Montclair State University
Montclair State University Digital Commons

Theses, Dissertations and Culminating Projects

8-2017

Molecular Cloning and Characterization of a Hemolytic Lectin from Venom of the Stinging Sea Nettle, Chrysaora quinquecirrha

Anup Khanal
Montclair State University

Follow this and additional works at: https://digitalcommons.montclair.edu/etd

 Part of the Biology Commons

Recommended Citation
Khanal, Anup, "Molecular Cloning and Characterization of a Hemolytic Lectin from Venom of the Stinging Sea Nettle, Chrysaora quinquecirrha" (2017). *Theses, Dissertations and Culminating Projects*. 14.
https://digitalcommons.montclair.edu/etd/14

This Thesis is brought to you for free and open access by Montclair State University Digital Commons. It has been accepted for inclusion in Theses, Dissertations and Culminating Projects by an authorized administrator of Montclair State University Digital Commons. For more information, please contact digitalcommons@montclair.edu.

Abstract

I have identified a novel pore-forming toxin (PFT) in the venom of the Sea Nettle (*Chrysaora quinquicherra*), a protein I have named Chrysaoralin. This protein is discharged from specialized organelles called cnidocysts (nematocysts) found primarily in the tentacles of this jellyfish. Chrysaoralin was first identified by Nextgen sequencing (RNA-Seq) of libraries made from mRNA isolated from tentacles of mature medusa collected from Barnegat Bay, NJ. The full-length of the Chrysaoralin gene is 1365 bp, encoding a protein of 454 AA (50.695 kD; pI = 6.58). The SignalP 4.1 algorithm (http://www.cbs.dtu.dk/services/SignalP/) predicts a signal peptide of 22 AA. The mature protein (minus the putative signal peptide) is 432 AA (48.321 kD; pI = 6.58). This protein shows strong homology (66%) to a hemolytic lectin from the sea cucumber, *Cucumaria echinata* (Phylum Echinodermata). In support of this fact hemolytic activity was detected in the purified nematocyst preparations, which demonstrates sensitivity to both boiling and Proteinase K digestion, suggesting this activity is proteinaceous. The RNA-Seq data was verified by generating PCR amplicons using 9 sets of primers that span the full gene. Genomic sequences from both Barnegat Bay and Chesapeake Bay Chrysaoralin were intron-less. I also modified and subcloned the full-length Chrysaoralin gene into a pET SUMO expression vector and transformed into *E. coli*. Future expression of this recombinant protein in *E. coli* may further the understanding of the physiological role of Chrysaoralin in human envenomations.

MONTCLAIR STATE UNIVERSITY

MOLECULAR CLONING AND CHARACTERIZATION OF A HEMOLYTIC LECTIN FROM VENOM OF THE STINGING SEA NETTLE, *CHRYSAORA QUINQUECIRRHA*

by

ANUP KHANAL

A Master's Thesis Submitted to the Faculty of

Montclair State University

In Partial Fulfillment of the Requirements

For the Degree of

Master of Science

August 2017

College of Science and Mathematics

Biology

Thesis Committee

[redacted]

John J. Gaynor, Ph.D.
Thesis Sponsor

[redacted]

Paul A. X. Bologna, Ph.D.
Committee Member

[redacted]

Vladislav Snitsarev, Ph.D.
Committee Member

MOLECULAR CLONING AND CHARACTERIZATION OF A HEMOLYTIC LECTIN

FROM VENOM OF THE

STINGING SEA NETTLE, *CHRYSAORA QUINQUECIRRHA*

A THESIS

Submitted in partial fulfillment of the requirements

For the degree of Master of Science

by

ANUP KHANAL

Montclair State University

Montclair, NJ

2017

Copyright ©2017 by *Anup Khanal*. All Rights Reserved.

Acknowledgements

I would like to give special acknowledgement to my advisor, Professor John Gaynor. Without his patience and guidance, this work would not have been possible. I am incredibly grateful for the support I received from him while working on this research. I would also like to thank Professors Paul Bologna and Vladislav Snitsarev for serving on my thesis committee. Your time, effort, and feedback are truly appreciated. Lastly, I would like to extend my sincere gratitude to my friends and family.

Table of Contents

List of Figures

List of Tables

List of Appendices

Introduction to *Chrysaora quinquecirrha*

A common jellyfish found along the Atlantic and Gulf coasts of the United States is the cnidarian, *Chrysaora quinquecirrha*, commonly known as the Atlantic or Stinging Sea Nettle. *C. quinquecirrha* are found in waters of low to moderate salinity, mainly in estuaries, off the coast of the Atlantic, Indian, and western Pacific oceans (Figure 1; Mayer, 1910). In the United States, they are found seasonally, from July through August, in the Chesapeake Bay (Calder, 1972), which is the largest estuary in the United States bordering Maryland and Virginia. In Barnegat Bay, NJ high reproductive potential of the *C. quinquecirrha* adults has been reported and establishment of polyp colonies in new habitats of Barnegat Bay could be detrimental to fisheries in the region (Bologna and Gaynor, 2013). Additionally, a sudden rise in jellyfish population can cause a significant impact on human activities and marine ecosystems (Bordeur *et al*., 2008). Therefore, a detailed understanding of the physical and biochemical features of this organism and its venom is much warranted, and will enable us to be better prepared during events of human envenomation.

As is characteristic with all Scyphozoans, sea nettles are radially symmetrical, diploblastic organisms consisting a dome shaped bell and long silky tentacles (Figure 2). The dome-shaped bell measures approximately 25 centimeters in diameter and contains 8 scalloped, flower-petal shaped lobes with 7 to 10 silk thread like tentacles lined with specialized stinging organelles called nematocysts that extend outwards from each lobe. These tentacles can grow up to 50 centimeters in length. Additionally, four long, ribbon-like oral arms or

Figure 1: Global distribution of *Chrysaora quinquecirrha* (shown in gold). Image source: Global biodiversity information facility (Url: https://demo.gbif.org/species/5185413).

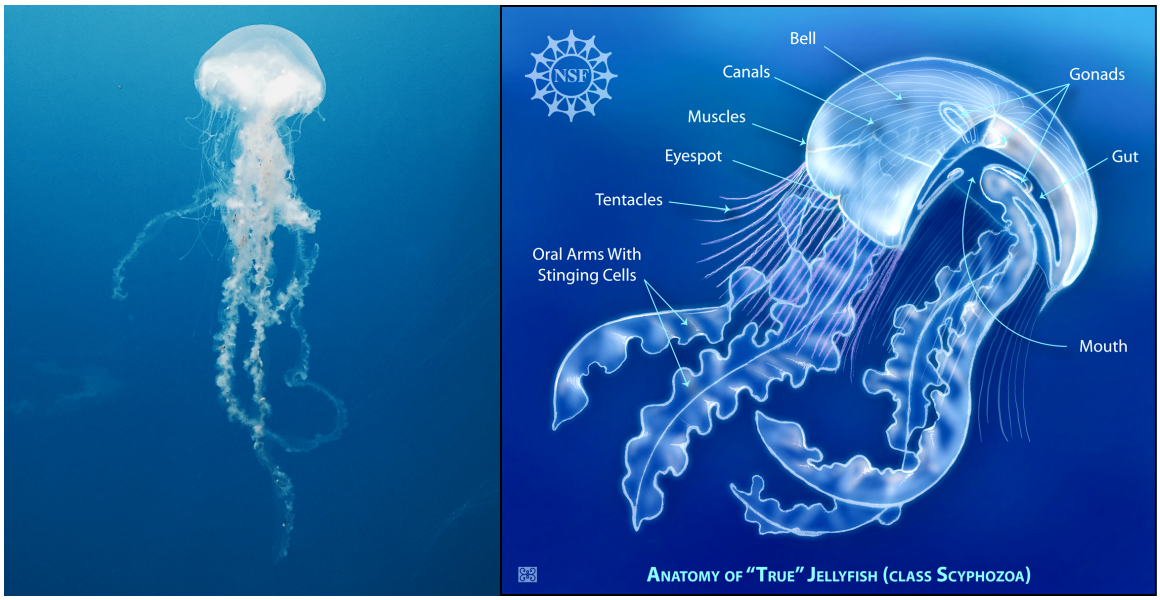


Figure 2.1 left: An adult *Chrysaora quinquecirrha* medusa. Image Credit: Dena Restaino. *Figure 2.2 right:* Anatomy of a true jellyfish. Image credit: Zina Deretsky, National Science Foundation.

lappets extend outward from the middle of the jellyfish dome, which serve to bring food up to the mouth (Ford *et al*, 1997).

Chrysaora quinquecirrha exist in two distinct life stages: sessile polyp form and free-swimming medusa form (Figure 3), exhibiting an alternation of generation, and both stages are able to survive in waters of low salinity and low dissolved oxygen (Condon *et al*., 2001). This tolerant adaptation, coupled with toxic venom, gives *C. quinquecirrha* an unparalleled advantage over other organisms for procurement of food and nutrients. Alternating between the polyp and medusa life stages, *C. quinquecirrha* are able to produce colonies that persist for extended lengths of time and survive in cold and hypoxic water that are unfavorable for other organisms in its habitat (Purcell, 1999; Purcell, 2001). Specialized stinging cells capable of delivering potent venom are found in both polyp and medusa stages of *C. quinquecirrha* (Calder 1972b). A combination of adaptive nature and toxic venom enable *C. quinquecirrha* to outcompete other organisms and cause massive seasonal blooms in the estuaries where they are found.

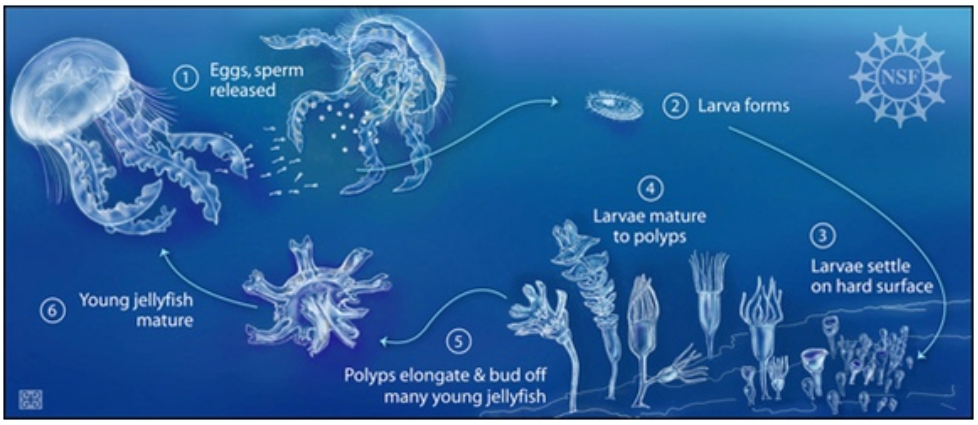
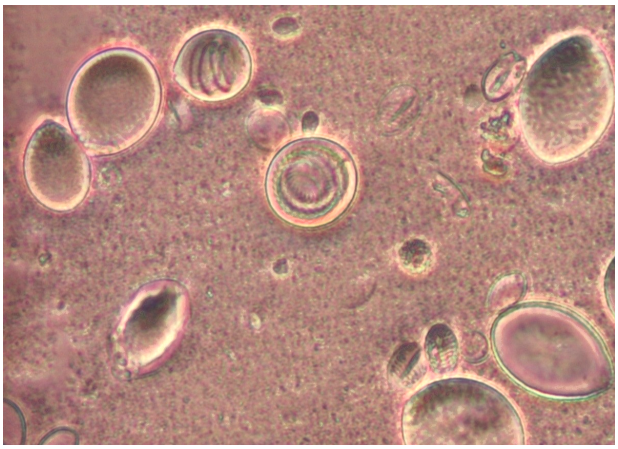


Figure 3: The life cycle of scyphozoan sea jellies. Image credit: Zina Deretsky, National Science Foundation.

Nematocysts

The origin of cnidarians is dated to more than 750 million years ago, during the Ediacaran period (Technau and Steele, 2011). Cnidarians derive their name from the stinging organelle called the nematocyst (or cnidocyst), common to more than 10,000 cnidarian species, which they primarily use for procurement of their prey. Based on their morphology, nematocysts can be of 25 to 30 different types (Özbek, 2009). The nematocysts exist in a variety of sizes ranging between 5 and 100 μm, and shapes ranging from round to cylindrical (see Figure 4). However, structurally all nematocysts contain a wall and a tubule that may be further enhanced with spines and appendices (Teragawa and Bode, 1995). A mechanosensory apparatus called the cnidocil is present at the tip of the nematocyte (or cnidocyte), which is a critical structure during the discharge process (Figure 5). The nematocyst itself is tightly attached to the cytoplasm of the nematocyte by microtubules surrounding the outer capsule (Engel *et al*., 2002). When nematocysts are triggered to fire and their polar tubule is injected into the integument of a prey or a victim, a mixture of proteins, polypeptides and enzymes are released that modify cellular processes by disruption of ion channels, formation of membrane pores or through enzymatic mechanisms (Ponce *et al*., 2015).

During envenomation events, nematocysts are discharged because of physical and/or chemical stimuli, however the discharge mechanism is not fully understood (Jouiaei *et al*., 2015). Nematocyst discharge is notable because it has been recorded to be one of the fastest biophysical events ever recorded



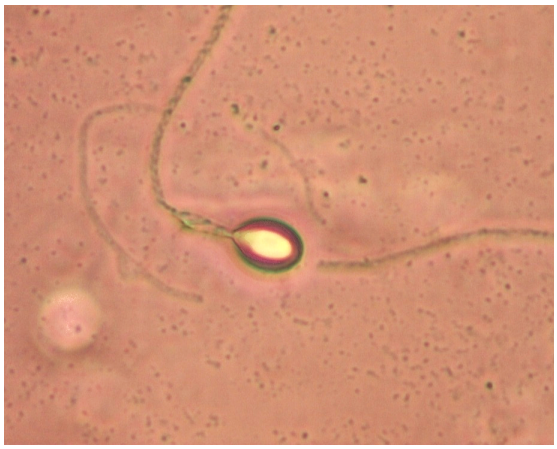


Figure 4.1 left: Nematocysts of different shapes and sizes under the microscope at 400X magnification. *Figure 4.2 right:* A discharged nematocyst. Image credit: John Gaynor.

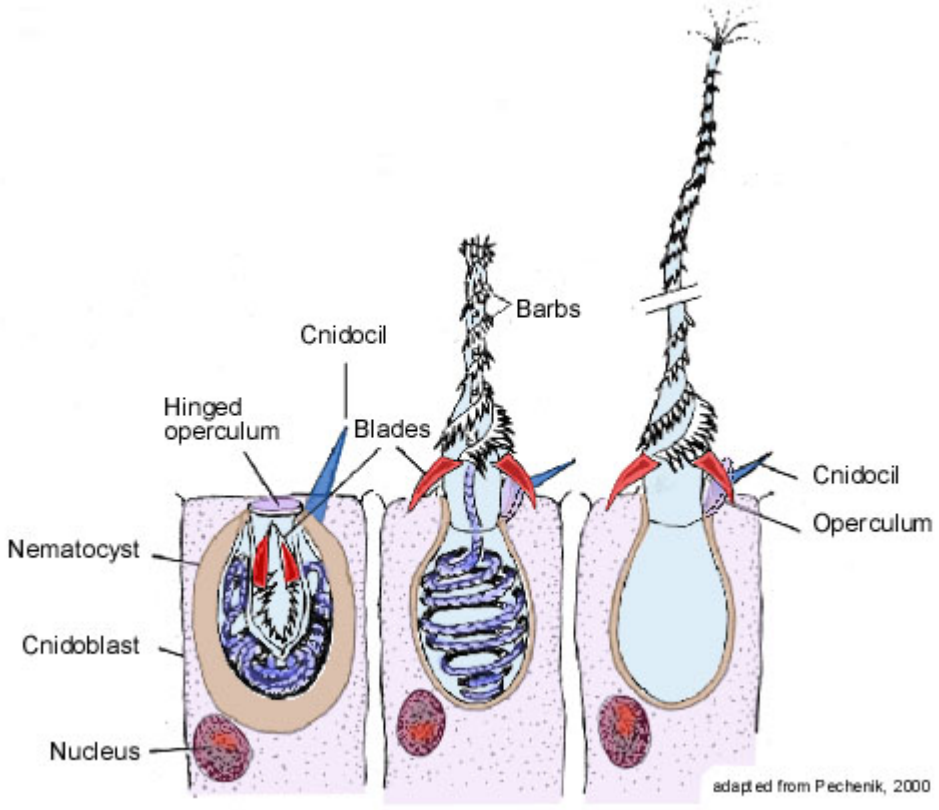


Figure 5: Nematocyst discharge mechanism. Image source: Pechenik, 2000.

(Holstein and Tardent, 1984; Fautin, 2009). Early experiments on the nematocyst discharge by Holstein and Tardent (1984) recorded the discharge event to take only around 3 milliseconds. Nematocyst discharge is dependent on the inherent structural contractional energy of the microtubule scaffold of the nematocyst, fueled by the intramembranous charge and chemical flux, and finally triggered by a chemical and/or mechanical stimuli (Holstein and Tardent, 1984; Watson and Hessinger, 1988; Cannon and Wagner, 2003). Additionally, the diversity in nematocyst morphology in cnidarians, and even between organisms of the same species, can be attributed to phenotypic plasticity driven by their environmental forces. Therefore, a careful look into the genome of these organisms and particularly a study of their toxins may provide us more insight about their feeding behavior and stinging mechanism.

Cnidarian Venom

Cnidarians use a mixture of proteinaceous and non-proteinaceous toxin components that have evolved over the last several hundred million years to subdue their prey and to escape from their predators (Mariotini, 2014; Jouiaei *et al*., 2015). Understanding the composition and characteristics of cnidarian venom is essential because cnidarian venoms provide us with a repertoire of bioactive compounds with therapeutic promise against many human neurological, hematological, infectivological, and oncological maladies (Mariottini and Pane, 2013). The compounds in cnidarian venom can be broadly categorized as enzymes, pore-forming toxins, and neurotoxins (Lee *et al*., 2011; Jouiaei *et al*.,

2015). These compounds can be further subcategorized into C-type lectins, phospholipase A2, potassium channel inhibitors, protease inhibitors, metalloproteases, hemolysins, and other toxins. Acting singly or in conjunction, these toxins produce a myriad of localized and systemic effects (Frazao and Antunes, 2016; Ponce *et al*., 2016).

In general, most jellyfish from the genus *Chrysaora* (sea nettles) inflict stings that can cause injurious reactions in humans. Some common physiological effects include a burning sensation, blisters, skin redness, localized edema, headaches, cramps, and lachrymation (Newman-Martin, 2007; Cegolon *et al*., 2013). Based on clinical studies, the venom of *C. quinquecirrha* is seen to cause cessation of spontaneous beating of a primary culture of embryonic chick cardiocytes (Cobbs *et al*., 1983; Kelman *et al*., 1984). *C. quinquecirrha* venom has also been seen to induce mitogenic activity. It was able to produce nuclear alterations and dissolution of intercellular collagen in Chinese hamster ovary K-1 cells (CHO K-1) (Neeman *et al*., 1980a, 1980b). When tested against human hepatocytes, *C. quinquecirrha* venom caused an initial increase in metabolic activity, followed by a sharp decrease and cell death within minutes (Cao *et al*., 1998). However, there are few studies regarding *C. quinquecirrha* venom. Although, hemolytic activity of *C. quinquecirrha* venom has been demonstrated and evaluated by Ponce *et al.* in 2015, there is no sequence data available for the venom proteins yet.

Pore-Forming Toxins

The presence of a plasma membrane is probably the fundamental difference between life and non-life. When different biochemical components are surrounded within a membrane, not only does it define a boundary, but it also creates two different environments, each partitioned with completely different concentrations and chemical properties (Bischofberger *et al*., 2012). A cell membrane is also the key target during intercellular conflicts. It may be one reason why organisms from all kingdoms have evolved molecules that can alter membrane permeability and cause lysis of cells.

Pore-forming proteins are such membrane altering molecules and are frequently components of the toxin repertoire of many organisms in our biosphere. As the name implies, pore-forming proteins form a transmembrane pore in the cell membrane and disrupt the permeability barrier that quickly leads to cell death. Hemolytic activity has been demonstrated with *C. quinquecirrha* venom (Long-Rowe and Burnett, 1994; Bloom *et al*., 2001; Lozanno, 2013) and it is predicted that the erythrocyte membrane is ruptured by a pore-forming mechanism of one of its peptide toxins.

Although pore-forming proteins are produced as simple water-soluble peptides, they transform into pore penetrating membrane proteins only when they reach their intended target (Parker and Feil, 2005). Out of 300 protein toxins characterized to date, around 100 were responsible for disrupting the cell membrane by formation of some kind of pore (Feil *et al*., 2010). Bacterial pore-forming toxins are the best-characterized and also the largest class of pore-forming

proteins. Bacterial pore-forming toxins are employed either to kill other bacteria (Lakey *et al*., 1994) or to affect their hosts in order to promote colonization and spread during pathogenesis (Bischofberger *et al*., 2012). A comparison of pore-forming proteins is summarized in Table 1.

Many eukaryotic organisms produce pore-forming proteins either for defense, or for procurement of food (Sousa *et al*., 1994; Tomita *et al*., 2004; Sher *et al*., 2005; Kafsack *et al*., 2009; Kristan *et al*., 2009) including protozoan parasites (PLP1 - perforin-like protein 1), fungi (pleurotolysin), sea anemones (equinatoxin II), hydra (hydralysin), or plants (enterolobin). Vertebrates use pore-forming proteins to kill bacteria (complement membrane attack complex, MAC), to kill infected or malignant cells (perforin), or to permeabilize mitochondria in order to trigger apoptosis (members of the Bak family) (Bischofberger *et al*., 2012). Sometimes pore-forming proteins are seen to induce unintended consequences such as the proteins involved in neurodegenerative diseases. For example, α-synuclein or the ß-amyloid peptide of Alzheimer's can assemble into pore-forming aggregates that are similar to pore-forming toxins (Kagan, 2012).

The dimensions of the pore, duration of pore formation, and localization of pore forming effect depend on multiple variables. These variables include concentration of the pore-forming peptide, pore diameter, amino acid residues that line the pore lumen, number of pores per cell, and stability of the pore (Bischofberger *et al*., 2012). Among the different variables, the diameter and size of the pore varies greatly between different organisms with the pore lumens ranging between 10Å to 150Å (Feil *et al*., 2010). Pore-forming toxins can be

Group	*Organism*	Pore forming Protein	Reference	Pdb ID	AA Identity	E value	Query Cover
Bacteria	*V. cholerae*	Vibrio cholerae	Olson and Gouaux, 2005	1XEZ	40%	0.38	32%
	A. hydrophila	Aerolysin	Abrami et al., 2000	5JZH	33%	0.49	19%
	C. absonum	Alpha-toxin	Clark et al., 2003	1OLP	9%	2.9	9%
	B. anthracis	Protective Antigen PA	Petosa et al., 1997	1ACC	35%	1.2	31%
	C. perfringens	Perfringolysin O	Feil et al., 2012	1M3I	25%	9.0	1%
	L. monocytogenes	Listeriolysin O	Koester et al., 2014	4CDB	50%	9.0	1%
	S. intermedius	Intermedilysin	Polekhina et al., 2005	1S3R	50%	9.0	1%
	V. cholerae	HlyA	Linhartova et al., 2010	3O44	40%	0.34	32%
	S. aureus	staphylococcal alpha-hemolysin	Song et al., 1996	7AHL	23%	3.7	10%
Hydra	*C. viridissima*	Hydralysin	Sher et al., 2005	N/A	27%	0.18	26%
Sea Anemone	*A. equina*	Equinatoxin II	Athanasiadis et al., 2001	1IAZ	39%	0.57	23%
Mouse	*M. musculus*	Perforin	Law et al., 2010	3NSJ	50%	2.7	15%
Human	*H. sapiens*	C9 complement	Dudkina et al., 2016	5FMW	26%	1.8	34%
Plant	*E. contortisiliquum*	Enterolobin	Fontes et al., 1997	N/A	35%	0.31	11%
Fungi	*P. ostreatus*	Pleurotolysin	Lukoyanova et al., 2015 Sakurai et al., 2004	4V2T	29%	1.9	17%
Parasite	*T. gondii*	Perforin	Yan et al., 2011	N/A	24%	4.7	16%
Mollusc	*B. glabrata*	Biomphalysin	Galinier et al., 2013	N/A	56%	2.0	12%
Sea Cucumber	*C. echinata*	Hemolytic Lectin	Uchida et al., 2004	1VCL	64%	0.0	100%

Table 1: Representative pore-forming proteins. The table shows protein source, common names and 4-character unique Protein Data Bank (PDB) identifier of the protein entries in the database. The extent to which Chrysaoralin and other protein sequences share the same residues at the same positions in an alignment are listed as percent identity. The lower the E value, the more significant is the alignment. Query cover shows the percentage of Chrysaoralin sequence aligned to the protein sequences in GenBank.

categorized into two groups based on the structural feature they utilize to cross the cell membrane: either as an alpha helix (α-PFT) or as a ß-barrel (ß-PFT).

