# Discovery and Characterization of Novel Thanatin Orthologs Against *Escherichia coli* LptA and *Pseudomonas aeruginosa* LptH

by

Kelly Huynh

Department of Biochemistry
Duke University

Date:_______________________
Approved:

___________________________
Bruce Donald, Supervisor

___________________________
Pei Zhou

___________________________
Terrence Oas

___________________________
Richard Brennan

___________________________
Jane Richardson

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Biochemistry in the Graduate School
of Duke University

2023

# ABSTRACT

# Discovery and Characterization of Novel Thanatin Orthologs Against *Escherichia coli* LptA and *Pseudomonas aeruginosa* LptH

by

Kelly Huynh

Department of Biochemistry
Duke University

Date:_______________________
Approved:

___________________________
Bruce Donald, Supervisor

___________________________
Pei Zhou

___________________________
Terrence Oas

___________________________
Richard Brennan

___________________________
Jane Richardson

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Biochemistry in the Graduate School of
Duke University

2023

Copyright by
Kelly Huynh
2023

# Abstract

Multidrug resistance (MDR) in bacteria is ever growing and complicates treatment of infections, especially in patients who are critically ill and immunocompromised. Treatment often utilizes a regiment of small molecule drugs, however resistance against them develops after prolonged usage. An alternative class of molecules, antimicrobial peptides (AMPs), has remained of interest due to its vast potential of becoming pharmacological agents. AMPs, or host defense peptides, are naturally expressed in many organisms, including microbes, plants, and humans. AMPs are expressed to control the population of bacteria, fungi, and viruses as a defense mechanism. Mining host genomes for AMPs will prove to be a valuable source of novel alternative drug molecules. Characterization of AMPs will lead to be a better understanding of their mechanism of action and allow for applications to novel targets. Here in this dissertation, we apply these methods to thanatin, an AMP identified from the spined soldier bug (*Podisus maculiventris*) that was reported to regulate the gut microbiome population by targeting Gram-positive bacteria, Gram-negative bacteria, and fungi.

First, we mined genomic databases to discover novel thanatin orthologs. We generated these orthologs and characterized their binding against *Escherichia coli* LptA, a known target of thanatin, via bio-layer interferometry (BLI) and their antimicrobial

activity against several *E. coli* strains via minimum inhibition concentration (MIC) assays. We found a subset of thanatin sequences that target *E. coli* better than *P. maculiventris* thanatin, as shown with increased binding affinity, cell permeability, and overall potency. We crystallized and determined the structures of *Chinavia ubica* thanatin and *Murgantia histrionica* thanatin, the two most improved thanatin orthologs, in complex with *E. coli* LptA to better understand the interaction. We performed mutagenesis studies to show that thanatin residues A10 and M21 interacts with the hydrophobic core of LptA and improves binding and synergistically improves antimicrobial activity against *E. coli*. We redesigned *M. histrionica* thanatin to truncate the sequence and remove the need for a disulfide bond. Our stapled peptide retained binding affinity to LptA, however potency was hampered. Despite seeing no improvement in antimicrobial activity, we present a novel scaffold for the next generation of thanatin-based AMPs.

Next, we characterized thanatin against *Pseudomonas aeruginosa*, a known but weaker target of thanatin. We confirmed binding of *P. maculiventris* thanatin to LptH, the *P. aeruginosa* homolog of *E. coli* LptA, via BLI and isothermal titration calorimetry (ITC) and showed inhibition of *P. aeruginosa* strain RP73 via MIC assays. We used homology modeling and an *E. coli* model system to identify the resistance factor of thanatin to be LptH Y51 at the predicted binding interface. We attempted to overcome the hinderance of LptH Y51 by modeling thanatin to accommodate it. Our designs cooperated in the *E. coli* model system, however they did not translate to improve binding with LptH. Interestingly, we discovered that thanatin Y10 is essential to binding LptH. We applied the small library of thanatin orthologs to LptH and *P. aeruginosa* and did not discover any sequences with improved binding or antimicrobial activity. Our small library screening highlighted the necessity of thanatin Y10 and the resistance factor LptH Y51 again. We investigated the role of improved potency of thanatin with *P. aeruginosa* through C-amidation. Our *E. coli* model system shows a key rescued interaction between the C-amidated terminus of thanatin and LptA R76Q that mimics LptH. We crystallized and determined the structure of C-amidated truncated thanatin and LptA R76Q to gain insight on the interaction. However, we did not observe the hypothesized rescued interaction. When translating our findings to LptH, we did not observe improved binding due to C-amidation via BLI, but we did via ITC. Conflicting data about how thanatin interacts with LptH could be clarified with a high-resolution protein:peptide complex structure, however attempts to experimentally obtain one has been difficult. Overall, we provide some insight on the mechanism of how thanatin targets LptH in *P. aeruginosa*, but further studies will be needed to fully elucidate its mechanism of action.

Collectively, this dissertation provides an example of how natural sources can be mined to uncover novel AMPs to target bacteria with MDR on the rise. We present various insights gained on the mechanism of action of thanatin by characterizing thanatin and its novel orthologs against *E. coli* LptA and *P. aeruginosa* LptH. The characterization of thanatin will allow for improved AMPs to be designed in the next generation of thanatin peptides to target pathogens.

# Contents

Abstract ..... iv

List of Tables ..... xiii

List of Figures ..... xiv

List of Abbreviations ..... xvi

1. Introduction ..... 1

    1.1 Antimicrobial peptides ..... 1

        1.1.1 Discovery and sources ..... 1

        1.1.2 Classes and mechanisms of AMPs based on structure ..... 4

        1.1.3 Structural modifications of AMPs for clinical usage ..... 8

    1.2 Lipopolysaccharide transport pathway ..... 9

        1.2.1 Introduction ..... 9

        1.2.2 Inner membrane ..... 10

        1.2.3 Periplasm ..... 12

        1.2.4 Outer membrane ..... 13

        1.2.5 Inhibition of lipopolysaccharide transport pathway ..... 15

    1.3 Thanatin ..... 17

        1.3.1 Discovery and activity profile ..... 17

        1.3.2 Structure and functional mapping of thanatin ..... 18

        1.3.3 Mechanisms of action against Gram-negative bacteria ..... 23

        1.3.4 Thanatin hybrid peptides ..... 25

1.4 Summary ........................................................................................................27

1.5 Outline of dissertation ........................................................................................28

2. Characterizing thanatin against *Escherichia coli* LptA ........................................................29

2.1 Background ........................................................................................................29

2.2 Novel thanatin orthologs reveal consensus sequence ..........................................31

2.3 Thanatin orthologs bind tighter to *E. coli* LptA and display stronger antibiotic activity than *P. maculiventris* thanatin ........................................................................33

2.4 Crystal structure of P. maculiventris thanatin bound to *E. coli* LptAm reveals a domain-swapped dimer ........................................................................................35

2.5 Shared features of LptAm-thantin ortholog complexes ........................................38

2.6 Molecular basis of the enhanced LptAm binding and antibiotic activity by thanatin orthologs from *C. ubica* and *M. histrionica*..................................................44

2.7 Redesigning *M. histrionica* thanatin ........................................................................46

2.8 Discussion..........................................................................................................49

2.9 Materials and methods ........................................................................................52

2.9.1 Sequence analysis............................................................................................52

2.9.2 Cloning, expression, and purification of *E. coli* LptA ..........................................53

2.9.3 Cloning, expression, and purification of thanatin peptides ..................................54

2.9.4 Bio-layer interferometry binding assay ..............................................................56

2.9.5 Minimal inhibition concentration assay ..............................................................57

2.9.6 Crystallization and structure determination of *E. coli* LptA with *P. maculiventris* thanatin..........................................................................................57

2.9.7 Crystallization and structure determination of *E. coli* LptA with *C. ubica* thanatin..............................................................................................................58

2.9.8 Crystallization and structure determination of *E. coli* LptA with *M. histrionica* thanatin ........ 59

3. Characterizing thanatin against *Pseudomonas aeruginosa* LptH ........ 60

3.1 Background ........ 60

3.2 Thanatin targets LptH in *P. aeruginosa* ........ 61

3.3 Identifying the resistance factor of thanatin in *P. aeruginosa* LptH ........ 64

3.4 Improving thanatin to overcome LptH Y51 resistance factor ........ 68

3.5 Characterizing thanatin orthologs against *P. aeruginosa* LptH ........ 71

3.6 Role of C-amidated thanatin with *P. aeruginosa* ........ 73

3.7 Discussion ........ 77

3.8 Materials and methods ........ 81

3.8.1 Generating AlphaFold2-multimer models ........ 81

3.8.2 Molecular cloning, expression, and purification of *P. aeruginosa* LptH and mutants ........ 81

3.8.3 Molecular cloning, expression, and purification of thanatin peptides ........ 82

3.8.4 Molecular cloning, expression, and purification of *E. coli* LptA mutants ........ 83

3.8.5 Bio-layer interferometry binding assay for *P. aeruginosa* LptH and mutants ........ 83

3.8.6 Bio-layer interferometry binding assay for *E. coli* LptA mutants ........ 84

3.8.7 Isothermal calorimetry titration binding assay for *P. aeruginosa* LptH and mutants ........ 84

3.8.8 Minimum inhibition concentration assay for *P. aeruginosa* RP73 ........ 85

3.8.9 Crystallization and structure determination of *E. coli* LptA R76Q with *P. maculiventris* thanatin Δ1-3 C-amidation ........ 86

4. Conclusions........................................87

  4.1 Discovery of novel thanatin more potent towards *E. coli* ........................................87

  4.2 Insight towards thanatin with *P. aeruginosa* LptH ........................................88

  4.3 Future work........................................90

  4.3 Conclusion........................................93

Appendix A. Effect of lipid composition on the function of hepatitis C virus p7 ........................................94

  A.1 Introduction ........................................94

    A.1.1 Hepatitis C virus background........................................94

    A.1.2 Reported data supporting C6 and C7 oligomeric state ........................................95

  A.2 Alternative C6 and C7 models of p7 ........................................99

  A.3 p7 in DMPC liposomes is not functional........................................105

  A.4 p7 in PA:PC:PE liposomes is functional ........................................107

  A.5 Discussion ........................................109

  A.6 Conclusions........................................114

  A.7 Materials and methods........................................115

    A.7.1 Determination of alternative C6 and C7 models of p7........................................115

    A.7.2 Molecular cloning, expression, and purification of p7........................................116

    A.7.3 Liposome formation ........................................118

    A.7.4 TEM imaging of liposomes........................................119

    A.7.5 p7 incorporation into liposomes........................................120

    A.7.6 Dye-release liposome assay........................................121

References ........................................123

Biography ....................................................................................................................139

# List of Tables

Table 1: X-ray data collection and refinement statistics of *E. coli* LptA-Thanatin complexes......41

Table 2: X-ray data collection and refinement statistics of *E. coli* LptA R76Q - *P. maculiventris* Thanatin ΔG1-K3 C-amidated complex......74

Table 3: Summary of NMR restraint statistics......102

# List of Figures

Figure 1: Diversity of hosts and targets of antimicrobial peptides. ......................................3

Figure 2: Diversity of antimicrobial peptides based on structure. ........................................5

Figure 3: Model of the lipopolysaccharide transport pathway............................................14

Figure 4: Structures of thanatin. ..........................................................................................21

Figure 5: Sequence alignment of thanatin orthologs reveals a consensus sequence. ........32

Figure 6: Identification of thanatin orthologs that improve binding and inhibition to *E. coli*..................................................................................................................................34

Figure 7: Representative BLI binding curves for each thanatin ortholog binding to *E. coli* LptA. ..................................................................................................................................35

Figure 8: Structural comparison of LptAm in complex with *P. maculiventris* thanatin by NMR and crystallography...................................................................................................36

Figure 9: Truncated LptA is monomeric in solution as shown by SECMALS. ..................37

Figure 10: Crystal structures of LptAm in complex with thanatin orthologs....................42

Figure 11: Characteristics of *P. maculiventris* thanatin binding to *E. coli* LptAm. ..............43

Figure 12: Differences between *P. maculiventris*, *C. ubica*, and *M. histrionica* thanatin. .....45

Figure 13: Redesigning *M. histrionica* thanatin into a novel scaffold. .................................48

Figure 14: *P. maculiventris* thanatin targets *P. aeruginosa* LptH............................................63

Figure 15: Identifying LptH Y51 as resistance factor through LptA model system. .........65

Figure 16: LptH Y51F induces higher-ordered self-oligomerization. .................................67

Figure 17: Thanatin Y10 is essential for binding LptH.........................................................69

Figure 18: Thanatin orthologs do not target *P. aeruginosa* LptH better than *P. maculiventris* thanatin. .........................................................................................................72

Figure 19: C-amidation of thanatin may not play a role in improved interactions with LptAm R76Q. ....75

Figure 20: Poor fit of $^1$H-$^{15}$N RDCs from PDB ID: 2M6X. ....101

Figure 21: Alternative C6 and C7 structures of p7. ....103

Figure 22: p7 in DMPC liposomes is not functional. ....106

Figure 23: p7 in PA:PC:PE liposomes is functional. ....109

## List of Abbreviations

AA1 – *Aelia acuminata* (isoform 1)

AA2 – *Aelia acuminata* (isoform 2)

ABC – ATP-binding cassette

AMP – antimicrobial peptide

APD – antimicrobial peptide database

ATP – adenosine 5′-triphosphate

ATCC – American Type Culture Collection

BLI – bio-layer interferometry

BMRB – biological magnetic resonance bank

BN-PAGE – blue native – polyacrylamide gel electrophoresis

CAMHB – cation-adjusted mueller-hinton broth

CF – carboxyfluorescein

CFU – colony forming unit

CLSI – clinical & laboratory standards institute

CNBr – cyanogen bromide

CU – *Chinavia ubica*

Da – dalton

DAA – direct-acting antiviral

DBAASP – database of antimicrobial activity and structure of peptides

DHPC – 1,2-dihexanoyl-sn-glycero-3-phosphocholine

DMPC – 1,2-dimyristoyl-sn-glycero-3-phosphocholine

DNA – deoxyribose nucleic acid

DTT – dithiothreitol

EDTA – ethylenediaminetetraacetic acid

FDA – food and drug administration

FOT – fold operator theory

HBS – HEPES buffered saline

HCl – hydrochloric acid

HCV – hepatitis C virus

HEPES – N-2-hydroxyethylpiperazine-N′-2-ethanesulfonic acid

HH – *Halyomorpha halys*

HIV – human immunodeficiency virus

HPLC – high-performance liquid chromatography

ID – identification

IPTG – isopropyl-β-D-thiogalactopyranoside

kDa – kilodalton

LB – Luria-Bertani

LC-MS – liquid chromatography – mass spectrometry

LPS – lipopolysaccharide

Lpt – lipopolysaccharide transport

MALDI-TOF – matrix-assisted laser desorption/ionation – time of flight

MDR – multidrug resistance

MH – *Murgantia histrionica*

MIC – minimal inhibition concentration

MRSA – methicillin-resistant *Staphylococcus aureus*

MSP – membrane scaffold protein

MSR – maximally satisfying region

MTT – 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide

MW – molecular weight

MWCO – molecular weight cut-off

NDM-1 – New Delhi metallo-β-lactamase

Ni-NTA – nickel nitrilotriacetic acid

NMR – nuclear magnetic resonance

NOE – nuclear Overhauser effect

NZ – *Nezara viridula*

OD – optical density

PA – phosphatidic acid

PBS – phosphate buffered saline

PC – phosphatidylcholine

PDB – protein databank

PE – phosphatidylethanolamine

PEG – polyethylene glycol

PM – *Podisus maculiventris*

PMBN – polymyxin B nonapeptide

PRE – paramagnetic relaxation enhancement

RDC – residual dipolar coupling

RMSD – room mean squared deviation

RNA – ribonucleic acid

RP – *Riptortus pedestris*

SDS-PAGE – sodium dodecyl-sulfate polyacrylamide gel electrophoresis

SEC-MALS – size-exclusion chromatography – multi-angle light scattering

SEM – standard error of the mean

SMALP – styrene maleic acid lipid particle

TEM – transmission electron microscopy

TEV – tobacco etch virus

TFA – trifluoroacetic acid

TFE – 2,2,2-trifluoroethanol

VRE – vancomycin-resistant enterococci

# Acknowledgements

I would like to thank all past and current mentors who have helped me become the person and scientist that I am today. I would like to thank Bruce Donald, Pei Zhou, Terrence Oas, Richard Brennan, and Jane Richardson for serving on my committee and offering their advice and expertise throughout my time at Duke University. I would like to thank Maria Schumacher for all the life and science advice I have received throughout the last couple of years. Thank you to all who have welcomed me and given me a desk in their laboratory as I bounced between FFSC, LSRC, NanDuke, and Sands.

I would like to thank all current and past members of the Donald Laboratory and Zhou laboratory for their advice, training, feedback, and friendship. Thank you to everyone for making the labs enjoyable places to learn from and be in. I would especially like to thank Amanuel Kibrom for being a close collaborator and great friend. I would also especially like to thank Catherine Ehrhart for being with me since Day 1 at Duke and experiencing all the highs and lows of graduate school with me.

I would like to thank all my friends for their unwavering support throughout this journey. I truly treasure every memory from indulging in every farmers market during muscadine season, to going on endless pistachio latte runs, to throwing pottery at the Art Annex, to living together or next door, and so much more. Thank you for all the phone calls, sanity checks, care packages, and food deliveries throughout the years. I

am grateful for you all for letting me be in your orbit and I look forward to our paths crossing again.

Most importantly, I would like to thank my family, especially my parents, sister, and brother for their endless support. I am grateful for all the sacrifices they made to allow me to become the first doctorate in the family.

# 1. Introduction

## *1.1 Antimicrobial peptides*

### 1.1.1 Discovery and sources

After the boom of antibiotic discovery from 1940 – 1970, interest and discovery of antimicrobial peptides (AMP) began to emerge. Some of the earliest reported AMPs included melittin, a 26-residues long peptide isolated from bee venom in the 1970s; bombesin, a 14-residues long peptide isolated from frog skin in 1971; cecropins, 37-residues long peptides isolated from the silk moth hemolymph in 1981; and magainins, 23-residues long peptides isolated from frog skin in 1987 [1]–[4]. These short biological peptides were discovered to have benefits such as being water soluble, nonhemolytic, and having effective antimicrobial activity [1]–[4]. The link between AMPs and host immunity was first reported with the discovery of inducible cecropins caused by challenging the immune system of the silk moth [5]. It has later been shown that downregulating genes expressing AMPs can increase the host susceptibility to infection, especially if the host lacks an adaptive immune system [6]. This shows that nature has evolved a defense mechanism for innate immunity through self-regulation of the microbiome population.

AMPs have been discovered in a wide range of hosts from wildly different environments, most likely reflective of the high diversity of the microbiota. Insects and amphibians are common host organisms of AMPs, as highlighted above, however AMPs can also be found in humans, plants, fungi, and bacteria and target a wide variety of groups **(Figure 1A, B)**. LL-37/hCAP-18 is a 37-residues long peptide found in humans in 1995 showing antimicrobial and anti-inflammatory properties [7]. Plants contain defensins, which are a family of small cysteine-rich peptides that serve as the innate immune system primarily against fungi [8]. Effective AMPs for specific targets can also arise from their own family. Caspofungin is an anti-fungal AMP isolated from another fungal host, *Glarea lozoyensis*, shown to have successful derivatives [9]. Gramicidin D is a 15-residues long peptide targeting Gram-positive bacteria isolated from another Gram-positive bacterial host, *Brevibacillus brevis* [10]. AMPs are often expressed as a precursor to protect the host from self-harm. The AMP only becomes active after cleavage from the precursor and proper delivery to the site of release [11]. These examples all highlight instances of how various organisms across all forms of life balance their microbiome population through a common tactic.

Over the years, efforts to discover and characterize more AMPs to exploit their benefits have only increased. Valuable databases such as the Antimicrobial Peptide Database (APD) and the Database of Antimicrobial Activity and Structure of Peptides (DBAASP) have built libraries of over 3,500 and 20,000 peptides, respectively [12], [13]. Since AMPs have been found across all forms of life, mining the deposited genomes for novel AMPs would be beneficial. Efforts to mine the human and amphibian genome and insect transcriptome for novel AMPs have produced new computational tools and

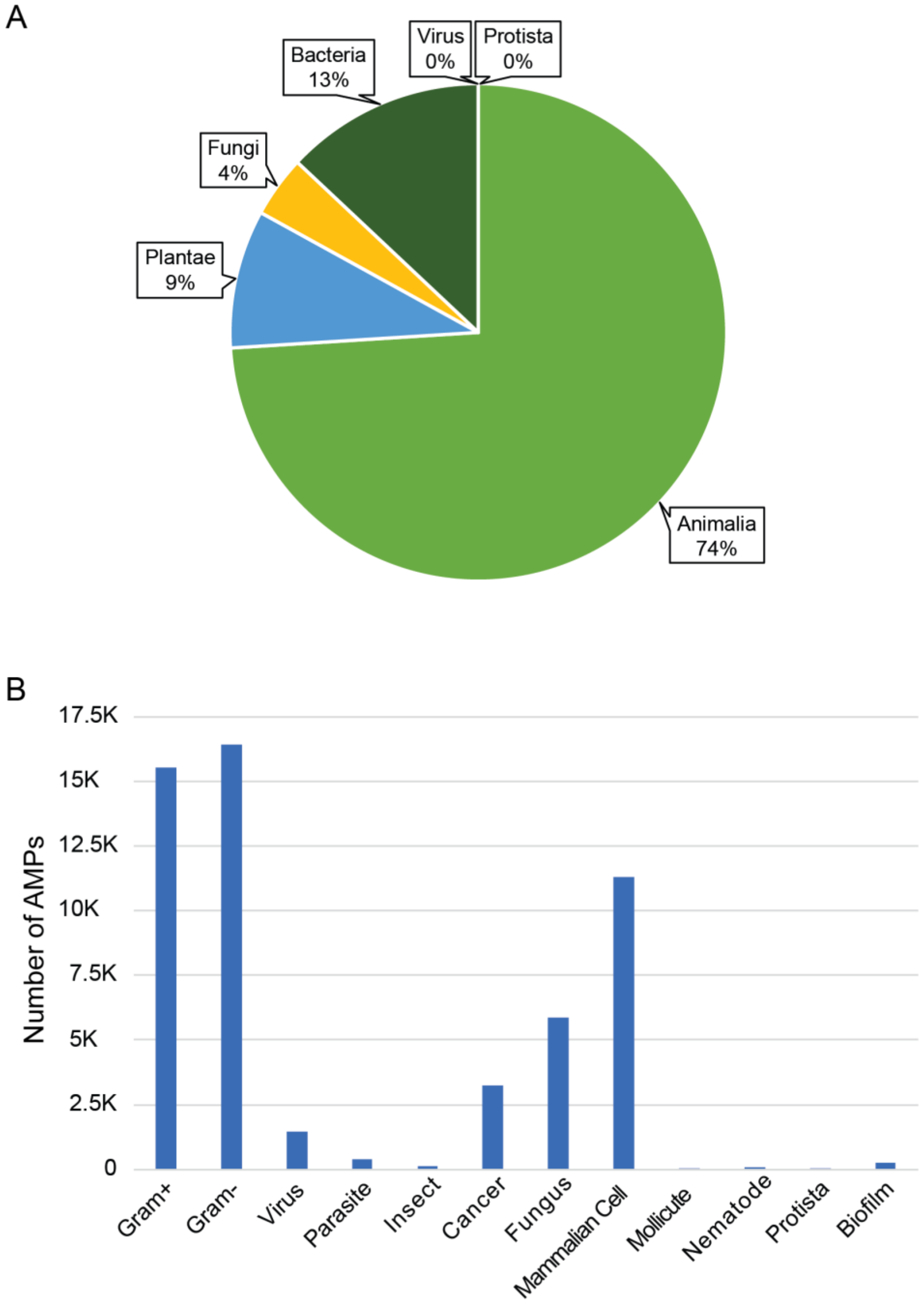

**Figure 1: Diversity of hosts and targets of antimicrobial peptides.**

Data was adapted from the Database of Antimicrobial Activity and Structure of Peptides (DBAASP) [13]. **A**, DBAASP library represented by the host kingdom. **B**, DBAASP library represented by what the antimicrobial peptide targets.

effective AMPs to further develop [14]–[16]. Computational tools for mining cryptic peptides, such as an AMP encoded within a precursor, or predicting an AMP based on sequence have also been recently developed to search for new antibiotics [14], [17]. As multidrug resistance in pathogens is on the rise, the search for new antimicrobial agents to combat infection will be in high demand.

### 1.1.2 Classes and mechanisms of AMPs based on structure

In general, AMPs have been defined to be around 100 residues or less, cationic, amphipathic, and act upon membranes. The first step in the mechanism of action usually involves binding to the negatively charged components of the cellular membrane. Gram-negative bacteria contain negatively charged lipopolysaccharide (LPS) molecules. Gram-positive bacteria contain negatively charged teichoic acid and lipoteichoic acid. Fungi, such as *C. albicans*, generally contain a negatively charged surface due to phosphomannosylation of surface proteins [18]. The cationic nature of AMPs creates a natural attraction between the AMPs and cell membranes. A study on magainin showed increased membrane affinity correlated with magainin analogs of increasing net positive charges [19]. The next step in the mechanism of action varies based on AMP structure. Most AMPs can be grouped into the following three classes based on their secondary structure: 1) amphipathic α-helical, 2) β-sheet stapled by disulfide bonds, and 3) extended or random coil. Despite being highly active against an overlapping population of microbes, these AMP classes have various mechanisms of action.

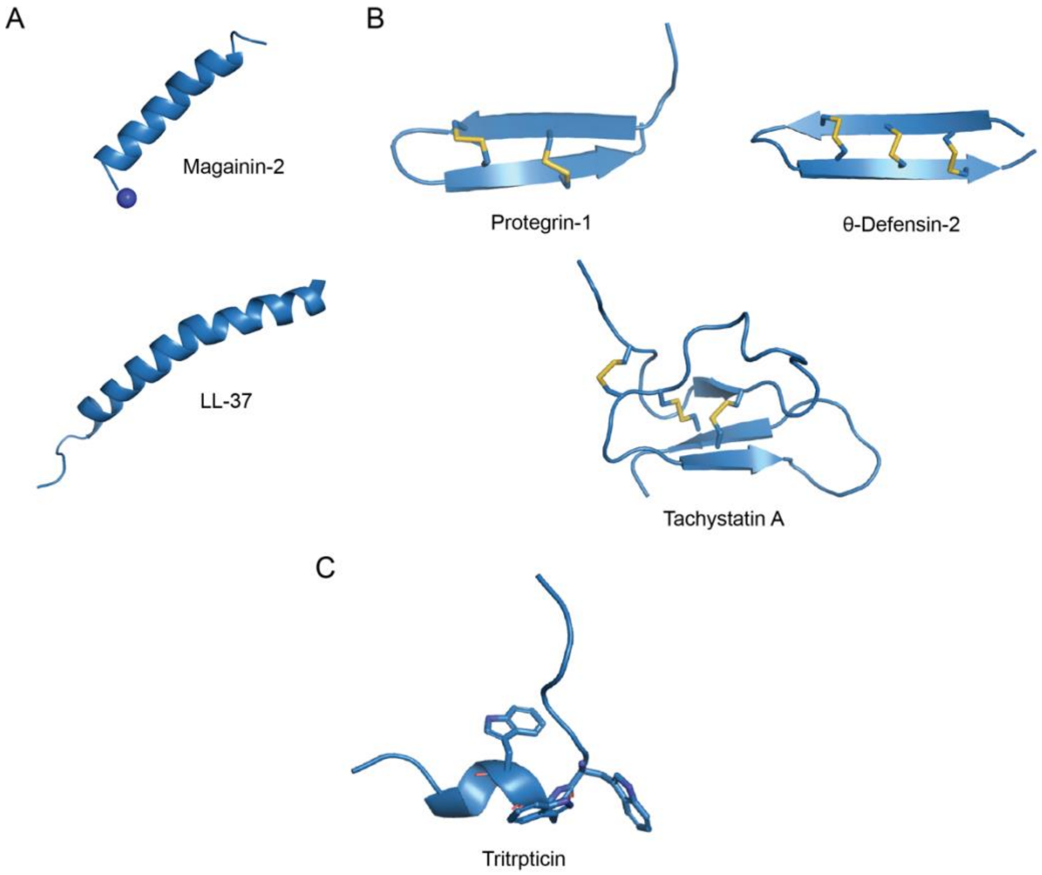

**Figure 2: Diversity of antimicrobial peptides based on structure.**
**A**, Examples of α-helical AMPs include magainin-2 (PDB ID: 2MAG) and LL-37 (PDB ID: 2K6O). **B**, Examples of β-sheet AMPs include protegrin-1 (PDB ID: 1PG1), θ-defensin-2 (PDB ID: 5INZ), and tachystatin A (PDB ID: 1CIX). **C**, Example of extended AMPs such as tritrpticin (PDB ID: 1D6X).

Amphipathic α-helical AMPs are some of the most common AMPs reported in literature. Of the known structures deposited in the APD, α-helical AMPs make up about 60% of the database, while another roughly 15% of the data include a mix of α-helices and β-strands [12]. Examples of peptides in this class are melittin, cecropins, LL-37, and magainins **(Figure 2A)**. The mechanism of action for this class of AMPs is through membrane disruption. α-helical AMPs, such as melittin and magainin-2, insert themselves into the lipid bilayer through its amphipathic properties, induce pore formation, and cause leakage of ions causing depolarization, thus leading to cellular lysis [20], [21]. Some linear AMPs adopt an α-helical structure after insertion into the lipid membrane, stabilized by the hydrophobic environment [21]. One disadvantage of α-helical AMPs often observed is the high rate of hemolytic activity and toxicity to mammalian cells with increased helicity [22]. Developing α-helical AMPs into therapeutic agents will need to overcome this hurdle.

β-sheet AMPs consist of at least two β-strands or more and are stabilized by at least one disulfide bond. They are often amphipathic and have not been shown to undergo conformational change upon binding to the membrane. Of the known structures deposited in the APD, β-sheet AMPs make up about 10% of the database, in addition to the roughly 15% of mixed α-helices and β-strands as previously mentioned [12]. Examples of this class include protegrin-1, an AMP isolated from porcine leukocytes; tachystatin A, an AMP isolated from horseshoe crabs; and θ-defensins, a cyclic AMP isolated from monkey leukocytes **(Figure 2B)**. In addition to destabilizing membranes, these AMPs were also shown to affect cellular pathways. While protegrin-1 was shown to insert into membrane and form pores, it was also shown to regulate immune activity by binding to receptors [23], [24]. The exact mechanism of action of tachystatin A has not been elucidated, however it has high structural similarity with agatoxin from spider venom which was shown to antagonize calcium channels through binding [25]. θ-defensins were shown to destabilize the membrane and induce cell lysis, but also play a role in suppressing inflammatory cytokines in the host [26]. These examples show that AMPs have multiple targets on top of what is commonly seen, thus further building on their broad profile. Multiple mechanisms of actions and targets would slow down the emergence of multidrug resistance by creating a larger barrier for pathogens to overcome.

Extended AMPs consist of AMPs with no clear secondary structures, however they are often identified through enriched amino acids, such as glycine, proline, histidine, and tryptophan. Of the known structures deposited in the APD, extended AMPs make up about 15% of the database [12]. Examples of this include armadillidin, a glycine-rich AMP isolated from pill bugs; SpPR-AMP1, a proline-rich AMP isolated from mud crabs; and tritrpticin, a tryptophan-rich AMP isolated from porcine bone marrow **(Figure 2C)**. Armadillidin, was shown to inhibit bacteria through membrane destabilization, despite not having any secondary structure in solution or membrane mimicking conditions [27]. Interestingly, proline-rich AMPs, such as SpPR-AMP1, were shown to not induce bacterial lysis, however they are transported into the cytoplasm through the inner membrane transporter SbmA, where they inhibit protein synthesis [28], [29]. Tritrpticin was shown to inhibit bacteria through membrane destabilization, due to the favorable insertion of tryptophan into hydrophobic environments, but an analog was shown to act through a different mechanism that does not depolarize the membrane [30], [31]. Even without much structure, extended AMPs are shown to still be effective at targeting microbes, often through multiple mechanisms of action.

### 1.1.3 Structural modifications of AMPs for clinical usage

AMPs have shown a wide display of activity in nature, however translating them into therapeutic use in humans has proven to be difficult due to hemolytic effects, cytotoxicity, and low metabolic stability. AMPs have been able to overcome these difficulties by using chemical modifications to improve their drug-like properties, such as solubility, membrane permeability, and bioavailability. Pexiganan, a derivative of magainin, was the first AMP to be submitted for Food and Drug Administration (FDA) approval [32]. It was developed to be a topical cream to treat diabetic foot ulcer infections. Pexiganan was C-amidated to increase the metabolic stability by reducing proteasomal degradation and increase the net positive charge to increase membrane permeability. Unfortunately, pexiganan did not pass the phase III clinical trial stage due to not showing improvement over standard of care [32]. However, AMPs have improved over the years to enter the clinic with smarter and more creative modifications. Daptomycin, a cyclic lipopeptide, was FDA approved in 2003 for the treatment of Gram-positive hospital acquired infections, such as methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant enterococci (VRE) [33], [34]. Daptomycin makes use of cyclization, nonnatural amino acids, and D-amino acids to improve metabolic stability and reduce proteasomal degradation. Daptomycin also makes use of hybridization with a lipid tail to improve membrane permeability and delivery. Together, these modifications aided in overcoming barriers to lead to FDA approval and clinical use, under the name Cubicin [33], [34]. As more AMPs are discovered with potential clinical application, synthetic modifications to enhance drug-like characteristics will need to be further developed to add to the toolbox to combat pathogens.

## *1.2 Lipopolysaccharide transport pathway*

### 1.2.1 Introduction

AMPs have a wide range of targets **(Figure 1B)**, however this dissertation will be focusing on Gram-negative bacteria. Gram-negative bacteria contain a complex cell wall to shield the bacteria from foreign materials and combat environmental stresses. The cell wall structure is composed of an inner membrane, a periplasm, and an outer membrane [35]. The inner membrane serves as a barrier between the cytoplasm and periplasm and is imbedded with various proteins that serve as signal receptors, transporters, enzymes, or anchors. The periplasm, located between the inner and outer membrane, contains a soluble matrix of peptidoglycans and proteins with a range of functions, such as protein folding, oxidation, and secretion. The outer membrane encapsulates the cell and is imbedded with various proteins that serve as efflux pumps, transporters, or porins. Additionally, the outer membrane is coated in a dense layer of LPS molecules that provide an added layer of protection by creating an additional permeability barrier for small hydrophobic molecules like antibiotics and detergents. LPS molecules are composed of a lipid A domain, a core oligosaccharide, and an O-antigen [36]. The lipid A domain anchors the LPS to the outer membrane and is the toxic component of LPS. The core oligosaccharide links the lipid A domain to the O-antigen, which is a long chain of repeated sugars. When mammalian hosts detect LPSs, the immune system is triggered to clear the bacteria [37]. However, high levels of LPSs, especially in immunocompromised patients, can cause sepsis and is potentially life-threatening. Adding to the difficulty of host clearance, bacteria have evolved to highly vary the O-antigen composition and structure to evade detection [35]. LPS is essential for most Gram-negative bacteria, thus making it a promising target to study. Here, we review the lipopolysaccharide transport (Lpt) pathway in Gram-negative *E. coli* that is responsible for transporting LPS from the inner membrane to the outer membrane for display.

### 1.2.2 Inner membrane

LPS synthesis occurs at the inner membrane. On the cytoplasmic side of the inner membrane, the lipid A and core oligosaccharide portions are synthesized and joined together through the Raetz pathway [36]. Also synthesized in the cytoplasmic side of the inner membrane, is the O-antigen oligosaccharide unit attached to a carrier molecule, undecaprenyl phosphate (UndP), through the Wzx/Wzy-dependent pathway [38]. Both components are flipped to the periplasmic side of the inner membrane in an ATP-dependent manner where final assembly occurs. The lipid A and oligosaccharide precursor molecule are transported to the periplasmic side of the inner membrane by an ATP-binding cassette (ABC) transporter, the MsbA flippase [39]. The O-antigen is flipped to the periplasmic side of the inner membrane by an ABC transporter, the Wzx flippase [38], [40]. Elongation of the O-antigen sugar chain occurs after transfer across the inner membrane. The O-antigen is attached to the core oligosaccharide by a ligase, the WaaL ligase, to form the mature LPS molecule ready for transport across the cell wall through the Lpt pathway [41].

