BNL-112724-2016-JAAM

Identification of small molecule inhibitors of botulinum neurotoxin serotype E via footprint similarity

Y. Zhou, S. Swaminathan

To be published in "Bioorganic & Medicinal Chemistry"

October 2016

Biology Department

Brookhaven National Laboratory

U.S. Department of Energy

US Department of Defense (DoD),

US Department of Defense (DoD), Defense Threat Reduction Agency (DoD) (DTRA)

Notice: This manuscript has been authored by employees of Brookhaven Science Associates, LLC under Contract No. DE-SC0012704 with the U.S. Department of Energy. The publisher by accepting the manuscript for publication acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or any third party's use or the results of such use of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof or its contractors or subcontractors. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

BNL-112724-2016-JA

Identification of Small Molecule Inhibitors of Botulinum Neurotoxin Serotype E via Footprint Similarity

Yuchen Zhou,[§,a] Brian E. McGillick,[§,b,e] Yu-Han Gary Teng,[c, d] Krupanandan Haranahalli,[d] Iwao Ojima,[c, d] Subramanyam Swaminathan,[e] and Robert C. Rizzo[, a, c, f]*

Submitted to Bioorganic & Medical Chemistry

October 15, 2015

Department/Division/Office

Brookhaven National Laboratory

U.S. Department of Energy
DOE Office of Basic Energy Sciences

Notice: This manuscript has been authored by employees of Brookhaven Science Associates, LLC under Contract No. DE- SC0012704 with the U.S. Department of Energy. The publisher by accepting the manuscript for publication acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or any third party's use or the results of such use of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof or its contractors or subcontractors. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Identification of Small Molecule Inhibitors of Botulinum Neurotoxin Serotype E via Footprint Similarity

Yuchen Zhou,[§, a] *Brian E. McGillick,*[§, b,e] *Yu-Han Gary Teng,*[c, d] *Krupanandan Haranahalli,*[d]

Iwao Ojima,[c, d] *Subramanyam Swaminathan,*[e] *and Robert C. Rizzo**[, a, c, f]

[a] Department of Applied Mathematics & Statistics, Stony Brook University, Stony Brook, NY 11794

[b] Graduate Program in Biochemistry & Structural Biology, Stony Brook University, Stony Brook, NY 11794

[c] Institute of Chemical Biology & Drug Discovery, Stony Brook University, Stony Brook, NY 11794

[d] Department of Chemistry, Stony Brook University, Stony Brook, NY 11794

[e] Biology Department, Brookhaven National Laboratory, Upton, New York 11973

[f] Laufer Center for Physical & Quantitative Biology, Stony Brook University, Stony Brook, NY 11794

Received XX, 2016

[§] These authors contributed equally to this work

* Corresponding author e-mail: rizzorc@gmail.com

Keywords: Botulinum Neurotoxin, BoNT/E, Docking, Virtual screening, DOCK, Footprint similarity, Scoring functions, Molecular dynamics

Abbreviations: BoNT/E, Botulinum neurotoxin serotype E; FPS, footprint similarity; DCE, DOCK Cartesian energy; VDW, van der Waals; ES, electrostatic; MD, molecular dynamics;

ABSTRACT: Botulinum neurotoxins (BoNT) are among the most poisonous substances known, and of the 7 serotypes (A-G) identified thus far at least 4 can cause death in humans. The goal of this work was identification of inhibitors that specifically target the light chain catalytic site of the highly pathogenic but lesser-studied E serotype (BoNT/E). Large-scale computational screening, employing the program DOCK, was used to perform atomic-level docking of 1.4 million small molecules to prioritize those making favorable interactions with the BoNT/E site. In particular, "footprint similarity" (FPS) scoring was used to identify compounds that could potentially mimic features on the known substrate tetrapeptide RIME. Among 92 compounds purchased and experimentally tested, compound C562-1101 emerged as the most promising hit with an apparent IC_{50} value three-fold more potent than that of the first reported BoNT/E small molecule inhibitor NSC-77053. Additional analysis showed the predicted binding pose of C562-1101 was geometrically and energetically stable over an ensemble of structures generated by molecular dynamic simulations and that many of the intended interactions seen with RIME were maintained. Several analogs were also computationally designed and predicted to have further molecular mimicry thereby demonstrating the potential utility of footprint-based scoring protocols to help guide hit refinement.

1. Introduction

Botulinum neurotoxins (BoNT), secreted by the anaerobic bacterium *Clostridium botulinum*, are among the most lethal biological substances known to humans and estimated to have a median lethal dose (LD_{50}) of 1 ng/kg of body weight.[1] There are seven BoNT serotypes (labeled A-G) and all target members of the Soluble N-ethylmaleimide-sensitive factor Attachment Protein Receptor (SNARE) protein family for cleavage which results in the paralysis of muscle tissue.[2,3] In particular, serotypes A, B, E, and F all cause human botulism with A, B, and E being the most common.[1] Given their potential for use as bioweapons,[3] the Centers for Disease Control and Prevention (http://www.cdc.gov) have designated BoNTs as a Category A biological agent, underscoring the importance of studies aimed at developing effective therapeutics.

All of the serotypes share high amino acid sequence homology and activity mechanism.[4] As shown in Figure 1 the toxin is organized as two chains comprised of: (i) a large ~100 kDa "heavy chain" (Figure 1a) consisting of a binding domain (BD) and translocation domain (TD), and (ii) a smaller ~50 kDa "light chain" (Figure 1b) catalytic domain (CD) that includes a zinc dependent active site for substrate cleavage dependent on serotype.[4,5] In terms of mechanism, under normal physiological conditions, members of the SNARE family that include synaptobrevin, syntaxin, and synaptosome-associated protein of 25 kDa (SNAP-25) help assemble the synaptic fusion complex thereby allowing synaptic vesicles to fuse with neuron cell membranes resulting in ordered release of neurotransmitters. In the presence of the toxin, BoNT heavy chains target gangliosides and a specific protein receptor at the presynaptic membrane which results in endocytosis of the toxin. A significant conformational change in the toxin[4] enables the release of BoNT light chains into the surrounding neuronal cytosol resulting in serotype specific cleavage of SNARE proteins. As a consequence the synaptic fusion complex cannot be formed and neurotransmitter release is blocked. Serotypes A, C, and E cleave different sites on SNAP-25 while serotypes B, D, F, and G cleave various sites on synaptobrevin. Serotype C also has activity against syntaxin.[3,6]

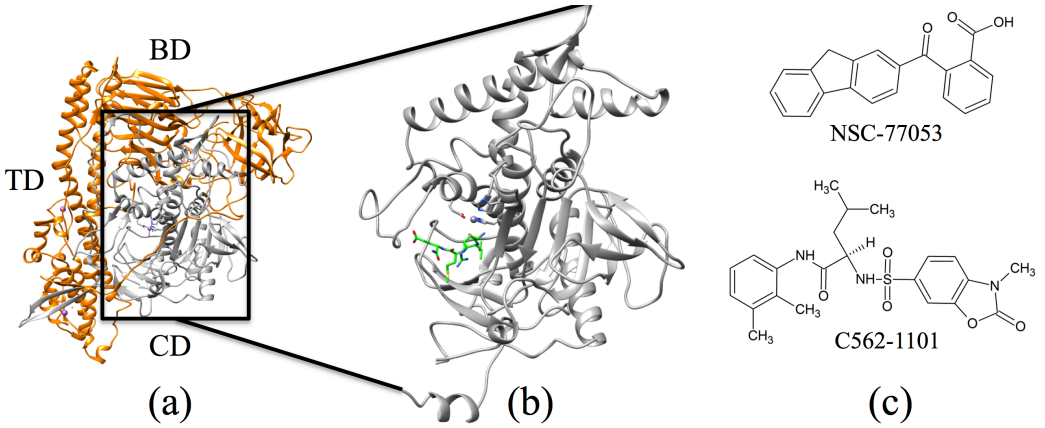


Figure 1. (a) Ribbon structure of BoNT/E showing the heavy chain (orange) consisting of the binding domain (BD) and the translocation domain (TD), and the light chain (gray) comprised of the catalytic domain (CD), coordinates from PDB code 3FFZ.[7] (b) Close-up view of a BoNT/E light chain (gray) co-crystallized with a natural substrate derivative tetrapetide RIME (green), coordinates from PDB code 3D3X.[8] (c) Structures of NSC-77053 (top) and C562-1101 (bottom).

While a number of prior studies focused on vaccine development,[9-12] considerable efforts to develop BoNT inhibitors have also been undertaken. These previous studies focused primarily on serotypes A and B and resulted in the discovery of several classes of inhibitors.[2,13] One such class are polypeptides designed to bind to the BoNT/A light chain active domain. Highlighting the importance of a P1' arginine position for tight binding, Schmidt et al[14] designed the peptide inhibitor CRATKML with a K_i of 2 μM, and subsequently improved the K_i to 330 nM by replacing the cysteine with 2-mercapto-3-phenylpropionyl.[15] Further refinement led to the design of an extremely potent peptide inhibitor DNP-DAB-RWT-DAB-ML with a K_i of 41 nM. Kumaran et al[16] as well as Kumar et al[17] developed a series of tetrapeptide inhibitors displaying low micromolar activity,[13] the most potent comprised of the sequence RRGF, with an IC_{50} of 0.9 μM. Other efforts included the identification of two natural product inhibitors that target the BoNT/A exosite, D-Chicoric acid by Silhar et al[18] with a K_i of 0.7 μM, and Lomofungin identified by Eubanks et al[19] with a K_i of 6.7 μM.

Past efforts to develop small molecules targeting the light chain active site have focused primarily on serotype A.[2,13] Examples include the discovery of para-chloro-cinnamic hydroxamate with an IC_{50} of 15 μM and a subsequently improved analog ortho-para-chloro-cinnamic hydroxamate with an IC_{50} of 0.41 μM.[20] Silvaggi et al[21] determined the crystal structure of ortho-para-chloro-cinnamic

hydroxamate revealing the hydroxamic acid chelates the active site zinc. Building on this finding, Thompson et al[22] designed and tested a series of hydroxamic acid derivatives, among which the most potent compound showed an IC_{50} of 4.6 μM. Quinolinol based small molecules are another important class of compounds showing low micromolar inhibitory activity, with recent examples identified by Burnett et al[23] showing IC_{50} values as low as 10 μM. Additionally, Videnovic et al[24] reported a second generation 4-amino-7-chloroquinoline inhibitor with a K_i value of 103 nM and Caglic et al[25] recently tested a series of quinolinol inhibitors with the most active compound having an IC_{50} of 0.8 μM.

Despite its importance, efforts to develop potent BoNT/E inhibitors have been minimal. Exceptions include work by Agarwal et al[8] who reported a co-crystal structure of the BoNT/E light chain catalytic domain in complex with the tetrapeptide RIME (see Figure 1b) and Kumar et al[26] who subsequently used this complex to identify candidate small molecule inhibitors through computer-based virtual screening. Among the computational candidates, compound NSC-77053 (Figure 1c top) showed an experimental K_i of 1.29 μM using an HPLC-based assay and represents the first reported small molecule inhibitor identified that targets the BoNT/E catalytic site.[26] Recent follow up work by the same group,[27] employing pharmacophore modeling, led to the identification of four roughly equipotent analogs. Importantly, although the RIME peptide is a derivative of the natural substrate SNAP-25 (residues 180 through 183 of SNAP-25 spanning the scissile bond) it is not cleavable by BoNT/E.[28] Thus, the BoNT/E-RIME crystallographic complex[8] represents a native-like substrate-bound state and is a reasonable starting point for continued structure-based inhibitor design.

The goal of the present work is to expand upon previous BoNT/E studies, utilizing a much larger virtual screen against the BoNT/E catalytic site in order to identify new small molecule inhibitor candidates. The primary objectives are fourfold: (1) perform flexible ligand docking of ~1.4 million commercially available small molecules from the ZINC[29] database, (2) employ a recently updated version of the screening program DOCK[30,31] containing enhanced scoring functions[32,33] that allow for identification of compounds that make similar interaction patterns as the bound RIME substrate, (3) purchase and experimentally test a subset of the top-scoring compounds, and (4) perform detailed

structural and computational analysis on compounds showing experimental activity. As described below, this workflow has led to the identification of a novel BoNT/E small molecule inhibitor C562-1101 / ZINC20284316 (Figure 1c bottom) which is about three-fold more potent than inhibitor NSC-77053 (Figure 1c top) when tested in the same GFP-based assay.

2. Methods

2.1 Receptor selection. The crystallographic coordinates selected for virtual screening were based on the structure reported by Agarwal et al[8] consisting of a BoNT/E light chain at a resolution 2.25 Å (PDB: 3D3X) co-crystalized with a substrate SNAP-25 derivative, tetrapeptide RIME. Importantly, use of the substrate-bound structure for virtual screening helps to ensure that the target sites are "preformed" to accommodate a bound ligand. And in the case of RIME, the substrate also provides a baseline reference to help guide identification of compounds that may make similar interactions with the active site zinc ion and surrounding residues as outlined below. For screening, chain B of the structure was selected over chain A given that the tetrapeptide ligand RIME in chain A was not completely resolved, there were 4 missing residues (235-238) in chain A that were in close proximity to the binding site, and chain B had only two missing residues (59, 60) and these were relatively distant from the binding site (> 15 Å). It should be noted that the experimental BoNT/E structure[8] shows the N-terminus of RIME coordinated directly to the active site zinc and therefore unlikely to be protonated. Consequently, the coordinating RIME substrate was modeled as having an uncharged (free NH_2) terminal amine group.

