scientific reports

Check for updates

OPEN

Reengineering the specificity of the highly selective *Clostridium botulinum* protease via directed evolution

Rebekah P. Dyer[2], Hariny M. Isoda[1], Gabriela S. Salcedo[2], Gaetano Speciale[1], Madison H. Fletcher[1], Linh Q. Le[4], Yi Liu[4], Karen Brami-Cherrier[4], Shiazah Z. Malik[4], Edwin J. Vazquez-Cintron[4], Andrew C. Chu[2], David C. Rupp[4], Birgitte P. S. Jacky[4], Thu T. M. Nguyen[1], Benjamin B. Katz[1], Lance E. Steward[4], Sudipta Majumdar[1], Amy D. Brideau-Andersen[4]✉ & Gregory A. Weiss[1,2,3]✉

The botulinum neurotoxin serotype A (BoNT/A) cuts a single peptide bond in SNAP25, an activity used to treat a wide range of diseases. Reengineering the substrate specificity of BoNT/A's protease domain (LC/A) could expand its therapeutic applications; however, LC/A's extended substrate recognition (≈ 60 residues) challenges conventional approaches. We report a directed evolution method for retargeting LC/A and retaining its exquisite specificity. The resultant eight-mutation LC/A (omLC/A) has improved cleavage specificity and catalytic efficiency (1300- and 120-fold, respectively) for SNAP23 versus SNAP25 compared to a previously reported LC/A variant. Importantly, the BoNT/A holotoxin equipped with omLC/A retains its ability to form full-length holotoxin, infiltrate neurons, and cleave SNAP23. The identification of substrate control loops outside BoNT/A's active site could guide the design of improved BoNT proteases and inhibitors.

Therapeutic proteases must select a narrow range of substrates to ensure safety and efficacy[1]. For example, BoNT/A hydrolyses a single peptide bond in its sole substrate SNAP25[2]. This cleavage of SNAP25 disrupts vesicle fusion, neurotransmitter release, and neuro-muscular communication. The remarkable specificity of BoNT/A and its months-long duration of clinical benefits allows its application to a broad and expanding list of neuronal indications, ranging from spasticity to facial wrinkles and migraines[3]. The complete toxin, or holotoxin, has a heavy chain (HC/A) and light chain (LC/A), which cooperate to infiltrate peripheral neurons and cleave SNAP25, respectively. The HC/A binding domain enters the cell through receptor-mediated endocytosis[4,5]. Upon endocytic escape mediated by the translocation domain within HC/A and subsequent disulfide reduction, LC/A recognizes and cleaves SNAP25[6].

Retargeting the delivery or substrate selectivity of BoNT/A could expand its therapeutic uses. For example, antibody fragments[7,8], growth factors[9], lectins[10], or other proteins[11] fused to the HC/A translocation domain could successfully deliver BoNT subtypes to new cell types to potentially treat pain, neuroendocrine disorders, and cancers[12]. Several non-neuronal SNARE proteins offer attractive therapeutic targets for reengineering the proteolytic specificity of LC/A. For instance, SNAP23 mediates the release of pro-inflammatory molecules[13,14], secretion of matrix metalloproteinases that promote tumor invasion[15], and hypersecretion of mucin in asthma[16]. Though not a substrate for wild-type LC/A (wtLC/A), SNAP23 is the closest homolog to SNAP25 (Fig. 1A). Additionally, the related protease LC/E was successfully engineered to cleave SNAP23 with minimal engineering[17]. Recently, a phage-assisted evolution system with negative selection (termed PANCE) was used to re-engineer three different BoNT LC proteases to cleave new substrates[18]. Here, we report methods to access large numbers of BoNT variants and adapt a simple BoNT LC assay for high-throughput screening. Additionally, this report tests

[1]Departments of Chemistry, University of California, Irvine, 1102 NS-2, Irvine, CA 92697-2025, USA. [2]Molecular Biology and Biochemistry, University of California, Irvine, 1102 NS-2, Irvine, CA 92697-2025, USA. [3]Pharmaceutical Sciences, University of California, Irvine, 1102 NS-2, Irvine, CA 92697-2025, USA. [4]Allergan Aesthetics, An AbbVie company, 2525 Dupont Drive, Irvine, CA 92612, USA. ✉email: amy.brideauandersen@abbvie.com; gweiss@uci.edu
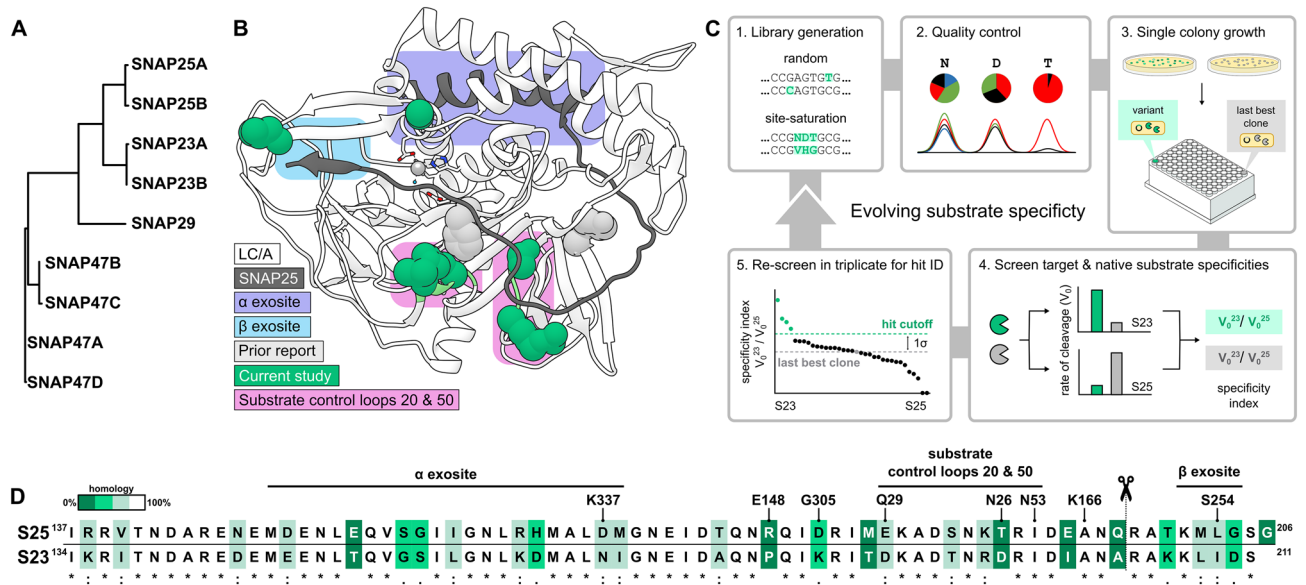
Figure 1. Directed evolution overview. (**A**) SNAP25 sub-family proteins and their isoforms (CLUSTAL 2.1[40]). (**B**) Co-crystal structure (PDB: 1XTG[27]) of LC/A (white) and SNAP25 (dark gray). Eight substitutions in LC/A drive SNAP23 specificity (teal) through substrate control loops (pink) alongside prior substitutions (light gray). (**C**) Platform for the directed evolution of SNAP23 substrate specificity. 1. Random or site-directed mutagenesis (e.g., site-saturation); 2. QC by DNA sequencing; 3. High-throughput protein production; 4. Measure V_0^{23} and V_0^{25} for substrate specificity; 5. Confirmation screens. The most specific and consistent variant from the DARET assay entered the next round of directed evolution. (**D**) Sequence alignment of substrates used for screening (UniProt: P60880, O00161). The SNAP binding exosites in LC/A (residue numbers above) and cleavage site (scissors) are shown. The gradient of color indicates homology from identical (white, *), to strongly similar (light gray, :), weakly similar (teal, .), or dissimilar (dark teal, space).

whether a large number of mutations required for substrate redirection could affect the folding and production of full-length holotoxin.

