BOTDB: A DATABASE FOR THE CLOSTRIDIAL NEUROTOXINS

Frank J. Lebeda
United States Army Medical Research Institute of Infectious Diseases
Division of Toxinology and Aerobiology; Department of Cell Biology and Biochemistry,
Fort Detrick, Maryland USA 21702-5011

ABSTRACT

BotDB is designed to encapsulate the rapidly expanding amount of information about the structure and function of the botulinum (BoNT) and tetanus (TeNT) neurotoxins and to track a variety of basic and applied research efforts. The AceDB management system was chosen for this project because of its flexibility in manipulating semi-structured data sets and for its information retrieval query languages. Besides storing amino and nucleic acid sequences of the clostridial neurotoxin genes and proteins, BotDB provides sequence data for new classes of objects including neurotoxin mutants, substrates and their mutants, associated non-toxic proteins, and C-fragment vaccine candidates. New data types provide information on detection assays for the neurotoxins, and on structural data from X-ray crystallographic and circular dichroism spectroscopic studies. Kinetic parameters from biochemical experiments include reaction rates for substrate cleavage, and block of neurotransmission. The structures and kinetic characteristics of presently known chemical inhibitors are also being archived. All of these data are associated with citations of the relevant literature for on-line annotation. Graphics viewer programs are provided to display stored images and three-dimensional representations of protein structures. BotDB is in the alpha-test phase of development and will become a publicly available web site.

INTRODUCTION

The neurotoxins from *Clostridium botulinum* and related species represent some of the most lethal substances known and are the subject of general and specialized reviews[1-6]. Only those structural and functional features of these proteins that pertain to the description of BotDB will be mentioned here.

The seven immunologically distinguishable serotypes (BoNT/A-G) cause flaccid paralysis in humans and experimental animals by preventing acetylcholine-containing vesicles from releasing their contents in a calcium-mediated response to chemical or electrical stimuli at peripheral cholinergic presynaptic endings. The structurally related tetanus neurotoxin (TeNT) from *C. tetani*, in contrast, causes spastic paralysis due to its net disinhibitory effect within the central nervous system.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01 JUL 2003	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
BOTDB: A Database For The Clostridial Neurotoxins	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **United States Army Medical Research Institute of Infectious Diseases Division of Toxinology and Aerobiology; Department of Cell Biology and Biochemistry, Fort Detrick, Maryland USA 21702-5011**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM001523., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**11**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Within this family, each neurotoxin is expressed as a single polypeptide chain having a molecular mass (M_r) of ~150 kDa. For the toxicity to occur, the polypeptide must be enzymatically 'nicked' to form a heterodimer of the light (L) chain (M_r ~50 kDa) and of the heavy (H) chain (M_r ~100 kDa) that are covalently linked by a disulfide bridge. The reduction of this inter-chain bridge is also required for toxicity. The three-dimensional (3-D) crystal structures of some these neurotoxins and their fragments have been recently solved including the BoNT/A[7] and the BoNT/B[8] holotoxins.

The L-chains are zinc-dependent proteases that cleave soluble N-ethylmaleimide sensitive factor (NSF) attachment protein receptors (SNAREs). BoNT/A, /C1,and /E cleave the synaptosome-associated protein of 25 kDa (SNAP-25). BoNT/C1 is the only serotype known to cleave both syntaxin-1a and SNAP-25[9,10]. The remaining serotypes and TeNT cleave synaptobrevin (or vesicle-associated membrane protein, VAMP[4]. Isoforms of synaptosomal-associated proteins that occur in non-neural tissue (e.g., cellulobrevin and SNAP-23) are also susceptible to cleavage if exposed to the appropriate neurotoxin serotype.

The H-chain is associated with a high-affinity binding of its C-terminal domain to as yet unidentified ectoacceptors on cholinergic nerve terminals[11]. After receptor-mediated endocytosis, the translocation of the toxic moiety into the neuroplasm is caused by the N-terminal domain of the H-chain. Cation-selective ion channels are formed by from these domains that are believed to allow the escape of the toxic moiety from low-pH, endocytotic compartments into the neuroplasm.

