Journal of Structural Biology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

High resolution crystal structures of *Clostridium botulinum* neurotoxin A3 and A4 binding domains

Jonathan R. Davies[a], Jay Rees[a], Sai Man Liu[b], K. Ravi Acharya[a,*]

[a] Department of Biology and Biochemistry, Claverton Down, University of Bath, Bath BA2 7AY, UK
[b] Ipsen Bioinnovation Limited, 102 Park Lane, Milton Park, Abingdon OX14 4RY, UK

ARTICLE INFO

Keywords:
Clostridium botulinum
Botulinum neurotoxin
Subtypes
Binding domain structure
Crystallography

ABSTRACT

Clostridium botulinum neurotoxins (BoNTs) cause the life-threatening condition, botulism. However, while they have the potential to cause serious harm, they are increasingly being utilised for therapeutic applications. BoNTs comprise of seven distinct serotypes termed BoNT/A through BoNT/G, with the most widely characterised being sub-serotype BoNT/A1. Each BoNT consists of three structurally distinct domains, a binding domain (H_C), a translocation domain (H_N), and a proteolytic domain (LC). The H_C domain is responsible for the highly specific targeting of the neurotoxin to neuronal cell membranes. Here, we present two high-resolution structures of the binding domain of subtype BoNT/A3 (H_C/A3) and BoNT/A4 (H_C/A4) at 1.6 Å and 1.34 Å resolution, respectively. The structures of both proteins share a high degree of similarity to other known BoNT H_C domains whilst containing some subtle differences, and are of benefit to research into therapeutic neurotoxins with novel characteristics.

1. Introduction

Botulinum neurotoxins (BoNTs) comprise a family of proteins produced predominantly by *Clostridium botulinum*. There are seven distinct serotypes termed BoNT/A through BoNT/G that cause the deadly condition botulism in vertebrates (Schiavo et al., 2000, Johnson and Montecucco, 2008, Rossetto et al., 2014). Of these serotypes, /A, /B, /E and /F are most commonly associated with human botulism (Coffield et al., 1997). BoNTs are composed of three major domains that are organised into two chains: a 50 kDa light chain (LC) and a 100 kDa heavy chain (HC), that are linked by a disulphide bond. The LC comprises a single zinc-endopeptidase domain while the HC comprises the other two domains – a translocation domain (H_N) responsible for transporting the LC across a plasma membrane, and a binding domain (H_C) responsible for forming a receptor complex with the target neuronal cell. All BoNTs bind polysialogangliosides found on neuronal cell membranes. In addition, BoNT/A, /D, /E and /F also bind the luminal domain of the transmembrane synaptic vesicle 2 (SV2) protein, while BoNT/B, /DC (a mosaic toxin) and /G instead bind to the luminal domain of the protein synaptotagmin I and II (Syt).

Each BoNT serotype can be further categorised into subtypes based on amino acid sequence identity. For BoNT/A there are currently eight known subtypes (BoNT/A1-8) which share between 84% and 97% sequence identity amongst them. The most widely used therapeutic BoNT is the subtype A1. Some subtypes have been reported to exhibit different properties to BoNT/A1 such as onset of symptoms and duration of action (Tepp et al., 2012; Whitemarsh et al., 2014; Henkel et al., 2009). For example, BoNT/A3 has been shown to be less effectively neutralised by anti-BoNT/A1 antibodies whilst also inducing symptoms of intoxication in mice that are distinct from BoNT/A1 (Mazuet et al., 2010; Tepp et al., 2012). It has been suggested that these differences are due to structural differences within the binding domain of BoNT/A3 (H_C/A3) (Tepp et al., 2012). Here we provide structural insight into the subtype differences of the H_C domains from BoNT/A3 and also BoNT/A4, and provide a plausible explanation on how structural differences may play a role in receptor binding.

