

2011

Expanding the Horizons of Enzybiotic Identification

Jonathan Edward Schmitz

Follow this and additional works at: http://digitalcommons.rockefeller.edu/student_theses_and_dissertations

 Part of the [Life Sciences Commons](#)

Recommended Citation

Schmitz, Jonathan Edward, "Expanding the Horizons of Enzybiotic Identification" (2011). *Student Theses and Dissertations*. Paper 138.

This Thesis is brought to you for free and open access by Digital Commons @ RU. It has been accepted for inclusion in Student Theses and Dissertations by an authorized administrator of Digital Commons @ RU. For more information, please contact mcsweej@mail.rockefeller.edu.



EXPANDING THE HORIZONS OF ENZYBIOTIC IDENTIFICATION

*A Thesis Presented to the Faculty of
The Rockefeller University
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy*

by
Jonathan Edward Schmitz
June 2011

Expanding the Horizons of Enzybiotic Identification

Jonathan Edward Schmitz, Ph.D.

The Rockefeller University 2011

Recently, phage lytic enzymes (also known as endolysins or, simply, lysins) have received considerable attention as potential antibacterial agents. During the infective cycle of double-stranded DNA phage, these peptidoglycan hydrolases are responsible for digesting the cell wall of the host bacterium and freeing newly-assembled viral particles. At the same time, an increasing body of evidence has demonstrated that recombinantly-purified phage lysins – when added exogenously – can potently kill Gram-positive bacteria, whose peptidoglycan is accessible from the extracellular space. Consequently, lysins have been proposed as novel *enzybiotic* (i.e. enzyme-antibiotic) molecules that could serve as novel weapons in the fight against drug-resistant bacteria. Most lysins characterized to date were initially identified through either recombinant screening or DNA-sequencing of phage genomes. Recent technological and methodological advances, however, have drastically increased the potential avenues for lysin identification. The goal of the work presented here to exploit and expand upon these advances so that the identification of new lysins is increasingly rapid and straightforward.

This thesis is subdivided into four interrelated sections, each of which represents a distinct study into a novel approach/method for cloning phage

lysins. The first study (Chapter 2) addresses the issue of bacterial genomic sequencing and how the rapidly expanding database of bacterial genomes represents a vast source of proviral lysins. Focusing on the anaerobic pathogen *Clostridium perfringens*, the genomes of 9 recently-sequenced strains were computationally mined for prophage lysins and lysin-like ORFs (open reading frames), revealing several dozen proteins of various enzymatic classes. Of these lysins, a muramidase (termed PlyCM) from strain ATCC 13124 was chosen for recombinant analysis based on its dissimilarity to the only other previously-characterized *C. perfringens* lysin. Following expression and purification, various biochemical properties of PlyCM were determined *in vitro*, including pH/salt-dependence and temperature stability. The enzyme exhibited activity at low $\mu\text{g/ml}$ concentrations, and it was active against 23/24 strains of *C. perfringens* tested.

Chapters 3 and 4 focus on the emerging field of *viral metagenomics*, a term which refers to the bulk extraction and analysis of DNA from environmental phage without prior laboratory culture of any particular virus. Phage metagenomes have been shown to be incredibly complex and diverse, and the goal of these chapters was to tap into this diversity through functional metagenomic screens for lytic enzymes. Chapter 3 first addresses a preliminary methodological issue, namely the fact that uncultured phage samples generally do not provide sufficient quantities of DNA for ready screening. A novel E-

LASL protocol (for expressed linker amplified shotgun library) was developed that combines linker amplification of enzyme-digested DNA with subsequent topoisomerase cloning into linearized expression plasmids. As proof-of-principle, genomic and metagenomic E-LASLs were constructed and screened for antibacterial and hemolytic activity in an *Escherichia coli* host. Six *Bacillus anthracis* phage lysins were cloned in the process, along with a virulence factor of the aerolysin gene family.

Chapter 4 proceeds to address an additional methodological issue surrounding metagenomic lysin identification: the question of how to identify lysin-encoding clones in a functional screen when the targeted bacteria are not pre-defined. A novel two-step screening technique was devised for this purpose. It involves a primary screen in which transformed *E. coli* clones were identified that demonstrated colony lysis following exposure to nebulized inducing agent. This effect, which can be due to the expression of membrane-permeabilizing phage holins, was discerned by the development a hemolytic-effect in surrounding blood agar. The selected clones were then overlaid with autoclaved Gram-negative bacteria (specifically *Pseudomonas aeruginosa*) to assay directly for recombinant expression of lytic enzymes, which are often encoded proximally to holins in phage genomes. This method was combined with the aforementioned E-LASL technique and applied to a viral metagenomic library constructed from mixed animal feces. Twenty-six lytic enzymes were cloned in this screen,

including both Gram-positive-like and Gram-negative-like enzymes, as well as several atypical lysins whose predicted structures are less common among known phage.

Finally, Chapter 5 takes the above techniques and reapplies them outside the context of metagenomics, returning to individual genomes as sources of lytic enzymes. Specifically, 2 lysins were cloned from prophage of *Streptococcus suis*, an important veterinary and emerging zoonotic pathogen. One of these *S. suis* enzymes (PlySs1) was identified by applying the two-step screen to the genome of an unsequenced clinical strain. The other (PlySs2) was identified in a manner similar to the clostridial lysin PlyCM, by analyzing the published genomes of various sequenced strains. Finally, PlySs1 was subject to chromatographic purification and *in vitro* analysis against numerous suis and non-suis strains of streptococci. Currently, both PlySs1 and PlySs2 are involved in a collaborator's ongoing *in vivo* trial employing experimentally-infected pigs.

Alla mia bella Bamboulaine...
La ragazza più speciale del mondo.

ACKNOWLEDGEMENTS

I would like to begin by acknowledging the mentorship of my advisor, Prof. Vincent Fischetti. Of course, I would like to thank him for all of his advice over the past four years. Even more so, I am grateful to Prof. Fischetti for fostering a research environment that was *always* open to new ideas, directions, and collaborations. He provided a wonderful model of how a biomedical investigator should strive for a balance between basic academic questions and applied science that translates those questions into real-world tools. And he demonstrated how all of this science can (and should) take place in a laboratory that is creative, collegial, and downright fun! The Laboratory of Bacterial Pathogenesis at The Rockefeller University is truly a unique place, and I am fortunate to have received my graduate training within it.

I am also incredibly grateful for the opportunity to have spent several months during my PhD as a guest researcher abroad at the University of Parma in Italy. I would like to thank the members of that laboratory, especially its directors, Prof. Maria Cristina Ossiprandi and Prof. Ezio Bottarelli, for welcoming me so kindly and for helping develop such a fruitful collaboration (as well as a wonderful formative experience for me personally).

I would like to thank the members of my faculty advisory committee who helped steer my thesis project from beginning to end: Profs. Sean Brady and Charles Rice of The Rockefeller University and Prof. Sabine Ehrt of The Weill-Cornell Graduate School of Biomedical Sciences. I am likewise grateful to Dr.

Jonathan Dworkin of Columbia University for serving as the external examiner at my dissertation defense. Also instrumental in this entire process have been the members of the Tri-Institutional MD-PhD Program—in particular its director, Prof. Olaf Anderson, and the program administrator, Ms. Ruthie Gotian—as well as The Rockefeller University Graduate School Office. Although not directly tied to my thesis project, I must also give profound thanks to Prof. Steven Ackerman of the University of Illinois-Chicago, who first helped spark my interest in research and who has always served as an invaluable mentor and role model.

Within the Fischetti laboratory, there are numerous individuals to whom I am indebted for their advice, their time, and their friendship. In particular, the guidance of Prof. Raymond Schuch was central in getting my project off the ground and troubleshooting hurdles as they arose. Although a long list of names hardly does justice to their individual contributions, the following people all assisted in no small way during my time at Rockefeller: Prof. John Zabriski, Prof. Daniel Nelson, Prof. Patricia Ryan, Prof. Robert Boylan, Dr. Anu Daniel, Dr. Gregory Resch, Dr. Mina Pastagia, Dr. Sung Lee, Dr. Michael Deshotel, Dr. Chad Euler, Dr. Assaf Raz, Dr. Monica Fazzini, Dr. Manu Kapoor, Ms. Barbara Juncosa, Mr. Dennis Spencer, Mr. Daniel Gilmer, Mr. Curtis O’Neal, Ms. Sherry Kahn, Mr. Rashid Rumah, Mr. Peter Chahales, Mr. Adam Pelzak, Mr. Chris Fields, Mr. Tommie Hata, Ms. Clara Eastbay, Ms. Mary Windels, Ms. Shi-wei Zhu, Mr. Jessie Afriye, and Ms. Jennifer Groves. There are other individuals, both inside and outside Rockefeller, who assisted with a particular aspect of my doctoral

research. As it is difficult to mention them here in a way that meaningfully describes their help, each is acknowledged in the particular chapter of this thesis related to his contributions.

I would like to acknowledge several organizations for their generous financial assistance which made possible both my graduate research at Rockefeller and my MD-PhD studies overall. I must thank the MSTP program of the National Institutes of Health for their general support, as well as the PhRMA Foundation and its administrator, Ms. Eileen McCarron, for their 2007 Paul Calabresi Fellowship. I am also indebted to the Rockefeller University Center for Clinical and Translation Science for providing a 2008-2009 pilot grant that funded my benchtop efforts.

Finally, I would like to thank my family members who have given so much love and encouragement throughout this long process. I am eternally grateful to my brilliant, beautiful, and loving wife Francesca for being so open to this amazing and unforeseen adventure of ours which is only just beginning. At the start of graduate school, I could never have guessed the direction that my life would take due to a chance encounter with an Italian girl studying abroad in New York. Now, in retrospect, I could not imagine my life being any different or any better. My parents have likewise been endless sources of support and I could not be where I am now without their years of love and understanding – it is amazing to think that almost thirty years ago to this day, I was thumping around *in utero* at a lecture hall at Notre Dame while my dad defended his PhD

thesis! I am equally grateful to my incredible sister Jessica, without whom I would never have come to New York and without whom the last six years would not have been so special. There are so many other people whom I would thank if space here permitted, but it suffices to say that I have always been abundantly blessed with loved ones and friends.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iv
TABLE OF CONTENTS	viii
LIST OF FIGURES	xi
LIST OF TABLES	xiii

CHAPTER 1 – *Phage, Phage Lysins, and their Molecular Identification*

•Bacteriophage and their Biology	1
Bacteriophage Taxonomy	2
Phage Life-Style	7
Phage Ecology	12
•Phage in Biotechnology and Biomedicine	14
Phage and Molecular Biology	14
Phage Typing	16
Phage Therapy	17
•The Lytic Enzymes of Bacteriophages	24
The Concept of Lysin Therapy	25
Lysin Enzymology	26
Lysin Architecture	34
Lysin-Associated Proteins	41
Genomic Arrangement of Phage Lysins	45
Non-Caudoviral Lytic Proteins	47
Phage Lysins into the Spotlight	49
•Phage-Lysin Identification throughout the Years	52
Historical Perspectives on Lysin Identification	53
Into the Age of Molecular Cloning	56
Phage Lysins and Functional Screening	60
Recent Additions to Lysin Screening	67

Lytic Enzymes and Whole-Genome Sequencing	70
Sequence-Based versus Functional: Pros and Cons	74
•Objectives and Overview of the Thesis Project	77
The Present Body of Research	77
Dissertation Format	81
•Endnotes	83
CHAPTER 2 – <i>Lysins without Phage: Bacterial Genomes and Proviral Enzymes</i>	
•Introduction	86
•Materials and Methods	92
•Results	97
•Discussion	125
•Acknowledgments	130
•Endnotes	131
CHAPTER 3 – <i>Functional Viral Metagenomics I: The E-LASL Technique</i>	
•Introduction	132
•Materials and Methods	138
•Results	147
•Discussion	170
•Acknowledgments	176
•Endnotes	176

CHAPTER 4 – *Functional Viral Metagenomics II: A Novel Phage Lysin Screen*

•Introduction	179
•Materials and Methods	185
•Results	187
•Discussion	206
•Acknowledgments	211

CHAPTER 5 – *Back to Basics: Applying New Techniques to Genomic Samples*

•Introduction	212
•Materials and Methods	217
•Results and Discussion	224
•Acknowledgments	252
•Endnotes	253

CHAPTER 6 – *Future Prospects for Enzybiotic Identification*

•Reflections on the Current Project	254
•The Future of Genomics and Metagenomics	258
•Enzybiotics beyond Phage Lysins	262
•Concluding Remarks	268
•Endnotes	269

REFERENCES	270
-------------------	------------

APPENDIX OF DNA AND AMINO ACID SEQUENCES	329
---	------------

LIST OF FIGURES

•1.1 Phage Morphologies	3
•1.2 Phage Plaques	9
•1.3 Lysogeny versus Lysis	10
•1.4 Bacteriolytic Activity of Phage Lysins	28
•1.5 Bond Specificity of Phage Lysins	29
•1.6 Modular Architecture of Phage Lysins	35
•1.7 Genomic Architecture of Lysin-Associated Genes	46
•1.8 Lysin-Induced Clearing Zones	63
•2.1 PlyCM Purification	96
•2.2 Lysin-Like ORFs in <i>C. perfringens</i> Genomes	99
•2.3a Architectural Diversity among GH25 Phage Lysins	103
•2.3b GH25 Domain Alignment	106
•2.4 Purified PlyCM	107
•2.5 PlyCM-Induced Lysis	108
•2.6a Broad pH Dependence	111
•2.6b Narrow pH Dependence	112
•2.6c Irreversible Acid Denaturation	114
•2.7 Temperature Stability of PlyCM	115
•2.8a/b NaCl Effect on Lysis in ATCC 13124	117
•2.8c Light Microscopy of Lysin-Treated Cells in 150 mM NaCl	119
•2.8d SEM of Lysin-Treated Cells in 150 mM NaCl	120
•3.1 Metagenomics Flowchart	136
•3.2 Constructing an E-LASL	148
•3.3a E-LASL Amplification Products	150
•3.3b Digested DNA Prior to Linker Amplification	151

•3.4 Sequence Comparison of PlyB and PlyBeta	157
•3.5 PlyBeta Clearing Zones	159
•3.6a Positive Hemolysin Clone	164
•3.6b Aerolysin Phylogeny	166
•3.6c Purification of Aer _M	167
•3.6d Purified Aer _M	168
•3.6e Edman Sequencing of Aer _M	169
•3.6f Aer _M -Induced Hemolysis of RBCs	171
•4.1 Preliminary Hemolysis Screen	189
•4.2a Lysin-Encoding Clones with <i>Pseudomonas</i> Overlay (Genomic)	193
•4.2b Lysin-Encoding Clones with <i>Pseudomonas</i> Overlay (Metagenomic)	194
•4.3a Similarity among Gram-Positive Hits (Non-Bootstrapped)	202
•4.3b Similarity among Gram-Positive Hits (Bootstrapped)	203
•5.1 ΔPlySs1 Purification	221
•5.2 PlySs1 Prophage Lysin	225
•5.3 Purified ΔPlySs1	229
•5.4 PlySs2 Prophage Lysin	230
•5.5 ΔPlySs1 Enzymatic Specificity	232
•5.6 ΔPlySs1 pH Dependence	234
•5.7 ΔPlySs1 NaCl Dependence	236
•5.8 ΔPlySs1 DTT and EDTA Susceptibility	237
•5.9 ΔPlySs1 Temperature Stability	239
•5.10 ΔPlySs1 Bacteriolytic Activity	242
•5.11 ΔPlySs1 Growth Inhibition of <i>S. suis</i> 7711	245
•5.12 ΔPlySs1 Bacterial Strain Panel	249

LIST OF TABLES

•1.1 Advantages and Disadvantages of Phage Therapy	22
•1.2 Advantages and Disadvantages of Lysin Therapy	27
•1.3 Conserved Enzymatic Domains within Phage Lysins	32
•1.4 Conserved Binding Domains within Phage Lysins	36
•1.5 Lysins Cloned Through Functional Genomic Screening	65
•1.6 Lysins Cloned Through Genomic Sequencing	72
•2.1 PlyCM Species/Strain Specificity	123
•3.1a Taxonomic Distribution of Metagenomic BLAST Hits	161
•3.1b Top BLAST Hits by Genus	163
•4.1 Cloned Metagenomic Lysins	198
•5.1 CFU Analysis of Strains 7711, S735, 7997	244
•5.2a Analysis of Other <i>S. Suis</i> Strains	246
•5.2b Analysis of Other Gram-Positive Bacteria	248

CHAPTER 1

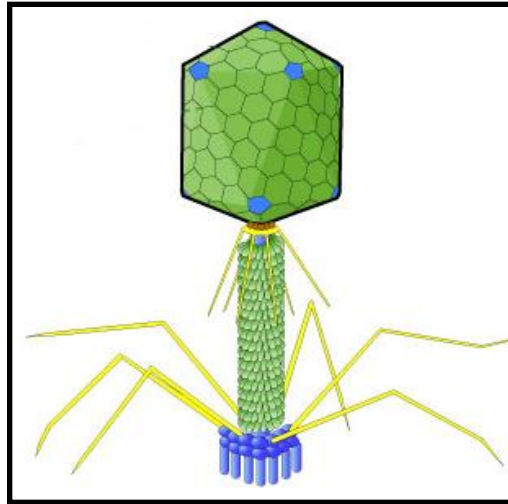
Phage, Phage Lysins, and their Molecular Identification

BACTERIOPHAGE AND THEIR BIOLOGY

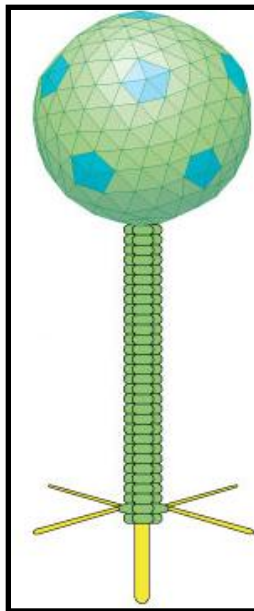
Nearly a century has now passed since the initial discovery of *bacteriophages*, a term that has come to encompass any virus that infects a bacterial host. In 1915, Frederick Twort reported that filtered suspensions of environmental samples were capable of producing “glassy areas” on plates of *Micrococcus*, and he proposed a bacterial virus as one of several possible causes (Twort 1915). Two years later, similar findings were reported for enteric bacilli by Felix d’Herelle, who was more adamant of the viral nature of the

phenomenon (1917). It was d'Herelle, in fact, who originally coined the term *bacteriophage*, derived from Greek and signifying *bacteria-eater*. At that time, it would have been difficult to predict the global significance of these viruses and the massive body of research that would go into studying their biology and applications. Bacteriophages (or phages) are crucial players in bacterial ecology and pathogenesis; they were important experimental tools in the development of the modern field of molecular biology; and they – along with the proteins they produce – can be utilized for various biotechnological and biomedical purposes, many of which have become evident only recently. It is this latter area of applied phagology that is the focus of the present thesis.

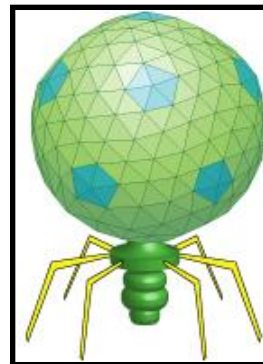
Bacteriophage taxonomy. Phages can be subdivided into several families (see Figure 1.1) based on their structural morphology and the nature of their genetic material (Fauquet et al. 2005). The large majority of the 5000+ phages examined to date belongs to the order *Caudovirales* (Ackermann and Abedon 2001), defined as non-enveloped viruses with protein tails and linear, dsDNA genomes (ranging from tens to 100+ kb in length). The caudoviruses are further subdivided into three families based on their specific tail-architecture: *Siphoviridae* – long, noncontractile tails; *Myoviridae* – intermediate length, contractile tails; and *Podoviridae* – short, noncontractile tails. Outside the caudoviruses, there exist several non-tailed families of phages with more divergent biological properties.



Myovirus



Siphovirus



Podovirus

Figure 1.1a Phage Morphologies – Caudoviruses

Phage in the order *Caudovirales* possess linear, dSDNA genomes and proteinaceous tails. The order is divided into the above three families based on the tail morphology (see text).

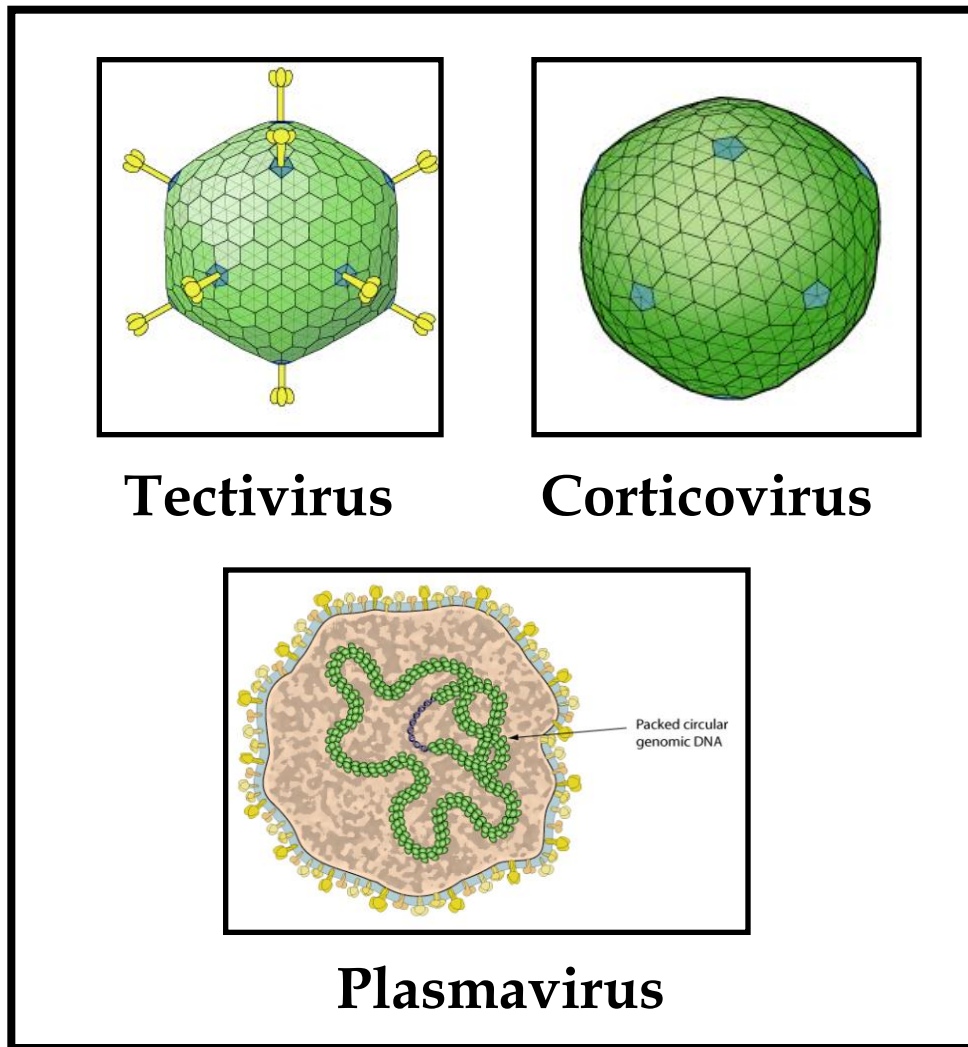
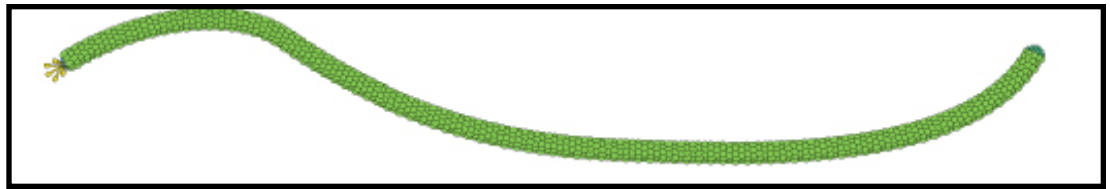


Figure 1.1b Phage Morphologies – Lipidic dsDNA Phage

The tectiviruses and corticoviruses both possess external icosahedral capsids with a lipid membrane lying directly beneath (not evident in the above images). The two families differ in terms of capsid and genome organization. Whereas the corticovirus genome is circular and highly supercoiled, the tectivirus genome is linear with terminal inverted repeats.

By contrast, the plasmaviruses possess an external lipid envelope, a pleomorphic geometry, and a circular genome. They are only known to infect the mycoplasmal genus *Acholeplasma*.



Inovirus (Filamentous Phage)

Microvirus

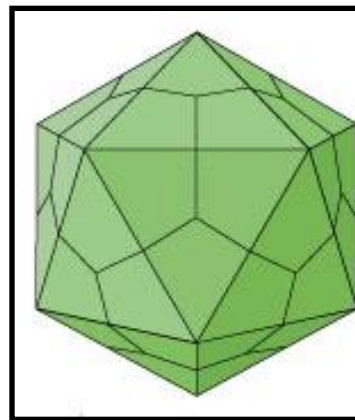


Figure 1.1c Phage Morphologies – ssDNA Phage

The inoviruses consist of a non-enveloped rod of filaments surrounding a circular, ssDNA genome. The microviruses possess a linear, ssDNA genome and a non-enveloped, icosahedral capsid.

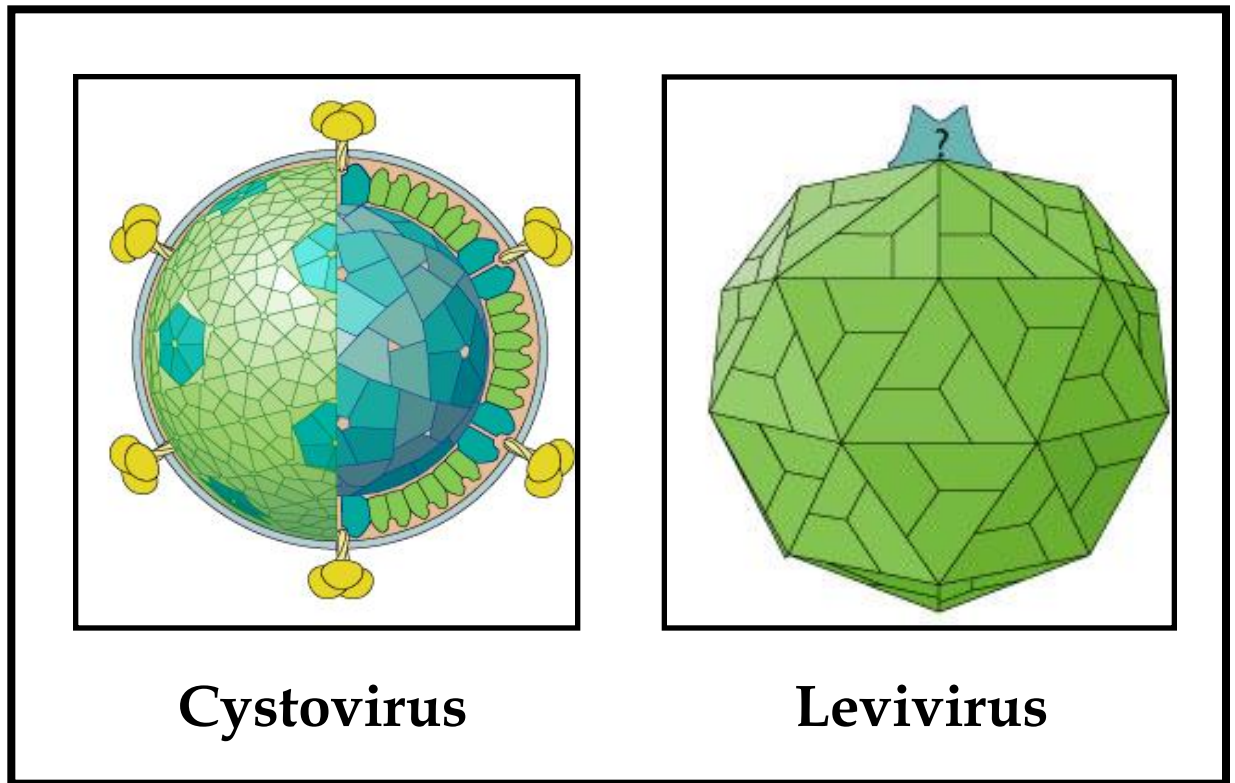


Figure 1.1d Phage Morphologies – RNA Phage

The cystoviruses possess a linear, segmented, dsRNA genome. They are characterized by a double capsid with a surrounding lipid envelope. The leviviruses possess a linear, positive-stranded, ssRNA genome and a non-enveloped, spherical capsid.

NOTE: The preceding images were taken, with permission, from the *ViralZone* website (www.expasy.org/viralzone, Swiss Institute of Bioinformatics). It should also be noted that three other phage families are known to exist in addition to the ones pictures above (*Lipothrixoviridae*, *Fuselloviridae*, and *Rudiviridae*). These viruses, however, are only known to infect archaea and are not considered here with the bacteriophage proper.

These include several other DNA phages: *Tectiviridae* and *Corticoviridae* – dsDNA, possessing a lipid membrane internal to the capsid; *Plasmaviridae* – dsDNA, possessing an external lipid envelope; *Inoviridae* – ssDNA with a filamentous morphology; and *Microviridae* – ssDNA with an icosahedral morphology. Phage families with RNA genomes include: *Cystoviridae* – dsRNA with an enveloped capsid; and *Levivirus* – ssRNA with a non-enveloped capsid.

The caudoviruses possess by far the broadest range of known hosts, and examples have been defined that infect virtually all bacterial phyla. For any single caudovirus, however, the host range is often limited to a particular bacterial species or a subset of strains within that species. Nonetheless, a number of phages have been observed that infect various species within a given genus or even across genera (Hyman and Abedon 2010). For the non-caudoviral phage families, the taxonomic range of hosts (at least among phage isolated to date) is far narrower. For example, the tectiviruses are the only non-caudoviral family known to infect Gram-positive bacteria, and several phage families (plasmaviruses, cystoviruses, corticoviruses) have only been defined for a single bacterial genus (Fauquet et al. 2005).

Phage life-style. In terms of the bacteriophage lifecycle, various examples have been observed among the above taxonomic families. These range from replication and extrusion in the absence of host death (in the case of the filamentous inoviruses, Calendar and Inman 2005) to rapid and complete host

lysis. For some of the rarer phage families, relatively few details are known on the molecular level. And even for the ubiquitous caudoviruses, the majority of mechanistic information is defined for model viruses (often *E. coli* λ phages), and these details do not necessarily apply universally. Broadly speaking, though, the lifecycle of the tailed phages can be broken down into two possible categories: lytic and lysogenic. The lytic (or virulent) phages co-opt the host's transcriptional apparatus following DNA injection, replicating and assembling within in the cytoplasm and inducing bacteriolysis from within. Within 30 minutes of initial infection, up to several hundred progeny viruses are released and the life-cycle begins anew (Karam 1994). Macroscopically, the tell-tale sign of actively-lysing phages is the clearing zone, or *plaque*, that results on a lawn of host bacteria (Figure 1.2). (The molecular mediators of host lysis – phage lytic enzymes – are the principle focus of this thesis and are described later in much greater detail).

By contrast, lysogenic (or temperate) phages are faced with a decision following infection: either they can [1] begin the above lytic cycle or [2] enter into a proviral state in which their DNA is incorporated into the host genome (Figure 1.3). The latter pathway is possible because these phages encode regulators that repress transcription of lytic-pathway proteins (Little 2005). Lysogenic phages also typically encode an integrase that is responsible for the site-specific recombination of their genome into the bacterial chromosome (van Duyne 2005).

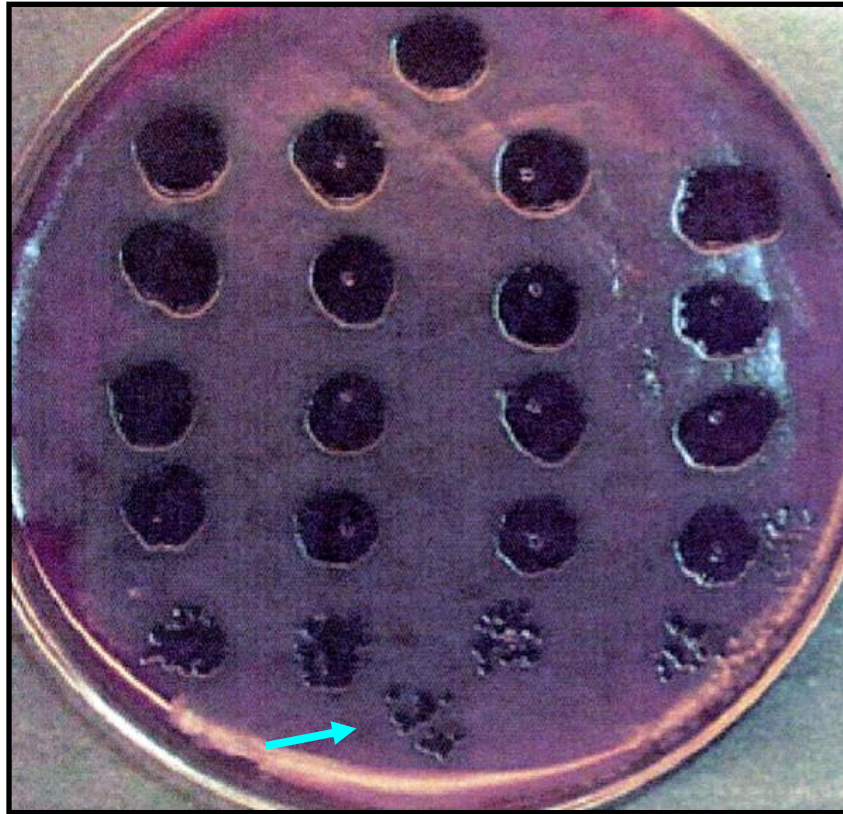


Figure 1.2 Phage Plaques

In this image, purified lytic phage have been dropped onto a lawn of their host bacteria. The resulting clearing zones are known as *plaques*. The majority of the above plaques are considered *macro-plaques*; here, the initial titer of phage was so high that it is impossible to differentiate individual plaque-forming-units (PFUs). At the bottom of the image, however, several *micro-plaques* are evident (indicated with arrow). Each one is derived an single ancestral phage (i.e. each micro-plaque is clonal). **NOTE:** The above image originally appeared in Abshire et al. (2005). It has been reproduced here with permission (American Society for Microbiology, Licensing Number 2482280006256).

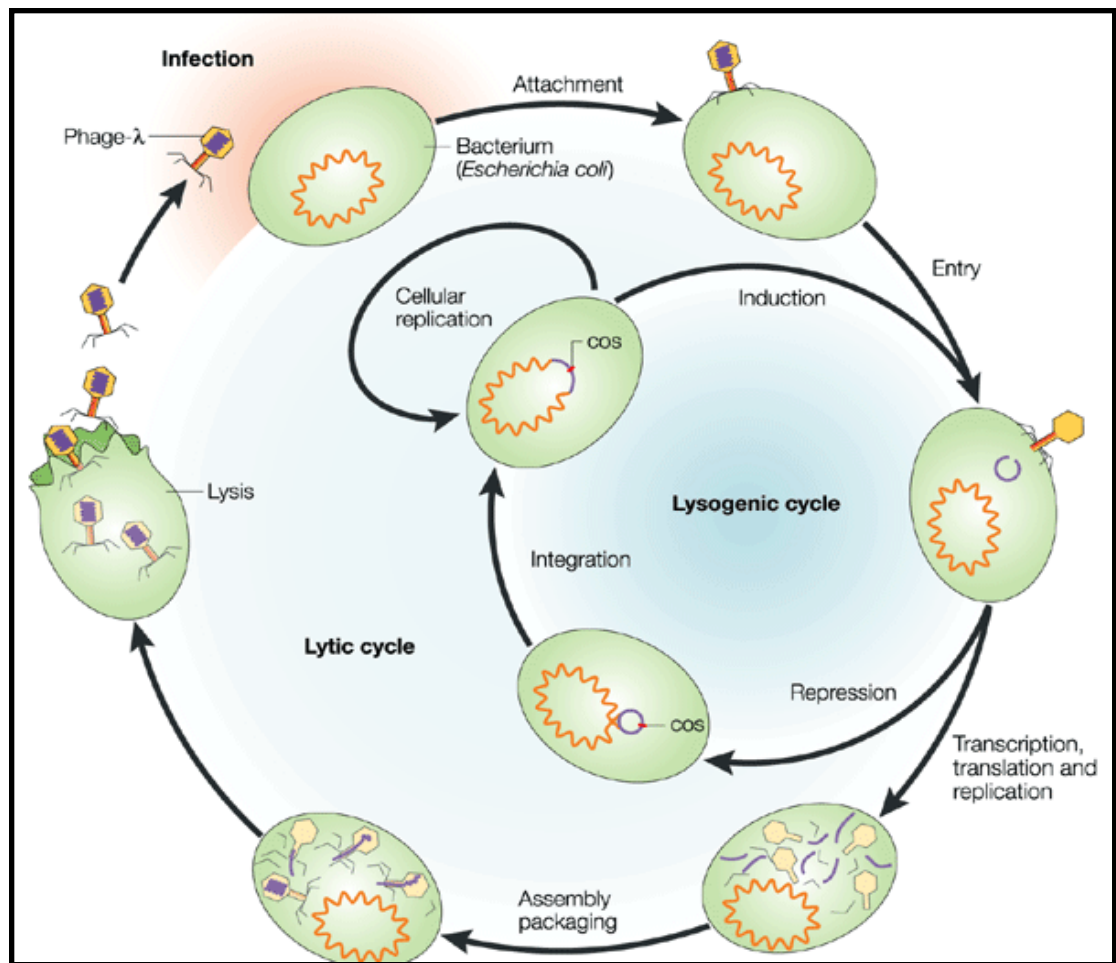


Figure 1.3 Lysogeny Versus Lysis

Upon initial infection, lysogenic phage may either begin the lytic cycle or enter into a proviral state (the lysogenic cycle). These alternate pathways are depicted above for *E. coli* and its λ-phage. In the image, the *cos* site of the phage genome is noted; this refers to the terminal portions of the linear DNA molecule that circularize prior to recombination with the host chromosome. **NOTE:** The above image originally appeared in Campbell (2003). It has been reproduced here with permission (Nature Publishing Group, Licensing Number 2482130951344).

Alternatively, some lysogenic phages do not physically integrate their DNA; rather, their genomes independently propagate in the host cytoplasm as linear (Ravin 2003) or circularized (Bourhy et al. 2005) episomes. It must be emphasized that the individual molecular steps governing lysogenization represent an expansive field unto itself (Ptashne 1994). Very broadly stated, the lysogeny-versus-lysis decision is governed by a combination of probabilistic factors, the number of infecting viral particles, and the host bacterium's metabolic status at the time of infection (Zeng et al. 2010).

Prophages can replicate for numerous generations within the host without synthesis of progeny viral particles. At some point, however, reactivation may take place: the phage genome is excised from the chromosome and viral replication commences, culminating in host lysis. The stimulus for reactivation is typically DNA damage (chemical or UV-induced) or other physiological stressors on the host. The best-characterized molecular pathway involves the bacterial SOS system, a response to damage-induced ssDNA in the cell (Little and Mount 1982). Even in the absence of stressors, some prophages will reactivate spontaneously within their host, albeit at a far less frequent rate ($\sim 10^{-4}$ to 10^{-5} cells, Little 2005).

Conversely, other prophages may remain incorporated for such an extended period that they decay and lose their ability to reactivate, becoming defective (or cryptic) (Casjens 2003). More recent research has indicated that prophage DNA, both intact and defective, may occasionally excise itself from the

chromosome without completion of the lytic pathway. This phenomenon was exploited as an elegant experimental means of curing a polylysogenized bacterial strain of its prophages *in vitro* (Euler 2010).

Phage ecology. With the complex interplay between phage and host, it is hardly surprising that phages are crucial players in the field of microbial ecology. In fact, the importance of phage in shaping the global microbiome cannot be understated, especially when one considers the ubiquity of these viruses within the environment. The numbers that have been reported on the topic are truly staggering: it is thought that 10^{31} phage particles exist globally (the majority residing in the oceans); there are up to 10 phages for every single bacterium on earth; an estimated 10^{23} phage infections occur every second; and ~20% of the marine biomass is turned over *every day* by phages (Suttle 2007; Hatfull 2008). Resultantly, phages are major participants in biogeochemical pathways, particularly when it comes to the solubilization of biomass and the dynamics of carbon respiration/fixation (Abedon 2006).

On the level of individual bacterial species and their genotypic and/or phenotypic properties, phages likewise serve crucial roles. In general, the movement of genetic information between cells via phages, *transduction*, is one major modes of horizontal gene transfer that exist for bacteria (Birge 2006). This includes both *generalized* transduction, in which bacterial DNA is mistakenly packaged into the phage capsid, as well as *specialized* transduction, in which a

prophage DNA excises itself from the bacterial genome but takes with it a stretch of adjacent chromosome. Lysogenization in itself can also lead to the acquisition of important genetic material. Although many viral genes are repressed in the proviral state (e.g. structural and lytic cassettes), this is not true for all prophage genes (Canchaya et al. 2004). Various prophage transcripts are expressed, including ones that can affect the behavior of the host and its ability to survive, proliferate, and infect (Broudy and Fischetti 2003).

The relationship between lysogenization and bacterial phenotype is particularly characterized as it relates to the virulence of bacterial pathogens (Breitbart et al. 2005). A number of pathogenic bacterial species have been identified for which a crucial virulence factor is prophage-encoded or included on chromosomal islands that have evolved from prophage (Novick et al. 2010). These include: shiga toxin for enterohemorrhagic *E. coli*, cholera toxin for *Vibrio cholera*, and superantigens for toxic shock-inducing *Streptococcus pyogenes*. The preceding list is far from exhaustive, however, and new phage-encoded virulence factors are often identified (for a more complete review of the topic, the reader is referred to Boyd and Brüssow 2002 or Chapters 7 - 18 of Waldor et al. 2005). Prophages could likewise enhance the ability of bacterial pathogens to persist in environmental reservoirs, as was recently proposed for *Bacillus anthracis* by Schuch and Fischetti (2009).

PHAGE IN BIOTECHNOLOGY AND BIOMEDICINE

The field of phage biology is clearly diverse, incorporating aspects of molecular genetics, ecology, and infectious disease. (The discussion to this point has not even touched upon the areas of bacteriophage evolution and interphage recombination, see Hendrix 2002 or 2005). At the same time, the above topics are all connected in that each one attempts to study phages *in their own right*. Basic phage research, however, is only one side of the coin as to why investigators are so interested in these microscopic juggernauts. Since their discovery, phages have also been utilized as *experimental tools* for answering broader questions facing the scientific community. *Applied phage research* is equally expansive, as phages and phage-products can serve in a variety of manners as industrial and biomedical agents. Illustrated in the following section, these other aspects of phage science have their own rich histories, ones that are still actively unfolding.

Phages and molecular biology. First and foremost, any discussion of phage-as-tools would be incomplete without mentioning their indispensable role in the development of the field of modern molecular biology. A number of classic experiments establishing the molecular principles of genetics were dependent on bacteriophages as model “organisms.” For instance, the nature of genetic mutations was revealed by experiments involving the *de novo* generation of phage resistance in *E. coli* (Luria and Delbruck 1943; Lederberg and Lederberg 1952). Following the initial identification of DNA as the heritable genetic

material (Avery et al. 1944), important confirmatory evidence was provided by Hershey and Chase using selective radiolabelling of the T2 phage (1952). And the triplet nature of DNA code was subsequently established by Crick et al. by observing the altered plaque morphologies following mutagenesis of the T4 phage (1961).

Moving ahead, while the discipline of molecular biology is now well-established, phages are still intimately involved in day-to-day research efforts. A variety of common laboratory reagents (primarily involved in nucleic acid modification) are recombinantly-expressed enzymes originating from phages. These include benchtop staples such T4 DNA ligase and polynucleotide kinase, as well as more recent additions such as *Bacillus* phage ϕ 29 polymerase (a highly processive polymerase used in whole-genome amplification). A review of phage-encoded proteins that could serve as applied tools was recently provided by Schoenfeld et al. (2010).

It is important to emphasize that a great deal of current research in molecular biology (including this thesis) involves the functional screening of DNA libraries to identify proteins with targeted properties. Phage-based cloning systems are commonplace in this regard. In fact, one of the first widespread cloning vectors was the *E. coli* λ -phage (Chauthaiwale et al. 1992). By inserting a fragmented DNA library into a specific position in the λ -genome (followed by *in vitro* viral packaging and host infection), one could generate a large set of clonal plaques, each encoding and expressing a unique recombinant insert.

In large part, λ -libraries have given way to other generalized cloning methods (involving plasmids, cosmids, and artificial chromosomes). Nevertheless, a related screening technique is still widely used for elucidating protein-ligand and peptide-ligand interactions: *phage display* (Paschke 2006; Gupta et al. 2005). Here, a DNA-library is ligated in-frame with a phage's coat protein (both filamentous and tailed phages have been used). This generates a set of recombinant phages, which are exposed in bulk to a molecular target immobilized on a solid interface. Phages that bind the target via their surface fusion-protein are isolated, enriched, and identified through subsequent rounds of infection and selection. Outside of basic scientific discovery, applications of phage display include the identification of drugs and drug targets, as well as the development of molecular imaging agents (Newton and Deutscher 2008).

Phage typing. While λ -cloning and phage display involve the generation of genetically-modified viruses, other important applications utilize wild-type phages for generating information. In the fields of epidemiology and clinical microbiology, susceptibility to phage infection is a classic method of typing bacterial strains on the sub-species level. By analyzing which phages (from a predefined panel) are capable of infecting a bacterial isolate, it is possible to track the flow of pathogens within the human population. Most notably, phage typing is used in the study of *Salmonella* strains (Threllfall and Frost 1990), although schemes for other bacteria have been reported (Engel 1978, Mokrousov 2009). In

a related application, the specificity that a phage demonstrates for a particular host can be utilized to identify an unknown bacterial isolate. For example, while clinical isolations are admittedly rare, a classic test for the positive identification of *B. anthracis* involves its susceptibility to infection by γ -phage (Brown and Cherry 1955; Abshire 2005). More recently, culture-free methods for the rapid identification of bacterial pathogens have been developed that rely upon ELISA-based detection of selectively-amplified phages (Rees and Dodd 2006; www.microphage.com).

Phage therapy. With these diverse applications, researchers have clearly exploited bacteriophages in creative and sophisticated ways. At the same time, there exists another avenue of applied phage research that – while tremendous in its potential benefits – is theoretically quite straightforward. Simply put, phages have the ability to kill bacteria, and bacteria have the ability to cause infections... therefore, phages might be used to treat infections! In fact, the idea of phage therapy dates back nearly as long as the discovery of phages themselves. Shortly after his (co)discovery of these viruses, d’Herelle began employing them to treat human infections, often with a great deal of reported success. These include his oft-cited application of phages to treat cases of *Shigella* dysentery in Paris in 1919, as well as later work with cholera and plague patients in India (Summers 1999). In 1923 he founded along with Georgian microbiologist Giorgi Eliava (who shortly after was executed by Stalin-era Soviet agents) the International

Bacteriophage Institute in Tbilisi, a center dedicated to phage-based treatment of infectious disease.

In the years that followed, the future for phage-as-pharmaceuticals seemed incredibly bright. (For more detailed historic and scientific information on phage therapy – from its beginnings to the present day – the reader is referred to numerous recent review articles and book chapters: Abedon et al. 2010 and accompanying articles; Górski et al. 2009; Housby and Mann 2009; Górski 2007; Merrill et al. 2006; Sulakvelidze and Kutter 2005; McKinstry and Edgar 2005; Sulakvelidze et al. 2001; Chanishvili et al. 2001; Summers 2001; Carlton 1999). By the early 1940's, more than 100 publications had been devoted to phage therapy (Krueger and Scribner 1941; Eaton and Bayne-Jones 1934), and pharmaceutical companies had begun marketing phage-based products in several countries (L'Oréal in France, Behringwerke in Germany, and Eli Lilly in the United States) (Häusler 2006).

Unfortunately, the results and reception of this initial work was mixed at best, with variable treatment success-rates and much skepticism among the scientific community. In retrospect, many of the initial problems can be attributed to poor purification quality, a lack of proper control groups in clinical trials, and insufficient understanding of phage biology (e.g. using lysogenic phage instead of strictly lytic phage) (Housby and Mann 2009; Sulakvelidze et al. 2001). Nevertheless, these results soon lead to the abandonment of phage-

therapy research in Western countries, especially in light of the development of small-molecule antibiotics that occurred around the same time.

By contrast, phage therapy continued unabated in Eastern Europe during the decades that followed. This work was especially prominent at the aforementioned International Bacteriophage Institute in Georgia (now known as the Eliava Institute, www.eliava-institute.org) as well as at the Hirszfeld Institute of Immunology and Experimental Therapy in Wroclaw, Poland (www.aite.wroclaw.pl/phages/phages.html). Over the years, these centers reported numerous successful trials for various combinations of Gram-positive and Gram-negative pathogens and anatomic sites of infection (Sulakvelidze et al. 2001). Due to language barriers and cold-war era geopolitical realities, however, few of the results reached (or were evaluated by) Western scientists at the time the research was conducted. It is only recently, in fact, that many written accounts of Eastern European phage therapy are starting to become available to the English-speaking world (Chanishvili 2009).

This broadened access to data has corresponded to a rekindled global interest in phage therapy. Changing attitudes have been motivated largely by the emergence of antibiotic-resistant bacterial pathogens (Wax et al. 2008), along with the decreasing efficiency and profitability associated with the development of new small-molecule antibiotics (Donadio et al. 2010). In this light, Western scientists began reconsidering phage therapy in earnest during the 1980's and 90's. The initial focus was on animal models of human disease, including rodent

and ruminant models of *E. coli* and *Pseudomonas aeruginosa* infections (Smith et al. 1982; Smith et al. 1983; Soothill 1994). Subsequent model studies dealt with other pathogen-animal combinations; a summary of this work was recently provided by O’Flaherty et al (2009).

Moving ahead rapidly, phage-therapy research in Western nations has now progressed to actual human trials. The first double-blinded clinical study involving phage therapy was recently completed in the United Kingdom (Wright et al. 2009). This phase II trial, which involved 24 patients with chronic *P. aeruginosa* otitis media, demonstrated both efficacy and safety in treated patients versus controls. Although efficacy data is still pending, a phase I safety trial was likewise recently completed in the United States involving phage treatment of venous leg ulcers (Rhoads et al. 2009). In addition to these studies, done in association with (respectively) BioControl Limited and Intralytix Incorporated, a number of other biotechnology companies have recently been established that are developing phage-therapy products in pre-clinical phases (reviewed in Housby and Mann 2009).

One should note that this research is taking place alongside – and, often, in collaboration with – the continued use of phage therapy in its traditional Eastern European bastions. For instance, a clinical trial is currently underway involving the Eliava Institute and Belgian physicians to treat burn infections with a cocktail of *Staphylococcus* and *Pseudomonas* phages, and initial quality-control data has already been reported (Merabishvili et al. 2009). Various other

examples of current phage therapy can be found in the above-mentioned review articles. A corollary message to all this work is that, if an individual patient is interested in phage therapy – and is willing to travel and pay – various avenues presently exist for obtaining it (www.phagetherapycenter.com).

It is important to emphasize that the potential utility of phage therapy extends beyond the treatment of human disease. Naturally, the same basic principles that apply to infections in man also apply to veterinary infections, and phage therapy has been explored as a novel means of treating disease in livestock and poultry (Johnson et al. 2008). The use of phage to treat plant pathogens is likewise an intriguing possibility (Balogh et al. 2010), and EPA-approved phage cocktails are already commercially available for controlling *Xanthomonas* and *Pseudomonas* infections (i.e. bacterial speck) in tomato and pepper plants (omnilytics.com/products/agriphage/agriphage4.html). In general, phage therapy shows great promise in the field of food science, as the viruses could be used to prevent spoilage or to curtail the transmission of food-borne illness. Just recently, in fact, the Food and Drug Administration approved a cocktail of *Listeria* phage for use as an additive to commercial meat and cheese products (FDA Code of Federal Regulations 21CFR172.785).

At this point, it is worthwhile to consider briefly the relative pros and cons of using purified phages as antibacterial agents. These points are summarized in Table 1.1.

Phage advantages	Phage disadvantages
•Easy to isolate and propagate	•More complex pharmacology than traditional small-molecule antibiotics
•Can overcome resistance	•Potentially immunogenic
•Self-replicating	•Subject to degradation as proteinacious compounds
•Act synergistically in a cocktail or in combination with other antibiotics	•Need to select for virulent phage to prevent genetic transfer
•Inhibits Gram-positive and Gram-negative organisms	•Bacterial strains can develop resistance
•Some phage products already have regulatory approval	•Many phages can have a limited host range
•Potential for use in numerous environments (human, animal, food, biofilm, etc.)	•Broader regulatory and consumer acceptance still required
•Historically have been in use for nearly a century	
•Possibility to genetically engineer phage	
•Specific bacterial targets	

Table 1.1 Advantages and Disadvantages of Phage Therapy

This table was adapted from O’Flaherty et al. (2009) with permission (John Wiley and Sons, Licensing Number 2486030498200).

The biggest advantage to phage therapy is largely self-evident: namely, that phages provide an alternate mechanism for killing bacterial pathogens, particularly those that already demonstrate antibiotic resistance. Phages also represent a unique sort of *smart antibacterial agent* whose concentration naturally amplifies itself at the relevant site of infection. Moreover, given their host specificity, phages should not affect commensal and probiotic human flora in the same manner as traditional antibiotics, a relevant issue given the increasing prevalence of antibiotic-associated illnesses such as *Clostridium difficile*-associated enteritis (Kuijper et al. 2006).

An inherent corollary to the latter point is that phage therapy is inherently not broad-spectrum, and could not be employed for empiric therapy before the causative species of an infection is defined. Likewise, even for a particular bacterial species, not all strains are equally susceptible to different phages. This would necessitate the use of phage cocktails to ensure coverage, a practice that is already common in ongoing trials. If phage therapy were to become widespread, an additional concern would involve the potential development of resistance among bacteria. As with small-molecule antibiotics, bacteria are known to possess several mechanisms for acquiring resistance to phage infection. These include the variation of surface epitopes (for initial phage binding), the horizontal acquisition of restriction enzymes, and the use of CRISPR genomic sequences (Hyman and Abedon 2010).

Outside these hurdles, the fact the phages are proteinacious particles naturally confers more complicated pharmacological parameters (involving application, biodistribution, biodegradation, and possible immunogenicity). Overall, while such drawbacks would likely preclude it from supplanting the role of antibiotics entirely, phage therapy could still serve as an important supplementary treatment strategy, particularly in the case of chronic infections or when initial options fail.

THE LYTIC ENZYMES OF BACTERIOPHAGES

Although it could be considered in a state of *product development* for nearly a century, phage therapy (at least the idea behind it) is a long-established antimicrobial strategy. At the same time, it is not the only phage-based approach for fighting bacterial pathogens. More recently, another strategy has arisen that does not utilize phages in their entirety, but rather a particular class of phage-encoded proteins: the *phage lytic enzymes* (also known as *phage lysins*, *endolysins*, or just *lysins*). As described previously, the final stage of infection for tailed phages involves the lysis of the host cell with release of viral progeny – the phage lytic enzymes are molecular facilitators of this event.

These enzymes are expressed late in the cycle of phage infection, and they are responsible for digesting the peptidoglycan of the host bacterium. This compromises the strength of the cell wall and subjects the bacterium (which experiences positive turgor pressure up to 25 atmospheres) to immediate osmotic

lysis (i.e. *bacteriolysis-from-inside*). Biotechnological interest in phage lysins has peaked in recent years after it was shown that exogenously-applied lysin can effectively kill Gram-positive bacteria (i.e. *bacteriolysis-from-outside*). This phenomenon has forced phage lysins into the spotlight as potential anti-infective agents in their own right, and has sparked a general interest in enzyme-based antibiotics, or *enzybiotics*. (As with phage therapy, a number of recent reviews have examined the biology and applications of phage lysins, including: Fischetti 2010; Villa and Veiga-Crespo 2010; Courchesne et al. 2009; O’Flaherty et al. 2009; Fischetti 2008; Hermoso et al. 2007; Fischetti et al. 2006; Borysowski et al. 2006; Fischetti 2005a; Fischetti 2005b; Young 2005; and Loessner 2005).

The concept of lysin therapy. The ability of phage lysins to act as antibacterial agents is fundamentally linked to the structure of the bacterial cell envelope. As noted above, the potential targets for lysin therapy would be limited to Gram-positive species. Without an outer membrane, their peptidoglycan layer is directly accessible to lysin treatment from the extracellular space. Gram-negative bacteria, by contrast, are generally insensitive to lysin treatment due to the protective effect of their outer membrane¹. Despite a narrower target range, lytic enzymes do offer an important advantage over intact phage: while bacteria have co-evolved with predatory phage for millions of years (hence their resistance mechanisms), exogenous lysin treatment is an unnatural phenomenon. Consequently, the development of resistance should not occur readily – to date,

acquired lysin resistant has not been observed in experimental trials (Loeffler et al. 2001; Schuch et al. 2002). Other relative advantages and disadvantages of phage lysin therapy are listed in Table 1.2.

Like intact phage, purified lytic enzymes would provide an alternative bactericidal mechanism for antibiotic-resistant pathogens. Much of the excitement surrounding these enzymes stems from the potency and specificity they demonstrate toward their particular Gram-positive targets. In buffered solutions (referred to as *in vitro* conditions), lysins exert a rapid lethal affect at low $\mu\text{g}/\text{ml}$ -concentrations (Schuch et al. 2002; Porter et al. 2007), sometimes even less (Nelson et al. 2001). Within minutes or seconds of lysin addition, a concentrated solution of live Gram-positive bacteria will undergo a several log decrease in viability with complete loss of turbidity (see Figure 1.4). Typically, a given lysin demonstrates activity against the bacterial species that the encoding phage infects (Loeffler et al. 2001), although some cross-reactivity can exist toward other species within the same genus or related genera (Nelson et al. 2001; Yoong et al. 2004). In the following paragraphs, I will provide a broad overview of the lysin field, including their biochemistry, genetics, and development as applied biomedical tools.

Lysin enzymology. A particular phage lysin can target one of several bonds within the peptidoglycan (or murein) macromolecule (Figure 1.5).

Lysin advantages	Lysin disadvantages
•Not self-replicating, more targeted defined control	•Not self-replicating
•Protein therapeutic	•Protein; therefore, susceptible to inactivation
•Resistance not yet reported	•To date not yet successfully applied against Gram-negative bacteria
•Possibility to genetically engineer lysins	•Potentially immunogenic
•Specific bacterial targets	
•Could be used as a prophylactic and for treatment	

Table 1.2 Advantages and Disadvantages of Lysin Therapy

This table was adapted from O'Flaherty et al. (2009) with permission (John Wiley and Sons, Licensing Number 2486030498200).

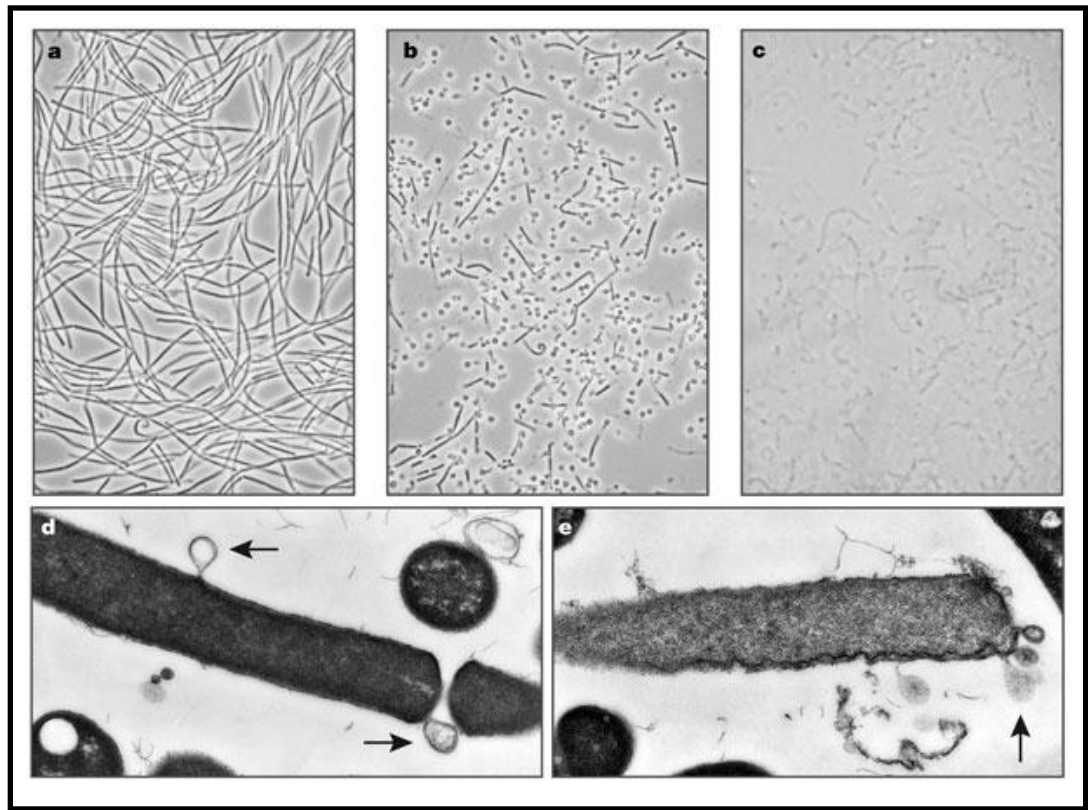


Figure 1.4 Bacteriolytic Activity of Phage Lysins

Depicted here is the effect of the PlyG lysin (encoded by the γ -phage) on *B. anthracis*. [A] Untreated cells, phase contrast microscopy (PCM); [B] 1 min post-treatment, PCM; [C] 15 min post-treatment, PCM; [D] 1 min post-treatment, transmission electron microscopy (TEM); [E] 10 min post-treatment, TEM. Overall, by 1 min post-treatment, the cells' gross morphology is compromised and cytoplasmic contents are budding through the cell wall (arrows). By 10 min, the cell wall is largely digested and only cellular ghosts remain. **NOTE:** The above image originally appeared in Schuch et al (2002). It has been reproduced here with permission (Nature Publishing Group, Licensing Number 2482150387937).

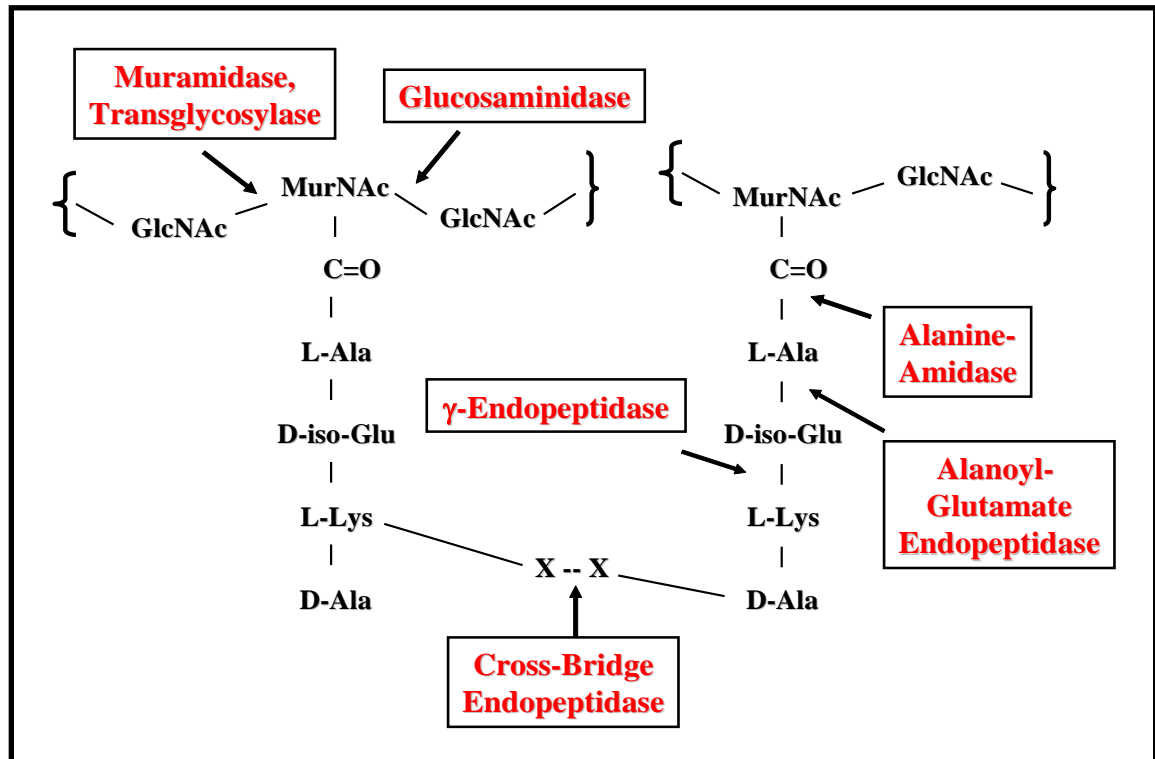


Figure 1.5 Bond Specificity of Phage Lysins

A summary of the chemical structure of peptidoglycan is depicted above (MurNAc = N-acetyl-muramic acid; GlcNAc = N-acetyl-glucosamine). Bonds targeted by known phage lysins are designated with an arrow. To date, lysins have been identified that target nearly every backbone, stem, and cross-bridge position within the macromolecule. The reader should note, however, that the above image is oversimplified: the identity of the cross-bridge (denoted here as X-X) can vary considerably among bacterial species. In Gram-negative organisms (and some Gram-positive ones), there is a direct cross-bridge between the position #3 and position #4 stem peptides. The identity of the stem peptides themselves can likewise differ after the position #1 L-Ala (see Vollmer et al. 2008 for examples).

