Crystal Structure of Botulinum Neurotoxin Type G Light Chain: Serotype Divergence in Substrate Recognition[†,‡]

Joseph W. Arndt, Wayne Yu, Fay Bi,[§] and Raymond C. Stevens*

Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received March 31, 2005; Revised Manuscript Received May 13, 2005

ABSTRACT: The seven serotypes (A−G) of botulinum neurotoxins (BoNTs) block neurotransmitter release through their specific proteolysis of one of the three proteins of the soluble *N*-ethylmaleimide-sensitive-factor attachment protein receptor (SNARE) complex. BoNTs have stringent substrate specificities that are unique for metalloprotease in that they require exceptionally long substrates (*1*). To understand the molecular reasons for the unique specificities of the BoNTs, we determined the crystal structure of the catalytic light chain (LC) of *Clostridium botulinum* neurotoxin type G (BoNT/G-LC) at 2.35 Å resolution. The structure of BoNT/G-LC reveals a C-terminal β-sheet that is critical for LC oligomerization and is unlike that seen in the other LC structures. Its structural comparison with thermolysin and the available pool of LC structures reveals important serotype differences that are likely to be involved in substrate recognition of the P1′ residue. In addition, structural and sequence analyses have identified a potential exosite of BoNT/G-LC that recognizes a SNARE recognition motif of VAMP.

Botulinum neurotoxins (BoNTs),[1] types A−G, cause the paralytic disease known as botulism by preventing neurotransmitter release through the specific cleavage of proteins of the neuorocytosis apparatus known as the soluble *N*-ethylmaleimide-sensitive-factor attachment protein receptor (SNARE). The causative agent of the toxicity is a 150 kDa multidomain protein that is processed by endogenous proteases into a 50 kDa light chain (LC) and a 100 kDa heavy chain (HC) that tightly associate via an interchain disulfide bond and extensive noncovalent interactions (*2*, *3*). The HC has two functional domains that facilitate selective binding to the neuronal presynaptic membrane and translocation of the LC following receptor-mediated endocytosis into the cytosol. The LC is a zinc metalloprotease that contains all of the molecular components for substrate recognition and proteolysis.

BoNT-LCs share significant sequence similarity ranging from 31% to 61% identity (*4*), yet each has exclusive substrate specificity. BoNT/A and BoNT/E cleave the synaptosomal-associated 25 kDa protein (SNAP-25); BoNT/B, BoNT/D, BoNT/F, and BoNT/G cleave the vesicle-associated membrane protein (VAMP), which is also referred to as synaptobrevin. BoNT/C is broader in its specificity and can cleave both SNAP-25 and syntaxin. BoNT-LCs are highly specific for their substrates but allow amino acid substitutions near the site of peptide bond cleavage with the exception of the amino acid adjacent to the scissile bond on the C-terminal side (P1′ position) (*5*−*7*). The high specificity of BoNTs has been attributed to additional substrate recognition sites that are remote from the catalytic active site that were identified in the crystal complex structures of BoNT/B-LC with VAMP (*8*) and BoNT/A1-LC with SNAP-25 (*9*), as well as SNARE secondary recognition (SSR) sites identified by mutagenesis studies (*10*−*12*).

If we are to understand the molecular basis for the unique specificities of BoNTs, it is essential to expose the differences in their structures that give rise to their unique characteristic. Therefore, structures of all serotypes are required, and toward achieving this goal we have determined the crystal structure of *Clostridium botulinum* neurotoxin type G LC (BoNT/G-LC) to 2.35 Å resolution. Comparison of BoNT/G-LC with the available neurotoxin structures reveals subtle yet distinct serotype variations that likely influence LC oligomerization and substrate recognition. This structure helps to elucidate the complete function of BoNT/G, as well as complements the structure collection of BoNT-LCs (serotypes A, B, and E) that is known.

EXPERIMENTAL PROCEDURES

Protein Production and Crystallization of BoNT/G-LC. Overexpression of BoNT/G-LC was induced in the presence of 1 mM $ZnCl_2$ with isopropyl α-thiogalactopyranoside (IPTG) under control of the T5 promoter in plasmid pBN13, which encodes an expression and purification tag consisting

[†] This work was supported by Contract DAMD17-00-C-0040 from the Department of the Army. Portions of this research were carried out at the General Medicine and Cancer Institutes Collaborative Access Team (GM/CA-CAT) beamlines of the Advanced Photon Source (APS). GM/CA-CAT has been funded in whole or in part with Federal funds from the National Cancer Institute (Y1-CO-1020) and the National Institute of General Medical Sciences (Y1-GM-1104). Use of the APS was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract W-31-109-ENG-38.

[‡] The coordinates for the structure have been deposited with the Protein Data Bank as entry 1ZB7.

* To whom correspondence should be addressed. Fax: (858) 784-9483. E-mail: stevens@scripps.edu.

[§] Current address: Princeton University, Princeton, NJ.

[1] Abbreviations: BoNT, botulinum neurotoxin; CV, column volumes; DTT, dithiothreitol; HC, heavy chain; IPTG, isopropyl α-thiogalactopyranoside; LC, light chain; SNAP-25, synaptosomal-associated 25 kDa protein; SNARE, soluble *N*-ethylmaleimide-sensitive-factor attachment protein receptor; SSR, SNARE secondary recognition; TCEP, tris(2-carboxyethyl)phosphine hydrochloride; TeNT, tetanus neurotoxin; TSA, transition state analogue; VAMP, vesicle-associated membrane protein.

