A STUDY OF BACTERIAL PRODUCED ENDOLYSINS ON THE PEPTIDOGLYCAN LAYER OF CLOSTRIDIUM DIFFICILE

An Undergraduate Research Scholars Thesis

by

WILLIAM MEYERS

Submitted to the LAUNCH: Undergraduate Research office at
Texas A&M University
in partial fulfillment of requirements for the designation as an

UNDERGRADUATE RESEARCH SCHOLAR

Approved by
Faculty Research Advisor: Dr. Joseph Sorg

May 2021

Major: Microbiology

RESEARCH COMPLIANCE CERTIFICATION

Research activities involving the use of human subjects, vertebrate animals, and/or biohazards must be reviewed and approved by the appropriate Texas A&M University regulatory research committee (i.e., IRB, IACUC, IBC) before the activity can commence. This requirement applies to activities conducted at Texas A&M and to activities conducted at non-Texas A&M facilities or institutions. In both cases, students are responsible for working with the relevant Texas A&M research compliance program to ensure and document that all Texas A&M compliance obligations are met before the study begins.

I, William Meyers, certify that all research compliance requirements related to this Undergraduate Research Scholars thesis have been addressed with my Research Faculty Advisor prior to the collection of any data used in this final thesis submission.

This project did not require approval from the Texas A&M University Research Compliance & Biosafety office.

TABLE OF CONTENTS

ABSTRACT

A Study of Bacterial Produced Enzymes on the Peptidoglycan Layer of Clostridium difficile

William Meyers
Department of Biology
Texas A&M University

Research Faculty Advisor: Dr. Joseph Sorg
Department of Biology
Texas A&M University

Endospore formation by *Clostridium difficile* is a vital part of the life cycle for this Gram-positive bacterium. During this process, the mothercell packages proteins into the developing endospore. However, the mechanisms by which many of these proteins are secreted into the spore cortex layer is unknown. Using an endolysin as a reporter protein to translationally fuse to potential sporulation secretion substrates would help determine if the secreted proteins is destined for the cortex layer or if it remains in the outer coat. This thesis uses a bioinformatics approach to identify potential endolysins that could be used as a reporter protein in the future. By searching primary literature and online databases, a variety of naturally-produced endolysins derived from several C. *difficile* strains, similar bacteria species and phages were identified. From this list, these proteins were fed into Basic Local Alignment Tool (BLAST) to determine which of these were present in multiple organisms. Next, the potential cellular location of these candidate proteins was determined using the PSORT-algorithm. While PSORT was unable to determine where many of the putative reporter proteins were located, it did give results for some that could be of use. These new, potential reporter proteins would give new insight on the

process of sporulation, allowing for the study of the secretion process that occur during *C. difficile* spore formation.

DEDICATION

To my family, friends and mentors who have guided me and supported my endeavors in science.

ACKNOWLEDGEMENTS

Contributors

I would like to thank my faculty advisor, Dr. Joseph Sorg for his guidance and support throughout the course of this research.

Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

Finally, thanks to my parents for their encouragement and to my girlfriend for her patience and love.

Funding Sources

This undergraduate research did not require any funding.

NOMENCLATURE

B. subtilis	*Bacillus subtilis*
C. difficile	*Clostridium difficile*
CDI	*Clostridium difficile* infection
PSORT	Protein sorting algorithm
BLAST	Basic Local Alignment Tool

1. INTRODUCTION

Clostridium difficile is a Gram-positive bacterium and the primary causative agent of antibiotic-associated diarrhea in hospitals. In response to unknown signals, the strictly anaerobic C. *difficile* vegetative form produces dormant endospores that permit survival of the organism outside of a host. *C. difficile* spores are transmitted to susceptible hosts and, upon germination begins to produce endotoxins which cause the characteristic symptoms of *C. difficile* infections (CDI) (*i.e.*, diarrhea or pseudomembranous colitis) (1). Due to the ability of *C. difficile* to sporulate and become insensitive to antibiotic treatment, there is a high recurrence rate for CDI.

Due to the metabolic dormancy characteristic of bacterial endospores, it is the *C. difficile's* vegetative form that produce the toxins that result in the symptoms of CDI (2). During the sporulation process, the mothercell divides in an asymmetric septation event. This results in the generation of a larger mothercell compartment and a smaller forespore compartment. These two compartments communicate using a cascade of sigma factor activation leading to the engulfment of the forespore compartment by the mothercell and maturation of the forespore into a dormant endospore. Once the forespore matures it is released. In a host, the spore's germination is in response to host-derived bile acids.