The nomenclature of these enzymes is admittedly rather complex and (at times) inconsistent, but can generally be categorized as follows (Fischetti 2008). [1] *Muramidases* (also referred to as *muraminidases* or *lysozymes*) cleave the β 1-4 bond between N-acetylmuramic acid (MurNAc) and N-acetyl-glucosamine (GlcNAc). [2] *Glucosaminidases* cleave the β 1-4 GlcNAc-MurNAc bond. [3] *Lytic transglycosylases* (LTs) cleave the β 1-4 MurNAc-GlcNAc bond and reform a 1,6-anhydrobond within MurNAc (i.e. they do not generate a reducing end). [4] *Alanine-amidases* cleave the amide bond between the C3-lactate of MurNAc and the α -position L-alanine of the pentapeptide stem. [5] *Endopeptidases* cleave subsequent peptide bonds within the pentapeptide stem (β - or γ -positions) or the interstem cross bridge. The specific designation of an endopeptidase (e.g. glycine-glycine endopeptidase or alanoyl-glutamate endopeptidase) depends on its site of cleavage and the identity of the particular peptides, which vary from bacteria to bacteria (Vollmer et al. 2008a).

Muramidases, glucosaminidases, and LTs are collectively referred to as *endoglycosylases*, as they target the carbohydrate backbone of peptidoglycan at internal positions of the chain. Moreover, with the exception of the LTs, all phage lysins are also considered *hydrolases* based on their mechanism of bond cleavage. Collectively, people often refer to lysins as *peptidoglycan hydrolases* or *murein hydrolases* – rigorously speaking, this is incorrect due to the existence of LTs. Practically, though, this is of minor consequence concerning the role of lysins as enzymatic agents.

Another noteworthy distinction involves the difference between *enzymatic activity* and *protein domain*. The enzymatic region of most phage lysins can be categorized into one of various conserved domains. The bond specificity of the lysin, however, does not always correspond one-to-one with the identity of this domain. For instance, among the muramidase lysins that target the MurNAc-GlcNAc bond, some possess enzymatic domains that belong to the “phage lysozyme” family (Weaver and Matthews 1987), while others possess domains of the “glycosyl-hydrolase type 25” family (Porter et al. 2007). These two domains diverge sequentially and structurally, but still catalyze the same reaction. Conversely, CHAP domains have been described among different lysins that function alternatively as alanine-amidases (Nelson et al. 2006) or endopeptidases (Becker et al. 2009).

A complete list of enzymatic domains found in phage lysins is provided in Table 1.3, along with their corresponding bond specificities and representative examples. The domains are listed according to their Pfam nomenclature (Pfam is a central database of conserved protein domains – Finn et al. 2010, pfam.sanger.ac.uk). One should note that, even for a particular enzymatic domain, primary sequence diversity does exist from lysin to lysin, reflecting the evolutionary accumulation of mutations. Generally speaking, the degree of sequence homology between two domains of the same family reflects the phylogenetic closeness between the encoding phages.

Domain	Pfam Accession	Bond Specificity	Example	Reference
Phage Lysozyme	PF00959	Muramidase Lytic Transglycosylase	<i>E. coli</i> T4 lysin [NP_049736] <i>E. coli</i> λ lysin [NP_040645]	Inouye and Tsugita 1966 Taylor et al. 1975.
Soluble Lytic Transglycosylase	PF01464	Lytic Transglycosylase	<i>P. aeruginosa</i> KZ144 lysin [AAL83045]	Briers et al. 2007
Glucosaminidase	PF01832	Glucosaminidase	<i>S. agalactiae</i> LambdaSA2 lysin [NP_688827]	Pritchard et al. 2007
Glycosyl Hydrolase, Type 25	PF01183	Muramidase	<i>S. pneumoniae</i> Cp-1 lysin (CPL-1) [NP_044837]	García et al. 1988
Amidase, Type 2	PF01510	Alanine-amidase	<i>B. anthracis</i> γ lysin (PlyG) [YP_338200]	Schuch et al. 2002
Amidase, Type 3	PF01520	Alanine-amidase	<i>C. perfringens</i> phi3626 lysin (Ply3626) [NP_612849]	Zimmer et al. 2002b
Amidase, Type 5	PF05382	Alanine-amidase Endopeptidase (γ -position)	<i>S. pneumoniae</i> Dp-1 lysin (PAL) [O03979] <i>S. agalactiae</i> LambdaSA1 lysin [NP_687631]	Sheehan et al. 1997 Pritchard et al. 2007
CHAP	PF05257	Alanine-amidase Endopeptidase (various positions)	<i>S. dysgalactiae</i> C1 lysin (PlyC) [AAP42310] <i>S. aureus</i> K lysin (LysK) [YP_024461]	Nelson et al. 2006 Becker et al. 2009
VanY	PF02557	Endopeptidase (β -position)	<i>L. monocytogenes</i> A500 lysin (Ply500) [YP_001468411]	Korndörfer et al. 2008
M23 Peptidase	PF01551	Endopeptidase (cross-bridge)	See Caption	See Caption
U40 Peptidase	PF10464	Endopeptidase (unknown specificity)	<i>P. aeruginosa</i> ϕ 6 lysin [P07582]	Caldentey and Bamford 1992
Predicted Lysozyme (DUF847)	PF05838	Muramidase	<i>P. aeruginosa</i> ϕ 8 lysin [NP_524573]	Pei and Grishin 2005

Table 1.3 Conserved Enzymatic Domain within Phage Lysins

Table 1.3, continued

This table lists the enzymatic domains commonly encountered in phage lysins. For each domain, the pfam accession number is provided, along with a representative example among lysins whose properties have been studied recombinantly (the GenBank accession number of these is given in brackets). The table also lists the enzymatic specificity for each domain. As is evident, certain reactions can be catalyzed by more than one domain. In some instances, a conserved given domain is known to catalyze two different reactions, depending on the particular enzyme in which it is found. In these instances, an example is given for each reaction type. For the cross-bridge endopeptidases, a particular bond is not specified, as the nature of this bond varies considerably among different species.

For one domain (M23 endopeptidase), no representative example is given. No *definitive* phage lysins have (as of yet) been characterized that possess this domain. Nonetheless, M23 demonstrates peptidoglycan hydrolase activity in well-characterized non-phage enzymes (for instance, lysostaphin – see Kumar 2008), and the domain can be identified in the sequences of *putative* phage lysins. To date, the U40 endopeptidase has only been identified in the genome of a single dsRNA cystovirus. The DUF847 domain has been identified in a cystovirus and several caudoviruses (all Gram-negative), and muramidase activity was recently confirmed in one of the latter (Stojković and Rothman-Denes 2007).

Lysin architecture. For Gram-positive lysins, the enzymatic domain is one component of a larger overall architecture. As shown in Figure 1.6, these proteins possess modular structures that typically combine an N-terminal enzymatic domain with a C-terminal binding domain (Fischetti 2008). The latter recognizes (at nanomolar affinities) one of various epitopes within the target-cell envelope. These include surface carbohydrates (Loessner et al. 2002), choline moieties (García et al. 1990), or peptidoglycan itself (Buist et al. 2008). Like the enzymatic domains, the binding domains are also categorized into conserved protein families, a list of which is provided in Table 1.4. For certain domains shown in the table, the specific molecular target has not yet been defined experimentally. For some phage lysins, moreover, an extended C-terminal region is present, but cannot be identified computationally as a conserved domain. Most likely, these regions do serve binding functions, albeit ones that have not yet been identified/classified in an organized manner.

For a given bacterial species, it is not uncommon for its phage lysins to include various combinations of enzymatic and binding domains. For instance, one lysin might combine an N-terminal muramidase domain with a C-terminal SH3 domain, while another will combine an alanine-amidase with the SH3 domain (see Chapter 2). In general, phage genomes are modular and are known to evolve through extensive inter-phage recombination (Hendrix 2005).

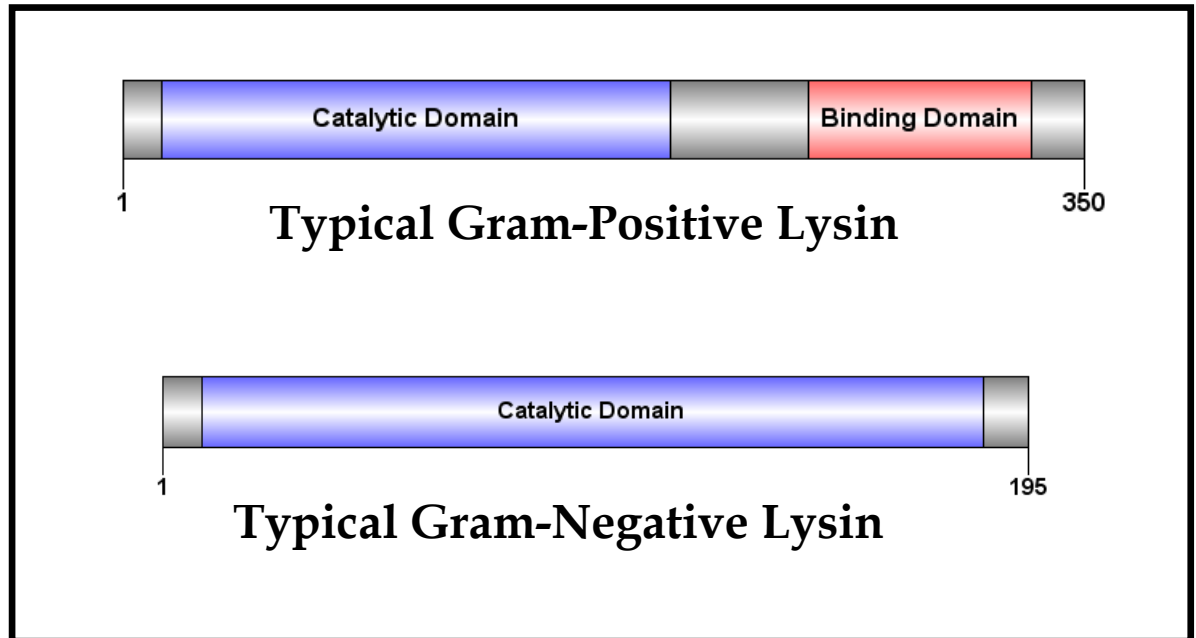


Figure 1.6 Modular Architecture of Phage Lysins

Typically, lysins for Gram-positive bacteria are characterized by an N-terminal enzymatic region and a C-terminal binding region. There are a variety of possibilities as to the particular identity of these domains (see Tables 1.1a and 1.1b for examples). Gram-negative bacteria are generally comprised of an enzymatic domain alone. It should be noted, however, that examples of lysins do exist (both Gram-positive and Gram-negative) that do not conform to these *standard* architectures. These atypical enzymes are discussed at various points throughout the text (in particular, see Chapters 4 and 5).

Domain	Pfam Accession	Binding Target	Example	Reference
Cell Wall Binding Repeat	PF01473	Choline	<i>S. pneumoniae</i> Cp-1 lysin (CPL-1) [NP_044837]	García et al. 1988
CPL-7	PF08230	Unknown	<i>S. pneumoniae</i> Cp-7 lysin (CPL-7) [P19385]	García et al. 1990
LysM	PF01476	Peptidoglycan	<i>L. fermentum</i> ϕ PYB5 lysin (Lyb5) [ABP88927]	Hu et al. 2010 ; Buist et al. 2008
Amidase-2 Associated domain	PF12123	Surface carbohydrate (GlcNAc-ManNAc)	<i>B. anthracis</i> γ lysin (PlyG) [YP_338200]	Schuch et al. 2002 and personal communications
SH3, Type 3	PF08239	Unknown, perhaps surface protein	<i>C. perfringens</i> 13124 prophage lysin (PlyCM) [YP_685420]	Chapter 2
SH3, Type 5	PF08460	Unknown, perhaps surface protein	<i>B. anthracis</i> BG-1 lysin (PlyBeta) [EU258891]	Schmitz et al. 2008
PG-1	PF01471	Peptidoglycan	<i>P. aeruginosa</i> KZ144 lysin [AAL83045]	Briers et al. 2007
PG-3	PG09374	Presumptively peptidoglycan	<i>P. aeruginosa</i> ϕ 8 lysin [NP_524573]	Pei and Grishin 2005

Table 1.4 Conserved Binding Domains within Phage Lysins

Table 1.4, continued

Listed here are binding domains (both experimentally-confirmed and presumptive ones) utilized by phage lysins. Domain accession numbers, representative examples, and corresponding references are provided. For cases in which a specific molecular target has been identified, this moiety is also given. Overall, several caveats must be made about this list. First, it is important to emphasize that it is far from complete. Numerous Gram-positive lysins – both confirmed enzymes and putative ones for which only a genetic sequence is known – possess extended C-termini for which no pre-classified domain is (yet) identifiable bioinformatically.

Second, for a given domain, the precise molecular target is not necessarily identical in every instance. For example, the *amidase 2-associated domain* targeted GlcNAc-ManNAc residues in the *B. anthracis* strain in which it was characterized. Although it is likely a carbohydrate-binding domain universally, the precise glycosidic signature recognized could vary from species to species (reflecting the primary sequence diversity within the domain itself). Finally, for the last two domains listed, the specific examples represent atypical examples of Gram-negative lysins with binding domains. For PG-1, Gram-positive lysins have also been identified with the domain (for instance, see Chapter 4). For PG-3, however, the only lysins known to possess it are Gram-negative (in fact, the example given is encoded by a dsRNA phage, although known caudoviral sequences also exist).

Presumably, these cut-and-paste lysin structures resulted from successful recombination events that occurred within the coding region of the lysin gene itself.

A general paradigm in lysin research is that the binding domain is largely responsible for the specificity an enzyme demonstrates toward its particular Gram-positive target. This concept derives from two different lines of experimentation. First, it has been demonstrated with various lysins that truncating the protein after its enzymatic domain greatly diminishes its bacteriolytic activity (Schuch et al. 2002; Porter et al. 2007). Second, in experiments with *Lactococcus lactis* and *Streptococcus pneumoniae* lysins, the enzymatic domain of one protein was recombinantly fused with the binding domain of the other (Sheehan et al. 1996). The authors observed that the latter domain was sufficient to maintain bacteriolytic activity, even when paired with an unnatural partner.

It should be emphasized, however, that these statements regarding the molecular basis of lysin specificity should be taken as *common principles* and not *universal facts*. Given the tremendous diversity of phage, one generally finds that every so-called *rule* in phage research has its noteworthy exceptions.

Accordingly, some lysins maintain high activity even with a deleted binding domain (Horgan et al. 2009). And chimeric lysins have been observed that fail to effectuate the lysis of either original bacterial species (unpublished observations and personal communications).

Although Gram-negative lysins have not (as of yet) demonstrated great potential as enzymatic agents, it is worth considering their structural architecture. Unlike Gram-positive lysins (250 - 400 amino acids), Gram-negative lysins are typically smaller (150 - 200 amino acids) and consist of an enzymatic domain alone (see Figure 1.6, Young et al. 2005). In this regard, they more closely resemble the nonspecific peptidoglycan hydrolases encoded by eukaryotic organisms, such as animal and plant lysozyme. Overall, there are several possible explanations for the lack of Gram-negative binding domains. Given the relative thinness of Gram-negative peptidoglycan (as little as a single layer) and its lack of decoration with associated molecules, a non-specific lysin might be all that is required to compromise envelope integrity (Schmitz et al. 2010b). By contrast, Gram-positive peptidoglycan is sufficiently thick (~40 layers) that the extra lytic potency/specificity afforded by the binding domain might be necessary for lysis.

An alternative (or complementary) explanation could involve the potential effect of free lysin in a phage-infected Gram-positive population (Loessner et al. 2002). After bacterial lysis, it is to the advantage of the progeny phage to locate other viable bacteria so that the infective cycle can expand. Soluble lytic enzyme, however, would compromise the viability of nearby cells and prevent them from becoming new hosts. In this regard, the enzyme's binding domain could keep it tethered to the remnants of lysed bacteria, functionally inactivating it. For Gram-negative bacteria, the outer membrane

desensitizes the cells to the effect of exogenous lysin in the first place. So even if molecules of lytic enzyme are diffusing in their vicinity, the cells would remain viable.

One should note that the above scenarios remain only theories and have yet to be confirmed with experimental evidence. On the whole, the *in vivo* function of the binding domain remains somewhat obscured by the fact that Gram-positive lysins have only been studied in the context of recombinant expression and exogenous treatment. To date, the effect of binding-domain deletions and chimeras has never been investigated in actively-propagating Gram-positive phages (Young 2005). The situation is further complicated because – as more and more enzymes are characterized – Gram-negative lysins have been identified that do, in fact, encode short N-terminal (Briers et al. 2007) or C-terminal (Pei and Grishin 2005) binding domains. Just as before, the traditional rules on lysin architecture (as depicted in Figure 1.6) seem full of apparent exceptions. The same is true for Gram-positive lysins, some of which contain multiple enzymatic domain (Baker et al. 2006; Pritchard et al. 2007; Cheng and Fischetti 2007; Becker et al. 2009) or even multimeric subunits (Nelson et al. 2006).

It is important to emphasize that the component domains listed in Tables 1.3 and 1.4 (both enzymatic and binding) are not exclusive to phage lytic enzymes. Bacteria themselves encode chromosomal peptidoglycan hydrolases that are involved in processes such as bacterial growth, division, sporulation,

and signaling (Vollmer et al. 2008b). Collectively, these enzymes are known as *autolysins*, and they can share many of the same domains as phage lysins proper. In fact, it is thought that phage lysins co-evolved with (or perhaps co-opted) the autolysins of their bacterial hosts (López et al. 1997). In terms of their modular architectures, some Gram-positive autolysins share an N-terminal-enzymatic-C-terminal-binding arrangement (García et al. 1985); for many others, however, the combination of domains is more diverse. Finally, even within phage genomes themselves, other structural proteins can include peptidoglycan-hydrolase domains. In particular, tail fibers often contain lytic components that participate in the process of initial DNA injection (Kanamaru et al. 2004; Kenny et al. 2004; Piuri and Hatfull 2006). Enzymatic head proteins have likewise been identified (Moak and Molineux 2004).

Lysin-associated proteins. Although lytic enzymes are ultimately responsible for affecting bacterial lysis during phage infection, they are not capable of this phenomenon by themselves. As with all bacterial proteins whose site of activity is external to the cytoplasmic membrane, the lysins need a mechanism for traversing this hydrophobic barrier. Unlike bacterial autolysins, which are typically secreted by the type II sec-mediated pathway, the phage lysins depend on accessory proteins known as *holins* (for reviews, see Gasset 2010; Young and Wang 2006; Young 2005; Wang et al. 2000).

These transmembrane proteins (typically 50 – 150 amino acids in length) insert into the cytoplasmic membrane and create lesions (i.e. *death rafts*) through which the lysin can diffuse to access the peptidoglycan. The holins, in fact, are ultimately responsible for the timing of lysis during phage infection. During the late stage of phage-gene transcription, lytic enzyme accumulates for a period of time within the bacterial cytoplasm. The actual moment of lysis does not correspond to a particular lysin concentration, but rather the rapid formation of holin-induced pores.

While the holins have a conserved (and relatively simple) biological function, their biochemical specifics are notably complex. Structurally, holins are organized into three classes based on whether they are comprised of one, two, or three transmembrane α -helices (four-TM holins have also been observed). On the level of primary sequence, however, these proteins are remarkable for the dissimilarity that individual holins can demonstrate toward one another, so much so that holins have been called “arguably the most diverse functional class of proteins known in biology” (Young 2005).

The regulation of holin activity is an intricate process involving the physical interaction of the holin with an additional protein, the *antiholin*, within the cytoplasmic membrane. In many cases, the antiholin is encoded by the same gene as the holin itself, and simply represents transcription from an alternate start codon (Bläsi and Young 1996); in other cases, it is encoded by an independent gene (Ramanculov and Young 2001). Overall, the formation of

functional membrane lesions is thought to depend on the holin-antiholin ratio, although the mechanistic specifics are variable from phage to phage, often poorly defined, and beyond the scope of the present text (see the review articles cited above for additional information).

Once again, it is worth noting that certain phage have been identified where the typical relationship between lysin and holin does not necessarily apply. A phage infecting the Gram-positive bacterium *Oenococcus oeni* encodes a lysin with a canonical signal peptide for the sec-mediated secretion pathway (Parreira et al. 1999). Likewise, a secreted autolysin of *Enterococcus faecalis* was recently shown to correspond to the lytic enzyme of an integrated prophage; it, too, includes an N-terminal signal peptide (Mesnage et al. 2008). Interestingly, both the *Oenococcus* phage and the *Enterococcus* prophage also encode putative holins, and it remains unclear the role of this apparent secretory redundancy.

Another variant situation exists for a group of Gram-negative phages (first defined for the *E. coli* P1 phage) whose lysins encode an N-terminal signal-arrest-and-release (SAR) sequence (Xu et al. 2004). Although the host's sec system initially engages and exports these lysins, signal peptide cleavage does not occur. Instead, the lysins remain tethered to the membrane as inactive, periplasmic proteins. For these phages, the holins (known as pinholins) function by depolarizing the membrane without the formation of macromolecular lesions (Park et al. 2007). This effect, however, is sufficient to release the lysins from

their membrane-bound state, thereby activating the enzymes and inducing bacterial lysis.

In addition to the holin and lysin, Gram-negative phages can encode two additional lysis-related proteins that are responsible for destabilizing the outer membrane (Young 2005). Generically, these proteins are referred to as Rz/Rz1-like proteins, after the first such examples to be identified in the *E. coli* λ phage (Young et al. 1979; Zhang and Young 1999). Rz is anchored to the cytoplasmic membrane, while Rz1 is an outer membrane-associated protein. The two are believed to physically interact in the periplasmic space, leading to the juxtaposition and fusion of the two membranes (Berry et al. 2008). Phages with a defective Rz/Rz1 cassette are capable of lysis and propagation, but are inhibited by the extracellular presence of divalent cations (Young et al. 1979).

In terms of their primary sequence, these proteins (much like the holins) are quite divergent from one another. After the initial identification of Rz/Rz1, a paucity of homologues was identified in other Gram-negative phages. However, more recent bioinformatic predictions (combining topology and genomic-architecture analysis) have revealed that orthologous proteins are nearly ubiquitous in sequenced Gram-negative phage (Summer et al. 2007). For some phages, the Rz/Rz1-like proteins seem to be combined as a single polypeptide (a *spanin*) that interacts simultaneously with the inner and outer membrane.

Genomic arrangement of phage lysins. As the preceding sections indicate, at least two (and sometimes more) proteins are involved in host-cell lysis during phage infection. It is worth considering the genes that encode these proteins and, specifically, their proximity within the viral genome. This issue is particularly relevant to Chapter 4 of the present thesis. If one had to designate a *standard* genomic arrangement for lysin-associated genes, it would be as follows (Wang et al. 2000). The holin/antiholin-encoding ORF is located immediately upstream and in the same orientation as the lysin-encoding ORF (sometimes the two reading frames overlap by several nucleotides). These two genes are thus under control of the same late-stage promoter. In Gram-negative phage, the Rz and Rz1 genes would also be located nearby (either upstream or downstream), and often the Rz1 gene is embedded within the Rz gene in a different reading frame (Summer et al. 2007). These *standard* genomic architectures are depicted in Figure 1.7.

Yet again, though, various other arrangements have been reported for both Gram-positive and Gram-negative phages. These include examples where the lysin is encoded immediately upstream of the holin, where the lysin and holin are separated from one another (Dunn and Studier 1983; Schuch et al. 2004), where the holin gene is imbedded within that of the lysin (Loessner et al. 1999), and where the phage encodes two putative holins (Delisle et al. 2006) or two lysins (van der Ploeg 2007).

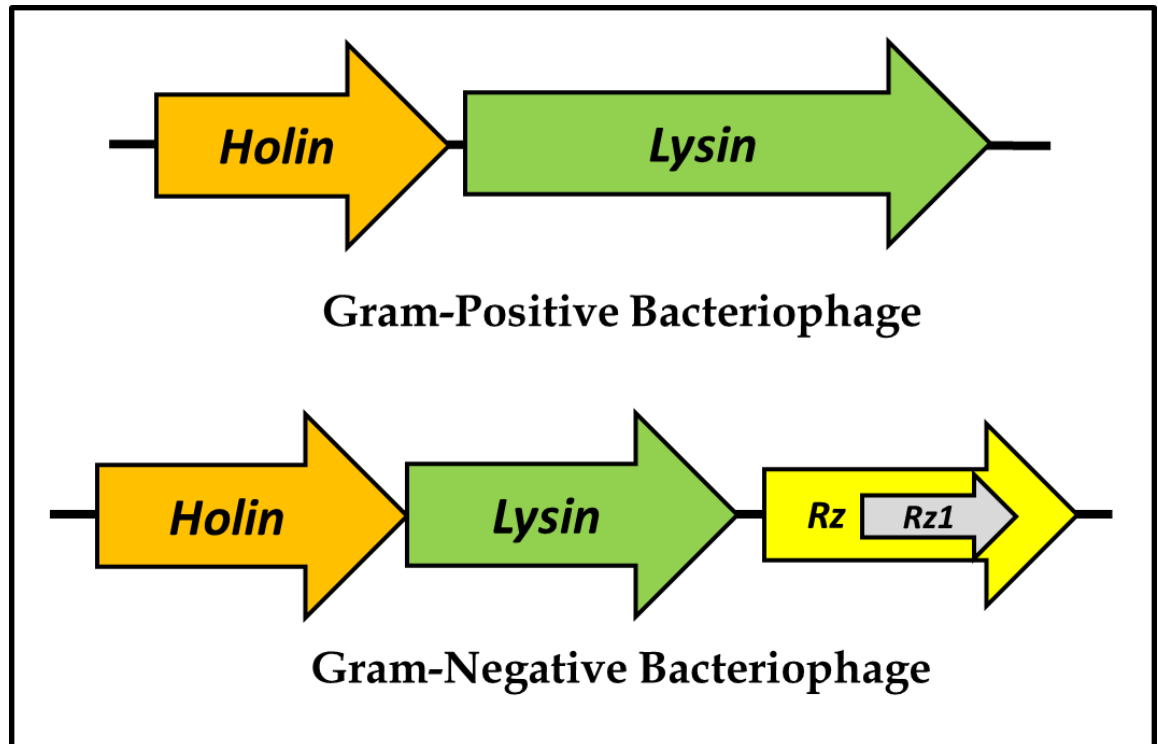


Figure 1.7 Genomic Architecture of Lysin-Associated Genes

The standard genomic relationship among lysin-associated genes is depicted here. The holin-encoding gene is encoded immediately upstream from the lysin-encoding gene, and both are under control of the same late-stage promoter. Often, the holin-encoding gene also encodes the regulatory antiholin: transcription of the entire ORF yields the antiholin, while transcription from an alternate start codon (anywhere from 2 - 10+ amino-acid residues downstream) yields the holin. In Gram-negative phage, the OM-disrupting genes (Rz/Rz1) are also frequently encoded nearby. Here, they are depicted immediately downstream from the lysin, the arrangement found in the *E. coli* λ phage (Young 2005). Overall, it is important to emphasize that the above arrangements for these genes do not hold true universally. Various other potential architectures are cited in the text.

In addition, several phage genomes have been sequenced that (based on homology analysis) do not seem to encode a lytic enzyme at all (Scholl et al. 2004). Several possibilities could explain this apparent omission. These phages could encode completely novel lysins that do not correspond to any known enzymes. Alternatively, they might utilize a structural peptidoglycan hydrolase (i.e. a tail-lysin) as the lytic agent, or rely upon an endogenous bacterial enzyme to digest the cell wall. Finally, lysin-encoding genes have been identified among certain streptococcal and staphylococcal phage that include a self-splicing, group I intron (Foley et al. 2000; O'Flaherty et al. 2005). Truly, it appears that the potential variations that have evolved in phages are only limited by one's own imagination!

Non-caudoviral lytic proteins. The discussion to this point has focused exclusively on the lytic mechanisms of the caudoviruses. This is understandable given their overwhelming predominance among known phage. Nonetheless, a dedicated body of research has also studied lysis for non-tailed phage; this work will be summarized briefly here. As mentioned previously, all phage families except the filamentous inoviruses release progeny viruses through host-cell lysis. For several of these families, the mechanism is reminiscent of the caudoviruses and involves phage-encoded peptidoglycan hydrolases. This is known to be the case for the dsRNA cystoviruses (Caldentey and Bamford 1992; Pei and Grishin 2005) and the lipid-associated dsDNA tectiviruses (Verheust et al. 2004). For

another lipid-associated dsDNA family, the corticoviruses, the mechanism remains ambiguous. The genome of a single corticovirus (the only known isolate) has been sequenced and it fails to include any proteins with predicted murlaytic activity (Krupocič et al. 2007a). Nevertheless, genomic analyses of various marine bacteria have revealed corticovirus-like prophage elements with putative lytic enzymes (Krupocič et al. 2007b).

A divergent, and rather intriguing, strategy of host lysis is employed by the ssRNA leviviruses and the ssDNA microviruses. These phages encode proteins that are collectively referred to as *amurins*. Rather than hydrolyze peptidoglycan, the amurins inhibit the biosynthetic machinery that is responsible for its synthesis (Bernhardt et al. 2001; Bernhardt et al. 2000). In a mechanism similar to small-molecule antibiotics, they comprise the cell-wall strength of the nascent bacteria and induce osmotic lysis. For the dsDNA plasmaviruses, the mechanism of lysis is not well-established. Given that these phages are only known to infect mycoplasma (which lack a cell wall), the mechanism almost certainly does not involve a peptidoglycan hydrolase. Overall, when all the non-tailed phages are considered, only the tectiviruses have been observed to infect Gram-positive species (including the pathogen *B. anthracis*). So – at least at current state of the technology – they are the only ones whose lysins seem compatible with enzybiotics usage.

Phage lysins into the spotlight. As the discussion to this point illustrates, phage-lysin research is clearly an expansive field. At the same time, one important question has not yet been addressed: *when* exactly were these proteins first viewed as potential anti-infective agents? The truth is that researchers were aware of lysins for decades before they were first considered pharmacologically. (A detailed history of the “nuts and bolts” of lysin identification will be provided a little later, as the last main section of this introduction.) It was not until the late 1980’s, in fact, that lysins were first viewed through the prism of applied microbiology. Interestingly, some of this initial work did not have biomedical motivations. Several lysins were studied that targeted species of lactic acid bacteria (LAB) commonly employed by the dairy industry (reviewed in Sable and Lortal 2005). It was proposed that lysins could facilitate the release of intracellular enzymes from LAB, enhancing the ripening of fermented food products.

Throughout the 1990’s, two other groups – García, López and colleagues in Madrid and Loessner, Scherer, and colleagues in Munich – began investigating phage lysins *in vitro* against various human pathogens, including pneumococci (García et al. 1988 and 1990), staphylococci (Loessner et al. 1998), and *Listeria* (Loessner et al. 1995). While their work demonstrated bacterial killing, the lysin field did not truly expand until after the turn of the millennium. It was at this point that Fischetti and colleagues at Rockefeller began utilizing phage lysins *in vivo* with animal models of bacterial pathogenesis. Published in 2001, two of the

first such studies involved pharyngeal colonization of mice with *S. pneumoniae* (Loeffler et al.) and *S. pyogenes* (Nelson et al.). In both cases, a single treatment with the respective lysins PAL and PlyC abolished bacterial titers hours after dosing. Similar results were later reported for the PlyGBS lysin following the pharyngeal and vaginal colonization of mice with *S. agalactiae* (Cheng et al. 2005).

A common feature of these trials is that they involved selective decolonization of mucous membrane sites. This is highly relevant given the pathogenesis of many common Gram-positive bacteria. For virtually all streptococcal and staphylococcal pathogens, the presence of disease is not simply a binary issue of “having the bacteria” or “not having the bacteria”. Countless individuals are colonized with these organisms, generally at mucous membranes, without any clinical signs of infection. The progression to actual pathology (either at the initial mucous membrane or at distal sites) depends on this initial colonization, but it is not synonymous with it. In this light, phage lysins – or, for that matter, intact phages – could be used as targeted prophylactic agents that prevent the colonization/infection transition. This application is be relevant even outside the discussion of burgeoning antibiotic resistance, although the latter issue imparts an even greater sense of priority.

Subsequent *in vivo* trials have investigated the effect of lysin treatment against fulminant Gram-positive infections. For example, Schuch et al. demonstrated the ability of the PlyG lysin to prevent lethality in mice with *B. cereus* peritonitis (2002). This work generated significant attention at the time,

especially in light of heightened bioterrorism concerns over the closely-related species *B. anthracis*. In other work, the pneumococcal lysin CPL-1 (different than the PAL lysin mentioned above) has been used successfully in rodent models of bacteremia (Loeffler et al. 2003), endocarditis (Entenza et al. 2005), otitis media (McCullers et al. 2007), and meningitis (Grandgirard et al. 2008).

Just recently, an anti-staphylococcal lysin (ClyS) was utilized synergistically with traditional antibiotics to treat cutaneous *S. aureus* infections in mice. This enzyme is currently employed in an *ex vivo* trial involving skin scraping from psoriasis patients (Daniel et al. 2010; Pastagia and Fischetti, personal communications). Admittedly, the ability to treat particular anatomic sites with lysin could be limited by the fact that it is a protein. Nevertheless, a growing body of research (reviewed in Fischetti 2008) has addressed the issues of biodistribution and immunogenicity of lytic enzymes. The results provide cautious optimism that these agents might be useful even beyond their most straightforward application as topical agents. Moreover, just as with bulk phage therapy, lysins also have potential applications as food additives (see Chapter 2) and veterinary pharmaceuticals (see Chapter 5). They could likewise be used as industrial decontaminating agents for surfaces that have been colonized with planktonic bacteria or biofilms (Donlon 2009).

PHAGE-LYSIN IDENTIFICATION THROUGHOUT THE YEARS

Despite the work dedicated to phage lytic enzymes of late, the number of lysins that have been studied functionally is still only a small fraction of the total lysin pool encoded by global phage. With the numbers alluded to above (10^{31} worldwide phage particles), bacteriophage are thought to be the single greatest source of genetic information on the planet. Even among the known phage/prophage of Gram-positive pathogens, only a small minority have seen their lytic enzymes cloned, expressed, and examined as anti-infective agents. In itself, the sheer number of lysins with the potential for development as enzybiotics makes this class of proteins extremely attractive pharmacologically.

Considering this magnitude, future research will undoubtedly continue to isolate new lysins to complement those already in development. And while it is the activity of these enzymes that will ultimately garner attention, the success of such work is fundamentally dependent on the techniques employed to identify the proteins in the first place. The purpose of the current section, therefore, is to address the issue of lysin identification. Following a summary of lysin isolation prior to modern molecular cloning, I will outline the various techniques currently available for cloning lysin-encoding genes within phage genomes. The relative advantages and disadvantages of these approaches will be discussed, especially in the context of ongoing technological advances. Overall, through the efficient identification of novel phage lysins, researchers can only broaden the

potential impact of these proteins and hasten their development into true clinical and industrial tools.

Historical perspectives on lysin identification. While medical interest in phage lysins is a more recent phenomenon, initial observations of these enzymes and attempts to purify them date back much longer. In 1921, Felix d'Hérelle first proposed the existence of a phage-associated enzyme that was capable of lysing bacteria independently from total phage action. His idea was based on experiments with alcohol-denatured phage; d'Hérelle was able to extract an active agent that could lyse bacilli but not propagate between cultures. Soon after, Vladimir Sertic reported the isolation of a *lysine d'une race du bactériophagie* that was responsible for creating altered morphological zones that surrounded *Escherichia coli* plaques proper (1929). The lysin hypothesis was contested at the time, and it is difficult to judge whether these initial observations were truly due to the activity of what we now know as phage lytic enzymes. Nevertheless, other contemporary studies did report the tendency of nonviable Gram-positive bacteria to lyse when in the presence of live bacteria and their corresponding phage (Gratia and Rhodes, 1923; Twort 1925; Bronfenbrenner and Muckenfuss, 1927). In retrospect, this activity could be attributed to the diffusion of lysin from dead cells.

It was not until several decades later that the existence and activity of lysins were broadly accepted by the scientific community. In 1955, Ralston et al.

described a lytic agent that appeared in the supernatant of a *Staphylococcus aureus* culture following infection by bacteriophage P₁₄. This protein, which they termed *virolysin*, could be separated from intact phage through ultracentrifugation and concentrated by ammonium sulfate precipitation (Ralston et al., 1957). Their work was notable in that the phage enzyme was isolated and characterized alongside an endogenous lytic enzyme produced by the host staphylococci – the distinction between these molecules allowed the phage origin of virolysin to be established. It was subsequently demonstrated that antigenically-distinct lytic enzymes could be isolated when the *S. aureus* host was infected with diverse phage (Ralston and McIvor 1964). In later works, two staphylococcal lysins (from phages 80 and 53) were purified beyond ammonium sulfate precipitation by ion exchange chromatography (Doughty and Mann 1967; Sonstein et al. 1971).

One should note that, in these initial attempts to isolate *S. aureus* lysins, the final products did not demonstrate the degree of activity currently associated with lytic enzymes. Of the above (semi-pure) enzymes, only the phage-53 lysin could successfully lyse viable staphylococci; the phage-80 lysin was only active against isolated cell walls and the P₁₄-virolysin required the bacteria to be “sensitized” by one of several additional agents. It is unclear whether this diminished activity was due to less-than-ideal purification conditions at the time, or simply reflect the efficacy of these particular enzymes.

During the same period as this work on staphylococcal enzymes, similar research was progressing on other bacteria-bacteriophage combinations. Again with the strategy of isolating the proteins from culture supernatant, lysins were investigated from phage whose hosts included enterobacteria (Inouye and Tsugita 1966; Maass and Weidel 1963; Rao and Burma 1971), bacilli (Murphy 1957; Welker 1967), lactococci (Tourville and Johnstone 1966), and streptococci (Reiter and Oram 1963; Oram and Reiter 1965). Overall, one of the most extensively studied lysins of this time was that of the C₁ phage infecting Lancefield group C Streptococci. Two 1957 publications documented its ability to lyse not only live Streptococci of the same type, but also live Group A and E Streptococci (Maxted; Krause).

Over the following years, several increasingly sophisticated attempts were made to purify the C₁ lysin using a combination of ammonium sulfate precipitation, gel filtration, and calcium phosphate adsorption (Krause 1958; Doughty and Hayashi 1961). In the process, it became a valuable tool for selectively removing and characterizing antigenic components of the streptococcal cell envelope. The final obstacle to achieving a highly pure protein preparation (the presence of reactive sulfhydryl groups) was overcome by Fischetti et al., who utilized reversible sodium tetrathionate protecting groups to stabilize the lysin during chromatography (1971). Interestingly, these non-recombinant purification schemes for the C₁-phage lysin, currently referred to as PlyC, endured into the modern era of lysins as enzybiotics. In 2001, PlyC was

one of the first lysins to effectively decolonize a bacterial pathogen *in vivo* (Nelson et al.), even though the encoding genomic region was not cloned and characterized until several years later (Nelson et al. 2006).

Into the age of molecular cloning. Following the elucidation of the central dogma of molecular biology, attempts to characterize (and ultimately identify) phage lytic enzymes through genetic techniques began in earnest by the late 1960's. The first lysin amino acid sequence was reported in 1966 by Inouye and Tsugita for the *E. coli* T4 phage muramidase. It was determined through Edman analysis of culture-purified enzyme, and it suggested an approximate nucleotide sequence by reverse translation. This work on the T4 lysin, in fact, provided important evidence confirming the very nature of the triplet genetic code. By inducing mutations in the T4 genome and observing the corresponding frame-shifts in the purified lysin, the authors were able to verify Crick's hypothesis regarding the language of DNA codons (Terzaghi et al. 1966; Okada et al. 1968).

The definitive nucleotide sequence for the T4 lysin was not published until 1983, when Owen et al. successfully cloned the gene by its ability to rescue a lysis-defective phage strain. This sequence, in turn, allowed several additional lysin genes to be recognized through nucleotide homology. In 1985, the *Salmonella* P22-lysin was identified in this manner from a sequenced fragment of the viral genome (Rennell and Poteete). The first nucleotide sequence of a Gram-positive lysin, from *Bacillus subtilis* phage ϕ 29, was likewise reported one year

later (Garvey et al. 1986). It is important to note that the above work was conducted at the same time as the advent of modern protein expression technology; shortly after their cloning, the T4 and ϕ 29 lysins were the first such enzymes to be expressed and purified recombinantly from *E. coli* (Perry et al. 1985; Saedi et al. 1987).

With molecular techniques at their disposal, researchers could now identify lysins directly from phage DNA, and this ability was soon utilized to search for enzymes against Gram-positive pathogens. One of the first bacterial species targeted in this regard was *S. pneumoniae*, as the 1980's and 90's saw the cloning, expression, and functional analysis of a number of pneumococcal phage lysins. The same general strategy was employed in identifying each of these proteins: following the cloning of an initial prototype enzyme, the gene was utilized to identify related lysins through Southern blot analysis. Ironically, the prototype enzyme for these pneumococcal lysins was not actually viral in origin, but rather a genomic peptidoglycan hydrolase encoded by *S. pneumoniae*, LytA. Like the endogenous *S. aureus* lysin from Ralston et al. (1957), LytA is not a phage enzyme but rather a chromosomal autolysin. Specifically, LytA is an alanine-amidase with various roles in pneumococcal physiology, including the release of cytoplasmic virulence factors during pathogenesis (Jedrzejewski 2001) and the predation of non-competent cells by competent ones within pneumococcal communities (Guiral et al. 2005).

García et al. cloned LytA in 1985 through a complementation strategy in which an *S. pneumoniae* genomic fragment was identified by its ability to rescue autolytic activity in a mutant *lytA*⁻ strain. The resultant sequence was subsequently used to probe genomic DNA from several *S. pneumoniae* phages in order to isolate their lysin genes. Successful hybridizations were noted in many instances, and the corresponding bands were cloned and sequenced to reveal viral LytA-homologues. Specifically, the lysins from the following phage were identified [with corresponding enzyme names]: Cp-1 [CPL-1] (García et al. 1988); Cp-7 and Cp-9 [CPL-7 and CPL-9] (García et al. 1990); HB-3 [HBL] (Romero et al. 1990); EJ-1 [EJL] (Díaz et al. 1992); and Dp-1 [Pal] (Sheehan et al. 1997).

These enzymes demonstrated varied overall sequence homology to LytA itself. HBL and EJL are highly homologous to the autolysin throughout their entire sequence, as they share both LytA's N-terminal amidase domain and its C-terminal choline-binding domain. By contrast, CPL-1, CPL-7 and CPL-9 possess muramidase activity and are homologous to LytA only at the C-terminal binding end. Pal likewise demonstrates only C-terminal homology. Although both lysins possess alanine-amidase activity, Pal encodes a type 5 amidase domain while LytA encodes a type 2 amidase domain. Of all these pneumococcal enzymes, CPL-1 has received by the far most attention in subsequent enzymatic trials (as described in the preceding section).

As for the hybridization strategy itself, the technique is no longer commonly used for identifying lytic enzymes. A potential reason for this is

logistical: with more recent cloning strategies (described in the following sections), Southern blotting can be relatively cumbersome by comparison. There also exists the possibility of limiting the results due to the sequence of the original DNA probe. With this approach, it is only possible to identify genes with some homology to a prototype sequence. But if a phage is relatively novel – for instance, in terms of its host bacterium or viral morphotype – suitable homologues might not exist for its lysin. This is especially problematic considering that enzymes with highly novel sequences represent some of the most attractive targets for future discovery.

Nevertheless, if one has good reason to suspect that a desired lysin is similar in sequence to one already characterized, techniques based on nucleotide homology can still be quite effective. In this regard, several studies have utilized PCR-based approaches as more rapid alternatives to Southern blotting. For example, Morita et al. designed primers from the genomic regions surrounding the lysin of a *B. subtilis* phage, which they used to amplify a related enzyme for *B. amyloliquefaciens* (2001b). Romero et al. likewise synthesized various primers based on known LytA-like sequences when attempting to clone the lysins from two *S. mitis* phage (2004). These authors successfully identified a primer-pair that amplified a portion of both enzymes; these partial sequences were subsequently used to characterize the remainder of the genes by genomic primer walking (i.e. chain termination sequencing with the genome as the direct template). It should be mentioned that, although they have not been applied

specifically to lytic enzymes, several other techniques are available that could identify a complete lysin gene from only a partial sequence. These include inverse PCR (Ochman et al. 1988) and semi-random PCR (Hermann et al. 2000).

Phage lysins and functional screening. To avoid the possibility of sequence-based bias, it is ultimately necessary to identify lysin-encoding genes by the enzymatic activity of their translated proteins. This approach is the foundation of *functional genomic screening*, and it has become a common tool for lysin identification over the past decade. The experimental specifics of lysin-screening can differ slightly, and these variables are reviewed in the proceeding paragraphs. Overall, however, such methods represent variations on the same general theme: [I] phage genomic DNA is isolated and digested into fragments; [II] the fragments are ligated into an expression vector and transformed into a host organism; [III] the transformants are clonally propagated and exposed to an inducing agent to force transcription of the genomic inserts; and [IV] the clones are analyzed for the acquisition of a phenotype that indicates the presence of a lysin-encoding gene.

For the first three steps, the experimental considerations are fairly general in nature and unrelated to the activity of the targeted enzymes. The original source DNA can be derived from either lytic phage, isolated from the environment or purchased from commercial sources, or lysogenic prophage, induced from host bacteria with an appropriate stressing agent (e.g. mitomycin,

phosphomycin, or UV-treatment). For the fragmentation step, most lysin screens have utilized a standard *shotgun* approach to create a random array of genomic fragments. Here, the phage DNA is partially digested with restriction enzyme, usually one with a 4-bp consensus sequence. A final length distribution of 1.5 - 3 kb is ideal, as it represents 2 - 3 times the length of typical Gram-positive lysins.

While alternate methods have been used on occasion², the second step typically involves ligation of these fragments into an expression plasmid with transformation of an *E. coli* host. One should note that other vectors, in theory, could support longer DNA inserts (i.e. cosmids or bacterial artificial chromosomes). However, these systems suffer from the fact that they would rely upon native promoters for recombinant expression. Given the small size of Caudoviral genomes (several dozen to several hundred kilobases), plasmid-based screens are still capable of identifying lytic clones with high efficiency. In past studies, for example, hits have typically been observed at a frequency of 0.1-2% of total colonies (Schuch et al., 2008).

When choosing a particular plasmid for lysin screening, the same variables must be taken into consideration as during the recombinant expression of any protein. These include the type of promoter, induction conditions, codon usage, and the host *E. coli* strain. For a given lysin, the expression level and solubility can vary significantly from one system to the next, often in unpredictable ways. Nevertheless, for a detailed protocol that we have found generally reliable, the reader is referred to Schuch et al. (2009). The screen

outlined here utilizes an arabinose-inducible pBAD plasmid, which is attractive for its tight transcriptional control and the cost-effective nature of the inducing agent.

For the final steps of a lysin screen, induction and selection of positive clones, the following experimental manipulations are typically involved. Transformed *E. coli* are spread onto agar plates that lack inducing agent, allowing clones to proliferate without transcription of genomic inserts. These *master plates* are replicated onto *screening plates* whose agar has been supplemented with inducing agent. After propagating with forced transcription, the clones are exposed to chloroform vapor to permeabilize the *E. coli* and allow free diffusion of the expressed proteins. The clones are overlaid with a soft agar media containing Gram-positive cells. The plates are observed over time for clones over which there develops a zone of diminished bacterial density, indicating the presence of a lysin-encoding gene (see Figure 1.8). The corresponding clone on the master plate is subsequently identified and expanded for sequencing and large-scale expression. In early functional screens, the process occasionally differed in minor aspects (for instance, in the logistics of replica-plating or cell permeabilization). In current studies, however, this procedure has become the norm.

The main enduring variable in the process of lysin identification lies at the end of the procedure, as several options are available regarding the type and quantity of Gram-positive bacteria in the soft-agar overlay step.

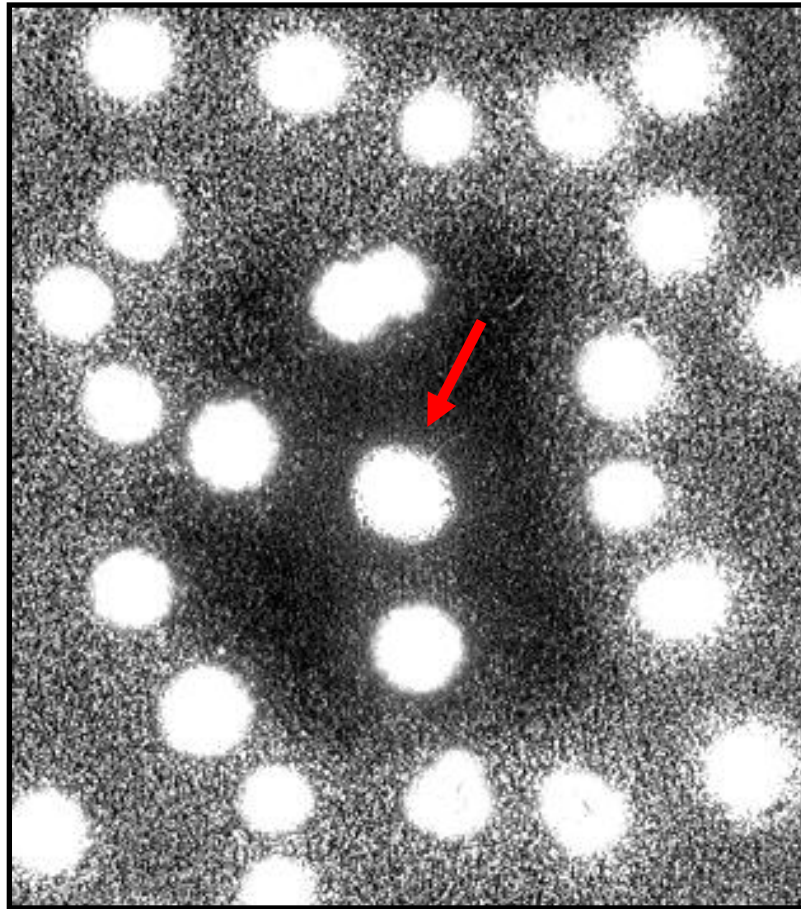


Figure 1.8 Lysin-Induced Clearing Zone

To identify a phage lytic enzyme in a functional screen, Gram-positive bacteria are overlaid on permeabilized *E. coli* clones expressing phage genomic inserts. The desired clone (indicated above with an arrow) is identified by the development of a surrounding halo of Gram-positive lysis. The particular example shown here involves *B. anthracis* (strain 222) cells and the PlyB lysin from the BcpI phage (Porter et al. 2006). One should note that the relative size and intensity of the halo can vary depending on the particular lysin and the overlaid bacterial species/strain.

Table 1.5 summarizes these techniques and the individual studies in which they have been employed over the years. In a few studies, permeabilized clones were overlaid with concentrated *Micrococcus* cells (Jayaswal et al. 1990; Bon et al. 1997). Decreased micrococcal turbidity is a classic method of quantifying the activity of eukaryotic peptidoglycan hydrolases (Shugar 1952). The majority of screens, however, have utilized the host bacteria of the phage whose genome was being screened. This makes intuitive sense considering the specificity that lytic enzymes demonstrate toward host organisms.

At the same time, several variations do exist as to how these cells are applied. The permeabilized clones can be overlaid with either [I] concentrated-viable bacteria, [II] concentrated-nonviable bacteria (typically autoclaved), or [III] dilute-viable bacteria. In the first two cases, plates are observed for clones around which bacterial density decreases, while, in the third, they are observed for clones around which bacteria fail to proliferate. Previous examples of each approach include, respectively: several *L. monocytogenes* lysins (Loessner et al. 1995); an enzyme targeting a novel *Staphylococcus* strain (Yokoi et al. 2005b); and the amidase of the γ diagnostic phage of *B. anthracis* (Schuch et al. 2002).

While there has never been a dedicated study comparing the relative efficacy of these three variations, each type of overlay has been utilized within the Rockefeller University Laboratory of Pathogenesis with general effectiveness. One notable benefit, however, of employing dilute-viable cells is its particularly high level of sensitivity.

Year	Screening Strategy	Target Bacteria	Reference
1990	Micrococcal overlay	<i>Staph. aureus</i>	Jayaswal et al.
1997	Micrococcal overlay	<i>Staph. aureus</i>	Bon et al.
1989	λ -phage expression; Target GPB: Conc. viable	<i>Lactococcus lactis</i>	Shearman et al.
1990	Target GPB: Conc. viable	<i>Lactobac. bulgaricus</i>	Boizet et al.
1992	Target GPB: Conc. viable	<i>Lactococcus lactis</i>	Platteeuw and de Vos
1993	Target GPB: Conc. viable	<i>Lactococcus lactis</i>	Ward et al.
1995	Target GPB: Conc. viable	<i>List. monocytogenes</i>	Loessner et al.
1997	Target GPB: Conc. viable	<i>Bacillus cereus</i>	Loessner et al.
1998	Target GPB: Conc. viable	<i>Staph. aureus</i>	Loessner et al.
1999	Target GPB: Conc. viable	<i>Staph. aureus</i>	Loessner et al.
2002	Target GPB: Conc. viable	<i>Bacillus anthracis</i>	Schuch et al.
2005	Target GPB: Conc. viable	<i>Strep. agalactiae</i>	Cheng et al.
2004	Target GPB: Conc. nonviable	<i>Lactobac. helveticus</i>	Deutsch et al.
2005	Target GPB: Conc. nonviable	<i>Lactobac. gasseri</i>	Yokoi et al. [a]
2005	Target GPB: Conc. nonviable	<i>Staph. warneri</i>	Yokoi et al. [b]
1986	Target GPB: Dilute. viable	<i>Lactobac. delbrueckii</i>	Trautwetter et al.
2004	Target GPB: Dilute. viable	<i>Enterococcus faecalis</i>	Yoon et al.
2006	Target GPB: Dilute. viable	<i>Bacillus anthracis</i>	Portet et al.
2008	Target GPB: Dilute. viable	<i>Bacillus anthracis</i>	Schmitz et al. [a]
2001	Target GPB: ?	<i>Lactobac. plantarum</i>	Yoon et al.
1995	λ -phage complementation	<i>Lactobac. gasseri</i>	Henrich et al.

Table 1.5 Lysins Cloned Through Functional Genomic Screening

Table 1.5, continued

Provided here is a complete list (as of 2009) of recombinantly-expressed phage lytic enzymes against Gram-positive bacteria for which the original means of identification was a functional genomic screen. They are organized by screening methodology, and include the corresponding target bacteria and year of identification. The methodologies include: overlay by micrococcal cells; overlay by concentrated, viable Gram-positive bacteria (GPB) of the target species; overlay by concentrated, nonviable GPB; and overlay by dilute, viable GPB. For two entries (Shearman et al. 1989; Henrich et al. 1995), a λ -phage screening strategy was used in place of a plasmid-based approach (see Endnotes for more information).

Since only enough lysin is required to prevent the growth of a small initial population, clearing zones are often evident even when an enzyme is not well expressed under the screening conditions. Using this method, clearing zones have been observed even in instances when a particular clone does not produce sufficient lysin for recombinant purification or detection by Coomassie-staining (unpublished observations).

Several other factors must likewise be considered when selecting a particular overlay technique. For instance, certain bacteria can react poorly to heat-killing (e.g. with aggregation or lysis), rendering the concentrated-nonviable approach ineffective. Other bacterial organisms can interact non-specifically with the *E. coli* library clones when proliferating in soft agar, leading to widespread *pseudo*-clearing clones for the dilute-viable method. While it is difficult to predict what approach is ideal for a given species/strain, this should not prevent one from successfully cloning a lytic enzyme. All three techniques are ultimately straightforward, and (as replica plating is not a time-consuming step) one can readily conduct multiple types of overlays from each master plate.

Recent additions to functional screening. Despite the general success of the proceeding techniques, they are not the only functional methods available for cloning phage lysins. Indeed, several additional strategies have been devised recently that complement or expedite these approaches. One such example is *holin-based* screening for lytic enzymes. As described above, it is the combined

action of lysin and holin that leads to host-cell lysis and the release of progeny viral particles. The necessity of both proteins during phage infection explains why lytic enzymes can be overexpressed recombinantly in *E. coli*: even if a given lysin possesses activity against the *E. coli* peptidoglycan, it cannot exert a toxic effect as long as it is sequestered in the cytoplasm. By contrast, co-expression of both lysin and holin can lead to marked toxicity, as holins can integrate non-specifically into cytoplasmic membranes (including those of an *E. coli* host).

Due to the genomic adjacency of holins and lysins, lysin-containing fragments in shotgun libraries commonly encode holins as well. This creates the potential for selective toxicity of exactly the clones one hopes to identify. When a lysin-encoding clone is identified in an enzyme-based screen, it is generally one in which either [I] the holin and lysin happen not to be encoded adjacently in the particular genome, [II] the holin (fortuitously) is not sufficiently expressed, or [III] only a limited amount of genomic DNA surrounds the lysin, excluding the complete holin. Due to the small size of phage genomes and the resultant high proportion of lysin-encoding clones, the issue of holin toxicity has not proven a tremendous obstacle in past screens. Nevertheless, several studies have looked to avoid the situation altogether by *selecting for* holin-encoding clones. By targeting the holin genes, it is possible to identify adjacent lysins without actually observing lysin activity.

In this regard, Delisle et al. utilized a *plasmid release* protocol to identify a phage lytic enzyme for the dental pathogen *Actinomyces naeslundii* (2006). For

this study, mixed *E. coli* transformants were grown in a single liquid culture. Following induced expression, holin-encoding cells would undergo lysis, releasing their plasmid into the culture media. The plasmids were then purified and used to retransform a new set of competent *E. coli*. Through several rounds of this procedure, the authors were able to enrich for a set of clones encoding the holin-lysin region, ultimately allowing them to sub-clone and express the lytic enzyme.

One additional technology with potential relevance to lysin identification is that of *whole-genome amplification*. PCR-based methods have recently come into prominence that allow for general amplification of viral (along with bacterial or eukaryotic) DNA. These include the use of linker-based amplification (see Chapter 3), as well as high-diversity primers in combination with the ultra-processive $\phi 29$ polymerase. For a review that discusses genome amplification specifically in the context of viruses, the reader is referred to Delwart et al. (2007).

The significance of these techniques lies in the access they provide to exceedingly small biological samples, as once-undetectable genetic material is now available for analysis. From the perspective of lysin screening, whole-genome amplification can significantly expedite the preparation time for library construction. Purification of phage DNA is a relatively time-intensive process compared to that of cellular organisms. Depending on the phage and the growth properties of the host, obtaining a high-enough viral titer for microgram quantities of DNA (typically required for shotgun cloning) can represent several

days to several weeks of work. By amplifying a small initial quantity of phage DNA, one can side-step this issue and obtain essentially a limitless supply of genomic material.

Lytic enzymes and whole-genome sequencing. Despite this growing array of techniques, functional screening still represents only one side of the ongoing effort to clone lytic enzymes. In fact, only ~50% of phage lysins that have been recombinantly expressed to date were first identified in this manner. Many others, at the same time, have been the result of genomic sequencing and nucleotide homology analysis. The field of whole-genome sequencing has expanded rapidly in recent years (for viruses, bacteria, and eukaryotes), driven in large part by the development of high-throughput technologies like 454 pyrosequencing and Illumina/Solexa sequencing-by-synthesis (Strausberg et al. 2008). Publications documenting complete phage genomes have become commonplace, and a growing number of studies now focus on the genomes of numerous, interrelated phage. To date, 500+ complete phage genomes are present within the NCBI database (not counting prophage), and this number is expected to increase substantially in the near future (Hatfull 2008).

The impact of genomic sequencing on lysin research is simple: whenever the genome of a new phage is reported, another lysin gene is uncovered. Their enzymatic motifs and overall structure are sufficiently conserved that standard algorithms (e.g. Blast, Pfam) generally make it possible for one to recognize a

lytic enzyme from a sequence alone. Overall, identifying these enzymes through nucleotide homology is not a new concept; it was mentioned before that two of the earliest-known lysin genes (from the *Salmonella* P22 and *B. subtilis* ϕ 29 phage) were discovered based on their similarity to the *E. coli* T4 lysin. What distinguishes today's bioinformatic analyses is the large (and ever growing) size of public databases, which allow one not only to locate a lysin gene itself, but often predict its enzymatic mechanism, domain phylogeny, or even catalytic residues³. These genes can then be PCR-cloned and tested for activity.

Table 1.6 provides a chronological list of recombinantly-expressed lysins whose genes were initially identified through DNA sequence analysis. With a single exception (O'Flaherty 2005), the genes were PCR-cloned directly from genomic DNA. In the one instance, a staphylococcal lysin was interrupted by an intronic sequence, requiring the investigators to extract mRNA from infected bacteria and prepare cDNA as an amplification template. Aside from the increasing number of lysins in recent years, Table 1.6 reveals several important trends. For instance, whereas earlier lysins were often discovered within partial genomic sequences (i.e. a fortuitous sampling of restriction fragments), recently-cloned enzymes have generally been identified from complete phage genomes. Presumably, this reflects both gains in sequencing technology, as well as improvements in functional techniques that make it possible to clone a lysin without resorting to partial sequencing.

Year	Sequecning Description	Target Bacteria	Reference
1986	Partial phage genome	<i>Bacillus subtilis</i>	Garvey et al.
1994	Partial phage genome	<i>Lactococcus lactis</i>	Arendt et al.
1994	Partial phage genome	<i>Lactococcus lactis</i>	Birkeland
1996	Partial phage genome	<i>Lactobacillus sp.</i>	Oki et al.
1997	Complete phage genome	<i>Lactococcus lactis</i>	Chandry et al.
1999	Partial phage genome	<i>Oenococcus oeni</i>	Parreira et al.
2000	Complete phage genome	<i>Lactobacillus casei</i>	Kashige et al.
2002	Complete phage genome	<i>Clostridium perfringens</i>	Zimmer ett al.
2004	Partial phage genome	<i>Strep. agalactiae</i>	Pritchard et al.
2004	Complete phage genome	<i>Bacillus thuringiensis</i>	Verheust et al.
2005	Complete phage genome	<i>Staph. aureus</i>	Takác et al.
2005	Prophage/host genome	<i>Bacillus anthracis</i>	Low et al.
2006	Prophage/host genome	<i>Bacillus anthracis</i>	Yoong et al.
2006	Complete phage genome	<i>List. monocytogenes</i>	Korndörfer et al.
2007	Complete phage genome	<i>Staph. aureus</i>	Rashel et al.
2007	Complete phage genome	<i>Staph. aureus</i>	Sass and Bierbaum
2007	Prophage/host genome	<i>Strep. agalactiae</i>	Pritchard et al.
2008	Partial phage genome	<i>Streptococcus uberis</i>	Celia et al.
2008	Complete phage genome	<i>Clostridium difficile</i>	Mayer et al.

Table 1.6 Lysins Cloned Through Genomic Sequencing

Table 1.6, continued

Organized by date, this table lists recombinantly-expressed lysins against Gram-positive bacteria which were originally identified through DNA sequencing and bioinformatic comparison to known proteins. Three categories are present within the table: enzymes identified through partial sequencing of a phage genome, those identified through complete sequencing of a phage genome, and prophage lysins identified through complete sequencing of the bacterial host genome.

One should likewise note that four lytic enzymes were amplified not from the genomes of individual phage, but rather from the genomes of their bacterial hosts. These include PlyL (Low et al. 2005) and PlyPH (Yoong et al. 2006) from the Ames strain of *B. anthracis*, along with the LambdaSa1 and LambdaSa2 lysins from *S. agalactiae* 2603 V/R (Pritchard et al. 2007).

Bacterial genomes in themselves represent attractive sources of lytic enzymes in the form of integrated viral DNA (this, in fact, is the focus of the next chapter of this thesis). As the genomes of additional bacterial species/strains are sequenced, the number of prophage lysins that could serve as potential enzybiotics will only grow further.

Sequence-based versus functional: pros and cons. At this point, it is worthwhile to consider the relative advantages and disadvantages of identifying lysins through functional screening versus genomic sequencing and PCR-cloning. One of the most attractive features of the latter approach is its straightforward nature. If one has isolated a phage and wishes to identify its lysin, genomic sequencing is a method for which success is highly likely. The primary obstacle here is not cloning the gene itself, but rather identifying a suitable vector for active expression. This challenge is true for all recombinant proteins, however, and is equally applicable to functional lysin screens.

A sequence-based approach likewise allows one to mine genomic databases for lysins that have not yet been studied in detail. In the process, one

can compare their putative biochemical properties and overall uniqueness through bioinformatic analysis (again, see Chapter 2). For example, a lysin with a catalytic domain that is unique for a particular host bacteria is far more attractive than a lysin that is highly similar to a previously-expressed enzyme. Overall, while it is not always the case that a desired sequence is present within public databases, it is always advantageous to check, as this represents easily the most rapid singular approach to expressing a novel lysin.