10.1021/bi0505924 CCC: $30.25
Published on Web 06/18/2005

of APPTPGHHHHHH at the C-terminus of the LC protein. Bacteria were lysed by sonication in lysis buffer (50 mM K_2HPO_4, pH 7.8, 300 mM NaCl, 10% glycerol, 5 mM imidazole) supplemented with Roche EDTA-free protease inhibitor tablets and 0.5 mg/mL lysozyme. In addition, 0.25 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP) was added throughout the purification. Immediately after sonication, the cell debris was pelleted by ultracentrifugation. The soluble fraction was applied to a gravity flow metal chelation column (Talon resin charged with cobalt; Clontech) equilibrated in lysis buffer. The column was then washed with wash buffer (25 mM Tris, pH 7.8, 300 mM NaCl, 10% glycerol, 10 mM imidazole) and eluted with 3 column volumes (CV) of elution buffer (25 mM Tris, pH 7.8, 15 mM NaCl, 150 mM imidazole). The protein was then applied to a Poros 20 HQ column (Applied Biosystems) equilibrated in anion-exchange buffer (25 mM Tris, pH 7.8). The column was washed with equilibration buffer and eluted with a 10 CV gradient elution (0−500 mM NaCl in 25 mM Tris, pH 7.8). The protein was then concentrated to ~10 mg/mL by centrifugal ultrafiltration (Orbital), applied to a sizing column (Biocep SEC-S 3000, Phenomenex) equilibrated in size exclusion buffer (150 mM NaCl in 25 mM Tris, pH 7.4), and eluted. The purified protein was concentrated to 13 mg/mL (Orbital) and either frozen in liquid nitrogen for later use or used immediately for crystallization trials. The protein was crystallized using the nanodroplet vapor diffusion method (*13*) at a temperature of 293 K. The crystallization solution contained 6% PEG 6000, 0.65 M LiCl, and 100 mM sodium citrate at pH 5.5.

Activity Assay. BoNT solid-phase activity assay optimized for BoNT/B was performed with a substrate-coated plate immobilized with a fluorescein-derivatized VAMP peptide (residues 60−94) as described (*14*). Briefly, assays were conducted in triplicate in 50 mM Hepes at pH 7.4, 1 mM DTT, 0.25 mM $ZnCl_2$, and 0.05% Tween-20 with 90 μg/mL BoNT/G-LC and 1 μg/mL BoNT/B-LC as a positive control. Reactions were incubated on a substrate-immobilized plate at 37 °C for 5 h. Fluorescence of the cleavage reaction was measured with a GENios Pro fluorometer (Tecan) with excitation and emission wavelengths of 485 and 535 nm, respectively (Supporting Information, Figure S1).

Data Collection. Diffraction data were collected at the Advanced Photon Source (APS, Argonne, IL) on beamline GM/CA-CAT-23ID at a wavelength of 1.009 Å using the MAR data collection environment (Table 1). The data sets were collected at 100 K using a MAR 225 CCD detector. Data were integrated, reduced, and scaled using HKL2000 (*15*). The crystals were indexed in the hexagonal space group *P*622; data statistics are summarized in Table 1.

Structure Determination and Refinement. The structure was determined with MOLREP (*16*) using the two possible space groups, $P6_222$ and $P6_422$, predicted from the observed systematic absences. Only the $P6_222$ space group yielded the correct solution. ARP/wARP (*17*) was used to build 94% of the model followed by manual building of the rest of the structure with O (*18*). Structure refinement was performed using REFMAC5 (*19*). Refinement statistics are summarized in Table 1. The final model includes a single protein molecule, a citrate molecule in coordination to the active site zinc ion, and 197 water molecules. No electron density was observed for residues 53−67, 211−213, 252−256, and 439−445. Analysis of the stereochemical quality of the model was accomplished using the AutoDepInputTool (http://deposit.pdb.org/adit/). Figures were prepared with PyMOL (DeLano Scientific). Atomic coordinates and experimental structure factors of BoNT/G-LC have been deposited with the PDB and are accessible under the code 1ZB7.

Substrate Modeling. The potential S1′ binding subsite of BoNT/G-LC, as well as the other LC serotypes, for recognition of the P1′ residue was identified by comparing the structure of BoNT/G-LC with that of metalloprotease thermolysin (PDB code 4TMN) (*20*) bound to a transition state analogue (TSA). The modeling strategy involved superimposing the active site HExxH motif of this protein with that of BoNT/G-LC using the web server C-alpha Match (http://bioinfo3d.cs.tau.ac.il/c_alpha_match/).

The probable SSR site of BoNT/G-LC that is specific for the V2 SSR of VAMP was identified by comparison of BoNT/G-LC with the structure of the mutant E224Q/Y366F of BoNT/A1-LC in complex with residues 141−204 of SNAP-25 (PDB code 1XTG) (*9*) with the program TOP (*21*). A five-residue region containing the three essential acidic residues of the V2 SSR motif of VAMP (residues 64−68) was modeled to BoNT/G-LC based on the analogous residues in SNAP-25 (residues 180−184) that are 14 residues upstream from the proteolytic cleavage site.

RESULTS AND DISCUSSION

Structure of BoNT/G-LC. The structure of BoNT/G-LC (Figure 1A) was determined to 2.35 Å resolution by molecular replacement using the BoNT/B-LC structure [PDB code 1F82 (*8*) with 61% sequence identity] as the search model. Data collection, model, and refinement statistics are summarized in Table 1. The final model of the BoNT/G-LC structure is composed of 14 β-strands (β1−β14), eight α-helices (α1−α8), and four 3_{10}-helical segments that form a homodimer through crystallographic symmetry. Homodimeric structures are also found in the LC structures of BoNT/A2 (*22*) and BoNT/E (*23*), suggesting a potential importance of LC oligomerization. The dimensions of the BoNT/G-LC homodimer are 115 Å × 56 Å × 38 Å, with an overall surface area of 27800 Å^2. The overall conformation of the BoNT/G-LC monomer is similar to the LC structures of BoNT/A1 (*9*), BoNT/A2 (*22*), BoNT/B (*8*), and BoNT/E (*23*) with minor differences observed for the loop regions (25, 145, 200, 250, and 310 loops). However, two distinguishing features set it apart from the other LC structures, those being an extra β-sheet and a channel that connects the two active sites of the homodimer. The additional antiparallel β-sheet is located near the C-terminus and is composed of β-strands β4 (residues 69−73) and β14 (residues 429−436) (Figure 1A). This region is directly before the interchain disulfide bond of the holotoxin and contributes to the dimerization interface. A similar β-sheet is seen in the catalytic domain of BoNT/B holotoxin (*2*) between β-strands β4 (residues 71−72) and β20 (residues 429−430), though the β-sheet in the BoNT/G-LC structure is more extensive. Consistent with this observation, DasGupta reported that the single chain holotoxin significantly alters its structure upon nicking as observed in the increase in the prevalence of β-sheet from 37% to 41% (*24*). The BoNT/A1, BoNT/A2, and BoNT/B-LC structures lack this secondary structure element, likely

