% PEG3350. However, scLHn/D crystals failed to diffract to better than 6 Å.

A single LHn/D/Xa (long needle-shape) was obtained in 0.05 M sodium citrate, 20% PEG3350 in a single drop from primary sitting-drop crystallisation screens, observed after a one-year incubation period.

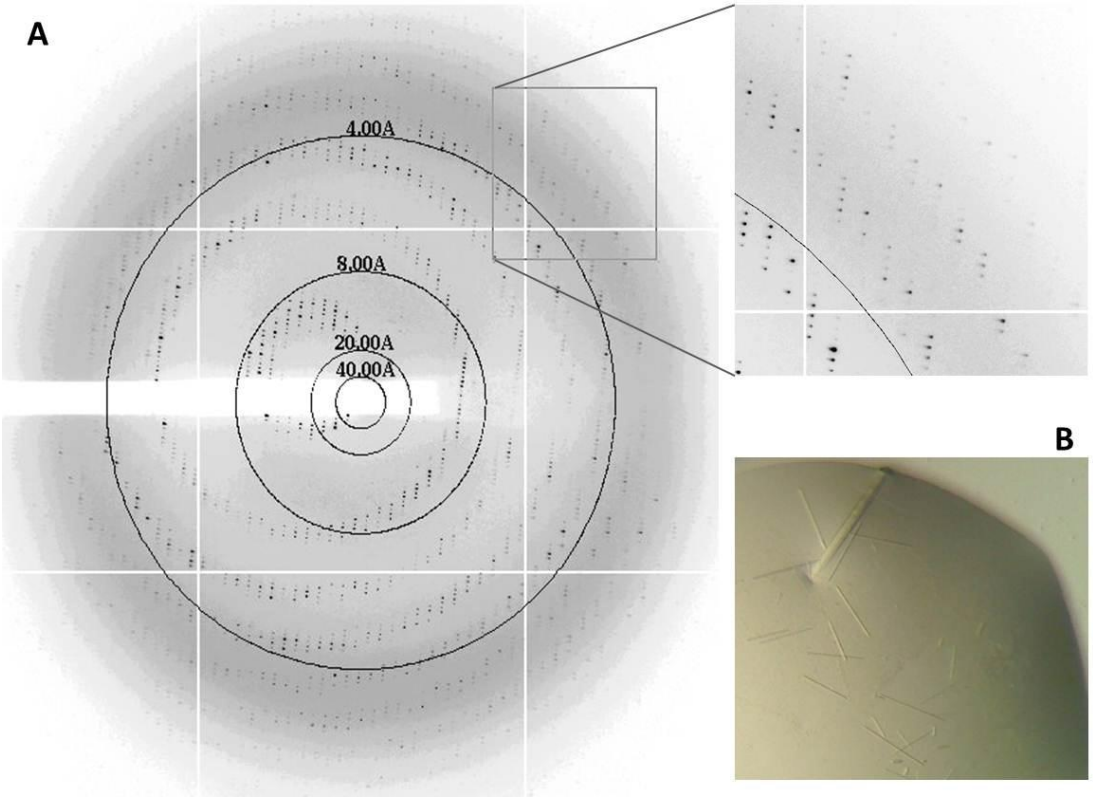

Figure 3.11 Crystals and x-ray diffraction of LHn/D. (A) Diffraction image collected at DLS IO3 where LHn/D crystals diffracted at 2.2 Å. **(B)** Crystals of LHn/D after optimisation in 10% PEG1500, 20% glycerol.

Data collection and structure determination

X-ray diffraction data for LHn/D were collected at the Diamond Light Source, UK, beamline IO3. A complete dataset to 2.2 Å was collected from a single crystal at 100K (no cryoprotectant added) using a Quantum-4 CCD detector (ADSC Systems, CA). The data were processed and scaled in hexagonal space group $P6_5$ using MOSFLM and SCALA (CCP4, 1994; Leslie, 2006) (Table 3.3).

Initial phases were obtained by molecular replacement (MR) using PHASER (McCoy *et al.*, 2007). MR trials were first tested with the coordinates of LHn/A and B (PDB code 2W2D, 2XHL respectively, Masuyer *et al.*, 2009, 2011) but failed to give a solution. Then search parameters were set for a single LC domain (PDB code 2FPQ, Arndt *et al.*, 2006). A convincing solution was found (Z score = 34.0) although it represented residues 1-400. Hn

searches with this LC solution were unfruitful. The final structure was solved after searching for two LC molecules per asymmetric units (Z score = 71.6).

Crystallographic refinement was carried out using REFMAC5 (version 5.5) (Vagin *et al.*, 2004). Manual adjustments and model fitting was done using Coot (version 0.6.1) (Emsley and Cowtan, 2004). Water molecules were added at positions where F_o–F_c electron density peaks exceeded 3σ and potential H-bonds could be made. The structure was validated using MolProbity (Davis *et al.*, 2007). Structure figures were drawn with PyMOL (DeLano Scientific LLC).

X-ray diffraction data for LHn/D/Xa were collected at the Diamond Light Source, UK, beamline I24 (microfocus beamline). A complete dataset to 3.2 Å was collected from a single crystal at 100K (after transferring the crystal to 25% PEG3350) using a PILATUS-6M (Dectris) detector. The data were processed and scaled in hexagonal space group $P6_5$ using MOSFLM and SCALA (CCP4, 1994; Leslie, 2006). The unit cell parameters (Table 3.4) being similar to the ones from the LHn/D diffraction, 2 LC molecules were used in the search parameters with Phaser (McCoy *et al.*, 2007). This led to a straightforward solution that was set through refinement with REFMAC5 (version 5.5; Vagin *et al.*, 2004) and confirmed that again, only LC was present in the crystal.

Table 3.3. LHn/D X-ray data collection and LC/D refinement statistics

A. Data collection statistics	
Space group	$P6_5$
Number of molecules/asymmetric unit	2
Cell dimensions	a =99.2, b =99.2, c =223.9 Å; α, β =90,γ =120°
Resolution range (Å)	50-2.2
R_{sym}[1] (%)	9.2 (63.7)
I/σI (outer shell)	9.9 (2.1)
Completeness (outer shell) %	89.7 (86.3)
Total no. of reflections	229,707
Unique no. of reflections	32,013
Redundancy	4.1 (4.0)
Wilson B-factor (Å^2)	33.2
B. Refinement statistics	
Resolution range (Å)	50-2.2
R_{cryst}[2] (%)	22.8
R_{free}[3] (%)	25.5
Number of non-H atoms	
Protein	6,266
Metal	2 Zinc atoms
Water molecules	233
Average temperature factor (Å^2)	24.5
RMSD in bond lengths (Å)	0.006
RMSD in bond angles (°)	0.83

[1] $R_{sym} = \Sigma_h \Sigma_i |I(h) - I_i(h)| / \Sigma_h \Sigma_i I_i(h)$, where $I_i(h)$ and $I(h)$ are the i^{th} and the mean measurements of the intensity of reflection *h*, respectively.
[2] $R_{cryst} = \Sigma_h |F_o| - |F_c| / \Sigma_h F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection *h*, respectively.
[3] R_{free} is equal to R_{cryst} for a randomly selected 5.0% subset of reflections not used in the refinement.

Table 3.4. LHn/D/Xa X-ray data collection and LC/D refinement statistics

A. Data collection statistics	
Space group	$P6_5$
Number of molecules/asymmetric unit	2
Cell dimensions	a =100.4, b =100.4, c =224.0 Å; α, β =90, γ =120°
Resolution range (Å)	69-3.2
R_{sym}[1] (%)	13.9 (29.1)
I/σI (outer shell)	10.3 (5.5)
Completeness (outer shell) %	93.3 (94.0)
Total no. of reflections	89,970
Unique no. of reflections	19,625
Redundancy	4.6 (4.5)
Wilson B-factor (Å^2)	32.6
B. Refinement statistics	
Resolution range (Å)	50-3.2
R_{cryst}[2] (%)	21.9
R_{free}[3] (%)	24.9

[1] $R_{sym} = \Sigma_h \Sigma_i |I(h) - I_i(h)| / \Sigma_h \Sigma_i I_i(h)$, where $I_i(h)$ and $I(h)$ are the i^{th} and the mean measurements of the intensity of reflection h, respectively.
[2] $R_{cryst} = \Sigma_h ||F_o| - |F_c|| / \Sigma_h F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection h, respectively.
[3] R_{free} is equal to R_{cryst} for a randomly selected 5.0% subset of reflections not used in the refinement.

3.3.3. Results and discussion

Structure of LC/D

The crystal structure obtained from the crystallisation trials with LHn/D actually corresponds to a truncated version of LC/D, residues 1-400 (LC/D). LC/D crystallised as a dimer in space group $P6_5$ with two molecules per asymmetric unit and the structure was solved at 2.2 Å resolution. The structure was refined to a final R_{free} = 25.5%, and R_{cryst} = 22.8% (Table 3.3, Figure 3.12), with 99.7% of amino acids in the Ramachandran plot allowed region.

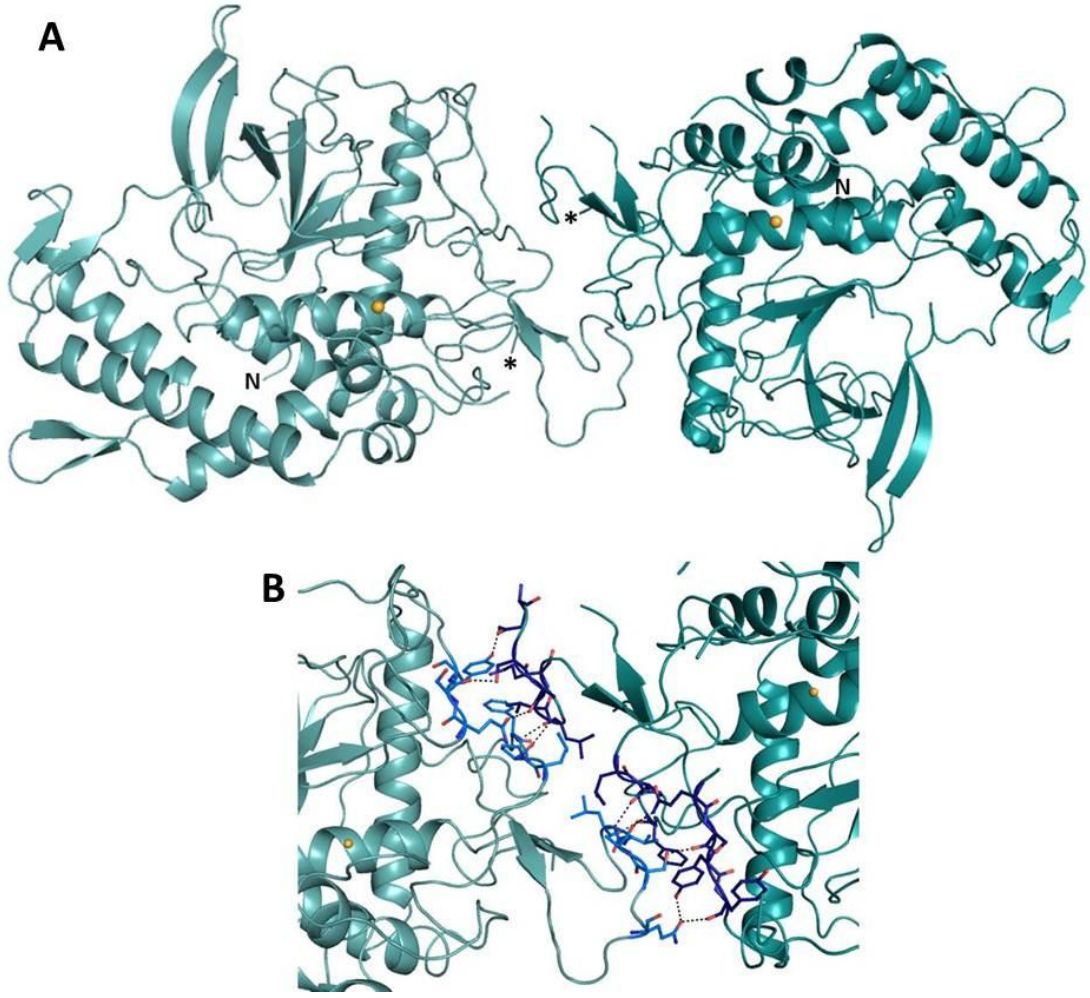

Figure 3.12. Crystal structure of LC/D. (A) Ribbon diagram representation of LC/D structure, LC in cyan (molecule A and B in light and dark cyan respectively). Zinc ion is shown as orange spheres. N- and C-termini are marked (N and *, respectively). (**B)** Dimeric interface between the two LC molecules. Residues involved in the interaction are shown in stick and ball representation (light and dark blue for molecules A and B respectively), potential hydrogen bonds are linked by black dashes.

The structure presents two LCs interacting in a crystallographic dimer through 11 potential hydrogen bonds and 17 van der Waals interactions, corresponding to an interface area of 892 $Å^2$ (Figure 3.12B). The two LCs are nearly identical and superpose with a rmsd of 0.2 Å (over 383 C_α). No electron density was observed for residues 174-175, and 254-261 of chains A and B. Regions 64-67 and 206-208 could not be modelled in chain B only. LC/D is a metalloprotease and the zinc atom could be seen in both molecules, with a classic tetrahedral coordination by His 229, His 233, Glu 269 and a water-mediated Glu 230 (Figure 3.13).

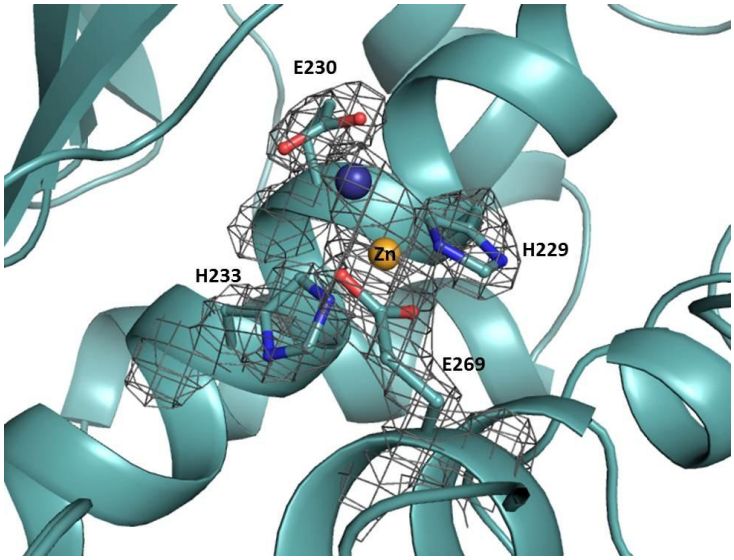

Figure 3.13. Zinc coordination in LC/D. Cartoon diagram representation of LC/D (cyan). Zinc ion is shown as an orange sphere, water molecule in dark blue. Residues involved in the zinc coordination are shown in stick and ball representation. 2|Fo|-|Fc| map contoured at 1σ.

Overall, the LC/D structure solved here confirms the LC fold common to all BoNT, and is identical to the already published LC/D structure (PDB 2FPQ, Arndt *et al.*, 2006) with an rmsd of 0.6 Å (over 391 Cα-atoms). However, comparison of the two structures (Figure 3.14) highlighted some interesting differences. PDB 2FPQ was solved to 1.6 Å and corresponds to residues 1-424, thus with 24 additional C-terminal amino acids. Superposition of 2FPQ with our structure indicates that this C-terminal region would clash with the crystallographic dimer interaction, therefore the truncated LC molecule favoured the new crystal form.

In 2FPQ, residues 166-170 appear in two possible conformations, one is a random coil close to the zinc atom (Figure. 3.14B) and in the second case this region makes up for the short β-strand (β8) of the active site. This latter conformation corresponds to what is

observed in our LC/D, confirming its more likely prominence and physiological relevance as it is also observed in the other LC structures (Arndt *et al.*, 2006). This may indicate a degree of flexibility of this segment in vicinity of the active site, to provide space for substrate binding in the catalytic pocket.

The LC/D structure presented here offers additional information regarding loop 206-214 which was not visible in 2FPQ. The interactions of residues 208-216 (chain A) with the second LC/D of the crystallographic dimer (chain B) stabilises this loop into a random coil otherwise in a solvent accessible surface (such as in chain B). The full extent of the binding interaction between LC/D and its VAMP substrate is not known although it is likely to involve several exosites. It is therefore possible that several of the unstructured loops present around the catalytic pocket may be implicated, in particular loops 175, 210, and 250 (Figure 3.14). Furthermore, loop 145 linking the β5 and β6 strands shows a slight shift compared to 2FPQ, this may be explained by the crystallographic packing.

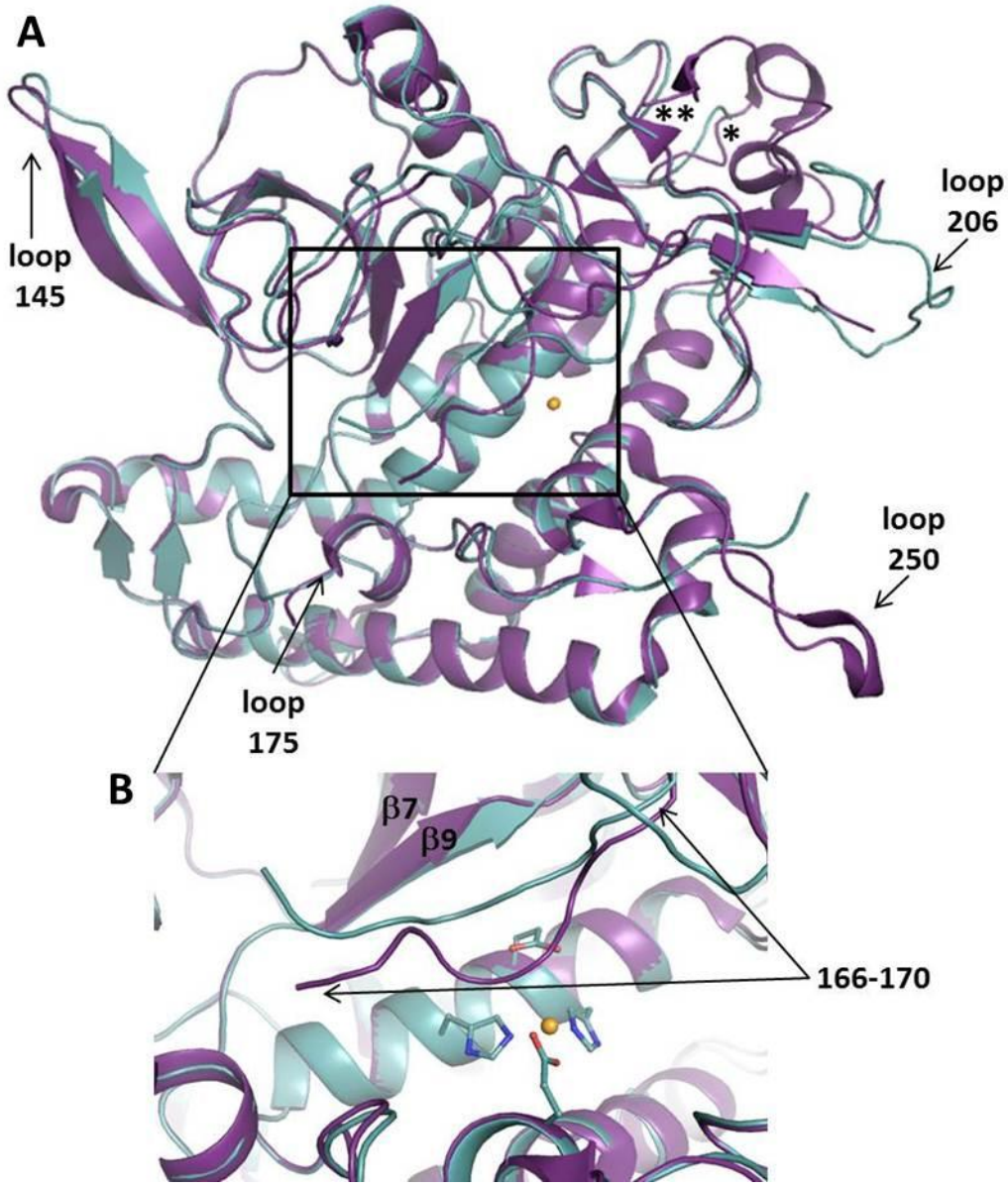

Figure 3.14. Comparison of LC/D structures. (A) Superposition of LC/D structures, LC/D in cyan, PDB 2FPQ in purple, with their C-termini marked by * and ** respectively. Zinc ion is shown as orange spheres. (**B)** Alternative conformation of residues 166-170 in 2FPQ. Catalytic site shown in stick and ball representation.

Truncation of LHn/D – possible explanation

The crystal structure described here exposed an unexpected result in only showing the LC domain. The nucleotide sequence of the clone from which LHn/D was expressed had been checked. Analysis of cell expression indicated an overexpressed protein at the expected size of approximately 100 kDa (Figure 3.10). Purification of this protein was carried out. Specifically, a nickel-affinity column was used, and thus successfully utilised the poly-histidine tag on the C-terminal end of the translocation domain. Activation of LHn/D into a

di-chain molecule was achieved using the engineered enterokinase cleavage site. A faint band corresponding to inactivated material is still visible at 100 kDa in reducing conditions but the activation yield was considered acceptable to complete the purification by hydrophobic interaction. The final concentrated material was over 90% pure (Figure 3.15). Integrity of the final sample was examined by Western blotting with domain specific antibodies. The blots confirmed the presence of both LC and Hn domains in reducing and oxidising conditions. The poly-His detection also demonstrated the presence of the full-length Hn chain.

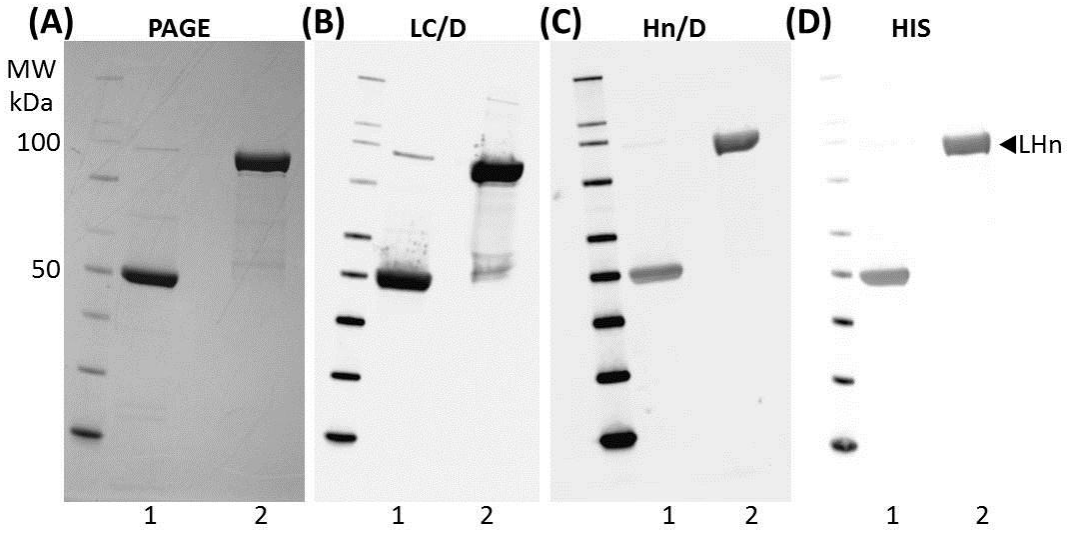

Figure 3.15. Western blot analysis of LHn/D. Purified LHn/D sample in reduced and oxidised conditions (lanes 1 and 2 respectively) in **(A)** SDS-PAGE; (**B)** anti-LC/D Western blot (W.B); **(C)** anti-Hn/D W.B; **(D)** anti-His W.B.

The solved LC/D structure leaves little ambiguity regarding the composition of the crystal from the LHn/D trials. Despite the starting sample containing an intact LHn/D, only a truncated LC was observed. At first it was hypothesised that truncation could result from unspecific cleavage by enterokinase which is used in the activation step of LHn/D. However from the SDS PAGE and Western blot analysis performed (Figure 3.15), there is no clear indication of the presence of truncated material, which would correspond to a protein of 45.8 kDa. Also, analysis of the primary sequence did not indicate any potential cleavage site for enterokinase. To investigate this possibility further, a single-chain, non-activated, scLHn/D was purified (Figure 3.10) and crystallisation trials were performed. However, this construct failed to crystallise in the same condition as LC/D. Crystals were nonetheless obtained in a different set of conditions but did not diffract to better than 6 Å at the Diamond Light Source (UK). Additionally, a protein based on the LHn/D backbone but activated by an alternative exoprotease (Factor Xa) was provided (LHn/D/Xa). A single crystal was obtained from new conditions and data collected to 3.2 Å. Analysis of the resulting x-ray diffraction data

showed it corresponds to the same crystal form as LC/D with nearly identical cell parameters (Table 3.4). This structure did not provide any additional information than the one described previously. However, it did present a truncated LC molecule with the electron density ending at residue 400.

In view of the structures solved from the various LHn/D constructs, it seems unlikely that the truncation observed in LC/D was due to the activation treatment by exoproteases. Despite not noticing any cleavage product in the stored LHn/D sample, the possibility of a proteolytic degradation during the crystallisation incubation period could not be ruled out. LHn/D is indeed a metallo-protease with high specificity for synaptobrevin (Schiavo *et al.,* 2000). Evidence of BoNT autoproteolysis has been described for several serotypes (DasGupta *et al.,* 2005) and provided a basis for the identification of potential cleavage sites.

Substrate recognition by BoNTs involves an essential nine-residue long motif common to all SNARE proteins (Rossetto *et al.,* 1994). DasGupta *et al.* (2005) identified and characterised analogous SNARE motifs within BoNTs. The motifs discovered correlated with experimental results of autoproteolysis that were available for serotypes A, B and E. From these observations, DasGupta *et al.* (2005) predicted several such binding sites in other clostridial neurotoxins. These motifs were located in regions of the toxins that adopted a α-helical structure, were surface exposed and contained negatively charged residues. One of the predicted segments in LC/D was at position 387-395 (Figure 3.16). Noticeably, experimental observations (for types A, B and E) indicated that autoproteolytic cleavage occurred two to four amino acids away from the putative recognition site. This might therefore provide a possible explanation for the truncation observed at position 400 in the crystal structure.

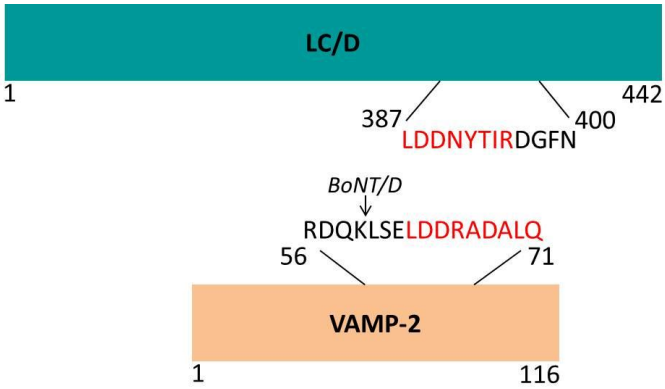

Figure 3.16. SNARE motif in LC/D. Sequence alignment of LC/D with VAMP-2. DasGupta *et al.,* (2005) identified the putative SNARE motif "LDDNYTIR" in LC/D with "Y" the only disagreement for substrate recognition.

Other possible explanations for the observed truncation such as chemical proteolysis and dehydrogenase-like activity (DasGupta *et al.,* 2005) may be discarded by considering the conditions under which LC/D crystallised here.

Further characterisation of the crystallised molecule may be necessary to fully understand the cause of the truncation. Unfortunately crystallisation was not readily reproducible from a different purification batch of LHn/D. N-terminal sequencing or mass spectrometry would provide useful information on the precise cleavage site. Even though the electron density map was not present beyond residue 400 in the two structures presented here, an unobservable segment beyond this point, in the solvent accessible area, may not be excluded.

Towards the crystallisation of LHn/D

Determination of the LHn/D crystal structure may rely on preventing potential autoproteolysis. This may be achieved using several different methods. Firstly, crystallisation in the presence of a metal-chelating agent (like for example EDTA) could be tested. The zinc atom, although essential for enzymatic activity, is not involved in any structural role. Removal of the catalytic centre may thus lead to the structure of the apoprotein. Secondly, a more complex approach would be to mutate key residues involved in the coordination of the metal ion, so as to hinder the enzymatic activity of LC/D. Such method was successfully used to determine the structure of LC/A in complex with its substrate (Breidenbach and Brünger, 2004). Finally, mutations in the putative SNARE motif recognition site described above may offer an opportunity to crystallise a catalytically active form of LHn/D. However this latter option might disrupt the enzyme's tertiary structure since it would involve a change of the local α-helical conformation.

3.4. Small-angle x-ray scattering (SAXS) analysis of LHn fragments

3.4.1. Introduction

The crystal structures of BoNTs have highlighted a stable fold of their components, with each domain retaining its typical fold (Swaminathan, 2011). However there is little information available regarding the structure of these molecules in a more biologically relevant environment, i.e. in solution. Small angle x-ray scattering (SAXS) offers an alternative approach to study the molecular structure of proteins. Although it provides limited information on the size and shape of proteins, it may prove useful, in combination with other methods, to study inter- and intra-molecular interactions.

One of the key elements for BoNT toxicity is the toxin's ability to transport its catalytic domain (LC) into the cytosol of cholinergic motor neurons. This step is mediated by the translocation domain (Hn) and importantly is dependent upon the pH gradient between the endosome (pH 5.0) and the cytosol (pH 7.0) (Fischer *et al.,* 2008b). There is therefore an essential pH-dependent structural transformation occurring at this stage of the intoxication process. Koriazova and Montal (2003) demonstrated that LC/A undergoes a complete secondary structure unfolding at pH 5.0, which allows for its passage through the putative narrow pore formed by Hn. However, analysis by circular dichroism spectroscopy (Li and Singh, 2000) showed low pH (4.7) had little consequences on LC/A secondary structure. More recently, another CD spectroscopy study concluded that the endosomal pH induced the transformation of LC/A into a molten globule state (Cai *et al.,* 2006). This would result in a protein with similar secondary structure characteristics to its native counterpart, but with a different tertiary fold. Additionally, the crystal structures of BoNT/B, solved in different pH buffer conditions (pH 4.0 to 7.0) did not show any significant alterations (Eswaramoorthy *et al.,* 2004).

Small angle x-ray scattering is a pertinent technique to study the domain interactions within BoNT and identify potential structural differences between serotypes. Furthermore, the LHn fragment, which holds the functional domains responsible for the toxin's intracellular activity, is a particularly interesting molecule to study to check the possible pH-dependent mechanisms. Here, the details of the first solution scattering experiments carried on BoNT and an investigation into the effect of pH on the tertiary structure are described. Although data collection was not optimal, this study produced useful information on the structure of LHn/A, /B and /D in solution, including low resolution *ab initio* models. No structural modifications were observed under endosomal pH conditions.

3.4.2. Materials and methods

Sample preparation

Protein samples corresponding to LHn/A, LHn/B, LHn/D and scLHn/D were prepared for SAXS analysis. All samples were purified as described previously, and produced protein crystals. Monodispersity was checked by dynamic light scattering (Malvern Instruments).

The four proteins were stored at -20°C in 0.05 M HEPES 7.2, 0.2 M sodium chloride. Buffer exchange was performed by dilutions in the appropriate buffers followed by a concentrating stage using Vivaspin 50000 MWCO concentrator (Millipore). Three different pH were tested corresponding to 0.05 M HEPES pH 7.0, 0.1 M TRIS acetate pH 5.0, and 0.1 M TRIS acetate pH 4.0 respectively. The proteins were kept at 5.0 mg/ml and stored at 4°C until day of testing. On the day of the experiment, samples were diluted in the appropriate buffer to give three final concentrations at 5, 2.5 and 1 mg/ml. These dilutions were checked by absorbance measurement at 280 nm using a NanoDrop ND-1000 (Thermo Scientific), and concentration discrepancies were corrected for data analysis.

Data collection and analysis

Solution scattering data for the four LHn molecules were collected at DESY, Germany, EMBL Hamburg, beamline X33 (BioSAXS). For each sample, a buffer measurement was performed before and after the sample at room temperature, using a PILATUS-1M (Dectris) detector. SAXS measurements were performed for each sample at three concentrations. Calibration was carried out with BSA at 5mg/ml in 0.05 M HEPES pH 7.5 as a molecular weight standard. The data were processed and averaged with PRIMUS (Konarev *et al.,* 2003) followed by particle distance distribution function *P(r)* analysis using GNOM (Svergun, 1992). Data considered of good quality were further treated by using the on-line server for DAMMIF (Franke and Svergun, 2009) to produce twenty *ab initio* models that were averaged with DAMAVER (Volkov and Svergun, 2003). Successful models were compared to known crystal structures of LHn by manual docking (PyMOL, DeLano Scientific LLC).