2. Materials and methods

2.1. Protein expression and purification

The binding domains of BoNT/A3 (residues 866-1292) and BoNT/A4 (residues 870-1296) were cloned into the pJ401 vector (Atum) from their respective full-length sequences (Genbank: ABA29017.1 and ABA29018.1) with an N-terminal 6xHis tag. Constructs were transformed into *E. coli* BL21 cells, which were grown in 0.5 L LB at 37 °C, and induced at an OD_{600} of 0.6 with 1 mM IPTG at 16 °C for 16 h. For purification, cells expressing H_C/A3 were lysed in 50 mM Tris pH 7.4,

* Corresponding author.
E-mail address: bsskra@bath.ac.uk (K.R. Acharya).

https://doi.org/10.1016/j.jsb.2017.12.010
Received 21 November 2017; Received in revised form 20 December 2017; Accepted 23 December 2017

0.5 M NaCl, whereas those expressing H_C/A4 were lysed in 50 mM Tris pH 7.4, 0.2 M NaCl, 10 mM Trehalose. Each protein was captured on a HisTrap column (GE Healthcare) and further purified by size-exclusion chromatography on a Superdex 200 column (GE Healthcare).

2.2. Crystallisation

Protein crystallisation conditions were screened using the sitting-drop vapour diffusion method at 16 °C. H_C/A3 (6 mg/mL) and H_C/A4 (4 mg/mL) were dispensed using an Art Robbins Phoenix crystal screening nano dispenser into 96-well 3-drop Intelliwell plates (Molecular Dimensions, England). Multiple screening kits from Molecular Dimensions were used. H_C/A3 crystals were obtained after one week with 0.1 M MIB pH 4.0 25% w/v PEG 1500 from the PACT premier screen (condition B1). Individual crystals were mounted in a loop and dipped in a solution containing reservoir and 50% glycerol in a 1:1 ratio for cryoprotection before vitrification and storage in liquid nitrogen. H_C/A4 crystals were also obtained after one week, but using the Morpheus II screen (Gorrec, 2015) (Molecular Dimensions, England). The best crystal of approximately 60 μm^3 was produced in 10 mM Spermine tetrahydrochloride, 10 mM Spermidine trihydrochloride, 10 mM 1,4-Diaminobutane dihydrochloride, 10 mM D-L-Ornithine monohydrochloride, 0.1 M MOPSO/bis-tris pH 6.5, 15%(w/v) PEG 3 K, 20%(v/v), 1,2,4-butanetriol, 1%(w/v) NDSB 256 (condition H1). This was mounted in a loop and then flash-cooled for storage in liquid nitrogen.

2.3. X-ray data collection and structure determination

Diffraction data for H_C/A3 and H_C/A4 (7200 images total) were collected using 0.1° oscillations and exposures of 0.02 s at beamline IO3 (Diamond Light Source, Didcot, UK) and processed using DIALS (Waterman et al., 2016). The structures of both proteins were solved by molecular replacement with PHASER (McCoy et al., 2007) using the binding domain of BoNT/A1 (PDB 2VUA; Stenmark et al., 2008) as a search model. Both models were manually fitted using COOT (Emsley et al., 2010) and refined with REFMAC (Murshudov et al., 1997) in the CCP4 suite of programs (Winn et al., 2011). The structures were validated with MolProbity (Chen et al., 2010). Crystallographic data processing and refinement statistics are given in Table 1. Figures were prepared using CCP4MG (McNicholas et al., 2011).

Table 1
X-ray data collection and refinement statistics. Inner and outer shell statistics are given in square and curved brackets respectively.

	H_C/A3	H_C/A4
Diamond Beamline	I03, DLS	I03, DLS
Wavelength (Å)	0.9763	0.9763
Crystallographic Statistics		
Space Group	$P2_12_12_1$	$P2_13$
Unit-cell parameters:		
a,b,c (Å)	39.8, 96.5, 111.0	116.0, 116.0, 116.0
$\alpha = \beta = \gamma$ (°)	90	90
Resolution (Å)	[96.47–8.76] 96.47–1.60 (1.63–1.60)	[81.95–7.33] 81.95–1.34 (1.36–1.34)
R_{merge} (%)	[0.076] 0.126 (1.352)	[0.038] 0.091 (2.306)
R_{meas} (%)	[0.080] 0.131 (1.416)	[0.039] 0.093 (2.424)
R_{pim} (%)	[0.024] 0.036 (0.408)	[0.008] 0.020 (0.740)
$CC_{1/2}$ (%)	[0.998] 0.997 (0.639)	[1.000] 0.999 (0.421)
Mean < I/σ(I) >	[24.3] 10.3 (1.9)	[68.1] 17.5 (1.1)
Completeness (%)	[99.6] 96.9 (79.6)	[100.0] 100.0 (100.0)
No. observed reflections	[4680] 715,684 (25075)	[16955] 2516326 (60014)
No. unique reflections	[4 3 0] 55445 (2169)	[7 9 4] 116165 (5700)
Multiplicity	[10.9] 12.9 (11.6)	[21.4] 21.7 (10.5)
Refinement Statistics		
R_{work}/R_{free}	0.179/0.214	0.144/0.158
RMSD bond lengths (Å)	0.016	0.001
RMSD bond angles (°)	1.819	1.490
Ramachandran statistics (%)		
Favoured	97.74	96.62
Allowed	2.26	3.38
Outliers	0	0
Average B-factors (Å^2)		
Protein	23.5	20.7
Water	33.8	30.8
No. of atoms		
Protein	3319	3504
Water	364	309
PDB Code	**6F0O**	**6F0P**