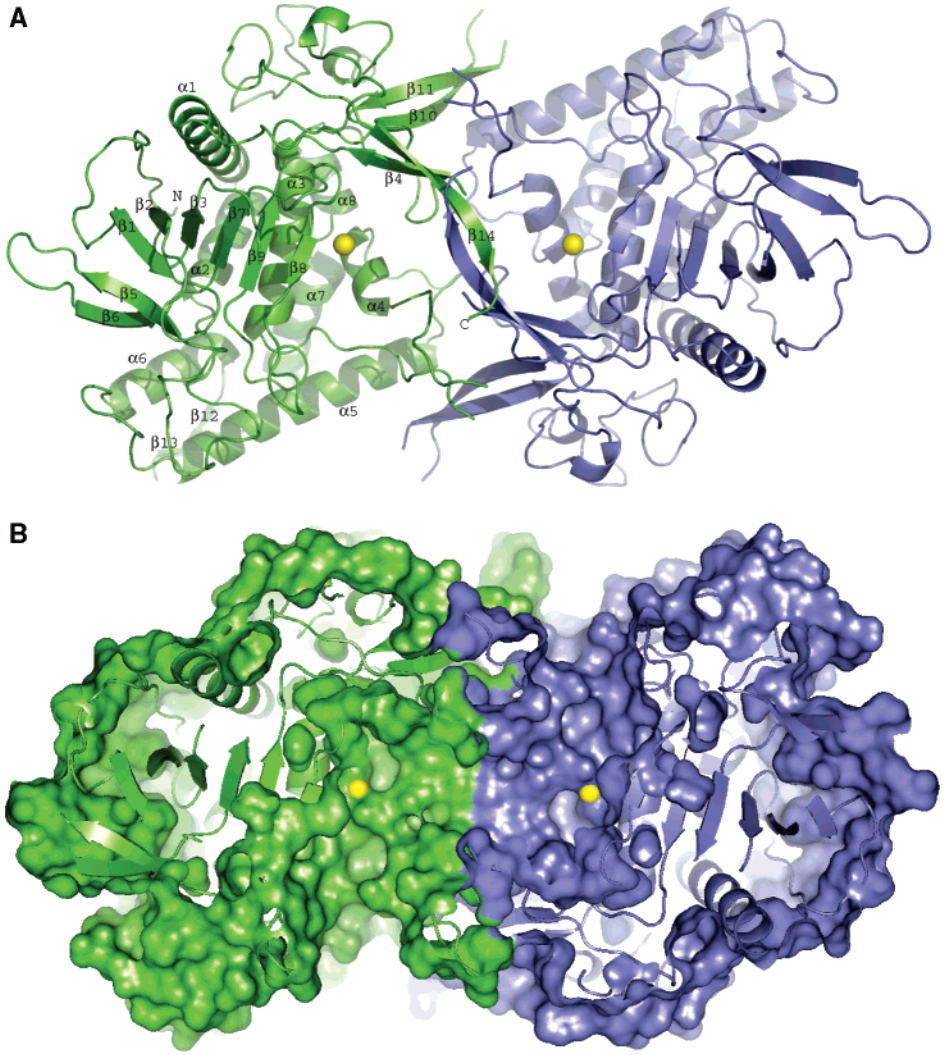

FIGURE 1: Crystal structure of the BoNT/G-LC homodimer. (A) Ribbon diagram of *C. botulinum* BoNT/G-LC with one monomer of the crystallographic dimer colored green and the second monomer in slate blue with N- and C-termini indicated. β-Strands ($\beta1-\beta14$) and α-helices ($\alpha1-\alpha8$) of one of the monomers are labeled. The active site zinc ion is shown as a yellow sphere. (B) Same as (A) but a cross section of the molecular surface of the BoNT/G-LC homodimer overlaid with the ribbon diagram looking down onto the serpentine channel that connects the two active sites.

because they are C-terminal truncated variants and not full-length LCs (Table 2).

The second structural feature unique to BoNT/G-LC is that the active sites are connected via a serpentine-like channel (Figure 1B) that is partially covered but not blocked by β-strand $\beta14$ and terminal residues of the C-terminus (residues 437−438) of each subunit (Figure 1A). The individual active sites of the BoNT/G-LC homodimer are solvent accessible and joined by a contiguous channel 28.5 Å in length. The channel takes an abrupt 90° turn at each of the active sites of the homodimer where it continues another ~17 Å along the groove primarily formed by the 175 and 250 loops. Intriguing as it may be, to our knowledge there is no evidence that suggests that these active sites work in a cooperative manner. Thus, further studies should be conducted to determine if there is a functional significance to the observed channel or if it is merely a packing artifact.

Evidence of the importance of the C-terminus in LC transport and proteolytic activity is increasingly mounting. Fernandez-Salas and colleagues recently reported that the N- and C-termini of BoNT/A- and BoNT/E-LCs are critical in its localization (*25*). In particular, they showed that mutations and deletions of the C-terminus of BoNT/A-LC suggest that it is likely involved in trafficking and/or interaction with membrane adaptor proteins through a diluecine motif ($\underline{E}^{423}XXX\underline{LL}^{428}$). A dileucine to dialanine mutant (L427A and L428A) revealed that disruption of this motif caused changes in the steady-state distribution of BoNT/A-LC in the plasma membrane, resulting in the mutant being distributed in the periplasmalemmal space instead of plasma membrane localized, as well as being 26-fold less active than the wild-type BoNT/A-LC. Moreover, a C-terminal truncated construct of BoNT/A-LC (ΔC22), lacking 22 C-terminal acids, displayed similar altered distribution with ~80-fold decreased activity. Consistent with this observation of reduced catalysis are the reports of ~10-fold decreased activity of C-terminal truncated LC variants of BoNT/A1 ΔC30 and ΔC50 (*26*) and BoNT/A2 ΔC32 (*27*), thus further implicating the significance of the LC C-terminus. Incidentally, the ΔC32 BoNT/A2-LC variant is the protein construct used to obtain the crystal structure of BoNT/A2 while the ΔC22 variant is similar in size to the ΔC18 construct used in the BoNT/A1 structure (Table 2). It is clear from these separate studies that the C-terminus can influence