2.2 Docking setup. Docking setups, procedures, and protocols for DOCK generally followed those previously reported.[30,31,34] Briefly, water molecules and sulfate ions were deleted from the initial x-ray coordinates and the *antechamber* and *tleap* programs in the AMBER 14 package[35] were used to assign ff99SB[36] (protein) and GAFF[37] plus AM1-BCC[38,39] (RIME reference) force field parameters. Hydrogen atoms in the complex were then minimized using the *sander* module in AMBER in conjunction with heavy restraints (1000.0 kcal mol^{-1} $Å^{-2}$) on all non-hydrogen atoms. Protein and ligand coordinates

were then extracted and separately saved in MOL2 format for use with DOCK. A final energy minimization allowing the RIME ligand to relax in the context of the DOCK infrastructure was then performed using a tethered restraint (in this case 5.0 kcal mol^{-1} $Å^{-2}$). As expected, the DOCK preparation protocol yielded little deviation in coordinates relative to the initial x-ray structure for residues near (defined as ≤ 6 Å from RIME) the targeted binding site (0.008 Å rmsd) or for the RIME reference itself (0.39 Å rmsd).

2.3 Binding site preparation. Following the receptor preparation and energy minimizations steps, the *sphgen*[40] module in the program DOCK6[31] was used to generate spheres to help guide anchor placement for docking, and only spheres within 8.0 Å to the reference ligand were retained. An energy grid consisting of van der Waals (VDW) and electrostatic (ES) terms was generated using the program *grid*[41] which pre-computes the receptor contributions which speeds up the docking. Grid calculations employed a box surrounding spheres within an 8.0 Å margin in each dimension from the RIME reference and a grid point separation of 0.4 Å. Following standard practice,[34] VDW terms were computed using a 6-9 Lennard-Jones potential which softens the intermolecular energy landscape and ES terms were computed using Coulombs law with a distance dependent dielectric (ddd) = 4r which improves DOCK outcomes.

2.4 Virtual screening protocol. A subset of the ZINC library,[29] consisting of ~ 1.4 M commercially available drug-like small organic molecules, was then screened against the BoNT/E light chain active site using the program DOCK6,[31] with the best grid-based energy pose being retained for each successfully docked compound. The grid-based poses were then energy minimized using the standard DOCK Cartesian energy (DCE) function consisting of Coulombic (ddd =4r) plus Lennard-Jones (in this case 6-12) terms in Cartesian space to reach a more refined local minimum energy binding pose and permit footprint-based terms to be computed as described below.

It should be emphasized that all ZINC libraries were docked as provided in terms of their protonation states or isomeric form. Molecules with different protonation states may exist as separate entries in ZINC libraries but these typically will have the same ZINC code. As standard practice, all are docked individually, however only the best scored molecule with a given ZINC code is retained in later post-processing steps. In terms of optical isomers, these typically have unique ZINC codes and thus from the standpoint of the virtual screening protocol are treated as unique molecules. However, if a compound is identified as active through experimental testing, and the physical vendor sample contains multiple isomeric states, all of the isomers should be more thoroughly studied as exemplified by the discussion presented in sections 2.10 and 3.2 regarding R versus S enantiomers.

2.5 Footprint similarity (FPS) scoring. A recently-developed footprint similarity (FPS) scoring function[32,33] was subsequently used to rescore the docked compounds. FPS score quantifies the similarity of interaction patterns made by candidate compounds with that of a reference compound known to bind to the target, using a per-residue decomposition of the intermolecular VDW and ES energies (based on the DCE function). The goal in this case was to identify compounds that make similar footprint patterns as the tetrapeptide RIME. Prior work in our laboratory has employed FPS scoring alone, and in combination with other functions, to successfully identify active compounds targeting HIVgp41,[42-44] FABP,[45,46] and BoNT/A.[47] Figure 2 graphically illustrates the overall DOCK setup comprised of (a) the BoNT/E light chain (gray) with active site region bounded by the docking grid (magenta), docking spheres (orange), and RIME reference (green), as well as (b) the VDW (top) and ES (bottom) footprints made by bound RIME used to guide compound selection.

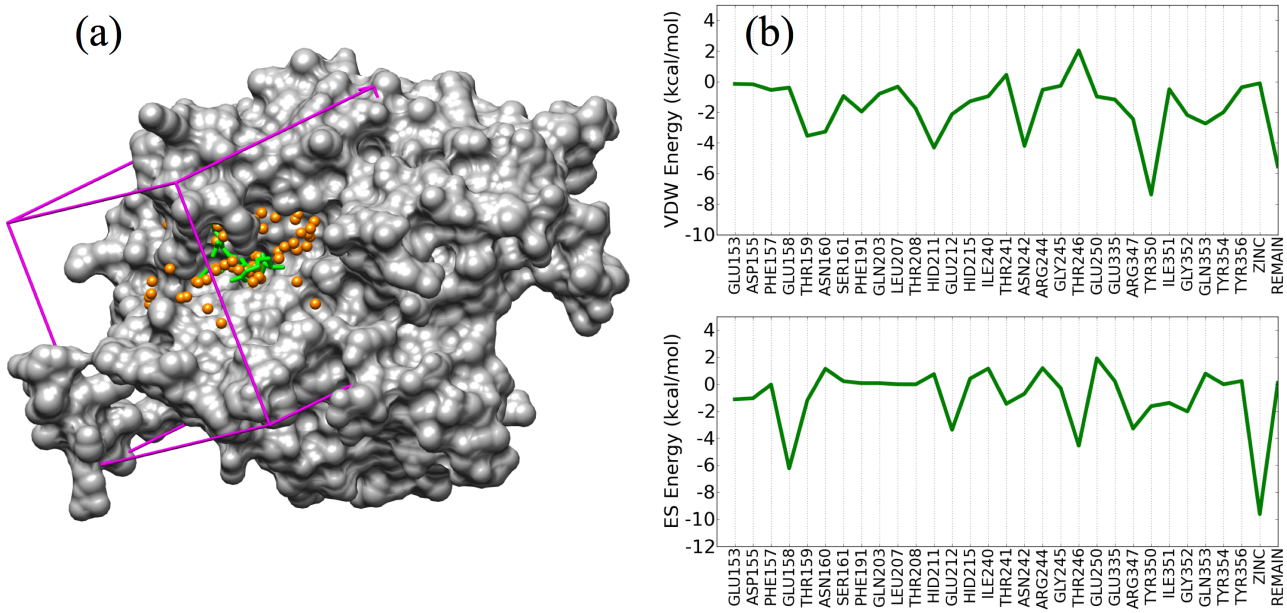


Figure 2. DOCK setup showing (a) the BoNT/E light chain site being targeted through virtual screening. Protein surface in gray, tetrapeptide RIME in green stick model, docking grid in magenta, and docking spheres in orange. Panel (b) shows the per-residue footprints (top 30 residues) made by bound RIME decomposed into VDW (top) and ES (bottom) contributions.

2.6 Compound selection. To improve diversity in the types of compounds chosen for experimental testing, the top-scoring compounds from standard docking were clustered using a pairwise clustering algorithm based on MACCS[48] fingerprint similarity score, as implemented in the program MOE,[49] with a Tanimoto similarity cutoff of 0.75. To help prioritize molecules for purchase, five different rank-ordered lists were then prepared by re-ranking the resulting clusterheads: (1) DCE_{VDW+ES}, DOCK Cartesian energy score consisting of van der Waals plus electrostatic terms, (2) FPS_{VDW+ES}, footprint similarity score consisting of van der Waals plus electrostatic terms, (3) FPS_{VDW}, footprint similarity score consisting of only the van der Waals term, (4) FPS_{ES}, footprint similarity score consisting of only the electrostatic term, and (5) Total Score, a linear combination of DCE_{VDW+ES} and FPS_{VDW+ES}. Following 3D visual inspection of compounds in the binding site ~ 29 representative members from each list were purchased (vendor: ChemDiv, Inc.) for experimental testing.

2.7 Experimental SNAP-Etide assay. Testing of compounds identified through the virtual screen was performed using a commercially available SNAP-Etide substrate (List Biological Inc., CA) as employed

in prior work,[50] following the manufacture recommendations.[51] Assays were carried out in a 100 μL reaction mixture containing an optimized buffer solution containing 20 mM Hepes pH 7.4, 2 mM DTT, 2.5 nM of BoNT/E light chain, and 100 μM of each test compound. A few compounds were tested at lower concentrations due to reduced solubility although none of these showed activity and were not pursued further. Compounds were dissolved in 100% DMSO and added to the reaction mixture to a final concentration of ⩽ 1% DMSO and incubated with the BoNT/E light chain for 15 minutes. The reaction was initialized by the addition of 10 μM SNAP-Etide substrate and conducted at 37°C in 96-well, black, flat-bottomed, non-binding plates (Corning® Costar®) and monitored using a Tecan Saphire2 microplate reader. The excitation and emission wavelengths were set at 320 nm and 420 nm, respectively. The compounds were tested in sets of 10 and were run simultaneously with two DMSO controls. Initial velocities (Vi) of the control runs were averaged and compared to the initial velocity of each test compound with percent inhibition of each compound being calculated as $[1 - (TEST_{(Vi)} \div DMSO_{(Vi)})] \times 100\%$.

2.8 Experimental GFP based assay. Compounds showing activity in the SNAP-Etide assay were re-tested using a secondary assay based on the work of Hines et al[52] that contains a Green Fluorescent Protein (GFP) reporter and employs a larger substrate than SNAP-Etide. Briefly, the assay utilizes a hybrid VAMP2/SNAP-25 substrate with the SNAP-25 portion consisting of residues Ala128-Gly206. The VAMP2 and SNAP-25 domains are connected to one another by a short linker sequence and the entire construct contains an N-terminal AviTag and C-terminal GFP. After biotinylation of the AviTag, the substrate is fixed to Streptavidin coated plates. Subsequent cleavage of substrate by addition of buffer containing BoNT/E light chain liberates GFP into bulk solution allowing for the visualization of a normal enzymatic progress curve. Reactions were performed in 96-well, black, streptavidin coated plates (Pierce™) in a buffer solution containing 20 mM Hepes pH 7.4, 2 mM DTT, 25 nM BoNT/E light chain, and compound. Optimization of the reaction required an approximate 10-fold increase in the BoNT/E LC concentration compared to the SNAP-Etide assay which is likely due to differences

between the two substrates, fluorophores, and other conditions. As before, potential inhibitors were allowed to incubate with BoNT/E light chain for 15 minutes and the solutions were subsequently added to the plate wells to initiate the reactions. All reactions were conducted at 37°C and were monitored using a Tecan Saphire2 microplate reader at excitation and emission wavelengths of 460 nm and 506 nm, respectively.

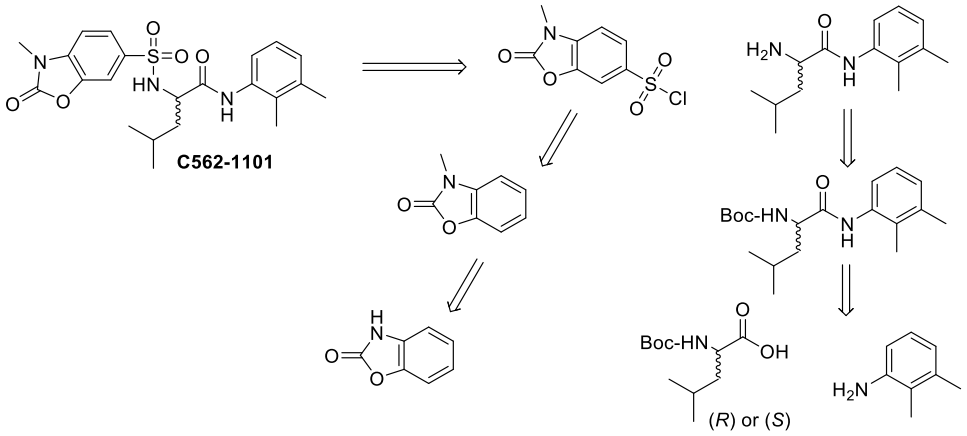
2.9 Molecular dynamics simulations and free energy calculations. Molecular dynamics (MD) simulations were also performed for the most promising hits identified in both assays, as well as several analogs, to examine the overall stability of the predicted binding geometries over an ensemble of structures. Analogous to the DOCK setup protocols described above, MD-ready models were similarly constructed using the AMBER 14 package[35] with the protein being assigned ff99SB[36] parameters and docked ligands being assigned GAFF[37] parameters augmented by AM1-BCC[38,39] partial charges. Systems were solvated using the TIP3P[53] explicit water model and the AMBER *pmemd* module was used to perform energy minimizations and MD in the NPT ensemble at a temperature of 298.15 K to mimic the experimental conditions.

Prior to data collection (20 ns in the present work), a nine-step minimization and dynamics equilibration protocol was employed that enforced gradually decreasing restraints on both the receptor and ligand to relax the complex in a controlled manner. Similar to past MD protocols employed by our laboratory,[43,54,55] the data collection phase employed no restraints on the ligand and only a relatively weak restraint (0.1 kcal mol^{-1} $Å^{-2}$) on protein backbone heavy atoms. To evaluate the geometric stability of predicted binding poses, post-MD analysis included ligand root mean squared deviation (RMSD), distance measurements, and ensemble-based footprint similarity calculations (i.e. FPS scores) which were all calculated using the AMBER *cpptraj*[56] utility.