METHODS

SOURCES OF DATA

A critical amount of functional and structural information about the BoNTs and TeNT exits. Presently, 23 precursor sequences from various strains of the seven BoNT serotypes and TeNT have been deposited in public databases (SwissProt[12], GenBank[13]). At this time, four sets of crystal structure coordinates for the holotoxins, three sets for the L- chain and eight sets for H- chain C-fragments (with and without ligands) are in the Protein DataBank[14] (PDB). At least three complete neurotoxin progenitor genes are available from GenBank[13] which contain four to six non-toxic genes for proteins that are associated with each of the BoNTs[5]. The number of protein substrates and cleavable peptides that have been examined in the literature is rapidly increasing[15,16] and the list of neurotoxin active-site inhibitors is growing[17,18].

DESCRIPTION OF BOTDB FEATURES

BotDB was patterned after aCHEdb, a specialized archive of protein structural data for the α/β hydrolase fold family that are structurally similar in 3-D space, yet have a variety of sequences and a diverse set of catalytic and non-catalytic functions[19,20]. Both of these databases are controlled by AceDB, an object-oriented-like database management system originally written by Richard Durbin and Jean Thierry-Mieg for the genomic data of the nematode *C. elegans*,[23]. In contrast to relational database systems where data are stored in tables[24], AceDB has classes of objects that store the data. Objects are defined by models that, taken together, comprise the schema for the database.

A key advantage that AceDB offers over relational tables is that it is more flexible. Specifically, AceDB allows incomplete sets of diverse types of data, a feature that saves disk space and does not degrade the speed or efficiency of the program. Also, the models of AceDB can be readily changed or expanded without rebuilding the database, a characteristic that is especially important in the early, dynamic stages of development when the data structures are not yet fully defined. Moreover, AceDB is open-source software in contrast to commercial-grade relational systems that can be prohibitively expensive.

BotDB has thus far been tested with the Microsoft WINDOWS 95, 98, NT and 2000 operating systems using 'ACEDB for Windows'[25] and the aCHEdb schema file as an initial template for the classes and objects. The main window of BotDB is illustrated in Figure 1 and shows a partial list of the 25 data classes that are presently available. Medline citations available through PubMed searches (www.ncbi.nlm.nih.gov/entrez/query.fcgi) are linked to each data class and for each amino or nucleic acid sequence.

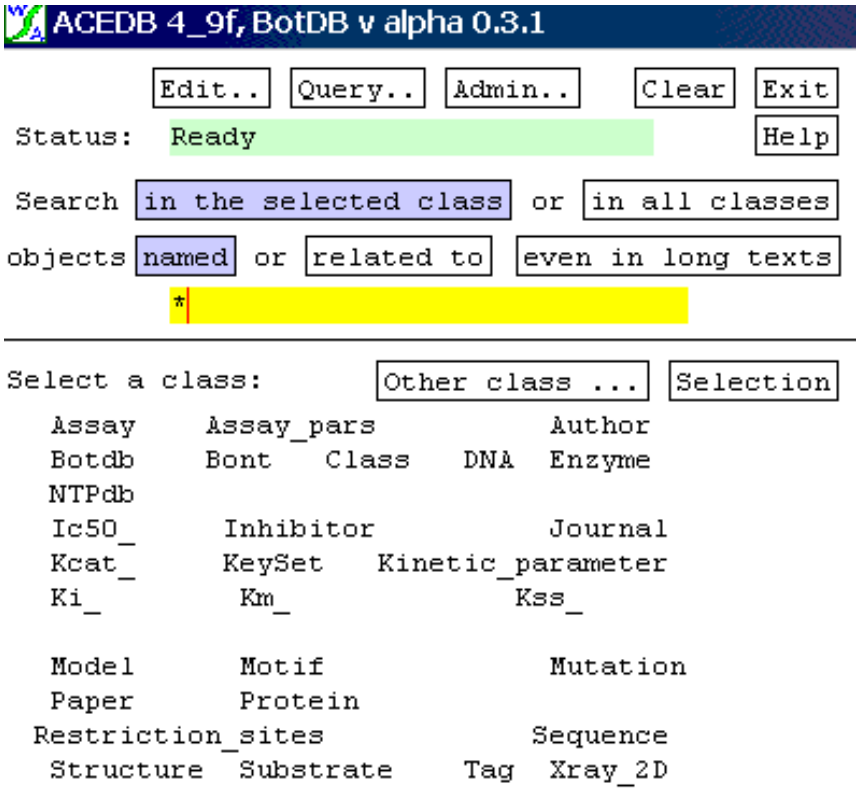

Figure 1. The main widow of BotDB. Top: boxes can be clicked on for database searching and general administrative operations. Bottom: partial list of Classes that the user can choose for more detailed information.