Table 1: Summary of Crystallographic Parameters, Data Collection, and Refinement Statistics for BoNT/G-LC (PDB Code 1ZB7)

Data Collection	
space group	$P6_222$
unit cell parameters (Å)	$a = 179$, $b = 179$, $c = 81$
resolution range (Å)	50.00−2.35
highest resolution shell (Å)	2.43−2.35
no. of observations	225818
no. of reflections	30233
completeness (%)	93.9 (96.0)[a]
mean $I/\sigma(I)$	24.4 (4.8)[a]
R_{sym}[b] on I	0.053 (0.395)[a]
Model and Refinement Statistics	
resolution range (Å)	24.81−2.35
no. of reflections (total)	30226
no. of reflections (test)	1536
completeness (% total)	94.0
R_{cryst}[c]/R_{free}[d]	0.174/0.222
Stereochemical Parameters	
restraints (RMS observed)	
bond length (Å)	0.027
bond angle (deg)	2.03
protein residues/atoms	407/3324
solvent molecules	197
heterogen atoms	14

[a] Highest resolution shell. [b] $R_{sym} = \sum|I_i - \langle I_i\rangle|/\sum|I_i|$, where I_i is the scaled intensity of the ith measurement and $\langle I_i\rangle$ is the mean intensity for that reflection. [c] $R_{cryst} = \sum||F_o| - |F_c||/\sum|F_o|$, where F_c and F_o are the calculated and observed structure factor amplitudes, respectively. [d] R_{free} = as for R_{cryst}, but for 5.0% of the total reflections chosen at random and omitted from refinement.

Table 2: Summary of BoNT-LC Crystal Structures

LC type	PDB code	resolution (Å)	truncation construct[a]	C-terminus disorder	oligomeric state
A1 (*9*)	1XTF	2.2	ΔC18	none	monomer
A2 (*22*)	1E1H[b]	1.8	ΔN8, ΔC32	none	dimer
B (*8*)	1F82	2.2	ΔC19	none	monomer
E (*23*)	1T3A	2.1	full length	residues 1−10	dimer
G	1ZB7	2.35	full length	residues 1−5	dimer

[a] All LCs contain C-terminal His tags. [b] Contains 47 additional N-terminal residues consisting of His and S tags and a protease cleavage site.

activity. However, it is not yet apparent if different localization motifs are present in the other LC serotypes, since the dileucine motif is unique to BoNT/A, or if the C-terminal β-sheet is common to all BoNT serotypes, since the BoNT/E-LC structure is full length and it is helical in this equivalent position. Thus, further structural studies of other full-length BoNT-LCs are needed.

BoNT/G-LC Oligomerization. The crystal packing in the BoNT/G-LC structure shows a homodimer, which is noncovalent as determined by SDS−PAGE under nonreducing conditions, and is formed through mostly hydrophobic interactions (69% nonpolar) that associate via a 2-fold rotation. The dimer interface is composed of β-strands β10 and β11, the 250 loop, and the C-terminal β-strands β14 of each subunit, which form an interchain antiparallel β-sheet (Figure 1A). Additional interactions are formed around α-helix α5 and the 370 loop from each subunit. This intimate dimer interface accounts for a buried surface area of 2054 Å^2 (10.6% of the total surface area) for each monomer and includes 16 hydrogen bonds and 2 salt bridges (R217−D262 from each monomer) as determined by the Protein−Protein Interaction Server (http://www.biochem.ucl.ac.uk/bsm/PP/server/) (*28*). It is anticipated that the oligomerization state of BoNT/G-LC may be more biologically relevant than those seen in BoNT/A2-LC (*22*) and BoNT/E-LC (*23*), since its interface is more hydrophobic and considerably more extensive as indicated by the lower buried surface areas of BoNT/A2- and BoNT/E-LCs, 1512 and 1181 Å^2, respectively (*28*). Furthermore, the BoNT/G-LC structure has a quaternary structure similar to that of BoNT/A2-LC (PDB code 1E1H) (*22*) with equivalent structural elements forming its dimer interface, with the exception that the 250 loop of each of the subunits of BoNT/A2-LC covers the active site of the homodimer partner. This altered interface results in one of the subunits of BoNT/A2-LC being rotated (∼120°) relative to that of the BoNT/G-LC dimer. As a result, BoNT/A2-LC was observed to be autoproteolytic in solution and crystalline states (*22*). Analysis of a dissolved crystal of the BoNT/G-LC by SDS−PAGE yielded a single 50 kDa band indicating that the self-cleavage observed in the BoNT/A2-LC crystal does not occur in the BoNT/G-LC crystal (data not shown). Interestingly, crystal packing shows that the dimeric BoNT/G-LC is covalently linked to another dimer molecule by two interchain disulfides to form a homotetramer. The observed dimer−dimer interface is relatively small (1672 Å^2, 4.8% of the total surface area) (*28*, *29*) and, therefore, may be an artifact of the crystal packing. However, it could also suggest a model of redox-dependent oligomeric states like that seen with the EF-hand protein P11 (*30*), since there is evidence that the tetramer exists in solution, but only under oxidizing conditions, as determined by size exclusion chromatography. This issue of oligomerization is interesting, since it may be possible that some of the serotypes form higher order oligomeric states *in vivo*. The differences reported in oligomerization of the LC crystal structures may explain functional variations observed in the serotypes: for instance, the dissimilarity observed in duration of the therapeutic effects in the long-acting BoNT/A compared to the short-acting BoNT/E (*25*) by influencing the exposure (or protection) of labile regions of the LC, such as epitopes and loops, that are susceptible to immunoresponse and proteolytic digestion.