The Lpt pathway consists of seven Lpt proteins, LptABCDEFG, that form a complex spanning from the inner membrane to the outer membrane **(Figure 3)**. The inner membrane portion of the Lpt pathway consists of LptB, LptC, LptF, and LptG. $LptB_2FGC$ forms an ABC transporter that extracts LPS from the inner membrane and into the complex [42], [43]. LptB is a cytoplasmic protein with an ATP and $Mg^{2+}$ ion binding site and distinct hydrophobic groove [44]. LptB dimerizes and exposes the two hydrophobic grooves towards the inner membrane, where LptF binds to one groove and LptG binds to the other. LptF and LptG both contain six transmembrane helices and a periplasmic β-jellyroll domain [42], [43]. They dimerize to form the core of the ABC transporter. LptC is composed of a single transmembrane helix and a periplasmic β-jellyroll domain [42]. The transmembrane helix of LptC was found to be inserted between the transmembrane domain of LptF and LptG, thus widening the cavity [42]. The β-jellyroll domain of LptC was found to interact with the β-jellyroll domain of

LptFG to extend above the complex into the periplasm [42]. Structures of the active and inactive form of $LptB_2FGC$ show large conformational rearrangements upon ATP hydrolysis for the extraction and expulsion of LPS [42]. LPS is presumed to be able to enter the cavity through the membrane. It has weak affinity with the transmembrane helices surrounding the cavity. When the transmembrane helix of LptC dissociates away from the cavity, thus narrowing the cavity, LPS is pushed towards the top of the cavity due to favorable electrostatic interactions [42]. The released transmembrane helix of LptC places less strain on the linker between the helix and β-jellyroll domain to allow LptC to stably bind to the β-jellyroll domains of LptFG by continuing the β-jellyroll motif [42]. Upon ATP binding and hydrolysis in LptB, large conformational changes occur that push LPS out of the LptFG cavity and into the hydrophobic groove of the LptFGC β-jellyroll motif directly above the cavity [42]. LPS is now able to be transported across the periplasm.

### 1.2.3 Periplasm

The periplasmic portion of the Lpt pathway mainly involves LptA, however LptFGC and LptD are also involved. All binding interactions between these proteins involve the β-jellyroll fold. The β-jellyroll fold consists of multiple anti-parallel β-strands arranged in a sheet, where one side of the sheet is highly hydrophobic and folds onto itself in a V-shape to bury the hydrophobic residues in a groove. This motif can build upon itself by continuing the β-sheet, as seen in the Lpt pathway. LptGFC contain a periplasmic β-jellyroll fold that is anchored to the inner membrane through the transmembrane parts of the proteins. LptA is solely a β-jellyroll fold [45]. LptD contains a periplasmic β-jellyroll fold, while the rest of the protein is imbedded in the outer membrane [46]. The proteins bind in a N-to-C fashion spanning the entire periplasm to form LptFGCA$_n$D [47]. LptA has been shown to self-oligomerize with itself, however the number of copies of LptA involved in the complex remains unknown [45], [48]. When these proteins oligomerize, the hydrophobic groove of the β-jellyroll fold forms a continuous channel that slightly twists with each addition [45]. LPS has been shown to bind to these proteins by inserting the acyl chains of the lipid A moiety into the hydrophobic groove [49], [50]. Once LPS is in the hydrophobic groove of the β-jellyroll fold of LptC, LPS is shuttled through the LptA bridge by interacting with hydrophobic binding pockets formed at each protein:protein interface [49]. The LptA bridge runs the length of the periplasm and arrives at LptD in the outer membrane. LPS is now positioned at the periplasmic side of the outer membrane ready to be transported across the outer membrane for display.

### 1.2.4 Outer membrane

The outer membrane portion of the Lpt pathway consists of LptD and LptE. LptD consists of the periplasmic β-jellyroll fold and a large bilobal β-barrel of 26 transmembrane antiparallel β-strands [46]. LptE is a lipoprotein composed of 4 β-strands and 2 α-helices [46]. LptE inserts itself into the larger lobe of LptD to further stabilize the

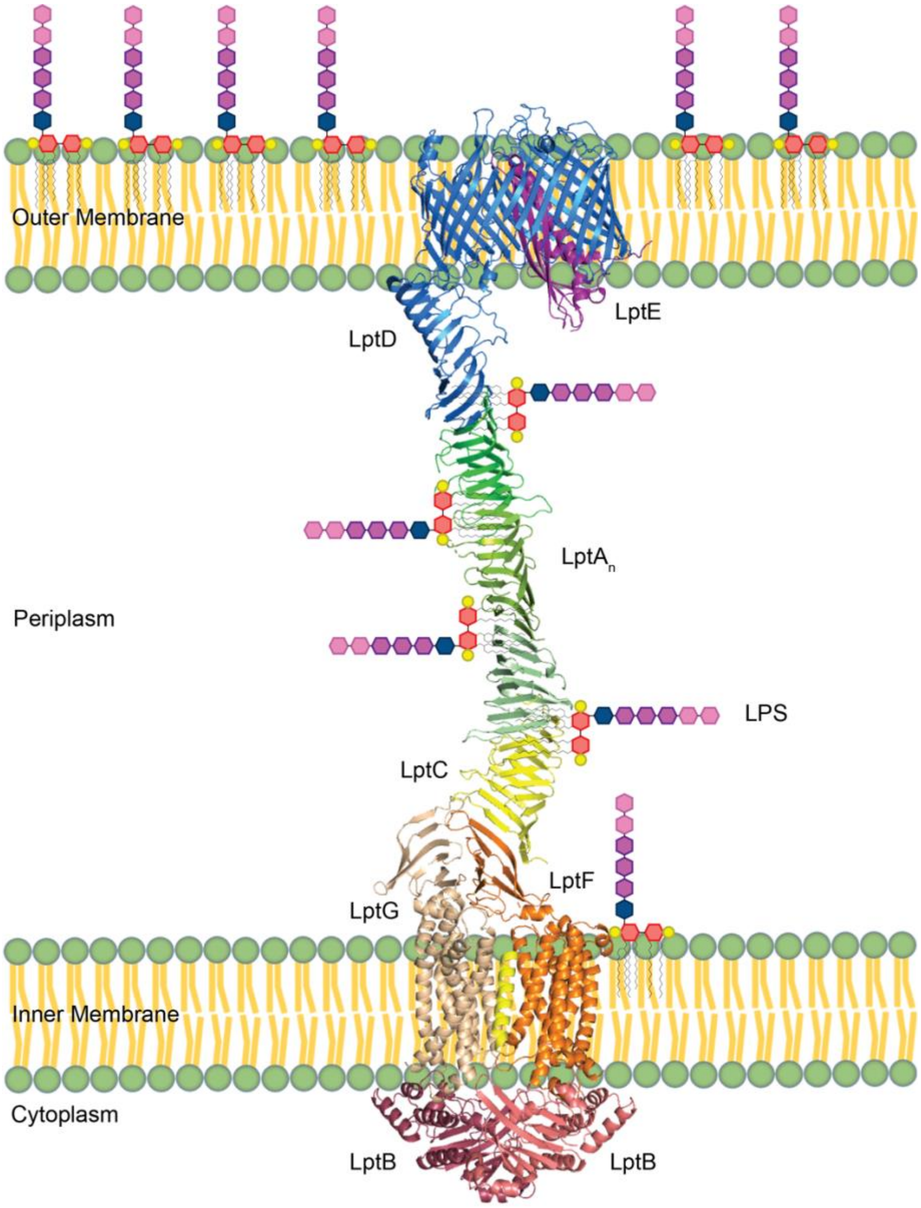
Outer Membrane
LptE
LptD
LptA$_n$
Periplasm
LPS
LptC
LptF
LptG
Inner Membrane
Cytoplasm
LptB
LptB

**Figure 3: Model of the lipopolysaccharide transport pathway.**
Lipopolysaccharide is transported from the inner membrane, through the periplasm, to the outer membrane through the Lpt pathway. The Lpt pathway consists of LptABCDEFG. The inner membrane portion contains LptB$_2$FGC (PDB ID: 6MIT) which extracts LPS from the inner membrane. The periplasmic portion contains an unknown number of LptA (PDB ID: 2R1A) forming a bridge to transport LPS across the periplasm. The outer membrane portion contains LptDE (PDB ID: 5IV9) which translocates LPS to the surface. The efficient transport of LPS through the machinery has been termed the "PEZ Model".

complex [51], [52]. Once LPS arrives at the β-jellyroll fold of LptD, it is unloaded into the β-barrel of LptD through a lateral gate between β1 and β26 [53]. The core oligosaccharide and O-antigen are inserted into the smaller lobe of the β-barrel while the lipid A moiety is inserted into the outer membrane [54]. The role of LptE, located inside of LptD, has yet to be fully elucidated. However, LptE was shown to bind LPS and is proposed to serve as the recognition site for LPS in the LptDE complex [55]. LPS is translocated through LptDE to the external side of the outer membrane for display. The constant loading of a new LPS molecule into the pathway is proposed to drive the efficient and steady stream of LPS molecules to be translocated across the cell wall. This model has been termed the "PEZ Model" due to similarities to the mechanical PEZ candy dispenser **(Figure 3)** [56].

### 1.2.5 Inhibition of lipopolysaccharide transport pathway

The Lpt pathway is essential for many Gram-negative bacteria to ensure proper cell wall development during infectivity, thus inhibiting it at various points would be beneficial to stopping bacterial spread. Various molecules have already been reported to do so, ranging from early *in vitro* characterization to phase III clinical trials. Two small molecule inhibitors of LptB, found through screening a small library of kinase inhibitors against the ATPase activity, were presented as scaffolds to improve upon [57]. Two small molecules, IMB-881 and IMB-0042, and one AMP, thanatin, were found to disrupt the LptC-LptA interface. IMB-881 and IMB-0042 were found through yeast two-hybrid screening of a small molecule library [58], [59]. IMB-881 was reported to bind to LptA, while IMB-0042 was reported to bind to both LptA and LptC [58]–[60]. Thanatin, which will be extensively covered in Section 1.3, was reported to bind to LptA and and also disrupt LptA-LptA interactions [61]. Thanatin was also reported to bind to LptD, but its role has been uncharacterized so far [61]. Murepavidin, a macrocyclic AMP, has been reported to disrupt the LptA-LptD interface by binding to LptD [62]–[64]. The scaffold for murepavadin was discovered from a library of β-hairpin-shaped macrocycles based on the AMP protegrin I structure [65]. The library was screened for bactericidal activity of Gram-positive and Gram-negative bacteria. After several rounds of optimization, one lead compound specific to *Pseudomonas* was identified, which was then licensed to various pharmaceutical companies for further development into murepavadin. Murepavadin was in two phase III clinical trials for intravenous treatment of broad *Pseudomonas* infection (ClinicalTrials.gov identifier: NCT03582007) and *P. aeruginosa* infection specifically (ClinicalTrials.gov identifier: NCT03409679) [66], [67]. However, these trials were terminated due to high nephrotoxicity. Murepavadin was reformulated to be an orally inhaled AMP and is currently in phase I clinical trial for treating *P. aeruginosa* infections in patients with bronchiectasis, including cystic fibrosis patients [68]. Further studying these recent advancements on targeting the Lpt pathway through AMPs will give insight on how to tackle other undruggable targets in a similar manner.

## *1.3 Thanatin*

### 1.3.1 Discovery and activity profile

Thanatin, an antimicrobial peptide (AMP) isolated from *Podisus maculiventris* (spined soldier bug), was first reported by Fehlbaum and colleagues after challenging the immune system of *P. maculiventris* through sepsis [69]. Thanatin was isolated, purified, and sequenced from the hemolymph of *P. maculiventris*. Thanatin was identified to be a cationic 21-residue long antimicrobial peptide (GSKKPVPIIYCNRRTGKCQRM) with activity against Gram-positive bacteria (*Aerococcus viridans*, *Micrococcus luteus*, *Bacillus megaterium*, *Bacillus subtilis*, *Staphylococcal aureus*, and *Pediococcus acidolactici*), Gram-negative bacteria (*Escherichia coli, Salmonella typhimurium, Klebsiella pneumoniae, Enterobacter cloacae, Erwinia carotovora*, and *Pseudomonas aeruginosa*), and fungi (*Neurospora crassa, Botrytis cinerea, Nectria haematococca, Trichoderma viride, Alternaria brassicola, Fusarium culmorum, Ascochyta pisi, Fusarium oxysporum*) at biological concentrations [69]. The broad activity against multiple microbes indicates multiple targets and mechanisms of thanatin, that have yet to be fully explored. To begin building an activity profile, the antimicrobial activity of thanatin and a D-enantiomer of thanatin were assessed. Thanatin was highly effective against the majority of these microbes assessed with most MIC values <25 µg/mL. D-thanatin maintained its level of antimicrobial activity against Gram-positive bacteria and fungi, except for *P. acidolactici* where potency was lost. D-thanatin was largely ineffective at targeting Gram-negative bacteria. This indicates that there is a chiral preference in Gram-negative bacteria, which is not the case for Gram-positive bacteria and fungi. Thanatin has also been shown to have antiviral activity against the chikungunya virus and tobacco mosaic virus [70], [71]. This broad but potent activity and unique profile makes thanatin an interesting molecule to study.

Thanatin from *Riptortus pedestris* (bean bug) was also recently identified as an AMP and regulator of the midgut microbiome. *R. pedestris* thanatin was found to have strong antimicrobial activity against *E. coli* and *S. aureus*, however it was found to have mixed activity against *Burkholderia* which was dependent on the cultured sources [72], [73]. As seen with other AMPs in insects, the expression level of thanatin was found to be elevated upon bacterial infection of *R. pedestris* [73]. In addition, when the genes for thanatin were silenced in *R. pedestris*, the midgut population of *Burkholderia* increased [73]. This example shows how a host organism can express thanatin to regulate the internal microbiome population.

### 1.3.2 Structure and functional mapping of thanatin

A solution NMR structure of *P. maculiventris* thanatin (PDB ID: 8TFV) reported thanatin to be composed of a flexible N-terminal region (G1 – P7), two well-defined β-strands (I8 – N12 and K17 – M21) held together by a disulfide bond (C11 and C18), and a cationic loop (R13 – G16) **(Figure 4A)** [74]. This was similarly seen with other β-sheet AMPs also consisting of an extended flexible arm and a β-hairpin structure held together by disulfide bonds, such as brevinins, protegrins, and tachyplesins [75]–[77]. Various studies have been performed to map the functionality of thanatin.

The N-terminal region of thanatin has been shown to be flexible and dispensable. The solution NMR structure (PDB ID: 8TFV) reported a poorly defined region from G1 – P7, especially from G1 – K3, indicating high flexibility in solution [74]. Truncation of G1 – K3 was shown to not affect the majority of antimicrobial activity against the tested Gram-positive bacteria, Gram-negative bacteria, and fungi [69]. Truncation of G1 – P5 also did not negatively affect the potency against the majority of tested Gram-positive bacteria, however it moderately reduced the activity against tested Gram-negative bacteria and fungi [69]. Truncation of G1 – P7 further reduced the activity against Gram-negative bacteria [69]. These results indicated that while G1 – K3 are dispensable for activity, the latter half of the N-terminal region is important. Characterization of thanatin with its target to observe the changes between apo and bound thanatin will be necessary to confirm the role of the N-terminal region.

The role and necessity of the disulfide bond has been questioned in literature. Mainly, the consensus is that the disulfide bond is necessary for *in vivo* activity. Disruption of the disulfide bond through single point mutations such as C11Y, C11S, C18S, and C18Y and double point mutations such as C11A/C18A and C11S/C18S were reported to greatly increase the MIC values or show no inhibition in *E. coli* [78]–[80]. One group reported that the disulfide bond was dispensable *in vivo* for various Gram-positive and Gram-negative bacteria, however linear thanatin was defined as thanatin still containing both cysteines but omitting an oxidation step during synthesis [81]. Linear thanatin was added into MIC assays, where oxidizing conditions of the periplasm of Gram-negative bacteria, or proximity of the two thiols most likely caused the formation of a disulfide bond in the duration of the experiment [82]. The role of the disulfide bond has yet to be investigated *in vitro* to quantify the effect on specific targets. The cationic loop of thanatin was shown to be necessary for activity and a place for improvement. Single point mutants such as R13C, R13H, and R14C in *E. coli* and double point mutant R13A/R14A in various Gram-positive and Gram-negative bacteria were shown to abolish antimicrobial activity [78], [83]. AMPs are overall cationic to help disrupt the membrane by binding to negatively charged lipopolysaccharides in Gram-negative bacteria and negatively charged teichoic acid and lipoteichoic acid in Gram-positive bacteria. Loss of overall net positive charge would reduce the affinity towards the membrane, thus decreasing membrane perturbation. Varying the length of the cationic loop was also explored and shown to not affect the antimicrobial activity of thanatin much. Deletion of T15 and/or G16 showed slight improved potency in various Gram-positive bacteria and minimal effect in various Gram-negative bacteria [79]. Insertion of one or two alanine residues after R14 showed a slightly reduced antimicrobial activity against various Gram-negative and Gram-positive bacteria [79]. Thanatin T15S has been a mutation of interest due to broad antimicrobial activity, low

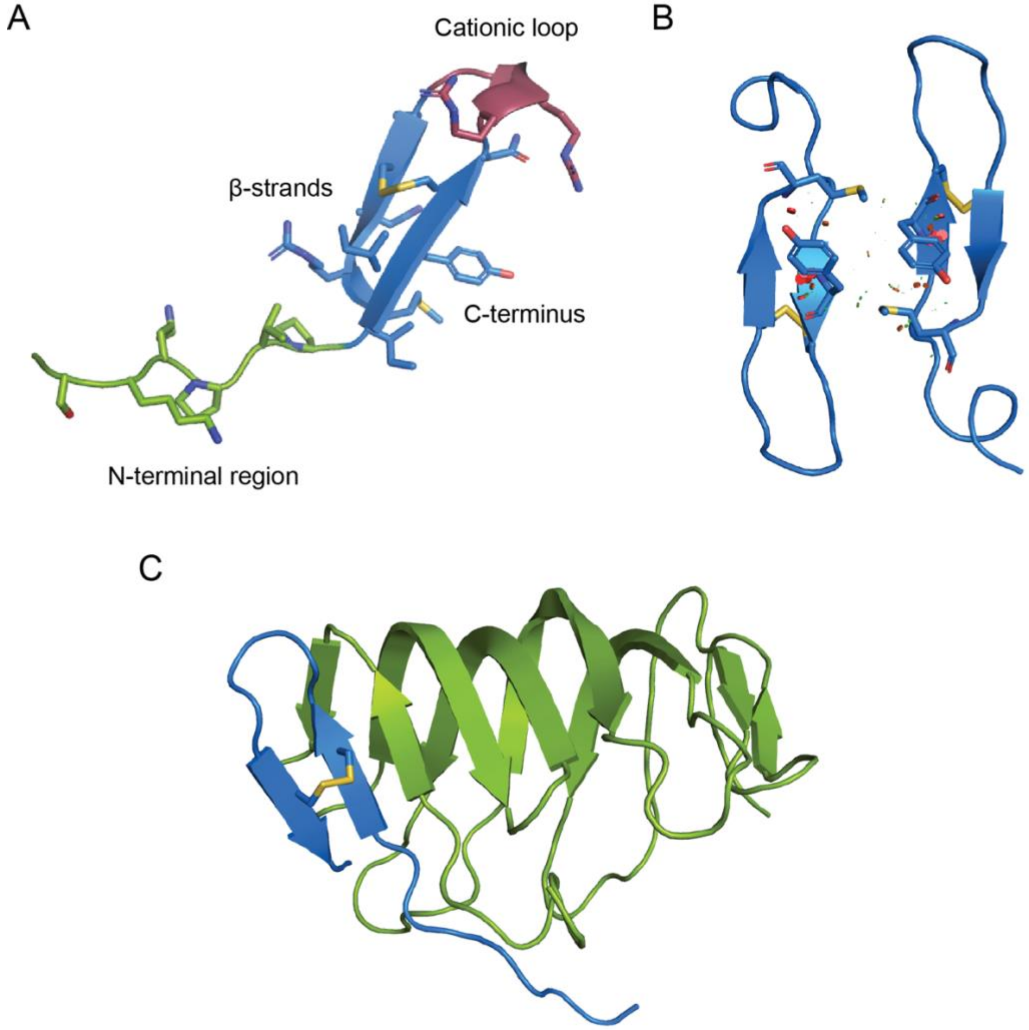

**Figure 4: Structures of thanatin.**

**A**, Solution NMR structure of thanatin (PDB ID: 8TFV) highlighting the N-terminal region (green), two anti-parallel β-strands held together by a disulfide bond (blue), and the cationic loop (maroon). The N-terminal region is highly flexible, with the first several residues being dispensable. The β-sheet core is amphipathic with hydrophobic residues lining the side opposite of the disulfide bond. The cationic loop is positively charged from several arginine residues. **B**, Solution NMR structure of thanatin dimer in LPS micelles (PDB ID: 5XO9) highlighting efficient packing at the hydrophobic interface. Dots represent favorable (green) and unfavorable (red) van der Waals interactions. **C**, Solution NMR structure of thanatin in complex with *E. coli* LptAm (PDB ID: 6GD5). Thanatin binds to the N-terminal of LptAm by continuing the β-sheet.

hemolytic activity, and high membrane perturbation [84]–[90]. However, characterization of thanatin T15S with its target has yet to be reported to fully understand its improvements over wild type thanatin.

The C-terminal region of thanatin was shown to be important for antimicrobial activity. Deletion of M21 moderately reduced the activity against tested Gram-positive bacteria and fungi, however it severely reduced the activity against tested Gram-negative bacteria [69]. Deletion of R20 – M21 and Q19 – M21 did not further reduce the activity in general, indicating a key role of M21 in activity [69]. Structural characterization of M21 with its target will be needed to further understand its role. Single point mutations such as Q19H and M21R abolished the antimicrobial activity against *E. coli* [78]. C-amidation of thanatin was also explored and shown to not negatively affect the antimicrobial activity against Gram-positive bacteria, Gram-negative bacteria, and fungi, except for *S. aureus* (improvement from no inhibition to MIC: 50 - 100 μg/mL) and *P. aeruginosa* (4 – 16-fold improvement in MIC) where significant improvement in potency was observed [69]. The role of C-amidation improving potency for these microbes has not been explored. Despite C-amidated thanatin not improving the antimicrobial against *E. coli*, it still maintains high efficacy (MIC: 1.5 – 16 μg/mL) against *E. coli*, especially in clinical isolates of extended-spectrum-β-lactamase-producing and multidrug resistant strains, highlighting the potential for clinical application [91].

### 1.3.3 Mechanisms of action against Gram-negative bacteria

The mechanisms of action of thanatin against Gram-positive bacteria, fungi, and viruses have yet to be fully elucidated. However, the mechanism of action against Gram-negative bacteria has been well characterized, especially for *E. coli*. Thanatin has been proposed to have four distinct mechanisms of action: 1) binding to LPS, thus inhibiting the Lpt pathway, 2) binding to LptA, 3) binding to LptD, and 4) binding to New Delhi metallo-β-lactamase (NDM-1).

Thanatin has been reported to bind to LPS molecules that densely coat the outer membrane of Gram-negative bacteria. LPS molecules consist of a lipid A moiety that anchors it to the outer membrane, an inner core region containing many negatively charged phosphate groups, an outer core region, and a variable O-antigen sugar chain. The cluster of negative charges in the inner core region is stabilized by the binding of $Ca^{2+}$ and $Mg^{2+}$ ions. LPS molecules provide protection against antimicrobial compounds, such as antibiotics and detergents, by densely packing the surface and creating a protective barrier [92]. LPS molecules also aid in evading host detection by modifying the O-antigen sugar chain [92]. Thanatin, cationic in nature, has been shown to bind to LPS and displace the $Ca^{2+}$ and $Mg^{2+}$ ions to destabilize LPS layer [93]. Binding studies have reported that thanatin binds to LPS about 100-fold tighter than $Ca^{2+}$ ions and about 200-fold tighter than $Mg^{2+}$ ions [83], [93]. It was shown that thanatin dimerizes at the hydrophobic interface in the presence LPS micelles (PDB ID: 5XO9), with almost all of the residues in contact with a LPS molecule, thus creating large LPS micelle aggregates **(Figure 4B)** [83]. Bacterial cells treated with thanatin have been shown to have physical morphological changes such as elongation, roughing of the surface, and increased agglutination, which would aid in host clearance through phagocytosis [83], [87], [93]. In addition, the destabilization of the LPS layer would increase membrane permeability.

Thanatin has been shown to inhibit the Lpt pathway by binding to the lipopolysaccharide export system protein A (LptA) and lipopolysaccharide export system protein D (LptD). Both LptA and LptD were identified as targets of thanatin through a pulldown assay in *E. coli* [61]. The interaction for thanatin has been well characterized against LptA. Binding experiments have reported high affinity of thanatin towards LptA in the nanomolar range [61]. A solution NMR structure (PDB ID: 6GD5) showed that thanatin binds to the N-terminal edge of LptA by continuing the β-jellyroll motif **(Figure 4C)** [61]. This appendage of thanatin prevents complex formation between LptC and LptA and LptA self-oligomerization, thus disrupting the Lpt pathway [60]. The interaction for thanatin with LptD has been poorly understood. However, binding experiments have also reported high affinity of thanatin towards LptD [61]. Further characterization and a high-resolution structure of thanatin with LptD have yet to be reported. However, it is plausible that thanatin binds to the periplasmic portion of LptD which also contains the β-jellyroll motif with the N-terminus exposed. The full-length structure of *E. coli* LptD/E that includes the periplasmic portion has yet to be reported, however the structure has been reported for *K. pneumoniae* (PDB ID: 5IV9), a highly similar Gram-negative bacterium [46]. Thanatin binding to LptD here would disrupt the complex formation between LptA and LptD. Overall, thanatin has been shown to disrupt the Lpt pathway in multiple locations, contributing to its strong antimicrobial activity.

Thanatin has also been shown to bind to and inhibit NDM-1 [93]. NDM-1 is an enzyme produced by Gram-negative bacteria that hydrolyzes almost all clinically used β-lactam antibiotics. The innate ability to degrade β-lactams serves as an internal defense against antibiotics. Inhibiting NDM-1 would allow β-lactam antibiotics to be more effective by maintaining stability. The mechanism of action has been shown to be through the displacement of the two $Zn^{2+}$ cofactors from the active site of NDM-1 [93]. However, a high-resolution complex structure of thanatin with NDM-1 to define the binding site has yet to be reported. Combination treatment of thanatin with β-lactam antibiotics in *E. coli* showed reversed antibiotic resistance and increased survival rate in murine models [93]. These studies indicate that thanatin has synergistic benefits with existing pharmacological agents, that could aid in restoring clinical usage of drugs that microbes have already become resistant to.

### 1.3.4 Thanatin hybrid peptides

Recent efforts to improve thanatin by combining it with other AMPs have been made. A hybrid peptide combining the N-terminal half of cecropin B, a 33-residue long α-helical AMP from silkworms, with the C-terminal half of *P. maculiventris* thanatin, maintained antimicrobial activity against Gram-positive *B. subtilis* and fungal *C. albicans*, however no overall improvements were reported [94]. A hybrid of attacin A, a 190-residue long random coil AMP from the silk moth, with a truncated *P. maculiventris* thanatin on the C-terminus, also showed retained potency in *E. coli* and *S. aureus*, but no synergistic improvements [95]. A hybrid of the C-terminal half of melittin, a 26-residue long α-helical AMP from bee venom, with the N-terminal half of *P. maculiventris* thanatin also reported retained potency in Gram-positive and Gram-negative bacteria, but no synergistic improvements [96]. A hybrid of the N-terminal half of LfcinB, a 25-residue long β-hairpin AMP from bovine, combined with the C-terminal half of *P. maculiventris* thanatin showed improved antimicrobial activity against *P. aeruginosa, E. aerogenes*, and *K. pneumoniae* ranging from 2-4 fold over LfcinB and thanatin alone [97]. A hybrid of the N-terminal half of nisin, a 34-residue long lantibiotic from bacterium *Lactococcus lactis*, with ripcin, a macrocyclic lanthipeptide derived from *R. pedestris* thanatin, was shown to improve antimicrobial activity in *S. aureus*, along with various other Gram-positive and Gram-negative bacteria, over the nisin fragment and ripcin alone [98]. Interestingly, a triple hybrid of latarcin, an α-helical AMP from spider venom, PAP1 protein, a ribosome-inactivating protein from plants, and thanatin was shown to have improved anti-viral activity against the Chikungunya virus over each individual component alone [71]. The fusion of thanatin to molecules for improved activity may indicate a biological tool application as a cell permeating peptide due to its cationic nature and effectiveness at targeting membranes [99]. Future mechanistic studies on how these hybrid peptides synergistically improve potency will be needed to better understand them. Even for the hybrid peptides that maintained but did not improve potency, studies on how to improve their scaffolds will be worthwhile for the next generation of thanatin derived peptides.

## 1.4 Summary

AMPs have proven to be a valuable tool in developing new agents to combat pathogens. AMPs come from all forms of life and are diverse in structure and function, however they all have the same purpose of regulating the microbiome population. Discovery of novel AMPs is rapidly increasing as new tools are developed to identify them. Thanatin is a β-hairpin AMP from the spined soldier bug that has proven to be multi-faceted and effective in multiple ways, such as destabilizing membranes and targeting the Lpt pathway. Structural studies of thanatin have mapped out the function of various parts of the 21-residue long peptide. Further characterization of thanatin with specific targets will be insightful on its various mechanisms of action, which have been well characterized in Gram-negative bacteria. The profile of thanatin will only become more impressive as undiscovered mechanisms of actions, such as those against Gram-positive bacteria, fungi, and viruses, have yet to be elucidated. Gram-positive bacteria, fungi, and viruses do not contain LPS, the Lpt pathway, or NDM-1, hence new targets for thanatin will need to be identified. The continued efforts to harness the benefits of thanatin being a broad but highly effective AMP and further improve upon its scaffold demonstrates the growing potential of developing thanatin as a pharmacological agent.

## 1.5 Outline of dissertation

This dissertation focuses on the discovery of novel thanatin orthologs and their characterization against the Lpt pathway in Gram-negative bacteria. Additional work outside of this topic is presented in the appendix.

Chapter 2 focuses on the discovery and application of these thanatin orthologs against *E. coli* LptA, structural and binding studies of their interactions, inhibition studies, and design of novel scaffolds for the next generation of thanatin-based AMPs. Chapter 2 is adapted from the following manuscript in progress:

> **Huynh, K**.*, Kibrom, A.*, Donald, B.R., and Zhou, P. Discovery, characterization, and redesign of potent antimicrobial thanatin orthologs from *Chinavia ubica* and *Murgantia histrionica* targeting *E. coli* LptA. *Manuscript in preparation*.
> *These authors contributed equally.

Chapter 3 focuses on the application of these thanatin orthologs against *P. aeruginosa* and structural and binding studies of their interactions, and inhibition studies.

The Conclusions chapter summarizes the work presented, their implications, and future directions.

Appendix A focuses on a computational re-evaluation of the hepatitis c virus viroporin, p7, and the effect of lipid composition on the function of p7.

# 2. Characterizing thanatin against *Escherichia coli* LptA

The following chapter is adapted from a manuscript in progress:

**Huynh, K**.*, Kibrom, A.*, Donald, B.R., and Zhou, P. Discovery, characterization, and redesign of potent antimicrobial thanatin orthologs from *Chinavia ubica* and *Murgantia histrionica* targeting *E. coli* LptA. *Manuscript in preparation*.
*These authors contributed equally.

The following statements reflect the contribution of work mentioned in this chapter. The consensus sequence search was done by K. Huynh after helpful discussions with Dr. Abhishek Chhetri. Molecular cloning and protein purification were shared between K. Huynh and A. Kibrom. All BLI binding assays and MIC inhibition assays were done by K. Huynh. The crystallization and structure determination of *P. maculiventris* thanatin with E. coli LptA and *C. ubica* thanatin with *E. coli* LptA were done by K. Huynh. The crystallization and structure determination of *M. histrionica* thanatin with *E. coli* LptA was done by A. Kibrom. The stapled and cyclized peptide designs of thanatin were done by K. Huynh. All written text and analysis presented here were done by K. Huynh.

## 2.1 Background

Antimicrobial peptides (AMPs) are important components of the host innate defense system against pathogens. Despite their small sizes, AMPs can possess potent antiviral, antibacterial, and antifungal activity. *Podisus maculiventris* thanatin, an AMP isolated from the spined soldier bug, is a 21-residue peptide with broad inhibitory properties towards bacteria and fungi [69]. Although the mechanism of action for *P. maculiventris* thanatin towards Gram-positive bacteria and fungi remains to be elucidated, recent studies suggest that *P. maculiventris* thanatin can interfere with multiple pathways in Gram-negative bacteria. In *E. coli*, binding of *P. maculiventris* thanatin to lipopolysaccharide (LPS) on the outer layer of the outer membrane dislodges $Ca^{2+}$ and $Mg^{2+}$ ions, causing membrane instability and bacterial agglutination for host clearance [83], [93]. Likewise, *P. maculiventris* thanatin has been reported to disrupt the LPS translocation from the inner membrane to the outer membrane by binding to components of the lipopolysaccharide transport system, LptA and LptD [60], [61], ultimately disrupting the assembly of the bacterial outer membrane.

Apo *P. maculiventris* thanatin has been shown to adopt an architecture of β-hairpin held together by a disulfide bond in solution [74]. Using a truncated *E. coli* LptA protein that maintains a monomeric state in solution (LptAm), Vetterli and colleagues reported that *P. maculiventris* thanatin pairs its own β-hairpin with the N-terminal edge of the LptAm β-sheet, thus disrupting the LptA oligomer required for LPS transport [49], [61]. Despite conflicting reports about the requirement of a cysteine disulfide bond for thanatin activity [78]–[81], truncation and mutagenesis studies in general support the observations that the N-terminal loop of thanatin is dispensable [69] and the cationic loop [78], [83] and the C-terminal region are required for *in vivo* inhibition [69].

In this study, we report the discovery of distinct thanatin orthologs from mainly diverse insect species by mining publicly available genomic databases. We show that thanatin orthologs from *Chinavia ubica* and *Murgantia histrionica* bind *E. coli* LptAm more tightly *in vitro* and display stronger antibiotic activity than the widely studied thanatin from *P. maculiventris* [61], [69], [81], [93]. Our analysis from high-resolution crystal structures of thanatin orthologs from *P. maculiventris*, *C. ubica* and *M. histrionica* in complex with *E. coli* LptAm reveal key residues in *M. histrionica* thanatin that contribute to the tighter LptA binding affinity and superior antibiotic activity towards *E. coli*.

## 2.2 Novel thanatin orthologs reveal consensus sequence

Thanatins from *P. maculiventris* and *R. pedestris* have previously been reported with potent inhibitory properties against Gram-negative bacteria [69], [72]. To search for additional thanatin orthologs, we used the amino acid sequence of thanatin from *P. maculiventris* to query genomic databases deposited at the National Center for Biotechnology. A total of 16 thanatin-like sequences from 13 unique species were discovered (**Figure 5A**), which can be aligned and grouped into 11 unique sequences. Analysis of the consensus sequence logo shows that thanatin peptides share a pattern of $G_2S_3K_4P_5V_6P_7I_8I_9(A/Y)_{10}C_{11}N_{12}R_{13}K_{14}T_{15}G_{16}K_{17}C_{18}(T/R/Q/K)_{19}R_{20}(I/M/F/L)_{21}$ (**Figure 5B**). It is important to note that A10 only occurs when position 21 is an isoleucine (I21), whereas Y10 is observed with multiple types of hydrophobic residues at position 21, such as methionine, phenylalanine, and leucine.