For instance, the alpha-pore forming toxins are predicted to form pores using their alpha helices. Colicins, produced by *E. coli*, are a classic example of alpha pore-forming toxins (Cascales *et al*., 2007). This category also includes *Pseudomonas aeruginosa* exotoxin A, some insecticidal delta-endotoxins (Cry), and diphtheria toxin (Allured *et al*., 1986; Lee *et al*., 1991; Choe *et al*., 1992). Some apoptotic proteins of the Bcl-2 family have also been seen to possess structural similarity and form ion channels similar to other alpha pore-forming toxins (Feil *et al*., 2010). The *Escherichia coli* hemolysin E is a representative example of the alpha-pore forming toxin (Feil *et al*., 2010). The crystal structure of this toxin in its water-soluble state is a predominantly helical molecule with its core formed by four helix bundles (Figure 6.1) (Wallace *et al*., 2000). In another case, the crystal structure of the *Pseudomonas aeruginosa* exotoxin A contains six alpha-helical structures in its membrane translocation domain (Figure 6.2) (Wedekind *et al*., 2001). Details on alpha-pore forming toxins is available in a review by Iacovache *et al.* (2008).

The beta pore-forming toxins are the second major class of pore-forming toxins. This category of toxins attaches to the membrane of their intended target and form a beta-barrel upon insertion into the lipid bilayer. This type of pore formation is seen in aerolysin, *Clostridium septicum* alpha-toxin, *Staphylococcus* alpha-hemolysin, anthrax protective antigen, *Bacillus thuringiensis* Cyt delta-endotoxins, and cholesterol dependent cytolysis, among many others (Tweten *et*

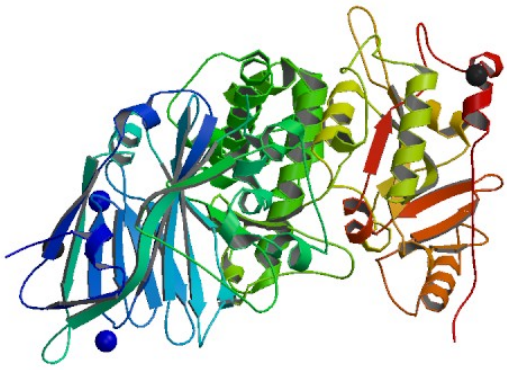


Figure 6.1 (left): Hemolysin E (HlyE) is a pore-forming toxin of *Escherichia coli*, *Salmonella typhi*, and *Shigella flexneri* (Wallace *et al*., 2000). *Figure 6.2 (right):* Exotoxin A of *Pseudomonas aeruginosa* (*Wedekind et al.,* 2001).

al., 2001; Menestrina *et al*., 2003; Iacovache et al., 2008). The membrane attack complex/perforin superfamily, which include proteins in the complement cascade of higher organisms such as C6, C7, C8-alpha, C8-beta, and C9 have also been seen to possess structural similarities to beta-pore-forming toxins (Hadders *et al*., 2007). The crystal structure of many of these pore forming toxins have been resolved. In some instances, the pore forming toxins from different organisms are structurally and functionally similar, although they possess no sequence homology.

Unlike the alpha-pore forming toxins, the domains of the beta-pore forming toxins do not perform distinct biological activity, are more diverse, and the translocation domain appear to be more cryptic (Feil *et al*., 2010). One of the earliest crystal structure of the ß-pore forming toxins was of the *Aeromonas hydrophila* proaerolysin. Different from the previously determined structures of the

alpha-pore forming toxins, the proaerolysin predominantly contained beta-sheets (Figure 7.1) (Parker *et al*., 1994).

Figure 7.1 (left): Aeromonas toxin proaerolysin predominantly contains beta-sheets that undergo a multi-step transformational change to form a transmembrane channel (Parker *et al*., 1994). *Figure 7.2 (right): Staphylococcus aureus* alpha-hemolysin. The transmembrane domain responsible for cytolysis comprises the lower half of a 14-strand antiparallel beta barrel (Song *et al*., 1996).

The crystal structure of another ß-pore forming toxin is that of the monomeric *Staphylococcus aureus* alpha-hemolysin (Figure 7.2). This peptide is a water soluble 33 kDa component of the hemolysin hexamer/heptamer which is formed when the monomer oligomerizes upon binding to the target membrane surface (Song *et al*., 1996; Montoya and Gouaux, 2003). The toxin upon oligomerization constitutes a distinct mushroom like shape with a cap, rim and stem domains (Kaneko and Kameo, 2004). The stem makes up the transmembrane domain and is comprised of 14 ß-strands, formed from the 7 ß-hairpins, where each hairpin is contributed from a single monomer. The resulting

structure is a ß-barrel or ß-can motif. The alternating hydrophilic and hydrophobic residues in the hairpin assemble in such a way that the hydrophilic residues formed the lumen of the pore, while the hydrophobic residues communicated with the membrane to which it was bound (Song *et al*., 1996; Pédelacq *et al*., 1999). Such a structure is not only stable in the lipid bilayer, but the hydrophilic pore interior facilities the movement of water, ions and small molecules between the cell interior and the outside world.

Strategies of venom study

Study of venom components have interested scientists over the years and there have been significant changes in the way venom is studied. Common strategies of venom study comprise cell based assays using different cell lines and tandem mass spectrometry of the crude venom (Ponce *et al*., 2015*).* Some venom peptides that have been identified and studied in other venomous organisms are also constituents of the cnidarian venom repertoire (Frazao and Antunes, 2016). Indirect study of venom using deep sequencing of the mRNA transcriptome and *in silico* analyses of the venomous organism's genome can reveal components that would otherwise go unnoticed using more traditional methods. Polymerase chain reaction and DNA sequence analysis, which are common techniques in molecular biology, can be employed to study the venom gland components at their nucleotide level.

Although no transcriptome data is publicly available for *Chrysaora quinquecirrha* venom yet, a recent study identified a total of 163 proteins in the

venom proteome of *Chrysaora fuscescens,* a scyphozoan from the same family, using an integrated transcriptomic and proteomic approaches (Ponce *et al.*, 2016). Of the total 163 proteins identified, 27 were classified as potential toxins. The toxins were broadly categorized into 6 protein families namely proteinases, venom allergens, C-type lectins, pore-forming toxins, glycoside hydrolases, and enzyme inhibitors. Some potential toxins comprising of proteinases, lipases, and deoxyribonuclease were identified in the transcriptome data, but not in the proteome data (Ponce *et al*., 2016). All components were not identified using proteomic techniques because some components are likely present in the venom at very low concentrations. This suggests that a venom gland transcriptomic analysis can offer a comprehensive solution in the identification of all protein content and toxin-like peptides in the venom glands. The obvious problem with scyphozoans is that jellyfish do not have a centralized venom gland like other venomous organisms. Their venom is partitioned into myriad small organelles distributed throughout their tentacles. So, finding a localized cache of cells synthesizing and accumulating only venom components is impossible. But by understanding how cnidocytes develop, and the developmental pathway of cnidocysts in these cells, might permit us to discern the import pathway and, thus, identify venom genes by a computational approach.

Availability of transcriptome data will enable creation of cDNA libraries which may enhance our ability to identify individual venom components. Using DNA sequencing in conjunction with the bioinformatic approaches, functional linkages between fully sequenced genomes and their expressed protein products

can be established. There is a great deal of data on the DNA sequences of many venomous organisms, therefore available sequence data could be used to answer our biological questions of interest.

Research **Objectives**

Previous research has demonstrated that the venom of *Chrysaora quinquecirrha* contains a hemolytic protein capable of lysing erythrocytes, although the specific gene and protein have not been identified. Accordingly, this research project has 3 specific objectives: 1. Using data initially gleaned from an RNA-seq library generated from sea nettle tentacle RNA, to isolate and sequence the full length genomic clone of the hemolytic lectin gene of *Chrysaora quinquecirrha*; 2). to compare the sequence and structure of this gene and putative protein from populations isolated from both Barnegat Bay, NJ and Chesapeake Bay, MD; and 3). to construct pET-SUMO plasmids containing both the full-length gene of this hemolytic lectin as well as a sub-domain encoding the pore-forming domain of the gene for future expression studies in *E. coli*.

Materials and Methods

1. C*hrysaora quinquecirrha* Transcriptome by RNA-Seq Analysis

1.1 Isolation of Total RNA from Chrysaora quinquecirrha.

Total RNA was isolated from the tentacles of a single medusa collected from the Cattus Island region of Barnegat Bay (collected August 10, 2013). The jellyfish was transported back to the laboratory in Montclair, NJ and washed several times in sterile artificial seawater (19 ppt). The individual was kept alive for 2 days to allow time for all gut contents to be expelled. It was then rinsed again with artificial seawater to remove any other (non-jellyfish) DNA/RNA. Tentacles were frozen in liquid nitrogen then ground to a fine powder with a homogenizer. Total RNA was isolated using the Qiagen RNAeasy Plus MicroKit (Cat No./ID: 74034) following the manufacturer's instructions.

1.2 Preparation of NGS Library

Library preparation was performed by GeneWiz, Inc. (South Plainfield, NJ) and included separating out poly A+ RNA (to eliminate or minimize the inclusion of rRNA and tRNA), construction of a cDNA (complementary DNA) library by reverse transcription, and shearing of cDNAs to produce fragments ranging from 100 to 200 bp in length. Ends of dsDNA were repaired and adaptors ligated to ends to permit multiplexing of samples.

1.3 NGS Sequencing

DNA was sequenced on an Illumina HiSeq 2500 platform using 2 x 100 paired ends. Approximately 380,000,000 reads were generated from this run from triplicate samples.

1.4 Contig Assembly

Raw sequence data were processed by eliminating sequences with low quality scores, removal of adaptor sequences, and then assembling using CLC Workbench to generate a file of 87,600 contigs (JG01-CQTTotalRNA-Contigs.fasta). The data were organized as a series of fasta files, with the first line indicating the contig number and the approximate coverage of the assembled sequence.

1.5 BLAST Search

This file of assembled contigs was BLASTed against the nr database of Genbank (this is the complete Genbank collection of all known sequences) and the best hit (highest score match or lowest E or Expect value) was recorded in a second file (rna.nr.best.hit.complete.xlsx). Subsequently, these BLAST hits were cross-indexed to the UniProt Venom Database (http://www.uniprot.org/program/Toxins) to identify putative genes encoding venom proteins in *Chrysaora quinquecirrha*.

2. Genomic DNA analysis

2.1 Sample Preparation

Adult sea nettle medusa were collected (from both the Tom's River in Barnegat Bay, NJ and from St. Mary's River in Chesapeake Bay, MD) and transported to the laboratory in 70% (v/v) ethanol for genomic DNA analysis. The tube was centrifuged at 13,200 rpm for 2 minutes and the supernatant was decanted. The jellyfish was then transferred to a petri dish and cut into pieces small enough to fit into a 1.5 mL centrifuge tube. These samples were spun in a Speed Vac until all visible solvent had been removed. The sample was then stored at -80ºC.

2.2 *Extracting Jellyfish DNA*

2.2.1 Sample Homogenization in Buffer

CTAB (hexadecyltrimethylammonium bromide) isolation buffer with NaCl was used to isolate high molecular weight DNA from jellyfish using a modified Winnepennickx *et al*. 1993 protocol and as described by Gaynor *et al*., 2016. Typically, extractions were done in 1.5 mL Eppendorf tubes. Jellyfish tissue was introduced into tubes containing 500 µL of CTAB isolation buffer and ground using separate micro pestles into a homogenous slurry. The tubes were incubated at 60ºC for 60 minutes while mixing them occasionally by inverting the tubes. After incubation was completed, 0.5 mL of chloroform:isoamyl alcohol (24:1) mixture was added to each tube and gently mixed for 2 minutes by inverting the tubes.

2.2.2 Centrifugation and Washing

The tubes were centrifuged for 10 minutes at maximum speed (14,000 x g) in a micro centrifuge at 4 °C. The upper aqueous phase was transferred into new 1.5 mL tube, while discarding the solid cellular debris. RNase A (1 µL of 10 mg/mL) was added to each tube containing the supernatant and incubated for 30 minutes at 37°C. Isopropanol (2/3 volume) was added to each tube, capped and gently inverted to mix. The tubes were allowed to sit at room temperature (20ºC) for 2 hours. The tubes were then centrifuged for 15 minutes at 14,000 x g at 4 °C to pellet the DNA. The supernatant was carefully removed, followed by 2X washings of the pellet with 500 µL of 70% ethanol. Between the washes, the tubes were briefly re-centrifuged at 14,000 x g at 4 °C to pellet the DNA. After the second wash, the pellets were dried briefly (5 minutes, or depending on remaining liquid volume) in a Speed-Vac without heating.

2.2.3 Resuspension of DNA

The DNA was resuspended in a minimum volume (50 µL) of TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) buffer. Concentration and purity of the DNA were determined by UV absorption using a NanoDrop ND 1000. An aliquot of the isolated DNA product (10 µL) was run on a 1.0% (w/v) agarose gel containing SYBR-Safe to check for quality and size of DNA. Isolated DNA was aliquoted into labeled 1.5 mL tubes and stored at -20 ºC.

3. PCR preparation and amplification of the Hemolytic lectin gene

3.1 Primer Design

The primers were designed to amplify the full-length and truncated versions of the hemolytic lectin gene from *Chrysaora quinquecirrha*. The region between the start and stop codon were selected for primer design. PrimerQuest Tool from the Integrated DNA Technologies (http://www.idtdna.com/PrimerQuest/Home/Index) was used to generate 9 sets of primers. When designing the primers, the melting temperature (Tm) of the primers were within 5°C of each other. During the design process, intra- and inter-primer homologies were avoided to prevent self-dimers or primer-dimer formation instead of annealing to the desired DNA sequences. The primers designed based on the RNA transcriptome data is listed in Figure 10. Additionally, other primers used for sequence analysis are listed in Appendix B.

3.2 PCR preparation

Choice-Taq™ DNA Polymerase manufactured by Denville Scientific Inc (Denville Scientific, Denville, NJ; http://www.denvillescientific.com) (10 µL) was pipetted into each labeled, sterile 200 µL polypropylene thin-wall PCR tubes. The Choice-Taq™ DNA Polymerase was supplied with 10X PCR reaction buffer, containing 15 mM $MgCl_2$, 100 mM KCl, 80 mM $(NH4)_2SO_4$, 100 mM Tris-HCl, pH 9.0, 0.5% NP-40. The buffer produced a final Mg^{2+} concentration of 1.5 mM. To every reaction tube containing 10 µL Choice-Taq™, 1 µL each of forward and reverse primer (10 µM stocks), 1 µL of the DNA sample and 7 µL of sterile

deionized water was pipetted in to obtain a final reaction volume of 20 µL. See Appendix B for a complete list of primers used. In each run, a negative control (NTC, or no template control) received 8 µL of sterile deionized water, but no DNA.

3.3 Cycling parameters

PCR tubes containing 20 µL of reaction mixture were inserted into the Applied Biosystems ProFlex™ 3 x 32-well PCR System thermocycler. Routinely, initial denaturation was carried out one time at 95ºC for 1 minute. Then, amplification cycles comprised denaturation, annealing and extension phases were run for 30 cycles at 94 ºC for 20 seconds, 55 ºC for 20 seconds, and 72ºC for 1 minute, respectively. After the 30 amplification cycles, a final extension was run at 72ºC for 10 minutes. The samples were held at 4ºC until they were removed from the thermocycler.

4. Agarose Gel Electrophoresis

4.1 Gel Preparation

Agarose gel electrophoresis was used as a standard method to assess both purity and size of amplicons generated by PCR. To prepare a 1.0% (w/v) agarose gel 0.40 g of agarose was mixed with 40 mL 1X TAE buffer (40 mM Tris - Acetate, 1 mM EDTA) in a 250 mL Erlenmeyer flask. To avoid any spillage and over boiling, the agar was boiled in a microwave at low power (3) for 3 minutes and 15 seconds. It was ensured that the agar had fully dissolved to a clear solution. The solution was allowed to cool briefly. While the solution was still a liquid, 4 µL of Invitrogen™

SYBR™ Safe DNA Gel Stain (10,000X in DMSO, Invitrogen) was added to the flask and gently swirled avoiding any bubble formation. A casting tray with 10 well comb was readied in the gel box. The agar was gently poured into the casting tray and allowed to harden for about 15 minutes. When the gel hardened, the comb was firmly removed and oriented with the wells on anode. Then, the running buffer 1X TAE was poured into the gel box until gel was submerged about 5 millimeters.

4.2 Preparing Samples for Gel Electrophoresis

2 µL of 6X loading dye (0.25% (w/v) Bromophenol Blue, 0.25% (w/v) Xylene Cyanol, 30% (v/v) Glycerol) was pipetted into each labeled 1.5 mL Eppendorf tube. Of the 20 µL PCR reactions, only 10 µL of PCR amplicon was pipetted into each tube and mixed with the loading dye (remaining sample was stored at –20ºC for subsequent DNA sequence analysis or cloning). The mixture was centrifuged briefly. Ten µL of HiLo DNA ladder (HiLo DNA Ladder, Minnesota Molecular; http://www.mnmolecular.com) was pipetted into the end wells flanking the samples. Each sample (12 µL) was carefully pipetted into individual wells. The gel box was covered and plugged into the power supply and run at 105 volts for 45 minutes. Gel running time was adjusted based on the length of the PCR amplicon.

4.3 Digital Gel Imaging

The gels were visualized in a Kodak Imager System (GL100) under blue light (470 nm). Gel image was taken and saved on the Kodak 1D image analysis

software (v 3.6). The bands observed on the gel were compared with the HiLo DNA ladder to determine DNA length.

5. Automated Sanger Dideoxy Sequencing

Amplicons deduced to be both clean (*i.e.,* a single band) and of sufficient quantity (based on intensity of the band) were processed for DNA sequence analysis. If there was more than 1 µg of an amplicon in a lane, then that sample was diluted accordingly for DNA sequence analysis (typically between 10- and 100-fold depending on intensity of the band). All dilutions were done with sterile deionized water. Samples submitted for sequencing contained 1 µL of amplicon (or diluted amplicon), 1 µL of forward or reverse primer (10 µM stock), and 8 µL of sterile deionized water. Samples were sequenced in both the forward and reverse directions. Sequencing was performed using the BigDye Terminator Cycle Sequencing Kit Version 3.1 (Applied Biosystems Inc., Foster City, CA 94404; https://tools.thermofisher.com/content/sfs/manuals/cms_081527.pdf) following the manufacturer's instructions with the exception that we routinely ran 1/16 reactions. Cleanup was performed using an EdgeBio Performa DTR Gel Filtration Cartridges (Gaithersburg, MD; https://www.edgebio.com). The samples were analyzed using an ABI3130 Genetic Analyzer from Applied Biosystems (Foster City, CA) using a 36-cm column array and NANO POP7 polymer (MCLAB, South San Francisco, CA 94080, NP7-100; http://www.mclab.com). Sequence calls were made using the KB basecaller.

6. Analyzing DNA sequences

Sequence data was generated as ABI chromatogram files. The ABI file format is a binary file produced by ABI sequencer software. Macintosh compatible software applications 4Peaks (http://4peaks.en.softonic.com/mac) and Geneious (http://www.geneious.com/) were used to trim, assemble and view Sanger sequencing trace files, correct base calls and create consensus sequences. All sequencing data generated locally was assembled both *de novo* and by comparing with a reference sequence generated by RNA transcriptome analysis. Sequences were searched against Genbank using the BLASTn algorithm. BLAST2Seq algorithm was also used to produce alignments between overlapping sequences and to help resolve inconsistencies between forward and reverse sequencing reads. BLASTn searches, unless otherwise specified, were done using standard default values. A match with an e (expect) value of $< 10^{-4}$ was considered a match.

7. Gene Alignment and Assembly

Sequencing data was analyzed both manually and with the help of Geneious sequence analysis software for *de novo* gene assembly. For manual assembly, sequencing data was extracted from the chromatogram files and copied into a word processor. A short segment of the nucleotide was selected from a sequence and searched for relative abundance. Sequences with similarities were aligned to create a scaffold and other sequences were built around it by joining overlapping fragments. A consensus sequence generated by the manual assembly of the sequence data was aligned to the mRNA seq data for validation.

The built-in Geneious assembler, Tadpole, SPAdes, Velvet, MIRA and CAP3 assemblers available in the Geneious sequence analysis software were also used to validate the assembled sequences.

8. Cloning into the pMiniT vector

8.1 Ligation Reaction

NEB PCR Cloning Kit (NEB #E1202) was used to clone the full-length gene into the pMiniT vector. The insert DNA was mixed with 1 µL linearized vector and brought to a 5 µL volume by adding with sterile deionized water. Then 4 µL of Cloning Mix 1 and 1 µL of Cloning Mix 2 were added to the mixture and incubated at room temperature (25°C) for 15 minutes followed by incubation on ice for 2 minutes. The ligation reaction (2 µL) was then mixed with 50 µL NEB 10-beta Competent *E. coli* (NEB #C3019) and incubated on ice for 20 minutes. The cells were then heat shocked at 42°C for 30 seconds and immediately transferred to ice, where they were incubated for 5 minutes. S.O.C (Super Optimal broth with Catabolite repression) medium (950 µL) was added to the transformants and incubated at 37ºC for an hour. The transformants (50 µL and 50 µL of a 1:10 dilution) were spread on 37°C pre-warmed LB selection plate containing 100 µg/mL ampicillin (LB-AMP) and incubated overnight at 37ºC.

8.2 Insert Screening

Screening for inserts was carried out by colony PCR, restriction enzyme digestion, PCR and sequencing of the isolated plasmid DNA.

8.2.1 Colony PCR

Individual *E. coli* colonies were picked from the LB-AMP plates using a sterile inoculation loop and subcultured into new gridded LB-AMP plates. A loopful of colonies generated by overnight subculture was then introduced into a PCR tube containing 100 µL of 5% (w/v) Chelex prepared in 100 mM of Tris Buffer (pH 11). The tubes were vortexed briefly to mix the bacteria in Chelex and boiled in a heat block for 10 mins at 100 ºC. Tubes were spun at 13,200 rpm for 2 minutes. Twenty µL of the supernatant was removed and 1µL was used for colony PCR (using 1µL of pMiniT F and pMiniT R primers (see Appendix B)). Amplification was carried out using manufacturer's directions. The PCR amplicons were electrophoresed on agarose gels (see Materials and Methods Section 4). Bands visible on the gel were screened and prepared for DNA sequencing analysis as explained in section 5 of Materials and Methods. Sequence data was used to determine the direction and correct reading frame of the insert. Colonies with positive inserts were sub-cultured on fresh LB-AMP plates and stored in 4ºC for further use. A glycerol stock of bacteria was also prepared for each clone by mixing 15µL glycerol and 85µL cells in broth for long term storage at -80ºC. Positive colonies were also grown in LB broth with 100 µg/mL ampicillin for mini-prep of the plasmid DNA (see Appendix C). Plasmid DNA was extracted using the QIAGEN® Plasmid Purification kit (Catalog number: 12123) using the manufacturer’s protocol. The extracted plasmid was stored at -20ºC until further use.

9. Cloning into the pET SUMO expression vector

9.1 Ligation Reaction

The Champion™ pET SUMO Expression System by Life Technologies (Catalog # K30001) was used for cloning and expression of our gene. A 1:1 molar ratio of vector: insert was recommended by the manufacturer for optimum ligation efficiency. One µL of PCR sample was determined to be optimal for ligation. In a 200 µL tube, 10 µL ligation reaction was carried out with 1µL of the fresh PCR product, 1 µL of the 10X Ligation buffer, 2 µL of the 25 ng/µL pET SUMO vector, 5 µL of deionized water and 1 µL of the 4.0 Weiss units T4 DNA Ligase. The reaction was incubated at room temperature (20ºC) for 30 minutes for the truncated (444 bp) insert and incubated overnight at 15ºC in a thermocycler for the full length (1299 bp) insert.

9.2 Transforming One Shot® Mach1™-T1^R Competent Cells

Before transformation was carried out, the S.O.C medium (Invitrogen, Catalog No. 15544-034) was equilibrated to room temperature. LB plates containing 50 µg/mL Kanamycin (LB-KAN) were warmed in an incubator at 37ºC. The One Shot® Mach1™-T1^R Competent Cells were removed from -80ºC freezer and thawed on ice. Each ligation product (2 µL) was pipetted into individual vials containing One Shot® Mach1™-T1^R Competent *E. coli* (Invitrogen, Cat. No. C8620-03). The cells and ligation mixture were mixed by gently flicking the walls of the tube and incubated on ice for 15 minutes. A water bath at 42ºC was prepared to heat shock the cells for 30 seconds. Vials containing the cells were immediately

transferred to ice. Room temperature S.O.C medium (250 µL) was added to the tube containing cells. The tubes were tightly capped and horizontally shaken in a 37ºC shaking incubator for 1 hour at 200 rpm. Two separate volumes, 100 µL and 200 µL, of the transformants were aseptically spread on a pre-warmed LB-KAN plates and incubated overnight at 37ºC.