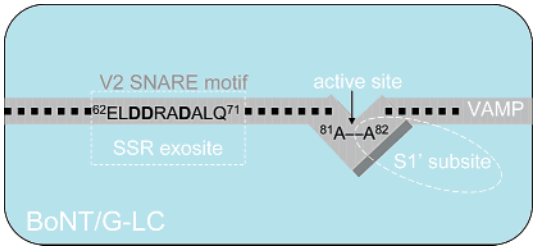

FIGURE 2: Schematic of the proposed interaction of BoNT/G-LC (light blue) and VAMP (gray) with the S1′ subsite and SSR exosite indicated. The cleavage site is identified with the arrow, and essential acidic residues of the V2 SSR are in bold.

Substrate Recognition. Each BoNT serotype is known to require substrates with a minimum length and proper sequence. The unique specificity of each of the serotypes is likely due to the fact that each LC possesses two or more substrate recognition sites as proposed for BoNT/G-LC with its substrate VAMP in Figure 2. The first consists of the subsites located near the active site that are required for discrimination of SNARE residues in proximity to the cleavage site, such as the strictly required P1′ substrate

residue. The second site, termed exosite, is far removed from the active site and is involved in recognition of the nine-residue SSR sequence motif. Substrate cleavage occurs only when the binding of the SNARE substrate to each of these sites takes place (*11*, *31*). Therefore, substrate specificity is likely to be a result of the spatial relationship between an exosite of the LC that recognizes a SSR motif and the active site. The active site is composed of one or more subsites including, but not necessarily limited to, the S1′ subsite for recognition of the essential P1′ amino acid on the different SNARE proteins (*8*, *11*). Hence, substrate proteolysis only occurs upon binding of the SSR motif that is the proper distance away from the cleavage site.

It has been very challenging to obtain LC−substrate crystal structures, and consequently, the details of substrate recognition are limited and sometimes contradictory. Indeed, our attempts to obtain a cocrystal structure of BoNT/G-LC have yet to be successful. Cocrystal structures of BoNT/A1 (*9*) and BoNT/B-LC (*8*) reveal that BoNTs bind their substrate SNARE proteins at a channel formerly occupied by 54 residues known as the "belt" region connecting the LC to the translocation domain, as previously proposed (*8*). Unfortunately, these structures do not offer insight into the intimate details of substrate recognition of the P1′ residue and the SSR motif. The BoNT/A1-LC (E224Q/Y366F double mutant) SNAP-25 complex structure does not allow the mapping of the S1′ subsite, since the substrate is disordered in the active site, nor does it reveal any details in the recognition of the conserved residues of the required SSR motif (*9*). Furthermore, the BoNT/B-LC VAMP complex structure has been the subject of debate (*32*, *33*) due to the low occupancy and atypical orientation of substrate binding as compared to other HExxH proteases. In the absence of a BoNT/G-LC VAMP complex structure, two probable recognition sites of BoNT/G-LC, a S1′ subsite and a SSR exosite, are identified on the basis of homology modeling and are discussed individually in the following subsections.

S1′ Subsite Comparison with Other Serotype LC Structures. Despite the unique substrate specificities of the BoNTs, their active sites are highly similar to each other. In BoNT/G-LC the active site zinc is coordinated by His 230 and His 234 of the HExxH motif and Glu 268 of α-helix α4, as does its counterparts in the other serotypes. However, the high active site residue conservation of BoNTs is not expected to extend to the S1′ subsite given the strict dependence for the P1′ residues of their SNARE substrates (*5*−*7*, *34*). The P1′ residues are chemically diverse; they vary in size and also range in increasing order of hydrophobicity from Arg (BoNT/A), Lys (BoNT/F), Ala (BoNT/C and BoNT/G), Phe (BoNT/B), Leu (BoNT/D), and Ile (BoNT/E) (*35*). Given the diverse chemical nature of the P1′ residues, it is expected that each BoNT serotype would have a complementary S1′ subsite for recognition of this structurally varied pool of P1′ residues of their SNARE substrates.

Even with the number of different neurotoxin LC structures now available (serotypes A, B, E, and G and tetanus), surprisingly the S1′ subsite of the BoNT-LCs has yet to be convincingly mapped. Here we attempt to define the S1′ subsite of BoNT/G-LC by a structural comparison with that of thermolysin, considered to be the prototypical HExxH motif protease, in complex with a TSA containing leucine at the P1′ residue. BoNT-LCs show significant structural