LC/A obtains its tremendous enzymatic specificity by interrogating ≈60 residues of its substrate (Fig. 1B). Conventional platforms for protease reengineering are limited to short recognition sequences[19–21]. Therefore, progress towards evolution of LC/A to cleave SNAP23 has largely relied on in silico studies[22]. For example, adding four computationally derived mutations (E148Y/K166F/S254A/G305D) to LC/A generates some SNAP23 cleavage activity[23,24]. However, this quadruple mutant of LC/A (qmLC/A) remains highly specific for its native substrate, SNAP25 (Fig. S1). Thus, directed evolution was applied to retarget qmLC/A to specifically cleave SNAP23 over SNAP25.

Results and discussion

To screen for substrate specificity, depolarization after resonance energy transfer (DARET)[25,26] was used to measure LC/A cleavage of SNAP23 and SNAP25. Here, > 70 residues of each substrate were sandwiched between GFP and BFP (shown schematically in Fig. S2). This assay features exceptionally low background compared to a FRET-only assay and increased polarization upon proteolysis[25]. Achieving consistent growth of individual LC/A variants overexpressed in deep-well microtiter plates required high-speed shaking at 900 rpm for more consistent aeration, along with optimization of the growth media and the fermentation time and temperature to favor LC/A expression while limiting its proteolysis (Materials and Methods). During each round of directed evolution, normalizing the initial rate of proteolysis for SNAP23 (V_0^{23}) to the analogous rate for SNAP25 (V_0^{25}) yielded a substrate specificity index, which accommodated variation in enzyme concentration (Fig. 1C). The assay proved effective at revealing even small gains in substrate specificity.

Error-prone PCR of the gene encoding qmLC/A in Round 1 uncovered three variants with potentially improved SNAP23 specificity: N240S, E201D/D203V, and K364R/Y387N. The latter two variants exhibited inconsistent SNAP23 specificities in triplicate follow-up screens, perhaps due to protein insolubility. Only N240S qmLC/A yielded a consistent increase in SNAP23 specificity compared to qmLC/A. Additional site-saturation at this position in Round 2 identified N240A as improving SNAP23 specificity approximately 1.3-fold or > 1σ above the mean specificity for qmLC/A data at the same screening dilution (σ = standard deviation) (Fig. 2A). This residue exists in a semi-ordered loop approximately 6 Å away from the beta hairpin of the beta exosite, an important contributor to binding SNAP25.

Subsequent rounds of directed evolution were guided by rational design, including inspection of the LC/A-SNAP25 co-crystal structure (Fig. S3)[27] and prior work[23,24]. Rounds 3 to 7, for example, included residues with high B-factors, which can be productive sites for substitution[28]. Investigating these sites with an iterative mutagenesis method[29] improved specificity for SNAP23 up to 80-fold over qmLC/A in the DARET screens
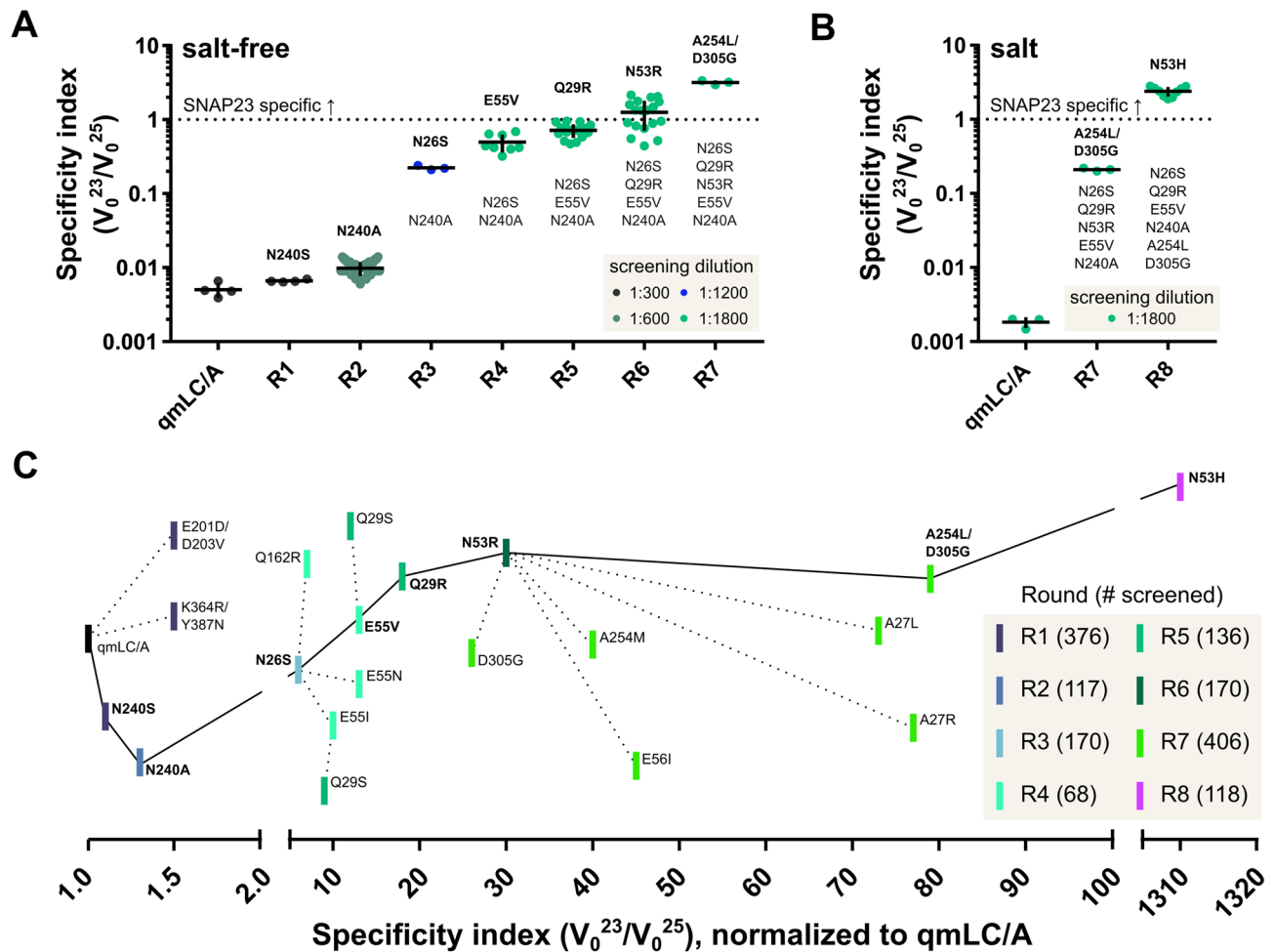
Figure 2. Directed evolution of a SNAP23-specific LC/A. (**A**) Improvement in specificity index through directed evolution in salt-free buffer. The most specific variant from each round was assayed multiple times (replicates shown). The average specificity index (horizontal bars) and standard deviation (error bars) are shown. The final dilution of cell lysates for screening is indicated by color code. (**B**) Increased specificity index achieved through directed evolution in salt buffer. (**C**) Roadmap for directed evolution showing the fold increase in specificity indices over eight rounds (R1 to R8). As shown in Fig. S1, the substrate specificity of the qmLC/A variant is sensitive to screening dilution; therefore, each point represents the variant's average specificity normalized to the qmLC/A average specificity at the corresponding screening dilution. Advancement to the next round (solid line) weighed specificity index, protein solubility, and stability.

(Fig. 2A). In each round, Q_{pool} analysis[30] was used to assess library quality, and the TopLib program estimated the numbers of variants required for screening[31].

Later rounds of directed evolution identified a sensitivity to the screening and assay conditions. In early rounds, variants were diluted and screened in a "salt-free" assay buffer (50 mM HEPES pH 7.4, 0.05% Tween with no added salt) chosen to maximize LC/A cleavage activity in vitro[25]. These screening conditions yielded LC/A variants that cleaved SNAP23 at over three times the rate of SNAP25 cleavage (Fig. 2A). However, the presence of salt reversed these gains in SNAP23 specificity for the selectant from Round 7 (Fig. 2B, Fig. S4). The results vividly illustrate the First Law of Directed Evolution, "you get what you screen for"[32], as screening took place in salt-free buffer.