TYPES OF DATA WITHIN BOTDB

Comprising one portion of this database are the nucleic and amino acid sequences of the neurotoxins and their corresponding SNARE substrates. In Figure 2, the location of a hydrophobic, potential channel-forming region[2,3,26-28], either in absolute

terms of the highlighted amino acid sequence display, or by its relative position along the vertical bar, represents just one protein feature that can be viewed in this type of graphic display.

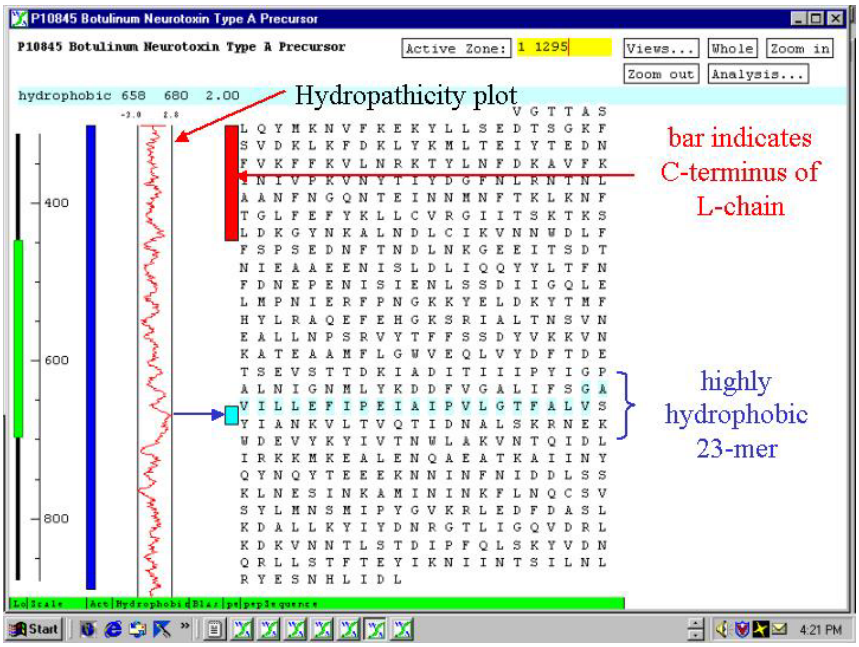

Figure 2. Amino acid sequence display.

This search result shows a portion of the 1295 residue sequence of BoNT/A with a vertically displayed hydropathicity plot and a 23-mer region corresponding to a channel-forming peptide.

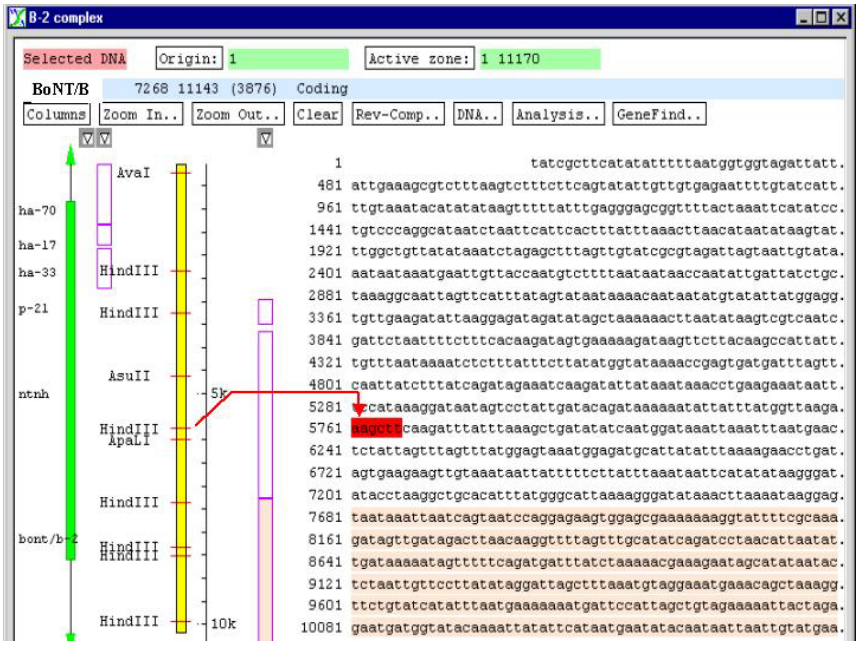

Figure 3. Nucleic acid sequence display.