The spore has multiple defined layers, from the protein coat, the outer spore membrane, the cortex layer, the inner spore membrane, and the genome-containing core. During the sporulation process, proteins are hypothesized to be secreted from the mothercell, across the outerspore membrane, to the peptidoglycan-containing cortex layer via unknown secretion methods. Due to the difficulty in separating the coat layers from the cortex layer, it is difficult to separate the proteins destined for the cortex layer versus the proteins destined for the coat layer

(which do not cross a membrane). The ability to determine whether the protein stops in the coat or the cortex would be the first step to begin understanding this secretion pathway and give valuable information that could lead to isolation of secreted proteins. Though it is possible to tag potentially secreted proteins using histidine tags, or other small tags, a hybrid protein that has specific activity against the peptidoglycan-containing cortex layer would permit easy screening of transported substrates. I propose using an endolysin as this reporter protein.

Endolysins are enzymes that degrade peptidoglycan. Commonly, endolysins have two domains: an N-terminal domain that provides the catalytic activity and hydrolyzes the murein cell wall and a C-terminal domain that binds to the cell wall substrates (3). The use of an endolysin as a reporter protein attached to a protein secreted from the mothercell should result in the degradation of the cortex layer and prevent spore formation. This would then permit the study of protein transport from the mothercell and the identification of secretion signals for the transported substances – a non-transported hybrid protein would not result in loss of sporulation. To identify such proteins, I used a bioinformatics approach to look through the genomes of *C. difficile*, *Bacillus subtilis*, *Clostridium perfringens*, and several bacteriophages. Due to the requirement that the endolysin transport must be dependent on the spore protein to which it is fused, the endolysin itself should not contain a secretion signal. Thus, to determine the location of the identified endolysins in their host organisms they were analyzed through a protein sorting algorithm. The endolysins collected were analyzed by the PSORT algorithm that provides an analysis of protein localization. Using these locations, it is possible to determine the best endolysin by looking at their predicted location. The generated list serves as a resource for future work on protein secretion during sporulation.

2. METHODS

2.1 Identification of endolysins

The first step to determine a compatible endolysin was to define the search parameter in which to look for potential endolysins. This was done by creating a list (Appendix 1) of organism and strains that are varied yet close enough to each other to have potential crossover proteins. This list consisted of several *C. difficile's* ribotypes: ribotypes 078, 027, 017. The ribotype 078 is associated in the Netherlands and parts of Europe, the strain examined from this ribotype was M120. Ribotype 027 is an epidemic ribotype that has emerged in the past 10 years and accounts for a large number of cases in the United States. Because this ribotype is so prevalent, it accounted for the majority of strains researched: R20291, 196, Bi1, 885. Ribotype 017 is another strain which is responsible for infections across the globe but heavily concentrated in Asia. The final was ribotype106 which accounts for the majority of infections in North America as it produces more spores than even 027 (4).

2.2 Acquiring sequences

Each genome was uploaded into the free DNA analysis program, benchling.com, and was used to search the genome for genes annotated as endolysins. This strategy yielded a few results as few endolysins were already annotated in the genome. To expand the list, the protein sequences found in the benchling analysis were fed into the BLAST algorithm, maintained by the National Center for Biotechnology Information, to reveal potential homologues in other organisms. This gave rise to more homologous protein that were conserved across the strains. These proteins were also cross searched into each strain and the same was done for all of the sequences found in the literature. Each strain was cross referenced into one another resulting in

endolysins being found across the strains. The search continued into *C. perfringens*, *B. subtilis*, and *C. sordellii.* The same process was done for each of the sister organisms, starting with uploading to benchling, to searching literature for endolysins and eventually cross referencing for more results. Lastly this search process was done for a few phage groups as well, this resulted in more results as well that had cross over. Each of the putative endolysins were compiled into a list where their origins were recorded as well as the classification of the endolysin. Once satisfied the list was completed, each endolysin was put into a program that sorted them and gave an analysis report. Psort is able to give the predicted location of proteins in bacteria. The possible locations for a Gram-positive cell listed by PSORT were: cell wall, extracellular, cytoplasmic membrane and cytoplasmic. The algorithm determined a localization score where a greater value represented a greater confidence in the prediction. If unable to determine the location, the program would give an unknown final prediction.

2.3 Putting it all together

Once the locations were determined, the size and classifications of these endolysins were determined to understand their role in the cell. Each protein has a unique identifier known as an accession number to help identify it. The accession number for each protein was logged in the table and searched back into BLAST to give the classification of the protein. The classification of the endolysin helps determine its activity. The size was determined from the BLAST analysis as well by looking and the amino acid count. Table 1 logged all of these stats including: potential location, size, accession number and classification. The table could then be utilized to identify the best reporter protein by looking at the predicted location and how many organisms it appeared in.