Since we sought protease variants having specificity for SNAP23 within the salty environment of cells, we addressed this issue with a final round of mutagenesis and screening in salt buffer. During earlier rounds, positions Q29 and N53 were mutated to positively charged Arg, introducing a potential source of salt sensitivity. Therefore, the two sites were revisited with 19 AA substitution libraries and screened in a salt-containing buffer (50 mM KH_2PO_4, pH 7.4). Screening these libraries revealed that substituting His at residue 53 confers SNAP23 specificity to LC/A even in salt buffer, improving screening specificity 1300-fold compared to qmLC/A (Figs. 2B,C, S4). This selectant was termed the octuple mutant LC/A (omLC/A).

Prior work attributed substrate specificity in LC/A to primarily the alpha and beta binding exosites[22,27,33,34]. These binding sites comprise the first and last interactions between substrate and protease, respectively. The results presented here identify a new series of residues between these exosites exerting control over SNAP23/

	SNAP23				SNAP25			
	k_{cat} (s^{-1})	K_m (µM)	V_{max} (mP*µM*s^{-1})	k_{cat}/K_m (µM^{-1}*s^{-1})	k_{cat} (s^{-1})	K_m (µM)	V_{max} (mP*µM*s^{-1})	k_{cat}/K_m (µM^{-1}*s^{-1})
wtLC/A	Und	Und	Und	Und	52 ± 2	3.4 ± 0.3	0.52 ± 0.02	15 ± 1
qmLC/A	3.3 ± 0.4	18 ± 4	0.033 ± 0.004	0.18 ± 0.05	140 ± 9	0.7 ± 0.2	1.40 ± 0.09	200 ± 60
omLC/A	8.1 ± 0.3	7.4 ± 0.6	0.081 ± 0.003	1.1 ± 0.1	3.2 ± 0.1	0.32 ± 0.09	0.032 ± 0.001	10 ± 3

Table 1. Enzyme kinetics of LC/A variants. Kinetic constants of recombinantly expressed and purified LC/A variants were determined via the DARET assay[25] with modifications as described in the *Materials & Methods*. Values are means ± SEM (n = 3) and und = undetectable.

SNAP25 specificity. We term these substrate control loops 20 (LC/A residues 26 to 29) and 50 (residues 52 to 56) (Fig. 1B,D).

Two categories of substitutions in these substrate control loops improved SNAP23 specificity. First, the enzyme's overall charge became more positive. For instance, the loss of E55 and E56 in control loop 50 increased SNAP23 specificity up to twofold (Fig. 2C). Similarly, substituting positively charged Arg residues at positions in both control loops (A27, Q29, N53) improved SNAP23 specificity up to two-fold for each substitution. Increasing the positive charge in substrate control loop 20 likely accommodates D191 and D198 in SNAP23 (Fig. S5). Indeed, mutating SNAP25 residues to SNAP23 residues at these sites (E183D, T190D) reduces cleavage by wtLC/A up to 60% compared to cleavage of the native SNAP25[22].

The second category of substitutions conferring greater SNAP23 specificity in these substrate control loops focuses on sidechain size. For example, substituting a small, hydrophilic serine for N26 in Round 3 increased specificity by nearly five-fold compared to the best Round 2 variant (Fig. 2C). Conversely, mutating the neighboring A27 to the larger Leu sidechain increased specificity 2.4-fold from Round 6 to Round 7. Position 26 could directly contact the SNAP23 substrate, whereas position 27 is directed towards control loop 50 (Fig. S5). Changing the size and flexibility of these residues therefore may reshape this SNAP-LC/A binding region to favor SNAP23 binding over SNAP25.

In addition to the discovery of substrate control loops 20 and 50 in LC/A, directed evolution revealed new information on established binding sites. At the alpha exosite, the SNAP25 residues in contact with LC/A are highly homologous to those in SNAP23, which suggests both substrates bind in a similar manner to this region. Indeed, generating libraries at alpha exosite residues in LC/A, such as K337, yielded few variants with proteolytic activity, and none with SNAP23 specificity (Fig. S6). Targeting residues near or within the beta exosite, however, was more productive. For example, libraries at N240 and S254 generated variants with up to three-fold improved SNAP23 specificity (Fig. 2C).

Comparing the enzyme kinetics for omLC/A and qmLC/A reveals the basis for its improved specificity. The main driver for omLC/A's improved SNAP23 specificity is a 20-fold reduction in catalytic efficiency (k_{cat}/K_M) for cleaving SNAP25 (Table 1, Figs. 3A,B, S7). Additionally, omLC/A improves the catalytic efficiency for cleaving SNAP23 by sixfold. Notably, our directed evolution strategy assessed both SNAP23 and SNAP25 cleavage activities, driving these trends. The catalytic rate (k_{cat}) for hydrolysis of SNAP25 is 40-fold lower for omLC/A compared to qmLC/A. In line with previous studies[27,35,36], the eight added mutations in omLC/A impact catalytic rate more than substrate affinity. Overall, the results represent a 120-fold improvement in SNAP23-specific catalytic efficiency from qmLC/A to omLC/A. The in vitro screens applied here offer an imperfect model for the human SNAP23 target. In cells, both SNAP23 and BoNT/A are membrane-associated[42], which could dramatically affect the enzyme's activity; these cellular conditions could diminish the importance of K_M and promote k_{cat} as more important for LC/A engineering.

Interestingly, the omLC/A variant exhibits strong dependence on Zn^{2+} for its substrate specificity, but not its activity. Without added Zn^{2+} during purification, the enzyme remains proteolytic, presumably due to endogenous Zn^{2+}, but has a higher rate of cleavage for SNAP25 than SNAP23. The addition of Zn^{2+} (2 µM) during protein isolation reverses this specificity, and restores the enzyme's preference for SNAP23. Importantly, once treated with Zn^{2+}, the enzyme remains specific for SNAP23 (Fig. S8).

We next determined how omLC/A recognizes and cleaves the full-length SNAP23 (fl-S23) protein using mass spectrometry. Loss of the eight C-terminal residues of fl-S23 plus water (RAKKLIDS) in a cleavage product with a putative charge state of [M + 2H]$^{2+}$ is observed upon treatment with omLC/A (Fig. 3C, Fig. S14). This cleavage site in fl-S23 aligns with the expected SNAP25 site for wtLC/A hydrolysis (Fig. 1D). Therefore, the stringent cut-site specificity of this class of proteases is preserved in our engineered omLC/A variant for SNAP23; indeed, omLC/A fails to cleave the C-terminal domain of SNAP29 (Fig. S9), which is highly homologous to both SNAP23 and SNAP25 (Fig. 1A). The results suggest omLC/A interrogates its substrates and binds similarly to wtLC/A binding SNAP25 (as shown in Fig. S5).

We next incorporated omLC/A into the full-length BoNT/A holotoxin, termed omBoNT/A. In the experiments reported here, purified omLC/A and omBoNT/A exhibited comparable stability and purity (Fig. S11) to wild-type LC/A and BoNT/A, respectively. Note that the omBoNT/A reported here has a C-terminal His$_6$-tag, is "scar-free," and lacks any sequences required for ligation-mediated assembly. Recombinant expression of omBoNT/A yielded a soluble 150 kDa polypeptide; as expected, reduction by DTT released the 100 kDa (HC/A) and the 50 kDa (omLC/A) proteins (Fig. 3D, Fig. S12A). Release of BoNT LC proteases through disulfide bond reduction is vital to successful cytosolic delivery and SNAP cleavage[6]. The omBoNT/A retains its SNAP23 cleavage activity in vitro and *in cellulo* (Figs. 3E,F, 12B-C, S13). Additionally, the omBoNT/A exhibits at least 25-fold

reduced muscle paralysis associated with SNAP25 cleavage compared to wtBoNT/A (Fig. S10A); the residual paralysis observed demonstrates successful cellular delivery of active omLC/A into motor nerve terminals. Furthermore, no systemic toxicity effects were observed for mice treated with 5 ng/kg omBoNT/A (Fig S10B). Additional engineering of the omLC/A holotoxin is needed to enable efficient uptake into non-neuronal cells for useful SNAP23 cleavage activity in vivo.

This report demonstrates the plasticity of both wtLC/A's hyperspecific substrate preferences and folding of the full-length, non-ligation-mediated holotoxin. omLC/A preserves wtLC/A's vaunted selectivity for cleaving a single targeted peptide bond with no additional cleavage products observed by MS. Since the belt acts as a pseudosubstrate, incorporation of multiple, specificity-controlling mutations near and in the binding cleft could also disrupt proper folding and function of the holotoxin. However, the eight mutations of omLC/A do not interfere with the omBoNT/A's ability to fold, bind, and then release its peptide "belt," as required for the holotoxin's successful assembly and function; the result validates the approach of focusing substrate-tailoring efforts on the LC domain.