Free energies of binding (ΔG_{bind}) were also estimated using the implicit solvent MM-GBSA[57,58] "single-trajectory" approach, from periodically saved protein-ligand snapshots extracted from the original explicit-solvent MD trajectories, following protocols outlined in previous work.[54,55,59] It should

be emphasized that since the trajectories were generated with explicit solvent, the effects of structural or ordered waters should be reflected in the resultant ensemble of protein binding site and ligand conformations used for the free energy calculations, although, no detailed water analysis was performed in this work.

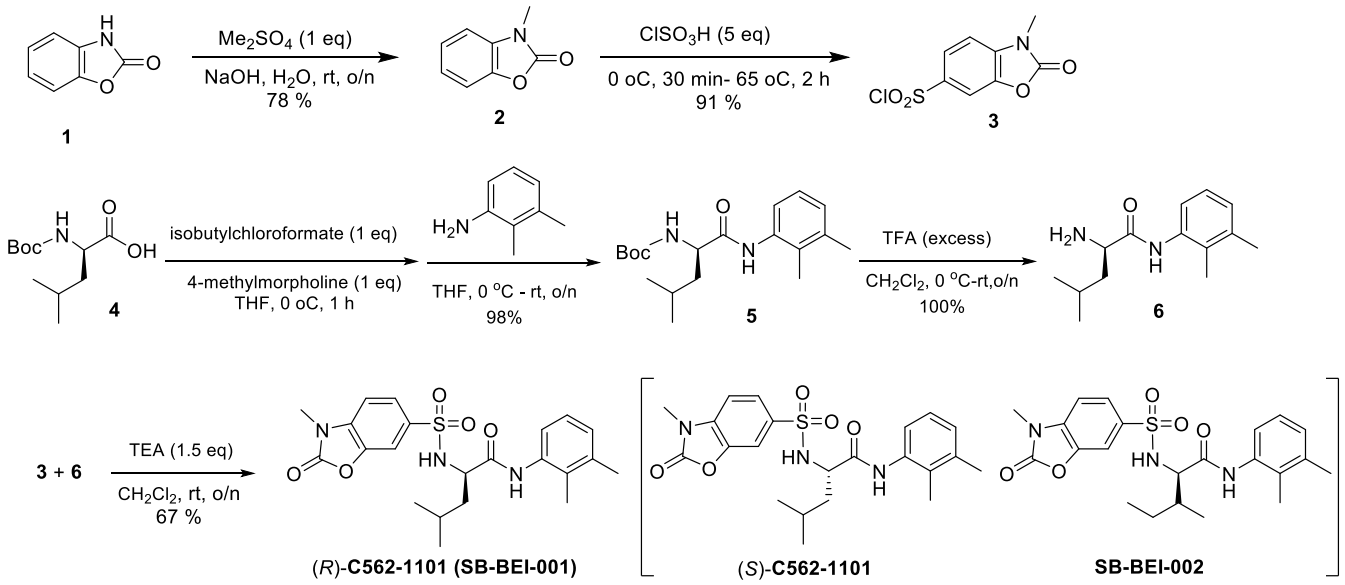
2.10 Chemical synthesis of C562-1101. Although the *R* enantiomer of C562-1101 (ZINC20284316) was the isomer originally identified from the DOCK virtual screen, it was unclear if the sample provided by the vendor for experimental testing was a racemic mixture, or the *R* or *S* enantiomer. Thus, in order to clarify the stereochemical integrity of the sample from the vendor and to examine the activity of *R* and *S* enantiomers, as well as to establish a feasible synthetic route to new analogs for optimization, (*R*)- and (*S*)-C562-1101 were synthesized starting from commercially available (*R*)- and (*S*)-Boc-Leu-OH, and 2-benzoxazolinone based on the retrosynthetic analysis shown in Scheme 1.



Scheme 1. Retrosynthetic analysis of (*R*)- and (*S*)-C562-1101

Synthesis of (*R*)-C562-1101 is illustrated in Scheme 2. 3-Methylbenzoxazolinon-6-ylsulfonyl chloride (**3**), was synthesized from 2-benzoxazolinone (**1**), in two steps in high yield. First, the methylation of **1** using dimethyl sulfate gave 3-methylbenzoxazolinone (**2**), which was then reacted with chlorosulfonic acid to afford **3** in excellent yield. The coupling of (*R*)-Boc-Leu-OH (**4**) with 2,3-

dimethylaniline via mixed anhydride gave (*R*)-Boc-Leu-anilide **5** in virtually quantitative yield. (*R*)-H-Leu-anilide **6**, obtained by deprotection of the Boc group with trifluoroacetic acid (TFA), was reacted with **3** in the presence of trimethylamine (TEA) to give the desired (*R*)-C562-1101 (SB-BTE-001) in good yield. The corresponding *S* enantiomer and (*R*)-Ile analog of C562-1101 (SB-BTE-002) were also synthesized using the same procedure in similar overall yields. See Supplementary Information for details.



Scheme 2. Synthesis of (*R*)-C562-1101 (SB-BEI-001)

2.11 Chiral HPLC analysis of C562-1101 obtained from the vendor. Chiral HPLC analysis of the purity and stereochemistry of the sample obtained from the vendor was carried out using a chiral HPLC column and *R* and *S* enantiomers of C562-1101, synthesized above, as standards. The analysis indicated that the sample of C562-1101 from the vendor was a ~90:10 mixture of *S* and *R* enantiomers (~80% e.e.). Thus, the original C562-1101 vendor sample was the *S* enantiomer arising from (*S*)-Leu-OH. See the Supplementary Information for details.

3. Results and Discussion

3.1 Virtual screening. Figure 3 outlines the overall virtual screening procedure consisting of four primary steps (gray boxes) used to obtain smaller, more focused, sets of results ultimately resulting in 92 compounds being purchased. From the initial ZINC library of ~1.4M compounds docked to BoNT/E, the 100,000 top-scoring molecules (DCE_{VDW+ES} score) were then processed and clustered into groups having similar chemically identity according to the MACCS[48] fingerprinting method. Post processing also included computation of per-residue footprints for each compound and a variety of drug-like descriptors including Lipinski-like terms, number of rotatable bonds, and formal charge, among others, to assist with compound prioritization. Using the five different scoring criteria described above in methods (DCE_{VDW+ES}, FPS_{VDW+ES}, FPS_{VDW}, FPS_{ES}, Total Score) the resultant clusterheads were then re-ranked and the top 2,000 from each list were retained to aid compound selection (see Figure 4). The goal is twofold: (i) maximize the diversity of compounds prioritized for purchase by using clusterheads of a given family of related compounds, and (ii) employ different ranking functions to eliminate bias from any one particular scoring method. In general, each of the 5 lists should contain a diverse set of compounds that will be unique from those in different ranked lists.

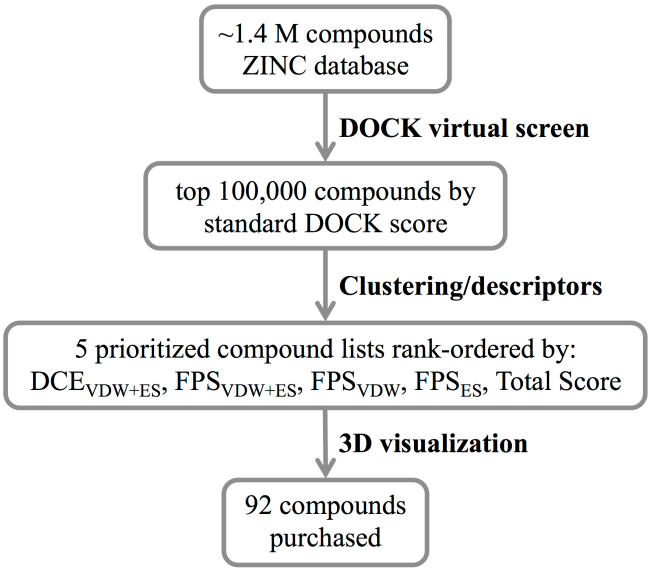


Figure 3. Virtual screening protocol for targeting the BoNT/E light chain active domain.

3.1.1 Ligand properties. Figure 4 emphasizes the unique characteristics of each of the five ranked sets of ligands in terms of nine different descriptors: (a) DCE_{VDW+ES} energy, (b) DCE_{VDW} energy, (c) DCE_{ES} energy, (d) FPS_{VDW+ES} distance, (e) FPS_{VDW} distance, (f) FPS_{ES} distance, (g) Total Score, (h) molecular weight, and (i) number of rotatable bonds. As expected, use of a specific scoring function will show bias for specific chemical or physical properties. For example, clusterheads selected by the DCE_{VDW+ES} function always have the most favorable DCE_{VDW+ES} scores (Figure 4a black line) as well as strongly favorable DCE_{VDW} (Figure 4b black line) and DCE_{ES} (Figure 4c black line) terms which are the components of the DCE_{VDW+ES} score. Similarly, compounds selected using FPS_{VDW+ES} (Figures 4d green), FPS_{VDW} (Figures 4e blue), or FPS_{ES} (Figures 4f red) functions yield more compounds with more favorable scores in each of their respective histograms. Compounds selected with Total Score, intended to provide a balance between DCE_{VDW+ES} (which favors compounds with the most favorable receptor interactions) and FPS_{VDW+ES} (which favors compounds with the most similar interactions as a reference molecule), analogously yields the "largest peak" in its respective histogram (Figure 4g orange). The clusterheads selected using Total Score yield population distributions roughly in between that obtained with DCE_{VDW+ES} (Figure 4a orange versus black or green) and FPS_{VDW+ES} (Figure 4d orange versus black or green) indicative of docked poses with both favorable energy scores and reasonable molecular mimicry in terms of the RIME reference.

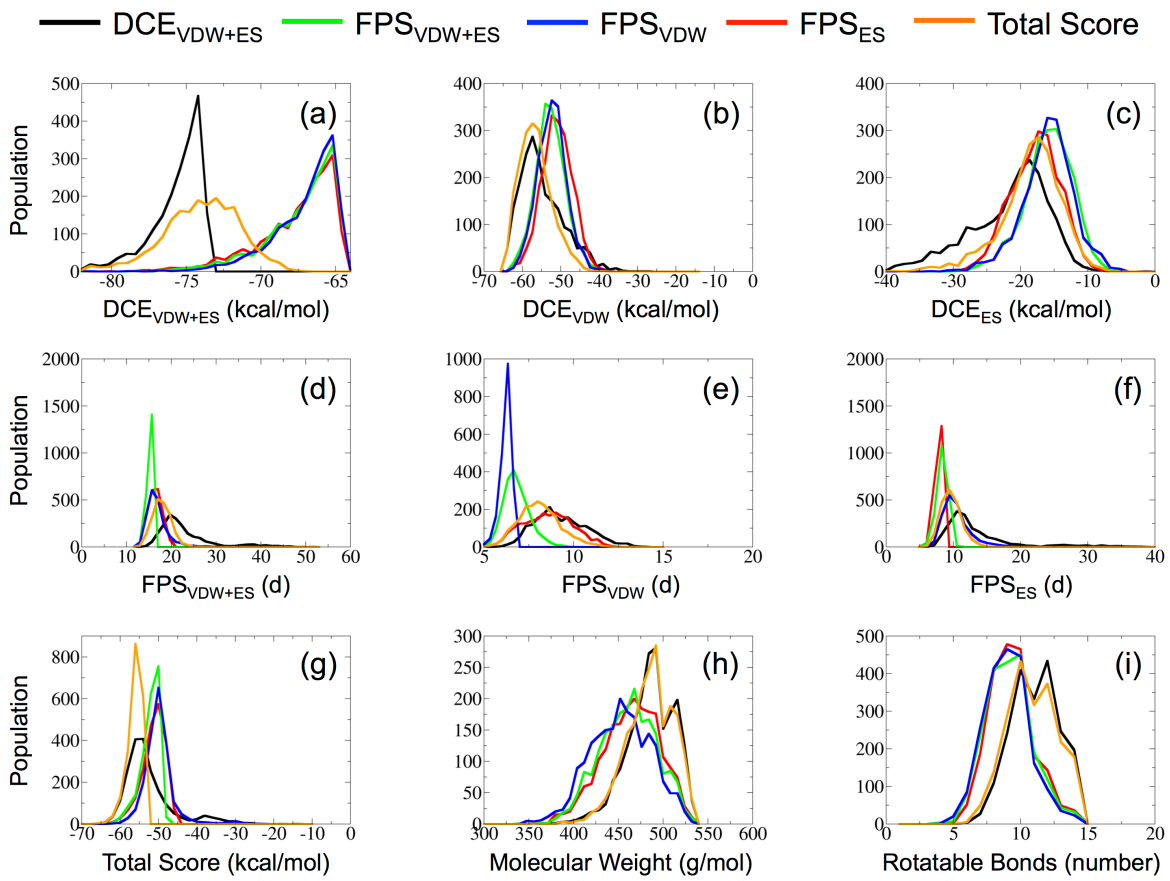


Figure 4. Histograms for top-ranked compounds (N=2000 clusterheads) prioritized by 1 of 5 different DOCK scoring functions (DCE_{VDW+ES} = black, FPS_{VDW+ES} = green, FPS_{VDW} = blue, FPS_{ES} = red, Total Score = orange) plotted as 9 different descriptors: (a) DCE_{VDW+ES} energy, (b) DCE_{VDW} energy, (c) DCE_{ES} energy, (d) FPS_{VDW+ES} distance, (e) FPS_{VDW} distance, (f) FPS_{ES} distance, (g) Total Score, (h) molecular weight, and (i) number of rotatable bonds.

