Annotation of the *Clostridium botulinum* Genome Strain 657/Type Ba4: Locus Tags: CBO_0025, CBO_0032, CBO_0029 and CBO_0033

Lauren H. Grover, Cassandra A. Pillard, Rachel M. Radack, Briget M. Whipple and Margaret S. Jolles
Silver Creek Central School District and The Western New York Genetics Research and Health Care Partnership

SEPA SCIENCE EDUCATION PARTNERSHIP AWARD
Supported by the National Institutes of Health

University at Buffalo

Abstract

Four students from Silver Creek Central School District High School participated in the Western New York Genetics in Research and Health Care Partnership Gene Annotation Research Study. Gene Annotation is the process of assigning function to genes using the Geni-Act website. Modules within the website that were utilized cover the scope of Basic Information, Sequence Based Similarity, Structure Based Evidence, Cellular Localization Data and Alternative Open Reading Frame. The data derived from the work allowed students to assign probable protein structure, location and function to their gene products. In all cases it is believed that the computer has called the protein correctly.

Introduction

Clostridium botulinum is a gram positive, motile, rod shaped, anaerobic bacteria that is present in soil and marine sediments. During adverse conditions the bacteria is able to survive by forming spores, which protect it and allow it to survive, dormant, until conditions become more favorable. The free-living form of the bacteria produces a neurotoxin that is the most lethal known to man. This neurotoxin, botulinum toxin, is a paralytic. The release of the neurotransmitter acetylcholine is inhibited and cells that contain the acetylcholine receptors are not able to respond, therefore voluntary movement of the skeletal system are diminished or not possible.

Clostridium botulinum can produce up to seven different types of toxins designated A-G. Types A, B, E & F are known to affect humans while C & D affect animals.
The most common way to encounter *Clostridium botulinum* and then fall victim to the Botulinum toxin is through consuming improperly canned or refrigerated food. Home canning of low acid foods can be problematic and mass produced canned goods, if in a deformed or bulging can, may carry the toxin.

Our team of 4 analyzed Strain 657/Type Ba4, Locus Tags CBO_0025, CBO_0032, CBO_0029 & CBO_0033.

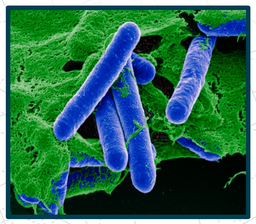


Figure 1 – left

Clostridium botulinum

Source:
https://microbewiki.kenyon.edu/index.php/Clostridium_botulinum_Neurotoxins

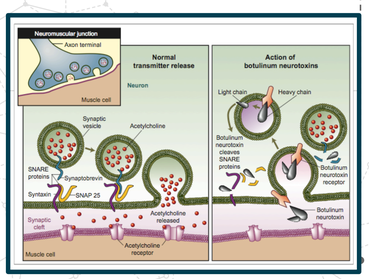


Figure 2- Right

C. botulinum mode of action

Source:
http://www.ebi.ac.uk/biomodels-main/static-pages.do?page=ModelMonth%2F2010-08

Methods

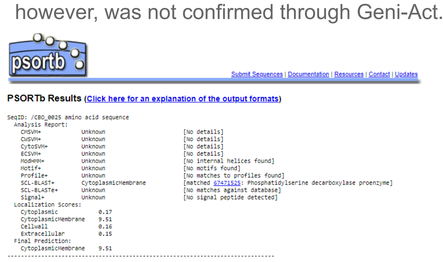
Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Clostridium botulinum's* genome annotation. The modules are described below:

Modules	Activities	Questions Investigated
Module 1- Basic Information Module	DNA Coordinates and Sequence, Protein Sequence	What is the sequence of my gene and protein? Where is it located in the genome?
Module 2- Sequence-Based Similarity Data	Blast, CDD, T-Coffee, WebLogo	Is my sequence similar to other sequences in Genbank?
Module 3- Cellular Localization Data	Gram Stain, TMHMM, SignalP, PSORT, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?
Module 4- Alternative Open Reading Frame	IMG Sequence Viewer For Alternate ORF Search	Has the amino acid sequence of my protein been called correctly by the computer?
Module 5- Structure-Based Evidence	TIGRfam, Pfam, PDB	Are there functional domains in my protein?

