

# EXPRESSION AND PURIFICATION OF AUGERTOXINS: SEARCHING FOR NOVEL PROTEIN FOLDS IN VENOMOUS MARINE SNAILS

Kacey A. Davis (Martin P. Horvath, PhD) Department of Biology

## ABSTRACT

This study describes a method for bacterial expression and purification of previously uncharacterized proteins. The proteins chosen for this study come from auger snail toxins, which have evolved to help the snail hunt and kill their prey. Evolutionary pressure between predator and prey selects for diverse toxin proteins with new functions. New functions could be accomplished by repurposing pre-existing proteins within the snail to become toxins or by developing completely new proteins, potentially with novel folded structures. Well over 10,000 known species of venomous marine snails, each with distinct toxins containing hundreds of protein components, represent a rich source of potentially novel protein folds (Olivera et al. 2014). However, toxin proteins are difficult to harvest from small snails and challenging to chemically synthesize if larger than ~35 residues. Overexpression of toxin protein genes in bacteria allows for large amounts of folded, functional protein, without predicted limitations on size. This study selected five auger snail toxins, augertoxins, for expression and purification. The augertoxins ranged in size from 40 residues to 150 residues for the mature toxin. Four of five chosen toxins were expressed successfully, and one of those was further purified to give pure, testable toxin protein. Future work will further characterize pure protein using x-ray crystallography to determine the folded structure and biological assays to explore relevant functions. My foundational work on an optimal bacterial expression system can be applied to other uncharacterized toxin proteins and will help the search for new folds and functions.

ii

# TABLE OF CONTENTS

ABSTRACT	ii
INTRODUCTION	1
METHODS	9
RESULTS	13
DISCUSSION	33
ACKNOWLEDGMENTS	38
REFERENCES	39

#### INTRODUCTION

Predatory marine snails are found in every ocean and produce venoms that can kill worms, fish, other snails, and even humans. The venoms are made of a medley of toxins, small molecules and proteins that work together to disable the snail's prey. As prey evolves to evade the snail, evolution selects for snails with more effective and innovative toxins (Olivera et al. 2017). Evolutionary pressure for snails to outcompete prey results in a wealth of new proteins with novel functions. Some proteins have been explored, but the vast majority of venomous marine snail toxins have yet to be characterized. This project investigates a method for expression and purification of uncharacterized marine snail toxin proteins using *E. coli* as an expression host.

Marine snails have evolved venoms to defend against predators and become predators themselves (Duda and Palumbi 1999; Casewell et al. 2013). The protein and peptide components of snail venom have responded to the evolutionary pressure to catch new prey by targeting diverse receptors in the prey's nerve cells more efficiently (Olivera et al. 2012; Espiritu et al. 2001). The combination of diversity and efficiency results in proteins that selectively target specific subtypes of receptors (Terlau and Olivera 2004; Cruz, Johnson, and Olivera 1987). One protein from a fish hunting snail toxin, for example, targets a single receptor subtype expressed in the muscular system of fish. Many such proteins and peptides work together to hunt and paralyze the snail's prey.

Some marine snail toxins have gathered interest as potential drug candidates.  $\omega$ conotoxin MVIIA, a protein from *Conus magus*, targets a type of calcium channel responsible for pain signaling in humans and is an FDA-approved non-narcotic pain reliever that is sold under the names Ziconotide or Prialt, meaning primary alternative to

1

morphine (Puillandre and Holford 2010; Terlau and Olivera 2004; Richard J. Lewis and Garcia 2003). Other toxin proteins from *Conus geographus* and *Conus tulipa* act like a "weaponized insulin" to induce hypoglycemic shock in their prey and could be useful to treat diabetes in humans (Safavi-Hemami et al. 2015; Safavi-Hemami, Lu, et al. 2016; Robinson and Safavi-Hemami 2016). These examples hold the promise that uncharacterized snail toxins, the vast majority of the hundreds of thousands known to exist, represent a goldmine of undiscovered drugs and research tools.

The specificity of these proteins can help classify neurons according to the receptor subtypes they express, using a method developed by the Olivera Lab at the University of Utah called constellation pharmacology (Teichert et al. 2012; Teichert, Schmidt, and Olivera 2015). Constellation pharmacology can determine each neuron's "constellation" of receptors based on the molecules or proteins it responds to. Precise classification of neurons with this method is possible in part because of the specificity of marine snail toxin proteins. For example, proteins from *Conus catus* have been shown to discriminate between subtypes of calcium channel receptors (R J Lewis et al. 2000). Understanding the receptors of a specific cell type could help researchers design more effective drugs that target only the cells implicated in disease, rather than the current practice of using one molecule that broadcasts to many cell types (Teichert et al. 2012). Marine snail toxins enable valuable research about types of neurons, which helps map the human brain and refine treatments for everything from Alzheimer's disease to depression.

This goldmine of newly evolved proteins found in the venom of marine snails holds promise for discovering new protein folds as well as new functions. The function of a protein depends on the folding pattern, which is encoded by the sequence of amino acids (Anfinsen 1973). Toxin proteins are notable for their highly conserved framework of cysteine residues that form disulfide bonds to help direct and maintain interesting folds (Robinson et al. 2016). Predatory marine snail's rapidly evolving toxin sequences and stable disulfide bonds make them an evolutionary incubator for potentially new protein folds (Olivera et al. 1999, 2012).

Most snail toxin research has focused on toxins isolated from snails in the Conidae family that includes *Conus* species; these are called conotoxins. The abundance and diversity of the ~800 species of cone snails in Conidae provides ample research material to hunt for new drugs, neuron types, and protein folds. However, many other species of mollusks produce toxins. The 300-400 species of snails in the neighboring Terebridae family also have exciting research potential and are surprisingly understudied (Imperial et al. 2003; Verdes et al. 2016; Gorson et al. 2015). If novel folds are waiting to be characterized, uncharted territory seems a logical place to look. To explore the idea that Terebridae snail venom may harbor novel protein folds, this study chose five toxin protein genes isolated from the terebrids *Hastula solida* or *Terebra subulata* (see Figure 1) to express and purify.

Terebrid snails are found in all tropical waters and mostly feed on worms (Chang, Lin, and Chen 2007). They are nicknamed auger snails for their shell's resemblance to an auger drill bit – their shells are long with an exaggerated spiral, quite distinct from the cone snail's compact shell, seen in Figure 1 (Imperial et al. 2003). Terebrids are evolutionary neighbors to cone snails and their toxin proteins, called augertoxins or occasionally teretoxins, are stabilized by disulfide bonds and face the same pressure to fold and function in new ways (Puillandre and Holford 2010; Gorson et al. 2015).



**Figure 1:** The sources of selected augertoxins compared to other mollusks in the Conoidea superfamily. (A) Conidae species. (B) Turridae species. (C) Drillidae species. (D) Terebridae species. *Terebra subulata* and *Hastula solida*, the sources for the five candidate toxins, are circled. Scale bar is 1 cm, all specimen to scale. Credit J.S. Imperial *et al.*, 2003 and CF Chang *et al.*, 2007.

Research on augertoxins is limited, but studies testing for biological activity indicate that they are also a rich source of novel proteins with potentially therapeutic benefits. For example, fractions from *Terebra subulata* (Figure 1) venom induced uncoordinated twisting when injected into the nematode *Caenorhabditis elegans* (Imperial et al. 2003). The same compound had no effect when injected into mice. Venom extracts from *T. consobrina* and *T. argus* showed strong inhibitory effects on acetylcholine receptors expressed in frog oocytes when tested with a two-electrode voltage clamp technique (Kendel et al. 2013). Biological activity assays such as those performed in these two studies show that some fractions of auger snail venom have biological function.

Both of these augertoxin studies were done with samples collected from the venom ducts of the snails themselves and were limited by availability of the venom (Kendel et al. 2013; Imperial et al. 2003). Terebrids can be quite small – they are difficult to find in the ocean and it is nearly impossible to dissect any useable venom duct sample. *H. solida*, the source of four candidate proteins for this study, is only 1 cm long (Figure 1)! Also, snail venom is made up of many different components, only some of which are active proteins. Isolating active proteins from the medley of toxin components by fractionating venom is inefficient, nonspecific, and relies on scarce material. Venom fractionation is an impractical method of testing for active toxin components among the tens of thousands of unexplored venomous snail toxins.

Venom components that are difficult to isolate from the snail that makes them can now be studied due to recent advances in sequencing technologies (Fry 2005; Gorson et al. 2015). Small amounts of mRNA from a snail's venom duct can be sequenced to give a snapshot of the proteins usually expressed there. mRNA sequences are reversetranscribed to give a cDNA library full of sequences that can then be used to make the toxin proteins in the lab. Traditionally, conotoxins were made by chemical synthesis, but that route is impractical for larger proteins more than ~35 residues or toxins with extensive permutations of disulfide bonds between cysteines (Bulaj 2005; Safavi-Hemami et al. 2015; Imperial et al. 2003). Another option is to express cDNA sequences in bacteria, where the endogenous translation machinery favors the proper combination of cysteines in large enough quantities to allow further research (Bulaj 2005).