9.3 Screening for Inserts

Screening for the correct insert was carried out by colony PCR, restriction enzyme digestion, PCR and sequencing of the isolated plasmid DNA (as described in section 8.2 of the Materials and Methods). Only those clones that contained an insert in the 5' to 3' direction and correct reading frames were sub-cultured into fresh LB-KAN plates and LB-KAN broth for plasmid mini-prep. See Appendix C for the Alkaline Lysis plasmid mini-prep protocol.

10. Transforming BL21(DE3) One Shot® Cells for Expression

10.1 Ligation and Transformation

BL21(DE3) One Shot® cells (Invitrogen, Cat. No. C6000-03) were removed from -80ºC freezer and quickly thawed on ice. The volume of the plasmid DNA was adjusted so the final concentration was 10 ng/µL. Plasmid DNA (1 µL) was added to each vial of BL21(DE3) One Shot® cells and stirred gently with the pipette tip. The mixture was incubated on ice for 30 minutes. The tubes containing the ligation mixture was then introduced to a 42ºC water bath to heat-shock the cells for 30 seconds. Tubes were then immediately transferred to ice. To the mixture,

250 µL of room temperature S.O.C. medium was added and incubated at 37°C for 1 hour with shaking at 200 rpm.

The entire transformation reaction was added to 10 mL of LB broth containing 50 µg/mL Kanamycin. The culture was grown overnight at 37°C with shaking at 200 rpm.

10.2 IPTG Induction and Sample Processing

An aliquot (500 µL) of the overnight culture was inoculated into 10 mL of LB broth containing 50 µg/mL kanamycin (in a 125-mL Erlenmeyer flask) and allowed to grow for two hours at 37°C with shaking at 200 rpm. The optical density (OD_{600}) reading was taken at the end of the two hours. The sub-culture was then split into two 5 mL cultures. To one of the 5 mL cultures, IPTG was added to a final concentration of 1 mM to induce protein expression, while the other culture was left uninduced. From each culture, 500 µL aliquot was taken immediately and centrifuged at maximum speed (13,200 rpm) in a microcentrifuge for 1 minute. The supernatant was aspirated and the cell pellets were immediately frozen at -20°C. These were marked as the zero-time point samples. The remaining cultures were incubated continuously at 37°C with shaking at 200 rpm. Time points for each culture was taken every hour for 4 hours. For each time point, 500 µL of both the induced and uninduced cultures were taken and processed as described above.

10.3 Sample Analysis

A Mini-PROTEAN® TGX™ Precast Gel (4-15%), purchased from Bio-Rad Laboratories, Inc., was used to analyze the proteins in our expressed samples. Frozen zero-time point, induced and uninduced samples were thawed and resuspended in 80 µL of 1X Laemmli sample buffer (60 mM Tris-Cl pH 6.8, 2% SDS, 10% glycerol, 5% β-mercaptoethanol, 0.01% bromophenol blue) containing 2-mercaptoethanol. Samples were boiled for 10 minutes in a water bath and centrifuged at 13,200 rpm for 2 minutes to remove any insoluble material. Precision Plus Protein™ Dual Color Standards (161-0374, Bio-Rad) (10 µL) and samples (10 µL) were loaded on an SDS-PAGE gel and electrophoresed in a 1X Tris/Glycine/SDS buffer for 50 minutes at 170V. Remaining samples were stored at -20ºC. Upon completion of electrophoresis, the gel was transferred to a tray containing destaining solution (40% methanol and 10% glacial acetic acid). After the initial fixing of the gel for 2 minutes destaining solution was poured off, then the gel was stained with Coomassie Blue stain (0.1% Coomassie Brilliant Blue R-250, 50% methanol and 10% glacial acetic acid) for 30 minutes on a rocking platform. The gel was destained with a de-staining solution containing methanol, water and glacial acetic acid for 3 hours on a rocking platform. The destaining solution was changed every 30 minutes. The bands on the gel was visualized by placing the gel on a white light box and photographed for band size analysis.

Results

Analysis of an RNA-Seq library created from the tentacles of a mature *Chrysaora quinquecirrha* isolated from Barnegat Bay (Gaynor, Meredith, and Shchegolev, unpublished data) suggested the presence of a hemolytic lectin. Contig 22835 (out of 87,600 total contigs generated in this library), with an average coverage of 45.40X, had a strong match to the hemolytic lectin S-1 from *Cucumaria echinata*, a sea cucumber native to the Indo-Pacific ocean. This match was generated using the BLASTX algorithm (Altschul *et al*., 1997), using the standard default settings, and demonstrated a 62.17% identity, an alignment length of 454 AA (292 matches, 168 mismatches) in one continuous reading frame, and an e (expect) value of 0 and a bit score of 573. The annotated RNA-seq sequence is shown in Figure 8.

As can be seen in Figure 8, the contig is 1645 nt in length, and contains an ATG start codon at position 85, an in-frame TGA stop codon at position 1447, and a downstream polyadenylation consensus signal - AATAAA - or Proudfoot Box (Proudfoot and Brownlee, 1976), starting at position 1519. The polyadenylated tail (which is 63 nts long in this contig) starts 57 nts downstream from the 3' end of the polyadenylation consensus signal, consistent with reports from other eukaryotic genes (Chang *et al*., 2014). The single open reading frame (ORF) in this contig predicts a protein with a length of 454 AA.

Bioinformatic analysis of the predicted protein using the SignalP algorithm demonstrates a potential signal peptide of 22 AA. This program calculates three different scores (C-, S-, and Y-) to predict putative signal peptide and, in this case,

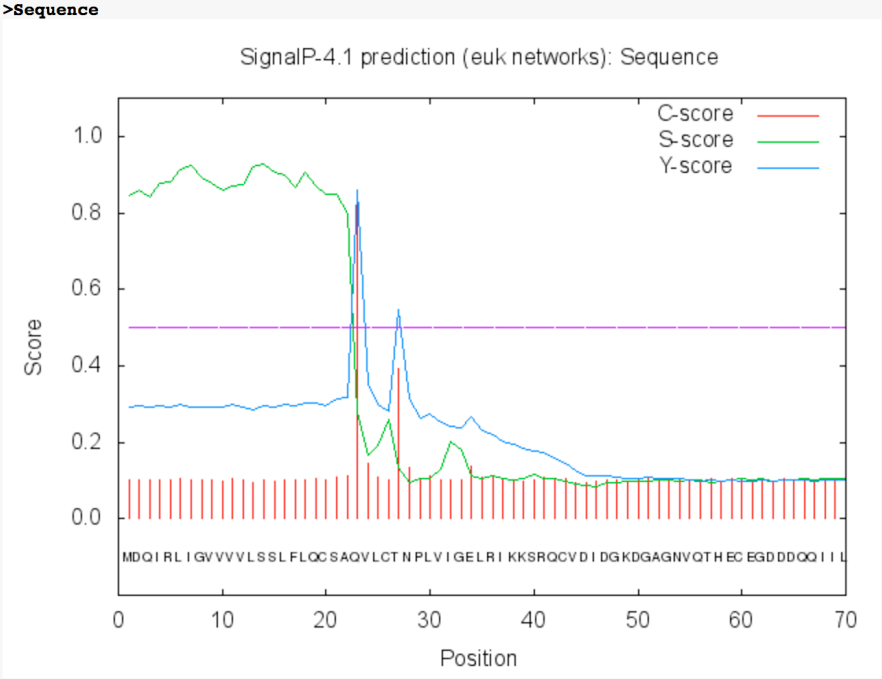
AAAGCAACTGAGTGCAGTGAAGCTGTAGCAGTTGCACATAGGCAAAAGCT
CGAAGAACGCTTGC**AGAAGTTAAGGTTGATAAAAC**ATGGATCAAATACGC
TTGATTGGTGTGATCGTTGTACTTTCGTCATTGTTTTTGCAATGCTCTGC
TCAAGTCCTGTGCACCAATCCGTTGGTAATTGGAGAGCTTCGAATCAAGA
AGTCAAGACAATGTGTTGACATTGATGGAAAAGACGGAGCTGGAAATGTG
CAGACACATGAATGTGAAGGAGATGACGATCAACAAATCATCCTATGTGG
TGATGGCACAATTCGCAACGAGGCTAGAAATTACTGCTTCACACCACGTG
GCAGTGGCAACGACAATGTTGAATCGTCAGCCTGTCAGCATTACCCAAGA
ATTCCTACAAGACAGAAGTGGAGACTTGGAAGGTCAAAGAAATTCTATGA
CATGGGAGGAATCTTACAGGAAGCAAGAGAAATCATCAACGTTGAATCAA
ATAGATGCCTTGATGTTAGTGGCTACGATGGAACTGGCAACATTGGCGTG
TATCATTGCGAAAACAAAGATGACCAGTACTTTTATTTCCGATCAAGAGG
AAAAGAAGTCGCTTTCGGGAGGCTCAGGAATGAGAAATCAAGTCAATGCC
TTGATGTCAGTGGGTATGATGGCAAAGGAAATGTACAAATGTACGACTGT
GAAGATAAGAAGGACCAATGGTTTAAATTTTATGAGAATGGAGAAGTCGT
CAATGAGCAGTCAAGACGTTGTTTGGACGTATCTGGCTATGATGGAACAG
GCAACATTGGTACATATTGCTGTGAAGACAAGCATGATCAGATGTGGTCT
CGACCATCTCAGCTTTGCAACGGCGAATCGTGTTCTTTTGTCAACAAAAA
ATCAGGCCAATGTCTGGATGTGTCAGGATACGATGGACGAGGCGGTGTGG
CTACCTATCATTGTGAAGGACTTGCTGATCAACGACTGAAATGGGTGACT
GACAAATGGACAGCTCCTAATGCTGTTTGGGTGATGGTTGGCTGCAATCA
AAACGGAAAGGTTTCTCAGTGGCTTTCCAACACTGTTTCATATTCATCTA
CAATTACACACACTGTCACTGTTGAAGTTGGTGCATCCATGGAAGCAGAT
CTTGTGTTTGCAAAAGCAACAGTGTCAACCAAAGTTTCTACATCACTTTC
AACTGCCTGGACCAAGAGCCAGAGTGGAACAACTCGTATCGTCTTCACCT
GTGAGTATTACGACAACCAGGAAGCATTTACAGGAGGATGCATGTGGCAG
CTTCGGGTCGACACCAAGCATGTCAACTCTGGCCGTCTACTTACATGGAG
TCCACAGATCACGAGGTGCACAACGTCAAACACCCAGCCAAGATGCCCAC
CGTTCACAAAATGTGTCGATAAGGCCTGTTCTCTTTGCCAAGAAAT**CTGA**
CATTAATTGCTGCCGTTTCTCTTTCTTTTAACTGCTTGTTTTACTTTTGA
CTTTGATTAAACATGATTCAATAAAAATATCTGCTGTTGCTGCTACTTTA
ATAACAAATATAATTATTATTGATGAACTTGCAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Figure 8: Contig 22835 from RNA-Seq dataset. This sequence generated by the mRNA seq contains 1645 bases. Sequence in bold and underlined type-face are the forward and reverse primers. Initially identified sequence features are highlighted in green (start codon), red (stop codon) and blue (poly-A tail).

there was excellent agreement of the methods, indicating that the cleavage of the nascent protein likely occurs between the 22nd (alanine - A) and 23rd (glutamine

- Q) amino acids (Figure 9). Thus, when cleaved this produces a mature protein of 438 AA with an N-terminus of Q instead of M.

```
# SignalP-4.1 euk predictions
>Sequence
```




```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    23        0.838
  max. Y    23        0.858
  max. S    14        0.927
  mean S     1-22     0.878
       D     1-22     0.869   0.450    YES
Name=Sequence   SP='YES' Cleavage site between pos. 22 and 23: CSA-QV D=0.869 D-cutoff=0.450 Networks=SignalP-noTM
# data
# gnuplot script
```

Figure 9: The SignalP 4.1 algorithm (http://www.cbs.dtu.dk/services/SignalP/) predicts a signal peptide of 22

Contig 22835 was used as a scaffold to generate 9 unique sets of primers (Figure 10) using to the Primer Quest Tool (https://www.idtdna.com/Primerquest/Home/Index) of Integrated DNA Technologies (https://www.idtdna.com/site). As seen in Figure 10, each primer was assigned a unique identifying letter, with start (5') and end (3') position indicated relative to the

start codon (ATG) of the gene, plus the calculated Tm value (in ºC). Forward and reverse primers can be distinguished by whether they had ascending or descending start and end positions relative to the ATG start codon. Generally, the strategy used to verify the assembled RNA-Seq contig was to use multiple combinations of forward and reverse primers in PCR reactions to generate specific amplicons. Agarose gel electrophoresis was used to verify that a clean amplicon of the expected size was generated. Then the amplicon was subjected to Sanger dideoxy sequencing of both strands using the forward and reverse primers (same ones used to generate that specific amplicon), respectively. These primers , taken together, bridged the full length of the ORF (minus the poly A tail).

All primers were mapped (Figure 11) to the full length gene using SnapGene Viewer (www.snapgene.com) and Geneious sequence alignment software (www.geneious.com). The DNA sequence of the putative gene is annotated with restriction enzymes and primers at their respective binding sites. As shown in Figure 11.1, primers M, C, K, O, E, I, HemoF, G, and A are forward primers (dark green arrows), while D, HemoR, B, H, F, L, N, P, and J are reverse primers (light green arrows). A corresponding restriction map is seen in Figure 11.1 (restriction sites and positions in black) along with primer order and nucleotide positions (in purple). The Geneious software was was used to test various primer pairs against the putative gene to import a primer pair map (Figure 11.2).

The various primer sets were used in combination to generate amplicons of various sizes that targeted different sections of the putative gene. All PCR reactions were verified by agarose gel electrophoresis to assess both amplicon

Primers based on the RNA Seq. Data _A of ATG is 1					
TUBE	PRIMER SEQUENCE	ID	Start	End	Tm °C
C1	CAAGACGTTGTTTGGACGTAT	A	677	699	56.9
C2	CCATTTGTCAGTCACCCATTT	B	925	903	56.7
C3	TCAAGTCCTGTGCACCAAT	C	66	86	56.9
C4	CAGTTCCATCGTAGCCACTAA	D	452	430	57.5
C5	TTCGCAACGAGGCTAGAAATT	E	227	249	58.3
C6	CAGGTGAAGACGATACGAGTT	F	1167	1145	57.5
C7	GGAGGCTCAGGAATGAGAA	G	533	553	56.1
C8	CCATCACCCAAACAGCATTA	H	953	932	56.0
C9	CGCAACGAGGCTAGAAATT	I	229	249	56.0
C10	ATCTTGGCTGGGTGTTT	J	1310	1293	53.1
C11	AAGTCCTGTGCACCAAT	K	68	86	53.2
C12	GCTTCCTGGTTGTCGTAAT	L	1191	1172	55.2
C13	ATTGGTGTGATCGTTGTACT	M	19	40	54.7
C14	CCTCCTGTAAATGCTTCCT	N	1203	1184	54.1
C15	AATCCGTTGGTAATTGGAGA	O	82	103	54.3
C16	AAGTAGACGGCCAGAGT	P	1258	1242	54.1
HemoF	CCTGTCAGCATTACCCAAGAA	HemoF	296	317	57.9
HemoR	CCTGTTCCATCATAGCCAGATAC	HemoR	717	694	58.4

Figure 10: List of primers that span the full gene. These primers were generated to verify the RNA-Seq dataset.

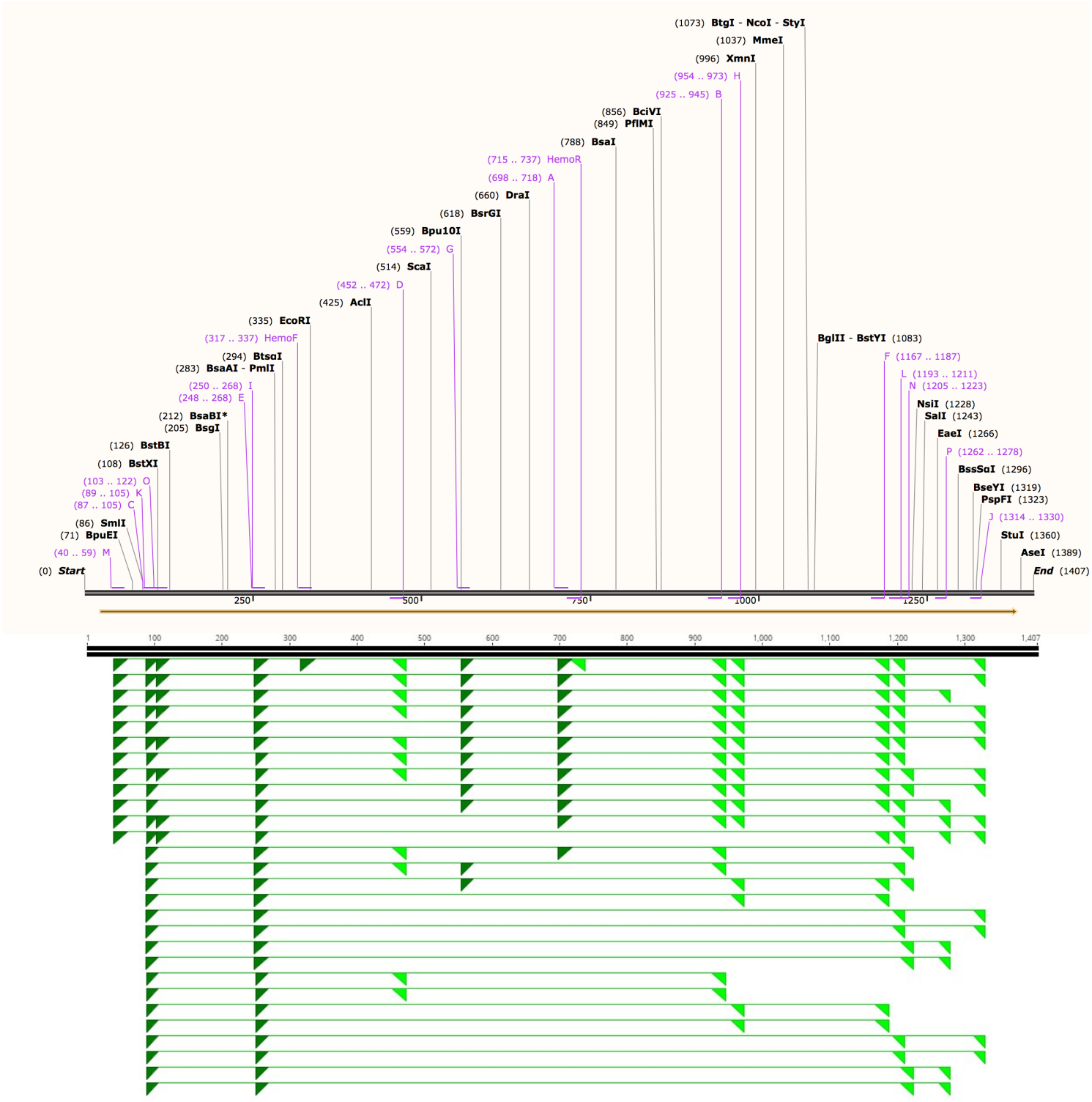


Figure 11.1 top: Primer map generated using Snapgene viewer. *Figure 11.2 bottom*: Potential PCR amplicons generated using Geneious sequence analysis software. Dark green represents forward primers and light green represents reverse primers. The primer map on the top image correspond to the amplicons on the bottom.

size and purity. A representative gel image (Figure 12) shows amplicons of different sizes generated by combinations of forward and reverse primers. Hi-Lo™ DNA Markers were used to calculate amplicon size. All amplicons that produced a clean band on an agarose gel were sequenced and edited, aligned and assembled into larger contigs using the Geneious R10 package.

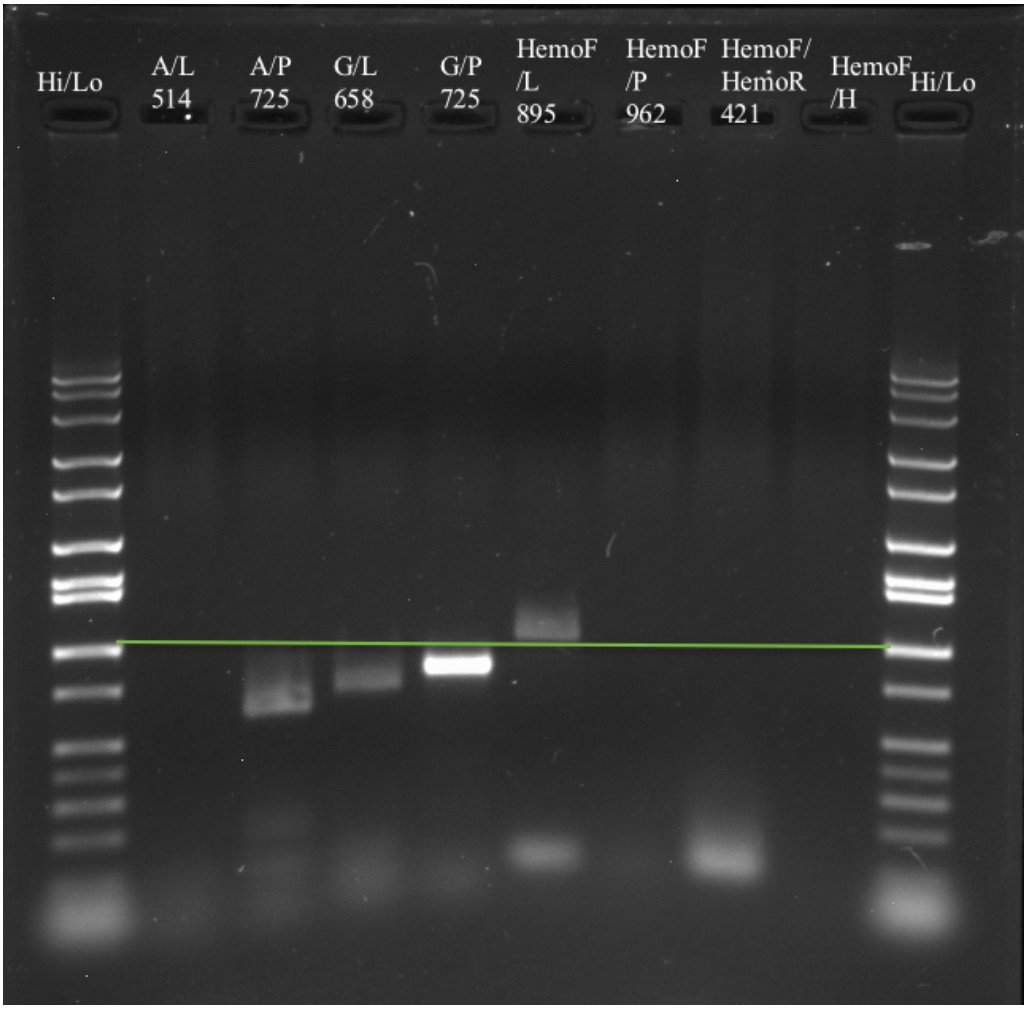


Figure 12: A representative gel image that shows different primer sets and PCR amplicons of various sizes. In the above gel image, amplification is not seen with A/L, and HemoF/P primer sets. HemoF/HemoR and HemoF/H are no template controls.

As demonstrated in Figure 12, most of the PCR amplifications were clean and robust. Typically we would see small amounts of residual primers co-migrating

with 50 to 100 bp markers. But these were not generally an issue in subsequent Sanger sequencing reactions since large molar excess of forward or reverse primer was added to sequencing reactions that greatly reduced the impact of the competing primer.

Alignment of all sequenced amplicons produced a full-length assembly as seen graphically in Figure 13. Each amplicon was sequenced (both forward and reverse strands) multiple times and there was sufficient overlap between amplicons to permit easy alignment and assembly. Figure 14 also shows a similar alignment and assembly using just 3 long amplicons that span the entire coding region of the gene. This process was repeated for amplicons generated from gDNA isolated from *C. quinquecirrha* from both the Barnegat Bay and Chesapeake Bay populations. This permitted the generation of a consensus sequence for the gene from both populations. These sequences were submitted to Genbank (https://www.ncbi.nlm.nih.gov) and are shown in Figures 15 (Barnegat Bay) and Figure 16 (Chesapeake Bay) below. Since these represented the first known sequences for venom proteins from *C. quinquecirrha*, this hemolytic lectin was renamed Chrysaoralin.

A BLASTp (Altschul *et al*., 1997) search of these putative proteins reveals interesting homologies. As seen in Figure 17, this generates Conserved Domain Database (CDD) hits to both Ricin superfamily and Ricin B lectin superfamily in the first half of the protein sequence (Marchler-Bauer *et al*., 2012). The homologies are significant with expect (e) values of 2.16 e^{-19}. In addition, putative sugar-binding domains and Q-X-W motifs (Mancheno *et al*., 2010), both characteristic of

lectins and other sugar-binding proteins, are also found repeated several times in Chrysaoralin.

Figure 13: A representative alignment of *Chrysaora quinquecirrha* genomic DNA sequencing data on Geneious sequencing analysis software. These amplicons span the full length of the gene.