Of course, identifying lysins through nucleotide homology does not always present the best option—quite frequently, the reason is purely logistical. Despite ongoing technological improvements, genomic sequencing does not yet represent an insignificant investment of time and resources, and high-throughput access is far from universal. For a majority of laboratories, functional screening is still the most cost-effective and rapid method for identifying lytic enzymes from individual phage genomes. With sequence analysis alone, moreover, there exists a small chance of overlooking a lytic enzyme due to an atypical sequence. While they are uncommon, lysins are occasionally identified whose genes deviate significantly from the norm.

The most prominent example is the streptococcal C₁ lysin (PlyC), described previously in this section for its early role in lysin purification. The C₁ genomic sequence, reported in 2003, revealed an abnormal lytic region in which a putative holin (ORF8) lied adjacent to several ambiguous open reading frames (Nelson et al.). While the theoretical translation of ORF9 corresponded to a

partial Edman sequence of the purified enzyme, none of ORFs demonstrated typical lysin features or possessed a molecular mass corresponding to PlyC. Only when the authors functionally screened the C₁ genome and dissected the ORFs of the lytic clone were they able to determine that PlyC represents a unique multimeric phage lysin (resulting from ORFs 9 and 11) (Nelson et al. 2006). In fact, PlyC remains without any homologues among sequenced phage, and it is a remarkable coincidence that one of the first lysins ever purified is so atypical.

Just recently, another highly novel lysin was reported for the ϕ IN93 phage of the extremophilic bacterium *Thermus aquaticus* (Matsushita and Yanase, 2008). Employing a more classical approach, the authors first purified the enzyme from infected culture supernatant and subjected it to Edman sequencing. The N-terminal amino acids were then cross-referenced to the sequenced genome to locate the lysin-encoding ORF, which until that point had remained unidentified due its lack of recognized domains or homology to other proteins. While the authors did not utilize a functional screen to identify the enzyme, per se, this case again demonstrates how genomic sequences alone can occasionally be insufficient for lysin identification. Overall, the value of this particular enzyme lies not in its anti-infective potential—*T. aquaticus* is neither Gram-positive nor pathogenic—but in its ability to retain activity at high temperatures, an attractive industrial feature.

OBJECTIVES AND OVERVIEW OF THE THESIS PROJECT

The present body of research. Between sequence-based and functional cloning approaches, the number of phage lysins that have been subject to recombinant analysis has steadily increased in recent years. This trend should only continue, especially if pre-clinical work these enzymes successfully progresses to human trials. At the same time – and despite their widespread use – the techniques described above are no longer the only (or even, necessarily, the most efficient) avenues for identifying new enzymatic candidates. My doctoral research at Rockefeller has addressed this very issue. Broadly stated, the work presented in this thesis explores new approaches for cloning novel phage lytic enzymes. The underlying motivation for this research is straightforward: by increasing the number characterized lysins, one will only hasten the development of these molecules into effective biomedical tools.

The work presented here does not focus on any particular bacterial pathogen or phage lysin. Rather, the focus is on the *process*, and how researchers can move beyond traditional approaches in the lysin field – this includes a combination of [1] strategy and [2] methodology. *Strategy* refers to potential genetic sources of lysins and whether it is possible to look for these enzymes elsewhere than individual phage genomes. *Methodology* refers to experimental techniques and the creation of new protocols that will make cloning lytic enzymes more efficient and expansive. This thesis is divided into four

individual sections (Chapters 2 - 5), each of which represents an independent (yet interrelated) investigation into strategy and/or methodology.

Chapter 2 focuses on bacterial genomic sequencing and its impact on the lysin field. The recent explosion of sequenced bacterial genomes – which naturally include integrated prophages – has revealed numerous ORFs that can be assigned putative lysin functionality based on homology analysis. In the past several years, in fact, the number of theoretical prophage lysins in computational databases has come to dwarf the number of lysins whose activity has been studied experimentally. Identifying prophage lysins via bacterial genomic sequencing is no different than the phage-sequencing strategy discussed above. The difference now, however, is that high-throughput sequencing technology has made entire bacterial genomes (several Mb) as accessible as their smaller phage counterparts (tens of kb). As a result, numerous prophage enzymes can be rationally compared and rapidly cloned without ever having to work with phage itself.

Chapter 2 takes advantage of this information for the pathogenic anaerobe *Clostridium perfringens*. The recently-sequenced genomes of 9 strains of *C. perfringens* were computationally mined for proviral lytic enzymes, and a comprehensive list of lysins was constructed and compared. From this list, a particular enzyme was chosen for cloning, expression, and *in vitro* analysis based on its dissimilarity with the only other previously-studied *C. perfringens* lysin. Various biochemical properties of this lysin were determined, as well as its

activity profile against a panel of *C. perfringens*, non-perfringens clostridia, and other Gram-positive bacterial species.

By contrast, Chapter 3 of this thesis changes course; it is the first of two chapters to explore the rapidly expanding field of metagenomics as it relates to lytic enzymes. *Metagenomics* refers to the direct expansion of DNA/RNA from environmental samples without first isolating and cultivating any of the individual component microbes (the reader is referred to the individual chapter for a more detailed review). While the initial focus of metagenomics research was on environmental bacteria, the field has since expanded to include investigations of uncultured viruses. In these studies, viral particles (consisting primarily of phage) are separated in bulk from environmental bacteria and other debris prior to DNA extraction and analysis.

To date, viral metagenomics has been primarily sequence-based in nature; here, uncultured phage sequences are compared to known genetic databases to address broad questions of phage biology and ecology. On the other hand, functional screening of viral metagenomes could provide direct access to targeted classes of recombinant molecules. Phage lysins, of course, represent an intriguing example. Nevertheless, functional metagenomic screens for lytic enzymes would present several distinct challenges. The goal of Chapters 3 and 4 is to address these challenges and make metagenomic lysin screens a reality.

Chapter 3 focuses on the issue of *DNA quantity*, as environmental phage samples do not typically provide a sufficient quantity of genetic material for

functional screening. A technique known as E-LASL (for expressed linker amplified shotgun library) was developed that combines linker amplification of fragmented DNA with commercial topoisomerase cloning. This creates a rapid protocol for constructing (in an *E. coli* host) expressible plasmid-based libraries from nanogram quantities of DNA (a typical yield for viral metagenomic extractions). To validate the technique, various genomic libraries were constructed from *Bacillus* phage, and six lytic enzymes for *B. anthracis* were successfully cloned. Additionally, a bacterial metagenomic E-LASL was constructed from the gut contents of an earthworm, and an active virulence factor was cloned in a subsequent hemolysis screen.

With this method for library construction in place, Chapter 4 develops a procedure for selecting *E. coli* transformants that encode and actively express metagenomic lysins. The method circumvents additional challenges that face metagenomic lysin screening, ones involving holin-induced toxicity of lysin-encoding clones and the choice of an appropriate selection agent (see Chapter 4 for additional information). The technique utilizes two sequential steps: a toxicity screen is first employed in which colonies are selected for holin-induced lysis following induced expression. In a secondary step, then, the initial hits are overlaid with autoclaved Gram-negative bacteria (specifically *Pseudomonas aeruginosa*) to assay directly for the production of lytic enzyme. As proof-of-principle, a viral metagenomic library was constructed from mixed animal feces

and subjected to the screen. It was validated by the successful cloning of 26 lysins of various enzymatic activities and diverse modular architectures.

Finally, Chapter 5 takes the techniques developed in the preceding chapter and reapplies them back to a genomic DNA library. While the two-step screen was designed with metagenomes in mind, it is still applicable to more traditional sources of lysins. In this respect, the genome of a clinical strain of *Streptococcus suis* (a veterinary and opportunistic human pathogen) was subject to the screen, and a single prophage lysin was successfully cloned (PlySs1). An additional prophage lysin (PlySs2) was likewise identified and cloned via inspection of various published *S. suis* genomes (i.e. in the same manner as PlyCM). The PlySs1 lysin was subsequently purified and characterized so that it could be included in an ongoing veterinary trial involving experimentally-infected pigs.

Dissertation format. At this point, before going on to present the research itself, it is necessary to provide some brief logistical information on the format of this thesis. Although the work presented here is designed to be read as a single coherent document from beginning to end, Chapters 2 - 5 each represent a distinct sub-study in its own right. For every individual chapter, the work described has been drafted into a separate manuscript for publication in a peer-reviewed journal. One of the studies (corresponding to Chapters 3) has already been accepted (Schmitz et al. 2008); the work in Chapter 4 has been submitted

and is currently under review (Schmitz et al. 2010b); and the work in Chapters 2 and 5 will be submitted shortly. Moreover, it should be noted that portions of this Introduction and have appeared in similar form in a recently published text (Schmitz et al. 2010a). As a result of this breakdown, Chapters 2 -5 have been formatted so that (if desired) each can be read individually without loss of comprehension.

The layout of Chapter 2 - 5 is each essentially the same. They begin with an *Introduction* that provides additional background information on the particular research-related aspects of the chapter (for instance, this is where the reader would find an overview of *C. perfringens*, *S. suis*, or the field of metagenomics). The *Materials and Methods* section details the experimental protocols employed in the chapter. The data itself is reported in *Results*, and the *Discussion* analyzes the significance of that data in light of the current state of the field. Several chapters include endnotes that provide additional commentary on particular points in the text. Chapter 6 reflects upon the entire body of work and comments on future avenues for lysin discovery (as well as the discovery of enzybiotics other than phage lysins). A single *References* section is included at the end of the document, along with an appendix that provides sequence information (DNA and amino acid) for all genes cloned here.

From reading this thesis, one should gain an appreciation of just how rapidly the study of phage lysins and other enzybiotics has accelerated in recent years. Considering the unrelenting emergence of antibiotic resistance, this work

should continue unabated into the foreseeable future. Its eventual goal, of course, is to develop these agents to the point where they can have a positive impact on human health. The process will take time, effort, good science, and a little luck. But, fundamentally, it is still dependent on the efficient and intelligent identification of candidate enzymes. It is my hope that the work presented here will contribute to that process, and (at least in a small way) will help bring the term *enzybiotic* from a promising idea to a proven reality.

ENDNOTES

1. While it is true that several reports have documented the use of lytic enzymes against viable Gram-negative organisms (Alakomi et al. 2000; Morita et al. 2001a; Kim 2004; Briers et al. 2008), this activity was either quantitatively weak or required the addition of a general membrane-disrupting agent. One should note, however, that co-treatment of Gram-negative bacteria with a membrane-permeabilizing agent and a peptidoglycan hydrolase is already a common practice: it is the basis of commercial genomic extraction techniques and has various shortcomings from a drug-development perspective.

2. In the body of the text, we discuss the prevalent strategy for fragmenting and transforming phage genomic DNA for lysin screening (i.e. shotgun cloning into plasmid vectors). Although no longer generally employed, alternative screening techniques were occasionally used in early studies. For example, when cloning the lysin of a lactococcal phage, Shearman et al. utilized a λ -phage system that

relied upon infected *E. coli* for recombinant lysin expression (1989). Henrich et al. also employed λ -phage when cloning a *Lactobacillus* lysin. This study involved a defective λ -strain that required ligation of a complementary lysis cassette to form plaques on *E. coli* (1995).

Furthermore, one should note that *E. coli* is not the only host that, in theory, could be used for lysin screening. While it has been utilized exclusively to date, there is nothing about this species that makes it particularly well-suited for identifying lytic clones (other than the commercial prevalence of *E. coli* expression systems). Lysins have been expressed recombinantly from other bacterial species for purification purposes—and, in certain cases, these alternate species offered superior expression to *E. coli* (Yoong et al. 2004).

3. It is not always possible, however, to determine such information from a nucleotide sequence alone. While analysis of a lysin's sequence typically reveals a conserved domain—which, in turn, implies an enzymatic mechanism—lytic enzymes with ambiguous domains are still reported (see the section “Sequence-Based versus Functional: Pros and Cons”). Moreover, computational analysis alone can occasionally make mistakes regarding a lysin's biochemical properties. For example, one recent study describes two *S. agalactiae* lysins that were predicted to encode alanine-amidase enzymatic domains. Following expression and functional testing, they were instead determined to possess endopeptidase and glucosaminidase activity (Pritchard et al. 2007).

It is important to note that, for most phage lysins that have been studied recombinantly, the reported enzymatic activity was based on computational predictions rather than experimental observations. While these predictions should be correct in a majority of circumstances, it still leaves open the possibility for occasional errors.

CHAPTER 2

Lysins without Phage: Bacterial Genomes and Proviral Enzymes

INTRODUCTION

Over the past decade, phage lytic enzymes (a.k.a. lysins or endolysins) have been the focus of a great deal of applied microbiological research (Villa and Veiga-Crespo 2010; Fischetti 2008; Borysowski et al. 2005; Loessner 2005). Encoded by virtually all double-stranded DNA phage, lysins are expressed late in the cycle of phage infection. They are responsible for hydrolyzing the peptidoglycan of the host bacterium and – along with membrane permeablizing

proteins known as holins – they allow progeny viral particles to escape the cytoplasm and reinfect new cells (Gasset 2010).

Biotechnological interest in these proteins stems from their ability to lyse Gram-positive species when applied exogenously, as the peptidoglycan of these organisms is directly accessible from the extracellular space. Phage lysins have been proposed as novel *enzymibiotic* (enzyme-antibiotic) agents against human and veterinary pathogens, in addition to other potential uses in the areas of food (Deutsch et al. 2004), agricultural (Kim et al. 2004), and industrial (Ye and Zhang 2008) science. Numerous individual lysins have been recombinantly expressed and characterized to date, including successful *in vivo* trials involving various animal models of colonization and infection (Daniel et al. 2010; Grandgirard et al. 2008; Nelson et al. 2001). Their appeal lies in both their potency and their specificity toward individual bacterial species, typically the host organism of the encoding phage.

Both enzymatically and architecturally, the lysins are a highly diverse group of proteins. Gram-positive lysins are classically modular, 250-400 amino acids in length, with an N-terminal enzymatic domain and a C-terminal binding domain (Fischetti 2008). The enzymatic domain itself can target numerous bonds within the peptidoglycan macromolecule. These include: muramidase (glycan backbone, Porter et al. 2007), glucosaminidase (glycan backbone, Pritchard et al. 2007), alanine-amidase (stem position α ; Schuch et al. 2002), alanoyl-glutamate endopeptidase (stem position β ; Korndörfer et al. 2008), glutaminyl-lysine

endopeptidase (stem position γ ; Pritchard et al. 2007), and cross-bridge endopeptidase (Baker et al. 2006). The C-terminal domain likewise binds one of various epitopes within the target cell envelope (e.g. surface carbohydrate, choline, or peptidoglycan itself), and it is largely responsible for the specificity of a lysin toward particular bacteria (Diaz et al. 1991). Gram-negative lysins, by contrast, are typically smaller and are comprised of an enzymatic domain alone. Occasionally, lysins are identified that do not conform to standard architectures, including Gram-positive lysins with multiple enzymatic domains (Baker et al. 2006), Gram-negative lysins with an N-terminal binding domain (Briers et al. 2007), and a multimeric Gram-positive lysin (Nelson et al. 2006).

Most lysins characterized to date have been cloned through traditional phage genomic techniques (Schmitz et al. 2010a). Here, a phage is first isolated through environmental sampling or prophage induction, and its genomic DNA is extracted following laboratory culture. At this point, either the phage genome may be sequenced and the lysin-encoding ORF identified through sequence homology with known enzymes, or the DNA can be fragmented and subjected to a recombinant functional screen. In the latter case, the lysin-encoding fragment is identified by its ability to confer a bacteriolytic phenotype on a host clone. While generally successful, these methods are still rather time-consuming and wholly dependant on the successful isolation/propagation of the initial phage. Moreover, they cannot predict *a priori* how novel the identified lysin will be

relative to previously-characterized enzymes, nor can they assess total lysin diversity for a particular bacterial host.

In this regard, the rapidly expanding number of published bacterial genome sequences (1000+ at the time of writing) could prove quite valuable. Representative genomes have now been sequenced for hundreds of bacterial species; for many medically/technologically-relevant bacteria, numerous individual strains have been sequenced (www.ncbi.nlm.nih.gov/genomes/lproks.cgi). If a particular strain is lysogenized with prophage, its genome becomes an easy source of lytic enzymes, which can be rapidly cloned through PCR. In the case of polylysogenized strains or species with multiple sequenced strains, one can systematically compare the different lysins and their predicted properties (particularly enzymatic and binding domains) before choosing which one(s) to express. In the current study, we wished to do precisely this – computationally analyze all known prophage lysins for a particular Gram-positive pathogen, and use this information as a guide for subsequent recombinant analysis.

For this work, we chose to focus on the spore-forming, anaerobic rod *Clostridium perfringens*, one of the most frequently encountered species of clostridia in clinical and environmental laboratories. *C. perfringens* is a common component of healthy human feces, and it is found in the digestive tracts of many other vertebrate and non-vertebrate animals, as well as environmental soil samples (Matches et al. 1974). Despite its ubiquity, *C. perfringens* can

nevertheless be the etiological agent of various pathologies. The species is taxonomically subdivided into five toxinotypes (A-E) based on the combinatorial presence of at least 15 exotoxins gene, including *alpha*, *beta*, *beta-2*, *epsilon*, *iota*, and enterotoxin/CPE (Smedly et al. 2004). These toxins may be encoded on mobile genetic elements and they help determine a strain's pathogenic potential.

The most commonly encountered toxinotype of the healthy human bowel (enterotoxin-negative type A, Carman et al. 2008) can induce myonecrosis in the context of wound infection (Bryant and Stevens 1997). Enterotoxin-positive type A strains are prevalent agents of food-borne illness, causing up to 600,000 cases of self-limited enteritis annually in the United States (McClane 1997). And while cases are uncommon, type C strains are the agents of the potentially-fatal gut condition *enteritis necroticans* (Lawrence 1997). In addition to these human illnesses, all 5 toxinotypes of *C. perfringens* have been implicated in various maladies of poultry and livestock (van Immerseel et al. 2009; Uzal and Songer 2008).

To date, a single phage lysin for *C. perfringens* has been subject to recombinant expression and analysis. Zimmer et al. (2002a) induced a temperate siphovirus (ϕ 3626) from toxinotype B strain ATCC 3626 and sequenced the phage's 33.5 kB genome. The ORF encoding the lytic enzyme (Ply3626) was subsequently identified by sequence homology and cloned into an inducible *E. coli* expression plasmid (Zimmer et al. 2002b). The authors demonstrated that crude extracts of the induced *E. coli* clone were capable of lysing buffered

suspensions of viable *C. perfringens*. Although the degree of lysis varied from strain to strain, every isolate of *C. perfringens* tested (48/48) demonstrated susceptibility to Ply3626, even ones that were not susceptible to infection by the phage itself. In comparison, non-perfringens clostridia and other Gram-positive genera demonstrated virtually no susceptibility to the lysin. It has subsequently been suggested that Ply3626 could be used as a topical agent or food additive to prevent associated illnesses (Jay et al. 2005; Zimmer et al. 2008).

Since the above work on Ply3626, the genomes of 9 different strains of *C. perfringens* have been sequenced, all of which contain identifiable prophage regions (Shimizu et al. 2002; Myers et al. 2006; gsc.jcvi.org/projects/msc/clostridium/). In the present study, we systematically examined the prophage lysins in these sequences, comparing them based on sequence homology and domain composition. In the process, we identified 14 lysin genes and 31 lysin-like genes (the latter being defined as having high homology to the lysins throughout their entire sequence, lacking a signal peptide, but not being encoded in a recognizable prophage region). These proteins could be categorized into three enzymatic classes: type 2 alanine-amidase, type 3 alanine-amidase, and muramidase (glycosyl hydrolase, type 25).

Of the identified genes, a muramidase lysin from strain ATCC 13124 was selected for cloning and expression based on its divergent catalytic mechanism from Ply3626. Following purification, the enzyme (termed PlyCM, for clostridial muramidase) was subject to a variety of *in vitro* tests to determine its potency,

specificity, and biochemical properties. Overall, this work adds another phage lysin to the growing list of candidate enzybiotics, demonstrating an important role for comparative genomics in the development of these proteins.

MATERIALS AND METHODS

Computational identification/analysis of lytic enzymes. The genomes of the following sequenced strains of *C. perfringens* were included in this study: ATCC 13124 (toxino type A, NCBI genome project #304); SM101 (A, #2521); 13 (A, #79); F4969 (A, #20031); NCTC 8239 (A, #20033); ATCC 3626 (B, #20027); JGS1495 (C, #20025); JGS1721 (D, #28587); JGS1987 (E, #20029). The first step in compiling a comprehensive list of prophage lysins was to identify obvious proviral regions within each genome. This was accomplished through a combination of manual inspection of the annotated ORFs and use of the prophage prediction algorithm Prophinder (Lima-Mendez et al. 2008).

The lysin-encoding ORFs within each prophage regions were next selected. In many cases, these ORFs were already designated explicitly (e.g. “endolysin”) by the database annotation. For other prophage, the existing annotations were insufficient – in these cases, we performed Pfam domain analysis on the individual proteins in the region (Finn et al. 2009; pfam.sanger.ac.uk/). A set of several criteria (described in detail in the Results section) were applied to the Pfam predictions in order to designate a lysin for each case, and a preliminary list of probable lysin-encoding ORFs was compiled.

To ensure that no enzymes were overlooked in the process (for instance, ones encoded within short prophage remnants) the list was subjected to iterative Blast analysis (Altschul et al. 1990; blast.ncbi.nlm.nih.gov/Blast.cgi). The protein sequences on the initial list were BlastP-queried against the nine sequenced *C. perfringens* genomes. Any newly-identified homologs were added to the original list, and the process was repeated until no new ORFs were revealed. The aforementioned criteria were then applied to the expanded list to eliminate any ORFs that were included erroneously.

The sequences comprising this final list were phylogenetically compared to one another through the Phylip v3.67 package (Felsenstein 1989; evolution.gs.washington.edu/phylip.html). The translated protein sequences were subject to multiple sequence alignment through the *ClustalX* algorithm, followed by 1000 rounds of boot-strapping with *seqboot*. These alignments were analyzed in turn by the *protdist* (distance matrix generating) and *fitch* (tree generating) algorithms with default settings, and an unrooted consensus tree was generated with *consense*. The enzymes were compared with other known proteins with the BlastP algorithm (blast.ncbi.nlm.nih.gov/Blast.cgi), using default search settings and the non-redundant protein database (nr). Signal peptide predictions were made via the SignalP v3.0 server (Bendtsen et al. 2004; www.cbs.dtu.dk/services/SignalP/), employing both hidden Markov and neural network methods and the *Gram-positive bacteria* organism group.

Expression and purification of PlyCM. Extracted genomic DNA from *C. perfringens* ATCC 13124 was subjected to taq-based PCR amplification with the following PlyCM-targeted primers: ACCATGGAAAGTAGAAACAATAATAATTAAAAGG (fwd) and GTCAGATATTACTCTAACTAACCTTAAAA (rev). The underlined G in the forward primer was intentionally altered from the wild-type C at that position (a Q2E mutation) in order to introduce an NcoI restriction site that overlaps with the start codon. The PCR amplicon was topoisomerase cloned into pBad-TOPO (Invitrogen), an arabinose-inducible *E. coli* expression plasmid. This construct was subsequently purified, NcoI-digested, and re-circularized to eliminate a plasmid-encoded N-terminal leader sequence and ensure translation from the native start codon. This final construct was maintained and expressed in TOP10 *E. coli*.

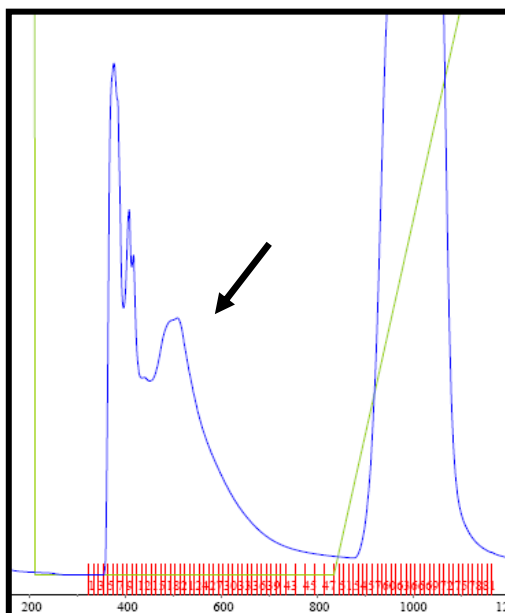
To express PlyCM, the clone was grown in LB broth to $OD_{600} \approx 0.5$ and induced with 0.2% L-(+)-arabinose. Following induction, the culture was shaken rigorously overnight at 30°C; this specific temperature was important, as inclusion bodies preferentially formed at 37°C. The expressing cells were pelleted, resuspended in 15 mM phosphate buffer pH 7.4, and lysed by three passages through an EmulsiFlex C-5 homogenizer. Cellular debris was removed by ultracentrifugation (1 hr, 35,000 X g), and the supernatant was $(NH_4)_2SO_4$ -precipitated to 40% saturation (226 g/L). The precipitated protein (which included the predominant amount of PlyCM) was pelleted, resolubilized in 15

mM phosphate buffer pH 7.4 (buffer A), and dialyzed against this buffer overnight.

The protein solution was next passed through a DEAE anion-exchange column equilibrated against buffer A (fast flow resin, General Electric). Based on PlyCM's theoretical isoelectric point of 5.1, one would predict it to bind to DEAE at a pH of 7.4. Nevertheless, the protein demonstrated an unusual chromatographic response: it neither bound the column in earnest nor flowed directly through it. Rather, there was a transient interaction in which PlyCM would initially bind the resin, but then slowly elute over 5+ column volumes as the column was washed with buffer A (see Figure 2.1). Although atypical, this property provided a fortuitous purification step, as PlyCM could be separated from the proteins in both the flow-through and tightly-bound fractions. The PlyCM-containing "slow wash fraction" was immediately passed through a ceramic hydroxyapatite column (Macro-Prep type II, 40 μ m, Bio-Rad) equilibrated against the same buffer A. The lysin demonstrated non-transient binding to this resin, and it was subsequently concentrated/purified through a 20 column-volume elution with buffer B (500 mM phosphate buffer pH 7.4), eluting at ~80 mM.

In the hours following purification, we observed that PlyCM would undergo irreversible precipitation in simple phosphate buffer. Addition of L-arginine to the lysin solution mitigated the precipitation, as has been reported for other recombinant proteins (Hamada et al. 2009).

DEAE



CHT

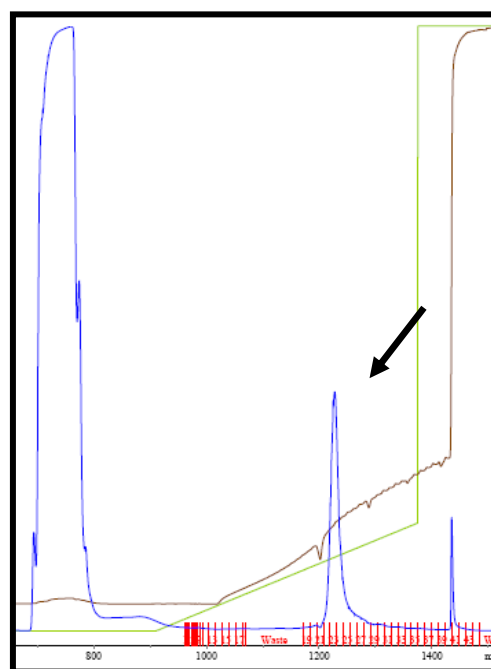


Figure 2.1 PlyCM Purification

PlyCM was purified through a combination of DEAE and CHT chromatography, as described in the *Materials and Methods* section.

Chromatograms for each successive step are presented here. In both cases, the peak corresponding to PlyCM is denoted with an arrow.

As a result, 100 mM L-arginine (pH = 7.4) was included in the PlyCM stock-preparation prior to freezing, lyophilization, and storage at -20°C. Overall, the above protocol generated ~10 mg of purified PlyCM per liter of original *E. coli* culture.

***In vitro* analysis of PlyCM.** The activity of PlyCM was examined predominately through optical density analysis. The various strains/species of bacteria were grown on agar plates at 37°C. Schaedler agar with vitamin K₁ and 5% sheep's blood (Becton Dickinson) was employed for clostridial strains (perfringens and non-perfringens), and the bacteria were grown under anaerobic conditions. For all non-clostridial strains, brain-heart infusion agar was employed with aerobic culture. Following overnight growth, bacteria were scraped from the plates and suspended directly in buffer (which varied depending on the particular experiment) to the desired optical density. PlyCM or lysin vehicle was added immediately prior to the start of each experiment, and OD₆₀₀-measurements were conducted in 96-well plate format at 22°C. For experiments that involved CFU-counts, each *C. perfringens* sample was diluted over five orders of magnitude, with triplicate plating at each dilution onto Schaedler agar.

RESULTS

We sought to compile a comprehensive list of phage lysins within all sequenced genomes of *C. perfringens*. In doing so, the following criteria were used for designating a given ORF as a probable lysin: [1] The presence of an N-

terminal enzymatic domain; [2] the presence of a C-terminal binding region¹; [3] the absence of an N-terminal signal peptide; and [4] the absence of any additional domains with non-lysin function. The third criterion is important because bacteria encode chromosomal peptidoglycan hydrolases (autolysins) that are involved in processes such as cell wall turnover, sporulation, and programmed death (Vollmer et al. 2008b). Some autolysins possess the same domain architecture as phage lysins, except that they also typically include a signal peptide and transverse the plasma membrane via the type-II pathway (rather than with the aid of holins). The fourth criterion is significant because phage themselves can encode other proteins (i.e. not lysins proper) that include peptidoglycan hydrolase domains, particularly structural proteins involved with DNA-injection (Rashel et al. 2008). Usually, however, these structural lysins are readily discernable by their much greater size (upwards of 1000 amino acids) and the presence of additional domains.

In total, 45 ORFs were identified with the above criteria. The GenBank accession numbers for their protein translations are provided in Figure 2.2, which depicts the predicted evolutionary relationships among the enzymes as a consensus phylogram. Of the 45 proteins, 23 were predicted to possess N-acetylmuramidase activity with an N-terminal GH-25 domain (glycosyl hydrolase type 25, Pfam accession number PF01183). The 22 others were predicted to possess N-acetylmuramoyl-L-alanine-amidase activity.

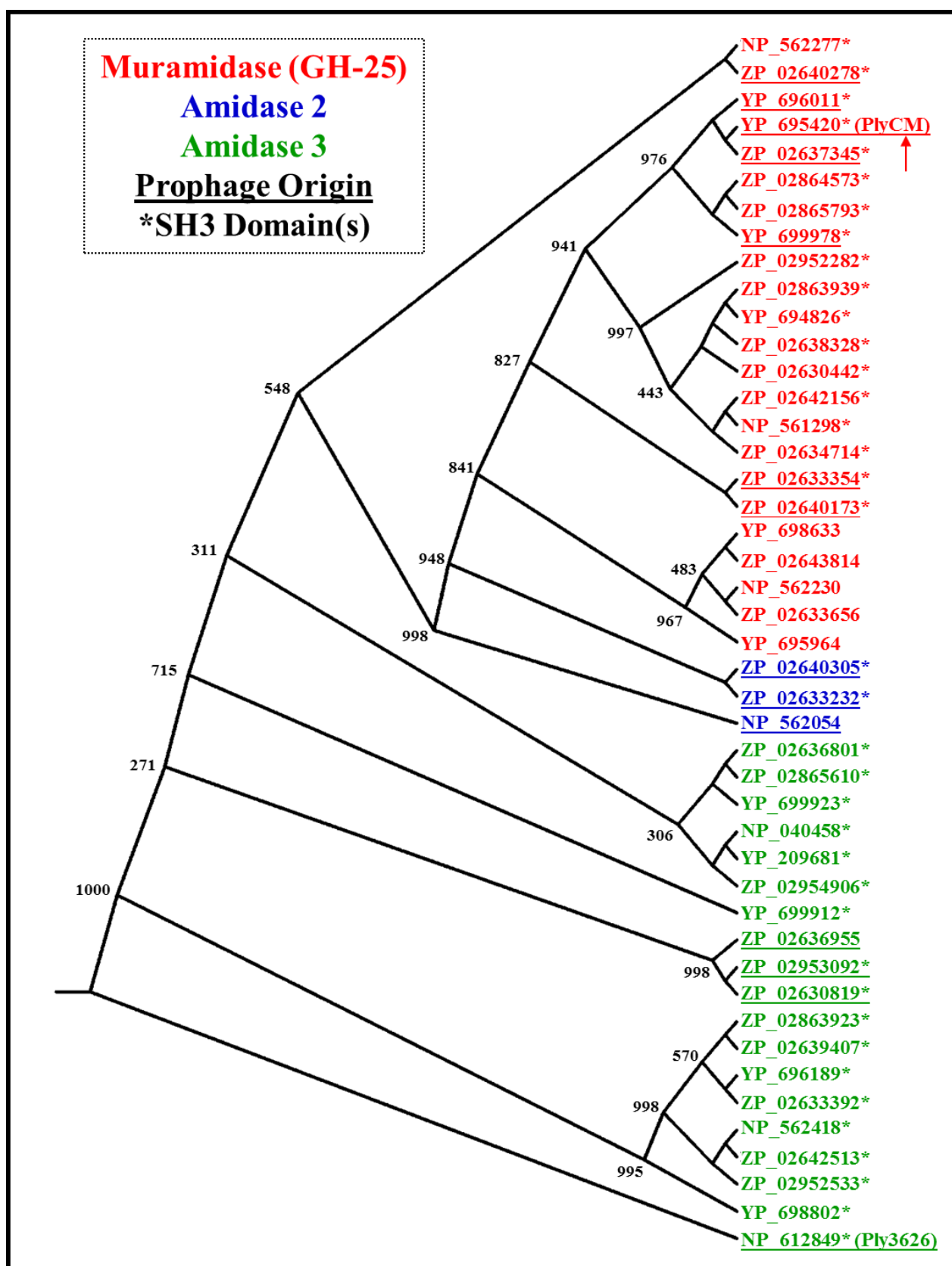


Figure 2.2 Lysin-Like ORFs in *C. perfringens* Genomes

Figure 2.2, continued

The phylogenetic relationship among *C. perfringens* enzymes with phage-lysine-like architectures is depicted here as a distance-based phylogenetic tree; the numbers at select nodes represent the consensus values following 1000 rounds of bootstrap analysis. The catalytic domain of each protein is indicated with the corresponding color scheme (see inset). The proteins for which pfam analysis predicted a C-terminal SH3 type 3 domain (either a single or a dual) are marked with an asterisk*. Of these 45 proteins, 14 appear to be phage lysins proper, in that they are encoded within proviral regions within the *C. perfringens* genomes. These are denoted with an underline. The other *lysine-like* proteins are highly homologous to the phage enzymes, but reside elsewhere in the chromosome or in an associated plasmid.

Three encoded a type 2 alanine-amidase domain (PF01510), while 19 encoded a type 3 alanine-amidase domain (PF01520). Although these two alanine-amidase families diverge sequentially, they target the same bond at the beginning of peptidoglycan's pentapeptide stem. The only other characterized *C. perfringens* phage lysin, Ply3626, is a type 3 alanine-amidase and is specified in the figure.

At their C-termini, 38/45 enzymes (also denoted in Figure 2.2) contained either a single or a double SH3-3 binding domain (PF08239). For some of the proteins, the degree of alignment with the SH3-3 consensus sequence was rather low (E-value range: 10^{-17} – 0.03). Although the molecular epitope is not well characterized, bacterial SH3 domains are commonly found in autolysins and phage lysins, and they presumed to function in binding the bacterial cell wall (Xu et al. 2009). The other 7 enzymes possessed extended C-terminal regions (see Footnote 1), although Pfam failed to recognize any conserved domains. One of these proteins, ZP_02636955, demonstrated Blast homology with the C-terminal regions of several of the other 45 enzymes. Most likely, ZP_02636955 possesses the same binding functionality as the others, although its C-terminal domain differs too greatly from the SH3-3 consensus for Pfam recognition. Another enzyme, NP_562054, demonstrated closest C-terminal alignment (E-value $\approx 10^{-35}$) to a segment found within the *C. perfringens* bacteriocin BCN5 (Garnier and Cole 1986). Finally, a group of 5 muramidases (all clustered together in Figure 2.2), contained C-terminal regions that – while nearly identical to one another – showed no homology to any other *C. perfringens* proteins.

Based on the selection criteria, we initially assumed all 45 of these proteins to represent legitimate phage lysins of proviral origin. After inspecting the position of the ORFs within their respective genomes, however, it soon became apparent that this was not the case. Many of the proteins (31/45) were encoded in genomic regions that did not correspond to prophage or prophage remnants. In fact, several patterns existed as to where these genes were encoded within the 9 sequenced strains, including in the vicinity of UV-inducible bacteriocins. The common occurrence of these *lysine-like* proteins within *C. perfringens* raises several interesting questions, and the issue will be discussed further in the Discussion section. Fourteen of the enzymes did reside in clear prophage regions (i.e. adjacent to other viral proteins). Underlined in Figure 2.2, these enzymes include a combination of muramidases, type 2 alanine-amidases, and type 3 alanine-amidases. The first two categories are of particular interest, as lysins of these classes have never been purified and characterized for *C. perfringens*.

We decided to focus our subsequent efforts on a prophage muramidase encoded by the CPE-negative toxinotype A strain ATCC 13124 (which is also the original type-strain for the species). We hereafter refer to this protein, YP_695420, as PlyCM (for *C. perfringens* muramidase). It is one of 2 prophage lysins encoded by ATCC 13124, the other being muramidase YP_696011. Several other non-perfringens phage lysins with GH-25 lytic domains have been characterized to date. These enzymes are summarized in Figure 2.3a, along with the binding domains with which they are paired.

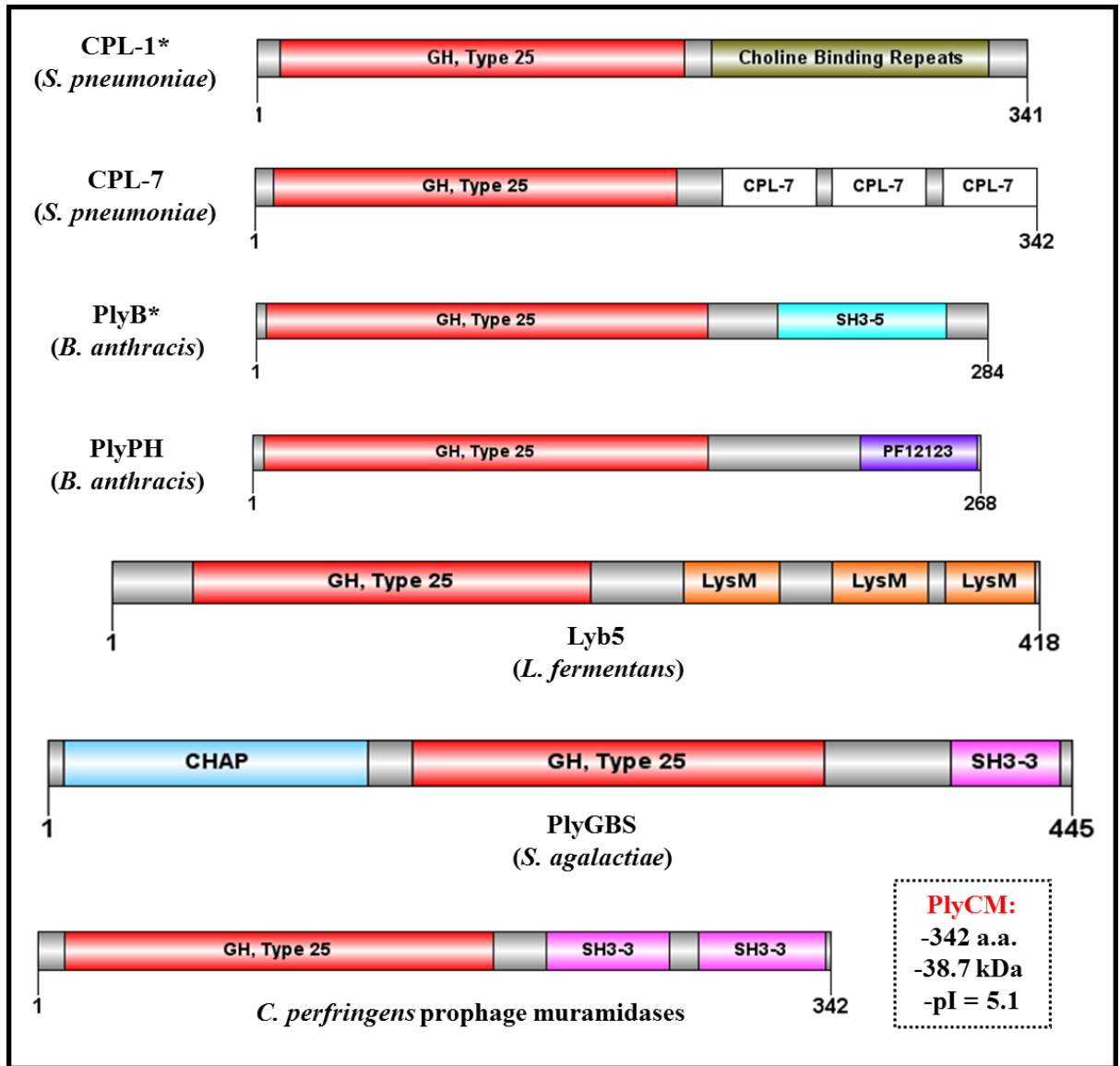


Figure 2.3a Architectural Diversity among GH25 Phage Lysins

Figure 2.3a, continued

Several phage lytic enzymes containing a GH-25 muramidase domain (all from non-clostridial phage) have been recombinantly expressed and studied to date. These are depicted here along with the variety of C-terminal binding domains with which they are paired. Lysins for which a crystal structure has been solved are indicated with an asterisk*. Included are: CPL-1 from *Streptococcus pneumoniae*, choline bindings repeats (pfam PF01473; GenBank ABC88204; Hermoso et al 2003); CPL-7 from *S. pneumoniae*, eponymous binding repeats (PF08230; AAA72844; García et al. 1990); PlyB from *Bacillus anthracis*, SH3-5 binding domain (PF08460; 2NW0_A; Porter et al. 2007); PlyPH from *B. anthracis*, conserved hypothetical binding domain (PF12123; NP_845154; Yoong et al. 2006); Lyb5 from *Lactobacillus fermentum*, LysM binding domains (PF01476; ABP88927; Wang et al. 2008); and PlyGBS from *Streptococcus agalactiae*, single SH3-3 binding domain (PF08239; AAR99416; Cheng et al. 2005). Also listed (in the inset) are several basic characteristics of the PlyCM lysin studied here.

In Figure 2.3b, the GH-25 domain of PlyCM is aligned with that of 2 non-clostridial lysins, ones for which crystal structures have been solved (Hermoso et al. 2003; Porter et al. 2007). As shown, the catalytic residues are conserved in all 3 cases.

PlyCM was recombinantly expressed in *E. coli* and purified by column chromatography (Figure 2.4). The lysin showed clear bacteriolytic activity *in vitro* against the encoding strain ATCC 13124. When added to a buffered suspension (20/10 mM phosphate/citrate pH 6.4) of live cells for 1 hr, PlyCM generated a ~70% reduction in bacterial turbidity at low nM (low-to-sub $\mu\text{g/ml}$) concentrations (Figure 2.5). The time-versus-OD response for PlyCM is very similar to that which was observed for Ply3626 against its host strain (Zimmer et al. 2002b). The required concentration of PlyCM is also commensurate with values observed for other Gram-positive bacteria and their respective lysins (Fischetti 2008). The residual turbidity after 1 hr of treatment can be considered a *baseline* value for these buffering conditions – neither increasing the lysin concentration nor the incubation time lead to further OD-declines. The reader should note that this baseline OD does not correspond to quantitative viability levels (see CFU analysis below).

Acidity represents one of the most important variables affecting lysin activity, so we sought to determine the effect of pH on PlyCM. Two sets of OD-drop experiments were conducted in which pH was varied at a constant lysin concentration.

Cpl-1	1	MASVKKNDLF	-VDVSSHNGY	DITGILEQMG	TTNTIIKISE	STTYLNPC--
PlyB	1	MGYI-----	-VMSKWNGS	PDWDTAKGQ-	LDLVTIARVQD	GSNYVDEVYK
PlyCM	1	MESRNNNNLK	GIDVSNWKGN	INFQSVKNDG	VEVVYIKATE	GNYFKDKYAK
			↑			
Cpl-1	48	--LSAQVEQS	NPFGFYHFR	FGGDVAEAER	EAQFFLDNVP	MQVKY-----
PlyB	43	DYVAAMKARN	IFFGSYAFCR	FVS-VEDAKV	EARDFWNRGD	KDSLFI-----
PlyCM	51	QNYERAKEQG	LRVGFYHEFR	A-----	-----NKGA	KDQANFFVNY
Cpl-1	91	-----	--LVLDYEDD	PSGLAQANTN	ACLRFMQMIA	D-AGYKPIYY
PlyB	87	-----	--WVADVEVT	TMSDMRAGTQ	A----FIDELY	R-LGAKKVGL
PlyCM	86	LNEIGAVNYD	CKLALDIETT	EGVGARDLTS	MCIEFLEEVK	RITGKEVVVY
			↑ ↑			
Cpl-1	128	SYKPFTHDNV	DYQQTIAQFP	NSLWIAGYGL	NDGTANFEYF	PSMDGIRWW-
PlyB	121	YVGHHKYEEF	GAAQIKCDF-	--TWIPRYGA	KPAYPCDL--	-----W-
PlyCM	136	TYTSFANNNL	DSR--LSSYP	--VWIAHYGV	NTPGANNI--	-----WS
Cpl-1	177	-----QYSSN	-----PF	DKNIV		
PlyB	157	-----QYDEY	GQVPGI-GKC	DLNRL		
PlyCM	172	EWVGFQYSEN	GSVAGVSGGC	DMNEF		
				↑		

Figure 2.3b GH25 Domain Alignment

The GH-25 catalytic domain of PlyCM is aligned with those of CPL-1 and PlyB.

Three-way and pair-wise identities are indicated with a blue/pink/

yellow/gray color-scheme. Putative catalytic residues (see Porter et al. 2007) are

denoted with arrows.

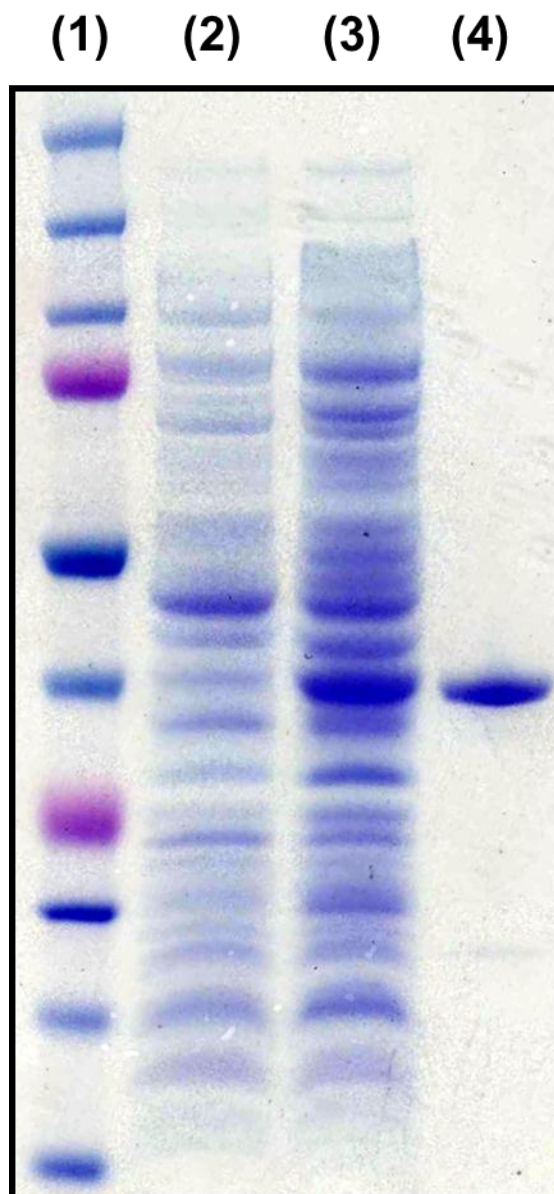


Figure 2.4 Purified PlyCM

PlyCM was purified following recombinant expression in *E. coli*. Lane 1: molecular weight ladder. Lane 2: crude extract of encoding strain prior to induction. Lane 3: crude extract of encoding strain ~16 hours after induction. Lane 4: final product following isolation protocol. By visual approximation, PlyCM is > 90% pure; its band appears at the level of the 37 kDa marker (predicted MW = 38.7 kDa).

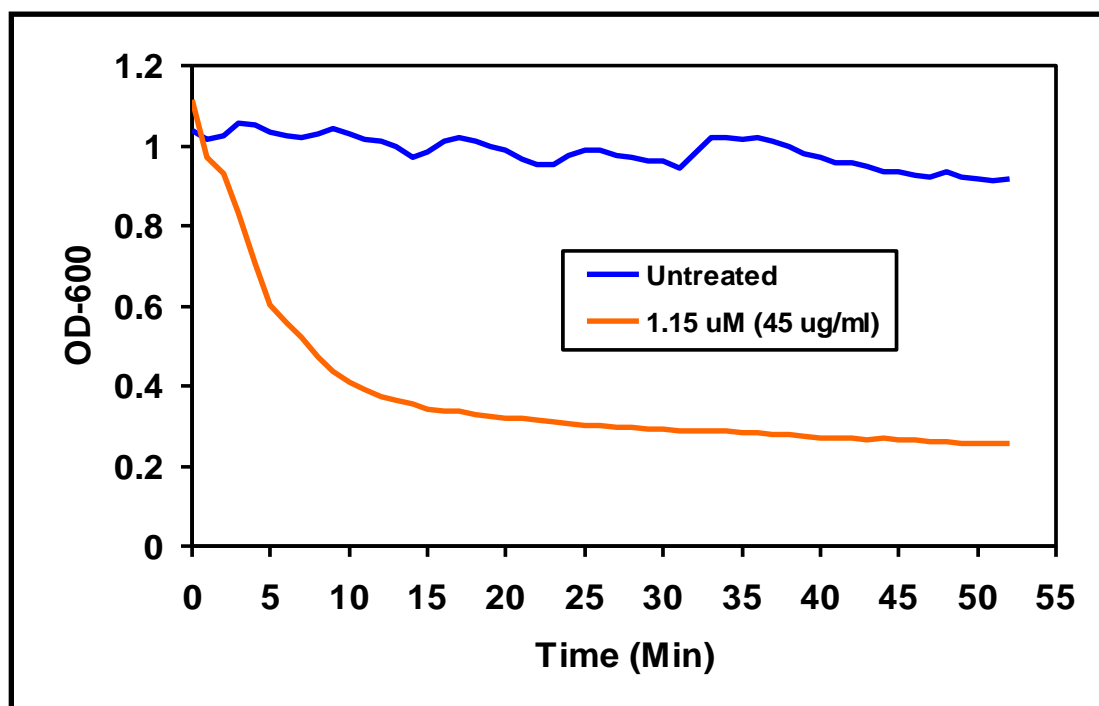


Figure 2.5a PlyCM-Induced Lysis (Time Course)

Depicted here is a representative example of an OD-drop experiment in which lysin was added to a buffered suspension (20/10 mM phosphate/citrate pH 6.4) of live host strain ATCC 13124. A 1.15 μM (45 $\mu\text{g}/\text{ml}$) enzyme-treated sample and an untreated control are shown.

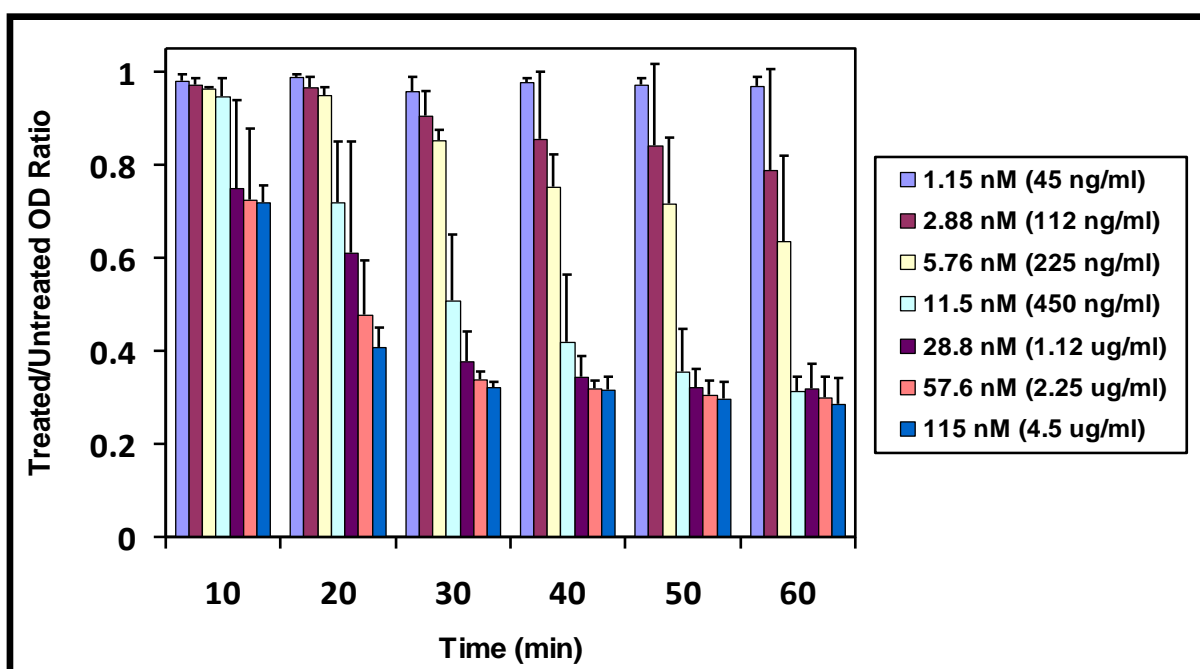


Figure 2.5b PlyCM-Induced Lysis (Variable Concentration)

The PlyCM concentration was decreased 10-1000 fold from the level shown in (A), and cell lysis was monitored over an hour. The figure reports treated/untreated OD-ratios (to account for slight possible fluctuation in the untreated) at 10-min intervals; the average of three independent experiments is depicted. As the graph indicates, concentrations as low as 11.5 nM (450 ng/ml) were able to bring the OD to a baseline level within 1 hr. In comparison, hen egg white lysozyme (HEWL, a non-specific eukaryotic muramidase, NP_990612, PF00062) failed to affect the cells' turbidity or microscopic morphology after 1 hr at 500 μ g/ml.

First, a broad-range buffer (25 mM boric/phosphoric acid) was utilized with a variable pH of 3.0 - 10.5 and a PlyCM concentration of 115 nM (4.5 µg/ml). Enzymatic activity was observed between pH = 4.0 - 9.5, with a maximum from 6.5 – 8.0 (Figure 2.6a). To confirm these observations (and fine-tune an optimal pH), a narrow-range buffer (20/10 mM phosphate/citrate) was utilized with smaller pH-degradations and a lower PlyCM concentration (11.5 nM, 450 ng/ml). A similar profile was observed here (Figure 2.6b), with pH = 6.4 as the center of activity.

Considering the acidic environment of the stomach (as well as many food items), we wished to determine whether the loss-of-activity at low pH represented a mechanistic inhibition or an irreversible denaturation. The latter scenario turned out to be the case. PlyCM was buffered at a range of acidic pHs, followed by titration back to pH = 6.5. Activity was subsequently lost for samples at pH = 3.3 and below. The pH = 4.3 sample, however, exhibited an identical lytic profile to the positive control maintained at pH = 6.5 (Figure 2.6c).

Several other biochemical characteristics of PlyCM were likewise evaluated, in particular the temperature stability of the enzyme (this is important given the potential use of a *C. perfringens* lysin in food science). Following incubation at various temperatures for 30 min, PlyCM demonstrated a sharp loss of activity (with concomitant flocculation) between 50°C and 55°C (Figure 2.7).

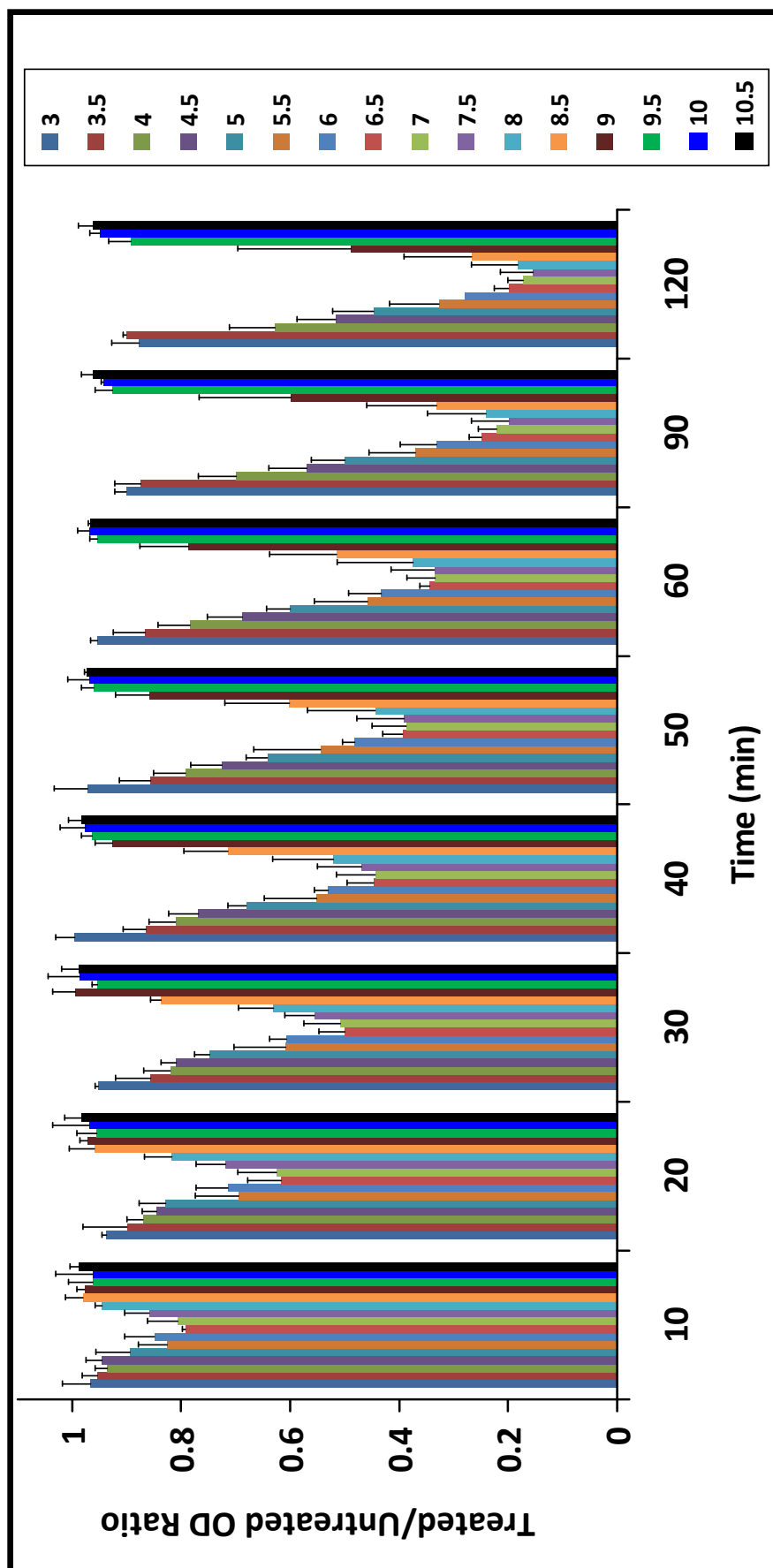


Figure 2. 6a Broad pH Dependence

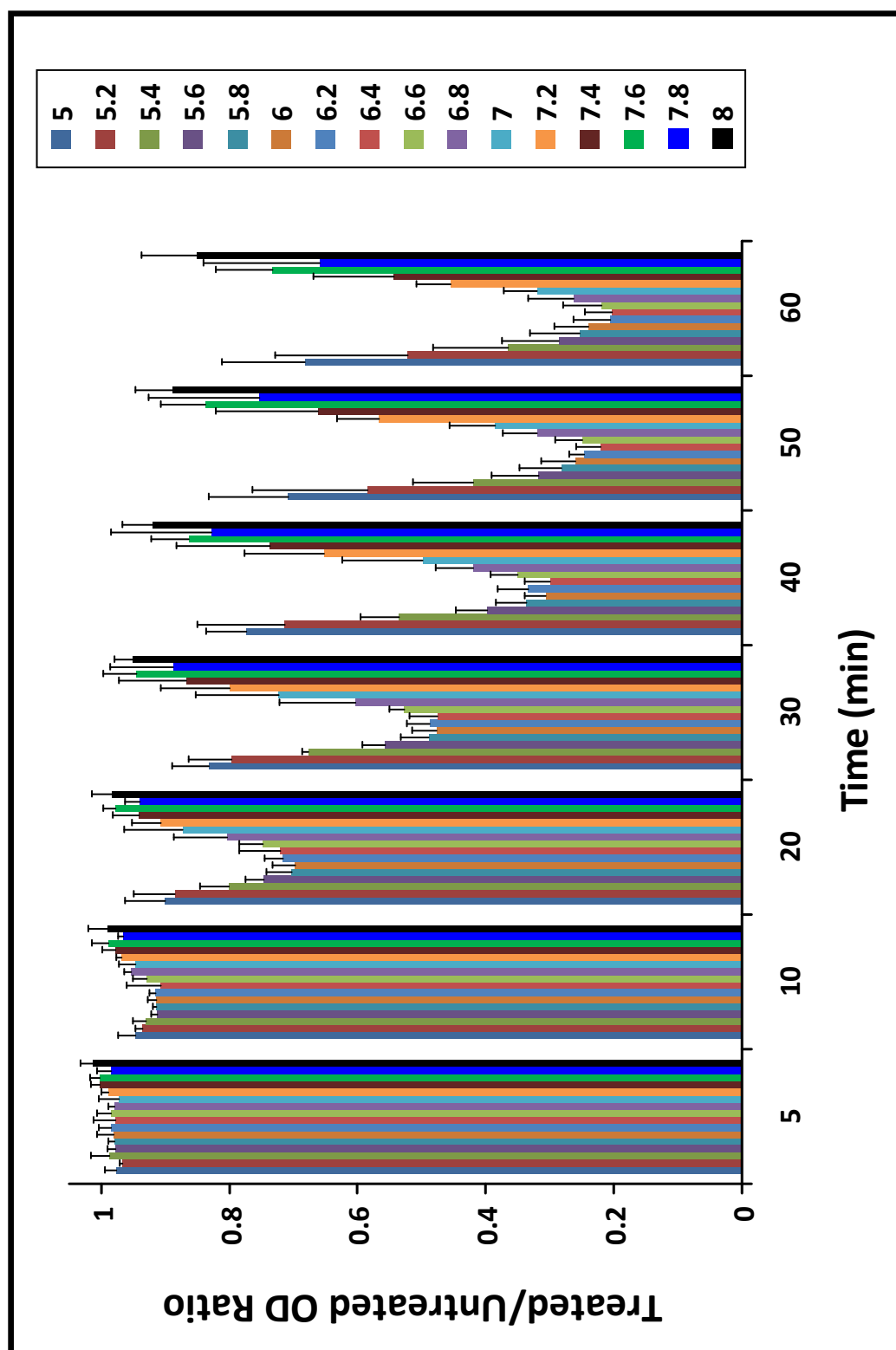


Figure 2.6b Narrow pH Dependence

Figure 2.6a/b, continued

[A] 115 nM PlyCM was incubated for 2 hr with suspensions of ATCC 13124 in a variable-pH boric/phosphoric acid buffer (pH: 3.0 – 10.5, 0.5 intervals, n = 3). By the end of the experiment, some degree of activity was observed between pH = 4 - 9.5, although lysis was maximal from 6.5 - 8. **[B]** The experiments in [A] were repeated under more sensitive digestion conditions. This include a 10-fold reduction in PlyCM concentration and the use of a phosphate/citrate buffer (pH = 5.0 – 9.0, 0.2 intervals, n = 3). After 1 hr, maximal activity was centered around pH = 6.4.