My research used bacteria to express five augertoxins. The toxins chosen vary in species of origin, size, and cysteine framework. Four of the toxin protein sequences were inferred from cDNA of *Hastula solida* and the proteins were completely uncharacterized. One toxin was from *Terebra subulata* and has previously been fractionated from the venom duct, sequenced, and found to have activity in *C. elegans* (Imperial et al. 2003). The proteins are named for species of origin and cysteine framework and they range in size from 4.3 kDa to 17 kDa. Table 1 lists each toxin protein with its sequence and characteristics including expected mass.

Name and species of origin	Protein sequence <sup>1</sup>	Amino acids <sup>2</sup>	Molecular weight <sup>2</sup> (Da)	pI
HSd 6.3 from Hastula solida	ENLYF <b>Q^G</b> VPEETDGLLELYKNYARR <b>M^</b> GK NCHSSGQVCDIGSVAFKCCRGYECRPDTTK GTCQTEN	59	6521.3	6.73
HSd 9.1 from Hastula solida	ENLYFQ^GRDLDTDGPARRDRADRNLLSIL TRRDYVPLLRSQRTHEAVKPPRIQRM^AYV PASTTTTAAPTPDPYSECTKCEEKTADDCP SLEYDCKPTVYKECSPC	99	11176.5	6.76
HSd 29.1 from Hastula solida	ENLYFQ^GFEQNCTKHKYIRPCGNTEPCSH KKSGPDGCDVYINCKCGGGRRCQDNQDRSV QARHCKKQGVSHTFTFCSELNDLSLATCMT GSAIVIEGNKSESRPNSEINCLCTEENELV YIETGRKWDVYCQPFGVSNICN	135	15023.8	6.99
HSd 29.8 from Hastula solida	ENLYFQ^GDPCEKAVFGCLESHFKLTSHVQ EEVDTKCRVFHEQGFNACTKDLIAKCRDGY QWAMGLLNEIGRCYCTEDVVTAVKENLACR IGDEYVAAVAPCYTLEKTSKCSFAKALRDC IFEEVDNYCPTYRKVLEISHDFVFRLLKCH ITDSPMC	150	17041.5	5.53
s7a from Terebra subulata	ENLYFQ^GPGVSLNLM^ATNRHQCDTNDDC EEDECCVLVGGNVNNPGVQTRICLACS	40	4297.68	4.14

Table 1: Protein sequences and characteristics for the five selected augertoxin proteins.

<sup>1</sup> Each character is one amino acid residue according to the single-letter abbreviation system. ENLYFQ<sup>A</sup>G (green) define the TEV protease recognition site with peptide bond cleavage site marked <sup>^</sup>, pink is toxin sequence with pro-region underlined, CNBr cleavage site marked <sup>^</sup>, and cysteine residues are in black. <sup>2</sup> For mature toxin obtained after cleavage with TEV protease or after CNBr cleavage (s7a).

The toxins were also chosen for the presence of a pro-region, a sequence of amino acids that is predicted to aid in folding but is not present in the mature toxin (Olivera et al. 1999). The pro-region is conserved among families of toxins in different species of snails even when the mature toxin sequence is highly divergent; more on this idea will follow in the Discussion. Smaller toxins in cone and auger snails nearly always have a pro-region, but larger augertoxins do not. HSd 29.1 and HSd 29.8 selected for this study lack a pro-region, while a pro-region was found in HSd 6.3, HSd 9.1, and s7a (see Table 1). The exact function of the pro-region is a topic of much debate and speculation and this open question was one of the motivating factors for the selection of augertoxins in my thesis project.

Augertoxins represent a goldmine of undiscovered proteins that could hold answers to open questions about novel protein folds and functions. My thesis describes a complete method for bacterial expression and purification of HSd 6.3 that can be adapted to express other uncharacterized augertoxins. The framework developed in my research was designed to explore the pro-region and biological activity of selected augertoxins. Efficient production of these understudied proteins helps answer questions about the folds and function of the toxins themselves, but my research also has broad applications for the evolution of proteins – these ideas will be further explored in the Discussion. My thesis lays the foundation that enables further research about predatory marine snail toxins and the origins of protein folds.

#### **METHODS**

#### Construction of plasmid

Candidate toxin gene sequences were identified from Terebridae cDNA libraries with the help of Maren Watkins from Baldomero Olivera's lab in the Biology department. The genes were formatted into GeneBlocks, selecting codons that are frequently encountered in bacteria (Table 2) and these GeneBlock sequences were ordered from Integrated DNA Technologies (IDT, Coralville Iowa). GeneBlock DNA was incorporated into a T7promoter driven expression plasmid by ligation independent cloning (LIC). Briefly, left hand and right hand fragments of the MBP-pET22b bacterial expression plasmid were amplified via PCR with use of DNA primers carrying sequences that overlapped with each GeneBlock. For the left hand fragment, PCR buffer included 4% DMSO, and the reaction proceeded through 30 cycles of 96°C 10 sec denature, 58°C 10 sec anneal, and 73°C 210 sec elongation. For the right hand fragment, PCR buffer included 2% DMSO, and the reaction proceeded through 30 cycles of 96°C 10 sec denature, 58°C 10 sec anneal, and 72°C 150 sec elongation. Products were purified from a 0.8% agarose gel using a GeneJet gel extraction kit (Thermofisher) by following the instructions provided by the manufacturer. Pure PCR products ( $5\mu L$  of combined left hand and right hand fragments) and GeneBlock DNA (5µL of DNA diluted 1/10,000) were added to 80µL of heat-shock competent DH5a on ice. After heat-shock for 45 sec, cells were briefly placed on ice, transferred to SOB liquid media with 5mM MgCl2, and shaken at 37°C for 2 hours. The cells were plated onto LB plates with  $100\mu g/mL$  ampicillin and incubated at 37°C until colonies formed. Colonies were transferred to liquid media, grown at 37°C, and harvested for DNA. Plasmids were harvested using standard mini prep procedure. 
 Table 2: GeneBlock sequences ordered for each toxin.

Toxin	GeneBlock <sup>1</sup>
HSd 6.3	gcagactcgtatcaccaagtgggcgaagaaatattcaaacgcaatcctgagc <u>GA</u> AAACCTGTATTTCCAGGGCGTCCCAGAGGAGACCGACGGCCTGCTGGAACTGTA CAAAAACTACGCGCGCCGCATGGGTAAAAACTGCCACTCTTCCGGTCAGGTCTG CGATATCGGCTCTGTGGCTTTCAAATGCTGCCGCGGGTTATGAATGCCGCCCGGA <u>CACCACCAAGGGTACTTGCCAGACTGAAAACTAA</u> ggatccgaattcgagctccg tcgacaagcttgcggccgcactcgagcaccaccac
HSd 9.1	gcagactcgtatcaccaagtgggcgaagaaatattcaaacgcaatcctgagc <u>GA</u> AAACCTGTACTTTCAGGGTCGTGATCTGGACACGGACGGCCCAGCACGTCGTGA CCGTGCGGATCGTAACCTGCTGTCCATCCTGACCCGCGTGACTACGTGCCGCT GCTGCGTTCCCAGCGCACCCATGAGGCAGTGAAGCCGCCGCGGTATCCAGCGCAT GGCATACGTTCCGGCCAGCACCACCACCACCACCGCCGCGCCGACCCCAGATCCATA TAGCGAATGTACCAAATGTGAAGAAAAAACCGCAGACGATTGCCCGTCTCTGGA ATATGACTGTAAACCAACCGTTTATAAAGAATGCTCTCCGTGTTAAggatccga attcgagctccgtcgacaagcttgcggccgcactcgagcaccacca
HSd 29.1	gcagactcgtatcaccaagtgggcgaagaaatattcaaacgcaatcctgagc <u>GA</u> AAACCTGTACTTCCAGGGTTTCGAACAGAACTGCACCAAACACAAGTACATCCG TCCGTGTGGCAACACTGAACCGTGTTCTCACAAAAAATCCGGTCCGGATGGCTG TGACGTTTACATCAACTGTAAATGCGGCGGTGGCCGTCGTTGCCAGGATAACCA GGACCGTAGCGTGCAAGCACGTCACTGCAAGAAACAGGGTGTTTCGCATACTTT TACTTTTTGCAGCGAACTGAACGACTTGAGCCTGGCGACCTGCATGACTGGCTC CGCAATCGTGATTGAGGGCAACAAGAGCCGAAAGCCGCCCGAACTCTGAAATTAA CTGCCTGTGTACCGAAGAACGAACTGGTTTACATTGAAACCGGGCGTAAATG GGATGTGTACCGAAGAACGAACTGGTGTCTAACATTGAAACCGGGCGTAAATG attcgagctccgtcgacaagcttgcggccgcactcgagcaccacca
HSd 29.8	gcagactcgtatcaccaagtgggcgaagaaatattcaaacgcaatcctgagc <u>GA</u> GAACCTGTACTTTCAGGGTGACCCGTGTGAAAAAGCTGTCTTCGGCTGCCTGGA ATCGCACTTCAAGTTAACCTCTCATGTTCAAGAAGAGGTGGATACCAAATGCCG TGTATTTCACGAACAAGGGTTCAACGCATGCACCAAAGACCTCATTGCGAAATG TCGTGATGGTTATCAGTGGGCTATGGGCCTGCTGAATGAA
s7a	gcagactcgtatcaccaagtgggcgaagaaatattcaaacgcaatcctgagc <u>GA</u> GAACCTGTATTTCCAGGGCCCAGGTGTGTCCCTGAACTTAATGGCTACTAATCG TCACCAGTGCGACACCAATGACGATTGTGAAGAAGATGAATGCTGCGTTCTGGT AGGGGGTAATGTGAACAACCCAGGCGTTCAGACGCGTATCTGCCTGGCGTGTAG <u>CTAA</u> ggatccgaattcgagctccgtcgacaagcttgcggccgcactcgagcacc accac

