Spring 2017

Recombinant Expression and Potential Autocatalysis of *Aedes aegypti* Trypsin-like Serine Proteases (AaSPII and AaSPIV)

Diane Francine Eilerts
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_theses

Recommended Citation

Eilerts, Diane Francine, "Recombinant Expression and Potential Autocatalysis of *Aedes aegypti* Trypsin-like Serine Proteases (AaSPII and AaSPIV)" (2017). *Master's Theses*. 4795.
DOI: <https://doi.org/10.31979/etd.5tk2-fsj5>
https://scholarworks.sjsu.edu/etd_theses/4795

This Thesis is brought to you for free and open access by the Master's Theses and Graduate Research at SJSU ScholarWorks. It has been accepted for inclusion in Master's Theses by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.

RECOMBINANT EXPRESSION AND POTENTIAL AUTOCATALYSIS OF
Aedes Aegypti TRYPSIN-LIKE SERINE PROTEASES (AASPII AND AASPIV)

A Thesis

Presented to

The Faculty of the Department of Chemistry

San José State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Diane Eilerts

May 2017

© 2017

Diane Eilerts

ALL RIGHTS RESERVED

The Designated Thesis Committee Approves the Thesis Titled

RECOMBINANT EXPRESSION AND POTENTIAL AUTOCATALYSIS OF
AEDES AEGYPTI TRYPSIN-LIKE SERINE PROTEASES (AASP II AND AASP IV)

by

Diane Eilerts

APPROVED FOR THE DEPARTMENT OF CHEMISTRY

SAN JOSÉ STATE UNIVERSITY

May 2017

Alberto A. Rascón, Jr., Ph.D.	Department of Chemistry
-------------------------------	-------------------------

Laura Miller Conrad, Ph.D.	Department of Chemistry
----------------------------	-------------------------

Marc d'Alarcao, Ph.D.	Department of Chemistry
-----------------------	-------------------------

ABSTRACT

RECOMBINANT EXPRESSION AND POTENTIAL AUTOCATALYSIS OF *Aedes Aegypti* TRYPSIN-LIKE SERINE PROTEASES (AaSPII AND AaSPIV)

by Diane Eilerts

Aedes aegypti mosquitoes can be found globally in tropical and subtropical urban areas and spread Zika, Dengue fever, yellow fever, and Chikungunya viruses. Current vector control methods are limited and nonspecific. The female *Ae. aegypti* mosquito uses blood meal proteins to obtain nutrients required for oogenesis; inhibition of the midgut trypsin-like serine proteases responsible for blood meal digestion may provide a novel method of vector control. *Ae. aegypti* blood meal digestion is complex and the role of uncharacterized serine proteases in blood digestion is unclear; specifically, a group of trypsin-like serine proteases (AaSPII–V) is expressed at constant levels before and following *Ae. aegypti* blood meal acquisition. This research focuses on the *in vitro* biochemical study of two specific *Ae. aegypti* trypsin-like serine proteases (AaSPII and AaSPIV) in order to gain further understanding of their role in blood meal digestion. The approach involved the successful cloning and bacterial expression of these soluble, recombinant proteases. Results from attempts to purify these proteases were unsuccessful but indicative of potential autocatalytic and autodigestive behavior. Future studies will focus on obtaining purified recombinant proteases for further study. The study of AaSPII and AaSPIV, as well as other midgut *Ae. aegypti* proteases, will aid in understanding the overall role proteases play in blood meal digestion and may eventually allow for the development of mosquito-specific enzyme inhibitors.

ACKNOWLEDGMENTS

First and foremost, I would like to thank Prof. Alberto A. Rascón Jr., my research advisor, for the opportunity to perform this research in his laboratory and for the expert mentorship and guidance I received throughout my research endeavors. While working with Prof. Rascón, I gained invaluable experience and learned both technical and practical skills to help me succeed in the future as a research scientist. I would also like to extend my gratitude to the members of my graduate committee, Prof. Laura Miller Conrad and Prof. Marc d'Alarcao, for their support and assistance in developing this thesis. I also thank my friends and colleagues at San Jose State University and in the Rascón Lab for the scientific discourse, encouragement, and laughter we've shared.

For their constant love and support, I want to thank my family and friends. My mother and father, Valerie and Gregg Eilerts, have provided a lifetime of support and encouragement and prepared me for a successful future. I thank my brother, Tom Eilerts, for always making me laugh and for his endearing attempts to help me overcome research struggles via jelly donut metaphors. I thank my sister, Renee Eilerts, for always welcoming me in her home and making sure I remembered to occasionally have fun outside of the lab (Go Sharks!). I thank my best friend, Chelsea Carman, for listening and supporting me through both my undergraduate and graduate education.

Lastly, I want to extend my deepest gratitude to my partner, Morgan Hoffman, for his unwavering patience, understanding, love, and support throughout my educational endeavors.

TABLE OF CONTENTS

List of Tables	viii
List of Figures	ix
Chapter 1: Introduction	1
1.1 Role of <i>Aedes aegypti</i> in Viral Pathogen Transmission.....	1
1.1.1 Zika	3
1.1.2 Dengue	7
1.1.3 Yellow Fever.....	9
1.1.4 Chikungunya	12
1.2 Vector Control Strategies.....	14
1.2.1 Current Methods	14
1.2.2 <i>Ae. aegypti</i> Midgut Serine Proteases	17
1.3 Recombinant Protein Expression in <i>Escherichia coli</i>	23
1.4 Protein Purification Techniques.....	24
1.5 Protease Characterization Techniques	26
1.6 Specific Research Aims	27
Chapter 2: Methods.....	28
2.1 Preparation of Protease Plasmid Constructs	28
2.1.1 General Cloning Strategy.....	28
2.1.2 pET28a Plasmid Constructs	34
2.1.3 Enterokinase (EK)-Pseudo Propeptide Plasmid Constructs	36
2.1.4 pMAL-c5E Plasmid Constructs	37
2.1.5 pET29b Plasmid Constructs.....	38
2.1.6 Site-Directed Mutagenesis	39
2.2 Recombinant Protein Expression.....	42
2.2.1 Induction by IPTG	42
2.2.2 Auto-Induction.....	43
2.2.3 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).....	44
2.2.4 His-Stain	45
2.2.5 Western blotting.....	46
2.3 Protein Purification	48
2.3.1 Wheat Starch Purification for MBP-Tagged Proteins	48
2.3.2 Fast Protein Liquid Chromatography.....	49
2.4 Qualitative Activity Assay.....	51
Chapter 3: Results and Discussion.....	52
3.1 Molecular Cloning	52
3.2 Recombinant Protein Expression.....	57
3.3 Initial Purification Attempts and Qualitative Activity Assay	77
3.4 Alternative Purification Approaches.....	89

Chapter 4: Conclusions	98
4.1 AaSPIV Genome Sequence	98
4.2 Recombinant Protein Expression	98
4.3 Autocatalytic Activity of AaSPII and AaSPIV	100
4.4 Future Studies	101
References	106

LIST OF TABLES

Table 1. Primers used for PCR amplification and cloning of recombinant forms of the full length zymogen (with leader), mature form (no leader), MBP-fusion, EK-pseudo propeptide, and mutant (Ser to Ala and Lys to Ala) AaSPII	29
Table 2. Primers used for PCR amplification and cloning of recombinant forms of the full length zymogen (with leader), mature form (no leader), MBP-fusion, EK-pseudo propeptide, and mutant (Ser to Ala and Lys to Ala) AaSPIV	30
Table 3. Cycling Parameters for PCR amplification of recombinant forms of the full length zymogen (with leader), mature form (no leader), MBP-fusion, EK-pseudo propeptide, and mutant (Ser to Ala and Lys to Ala) AaSPII and AaSPIV	31
Table 4. General PCR Reaction for Site-Directed Mutagenesis Reaction – Single Base Point Mutation	40
Table 5. Thermal Cycler Parameters for Site-Directed Mutagenesis Reaction – Single Base Point Mutation	40
Table 6. Summary of plasmid constructs developed for expression of recombinant forms of the full length zymogen (with leader), mature form (no leader), MBP-fusion, EK-pseudo propeptide, and mutant (Ser to Ala and Lys to Ala) AaSPII and AaSPIV	41
Table 7. Summary of purification columns and conditions used indicating the type of resin, whether anion or cation exchange, and buffers used.....	51

LIST OF FIGURES

Figure 1. The <i>Ae. aegypti</i> life cycle showing aquatic and terrestrial stages	2
Figure 2. Global Distribution of Zika virus transmission	5
Figure 3. Distribution of Zika virus in the United States.....	6
Figure 4. Distribution of Dengue in the United States.....	8
Figure 5. Distribution of yellow fever in Africa	10
Figure 6. Distribution of yellow fever in South America	11
Figure 7. Global distribution of Chikungunya.....	13
Figure 8. Catalytic Triad	18
Figure 9. Serine protease enzyme mechanism	20
Figure 10. General molecular cloning procedure.	34
Figure 11. Regulation of protein expression with vectors utilizing the T7 expression system	35
Figure 12. Amino acid sequence of the zymogen polypeptide for AaS ^{PII} and AaS ^{PIV}	37
Figure 13. MBP–His ₆ –tag–Protein fusion expected product.....	38
Figure 14. 1% Agarose gel of AaS ^{PIV} PCR products using primers designed for cloning into the pET28a vector.....	52
Figure 15. 1% Agarose gel of digested and dephosphorylated pET28a vector	53
Figure 16. 1% Agarose gel of Colony PCR reactions showing positive hits of the AaS ^{PIV} gene.....	54
Figure 17. 4-12% BIS-TRIS Gel analysis of samples collected from a 64 h growth experiment of AaS ^{PII} -NL-pET28a in SHuffle T7 <i>E. coli</i> cells grown at 10 °C using TB media	59
Figure 18. 4-12% BIS-TRIS Gel analysis of samples collected from a 64 h growth experiment of AaS ^{PIV} -pET28a in SHuffle T7 <i>E. coli</i> cells grown at 10 °C using TB media	60

Figure 19. 4-12% BIS-TRIS Gel analysis of samples collected from a 96 h growth experiment of AaS ^{PII} -NL-pET28a in SHuffle T7 <i>E. coli</i> cells grown at 10 °C using auto-induction.....	63
Figure 20. 4-12% BIS-TRIS Gel analysis of samples collected from a 96 h growth experiment of AaS ^{PIV} -pET28a in SHuffle T7 <i>E. coli</i> cells grown at 10 °C using auto-induction.....	64
Figure 21. 4-12% BIS-TRIS Gel analysis of samples collected from a 24 h growth experiment of EK-AaS ^{PII} -pET28a in SHuffle T7 <i>E. coli</i> cells grown at 30 °C using auto-induction.....	66
Figure 22. 4-12% BIS-TRIS Gel analysis of samples collected from a 72 h growth experiment of EK-AaS ^{PII} -pET28a in SHuffle T7 <i>E. coli</i> cells grown at 15 °C using auto-induction.....	67
Figure 23. Western blot analysis of samples collected from a 24 h growth experiment of EK-AaS ^{PII} -pET28a in SHuffle T7 <i>E. coli</i> cells grown at 30 °C using auto-induction.....	70
Figure 24. 4-12% BIS-TRIS Gel analysis of samples collected from a 24 h growth experiment of AaS ^{PII} -pMAL-c5E in SHuffle T7 <i>E. coli</i> cells grown at 15 °C using LB media	73
Figure 25. 4-12% BIS-TRIS Gel analysis of samples collected from a 48 h growth experiment of AaS ^{PIV} -pMAL-c5E in SHuffle T7 <i>E. coli</i> cells grown at 15 °C using LB media	74
Figure 26. 4-12% BIS-TRIS Gel analysis of soluble samples collected from a 77 h growth experiment of AaS ^{PIV} -pMAL-c5E in SHuffle T7 <i>E. coli</i> cells grown at 12 °C using TB media.....	75
Figure 27. 4-12% BIS-TRIS Gel analysis of fractions collected from wheat starch purification of MBP-AaS ^{PIV} recombinant protease expressed in SHuffle T7 cells grown at 15 °C for 48 h in LB media	77
Figure 28. Scheme showing cleavage of the MBP–His ₆ -tag–Protein fusion product following autoactivation, resulting in free His ₆ -tagged MBP and free active target protein.....	79
Figure 29. 4-12% BIS-TRIS Gel analysis of fractions collected from purification via FPLC using DEAE anion exchange column in 20 mM Tris-HCl, pH 8 buffer of MBP-AaS ^{PII} recombinant protease expressed in SHuffle T7 cells grown at 15 °C for 48 h in LB media	80

Figure 30. 4-12% BIS-TRIS Gel analysis of fractions collected from purification via FPLC using DEAE anion exchange column in 20 mM Tris-HCl, pH 8 buffer of MBP-AaSPIV recombinant protease expressed in SHuffle T7 cells grown at 15 °C for 48 h in LB media	81
Figure 31. Western blot analysis of fractions obtained from purification attempt via FPLC using DEAE anion exchange column in 20 mM Tris-HCl, pH 8 buffer of MBP-AaSPIV expressed in SHuffle T7 cells grown at 15 °C for 48 h in LB media	83
Figure 32. Overlay of Western blot (Figure 31) and complementary SDS-PAGE (Figure 30) performed on fractions obtained from purification attempt via FPLC using DEAE anion exchange column in 20 mM Tris-HCl, pH 8 buffer of AaSPIV-pMAL expressed in SHuffle T7 cells grown at 15 °C for 48 h in LB media	85
Figure 33. Qualitative enzyme activity assay reaction samples using BApNA to detect the presence of trypsin-like activity in fractions collected from purification via FPLC using DEAE anion exchange column in 20 mM Tris-HCl, pH 8 buffer of AaSPIV-pMAL expressed in SHuffle T7 cells grown at 15 °C for 48 h in LB media.....	87
Figure 34. 4-12% BIS-TRIS Gel analysis of fractions collected from purification via FPLC using a CM weak cation exchange column in 50 mM sodium acetate, pH 5.2 buffer of MBP-AaSPII expressed in SHuffle T7 cells grown at 15 °C for 48 h in LB media.....	90
Figure 35. 4-12% BIS-TRIS Gel analysis of fractions collected from purification via FPLC using a SP cation exchange column in 50 mM acetic acid, pH 4.3 buffer of MBP-AaSPIV expressed in SHuffle T7 cells grown at 15 °C for 48 h in LB media	91
Figure 36. 4-12% BIS-TRIS Gel analysis of fractions collected from purification via FPLC using DEAE anion exchange column in 20 mM Tris-HCl, pH 8 buffer of MBP-AaSPIV recombinant protease expressed in SHuffle T7 cells grown at 15 °C for 72 h in LB media and prepared with 1 mM PMSF added to the crude lysate preparation.....	93
Figure 37. Western blot analysis of fractions obtained from purification attempt via FPLC using DEAE anion exchange column in 20 mM Tris-HCl, pH 8 buffer of MBP-AaSPIV (with leader) expressed in SHuffle T7 cells grown at 15 °C for 72 h in LB media and prepared with 1 mM PMSF added to the crude lysate preparation	94

CHAPTER 1: INTRODUCTION

1.1 Role of *Aedes aegypti* in Viral Pathogen Transmission

Aedes aegypti mosquitoes have a complex life cycle with both aquatic and terrestrial stages (Figure 1).¹ Adult female mosquitoes lay eggs in the walls of water-filled containers above the water's surface.¹ Even small amounts of water are sufficient for attracting female mosquitoes and eggs can survive desiccation and remain viable through cold and dry conditions for several months.¹ Once the water level rises to cover the eggs, larvae emerge and feed on microorganisms present in the water. The larvae molt 3 times before transforming into pupae and continue to develop in the water.¹ Adult flying mosquitoes (imago) emerge from pupal skin and leave the water environment.¹ Both male and female mosquitoes feed on nectar, but only female mosquitoes require blood for egg production and after obtaining a blood meal will seek out water sources for an ideal environment to lay eggs.¹

Figure 1. The *Ae. aegypti* life cycle showing aquatic and terrestrial stages. Female mosquitoes lay eggs in the walls of water-filled containers. When eggs are covered with water, larvae emerge. Larvae eventually transform into pupae and develop into adult mosquitoes. Public domain image by Mariana Ruiz Villarreal.

Ae. aegypti mosquitoes act as vectors of the viral pathogens Zika, Dengue fever, yellow fever, and Chikungunya viruses, which can cause numerous human diseases.² When mosquitoes acquire a blood meal from an infected host, they too become infected; after an incubation period in which the pathogen can develop and replicate within mosquitoes, they can transmit these pathogens to a new host during later blood meal

acquisition.² Humans remain the primary amplifying host for such viruses.³ Several factors are thought to contribute to the anthropophilic nature of *Ae. aegypti* mosquitoes. Research suggests that the evolution of *Ae. aegypti* mosquitoes to preferentially feed on humans may be linked to increases in the expression and ligand-sensitivity of a specific odorant receptor for sulcatone, an odorant primarily associated with humans.⁴ Feeding on humans (as opposed to primates or other vertebrates) benefits mosquito fitness for several reasons: (1) humans are hairless, making it easier to obtain a blood meal, (2) humans have a tendency to live in large groups, making food sources abundant for mosquitoes, and (3) humans typically provide adequate environments for mosquito eggs by collecting water, storing water, and producing debris that collects water.⁴ Therefore, feeding preferentially on humans may provide *Ae. aegypti* mosquitoes with greater reproductive fitness. *Ae. aegypti* mosquitoes live and thrive in tropical, subtropical, and temperate urban and semi-urban areas.³ Due to its anthropophilic nature, tendency to bite multiple times before oogenesis completion, and propensity to live in highly populated areas, the *Ae. aegypti* mosquito is a highly efficient vector.³

1.1.1 Zika

The Zika flavivirus (ZIKV) is primarily transmitted by infected *Ae. aegypti* mosquitoes; however, sexual transmission of ZIKV has also been reported.^{5,6} There is currently no vaccine available but symptoms are typically mild, usually lasting for 2 to 7 days and consisting of fever, rash, conjunctivitis, pain, malaise, and headache.⁵ Additionally, Zika infection in pregnant women has recently been associated with various birth defects in offspring including microcephaly and Guillain-Barré syndrome.^{7,8} Microcephaly is a birth

defect causing a baby's head to be much smaller in size than expected.⁹ Microcephaly occurs when a baby's brain does not develop properly during pregnancy or stops growing after birth. Several problems can be associated with microcephaly, including seizures, developmental delays, intellectual disability, problems swallowing, hearing loss, and vision issues. Severity ranges but problems are typically lifelong and often life threatening.⁹ Guillain-Barré syndrome is a disorder in which the body's immune system attacks the peripheral nervous system; it is the most severe acute paralytic neuropathy and consists of several variants.¹⁰ Up to 30% of all cases manifest as severe Guillain-Barré syndrome with respiratory failure. Initial symptoms include weakness and tingling in extremities and can quickly worsen, leading to paralysis. While most people recover from Guillain-Barré syndrome, some people experience continued symptoms.¹⁰ There is no cure for microcephaly or for Guillain-Barré syndrome.^{9,10} Zika virus is endemic in several areas within Africa and Asia (Figure 2).⁵ More recently, locally acquired Zika has also been reported in the Americas and even in the continental United States in Florida and Texas (Figure 3).^{5,11}

Figure 2. Global Distribution of Zika virus transmission. Figure obtained from the Centers for Disease Control and Prevention.