3.4.3. Results and discussion

The first experiments analysing the structure of LHn fragments in solution by SAXS resulted in the production of several models. A few of the conditions tested did not yield reliable data. This was the case for most of the proteins tested at pH 4.0 that showed a high amount of aggregation during manipulation. All the molecules prepared have a theoretical pI of approximately 5.1, and thus did not survive under such acidic conditions. Furthermore, a

key element for SAXS experimentation is to have a background buffer control that matches exactly with the solution the sample is in. The method for sample preparation described above was not optimal and thus led to a difficult interpretation for some of the data collected (Table 3.5).

Table 3.5. Summary of SAXS data collected at DESY (X33).

	pH 4.0	pH 5.0	pH 7.0
LHn/A	model (*ag*[1].)	model	*buf. mis.*[2]
LHn/B	*ag.*	model	model (*buf. mis.)*
LHn/D	*ag.*	model	model (*buf. mis.)*
scLHn/D	*ag.*	model	model (*buf. mis.)*

[1] *aggregation;* [2] *buffer mismatch.*

Despite the data quality not being optimal, the overall features of LHn in solution could be observed. The experiment was aimed at studying differences in solution scattering between the serotypes and the potential effect of acidic pH on LHn. An example of data collected for LHn/A at pH 5.0 is shown in figure 3.17. While the estimations for the maximum dimension D_{max}, and the radius of gyration R_g should be treated carefully, it is interesting to notice their consistency for all of the proteins tested (Table 3.6). Comparison of the data indicated a good correlation between the four LHn molecules analysed. *Ab initio* models produced were consistent with these data and showed a common shape and size of the samples (Figure 3.18).

Table 3.6. Small angle x-ray scattering data statistics for LHn samples at pH 5.0.

	P(r) function[1]		Structural modelling[2]	
	D_{max} (Å)	R_g (Å)	Discrepancy (χ^2)	NSD
LHn/A	116±5	34.6±1.0	0.96	0.90
LHn/B	118±6	34.2±0.1	1.20	0.81
LHn/D	120±1	34.6±0.7	0.97	0.70
scLHn/D	114±1	34.0±0.6	1.10	0.84

[1] The distance distribution function (*P(r)*) function from GNOM was used to calculate the maximum dimension (D_{max}) and the radius of gyration (R_g).
[2]The discrepancy (χ^2) between experimental and *ab initio* data was calculated by the program DAMMIF, DAMAVER was used to average 20 models and gave the above value for normalised spatial discrepancy (NSD).

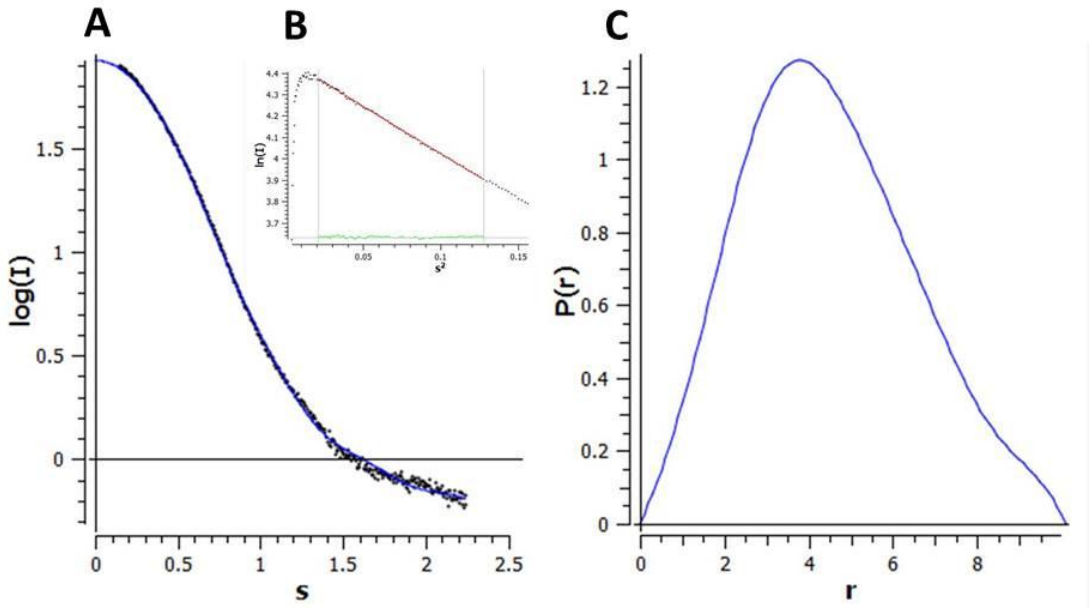

Figure 3.17. SAXS analysis of LHn/A at pH 4.0. (A) Scattering curve of LHn/A. (*I* is a relative unit, *s* is expressed in nm^{-1}) **(B)** The radius of gyration (R_g) was calculated using the Guinier approximation (Guinier, 1939) and was found to be 35.0 Å. **(C)** The interatomic distance function (*P(r)*) gave a maximum dimension D_{max} of approximately 110 Å (*x axis*, *r* in nm). These values were calculated in GNOM (Svergun, 1992).

Rigid body modelling with the coordinates from the crystal structures of LHn/A and /B was attempted. However analysis with CRYSOL (Svergun *et al.*, 1995) and trials with SASREF (Petoukhov and Svergun, 2005) did not produce realistic results. Manual docking allowed the drawing of a relatively convincing comparison between the *ab initio* bead models and the crystal structures (Figure 3.18). The rotational averaging of the molecules in solution limits the interpretation of the models. However, the distinctive shape of the two domains forming the LHn molecule can be clearly noticed. Indeed the atypical α-helices of Hn give the models an extended form over 110 Å long whereas LC's structure accounts for the globular central shape.

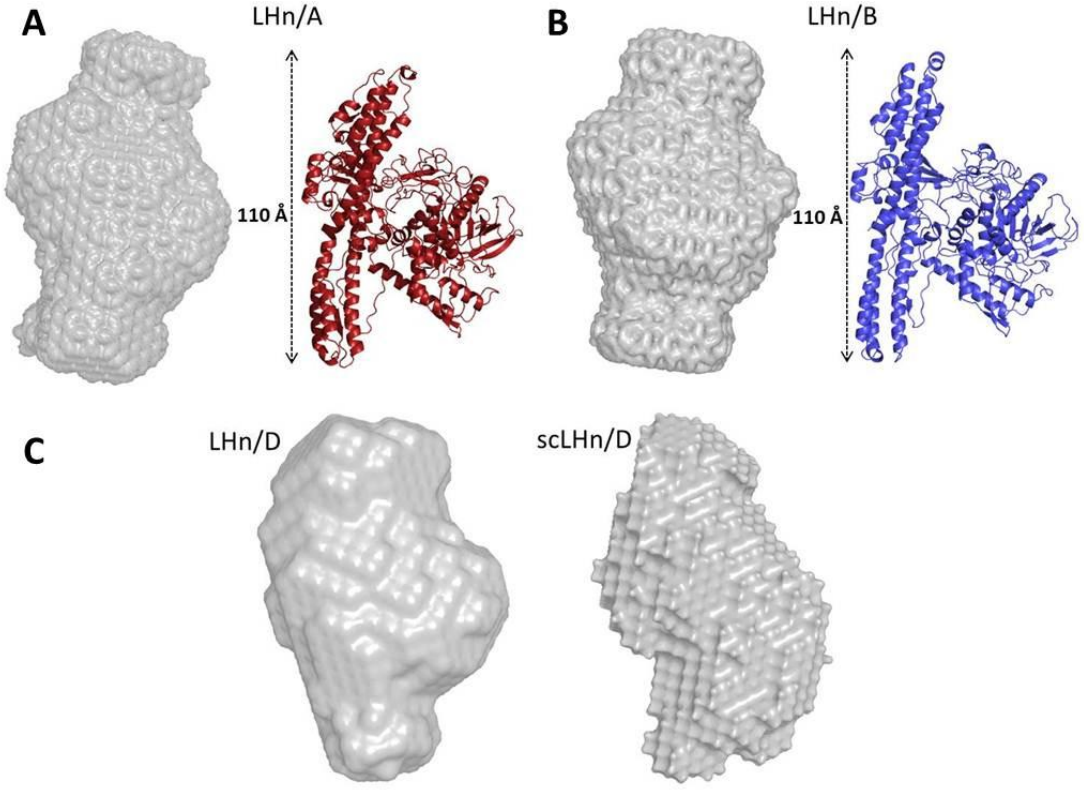

Figure 3.18. *Ab initio* solution scattering models of LHn/A, /B and /D at pH 5.0. (A) and **(B) Left**, SAXS model of LHn generated by DAMMIF (Franke and Svergun, 2009) and averaged in DAMAVER (Volkov and Svergun, 2003). **Right**, crystal structure of LHn/A (red) and /B (blue). **(C)** SAXS model of LHn/D (**left**) and scLHn/D (**right**).

Overall, the LHn molecules tested seem to conserve their overall structural features in solution. The change of pH did not appear to have any dramatic influence on the molecules. Although the resolution of the data may not allow to visualise small structural details, the information collected for LHn/A and /B showed that the fragments kept their structural integrity from neutral to acidic pH. This may be expected as these fragments have low theoretical pI, just below the endosomal pH (5.0-6.0) where BoNTs are active.

The first structural information obtained from LHn/D (in both forms, i.e. single chain and activated) seemed to indicate a conserved LHn fold. The scattering curves superposed well with LHn/A and /B, and the similarity is illustrated by comparable *ab initio* bead models (Figure 3.18C). Additional processing of the data by collaborators at EMBL Hamburg indicated a significant difference with the data collected for LHn/D at pH 7.0, compared to all the other curves. Further experiments would be necessary to characterise this variation. In particular, refinement of the experimental conditions and a narrower pH range may help determine any modifications in the LHn structures.

3.5. Discussion

Structure of LHn

Botulinum neurotoxins are complex molecules that present a deadly machinery able to specifically inhibit synaptic transmission of cholinergic motor neurons. Each of the toxin's domains is essential for optimal potency and understanding their mechanism will allow to better harness this activity towards novel applications. Fragments of the toxin (LHn) comprising the catalytic and translocation domains of BoNT have been described as potential functional molecules from which could be designed a range of novel biological products aiming at targeted cell secretion inhibition (Foster and Chaddock, 2010).

The production of recombinant engineered LHn fragments was achieved following the method described by Chaddock *et al.* (2002). As a first step towards the production of biological molecules, it is important to obtain reliable and consistent yields of active material. Engineering of the LHn fragment for optimal expression in *Escherichia coli* was achieved for several serotypes. Importantly it allowed the insertion of exoprotease recognition (i.e. cleavage) sites in order to mimic the post-translational "activation" of BoNT into active di-chain toxins. Additionally, the molecules could be designed to include for peptidic tags (poly-histidine) which are utilised in classical affinity chromatography techniques and thus facilitate the purification process. The aim of our study was to conduct biochemical and structural characterisation of several LHn constructs and hence required large quantities of highly pure material. Optimisation of the purification process by affinity and hydrophobic chromatography led to the production of LHn/A, /B and /D to standards required for protein crystallography.

Determination of the crystal structures of LHn from serotypes A and B demonstrated the stability of this BoNT fragment. In both cases, the structures were similar to their corresponding parent neurotoxin lacking the binding domain (Hc). The catalytic domain (LC), a zinc metalloprotease with a globular structure composed of a mixture of α-helices and β-sheets, was particularly conserved with only minor differences observed in flexible loop regions. These loops are located in solvent accessible surfaces of LC, a number of which have been linked to substrate binding (Breidenbach and Brünger, 2004). In all the structures solved, the zinc ion was present and tetrahedrally coordinated by the HEXXH motif. LHn/A and /B were submitted to Factor Xa treatment for activation into a di-chain form. As a result, LC is linked to Hn by a single disulphide bridge that was clearly visible. Furthermore, the corresponding cleavage sites were confirmed by the lack of electron density at the expected sites.

The relationship between the two domains forming LHn is essential for BoNT's activity and relies on a complex interaction, illustrated by the atypical "belt" region. This segment of Hn surrounds LC and by doing so prevents access to the catalytic site (Brünger *et al.,* 2007) and potentially promotes membrane insertion (Galloux *et al.,* 2008) in BoNT/A. Similar roles in other serotypes have not been demonstrated yet. Although the belt region partially occludes the active site in BoNT/B (LHn/B) and /E, the catalytic zinc is more accessible compared to BoNT A. Elucidation of the structures from other serotypes may help understand the function of this segment and why it represents the region with highest sequence variation (Lacy and Stevens, 1999).

The translocation domain is constituted of long α-helices believed to be responsible for pore-formation and allowing transport of LC within the cytosol (Fischer *et al.,* 2008b). Surprisingly, the structures of Hn in LHn (/A and /B) are identical to the ones observed in the full length holotoxins. Removal of the BoNT binding domain left Hn mostly unperturbed despite presenting a new and consequent solvent accessible surface. Only small unstructured loop regions connecting these helices, and normally in direct contact with Hc, showed some changes and were not visible in the crystal structures. Strong inter-helical bonds hold this unusual fold together and make LHn an ideal framework on which to build novel molecules. The LHn C-termini would likely accommodate polypeptide partners without damaging the fragment's structure.

Trials with LHn/D led to the crystallisation of a truncated form with only the light chain. Crystallisation parameters were different from the previously published LC/D structure (Arndt *et al.,* 2006) and noticeable differences were observed. The overall globular LC fold was conserved but details in several flexible loop regions that may be involved in substrate binding were noticed. Biochemical characterisation of several LHn/D constructs confirmed that full-length LHn/D was purified and represented the main component in all the samples. Truncation of LC/D is unlikely to be caused by unspecific cleavage from the exoprotease used for activation. Instead it is believed that crystallisation conditions, for example the high protein concentration and prolonged incubation period, would have favoured autoproteolysis. Such autoproteolysis phenomenon has been observed in several BoNT serotypes (DasGupta *et al.,* 2005). A SNARE-like motif was identified in the primary sequence of LC/D and could provide additional explanation for cleavage of the molecule. This occurrence may be prevented by crystallisation trials in the presence of a chelating agent or by key mutations, thus enabling the presence of the full length LHn/D in the crystal, although this needs to be tested with further experiments.

The first observations based on small angle x-ray scattering experiments on BoNT molecules demonstrated the significance of the LHn crystallographic structures in a more biologically relevant environment. The data collected for LHn/A, /B and /D allowed for the determination of low resolution models. While direct rigid body modelling could not be performed, a convincing shape could still be obtained for each molecule. As observed in the LHn crystal structures, the SAXS models produced a consistent size and shape, in a neutral to acidic pH range. The individual LHn domains could be placed in the SAXS model without any ambiguity. More particularly, the longer axis fitted conclusively with Hn, and the length of the translocation domain's helices corresponded to the experimental dimension measured by Guinier approximation (approximately 100-115 Å). The central globular shape matched with the LC fold seen in crystal structures. These structural features were also observed with LHn/D. For the first time, experimental evidence offered a glimpse of the structure of the Hn/D domain and indicated a conserved long α-helical arrangement. The translocation domain, excluding the belt region, shows high sequence conservation. Organisation of the conserved residues across Hn was predicted to maintain the inter-helical interactions holding the domain together (Lacy and Stevens, 1999). In addition, the SAXS data confirmed the presence of the full length LHn/D in the sample that generated crystals of a truncated LC molecule.

Activity of LHn

The botulinum neurotoxins cause inhibition of neurotransmission in cholinergic motor neurons by cleaving one of the SNARE proteins. In order to utilise the LHn molecules as a biologically active material, they need to retain this critical catalytic property. Chaddock *et al*. (2002) demonstrated LHn/A had a similar enzymatic activity to BoNT/A. *In vitro* assays performed on a recombinant SNAP25-GFP substrate over time confirmed that the crystallised molecule was also catalytically active, and stable in solution for up to 36 days. Substrate cleavage assays were performed with LHn/B and BoNT/B using recombinant GFP-tagged VAMP-1, -2 and -3. LHn/B was able to cleave all of the three substrates tested although some differences in maximal velocity (*Vmax*) were noticed. Similarly to the full length toxin, LHn/B presented a better affinity for VAMP-1 and -2 over -3. Minor differences in the primary structures of these homologous substrates may explain this result and provide further information on the BoNT/B-VAMP interaction (Chen *et al.,* 2008).

Toxicity of the botulinum neurotoxins rely on these molecules reaching their intracellular SNARE substrates. BoNT associates with the vesicular membrane via its receptor binding domain and endocytosis. Once inside the vesicle, a conformational change is believed to allow membrane pore-formation by Hn and transport of LC within the cytosol.

Despite lacking the receptor binding domain, LHn/A was shown to cleave intracellular (embryonic spinal cord neurons) SNAP25 and prevent neurotransmission (Chaddock *et al.* 2002). The study described here demonstrated LHn/B was able to cleave intracellular VAMP-1 and -2. The mechanism for LHn entry into neuronal cells is unknown but likely depends on Hn's inherent pore-forming capabilities. Using single channel current measurements, Fischer *et al.* (2008) proved that LHn/A was sufficient to transport LC into the cytosol of target cells through channel formation. Interactions between LHn and the neuronal cells are likely to be unspecific considering the conditions of the *in vitro* assays described above, i.e. high cell density and LHn concentration, and it should be highlighted that LHn's activity is greatly reduced (by a factor over 100,000) compared to BoNT. Nevertheless LHn appears as a fully functional molecule capable of LC translocation into neuronal cells where it can cleave cytosolic SNARE proteins. Further study would be necessary to determine whether LHn can act on other non-neuronal cell types and thus determine its utility as a pharmacological tool for the study of cellular SNARE-mediated processes.

Conclusion

LHn consists of the catalytic (LC) and translocation (Hn) domains of the botulinum neurotoxin but is devoid of the receptor binding domain. Crystallographic studies of serotypes A and B have established that the fragment is a structurally stable scaffold, with Hn constituted of long α-helices and LC, a zinc protease presenting a globular fold. Preliminary experimental evidence suggests that this global structure is also conserved in a solution environment and extend to serotype D.

LHn's function was also verified *in vitro* by biochemical characterisation of recombinant SNARE substrate cleavage. Additionally, LHn maintains an ability to cleave intracellular SNARE proteins in neuronal cells, thereby showing its innate ability to transport the catalytic light chain within the cytosol of target cells.

The recombinant LHn fragments described are soluble proteins with attenuated toxicity but retaining the capabilities of the botulinum neurotoxins. These fragments are proving to be significant tools to study the mechanisms responsible for BoNT potent toxicity (Fischer *et al.,* 2008b). Finally they represent a fantastic framework on which to build novel engineered molecules with added therapeutic applications (Chaddock and Acharya, 2011).

Chapter 4. Engineering self-activating *Clostridium botulinum* neurotoxin fragments using SNARE-peptides

4.1. Introduction

BoNTs have been a point of focus for protein engineering since it was demonstrated that its functionality as secretion inhibitor could be extended to non-neuronal targets by engineering a new cell targeting domain onto the LHn fragment of BoNT. The proof-of-concept was first provided with the use of lectin and growth factor conjugates (Chaddock *et al.,* 2000a, 2004; Duggan *et al.,* 2002). Subsequently, novel recombinant BoNT derivatives consisting of the catalytic LC and the Hn translocation domain, together forming the LHn fragment (Figure 4.1), coupled with retargeting polypeptides have been developed as a novel biologics platform - Targeted Secretion Inhibitors (TSI).

An essential step in the production of fully active BoNT is the post-translational cleavage from a single chain polypeptide into its di-chain activated form (Wey *et al.,* 2006). Treatment of BoNT with a non-specific protease such as trypsin can result in undesirable digestion and a mixture of heterogeneous proteins (Shone *et al.,* 1985). Recombinant derivatives of BoNT can be engineered to incorporate specific recognition sites for exoproteases such as Factor Xa or enterokinase (Sutton *et al.*, 2005). This process allows for a more controlled approach of activation. This is however a delicate and limiting process in terms of protein production.

Thus the use of the BoNT LC proteolytic activity towards SNARE peptides was investigated as a tool for engineering both self-activating and LC-activating molecules. This resulted in several novel derivatives in which an optimum SNARE protein region was inserted in the LHn/A backbone. More precisely, a SNAP25 peptide was introduced between the carboxyl terminal cysteine of LC/A and the amino terminal cysteine of Hn/A. This substitution of the native activation site created a backbone with domains denoted LC/A-SNAP25-Hn/A. LC/A-SNAP25-Hn/A is a novel, engineered backbone susceptible to activation by ‘self-cleavage’ with LC.

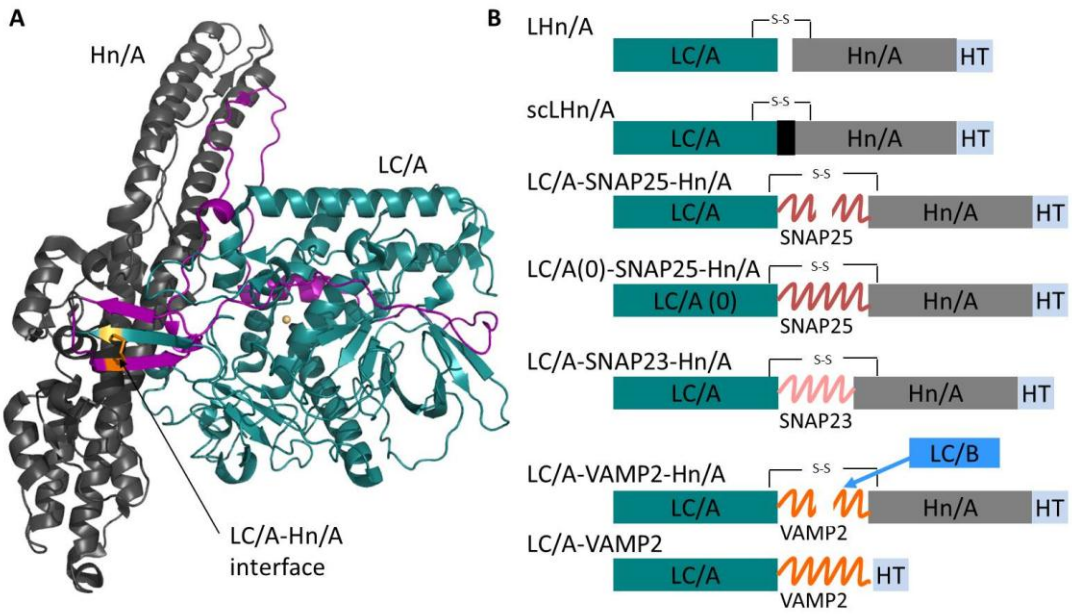

Figure 4.1. Representation of LHn/A domain structure and constructs engineered with SNARE. (A) Crystal structure of LHn/A based on PDB code 2W2D (Masuyer *et al.*, 2009). The LC/A domain is shown in cyan, Hn/A domain in grey, Hn/A belt region in magenta and the disulphide bridge in gold. **(B)**. Domain structure of LHn/A: LC/A (cyan), Hn/A (grey), His tag (HT, light blue), SNAP25 (red), SNAP23 (pink) and VAMP2 (orange). LC/B activating protease is shown in blue and disulphide bonds are represented (S-S).

To investigate the broad applicability of this concept to other LHn serotypes and beyond self-cleavage, a VAMP2 substrate peptide was substituted in place of SNAP25 to create LC/A-VAMP2-Hn/A. VAMP2 is not cleaved by LC/A but is a substrate for LC/B. A backbone in which LC/A is coupled to Hn/A with a VAMP2 peptide therefore remains as a single-chain until it is activated by LC/B, which behaves as the exogenous activating protease.

This study demonstrated that extended SNARE peptides corresponding to the carboxyl coiled coil region of SNAP25 and VAMP2 can be inserted between the LC/A and Hn/A of LHn/A and are effective substrates for activation of the LHn backbone. LC/A-SNAP25-Hn/A acted as a self-processing enzyme that retained the ability to assemble into a functional protein with translocation and proteolytic properties. LC/A-VAMP2-Hn/A could be activated by LC/B, under controlled conditions, to maintain the integrity of the molecule. This novel engineering strategy has potential for alternative manufacturing approaches for TSI proteins and for understanding their enzymatic properties.

The crystal structures for BoNT/A (Lacy *et al.*, 1998), -B (Swaminathan and Eswaramoorthy, 2000) and -E (Kumaran *et al.*, 2009) and more recently that of LHn/A and

LHn/B (Masuyer *et al.*, 2009, 2011) have been determined. In all the structures available, the botulinum neurotoxin domains conserve their singular fold (Swaminathan, 2011). In particular, LHn (A and B) retain the structural stability of the LC-Hn interactions seen in the holotoxin despite lacking the Hc binding domain. This highlighted the relevance of LHn as useful and safe models for functional and structural studies of the botulinum neurotoxins.

Molecular structures of the synaptosome-associated proteins (SNAP) have been reported as either functional heteromeric SNARE complexes (Sutton *et al.*, 1998, Fiebig *et al.*, 1999), self-assembling homotetrameric molecules (Freedman *et al.*, 2003), or in complex with LC/A (Breidenbach and Brünger, 2004). While the crystal structures demonstrated the SNARE proteins tendency to form a helix bundle mediated by the coiled-coil SNAP domains, they are expected to be mainly unstructured in solution (Margittai *et al.*, 2001). This is in accordance with the low activity of BoNT/A on the structurally stable SNARE complex compared to free SNAP25. The flexibility of SNAP25 is a requirement for substrate cleavage by LC/A (Breidenbach and Brünger, 2004).

The crystal structures of two engineered derivatives of the LHn backbones from serotype A were solved to 2.7-3.0 Å and are described in this study. They correspond to constructs used as single chain controls in the self-cleaving LC-SNARE-Hn molecules study namely LC/A-SNAP23-Hn/A and LC/A(0)-SNAP25-Hn/A. SNAP23 is a non-cleavable SNAP25 homologue. LC/A(0) was mutated at key residues to impair LC/A proteolytic activity. Although the engineered fragment could not be modelled due to disorder, the structures demonstrate the remarkable stability of the LHn fold despite the addition of significant peptide segments at the LC-Hn interface. This strengthens the relevance of LHn as safe and useful molecules for the study of botulinum neurotoxin engineering.

4.2. Materials and methods

Plasmid construction, Protein Expression, Purification and Detection

Synthetic DNA, codon optimised for *E. coli* expression, encoding human SNAP25 (145-206), human SNAP23 (151-211) and human VAMP2 (24-94) was obtained from Entelechon GmbH. SNARE encoding DNA was cloned into pET26b expression vector (Novagen) in which the pelB signal leader is removed. These plasmids contained open reading frames (ORF) for LHn/A encoding 842 amino acids, or for LC/A encoding 430 amino acid and LC/B encoding 437 amino acids. In frame fusion was achieved by use of unique restriction sites introduced into the DNA encoding the C terminus of the LC, N terminus of Hn and the SNARE peptide linker. LC/A(0)-SNAP25-Hn/A contained mutations in the active site at positions E224Q and H227Y as well as GS linkers flanking the SNAP25 fragment. Poly-GS linkers correspond to GGGGS repeats. The clones were provided by Syntaxin Ltd (Stancombe *et al.*, 2011).

Plasmids encoding open reading frames were expressed in *E. coli* (BL21 (DE3), Novagen) and purified as previously described (Masuyer *et al.*, 2009). Briefly, expression cultures were grown to mid log growth phase in enriched terrific broth, the temperature reduced from 37°C to 16°C and protein expression induced with 1 mM IPTG. Cells were harvested after 18h, and stored at -80°C until further use.

Homogenised cell lysate, clarified by centrifugation, was loaded onto a Ni^{2+}-charged chelating sepharose column (GE Healthcare). Proteins eluting at either 100 or 200 mM imidazole in 50 mM HEPES (pH 7.2), 0.2 M NaCl buffer were dialysed overnight at 4°C against the same buffer to remove salt. Protein sample volume was reduced using Vivaspin 50000 MWCO concentrators (Millipore) and stored at -20°C. Protein concentration was determined by A_{280} measurement using a NanoDrop 2000 instrument (Thermo Scientific).

Rabbit polyclonal antibodies specific to LC or Hn, obtained from Eurogentec, were evaluated by cross reaction with a panel of recombinant A, C, D and F serotype LHn proteins and the *E. coli* lysate used for their expression. Antisera raised against either LC/A protein or Hn/A peptide (CQLSKYVDNQRLLSTL), displayed high specificity to antigen and low reactivity to other serotypes and to *E. coli* proteins.

Table 4.1. Summary of domain structure and function in SNARE-LHN constructs.

Construct	Serotype	Domain linker	Activation status	Molecule function
LHn/A	A	Ek[1]	Activated	Non-SNARE LHn
scLHn/A	A	Ek[1]	Non-activated	Non-SNARE LHn
LC/A-SNAP25-Hn/A	A	Human SNAP25 (145-206)[2]	Activated	Self-activating
LC/A(0)-SNAP25-Hn/A	A	Human SNAP25 (145-206)[2]	Non-activated	Endopeptidase inactive control
LC/A-SNAP23-Hn/A	A	Human SNAP23 (151-211)[2]	Non-activated	Non-substrate homologue
LC/A-VAMP2-Hn/A	A	Human VAMP2 (24-94)[2]	Activated	Activated by LC/B
LC/A-VAMP2	A	Human VAMP2 (24-94)[2]	n/a	Non-translocating control
LC/B-SNAP25	B	Human SNAP25 (145-206)[2]	n/a	Activation protease

[1] Engineered enterokinase site for activation.
[2] Numbers correspond to region of SNARE used as a linker.

Recombinant SNAP25 cleavage

The catalytic activity of the endopeptidase component of engineered LHn backbones was assessed using a recombinant SNAP25 assay. Cleavage of recombinant SNAP25 (2-206) substrate with a carboxyl-terminal GFP was assayed in triplicate in three experiments for each molecule (Masuyer *et al.*, 2009). Briefly, test enzymes were diluted to 1.0 mg/ml in buffer containing 50 mM HEPES pH 7.2, 20 μM $ZnCl_2$, 1 mg/ml BSA, 10 mM DTT, and pre-incubated at 37°C for 30 minutes. SNAP25-GFP was diluted to 8 μM and incubated with the diluted test enzyme in the range from 1 mg/ml to 10 ng/ml at 37°C for 1 hour. In the case of the competition experiment, the procedure included pre-incubation of the test enzyme with 30 μM LC/B-SNAP25 at 37°C for 30 minutes, before addition of substrate. SNAP25 cleavage was measured by comparison of cleaved product to un-cleaved substrate and percentage cleavage calculated by densitometry of coomassie stained SDS PAGE using GeneSys software (SynGene Bioimaging). Data was normalised to account for signal intensity from molecular weight differences between cleaved and un-cleaved substrate. Calculations and curve fitting were done with Prism5 (GraphPad Software).

In vitro spinal cord neuron SNAP25 cleavage

The functionality of engineered LHn backbones was also assessed in embryonic spinal cord neurons (eSCNs). eSCNs prepared from embryos of 14–15 day old foetal Sprague

Dawley rats were cultured for 21 days as described previously (Duggan *et al.*, 2002). Briefly, cell preparations were treated in triplicate with serial dilutions (range 3 μM to 3 nM) of LHn/A and LC/A (Syntaxin Ltd) proteins and incubated at 37°C with 10% CO_2 for 24h. Cells were lysed by removing media and adding 25% NuPAGE buffer in 10 mM DTT. Samples were heated at 95°C for 5 minutes and resolved by SDS-PAGE on 12% BIS-TRIS NuPAGE gels (Invitrogen) before transfer to nitrocellulose membrane (Millipore). SNAP25 cleavage was measured by comparison of cleaved product to un-cleaved substrate and percentage cleavage calculated by densitometry of enhanced chemiluminescent (ECL) signal from specifically immunoreactive primary anti-SNAP25 antibody (Sigma catalog. No. S9684). Each assay was performed three times. Data were normalised to account for signal intensity from molecular weight differences between cleaved and un-cleaved substrate. Calculations and curve fitting were done with Prism5 (GraphPad Software).

Crystallisation of LC/A-SNAP23-Hn/A and LC/A(0)-SNAP25-Hn/A

Further purification was necessary for crystallisation trials performed on two of the constructs previously prepared. Briefly, LC/A-SNAP23-Hn/A was further purified by hydrophobic interaction chromatography (Phenyl Sepharose HP, GE Healthcare). LC/A(0)-SNAP25-Hn/A was lastly purified by gel filtration (Superdex 75, GE Healthcare). Final samples were concentrated to 6.1 mg/ml and 5.1 mg/ml respectively.