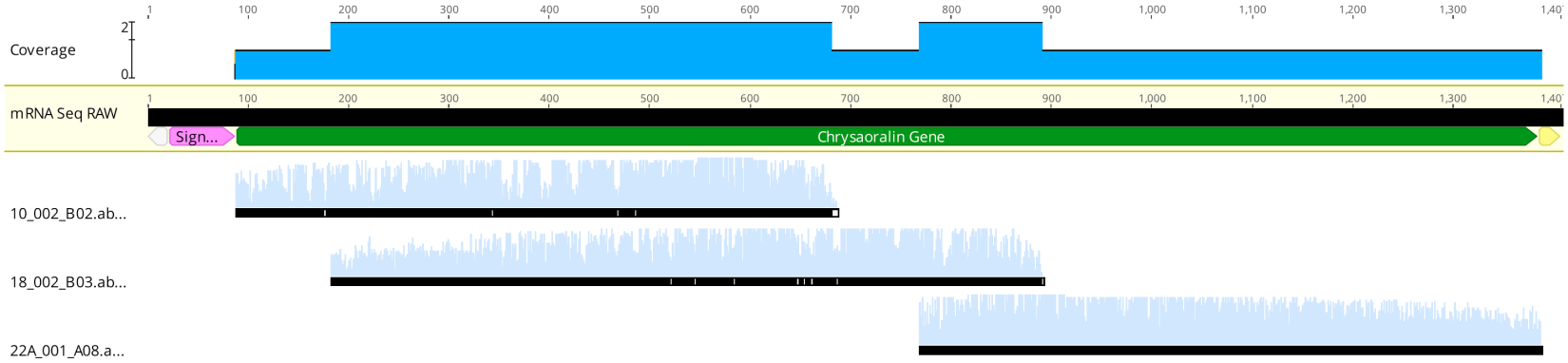


Figure 14: Sequence alignment of three amplicons generated full length Chrysaoralin gene.

Barnegat Bay Chrysaoralin

>KX656922.1 *Chrysaora quinquecirrha* chrysaoralin gene, complete cds

ATGGATCAAATACGCTTGATTGGTGTGATCGTTGTACTTTCGTCATTGTTTTTGCAATGCTCTGCTCAAG
TCCTGTGCACCAATCCGTTGGTAATTGGAGAGCTTCGAATCAAGAAGTCAAGACAATGTGTTGACATTGA
TGGAAAAGACGGAGCTGGAAATGTGCAGACACATGAATGTGAAGGAGATGACGATCAACAAATCATCCTA
TGTGGTGATGGCACAATTCGCAACGAGGCTAGAAATTACTGCTTCACACCACGTGGCAGTGGCAACGACA
ATGTTGAATCGTCAGCCTGTCAGCATTACCCAAGAATTCCTACAAGACAGAAGTGGAGACTTGGAAGGTC
AAAGAAATTCTATGACATGGGAGGAATCTTACAGGAAGCAAGAGAAATCATCAACGTTGAATCAAATAGA
TGCCTTGATGTTAGTGGCTACGATGGAACTGGCAACATTGGCGTGTATCATTGCGAAAACAAAGATGACC
AGTACTTTTATTTCCGATCAAGAGGAAAAGAAGTCGCTTTCGGGAGGCTCAGGAATGAGAAATCAAGTCA
ATGCCTTGATGTCAGTGGGTATGATGGCAAAGGAAATGTACAAATGTACGACTGTGAAGATAAGAAGGAC
CAATGGTTTAAATTTTATGAGAATGGAGAAGTCGTCAATGAGCAGTCAAGACGTTGTTTGGACGTATCTG
GCTATGATGGAACAGGCAACATTGGTACATATTGCTGTGAAGACAAGCATGATCAGATGTGGTCTCGACC
ATCTCAGCTTTGCAACGGCGAATCGTGTTCTTTTGTCAACAAAAAATCAGGCCAATGTCTGGATGTGTCA
GGATACGATGGACGAGGCGGTGTGGCTACCTATCATTGTGAAGGACTTGCTGATCAACGACTGAAATGGG
TGACTGACAAATGGACAGCTCCTAATGCTGTTTGGGTGATGGTTGGCTGCAATCAAAACGGAAAGGTTTC
TCAGTGGCTTTCCAACACTGTTTCATATTCATCTACAATTACACACACTGTCACTGTTGAAGTTGGTGCA
TCCATGGAAGCAGATCTTGTGTTTGCAAAAGCAACAGTGTCAACCAAAGTTTCTACATCACTTTCAACTG
CCTGGACCAAGAGCCAGAGTGGAACAACTCGTATCGTCTTCACCTGTGAGTATTACGACAACCAGGAAGC
ATTTACAGGAGGATGCATGTGGCAGCTTCGGGTCGACACCAAGCATGTCAACTCTGGCCGTCTACTTACA
TGGAGTCCACAGATCACGAGGTGCACAACGTCAAACACCCAGCCAAGATGCCCACCGTTCACAAAATGTG
TCGATAAGGCCTGTTCTCTTTGCCAAGAAATCTGA

```
Info:/country="USA: Barnegat Bay, Metedeconk River, Brick, NJ"
     /lat_lon="40.0502 N 74.1131 W"
     /collection_date="10-Jul-2014"
     /collected_by="John Gaynor"
Assembly Method       :: CLC Genomics Workbench v. 7.5
      Coverage        :: 45.4X
Sequencing Technology :: Illumina
                     /codon_start=1
                     /product="chrysaoralin"
                     /protein_id="AOO35153.1"
                     /translation="MDQIRLIGVIVVLSSLFLQCSAQVLCTNPLVIGELRIKKSRQCV
                     DIDGKDGAGNVQTHECEGDDDQQIILCGDGTIRNEARNYCFTPRGSGNDNVESSACQH
                     YPRIPTRQKWRLGRSKKFYDMGGILQEAREIINVESNRCLDVSGYDGTGNIGVYHCEN
                     KDDQYFYFRSRGKEVAFGRLRNEKSSQCLDVSGYDGKGNVQMYDCEDKKDQWFKFYEN
                     GEVVNEQSRRCLDVSGYDGTGNIGTYCCEDKHDQMWSRPSQLCNGESCSFVNKKSGQC
                     LDVSGYDGRGGVATYHCEGLADQRLKWVTDKWTAPNAVWVMVGCNQNGKVSQWLSNTV
                     SYSSTITHTVTVEVGASMEADLVFAKATVSTKVSTSLSTAWTKSQSGTTRIVFTCEYY
                     DNQEAFTGGCMWQLRVDTKHVNSGRLLTWSPQITRCTTSNTQPRCPPFTKCVDKACSL
                     CQEI"
```

Figure 15: GenBank entry of Chrysaoralin gene from *Chrysaora quinquecirrha* captured from Metedeconk River, Barnegat Bay, NJ. Accession number of the gene is KX656922.1.

Chesapeake Bay Chrysaoralin

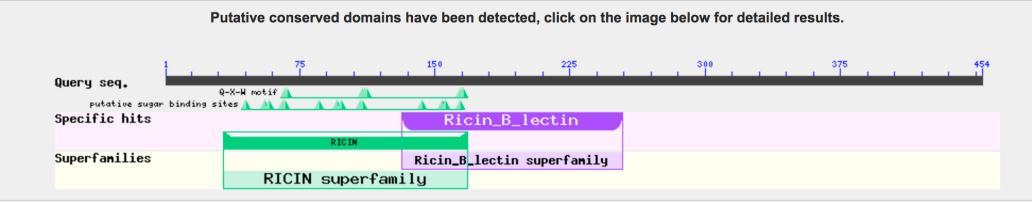
>KX356909.1 *Chrysaora quinquecirrha* chrysaoralin gene, complete cds

ATGGATCAAATACGCTTGATTGGTGTGGTCGTTGTACTTTCGTCATTGTTTTTGCAATGCTCTGCTCAAG
TCCTGTGCACCAATCCGTTGGTAATTGGAGAGCTTCGAATCAAGAAGTCAAGACAATGTGTTGACATTGA
TGGAAAAGACGGAGCTGGAAATGTGCAGACACATGAATGTGAAGGAGATGACGATCAACAAATCATCCTA
TGTGGTGATGGCACAATTCGCAACGAGGCTAGAAATTACTGCTTCACACCACGTGGCAGTGGCAACGACA
ATGTTGAATCGTCAGCCTGTCAGCATTACCCAAGAATTCCTGCAAGACAGAAGTGGAGACTTGGAAGGTC
AAAGAAATTCTATGACATGGGAGGAATCTTACAGGAAGCAAGAGAAATCATCAACGTTGAATCAAATAGA
TGCCTTGATGTTAGTGGCTACGATGGAACTGGCAACATTGGCGTGTATCATTGCGAAAACAAAGATGACC
AGTACTTTTACTTCCGATCAAGAGGAAAAGAAGTGGCTTTCGGGAGGCTCAGGAATGAGAAGTCAAGTCA
ATGTCTTGATGTCAGTGGGTATGATGGCAAAGGAAATGTACAAATGTACGACTGTGAAGATAAGAAGGAC
CAATGGTTTAAATTTTATGAGAATGGAGAAGTCGTCAATGAGCAGTCAAGACGTTGTTTGGACGTATCTG
GCTATGATGGAACAGGCAACATTGGTACATATTGCTGTGAAGACAAGCATGATCAGATGTGGTCTCGACC
ATCTCAGCTTTGCAACGGCGAATCGTGTTCTTTTGTCAACAAAAAATCAGGCCAATGTCTGGATGTGTCA
GGATACGATGGACGAGGCGGTGTGGCTACCTATCATTGTGAAGGACTTGCTGATCAACGACTGAAATGGG
TGACTGACAAATGGACAGCTCCTAATGCTGTTTGGGTGATGGTTGGCTGCAATCAAAACGGAAAGGTTTC
TCAGTGGCTTTCCAACACTGTTTCATATTCATCTACAATTACACACACTGTCACTGTTGAAGTTGGTGCA
TCCATGGAAGCAGATCTTGTGTTTGCAAAAGCAACAGTGTCAACCAAAGTTTCTACATCACTTTCAACTG
CCTGGACCAAGAGCCAGAGTGGAACAACTCGTATCGTCTTCACCTGTGAGTATTACGACAACCAGGAAGC
ATTTACAGGAGGATGCATGTGGCAGCTTCGGGTCGACACCAAGCATGTCAACTCTGGCCGTCTACTTACA
TGGAGTCCACAGATCACGAGGTGCACAACGTCAAACACCCAGCCAAGATGCCCACCGTTCACAAAATGTG
TCGATAAGGCCTGTTCTCTTTGCCAAGAAATCTGA

Info: /country="USA: Chesapeake Bay, St, Mary's River, MD"
/lat_lon="38.1326 N 76.4501 W"
/collection_date="10-Oct-2012"
/PCR_primers="fwd_name: chrysf, fwd_seq:atggatcaaatacgcttgattggtg,
rev_name: cqr, rev_seq:gagaaacggcagcaattaatgtcag"

Sequencing Technology :: Sanger dideoxy sequencing
/codon_start=1
/product="chrysaoralin"
/protein_id="AOO35152.1"
/translation="MDQIRLIGVVVVLSSLFLQCSAQVLCTNPLVIGELRIKKSRQCV
DIDGKDGAGNVQTHECEGDDDQQIILCGDGTIRNEARNYCFTPRGSGNDNVESSACQH
YPRIPARQKWRLGRSKKFYDMGGILQEAREIINVESNRCLDVSGYDGTGNIGVYHCEN
KDDQYFYFRSRGKEVAFGRLRNEKSSQCLDVSGYDGKGNVQMYDCEDKKDQWFKFYEN
GEVVNEQSRRCLDVSGYDGTGNIGTYCCEDKHDQMWSRPSQLCNGESCSFVNKKSGQC
LDVSGYDGRGGVATYHCEGLADQRLKWVTDKWTAPNAVWVMVGCNQNGKVSQWLSNTV
SYSSTITHTVTVEVGASMEADLVFAKATVSTKVSTSLSTAWTKSQSGTTRIVFTCEYY
DNQEAFTGGCMWQLRVDTKHVNSGRLLTWSPQITRCTTSNTQPRCPPFTKCVDKACSL

Figure 16: GenBank entry of Chrysaoralin gene from *Chrysaora quinquecirrha* captured from St. Mary's River, Chesapeake Bay, NJ. Accession number of the gene is KX356909.1.



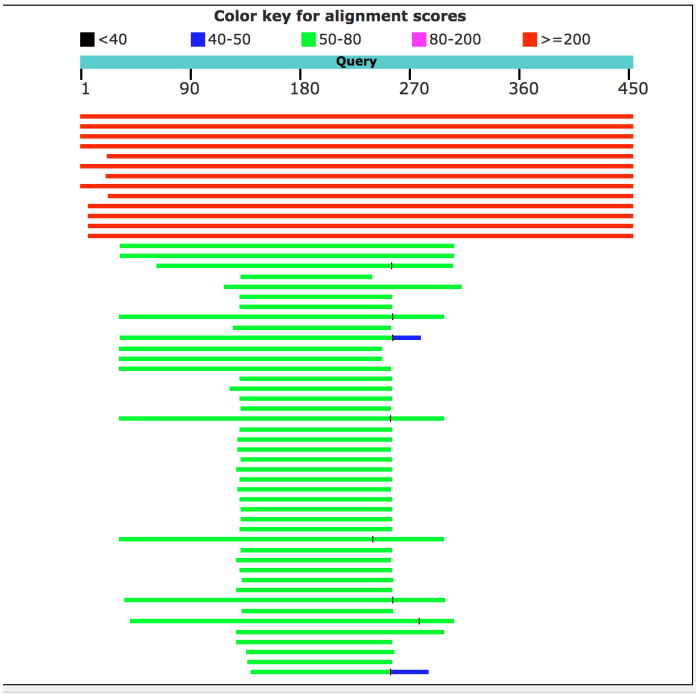


Figure 17.1 top: A graphical representation of the BLASTp result of *Chrysaora quinquecirrha* Chrysaoralin gene detecting conserved domains with the Ricin and Ricin B lectin superfamilies, sugar binding sites, as well as the Q-x-W motifs. *Figure 17.2 bottom:* Distribution of the top 113 BLAST hits.

A search of Genbank using the BLASTp algorithm for protein homologies to the Chrysaoralin gene is shown in Figure 18. Only the top 30 hits are listed. The best matches are to a family of hemolytic lectin proteins from the Sea Cucumber, *Cucumaria echinata,* with 62-64% amino acid identity. The search results also show homologies to the hypothetical proteins from other Cnidarians such as *Acropora millepora* and *Acropora digitifera*, a branching stony coral and an acroporid coral, respectively. The other BLAST hits show homology to

	Max score	Total score	Query cover	E value	Ident	Accession
chrysaoralin [Chrysaora quinquecirrha]	943	943	100%	0.0	100%	AOO35153.1
chrysaoralin [Chrysaora quinquecirrha]	940	940	100%	0.0	99%	AOO35152.1
hemolytic lectin-S1 [Cucumaria echinata]	597	597	100%	0.0	62%	BAK19472.1
hemolytic lectin-S2 [Cucumaria echinata]	597	597	100%	0.0	62%	BAK19473.1
hemolytic lectin-LS1 [Cucumaria echinata]	595	595	95%	0.0	65%	BAK19476.1
hemolytic lectin-L1 [Cucumaria echinata]	593	593	100%	0.0	62%	BAK19474.1
hemolytic lectin CEL-III [Cucumaria echinata]	592	592	95%	0.0	64%	BAC75827.1
hemolytic lectin-L2 [Cucumaria echinata]	588	588	100%	0.0	61%	BAK19475.1
Chain A, Crystal Structure Of Hemolytic Lectin Cel-iii	586	586	94%	0.0	64%	1VCL_A
hypothetical protein A036-E7 [Acropora millepora]	499	499	98%	6e-172	53%	ACJ64656.1
PREDICTED: uncharacterized protein LOC107342960 [Acropora digi	486	486	98%	1e-166	54%	XP_015763973.1
hypothetical protein A049-E7 [Acropora millepora]	483	483	98%	2e-165	50%	ACJ64657.1
PREDICTED: uncharacterized protein LOC107347166 [Acropora digi	479	479	98%	1e-163	50%	XP_015768531.1
hypothetical protein KCH_76140 [Kitasatospora cheerisanensis KCT(	76.6	76.6	60%	3e-11	26%	KDN80638.1
hypothetical protein [Kitasatospora cheerisanensis]	76.3	76.3	60%	4e-11	26%	WP_084224116.1
ricin-type beta-trefoil lectin domain protein [Pseudoalteromonas luteo	70.9	124	53%	2e-09	31%	WP_023398035.1
hypothetical protein [Prosthecomicrobium hirschii]	65.1	65.1	23%	5e-09	35%	WP_054361069.1
hypothetical protein [Pseudoalteromonas piscicida]	67.0	132	42%	6e-09	45%	WP_010371518.1
1,4-beta-xylanase [Streptomyces sp. JHA19]	67.8	67.8	27%	1e-08	33%	WP_055619031.1
xylanase [Streptomyces tendae]	67.0	67.0	27%	2e-08	33%	AAP72963.1
Glycosyl hydrolase family 59 [Streptomyces rubidus]	67.4	132	58%	2e-08	26%	SEN77595.1
hypothetical protein [Cellulomonas sp. URHD0024]	67.4	67.4	28%	3e-08	33%	WP_028049199.1
hypothetical protein [Streptomyces sp. NRRL S-1022]	66.2	113	54%	7e-08	26%	WP_078894836.1
esterase [Streptomyces venezuelae]	65.9	65.9	47%	7e-08	23%	ALO06367.1
putative secreted esterase [Streptomyces venezuelae]	65.9	65.9	47%	7e-08	23%	CUM43375.1
hypothetical protein [Streptomyces venezuelae]	65.9	65.9	49%	7e-08	23%	WP_079037757.1
1,4-beta-xylanase [Streptomyces katrae]	65.1	65.1	27%	1e-07	33%	WP_030294642.1
SGNH hydrolase [Amycolatopsis mediterranei]	64.7	64.7	29%	1e-07	32%	WP_013225599.1
MULTISPECIES: 1,4-beta-xylanase [Streptomyces]	65.1	65.1	27%	1e-07	33%	WP_062671095.1
hypothetical protein [Micromonospora lupini]	65.5	65.5	27%	1e-07	32%	WP_007461673.1
galactosylceramidase [Streptomyces hyaluromycini]	65.5	123	58%	1e-07	25%	WP_089106345.1
1,4-beta-xylanase [Actinobacteria bacterium OV450]	64.3	64.3	27%	2e-07	33%	WP_054227866.1

Figure 18: BLASTp search result of *Chrysaora quinquecirrha* Chrysaoralin gene. Only top 30 hits are shown. First two hits are Chrysaoralin gene sequences. Chrysaoralin gene shows 66% amino acid identity with the hemolytic lectin gene from the sea cucumber, *Cucumaria echinata.*

hypothetical proteins from different bacterial species. Curiously, after the hemolytic lectins from *Cucumaria echinata* and the unidentified proteins in corals, the homology drops precipitously. The next closest matches are primarily unidentified proteins from prokaryotes which match in the 30% to 20% homology range (with e values between 10^{-8} and 10^{-7}). Although I have only demonstrated the BLASTp results here, a similar pattern is seen at the nucleotide level using the BLASTn algorithm (data not shown). The lack of significant hits to Chrysaoralin from other cnidarians is both striking and surprising.

A comparison of the Chrysaoralin gene sequences isolated from two separate locations, Barnegat Bay in New Jersey and Chesapeake Bay in Maryland, was carried out using CLUSTAL Omega (Sievers *et al*., 2011), as seen in Figure 19. The comparison resulted in 6 single nucleotide polymorphisms (SNP's). The differences were seen between 28 G, 322 G, 501 C, 525 G, 552 G, 564 T of Chesapeake Bay Chrysaoralin and 28 A, 322 A, 501 T, 525 C, 552 A, 564 C of Barnegat Bay Chrysaoralin. However, these six nucleotide differences resulted in only two amino acid changes (10 Valine/Isoleucine and 108 Alanine/Threonine) between the genes from two locations, as can be seen in Figure 20. Thus, these two genes are 99.56% homologous at both the nucleotide and amino acid level.

A CLUSTAL Omega multiple sequence alignment of the Sea Cucumber Hemolytic Lectin, Chrysaoralin from Chesapeake Bay and Chrysaoralin from Barnegat Bay is shown in Figure 20. In the figure, the region shaded in grey is the signal peptide, shading in blue is domain 1, green is domain 2, and red is domain

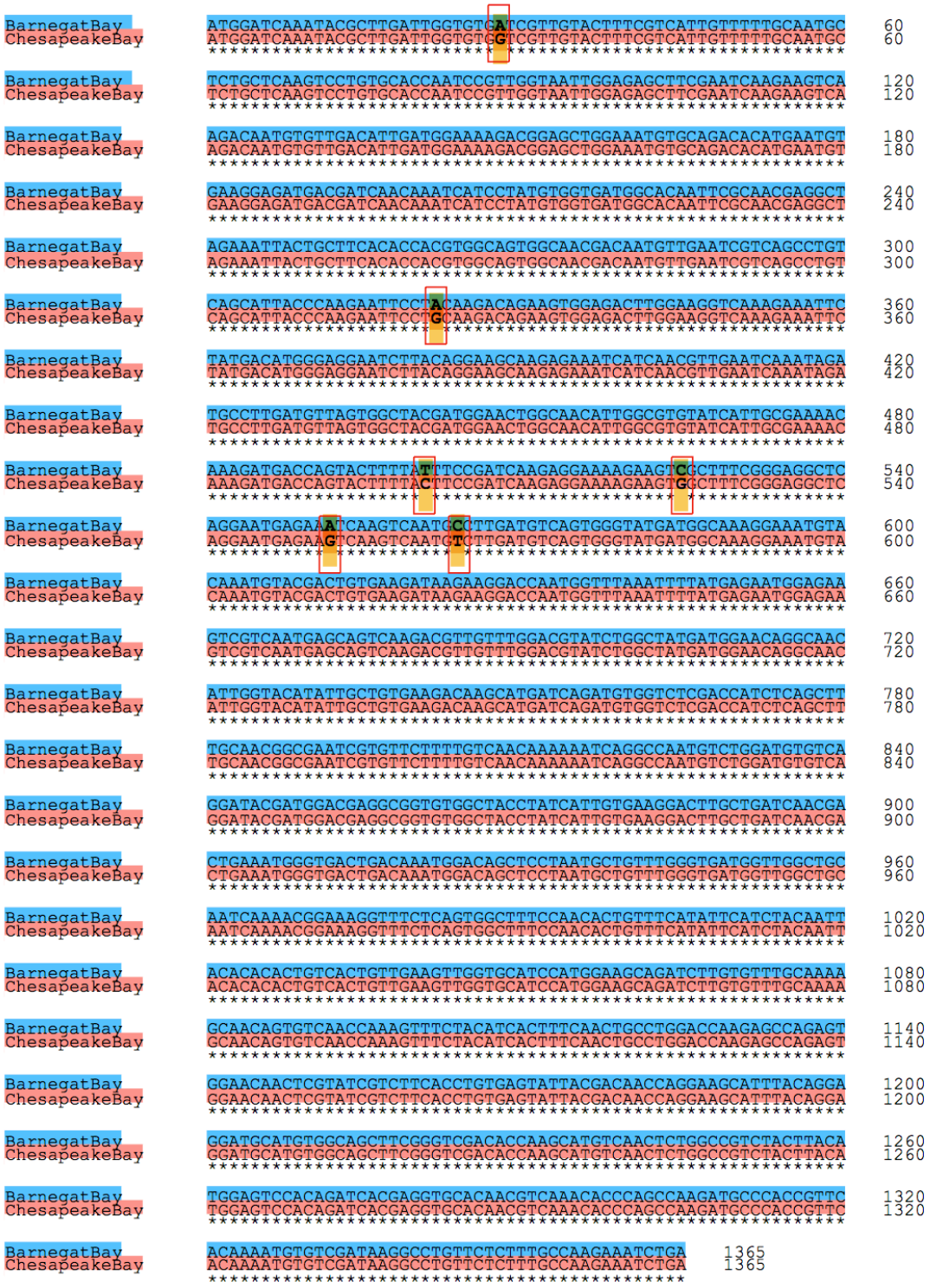


Figure 19: Comparison of Barnegat Bay and Chesapeake Bay Chrysaoralin sequences demonstrates a difference of six nucleotides. However, the difference result in only two changes in the amino acid sequences.

3. Domains 1 and 2 are putative carbohydrate-binding domains and domain 3 is the putative pore-forming domain (PFD). Highlighted within the boxes are the two amino acid differences (V/I and A/T) seen between Chesapeake Bay Chrysaoralin and Barnegat Bay Chrysaoralin, and “TVTVEVGASM” and “SVKVSTLSTA” sequences of the two alpha helices found in domain 3 that are responsible for formation of the transmembrane beta-barrel during pore formation (Unno *et al*., 2014).