The SNAP25 substrate wraps around nearly the entire circumference of LC/A, complicating rational attempts to alter substrate specificity. We address this with a novel fluorescence polarization-based platform that expands substrate screening from peptide-sized to include an entire domain of SNARE proteins. The approach offers general applicability, and can be applied to non-BoNT proteases. However, the SNARE proteins are well-suited to the extreme specificity testing demanded by the BoNT family proteases; for example, the SNAREs are structurally malleable. For non-BoNT proteases, which might not feature extremely long half-lives, catalytic rates should be considered with the substrate specificity index to identify the most active and selective proteases.

The advances reported here include a robust, readily implementable high-throughput assay system for evolution of LC/A, which uncovered substrate control loops. Furthermore, *in cellulo* and in vivo (mice) application of an engineered, full-length omBoNT/A holotoxin accomplish key preclinical milestones along the pathway for drug development. Together, the experiments demonstrate the compatibility of omBoNT/A's mutations in the context of the full-length holotoxin and provide a blueprint for redirecting the enzyme's vaunted substrate specificity.

With respect to general lessons for protease engineering, conventional approaches typically use a short substrate analog as a proxy for cleavage of the target substrate. A short substrate, however, would likely fail to identify the substrate control loops uncovered here. Furthermore, the exceptionally low background of the DARET assay proved critical to the success of this project, as the minimal background allowed high-throughput screening of crude, cell-lysates; assay development for screening should focus on these criteria initially and then on identifying conditions as similar to the target environment as possible. For example, the reported experiments illustrate the unexpected sensitivity of LC/A variants to Zn^{2+} and salt concentration in the absence of such screening criteria.

Our data suggests avoiding the alpha exosite for reengineering substrate specificity and focusing on the substrate control loops uncovered here. Notably, substrate control loops 20 and 50 are structurally conserved amongst BoNT serotypes, yet feature divergent sequences (Fig. 4), which suggests their importance to directing the diverse substrate specificities of these proteases. However, pan-BoNT/A inhibitors (e.g., for anti-botulism therapies) could target the alpha exosite due to its recognition of a highly conserved region of SNAP substrates. Therefore, the data reported here could guide engineering of both catalysts and inhibitors.

BoNT represents a unique biotherapeutic as it catalytically ablates its target. Such capabilities are becoming increasingly important with engineered proteins, including proteases, emerging in pharmaceutical pipelines[1]. We hope this work inspires new uses for BoNT proteases to treat a wide range of diseases.

Materials and methods

Construction of the qmLC/A encoding plasmid. The gene encoding the qmLC/A variant (E148Y, K166F, S254A, G305D) was constructed through overlap extension PCR using the gene encoding wild-type LC/A (WT, amino acids 1 to 430) as the template. The resulting PCR products were *DpnI*-digested, column purified and concentrated, then ligated and subcloned via Gibson Assembly into a pET-29b(+) vector featuring a C-terminal His$_6$ tag for recombinant protein expression and purification. Sanger sequencing (Genewiz LLC) confirmed construction of qmLC/A-pET-29b(+).

