Annotation of the Clostridium botulinum Genome Strain 657/Type Ba4:

Locus Tags CBO_0028, CBO_0026, CBO_0031 and CBO_0030

Alexis M. Murawski*, Jenna L. Miller, Samantha C. Page, Ryan T. Rocque and Margaret S. Jolles Silver Creek Central School District & The Western New York Genetics in Research and Health Care Partnership







Figure 6 -

PDB 2B3J

of S. aureus

deaminase

Crystal Structure

tRNA adenosine

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, Alternative Open Reading Frame and Enzymatic Function. 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_0028, CBO_0026, CBO_0031 & CBO_0030.

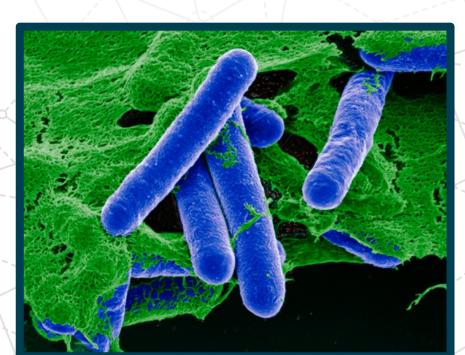


Figure 1 – left

Clostridium botulinum

Source: https://mi

Neuromuscular junction

https://microbewiki.kenyon.edu/ind ex.php/Clostridium_botulinum_Ne urotoxins

botulinum neurotoxins

Figure 2 - Right

C. botulinum mode of action

Source:
http://www.ebi.ac.uk/biomo
dels-main/staticpages.do?page=ModelMo
nth%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?
Module 6- Enzymatic Function	KEGG, MetaCyc, E.C. Number,	In what process does my protein take part?

Results

CBO 0028:

The gene at Locus tag CBO_0028 contains 512 amino acids. Using Swiss Protein curated database, the top hit was Transcriptional regulatory protein, TyrR. This protein is involved in transcriptional regulation of aromatic amino acid biosynthesis and transport. The TyrR protein of *E. Coli* has ATPase activity, stimulated by tyrosine, and resembles the transcriptional activator NtrC. (Biol Chem. Journal, 1993) The first TigrFam hit, Tigr01818 confirms this by indicating the presence of NtrC, a nitrogen regulatory protein.

Pfam's top hit, #PF00158, indicates a sigma-54 interaction domain, which contains NtrC proteins and can direct the transcription of a wide variety of genes. Clan #CL0023 shows that P-loop NTPase is a member of AAA family proteins that assist in the assembly, operation, or disassembly of protein complexes. Protein Data Bank #5EXX indicates homology with an AAA+ ATPase FleQ from Psuedemonas aeruginosa, which can be divided into subdomain N-terminal P-loop NtPase. COG result #COG3829 supports previous information found through BLAST, TigrFam, Pfam and Clan. The domain family is RocR, which are transcriptional regulator containing PAS, AAA-type ATPase and DNA binding Fis domains.

Location of this protein is believed to be cytoplasmic based on results from PSORTb. The TMHMM results indicate no helices and Signal P shows no signal peptide present. The start codon appears to have been called correctly based on Shine-Dalgarno sequences and KEGG results indicate that this protein is an enzyme associated with nitrogen metabolism in the cell.

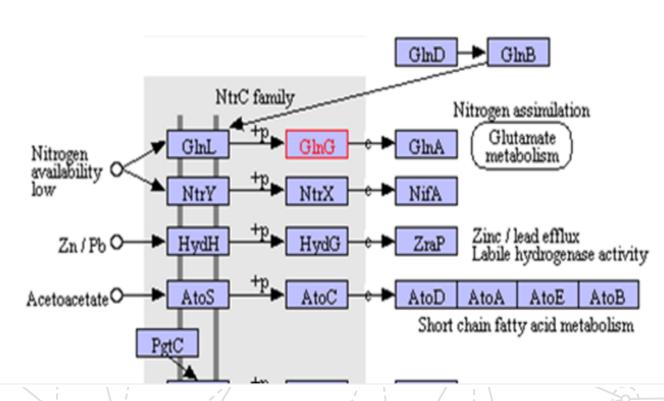


Figure 3 –

indicating participation in NtrC family of enzymes and association with nitrogen metabolism.