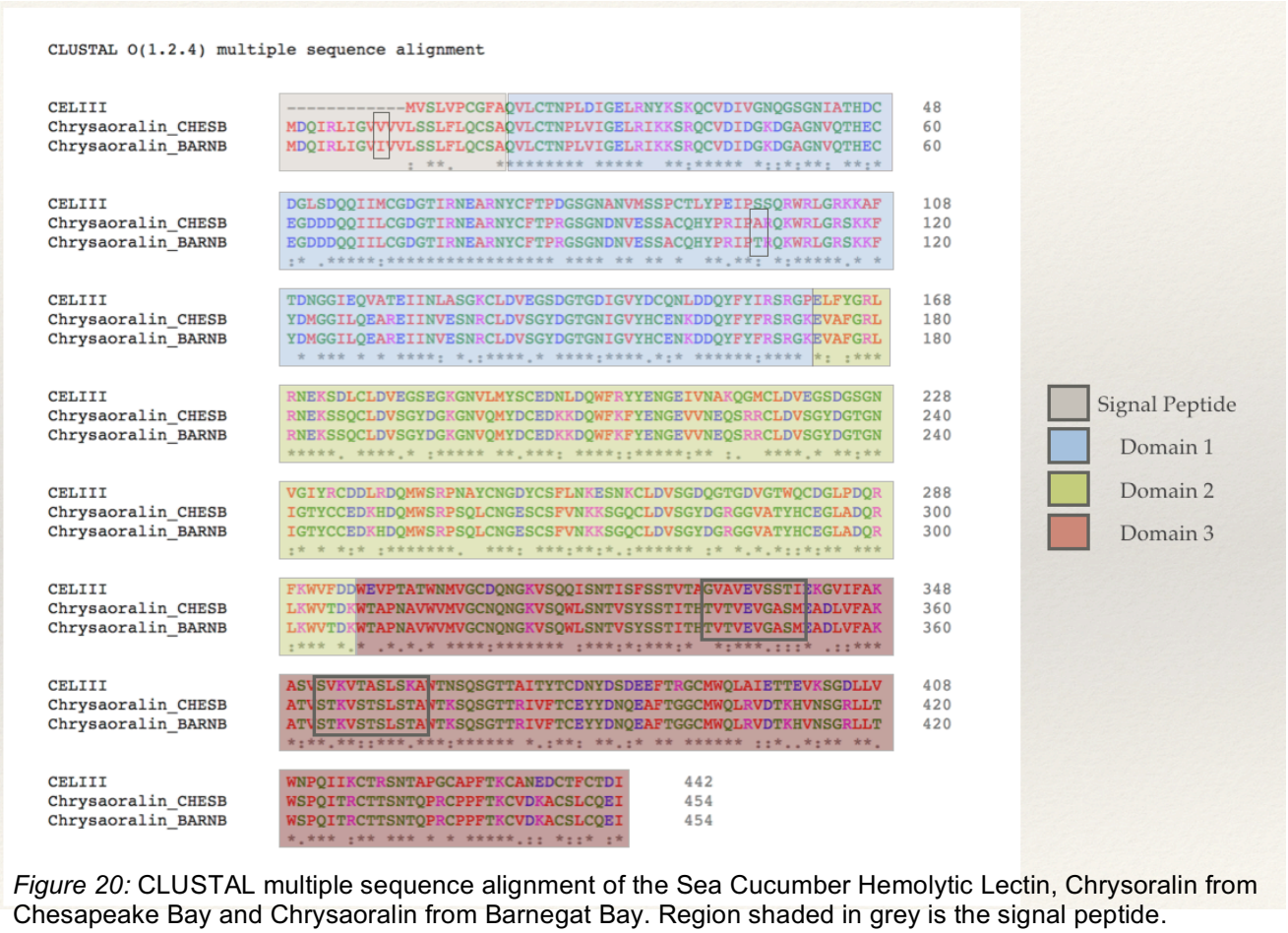


Figure 20: CLUSTAL multiple sequence alignment of the Sea Cucumber Hemolytic Lectin, Chrysoralin from Chesapeake Bay and Chrysaoralin from Barnegat Bay. Region shaded in grey is the signal peptide. Domain1 is shaded in blue, domain 2 in green and domain 3 in red. The two amino acid differences (V/I and A/T) between Chesapeake Bay Chrysaoralin and Barnegat Bay Chrysaoralin are highlighted. TVTVEVGASM and SVKVSTLSTA are sequences of the two alpha helices in domain 3 that are responsible for formation of the beta barrel.

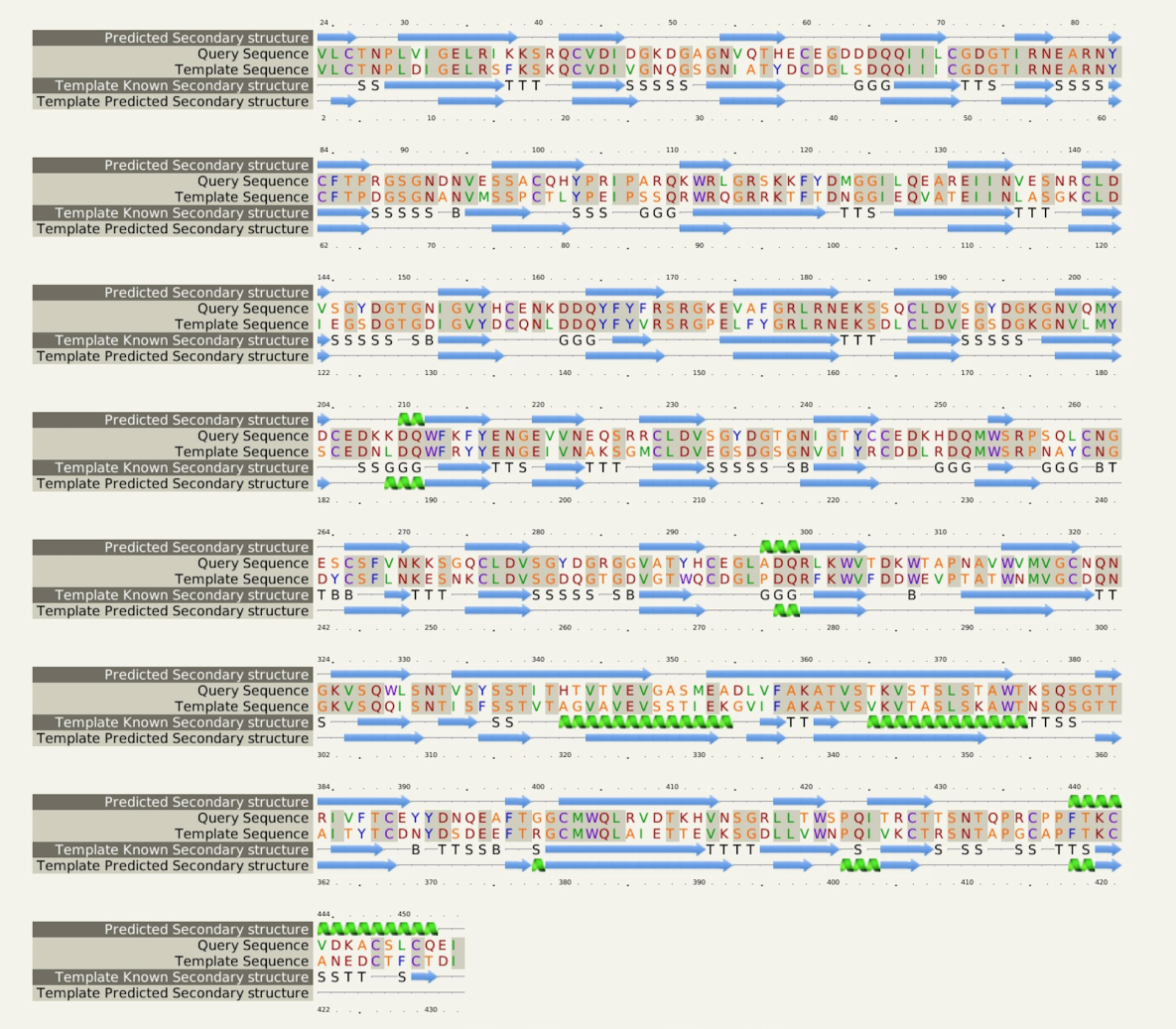


Figure 21: Secondary structure prediction of Chesapeake Bay Chrysaoralin protein using the Phyre2 web portal.

In addition to the sequence alignments, a comparison of the secondary protein structure of the known protein structure of sea cucumber was carried out with the predicted secondary structure of the Chrysaoralin gene as demonstrated in Figure 21. The different domains in sea cucumber were characterized by Unno *et al*. in 2014, which was utilized as a scaffold for the Phyre2 (www.sbg.bio.ic.ac.uk/~phyre2) web-based software to model, predict and analyze the secondary and tertiary structure of Chrysaoralin. This software predicted the folds in Chrsaoralin based on the amino acid homologies with the

sea cucumber hemolytic lectin. Phyre2 web tools was also used to build 3D models of the Chesapeake and Barnegat Bay Chrysaoralin, and sea cucumber hemolytic lectin as show in Figure 22. This software makes use of the predicted ligand binding sites and analyzes the effect of amino acid variants on the protein folding, thereby generating a confident 3D model of the protein structure. Based

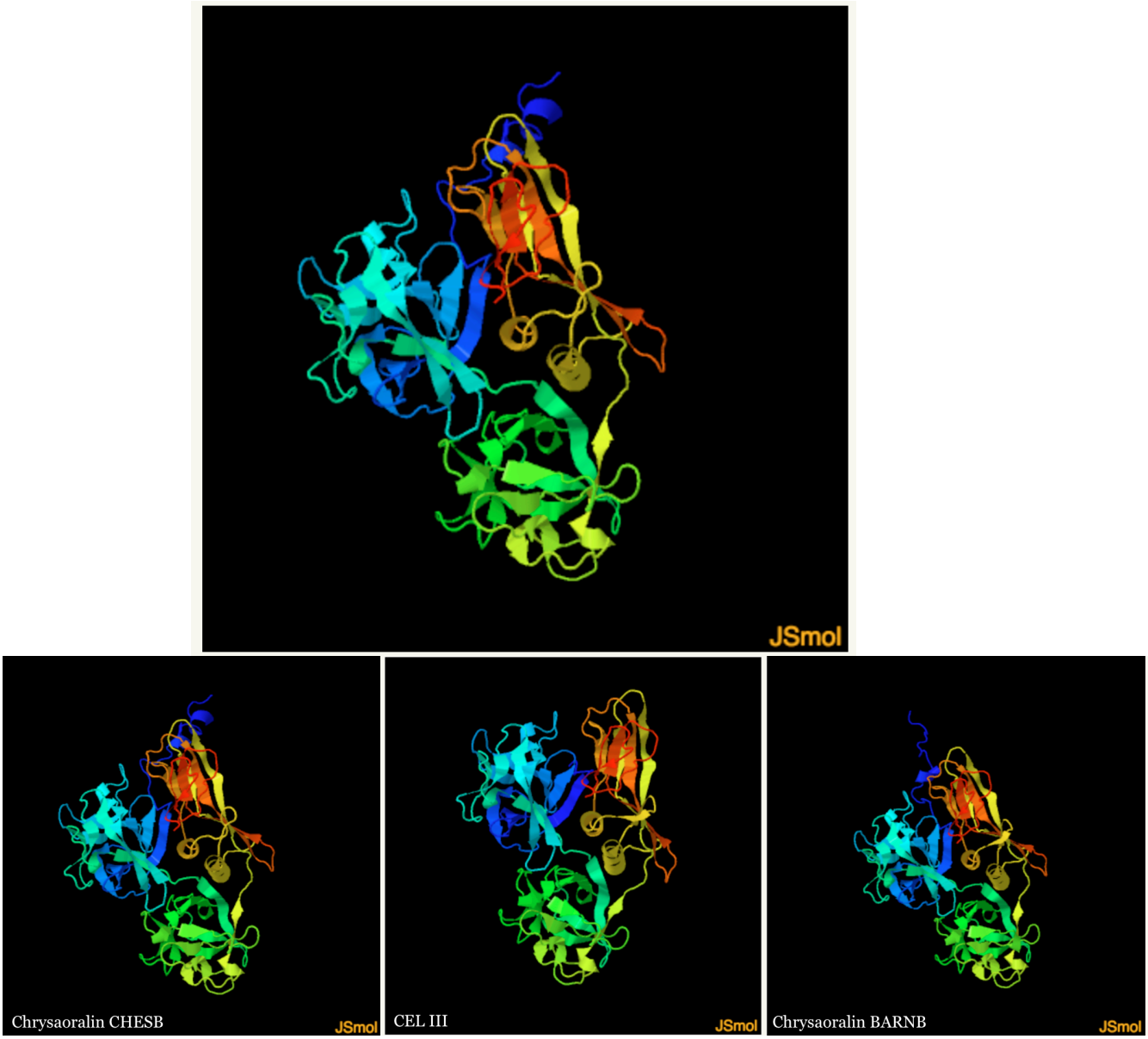


Figure 22.1 top: A Chrysaoralin protein monomer as predicted by the Phyre2 automatic fold recognition server. Phyre2 servers predict the three-dimensional structure of a protein sequence using homology modeling. *Figure 22.2 bottom:* Comparison of the models of Chesapeake Bay Chrysaoralin protein, Hemolytic Lectin from Sea cucumber and Barnegat Bay Chrysaoralin protein.

on the 3D models generated by Phyre2 web tool, these three proteins look strikingly similar in their appearance.

Geneious sequence analysis software was used to create an amino acid sequence alignment of the Chesapeake Bay Chrysaoralin and sea cucumber Hemolytic Lectin as shown in Figure 23. The important features of the two proteins are annotated on the bottom of the sequences. Highlighted in brown are metal ion binding sites in the following amino acid positions 45, 46, 48, 54, 55, 65, 94, 95, 143, 146, 149, 150, 152, 153, 159, 163, 190, 191, 193, 199, 200, 210, 231, 232, 234, 240, 241, 251, 278, 279, 281, 287, 288, 298, and 395. Highlighted in yellow are the Cysteine (C) residues and disulfide bonds between 26-71, 43-60, 84-100, 141-158, 188-205, 229-246, 261-266, 276-293, 320-402, 389-428, 437-451, and 443-448. Both metal ion binding sites and cysteine residues are conserved between *Chrysaora quinquecirrha C*hrysaoralin and the *Cucumaria echinata* hemolytic lectin. The conserved Ricin B-type lectin domains are highlighted by a red arrow that extends from residues 40-114, 127-257, 273-305 in Chrysaoralin. These structural features are also demonstrated in the 3D protein model using the UCSF Chimera (Pettersen *et al.*, 2004) software as seen in Figure 24.

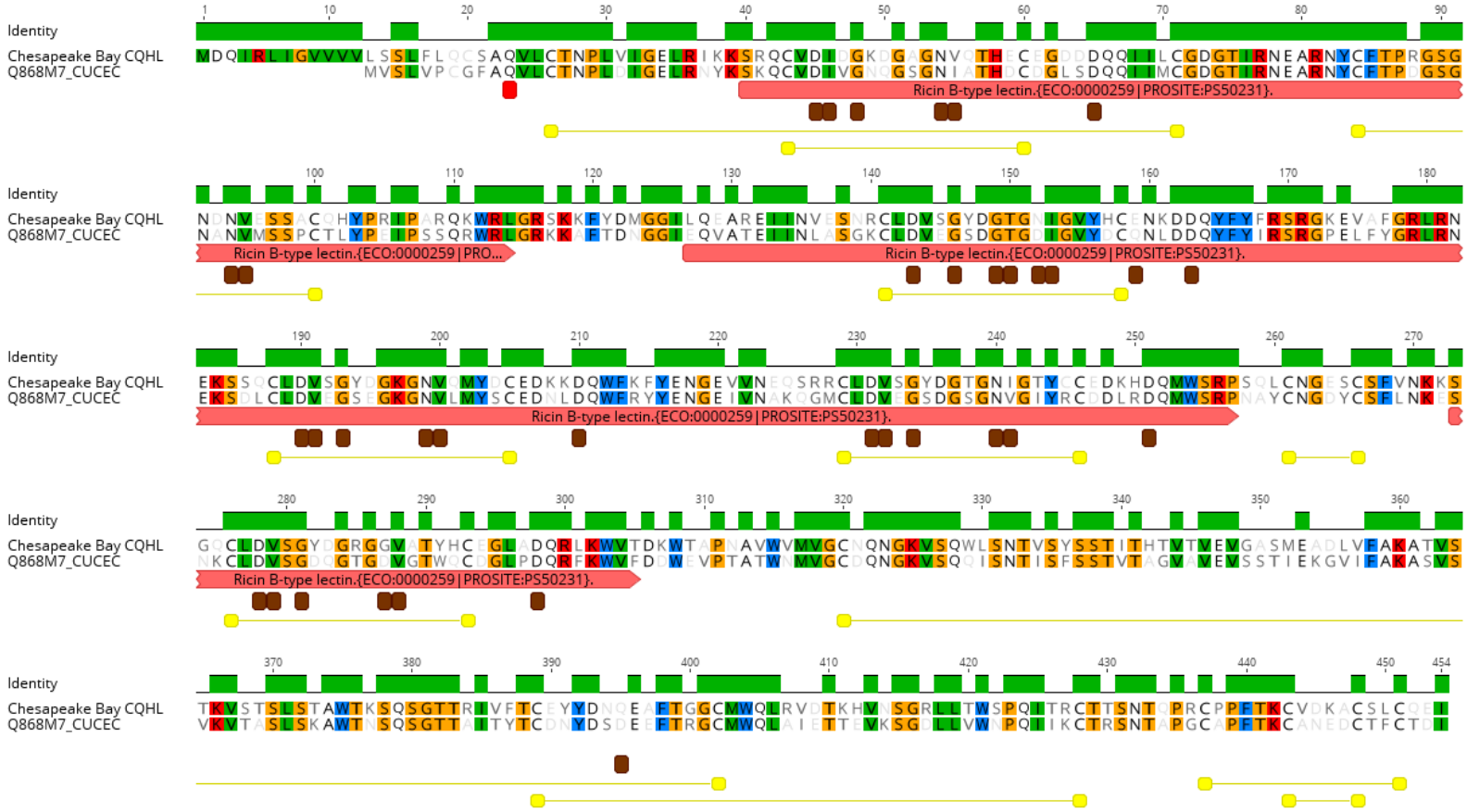


Figure 23: Amino acid sequence alignment of Chesapeake Bay Chrysaoralin and Sea Cucumber Hemolytic Lectin. The important features of the two proteins are annotated. The alignment was generated using Geneious sequence analysis software. Highlighted in brown are metal ion binding sites. Highlighted in yellow are the Cysteine residues and disulfide bonds between them. Both metal ion binding sites and cysteine residues are conserved between *Chrysaora quinquecirrha* chrysaoralin and *Cucumaria echinata* hemolytic lectin.

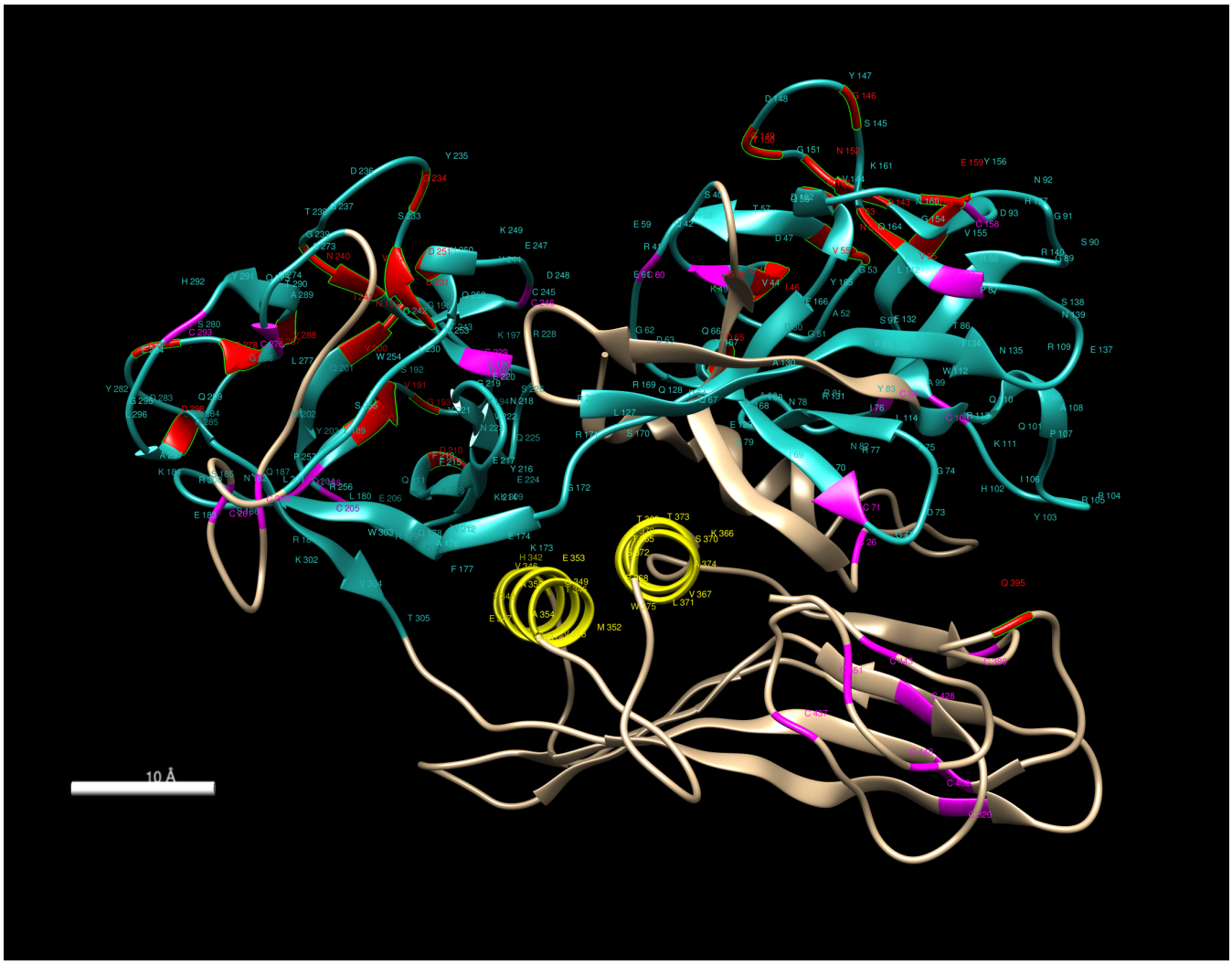


Figure 24: Model of the Chrysaoralin protein using UCSF Chimera. Highlighted in yellow are the two alpha helical regions (342-354, 365-376) in domain 3 of the Chrysaoralin protein that are responsible for pore formation. Highlighted in pink are the cysteine residues that form disulfide bonds (26-71, 43-60, 84-100, 141-158, 188-205, 229-246, 261-266, 276-293, 320-402, 389-428, 437-451, 443-448), in red are the metal ion binding sites (45, 46, 48, 54, 55, 65, 94, 95, 143, 146, 149, 150, 152, 153, 159, 163, 190, 191, 193, 199, 200, 210, 231, 232, 234, 240, 241, 251, 278, 279, 281, 287, 288, 298, 395), and in teal are the conserved ricin b lectin domains (40-114, 127-257, 273-305).