Figure 3. Distribution of Zika virus in the United States. Laboratory-confirmed Zika virus disease cases by state as of January 11, 2017. Degree of shading indicates density of cases reported. Figure obtained from the Centers for Disease Control and Prevention.

1.1.2 Dengue

Dengue virus (DENV) is an RNA flavivirus consisting of five serotypes (DEN-1, DEN-2, DEN-3, DEN-4, and the recently discovered DEN-5) primarily transmitted by *Ae. aegypti* mosquitoes.^{3,12} Dengue fever, which can result from infection by DENV, causes flu-like symptoms (fever, headache, severe pain, rash, bleeding) and has potential to develop into severe dengue (or Dengue Hemorrhagic Fever), a lethal complication characterized by massive bleeding and shock.¹³ There is extensive genetic variability among the five serotypes, making vaccine development challenging.¹³ Because the virus has multiple serotypes, those who recover from dengue fever and gain immunity to a particular serotype can still fall victim to subsequent infection by another serotype; unfortunately, consecutive dengue infection by multiple virus serotypes increases the risk of severe dengue.¹³

Because dengue cases are severely underreported and often misdiagnosed, actual dengue infection incidence cannot be accurately determined.¹³ A 2012 study estimates that approximately 3.9 billion people in 128 countries are at risk of dengue infection.¹⁴ Furthermore, a 2013 estimate suggests that approximately 390 million dengue infections occur annually, with approximately 96 million manifesting clinical symptoms.¹⁵ Cases of imported dengue infections have been reported throughout Mexico and the United States; instances of locally acquired dengue confirmed by the CDC have occurred in Florida, Texas, Hawaii, and even New York.¹⁶ Increasing air travel and globalization will continue to impact the distribution of dengue throughout the United States, especially in southern border states like California, Arizona, and Texas (Figure 4).¹⁶

Figure 4. Distribution of Dengue in the United States. Shading indicates regions of imported and/or locally acquired dengue cases in the United States confirmed by the CDC between 2010 and 2014. Values indicate the number of imported dengue cases reported and locally acquired cases (parentheses) reported. Hashed area represents areas at risk based on 2013 vector distribution estimates. Reprinted with permission from Fredericks and Fernandez-Sesma¹⁶; DOI: <http://dx.doi.org/10.1016/j.aogh.2015.02.006> in accordance with Creative Commons user license: <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>.

1.1.3 Yellow Fever

The yellow fever flavivirus (YFV) is another viral pathogen transmitted by infected *Ae. aegypti* mosquitoes. Fever, headache, jaundice, pain, nausea, vomiting, and fatigue are symptoms of yellow fever infection.¹⁷ Yellow fever virus is endemic in tropical regions of Africa as well as in Central and South America (Figures 5 and 6).¹⁷ While most patients see symptoms disappear within 3 to 4 days, a small portion of those infected succumb to a second, more toxic phase following the initial recovery in which high fever, jaundice, and bleeding occur.¹⁷ Approximately half of the patients who enter the toxic phase will die within 7 to 10 days.¹⁷ Fortunately, a highly effective, safe, and affordable vaccine against the yellow fever virus is available and provides lifelong immunity to yellow fever disease.¹⁷

Figure 5. Distribution of yellow fever in Africa. Figure obtained from the Centers for Disease Control and Prevention.

Figure 6. Distribution of yellow fever in South America. Figure obtained from the Centers for Disease Control and Prevention.

1.1.4 Chikungunya

Chikungunya virus (CHIKV) is an alphavirus transmitted by infected *Ae. aegypti* mosquitoes; the disease shares many of the symptoms associated with dengue (fever, joint pain, headache, nausea, fatigue, and rash).¹⁸ There is no vaccine or antiviral treatment available for chikungunya so treatment focuses only on symptoms of the disease.¹⁸ Joint pain associated with chikungunya can be extremely debilitating and can persist for months or even years. The name “chikungunya” originates from Makonde (or Kimakonde) root word *kungunyala*, meaning “to become contorted,” and describes the stooped posture of those suffering from the associated extreme joint pain. Chikungunya has been identified in more than 60 countries within Asia, Africa, Europe, and North and South America (Figure 7).¹⁸ Many individuals living in areas with endemic chikungunya depend on manual labor to support themselves and their families. While patients usually recover from chikungunya infection, the temporary inability to work can have detrimental effects on these impoverished populations. Cases of imported chikungunya have been reported in Canada, Mexico, and the United States.¹⁸ Twelve locally acquired cases of chikungunya were reported in Florida in 2014 and one was reported in Texas in 2015.¹⁹

Figure 7. Global distribution of Chikungunya. Countries and territories in which locally acquired chikungunya cases have been reported as of April 22, 2016. Figure obtained from the Centers for Disease Control and Prevention.

1.2 Vector Control Strategies

Ae. aegypti mosquito populations are spreading worldwide and are predicted to continue to spread as temperatures rise due to climate change, creating more environments ideal for mosquitoes and increasing human exposure to the mosquito.^{20,21} With no effective treatment for Dengue, Zika, or Chikungunya infection, no vaccines available to combat these viruses, and a relentless increase in vector-borne disease cases reported globally, there is a vital need to control the disease through the only available method: vector control.^{3,15,22,23}

1.2.1 Current Methods

Current vector control strategies include environmental management, and biological and chemical methods.³ Environmental management techniques aim to alter controllable factors that contribute to the presence of available larval habitats for *Ae. aegypti* populations.³ As such, mosquito population prevalence is often correlated to poor water supply, sanitation, and waste disposal, which these environmental management techniques are designed to address.³ When properly applied, these methods can significantly reduce mosquito populations. However, these methods have obvious limitations: they are dependent on human behavior and awareness, requiring that all individuals of a population be knowledgeable of these practices, and can be ineffective in areas containing natural water reservoirs (tree or rock holes). Additionally, environmental management can be difficult to successfully implement in underdeveloped areas or in areas with limited resources.

Biological control of vector populations utilizes known ecological species interactions to decrease target mosquito populations by introducing predatory, competing, parasitic, or pathogenic organisms into the ecosystem.³ However, these methods are often impractical to employ in many urban settings and can have adverse effects on ecosystems.³ Recent studies suggest that release of genetically modified male *Ae. aegypti* mosquitoes carrying a homozygous dominant lethal gene can reduce mosquito populations; the transgenic males can compete with native male mosquitoes and procreate with female mosquitoes to produce offspring with an expressed lethal gene that causes death in embryonic or larval stages.²⁴ While initial studies appear promising, research has been limited and public opinion of the practice can be extremely controversial; there is also uncertainty in the future of research and implementation of this method due to issues associated with regulation.^{25,26} Until further research is conducted, it is unclear whether the use of genetically modified mosquitoes would be a viable option to control mosquito populations. Another biological method gaining in notoriety involves releasing *Ae. aegypti* mosquitoes that have been experimentally infected with specific strains of *Wolbachia* (a bacterial symbiont) to establish long-term infection in the mosquito population.²⁷ Infection by *Wolbachia* has been shown to reduce *Ae. aegypti* life span, egg viability, and blood feeding success in addition to increasing mosquito resistance to DENV infection.²⁷ While this method of vector control may have potential to positively impact human health, the effectiveness varies for various DENV serotypes and it would likely be insufficient in high-transmission areas.²⁷ Biological methods are typically expensive and impractical to implement at large scales and benefits

outweighing potential damage to ecosystems are often short-sighted since mosquito populations usually recover.

Chemical vector control methods depend on the use of insecticides, which can either target mosquito larvae (larvicides) or adult mosquitoes (adulticides). Insecticides act as neurotoxins, disturbing calcium homeostasis and other molecular mechanisms, which are not species-specific.²⁸ There are numerous types of insecticides, classified by chemical structure and mode of action: pyrethroids cause voltage-gated sodium channels to remain activated; organophosphates irreversibly inhibit the enzyme acetylcholinesterase; organochlorine insecticides can target voltage-gated sodium channels or receptors that respond to either glycine or to the neurotransmitter gamma-aminobutyric acid (GABA); neonicotinoids target nicotinic acetylcholine receptors.²⁸ While insecticides currently act as the most potent tool available for controlling mosquito populations, insecticides should only be used in emergencies or in situations in which no other methods would be beneficial. Since insecticides are not species-specific, they can be toxic to humans and other ecologically significant populations. Pesticide use has been known to have devastating off-target effects on other animals, including honeybees and other pollinators.^{29,30} Another concern associated with reliance on insecticides is the growing trend of mosquitoes exhibiting insecticide resistance. The increasing frequency of *Ae. aegypti* insecticide resistance necessitates research towards the development of novel vector control methods that do not drive mosquito evolution towards resistance.^{2,3,31-34} New methods to address vector control are

vitally needed due to increased pathogen drug resistance, lack of effective disease treatment or vaccine options, and increased insecticide resistance.²

1.2.2 Ae. aegypti Midgut Serine Proteases

A human blood meal is required by female *Ae. aegypti* mosquitoes to supply the nutrients necessary for metabolism and oogenesis.³⁵ Digestion of blood proteins occurs in the midgut via proteases: a dramatic increase in protease activity occurs in the *Ae. aegypti* midgut upon blood intake.^{36,37} Proteases are proteolytic enzymes that release an amino acid or peptide from a larger peptide or protein; endopeptidases act on internal peptide bonds while exopeptidases act on the peptide bonds of amino acids at the N- or C-terminus.³⁸ Within the *Ae. aegypti* midgut, serine proteases, in addition to exopeptidases, are responsible for the rapid digestion of blood and, thus, contribute to mosquito fecundity.³⁵ Inhibiting the activity of proteases necessary for blood meal digestion may be an effective way to manage mosquito populations and consequently, transmission of associated viral pathogens. Previous research confirms that reducing the percentage of blood in the female mosquito blood meal significantly decreases the amount of viable eggs produced.²² By effectively inhibiting the digestion of blood meal proteins with synthetic peptide analog inhibitors, the mosquito would be unable to render nutrients from the blood necessary for egg development, reducing the population of future mosquito generations.

Proteases have catalytic sites consisting of either amino acid or water nucleophiles.³⁹ Serine proteases are named for the nucleophilic serine residue in the enzyme active site.⁴⁰ The nucleophilicity of the serine hydroxyl group is dependent on aspartate and histidine

residues, which, along with serine, make up the catalytic triad or charge relay system (Ser-195, His-57, and Asp-102 according to bovine chymotrypsin numbering convention) (Figure 8).⁴⁰ Compression of the hydrogen bond between the histidine and aspartate residue results in a low-barrier hydrogen bond, increasing the pK_a of the histidine residue and allowing the histidine residue to act as an enhanced general base and hydrogen ion acceptor; the aspartate residue orientates the imidazolium ring of the histidine residue, making histidine a better proton acceptor and bringing serine and histidine into closer proximity.⁴¹ While the pK_a value of a serine hydroxyl group is generally too high for deprotonation to be favorable in physiological conditions, this hydrogen bonding network results in polarization of the hydroxyl group of the serine residue side chain, which can then be deprotonated to prevent formation of an unfavorable positive charge on the serine hydroxyl group and makes the serine residue an extremely powerful nucleophile.^{38,41,42}

Figure 8. Catalytic Triad. The charged aspartic acid side chain orientates the imidazolium ring of the histidine residue, which acts as a general base and can deprotonate the hydroxyl group on the serine residue, generating the alkoxide ion that can act as an extremely powerful nucleophile.

The mechanism of peptide hydrolysis (Figure 9) proceeds in the following steps, beginning with (1) substrate binding, followed by (2) a nucleophilic attack on the substrate carbonyl carbon of the peptide bond by the activated hydroxyl group of serine 195. This alters the geometry of atoms around the carbon, resulting in an unstable

tetrahedral intermediate with a negatively charged oxygen atom stabilized by the oxyanion hole (hydrogen bonding by backbone NH groups). Proton transfer from the histidine residue to the newly formed amino group (at the substrate cleavage site) leads to (3) collapse of the tetrahedral intermediate, producing the acyl-enzyme. The N-terminal region of the cleaved substrate can then (4) leave the enzyme active site and is then replaced by a water molecule. The water molecule attacks the carbonyl group and the histidine residue removes a proton, (5) forming a second tetrahedral intermediate. Collapse of the second tetrahedral intermediate leads to (6) formation and release of the carboxylic acid product. Afterwards, the enzyme catalyst is regenerated and ready for subsequent catalysis.⁴² This mechanism is common among all serine proteases, which include trypsin, chymotrypsin, elastase, and serine collagenase. Studies focusing on the *Ae. aegypti* mosquito have revealed that the major serine protease involved in the blood meal digestion process are trypsin-like enzymes.^{22,23}

Figure 9. Serine protease enzyme mechanism.

Blood meal digestion in *Ae. aegypti* mosquitoes occurs in a biphasic manner; midgut trypsin-like activity increases during the early phase of digestion, 0 to 6 h post blood meal (PBM), and the late phase of digestion shows a more dramatic increase 12-18 h PBM.^{22,43} Three serine proteases showing the highest expression levels in female *Ae. aegypti* were initially characterized: early trypsin (AaET), late trypsin (AaLT), and 5G1 (AaSPVI).^{43,44} Completion of the *Ae. aegypti* genome and subsequent bioinformatics analyses identified another abundant protease, AaSPVII, in addition to other midgut proteases.^{22,35,45} Of the hundreds of serine protease-like genes in the *Ae. aegypti* genome, 66 were identified as potential trypsins and just nine of these were found to be expressed in the midgut.^{35,45} Trypsin belongs to the serine protease family [subclan PA(S)] and is vital in *Ae. aegypti* blood meal digestion.^{38,46} Trypsins cleave peptide substrates on the C-terminal side of lysine or arginine. Mammalian trypsins have the ability to activate trypsinogen, the zymogen form of trypsin; it has been proposed that mosquito trypsin-like serine proteases may also display autocatalytic behavior (autoactivation) but has not yet been shown definitively.⁴⁴

Knockdown studies of the four most abundant *Ae. aegypti* proteases have shown that only three have an effect on mosquito fecundity: AaLT, AaSPVI, and AaSPVII.²² While mRNA knockdown of each individual protease resulted in a decrease in the amount of eggs laid by the mosquito, knockdown of all three of these proteases at once did not have an overall additive decrease as expected.²² This suggests that other *Ae. aegypti* midgut enzymes may compensate for this loss of activity.²²

In order to determine which proteases may play a significant role in this process, expression profiles were determined for the other nine *Ae. aegypti* midgut serine proteases.³⁵ Expression profiles for *Ae. aegypti* serine proteases II–V (AaSP group I) show that these genes are expressed at constant levels before and after blood meal acquisition.³⁵ Therefore, it is possible that AaSP group I proteases may be responsible for protease activity observed following knockdown of abundant proteases.³⁵ High sequence similarity between AaSP group I genes prevented researchers from determining individual gene transcript levels; instead, the cumulative level of AaSP group I genes were quantified from whole body *Ae. aegypti* genomic DNA libraries.³⁵ Initial sequence analyses show that these enzymes are trypsin-like serine proteases each containing six cysteine residues corresponding to three highly conserved disulfide bonds; serine proteases II and IV (AaSPII and AaSPIV) sequences contain two additional cysteine residues for a total of four disulfide bonds in the native conformation.³⁵ Sequence alignments show the AaSP group I genes belong to the same clade, with highest sequence similarity between AaSPII and AaSPIV (96% sequence identity).³⁵ Therefore, the degree to which individual proteases belonging to the AaSP group I are expressed in the mosquito midgut remains unknown.³⁵

Results from mRNA knockdown experiments suggest that inhibiting blood meal metabolism in *Ae. aegypti* may additionally require inhibition of compensatory proteases. Such development of specific inhibitors requires prior biochemical studies aimed at enzyme characterization, which have not yet been performed for AaSP group I proteases and other proteases with lower expression levels. Specifically, AaSPII and AaSPIV are of

particular interest in this study. These genes will be cloned and studied *in vitro* to gain insight into their function and determine their potential roles in the *Ae. aegypti* blood meal process.

1.3 Recombinant Protein Expression in *Escherichia coli*

Natural protein quantities are often insufficient for biochemical study; instead, recombinant technologies are frequently used in biochemical research to obtain pure, soluble, and functional proteins of interest.⁴⁷ Large quantities of recombinant protein for biochemical studies can be produced by utilizing simple prokaryotic organisms such as *Escherichia coli* and bacterial specific plasmid expression vectors.⁴⁸ Expression using *E. coli* as the host is often employed; however, using prokaryotic expression systems for eukaryotic proteins may not always be successful and can cause expression of insoluble protein.⁴⁷ The *E. coli* cytoplasm is maintained in a reduced state primarily through thioredoxin and glutaredoxin systems, which makes disulfide bond formation in proteins unfavorable.⁴⁹⁻⁵¹ The thioredoxin system comprises of thioredoxin reductase and thioredoxin and the glutaredoxin system consists of glutathione reductase, glutathione, and three glutaredoxins.⁵¹ Both systems use NADPH to reduce disulfide bonds in enzymes expressed in the *E. coli* cytoplasm.⁵¹ Therefore, using *E. coli* to express proteins requiring multiple disulfide bonds, such as AaS^{PII} and AaS^{PIV}, to maintain structure and function may result in the expression of misfolded proteins (or inclusion bodies) when necessary disulfide bonds are reduced.⁴⁷ Several strategies have been developed in an effort to overcome such obstacles in expressing eukaryotic recombinant protein and include both preventative and consequent techniques.⁴⁷ Specifically, three subsets of

methods are commonly used to obtain soluble, active protein from *E. coli*: those that attempt to modify the *in vivo* parameters without modifying the target protein; those that involve modifying the target protein; and those that employ *in vitro* refolding methods.⁴⁷

Numerous strategies in which target modification is avoided exist and can include: cultivating bacteria at reduced temperatures, adjusting growth media components, using mutant *E. coli* strains engineered to promote solubility, or addition of appropriate binding cofactors, binding partners, or molecular chaperones.⁴⁷ When these methods are unsuccessful, it may be necessary to engineer the target protein by attaching a more soluble species as a fusion protein.⁴⁷ In cases of unavoidable protein aggregation and inclusion body formation, solubilization and refolding methods can be used; however, this often results in poor recovery or inactive enzyme.⁵²

1.4 Protein Purification Techniques

Upon successful soluble protein expression, it is necessary to purify the target enzyme. Often, a series of purification steps are executed to ensure homogeneity of the final product.⁵³ Ion exchange chromatography is one purification method that can be employed. Ion exchange chromatography separates proteins based on charge type and relative strength; it is dependent on the interaction between charged amino acid residues in the sample and ions in the column stationary phase.⁵⁴ At physiological pH, the amino acids lysine, arginine, and histidine are positively charged while aspartate and glutamate are negatively charged. The overall charge of the target protein is dependent on pH and the amino acid sequence, which determines the protein's isoelectric point (pI).