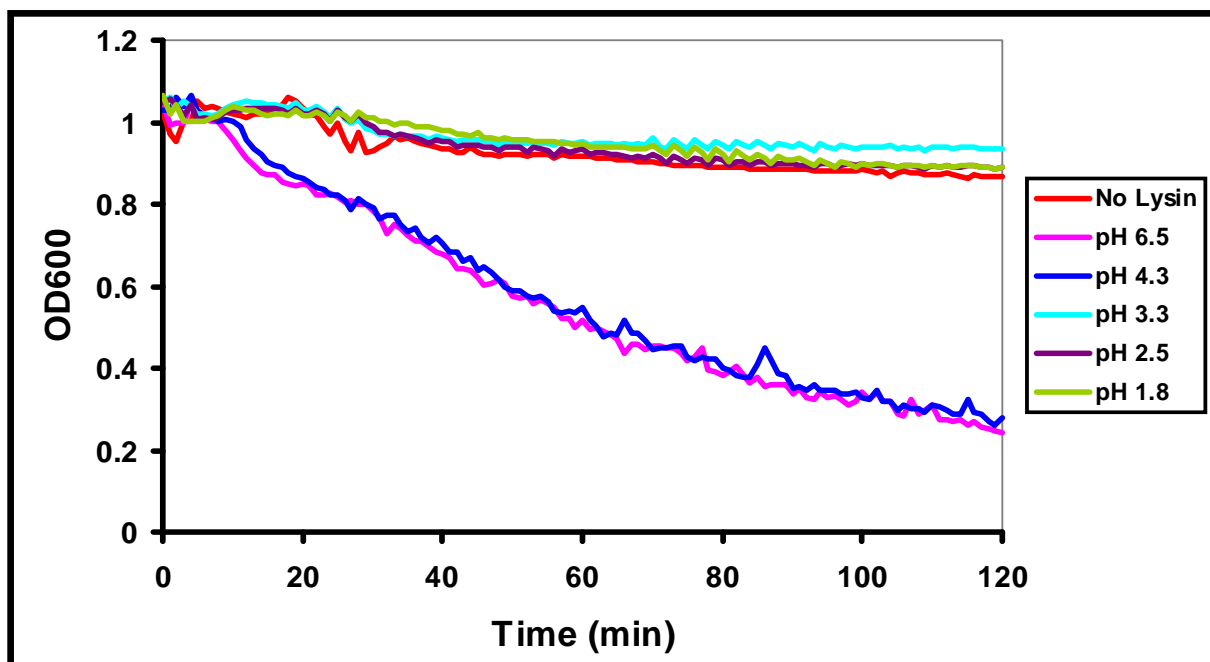


Figure 2.6c Irreversible Acid Denaturation

Aliquots of PlyCM were diluted in 20/10 mM phosphate/citrate buffers of various pH-values, yielding solutions whose final pHs are indicated in the graph. Each aliquot was titrated back to pH = 6.5 with dibasic phosphate; a positive control was maintained at pH = 6.5 for the entire experiment.

Following volume normalization (with phosphate/citrate pH = 6.5), lytic enzyme was added to the cells (final PlyCM concentration = 58 nM, final pH = 6.5) and lysis was observed for 2 hr. The sample brought to a pH = 4.3 showed identical activity to the positive control, while the others failed to induce lysis.

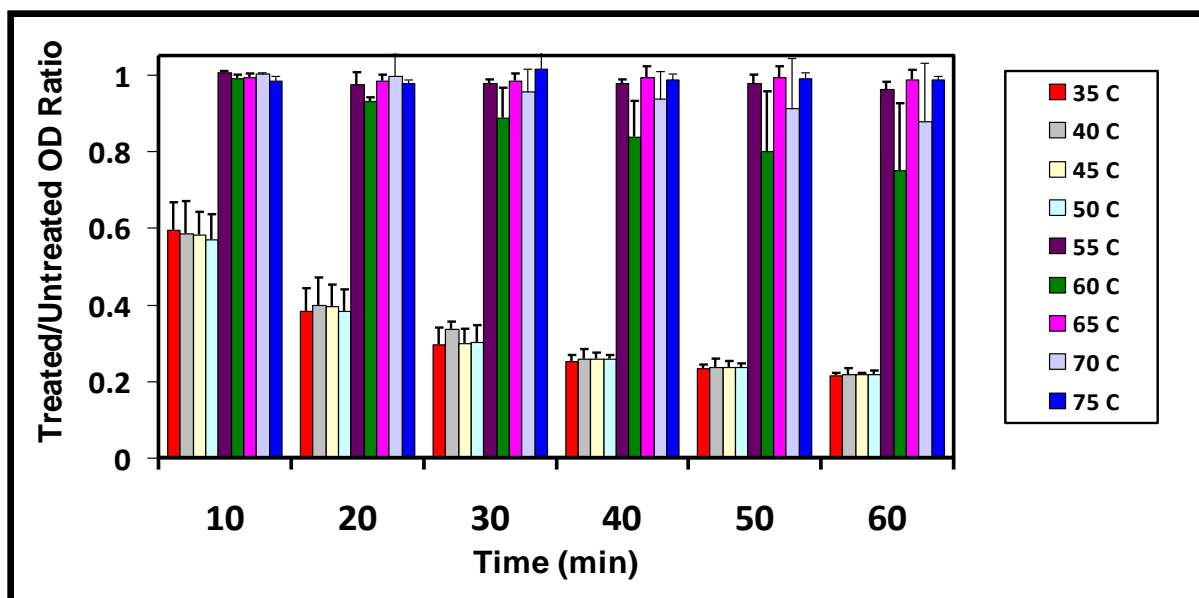


Figure 2.7 Temperature Stability of PlyCM

Stock concentrations of PlyCM were incubated for 30 min at the indicated temperatures, after which they were added to buffered suspensions (pH 6.4) of *C. perfringens* ATCC 13124 at 11.5 nM (450 ng/ml) for 1 hr. The figure reports the treated/untreated OD-ratio for each temperature at 10-min intervals (n = 4).

In separate experiments, molar excesses of either EDTA or DTT failed to inhibit lysin activity (data not shown), indicating that PlyCM is not dependent upon chelatable cations or intramolecular disulfide bonds for activity.

The above experiments were all conducted in low-osmolarity suspensions in which the only salt was the buffering agent. Seeing as any real-world application would likely occur in a less hypotonic environment, we evaluated the effect of salt on PlyCM-induced lysis. The experiments in Figure 2.5 were repeated with 150 mM NaCl in the lysis buffer. Although the concentration-dependence of PlyCM was virtually identical, the baseline OD of the treated cells jumped to ~60% of the untreated value (Figures 2.8a and 2.8b). Microscopic inspection of the cells indicated the reason: in the presence of 150 mM NaCl, many of the cells collapsed to spheroplast forms without extrusion of their cytoplasmic contents (see Figures 2.8c and 2.8d, respectively, for phase contrast and scanning electron micrographs).

In rod-shaped bacteria, the peptidoglycan layer serves two gross morphological functions. It prevents osmotic lysis of hypertonic intracellular contents and it maintains the bacilloid morphology, the natural lowest energy form being coccoid (Pichoff and Lutkenhaus 2007). Presumably, the addition of NaCl reduced the pressure on the cells to lyse, without affecting the pressure-independent collapse to spheroplasts. Indeed, when these lysin-treated cells were pelleted and resuspended in salt-free solution, immediate lysis was observed.

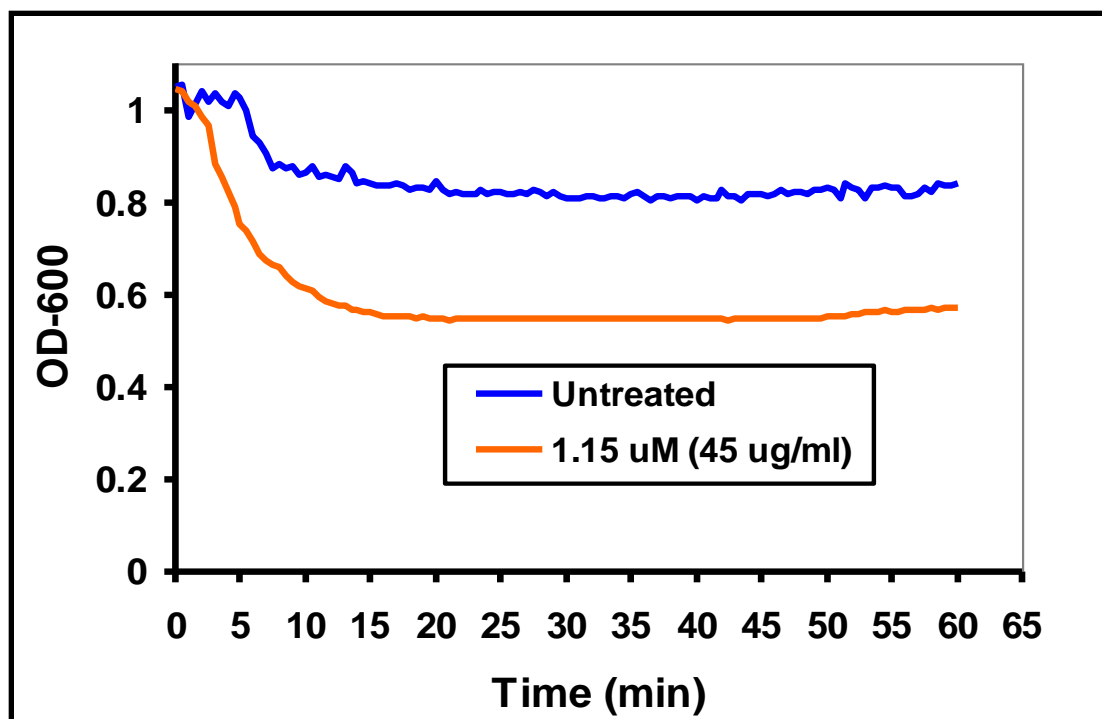


Figure 2.8a NaCl Effect on Lysis in ATCC 13124 (Time Course)

A single lytic time-course is shown with 1.15 mM PlyCM (45 μ g/ml) and buffering conditions that include 150 mM NaCl (pH = 6.4).

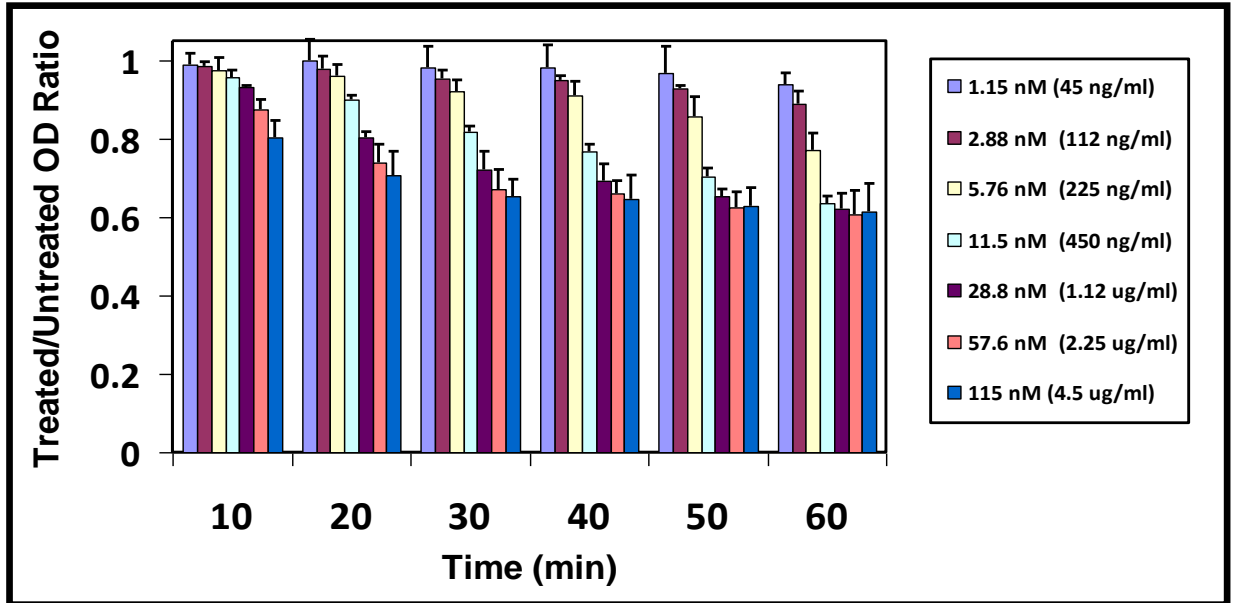


Figure 2.8b NaCl Effect on Lysis in ATCC 13124 (Variable Concentration)

The experiment in Figure 4b was repeated with the addition of 150 mM NaCl in the final lysis buffer (for B, n = 3). The clostridia responded to PlyCM at nearly identical concentrations as in Figure 4b, except that the OD would fall to a maximum of 60% the untreated value.

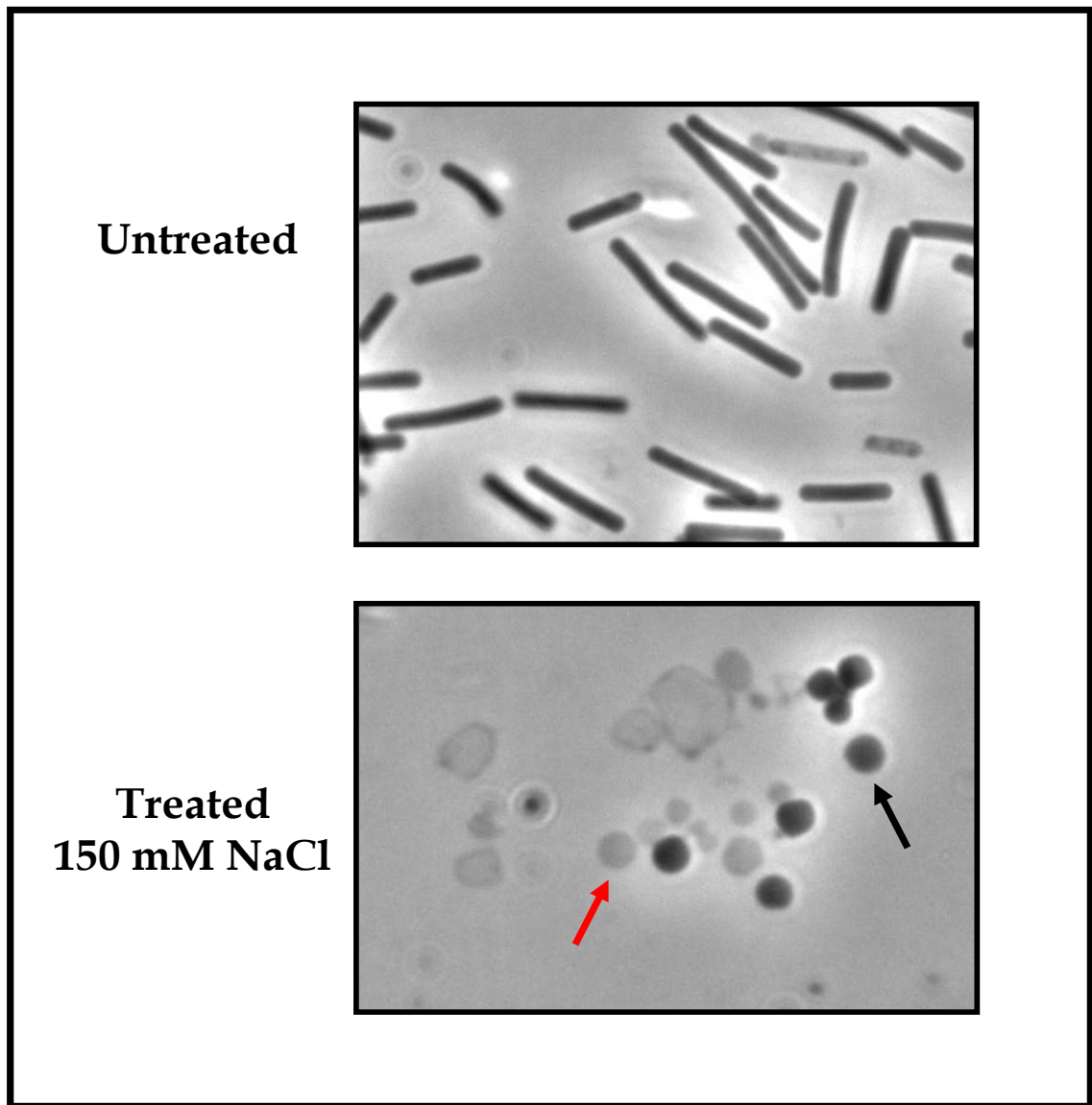
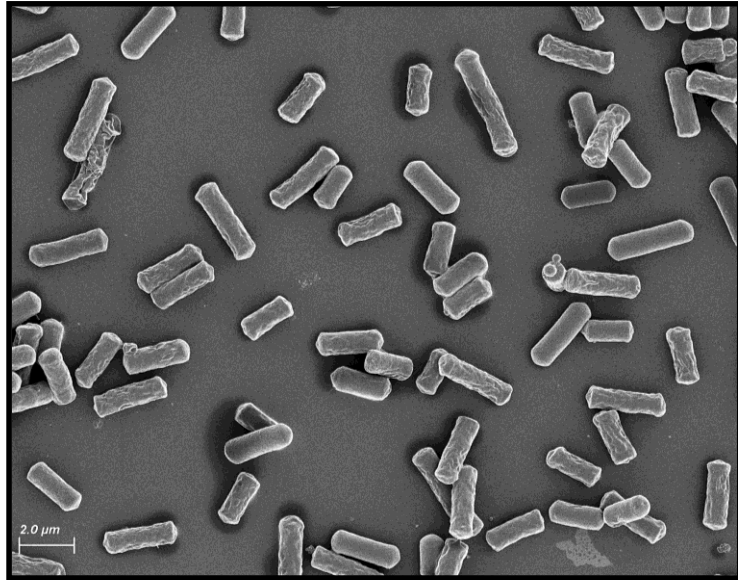


Figure 2.8c Light Microscopy of Lysin-Treated Cells in 150 mM NaCl

The image compares (1000X magnification) the gross morphology of untreated ATCC 13124 with that of cells treated with 115 nM PlyCM for 1 hr in the presence of 150 mM NaCl. Within the treated samples, there exists a combination of collapsed spheroplasts (black arrow) that have maintained their cytoplasmic contents and contribute to optical density, as well as clostridial ghosts and ghost-fragments (red arrow) that have undergone complete lysis.

Untreated



**Treated
150 mM NaCl**

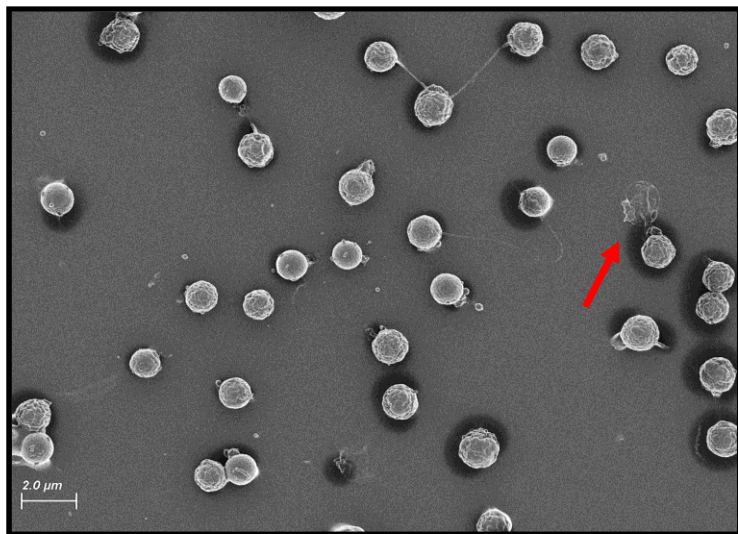


Figure 2.8d SEM of Lysin-Treated Cells in 150 mM NaCl

Scanning electron microscopy was also conducted on untreated and PlyCM-treated ATCC 13124 cells in the presence of 150 mM NaCl. The treated image is dominated by clostridial spheroplasts; a presumptive ghost is denoted with the red arrow.

The preceding observations raise the question of whether the spheroplast-like clostridia are viable. Ultimately, bacterial death following lysin treatment is a more significant metric than simple optical density. Accordingly, CFU-analysis was conducted on cells treated in the presence of 150 mM NaCl. Despite the residual turbidity, the cells were overwhelmingly non-viable. The following percentages-killed were observed after 1 hr (relative to untreated, n = 5 independent experiments): 11.5 nM (86 – 99.2%); 115 nM (> 99.7%); 1.15 μ M (> 99.99%). These data indicate that, even in environments that are incompatible with osmotic lysis, PlyCM exerts a potent lethal effect on ATCC 13124.

We next attempted to gauge the effect of PlyCM on actively-dividing cells. Unfortunately, minimum-inhibitory-concentration (MIC) analysis in liquid culture proved unsuccessful. At the concentrations employed in Figures 2.5 and 2.8, PlyCM failed to prevent proliferation of ATCC 13124. (An inoculum of $\sim 5 \times 10^5$ cells/ml was used, the minimal value for standard CFU analysis, Wiegand et al. 2008). Although we would have liked to repeat these experiments at higher lysin concentrations, a technical barrier prevented it. As mentioned in Materials and Methods, 100 mM L-arginine was added to the PlyCM stock solution to prevent enzyme precipitation. Higher lysin concentrations would have likewise necessitated elevated amounts of L-arginine in the media. In preliminary experiments, however, we observed that mM-concentrations of this amino acid could itself impact actively-dividing *C. perfringens* (the cells would proliferate

with defective division, yielding elongated rod forms²). With this confounding variable, we did not feel MIC analysis would be rigorously meaningful.

Finally, the activity of PlyCM was tested *in vitro* against a panel of other bacterial species and strains. These include: 24 additional isolates of *C. perfringens* (with representative examples of each toxinotype); 10 non-*perfringens* clostridia; and 16 non-clostridial species of Gram-positive bacteria. We found it necessary to conduct these experiments in a buffer (phosphate/citrate pH 6.4) that included 150 mM NaCl; without this, many strains of *C. perfringens* would self-adhere into macroscopic aggregates. Given the above observations involving spheroplast-formation at 150 mM NaCl, we did not want to rely exclusively on OD-measurements in evaluating the panel (as is typically done for assessing a lysin's target range). Consequently, a semi-quantitative scoring system was devised that relied upon microscopic inspection of the samples. Strains were ranked on scale from 1 (equally sensitive or more sensitive than host train ATCC 13124) to 4 (insensitive). The results of this panel are summarized in Table 2.1 (see the table caption for the specific details of the scoring system).

Overall, 23/24 strains of *C. perfringens* demonstrated susceptibility to the lysin, although the level varied quantitatively from strain to strain. Outside of *C. perfringens*, only 3/10 other clostridia demonstrated susceptibility, and these were ranked at the lowest level. Both of these findings are highly reminiscent of patterns observed for Ply3626 (although its panel of strains was evaluated on OD alone, Zimmer et al. 2002b).

Species/Strain	Sensitivity
<i>C. perfringens</i> ATCC 13124 (Type A, CPE -)	I (reference)
<i>C. perfringens</i> ATCC 3624 (Type A, CPE -)	II
<i>C. perfringens</i> ATCC 12915 (Type A, CPE +)	II
<i>C. perfringens</i> ATCC 12916 (Type A, CPE +.)	III
<i>C. perfringens</i> ATCC 12917 (Type A, CPE +)	III
<i>C. perfringens</i> ATCC 12919 (Type A, CPE +)	II
<i>C. perfringens</i> ATCC 3626 (Type B)	I
<i>C. perfringens</i> ATCC 3628 (Type C)	I
<i>C. perfringens</i> NCTC 8346 (Type D)	II
<i>C. perfringens</i> ATCC 27324 (Type E)	III
<i>C. perfringens</i> : 3 untyped human fecal isolates	Two— I , One— II
<i>C. perfringens</i> : 12 untyped canine fecal isolates	Three— I Seven— II One— III One— IV
<i>C. tetani</i> ATCC 19406, <i>C. septicum</i> ATCC 12464, <i>C. beijerinckii</i> ATCC 8260	III
<i>C. difficile</i> ATCC 43593, <i>C. difficile</i> ATCC 700057, <i>C. difficile</i> ATCC 9689, <i>C. histolyticum</i> ATCC 19401, <i>C. sordelli</i> ATCC 9714, <i>C. sporogenes</i> ATCC 3584, <i>C. bifermentans</i> ATCC 638	IV
<i>S. pyogenes</i> D471, <i>S. agalactiae</i> 090R, <i>S. pneumoniae</i> R36, <i>S. mutans</i> in-house strain, <i>E. faecalis</i> V583, <i>E. faecium</i> EFSK-2, <i>S. aureus</i> RN4220, <i>S. epidermidis</i> ATCC 12228, <i>M. lysodeikticus</i> ATCC 4698, <i>L. monocytogenes</i> HER1083, <i>B. anthracis</i> 222, <i>B. cereus</i> ATCC 14579, <i>B. subtilis</i> SL4, <i>B. thuringiensis</i> HD-73, <i>B. megaterium</i> in-house strain, <i>G. vaginalis</i> ATCC 14018, <i>P. aeruginosa</i> PAO1, <i>E. coli</i> TOP10	IV

Table 2.1 PlyCM Species/Strain Specificity

Table 2.1, continued

The above species/strains of bacteria were evaluated for their susceptibility to PlyCM. If known, the toxinotype and CPE-status (for toxinotype A) are listed for *C. perfringens* strains. Each bacteria was suspended in phosphate/citrate buffer (pH = 6.4) + 150 mM NaCl and exposed to three PlyCM-concentrations (11.5 nM, 115 nM, and 1.15 μ M). 11.5 nM represents the lowest concentration that brought the host strain ATCC 13124 to baseline OD (see Figures 4 and 7). After 1 hr, the samples were visualized at 1000X-magnification and the OD was measured. Each bacteria was assigned a *rank* based on the following semi-quantitative criteria: **[I]** 11.5 nM PlyCM induced complete lysis, or > 9/10 cells had converted from rod-forms to spheroplasts; **[II]** 115 nM PlyCM lead to either preceding observation; **[III]** 1.15 μ M PlyCM induced lysis/spheroplast conversion (complete or partial); **[IV]** No lysis or spheroplast conversion was observed after 1 hr, even at 1.15 μ M.

The other non-clostridial bacteria were all insensitive to PlyCM treatment. In total, this enzyme exhibits the kind of organism-specificity that has come to characterize phage lysins in general.

DISCUSSION

The work presented here illustrates the general utility of bacterial genomic sequencing in the identification of candidate enzybiotics. Moreover, the specific protein expressed and characterized (PlyCM) exhibits potent bacteriolytic ability *in vitro*, demonstrating potential as a novel antibacterial agent. In theory, phage lysins seem well-suited as food additives to combat *C. perfringens*. It is true that clostridia are spore-forming species, and that phage lysins are generally not effective against dormant spores (due to the external exosporium). Nevertheless, *C. perfringens* enteritis is not caused by the ingestion of spores, but rather vegetative cells that have proliferated in food (most commonly meats and meat broths). Rigorously-speaking, it does not represent *food poisoning* (i.e. involving preformed toxin), but rather foodborne infection. It is estimated that ingestion of $> 10^8$ vegetative bacteria are required for clinical symptoms (USFDA 2009). As a result, an agent that could selectively reduce the overall bacterial load or prevent its initial proliferation could prove quite useful.

Despite different enzymatic mechanisms, Ply3626 and PlyCM demonstrated several noteworthy similarities to one another. The time-dependant response of the host strain to treatment was nearly identical in both

cases. And the two enzymes possessed broad activity against a panel of *C. perfringens* strains, with little effect against other species and genera.

Quantitatively speaking, however, this level of sensitivity varied somewhat from isolate to isolate. For both Ply3626 and PlyCM, the treatment-response of the host strain was more pronounced than certain others. This observation underscores the importance of developing complimentary lysins with varied sequences and component domains. In fact, the ability to accomplish this rationally and rapidly for a given pathogen is an inherent strength of multigenomic sequence analysis.

In this regard, it would be informative for future work to consider the combined effect of a muramidase like PlyCM and an alanine-amidase like Ply3626. For instance, one could observe whether they normalize the strain-to-strain response or act synergistically with one another (as was observed for pneumococci in Loeffler and Fischetti 2003). Going further, one could even include non-viral enzymes in a combined treatment pool. Just recently, Camiade et al. characterized an endogenous *C. perfringens* peptidoglycan hydrolase (an autolysin) involved in cell division and stress responses (2010). This protein (Acp) encodes a glucosaminidase domain (Pfam family PF01832), which is distinct from all the sequences considered here. While autolysins have not received as much attention from the applied research community, nothing in theory would prevent them from functioning as enzybiotics alongside phage lysins proper.

In terms of future work on PlyCM, the immediate next steps must involve the fine-tuning of certain biochemical properties. In particular, the issue of solubility must be addressed; this would allow for increased concentrations of lysin to be tested against actively-proliferating cells (either in culture media or in the presence of actual food products). Possible strategies include the use of additives other than arginine or the re-engineering of its sequence to increase solubility. Specifically, the latter option could involve mutagenesis of the encoding gene or the design of a chimeric lysin that combines PlyCM's enzymatic domain with a different binding domain (either an SH3 domain with a somewhat varied sequence or another domain type altogether). These strategies have been employed successfully in the past to optimize the activity of other lytic enzymes (Cheng and Fischetti 2007; Daniel et al. 2010).

Aside from the goal of enzybiotic development, one other relevant issue in this chapter warrants further discussion. The initial genomic analysis revealed numerous enzymes that possessed all the architectural properties of a lysin (N-terminal lytic region, C-terminal binding region, no signal peptide), but that did not reside in a recognizable prophage region. Conceivably, they could represent small prophage remnants from their hosts' evolutionary history. Nevertheless, many of them demonstrated conserved genomic arrangements from strain to strain, suggesting a dedicated *in vivo* role. In fact, these lysin-like proteins are probably better classified as host autolysins, and their existence raises several noteworthy questions. For instance, what biological purpose do they serve for

the bacteria? And how do they gain access to the peptidoglycan layer without a secretion signal?

Several of the enzymes (ZP_02863923, ZP_02639407, YP_696189, ZP_02633392, NP_562418, ZP_02642513, ZP_02952533, YP_698802) are each encoded in the vicinity of a putative histidine kinase and metallo-beta-lactamase. They are classified as type 3 alanine-amidases, and are clustered together at the bottom of the phylogram in Figure 1. Also clustered together in the figure are several additional type 3 alanine-amidases (YP_699923, NP_040458, YP_209681, ZP_02954906, YP_699912); for these enzymes, the genomic positioning suggests an intriguing biological function. They are all encoded near UV-inducible *C. perfringens* bacteriocins known as BCN5 proteins³. Four of these lysins are located on plasmids, while one is chromosomal (ZP_02954906, strain JGS1721).

The BCN5 protein was first identified many years ago (Wolff and Ionesco 1975), as was its encoding gene within pIP404 plasmid of *C. perfringens* strain CPN50 (Garnier and Cole 1986; Garnier and Cole 1988). When CPN50 is irradiated, it undergoes partial autolysis with the induction and release of BCN5. The resulting culture supernatant (or purified BCN5) is subsequently inhibitory toward other *C. perfringens*. The genetic architecture of the BCN5-encoding region is conserved within the several strains in which it has (now) been sequenced. In each case, the ~2.7 kb *bcn* gene is immediately preceded by two ORFs in the same orientation: the ~500 bp *uviA* followed by the ~200 bp *uviB*. Dupuy et al. recently demonstrated that *uviA* encodes a DNA-damage inducible

σ -factor responsible for *bcn* transcription (2005), although other aspects of BCN5 activity remain undefined. These include its antibacterial mechanism, as well as the molecular events responsible for autolysis and concomitant BCN5 externalization (the bacteriocin contains no signal peptide). The role of *uviB* also remains unknown, although it has been noted that it demonstrates Blast-homology to a *Bacillus subtilis* phage holin (Dupuy et al. 2005).

In this regard, the above group of 5 lysin-like proteins offers potentially-valuable insights. Each is encoded immediately upstream of *uviA* on the opposite strand. When one considers the two ORFs between which *uviA* is sandwiched, a hypothetical mechanism for BCN5 release becomes evident: a phage-like process combining the murlaytic activity of the lysin with membrane permeabilization by the *uviB* “holin”. The release of the lytic enzyme during BCN5 induction could also contribute to the anti-clostridial effect attributed to BCN5 itself.

Although not identical, this mechanism would be reminiscent of several other “non-traditional” methods by which bacteria externalize intercellular contents. These include the damage-induced release of colicin from *Escherichia coli* (Cascales et al. 2007; Lin et al. 2009), the secretion of virulence factors by *C. difficile* (Tan et al. 2001); and the autolysis/allolysis of *Streptococcus pneumoniae* during competence-programmed predation (Guiral et al. 2005). The latter example involves the phage-lysin-like autolysin LytA; this enzyme has been studied at length, although it too lacks a signal peptide and its export mechanism

remains undetermined (Novak et al. 2000). Ongoing research will explore these possibilities, as well as analyze whether similar elements exist in the published genomes of other clostridia and Gram-positive bacteria.

In conclusion, the increasing ease of genomic sequencing is clearly transforming the fields of both basic and applied microbiology. These genomes are sources of both mechanisms and molecules, and it is through their careful inspection that researchers might uncover new avenues for improving human health and increasing general knowledge.

ACKNOWLEDGMENTS

This research was funded by NIH/NIAID grants AI057472 and AI11822 to VAF. JES acknowledges the kind support of the NIH MSTP program (Weill Cornell/ Rockefeller/Sloan-Kettering grant GM 07739), as well as the insights and efforts of Mr. Kareem Rumah. This research would not have been possible without the accommodation and advice of Drs. Maria Cristina Ossiprandi and Ezio Bottarelli of the University of Parma. We would also like to thank Dr. Davise Larone of New York-Presbyterian Hospital/WCMC for providing clostridial strains, as well as Ms. Cinzia Riverbi and Mr. Roberto Lurisi of the University of Parma for technical assistance. Electron micrographs were collected by Ms. Eleana Sphicas of the Rockefeller University Electron Microscopy Core Facility.

ENDNOTES

1. ORFs that encoded an extended C-terminal region without a Pfam-recognized binding domain were not excluded on this fact alone. There exist a number of Gram-positive phage lysins for which this is indeed the case. Most likely, these C-termini do possess a cell-wall binding functions, albeit ones that have not yet been characterized and organized into conserved protein families.
2. It should be emphasized that L-arginine, in itself, had no effect on bacterial turbidity or viability when cells were exposed to the amino acid for a short time in buffered solution (as in previous experiments).
3. This is the same BCN5 mentioned previously in the text. It demonstrated sequence-homology to the C-terminus of another *C. perfringens* lysin considered here (NP_562054, a type 2 alanine-amidase of evident proviral origin).

****NOTE:** In the manuscript to be submitted based on this chapter, the following individuals will be listed as co-authors (in the stated order): Maria Cristina Ossiprandi, Kareem Rumah, and Vincent Fischetti.

CHAPTER 3

Functional Viral Metagenomics I: The E-LASL Technique

INTRODUCTION

The rapidly growing field of metagenomics provides powerful new methods for studying environmental microorganisms and the natural products they synthesize (for reviews, see Green and Keller 2006; Ward 2006; Tringe and Rubin 2005; Daniel 2005; Handelsman 2004). Metagenomics involves the direct extraction of DNA or RNA from environmental samples (e.g. soil, sediment, water, feces), without prior laboratory cultivation of individual species. The appeal of metagenomics lies in the access it provides to genes and gene products

that cannot be isolated through typical culture-based techniques. This is significant considering that only a small percentage of environmental microbial species (~1% of soil species, for instance) are thought to be culturable under standard laboratory conditions (Torsvik and Ovreas 2002). Metagenomics is particularly well suited for the analysis of extreme environments whose conditions are notably difficult to reproduce *in vitro* (Baker et al. 2010).

Metagenomics research can be subdivided into two general categories: sequence-based and functional. In sequence-based studies, the environmental DNA (eDNA) is sequenced and compared to genetic databases to gain insight into environmental diversity, phylogeny, and ecology. Notable examples include a 76-megabase analysis of an acid mine biofilm (Tyson et al. 2004), a 1-gigabase analysis of the Sargasso Sea (Venter et al. 2004), and an extensive analysis of several oceanic samples (Rusch et al. 2007; Yooseph et al. 2007).

In functional metagenomics, by contrast, eDNA is transcribed and translated directly within a host organism, and clones are screened for the acquisition of a desired phenotype¹. Although functional metagenomics has constraints—such as the potential for low expression, misfolding, or clonal toxicity—it has already identified a number of molecules with biotechnological potential. These include regio- and enantio-selective biocatalysts (Ferrer et al. 2005), enzymes involved in vitamin biosynthesis (Eschenfeld et al. 2001), and antibiotic-resistance mediators (Donato et al. 2010). These examples are far from exhaustive, however, and the sorts of agents that could be targeted are

essentially only limited by one's imagination and ability to devise a read-out for identifying transformants. For several reviews that cover the broad goals of functional metagenomics, the reader is referred to Li et al. (2009), Uchiyama and Miyazaki (2009), Ferrer et al. (2005), and Voget et al. (2005).

It is important to note that functional metagenomic targets can be either proteins transcribed from individual genes or small molecules synthesized by several enzymes encoded by a contiguous cluster. Antibiotics are the most prominent example of the latter (Brady et al. 2009; King et al. 2009; Courtois et al. 2003; Gillespie et al. 2002; Brady et al. 2001). In this case, cosmids are typically utilized as the cloning vector (Brady 2007). Although they suffer (relative to plasmids) in that induced transcription is not possible, cosmids are capable of maintaining the lengthier inserts. One challenge regarding this sort of cloning is that gene clusters are often *so large* that traditional techniques fail to capture them in their entirety. This issue was recently addressed, however, by the use of transformation-associated recombination (TAR), in which overlapping contigs are reassembled *in vivo* (Kim et al. 2010). Typically, *E. coli* is used as the host organism in metagenomic screens (both for single genes and gene clusters), although other species have recently been employed with particular advantages (Craig et al. 2010 and 2008; Kim et al, 2010).

It should also be noted that hybrid sequence-based/functional approaches have also been used to identify novel agents (Bell et al. 2002, Marchesi and Weightman 2003). Here, eDNA is amplified with degenerate

primers targeted against conserved regions of known genes; the resultant amplicons can then be sequenced and tested for activity. Single-primer genome walking techniques are also possible, which allow for amplification of targeted genes without the burden of having to specify a reverse sequence downstream the gene of interest (Kotik 2009).

In addition to targeted amplification, random amplification has also proved an important metagenomic tool, as it allows for analysis of small initial quantities of eDNA (Rohwer et al. 2001; Delwart 2007). This is valuable in situations where only a small amount of sample is available, where the purity of the sample is low (Abutencia et al. 2006), or where the analysis involves only a subset of an environmental population. An example of the latter is the growing field of viral metagenomics (reviewed in Casas and Rohwer 2007; Delwart 2007, Edwards and Rohwer 2005), in which bacteriophage particles or other viruses are isolated prior to DNA (Angly et al. 2007; Breitbart et al. 2002 and 2003) or RNA (Culley et al. 2006, Zhang et al. 2006) extraction. (The flowchart in Figure 3.1 depicts how the field of metagenomics can be broken-down into its individual subfields: *sequence-based* versus *functional*, *bacterial* versus *viral*).

One previously-described method of amplification in viral metagenomics is the *linker-amplified shotgun library* (LASL) approach (Breitbart et al. 2002). Here, viral eDNA is mechanically fragmented, end-modified, and ligated to short double-stranded linkers; PCR is then performed with primers directed against the linkers.

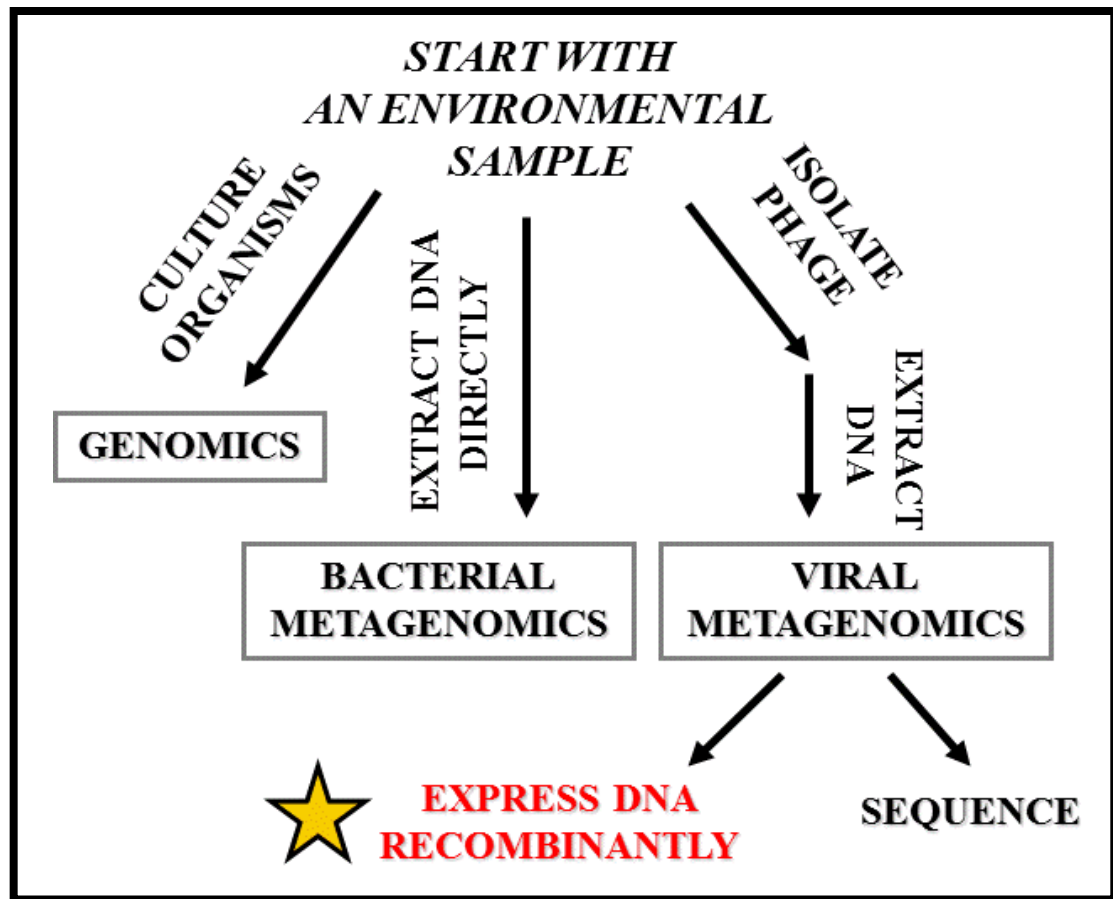


Figure 3.1 Metagenomics Flowchart

In metagenomic analyses, DNA is extracted directly from bulk environmental samples without the initial isolation and laboratory culture of any individual organisms. Metagenomics itself can be subdivided into several individual categories. A primary distinction is between bacterial metagenomics and viral metagenomics. In the former, total DNA (overwhelmingly bacterial in origin) is purified from a sample, whereas in the latter, the viral (predominantly bacteriophage) fraction is separated from the bacterial fraction prior to nucleic acid extraction. Bacterial and viral metagenomics, in turn, can each be subdivided into sequence-based and functional (i.e. involving screening) studies.

Although typically applied to metagenomic samples, the technique is equally applicable to genomic samples and could be particularly useful for genomes where the collection of large quantities of DNA is difficult or time-intensive (such as with certain environmental bacteriophages or slow-growing microbes).

In this study, we have combined the LASL approach with topoisomerase cloning to develop a rapid method of constructing expressible libraries with gene-sized inserts. Topoisomerase cloning utilizes the ability of vaccinia-virus topoisomerase to associate with terminal CCCTT motifs and ligate single-stranded threonine overhangs to complimentary adenines (Shuman 1994). 3'-adenine overhangs are generated automatically during *Taq*-based PCR by the template-independent terminal transferase activity of the polymerase.

Amplification of genomic or metagenomic fragments with *Taq* polymerase, therefore, allows for ligation in this manner. Our method is particularly well suited for functional screens, as commercially-available topoisomerase cloning kits allow for rapid ligation of PCR amplicons with subsequent inducible transcription in *E. coli*. Consequently, we refer to these libraries as “E-LASLs” (expressed linker-amplified shotgun libraries).

The utility of the technique was first demonstrated genomically with libraries constructed from the DNA of several recently-isolated *B. anthracis* bacteriophages. Antibacterial screening led to the identification of six new phage lytic enzymes with activity against this species. One of these lysins is the closest known homologue of a heretofore novel *Bacillus* muramidase, PlyB. Second, a

metagenomic DNA library was constructed from the gut contents of the European night crawler *Eisenia hortensis*. Functional screening of this library identified a positive clone in a hemolysis screen. The resultant protein – a new member of the aerolysin family – was purified and its activity confirmed against erythrocytes from several species, as well as human epithelial cells. It is the first example (to our knowledge) of a hemolysin isolated through functional metagenomic screening.

In total, the E-LASL approach may be used to mine for numerous classes of proteins from environmental bacteria and viruses. It is especially useful in any scenario (genomic or metagenomic) in which larger quantities of DNA are difficult or impossible to obtain, or where speed of library construction is of particular importance.

MATERIALS AND METHODS

Isolation of *Bacillus* phage DNA. The following commercially-available brands of bat and bird guano were screened for the presence of anti-*Bacillus* lytic phages:

[I] High Phosphorous Bat Guano 0-4-0, Fox Farm Soil and Fertilizer Company; [II] Superswell 0-7-0, Guano Company International, Inc. (GCI); [III] Bat Guano 4-6-2, e-Bio Organic; [IV] Jamaican Bat Guano 1-10-0.2, Sunleaves.com; [V] Dry-Bar Cave Bat 3-10-1, GCI; [VI] Desert Bat 8-4-1, GCI; [VII] Indonesian Bat Guano 0.5-12-0.2, Sunleaves.com; [VIII] Budswell 0-7-0, The Guano Company; [IX] Peruvian Seabird Guano 10-10-2, Sunleaves.com; [X] Mexican Bat Guano 10-2-1,

Sunleaves.com; [XI] Original Sea Bird 13-12-2, GCI; [XII] Fossilized Sea Bird 1-10-1, GCI; [XIII] Peruvian Seabird 10-10-2.5, GCI.

For each sample, ~5 g of guano were added to an equal volume of PBS (pH = 7.4) and shaken overnight at room temperature. The emulsions were centrifuged (20 min, 4,000 X g), and the supernatant passed twice through 0.22-micron PES filters (Millipore). Samples were loaded into 15-ml spin concentrators (15 kDa MWCO, Amicon) and reduced to ~100 μ l. Each concentrate was spread on freshly solidified fields of BHI soft-agar impregnated with *B. anthracis* strain Δ Sterne (3 ml molten soft-agar, 200 μ l stationary-phase culture, 10-cm plates).

Plates were incubated overnight at 30°C and inspected for plaques, which were noted for samples I - VII. Of these, isolates I - IV were chosen for further use in this study (designated BG-1, BG-2, BG-3, and BG-4). Phages were propagated for three generations on soft-agar, followed by two days of liquid culture (20 ml BHI, mid-log phase *B. anthracis*, 30°C). Following bacterial pelleting and sterile filtration, the culture solutions were subject to ultracentrifugation in a cesium chloride step gradient (150 min, 22,000 X g) and the phage-containing $\rho(1.45)/\rho(1.5)$ interface was collected (Sambrook et al. 1989). DNA was extracted directly by phenol-chloroform treatment, with chloroform wash and ethanol precipitation.

In addition to these phages, several additional *Bacillus cereus sensu latu* phages (A10, A14, A17, and TSH) were provided by Mr. Tommie Hata of the

Pingry School (Martinsville, New Jersey). They were isolated from environmental sources by local high-school students as part of a science-education curriculum (Hata 2010). As they were already purified as high-titer stocks; these phages were subjected directly to genomic extraction without further purification.

Isolation of worm-gut DNA. The method was adapted from Steffan and Atlas (1988). Live earthworms (species *Eisenia hortensis*) were isolated from forest litter from Stroudsburg, Pennsylvania. A single large adult was washed twice in phosphate buffered saline (pH = 7.4) and twice in dH₂O. The worm was transected into four equal sections to ease isolation of gut contents, which were manually extracted and resuspended in 1X PBS. The contents were pooled, sterile dH₂O was added, and the suspension was vortexed vigorously. The sample was centrifuged for 10 min at 1000 x g to remove large debris and soil particles. This step was repeated twice and the supernatants were pooled and centrifuged at 10,000 x G for 30 min.

The resulting bacterial pellet was resuspended in buffer (350 mM Tris-Cl, pH 8.0 + 10 mM EDTA) and centrifuged at 10,000 x g for 30 min. The pellet was resuspended in 1 ml of the same buffer, lysozyme was added (5 mg/ml final concentration), and the suspension was incubated at 37°C for 1 hr. Following a 5-min incubation at 60°C, SDS was added (1% final concentration) and the suspension was incubated at 60°C for 10 additional min. The suspension was

cooled on ice for 1 hr and centrifuged at 12,000 x g for 20 min at 4°C. Solid ammonium acetate was added to the lysate to a final concentration of 2.5 M and centrifuged at 12,000 x g for 30 min at 4°C. DNA in the supernatant was precipitated by adding 2.5 volumes of ice-cold ethanol and incubated at -70°C for 1 hr followed by centrifugation at 12,000 x g for 30 min at 4°C. The pellet was washed with 70% ethanol, air-dried, and resuspended in appropriate amount of TE buffer (10 mM Tris-Cl, pH 8.0, 1 mM EDTA).

Construction of E-LASLs. 0.01 or 0.1 units Tsp509I (diluted in 1X NEB Buffer II) were added to 100 ng genomic or metagenomic DNA in 50 µl total volume 1X NEB Buffer II. Samples were incubated at 65°C for 1 min and an equal volume of 25:24:1 phenol:chloroform:isoamyl alcohol (Roche) was added to terminate the digestion. The aqueous layer was isolated and washed with 50 µl chloroform. 50 µl 5 M ammonium acetate and 200 µl 100% ethanol were added to precipitate DNA, which was washed twice in 200 µl 70% ethanol. The DNA was desiccated and the following reagents were added: 6.5 µl dH₂O, 2 µl 20 ng/µl linker DNA (AATTCGGCTCGAG), 1 µl 10X ligase buffer, and 0.5 µl T4 DNA ligase (Fermentas). The underlined portion of the linker consists of single-stranded overhang, complimentary to the overhang generated by Tsp509I.

Based on the average size of the fragmented DNA (~1.5 kb) and assuming a DNA recovery of 100% following phenol/chloroform extraction (greater than the actual value), this mixture contains a ~55-fold molar excess of linker-to-insert

DNA. Following overnight ligation at 16°C, 4 µl of the reaction mixture were added to 44 µl dH₂O, 2 µl 10 µM linker-targeted primer (CCATGACTCGAGCCGAATT), and 50 µl *Taq* PCR Master Mix (Invitrogen). DNA was amplified with the following thermocycling conditions: 95°C-1 min.; (95°C-30 s.; 55°C-30 s.; 72°C-5 min.) X 40; 72°C-10 min.

Topoisomerase cloning of genomic/metagenomic inserts. 1 µl of the above PCR reaction mixture was added directly to 3 µl dH₂O, 1 µl 1X salt solution, and 1 µl linearized, topoisomerase-conjugated pBAD expression vector (pBAD-TOPO TA Expression Kit, Invitrogen). The mixture was incubated at room temperature for 1 min and divided (2 µl each) between three 50-µl vials of TOP10 chemically competent *E. coli* (Invitrogen). Cells were placed on ice for 20 min, heat-shocked at 42°C for 30 s, and placed back on ice for 5 min. 250 µl SOC media was added and the cells were incubated at 37°C for 30 min (as opposed to the manufacturer-suggested 60 min, to minimize the possibility of clonal expansion). Cells were spread onto LB agar with ampicillin (500 µg/ml), and these master plates were incubated overnight at 37°C.

Colony PCR. Individual colonies were picked and suspended in 25 µl 0.5 M NaOH, to which 25 µl 1 M Tris (pH = 8.0) and 450 µl dH₂O were added. 1 µl of this mixture was added to 1 µl 10 µM BadF primer (ATGCCATAGCATTTTATCC), 1 µl 10 µM Bad3 primer

(GCAGTTCCCTACTCTGCC), 12 μ l dH₂O, and 50 μ l *Taq* PCR Master Mix (Invitrogen). The BadF and Bad3 target sites are located, respectively, 160 bp upstream and 284 bp downstream from the point of topoisomerase-mediated cloning. Thermocycling was conducted under the same conditions as during library construction. Certain amplified inserts were sent for sequencing with BadF and/or Bad3 primers to Genewiz, Inc. (South Plainfield, NJ).

Functional screening of clones. To identify clones with antibacterial activity, two different screening methodologies were used. [1] For the BG-1, BG-2, BG-3, and BG-4 libraries, master plates were replicated with sterile velvet onto LB-agar with 0.2% arabinose (the pBAD-TOPO plasmid is arabinose inducible) and incubated overnight at 30°C. *E. coli* cell membranes were permeabilized by a 15-min exposure to chloroform vapor. The cells were overlaid with 7 ml molten BHI soft-agar containing 5 μ l overnight culture of *B. anthracis* (Δ Sterne), *Pseudomonas aeruginosa* (PAO1), or *Staphylococcus aureus* (RN4220). Plates were left at room temperature for 5 hr, transferred to 4°C for overnight storage, returned to room temperature the following day, and placed back at 4°C for one more night. Plates were inspected for clearing zones devoid of bacteria. [2] For phages BG-3, BG-4, A10, A14, and A17, the E-LASL were also screened using a novel hemolysis-based method that relies upon the toxic effect of the adjacent holin. This technique is the focus of Chapter 4, and the reader is referred there for experimental details.

For hemolysis screens, master plates were constructed and replicated as in method [1] above. Permeabilized clones were overlaid with 6 ml molten BHI soft-agar containing 1 ml defibrinated sheep blood (Cleveland Scientific). Plates were placed directly at 37°C and monitored for hemolytic zones. For any observed hit in the above screens, the corresponding colony on the master plate was identified, expanded, and retested in a second experiment. Confirmed hits were subject to PCR and sequence analysis.

Qualitative analysis of lysin activity. An *E. coli* clone expressing PlyBeta was streaked onto LB-agar with 0.2% arabinose and incubated overnight at 30°C. Cells were chloroform permeabilized and soft agar overlays were conducted as described above with a variety of bacterial species/ strains. These were: *B. anthracis* ΔSterne, *B. anthracis* RS222, *B. cereus* ATCC 10987, *B. cereus* ATCC 14579, *B. cereus* 13100; *B. cereus* 4429/73 FRI-16, *B. cereus* 03BB87, *B. cereus* E33L ZK, *B. thuringiensis* HD73, *B. thuringiensis* HD866, *B. thuringiensis* Al Hakam, *B. subtilis* SL4, *B. mycoides* 6462, *B. megaterium* in-house strain, *Staphylococcus aureus* RN4220, *L. monocytogenes* HER 1083, *Streptococcus pyogenes* D471, and *S. agalactiae* O90R. All species were grown in BHI soft agar at 37°C in ambient atmosphere. Plates were inspected for clearing zones around the PlyBeta expressing patches; permeabilized patches of non-transformed *E. coli* TOP10 were used as negative controls.

Purification of environmental aerolysin. The aerolysin-containing *E. coli* clone was expanded in 2 liters of LB at 37°C. At OD₆₀₀ = 0.5, arabinose was added to a final concentration of 0.2%, and the cells were incubated for 4 more hr. Cells were pelleted and the culture supernatant was subject to ammonium sulfate precipitation. (NH₄)₂SO₄ was slowly added to 40% saturation, and the solution gently rotated for 12 hr at 4°C. The precipitate was spun down (20 min, 7,000 X g), dissolved in 10 ml 10 mM phosphate buffer (pH = 7.4), and dialyzed for 24 hr against 100-fold excess of this buffer with three buffer changes.

The dialysate was loaded onto a MonoQ anion exchange column (Amersham), and eluted with a gradient of 10 mM phosphate buffer (pH = 7.4), 1M NaCl. Hemolytic activity was observed in the fractions corresponding to ~33 mM NaCl. These fractions were pooled together and subject to SDS-PAGE, which revealed a single ~48-kDa band upon Coomassie staining. Protein concentration was determined using BCA analysis with albumin standards. 50 µl of protein at 117 ng/µl was submitted to the Rockefeller University Proteomics Facility for 10 cycles of N-terminal sequence analysis.

Quantitative hemolysis assay. Hemolysis was quantified using a variation of the protocol of Eschbach et al. (2001). Defibrinated sheep and rabbit blood and alsevers-treated chicken blood were purchased from Cleveland Scientific (Cleveland, OH). Human blood was drawn from a healthy donor (with informed consent) into a heparinized vacuum tube (BD) immediately prior to

use. All samples were centrifuged (10 min; 2,500 X g), and the plasma and leukocyte layers removed. Erythrocytes were washed 4 times and resuspended in PBS to 10^9 cells/ml. 10- μ l aliquots of this suspension were diluted to 100 μ l with PBS containing various concentrations of aerolysin. Samples were incubated for 1 hr at 37°C. A 0% hemolysis standard was created by adding 10 μ l of the 10^9 cells/ml suspension to 90 μ l PBS, which was incubated for 1 hr at 37°C. A 100% hemolysis standard was created by adding 10 μ l of the suspension to 80 μ l dH₂O, followed by 37°C incubation and brief vortexing; 10 μ l 10X PBS were then added.

Following incubation, 900 μ l PBS were added to all experimental samples and standards, and 250 μ l of the supernatant were loaded into the wells of a 96-well plate. The supernatants of the 0% and 100% hemolysis standards were mixed to create intermediate standards. Absorbance was measured at $\lambda = 410$ and hemolysis of the experimental samples was calculated through linear reduction versus standards. The aerolysin itself exhibited negligible absorbance at this wavelength.

Computational analysis. Metagenomic sequences were analyzed with the TBlastX algorithm (translated query versus translated database; www.ncbi.nlm.nih.gov/Blast/). Searches were conducted against the non-redundant database (nr) with a maximum reported E-value of 0.001 and all other parameters at default value (Altschul et al. 1997). For the genes cloned in this

study, homologues were identified through the BlastX algorithm (translated query versus protein database) with default parameters. Predicted enzymatic and binding domains were assigned through Pfam analysis (Finn et al. 2009; pfam.sanger.ac.uk/)

Multiple sequence alignment of known aerolysins was conducted with ClustalX, v1.81 (ftp-igbmc.u-strasbg.fr/pub/ClustalX/; Thompson et al. 1997). From this alignment, phylogenetic analysis was performed with PHYLIP v3.67 (evolution.genetics.washington.edu/phylip.html; Felsenstein 1989); DNA distance and parsimony methods were employed with 1,000 rounds of bootstrapping and all other parameters at default value. The sequences cloned in this study were submitted to GenBank with the following accession numbers: BG-1 lysin (PlyBeta), EU258891; BG-2 lysin, EU258892; BG-3 lysin, EU258893; environmental aerolysin (Aer_M), EU258894. The sequences for lysins BG-4, A14, and TSH lysins have not yet been submitted to GenBank, although they are reported in the appendix of this thesis.

RESULTS

Construction of E-LASLs. To conduct functional genomic and metagenomic screens, a rapid technique was devised that combines linker amplification with topoisomerase cloning. The overall E-LASL procedure is outlined in Figure 3.2.

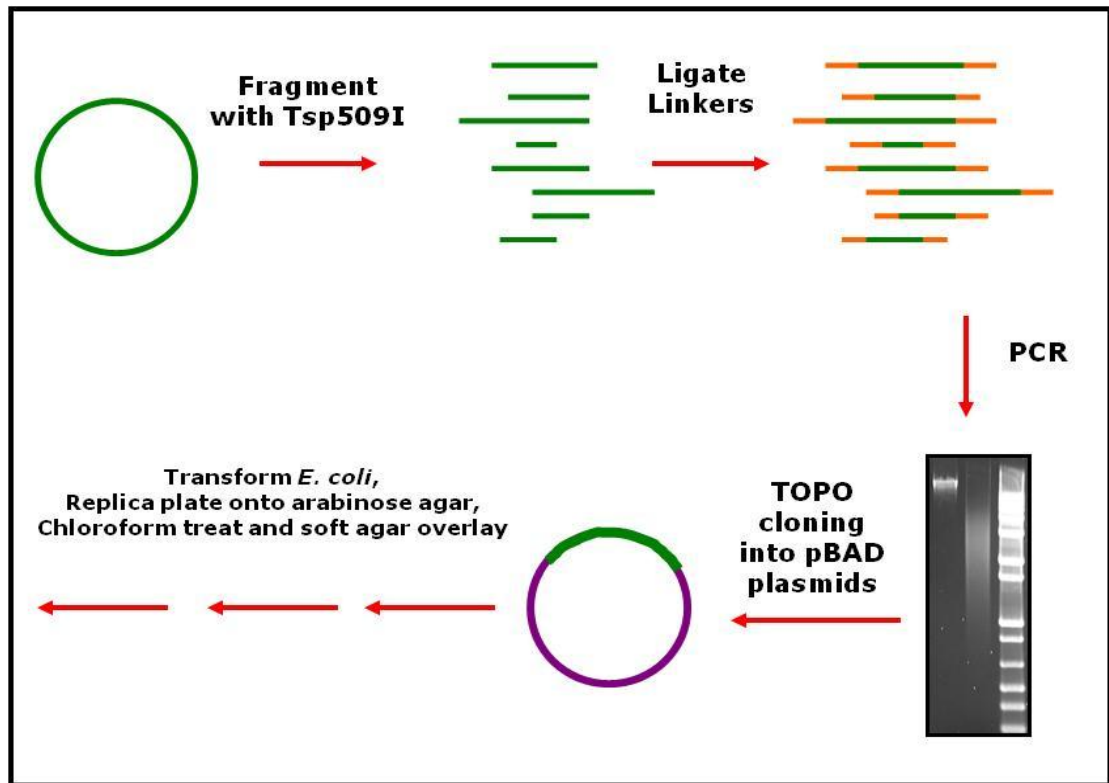


Figure 3.2 Constructing an E-LASL

Genomic or metagenomic DNA (represented by the circle at the upper-left) is enzymatically fragmented by brief exposure to Tsp509I. A short segment of linker DNA with a complementary overhang is ligated to both ends of the fragments. PCR is conducted with *Taq* polymerase and a primer targeted against the linker sequence. The resultant amplicons have a range of molecular weights, but all possess 3'-adenine overhangs. The agarose gel at bottom-right depicts bacteriophage genomic DNA (left lane) alongside its E-LASL amplification products (middle lane). The amplicons are cloned directly into linearized pBAD plasmids using commercial topoisomerase cloning. *E. coli* are transformed with the plasmids and screened for the acquisition of a desired phenotype (in our case, using soft agar overlay experiments).

Rather than fragmenting DNA with a mechanical shearing apparatus (as in previous studies, Breitbart et al. 2002 and 2003), we instead relied upon brief enzymatic digestion with Tsp509I. The four-nucleotide consensus sequence (AATT) would be expected to appear once every 256 bases of random DNA. As shown in Figure 3.3a, the following conditions for partial digestion lead to the desired gene-sized lengths after amplification: 100 ng genomic or metagenomic DNA; 0.01 or 0.1 units Tsp509I; 50 μ l reaction volume; and a 1 min digestion at 65°C.

The length distributions of the amplicons ranged from 500 bp to >4 kb, with the maximum intensity centered around ~2 kb (for 0.1 units of enzyme) or ~3 kb (for 0.01 units of enzyme). The length distribution of pooled, unamplified genomic DNA digested under the same conditions differed in that it extended to a slightly lower mass (~300 bp) and contained prominent amounts of undigested DNA (Figure 3.3b). The undigested DNA, however, is not amplified during linker-based PCR and is non-contributory to the final libraries.

We should note that it is possible that the amplified libraries contained ligated chimeras of two or more digested fragments (possibly the origin of the longest E-LASL fragments). Such chimeras are of little concern, however, given the functional nature of the screens and the fact that the majority of digested, unamplified DNA was gene-sized or greater in length. In preliminary experiments, E-LASLs were also constructed with 10 ng and 1 ng of starting DNA.

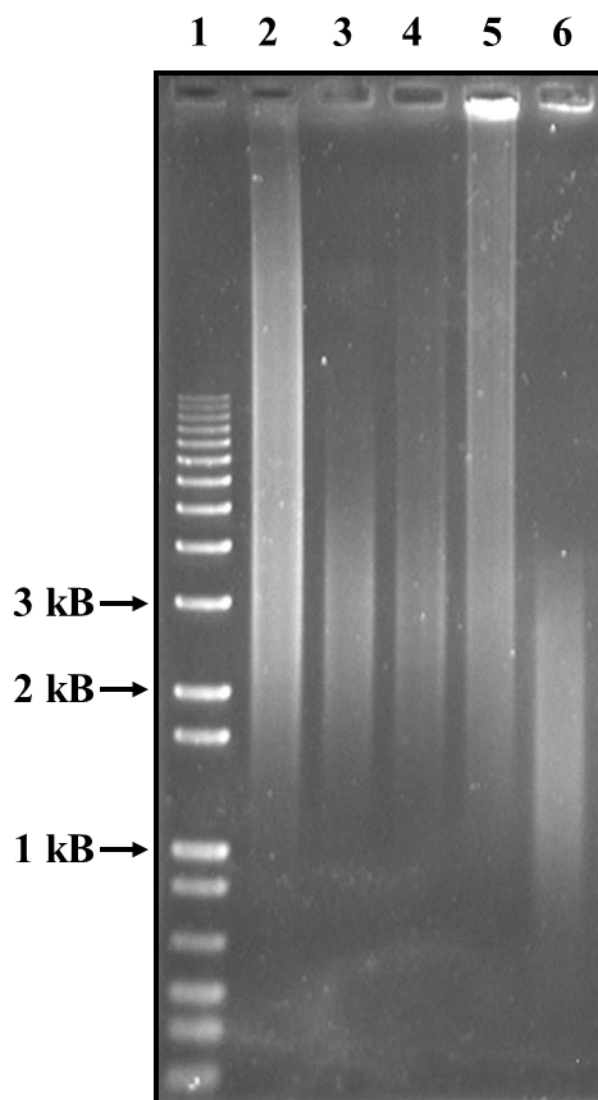


Figure 3.3a E-LASL Amplification Products

Lanes 2-5 depict E-LASLs constructed from four *Bacillus* phage genomic samples (100 ng DNA/sample; 1 min digestion; 0.01 units Tsp509I; 2 μ l reaction product/well). The lane 6 E-LASL was constructed from metagenomic DNA extracted from earthworm gut contents (100 ng DNA; 1 min digestion; 0.1 units Tsp509I; 2 μ l reaction product/well).