Batch expression and purification of LC/A variants. The LC/A-pET-29b(+) WT and mutant plasmids were transformed into BL21 (DE3) *E. coli* (One Shot™ Star™, ThermoFisher) cells and spread on LB$_{kan}$ agar plates at 37 °C overnight. A single colony was used to inoculate an LB$_{kan}$ seed culture incubated for 6 to 18 h at 37 °C with shaking at 225 rpm. The LB$_{kan}$ expression cultures were supplemented with glucose (1% w/v) and inoculated with seed culture (1:1000), then incubated at 37 °C, with shaking at 225 rpm until the culture reached an OD$_{600}$ of 0.6. The LC/A protein production was induced by addition of isopropyl β-D-1thiogalactopyranoside (IPTG, 1 mM final concentration) and incubated at 25 °C with shaking at 225 rpm for 20 to 24 h. The cells were harvested by centrifugation (6084×*g*, 4 °C, 20 min) and resuspended while on ice in LC Lysis Buffer (100 mM HEPES, 25 mM imidazole, 500 mM NaCl, pH 7.4) for storage at −80 °C or immediately subjected to lysis. The soluble protein was extracted from cells via sonication followed by centrifugation (26,891×*g*, 4 °C, 45 min), and the supernatant then batch bound overnight at 4 °C to nickel-charged IMAC resin (Profinity, BioRad) which had been equilibrated with LC Lysis Buffer. The LC/A variants were purified via a gravity column, which was first washed with 20 column volumes (c.v.) of LC Lysis Buffer, then 12 c.v. LC Lysis Buffer supplemented with 50 mM imidazole, and finally eluted with 16 c.v. LC Lysis Buffer supplemented with 500 mM imidazole, collecting four 4-mL fractions. The fractions were analyzed by SDS PAGE (12% acrylamide) and visualized by Coomassie blue stain. LC/A-containing fractions with a purity of 95% or higher were combined and dialyzed into chilled Storage Buffer at 4 °C. After sterile filtration through a 0.22 μm filter, the purified LC/A variants were subjected to
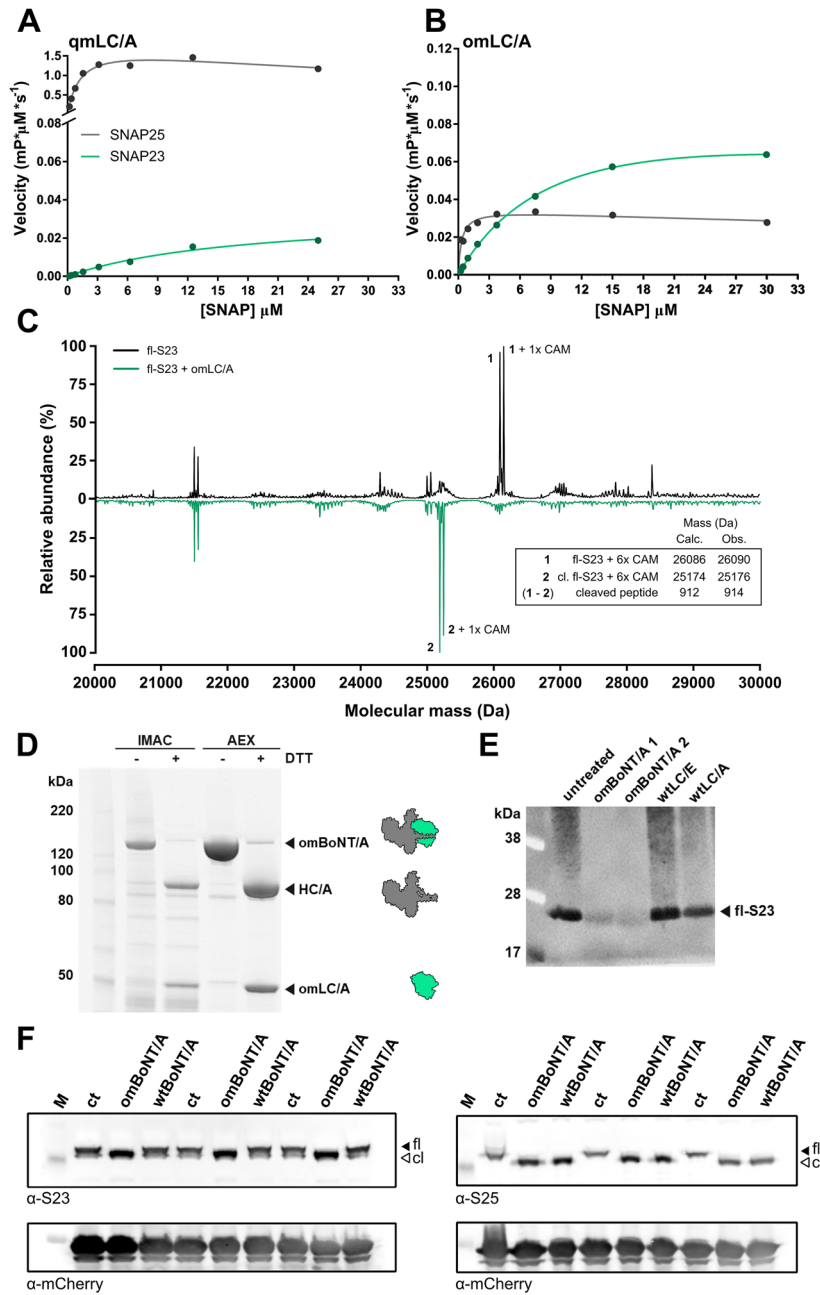
Figure 3. Characterization of omLC/A cleavage. (**A**) Rates of SNAP cleavage by batch-expressed, purified qmLC/A and (**B**) omLC/A at the indicated DARET substrate concentrations (n = 3). (**C**) Deconvoluted ESI–MS of recombinant, full-length SNAP23 (fl-S23) treated with DTT and iodoacetamide to carbamidomethylate its six cysteines (6×CAM). The mass spectrum of intact fl-S23 incubated with buffer (top, black) is compared to that of fl-S23 incubated with omLC/A (bottom, teal). Intact fl-S23 (1) and cleaved fl-S23 (cl. fl-S23, 2) peaks are labeled. Additional marked peaks correspond to the masses of peaks 1 and 2 plus one additional CAM (+ 57 Da), likely resulting from overalkylation by iodoacetamide[41]. The cleaved peptide was not directly observed, but inferred from the mass difference of peaks 1 and 2. The deconvolution error is +/- 2 Da. (**D**) Recombinant omBoNT/A was purified by IMAC followed by anion exchange (AEX) chromatography. The omBoNT/A is ≈95% nicked upon DTT reduction as demonstrated by the presence of the HC/A and omLC/A bands. (**E**) In vitro cleavage of recombinant fl-S23 by two independent preparations of omBoNT/A (1 and 2) visualized with a C-terminal anti-SNAP23 antibody; before proteolysis, omBoNT/A was reduced with TCEP. The untreated, wtLC/A, and wtLC/E lanes provide negative controls. (**F**) In cellulo cleavage of SNAP23 and SNAP25 in SiMA-H1 cells infected with adenovirus delivering DNA encoding mCherry/SNAP23. Cells were treated with omBoNT/A or wtBoNT/A or without toxin (ct). Proteins, full-length (fl) or cleaved (cl), were identified by Western blotting with anti-SNAP23, -SNAP25, or -mCherry antibodies (M, MW marker). Full-length images of these gels with multiple exposures where necessary are shown in Fig. S12 and S13.
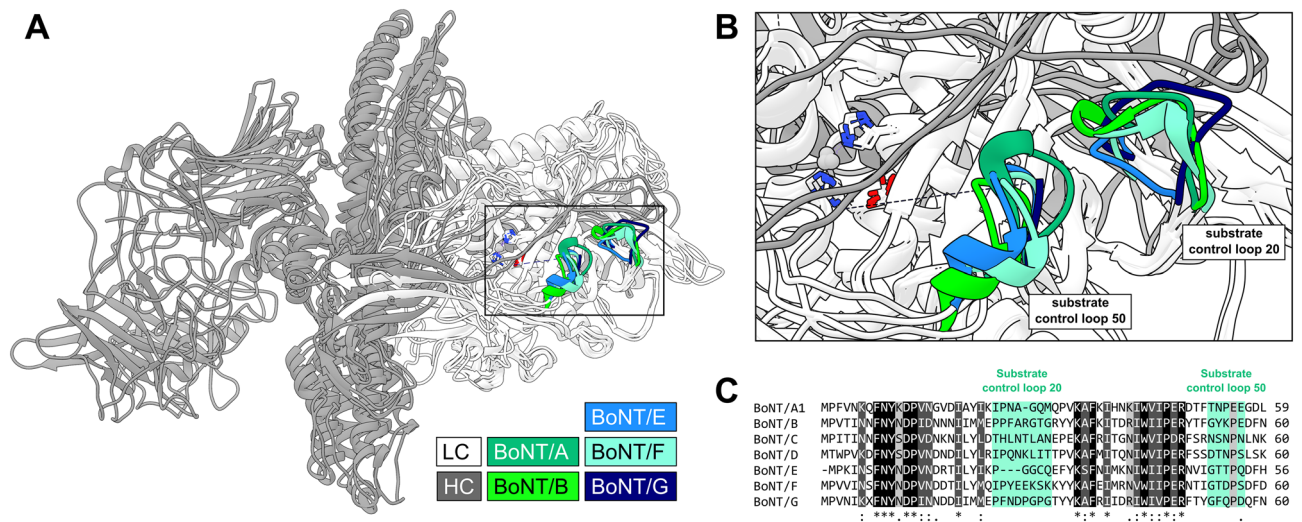
Figure 4. Substrate control loops are structurally present in other BoNT LC serotypes. (**A**) Multiple BoNT holotoxins were structurally aligned with UCSF Chimera MatchMaker (PDB IDs: 3BTA, 1S0G, 3FFZ, 1ZB7, & 2A8A). The heavy chain (HC) of each BoNT is colored dark grey, the light chain (LC) is colored in white, and the substrate control loops are colored as indicated in the key. (**B**) Larger image of the boxed section in A highlighting the substrate control loops. (**C**) The primary sequences of LC serotypes A through G were aligned using Clustal Omega. The substrate control loops are highlighted and annotated in teal. Identical residues are color-coded black (*), strongly similar residues in dark grey (:), weakly similar residues in light grey (.), and dissimilar residues are transparent. The poor primary sequence conservation in these structurally conserved loops suggests they help inform the diverse substrate specificities of other LC serotypes.

BCA assay to quantify protein concentration through comparison to BSA standards (Pierce™ BCA Assay, ThermoFisher). The proteins were stored at 4 °C before assays.

Expression and purification of SNAP23 & SNAP25 substrates. Substrates for the depolarization after resonance energy transfer (DARET) assay comprising of the C-terminal domain of SNAP23 (amino acids 137 to 211) or SNAP25 (amino acids 134 to 206) fused between GFP and BFP were subcloned into pET-28c(+) via ligation independent cloning. The DARET-pET-28 plasmids were transformed into BL21 (DE3) *E. coli* cells and spread on LB agar plates supplemented with 40 µg/mL kanamycin (LB$_{kan}$) then incubated at 37 °C overnight. A single colony was used to inoculate an LB$_{kan}$ seed culture and incubated for 6 to 18 h at 37 °C with shaking at 225 rpm. Next, LB$_{kan}$ expression cultures were inoculated with 1:1000 of the seed culture and incubated at 37 °C with shaking at 225 rpm until the cultures reached an OD$_{600}$ of 0.6. The DARET substrate production was induced through addition of IPTG (final concentration of 0.5 mM) and incubated at 16, 18, or 22 °C with shaking at 225 rpm for 18 to 24 h. The cells were harvested by centrifugation (6084×*g*, 4 °C, 20 min) and resuspended while on ice in DARET Lysis Buffer (25 mM HEPES, 300 mM NaCl, pH 8.0) for storage at -80 °C or immediately subjected to lysis. The soluble protein was extracted from cells via sonication followed by centrifugation (26,891×*g*, 4 °C, 45 min), and the supernatant was mixed with urea (0.5 M final concentration) before batch binding overnight at room temperature (RT) to nickel-charged IMAC resin (Profinity, BioRad) that had been previously equilibrated with the DARET Lysis Buffer. The DARET substrates were purified in a gravity column, first washing with 20 (c.v.) DARET Lysis Buffer, then 12 c.v. of DARET Lysis Buffer supplemented with 20 mM imidazole, and finally eluted with 16 c.v. of DARET Lysis Buffer supplemented with 250 mM imidazole. During elution, fractions with visibly green DARET substrates were collected. These fractions were combined and transferred to dialysis tubing (SnakeSkin, ThermoFisher) and dialyzed into chilled Storage Buffer (50 mM HEPES pH 7.4) at 4 °C. After sterile filtration through a 0.22 µm filter, the purified DARET substrates were subjected to BCA assay to quantify protein concentration through comparison to BSA standards. The DARET substrates were then aliquoted, protected from light, and stored at 4 °C or flash frozen for storage at − 80 °C before assays.