Initial crystallisation condition for LC/A-SNAP23-Hn/A was identified using the Stura Footprint Screen + MacroSol (Molecular Dimensions) corresponding to 0.1 M imidazole malate pH 7.0, 15% PEG4000. Optimisation was carried out by the hanging-drop vapour diffusion method at 16°C, with 3 μl drops (2 μl protein and 1 μl mother liquor against a 500 μl reservoir). LC/A-SNAP23-Hn/A crystals were grown in 2-3 months in 0.1 M imidazole malate pH 6.0, 15% PEG3350. X-ray diffraction data to 2.95 Å from a single crystal cryo-protected with 25% PEG3350 were collected at the Diamond Light Source, UK, beamline IO3. Data were processed and scaled in monoclinic space group $P2_1$ using MOSFLM and SCALA (Leslie, 2006; CCP4, 1994) (Table 4.3). Initial phases were obtained by molecular replacement using Phaser (McCoy *et al.*, 2007) with the coordinates of LHn/A (PDB 2W2D, Masuyer *et al.*, 2009).

Initial crystallisation condition for LC/A(0)-SNAP25-Hn/A was identified using Heavy + Light Twin Pack (Molecular Dimensions) corresponding to 0.1 M TRIS pH 8.5, 21% PEG4000. Optimisation was performed by the hanging-drop vapour diffusion method (2 μl protein and 1 μl mother liquor against a 500 μl reservoir) at 16°C. LC/A(0)-SNAP25-Hn/A crystals were grown in 2-3 months in 0.1 M TRIS pH 7.5, 15% PEG3350, 0.2 M

lithium sulphate. X-ray diffraction data to 2.7 Å from a single crystal cryo-protected with 2 M lithium sulphate were collected at the Diamond Light Source, UK, beamline IO2. Data were processed and scaled in orthorhombic space group $P2_12_12_1$ using MOSFLM and SCALA (Leslie, 2006; CCP4, 1994; Table 4.3). Initial phases were obtained by molecular replacement using Phaser (McCoy *et al.*, 2007) with the coordinates of LHn/A.

The atomic coordinates and the structure factors have been deposited with the RCSB Protein Data Bank under accession codes 3ZUS, 3ZUR for LC/A-SNAP23-HA, LC/A(0)-SNAP25-HA respectively.

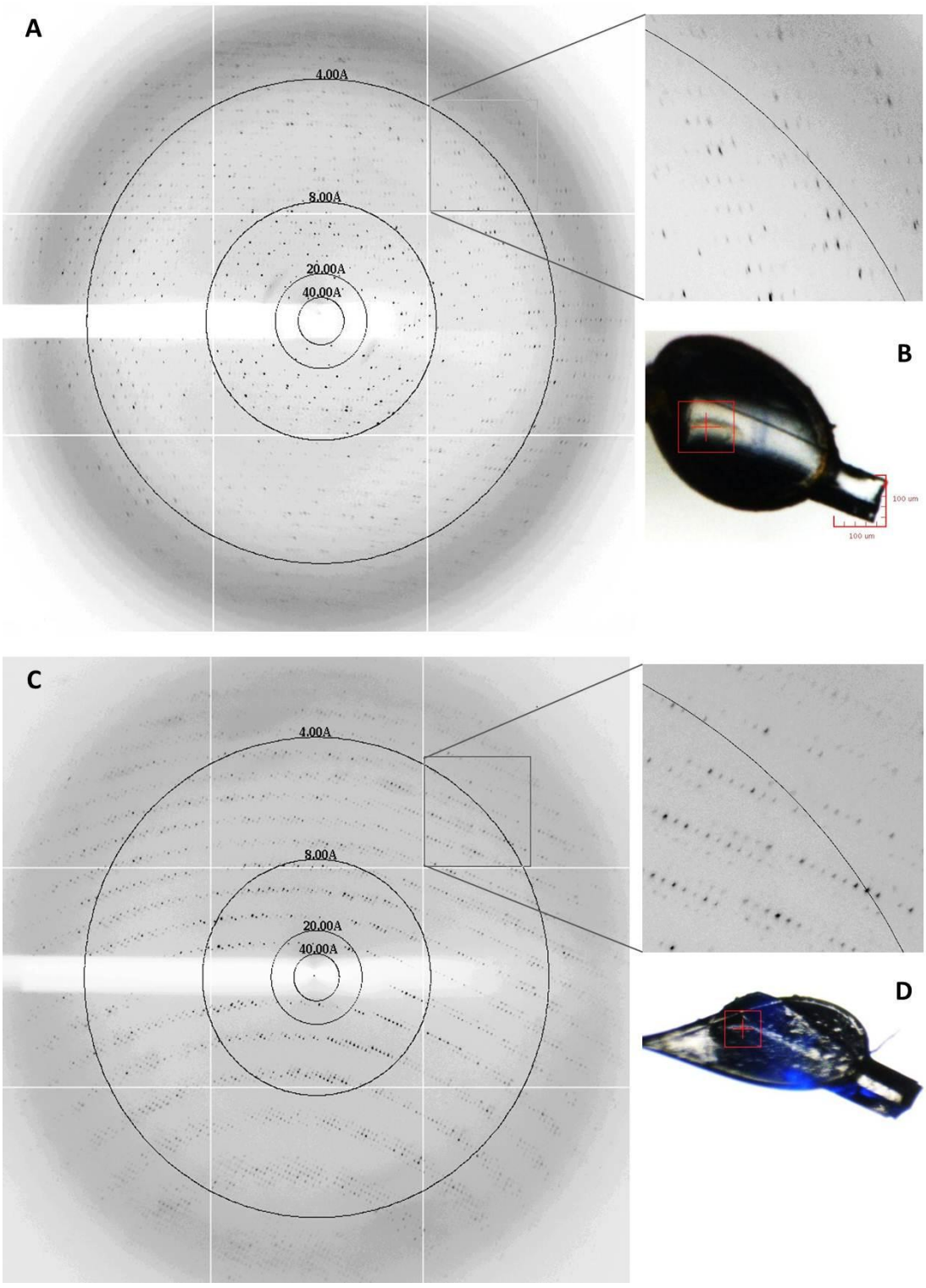

Figure 4.2. Crystals of LC/A-SNAP23-Hn/A and LC/A(0)-SNAP25-Hn/A. (A) X-ray diffraction of LC/A-SNAP23-Hn/A crystal at IO3 (Diamond Light Source, UK). **(B)** LC/A-SNAP23-Hn/A crystal mounted for data collection (red square is 100x100 μm). **(C)** X-ray diffraction of LC/A(0)-SNAP25-Hn/A crystal at IO2 (Diamond Light Source, UK). **(D)** LC/A(0)-SNAP25-Hn/A crystal mounted for data collection (red square is 100x100 μm).

4.3. Results

4.3.1. Production of engineered LHn-SNARE molecules

Use of SNARE peptide sequences as activation linkers for LHn backbones

Multiple species alignments of sequences for R-SNARE and Q-SNARE proteins compiled in FunShift (Abhiman and Sonnhammer, 2005) were used to identify the extent of homology within the soluble carboxyl region for each subfamily (Stancombe *et al.,* 2011). The cleavage kinetics of different length SNARE peptides as substrates for LC was also considered (Chen and Barbieri, 2006; Shone *et al.,* 1993). The minimum SNARE segment selected was based on retaining full cleavage properties of the SNARE peptide relative to the full length protein and included at least one, nine residue recognition SNARE motif, proximal to the cleavage site (Washbourne *et al.*, 1997). 62 amino acid SNAP25 (145-206), 61 amino acid SNAP23 (151-211) and 67 amino acid VAMP2 (28-94) peptides were synthesised based on these criteria. Figure 4.1 and Table 4.1 illustrate the domain structure of the LHn/A backbones used in this study.

Purification of self-cleaving LHn backbones

Lysate from LC/A-SNAP25-Hn/A *E. coli* expression media showed prominent over-expressed proteins that migrated at approximately 50 and 56 kDa when analysed by reducing SDS PAGE. These are the expected sizes of Hn/A and LC/A-SNAP25 protein domains respectively. Western blot analysis using anti-LC/A and anti-His antibodies confirmed these molecules were cleavage products derived from the 106 kDa LC/A-SNAP25-Hn/A protein (with a C-terminal His-tag) (Figure 4.3.H). It can therefore be deduced that LC/A-SNAP25-Hn/A had been cleaved within the host cell or during cell lysis.

Purification of intact LC/A-SNAP25-Hn/A by affinity chromatography using Hn/A's His tag was demonstrated by Western blot with the detection of a protein that migrated at 106 kDa and cross reacted with anti-LC/A, anti-Hn/A and anti-His antibodies (Figure 4.3.C). Reduction of LC/A-SNAP25-Hn/A by the addition of 10 mM DTT fully dissociated the LHn backbone, indicating that all single-chain polypeptide had been cleaved. This also indicated that the assembled LHn backbone was covalently associated by disulphide bonding (Figure 4.3.C). Western blotting confirmed the presence of 56 kDa LC/A-SNAP25 and 50 kDa Hn/A monomers.

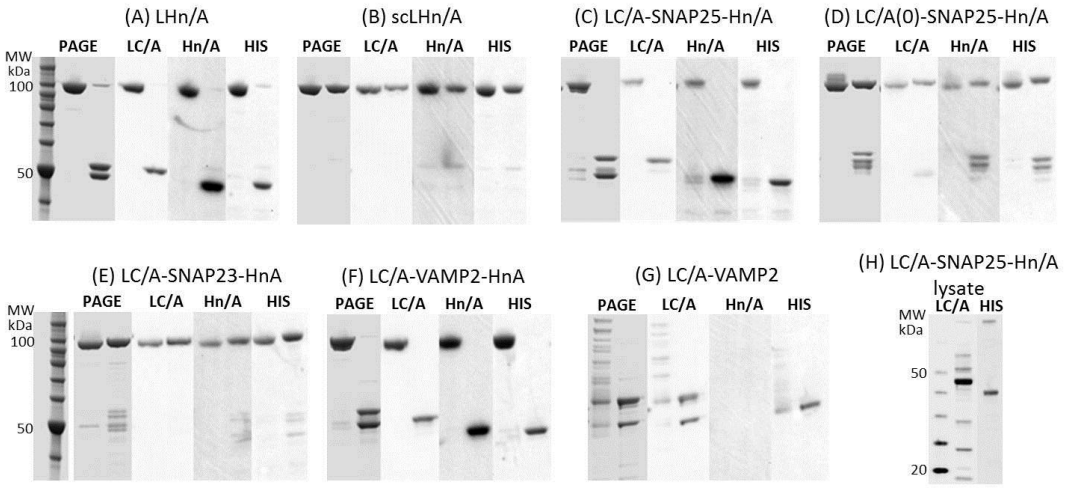

Figure 4.3. SDS PAGE and western blot of purified LHn/A proteins and host cell lysate. SDS PAGE and western blot of purified LHn/A proteins (**A-G**) and host cell lysate (**H**). Comparison of oxidised and reduced purified proteins (**left** and **right** lanes respectively, **A-G**) and lysate (oxidised only, **H**) by coomassie staining (PAGE), anti-LC/A (LC/A), anti-Hn/A (Hn/A) and anti-His (HIS) antibodies.

LHn backbone dissociation is dependent upon LC endopeptidase

The assembled LC/A-SNAP25-Hn/A backbone was reduced and analysed by N-terminal sequencing. The reduced proteins were found to correspond to LC/A and SNAP25. Cleavage had occurred between amino acids Q197 and R198 in the SNAP25 peptide; precisely the expected proteolytic site for BoNT/A endopeptidase. Non-specific cleavage products were not detected.

To control for non LC/A mediated cleavage of the SNAP25 peptide, enzymatically inactive LC/A(0)-SNAP25-Hn/A and non-substrate SNAP23, LC/A-SNAP23-HN/A, LHn backbones were created. The LC of LC/A(0)-SNAP25-HN/A contained mutations in the active site at positions E224Q (Kukreja *et al.*, 2007) and H227Y (Duggan *et al.*, 2002). LC/A-SNAP23-Hn/A contained the non-neuronal homolog of human SNAP25, SNAP23, which is not a substrate for LC/A. LC/A(0)-SNAP25-Hn/A and LC/A-SNAP23-Hn/A (Figures 4.3.D and E respectively) did not display the same pattern of dissociation when reduced to that of either LC/A-SNAP25-Hn/A or LC/B treated LC/A-VAMP2-Hn/A (Figures 4.3.C and 4.3.F respectively). Lack of specific dissociation of LC/A(0)-SNAP25-Hn/A and LC/A-SNAP23-Hn/A was linked to a reduced level of *in vitro* SNAP25 cleavage and was indicative of a non-activated state (Figure 4.4).

In order to investigate the function of LC as an exogenous activating protease, serotype A LHn backbone was constructed with a non-substrate SNARE linker. LC/A-

VAMP2-Hn/A remained as a single-chain polypeptide when expressed and purified. This was to be expected as VAMP2 is the substrate for LC/B and SNAP25 is substrate for LC/A. This also confirmed the specificity of LC and SNARE pairing required to achieve activation.

A methodology was developed to capture intact, serotype homogenous di-chain molecules. Exogenous LC protease activation was achieved by immobilisation of oxidised, single-chain, LC/A-VAMP2-Hn/A on an affinity matrix followed by a washing step with LC/B-SNAP25. This treatment was sufficient to completely cleave the VAMP2 linker in the target LC/A-VAMP2-Hn/A protein. LC/B protease was not retained on the affinity column, thus allowing for the recovery of homogenous, di-chain LC/A-VAMP2-Hn/A (Figure 4.3.F).

4.3.2. Activity of engineered LHn-SNARE molecules

LHn backbones cleave recombinant SNAP25

The enzymatic activities of LHn backbone proteins were measured in cell free cleavage assays. Characterisation of the molecules was assessed by cleavage of a recombinant SNAP25-GFP substrate. The assay was performed under reducing conditions to allow release of LC from Hn in activated proteins.

LHn backbones clustered into two groups based on the level of substrate cleavage activity detected. The more active group consisted of activated LHn/A and LC/A-VAMP2-Hn/A as well as LC/A-VAMP2. A less active group was represented by scLHn/A, self-activated LC/A-SNAP25-Hn/A and non-activated LC/A-SNAP23-Hn/A. The effectiveness of each LHn backbone was established by plotting concentration against percentage cleavage (Figure 4.4) and calculating half maximal effective concentrations (EC50) for cleaved SNAP25-GFP (Table 4.2). Compared to activated LHn/A (0.01 μM), observed reductions in enzymatic activity were 60 fold for self-activated LC/A-SNAP25-Hn/A (0.46 μM), 30-fold for non-activated LC/A-SNAP23-Hn/A (0.2 μM), and 70 fold for scLHn/A (0.51 μM). Enzymatically inactive LC/A(0)-SNAP25-Hn/A was greater than 1000-fold reduced in proteolytic activity in comparison to activated LHn/A.

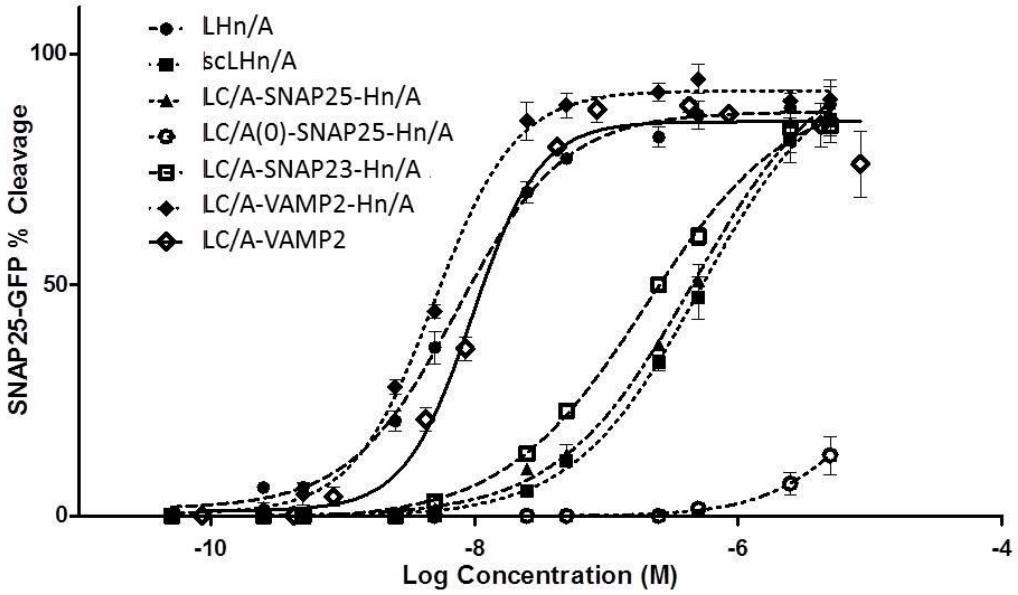

Figure 4.4. *In vitro* SNAP25-GFP assay. Comparison of SNARE-LHn molecules in a recombinant SNAP25-GFP cleavage assay. Reduced purified protein was assayed for activated LHn/A (closed circle/dashed line), scLHn/A (closed square/dotted line), self-activated LC/A-SNAP25-Hn/A (closed triangle/dashed line), non-activated LC/A(0)-SNAP25-Hn/A (open circle/dashed line), activated LC/A-VAMP2-Hn/A (closed diamond/dotted line), LC/A-VAMP2 (open diamond/solid line) and non-activated LC/A-SNAP23-Hn/A (open square/dashed line).

The lower activity of self-activated LC/A-SNAP25-Hn/A compared to other forms of activated LHn/A was further investigated. Possible competition between the SNAP25 linker and the SNAP25-GFP assay substrate for LC was considered. Comparison of activated LHn/A and self-activated LC/A-SNAP25-Hn/A activity in the presence and absence of a SNAP25 competitor (LC/B-SNAP25) is shown in figure 4.5. SNAP25 spiked LHn/A (0.15 μM) and LC/A-SNAP25-Hn/A (33 μM) showed a 5 and 9 fold increase in EC50 concentration respectively, compared to un-spiked LHn/A (0.01 μM) and LC/A-SNAP25-Hn/A (0.46 μM) (Table 4.2). Performing the substrate cleavage assay with increasing dose of LC/B-SNAP25 resulted in a dose-dependent loss of potency (data not shown). This indicated that SNAP25 peptide present within LC/A-SNAP25-Hn/A is likely to be competing with SNAP25-GFP assay substrate for enzyme catalysis. This might also explain the reduced activity of self-activated LC/A-SNAP25-Hn/A in this assay.

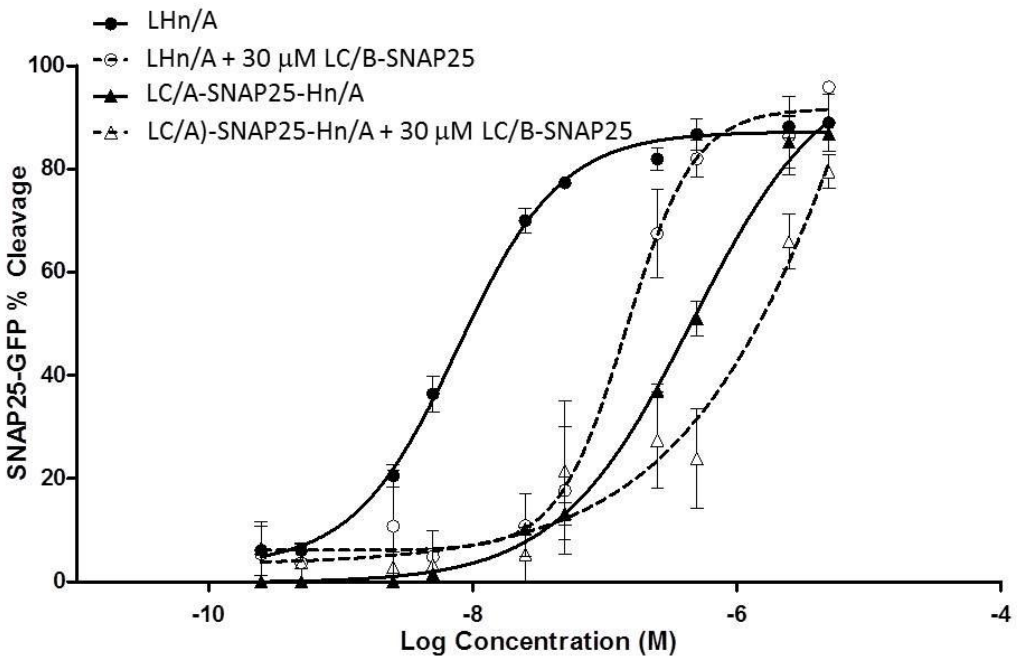

Figure 4.5. *In vitro* SNAP25-GFP competition assay. Comparison of LHn/A and self-activated LC/A-SNAP25-Hn/A in presence and absence of LC/B-SNAP25. Reduced purified protein was assayed for LHn/A only (closed circle/solid line), LHn/A with LC/B-SNAP25 (open circle/dashed line), LC/A-SNAP25-Hn/A only (closed triangle/solid line) and LC/A-SNAP25-Hn/A with LC/B-SNAP25 (open triangle/dashed line).

Table 4.2. Half maximal effective concentration (EC50) values (μM) for *in vitro* SNAP25 cleavage.

Construct	SNAP25-GFP assay[1]	SNAP25-GFP competition assay[1]	eSCN SNAP25 assay[1]
LHn/A	0.01	0.01	0.6
LHn/A + LC/B-SNAP25	-	0.15	-
scLHn/A	0.51	-	9.6
LC/A-SNAP25-Hn/A	0.46	0.46	0.5
LC/A-SNAP25-Hn/A + LC/B-SNAP25	-	33	-
LC/A(0)-SNAP25-Hn/A	>10	-	0
LC/A-SNAP23-Hn/A	0.20	-	ND[2]
LC/A-VAMP2-Hn/A	0.01	-	ND[2]
LC/A-VAMP2	0.01	-	ND[2]
LC/A	-	-	54.4

[1] Values are the average of three experiments performed in triplicate. Data calculated using Prism5 (GraphPad).
[2] ND, not determined.

LHn backbones cleave eSCN SNAP25

The divergence in enzyme activity between the methods used to activate LHn backbones observed in the cell free assay was not reproduced in eSCN cells. Cleavage of intracellular SNAP25 by LHn backbones was independent of the protease used for activation. Half-maximal effective concentrations (EC50) for eSCN SNAP25 cleavage by each molecule are shown in table 4.2. Self-activated LC/A-SNAP25-Hn/A (0.5 μM) cleaved SNAP25 in eSCN cells at a rate equivalent to activated LHn/A (0.6 μM) (Figure 4.6). SNAP25 cleavage activity for single-chain, scLHn/A (9.6 μM) was 16 fold reduced compared to activated LHn/A. SNAP25 cleavage activity for LC/A (translocation negative control) (54 μM), was 90-fold reduced compared to activated LHn/A. Enzymatic inactive LC/A(0)-SNAP25-Hn/A had no detectable SNAP25 cleavage activity.

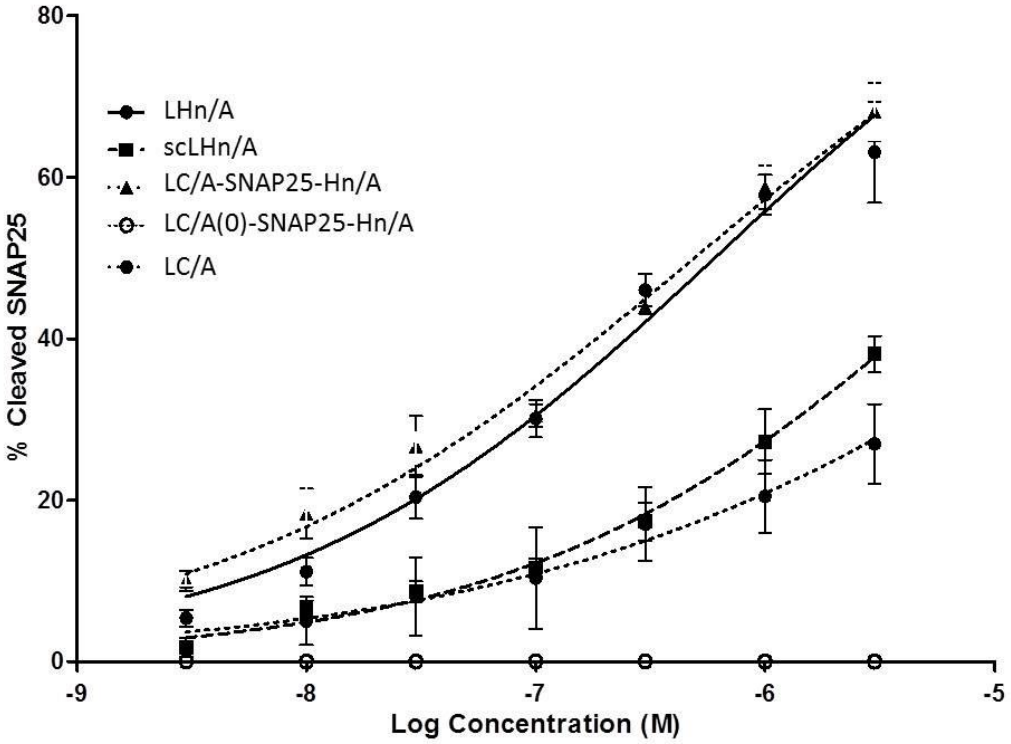

Figure 4.6. *In vitro* eSCN SNAP25 cleavage assay. Comparison of SNARE-LHn molecules in a cell based assay. Activated LHn/A (closed circle/solid line), scLHn/A (closed square/dashed line), self-activated LC/A-SNAP25-Hn/A (closed triangle/dotted line), non-activated LC/A(0)-SNAP25-HN/A (open circle) and LC/A (closed circle/dotted line).

4.3.3. Crystal structures of engineered LHn-SNARE molecules

Structures of LC/A-SNAP23-Hn/A and LC/A(0)-SNAP25-Hn/A

The crystal structure of LC/A-SNAP23-Hn/A was determined at 2.95Å resolution (Figure 4.7.A). A straightforward molecular replacement solution was found with Phaser (McCoy *et al.*, 2007) using LHn/A as a search model, with 4 molecules per asymmetric unit. While the whole LHn/A backbone could be fitted, no electron density was visible for SNAP23 at the LC-Hn interface, corresponding to a 63 residue long peptide due to disorder (Figure 4.8.A). The structure was refined to a final R_{free} = 29.3%, and R_{cryst} = 25.0% (Table 4.3), with 96.2% of amino acids in the Ramachandran plot favoured region, and 71 water molecules in the asymmetric unit.

LC/A(0)-SNAP25-Hn/A crystal structure was solved at 2.7Å resolution (Figure 4.7.B). Again, LHn/A was used as a search model in Phaser (McCoy *et al.*, 2007) and led to a molecular replacement solution, with 2 molecules per asymmetric unit. However, a 98-residues long region corresponding to the engineered SNAP25 peptide flanked by GS linkers between LC and Hn, could not be located due to disorder (Figure 4.8.B). The structure was refined to a final R_{free} = 26.6%, and R_{cryst} = 21.5% (Table 4.3), with 94.8% of amino acids in the Ramachandran plot favoured region. Solvent molecules (154 water molecules) and 5 sulphate ions were included in the model.

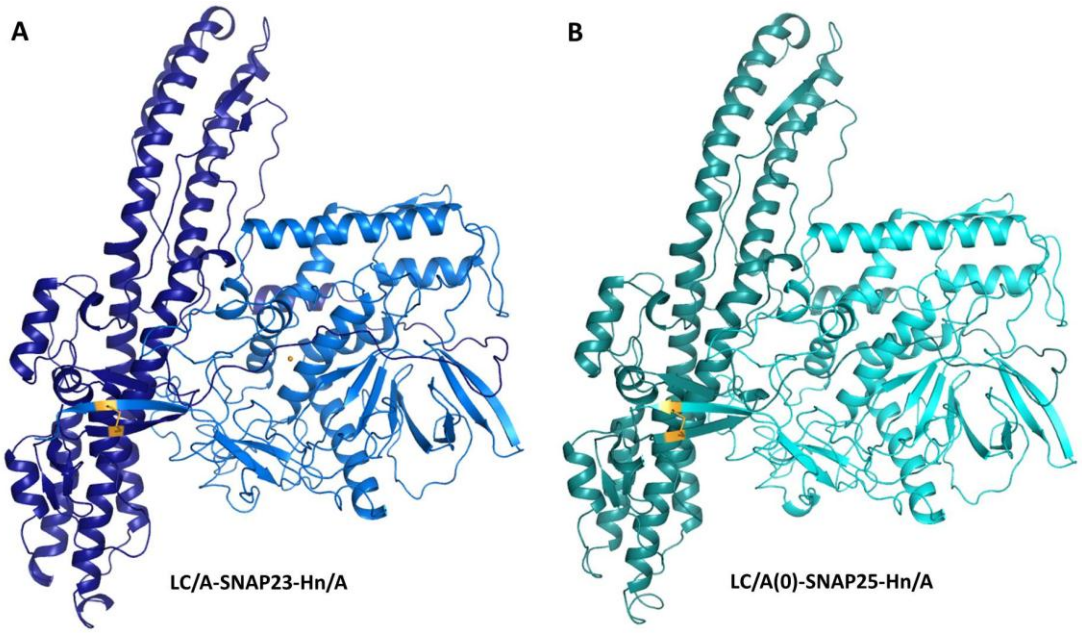

Figure 4.7. Crystal structures of LC/A-SNAP23-Hn/A and LC/A(0)-SNAP25-Hn/A. (A) Structure of LC/A-SNAP23-Hn/A, LC/A (light blue), Hn/A (dark blue), Zinc ion (orange sphere), disulphide bridge (gold sticks). **(B)** Structure of LC/A(0)-SNAP25-Hn/A, LC/A (light cyan), Hn/A (dark cyan), disulphide bridge (gold sticks).

Table 4.3. Data collection and refinement statistics of LC/A-SNAP23-Hn/A and LC/A(0)-SNAP25-Hn/A.

A. Data Collection Statistics	LC/A-SNAP23-Hn/A	LC/A(0)-SNAP25-Hn/A
Space Group	$P2_1$	$P2_12_12_1$
Cell	a =89.2, b=205.0, c=130.9 Å; $\alpha, \gamma = 90°, \beta = 91.9°$	a =79.0, b=157.5, c=209.4 Å; $\alpha, \beta, \gamma = 90°$
Resolution range (Å)	50-2.95	30-2.7
R_{sym}[1] (%)	13.0 (55.0)	10.0 (49.6)
I/σI (outer shell)	10.4 (3.0)	14.5 (2.9)
Completeness %	99.6 (99.2)	96.6 (79.7)
Redundancy	5.3 (5.1)	7.3 (4.7)
B. Refinement Statistics		
Resolution range (Å)	130-2.95	126-2.7
No. of reflections	515,451	508,979
Unique no. of reflections	97,807	69,459
R_{cryst}[2] (%)	25.0	21.5
R_{free}[3] (%)	29.3	26.6
Number of non-H atoms		
Protein	27,586	13,649
Ligand	4 Zinc atoms	5 Sulphate ions
Solvent	71	149
Average temperature factor ($Å^2$)	54.5	41.0
RMSD for bond lengths (Å)	0.006	0.008
RMSD for bond angles (°)	0.821	1.094

[1] $R_{sym} = \Sigma_h \Sigma_i |I(h) - I_i(h)| / \Sigma_h \Sigma_i I(h)$, where $I_i(h)$ and $I(h)$ are the *i*th and the mean measurements of the intensity of reflection *h*, respectively

[2] $R_{cryst} = \Sigma_h |F_o - F_c| / \Sigma_h F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection *h*, respectively.

[3] R_{free} is equal to R_{cryst} for a randomly selected 5.0 % reflections not used in the refinement.