<sup>1</sup>Underlined sequence indicates Geneblock, lowercase is overhang with left or right hand fragments.

## Expression of MBP-toxin fusion protein

Pure DNA was transformed into heat-shocked competent Origami B *E. coli* cells. Cloning was tested by inducing protein expression with 100µM isopropyl β-D-1thiogalactopyranoside (IPTG) and visualizing protein via 9.5% SDS-PAGE using Horvath lab protocols. The complete circular plasmid was purified from individual colonies using standard mini prep protocol and sequenced to confirm results (HSC core facility). The Origami B cell cultures were grown overnight in 2mL of 2XYT media in a 37°C shaking incubator from glycerol stocks (overnight cultures from individual colonies combined with 30% glycerol, stored at -80°C). Cells were pelleted, washed, and resuspended in 60mL 2XYT for 3-4 hours. Cell cultures were again pelleted, washed, and resuspended in 750mL or 1000mL 2XYT and cultured until optical density reached 0.8 at 600nm measured on a UV spectrophotometer. Expression of plasmid genes was induced with 100µM IPTG for 15 hours at 17°C. Cells were harvested by washing and centrifuging and stored at -20°C for up to several months.

## Amylose affinity chromatography

Cell pellets were sonicated in 35 mL of Buffer A-ACB (25mM Tris, 100mM NaCl, 0.1mM EDTA, 0.02% azide, pH 8) and centrifuged at 15,000 rpm for 20 minutes. The supernatant was processed through a 0.45µM filter and dripped slowly using a serological pipet onto a freshly packed or regenerated 1.5 mL amylose affinity column. The supernatant was divided into two halves, each ~18 mL, and processed on the same column sequentially. Buffer M (Buffer A-ACB plus 10mM maltose, pH 8) eluted the MBP-toxin fusion protein. Fractions M1 and M2 from each half were pooled. Yield was estimated via UV/VIS spectrophotometer measuring at 280nm and 220nm.

#### Cleavage reaction with TEV or CNBr

TEV protease was provided by Heidi Schubert and Katherine Ferrell in the Hill lab at the University of Utah. To digest with TEV, 1/30 ratio by weight TEV to protein, 1 M urea, and 0.3 M DTT was added to the pooled M fractions and allowed to proceed for 24 hours at room temperature. For proteins with chemical cleavage, cyanogen bromide was added to pooled M fractions in a 1/50 ratio by weight, digested for 24 hours in the dark at room temperature and the reaction was stopped by lyophilization at HSC Core Peptide facilities (Andreev et al. 2010).

## Purification of the toxin

To purify the putative toxin, the cleavage products were separated by size using a HiLoad Superdex 75 16/60 size exclusion column on an Akta FPLC. Buffer A-ACB was used as the chromatography buffer. A C18 reverse phase HPLC column and MonoQ ion exchange FPLC were also used in attempts to purify s7a. Protein identity and the status of disulfide bonds was confirmed by mass spectrometry (HSC core facility, University of Utah).

#### Biological assay using C. elegans

Bus-8(e2887) worms were ordered from the Worm Bank and kept on OP50 bacteria plates at 22.5°C. The worms were passaged every 3-4 days. Thrashing assays were performed on gravid hermaphrodites by picking up a worm into a 50 µL drop of M9 buffer (22 mM KH2PO4, 42 mM Na2HPO4, 86 mM NaCl) and counting thrashes, according to standard procedure (Nawa and Matsuoka 2012).

#### RESULTS

Characterization of novel protein folds begins with large amounts of the pure protein. Part of my thesis explored a way to hijack bacteria to manufacture an excess of target protein. The five selected augertoxin proteins were overexpressed as part of a much larger protein, called an affinity tag, to create a toxin fusion protein. The affinity tag helped separate the toxin fusion protein from other proteins expressed in bacteria. The toxin was cut from the affinity tag by cleaving a bond at the beginning of the toxin protein. To purify the toxin and remove other products from the cleavage reaction, the products were separated based on size. Expression of candidate proteins in bacteria and purification using affinity and size exclusion columns yielded pure protein that could be tested on an animal model.

#### Selection of host bacteria strain

The host bacterium is in charge of overexpressing and accurately folding the protein of interest. Augertoxins are usually expressed in venom ducts of snails, where an oxidizing environment and chaperone proteins help form stabilizing disulfide bonds between cysteine residues (Safavi-Hemami, Li, et al. 2016; Safavi-Hemami et al. 2010). Bacterial cytoplasm is typically reducing and prevents disulfide bond formation (Bulaj 2005). Routing expression to the oxidizing periplasm of a wild-type *E. coli* would overcome the issue, but rerouting is inefficient (Klint et al. 2013). Previous work determined that expression of marine snail toxins in Origami B (OriB), a mutant *E. coli* with an oxidizing cytoplasm, yielded significantly more soluble product than expression in BL21, a different host strain with a wild-type cytoplasm (Allie Fredo, personal

communication). OriB was chosen as the host bacteria strain to help ensure the correctly folded, biologically active toxin protein was favored in this study.

## Design of bacterial expression plasmid

Plasmids were designed to include several elements, illustrated in Figure 2: an inducible promoter to easily over-express the toxin protein; an affinity tag to aid in purification of the protein; a cleavage recognition site to cut the affinity tag off; and the toxin protein-encoding gene. The plasmid also contained an ampicillin resistance gene that enabled the bacteria with the plasmid to grow in selective conditions. The plasmid was designed by Alyssa Fredbo in the Horvath lab. A fusion protein design previously worked well to produce large amounts of properly folded conotoxin from *C. striatus*, expressed by Alyssa Fredbo, and another toxin from a spider (Klint et al. 2013; Bende et al. 2013).

## Selection of affinity tag

Expressing a protein in high concentrations often results in insoluble protein, which is difficult, if not impossible, to work with. To address this concern, I included an affinity tag attached to the toxin protein to help maintain solubility and aid in the folding process, overcoming challenges expected if the toxin was expressed without an affinity tag. Maltose binding protein (MBP) is an affinity tag that also assists during purification of the augertoxin by binding amylose attached to a solid matrix while bacterial proteins are washed away. Expressing the affinity tag MBP fused to the toxin protein resulted in an MBP-toxin fusion protein that was predicted to be soluble, folded, and easily purified.



**Figure 2:** Constructing a bacterial expression system for MBP-toxin fusion proteins. (A) Three overlapping sections of plasmid were joined together during replication in the host bacterium, DH5 $\alpha$ . After the full plasmid containing genes for the full MBP-toxin fusion protein and ampicillin resistance is purified from DH5 $\alpha$ , it is transformed into the expression host, OriB. Transcription and translation in OriB results in fully folded MBP-toxin fusion protein. (B) Agarose gel showing the products of PCR for the left hand and right hand fragments. (C) Confirmation of cloning via SDS-PAGE. The lanes marked MBP express only the MBP protein without the toxin. Lanes numbered 1-8 each analyze the crude extract from a different clone transformed with MBP-toxin plasmid DNA. Positive clones were inferred by presence of a dark protein band at a higher molecular weight than MBP, indicating an MBP fusion protein may be expressed. Clones in lanes labeled with a star were further confirmed as positive by DNA sequence.