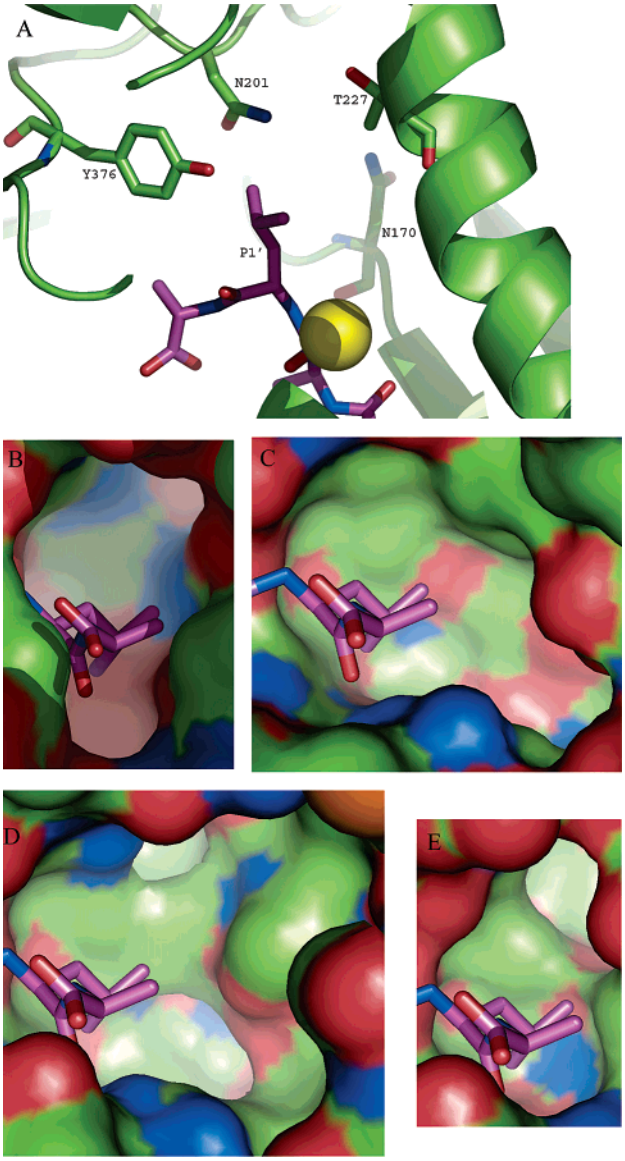

FIGURE 3: (A) Close-up view of the putative S1′ subsite of BoNT/G-LC with a modeled thermolysin transition state analogue (TSA). Putative S1′ binding site residues as observed in BoNT/G-LC (green) with a modeled thermolysin TSA (violet) with residue labels and the substrate P1′ residue indicated. (B) Same as (A) but a surface diagram looking down onto the S1′ subsite of BoNT/G-LC (green) with oxygens (red) and nitrogens (blue) with a modeled TSA (magneta). (C) Same as (B) but of the S1′ subsite of BoNT/A1-LC (*9*) with the C-terminal His tag not shown for clarity. (D) Same as (C) but of the S1′ subsite of BoNT/B-LC (*8*). (E) Same as (C) but of the S1′ subsite of BoNT/E-LC (*23*).

homology to thermolysin as indicated by the fact that BoNT/G-LC and thermolysin (PDB code 4TMN) (*20*) can be superposed with a C_α RMSD of 2.05 Å for 120 structurally equivalent C_α atoms. The superposition reveals that the P1′ leucine residue of the TSA is directed toward a shallow pocket believed to be the S1′ subsite of BoNT/G-LC that is composed of Asn 170, Asn 201, Thr 227, and Tyr 376 (Figure 3A). Importantly, LC structures of BoNT/A1, BoNT/B, and BoNT/E also have binding cavities at this same location, but their putative S1′ subsites are complementary in size and hydrophobicity to their cognate P1′ substrate residues (Table 3). The S1′ subsite of BoNT/G-LC (Figure 3B) is relatively small and shallow compared to the S1′ subsites of the other serotypes (Figure 3C−E) and, therefore,

Table 3: Residues Composing the Putative S1′ Subsite of BoNTs with Their SNARE Substrates

BoNT serotype	substrate: cleavage point[a] P3−P2−P1−**P1′**−P2′−P3′	probable residues of the S1′ recognition site			
A	SNAP-25: ^{195}A−N−Q−**R**−A−T^{200}	F163	F194	T220	D370
B	VAMP: ^{74}A−S−Q−**F**−E−T^{79}	N169	S200	I226	S376
E	SNAP-25: ^{178}I−D−R−**I**−M−E^{183}	T159	F191	T208	Y354
G	VAMP: ^{79}T−S−A−**A**−K−L^{84}	N170	N201	T227	Y376

[a] P1′ residue indicated in bold.

is well suited for the P1′ Ala of VAMP. The pocket is made smaller than the S1′ subsites of BoNT/A1, BoNT/B and BoNT/E primarily due to the steric bulk of Tyr 376 that partially fills this cavity. The S1′ subsite of BoNT/E-LC is also small, but deeper and more hydrophobic, reflecting the nature of the P1′ Ile of SNAP-25 (Figure 3E). In contrast, the S1′ subsites of BoNT/A1-LC (Figure 3C) and BoNT/B-LC (Figure 3D) are considerably larger than their BoNT/G-LC and BoNT/E-LC counterparts. The BoNT/A1-LC S1′ subsite is specific for the P1′ Arg of SNAP-25 and likely forms a buried salt bridge with Asp 370 at the bottom of its pocket. The BoNT/B-LC S1′ subsite is the largest of the four serotypes investigated and is well matched for the P1′ Phe of VAMP. Homology modeling reveals the anticipated disparity in the S1′ subsites of the different LCs that likely account for their unique P1′ specificities; however, additional studies are clearly required to confirm this.

SSR in VAMP-Specific Neurotoxins. It is clear that P1′ recognition alone cannot account for the high specificity of BoNTs. Indeed, substrate recognition is enhanced by the contribution of high-affinity exosites that are far removed from the active site, such as the yet to be identified exosite for binding of the consensus SSR motif found in each of the three SNARE proteins (*10*). The SSR motif is composed of nine residues and contains three conserved acidic residues at positions 3, 4, and 7 (Figure 2). This motif is present in two copies in VAMP, termed V1 (residues 38−47) and V2 (residues 62−71) (*10*). The essential role of the three conserved acidic residues in this motif in substrate recognition by VAMP-specific neurotoxins was demonstrated by mutagenesis of residues comprising this motif. These studies revealed that the V2 copy is critical for recognition by BoNT/B and BoNT/G, while in contrast the V1 is imperative for recognition by tetanus neurotoxin (TeNT) (*11*), as observed in the dramatic decrease in the rate of proteolysis of mutant VAMP substrates. Our structural and sequence analysis of these VAMP-specific neurotoxins helps to reconcile the differences observed in their SSR properties. The BoNT/B-LC VAMP crystal structure identified key residues involved in the recognition of the V2 SSR motif of VAMP. These LC residues include Arg 67, Tyr 72, Asp 375, Ser 376, and Glu 422. The equivalent residues in BoNT/G-LC are only partially conserved (Supporting Information, Figure S2). Furthermore, a superposition of the two structures (not shown) reveals that this motif clashes with β-strand β4 of BoNT/G-LC structure. Therefore, unless there is a conformational rearrangement in this region, it is unlikely that BoNT/G-LC shares the same SSR exosite as that defined in the BoNT/B-LC VAMP structure. However, on the basis of structural comparison of the BoNT/G-LC to BoNT/A-LC SNAP-25 complex structure (PDB code 1XTG, with an

