Microbiology (2003), **149**, 3361–3370 DOI 10.1099/mic.0.26586-0

Structural analysis by X-ray crystallography and calorimetry of a haemagglutinin component (HA1) of the progenitor toxin from *Clostridium botulinum*

Kaoru Inoue,[1]† Mack Sobhany,[1]† Thomas R. Transue,[2] Keiji Oguma,[3] Lars C. Pedersen[1,2] and Masahiko Negishi[1]

Correspondence
Lars C. Pedersen
pederse2@niehs.nih.gov

[1,2]Pharmacogenetic Section Laboratory of Reproductive and Developmental Toxicology[1] and Laboratory of Structural Biology[2], National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, USA

[3]Department of Bacteriology, Okayama University Medical School, 2-5-1 Shikata-cho, Okayama 700-8558, Japan

Received 20 June 2003
Revised 29 August 2003
Accepted 2 September 2003

Botulism food poisoning is caused primarily by ingestion of the *Clostridium botulinum* neurotoxin (BoNT). The 1300 amino acid BoNT forms a progenitor toxin (PTX) that, when associated with a number of other proteins, increases its oral toxicity by protecting it from the low pH of the stomach and from intestinal proteases. One of these associated proteins, HA1, has also been suggested to be involved with internalization of the toxin into the bloodstream by binding to oligosaccharides lining the intestine. Here is reported the crystal structure of HA1 from type C *Clostridium botulinum* at a resolution of 1·7 Å. The protein consists of two β-trefoil domains and bears structural similarities to the lectin B-chain from the deadly plant toxin ricin. Based on structural comparison to the ricin B-chain lactose-binding sites, residues of type A HA1 were selected and mutated. The D263A and N285A mutants lost the ability to bind carbohydrates containing galactose moieties, implicating these residues in carbohydrate binding.

INTRODUCTION

The botulinum neurotoxin (BoNT) produced by *Clostridium botulinum* is one of the most potent toxins known. Exposure to the toxin can result in botulism food poisoning leading to blepharoptosis, dysphasia followed by dyspnoea and finally death (Lund, 1990). *C. botulinum* strains produce seven immunologically distinct neurotoxins labelled A–G (Sakaguchi *et al.*, 1984). Human botulism is primarily caused by the strains that produce toxin types A, B, E and F. Types C and D strains for the most part cause botulism only in non-human species (bovine, mink and bird) (Shapiro *et al.*, 1998). The neurotoxins (A–G) block the release of acetylcholine at neuromuscular junctions and synapses by cleaving protein components of the neuroexocytosis apparatus (Jahn & Niemann, 1994; Montecucco & Schiavo, 1994). In culture fluids these neurotoxins associate with non-toxic components to form larger complexes that are designated progenitor toxins (PTXs). Three different sizes of PTXs are produced by type A strains: 19S toxin (900 kDa), 16S toxin (500 kDa) and 12S toxin (300 kDa). Type B, C and D strains produce both the 16S and the 12S toxins, whereas type E and F strains produce only the 12S toxin. The type G strain produces only the 16S toxin (Sakaguchi *et al.*, 1984). The non-toxic component of 12S toxin was characterized and shown to lack haemagglutination activity, while the 19S and 16S toxins were shown to contain the 12S toxin plus a haemagglutinin component (HA) consisting of four proteins: HA1, HA2, HA3a and HA3b (Fujinaga *et al.*, 1994; Inoue *et al.*, 1996). Thus, the 19S and 16S toxins are referred to as HA-positive progenitor toxins (HA^+-PTXs). The non-toxic components increase the oral toxicity of the neurotoxin by protecting the neurotoxin from the acidic condition in the stomach and protease digestion in the intestine (Sugii *et al.*, 1977). In addition to protecting the structural integrity of the neurotoxin, HA may increase the internalization of the neurotoxin into the bloodstream of the host (Fujinaga *et al.*, 1997, 2000); however, it is not clear which protein of the HA complex directly interacts with the neurotoxin. Depending on the strain (A–G), different components of HA may play more dominant roles. It has been demonstrated that HA of type C 16S toxin is involved in binding to the small intestine of guinea pigs through interactions with glycolipids and glycoproteins containing sialic acid moieties (Fujinaga *et al.*, 1997; Inoue *et al.*, 1999). This

†These authors made equal contributions to this work.

Abbreviations: ACA, *Amaranthus caudatus* agglutinin; BoNT, botulinum neurotoxin; Gal, galactose; ITC, isothermal titration calorimetry; PTX, progenitor toxin.

binding ability has been attributed to HA3b (Fujinaga *et al.*, 2000). Type A HA^+-PTX does not appear to bind carbohydrates containing sialic acid but instead binds mainly glycolipids and glycoproteins containing β-D-Gal-(1→4)-D-GlcNAc (*N*-acetyllactosamine). This recognition has been attributed to the HA1 component (Inoue *et al.*, 2001). These differences may contribute to strain specificity for a particular host.

To better understand how type A HA1 interacts with sugars and thus may be involved in internalization of the neurotoxin, the techniques of isothermal titration calorimetry (ITC) and protein crystallography have been employed. Although we were unable to obtain diffraction quality crystals of type A HA1, we were able to crystallize and determine the X-ray structure of type C HA1. Type C HA1 is 37 % identical in sequence to type A HA1 and thus should provide a good structural model for type A HA1. Type C HA1 consists of two β-trefoil domains and bears structural similarities to the crystal structure of the ricin B-chain with lactose bound. ITC analysis was successfully applied to characterize carbohydrate-binding specificity of type A HA1, suggesting a single binding site with preference for galactose (Gal) and oligosaccharides containing Gal at the non-reducing end. Based on the sequence alignment of type A HA1 to type C HA1 and the structural comparison of type C HA1 to the ricin B-chain lactose-binding sites, site-directed mutagenesis was performed on type A HA1 to determine the carbohydrate-binding site. ITC analysis of these mutants revealed that the D263A and N285A mutants completely abrogated carbohydrate binding, implicating these residues in carbohydrate binding.