## Selection of cleavage agents

The covalently linked MBP and toxin protein needed to be cleaved and separated by further purification steps to produce pure toxin. My research employed two methods in an orthologous cleavage system where two or more products were possible depending on the toxin and the cleavage method used. Cleavage could be accomplished using either tobacco etch virus (TEV) protease, an enzyme, or cyanogen bromide (CNBr), a chemical agent. TEV protease was used to cleave between MBP and the toxin, leaving the proregion, if present, still attached. TEV protease is highly selective for the sequence ENLYFQ $^X$ , where X is a glycine or serine that remains on the N-terminus of the toxin. Other equally selective proteases, such as human rhinovirus 3C (HRV3C, or PreScission) protease, leave two, larger amino acid residues that are more likely to impact folding and function of the final product, which is why I opted for the TEV protease recognition site. CNBr drives a chemical reaction that cuts the peptide bond C-terminal to methionine residues (Gross 1967). In my system, an added methionine encoded between the proregion and the toxin protein gave the option to produce toxin without the pro-region (see Table 1). Both CNBr and TEV protease acted as molecular scissors to cleave the peptide bond between MBP and the toxin.

The pro-region of predatory marine snail toxins is predicted to help fold the mature toxin, but the exact function is yet unclear. The orthologous cleavage design was meant to help understand the purpose of the pro-region. By having the option to explore the protein folds and functions of the toxins with or without the pro-region, the exact role in folding structure or biological function could be examined. The orthologous cleavage system ended up being quite problematic, as will be described in later sections.

#### Construction of expression plasmid DNA via ligation independent cloning

The plasmid was constructed in three fragments in a process known as ligation independent cloning. The 'left hand' fragment contained the gene for MBP, the 'right hand' contained an ampicillin resistance gene, and the GeneBlock contained the toxin (Figure 2A). The GeneBlock was designed to encode the toxin sequence using codons that are commonly encountered in bacteria to ensure efficient expression. The left hand and right hand fragments were amplified using PCR; the products are shown in Figure 2B.To construct the complete expression plasmid, the three DNA fragments were transformed into DH5 $\alpha$  *E. coli*. Inside the bacteria, DNA repair enzymes pasted the three fragments into one continuous circular DNA molecule. Constructed DNA expression plasmids were purified and transformed into OriB *E. coli* prior to confirmation by testing for protein expression and DNA sequencing.

Only bacteria containing the complete plasmid – both the ampicillin resistance gene (right hand fragment) and the full MBP-toxin gene (left hand and GeneBlock) – could grow in ampicillin-containing media and produce the fusion protein. OriB bacteria that contained a correctly cloned plasmid isolated from DH5 $\alpha$  were selected by their ability to express a fusion protein. The MBP-toxin fusion protein is slightly larger than MBP without the toxin. An SDS-PAGE with all proteins stained by Coomassie Blue indicated some bacteria expressed a protein with a slightly higher molecular weight than MBP, as shown in Figure 2C. Multiple bacterial clones were tested for each candidate toxin. OriB that produced a fusion protein supposedly contained the complete plasmid, which was later confirmed by DNA sequencing and mass spectrometry of purified protein.

#### *Making a monster*

Verifying the bacteria could produce a fusion protein was one indication of successful cloning but sequencing the plasmid DNA was crucial to avoid making 'monsters.' Figure 3A shows an SDS-PAGE of bacterial clones transformed with the HSd 9.1 GeneBlock. A shift in molecular weight (red arrow) indicated expression of a fusion protein. The sequencing results of DNA from these apparently successful bacterial clones is provided in Figure 3B. The translated sequence shows the expected WAKKYSNAILS linker, but no TEV recognition site or toxin. The plasmid was ligated with left hand and right hand fragments only and appears to have completely skipped the GeneBlock fragment. Closer inspection of the SDS-PAGE in Figure 3B reveals that the putative fusion protein actually had a molecular weight of ~65 kDa, much too high to be the MBP-HSd 9.1 fusion protein (expected mass = 54 kDa). Interestingly, the putative 65kDa MBP fusion protein was effectively cleaved with TEV, shown in Figure 3C. The cleavage reaction gave a product similar in size to MBP and a product  $\sim 20$  kDa. The monsters were discarded after sequencing. The two-step verification process – protein induction and DNA sequencing – was crucial, as seen from the monster cloning attempt of HSd 9.1.



**Figure 3:** Making a monster. (A) Visualization of protein induction test via SDS-PAGE for 12 bacterial clones containing the three DNA fragments to make MBP-HSd 9.1. Red arrow indicates putative fusion protein. (B) Electropherogram for plasmid DNA from clone shown in lane 10 of (A). Translated protein sequence is above the DNA sequence. The linker sequence, DNA coding for the amino acids WAKKYSNAILS, is present, but in place of toxin-encoding DNA, DNA corresponding to the right hand fragment is linked, encoding amino acids that do not correspond with toxin protein (shown in red). (C) Putative fusion protein from clone 10 in (B) was purified and cleaved with TEV to yield a protein with a similar molecular weight as MBP and a very faint blurred band at ~20 kDa.

#### Optimization of expression

Plasmids containing each MBP-toxin protein gene were expressed at various times and temperatures to test parameters important for optimal expression. Protein expression was quantified by the band size and stain on SDS-PAGE. Timepoints were taken for two MBP-toxin proteins, HSd 29.1 and HSd 29.8, at 17°C, 25°C (room temperature), and 37°C. Prior to my work expressing augertoxins, bacterial expression at 17°C for 24 to 28 hours was the standard Horvath lab practice for MBP-toxin fusion proteins and was predicted to give the most protein. However, Figure 4 shows that a 15 hour induction period at 17°C gave noticeably more MBP-toxin fusion protein than any other timepoint or temperature tested. Protein expression for 15 hours at 17°C became the new standard condition for optimal expression of MBP-toxin fusion proteins in bacteria. *Purification via affinity chromatography* 

The overexpressed MBP-toxin fusion proteins were purified by affinity chromatography. Figure 5A illustrates an MBP-toxin fusion protein attached to the beads of the amylose column while contaminants are washed away in the Flowthrough (FT) fraction. Each fraction was analyzed by SDS-PAGE, shown in Figure 5B. The MBPtoxin fusion protein is bound to the amylose column, the stationary phase, until eluted with maltose-containing buffer (fractions M1-M3). The maltose buffer binds to the maltose binding protein and carries it into the mobile phase and off of the column. Some MBP-toxin fusion protein was lost in the wash (W) step before eluting with maltose buffer. The loss was minimized when samples were loaded very slowly on either newly packed columns or freshly regenerated columns.



**Figure 4:** Optimization of expression time at 17°C for MBP-HSd 29.1 fusion protein. Samples were taken after induction with IPTG at the timepoints indicated. For each timepoint, the sample was sonicated to yield total protein from the cell (TP) and treated with amylose beads to isolate the MBP-toxin fusion protein (beads). A darker band indicates a higher concentration of protein. Although much of the MBP-toxin protein did not bind the beads in this small scale assay, the TP indicates highest expression at 15 hours after induction.



**Figure 5:** Purification by affinity chromatography. (A) The diagram shows soluble protein isolated from OriB loaded onto an amylose column. MBP-toxin fusion protein binds to amylose beads while other proteins are washed away in the Flowthrough fraction. (B) Each fraction from the purification of MBP-HSd 6.3 compared to MBP alone visualized via SDS-PAGE. Total protein (TP) is all protein from the bacteria. SP is soluble protein, including MBP-toxin fusion protein (darkest band in each lane). The Flowthrough (FT) fraction contained most of the contaminating bacterial proteins. The column was washed (W), and MBP-toxin protein was eluted with maltose buffer (M1, M2, M3). The high molecular weight band in fractions M1 M2 M3 probably reflect the tendency of MBP to dimerize at high concentration, even under denaturing conditions. Fractions M1 and M2 were pooled, being most enriched for MBP-toxin fusion protein.

#### Optimization of MBP-toxin peptide bond cleavage

Purified MBP-toxin fusion protein was cleaved with TEV protease. Varying amounts of TEV protease were added to the MBP-toxin fusion protein in three different buffer solutions to optimize cleavage of the MBP-toxin fusion proteins. Cleavage efficiency was measured by the intensity of the MBP and toxin bands on SDS-PAGE after treatment with TEV. The main determinant of final toxin concentration was the amount of TEV protease added to the reaction: Figure 6 shows that as concentration of TEV increased, cleavage became more complete and more HSd 29.1 toxin was in solution. A second parameter that impacted cleavage efficiency related to the redox potential of the buffer solution. TEV requires some reducing agent to effectively cleave the peptide bond, but the disulfide bonds in the toxin will denature if the buffer is too reducing. Two different reduction/oxidation reagents were added: a mixture of reduced glutathione (GSH) and oxidized glutathione (GSSG) was compared to a buffer containing dithiothreitol (DTT). Each reaction gave similar amount of product and both were more effective than no redox reagents at all. Optimization of the two parameters for cleavage of HSd 29.1 and HSd 29.8 led to standard conditions of 1/30 TEV in a buffer with 0.3mM final concentration of DTT for each MBP-toxin protein. Time, temperature, and pH conditions were previously optimized by Heidi Schubert in the Hill lab to be overnight at room temperature in pH 8 buffer.