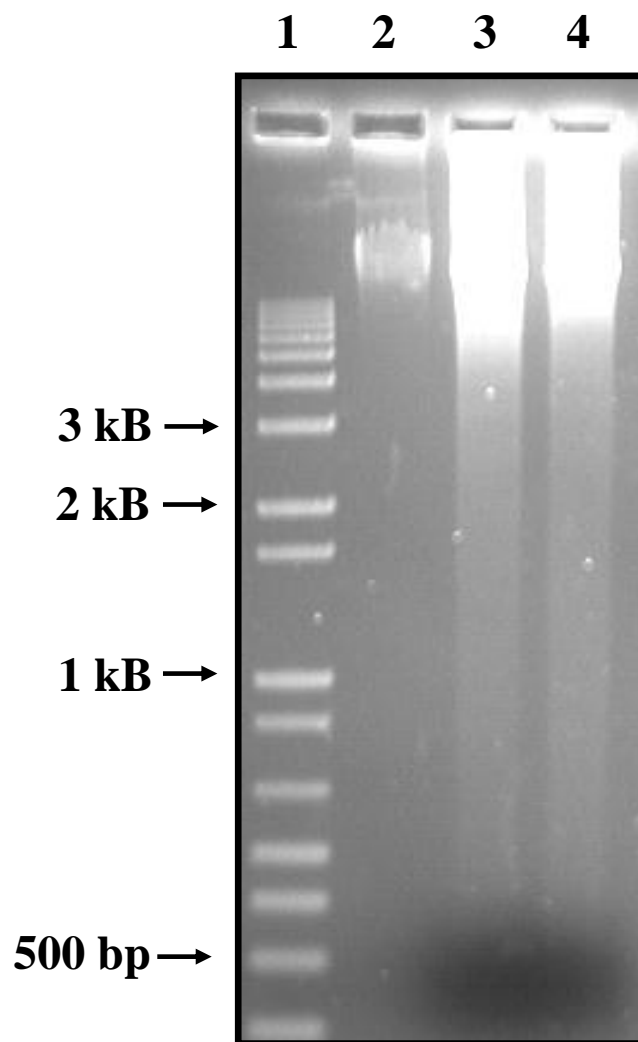


Figure 3.3b Digested DNA Prior to Linker Amplification

Lane 2 contains undigested phage genomic DNA. 1 µg was digested under the same conditions as during E-LASL construction: 0.1 units Tsp509I/100 ng DNA/50 µl reaction volume (lane 3) and 0.01 units Tsp509I/100 ng DNA/50 µl reaction volume (lane 4). The undigested DNA, however, is not amplified during PCR and is non-contributory to the final libraries. We should note that the amplified libraries could contain ligated chimeras of two or more digested fragments (the likely origin of the longest E-LASL components of 1a). These are of little concern, however, given that these are functional screens and that the majority of unamplified DNA is gene-sized or greater in length.

With these quantities, however, the success and efficiency of library construction were variable (most likely due to the physical loss of DNA during laboratory manipulations involved in E-LASL construction). Consequently, all libraries used in this study were constructed from 100 ng starting DNA.

Following *E. coli* transformation, the mean insert size of the library was determined through colony PCR and gel electrophoresis of randomly selected clones. The average insert size for the worm-gut metagenomic library digested with 0.1 units of Tsp509I was 1.99 ± 0.61 kb ($n = 65$), while for the phage genomic libraries digested with 0.01 units it was 2.27 ± 0.74 kb ($n = 97$, across all libraries). During this gel electrophoresis, a number of bands (~34%) ran at ~450 bp, corresponding to a circularized vector without any insert (verified through sequencing). These bands were not included as part of the above mean length calculations. Such clones also appeared in control experiments in which *E. coli* were transformed with a topoisomerase reaction mixture from which the insert DNA was omitted.

While the presence of these self-ligated clones in the final libraries was an unavoidable byproduct of the cloning kit, they did not affect the subsequent functional screens. Fortuitously, any clones containing this plasmid failed to proliferate in the presence of arabinose. The translated 36 amino-acid polypeptide encoded by these plasmids (consisting of an N-terminal enterokinase domain, a V5 epitope, and a C-terminal His-tag) apparently hindered *E. coli* growth when overexpressed by itself. Conversely, all clones that

did successfully transfer from the master plates to the arabinose plates were observed to contain a plasmid with a genomic/metagenomic insert.

In terms of colony yield, the number of individual clones that could be generated per topoisomerase cloning reaction (1 µl enzyme-conjugated plasmid, 1 µl E-LASL, 4 µl buffer) varied depending on the given cloning reaction and decreased with the age of the cloning kit (we recommend using freshly-purchased kits). For all libraries considered², the average number of clones per 6 µl reaction was 1187 (range: 596 – 2713). This number reflects the colonies that successfully transferred onto arabinose plates, thus specifically excluding colonies whose plasmids lacked an insert. We should note that increasing the E-LASL volume in a given reaction mixture or gel-purifying and concentrating the E-LASL prior to ligation did not improve (and, in fact, diminished) the colony yield.

Screening of bacteriophage genomic libraries. Various lytic bacteriophage capable of infecting *B. anthracis* (strain ΔSterne) were isolated from commercially available brands of bat guano. E-LASLs derived from four of these phages (BG-1, BG-2, and BG-3, and BG-4) were subjected to soft agar overlay screens in search of colonies with acquired antibacterial activity against *B. anthracis*. In general, such screens of bacteriophage genomes are ideal for testing the utility of the E-LASLs. All double-stranded DNA phage encode a cell wall hydrolase (known as a bacteriophage lytic enzyme, or lysin) with activity against the host

species. These enzymes are expressed at the end of a phage infective cycle, and, along with a pore-forming protein known as a holin, are responsible for lysing the host and freeing newly assembled viral particles. Lysins have generated considerable pharmacological interest due to their ability to lyse Gram-positive pathogens when added exogenously (Fischetti 2008 and 2005).

In bacterial-overlay screens, we were able to clone the lysin genes from three of the four libraries tested: BG-1, BG-2, and BG-3. In screens conducted during the development of the E-LASL protocol, we likewise cloned the well-characterized PlyG lysin from genomic DNA of the *Bacillus* γ phage (Brown and Cherry 1955, Schuch et al. 2002). The proportion of positive lysin hits among all clones varied for each library: BG-1: 8 confirmed hits, 640 clones screened; γ phage: 2 confirmed hits, 1600 clones screened; BG-2: 1 confirmed hit, 540 clones screened; BG-3: 1 confirmed hit, 2710 clones screened. For the BG-4 phage, 1220 clones were screened, but no hits were generated.

Each of the 8 confirmed hits for BG-1 encoded the full-length lysin, although no two contained flanking DNA of identical lengths. Six of these 8 clones, moreover, encoded the lysin on the strand whose transcription was not under control of the arabinose-inducible promoter (indicating the presence of an endogenous promoter with strong activity in *E. coli*). When considering only the two clones with the lysin in the forward orientation, the proportion of hits for BG-1 (2/640) is closer to that of BG-2 and γ phage. For BG-3, the one confirmed lysin hit encoded only the N-terminal fragment of the complete protein. This

fragment was still sufficient to create a zone of mildly decreased *Bacillus* density and altered coloration, allowing for identification. The remainder of the gene was sequenced through genomic primer walking.

Additionally, E-LASLs from BG-3, BG-4, and four other *Bacillus* phages (A10, A14, A17, and TSH) were screened for lytic enzymes using a novel technique described in Chapter 4. With this approach, the BG-3 lysin was successfully re-cloned (in its entirety, without the need for primer walking) and the BG-4 lysin was identified for the first time. Of the three new phages, the lysin for the A14 phage was successfully cloned. The A10 and A17 E-LASLs were not subsequently rescreened using the standard cellular-overlay technique, and their lysins remain undetermined.

Based on Pfam domain prediction, the majority of the cloned lysins (BG-2, BG-3, BG-4, A14, and TSH) are N-acetylmuramoyl-L-alanine-amidases. The first five of these enzymes possess an N-terminal type 2 alanine-amidase motif, while the TSH lysin possesses a type 3 alanine-amidase motif. Blast analysis suggests that this enzymatic specificity is very common among *Bacillus* phage and prophage lysins, with PlyG lysin as a notable example (Schuch et al. 2002). By contrast, the BG-1 lysin is predicted to have N-acetylmuramidase activity with a glycosyl hydrolase, type 25 domain; this motif appears less common among known *Bacillus* phages. At their C-termini, all the above lysins possess either a single or dual SH3 binding domain; BG-1 and the TSH lysin encode type 5 SH3

domains, while the others encode type 3 SH3 domains (see the Appendix for exact positional details).

The BG-1 lysin (henceforth referred to as PlyBeta) demonstrates homology throughout its entire sequence to only one other known lysin³ – PlyB lysin from the *Bacillus* BcpI phage, which was also recently characterized by the Fischetti laboratory (Porter et al. 2008). PlyB and PlyBeta are each other's closest known homologues (E-value = 10^{-143}). They share 78% nucleotide identity and 81% amino acid identity (Figure 3.4). The crystal structure of PlyB's catalytic domain (residues 1-190) was recently solved (Porter et al. 2008). Based on this structure, all active site residues are conserved between PlyB and PlyBeta, with a single exception. The tyrosine residue at position 160 of PlyB (part of the proposed substrate-binding interface) corresponds to a threonine residue at position 161 of PlyBeta. The most noted region of divergence between the two lysins is a 9 residue Asn-rich stretch beginning at position D189 of PlyBeta, immediately following the end of the catalytic domain. This region of PlyBeta also encodes an additional cysteine residue, increasing the total number of cysteines in PlyBeta to 3 (PlyB encodes 2 cysteines).

The activity of PlyBeta was examined qualitatively against a variety of *Bacillus* species. Soft agar overlay experiments were conducted with a PlyBeta-expressing *E. coli* clone and 14 strains composed of *B. anthracis*, *B. cereus*, *B. thuringiensis*, *B. subtilis*, *B. mycoides*, and *B. megaterium*.

		10	20	30	40			
PlyBeta	1	MGYIVDI	SKWNGDINWDV	AAGQLDLA	IARVQDGSNYVD	PMYKSYVASMKA		
PlyB	1	MGYIVDM	SKWNGSPDWD	TAKGQLDLV	IARVQDGSNYVD	PVYKDYVAAMKA		
		▲						
		60	70	80	90			
PlyBeta	51	RNV	PFGNYAF	CRFVS	VEDAKVE	ARDFWARGDKDALFWVADVEVK	TMGDMQ	
PlyB	51	RNI	PFGSYAF	CRFVS	VEDAKVE	ARDFWNRGDKDSLFWVADVEVT	TMSDMR	
						▲	▲	
		110	120	130	140			
PlyBeta	101	AGTQAF	IDELYRLG	AKKVGLY	VGHHTY	VAFGAKNIR	CDFTWIPRYGGLKP	
PlyB	101	AGTQAF	IDELYRLG	AKKVGLY	VGHHKY	EFGAAQIK	CDFTWIPRYG - AKP	
		160	170	180	190			
PlyBeta	151	DF	PCDLWQYT	TETGNV	PGIGKCD	INSLNSDKT	LEWFTGKDCNNGNVT	PPPPQ
PlyB	150	AY	PCDLWQY	DEYGV	QVPGIGKCD	LNRLNGDKS	LDWFTGK - - -	GEEAVQPPQ
				▲				
		210	220	230	240			
PlyBeta	201	GAYDSSWFTK	QTVFTLDRT	INLRTAPFP	NAPLIAQLN	NAGDNVNY	EAYGY	
PlyB	197	GNYDSSWFTK	QNGVFTLDRT	INLRTAPFP	TAPLIAQLN	NAGDNVTY	DGYGY	
		260	270	280				
PlyBeta	251	EKDGYVWLRQ	HARGNGNF	GYIAS	GETKNGQR	ISTWGTFK		
PlyB	247	EKDGYVWLRQ	NARGNGNY	GYIAS	GETKNGQR	ISTWGTFK		

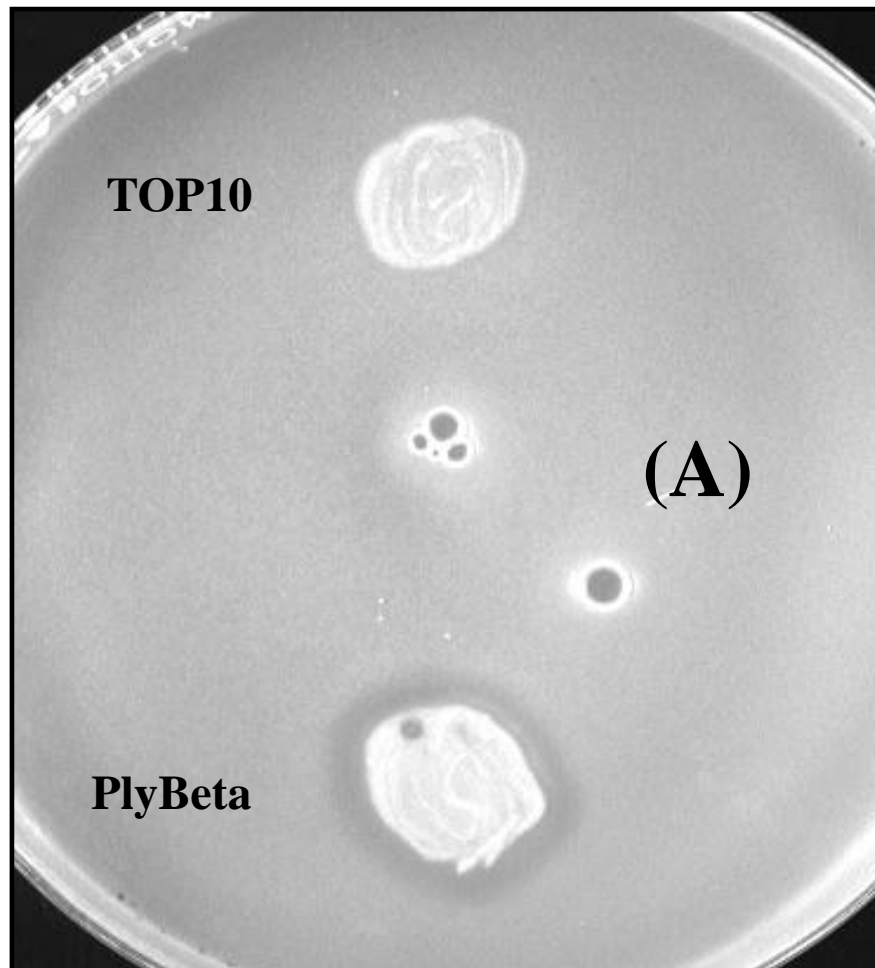
Figure 3.4 Sequence Comparison of PlyB and PlyBeta

The lysins share 81% amino acid identity, including all putative catalytic residues (noted with arrowheads). The most prominent area of divergence is a 9-residue, Asn-rich stretch immediately following the catalytic domain at position 189.

PlyBeta (like PlyB) is active against *B. anthracis* and other bacilli, although the activity varies considerably between species/strains. Clearing zones were noted for 7/14 strains: 2/2 *B. anthracis*, 3/6 *B. cereus*; 1/3 *B. thuringiensis*, 1/1 *B. subtilis*, 0/1 *B. mycoides*, and 0/1 *B. megaterium*. Representative examples of susceptible and non-susceptible strains are shown in Figure 3.5. Clearing zones were not observed for soft-agar overlays conducted with PlyBeta and strains of *Streptococcus*, *Staphylococcus*, and *Listeria*.

Screening of worm gut metagenomic library. An E-LASL was next constructed with DNA extracted from the gut contents of a single European night crawler (*Eisenia hortensis*). Prior to functional screening, a brief sequence-based analysis was performed to verify the bacterial origin of the extracted DNA. The inserts of 50 clones were subjected to a single sequencing read (37.8 kb total). When analyzed with the TBLASTX algorithm, 46 of the 50 reads returned known homologues. For each read, we calculated the relative contribution of different taxa to the total number of hits. The average of these values across the 46 reads is shown in Table 3.1a.

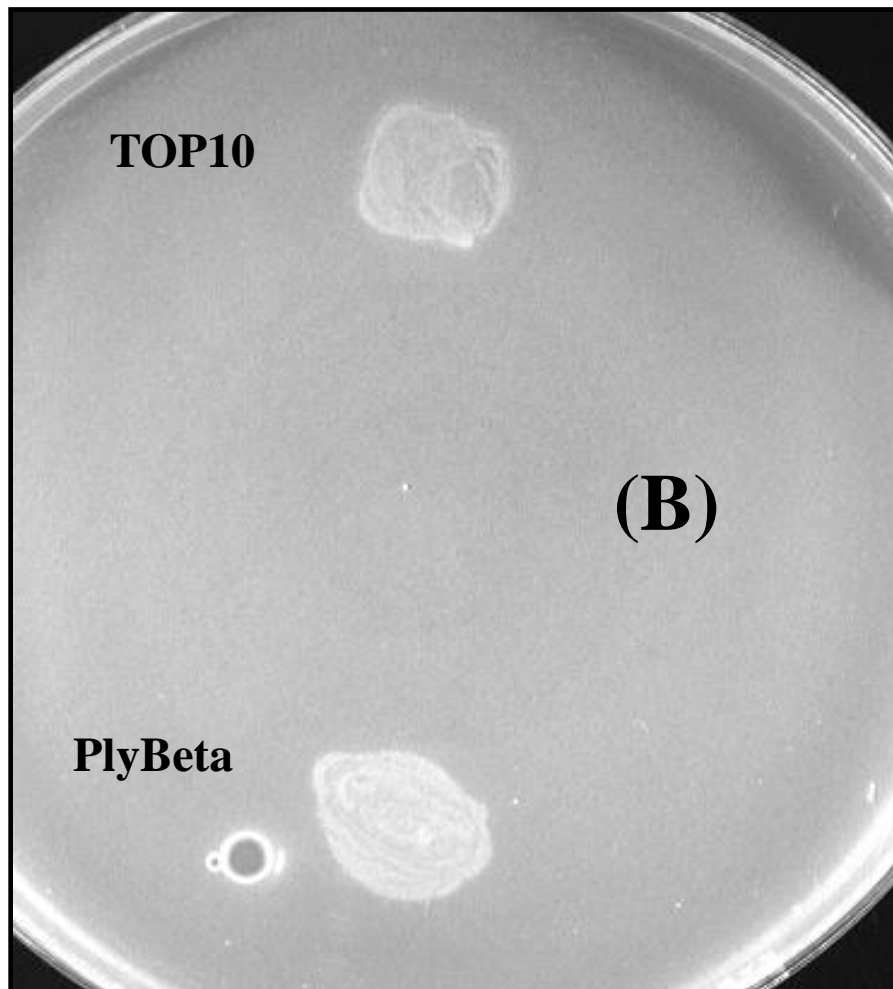
The most highly represented group is the proteobacteria (66%), with γ -proteobacteria making the largest contribution (38%). Firmicutes contributed ~13% of hits, with all other bacterial phyla each contributing less than 5%. Viruses, archaea, and the eukaryotic kingdoms likewise contributed less than 5% each.



B. thuringiensis HD73

Figure 3.5 PlyBeta Clearing Zones

In soft agar overlay experiments, PlyBeta-expressing *E. coli* clones created clearing zones for some, but not all, species/strains of *Bacillus*. Depicted here are representative examples of strain that [A] did and [B] did not demonstrate susceptibility.



B. cereus 13100

Figure 3.5, continued

BACTERIAL TAXA	Proportion of TBlastX Hits
Proteobacteria	0.663
<i>γ-Proteobacteria</i>	0.381
<i>α-Proteobacteria</i>	0.122
<i>β-Proteobacteria</i>	0.010
<i>ε/δ-Proteobacteria</i>	0.0567
Firmicutes	0.134
Actinobacteria	0.038
Bacteroides/Chlorobi	0.025
Cyanobacteria	0.021
Spirochaetes	0.019
Other Bacteria	0.020
Animal	0.032
Plant	0.011
Fungi	0.006
Protozoa	0.005
Archaea	0.015
Virus	0.011

Table 3.1a Taxonomic Distribution of Metagenomic BLAST Hits

Forty-six of 50 metagenomic sequences generated BLASTX homologies, for which the encoding organisms were noted. The proportion of individual taxa among all hits was calculated for each sequence. Listed above are the average of these values across the 46 sequences.

The top hit for each of the 46 reads was bacterial in origin (Table 3.1b), and 11 of the 46 top hits were from various species of the genus *Aeromonas*, suggesting a gut environment dominated by aeromonads.

The metagenomic library was subject to soft agar overlay screens in which 5005 permeabilized clones were each overlaid with *B. anthracis* (Δ Sterne), *P. aeruginosa* (PAO1), *S. aureus* (RN4220), and sheep's blood. The blood-agar screen was included here due to the bacterial origin of the extracted DNA, with its potential for identifying encoded hemolytic virulence factors. Indeed, while no clearing zones were observed in the bacterial screens, a zone of β -hemolysis was noted around a single clone in the blood-agar screen (see Figure 3.6a). The clone's metagenomic insert was sequenced and BlastX analysis showed it to encode a new member of the aerolysin gene family.

Aerolysin is a pore-forming exotoxin encoded by various *Aeromonas* species. The protein is secreted through the type II pathway, and – following a C-terminal cleavage event in solution or at the target cell surface – it multimerizes and inserts into eukaryotic plasma membranes (Fivas et al. 2001). When the hemolytic clone was further tested, it was found that chloroform permeabilization was not required for hemolysis. The clone had acquired a β -hemolytic phenotype when grown on blood agar (Figure 3.6a, inset), indicating the aerolysin is recognized by the *E. coli* type II secretion system (both inner and outer membrane components).

TAXA	Top Hits
<i>Aeromonas</i>	11
<i>Salmonella</i>	5
<i>Escherichia</i>	4
<i>Clostridium</i>	3
<i>Shewanella</i>	3
<i>Psuedomonas</i>	3
No Match	4
<i>Bacillus, Branchiostoma, Rhizobium, Enterococcus, Methylococcus, Salinibacter, Pseudoalteromonas Streptococcus, Bordetella, Synechocystis, Rhodopseudomonas, Oceanobacillus, Rhodopirellula, Parachlamydia, Yersinia, Nitrobacter, Psychromonas</i>	1 Each

Table 3.1b Top BLAST Hits by Genus

For each of metagenomic sequences, the encoding organism of the hit with the lowest e-value was noted. The table compiles these hits organized by genus, with *Aeromonas* being the most highly represented.

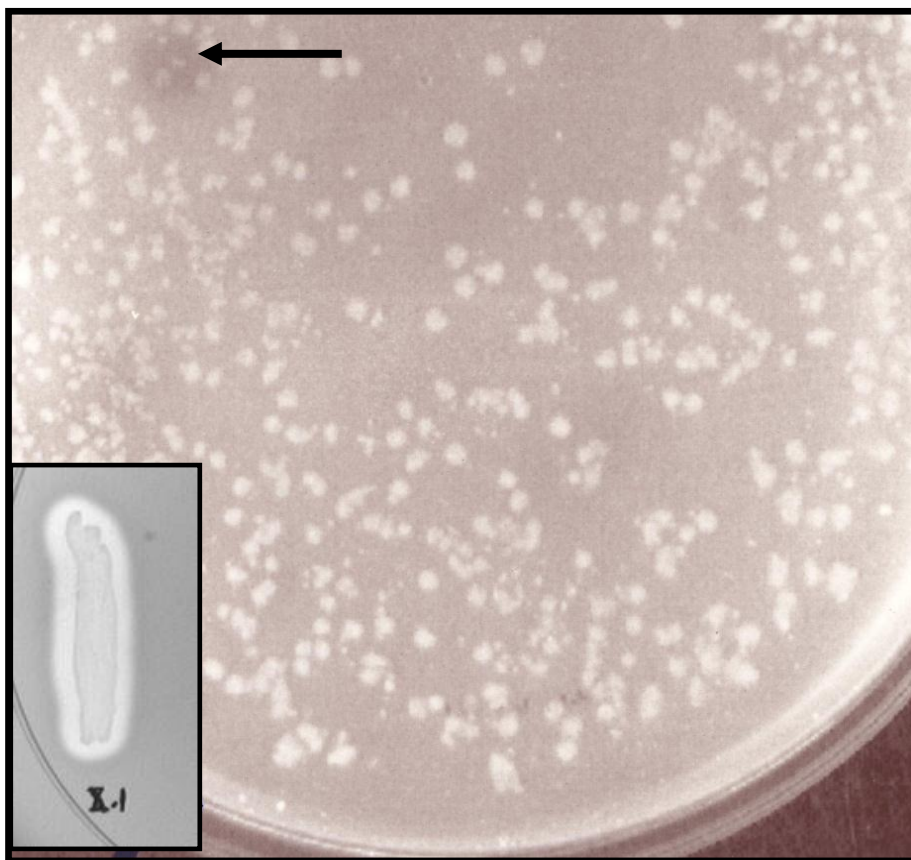


Figure 3.6a Positive Hemolysin Clone

During the metagenomic screen of worm-gut contents, a single positive clone was observed in blood agar overlay screens (indicated by arrow). The corresponding colony on the master plate was identified, amplified, and shown to contain a metagenomic DNA insert encoding a novel aerolysin (Aer_M). When propagated on LB-blood agar, this *E. coli* clone demonstrated β -hemolysis without chloroform permeabilization (depicted in the inset).

Overall, aerolysin sequences have been reported for a number of *Aeromonas* species/strains (Epple et al. 2004; Wang et al. 2003 and 1996; Fujii et al. 1998; Khan et al. 1998; Husslein et al. 1998; Hirono and Aoki 1993; Chopra et al. 1993; Hirono et al. 1992; Howard et al. 1987; Chakraborty et al. 1986; Howard and Buckley 1986). When compared to these proteins, the new aerolysin (subsequently referred to as Aer_M) demonstrated an average Clustal nucleotide alignment score of 71.1 (range: 68-78) and an average amino acid alignment score of 73.2 (range: 68-79) (all pairwise E-values < 10⁻¹²⁰). Phylogenetic analysis, moreover, showed it to occupy a relatively unique evolutionary position relative to other aerolysins (Figure 3.6b). Considering that the other aerolysins did not always group according to their species, however, one cannot surmise whether Aer_M is encoded by a different strain of one of the species listed in Figure 5b or a different aeromonad species entirely.

Aer_M was purified so that its activity could be confirmed quantitatively. In SDS-PAGE, the protein ran at a molecular weight of ~48 kDa (Figure 3.6c and d). This is the predicted mass of Aer_M following C-terminal cleavage. To verify this assumption – and not the alternate possibility of a nonspecific degradation product of coincidental molecular weight – Edman sequencing was conducted. The N-terminus was shown to begin at residue A24 (immediately following the predicted position of signal peptide cleavage), implying the presence of the final processed protein (Figure 3.6e).

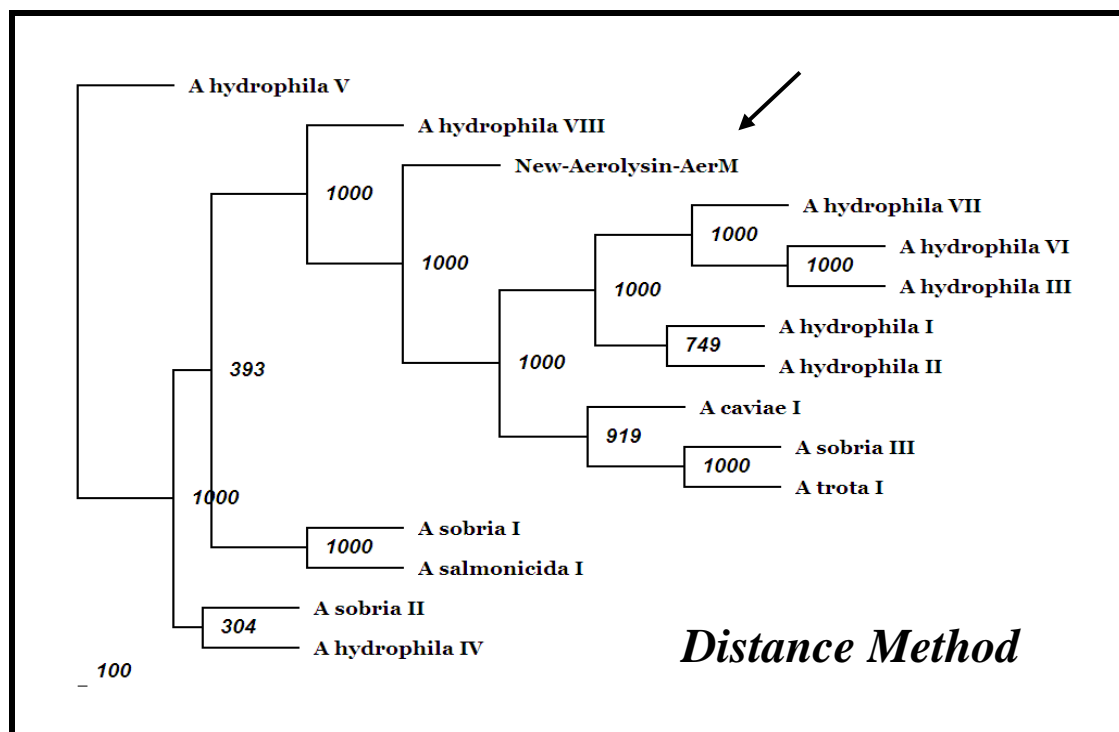


Figure 3.6b Aerolysin Phylogeny

Multiple sequence alignment was conducted with Aer_M and all other non-redundant aerolysin sequences in the NCBI database: *A. hydrophila* I (GeneBank accession number M16495); *A. hydrophila* II (X65044); *A. hydrophila* III (X65045); *A. hydrophila* IV (AY611033); *A. hydrophila* V (AF41110466); *A. hydrophila* VI (M84709); *A. hydrophila* VII (DQ40826); *A. hydrophila* VIII (AY378303); *A. sobria* I (X65046); *A. sobria* II (AY157998); *A. sobria* III (Y00559); *A. salmonicida* I (X65048); *A. trota* I (AF064068); *A. caviae* (AAC44637).

Phylograms constructed from these alignments (using DNA *parsimony* and *distance* methods, *distance* shown here) indicate that Aer_M occupies a relatively unique evolutionary position relative to the other aerolysins. Bootstrap consensus values (out of 1000) are indicated at their respective nodes.

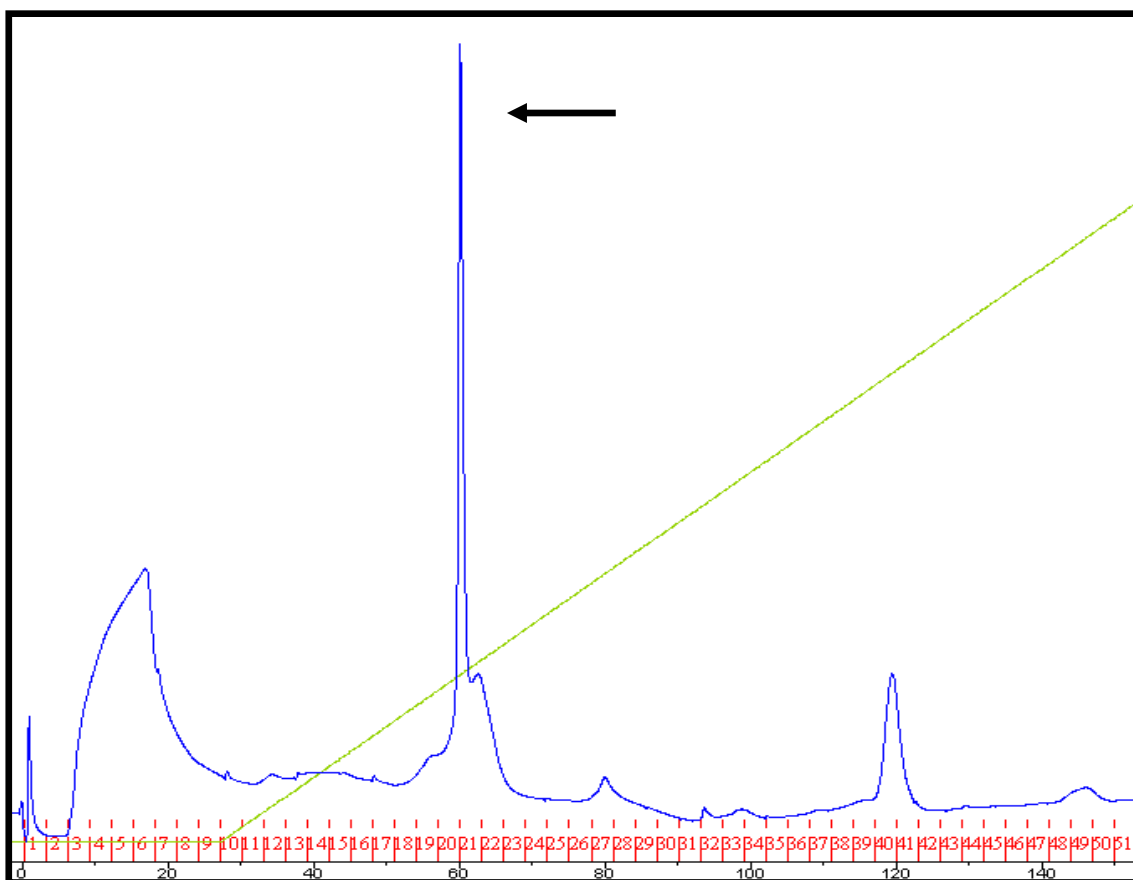


Figure 3.6c Purification of Aer_M

Aer_M was secreted by the encoding *E. coli* into the culture supernatant. As a result, the cell-cleared media was ammonium sulfate-precipitated and the total extracellular protein fraction was collected, resuspended, and subject to Mono-Q anion exchange chromatography. The peak corresponding to Aer_M (as determined by drop tests onto blood agar) is indicated with an arrow.

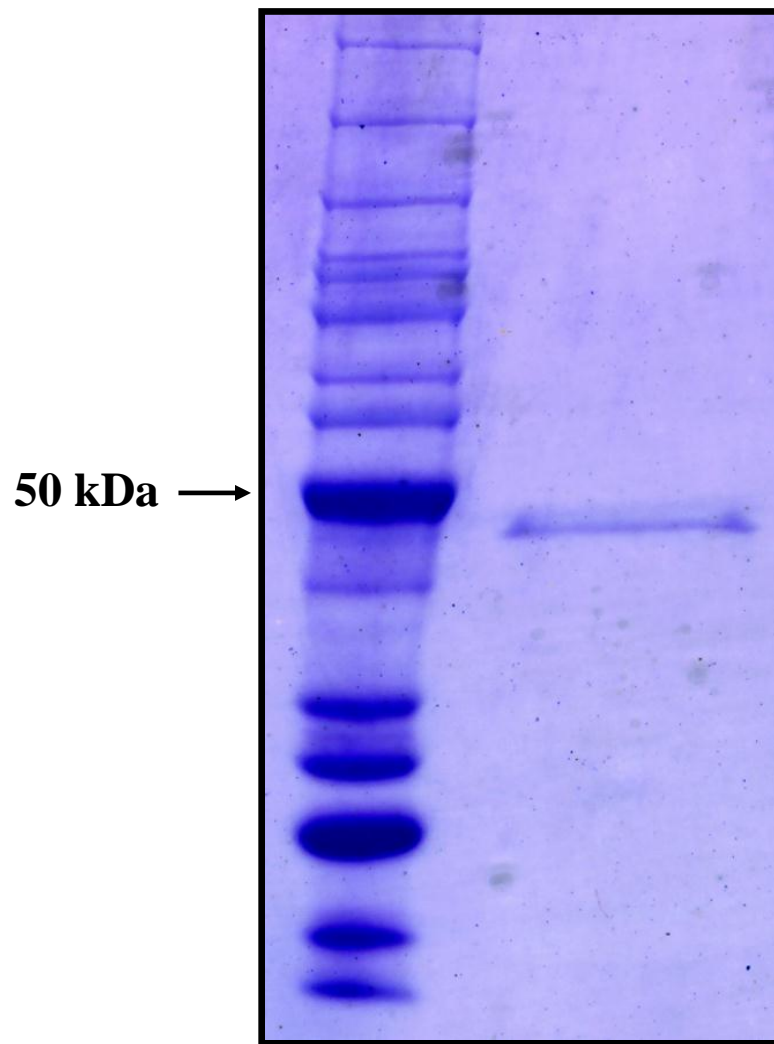


Figure 3.6d Purified Aer_M

In SDS-PAGE, Aer_M ran at a molecular mass slightly below 50 kDa. The predicted molecular mass of the initial translated protein (preproaerolysin) is 54.5 kDa. Following inner membrane translocation and N-terminal cleavage, the protein (proaerolysin) has a predicted mass of 52.0 kDa. The final, processed protein (following C-terminal cleavage) has a predicted mass of 47.5 kDa. The mass of the purified protein above suggests that it was isolated in this form.

Raw Data Tabulation ... (in pmols)																			
	D	N	S	Q	T	G	E	H	A	R	Y	P	M	V	W	F	I	L	L
1	9.5	18.5	17.0	1.4	10.2	13.6	7.3	4.2	35.8	7.4	4.1	2.5	4.1	12.4	0.0	7.3	6.8	5.4	
2	9.5	7.0	11.4	10.1	6.9	9.5	23.1	2.0	12.1	3.4	3.2	5.1	3.3	19.7	0.0	4.6	11.0	5.5	
3	8.8	4.6	10.8	7.8	6.0	11.2	14.7	1.5	8.9	6.4	3.3	25.5	1.7	7.8	0.0	2.5	6.7	3.3	
4	4.7	4.8	10.5	5.9	4.0	11.4	12.8	1.5	5.9	2.7	4.1	15.3	1.0	17.8	0.0	4.0	5.1	4.2	
5	4.2	5.3	8.9	6.0	4.2	7.7	13.0	1.4	7.0	2.6	13.4	12.4	0.9	11.4	0.0	3.0	5.4	4.5	
6	9.1	4.3	6.6	7.7	5.0	6.6	9.1	1.4	6.1	2.3	6.6	14.4	1.0	8.0	0.0	4.2	8.8	4.1	
7	11.2	4.2	7.8	6.7	4.3	6.3	7.0	1.0	6.6	2.1	3.9	11.3	0.9	8.1	0.0	3.0	6.5	4.0	
8	10.5	3.8	5.7	11.7	6.2	5.9	8.1	1.0	5.5	2.4	3.0	10.6	2.5	5.8	0.0	2.7	5.1	3.6	
9	5.1	3.4	5.1	8.4	5.2	5.7	5.6	0.9	5.5	2.6	3.1	8.4	1.1	5.8	0.0	2.4	4.9	7.4	
10	3.6	3.4	4.2	6.2	5.0	5.4	4.7	0.9	5.3	7.4	2.7	7.4	1.1	5.1	0.0	2.8	4.0	5.9	
11																			

Major	Minor
A	
E/V	G/P/I
P	G/R
V	
Y	
P	D/Q/I
D	
Q	
L	
R	

Figure 3.6e Edman Sequencing of Aer_M

To confirm the that Aer_M was recognized and secreted by the E. coli sec-mediated pathway – and to verify the predicted position of signal peptide cleavage – the purified protein was subject to N-terminal sequencing by the Rockefeller Proteomics Facility. Indeed, both were confirmed, as the predicted N-terminal position of purified Aer_M was A24.

The activity of Aer_M against various erythrocytes was tested with an *in vitro* hemolysis assay. While the toxin showed activity against all cells tested (sheep, rabbit, chicken, human), the specific sensitivity varied among the species (Figure 3.6f). Under the assay conditions (10⁸ RBC/ml in PBS, 1 hr incubation, 37°C), the erythrocytes could be ranked by increasing sensitivity as follows: sheep (complete hemolysis at 200 nM) < human < chicken < rabbit (complete hemolysis at 5 nM). The activity of Aer_M was also confirmed against an actively dividing mammalian cell line, Detroit 562 (human pharyngeal carcinoma). When added to a 75% confluent culture, 2 nM Aer_M induced near total cellular dissociation with loss of viability by 48 hr (images not shown). Boiling Aer_M prior to its addition to culture media abolished the cytotoxic affect, indicating that the cellular response was attributable to the protein itself and not residual endotoxin contamination from the purification protocol.

DISCUSSION

We describe here a novel technique for screening genomic or metagenomic DNA that utilizes a combination of PCR amplification and topoisomerase cloning. This expression-based approach has several strengths that make it a useful tool in the search for metagenomic proteins of interest. Library construction is rapid and straightforward, and only nanogram quantities of DNA are required.

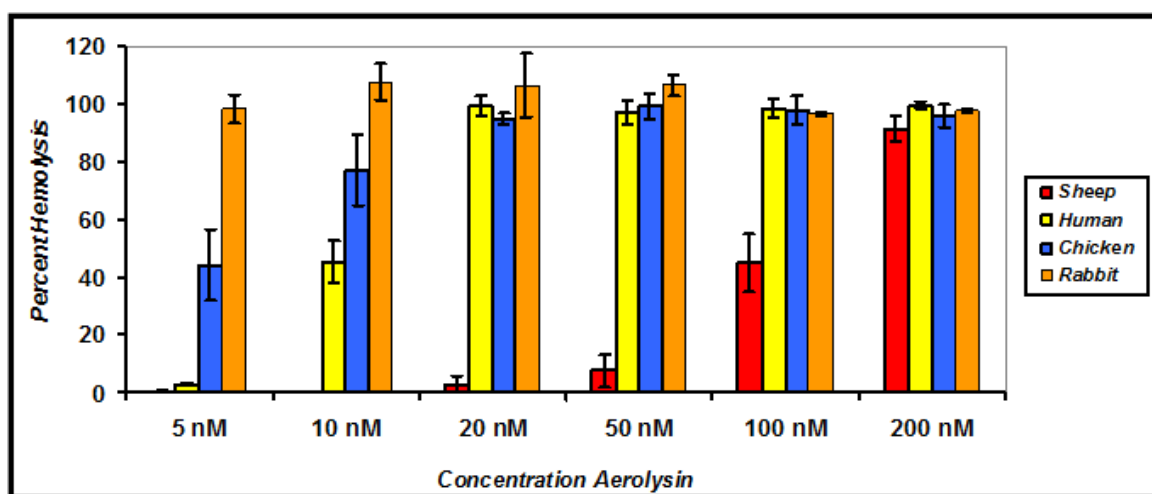


Figure 3.6f Aer_M-Induced Hemolysis of RBCs.

Erythrocytes from several species all showed susceptibility to Aer_M in a quantitative *in vitro* hemolysis assay. Rabbit erythrocytes demonstrated the highest sensitivity, sheep erythrocytes demonstrated the lowest sensitivity, while human and chicken erythrocytes demonstrated intermediate sensitivity. For each measurement, n = 6, except human (n = 3). Several hemolysis measurements extend slightly above the 100% value. This is merely an artifact of the experimental standards; in these cases, Aer_M-induced hemolysis exceeded hypotonic hemolysis (defined as 100%).

Moreover, the use of commercially-available cloning products makes the technique desirable for laboratories that are entering the field of metagenomics and might lack dedicated expression systems. Overall, the great potential of metagenomics is attracting much attention, so the development of simple, functional screens could broaden accessibility to an already rapidly growing science.

A particularly well-suited application for the technique is the functional screening of viral metagenomes. Viral metagenomics is an emerging area in which bacteriophage particles (or other viruses) are purified from environmental samples prior to library construction. Given that bacteriophages are likely the most abundant genetic entities on Earth (Hatfull 2008), metagenomics provides a powerful tool for investigating their ecology and diversity. Previous viral metagenomic studies have utilized linker amplification, as phage DNA is relatively scarce relative to the total metagenomic pool. Nevertheless, this work has been entirely sequence-based in nature (Angly et al. 2007; Culley et al. 2006; Zhang et al. 2006; Breitbart et al. 2002 and 2003). Functional viral metagenomics could provide an important complement, especially in the search for antibacterial proteins (e.g. phage lysins) or ones involved in bacterial pathogenesis (e.g. phage-encoded virulence factors).

The E-LASL approach is likewise applicable to functional screening of bacterial metagenomes (such as the worm-gut contents screened here). Although used here to identify a new hemolysin, the method is equally applicable in

mining for other classes of active proteins. By simply changing the readout of the soft agar overlay step, for instance, one could search for various bacteria-encoded antibiotics or biocatalysts. In the case of antibacterial proteins, the screens could target cell wall remodeling enzymes (autolysins), peptide antibacterial compounds (bacteriocins), or lysins from integrated prophage. Given the size of bacterial genomes, such genes would comprise a smaller proportion of the total DNA and would likely necessitate that a greater number of clones be screened to generate hits. It is therefore not surprising that the relatively small metagenomic screen conducted here against *B. anthracis*, *S. aureus*, and *P. aeruginosa* did not identify any antibacterial proteins.

One should note that the E-LASL method admittedly has several limitations that must be taken into consideration. Like all functional metagenomic screens, there is always the possibility that foreign proteins may misfold, not express, or prove toxic to the host species. Moreover, the technique would not be expected to identify secondary metabolites from environmental organisms, as the DNA inserts are too short. Ultimately, the insert-length cannot exceed the substrate length limit of the polymerase itself. Considering current polymerase technology, it is theoretically possible to construct an E-LASL with inserts in the range of tens of kilobases (i.e. for use in cosmids). At the same time, commercially-available systems do not yet exist that would make such libraries compatible with topoisomerase cloning.

It is important to mention that there are several other ways in which the method presented here could be adapted to fit specific situations. Although Tsp509I was employed here to fragment DNA, other restriction enzymes exist with suitable 4-bp consensus sequences. For instance, one might employ BfuCI (/GATC) or MspI (C/CGG), especially if dealing with a more GC-rich organism or environmental sample. Enzymatic digestion of DNA could even be omitted entirely in favor of mechanical fragmentation, although this would create a need for overhang digestion and blunt-ended ligation (as in previous studies, Breitbart et al. 2002 and 2003). We also envision a protocol in which RNA is extracted in bulk from an uncultured sample. An E-LASL could then be constructed from reverse-transcribed metagenomic cDNA.

In the current study, several functional proteins were identified through the E-LASL technique. Although they are presented more as proof-of-principal, these proteins raise several issues that, in themselves, merit further discussion. For instance, the *B. anthracis* lysins cloned here further expand the list of potential enzybiotics that have been identified for this important bioterrorism-related pathogen (Porter et al. 2008; Yoong et al. 2006; Low et al. 2005; Schuch et al. 2002). Moreover, the Aer_M protein represents the first example of a hemolytic virulence factor characterized through metagenomics. Although hemolysis of blood agar is a classic method of identifying bacterial species, surprisingly few metagenomic screens have focused on an acquired hemolytic phenotype.

To our knowledge, the only previous examples are the studies of Rondon et al. (2000) and Gillespie et al. (2002). In the former, the authors identified a number of hemolytic BAC clones from a soil-derived metagenomic sample. This screen was performed as proof-of-principle of their screening technique, however, and the responsible molecules were not identified. In Gillespie et al. (a follow-up of the previous paper), one of these clones was investigated further and was demonstrated to encode a biosynthetic enzyme leading to the production of the small molecule turbomycin.

Admittedly, Aer_M is not a protein whose cloning would have absolutely required a metagenomic approach. Aeromonads are generally culturable, so the Aer_M-encoding species presumably could have been isolated from the worm gut first, and the aerolysin cloned through genomic techniques. Nevertheless, this approach would have required considerably more effort (as well as the specific original intent of looking at this genus). In general, environmental virulence-factor screening represents an interesting avenue for future metagenomic screening. Given the role they play in pathogenesis and inter-species competition, the identification of such proteins could prove valuable from an ecological or infectious disease perspective.

In summary, we have presented here a novel method for constructing genomic and metagenomic expression libraries from small initial quantities of extracted DNA. This plasmid-based approach utilizes linker-amplification and topoisomerase cloning, and provides a rapid, technically straightforward, and

adaptable means of functional screening in an *E. coli* host. Overall, the E-LASL approach may be utilized to mine uncultured bacterial or viral populations for novel antibiotics, as well as other protein compounds with pharmacological, industrial, or pathogenic potential.

ACKNOWLEDGMENTS

This research was funded by NIH/NIAID grants AI057472 and AI056510 to VAF. JES acknowledges the support of the Pharmaceutical Research and Manufacturers of America Foundation and the NIH MSTP program (Weill Cornell/Rockefeller/Sloan-Kettering grant GM 07739). JES would like to thank Dr. Daniel Nelson for his assistance with protein purification and Ms. Clara Eastbay for her help with cell culture. We thank Dr. Forrest Rohwer of San Diego State University for his discussions and helpful advice on the topic of linker amplification in metagenomics. We also acknowledge Dr. Joseph Fernandez of the Rockefeller University Proteomics Facility for performing the Edman sequencing included here.

ENDNOTES

1. The distinction between sequenced-based and functional metagenomics is blurred somewhat by certain publications that refer to the homology-based prospecting for desired enzymes as *functional* (Schoenfeld et al. 2010). In the current text, however, the term *functional* is reserved for recombinant screens, and

not sequence-based searches that subsequently lead to PCR-cloning and functional analysis (as in Chapter 2).

2. The numbers regarding transformation efficiency are those that were originally reported in Schmitz et al. (2008). They were derived from the E-LASLs for phages BG-1, BG-2, BG-3, BG-4, and the worm-gut metagenome. Colonies were not counted when the phage-genomic E-LASLs were used in combination with the holin-based hemolysis screen. The latter screens, in fact, transpired after this data was originally published. They represent the preliminary work that led to the development of screening technique itself (see Chapter 4). The information is included in this chapter so that Chapter 4 can best reflect the format that appeared in Schmitz et al. (2010b).

3. This was true at the time of original publication in 2008. Since that time, genomic sequencing of numerous *Bacillus* strains has revealed various other lysin-genes with an N-terminal GH-25 domain and a C-terminal SH3-5 domain. PlyB and PlyBeta remain each other's closest known homologue, but their overall architecture can no longer be considered unique among *Bacillus*. To date, none of these other enzymes has been studied recombinantly.

****NOTE:** In the published manuscript based on this chapter (Schmitz et al. 2008), the following individuals were included as co-authors (in the stated order): Anu Daniel, Mattias Collin, Raymond Schuch, and Vincent Fischetti.

CHAPTER 4

Functional Viral Metagenomics II: A Novel Phage Lysin Screen

INTRODUCTION

The field of *metagenomics* has expanded rapidly in recent years, providing access to environmental microorganisms that would remain unapproachable by standard, culture-based methods. The foundation of metagenomics lies in the direct extraction of DNA/RNA from environmental samples (e.g. soil, water, or feces), without prior isolation of individual microbial species (reviewed in Green

and Keller 2006, Tringe and Rubin 2005). It has been estimated that only a small proportion of naturally occurring microbes—approximately 1% of soil bacteria, for instance—are culturable under standard laboratory conditions (Torsvik and Ovreas 2002). In this light, metagenomics has become an increasingly common tool for studying diverse ecosystems, from around the globe to within the human body.

Overall, metagenomics research can be divided into two general categories: sequence-based and functional. In the former, environmental DNA is sequenced in mass and compared with genetic databases to address broad questions of ecology, taxonomy, and diversity. Some of the most extensive metagenomic studies to date have been sequence-based in nature, benefiting greatly from the development of high-throughput sequencing technologies. Notable examples include a 76-megabase study of an acid-mine biofilm (Tyson et al. 2004), a 1-gigabase analysis of the Sargasso Sea (Venter et al. 2004), and a 6.3-gigabase sampling of global oceanic samples (Rusch et al. 2007).

In functional metagenomics, by contrast, environmental genes are recombinantly expressed within a host organism, which is monitored for the acquisition of a desired phenotype. Rather than provide insight into entire ecosystems, functional studies aim to identify individual molecules with biomedical or industrial value. Such compounds may either be proteins (encoded directly by environmental genes) or small molecules (synthesized by several enzymes of a gene cluster). Numerous classes of molecules have been

identified to date, with particular interest in the areas of biosynthesis, biomass degradation, and antibiotic discovery (reviewed in Brady et al. 2009, Uchiyama and Miyazaki 2009, Voget et al. 2005).

While bacteria provide the majority of DNA to most metagenomic pools, recent studies have begun focusing on subsets of total environmental populations. A prominent example is *viral metagenomics*, in which viral particles (predominately bacteriophage) are purified from cellular material prior to DNA extraction (reviewed in Delwart 2007, Edwards and Rohwer 2005). Although the yield of DNA from environmental phage isolates is generally low, PCR-amplification techniques have been developed to overcome this issue (Breitbart et al. 2002, Schmitz et al. 2008). Viral metagenomic analyses have been conducted on a growing number of samples, including ones purified from soil (Fierer et al. 2007), sea-water (Breitbart et al. 2002, Williamson et al. 2008), and human feces (Breitbart et al. 2003).

These studies have revealed a remarkable abundance of novel sequences, supporting the notion that phage represent the largest source of untapped genetic diversity on the planet (Hatfull 2008). Despite this wealth of information, however, viral metagenomic studies to date have remained predominantly sequence-based in nature. In this regard, functional screens of viral metagenomes could provide a large source of recombinant molecules.

Recently, one class of phage-encoded protein has received particular attention from the biotechnology field: phage lytic enzymes (also referred to as

endolysins or lysins) (reviewed in Fischetti 2005, Fischetti et al. 2006). These peptidoglycan hydrolases are expressed late in the infective cycle of double-stranded DNA phage, and—along with a membrane permeablizing protein known as a holin—they are responsible for disrupting the bacterial cell envelope and freeing progeny viral particles.

Despite this conserved biological function, phage lysins (especially Gram-positive ones) are a tremendously diverse group of proteins whose enzymatic specificity includes various bonds within the peptidoglycan macromolecule. They include glycosyl hydrolases that target the polysaccharide backbone (muramidases/lysozymes and glucosaminidases), alanine-amidases that target the initial L-alanine of the pentapeptide stem, and endopeptidases that target subsequent peptide bonds in the stem or cross-bridge. While lysins of Gram-negative phage generally consist of an enzymatic domain alone, Gram-positive lysins are modular and combine an N-terminal lytic domain with a C-terminal binding domain that can recognize various epitopes within the target cell envelope.

Although researchers have known of lysins for decades, interest has increased markedly in recent years after it was proposed that they could act as novel anti-infective agents against Gram-positive pathogens, whose peptidoglycan is directly accessible from the extracellular space (for instance in Nelson et al. 2001, Schuch et al. 2002, Cheng et al. 2005). A growing number of *in vitro* and *in vivo* studies have confirmed the ability of recombinantly expressed

lysins to kill such organisms, and their appeal lies in both the potency and the specificity they demonstrate toward individual Gram-positive species. This *enzymatic* value of phage lysins goes alongside additional proposed applications in the areas of food (Deutsch et al. 2004), agricultural (Kim et al. 2004), veterinary (Celia et al. 2008), and industrial science (Matsushita and Yanase 2008, Ye and Zhang 2008).

Considering this potential, lytic enzymes represent an intriguing functional target for viral metagenomic screens. At the same time, identifying lysins in this manner would present several distinct challenges. Aside from general concerns common to all functional screens (e.g. protein expression and solubility), metagenomic lysin identification would face the following particular issues.

(I) *Clonal toxicity*: Recombinant lysin expression is typically well-tolerated by host bacteria, as the enzymes are sequestered in the cytoplasm away from the peptidoglycan layer. Holins, on the other hand, interact nonspecifically with plasma membranes and are generally toxic to an *E. coli* host, inducing bacteriolysis from within (9). Since holins are short (~100 residues) and often encoded adjacent to lysins, they can lead to selective toxicity of many of the clones one hopes to identify. In a metagenomic screen, where numerous lysins are present within a single library, this effect could lead to a significant loss of positive hits. (II) *Target bacterial species*: In standard phage genomic screens, lysin-encoding clones are selected by their ability to kill the host bacterium of the

encoding phage, which generally demonstrates the highest sensitivity. In a metagenomic screen, however, numerous host species of unknown origin could be present, confounding this choice of screening agent.

To address these issues, we have devised a novel functional strategy for the general cloning of lytic enzymes from uncultured phage DNA. It utilizes a plasmid-based *E. coli* expression system and consists of a two-step process. Following induction by nebulized arabinose, clones are first screened for holin-mediated lysis by a hemolytic effect they create in the surrounding blood agar. These initial hits are then restreaked as patches and overlaid with Gram-negative cells whose outer membranes have been permeabilized by autoclaving, serving as a general source of peptidoglycan. The clones are observed for surrounding Gram-negative clearing zones to assay directly for the recombinant production of lytic enzymes encoded adjacent to the holins.

As proof-of-principle, we applied our methodology to a viral metagenomic library constructed from mixed animal feces, identifying 26 actively-expressed lysins of diverse molecular architectures. The first of its kind, this study presents a general model for lysin identification through viral metagenomics, highlighting the potential of this field for cloning proteins of biotechnological or academic value.

MATERIALS AND METHODS

DNA library construction. Fecal specimens were collected at the Long Island Game Farm Wildlife Park and Children's Zoo (Manorville, New York) from the following species: giraffe, zebra, donkey, domestic goat, llama, lion, and bison. Additionally, two dried fecal specimens were obtained from commercial sources: bat guano (Fox Farm Soil and Fertilizer Company) and cricket droppings (www.cricketpoo.com).

Viral fractions were purified by an adaptation of the procedure of Casas and Rohwer (2007). In summary, fecal samples (~100 g each) were suspended in an equal volume of PBS (pH 7.4) and agitated overnight at 4°C. Particulate/cellular material was removed by centrifugation, followed by two passages through a 0.22-micron filter. Phages were precipitated by addition of polyethylene glycol 10,000 MW (10% w/v). Centrifuged precipitates were pooled to form a collective phage library, which was subject to phenol/chloroform extraction and ethanol precipitation. Phage DNA was separated from co-precipitated compounds by agarose gel electrophoresis and extraction of high molecular weight DNA. From this material, an expressible linker amplified shotgun library (E-LASL) was constructed, as previously described (see Chapter 3).

Lysin screening methodology. Amplified metagenomic inserts were ligated into the arabinose-inducible pBAD plasmid using the TOPO-TA Expression Kit

(Invitrogen). Transformed clones were initially plated onto LB-agar supplemented with 100 µg/ml ampicillin and 5% defibrinated sheep's blood. Following overnight growth at 37°C, clones were subject to an initial toxicity screen to isolate potential holin-encoding clones. Plates were sealed in a container into which a nebulized solution of aqueous arabinose (20% w/v) was pumped for 1 hr. Following induction, the plates were returned to 37°C and observed for colony lysis, as indicated by the development of a zone of hemolysis in the surrounding blood agar. Hits were identified over the subsequent 6-8 hr period, as nonspecific blood-agar oxidation (i.e. alpha-hemolysis) would often appear around colonies at longer times (~16 hrs). Chosen clones were streaked onto separate LB-ampicillin plates (lacking arabinose) and allowed to repropagate without induced expression.

For the secondary (i.e. lysin-targeting) screen, the above hits were streaked as patches onto LB-ampicillin plates supplemented with 0.2% arabinose. Following overnight expression at 37°C, cells were exposed to chloroform vapor (15 min) to kill and permeabilize any still-viable *E. coli* (by this point, many cells had already undergone lysis due to holin expression). The patches were then overlaid with molten soft agar containing autoclaved *P. aeruginosa* (see below), and observed for clearing zones for up to 24 hr. For all lytic clones, the encoded metagenomic insert was sequenced with plasmid-targeted primers (Genewiz; South Plainfield, NJ). When primer-walking was required to sequence an insert in entirety, appropriate oligonucleotides were designed and ordered (Operon).

Preparation of Gram-negative overlay. Gram-negative soft agar was prepared as follows: *P. aeruginosa* strain PAO1 was grown to stationary density in Brain-Heart Infusion media. Cells were pelleted and resuspended in PBS pH 7.4 to half the volume of the original liquid culture. Agar was added directly to this suspension (7.5 g/L), which was autoclaved for 15 min at 122°C, 15 psi. Solidified aliquots were stored at 4°C until the time of use, at which point they were melted and equilibrated at 55°C. For a single 150-mm Petri dish, 15 ml of soft agar was overlaid.

Computational analysis. Protein sequences of the cloned lysins were subject to BlastP analysis to identify known homologues among defined organisms (blast.ncbi.nlm.nih.gov/Blast.cgi). As a reference database, the NCBI non-redundant sequence collection (nr) was utilized. Putative catalytic and binding domains were assigned via Pfam v24.0 (Finn et al. 2010; pfam.sanger.ac.uk). Multiple sequence alignment of cloned lysins (amino acid sequence) was conducted with the ClustalX algorithm (Thompson et al. 1997). From this, a phylogenetic tree was constructed with the PHYLIP v3.67 software package (Felsenstein 1989) using the ProtDist and Kitch programs (default settings).

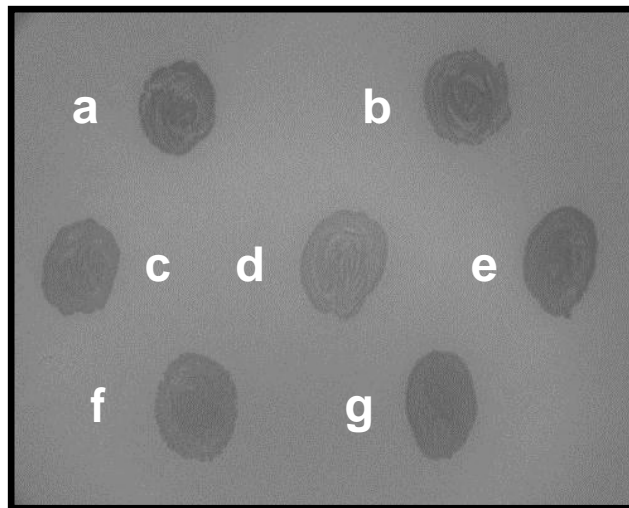
RESULTS

A plasmid-based shotgun library was constructed with pooled DNA extracted from the phage fraction of multiple animal fecal samples. In order to

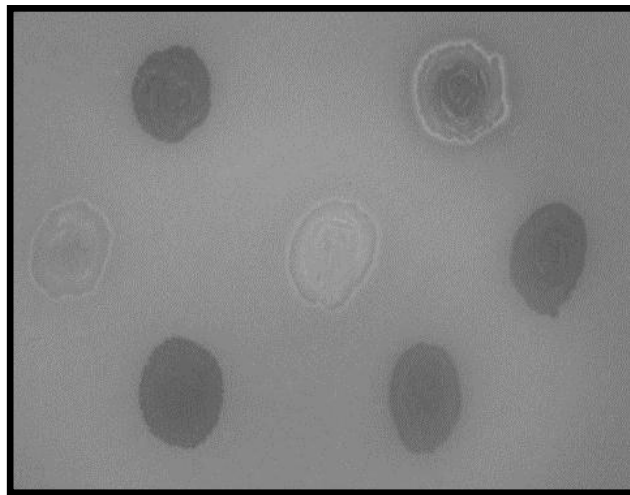
clone lytic enzymes from this metagenomic pool, we first sought to address the issue of holin-based clonal toxicity. Rather than attempt to mitigate the effect, we instead chose to exploit it by *selecting for* toxic clones. In the literature, there exists one example in which a lysin-encoding clone was identified through the lytic activity of its adjacent holin. When screening a genomic library from *Actinomyces naeslundii* phage AV-1, Delisle et al. utilized a “plasmid release” protocol in which mixed *E. coli* transformants were grown in a single liquid culture (2006). Holin-encoding cells would undergo lysis following induced expression, releasing their plasmids into the culture media, from which they could be purified and characterized. Their approach was adapted here for metagenomic libraries, in which numerous targeted clones are present within a single library.

Following transformation, *E. coli* clones were plated onto agar media supplemented with 5% sheep’s blood, but lacking arabinose inducing agent. The latter was supplied via a nebulized mist only once the clones had proliferated to visible colonies. In the hours following arabinose induction, colonies undergoing lysis could be visualized by the appearance of subtle, yet definitive, hemolytic zones in the surrounding blood agar (Figure 4.1). This effect was often accompanied by the development of a viscous colony-phenotype: when a pipette tip was touched to a hemolytic colony and lifted up, the bacterial mass would adhere in a string-like manner between the tip and the agar surface.