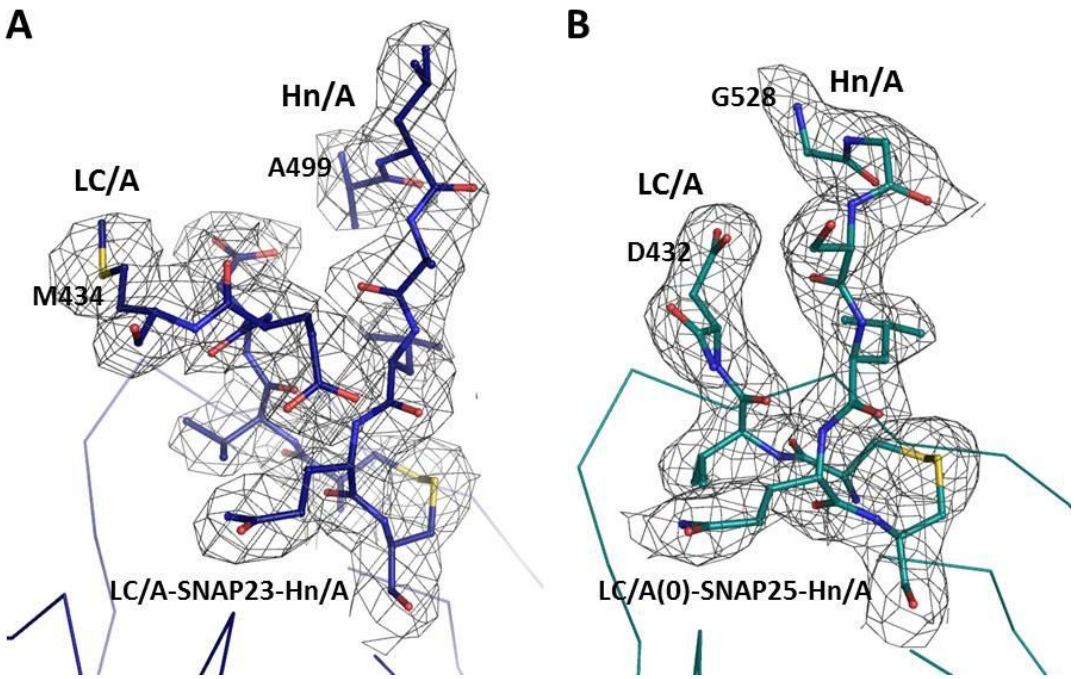

Figure 4.8. Missing segment at the LC-Hn interface in LC/A-SNAP23-Hn/A and LC/A(0)-SNAP25-Hn/A. (A) Structure of LC/A-SNAP23-Hn/A (blue), LC-Hn interface highlighting the disulphide bridge and missing electron density above the LHn backbone. Fo-Fc map at 1σ. **(B)** Structure of LC/A(0)-SNAP25-Hn/A (cyan), LC-Hn interface highlighting the disulphide bridge and missing electron density above the LHn backbone. Fo-Fc map at 1σ. Stick and ball representation with Oxygen atoms in red, Nitrogen in blue, Sulphur in yellow.

Comparison with LHn/A structure

The structures present the two domains of LHn interacting strongly. Despite inclusion of the SNARE peptides between the LC and Hn domains, each retains the fold seen in the structures of the full length toxin and LHn/A (Figure 4.9), including the "belt" region of Hn surrounding LC (Leu 500 – Leu 595, and Leu 534 – Leu 627, respectively for LC/A-SNAP23-Hn/A and LA(0)-SNAP25-HA) which is stabilised by a single disulphide bridge. This is highlighted by root mean square deviations with LHn/A of 0.8 and 0.7 Å respectively (over 847 C_{α}-atoms). In contrast to the structure obtained for LHn/A, the SNAP-engineered constructs are single chain molecules. Analysis by SDS-PAGE (Figure 4.2.D and E) confirmed the two molecules are not subject to self-proteolysis since SNAP23 is not a natural substrate for BoNT/A, and the SNAP25 construct proteolytic activity is prevented by a double mutation at the active site.

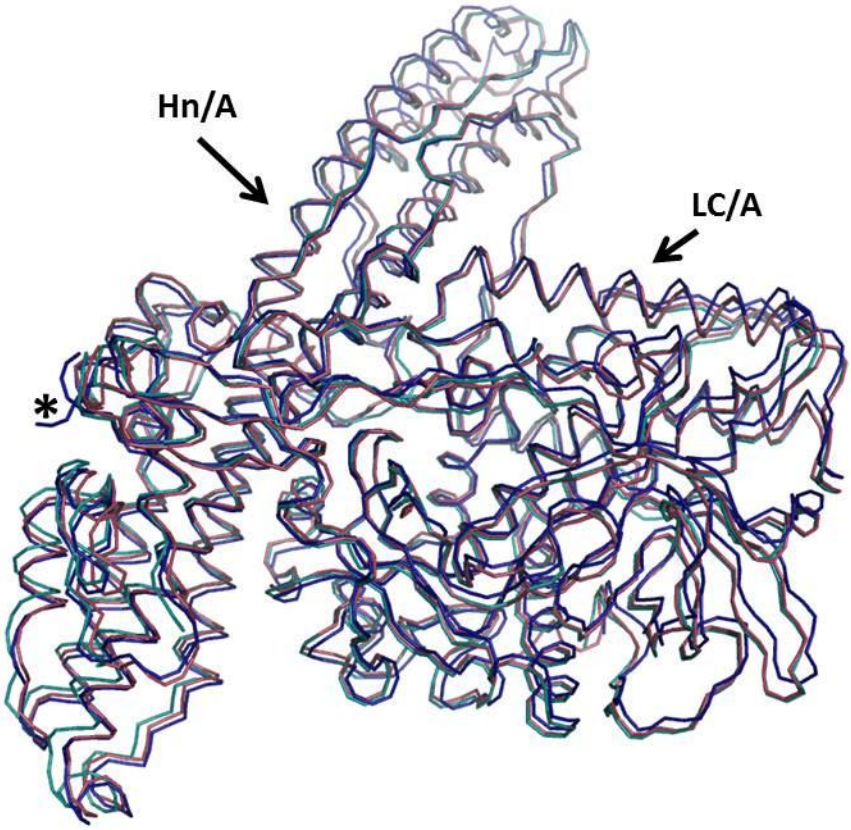

Figure 4.9. Crystal structures of LHn/A backbones. Ribbon representations of LHn/A (pink, PDB 2W2D, Masuyer *et al.*, 2009), LC/A-SNAP23-Hn/A (blue) and LC/A(0)-SNAP25-Hn/A (cyan) Asterisk marks location of the LC-Hn bridge.

In native BoNT/A, the zinc ion is tetrahedrally coordinated by His 223, His 227, Glu 262 and water-bounded Glu 224 (Figure 4.10.A, Table 4.4). This was observed in the LC/A-SNAP23-Hn/A structure. The mutations engineered into LC/A(0)-SNAP25-Hn/A, although different from the previously reported inactive LC/A mutant structure (Glu224Gln and Tyr366Phe, Breidenbach and Brünger, 2004), prevent binding of the catalytic zinc ion necessary for proteolytic activity. More particularly the hydroxyl group of Tyr 227 of LC/A(0) is in the position held by the zinc atom and is stabilised by a hydrogen bond with His 223. Interestingly this causes the side chain of Glu 262 to change orientation and form hydrogen bonds with residues Glu 257 and Ser 259. Mutation at position 224 from Glu to Gln prevents interaction of this amino acid with residues previously involved in the active site. Gln 224 is stabilised by direct interactions of its carboxyl group with Phe 163, and 2 water molecules, one of which is mediating a potential link with Thr 220 (Figure 4.10.B, Table 4.4).

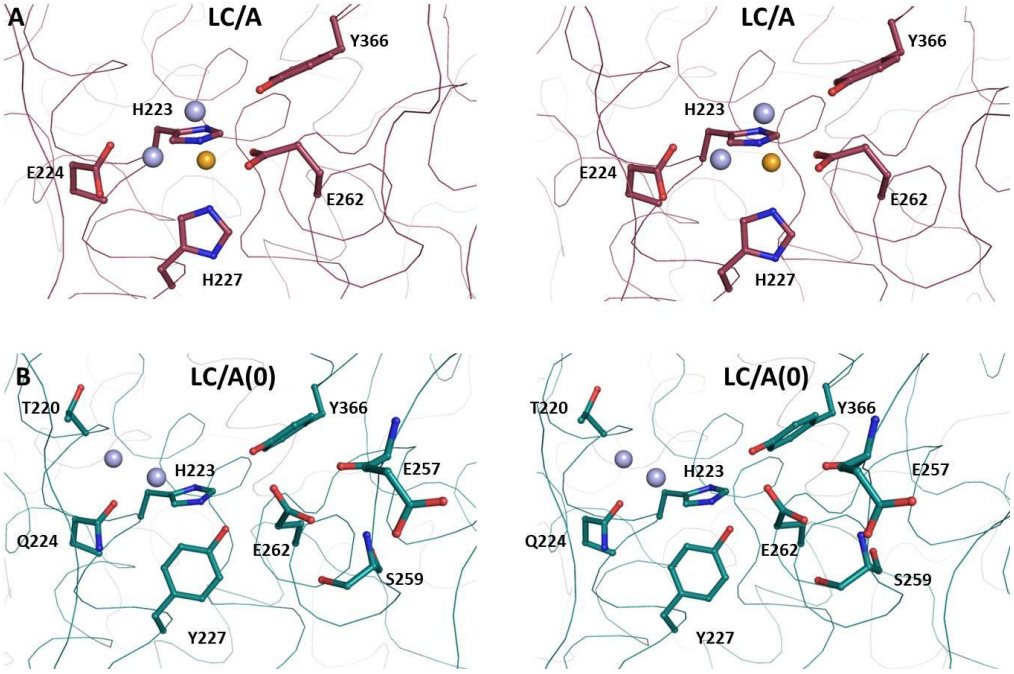

Figure 4.10. Structural consequences of mutations at LC/A catalytic site. Stereo figures of the crystal structures of **(A)** LHn/A (PDB code 2W2D, magenta), and **(B)** LC/A(0)-SNAP25-Hn/A (cyan). Residues involved in catalytic site interactions are highlighted in stick and ball, along with water molecules (light blue spheres) and zinc ion (orange sphere).

Table 4.4. Differences in catalytic sites interaction due to mutations in LC/A(0)-SNAP25-Hn/A

Potential bonds in LC/A (distance Å)			Potential bonds in LC/A(0) (distance Å)		
His 223 NE2	Zinc	2.7	His 223 NE2	Y227 OH	3.0
His 227 NE2	Zinc	2.3	Gln 224 OE1	Water 1	2.8
Glu 224 OE2	Water	2.5	Gln 224 OE1	Water 2	3.1
Water	Zinc	2.	Thr 220 CG2	Water 1	3.2
Glu 262 OE1	Zinc	2.1	Glu 262 OE2	Ser 259 OG	3.0
Glu 262 OE2	Tyr 366 OH	3.3	Glu 262 OE2	Ser 259 N	2.6
			Glu 262 OE2	Glu 257 O	2.7
			Glu 262 OE1	Tyr 366 OH	2.7

Superposition of the three LHn/A backbones (Figure 4.9) indicated a slight shift at the extremities of the translocation domain long α-helices. These areas of Hn seem to be located in solvent pockets and are not involved with symmetry related molecules (Figure 4.11). The most pronounced change is for LC/A-SNAP23-Hn/A which presents the lowest solvent

fraction. It is likely that some of these regions are involved in membrane insertion upon pH-mediated changes. Recent evidence pointed to conformational changes of the 659-681 region, while residues 826-835 were identified in membrane association experiments with LHn/A (Mushrush *et al.*, 2011). The solvent accessible loop corresponding to residues 561-564 is not visible in LC/A(0)-SNAP25-Hn/A.

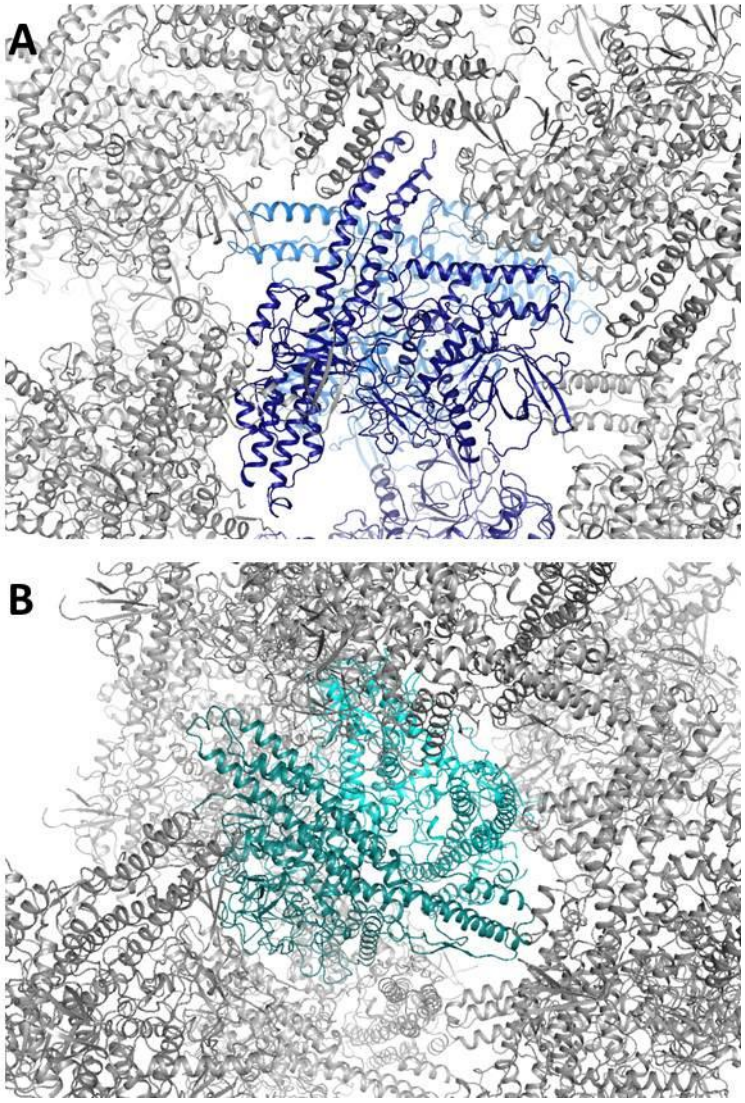

Figure 4.11. Crystal packing in LC/A-SNAP23-Hn/A and LC/A(0)-SNAP25-Hn/A structures. (A) Structure of LC/A-SNAP23-Hn/A (molecules of the crystallographic dimer in dark and light blue). **(B)** Structure of LC/A(0)-SNAP25-Hn/A (molecules of the crystallographic dimer in dark and light cyan). Symmetry-related molecules in grey.

Common dimeric arrangement in the LHn/A backbone structures

Despite the different crystallisation conditions and crystallographic unit cells (Table 4.3), dimeric interactions of the crystallographic LHn/A backbones appeared similar in the

three LHn/A-backbone structures (Figure 4.12). In this packing, orientation of the short α-strands at the LC-Hn interface is in a solvent accessible pocket. This may explain why the SNAP peptides linking the two domains are not visible. SNAP have been shown to be of a flexible nature in solution (Margittai *et al.*, 2001) and as a coiled-coil fold when in complex with its SNARE partners (Sutton *et al.*, 1998). Localisation of this peptide in the two constructs studied along with the favoured LHn crystal packing prevented SNAP-SNAP interaction into a coiled-coil structure.

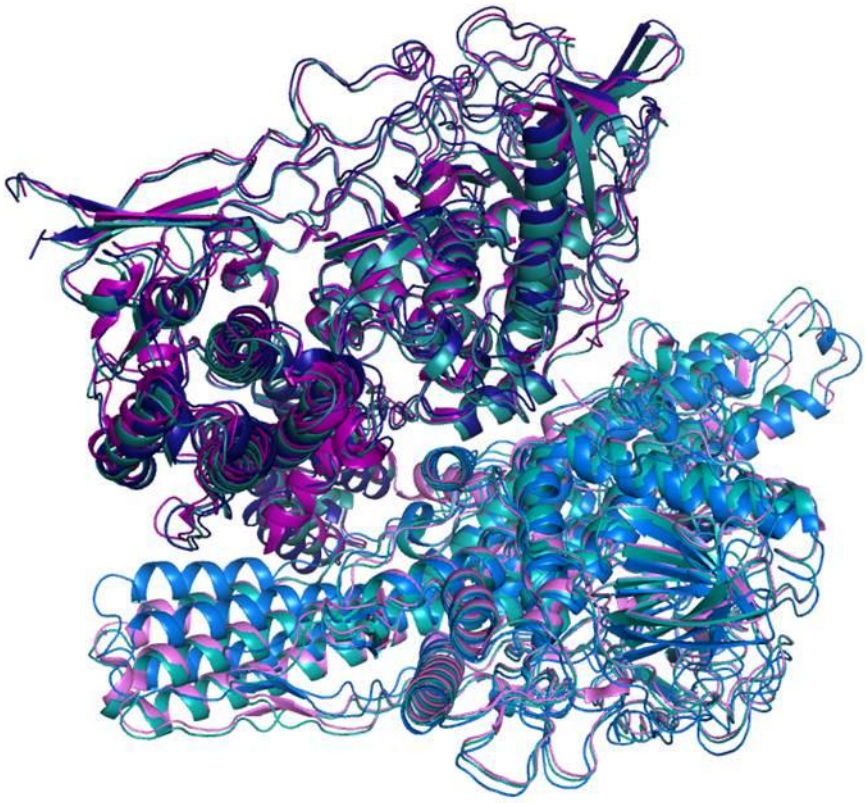

Figure 4.12. Dimeric interaction in LC/A-SNAP23-Hn/A and LC/A(0)-SNAP25-Hn/A crystal structures. Superposition of dimers from LHn/A (PDB 2W2D, magenta), LC/A-SNAP23-Hn/A (blue), and LC/A(0)-SNAP25-Hn/A (cyan) crystal structures. Molecules of the crystallographic dimer in dark and light colours.

SNAP25 forms partial secondary structural arrangements when bound to LC/A (Breidenbach and Brünger, 2004) which explain the long substrate requirements for optimal activity. Brünger *et al.* (2007) demonstrated that the belt region of Hn and SNAP25 superpose well. Competition between these two components for binding to LC might be expected. The structures presented here show that SNAP peptides do not disturb the strong domain interactions within LHn. LC/A-SNAP23-Hn/A and LC/A(0)-SNAP25-Hn/A are single chain proteins and crystallised in non-reducing conditions, thus favouring the stability of the two domains.

4.4. Discussion

Engineering self-activating LHn backbones

LHn backbones incorporating a SNARE linker were specifically cleaved by LC. Cleavage of the SNAP25 linker by LC/A accounted for the dissociation of reduced LC/A-SNAP25-Hn/A. It is likely that structural constraints would prevent intra-molecular enzymatic cleavage of the SNAP25 linker, and that inter-molecular cleavage accounts for proteolysis. Support for inter-molecular cleavage was seen by the cross-activating properties of LC/A-VAMP2-HN/A and LC/B-SNAP25 proteins when incubated together.

Engineering multi-domain proteins for optimal expression often involves empirical testing of polypeptide variants. It is well known that BoNT/A requires an extended substrate interaction for optimal cleavage kinetics (Breidenbach and Brünger, 2004; Chen *et al.*, 2007). SNARE inter-domain linkers that correspond to the complete coiled coil region underwent complete activation by self-cleavage or when treated with an exogenous LC protease. The SNAP25 and VAMP2 peptide linkers used in this study included the binding and recognition sites for all of the botulinum serotypes known. The minimum SNAP25 peptide length effective for complete activation was not determined, but shorter SNAP25 peptide linkers were tested that either failed to self-activate or to do so to completion (results not shown). This is in agreement with previous observations that shorter SNARE peptides are less effective substrates for LC (Chen and Barbieri 2006, Shone *et al.*, 1993).

Activated LHn are di-chain molecules that take the form of separate LC and Hn proteins paired by a di-sulphide bond. However, self-cleaving LHn backbones were seen to dissociate into LC and Hn domains during protein expression in *E.coli*. LC and Hn subsequently assembled spontaneously upon oxidisation during purification. Disulphide bond formation remained possible with the cysteines at the carboxyl end of LC/A and amino-terminal side of Hn/A. Nevertheless, as could be expected from uncontrolled association, both homomeric and heteromeric dimers were observed when reconstituted from LC and Hn monomers. Assembly of LC/A and HC/A to form functionally active, disulphide bonded BoNT/A has been reported previously (Zhou *et al.*, 1995). In addition, a disulphide bonded carboxyl heavy chain (HC) dimer has also been observed during vaccine development for the recombinant HC binding domain of BoNT/A (Bouvier *et al.*, 2003).

Use of different expression cell lines which could enhance disulphide bond formation in the cytoplasm, may help maintain oxidised cysteines and retain an intact disulphide bond in self-activating LC/A-SNAP25-Hn/A. Alternatively, LC activation under oxidising

conditions, as performed with other exogenous proteases, preserved molecular integrity of the disulphide bond.

To improve the recovery of intact LHn backbone, LC was employed as an exogenous protease for activation. Oxidised LC/A-VAMP2-Hn/A single-chain polypeptide was activated by LC/B, while immobilised in an affinity matrix. The use of immobilised proteases for post translational modification of target proteins is an established technique (Kubitzki *et al.,* 2008). However, use of BoNT LC and SNARE sequences in such a role has not previously been reported. Rapid, on column activation of single-chain LC/A-VAMP2-Hn/A polypeptide by LC/B was achieved. The high specificity of LC/B for VAMP2 means that use of large quantities of enzyme does not lead to secondary cleavage events within the LC/A-VAMP2-Hn/A target backbone. Use of the inherent endopeptidase property of LHn backbones for post translational intermolecular self-processing offers a distinct benefit to that of other commonly used proteases. The exquisite interaction between LC and SNARE proteins for optimal substrate binding and catalysis (Breidenbach and Brünger, 2004), reduces the probability of secondary cleavage events occurring using these enzymes.

Differences in the relative potency of self-activated LC/A-SNAP25-Hn/A compared to activated LHn/A were observed. More particularly, LC/A-SNAP25-Hn/A presented a lower potency in cell free cleavage assays whereas it showed equivalent activity in spinal cord neurons (Figures 4.4 and 4.6). The control molecule, non-activated, scLHn/A, here and reported by Fischer *et al*, (2008) cleaved SNAP25 at a much reduced rate compared to activated LHn/A in both cell free and cellular assays. This suggests the difference in relative potency of self-activated LC/A-SNAP25-Hn/A to activated LHn/A between the two assays is a consequence of the molecule itself rather than assay conditions. The reason for reduced potency of LC/A-SNAP25-Hn/A in cell free assays is most likely caused by the SNAP25 peptide that remains part of the LC/A and Hn/A after self-cleavage (LC/A-SNAP25[145-197] and SNAP25[198-206]-Hn/A). Indeed the potency of LHn/A in cell free SNAP25 cleavage assays can also be reduced by the addition of LC/B-SNAP25 as a source of SNAP25[145-197] (Figure 4.5). Further support for this hypothesis is that LC/A-VAMP2-Hn/A (di-chain) did not exhibit reduced potency in the SNAP25-GFP assays. This is in agreement for an LHn backbone containing a SNARE peptide that is not susceptible to binding or cleavage by LC/A.

There are several noticeable differences between the two functional assays performed in this study that may have an impact on the results described above. Firstly, the two assays measure the cleavage of different SNAP25 substrates. In the cell free assay, catalysis of an artificial SNAP25-GFP substrate was assessed, while in the eSCN assay endogenous

SNAP25 cleavage was measured. SNAP25 fusion to the 27 kDa GFP tag might cause alterations in substrate structure and its interaction with the enzyme. Lack of post translational modification of the recombinant substrate compared to native SNAP25 might also contribute to the differences observed between assays. Secondly, other proteins interacting with SNAP25, such as other SNARE proteins that are present in eSCNs but are not present in the cell free assays, may influence the activity of the LHn-SNARE derivatives. Thirdly, the mechanism by which the LHn backbone delivers LC inside neuronal cells is not known. Thus the interactions between LC and Hn may impact on the molecules activity depending on their redox state and the ability of the two domains to separate. A study by Gul *et al.* (2010) demonstrated the higher specific activity of LC/A moiety in comparison to LHn/A and BoNT/A.

Crystal structures of LHn-SNARE derivatives

The crystal structures of two constructs designed around the LHn/A scaffold and engineered to contain an extended SNARE peptide at the LC-Hn interface were determined. In order to maintain an intact protein assemblage, these constructs were engineered to prevent self-activation either by use of a non-cleavable substrate homologue (SNAP23) or an enzymatically inactive light chain mutant. They therefore offer interesting models to study the stability of engineered LHn derivatives. Despite not observing the entire molecules, the structure of the main framework could be solved. These demonstrated the stability of LHn/A, and also its flexibility in supporting additional engineered peptide segments.

The light chain structure was unchanged. Mutations at key residues of the catalytic site had for consequences a movement of the residues normally responsible for the zinc coordination, explaining the loss of LC/A proteolytic activity in LC/A(0)-SNAP25-Hn/A. The long helices of Hn/A did not present any significant conformational change and are stabilised by strong inter-helical interactions, although a slight shift in the solvent accessible extremities was observed. Interestingly, there appear to be a strong dimeric interaction due to crystallographic packing between LHn/A molecules which is similar in all the LHn/A structures solved. This interaction has no biological implications but seemed to favour the crystallisation of LHn/A-based molecules.

Engineered peptides at the LC-Hn interface were not visible. The localisation of such peptides in solvent pockets of the crystal, as well as the flexible structure of these peptides may explain this phenomenon. It would be interesting to design larger inserts with known secondary structures to study their impact on the LHn fold.

Conclusion

This study demonstrated that it is feasible to engineer LHn backbones to incorporate highly specific protease recognition sites based on SNARE polypeptides. These engineered backbones were proven to be susceptible to both self-activation and activation by treatment with exogenous LC. Importantly they are functional molecules *in vitro*. LHn backbones incorporating SNARE peptides provide an opportunity to optimise their manufacture process by an alternative activation strategy to traditional methods. They were shown to retain the ability to translocate LC across the endosome membrane and deliver it into the cytosol to cleave endogenous SNARE proteins.

LHn fragments are safe and reliable tools to study the mechanism of action of the botulinum neurotoxins (Fischer *et al.,* 2008b; Mushrush *et al.,* 2011). The crystal structures presented in this study show that LHn provides a fantastic framework for protein engineering where functional peptides can be added without disturbing the overall structure of these proteins.

Chapter 5. Purification and crystallisation of *Clostridium botulinum* neurotoxin derivatives

5.1. Other LHn molecules

5.1.1. Introduction

The structures of the LHn fragment from botulinum neurotoxin serotypes A and B have been described and demonstrated the stability of these molecules. Furthermore, both serotypes showed that they are functional fragments, retaining the catalytic properties of BoNT as well as the ability to cleave intracellular SNARE proteins in neurons. Retargeted toxins with therapeutic potential based on the LHn backbones are therefore likely to maintain some of the biochemical properties of their parent neurotoxins, in terms of catalytic activity and translocation of the light chain within the target cells cytosol.

Interestingly, each serotype presents variations in specificity and toxicity (Montal, 2010). For example, BoNT/C was shown to be particularly cytotoxic compared to serotype A (Williamson and Neale, 1998). Additionally, serotype C is the only BoNT able to cleave two of the SNARE proteins, SNAP25 and syntaxin (Schiavo *et al.*, 2000). On the other hand, BoNT/E was demonstrated to have a faster translocation rate and thus the ability to act more rapidly (Wang *et al.,* 2008).

The level of primary sequence homology between the seven serotypes (approximately 35% identity, Lacy and Stevens, 1999) indicates that the structure for all serotypes is likely to be conserved. While this is true of the domain structures determined so far (Swaminathan, 2011), the crystal structure of BoNT/E showed that the domains' spatial organisation may differ (Kumaran *et al.,* 2009).

The LHn fragments, composed of the catalytic (LC) and translocation (Hn) domains, represent useful tools to study the domain interactions underlying the mechanism of BoNT toxicity. The production of recombinant LHn has been described previously for LHn/A, /B and /C (Sutton *et al.,* 2005, Chapter 3). These engineered fragments were optimised for expression in *E. coli*. They include a specific exoprotease cleavage site between LC and Hn, to allow for activation into biologically active di-chain molecules.

Harnessing the different LHn serotypes may provide further flexibility in the design of novel molecules derived from BoNTs. The variations seen between the serotypes offer a selection of biochemical properties that could prove useful for tailoring a therapeutic

molecule against a particular disease. The production and characterisation of the LHn molecules should therefore provide valuable information for protein engineering.

In this study, the crystal structure of a molecule representing an LHn/B backbone was solved at 2.7 Å. This construct included peptide linker segments around the exoprotease activation site, with the aim of analysing the differences made to primary sequence changes in this essential region of the botulinum neurotoxins. Although the additional engineered fragment could not be modelled due to disorder, the structure confirmed the significant stability of the LHn fold. Furthermore, the production of recombinant LHn serotypes C and E was achieved. Both serotypes were successfully crystallised and preliminary x-ray diffraction data were obtained for LHn/C.

5.1.2. Methods

Protein expression and purification

The synthetic gene encoding 863 amino acids of LHn/B was cloned into modified pET vector (Novagen, UK) with a C-terminal 6 x His-tag and transformed into *E. coli* BL21 expression cells. The LHn/B gene was engineered to encode for a Factor Xa cleavage site (IEGR) flanked by three GGGGS repeats between the LC and Hn domains (LC/B Asp 439 – Hn/B Leu 477). The clone was provided by Syntaxin Ltd. Expression and purification of LC/B-GS-Hn/B was performed as described previously for LHn/B (Masuyer *et al.*, 2011). Briefly, expression cultures were grown in terrific broth and induced with IPTG (1mM). After cell lysis, LHn/B in the soluble fraction was purified by affinity chromatography (HisTrap HP, GE Healthcare), activated by Factor Xa cleavage (New England BioLabs) and finally hydrophobic interaction chromatography (Phenyl Sepharose HP, GE Healthcare), prior to concentration to 6.5 mg/ml (Figure 5.1). All concentrations were determined by A_{280} measurement.

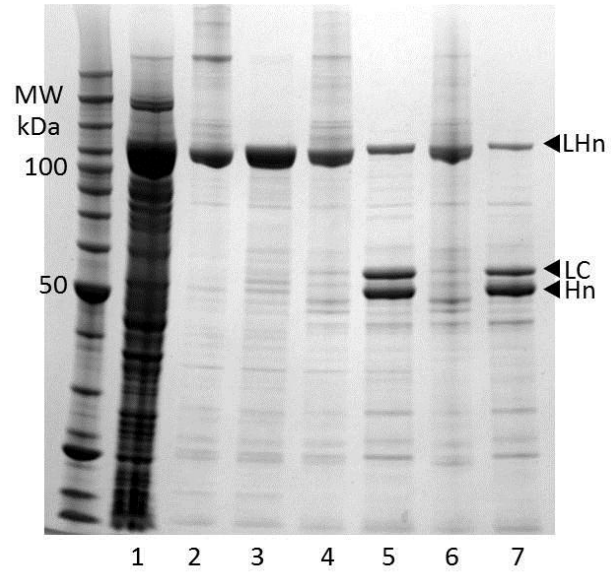

Figure 5.1. SDS-PAGE analysis of LC/B-GS-Hn/B purification. Lane 1, cell lysate supernatant; 2, 3, eluate from first affinity purification step in oxidised (O) and reduced (R) conditions (respectively); 4, 5, sample activated by Factor Xa (O and R); 6, 7, eluate from hydrophobic interaction purification.

The synthetic gene encoding 868 amino acids of LHn/C was cloned into a modified pET vector (Novagen, UK) with an N-terminal (cleavable) 6 x His-tag and transformed into *E. coli* BL21 expression cells. The LHn/C gene was engineered to encode for a Factor Xa cleavage site (IEGR) between the LC and Hn domains at position 440-445. The clone was provided by Syntaxin Ltd. Expression cultures and cell preparation of LHn/C were performed as per previously reported for other LHn (Chapter 3). Soluble fraction was loaded onto a Ni^{2+}-charged chelating sepharose column (GE Healthcare). LHn/C eluted at 100 and 250 mM imidazole (dissolved in buffer A) and was dialysed overnight at 4°C against buffer A. Activation of purified LHn/C was achieved by Factor Xa (New England BioLabs) treatment. The cleaved fusion protein sample was loaded onto a HisTrap column (GE Healthcare) and the flow-through collected to be dialysed against 50 mM HEPES, pH 7.2. This material was then applied to an anion exchange Q HiTrap FF column (GE Healthcare). LHn/C eluted at 200 mM NaCl. (in 50 mM HEPES, pH 7.2). The sample was finally concentrated using Vivaspin 50000 MWCO concentrator (Millipore) to 12.0 mg/ml (Figure 5.2A).

The synthetic gene encoding 859 amino acids of LHn/E was cloned into a modified pET vector (Novagen, UK) with a C-terminal 6 x His-tag and transformed into *E. coli* BL21 expression cells. The clone was provided by Syntaxin Ltd. The LHn/E gene was engineered to encode for a Factor Xa cleavage site (IEGR) between LC and Hn at positions 414-419.