The Chrysaoralin gene (minus signal peptide, 1299 bases) was cloned into a pET SUMO vector (5643 bases) and the proof of cloning is shown in the agarose gel image in Figure 25. Lanes P12 and P15 were loaded with uncut plasmids, but no bands were observed. However, in the lanes that contained the Hind III

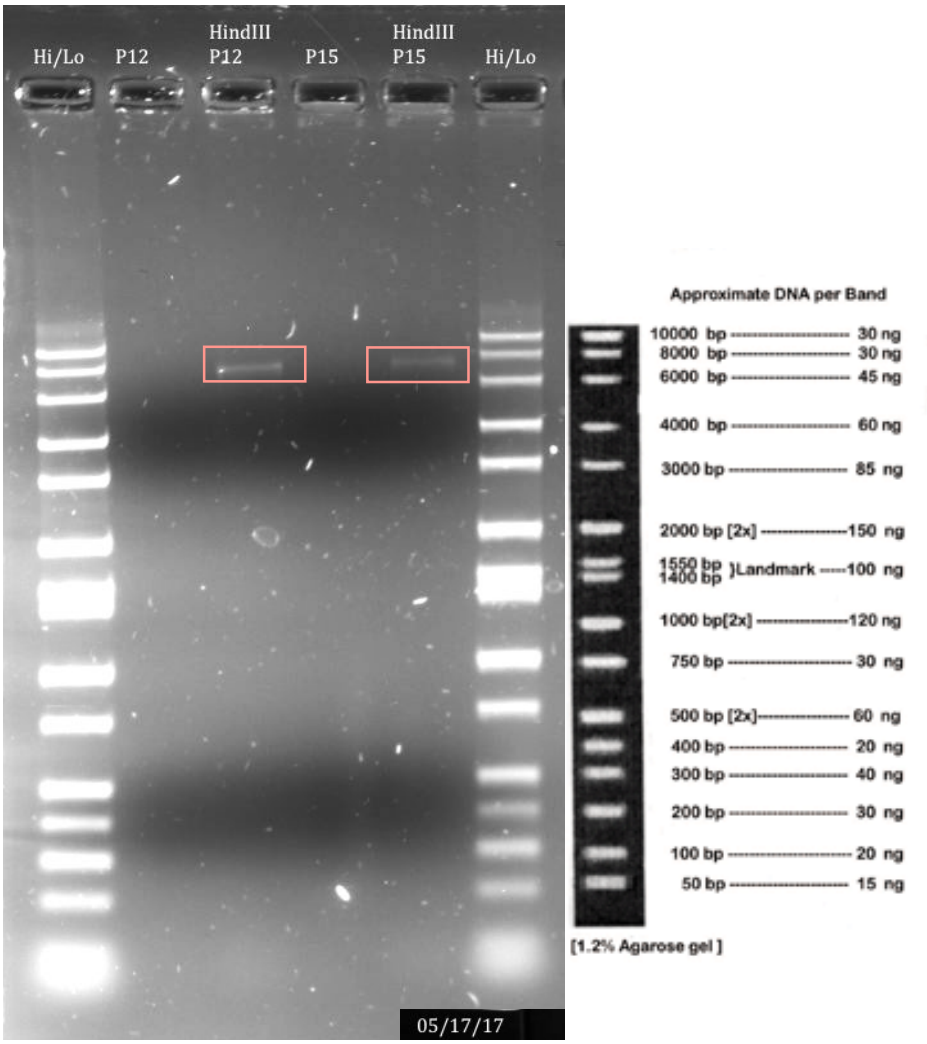


Figure 25: Restriction digest analysis of plasmid DNA. After ligation of Chrysaoralin gene (1299 bp, minus signal peptide) into the pET SUMO vector (5642 bp), the presence of insert in plasmid was analyzed by performing a single cutter restriction digest of the pET SUMO vector using HindIII restriction enzyme. Correct size of band of the gene in vector (6941 bp) is observed in the gel above.

restriction enzyme digested plasmids, bands are visible at the expected region of 6941 bases.

The section of pET SUMO vector containing Chrysaoralin gene is demonstrated in Figure 26.1. The SUMOF primer forward binding site is highlighted in pink, ChrysF primer binding site in blue, DO3F binding site in green, and the reverse complemented T7R primer binding site is highlighted in orange.

5'– AGA TTC TTG TAC GAC GGT ATT AGA ATT CAA GCT GAT CAG ACC
CCT GAA GAT TTG GAC ATG GAG GAT AAC GAT ATT ATT GAG GCT CAC
AGA GAA CAG ATT GGT GGT CAA GTC CTG TGC ACC AAT CCG TTG GTA
ATT GGA GAG CTT CGA ATC AAG AAG TCA AGA CAA TGT GTT GAC ATT
GAT GGA AAA GAC GGA GCT GGA AAT GTG CAG ACA CAT GAA TGT GAA
GGA GAT GAC GAT CAA CAA ATC ATC CTA TGT GGT GAT GGC ACA ATT
CGC AAC GAG GCT AGA AAT TAC TGC TTC ACA CCA CGT GGC AGT GGC
AAC GAC AAT GTT GAA TCG TCA GCC TGT CAG CAT TAC CCA AGA ATT
CCT GCA AGA CAG AAG TGG AGA CTT GGA AGG TCA AAG AAA TTC TAT
GAC ATG GGA GGA ATC TTA CAG GAA GCA AGA GAA ATC ATC AAC GTT
GAA TCA AAT AGA TGC CTT GAT GTT AGT GGC TAC GAT GGA ACT GGC
AAC ATT GGC GTG TAT CAT TGC GAA AAC AAA GAT GAC CAG TAC TTT
TAC TTC CGA TCA AGA GGA AAA GAA GTG GCT TTC GGG AGG CTC AGG
AAT GAG AAG TCA AGT CAA TGT CTT GAT GTC AGT GGG TAT GAT GGC
AAA GGA AAT GTA CAA ATG TAC GAC TGT GAA GAT AAG AAG GAC CAA
TGG TTT AAA TTT TAT GAG AAT GGA GAA GTC GTC AAT GAG CAG TCA
AGA CGT TGT TTG GAC GTA TCT GGC TAT GAT GGA ACA GGC AAC ATT
GGT ACA TAT TGC TGT GAA GAC AAG CAT GAT CAG ATG TGG TCT CGA
CCA TCT CAG CTT TGC AAC GGC GAA TCG TGT TCT TTT GTC AAC AAA
AAA TCA GGC CAA TGT CTG GAT GTG TCA GGA TAC GAT GGA CGA GGC
GGT GTG GCT ACC TAT CAT TGT GAA GGA CTT GCT GAT CAA CGA CTG
AAA TGG GTG ACT GAC AAA TGG ACA GCT CCT AAT GCT GTT TGG GTG
ATG GTT GGC TGC AAT CAA AAC GGA AAG GTT TCT CAG TGG CTT TCC
AAC ACT GTT TCA TAT TCA TCT ACA ATT ACA CAC ACT GTC ACT GTT
GAA GTT GGT GCA TCC ATG GAA GCA GAT CTT GTG TTT GCA AAA GCA
ACA GTG TCA ACC AAA GTT TCT ACA TCA CTT TCA ACT GCC TGG ACC
AAG AGC CAG AGT GGA ACA ACT CGT ATC GTC TTC ACC TGT GAG TAT
TAC GAC AAC CAG GAA GCA TTT ACA GGA GGA TGC ATG TGG CAG CTT
CGG GTC GAC ACC AAG CAT GTC AAC TCT GGC CGT CTA CTT ACA TGG
AGT CCA CAG ATC ACG AGG TGC ACA ACG TCA AAC ACC CAG CCA AGA
TGC CCA CCG TTC ACA AAA TGT GTC GAT AAG GCC TGT TCT CTT TGC
CAA GAA ATC **TGA** AGA CAA GCT TAG GTA TTT ATT CGG CGC AAA GTG
CGT CGG GTG ATG CTG CCA ACT TAG TCG AGC ACC ACC ACC ACC ACC
ACT GAG ATC CGG CTG CTA ACA AAG CCC GAA AGG AAG CTG AGT TGG
CTG CTG CCA CCG CTG AGC AAT AAC TA –3'

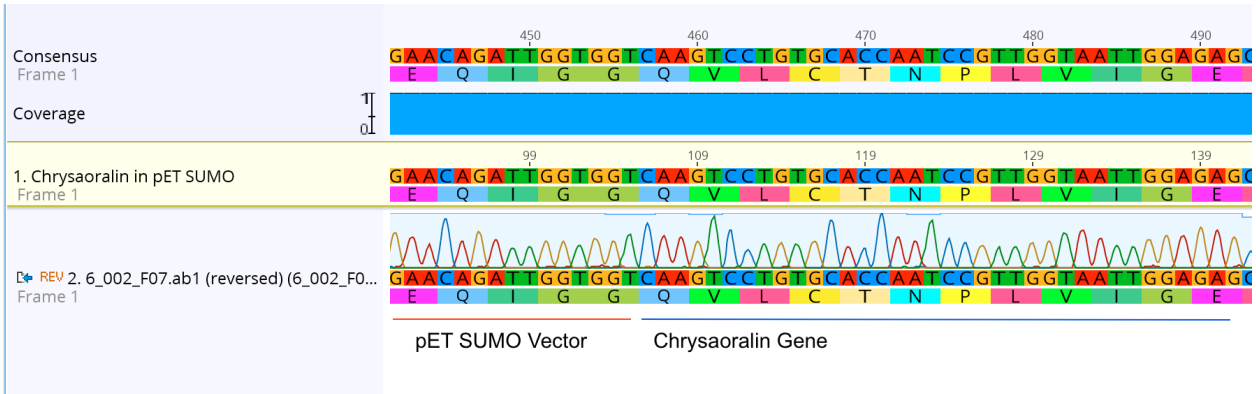


Figure 26.1 top: Chrysaoralin gene (minus signal peptide) in pET SUMO vector. Highlighted in pink is the SUMOF primer forward binding site, in blue is ChrysF binding site, in green is DO3F binding site, and in orange is the binding site for reverse complement of T7R primer. Highlighted in red is the stop codon. *Figure 26.2 bottom:* An electropherogram demonstrating the presence of Chrysaoralin gene (minus signal peptide) in the pET SUMO vector, in the correct reading frame and right direction (5' to 3').

Highlighted in red is the stop codon. The proof that the Chrysaoralin gene is correctly cloned into the pET SUMO vector is demonstrated by the electropherogram in Figure 26.2. In the figure, a sequence alignment performed between the putative gene/vector sequence and the electropherogram of the sequencing of the amplicon generated by the colony PCR of *E. coli* containing the inserts is demonstrated. The junction between vector and the gene, vector sequences AAT GGT GGT // Chrysaoralin sequences CAA GTC CTG, are in the correct reading frame and orientation. This translates to IGG amino acid in the vector and QVL in the gene, which demonstrates the successful cloning of the full-length Chrysaoralin gene (minus the signal peptide) into the pET SUMO expression vector.

Full-length (minus signal peptide) and truncated version (only Domain 3) of the Chrysaoralin gene were successfully amplified from the pET SUMO vector as demonstrated in Figure 27. The amplicons of correct sizes, 1299 bases for full-length gene and 444 bases for domain 3, are demonstrated in the agarose gel image in Figure 27.

A phylogenetic tree of 19 pore forming proteins from Table 1 created using the amino acid sequences of the pore-forming proteins as available on GenBank and Protein Database is demonstrated in Figure 28. The phylogenetic tree was generated using Jukes-Cantor genetic distance model, Blosum62 cost matrix and Neighbor-Joining method in Geneious sequence analysis software. Distance between the pore forming proteins (prokaryotic in black, eukaryotic in blue

typeface) were obtained by pairwise alignment of the sequences. The genus and common name of the pore forming protein is represented in the figure.

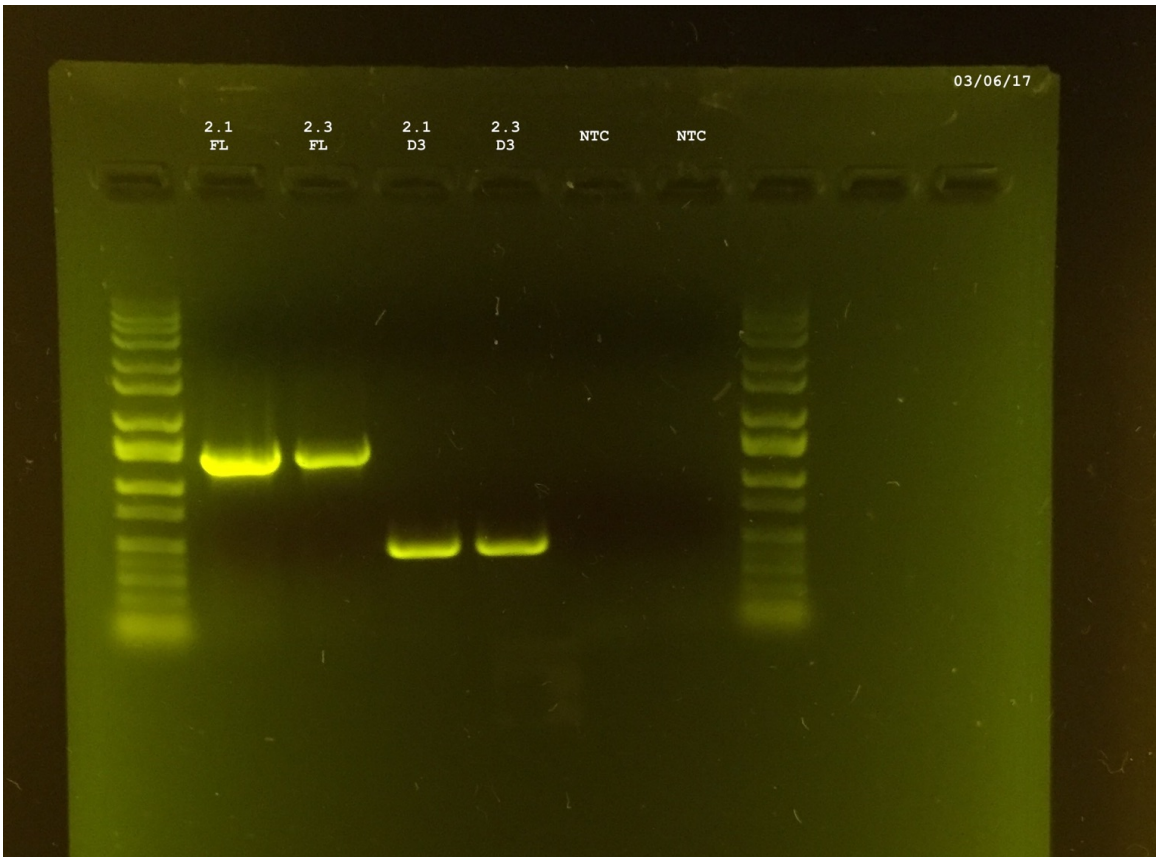


Figure 27: Full length Chrysaoralin (2.1 FL and 2.3 FL) and truncated version (2.1 D3 and 2.3 D3) of the gene in pET SUMO vector. Amplification of the gene was carried out using Colony PCR.

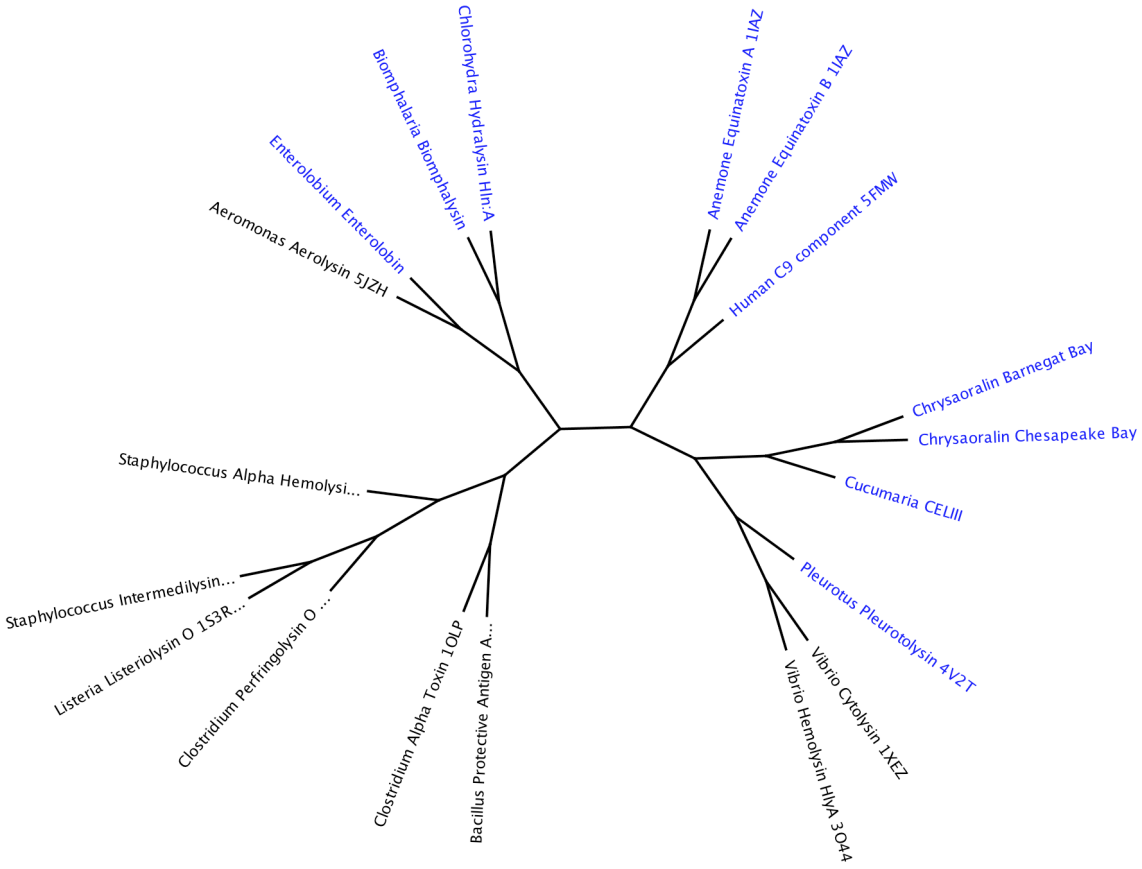


Figure 28: A phylogenetic tree of pore forming proteins from Table 1 constructed using Jukes-Cantor genetic distance model and Neighbor-Joining method in Geneious sequence analysis software. Distance between the pore forming proteins (prokaryotic in black, eukaryotic in blue typeface) were obtained by pairwise alignment of the sequences. Blosum62 cost matrix was employed. The genus and common name of the pore forming protein is represented in the figure.

Discussion

Chrysaoralin Gene Assembly

In the preliminary stages of my study, RNA sequencing (RNA-Seq) was used for rapid gene profiling and investigation of the transcriptome of *Chrysaora quinquecirrha* adult medusa captured from Barnegat Bay, NJ. The inquiry objectives were to investigate protein composition of *Chrysaora quinquecirrha* venom and identify specific biological function of its protein constituents. Proteins in the venom were traditionally identified and studied by raising antibodies in expression libraries and by other alternative methods such as Edman degradation. These days high-throughput “omics” methodologies that incorporates proteomics, transcriptomics and genomics have been the method of choice for identification and characterization of novel peptides in venom (Fox *et al*., 2008, Brahma *et al.*, 2015). This has led to a new term “venomics”, which is considered the proteomic characterization of venom proteomes. The significance of “omics" methodologies for venom research has been possible by the pace of genome sequencing projects and availability of fully sequenced genomes of organisms. However, overwhelming majority of organisms, including *Chrysaora quinquecirrha*, still contain unsequenced genomes. Therefore, my study attempts to fulfill this knowledge gap and embraces both experimental and *in silico* techniques such as PCR, Sanger DNA sequencing, RNA sequencing technology, and bioinformatics to identify and characterize a toxin peptide from *Chrysaora quinquecirrha* venom.

Transcriptome analysis for protein coding sequence detection has been used before for toxin identification in the scyphozoan jellyfish *Chrysaora*

fuscescens (Ponce *et al.*, 2016), the cubozoan jellyfish *Chironex fleckeri* (Brinkman *et al*., 2015), and the scyphozoans *Stomolophus meleagris* (Li *et al*., 2014) and *Cyanea capillata* (Liu *et al*., 2015). A mRNA transcriptome reflects the genes that are being actively expressed at a given time and is considered one of the key tools in identifying candidate genes that may code for proteins with therapeutic potential (Li *et al*., 2014, Brinkman *et al*., 2015, Ponce *et al.*, 2016). In my study, RNA extracted from tentacles of a single Sea Nettle captured from Barnegat Bay in NJ generated a total of 380,000,000 pairs of 100-base-length paired-end reads (data unpublished). The raw sequencing data was processed by eliminating sequences with low quality scores and removing adaptor sequences, and assembled using CLC Workbench to generate a file of 87,600 contigs. BLASTX search of all the contigs resulted in 30,817 significant hits to known proteins on GenBank. Contig 22,835 from the transcriptome data as shown in Figure 8 matched a hemolysin, hemolytic lectin, from *Cucumaria echinata*, a sea cucumber from the phylum Echinodermata. This protein is a calcium dependent and galactose-specific lectin that exhibits hemolytic and hemagglutinating activities (Uchida *et al.*, 2004). Similarity of contig 22,835 with the sea cucumber hemolytic lectin protein warranted further investigation of the protein coded by the gene in contig 22,835.

The contig 22835 was processed through a signal peptide detecting algorithm SignalP 4.0. The SignalP 4.0 algorithm (http://www.cbs.dtu.dk/services/SignalP/) predicted a signal peptide of 22 AA (MDQIRLIGVIVVLSSLFLQCSA) (Figure 9 and Figure 20). The primary function of

the signal peptide is to translocate the protein downstream of the internal signal peptide (Coleman *et al*., 1985). The signal peptides direct the protein to the endoplasmic reticulum, where the protein matures before secretion (Duffaud *et al*., 1985).

Contig 22,835 was used as a scaffold to generate 9 sets of primers (Figure 10) to verify the integrity of the contig assembly using PCR amplification, dideoxy Sanger sequencing and sequence analysis. These primers spanned the full length of the contig (minus the poly A tail). PCR amplification with the mixture of different primer sets from the list in figure 9 would allow to generate amplicons of permissible length for reliable sequencing. The primers were mapped against the Barnegat bay jellyfish mRNA transcriptome data using SnapGene viewer (Figure 11.1) and Geneious sequencing analysis software (Figure 11.2) for an *in silico* analysis. These platforms enabled me to visualize and locate primers on the putative gene, and helped to design the amplification reactions. The RNA transcriptome was verified using the genomic DNA from *Chrysaora quinquecirrha* captured from Chesapeake Bay in Maryland.

Using jellyfish samples from two separate locations, Barnegat Bay and Chesapeake Bay, I intended to verify the transcriptome analysis and additionally investigate the presence of any variations in my gene of interest. A pair-wise sequence alignment was carried out for all genomic sequencing data to look for any ambiguities within the forward and reverse (reverse-complemented) sequences from Chesapeake Bay Chrysaoralin. The genomic sequencing data was visualized using 4Peaks and Geneious sequencing analysis software for

preliminary sequence analysis. During this course, the forward and reverse sequences were lined up against each other so the complementary nucleotides generated a perfect match. The poor reads in the beginning and end of the sequence (approximately 50 nucleotides on both ends) were trimmed out before sequence alignments were made, because alignment of untrimmed sequences yielded unpredicted and unreliable results. The first 50 bases are of low quality because of uneven separation by electrophoresis and noise created by the unincorporated primers and primer dimer artifacts. Similarly, the last few bases at the end of the sequence are of low quality because of the reduced sequence strength, low availability of fragments towards the end of the sequencing reaction, and poor separation of fragments due to smaller relative differences in electrophoretic mobility (Brown, 2013). In some cases, the forward and reverse-complemented sequences yielded subtle variations in more than one site. At least 8 heterozygosities were detected in the full length Chrysaoralin gene from Chesapeake Bay at the following nucleotide positions 348, 416, 459, 561, 567, 575, 600, and 660 (starting at ATG, A as 1). This could mean that there is presence of more than one copy of the Chrysaoralin gene in the organism's genome. A consensus sequence was generated with the most frequently occurring nucleotides and the consensus sequence was submitted to GenBank. These variations were noted as ambiguities in the sequence. However, more sequencing data is required to reliably investigate any heterozygosities in the gene.

The consensus sequence generated from the assembly of multiple overlapping reads in both directions was used as the scaffold to generate a new

set of primers (ChrysF and CQR) that would result in the full length of the gene; nevertheless, the ABI sequencer is able to generate high quality reads of only 500-700 bases. Considering this sequencing limitation, an additional inner primer set is needed to obtain an overlapping sequence to produce full length of the gene. The permutation of the 9 primer sets generated sufficient amplicons to span the full length of the gene (Figure 13) and my initial project to obtain full length of the gene was completed. The genomic sequence for the Chesapeake Bay chrysaoralin gene was assembled successfully and full coverage of the gene with as less as 3 sequence files is demonstrated in figure 14.

Based on the RNA-seq data in conjunction with the direct DNA sequencing of genomic DNA, the Chrysaoralin gene from both Barnegat Bay and Chesapeake Bay populations of *Chrysaora quinquecirrha* was found to be intronless. It is generally accepted that introns are common in eukaryotic genes, especially in multicellular eukaryotes. The existence of introns within hemolytic genes in Cnidarians have been previously reported in the sea anemones *Actineria villosa* and *Phyllodiscus semoni.* Both of these genes were identified to be highly toxic and to contain introns (Uechi *et al*., 2010). A small number of venom peptide encoding genes from cone snails, scorpions, and sea anemones also contain an intron-exon architecture, however there are a few exceptions (Pineda *et al*., 2012). It is therefore, somewhat unusual, to find that my Chrysaoralin toxin genes from sea nettles are intron-less.

There is evidence supporting the fact that rapidly regulated genes, such as the heat shock protein genes, lack introns (Jeffares *et al*., 2008). Introns may delay

regulatory responses and therefore some eukaryotic lineages lose their introns to rapidly synthesize proteins in response to various intracellular stresses. It is thus speculated that the presence of introns are selected against in genes whose proteins are required for rapid adjustments to cope with environmental difficulties. The evolution of Chrysaoralin gene sequence without any introns may have occurred to minimize the delay in transcript processing and to permit rapid translation.