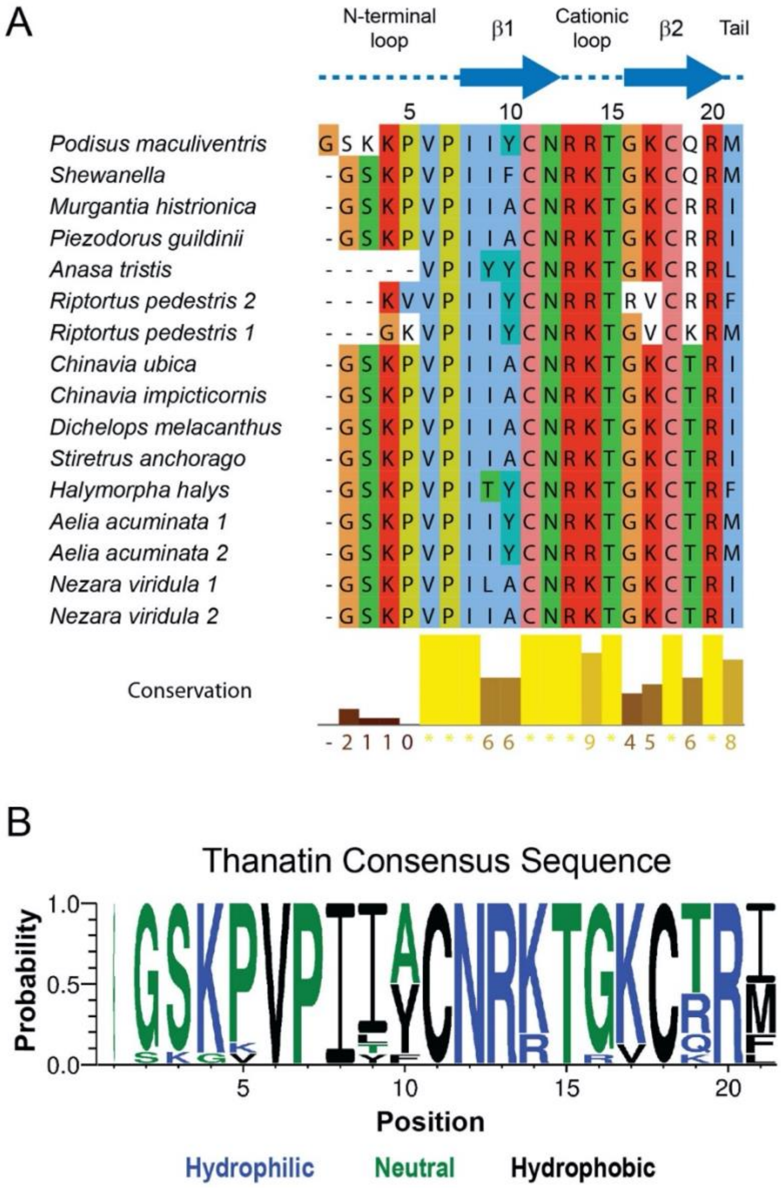

**Figure 5: Sequence alignment of thanatin orthologs reveals a consensus sequence.**

**A**, A sequence alignment of 16 thanatin sequences identified through genomic databases, with sequence conservation shown as a bar graph underneath the sequences. There are 11 unique sequences from 13 distinct species. **B**, A consensus logo generated from the thanatin sequences shows overall high conservation except for N-terminal loop (positions 1-5) and positions 9, 10, 19 and 21.

## *2.3 Thanatin orthologs bind tighter to E. coli LptA and display stronger antibiotic activity than P. maculiventris thanatin*

In order to assess the interaction of these thanatin orthologs towards *E. coli* LptA, we developed a bio-layer interferometry (BLI) binding assay. Eight thanatin orthologs from *Aelia acuminata* (two isoforms), *Chinavia ubica*, *Halyomorpha halys*, *Murgantia histrionica*, *Nezara viridula*, *Podisus maculiventris,* and *Riptortus pedestris* together with *E. coli* LptAm [49] were cloned and purified. LptAm was biotinylated and immobilized onto streptavidin biosensors, which were then immersed into solutions containing different concentrations of target thanatin peptides. Representative titration and binding curves of *P. maculiventris* thanatin with LptAm are shown in **Figure 6A and 6B**. The steady-state binding association was fit to a 1:1 binding model, yielding a $K_d$ value of 1.8 ± 0.2 nM. Representative titration curves and steady-state binding analyses of individual thanatin orthologs are summarized in **Figure 6C** (raw data shown in **Figure 7**). Surprisingly, with the exception of *R. pedestris* thanatin, which binds similarly to *P. maculiventris* thanatin ($K_d$ = 2.0 ± 0.4 nM versus $K_d$=1.8 ± 0.2 nM), all of the remaining thanatin peptides bind LptAm at least 2-fold more tightly ($K_d$ = 0.4 - 0.8 nM).

Two of these thanatin orthologs also displayed stronger antibiotic activity against *E. coli* by at least 2-fold **(Figure 6D)**. *P. maculiventris* thanatin had an MIC of 2.9 ± 0.2 µg/mL against the *E. coli* ATCC 25922 strain, whereas thanatin from *C. ubica* had a MIC value of 1.4 ± 0.3 µg/mL. Thanatin from *M. histrionica* appeared to be the most

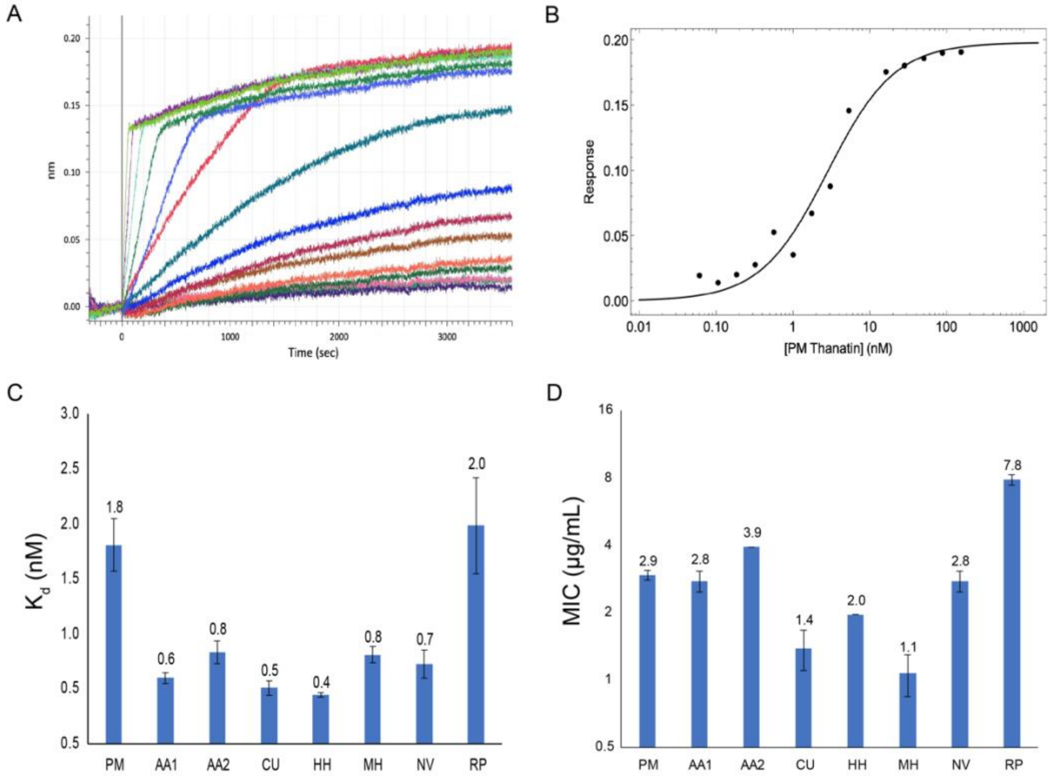

**Figure 6: Identification of thanatin orthologs that improve binding and inhibition to *E. coli*.**

**A**, Representative BLI binding curves of the *P. maculiventris* thanatin titration with LptAm immobilized on the BLI biosensor tip. **B**, Steady-state binding curve of *P. maculiventris* thanatin with LptAm. **C**, Comparison of binding affinities of thanatin orthologs with LptAm. Error bars indicate SEM (minimum of at least n=3). **D**, MIC values of thanatin orthologs against *E. coli* ATCC 25922. MIC values are calculated as geometric means, and the error bars represent SEM (minimum of at least n=3).

potent antibiotic, showing an MIC of 1.1 ± 0.2 µg/mL, improving the potency of *P. maculiventris* thanatin by approximately 2.8-fold.

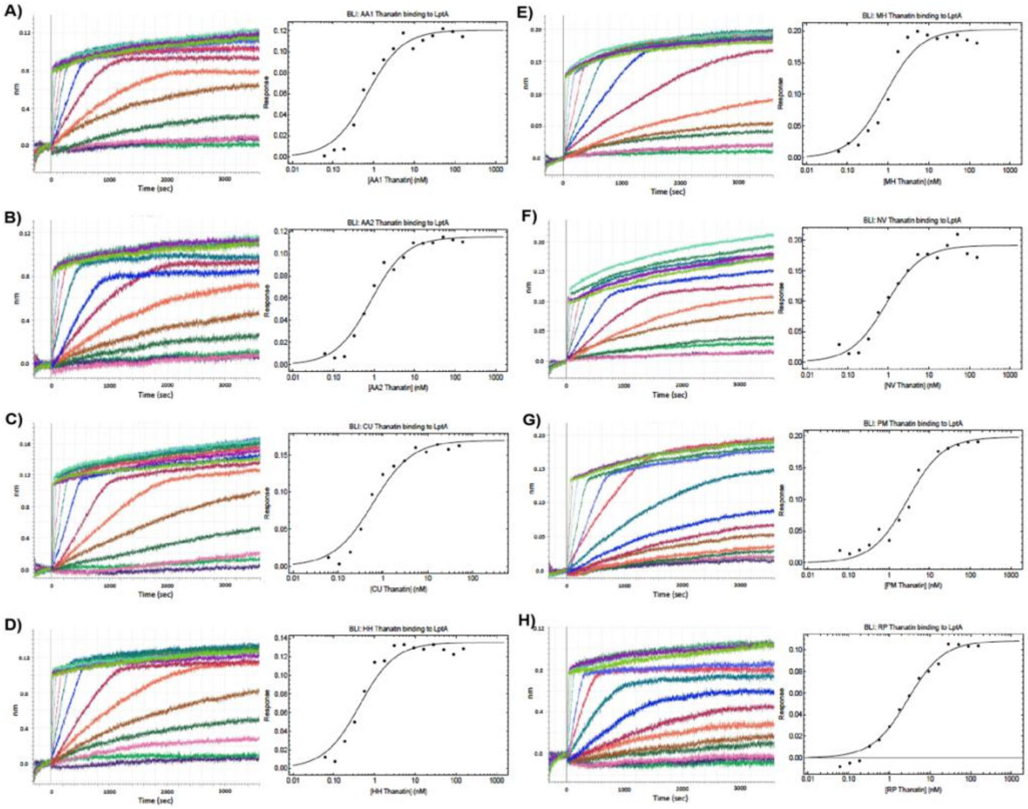

**Figure 7: Representative BLI binding curves for each thanatin ortholog binding to *E. coli* LptA.**

BLI association traces of a thanatin titration (left) and binding curve fit to steady state 1:1 binding model (right) are shown for the following thanatin orthologs: A) *Aelia acuminata (1)*, B) *Aelia acuminata (2)*, C) *Chinavia ubica*, D) *Halyomorpha halys*, E) *Murgantia histrionica*, F) *Nezara viridula*, G) *Podisus maculiventris*, H) *Riptortus pedestris (1)*.

## *2.4 Crystal structure of P. maculiventris thanatin bound to E. coli LptAm reveals a domain-swapped dimer*

In order to obtain a molecular understanding of the *P. maculiventris* thanatin interaction with *E. coli* LptA, we crystallized the LptAm - *P. maculiventris* thanatin complex in the $P2_12_12_1$ space group, which diffracted to 2.43 Å resolution. Unexpectedly, the crystal structure revealed a domain-swapped dimer of LptAm in the shape of a

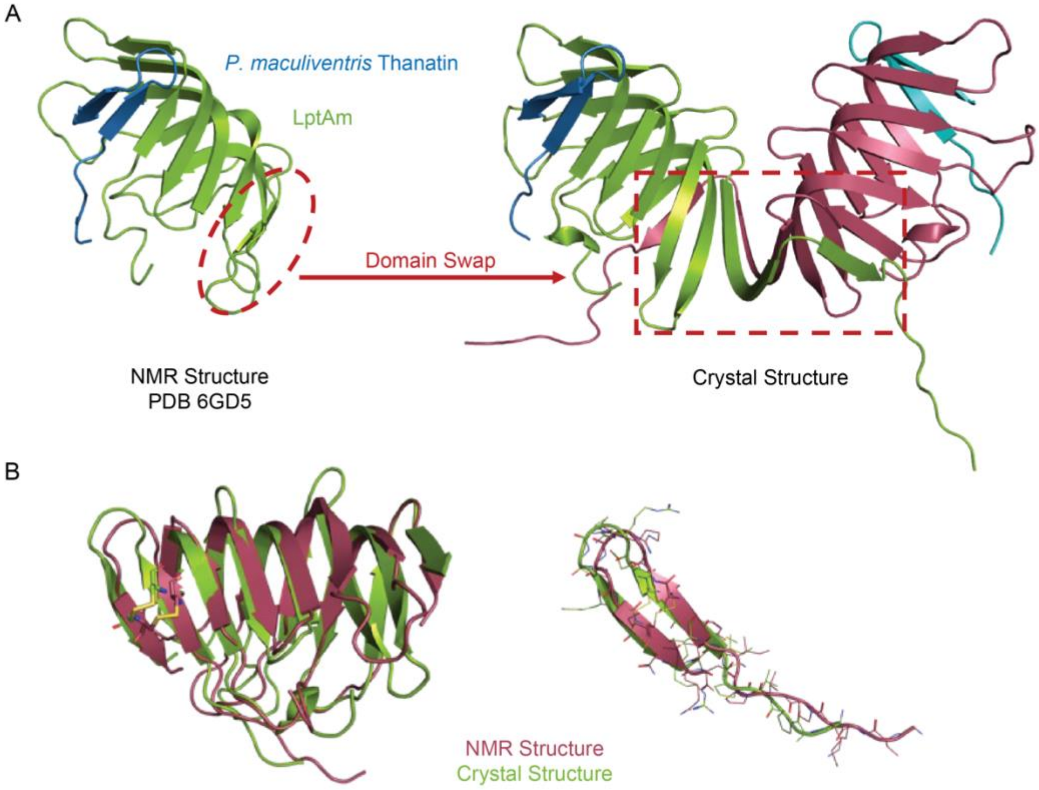
A
P. maculiventris Thanatin
LptAm
Domain Swap
NMR Structure
PDB 6GD5
Crystal Structure
B
NMR Structure
Crystal Structure

**Figure 8: Structural comparison of LptAm in complex with *P. maculiventris* thanatin by NMR and crystallography.**

**A**, Previously published solution NMR structure (PDB ID: 6GD5) showed a monomeric LptAm-thanatin complex. Crystal structure reveals a domain swapped dimer of the heterodimeric LptAm-thanatin complex. **B**, Overlay of the NMR structure and crystal structure show good backbone agreement for the β-strands, but poor agreement in the loop regions and side chain rotamers.

butterfly, with the terminal two β-strands on each protomer pairing up with the edge of the central β-sheet in the adjacent protomer **(Figure 8A)**. *P. maculiventris* thanatin forms a β-hairpin held together by a disulfide bond and appends to LptAm via parallel β-strand interactions through the N-terminal β-strand of LptAm, similar to the previously reported solution NMR structure (PDB 6GD5). As LptAm and LptAm in complex with

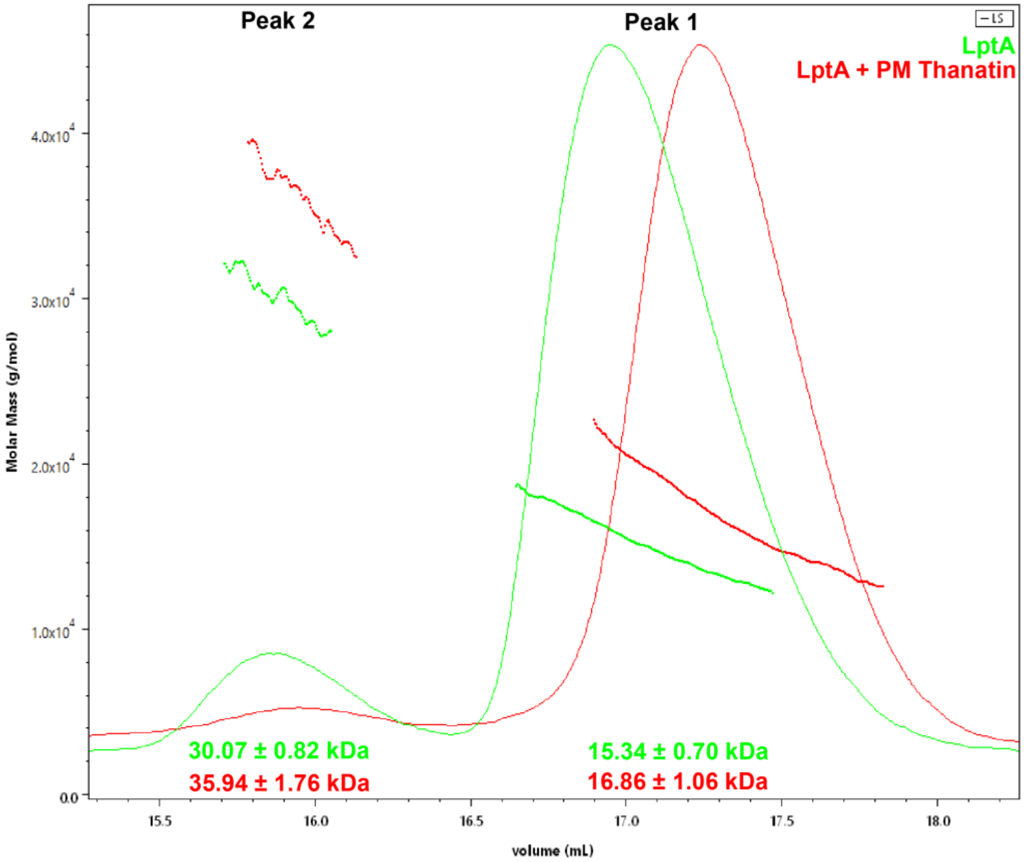

**Figure 9: Truncated LptA is monomeric in solution as shown by SECMALS.** A 500 μL sample of truncated LptA (15 mg/mL) (green) and truncated LptA in complex with *P. maculiventris* thanatin (15 mg/mL) (red) were separately injected onto a Superdex200 Increase 10/300 GL column in buffer (50 mM sodium phosphate pH 7.5, 150 mM NaCl). Truncated LptA is primarily monomeric (expected MW: 14.52 kDa) in solution with a small dimeric (expected MW: 29.04 kDa) population. Truncated LptA in complex with *P. maculiventris* thanatin is primarily monomeric (expected MW: 16.94 kDa) in solution with a small dimeric (expected MW: 33.87 kDa) population.

*P. maculiventris* thanatin are primarily monomeric in solution with a small dimeric population, shown via SEC-MALS **(Figure 9)**, the dimeric state of LptAm likely results from the high protein concentrations under crystallization conditions.

The overall features of the *P. maculiventris* thanatin interaction with LptAm in the crystal structure are similar to that of the NMR structure (PDB 6GD5). When disregarding the terminal two β-strands, the backbones are in good agreement with a RMSD of ~1.9 Å **(Figure 8B)**. The areas of poor agreement are largely in the loop regions, along with great variance in the side chain rotamers (overall total RMSD of ~2.4 Å), which is expected due to dynamic movement in solution **(Figure 8B)**. Furthermore, the domain-swapping occurs at the C-terminus of LptAm which is far removed from the binding interface between LptAm and thanatin, thus the thanatin binding interface is entirely unperturbed.

## 2.5 Shared features of LptAm-thantin ortholog complexes

We next determined the crystal structures of *C. ubica* thanatin and *M. histrionica* thanatin in complex with *E. coli* LptAm to understand why these thanatin orthologs bind more tightly to *E. coli* LptA than *P. maculiventris* thanatin. Both *C. ubica* thanatin and *M. histrionica* thanatin were co-crystallized with *E. coli* LptAm and took the same $P2_12_12_1$ space group as *P. maculiventris* thanatin in complex with *E. coli* LptAm, and their crystal structures were resolved to 1.90 Å and 1.80 Å, respectively. Each unit cell similarly contains two domain-swapped LptAm - thanatin complexes as seen for the LptAm - *P. maculiventris* thanatin complex (**Figure 10;** refinement statistics for all three structures can be found in the **Table 1**).

In all three structures, LptAm forms a β-jellyroll domain and thanatin forms a β-hairpin held together by an invariant disulfide bond **(Figure 11A)**. Removal of the disulfide bond in *P. maculiventris* thanatin via a C11A/C18A mutation reduced the LptAm binding affinity by 26-fold ($K_d$: 46.9 ± 4.9 nM versus WT $K_d$: 1.8 ± 0.2 nM) **(Figure 11E)** and completely abolished the antibiotic activity (MIC >125 (μg/mL) **(Figure 11F)**, showing that the disulfide bond stabilized β-hairpin structure is an important feature for maintaining the biological activity of thanatin.

The thanatin peptides bind to the N-terminus of LptAm by aligning their own N-terminal β-strands at the interface parallel to the N-terminal β-strand of LptAm and forming backbone interactions **(Figure 11A)**. As a result, thanatin disrupts the LptA oligomerization "bridge" in lipid A transport by binding asymmetrically to the N-terminus of LptAm, causing an uneven surface to disfavor another β-jellyroll fold from binding.

All thanatin peptides contain a dispensable and flexible N-terminal tail that is not well defined for the first several residues **(Figure 11B)**. Deletion of first three residues (ΔG1-K3) in *P. maculiventris* thanatin did not negatively affect its interaction with LptAm ($K_d$: 0.7 ± 0.1 nM) **(Figure 11E)** nor its antibiotic activity (MIC: 3.1 ± 0.3 μg/mL) **(Figure 11F)** suggesting that this region is overall dispensable for efficient binding. This is in agreement with previously reported results of minimal effect observed between full length and truncated (ΔG1-K3) *P. maculiventris* thanatin on *E. coli* inhibition [69]. Similarly, deletion of K3 (ΔK3) in *P. maculiventris* thanatin, which mimics the shorter peptide length observed in several orthologs, has no major consequence due to its location in the dispensable N-terminal tail ($K_d$: 1.1 ± 0.03 nM, MIC: 4.9 ± 0.3 μg/mL) **(Figure 11E, F)**.

The cationic loop, specifically R13, in all three thanatin peptides is involved in extensive intermolecular interactions with LptAm **(Figure 11C)**, suggesting that it is an important residue for LptAm binding. Mutations of R13 that would break contact to LptAm, such as R13C or R13H, were reported to abolish inhibition of *E. coli* growth [78]. In contrast to R13, the residue at position 14 has no obvious contact with LptAm, though these two residues are invariantly conserved as basic residues (Arg/Lys). R14K has no significant changes ($K_d$: 1.5 ± 0.3 nM, MIC: 3.9 ± 0.0 μg/mL) **(Figure 11E, F)**. However, removal of positively charged residues in the cationic loop via *P. maculiventris* thanatin R13A/R14A, was reported to abolish inhibition of *E. coli* growth [83], showing that it plays a critical role in the efficacy of thanatin presumably due to enhanced membrane permeability.

The C-terminal carboxyl group in all three thanatin peptides forms bidentate hydrogen bonds with sidechain of LptAm R76 to position thanatin with the hydrophobic core of LptAm **(Figure 11D)**. C-amidation of *P. maculiventris* thanatin was reported to increase the MIC value by 2-fold, most likely due to the loss of charged interaction with LptA R76 [69]. Deletions of residues 19 – 21, 20 – 21, or 21 in

**Table 1: X-ray data collection and refinement statistics of *E. coli* LptA-Thanatin complexes.**

| | *E. coli* LptAm-<br>*P. maculiventris*<br>Thanatin | *E. coli* LptAm-<br>*Chinavia ubica*<br>Thanatin | *E. coli* LptAm-<br>*Murgantia histrionica*<br>Thanatin |
|---|---|---|---|
| PDB | XXXX | YYYY | ZZZZ |
| Data collection | | | |
| Wavelength (Å) | 0.9792 | 0.9792 | 0.9792 |
| Space group | P 21 21 21 | P 21 21 21 | P 21 21 21 |
| Cell dimensions | | | |
| *a*, *b*, *c* (Å) | 44.95, 61.83, 150.33 | 45.51, 62.13, 150.08 | 45.60, 62.47, 147.12 |
| α, β, γ(Å) | 90.0, 90.0, 90.0 | 90.0, 90.0, 90.0 | 90.0, 90.0, 90.0 |
| Resolution (Å) | 47.75 - 2.43<br>(2.52 - 2.43) | 47.86 – 1.90<br>(1.97 – 1.90) | 47.62 – 1.80<br>(1.86-1.80) |
| $R_{\mathrm{meas}}$ | 0.06248 (0.5848) | 0.067 (0.852) | 0.062 (0.708) |
| $CC_{1/2}$ | 0.999 (0.98) | 0.999 (0.974) | 0.998 (0.936) |
| Mean $I/\sigma I$ | 23.27 (3.94) | 20.00 (3.22) | 14.17 (2.83) |
| Completeness (%) | 99.90 (99.94) | 99.61 (99.20) | 97.89 (98.59) |
| Redundancy | 11.5 (11.3) | 13.1 (13.7) | 5.5 (5.7) |
| Total reflections | 189,877 (18,210) | 450,466 (45,983) | 216,971 (22,188) |
| Unique reflections | 16,463 (1,614) | 34,382 (3,364) | 39,095 (3,874) |
| | | | |
| Refinement | | | |
| $R_{\mathrm{work}}$ / $R_{\mathrm{free}}$ | 0.1938/0.2416 | 0.197/0.239 | 0.181/0.212 |
| No. atoms | 2406 | 2384 | 2608 |
| Protein | 2237 | 2245 | 2274 |
| Water | 169 | 139 | 334 |
| Average *B*-factors | 23.68 | 49.39 | 39.18 |
| Protein | 23.48 | 49.21 | 38.16 |
| Water | 26.42 | 52.28 | 46.11 |
| R.m.s. deviations | | | |
| Bond lengths (Å) | 0.004 | 0.007 | 0.015 |
| Bond angles (°) | 0.62 | 0.92 | 1.37 |
| Ramachandran | | | |
| Favored (%) | 98.24 | 97.21 | 98.24 |
| Allowed (%) | 1.76 | 2.79 | 1.76 |
| Outliers (%) | 0.00 | 0.00 | 0.00 |

*Values in parentheses are for highest-resolution shell.

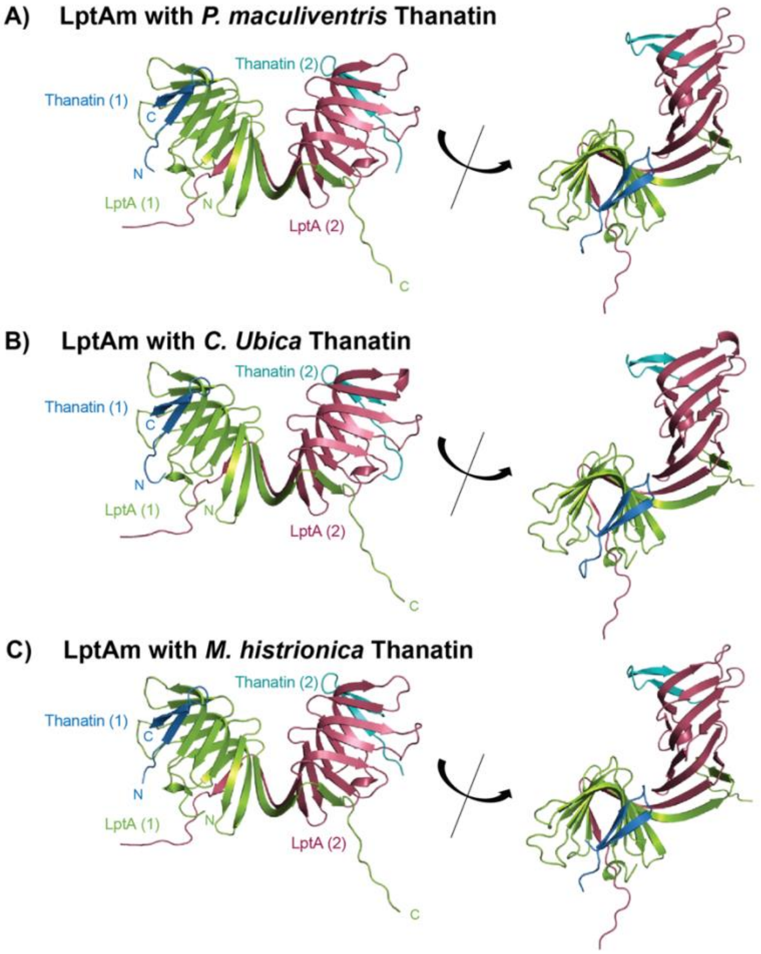
A) LptAm with *P. maculiventris* Thanatin
Thanatin (2)
Thanatin (1)
C
N
LptA (1)
N
LptA (2)
C
B) LptAm with *C. Ubica* Thanatin
Thanatin (2)
Thanatin (1)
C
N
LptA (1)
N
LptA (2)
C
C) LptAm with *M. histrionica* Thanatin
Thanatin (2)
Thanatin (1)
C
N
LptA (1)
N
LptA (2)
C

**Figure 10: Crystal structures of LptAm in complex with thanatin orthologs.** Ribbon diagram of **A**, LptAm in complex with P. maculiventris, **B**, LptAm in complex with *C. ubica* thanatin, and **C**, LptAm in complex with *M. histrionica* thanatin in two different orientations. All three structures contain two copies of LptAm (green and maroon) and thanatin (blue and teal), where the LptAm monomers are domain swapped in the C-terminal region. Thanatin binds to the N-terminal region of LptAm by aligning the β-strands at the interface.

*P. maculiventris* thanatin were also reported to abolish inhibition of *E. coli* growth, highlighting the importance of the C-terminal region [69].

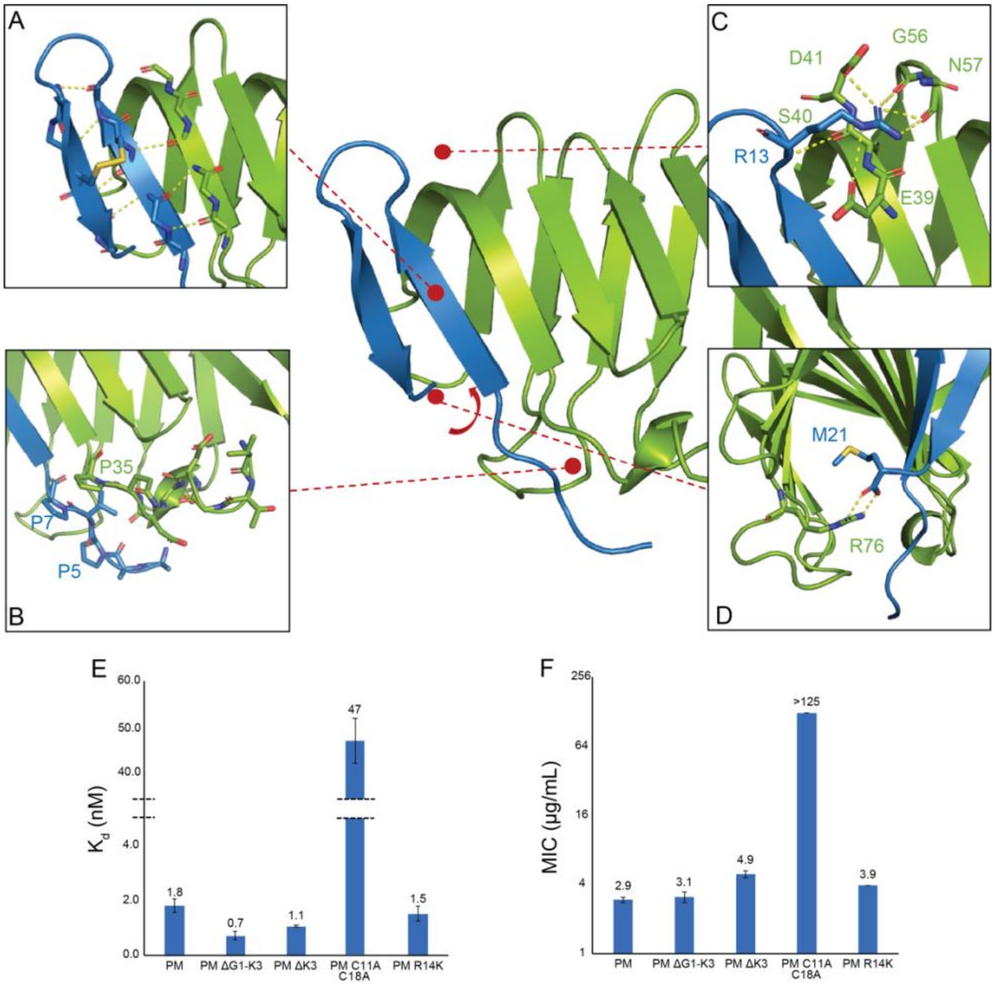
A
C
D41
G56
N57
S40
R13
E39
B
P35
P7
P5
M21
R76
D
E
60.0
50.0
40.0
4.0
2.0
0.0
$K_d$ (nM)
1.8
0.7
1.1
47
1.5
PM
PM ΔG1-K3
PM ΔK3
PM C11A C18A
PM R14K
F
256
64
16
4
1
MIC (μg/mL)
2.9
3.1
4.9
>125
3.9
PM
PM ΔG1-K3
PM ΔK3
PM C11A C18A
PM R14K

**Figure 11: Characteristics of *P. maculiventris* thanatin binding to *E. coli* LptAm.**

**A**, Thanatin binds to the N-terminal β-strand of LptA in a parallel orientation mediated through backbone interactions. Thanatin forms a β-hairpin held together by a disulfide bond and provides a β-strand edge to the complex that disfavors further oligomerization. **B**, Thanatin contains a flexible N-terminal tail not well defined in the electron density map, thus missing the first several residues and side chains. **C**, Thanatin R13 of the cationic loop is extensively involved in intermolecular interactions. **D**, LptA R76 stabilizes the C-terminus of thanatin through bidentate hydrogen bonds. **E**, Probing importance of shared structural regions by binding. Error bars indicate SEM (minimum of at least n=3). **F**, MIC values of thanatin mutants against *E. coli* ATCC 25922. MIC values are calculated as geometric means, and the error bars represent SEM (minimum of at least n=3).

## 2.6 Molecular basis of the enhanced LptAm binding and antibiotic activity by thanatin orthologs from C. ubica and M. histrionica

When comparing the sequences of thanatin from *P. maculiventris*, *C. ubica*, and *M. histrionica*, Y10A, Q19R, Q19T, and M21I highlight the major differences at the interface of thanatin with the hydrophobic core of LptAm **(Figure 12A, B)**. The hydrophobic surface of LptAm consists of I38, L45, V52, and F54, which is the main interface that residues 10 and 21 of thanatin pack against. *P. maculiventris* thanatin contains Y10 and M21, while *C. ubica* thanatin and *M. histrionica* thanatin contain A10 and I21. Residues A10 and I21 provide a better interface, as reflected in tighter binding to LptAm (*C. ubica* $K_d$ of 0.5 ± 0.1 nM and *M. histrionica* $K_d$ of 0.8 ± 0.1 nM vs *P. maculiventris* $K_d$ of 1.8 ± 0.2 nM) **(Figure 6C)**. The tighter binding of *C. ubica* thanatin and *M. histrionica* thanatin to LptAm also contributes to lowering the MIC values **(Figure 6D)**.

To evaluate the contributions of the variant surface residues of thanatin orthologs to LptAm binding, we generated single point mutations of Y10A, Q19R, and M21I and double mutant Y10A/M21I on the *P. maculiventris* backbone. Binding for all mutants to LptAm improved by roughly 3-fold **(Figure 12C)**. However, a gradual improvement of inhibition was observed through MIC values **(Figure 12D)**. Single point mutants Y10A, Q19R, and M21I did not significantly lower the MIC more than 2-fold, whereas a double point mutant Y10A/M21I lowered the MIC 3-fold, suggesting a synergistic relationship between the two positions.