METHODS

Expression and purification of type C HA1. Type C HA1 was amplified from a previously described template using the 5′ PCR amplification primer 5′-CCG CGT GGA TCC ATG TCT CAA ACA AAT GCA AAT-3′ which adds a *Bam*HI restriction site to the 5′ end (Fujinaga *et al.*, 1994; Tsuzuki *et al.*, 1990). An *Xho*I site was added to the 3′ end using the 3′ PCR amplification primer 5′-CGG CCG CTC GAG TTA TAT TAA ATT TAT AAT CAT-3′. The PCR product was ligated into the pGEX-4T-3 vector (Amersham Pharmacia Biotech) containing the glutathione-*S*-transferase (GST) coding sequence using the *Bam*HI and *Xho*I restriction sites. A 10 × histidine tag was inserted on the N-terminal end of HA1 (between GST and HA1), using the oligomers 5′-GAT CCC ATC ATC ATC ATC ATC ATC ATC ATC ATC ATA-3′ and 5′-GAT CTA TGA TGA TGA TGA TGA TGA TGA TGA TGA TGG-3′. These oligomers were phosphorylated by polynucleotide kinase and annealed. The pGEX-4T-3 cloned with HA1 was treated with *Bam*HI and ligated with annealed oligomers. The nucleotide sequence of the clone was determined using pGEX 5′ and 3′ primers and found to be identical to the published sequence.

The HA1–GST fusion protein was expressed in *Escherichia coli* BL21(DE3) cells (Invitrogen). The transformed bacteria were grown in Luria–Bertani (LB) medium containing 100 μg ampicillin ml^{-1} at 37 °C. Growth was monitored by optical density measurements at 600 nm; at an OD_{600} value of 0·8, expression was induced with 0·1 mM IPTG. The cells were grown overnight at room temperature in 18 l of LB medium and harvested by centrifugation (5000 ***g***, 20 min). Cells were resuspended in PBS, sonicated and spun down at 38 000 r.p.m. for 30 min at 4 °C. The supernatant was collected and incubated with Glutathione Sepharose 4B (Amersham Pharmacia Biotech) resin (1 ml resin per 1 l induced cell culture), equilibrated with PBS and shaken gently for 30 min at 4 °C. The resin was washed five times in batch with a volume of PBS 25 times that of the resin for each wash. The fusion protein was cleaved on the resin with 50 units of thrombin (Sigma) overnight at room temperature. The supernatant/Sepharose mixture was poured into a column with a fritted filter and the resin was washed with PBS until the OD_{280} value became less than 0·1. The eluted solution was applied to a Benzamidine Sepharose 6B (Amersham Pharmacia Biotech) column to remove thrombin. The resulting solution was collected and then applied onto a Ni-NTA agarose (Qiagen) column pre-equilibrated with 20 mM Tris/HCl (pH 7·5) containing 0·1 M NaCl. After washing the column with equilibration buffer, bound protein was eluted using a 0–500 mM imidazole gradient. Fractions that contained pure HA1 were pooled, dialysed against 20 mM MES (pH 5·5), concentrated to 25 mg ml^{-1} and stored at −80 °C.

Expression and purification of type A HA1. The expression vector pGEX-5X-3 type A HA1 constructed previously (Fujinaga *et al.*, 2000) was used as a template. Primers 5′-CCGCGTGGATCCATGGAACACTATTCAGTAATC-3′ and 5′-CGGCCGCTCGAGTTATGGGTTACGAATATTCCA-3′ were designed for amplification with PCR. The product was restricted with *Bam*HI and *Xho*I and, after purification by agarose gel electrophoresis, was inserted into the expression vector pGEX-4T-3 and transformed into *E. coli* BL21(DE3) cells. Using a similar protocol as for type C HA1, the protein was expressed and then purified on Glutathione Sepharose 4B resin. The protein was eluted from the resin by overnight thrombin cleavage (6 units per millilitre of resin). The eluted protein was then dialysed overnight against 10 mM phosphate buffer, pH 5·5. All type A HA1 mutants were purified using the same protocol as the wild-type protein.

Site-directed mutagenesis of type A HA1. Site-directed mutagenesis was performed using the QuickChange Site-Directed Mutagenesis kit (Stratagene) following the protocols described in the accompanying instruction manual. The primers used were as follows: for D171A mutation, 5′-GTCGTACAACAAGTGGCTGTGACAAATCTAAAT-3′ and 5′-ATTTAGATTTGTCACAGCCACTTGTTGTACGAC-3′; for N187A mutation, 5′-TGGGACTATGGTCGCGCTCAAAAATGGACAATT-3′ and 5′-AATTGTCCATTTTTGAGCGCGACCATAGTCCCA-3′; for D263A mutation, 5′-ACAACTAAAGCTCTAGCTTTATATGGCGGCCAA-3′ and 5′-TTGGCCGCCATATAAAGCTAGAGCTTTAGTTGT-3′; for N285A mutation, 5′-TATCATGGAGATGATGCTCAGAAATGGAATATT-3′ and 5′-AATATTCCATTTCTGAGCATCATCTCCATGATA-3′. Sequences were confirmed, following mutagenesis, with the Big Dye Terminator Cycle Sequencing Reaction Kit (Applied Biosystems) in accordance with the protocol found in the accompanying instruction manual.

ITC. ITC measurements were carried out using a VP-ITC MicroCalorimeter (Micro Cal) at 30 °C. Solutions containing carbohydrate concentrations of 60 mM were injected into a reaction cell containing about 200 μM protein. Thirty injections of 6 μl at 180 s intervals were performed. Data acquisition and analyses were performed using the MICROCAL ORIGIN software package.

Crystallization of type C HA1. Crystals of type C HA1 were obtained using the vapour-diffusion hanging-drop method. Purified protein was mixed with equal volumes of a reservoir solution consisting of 0·1 M MES (pH 5·5), 100 mM $MgCl_2$ and 10 % PEG-8000 and placed above the reservoir. For data collection, harvested crystals were transferred to a cryoprotectant consisting of 0·1 M

Table 1. Crystallographic data statistics

Data set	Native 2*	Native 1	$HgCl_2$ derivative
Unit cell dimensions	a = 138·17, b = 62·33, c = 82·32 $\alpha = 90^\circ$, $\beta = 104{\cdot}53^\circ$, $\gamma = 90^\circ$	a = 138·57, b = 62·49, c = 82·68 $\alpha = 90^\circ$, $\beta = 104{\cdot}42^\circ$, $\gamma = 90^\circ$	a = 139·01, b = 62·57, c = 82·11 $\alpha = 90^\circ$, $\beta = 104{\cdot}29^\circ$, $\gamma = 90^\circ$
Space group	C2	C2	C2
No. of observations	203 352	119 630	140 918
Unique reflections	71 843	31 746	59 824
Rsym (%) (last shell)†	4·2 (15·0)	4·2 (11·5)	3·3 (6·7)
I/σI (last shell)	23·5 (8·0)	23·2 (4·9)	32·8 (9·3)
Mosaicity	0·31	0·50	0·58
Completeness (%) (last shell)	96·4 (87·6)	96·7 (94·7)	93·5 (89·1)
Refinement statistics			
Resolution (Å)	25·0–1·7	25–2·25	25·0–2·25
Rcryst (%)‡	17·7		
Rfree (%)§	20·1		
No. of waters	651		
Mean B value (Å)	17·2		
Root mean square deviation from ideal values		**Phasing statistics**	
Bond length (Å)	0·005	No. of heavy atom sites	6
Bond angle (°)	1·4	Riso (%)‖	27·9
Dihedral angle (°)	25·1	Rcullis (ano)¶	0·69 (0·80)
Improper angle (°)	0·67	Phasing power#	1·81
Ramachandran statistics			
Residues in:			
Most favoured regions (%)	87·0		
Additionally allowed regions	12·8		
Generously allowed regions	0·2		
Disallowed regions	0·0		

*This data set was collected with the crystal soaked in lacto-*N*-neo-tetraose; however, no density was visible for this molecule.