Pre-Induction



**6 hr
Post-
Induction**



**24 hr
Post-
Induction**

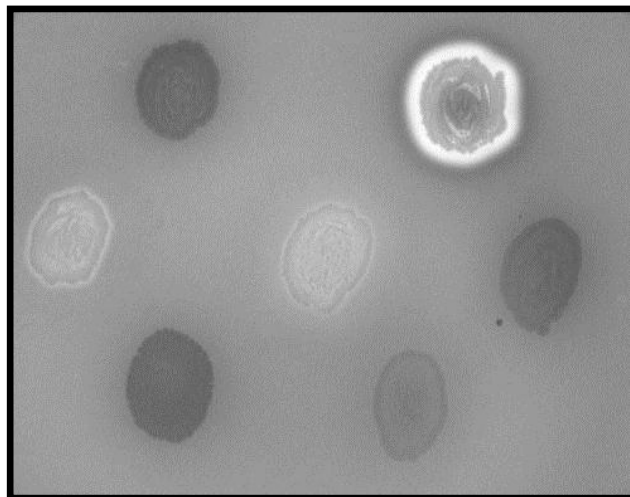


Figure 4.1 Preliminary Hemolysis Screen

Figure 4.1, continued

Depicted here are seven *E. coli* Top-10 clones on LB-ampicillin supplemented with 5% sheep's blood. Each clone is transformed with a pBAD plasmid containing a different protein-encoding insert. A-Top left: empty plasmid; B-Top right: aerolysin, a hemolytic bacterial exotoxin encoded by the genus *Aeromonas*; C-Center left: *Bacillus anthracis* phage alanine-amidase with adjacent holin; D-Center middle: lysin PlyM14 from this study with adjacent holin; E-Center right: arbitrary metagenomic insert from this study; F-Bottom left: *Bacillus anthracis* phage alanine-amidase without adjacent holin; G-Bottom right: *Bacillus anthracis* phage muramidase without adjacent holin. Prior to induction with nebulized arabinose, the clones are indistinguishable against the blood agar background. Following induction (apparent at 6 hrs and more prominent at 24 hrs), a zone of hemolysis is present around the clones encoding holin/lysin combinations (C and D). No such zones are present surrounding the clones encoding empty plasmid (A), random sequence (E), or lysin only (F and G). The aerolysin-encoding clone (B) is included as a positive control of a dedicated hemolytic transformant.

Presumably, both the hemolysis and the viscosity were due to the release of intracellular contents from the lysing colonies. The hemolytic effect could likewise be attributable to interactions between the expressed holins and the blood-agar erythrocytes, as holins have been shown capable of permeablizing eukaryotic membranes *in vitro* (Agu et al. 2007).

Approximately 200,000 clones were screened in this manner, and 502 preliminary hits were identified and repropagated on media lacking inducing agent. To confirm the identity of the cloned inserts, 52 of these hits were subject to DNA sequencing. Forty-one unique clones were observed in this analysis—of these, 17 contained ORFs that encoded both a complete holin and a complete lysin gene (as determined through Blast-homology analysis); 4 encoded a complete holin with only a partial (i.e. truncated) lysin; 2 encoded complete holins without any recognizable lysin; and 18 did not demonstrate homology to either holins or lysins. The latter 18 inserts encoded ORFs with a variety of predicted functions, including a number phage structural proteins and DNA-interacting proteins. For 5 inserts, Blast analysis did not reveal any putative genes of known function.

As these sequencing results indicate, the hemolysis screen was not absolutely specific for holin/lysin cassettes, although this finding is not unexpected. In theory, the technique could select for any toxic proteins that compromise the viability (and, resultantly, the envelope integrity) of the host *E. coli*. Moreover, even among those clones encoding putative lysins, this initial

screen does not reveal which ones express soluble, active enzymes in the recombinant system. A secondary screen was thus necessary to specifically identify these clones.

For this step, we exploited the cell-envelope properties of Gram-negative bacteria. To explain, viable Gram-negative cells are generally resistant to exogenous lysin treatment, as their peptidoglycan layer is surrounded by the lipophilic outer membrane. Once this membrane is compromised, however, they become highly sensitized to enzyme action. This includes lysis by both non-specific eukaryotic lysozymes (forming the basis of many commercially-available extraction kits), as well as by phage lytic enzymes from viruses that infect other bacterial species (Briers et al. 2007). This sensitivity is attributable to the thinness of the Gram-negative peptidoglycan layer, and it is reflected in the very structure of Gram-negative phage lysins. As mentioned above, Gram-positive lysins possess both an N-terminal enzymatic domain and a C-terminal binding domain (which leads to target specificity), while Gram-negative lysins are classically comprised of an enzymatic domain alone (Fischetti 2005).

Consistent with these properties, we have previously observed (unpublished findings) that clearing zones appear in soft-agar experiments in which autoclaved Gram-negative cells are exposed to any number of diverse Gram-positive lysins with muramidase or alanine-amidase activity (Figure 4.2a).

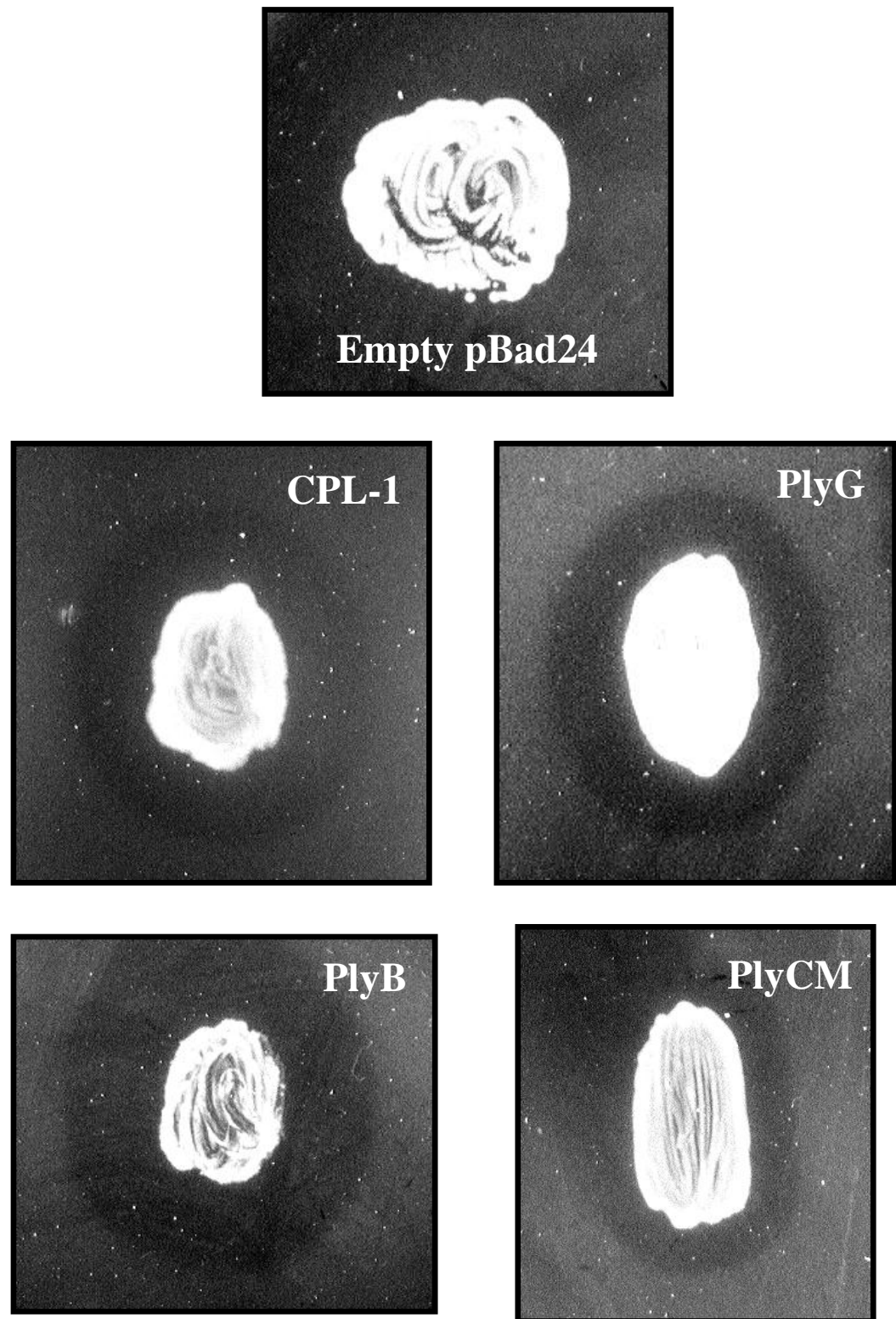


Figure 4.2a Lysin-Encoding Clones with
Pseudomonas Overlay (Genomic)

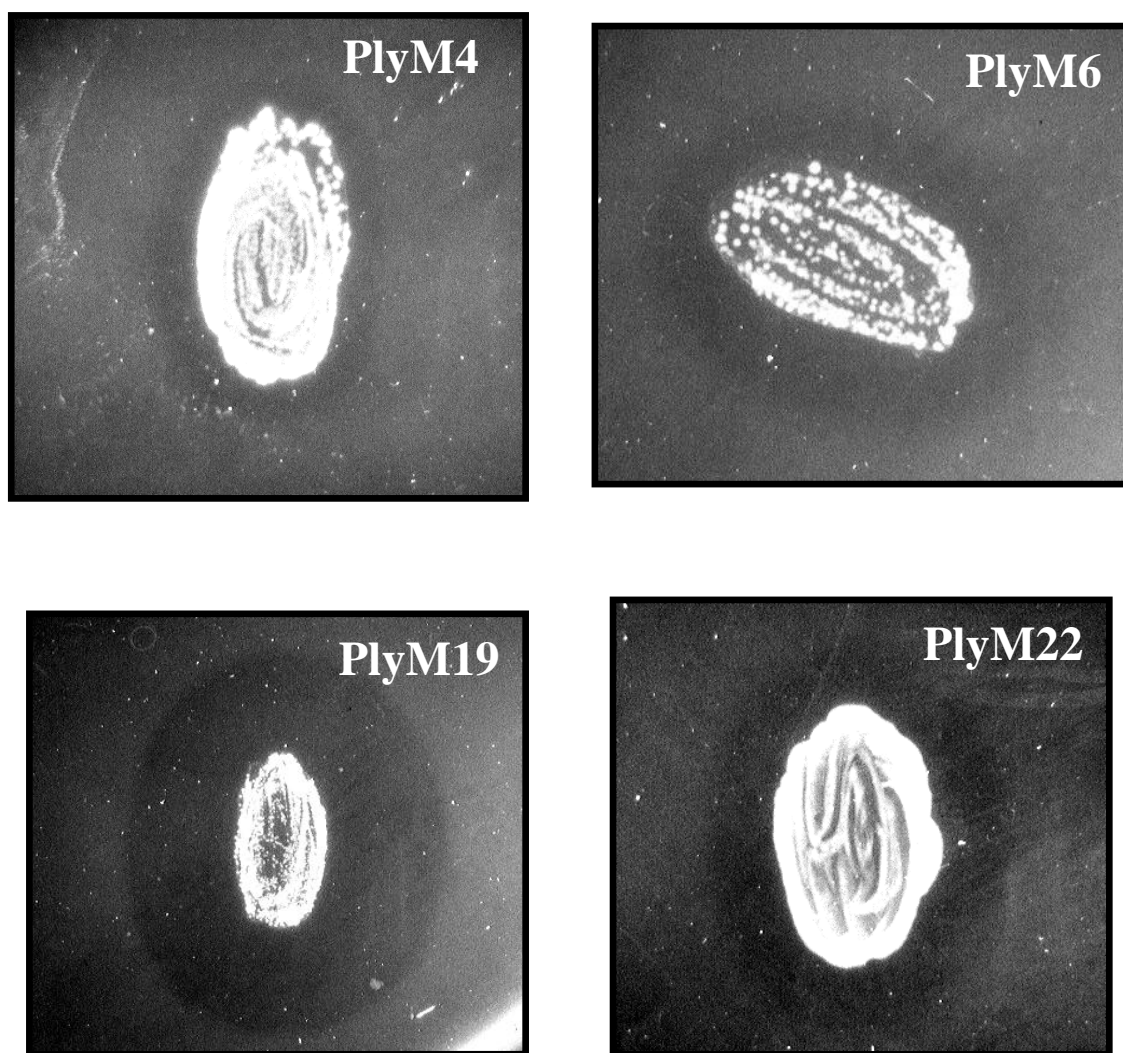


Figure 4.2b Lysin-Encoding Clones with
Pseudomonas Overlay (Metagenomic)

Figure 4.2, continued

Various lysin-encoding clones are overlaid here with autoclaved *P. aeruginosa* PAO1. **[A]** Previously characterized phage lysins. As opposed to the negative control (top), Gram-negative clearing zones are apparent around *E. coli* clones encoding CPL-1 (*Streptococcus pneumoniae* muramidase), PlyG (*Bacillus anthracis* alanine-amidase), PlyB (*Bacillus anthracis* muramidase), and PlyCM (*Clostridium perfringens* muramidase). In each case, the cells were streaked onto LB-agar containing arabinose inducing agent, allowed to proliferate overnight, and chloroform permeabilized prior to Gram-negative overlay. **[B]** Metagenomic lysin hits. Four representative examples of the 26 enzymes cloned during this study are depicted here (PlyM4, PlyM6, PlyM19, and PlyM22). As mentioned in the text, these clones (which encode holin/lysin combinations) often demonstrated toxicity when streaked onto arabinose-containing agar. This is particularly evident for the clones encoding PlyM4, PlyM6, and PlyM19, in which punctate colonies (presumably down-regulated mutants) have proliferated upon a lysed *E. coli* patch.

While this phenomenon was of little practical significance when cloning lytic enzymes by standard genomic techniques—here, the encoding phage and target species are known *a priori*—it could be useful for identifying unknown lysins from a metagenomic pool. (We should note that permeabilized Gram-negative cells are not broadly susceptible to endopeptidase lysins, as the peptidoglycan cross-bridge varies considerably among bacterial species; a more detailed discussion of this point is provided in the Discussions section of this chapter).

To these ends, the 502 hits identified in the initial hemolysis screen were generously restreaked as patches onto arabinose-containing agar and tested for their ability to lyse a soft-agar overlay of autoclaved *P. aeruginosa*, strain PAO1. This particular Gram-negative species/strain was chosen as the screening agent based on the visual clarity of the clearing zones it produced in preliminary experiments. Seeing that the 502 strains were selected for the toxicity of their encoded insert, the streaked patches obviously did not proliferate well on the arabinose plates; following overnight incubation, they were often comprised of lysed cells with occasional punctate colonies (presumably recombinant mutants with down-regulated insert expression, see Figure 4.2b). Nevertheless, even with little growth, enough recombinant protein was synthesized to produce distinct clearing zones in the overlay.

In total, 65 positive clones were identified in the secondary screen (Figure 4.2b), which were shown by sequencing to encode 26 unique lysins (denoted as PlyM1-PlyM26, GenBank accession numbers HM011589-HM011614). Of these

enzymes, 15 were observed for the first time during the secondary screen, while 11 had had been identified previously during the aforementioned sequencing of 52 hemolytic clones. Conversely, 6 of the putative lysins identified during the initial sequencing (PlyM27-PlyM32, GenBank accession numbers HM011615-HM011620) were not detected by the secondary screen, attributable to insoluble or insufficient recombinant expression.

Based on Pfam analysis, the 26 actively-expressed enzymes comprise a variety of molecular architectures and enzymatic motifs. They are summarized in Table 4.1 along with their putative domain assignments and known homologues among defined organisms. (For a complete list of Blast homologues, refer to the on-line supplementary information of the associated article.) The majority of the genes (PlyM1-PlyM20) encode typical Gram-positive lysins, with an N-terminal catalytic region and a C-terminal binding region. Of these 20 genes, 15 encode full-length lysins and 5 encode slightly truncated proteins (as an artifact of library construction) that lack the final portion of the C-terminal region but retain detectable catalytic activity. Of the 6 remaining lysins, 3 represent typical Gram-negative lysins (comprising only a catalytic domain, PlyM22-PlyM24) while 3 possess atypical architectures uncommon among bacteriophage (PlyM21, PlyM25-PlyM26). As expected, a large majority of the lysin-expressing clones (24/26, exceptions being PlyM18 and PlyM19) also encode a short, adjacent ORF that can be assigned putative holin functionality based on its database homologues and/or predicted transmembrane topology.

A	B	C	D	E	F
PlyM1	HM011589	197+ (trunc.)	Amidase 2	--	Numerous hits against genus <i>Bacillus</i> (10^{-65} – 10^{-20})
PlyM2	HM011590	249	Amidase 2	Amidase 2-associated	Same as above
PlyM3	HM011591	307+ (trunc.)	Amidase 2	PG-1	Same as above
PlyM4	HM011592	329	Amidase 2	PG-1	Same as above
PlyM5	HM011593	284+ (trunc.)	Amidase 3	PG-1	Two hits against <i>Desulfotomaculum reducens</i> and <i>Alkaliphilus metalliredigens</i> ($\sim 10^{-16}$)
PlyM6	HM011594	357	Amidase 3	PG-1	Same as above
PlyM7	HM011595	356	Amidase 3	PG-1	Same as above
PlyM8	HM011596	357	Amidase 3	PG-1	Same as above
PlyM9	HM011597	343+ (trunc.)	Amidase 3	PG-1	Same as above
PlyM10	HM011598	282+ (trunc.)	Amidase 3	PG-1	Same as above
PlyM11	HM011599	356	Amidase 3	PG-1	Same as above
PlyM12	HM011600	338	Amidase 2	PG-1 and PG-3	Numerous hits against broad variety of firmicutes (all $> 10^{-42}$)
PlyM13	HM011601	232	Amidase 3	SPOR	Numerous hits against genera <i>Bacillus</i> and <i>Geobacillus</i> (all $> 10^{-60}$)

Table 4.1 Cloned Metagenomic Lysins

A	B	C	D	E	F
PlyM14	HM011602	263	Amidase 3	SPOR	Same genera as above (all > 10 ⁻⁴⁹)
PlyM15	HM011603	230	Amidase 3	SPOR	Same genera as above (all > 10 ⁻²⁶)
PlyM16	HM011604	233	Amidase 3	SPOR	Same genera as above (all > 10 ⁻²⁸)
PlyM17	HM011605	256	Amidase 2	Not Predicted	Numerous hits against genus <i>Bacillus</i> (all > 10 ⁻⁶⁵); homologies do not extend into C-term binding region
PlyM18	HM011606	253	Amidase 2	LysM	Three hits against <i>Renibacterium salmoninarum</i> , <i>Actinomyces urogenitalis</i> (10 ⁻³⁰ - 10 ⁻²⁶)
PlyM19	HM011607	272	Amidase 3	SH3-5	Numerous hits against genus <i>Bacillus</i> (all > 10 ⁻⁷⁷)
PlyM20	HM011608	302	Amidase 2	SPOR	~10 hits against clostridia-like genera (10 ⁻⁶⁷ - 10 ⁻⁵¹)
PlyM21	HM011609	363	Endopeptidase, muramidase	Not Predicted	Numerous hits against diverse G+ and G- bacteria (all > 10 ⁻²⁵)
PlyM22	HM011610	159	Muramidase	--	Numerous hits against diverse G- bacteria (all > 10 ⁻³⁸)
PlyM23	HM011611	149	Muramidase	--	Numerous hits against diverse γ -proteobacteria (all > 10 ⁻⁴¹)
PlyM24	HM011612	181	Muramidase	--	Thirteen hits against <i>Acinetobacter</i> (10 ⁻⁶⁴ - 10 ⁻⁴⁵)
PlyM25	HM011613	256	?	PG-1 (N-term.)	Numerous hits against <i>Burkholderia</i> , other G- bacteria (all > 10 ⁻⁴⁰)
PlyM26	HM011614	188	?	--	Numerous 100-200 hits against large variety of bacterial species (all > 10 ⁻⁴⁹)

Table 4.1, continued

Table 4.1, continued

The 26 actively-expressed lysins cloned in this study are summarized here. [A] Lysin name. [B] GenBank accession number. [C] Protein length in amino acid residues. Five lysins (indicated with *trunc.*) were cloned as enzymatically active C-terminal truncations; for these proteins, the indicated length is that which was included on the plasmid insert. [D] Predicted enzymatic domains. The reader is referred to the text or Supplementary Document 1 for corresponding Pfam accession numbers. [E] Predicted cell-wall binding domains. For PlyM1, the protein was truncated within the C-terminal region, preventing an accurate prediction of a binding domain. Pfam analysis did not recognize conserved binding domain for the Gram-positive lysins PlyM17 and PlyM21, although a distinct C-terminal region exists that presumably serves this purpose. Binding domains were not predicted for 4/5 Gram-negative lysins (PlyM22-PlyM24, PlyM26). [F] Database homologues. Cloned lysins were subject to BLASTP analysis to identify homologues among sequenced bacteria/phage. In many instances, a lysin demonstrated highest homology to proteins encoded by a particular bacterial genus/species (or its phage). The identities of these taxa are identified here, and the degree of homology is indicated numerically with E-values (or a minimum E-value if numerous proteins demonstrated a continuous range of homology). In other cases, as noted, no genus/species was preferentially represented among the closest BLASTP hits.

Among the Gram-positive lysins, Pfam domain analysis suggests that all 20 possess N-acetylmuramoyl-L-alanine-amidase activity: 8 are predicted to be type 2 amidases (Pfam family PF01510), and 12 are predicted to be type 3 amidases (PF01520). Although these two protein families diverge sequentially, they target the same bond at the beginning of peptidoglycan's pentapeptide stem. At the C-terminal ends of the Gram-positive lysins, a variety of cell-envelope binding domains are likewise predicted. These include: 10 PG-1 motifs (PF01471); 1 PG-3 motif (PF09374); 5 SPOR motifs (PF05036); 1 SH3 type 5 motif (PF08460), 1 LysM motif (PF01476), and 1 amidase II-associated domain (PF12123). For two of the Gram-positive enzymes (PlyM17 and PlyM21), a clear C-terminal region is present, even though Pfam analysis fails to predict a binding motif. Most likely, these regions do possess binding functions, albeit ones that have not yet been categorized as conserved protein families.

Multiple sequence alignment of PlyM1-PlyM20 reveals strong similarities among some of the cloned proteins, summarized phylogenetically in Figures 4.3a and b. In particular, PlyM5 – PlyM11 demonstrate high sequence homology with one another, with pairwise sequence identities of 91-95% on the nucleotide level and 95-98% on the amino acid level (all E-values < 10^{-159}). PlyM3 and PlyM4 share this homology with PlyM5-PlyM11 at the C-terminus, but diverge in their enzymatic regions. These lysins were presumably derived from a group of highly similar phage infecting one of the component bacterial species of the fecal sample.

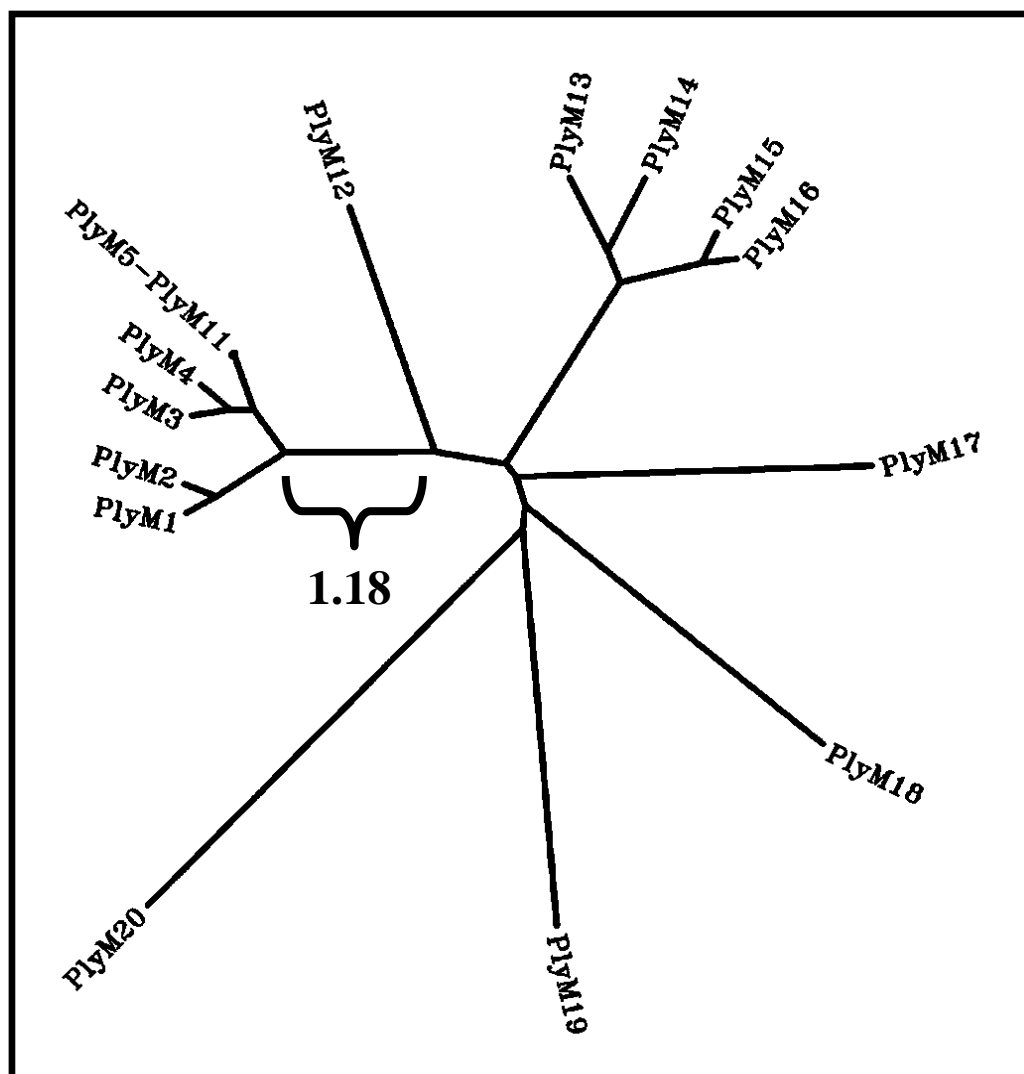


Figure 4.3a Similarity among Gram-Positive Hits (Non-Bootstrapped)

The sequence homology among the typical Gram-positive lysins (PlyM1-PlyM20) is organized here as a contemporaneous-tip phylogenetic tree (one branch length is included as reference). Several clusterings are apparent, especially PlyM5-PlyM11; these enzymes are depicted as a single node in the image, as their degree of homology (90+ pairwise nucleotide/amino acid identity) prevents individual visual differentiation. The reader should note that this tree was generated from a single multiple sequence alignment (MSA) so that the distance-values between proteins would be most visually apparent as the branch lengths.

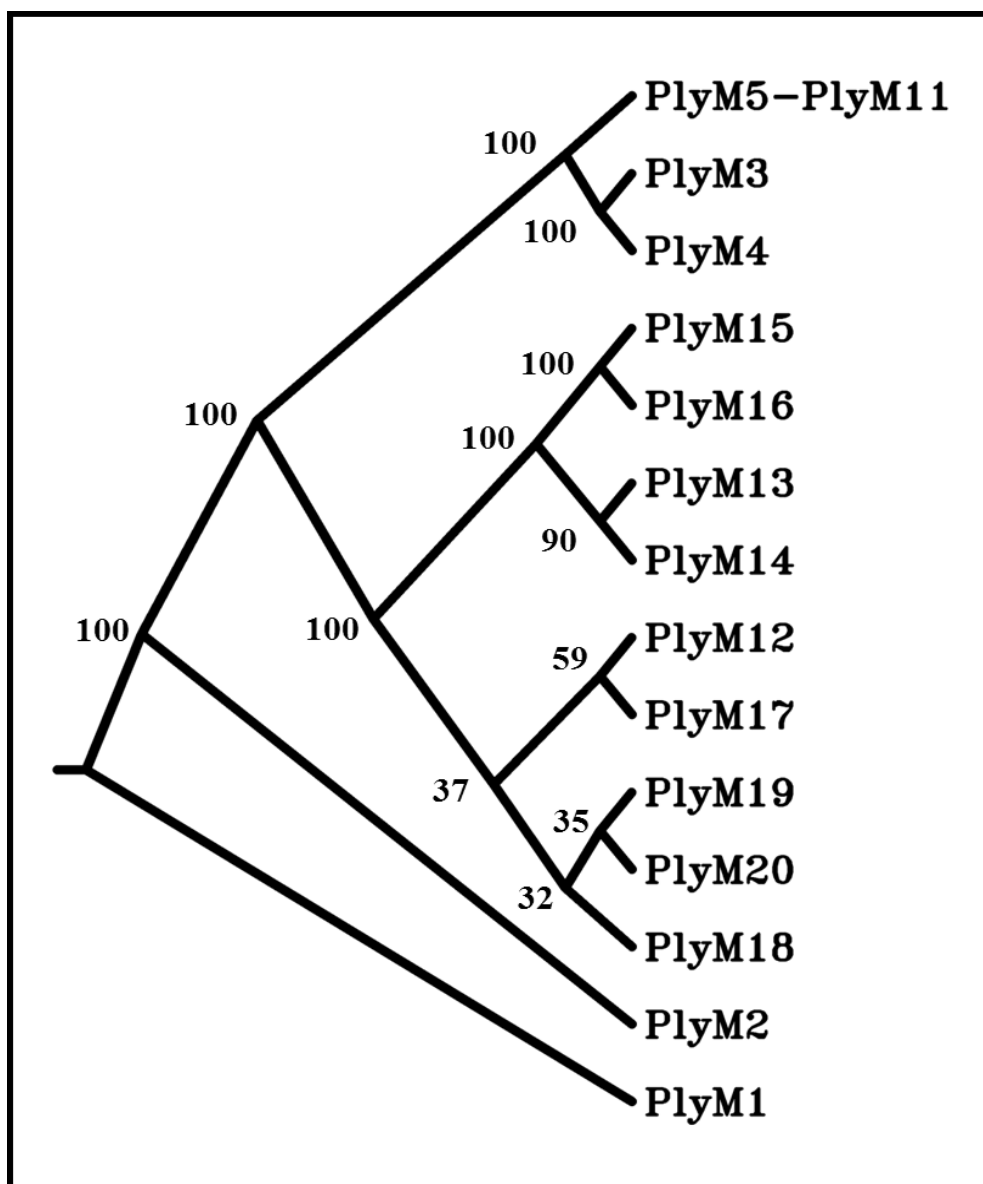


Figure 4.3b Similarity among Gram-Positive Hits (Bootstrapped)

Bootstrap analysis (100 rounds) was also conducted on the original MSA of PlyM1-PlyM20 and a consensus cladogram was generated. As in [A], PlyM5-PlyM11 have been collapsed here to a single node. The bootstrap consensus values are indicated at their respective nodes.

When the sequences of PlyM5-PlyM11 are compared to known proteins via Blast analysis, they demonstrate closest homology (E-values $\approx 10^{-16}$) to two putative prophage lysins from the sequenced genomes of *Delsulfotomaculum reducens* MI-1 and *Alkaliphilus metalliredigins* QYMF, both spore-forming organisms of the class *Clostridia*.

The remainder of the Gram-positive lysins were likewise subject to Blast analysis; for many of the proteins (PlyM1-PlyM4, PlyM13-PlyM17, PlyM19), the closest homologues are encoded by phage/prophage infecting *Bacillus* and related genera. In general, these homology findings are consistent with the fecal origin of the library, although one should not draw conclusions on the ecology of the sample based on this information alone. Aside from potential biases inherent in functional screening, it is important to emphasize that the sample was not a natural ecosystem in the first place, but rather a combination of several fecal samples pooled to increase diversity. Overall, if one is interested in viral metagenomics as a tool for studying microbial ecology, sequenced-based approaches are superior to functional screens.

Among the lysins that do not demonstrate typical Gram-positive structures, PlyM21 possesses two distinct catalytic regions – an N-terminal M23 endopeptidase domain (PF01551) and a central lambda phage-like muramidase domain (PF00959). While it is uncommon, several lytic enzymes from Gram-positive phage have been characterized that possess multiple catalytic domains (Cheng et al. 2005, Wang et al. 2009). It is currently unclear what, if any,

advantage is offered by this extended architecture. By contrast, lysins PlyM22-PlyM24 are relatively short in length and consist of single catalytic domain without any C-terminal binding region, an arrangement typical for Gram-negative phage. These three enzymes are each predicted to have the same muramidase domain (PF00959), although they share only moderate primary sequence homology with one another.

Finally, PlyM25 and PlyM26 differ significantly from the other enzymes cloned in this study and, in general, represent poorly characterized varieties of lytic enzymes. Domain analysis of PlyM25 predicts an N-terminal PG-1 binding motif (residues 3-38), but fails to recognize any catalytic domain. Rather, residues 76-249 correspond to a *domain of unknown function* (PF11860), from which the enzymatic activity presumably originates. Blast analysis of PlyM25 reveals several dozen ORFs of moderate homology (E-values: 10^{-20} – 10^{-40}) within the genomes of sequenced Gram-negative phage/prophage. Virtually none of these homologues, however, are currently annotated as lysins, exceptions being ORF12 of *P. aeruginosa* phage phiCTX (Nakayama et al. 1999) and ORF27 of *Burkholderia cepacia* phage Bcep781 (Summer et al. 2006). The latter two enzymes share the PG-1/DUF architecture of PlyM25, although only the *Burkholderia* protein has been assigned lytic function experimentally (its enzymatic class remains undetermined).

For PlyM26, Pfam analysis predicts neither a cell-wall binding nor a peptidoglycan hydrolase motif, but rather a chitinase enzymatic domain

(PF00182). While it is conceivable that PlyM26 represents a dedicated phage-encoded chitinase with cross-reactivity to *Pseudomonas* peptidoglycan (and not a lysin proper), this is an unlikely scenario. The first four nucleotides of PlyM26 overlap with an upstream 101-amino acid ORF containing three predicted transmembrane domains, making it a strong candidate for an adjacent holin. Moreover, given the structural similarity of chitin to peptidoglycan – and the fact that chitinases belong to the same protein clan (Pfam CL0037) as muramidases, glucosaminidases, and transglycosylases – it is more probable that PlyM26 represents a lysin of one of the later functionalities whose sequence merely deviates from the norm.

As with PlyM25, Blast analysis of PlyM26 reveals a number of homologues among sequenced Gram-negative phage/prophage. Again, however, few are annotated as phage lytic enzymes and (to our knowledge) none have been recombinantly expressed and characterized biochemically. Taken together, lysins PlyM25/PlyM26 and their database homologues argue that the complexity of many Gram-negative lysins is more complex than the classical picture suggests, reinforcing the need for further research in the field.

DISCUSSION

Overall, the above study represents one of the first functional screens of a viral metagenomic sample, and the technique serves as a general method for cloning lytic enzymes from uncultured phage. This naturally raises the

question—what applied or basic purposes could such screens serve? One possibility involves the identification of new enzymes with antibiotic activity against medically-relevant bacteria. To these ends, the 26 lysin-expressing clones were overlaid with various species of autoclaved Gram-positive pathogens. One of the enzymes (PlyM12) did produce clearing zones against *Bacillus anthracis* and *Enterococcus faecium*. It was subsequently sub-cloned to exclude the adjacent holin and subjected to preliminary chromatographic purification (data not shown). Nevertheless, when tested against live organisms in OD-drop experiments, PlyM12 failed to show activity comparable to previously-characterized lysins, and it was deemed a poor candidate for further study. Given the specificity that lysins show toward individual Gram-positive bacteria, finding a medically-relevant lysin would likely depend on whether a phage infecting that pathogen was represented in the original metagenomic pool. When one considers the origin and relative simplicity of the current library, it is not overly surprising that no such enzymes were identified here.

Metagenomic lysin screens could also be useful for identifying enzymes that are active under a desired set of biochemical conditions (temperature, pH, salts, etc...). For instance, several proviral lytic enzymes have been characterized from thermophilic species with the motivation that temperature resistance is an industrially attractive feature for peptidoglycan hydrolases (Ye and Zhang 2008, Matsushita and Yanase 2008). In this regard, a strength of metagenomics is its ability to study extreme environments whose bacteria/phage

are difficult to culture under laboratory conditions (Ferrer et al. 2007). One could easily envision isolating the phage fraction from an extreme environment (such as in Santos et al. 2010) and screening it for lytic enzymes with the above technique, circumventing the need for laboratory culture or prophage induction. For such cases, the buffered agar in which the secondary Gram-negative detection agent is suspended (here, PBS pH 7.4) could be varied to reflect the biochemical conditions under consideration.

That said, one must note that a sequence-based study of an environment (extreme or otherwise) could also identify lytic enzymes, as the assembled contigs of sequenced viral metagenomes have revealed numerous putative examples (Breitbart et al. 2002, Breitbart et al. 2003, Fierer et al. 2007, Williamson et al. 2008). Nevertheless, if the stated goal is to identify proteins of a particular functionality, clonal screening represents a far more rapid and cost-effective approach. Sequencing data alone does not predict what proteins are actively expressed in a recombinant host, and any candidate genes identified with high-throughput sequencing would have to be re-cloned via PCR into an appropriate expression vector.

All of this, moreover, presupposes that a DNA-sequence shows sufficient homology to other lysins already annotated in the database. In fact, metagenomic lysin screening is perhaps most useful in a very academic sense — the identification and characterization of new classes of catalytic and binding domains. Although lysins are already known to be remarkably diverse, enzymes

are still identified with completely novel sequences (Matsushita and Yanase 2008, Nelson et al. 2006). Even the simple library employed here yielded several clones whose identity as phage lysins would not have been obvious from their sequences alone. Accordingly, the characterization of PlyM25 and PlyM26 remains a subject of ongoing study, as does the screening of new libraries for additional novel enzymes.

Finally, it is important to note one limitation of this screening protocol and mention several ways it could be adapted to fit particular needs. Clearly, the method is capable of identifying lysins with glycosyl-hydrolase or alanine-amidase activity. At the same time, only one endopeptidase was cloned here (PlyM21), and this lysin possessed a secondary muramidase domain. In fact, endopeptidases present a particular obstacle for the technique, as it would be impossible to identify all endopeptidases using a single bacterial species in the soft agar overlay step.

While the polysaccharide backbone of peptidoglycan and the initial L-alanine of the pentapeptide stem are highly conserved among bacteria, considerable variability exists at the other stem positions and within the peptide cross-bridge (Vollmer et al. 2008a). Without the targeted bond, Gram-negative peptidoglycan would not be susceptible to a Gram-positive endopeptidase. For instance, as opposed to the clearing zones in Figure 4.2a (involving previously-characterized amidases and muramidases), no clearing zones were observed when several defined endopeptidases were applied to autoclaved *P. aeruginosa*.

To ensure endopeptidase coverage against a particular Gram-positive species, therefore, one could conduct the secondary screen in duplicate, including that species along with the general Gram-negative screening agent.

Another potential variation in the protocol could involve omitting the preliminary hemolysis screen entirely and progressing directly to the Gram-negative overlay step. In this case, one would replica-plate the original transformants onto additional LB-ampicillin plates (without arabinose, to avoid clonal toxicity), which could then be nebulized with inducing agent, chloroform treated, and overlaid with heat-killed *P. aeruginosa*. Clearing zones could then be traced to the respective colonies on the master plates. One potential advantage of this approach is that it could identify lysins from phage where the lysin and holin are not adjacent, which is the case in a minority of circumstances. Our rationale for not utilizing this approach here is that, on a single-colony scale, the hemolytic phenotype is more visually detectable than a clearing zone in a soft-agar overlay, which is better suited for the larger *E. coli* patches employed in the secondary screen. Nevertheless, the two approaches are hardly mutually exclusive and one could easily include both while screening a viral library to maximize lysin discovery.

Overall, the functional screening process described here is both straightforward and generalizable, allowing one to mine for biotechnologically- and academically-relevant enzymes from the largest genetic pool on the planet.

ACKNOWLEDGMENTS

This research was funded by NIH/NIAID grants AI057472 and AI11822 to VAF. JES acknowledges the support of the Pharmaceutical Research and Manufacturers of America Foundation and the NIH MSTP program (Weill Cornell/Rockefeller/Sloan-Kettering grant GM 07739). We would also to thank the Long Island Game Farm Wildlife Park and Children's Zoo for providing animal fecal specimens.

****NOTE:** In the submitted manuscript based on this chapter (Schmitz et al. 2010b), the following individuals were included as co-authors (in the stated order): Raymond Schuch and Vincent Fischetti.

CHAPTER 5

Back to Basics: Applying New Techniques to Genomic Samples

INTRODUCTION

Streptococcus suis is a Gram-positive, α - or β -hemolytic bacteria that is known for the variety of infections that it causes in pigs, including meningitis, septic arthritis, pneumonia, and bacteremia (Segura and Gottschalk 2004). It frequently leads to fatalities among commercial swine, and is especially problematic among newborn animals (Staats et al. 1997). *S. suis* is likewise an emerging zoonotic agent and can induce similar pathologies in humans as seen

in pigs (almost exclusively in those with occupational exposure to the animals) (Feng et al. 2010).

Reported veterinary and human cases of *S. suis* infection date back to the 1950's and 1960's (Field et al. 1954; Arends and Zanen 1988), although the species was not formally defined until 1987 (Kilpper-Bälz and Schleifer). Taxonomically, *S. suis* does not belong to the seven standard clusters of streptococci, nor does it group conveniently among Lancefield antigenic classifications (Facklam 2002). Instead, it is subdivided into 35 serotypes that depend on the nature of its polysaccharide capsule (Higgins et al. 1995). Various biochemical parameters have been defined for clinical identification of *S. suis* (Higgins and Gottschalk 1990). In practice, however, presumptive isolates are typically confirmed by antigenic or genetic methods (Lun et al. 2007).

The pathogenesis of *S. suis* infections in swine involves the initial colonization of the nasopharynx and palatine tonsils. The bacteria are thought to invade the tonsils, followed by dissemination via the lymphatics and/or efferent blood vessels; intracellular dissemination within host phagocytes has also been proposed (Segura and Gottschalk 2004). A number of potential *S. suis* virulence factors have been identified (reviewed in Staats et al. 1997), although the precise molecular mediators of pathogenesis are not well-defined. The extracellular capsule is necessary for infection, and serotype 2 strains are most commonly observed worldwide. It is important to note, however, that not all serotype 2 strains are virulent and other serotypes are frequently seen to cause disease. In

particular, serotype 9 infections are becoming more prominent in certain geographic regions (Blume et al. 2009). Complicating the situation further, colonized animals are often observed without signs of infection, and they are believed to play an important role in the epidemiology of *S. suis* (Luque et al. 2009). Dedicated environmental reservoirs have not been identified, although the isolation of the bacteria from other animal species suggests a more widespread distribution (Devriese et al. 1991).

The etiology of disease in humans likewise remains poorly understood. In all, there have been ~550 clinical cases of *S. suis* infection reported worldwide, with ~75 associated deaths (Feng et al. 2010; Lun et al. 2007). Many of these deaths stemmed from 1998 and 2005 outbreaks in China, which were particularly troubling given their association with a novel streptococcal toxic shock syndrome (Tang et al. 2006). It has been calculated that the annual risk of infection for employees of pig farms and slaughterhouses is roughly 3/100,000 (Arends and Zanen 1988). These estimates, however, are confounded by more recent immunological analyses of asymptomatic swine workers in the USA, ~10% of whom demonstrated high-titer seroconversion for *S. suis* antibodies (Smith et al. 2008). Moreover, it has been suggested that most hospital laboratories would misidentify *S. suis* isolates without a particular reason for suspecting the pathogen, as it falls outside typical diagnostic algorithms for human infections (Gottschalk 2004).

Considering the broad relevance of *S. suis*, there is great deal of interest in strategies for preventing its transmission within animal populations. While most *S. suis* isolates are sensitive to traditional antibiotics, there are significant downsides to large-scale antibiotic prophylaxis of swine (Staats et al. 1997). Vaccine development is naturally an area of focus, and positive results have been obtained during previous *in vivo* trials (Swildens et al. 2007). Commercially, it is not uncommon for animals to be vaccinated empirically with inactivated bacteria (Baums and Valentin-Weigand 2009; Haesebrouck et al. 2004). Nevertheless, the safety and overall efficacy of this practice are uncertain, and concerns remain regarding immunity across *S. suis* serotypes. As a result, innovative approaches for curtailing infections would be highly welcome.

In this light, phage lytic enzymes (also known as endolysins or lysins) have garnered much attention recently as novel antibacterial agents (reviewed in Villa and Veiga-Crespo 2010; O'Flaherty et al. 2009; Fischetti 2008). These peptidoglycan hydrolases (targeting a variety of specific bonds) are encoded by virtually all dsDNA phages and are responsible for digesting the host cell wall during phage infection. This leads to osmotic lysis of the bacteria and release of viral progeny. Biotechnological interest in these proteins stems from their ability to lyse Gram-positive bacteria when applied exogenously, as the peptidoglycan of these species is continuous with the extracellular space. Lysins have been proposed as potential *enzybiotic* agents, and are notable for the potency and specificity they demonstrate toward particular bacteria (generally, species that

the encoding phage infects or closely-related organisms). In a number of *in vivo* trials, lysins have successfully eliminated targeted pathogens from colonized and/or infected mucous membranes (Daniel et al. 2010; Cheng et al. 2005; Loeffler et al. 2001; Nelson et al. 2001).

To date, two phages infecting *S. suis* have been isolated and studied. Harel et al. (2003) induced a siphoviral prophage from the genome of a serotype 2 strain, although the identity of its lysin remains undetermined. More recently, Ma and Lu (2008) isolated a lytic phage from nasal swabs of healthy pigs, sequencing its 36 kb genome. This phage, termed SMP, demonstrated a limited host range, infecting only 2/24 *S. suis* strains within serotype 2. The same group later PCR-cloned and recombinantly expressed the SMP lysin (LySMP); the enzyme demonstrated broad bacteriolytic activity *in vitro* against several serotypes (albeit with the addition of protease inhibitors and at somewhat higher concentrations, 50+ µg/ml, than typically associated with phage lysins). The main drawback of LySMP was technical: the recombinant protein did not fold properly by itself and was only active in the presence of reducing agents, which could complicate potential *in vivo* trials (Wang et al. 2009).

In the current study, we sought to expand and evaluate the collection of known *S. suis* lysins. Through a combination of functional screening and sequence analysis of recently-published genomes, we identified two new prophage lytic enzymes (termed PlySs1 and PlySs2). PlySs1 is a homologue of the above-mentioned LySMP (but demonstrating superior expression

properties), while PlySs2 represents a divergent class of enzyme altogether. Both lysins were purified and characterized biochemically, and they are currently being employed in an ongoing *in vivo* trial involving experimentally-infected pigs. In the present thesis, I describe the cloning of PlySs1 and PlySs2, as well as the *in vitro* characterization of the former. Overall, through agents like these, it is possible to expand the potential of enzybiotics into the often overlooked – but, nevertheless, crucial – world of veterinary microbiology.

MATERIALS AND METHODS

Prophage functional screening. The genomes of several clinical strains of *S. suis* were subjected to functional shotgun screens in search of lytic enzymes from incorporated prophage. Microgram quantities of gDNA were briefly subjected to restriction digestion with Tsp509I (NEB). Fragments 1.5 – 4 kb in length were isolated via agarose-gel electrophoresis and ligated into EcoRI-linearized pBAD24 plasmid (Guzman et al. 1995). This plasmid confers ampicillin resistance and allows for arabinose induction of the recombinant insert.

To identify lysin-encoding clones, libraries were subject to a novel screening technique that relies upon the toxicity of adjacently-encoded holin proteins (see Schmitz et al. 2010b or Chapter 4 for details). Briefly, *E. coli* TOP10 transformants were plated onto LB-agar supplemented with ampicillin and sheep's blood. Following proliferation to macroscopic colonies, the plates were exposed to a mist of arabinose to induce recombinant transcription. Toxic clones

were revealed by the development of surrounding zone of hemolysis. These colonies were identified, re-propagated and subject to a secondary screen in which they were overlaid with heat-killed bacteria (to assay directly for the production of lytic enzyme). For the *S. suis* strain (7711) that yielded the PlySs1 lysin, ~3,500 clones were subjected to the original hemolysis screen; 100 of these were selected for the secondary screen, 2 of which encoded the lytic enzyme.

Sub-cloning of truncated PlySs1. For the positive hits identified above, recombinant inserts were amplified and sequenced with pBAD-targeted primers (Genewiz, Inc., South Plainfield, NJ). The specific position of the lysin gene was located by analysis with ORF Finder (www.ncbi.nlm.nih.gov/projects/gorf/) and BlastX (Altschul et al. 1990; blast.ncbi.nlm.nih.gov/Blast.cgi). For the translated protein, putative enzymatic and binding domains assignments were made via Pfam (Finn et al. 2009; pfam.sanger.ac.uk).

Based on this information, primers were designed for synthesizing a truncated constructed (Δ PlySs1) with an inserted stop codon preceding the C-terminal glucosaminidase domain (specifically, after D254). The primer sequences were:

TTTGAATTCTTTATGACAATCAATCTTGAAACATCCATTCGT – *fwd* and
TTTGCATGCTTAGTCCCCGTCCTCCTTGAATTGC – *rev*. The engineered stop codon is double-underlined, while the single-underlined nucleotides correspond

to EcoRI and SphI restriction sites for ligation into pBAD24. The resultant plasmid was cloned and maintained in TOP10 *E. coli*.

Genomic sequence analysis and cloning of PlySs2. The genomes of 8 sequenced isolates of *S. suis* were inspected for the presence of lysin-encoding genes within integrated prophage (Holden et al. 2009; Chen et al. 2007). These strains were: 05ZYH33 (NCBI Genome Project #17153); 98HAH33 (#17155); BM407 (#32237); GZ1 (#18737); P1/7 (#352); SC84 (#32239); 05HAS68 05HAH33 (#17157); and 89/1591 (#12417). For each genome, the topologically-arranged list of annotated ORFs was manually inspected for potential prophage regions. If a prophage was suspected, the theoretical translations of each ORF in that region were subject to BlastP and Pfam analysis. Putative lysin-status was assigned based on the combination of predicted enzymatic and binding domains.

The only lysin gene identified in this manner (PlySs2 from strain 89/1591) was PCR-cloned from genomic DNA with the following primers:

AATGCTAGCCTGATACACAGTTAGAGACC – *fwd* and

CCTAAGCTTCTTTTCACAAATCATAATCCCCAG – *rev*. The underlined

nucleotides again represent restriction sites (NheI and HindIII) for cloning into

pBAD24. It should be noted that the forward primer corresponds to a position

~60 bp upstream the start of the gene. For *plySs2*, several in-frame ATG triplets

are situated near one another at the 5'-end. To avoid choosing the incorrect start

codon, the upstream region was included so that transcription could be guided by the native ribosome binding site (instead of the engineered RBS of pBAD24).

Recombinant expression and purification of Δ PlySs1. To express Δ PlySs1, the clone was grown in Power Broth + LB-Booster (Athena Enzyme System) to $OD_{600} \approx 1.0$ and induced with 0.2% arabinose. The culture was shaken for 4 hr at 37°C (inclusion bodies would form at longer times). The expressing cells were pelleted, resuspended in 15 mM phosphate buffer pH 6.2, and lysed by three passages through an EmulsiFlex C-5 homogenizer. Residual debris was removed by centrifugation (1 hr, 35,000 X G), and ammonium sulfate was added at 225 g/L (40% saturation). The precipitated protein was pelleted and resolubilized in 15 mM phosphate pH 7.4, and dialyzed against this buffer overnight.

The dialysate was next passed through a DEAE anion-exchange column equilibrated against the same buffer (fast flow resin, General Electric). Quite surprisingly, Δ PlySs1 demonstrated the same sort atypical binding response observed for PlyCM in Chapter 2. With a predicted pI of 7.7, one would expect Δ PlySs1 to bind weakly to DEAE at pH = 7.4, or perhaps flow directly through it. Again, however, there was a transient interaction in which PlyCM would initially bind the resin, but then slowly elute as excess Buffer A was passed over the column (see Figure 5.1). The effect here was even more pronounced than with PlyCM, as the initial flow-through trace would return almost completely to baseline before Δ PlySs1 began eluting.

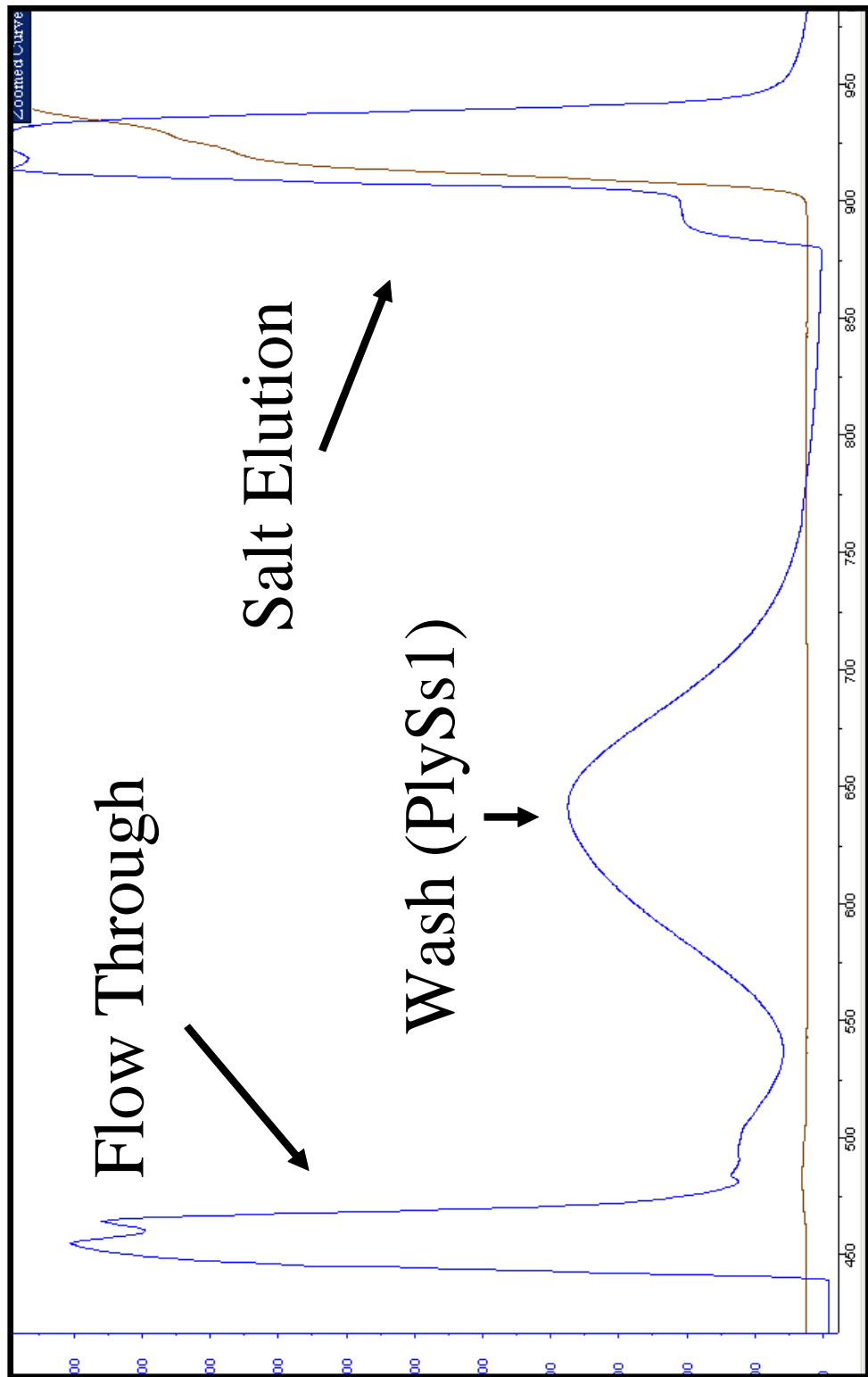


Figure 5.1 Δ PlySs1 Purification

Figure 5.1, continued

The dialyzed 40% ammonium sulfate cut of Δ PlySs1-lysate was loaded onto DEAE resin at pH = 7.4. Non-binding proteins flowed through the column as the sample was loaded. After continued wash in the same buffer, an extended secondary peak eluted. Non-transiently binding proteins eluted with the addition of 1M NaCl. The *extended-wash* fraction, when analyzed by SDS-PAGE, was observed to be high-purity Δ PlySs1 (see Figure 5.3).

This phenomenon led to a highly pure lysin preparation in only a single chromatographic step. For every liter of original *E. coli* culture, 50-100 mg of final Δ PlySs1 could be obtained.

Δ PlySs1 enzymatic specificity. To determine the bond specificity of the N-terminal enzymatic domain of Δ PlySs1, cell-wall preparations were made from *S. suis* type-strain S735. The *Gram-positive* protocol of Rosenthal and Dziarski (1994) was utilized with only slight modifications. Purified cell walls were suspended in phosphate buffered saline (pH 7.4); Δ PlySs1 and hen egg-white lysozyme (Sigma) were both added to 500 μ g/ml, and the solution was shaken gently overnight at room temperature. The digest was centrifuged (20 min, 16,000 X g), and the supernatant passed through a 10,000 MWCO micro-filter (Ultracel YM-10, Millipore) to remove the lytic enzymes and remaining macromolecular cell-wall components. This sample and an undigested negative control were subject to MALDI-TOF mass spectrometry (positive ion mode, Applied Biosystems DE-STR spectrometer) by the Rockefeller University Proteomics Core Facility.

***In vitro* characterization of Δ PlySs1.** The biochemical properties and strain-specificity of Δ PlySs1 were examined mainly through optical density-drop experiments. All bacteria were maintained on Brain Heart Infusion (BHI) agar plates. For each experiment, individual colonies were inoculated into 50 ml liquid BHI, which were gently shaken (125 rpm) for several hours at 37°C. When the cells had reached mid-log phase (OD \approx 0.5), they were pelleted, washed and

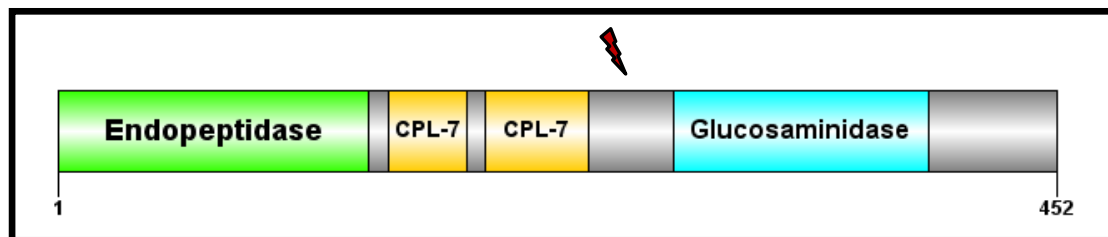
resuspended in the appropriate buffer (which varied from experiment to experiment) to an $OD_{600} \approx 0.8$. Optical density measurements were performed in 96-well plates at 37°C. Δ PlySs1 or lysin vehicle was added immediately prior to the start of each experiment, and measurements were taken every minute.

For CFU analysis, cells were treated with Δ PlySs1 in a similar manner as above. After 1 hr at 37°C, samples were each diluted over five orders of magnitude, with triplicate plating at each dilution onto BHI agar. For (attempted) MIC-analysis, cells were suspended in 2X BHI to $\sim 10^6$ cells/ml (determined by comparison with a McFarland standard). Sterile-filtered lysin and/or vehicle was added at the appropriate concentrations, yielding a final suspension of $\sim 5 \times 10^5$ cells/ml in 1X broth (Wiegand et al. 2008). Cells were distributed within a 96-well plate, and OD_{600} -measurements were taken overnight (every 2 minutes) with the plate maintained at 37°C.

RESULTS AND DISCUSSION

Identification of PlySs1 and sub-cloning of truncated construct. A prophage lytic enzyme, PlySs1, was cloned from a functional genomic screen of *S. suis* strain 7711, a serotype 7 isolate originating from the Netherlands¹. PlySs1 is a 452-residue protein: Pfam analysis predicts a type 5 alanine-amidase domain (PF05832) at the N-terminus, followed by a double CPL-7 cell-wall binding domain (PF08230) in the central region, and a secondary glucosaminidase domain (PF01832) at the C-terminus (Figure 5.2a).

(A)



(B)

PlySs1	MTINLET SIRWMS DRVGVSY SMDYRNGPNSYDCSSAVYYALMAGGAISAGWAVNTEYMH	60
LySMP	MTINIETAIRWMTDRVGLVKYSMDYRNGPNSFDCSSSVYYALMAGGAISAGWAVNTEYEH	60
	****:*:*:*:*:*:* * .*****:*:*:*:*:*:* *	
PlySs1	DWLIRNGYVLVAENKPFNAQRHDVCILKRGYSSGAGGHVVFVDNVNVIHCNYARNGIS	120
LySMP	DWLIKNGYKLI AENQDWD AKRGDIFIWGRRGQSSGAGGHTGIFVDPDNI IHCNYANNSIT	120
	****:*:* *:*:*: :*: * *: * *:* *****. **** *:*****.*:*:	
PlySs1	IDNYNQVHRG--MYYYLYR PANQPSIS--NKS LDQLVKETLAGVHGNGDTRKASLGSQY	175
LySMP	INNYNQTAAASGWMYCYVYRLGNQPTTSPAGKTLDTLVKETLAGKYGNQDQKKAALGNQY	180
	*:*****. . ** *:* *:*:* * .*:** ***** :**** *:*:*:*:	
PlySs1	EAVMAVINGKASASEKSDEELAREVLAKHGAGEDRKRSLSGPRYEPVQAKVNELLK----	231
LySMP	EAVMAVINGKATAPKKTVDQLAQEVIQKGHGNGEDRKKSLSGPDYDAVQKRVTEILQGSTS	240
	*****:*:*:*: :*:**:* *****:*:*:*:* *:*:*:	
PlySs1	--AKEKPSETAKN-----EPQTVQ-----FKEDGDL SFNGAILKKS VLEIT	270
LySMP	GNAPKLASDAPKNEVNSSTEPKTEETWATGKATDTKITKEDGDL SFNGAILKKS VLDVI	300
	* : .*: :.* ** *:* * : *****:*:*:	
	↑	
PlySs1	LKKCKEHDILPSYALTILHYEGLWGTS AVGKADNNWGGMTWTGQGNRPSGVIVTQGLARP	330
LySMP	IANCKKHDILPSYALTILHYEGLWGTS AVGKADNNWGGMTWTGKGERPSGVITVTQGTARP	360
	* :*:*****:	
PlySs1	SNEGGHYMHYATVDDFLTDW FYLLRKDGSYKVS GALTFSES IKGMFQVGGAKYDYAAAGY	390
LySMP	ACEGGHYMHYASVDDFLTDW FYLLRSGSYKVS GAKTFSDAVKGMFKIGGAVDYAASGF	420
	: *****:*:*:*:*:*:*:*:*:****** ***: :*:*****:*:* *	
PlySs1	DSYLVGATSRLKAIESENGSLTRFDATSNNVHSVD-PDKISVDIDGIEVTINGV VYKLEK	449
LySMP	DSYIIGASSRLKAIEAENGSLDKFDKQT--VTDVGQSDKIEVTIEGIEISINGV TTYTLSK	478
	:*:*:***:*:*:* ** : * *.***.* *:*****:*:*:*:*:	
PlySs1	KPV 452	
LySMP	KPV 481	

Figure 5.2 PlySs1 Prophage Lysin

Figure 5.2, continued

[A] The molecular architecture of PlySs1. The N- and C-terminal enzymatic domains are shown, along with the central cell wall-binding region. The N-terminal enzymatic domain is annotated according to its observed activity (γ -endopeptidase), rather than its predicted activity (alanine-amidase). Prior to purification and functional analysis, PlySs1 was recombinantly truncated to exclude its C-terminal glucosaminidase domain (Δ PlySs1). The position of that truncation, D254, is likewise denoted (red lightning-bolt). **[B] *S. suis* lysin alignment.** The protein sequences of PlySs1 and LySMP (YP_950557) are compared here. Residue identities are denoted with underlying asterisks and (in most cases) highlighting. The color of the highlighting indicates the particular domain to which the sequence corresponds (endopeptidase-yellow, CPL-7-blue, glucosaminidase-pink). Residue identities that do not fall within a predicted domain are not highlighted (only denoted with an asterisk). The position of the engineered C-terminal truncation of Δ PlySs1 is designated with an arrow.

PlySs1 demonstrates high homology to the previously-characterized LySMP lysin throughout its entire sequence (Figure 5.2b). The proteins share ~70% nucleotide and amino-acid identity with one another (all E-values < 10⁻⁶⁰).

Architecturally, the domain arrangement of PlySs1 (and LySMP) is highly atypical. Gram-positive lysins typically consist of an N-terminal enzymatic domain and a C-terminal binding domain (Fischetti 2008). While occasionally lysins are seen with two N-terminal lytic domains (for instance, Baker et al. 2006; Cheng and Fischetti 2005), it is rare for a second enzymatic functionally to be encoded *after* the binding domain. Besides PlySs1 and LySMP, the only other example characterized to date is the LambdaSa2 lysin of *S. agalactiae* (a few more putative examples exist within nucleotide databases) (Pritchard et al. 2007). Interestingly, the C-terminal enzymatic domains of these lysins demonstrate homology to several annotated bacteriophage tail proteins; the latter are known to possess muralytic activity for purpose of viral DNA injection (Kenny et al. 2004; Piuri and Hatfull 2006). Conceivably, the PlySs1/LySMP/LambdaSa2 group of enzymes could have evolved from a recombination event that juxtaposed a tail enzyme with a phage lysin proper.

Working with LambdaSa2, Donovan and Foster-Frey surprisingly observed *increased* enzymatic activity following removal of the C-terminal glucosaminidase domain (2008). With this motivation, we engineered a truncated PlySs1 construct with only the N-terminal enzymatic and central

binding domains. This construct was expressed and purified for subsequent functional analysis (Figure 5.3); it will henceforth be referred to as Δ PlySs1.

Identification of PlySs2. A second lysin, PlySs2, was identified (and, subsequently, PCR-cloned) through sequence analysis of 8 published *S. suis* genomes. PlySs2 is encoded in a prophage region of serotype 2 strain 89/1591 (the lysin-encoding ORF was originally annotated as *SH3-type 5 domain protein*, ZP_03625529; Lucas et al. 2004); it was the only lysin identified through our database search. The structure of PlySs2 is quite unlike that PlySs1 and LySMP. It encodes a predicted N-terminal CHAP domain (cysteine-histidine amidohydrolase/ peptidase, PF05257) and a C-terminal SH3-type 5 domain (PF08460) (Figure 5.4a).