CBO 0026:

The gene at Locus Tag CBO 0026 has 532 amino acids. The first BLAST hit from database NR is AarF/ABC1/UbiB kinase family protein. The second hit from NR confirms a protein kinase. Results from Swiss Protein curated database also confirms protein kinase, with the full name being Ubiquinone biosynthesis protein UbiB. Protein kinases are a family of enzymes whose function is to transfer a phosphate to another molecule thereby modifying function of a protein. COG results #0661 and TIGR01982 also indicate a protein kinase regulating ubiquinone biosynthesis. Pfam #PF03109 results show an ABC1 family association from yeast and AarF from E. coli. CLAN0016 also indicates a protein kinase family. TMHMM predicts two membrane helices with no signal peptide. PSORT-B predicts a cytoplasmic membrane protein with a score of 9.96 and Phobius confirms the 2 helices. The protein start location was called correctly and even though this is believed to be an enzyme, this was not confirmed through Geni-Act.

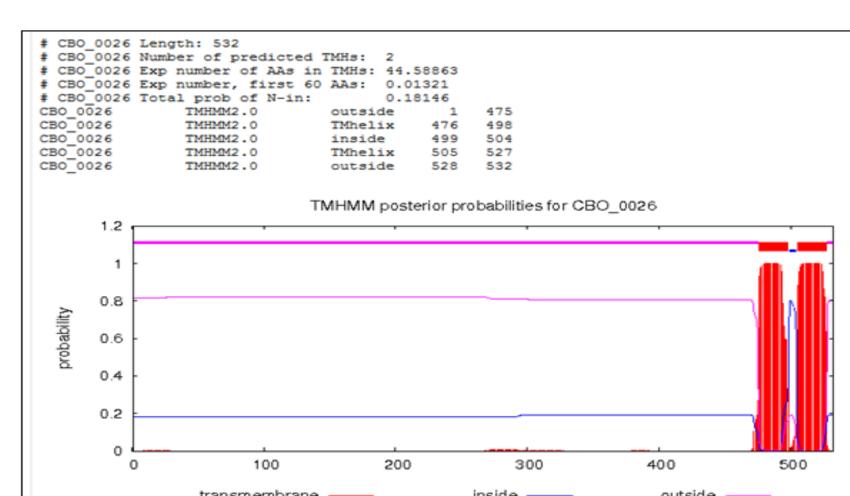


Figure 4 - TMHMM results indicate 2 transmembrane helices for Locus Tag CBO_0026

CBO_0031:

CBO 0031 is located at coordinates 38095 - 39720 and contains 541 amino acids. BLAST results from both Non Redundant and curated Swiss Protein Databases indicate a DNA Polymerase III, subunit gamma/tau. The enzyme DNA Polymerase III is a complex multichain enzyme responsible for replication in bacteria. COG2812 resulted in a DnaX Protein Domain involved with replication, Family recombination and repair of DNA. WebLogo indicates a high level of conservation in the first 2/3rd of the protein. TIGRFAM resulted in 2 hits, Tigr02397 & Tigr00678. Tigr02397 confirms the BLAST results and Tigr00678 shows a protein that is homologous to the gamma/tau subunit, known as DNA polymerase III, delta' subunit. Pfam, PF13177 also identifies the delta subunit and Pfam, PF12169 confirms the gamma/tau subunit. Two Clans were identified, CL0023 & CL0604. CL0023 indicates a P-Loop NTPase family and CL0604 results in a Post-AAA+ oligomerization domain like superfamily. P-Loop NTPase is a commonly seen fold that allows binding between nucleotides and proteins. TMHMM indicates not transmembrane helices and SignalP results show no signal peptide. PSORTb describes the location of this protein as cytoplasmic with a score of 9.97. Phobius contradicts this result however. PDB results in 3GLF which is the 'Crystal Structure of the E. coli Clamp Loader Bound to Primer-Template DNA' reinforcing the idea that this protein is involved with DNA replication. It is believed this protein's start has been called correctly.