†Rsym = $\Sigma(|I_i - \langle I \rangle|)/\Sigma(I_i)$, where I_i is the intensity of the ith observation and $\langle I \rangle$ is the mean intensity of the reflection.

‡Rcryst = $\Sigma||Fo| - |Fc||/\Sigma|Fo|$ calculated from working data set.

§Rfree is calculated from 5 % of data randomly chosen not to be included in refinement.

‖Riso = $\Sigma||F_{PH}| - |F_P||/\Sigma|F_P|$, where F_{PH} and F_P are the derivative and native structure factors, respectively.

¶Rcullis = $\Sigma||F_{PH} \pm F_P| - F_{Hcalc}|/\Sigma|F_{PH} \pm F_P|$.

#Phasing power = $\Sigma \langle F_h \rangle / E$, where $\langle F_h \rangle$ is the root mean square heavy atom structure factor and E is the residual lack of closure error.

MES (pH 5·5), 100 mM $MgCl_2$, 12 % PEG-8000 and 15 % ethylene glycol and then flash-frozen at the detector in a stream of nitrogen cooled to −170 °C. Crystals were transferred to the cryosolution in four steps of increasing ethylene glycol and PEG concentration. For the $HgCl_2$ derivatized crystal, the crystal was soaked in the reservoir solution containing saturated $HgCl_2$ (< 1 mM) overnight then back-soaked into the same cryoprotectant used for native crystals.

The higher resolution native dataset (native 2) was collected from a crystal soaked in the 11 mM tetra-saccharide lacto-*N*-neo-tetraose; however, no electron density was detected for the tetra-saccharide. All datasets were collected with an RAXIS IV area detector system with Yale/MSC mirrors using a RU3H Rigaku generator. Data were indexed and integrated with DENZO and scaled using SCALEPACK (Otwinowski & Minor, 1997) (Table 1). Phases were obtained using the single isomorphous replacement with anomalous scattering (SIRAS) technique. SHELX was used to determine the position of the heavy metal sites (Sheldrick *et al.*, 1993), MLPHARE was used to calculate phases, and DM was used to improve the quality of the electron density map (Bailey, 1994). Automated model building using WARP produced half of the model (Perrakis *et al.*, 1999). Two molecules were found in the asymmetric unit. Both molecules contain electron density for residues 4–286 out of the 286 expected and both have good geometry as determined by PROCHECK (Laskowski *et al.*, 1993). All programs, with the exception of DENZO and SCALEPACK, were from the CCP4 interactive package (Bailey, 1994).

RESULTS AND DISCUSSION

Overall description of type C HA1 crystal structure

The crystal structure of type C HA1 is composed of two β-trefoil domains linked by an α-helix (McLachlan, 1979; Murzin *et al.*, 1992; Notenboom *et al.*, 2002; Rutenber *et al.*, 1987) (Fig. 1a). The β-trefoil fold consists of a six-stranded anti-parallel β-barrel capped on one end by three β-hairpins (Fig. 1a). The formation of this structure is based on a repeating subdomain motif in which four β-strands, separated by three variable sized loops, come together to form two anti-parallel β-sheet fragments (Fig. 1b, c). The first and fourth strands in this motif form part of the β-barrel, while the second and third strands form an independent β-hairpin. The three variable-sized loops separating the four strands occasionally possess secondary structure elements, but generally exhibit no characteristic sequence or structure. While it is these three loops that lend the name 'trefoil' to the fold, it is understandably easy to confuse the name with the fact that three of these motifs are repeated around a pseudo-threefold