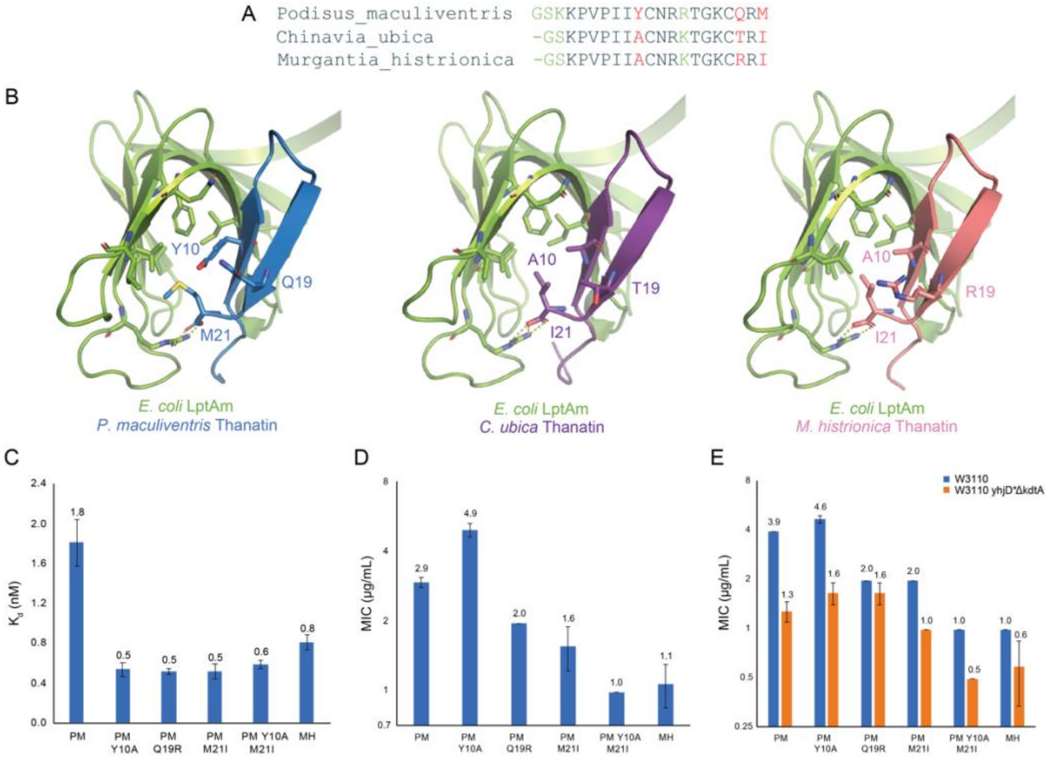

**Figure 12: Differences between *P. maculiventris*, *C. ubica*, and *M. histrionica* thanatin.**

**A**, Sequence alignment of thanatin peptides highlights neutral differences (ΔK3 and R14K) in green and major differences (Y10A, Q19T, Q19R, and M21I) in red. **B**, Major differences are found at the interface of thanatin with the hydrophobic core of LptA. **C**, Assessing effects on binding to *E. coli* LptAm of single- and double-point mutations. Values given are averages of $K_d$ (nM) with SEM as error bars (minimum of at least n=3). D, Assessing effects on inhibition to *E. coli* ATCC 25922, a clinical strain, of single- and double-point mutations. Values given are geometric mean of MIC (µg/mL) with SEM as error bars (minimum of at least n=3). **E**, Assessing effects in membrane permeability to *E. coli* W3110, a laboratory strain, and *E. coli* W3110 *yhjD* ΔkdtA*, a mutation that compromises the outer membrane. Values given are geometric mean of MIC (µg/mL) with SEM as error bars (minimum of at least n=3).

The synergy of Y10A and M21I cannot be attributed to binding due to the single point mutants also improving the $K_d$ to similar levels. To explore if the synergistic effect is due to membrane permeability, we tested the mutants against *E. coli* W3110, the parent laboratory strain, and *E. coli* W3110 *yhjD* ΔkdtA*, a strain with a compromised membrane **(Figure 12E)**. A deletion of kdtA causes disruption of LPS biosynthesis and results in only lipid A being displayed on the outer membrane instead of a complete LPS molecule. All single point mutations do not improve the MIC >2-fold in either the wild-type or membrane compromised strain. However, the double point mutation Y10A/M21I and *M. histrionica* thanatin improves 4-fold over *P. maculiventris* thanatin in the parent strain. When the membrane is compromised, the double point mutation Y10A/M21I and *M. histrionica* thanatin improves roughly 2-fold over *P. maculiventris* thanatin. The decrease in fold-improvement indicates that there is increased membrane permeability with the Y10A/M21I combination, present in both peptides here. The *M. histrionica* thanatin can be represented as the double point mutant Y10A/M21I with the additional Q19R mutation. Despite seeing no membrane permeability problems with the single point mutant Q19R, Q19R alone is not significant enough to further lower the MIC.

## 2.7 Redesigning M. histrionica thanatin

Next, we sought to improve *M. histrionica* thanatin, the most potent thanatin ortholog identified from our studies. We showed that the N-terminal region of *P. maculiventris* thanatin is dispensable **(Figure 11E, F)** and was not always well defined in any of the crystal structures, suggesting that the N-terminal flexibility applied to the other thanatin orthologs as well. We showed that the disulfide bond was necessary for binding and inhibition **(Figure 11E, F)**. We approached the redesign of thanatin in two manners, a stapled version and a cyclized manner. Our approach focused on truncating *M. histrionica* thanatin and removing the need of the disulfide bond, which are prone to environmental oxidizing and reducing conditions.

For the stapled thanatin, we mutated the disulfide bond to C11E and C18K and created a lactam bridge across the peptide (peptide synthesized by Genscript) **(Figure 13A, B)**. This created a staple across thanatin that retained binding to *E. coli* LptAm ($K_d$: 0.7 ± 0.0 nM vs. 0.8 ± 0.1 nM for *M. histrionica* thanatin) **(Figure 13D)**, However, potency against E. coli was 6-fold worse (MIC: 6.6 ± 0.2 μg/mL vs. 1.1 ± 0.2 μg/mL for *M. histrionica* thanatin) **(Figure 13E)**. This discrepancy in results could be due to how thanatin interacts with the cell membrane. The effect of the staple on membrane permeability would be interesting to explore to further understand how thanatin is transported across the membrane.

For the cyclized thanatin, we mutated the disulfide bond to C11A and C18A and created a lactam bridge across mutations P7D and R20K (peptide synthesized by Genscript) **(Figure 13A).** This created a bridge across thanatin to mimic a head-to-tail cyclization of the peptide **(Figure 13C)**. Unfortunately, binding to *E. coli* LptAm was drastically decreased ($K_d$: 986 ± 154 nM vs. 0.8 ± 0.1 nM for *M. histrionica* thanatin) and potency was abolished (MIC: >62.5 μg/mL) **(Figure 13D, E)**. These weakened qualities

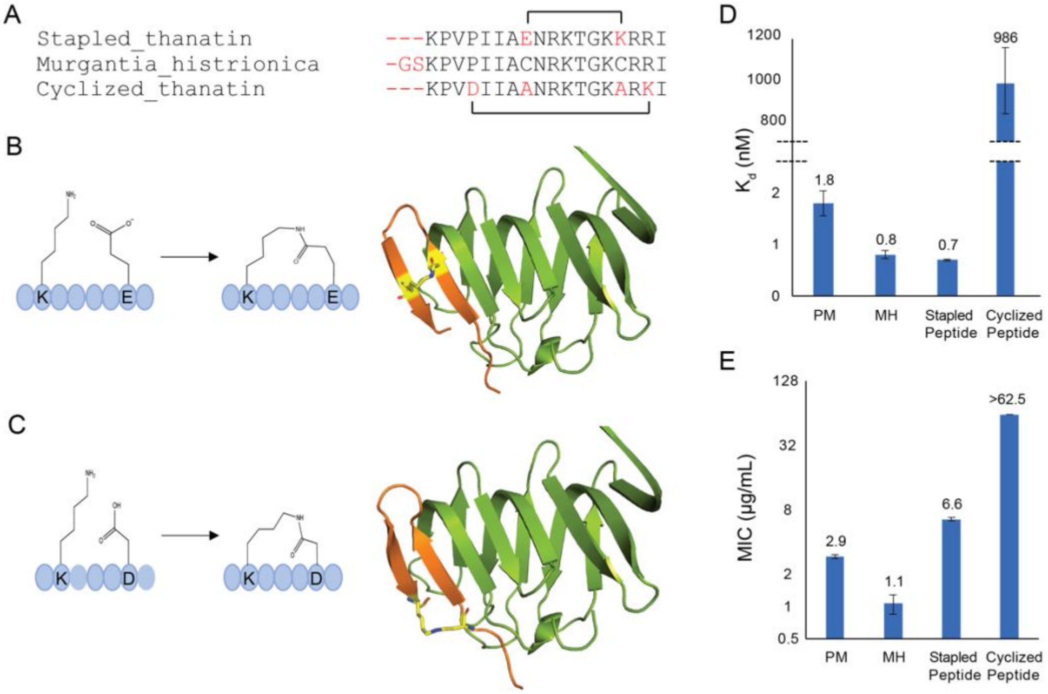

**Figure 13: Redesigning *M. histrionica* thanatin into a novel scaffold.**
**A**, Novel thanatin scaffolds were derived from the *M. histrionica* thanatin sequence. **B**, The stapled thanatin was truncated on the N-terminus and the disulfide bond was replaced with a lactam bridge. Modeling predicts stapled thanatin can retain its structure with the modifications. **C**, The cyclized thanatin was truncated on the N-terminus, the disulfide bond was mutated to alanine residues, and a lactam bridge was used to link thanatin in a head-to-tail manner. Modeling predicts cyclized thanatin can retain its structure with the modifications. **D**, Assessment of modified thanatin peptides binding to *E. coli* LptAm (minimum of at least n=3). **E**, Assessment of modified thanatin peptides potency against *E. coli* ATCC 25922. Values given are geometric mean of MIC (μg/mL) with SEM as error bars (minimum of at least n=3).

of thanatin could be due to a larger rearrangement of thanatin that our modeling did not predict.

Although we failed in improving thanatin via cyclization, we show promising results with a stapled thanatin scaffold. The stapled peptide retained binding with a slight reduction in potency. Structural studies will be needed to confirm the binding mechanisms of this thanatin derivative in order to further improve upon. Overall, we present a novel thanatin scaffold to use as a platform to create better antibiotics.

## 2.8 Discussion

Here, we showed that mining genomic databases for thanatin orthologs yielded more effective thanatin peptides that potently target *E. coli*. We developed a BLI assay to measure binding to *E. coli* LptA and a MIC assay to measure inhibition of *E. coli* growth and assess membrane permeability. We crystallized and determined the structures of *P. maculiventris* thanatin and two of the best thanatin orthologs, from *C. ubica* and *M. histrionica*, bound to *E. coli* LptA for structural characterization. We conducted mutagenesis studies to better understand the mechanism of improved binding and inhibition. Overall, Y10A, Q19R, and M21I increased binding affinity, however the improvement in antibiotic activity is additionally driven by the increased membrane permeability of Y10A and M21I. Furthermore, we redesigned a scaffold from *M. histrionica* thanatin, the most potent thanatin ortholog, to shorten the peptide and remove the need for a disulfide bond.

The discovery of novel thanatin orthologs adds to the recent and growing field of microproteomics and cryptic peptides. Recent efforts have been made to mine the human genome for microproteins (around 100 amino acids or less) that have been largely ignored until recently. Microproteins have been found to play various roles in regulating stress signaling (PIGBOS), muscle and fat metabolism (MOTS-c), and mitochondrial activity (SHMOOSE) [100]–[102]. They have also been found to be linked to various diseases such as cancer, diabetes, and Alzheimer's disease [100]–[102]. In addition, recent efforts have been made to mine the human genome for cryptic peptides that serve as antibiotics and better identify them [14], [17]. Cryptic peptides are active peptides created from the degradation of or cleavage from a larger protein. *P. maculiventris* thanatin is expressed within a precursor of 109 residues (Accession number ATG84180.1). After cleavage from the precursor, *P. maculiventris* thanatin becomes an active antimicrobial peptide of 21 residues. Thanatin sequences found in our genomic search were mostly from whole genome sequences and their corresponding precursors were not identified. However, the precursor for *R. pedestris* thanatin has been reported, suggesting similar processing and maturation of the peptides (Accession number BAN67668). Our studies show that mining non-human genomes, such as insects in our case, could identify a valuable library of unexplored microproteins with antimicrobial activity.

Variance in the thanatin peptides can reveal what role these peptides play in their host by elucidating their microbial targets. The search for thanatin orthologs yielded 16 thanatin-like sequences, which can be grouped into 11 unique sequences, from 13 species. The majority of these orthologs are from stink bugs and other insects. It has been reported that thanatin in *P. maculiventris* (spined soldier bug) and *R. pedestris* (bean bug) are found in the insect's midgut region and help regulate the microbiome to promote symbiosis [69], [72]. It is plausible that the newly reported thanatins in other insects play a similar role in their respective hosts. Different thanatin sequences could target different pathogens to various degrees, which may reflect the microbiome population of each host and helps build the broadly inhibitory profile of thanatin. Interestingly, one thanatin ortholog was found in *Shewanella*, an anaerobic Gram-negative bacteria found in marine sponges. It is plausible that *Shewanella* produces thanatin as an anti-toxin and secretes it into the environment as a defense mechanism against competing bacteria or fungi. Further studies will be needed to confirm the role of thanatin in these other species. Overall, as more genomic data is deposited, additional thanatin orthologs will be identified and allow us to update the consensus sequence.

Our early design of a stapled thanatin retains binding but slightly reduces potency. It would be worthwhile to apply computational tools, such as OSPREY, to our stapled thanatin to increase binding affinity and regain potency [103]. Another design approach would be to create hybrid peptides to combine our improved thanatins with other antimicrobial peptides to synergistically attack pathogens. Due to the effective membrane permeability and short cationic sequence of thanatin, thanatin could be reimagined to be like a cell-penetrating peptide to delivery molecules to targets in the periplasm. Examples of this have already been reported in literature, though the mechanisms of action have not been explored [97], [98].

Overall, our findings show that nature and evolution are valuable sources for antimicrobial peptides. We identified a more potent thanatin peptide from *M. histrionica* that targets *E. coli* through strong binding to LptA to disrupt the the lipopolysaccharide transport system. High-resolution crystal structures allowed us to better understand the mechanism of action between thanatin and LptA. We identified that the key to improved potency lies in the coordination between thanatin A10 and I21 to increase membrane permeability. We also redesigned *M. histrionica* thanatin to remove the need for a disulfide bond and present a scaffold for the next round of peptide design. Future work includes utilizing computational protein design to improve the affinity and potency of our redesigned *M. histrionica* thantin scaffold and to apply our thanatin library to other Gram-negative bacteria.

## 2.9 Materials and methods

### 2.9.1 Sequence analysis

Identification of novel thanatin orthologs was performed with the tblastn and blastp suites (https://blast.ncbi.nlm.nih.gov) [104]. The amino acid sequence from *P. maculiventris* thanatin was used as the query sequence. Sequence identities and positive substitutions were used to determine sequences of interest. A multiple sequence alignment was generated with Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) [105]. A consensus sequence was generated with WebLogo (https://weblogo.berkeley.edu) [106]. The accession numbers corresponding to all thanatin orthologs mentioned in this publication are: *Aelia acuminata* (CAJVQN010000565), *Anasa tristis* (GEII01018028), *Chinavia impicticornis* (GIVF01040023), *Chinavia ubica* (GBFA01023333), *Dichelops melacanthus* (GBES01007078), *Halyomorpha halys* (XP_014293651), *Murgantia histrionica* (GECQ01170005), *Nezara viridula* (GIBW01225186 and GGPJ01242699), *Piezodorus guildinii* (CM041296), *Podisus maculiventris* (AAB36066), *Riptortus pedestris* (BAN67668 and JAAEAN010010806), *Shewanella* (WP_238898582), and *Stiretrus anchorago* (WUAS01076668).

### 2.9.2 Cloning, expression, and purification of *E. coli* LptA

A monomeric mutant (Δ159-185) of LptA [49] without the signal peptide sequence (Δ1-27) was inserted into a modified pET24b vector through Takara Bio In-Fusion cloning to yield the final construct of His$_6$-SUMO-LptA. For BLI binding studies, a C-terminal flexible linker and AviTag were incorporated to yield the final construct of His$_6$-SUMO-LptA-GSGGSGSG-AviTag. After confirmation through DNA sequencing, the plasmid was transformed into BL21 (DE3) competent cells on LB plates containing kanamycin. Cells were grown in Luria-Bertani (LB) media with kanamycin at 37 °C until OD600 ≈0.6 with shaking and then induced with 1 mM IPTG at 18 °C for 18 hours. The cells were harvested by centrifugation and frozen at -80 °C until purification.

The cell pellet was thawed, resuspended in Buffer A (20 mM Tris pH 8.0, 250 mM NaCl) with a protease inhibitor tablet, and lysed with a microfluidizer. The lysate was centrifuged at 20,000 g for 20 mins to remove cellular debris. The supernatant was passed through a Ni-NTA column, washed with Buffer A + 25 mM imidazole, eluted with Buffer A + 250 mM imidazole, and further eluted with Buffer A + 500 mM imidazole. Elution fractions containing protein were pooled and digested with SENP1 protease (1:100 M protease:protein) while being dialyzed (MWCO 3.5 kDa) against 20 mM Tris pH 8.0, 150 mM NaCl, 1 mM DTT for 20 hours at 4 °C. The dialyzed and digested solution was passed through a Ni-NTA column, washed with Buffer A + 25 mM imidazole, eluted with Buffer A + 250 mM imidazole, and further eluted with Buffer A + 500 mM imidazole. The flow through and wash fractions containing protein were concentrated (MWCO 3K) and passed through a Superdex200 column in Buffer B (50 mM Tris pH 8.5, 200 mM NaCl). The fractions containing purified LptA were concentrated (MWCO 3 kDa) and flash frozen in liquid nitrogen for storage at -80 °C. For BLI binding studies, the protein was biotinylated with the BirA biotin-protein ligase standard reaction kit from Avidity and passed through a S75 column in Buffer C (1X PBS) before being flash frozen for storage at -80 °C.

### 2.9.3 Cloning, expression, and purification of thanatin peptides

The thanatin sequence was inserted into a pET15b vector through Takara Bio In-Fusion cloning to yield the final construct of $\text{His}_{10}$-GB1-TEV Cleavage Site-Thanatin. Single point and double point mutations were made with a Q5 Site-Directed Mutagenesis Kit from New England BioLabs. After confirmation through DNA sequencing, the plasmid was transformed into BL21 (DE3) competent cells on LB plates containing ampicillin. Cells were grown in LB media with ampicillin at 37 °C until OD600 ≈0.6 with shaking and then induced with 1 mM IPTG at 18 °C for 18 hours. The cells were harvested by centrifugation and frozen at -80 °C until purification.

The cell pellet was thawed, resuspended in Buffer D (50 mM Tris pH 8.0, 200 mM NaCl) with an added protease inhibitor tablet, and lysed with a microfluidizer. The lysate was centrifuged at 20,000 g for 20 mins to remove cellular debris. The supernatant was passed through a Ni-NTA column, washed with Buffer D + 25 mM imidazole, eluted with Buffer D + 250 mM imidazole, and further eluted with Buffer D + 500 mM imidazole. Elution fractions containing protein were pooled, concentrated (MWCO 3 kDa), and passed through a Superdex200 column in Buffer B. The fractions containing protein were pooled and digested with TEV protease (1:100 M protease:protein) for 20 hours at 4 °C. The digested solution was passed through a Ni-NTA column, washed with Buffer B + 25 mM imidazole, eluted with Buffer B + 250 mM imidazole, and further eluted with Buffer B + 500 mM imidazole. The flow through and wash fractions containing purified peptide were concentrated (MWCO 3K) and frozen for storage at -20 °C.

Synthesis of the stapled and cyclized thanatin peptide were carried out by Genscript.

### 2.9.4 Bio-layer interferometry binding assay

A BLI binding assay was developed on an Octet Red96e System. All assays were performed at 25 °C, at 700 r.p.m., and in buffer (1X Octet Kinetics Buffer: 1X PBS, 1% BSA, 0.1% Tween-20, 0.1% Kathon). Octet Streptavidin (SA) Biosensors were pre-wet for 15 minutes in buffer. The order of steps for each assay were as follows: (1) baseline: immersion in buffer for 300 seconds, (2) loading: immersion in biotinylated *E. coli* LptA-Avi at 1 μg/mL for 300 seconds, (3) baseline: immersion in buffer for 300 seconds, (4) association: immersion in peptide for 3600 seconds. A 15-point titration of peptide from 153.1 nM to 0.6 nM was used. Controls for double-referencing each point were the following: (1) reference sample well: SA biosensor with LptA-Avi loaded immersed into buffer, (2) reference sensor: SA biosensor with no LptA-Avi loaded immersed into various peptide concentrations. Curve fitting and data analysis for reporting dissociation constants ($K_d$) were performed using the following steady-state 1:1 binding model: $Response = \frac{[Thanatin]/K_d}{1+[Thanatin]/K_d} Rmax$. The SA Biosensors loaded with LptA-Avi were regenerated for multiple uses as follows: (1) baseline: immersion in buffer for 100 seconds, (2) regeneration: immersion in 10 mM glycine pH 1.7 for 10 seconds, (3) neutralization: immersion in buffer for 10 seconds, repeat steps (2) and (3) for a total of 3 times, (4) baseline: immersion in buffer for 100 seconds.

### 2.9.5 Minimal inhibition concentration assay

A MIC assay was developed with *E. coli* ATCC 25922, *E. coli* W3110, and *E. coli* W3110 *yhjD* ΔkdtA* based on the Clinical & Laboratory Standards Institute (CLSI) standard protocol [107]. *E. coli* was streaked onto an antibiotic free LB plate from a cryo stock and incubated at 37 °C overnight. A single colony from this plate was streaked onto an antibiotic free LB plate and incubated at 37 °C overnight. For *E. coli* ATCC 25922 and W3110, a single colony was added to LB media (10 mL) and grown at 37 °C to OD600 ≈0.1 (corresponding to ≈1 – 2 x $10^8$ CFU/mL) with shaking. Cells were plated at a final concentration of ≈5 x $10^5$ CFU/mL in Cation-Adjusted Mueller-Hinton Broth (CAMHB) over a titration of peptide inhibitor in clear 96 well plates. Due to difficulties growing in CAMHB, *E. coli* W3110 *yhjD* ΔkdtA* was grown in LB at 37 °C to OD600 ≈0.6 and plated at a final dilution of 1:100 in LB. The plate was incubated at 37 °C for 20 hours. MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide) (10 µL of 5 mg/mL) was added to each well and incubated at 37 °C for 3 hours. Isopropanol + 0.1 M HCl (100 µL) were added to each well. The plate was imaged and read at OD690 and OD5750 for analysis. The MIC was determined to be the minimum peptide concentration where no growth was observed.

### 2.9.6 Crystallization and structure determination of *E. coli* LptA with *P. maculiventris* thanatin

Purified *E. coli* LptA and *P. maculiventris* thanatin were incubated together and co-eluted through a Superdex200 column in Buffer B. The fractions containing the complex were concentrated (MWCO 3 kDa) to 5 and 10 mg/mL. Crystal trays were set up at each concentration using the sitting drop method with 1 μL protein and 1 uL of crystallization solution, along with seeding from *E. coli* LptA with *P. maculiventris* M21W crystals using the scratch method. Crystallization solutions that yielded crystals ranged from 60 – 70% MPD at 0.1 M MES pH 5.8. The crystallization condition of the reported dataset was 10 mg/mL LptA:thanatin in 0.1 M MES pH 5.8, 60% MPD.

Diffraction data were collected at NE-CAT 24-ID-C beamline. The 3D model was constructed using molecular replacement from the PHASER module in PHENIX with a structure of *E. coli* LptA and *P. maculiventris* thanatin M21W as the input model. Coot and PHENIX were used for iterative model building and refinement.

### 2.9.7 Crystallization and structure determination of *E. coli* LptA with *C. ubica* thanatin

Purified *E. coli* LptA and *C. ubica* thanatin were incubated together and co-eluted through a Superdex200 column in Buffer B. The fractions containing the complex were concentrated (MWCO 3 kDa) to 6, 8, and 15 mg/mL. Crystal trays were set up at each concentration using the sitting drop method with 1 μL protein and 1 uL of crystallization solution. Crystallization solutions that yielded crystals ranged from 60 – 77.5% MPD and pH 5.8 – 8.5 (0.1 M MES, HEPES, or Tris). The crystallization condition of the reported dataset was 6 mg/mL LptA:thanatin in 0.1 M MES pH 5.8, 70% MPD.

Diffraction data were collected at NE-CAT 24-ID-E beamline. The 3D model was constructed using molecular replacement from the PHASER module in PHENIX with an

AlphaFold model of *E. coli* LptA and *P. maculiventris* thanatin as the input model. Coot and PHENIX were used for iterative model building and refinement.

### 2.9.8 Crystallization and structure determination of *E. coli* LptA with *M. histrionica* thanatin

This section was done in collaboration with Amanuel Kibrom of the Zhou Laboratory. Purified *E. coli* LptA and *M. histrionica* thanatin were incubated together and co-eluted through a Superdex200 column in Buffer B. The fractions containing the complex were concentrated (MWCO 3 kDa) to 10.5 mg/mL. Crystal trays were set up at each concentration using the sitting drop method with 1 μL protein and 1 uL of crystallization solution. Crystallization solutions that yielded crystals ranged from 0.095 - 0.1M sodium citrate pH 5.6, 19 - 20% isopropanol, 19 - 20% PEG 4000, and 0 - 5% glycerol. The crystallization condition of the reported dataset was 10.5 mg/mL LptA:thanatin in 0.095 M sodium citrate pH 5.6, 19% isopropanol, 19% PEG 4000, and 5% glycerol.

Diffraction data were collected at NE-CAT 24-ID beamline. The 3D model was constructed using molecular replacement from the PHASER module in PHENIX with an AlphaFold model of *E. coli* LptA and *Murgantia histrionica* as the input model. Coot and PHENIX were used for iterative model building and refinement.

# 3. Characterizing thanatin against *Pseudomonas aeruginosa* LptH

## *3.1 Background*

*Podisus maculiventris* thanatin, a 21-residue β-hairpin held together by a disulfide bond, was reported as a broadly inhibitory antimicrobial peptide that could inhibit Gram-positive bacteria, Gram-negative bacteria, and fungi [69]. The mechanism of action of thanatin in Gram-positive bacteria and fungi has yet to be elucidated. However, the mechanism in Gram-negative bacteria has been well characterized in *Escherichia coli*, a microbe strongly inhibited by thanatin. Thanatin has been shown to target the Lipopolysaccharide transport (Lpt) system by disrupting complex formation [60], [61]. The Lpt system is composed of seven essential proteins, LptABCDEFG, that assemble into a complex spanning from the inner membrane to the outer membrane. The Lpt system is responsible for transporting lipopolysaccharide (LPS) to the outer membrane for display where a dense layer offers protection from antibiotics and evasion from host detection. Thanatin has been shown to bind to LptA, LptD, and lipopolysaccharides to disrupt the Lpt pathway [60], [61], [83], [93]. Thanatin has been well characterized with LptA, where it was shown to inhibit self-oligomerization by binding to the N-terminus of LptA and preventing β-sheet elongation [61].

*Pseudomonas aeruginosa*, a Gram-negative bacterium also containing the Lpt system, was reported as one of thanatin's targets. However, the potency drops roughly between 20- and 100-fold when compared to *E. coli* [69]. LptH, the *P. aeruginosa* ortholog of *E. coli* LptA, has been shown to be essential for the viability and infectivity of *P. aeruginosa* [108]. A crystal structure of apo LptH (PDB ID: 4UU4) has been reported which shows the same β-jellyroll fold as that of apo LptA (PDB ID: 2R19) and is also predicted to self-oligomerize in an N-to-C polarity [109], [110]. The two apo structures are similar with a backbone RMSD of 4.19 Å, a 45.9% sequence similarity, and 27.1% sequence identity. When disregarding the C-terminal end of LptH and LptA, where LptH contains a longer loop region, the backbone RMSD improves to 1.92 Å. This highlights the high tertiary similarities in the N-terminal end of LptH and LptA. Hence, it is plausible that thanatin also targets LptH in *P. aeruginosa* through a similar mechanism of action. Here, we confirm that thanatin binds to LptH. We identify the resistance factor Y51 in LptH and show that thanatin position 10 is vital for binding. We also provide insight on the role of C-amidation improving the potency of thanatin against *P. aeruginosa*. Designs made to improve thanatin towards *P. aeruginosa* based on modeling were not successful, therefore highlighting the need for a high-resolution structure of thanatin in complex with LptH to better understand the mechanism of binding.

## 3.2 Thanatin targets LptH in P. aeruginosa

We confirmed that *P. maculiventris* thanatin does indeed bind LptH. To predict where thanatin binds, we used AlphaFold2-multimer to model thanatin in complex with LptH [111], [112]. AlphaFold2-multimer predicted thanatin to bind to the N-terminus of

LptH in a similar manner as reported with *E. coli* LptA **(Figure 14A)**. Thanatin is predicted to append to the N-terminal β-strand of LptH to inhibit self-oligomerization. Crystallization studies of this complex were attempted, however no protein crystals were obtained from multiple rounds of general screening. Structure determination to confirm the binding site has yet to be elucidated. To measure the interaction, we developed a bio-layer interferometry (BLI) binding assay. Biotinylated monomeric LptH, LptHm, is immobilized on BLI streptavidin biosensors. The Avi-tag, the site-specific sequence for biotinylation, was placed on the C-terminus to orient the protein with the N-terminus available for binding. To prevent LptH from self-oligomerizing in solution, we modeled a monomeric LptH (LptHm) after a reported monomeric *E. coli* LptA (LptAm) construct and deleted the terminal β-strand of LptH [49]. The biosensors containing immobilized LptHm were then immersed into a titration of thanatin to record association for 1 hour **(Figure 14B)**. The steady-state data was fit to a 1:1 binding model, which yielded a $K_d$ of 12.9 ± 1.1 nM **(Figure 14C)** for *P. maculiventris* thanatin. We assessed the role of the disulfide bond by generating a double mutant C11A/C18A. Removal of the disulfide bond abolished binding, indicating the necessity of it ($K_d$: >4398 ± 0 nM) **(Figure 14D, E)**.

We developed a minimum inhibition concentration (MIC) assay and confirmed that *P. maculiventris* thanatin can inhibit *P. aeruginosa* growth, however we observed some discrepancy from literature. Literature reported MIC values of 12.5 - 100 μg/mL

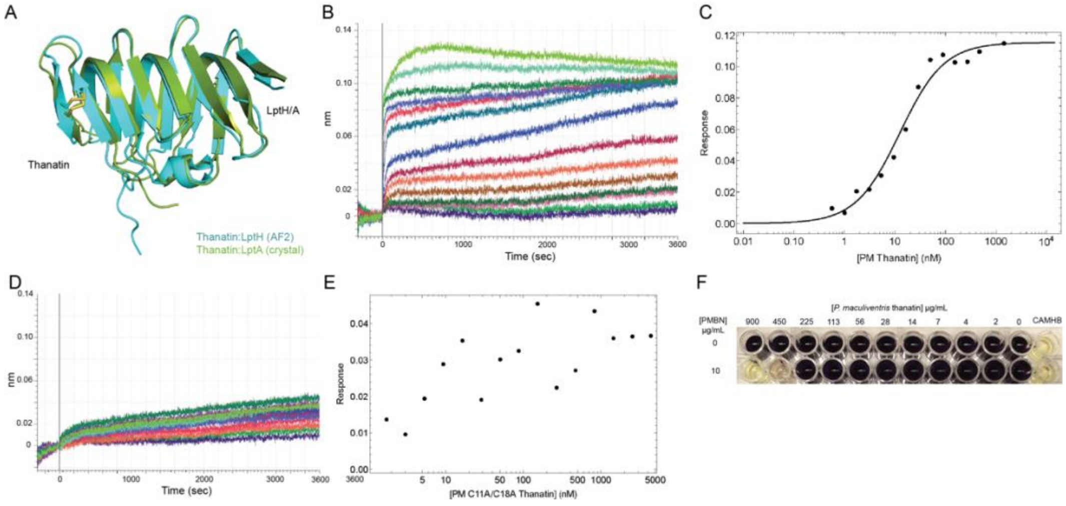

**Figure 14: *P. maculiventris* thanatin targets *P. aeruginosa* LptH.**
**A**, AlphaFold2-multimer predicts thanatin to bind to the N-terminal end of LptH in a similar manner as seen with the crystal structure of thanatin with *E. coli* LptA. Due to domain-swapping and dimerization observed in the crystal structure, alignment was truncated on the C-terminal end. **B**, Representative BLI trace of a thanatin titration over LptH showing binding **C**, Representative steady state binding curve fit to a 1:1 binding model for thanatin binding to LptH. **D**, Representative BLI trace of thanatin C11A/C18A titration over LptH showing loss of binding without a disulfide bond. **E**, Representative steady state binding curve of thanatin C11A/C18A indicating no binding. **F**, Thanatin inhibits *P. aeruginosa* RP73 with the aid of PMBN.

With clinical strain ATCC 9027, 2.5 – 16 μg/mL with clinical strain ATCC 27853, and 1 – 16 μg/mL with clinical strain XJ75315 [69], [81], [83], [91]. We did not observe inhibition of *P. aeruginosa* with up to 1000 μg/mL with ATCC 27853. We added polymyxin B nonapeptide (PMBN), a cationic peptide that induces increased outer membrane permeability, to sensitize the strain [113]. However, we observed undesired toxicity effects. We were able to develop MIC assay conditions for *P. aeruginosa* clinical strain RP73 with the aid of 10 μg/mL PBMN **(Figure 14F)**. No toxicity effects were observed at this concentration of PMBN. The MIC value of *P. maculiventris* thanatin under these conditions was 450 µg/mL.