As expected, use of the standard DOCK function alone (i.e. DCE_{VDW+ES}) biases towards larger compounds compared to that of FPS functions alone as illustrated by higher molecular weights (Figure 4h black) and greater number of rotatable bonds (Figure 4i black). Somewhat surprisingly, use of Total Score, containing both DCE and FPS terms, yields nearly the same trends for both descriptors (Figure 4h,i black vs orange) despite the fact the compound sets have quite different energetic properties (Figure 4h,i black vs orange). In sharp contrast, use of the three FPS functions alone shows much broader distributions of molecular weight and down shifted numbers of rotatable bonds indicative of smaller molecules.

3.1.2 Compounds in common. Table 1 shows the number of molecules in common between different rank-ordered lists for a smaller subset consisting of the top-scoring 200 compounds (a smaller more practical number that one might reasonably consider for purchase). For these experiments, it is apparent that use of footprint similarity scoring provides greater diversity in terms of which compound would be chosen, as opposed to if only the DCE_{VDW+ES} function was used. Specifically, there are no compounds in common between the top 200 DCE_{VDW+ES} compounds and any of lists derived from the three footprint similarity scores. Not surprisingly, there is overlap between the DCE_{VDW+ES} and Total Score lists (N=72 compounds). Among the three different footprint similarity derived lists, there is also diversity although not nearly as extreme. Use of FPS_{VDW+ES} yields between ~20-25% in common with either the van der Waals or electrostatic interaction footprint similarity scores alone. Interestingly, use of FPS_{VDW+ES} yields only a small number of compounds in common with Total Score (N=5), in sharp contrast to the value of 72 noted above for the other-way-round, despite the fact that Total Score is a combination of DCE_{VDW+ES} plus FPS_{VDW+ES}. Given these observations, with the present protocols, Total Score appears to be dominated by the standard DCE_{VDW+ES} score although this could easily be adjusted in future experiments by increasing the "weight" coefficient used when combining the two values.

Table 1
Number of overlapping compounds between top 200 compounds from each rank ordered list.

	DCE_{VDW+ES}	FPS_{VDW+ES}	FPS_{VDW}	FPS_{ES}	Total Score
DCE_{VDW+ES}	200	0	0	0	72
FPS_{VDW+ES}		200	48	38	5
FPS_{VDW}			200	2	0
FPS_{ES}				200	6
Total Score					200

Figure 5 shows the ensemble results visually in terms of binding geometries. As expected, the ensemble prioritized using DCE_{VDW+ES} (Figure 5a), while still within the general area of the binding site, shows compounds significantly spread out over a larger surface. In contrast, compounds identified by

FPS_{VDW+ES} (Figure 5b) are more tightly clustered in the BoNT/E binding site, as they are more likely to make contacts with amino acid side chains in a similar manner as the RIME reference used in prioritization. Notwithstanding the discussion above regarding the importance of individual score components, the use of Total Score to derive an ensemble of 200 compounds (Figure 5c) qualitatively leads to molecules occupying space in the binding site roughly "in-between" that of the other two functions (Figures 5a or 5b), as was the original intent.

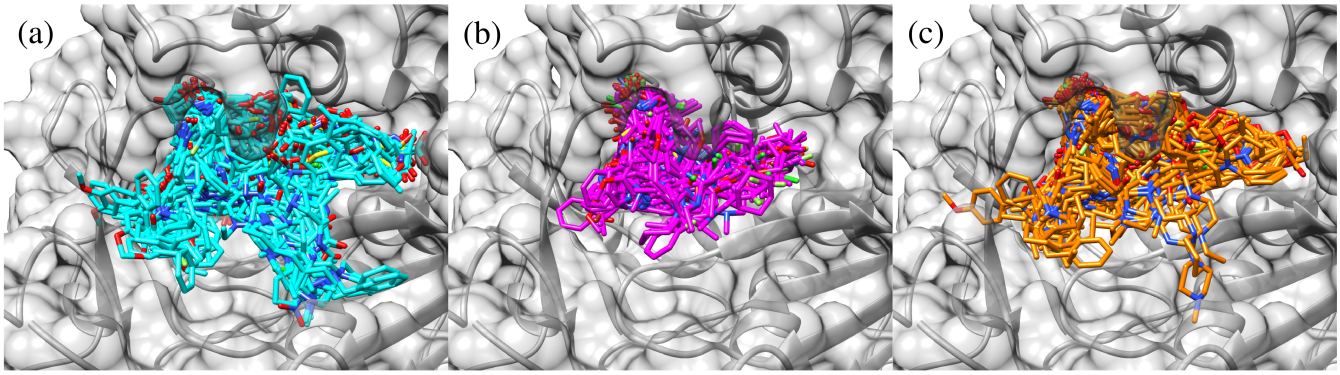


Figure 5. Overlay of the top 200 scoring compounds in the BoNT/E site prioritized using (a) DCE_{VDW+ES}, (b) FPS_{VDW+ES}, or (c) Total Score functions.

3.1.3 Compound selection. BoNT/E is a zinc dependent protease thus good coordination to the active site zinc ion was emphasized during compound prioritization in an attempt to identify compounds that "mimic" known substrates. This was primarily done through visual inspection of DOCK predicted binding poses for candidate compounds as well as requiring a strong electrostatic interaction peak with zinc ion in footprint similarity (FPS) scoring plots. Observed zinc coordinating groups included carbonyl oxygens, nitrogens, sulfurs, sulfones, and carboxylic acids. Additional criteria for eliminating compounds included number of stereo centers, formal charges, molecular weight, and number of rotatable bonds. In total, 122 compounds were prioritized for experimental testing from a roughly even breakdown in terms of which of the five scoring functions was used for prioritization: DCE_{VDW+ES} = 29, FPS_{VDW+ES} = 26, FPS_{ES}= 27, FPS_{VDW}= 26, and Total Score = 30. As was illustrated in Table 1, in some

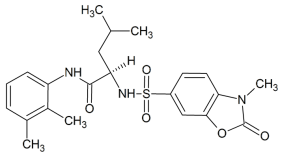
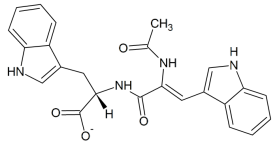
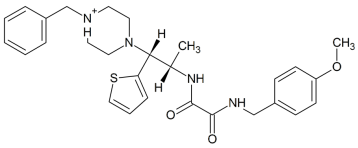
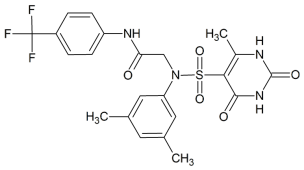
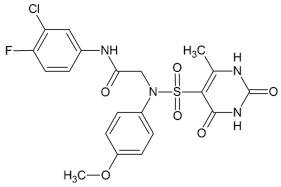
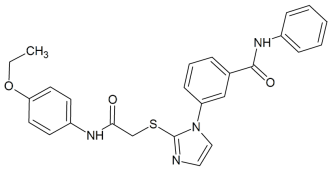
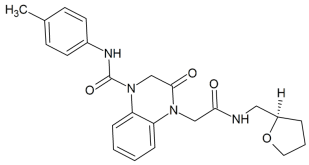
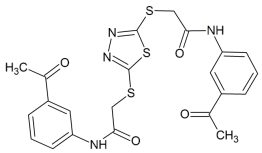
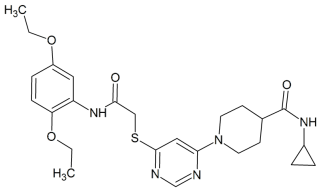
cases compounds are present in more than one group selected by different prioritizing criteria. Ultimately, 92 unique commercially available compounds were purchased for experimental testing.

3.2 Experimental testing. During optimization of the experimental SNAP-Etide assay, run-to-run deviations of initial velocity data between control runs was up to ±10%, therefore a cutoff of >20% inhibition at 100 μM was used as the criteria for a "positive hit". We would also emphasize that the goal of the initial experimental tests was to rapidly eliminate compounds showing no inhibition, even at 100 μM, and our use of a 20% cutoff helps to ensure we will not miss any potential hits. However, at this relatively high concentration there is an increased risk of false positives due to non-specific interactions. Therefore, as described below, the initial hits were tested in a secondary assay at a range of concentrations.

Of the 92 compounds purchased from the vendor based on the computational screen, 9 compounds showed 25-71% inhibition in the initial assay when tested at a concentration of 100 μM as shown in Table 2 which displays compound identifiers, in silico descriptors, % activity, and 2D structures. Among the group, two repeating features include a sulfonyl moiety which was present in 3 of the compounds as well as substituted phenylacetamide-like groups in 7 of the compounds. Interestingly, the compound with the largest % activity (C562-1101) has the second best footprint overlap (FPS_{VDW+ES} = 14.29) although this is likely to be fortuitous as FPS scores (like DOCK scores) are not expected to, in a quantitative sense, have strong correlation with experimental activities across a group of molecules. A noteworthy observation is that the relatively poorer DOCK score for C562-1101 (-65.89 kcal/mol), versus other docked compounds (see Figure 4a), indicates this hit would likely not have been selected for experimental testing if only the standard DCE_{VDW+ES} function were employed in the selection process.

Table 2

Compound identifiers, in silico descriptors, % activity, and 2D structures for 9 hits out of 92 experimentally tested compounds using the SNAP-Etide assay.

Code	DCE_{VDW+ES} (kcal/mol)	FPS_{VDW+ES} (d)	FPS_{VDW} (d)	FPS_{ES} (d)	Tot Score (kcal/mol)	Activity (%)	Structure
C562-1101	-65.89	14.29	5.96	8.33	-51.60	71	
8010-8498	-80.20	41.40	9.61	31.80	-38.80	52	
C609-0236	-78.77	20.73	8.60	12.13	-58.04	47	
C548-1327	-65.21	16.22	8.83	7.39	-48.99	42	
C548-0436	-67.80	16.42	9.40	7.02	-51.39	41	
F985-1368	-74.87	15.93	6.88	9.06	-58.94	40	
G741-1024	-66.56	14.11	5.62	8.48	-52.45	33	
8014-7190	-76.75	17.49	7.75	9.74	-59.27	31	
M779-0919	-75.28	15.33	7.28	8.05	-59.95	25	

Visualization of the docked poses for the actives (Figure 6 orange ligands) show that all make interactions with the catalytic ion via an O --- Zn motif with 7 of the 9 showing very similar coordination geometries as is made by the C-terminal backbone carbonyl of the native substrate RIME. Ligand O --- Zn distances (Figure 6 orange ligands) range from 1.8 to 2.5 Å in comparison to RIME (Figure 6 green ligand) at 2.4 Å.

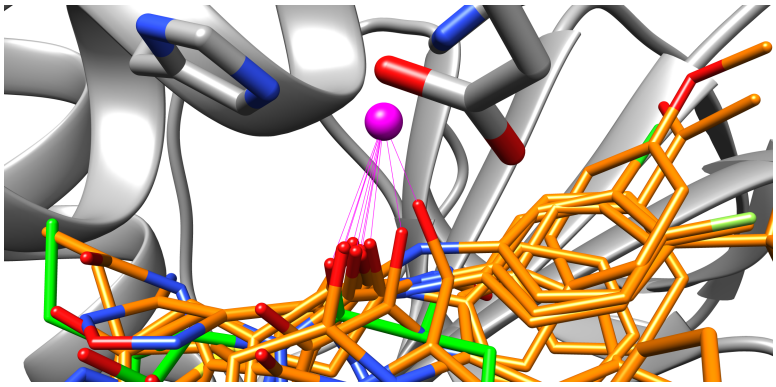


Figure 6. Three dimensional pose overlays for 9 docked compounds (orange) showing similar interaction distances (1.8 to 2.5 Å) between oxygen (red) and zinc (magenta) as the native tetrapeptide RIME ligand (green, 2.4 Å).

These 9 hits were also examined using a secondary GFP-based assay (see Methods), in triplicate, to help alleviate false positives common to fluorescence assays. By using a fluorescent molecule with a different excitation and emission wavelength than that employed in SNAP-Etide, activity demonstrated by a candidate inhibitor in both assays is unlikely to be due solely to interference with the fluorophore. The second assay also employed a larger SNAP-25 construct which is potentially more indicative of the biologically-relevant substrate. Somewhat surprisingly, only 1 of the 9 compounds (Table 2, compound C562-1101) active in the first SNAP-Etide assay displayed activity in the second GFP assay. As expected, the positive control compound NSC-77053 also showed activity. Interestingly, C562-1101 (percent inhibition 71%) was the most potent of all 9 compounds tested in both assays.

As described in the Methods section (*vide supra*), the most promising hit (C562-1101), was re-synthesized in-house in its enantiomerically pure (*R*)- and (*S*)-forms so that each enantiomer could be

tested in the second GFP-based assay. The chiral HPLC analysis of the initial C562-1101 sample provided by the vendor, using the pure *R* and *S* enantiomers as standards, indicated that it was an *S* enantiomer with ca. 80% enantiomeric purity (Supplementary Information). In contrast, the form originally identified from the computational screen was for the R enantiomer. Unexpectedly, the ZINC database contained unique entries for *(R)*-C562-1101 (ZINC20284316) and *(S)*-C562-1101 (ZINC20284319) although both entries code to the same ChemDiv ID of C562-1101. Nevertheless, in terms of experimental activity, it appears that there is no significant difference in activity between the pure *R* and *S* enantiomers when tested at three different concentrations (Table 3). Interestingly however, the predicted binding geometries for R and S were very different.