**Figure 6:** Optimization of proteolysis. Top: Cleavage of HSd 29.1 from MBP with increasing concentrations of TEV protease, shown on SDS-PAGE. Bottom: an illustration of MBP-toxin fusion protein cleaved with TEV, freeing the toxin protein from its affinity tag.

Cyanogen bromide was used to cleave MBP-toxin s7a, which did not cleave efficiently with TEV. Length of reaction and relative amount of CNBr was optimized. Although s7a was the smallest of the five candidates and was not visible via SDS-PAGE, efficiency of cleavage could be estimated by amount of other fragments expected after CNBr treatment, as seen in Figure 7. Typically, 20-100 fold excess CNBr by weight is recommended, and approximately 50 fold excess produced apparent cleavage and became the standard in this study (Andreev et al. 2010). Several timepoints were tested; a 24 hour reaction in a dark cabinet at room temperature seemed to give the most product. CNBr was difficult to work with due to safety concerns, which will be further discussed in the Discussion.

#### *Purification of toxin by size-exclusion chromatography (SEC)*

Fast protein liquid chromatography (FPLC) was used to separate proteins based on size. Figure 8 shows an FPLC trace and SDS-PAGE for the separation of MBP and putative HSd 6.3 toxin after cleavage with TEV protease. The blue trace is absorbance at 280 nm, which detected multiple unresolved peaks of protein in fractions 6 through 13 and one small peak in fractions 16 and 17. This complex FPLC profile is consistent with components of the protease reaction: the undigested MBP-toxin fusion protein most likely eluted first, then MBP and TEV, and finally the toxin eluted in fractions 16 and 17 (labeled in Figure 8). The five toxins selected for this study had very few, if any, aromatic tryptophan and tyrosine residues and were difficult to detect at 280 nm, but further analysis by SDS-PAGE helped determine the presence and molecular weight of the eluted protein. SEC worked well to purify toxins with molecular weights of 6-12 kDa.



**Figure 7:** Optimization of CNBr cleavage. The 4.3 kDa s7a toxin was too small to visualize via SDS-PAGE, but a 16.5 kDa fragment of MBP may indicate positive cleavage. Other expected fragments are 10.5 kDa, 6.7 kDa, 6.3 kDa, 2.1 kDa, 1 kDa, and 0.7 kDa.



**Figure 8:** Purification of HSd 6.3. (A) FPLC size exclusion chromatography. Blue trace indicates absorbance at 280 nm. Each fraction collected is 3 mL, numbered and shown in red on the x axis. HSd 6.3, the smallest protein in the mix, presumably eluted last. MBP-toxin, MBP, and TEV are larger and eluted earlier. Magenta trace is conductance and indicates included volume. (B) Analysis of eluted fractions by SDS-PAGE. FT is a sample of the Flowthrough from the concentrator, +TEV is after proteolysis but before SEC. Lanes are numbered according to fractions as labeled in the FPLC chromatogram (Panel A). A very faint band corresponding to the molecular weight of the smallest product from the TEV reaction suggests HSd 6.3 eluted in fractions 16 and 17.

#### Attempted purification of s7a

Other methods of chromatography attempted to purify s7a, which was not successfully purified in SEC. The putative toxin was resolubilized after treatment with CNBr. Two trials with reversed-phase high performance liquid chromatography (HPLC) showed a peak for pure MBP, but no peak corresponding to s7a toxin protein – and only MBP was visible when the fractions were visualized on SDS-PAGE. Ion exchange FPLC was also not effective at separating putative s7a from MBP. The products from the resolubilized CNBr reaction eluted before binding to the ion exchange column in all three trials. An SDS-PAGE of fractions showed no protein. Interpretations of these attempted purifications of s7a will be further explored in the Discussion section.

## Confirmation of HSd 6.3 identity and quantification of yield

Fractions 16 and 17 from the SEC of putative HSd 6.3 was further analyzed by mass spectrometry to confirm the identity of the protein (Figure 9). The results show an observed mass of 6510.8899 Da, which is not a significant difference from the expected mass of 6510.9 Da if HSd 6.3 was folded with all cysteines in disulfide bonds. If all 6 cysteines were reduced, a mass of 6516.9 Da would be expected. The minor contaminant with a mass of 6354.8 Da matches the expected mass of the toxin if the first two amino acids (GV) were cleaved or degraded after proteolysis with TEV. Quantification with UV/VIS spectrophotometry at a wavelength of 220 nm indicated an average yield of 390 nmoles per 1 liter of bacterial culture. The expression and purification scheme developed during my honors research work resulted in mostly pure, folded HSd 6.3 toxin in pharmacologically relevant amounts.



Figure 9: Mass determination and identity confirmation of HSd 6.3. The expected monoisotopic mass with all 6 cysteines oxidized (folded, in disulfide bonds) is 6510.907 Da, which is in agreement with the observed mass for the major peak, 6510.8899 Da. The minor peak indicates a mass of 6354.8 Da, which would be expected if the first two amino acids (GV) were cleaved or degraded after TEV treatment. Mass spectrometry indicates HSd 6.3 was pure and correctly folded.

## Final status of each augertoxin

Final status of each augertoxin is summed in Table 3 with an arrow. HSd 6.3, 6.5 kDa and the smallest protein selected from *Hastula solida*, was the most successfully expressed and purified and is ready for biological activity assays. The next largest, HSd 9.1, was cloned incorrectly, creating a monster that was discarded after identified by sequencing. Future work on HSd 9.1 requires recloning and resequencing before expression in OriB. Attempts to sequence HSd 29.1 and 29.8 gave garbled reads and also need to be verified before continuing purification. s7a was lost at some point during CNBr treatment, resuspension, or HPLC/FPLC. Further work on s7a should focus on getting sufficient cleavage while maintaining solubility of the toxin.

	H9 6.1	Sd 3	H\$ 9.	Sd 1	H3 29	Sd 9.1	HS 29	5d .8	s7	a
Ligation independent cloning										
Protein expression test								,		
Sequencing					2					
Large scale expression										
Affinity chromatography										
Cleavage from MBP			ST	0P						
Size exclusion chromatography										
Mass spec		,								
Biological activity assay	G	0								

**Table 3:** Furthest progress for all five augertoxins studied.

#### Biological activity assay

A biological activity assay to monitor a change in an animal's behavior when exposed to the toxin could confirm the protein was folded and functional. *Caenorhabditis elegans* are microscopic worms that feed on bacteria; although they are not the same worms that Terebridae snails hunt in the ocean, they are a good place to begin to test for activity. The number of wiggles, or thrashes, is counted for one minute after the worm is placed in a solution containing the toxin (Nawa and Matsuoka 2012). A biologically active toxin targeting receptors in the nervous or muscular system is expected to change the worm's ability to thrash their body back and forth. An increase in the frequency of thrashes would indicate the neurons are overstimulated, for example, while a decrease would indicate the muscles are paralyzed. A challenge with this approach is that C. *elegans* are protected by a thick cuticle that is impermeable to all but small molecule drugs (O'reilly et al. 2014; Page and Johnstone 2007). To address this challenge, worms with a mutation in the bus-8 gene (e2889 allele) and a 'leaky cuticle' phenotype that may be more permeable were used (Gravato-Nobre et al. 2005; Partridge et al. 2008). The bus-8 mutant worms had severely impaired motor function and high mortality rates. Many worms died as a result of the transfer to the control buffer solution, and the rest died before being subjected to the HSd 6.3 toxin. Although my current bioactivity results are inconclusive, this assay has promise for determining the function of predatory marine snail toxins.

#### DISCUSSION

Hundreds of toxin proteins from over 10,000 venomous snails are waiting to be characterized. Previously impossible-to-study toxins from the miniscule venom ducts of Terebridae snails are now ready to express or synthesize thanks to recent advancements in sequencing technologies (Casewell et al. 2013). Now more than ever, efficient expression systems are a valuable way to generate relevant amounts of difficult to synthesize proteins. My thesis has explored a way to express augertoxins in bacteria, purify using affinity and size exclusion chromatography, and apply biological activity assays or x-ray crystallography to discover novel protein folds and functions.

The system of expression and purification discussed in my thesis requires some optimization for each target toxin protein. Toxins approximately 6-12 kDa with fewer cysteines seem to be optimal candidates for bacterial expression. The bacterial expression system was innovated in part to help overcome the size restriction in place for chemically synthesized toxins; while my research suggests that efficient expression and purification may be limited to toxins fewer than ~100 amino acid residues, this restriction is still significantly higher than the ~35 residue maximum for chemically synthesized toxins. The extra room to express larger proteins gives researchers the ability to explore even more predatory marine snail toxins and ask more questions.