Results

CBO_0025

Locus Tag CBO_0025 is a gene containing 295 amino acids. BLAST searches using NR and Swiss Protein curated database identified the gene as coding for the proenzyme phosphatidylserine decarboxylase. Phosphatidylserine decarboxylase (PSD) is an important enzyme in the synthesis of phosphatidylethanolamine in both prokaryotes and eukaryotes. (Voelker, 1997 BBA) COG0688 also supports the BLAST results. The WebLogo for CBO_0025 shows several high conservation regions, reinforcing the BLAST results. TIGRFAM and Pfam also supporting BLAST with TIGR00163 and PF02666 as the same phosphatidylserine decarboxylase. TMHMM indicate no transmembrane helices and there is no signal peptide according to SignalP. PSORT-b, with a cytoplasmic membrane score of 9.51 predicts a protein in the cytoplasm of the cell; Phobius is in agreement. It appears that the start of the gene has been called correctly. All of these computer generated results, when combined shows that the gene at Locus Tag CBO_0025 most likely codes for an enzyme found in the cytoplasm. The fact that this is an enzyme, however, was not confirmed through Geni-Act.


psortb

Submit Sequences | Documentation | Resources | Contact | Updates

PSORTb Results (Click here for an explanation of the output formats)

```
SeqID: /CBO_0025 amino acid sequence
  Analysis Report:
    CytoSVM=          Unknown                 [No details]
    CuDuSVM=          Unknown                 [No details]
    CytoSVM=          Unknown                 [No details]
    ECSVM=            Unknown                 [No details]
    HookPRN=          Unknown                 [No internal helices found]
    Motif+=           Unknown                 [No motifs found]
    Profile+          Unknown                 [No matches to profiles found]
    SCL-BLAST+        CytoplasmicMembrane     [matched 67421123: Phosphatidylserine decarboxylase proenzyme]
    SCL-BLASTe+       Unknown                 [No matches against database]
    SignalP=          Unknown                 [No signal peptide detected]
  Localization Scores:
    Cytoplasmic               0.17
    CytoplasmicMembrane       9.51
    Cellwall                  0.16
    Extracellular             0.15
  Final Prediction:
    CytoplasmicMembrane       9.51
```

Figure 3 –

PSORT-b results for Locus Tag CBO_0023 indicating a cytoplasmic location

CBO_0032:

CBO_0032 has a sequence of 113 amino acids. When put through BLAST in both the Non-Redundant Data Base and the curated Swiss Protein Data Base, the top hit for the gene product name was YbaB/EbfC family nucleoid-associated protein. When researching YbaB, a Journal of Bacteriology article concluded that it is a 'new type of bacterial nucleoid associated protein and a global regulator of gene expression in Lyme Disease Spirochete' (Jutras et al, July 2012). When looking at the WebLogo it shows that this locus is highly conserved meaning that it is very similar in a lot of bacteria. COG0718, TIGRFAM #00103 and Pfam #02575 support the original BLAST results. It was determined that the function of this protein is unknown however with it being a nucleoid associated protein it can be inferred that it may interact with DNA. TMHMM indicate no transmembrane helices and Signal P indicates no signal peptide. PSORT-b, with a cytoplasmic score of 7.5 shows a probable cytoplasmic location, but Phobius contradicts that result. Throughout the research of CBO_0032 each piece of evidence reinforces the previous conclusions which implies that the computer has called this protein correctly.