3. Results and discussion

3.1. Structure of the BoNT/A3 binding domain (H_C/A3)

Crystals of H_C/A3 diffracted to 1.6 Å resolution, and belonged to space group $P2_12_12_1$ with one molecule in the asymmetric unit. A resolution cut-off was imposed based on $CC_{1/2}$ (Evans and Murshudov, 2013). The electron density map obtained following molecular replacement was very good except for three loop regions 1198-1201, 1268-1277 and 1227-1231, which could not be modelled completely. The crystal structure of H_C/A3 contains two motifs common to all BoNTs – a jelly-roll fold consisting of 14 β-strands, and a β-trefoil fold which are predicted to contain a conserved ganglioside and SV2 receptor binding site (Fig. 1a). The H_C/A3 sequence is most similar to that of the BoNT/A2 subtype, where both binding domains share 98.8% identity. This is reflected with a high structural similarity where the RMSD with H_C/A2 is 0.83 Å between Cα atoms.

The ganglioside binding site of BoNT/A1 is mostly conserved in the H_C/A3 structure (Fig. 2i) except for a tyrosine at position 1117 (instead of phenylalanine) and an undetermined conformation of R1276 due to poor electron density. Considering that BoNT/A1 binds to GT1b, it is predicted that BoNT/A3 might do so as well; however, the different residue at position 1117 may affect this since it would no longer be able to form a hydrogen bond to a sialic acid of GT1b (Stenmark et al., 2008).

Recent structures of both H_C/A1 and H_C/A2 bound to their protein receptor, SV2 isoform C (SV2C), suggest that the overall binding mode is generally conserved (Benoit et al., 2014; Yao et al., 2016; Benoit et al., 2017). The high conformational similarity with our structure suggests that H_C/A3 may also bind SV2C in the same manner (Fig. 2b and c). One notable point of variation, however, is position 1152 which is a methionine (and also for subtypes /A4 to /A8) that has the potential to form an electrophilic interaction with H563 of SV2C (Fig. 2c) (Pal and Chakrabarti, 2001). BoNT/A1, however, has an arginine at the equivalent position (Fig. 2a) that forms a cation-π stacking interaction, whereas BoNT/A2 has a glutamate that forms a salt bridge with the equivalent histidine of SV2C. Under normal physiological conditions SV2 is glycosylated, and N-linked glycosylation of SV2C is important for binding of BoNT/A1 to the neuronal cell membrane (Yao et al., 2016; Mahrhold et al., 2016). The glycans form an array of interactions with BoNT residues that neighbour those which interact with the protein chain directly. Within BoNT/A3 (and subtypes /A2, /A6 and /A8), an arginine residue at position 1060 (Fig. 2f), may cause a change to glycan binding specificity compared to BoNT/A1, which has a histidine in the equivalent position (Fig. 2e).

3.2. Structure of the BoNT/A4 binding domain (H_C/A4)

Despite the lower protein concentration of H_C/A4 (due to a lower solubility than H_C/A3) used in the crystal screens, it was possible to grow high quality crystals that diffracted to 1.34 Å resolution, and belonged to space group $P2_13$ with one molecule in the asymmetric unit. The exceptional quality of the electron density map allowed modelling