## 3.3 Identifying the resistance factor of thanatin in P. aeruginosa LptH

We identified several residues of interest in LptH that could account for the difference in binding when compared to *E. coli* LptA. We performed a sequence alignment of LptH with LptA to identify differences in the binding interface **(Figure 15A)**. The majority of residues identified to be important for thanatin binding to LptA are not conserved. The hydrophobic core of LptA that interacts with thanatin is defined by I38, L45, V52, F54, and V74. These residues correspond to V35, L42, A49, Y51, L71 in LptH, respectively. LptA residues identified to be involved in intermolecular interactions wih thanatin are E39, D41, Q43, G56, N57, and R76. These residues correspond to Q36, D38, A39, G53, D54, and Q73 respectively. To predict what structural effect these changes would have on thanatin binding, we overlayed apo LptH (PDB ID: 4UU4) with our structure of LptA in complex with *P. maculiventris* thanatin **(Figure 15B, C, D, E)**. Thanatin R13 is involved in extensive intermolecular contacts with LptA at E39, D41, and N57 **(Figure 15B)**. Corresponding mutations to reflect LptH would be E39Q and N57D. We predict that the intermolecular interactions involving thanatin R13 would be preserved due to E39Q both containing a carbonyl in their side chain, D41 is conserved, and N57D side chains are not involved. Thanatin N12 is interacting with LptA Q43 **(Figure 15C)**. An LptA Q43A mutation to reflect LptH would result in the loss

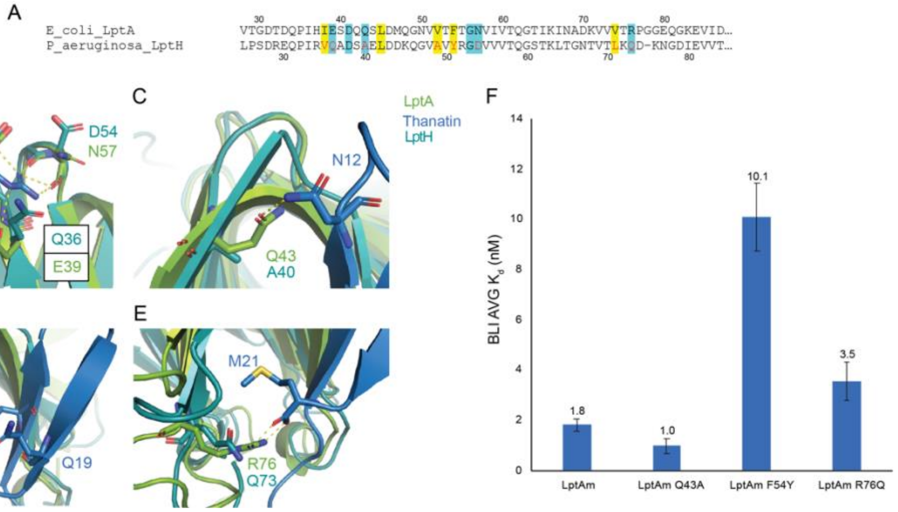
A
E_coli_LptA
P_aeruginosa_LptH
VTGDTDQPIHIESDQQSLDMQGNVVTFTGNVIVTQGTIKINADKVVVTRPGGEQGKEVID…
LPSDREQPIRVQADSAELDDKQGVAVYRGDVVVTQGSTKLTGNTVTLKQD-KNGDIEVVT…
30
40
50
60
70
80
B
D38
D41
D54
N57
R13
Q36
E39
C
N12
Q43
A40
LptA
Thanatin
LptH
D
F54
Y51
Y10
Q19
E
M21
R76
Q73
F
BLI AVG K_d (nM)
14
12
10
8
6
4
2
0
1.8
1.0
10.1
3.5
LptAm
LptAm Q43A
LptAm F54Y
LptAm R76Q

**Figure 15: Identifying LptH Y51 as resistance factor through LptA model system.**

**A**, Partial sequence alignment of the N-terminal end of LptA and LptH (top numbers LptA, bottom numbers LptH) with residues involved with thanatin binding highlighted. Yellow residues are involved with the hydrophobic pocket of LptA. Blue residues are involved in intermolecular interactions involving side-chains. Red residues are differences between LptA and LptH. **B**, Thanatin R13 interactions with LptA E39, D41, and N57 overlayed with apo LptH. **C**, Thanatin N12 interaction with LptA Q43 overlayed with apo LptH. **D**, Thanatin Y10 interaction with LptA F54 with apo LptH overlayed. **E**, Thanatin C-terminus interaction with LptA R76 with apo LptH overlayed. **F**, BLI binding $K_d$ values of *P. maculiventris* thanatin with LptAm and mutants corresponding to LptH interface. Values are averages with error bars representing SEM.

of interaction, perhaps weakening binding. Thanatin Y10 and M21 interact with the hydrophobic core of LptA, represented by I38, L45, V52, F54, and V74 **(Figure 15D)**. The corresponding mutations to reflect LptH, I38V, V52A, and V74L, are not predicted to affect the binding pocket and L45 is conserved. However, a F54Y mutation is predicted to clash with thanatin Y10 due to the addition of a hydroxyl group in the binding pocket. The C-terminus of thanatin interacts with LptA R76 **(Figure 15E)**. An LptA R76Q mutation to reflect LptH would result in the loss of a charged interaction, perhaps weakening binding. Overall, we predicted that LptH A40, Y51, and Q73 are residues of interest due to loss of interaction with thanatin as seen in LptA.

To test our predictions, we generated the following mutations, Q43A, F54Y, and R76Q, in LptAm and assessed binding with thanatin via a BLI binding assay **(Figure 15F)**. Due to having a high-resolution crystal structure of thanatin in complex with LptAm, but none with LptH, we used the LptAm system as a model for initial proof of concept. We reported a $K_d$ of 1.8 ± 0.2 nM for thanatin with LptAm. A LptAm Q43A mutation (1.0 ± 0.3 nM) and R76Q mutation (3.5 ± 0.8 nM) did not alter the $K_d$ >2-fold. However, A LptAm F54Y mutation caused a significant increase in $K_d$ to 10.1 ± 1.4 nM, aligning with our prediction that the addition of a hydroxyl group to the binding interface would be unfavorable.

To confirm if LptH Y51 is the resistance factor for thanatin, we made a LptHm Y51F mutant to assess if binding of thanatin improved via BLI. Unexpectedly, we did not observe any binding between thanatin and LptHm Y51F **(Figure 16A, B)**. We developed an orthogonal binding assay via isothermal calorimetry titration (ITC) to confirm this unexpected observation. We titrated thanatin into LptHm and LptHm Y51F to measure the change of heat and fit for the $K_d$ values. We were able to rescue binding of thanatin between LptHm ($K_d$: 149 ± 31 nM) and LptHm Y51F ($K_d$: 23 nM) **(Figure 16C,**

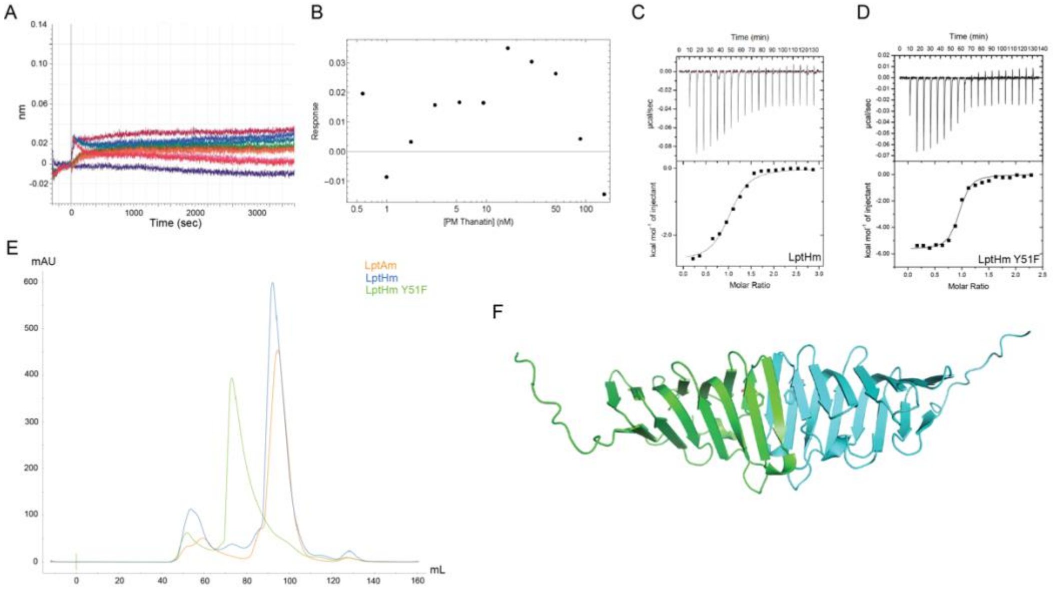

**Figure 16: LptH Y51F induces higher-ordered self-oligomerization.**
**A**, *P. maculiventris* thanatin does not bind to LptHm Y51F shown with BLI binding traces and **B**, data fit. **C**, *P. maculiventris* thanatin isotherm and binding curve with LptHm ($K_d$: 149 nM) and **D**, *P. maculiventris* thanatin isotherm and binding curve with LptHm Y51F ($K_d$: 23 nM) show recovery of binding via ITC. **E**, Size exclusion chromatography of LptAm (monomeric reference), LptHm, and LptHm Y51F on a superdex200 column show a shift of LptHm Y51F to a higher molecular weight. **F**, AlphaFold2-multimer modeling predicts that LptHm Y51F would form a N-to-N dimer. Two LptHm Y51F monomers are presented in cyan and green, with Y51F shown in the interface.

**D)**, hence confirming our prediction that LptH Y51 is a major resistance factor of thanatin. This discrepancy prompted us to inspect the construct more closely and we observed that a Y51F mutation shifts the population to a higher order oligomerization state **(Figure 16E)**. Wild-type LptHm was observed to be predominately monomeric, based on using monomeric LptAm as a reference in size exclusion chromatography, however a small peak indicating higher order oligomerization is present. We reason that the higher order oligomerization state is blocking the N-terminal edge of LptHm Y51F which would prevent thanatin from binding. To assess if this could be plausible, we used AlphaFold2-multimer to model a dimer of LptHm Y51F **(Figure 16F)**. AlphaFold2-multimer predicted N-to-N dimerization of LptHm Y51F that occluded the N-terminal interface. This oligomerization state could be concentration dependent, which could explain the discrepancy in the BLI and ITC binding data. The ITC experiment utilizes 2.5 μM LptH Y51F in the sample cell, however the BLI experiment immobilizes LptHm Y51F to the biosensor surface and creates a high local concentration. The binding affinity of thanatin to LptHm Y51F is not strong enough to disrupt the LptHm Y51F dimer, hence no binding is observed via BLI. Structure determination will be necessary to confirm and visualize these interactions. However, we identified a resistance factor in LptH that weakens binding to thanatin when compared to LptA by using the LptA model for analysis of the binding interface.

## 3.4 Improving thanatin to overcome LptH Y51 resistance factor

We attempted to mutate *P. maculiventris* thanatin to enhance binding to LptH. Using our overlay of apo LptH with LptAm in complex with thanatin, we predicted that LptH Y51 sterically clashes with thanatin Y10 **(Figure 17A)**. To reduce this clash, we generated thanatin Y10A and assessed binding with our LptAm F54Y model system **(Figure 17B)**. Thanatin Y10A improved binding to LptAm F54Y ($K_d$: 5.6 ± 0.4 nM) compared to wild type thanatin ($K_d$: 10.1 ± 1.4 nM), indicating that it would be beneficial to create more space in the binding pocket **(Figure 17D)**. To improve the interaction in

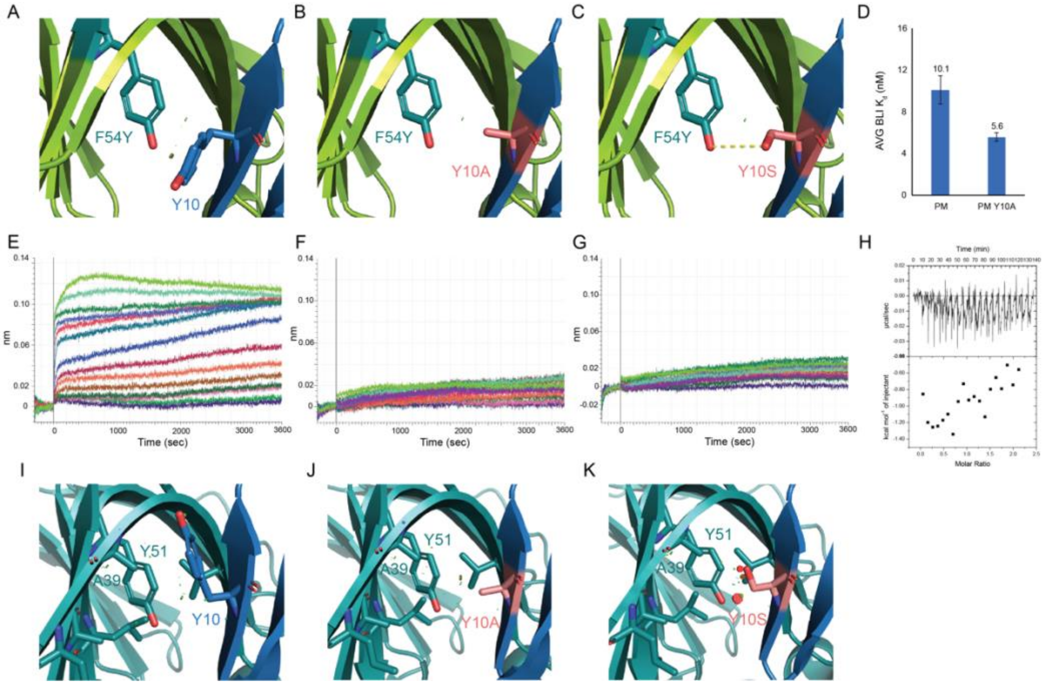

**Figure 17: Thanatin Y10 is essential for binding LptH.**

**A**, Homology modeling using *E. coli* LptAm in complex with *P. maculiventris* thanatin with a LptAm F54Y mutation to mimick LptH. Disks represent good (green) and bad (red) Vander Waals interactions. **B**, Homology modeling of thanatin Y10A with LptAm F54Y and **C**, thanatin Y10S with LptA F54Y to overcome resistance factor LptH Y51. **D**, Average BLI $K_d$ values for thanatin and thanatin Y10A with LptAm F54Y show improved binding. **E**, BLI trace for thanatin with LptHm as reference. BLI traces for **F**, Thanatin Y10A with LptHm and **G**, thanatin Y10S with LptHm show no binding. **H**, ITC isotherm and binding curve for Y10A with LptHm confirms no binding. **I**, AlphaFold2-multimer model of thanatin with LptH predicts Y10 to adopt a different rotamer. **J**, AlphaFold2-multimer model of thanatin Y10A with LptH and **K**, thanatin Y10S with LptH still predict complex formation. Y10A appears to be favorable with minimal clashes, however Y10S appears to be unfavorable with severe clashes with F54Y.

that region, we modeled thanatin Y10S to create a favorable interaction between the hydroxyl groups of LptAm F54Y and thanatin Y10S **(Figure 17C)**.

To confirm these findings, we applied thanatin Y10A and thanatin Y10S to LptHm in our BLI binding assay. Unexpectedly, we did not observe any binding of thanatin Y10A or Y10S with LptHm **(Figure 17F, G)**. We assessed thanatin Y10A with LptAm in our orthogonal ITC assay and confirmed the results of no binding **(Figure 17H)**. These results indicate to us that thanatin Y10 is essential to binding in a way that the LptA model does not predict. To further assess this observation, we generated AlphaFold2-multimer models of wild type thanatin, thanatin Y10A, and thanatin Y10S with LptH **(Figure 17I, J, K)**. AlphaFold2-multimer predicted all three peptides to bind at the N-terminus of LptH. Thanatin Y10 is predicted to adopt a rotamer pointing towards the turn in the LptA β-jellyroll fold, opposite of the direction as that observed in thanatin in complex with LptA **(Figure 17I)**. This explains the difference between LptH A39 and LptA Q43 in the corresponding position. LptH A39 allows for thanatin Y10 to adopt this conformation, where a glutamine would cause a steric clash. However, the AlphaFold2-multimer model of thanatin Y10A does not indicate any bad clashes **(Figure 17J)**. The AlphaFold2-multimer model of thanatin Y10S indicates that all three rotamers of serine would have steric clashes with LptH Y51 (only one rotamer shown here) **(Figure 17K)**. Despite our predictions being experimentally validated with the LptA model, our predictions do not hold true with LptH. AlphaFold2-multimer modeling did not predict the necessity of thanatin Y10, highlighting that experimental structure determination is required to better understand the binding interface.

## *3.5 Characterizing thanatin orthologs against P. aeruginosa LptH*

A small library of thanatin peptides was discovered by mining genomic databases for thanatin orthologs. To assess if any of these sequences better target LptH than *P. maculiventris* thanatin, we generated eight sequences and measured binding affinity with LptH in a BLI and ITC binding assay **(Figure 18A, B)**. The ITC assay parameters needed to be adjusted to use 10 μM LptH in the sample cell to accommodate for weak binders. Overall, the BLI and ITC binding data corroborate each other. None of the eight sequences significantly improve binding to LptH. Three sequences, from *Aelia acuminata* isoform 1 (BLI $K_d$: 23.3 ± 2.0 nM) and 2 (BLI $K_d$: 30.7 ± 0.9 nM) and *Riptortis pedestris* isoform 1 (BLI $K_d$: 17.7 ± 1.8 nM), bind to LptH roughly the same as *P. maculiventris* (BLI $K_d$: 12.9 ± 1.1 nM). One sequence, from *Halymorpha halys*, significantly weakens binding to LptH (BLI $K_d$: 58.7 ± 4.3 nM). Three sequences, from *Chinavia ubica*, *Murgantia histrionica*, and *Nezara viridula* isoform 1, do not bind LptH. These three sequences contain A10, which agrees with our loss of binding observation with *P. maculiventris* thanatin Y10A. Interestingly, when binding of *M. histrionica* thanatin with LptH Y51F is assessed via ITC, we observed recovery of binding ($K_d$: 287 nM), giving further evidence of LptH Y51 interfering with the binding of thanatin **(Figure 18C, D)**. It is also interesting to note that AlphaFold2-multimer predicted protein-peptide complex models of all thanatin orthologs presented here to bind (data not shown), despite clear experimental data showing a subset do not bind at all.

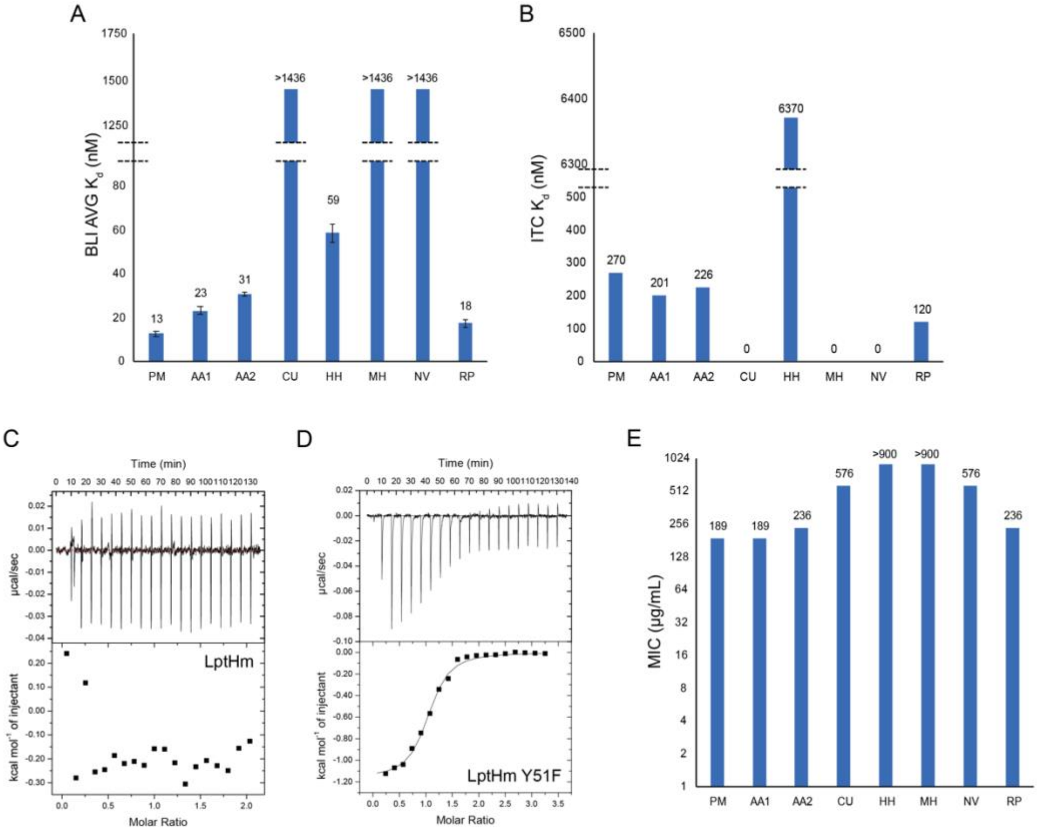

**Figure 18: Thanatin orthologs do not target *P. aeruginosa* LptH better than *P. maculiventris* thanatin.**

**A**, BLI K$_d$ values of thanatin orthologs do not show improvement in binding. Values are averages with error bars representing SEM. *H. halys* thanatin is a weaker binder. *C. ubica*, *M. histrionica*, and *N. viridula* thanatin do not bind at all. **B**, ITC K$_d$ values of thanatin orthologs do not show improvement in binding. *H. halys* thanatin is a weaker binder. *C. ubica*, *M. histrionica*, and *N. viridula* thanatin do not bind at all. **C**, ITC isotherm and binding curve for *M. histrionica* thanatin with LptHm showing no binding. **D,** ITC isotherm and binding curve for *M. histrionica* thanatin with LptH Y51F showing "rescue" of binding. **E**, Thanatin orthologs do not effectively inhibit *P. aeruginosa* RP73 with the aid of PMBN better than *P. maculiventris* thanatin.

To assess if any of the thanatin orthologs were more potent against *P. aeruginosa* than *P. maculiventris* thanatin, the peptides were tested against RP73 in a MIC assay supplemented with PMBN **(Figure 18E)**. Correlating with the binding data, thanatin from *A. acuminata* isoform 1 (MIC: 188 μg/mL) and 2 (MIC: 236 μg/mL) and *R. pedestris* (MIC: 236 μg/mL) had similar potency as *P. maculiventris* (MIC: 188 μg/mL). Thanatin from *C. ubica* and *N. viridula* both had a MIC value of 576 μg/mL. Thanatin from *H. halys* and *M. histrionica* both had MIC values >900 μg/mL. These weaker thanatin peptides inhibiting growth could be an off-target or cytotoxic effect due to high concentration of peptide. Overall, none of the thanatin orthologs screen in this subset of the library improve binding to LptH or inhibition of *P. aeruginosa* in cultures.

## 3.6 Role of C-amidated thanatin with P. aeruginosa

Previous studies have revealed that C-amidation of *P. maculiventris* thanatin improves the potency against *P. aeruginosa* by 4-fold at minimum [69]. We turned to our structural overlay of apo LptH with *E. coli* LptA in complex with *P. maculiventris* thanatin for insight. The C-terminus of thanatin interacts with LptA R76 through bidentate hydrogen bonds. In LptH, Q73 aligns with that position and would cause a loss of interaction with the C-terminus of thanatin. However, C-amidation of thanatin would recover the bidentate hydrogen bonds. To test this prediction, we generated LptAm R76Q to mimic LptH **(Figure 19A)**. LptAm R76Q increased the $K_d$ (3.5 ± 0.8 nM) of *P. maculiventris* thanatin compared to wild type LptAm ($K_d$ of 1.8 ± 0.2 nM). However, the binding affinity was rescued with C-amidated *P. maculiventris* thanatin and LptAm R76Q ($K_d$ of 1.1 ± 0.1 nM). To visualize this interaction, we crystallized and determined the structure of *E. coli* LptAm R76Q in complex with C-amidated *P. maculiventris* thanatin ΔG1-K3 in

**Table 2: X-ray data collection and refinement statistics of *E. coli* LptA R76Q - *P. maculiventris* Thanatin ΔG1-K3 C-amidated complex.**

| | *E. coli* LptAm R76Q-<br>*P. maculiventris*<br>Thanatin ΔG1-K3 C-amidated |
|---|---|
| PDB | XXXX |
| Data collection | |
| Wavelength (Å) | 0.9792 |
| Space group | P 21 21 21 |
| Cell dimensions | |
| *a*, *b*, *c* (Å) | 45.18, 61.17, 148.29 |
| α, β, γ(Å) | 90.0, 90.0, 90.0 |
| Resolution (Å) | 47.18 – 1.64<br>(1.70 – 1.64) |
| $R_{\text{meas}}$ | 0.02353 (0.69) |
| $CC_{1/2}$ | 1 (0.86) |
| Mean *I*/σ*I* | 17.76 (0.98) |
| Completeness (%) | 99.63 (99.31) |
| Redundancy | 2.0 (2.0) |
| Total reflections | 102,546 (10,130) |
| Unique reflections | 51,275 (5065) |
| | |
| Refinement | |
| $R_{\text{work}}$ / $R_{\text{free}}$ | 0.1994/0.2294 |
| No. atoms | 2445 |
| Protein | 2248 |
| Water | 195 |
| Average *B*-factors | 40.92 |
| Protein | 40.52 |
| Water | 45.38 |
| R.m.s. deviations | |
| Bond lengths (Å) | 0.010 |
| Bond angles (°) | 1.13 |
| Ramachandran | |
| Favored (%) | 98.23 |
| Allowed (%) | 1.77 |
| Outliers (%) | 0.00 |

*Values in parentheses are for highest-resolution shell.

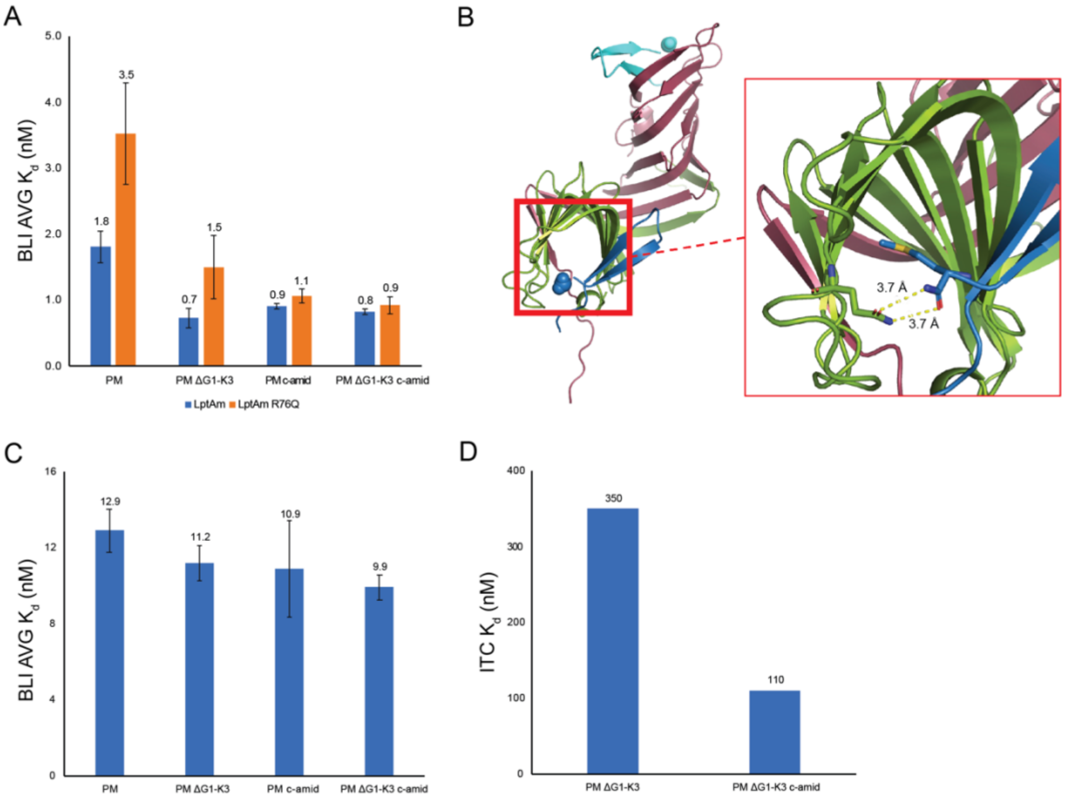

**Figure 19: C-amidation of thanatin may not play a role in improved interactions with LptAm R76Q.**

**A**, BLI $K_d$ values for various thanatin peptides with and without C-amidation in an *E. coli* LptAm model system mimicking LptH. C-amidation of thanatin rescues binding with LptAm R76Q. Values are averages with error bars representing SEM. **B**, Crystal structure of *E. coli* LptAm R76Q with C-amidated *P. maculiventris* thanatin ΔG1-K3. Distance between the C-amidated terminus and R76Q are too long to form hydrogen bonds. **C**, BLI $K_d$ values for various thanatin peptides with and without C-amidation with LptHm. C-amidation of thanatin does not improve binding with LptHm. Values are averages with error bars representing SEM. **D**, ITC $K_d$ values of truncated thanatin with and without C-amidation binding to LptHm. C-amidation of thanatin improves binding with LptHm.

the $P2_12_12_1$ space group, which diffracted to 1.64 Å resolution **(Figure 19B)** (data collection and refinement statistics can be found in **Table 2**). We have previously shown that the N-terminal region of *P. maculiventris* thanatin is dispensable. As seen with the wild type LptAm in complex with thanatin structures, the complex formed a domain-swapped dimer at the C-terminal end of LptAm R76Q due to high concentrations under crystallization conditions. C-amidated thanatin bound to the N-terminal end of each LptAm R76Q protomer. Upon closer inspection of the structure, we unexpectedly did not find the C-amidated end of thanatin proximal enough to the LptAm R76Q side chain to form hydrogen bonds **(Figure 19B)**. The distance between the C-amidated terminus and LptAm R76Q was 3.7 Å, when ideally a hydrogen bond is between 2.7 and 3.3 Å. This could be due to the LptA model system simplifying the differences at the binding interface to one residue change at a time along with the usage of an apo structure of LptH. Multiple residue changes may be needed to model the LptH interface more accurately and thus, the overall conformational change in LptH upon thanatin binding. These questions remain difficult to answer without an experimentally determined structure of LptH in complex with thanatin.

To confirm the improved effect of C-amidation on binding with LptH, we assessed C-amidated thanatin and C-amidated thanatin ΔG1-K3 in our BLI binding assay **(Figure 19C)**. Unexpectedly, we did not observe any significant changes in $K_d$ via BLI with the wild-type and C-amidated thanatin or thanatin ΔG1-K3. We assessed C-amidated thanatin ΔG1-K3 with LptH in our ITC binding assay and observed a tighter $K_d$ (110 μM) than wild type thanatin ΔG1-K3 ($K_d$: 350 μM) by roughly 3-fold **(Figure**

**19D)**. These discrepancies along with the lack of hydrogen bonds between LptAm R76Q and C-amidated thanatin perhaps indicate that the improved potency of C-amidated thanatin does not affect binding, but cell permeability instead. C-amidation would remove a negative charge from the peptide, thus increasing the net positive charge. Positive charges have been shown to help increase membrane permeability, such as seen with cell permeating peptides [99]. Further MIC experiments with *P. aeruginosa* with and without a compromised membrane will be needed to explore this.

## 3.7 Discussion

*P. aeruginosa*, an opportunistic pathogen, remains complex and difficult to study and treat. Current treatments of *P. aeruginosa* infections utilize small molecules that disrupt cell wall synthesis, RNA translation, DNA replication, and inhibit β-lactamases [114]. However, as multi-drug resistance is on the rise, a new class of inhibitors, such as peptide inhibitors, will be necessary. Here, we provided a characterization study on thanatin, an antimicrobial peptide, against LptH of *P. aeruginosa*. LptH remains to be greatly understudied with currently less than ten publications on it. LptH was shown to be essential for LPS transport and oligomeric, however only an apo monomeric crystal structure was reported [108]–[110]. We used *in silico* modeling to predict the binding site of thanatin on LptH and confirmed via *in vitro* binding assays that thanatin does bind LptH. Using structural homology analysis and mutagenesis, we identified LptH Y51 as a main resistance factor of thanatin. We attempted to improve thanatin towards targeting

LptH, however none of our designs improved binding. Interestingly, we found that thanatin Y10 is critical for binding LptH. Our findings were also reflected in our screening of novel thanatin orthologs against LptH, where none of them were found to better target LptH. Lastly, we used structural homology analysis and mutagenesis to gain insight on the role of C-amidation improving potency. Difficulties arose when applying our findings with an *E. coli* LptA model system to LptH, therefore highlighting the need of a high-resolution structure of LptH in complex with thanatin to better understand the mechanism of binding. Our attempts at crystallizing LptH in complex with thanatin through general screening did not yield any crystals. It is quite plausible that upon thanatin binding, LptH undergoes a large conformational change that is not reflected in any of our modeling.

In lieu of a protein-peptide complex structure, we turned to several *in silico* techniques to generate a model. We originally attempted to model LptH with thanatin (data not shown) using AutoDock Vina and HADDOCK [115], [116]. However, many of these results did not resemble realistic structures. Thanatin would lose its secondary and tertiary structure, often resembling an unfolded peptide. If thanatin did resemble a β-hairpin, it would be positioned in unexpected ways such as perpendicular to or upside down to the binding interface with very poor contacts. Continuing the LptH β-sheet through backbone mediated interactions with thanatin is a main feature of binding that was often not captured. We assessed that AutoDock Vina did not allow for much backbone flexibility, which could be useful for small molecule docking but not peptide docking. HADDOCK did incorporate more backbone flexibility, however we were unable to take advantage of their distinct feature of incorporating experimentally determined restraints due to the lack of data. Both AutoDock Vina and HADDOCK required known input structures, which could also complicate the results by not accounting for conformation change upon binding which requires extensive backbone flexibility. Both AutoDock Vina and HADDOCK also required the definition of the binding interface, placing bias on where the ligand could be docked to and possibly missing other binding sites if no previous knowledge was known. The recent releases of AlphaFold2 and AlphaFold2-multimer have made great advances in the field of protein structure prediction [111], [112]. AlphaFold2 and AlphaFold2-multimer uses artificial intelligence to predict 3D protein structures or protein complex structures based on the amino acid sequence alone. The algorithm was trained on a large library of crystal structures from the Protein Data Bank and was shown to be highly accurate in their predictions. In our own hands, AlphaFold2 and AlphaFold2-multimer have proved to be useful in generating initial models before structure determination. However, experimentally determined structures will still be necessary. Highlighted in our work here, is the discrepancy between AlphaFold2-multimer predictions and *in vitro* binding data. AlphaFold2-multimer predicted single point mutants thanatin Y10A and Y10S, and thanatin from *C. ubica*, *M. histrionica*, and *N. viridula* to form a complex with LptH, however our orthogonal binding studies do not indicate binding. This discrepancy demonstrates that despite recent advances in the field, the necessity to experimentally validate *in silico* models remain to be standard.

Future experiments to further understand the mechanism of action of thanatin with LptH will be needed. Our studies highlight the importance of thanatin Y10, which did not vary much in our small library of thanatin orthologs. We unexpectedly saw elimination of binding with a Y10A and Y10S mutation. Thanatin from *Shewanella* was not tested in this study due to later discovery and deposition in the genomic data bank. However, thanatin from *Shewanella* could be of interest for future studies. *Shewanella* thanatin contains F10 which would allow us to interrogate if the aromatic ring of F10 and Y10 or if the hydroxyl group on Y10 is critical for binding LptH. Without a current structure, this mutagenesis study could further map out the interaction. The thanatin library was mostly generated from insect hosts, however the potential symbiotic link between *Shewanella* and thanatin could indicate a new source of thanatin hosts or new group of targets. *Shewanella*, *Pseudomonas*, and *Burkholderia* (previously of the *Pseudomonas* genus), are all found in marine sponges [117], [118]. It is plausible that the microbial population in the marine sponge is regulated by an antimicrobial peptide similar to thanatin, however it has evaded discovery due to divergence over the years. Marine sponges have already been reported as a rich source of antibiotics and antimicrobial peptides [119]–[121]. Sequencing and mining of the marine sponge genome for cryptic peptides could produce more thanatin orthologs or a novel group of antimicrobial peptides, similar to how the thanatin orthologs were discovered, that better target *Pseudomonas* and similar pathogens.

## 3.8 Materials and methods

### 3.8.1 Generating AlphaFold2-multimer models

Models were generated by using the online notebook of AlphaFold2-multimer hosted by DeepMind on Google Colaboratory (https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb) [111], [112]. The protein and peptide sequences were input separately to select for the multimer model. The multiple sequence alignment against the genetic databases were set to their default parameters with jackHMMER. Multimer models were refined for a maximum of three cycles and relaxed by AMBER.

### 3.8.2 Molecular cloning, expression, and purification of *P. aeruginosa* LptH and mutants

A monomeric mutant (Δ164-175) of LptH, modeled after a monomeric mutant of *E. coli* LptA [49], without the signal peptide sequence (Δ1-24) was inserted into a modified pET24b vector through Takara Bio In-Fusion cloning to yield the final construct of His$_6$-SUMO-LptH. For BLI binding studies, a C-terminal flexible linker and AviTag were incorporated to yield the final construct of His$_6$-SUMO-LptH-GSGGSGSG-AviTag. Single point mutations were made with a Q5 Site-Directed Mutagenesis Kit from New England BioLabs.

Purification and expression of *P. aeruginosa* LptH and mutants were identical to the purification of *E. coli* LptA outlined in Section 2.9.2. In brief, all constructs were confirmed by DNA sequencing before transformation and expression in BL21 (DE3) competent cells. Cells were grown in LB media with kanamycin and induced with 1 mM IPTG for an overnight growth at low temperature. The cell pellet was lysed and the supernatant was applied to a Ni-NTA column. The His-tagged LptH was eluted with imidazole on a step gradient. The SUMO tag was removed through SENP1 protease digest. The digest was applied to a Ni-NTA column to collect the flow through. The cleaved LptH was further purified on a Superdex 200 column before being prepped for storage. For BLI binding studies, LptH was biotinylated with BirA and further purified on a S75 column before prepped for storage.