CHAP domains are included in several previously-characterized streptococcal (Nelson et al. 2006; Baker et al. 2006) and staphylococcal (Daniel et al. 2010; Becker et al. 2009) lysins. On a primary sequence level, however, the CHAP domain of PlySs2 is rather divergent from other database CHAP domains (all pairwise E-values $> 10^{-15}$). In Figure 5.4b, CHAP domain of PlySs2 is aligned with that of the well-characterized streptococcal PlyC lysin, demonstrating conserved catalytic residues but only a modest level of identity overall (28% sequence identity, E-value = 10^{-8}) (Nelson et al. 2006). SH3 domains are commonly seen in viral and bacterial cell wall-binding proteins, although the exact molecular target remains unknown (Xu et al. 2009).

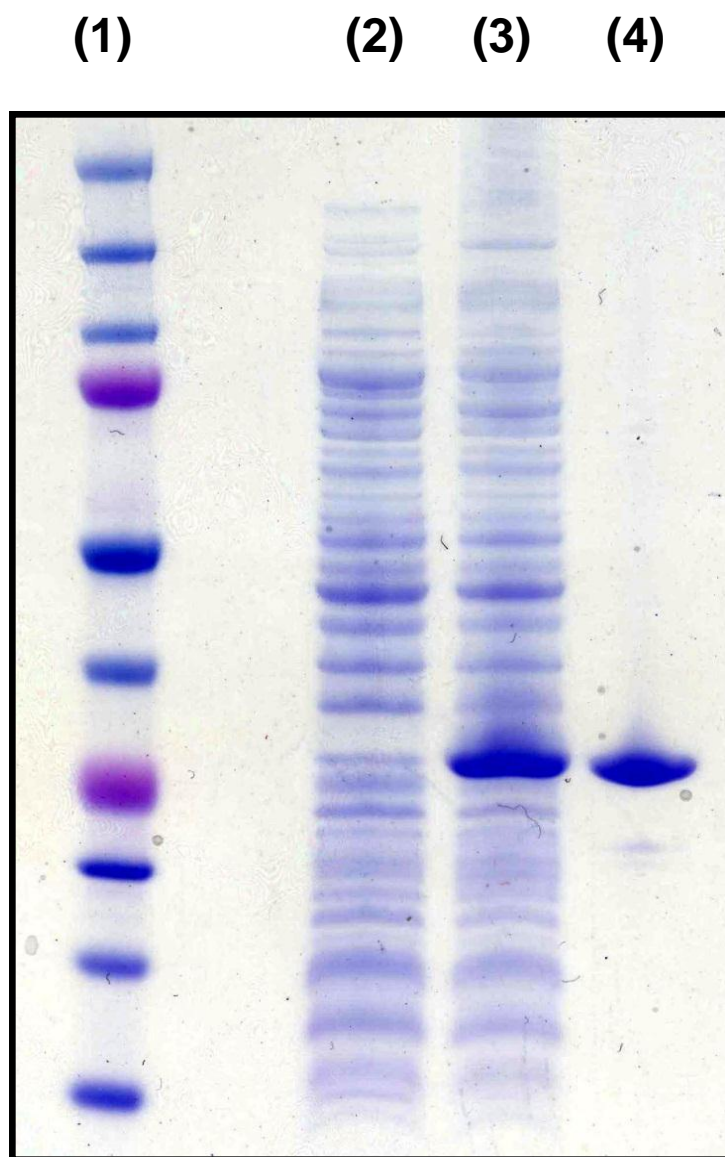


Figure 5.3 Purified Δ PlySs1

Δ PlySs1 was chromatographically purified after recombinant expression. Lane 1: molecular weight ladder. Lane 2: crude extract of encoding strain prior to induction. Lane 3: crude extract of encoding strain 4-hr after induction. Lane 4: final product following isolation protocol. By visual approximation, Δ PlySs1 is > 90% pure; its band appears just above 25-kDa marker (predicted MW = 28.1 kDa).

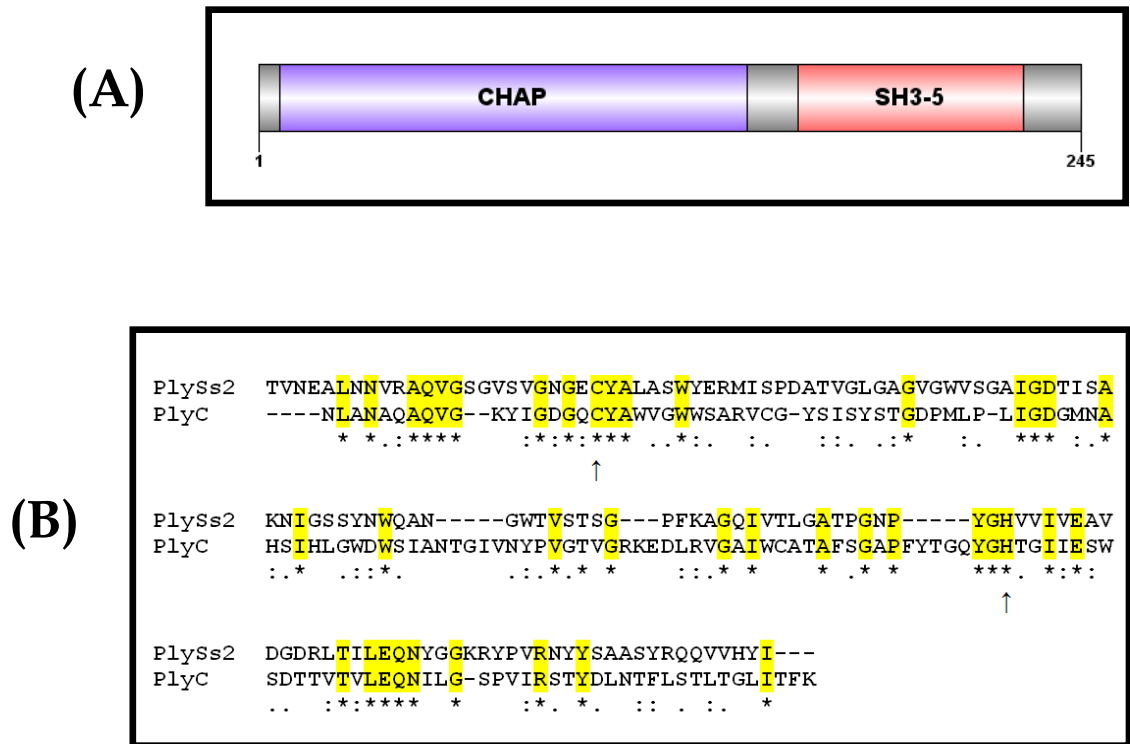


Figure 5.4 PlySs2 Prophage Lysin

[A] The molecular architecture of PlySs2. [B] Enzymatic domain

alignment. The CHAP domains of the streptococcal lysins PlySs2 and PlyC (subunit A, GenBank no. AAP42310) are aligned here. Amino-acid identities are indicated with underlying asterisks and highlighting. The positions of the presumptive catalytic residues (cysteine and histidine, for which the domain is named) are indicated with arrows.

Enzymatic characterization of Δ PlySs1. The first issue to be addressed for Δ PlySs1 was its enzymatic specificity. Although Pfam predicts a type 5 alanine-amidase domain, this motif is historically associated with false computational assignments. The same domain was predicted at the N-terminus of LambdaSa2, but was instead shown to be a γ -glutaminyll-L-lysine-endopeptidase experimentally (Pritchard et al. 2007). To clarify the situation here, purified *S. suis* cell walls (from type strain S735) were subject to double digestion with HEWL (a muramidase) and Δ PlySs1. Mass spectrometric analysis of the filtered digest revealed the expected peaks for combined muramidase/ γ -endopeptidase activity, demonstrating that Δ PlySs1 hydrolyzes the same bond as LambdaSa2 (Figure 5.5).

The enzymatic specificity of PlySs2 currently remains in doubt. CHAP domains are catalytically diverse and can possess either alanine-amidase (Nelson et al. 2001) or cross-bridge endopeptidase activity (Daniel et al. 2010), depending on the particular lysin. Further complicating the situation here is the fact that the molecular nature of the peptidoglycan cross-bridge in *S. suis* can vary between strains (Kilpper-Bälz and Schleifer 1987). Accordingly, mass-spectrometric experiments with PlySs2 are ongoing.

Optimization of Δ PlySs1 activity. The optimal biochemical conditions for Δ PlySs1 were determined against live cells of the encoding *S. suis* strain (7711).

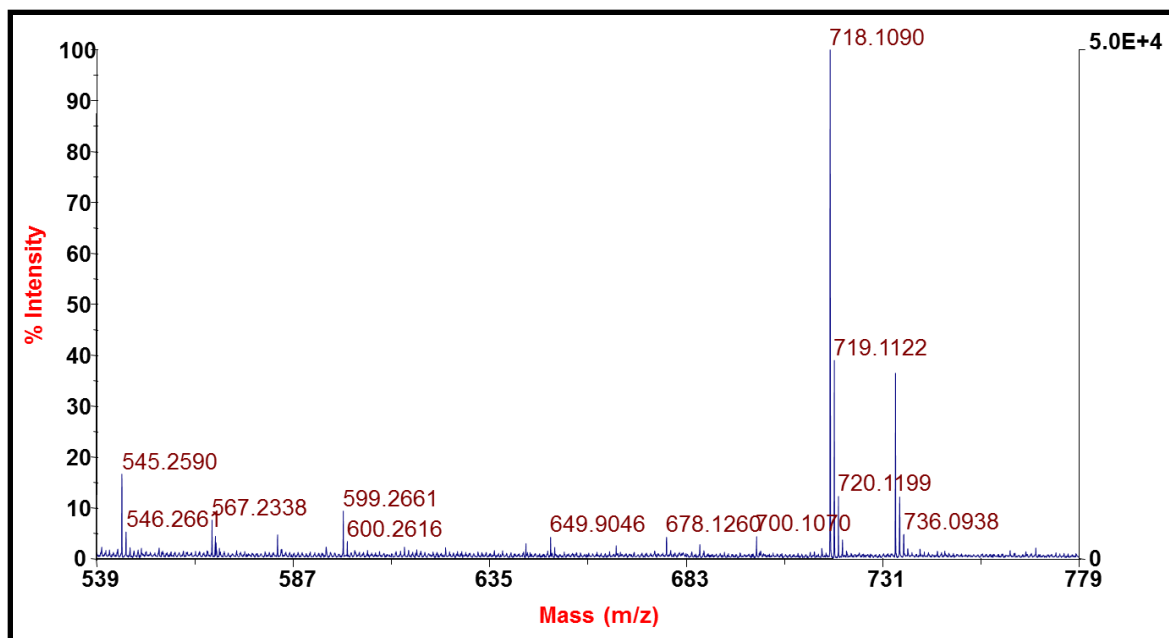


Figure 5.5 Δ PlySs1 Enzymatic Specificity

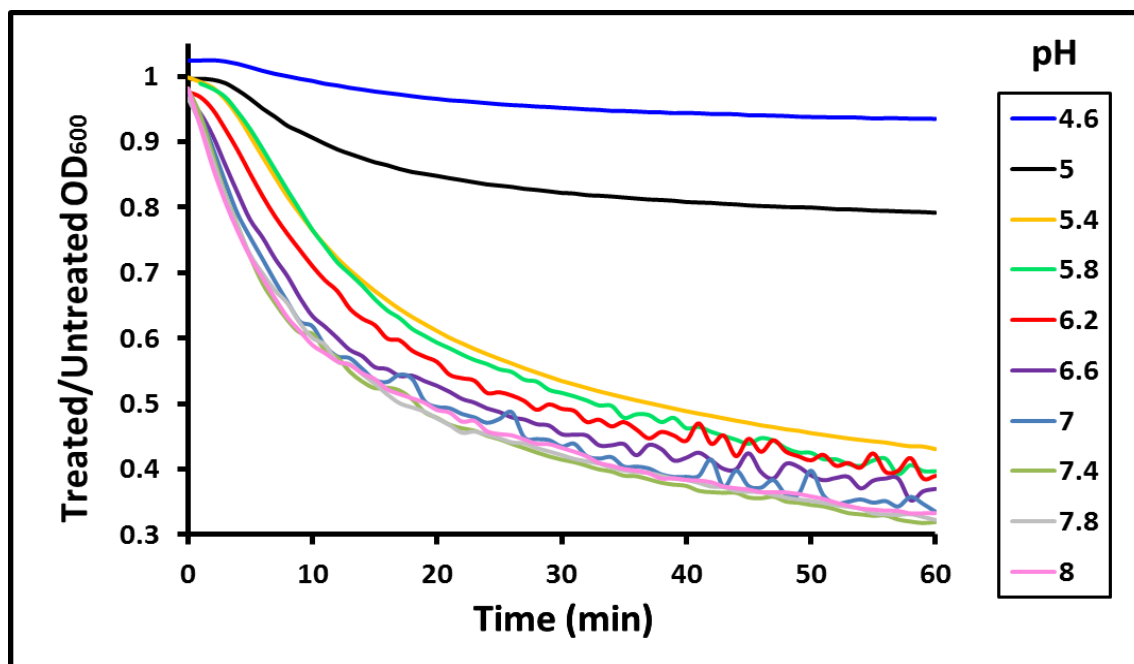
Depicted here is the mass spectrum of digested (and filtered) *S. suis* S735 cell wall. Hen egg-white lysozyme (a eukaryotic muramidase) was used in combination with Δ PlySs1. The two predominant peaks (with respective isotopic tails) are $m/z = 718$ and $m/z = 734$. This corresponds exactly to the predicted masses of the $[\text{Na-M}]^+$ and $[\text{K-M}]^+$ adducts of GlcNAc-MurNAc-L-Ala-D-Gln. This suggests that Δ PlySs1 possesses gamma-endopeptidase activity, cleaving the peptidoglycan stem between D-Gln and L-Lys. When a mass spectrum was taken of undigested cell wall (not shown), the above two peaks were absent.

For these experiments, activity was gauged through the degree of turbidity reduction (OD_{600}) of an aqueous bacterial suspension following the addition of lysin. The pH-dependence of the enzyme was first addressed using two buffer sets with adjacent pH ranges, citrate/phosphate: 4.6 – 8.0; and bis-tris-propane (BTP): 7.0 – 9.7. An extended spectrum of lysis was observed, from 5.4 – 9.4 (Figure 5.6). In BTP, lysis was maximal from 8.2 – 9.0; at commensurate pH-values, however, the magnitude of the OD-drop was slightly more pronounced in citrate/phosphate. The role of salt concentration was likewise considered, although it did not greatly affect Δ PlySs1-induced lysis. At constant enzyme concentrations, bacteriolysis varied little from 0 – 1000 mM NaCl, with only small numeric increases under the most hypotonic conditions (Figure 5.7).

Exposure of Δ PlySs1 to an excess of DTT had no impact (either positive or negative) on activity (Figure 5.8a). This indicates both that [1] the lysin does not rely on intramolecular disulfide bridges, as well as [2] that it was properly folded following recombinant expression and purification. The latter point is significant given that LySMP had to be treated with reducing agents prior to use (Wang et al. 2009). The reason for this discrepancy between two homologous lysins is unclear, although (most likely) it involves the numerous variable cysteine residues between the proteins (see Figure 5.2b).

Interestingly, treatment with EDTA *enhanced* Δ PlySs1-induced lysis of *S. suis* (Figure 5.8b).

(A)



(B)

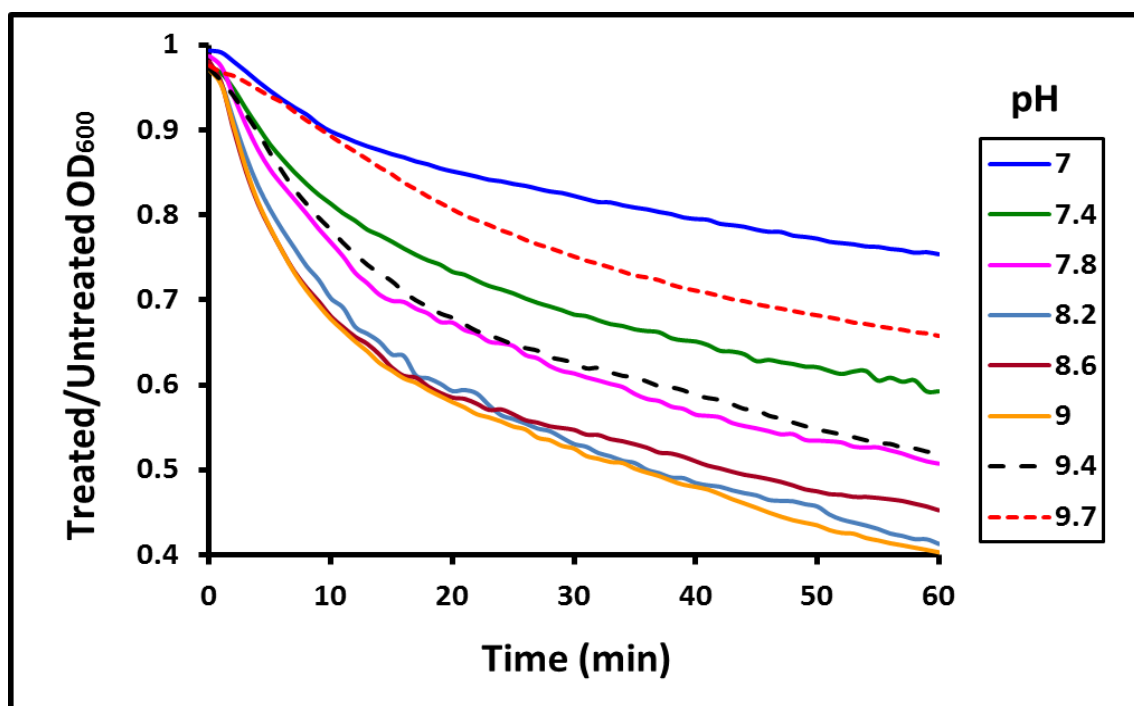


Figure 5.6 Δ PlySs1 pH Dependence

Figure 5.6, continued

[A] Cells of host strain 7711 were suspended in phosphate-citrate buffer (40/20 mM) at a range of pH-values from 4.6 to 8.0. Δ PlySs1 was added (110 μ g/ml) and OD₆₀₀ was measured over 60 min (horizontal axis) at 37°C. The vertical axis represents the treated/untreated OD₆₀₀-ratio at each time-point. For each pH-value, the curve depicts the running average of 3 independent experiments. Overall, activity was maximal at the upper end of the buffering range. **[B]** Here, bis-tris-propane (40 mM) was employed as the buffering agent with a pH-range from 7.0 to 9.7; Δ PlySs1 was again added to 110 μ g/ml. Each curve depicts the running average of 3 experiments. Maximal activity was observed at pH = 9.0, although the quantitative degree of OD-decline was, in general, less than in phosphate-citrate.

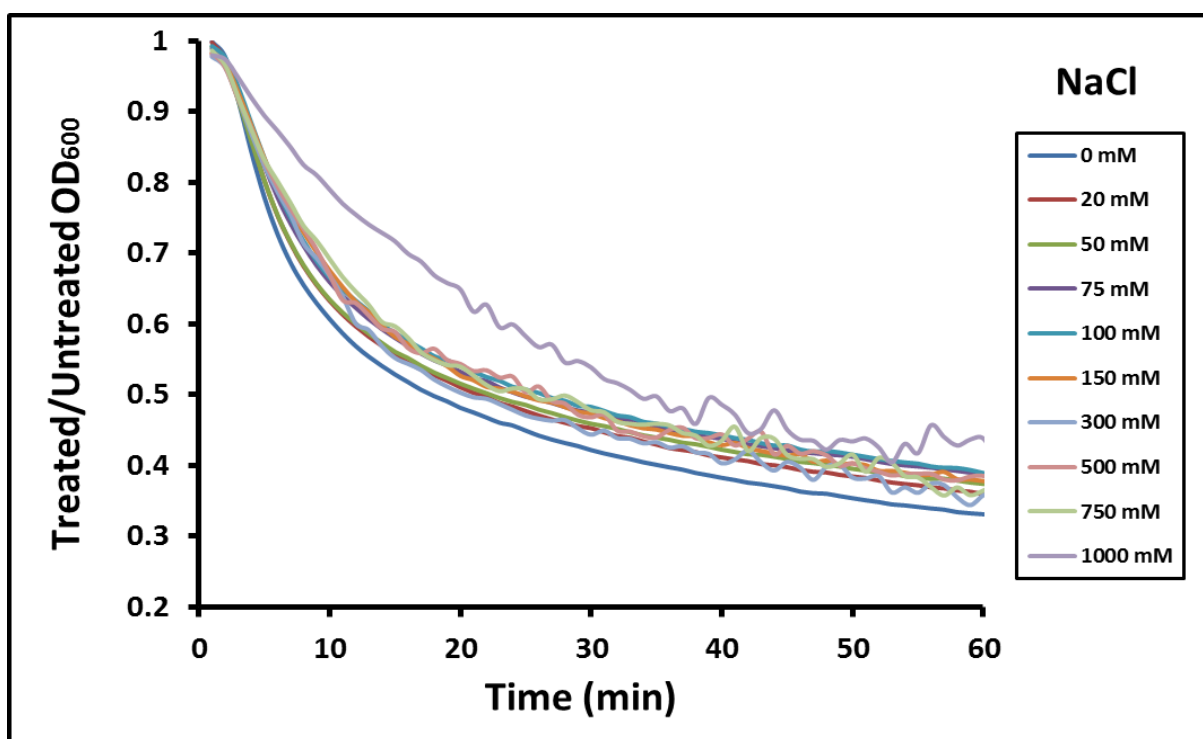
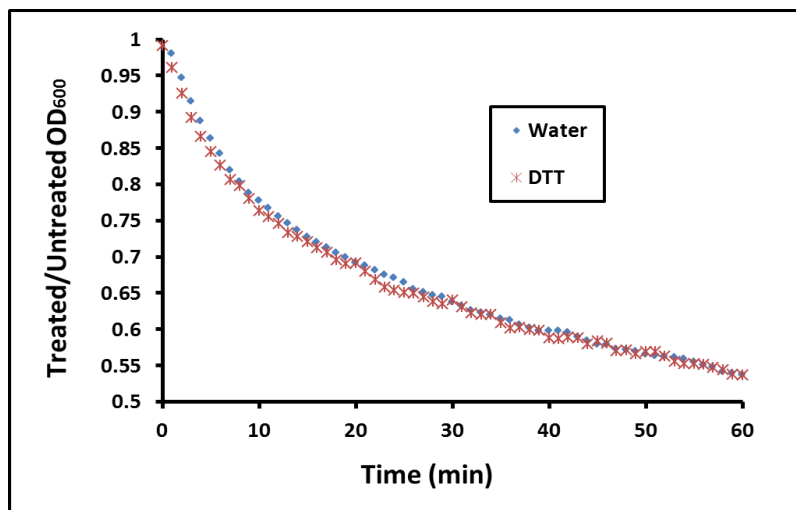


Figure 5.7 Δ PlySs1 NaCl Dependence

S. suis 7711 cells were suspended in phosphate-citrate buffer pH = 7.8 (40/20 mM). NaCl was added to the above concentrations, followed by Δ PlySs1 at 110 μ g/ml. Optical density at 600 nm was observed over 60 min at 37°C. In this figure, the vertical axis represents the treated-untreated OD₆₀₀-ratio for each NaCl concentration, averaged over 3 independent experiments.

(A)



(B)

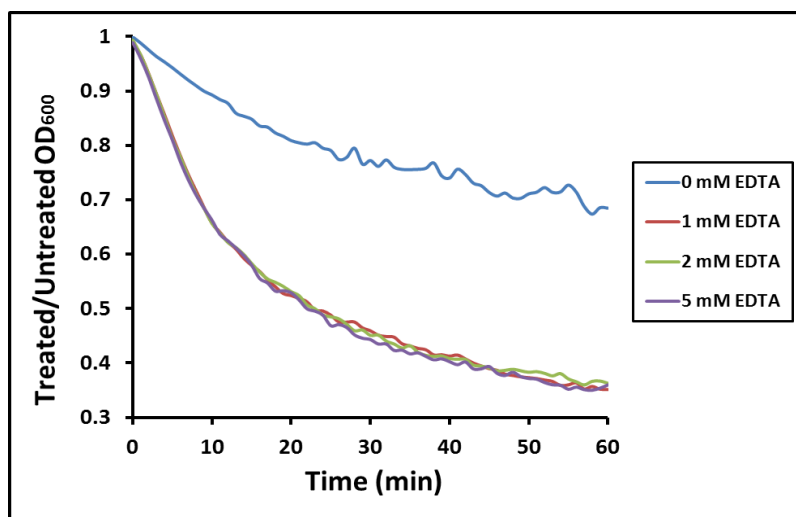


Figure 5.8 Δ PlySs1 DTT and EDTA Susceptibility

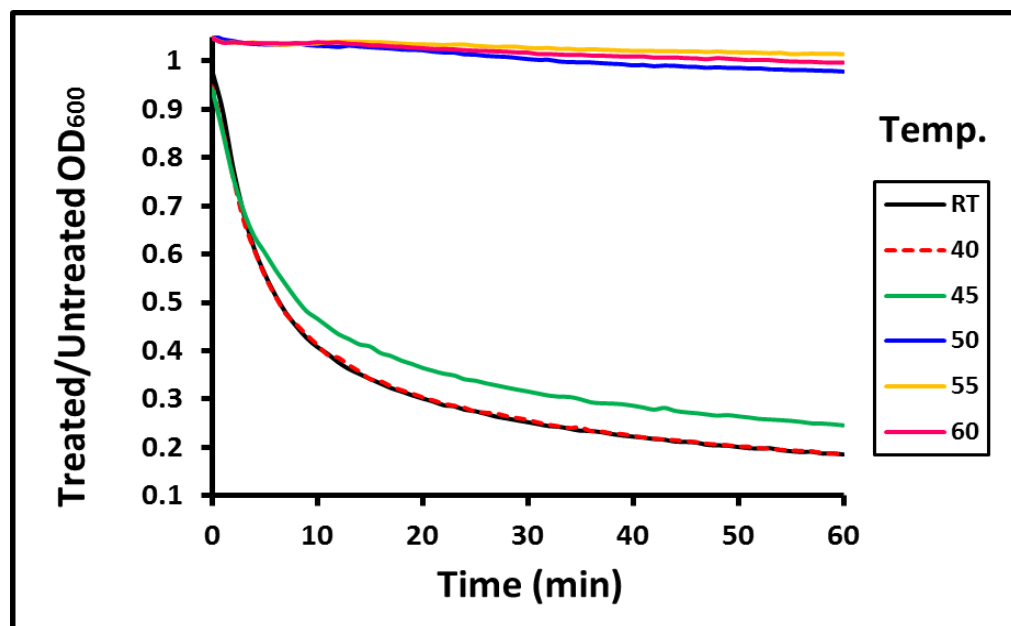
[A] Δ PlySs1 was pre-incubated for 1 hr with 5 mM DTT (a large molar excess) prior to addition to 7711 cells; activity was unchanged. [B] Here, various concentrations of EDTA were included in the buffered suspension of cells prior to addition of Δ PlySs1 (110 μ g/ml lysin). For both images, the vertical axis represents the treated/untreated OD₆₀₀-ratio for each condition, averaged over 3 independent experiments.

The lack of inhibition naturally suggests that Δ PlySs1 does not rely upon divalent cations as cofactors, and several potential mechanisms could underlie the increased activity. A particularly attractive scenario is that the lysin is susceptible to metal-dependent proteases on the *S. suis* surface. Wang et al. previously demonstrated that bacterial proteases desensitize *S. suis* to lysin activity (2009). Under this hypothesis, EDTA would mediate bacteriolysis by inhibiting lysin-degradation and increasing the effective Δ PlySs1 concentration.

The thermal stability of Δ PlySs1 was examined by incubating the enzyme at various elevated temperatures prior to use (the OD-drop experiment itself was always conducted at 37°C). When held at 35°C – 60°C for 30 min, lysin activity was virtually unaffected until 50°C, at which point it was completely abolished (Figure 5.9a). For a 6-hr incubation, a partial decrease in activity was observed at 45°C, while the 40°C sample was unaffected (Figure 5.9b). The latter corresponds to typical porcine body temperature.

Δ PlySs1 activity against *S. suis* and other bacteria. Given the above experiments, the following *optimal* buffering conditions were employed for all further *in vitro* experiments with Δ PlySs1: 20 mM phosphate buffer, pH = 7.8, 2 mM EDTA. A range of lysin concentrations, from 6.5 – 130 μ g/ml, were introduced to live *S. suis* cells in this buffer. Three strains were considered particularly relevant: 7711, the serotype 7 strain that encodes PlySs1; S735, the serotype 2 reference strain; and 7997, a highly virulent serotype 9 strain.

(A)
30 min



(B)
6 hr

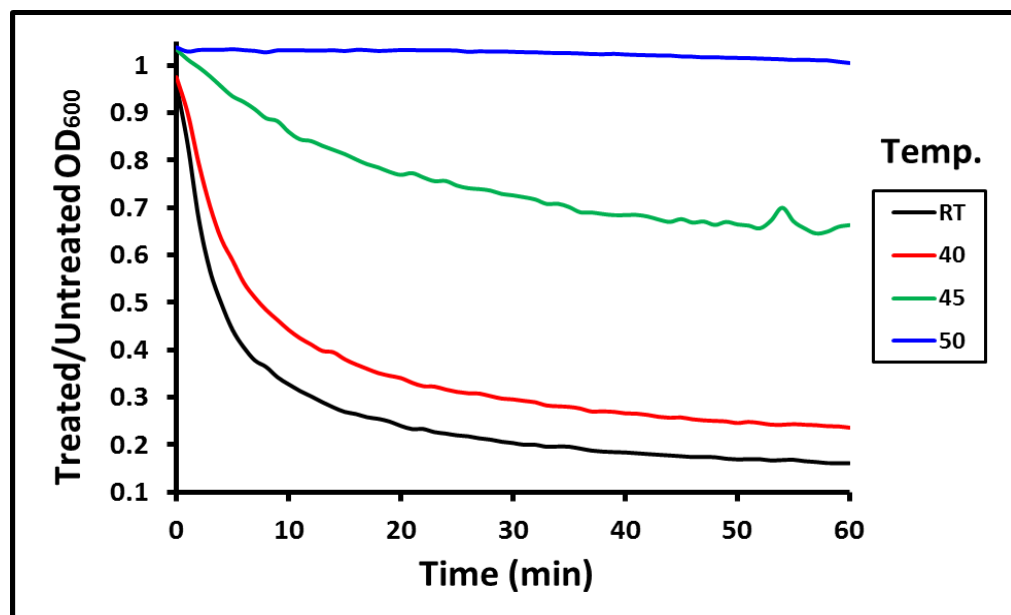


Figure 5.9 Δ PlySs1 Temperature Stability

Figure 5.9, continued

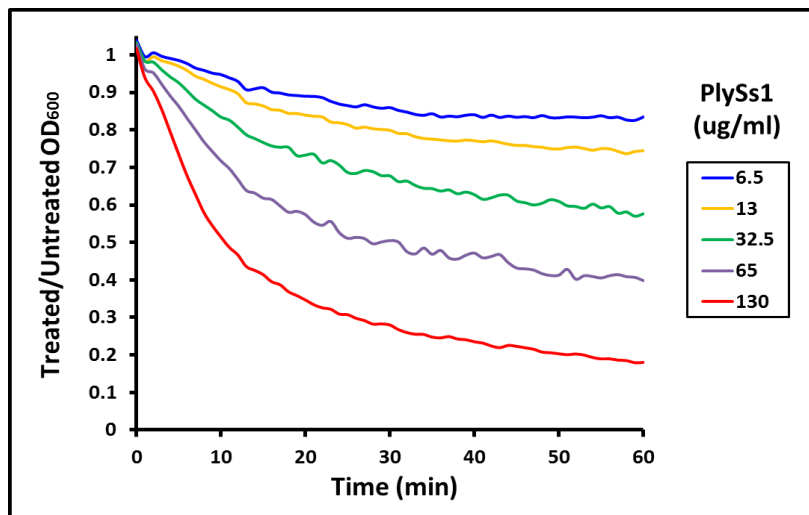
[A] A Δ PlySs1 stock solution was held at each of the above temperature for 30 minutes, followed by addition to 7711 cells (270 μ g/ml final enzyme concentration, final temperature = 37°C, ideal buffering conditions). The curves in this image represent running averages of 3 individual experiments. In each case, complete loss of activity was observed between the 45°C and 50°C samples. The 3 hottest samples show a slightly higher OD₆₀₀ reading than the untreated control due to flocculation of Δ PlySs1 upon denaturation. **[B]** The above experiment was repeated, but with 6 hours of heat-treatment prior to the assay. At this longer incubation time, the 45°C sample showed some loss of activity, though not complete. The 40°C sample maintained essentially native activity.

For each of these strains, the time-dependent OD₆₀₀ response at various ΔPlySs1 dosages is given in Figure 5.10.

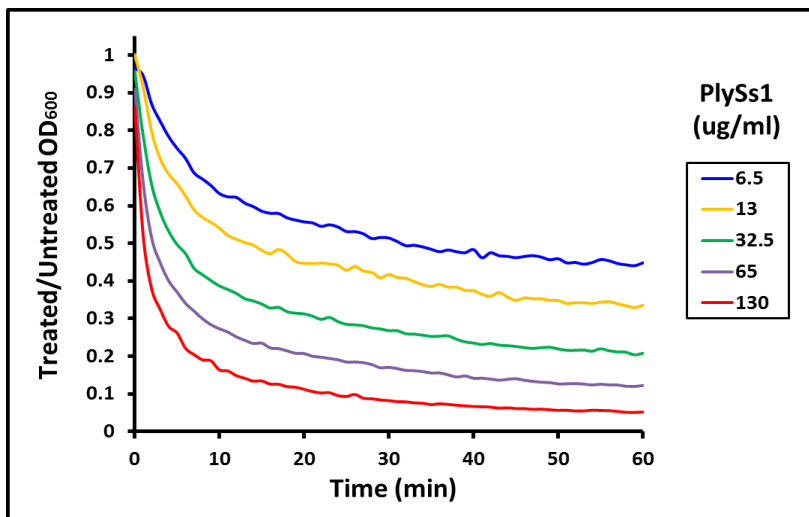
Overall, these data are notable for the relatively high concentration of ΔPlySs1 needed to induce efficient bacteriolysis (> 65 µg/ml). For many phage lysins, only low µg/ml-concentrations are required to lyse their respective bacteria *in vitro* (see Nelson et al, 2001; Schuch et al. 2002; or Chapter 2 of this thesis). At the same time, the higher concentrations needed here are highly commensurate with the amount of LySMP that was needed in that study (in fact, the magnitude of the turbidity-decreases is slightly more favorable here). In terms of bacterial viability, only the highest ΔPlySs1-concentration (130 µg/ml) led to a >90% decrease in CFUs for 7711, S735, and 7997 after 1 hr treatment (Table 5.1). The lysin was also tested against actively-dividing cells in broth culture (strain 7711). Although it delayed bacterial proliferation in a dose-dependent manner (Figure 5.11), these effects were generally mild and ΔPlySs1 could not inhibit *S. suis* growth outright.

ΔPlySs1 was further tested against a panel of 19 other *S. suis* strains of diverse serotypes, as well as other species of Gram-positive bacteria. The same lysin concentrations were used as above. For each dosage, the observed lysis-values after 1 hr are listed in Table 5.2, and the information is summarized graphically in Figure 5.12. All *S. suis* strains demonstrated some degree of susceptibility – for certain isolates, lysis was more pronounced than that observed in Figure 5.10.

7711:
Serotype 7,
(encoding strain)



S735
Serotype 2
(type strain)



7997
Serotype 9

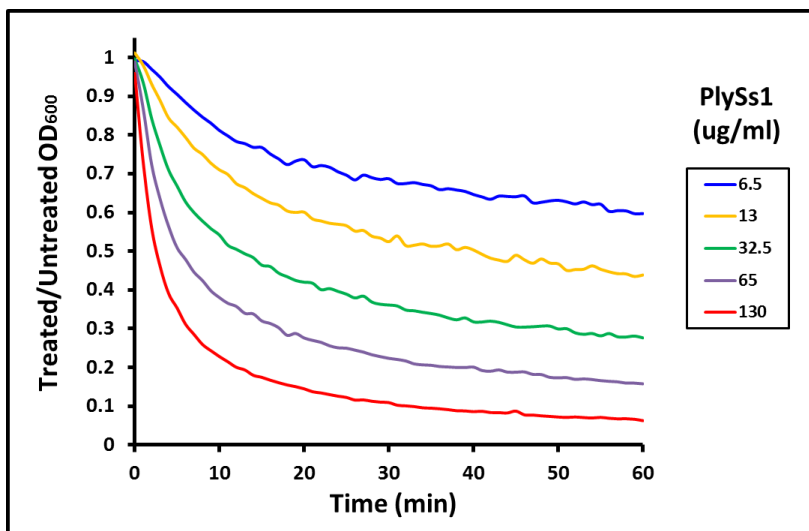


Figure 5.10 Δ PlySs1 Bacteriolytic Activity

Figure 5.10, continued

Depicted here are OD-drop curves for three strains of *S. suis*: 7711, the serotype 7 strain from which PlySs1 was originally cloned (i.e. the host strain); S735, a serotype 2 isolate that is the type-strain for the species; and 7997, a virulent serotype 9 strain. Bacteria were suspended in 20 mM phosphate buffer pH 7.8, 2 mM EDTA (defined as *optimal* conditions). Δ PlySs1 was added to the cells at a range of concentrations (indicated by the inset). For each sample, optical density at 600 nm (vertical axis) was measured over the course of an hour (horizontal axis) at 37°C. In this image, all curves represent running averages of 3 or 4 independent experiments.

Strain	13 µg/ml	130 µg/ml
S735 (ST2)	80.4% – 92.6%	95.4% – 99.5%
7997 (ST9)	16.8% – 30.3%	89.9% – 93.9%
7711 (ST7)	0% – 35.6%	95.3% – 99.2%

Table 5.1 CFU Analysis of Strains 7711, S735, 7997

For two Δ PlySs1 concentrations (130 and 13 µg/ml), CFU analysis was conducted on *S. suis* strains S735, 7997, and 7711 after 1 hr treatment (optimal buffering conditions). In each experiment, the percentage-decrease in CFUs was determined for the treated sample versus the untreated. The range of the values observed (across 3 independent experiments) is reported here for each strain. The serotype of each strain is indicated in parentheses.

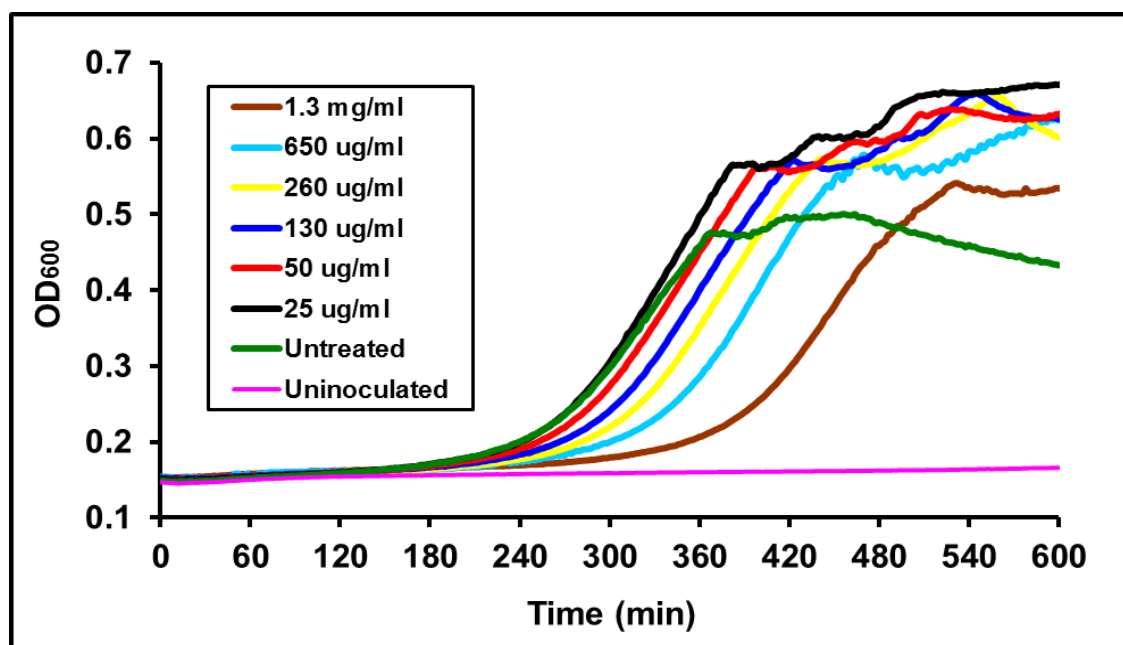


Figure 5.11 Δ PlySs1 Growth Inhibition of *S. suis* 7711

Δ PlySs1 was added at the above final concentrations to a dilute suspension of *S. suis* strain 7711 in BHI broth (see *Materials and Methods* for experimental details). The optical density of each sample was measured continuously overnight in 96-well plate format. Overall, bacterial growth was delayed in a dose-dependent manner. However, for enzyme-concentrations that were sufficient to induce lysis in buffered solutions (130 and 50 μ g/ml), the effect was quite minimal here. Moreover, none of the above Δ PlySs1 concentrations inhibited growth outright—hence, a MIC could not be assigned. For all of the treated samples, one will note that the final optical densities are actually higher than that of the untreated sample. This is an artifact of the accumulation of aggregated bacterial debris that occurred in the presence of lytic enzyme.

Strain	6.5 µg/ml	13 µg/ml	30 µg/ml	65 µg/ml	130 µg/ml
ST13	0.32	0.17	0.04	0.02	0.02
6112 (ST1)	0.14	0.11	0.06	0.02	0.01
ST8	0.25	0.12	0.06	0.03	0.03
6388 (ST1)	0.15	0.13	0.06	0.03	0.02
10 (ST2)	0.29	0.18	0.10	0.05	0.02
8076 (ST9)	0.52	0.40	0.21	0.14	0.04
ST9	0.50	0.30	0.23	0.13	0.05
ST4	0.63	0.47	0.32	0.22	0.12
ST11	0.64	0.47	0.32	0.19	0.07
ST14	0.79	0.57	0.33	0.15	0.06
ST7	0.65	0.47	0.34	0.22	0.11
ST1	0.80	0.34	0.36	0.19	0.06
ST5	0.78	0.59	0.39	0.22	0.10
7197 (ST7)	0.64	0.49	0.39	0.16	0.07
ST6	0.76	0.56	0.40	0.21	0.06
ST3	0.81	0.71	0.48	0.32	0.16
ST2	0.79	0.70	0.49	0.34	0.17
ST10	0.85	0.72	0.55	0.44	0.28
ST12	See Caption				

Table 5.2a Analysis of Other *S. Suis* Strains

Table 5.2a, continued

Various isolates of *S. suis* were exposed (at optimal buffering conditions) to Δ PlySs1 at the above concentrations. The majority of these bacteria are unnamed clinical isolates of the indicated serotype (e.g. ST1, ST2, etc...). For the named strains, the serotype is given in parentheses. The 1-hour treated/untreated OD₆₀₀-ratio is given for each Δ PlySs1 concentration (representing a single experiment), and the strains are listed in the order of decreasing sensitivity. For strain ST12, it was not possible to conduct OD analysis. Upon the addition of Δ PlySs1 (all above concentrations), the cells would rapidly self-adhere and fall out of suspension. This phenomenon was not observed for untreated ST12-cells.

Strain	6.5 µg/ml	13 µg/ml	30 µg/ml	65 µg/ml	130 µg/ml
<i>S. oralis</i> 35037	0.30	0.13	0.08	0.07	0.04
<i>S. agalactiae</i> type II	0.61	0.21	0.11	0.08	0.04
<i>S. dysgalactiae</i> 21597	0.26	0.18	0.12	0.10	0.09
<i>S. pyogenes</i> A486	0.12	0.13	0.13	0.11	0.10
<i>S. pneumoniae</i> R36	0.25	0.22	0.14	0.16	0.12
<i>S. dysgalactiae</i> GGS	0.30	0.27	0.15	0.11	0.14
<i>S. equi</i> 700400	0.48	0.25	0.15	0.07	0.09
<i>S. uberis</i> 27598	0.42	0.23	0.16	0.14	0.12
<i>S. pyogenes</i> D471	0.39	0.27	0.17	0.13	0.09
<i>S. gordonii</i> 10558	0.76	0.32	0.19	0.09	0.06
<i>S. equi</i> 9528	0.66	0.45	0.25	0.19	0.16
<i>L. monocytogenes</i> HER1084	0.63	0.52	0.26	0.14	0.04
<i>S. sanguinis</i> 10556	0.48	0.44	0.28	0.21	0.11
Group E streptococci K131	0.69	0.50	0.33	0.22	0.15
<i>S. sobrinus</i> 6715	0.64	0.48	0.39	0.32	0.23
<i>E. faecium</i> EFSK2	0.85	0.67	0.52	0.32	0.13
<i>S. aureus</i> RN4220	0.89	0.78	0.55	0.31	0.10
<i>S. salivarius</i> 9222	0.80	0.76	0.56	0.53	0.37
<i>S. rattus</i> BHT	0.82	0.84	0.82	0.83	0.79
<i>M. luteus</i> 4698	0.84	0.90	0.83	0.87	0.82
<i>E. faecalis</i> V583	0.98	0.93	0.84	0.71	0.52
<i>B. cereus</i> 14579	0.93	0.92	0.86	0.90	0.86
<i>B. thuringiensis</i> HD73	0.99	0.98	0.93	0.86	0.60
<i>S. mutans</i> U159	0.95	0.99	0.94	0.76	0.85
<i>S. epidermidis</i> HN1292	1.04	1.00	0.96	0.94	0.87
<i>S. agalactiae</i> 090R	0.97	0.99	0.97	0.98	0.93
<i>S. simulans</i> TNK3	0.96	1.00	1.00	1.00	0.96
<i>B. anthracis</i> ΔSterne	1.02	1.03	1.02	0.98	0.90
<i>B. subtilis</i> SL4	1.07	1.05	1.04	1.03	0.96

Table 5.2b Analysis of Other Gram-Positive Bacteria

The experiments summarized in 5.2a were repeated here for various Gram-positive bacterial species other than *S. suis*. As before, 1-hr OD₆₀₀-ratios are given (again, each representing a single experiment).

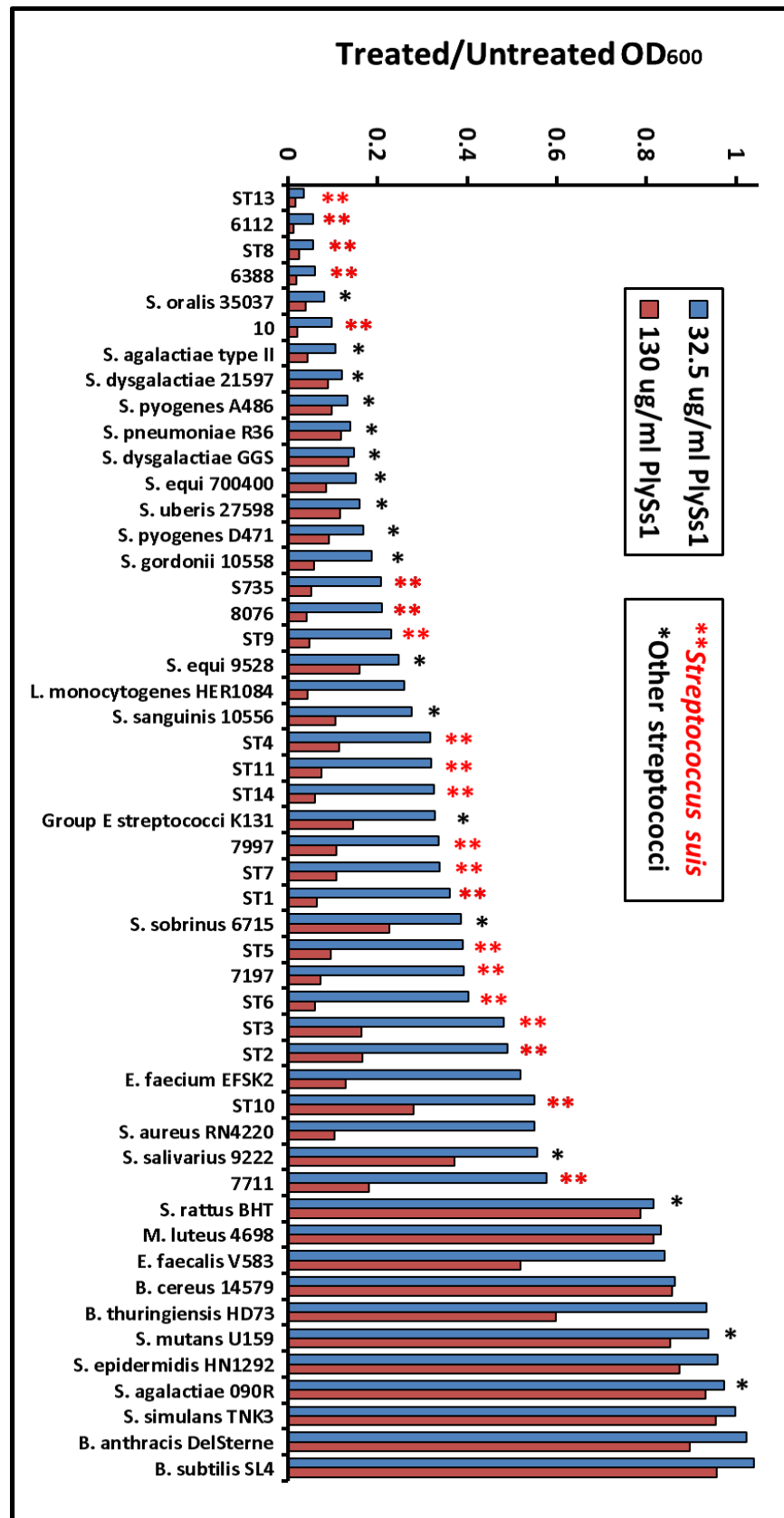


Figure 5.12 Δ PlySs1 Bacterial Strain Panel

Figure 5.12, continued

The information provided in Figure 5.10 and Tables 5.2a and 5.2b is summarized graphically for two Δ PlySs1 concentrations, 130 $\mu\text{g/ml}$ and 32.5 $\mu\text{g/ml}$. In the image, strains of *S. suis* are denoted with double red asterisks and non-suis streptococci are denoted with single black asterisks. The optical density response (treated-versus-untreated OD₆₀₀ ratio) after 1 hr is shown. The reader is referred to Table 5.2a for the serotype definitions of the *S. suis* strains.

Interestingly, many of the non-*suis* streptococci (and even some non-streptococci) also lysed at commensurate enzyme concentrations. Classically, a phage lysin demonstrates a marked decrease in activity when going from within its host species to outside of it. Here, however, a broad range of susceptibility was seen among non-*suis* bacteria, with some demonstrating identical lysis to *S. suis* itself. This phenomenon raises interesting questions regarding the specificity of the dual CPL-7 binding domain of Δ PlySs1.

Comments on ongoing work. Admittedly, the characterization of Δ PlySs1 does not represent a tremendous *theoretical* step forward (given its similarity to the previously-characterized LySMP). And we would prefer to have seen bacteriolytic activity at lower overall concentrations. From a technical perspective, however, Δ PlySs1 can be purified rapidly (in a single chromatographic step), solubly (without additional reducing agent), and in large quantities (100's of mg in a single protocol), all of which offer practical advantages over LySMP.

Although it was not discussed in this thesis – see Chapter 6 for further explanation – PlySs2 has also been purified and tested *in vitro* against the same panel of *S. suis* isolates. Fortunately, the potency of this enzyme seems far greater than that of Δ PlySs1 (i.e. similar bacteriolytic responses at 10-fold lower concentration). Among the two lysins, PlySs2 is clearly the superior enzymatic candidate. In this regard, combination therapy with both lysins (by intranasal

lavage) is currently being tested as a means of preventing colonization/infection of newborn piglets by serotype 2 and 9 strains of *S. suis*. This work is being undertaken by collaborators at the University of Utrecht in the Netherlands (see Acknowledgements).

Overall, an inherent advantage of applying phage lysins to veterinary pathogens is that, ethically, one can rapidly progress to applying the agents in their intended clinical setting. It is true that several previous reports have investigated lysins from, broadly speaking, a veterinary perspective (Hoopes et al. 2009; Celia et al. 2008; Nelson and Fischetti 2004). Nevertheless, the ongoing trials with Δ PlySs1 and PlySs2 represent the first instance of enzybiotic agents being used to treat a bacterial disease in its natural context (as opposed to animal models of human infections). The fact that the progression of this research – from initial identification to *in vivo* experiments – took only slightly more than half a year makes the work particularly exciting. From this perspective, we feel that Δ PlySs1 and PlySs2 are indeed expanding the horizons of the enzybiotics field in a significant way.

ACKNOWLEDGMENTS

This research was funded by NIH/NIAID grants AI057472 and AI11822 to VAF. JES acknowledges the kind support of the NIH MSTP program (Weill Cornell/ Rockefeller/Sloan-Kettering grant GM 07739), as well as the important intellectual contributions of Dr. Gregory Resch and Mr. Daniel Gilmer. Genomic

DNA from strain 89/1591 was generously provided by Dr. Marcelo Gottschalk of the University of Montreal. We would like to acknowledge our valuable collaborators at the University of Utrecht – Dr. Jaap Wagenaar, Dr. Hilde Smith, Mr. Niels Dekker, and Mr. Koen Verstappen, who are conducting ongoing animal trials.

ENDNOTES

1. There was no pre-designed rationale for choosing this particular strain or serotype for functional genomic screening. A variety of clinical isolates were available for this purpose, and 7711 just happened to be the first (out of 4 tested) to successfully yield a lysin. After the identification of PlySs1 (and, concomitantly, PlySs2), no further strains were subject to functional screening.

CHAPTER 6

Future Prospects for Enzybiotic Identification

REFLECTIONS ON THE CURRENT PROJECT

Looking back at the work presented in this thesis, it is tempting to consider how I would expand on the various topics given the opportunity to pursue them further. In the following paragraphs, I will examine briefly for each chapter what I consider the most promising *next steps* for ongoing research. Some of these possibilities were already mentioned in the Discussion sections of the individual chapters. Here, however, I will focus exclusively on the most

practical priorities in the context of an academic research lab (and, specifically, the Laboratory of Bacterial Pathogenesis at Rockefeller).

For Chapter 2, one of the most straightforward extensions would be to apply a similar *in silico* lysin screen to the published genomes of other Gram-positive bacterial species. As the results of the chapter demonstrated, this information can be valuable both in the identification of novel enzymatic candidates, as well in the generation of new hypotheses regarding the biology of the host bacteria (with *C. perfringens*, for example, involving BCN5 externalization). In the present work, however, the search strategy was essentially manual. In theory, however, the same set of steps (i.e. the algorithm) could be conducted computationally, expediting the process tremendously. A program such as this could readily be applied to all known genomes – as well as new sequencing data, assembled or unassembled – with the goal of generating a single, organized *database of lysins*.

Experimentally speaking, one of the most intriguing next steps does not actually involve work with PlyCM itself, but rather the non-viral “lysins” in the *C. perfringens* genomes. In particular, the BCN5-associated lysins are a tempting area for investigation. Although this possibility was mentioned in the chapter as more of an afterthought, its potential implications (i.e. secretion-by-suicide) and the relative ease with which the hypothesis could be tested make it an extremely attractive project. Since these genes are already plasmid-encoded, it should be straightforward to introduce them into a *C. perfringens*/*E. coli* shuttle plasmid

while maintaining natural function. Knockouts or mutations of the lysin-like ORF could then be constructed; the plasmid would be reintroduced to *C. perfringens*, which would be observed phenotypic differences in response to UV-radiation.

In contrast, the work in Chapter 3 could probably be considered a closed topic from a theoretical perspective. This not to say that E-LASLs are no longer relevant... to the contrary, the technique is still widely used within the Laboratory of Bacterial Pathogenesis for screening genomes and metagenomes. It is just that – from a methodological standpoint – any developments of the E-LASL protocol would likely just represent incremental adjustments rather than major advances.

In this regard, a more relevant issue involves the actual screens that are conducted once the libraries are synthesized. Phages encode numerous biotechnologically-relevant proteins other than lysins (Schoenfeld et al. 2010). From a screening perspective, the biggest issue is how to readily identify their associated properties within transformed clones. While a bacteriolytic or hemolytic phenotype is visually straightforward, the same might not be true for a colony that expresses, say, an interesting phage polymerase. Overall, as more and more screening techniques are devised, the utility of the E-LASL can only expand.

For Chapter 4, the most pressing avenue for ongoing research is abundantly clear: applying the protocol to larger and more diverse viral

metagenomes so that the holin-based method can be exploited for its full worth. The 26 enzymes identified here demonstrate the validity of the technique, and several possess interesting sequences that could merit future purification and biochemical analysis. At the same, the true value of the method can only be realized in the identification of novel enzymes that would not have been approachable by genomic techniques alone.

Based on the extreme abundance of phage in the biosphere and the results of previous viral metagenomic sequencing studies, there is good reason to believe that we have only seen the *tip of the iceberg* when it comes to the diversity of proteins that phages encode. These studies have been remarkable for the lack of similarity between the new metagenomic sequences and genetic information already in the databases (Edwards and Rohwer 2005). Personally, I am optimistic that the same holds true specifically for lytic enzymes. In this regard, our laboratory has been in contact with other research groups about potential collaborations, so that we can apply our functional techniques to existing phage libraries that have only been examined via sequencing to date.

Finally – regarding the *S. suis* lysins of Chapter 5 – it is admittedly difficult to surmise where the work presented here will lead in the future. Much of this is dependent on the success of ongoing *in vivo* trials. At this point, we are cautiously hopeful that combination treatment with PlySs1 and PlySs2 will have a positive impact on experimentally-infected pigs. In a small cohort of animals, lysin treatment prevented clinical progression of disease in 4/5 subjects, while

4/5 untreated subjects experienced fulminant lethal infection (Dekker and Wagennar, personal communications). At the same time, more data analysis and further experimental trials are pending, and these will assuredly affect future decision making on how intensely PlySs1 and PlySs2 will be pursued.

It is also important to emphasize that, of the two *S. suis* lysins, PlySs1 actually represents the *less promising* enzybiotic candidate. Although only the characterization PlySs1 is presented in this thesis, it is the weaker enzyme quantitatively. PlySs2 demonstrates similar anti-streptococcal activity, but at ~10-fold lower concentrations than PlySs1 (values that are more representative of the typical potency of phage lysins). The reason that PlySs2 data is not presented in this thesis is purely logistical. The PlySs1 and PlySs2 enzymes were initially identified and cloned around the same time (autumn/winter 2009). Without prior knowledge of their behavior, I chose to focus on the characterization of PlySs1, while similar work on PlySs2 was undertaken by Mr. Daniel Gilmer (a rotation student in the Laboratory of Bacterial Pathogenesis, who has since joined as a permanent member). Mr. Gilmer has assumed PlySs2 as part of his own doctoral research, and is actively pursuing its characterization. We plan to co-author a manuscript that combines the *in vitro* characterization of both enzymes.

THE FUTURE OF GENOMICS AND METAGENOMICS

Looking beyond the specific work conducted here, it is also valuable to consider other potential developments that might impact lysin identification in

the future. Overall, the ideal lysin screen (from an enzybiotic perspective) would be one that combines the large number of enzymes encoded by metagenomic samples with the specificity that individual lysins demonstrate toward their host bacteria. One possible solution could involve screens in which genomic DNA from numerous strains of a single pathogen is pooled to form a *multi-genomic* library. Each strain would likely contain integrated prophage, such that a combination of diverse strains should encode a large collection of lysins. A functional screen of this library would generate many hits, like a metagenomic screen, with the exception that all enzymes would target the same bacteria *de facto*.

An alternative multi-genomic approach could involve large-scale proviral induction. A collection of strains could be grown in liquid culture and treated with an agent such as mitomycin or phosphomycin (Ryan and Hébert 2009). This would induce lysogen activation, and the resultant viral particles would be released into the culture supernatant. These could readily be purified by a combination of filtration, nuclease treatment (to remove free bacterial DNA/RNA), and polyethylene glycol precipitation. The DNA from a pooled lysogen library could then be subject to a single functional screen. This method, in fact, is currently being employed with success by other members of the Fischetti Laboratory (Schuch and Pelzak, personal communications).

Comparing these two functional multi-genomic approaches, the main advantage of the former is its ability to identify all proviral lysins, including ones

from cryptic prophages and those that do not induce well *in vitro*. At the same time, considering the much larger size of bacterial genomes, the lysin-encoding clones would represent a much smaller proportion of the overall library population. With induced phages, by contrast, the total diversity of lysins might be diminished somewhat, but the desired transformants would be relatively abundant within the library.

Of course, both of these methodologies assume that whole-genome sequencing is beyond the reach (and cost) of a typical academic laboratory. While this technology is vastly more accessible than a decade ago (500+ phage and 1,000+ bacteria been sequenced to date), the preceding statement still largely rings true. For instance, at the time of writing, the in-house cost of a single run of Solexa high-throughput sequencing at the Rockefeller University core facility is upwards of \$1,000 per sample. Although affordable for a limited number of isolates, this figure is still likely excessive when numerous bacterial or phage isolates are considered. And this does not even include the genomic-assembly process from the short sequencing reads, or the time needed to mine the assembled genomes for lysin-encoding ORFs (in this regard, the sort of computational algorithm mentioned above would be valuable). At the present time, while published genomes are valuable sources of enzymes, a search-for-lysins in itself probably does not merit the cost of sequencing numerous additional strains.

Fortunately, high-throughput sequencing technology is still considered to be in its infancy, and the associated efficiency and cost are only expected to become more favorable in the future. Commentators have suggested the eventual possibility of the \$1 bacterial genome (Ussery et al. 2009). In this circumstance, the sequencing of numerous genomes would definitely seem warranted, even if it is only to identify a single class of proteins (lysins or otherwise). The same holds true, of course, for metagenomic sequencing. For metagenomics, in fact, sequencing advances could prove even more valuable due to the extra complexity associated with assembling metagenomes beyond the short initial sequencing reads, which – from a perspective of cloning recombinant proteins – are not useful (Schoenfeld 2010).

All of this is not to say that DNA sequencing can completely replace functional screening as a method for identifying new enzymes. No matter how powerful the technology becomes, a sequence is only able to identify a protein that demonstrates homology to something that has already been characterized. To find something completely novel, functional screening is still a necessity. Ultimately, in this regard, the two approaches are vital complements to one another. Whenever a novel protein is uncovered, the vast database of existing sequences makes it all the more likely that homologues can be recognized immediately, facilitating rapid progression beyond the initial discovery.

ENZYBIOTICS BEYOND PHAGE LYSINS

For this discussion to be complete, it is important to emphasize that phage lytic enzymes are not the only molecules that can be included under the umbrella of *enzybiotics*. The two terms are often used interchangeable, and understandably so. Phage lysins have generated considerably more attention than any other class of enzyme as far as antibiotic potential is concerned. Rigorously speaking, however, any enzyme with antimicrobial properties could be considered an enzybiotic. These other proteins will be considered briefly here, so that the field – and ongoing developments within it – can be appreciated in their entirety.

First, bacteriophages encode enzymes other than lysins proper that have been examined as potential antibacterial agents. Like the lysins, some of these other proteins function as peptidoglycan hydrolases. It was mentioned briefly in Chapter 1, for instance, that hydrolase motifs are often present within proteins of the phage tail assembly (Kanamaru et al. 2004; Kenny et al. 2004; Piuri and Hatfull 2006) or even the head (Moak and Molineux 2004). Presumably, these structural enzymes facilitate the initial injection of viral DNA. Although it is not their natural purpose, recent evidence has shown that these enzymes can lead to Gram-positive lysis when added exogenously. Rashel et al. identified a tail-associated protein from the genome of an *S. aureus* phage that contains two putative lytic motifs; when individually expressed and purified, both domains induced staphylococcal death (2008).

Despite these findings, no tail enzymes have ever been identified during a functional screen of a phage genome. Several reasons could underline this discrepancy. For instance, some enzymes might require the presence of the entire macromolecular tail assembly for proper activity. Moreover, structural lysins often represent component domains within much larger polypeptides; in the context of a functional screen, these domains would not be expressed individually and the bulk proteins could prove difficult to obtain due to their size. Nevertheless, if future research could reliably harness the enzymatic activity of tail lysins, it would represent another promising avenue within the enzybiotics field.

Outside of their enzymatic activity, phage tail structures are additionally remarkable for their ability to traverse the other components of the bacterial cell envelope, including the outer membrane of Gram-negative bacteria. In this regard, various Gram-negative species encode protein complexes (“monocins”) that kill closely-related organisms through a cell-envelope depolarization effect (Zink et al. 1995; Strauch et al. 2001; Jabrane et al. 2002). These multimeric structures resemble isolated phage-tail assemblies, and are believed to represent proviral remnants that have been co-opted by the host. In a recent study, Williams et al. demonstrated that, by exchanging a component gene of one such complex with the tail fiber genes of another phage, they could engineer complexes with species-targeted activity (2008). While not examined specifically,

the study raises interesting questions about potential synergies between these complexes and recombinant lysins.