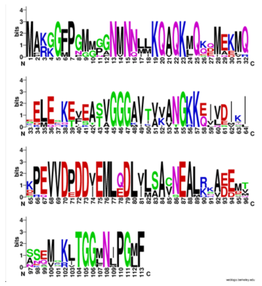


Figure 4 –

WebLogo for Locus Tag CBO_0032 showing areas of high conservation across the entire protein

CBO_0029:

CBO_0029 is a protein with 400 amino acids. The BLAST search, in both Non Redundant and curated Swiss Protein Data Bases show this is an enzyme called methionine gamma-lyase with strong e-values of 0.0. This enzyme is found from archaea to plants, but not in humans. It functions to degrade sulfur containing amino acids. (Sato & Nozaki, Nov. 2009, IUBMB Life) COG0626 indicated a Cystationine Beta Lyase/Gamma Synthase. TIGRFAM hit Tigr01328 also confirms this type of enzyme. T-Coffee and WebLogo show many areas of high conservation. Pfam yielded PF01053 which is a Cys/Met metabolism PLP-dependent enzyme family. Clan #CL0061 is a PLP dependent aminotransferase superfamily. All of these results indicate CBO_0029 produces and enzyme which is in the lyase family of enzymes. Lyases are enzymes that catatyze the cleavage of C-C, C-O, C-N bonds by other means than by hydrolysis or oxidation. Enzymatic pathways were not established through Geni-Act. Protein Data Bank results in 5DX5, illustrating the crystal structure of methionine gamma-lyase from Clostridium

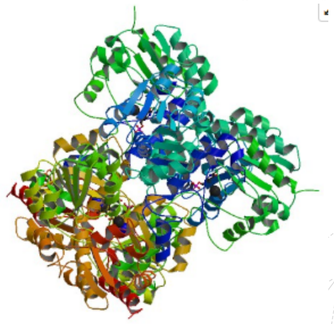


Figure 5 –

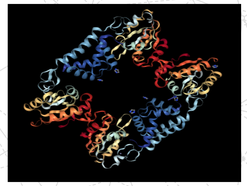
PDB5DX5
Crystal structure of methionine gamma-lyase from *Clostridium sporogenes.*

CBO_0033:

The gene CBO_0033 codes for 198 amino acids and all of the hits from BLAST using Non Redundant and curated Swiss Protein Data Base showed that this gene coded for a recombination protein, RecR. The protein is likely involved in the repair of damaged DNA strands. It could also be involved in the initiation of recombination or the actual recombination and repair process. WebLogo showed high conservation of amino acids in this gene. It is also likely that the protein has an aspect that has to do with helping keep DNA from knotting up during the replication process. Pfam hits PF02132 and PF13662 indicate a RecR domain and a Toprim 4 domain respectively. The Toprim 4 domain includes proteins that repair DNA as well as replicate it. Even though only one database came up with this domain, it would make sense that this protein was in it since all the other databases indicated that this protein was involved in DNA repair of some kind. TMHMM indicates no transmembrane helices and Signal P shows no signal peptide. This protein was at the exact cut off point of 7.5 for being cytoplasmic according to PSORT-b, so it is unlikely that it is in the cytoplasm. The evidence shows that this protein is responsible for DNA recombination, repair and possibly replication and implies that it would be an enzyme, but this was not proven using Geni-Act.

Figure 6 –

PDB3VDP – Structure of the recombination mediator protein RecR in RecFOR Pathway



Conclusion

The GENI-ACT proposed gene product did not differ significantly from the proposed gene annotation for each of the genes in the group and as such, the genes appear to be correctly annotated by the computer database.

Gene Locus	Geni-Act Gene Products	Proposed Annotation
CBO_0025	Phosphatidylserine Decarboxylase	Phosphatidylserine Decarboxylase
CBO_0032	Nucleoid Associated Protein	Nucleoid Associated Protein
CBO_0029	Methionine Gamma Lyase	Methionine Gamma Lyase
CBO_0033	Recombination Protein RecR	Recombination Protein RecR

References

https://www.sciencedirect.com/science/article/pii/S000527609700101X
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3434759/
https://www.ncbi.nlm.nih.gov/pubmed/19859976
https://en.wikibooks.org/wiki/Structural_Biochemistry/Enzyme_Catalytic_Mechanism/Enzyme_Classification/Lyases

Acknowledgments

- Supported by an NIH Science Education Partnership (SEPA) Award - R25ODO10536
- Stephen Koury, PhD - Dept. of Biotechnical & Clinical Laboratory Sciences
- www.buffalo.edu