### 3.8.3 Molecular cloning, expression, and purification of thanatin peptides

Cloning, expression, and purification of the thanatin orthologs were identical to that outlined in Section 2.9.3. In brief, cloning of all the orthologs was achieved through In-Fusion Cloning from Takara Bio. All constructs were confirmed by DNA sequencing before transformation and expression in BL21 (DE3) competent cells. Cells were grown in LB media with ampicillin and induced with 1 mM IPTG for an overnight growth at low temperature. The cell pellet was harvested, lysed, and applied to a Ni-NTA column. The His-tagged thanatin was eluted with imidazole on a step gradient. The His-tagged thanatin was further purified on a Superdex200. The His-tag was removed by TEV protease digest. The digest was applied to a Ni-NTA column before being prepped for storage.

### 3.8.4 Molecular cloning, expression, and purification of *E. coli* LptA mutants

Single point mutations to the wild-type *E. coli* LptA construct outlined in Section 2.9.2 were made with a Q5 Site-Directed Mutagenesis Kit from New England BioLabs. Expression and purification of the E. coli LptA mutants were identical to that outlined in Section 2.9.2. In brief, all constructs were confirmed by DNA sequencing before transformation and expression in BL21 (DE3) competent cells. Cells were grown in LB media with kanamycin and induced with 1 mM IPTG for an overnight growth at low temperature. The cell pellet was harvested, lysed, and applied to a Ni-NTA column. The His-tagged LptA mutant was eluted with imidazole on a step gradient. The SUMO tag was removed through SENP1 protease digest. The digest was applied to a Ni-NTA column to collect the flow through. The cleaved LptA was further purified on a Superdex 200 column before being prepped for storage. For BLI binding studies, LptA was biotinylated with BirA and further purified on a S75 column before prepped for storage.

### 3.8.5 Bio-layer interferometry binding assay for *P. aeruginosa* LptH and mutants

BLI binding assays for *P. aeruginosa* LptH and mutants were identical to that outlined in Section 2.9.4. In brief, biotinylated LptH was loaded onto streptavidin biosensors. The loaded biosensors were then immersed into a titration of thanatin peptide. Association of peptide was recorded for 1 hour. A 15-point titration of peptide from 1436 nM to 0.6 nM was used. At times, a higher concentration range was needed for weaker binders. All experiments were double-referenced. Curve fitting and data analysis was performed with a steady state 1:1 binding model. Biosensors were regenerated for multiple uses.

### 3.8.6 Bio-layer interferometry binding assay for *E. coli* LptA mutants

BLI binding assays for *E. coli* LptA mutants were identical to that outlined in Section 2.9.4. In brief, biotinylated LptA mutant was loaded onto streptavidin biosensors. The loaded biosensors were then immersed into a titration of thanatin peptide. Association of peptide was recorded for 1 hour. At times, a higher concentration range was needed for weaker binders. All experiments were double-referenced. Curve fitting and data analysis was performed with a steady state 1:1 binding model. Biosensors were regenerated for multiple uses.

### 3.8.7 Isothermal calorimetry titration binding assay for *P. aeruginosa* LptH and mutants

ITC binding assay for LptH and thanatin orthologs was developed on a MicroCal VP-ITC. All samples were degassed before each experiment. Thanatin peptide was loaded into the syringe at 100 μM and titrated into a sample cell containing 10 μM LptH at 25 °C in Buffer (50 mM Tris pH 8.5, 200 mM NaCl). For tighter binders, 25 μM thanatin and 25 μM LptH were used. After an intial delay of 600 seconds, the first injection was 7.5 μL for 7.5 seconds, followed by 15 μL injections for 15 seconds followed by an equilibration time of 400 or 600 seconds. The reference power was 10 μCal/sec, stirring speed of 307, and filter of 2 seconds. Data analysis and curve fitting were done with Origin with a 1:1 binding model.

### 3.8.8 Minimum inhibition concentration assay for *P. aeruginosa* RP73

A MIC assay was developed with *P. aeruginosa* RP73 based on the Clinical & Laboratory Standards Institute (CLSI) standard protocol [107]. *P. aeruginosa* was streaked onto an antibiotic free LB plate from a cryo stock and incubated at 37 °C overnight. A single colony from this plate was streaked onto an antibiotic free LB plate and incubated at 37 °C overnight. A single colony was added to LB media (10 mL) and grown at 37 °C to OD600 ≈0.1 (corresponding to ≈1 – 2 x $10^8$ CFU/mL) with shaking. Cells were plated at a final concentration of ≈5 x $10^5$ CFU/mL in Cation-Adjusted Mueller-Hinton Broth (CAMHB) supplemented with 10 μg/mL polymyxin B nonapeptide (PMBN) over a titration of peptide inhibitor in clear 96 well plates. The plate was incubated at 37 °C for 20 hours. MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide) (10 μL of 5 mg/mL) was added to each well and incubated at 37 °C for 3 hours. Isopropanol + 0.1 M HCl (100 μL) were added to each well. The plate was imaged and read at OD690 and OD5750 for analysis. The MIC was determined to be the maximum peptide concentration where no growth was observed.

### 3.8.9 Crystallization and structure determination of *E. coli* LptA R76Q with *P. maculiventris* thanatin Δ1-3 C-amidation

Crystallization and structure determination of *E. coli* LptA R76Q with *P. maculiventris* thanatin Δ1-3 C-amidation were similar to that outlined in Section 2.9.6, with the following modifications. In brief, the LptA mutant and thanatin construct were co-eluted through a Superdex200 column and concentrated to 5 and 10 mg/mL for crystal tray assembly via sitting drop method in a 1:1 mixture. Crystallization solutions that yielded crystals ranged from 60 – 72.5% MPD at 0.1 M MES pH 5.8. The crystallization condition of the reported dataset was 10 mg/mL protein complex in 0.1 M MES pH 5.8, 60% MPD.

Diffraction data were collected at NECAT 24-ID-C beamline. The 3D model was constructed using molecular replacement from the PHASER module in PHENIX with a structure of *E. coli* LptA in complex with *P. maculiventris* thanatin as the input model. COOT and PHENIX were used for iterative model building and refinement.

# 4. Conclusions

## *4.1 Discovery of novel thanatin more potent towards E. coli*

Here, we reported a small library of novel thanatin orthologs by mining genomic databases. We characterized the small library of thanatin orthologs against *E. coli*, which led to the discovery of more potent thanatins from *Chinavia ubica* and *Murgantia histrionica* when compared to *Podisus maculiventris* thanatin, the most reported on thanatin. Improved potency was due to better binding and cell membrane permeability from residues A10 and I21 in *C. ubica* and *M. histrionica* thanatin. We present a novel stapled thanatin scaffold based on the redesign of *M. histrionica* thanatin that removes the necessity of a disulfide bond. Although binding affinity of the stapled thanatin was retained, potency was slightly decreased, thus further optimization is needed.

The work presented in this chapter highlights the wealth of knowledge available in genomes across all forms of life waiting to be discovered. Only a small library of thanatin sequences were discovered here, however many more orthologs are likely to exist. Taking full advantage of the genomic knowledge available to us will require some sophisticated and creative protocols to efficiently mine through the vast amount of data. This work also presented several high-resolution crystal structures of thanatin with *E. coli* LptA. A wealth of mutagenesis, binding, and inhibition data were reported in literature, however understanding the results were not always definitive without a structure. Our structures help provide a better understanding of previously reported results. This work also presented a better understanding of the mechanism of action of thanatin. We found that potency of thanatin is not only driven by binding, but also cell permeability, thus highlighting the multiple mechanisms of action of thanatin reported in literature. This work also presents a novel approach to redesigning thanatin to make it more drug-like. *In vitro* characterization of the disulfide bond was not reported in literature until our findings. Despite not improving potency, we present a novel scaffold for the next generation of thanatin-based AMPs.

## *4.2 Insight towards thanatin with P. aeruginosa LptH*

Here, we characterized *P. maculiventris* thanatin against *P. aeruginosa* LptH to gain insight on the mechanism of action of thanatin. We used *E. coli* LptA for homology modeling and identified key differences between LptA and LptH in the binding interface with thanatin. We assessed these differences in our *E. coli* model system and LptH system to identify LptH Y51 as a resistance factor against thanatin. We attempted to overcome this resistance factor by mutating thanatin Y10 and showed success in our *E. coli* model system, however they did not translate to our LptH system. We used AlphaFold to model the interaction, however they agreed with our inaccurate predictions, indicating that an experimentally determined structure of the complex was needed to better understand the interaction. We applied our small library of thanatin orthologs to *P. aeruginosa* and did not find any thanatin orthologs better in binding LptH or inhibiting *P. aeruginosa* growth than *P. maculiventris* thanatin. However, our screening of thanatin orthologs highlighted the importance in thanatin Y10 for binding LptH again. We also used our *E. coli* LptA homology modeling to gain insight on why C-amidation of thanatin improves potency against *P. aeruginosa*. We reasoned that a C-amidated terminus rescues interaction with LptH Q73. Our findings were supported by binding assays with our *E. coli* model system, however when we crystalized and determined the structure of LptA R76Q with C-amidated thanatin ΔG1-K3, we did not observe the recovered interaction. In addition, when we applied our reasoning to LptH, we did not observe recovery of binding with c-amidated thanatin and LptH via BLI, however we did via ITC. These multiple instances of conflicting data indicate further experiments will be needed to fully understand the mechanism of action of thanatin with LptH.

This work presented here highlights how different *P. aeruginosa* LptH is from *E. coli* LptA, despite high sequence and structure homology. We thought that modeling the LptA interface to mimick LptH would be sufficient to gain initial understanding of the binding interface differences. Despite correlation with our predictions and *E. coli* system model, we could not translate the majority of our results to LptH. Even computationally modeling the interaction of thanatin with LptH proved to be difficult due to misleading models. This highlights that despite recent rapid advances in model prediction, experimentally determined structures are still needed as a final confirmation. Despite conflicting data, we were able to identify LptH Y51 and thanatin Y10 as key residues in binding. This work also highlights one possible explanation for improved potency of C-amidated thanatin through better binding to LptH. However, conflicting data may indicate that improvement is due to another mechanism of action, such as cell membrane permeability. Unsuccessful attempts to crystallize and determine the structure of thanatin in complex with LptH have been made. Further experiments will be needed to elucidate how thanatin interacts with *P. aeruginosa* in general and specifically with LptH.

## 4.3 Future work

This dissertation presents further insight on the mechanisms of action of thanatin interacting with *E. coli* LptA and *P. aeruginosa* LptH. Further questions in multiple directions also arise from this dissertation. It would be interesting to update the thanatin library and consensus sequence as more genomes are deposited, but also to discover more distantly related AMPs similar to thanatin. We used a very stringent percent similarity cutoff while searching for thanatin orthologs, but perhaps it will be meaningful to look at fragments of thanatin that have evolved into different branches of AMPs to further diversify the library. Difficulties may arise from trying to align even shorter lengths of sequences than full length thanatin, however development of computational tools could help in determining what is a significant match or not. This library could be screened and characterized against other pathogens to find effective AMPs. Our screening against *P. aeruginosa* unfortunately did not provide any hits.

Without a structure of LptH and thanatin, our studies were at a disadvantage. Since general screening did not produce any crystals, it would be worthwhile to attempt a solution NMR structure of the complex like the reported structure of *E. coli* LptA in complex with thanatin. Further structural characterization will be needed to understand the relationship between LptH Y51 and thanatin Y10.

Characterizing the thanatin orthologs against other reported mechanisms of actions could also give insight on why specific sequences are improved in potency. For example, it would be interesting to study the interactions of *M. histrionica* thanatin with LPS to see if binding affinity is improved to better characterize the membrane perturbation. *P. maculiventris* thanatin was reported to be a dimer in the presence of LPS micelles. Perhaps the sequence of *M. histrionica* thanatin breaks the dimer in LPS micelles, thus lowering the energy barrier to disrupt the membrane more effectively. Structurally determining and characterizing the binding affinity of *M. histrionica* thanatin with LptD and NDM-1 could also give further insight on their contributions. It would also be interesting to further characterize the role of the disulfide bond regarding transport across the membrane. Investigating whether thanatin is transported through an outer membrane protein or through pore formation in a linear form or β-hairpin structure would further improve understanding the mechanism of thanatin. *M. histrionica* thanatin could be acting upon *E. coli* in multiple synergistic ways, however we have only explored part of the mechanism of this molecule.

In addition, computational design tools could be applied to redesign thanatin in multiple ways. Peptide design can be used to improve the binding affinity of thanatin towards a specific LptA ortholog, such as LptH. It would also be interesting to see if any clinical strain of Gram-negative bacteria has mutations in LptA or LptD at the binding interface with thanatin. Peptide design can also be used to help thanatin overcome those resistance mutations. Experimentally determined structures of the complex would be needed for accurate designs though. Peptide design can also be used to synthetically improve thanatin, such as the lactam bridge presented in this work. Different staples and linkers can be modeled in to see if potency or membrane permeability can be improved or to make it more like a macrocycle. The unmodified residues can be swapped out for more favorable amino acids, non-natural amino acids, or D-amino acids, such as those seen in other successful drug-like AMPs. Once thanatin is more drug-like, it would be interesting to test them in clearing bacterial infections in a murine model. For example, bacteria of interest can be injected into a mouse to induce sepsis and intravenously administered thanatin can be provided as treatment. Survival rate can be the monitored outcome. Another example would be to induce bacterial infection on a skin lesion and topically apply thanatin as treatment. Wound healing can be the monitored outcome. The design of animal model studies will depend on future findings about thanatin to narrow its specific application. Overall, much more can be learned about thanatin to help develop it into a successful therapeutic agent to combat pathogens in a clinical setting.

### 4.3 Conclusion

The work presented in this dissertation adds to our overall understanding of AMPs. AMPs are a largely underutilized source of natural products with promising therapeutic potential. As more AMPs are characterized, we advance our understanding of how to effectively target pathogens in new ways. Adding new options to our toolbox is important due to many pathogens developing multi-drug resistance in clinical settings. Treatment of primary diagnoses is often prolonged and complicated due to a difficult hospital acquired infection that must be addressed first. Thus, developing broadly inhibiting AMPs for therapeutic usage could have its rare advantage of being a one-treatment-fits-all as a first line of defense. Our studies on thanatin provide insight on the versatility and potency of thanatin to serve as a platform for future studies to further develop thanatin into a clinical drug and for future applications to other AMP systems.

# Appendix A. Effect of lipid composition on the function of hepatitis C virus p7

## *A.1 Introduction*

### A.1.1 Hepatitis C virus background

Upon infection, the hepatitis C virus (HCV) causes acute and chronic inflammation of the liver. Severe infections often lead to cirrhosis, hepatocellular carcinoma, and death through liver failure, especially in vulnerable populations such as those with human immunodeficiency virus (HIV) co-infection or immunosuppression. HCV is a bloodborne pathogen often transmitted through improper use of contaminated medical equipment in a way that exposes an individual to infected blood. In 2019, the World Health Organization estimated that 58 million people globally have chronic HCV infection and attributed ~300,000 deaths that year due to complications caused by HCV [122]. There is no vaccine against HCV due to the difficulties of having to target multiple evasive genotypes, however effective treatment through an oral regiment of pan-genotypic direct-acting antivirals (DAAs) for 12 to 24 weeks is available. Despite great progress in treatment of hepatitis C, access to these DAAs is limited in low-income countries and new antiviral treatments will be necessary as the virus evolves and acquires resistance to current treatment.

HCV is comprised of a ribonucleic acid (RNA) genome that is directly translated into a polyprotein precursor and processed into three structural and seven nonstructural proteins. Current HCV treatment targets the viral proteases and polymerase, however the nonstructural protein p7 could be of druggable interest. p7 (monomeric molecular weight 6.7 kDa, 63 amino acids) is a small hydrophobic protein that oligomerizes into an ion channel that transports $Ca^{2+}$, $Na^{+}$, $K^{+}$, and $H^{+}$ ions [123]–[125]. p7 is localized in the endoplasmic reticulum and has been shown to play a necessary role in virus propagation through viral replication, release of virions, and capsid assembly [126], [127]. It can be inhibited by small molecules such as amantadine, rimantadine, and hexamethylene amiloride [123], [125], [128]. However, the inhibition response varies with the genotype of HCV [128]. Despite recent progress in research, the mechanism of how p7 functions is still unknown. To gain further insight on p7, several groups have performed structural studies. However, these studies present conflicting data and lack consensus. Thus, understanding p7 remains to be a challenge.

### A.1.2 Reported data supporting C6 and C7 oligomeric state

The oligomerization state of p7 has been reported as both C6 and C7, but never as a heterogenous solution. p7 (genotype 1b) was first proposed to form hexameric channels through cross-linking and TEM [123], [129]. A cross-linked p7 (genotype 1b) sample was shown to migrate at 42 kDa through sodium dodecyl sulfate - polyacrylamide gel electrophoresis (SDS-PAGE) and a GST-p7 (genotype 1b) sample in liposomes was shown to form hexameric rings through TEM and immunogold labelling that detected the GST tag [123]. Further evidence was established when a 16 Å TEM structure of p7 (genotype 2a) in 1,2-dihexanoyl-sn-glycero-3-phosphocholine (DHPC) micelles showed that p7 formed hexameric channels [129]. The low-resolution structure suggested that the N and C terminus of each p7 monomer pointed towards the ER lumen through antibody binding. Cross-linking of p7 (genotype 2a) and blue native – polyacrylamide gel electrophoresis (BN-PAGE) also showed six bands indicative of a hexamer [129]. The structure was too low of a resolution to give further insight into p7 though. More recently, a solution NMR structure of p7 (genotype 5a*) in dodecylphosphocholine (DPC) micelles was published (PDB ID: 2M6X) [130]. The structure was determined through simulated annealing (SA) with nuclear Overhauser effect (NOE) and residual dipolar coupling (RDC) constraints. The structure positions the N and C terminus of each p7 monomer towards the cytosol, opposite of what the TEM structure proposed. Additional negatively stained TEM images were published that also showed hexameric channel formation [130]. The same group then reported that p7 (genotype 5a*) in 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC):DHPC bicelles also formed hexamers through an oligomer-labeling and cross-linking method [131]. SDS-PAGE showed six bands indicative of a hexamer. Overall, the NMR structure displaying C6 symmetry is the most detailed p7 structure reported so far.

p7 has also been reported as a heptamer in literature. The first TEM images of p7 reported were GST-FLAG-p7 (genotype 1b) in phosphatidic acid (PA):phosphatidylcholine (PC) liposomes [132]. The sample formed heptameric rings. FLAG-p7 (genotype 1b) was shown to migrate in seven different bands in SDS-PAGE to further validate heptameric formation. Cross-linked FLAG-p7 (genotype 1b) was also subjected to matrix assisted laser desorption/ionization – time of flight (MALDI-TOF) and peaks equivalent to the heptameric form could be identified [132]. Extensive modeling has also been done with a heptameric p7 structure. A heptameric p7 (genotype 1b) structure was built with Maestro to support the results observed with point mutations in a functional assay [133]. A heptameric p7 (genotype 2a) structure was also built with Maestro. Both heptameric p7 models were used to model binding sites for drugs to help understand the results from inhibition assays [134]. Lastly, a solution NMR structure of monomeric FLAG-p7 (genotype 1b) in methanol was reported (PDB ID: 3ZD0) [135]. The monomeric hairpin structure was generated from paramagnetic relaxation enhancement (PRE) and NOE restraints. The monomeric structure was then modeled into a heptamer, used for *in silico* screening of inhibitors, validated *in vitro* in a dye release assay, and validated *in vivo* in HCV infected Huh7 cells [135]. These results have provided promising data towards drug design, but a comprehensive structure of p7 is still lacking.

Data from crosslinking experiments, TEM, NMR, and modeling have supported both C6 and C7 structures of p7. The only oligomeric structure of p7 deposited in the protein data bank (PDB) is the solution NMR structure that supports C6 symmetry [130]. Despite the amount of data supporting C6 symmetry, concerns over the proposed NMR structure have been raised [136]. His17 has been shown to be involved in ion conductance, hence it was predicted to point inside the channel [137]. The NMR structure shows His17 pointed away from the channel. The N and C termini in the NMR structure point towards the cytosol, contradicting the antibody binding data that show the N and C termini pointing towards the lumen [129]. The shorter outer transmembrane helix exposes polar residues to the hydrophobic inner membrane, which was simulated to cause improper insertion into a lipid bilayer [136]. The NMR structure was recently shown to cause thinning of the surrounding lipid bilayer via PRE analyses when hexameric p7 was inserted into DMPC:DHPC bicelles [131]. The sample conditions of the NMR structure were also questioned to be monomeric via calculations from $R_1$ and $R_2$ relaxation rates and size exclusion chromatography – multi angle light scattering (SEC-MALS) [136]. However, Chen and colleagues, the authors of the solution NMR structure, recently reported that the discrepancy was due to differences in the assembly protocol and detergent:protein ratio [138]. Concerns about artefactual crosspeaks in the NOEs collected due to incomplete deuteration were also brought to attention [136]. This suggested that false intermolecular contacts were reported. It was also noted that the alignment tensor calculated from the RDCs was large with non-zero rhombicities, which is unexpected for a symmetric oligomer [136]. These observations could allow for multiple assignments and several structures to fit the data.

Discrepancy in the data could also arise from differences due to construct used, genotype used, reconstitution environment, and purification method. Further experiments are needed to better define and understand the true structure and behavior of p7 in its native environment. Here, we show data on reanalyzing the solution NMR structure and determining the effects of the membrane mimicking environment. We show that although the proposed NMR structure could fit the data, there are also alternative C6 and C7 structures that satisfy the data as well. This shows that the data is ambiguous and the conformation space needs to be sampled more to address subunit ambiguity. A functional assay for p7 was developed and we show that p7 is not active in a DMPC environment. These are the same lipids used for determining the NMR structure. However, we show that p7 activity is observed in PA:PC:phosphatidylethanolamine (PE) liposomes and can be inhibited with amantadine. Different reconstitution conditions are needed to better mimic the natural environment of p7. Our findings show that a different interpretation of the NMR data in combination with a functional assay can provide further insight on the structure of p7.

## A.2 Alternative C6 and C7 models of p7

Several concerns about the deposited solution NMR structure of p7 (PDB ID: 2M6X) have been raised in literature. Our own internal analysis, done in collaboration with Jeff Martin and Hong Niu (former members of the Donald Laboratory), of the deposited structure raises several other concerns. To assess the fit of RDCs measured, the $^{1}$H-$^{15}$N RDCs from the deposited structure were back calculated and plotted against the experimentally measured $^{1}$H-$^{15}$N RDCs reported **(Figure 20)**. Poor agreement between the RDCs is observed, as shown by a RMSD of 15.42 Hz. The deposited structure was also assembled using SA, which has an inherent weakness of possibly being trapped in local minima. Therefore, SA does not guarantee that the absolute energy minimum is found. Adequate sampling of the conformational space is necessary but is not guaranteed through this method. Also, due to the symmetric oligomerization of p7, subunit ambiguity increases the complexity in assigning the NMR peaks and assembling the structure. If the conformational space was not sampled enough, multiple models could be generated to fit the data. This is plausible especially given that multiple oligomeric states have been reported in literature.

To assess if the deposited p7 data could support several models, we generated alternative C6 and C7 models of p7 using the deposited data from the solution NMR structure (PDB ID: 2M6X). A monomeric p7 subunit consists of three distinct α-helices and two linker regions. A p7 subunit was assembled by constructing each helix and linker separately using intramolecular NOEs, RDCs, and dihedral restraints. We found that the helices were well defined, but the linkers could not be fully orientated by the restraints alone. We used intermolecular NOEs to orient the first helix around a central symmetry axis, grafted on the first linker region, then incorporated the second helix, and so on until the end of the subunit. Each time a new section was grafted on, the partial structure was oligomerized and refined using XPLOR-NIH with corresponding intramolecular and intermolecular NOEs, NH RDCs, and dihedral angles until all

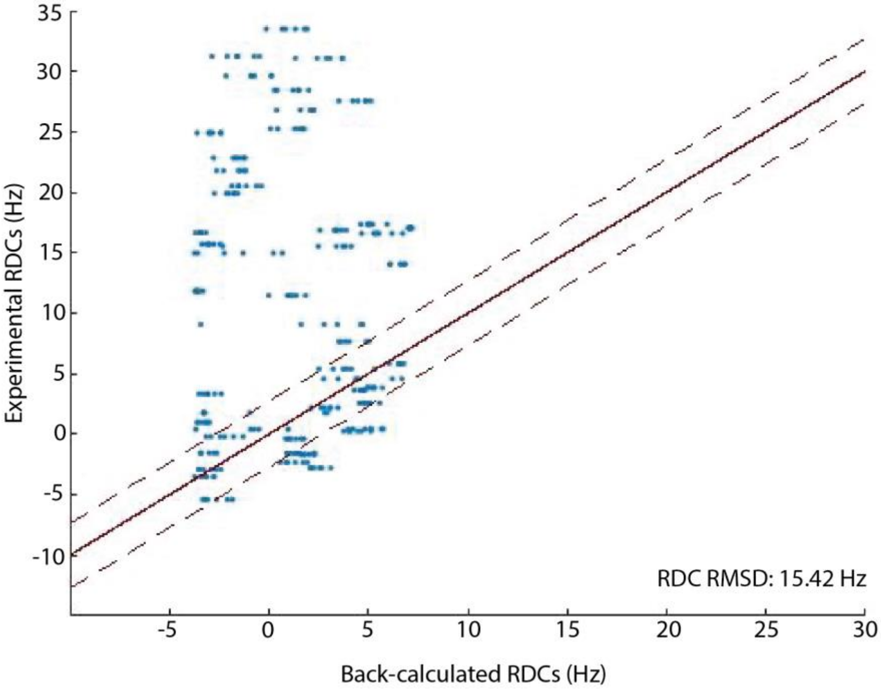

**Figure 20: Poor fit of $^1$H-$^{15}$N RDCs from PDB ID: 2M6X.**

Experimental $^1$H-$^{15}$N RDCs vs back-calculated $^1$H-$^{15}$N RDCs with dashed lines show ± 2.7 Hz error threshold. When comparing the experimental RDCs to the back-calculated RDCs of PDB ID: 2M6X, there is a large disagreement between the RDC values (RMSD: 15.42 Hz).

sections were combined to form a fully oligomerized structure. We found multiple scaffolds at each step, but considered the best structures to be those that most satisfied the NMR data and had good overall XPLOR-NIH energy.

When restricting the data to a C6 symmetry, we found two alternative structures that satisfied the NMR data as well as the deposited NMR structure (PDB ID: 2M6X) **(Figure 21A, B, C)**. Both alternative structures maintained similar levels of satisfaction of NOE and dihedral angle restraints, however the RDC restraints were better satisfied when compared to the deposited NMR structure (PDB ID: 2M6X) **(Table 3)**. One

**Table 3: Summary of NMR restraint statistics.**

| | 2M6X | C6 - A | C6 - B | C7 |
|---|---|---|---|---|
| **Intramolecular-NOEs** | | | | |
| **Total Number of Restraints** | 1350 | 1350 | 1350 | 1575 |
| **Number Satisfied** | 1127 | 1002 | 978 | 1225 |
| **% Satisfied** | 83.5 | 74.2 | 72.4 | 77.8 |
| **RMSD (Å)** | 0.06 | 0.06 | 0.08 | 0.08 |
| **Number of Differences > 0.05 Å** | 1 | 0 | 0 | 7 |
| **Intermolecular-NOEs** | | | | |
| **Total Number of Restraints** | 138 | 138 | 138 | 161 |
| **Number Satisfied** | 135 | 120 | 108 | 147 |
| **% Satisfied** | 97.8 | 87.0 | 78.3 | 91.3 |
| **RMSD (Å)** | 0.02 | 0.10 | 0.11 | 0.20 |
| **Number of Differences > 0.05 Å** | 0 | 0 | 0 | 14 |
| **RDCs** | | | | |
| **Total Number of Restraints** | 294 | 294 | 294 | 294 |
| **RMSD (Hz)** | 15.42 | 0.35 | 0.46 | 0.46 |
| **Q-factor** | 0.97 | 0.02 | 0.03 | 0.03 |
| **Rhombicity** | 0.03 | 0.00 | 0.00 | 0.00 |
| **Dihedral Angles** | | | | |
| **Total Number of Restraints** | 468 | 468 | 468 | 546 |
| **Number Satisfied** | 410 | 396 | 381 | 448 |
| **% Satisfied** | 87.6 | 84.6 | 81.4 | 82.1 |
| **RMSD (°)** | 17.8 | 2.2 | 2.9 | 2.1 |
| **Number of Differences > 0.05°** | 37 | 24 | 29 | 14 |

Structures include deposited NMR structure (PDB ID: 2M6X), alternative C6 structure (C6 – A) similar to 2M6X, alternative C6 structure (C6 – B) with a novel fold, and alternative C7 structure with a novel fold. Total restraints include repeated measurements for symmetry.

alternative structure, C6 – A, had a similar fold to the deposited NMR structure (PDB ID: 2M6X), suggesting that the deposited NMR structure could be a real solution **(Figure 21B)**. However, the other alternative structure, C6 – B, represented a novel fold where

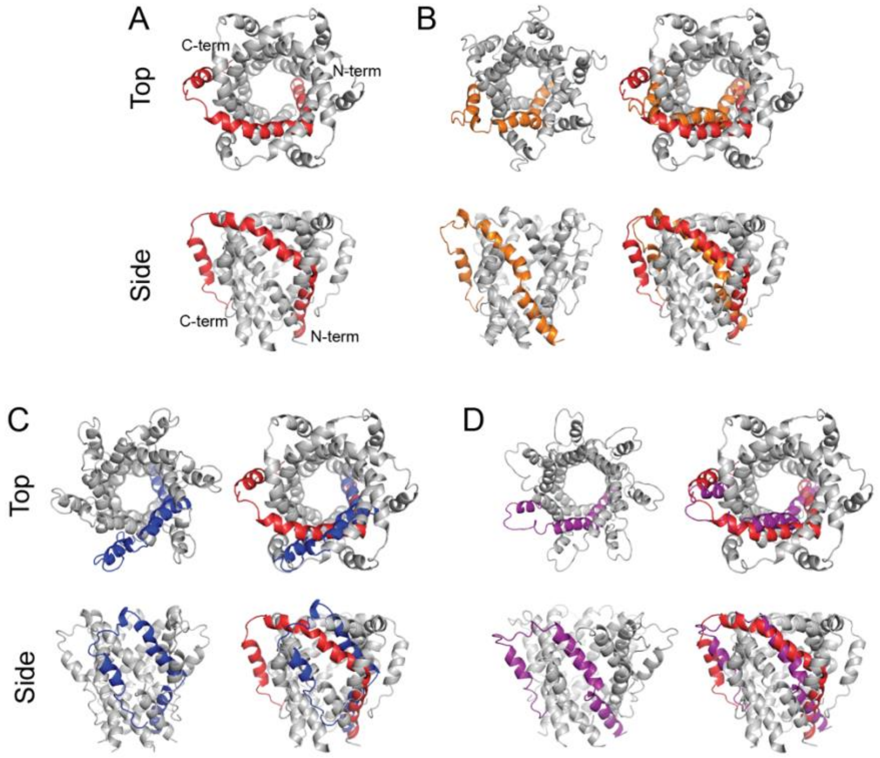

**Figure 21: Alternative C6 and C7 structures of p7.**
**A**, Published structure PDB ID: 2M6X showing C6 symmetry with a monomer highlighted in red. **B**, Alternative C6 - A structure, similar to 2M6X, with monomer highlighted in orange. Orange monomer is also overlayed with 2M6X. **C**, Alternative C6 - B structure, a novel fold, with monomer highlighted in blue. Blue monomer is also overlayed with 2M6X. The C-terminal helix is positioned with a 60° difference around the symmetry axis. **D**, Alternative C7 structure with monomer highlighted in purple. Purple monomer is also overlayed with 2M6X. The N-terminal helix and middle helix are more angled to allow for a seventh subunit.

the main difference lies in the orientation of the third helix. The third or C-terminal helix is positioned with a 60° difference in rotation around the symmetry axis in respect to the C-terminal helix of the deposited structure **(Figure 21C)**.

When restricting the data to a C7 symmetry, we also found an alternative structure, C7 – A, that satisfied the NMR data as well as the deposited NMR structure (PDB ID: 2M6X) **(Figure 21D)**. The alternative structure maintained similar levels of satisfaction of NOE and dihedral angle restraints, however the RDC restraints were better satisfied when compared to the deposited NMR structure (PDB ID: 2M6X) **(Table 3)**. The main difference between the alternative C7 structure and the deposited C6 structure is in the first or N-terminal helix. The N-terminal helix is slightly shifted over to accommodate an additional subunit **(Figure 21D)**.

One caveat to this analysis is that the genotype of p7 may play a crucial role in what oligomerization state is observed. All *in vitro* experiments suggesting hexameric or heptameric oligomerization states are from a mixture of p7 genotypes. Hexameric p7 has been reported from genotype 1b, 2a, and a modified 5a [123], [129], [130]. Heptameric p7 has only been reported from genotype 1b, which was also reported as hexameric [132], [135]. The NMR structure was solved from the modified genotype 5a, which has not been reported to have any heptameric evidence. However, genotype 1b suggests that p7 can be either hexameric or heptameric under certain conditions. Overall, multiple solutions supporting a C6 or C7 symmetry show that the deposited NMR data is ambiguous and further sampling of the configuration space is necessary to elucidate the true structure of p7.

## A.3 p7 in DMPC liposomes is not functional

We developed a functional assay for p7 genotype 5a with DMPC liposomes by modifying a reported dye-release assay that uses p7 genotype 1b and PA:PC:PE liposomes [128], [139]. p7 was inserted into the lipid bilayer of liposomes containing carboxyfluorescein (CF). p7 self-oligomerized into a viroporin, which should allow CF to be released from the liposomes. The CF dye was encapsulated in the liposomes at a self-quenching concentration and became detectable upon dilution into the surrounding solution. The activity was measured by monitoring the fluorescence at $\lambda_{ex}$ 485 nm/$\lambda_{em}$ 520 nm.