Construction, expression, and purification of SNAP29 substrate. SNAP29 DARET was constructed by inserting a gene fragment of human SNAP29 (amino acids 185 to 258, Genewiz) between eGFP and eBFP2 via overlap extension PCR, then cloning into pET-28c(+) through Gibson assembly. The SNAP29 DARET-pET28-c(+) construct was confirmed by Sanger sequencing (Genewiz). The SNAP29 DARET-pET-28 plasmid was transformed into BL21 (DE3) *E. coli* cells and spread on an LB$_{kan}$ agar plate then incubated at 37 °C overnight. The SNAP29 DARET substrate was expressed as described for the SNAP23 and SNAP25 DARET substrates with minor alterations (OD$_{600}$ of 0.8, 0.25 mM final IPTG). Purification and storage of the SNAP29 substrate were performed as described for SNAP23 and SNAP25 DARET substrates.

Library generation. The qmLC/A encoding plasmid served as a template for the first round of mutagenesis, and the qmLC/A provided a positive control and standard for each screen of SNAP23/25 specificity. The

PCR products were *DpnI*-digested, column purified and concentrated, and then subcloned into pET29b(+) via Gibson Assembly and evaluated by Sanger sequencing. The error-prone PCR (epPCR) library of the gene encoding qmLC/A was created with the GeneMorph II kit (Agilent) according to the manufacturer's instructions. The Round 1 epPCR library yielded the following mutants having SNAP23/SNAP25 specificity greater than qmLC/A: E201D/D203V, N240S, K364R/Y387N. Site-directed mutagenesis libraries were generated with overlap extension PCR using oligonucleotides featuring degenerate sequences at selected residues. In Round 2, the saturated amino acid substitution library (20 amino acids) was constructed using Tang's mutagenesis method[37] at position 240. In Rounds 3 to 6 and part of Round 7, iterative saturation mutagenesis using an NDT codon (encoding C, D, F, G, H, I, L, N, R, S, V, and Y) was employed at residues N26, A27, G28, Q29, M30, T52, N53, E55, E56, E171, S143, Q162, M253, T327, and K337. These rounds yielded the N26S, Q29R, N53R, and E55V variants. In the second part of Round 7 and the entirety of Round 8, libraries featuring NDT and VHG codons (substitution with all AAs except Trp) were constructed at residues Q29, N53, Y148, F166, A254, and D305. These rounds yielded the N53H, A254L, and D305G variants. Each saturation library was subject to Q_{pool} analysis for quality control[30]; libraries with a $Q_{pool} < 0.7$ were subcloned again from PCR fragments and further analyzed before screening.

Library expression, harvest, and preparation for screening.

The qmLC/A libraries were transformed into BL21 (DE3) *E. coli* cells then spread onto LB$_{kan}$ agar plates for overnight incubation at 37 °C. Single colonies were selected and inoculated single 96well deep-well plate (DWP) seed cultures (300 µL LB$_{kan}$ per well), then incubated at 37 °C with shaking at 225 rpm for 18 h. A qmLC/A and uninoculated culture were included as single wells on each DWP as positive and negative controls, respectively. An expression culture (630 µL LB$_{kan}$ per well) was subsequently inoculated with 20 µL of the seed culture and incubated for 3 h at 37 °C with shaking at 225 rpm. To generate glycerol stocks, the cells remaining from the seed culture DWP were harvested via centrifugation (2056×*g*, 4 °C, 1 h), then resuspended in 50 to 100 µL ultrapure glycerol (50% in autoclaved nanopure water) and transferred to a 96-well plate for storage at −80 °C. The expression DWP culture was chilled on ice for 10 min, then induced by adding IPTG (1 mM final concentration) before incubating for 22 to 24 h at 23 °C with shaking at 900 rpm. The cells were harvested via centrifugation (2056×*g*, 4 °C, 1 h) then stored at −80 °C for 20 min or overnight. For screening, cells were thawed at RT, then lysed at 23 °C with shaking at 500 rpm for 30 min in DWP Lysis Buffer: 100 µL per well of SoluLyse (Genlantis) plus benzonase nuclease (75 U/mL, NEB). The insoluble debris was removed via centrifugation (2056×*g*, 4 °C, 1 h), then an aliquot (75 µL) of each lysate was transferred to a 96-well plate (Celltreat). Lysates were diluted 1:100, 1:200, 1:400, or 1:600 in Assay Buffer (50 mM HEPES, 0.05% v/v Tween, pH 7.4) or Intracellular Buffer (50 mM KH$_2$PO$_4$, pH to 7.4 with KOH) before screening.

DARET assay to screen for SNAP23/SNAP25 specificity.

Recombinantly expressed and purified SNAP25 DARET substrate was diluted to 3 µM in Assay Buffer or Intracellular Buffer, then 100 µL substrate added to each well of a 96-well flat, black, non-binding surface microtiter plate (Corning). From the diluted lysate plate, 50 µL of the blank was added to its corresponding well in the black plate and used to optimize the gain and Z position of a Spark fluorescence polarization plate reader (Tecan). The sample was excited with polarized light at 380 nm with a 20 nm bandwidth, written here as 380(20) nm, and the polarized emission detected at 535(25) nm. For the remaining 95 wells, 50 µL from each well of the lysate dilution plate was added to the black plate containing substrate and the entire plate screened kinetically for 50 min to 14 h at 28 ± 1 °C. The assay steps were then repeated for 3 µM SNAP23 DARET using the same lysate dilution plate. The final concentration of each substrate screened was 2 µM. The changes in polarization over time were visualized using Prism (GraphPad) and initial rates (V_0) were derived from fitting trendlines to the initial, linear portion of the raw data. The rates of negative controls (no enzyme) for each DARET substrate were subtracted from the rates of each variant to account for nonenzymatic changes in polarization. The specificity indices were calculated via the ratio of SNAP23 and SNAP25 initial rates for each clone according to Eq. (1).

$$Specificity\ index \ = \left(V_0^{23}\right) / \left(V_0^{25}\right) \tag{1}$$

In Rounds 2 to 8, LC/A variants with specificity indices at least 1.5 times higher than qmLC/A or the most specific variant from the previous round of directed evolution were subject to further screening in triplicate. From the glycerol stocks, the variants and controls (qmLC/A, last best variant) were streaked onto LB$_{kan}$ plates for overnight incubation at 37 °C. Three colonies from these streaks were used to inoculate three wells of a seed culture DWP, then three wells of an expression culture DWP for LC/A variant production, harvesting, lysis, and screening as described above. The specificity index of each well was calculated first, then indices averaged together for the same variant. LC/A variants demonstrating consistent, improved specificity over the last best variant (> 1σ, where σ = standard deviation) were selected as starting points for the next round of mutagenesis and screening.

Mass spectrometry.

To determine the SNAP23 cleavage site of the LC/A variants, recombinant full-length SNAP23 (fl-S23) with N-terminal hexa-His and TEVp tags was expressed and purified to > 95% purity by SDS PAGE, then dialyzed into 20 mM Tris pH, 8.0 at 4 °C overnight. For mass spectrometry (MS) assays, fl-S23 was treated with dithiothreitol (7.5 mg/mL) at 80 °C for 10 min followed by iodoacetamide (9 mg/mL) at 25 °C for 1 h in the dark. This treatment functionalized the six free cysteines between the N- and C-terminal domains to prevent inter- and intra-disulfide bonding, thus simplifying any subsequent mass spectra. The functionalized fl-S23 was then incubated overnight at RT with the omLC/A variant at final concentrations of 2 µM substrate, 50 nM protease in Storage Buffer. The negative control was fl-S23 incubated with Storage Buffer. The samples were analyzed by liquid chromatography-electrospray ionization mass spectrometry using a Xevo G2-XS mass

spectrometer (Waters). The injected samples were subjected to chromatography with a Waters Acquity UPLC Protein BEH C4 Column 300 Å, 1.7 mm, 2.1 mm × 50 mm using the following solvents with a flow-rate of 0.2 mL/min: 100% aq. formic acid (0.1% v/v in water) for 30 s, then a gradient to 97% acetonitrile in aqueous formic acid for 90 s, followed by 97% acetonitrile for 30 s, then a gradient to 100% aq. formic acid for 90 s, followed by 100% aq. formic acid for 60 s. The resulting mass spectra were deconvoluted in MassLynx using the MaxEnt1 algorithm.