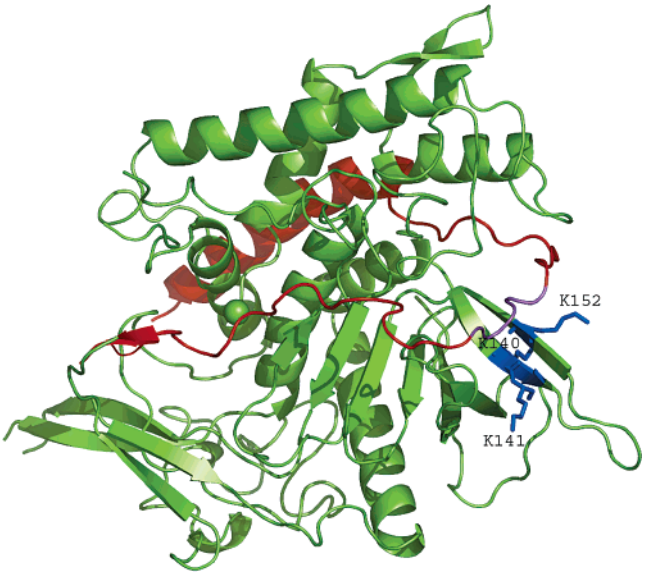

FIGURE 4: Putative SSR motif recognition exosite of BoNT/G-LC. Ribbon diagram of BoNT/G-LC (green) superimposed SNAP-25 (red), the substrate of BoNT/A. Basic residues in the putative BoNT/G-LC exosite involved in SSR motif recognition are colored (blue), and essential substrate residues 64−68 corresponding to the V2 SSR motif of VAMP are highlighted (violet).

overall sequence identity of 31%), Lys 140 and Lys 152 of the BoNT/G-LC are likely to influence the SSR of the V2 carboxylates, presumably by salt bridge formation. An overlay of the BoNT/G-LC structure with the structure of SNAP-25 from the BoNT/A-LC SNAP-25 complex (*9*) shows that three basic residues of the BoNT/G-LC, Lys 140, Lys 141, and Lys 152, are in close proximity to the SNAP-25 residues that correspond to the three acidic residues of the V2 SSR of VAMP (Figure 4). An alignment of BoNT/G-LC with the other BoNT and TeNT serotypes and strains reveals that two of these lysine residues, Lys 140 and Lys 152, are also conserved in the BoNT/B strains, which also require the V2 SSR, but not TeNT which does recognize the V2 SSR although is specific for the V1 SSR (Supporting Information, Figure S2) (*11*). Further functional and structural studies will be needed to determine if this putative SSR recognition site is correct.

CONCLUSION

The crystal structure of BoNT/G-LC presented complements the structural pool of neurotoxin structures and, in doing so, highlights the importance of the C-terminus. The identification of the putative S1′ subsite is consistent with the available LC structures of BoNT/G, BoNT/A, BoNT/B, and BoNT/E and thereby provides a molecular basis for the strict specificity for P1′ residues of their SNARE protein substrates. A putative SSR site is identified in the BoNT/G-LC structure that is consistent with the biochemical evidence of the importance of this SSR motif in substrate recognition. Collectively, our structural analysis may lead to development of highly specific inhibitors for BoNT/G and other serotypes.

NOTE ADDED IN PROOF

During the review of this article, the TeNT LC structure was published by two groups: Breidenbach and Brunger (*37*) and Rao et al. (*38*).

ACKNOWLEDGMENT

We thank Jim Schmidt for providing the VAMP-immobilized plates for the activity assay and Angela Walker for assistance in manuscript preparation.