Another common method of protein purification is immobilized metal affinity chromatography (IMAC). Specific amino acids located on the protein surface can bind to certain metal ions; this interaction is exploited in IMAC for protein purification. Protein binding in IMAC is dependent on the histidine residues accessible on the protein's surface, as histidine shows the strongest interaction with immobilized metal matrices due to the electron donating groups on the imidazole ring of histidine that readily coordinate with the metal.⁵⁵⁻⁵⁷ Histidine-tagged fusion proteins can be purified in this manner with the use of Ni(II) or other divalent ions. Histidine-tagged proteins can be applied to IMAC column matrices and impurities can be removed with washing; the target protein can then be eluted simply by adjusting the pH or adding free imidazole to the column.⁵⁷ There are several advantages in using IMAC for purification; for example, there is usually low contamination from bacterial proteins due to the high selectivity of histidine-tagged protein to the metal.⁵⁸

Size-exclusion chromatography is another technique that can be used to purify proteins and is often employed as the final step in the purification process. In size-exclusion chromatography, sample components separate based on molecular size and shape; separation performed by size-exclusion chromatography is determined by the pore size of the column matrix.⁵⁹ Larger proteins that cannot move through the pores instead travel around the matrix and elute prior to smaller proteins; because small proteins can fit through all of the pores in the matrix, they take a longer path and thus elute last.⁵³ Because purification is not dependent on binding of the target protein, this method is generally less susceptible to protein loss.⁵³

1.5 Protease Characterization Techniques

Upon successful purification of the target protein, biochemical methods of enzyme characterization can be performed. Trypsin-like activity can be efficiently measured using N α -benzoyl-D,L-arginine-p-nitroanilide hydrochloride (BAPNA), a synthetic colorimetric substrate for trypsin.^{43,60,61} Trypsin-like proteases will cleave BAPNA to generate *p*-nitroaniline (pNA), which is yellow in color and can be detected at 410 nm; the color intensity of pNA is directly proportional to concentration, allowing for an accurate measurement of protease activity.⁶⁰ Total enzyme activity can be determined by observing the production of pNA from BAPNA via a UV-visible spectrophotometer and using the extinction coefficient, $\epsilon_{410\text{nm}} = 9960 \text{ M}^{-1}\text{cm}^{-1}$.⁶⁰

Characterization typically also involves the determination of the three-dimensional enzyme structure in order to understand the catalytic mechanism, identify the enzyme active site, and determine the substrate binding mechanism for the potential development of inhibitors. The primary methods to determine protein structure at atomic resolution are protein crystallography and nuclear magnetic resonance (NMR) spectroscopy. Protein crystallography involves collecting data of an X-ray beam being diffracted by a protein crystal and can provide an electron density model of the crystal.⁶² Often, the most difficult aspect of protein crystallography is forming the crystal; proteins are large, irregularly shaped, and dynamic macromolecules that do not easily form an ordered crystal lattice.⁶³ The greatest limitation of protein crystallography is that it requires crystallization of the protein, which renders the protein target immobile and may not be representative of dynamic protein structures in solution.

NMR spectroscopy can be applied to proteins in solution and uses radiowaves that interact with the protein in a strong magnetic field and can be used for proton localization, which can be difficult to determine via crystallography.^{62,64} Because it does not require protein crystallization, NMR spectroscopy can be used to study reaction kinetics and dynamics.⁶⁵ However, proteins large in size can be extremely difficult to study using protein NMR as spectra become quite complex. Often, complementary techniques can be used for structure determination to resolve various aspects of protein structure.

1.6 Specific Research Aims

Research herein aims to understand the role of *Ae. aegypti* trypsin-like midgut serine proteases II and IV (AaSPII and AaSPIV). Specifically, *in vitro* biochemical studies of AaSPII and AaSPIV will involve successful cloning and bacterial expression of soluble, recombinant enzyme and protein purification in preparation for biochemical analyses. Studying AaSPII and AaSPIV, as well as other midgut *Ae. aegypti* proteases, will aid in understanding the overall role of proteases involved in blood meal digestion and may eventually allow for the development of potential enzyme inhibitors to inhibit egg production, minimize mosquito populations, and hopefully lead to a decrease in viral pathogen transmission.

CHAPTER 2: METHODS

2.1 Preparation of Protease Plasmid Constructs

2.1.1 General Cloning Strategy

Gene-specific polymerase chain reaction (PCR) primers for serine proteases AaS^{PII} and AaS^{PIV} were designed manually using a modified cloning PCR strategy to generate a gene product containing the appropriate base pair region encoding the AaS^{PII} or AaS^{PIV} gene from cDNA isolated from blood-fed mosquitoes (Table 1 and Table 2).⁶⁶

All primers were checked using NetPrimer

(<http://www.premierbiosoft.com/netprimer/netplaunch/netplaunch.html>) for appropriate parameters including: length, melting temperature (T_m), potential hairpin and primer-dimer formation, and G-C content.

Table 1. Primers used for PCR amplification and cloning of recombinant forms of the full length zymogen (with leader), mature form (no leader), MBP-fusion, EK-pseudo propeptide, and mutant (Ser to Ala and Lys to Ala) AaS^{PII}.

Protease	Leader	Target Vector for Cloning	Primer	Primer Sequence
AaS ^{PII}	With Leader	pET28a	Fwd	5'-AAAAAGCTAGCATGTGG TTTTCTGTGAAAATATCG -3'
			Rev	5'- AAAAAGCTCGAGTTATT ACACAACCCAAACCGCTC -3'
AaS ^{PII}	No Leader	pET28a	Fwd	5'- AAAAAGCTAGCATGTC GGTGGACAACGACGTG -3'
			Rev	5'- AAAAAGCTCGAGTTATT ACACAACCCAAACCGCTC -3'
EK-AaS ^{PII}	No Leader	pET28a	Fwd	5'- AAAAAGCTAGCAACAACA ACCTCGGCGATGACGATGACA AGATCATCGGTGGTTTTCCCG -3'
			Rev	5'- AAAAAGCTCGAGTTATTAC ACAACCCAAACCGCTC -3'
MBP-AaS ^{PII}	No Leader	pMAL-c5E	Fwd	5'- AAAAAGCATGGGCGAGC AGCCATCATCATC -3'
			Rev	5'- AAAAAGAATTCTTATTACA CAACCCAAACCGCTCT -3'
AaS ^{PII} Ser to Ala Mutant	No Leader	pET28a	Fwd	5'- CTCTTGCCAGGGTGATGC CGGGGGTCCTTTGG -3'
			Rev	5'- CCAAAGGACCCCGGC ATCACCTGGCAAGAG -3'
AaS ^{PII} Lys to Ala Mutant	No Leader	pET28a	Fwd	5'- CGGTGGACAACGACGTG GCGATCATCGGTGGTTTTC -3'
			Rev	5'- GAAAACCAACGATGATC GCCACGTCGTTGTCCACCG -3'
AaS ^{PII}	No Leader	pET29b	Fwd	5'- AAAAAGCATGGGCTCGG TGGACAACGACGTG -3'
			Rev	5'- AAAAAGCTCGAGCACAAC CCAAACCGCTCTG -3'

Table 2. Primers used for PCR amplification and cloning of recombinant forms of the full length zymogen (with leader), mature form (no leader), MBP-fusion, EK-pseudo propeptide, and mutant (Ser to Ala and Lys to Ala) AaSPIV.

Protease	Leader	Target Vector for Cloning	Primer	Primer Sequence
AaSPIV	With Leader	pET28a	Fwd	5'- AAAAAGCTAGCATGTGGTTTT CTGTGAAAATATCG -3'
			Rev	5'- AAAAAGCTCGAGTTACACAACC CAAACCGCT -3'
AaSPIV	No Leader	pET28a	Fwd	5'- AAAAAGCTAGCATGTCGGTGG ACAACGACATGA -3'
			Rev	5'- AAAAAGCTCGAGTTACACAACC CAAACCGCT -3'
EK-AaSPIV	With Leader	pET28a	Fwd	5'-AAAAAGCTAGCAACAACAACC TCGGCGATGACGATGACAAGATCA TCGGCGGTTTTCC-3'
			Rev	5'- AAAAAGCTCGAGTTACACAACC CAAACCGCT -3'
MBP-AaSPIV	With Leader	pMAL-c5E	Fwd	5'- AAAAACCATGGGCAGCAGCCA TCATCATC -3'
			Rev	5'- AAAAAGAATTCTTACACAACC CAAACCGCT -3'
AaSPIV Ser to Ala Mutant	With Leader	pET28a	Fwd	5'- CTCTTGCCAGGGTGATGCCGG GGGTCCTTTGG -3'
			Rev	5'- CCAAAGGACCCCCGGCATCAC CCTGGCAAGAG -3'
AaSPIV Lys to Ala Mutant	With Leader	pET28a	Fwd	5'- CGGTGGACAACGACATGGCGA TCATCGGCGGTTTTC -3'
			Rev	5'- GAAAACCGCCGATGATCGCCA TGTCGTTGTCCACCG -3'
AaSPIV	No Leader	pET29b	Fwd	5'- AAAAACCATGGGCTCGGTGGA CAACGACATGAA -3'
			Rev	5'- AAAAAGCTCGAGCACAACCCAA ACCGCTCTG -3'

PCR reactions were performed to obtain the target gene region. Each reaction contained 1X GoTaq Green Master Mix (Promega # M7128), 0.5 μ M Forward Primer, 0.5 μ M Reverse Primer, 25 ng DNA template, and nuclease-free water to 15 μ L. All PCR reactions were performed in an Eppendorf 5331 MasterCycler Gradient Thermal Cycler with cycling parameters as listed (Table 3).

Table 3. Cycling Parameters for PCR amplification of recombinant forms of the full length zymogen (with leader), mature form (no leader), MBP-fusion, EK-pseudo propeptide, and mutant (Ser to Ala and Lys to Ala) AaSPII and AaSPIV.

Steps	Cycles	Temperature	Time
Initial Denaturation	1	95 °C	5 min
Denaturation	25	95 °C	20 s
Annealing		58 °C	20 s
Extension		72 °C	20 s
Final Extension	1	72 °C	5 min
Hold	1	20 °C	HOLD

PCR reaction success was evaluated by performing electrophoresis on reaction samples in a 1% agarose gel containing a 1:10000 dilution of GelRed Nucleic Acid Gel Stain (Biotium # 41002). The gel was then visualized by UV transillumination. Success of the PCR reactions was determined by the presence of a band near the correct molecular weight (predicted size of the PCR product) using a standard DNA ladder (Thermo Scientific GeneRuler 1 kb DNA Ladder, # SM0311). Successful PCR products were carefully excised and extracted from the gel using the Thermo Scientific GeneJET Gel Extraction Kit (# K0692) and eluting with 20 μ L ultrapure water. The PCR products extracted from the gel were digested with 1 μ L of each restriction enzyme (Thermo Scientific FastDigest restriction enzymes, details listed in following sections) and 1X FastDigest Buffer (Thermo Scientific # B64) prior to incubation at 37 °C for 1.5 h.

Vector digestion was performed in preparation for insertion of PCR products by combining 1000 ng vector, 1 μ L of each restriction enzyme, 1X FastDigest Buffer (Thermo Scientific # B64), and nuclease-free water to 20 μ L. The mixture was incubated at 37 °C for 1.5 h. Digestion mixture was then incubated at 90 °C for 10 min and then on ice for 1 min. Dephosphorylation of the digested vector was performed by adding 1 μ L FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific # EF0654) and incubating for 1.5 h at 37 °C. Digested and dephosphorylated vector solutions were also run on a 1% agarose gel containing a 1:10000 dilution of GelRed Nucleic Acid Gel Stain (Biotium # 41002). Vector bands were excised from the gel and combined with digested PCR products and co-purified using the Thermo Scientific GeneJET Gel Extraction Kit (# K0692) and eluted with 20 μ L nuclease-free water.

Both vector and PCR product were digested with the same restriction enzymes in order to introduce the complimentary “sticky ends” to aid in successful ligation forming the plasmid of interest. PCR products and vector ligation was performed by adding 1 μ L T4 DNA Ligase (NEB # M0202S) and 1X T4 DNA Ligase Reaction Buffer (NEB # B0202S) and incubating at room temperature for 20 min.

NEB 10-beta Competent *E. coli* cells (# C3019H) were briefly thawed on ice in preparation for transformation of the ligated plasmid. Immediately after cells were thawed, 10 μ L from the ligation reaction was added to the cells. The mixture was mixed by flicking the tube 4–5 times and incubated for 30 min on ice. Cells were heat shocked at 42 °C for 30 s and incubated on ice for 5 min. Cells were combined with 950 μ L of SOC Outgrowth Media (NEB #B9020S) and incubated for 1 h in a 37 °C shaker

(250 RPM). Centrifugation at 13000 RPM for 1 min was then performed to pellet the cells and remove 900 μ L of growth media. Cells were resuspended in remaining media and plated on agar containing an appropriate antibiotic [30 mg/mL kanamycin (the pET28a plasmid construct) or 100 μ g/mL ampicillin or 50 μ g/mL carbenicillin (either antibiotic is appropriate for selection of the pMAL-c5E plasmid construct)] to select for cells containing desired plasmid. Transformation plates were incubated overnight at the optimal growth temperature. Colonies were then restreaked onto fresh transformation plates and incubated overnight at the optimal growth temperature. The general cloning procedure is summarized in Figure 10. Colony PCR was performed for new “stock” bacterial plates from original transformation colonies to check for successful molecular cloning. For each “stock”, one PCR reaction was prepared by combining 1X GoTaq Green Master Mix (Promega # M7128), 0.5 μ M Forward Primer, 0.5 μ M Reverse Primer, and nuclease-free water to 15 μ L with one colony from the restreaked plate. All PCR reactions were performed in an Eppendorf 5331 MasterCycler Gradient Thermal Cycler with cycling parameters as listed in Table 3 and reaction success was assessed via 1% agarose gel electrophoresis.

For each sample with successful Colony PCR products, one colony was added to overnight cultures containing 25 mL Luria-Bertani (LB) broth with appropriate antibiotic for selection. Overnight cultures were incubated overnight (16–18 h) at the optimal growth temperature with shaking (250 RPM). Plasmids were isolated from cells grown overnight using the Thermo Scientific GeneJET Plasmid Miniprep Kit (# K0502) and eluting with 30 μ L nuclease-free water. Plasmid concentrations were determined using a

Thermo Scientific NanoDrop 1000 Spectrophotometer. Plasmid construct DNA sequences were verified through sequencing.

Figure 10. General molecular cloning procedure. Target genes are PCR-amplified from *Ae. aegypti* cDNA and digested with restriction enzymes. The expression vector is digested with the same restriction enzymes to facilitate ligation of the target gene into the expression vector. The vector plasmids are then transformed into *E. coli*.

2.1.2 pET28a Plasmid Constructs

Primers were designed to include the *Nhe*I and *Xho*I restriction sites to enable cloning of the gene into the pET28a expression vector (Novagen # 69864-3) in addition to a 5'-extension to the restriction site (consisting of 5 adenine bases) to increase cleavage efficiency of the restriction enzymes.⁶⁶ Two sets of primers were designed for each target protein. The gene fragment obtained was ligated into the bacterial expression vector pET28a downstream of the T7 promoter, lac operator, and N-terminal sequence encoding six histidine residues (His₆-tag) for AaS^{PII} and AaS^{PIV}. Restriction enzymes used in

cloning process were Fisher Scientific FastDigest *NheI* (# FD0973) and *XhoI* (# FD0694).

The pET28a vector utilizes the inducible T7 expression system in order to express the cloned gene sequences in *E. coli* cells (Figure 11). The T7 RNA polymerase coding sequence is present in *E. coli* cells and is controlled by the lac promoter and induced by lactose; the pET28a plasmid is under the control of a promoter recognized by T7 RNA polymerase.⁶⁷ Production of T7 RNA polymerase can be artificially induced by isopropyl- β -D-thiogalactoside (IPTG, a non-hydrolyzable analog of allolactose) and DNA transcription controlled by the T7 promoter.

Figure 11. Regulation of protein expression with vectors utilizing the T7 expression system. Binding of IPTG to the *lac* repressor induces a conformational change preventing the *lac* repressor from binding to the *lac* promoter. This allows *E. coli* RNA polymerase to bind and begin transcribing the T7 RNA polymerase. Once T7 RNA is expressed, it can bind to the T7 promoter and begin transcribing the gene of interest. Figure adapted by author from Novagen pET System Manual.

2.1.3 Enterokinase (EK)-Pseudo Propeptide Plasmid Constructs

Since it is unknown whether AaS^{PII} and AaS^{PIV} proteases are autocatalytic or have trypsin-like activity, an enterokinase (EK) cleavage site (DDDDK) was incorporated into a pseudo propeptide region introduced via PCR to allow for *in vitro* activation by commercially available recombinant enterokinase (Figure 12).²³ The predicted activation site was identified by sequence analysis using SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP/>). Although the sequence analysis suggested that the preferred site for activation is at a lysine residue, that site was chosen for design of the EK Forward Primer. Primers were designed to include the *NheI* and *XhoI* restriction sites to enable cloning of the gene into the pET28a expression vector (Novagen) in addition to incorporation of the EK cleavage site and 5'-extension to the restriction site (5 adenine bases). Restriction enzymes used in cloning process were Fisher Scientific FastDigest *NheI* (# FD0973) and *XhoI* (# FD0694). The gene fragment obtained was ligated into the bacterial expression vector pET28a downstream of the T7 promoter and lac operator.

Figure 12. Amino acid sequence of the zymogen polypeptide for AaSPII and AaSPIV. The protein leader sequence and propeptide region (top) were predicted via sequence analyses based on the known sequence of bovine trypsin. Hydrophobic residues of the leader sequence and propeptide region are shown in green, negatively-charged residues in red, positively-charged residues in blue, and polar uncharged residues in gray. The amino acid sequence of the engineered EK-pseudo propeptide forms of the recombinant proteases (bottom) is shown. The EK-pseudo propeptide contains an enterokinase cleavage site (pink).

2.1.4 pMAL-c5E Plasmid Constructs

Additionally, maltose binding fusion protein constructs were designed and cloned into the pMAL-c5E vector (NEB # N8110S), which contains the *malE* gene of *E. coli* that encodes the maltose binding protein (MBP). This *E. coli* strain lacks the T7 RNA polymerase and instead uses the naturally produced *E. coli* RNA polymerase from the bacterial expression cells; expression can still be induced using IPTG. While the exact mechanism is unknown, MBP has been shown to act as a highly effective enhancer of target protein solubility when used as a fusion protein.⁴⁷ *Escherichia coli* MBP is a periplasmic protein involved in sugar chemotaxis and transport across the bacterial envelope.⁶⁸ AaSPII (no leader) and full length AaSPIV (with leader) were each cloned into the pMAL-c5E vector using each pET28a vector plasmid construct as a template for

PCR so that constructs included the His₆-tag encoded portion from pET28a constructs for later use in purification steps (Figure 13). Primers were designed to contain *Eco*RI and *Nco*I restriction sites. Restriction enzymes used in the cloning process were Fisher Scientific FastDigest *Eco*RI (# FD0275) and *Nco*I (# FD0573). The cloned gene was inserted downstream the *malE* gene so that the fusion protein contained MBP N-terminally fused to the target protein (with the His₆-tag coding sequence from the pET28a plasmid construct).

Figure 13. MBP–His₆-tag–Protein fusion expected product. To improve soluble expression of the target protein of interest, the pMAL-c5E vector was utilized to clone the target gene downstream of the Maltose Binding Protein. However, the target gene included the N-terminal His₆-tag sequence in order to produce the above fusion.

2.1.5 pET29b Plasmid Constructs

Additional plasmid constructs using the pET29b vector were developed by students at San Jose State University as part of the Fall 2016 Chem 131B lab course using methods described here. Primers were designed to include the *Nco*I and *Xho*I restriction sites to enable cloning of the gene into the pET29b expression vector (Novagen) in addition to a 5'-extension to the restriction site (5 adenine bases). Unlike the pET28a vector, inserting the gene fragment into the pET29b vector incorporates a C-terminal sequence encoding six histidine residues (His₆-tag) for AaS^{PII} and AaS^{PIV}. Therefore, reverse primers

designed did not encode stop codons to allow for incorporation of a C-terminal His₆-tag. The gene fragment obtained was ligated into the bacterial expression vector pET29b downstream of the T7 promoter and lac operator. The pET29b vector also utilizes the inducible T7 expression system in order to produce proteins from cloned sequences in *E. coli* cells and production of T7 RNA polymerase can be experimentally induced by IPTG (Figure 11).