Kinetics. LC/A variant kinetic parameters (K_m, k_{cat}) were measured by monitoring the increase in millipolarization units (mP) in a Spark M10 fluorescence polarization plate reader (Tecan) at 28 ± 1 °C over 2 h, excited with polarized light at 380(20) nm and detected with polarized emission at 535(25) nm. The Michaelis–Menten kinetic experiments for DARET SNAP23 and SNAP25 substrates employed a substrate concentration range of 0.2 to 30 µM, with a fixed concentration of 10 nM LC/A variant in 50 mM KH_2PO_4 0.2 nM $ZnCl_2$ pH 7.4. Each concentration of substrate with LC/A variant or buffer (blank or negative control) was assessed in triplicate (n = 3). The average change in polarization over time observed for the buffer was subtracted from the rate observed for each LC/A-incubated well at the substrate concentration; the change in polarization monitored by the blank was exceptionally low. The initial rates (V_0) were derived by fitting the first 25 to 35 min of blank-subtracted rates to a linear regression model in Prism (GraphPad) and calculating the slope. Next, V_0 values were multiplied by their corresponding substrate concentrations as previously described[25], and fit to Michaelis–Menten curves in Prism. Kinetic parameters for wild-type LC/A were derived from a k_{cat} model, and parameters for qmLC/A and omLC/A were fit to a substrate inhibition model. Goodness of fit (R^2) values for wtLC/A (0.99), qmLC/A (0.97 for SNAP25, 0.99 for SNAP23), and omLC/A (0.87 for SNAP25, 0.99 for SNAP23) data guided which models were used. The k_{cat} values were derived by dividing V_{max} by the concentration of protease. For all assays, the blank with the highest substrate concentration was used to optimize the gain and Z-position of the plate reader before addition of protease.

Molecular modeling/visualization. Molecular graphics and structural analyses were performed with UCSF Chimera.

Batch expression and purification of omBoNT/A. The omBoNT/A variant was expressed in 1.5 L of Terrific Broth supplemented with 1% glucose (w/v) until the culture reached an O.D. of 0.6. The culture was then induced with 0.2 mM IPTG and incubated at 16 °C, 265 rpm for 20 h. After cell harvest and lysis, omBoNT/A was purified by IMAC on MagneHis™ resin followed by anion exchange chromatography on a Hitrap Q HP column. Pooled anion exchanged fractions were exchanged into 50 mM Tris pH 8.0, 120 mM NaCl, 0.1 mM ZnCl2 and 5% PEG (v/v) 400 (Sigma). Fractions were analyzed for nicking in the presence or absence of 100 mM DTT by SDS PAGE with Sypro Ruby staining.

SNAP23 in vitro cleavage assay. Human SNAP23 in vitro cleavage was evaluated by incubating 30 µg of human recombinant SNAP23 protein (Novus Biologicals) with 400 nM of either wild-type LC/A, wild type LC/E, or reduced full length omBoNT/A at 37 °C for 1 h in PBS, pH 7.2 (ThermoFisher). The omBoNT/A holotoxin was reduced by incubation with 2 mM TCEP (Tris-(2-carboxyethly) phosphine, hydrochloride (ThermoFisher) at 37 °C for 4 h. The reaction was quenched by addition of NuPAGE™ LDS Sample Buffer (ThermoFisher) and the amount of SNAP23 was assessed by Anti-SNAP-23 Western blot analysis. Reaction samples (≈24 µg) were separated by SDS-PAGE (12% acrylamide) and transferred to nitrocellulose membrane, 0.45 µm pore size (ThermoFisher) in 1X Transfer buffer (ThermoFisher) containing 20% v/v methanol (Honeywell). Membranes were blocked in 2% ECL Prime™ Blocking Agent (GE Healthcare) in TBST (Tris-Buffered Saline (Bio-Rad) with 0.1% Tween-20) for 1 h at RT. Intact SNAP-23 protein was detected with anti-SNAP23 polyclonal antibody (Synaptic Systems) diluted 1:1000 in 2% blocking buffer. Blots were incubated overnight with primary antibody at 4 °C with gentle agitation and then washed in TBST. The bound antibody was detected after 1 h incubation at RT with HRP-conjugated goat anti-rabbit IgG (HL) (ThermoFisher) diluted 1:4000 in 2% blocking buffer. After washing with TBST, the membranes were reacted with Pierce™ ECL Plus Western Blotting Substrate (ThermoFisher) and scanned using the Typhoon 9410 Imager (GE Healthcare).

SNAP23 cleavage assay in cells. Human neuroblastoma SiMa-H1 cells (clonal cell line derived from SiMa cells purchased from DSMZ-German Collection of Microorganisms and Cell Cultures, GmbH, catalogue #ACC 164) were plated at a density of 3 million of cells/well onto collagen IV-coated 6-well plates in serum free media (SFM) supplemented with ganglioside GT1b trisodium (Enzo) at 25 µg/mL. After 3 days of differentiation, SiMa-H1 cells were incubated for 24 h with an adenovirus- human type 5 (dE1/E3) vector encoding ORFs for m-Cherry and human SNAP23 under two independent CMV promoters (Vector biolabs) in SFM supplemented with GT1b. To express mCherry and human SNAP23 in ∼ 100% of cells, the adenovirus was diluted at 75×10^6 PFU/ well in 1 mL. After 24 h, the virus containing media was removed and replaced with fresh SFM (not supplemented with GT1b) in the presence or absence of 50 nM of omBoNT/A or wtBoNT/A toxin (Metabiologics). After 48 h of incubation, cells were lysed in 120 µL of ice-cold lysis buffer containing 20 mM Tris, pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100 and 1X of Halt™ Protease and Phosphatase Inhibitor Cocktail (ThermoFisher). After migration on 12% SDS-PAGE, the proteins were transferred to nitrocellulose membranes, followed by saturating for 1 h at RT with Intercept Blocking Buffer (TBS, Li-Cor), and incubating with a mouse anti-mCherry antibody (1:2000, Abcam,) together with either a rabbit anti-SNAP25 (1:5000, Sigma) or SNAP23 antibody (1:1000, Abcam) in Intercept Antibody diluent (TBS-Tween 20, Li-Cor) overnight at 4 °C. After 3 washes in TBS-Tween (0.05% Tween), membranes were incubated with secondary

antibodies for 1 h at RT (goat anti-mouse IRDye 680RD, 1:5000, Li-Cor; goat anti-rabbit IRDye 800CW, 1: 5000, Li-Cor). The Membrane was then scanned using the Odyssey-CLx imaging system (Li-Cor).

Mouse digital abduction (DAS) assay. All procedures and experiments involving animals were conducted in accordance with relevant guidelines and regulations, and approved by Allergan Animal Care and Use Committee (AACUC; approved protocol #225–100,051-2019). The DAS assay was performed as previously reported[38,39]. In summary, female CD-1 mice (Charles River), with an average weight of 30.2 g and age range of 6 – 10 weeks old, were used in this study. The omBoNT/A and purified BoNT/A1 neurotoxin (Metabiologics Inc., referred to as wtBoNT/A) were diluted in 0.5% human serum albumin in 0.9% saline (Fresenius Kabi, 918,620). For the assay, 0.005 mL of each diluted holotoxin were injected in the right gastrocnemius muscle. Three mice per dose (n = 3) were tested in triplicates (N = 3). The DAS, well-being score, and weight were recorded daily for 4 days. The results were plotted using Prism (GraphPad). Each mouse's well-being was scored on a 4-point system (0 = activity level normal; 1 = slightly diminished activity level and/or slight weight loss (5—10%); 2 = moderately diminished activity level and moderate weight loss (10—15%); 3 = severely diminished activity level, little to no reaction to outside stimuli, inability to ambulate, agonal or labored respiration). To the best of our abilities, the studies were carried out in compliance with the ARRIVE guidelines.