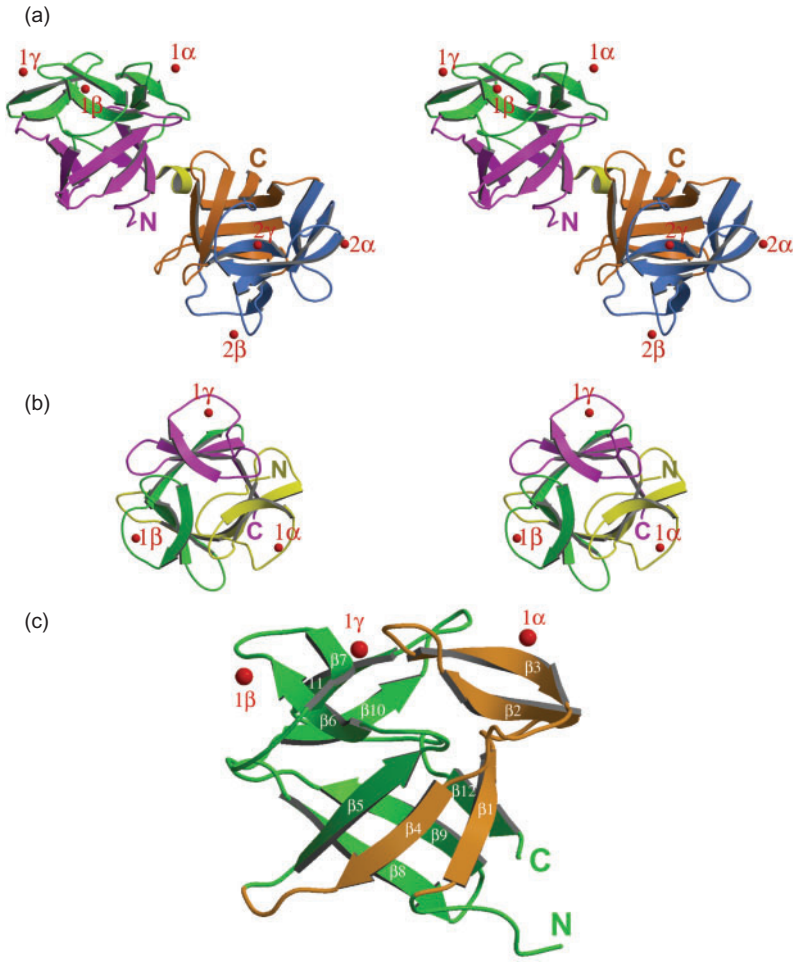

Fig. 1. (a) Stereo ribbon diagram of the crystal structure of HA1 displaying the N-terminal β-trefoil domain in magenta and green (residues 4–141) and the C-terminal β-trefoil domain in orange and blue (residues 146–286). The two domains are connected by a short α-helix (yellow). Red dots are positioned at the possible carbohydrate-binding positions for each trefoil repeat based on the structures of the ricin B-chain and the xylan-binding domain of xylanase. The β-barrel of the two trefoil domains are coloured magenta and orange, and the β-hairpin and loops of the repeats are coloured green and blue. (b) Stereo ribbon diagram of the N-terminal domain of HA1 looking down the pseudo-threefold axis. The three repeats are coloured yellow (1α), green (1β) and magenta (1γ). (c) Ribbon diagram of N-terminal β-trefoil domain of HA1 with the 1α repeat shown in orange and the rest of the domain shown in green. Each secondary structural element is labelled. These images were created using MOLSCRIPT and RASTER3D (Kraulis, 1991; Merritt & Bacon, 1997).

axis to form the entire β-trefoil domain (Fig. 1b). The three motifs are termed α-, β- and γ-repeats in sequence from the N- to the C-terminal end of the domain.

In the type C HA1 structure presented here, the first or N-terminal β-trefoil domain is made up of residues 4–141 and the second, C-terminal β-trefoil is made up of residues 146–286. The two domains are connected by a short α-helix (Fig. 1a). Superposition of the two β-trefoils reveals a root mean square deviation of 1·4 Å for 121 common Cα atoms. Based on this structural alignment, the two domains share 22 % sequence identity, suggesting that the protein is the result of gene duplication as has been suggested for other proteins containing two β-trefoil domains (Rutenber *et al.*, 1987; Transue *et al.*, 1997).

The β-trefoil domain is a structural fold found in other proteins including lectins such as the ricin B-chain (Rutenber *et al.*, 1987) and *Amaranthus caudatus* agglutinin (ACA) (Transue *et al.*, 1997), cytokines such as fibroblast growth factor (FGF) (Faham *et al.*, 1996; Ornitz *et al.*, 1995; Zhang *et al.*, 1991) and interleukin-1 (IL-1) (Finzel *et al.*, 1989), Kunitz trypsin inhibitors such as that from soybean (Sweet *et al.*, 1974), the actin cross-linking protein hisactophilin (Habazettl *et al.*, 1992), the xylan-binding domain of xylanase 10A (XBD-10A) (Notenboom *et al.*, 2002), the cysteine-rich domain of the mannose receptor (Liu *et al.*, 2000), and even in the BoNT structures (Lacy *et al.*, 1998; Swaminathan & Eswaramoorthy, 2000). In many instances, the β-trefoil domain acts as an oligosaccharide-binding unit supporting the suggestion that the β-trefoil fold itself is a result of a gene triplication of an ancient carbohydrate-binding peptide (the four-strand trefoil motif) of approximately 40 residues (Rutenber *et al.*, 1987).

Position, orientation and interactions between β-trefoil domains and carbohydrates vary among different proteins. In the case of ligand binding to the BoNT/B protein, the majority of the interactions between the protein and the sialylactose ligand are with the sialic acid substituent (Swaminathan & Eswaramoorthy, 2000). Binding of a substituent also dictates carbohydrate binding to the cysteine-rich domain of the mannose receptor where the majority of the interactions between ligand and protein are backbone amides forming hydrogen bonds with a sulfate moiety (Liu *et al.*, 2000). In addition, sulfate moieties are responsible for heparan sulfate binding to the FGF1 β-trefoil protein. In the crystal structure of the FGF1–FGFR2–heparan sulfate complex, the heparan sulfate molecule binds two FGF1s to create the FGF1 dimer required for binding to the FGFR2 (Pellegrini *et al.*, 2000). The non-toxic plant lectin ACA, like HA1, contains two β-trefoils per subunit. This protein is a dimer in solution even when carbohydrate is not bound. ACA contains two sugar-binding sites specific for the T-disaccharide [β-D-Gal-(1→3)-D-GalNAc] and

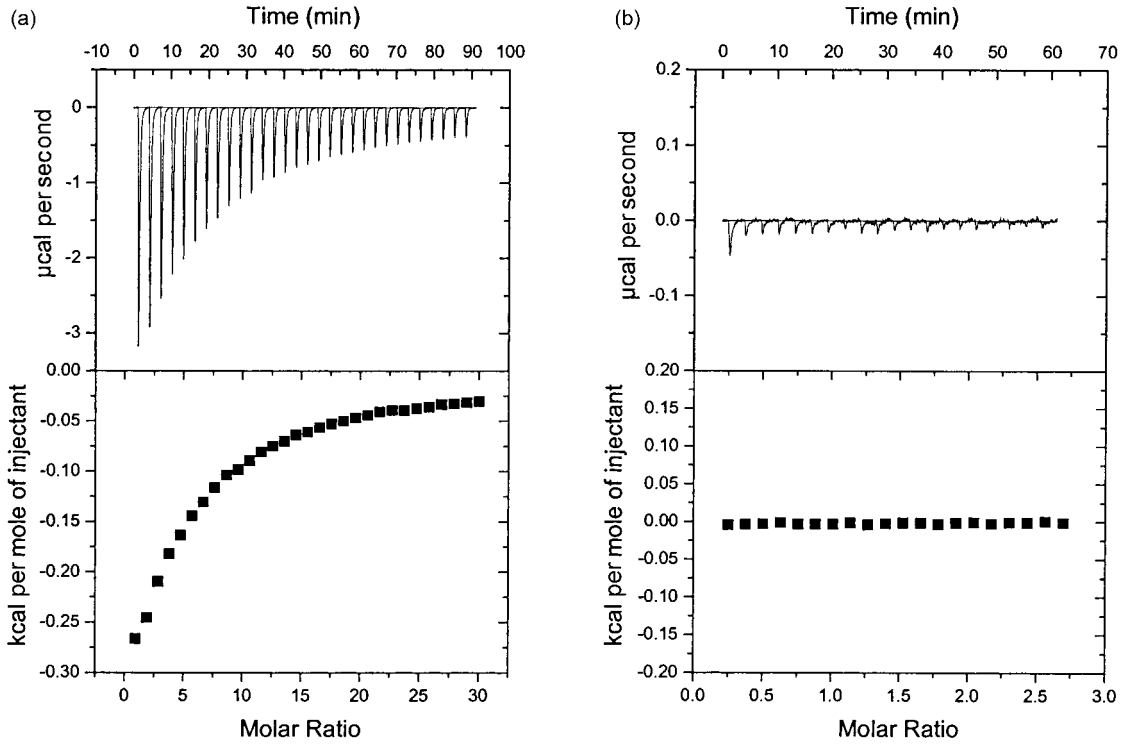

Fig. 2. Titration microcalorimetric data with lactose (60 mM) in the presence (a) and absence (b) of type A HA1 (200 µM). Upper panel, raw data obtained from 25 injections, 6 µl each of lactose. Lower panel, integrated curve showing experimental points (■). 1 cal = 4·184 J.