The sequences are obtained from publicly accessible databases: NCBI's Genbank and ExPASy's SwissProt. Gene and protein fragments and the C-fragment vaccine candidates[30,31], are included along with sequences of the nontoxic proteins that form large complexes with the BoNTs. A limited set of restriction enzymes that target sites on DNA are also featured (Figure 3).

This output shows the partial sequence of the progenitor gene of BoNT/B with the toxin and the five non-toxin genes (far left). The short horizontal lines on the vertical yellow bar represent cleavage sites for the indicated restriction enzymes (Ava I, Hind III, AsuII, and ApaLI). The vertical boxes on either side of the yellow bar indicated that the genes for 'bont/b-2',''ntnh', and 'p-21' are coded on the opposite strand with respect to the other three genes. The highlighted portion of the DNA sequence at the bottom represents a portion of the 'bont/b-2' gene. The 'ntnh' sequence "aagctt" highlighted in red corresponds to the motif recognized by the HindIII restriction enzyme.

As in the aCHEdb, inhibitors of enzymatic activity of the neurotoxins (e.g., Figure 4) will also constitute a continually expanding segment of BotDB. Data from coordinate files for 3-D structures will also be included for display (Figure 5).

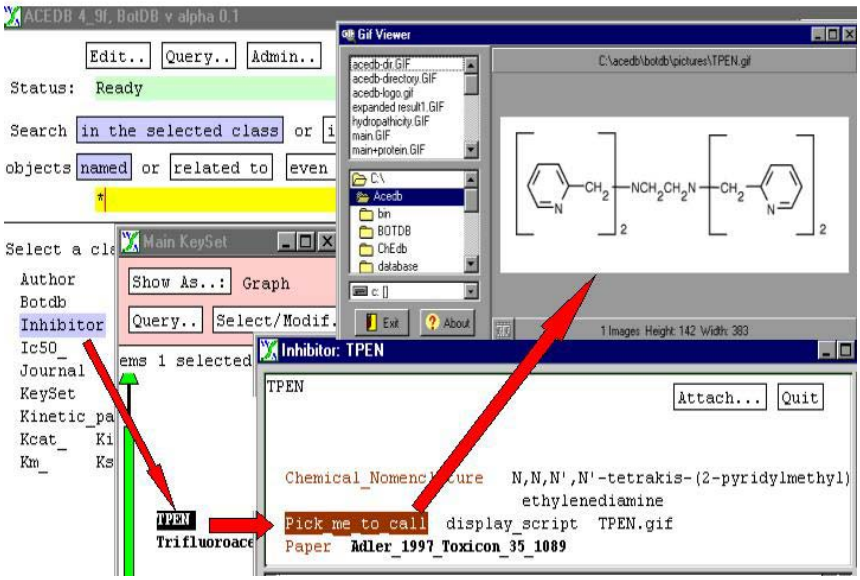

Figure 4. Candidate inhibitor of BoNT activity.

The arrows between the overlapping windows indicate the user's order of selection. The cascade of windows starts in the back. The desired class, INHIBITOR in this case, is highlighted in the main BotDB window, followed by TPEN in the MAIN KEYLIST window. Selecting the 'pick_me_to_call' label opens the window containing the chemical representation of TPEN within the GifViewer.