2.1.6 Site-Directed Mutagenesis

Site-directed mutagenesis reactions were used to introduce single base point mutations into the known DNA plasmid templates using the Thermo Scientific Phusion Site-Directed Mutagenesis Kit (# F541). Two complimentary oligonucleotide primers were designed for each reaction to incorporate the desired mutation using guidelines described for the Stratagene QuikChange Site-Directed Mutagenesis Kit (# 200518).

Two sets of mutagenesis primers were designed for different mutations: (1) primers for mutating the active site serine residue to an alanine residue and (2) primers for mutating the predicted cleavage site residue (lysine) to an alanine residue. Both mutants were designed to prevent potential autoactivation of the recombinant enzymes. Reactions were prepared by mixing components to a final volume of 50 μ L (Table 4). Reactions were performed in an Eppendorf 5331 MasterCycler Gradient Thermal Cycler with parameters described (Table 5).

Table 4. General PCR Reaction for Site-Directed Mutagenesis Reaction – Single Base Point Mutation

Component	Quantity
5X Phusion HF Buffer*	10 μ L
10 mM dNTPs*	1 μ L
10 mM Forward Primer	2.5 μ L
10 mM Reverse Primer	2.5 μ L
1 ng/ μ L Plasmid Template DNA	1 μ L
2 U/ μ L Phusion Hot Start DNA Polymerase*	0.5 μ L
Nuclease-Free Water	32.5 μ L

Table 5. Thermal Cycler Parameters for Site-Directed Mutagenesis Reaction – Single Base Point Mutation

Steps	Cycles	Temperature	Time
Initial Denaturation	1	98 °C	30 s
Denaturation	25	98 °C	10 s
Annealing		65 °C	30 s
Extension		72 °C	3 min
Final Extension	1	72 °C	5 min
Hold	1	20 °C	HOLD

All plasmid constructs developed are summarized in Table 6. Plasmid constructs are displayed by the assigned annotation, which indicates: the target protein it encodes (wild-type AaS^{PII}, wild-type AaS^{PIV}, mutated Ser to Ala AaS^{PII}, or mutated Ser to Ala AaS^{PIV}); the plasmid used to facilitate expression; whether the inserted gene sequence includes the leader sequence or not; and notable features of each plasmid construct. Primers for mutating the predicted cleavage site lysine residue were designed but have not yet been cloned into a vector plasmid; mutants listed in the table are those with mutations aimed at changing the active site serine residue.

* Included in Thermo Scientific Phusion Site-Directed Mutagenesis Kit (# F541)

Table 6. Summary of plasmid constructs developed for expression of recombinant forms of the full length zymogen (with leader), mature form (no leader), MBP-fusion, EK-pseudo propeptide, and mutant (Ser to Ala and Lys to Ala) AaS^{PII} and AaS^{PIV}.

Annotation	Target Protein	Vector	Leader Sequence	Notable Feature
AaS ^{PII} -pET28a	WT AaS ^{PII}	pET28a	With Leader	N-terminal His ₆ -tag
AaS ^{PII} -NL-pET28a		pET28a	No Leader	N-terminal His ₆ -tag
EK-AaS ^{PII} -pET28a		pET28a	No Leader	N-terminal His ₆ -tag and EK cleavage site
AaS ^{PII} -NL-pMAL		pMAL-c5E	No Leader	N-terminal MBP and His ₆ -tag
AaS ^{PII} -NL-pET29b		pET29b	No Leader	C-terminal His ₆ -tag
AaS ^{PII} -NL-Ala-pET28a ^{**}	Mutant AaS ^{PII}	pET28a	No Leader	N-terminal His ₆ -tag and Mutated Ser to Ala
AaS ^{PIV} -pET28a	WT AaS ^{PIV}	pET28a	With Leader	N-terminal His ₆ -tag
EK-AaS ^{PIV} -pET28a		pET28a	With Leader	N-terminal His ₆ -tag and EK cleavage site
AaS ^{PIV} -pMAL		pMAL-c5E	With Leader	N-terminal MBP and His ₆ -tag
AaS ^{PIV} -NL-pET29b		pET29b	No Leader	C-terminal His ₆ -tag
AaS ^{PIV} -Ala-pET28a ^{**}	Mutant AaS ^{PIV}	pET28a	With Leader	N-terminal His ₆ -tag and Mutated Ser to Ala

^{**} Plasmids were constructed for mutation of the active site serine residue to an alanine residue and cloning was successful but experiments for the successful recombinant protein expression of the mutant proteases in *E. coli* have not yet been performed and are therefore not discussed in the Results section (Chapter 4).

2.2 Recombinant Protein Expression

SHuffle T7 competent *E. coli* cells were chosen for bacterial expression of AaSPII and AaSPIV because they were engineered to express cytoplasmic proteins containing disulfide bonds and are compatible with expression using the T7 promoter system.⁶⁹ Plasmid constructs were transformed into SHuffle T7 competent *E. coli* cells (NEB # C3029H) and grown overnight on LB agar with 30 mg/mL kanamycin (the pET28a plasmid construct) or 100 µg/mL ampicillin or 50 µg/mL carbenicillin (either antibiotic is appropriate for selection of the pMAL-c5E plasmid construct) at 30 °C. One colony from the overnight plate growth was added to 50 mL LB with appropriate antibiotic and grown overnight (16–18 h) at 30 °C with shaking (250 RPM). The optical density at 600 nm (OD₆₀₀) of a ten-fold diluted sample was used to measure culture growth in a Hewlett Packard 8452A Diode Array Spectrophotometer and a 1-cm path-length cuvette.

2.2.1 Induction by IPTG

Six 1-L Erlenmeyer flasks each containing either 500 mL LB or 500 mL Terrific Broth (TB) media with proper antibiotic (based on plasmid antibiotic resistance) and an appropriate volume of overnight culture to start at an OD₆₀₀ of 0.05 was incubated at 30 °C with shaking (250 RPM) until reaching OD₆₀₀ between 0.8 and 1.0. At this time, recombinant expression of the target protein (AaSPII or AaSPIV) was induced with 0.1 M IPTG, the growth temperature dropped to 10 °C or 15 °C (250 RPM), and grown for up to 96 hours. 1 mL samples were collected at various time-points throughout the growth (determined by hours post induction), centrifuged at 13000 RPM for 5 min, and pellets

stored at $-20\text{ }^{\circ}\text{C}$. At the growth experiment endpoint, cell paste was collected by centrifugation at 6000 RPM and flash frozen in liquid nitrogen before storage at $-80\text{ }^{\circ}\text{C}$.

2.2.2 Auto-Induction

Methods used for bacterial growth and recombinant protein expression by auto-induction was designed with parameters according to the auto-induction process outlined by Studier.⁷⁰ Auto-induction processes use rich media in order to accomplish high cell densities, auto-induction (without the use of IPTG), and maximum target protein yields.^{70,71} Studier's ZYM-5052 auto-induction media, which was also used for this study, contains: N-Z-amine (an enzymatic digest of casein), glucose and glycerol to support growth to high densities, carbon sources, buffer to maintain pH, critical nitrogen and magnesium sources.⁷⁰ Glucose and lactose are both included in the media: glucose prevents induction by lactose in early stages of growth and, once glucose is depleted, lactose automatically induces expression of the T7 RNA polymerase.⁷⁰ Auto-induction only works with vectors and bacterial strains containing the T7 expression system, such as pET28a, as they require the *lacUV5* promoter.^{67,70,71}

A 125-mL Erlenmeyer flask containing 50 mL Auto-Induction Media (corresponding to Studier's ZYM-5052 medium) with 0.05% w/v lactose, 0.2% w/v glucose, proper antibiotic, and an appropriate volume of overnight culture to reach final OD₆₀₀ of 0.05 was incubated at $30\text{ }^{\circ}\text{C}$ with shaking (250 RPM) until reaching OD₆₀₀ between 0.5 and 0.8. At this time, the reaction flask was incubated on ice for 5 min and then at $15\text{ }^{\circ}\text{C}$ with shaking (250 RPM) and grown for 72 hours. Bacterial cells were obtained at various

time-points throughout the growth by centrifugation of 1 mL samples from growth media at 13000 RPM for 5 min and pellets were stored at -20°C for further analysis.

2.2.3 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Time-points collected throughout the growth were assessed for successful recombinant protein expression by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Sodium dodecyl sulfate (SDS) is an anionic detergent that binds to proteins, denaturing them and resulting in a linear structure with a net negative charge that is approximately proportional to the protein mass.⁴² SDS-protein complexes can then be subjected to electrophoresis within a polyacrylamide gel and will separate based on size: polypeptide chain mobility is inversely linearly proportional to the logarithm of their mass so small complexes move more rapidly through the gel.⁴²

Cells pelleted for each time-point sample were resuspended in 20 mM Tris-HCl buffer pH 8. Buffer amount was determined qualitatively based on the size of the largest pellet (typically between 250 and 350 μL for growth experiments using induction by IPTG and between 600 and 700 μL for growth experiments using auto-induction). Samples were maintained on ice and sonicated for 10 s at 20% for a total of three times using an ultrasonic probe (Fisher Scientific Model 120 Sonic Dismembrator with 120 Watt power requirement and 20 kHz operating frequency; QSonica Sonicators converter # CL18 and probe # 4423). In new 1.5 mL microcentrifuge tubes, 20 μL from each of the total sonicated samples was combined with 3.4 μL 6x SDS Loading Dye. Remaining sonicated samples were centrifuged at 13000 RPM for 5 min to pellet in fractions and obtain soluble sample fractions. Then, 20 μL from each of the soluble sonicated samples

was combined with 3.4 μ L 6x SDS Loading Dye. Total and soluble samples with SDS Loading Dye added were incubated at 90 °C for 4 min intervals, briefly centrifuging between incubation periods until no visible bubbles remained. Samples were subsequently loaded into the wells of a NuPAGE Novex 4-12% bis-Tris polyacrylamide gel (Thermo Scientific # NP0321BOX, # NP0322BOX, # NP0323BOX) along with 6 μ L of a protein standard ladder (Thermo Scientific PageRuler Prestained Protein Ladder, # 26616). The gel was subjected to electrophoresis in NuPAGE 1x MES SDS Running Buffer (Thermo Scientific # NP000202; 50 mM MES, 50 mM Tris Base, 0.1% SDS, 1 mM EDTA, pH 7.3) at 180 V for 55 min. Following electrophoresis, the gel was rinsed three times with approximately 20 mL ultrapure water (purity above 18 M Ω ·cm at 25 °C) for 5 min. The gel was then stained with 20 mL SimplyBlue SafeStain (Thermo Scientific # LC6065) for at least 1 h and then destained with 20 mL ultrapure water for at least 1 h. Gels images were obtained using the Alpha Innotech FluorChem 8900 Transillumination Imaging System Light Cabinet.

2.2.4 His-Stain

To aid in identifying expression of His₆-tagged proteins, InVision His₆-tag In-gel Stain (Novex Life Technologies # LC6030) was also used. The stain contains a dye that is conjugated to a Ni²⁺-nitrilotriacetic acid complex, which binds specifically to the six-histidine residues incorporated into the protein of interest. The fluorescent dye can be detected using a UV transilluminator.

Following SDS-PAGE procedures (as previously described), gels were fixed by adding 100 mL Fixing Solution (40% v/v ethanol, 10% v/v acetic acid solution),

microwaving (~1000 W) for 1 min, and incubating with agitation for 10 min at room temperature. The gels were washed by decanting the Fixing Solution and adding 100 mL ultrapure water, microwaving for 1 min, and incubating for 5 min with agitation. Then, the water was decanted and 25 mL InVision His₆-tag In-gel Stain (Novex # LC6030) was added, microwaved for 1 min and incubated with agitation for 40 min at room temperature. The stain was decanted and gels were then washed by adding 100 mL 20 mM phosphate buffer, pH 7.8, microwaving for 1 min, and incubating at room temperature for 10 min with agitation (three times). Gels were imaged immediately using a UV transilluminator, followed by washing with ultrapure water and stained with SimplyBlue Safe Stain.

2.2.5 Western blotting

The immunoassay technique of Western blotting was used to confirm the successful expression of the target recombinant proteins. Western blotting involves detection of proteins on the surface of a polymer sheet by exploiting antibody-antigen interactions.⁴² A primary antibody specific to the expressed protein is added to the sheet for interaction; a secondary antibody labeled with enzyme reporters specific to the primary antibody will complex with the primary antibody.⁴² Light produced by the enzymatic reaction allows for detection of the complex; chemiluminescent detection in this manner is extremely specific and sensitive.

In order to verify the presence of the target recombinant protein containing a His₆-tag, SDS-PAGE was performed as previously described. Following electrophoresis, the gel was rinsed three times in approximately 20 mL ultrapure water for 5 min and then rinsed

in 20 mL NuPAGE 1x Transfer Buffer (Thermo Scientific # NP0006) with 10% methanol for 10–15 min.

Proteins from the gel were transferred to a nitrocellulose membrane (EMD Millipore # ISEQ00010) via electrophoresis with the membrane on the cathode and the gel on the anode at 30 V for 1.5 h in NuPAGE 1x Transfer Buffer (Thermo Scientific # NP0006). The blot (membrane) was briefly rinsed with water and then rinsed with 10 mL 1x TBS (Tris-buffered saline: 20 mM Tris, 137 mM NaCl, pH 7.5) buffer for 5 min. The blot was then incubated with Blocking Buffer (3% w/v BSA in 1x TBS with 0.1% Tween 20) at room temperature for 1 h with shaking. Three 5 min rinses with 10 mL 1x TBST (1x TBS with 0.1% Tween 20) were performed, followed by an overnight incubation at 4 °C in 10 mL Primary Antibody Solution prepared by diluting (1:1000) His₆-tag Polyclonal Antibody (Cell Signaling Technology # 2365) in 3% w/v BSA in 1x TBST. Three subsequent 5 min rinses with 10 mL 1x TBST were performed and the membrane was then incubated in 10 mL Secondary Antibody Solution prepared by diluting (1:2000) secondary antibody (GE Healthcare # NA934-1ML, Amersham enhanced chemiluminescent Anti-Rabbit IgG, horseradish peroxidase linked whole antibody from donkey) in 3% w/v BSA in 1x TBST for 1 h at room temperature with shaking. Three final 5 min rinses with 10 mL 1x TBST were performed before imaging the Western blot.

The membrane was incubated with SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific # 34077), an enhanced chemiluminescent (ECL) substrate for low-picogram-level detection of horseradish peroxidase (HRP), for 1 min and visualized in the Alpha Innotech FluorChem 8900 system equipped with a

chemiluminescent filter. The image was collected with varying exposure times between 1 and 10 min.

2.3 Protein Purification

2.3.1 Wheat Starch Purification for MBP-Tagged Proteins

In order to purify MBP-tagged target proteins, a purification method using wheat starch was attempted. The binding of MBP to wheat starch is successful due to the presence of amylose, which can then be exploited to immobilize the target protein, wash away contaminants, and elute the MBP-tagged protein with maltose. Maltose binds with stronger affinity to the MBP tag than the wheat starch, facilitating elution of the MBP-fusion protein of interest.

Collected cell paste from previous growth was resuspended in 5 mL/g Resuspension Buffer (20 mM Tris-HCl, 250 mM NaCl, pH 7.2). The cell suspension was maintained on ice and lysed by sonication at 20% amplitude for a total of 8 min using a pulsing (10 s on, 15 s off) ultrasonic probe (Fisher Scientific Model 120 Sonic Dismembrator with 120 Watt power requirement and 20 kHz operating frequency; QSonica Sonicators converter # CL18 and probe # 4423). At this point, 20 μ L of “total” cell sample was collected. The lysed cell solution was centrifuged at 16000 RPM for 30 min at 4 °C and 20 μ L of the soluble fraction (crude lysate) was collected.

Prior to purification, wheat starch is subjected to washing and equilibration. First, 50 g wheat starch was rinsed three times in water: the wheat starch was suspended in 200 mL cold ultrapure water and the solution was centrifuged at 12000 RPM for 10 min

at 4 °C. The supernatant was discarded after each rinse. Three additional rinses with Resuspension Buffer were then performed.

After equilibration, 100 mL of the diluted crude lysate solution was added to the equilibrated wheat starch and the mixture was incubated at 4 °C for 30 min with shaking (200 RPM). The mixture was centrifuged at 12000 RPM for 6 min at 4 °C. The supernatant (Flow Through 1) was decanted and stored at 4 °C. The wheat starch was then resuspended in 200 mL Resuspension Buffer and centrifuged at 12000 RPM for 6 min at 4 °C; collected supernatant was stored at 4 °C as “Flow Through 2.” Again, the wheat starch was resuspended in 200 mL Resuspension Buffer and centrifuged at 12000 RPM for 6 min at 4 °C; collected supernatant was stored at 4 °C as “Flow Through 3.” Wheat starch pellet was then resuspended in 50 mL cold Elution Buffer (20 mM Tris-HCl, 250 mM NaCl, 25 mM maltose, pH 7.2) and incubated at 4 °C for 15 min with shaking (200 RPM). The solution was centrifuged at 12000 RPM for 6 min at 4 °C and the supernatant stored at 4 °C as “Elution.” The collected elution was again centrifuged at 16000 RPM for 10 min at 4 °C to ensure removal of any remaining wheat starch. Total, soluble, flow through, and elution samples were then analyzed via SDS-PAGE.

2.3.2 Fast Protein Liquid Chromatography

In preparation for purification of the target protein, cell paste was resuspended in Buffer A (20 mM Tris-HCl, pH 8); resuspension volume was determined by mass of the cell paste (5 mL buffer were used per 1 g cell paste). The solution was then sonicated for 10 min total (pulse: 10 s on, 15 s off) at 20% amplitude while on ice to lyse cells.

Purification of the target protein was performed via Fast Protein Liquid Chromatography (FPLC; GE Healthcare AKTA Amersham Pharmacia Biotech) using a weak anion exchanger column: HiTrap DEAE Sepharose FF 5 mL (GE Healthcare # 17-5154-01). Components of the system include: Monitor UPC-900 (GE Healthcare # 18-1129-74) for monitoring of UV absorption, pH, and conductivity; Pump P-920 (Amersham Biosciences # 18-1125-54) high precision laboratory pump; Fraction Collector Frac-950 (GE Healthcare # 18608300) for collection in 50 mL polypropylene conical tubes (Corning Falcon # 352070) and 96-well DeepWell polypropylene 2 mL microplates (Fisher Scientific # 12-566-121); a 10 mL Superloop (GE Healthcare # 19-7850-01) for loading sample volumes; and Motor Valve INV 907 for use as an injection valve (GE Healthcare # 18110840). The procedure was designed and monitored using UNICORN Control Software (GE Healthcare). Methods used for purification included the following steps: pump wash, column wash, sample loading, and a controlled linear gradient elution with Buffer B (20 mM Tris-HCl, 1 M NaCl, pH 8). Following completion of the run, the chromatogram was analyzed based on UV detection in order to select appropriate fractions to further assess via SDS-PAGE and Western blotting.

Later purification attempts used alternate columns and buffers; each combination used is summarized in Table 7.