The solution NMR structure was solved in DPC micelles, which is not biologically relevant due to the lack of a lipid bilayer [130]. Further NMR studies were reported in DMPC:DHPC bicelles, where the disk was composed of a bilayer of DMPC and the edge of the disk was composed of DHPC [131]. Bicelles, despite providing a lipid bilayer, would not work with this functional assay due to not having an inner and outer compartment, hence liposomes were chosen for the lipid construct instead. We made DMPC liposomes by extrusion and confirmed their size and homogeneity through negatively stained TEM **(Figure 22A)**. To confirm p7 incorporation into the DMPC liposomes, we separated our sample mixture on a density gradient and used a Stewart Assay for lipid detection and SDS-PAGE for protein detection **(Figure 22B, C, D)**. The Ficoll step-gradient separated unincorporated p7, proteoliposomes, and liposomes into

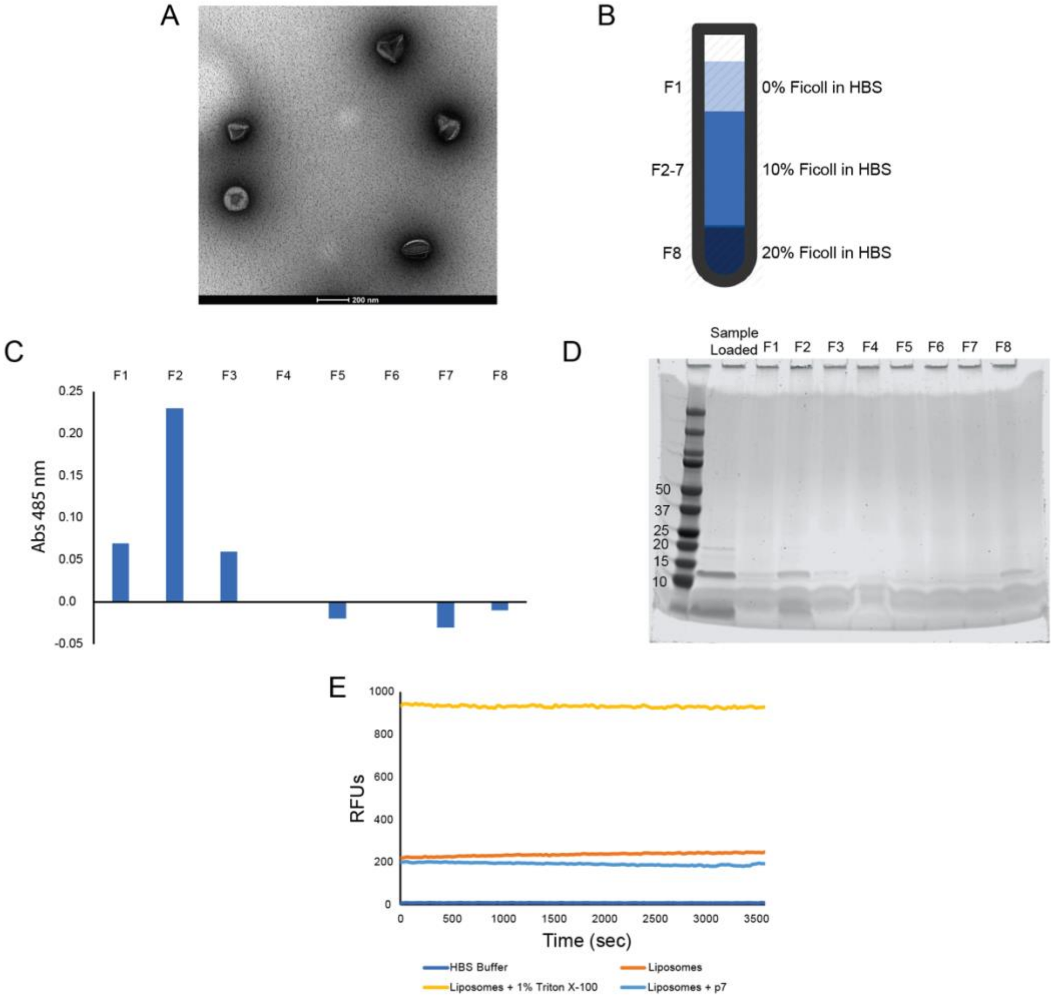

**Figure 22: p7 in DMPC liposomes is not functional.**

**A**, DMPC liposomes were made by extrusion and confirmed by negatively stained TEM. **B**, Schematic of Ficoll step gradient used to separate and fractionate DMPC liposomes incorporated with p7. **C**, Detecting DMPC liposomes with the Stewart Assay from fractions of the Ficoll step gradient. Fractions were reacted with ammonium ferrothiocyanate and phase separated with chloroform. Lipids in complex with ammonium ferrothiocyanate were detected at Abs 485 nm. Lipids were mainly detected in F2. **D**, Detecting p7 with SDS-PAGE from fractions of the Ficoll step gradient. Fractions were visualized for protein on a 4-20% gradient SDS-PAGE gel. p7 was detected in F2 and F8. **E**, Incorporation of p7 into DMPC liposomes did not cause release of CF dye over time.

distinct fractions. Each fraction was tested for lipid and protein detection. Fractions containing only lipids corresponded to liposomes, which was mostly observed in F1 at the top of the gradient. Fractions containing only protein corresponded to unincorporated p7, which was mostly observed in F8 at the bottom of the gradient. Fractions containing both lipids and proteins corresponded to proteoliposomes, which was mostly observed in F2. To confirm that CF was properly encapsulated, we measured the fluorescence with and without disruption from 1% Triton X-100 **(Figure 22E)**. To confirm activity, we added p7 to DMPC liposomes containing CF and monitored the fluorescence at $\lambda_{ex}$ 485 nm/$\lambda_{em}$ 520 nm every 30 seconds for 1 hour **(Figure 22E)**. Surprisingly, we did not observe any activity under these conditions, suggesting an inactive structure of p7 is present with DMPC lipids.

## *A.4 p7 in PA:PC:PE liposomes is functional*

We next assessed the activity of p7 genotype 5a in PA:PC:PE liposomes, the original liposome conditions published in the reported assay [139]. We made PA:PC:PE liposomes by extrusion and confirmed their size and homogeneity through negatively stained TEM **(Figure 23A)**. To confirm p7 incorporation into the PA:PC:PE liposomes, we separated our sample mixture on a density gradient and used a Stewart Assay for lipid detection and SDS-PAGE for protein detection **(Figure 23B, C)**. The Ficoll step-gradient separated unincorporated p7, proteoliposomes, and liposomes into distinct fractions. Each fraction was tested for lipid and protein detection. Fractions containing only lipids corresponded to liposomes, which was not observed in the gradient. Fractions containing only protein corresponded to unincorporated p7, which was mostly observed in F6-F8 at the bottom of the gradient. Fractions containing both lipids and proteins corresponded to proteoliposomes, which was mostly observed in F1-F2. Overall, this could indicate that the liposomes were saturated with p7 so no free liposomes were detectable. To confirm that CF was properly encapsulated, we measured the fluorescence with and without disruption from 1% Triton X-100 **(Figure 23D)**. To confirm activity, we added p7 to PA:PC:PE liposomes containing CF and monitored the fluorescence at $\lambda_{ex}$ 485 nm/$\lambda_{em}$ 520 nm every 30 seconds for 1 hour **(Figure 23D)**. We were able to observe an increase in fluorescence over time, indicating dye release and activity. Additionally, we were able to inhibit the activity of p7 with amantadine, a known small molecule inhibitor of p7 **(Figure 23D)**. Our results indicate that p7 is active in a PA:PC:PE lipid environment, so structural studies under these conditions could provide better insight into the biological structure and function of p7.

We reasoned that the negatively charged polar head group of PA could better interact with the cationic p7 monomer. The pI of p7 genotype 5a is 12.13 with a net charge of 6.3 at pH 7, thus being highly charged under assay conditions. DMPC, PC, and PE contain neutral polar head groups, so the advantage of a charged attraction is not applicable. In addition, p7 is found in the endoplasmic reticulum which is estimated to contain ~15% negatively charged lipids and 85% neutral lipids [140]. Including PA into

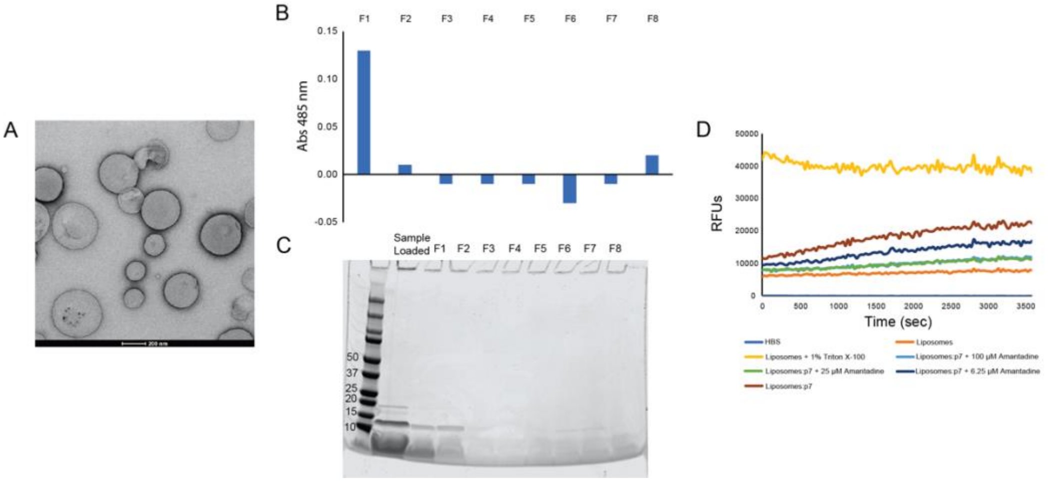

**Figure 23: p7 in PA:PC:PE liposomes is functional.**

**A**, PA:PC:PE liposomes were made by extrusion and confirmed by negatively stained TEM. **B**, PA:PC:PE liposomes incorporated with p7 were centrifuged through a Ficoll step gradient and fractionated into 8 fractions, as similarly shown before with DMPC. Lipid detection with a Stewart Assay was done as previously described. Lipids were mainly detected in F1. **C**, Detecting p7 with SDS-PAGE was done as previously described. p7 was detected in F1-F2. **D**, Incorporation of p7 into PA:PC:PE liposomes causes release of CF dye over time. Amantadine can inhibit the activity of p7.

the lipid mixture better mimics the natural environment of p7 than DMPC alone. However, p7 insertion still occurs most likely due to the hydrophobic properties of p7 and the lipid bilayer. Our activity assay results suggests that the lipid environment is crucial to ensure proper insertion and structure assembly.

## A.5 Discussion

Our analysis and data generate more questions regarding the structure and function of p7. Further experiments are needed to better define and understand p7, where variables in the experimental conditions are carefully selected. A new approach at experimental conditions and data processing could prove to be beneficial in this case.

The deposited solution NMR structure (PDB ID: 2M6X) of p7 in DPC micelles was solved with XPLOR-NIH, which uses simulated annealing [130]. DPC micelles do not provide a lipid bilayer, which is necessary for replicating the biological environment of p7. Simulated annealing has an intrinsic weakness of being trapped in local energy minima, which could lead to inaccurate structures. The symmetric nature of p7 also creates subunit ambiguity in the data. To overcome these three weaknesses, it would be advantageous to reconstitute p7 in nanodiscs and apply Fold Operator Theory (FOT) and DISCO, two algorithms from the Donald Laboratory, to the structure determination process [141]–[143].

Nanodiscs are a recent development in the field that mimics membranes [144]. Nanodiscs are composed of a lipid bilayer enclosed by a membrane scaffold protein (MSP). The MSP determines the diameter of the nanodisc and should be appropriately selected based on the size of the protein. Various constructs have been generated that allow for a diameter from 6 – 30 nm [145], [146]. The lipid bilayer can be swapped out for a range of appropriate lipids. We estimate that a nanodisc with a 9 nm diameter, the MSP1D1 construct, would be appropriate for a hexameric or heptameric p7. Based on our results above, we suggest loading PA:PC:PE lipids into the nanodisc to provide an active environment. Structure determination can be performed in a similar manner via NMR as previously published, however the increase in molecular weight due to the nanodisc (hexameric p7 in MSP1D1: 62.6 kDa, heptameric p7 in MSP1D1: 69.3 kDa) could create some difficulties such as signal overlap and line broadening. The innate symmetric nature of p7 would help reduce signal overlap, however line broadening from a large molecular weight would still require sophisticated and careful measurements. Another recent development, styrene maleic-acid lipid particles (SMALPs), could reduce the size problem [147]. SMALPs are similar to nanodiscs in that they enclose a disk of lipid bilayer, however they do so by using a polymer ring instead of MSPs. SMALPs can achieve similar disk sizes, however they are much smaller in molecular weight, which would be advantageous in an NMR-based approach.

If NMR measurements of p7 in a MSP1D1 nanodisc are obtainable, then the data analysis for structure assembly can be improved upon. FOT, which avoids local energy minima traps, and DISCO, which avoids subunit ambiguity, can be applied to the acquired NOEs and RDCs. FOT is a computational tool that systematically searches the entire energy landscape for topologically distinct folds [143]. It uses mathematical operators to transform an initial fold into alternative folds. The algorithm can be broken down into the following four steps. First, an initial 3D fold is reduced to a 2D schematic to help visualize packing. A p7 monomer can be reduced to 3 helices. This is a rigid oversimplification of the 3D structure, however flexibility in search space will be allowed back in when a 3D structure is reconstructed. Second, operators are defined to search the packing space. Operators can be applied in any order and the result is the same. Third, the operators are applied to the initial fold to generate the search space. The entire search space is evaluated, therefore identifying and avoiding local energy minima. Intermolecular restraints are used here to remove unsatisfied folds. Fourth, a model of each remaining fold is constructed and annealed with XPLOR-NIH. A rigid backbone is used to preserve the fold. Intramolecular restraints are used to guide the construction. Applying FOT to the data should produce an accurate p7 subunit structure.

The subunit is then oligomerized with DISCO in the following five steps [141], [142]. DISCO is an algorithm that computes oligomeric structures by analyzing intermolecular distance restraints and RDC constants, even when atom and subunit ambiguity is present. DISCO interprets an oligomeric structure as a subunit rotated $n$ times around its symmetry axis that can be computed from the RDC constants. It uses the observation that the symmetry axis must be parallel to one of the eigenvectors of the alignment tensor. First, the central orientation of the alignment tensor is computed from the RDC constants and subunit structure. The rhombicity and RMSD are used to evaluate the structural agreement. For p7, a higher-order oligomer, zero rhombicity is expected and the symmetry axis is parallel to the eigenvector with the largest magnitude, often the Z-axis. Second, the uncertainty in the symmetry axis orientation is computed based on the experimental errors of RDCs. A normal distribution with the mean equal to the RDC value is created and fit for an alignment tensor, hence the symmetry axis orientation. This is done for each set of RDC defined in the distribution. This process is repeated 10,000 times to generate a large sample of possible symmetry axis orientations, which is then sampled with a uniformed grid. Third, the maximally satisfying regions (MSRs) are computed for each grid orientation sampled. This region represents a continuous set of symmetry axis positions that represents the oligomeric structures that satisfy the greatest number of intermolecular distance restraints. It is guaranteed that the MSR contains all satisfied structures or at least the structure of best fit. DISCO computes one annulus for each distance restraint that describes an oriented oligomer structure that satisfies the assignment. The region with the most overlapping annuli is the MSR. DISCO can also find the MSR with atom ambiguity, however at a slightly lower precision. Fourth, discrete oligomeric structures are computed from the MSR. The MSR can also be sampled in a uniform grid to evaluate the symmetry axis positions. Fifth, each computed structure is evaluated after energy minimization in XPLOR-NIH with rigid backbone and flexible side chains. RMSD restraint violations are used to measure structural agreement with the intermolecular distance restraints. Van der Waals energy are used to score the structure for intermolecular packing. Violations in these parameters are then used to prune out ill-fitting structures. Applying DISCO to the data should produce an ensemble of well-satisfied oligomeric p7 structures.

As the field of structural biology advances, the use of cryo-EM for structure determination could also provide an alternative solution, especially if p7 in a nanodisc (MW: 62.6 - 69.3 kDa) is too large to measure by NMR. Electron microscopes and data processing techniques are rapidly improving and pushing the lower limit of observable molecular weight. Recently, cryo-EM structures of similarly sized proteins, such as the 64 kDa methemoglobin, 52 kDa streptavidin, and 43 kDa catalytic domain of protein kinase A, have been published [148]. The cyclic symmetry of p7 would also provide an advantage in sorting particles into 2D-class averages and increasing resolution. If structure determination by NMR presents difficulties, then cryo-EM could provide a feasible alternative approach.

Overall, there are multiple approaches to address the conflicting and ambiguous data surrounding p7. Until further experiments are conducted to more thoroughly understand p7, the field remains to be under sampled. Experimental and computational developments will be key to determining a biologically relevant structure of p7. If proven successful, the p7 system can be used as a model system for studying other membrane proteins or homo-oligomeric protein complexes that remain to be a challenge.

## A.6 Conclusions

The structure and function of p7 has been poorly understood due to much conflicting data. p7 has been reported to self-oligomerize into hexamers and heptamers, however never as a heterogenous sample. The most detailed structure of p7 is a solution NMR structure in DPC micelles that supports C6 symmetry. However, several concerns have been raised surrounding the experimental conditions and computational approach used for assembling the structure. Here, we reanalyzed the deposited NMR data and showed that the data are ambiguous. We fit multiple C6 and C7 structures to the data that satisfied the restraints just as well as the deposited structure. To determine the true structure of p7 more accurately, further experiments and sampling of the data will be needed to rule out alternative structures. Conflicting reports of the oligomerization state of p7 could also arise from differences in the construct, reconstitution condition, purification method, and other experimental conditions used. Here, we explored the effect of the lipid environment on p7 by developing an activity assay. We showed that p7 is not active in DMPC liposomes, however, it is active in PA:PC:PE liposomes. Our findings highlight that the variation in experimental conditions could severely affect the structure and function of p7, hence the most biologically relevant conditions should be carefully selected when available.

## *A.7 Materials and methods*

### A.7.1 Determination of alternative C6 and C7 models of p7

This section was done in collaboration with Jeff Martin and Hong Niu (former members of the Donald Laboratory). Deposited NMR data was downloaded from Biological Magnetic Resonance Data Bank (BMRB) Entry 19162. Intramolecular and intermolecular NOEs, RDCs, and dihedral angles were used as restraints. XPLOR-NIH was used to fold the p7 subunit in separate components: helix 1, helix 2, helix 3, linker 1-2, linker 2-3, and C-terminal region. Each component was sequentially appended together around a symmetry axis. The partial subunit was homo-oligomerized and refined with XPLOR-NIH after each addition. Refinements were performed at high temperature, followed by a period of cool, and then low temperature. Top candidates were selected at each step to further build upon. We considered the overall XPLOR-NIH energy, NMR data fit, and diversity in quaternary structure when selecting for the best structure. Local refinement were performed for each final structure by XPLOR-NIH.

### A.7.2 Molecular cloning, expression, and purification of p7

p7 genotype 5a* was cloned into a pWFLE vector to yield the final construct of TrpLE-His$_{11}$-Met-p7. The TrpLE leader sequence brings the protein to inclusion bodies. The His$_{11}$ tag allows for affinity purification. The methionine provides a cleavage site for cyanogen bromide (CNBr) to separate the tag from p7. The gene and vector were cut at NdeI and EcoRI restriction sites and amplified with a Phusion High-Fidelity PCR Kit (New England BioLabs). The PCR products were ligated together with a Quick Ligation Kit (New England BioLabs). After confirmation through DNA sequencing, the plasmid was transformed into c41 (DE3) competent cells on Luria-Bertani (LB) plates containing ampicillin. Cells were grown in LB media with ampicillin at 37 °C until OD600 ≈0.6 with shaking and then induced with 0.5 mM IPTG at 37 °C for 5 hours. The cells were harvested by centrifugation and frozen at -80 °C until purification.

The cell pellet was resuspended in p7 Lysis Buffer (50 mM Tris pH 7.8, 25% sucrose, 1 mM EDTA, 20 mM DTT). The cells were incubated with 1 mg/mL lysozyme for 30 mins, and then with 0.2 mg/mL DNase I, 33 mM $MgCl_2$, 666 μM $MnCl_2$ for 30 mins. A series of detergent washes were used to purify the inclusion bodies. The lysed cells were diluted 3-fold with Detergent Buffer (20 mM Tris pH 7.8, 0.2 M NaCl, 1% deoxycholate, 1% nonidet p-40, 20 mM DTT, 2 mM EDTA) and pelleted. The pellet was washed once with Triton Wash Buffer (0.5% Triton X-100, 20 mM DTT, 1 mM EDTA), twice with Deoxycholate Wash Buffer (0.1% deoxycholate, 20 mM DTT, 1 mM EDTA), and up to five times with $dH_2O$. The final pellet was resuspended in p7 Buffer A (6M guanidine HCl, 50 mM Tris pH 8.0, 200 mM NaCl) and loaded onto a nickel column. The sample was eluted off with an imidazole gradient of 0 – 100% p7 Buffer B (6M guanidine HCl, 50 mM Tris pH 8.0, 200 mM NaCl, 400 mM imidazole). The peak of interest was dialyzed (MWCO: 1 kDa) against $dH_2O$, pelleted, and then lyophilized. The cyanogen bromide cleavage reaction was initiated by incubating the lyophilized pellet in 70% trifluoroacetic acid (TFA) with 0.1 g/mL CNBr for 2 hours. The reaction was terminated under vacuum until all liquid was transferred to a cold trap. The reaction products were dialyzed (MWCO: 3.5 kDa) against $dH_2O$, pelleted, and then lyophilized. The lyophilized mixture was then resuspended in 70% acetic acid and loaded onto a C18 semi-prep column for reversed-phase high performance liquid chromatography (HPLC). The mixture was separated with 46% acetonitrile at 5 mL/min (Buffer A: $dH_2O$ + 0.1% TFA, Buffer B: 100% acetonitrile + 0.1% TFA). The peak corresponding to purified p7 was collected and dried with a rotary evaporator. The purity and identity of each sample was checked with MALDI-TOF and liquid chromatography - mass spectrometry (LC-MS).

### A.7.3 Liposome formation

Two types of liposomes, DMPC and PA:PC:PE, were used in these studies. PA, PC, and PE mixtures were sourced from chicken egg. Both liposomes were made similarly with and without incorporating dye. Empty liposomes were used for imaging and lipid detection. Liposomes loaded with 5(6)-carboxyfluorescein (CF) dye were used for a functional dye-release assay.

The protocol for empty liposomes is as follows. DMPC or PA:PC:PE (1:1:0.5 molar ratio) powder was suspended in chloroform, dried over nitrogen gas, and further dried under vacuum overnight in a glass vial. The thin film of lipids was resuspended in HEPES Buffer Saline (HBS) (10 mM HEPES pH 7.4, 100 mM NaCl) and incubated in a 37 °C water bath for 1 hour. The sample was freeze-thawed four times where one freeze-thaw cycle consisted of 5 mins on dry ice and 10 mins in a 37 °C water bath. The sample was extruded through a 200 nm filter paper at 37 °C with an Avanti Mini Extruder for 21 passes. The sample starts as a cloudy solution and becomes clear with more passes through the extruder. The liposomes can be pelleted at this step by ultracentrifugation at 100,000 g at 10 °C for 20 mins to either buffer exchange or resuspend at a higher concentration for further experiments.

The protocol for liposomes containing CF dye is as follows. DMPC or PA:PC:PE (1:1:0.5 molar ratio) powder was suspended in chloroform, dried over nitrogen gas, and further dried under vacuum overnight in a glass vial. The thin film of lipids was resuspended in HEPES Buffer Saline (HBS) (10 mM HEPES pH 7.4, 100 mM NaCl) with 50 mM CF, covered from the light, and incubated in a 37 °C water bath for 1 hour. CF self-quenches at 50 mM. The sample was freeze-thawed four times where one freeze-thaw cycle consisted of 5 mins on dry ice and 10 mins in a 37 °C water bath. The sample was extruded through a 200 nm filter paper at 37 °C with an Avanti Mini Extruder for 21 passes. The sample starts as a cloudy solution and becomes clear with more passes through the extruder. The liposomes were pelleted by ultracentrifugation at 100,000 g at 10 °C for 20 mins. The supernatant was discarded and the pellet was washed with HBS. This pellet and wash step was repeated for a total of 4 times to remove excess dye. The supernatant starts as a red-orange color, turns into a neon yellow color, and then becomes clear as the washes progress. The final pellet was resuspended in HBS to the desired concentration for further experiments.

### A.7.4 TEM imaging of liposomes

Sample preparation for negatively stained imaging of DMPC or PA:PC:PE liposomes were performed similarly. Liposome concentration could slightly vary, but 0.025 mg/mL DMPC and 0.5 mg/mL PA:PC:PE were typical. A 4 μL solution of liposomes was applied to the formvar/carbon coated copper grid (300 mesh) for 60 secs.

The solution was blotted away with filter paper. The grid was washed with 0.6% uranyl formate and blotted away with filter paper. This wash step was repeated for a total of 2 times. The grid was stained with 0.6% uranyl formate for 60 seconds and then blotted with filter paper. The grid was air dried. Grids were imaged on a FEI Tecnai G$^2$ Twin electron microscope at 80 kV. Images were collected at 25 kx magnification and defocused to -1.49 μm.

### A.7.5 p7 incorporation into liposomes

To confirm p7 incorporation into the liposomes, a liposome control and proteoliposome sample were separated on a Ficoll gradient. The gradient was fractionated and tested for lipid and protein prescence. The Stewart Assay was used to detect lipids. Coomassie stain and SDS-PAGE were used to detect protein. Fractions containing both lipids and protein indicated p7 association with liposomes.

Lyophilized p7 was resuspended in 50% trifluoroethanol (TFE) + HCl, which has been shown to induce α-helical structures. Dye-free DMPC and PA:PC:PE liposomes were prepared as mentioned in section A7.2. Liposome formation. DMPC liposomes were incubated with p7 at a 200:1 molar ratio for 1 hour at room temperature. PA:PC:PE liposomes were incubated with p7 at a 200:1 molar ratio for 1 hour at 37 °C.

A discontinuous Ficoll gradient was generated as follows. The liposome or proteoliposome samples were mixed 40% Ficoll in 1:1 ratio to bring the final Ficoll concentration to 20%. This sample was placed at the bottom of a tube. The volume is 1X.

A 10% Ficoll layer was placed on top of the 20% Ficoll layer containing the sample. The volume is 6X. A layer of HBS (0% Ficoll) was placed on top of the 10% Ficoll layer. The volume is 1X. The tube was ultracentrifuged at 100,000 g for 30 mins at 10 °C. The tube was fractionated into 8 parts of equal volume. Each fraction was run on an SDS-PAGE gel and visualized with Coomassie stain for protein detection. Each fraction was tested via the Stewart Assay for lipid detection.

The Stewart Assay utilizes ammonium ferrothiocyanate which binds to phospholipid head groups to form a colored complex. Ammnoium ferrothiocynante is water soluble, but not soluble in chloroform. However, the lipid complex can be extracted into a chloroform layer using phase separation and be optically detected. Each fraction is mixed with ammonium ferrothiocynante and chloroform in a 1:5:5 ratio. The tubes are centrifuged to separate the aqueous layer and organic solvent layer. The chloroform layer is measured at $\lambda_{abs}$ = 485 nm. The absorbance value is then compared to a standard curve to determine lipid concentration. Using a known lipid stock identical to the lipid being detected, a standard curve can be generated using the same steps mentioned above.

### A.7.6 Dye-release liposome assay

p7 forms a channel upon insertion of a double membrane. To test the function of p7, a dye release assay was developed. p7 was inserted into liposomes containing CF dye at a self-quenching concentration. Upon p7 insertion, the CF dye was released into the supernatant through the viroporin and diluted to detectable concentration. The release of CF dye was monitored by fluorescence at $\lambda_{ex}/\lambda_{em}$ = 485/510 nm.

Lyophilized p7 was resuspended in 50% trifluoroethanol (TFE) + HCl, which has been shown to induce α-helical structures. DMPC and PA:PC:PE liposomes containing 50 mM CF dye were prepared as mentioned in section A7.2. Liposome formation. A 5 uL solution of inhibitor, amantadine HCl, was added to a black 96 wells clear bottom plate. A titration of inhibitor was prepared in a 1:1 serial dilution. A 5 μL solution of p7 in 50% TFE was added to the plate. The mixture of p7 with inhibitor was incubated together for 15 mins at room temperature. A 90 uL solution of liposomes was added to the plate. The final concentration of p7 was 50 nM and the final concentration of PA:PC:PE liposomes was 50 μM. Various concentrations of p7 and DMPC liposomes were tested, however no activity was observed. The fluorescence was read from the bottom at $\lambda_{ex}/\lambda_{em}$ = 485/510 nm every 30 seconds for 1 hour at room temperature.

# References

[1] A. Anastasi, V. Erspamer, and M. Bucci, “Isolation and Structure of Bombesin and Alytesin, two Analogous Active Peptides from the Skin of the European Amphibians Bombina and Alytes,” *Experientia*, vol. 27, pp. 166–167, Aug. 1970.

[2] E. Habermann, “Bee and Wasp Venoms,” *Science*, vol. 177, no. 4046, pp. 314–322, Jul. 1972.

[3] H. Steiner, D. Hultmark, Å. Engström, H. Bennich, and H. G. Boman, “Sequence and specificity of two antibacterial proteins involved in insect immunity,” *Nature*, vol. 292, no. 5820, pp. 246–248, Jul. 1981, doi: 10.1038/292246a0.

[4] M. Zasloff, “Magainins, a class of antimicrobial peptides from Xenopus skin: isolation, characterization of two active forms, and partial cDNA sequence of a precursor.,” *Proc. Natl. Acad. Sci. USA*, 1987.

[5] D. Hultmark, H. Steiner, T. Rasmuson, and H. G. Boman, “Insect Immunity. Purification and Properties of Three Inducible Bactericidal Proteins from Hemolymph of Immunized Pupae of *Hyalophora cecropia*,” *European Journal of Biochemistry*, vol. 106, no. 1, pp. 7–16, May 1980, doi: 10.1111/j.1432-1033.1980.tb05991.x.

[6] M. A. Hanson, L. B. Cohen, A. Marra, I. Iatsenko, S. A. Wasserman, and B. Lemaitre, “The Drosophila Baramicin polypeptide gene protects against fungal infection,” *PLoS Pathog*, vol. 17, no. 8, p. e1009846, Aug. 2021, doi: 10.1371/journal.ppat.1009846.

[7] J. B. Cowland, A. H. Johnsen, and N. Borregaard, “hCAP-18, a cathelin/pro-bactenecin-like protein of human neutrophil specific granules,” *FEBS Letters*, vol. 368, no. 1, pp. 173–176, Jul. 1995, doi: 10.1016/0014-5793(95)00634-L.

[8] A. F. Lacerda, Ã. A. R. Vasconcelos, P. B. Pelegrini, and M. F. Grossi de Sa, “Antifungal defensins and their role in plant defense,” *Front. Microbiol.*, vol. 5, Apr. 2014, doi: 10.3389/fmicb.2014.00116.

[9] J. M. Balkovec, D. L. Hughes, P. S. Masurekar, C. A. Sable, R. E. Schwartz, and S. B. Singh, “Discovery and development of first in class antifungal caspofungin (CANCIDAS®)—A case study,” *Nat. Prod. Rep.*, vol. 31, no. 1, pp. 15–34, 2014, doi: 10.1039/C3NP70070D.

[10] R. J. Dubos, "Studies on a bactericidal agent extracted from a soil bacillus," *J Exp Med*, vol. 70, no. 1, pp. 1–10, Apr. 1939, doi: https://doi.org/10.1084/jem.70.1.1.

[11] S. Kozlov, A. Vassilevski, and E. Grishin, "Antimicrobial Peptide Precursor Structures Suggest Effective Production Strategies," *IAD*, vol. 2, no. 1, pp. 58–63, Jan. 2008, doi: 10.2174/187221308783399261.

[12] G. Wang, X. Li, and Z. Wang, "APD3: the antimicrobial peptide database as a tool for research and education," *Nucleic Acids Res*, vol. 44, no. D1, pp. D1087–D1093, Jan. 2016, doi: 10.1093/nar/gkv1278.

[13] M. Pirtskhalava *et al.*, "DBAASP v3: database of antimicrobial/cytotoxic activity and structure of peptides as a resource for development of new therapeutics," *Nucleic Acids Research*, vol. 49, no. D1, pp. D288–D297, Jan. 2021, doi: 10.1093/nar/gkaa991.

[14] M. D. T. Torres, M. C. R. Melo, O. Crescenzi, E. Notomista, and C. de la Fuente-Nunez, "Mining for encrypted peptide antibiotics in the human proteome," *Nat Biomed Eng*, vol. 6, no. 1, pp. 67–75, Nov. 2021, doi: 10.1038/s41551-021-00801-1.

[15] D. Lin *et al.*, "Mining Amphibian and Insect Transcriptomes for Antimicrobial Peptide Sequences with rAMPage," *Antibiotics*, vol. 11, no. 7, p. 952, Jul. 2022, doi: 10.3390/antibiotics11070952.

[16] C. D. Santos-Júnior, S. Pan, X.-M. Zhao, and L. P. Coelho, "Macrel: antimicrobial peptide screening in genomes and metagenomes," *PeerJ*, vol. 8, p. e10555, Dec. 2020, doi: 10.7717/peerj.10555.

[17] K. Pane *et al.*, "Antimicrobial potency of cationic antimicrobial peptides can be predicted from their amino acid composition: Application to the detection of 'cryptic' antimicrobial peptides," *Journal of Theoretical Biology*, vol. 419, pp. 254–265, Apr. 2017, doi: 10.1016/j.jtbi.2017.02.012.

[18] R. J. González-Hernández *et al.*, "Phosphomannosylation and the Functional Analysis of the Extended Candida albicans MNN4-Like Gene Family," *Front. Microbiol.*, vol. 8, p. 2156, Nov. 2017, doi: 10.3389/fmicb.2017.02156.

[19] K. Matsuzaki, A. Nakamura, O. Murase, K. Sugishita, N. Fujii, and K. Miyajima, "Modulation of Magainin 2−Lipid Bilayer Interactions by Peptide Charge," *Biochemistry*, vol. 36, no. 8, pp. 2104–2111, Feb. 1997, doi: 10.1021/bi961870p.

[20] G. van den Bogaart, J. V. Guzmán, J. T. Mika, and B. Poolman, "On the Mechanism of Pore Formation by Melittin," *Journal of Biological Chemistry*, vol. 283, no. 49, pp. 33854–33857, Dec. 2008, doi: 10.1074/jbc.M805171200.

[21] K. Matsuzaki, O. Murase, N. Fujii, and K. Miyajima, "An Antimicrobial Peptide, Magainin 2, Induced Rapid Flip-Flop of Phospholipids Coupled with Pore Formation and Peptide Translocation," *Biochemistry*, vol. 35, no. 35, pp. 11361–11368, Jan. 1996, doi: 10.1021/bi960016v.

[22] A. Giangaspero, L. Sandri, and A. Tossi, "Amphipathic $\alpha$ helical antimicrobial peptides.: A systematic study of the effects of structural and physical properties on biological activity," *European Journal of Biochemistry*, vol. 268, no. 21, pp. 5589–5600, Nov. 2001, doi: 10.1046/j.1432-1033.2001.02494.x.

[23] D. Gidalevitz *et al.*, "Interaction of antimicrobial peptide protegrin with biomembranes," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 100, no. 11, pp. 6302–6307, May 2003, doi: 10.1073/pnas.0934731100.

[24] J. Penney and J. Li, "Protegrin 1 Enhances Innate Cellular Defense via the Insulin-Like Growth Factor 1 Receptor Pathway," *Front. Cell. Infect. Microbiol.*, vol. 8, p. 331, Sep. 2018, doi: 10.3389/fcimb.2018.00331.

[25] N. Fujitani *et al.*, "Structure of the Antimicrobial Peptide Tachystatin A," *Journal of Biological Chemistry*, vol. 277, no. 26, pp. 23651–23657, Jun. 2002, doi: https://10.1074/jbc.M111120200.

[26] J. B. Schaal *et al.*, "Rhesus Macaque Theta Defensins Suppress Inflammatory Cytokines and Enhance Survival in Mouse Models of Bacteremic Sepsis," *PLoS ONE*, vol. 7, no. 12, p. e51337, Dec. 2012, doi: 10.1371/journal.pone.0051337.

[27] J. Verdon *et al.*, "Armadillidin H, a Glycine-Rich Peptide from the Terrestrial Crustacean Armadillidium vulgare, Displays an Unexpected Wide Antimicrobial Spectrum with Membranolytic Activity," *Front. Microbiol.*, vol. 7, Sep. 2016, doi: 10.3389/fmicb.2016.01484.

[28] C. Imjongjirak, P. Amphaiphan, W. Charoensapsri, and P. Amparyup, "Characterization and antimicrobial evaluation of Sp PR-AMP1, a proline-rich antimicrobial peptide from the mud crab Scylla paramamosain," *Developmental & Comparative Immunology*, vol. 74, pp. 209–216, Sep. 2017, doi: 10.1016/j.dci.2017.05.003.

[29] M. Mattiuzzo *et al.*, "Role of the Escherichia coli SbmA in the antimicrobial activity of proline-rich peptides," *Mol Microbiol*, vol. 66, no. 1, pp. 151–163, Oct. 2007, doi: 10.1111/j.1365-2958.2007.05903.x.

[30] S.-T. Yang, S. Y. Shin, K.-S. Hahm, and J. I. Kim, "Different modes in antibiotic action of tritrpticin analogs, cathelicidin-derived Trp-rich and Pro/Arg-rich peptides," *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1758, no. 10, pp. 1580–1586, Oct. 2006, doi: 10.1016/j.bbamem.2006.06.007.

[31] C. Lawyer *et al.*, "Antimicrobial activity of a 13 amino acid tryptophan-rich peptide derived from a putative porcine precursor protein of a novel family of antibacterial peptides," *FEBS Letters*, vol. 390, no. 1, pp. 95–98, Jul. 1996, doi: 10.1016/0014-5793(96)00637-0.

[32] B. A. Lipsky, K. J. Holroyd, and M. Zasloff, "Topical versus Systemic Antimicrobial Therapy for Treating Mildly Infected Diabetic Foot Ulcers: A Randomized, Controlled, Double-Blinded, Multicenter Trial of Pexiganan Cream," *CLIN INFECT DIS*, vol. 47, no. 12, pp. 1537–1545, Dec. 2008, doi: 10.1086/593185.

[33] R. D. Arbeit, D. Maki, F. P. Tally, E. Campanaro, B. I. Eisenstein, and Daptomycin 98-01 and 99-01 Investigators, "The Safety and Efficacy of Daptomycin for the Treatment of Complicated Skin and Skin-Structure Infections," *Clinical Infectious Diseases*, vol. 38, no. 12, pp. 1673–1681, Jun. 2004, doi: 10.1086/420818.