Table 3
Percent inhibition for enantiomers of C562-1101 in the GFP-based assay.

Concentration	***R* enantiomer**	***S* enantiomer**
100 μM	84.0% ± 4.0%	77.5% ± 3.5%
25 μM	58.0% ± 7.0%	62.5% ± 7.5%
12.5 μM	46.5% ± 5.5%	52.0% ± 2.0%

To investigate why docking the two enantiomers yielded different predicted binding geometries, the opposite isomer was manually constructed in each case starting from each respective R and S pose. Manual construction of S, based on the predicted geometry for R, resulted in unfavorable clashes in the binding site that could not be ameliorated through simple rotamer exploration of the isobutyl group. At this time, DOCK does not directly support receptor flexibility, thus, with the current binding site conformation derived from the x-ray coordinates, only the R enantiomer can adopt the predicted geometry. Conversely, although manual construction of R based on the DOCK-predicted geometry for S did not result in an unfavorable clash with the protein, the DCE score at only -38.8 kcal/mol was significantly less favorable than either the docked S (-60.3 kcal/mol) or docked R (-65.9 kcal/mol) forms. In terms of the experimental results, one possible explanation is that induce fit effects not currently captured in our molecular modeling approach allow both R and S enantiomers to adopt the

same basic pose in the BotNT/E site which leads to similar activity. Ideally, co-crystallization could be used to help resolve this question.

From multiple structural and energetic standpoints, the docked *R* enantiomer was more favorable than *S* in terms of: (1) energy score (-65.9 vs -60.3 kcal/mol), (2) footprint similarity with RIME (14.3 vs 27.3 Euclidian distance), (3) volume overlap with RIME (52 vs 30 % overlap), and (4) number of matched pharmacophore features[60] with RIME (5 vs 0). In addition, subsequent molecular dynamics (MD) simulations of the *R* enantiomer showed good geometric and energetic stability (see discussion below). For these reasons, follow-up analysis presented below and computational optimization efforts have focused on the *R* enantiomer unless otherwise noted.

Figure 7 plots the dose response curves for (*R*)-C562-1101 and for comparison NSC-77053, which were performed in triplicate over a range of compound concentrations under the same assay conditions. IC_{50} values were obtained by fitting to a sigmoid dose-response equation using a hillslope = 1. Encouragingly, in the GFP-based assay, the estimated IC_{50} value for (*R*)-C562-1101 at 14.2 ± 1.7 μM is about three-fold lower than that of NSC-77053 at 48.5 ± 6.5 μM. It is important however to emphasize the identification of (*R*)-C562-1101 represents only a first-stage lead discovery outcome and that follow-up work to evaluate properties other than activity should be performed including metabolic, toxic, pharmacokinetic, and related physicochemical descriptors.

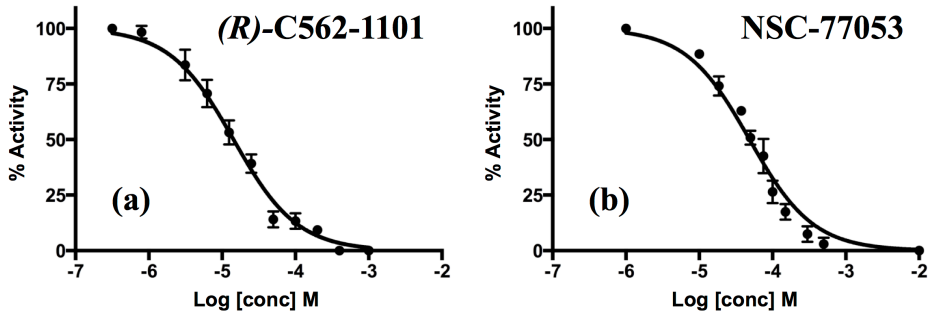


Figure 7. Dose response plots for (a) (*R*)-C562-1101 and (b) control NSC-77053.

Overall, the current results suggest that (*R*)-C562-1101 is a promising hit compound for optimization. Since the ligand is considered a mimic of the known substrate tetrapeptide RIME (see

Figure 9a) that includes isoleucine in the sequence, a (*R*)-Ile analog of (*R*)-C562-1101 (SB-BEI-002) was also synthesized (see the Methods section). It was found that the inhibitory activity of this analog was in the same level (IC_{50} ~16 μM) as that of (*R*)-C562-1101 (IC_{50} ~14 μM), i.e., no improvement was observed. In parallel, as additional early steps towards refinement, several hypothetical analogs were constructed and computationally evaluated using the predicted DOCK structure as a guide (see section 3.5).

3.3 MD analysis of active compounds. In conjunction with experimental testing, fully flexible explicit-solvent MD simulations were carried out for each of the 9 compounds identified in the initial SNAP-Etide assay to help gauge the overall steric and electrostatic stability of the compounds in the binding site in terms of their predicted binding geometry. Figure 8 shows running block-averaged RMSD, and estimated free energies of binding (ΔG_{bind}) using the MM-GBSA method (see section 2.9) over a 20 ns simulation for each ligand initiated from the docked pose. Interestingly, although a number of compounds reasonably maintain their predicted binding pose and show stable ΔG_{bind} values, compound (*R*)-C562-1101, which was the most active compound in the SNAP-Etide assay and the only one that displayed activity in the GFP-based assay, was particularly well-behaved. Specifically, (*R*)-C562-1101 maintains a low ligand RMSD of ~2 Å (Figure 8a, orange line) during the 20 ns simulation and had one of the most favorable ΔG_{bind} values (Figure 8b, orange line). Visualization of (*R*)-C562-1101 geometries taken along the trajectory (Figure 8c) confirms the stability of the binding pose. It should be emphasized that although such ligand behavior during MD simulations in-and-of-itself does not guarantee a given compound will be an inhibitor, it does provide additional supporting evidence for compatibility with the target. A recent study[43] by our group examined the effects of using MD simulations to gauge structural and energetic variability of binding geometries through comparison of experimentally observed poses vs low-energy decoys over a large test set. The study also examined the behavior of active and inactive compounds. While additional investigations should be pursued, taken

together, the results suggest that, on average, active-like compounds show good overall behavior in terms of stable and low ligand RMSDs and ΔG_{bind} values.

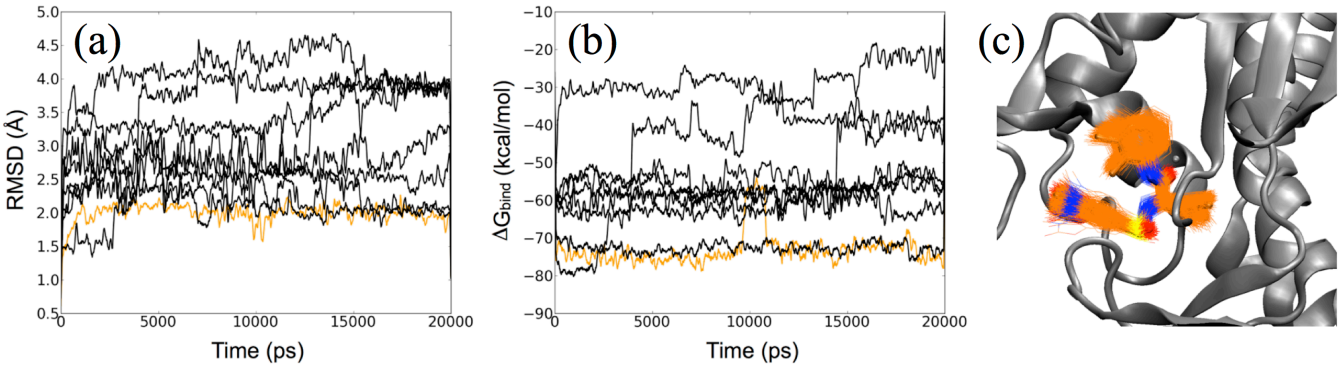


Figure 8. (a) Ligand RMSDs and (b) estimated ΔG_{bind} values vs time for 9 compounds showing activity in the SNAP-Etide assay. (*R*)-C562-1101 shown with an orange line. (c) Structural overlays for (*R*)-C562-1101 from 100 evenly spaced snapshots along the trajectory.

3.4 Binding analysis of (*R*)-C562-1101. Together, the experimental assay results and computational analysis suggest (*R*)-C562-1101 is most promising compound identified from the virtual screen and will be the focus of the discussions that follow. Notably, (*R*)-C562-1101 was selected from among top-ranked compounds in the FPS_{VDW+ES} ranked list and therefore expected to share similar geometric features and make similar energetic interactions as the reference tetrapeptide RIME in the BoNT/E catalytic site. Visual overlay of the two binding poses (docked (*R*)-C562-1101, x-ray RIME) confirms significant structural overlap (Figure 9a) with several notable features in common (Figure 9b vs 9c): (1) the carbonyl oxygen of the pentanamide moiety of (*R*)-C562-1101 occupies a nearly identical position as the backbone oxygen of arginine on RIME, which displaces the nucleophilic water present seen in the BoNT/E apo structure, and is situated to make similar contacts with both the catalytic zinc ion and the hydroxyl of Tyr350 (Figure 9 dashed lines), (2) the hydrophobic isobutyl group of (*R*)-C562-1101 occupies a nearly identical position as the isoleucine side chain of RIME for which both are buried in the S1' hydrophobic pocket[8] formed by residues Thr159, Phe191, and Thr208, (3) a sulfonamide oxygen of (*R*)-C562-1101 overlays with the backbone oxygen of isoleucine on RIME and positioned to

make a similar H-bond with Arg347 (Figure 9 dashed lines), and (4) the substituted benzoxazole group of (*R*)-C562-1101 overlaps structurally with carboxylic acid on RIME and positioned to make similar H-bonds with the backbone amide hydrogen of Gly352 and the hydroxyl side chain of Thr246 (Figure 9 dashed lines). Major differences include the ability of RIME to make a salt bridge (Glu158) and extra hydrogen bond (Glu212) as discussed further below (Figure 9b vs 9c).

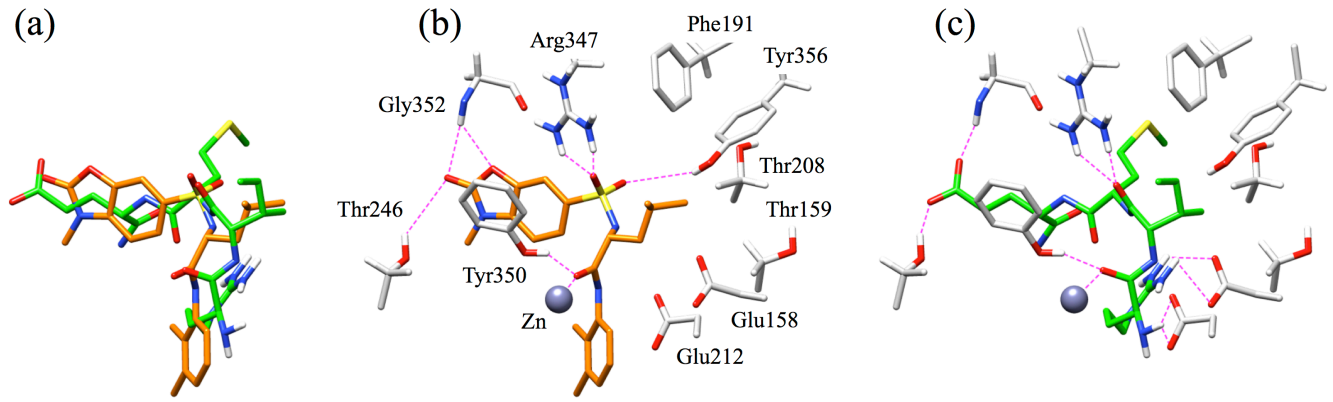


Figure 9. (a) (*R*)-C562-1101 (orange) and tetrapeptide RIME (green) in the BoNT/E light chain active domain binding pocket with protein residues hidden for clarity. (b) Subset of residues making key interactions with (*R*)-C562-1101 (orange). (c) Subset of residues making key interactions with RIME (green). Interactions within reasonable H-bonding distance (1.7 - 3.3 Å) shown in magenta.

In terms of quantitative energetic similarities, Figure 10 compares the per-residue energy footprints made by docked (*R*)-C562-1101 (Figure 10, orange) versus x-ray RIME (Figure 10, green) for the fifty BoNT/E residues with the most significant van der Waals (VDW) interactions and electrostatic (ES) interactions shown explicitly, and lesser contributions summed together into the residue labeled REMAIN. For comparison, the plot also shows an "average" footprint made by (*R*)-C562-1101 over the 20 ns MD simulation (Figure 10, blue). In terms of the VDW similarities (Figure 10 top, orange vs green), (*R*)-C562-1101 well-matches the overall RIME footprint with near perfect overlap in terms of magnitude at residues Thr159, Asn160, Phe191, His211, Asn242, Glu250, Tyr350, and Gly352. The principal difference is the favorable VDW peak at Thr246 observed for (*R*)-C562-1101 for which RIME is unfavorable (discussed further below). Importantly, the MD-averaged footprint for (*R*)-C562-1101 (Figure 10 top, blue) shows only minor VDW differences versus the originally docked pose (Figure 10

top, orange) thereby demonstrating that RIME-like steric packing interactions can be maintained by a small organic ligand during normal thermal fluctuations.