MaYTALYREWRPesF-eVyGQEHLT.TL Net _____BleHAYLEGGTRGTGKTSIAKILEKA UNCLUC BGEPCN-C-MC-KIN-G-C-DV-E-DAAson-yeelbe-leevayePeeckxKVYI. DEVHMLS & AVNAELKTLEEPP & V. V. LLATT DPQKuPaTILSRCQREDEKRISASSIA BLR. ╚╗┡╏┈┇┇Х╒┇┇╇┱╏╏Б┋ХХ╒┋<mark>С</mark>Қ<u>Б</u>┾ऱ┈<u>Е</u>ТҚБй I HERN-L-XKY--NPEEY DUS ES ... KE Q≈≈≈≈≈≈E≈IMB~Ie:\F9EAE≈≈≈K≈≈KQ≈Rı ⋛<mark>⋌┞⋶⋫</mark>⋞⋋⋷⋷⋐⋵<mark>₭⋷⋶⋋</mark>ⅅ⊥⋛⋍<mark>⋿⋍⋷⋷⋼</mark>₿⊁⋻⋷⊦<mark>⋿</mark>⋿⋷⋷ ┖╒╒╼┺┰╘╒╾Үϗ╍╼╟⋳<mark>₽</mark>╽┢⋹с∊⋉⋳⋻₿┈╟∊⋎⋨⋪⋷ _⋽⋍⋍⋞⋧⋍⋷⋍∊⋵⋴⋴⋳⋷**⋷⋍⋷⋍**⋦⋍<mark>⋉⋷</mark>Х⋬**⋸**⋳⋉⋉⋻<mark>⋷⋶</mark>⋉

Figure 5 - WebLogo for CBO_0031 indicating high conservation in the first 2/3rds of the protein

CBO_0030:

The gene at Locus Tag CBO 0030 has 430 nucleotides and 145 amino acids in the protein. Curated database Swiss Protein confirmed that this gene codes for a tRNA - specific adenosine deaminase. This protein catalyzes the deamination of adenosine to inosine at the wobble position 34 of tRNA. COG0590 supports the BLAST results. WebLogo showed many areas of polar amino acids and generally good conservation of sequence. Pfam, PF14437 provided information that the gene was part of a deaminase like family; Clan #CL0109 results show a MAF-B19 like deaminase. Protein Data Bank provided the hit 2B3J which is the Crystal Structure of Staphylococcus aureus tRNA Adenosine Deaminase, TadA, in Complex with RNA. TMHMM and Phobius agree there are no transmembrane helices. SignalP states no signal protein exists. With all the information listed Psortb and Phobius were not able to predict the location for this protein. The computer has correctly called the start codon. This gene codes for an enzyme but was not confirmed through Geni-

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_0028	Transcriptional Regulatory Protein, TyrR	Transcriptional Regulatory Protein, TyrR
CBO_0026	Protein kinase	Protein kinase
CBO_0031	DNA Polymerase III, subunit gamma/tau	DNA Polymerase III, subunit gamma/tau
CBO_0030	tRNA – specific adenosine deaminase	tRNA – specific adenosine deaminase

References

http://www.ebi.ac.uk/interpro/entry/IPR027417
https://www.uniprot.org/uniprot/P09122
http://www.uniprot.org/uniprot/P68398

Acknowledgments

- Supported by an NIH Science Education Partnership (SEPA)
 Award R250D010536
- Stephen Koury, PhD RDept. of Biotechnical & Clinical Laboratory Sciences
- www.buffalo.edu