Expression cultures and cell preparation of LHn/E were carried out as described for other LHn (Chapter 3). Soluble LHn/E was purified by affinity chromatography (HisTrap HP, GE Healthcare), followed by hydrophobic interaction chromatography (Phenyl Sepharose FF, GE Healthcare). The final sample consists of a single chain LHn/E and was concentrated to 5.3 mg/ml (Figure 5.2B).

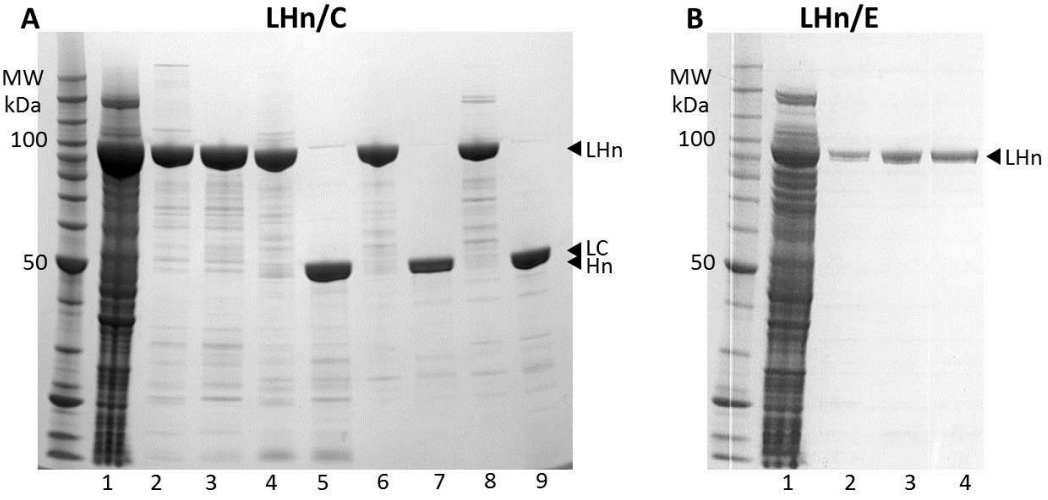

Figure 5.2. SDS-PAGE analysis of LHn/C and /E purifications. (A) Summary of LHn/C purification. Lane 1, cell lysate supernatant; 2, 3, eluate from first affinity purification step in oxidised (O) and reduced (R) conditions (respectively); 4, 5, sample activated by Factor Xa (O and R); 6, 7, eluate from second affinity purification (O and R); 8, 9, eluate from hydrophobic interaction purification (O and R). **(B)** Summary of LHn/E (single chain) purification. Lane 1, cell lysate supernatant; 2, eluate from first affinity purification step (O); 3, 4, eluate from hydrophobic interaction purification (O and R).

Crystallisation

Initial crystallisation condition for LC/B-GS-Hn/B was identified using the JCSG-*plus* screen (Molecular Dimensions) corresponding to 0.1 M TRIS pH 8.5, 0.2 M magnesium chloride, 20% PEG8000. Optimised crystals were grown manually in 2-3 weeks by the hanging drop method (2 µl protein and 1 µl mother liquor against a 500 µl reservoir) at 16°C in 0.1 M TRIS acetate pH 8.5, 0.2 M magnesium chloride, 12% PEG6000. X-ray diffraction data were collected at the Diamond Light Source, UK, beamline IO3 (Figure 5.3). A complete dataset to 2.7 Å was obtained from a single crystal. The data were processed and scaled in orthorhombic space group $P2_12_12_1$ using MOSFLM and SCALA (Leslie, 2006; CCP4, 1994; Table 5.1). Initial phases were obtained by molecular replacement using Phaser (McCoy *et al.*, 2007) with the coordinates of LHn/B (PDB 2XHL, Masuyer *et al.*, 2011). Crystallographic refinement was carried out using REFMAC5 (version 5.5) (Vagin *et al.*, 2004) and manual model fittings were done using Coot version (0.6.1) (Emsley and Cowtan,

2004). The structures were validated using MolProbity (Davis *et al.*, 2007). Figures were drawn with PyMOL (DeLano Scientific LLC).

The atomic coordinates and the structure factors have been deposited with the RCSB Protein Data Bank under accession code 3ZUQ.

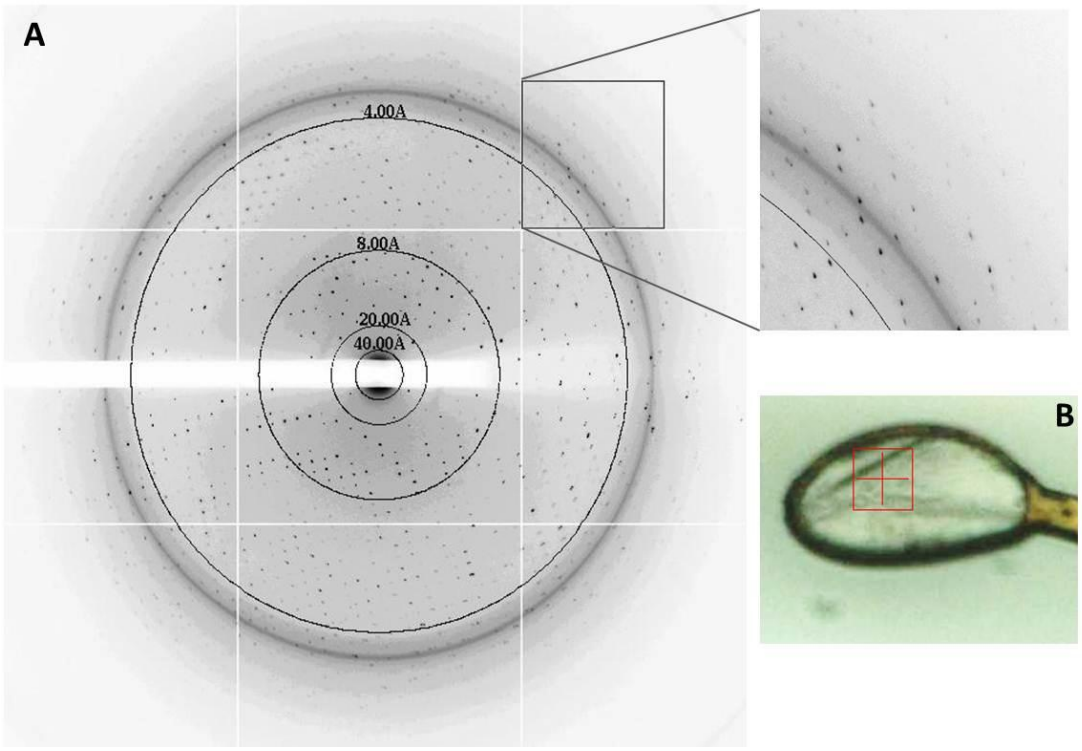

Figure 5.3. Crystal and x-ray diffraction of LC/B-GS-Hn/B. (A) Diffraction image collected at DLS IO3 where LC/B-GS-Hn/B crystals diffracted at 2.7 Å. **(B)** Crystals of LC/B-GS-Hn/B mounted for data collection, optimised crystal grew with 0.1 M TRIS acetate pH 8.5, 0.2 M magnesium chloride, 12% PEG6000. (Red square is 167x167 μm).

Table 5.1. Data collection and refinement statistics for LC/B-GS-Hn/B

A. Data Collection Statistics	LC/B-GS-Hn/B
Space Group	$P2_12_12_1$
Molecules/asymmetric unit	1
Cell	a=89.4, b=103.8, c=115.0 Å; α, β, γ = 90°
Resolution range (Å)	58-2.7
R_{sym}[1] (%)	13.4 (49.2)
I/σI (outer shell)	8.5 (3.3)
Completeness %	93.3 (100.0)
Redundancy	5.2 (5.4)
B. Refinement Statistics	
Resolution range (Å)	77-2.2
No. of reflections	146,733
Unique no. of reflections	28,039
R_{cryst}[2] (%)	25.0
R_{free}[3] (%)	28.2
Number of non-H atoms	
Protein	6953
Ligand	1 Zinc atom
Solvent (water)	44
Average temperature factor (Å^2)	36.5
RMSD for bond lengths (Å)	0.006
RMSD for bond angles (°)	0.844

[1] $R_{sym} = \Sigma_h \Sigma_i |I(h) - I_i(h)| / \Sigma_h \Sigma_i I(h)$, where $I_i(h)$ and I(h) are the *i*th and the mean measurements of the intensity of reflection *h*, respectively

[2] $R_{cryst} = \Sigma_h |F_o - F_c| / \Sigma_h F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection *h*, respectively.

[3] R_{free} is equal to R_{cryst} for a randomly selected 5.0 % reflections not used in the refinement.

For LHn/C, several crystallisation conditions were identified in the PACT *premier* and Heavy + Light Twin Pack screens (Molecular Dimensions). The conditions were repeated manually by the hanging-drop method and optimised. Crystals could be readily grown in a few days in 0.1 M HEPES pH 7.2, 1.5 M lithium sulphate. Crystals grew only as stacks of very thin needles. Streak and macro-seeding was performed to optimise the crystal size. Single needles were finally obtained but were too weak to allow for collection of a

complete dataset. Few low resolution (4.0 Å) images were taken at the Diamond Light Source, UK, beamline I24 (Figure 5.4). The images could not be indexed precisely due to the weak x-ray diffraction. The cell dimensions obtained using MOSFLM (CCP4, 1994; Leslie, 2006) were a= 139, b= 150, c= 239 Å; α= 94, β= 90, γ = 101°.

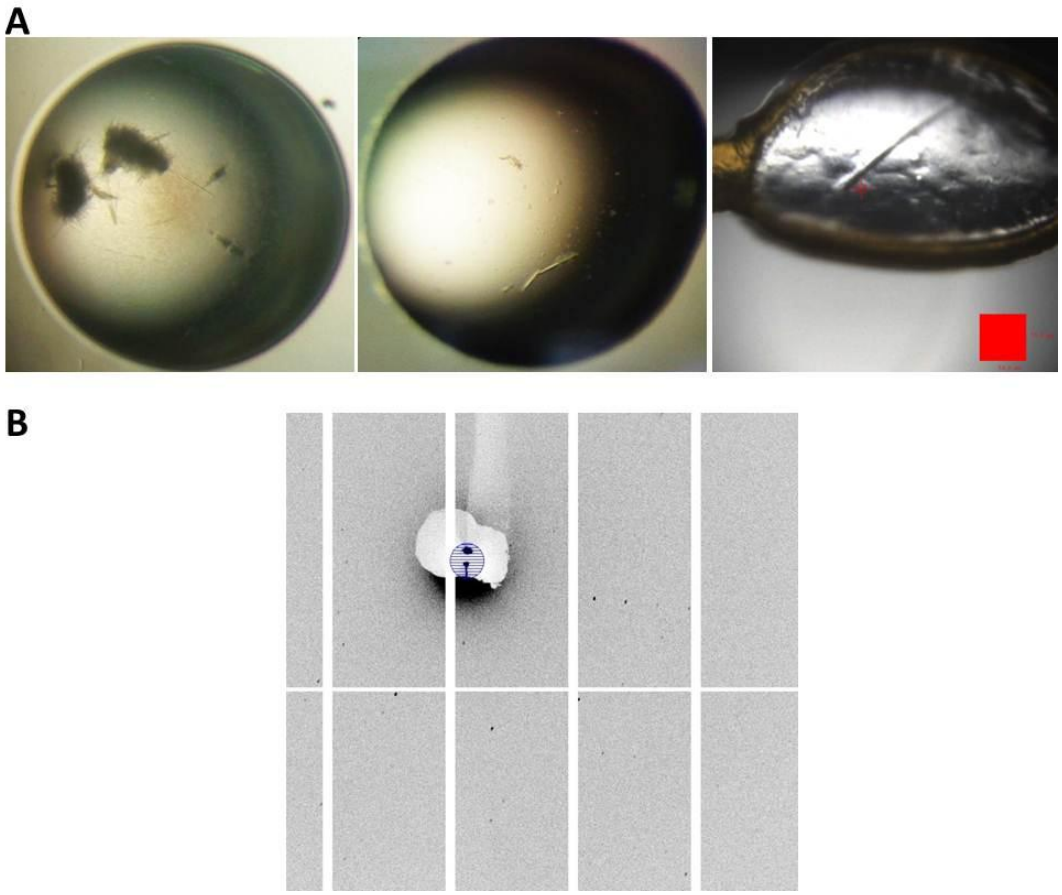

Figure 5.4. Crystal and x-ray diffraction of LHn/C. (A) Optimisation of LHn/C crystallisation. Crystals grew in 0.1 M HEPES pH 7.2, 1.5 M lithium sulphate. From **left** to **right** respectively, thin needle stacks, single needles, and crystal of LHn/C mounted for data collection at DLS I24 (red square is 55x56 μm). **(B)** Diffraction image collected at DLS I24 where LHn/C crystals diffracted at >4.0 Å. Zoom around beam centre (blue circle).

Concerning LHn/E, a few crystallisation conditions were identified in the primary screening. Some of the conditions could be repeated manually by the hanging-drop method, particularly with 3.0M sodium formate or 0.1 M TRIS pH 8.5, 0.2 M ammonium sulphate, 15% PEG3350. Extremely thin single needles were visible, and crystallisation optimisation trials were set up. The crystals however disappeared after a few days and none were of good enough quality for x-ray diffraction.

5.1.3. Results and discussion

Crystal structure of LC/B-GS-Hn/B

The crystal structure of LC/B-GS-Hn/B has been determined at 2.7Å resolution. A straightforward molecular replacement solution was found with PHASER using LHn/B as a search model, with 1 molecule per asymmetric unit and 50.2% solvent content. The LHn/B backbone, composed of LC-Hn in its di-chain form linked by a disulphide bridge, was observable but no electron density could be seen for the GS linkers (GGGGS repeats), inserted on both side of the Factor Xa cleavage site between LC and Hn. This corresponds to 32 missing residues (excluding the "IEGR" protease site). The structure was refined to a final R_{free} = 28.1%, and R_{cryst} = 25.0% (Table 5.1, Figure 5.5), with 96.1% of amino acids in the Ramachandran plot allowed region and 44 water molecules per asymmetric unit.

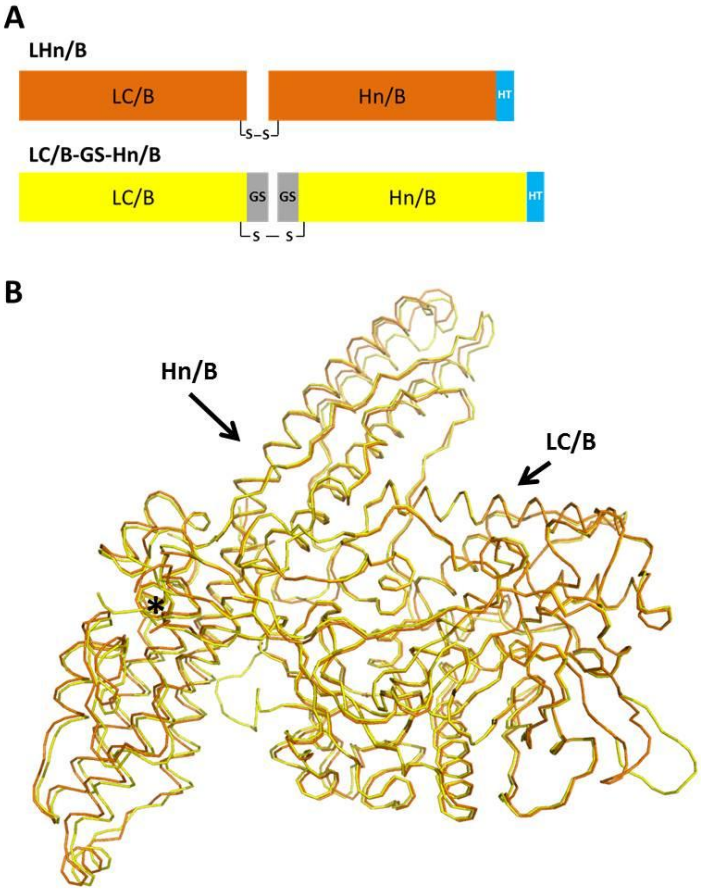

Figure 5.5. Crystal structure of LC/B-GS-Hn/B. (A) Representation of LHn/B domain structures. With LHn/B (orange, PDB 2XHL, Masuyer *et al.*, 2011) and LC/B-GS-Hn/B (yellow), His tag (HT, light blue), disulphide bonds are represented (S-S). **(B)** Ribbon representation of the crystal structures of LC/B-GS-Hn/B (yellow) superposed with LHn/B (orange). Asterisk marks the LC-Hn interface where the disulphide bridge and GS linkers (not visible) are located.

Overall, LC/B-GS-Hn/B superposed well with the solved structures of LHn/B and BoNT/B (Figures 5.5 and 5.7A), with a root mean square deviation of 1.2Å over 839 Cα-atoms with LHn/B. SDS-PAGE analysis of the purified LC/B-GS-Hn/B confirmed its Factor Xa activation into a di-chain molecule (Figure 5.1). Addition of GS-linker sequences between LC and Hn did not alter the structure of the protein nor the two domains interaction. Indeed, the Hn belt region surrounds LC at similar position to BoNT/B, and ends in a short β-sheet arrangement that stabilises the disulphide bridge linking the two domains. Electron density is only visible around the residues delimited by the LHn/B backbone, with nothing visible between LC – Gly 440 and Hn – Leu 477 (Figure 5.6). GS linkers are flexible protein segments by nature, and in the context of the crystal packing are within a solvent accessible area, it is therefore not surprising to be missing this region of the structure.

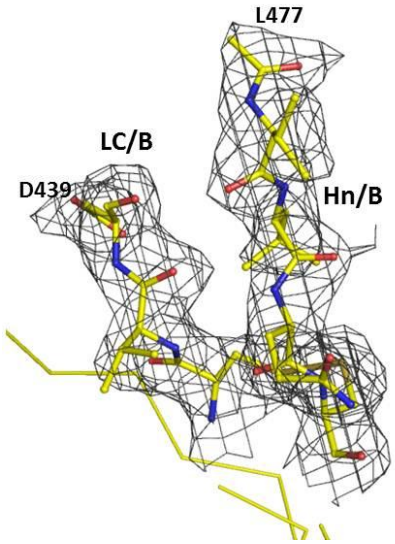

Figure 5.6. Missing segment at the LC-Hn interface in LC/B-GS-Hn/B. Structure of LC/B-GS-Hn/B (yellow), LC-Hn interface highlighting the disulphide bridge and missing electron density above the LHn backbone. Fo-Fc map at 1σ.

Comparison of LC/B-GS-Hn/B with previous structures highlights several noticeable differences, particularly around three loop regions (Figure 5.7). (i) Firstly, the loop corresponding to residues 142-150 is protruding off the globular fold of LC and has a different orientation to that seen in the previous LC/B structures. The electron density in this area is weak, indicative of the loop's movement. (ii) Secondly, residues 208-218 downstream of the active site, which could not be modelled in the first LHn/B structure, present a different arrangement to that seen in BoNT/B where it was stabilised by interactions with helices α17 and α19 of Hn. In the LC/B-GS-Hn/B structure, it interacts mainly with α19 through different potential hydrogen bonds and van der Waals interactions.

The alternative orientation of this loop confirms its flexibility and may have some implications for substrate binding or access to the nearby catalytic site. (iii) Additionally, loop 625-631 of Hn could not be modelled due to the lack of electron density, similarly to LHn/B. This loop is stabilised by the binding domain in BoNT/B, which is missing in this structure. On another hand, the Hn domain showed little difference with what had been observed in previous structures with only a slight shift in its long helices extremities that are not involved in crystal packing interactions (Figure 5.7B). In a parallel with BoNT/A these regions might be involved in pH-dependent conformational changes and membrane insertion (Mushrush *et al.*, 2011).

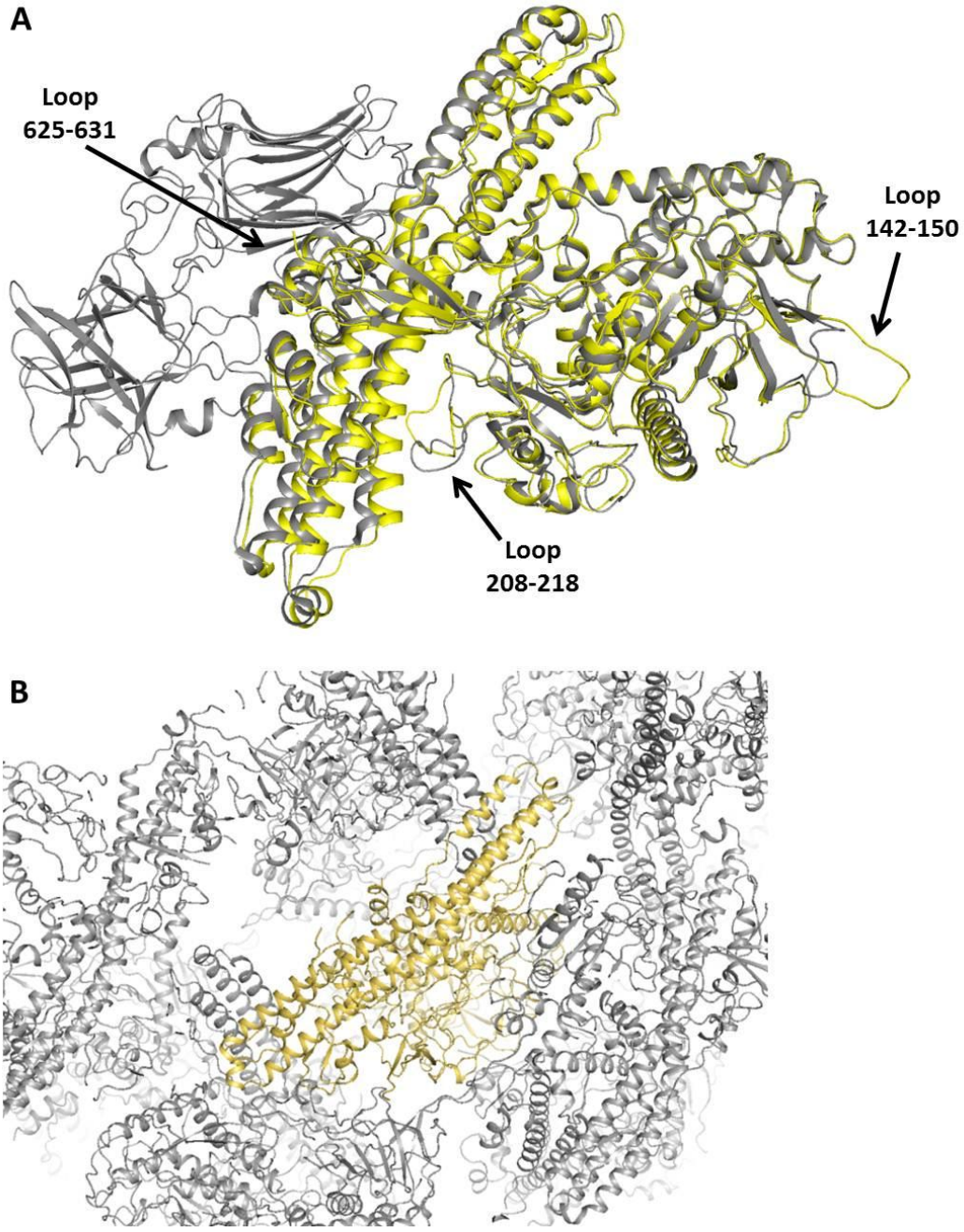

Figure 5.7. Structure of LC/B-GS-Hn/B, compared to BoNT/B, and crystal packing. **(A)** Cartoon representation of LC/B-GS-Hn/B (yellow), and BoNT/B (grey; PDB 1EPW, Swaminathan and Eswaramoorthy, 2000) superposed crystal structures. Loops with a different conformation are highlighted. **(B)** Lattice structure of LC/B-GS-Hn/B. Symmetry-related molecules are shown in grey.

Purification and crystallisation of LHn/C and /E

Recombinant LHn/C was successfully purified from *E. coli*. A high expression yield was obtained and purification using metal chelate affinity chromatography was sufficient to obtain the purity required for crystallography (Figure 5.2A). A final anion exchange

purification step was performed but did not appear to eliminate some of the minor contaminants visible by SDS PAGE. Activation by Factor Xa treatment was highly effective with only a very faint band visible at approximately 100 kDa in the reduced sample corresponding to single-chain LHn/C. Noticeably, LHn/C was concentrated up to 12 mg/ml. Only a slight precipitate was observable at this concentration while other LHn (/A, /B, /D and/E) would typically not subsist at concentration higher than 5 mg/ml in similar buffer conditions.

Crystallisation trials with LHn/C led to several hits. All were very thin needles as presented in figure 5.4A. It took several rounds of optimisation using streak seeding first with a lower LHn/C concentration and different levels of precipitants to obtain single crystals. All the crystals obtained were needle-like and the best ones only showed weak x-ray diffraction on a microfocus beamline at the DLS synchrotron (Figure 5.4B). The couple of images obtained could not be indexed accurately, however the presumed cell dimensions (a= 139, b= 150, c= 239 Å) were consistent with crystals from other LHn molecules. Optimisation of these crystals is still on-going.

The production of a recombinant LHn/E was achieved and is reported for the first time here. The expression level in *E. coli* was lower than for other proteins (Figure 5.2B) but yielded enough material for purification. A first affinity purification step using the poly-His tag combined with hydrophobic interaction chromatography produced a highly pure protein when analysed by SDS-PAGE. Despite having a Factor Xa cleavage site engineered between LC and Hn, the sample was not activated. Previous activation experiments had proven unsuccessful with Factor Xa displaying poor cleavage results, and trypsin yielding to non-specific cleavage products (results not shown). While LHn/E in its single chain form might have reduced activity, it still represented an interesting molecule for crystallographic studies. Preliminary crystallisation conditions were determined and showed long and very thin single needles. The crystals however disappeared after a few days and were difficult to reproduce. Optimisation of these crystals is still on-going.

Conclusion

Production and crystallisation of recombinant LHn/B (LC/B-GS-Hn/B), /C and E/ were accomplished. Preliminary x-ray diffraction data were collected for LHn/C and further crystal optimisation will be necessary to obtain the protein structure. Additionally, the crystal structure of LC/B-GS-Hn/B was solved at 2.7 Å. While it confirmed the structural stability of the LHn conformation, the engineered peptides (GS) at the domain interface were not visible.

5.2. Recombinant ZZ-LHn derivatives

5.2.1. Introduction

Protein A is a polypeptide from *Staphylococcus aureus* known for its capacity to bind immunoglobulin G (IgG). Nilsson *et al.* (1987) described a synthetic IgG binding domain, Z, based on an optimised B domain of protein A. This molecule subsequently proved useful as a protein engineering tool by showing enhanced expression and solubility of proteins expressed with a ZZ-tag, as well as offering a capacity for protein purification by IgG affinity chromatography (Moks *et al.*, 1987). The IgG-binding property of ZZ-fusion proteins may also be used for functional studies. Nizard *et al.* (2001) demonstrated the successful anchoring of a ZZ protein at the cell surface when attached to the diphtheria toxin transmembrane domain.

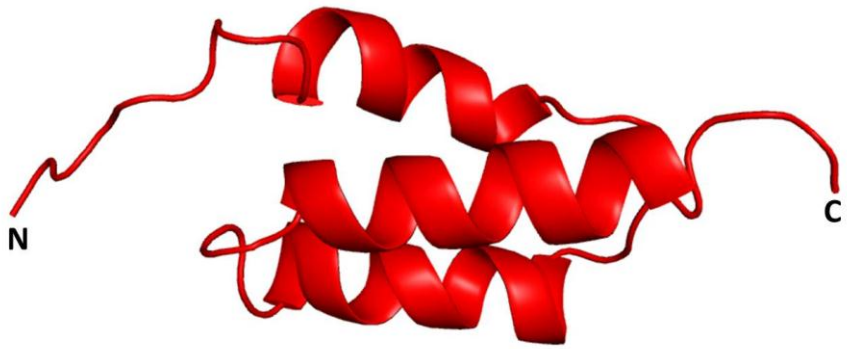

Figure 5.8. Crystal structure of the B domain of protein A. Cartoon representation of the B domain structure of protein A (red, PDB 1BDD, Gouda *et al.*, 1992). N and C termini are labelled. Z is homologous to B and is expected to retain the three helices bundle conformation.

Each of the Z domains is about 7.5 kDa. The three-dimensional structure of the B domain of staphylococcal protein A (Gouda *et al.*, 1992), on which Z is based, is a bundle of three short anti-parallel α-helices (Figure 5.8). Association of a ZZ tag with the LHn fragment of the botulinum neurotoxin was achieved to assess its impact on protein expression and solubility. It may also confer to the fragment further functionality by binding the Fc domain of IgG. The recombinant proteins produced represented interesting molecules to study the structure and stability of LHn when associated with a large peptide. Constructs corresponding to ZZ-LHn serotypes A and D were purified and successfully crystallised. Purification, crystallisation and preliminary analysis are described here.

5.2.2. Methods

Protein expression and purification

The genes encoding LHn/A and LHn/D (Chapter 3) were cloned into modified pET vector (Novagen, UK) along with a synthetic gene corresponding to two Z domains (129 amino acids) at the C-terminal of LHn and with N and C-terminal poly-His tags. The constructs were transformed into *E. coli* BL21 expression cells. The sequence was engineered to encode for an enterokinase cleavage site between LC and Hn. The clone was provided by Syntaxin Ltd. Expression was performed as described previously. Both constructs are approximately 115 kDa. ZZ-LHn/A in the soluble fraction was purified by affinity chromatography (HisTrap HP, GE Healthcare), activated by enterokinase cleavage (New England BioLabs) and a final chromatography step by HisTrap HP (GE Healthcare). Sample was concentrated to 11.3 mg/ml (Figure 5.9B).

ZZ-LHn/D protein was provided by Syntaxin Ltd for crystallisation trials and had been purified following a similar method to ZZ-LHn/A. The sample was at a concentration of 7.3 mg/ml.

Crystallisation

For ZZ-LHn/A, one crystallisation condition was identified in the ProPlex screen (Molecular Dimensions). The condition could not be repeated manually but was obtained twice with the crystallisation robot (Art Robbins). The crystals took over 10 months to grow in 1.4 M sodium malonate pH 6.0. Crystals grew as small prisms. The crystal obtained the first time showed x-ray diffraction at low resolution (7.0 Å) on beamline IO3 (Figure 5.9D) and the second one diffracted to 4.5 Å on the I24 microfocus beamline at the Diamond Light Source, UK. Diffraction images could be indexed in orthorhombic space group I222 using MOSFLM and scaled with SCALA (CCP4, 1994; Leslie, 2006) with dimensions a= 220, b= 245, c= 254 Å, α, β, γ = 90° (Table 5.2).