Comparison of the Barnegat Bay and Chesapeake Bay Chrysaoralin Genes

Chesapeake and Barnegat Bay Chrysaoralin gene sequences were aligned and compared using Geneious and CLUSTAL omega sequence alignment tools (Figure 19) to examine any differences in the gene sequences. The Barnegat Bay *Chrysaora quinquecirrha* gene captured from Barnegat Bay, Metedeconk River in Brick, NJ was assembled using the CLC Genomics Workbench v.7.5 platform and submitted to GenBank (accession number (KX656922.1, Figure 15) after sequence analysis. The Chesapeake Bay *Chrysaora quinquecirrha* was captured from St. Mary's River that flows into the Chesapeake Bay in Maryland. The Chesapeake Bay gene sequence was assembled *de novo* and verified using 4Peaks and Geneious sequence analysis software, and then submitted to GenBank (accession number (KX356909.1, Figure 16).

Comparison of Barnegat Bay and Chesapeake Bay Chrysaoralin sequences resulted in a difference of 6 nucleotides. On the protein level, the 6 nucleotide difference translated to difference of only 2 amino acids. The first amino

acid difference lies in the signal peptide region of the Chrysaoralin gene, which eventually gets cleaved off in the mature peptide. The signal peptide Valine (V, GTC) residue in Chesapeake Bay Chrysaoralin varies with the Isoleucine (I, ATC) residue in the Barnegat Bay Chrysaoralin. Another difference lies between Alanine (A, GCA) from Chesapeake Bay Chrysaoralin and Threonine (T, ACA) residue from Barnegat Bay Chrysaoralin in the carbohydrate binding domain of the Chrysaoralin gene. In both cases, the two amino acid variations have resulted from mutations in the first nucleotide of the codons that code for their respective amino acids.

The foundation of biodiversity lies in the genetic variation within species that result from one or more variations in their genetic composition. In this case, there were six nucleotides differences within a single gene from the two separate locations. Although yet to be proven, these differences may have been environmentally induced. A single mutation at a decisive location in a gene can have a significant effect. These variations have resulted in a difference of two amino acid residues, however, the effect of two amino acid residue variation on the functionality of the toxin was not investigated in this study.

The sequences of the Barnegat and Chesapeake bay Chrysaoralin were identical, except for the two aforementioned positions. The occurrence of the subtle variation in the gene sequence of the two locations may have been due to genetic adaptation to the different environmental conditions in Chesapeake and Barnegat Bays. However, the functional features of this polymorphism will be

known only upon expression of the gene and testing the effect of the protein using biological assays.

Conserved Motifs in Chrysaoralin

Using the NCBI BLASTx tool that queries protein databases using translated nucleotides, I found that Chrysaoralin protein is 64% identical to a hemolytic lectin protein (CEL III) from the Sea Cucumber, *Cucumaria echinata*. In addition to the isolation, cloning, and characterization of the hemolytic CEL III protein, crystal structure of CEL III has also been determined by Uchida *et al.* (2004) at 1.7 Å resolution. The availability of CEL III crystal structure and a strong homology between Chrysaoralin and CEL III not only enhance our understanding on the structural and functional features of Chrysaoralin protein, but also help to elucidate its pore forming mechanism.

BLAST search of the Chrysaoralin sequence reveals some highly conserved sequence features in the N-terminal region, particularly to the B-chains of ricin (Figure 19.1, Figure 29). The Ricin-type beta-trefoil domains are Carbohydrate-binding domains formed from presumed gene triplication. The domain is found in a variety of molecules from a wide range of organisms serving diverse functions such as enzymatic activity, inhibitory toxicity and signal transduction (Uchida *et al.,* 2004). Conserved domain homology on NCBI BLAST reveals that the Ricin-type beta-trefoil domains extends from 97 to 762 nucleotide positions on the N-terminal region of the Chrysaoralin protein. Contrarily, the C-

terminal region revealed no conservation to any known domains in the NCBI database.

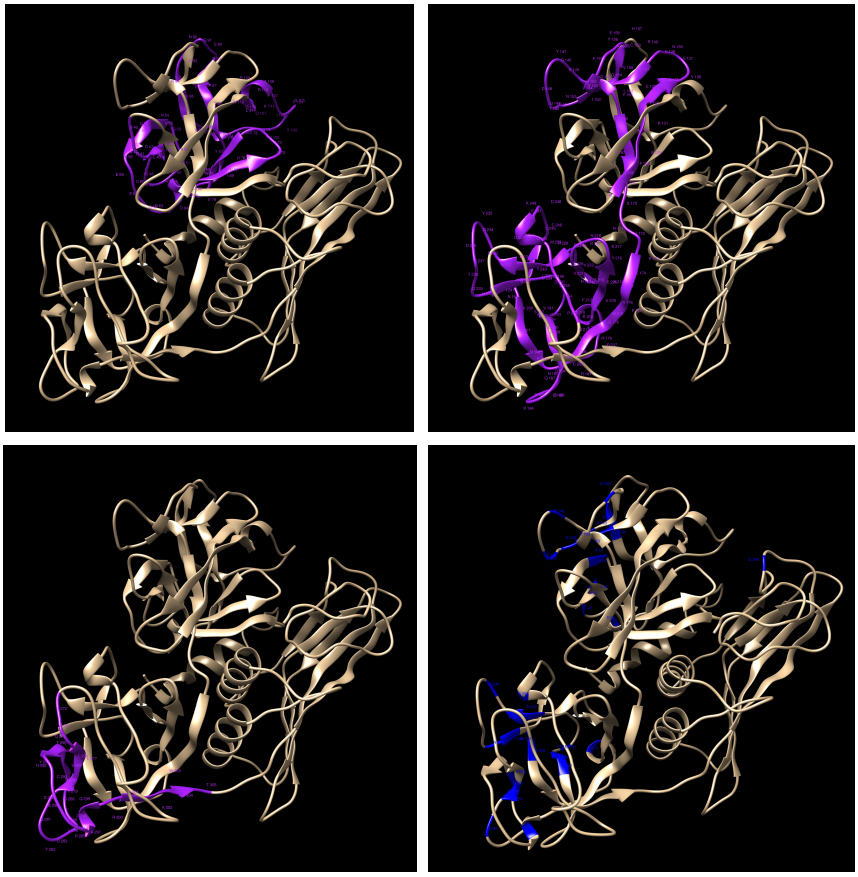


Figure 29: Three B-chains of Ricin (40-114, 127-257, 273-305) are highlighted in purple. Highlighted in blue are metal ion binding sites (45, 46, 48, 54, 55, 65, 94, 95, 143, 146, 149, 150, 152, 153, 159, 163, 190, 191, 193, 199, 200, 210, 231, 232, 234, 240, 241, 251, 278, 279, 281, 287, 288, 298, 395). Protein models created using UCSF Chimera.

The homologous hemolytic lectin from sea cucumber, CEL III, also belongs to the ricin-type (R-type) lectin family. Their carbohydrate-recognition pattern resembles to that of the C-type lectins that bind specific carbohydrates through coordinate bonds with Ca2+ located at their binding sites (Drickamer, 1999). The CEL III are members of the R-type lectins, the lectin family found in both

prokaryotes (bacteria) and eukaryotes (*C. elegans*, *Drosophila*, vertebrates, and plants) (Mancheno *et al*., 2010).

Amino acid residues in the ricin B-chain involved in carbohydrate-binding that are conserved in CEL III are conserved in Chrysaoralin as well (Figure 22). The Ricin B-type lectin homologs are spread across three different locations in both Chrysaoralin and CEL-III genes (Figure 29). These three homologs are 75, 131, and 33 amino acids in length respectively, and on average approximately 70% identical to each other in terms of their amino acid residues. The high degree of conservation in structure and in sugar-binding function of this carbohydrate binding domain could mean that the gene encoding an R-type lectin has moved laterally between species (Mancheno *et al*., 2010). Highly conserved Ricin B-lectin domain in the two proteins may mean that these proteins are similar in their specificities to the carbohydrate moieties they bind on the target cell membrane however, this claim is yet to be verified experimentally. Additionally, the selectivity of Chrysaoralin protein to specific carbohydrates can be utilized during the construction of protein purification columns.

Two (QxW) sub-domains are present within the B-chains of ricin in Chrysaoralin gene (QKW, aa110-112, QMW aa252-254) (Figure 30.1). These domains are known to accommodate considerably differing amino acids at multiple positions in the protein structure (Mancheno *et al*., 2010). Another sub-domain called (QxF), similar to the (QxW) sub-domain, is also present within the B-chains of ricin in both CEL III and Chrysaoralin genes (QYF aa164-166, QWF aa211-213) (Figure 30.2). (QxW) and (QxF) sub-domains are laterally shared by many

unrelated proteins and are suitable evolutionary building blocks, which by gene fusion add carbohydrate-binding functionality to other proteins (Mancheno *et al*., 2010). There is also an assumption by Mancheno *et al* that these motifs might have appeared early in evolution and thus been available to most evolving organisms to create proteins of varying functionality and exotic properties.

The presence of the carbohydrate binding motifs such as (QxW) and (QxF) in Chrysaoralin adds to the protein's ability to bind membrane surface carbohydrates and enhance its functionality as a pore forming protein.

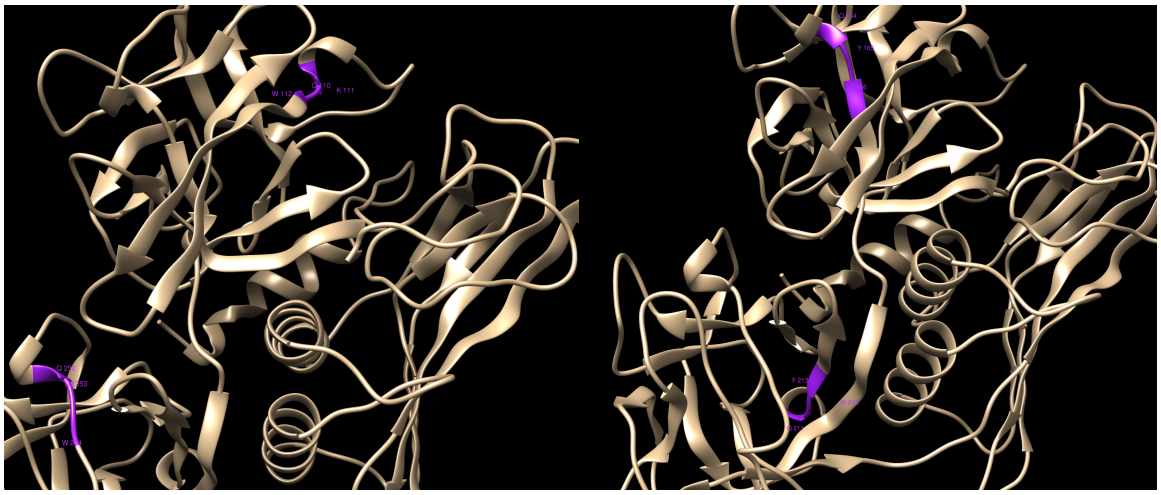


Figure 30.1 left: (QxW) sub domains (110-112, 252-254) in Chrysaoralin. *Figure 30.2 right:* (QxF) sub domains (164-166, 211-213) in Chrysaoralin. Protein model generated using UCSF Chimera.

Pore Forming Mechanism

A study by Hatakeyama *et al.* (1995) revealed that the CEL III protein exerted hemolytic activity by damaging the erythrocyte membrane upon binding to specific carbohydrate moieties on the cell membrane. Their experiments were carried out with the galactose and GalNAc containing carbohydrates, exploiting CEL III protein's specificity to these carbohydrate moieties. Hemolysis was tested with rabbit erythrocytes under different temperature conditions and highest hemolytic activity was observed at 10ºC. This suggested that pore formation is a temperature sensitive non-enzymatic process (Hatakeyama *et al.*, 1995). Until this experiment, the exact pore forming mechanism of the CEL III protein was not known. So, they postulated that the hemolytic activity was caused by the formation of a transmembrane pore and conducted immunoblotting experiments on the proteins bound to the CEL III treated membrane. They discovered higher molecular weight of the irreversibly bound anti-CEL III antiserum in the susceptible erythrocytes treated with CEL III (Hatakeyama *et al.*, 1995). This finding led them to conclude that there was aggregation of the CEL III protein on the erythrocyte membrane.

Subsequent experiments by Hatekayama's group on the CEL III protein to determine its molecular weight revealed an interesting feature. They observed multiple high molecular weight bands than expected on the SDS page membrane. This observation was extrapolated to similar feature observed in other beta pore forming toxins, such as the *Staphylococcus aureus* alpha-toxin, that forms transmembrane pores with the integration of toxin hexamers. Therefore, from

these experiments it was understood that the hemolytic activity of the CEL III protein was a temperature dependent, non-enzymatic process that takes place upon oligomerization on the target erythrocyte membrane. Although experiments have not yet been conducted on the pore forming mechanism of Chrysaoralin protein, it can be predicted by simply comparing its sequence with those of previously characterized CEL III protein. It has been established that amino acid sequence determines protein structure and structure dictates biochemical function. Therefore, proteins that contain amino acid sequence homology usually perform similar biochemical functions, even when they are found in distantly related organisms, such as in the case of Sea Cucumber CEL III and Sea Nettle Chrysaoralin.

Revelation of the Sequence Features

In 2014, Unno *et al.* revealed that the CEL III protein undergoes spontaneous oligomerization and conformational changes to create a transmembrane heptameric beta-barrel pore. The CEL III beta-barrel has a 75 Å height and 25 Å diameter. These dimensions are sufficient dimensions to make a transmembrane pore permitting small ions and molecules to pass across the cell membrane (Unno *et al.*, 2014). Because CEL III and Chrysaoralin share a high degree of sequence homology, the residues involved in oligomerization and transmembrane pore formation in CEL III may add to our understanding of Chrysaoralin pore formation. An alignment of annotated CEL III protein sequence with the Chrysaoralin protein sequence uncovers the domains conserved in both

CEL III and Chrysaoralin (Figure 22). The Sea Cucumber CEL III protein comprises of two ricin B-chain-like carbohydrate recognition domains and a C-terminal domain that is responsible for oligomerization in target cell membranes.

The N-terminal Ricin B-chain-like carbohydrate recognition domains are termed domains 1 and 2, while the C-terminal domain responsible for oligomerization is termed domain 3. Based on the CLUSTAL alignment and Phyre2 protein structure prediction, the domain 1 in Chrysaoralin extends from 23 to 173 amino acids, domain 2 from 164 to 307 amino acids, and domain 3 from 308 to 454 amino acids (Figure 20). The signal peptide constitutes the first 22 amino acids. Domains 1 and 2 in both CEL III and Chrysaoralin comprise of conserved metal ions and carbohydrate binding sites.

The CEL III protein is secreted as a monomer and during oligomerization domains 1 and 2 bind to the carbohydrates on the erythrocyte cell membrane and stabilize the pore by forming a large outer ring on the cell surface (Unno *et al.*, 2014). From Unno *et al*'s study on the CEL III protein, it was also discovered that domain 3 of monomeric CEL III contacts the side of domains 1 and 2 with its two alpha helices. Based on the sequences homology and Phyre2 prediction, "TVTVEVGASM" and "SVKVSTLSTA" amino acid sequences of the two alpha helices in domain 3 encode the two helices in Chrysaoralin (Figure 20, Figure 31). During heptamerization of the monomers, it is predicted that, the two helices transform to two long strands that assemble into a 14 stranded beta-barrel. The oligomerization process is described in great detail by Unno *et al* and a

supplemental movie is also available that illustrates oligomerization and pore formation (Unno *et al.*, 2014).

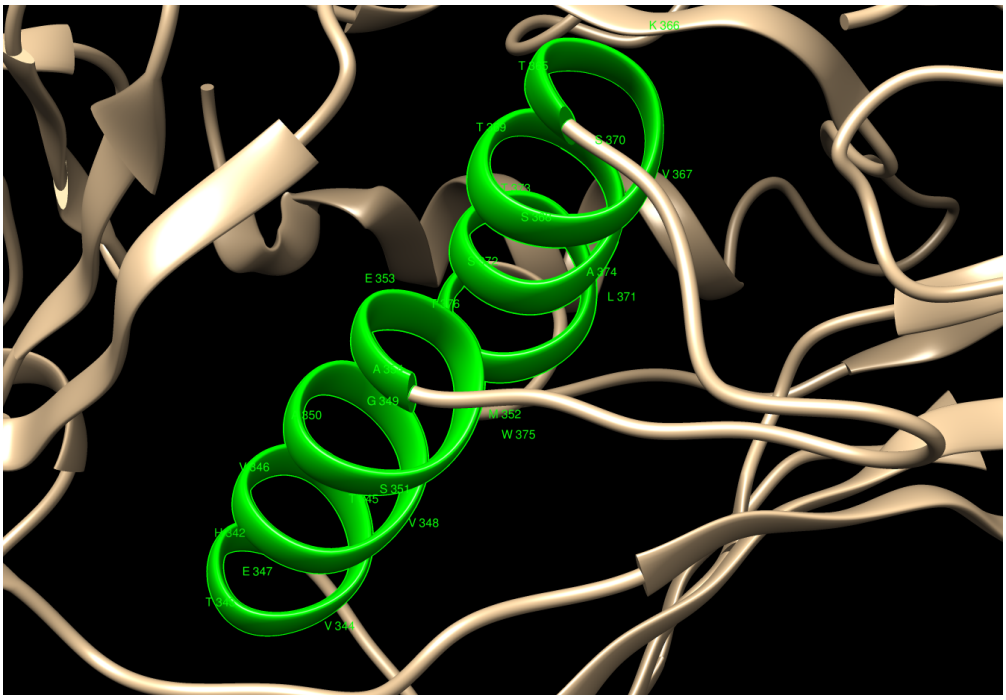


Figure 31: Protein model created using UCSF Chimera depicting two alpha-helices TVTVEVGASM (342-354) and SVKVSTLSTA (365-376) in domain 3 of Chrysaoralin.

Chrysaoralin protein was also compared to other pore forming beta-toxins such as hemolysin from *Staphylococcus aureus* hemolysin, *Vibrio Cholerae* cytolysin, Aeromonas hydrophila aerolysin, *Streptococcus pneumoniae* *p*neumolysin and *Bacillus anthracis* anthrax toxin for sequence similarity (data not shown). Although these proteins are similar to CEL III in terms of their pore size and mechanism of action, conservation of any functional domains of Chrysaoralin and CEL III with these toxins was not observed.

Although pore formers, the *C. echinata* CEL III protein differs from aforementioned pore forming toxins because of its property to undergo secondary structural change from alpha-helices to beta-strands during the formation of the membrane-spanning beta-barrel. This mechanism of pore formation is also observed in cholesterol-dependent cytolysin family proteins and membrane attack complex/perforin domains (Tilley *et al.*, 2005, Rosado *et al.*, 2007). The conservation of Chrysaoralin sequence with the CEL III sequence in the alpha-helical region could mean Chrysaoralin's pore forming mechanism is similar to that of CEL III.

Origin of the Chrysaoralin gene in *Chrysaora quinquecirrha*

To investigate the evolutionary origin of the Chrysaoralin gene, I searched for homologs of Chrysaoralin gene in the NCBI database (Figure 18). I did not observe many significant results that matched my gene of interest. Along with the Hemolytic lectin from Sea Cucumber, other various distantly-related eukaryote taxa shared the BLAST search results. The uncharacterized and hypothetical proteins of the acroporid coral *Acropora digitifera* (Accession number: XM_015908487) and the branching stony coral *Acropora millepora* (Accession number: EU863776) shared some sequence homologies with Chrysaoralin. Such sporadic distribution of Chrysaoralin like genes in eukaryotic homologues could suggest that this gene was acquired through independent horizontal gene transfer events. Additionally, the presence of this gene in phylogenetically distant organisms suggests that this gene is highly mobile. Availability of more closely

related sequences on the NCBI database could further our understanding of the evolutionary lineage of this gene.

Horizontal gene transfer, otherwise known as lateral gene transfer, is the transfer of DNA between organisms outside of traditional reproduction, which often results in adaptive gains of novel genes and traits. The acquisition of the full length Chrysaoralin gene, or some parts of the gene may have important consequences on *Chrysaora quinquecirrha*'s evolution.

Future Research

Chrysaoralin, a putative pore-forming protein, is one of the many toxins in *Chrysaora quinquecirrha*'s venom repertoire. The presence of this toxin might have improved prey immobilization and defensive functions. However, some fields of inquiry such as, the reasons for phylogenetically distant organisms like the sea cucumber and the sea nettle to share this toxin, examination of this gene in other species of jellyfish and specifically, in other *Chrysaora* species are yet to be investigated.

As the NCBI and venom databases amass more sequences from closely related species, new sequence features could be revealed in Chrysaoralin which may have been overlooked at the moment. Cloning and recombinant expression of the Chrysaoralin protein and testing its biological effect on different cell types could shed more light on its pore forming and cytolytic properties. Conducting functional assays using Chrysaoralin genes from both Barnegat and Chesapeake

Bays could reveal the significance of the two amino acid variation seen in the Chrysaoralin protein.

The presence of Chrysaoralin gene in other *Chrysaora* species is not yet known. By obtaining tentacles or DNA samples from other *Chrysaora* species, I could investigate the presence of this gene in other species. The availability of this gene in other *Chrysaora* species could also answer how, and possibly when, this gene may have been acquired in the course of this organism's evolution.

Conclusions

This study embraced both experimental and *in silico* techniques to identify and characterize a unique toxin peptide, Chrysaoralin, from the venom of *Chrysaora quinquecirrha*. Using Barnegat Bay *Chrysaora* mRNA transcripts generated by Next Generation sequencing as a scaffold, the genomic sequence for the Chesapeake and Barnegat Bay Chrysaoralin gene was successfully sequenced and assembled. Comparison of Barnegat Bay and Chesapeake Bay Chrysaoralin sequences resulted in a difference of 6 nucleotides, which translated to two amino acid differences in the immature protein. A striking homology of 64% at the amino acid level was seen between Chrysaoralin and a sea cucumber hemolytic lectin, CEL III. Conservation was observed in the different domains of the pore-forming proteins, particularly in the sugar-binding sites responsible for adhering to the target cell membrane, as well as in the alpha-helices that undergo conformational change to beta sheets during transmembrane pore formation. Finally, the Chrysaoralin gene was successfully cloned into the pET SUMO vector, which will permit expression of a fusion protein in *E. coli* to aid future studies on this important protein toxin.

References

Abrami L, Fivaz M, van Der Goot FG. 2000. Adventures of a pore-forming toxin at the target cell surface. Trends Microbiol. 8:168–172.

Allured VS, Collier RJ, Carroll SF. 1986. Structure of exotoxin A of *Pseudomonas aeruginosa* at 3.0-Ångstrom resolution. Proc. Natl. Acad. Sci. USA. 83:1320-1324.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402.

Athanasiadis A, Anderluh G, Macek P, Turk D. 2001. Crystal structure of the soluble form of equinatoxin II, a pore-forming toxin from the sea anemone *Actinia equina*. Structure 9:341-346.

Bischofberger, M, Iacovache I, Van der Goot FG. 2012. Pathogenic Pore-Forming Proteins: Function and Host Response. Cell Host Microbe. 12(3):266-75.

Bloom DA, Radwin FFY, Burnett JW. 2001. Toxinological and immunological studies of capillary electrophoresis fractionated *Chrysaora quinquecirrha* (Desor) fishing tentacle and *Chironex fleckeri* Southcott nematocyst venoms. Comp Biochem Physiol C. 128:75–90.

Bologna P, Gaynor J. 2013. Assessment of the distribution and abundance of Stinging Sea Nettles (Jellyfish) in Barnegat Bay. Final project report: 2013. NJDEP.

Gaynor JJ, Bologna PAX, Restaino D, Barry CL. 2016. First occurrence of the invasive hydrozoan *Gonionemus vertens* A. Agassiz, 1862 (Cnidaria: Hydrozoa) in New Jersey, USA. Bioinvasions Rec. 5(4):233-237.

Brahma RK, McCleary RJ, Kini RM, Doley R. 2015. Venom gland transcriptomics for identifying, cataloging, and characterizing venom proteins in snakes. Toxicon. 93:1–10.

Brodeur R, Decker M, Ciannelli L, Purcell J, Bond N, Stabeno P, Acuna E, Hunt G. 2008. Rise and fall of jellyfish in the eastern Bering Sea in relation to climate regime shifts. Prog. Oceanogr. 77:103-111.

Brinkman DL, Jia X, Potriquet J, Kumar D, Dash D, Kvaskoff D, Mulvenna J. 2015. Transcriptome and venom proteome of the box jellyfish *Chironex fleckeri*. BMC Genomics 16(1):407.

Calder D. 1972. Development of the sea nettle *Chrysaora quinquecirrha*. Chesapeake Science 13: 40-44.

Calvete JJ, Sanz L, Angulo Y, Lomonte B, Gutiérrez JM. 2009. Venoms, venomics, antivenomics. FEBS Lett. 583(11):1736–1743.

Cannon Q, Wagner E. 2003. Comparison of Discharge Mechanisms of Cnidarian Cnidae and Myxozoan Polar Capsules. Rev. Fish. Sci. 11(3):185-219.

Cao CJ, Eldefrawi ME, Eldefrawi AT, Burnett JW, Mioduszewski RJ, Menking DE, Valdes JJ. 1998. Toxicity of sea nettle toxin to human hepatocytes and the protective effects of phosphorylating and alkylating agents. Toxicon 36, 269-281.