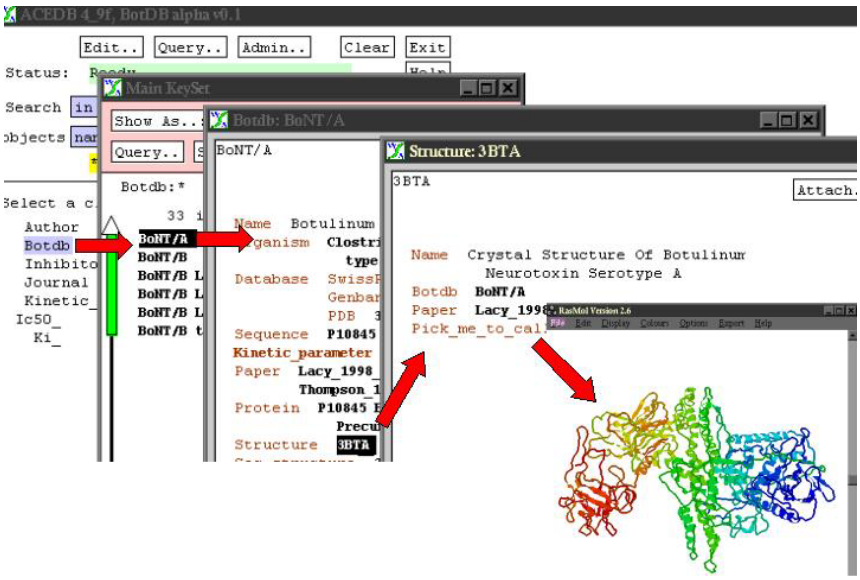

Figure 5. 3-D representation of BoNT/A.

In this example, the cascade of windows starts in the back while the red arrows indicate the item selected by the user. Following the red arrows within this cascade illustrates how the user can highlight the class Botdb in the main BotDB window, and highlight BoNT/A in the MAIN KEYSET window to obtain the Structure "3BTA". Selecting the 'pick_me_to_call' label brings up the final window that has the molecular representation of the PDB file 3BTA as visualized using RASMOL.

Unlike aCHEdb, lethality and detection assay data for the BoNTs comprise a new class of data objects for BotDB (Figure 6). Each bioassay is characterized by its serospecificity, sensitivity, the measures used (e.g., optical density; agglutination), matrices (e.g., buffers), animal species, and a key reference. An example output from this class of data is illustrated in the description of the query features below.

Beyond merely archiving PDB coordinates of atomic positions from X-ray crystallographic data, BotDB includes a new class of data objects for protein secondary structures. BotDB can store secondary structural information in several ways. BotDB can store images that display the locations of secondary structural elements (helices and strands) with respect to a given amino acid sequence. These "maps" are created from the 3-D structural data at the PDBsum website from where they can be downloaded[31]. Quantitative measurements are also included for secondary structure content (% helix, strand, coil) in aqueous solution of various peptides and proteins from circular dichroic studies that are available from the literature[32,33]. These results may be compared to those from the available crystallographic data. Finally, graphic displays of secondary structure predictions from outputs of artificial neural networks can be included[34,35].

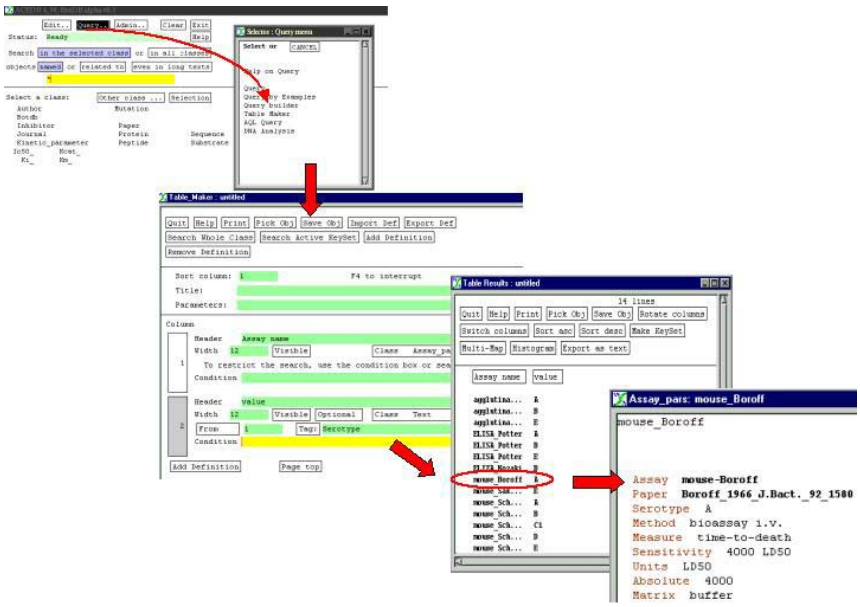

Figure 6. Querying the database using the Table Maker utility.