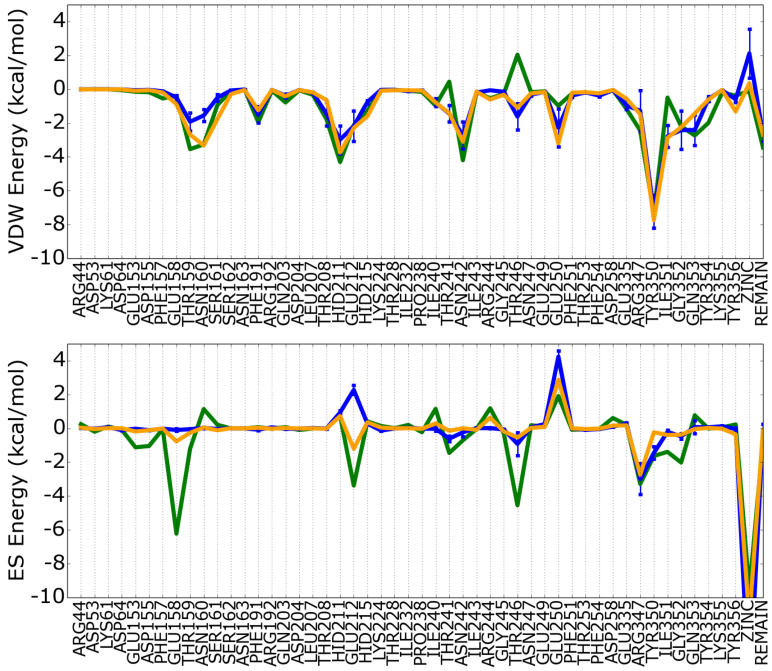


Figure 10. Footprint comparison plots with the BoNT/E light chain active domain for the crystallographic reference pose for tetrapeptide RIME (green), docked pose for (*R*)-C562-1101 (orange), and average footprint over an ensemble of conformations from 20 ns MD simulations (blue). The fifty most significant residues are shown explicitly with remaining interactions combined into the residue labeled REMAIN.

However, in contrast to the high overlap seen in the VDW footprints, (*R*)-C562-1101 (Figure 10 bottom, orange) does not show the same quantitative correspondence as RIME in terms of the magnitude of ES peaks with the notable exception of Arg347 and zinc. A primary cause of the difference is the presence of charged arginine and glutamic acid functionality on RIME which can make much stronger H-bonds than the neutral functionality of (*R*)-C562-1101 as well as form salt bridges. As an example, strong ES interactions (Figure 10 bottom, large green peaks) are observed for R with Glu158, the N-terminus of R with Glu212, for E with Thr246, and for E with Gly352 with computed contributions in the range of -2 to -6 kcal/mol (Figure 10 bottom, green peaks). Taken together, the visual analysis and ES footprints results suggests: (1) a lack of comparable functionality on (*R*)-C562-1101 explains the absence of strong interactions with Glu158 and Glu212, (2) a lack of charged

functionality explains the small ES peaks at Thr246 and Glu352, and (3) the reduced peak at Tyr350 is attributed to sub-optimal H-bonding geometry. As described in the next section, a series of (*R*)-C562-1101 analogs were devised to try and recapitulate the ES patterns made by RIME, for example, by adding charged ligand functionality to mimic RIME's H-bonding within the BoNT/E site. In general the "average" ES footprint for (*R*)-C562-1101, similar to that seen for the VDW patterns, shows only minor differences with regards to the originally docked pose (Figure 10 bottom, orange vs blue) indicating that the ES interactions are also maintained during the MD simulations.

An interesting result to note is that the MD results predict a smaller distance (2.00 Å ± 0.11 Å MD-average vs 2.39 Å docked) between the (*R*)-C562-1101 carbonyl oxygen with zinc thus yielding a stronger ES interaction (Figure 10c bottom, residue ZINC, blue). As a consequence, the isobutyl group of (*R*)-C562-1101 is pulled slightly away from the S1' pocket which leads to somewhat less favorable van der Waals interactions with Thr159 and Asn160 (Figure 10 top, orange vs blue), and the initially favorable ES interaction with Glu212 becomes unfavorable (Figure 10 bottom, orange vs blue). Such behavior for VDW footprints is often observed when very strong ES interactions are involved. Stated another way, although the total interaction energy for a given ligand may be favorable, energy decomposition can yield generally small unfavorable interactions at specific residues as a result of very favorable ES interactions often involving ions.

3.5 Computationally designed analogs. To computationally investigate how functional group alteration would alter ligand behavior in the binding pocket, and to predict if affinity of the initial hit could be improved, we generated 7 hypothetical analogs of (*R*)-C562-1101 (Figure 11, compounds SBA1 to SBA7) and performed MD simulations, footprint similarity analysis, and estimated free energies of binding (ΔG_{bind}). As discussed above, the predicted binding geometry of (*R*)-C562-1101 in comparison to RIME (Figure 9) shows that although the two molecules coincide at several positions, the ES patterns are different (Figure 10, green vs orange lines). Thus, while fashioning new analogs, the initial focus was on placement of functional groups at positions on (*R*)-C562-1101 that could more

directly mimic the *electrostatic interaction* patterns made by the native substrate RIME through the addition of H-bond donors and acceptors (compare molecules in Figure 11 to Figure 9c).

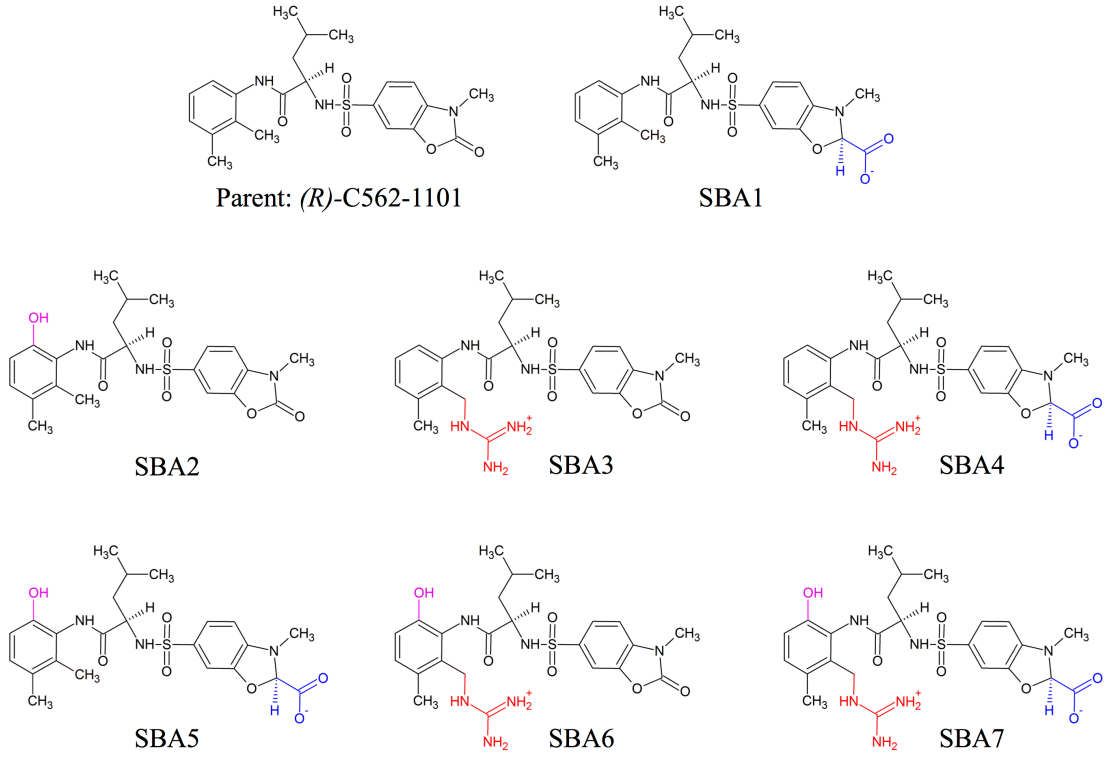


Figure 11. 2D structures for seven analogs derived from parent compound (*R*)-C562-1101.

Specifically (see Figure 11), analogs were designed based with the following hypotheses in mind: (1) SBA1 replaced the benzoxazole carbonyl oxygen with a carboxylic acid (Figure 11 blue) in an attempt to mimic the RIME glutamic acid sidechain that makes contacts with Thr244 and the backbone amide group of Gly352. (2) SBA2 was modified by adding a hydroxyl group (Figure 11 magenta) to the ortho position of the dimethylphenyl ring to mimic the H-bond observed between Glu212 and the free amine on the N-terminus of RIME. (3) SBA3 replaced one of the methyl groups at the ortho position of the dimethylphenyl ring with a guanidinium group (Figure 11 red) to mimic the RIME arginine side chain that makes a salt-bridge with Glu158. (4) SBA4 was modified to have both a carboxylic acid (as in SBA1) and a guanidinium group (as in SBA3) in an attempt to strengthen the ES interactions with Thr244, Gly352, and Glu158. (5) SBA5 combined the carboxylate functionality from

SBA1 with the hydroxyl functionality of SBA2. (6) SBA6 combined the guanidinium functionality of SBA3 with the hydroxyl functionality as SBA2. (7) SBA7 combined all three added functional groups in an effort to recreate all of the major electrostatic interaction peaks made by RIME not originally seen with (*R*)-C562-1101. Table 4 summarizes the result of the MD simulations using four time-averaged metrics: (1) van der Waals footprint similarity score using RIME as reference (FPS_{VDW}), (2) electrostatic footprint similarity score using RIME as reference (FPS_{ES}), (3) ligand RMSD relative to the initial starting structure (RMSD), and (4) estimated free energies of binding (ΔG_{bind}). Also listed are the specific BoNT/E residues for which a given analog is "hypothesized" to interact with favorably in an analogous manner as RIME, using a coloring scheme consistent with Figures 11 and 12.

Table 4

Ensemble-averaged properties for (*R*)-C562-1101 and seven analogs.

Code	Peak Hypotheses[a]	FPS_{VDW}[b]	FPS_{ES}[b]	RMSD (Å)	ΔG_{bind} (kcal/mol)
1101	–	5.16±1.10	9.91±0.33	2.00±0.22	-73.16±6.81
SBA1	Thr246, Gly352	5.34±1.18	10.24±0.32	2.11±0.24	-65.60±5.92
SBA2	Glu212	9.27±1.52	9.28±0.46	3.28±0.41	-71.32±7.65
SBA3	Glu158	5.55±1.27	7.07±0.43	1.83±0.28	-66.10±4.87
SBA4	Glu158, Thr246, Gly352	6.09±1.43	7.65±0.69	2.18±0.19	-85.30±7.09
SBA5	Glu212, Thr246, Gly352	9.46±1.96	10.15±0.51	2.60±0.74	-82.30±7.41
SBA6	Glu158, Glu212	10.35±1.78	7.38±0.42	3.02±0.18	-86.94±8.01
SBA7	Glu158, Glu212, Thr246, Gly352	6.98±1.70	7.22±1.24	2.35±0.37	-84.35±7.63

[a]Residues hypothesized to yield new favorable ES peaks, analogous to those seen with RIME, as a result of functional group modifications. Resides color-coded to match Figures 11 and 12. [b]Average footprint similar (FPS) scores for van der Waals (FPS_{VDW}) and electrostatic (FPS_{ES}) terms. Computed FPS scores neither penalize nor reward similarity when interactions at certain residues are more favorable than that of the reference. [c]Average ligand RMSD. [d]Average estimated free energies of binding (ΔG_{bind}) via the MM-GBSA method.

In most cases, analogs with one added functional group (SBA1, SBA2, SBA3) display an enhanced electrostatic peak at their intended residues as originally hypothesized. For example, the MD-averaged footprint for SBA1 re-creates the favorable peak seen at Thr246 with RIME (Figure 12 bottom, blue arrow). Although, because of the stereo chemistry introduced at the benzoxazole group, the carboxylic acid adopted a position that was not well-coordinated with Gly352 thus that particular peak was not re-created. In terms of geometry, SBA1 adopted a stable binding pose with a ligand RMSD of

2.11 Å (Table 4). For analog SBA2, the added hydroxyl group re-created (Figure 12 bottom, magenta arrow) the H-bond originally observed between RIME and the Glu212 sidechain (Figure 9c). However, the benzoxazole group shifted during the simulation resulting in a larger ligand RMSD of 3.28 Å (Table 4). As a result of this movement, the favorable van der Waals interactions originally observed with Tyr350 were also significantly reduced (Figure 12 top, magenta vs green) which resulted in an unfavorable increase in the FPS_{VDW} score (9.27, Table 4). For SBA3, the expected salt-bridge with Glu158 analogous to that observed with RIME (Figure 9c) was recreated (Figure 12 bottom, red arrow). This ligand was also geometrically stable in the binding site with an average RMSD of 1.83 Å. Overall, five of the seven analogs showed relatively low RMSDs (1.83 - 2.60 Å) with only two compounds (SBA2, SBA6) yielding above 3 Å RMSD (Table 4).