Table 7. Summary of purification columns and conditions used indicating the type of resin, whether anion or cation exchange, and buffers used.

Column	Type	Buffer A	Buffer B
HiTrap DEAE Sepharose FF 5 mL (# 17-5154-01)	Weak Anion	20 mM Tris-HCl, pH 8	Buffer A + 1 M NaCl
HiTrap CM Sepharose FF 5 mL (# 17-5155-01)	Weak Cation	50 mM sodium acetate, pH 5.2	Buffer A + 1 M NaCl
HiTrap SP Sepharose FF 5 mL (# 17-5157-01)	Strong Cation	50 mM acetic acid, pH 4.3	Buffer A + 1 M NaCl
HiTrap DEAE Sepharose FF 5 mL (# 17-5154-01)	Weak Anion	20 mM Tris-HCl, 1 mM PMSF, pH 8	20 mM Tris-HCl, 1 M NaCl pH 8

2.4 Qualitative Activity Assay

In order to detect trypsin-like activity in purified fractions, which would suggest that the recombinant enzymes are active, the synthetic colorimetric substrate for trypsin BApNA was used. Upon cleavage of BApNA by enzymes with trypsin-like activity, a yellow product (*p*-nitroaniline) is generated with color intensity proportional to enzyme concentration.⁶⁰ While BApNA can be used as a substrate to obtain kinetic constants once a protein is successfully purified, for this work, it was used qualitatively to determine whether or not activity was present in purified fractions.

Each reaction was prepared in a 1.5 mL microcentrifuge tube and contained 20 mM Tris-HCl, pH 7.2, 10 mM CaCl₂, 1 mM BApNA, and ultrapure water for a total volume of 990 μ L. For each reaction, 10 μ L from the sample fraction in question was added and the solution vortexed to mix and incubated at room temperature overnight. Reaction mixtures were observed to assess the color intensity of solutions that showed a yellow color following incubation to determine if any samples were positive for trypsin-like activity.

CHAPTER 3: RESULTS AND DISCUSSION

3.1 Molecular Cloning

As described in the previous section, PCR amplification (Figure 14) and vector digestion (Figure 15) steps in the cloning process were qualitatively assessed via agarose gel electrophoresis.

Figure 14. 1% Agarose gel of AaSPIV PCR products using primers designed for cloning into the pET28a vector. Lanes 1 and 2 are duplicate PCR reactions at 58 °C and lane 3 is the DNA Ladder in base pairs (Thermo Scientific GeneRuler 1 kb DNA ladder). The expected PCR product is 885 bp (indicated by the arrow).

Figure 15. 1% Agarose gel of digested and dephosphorylated pET28a vector. Lane 1 is the DNA Ladder in base pairs (Thermo Scientific GeneRuler 1 kb DNA ladder). Lane 2 contains the digested and dephosphorylated pET28a vector at 5296 bp (indicated by the arrow).

Following confirmation of successful PCR and vector digestion (Figures 14 and 15), the digested and dephosphorylated vector was combined with the PCR product, ligated, and transformed into bacterial cells. Successful cloning was assessed via colony PCR (Figure 16). The AaSPII/pET28a and AaSPIV/pET28a plasmid constructs were then isolated from the bacterial cells and sequenced to confirm successful molecular cloning.

Figure 16. 1% Agarose gel of Colony PCR reactions showing positive hits of the AaSPIV gene. Colony PCR was used to verify successful molecular cloning of the AaSPIV-pET28a plasmid construct. Lane 9 is the DNA Ladder in base pairs (Thermo Scientific GeneRuler 1 kb DNA ladder). Lanes 1, 2, 3, 5, 6, and 8 showed a PCR product in the expected region (885 bp).

Cloning of all constructs of AaSPII and AaSPIV, whether it was for full-length recombinant proteases, proteases without the leader sequence, incorporation of the EK-pseudo propeptide or the MBP-fusion tag, was performed as described in the methods (see Table 6). All constructs were confirmed by sequencing (data not shown).

Sequencing results obtained for the AaSPIV-pET28a plasmid constructs revealed a change at two specific nucleotides not found in the original AaSPIV gene sequence annotated in the National Center for Biotechnology Information (NCBI) website.³⁵ The first nucleotide change led to an amino acid change at position 162 (lysine) to a glutamate

residue. The second nucleotide that did not match the annotated sequence resulted in an arginine residue at position 288 (instead of a lysine residue).³⁵ Initially, these two nucleotide changes were thought to be mutations introduced during the cloning process, especially since *Ae. aegypti* cDNA was used as the template. However, these changes were observed consistently even after multiple cloning attempts were performed, indicating the possibility that the published *Ae. aegypti* gene sequence for AaSPIV was incorrectly annotated, which has been known to occur.⁷²⁻⁷⁶ Most cases of misannotation are associated with overprediction of molecular function.⁷⁶ Additional problems associated with genome annotation include: expressed sequence tags (ESTs; short mRNA sequences derived from cDNA libraries) are susceptible to artifacts, repetitive elements in cDNA can cause false homology matches, similarities between species do not often take into account evolutionary divergence or homologs, and alignments can be severely complicated by alternative splicing.⁷⁷ To ensure that this was the case, a fresh new batch of *Ae. aegypti* cDNA was prepared (by Dr. Jun Ioe, University of Arizona) and the cloning was repeated by Mr. Bryan Tracy Conner (an undergraduate research student in the Rascón lab). His experiments resulted in the same amino acid changes observed in the original AaSPIV-pET28a constructs.

It is also well established that high genetic variability between different *Ae. aegypti* strains or between populations in different geographic regions exists, which may also account for the two observed nucleotide discrepancies.⁷⁸⁻⁸⁰ Potential variation of these two amino acid residues between *Ae. aegypti* strains may indicate that the residues are not highly conserved suggesting that the properties of these specific amino acid residues

may not be essential for folding or function.⁴² Often, features of the protein structure necessary for structure and/or function can depend on general properties associated with tertiary structure as opposed to conserved protein sequences; these can include exposure of surface polar or charged residues, which can exchange protons with water.^{42,81,82} Because the amino acids in question were annotated at lysine residues (positively-charged) and the amino acids discovered in plasmid constructs were a glutamate residue (negatively-charged) and an arginine residue (positively-charged) it is possible that residues at these positions only need to be charged—and perhaps the amino acid at position 288 needs to specifically be positively-charged—for proper structure and function of the native AaSPIV protease.

3.2 Recombinant Protein Expression

Growth experiments were performed to determine the optimal conditions for maximizing soluble recombinant expression of both AaSPII and AaSPIV. As such, multiple approaches were taken initially to compare various effects on bacterial growth and recombinant protein expression. This included using different types of media, growth temperature, induction by IPTG or auto-induction, incorporation of fusion tags like MBP or a His₆-tag, and replacing the propeptide region with an EK-pseudo propeptide.

Initially, growth experiments were conducted with pET28a plasmid constructs to yield N-terminally His₆-tagged recombinant enzymes (AaSPII without leader, AaSPIV with leader) using SHuffle T7 *E. coli* cells grown at 10 °C for a total of 64 h in either LB media (results not shown) or TB media (See Figures 17 and 18). Effects of recombinant target protein expression using LB and TB media were assessed but expression with TB

media was clearly superior (higher expression levels of total recombinant proteases). TB media (12 g/L tryptone, 24 g/L yeast extract, 9.4 g/L K_2HPO_4 , 2.2 g/L KH_2PO_4 , and 0.4% v/v glycerol) is much richer than LB media (10 g/L tryptone, 5 g/L yeast extract, and 5 g/L NaCl) and provides more effective buffering against metabolic pH changes due to the inclusion of monobasic and dibasic potassium phosphate.⁸³ For this reason, growth experiments using LB media can limit target protein expression due to acid produced via glucose metabolism surpassing buffering capabilities of LB media and halting cell growth and protein synthesis.⁸³

Lowering the temperature has been shown to reduce protein aggregation in difficult-to-express proteins, but at the same time slows bacterial growth and metabolism, and so to further increase soluble expression, growth times were extended when lower temperatures used.⁴⁷ Factors contributing to increased solubility of recombinant proteins at reduced expression temperatures include the temperature-dependence of hydrophobic interactions that can lead to protein aggregation and the prevention of bacterial heat shock protein expression that could otherwise be induced in recombinant overexpression conditions.⁴⁷ Additionally, lower temperatures slow the bacterial protein expression machinery, leading to lower expression rates of proteins and thus more chaperone proteins are available to aid in proper folding (despite optimal chaperone activity and expression $\sim 30^\circ\text{C}$).^{47,83}

Another factor that is typically problematic in recombinant protein expression is the presence of disulfide bonds in the target protein required for proper folding and solubility; both AaSPII and AaSPIV are predicted to form four disulfide bonds in the

native conformation, which may not be properly forming during growth experiments.^{47,49,84} Because correctly forming the native disulfide bonds between these eight cysteine residues (of 105 possible isomers) is essential to proper folding, SHuffle T7 competent *E. coli* cells were chosen for expression in anticipation of potential expression difficulties associated with improper folding of disulfide-bonded proteins because they express the disulfide bond isomerase DsbC, which acts as a chaperone to aid in protein folding and the correction of mismatched disulfide bonds.^{47,49,69,84} The strain also contains deletions of the genes for glutaredoxin reductase and thioredoxin reductase leading to a more oxidizing cytoplasm and facilitating proper disulfide bond formation.^{49,50,69}

As shown in Figures 17 and 18, the recombinant enzymes (AaSPII without leader and AaSPIV with leader) were successfully expressed, but as in insoluble proteins. They were only found in the total samples, which represent all of the total proteins and enzymes expressed by the bacterial cells, including the recombinant enzyme of interest.⁸⁵ No visible bands corresponding to the enzymes of interest are observed in the soluble samples. This may be due to misfolding of the target enzymes, which is a common occurrence for disulfide-bonded proteins and would result as insoluble aggregated proteins known as inclusion bodies.⁴⁷ Unfavorable protein-folding conditions caused by macromolecular crowding can lead to the formation of inclusion bodies, which occurs as a response to stress caused by high rates of recombinant protein expression.⁸⁶

Figure 17. 4-12% BIS-TRIS Gel analysis of samples collected from a 64 h growth experiment of AaSPII-NL-pET28a in SHuffle T7 *E. coli* cells grown at 10 °C using TB media. Lane 1 is the protein Molecular Weight (MW) ladder (Thermo Scientific PageRuler (in kDa)). From left to right, lanes include total and soluble samples at various time-points (labeled by h). Total samples show target protein expression at the expected MW region (~32 kDa).

Figure 18. 4-12% BIS-TRIS Gel analysis of samples collected from a 64 h growth experiment of AaSPlV-pET28a in SHuffle T7 *E. coli* cells grown at 10 °C using TB media. Lanes 1 and 8 are the protein Molecular Weight (MW) ladder (Thermo Scientific PageRuler, in kDa). From left to right, other lanes include total and soluble samples at various time-points (labeled by h). Total samples show target protein expression at the expected MW region (~32 kDa).

Additionally, the metabolic burden of recombinant protein expression imparted on bacterial cells often limits cell growth; growth rates of bacterial cells containing vector plasmids suffer due to the added strain of plasmid replication and cellular translation and transcription mechanisms.^{85,87,88} Therefore, expression via auto-induction was pursued in an attempt to increase soluble protease expression; because auto-induction techniques enable bacterial cells to grow to high density before expression is induced, higher total target protein expression levels can be achieved by limiting the metabolic stress bacterial cells can undergo during recombinant protein expression.^{70,71} To determine if this would be the case for AaSPII (no leader) and AaSPIV (with leader), both enzymes were expressed under auto-induction conditions (auto-induction media) in SHuffle T7 *E. coli* cells at 10 °C for a total of 96 h (Figures 19 and 20).

Unfortunately, soluble expression of AaSPII (no leader) and AaSPIV (with leader) using SHuffle T7 cells and auto-induction media (Figures 19 and 20) showed no visible bands in the expected region associated with the size of target proteins. Both were found only in the total samples. Additionally, these bands appeared to be less intense (lower expression) than those associated with expression using TB media and IPTG induction (Figures 17 and 18). However, initial comparison of the expression of these recombinant enzymes in TB and in auto-induction media is somewhat misleading: while the expression level of target protein (based on band intensity) was higher for IPTG-induced growth samples, the overall yield of all expressed proteins was lower. Because auto-induced cultures grew to notably higher cell densities, the overall yield of all expressed proteins was much higher for auto-induced growth samples.⁷⁰ Because

recombinant protein expression is suppressed in the auto-induction process until glucose and glycerol in the media are depleted, cells can grow to high densities without suffering metabolic stress before recombinant protein production begins.⁷⁰ Additionally, the rich auto-induction media (1% N-Z-amine, 0.5% yeast extract, 25 mM Na₂HPO₄, 25 mM KH₂PO₄, 50 mM NH₄Cl, 5 mM Na₂SO₄, 2 mM MgSO₄, 0.5% glycerol, 0.05% glucose, and 0.2% lactose) is designed for higher availability of bacterial nutrient sources as well as better pH regulation for optimal bacterial cell growth until nutrient sources are depleted and protein expression is auto-induced by lactose present in the media.⁷⁰ Although growth experiments under auto-inducing conditions did not seem to promote soluble protein expression as hoped, increased yields of insoluble target proteins would be an improvement if later refolding methods are used.^{47,89} Because refolding techniques are consistently inefficient, higher initial protein yields would maximize final yields.^{47,89}

Figure 19. 4-12% BIS-TRIS Gel analysis of samples collected from a 96 h growth experiment of AaSPII-NL-pET28a in SHuffle T7 *E. coli* cells grown at 10 °C using auto-induction. Lane 1 is the protein Molecular Weight (MW) ladder (Thermo Scientific PageRuler, in kDa). From left to right, other lanes include total and soluble samples at various time-points (labeled by h). Total samples show target protein expression at the expected MW region (~32 kDa).

Figure 20. 4-12% BIS-TRIS Gel analysis of samples collected from a 96 h growth experiment of AaSPIV-pET28a in SHuffle T7 *E. coli* cells grown at 10 °C using auto-induction. Lanes 1 and 8 are the protein Molecular Weight (MW) ladder (Thermo Scientific PageRuler, in kDa). From left to right, other lanes include total and soluble samples at various time-points (labeled by h). Total samples show target protein expression at the expected MW region (~32 kDa)

Since the described attempts at expressing soluble recombinant AaS^{PII} and AaS^{PIV} (media, temperature, and bacterial cells) failed, the decision was made to determine if modifying the natural propeptide region with an EK-pseudo propeptide region would aid in the soluble expression of AaS^{PII} by preventing possible autoactivation of the recombinant enzymes, but also in removing eukaryotic nucleotides rarely found in *E. coli*.^{23,90,91} When foreign mRNA contains excessive codons that are rare in *E. coli*, less abundant tRNAs can be depleted, leading to incorporation of the wrong amino acid or truncation, both of which cause decreased target protein expression levels. The EK-pseudo propeptide plasmid construct was expressed in SHuffle T7 *E. coli* cells grown under two different sets of conditions: (1) auto-induced and grown in auto-induction media at 30 °C for 24 h (Figure 21) or (2) induced by IPTG and grown in TB media at 15 °C for 72 h (Figure 22).

It has been proposed that the lack of an enterokinase cleavage site in the zymogen form of *Ae. aegypti* trypsin-like serine proteases homologous to mammalian trypsinogen is indicative of autocatalytic behavior; but, this has not yet been shown definitively.⁴⁴ Additionally, other abundant *Ae. aegypti* serine proteases of study in the Rascón lab have been shown to be autocatalytic (results not published). Incorporation of the EK-pseudo propeptide region at the N-terminus of express recombinant target proteins was performed for two main purposes: (1) to overcome potential autoactivation of recombinant AaS^{PII} or AaS^{PIV} proteases in addition to autodigestion and thus contribute to difficulties in obtaining soluble expression and (2) to facilitate *in vitro* activation post purification.²³

Figure 21. 4-12% BIS-TRIS Gel analysis of samples collected from a 24 h growth experiment of EK-AaSPII-pET28a in SHuffle T7 *E. coli* cells grown at 30 °C using auto-induction. Lanes 1 and 8 are the protein Molecular Weight (MW) ladder (Thermo Scientific PageRuler, in kDa). From left to right, other lanes include total and soluble samples at various time-points (labeled by h). Total and soluble samples show target protein expression at the expected MW region (~32 kDa).

Figure 22. 4-12% BIS-TRIS Gel analysis of samples collected from a 72 h growth experiment of EK-AaSPII-pET28a in SHuffle T7 *E. coli* cells grown at 15 °C using auto-induction. Lane 1 is the protein Molecular Weight (MW) ladder (Thermo Scientific PageRuler, in kDa). From left to right, other lanes include total and soluble samples at various time-points (labeled by h). Total samples show target protein expression at the expected MW region (~32 kDa).

SDS-PAGE results from the two growth experiments (Figure 21, Figure 22) to express the EK-pseudo propeptide plasmid construct did not show high expression levels in either total or soluble samples. However, soluble expression of AaS_{PII} with the EK-pseudo propeptide at 30 °C using auto-induction media (Figure 21) was observed in the predicted region of the gel (~32 kDa). Since growth under 30 °C auto-induction conditions showed some soluble target protein expression (Figures 21 and 23), the growth experiment was repeated but, rather than grow at 30 °C only, the temperature was lowered to 15 °C after reaching an OD₆₀₀ of approximately 0.8 in an effort to increase soluble expression; unfortunately, there was no visible expression in SDS-PAGE or Western blot analyses of later growth experiments using auto-induction methods performed at 15 °C (results not shown). Because the growth temperature was lowered in an attempt to increase soluble expression, a loss of soluble expression at lower temperatures (Figure 20) was unexpected. However, this may be due to bacterial protein chaperone activity and expression, which is typically optimal in *E. coli* at 30 °C.⁴⁷ While heat shock proteins can function as chaperones at high temperatures, cold shock proteins typically function as RNA chaperones at low temperatures.^{92,93} Additionally, beneficial effects of low growth temperatures can become overwhelmed by the lowered transcription and translation efficiency of *E. coli*.^{92,93}

In order to determine if low levels of soluble AaS_{PII} with the EK-pseudo propeptide (Figure 20) were expressed at 30 °C using auto-induction techniques and confirm if bands associated with total samples in the appropriate molecular weight region could indeed be attributed to target protein expression, analysis of growth experiment time-points by

Western blot was performed (Figure 23). Because Western blotting uses antibody-antigen interactions to visualize expression of the target protein, it is a much more specific and sensitive technique (compared to visualization by SDS-PAGE, which is not specific for a particular protein); eliminating visualization of non-specific background bacterial protein expression would increase confidence in confirming expression of the target protein.⁴²

Figure 23. Western blot analysis of samples collected from a 24 h growth experiment of EK-AaSPII-pET28a in SHuffle T7 *E. coli* cells grown at 30 °C using auto-induction. Lane 1 is the protein Molecular Weight (MW) ladder (Thermo Scientific PageRuler, in kDa). From left to right, other lanes include total and soluble samples at various time-points (labeled by h). Total and soluble samples show target protein expression at the expected MW region (~32 kDa).