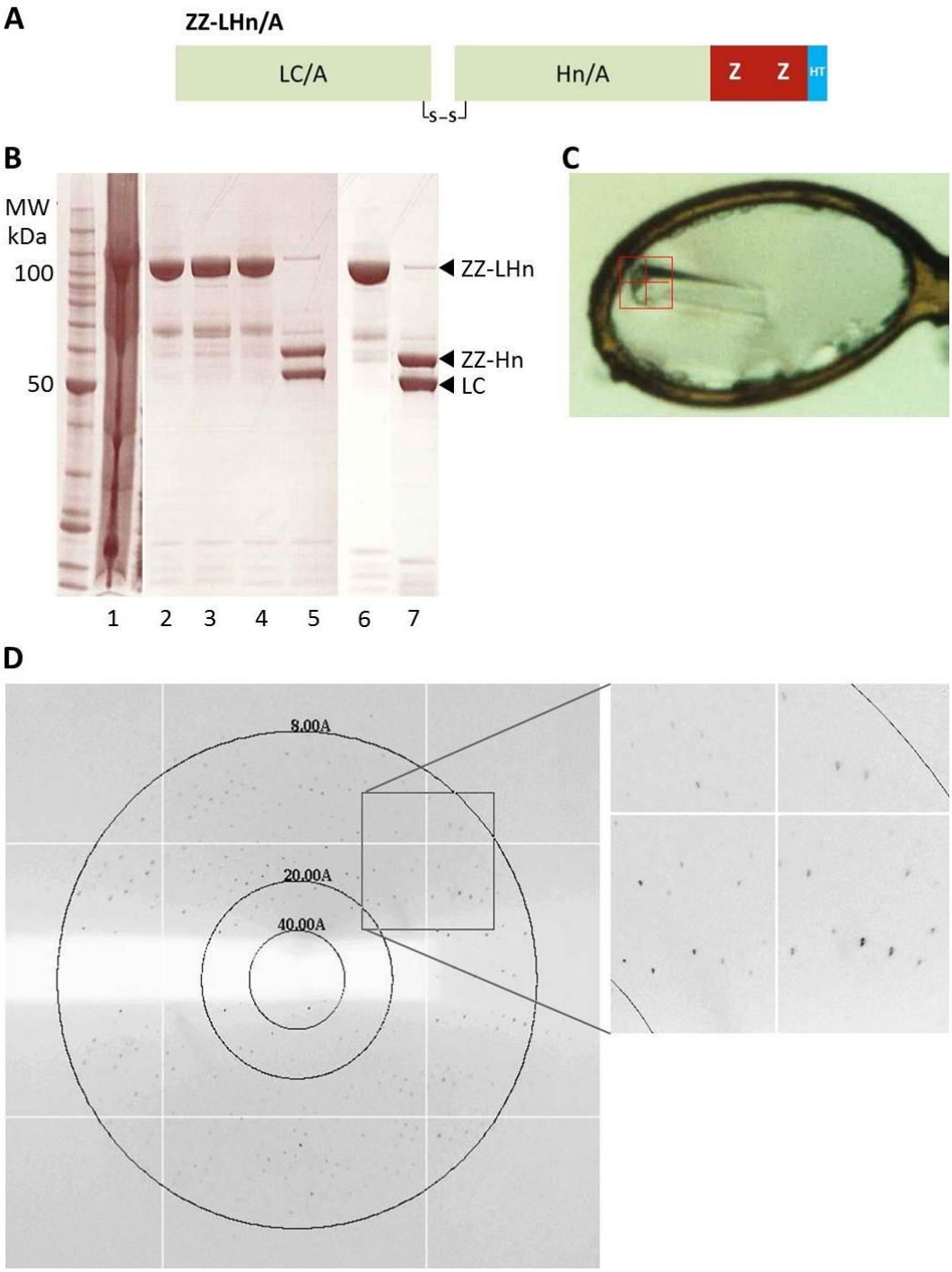

Figure 5.9. Purification and crystallisation of ZZ-LHn/A. (A) ZZ-LHn/A construct. LHn/A (green) Z domain (red), and poly-His tag (light blue). **(B)** Summary of ZZ-LHn/A purification. Lane 1, cell lysate supernatant; 2, 3, eluate from first affinity purification step in oxidised (O) and reduced (R) conditions (respectively); 4, 5, sample activated by enterokinase (O and R); 6, 7, eluate from second affinity purification (O and R). **(C)** Crystals of ZZ-LHn/A mounted for data collection, crystal grew in 1.4 M sodium malonate pH 6.0. (Red square is 167x167 µm). **(D)** Diffraction image collected at DLS IO3 where ZZ-LHn/A crystals diffracted >7.0 Å.

In the case of ZZ-LHn/D, two crystallisation conditions were identified in the JCSG-*plus* screen (Molecular Dimensions): 20% PEG3000, 0.1 M citrate pH 5.5; and 20% PEG6000, 0.1 M sodium citrate pH 5.0 (Figure 5.10A). Manual trials were set up and crystals were grown in citrate buffers (pH 5.0-6.0) with various PEGs (2000 to 6000). However, the best crystals were obtained with the solution identified in the screen and purchased from Molecular Dimensions (20% PEG6000, 0.1 M sodium citrate pH 5.0). Crystals were readily obtained in 2-3 days by the hanging drop method (2 µl protein and 1 µl mother liquor against a 500 µl reservoir at 16°C). The crystals grew in good sizes with an arrowhead shape (Figure 5.10A). Several crystals were tested and all showed limited x-ray diffraction. The best dataset obtained was at 5.0 Å, at the Diamond Light Source, UK, beamline I24 (Figure 5.10B). The crystal was quickly soaked in 25% PEG6000 for cryo-protection. Images could be indexed in a primitive tetragonal space group (likely $P4_12_12$) using MOSFLM and scaled with SCALA (CCP4, 1994; Leslie, 2006) with cell dimensions a= 108, b= 108, c= 478 Å, $\alpha, \beta, \gamma = 90°$. With the large cell dimension, images were collected with a small oscillation range, however data were difficult to process and yielded an unsatisfactory R merge (> 0.18) due to weak diffraction (Table 5.2).

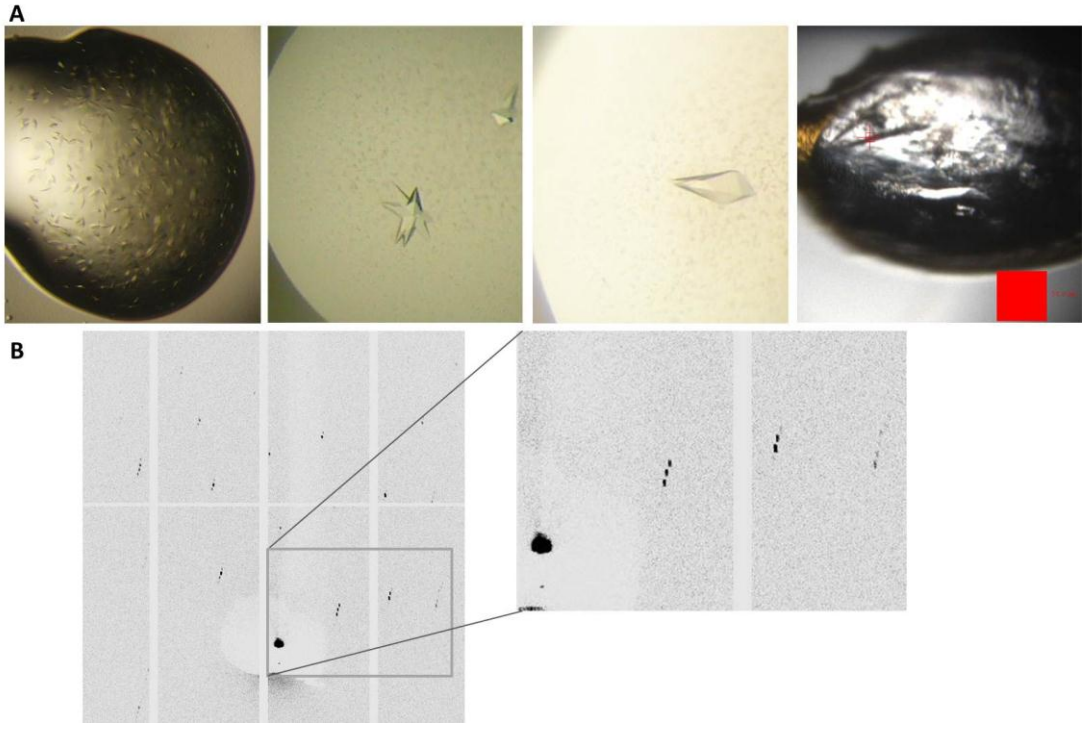

Figure 5.10. Crystal and x-ray diffraction of ZZ-LHn/D. (A) Optimisation of ZZ-LHn/D crystallisation. From **left** to **right** respectively, small arrowhead crystals grew in 20% PEG3000, 0.1 M citrate pH 5.5, single crystal grew in 20% PEG6000, 0.1 M sodium citrate pH 5.0, and crystal of ZZ-LHn/D mounted for data collection at DLS I24 (red square is 55x56 µm). **(B)** Diffraction image collected at DLS I24 where ZZ-LHn/D crystals diffracted at >5.0 Å. Zoom around beam centre.

Table 5.2. ZZ-LHn X-ray data collection

A. Data collection statistics	ZZ-LHn/A	ZZ-LHn/D
Space group	I222	$P4_12_12$
Molecules/asymmetric unit	$(6)^2$	$(2)^2$
Cell dimensions	a=220, b=245, c=254 Å; α, β, γ =90°	a=108, b=108, c=478 Å; α, β, γ =90°
Resolution range (Å)	30-4.5	41-5.0
R_{sym}[1] (%)	14	18
I/σI (outer shell)	10.4 (6.0)	6.6 (5.2)
Completeness (outer shell) %	99.2 (99.9)	99.6 (99.9)
Total no. of reflections	230,662	73,570
Unique no. of reflections	41,036	14,145
Redundancy	5.6 (5.7)	5.6 (5.8)

[1] $R_{sym} = \Sigma_h \Sigma_i |I(h) - I_i(h)| / \Sigma_h \Sigma_i I_i(h)$, where $I_i(h)$ and $I(h)$ are the i^{th} and the mean measurements of the intensity of reflection *h*, respectively.
[2] most likely value based on Matthews coefficient.

5.2.3. Results and discussion

Purification and crystallisation of ZZ-LHn/A

ZZ-LHn/A was successfully purified from *E. coli*. The ZZ tag had been demonstrated to enhance solubility and expression (Moks *et al.*, 1987) and the yield seemed higher than for LHn/A. This allowed a straightforward purification process by Ni^{2+}-chelate affinity chromatography, illustrated by SDS-PAGE analysis in figure 5.9B, leaving only a faint persistent protein band at approximately 65 kDa and low molecular weight residuals. Activation by enterokinase was not entirely complete as demonstrated by the faint band at approximately 115 kDa in reduced conditions. Overall, the sample was considered pure enough for crystallisation trials and could be concentrated up to 7.3 mg/ml without any stability issues. In comparison, LHn/A would not sustain concentration higher than 5 mg/ml (in similar buffer conditions).

Crystals of ZZ-LHn/A were obtained twice in the same condition from preliminary crystallisation screening. These crystals could not be repeated and took more than 10 months to grow. One of the crystals diffracted to 4.5Å and a full dataset was collected on the microfocus beamline (I24) at DLS. Despite a relatively high R merging value (14%), the scaled statistics were acceptable and showing a high completeness (Table 5.2). Since the

crystal structure of all the elements composing ZZ-LHn/A are known and available, molecular replacement was used for phasing. Several trials were set up using Phaser (McCoy *et al.*, 2007) and MolRep (Vagin and Isupov, 2001) with either ZZ domains and LHn/A, LHn/A on its own or LC/A as search models, however all attempts failed to deliver a solution. Further crystallisation trials are on-going in an effort to reproduce the crystals and optimise data collection at better resolution.

Crystallisation of ZZ-LHn/D

Purified ZZ-LHn/D protein was provided by Syntaxin Ltd for crystallisation trials. Crystallisation conditions were identified from the preliminary screens and could be readily reproduced (Figure 5.10A). Single crystals with sharp edges and good size were obtained. Crystals would typically diffract to over 6 Å. Optimisation of the cryo-protectant solution allowed for collection of a full dataset at 5.0 Å at the microfocus beamline (I24, DLS). Analysis of the data showed a primitive tetragonal space group with POINTLESS (CCP4, 1994) indicating $P4_12_12$ as the most likely option. The long cell dimension (c= 478 Å) with very close diffraction spots (Figure 5.10B) made the data difficult to process, particularly at higher resolution. This is illustrated by the high R_{symm} (18%) observed. Data were nonetheless used for molecular replacement. No solution was found when searched for ZZ and LHn, LHn or LC/D. Optimisation of the crystallisation conditions was carried out, including experiments with commercially available additive screens (Hampton Research) using the Phoenix crystallisation robot (Art Robbins Instruments). None of the crystals produced diffracted to better than 5 Å resolution.

Conclusion

The two ZZ-LHn constructs were crystallised and x-ray diffraction data were collected at low resolution. No molecular replacement solution could be found and further crystallisation optimisation will be necessary to obtain higher resolution data. Furthermore, while the ZZ domain may confer the molecules better solubility, its position at the C-terminal end of LHn, in an accessible surface area, may cause some instability in the crystals. Since the Z domain has the capacity to bind the Fc region of IgGs (Nilsson *et al.*, 1987), co-crystallisation in the presence of such fragment may help obtain crystals of a stable complex.

5.3. Recombinant VIP-LHn derivatives

5.3.1. Introduction

The potential of retargeting BoNT activity for therapeutic purposes was demonstrated by coupling the LHn fragment to *Erythrina cristagalli* lectin (Duggan *et al.,* 2002), a ligand capable of specific binding on nociceptive afferent neurons. This novel molecule was proved to inhibit neurotransmission release *in vitro* and *in vivo* thereby offering possibilities for pain relief. Consequently, BoNT activity was demonstrated on neuronal cells different from its natural target. It is therefore of interest to consider other ligands with which to associate LHn, so as to target neuronal populations of therapeutic relevance.

Vasoactive intestinal peptide (VIP) is a 28-residue neuropeptide which is part of the glucagon/secretin superfamily and acts through interactions with VPAC (1 and 2), class II G protein-coupled receptors (Laburthe *et al.,* 2007). VIP is widely present in the central and peripheral nervous systems. It is involved in a diverse range of pathological processes related to development, growth, and the control of neuronal and endocrine cells. Its role has also been highlighted in major physiological mechanisms such as the immune response and digestive, respiratory, and cardiovascular systems (Dickson and Finlayson, 2009). Several studies have highlighted the therapeutic potential of VIP and VIP agonist, for example in the treatment of inflammatory diseases (Smalley *et al.,* 2009).

The solution structure of VIP was solved by nuclear magnetic resonance (NMR; Tan *et al.,* 2006; Umetsu *et al.,* 2011) and reported to form a long α-helix and an unstructured N-terminal region (Figure 5.11). VIP structural features are similar to that of other ligands from the same family (Laburthe *et al.,* 2007) and modelling with the VPAC receptor was reported based on homologous interactions (Tan *et al.,* 2006). It was predicted that VIP would bind to the N-terminal extracellular domain of VPAC.

Figure 5.11. NMR structure of VIP. Cartoon representation of VIP (blue, PDB 2RRH, Umetsu *et al.,* 2011). N and C termini are labelled.

A novel molecule with therapeutic potential, VIP-LHn/A, was cloned, expressed and purified for structural analysis by x-ray crystallography. Information on the structure would help understand the interface between ligand and LHn. Crystallisation and analysis of x-ray diffraction data are presented. Furthermore, the VPAC1 extracellular domain was produced with the aim of studying the fusion molecule interaction with its receptor.

5.3.2. Methods

Cloning, expression and purification of VIP-LHn/A

Clones were provided by Syntaxin Ltd for LHn/A (Chapter 3) and VIP in modified pET vectors (Novagen, UK). Both genes had been obtained synthetically and were included in a cloning cassette. By digestion of the DNA vectors with appropriate restriction enzymes followed by ligation, the two genes were fused on a single vector so that VIP would be located at the C-terminal end of LHn/A in the same open reading frame. The construct also included a C-terminal poly-His tag and a Factor Xa cleavage site engineered between LC and Hn. The plasmid was transformed into *E. coli* BL21 expression cells. Expression was performed as described previously for LHn/A. The construct was approximately 100 kDa. Soluble VIP-LHn/A was purified by affinity chromatography (HisTrap HP, GE Healthcare), activated by Factor Xa cleavage (New England BioLabs) and a final hydrophobic interaction purification step (Phenyl-650M, Toyopearl). Sample was concentrated to 4.5 mg/ml.

Crystallisation of VIP-LHn/A

VIP-LHn/A crystallisation trials were set up manually by the hanging drop method with the conditions that had produced LHn/A crystals (i.e. 15% sucrose, 0.1 M TRIS Acetate pH 8.5, 1.5 M ammonium sulphate). Crystals grew in 3-4 days at 16°C with a similar morphology to the LHn/A ones (thin rod-shaped crystals; Figure 5.12B).

X-ray diffraction data were collected from a single crystal at the Diamond Light Source, UK, beamline IO2 (Figure 5.12A). The crystal diffracted to 3.25 Å. Analysis of the diffraction images using MOSFLM (CCP4, 1994; Leslie, 2006) allowed indexing with a cell similar to LHn/A, in the orthorhombic space group $P2_12_12_1$. The dimensions were a= 78.4, b= 156.1, c= 209.7 Å, $\alpha, \beta, \gamma = 90°$. Processing of the data appeared difficult since two lattices may have been present in the crystal. Scaling was done with SCALA (CCP4, 1994) and the resulting statistics are presented in table 5.3.

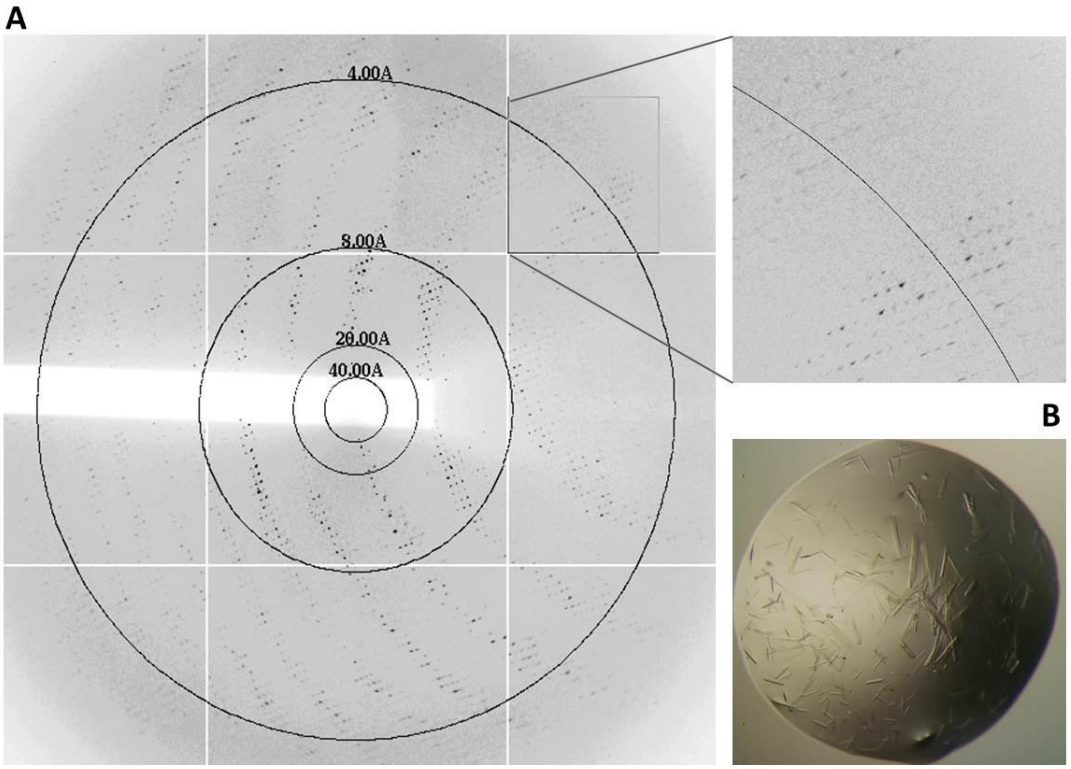

Figure 5.12. Crystal and x-ray diffraction of VIP-LHn/A. (A) Diffraction image collected at DLS IO2 where VIP-LHn/A crystal diffracted at 3.25 Å. **(B)** VIP-LHn/A crystallisation. Thin rode-shaped crystals grew in 15% sucrose, 0.1 M TRIS Acetate pH 8.5, 1.5 M ammonium sulphate.

Table 5.3. VIP-LHn/A x-ray data collection.

A. Data collection statistics	
Space group	$P2_12_12_1$
Molecules/asymmetric unit	2
Cell dimensions	a=78.4, b=156.1, c=209.7 Å; α, β, γ =90°
Resolution range (Å)	80-3.25
R_{sym}[1] (%)	16.8
I/σI (outer shell)	7.2 (2.1)
Completeness (outer shell) %	92.6 (93.3)
Total no. of reflections	111,165
Unique no. of reflections	38,096
Redundancy	2.9 (3.0)

[1] $R_{sym} = \Sigma_h \Sigma_i |I(h) - I_i(h)| / \Sigma_h \Sigma_i I_i(h)$, where $I_i(h)$ and $I(h)$ are the i^{th} and the mean measurements of the intensity of reflection *h*, respectively.

Western-blot analysis of VIP-LHn/A

The purified VIP-LHn sample was run on SDS-PAGE and blotted on a PVDF membrane (Millipore). Primary detection was performed using an anti-VIP antibody targeting the C-terminal epitope (#Sc7841, Santa Cruz Biotechnology). Secondary recognition was performed with a horseradish peroxidase-conjugated anti-goat antibody (Sigma). Enhanced chemiluminescent (ECL) substrate reagent (Thermo Scientific) was used for revelation.

Cloning, expression and purification of VIP receptor extracellular domain (VPAC1 ECD)

Couvineau *et al.,* (2008) achieved the production of recombinant VPAC1 ECD from an *E. coli* expression system and the resulting protein was able to bind VIP. A similar system was therefore used for the production of VPAC1 ECD suitable for crystallographic study. The sequence for residues 31-142 of VPAC1 (UniProt P32241) was back-translated and the corresponding gene optimised for *E. coli* codon usage (Entelechon tool). Restrictions sites were added to the 5' (*Nde*I and *EcoR*I) and 3' (*Hind*III) ends for sub-cloning. A C-terminal poly-His tag was also included. The designed synthetic gene was purchased (GeneArt) and cloned into modified pMal (New England BioLabs) and pGEX vectors (GE LifeSciences). These vectors allowed for inclusion of cleavable (by Factor Xa) N-terminal MBP and GST tags respectively (Figure 5.13A). The constructs were transformed into *E. coli* Origami expression cells (Novagen).

Expression was carried out as described previously and soluble VPAC1 ECD was observed with both constructs (results not shown). Cells were resuspended in 25 mM TRIS pH7.2, 200 mM NaCl (Buffer A). Purification was performed with the MBP-VPAC1 ECD protein and consisted firstly of an affinity chromatography step using an amylose resin (New England BioLabs). MBP-VPAC1 ECD was eluted with 10 mM maltose (in buffer A). The resulting sample was treated by Factor Xa for removal of the N-terminal MBP tag and dialysed against buffer A. This was followed by another amylose purification. The flow-through was collected and loaded on a HisTrap HP column (GE Healthcare). VPAC1 ECD was eluted in 250 mM imidazole (in buffer A). SDS-PAGE analysis of the purification is presented in figure 5.13B.

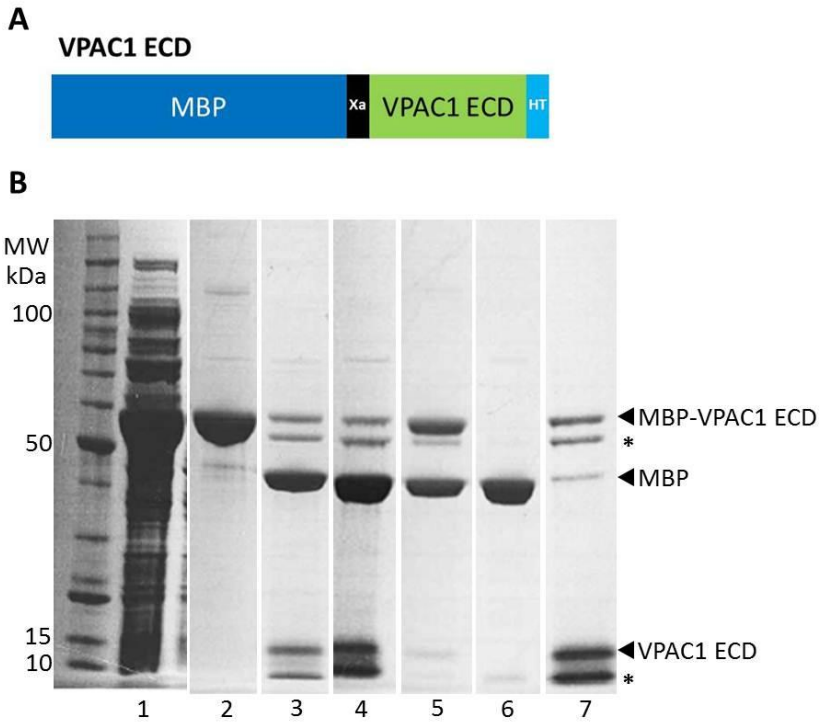

Figure 5.13. SDS-PAGE analysis of VPAC1 ECD purification. (A) VPAC1 ECD construct. MBP-tag (dark blue), Factor Xa cleavage site (black), VPAC1 ECD (green) and poly-His tag (light blue). **(B)** Summary of VPAC1 ECD purification. Lane 1, cell lysate supernatant; 2, eluate from first amylose purification step; 3, sample treated by Factor Xa; 4, flow-through from second amylose purification; 5, eluate from second amylose purification; 6, flow-through from HisTrap purification; 7, eluate from HisTrap purification. Asterisks mark unspecific cleavage products.

5.3.3. Results and discussion

Production and crystallisation of VIP-LHn/A

Recombinant VIP-LHn/A was successfully cloned and expressed in *E. coli*. The purification was optimised from the LHn/A process (Chapter 3) and consisted of two chromatography steps using the engineered poly-His tag and the protein's hydrophobic properties. VIP-LHn/A was completely activated by treatment with Factor Xa protease. This is illustrated by SDS-PAGE analysis of the purified di-chain sample presented in figure 5.14B. Furthermore, integrity of the molecule was investigated by Western blot. Detection with a specific anti-VIP antibody highlighted the presence of the peptide at the C-terminal end of the protein where it is associated to the translocation domain (upper band at approximately 50 kDa in the reduced sample). Although the yield obtained was lesser than for previous LHn molecules, the level of purity achieved (>95%) was suitable for crystallisation.

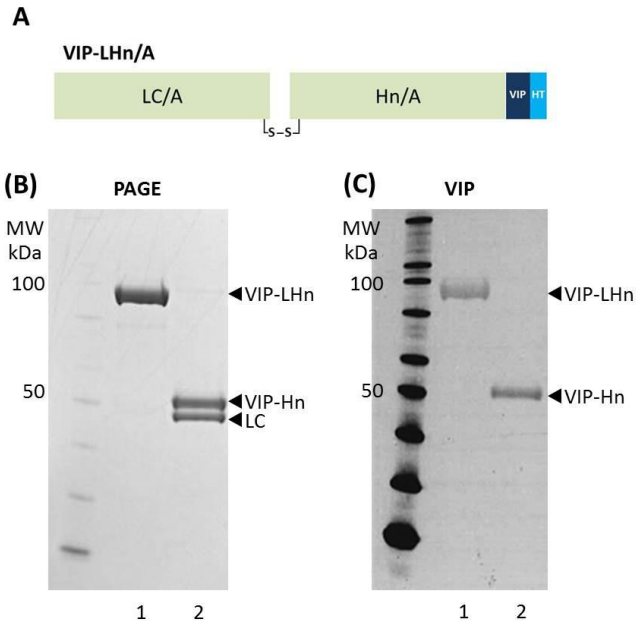

Figure 5.14. SDS PAGE and western blot of purified VIP-LHn/A. (A) VIP-LHn/A construct. LHn/A (green) VIP (dark blue), and poly-His tag (light blue). **(B)** SDS PAGE and **(C)** Anti-VIP Western blot; Lanes 1 and 2, VIP-LHn/A in oxidised and reduced conditions respectively.

The difference between LHn/A and VIP-LHn/A is the engineered VIP peptide corresponding to less than thirty residues. Since similar purification processes were used, it was not surprising to observe VIP-LHn/A crystals with the conditions identified for its LHn counterpart. Additionally, the crystals obtained had similar morphology to LHn/A as well as an identical crystallographic cell when tested at the synchrotron light source (Table 5.3). Processing of the 3.25 Å x-ray diffraction images was difficult and this is reflected by the high merging value (17%). A close inspection of the images (Figure 5.12) indicated the presence of two possible lattices in the crystal.

The data were nevertheless used for molecular replacement using LHn/A as a search model. While a solution was found with two LHn/A molecules per asymmetric unit (as seen in LHn/A, Masuyer *et al.*, 2009), initial refinement was not conclusive (R factor > 38%). This may be the result of the experimental at data low resolution. However, inspection of the unrefined model did not indicate any Fourier difference density at the C-terminal end of the translocation domain, and thus failed to show the VIP peptide. This part of the protein is solvent accessible, and the C-terminal region of the LHn/A structure was already observed to be missing (including 6 amino acids and the poly-His tag). VIP is expected to present with

an α-helical structure. Surface exposure to the solvent and flexibility of the peptide might explain the disorder in this region of the structure. Optimisation of the crystallisation and data collection could improve interpretation of these results. An alternative method to determine the structure of the VIP fusion protein may rely in stabilising the C-terminal region. This may be achieved by co-crystallisation of VIP-LHn/A with a molecule binding specifically to the peptide.

Production of VIP receptor extracellular domain (VPAC1 ECD)

VIP interacts with VPAC (1 and 2) class II G protein-coupled receptors by binding to their N-terminal extracellular domain (Laburthe *et al.,* 2007). A method for expression of recombinant VPAC1 extracellular domain in *E. coli* was reported (Couvineau *et al.,* 2008). A similar system was used in this study in which VPAC1 ECD was associated with a cleavable N-terminal MBP tag. Such protein fusion partner was added to improve protein solubility. Indeed, VPAC1 ECD contains three disulphide bridges and is prone to precipitation. To further address this issue, the *E. coli* Origami cell line (Novagen) was chosen for expression. These cells were engineered to promote disulphide bond formation in the host cytoplasm and thus supporting correct protein folding.

VPAC1 ECD was successfully cloned and expressed in soluble form. The expression yield was satisfactory and pure MBP-tagged material could be recovered by amylose affinity chromatography (Figure 5.13B). Removal of the MBP tag is however necessary as VIP is known to bind the N-terminal region of VPAC1 (Laburthe *et al.,* 2007). The construct included a Factor Xa cleavage site between the tag and the protein of interest. Treatment with the protease was carried out and led to cleavage of the MBP tag. SDS-PAGE analysis of the treated sample showed that tag removal was not complete (MBP-VPAC1 ECD at approximately 56 kDa). Along with the expected cleavage products (MBP at 43 kDa, and VPAC1 ECD at 13 kDa), the sample presented two unidentified protein bands at approximately 50 and below 10 kDa. These were likely the results of unspecific proteolysis by Factor Xa. N-terminal sequencing (AltaBioscience) was performed on the lower band and the residues identified were part of the VPAC1 ECD sequence corresponding to a fragment of 49 amino acids at the VPAC1 ECD carboxyl end. No potential protease sites were identified when verifying the protein sequence. Furthermore, the final purification step by Ni^{2+}-chelate affinity was efficient in removing most of the cleaved MBP but overall, intact and pure VPAC1 ECD could not be extracted. Further work is required to purify the VPAC1 ectodomain which would include optimisation of the purification process and trials with alternative proteases for efficient tag removal.

Conclusion

Cloning, purification and crystallisation of a novel VIP-LHn/A fusion protein were achieved. Crystallographic analysis did not allow the observation of the VIP fusion partner in the structure. The target receptor for this molecule, VPAC1 extracellular domain, was cloned and expressed in a soluble form. Further optimisation of the purification will be necessary to use it in crystallographic experiments.

5.4. Recombinant EGF-LHn derivatives

5.4.1. Introduction

Inhibition of neurotransmission by retargeting the botulinum neurotoxin has been established (Duggan et *al.*, 2002). Exocytosis is a universal mechanism of secretion mediated by SNARE proteins in eukaryotic cells (Bonifacino and Glick, 2004). Thus there is a potential to retarget BoNT not only on neuronal cells but on all other cell types, where inhibition of secretory events could be therapeutically useful. This broader principle was demonstrated by Foster *et al.* (2006). In their study, a recombinant molecule consisting of LHn/C in combination with the epidermal growth factor (EGF) domain was shown to inhibit mucus secretion in a respiratory epithelial cell line. Such molecule could have applicability in the treatment of chronic respiratory pathologies such as asthma.

The work by Foster *et al.* (2006) also represented the first report of a recombinant liganded-LHn molecule produced in *E. coli*. This molecule was engineered with a Factor Xa recognition site between LC and Hn for activation, and a cleavable C-terminal MBP (maltose binding protein) tag. Catalytic activity was demonstrated by cleavage of syntaxin which led to inhibition of mucin secretion in epithelial cells.

Human epidermal growth factor (hEGF) is a polypeptide of 53 amino acids in length with three internal disulphide bridges. Its crystal structure in a dimeric state has been reported (Lu *et al.*, 2001). The EGF structure is composed of two domains (Figure 5.15). The N-domain (residues 1–32) has an unstructured N-terminal segment and an antiparallel β-sheet. The C-domain (residues 33–53) consists of a short antiparallel β-sheet and a flexible C-terminal region.

With knowledge of the functional potential of EGF-LHn molecules, several constructs corresponding to serotypes A, C and D were expressed and purified in order to study their molecular structures. Such structural information would help understand the stability of the EGF ligand in fusion with LHn, and the interaction between the two partners. The three proteins were used in crystallisation trials and preliminary crystallisation and x-ray diffraction results are presented below.