Table 2. Ligand specificity for type A HA1

Carbohydrate*	ΔH	ΔS	K_d (mM)
β-D-Gal-(1→4)-α-D-Glc	−1743 ± 128·9	−5·282	3·9
β-D-Gal-(1→4)-β-D-Glc	−3863 ± 42·40	−1·392	3·3
α-D-Gal-(1→3)-D-Gal	−5606 ± 162·8	−9·067	8·7
α-D-Gal-(1→3)-β-D-Gal-(1→4)-GlcNAc	−5626 ± 491·1	−8·446	6·2
Stachyose tetrahydrate	−4323 ± 230·1	−4·665	8·0
Gal	−4130 ± 21·45	−3·569	6·3
GalNAc	−9900 ± 197·2	−22·63	6·5
Glc	ND	ND	ND
Fru	ND	ND	ND
GlcA	ND	ND	ND
GlcNAc	ND	ND	ND

ND, Not detected.

*β-D-Gal-(1→4)-α-D-Glc, α-lactose; β-D-Gal-(1→4)-β-D-Glc, β-lactose; GalNAc, *N*-acetylgalactosamine; Glc, glucose; Fru, fructose; GlcA, glucuronic acid; GlcNAc, *N*-acetylglucosamine.

related by dimer symmetry. The 1β- and 1γ-repeats from one molecule and the 2γ-repeat from the dimer-related molecule make contacts with the flat sides of the sugars to make up one binding site (Transue *et al.*, 1997).

Lactose binding to β-trefoil domains has been observed in the ricin B-chain and the XBD-10A structures (Notenboom *et al.*, 2002; Rutenber & Robertus, 1991). Like HA1 and ACA, the ricin B-chain comprises two β-trefoil domains. In this structure, each of the β-trefoil domains (1α and 2γ sites) binds one lactose molecule (Rutenber & Robertus, 1991). In both cases the Gal moiety forms hydrogen bonds with an Asp from β-strand 2 of the repeat, an Asn from just before β-strand 4 of the repeat, and forms van der Waals contacts with an aromatic residue from β-strand 3 of the repeat. Similar binding is also seen with lactose binding to XBD-10A (Notenboom *et al.*, 2002). The XBD-10A structure is a single β-trefoil that binds lactose by the Gal sugar at the α and γ positions in a manner similar to lactose binding to ricin B-chain. Interestingly, the reducing glucose sugar protruding from the γ-repeat is seen binding to the β-repeat of a crystal symmetry-related XBD-10A molecule. Thus, lactose is found binding at all three sites in the crystal structure.

ITC analysis of carbohydrate binding to type A HA1

Previous work demonstrated that lactose inhibited haemagglutinin activity of type A HA1(Inoue *et al.*, 2001), therefore ITC was utilized to investigate lactose binding to type A HA1. The heat released was measured during successive injections of lactose solution, and the integrated heats of binding were plotted against the molar ratio of lactose and type A HA1 to generate a binding isotherm. From this, the stoichiometry, affinity (dissociation constant K_d), entropy (ΔS) and enthalpy (ΔH) of the binding were determined. Heat was released as an indication of the

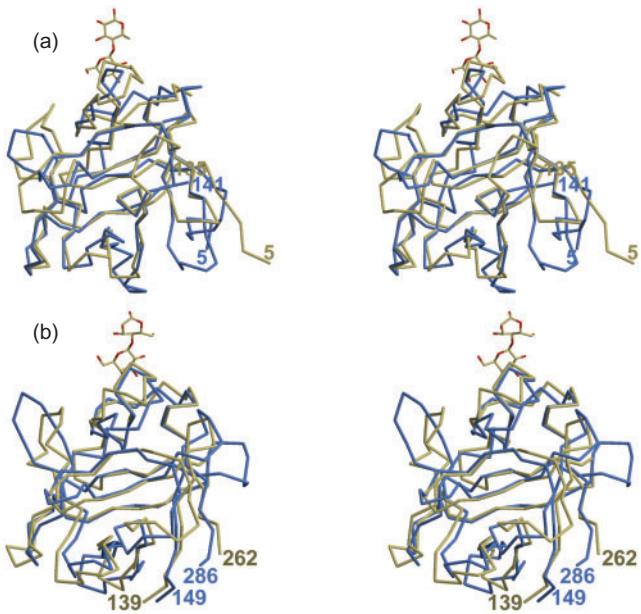

Fig. 3. (a) Stereo diagram of the superposition of the N-terminal β-trefoil domains of type C HA1 (blue) and ricin B-chain with lactose bound at the 1α-repeat (khaki). Superposition is based on 106 structurally equivalent Cαs with a root mean square deviation of 1·5 Å. (b) Stereo diagram of the superposition of the C-terminal β-trefoil domains of type C HA1 (blue) and ricin B-chain with lactose bound at the 2γ-repeat (khaki). This orientation is based on positioning the 2γ-repeat in the same orientation as the 1α-repeat in (a). Superposition is based on 115 structurally equivalent Cαs with a root mean square deviation of 1·8 Å. These images were created using MOLSCRIPT and RASTER3D (Kraulis, 1991; Merritt & Bacon, 1997).

binding of β-lactose to the type A HA1 protein (Fig. 2a). However, no heat was released either when lactose solution was successively injected into the buffer solution without type A HA1 (Fig. 2b) or when BSA was used instead of type A HA1 (data not shown). The stoichiometry of the heat release suggests that type A HA1 has a single binding site per molecule.