Although the orthologous cleavage design feature was attractive on paper, in practice it was a nightmare. CNBr cleavage was particularly troublesome due to chemical safety concerns and poor behavior of the final product. Handling the volatile and dangerous CNBr required caution and special equipment at the U's Peptide Core facility to stop the reaction by lyophilization. Final purification of s7a was complicated by the solubilization step after lyophilization. If soluble s7a did make it out of the CNBr reaction, it was even more difficult to purify due to its small size; FPLC size exclusion worked well with toxins that were big enough to elute in included volume and small enough to separate from other proteins. Other methods of chromatography such as ion exchange FPLC failed to purify the protein because the resuspended sample was too salty to bind the column, even after desalting procedures. Reversed-phase HPLC also gave little product, again possibly due to complications with solubility after CNBr treatment. The process of lyophilization could have decreased solubility by destabilizing the toxin. Lyophilization has been shown to rearrange disulfide bonds when an aromatic residue is nearby (I. Roy and Gupta 2004; Esfandiary et al. 2016). Adding sugar to the buffer or changing freezing conditions could improve recoverability in the future (Jiang and Nail 1998; Ugwu and Apte 2004). For future orthologous cleavage experiments to test the function of the pro-region, I recommend encoding two different protease sites instead of including a chemical cleavage site.

The biological activity assay in bus-8 mutant *C. elegans* relied on monitoring the movements of a single microscopic worm. The bus-8 mutants had a much lower survival and reproductive rate than wild-type worms, and their ability to move was also severely impaired. In the future, the worm's reaction to a solution containing the drug or protein could be monitored using a camera and computer algorithm, greatly increasing efficiency and decreasing human error (O 'reilly et al. 2014). Coupled with new technology, bus-8 mutant *C. elegans* could be advantageous for high-throughput biological function testing of many different marine snail toxin proteins. Other potential biological activity assays include testing the toxin on neurons as part of the constellation pharmacology studies

mentioned in the Introduction. Additionally, injecting the toxin into mice or zebrafish could uncover interesting biological functions.

The long term goal of my work was to determine the structures and contribute to mapping the protein folding universe. Crystallization of expressed and purified toxin proteins would generate a 3D picture of the protein folds. Although decoding the sequence of a protein is a simple matter, extrapolating the folded structure is more difficult – a challenge known as the protein folding problem (Dill and MacCallum 2012; Service 2008). Some protein structures can be predicted by comparing the sequence to known proteins, but that only works if there are known structures of proteins with a similar sequence. For many toxins, including the five selected for this study, there are no known sequence homologs. The folded pattern cannot be predicted. However, proteins can have a similar structure without having homologous sequences. Crystallization would identify structural homologs, if they exist. If they do not, then the toxins could be classified as novel folds. Discovery of novel protein structures can help address the protein folding problem by contributing new sequences and types of folds. Future work on characterizing toxins through x-ray crystallography would add to databases such as the Protein Data Bank (PDB), Homology-Derived Secondary Structure of Proteins (HSSP), and Structural Classification of Proteins (SCOP) that are collecting folds to help solve the protein folding problem (Zhang and Skolnick 2005; Sander and Schneider 1991; Murzin et al. 1995; Skolnick et al. 2003).

New sequences and folds added to the databases informs algorithms for the prediction of protein structure. A worldwide prediction conference held every two years invites researchers to submit structure predictions for proteins with unreleased structures. The 12<sup>th</sup> Critical Assessment of Structure Prediction (CASP12) in 2016 received a special issue of the journal PROTEINS and showed that prediction based on structure homologs has greatly improved compared to previous CASPs (Moult et al. 2017). CASP12 was more successful in part due to advances in deep-learning computer modeling and improved versions of programs such as Rosetta, which both require known structures to accurately predict new ones (Schaarschmidt et al. 2018; Ovchinnikov et al. 2018). CASP13 is opening in April 2018. Other algorithms use more pedestrian power: FoldIt is a game that empowers citizens to predict the structure of proteins. Interestingly, FoldIt players and researchers converged on a very similar algorithm to solve structures (Khatib et al. 2011). The most effective algorithms still require structural homologs – this is where novel toxin folds could add a missing piece to the puzzle. Groundbreaking work to crack the amino acid to fold code is currently happening and could be greatly aided by exploring uncharacterized predatory marine snail toxins.

If there are novel folds yet to be discovered, predatory marine snail toxins are a great place to look. Predatory marine snails and their toxins are rapidly evolving. Toxins from *C. abbreviatus* with highly dissimilar amino acid sequences have recently diverged from a common duplicated gene, reflecting a trend of rapid mutation and diversification seen throughout predatory marine snail toxins (Duda and Palumbi 1999). Not all parts of the toxin mutate at the same rate, however. The signal sequence and pro-region are highly conserved while the mature toxin sequence evolves quickly (Olivera et al. 1999, 2012). The variable rates of evolution could be a result of hypermutability, an increased rate of mutation, in the mature toxin compared to the pro-region. Alternatively, the rate of selection could make the difference between the regions, rather than rate of mutation (S.

W. Roy 2016). The pro-region is an area predicted to help fold the mature toxin but is cleaved off before the toxin reaches its target. So, while mature toxins are under strong positive selective pressure to diversify and reach more targets, the pro-region may lack pressure to select for new mutations. Whether increased mutation or increased selection is to blame, the sheer number and diversity of venomous marine snails and their toxins is an evolutionary breeding ground for new protein folds.

New folds could come from new combinations of structural motifs, common sections of protein that usually have a specific biological function (Fernandez-Fuentes, Dybas, and Fiser 2010; Jones and McGuffin 2003). Over the course of evolution, genes get shuffled, duplicated, or mutated to code for proteins with new combinations of motifs and eventually, new folds and functions (Lupas, Ponting, and Russell 2001; Söding and Lupas 2003). The building blocks for new proteins are smaller, functional proteins or peptides (Höcker 2014). Predatory marine snail venom is composed of proteins that fit that description. In fact, a novel fold has already been found from a marine snail toxin (Robinson et al. 2016). My thesis has laid the foundation for characterization of never-before-seen proteins with potentially novel folds.

#### ACKNOWLEDGEMENTS

I would like to thank Dr. Martin Horvath for his continued support and guidance. I have learned a lifetime in the past four years not only from research, but from your insight and advice. Thank you to Alyssa Fredbo for teaching me everything about conkunitzins (and for the skis). I am grateful to Dr. Toto Olivera for support for this project. I would also like to thank other members of the Olivera Lab, including Julita Imperial, Helena Safavi-Hemami, Pradip Bandyopadhyay, Maren Watkins, Peter Huynh, and countless others for their encouragement, advice, and expertise. Thank you to the Jorgenson Lab for assistance with C. elegans and my very own worm pick, specifically Robert Hobson and Patrick Mac. Thank you to my labmates in the Horvath Lab, who have made ASB my home: Evan Drage, Peyton Russelburg, Evan George, Sonia Sehgal, and Malika Kadirova. I owe my start in the Horvath Lab to Rosemary Gray and the ACCESS Program. The most significant experiences of my undergraduate career have been because of ACCESS and I am forever grateful to be a part of it. Thank you to my sisters, Bre and Natasha, for always being there for me. Finally, thank you to my roommates and friends who have helped me keep my sanity through the rollercoaster that is research.