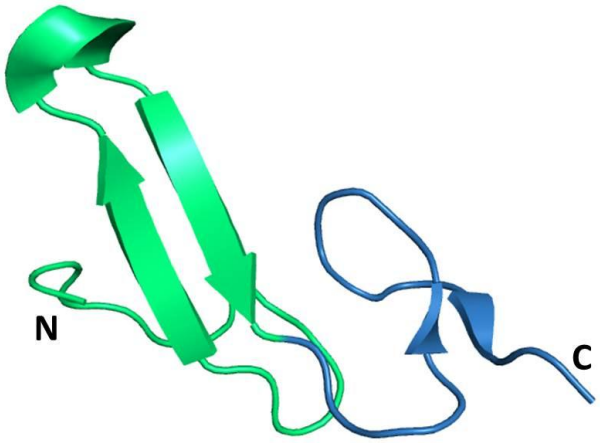

Figure 5.15. Crystal structure of human EGF. Cartoon representation of hEGF crystal structure, N-terminal domain (green), C-terminal domain (blue) (PDB 1JL9, Lu *et al.*, 2001). N and C termini are labelled.

5.4.2. Methods

Protein expression and purification

The genes encoding LHn/A, /C and /D (Chapters 3 and 5.1) were cloned into modified pET vector (Novagen, UK) along with a synthetic gene corresponding to EGF at the C-terminal of LHn (71 residues). EGF-LHn/A and /C have a C-terminal 10 x His-tag, while /D has a N-terminal one. The constructs were transformed into *E. coli* BL21 expression cells. The sequences were engineered to encode for an enterokinase cleavage site between LC and Hn. The clones were provided by Syntaxin Ltd. Expression was performed as described previously. The three preparations were performed following a similar method. The purification process consisted of affinity chromatography using a Ni^{2+}-charged chelating sepharose column (GE Healthcare), activation by enterokinase cleavage (New England BioLabs), second affinity chromatography (HisTrap HP, GE Healthcare) and then a hydrophobic interaction step (Phenyl Sepharose HP, GE Healthcare). Only EGF-LHn/A was further purified by size exclusion (Superdex 200, GE Healthcare). Samples were concentrated to 2.5 mg/ml (EGF-LHn/A), 2 mg/ml (EGF-LHn/C) and 2.9 mg/ml (EGF-LHn/D). A summary of EGF-LHn/C purification by SDS-PAGE analysis is presented in figure 5.16.

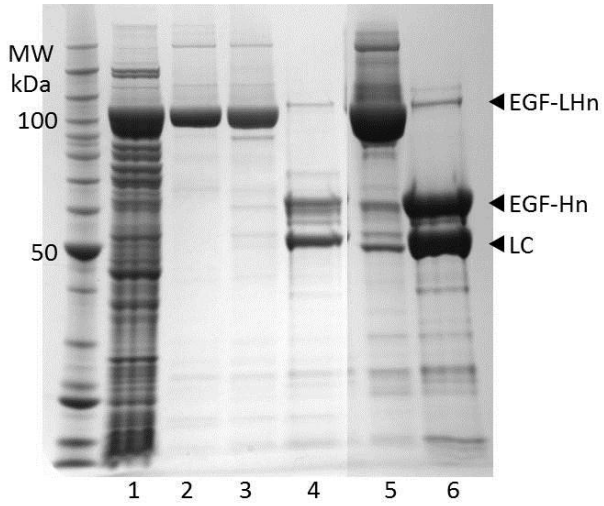

Figure 5.16. SDS-PAGE analysis of EGF-LHn/C purification. Summary of EGF-LHn/C purification. Lane 1, cell lysate supernatant; 2, eluate from first affinity purification step; 3, 4, sample activated by enterokinase (oxidised (O) and reduced (R) respectively); 5, 6, eluate from second affinity purification (O and R).

Western-blot analysis

The three purified EGF-LHn samples were run on SDS-PAGE and blotted on a PVDF membrane (Millipore). Primary detection was performed using an anti-EGF antibody (#Sc275, Santa Cruz Biotechnology), and a Tetra-His (QIAGEN) antibody was used for the tag detection. For secondary recognition, horseradish peroxidase-conjugated anti-rabbit and anti-mouse antibodies (respectively) were applied (Sigma). Enhanced chemiluminescent (ECL) substrate reagent (Thermo Scientific) was utilised for revelation.

Crystallisation

For EGF-LHn/A, automated crystallisation screens did not yield any hits. Manual crystallisation trials were set up in conditions that produced the LHn/A crystals (15% glycerol, 0.1 M TRIS pH 8.5, 1.5M ammonium sulphate), performing cross-streak seeding with LHn/A crystals. A few needle-like crystals were obtained after several weeks of incubation. The crystals were too small to be tested for diffraction.

In the case of EGF-LHn/C, several conditions were identified in the PACT *premier*, Heavy + Light Twin Pack and JCSG-*plus* screens (Molecular Dimensions). Manual crystallisation trials were set up in order to optimise these conditions. Most of the crystals obtained were stack of thin needles. The best looking crystals from the first screenings (with 15% glycerol, 0.1 M HEPES pH 7.5, 1.0 M lithium sulphate) were used for streak-seeding. Rod-shaped crystals were obtained manually in 1-2 weeks by the hanging drop method (2 μl

protein and 1 µl mother liquor against a 500 µl reservoir at 16°C; Figure 5.17B). Several crystals were tested in the x-ray beam and all showed diffraction. The best diffraction obtained was at 6.0 Å, at the Diamond Light Source, UK, beamline IO2. A couple of images were collected but could not be indexed precisely (Figure 5.17A). Using MOSFLM (CCP4, 1994; Leslie, 2006) a possible cell with dimensions a= 125, b= 278, c= 300 Å in P1 space group was obtained.

Concerning EGF- LHn/D, a single crystallisation condition was identified in the JCSG-*plus* screen (Molecular Dimensions) corresponding to 0.1 M CHES pH 9.5, 20% PEG8000. A single small and thin rod-shaped crystal was observed after several months of incubation. The condition could not be repeated despite setting up optimisation trials. The crystal was too small to be tested for diffraction and used for seeding in further crystallisation experiments.

Cloning of EGF receptor extracellular domain (EGFR ECD)

Kim *et al.* (2006) described the engineering, through directed evolution and yeast surface display, of the extracellular domain from the epidermal growth factor receptor (EGFR ECD) for optimal expression in *Saccharomyces cerevisiae*. The protein sequence (621 residues) made available was back-translated and the nucleotide sequence optimised for codon usage in yeast using the Entelechon server. The corresponding synthetic gene was purchased (GeneArt) which also included a C-terminal poly-His tag (x6) and the restriction sites (5' *Eco*RI, and 3'*Not*I) require for cloning into a pPICZα vector (Invitrogen). Cloning into pPICZα was achieved and checked by sequencing. Transformation into *Pichia pastoris* expression cells is under investigation.

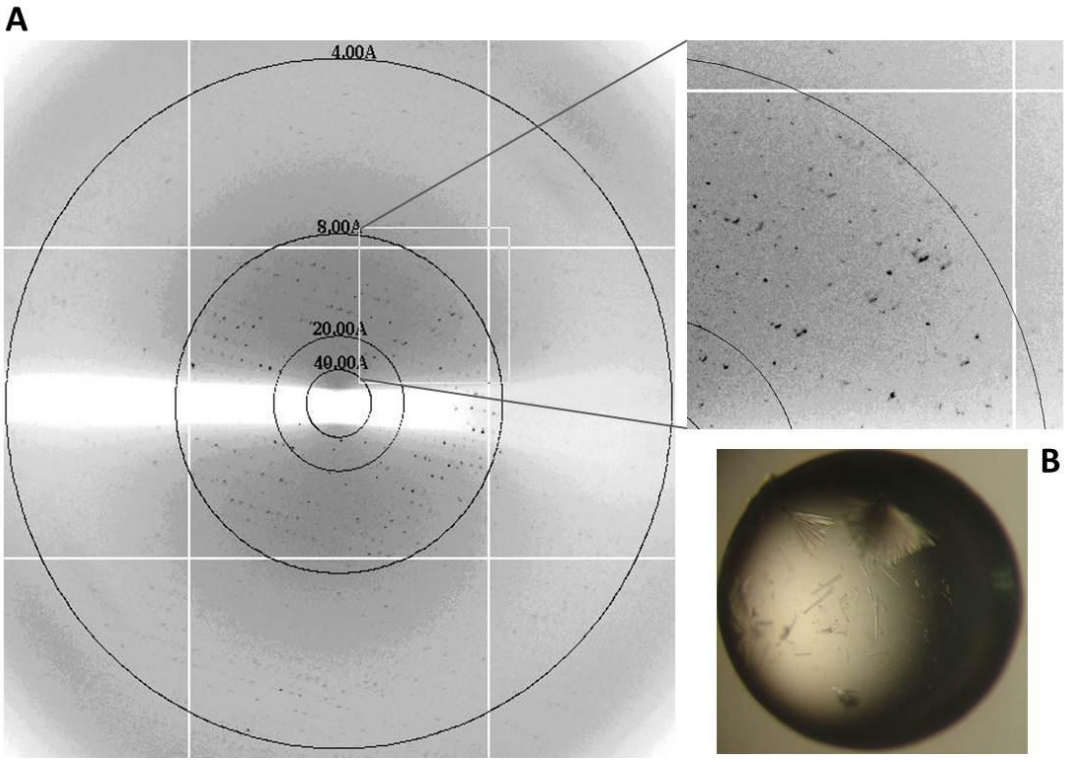

Figure 5.17. Crystal and x-ray diffraction of EGF-LHn/C. (A) Diffraction image collected at DLS IO2 where EGF-LHn/C crystals diffracted at >6.0 Å. **(B)** EGF-LHn/C crystallisation. Needles grew in 15% glycerol, 0.1 M HEPES pH 7.5, 1.0 M lithium sulphate.

5.4.3. Results and discussion

Purification of EGF-LHn

EGF-LHn fusion proteins of serotypes A, C and D were purified using a different method to the one previously reported for EGF-LHn/C (Foster *et al.*, 2006). Analysis by SDS-PAGE and Western blot (Figure 5.18) of the purified samples confirmed the presence of the EGF peptide. Furthermore, although activation was overall successful for the three proteins, the band at approximately 105 kDa in the reduced samples indicated a low level of single-chained material remained in EGF-LHn/A and /C. Other minor residual proteins were visible, with a persistent protein at 25 kDa in EGF-LHn/A and /C. The anti-EGF and anti-His Western blots highlighted several bands in EGF-LHn/A and /C that may be the results of truncated material, however some of the detection may be due to unspecific cross-reaction with the primary antibodies used. EGF-LHn still represented the main component of the purified samples which were used for crystallisation trials.

Crystallisation of EGF-LHn

Crystals were obtained for the three constructs prepared. EGF-LHn/A and /D only seldom crystallised in conditions that were not reproducible. Efforts were focused on EGF-LHn/C for which several crystallisation conditions were observed and replicated, even with different purification batches. Noticeably, the crystal morphology was very similar to the one observed for LHn/C crystals, i.e. thin needle clusters (Figure 5.4). After optimisation and seeding, larger crystals were obtained (Figure 5.17B). These crystals were tested for x-ray diffraction but they could not withstand the damage inflicted by x-ray radiation. Only a couple of images were obtained and spots were obtained only at low resolution (6 Å). The same crystallisation conditions also yielded thin plate-like crystals once, however these crystals also only showed weak diffraction to >8 Å. Further optimisation of the crystallisation is necessary to acquire data suitable for structure determination.

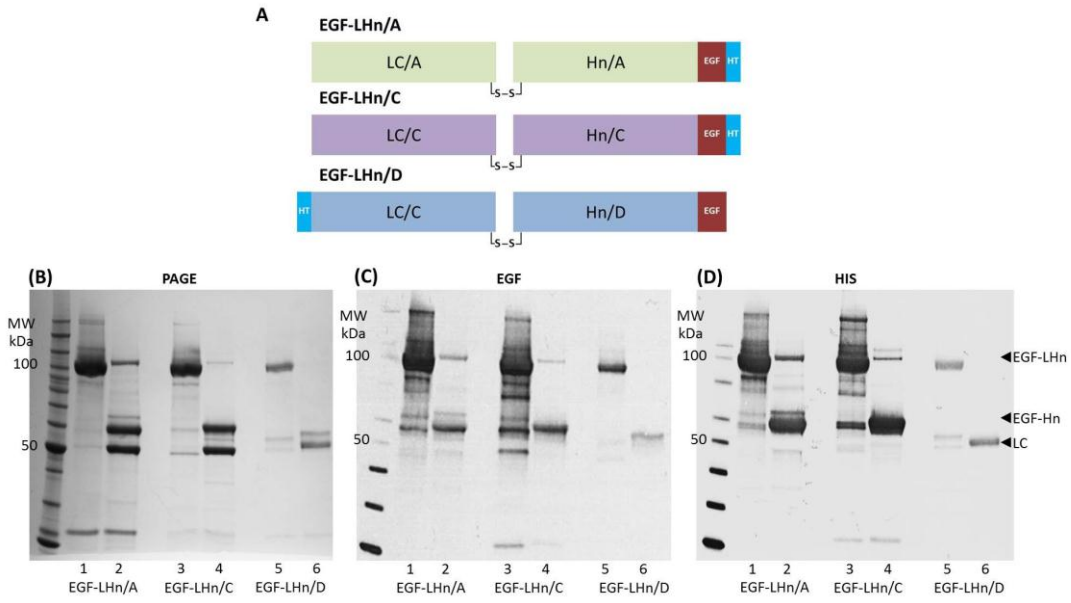

Figure 5.18. SDS PAGE and western blot of purified EGF-LHn proteins. (A) EGF-LHn constructs. LHn/A (green), LHn/C (purple), LHn/D (blue), EGF (red), and poly-His tag (light blue). **(B)** SDS PAGE, **(C)** Anti-EGF and **(D)** Anti-His Western blot; Lanes 1 and 2, EGF-LHn/A; 3 and 4, EGF-LHn/C; 5 and 6, EGF-LHn/D, with each sample in oxidised and reduced conditions respectively.

Towards the production of EGFR extracellular domain

Epidermal growth factor (EGF) was shown to regulate cell proliferation and differentiation by binding to the EGF receptor extracellular domain (EGFR ECD; Carpenter and Cohen, 1979). An alternative way to approach the structure determination of the EGF-liganded molecules would be to investigate their interaction with EGF receptor partner.

Crystallisation of EGF-LHn in complex with EGFR ECD may provide stability to the EGF domain (and hence crystals suitable for structural study), and also deliver valuable information regarding the retargeted toxin-receptor interaction. The structure of EGF in complex with the EGFR ECD has been reported by Ogiso *et al.* (2002). Details of the ligand-receptor interface and comparison with previously known structures may prove useful in optimising future EGF-LHn fusion molecules. With this knowledge in mind, production of EGFR ECD was considered. Based on the data available in literature, a yeast expression system was preferred (Kim *et al.*, 2006) in order to allow for post-translational modification and extracellular secretion. Furthermore the yield of expression in *P. pastoris* should provide enough material for crystallisation studies. While the gene was successfully cloned into the appropriate vector, transformation and expression trials are under investigation.

Conclusion

Three recombinant EGF-LHn fusion proteins were successfully purified and crystallised. X-ray diffraction was shown for EGF-LHn/C. Further optimisation of the crystallisation would be necessary to lead to structure determination. Purification and co-crystallisation with the protein's receptor was investigated and EGFR ECD was cloned into a yeast vector for future recombinant expression in *P. pastoris*.

Chapter 6. General discussion

The *Clostridium botulinum* neurotoxin is a fascinating protein with the potency to inhibit synaptic transmission of cholinergic motor neurons, making it both the most fearsome of poisons and yet also an attractive therapeutic molecule. Structural studies by x-ray crystallography have been fundamental in understanding the toxin's complex mechanism of action (Montal, 2010). Essentially, there is evidence for a modular structure of BoNT in which each domain is responsible for one of the key functions: cell binding, intracellular transport and proteolytic activity (Swaminathan, 2011). This unique modality, along with the universality of the SNARE-mediated secretion process, has led to the development of novel BoNT-derived molecules capable of inhibiting the secretion of a specific cell type (Foster, 2009). The common features of these molecules consist of the catalytic (LC) and translocation domains (Hn) of BoNT, a fragment that was first isolated by trypsinisation of the holotoxins (Shone *et al.,* 1985) and referred to as LHn. It was therefore essential to assess the structure and biological functions of LHn in order to provide the basis for future protein engineering.

The production of the LHn fragments from five of the seven serotypes of BoNT has been achieved by recombinant expression in *E. coli*. This was carried out using an optimised protocol following the work by Chaddock *et al.* (2002). The constructs were efficiently purified with the help of affinity chromatography techniques following the insertion of peptide tags in the constructs. Furthermore, the design of the constructs included a specific exoprotease cleavage site between the two domains, allowing for an effective and controlled activation of the molecules in their di-chain form. High levels of purity were obtained and allowed for structural studies of these proteins by x-ray crystallography. Crystals were obtained for all the LHn fragments after extensive crystallisation experiments using automated and manual methods.

The x-ray crystal structure of LHn/A at a resolution of 2.6 Å was first obtained. Interestingly it presented a conformation identical to its parent neurotoxin, with the exception of the deleted binding domain (Hc). The structural features common to other BoNT were observed with the globular fold of the catalytic light chain, including the tetrahedral coordination of the zinc ion, as well as the atypical coiled-coil fold formed by the long α-helices of the translocation domain. Additionally the structure of LHn was particularly useful to study the interactions between LC and Hn and demonstrated the activation of LHn/A into a di-chain form and the capacity of the belt region to stabilise

around LC. The belt region is believed to play an essential role in the translocation mechanism but is not as yet fully understood, thus LHn may be a beneficial tool in investigating parts of the toxin's intoxication processes. Secondly, the crystal structure of LHn/B was determined at 2.8 Å. Analysis of this structure led to similar observations to LHn/A in that it showed a stable molecule retaining the structural characteristics of BoNT/B lacking the Hc domain with minor differences observed in flexible loop regions. Furthermore, crystals were obtained for serotypes C, D and E although these studies did not lead to the determination of their structures. Interesting data were collected for LHn/D at 2.2 Å but a thorough analysis showed only the truncated catalytic domain had crystallised and thus provided limited information since the structure of LC/D at high resolution (1.6 Å) had been previously reported (Arndt *et al.,* 2006). An alternative approach using small-angle x-ray scattering showed that LHn/D is likely have a similar fold to the other LHn. Although further experiments are required to confirm this result, the solution structure of LHn (from serotypes A, B and D) correlated well with the crystal structures reported here.

The botulinum neurotoxins cause inhibition of neurotransmission by cleaving one of the SNARE proteins. The biological activity of the LHn fragment could therefore be assessed by testing their proteolytic activity *in vitro*. When LHn/A was first characterised Chaddock *et al*. (2002) demonstrated it retained the catalytic properties of BoNT/A towards SNAP-25. Here, a study of the stability of LHn/A showed the molecule was active over several weeks and thus confirmed its catalytic property on a recombinant SNAP-25 substrate. Additionally, the catalytic properties of LHn/B towards several VAMP substrates were determined following a similar assay. It demonstrated a preference of LHn/B for the neurotransmission-associated VAMP-1 and -2 over VAMP-3. However, the activity of the botulinum neurotoxins is dependent on their ability to reach their substrate intracellularly. Recent studies showed LHn/A is a sufficient molecule to allow pore formation and transport of LC inside the cells (Fischer *et al.,* 2008), as well as inhibiting neurotransmission release in cultured cells, albeit at high concentrations (Chaddock *et al.,* 2002). The results presented here showed LHn/B was able to cleave intracellular VAMP-1 and -2 in spinal cord neurons in a dose-dependent manner, although with less potency than the full length toxin. The process by which LHn gains entry into neuronal cells is unknown but certainly relies on the pore-forming capabilities of Hn and the high levels of proteins applied to the cells in this assay.

Characterisation of the structure and activity of the LHn from serotypes A and B have demonstrated the stability and functionality of the BoNT fragment. It has validated the relevance of using LHn as a safe and reliable tool to study the mechanism of action of the botulinum neurotoxins (Fisher *et al.,* 2008; Mushrush *et al.,* 2011). On another hand, the low

toxicity of LHn and its lack of a specific cell binding feature make it an interesting tool to study SNARE-mediated cellular mechanisms. However further work is required to determine the effect of LHn alone on non-neuronal cells. Importantly, the LHn structure provides a stable framework for protein engineering and the development of novel pharmaceutical molecules.

The production of LHn-derived molecules in their active di-chain form is particularly challenging although the use of specific exoprotease sites by recombinant techniques has been developed efficiently (Chaddock *et al.*, 2002). In this study, an alternative approach was investigated and consisted in using the BoNT's own proteolytic activity. Several constructs were produced in which a SNARE peptide region, substrate of BoNT, was inserted between the LC and Hn domains. These molecules were biochemically characterised by SDS-PAGE and Western blotting, and successfully demonstrated the production of integral di-chain LHn backbones and thus the principle of LHn activation by LC. The activity of these molecules was verified *in vitro* by substrate cleavage and spinal cord neuron assays. Although the LC/A-SNAP25-Hn/A construct was less efficient in the cell-free SNAP-25 cleavage assay, its activity in neurons was consistent with LHn/A on its own.

The crystal structures of two molecules produced for the investigation of SNARE-mediated activation, namely LC/A-SNAP23-HN/A and LC/A(0)-SNAP25-Hn/A, were determined to 3.0 and 2.7 Å respectively. These molecules proved useful in analysing the structure of LHn backbones fusioned to functional polypeptides. Interestingly, it demonstrated again the stability of the LHn fold, being stabilised by the strong interactions between LC and Hn. Unfortunately, the additional SNARE peptide located at the interface between the two domains could not be observed in these structures. This was likely due to the flexible nature of the SNARE proteins and their localisation in solvent accessible pockets. Importantly, they did not seem to interact with LHn, thus not interfering with its activity.

One of the limitations of using x-ray protein crystallography was not being able to observe some flexible regions in the structures that were solved. This was highlighted by the SNARE-LHn structures and also in the case of LC/B-GS-Hn/B, an alternative LHn/B backbone with an extended peptide linker region between LC and Hn, in which the 32-long additional peptide was not observed, despite determination of the structure at 2.7Å. In these examples, the fusion peptide was inserted between LC and Hn and did not interrupt the overall structure.

Several other LHn-based molecules were produced for structural investigations. These consisted in ligand-LHn fusion proteins designed with the potential to inhibit secretion in specifically targeted cells, and included a variety of ligands, each partnered with various LHn serotypes. In all molecules, the peptidic ligand was located at the C-terminal end of LHn, thus replacing the cell binding domain of BoNT. The crystal structure of VIP-LHn/A was determined at 3.25 Å but did not provide any information on the ligand's structure due to the low quality of the data. Furthermore, VIP is a short peptide (28 residues) expected to take a α-helical fold, and thus may be too flexible to observe in the crystal structure. In all the structures of LHn determined in this study, the C-terminus of LHn was in a widely open solvent accessible area. Indeed the Hn domain's long α-helical conformation is left unchanged by the removal of its Hc domain in the full length toxin, and leaves an open surface that may thus accommodate new polypeptide without disturbing the main backbone. Determination of the crystal structures of other fusion proteins with different ligands should provide useful information as to how these ligands interact with LHn. The crystallisation conditions for several of these molecules were described in this study and should be the basis for further optimisation. The difficulties seen in obtaining crystals diffracting to high resolution might be due to the flexibility of these additional C-terminal ligands. An alternative approach to investigate their structure may include stabilising this region to support crystallisation. Early work on preparing receptor targets for two of the fusion molecules was described. This may help determining the crystal structures of BoNT-derivatives, as well as providing valuable information on the ligand-receptor interactions.

One of the future priorities should be to determine the crystal structure of the other LHn serotypes, and LHn/D in particular since its potential use in protein engineering to develop new therapeutics has already been reported (Bade *et al.*, 2004). In this study, LHn/D was shown to be subject to auto-proteolytic activity under crystallisation conditions. Thus alternative approaches should be investigated, including crystallisation trials with metal chelating agents, or the production of a mutant inactive LHn/D construct. Additionally, further small-angle x-ray scattering experiments should provide useful structural information to model these molecules at low resolution and confirm or not the stability of the LHn fold across all the serotypes.

Finally, one of the most challenging aspects of the work in this study has been the crystallisation of fusion molecules with therapeutic potentials. Although crystallisation conditions were identified, none of the data collected from x-ray diffraction facilitated structure determination. Further investigations should now focus on co-crystallisation of the

fusion proteins with their molecular receptors. This would involve production of the binding region of the receptors, which in the case of the VIP and EGF fusion molecules, corresponds to soluble extracellular domains. Determination of the crystal structure of such a complex would be a unique opportunity to study the interaction between the protein and the receptor, and thus giving the basis to optimise future therapeutic molecules, particularly in terms of affinity and specificity.

The results presented in this study have demonstrated the stability and functionality of the LHn fragment from the botulinum neurotoxins. They have showed the multiple assets LHn can provide as a biological tool to safely investigate the mechanism of action of the most poisonous protein toxin, as well as the cellular secretion processes. Finally, this study has provided the structural basis for the design of future therapeutic molecules that harness the potency of the botulinum neurotoxin into powerful targeted secretion inhibitors.

References

Abhiman S, Sonnhammer EL (2005) FunShift: a database of function shift analysis on protein subfamilies. *Nucleic Acids Res.* 33, D197-D200.

Agarwal R, Eswaramoorthy S, Kumaran D, Binz T, Swaminathan S (2004) Structural analysis of botulinum neurotoxin type E catalytic domain and its mutant Glu212 Gln reveals the pivotal role of the Glu212 carboxylate in the catalytic pathway. *Biochemistry* 43, 6637-6644.

Agarwal R, Binz T, Swaminathan S (2005a) Analysis of active site residues of botulinum neurotoxin E by mutational, functional and structural studies: Glu335Gln is an apoenzyme. *Biochemistry* 44, 8291-8302.

Agarwal R, Binz T, Swaminathan S (2005b) Structural analysis of botulinum neurotoxin serotype f light chain: implications on substrate binding and inhibitor design. *Biochemistry* 44, 11758-11765.

Agarwal R, Schmidt JJ, Stafford RG, Swaminathan S. (2009) Mode of VAMP substrate recognition and inhibition of Clostridium botulinum neurotoxin F. *Nat. Struct. Mol. Biol.* 16, 789-794.

Agrawal A, Pulendran B (2004) Anthrax lethal toxin: a weapon of multisystem destruction. *Cell Mol. Life Sci.* 61, 2859-2865.

Ahuja N, Kumar P, Bhatnagar R. (2004) The adenylate cyclase toxins. *Crit Rev Microbiol.* 30, 187-196.

Aktories K (2000) Bacterial protein toxins. In Handbook of Experimental Pharmacology Volume 145. Edited by Aktories K, Just I, Springer-Verla, Berlin, Germany.

Aktories K (2006) Toxins as tools. In The Comprehensive Sourcebook of Bacterial Protein Toxins, Third Edition. Edited by Alouf JE, Popoff MR, Academic press, San Diego, California, 976-990.

Alouf J (2006) A 116-year story of bacterial protein toxins. In The Comprehensive Sourcebook of Bacterial Protein Toxins, Third Edition. Edited by Alouf JE, Popoff MR, Academic press, San Diego, California, 3-24.

Arndt JW, Yu W, Bi F, Stevens RC (2005) Crystal structure of botulinum neurotoxin type G light chain: serotype divergence in substrate recognition. *Biochemistry* 44, 9574-9580.

Arndt JW, Chai Q, Christian T, Stevens RC (2006) Structure of botulinum neurotoxin type D light chain at 1.65 Å resolution: repercussions for VAMP-2 substrate specificity. *Biochem. Cell Biol.* 45, 3255-3262.

Arnon SS, Schechter R, Inglesby TV, Henderson DA, Bartlett JG, Ascher MS, Eitzen E, Fine AD, Hauer J, Layton M, Lillibridge S, Osterholm MT, O'Toole T, Parker G, Perl TM, Russell PK, Swerdlow DL, Tonat K (2001) Botulinum toxin as a biological weapon: medical and public health management. *J. Am. Med. Assoc.* 285, 1059-1070.

Bade S, Rummel A, Reisinger C, Karnath T, Ahnert-Hilger G, Bigalke H, Binz T (2004) Botulinum neurotoxin type D enables cytosolic delivery of enzymatically active cargo proteins to neurones via unfolded translocation intermediates. *J Neurochem*. 91, 1461-1472.

Barash JR, Arnon SS (2004) Dual toxin-producing strain of Clostridium botulinum type Bf isolated from a California patient with infant botulism. *J. Clin. Microbiol.* 42, 1713–1715.

Barnes MP (2007) Introduction to the clinical use of Botulinum neurotoxins. In Treatments from toxins- the therapeutic potential of Clostridial neurotoxins. Edited by Foster KA, Hambleton P, Shone CF, CRC press, Boca Raton, Florida, 139-162.

Binz T, Bade S, Rummel A, Kollewe A, Alves J (2002) Arg362 and Tyr365 of the botulinum neurotoxin type A light chain are involved in transition state stabilization. *Biochemistry* 41, 1717-1723.

Binz T, Rummel A. (2009) Cell entry strategy of clostridial neurotoxins. *J Neurochem.* 109, 1584-1595.

Bouvier A, Chapline J, Boerner R, Jeyarajah S, Cook S, Acharya PS, Henderson I, Schrimsher JL, Shepard SR (2003) Identifying and modulating disulfide formation in the biopharmaceutical production of a recombinant protein vaccine candidate. *J. Biotechnol.* 103, 257-271.

Breidenbach MA, Brünger AT (2004) Substrate recognition strategy for botulinum neurotoxin serotype A. *Nature* 432, 925-929.

Brett MM, Hallas G, Mpamugo O (2004) Wound botulism in UK and Ireland. *J. Med. Microbiol.* 53, 555-561.

Briggs DC, Naylor CE, Smedley JG, Lukoyanova N, Robertson S, Moss DS, McClane BA, Basak AK. (2011) Structure of the Food-Poisoning Clostridium perfringens Enterotoxin Reveals Similarity to the Aerolysin-Like Pore-Forming Toxins. *J Mol Biol.* Published ahead of print.

Brünger AT (2005) Structure and function of SNARE and SNARE-interacting proteins. *Q. Rev. Biophys.* 38, 1-47.

Brünger AT, Breidenbach MA, Jin R, Fischer A, Santos JS, Montal M (2007) Botulinum neurotoxin heavy chain belt as an intramolecular chaperone for the light chain. *PLoS Pathog.* 3, 1191-1194.

Carpenter G, Cohen S (1979) Epidermal growth factor. *Ann. Rev. Biochem.* 48, 193-216.

Chaddock JA, Purkiss JR, Duggan MJ, Quinn CP, Shone CC, Foster KA. (2000a) A conjugate composed of nerve growth factor coupled to a non-toxic derivative of Clostridium botulinum neurotoxin type A can inhibit neurotransmitter release in vitro. *Growth Factors* 18, 147-155.

Chaddock JA, Purkiss JR, Friis LM, Broadbridge JD, Duggan MJ, Fooks SJ, Shone CC, Quinn CP, Foster KA (2000b) Inhibition of vesicular secretion in both neuronal and nonneuronal cells by a retargeted endopeptidase derivative of Clostridium botulinum neurotoxin type A. *Infect. Immun.* 68, 2587-2593.

Chaddock JA, Herbert MH, Ling RJ, Alexander FC, Fooks SJ, Revell DF, Quinn CP, Shone CC, Foster KA (2002) Expression and purification of catalytically active, non-toxic endopeptidase derivatives of Clostridium botulinum toxin type A. *Protein Expr. Purif.* 25, 219-228.

Chaddock JA, Purkiss JR, Alexander FC, Doward S, Fooks SJ, Friis LM, Hall YH, Kirby ER, Leeds N, Moulsdale HJ, Dickenson A, Green GM, Rahman W, Suzuki R, Duggan MJ, Quinn CP, Shone CC, Foster KA (2004) Retargeted clostridial endopeptidases: inhibition of nociceptive neurotransmitter release in vitro and antinociceptive activity in in vivo models of pain. *Mov. Disord.* 19, S42-47.

Chai Q, Arndt JW, Dong M, Tepp WH, Johnson EA, Chapman ER, Stevens RC, (2006) Structural basis of cell surface receptor recognition by botulinum neurotoxin B. *Nature* 444, 1096-100.

Chen S, Barbieri J.T. (2007) Multiple pocket recognition of SNAP25 by botulinum neurotoxin serotype E. *J. Biol. Chem.* 282, 25540-25547.