Binding studies with different mono- and oligosaccharides were performed and the results are summarized in Table 2. Only Gal or oligosaccharides containing Gal at the non-reducing end bound to type A HA1. The K_d values of HA1 to α- and β-lactoses were almost identical and the lowest among carbohydrates tested (Table 2). Gal, *N*-acetylgalactosamine and α-D-Gal-(1→3)-β-D-Gal-(1→4)-GlcNAc possessed slightly higher K_d values, followed by α-D-Gal-(1→3)-D-Gal and stachyose tetrahydrate. The other carbohydrates glucose, fructose, glucuronic acid and *N*-acetylglucosamine did not bind to HA1 (Table 2).

Ligand-binding site of type A HA1

Given the preference for lactose by type A HA1, it is possible that for at least type A, HA1 binding of carbohydrates could be more similar to ricin B-chain than to other β-trefoil-containing proteins. Based on 37 % sequence identity between type A HA1 and type C HA1, structurally these proteins are likely to be very similar. To determine which residues of type A HA1 may be interacting with carbohydrates, superpositions of type C HA1 domains were made to those of the ricin B-chain (Fig. 3a, b). These superpositions were then used to produce structural alignments of the repeats between ricin B-chain and type C HA1. Type A and B HA1 were then aligned based on sequence to type C HA1 (Fig. 4). From this figure it is observed that there are four conserved residues in ricin B-chain that form interactions with lactose in both the 1α- and 2γ-repeats. All the interactions between the ricin B-chain and the lactose molecules are with the Gal sugar. The conserved Asp residues (D22 from 1α, D234 from 2γ) from β-strand 2 and the Asn residues (N46 from 1α, N255 from 2γ) just prior to β-strand 4 form hydrogen bonds with the Gal, while the conserved aromatic residues (W37 from 1α, Y248 from 2γ) from β-strand 3 are likely involved in van der Waals contacts. Also conserved are Gln residues (Q47 from 1α, Q256 from 2γ) just after the conserved Asn residues that forms hydrogen bonds with the conserved Asp.

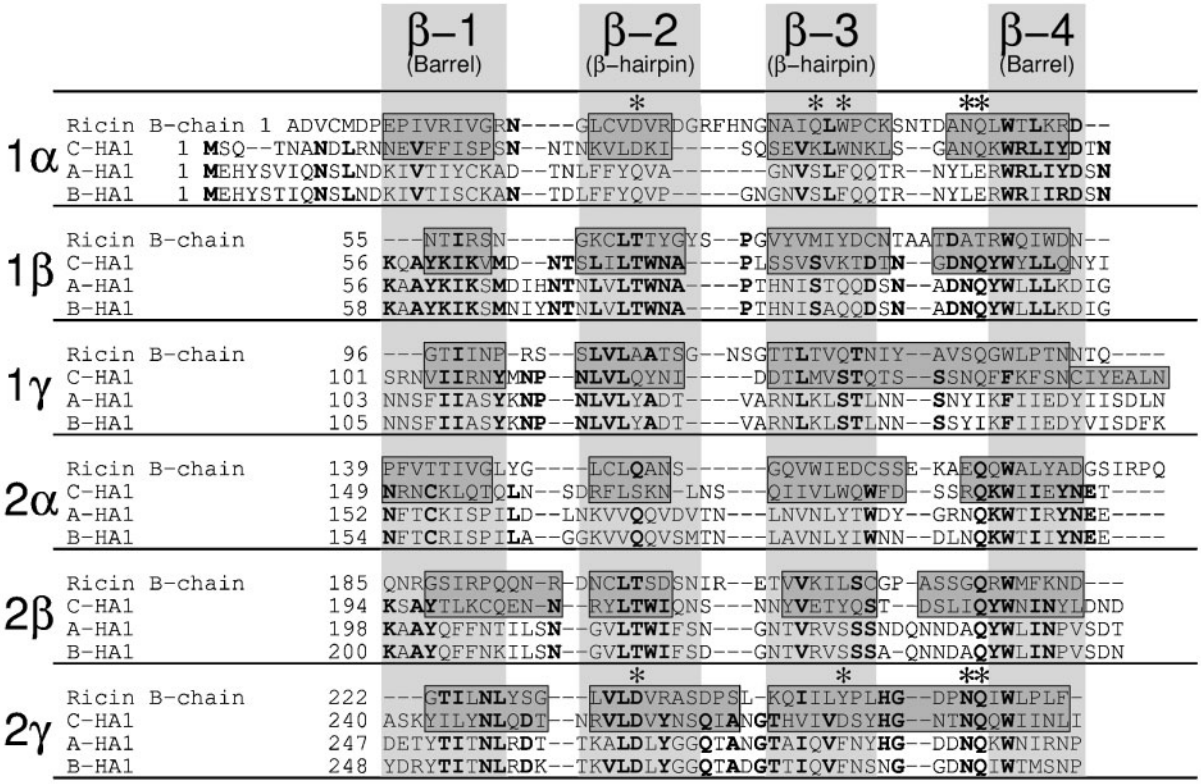

Fig. 4. Sequence alignments of the α-, β- and γ-repeats from type C HA1 domains with those of the ricin B-chain and with type A and B HA1 (type D HA1 is 100 % identical in sequence to C). Ricin B-chain and type C HA1 alignments are based on common residues (dark-grey boxes) after superposing each domain. Alignments of type A and B HA1 are based on sequence comparison to the type C protein. Grey columns define approximate residues of the four β-strands that comprise each repeat. The last seven residues in the 1γ row of HA1 (boxed) represent the α-helical linker between the two domains. Residues known to contact carbohydrate in ricin are marked with asterisks.

In the 1α site an additional residue, Gln45, also forms a hydrogen bond with the Gal. Although there are only two repeats in the ricin B-chain that interact with lactose, based on the structure, there is a possible binding site at each of the six repeats (1α, 1β, 1γ, 2α, 2β and 2γ).