As shown in Figure 23, Western blot analysis revealed successful expression of the His₆-tagged protein at low levels in both total and soluble fractions. In the later time-point samples (taken at 18 and 24 h), there are also bands present in lower molecular weight regions. This may indicate that, despite incorporation of the EK-pseudo propeptide region, the expressed proteins have autoactivated and begun autodigesting, resulting in the presence of smaller molecular weight peptides containing the His₆-tag. If this is the case for AaS^{PII} and AaS^{PIV}, despite incorporation of the EK-pseudo propeptide, it may be due to the presence of the lysine residue at the cleavage site, which is also present in the wild-type protease; if autoactivation of the proteases is facilitated by the presence of the lysine residue (which is consistent with trypsin substrate specificity), incorporation of the EK cleavage site (DDDDK) would not prevent autoactivation.⁹⁴ Aside from the amino sequence of the propeptide region, one factor that can contribute to autoactivation of mammalian trypsinogen includes modification of the protein surface charge since autoactivation is promoted by protein-protein interactions.⁹⁵

Additionally, comparison of the SDS-PAGE results (Figure 21) with the complementary Western blot (Figure 23) for expression of AaS^{PII} containing the EK-pseudo propeptide region in SHuffle T7 *E. coli* cells grown at 30 °C under auto-induction conditions shows that, despite achieving soluble expression, the expression level (band intensity) of the target protein is similar to expression levels of bacterial proteins. Because target protein yields were relatively minimal, the AaS^{PII} (no leader) and full-length AaS^{PIV} (with leader) genes were each cloned into the pMAL-c5E vector for expression of the protein using the pET28a plasmid constructs as templates. Expressing

recombinant enzymes as MBP-tagged proteins has been shown dramatically increase the expression levels of soluble recombinant proteins and use of this method was performed in hopes of achieving similar results for AaSPII (no leader) and full length AaSPIV (with leader).⁴⁷ Using the pET28a plasmid constructs as templates for insertion allowed for incorporation of the His₆-tag between the N-terminal MBP and C-terminal recombinant enzyme (Figure 13); incorporation of the His₆-tag should not interfere with expression.⁹⁶ Additionally, the use of MBP and the His₆-tag provide two potential downstream methods of purification (exploitation of MBP-amylose binding or of metal affinity to the His₆-tag); optimal purification methods for obtaining the higher purified protein yields can be determined by using one or both of the two methods.⁹⁶ In many previous instances, MBP for high soluble expression levels in combination with the robust method IMAC method of purification has been shown to be very effective for difficult-to-express proteins.⁹⁶

MBP-tagged AaSPII without the leader sequence (MBP-AaSPII) and MBP-tagged full length AaSPIV with the leader sequence (MBP-AaSPIV) were expressed in LB media at 15 °C for 24 h (MBP-AaSPII; Figure 24) or for 48 h (MBP-AaSPIV; Figure 25). Additionally, to determine if soluble expression could be increased by altering growth time, temperature, and media, the growth experiment to express MBP-AaSPIV was repeated using the AaSPIV-pMAL plasmid construct in SHuffle T7 cells grown in TB media at 12 °C for 77 h; results were assessed via His-Stain procedure to reduce visualization of background protein expression in the gel (Figure 26).

Figure 24. 4-12% BIS-TRIS Gel analysis of samples collected from a 24 h growth experiment of AaSPlI-pMAL-c5E in SHuffle T7 *E. coli* cells grown at 15 °C using LB media. Lanes 1 and 7 are the protein Molecular Weight (MW) ladder (Thermo Scientific PageRuler, in kDa). From left to right, other lanes include total and soluble samples at various time-points (labeled by h). Total and soluble samples show target protein expression at the expected MW region (~75 kDa).

Figure 25. 4-12% BIS-TRIS Gel analysis of samples collected from a 48 h growth experiment of AaSPIV-pMAL-c5E in SHuffle T7 *E. coli* cells grown at 15 °C using LB media. Lanes 1 and 8 are the protein Molecular Weight (MW) ladder (Thermo Scientific PageRuler, in kDa). From left to right, other lanes include total and soluble samples at various time-points (labeled by h). Total and soluble samples show target protein expression at the expected MW region (~75 kDa).

Figure 26. 4-12% BIS-TRIS Gel analysis of soluble samples collected from a 77 h growth experiment of AaSPIV-pMAL-c5E in SHuffle T7 *E. coli* cells grown at 12 °C using TB media. Lane 1 is the protein Molecular Weight (MW) ladder (Thermo Scientific PageRuler, in kDa). From left to right, other lanes include soluble samples at various time-points (labeled by h). Samples show target protein expression at the expected MW region (~75 kDa).

As expected, growth experiments with MBP-tagged recombinant enzymes (Figures 24, 25, and 26) significantly increased expression levels in both total and soluble samples. Results showed that expression of pMAL-c5E plasmid constructs in SHuffle T7 cells grown at 15 °C in LB media (Figures 24 and 25) resulted in similar expression of MBP-AaSPII and MBP-AaSPIV as compared to growth in TB media (results not shown). Expression with the MBP-fusion tag using LB media seemed to be sufficient; potential buffering problems with LB that could lead to poor soluble target protein yield were counterbalanced by the improved expression due to the MBP-fusion tag while expression of background bacterial proteins was limited, which could aid in downstream purification processes, and so LB media was used to express MBP-AaSPII or MBP-AaSPIV in later growths.⁸³

Extending growth times also did not seem to significantly increase soluble protein expression; expression of pMAL-c5E plasmid constructs in SHuffle T7 *E. coli* cells at 15 °C in LB media for 24 h (MBP-AaSPII; Figure 24) or for 48 h (MBP-AaSPIV; Figure 25) were sufficient to maximize soluble target protein yields as predicted, due to the robustness of the MBP-fusion tag for use as a solubility enhancer.^{47,96}

3.3 Initial Purification Attempts and Qualitative Activity Assay

For both enzymes (AaSII without the leader sequence and AaSIV with the leader sequence), expression using the MBP-fusion tag (pMAL-c5E plasmid constructs) resulted in soluble recombinant protein expression; and so, this method was used for large-scale protein expression for purification. As mentioned, one benefit of using the pMAL-c5E plasmid constructs is that we were able to clone the genes with the N-terminal His₆-tag in

place, thereby creating the MBP-fusion protein (Figure 13), both of which can be exploited in the purification process.⁹⁶ Initial purification attempts of MBP-AaSPIV (with leader) were performed using the wheat starch purification protocol previously outlined. Unfortunately, SDS-PAGE analysis of fractions obtained throughout the purification process indicated purification was unsuccessful (Figure 27). SDS-PAGE results show a visible band of appropriate size and intensity associated with the total sample; however, the band is no longer visible in the soluble sample or in any of the later flow through or elution samples (see Figure 27).

Figure 27. 4-12% BIS-TRIS Gel analysis of fractions collected from wheat starch purification of MBP-AaSPIV recombinant protease expressed in SHuffle T7 cells grown at 15 °C for 48 h in LB media. Lane 1 is the protein Molecular Weight (MW) ladder (Thermo Scientific PageRuler, in kDa). From left to right, other lanes include total and soluble samples three subsequent flow through samples (FT 1, FT 2, and FT3), and the elution sample. Only the total sample shows target protein expression at the expected MW region (~75 kDa).

The observed absence of MBP-AaSPIV in the soluble fractions (Figure 27), despite evident soluble expression in the large-scale growth (Figure 25), suggests that the MBP-fusion proteases may be autoactivating.⁴⁴ Interestingly, results of the His-Stain procedure used to visualize expression of the His₆-tagged MBP-AaSPIV recombinant fusion protein revealed potential autoactivation at low levels due to the presence of a protein species at around 50 kDa (Figure 26). From the His-Stained gel, high levels of soluble expression are observed at the expected molecular weight of ~75 kDa, but upon careful analysis, the presence of faint bands at the lower molecular weight (~50 kDa) becomes evident. The major reason visualization by His-Stain is used is because this stain eliminates bands associated with background bacterial protein expression, especially those that lack the His₆-tag, therefore making it easier to distinguish between native bacterial proteins and the target His₆-tagged recombinant enzymes.

Autoactivation leading to cleavage at the predicted cleavage site would result in free, active soluble recombinant AaSPIV (MW ~28 kDa) and free His₆-tagged MBP (MW ~47 kDa) (see scheme in Figure 28) (combined MW of the fusion protein ~75 kDa), effectively preventing purification of the target enzyme using wheat starch, which has a fair amount of amylose. As mentioned, MBP will naturally bind to amylose, and it is expected that free MBP should be visible in the elution fraction after the addition of maltose. However, there is nothing visible in the elution sample (Figure 27), which may support the autoactivation of AaSPIV, in which this active protease is cleaving and digesting the His₆-tagged MBP.⁴⁴

Figure 28. Scheme showing cleavage of the MBP–His₆-tag–Protein fusion product following autoactivation, resulting in free His₆-tagged MBP and free active target protein.

Initial failed attempts using the wheat starch batch purification method prompted purification trials using ion exchange chromatography to remove contaminants and purify based on the pI of the MBP-fusion protease. Typically, multiple purification steps are performed in succession to recover pure target proteins; in these cases, inclusion of the MBP (which can bind to wheat starch and amylose resin) and the His₆-tag (recoverable via IMAC) presents possibilities for future purification steps in addition to other options such as size-exclusion chromatography.⁹⁶ However, since the initial wheat starch method was unsuccessful, ion exchange was attempted using the following specifications: MBP-AaSPII and MBP-AaSPIV expressed in SHuffle T7 cells grown at 15 °C for 48 h in LB media purified via FPLC using a DEAE (weak) anion exchange column in 20 mM Tris-HCl, pH 8 buffer (Figures 29 and 30).

Figure 29. 4-12% BIS-TRIS Gel analysis of fractions collected from purification via FPLC using DEAE anion exchange column in 20 mM Tris-HCl, pH 8 buffer of MBP-AaSPII recombinant protease expressed in SHuffle T7 cells grown at 15 °C for 48 h in LB media. Lanes 1 and 21 are the protein Molecular Weight (MW) ladder (Thermo Scientific PageRuler, in kDa). From left to right, other lanes include total and soluble samples, two subsequent flow through samples (FT 1 and FT 2), and elution samples labeled according to the corresponding location in the 96-well collection microplate. Only the total sample shows target protein expression at the expected MW region (~75 kDa).

Figure 30. 4-12% BIS-TRIS Gel analysis of fractions collected from purification via FPLC using DEAE anion exchange column in 20 mM Tris-HCl, pH 8 buffer of MBP-AaSPIV recombinant protease expressed in SHuffle T7 cells grown at 15 °C for 48 h in LB media. Lanes 1 and 23 are the protein Molecular Weight (MW) ladder (Thermo Scientific PageRuler, in kDa). From left to right, other lanes include total and soluble samples, two subsequent flow through samples (FT 1 and FT 2), and elution samples labeled according to the corresponding location in the 96-well collection microplate. Only the total sample shows target protein expression at the expected MW region (~75 kDa).

Initial results from the anion exchange purification of MBP-AaSPII and MBP-AaSPIV recombinant enzymes (Figures 29 and 30) were similar to those observed using the wheat starch purification method (Figure 27): a band of the correct size and intensity (~75 kDa) is visible in the total sample (indicative of the MBP-fusion target enzyme), but not present in the soluble sample or in later fractions taken throughout the flow through or elution process. This outcome, in combination with corresponding previous results of wheat starch purification methods, provides consistent evidence in support of the potential autocatalytic behavior of AaSPII and AaSPIV.⁴⁴ In an attempt to confirm if autoactivation and/or autodigestion of the recombinant enzymes is indeed occurring, analysis of FPLC fractions by Western blotting was performed to locate any cleaved His₆-tagged peptides in the fractions (Figure 31).

Figure 31. Western blot analysis of fractions obtained from purification attempt via FPLC using DEAE anion exchange column in 20 mM Tris-HCl, pH 8 buffer of MBP-AaSPIV expressed in SHuffle T7 cells grown at 15 °C for 48 h in LB media. Lane 1 is the protein Molecular Weight (MW) ladder (Thermo Scientific PageRuler, in kDa). From left to right following the ladder, lanes include total and soluble samples, two subsequent flow through samples (FT 1 and FT 2), and elution samples labeled according to the corresponding location in the 96-well collection microplate. Total, soluble, and fractions B8, B12, and C4 show His₆-tagged protein expression at a much lower MW region than expected (~50 kDa observed, ~75 kDa expected).

Results of the Western blot (Figure 31) from collected purified fractions of MBP-AaSPIV using DEAE clearly show a protein species with a His₆-tag in the soluble sample and in later elution fractions at a much lower molecular weight than anticipated (~50 kDa observed, rather than the ~75 kDa that is expected). This provided further evidence that the enzyme is autocatalyzing at the propeptide cleavage region resulting in the loss of the His₆-tag on the target protease, but cleaving the site to remain on the MBP (Figure 28). This initial hypothesis is based on amino acid sequence analysis suggesting the likely location of protein cleavage leading to autoactivation: two protein species would potentially be produced: MBP with the His₆-tag and propeptide region of the enzyme (~47 kDa) and the active, recombinant enzyme (~28 kDa for both AaSPII and AaSPIV). The only probable conclusion that can be made is that autoactivation is occurring during the purification process, leading to free, activated AaSPIV recombinant enzyme (as shown in the scheme depicted in Figure 28) and the MBP with the His₆-tag on the C-terminus.⁴⁴ Overlaying the results of the Western blot (Figure 31) with those of the complementary SDS-PAGE (Figure 30) shows very clearly in which purification fractions the cleaved His₆-tagged protein appears and at what approximate concentrations based on band intensity (Figure 32).

Figure 32. Overlay of Western blot (Figure 31) and complementary SDS-PAGE (Figure 30) performed on fractions obtained from purification attempt via FPLC using DEAE anion exchange column in 20 mM Tris-HCl, pH 8 buffer of AaSPIV-pMAL expressed in SHuffle T7 cells grown at 15 °C for 48 h in LB media.

The results observed from the wheat starch and anion exchange purification is not at all too surprising since mammalian trypsins have been shown to autoactivate at slightly alkaline conditions (pH ~7–9) and can digest other proteins in the cytosol once active.⁴⁶ In addition, AaSPII and AaSPIV, as well as other *Ae. aegypti* serine proteases, lack the enterokinase cleavage site present in mammalian trypsins (the main autocatalytic

regulation mechanism) indicating that autoactivation is the likely mechanism responsible for activation of these digestive mosquito enzymes.^{22,36,44,46,61} Thus far, the Rascón lab has successfully determined that the most abundant midgut proteases are indeed autocatalytic (unpublished results).

While autoactivation of the MBP-fusion recombinant proteases presents extensive purification challenges, it indicates that the recombinant target enzymes are properly folded and active. Thus, based on these results, we can confidently suggest that successful soluble expression was achieved and the native disulfide bonds found in AaSPII and AaSPIV were successfully formed in the folding process. While it also seems reasonable to expect that the bulky MBP-fusion tag might inhibit autoactivation (which depends on successful protein-protein contacts) and any subsequent enzyme activity, this was surprisingly not the case. However, previous literature does implicate that other enzymes expressed with an MBP-fusion tag have also been shown to retain activity comparable to the free, active enzyme and were not inhibited by the presence of the MBP-fusion protein.^{95,97-99} In order to confirm the presence of trypsin activity in any of the fractions and to determine which fractions might contain the target proteases, a qualitative enzyme activity assay using the chromogenic trypsin peptide substrate analog BApNA was performed on each DEAE purified fraction, as described in the Methods section. As shown in Figure 33, the soluble fraction and later elution fractions (corresponding to C6–D5, Figure 30) were positive for trypsin activity (yellow).

Figure 33. Qualitative enzyme activity assay reaction samples using BApNA to detect the presence of trypsin-like activity in fractions collected from purification via FPLC using DEAE anion exchange column in 20 mM Tris-HCl, pH 8 buffer of AaSPIV-pMAL expressed in SHuffle T7 cells grown at 15 °C for 48 h in LB media. From left to right and top to bottom, reaction samples are the soluble fraction and elution samples labeled according to the corresponding location in the 96-well collection microplate. A yellow color indicative of trypsin-like activity is visible in the soluble sample and in reaction tubes labeled C6, C8, C11, D2, and D5.

While no negative control is shown here, similar assays have been performed by members of the Rascón lab using total and soluble bacterial protein samples without expression of the recombinant serine proteases resulting in no color change or indication of trypsin activity. Confirmation of trypsin activity in purification fractions seems promising but also indicates further challenges associated with obtaining purified enzyme. Specifically, cleavage at the predicted cleavage site results in free, active enzyme without the His₆-tag or MBP attached, both of which could have been used to facilitate further purification steps.⁹⁶ Additionally, the lack of any bands present with high intensity at the appropriate size (~28 kDa) is likely due to autodigestion, which would likely begin immediately following activation.⁴⁶ Potential autodigestion of AaSPII and AaSPIV in addition to autoactivation further contributes to difficulties associated both with purifying the enzymes and downstream experiments to perform activity assays and obtain kinetic data.⁴⁴

3.4 Alternative Purification Approaches

In an effort to achieve purification of the recombinant enzymes, while preventing previously observed autoactivation and autodigestion, several alternative purification approaches will be pursued. Previous purification methods used buffers with pH values within the same range as that associated with optimal trypsin activity (pH 7–8).⁴¹ Since enzyme activity is often dependent on pH, later purification processes will use low pH buffer systems in an effort to prevent the recombinant proteases from being active and avoid autoactivation that may occur during the purification preparation process.⁴¹ An initial attempt using this strategy focused on two cation exchange columns: (1) a CM

weak cation exchange column using 50 mM sodium acetate, pH 5.2 to purify MBP-AaSPII expressed in SHuffle T7 cells grown at 15 °C for 48 h in LB media (Figure 34) and (2) an SP strong cation exchange column using 50 mM acetic acid, pH 4.3 buffer to purify MBP-AaSPIV expressed in SHuffle T7 cells grown at 15 °C for 48 h in LB media (Figure 35).

Figure 34. 4-12% BIS-TRIS Gel analysis of fractions collected from purification via FPLC using a CM weak cation exchange column in 50 mM sodium acetate, pH 5.2 buffer of MBP-AaSPII expressed in SHuffle T7 cells grown at 15 °C for 48 h in LB media. Lanes 1 and 17 are the protein Molecular Weight (MW) ladder (Thermo Scientific PageRuler, in kDa). From left to right, other lanes include total and soluble samples, two subsequent flow through samples (FT 1 and FT 2), and elution samples labeled according to the corresponding location in the 96-well collection microplate. Total, soluble, and the first flow through (FT 1) samples show target protein expression at the expected MW region (~75 kDa).

Figure 35. 4-12% BIS-TRIS Gel analysis of fractions collected from purification via FPLC using a SP cation exchange column in 50 mM acetic acid, pH 4.3 buffer of MBP-AaSPIV expressed in SHuffle T7 cells grown at 15 °C for 48 h in LB media. Lane 1 is the protein Molecular Weight (MW) ladder (Thermo Scientific PageRuler, in kDa). From left to right, other lanes include total and soluble samples, two subsequent flow through samples (FT 1 and FT 2), and elution samples labeled according to the corresponding location in the 96-well collection microplate. Total, soluble, and the first flow through (FT 1) samples show target protein expression at the expected MW region (~75 kDa).