The user initially selects the 'Query' button in the main window to see the query menu. After selecting Table Maker, the user fills in the appropriate blanks in the form to make a list of BoNT assays. Selecting the 'Search Whole Class' button generates the search result. In this output there are fourteen BotDB entries and five types of assays for BoNT serotypes A-E. The whole-animal bioassay selected is given with the route of neurotoxin administration used and the observed lower-limit of sensitivity in mouse LD_{50} units.

A portion of BotDB will also be devoted to the kinetics of neurotoxicity, e.g., the onset of paralysis and persistence of symptoms. The kinetic parameters that are presently included are those calculated from enzymatic assays[18] (e.g., K_M, k_{cat}), the effect of inhibitors[36] (e.g., K_i), their binding affinity to ectoacceptors[11] (K_d), and the time course of motor paralysis[37] (e.g., time-to-50% block).

DATA VISUALIZATION

Two visualization programs are included as auxiliary software that accompanies BotDB. GifViewer is a freeware utility to view static GIF-formatted files on WINDOWS systems (DevelCor, www.develcor.com). An example is shown in Figure 4 of a two dimensional chemical representation of a small organic chelator molecule (TPEN)[38]. The other program, RasMol (www.umass.edu/microbio/rasmol/index2.htm), is a freeware package to visualize 3-D molecular representations using stored PDB-formatted files (Figure 5). This utility enables the user to move and rotate images of proteins and other molecules in a variety of display formats at the atomic level.

SEARCH ENGINES

AceDB contains two search engines, the original AceDB query language and the newer AQL, both of which allow sophisticated queries to be made. Results from the queries are displayed as relational tables that can be exported as text files to other programs, e.g., spreadsheets, for further computational analyses. An example is shown in Figure 6 in which the Table Maker utility builds a storable search query for a BoNT assay. An expanded version of this example is included in one of the tutorials that accompany this database package.

CONCLUSIONS

This database is designed for use in a variety of tasks. One of these will be to serve as a local repository of parameters derived from enzymatic analyses, inhibitor screenings, and toxicokinetic studies. Parameters will also include those from studies focused on the formation of toxin-induced ion channels and on SNARE complex formation[39].

Other features of this database could be used to address a number of questions that remain in this research area. For example, the difficulty in separating the L- and H-chains emphasizes the importance to understand these, as yet uncharacterized, non-covalent interactions, so that the structural identity of the toxic moiety can be explained. A related open question is how the cationic channels formed by the translocation domain[40] are associated with the escape of the toxic fragment into the neuroplasm[3]. Areas of further content development will include ganglioside structures and their role in neurotoxin binding at cholinergic nerve terminals[41]. Despite our long-held knowledge of the multi-step intoxication process[1], it is evident from the above remarks that much work remains to clarify further the functional roles played by the molecular machinery of these neurotoxins.

Future database improvements will include the port of BotDB to a UNIX-based server so that a web accessible version of this database can be made publicly available. Similar uses of auxiliary 3-D viewers exist at the SCORPION website[42] and the RDB receptor database[43]. Future releases of BotDB will provide additional choices for these viewers as they become available. It is also envisioned that specialized structure-function databases such as BotDB will be integrated into networks of databases, despite their diverse schema and data formats, by converting their data files into a universally compatible format such as XML or its successor[24].

ACKNOWLEDGEMENTS

I thank the developers of AceDB and aCHEdb and to Drs. Richard Gordon, Harry Hines, James Marks, Charles Millard, Mark A. Olson, George Oyler, Mark Poli, James Schmidt, Bal Ram Singh, and Raymond Stevens for their many useful suggestions and continuous support. The work described herein was funded in part by USAMRIID Research Plan ET-00-04.

DISCLAIMER

The mention of a company or a commercial product does not constitute an endorsement. The views expressed in this article are those of the author and are not necessarily endorsed by Department of Defense or the Department of the Army.