For type C HA1, the 1α-repeat is a likely candidate for carbohydrate binding. The conserved Asp, Trp, Asn and Gln from the ricin B-chain are all present in this repeat (Fig. 4). Interestingly, a molecule of ethylene glycol used as a cryoprotectant for data collection is found binding in the 1α site of HA1 (Fig. 5a). Previously, it has been demonstrated that the cryoprotectant glycerol (which structurally is very similar to ethylene glycol) binds in two of the three lactose-binding sites in the β-trefoil domain of the XBD-10A (Notenboom *et al.*, 2002). This supports the suggestion that these compounds are good at mimicking carbohydrate binding to β-trefoil domains. The ethylene glycol molecule in type C HA1 superposes almost exactly with the O-3, C-3, C-4 and O-4 atoms of the Gal seen in the ricin B-chain structure. This suggests that this region in type C HA1 may accommodate its yet unidentified carbohydrate of specificity in much the same manner as observed for the ricin B-chain. For type A HA1, conservative substitutions exist at the conserved Asp (Gln) and Trp (Phe) sites in the 1α-repeat. However, the Leu at the conserved Asn site suggests this may not be a carbohydrate-binding site. Although both types A and C HA1 contain the conserved Asn at the 1β site, neither has the conserved Asp or Trp residues, making this an unlikely binding site for either. In addition, the only conserved residue for either type A or C HA1 in the 1γ-repeat is the conserved Asn in type C HA1. Although type C HA1 only has the conserved Trp in the 2α-repeat, type A HA1 contains the conserved Asn residue and conservative substitutions at the Asp (Gln) and Trp (Tyr) sites, making this a possible site for binding. Neither protein contains any conserved residues in the 2β-repeat; however, both show strong possibilities for binding at the 2γ-repeat. Type C HA1 contains the conserved Asp and Asn residues as well as an additionally conserved Gln after the Asn. Fig. 5(b) displays a superposition of type C HA1 2γ residues with those of ricin B-chain involved in making contacts with Gal. The 2γ lactose-binding site of the ricin B-chain (D234 and N255) appears to be conserved and superposes well with D256 and N278 (respectively) of type C HA1 (Fig. 5b). Type A HA1 contains these same conserved residues as type C, plus a conservative substitution of Phe at the Trp position and the additional Gln found hydrogen-bonding to Gal in the 1α-repeat of the ricin B-chain.

Because we do not know the specific ligand for type C HA1, it makes it difficult to test the possible binding sites based on the crystal structure. However, using site-directed mutagenesis and ITC analysis, we have tested for carbohydrate binding to type A HA1 at the 2α and 2γ sites which showed the highest sequence similarity to conserved residues interacting with lactose in the ricin B-chain binding sites (Table 3). Substitution of an Ala for the conserved Asn in the 2α-repeat (N187A) had no significant effect on the ability to bind lactose. However, both the D263A and N285A mutants at the 2γ-repeat abrogated lactose binding to type A HA1. This is consistent with the results of the ITC data that suggested one binding site for type A HA1. These results suggest that for type A HA1 the 2γ-repeat is involved in carbohydrate binding.

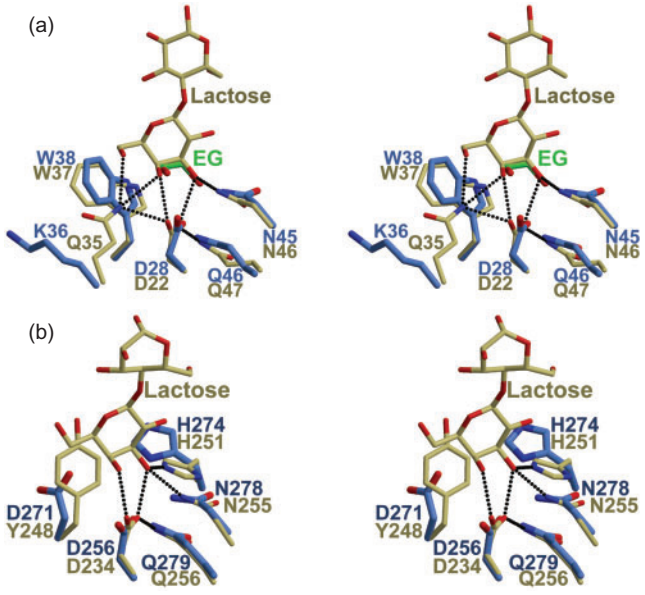

Fig. 5. (a) Stereo diagram of the superposition of residues from the 1α-repeat of type C HA1 (blue) binding ethylene glycol (green) to the residues from the 1α-repeat of ricin B-chain (khaki) binding lactose (khaki). (b) Superposition of residues from the 2γ-repeat of type C HA1 (blue) and ricin B-chain with lactose bound (khaki). Orientation of these figures are the same as seen in Fig. 3(a, b). These figures were created using MOLSCRIPT and RASTER3D (Kraulis, 1991; Merritt & Bacon, 1997).

Table 3. Mutational analysis of type A HA1 potential ligand-binding site

Type A HA1	Δ*H*	Δ*S*	K_d (mM)
Wild-type	−3 863 ± 42·40	−1·392	3·3
Asn285Ala	ND	ND	ND
Asp263Ala	ND	ND	ND
Asp171Ala	−4 388 ± 53·60	−3·144	3·3
Asn187Ala*	−12 270 ± 548·0	−30·20	5·6
Wild-type*	−11 360 ± 664·6	−26·90	4·8

ND, Not detected.
*Analysis performed at pH 8·5.

Final remarks

Although we have not yet obtained the crystal structure of a type C or a type A HA1 carbohydrate-bound complex, by comparing the crystal structure of type C HA1 to other lectins containing β-trefoil domains, and using ITC analysis coupled with site-directed mutagenesis, we have mapped the carbohydrate-binding site of type A HA1 to the 2γ-repeat. These results provide solid evidence that at least one of HA1's roles in the PTX is to bind oligosaccharides (Inoue *et al.*, 2001), supporting the hypothesis that HA1 may aid in internalization from the intestine of the BoNT.

REFERENCES

Bailey, S. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D* **50**, 760–763.

Faham, S., Hileman, R. E., Fromm, J. R., Linhardt, R. J. & Rees, D. C. (1996). Heparin structure and interactions with basic fibroblast growth factor. *Science* **271**, 1116–1120.

Finzel, B. C., Clancy, L. L., Holland, D. R., Muchmore, S. W., Watenpaugh, K. D. & Einspahr, H. M. (1989). Crystal structure of recombinant human interleukin-1β at 2·0 Å resolution. *J Mol Biol* **209**, 779–791.

Fujinaga, Y., Inoue, K., Shimazaki, S. & 8 other authors (1994). Molecular construction of *Clostridium botulinum* type C progenitor toxin and its gene organization. *Biochem Biophys Res Commun* **205**, 1291–1298.

Fujinaga, Y., Inoue, K., Watanabe, S., Yokota, K., Hirai, Y., Nagamachi, E. & Oguma, K. (1997). The haemagglutinin of *Clostridium botulinum* type C progenitor toxin plays an essential role in binding of toxin to the epithelial cells of guinea pig small intestine, leading to the efficient absorption of the toxin. *Microbiology* **143**, 3841–3847.