Analysis of fractions collected throughout the purification using these two cation exchange strategies [(1) CM weak cation exchange (Figure 34) and (2) SP strong cation exchange (Figure 35)] suggest that the recombinant, MBP-tagged enzymes remain soluble and do not autoactivate at low pH. However, the MBP-fusion proteins of interest eluted in the flow through (FT 1) indicating that the fusion proteins did not interact with either column. Interestingly, it appears that using pH 5.2 is enough to reduce and potentially eliminate protease autoactivation, resulting in a more positively-charged

histidine residue in the active, which would no longer act as an effective proton acceptor to generate the active site serine alkoxide ion needed for catalytic activity.⁴² While autoactivation appears to have been prevented, it also seems that other bacterial proteins in the lysate have also eluted in the flow through; despite the enzymes remaining inactive during both purification attempts (Figures 34 and 35), purification at low pH does not seem to be an effective method of removing excess bacterial proteins. Perhaps in combination with other purification methods, this could be used for an initial, crude purification step or purification attempts with even lower pH buffers could be used with a strong cation exchange column. Since the MBP-fusion complex has a calculated pI value of 5.50 (MBP-AaSPII) or 5.38 (MBP-AaSPIV), lowering the pH even further to achieve tighter binding of the complex to the column would result in a later elution time for the complex.

Because purification via FPLC using DEAE anion exchange column in 20 mM Tris-HCl, pH 8 buffer of MBP-tagged AaSPII or AaSPIV (Figures 29 and 30) resulted in both the MBP and the recombinant enzymes eluting in later fractions (not the flow through), it may be a more effective method of purification, only if autoactivation can be prevented. Therefore, this purification procedure was re-attempted under the same conditions with the addition of 1 mM phenylmethylsulfonyl fluoride (PMSF) to the crude lysate preparation. PMSF is an irreversible, covalent serine protease inhibitor and adding it to the purification buffer should prevent any potential autoactivation and subsequent digestion from occurring. Results of this purification were assessed via SDS-PAGE (Figure 36) and Western blot (Figure 37).

Figure 36. 4-12% BIS-TRIS Gel analysis of fractions collected from purification via FPLC using DEAE anion exchange column in 20 mM Tris-HCl, pH 8 buffer of MBP-AaSPIV recombinant protease expressed in SHuffle T7 cells grown at 15 °C for 72 h in LB media and prepared with 1 mM PMSF added to the crude lysate preparation. Lanes 1 and 24 are the protein Molecular Weight (MW) ladder (Thermo Scientific PageRuler, in kDa). From left to right, other lanes include total and soluble samples, two subsequent flow through samples (FT 1 and FT 2), and elution samples labeled according to the corresponding location in the 96-well collection microplate. Only the total sample shows target protein expression at the expected MW region (~75 kDa).

Figure 37. Western blot analysis of fractions obtained from purification attempt via FPLC using DEAE anion exchange column in 20 mM Tris-HCl, pH 8 buffer of MBP-AaSPIV (with leader) expressed in SHuffle T7 cells grown at 15 °C for 72 h in LB media and prepared with 1 mM PMSF added to the crude lysate preparation. Lane 1 is the protein Molecular Weight (MW) ladder (Thermo Scientific PageRuler, in kDa). From left to right following the ladder, lanes include total and soluble samples, two subsequent flow through samples (FT 1 and FT 2), and elution samples labeled according to the corresponding location in the 96-well collection microplate. Total, soluble, and fractions B5, B9, C2, and C7 show His₆-tagged protein expression at a much lower MW region than expected (~50 kDa observed, ~75 kDa expected).

As shown in Figures 36 and 37, results from this purification of MBP-AaSPIV (with leader, expressed in SHuffle T7 cells grown at 15 °C for 72 h in LB media via FPLC using a DEAE weak anion exchange column in 20 mM Tris-HCl, pH 8 buffer with the addition of 1 mM PMSF to the crude lysate preparation) were quite unexpected: analysis of sample fractions shows that cleavage of the MBP-tagged protein still occurs despite addition of PMSF, yielding a His₆-tagged protein region consistent in size with the 47 kDa His₆-tagged MBP expected after autoactivation. PMSF is an effective irreversible serine protease inhibitor so it is unclear why this recombinant enzyme would remain active after addition of PMSF to the purification buffer and appear to have no effect on limiting autoactivation. However, it has been suggested that PMSF is not as specific for serine proteases as is often assumed and there are examples of serine proteases that are not inhibited by PMSF.¹⁰⁰ Results from a qualitative activity assay (not shown) using BApNA showed that trypsin activity in fractions associated with this purification attempt (Figure 36) was not inhibited by the presence of PMSF. Also, predictions of protease function based on sequence analyses are sometimes found to be incorrect upon obtaining experimental functional data so it may be possible that AaSPII or AaSPIV are not trypsin-like enzymes despite sequence alignments suggesting so; while sequence analyses can be powerful tools for designing experimentation and predicting results, assumptions based on sequence homology assessments are somewhat uncertain until enzyme function and structure are experimentally determined.^{82,101} Previous study of one of the abundant *Ae. aegypti* serine proteases, late trypsin (AaLT), found that, despite identification of AaLT as a trypsin based on deduced amino acid sequence and sequence alignment data,

reduction of AaLT transcripts did not result in a loss of trypsin-like activity, as was predicted.²² It was later determined that AaLT lacked the necessary aspartate residue at position 189 necessary for substrate recognition and specificity of trypsins and, based on structural modeling data, AaLT is likely to be more accurately defined as a serine collagenase and not a classic trypsin.^{23,35}

CHAPTER 4: CONCLUSIONS

4.1 AaSPIV Genome Sequence

Initial attempts taken to isolate the AaSPIV gene from *Ae. aegypti* cDNA led to the presence of two specific nucleotide changes that resulted in specific amino acid changes with respect to the annotated sequence at position 162 (a lysine to a glutamate residue) and at position 288 (a lysine to an arginine).³⁵ Repeating the isolation of the AaSPIV gene from a freshly prepared batch of *Ae. aegypti* whole body mosquito cDNA confirmed that these changes are indeed consistent with results from initial cloning attempts. This indicated potential misannotation of the gene sequence in public databases or differences due to genetic variability of the *Ae. aegypti* species; it is possible that these two residues are nonessential for function or flexible in the electrostatic or steric property requirements for proper folding and activity and therefore are not highly conserved. Hence, these specific inconsistencies between the AaSPIV gene sequence cloned into each plasmid construct and the annotated AaSPIV gene sequence were disregarded for all later experiments and studies herein proceeded; upon solving the structure of AaSPIV, the identity of these two amino acid residues may be confirmed and their role in the enzyme's structure and function can be further assessed to determine if they are expected to be highly conserved.

4.2 Recombinant Protein Expression

All growth experiments were performed using SHuffle T7 *E. coli* cells to aid in proper folding of recombinant AaSPII or AaSPIV. Both of these mosquito proteases are predicted to contain four disulfide bonds and this engineered bacterial species is designed

to have a more oxidized cytosolic environment and it expresses a disulfide bond isomerase (DsbC) to aid in forming native disulfide bonds when cysteine residues are mis-bonded.⁶⁹ Several growth experiment conditions resulted in the insoluble expression of the target enzymes: AaSPII (no leader)-pET28a construct (N-terminally His₆-tagged) grown at 10 °C using TB media (Figure 17) or in auto-induction media (Figure 19); AaSPIV (with leader)-pET28a construct (N-terminally His₆-tagged) grown at 10 °C in TB media (Figure 18) or in auto-induction media (Figure 20); and AaSPII-pET28a with the EK-pseudo propeptide region construct (N-terminally His₆-tagged) grown at 15 °C in auto-induction media (Figure 22). Surprisingly, when this construct was grown at a higher temperature (30 °C for 24 h) soluble expression was observed (Figure 23). Additionally, total expression levels for each of these growth experiments did not appear to be very high compared to the naturally expressed bacterial proteins. In biochemically studying a target protein, large yields of solubly expressed protein are necessary since large quantities are required for activity assays or structure determination. Therefore, further experimentation was performed in an attempt to achieve solubly expressed AaSPII and AaSPIV recombinant proteases.⁴⁷

After a multitude of growth experiments aimed at optimizing conditions for recombinant protein expression, I was able to establish a highly effective method for expressing significant concentrations of soluble recombinant proteases. For both AaSPII (no leader) and AaSPIV (with leader), soluble recombinant protein expression was achieved by expressing each enzyme as an MBP-fusion tag using the pMAL-c5E vector. Specifically, soluble expression was maximized using the following conditions:

MBP-tagged AaS^{PII} without the leader sequence (MBP-AaS^{PII}) and MBP-tagged full length AaS^{PIV} with the leader sequence (MBP-AaS^{PIV}) were expressed using SHuffle T7 *E. coli* cells grown in LB media at 15 °C for 24 h (MBP-AaS^{PII}; Figure 24) or for 48 h (MBP-AaS^{PIV}; Figure 25). Growth experiments using the pMAL-c5E constructs were consistently successful in producing high levels of total and soluble target protein and were not drastically affected by lowering the growth temperature as far as 10 °C or extending growth times to up to 96 h (results not shown). For these reasons, we focused on the soluble expression of the MBP-fusion proteases for later purification experiments.

4.3 Possible Autocatalytic Activity of AaS^{PII} and AaS^{PIV}

Purification attempts seem to indicate that AaS^{PII} and AaS^{PIV} may be autoactivating⁴⁴: (1) results of the SDS-PAGE analysis of fractions collected from the purification of MBP-AaS^{PIV} using the wheat starch purification method (Figure 27); (2) results of the SDS-PAGE analysis of fractions collected from the purification of MBP-AaS^{PII} via FPLC using a DEAE anion exchange column in 20 mM Tris-HCl, pH 8 buffer (Figure 29); and (3) results of the SDS-PAGE (Figure 30) and (4) Western blot (Figure 31) analyses of fractions collected from the purification of MBP-AaS^{PIV} via FPLC using a DEAE anion exchange column in 20 mM Tris-HCl, pH 8 buffer. It is also likely that the proteases undergo autodigestion in addition to autoactivation. Trypsin activity was observed using the qualitative activity assay with BApNA (Figure 33) to assess fractions collected from the purification of MBP-AaS^{PIV} via FPLC using a DEAE anion exchange column in 20 mM Tris-HCl, pH 8 buffer, but in the complementary SDS-PAGE analysis (Figure 30) there were no discernable bands present in the expected

region of the gel associated with the free, active protease (~28 kDa), suggesting autodigestion. This, in combination with the known challenges associated with expressing disulfide-bonded proteins in *E. coli*⁴⁹, would explain much of the difficulty experienced in both the expression and the purification processes. Despite challenges presented in attempts to purify the recombinant enzymes, the results of purification experiments are suggestive of both autocatalysis and autodigestion. Additionally, purification attempts using lower pH buffers indicate the pH-dependence of protease activity. While purified proteases were not obtained for use in kinetic or structural studies, the potential autocatalysis, autodigestion, and pH-dependence of these proteases provide further insight into the potential roles of AaS^{PII} and AaS^{PIV} in *Ae. aegypti* blood meal digestion; because these enzymes are expressed constantly following blood meal acquisition³⁵, we hypothesize that they may be responsible for the initial digestion of blood meal proteins to promote accessibility of internal peptide bonds for further digestion by other midgut serine proteases.

4.4 Future Studies

Due to difficulties in obtaining purified recombinant proteases, which are likely due to autoactivation and autodigestion during the purification preparation process, immediate future studies will focus on growth experiments to express recombinant AaS^{PII} (no leader) and AaS^{PIV} (with leader) with the mutated active site serine to alanine using both pET28a and pMAL-c5E constructs. This should lead to inactive proteases incapable of autoactivating and autodigesting and prevent the associated obstacles observed during the purification of the wildtype AaS^{PII} and AaS^{PIV} fusion

proteins. However, mutating the active site serine residue would lead to inactive proteases, making downstream kinetic assays impossible, but could be useful for obtaining structural data.

Additionally, primers were developed to mutate the lysine residue to alanine at the propeptide region. This region is necessary in order to convert the inactive zymogen to the active mature enzyme, and for both AaS^{PII} and AaS^{PIV}, the propeptide cleavage amino acid is lysine. By mutating this lysine residue, this may prevent autoactivation and allow for later *in vitro* activation of the enzyme for functional studies.

Expression of the recombinant enzymes with the engineered EK-pseudo propeptide was also attempted to prevent autoactivation; however, trypsin-like enzymes prefer to cleave on the C-terminal peptide bonds of lysine or arginine residues, so the EK region may still be a suitable target for AaS^{PII} and AaS^{PIV} autoactivation. However, results of AaS^{PII} with the EK-pseudo propeptide construct in SHuffle T7 cells grown using auto-induction at 30 °C did show successful expression with limited or no autoactivation (Figure 21). However, repeated growth experiments using the same conditions or with lowered temperatures and extended growth times were all unsuccessful (results not shown). Further efforts to repeat the successful soluble expression observed in Figure 21 need to be taken. Issues obtaining soluble expression of the proteases containing the EK-pseudo propeptide could likely be overcome with the addition of an MBP-fusion tag. Each of the pET28a plasmid constructs encoding AaS^{PII} and AaS^{PIV} with the attached EK-pseudo propeptide can be used as templates for cloning into the pMAL-c5E vector as described in the Methods section for the AaS^{PII} (no leader) and AaS^{PIV} (with leader)

constructs. This would allow for expression of MBP-tagged recombinant proteases with both an internal His₆-tag and EK-pseudo propeptide region, which would hopefully result in highly expressed inactive recombinant enzymes that could be easily purified and later activated *in vitro* (this step could be used to both activate the enzyme and also remove the attached MBP and His₆-tag). In order to fully understand the biological role of AaSPII and AaSPIV in the *Ae. aegypti* blood meal digestion process, it will be necessary to circumvent certain difficulties in the expression process in order to obtain kinetic data and provide insight into the enzymes' functions.

Of the discussed experiments, purification via FPLC using a DEAE weak anion exchange column in 20 mM Tris-HCl, pH 8 buffer with the addition of 1 mM PMSF needs to be repeated with both MBP-tagged proteases to confirm consistency and repeatability of the results, especially based on the potential implications of ineffective inhibition of these proteases by PMSF. At this point, purification with PMSF added was only performed once and only with MBP-tagged AaSPIV (with leader). Because PMSF is an irreversible inhibitor of serine proteases, its use in the purification process would prevent later activity studies (as is the case for potential expressed proteases with an active site serine mutated to an alanine). Therefore, these methods could be used to obtain inactive but properly folded recombinant enzymes suitable for structural studies via crystallography. Additionally, purified inactive soluble recombinant proteases could be used in binding studies to determine binding specificity.

If later experiments confirm that PMSF is not an effective inhibitor against active AaSPII or AaSPIV, the addition of other inhibitors can be used to both achieve purified

inactive proteases and determine what type of inhibitors are effective against the proteases, which would also provide insight into the enzyme function. Another method to potentially inactivate the recombinant proteases expressed as MBP-fusion proteins would be to include a reducing agent such as dithiothreitol (DTT) in the purification buffer to counterbalance the effects of MBP as a solubility enhancer; expression with MBP could assure that solubility of the recombinant proteases is achieved while reduction of the disulfide bonds may be enough to minimize activity while maintaining solubility. Purification efforts at lower pH values (Figures 34 and 35) implied the pH-dependence of the proteases, which can be further explored to obtain pH profiles for each protease by determining if autocatalyzation also occurs at higher pH values or if pH-dependence is parallel to that of classic trypsin.

If the enzymes are autocatalyzing and potentially autodigesting, determining potential internal cleavage sites within the active proteases would be extremely informative. Cloning of the AaS^{PII} and AaS^{PIV} genes were both cloned into the pET29b vector, which encodes for a C-terminal His₆-tag. Further growth experiments will be performed to determine the optimal conditions for producing soluble, recombinant enzyme with a C-terminal His₆-tag; if subsequent purification results also suggest autocatalytic behavior, the cleaved protein product will still be purified by exploiting the His₆-tag for IMAC. Cleaved proteins and protein species could then be verified using N-terminal sequencing and confirm specific internal cleavage sites, providing valuable information on potential substrate recognition and specificity.

REFERENCES

1. Centers for Disease Control and Prevention (CDC) Mosquito Life-Cycle.
https://www.cdc.gov/dengue/entomologyecology/m_lifecycle.html (2012).
2. Takken, W.; Koenraadt, C. J. M. *Ecology of parasite-vector interactions*; Ecology and control of vector-borne diseases; Wageningen Academic Publishers: Wageningen, 2013. Vol. 3.
3. *Dengue: Guidelines for treatment, prevention and control*; World Health Organization (WHO) and Special Programme for Research and Training in Tropical Diseases (TDR): Geneva, 2009.
4. McBride, C. S.; Baier, F.; Omondi, A. B.; Spitzer, S. A.; Lutomiah, J.; Sang, R.; Ignell, R.; Vosshall, L. B. Evolution of mosquito preference for humans linked to an odorant receptor. *Nature* **2014**, *515*, 222-227.
5. World Health Organization (WHO) Zika virus.
<http://www.who.int/mediacentre/factsheets/zika/en/> (2016).
6. Moreira, J.; Peixoto, T. M.; Machado de Siqueira, A.; Lamas, C. C. Sexually acquired Zika virus: a systematic review. *Clin. Microbiol. Infect.* **2017**.
7. Krauer, F.; Riesen, M.; Reveiz, L.; Oladapo, O. T.; Martinez-Vega, R.; Porgo, T. V.; Haeffliger, A.; Broutet, N. J.; Low, N.; WHO Zika Causality Working Group Zika Virus Infection as a Cause of Congenital Brain Abnormalities and Guillain-Barre Syndrome: Systematic Review. *PLoS Med.* **2017**, *14*, e1002203.
8. Mlakar, J.; Korva, M.; Tul, N.; Popovic, M.; Poljsak-Prijatelj, M.; Mraz, J.; Kolenc, M.; Resman Rus, K.; Vesnaver Vipotnik, T.; Fabjan Vodusek, V.; Vizjak, A.; Pizem, J.; Petrovec, M.; Avsic Zupanc, T. Zika Virus Associated with Microcephaly. *N. Engl. J. Med.* **2016**, *374*, 951-958.
9. Centers for Disease Control and Prevention (CDC) Facts about Microcephaly.
<https://www.cdc.gov/ncbddd/birthdefects/microcephaly.html> (2016).
10. Willison, H. J.; Jacobs, B. C.; van Doorn, P., A. Guillain-Barré syndrome. *Lancet* **2016**, *388*, 717-727.
11. Centers for Disease Control and Prevention (CDC) Areas with Zika.
<https://www.cdc.gov/zika/geo/index.html> (2017).
12. Normile, D. Tropical medicine. Surprising new dengue virus throws a spanner in disease control efforts. *Science* **2013**, *342*, 415.
13. World Health Organization (WHO) Dengue and severe dengue.
<http://www.who.int/mediacentre/factsheets/fs117/en/> (2016).