[34] H. Y. Chow *et al.*, "Establishing the Structure–Activity Relationship of Daptomycin," *ACS Med. Chem. Lett.*, vol. 11, no. 7, pp. 1442–1449, Jul. 2020, doi: 10.1021/acsmedchemlett.0c00175.

[35] I. Lerouge and J. Vanderleyden, "O-antigen structural variation: mechanisms and possible roles in animal/plant–microbe interactions," *FEMS Microbiol Rev*, vol. 26, no. 1, pp. 17–47, Mar. 2002, doi: 10.1111/j.1574-6976.2002.tb00597.x.

[36] C. R. H. Raetz and C. Whitfield, "Lipopolysaccharide Endotoxins," *Annu. Rev. Biochem.*, vol. 71, no. 1, pp. 635–700, Jun. 2002, doi: 10.1146/annurev.biochem.71.110601.135414.

[37] B. S. Park and J.-O. Lee, "Recognition of lipopolysaccharide pattern by TLR4 complexes," *Exp Mol Med*, vol. 45, no. 12, pp. e66–e66, Dec. 2013, doi: 10.1038/emm.2013.97.

[38] Y. Hong, V. A. Morcilla, M. A. Liu, E. L. M. Russell, and P. R. Reeves, "Three Wzy polymerases are specific for particular forms of an internal linkage in otherwise

identical O units," *Microbiology*, vol. 161, no. 8, pp. 1639–1647, Aug. 2015, doi: 10.1099/mic.0.000113.

[39] W. T. Doerrler and C. R. H. Raetz, "ATPase Activity of the MsbA Lipid Flippase of Escherichia coli," *Journal of Biological Chemistry*, vol. 277, no. 39, pp. 36697–36705, Sep. 2002, doi: 10.1074/jbc.M205857200.

[40] Y. Hong, M. A. Liu, and P. R. Reeves, "Progress in Our Understanding of Wzx Flippase for Translocation of Bacterial Membrane Lipid-Linked Oligosaccharide," *J Bacteriol*, vol. 200, no. 1, Jan. 2018, doi: 10.1128/JB.00154-17.

[41] P. D. Abeyrathne, C. Daniels, K. K. H. Poon, M. J. Matewish, and J. S. Lam, "Functional Characterization of WaaL, a Ligase Associated with Linking O-Antigen Polysaccharide to the Core of *Pseudomonas aeruginosa* Lipopolysaccharide," *J Bacteriol*, vol. 187, no. 9, pp. 3002–3012, May 2005, doi: 10.1128/JB.187.9.3002-3012.2005.

[42] Y. Li, B. J. Orlando, and M. Liao, "Structural basis of lipopolysaccharide extraction by the LptB2FGC complex," *Nature*, vol. 567, no. 7749, pp. 486–490, Mar. 2019, doi: 10.1038/s41586-019-1025-6.

[43] H. Dong, Z. Zhang, X. Tang, N. G. Paterson, and C. Dong, "Structural and functional insights into the lipopolysaccharide ABC transporter LptB2FG," *Nat Commun*, vol. 8, no. 1, p. 222, Aug. 2017, doi: 10.1038/s41467-017-00273-5.

[44] Z. Wang *et al.*, "Structural and functional studies of conserved nucleotide-binding protein LptB in lipopolysaccharide transport," *Biochemical and Biophysical Research Communications*, vol. 452, no. 3, pp. 443–449, Sep. 2014, doi: 10.1016/j.bbrc.2014.08.094.

[45] M. D. L. Suits, P. Sperandeo, G. Dehò, A. Polissi, and Z. Jia, "Novel Structure of the Conserved Gram-Negative Lipopolysaccharide Transport Protein A and Mutagenesis Analysis," *Journal of Molecular Biology*, vol. 380, no. 3, pp. 476–488, Jul. 2008, doi: 10.1016/j.jmb.2008.04.045.

[46] I. Botos *et al.*, "Structural and Functional Characterization of the LPS Transporter LptDE from Gram-Negative Pathogens," *Structure*, vol. 24, no. 6, pp. 965–976, Jun. 2016, doi: 10.1016/j.str.2016.03.026.

[47] K. M. Schultz, J. B. Feix, and C. S. Klug, "Disruption of LptA oligomerization and affinity of the LptA-LptC interaction: LptA-LptC Affinity Studies," *Protein Science*, vol. 22, no. 11, pp. 1639–1645, Nov. 2013, doi: 10.1002/pro.2369.

[48] J. A. Merten, K. M. Schultz, and C. S. Klug, "Concentration-dependent oligomerization and oligomeric arrangement of LptA: Oligomerization of LptA Is Concentration Dependent," *Protein Science*, vol. 21, no. 2, pp. 211–218, Feb. 2012, doi: 10.1002/pro.2004.

[49] C. Laguri *et al.*, "Interaction of lipopolysaccharides at intermolecular sites of the periplasmic Lpt transport assembly," *Scientific ReportS*, vol. 7, no. 9715, p. 13, 2017, doi: 10.1038/s41598-017-10136-0.

[50] K. M. Schultz, T. J. Lundquist, and C. S. Klug, "Lipopolysaccharide binding to the periplasmic protein LptA: LPS Binding to LptA," *Protein Science*, vol. 26, no. 8, pp. 1517–1523, Aug. 2017, doi: 10.1002/pro.3177.

[51] G. Chimalakonda, N. Ruiz, S.-S. Chng, R. A. Garner, D. Kahne, and T. J. Silhavy, "Lipoprotein LptE is required for the assembly of LptD by the β-barrel assembly machine in the outer membrane of *Escherichia coli*," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 108, no. 6, pp. 2492–2497, Feb. 2011, doi: 10.1073/pnas.1019089108.

[52] S.-S. Chng, N. Ruiz, G. Chimalakonda, T. J. Silhavy, and D. Kahne, "Characterization of the two-protein complex in *Escherichia coli* responsible for lipopolysaccharide assembly at the outer membrane," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 107, no. 12, pp. 5363–5368, Mar. 2010, doi: 10.1073/pnas.0912872107.

[53] Y. Gu, P. J. Stansfeld, Y. Zeng, H. Dong, W. Wang, and C. Dong, "Lipopolysaccharide is Inserted into the Outer Membrane through An Intramembrane Hole, A Lumen Gate, and the Lateral Opening of LptD," *Structure*, vol. 23, no. 3, pp. 496–504, Mar. 2015, doi: 10.1016/j.str.2015.01.001.

[54] X. Li, Y. Gu, H. Dong, W. Wang, and C. Dong, "Trapped lipopolysaccharide and LptD intermediates reveal lipopolysaccharide translocation steps across the Escherichia coli outer membrane," *Sci Rep*, vol. 5, no. 1, p. 11883, Jul. 2015, doi: 10.1038/srep11883.

[55] G. Malojčić *et al.*, "LptE binds to and alters the physical state of LPS to catalyze its assembly at the cell surface," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 111, no. 26, pp. 9467–9472, Jul. 2014, doi: 10.1073/pnas.1402746111.

[56] S. Okuda, D. J. Sherman, T. J. Silhavy, N. Ruiz, and D. Kahne, "Lipopolysaccharide transport and assembly at the outer membrane: the PEZ model," *Nat Rev Microbiol*, vol. 14, no. 6, pp. 337–345, Jun. 2016, doi: 10.1038/nrmicro.2016.25.

[57] D. J. Sherman, S. Okuda, W. A. Denny, and D. Kahne, "Validation of inhibitors of an ABC transporter required to transport lipopolysaccharide to the cell surface in Escherichia coli," *Bioorganic & Medicinal Chemistry*, vol. 21, no. 16, pp. 4846–4851, Aug. 2013, doi: 10.1016/j.bmc.2013.04.020.

[58] X. Dai *et al.*, "Identification of a Small Molecule That Inhibits the Interaction of LPS Transporters LptA and LptC," *Antibiotics*, vol. 11, no. 10, p. 1385, Oct. 2022, doi: 10.3390/antibiotics11101385.

[59] X. Zhang *et al.*, "Identification of an anti-Gram-negative bacteria agent disrupting the interaction between lipopolysaccharide transporters LptA and LptC," *International Journal of Antimicrobial Agents*, vol. 53, no. 4, pp. 442–448, Apr. 2019, doi: 10.1016/j.ijantimicag.2018.11.016.

[60] E. C. C. M. Moura *et al.*, "Thanatin Impairs Lipopolysaccharide Transport Complex Assembly by Targeting LptC–LptA Interaction and Decreasing LptA Stability," *Front. Microbiol.*, vol. 11, p. 909, May 2020, doi: 10.3389/fmicb.2020.00909.

[61] S. U. Vetterli *et al.*, "Thanatin targets the intermembrane protein complex required for lipopolysaccharide transport in *Escherichia coli*," *Sci. Adv.*, vol. 4, no. 11, p. eaau2634, Nov. 2018, doi: 10.1126/sciadv.aau2634.

[62] J. Schmidt, K. Patora-Komisarska, K. Moehle, D. Obrecht, and J. A. Robinson, "Structural studies of β-hairpin peptidomimetic antibiotics that target LptD in Pseudomonas sp.," *Bioorganic & Medicinal Chemistry*, vol. 21, no. 18, pp. 5806–5810, Sep. 2013, doi: 10.1016/j.bmc.2013.07.013.

[63] M. Werneburg *et al.*, "Inhibition of Lipopolysaccharide Transport to the Outer Membrane in Pseudomonas aeruginosa by Peptidomimetic Antibiotics," *ChemBioChem*, vol. 13, no. 12, pp. 1767–1775, Aug. 2012, doi: 10.1002/cbic.201200276.

[64] G. Andolina *et al.*, "A Peptidomimetic Antibiotic Interacts with the Periplasmic Domain of LptD from *Pseudomonas aeruginosa*," *ACS Chem. Biol.*, vol. 13, no. 3, pp. 666–675, Mar. 2018, doi: 10.1021/acschembio.7b00822.

[65] N. Srinivas *et al.*, "Peptidomimetic Antibiotics Target Outer-Membrane Biogenesis in *Pseudomonas aeruginosa*," *Science*, vol. 327, no. 5968, pp. 1010–1013, Feb. 2010, doi: 10.1126/science.1182749.

[66] "Pivotal Study in Nosocomial Pneumonia Suspected or Confirmed to be Due to Pseudomonas (PRISM-UDR)," *ClinicalTrials.gov*, Aug. 28, 2019. https://clinicaltrials.gov/ct2/show/NCT03582007 (accessed Feb. 12, 2023).

[67] "Pivotal Study in VAP Suspected or Confirmed to be Due to Pseudomonas Aeruginosa (PRISM-MDR)," *ClinicalTrials.gov*, Aug. 28, 2019. https://clinicaltrials.gov/ct2/show/NCT03409679 (accessed Feb. 12, 2023).

[68] "Spexis reports solid safety and pharmacokinetics results from first-in-human study with inhaled murepavadin, a novel macrocycle compound," *Spexis*, Jan. 09, 2023. https://spexisbio.com/news-adhoc/news-detail/?newsid=2419969 (accessed Feb. 12, 2023).

[69] P. Fehlbaum *et al.*, "Structure-activity analysis of thanatin, a 21-residue inducible insect defense peptide with sequence homology to frog skin antimicrobial peptides.," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 93, no. 3, pp. 1221–1225, Feb. 1996, doi: 10.1073/pnas.93.3.1221.

[70] M. A. Sabokkhiz, A. Tanhaeian, and M. Mamarabadi, "Study on Antiviral Activity of Two Recombinant Antimicrobial Peptides Against Tobacco Mosaic Virus," *Probiotics & Antimicro. Prot.*, vol. 11, no. 4, pp. 1370–1378, Dec. 2019, doi: 10.1007/s12602-019-09539-4.

[71] H. A. Rothan, H. Bahrani, E. M. Shankar, N. A. Rahman, and R. Yusof, "Inhibitory effects of a peptide-fusion protein (Latarcin–PAP1–Thanatin) against chikungunya virus," *Antiviral Research*, vol. 108, pp. 173–180, Aug. 2014, doi: 10.1016/j.antiviral.2014.05.019.

[72] K.-E. Park *et al.*, "The roles of antimicrobial peptide, rip-thanatin, in the midgut of Riptortus pedestris," *Developmental & Comparative Immunology*, vol. 78, pp. 83–90, Jan. 2018, doi: 10.1016/j.dci.2017.09.009.

[73] J. Lee, W. H. Cha, and D.-W. Lee, "Multiple Precursor Proteins of Thanatin Isoforms, an Antimicrobial Peptide Associated With the Gut Symbiont of Riptortus pedestris," *Front. Microbiol.*, vol. 12, p. 796548, Jan. 2022, doi: 10.3389/fmicb.2021.796548.

[74] N. Mandard *et al.*, "Solution structure of thanatin, a potent bactericidal and fungicidal insect peptide, determined from proton two-dimensional nuclear magnetic resonance data," *Eur J Biochem*, vol. 256, no. 2, pp. 404–410, Sep. 1998, doi: 10.1046/j.1432-1327.1998.2560404.x.

[75] N. Morikawa, K. Hagiwara, and T. Nakajima, "Brevinin-1 and -2, unique antimicrobial peptides from the skin of the frog, Rana brevipoda porsa," *Biochemical and Biophysical Research Communications*, vol. 189, no. 1, pp. 184–190, Oct. 1992, doi: 10.1016/0006-291X(92)91542-X.

[76] A. Aumelas *et al.*, “Synthesis and Solution Structure of the Antimicrobial Peptide Protegrin-1,” *Eur J Biochem*, vol. 237, no. 3, pp. 575–583, May 1996, doi: 10.1111/j.1432-1033.1996.0575p.x.

[77] H. Tamamura *et al.*, “A comparative study of the solution structures of tachyplesin I and a novel anti-HIV synthetic peptide, T22 ([Tyr5,12, Lys7]-polyphemusin II), determined by nuclear magnetic resonance,” *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, vol. 1163, no. 2, pp. 209–216, May 1993, doi: 10.1016/0167-4838(93)90183-R.

[78] S. Taguchi, K. Kuwasako, A. Suenaga, M. Okada, and H. Momose, “Functional Mapping against Escherichia coli for the Broad-Spectrum Antimicrobial Peptide, Thanatin, Based on an In Vivo Monitoring Assay System,” *Journal of Biochemistry*, vol. 128, no. 5, pp. 745–754, Nov. 2000, doi: 10.1093/oxfordjournals.jbchem.a022811.

[79] M.-K. Lee, L.-N. Cha, S.-H. Lee, and K.-S. Hahm, “Role of Amino Acid Residues within the Disulfide Loop of Thanatin, a Potent Antibiotic Peptide,” *BMB Reports*, vol. 35, no. 3, pp. 291–296, May 2002, doi: 10.5483/BMBRep.2002.35.3.291.

[80] T. Imamura *et al.*, “NMR based structure–activity relationship analysis of an antimicrobial peptide, thanatin, engineered by site-specific chemical modification: Activity improvement and spectrum alteration,” *Biochemical and Biophysical Research Communications*, vol. 369, no. 2, pp. 609–615, May 2008, doi: 10.1016/j.bbrc.2008.02.057.

[81] B. Ma *et al.*, “The Disulfide Bond of the Peptide Thanatin Is Dispensible for Its Antimicrobial Activity *In Vivo* and *In Vitro*,” *Antimicrob Agents Chemother*, vol. 60, no. 7, pp. 4283–4289, Jul. 2016, doi: 10.1128/AAC.00041-16.

[82] B. Manta, D. Boyd, and M. Berkmen, “Disulfide Bond Formation in the Periplasm of *Escherichia coli*,” *EcoSal Plus*, vol. 8, no. 2, p. ecosalplus.ESP-0012-2018, Feb. 2019, doi: 10.1128/ecosalplus.ESP-0012-2018.

[83] S. Sinha, L. Zheng, Y. Mu, W. J. Ng, and S. Bhattacharjya, “Structure and Interactions of A Host Defense Antimicrobial Peptide Thanatin in Lipopolysaccharide Micelles Reveal Mechanism of Bacterial Cell Agglutination,” *Sci Rep*, vol. 7, no. 1, p. 17795, Dec. 2017, doi: 10.1038/s41598-017-18102-6.

[84] G. Wu *et al.*, “Effects of Cations and PH on Antimicrobial Activity of Thanatin and s-Thanatin Against Escherichia coli ATCC25922 and B. subtilis ATCC 21332,” *Curr Microbiol*, vol. 57, no. 6, pp. 552–557, Dec. 2008, doi: 10.1007/s00284-008-9241-6.

[85] G. Wu *et al.*, “Interaction of antimicrobial peptide s-thanatin with lipopolysaccharide in vitro and in an experimental mouse model of septic shock caused by a multidrug-resistant clinical isolate of Escherichia coli,” *International Journal of Antimicrobial Agents*, vol. 35, no. 3, pp. 250–254, Mar. 2010, doi: 10.1016/j.ijantimicag.2009.11.009.

[86] G. Wu *et al.*, “Selective toxicity of antimicrobial peptide S-thanatin on bacteria,” *Peptides*, vol. 31, no. 9, pp. 1669–1673, Sep. 2010, doi: 10.1016/j.peptides.2010.06.009.

[87] G. Wu *et al.*, “Membrane aggregation and perturbation induced by antimicrobial peptide of S-thanatin,” *Biochemical and Biophysical Research Communications*, vol. 395, no. 1, pp. 31–35, Apr. 2010, doi: 10.1016/j.bbrc.2010.03.107.

[88] G. Wu *et al.*, “The activity of antimicrobial peptide S-thanatin is independent on multidrug-resistant spectrum of bacteria,” *Peptides*, vol. 32, no. 6, pp. 1139–1145, Jun. 2011, doi: 10.1016/j.peptides.2011.03.019.

[89] G. Wu *et al.*, “Protective effects of antimicrobial peptide S-thanatin against endotoxic shock in mice introduced by LPS,” *Peptides*, vol. 32, no. 2, pp. 353–357, Feb. 2011, doi: 10.1016/j.peptides.2010.10.029.

[90] G. Wu *et al.*, “Application of S-thanatin, an antimicrobial peptide derived from thanatin, in mouse model of Klebsiella pneumoniae infection,” *Peptides*, vol. 45, pp. 73–77, Jul. 2013, doi: 10.1016/j.peptides.2013.04.012.

[91] Z. Hou *et al.*, “Underlying Mechanism of In vivo and In vitro Activity of C-terminal–amidated Thanatin Against Clinical Isolates of Extended-Spectrum β-lactamase–Producing Escherichia coli,” *The Journal of Infectious Diseases*, vol. 203, no. 2, pp. 273–282, Jan. 2011, doi: 10.1093/infdis/jiq029.

[92] B. W. Simpson and M. S. Trent, “Pushing the envelope: LPS modifications and their consequences,” *Nat Rev Microbiol*, vol. 17, no. 7, pp. 403–416, Jul. 2019, doi: 10.1038/s41579-019-0201-x.

[93] B. Ma *et al.*, “The antimicrobial peptide thanatin disrupts the bacterial outer membrane and inactivates the NDM-1 metallo-β-lactamase,” *Nat Commun*, vol. 10, no. 1, p. 3517, Dec. 2019, doi: 10.1038/s41467-019-11503-3.

[94] W. Hongbiao, N. Baolong, H. Lihua, S. Weifeng, and M. Zhiqi, “Biological activities of cecropin B-thanatin hybrid peptides: Thanatin hybrid peptides,” *The Journal of Peptide Research*, vol. 66, no. 6, pp. 382–386, Dec. 2008, doi: 10.1111/j.1399-3011.2005.00299.x.

[95] L. N. Wang, B. Yu, G. Q. Han, J. He, and D. W. Chen, "Design, expression and characterization of recombinant hybrid peptide Attacin-Thanatin in Escherichia coli," *Mol Biol Rep*, vol. 37, no. 7, pp. 3495–3501, Oct. 2010, doi: 10.1007/s11033-009-9942-3.

[96] X. Jiang *et al.*, "Design and activity study of a melittin–thanatin hybrid peptide," *AMB Expr*, vol. 9, no. 1, p. 14, Dec. 2019, doi: 10.1186/s13568-019-0739-z.

[97] X. Feng *et al.*, "Recombinant expression, purification, and antimicrobial activity of a novel hybrid antimicrobial peptide LFT33," *Appl Microbiol Biotechnol*, vol. 95, no. 5, pp. 1191–1198, Sep. 2012, doi: 10.1007/s00253-011-3816-z.

[98] X. Zhao and O. P. Kuipers, "Nisin- and Ripcin-Derived Hybrid Lanthipeptides Display Selective Antimicrobial Activity against *Staphylococcus aureus*," *ACS Synth. Biol.*, vol. 10, no. 7, pp. 1703–1714, Jul. 2021, doi: 10.1021/acssynbio.1c00080.

[99] H. Derakhshankhah and S. Jafari, "Cell penetrating peptides: A concise review with emphasis on biomedical applications," *Biomedicine & Pharmacotherapy*, vol. 108, pp. 1090–1096, Dec. 2018, doi: 10.1016/j.biopha.2018.09.097.

[100] Q. Chu *et al.*, "Regulation of the ER stress response by a mitochondrial microprotein," *Nat Commun*, vol. 10, no. 1, p. 4883, Oct. 2019, doi: 10.1038/s41467-019-12816-z.

[101] B. Miller *et al.*, "Mitochondrial DNA variation in Alzheimer's disease reveals a unique microprotein called SHMOOSE," *Mol Psychiatry*, Sep. 2022, doi: 10.1038/s41380-022-01769-3.

[102] C. Lee *et al.*, "The Mitochondrial-Derived Peptide MOTS-c Promotes Metabolic Homeostasis and Reduces Obesity and Insulin Resistance," *Cell Metabolism*, vol. 21, no. 3, pp. 443–454, Mar. 2015, doi: 10.1016/j.cmet.2015.02.009.

[103] M. A. Hallen *et al.*, "OSPREY 3.0: Open-source protein redesign for you, with powerful new features," *J. Comput. Chem.*, vol. 39, no. 30, pp. 2494–2507, Nov. 2018, doi: 10.1002/jcc.25522.

[104] S. F. AltschuP, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic Local Alignment Search Tool," *JMB*, vol. 215, no. 3, pp. 403–410, May 1990, doi: https://doi.org/10.1016/S0022-2836(05)80360-2.

[105] F. Sievers *et al.*, "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega," *Mol Syst Biol*, vol. 7, no. 1, p. 539, Jan. 2011, doi: 10.1038/msb.2011.75.

[106] G. E. Crooks, G. Hon, J.-M. Chandonia, and S. E. Brenner, "WebLogo: A Sequence Logo Generator," *Genome Res.*, vol. 14, no. 6, pp. 1188–1190, Jun. 2004, doi: https://10.1101/gr.849004.

[107] Clinical and Laboratory Standards Institute, *Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically*, 11th ed., vol. 2, 38 vols. 950 West Valley Road, Suite 2500, Wayne, Pennsylvania 19087, USA: Clinical and Laboratory Standards Institute, 2018.

[108] R. Fernández-Piñar, A. Lo Sciuto, A. Rossi, S. Ranucci, A. Bragonzi, and F. Imperi, "In vitro and in vivo screening for novel essential cell-envelope proteins in Pseudomonas aeruginosa," *Sci Rep*, vol. 5, no. 1, p. 17593, Dec. 2015, doi: 10.1038/srep17593.

[109] M. Bollati *et al.*, "Crystal structure of LptH, the periplasmic component of the lipopolysaccharide transport machinery from *Pseudomonas aeruginosa*," *FEBS J*, vol. 282, no. 10, pp. 1980–1997, May 2015, doi: 10.1111/febs.13254.

[110] R. Scala *et al.*, "Mutational analysis of the essential lipopolysaccharide-transport protein LptH of Pseudomonas aeruginosa to uncover critical oligomerization sites," *Sci Rep*, vol. 10, no. 1, p. 11276, Dec. 2020, doi: 10.1038/s41598-020-68054-7.

[111] J. Jumper *et al.*, "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol. 596, no. 7873, pp. 583–589, Aug. 2021, doi: 10.1038/s41586-021-03819-2.

[112] R. Evans *et al.*, "Protein complex prediction with AlphaFold-Multimer," Bioinformatics, preprint, Oct. 2021. doi: 10.1101/2021.10.04.463034.

[113] A. K. Duwe, C. A. Rupar, G. B. Horsman, and S. I. Vas, "In vitro cytotoxicity and antibiotic activity of polymyxin B nonapeptide," *Antimicrob Agents Chemother*, vol. 30, no. 2, pp. 340–341, Aug. 1986, doi: 10.1128/AAC.30.2.340.

[114] D. Ibrahim, J.-F. Jabbour, and S. S. Kanj, "Current choices of antibiotic treatment for Pseudomonas aeruginosa infections," *Current Opinion in Infectious Diseases*, vol. 33, no. 6, pp. 464–473, Dec. 2020, doi: 10.1097/QCO.0000000000000677.

[115] O. Trott and A. J. Olson, "AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *J. Comput. Chem.*, p. NA-NA, 2009, doi: 10.1002/jcc.21334.

[116] G. C. P. van Zundert *et al.*, "The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes," *Journal of Molecular Biology*, vol. 428, no. 4, pp. 720–725, Feb. 2016, doi: 10.1016/j.jmb.2015.09.014.

[117] L. A. Romanenko, M. Uchino, E. Falsen, G. M. Frolova, N. V. Zhukova, and V. V. Mikhailov, "Pseudomonas pachastrellae sp. nov., isolated from a marine sponge," *International Journal of Systematic and Evolutionary Microbiology*, vol. 55, no. 2, pp. 919–924, Mar. 2005, doi: 10.1099/ijs.0.63176-0.

[118] S. I. Paul, Md. M. Rahman, M. A. Salam, Md. A. R. Khan, and Md. T. Islam, "Identification of marine sponge-associated bacteria of the Saint Martin's island of the Bay of Bengal emphasizing on the prevention of motile Aeromonas septicemia in Labeo rohita," *Aquaculture*, vol. 545, p. 737156, Dec. 2021, doi: 10.1016/j.aquaculture.2021.737156.

[119] Y. S. Anteneh, Q. Yang, M. H. Brown, and C. M. M. Franco, "Antimicrobial Activities of Marine Sponge-Associated Bacteria," *Microorganisms*, vol. 9, no. 1, p. 171, Jan. 2021, doi: 10.3390/microorganisms9010171.

[120] Riyanti *et al.*, "Selection of sponge-associated bacteria with high potential for the production of antibacterial compounds," *Sci Rep*, vol. 10, no. 1, p. 19614, Nov. 2020, doi: 10.1038/s41598-020-76256-2.

[121] A. Falanga *et al.*, "Marine Antimicrobial Peptides: Nature Provides Templates for the Design of Novel Compounds against Pathogenic Bacteria," *IJMS*, vol. 17, no. 5, p. 785, May 2016, doi: 10.3390/ijms17050785.

[122] World Health Organization, "Hepatitis C," *Hepatitis C*, Jun. 24, 2022. https://www.who.int/news-room/fact-sheets/detail/hepatitis-c (accessed Jan. 29, 2023).

[123] S. D. C. Griffin *et al.*, "The p7 protein of hepatitis C virus forms an ion channel that is blocked by the antiviral drug, Amantadine," *FEBS Letters*, vol. 535, no. 1–3, pp. 34–38, Jan. 2003, doi: 10.1016/S0014-5793(02)03851-6.

[124] A. L. Wozniak *et al.*, "Intracellular Proton Conductance of the Hepatitis C Virus p7 Protein and Its Contribution to Infectious Virus Production," *PLoS Pathog*, vol. 6, no. 9, p. e1001087, Sep. 2010, doi: 10.1371/journal.ppat.1001087.

[125] A. Premkumar, L. Wilson, G. D. Ewart, and P. W. Gage, “Cation-selective ion channels formed by p7 of hepatitis C virus are blocked by hexamethylene amiloride,” *FEBS Letters*, vol. 557, no. 1–3, pp. 99–103, Jan. 2004, doi: 10.1016/S0014-5793(03)01453-4.

[126] G. Haqshenas, J. M. Mackenzie, X. Dong, and E. J. Gowans, “Hepatitis C virus p7 protein is localized in the endoplasmic reticulum when it is encoded by a replication-competent genome,” *Journal of General Virology*, vol. 88, no. 1, pp. 134–142, Jan. 2007, doi: 10.1099/vir.0.82049-0.

[127] E. Steinmann, F. Penin, S. Kallis, A. H. Patel, R. Bartenschlager, and T. Pietschmann, “Hepatitis C Virus p7 Protein Is Crucial for Assembly and Release of Infectious Virions,” *PLoS Pathog*, vol. 3, no. 7, p. e103, 2007, doi: 10.1371/journal.ppat.0030103.

[128] S. Griffin, C. StGelais, A. M. Owsianka, A. H. Patel, D. Rowlands, and M. Harris, “Genotype-dependent sensitivity of hepatitis C virus to inhibitors of the p7 ion channel,” *Hepatology*, vol. 48, no. 6, pp. 1779–1790, Dec. 2008, doi: 10.1002/hep.22555.

[129] P. Luik *et al.*, “The 3-dimensional structure of a hepatitis C virus p7 ion channel by electron microscopy,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 31, pp. 12712–12716, Aug. 2009, doi: 10.1073/pnas.0905966106.

[130] B. OuYang *et al.*, “Unusual architecture of the p7 channel from hepatitis C virus,” *Nature*, vol. 498, no. 7455, pp. 521–525, Jun. 2013, doi: 10.1038/nature12283.

[131] W. Chen, J. Dev, J. Mezhyrova, L. Pan, A. Piai, and J. J. Chou, “The Unusual Transmembrane Partition of the Hexameric Channel of the Hepatitis C Virus,” *Structure*, vol. 26, no. 4, pp. 627-634.e4, Apr. 2018, doi: 10.1016/j.str.2018.02.011.

[132] D. Clarke *et al.*, “Evidence for the Formation of a Heptameric Ion Channel Complex by the Hepatitis C Virus P7 Protein *in Vitro*,” *J. Biol. Chem.*, vol. 281, no. 48, pp. 37057–37068, Dec. 2006, doi: 10.1074/jbc.M602434200.

[133] C. StGelais *et al.*, “Determinants of Hepatitis C Virus p7 Ion Channel Function and Drug Sensitivity Identified In Vitro,” *Journal of Virology*, vol. 83, no. 16, pp. 7970–7981, Aug. 2009, doi: 10.1128/JVI.00521-09.

[134] T. L. Foster *et al.*, “Resistance mutations define specific antiviral effects for inhibitors of the hepatitis C virus p7 ion channel,” *Hepatology*, vol. 54, no. 1, pp. 79–90, Jul. 2011, doi: 10.1002/hep.24371.

[135] T. L. Foster *et al.*, "Structure-guided design affirms inhibitors of hepatitis C virus p7 as a viable class of antivirals targeting virion release: Hepatology, Vol. 00, No. 0, 2013," *Hepatology*, vol. 59, no. 2, pp. 408–422, Feb. 2014, doi: 10.1002/hep.26685.

[136] B. P. Oestringer *et al.*, "Re-evaluating the p7 viroporin structure," *Nature*, vol. 562, no. 7727, p. E8, Oct. 2018.

[137] C. F. Chew, R. Vijayan, J. Chang, N. Zitzmann, and P. C. Biggin, "Determination of Pore-Lining Residues in the Hepatitis C Virus p7 Protein," *Biophysical Journal*, vol. 96, no. 2, pp. L10–L12, Jan. 2009, doi: 10.1016/j.bpj.2008.10.004.

[138] W. Chen, B. OuYang, and J. J. Chou, "Critical Effect of the Detergent:Protein Ratio on the Formation of the Hepatitis C Virus p7 Channel," *Biochemistry*, p. acs.biochem.9b00636, Sep. 2019, doi: 10.1021/acs.biochem.9b00636.

[139] C. Stgelais, T. Tuthill, D. Clarke, D. Rowlands, M. Harris, and S. Griffin, "Inhibition of hepatitis C virus p7 membrane channels in a liposome-based assay system," *Antiviral Research*, vol. 76, no. 1, pp. 48–58, Oct. 2007, doi: 10.1016/j.antiviral.2007.05.001.

[140] F. Zambrano, S. Fleischer, and B. Fleischer, "Lipid composition of the golgi apparatus of rat kidney and liver in comparison with other subcellular organelles," *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism*, vol. 380, no. 3, pp. 357–369, Mar. 1975, doi: 10.1016/0005-2760(75)90104-6.

[141] J. W. Martin, A. K. Yan, C. Bailey-Kellogg, P. Zhou, and B. R. Donald, "A graphical method for analyzing distance restraints using residual dipolar couplings for structure determination of symmetric protein homo-oligomers: Structure Determination of Symmetric Homo-oligomers," *Protein Science*, vol. 20, no. 6, pp. 970–985, Jun. 2011, doi: 10.1002/pro.620.

[142] J. W. Martin, A. K. Yan, C. Bailey-kellogg, P. Zhou, and B. R. Donald, "A Geometric Arrangement Algorithm for Structure Determination of Symmetric Protein Homo-Oligomers from NOEs and RDCs," *Journal of Computational Biology*, vol. 18, no. 11, pp. 1507–1523, Nov. 2011, doi: 10.1089/cmb.2011.0173.

[143] J. W. Martin, P. Zhou, and B. R. Donald, "Systematic solution to homo-oligomeric structures determined by NMR: Systematic Solution to Homo-oligomers," *Proteins*, vol. 83, no. 4, pp. 651–661, Apr. 2015, doi: 10.1002/prot.24768.

[144] I. G. Denisov and S. G. Sligar, “Nanodiscs for structural and functional studies of membrane proteins,” *Nat Struct Mol Biol*, vol. 23, no. 6, pp. 481–486, Jun. 2016, doi: 10.1038/nsmb.3195.

[145] F. Hagn, M. Etzkorn, T. Raschle, and G. Wagner, “Optimized Phospholipid Bilayer Nanodiscs Facilitate High-Resolution Structure Determination of Membrane Proteins,” *J. Am. Chem. Soc.*, vol. 135, no. 5, pp. 1919–1925, Feb. 2013, doi: 10.1021/ja310901f.

[146] M. L. Nasr *et al.*, “Covalently circularized nanodiscs for studying membrane proteins and viral entry,” *Nat Methods*, vol. 14, no. 1, pp. 49–52, Jan. 2017, doi: 10.1038/nmeth.4079.

[147] T. J. Knowles, R. Finka, C. Smith, Y.-P. Lin, T. Dafforn, and M. Overduin, “Membrane Proteins Solubilized Intact in Lipid Containing Nanoparticles Bounded by Styrene Maleic Acid Copolymer,” *J. Am. Chem. Soc.*, vol. 131, no. 22, pp. 7484–7485, Jun. 2009, doi: 10.1021/ja810046q.

[148] M. A. Herzik, M. Wu, and G. C. Lander, “High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM,” *Nat Commun*, vol. 10, no. 1, p. 1032, Mar. 2019, doi: 10.1038/s41467-019-08991-8.

# Biography

Kelly Huynh graduated from Bridgewater-Raynham Regional High School in Bridgewater, Massachusetts in 2011. She then went on to attend Boston University in Boston, Massachusetts, where she received a Bachelor of Arts in Biochemistry and Molecular Biology and a Master of Arts in Biotechnology in 2015. During her time at Boston University, Kelly completed her thesis work under the mentorship of Dr. Karen N. Allen. After working in a biochemistry department in pharmaceutical industry in Cambridge, Massachusetts for two years, Kelly matriculated into the Biochemistry Ph.D. program at Duke University in Durham, North Carolina in 2017. In 2018, Kelly joined the laboratory of Dr. Bruce R. Donald and embarked on a close collaboration with Dr. Pei Zhou for her dissertation. During her time at Duke University, Kelly has been a mentor for first-generation and low-income undergraduates and first-year graduate students. Kelly was also a member of the Duke Scholars in Molecular Medicine program from 2021 – 2022. During her time at Duke University, Kelly has contributed to several publications, including "Understanding the characteristics of nonspecific binding of drug-like compounds to canonical stem-loop RNAs and their implications for functional cellular assays." in RNA in 2021 and "Discovery, characterization, and redesign of potent antimicrobial thanatin orthologs from *Chinavia ubica* and *Murgantia histrionica* targeting *E. coli* LptA.", which the manuscript is currently in preparation.