REFERENCES

1. Simpson LL. The origin, structure, and pharmacological activity of botulinum toxin. Pharmacol Rev 1981;33:155-188.
2. Singh BR, Li B, Read D. Botulinum versus tetanus neurotoxins: why is botulinum neurotoxin but not tetanus neurotoxin a food poison? Toxicon 1995;33:1541-1547.
3. Lebeda FJ, Singh BR. Membrane channel activity and translocation of tetanus and botulinum neurotoxins. J Toxicol-Toxinol Rev 1999;18:45-76.
4. Schiavo G, Matteoli M, Montecucco C. Neurotoxins affecting neuroexocytosis. Physiol Rev 2000;80:717-766.
5. Johnson EA, Bradshaw M. *Clostridium botulinum* and its neurotoxins: a metabolic and cellular perspective. Toxicon 2001;39:1703-1722.
6. Rossetto O, Seveso M, Caccin P, Schiavo G, Montecucco C. Tetanus and botulinum neurotoxins: turning bad guys into good by research. Toxicon 2001;39:27-41.
7. Lacy DB, Tepp W, Cohen AC, DasGupta BR, Stevens RC. Crystal structure of botulinum neurotoxin type A and implications for toxicity. Nature Struct Biol 1998; 5:898-902.
8. Swaminathan S, Eswaramoorthy S. Structural analysis of the catalytic and binding sites of Clostridium botulinum neurotoxin B. Nature Struct Biol 2000;7:693-699.
9. Williamson LC, Halpern JL, Montecucco C, Brown JE, Neale EA. Clostridial neurotoxins and substrate proteolysis in intact neurons: botulinum neurotoxin C acts on synaptosomal-associated protein of 25 kDa. J Biol Chem 1996;271:7694-7699.
10. Foran P, Lawrence GW, Shone CC, Foster KA, Dolly JO. Botulinum neurotoxin C1 cleaves both syntaxin and SNAP-25 in intact and permeabilized chromaffin cells: correlation with its blockade of catecholamine release. Biochemistry 1996.;35:2630-2636.
11. Daniels-Holgate PU, Dolly JO. Productive and non-productive binding of botulinum neurotoxin A to motor nerve endings are distinguished by its heavy chain. Journal Neuroscience Research 1996;44:263-271.
12. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res 2000;28:45-48.
13. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA, Wheeler DL. GenBank. Nucleic Acids Res 2000;28:15-18.
14. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, et al. The Protein Data Bank. Nucleic Acids Res 2000;28:235-342.
15. Shone CC, Quinn CP, Wait R, Hallis B, Fooks SG, Hambleton P. Proteolytic cleavage of synthetic fragments of vesicle-associated membrane protein, isoform-2 by botulinum type B neurotoxin. Eur J Biochem 1993;217:965-971.
16. Schmidt JJ, Bostian KA. Endoproteinase activity of type A botulinum neurotoxin: substrate requirements and activation by serum albumen. J Prot Chem 1997;16:19-26.
17. Martin L, Cornille F, Turcaud S, Meudal H, Roques BP, Fournie-Zaluski MC. Metallopeptidase inhibitors of tetanus toxin: A combinatorial approach. J Med Chem 1999; 42:515-525.