14. Brady, O. J.; Gething, P. W.; Bhatt, S.; Messina, J. P.; Brownstein, J. S.; Hoen, A. G.; Moyes, C. L.; Farlow, A. W.; Scott, T. W.; Hay, S. I. Refining the global spatial limits of dengue virus transmission by evidence-based consensus. *PLoS Negl Trop. Dis.* **2012**, *6*, e1760.
15. Bhatt, S.; Gething, P. W.; Brady, O. J.; Messina, J. P.; Farlow, A. W.; Moyes, C. L.; Drake, J. M.; Brownstein, J. S.; Hoen, A. G.; Sankoh, O.; Myers, M. F.; George, D. B.; Jaenisch, T.; Wint, G. R.; Simmons, C. P.; Scott, T. W.; Farrar, J. J.; Hay, S. I. The global distribution and burden of dengue. *Nature* **2013**, *496*, 504-507.
16. Fredericks, A. C.; Fernandez-Sesma, A. The burden of dengue and chikungunya worldwide: implications for the southern United States and California. *Ann. Glob. Health.* **2014**, *80*, 466-475.
17. World Health Organization (WHO) Yellow fever.
<http://www.who.int/mediacentre/factsheets/fs100/en/> (2016).
18. World Health Organization (WHO) Chikungunya.
<http://www.who.int/mediacentre/factsheets/fs327/en/> (2016).
19. Centers for Disease Control and Prevention (CDC) Chikungunya virus in the United States.
<https://www.cdc.gov/chikungunya/geo/united-states.html> (2017).
20. Monaghan, A. J.; Sampson, K. M.; Steinhoff, D. F.; Ernst, K. C.; Ebi, K. L.; Jones, B.; Hayden, M. H. The potential impacts of 21st century climatic and population changes on human exposure to the virus vector mosquito *Aedes aegypti*. *Clim. Change* **2016**, 1-14.
21. Butterworth, M. K.; Morin, C. W.; Comrie, A. C. An Analysis of the Potential Impact of Climate Change on Dengue Transmission in the Southeastern United States. *Environ. Health Perspect.* **2016**.
22. Isoe, J.; Rascon, A. A., Jr; Kunz, S.; Miesfeld, R. L. Molecular genetic analysis of midgut serine proteases in *Aedes aegypti* mosquitoes. *Insect Biochem. Mol. Biol.* **2009**, *39*, 903-912.
23. Rascon, A. A., Jr; Gearin, J.; Isoe, J.; Miesfeld, R. L. In vitro activation and enzyme kinetic analysis of recombinant midgut serine proteases from the Dengue vector mosquito *Aedes aegypti*. *BMC Biochem.* **2011**, *12*, 43-2091-12-43.
24. Seirin Lee, S.; Baker, R. E.; Gaffney, E. A.; White, S. M. Modelling *Aedes aegypti* mosquito control via transgenic and sterile insect techniques: endemics and emerging outbreaks. *J. Theor. Biol.* **2013**, *331*, 78-90.
25. Chakradhar, S. Buzzkill: Regulatory uncertainty plagues rollout of genetically modified mosquitoes. *Nat. Med.* **2015**, *21*, 416-418.
26. Lopes, R. J. Why transgenic insects are still not ready for prime time. *Nature* **2016**,
[doi:10.1038/nature.2016.19804](https://doi.org/10.1038/nature.2016.19804).

27. Ferguson, N. M.; Kien, D. T.; Clapham, H.; Aguas, R.; Trung, V. T.; Chau, T. N.; Popovici, J.; Ryan, P. A.; O'Neill, S. L.; McGraw, E. A.; Long, V. T.; Dui le, T.; Nguyen, H. L.; Chau, N. V.; Wills, B.; Simmons, C. P. Modeling the impact on virus transmission of Wolbachia-mediated blocking of dengue virus infection of *Aedes aegypti*. *Sci. Transl. Med.* **2015**, *7*, 279ra37.
28. Meijer, M.; Dingemans, M. M.; van den Berg, M.; Westerink, R. H. Inhibition of voltage-gated calcium channels as common mode of action for (mixtures of) distinct classes of insecticides. *Toxicol. Sci.* **2014**, *141*, 103-111.
29. Sanchez-Bayo, F.; Goulson, D.; Pennacchio, F.; Nazzi, F.; Goka, K.; Desneux, N. Are bee diseases linked to pesticides? - A brief review. *Environ. Int.* **2016**, *89-90*, 7-11.
30. Ross, S. M.; McManus, I. C.; Harrison, V.; Mason, O. Neurobehavioral problems following low-level exposure to organophosphate pesticides: a systematic and meta-analytic review. *Crit. Rev. Toxicol.* **2013**, *43*, 21-44.
31. Flores, A. E.; Ponce, G.; Silva, B. G.; Gutierrez, S. M.; Bobadilla, C.; Lopez, B.; Mercado, R.; Black, W. C., 4th Wide spread cross resistance to pyrethroids in *Aedes aegypti* (Diptera: Culicidae) from Veracruz state Mexico. *J. Econ. Entomol.* **2013**, *106*, 959-969.
32. Deming, R.; Manrique-Saide, P.; Medina Barreiro, A.; Cardena, E. U.; Che-Mendoza, A.; Jones, B.; Liebman, K.; Vizcaino, L.; Vazquez-Prokopec, G.; Lenhart, A. Spatial variation of insecticide resistance in the dengue vector *Aedes aegypti* presents unique vector control challenges. *Parasit. Vectors* **2016**, *9*, 67-016-1346-3.
33. Flores, A. E.; Gustavo Ponce-Garcia; Beatriz Lopez-Monroy; Olga, K. V.; Iram, P. R.; Juan, I. A.; Pablo Manrique-Saide. Current Status of the Insecticide Resistance in *Aedes aegypti* (Diptera: Culicidae) from Mexico. In *Insecticides Resistance*; Trdan, S., Ed.; InTech: Rijeka, Croatia, 2016; .
34. Maciel-de-Freitas, R.; Avendanho, F. C.; Santos, R.; Sylvestre, G.; Araujo, S. C.; Lima, J. B.; Martins, A. J.; Coelho, G. E.; Valle, D. Undesirable consequences of insecticide resistance following *Aedes aegypti* control activities due to a dengue outbreak. *PLoS One* **2014**, *9*, e92424.
35. Brackney, D. E.; Isoe, J.; W, C. B., 4th; Zamora, J.; Foy, B. D.; Miesfeld, R. L.; Olson, K. E. Expression profiling and comparative analyses of seven midgut serine proteases from the yellow fever mosquito, *Aedes aegypti*. *J. Insect Physiol.* **2010**, *56*, 736-744.
36. Noriega, F. G.; Edgar, K. A.; Bechet, R.; Wells, M. A. Midgut exopeptidase activities in *Aedes aegypti* are induced by blood feeding. *J. Insect Physiol.* **2002**, *48*, 205-212.
37. O'Farrell, T. J.; Guo, R.; Hasegawa, H.; Pourmotabbed, T. Matrix metalloproteinase-1 takes advantage of the induced fit mechanism to cleave the triple-helical type I collagen molecule. *Biochemistry* **2006**, *45*, 15411-15418.

38. Brix, K.; Stöcker, W., Eds.; In *Proteases: Structure and Function*; Springer-Verlag Wien: Heidelberg, 2013.
39. Barrett, A. J.; Rawlings, N. D.; Woessner, J. F., Eds.; In *Handbook of Proteolytic Enzymes, 3rd Edition*; Academic Press: Massachusetts, 2012. Vol. 1.
40. Page, M. J.; Di Cera, E. Serine peptidases: classification, structure and function. *Cell Mol. Life Sci.* **2008**, *65*, 1220-1236.
41. Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry*; W. H. Freeman and Company: New York, 2008.
42. Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry*; W H Freeman: New York, 2002.
43. Noriega, F. G.; Wang, X. Y.; Pennington, J. E.; Barillas-Mury, C. V.; Wells, M. A. Early trypsin, a female-specific midgut protease in *Aedes aegypti*: isolation, aminoterminal sequence determination, and cloning and sequencing of the gene. *Insect Biochem. Mol. Biol.* **1996**, *26*, 119-126.
44. Kalhok., S. E.; Tabak, L. M.; Prosser, D. E.; Brook, W.; Downe, A. E. R.; White, B. N. Isolation, sequencing and characterization of two cDNA clones coding for trypsin-like enzymes from the midgut of *Aedes aegypti*. *Insect Mol. Biol.* **1993**, *2*, 71-79.
45. Venancio, T. M.; Cristofaletti, P. T.; Ferreira, C.; Verjovski-Almeida, S.; Terra, W. R. The *Aedes aegypti* larval transcriptome: a comparative perspective with emphasis on trypsins and the domain structure of peritrophins. *Insect Mol. Biol.* **2009**, *18*, 33-44.
46. Rinderknecht, H. Activation of pancreatic zymogens. Normal activation, premature intrapancreatic activation, protective mechanisms against inappropriate activation. *Dig. Dis. Sci.* **1986**, *31*, 314-321.
47. Sorensen, H. P.; Mortensen, K. K. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb. Cell. Fact.* **2005**, *4*, 1.
48. Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell. 4th edition.*; Garland Science: New York, 2002.
49. Bessette, P. H.; Aslund, F.; Beckwith, J.; Georgiou, G. Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 13703-13708.
50. Berndt, C.; Lillig, C. H.; Holmgren, A. Thioredoxins and glutaredoxins as facilitators of protein folding. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **2008**, *1783*, 641-650.

51. Prinz, W. A.; Aslund, F.; Holmgren, A.; Beckwith, J. The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the Escherichia coli cytoplasm. *J. Biol. Chem.* **1997**, 272, 15661-15667.
52. Singh, A.; Upadhyay, V.; Panda, A. K. Solubilization and refolding of inclusion body proteins. *Methods Mol. Biol.* **2015**, 1258, 283-291.
53. Katoch, R. *Analytical Techniques in Biochemistry and Molecular Biology*; Springer New York: New York, NY, 2011.
54. Selkirk, C. Ion exchange chromatography. *Methods Mol. Biol.* **2004**, 244, 125-131.
55. Sulkowski, E. The saga of IMAC and MIT. *Bioessays* **1989**, 10, 170-175.
56. Arnold, F. H. Metal-affinity separations: a new dimension in protein processing. *Biotechnology (N. Y)* **1991**, 9, 151-156.
57. Bornhorst, B. J.; Falke, J. J. Reprint of: Purification of Proteins Using Polyhistidine Affinity Tags. *Protein Expr. Purif.* **2011**.
58. Sawadogo, M.; Van Dyke, M. W. Indirect use of immobilized metal affinity chromatography for isolation and characterization of protein partners. *Genet. Eng. (N. Y)* **1995**, 17, 53-65.
59. Striegel, A.; Yau, W. W.; Kirkland, J. J. *Modern Size-Exclusion Liquid Chromatography : Practice of Gel Permeation and Gel Filtration Chromatography (2nd Edition)*; Wiley: Hoboken, NJ, USA, 2009.
60. Erlanger, B. F.; Kokowsky, N.; Cohen, W. The preparation and properties of two new chromogenic substrates of trypsin. *Arch. Biochem. Biophys.* **1961**, 95, 271-278.
61. Felix, C. R.; Betschart, B.; Billingsley, P. F.; Freyvogel, T. A. Post-feeding induction of trypsin in the midgut of *Aedes aegypti* L. (Diptera: Culicidae) is separable into two cellular phases. *Insect Biochemistry* **1991**, 21, 197-203.
62. Lai, J.; Niks, D.; Wang, Y.; Domratcheva, T.; Barends, T. R. M.; Schwarz, F.; Olsen, R. A.; Elliott, D. W. M.; Fatmi, Q.; Chang, C. A.; Schlichting, I.; Dunn, M. F.; Mueller, L. J. X-ray and NMR Crystallography in an Enzyme Active Site: The Indoline Quinonoid Intermediate in Tryptophan Synthase. *J. Am. Chem. Soc.* **2011**, 133 (1), 4-7.
63. Lattman, E.; Loll, P. *Protein crystallography: a concise guide*; Johns Hopkins University Press: Baltimore, 2008.
64. Sharif, S.; Fogle, E.; Toney, M. D.; Denisov, G. S.; Shenderovich, I. G.; Buntkowsky, G.; Tolstoy, P. M.; Huot, M. C.; Limbach, H. H. NMR localization of protons in critical enzyme hydrogen bonds. *J. Am. Chem. Soc.* **2007**, 129, 9558-9559.

65. Cavanagh, J. *Protein NMR spectroscopy: principles and practice*; Academic Press: Amsterdam ; Boston, 2007.
66. European Molecular Biology Laboratory Cloning.
http://www.embl.de/pepcore/pepcore_services/cloning/pcr_strategy/primer_design/ (2015).
67. Studier, F. W.; Rosenberg, A. H.; Dunn, J. J.; Dubendorff, J. W. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **1990**, *185*, 60-89.
68. Boos, W.; Shuman, H. Maltose/Maltodextrin System of Escherichia coli: Transport, Metabolism, and Regulation. *Microbiology and Molecular Biology Reviews* **1998**, *62*, 204-229.
69. Levy, R.; Weiss, R.; Chen, G.; Iverson, B. L.; Georgiou, G. Production of correctly folded Fab antibody fragment in the cytoplasm of Escherichia coli trxB gor mutants via the coexpression of molecular chaperones. *Protein Expr. Purif.* **2001**, *23*, 338-347.
70. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **2005**, *41*, 207-234.
71. Grabski, A.; Mehler, M.; Drott, D. The Overnight Express Autoinduction System: High-density cell growth and protein expression while you sleep. *Nature Methods* **2005**, *2*, 233-235.
72. Stepanov, V. G.; Tirumalai, M. R.; Montazari, S.; Checinska, A.; Venkateswaran, K.; Fox, G. E. Bacillus pumilus SAFR-032 Genome Revisited: Sequence Update and Re-Annotation. *PLoS One* **2016**, *11*, e0157331.
73. Toker, L.; Feng, M.; Pavlidis, P. Whose sample is it anyway? Widespread misannotation of samples in transcriptomics studies. *F1000Res* **2016**, *5*, 2103.
74. McEntyre, J.; Ostell, J., Eds.; In *The NCBI Handbook [Internet]*; National Center for Biotechnology Information (US): Bethesda (MD), 2013-.
75. Klimke, W.; O'Donovan, C.; White, O.; Brister, J. R.; Clark, K.; Fedorov, B.; Mizrahi, I.; Pruitt, K. D.; Tatusova, T. Solving the Problem: Genome Annotation Standards before the Data Deluge. *Stand. Genomic Sci.* **2011**, *5*, 168-193.
76. Schnoes, A. M.; Brown, S. D.; Dodevski, I.; Babbitt, P. C. Annotation Error in Public Databases: Misannotation of Molecular Function in Enzyme Superfamilies. *PLOS Computational Biology* **2009**, *5*, e1000605.
77. Stein, L. Genome annotation: from sequence to biology. *Nat. Rev. Genet.* **2001**, *2*, 493-503.

78. Steffler, L. M.; Dolabella, S. S.; Ribolla, P. E.; Dreyer, C. S.; Araujo, E. D.; Oliveira, R. G.; Martins, W. F.; La Corte, R. Genetic variability and spatial distribution in small geographic scale of *Aedes aegypti* (Diptera: Culicidae) under different climatic conditions in Northeastern Brazil. *Parasit. Vectors* **2016**, *9*, 530.
79. Failloux, A. B.; Vazeille, M.; Rodhain, F. Geographic genetic variation in populations of the dengue virus vector *Aedes aegypti*. *J. Mol. Evol.* **2002**, *55*, 653-663.
80. Craig, G. B., Jr.; Vandehey, R. C.; Hickey, W. A. Genetic variability in populations of *Aedes aegypti*. *Bull. World Health Organ.* **1961**, *24*, 527-539.
81. Gitlin, I.; Carbeck, J. D.; Whitesides, G. M. Why are proteins charged? Networks of charge-charge interactions in proteins measured by charge ladders and capillary electrophoresis. *Angew. Chem. Int. Ed Engl.* **2006**, *45*, 3022-3060.
82. Russell, R. B.; Barton, G. J. Structural features can be unconserved in proteins with similar folds. An analysis of side-chain to side-chain contacts secondary structure and accessibility. *J. Mol. Biol.* **1994**, *244*, 332-350.
83. Wang, X.; Fu, M.; Ren, J.; Qu, X. Evaluation of different culture conditions for high-level soluble expression of human cyclin A2 with pET vector in BL21 (DE3) and spectroscopic characterization of its inclusion body structure. *Protein Expr. Purif.* **2007**, *56*, 27-34.
84. Qin, M.; Wang, W.; Thirumalai, D. Protein folding guides disulfide bond formation. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 11241-11246.
85. Rosano, G. L.; Ceccarelli, E. A. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* **2014**, *5*, 172.
86. Sorensen, H. P.; Mortensen, K. K. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J. Biotechnol.* **2005**, *115*, 113-128.
87. Bonomo, J.; Gill, R. T. Amino acid content of recombinant proteins influences the metabolic burden response. *Biotechnol. Bioeng.* **2005**, *90*, 116-126.
88. Bentley, W. E.; Mirjalili, N.; Andersen, D. C.; Davis, R. H.; Kompala, D. S. Plasmid-encoded protein: the principal factor in the "metabolic burden" associated with recombinant bacteria. *Biotechnol. Bioeng.* **1990**, *35*, 668-681.
89. Sarduy, E. S.; Munoz, A. C.; Trejo, S. A.; de los A Chavez Planes, M. High-level expression of Falcipain-2 in *Escherichia coli* by codon optimization and auto-induction. *Protein Expr. Purif.* **2012**, *83*, 59-69.
90. Kane, J. F. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* **1995**, *6*.

91. Kim, S.; Lee, S. B. Rare codon clusters at 5'-end influence heterologous expression of archaeal gene in *Escherichia coli*. *Protein Expr. Purif.* **2006**, *50*.
92. Graumann, P. L.; Marahiel, M. A. A superfamily of proteins that contain the cold-shock domain. *Trends Biochem. Sci.* **1998**, *23*, 286-290.
93. Vasina, J. A.; Baneyx, F. Expression of aggregation-prone recombinant proteins at low temperatures: a comparative study of the *Escherichia coli* cspA and tac promoter systems. *Protein Expr. Purif.* **1997**, *9*, 211-218.
94. Olsen, J. V.; Ong, S. E.; Mann, M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics* **2004**, *3*, 608-614.
95. Buettner, K.; Kreisig, T.; Strater, N.; Zuchner, T. Protein surface charge of trypsinogen changes its activation pattern. *BMC Biotechnol.* **2014**, *14*, 109-014-0109-5.
96. Pryor, K. D.; Leiting, B. High-level expression of soluble protein in *Escherichia coli* using a His6-tag and maltose-binding-protein double-affinity fusion system. *Protein Expr. Purif.* **1997**, *10*, 309-319.
97. Aboulwafa, M.; Saier, M. H., Jr Characterization of the *E. coli* glucose permease fused to the maltose-binding protein. *J. Basic Microbiol.* **2008**, *48*, 3-9.
98. Weller, U.; Muller, L.; Messner, M.; Palmer, M.; Valeva, A.; Tranum-Jensen, J.; Agrawal, P.; Biermann, C.; Dobereiner, A.; Kehoe, M. A.; Bhakdi, S. Expression of active streptolysin O in *Escherichia coli* as a maltose-binding-protein--streptolysin-O fusion protein. The N-terminal 70 amino acids are not required for hemolytic activity. *Eur. J. Biochem.* **1996**, *236*, 34-39.
99. Huijbers, M. M.; van Berkel, W. J. High yields of active *Thermus thermophilus* proline dehydrogenase are obtained using maltose-binding protein as a solubility tag. *Biotechnol. J.* **2015**, *10*, 395-403.
100. Sekar, V.; Hageman, J. H. Specificity of the serine protease inhibitor, phenylmethylsulfonyl fluoride. *Biochemical and Biophysical Research Communications* **1979**, *89*, 474-478.
101. Shah, I.; Hunter, L. Predicting Enzyme Function from Sequence: A Systematic Appraisal. *Proceedings / ...International Conference on Intelligent Systems for Molecular Biology; ISMB. International Conference on Intelligent Systems for Molecular Biology* **1997**, *5*, 276-283.