3. RESULTS

Analyzing these proteins gave three main data points: the predicted location, the size, and the classification. Each of these data points are logged in Appendix 1. The data gathered from the PSORT algorithm gave the predicted locations of the endolysins. The analysis gave mixed results with most endolysins having unknowns in the predicted location and only three of the nineteen proteins having a known value. Additionally, the only endolysins with predicted locations were found in the cell wall. The amino acid length of the proteins had a large variance ranging from 43 to 283 amino acids. Many different classifications of endolysins emerged in this study with the predominant being N-acetylmuramoyl-l-alanine amidase. The second largest representation of endolysins is glucosaminidase domain-containing protein, and the third largest class was 2 chain A endolysins and scattered hypothetical proteins.

4. CONCLUSION

4.1 Importance of endolysins as reporter proteins

C. difficile is an ongoing issue in many developed countries and most frequent in hospitals where heavy antibiotics are used. Their ability to form spores keeps *C. difficile* resistant to antibiotics and leads to an increase in recurring infections (1). Determining the method and destination of protein secretion during sporulation is vital to understanding the biology of the sporulation process. Using endolysins to act as a reporter protein could give insight on both unknowns. These endolysins are found naturally occurring from bacteria and trace their roots back to phages. Identifying the proper endolysin that will lyse a variety of *C. difficile* strains will provide a major tool to determine the secretion method and the destination of proteins secreted from the mothercell. As stated previously, endolysin degrade the peptidoglycan layer of the cell wall and are used to lyse the cell. Using endolysins as a reporter protein by tagging secreted proteins would cause the cell to lyse when the secreted protein passes into the cortex layer of the spore. The destination of the secreted protein also highlights difference in secretion systems. As the secretion system of many proteins are unknown this would give insight on narrowing down which secretion system that these proteins could be using. Lastly determining the location in the spore is the first step of being able to isolate a protein. An effective reporter protein that can work with different strains would save time and effort in procuring information about the sporulation method. Taking a bioinformatics approach this study was aimed to determine this effective reporter protein by cataloging and cross referencing multiple endolysins across various strains and analyzing where they are predicted to be found in the cell. The predicted location is the most important aspect looked at during this experiment due to this specificity of these

endolysins. The goal was to find a cytoplasmic endolysin with homology with multiple *C. difficile* strains. This is due to cytoplasmic endolysins being a part of a secretion system involved in cell wall synthesis and repair. As this study focuses on the identification of the secretion system during sporulation, a conflict in secretin signals would cause an issue in being able to use these as reporter proteins. Cytoplasmic endolysins do not have this issue due to them not being involved in cell wall repair and synthesis. It is this lack of secretion system that makes them an ideal target for use of a reporter protein in study of a secretion system.

4.2 Identifying the reporter protein

N-acetylmuramoyl-l-alanine amidase were the predominant endolysins discovered from this study, found in nearly all strains. This amidase is an endolysin whose role is in recycling peptidoglycan and cell wall synthesis (5). Another frequent class of endolysins found was the glucosaminidase domain-containing protein (6). These have some homology with NAG proteins in the murein cell wall leading to a possible role in cell wall synthesis. Each of these endolysins have domains on them that have potential to act as a reporter protein. As no cytoplasmic endolysins were found in this project this leads to future directions this project could lead. Expanding the search through other strains and increasing the amount of endolysins observed would result in a higher yield of results. The other issue of having many results from the Psort come back as unknown could potentially be resolved by further investigation into the primary literature on some of these endolysin categories.

In a non-virtual project, these different endolysins could be tested to determine if they are effective at degrading the spore cortex layer. This would show and determine if the endolysin is an effective reporter protein. With the amount of unknown predicted locations this will be the next step needed to continue this research project.

REFERENCES

1. Paredes-Sabja D, Shen A, Sorg JA. 2014. Clostridium difficile spore biology: sporulation, germination, and spore structural proteins. Trends Microbiol 22:406-16.

2. Zhu D, Sorg JA, Sun X. 2018. Clostridioides difficile Biology: Sporulation, Germination, and Corresponding Therapies for C. difficile Infection. Front Cell Infect Microbiol 8:29.

3. Gervasi T, Horn N, Wegmann U, Dugo G, Narbad A, Mayer MJ. 2014. Expression and delivery of an endolysin to combat Clostridium perfringens. Appl Microbiol Biotechnol 98:2495-505.

4. Knight DR, Elliott B, Chang BJ, Perkins TT, Riley TV. 2015. Diversity and Evolution in the Genome of Clostridium difficile. Clin Microbiol Rev 28:721-41.