18. Schmidt JJ, R.G. Stafford RG, C.B. Millard CB. High-throughput assays for botulinum neurotoxin proteolytic activity: serotypes a, b, d, and f. Anal Biochem 2001;296:130-137.
19. Chatonnet A, Cousin XA, Robinson A. Links between kinetic data and sequences in the alpha/beta-hydrolases fold database. Brief Bioinform 2001;2:30-37.
20. Cousin X, Hotelier T, Giles K, Toutant JP, Chatonnet A. aCHEdb: the database system for ESTHER, the alpha/beta fold family of proteins and the Cholinesterase gene server. Nucleic Acids Res 1998;26:226-228.
21. Martinelli SD, Brown CG, Durbin R. Gene expression and development databases for C. elegans. Seminars Cell Devel Biol 1997;:459-467.
22. Walsh S, Anderson M, Cartinhour SW. ACEDB: a database for genome information. Meth Biochem Anal 1998; 39:299-318.
23. Stein L, Sternberg P, Durbin R, Thierry-Mieg J, Spieth J. WormBase: network access to the genome and biology of *Caenorhabditis elegans*. Nucleic Acids Res 2001;29:82-86.
24. Lebeda FJ, Wolinsky M, Lefkowitz EJ. Information resources and database development for defense against biological weapons. In: Lindler L, Lebeda FJ, Korch G, editors. *Biological weapons defense; principles and mechanisms of infectious disease*. New York: Humana Press, in press.
25. A Caenorhabditis elegans database (ACEDB) for Microsoft Windows '95 ('98) or NT 4.0 (workstation or server). Version 4.9 (http://www.sanger.ac.uk/Users/rbsk/winace.html); 1999.
26. Montal MS, Tomich JM, Blewitt R, Montal M. Identification of an ion channel-forming motif in the primary structure of tetanus and botulinum neurotoxins. Febs Lett 1992;313:12-18.
27. Lebeda FJ, Olson MA. Secondary structural predictions for the clostridial neurotoxins. Proteins: Struct., Func. & Gen. 1994;20:293-300.
28. Lebeda FJ, Olson MA. Structural predictions of the channel-forming region of botulinum neurotoxin heavy chain. Toxicon 1995;35:559-567.
29. Clayton MA, Clayton JM, Brown DR, Middlebrook JL. Protective vaccination with a recombinant fragment of Clostridium botulinum neurotoxin serotype A expressed from a synthetic gene in Escherichia coli. Infect Immun 1995;63:2738-2742.
30. Smith LA. Development of recombinant vaccines for botulinum neurotoxin. Toxicon 1998;36:1539-1548.
31. Laskowski RA. PDBsum: summaries and analyses of PDB structures. Nucleic Acids Res 2001;29:221-222.
32. Singh BR, DasGupta BR. Structure of heavy and light chain subunits of type A botulinum neurotoxin analyzed by circular dichroism and fluorescence measurements. Mol Cell Biochem 1989; 85:67-73.
33. Fasshauer D, Otto H, Eliason WK, Jahn R, Brunger AT. Structural changes are associated with soluble N-ethylmaleimide-sensitive fusion protein attachment protein receptor complex formation. J Biol Chem 1997;272:28036-28041.
34. Rost B, Sander C. Prediction of protein secondary structure at better than 70% accuracy. J Mol Biol 1993;232:584-599.
35. Rost B, Sander C. Combining evolutionary information and neural networks to predict protein secondary structure. Proteins:Struct., Funct. & Gen. 1994;19:55-72.
36. Anne C, Cornille, F., Lenoir, C., and Roques, B.P. High-throughput fluorogenic assay for determination of botulinum type B neurotoxin protease activity. Anal Biochem 2001;291:253-261.
37. Amersdorfer P, Wong C, Chen S, Smith T, Deshpande S, Sheridan R, Finnern R, et al. Molecular characterization of murine humoral immune response to botulinum

neurotoxin type A binding domain as assessed by using phage antibody libraries. Infect Immun 1997;65:3743-3752.
38. Adler M, Dinterman RE, Wannemacher RW. Protection by the heavy metal chelator N,N,N,N-tetrakis (2- pyridylmethyl)ethylenediamine (TPEN) against the lethal action of botulinum neurotoxin A and B. Toxicon 1997;35:1089-1100.
39. Hua SY, Charlton MP. Activity-dependent changes in partial VAMP complexes during neurotransmitter release. Nature Neurosci1999;2:1078-1983.
40. Boquet P, Duflot E. Tetanus toxin fragment forms channels in lipid vesicles at low pH. Proc Natl Acad Sci USA 1982;79:7614-7618.
41. Schengrund CL, DasGupta BR, Hughes CA, Ringler NJ. Ganglioside-induced adherence of botulinum and tetanus neurotoxins to adducin. J Neurochem 1996;66:2556-2561.
42. Srinivasan KN, Gopalakrishnakone P, Tan PT, Chew KC, Cheng B, Kini RM, Koh JL, et al. SCORPION, a molecular database of scorpion toxins. Toxicon 2002;40:23-31.
43. Nakata K, Takai T, Kaminuma T. Development of the receptor database (RDB): application to the endocrine disruptor problem. Bioinformatics 1999;15:544-552.