#### REFERENCES

- Andreev, Yaroslav A., Sergey A. Kozlov, Alexander A. Vassilevski, and Eugene V. Grishin. 2010. "Cyanogen Bromide Cleavage of Proteins in Salt and Buffer Solutions." *Analytical Biochemistry* 407 (1). Elsevier Inc.: 144–46. doi:10.1016/j.ab.2010.07.023.
- Anfinsen, C B. 1973. "Principles That Govern the Folding of Protein Chains." *Science* (*New York, N.Y.*) 181 (4096): 223–30. http://www.ncbi.nlm.nih.gov/pubmed/4124164.
- Bende, Niraj S., Eunji Kang, Volker Herzig, Frank Bosmans, Graham M. Nicholson, Mehdi Mobli, and Glenn F. King. 2013. "The Insecticidal Neurotoxin Aps III Is an Atypical Knottin Peptide That Potently Blocks Insect Voltage-Gated Sodium Channels." *Biochemical Pharmacology* 85 (10): 1542–54. doi:10.1016/j.bcp.2013.02.030.
- Bulaj, Grzegorz. 2005. "Formation of Disulfide Bonds in Proteins and Peptides." *Biotechnology Advances* 23 (1). Elsevier: 87–92. doi:10.1016/J.BIOTECHADV.2004.09.002.
- Casewell, Nicholas R., Wolfgang Wüster, Freek J. Vonk, Robert A. Harrison, and Bryan G. Fry. 2013. "Complex Cocktails: The Evolutionary Novelty of Venoms." *Trends in Ecology and Evolution*. doi:10.1016/j.tree.2012.10.020.
- Chang, Chia-feng, Chih-ying Lin, and Ming-hui Chen. 2007. "Three New Records of Auger Shells (Mollusca : Gastropoda : Terebridae ) from Taiwan" Coll. and (20): 21–26.
- Cruz, Lourdes J, David S Johnson, and Baldomero M Olivera. 1987. "Characterization of the ω-Conotoxin Target. Evidence for Tissue-Specific Heterogeneity in Calcium Channel Types." *Biochemistry* 26 (3): 820–24. doi:10.1021/bi00377a024.
- Dill, Ken A., and Justin L. MacCallum. 2012. "The Protein-Folding Problem, 50 Years on." Science 338 (6110): 1042–46. doi:10.1126/science.1219021.
- Duda, T F, and S R Palumbi. 1999. "Molecular Genetics of Ecological Diversification: Duplication and Rapid Evolution of Toxin Genes of the Venomous Gastropod Conus." *Proceedings of the National Academy of Sciences* 96 (12). National Academy of Sciences: 6820–23. doi:10.1073/pnas.96.12.6820.
- Esfandiary, Reza, Shravan K. Gattu, John M. Stewart, and Sajal M. Patel. 2016. "Effect of Freezing on Lyophilization Process Performance and Drug Product Cake Appearance." *Journal of Pharmaceutical Sciences* 105 (4): 1427–33. doi:10.1016/j.xphs.2016.02.003.
- Espiritu, Doris Joy D, Maren Watkins, Virginia Dia-Monje, G. Edward Cartier, Lourdes J Cruz, and Baldomero M Olivera. 2001. "Venomous Cone Snails: Molecular Phylogeny and the Generation of Toxin Diversity." *Toxicon* 39 (12). Pergamon: 1899–1916. doi:10.1016/S0041-0101(01)00175-1.
- Fernandez-Fuentes, Narcis, Joseph M. Dybas, and Andras Fiser. 2010. "Structural Characteristics of Novel Protein Folds." Edited by Ruth Nussinov. *PLoS Computational Biology* 6 (4). Public Library of Science: e1000750. doi:10.1371/journal.pcbi.1000750.
- Fry, Bryan. 2005. "From Genome To 'venome': Molecular Origin and Evolution of the Snake Venom Proteome Inferred from Phylogenetic Analysis of Toxin Sequences

and Related Body Proteins." *Genome Research* 15 (3). Cold Spring Harbor Laboratory Press: 403–20. doi:10.1101/gr.3228405.

- Gorson, Juliette, Girish Ramrattan, Aida Verdes, Elizabeth M. Wright, Yuri Kantor, Ramakrishnan Rajaram Srinivasan, Raj Musunuri, et al. 2015. "Molecular Diversity and Gene Evolution of the Venom Arsenal of Terebridae Predatory Marine Snails." *Genome Biology and Evolution* 7 (6). Oxford University Press: 1761–78. doi:10.1093/gbe/evv104.
- Gravato-Nobre, Maria J., Hannah R. Nicholas, Reindert Nijland, Delia O'Rourke, Deborah E. Whittington, Karen J. Yook, and Jonathan Hodgkin. 2005. "Multiple Genes Affect Sensitivity of Caenorhabditis Elegans to the Bacterial Pathogen Microbacterium Nematophilum." *Genetics* 171 (3). http://www.genetics.org/content/171/3/1033#sec-1.
- Gross, Erhard. 1967. "The Cyanogen Bromide Reaction." *Methods in Enzymology* 11 (C): 238–55. doi:10.1016/S0076-6879(67)11029-X.
- Höcker, Birte. 2014. "Design of Proteins from Smaller Fragments-Learning from Evolution." *Current Opinion in Structural Biology* 27 (1): 56–62. doi:10.1016/j.sbi.2014.04.007.
- Imperial, Julita S., Maren Watkins, Ping Chen, David R. Hillyard, Lourdes J. Cruz, and Baldomero M. Olivera. 2003. "The Augertoxins: Biochemical Characterization of Venom Components from the Toxoglossate Gastropod Terebra Subulata." *Toxicon* 42 (4): 391–98. doi:10.1016/S0041-0101(03)00169-7.
- Jiang, Shan, and Steven L Nail. 1998. "Effect of Process Conditions on Recovery of Protein Activity after Freezing and Freeze-Drying." *European Journal of Pharmaceutics and Biopharmaceutics* 45 (3): 249–57. doi:10.1016/S0939-6411(98)00007-1.
- Jones, David T., and Liam J. McGuffin. 2003. "Assembling Novel Protein Folds From Super-Secondary Structural Fragments." In *Proteins: Structure, Function and Genetics*, 53:480–85. Wiley Subscription Services, Inc., A Wiley Company. doi:10.1002/prot.10542.
- Kendel, Yvonne, Christian Melaun, Alexander Kurz, Annette Nicke, Steve Peigneur, Jan Tytgat, Cora Wunder, Dietrich Mebs, and Silke Kauferstein. 2013. "Venomous Secretions from Marine Snails of the Terebridae Family Target Acetylcholine Receptors." *Toxins* 5 (5): 1043–50. doi:10.3390/toxins5051043.
- Khatib, Firas, Seth Cooper, Michael D Tyka, Kefan Xu, Ilya Makedon, Zoran Popovic, David Baker, and Foldit Players. 2011. "Algorithm Discovery by Protein Folding Game Players." *Proceedings of the National Academy of Sciences* 108 (47). National Academy of Sciences: 18949–53. doi:10.1073/pnas.1115898108.
- Klint, Julie K., Sebastian Senff, Natalie J. Saez, Radha Seshadri, Ho Yee Lau, Niraj S. Bende, Eivind A B Undheim, Lachlan D. Rash, Mehdi Mobli, and Glenn F. King. 2013. "Production of Recombinant Disulfide-Rich Venom Peptides for Structural and Functional Analysis via Expression in the Periplasm of E. Coli." *PLoS ONE* 8 (5). doi:10.1371/journal.pone.0063865.
- Lewis, R J, K J Nielsen, D J Craik, M L Loughnan, D A Adams, I A Sharpe, T Luchian, et al. 2000. "Novel Omega-Conotoxins from Conus Catus Discriminate among Neuronal Calcium Channel Subtypes." *The Journal of Biological Chemistry* 275 (45). American Society for Biochemistry and Molecular Biology: 35335–44.

doi:10.1074/jbc.M002252200.

- Lewis, Richard J., and Maria L. Garcia. 2003. "Therapeutic Potential of Venom Peptides." *Nature Reviews Drug Discovery* 2 (10): 790–802. doi:10.1038/nrd1197.
- Lupas, Andrei N., Chris P. Ponting, and Robert B. Russell. 2001. "On the Evolution of Protein Folds: Are Similar Motifs in Different Protein Folds the Result of Convergence, Insertion, or Relics of an Ancient Peptide World?" *Journal of Structural Biology* 134 (2–3). Academic Press: 191–203. doi:10.1006/JSBI.2001.4393.
- Moult, John, Krzysztof Fidelis, Andriy Kryshtafovych, Torsten Schwede, and Anna Tramontano. 2017. "Critical Assessment of Methods of Protein Structure Prediction (CASP)-Round XII." *Proteins: Structure, Function and Bioinformatics*. doi:10.1002/prot.25415.
- Murzin, Alexey G., Steven E. Brenner, Tim Hubbard, and Cyrus Chothia. 1995. "SCOP: A Structural Classification of Proteins Database for the Investigation of Sequences and Structures." *Journal of Molecular Biology* 247 (4). Academic Press: 536–40. doi:10.1016/S0022-2836(05)80134-2.
- Nawa, Mikiro, and Masaaki Matsuoka. 2012. "The Method of the Body Bending Assay Using Caenorhabditis Elegans." *BIO-PROTOCOL* 2 (17). doi:10.21769/BioProtoc.253.
- O 'reilly, Linda P, Cliff J. Luke, David H. Perlmutter, Gary A. Silverman, Stephen C.
   Pak, Linda P. O'Reilly, Cliff J. Luke, David H. Perlmutter, Gary A. Silverman, and
   Stephen C. Pak. 2014. "C. Elegans in High-Throughput Drug Discovery." Advanced
   Drug Delivery Reviews 69–70 (April): 247–53. doi:10.1016/j.addr.2013.12.001.
- Olivera, Baldomero M., Shrinivasan Raghuraman, Eric W. Schmidt, and Helena Safavi-Hemami. 2017. "Linking Neuroethology to the Chemical Biology of Natural Products: Interactions between Cone Snails and Their Fish Prey, a Case Study." *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology.* doi:10.1007/s00359-017-1183-7.
- Olivera, Baldomero M., Patrice Showers Corneli, Maren Watkins, and Alexander Fedosov. 2014. "Biodiversity of Cone Snails and Other Venomous Marine Gastropods: Evolutionary Success Through Neuropharmacology." *Annual Review of Animal Biosciences* 2 (1): 487–513. doi:10.1146/annurev-animal-022513-114124.
- Olivera, Baldomero M., Craig Walker, G. Edward Cartier, David Hooper, Ameurfina D. Santos, Robert Schoenfeld, Reshma Shetty, Maren Watkins, Pradip Bandyopadhyay, and David R. Hillyard. 1999. "Speciation of Cone Snails and Interspecific Hyperdivergence of Their Venom Peptides. Potential Evolutionary Significance of Introns." In Annals of the New York Academy of Sciences, 870:223–37. doi:10.1111/j.1749-6632.1999.tb08883.x.
- Olivera, Baldomero M., Maren Watkins, Pradip Bandyopadhyay, Julita S. Imperial, Edgar P.Heimer de la Cotera, Manuel B. Aguilar, Estuardo López Vera, Gisela P. Concepcion, and Arturo Lluisma. 2012. "Adaptive Radiation of Venomous Marine Snail Lineages and the Accelerated Evolution of Venom Peptide Genes." Annals of the New York Academy of Sciences 1267 (1). Blackwell Publishing Inc: 61–70. doi:10.1111/j.1749-6632.2012.06603.x.
- Ovchinnikov, Sergey, Hahnbeom Park, David E. Kim, Frank DiMaio, and David Baker. 2018. "Protein Structure Prediction Using Rosetta in CASP12." *Proteins: Structure*,