Cascales E, Buchanan SK, Duche D, Kleanthous C, Lloubes R, Postle K, Riley M, Slatin S, Cavard D. 2007. Colicin biology. Microbiol. Mol. Biol. Rev. 71:158–229.

Cegolon L, Heymann WC, Lange JH, Mastrangelo G. 2013. Jellyfish stings and their management: A review. Mar. Drugs 11:523–550.

Chang H, Lim J, Ha M, Kim VN. 2014. TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. Mol Cell. 53(6):1044-52.

Choe S, Bennett MJ, Fujii G. 1992. The crystal structure of diphtheria toxin. Nature 357:216-222.

Clark GC, Briggs DC, Karasawa T, Wang X, Cole AR, Maegawa T, Jayasekera PN, Naylor CE, Miller J, Moss DS, Nakamura S, Basak AK, Titball RW. 2003. Clostridium absonum alpha-toxin: new insights into clostridial phospholipase C substrate binding and specificity. J Mol Biol. 31;333(4):759-69.

Cobbs CS, Gaur PK, Russo AJ, Warnick JE, Calton GJ, Burnett JW. 1983. Immunosorbent chromatography of sea nettle (*Chrysaora quinquecirrha*) venom and characterization of toxins. Toxicon 21:385-391.

Condon RH, Decker MB, Purcell JE. 2001. Effects of Low Dissolved Oxygen on Survival and Asexual Reproduction of Scyphozoan Polyps (*Chrysaora quinquecirrha*). Hydrobiologia 451:85-95.

De S, Olson R. 2011. Crystal structure of the Vibrio cholerae cytolysin heptamer reveals common features among disparate pore-forming toxins. Proc. Natl. Acad. Sci. USA. 108:7385-7390.

Dudkina NV, Spicer BA, Reboul CF, Conroy PJ, Lukoyanova N, Elmlund H, Law RHP, Ekkel SM, Kondos SC, Goode RJA, Ramm G, Whisstock JC, Saibil JC, Dunstone MA. 2016. Structure of the Poly-C9 Component of the Complement Membrane Attack Complex. Nat.Commun. 7:10588.

Engel S, Jensen PR, Fenical W. 2002. Chemical ecology of marine microbial defense. J. Chem. Ecol.28:1971–1985.

Fautin DG. 2009. Structural diversity, systematics, and evolution of cnidae. Toxicon 54:1054–1064.

Feil SC, Polekhina G, Gorman MA, Parker MW. 2010. Proteins: Membrane Binding and Pore Formation. Landes Bioscience and Springer Science. New York.

Ford MD, Costello JH, Heilderberg KB. 1997. Swimming and feeding by the scyphomedusa *Chrysaora quinquecirrha*. Mar. Biol. 129:355–362.

Fontes W, Sousa MV., Aragao JB, Morhy L. 1997. Determination of the amino acid sequence of the plant cytolysin enterolobin. Arch. Biochem. Biophys. 347 (2):201-207.

Fox JW, Serrano SM. 2008. Exploring snake venom proteomes: multifaceted analyses for complex toxin mixtures. Proteomics. 8:909–920.

Frazão B, Antunes A. 2016. Jellyfish Bioactive Compounds: Methods for Wet-Lab Work. Mar. Drugs. 14(4):75.

Galinier R, Portela J, Mone Y, Allienne JF, Henri H, Delbecq S, Mitta G, Gourbal B, Duval D. 2013. Biomphalysin, a New beta Pore-forming Toxin Involved in Biomphalaria glabrata Immune Defense against Schistosoma mansoni. PLoS Pathog. 9(3):E1003216.

Hadders MA, Beringer DX, Gros P. 2007. Structure of C8a-MACPF reveals mechanism of membrane attack in complement immune defense. Science. 317:1552-1554.

Holstein T, Tardent P. 1984. An ultrahigh-speed analysis of exocytosis: nematocyst discharge. Science. 223:830–3.

Iacovache I, van der Goot FG, Pernot L. 2008. Pore formation: an ancient yet complex form of attack. Biochim. Biophys. Acta 1778:1611–1623.

Iacovache I, De Carlo S, Cirauqui N, Dal Peraro M, van der Goot FG, Zuber B. 2016. Cryo-EM structure of aerolysin variants reveals a novel protein fold and the pore-formation process. Nat. Commun. 7:12062.

Jeffares DC, Penkett CJ, Bähler J. 2008. Rapidly regulated genes are intron poor. Trends. Genet. 24(8):375-8.

Jouiaei M, Yanagihara AA, Madio B, Nevalainen TJ, Alewood PF, Fry BG. 2015. Ancient Venom Systems: A Review on Cnidaria Toxins. Toxins. 7:2251–2271.

Kafsack BF, Pena JD, Coppens I, Ravindran S, Boothroyd JC, Carruthers VB. 2009. Rapid membrane disruption by a perforin-like protein facilitates parasite exit from host cells. Science. 323:530–533.

Kagan BL. 2012. Membrane pores in the pathogenesis of neurodegenerative disease. Prog. Mol. Biol. Transl. Sci. 107:295–325.

Kaneko J, Kamio Y. 2004. Bacterial two-component and hetero-heptameric pore-forming cytolytic toxins: structures, pore-forming mechanism, and organization of the genes. Biosci. Biotechnol. Biochem. 68:981–1003.

Kelman SN, Calton GJ, Burnett JW. 1984. Isolation and partial characterization of a lethal sea nettle (*Chrysaora quinquecirrha*) mesenteric toxin. Toxicon. 22:139-144.

Koester S, Pee KV, Hudel M, Leustik M, Rhinow D, Kuehlbrandt W, Chakraborty T, Yildiz O. 2014. Crystal Structure of Listeriolysin O Reveals Molecular Details of Oligomerization and Pore Formation. Nat. Commun. 5:3690.

Kristan KC, Viero G, Dalla Serra M, Macek P, Anderluh G. 2009. Molecular mechanism of pore formation by actinoporins. Toxicon. 54:1125– 1134.

Lakey JH, van der Goot, FG, Pattus F. 1994. All in the family: the toxic activity of pore-forming colicins. Toxicology. 87:85–108.

Law RH, Lukoyanova N, Voskoboinik I, Caradoc-Davies TT, Baran K, Dunstone MA, D'Angelo ME, Orlova EV, Coulibaly F, Verschoor S, Browne KA, Ciccone A, Kuiper MJ, Bird PI, Trapani JA, Saibil HR, Whisstock JC. 2010. The structural basis for membrane binding and pore formation by lymphocyte perforin. Nature. 468:447-451.

Lee H, Jung ES, Kang C, Yoon WD, Kim JS, Kim E. 2011. Scyphozoan jellyfish venom metalloproteinases and their role in the cytotoxicity. Toxicon. 58:277–284.

Li J, Carrol J, Ellar DJ. 1991. Crystal structure of insecticidal delta-endotoxin from *Bacillus thuringiensis* at .5 Å resolution. Nature 353:815-821.

Li R, Yu H, Xue W, Yue Y, Liu S, Xing R, Li P. 2014. Jellyfish venomics and venom gland transcriptomics analysis of *Stomolophus meleagris* to reveal the toxins associated with sting. J. Proteomics. 106:17–29.

Liu G, Zhou Y, Liu D, Wang Q, Ruan Z, He Q, Zhang L. 2015. Global transcriptome analysis of the tentacle of the jellyfish *Cyanea capillata* using deep sequencing and expressed sequence tags: Insight into the toxin- and degenerative disease-related transcripts. PloS one. 10:102.

Long-Rowe KO, Burnett JW. 1994. Characteristics of hyaluronidase and hemolytic activity in fishing tentacle nematocyst venom of *Chrysaora quinquecirrha*. Toxicon. 32(2):165-74.

Lozano I. 2013. Studies on the Cnidocyst of the Atlantic Sea Nettle, *Chrysaora quinquecirrha*. Master's Thesis. Montclair State University.

Mancheño MJ, Tateno H, Sher D, Goldstein IJ. 2010. Proteins: Membrane Binding and Pore Formation, edited by Gregor Anderluh and Jeremy Lakey. Chapter 6. Landes Bioscience and Springer Science+Business Media.

Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Lu S, Marchler GH, Song JS, Thanki N, Yamashita RA, Zhang D, Bryant SH. 2012. CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Res. 41:D348.

Mariottini GL, Pane L. 2013. Cytotoxic and cytolytic cnidarian venoms. A review on health implications and possible therapeutic applications. Toxins. 6:108–151.

Mariottini GL. 2014. Hemolytic venoms from marine cnidarian jellyfish—an overview. J. Venom. Res. 5:22–32.

Mayer, AG. 1910. Medusae of the world, Vol. III. The scyphomedusae. Washington, DC: Carnegie Institution. 109(3): 585-588.

Menestrina G, Dalla Serra M, Comai M, Coraiola M, Viero G, Werner S, Colin DA, Monteil H, Prevost G. 2003. Ion channels and bacterial infection: the case of beta-barrel pore-forming protein toxins of *Staphylococcus aureus.* FEBS Lett. 552:54–60.

Montoya M, Gouaux E. 2003. Beta-barrel membrane protein folding and structure viewed through the lens of alpha-hemolysin. Biochim. Biophys. Acta. 1609:19–27.

Neeman I, Calton GJ, Burnett JW. 1980a. Cytotoxicity and dermonecrosis of sea nettle (*Chrysaora quinquecirrha*) venom. Toxicon. 18:55–63.

Neeman I, Calton GJ, Burnett JW. 1980b. An ultrastructural study of the cytotoxic effect of the venoms from the sea nettle (*Chrysaora quinquecirrha*) and Portugese man-of-war (*Physalia physalis*) on cultured Chinese hamster ovary K-1 cells. Toxicon. 18:495-501.

Newman-Martin G. 2007. Manual of envenomation and poisoning: Australian fauna and flora. Defence Publishing Service; Canberra, Australia.

Özbek S, Balasubramanian PG, Holstein TW. 2009. Cnidocyst structure and the biomechanics of discharge. Toxicon. 54(8):1038-1045.

Parker MW, Feil SC. 2005. Pore-forming protein toxins: from structure to function. Prog. Biophys. Mol. Bio. 88:91–142.

Parker MW, Buckley JT, Postma JP, Tucker AD, Leonard K, Pattus F, Tsernoglou D. 1994. Structure of the Aeromonas toxin proaerolysin in its water-soluble and membrane-channel states. Nature. 367(6460):292–5.

Pechenik JA. 2000. Biology of the Invertebrates. Edition, 4. Publisher, McGraw-Hill Education. ISBN-10: 0072484136.

Pédelacq J-D, Maveyraud L, Prévost G. 1999. The structure of a *Staphylococcus aureus* leucocidin component (LukF-PV) reveals the fold of the water-soluble species of a family of transmembrane pore-forming toxins. Structure. 7:277-287.

Petosa C, Collier RJ, Klimpel KR, Leppla SH, Liddington RC. 1997. Crystal structure of the anthrax toxin protective antigen. Nature. 385:833-838.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. J. Comput. Chem. 25:1605-1612.

Pineda SS, Wilson D, Mattick JS, King GF. 2012. The Lethal Toxin from Australian Funnel-Web Spiders Is Encoded by an Intronless Gene. PLoS ONE 7(8): e43699.

Polekhina G, Giddings KS, Tweten RK, Parker MW. 2005. Insights into the action of the superfamily of cholesterol-dependent cytolysins from studies of intermedilysin. Proc. Natl. Acad. Sci. USA. 102:600-605.

Ponce D, Brinkman DL, Luna-Ramírez K, Wright CE, Dorantes-Aranda JJ. 2015. Comparative study of the toxic effects of *Chrysaora quinquecirrha* (Cnidaria: Scyphozoa) and *Chironex fleckeri* (Cnidaria: Cubozoa) venoms using cell-based assays. Toxicon. 106:57–67.

Ponce D, Brinkman DL, Potriquet J, Mulvenna J. 2016. Tentacle Transcriptome and Venom Proteome of the Pacific Sea Nettle, *Chrysaora fuscescens* (Cnidaria: Scyphozoa). Toxins. 8(4):102.

Proudfoot NJ, Brownlee GG. 1976. 3′ Non-coding region sequences in eukaryotic messenger RNA. Nature. 263:211–214.

Purcell JE, White JR, Nemazie DA, Wright DA. 1999. Temperature, salinity, and food effects on asexual reproduction and abundance of the scyphozoan *Chrysaora quinquecirrha*. Mar. Ecol. Prog. Ser. 180:187-196.

Purcell JE, Arai, MN. 2001. Interactions of pelagic cnidarians and ctenophores with fish: a review. Hydobiologia. 451:27–44.

Purcell, JE. 2012. Jellyfish in Chesapeake Bay and Nearby Waters. The Shore Journal.

Rosado CJ, Buckle AM, Law RH, Butcher RE., Kan WT, Bird CH, Ung K, Browne KA, Baran K, Bashtannyk-Puhalovich TA, Faux NG, Wong W, Porter CJ, Pike RN, Ellisdon AM, Pearce MC, Bottomley SP, Emsley J, Smith AI, Rossjohn J, Hartland EL, Voskoboinik I, Trapani JA, Bird PI, Dunstone MA, and Whisstock JC. 2007. A common fold mediates vertebrate defense and bacterial attack. Science. 317:1548 –1551.

Sakurai N, Kaneko J, Kamio Y, Tomita T. 2004. Cloning, expression, and pore-forming properties of mature and precursor forms of pleurotolysin, a sphingomyelin-specific two-component cytolysin from the edible mushroom *Pleurotus ostreatus*. Biochim. Biophys. Acta 1679(1):65-73.

Sher D, Fishman Y, Zhang M, Lebendiker M, Gaathon A, Mancheno JM, Zlotkin E. 2005. Hydralysins, a new category of beta-pore-forming toxins in cnidaria. J. Biol. Chem. 280(24):22847-22855.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7;539.

Song L, Hobaugh MR, Shustak C, Cheley S, Bayley H, Gouaux JE. 1996. Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. Science. 274:1859-1866.

Sousa MV, Richardson M, Fontes W, Morhy L. 1994. Homology between the seed cytolysin enterolobin and bacterial aerolysins. J. Protein. Chem. 13: 659–667.

Technau U, Steele RE. 2011. Evolutionary crossroads in developmental biology: Cnidaria. Development (Cambridge, England). 138(8):1447–1458.

Teragawa CK, Bode HR. 1995. Migrating interstitial cells differentiate into neurons in hydra. Dev. Biol. 171:286-293.

Tilley SJ, Orlova EV, Gilbert RJ, Andrew PW, Saibil HR. 2005. Structural basis of pore formation by the bacterial toxin pneumo- lysin. Cell. 121:247–256.

Tomita N, Abe K, Kamio Y, Ohta M. 2011. Cluster-forming property correlated with hemolytic activity by staphylococcal g-hemolysin transmembrane pores. FEBS Lett. 585:3452–3456.

Tweten RK, Parker MW, Johnson AE. 2001. The cholesterol-dependent cytolysins, Curr. Top. Microbiol. Immunol. 257:15–33.

Uchida T, Yamasaki T, Eto S, Sugawara H, Kurisu G, Nakagawa A, Kusunoki M, Hatakeyama T. 2004. Crystal structure of the hemolytic lectin CEL-III isolated from the marine invertebrate *Cucumaria echinata*: implications of domain structure for its membrane pore-formation mechanism. J. Biol. Chem. 279:37133–37141.

Uechi G., Toma H., Arakawa T., Sato Y. 2010. Molecular characterization on the genome structure of hemolysin toxin isoforms isolated from sea anemone *Actineria villosa* and *Phyllodiscus semoni*. Toxicon. 56:1470–1476.

Unno H, Goda S, Hatakeyama T. 2014. Hemolytic Lectin CEL-III Heptamerizes via a Large Structural Transition from α-Helices to a β-Barrel during the Transmembrane Pore Formation Process. J. Biol. Chem. 289(18):12805-12812.

Wallace AJ, Stillman TJ, Atkins A, Jamieson SJ, Bullough PA, Green J, Artymiuk PJ. 2000. *E. coli* hemolysin E (HlyE, ClyA, SheA): X-ray crystal structure of the toxin and observation of membrane pores by electron microscopy. Cell. 100:265–276.

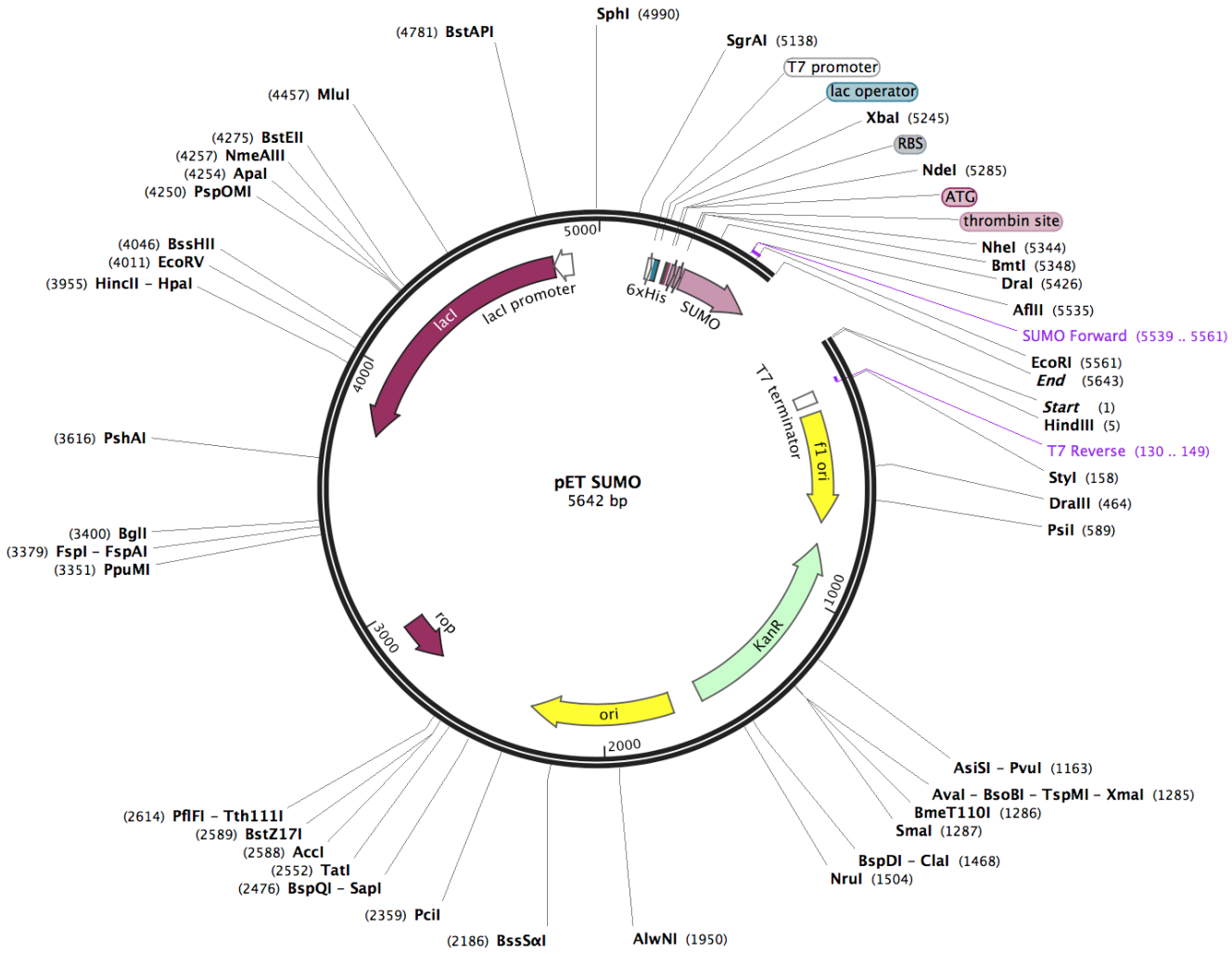
Watson GM, Hessinger DA. 1989a. Cnidocytes and adjacent supporting cells form receptor-effector complexes in anemone tentacles. Tissue Cell. 21:17–24.

Watson GM, Hessinger DA. 1989b. Cnidocyte mechanoreceptors are tuned to the movements of swimming prey by chemoreceptors. Science. 243:1589–1591.

Wedekind JE, Trame CB, Dorywalska M, Koehl P, Raschke TM, McKee M, FitzGerald D, Collier RJ, McKay DB. 2001. Refined crystallographic structure of Pseudomonas aeruginosa exotoxin A and its implications for the molecular mechanism of toxicity. J. Mol. Biol. 314:823-837.

Yan HK, Yuan ZG, Zhou Y, Zhu XQ. 2011. Sequence variation in *Toxoplasma gondii* perforin-like protein 1 gene. ACCESSION: AEO96991. Unpublished.

Appendix A | pET SUMO Vector



Appendix B | List of Primers for Sequence Analysis

Primer List		
SEQUENCE NAME	PRIMER SEQUENCE	Tm °C
pMiniTF	ACCTGCCAACCAAAGCGAGAAC	63.8
pMiniTR	TCAGGGTTATTGTCTCATGAGCG	60.4
ChrysF	ATGGATCAAATACGCTTGATTGGTG	60.2
CQ R	GAGAAACGGCAGCAATTAATGTCAG	61.2
SUMO F	AGATTCTTGTACGACGGTATTAG	55.7
T7 R	TAGTTATTGCTCAGCGGTGG	57.7
FL-SPF	CAAGTCCTTGTGCACCAATCCG	61.8
CTXD3F	ATGTGCATGCCGTGCTTTACC	62.1
CTXD3R	TTAAATTTCCTGGCACAGGCT	58.1
DOM3F	TGGACAGCTCCTAATGCTGTT	59.4
DOM3R	TCAGATTTCTTGGCAAAGAG	53.5

Appendix *B:* List of primers used for sequence analysis. pMiniTF and pMiniTR primers were used to analyze cloning of the gene within the pMiniT vector. ChrysF and CQR primers generate full length Chrysaoralin gene. SUMOF and T7R primers are used to analyze cloning of the gene within the pET SUMO vector. FL-SPF and CQR generate full length of the gene minus the signal peptide. DOM3F and DO3R primers amplify only domain 3 of the Chrysaoralin gene.

Alkaline Lysis Plasmid Mini-Prep Protocol

Harvesting:

1. Transfer a single bacterial colony into 2 ml of LB medium containing the appropriate antibiotic in a loosely capped 15 ml tube. Incubate the culture overnight at 37 °C with vigorous shaking.

2. Transfer 1.5 ml of the overnight culture into a microfuge tube. Centrifuge for 30 s at 12,000 x g in a microfuge. Store the remainder of the culture at 4°C.

3. Remove the medium, leaving the bacterial pellet as dry as possible.

Lysis by Alkali:

1. Resuspend the bacterial pellet (obtained in step 3 above) in 100 μl of ice-cold Solution I by vigorous vortexing.

2. Add 200 μl of freshly prepared Solution II. Close the tube tightly, and mix the contents by inverting the tube rapidly. Make sure that the entire surface of the tube comes in contact with Solution II. Do not vortex. Store the tube on ice for 5 minutes.

3. Add 150 μl of ice-cold Solution III. Close the tube and vortex it gently in an inverted position for 10 s to disperse Solution III through the viscous bacterial lysate. Store the tube on ice for 3-5 min.

4. Centrifuge at 12,000 g for 5 minutes in a microfuge. Transfer the **supernatant** to a fresh tube.

5. Add 450 μl of phenol : chloroform. Mix by vortexing. After centrifuging at 12,000 g for 2 minutes in a microfuge, transfer the aqueous phase to a fresh tube.

6. Add 450 μl of chloroform. Mix by vortexing. After centrifuging at 12.000 g for 2 minutes in a microfuge, transfer the aqueous phase to a fresh tube.

7. Precipitate the double-stranded DNA with 2 volumes of ethanol at room temperature. Mix by vortexing. Allow the mixture to stand for 5 minutes at 4°C.

8. Centrifuge at maximum speed for 5 minutes at 4 °C in a microfuge.

9. Remove the supernatant.

10. Rinse the pellet of double-stranded DNA with 1 ml of 70% ethanol. Remove the supernatant, and dry the pellet briefly in a Speed-Vac.

11. Resuspend the pellet in 50 μl of sterile TE (pH 8.0) containing DNAse-free RNAse A (20 μg/ml). Finger mix or vortex briefly. Store the DNA at -20°C.

12. Sacrifice 2 μL for a Nanodrop reading for OD_{260} and $OD_{260/280}$.

Solution I
50 mM glucose
25 mM Tris-Cl (pH 8.0)
10 mM EDTA (pH 8.0)
Solution I can be prepared in batches of approximately 100 ml, autoclaved for 15 min on liquid cycle and stored at 4° C.

Solution II (make fresh each time – do not store solution)
0.2 N NaOH (freshly diluted from a 10 N stock)
1% SDS (20 μl 10 N/ml, 50 μl 20%/ml)

Solution III (can be stored at 4° C)
5 M potassium acetate 60 ml
glacial acetic acid 11.5 ml
H_2O 28.5 ml