SUPPORTING INFORMATION AVAILABLE

Figure S1, time course of fluorescence release by BoNT/G-LC and BoNT/B-LC, and Figure S2, sequence alignment of BoNT-LCs. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Yamasaki, S., Binz, T., Hayashi, T., Szabo, E., Yamasaki, N., Eklund, M., Jahn, R., and Niemann, H. (1994) Botulinum neurotoxin type-G proteolyses the Ala(81)-Ala(82) bond of rat synaptobrevin-2, *Biochem. Biophys. Res. Commun. 200*, 829−835.
2. Swaminathan, S., and Eswaramoorthy, S. (2000) Structural analysis of the catalytic and binding sites of Clostridium botulinum neurotoxin B, *Nat. Struct. Biol. 7*, 693−699.
3. Lacy, D. B., Tepp, W., Cohen, A. C., DasGupta, B. R., and Stevens, R. C. (1998) Crystal structure of botulinum neurotoxin type A and implications for toxicity, *Nat. Struct. Biol. 5*, 898−902.
4. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) Clustal-W—Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res. 22*, 4673−4680.
5. Schmidt, J. J., Stafford, R. G., and Bostian, K. A. (1998) Type A botulinum neurotoxin proteolytic activity: development of competitive inhibitors and implications for substrate specificity at the S1′ binding subsite, *FEBS Lett. 435*, 61−64.
6. Vaidyanathan, V. V., Yoshino, K., Jahnz, M., Dorries, C., Bade, S., Nauenburg, S., Niemann, H., and Binz, T. (1999) Proteolysis of SNAP-25 isoforms by botulinum neurotoxin types A, C, and E: Domains and amino acid residues controlling the formation of enzyme−substrate complexes and cleavage, *J. Neurochem. 72*, 327−337.
7. Schmidt, J. J., and Stafford, R. G. (2005) Botulinum neurotoxin serotype F: identification of substrate recognition requirements and development of inhibitors with low nanomolar affinity, *Biochemistry 44*, 4067−4073.
8. Hanson, M. A., and Stevens, R. C. (2000) Cocrystal structure of synaptobrevin-II bound to botulinum neurotoxin type B at 2.0 angstrom resolution, *Na. Struct. Biol. 7*, 687−692.
9. Breidenbach, M. A., and Brunger, A. T. (2004) Substrate recognition strategy for botulinum neurotoxin serotype A, *Nature 432*, 925−929.
10. Rossetto, O., Schiavo, G., Montecucco, C., Poulain, B., Deloye, F., Lozzi, L., and Shone, C. C. (1994) SNARE motif and neurotoxins, *Nature 372*, 415−416.
11. Pellizzari, R., Rossetto, O., Lozzi, L., Giovedi, S., Johnson, E., Shone, C. C., and Montecucco, C. (1996) Structural determinants of the specificity for synaptic vesicle-associated membrane protein/synaptobrevin of tetanus and botulinum type B and G neurotoxins, *J. Biol. Chem. 271*, 20353−20358.
12. Washbourne, P., Pellizzari, R., Baldini, G., Wilson, M. C., and Montecucco, C. (1997) Botulinum neurotoxin types A and E require the SNARE motif in SNAP-25 for proteolysis, *FEBS Lett. 418*, 1−5.
13. Santarsiero, B., Yegian, D., Lee, C., Spraggon, G., Gu, J., Scheibe, D., Uber, D., Cornell, E., Nordmeyer, R., Kolbe, W., Jin, J., Jones, A., Jaklevic, J., Schultz, P., and Stevens, R. (2002) An approach to rapid protein crystallization using nanodroplets, *J. Appl. Crystallogr. 35*, 278−281.
14. Schmidt, J. J., Stafford, R. G., and Millard, C. B. (2001) High-throughput assays for botulinum neurotoxin proteolytic activity: Serotypes A, B, D, and F, *Anal. Biochem. 296*, 130−137.
15. Otwinowski, Z., and Minor, W. (1997) in *Macromolecular Crystallography*, Part A, pp 307−326.
16. Vagin, A., and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement, *J. Appl. Crystallogr. 30*, 1022−1025.
17. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement, *Nat. Struct. Biol. 6*, 458−463.
18. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron-density maps and the location of errors in these models, *Acta Crystallogr. A47*, 110−119.
19. Collaborative Computational Project Number 4 (1994) The CCP4 Suite: Programs for Protein Crystallography, *Acta Crystallogr. D50*, 760−763.
20. Monzingo, A. F., and Matthews, B. W. (1984) Binding of N-carboxymethyl dipeptide inhibitors to thermolysin determined by X-ray crystallography—a novel class of transition-state analogues for zinc peptidases, *Biochemistry 23*, 5724−5729.
21. Lu, G. G. (2000) TOP: a new method for protein structure comparisons and similarity searches, *J. Appl. Crystallogr. 33*, 176−183.
22. Segelke, B., Knapp, M., Kadkhodayan, S., Balhorn, R., and 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. U.S.A. 101*, 6888−6893.
23. Agarwal, R., Eswaramoorthy, S., Kumaran, D., Binz, T., and 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.
24. DasGupta, B. R. (1994) *The Structure of Botulinum Neurotoxins*, Marcel Dekker, New York.
25. Fernandez-Salas, E., Steward, L. E., Ho, H., Garay, P. E., Sun, S. W., Gilmore, M. A., Ordas, J. V., Wang, J., Francis, J., and Aoki, K. R. (2004) Plasma membrane localization signals in the light chain of botulinum neurotoxin, *Proc. Natl. Acad. Sci. U.S.A. 101*, 3208−3213.
26. Baldwin, M. R., Bradshaw, M., Johnson, E. A., and Barbieri, J. T. (2004) The C-terminus of botulinum neurotoxin type A light chain contributes to solubility, catalysis, and stability, *Protein Expression Purif. 37*, 187−195.
27. Kadkhodayan, S., Knapp, M. S., Schmidt, J. J., Fabes, S. E., Rupp, B., and Balhorn, R. (2000) Cloning, expression, and one-step purification of the minimal essential domain of the light chain of botulinum neurotoxin type A, *Protein Expression Purif. 19*, 125−130.
28. Jones, S., and Thornton, J. M. (1996) Principles of protein−protein interactions, *Proc. Natl. Acad. Sci. U.S.A. 93*, 13−20.
29. Henrick, K., and Thornton, J. M. (1998) PQS: a protein quaternary structure file server, *Trends Biochem. Sci. 23*, 358−361.
30. Rety, S., Sopkova, J., Renouard, M., Osterloh, D., Gerke, V., Tabaries, S., Russo-Marie, F., and Lewit-Bentley, A. (1999) The crystal structure of a complex of P11 with the annexin II N-terminal peptide, *Nat. Struct. Biol. 6*, 89−95.
31. Montecucco, C., and Schiavo, G. (1995) Structure and function of tetanus and botulinum neurotoxins, *Q. Rev. Biophys. 28*, 423−472.
32. Stevens, R., and Hanson, M. (2001) Questions about the structure of the botulinum neurotoxin B light chain in complex with a target peptide—Response to Rupp and Segelke, *Nat. Struct. Biol. 8*, 664.
33. Rupp, B., and Segelke, B. (2001) Questions about the structure of the botulinum neurotoxin B light chain in complex with a target peptide, *Nat. Struct. Biol. 8*, 663−664.
34. Washbourne, P., Bortoletto, N., Graham, M. E., Wilson, M. C., Burgoyne, R. D., and Montecucco, C. (1999) Botulinum neurotoxin E-insensitive mutants of SNAP-25 fail to bind VAMP but support exocytosis, *J. Neurochem. 73*, 2424−2433.
35. Schiavo, G., Matteoli, M., and Montecucco, C. (2000) Neurotoxins affecting neuroexocytosis, *Physiol. Rev. 80*, 717−766.
36. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) ESPript: analysis of multiple sequence alignments in PostScript, *Bioinformatics 15*, 305−308.
37. Breidenbach, M. A., and Brunger, A. T. (2005) *Biochemistry 44*, 7450−7457.
38. Rao, K. N., Kumaran, D., Binz, T., and Swaminathan, S. (2005) *Toxicon 45*, 929−939.

BI0505924