Chen S, Hall C, Barbieri JT (2008) Substrate recognition of VAMP-2 by botulinum neurotoxin B and tetanus neurotoxin. *J. Biol. Chem.* 283, 21153-59.

Chen S, Barbieri JT (2009) Engineering botulinum neurotoxin to extend therapeutic intervention. *Proc. Natl. Acad. Sci. USA.* 106, 9180-9184.

Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.* D50, 760-763.

Collier RJ (1975) Diphtheria toxin: mode of action and structure. *Bacteriol. Rev.* 39, 54-85.

Collins MD, East AK (1998) Phylogeny and taxonomy of the foodborne pathogen *Clostridium botulinum* and its neurotoxins. *J Appl. Microbiol.* 84, 5-17.

Couvineau A, Robert JC, Ramdani T, Lacapère JJ, Rouyer-Fessard C, Laburthe M (2008) Production and purification of large quantities of the functional N-terminal ectodomain of human VPAC1 receptor. *J. Mol. Neurosci.* 36, 249-253.

da Hora VP, Conceição FR, Dellagostin OA, Doolan DL (2011) Non-toxic derivatives of LT as potent adjuvants. *Vaccine* 29, 1538-1544.

Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, Snoeyink J, Richardson JS, Richardson DC (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35, W375-W383.

Dickerson TJ, Janda KD (2006) The use of small molecules to investigate molecular mechanisms and therapeutic targets for treatment of botulinum neurotoxin A intoxication. *A. C. S. Chem. Biol.* 1, 359-69.

Dickson L, Finlayson K (2009) VPAC and PAC receptors: From ligands to function. *Pharmacol. Ther.* 121, 294-316.

Dong M, Yeh F, Tepp WH, Dean C, Johnson EA, Janz R, Chapman ER (2006) SV2 is the protein receptor for botulinum neurotoxin A. *Science* 312, 592-596.

Donovan JJ, Simon MI, Montal M (1982) Insertion of diphtheria toxin into and across membranes: role of phosphoinositide asymmetry. *Nature* 298, 669-672

Drickamer, K. (1988) Two distinct classes of carbohydrate recognition domains in animal lectins. *J. Biol. Chem.* 263, 9557-9560.

Duggan MJ, Quinn CP, Chaddock JA, Purkiss JR, Alexander FC, Doward S, Fooks SJ, Friis LM, Hall YH, Kirby ER, Leeds N, Moulsdale HJ, Dickenson A, Green GM, Rahman W,

Suzuki R, Shone CC, Foster KA (2002) Inhibition of release of neurotransmitters from dorsal root ganglia by a novel conjugate of a Clostridium botulinum toxin A endopeptidase fragment and Erythrina cristagalli lectin. *J. Biol. Chem.* 277, 34846-34852.

Eleopra R, Tugnoli V, Rossetto O, Montecucco C, De Grandis D (1997) Botulinum neurotoxin serotype C: a novel effective botulinum toxin therapy in human. *Neurosci Lett.* 224, 91-94.

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr.* D60, 2126-2132.

Eswaramoorthy S, Kumaran D, Swaminathan S (2001) Crystallographic evidence for doxorubicin binding to the receptor-binding site in Clostridium botulinum neurotoxin B. *Acta Crystallogr. D Biol. Crystallogr.* 57, 1743-1746.

Fasshauer D, Bruns D, Shen B, Jahn R, Brünger AT (1997) A structural change occurs upon binding of syntaxin to SNAP-25. *J. Biol. Chem.* 272, 4582-4590.

Fasshauer D, Sutton RB, Brünger AT, Jahn R (1998) Conserved structural features of the synaptic fusion complex: SNARE proteins reclassified as Q- and R-SNAREs. *Proc. Natl. Acad. Sci. USA* 95, 15781-15786.

Feld GK, Thoren KL, Kintzer AF, Sterling HJ, Tang II, Greenberg SG, Williams ER, Krantz BA (2010) Structural basis for the unfolding of anthrax lethal factor by protective antigen oligomers. *Nat Struct Mol Biol.* 17, 1383-1390.

Fiebig KM, Rice LM, Pollock E, Brünger AT (1999) Folding intermediates of SNARE complex assembly. *Nat. Struct. Biol.* 6, 117-123.

Fischer A, Montal M (2007a) Crucial role of the disulfide bridge between botulinum neurotoxin light and heavy chains in protease translocation across membranes. *J. Biol. Chem.* 282, 29604-29611.

Fischer A, Montal M (2007b) Single molecule detection of intermediates during botulinum neurotoxin translocation across membranes. *Proc. Natl. Acad. Sci. USA* 104, 10447-10452.

Fischer A, Garcia-Rodriguez C, Geren I, Lou J, Marks JD, Nakagawa T, Montal M (2008a) Molecular architecture of botulinum neurotoxin E revealed by single particle electron microscopy. *J. Biol. Chem.* 283, 3997-4003.

Fischer A, Mushrush DJ, Lacy DB, Montal M (2008b) Botulinum neurotoxin devoid of receptor binding domain translocates active protease. *PLoS Pathog.* 4, e1000245.

Fischer A, Nakai Y, Eubanks LM, Clancy CM, Tepp WH, Pellett S, Dickerson TJ, Johnson EA, Janda KD, Montal M (2009) Bimodal modulation of the botulinum neurotoxin protein-conducting channel. *Proc. Natl. Acad. Sci. USA* 106, 1330-1335.

FitzGerald DJ, Kreitman R, Wilson W, Squires D, Pastan I (2004) Recombinant immunotoxins for treating cancer. *Int. J. Med. Microbiol.* 293, 577-582.

Foran PG, Mohammed N, Lisk GO, Nagwaney S, Lawrence GW, Johnson E, Smith L, Aoki KR, Dolly JO (2003) Evaluation of the therapeutic usefulness of botulinum neurotoxin B, C1, E, and F compared with the long lasting type A. Basis for distinct durations of inhibition of exocytosis in central neurons. *J. Biol. Chem.* 278, 1363-1371.

Foster KA, Adams EJ, Durose L, Cruttwell CJ, Marks E, Shone CC, Chaddock JA, Cox CL, Heaton C, Sutton JM, Wayne J, Alexander FC, Rogers DF (2006) Re-engineering the target specificity of clostridial neurotoxins: a route to novel therapeutics. *Neurotox. Res.* 9, 101-107.

Foster KA (2009) Engineered toxins: new therapeutics. *Toxicon*. 54, 587-592.

Franke D, Svergun DI (2009) DAMMIF, a program for rapid *ab-initio* shape determination in small-angle scattering. *J. Appl. Cryst.* 42, 342-346.

Fraser JD, Proft T (2008) The bacterial superantigen and superantigen-like proteins. *Immunol Rev.* 225, 226-243.

Freedman SJ, Song HK, Xu Y, Sun ZY, Eck MJ (2003) Homotetrameric structure of the SNAP-23 N-terminal coiled-coil domain. *J. Biol. Chem.* 278, 13462-13467.

Fu Z, Chen C, Barbieri JT, Kim JJ, Baldwin MR (2009) Glycosylated SV2 and gangliosides as dual receptors for botulinum neurotoxin serotype F. *Biochemistry* 48, 5631-5641.

Galloux M, Vitrac H, Montagner C, Raffestin S, Popoff MR, Chenal A, Forge V, Gillet D (2008) Membrane interaction of botulinum neurotoxin A translocation (T) domain. The belt region is a regulatory loop for membrane interaction. *J. Biol. Chem.* 283, 27668-27676.

Goodnough MC, Oyler G, Fishman PS, Johnson EA, Neale EA, Keller JE, Tepp WH, Clark M, Hartz S, Adler M (2002) Development of a delivery vehicle for intracellular transport of botulinum neurotoxin antagonists. *FEBS Lett.* 513, 163-168.

Gouda H, Torigoe H, Saito A, Sato M, Arata Y, Shimada I (1992) Three-dimensional solution structure of the B domain of staphylococcal protein A: comparisons of the solution and crystal structures. *Biochemistry* 31, 9665-9672.

Guinier, A. (1939) La diffraction des rayons X aux très petits angles: application à l'étude de phénomènes ultramicroscopiques. *Ann. Phys.* 12, 161-237.

Gul N, Smith LA, Ahmed SA (2010) Light chain separated from the rest of the type a botulinum neurotoxin molecule is the most catalytically active form. *PLoS One* 5, e12872.

Harnett MM (1994) Analysis of G-proteins regulating signal transduction pathways. *Methods Mol. Biol.* 27, 199-211.

Hatheway CL (1990) Toxigenic clostridia. *Clin. Microbiol. Rev.*3, 66-98.

Haug G, Leemhuis J, Tiemann D, Meyer DK, Aktories K, Barth H (2003) The host cell chaperone Hsp90 is essential for translocation of the binary Clostridium botulinum C2 toxin into the cytosol. *J. Biol. Chem.* 278, 32266-32274.

Hayashi T, McMahon H, Yamasaki S, Binz T, Hata Y, Südhof TC, Niemann H (1994) Synaptic vesicle membrane fusion complex: action of clostridial neurotoxins on assembly. *EMBO J.* 13, 5051-5061.

Hedeland M, Moura H, Båverud V, Woolfitt AR, Bondesson U, Barr JR (2011) Confirmation of botulism in birds and cattle by the mouse bioassay and Endopep-MS. *J. Med. Microbiol.* 60, 1299-1305.

Hill KK, Smith TJ, Helma CH, Ticknor LO, Foley BT, Svensson RT, Brown JL, Johnson EA, Smith LA, Okinaka RT, Jackson PJ, Marks JD (2007) Genetic diversity among Botulinum Neurotoxin-producing clostridial strains. *J Bacteriol.* 189, 818-832.

Hill KK, Xie G, Foley BT, Smith TJ, Munk AC, Bruce D, Smith LA, Brettin TS, Detter JC (2009) Recombination and insertion events involving the botulinum neurotoxin complex genes in Clostridium botulinum types A, B, E and F and Clostridium butyricum type E strains. *BMC Biol.* 5, 7-66.

Inoue K, Sobhany M, Transue TR, Oguma K, Pedersen LC, Negishi M (2003) Structural analysis by X-ray crystallography and calorimetry of a haemagglutinin component (HA1) of the progenitor toxin from Clostridium botulinum. *Microbiology* 149, 3361-3370.

Jahn R, Scheller RH (2006) SNAREs–engines for membrane fusion. *Nat. Rev. Mol. Cell Biol.* 7, 631-643.

Jin R, Sikorra S, Stegmann CM, Pich A, Binz T, Brünger AT (2007) Structural and biochemical studies of botulinum neurotoxin serotype C1 light chain protease: implications for dual substrate specificity. *Biochemistry* 46, 10685-10693.

Johannes L, Römer W (2010) Shiga toxins--from cell biology to biomedical applications. *Nat. Rev. Microbiol.* 8, 105-116.

Karalewitz AP, Kroken AR, Fu Z, Baldwin MR, Kim JJ, Barbieri JT (2010) Identification of a unique ganglioside binding loop within botulinum neurotoxins C and D-SA. *Biochemistry* 49, 8117-8126.

Keller JE, Cai F, Neale EA (2004) Uptake of botulinum neurotoxin into cultured neurons. *Biochemistry* 43, 526-532.

Kim YS, Bhandari R, Cochran JR, Kuriyan J, Wittrup KD (2006) Directed evolution of the epidermal growth factor receptor extracellular domain for expression in yeast. *Proteins* 62, 1026-1035.

Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI (2003) PRIMUS - a Windows-PC based system for small-angle scattering data analysis. *J. Appl. Cryst.* 36, 1277-1282.

Koriazova LK, Montal M (2003) Translocation of botulinum neurotoxin light chain protease through the heavy chain channel. *Nat. Struct. Biol.* 10, 13-18.

Kozaki S, Kamata Y, Watarai S, Nishiki T, Mochida S (1998) Ganglioside GT1b as a complementary receptor component for Clostridium botulinum neurotoxins. *Microb. Pathog.* 25, 91-99.

Kubitzki T, Noll T, Lütz S (2008) Immobilisation of bovine enterokinase and application of the immobilised enzyme in fusion protein cleavage. *Bioprocess Biosyst Eng.* 31, 173-182.

Kubota T, Yonekura N, Hariya Y, Isogai E, Isogai H, Amano K, Fujii N (1998) Gene arrangement in the upstream region of Clostridium botulinum type E and Clostridium butyricum BL6340 progenitor toxin genes is different from that of other types. *FEMS Microbiol. Lett.* 158, 215-221.

Kukreja RV, Sharma S, Cai S, Singh BR (2007) Role of two active site Glu residues in the molecular action of botulinum neurotoxin endopeptidase. *Biochim. Biophys. Acta* 1774, 213-222.

Kumaran D, Rawat R, Ahmed SA, Swaminathan S (2008) Substrate binding mode and its implication on drug design for botulinum neurotoxin A. *PLoS Pathog.* 4, e1000165.

Kumaran D, Eswaramoorthy S, Furey W, Navaza J, Sax M, Swaminathan S (2009) Domain organization in Clostridium botulinum neurotoxin type E is unique: its implication in faster translocation. *J. Mol. Biol.* 386, 233-245.

Kurazono H, Mochida S, Binz T, Eisel U, Quanz M, Grebenstein O, Wernars K, Poulain B, Tauc L, Niemann H (1992) Minimal essential domains specifying toxicity of the light chains of tetanus toxin and botulinum neurotoxin type A. *J Biol Chem.* 267, 14721-14729.

Laburthe M, Couvineau A, Tan V (2007) Class II G protein-coupled receptors for VIP and PACAP: structure, models of activation and pharmacology, *Peptides* 28, 1631-1639.

Lacy DB, Tepp W, Cohen AC, DasGupta BR, Stevens RC (1998) Crystal structure of botulinum neurotoxin type A and implications for toxicity. *Nat. Struct. Biol.* 5, 898-902.

Lacy DB, Stevens RC (1999) Sequence homology and structural analysis of clostridial neurotoxins. *J. Mol. Biol.* 291, 1091-1104.

Ladhani S (2003) Understanding the mechanism of action of the exfoliative toxins of Staphylococcus aureus. *FEMS Immunol. Med. Microbiol.* 39, 181-189.

Lemonnier M, Landraud L, Lemichez E (2007) Rho GTPase-activating bacterial toxins: from bacterial virulence regulation to eukaryotic cell biology. *FEMS Microbiol. Rev.* 31, 515-534.

Leslie AGW (2006). The integration of macromolecular diffraction data. *Acta Crystallogr.* D62, 48-57.

Lu HS, Chai JJ, Li M, Huang BR, He CH, Bi RC (2001) Crystal structure of human epidermal growth factor and its dimerization. *J. Biol. Chem.* 276, 34913-34917.

Margittai M, Fasshauer D, Pabst S, Jahn R, Langen R. (2001) Homo- and heterooligomeric SNARE complexes studied by site-directed spin labelling. *J. Biol. Chem.* 276, 13169-13177.

Masuyer G, Thiyagarajan N, James PL, Marks PM, Chaddock JA, Acharya KR (2009) Crystal structure of a catalytically active, non-toxic endopeptidase derivative of *Clostridium botulinum* toxin A. *Biochem. Biophys. Res. Commun.* 381, 50-53.

Masuyer G, Beard M, Cadd VA, Chaddock JA, Acharya KR (2011) Structure and activity of a functional derivative of *Clostridium botulinum* neurotoxin B. *J. Struct. Biol.* 174, 52-57.

McCoy JA, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J. Appl. Crystllogr.* 40, 658-674.

McMahon HT, Ushkaryov YA, Edelmann L, Link E, Binz T, Niemann H, Jahn J, Südhof TC (1993) Cellubrevin is a ubiquitous tetanus-toxin substrate homologous to a putative synaptic vesicle fusion protein. *Nature* 364, 346-349.

McPherson A (1985) Crystallization of macromolecules: General principles. *Method. Enzymol.* 114, 112-120.

Melville S (2007) Clostridia: diarrheal disease, tissue infection, botulism, and tetanus. In Schaechter's mechanisms of microbial disease, Fourth edition. Edited by Engleberg NC, Dermody TS, DiRita V, Lippincott Williams & Wilikins, Philadelphia, Pennsylvania, 230-235.

Mezaki T, Kaji R, Kohara N, Fujii H, Katayama M, Shimizu T, Kimura J, Brin MF (1995) Comparison of therapeutic efficacies of type A and F botulinum toxins for blepharospasm: a double-blind, controlled study. *Neurology* 45, 506-508.

Moks T, Abrahmsén L, Holmgren E, Bilich M, Olsson A, Uhlén M, Pohl G, Sterky C, Hultberg H, Josephson S (1987) Expression of human insulin-like growth factor I in bacteria: use of optimized gene fusion vectors to facilitate protein purification. *Biochemistry* 26, 5239-5244.

Montecucco C (1986) How do tetanus and botulinum toxins bind to neuronal membranes? *Trends Biochem. Sc.* 11, 314-317.

Moore AP (2007) Expanding clinical uses of Botulinum neurotoxins. In Treatments from toxins- the therapeutic potential of Clostridial neurotoxins. Edited by Foster KA, Hambleton P, Shone CF, CRC press, Boca Raton, Florida, 163-194.

Moriishi, K., M. Koura, N. Abe, N. Fujii, Y. Fujinaga, K. Inoue, and K. Oguma (1996) Mosaic structures of neurotoxins produced from Clostridium botulinum types C and D organisms. *Biochim. Biophys. Acta* 1307, 123-126.

Muraro L, Tosatto S, Motterlini L, Rossetto O, Montecucco C (2009) The N-terminal half of the receptor domain of botulinum neurotoxin A binds to microdomains of the plasma membrane. *Biochem. Biophys. Res. Commun.* 380, 76-80.

Neer EJ (1995) Heterotrimeric G proteins: organizers of transmembrane signals. *Cell* 80, 249-257.

Nilsson B, Moks T, Jansson B, Abrahmsén L, Elmblad A, Holmgren E, Henrichson C, Jones TA, Uhlén M (1987) A synthetic IgG-binding domain based on staphylococcal protein A. *Protein Eng.* 1, 107-113.

Nishiki T, Tokuyama Y, Kamata Y, Nemoto Y, Yoshida A, Sato K, Sekiguchi M, Takahashi M, Kozaki S (1996) The high-affinity binding of Clostridium botulinum type B neurotoxin to synaptotagmin II associated with gangliosides GT1b/GD1a. *FEBS Lett.* 378, 253-257.

Nizard P, Chenal A, Beaumelle B, Fourcade A, Gillet D (2001) Prolonged display or rapid internalization of the IgG-binding protein ZZ anchored to the surface of cells using the diphtheria toxin T domain. *Protein Eng.* 14, 439-446.

Nowack A, Yao J, Custer KL, Bajjalieh SM (2010) SV2 regulates neurotransmitter release via multiple mechanisms. *Am. J. Physiol. Cell Physiol.* 299, C960-C967.

Ogiso H, Ishitani R, Nureki O, Fukai S, Yamanaka M, Kim JH, Saito K, Sakamoto A, Inoue M, Shirouzu M, Yokoyama S (2002) Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* 110, 775-787.

Pang ZP, Südhof TC (2010) Cell biology of Ca^{2+}-triggered exocytosis. *Curr. Opin. Cell. Biol.* 22, 496-505.

Parker MW, Pattus F (1993) Rendering a membrane protein soluble in water: a common packing motif in bacterial protein toxins. *Trends Biochem. Sci.* 18, 391-395.

Pastan I, Hassan R, Fitzgerald DJ, Kreitman RJ (2006) Immunotoxin therapy of cancer. *Nat. Rev. Cancer* 6, 559-565.

Petoukhov MV, Svergun, DI (2005) Global rigid body modelling of macromolecular complexes against small-angle scattering data. *Biophys. J.* 89, 1237-1250.

Poulain B, Stiles BG, Popoff MR, Molgo J (2006) Attack of the nervous system by clostridial toxins: physical findings, cellular and molecular actions. In The Comprehensive Sourcebook of Bacterial Protein Toxins, Third Edition. Edited by Alouf JE, Popoff MR (2006) Academic press, San Diego, California, 348-389.

Proux-Gillardeaux V, Gavard J, Irinopoulou T, Mege RM, Galli T (2005) Tetanus neurotoxin-mediated cleavage of cellubrevin impairs epithelial cell migration and integrin-dependent cell adhesion. *Proc. Natl. Acad. Sci. USA* 102, 6362-6367.

Putnam CD, Hammel M, Hura GL, Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* 40, 191-285.

Ratts R, Zeng H, Berg EA, Blue C, McComb ME, Costello CE, vanderSpek JC, Murphy JR (2003) The cytosolic entry of diphtheria toxin catalytic domain requires a host cell cytosolic translocation factor complex. *J. Cell Biol.* 160, 1139-1150.

Rossetto O, Schiavo G, Montecucco C, Poulain B, Deloye F, Lozzi L, Shone CC (1994). SNARE motif and neurotoxins. *Nature* 372, 415-416.

Roux E, Yersin A (1888) Contribution à l'étude de la diphtérie. *Ann. Inst. Pasteur.* 2, 629-661.

Rummel A, Mahrhold S, Bigalke H, Binz T. The HCC-domain of botulinum neurotoxins A and B exhibits a singular ganglioside binding site displaying serotype specific carbohydrate interaction. *Mol. Microbiol.* 51, 631-643.

Sakurai J, Nagahama M, Oda M (2004) Clostridium perfringens alpha-toxin: characterization and mode of action. *J. Biochem.* 136, 569-574.

Sandvig, K. (2003). Transport of toxins across intracellular membranes. In Bacterial Protein Toxins. Edited by Burns DL, Barbieri JT, Iglewski BH, Rappuoli R (2003) ASM Press, Washington, D.C. 157-172.

Schiavo G, Benfenati F, Poulain B, Rossetto O, Polverino de Laureto P, DasGupta BR, Montecucco C (1992) Tetanus and botulinum B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin. *Nature* 359, 832-835.

Schmid MF, Robinson JP, DasGupta BR (1993) Direct visualization of botulinum neurotoxin induced channels in phospholipid vesicles. *Nature* 364, 827-830.

Schmitt J, Karalewitz AP, Benefield DA, Mushrush DJ, Pruitt RN, Spiller BW, Barbieri JT, Lacy DB (2010) Structural analysis of botulinum neurotoxin type G receptor binding. *Biochem. Cell Biol.* 49, 5200-5205.

Scott, A.B (1980) Botulinum toxin injection into extraocular muscles as an alternative to strabismus surgery. *Ophthalmology* 87, 1044-1049.

Segelke B, Knapp M, Kadkhodayan S, Balhorn R, Rupp B (2004) Crystal structure of Clostridium botulinum neurotoxin protease in a product-bound state: evidence for noncanonical zinc protease activity. *Proc. Natl. Acad. Sci. USA* 101, 6888-6893.

Sharma SK, Ramzan MA, Singh BR (2003) Separation of the components of type A botulinum neurotoxin complex by electrophoresis. *Toxicon* 41, 321-331.

Shone CC, Hambleton P, Melling J (1985) Inactivation of Clostridium botulinum type A neurotoxin by trypsin and purification of two tryptic fragments: proteolytic action near the COOH-terminus of the heavy subunit destroys toxin-binding activity. *Eur. J. Biochem.* 151, 75-82.

Shone CC, Hambleton P, Melling J (1987) A 50-kDa fragment from the NH2-terminus of the heavy subunit of *Clostridium botulinum* type A neurotoxin forms channels in lipid vesicles. *Eur. J. Biochem.* 167, 175-180.

Shone CC, Quinn CP, Wait R, Hallis B, Fooks SG, Hambleton P (1993) Proteolytic cleavage of synthetic fragments of vesicle-associated membrane protein, isoform-2 by botulinum type B neurotoxin. *Eur. J. Biochem.* 217, 965-971.

Sikorra S, Henke T, Galli T, Binz T, (2008) Substrate recognition mechanism of VAMP/synaptobrevin-cleaving clostridial neurotoxins. *J. Biol. Chem.* 283, 21145-21152.

Smalley SG, Barrow PA, Foster N (2009) Immunomodulation of innate immune responses by vasoactive intestinal peptide (VIP): its therapeutic potential in inflammatory disease. *Clin. Exp. Immunol.* 157, 225-234.

Smith TJ, Lou J, Geren I, Forsyth CM, Tsai R, LaPorte SL, Tepp WH, Bradshaw M, Johnson EA, Smith LA, Marks JD (2005) Sequence variation within botulinum neurotoxin serotypes impacts antibody binding and neutralization. *Infect. Immun.* 73, 5450-5457.

Smith JL, Bayles DO (2006) The contribution of cytolethal distending toxin to bacterial pathogenesis. *Crit. Rev. Microbiol.* 32, 227-248.

Smith LA (2009) Botulism and vaccines for its prevention. *Vaccine* 27, D33-D39.

Sobel J (2005) Botulism. Clin. Infect. Dis. 41, 1167-1173.

Staats HF, Fielhauer JR, Thompson AL, Tripp AA, Sobel AE, Maddaloni M, Abraham SN, Pascual DW (2011) Mucosal targeting of a BoNT/A subunit vaccine adjuvanted with a mast cell activator enhances induction of BoNT/A neutralizing antibodies in rabbits. *PLoS One* 6, e16532.

Stancombe PR, Masuyer G, Birch-Machin I, Beard M, Foster KA, Chaddock JA, Acharya KR (2011) Engineering Clostridium botulinum endopeptidase and translocation domains with SNARE peptides. *Submitted*.

Stenmark P, Dupuy J, Imamura A, Kiso M, Stevens RC (2008) Crystal structure of botulinum neurotoxin type A in complex with the cell surface co-receptor GT1b—insight into the toxin-neuron interaction. *PLoS Pathog.* 4, e1000129.

Stenmark P, Dong M, Dupuy J, Chapman ER, Stevens RC (2010) Crystal structure of the botulinum neurotoxin type G binding domain: insight into cell surface binding. *J. Mol. Biol.* 397, 1287-1297.

Strotmeier J, Lee K, Völker AK, Mahrhold S, Zong Y, Zeiser J, Zhou J, Pich A, Bigalke H, Binz T, Rummel A, Jin R (2010) Botulinum neurotoxin serotype D attacks neurons via two carbohydrate-binding sites in a ganglioside-dependent manner. *Biochem. J.* 431, 207-216.

Südhof TC, Rothman JE (2009) Membrane fusion: grappling with SNARE and SM proteins. *Science* 323, 474-477.

Sutton JM, Wayne J, Scott-Tucker A, O'Brien SM, Marks PM, Alexander FC, Shone CC, Chaddock JA. (2005) Preparation of specifically activatable endopeptidase derivatives of Clostridium botulinum toxins type A, B and C and their applications. *Protein Expr. Purif.* 40, 31-41.

Sutton RB, Fasshauer D, Jahn R, Brünger AT (1998). Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 Å resolution. *Nature.* 395, 347-353.

Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Cryst.* 25, 495-503.

Svergun DI, Barberato C, Koch MHJ (1995) CRYSOL - a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates *J. Appl. Cryst.* 28, 768-773.

Takamori S, Holt M, Stenius K, Lemke EA, Grønborg M, Riedel D, Urlaub H, Schenck S, Brügger B, Ringler P, Müller SA, Rammner B, Gräter F, Hub JS, De Groot BL, Mieskes G, Moriyama Y, Klingauf J, Grubmüller H, Heuser J, Wieland F, Jahn R (2006) Molecular anatomy of a trafficking organelle. *Cell* 127, 831-846.

Tan YV, Couvineau A, Murail S, Ceraudo E, Neumann JM, Lacapère JJ, Laburthe M (2006) Peptide agonist docking in the N-terminal ectodomain of a class II G protein-coupled receptor, the VPAC1 receptor. Photoaffinity, NMR, and molecular modelling. *J. Biol. Chem.* 281, 12792-1278.

Thompson AA, Jiao GS, Kim S, Thai A, Cregar-Hernandez L, Margosiak SA, Johnson AT, Han GW, O'Malley S, Stevens RC (2011). Structural characterization of three novel hydroxamate-based zinc chelating inhibitors of the *Clostridium botulinum* serotype A neurotoxin light chain metalloprotease reveals a compact binding site resulting from 60/70 loop flexibility. *Biochemistry*. 50, 4019-4028.

Umetsu Y, Tenno T, Goda N, Shirakawa M, Ikegami T, Hiroaki H (2011) Structural difference of vasoactive intestinal peptide in two distinct membrane-mimicking environments. *Biochim. Biophys. Acta* 1814, 724-730.

Vagin AA, Isupov MN (2001) Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps. *Acta Crystallogr.* D57, 1451-1456.

Vagin AA, Steiner RS, Lebedev AA, Potterton L, McNicholas S, Long F, Murshudov GN (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr.* D60, 2184-2195.

Volkov VV, Svergun DI (2003) Uniqueness of *ab-initio* shape determination in small-angle scattering. *J. Appl. Cryst.* 36, 860-864.

von Pawel-Rammingen U, Björck L (2003) IdeS and SpeB: immunoglobulin-degrading cysteine proteinases of *Streptococcus pyogenes*. *Curr. Opin. Microbiol.* 6, 50-55.

Wang J, Meng J, Lawrence GW, Zurawski TH, Sasse A, Bodeker MO, Gilmore MA, Fernández-Salas E, Francis J, Steward LE, Aoki KR, Dolly JO (2008) Novel chimeras of botulinum neurotoxin/A and /E unveil contributions from the binding, translocation and protease domains to their functional characteristics. *J. Biol. Chem.* 283, 16993-17002.

Wang J, Zurawski TH, Meng J, Lawrence G, Olango WM, Finn DP, Wheeler L, Dolly JO (2011) A dileucine in the protease of botulinum toxin A underlies its long-lived neuroparalysis: transfer of longevity to a novel potential therapeutic. *J. Biol. Chem.* 286, 6375-6385.

Washbourne P, Pellizzari R, Baldini G, Wilson MC, Montecucco C (1997) Botulinum neurotoxin types A and E require the SNARE motif in SNAP-25 for proteolysis. *FEBS Lett.* 418, 1-5.

Wey JJ, Tang SS, Wu TY. Disulfide bond reduction corresponds to dimerization and hydrophobicity changes of *Clostridium botulinum* type A neurotoxin. *Acta Pharmacol. Sin.* 27, 1238-1246.

Williamson C, Neale EA (1998) Syntaxin and 25-kDa synaptosomal associated protein: differential effects of botulinum neurotoxins C1 and A on neuronal survival. *J. Neurosci. Res.* 52, 569-583.

Yuan J, Inami G, Mohle-Boetani J, Vugia DJ (2011) Recurrent wound botulism among injection drug users in California. *Clin. Infect. Dis.* 52, 862-866.

Zhou L, de Paiva A, Liu D, Aoki R, Dolly JO (1995) Expression and purification of the light chain of botulinum neurotoxin A: a single mutation abolishes its cleavage of SNAP-25 and neurotoxicity after reconstitution with the heavy chain. *Biochemistry* 34, 15175-15181.

Zuniga JE, Hammill JT, Drory O, Nuss JE, Burnett JC, Gussio R, Wipf P, Bavari S, Brünger AT (2010) Iterative structure-based peptide-like inhibitor design against the botulinum neurotoxin serotype A. *PLoS one* 5, e11378.

Appendix

List of publications

Masuyer G, Thiyagarajan N, James PL, Marks PM, Chaddock JA, Acharya KR (2009) Crystal structure of a catalytically active, non-toxic endopeptidase derivative of *Clostridium botulinum* toxin A. *Biochem. Biophys. Res. Commun.* 381, 50-53.

Masuyer G, Beard M, Cadd VA, Chaddock JA, Acharya KR (2011) Structure and activity of a functional derivative of *Clostridium botulinum* neurotoxin B. *J. Struct. Biol.* 174, 52-57.

Masuyer G, Stancombe PR, Chaddock JA, Acharya KR (2011) Crystal structures of engineered Clostridium botulinum neurotoxin derivatives. *Acta Crystallogr.* F 67, 1466-1472.

Stancombe PR, **Masuyer G**, Birch-Machin I, Beard M, Foster KA, Chaddock JA, Acharya KR (2012) Engineering Clostridium botulinum endopeptidase and translocation domains with SNARE peptides. *FEBS J.* 279, 515-523.

*****

Akif M, **Masuyer G**, Schwager SL, Bhuyan BJ, Mugesh G, Isaac RE, Sturrock ED, Acharya KR (2011) Structural characterization of angiotensin I-converting enzyme in complex with a selenium analogue of captopril. *FEBS J.* 278, 3644-3650.

Masuyer G, Talat J, Öberg C, Leffler H, Nilsson U, Acharya KR (2012) Inhibition mechanism of human galectin-7 by a novel galactose-benzylphosphate inhibitor. *FEBS J.* 279, 193-202.