Data availability
All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

References

1. Craik, C. S., Page, M. J. & Madison, E. L. Proteases as therapeutics. *Biochem. J.* **435**, 1–16 (2011).
2. Blasi, J. *et al.* Botulinum neurotoxin a selectively cleaves the synaptic protein SNAP-25. *Nature* **365**, 160–163 (1993).
3. Montecucco, C. & Molgó, J. Botulinal neurotoxins: Revival of an old killer. *Curr. Opin. Pharmacol.* **5**, 274–279 (2005).
4. Dong, M. *et al.* SV2 is the protein receptor for botulinum neurotoxin A. *Science* **312**, 592–596 (2006).
5. Yowler, B. C., Kensinger, R. D. & Schengrund, C. L. Botulinum neurotoxin A activity is dependent upon the presence of specific gangliosides in neuroblastoma cells expressing synaptotagmin I. *J. Biol. Chem.* **277**, 32815–32819 (2002).
6. Koriazova, L. K. & Montal, M. Translocation of botulinum neurotoxin light chain protease through the heavy chain channel. *Nature* **10**, 13–18 (2003).
7. Ma, H. *et al.* Targeted delivery of a SNARE protease to sensory neurons using a single chain antibody (scFv) against the extracellular domain of P2X3 inhibits the release of a pain mediator. *Biochem. J.* **462**, 247–256 (2014).
8. Nugent, M. *et al.* Conjugate of an IgG binding domain with botulinum neurotoxin A lacking the acceptor moiety targets its SNARE protease into TrkA-expressing cells when coupled to anti-TrkA IgG or Fc-βNGF. *Bioconjug. Chem.* **28**, 1684–1692 (2017).
9. Leggett, J. *et al.* GHRH receptor-targeted botulinum neurotoxin selectively inhibits pulsatile GH secretion in male rats. *Endocrinology* **154**, 3305–3318 (2013).
10. Duggan, M. J. *et al.* Inhibition of release of neurotransmitters from rat dorsal root ganglia by a novel conjugate of a clostridium botulinum toxin A endopeptidase fragment and erythrina cristagalli lectin. *J. Biol. Chem.* **277**, 34846–34852 (2002).
11. Arsenault, J. *et al.* Stapling of the botulinum type A protease to growth factors and neuropeptides allows selective targeting of neuroendocrine cells. *J. Neurochem.* **126**, 223–233 (2013).
12. Fonfria, E., Elliott, M., Beard, M., Chaddock, J. A. & Krupp, J. Engineering botulinum toxins to improve and expand targeting and SNARE cleavage activity. *Toxins* **10**, 278 (2018).
13. Pagan, J. K. *et al.* The t-SNARE syntaxin 4 is regulated during macrophage activation to function in membrane traffic and cytokine secretion. *Curr. Biol.* **13**, 156–160 (2003).
14. Frank, S. P. C. C., Thon, K.-P.P., Bischoff, S. C. & Lorentz, A. SNAP-23 and syntaxin-3 are required for chemokine release by mature human mast cells. *Mol. Immunol.* **49**, 353–358 (2011).
15. Kean, M. J. *et al.* VAMP3, syntaxin-13 and SNAP23 are involved in secretion of matrix metalloproteinases, degradation of the extracellular matrix and cell invasion. *J. Cell Sci.* **122**, 4089–4098 (2009).
16. Ren, B. *et al.* SNAP23 is selectively expressed in airway secretory cells and mediates baseline and stimulated mucin secretion. *Biosci. Rep.* **35**, e00220 (2015).
17. Chen, S. & Barbieri, J. T. Engineering botulinum neurotoxin to extend therapeutic intervention. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 9180–9184 (2009).
18. Blum, T. R. *et al.* Phage-assisted evolution of botulinum neurotoxin proteases with reprogrammed specificity. *Science* **371**, 803–810 (2021).
19. Varadarajan, N., Rodriguez, S., Hwang, B.-Y., Georgiou, G. & Iverson, B. L. Highly active and selective endopeptidases with programmed substrate specificities. *Nat. Chem. Biol.* **4**, 290–294 (2008).
20. Guerrero, J. L., O'Malley, M. A. & Daugherty, P. S. Intracellular FRET-based screen for redesigning the specificity of secreted proteases. *ACS Chem. Biol.* **11**, 961–970 (2016).
21. Packer, M. S., Rees, H. A. & Liu, D. R. Phage-assisted continuous evolution of proteases with altered substrate specificity. *Nat. Commun.* **8**, 2 (2017).
22. Sikorra, S. *et al.* Identification and Characterization Of Botulinum Neurotoxin A substrate binding pockets and their re-engineering for human SNAP-23. *J. Mol. Biol.* **428**, 372–384 (2016).
23. Binz, T. *et al.* Mutations in light chain of botulinum neurotoxin a enable cleavage of human SNAP-23. *Toxicon* **156**, S10 (2018).
24. WO2019145577A1.pdf.
25. Gilmore, M. A. *et al.* Depolarization after resonance energy transfer (DARET): A sensitive fluorescence-based assay for botulinum neurotoxin protease activity. *Anal. Biochem.* **413**, 36–42 (2011).
26. Ross, J. A. *et al.* Characterization of Förster resonance energy transfer in a botulinum neurotoxin protease assay. *Anal. Biochem.* **413**, 43–49 (2011).
27. Breidenbach, M. A. & Brunger, A. T. Substrate recognition strategy for botulinum neurotoxin serotype A. *Nature* **432**, 925–929 (2004).
28. Sun, Z., Liu, Q., Qu, G., Feng, Y. & Reetz, M. T. Utility of B-factors in protein science: Interpreting rigidity, flexibility, and internal motion and engineering thermostability. *Chem. Rev.* https://doi.org/10.1021/acs.chemrev.8b00290 (2019).

29. Reetz, M. T. & Carballeira, J. D. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* **2**, 891–903 (2007).
30. Sullivan, B., Walton, A. Z. & Stewart, J. D. Library construction and evaluation for site saturation mutagenesis. *Enzyme Microb. Technol.* **53**, 70–77 (2013).
31. Nov, Y. When second best is good enough: Another probabilistic look at saturation mutagenesis. *Appl. Environ. Microbiol.* **78**, 258–262 (2012).
32. Schmidt-Dannert, C. & Arnold, F. H. Directed evolution of industrial enzymes. *Trends Biotechnol.* **17**, 135–136 (1999).
33. Chen, S., Kim, J.-J.J.P. & Barbieri, J. T. Mechanism of substrate recognition by botulinum neurotoxin serotype A. *J. Biol. Chem.* **282**, 9621–9627 (2007).
34. Silvaggi, N. R. *et al.* Structures of clostridium botulinum neurotoxin serotype a light chain complexed with small-molecule inhibitors highlight active-site flexibility. *Chem. Biol.* **14**, 533–542 (2007).
35. Mizanur, R. M. *et al.* The C terminus of the catalytic domain of type A botulinum neurotoxin may facilitate product release from the active site. *J. Biol. Chem.* **288**, 24223–24233 (2013).
36. Binz, T., Bade, S., Rummel, A., Kollewe, A. & Alves, J. Arg362 and tyr365 of the botulinum neurotoxin type a light chain are involved in transition state stabilization. *Biochemistry* **41**, 1717–1723 (2002).
37. Tang, L. *et al.* Construction of "small-intelligent" focused mutagenesis libraries using well-designed combinatorial degenerate primers. *Biotechniques* **52**, 149–158 (2012).
38. Broide, R. S. *et al.* The rat digit abduction score (DAS) assay: A physiological model for assessing botulinum neurotoxin-induced skeletal muscle paralysis. *Toxicon* **71**, 18–24 (2013).
39. Aoki, K. R. Preclinical update on BOTOX® (botulinum toxin type A)-purified neurotoxin complex relative to other botulinurn neurotoxin preparations. *Eur. J. Neurol.* **6**, s3–s10 (1999).
40. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
41. Boja, E. S. & Fales, H. M. Overalkylation of a protein digest with iodoacetamide. *Anal. Chem.* **73**, 3576–3582 (2001).
42. Fernández-Salas, E. *et al.* Plasma membrane localization signals in the light chain of botulinum neurotoxin. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 3208–3213 (2004).

Author contributions

Funding

Competing interests

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-13617-z.

Correspondence and requests for materials should be addressed to A.D.B.-A. or G.A.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.