Fujinaga, Y., Inoue, K., Nomura, T., Sasaki, J., Marvaud, J. C., Popoff, M. R., Kozaki, S. & Oguma, K. (2000). Identification and characterization of functional subunits of *Clostridium botulinum* type A progenitor toxin involved in binding to intestinal microvilli and erythrocytes. *FEBS Lett* **467**, 179–183.

Habazettl, J., Gondol, D., Wiltscheck, R., Otlewski, J., Schleicher, M. & Holak, T. A. (1992). Structure of hisactophilin is similar to interleukin-1β and fibroblast growth factor. *Nature* **359**, 855–858.

Inoue, K., Fujinaga, Y., Watanabe, T., Ohyama, T., Takeshi, K., Moriishi, K., Nakajima, H. & Oguma, K. (1996). Molecular composition of *Clostridium botulinum* type A progenitor toxins. *Infect Immun* **64**, 1589–1594.

Inoue, K., Fujinaga, Y., Honke, K., Yokota, K., Ikeda, T., Ohyama, T., Takeshi, K., Watanabe, T. & Oguma, K. (1999). Characterization of haemagglutinin activity of *Clostridium botulinum* type C and D 16S toxins, and one subcomponent of haemagglutinin (HA1). *Microbiology* **145**, 2533–2542.

Inoue, K., Fujinaga, Y., Honke, K. & 7 other authors (2001). *Clostridium botulinum* type A haemagglutinin-positive progenitor toxin (HA^+-PTX) binds to oligosaccharides containing Galβ1–4GlcNAc through one subcomponent of haemagglutinin (HA1). *Microbiology* **147**, 811–819.

Jahn, R. & Niemann, H. (1994). Molecular mechanisms of clostridial neurotoxins. *Ann N Y Acad Sci* **733**, 245–255.

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* **24**, 946–950.

Lacy, D. B., Tepp, W., Cohen, A. C., DasGupta, B. R. & Stevens, R. C. (1998). Crystal structure of botulinum neurotoxin type A and implications for toxicity. *Nat Struct Biol* **5**, 898–902.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK – a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26**, 283–291.

Liu, Y., Chirino, A. J., Misulovin, Z., Leteux, C., Feizi, T., Nussenzweig, M. C. & Bjorkman, P. J. (2000). Crystal structure of the cysteine-rich domain of mannose receptor complexed with a sulfated carbohydrate ligand. *J Exp Med* **191**, 1105–1116.

Lund, B. M. (1990). Foodborne disease due to *Bacillus* and *Clostridium* species. *Lancet* **336**, 982–986.

McLachlan, A. D. (1979). Three-fold structural pattern in the soybean trypsin inhibitor (Kunitz). *J Mol Biol* **133**, 557–563.

Merritt, E. A. & Bacon, D. J. (1997). RASTER3D: photorealistic molecular graphics. *Methods Enzymol* **277**, 505–524.

Montecucco, C. & Schiavo, G. (1994). Mechanism of action of tetanus and botulinum neurotoxins. *Mol Microbiol* **13**, 1–8.

Murzin, A. G., Lesk, A. M. & Chothia, C. (1992). β-Trefoil fold. Patterns of structure and sequence in the Kunitz inhibitors interleukins-1β and 1α and fibroblast growth factors. *J Mol Biol* **223**, 531–543.

Notenboom, V., Boraston, A. B., Williams, S. J., Kilburn, D. G. & Rose, D. R. (2002). High-resolution crystal structures of the lectin-like xylan binding domain from *Streptomyces lividans* xylanase 10A with bound substrates reveal a novel mode of xylan binding. *Biochemistry* **41**, 4246–4254.

Ornitz, D. M., Herr, A. B., Nilsson, M., Westman, J., Svahn, C. M. & Waksman, G. (1995). FGF binding and FGF receptor activation by synthetic heparan-derived di- and trisaccharides. *Science* **268**, 432–436.

Otwinowski, Z. & Minor, V. (1997). Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307–326.

Pellegrini, L., Burke, D. F., von Delft, F., Mulloy, B. & Blundell, T. L. (2000). Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin. *Nature* **407**, 1029–1034.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* **6**, 458–463.

Rutenber, E. & Robertus, J. D. (1991). Structure of ricin B-chain at 2·5 Å resolution. *Proteins* **10**, 260–269.

Rutenber, E., Ready, M. & Robertus, J. D. (1987). Structure and evolution of ricin B chain. *Nature* **326**, 624–626.

Sakaguchi, G., Kozaki, S. & Ohishi, I. (1984). Structure and function of botulinum toxins. In *Bacterial Protein Toxins*, pp. 435–443. Edited by J. E. Alouf, F. J. Fehrenbach, J. H. Freer & J. Jelijasawicz. London: Academic Press.

Shapiro, R. L., Hatheway, C. & Swerdlow, D. L. (1998). Botulism in the United States: a clinical and epidemiologic review. *Ann Intern Med* **129**, 221–228.

Sheldrick, G. M., Dauter, Z., Wilson, K. S., Hope, H. & Sieker, L. C. (1993). The application of direct methods and Patterson interpretation to high-resolution native protein data. *Acta Crystallogr Sec D* **49**, 18–23.

Sugii, S., Ohishi, I. & Sakaguchi, G. (1977). Correlation between oral toxicity and in vitro stability of *Clostridium botulinum* type A and B toxins of different molecular sizes. *Infect Immun* **16**, 910–914.

Swaminathan, S. & Eswaramoorthy, S. (2000). Structural analysis of the catalytic and binding sites of *Clostridium botulinum* neurotoxin B. *Nat Struct Biol* **7**, 693–699.

Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H. & Blow, D. M. (1974). Crystal structure of the complex of porcine trypsin with

soybean trypsin inhibitor (Kunitz) at 2·6-Å resolution. *Biochemistry* **13**, 4212–4228.

Transue, T. R., Smith, A. K., Mo, H., Goldstein, I. J. & Saper, M. A. (1997). Structure of benzyl T-antigen disaccharide bound to *Amaranthus caudatus* agglutinin. *Nat Struct Biol* **4**, 779–783.

Tsuzuki, K., Kimura, K., Fujii, N., Yokosawa, N., Indoh, T., Murakami, T. & Oguma, K. (1990). Cloning and complete nucleotide sequence of the gene for the main component of hemagglutinin produced by *Clostridium botulinum* type C. *Infect Immun* **58**, 3173–3177.

Zhang, J. D., Cousens, L. S., Barr, P. J. & Sprang, S. R. (1991). Three-dimensional structure of human basic fibroblast growth factor, a structural homolog of interleukin 1β. *Proc Natl Acad Sci U S A* **88**, 3446–3450.