*Function, and Bioinformatics* 86 (March). Wiley-Blackwell: 113–21. doi:10.1002/prot.25390.

- Page, Anthony, and Iain L. Johnstone. 2007. "The Cuticle." *WormBook*, 1–15. doi:10.1895/wormbook.1.138.1.
- Partridge, Frederick A., Adam W. Tearle, Maria J. Gravato-Nobre, William R. Schafer, and Jonathan Hodgkin. 2008. "The C. Elegans Glycosyltransferase BUS-8 Has Two Distinct and Essential Roles in Epidermal Morphogenesis." *Developmental Biology* 317 (2): 549–59. doi:10.1016/j.ydbio.2008.02.060.
- Puillandre, Nicolas, and Mandë Holford. 2010. "The Terebridae and Teretoxins: Combining Phylogeny and Anatomy for Concerted Discovery of Bioactive Compounds." *BMC Chemical Biology* 10 (1): 7. doi:10.1186/1472-6769-10-7.
- Robinson, Samuel D., Sandeep Chhabra, Alessia Belgi, Balasubramanyam Chittoor, Helena Safavi-Hemami, Andrea J. Robinson, Anthony T. Papenfuss, Anthony W. Purcell, and Raymond S. Norton. 2016. "A Naturally Occurring Peptide with an Elementary Single Disulfide-Directed β-Hairpin Fold." *Structure* 24 (2): 293–99. doi:10.1016/j.str.2015.11.015.
- Robinson, Samuel D., and Helena Safavi-Hemami. 2016. "Insulin as a Weapon." *Toxicon*. doi:10.1016/j.toxicon.2016.10.010.
- Roy, Ipsita, and Munishwar Nath Gupta. 2004. "Freeze-Drying of Proteins: Some Emerging Concerns." *Biotechnology and Applied Biochemistry* 39 (Pt 2): 165–77. doi:10.1042/BA20030133.
- Roy, Scott W. 2016. "Is Mutation Random or Targeted?: No Evidence for Hypermutability in Snail Toxin Genes." *Molecular Biology and Evolution* 33 (10): 2642–47. doi:10.1093/molbev/msw140.
- Safavi-Hemami, Helena, Grzegorz Bulaj, Baldomero M. Olivera, Nicholas A. Williamson, and Anthony W. Purcell. 2010. "Identification of Conus Peptidylprolyl Cis-Trans Isomerases (PPIases) and Assessment of Their Role in the Oxidative Folding of Conotoxins." *Journal of Biological Chemistry* 285 (17): 12735–46. doi:10.1074/jbc.M109.078691.
- Safavi-Hemami, Helena, Joanna Gajewiak, Santhosh Karanth, Samuel D Robinson, Beatrix Ueberheide, Adam D Douglass, Amnon Schlegel, et al. 2015. "Specialized Insulin Is Used for Chemical Warfare by Fish-Hunting Cone Snails." *Proceedings of the National Academy of Sciences* 112 (6): 1743–48. doi:10.1073/pnas.1423857112.
- Safavi-Hemami, Helena, Qing Li, Ronneshia L. Jackson, Albert S. Song, Wouter Boomsma, Pradip K. Bandyopadhyay, Christian W. Gruber, et al. 2016. "Rapid Expansion of the Protein Disulfide Isomerase Gene Family Facilitates the Folding of Venom Peptides." *Proceedings of the National Academy of Sciences* 113 (12): 3227–32. doi:10.1073/pnas.1525790113.
- Safavi-Hemami, Helena, Aiping Lu, Qing Li, Alexander E. Fedosov, Jason Biggs, Patrice Showers Corneli, Jon Seger, Mark Yandell, and Baldomero M. Olivera. 2016.
  "Venom Insulins of Cone Snails Diversify Rapidly and Track Prey Taxa." *Molecular Biology and Evolution* 33 (11). doi:10.1093/molbev/msw174.
- Sander, Chris, and Reinhard Schneider. 1991. "Database of Homology-Derived Protein Structures and the Structural Meaning of Sequence Alignment." *Proteins: Structure, Function, and Genetics* 9 (1). Wiley-Blackwell: 56–68. doi:10.1002/prot.340090107.

- Schaarschmidt, Joerg, Bohdan Monastyrskyy, Andriy Kryshtafovych, and Alexandre M J J Bonvin. 2018. "Assessment of Contact Predictions in CASP12: Co-Evolution and Deep Learning Coming of Age." *Proteins* 86 Suppl 1 (Suppl Suppl 1). Wiley-Blackwell: 51–66. doi:10.1002/prot.25407.
- Service, Robert F. 2008. "Problem Solved\* (\*sort of)." *Science*. American Association for the Advancement of Science. doi:10.1126/science.321.5890.784.
- Skolnick, Jeffrey, Yang Zhang, Adrian K. Arakaki, Andrzej Kolinski, Michal Boniecki, András Szilágyi, and Daisuke Kihara. 2003. "TOUCHSTONE: A Unified Approach to Protein Structure Prediction." In *Proteins: Structure, Function and Genetics*, 53:469–79. doi:10.1002/prot.10551.
- Söding, Johannes, and Andrei N. Lupas. 2003. "More than the Sum of Their Parts: On the Evolution of Proteins from Peptides." *BioEssays* 25 (9). Wiley-Blackwell: 837–46. doi:10.1002/bies.10321.
- Teichert, Russell W, Shrinivasan Raghuraman, Tosifa Memon, Jeffrey L Cox, Tucker Foulkes, Jean E Rivier, and Baldomero M Olivera. 2012. "Characterization of Two Neuronal Subclasses through Constellation Pharmacology." *Proceedings of the National Academy of Sciences of the United States of America* 109 (31). National Academy of Sciences: 12758–63. doi:10.1073/pnas.1209759109.
- Teichert, Russell W, Eric W Schmidt, and Baldomero M Olivera. 2015. "Constellation Pharmacology: A New Paradigm for Drug Discovery." Annual Review of Pharmacology and Toxicology 55 (1). Annual Reviews: 573–89. doi:10.1146/annurev-pharmtox-010814-124551.
- Terlau, Heinrich, and Baldomero M Olivera. 2004. "Conus Venoms: A Rich Source of Novel Ion Channel-Targeted Peptides." *Physiological Reviews* 84 (1). American Physiological Society: 41–68. doi:10.1152/physrev.00020.2003.
- Ugwu, Sydney O, and Shireesh P Apte. 2004. "The Effect of Buffers on Protein Conformational Stability Buffer Effects on Freeze Drying." *86 Pharmaceutical Technology MARCH*. www.pharmtech.com.
- Verdes, Aida, Prachi Anand, Juliette Gorson, Stephen Jannetti, Patrick Kelly, Abba Leffler, Danny Simpson, Girish Ramrattan, and Mandë Holford. 2016. "From Mollusks to Medicine: A Venomics Approach for the Discovery and Characterization of Therapeutics from Terebridae Peptide Toxins." *Toxins* 8 (4). Multidisciplinary Digital Publishing Institute (MDPI): 117. doi:10.3390/toxins8040117.
- Zhang, Yang, and Jeffrey Skolnick. 2005. "The Protein Structure Prediction Problem Could Be Solved Using the Current PDB Library." *Proceedings of the National Academy of Sciences of the United States of America* 102 (4): 1029–34. doi:10.1073/pnas.0407152101.

Name of Candidate:	Kacey Davis
Birth date:	January 8, 1997
Birth place:	Tampa, Florida
Address:	570 B Douglas St. Salt Lake City, UT, 84102