5. Prigozhin DM, Mavrici D, Huizar JP, Vansell HJ, Alber T. 2013. Structural and biochemical analyses of Mycobacterium tuberculosis N-acetylmuramyl-L-alanine amidase Rv3717 point to a role in peptidoglycan fragment recycling. J Biol Chem 288:31549-55.

6. Dhalluin A, Bourgeois I, Pestel-Caron M, Camiade E, Raux G, Courtin P, Chapot-Chartier MP, Pons JL. 2005. Acd, a peptidoglycan hydrolase of Clostridium difficile with N-acetylglucosaminidase activity. Microbiology (Reading) 151:2343-2351.

7. Ha E, Son B, Ryu S. 2018. Clostridium perfringens Virulent Bacteriophage CPS2 and Its Thermostable Endolysin LysCPS2. Viruses 10.

8. Perez-Dorado I, Campillo NE, Monterroso B, Hesek D, Lee M, Paez JA, Garcia P, Martinez-Ripoll M, Garcia JL, Mobashery S, Menendez M, Hermoso JA. 2007. Elucidation of the molecular recognition of bacterial cell wall by modular pneumococcal phage endolysin CPL-1. J Biol Chem 282:24990-9.

9. Sundram F, Guyot A, Carboo I, Green S, Lilaonitkul M, Scourfield A. 2009. Clostridium difficile ribotypes 027 and 106: clinical outcomes and risk factors. J Hosp Infect 72:111-8.

10. Tenover FC, Akerlund T, Gerding DN, Goering RV, Bostrom T, Jonsson AM, Wong E,

Wortman AT, Persing DH. 2011. Comparison of strain typing results for Clostridium difficile isolates from North America. J Clin Microbiol 49:1831-7.

11. Yoong P, Schuch R, Nelson D, Fischetti VA. 2004. Identification of a broadly active phage lytic enzyme with lethal activity against antibiotic-resistant Enterococcus faecalis and Enterococcus faecium. J Bacteriol 186:4808-12.

12. Yuan Y, Peng Q, Gao M. 2012. Characteristics of a broad lytic spectrum endolysin from phage BtCS33 of Bacillus thuringiensis. BMC Microbiol 12:297.

APPENDIX 1

Table A1: List of putative endolysins with accession number, origins, classification, and size.

Endolysins	Organism	Ribotype	Strain	Predicted location	Classification	Amino acid length
VIF68045.1	*C. difficile*	78	M120	Unknown	N-acetylmuramoyl-l-alanine amidase	85aa
WP_003415961.1	*C. difficile*	78	M120	Unknown	N-acetylmuramoyl-l-alanine amidase	276aa
VIG55126.1	*C. difficile*	78	M120	Unknown	N-acetylmuramoyl-l-alanine amidase	276aa
WP_003435466.1	*C. difficile*	27	196	Cell wall	N-acetylmuramoyl-l-alanine amidase	262aa
WP_021365488.1	*C. difficile*	27	R20291	Cell wall	N-acetylmuramoyl-l-alanine amidase	262aa
WP_015984863.1	phiCD38-2	Phage	Phage	Unknown	N-acetylmuramoyl-l-alanine amidase	270aa
WP_021359648.1	phiCD38-2	Phage	Phage	Unknown	N-acetylmuramoyl-l-alanine amidase	270aa
WP_131024832.1	phiCD38-2	Phage	Phage	Unknown	N-acetylmuramoyl-l-alanine amidase	270aa
OFU07532.1	*C. difficile*	27	R20291	Cell wall	N-acetylmuramoyl-l-alanine amidase	262a
EFH15356.1	*C. difficile*	78	M120	Unknown	Hypothetical protein HMPREF0219_2013	93aa
WP_018112549.1	phiCDIF1296T	Phage	Phage	Unknown	Glucosaminidase domain-containing protein	205aa
WP_009895117.1	phiCDIF1296T	Phage	Phage	Unknown	Glucosaminidase domain-containing protein	205aa
WP_131042004.1	phiCDIF1296T	Phage	Phage	Unknown	Glucosaminidase domain-containing protein	205aa
YP_009838148	*C. perfringens*	Phage	Phage	Unknown	hypothetical protein CPS2_25	43aa
3QAY_A	*C. Sordellii*	NA	NA	Unknown	Chain A endolysin	180aa
6AKV_A	*B. subtilis*	NA	NA	Unknown	Chain A, LysB4	282aa
WP_003435466.1	*C. diffivile*	NA	NA	Unknown	N-acetylmuramoyl-l-alanine amidase	270aa

WP_009889444.1	*C. difficile*	NA	NA	Unknown	N-acetylmuramoyl-l-alanine amidase	270aa
WP_003433910.1	*C. difficile*	NA	NA	Unknown	N-acetylmuramoyl-l-alanine amidase	268aa