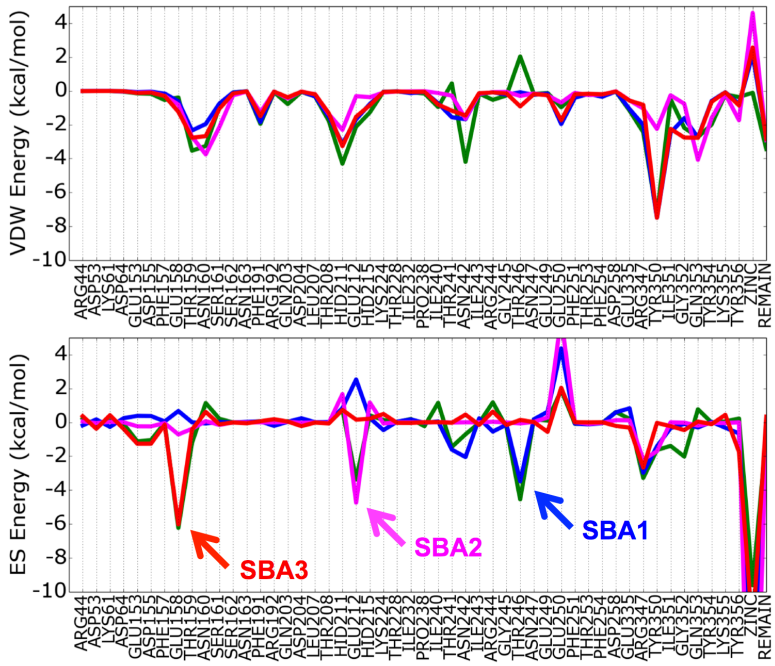


Figure 12. Average van der Waals (VDW, top) and electrostatic (ES, bottom) footprints for analogs SBA1 (blue), SBA2 (magenta), and SBA3 (red) over 20 ns MD simulations compared with the x-ray RIME reference (green). Colored arrows indicate key ES peaks re-created using the same coloring scheme as Table 4 and Figure 11. For clarity, error bars not shown.

Quantitatively (Table 4), four out of the seven analogs (SBA3, SBA4, SBA6, SBA7) yielded a predicted enhanced ES footprint overlap as indicated by smaller FPS_{ES} scores (7.07-7.65) compared to

(*R*)-C562-1101 (9.91) generally affirming the hypotheses that functionality placed at strategic scaffold positions could mimic the ES patterns made by RIME. For the three remaining analogs with roughly equal (SBA2) or slightly worse (SBA1, SBA5) FPS_{ES} scores than the parent, although the target peaks were in most cases re-created (e.g. Figure 12 colored arrows), no improvement in overall ES overlap was obtained due to peak changes occurring at other residues. An examination of the computed free energies of binding (ΔG_{bind}) shows that the addition of one functional group (SBA1, SBA2, and SBA3) does not improve affinity relative to the parent but adding two (SBA4, SBA5, SBA6) or three (SBA7) groups does yield a more favorable ΔG_{bind} energy by about 10 kcal/mol. Interestingly, SBA7 with three added groups is not computed to be more favorable than the analogs with two added groups, probably as a result of higher desolvation penalties.

Overall, SBA4 and SBA7 appeared to be most promising (Figure 13) with improved FPS_{ES} (7.65, 7.22) and ΔG_{bind} (-85.30, -84.25 kcal/mol) scores, respectively, relative to (*R*)-C562-1101 (Table 4). Good FPS_{VDW} overlap (6.09, 6.98) and stable binding geometries (2.18, 2.35 Å RMSD) are also observed over course of the MD simulations. While neither SBA4 or SBA7 mimic all RIME peaks (Figure 13 brown vs green lines), both show enhanced ES interactions at two out of the three targeted sites. Interestingly, SBA7 shows considerable variability at position Glu212 (Figure 13 right panel, error bars), with strong favorable and unfavorable interactions depending on the configuration. Visualization of the trajectory showed that the ligand can form an intermolecular salt bridge with the protein or an intramolecular salt bridge internally. Thus, depending on the conformation, SBA7 can in theory re-create all the dominant ES peaks made by RIME although the effect at position Glu212 is more transient than desired. Nevertheless, these hypothetical and computational investigations demonstrate the potential utility of using time-averaged footprints to aid structure-based design when starting from an experimentally verified hit.

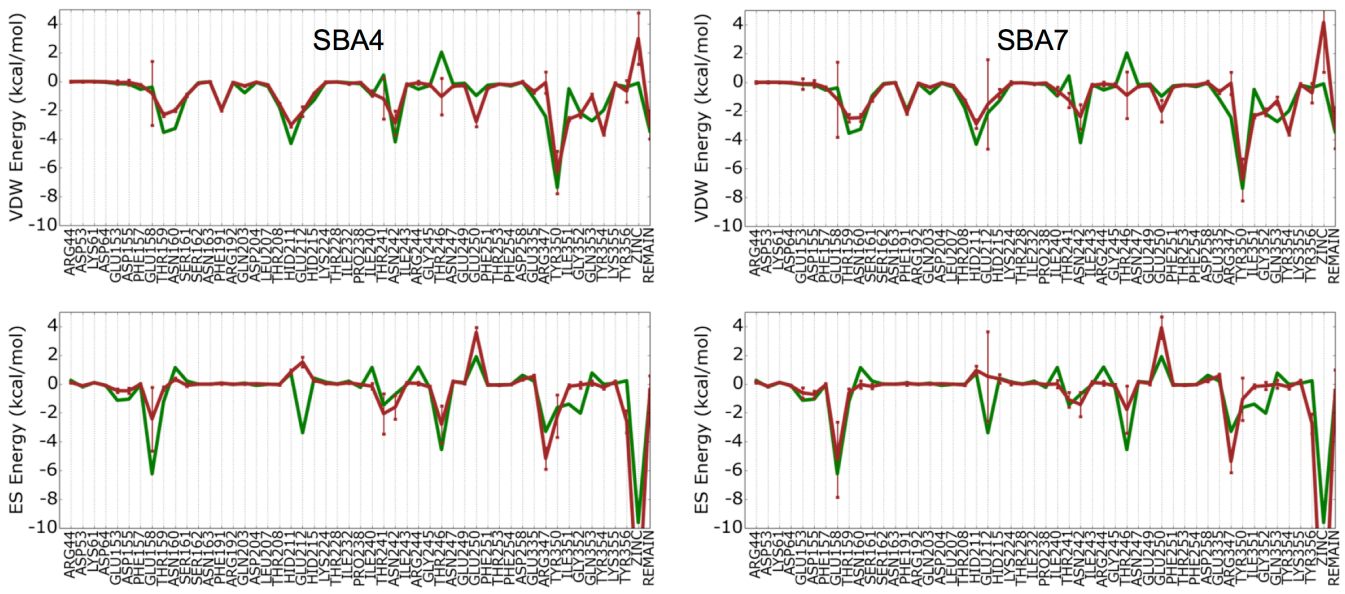


Figure 13. Average van der Waals (VDW, top) and electrostatic (ES, bottom) footprints for analogs SBA4 (left, brown) and SBA7 (right, brown) over 20 ns MD simulations compared with the x-ray RIME reference (green).

CONCLUSION

The goal of this work was to identify new small organic molecules with inhibitory activity against botulinum neurotoxin serotype E (BoNT/E). Using a computer-based screening strategy (DOCK program) ~1.4 M commercially available small molecules were docked to a crystallographic structure of the BoNT/E light chain catalytic domain, initially co-crystalized with the tetrapeptide RIME, with the peptide removed (Figure 2). Following the virtual screen, the 100,000 top scoring molecules (DCE_{VDW+ES} function) were retained and subsequently prioritized by five different DOCK scoring functions (see Methods for definitions) including: (1) DCE_{VDW+ES}, (2) FPS_{VDW+ES}, (3) FPS_{VDW}, (4) FPS_{ES}, and (5) Total Score (Figure 3).

Importantly, compounds prioritized by different functions showed different property distributions (see Figure 4) and thus top ranked members from the different lists showed diversity (see Table 1). A visual examination (see Figure 5) of docked compounds showed: (1) use of the DCE_{VDW+ES} function yielded binding poses that were more spread out among the binding site, (2) use of the FPS_{VDW+ES} function yielded binding poses that more closely resembled the pose adopted by the RIME

reference molecule, and (3) use of the Total Score function yielded binding poses which were in-between. Ultimately, 92 compounds prioritized from the virtual screen and subsequent analysis were purchased for experimental testing. Encouragingly, 9 of the 92 compounds tested (see Table 2) showed some level of activity using a SNAP-Etide assay. Follow-up studies, using a potentially more biologically relevant GFP-based assay, showed that only C562-1101 exhibited activity in both experiments. To further probe C562-1101, the compound was re-synthesized in-house to yield stereochemically well defined *R* and *S* enantiomers which revealed the vendor sample was primarily the *S* enantiomer (Supplementary Information). Experimentally however, it was found that the *R* and *S* enantiomers were roughly equipotent in the GFP-based bioassay (Table 3). Analysis of the two docked geometries showed that *R* enantiomer was more energetically favorable than *S* and had better overlap when compared to the RIME reference thus optimization efforts focused on the *R* form. Notably, enantiomerically pure (*R*)-C562-1101 (SB-BEI-001) showed good dose-response properties (Figure 7) and an estimated IC_{50} value of 14.2 ± 1.7 μM which is an approximately three-fold improvement over NSC-77053 (48.5 ± 6.5 μM) under the same conditions. Accompanying molecular dynamics simulations highlighted that (*R*)-C562-1101 had the best overall properties of the initial 9 hits in terms of good geometric stability and favorable estimated free energy of binding (see Figure 8).

Close structural and energetic examination of the predicted binding pose for (*R*)-C562-1101 suggests the compound makes key van der Waals (VDW) interactions in the BoNT/E binding site, and to a lesser extent, key electrostatic (ES) interactions, that are similar in magnitude to that seen with the native tetrapeptide RIME (see Figure 9, 10). The particularly good overlap in terms of VDW interactions, and key ES interactions with zinc (Figure 6), demonstrates the utility of using footprint similarity scoring in DOCK to help identify compounds when a "known" reference (e.g. RIME) is available. Footprint similarity analysis was also helpful in designing new analogs to further mimic RIME. Through strategic functional group placement, 7 new hypothetical analogs of (*R*)-C562-1101 (see Figure 11) were designed and simulated in an attempt to increase favorable ES interactions at specific residues Glu158, Glu212, Thr246, and Gly352. In general, the results of the hypothetical

analogs satisfied the design purposes given that analogs with one new functional group could computationally recreate single RIME interaction peaks (see Figure 12) and analogs with two or three new functional groups could recreate two or more RIME interaction peaks (see Figure 13) relative to the initial hit (*R*)-C562-1101. Of several properties examined, including total ES footprint similarity overlap, ligand RMSD, and computed free energies of binding, two analogs (SBA4 and SBA7) emerged as most promising (Table 4, Figure 13). Finally, this study has demonstrated the utility of using computer-aided modeling in conjunction with experimental testing for targeted drug discovery. The procedure has resulted in the identification of a new small molecule inhibitor against BoNT/E which provides a promising starting point for design of additional analogs, among which some are also discussed in this work.

Acknowledgements.

We thank Dr. William J. Allen for assistance with virtual screening setups, and Drs. Desigan Kumaran, Subramaniam Eswaramoorthy, Natasha Nesbitt, and Peter Tonge for helpful discussions. This work was funded in part through a grant from the Defense Threat Reduction Agency (DTRA), Department of Defense (BO742081 to S.S.). Research utilized resources at the New York Center for Computational Sciences at Stony Brook University/Brookhaven National Laboratory which is supported by the U.S. Department of Energy under Contract No. DE-AC02-98CH10886 and by the State of New York.

References

(1) Schantz, E. J.; Johnson, E. A. Properties and use of botulinum toxin and other microbial neurotoxins in medicine. *Microbiol Rev* **1992**, *56*, 80-99.

(2) Willis, B.; Eubanks, L. M.; Dickerson, T. J.; Janda, K. D. The strange case of the botulinum neurotoxin: using chemistry and biology to modulate the most deadly poison. *Angew Chem Int Ed Engl* **2008**, *47*, 8360-8379.

(3) Arnon, S. S.; Schechter, R.; Inglesby, T. V.; Henderson, D. A.; Bartlett, J. G.; Ascher, M. S.; Eitzen, E.; Fine, A. D.; Hauer, J.; Layton, M.; Lillibridge, S.; Osterholm, M. T.; O'Toole, T.; Parker, G.; Perl, T. M.; Russell, P. K.; Swerdlow, D. L.; Tonat, K.; Working Group on Civilian, B. Botulinum toxin as a biological weapon: medical and public health management. *JAMA* **2001**, *285*, 1059-1070.

(4) Rossetto, O.; Pirazzini, M.; Montecucco, C. Botulinum neurotoxins: genetic, structural and mechanistic insights. *Nat Rev Microbiol* **2014**, *12*, 535-549.

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

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

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

(8) Agarwal, R.; Swaminathan, S. SNAP-25 substrate peptide (residues 180-183) binds to but bypasses cleavage by catalytically active Clostridium botulinum neurotoxin E. *J Biol Chem* **2008**, *283*, 25944-25951.

(9) Hart, M. K.; Saviolakis, G. A.; Welkos, S. L.; House, R. V. Advanced Development of the rF1V and rBV A/B Vaccines: Progress and Challenges. *Adv Prev Med* **2012**, *2012*, 731604.

(10) Byrne, M. P.; Smith, T. J.; Montgomery, V. A.; Smith, L. A. Purification, potency, and efficacy of the botulinum neurotoxin type A binding domain from Pichia pastoris as a recombinant vaccine candidate. *Infect Immun* **1998**, *66*, 4817-4822.

(11) Ravichandran, E.; Al-Saleem, F. H.; Ancharski, D. M.; Elias, M. D.; Singh, A. K.; Shamim, M.; Gong, Y.; Simpson, L. L. Trivalent vaccine against botulinum toxin serotypes A, B, and E that can be administered by the mucosal route. *Infect Immun* **2007**, *75*, 3043-3054.

(12) Zichel, R.; Mimran, A.; Keren, A.; Barnea, A.; Steinberger-Levy, I.; Marcus, D.; Turgeman, A.; Reuveny, S. Efficacy of a potential trivalent vaccine based on Hc fragments of botulinum toxins A, B, and E produced in a cell-free expression system. *Clin Vaccine Immunol* **2010**, *17*, 784-792.

(13) Kumar, G.; Swaminathan, S. Recent developments with metalloprotease inhibitor class of drug candidates for botulinum neurotoxins. *Curr Top Med Chem* **2015**, *15*, 685-695.

(14) Schmidt, J. J.; Stafford, R. G.; Bostian, K. A. Type A botulinum neurotoxin proteolytic activity: development of competitive inhibitors and implications for substrate specificity at the S1' binding subsite. *FEBS Lett* **1998**, *435*, 61-64.

(15) Schmidt, J. J.; Stafford, R. G. A high-affinity competitive inhibitor of type A botulinum neurotoxin protease activity. *FEBS Lett* **2002**, *532*, 423-426.

(16) Kumaran, D.; Rawat, R.; Ludivico, M. L.; Ahmed, S. A.; Swaminathan, S. Structure- and substrate-based inhibitor design for Clostridium botulinum neurotoxin serotype A. *J Biol Chem* **2008**, *283*, 18883-18891.

(17) Kumar, G.; Kumaran, D.; Ahmed, S. A.; Swaminathan, S. Peptide inhibitors of botulinum neurotoxin serotype A: design, inhibition, cocrystal structures, structure-activity relationship and pharmacophore modeling. *Acta Crystallogr D Biol Crystallogr* **2012**, *68*, 511-520.

(18) Silhar, P.; Capkova, K.; Salzameda, N. T.; Barbieri, J. T.; Hixon, M. S.; Janda, K. D. Botulinum neurotoxin A protease: discovery of natural product exosite inhibitors. *J Am Chem Soc* **2010**, *132*, 2868-2869.

(19) Eubanks, L. M.; Silhar, P.; Salzameda, N. T.; Zakhari, J. S.; Xiaochuan, F.; Barbieri, J. T.; Shoemaker, C. B.; Hixon, M. S.; Janda, K. D. Identification of a Natural Product Antagonist against the Botulinum Neurotoxin Light Chain Protease. *ACS Med Chem Lett* **2010**, *1*, 268-272.

(20) Boldt, G. E.; Kennedy, J. P.; Janda, K. D. Identification of a potent botulinum neurotoxin a protease inhibitor using in situ lead identification chemistry. *Org Lett* **2006**, *8*, 1729-1732.

(21) Silvaggi, N. R.; Boldt, G. E.; Hixon, M. S.; Kennedy, J. P.; Tzipori, S.; Janda, K. D.; Allen, K. N. Structures of Clostridium botulinum Neurotoxin Serotype A Light Chain complexed with small-molecule inhibitors highlight active-site flexibility. *Chem Biol* **2007**, *14*, 533-542.

(22) Thompson, A. A.; Jiao, G. S.; Kim, S.; Thai, A.; Cregar-Hernandez, L.; Margosiak, S. A.; Johnson, A. T.; Han, G. W.; O'Malley, S.; Stevens, R. C. Structural characterization of three novel hydroxamate-based zinc chelating inhibitors of the Clostridium botulinum serotype A neurotoxin light chain metalloprotease reveals a compact binding site resulting from 60/70 loop flexibility. *Biochemistry* **2011**, *50*, 4019-4028.

(23) Burnett, J. C.; Opsenica, D.; Sriraghavan, K.; Panchal, R. G.; Ruthel, G.; Hermone, A. R.; Nguyen, T. L.; Kenny, T. A.; Lane, D. J.; McGrath, C. F.; Schmidt, J. J.; Vennerstrom, J. L.; Gussio, R.; Solaja, B. A.; Bavari, S. A refined pharmacophore identifies potent 4-amino-7-chloroquinoline-based inhibitors of the botulinum neurotoxin serotype A metalloprotease. *Journal of medicinal chemistry* **2007**, *50*, 2127-2136.

(24) Videnovic, M.; Opsenica, D. M.; Burnett, J. C.; Gomba, L.; Nuss, J. E.; Selakovic, Z.; Konstantinovic, J.; Krstic, M.; Segan, S.; Zlatovic, M.; Sciotti, R. J.; Bavari, S.; Solaja, B. A. Second generation steroidal 4-aminoquinolines are potent, dual-target inhibitors of the botulinum neurotoxin serotype A metalloprotease and P. falciparum malaria. *Journal of medicinal chemistry* **2014**, *57*, 4134-4153.

(25) Caglic, D.; Krutein, M. C.; Bompiani, K. M.; Barlow, D. J.; Benoni, G.; Pelletier, J. C.; Reitz, A. B.; Lairson, L. L.; Houseknecht, K. L.; Smith, G. R.; Dickerson, T. J. Identification of clinically viable

quinolinol inhibitors of botulinum neurotoxin A light chain. *Journal of medicinal chemistry* **2014**, *57*, 669-676.

(26) Kumar, G.; Agarwal, R.; Swaminathan, S. Discovery of a fluorene class of compounds as inhibitors of botulinum neurotoxin serotype E by virtual screening. *Chemical communications* **2012**, *48*, 2412-2414.

(27) Kumar, G.; Agarwal, R.; Swaminathan, S. Small molecule non-peptide inhibitors of botulinum neurotoxin serotype E: Structure-activity relationship and a pharmacophore model. *Bioorg Med Chem* **2016**, *in press*.

(28) Chen, S.; Barbieri, J. T. Unique substrate recognition by botulinum neurotoxins serotypes A and E. *J Biol Chem* **2006**, *281*, 10906-10911.

(29) Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. ZINC: a free tool to discover chemistry for biology. *J Chem Inf Model* **2012**, *52*, 1757-1768.

(30) Brozell, S. R.; Mukherjee, S.; Balius, T. E.; Roe, D. R.; Case, D. A.; Rizzo, R. C. Evaluation of DOCK 6 as a pose generation and database enrichment tool. *J Comput Aided Mol Des* **2012**, *26*, 749-773.

(31) Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C. DOCK 6: Impact of new features and current docking performance. *J Comput Chem* **2015**, *36*, 1132-1156.

(32) Balius, T. E.; Mukherjee, S.; Rizzo, R. C. Implementation and evaluation of a docking-rescoring method using molecular footprint comparisons. *J Comput Chem* **2011**, *32*, 2273-2289

(33) Balius, T. E.; Allen, W. J.; Mukherjee, S.; Rizzo, R. C. Grid-based molecular footprint comparison method for docking and de novo design: application to HIVgp41. *J Comput Chem* **2013**, *34*, 1226-1240.

(34) Mukherjee, S.; Balius, T. E.; Rizzo, R. C. Docking validation resources: protein family and ligand flexibility experiments. *J Chem Inf Model* **2010**, *50*, 1986-2000.

(35) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E., 3rd; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossvary, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. AMBER 14. *University of California, San Francisco* **2014**.

(36) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins-Structure Function and Bioinformatics* **2006**, *65*, 712-725.

(37) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J Comput Chem* **2004**, *25*, 1157-1174.

(38) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic Charges. AM1-BCC model: I. Method. *J Comput Chem* **2000**, *21*, 132-146.

(39) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem* **2002**, *23*, 1623-1641.

(40) DesJarlais, R. L.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. *Journal of medicinal chemistry* **1988**, *31*, 722-729.

(41) Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated Docking with Grid-Based Energy Evaluation. *J Comput Chem* **1992**, *13*, 505-524.

(42) Holden, P. M.; Kaur, H.; Goyal, R.; Gochin, M.; Rizzo, R. C. Footprint-based identification of viral entry inhibitors targeting HIVgp41. *Bioorg Med Chem Lett* **2012**, *22*, 3011-3016.

(43) Holden, P. M.; Allen, W. J.; Gochin, M.; Rizzo, R. C. Strategies for lead discovery: Application of footprint similarity targeting HIVgp41. *Bioorg Med Chem* **2014**, *22*, 651-661.

(44) Allen, W. J.; Yi, H. A.; Gochin, M.; Jacobs, A.; Rizzo, R. C. Small molecule inhibitors of HIVgp41 N-heptad repeat trimer formation. *Bioorg Med Chem Lett* **2015**, *25*, 2853-2859.

(45) Berger, W. T.; Ralph, B. P.; Kaczocha, M.; Sun, J.; Balius, T. E.; Rizzo, R. C.; Haj-Dahmane, S.; Ojima, I.; Deutsch, D. G. Targeting fatty acid binding protein (FABP) anandamide transporters - a novel strategy for development of anti-inflammatory and anti-nociceptive drugs. *PLoS One* **2012**, *7*, e50968.

(46) Kaczocha, M.; Rebecchi, M. J.; Ralph, B. P.; Teng, Y. H.; Berger, W. T.; Galbavy, W.; Elmes, M. W.; Glaser, S. T.; Wang, L.; Rizzo, R. C.; Deutsch, D. G.; Ojima, I. Inhibition of fatty acid binding proteins elevates brain anandamide levels and produces analgesia. *PLoS One* **2014**, *9*, e94200.

(47) Teng, Y. H.; Berger, W. T.; Nesbitt, N. B.; Kumar, K.; Balius, T. E.; Rizzo, R. C.; Tonge, P. T.; Ojima, I.; Swaminathan, S. Computer-Aided Identification, Synthesis, and Biological Evaluation of Novel Inhibitors for Botulinum Neurotoxin Serotype A. *Bioorg Med Chem* **2015**, *23*, 5489–5495

(48) Brown, R. D.; Martin, Y. C. Use of Structure-Activity Data to Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection. *J Chem Inf Comput Sci* **1996**, *36*, 572-584.

(49) MOE; Chemical Computing Group: Montreal, Quebec, Canada, 2012.

(50) Kumar, R.; Kukreja, R. V.; Cai, S.; Singh, B. R. Differential role of molten globule and protein folding in distinguishing unique features of botulinum neurotoxin. *Biochimica et biophysica acta* **2014**, *1844*, 1145-1152.

(51) Primak, Y.; Christian, T.; Shine, N. SNAP Etide™, a FRET Substrate for Botulinum Toxin Type E (poster). *43rd Interagency Botulism Research Coordinating Committee (IBRCC) Meeting* **2006**.

(52) Hines, H. B.; Kim, A. D.; Stafford, R. G.; Badie, S. S.; Brueggeman, E. E.; Newman, D. J.; Schmidt, J. J. Use of a recombinant fluorescent substrate with cleavage sites for all botulinum neurotoxins in high-throughput screening of natural product extracts for inhibitors of serotypes A, B, and E. *Appl Environ Microbiol* **2008**, *74*, 653-659.

(53) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **1983**, *79*, 926-935.

(54) Huang, Y. L.; Rizzo, R. C. A Water-Based Mechanism of Specificity and Resistance for Lapatinib with ErbB Family Kinases. *Biochemistry* **2012**, *51*, 2390-2406.

(55) Balius, T. E.; Rizzo, R. C. Quantitative Prediction of Fold Resistance for Inhibitors of EGFR. *Biochemistry* **2009**, *48*, 8435-8448.

(56) Roe, D. R.; Cheatham, T. E., 3rd PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* **2013**, *9*, 3084-3095.

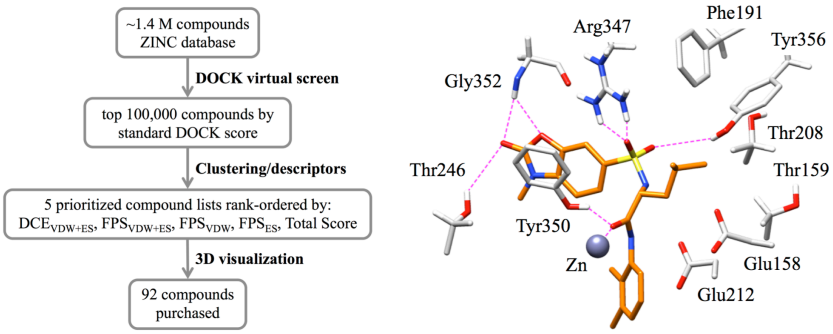
(57) Srinivasan, J.; Cheatham, T. E., 3rd; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401-9409.

(58) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., 3rd Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889-897.

(59) McGillick, B. E.; Balius, T. E.; Mukherjee, S.; Rizzo, R. C. Origins of Resistance to the HIVgp41 Viral Entry Inhibitor T20. *Biochemistry* **2010**, *49*, 3575-3592.

(60) Jiang, L.; Rizzo, R. C. Pharmacophore-based similarity scoring for DOCK. *J Phys Chem B* **2015**, *119*, 1083-1102.

Graphical Abstract

Identification of Small Molecule Inhibitors of Botulinum Neurotoxin Serotype E via Footprint Similarity

Yuchen Zhou, Brian E. McGillick, Yu-Han Gary Teng, Krupanandan Haranahalli, Iwao Ojima, Subramanyam Swaminathan, and Robert C. Rizzo