Other than peptidoglycan hydrolases, several other classes of phage-encoded enzymes have been examined for their overall antibacterial activity. For instance, some phages are known to produce depolymerases that digest the exopolysaccharide of bacterial capsules and biofilm matrices. (Although not discussed at length in this thesis, the dynamics of biofilm communities are thought to play a key role in pathogenesis of certain infections, as well as in the bacterial contamination of medical and industrial devices.) These enzymes may be associated with the actual viral particle or synthesized intracellularly during phage infection (Donlan 2009). Although phage-associated depolymerases have been known for some time (Eklund and Wyss 1962; Sutherland 1967), it is more recently that these enzymes have been proposed as recombinant agents for altering the course of bacterial growth and infection (Glonti et al. 2009; Lu and Collins 2007).

Another extracellular molecular complex with an important role in bacterial pathogenesis is the arabanogalactan layer of mycobacteria. Although this genus is considered a Gram-positive actinobacterium evolutionarily, it possesses a unique cell-envelope structure in which covalently-linked arabanogalactan and mycolic acid layers overlay the peptidoglycan. While mycobacteriophage have been studied extensively (Pedulla et al. 2003), it is never before been addressed whether specific phage-encoded proteins are responsible

for attacking these layers during lysis. Recently, however, Payne et al. identified a conserved arabanogalactan esterase that fulfills this role and is encoded adjacently to the lysin within the phages' genomes (2009). Although the authors did not demonstrate that exogenous enzyme was bacteriolytic (either by itself or in combination with lysin), their work raises the possibility that enzybiotics might still be possible against a group of pathogens for whom they were previously assumed to be irrelevant.

Although unrelated to host lysis, a variety of other phage proteins (many poorly described) exert an intracellular antibacterial effect (Liu et al. 2004; Sau et al. 2008). They are believed to be anti-host factors that allow the viruses to disrupt bacterial physiology following infection. From a perspective of drug-development, these proteins (not necessarily enzymes) do face the additional challenge of cytoplasmic delivery. Limited experimental evidence, however, also suggests that phages encode other poorly-defined proteins that are bactericidal from the outside. For instance, functional screening of the *B. cereus* Bcp-1 phage genome revealed two distinct clones that prevented the growth of *Bacillus* cells in soft-agar overlays. One clone encoded a typical modular lysin (PlyB), while the other encoded a short, difficult-to-purify protein termed KOA (for Killer of Anthrax). The latter is similar to only a limited number of hypothetical proteins from several other *Bacillus* phage genomes (Schuch, unpublished observations).

While it is half the size of standard Gram-positive lysins and does not contain any traditional sequences common to major lysin families, KOA

nevertheless possesses lytic activity against a range of *Bacillus cereus* organisms, including *B. anthracis*. It remains unclear what roles PlyB and KOA play (independently or in conjunction) during Bcp-1 infection, as the identification of this protein represents a fortuitous side-effect of the screening process. Overall, it is difficult to predict what sort of antibacterial compounds (enzymatic or otherwise) might be encoded by global phage, mainly because such a relative few have undergone functional analysis.

At the same time, phages are not the only genetic entities that can encode potential enzybiotic agents. As mentioned several times throughout this thesis, bacteria likewise encode autolytic proteins that mediate activities such as growth, division, and sporulation (Vollmer et al. 2008b). Some autolysins – like pneumococcal LytA and the *C. perfringens* enzymes discussed here – demonstrate high architectural homology to phage lysins, while others are more divergent. Although phage lysins have received more attention, various reports have documented the ability of recombinant autolysins to act as lytic agents against Gram-positive organisms (for instance, Dhalluin et al., 2005; Fukushima et al., 2008; Yokoi et al, 2008). And while their roles differ from those of phage lysins, the process of identifying and developing autolysins as antibacterial agents would follow essentially the same principals.

In fact, the categories of peptidoglycan hydrolases with enzybiotic potential extends even beyond these proteins. Many bacteria also encode proteins and peptides (broadly known as bacteriocins) that exert an antibiotic

effect on other species, often closely related ones. Of these molecules, several notable examples function through a cell wall-lytic mechanism¹. The most prominent example is lysostaphin, a poly-functional hydrolase originally identified in *Staphylococcus simulans* that targets rival staphylococci (Kumar 2008). In light of growing resistance to traditional antibiotics, lysostaphin has received considerable attention as a possible weapon against MRSA and non-MRSA isolates of *S. aureus*. Comparable enzymes, namely millericin B and zoocin A, have been isolated from streptococcal strains (Beukes et al. 2000; Akesson et al. 2007).

Going even further, other bacterial enzymes (glucanases and chitinases) have evolved that digest the cell wall of competing fungal species (Salazar and Asenjo 2007). Overall, biotechnological interest in these proteins stems more for their ability to manipulate yeast in laboratory settings. Nevertheless, given the existence of various fungal pathogens, these enzymes are still worth noting from an enzybiotics perspective. They are also worth considering evolutionarily, as mycolytic enzymes share certain structural and biochemical properties with their glycosyl hydrolase cousins among the peptidoglycan hydrolases (Veiga-Crespo and Villa 2010). All considered, the above proteins serve as important reminders that the enzybiotics field covers more than phage lysins, and necessitates that researchers cast a large net in search of novel antimicrobial agents.

CONCLUDING REMARKS

In the field of phage-lysin research (and enzybiotics, in general), one observation is unmistakably clear: between established strategies of enzyme identification and novel techniques, there will be no shortage of candidate molecules in the foreseeable future. As methods and technology continue to advance, what is already a dense field is only likely to become more crowded. Perhaps a greater challenge than merely identifying phage lysins, in fact, will be the ability to compare them systematically on a protein level. Enzyme kinetics, thermal and pH tolerance, immunogenicity, *in vivo* half-life, and biodistribution are only a few of quantities that could vary from one lysin to the next and, ultimately, they are the factors that will determine an enzyme's therapeutic promise.

Unfortunately, none of these properties can be measured (as of yet) with a mere shotgun screen or nucleotide sequence, as protein purification and old-fashioned pharmacological analyses are still required. In this regard, the enzybiotics field is in a unique position, as cloning technologies have created a pronounced gap between our ability to identify an enzyme-of-interest and the effort it takes to study it in detail and move it through the drug-development process. Nevertheless, if and when the first phage lysin progresses to human use—and an increasing body of experimental evidence suggests that day is coming—there will already be a large reserve of similar agents set to join it in the fight against infectious disease.

ENDNOTES

1. As mentioned in the text, the term *bacteriocin* is commonly used to describe bacteria-encoded proteins and peptides with antibacterial activity. The classification of peptidoglycan hydrolases as bacteriocins, however, is an issue on which a definitive consensus does not yet exist. Under some schemata, virtually all antibacterial proteins encoded by bacteria themselves should be considered bacteriocins (Heng and Tagg 2006). In this case, enzymes such as lysostaphin and zoosin A are categorized as class IIIa bacteriocins. At the same time, other systems reserve the term *bacteriocin* for distinct classes of peptides that function through a non-hydrolytic mechanism (Cotter et al. 2005). In either case, it should be emphasized that such schemata function only as organizational tools, and do not affect the *in vivo* function or biotechnological potential of these proteins.

REFERENCES

- Abedon ST, ed.** 2010. The 'Nuts and Bolts' of Phage Therapy. *Curr. Pharm. Biotechnol.* 11:1-86.
- Abedon ST.** 2006. Phage Ecology. In: *The bacteriophages*, Calendar R, ed. Oxford University Press, 2nd edition, 37-48.
- Abshire TG, Brown JE, and Ezzell JW.** 2005. Production and validation of the use of gamma phage for identification of *Bacillus anthracis*. *J. Clin. Microbiol.* 43:4780-4788.
- Abutencia CB, Wyborski DL, Garcia JA, et al.** 2006. Environmental whole-genome amplification to access microbial populations in contaminated sediments. *Appl. Environ. Microbiol.* 72:3291-3301.
- Ackerman H and Abedon ST.** 2001. Bacteriophage names 2000. *The Bacteriophage Ecology Group*, www.phage.org/names.htm.
- Agu CA, Klein R, Lengler J, et al.** 2007. Bacteriophage-encoded toxins: the lambda-holin protein causes caspase-independent non-apoptotic death of eukaryotic cells. *Cell. Microbiol.* 9:1753-1765.
- Akesson M, Dufour M, Sloan GL, and Simmonds RS.** 2007. Targeting of streptococci by zoocin A. *FEMS Microbiol. Lett.* 270:155-161.

- Alakomi HL, Skyttä E, Saarela M, Mattila-Sandholm T, Latva-Kala K, and Helander IM.** 2000. Lactic acid permeabilizes gram-negative bacteria by disrupting the outer membrane. *Appl. Environ. Microbiol.* 66:2001-2005.
- Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ.** 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, and Lipman DJ.** 1997. Gapped Blast and PSI-Blast: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389-3402.
- Angly FE , Felts B, Breitbart M et al.** 2006. The marine viromes of four oceanic regions. *PLoS Biol.* 4: e368.
- Arends JP and Zanen HC.** 1988. Meningitis caused by *Streptococcus suis* in humans. *Rev. Infect. Dis.* 10:131-137.
- Arendt EK, Daly C, Fitzgerald GF, and van de Guchte M.** 1994. Molecular characterization of lactococcal bacteriophage Tuc2009 and identification and analysis of genes encoding lysin, a putative holin, and two structural proteins. *Appl. Environ. Microbiol.* 60:1875-1883.
- Avery, OT, MacLeod CM, Maclyn McCarty M.** 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types:

induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. J. Exp. Med. 79:137-158.

Baker BJ, Comolli LR, Dick GJ, et al. 2010. Enigmatic, ultrasmall, uncultivated Archaea. Proc. Natl. Acad. Sci. USA. 107:8806-8811

Baker JR, Liu C, Dong S, and Pritchard DG. 2006. Endopeptidase and glycosidase activities of the bacteriophage B30 lysin. Appl. Environ. Microbiol. 72:6825-6828.

Balogh B, Jones JB, Iriarte FB, and Momol MT. 2010. Phage therapy for plant disease control. Curr. Pharm. Biotechnol. 11:48-57.

Baums CG and Valentin-Weigand P. 2009. Surface-associated and secreted factors of *Streptococcus suis* in epidemiology, pathogenesis and vaccine development. Anim. Health Res. Rev. 10:65-83.

Becker SC, Dong S, Baker JR, Foster-Frey J, Pritchard DG, and Donovan DM. 2009. LysK CHAP endopeptidase domain is required for lysis of live staphylococcal cells. FEMS Microbiol. Lett. 294:52-60.

Bell PJ, Sunna A, Gibbs MD, Curach NC, Nevalainen H, and Bergquist PL. 2002. Prospecting for novel lipase genes using PCR. Microbiology. 148:2283-2291.

Bendtsen JD, Nielsen H, von Heijne G, and Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340:783-795.

- Bernhardt TG, Roof WD, and Young R. 2000.** Genetic evidence that the bacteriophage ϕ X174 lysis protein inhibits cell wall synthesis. *Proc. Natl. Acad. Sci. USA.* 97:4297-4302.
- Bernhardt TG, Wang IN, Struck DK, and Young R. 2001.** A protein antibiotic in the phage Qbeta virion: diversity in lysis targets. *Science.* 292:2326-2329.
- Berry J, Summer EJ, Struck DK, and Young R. 2008.** The final step in the phage infection cycle: the Rz and Rz1 lysis proteins link the inner and outer membranes. *Mol. Microbiol.* 70:341-351.
- Beukes M, Bierbaum G, Sahl H, and Hastings JW. 2000.** Purification and partial characterization of a murein hydrolase, millericin B, produced by *Streptococcus milleri* NMSCC 061. *Appl. Environ. Microbiol.* 66:23-28.
- Birge EA. 2006.** Transduction. In: *Bacteria and bacteriophage genetics*. Springer, 5th edition, 305-326.
- Birkeland NK. 1994.** Cloning, molecular characterization, and expression of the genes encoding the lytic functions of lactococcal bacteriophage phi LC3: a dual lysis system of modular design. *Can. J. Microbiol.* 40:658-665.
- Bläsi U and Young R. 1996.** Two beginnings for a single purpose: the dual-start holins in the regulation of phage lysis. *Mol. Microbiol.* 21:675-682.

Blume V, Luque I, Vela AI, Borge C, Maldonado A, Domínguez L, Tarradas C, and Fernández-Garayzábal JF. 2009. Genetic and virulence-phenotype characterization of serotypes 2 and 9 of *Streptococcus suis* swine isolates. Int. Microbiol. 12:161-166.

Boizet B, Lahbib-Mansais Y, Dupont L, Ritzenthaler P, and Mata M. 1990. Cloning, expression and sequence analysis of an endolysin-encoding gene of *Lactobacillus bulgaricus* bacteriophage mv1 Gene. 94:61-67.

Bon J, Mani N, Jayaswal RK. 1997. Molecular analysis of lytic genes of bacteriophage 80 alpha of *Staphylococcus aureus*. Can. J. Microbiol. 43: 612-616.

Borysowski J, Weber-Dabrowska B, and Górski A. 2006. Bacteriophage endolysins as a novel class of antibacterial agents. Exp. Biol. Med. 231:366-377.

Bourhy P, Frangeul L, Couvé E, Glaser P, Saint Girons I, and Picardeau M. 2005. Complete nucleotide sequence of the LE1 prophage from the spirochete *Leptospira biflexa* and characterization of its replication and partition functions. J. Bacteriol. 187:3931-3940.

Boyd EF and Brüssow H. 2002. Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. Trends Microbiol. 10:521-529.

Brady SF. 2007. Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. *Nat. Protoc.* 2:1297-1305.

Brady SF, Chao CJ, Handelsmen J, and Clardy J. 2001. Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. *Org. Lett.* 3:1981-1984.

Brady SF, Simmons L, Kim JH, and Schmidt EW. 2009. Metagenomic approaches to natural products from free-living and symbiotic organisms. *Nat. Prod. Rep.* 26:1488-1503.

Breitbart M, Felts B, Kelley S, Mahaffy JM, Nulton J, Salamon P, and Rohwer F. 2004. Diversity and population structure of a near-shore marine-sediment viral community. *Proceedings of the Royal Society of London Series B.* 271:565-574.

Breitbart M, Hewson I, Felts B, Mahaffy JM, Nulton J, Salamon P, and Rohwer F. 2003. Metagenomic analyses of an uncultured viral community from human feces. *J. Bacteriol.* 185:6220-6223.

Breitbart M, Rohwer F, and Abedon ST. 2005. Phage Ecology and Bacterial Pathogenesis. In: *Phages: Their role in bacterial pathogenesis and biotechnology*, Waldor MK, Freidman DI, and Adhya SL, eds. American Society for Microbiology Press, 1st edition, 66-91.

Breitbart MI, Salamon P, Anderson B, Mahaffy JM, Segall AM, Mead D, Azam F, and Rowher F. 2002. Genomic analysis of uncultured marine viral communities. *Proc. Natl. Acad. Sci. USA.* 99:14250-14255.

Briers Y, Volckaert G, Cornelissen A, Lagaert S, Michiels CW, Hertveldt K, and Lavigne R. 2007. Muralytic activity and modular structure of the endolysins of *Pseudomonas aeruginosa* bacteriophages ϕ KZ and EL. *Mol. Microbiol.* 65:1334-1344.

Bronfenbrenner J and Muckenfuss R. 1927. Studies on the bacteriophage of d'Herelle: VII. The mechanism of lysis of dead bacteria in the presence of bacteriophage. *J. Exp. Med.* 45:887-909.

Broudy TB and Fischetti VA. 2003. In vivo lysogenic conversion of Tox(-) *Streptococcus pyogenes* to Tox(+) with Lysogenic Streptococci or free phage. *Infect. Immun.* 71:3782-3786.

Brown ER and Cherry WB. 1955. Specific identification of *Bacillus anthracis* by means of a variant bacteriophage. *J. Infect. Dis.* 96:34-39.

Bryant AE and Stevens DL. 1997. The pathogenesis of gas gangrene. In: *The clostridia: molecular biology and pathogenesis*, Rood JI et al., eds. Academic Press, 1st edition.

- Buist G, Steen A, Kok J, and Kuipers OP.** 2008. LysM, a widely distributed protein motif for binding to (peptido)glycans. *Mol. Microbiol.* 68:838-847
- Caldentey J and Bamford DH.** 1992. The lytic enzyme of the *Pseudomonas* phage phi 6: purification and biochemical characterization. *Biochim. Biophys. Acta* 1159:44-50.
- Calendar R and Inman R.** 2005. Phage Biology. In: *Phages: Their role in bacterial pathogenesis and biotechnology*, Waldor MK, Freidman DI, and Adhya SL, eds. American Society for Microbiology Press, 1st edition, 18-36.
- Camiade E, Peltier J, Bourgeois I, et al.** 2010. Characterization of Acp, a peptidoglycan hydrolase of *Clostridium perfringens* with N-acetylglucosaminidase activity that is implicated in cell separation and stress-induced autolysis. *J. Bacteriol.* 192:2373-2384.
- Campbell A.** 2003. The future of bacteriophage biology. *Nat. Rev. Genet.* 4:471-477.
- Canchaya C, Fournous G, and Brüssow H.** 2004. The impact of prophages on bacterial chromosomes. *Mol. Microbiol.* 53:9-18.
- Carlton RM.** 1999. Phage therapy: past history and future prospects. *Arch. Immunol. Ther. Ex.* 47:267-274.

- Carman RJ, Sayeed S, Li J, Genheimer CW, Hiltonsmith MF, Wilkins TD, and McClane BA.** 2008. *Clostridium perfringens* toxin genotypes in the feces of healthy North Americans. *Anaerobe*. 14:102-8.
- Casas V and Rohwer F.** 2007. Phage metagenomics. *Method. Enzymol.* 421:259-268.
- Cascales E, Buchanan SK, Duché D, et al.** 2007. Colicin biology. *Microbiol. Mol. Biol. Rev.* 71:158-229.
- Casjens S.** 2003. Prophages and bacterial genomics: what have we learned so far? *Mol. Microbiol.* 49:277-300.
- Celia L, Nelson KD, and Kerr DE.** 2008. Characterization of a bacteriophage lysin (Ply700) from *Streptococcus uberis*. *Vet. Microbiol.* 130:107-117.
- Chandry PS, Moore SC, Boyce JD, Davidson BE, and Hillier AJ.** 1997. Analysis of the DNA sequence, gene expression, origin of replication and modular structure of the *Lactococcus lactis* lytic bacteriophage sk1. *Mol. Microbiol.* 26:49-64.
- Chakraborty T, Huhle B, Bergbauer H, and Goebel W.** 1986. Cloning, expression, and mapping of the *Aeromonas hydrophila* aerolysin gene determinant in *Escherichia coli* K-12. *J. Bacteriol.* 167:368-374.

Chanishvili N. 2009. A literature review of the practical application of bacteriophage research. Eliava Institute of Bacteriophage, Microbiology and Virology, Tbilisi, Georgia.

Chanishvili N, Chanishvili T, Tediashvili M, and Barrow PA. 2001. Phages and their application against drug-resistant bacteria. J. Chem. Technol. Biot. 76:689-699.

Chauthaiwale VM, Therwath A, and Deshpande VV. Bacteriophage lambda as a cloning vector. Microbiol. Rev. 56:577-91.

Chen C, Tang J, Dong W, et al. 2007. A glimpse of streptococcal toxic shock syndrome from comparative genomics of *S. suis* 2 Chinese isolates. PLoS One. 2:e315.

Cheng Q, Nelson D, Zhu S, and Fischetti VA. 2005. Removal of group B streptococci colonizing the vagina and oropharynx of mice with a bacteriophage lytic enzyme. Antimicrob. Agents Chemother. 49:111-117.

Cheng Q and Fischetti VA. 2007. Mutagenesis of a bacteriophage lytic enzyme PlyGBS significantly increases its antibacterial activity against group B streptococci. Appl. Microbiol. Biotechnol. 74:1284-1291.

Chopra AK, Houston CW, Peterson JW, and Jin GF. 1993. Cloning, expression, and sequence analysis of a cytotoxic enterotoxin gene from *Aeromonas hydrophila*. Can. J. Microbiol. 39:513-523.

Cotter PD, Hill C, and Ross RP. 2005. Bacteriocins: developing innate immunity for food. Nat. Rev. Microbiol. 3:777-788.

Courchesne NM, Parisien A, and Lan CQ. 2009. Production and application of bacteriophage and bacteriophage-encoded lysins. Recent Pat. Biotechnol. 3:37-45.

Courtois S, Cappellano CM, Ball M, Francou FX, Normand P, and Helynck G. 2003. Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. Appl. Environ. Microbiol. 69:49-55.

Craig JW, Chang FY, and Brady SF. 2009. Natural products from environmental DNA hosted in *Ralstonia metallidurans*. ACS Chem. Biol. 4:23-28.

Craig JW, Chang FY, Kim JH, Obiajulu SC, and Brady SF. 2010. Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. Appl. Environ Microbiol. 76:1633-1641.

Crick FHC, Barnett L, Brenner S, and Watts-Tobin RJ. 1961. General nature of the genetic code for proteins. Nature. 192:1227-1232.

Culley AI, Lang AS, and Suttle CA. 2006. Metagenomic analysis of coastal RNA virus communities. *Science*. 312:1795-1798.

Daniel A, Euler C, Collin M, Chahales P, Gorelick KJ, and Fischetti VA. 2010. Synergism between a novel chimeric lysin and oxacillin protects against infection by methicillin-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 54:1603-1612.

Daniel R. 2005. The metagenomics of soil. *Nat. Rev. Microbiol.* 3:470-478.

Delisle AL, Barcak GJ, and Guo M. 2006. Isolation and expression of the lysis genes of *Actinomyces naeslundii* phage Av-1. *Appl. Environ. Microbiol.* 72:1110-1117.

Delwart EL. 2007. Viral metagenomics. *Rev. Med. Virol.* 17:115-131.

Deutsch S, Guezenec S, Piot M, Foster S, and Lortal S. 2004. Mur-LH, the broad-spectrum endolysin of *Lactobacillus helveticus* temperate bacteriophage phi-0303. *Appl. Environ. Microbiol.* 70:96-103.

Devriese LA, Ceyssens K, Hommez J, Kilpper-Balz R, and Schleifer KH. 1991. Characteristics of different *Streptococcus suis* ecovars and description of a simplified identification method. *Vet. Microbiol.* 26:141-150.

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

d'Herelle F. 1917. Sur un microbe invisible antagoniste des bacilles dysenteriques. *C. R. Acad. Sci. (Paris)* 165:373-375.

d'Hérelle F. 1921. Le bactériophage. Son rôle dans l'immunité. Monographies de l'Institut Pasteur, Masson et Cie.

Díaz E, López R, and García JL. 1992. EJ-1, a temperate bacteriophage of *Streptococcus pneumoniae* with a Myoviridae morphotype. *J. Bacteriol.* 174:5516-5525.

Díaz E, López R, and García JL. 1991. Chimeric pneumococcal cell wall lytic enzymes reveal important physiological and evolutionary traits. *J. Biol. Chem.* 266:5464-5471.

Díaz E, López R, and García JL. 1990. Chimeric phage-bacterial enzymes: a clue to the modular evolution of genes. *P. Natl. Acad. Sci. USA.* 87:8125-8129.

Donadio S, Maffioli S, Monciardini P, Sosio M, and Jabes D. 2010. Antibiotic discovery in the twenty-first century: current trends and future perspectives. *J. Antibiot.* In press, doi:10.1038/ja.2010.74.

- Donato JJ, Moe LA, Converse BJ, Smart KD, Berklein FC, McManus PS, and Handelsman J.** 2010. Metagenomic analysis of apple orchard soil reveals antibiotic resistance genes encoding predicted bifunctional proteins. *Appl. Environ. Microbiol.* 76:4396-4401
- Donlan RM.** 2009. Preventing biofilms of clinically relevant organisms using bacteriophage. *Trends Microbiol.* 17:66-72.
- Donovan DM, Dong S, Garrett W, Rousseau G, Moineau S, and Pritchard DG.** 2006. Peptidoglycan hydrolase fusions maintain their parental specificities. *Appl. Environ. Microbiol.* 72:2988-2996.
- Donovan DM and Foster-Frey J.** 2008. LambdaSa2 prophage endolysin requires Cpl-7-binding domains and amidase-5 domain for antimicrobial lysis of streptococci. *FEMS Microbiol. Lett.* 287:22-33.
- Doughty CC and Hayashi JA.** 1962. Enzymatic properties of a phage-induced lysin affecting group A streptococci. *J. Bacteriol.* 83:1058-1068.
- Doughty CC and Mann JA.** 1967. Purification and properties of a bacteriophage-induced cell wall peptidase from *Staphylococcus aureus*. *J. Bacteriol.* 93:1089-1095.
- Dunn JJ and Studier FW.** 1983. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J. Mol. Biol.* 166:477-535.

- Dupuy B, Mani N, Katayama S, and Sonenshein AL.** 2005. Transcription activation of a UV-inducible *Clostridium perfringens* bacteriocin gene by a novel sigma factor. *Mol. Microbiol.* 55:1196-1206.
- Eaton MD and Bayne-Jones S.** 1934. Bacteriophage therapy. Review of the principles and results of the use of bacteriophage in the treatment of infections. *J. Am. Med. Assoc.* 103:1769-1776.
- Edwards RA and Rohwer F.** 2005. Viral metagenomics. *Nat. Rev. Microbiol.* 3:504-510
- Eklund C and Wyss O.** 1962. Enzyme associated with bacteriophage infection. *J. Bacteriol.* 84:1209-1215
- Engle HW.** 1978. Mycobacteriophages and phage typing. *Ann. Microbiol. (Paris).* 129:75-90.
- Entenza JM, Loeffler JM, Grandgirard D, Fischetti VA, and Moreillon P.** 2005. Therapeutic effects of bacteriophage Cpl-1 lysin against *Streptococcus pneumoniae* endocarditis in rats. *Antimicrob. Agents Chemother.* 49:4789-4792.
- Epple HJ, Mankertz J, Ignatius R, Liesenfeld O, Fromm M, Zeitz M, Chakraborty T, and Schulzke JD.** 2004. *Aeromonas hydrophila* beta-hemolysin induces active chloride secretion in colon epithelial cells (HT-29/B6). *Infect. Immun.* 72:4848-4858.

Eschbach E, Scharsack JP, John U, and Medlin LK. 2001. Improved erythrocyte lysis assay in microtitre plates for sensitive detection and efficient measurement of haemolytic compounds from ichthyotoxic algae. *J. Appl. Toxicol.* 21:513-519.

Eschenfeld WH, Stols L, Rosenbaum H, Khambata ZS, Quaite-Randall E, Wu S, Kilgore DC, Trent JD, and Donnelly MI. 2001. DNA from uncultured organisms as a source of 2-5-Diketo-D-gluconic acid reductases. *Appl. Environ. Microbiol.* 67:4206-4214.

Euler CW. 2010. The role of lysogenic bacteriophage in virulence and survival of *Streptococcus pyogenes*. Doctoral dissertation, The Rockefeller University.

Facklam R. 2002. What happened to the streptococci: overview of taxonomic and nomenclature changes. *Clin. Microbiol. Rev.* 15:613-630

Fauquet CM, Mayo MA, Maniloff J, Desselberger U, and Ball LA, eds. 2005. *Virus taxonomy*, 2nd ed. Elsevier Academic Press, London.

Felsenstein J. 1989. PHYLIP - Phylogeny inference package (version 3.2). *Cladistics.* 5:164-166.

Feng Y, Zhang H, Ma Y, and Gao GF. 2010. Uncovering newly emerging variants of *Streptococcus suis*, an important zoonotic agent. *Trends Microbiol.* 18:124-131.

Ferrer M, Golyshina O, Beloqui A, and Golyshin PN. 2007. Mining enzymes from extreme environments. *Curr. Opin. Microbiol.* 10:207-214.

Ferrer M, Martinez-Abarca F, and Golyshin PN. 2005. Mining genomes and 'metagenomes' for novel catalysts. *Curr. Opinion Biotechnol.* 16:588-593.

Field HI, Buntan D, and Done JT. 1954. Studies on piglet mortality. I. Streptococcal meningitis and arthritis. *Veterinary Record.* 66:454-455.

Fierer N, Breitbart M, Nulton J, Salamon P, Lozupone C, Jones R, Robeson M, Edwards RA, Felts B, Rayhawk S, Knight R, Rohwer F, and Jackson RB. 2007. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Appl. Environ. Microbiol.* 73:7059-7066.

Finkbeiner SR, Allred AF, Tarr PI, Klein EJ, Kirkwood CD, and Wang D. 2008. Metagenomic analysis of human diarrhea: viral detection and discovery. *PLoS Pathog.* 4:e10000011.

Finn RD, Mistry J, Tate J, et al. 2010. The Pfam protein families database. *Nuc. Acids Res. Database Issue* 38:D211-222.

Fischetti VA. 2005a. Bacteriophage lytic enzymes: novel anti-infectives. *Trends Microbiol.* 13:491-496.

- Fischetti VA.** 2005b. The use of phage lytic enzymes to control bacterial infections. In: Bacteriophages: biology and applications, Kutter E and Sulakvelidze A, eds. CRC Press, 1st edition, 318-332.
- Fischetti VA.** 2008. Bacteriophage lysins as effective antibacterials. *Curr. Opin. Microbiol.* 11:393-400.
- Fischetti VA.** 2010. Bacteriophage endolysins: A novel anti-infective to control Gram-positive pathogens. *Int. J. Med. Microbiol.* In press, doi:10.1016/j.ijmm.2010.04.002.
- Fischetti VA, Gotschlich EC, and Bernheimer AW.** 1971. Purification and physical properties of group C streptococcal phage-associated lysin. *J. Exp. Med.* 133:1105-1117.
- Fischetti VA, Nelson D, and Schuch R.** 2006. Reinventing phage therapy: are the parts greater than the sum? *Nat. Biotechnol.* 24:1508-1511.
- Fivas M, Abrami L, Tsitrin Y, and van der Goot FG.** 2001. Aerolysin from *Aeromonas hydrophila* and related toxins. *Curr. Top. Microbiol. Immun.* 257:35-52.
- Foley S, Bruttin A, and Brüssow H.** 2000. Widespread distribution of a group I intron and its three deletion derivatives in the lysin gene of *Streptococcus thermophilus* bacteriophages. *J. Virol.* 74:611-618.

Fukushima T, Kitajima T, Yamaguchi H, Ouyang Q, Furuhashi K, Yamamoto H, Shida T, and Sekiguchi J. 2008. Identification and characterization of novel cell wall hydrolase CwlT: a two-domain autolysin exhibiting n-acetylmuramidase and DL-endopeptidase activities. *J. Biol. Chem.* 283:11117-11125.

Fujii Y, Nomura T, Kanzawa H, Kameyama M, Yamanaka H, Akita M, Setsu K, and Okamoto K. 1998. Purification and characterization of enterotoxin produced by *Aeromonas sobria*. *Microbiol. Immunol.* 42:703-714.

García E, García JL, García P, Arrarás A, and Sánchez-Puelles JM. 1988. Molecular evolution of lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages. *P. Natl. Acad. Sci. USA.* 85:914-918.

García E, García JL, García P, Arrarás A, and Sánchez-Puelles JM, and R. López. 1990. Modular organization of the lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages. *Gene.* 86:81-88

García E, García JL, Ronda C, García P, and López R. 1985. Cloning and expression of the pneumococcal autolysin gene in *Escherichia coli*. *Mol. Gen. Genet.* 201:225-230.

Garvert KJ, Saedi MS, and Ito J. 1986. Nucleotide sequence of *Bacillus* phage phi 29 genes 14 and 15: homology of gene 15 with other phage lysozymes. *Nucleic Acids Res.* 14:10001-10008.

Garnier T and Cole ST. 1986. Characterization of a bacteriocinogenic plasmid from *Clostridium perfringens* and molecular genetic analysis of the bacteriocin-encoding gene. *J. Bacteriol.* 168:1189-1196.

Garnier T and Cole ST. 1988. Complete nucleotide sequence and genetic organization of the bacteriocinogenic plasmid, pIP404, from *Clostridium perfringens*. *Mol. Microbiol.* 2:607-614.

Gasset M. 2010. Bacteriophage holins and their membrane-disrupting activity. In: *Enzybiotics: antibiotic enzymes as drugs and therapeutics*, Villa TG and Veiga-Crespo P, eds. John Wiley and Sons, 1st edition, 123-148.

Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, Rondon MR, Clardy J, Goodman RM, and Handelsman J. 2002. Isolation of antibiotics turbomycin A and B from a metagenomic library of soil DNA. *Appl. Environ. Microbiol.* 68:4301-4306.

Glonti T, Chanishvili N, and Taylor PW. 2009. Bacteriophage-derived enzyme that depolymerizes the alginic acid capsule associated with cystic fibrosis isolates of *Pseudomonas aeruginosa*. *J. Appl. Microbiol.* 108:695-702.

- Górski A, Borysowski J, Miedzybrodzki R, and Weber-Dabrowska B.** 2007. Bacteriophages in medicine. In: Bacteriophage: genetics and molecular biology, McGrath S and van Sinderen D, eds. Caister Academic Press, 1st edition, 125-158.
- Górski A, Miedzybrodzki R, Borysowski J, Weber-Dabrowska B, Lobočka M, Fortuna W, Letkiewicz S, Zimecki M, and Filby G.** 2009. Bacteriophage therapy for the treatment of infections. *Curr. Opin. Investig. Drugs.* 10:766-774.
- Gottschalk M.** 2004. Porcine *Streptococcus suis* strains as potential sources of infection in humans: an underdiagnosed problem in North America? *J. Swine Health Prod.* 12:197-199.
- Grandgirard D, Loeffler JM, Fischetti VA, and Leib SL.** 2008. Phage lytic enzyme Cpl-1 for antibacterial therapy in experimental pneumococcal meningitis *J. Infect. Dis.* 197:1519-1522.
- Gratia A and Rhodes B.** 1923. Action due principe lytique sur les émulsions de Staphylocoques vivants et de Staphylocoques tués. *Comptes Rendus Hebdomadaires des Séances et Mémoires de la Société de Biologie et des ses Filiales.* 89:1171-1172.
- Green BD and Keller M.** 2006. Capturing the uncultivated majority. *Curr. Opin. Biotech.* 17:236-240.

- Guiral S, Mitchell TJ, Martin B, and Claverys J.** 2005. Competence-programmed predation of noncompetent cells in the human pathogen *Streptococcus pneumoniae*: genetic requirements. P. Natl. Acad. Sci. USA. 102:8710-8715.
- Gupta A, Oppenheim A, and Chaudhary VK.** 2005. Phage display: a molecular fashion show. In: Phages: Their role in bacterial pathogenesis and biotechnology, Waldor MK, Freidman DI, and Adhya SL, eds. American Society for Microbiology Press, 1st edition, 415-429.
- Guzman LM, Belin D, Carson MJ, and Beckwith J.** 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J. Bacteriol. 177:4121-4130.
- Haesebrouck F, Pasmans F, Chiers K, Maes D, Ducatelle R, and Decostere A.** 2004. Efficacy of vaccines against bacterial diseases in swine: what can we expect? Vet. Microbiol. 100:255-268.
- Hamada H, Arakawa T, and Shiraki K.** 2009. Effect of additives on protein aggregation. Curr. Pharm. Biotechnol. 10:400-407.
- Handelsman J.** 2004. Metagenomics: application of genomics to uncultured microorganisms. Microbiol. Mol. Biol. Rev. 68:669-685.

Harel J, Martinez G, Nassar A, et al. 2003. Identification of an inducible bacteriophage in a virulent strain of *Streptococcus suis* serotype 2. Infect. Immun. 71:6104-6108.

Hata TS. 2010. Biotechnology in the high school classroom: A resource guide and lab manual for exploring biotechnology 44 minutes at a time (Volume 1). CreateSpace, 1st edition.

Hatful GF. 2008. Bacteriophage Genomics. Curr. Opin. Microbiol. 11:447-453.

Häusler T. 2006 Viruses vs. superbugs: a solution to the antibiotics crisis? Macmillan Science, 1st edition.

Hendrix RW. 2002. Bacteriophages: evolution of the majority. Theor. Popul. Biol. 61:471-480.

Hendrix RW. 2005. Bacteriophage evolution and the role of phages in host evolution. In: Phages: Their role in bacterial pathogenesis and biotechnology, Waldor MK, Freidman DI, and Adhya SL, eds. American Society for Microbiology Press, 1st edition, 55-65.

Heng NCK and Tagg JR. 2006. Nat. Rev. Microbiol. 4, doi: 10.1038/nrmicro1273-c1.

- Henrich B, Binishofer, and Bläsi U.** 1995. Primary structure and functional analysis of the lysis genes of *Lactobacillus gasseri* bacteriophage phi adh. J. Bacteriol. 177:723-732.
- Hermann SR, Miller JA, O'Neil S, Tsao TT, Harding RM, and Dale JL.** 2000. Single-primer amplification of flanking sequences. Biotechniques. 29:1176-1178.
- Hermoso JA, García JL, and García P.** 2007. Taking aim on bacterial pathogens: from phage therapy to enzybiotics. Curr. Opin. Microbiol. 10:461-672.
- Hershey AD, and Chase M.** 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. J. Gen. Physiol. 36:39-56.
- Higgins R and Gottschalk M.** 1990. An update on *Streptococcus suis* identification. J. Vet. Diagn. Invest. 2:249-252.
- Higgins R, Gottschalk M, Boudreau M, Lebrun A, and Henrichsen J.** 1995. Description of six new *Streptococcus suis* capsular types. J. Vet. Diagn. Invest. 7:405-406
- Hirono I and Aoki T.** 1993. Cloning and characterization of three hemolysin genes from *Aeromonas salmonicida*. Microb. Pathog. 15:269-282.

Hirono I, Aoki T, Asao T, and Kozaki S. 1992. Nucleotide sequence and characterization of haemolysin genes for *Aeromonas hydrophila* and *Aeromonas sobria*. Microb. Pathog. 13:433-46.

Holden MT, Hauser H, Sanders M, et al. 2009. Rapid evolution of virulence and drug resistance in the emerging zoonotic pathogen *Streptococcus suis*. PLoS One. 4:e6072.

Hoopes JT, Stark CJ, Kim HA, Sussman DJ, Donovan DM, and Nelson DC. 2009. Use of a bacteriophage lysin, PlyC, as an enzyme disinfectant against *Streptococcus equi*. Appl. Environ. Microbiol. 75:1388-1394.

Howard SP, and Buckley JT. 1986. Molecular cloning and expression in *Escherichia coli* of the structural gene for the hemolytic toxin aerolysin from *Aeromonas hydrophila*. Mol. Gen. Genet. 204:289-295.

Howard SP, Garland WJ, Green MJ, and Buckley JT. 1987. Nucleotide sequence of the gene for the hole-forming toxin for *Aeromonas hydrophila*. J. Bacteriol. 169:2869-2871.

Horgan M, O'Flynn G, Garry J, Cooney J, Coffey A, Fitzgerald GF, Ross RP, and McAuliffe O. 2009. Phage lysin LysK can be truncated to its CHAP domain and retain lytic activity against live antibiotic-resistant staphylococci. Appl. Environ. Microbiol. 75:872-874.

Housby JN and Mann NT. 2009. Phage therapy. *Drug Discovery Today*. 14:536-540.

Hu S, Kong J, Kong W, Guo T, and Ji M. 2010. Characterization of a novel LysM domain from *Lactobacillus fermentum* bacteriophage endolysin and its use as an anchor to display heterologous proteins on the surfaces of lactic acid bacteria. *Appl. Environ Microbiol.* 76:2410-2418.

Husslein V, Huhle B, Jarchau T, Lurz R, Goebel W, and Chakraborty T. 1998. Nucleotide sequence and transcriptional analysis of the aerCaerA region of *Aeromonas sobria* encoding aerolysin and its regulatory region. *Mol. Microbiol.* 2:507-517.

Hyman P and Abedon ST. 2010. Bacteriophage host range and bacterial resistance. *Adv. Appl. Microbiol.* 70:217-248.

Inouye M and Tsugita A. 1966. The amino acid sequence of the T4 bacteriophage lysozyme. *J. Mol. Biol.* 22:193-196.

Jabrane A, Sabri A, Compère P, Jacques P, Vandenberghe I, van Beeumen J, and Thenart P. 2002. Characterization of serracin P, a phage-tail-like bacteriocin, and its activity against *Erwinia amylovora*, the fire blight pathogen. *Appl. Environ. Microbiol.* 68:5704-5710.

- Jayaswal RK, Lee Y, and Wilkerson B.** 1990. Cloning and expression of a *Staphylococcus aureus* gene encoding a peptidoglycan hydrolase activity. J. Bacteriol. 172:5783-5788.
- Jedrzejewski M.** 2001. Pneumococcal virulence factors: structure and function. Microbiol. Mol. Biol. Rev. 65:187-207.
- Johnson RP, Gyles CL, Huff WE, Ojha S, Huff GR, Rath NC, and Donoghue AM.** 2008. Bacteriophages for prophylaxis and therapy in cattle, poultry and pigs. Anim. Health Res. Rev. 9:201-215.
- Johnson TL, Abendroth J, Hol WGJ, and Sandkvist M.** 2006. Type II secretion: from structure to function. FEMS Microbiol. Lett. 255:175-186.
- Jones BL and Wilcox MH.** 1995. *Aeromonas* infections and their treatment. J. Antimicrob. Chemoth. 35:453-461.
- Kanamaru S, Leiman PG, Kostychenko VA, Chipman PR, Mesyanzhinov VV, Arisaka F, and Rossman MG.** 2002. Structure of the cell-puncturing device of bacteriophage T4. Nature. 415:553-557.
- Khan AA, Kim E, and Cerniglia CE.** 1998. Molecular cloning, nucleotide, and expression in *Escherichia coli* of a hemolytic toxin (aerolysin) gene from *Aeromonas trota*. Appl. Environ. Microbiol. 64:2473-2478.

Karam JD, et al. 1994. Molecular biology of bacteriophage T4. American Society for Microbiology Press, 1st edition. ASM Press, Washington.

Kashige N, Nakashima Y, Miake F, and Watanabe K. 2000. Cloning, sequence analysis, and expression of *Lactobacillus casei* phage PL-1 lysis genes. Arch. Virol. 145:1521-1534.

Kenny JG, McGrath S, Fitzgerald GF, and van Sinderen D. 2004. Bacteriophage Tuc2009 encodes a tail-associated cell wall-degrading activity. J. Bacteriol. 186:3480-3491.

Kilpper-Bälz R and Schleifer KH. 1987. *Streptococcus suis* sp. nov., nom. rev. Int. J. System. Bacteriol. 37:160-162.

Kim JH, Feng Z, Bauer JD, Kallifidas D, Calle PY, and Brady SF. 2010. Cloning large natural product gene clusters from the environment: piecing environmental DNA gene clusters back together with TAR. Biopolymers. 93:833-44.

Kim W, Salm H, and Geider K. 2004. Expression of bacteriophage phiEa1h lysozyme in *Escherichia coli* and its activity in growth inhibition of *Erwinia amylovora*. Microbiology. 150:2707-2714.

King RW, Bauer JD, and Brady SF. 2009. An environmental DNA-derived type II polyketide biosynthetic pathway encodes the biosynthesis of the pentacyclic polyketide erdacin. Angew Chem. Int. Ed. Engl. 48:6257-6261.

Korndörfer IP, Danzer J, Schmelcher M, Zimmer M, Skerra A, and Loessner

MJ. 2006. The crystal structure of the bacteriophage PSA endolysin reveals a unique fold responsible for specific recognition of *Listeria* cell walls. J. Mol. Biol. 364:678-689.

Korndörfer IP, Kanitz A, Danzer J, et al. 2008. Structural analysis of the L-

alanoyl-D-glutamate endopeptidase domain of *Listeria* bacteriophage endolysin

Ply500 reveals a new member of the LAS peptidase family. Acta Cryst. D. 64:644-650.

Kotik M. 2009. Novel genes retrieved from environmental DNA by polymerase

chain reaction: current genome-walking techniques for future metagenome

applications. J Biotechnol. 144:75-82.

Krause RM. 1957. Studies on bacteriophages of hemolytic streptococci. I. Factors

influencing the interaction of phage and susceptible host cell. J. Exp. Med.

106:365-384.

Krause RM. 1958. Studies on the bacteriophages of hemolytic streptococci. II.

Antigens released from the streptococcal cell wall by a phage-associated lysin. J.

Exp. Med. 108:803-821.

Krueger AP and Scribner EJ. 1941. The bacteriophage: its nature and its

therapeutic use. J. Am. Med. Assoc. 116:2160-2167.

Krupovic M and Bamford DH. 2007. Putative prophages related to lytic tailless marine dsDNA phage PM2 are widespread in the genomes of aquatic bacteria. BMC Genomics. 8:236.

Krupovic M, Daugelavicius R, and Bamford DH. A novel lysis system in PM2, a lipid-containing marine double-stranded DNA bacteriophage. Mol. Microbiol. 64:1635-1648.

Kuijper EJ, Coignard B, Tüll P, et al. 2006. Emergence of *Clostridium difficile*-associated disease in North America and Europe. Clin. Microbiol. Infect. Suppl. 6:2-18.

Kumar JK. 2008. Lysostaphin: an antistaphylococcal agent. Appl. Microbiol. Biotechnol. 80:555-561.

Lawrence GW. 1997. The pathogenesis of enteritis necroticans. In: The clostridia: molecular biology and pathogenesis, Rood JI et al., eds. Academic Press, 1st edition, 197-210.

Lederberg J and Lederberg EM. 1952. Replica plating and indirect selection of bacterial mutants. J. Bacteriol. 63:399-406.

Li L, McCorkle¹ SR, Monchy S, Taghavi¹ S, and van der Lelie D. 2009. Bioprospecting metagenomes: glycosyl hydrolases for converting biomass. Biotechnology for Biofuels 2009, 2:10.

Lima-Mendez G, van Helden J, Toussaint A, and Leplae R. 2008. Prophinder: a computational tool for prophage prediction in prokaryotic genomes.

Bioinformatics. 24:863-865.

Lin LJR, Liao CC, Chen YR, and Chak KF. 2009. Induction of membrane permeability in *Escherichia coli* mediated by lysis protein of the ColE7 operon.

FEMS Microbiol. Lett. 298:85-92.

Little JW. 2005. Lysogeny, prophage induction, and lysogenic conversion. In: Phages: Their role in bacterial pathogenesis and biotechnology, Waldor MK, Freidman DI, and Adhya SL, eds. American Society for Microbiology Press, 1st edition, 37-54.

Little JW and Mount DW. 1982. The SOS regulatory system of *Escherichia coli*. Cell. 29:11-22.

Liu J, Dehbi M, Moeck G, et al. 2004. Antimicrobial drug discovery through bacteriophage genomics. Nat. Biotechnol. 22:185-191.

Loeffler JM and Fischetti VA. 2003. Synergistic lethal effect of a combination of phage lytic enzymes with different activities on penicillin-sensitive and -resistant *Streptococcus pneumoniae* strains. Antimicrob. Agents Chemother. 47:375-377.

Loeffler JM, Nelson D, and Fischetti VA. 2001. Rapid killing of *Streptococcus pneumoniae* with a bacteriophage cell wall hydrolase. Science. 294:2170-2172.

Loessner MJ. 2005. Bacteriophage endolysins--current state of research and applications. *Curr. Opin. Microbiol.* 8:480-487.

Loessner MJ, Gaeng S, Wendlinger G, Maier SK, and Scherer S. 1998. The two-component lysis system of *Staphylococcus aureus* bacteriophage Twort: a large TTG-start holin and an associated amidase endolysin. *FEMS Microbiol. Lett.* 162:265-274.

Loessner MJ, Gaeng S, and Scherer S. 1999. Evidence for a holin-like protein gene fully embedded out of frame in the endolysin gene of *Staphylococcus aureus* bacteriophage 187. *J. Bacteriol.* 181:4452-4460.

Loessner MJ, Kramer K, Ebel F, and Scherer S. 2002. C-terminal domains of *Listeria monocytogenes* bacteriophage murein hydrolases determine specific recognition and high-affinity binding to bacterial cell wall carbohydrates. *Mol. Microbiol.* 44:335-49.

Loessner MJ, Maier SK, Daubek-Puza H, Wendlinger G, and Scherer S. 1997. Three *Bacillus cereus* bacteriophage endolysins are unrelated but reveal high homology to cell wall hydrolases from different bacilli. *J. Bacteriol.* 179:2845-2851.

Loessner MJ, Wendlinger G, and Scherer S. 1995. Heterogeneous endolysins in *Listeria monocytogenes* bacteriophages: a new class of enzymes and evidence for

conserved holin genes within the siphoviral lysis cassettes. *Mol. Microbiol.* 16:1231-1241.

López R, García E, García P, and García JL. 1997. The pneumococcal cell wall degrading enzymes: a modular design to create new lysins? *Microb. Drug Resist.* 3:199-211.

Low LY, Yang C, Perego M, Osterman A, and Liddington RC. 2005. Structure and lytic activity of a *Bacillus anthracis* prophage endolysin. *J. Biol.Chem.* 280:35433-35439.

Lu TK and Collins JJ. 2007. Dispersing biofilms with engineered enzymatic bacteriophage. *Proc. Natl. Acad. Sci. USA.* 104:11197-11202.

Lucas S, Copeland A, Lapidus A, et al. 2004. Sequencing of the draft genome and assembly of *Streptococcus suis* 89/1591. Unpublished. NCBI Genome Project ID 12417, www.ncbi.nlm.nih.gov/nuccore/AAFA00000000.

Lun Z, Wang Q, Chen X, Li A, and Zhu X. 2007. *Streptococcus suis*: an emerging zoonotic pathogen. *Lancet Infect. Dis.* 7:201-209.

Luque I, Blume V, Borge C, Vela AI, Perea JA, Márquez JM, Fernández-Garayzábal JF, and Tarradas C. 2009. Genetic analysis of *Streptococcus suis* isolates recovered from diseased and healthy carrier pigs at different stages of production on a pig farm. *Vet. J.* doi:10.1016/j.tvjl.2009.09.005.

Luria SE and Delbrück M. 1943. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*. 28:491-511.

Ma YL and Lu CP. 2008. Isolation and identification of a bacteriophage capable of infecting *Streptococcus suis* type 2 strains. *Vet. Microbiol.* 132:340-347.

Maass D, and Weidel W. 1963. Final proof for the identity of enzymic specificities of egg-white lysozyme and phage T2 enzyme. *Biochim. Biophys. ACTA*. 78:369-70.

Marchesi JR and Weightman AJ. 2003. Comparing the dehalogenase gene pool in cultivated in alpha-halocarboxylic acid-degrading bacteria with the environmental metagenome pool. *Appl. Environ. Microbiol.* 69:4375-4382.

Matches JR, Liston J, and Curran D. 1974. *Clostridium perfringens* in the environment. *Appl. Microbiol.* 28:655-660.

Matsushita I and Yanase H. 2008. A novel thermophilic lysozyme from bacteriophage phiIN93. *Biochem. Biophys. Res. Comm.* 377:89-92.

Maxted WR. 1957. The active agent in nascent phage lysis of streptococci. *J. Gen. Microbiol.* 16:584-595.

- Mayer MJ, Narbad A, and Gasson MJ.** 2008. Molecular characterization of a *Clostridium difficile* bacteriophage and its cloned biologically active endolysin. J. Bacteriol. 190:6734-40.
- Myers GS, Rasko DA, Cheung JK, et al.** 2006. Skewed genomic variability in strains of the toxigenic bacterial pathogen, *Clostridium perfringens*. Genome Res. 16:1031-1040.
- McClane BA.** 1997. *Clostridium perfringens*. In: Food microbiology: fundamentals and frontiers, Doyle MP, Beuchat LR, and Montville TJ, eds. ASM Press, 1st edition, 305-326.
- Merabishvili M, Pirnay JP, Verbeken G, et al.** 2009. Quality-controlled small-scale production of a well-defined bacteriophage cocktail for use in human clinical trials. PLoS One. 4:e4944.
- Merril CR, Scholl D, and Adhya S.** 2006. Phage therapy. In: The bacteriophages, Calendar R, ed. Oxford University Press, 2nd edition, 725-742.
- Mesnager S, Chau F, Dubost L, and Arthur M.** 2008. Role of N-acetylglucosaminidase and N-acetylmuramidase activities in *Enterococcus faecalis* peptidoglycan metabolism. J. Biol. Chem. 283:19845-19853.

McCullers JA, Karlström A, Iverson AR, Loeffler JM, and Fischetti VA. 2007.

Novel strategy to prevent otitis media caused by colonizing *Streptococcus pneumoniae* PLoS Pathog. 3:e28.

McKinstry M and Edgar R. 2005. Uses of phages in therapy and bacterial detection. In: Phages: Their role in bacterial pathogenesis and biotechnology, Waldor MK, Freidman DI, and Adhya SL, eds. American Society for Microbiology Press, 1st edition, 430-440.

Moak M, and Molineux IJ. 2004. Peptidoglycan hydrolytic activities associated with bacteriophage virions. Mol. Microbiol. 51:1169-1183.

Mokrousov I. 2009. *Corynebacterium diphtheriae*: genome diversity, population structure and genotyping perspectives. Infect. Genet. Evol. 9:1-15.

Morita M, Tanji Y, Orito Y, Mizoguchi K, Soejima A, and Unno H. 2001a. Functional analysis of antibacterial activity of *Bacillus amyloliquefaciens* phage endolysin against Gram-negative bacteria. FEBS Lett. 500:56-59.

Morita M, Tanji Y, Orito Y, Mizoguchi K, Soejima A, and Unno H. 2001b. Antibacterial activity of *Bacillus amyloliquefaciens* phage endolysin without holin conjugation. J. Biosci. Bioeng. 91:469-473.

Murphy JS. 1957. A phage-associated enzyme of *Bacillus megaterium* which destroys the bacterial cell wall. Virology. 4:563-581.

- Myers GSA, Rasko DA, Cheung JK, et al.** 2006. Skewed genomic variability in strains of the toxigenic bacterial pathogen, *Clostridium perfringens*. *Genome Res.* 16:1031-1040.
- Nakayama K, Kanaya S, Ohnishi M, Terawaki Y, and Hayashi T.** 1999. The complete nucleotide sequence of phiCTX, a cytotoxin-converting phage of *Pseudomonas aeruginosa*: implications for phage evolution and horizontal gene transfer via bacteriophages. *Mol. Microbiol.* 31:399-419.
- Nelson D and Fischetti V.** 2004. Bacteriophage lysin, PlyC, may aid in detection and treatment of bovine mastitis. Abstracts of the General Meeting of the American Society for Microbiology. 104:394.
- Nelson D, Loomis L, and Fischetti VA.** 2001. Prevention and elimination of upper respiratory colonization of mice by group A streptococci by using a bacteriophage lytic enzyme. *Proc. Natl. Acad. Sci. USA.* 98:4107-4112.
- Nelson D, Schuch R, Chahales P, Zhu S, and Fischetti VA.** 2006. PlyC: a multimeric bacteriophage lysin. *Proc. Natl. Acad. Sci. USA.* 103:10765-10770.
- Nelson D, Schuch R, Zhu S, Tscherne DM, and Fischetti VA.** 2003. Genomic sequence of C1, the first streptococcal phage. *J. Bacteriol.* 185:3325-3332.
- Newton J, and Deutscher SL.** 2008. Phage peptide display. *Handb. Exp. Pharmacol.* 185 Pt 2:145-163.

Novak R, Charpentier E, Braun JS, Park E, Murti S, Tuomanen E, and Masure R. 2000. Extracellular targeting of choline-binding proteins in *Streptococcus pneumoniae* by a zinc metalloprotease. *Mol. Microbiol.* 36:366-376.

Novick RP, Christie GE, and Penadés JR. 2010. The phage-related chromosomal islands of Gram-positive bacteria. *Nat. Rev. Microbiol.* 8:541-551.

Ochman H, Gerber AS, and Hartl DL. 1998. Genetic applications of an inverse polymerase chain reaction. *Genetics.* 120:621-623.

O'Flaherty S, Coffey A, Meaney W, Fitzgerald GF, and Ross RP. 2005. The recombinant phage lysin LysK has a broad spectrum of lytic activity against clinically relevant staphylococci, including methicillin-resistant *Staphylococcus aureus*. *J. Bacteriol.* 187:7161-7164.

O'Flaherty S, Ross RP, and Coffey A. 2009. Bacteriophage and their lysins for elimination of infectious bacteria. *FEMS Microbiol. Rev.* 33:801-819.

Okada Y, Streisinger G, Tsugita A, and Inouye M. 1968. Frame shift mutations near the beginning of the lysozyme gene of bacteriophage T4. *Science.* 162:807-808.

Oki M, Kakikawa M, Yamada K, Taketo A, and Kodaira K. 1996. Cloning, sequence analysis, and expression of the genes encoding lytic functions of Bacteriophage phi g1e. *Gene*. 176:215-223.

Oram JD, and Reiter B. 1965. Phage-associated lysins affecting group N and group D streptococci. *J. Gen. Microbiol.* 40:57-70.

Owen JE, Schultz DW, Taylor A, and Smith GR. 1983. Nucleotide sequence of the lysozyme gene of bacteriophage T4. Analysis of mutations involving repeated sequences. *J. Mol. Biol.* 165:229-248.

Park T, Struck DK, Dankenbring CA, and Young R. 2007. The pinholin of lambdoid phage 21: control of lysis by membrane depolarization. *J. Bacteriol.* 189:9135-9139.

Parreira R, São-José C, Isidro A, Domingues S, Viera G, and Santos MA. 1999. Gene organization in a central DNA fragment of *Oenococcus oeni* bacteriophage fOg44 encoding lytic, integrative and non-essential functions. *Gene*. 226:83-93.

Paschke M. 2006. Phage display systems and their applications. *Appl. Microbiol. Biotechnol.* 70:2-11.

Payne K, Sun Q, Sacchettini J, and Hatfull GF. 2009. Mycobacteriophage Lysin B is a novel mycolylarabinogalactan esterase. *Mol. Microbiol.* 73:367-381.

- Pedulla ML, Ford ME, Houtz JM, et al.** 2003. Origins of highly mosaic mycobacteriophage genomes. *Cell*. 133:172-182.
- Perry LJ, Heyneker HL, and Wetzel R.** 1985. Non-toxic expression in *Escherichia coli* of a plasmid-encoded gene for phage T4 lysozyme. *Gene*. 38:259-264.
- Pei J and Grishin NV.** 2005. COG3926 and COG5526: a tale of two new lysozyme-like protein families. *Protein Sci.*14:2574-2581.
- Piuri M and Hatfull G.** 2006. A peptidoglycan hydrolase motif within the mycobacteriophage TM4 tape measure protein promotes efficient infection of stationary phase cells. *Mol. Microbiol.* 62:1569-1585.
- Platteeuw C, and de Vos MW.** 1992. Location, characterization and expression of lytic enzyme-encoding gene, *lytA*, of *Lactococcus lactis* bacteriophage phi US3. *Gene* 118:115-120.
- Porter CJ, Schuch R, Pelzak AJ, et al.** 2007. The 1.6 Å crystal structure of the catalytic domain of PlyB, a bacteriophage lysin active against *Bacillus anthracis*. *J. Mol. Biol.* 366: 540-550.
- Pritchard DG, Dong S, Baker JR, and Engler JA.** 2004. The bifunctional peptidoglycan lysin of *Streptococcus agalactiae* bacteriophage B30. *Microbiology*. 150:2079-2087.

Pritchard DG, Dong S, Kirk MC, Cartee RT, and Baker JR. 2007. LambdaSa1 and LambdaSa2 prophage lysins of *Streptococcus agalactiae*. Appl. Environ. Microbiol. 73:7150-7154.

Ptashne, M. A genetic switch: phage lambda revisited. Cold Spring Harbor Laboratory Press, 3rd edition, 1994.

Ralston DJ, Baer BS, Lieberman M, and Krueger AP. 1955. Virolysin: a virus-induced lysin from staphylococcal phage lysates. Proc. Soc. Exp. Biol. Med. 89:502-507.

Ralston DJ, Lieberman M, Baer B, and Krueger AP. 1957. Staphylococcal virolysin, a phage-induced lysin; its differentiation from the autolysis of normal cells. J. Gen. Physiol. 40:791-807.

Ralston DJ, and McIvor M. 1964. Lysis-from without *Staphylococcus aureus* strains by combination of specific phages and phage-induced lytic enzymes. J. Bacteriol. 88:676-681.

Ramanculov E and Young R. 2001. An ancient player unmasked: T4 rI encodes a t-specific antiholin. Mol. Microbiol. 41:575-583.

Rao GR, and Burma DP. 1971. Purification and properties of phage P22-induced lysozyme. J. Biol.Chem. 246:6474-6479.

Rashel M, Uchiyama J, Takemura I, Hoshiba H, Ujihara T, Takatsuji H, Honke K, and Matsuzaki S. 2008. Tail-associated structural protein gp61 of *Staphylococcus aureus* phage phi MR11 has bifunctional lytic activity. FEMS Microbiol. Lett. 284:9-16.

Rashel M, Uchiyama J, Ujihara T, Uehara Y, Kuramoto S, Sugihara S, Yagyu K, Muraoka A, Sugai M, Hiramatsu K, Honke K, and Matsuzaki S. 2007. Efficient elimination of multidrug-resistant *Staphylococcus aureus* by cloned lysin derived from bacteriophage phi MR11. J. Infect. Dis. 196:1237-1247.

Ravin NV. 2003. Mechanisms of replication and telomere resolution of the linear plasmid prophage N15. FEMS Microbiol Lett. 221:1-6.

Raya RR and Hébert EM. 2009. Isolation of phage via induction of lysogens. In: Bacteriophages: methods and protocols, Volume 1: Isolation, characterization, and interactions, Clokie RJ and Kropinski, eds. Humana Press, 1st edition, 23-32.

Rees CE and Dodd CE. 2006. Phage for rapid detection and control of bacterial pathogens in food. Adv. Appl. Microbiol. 59:159-86.

Reiter B, and Oram JD. 1963. Group N streptococcal phage lysin. J. Gen. Microbiol. 32:29-32.

Rennell D and Pette AR. 1985. Phage P22 lysis genes: nucleotide sequences and functional relationships with T4 and lambda genes. Virology. 143:280-289.

- Rhoads DD, Wolcott RD, Kuskowski MA, Wolcott BM, Ward LS, and Sulakvelidze A.** 2009. Bacteriophage therapy of venous leg ulcers in humans: results of a phase I safety trial. *J. Wound Care.* 18:237-243.
- Romero A, López R, and García P.** 2004. Characterization of LytA-like N-acetylmuramoyl-L-alanine amidases from two new *Streptococcus mitis* bacteriophages provides insights into the properties of the major pneumococcal autolysin. *J. Bacteriol.* 186:8229-8239.
- Romero A, López R, and García P.** 1990. Characterization of the pneumococcal bacteriophage HB-3 amidase: cloning and expression in *Escherichia coli*. *J. Virol.* 64:137-142.
- Rondon MR, August PR, Betterman AD, et al.** 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* 66:2541-2547.
- Rohwer F, Seguritan V, Choi DH, Segall AM, and Azam F.** 2001. Production of shotgun libraries using random amplification. *Biotechniques.* 31:108-118.
- Rosenthal RS and Dziarski R.** 1994. Isolation of peptidoglycan and soluble peptidoglycan fragments. *Methods Enzymol.* 235:253-285.

Rusch DB, Halpern AL, Sutton G, et al. 2007. The *Sorcerer II* global ocean sampling expedition: Northwest Atlantic through Eastern tropical Pacific. PLOS Biol. 5:e77.

Sable S and Lortal S. 1995. The lysins of bacteriophages infecting lactic acid bacteria. Appl. Microbiol. Biotechnol. 43:1-6.

Saedi MS, Garvey KJ, and Ito J. 1987. Cloning and purification of a unique lysozyme produced by Bacillus phage phi 29. Proc. Natl. Acad. Sci. USA. 84:955-958.

Salazar O and Asenjo JA. 2007. Enzymatic lysis of microbial cells. Biotechnol. Lett. 29:985-994.

Sambrook J, Fritsch EF, and Maniatis T. 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, 2nd edition.

Santos F, Yarza P, Parro V, Briones C, and Anton J. 2010. The metavirome of a hypersaline environment. Environ Microbiol. In press.

Sass P and Bierbaum G. 2007. Lytic activity of recombinant bacteriophage phi11 and phi12 endolysins on whole cells and biofilms of *Staphylococcus aureus*. Appl. Environ. Microbiol. 73:347-352.

Sau S, Chattoraj P, Ganguly T, Chanda PK, and Mandal NC. 2008. Inactivation of indispensable bacterial proteins by early proteins of bacteriophages: implication in antibacterial drug discovery. *Curr. Protein. Pept. Sci.* 9:284-290.

Segura M and Gottschalk M. 2004. Extracellular virulence factors of streptococci associated with animal diseases. *Front. Biosci.* 9:1157-1188.

Sertic V. 1929. The lysine origin of a race of the Bacteriophage. *Comptes Rendus Hebdomadaires des Séances et Mémoires de la Société de Biologie et des ses Filiales.* 100:477-479.

Schmitz JE, Daniel A, Collin M, Schuch R, and Fischetti VA. 2008. Rapid DNA library construction for functional genomic and metagenomic screening. *Appl. Environ. Microbiol.* 74:1649-1652.

Schmitz JE, Schuch R, and Fischetti VA. 2010a. Identifying phage lytic enzymes: past, present, and future. In: *Enzybiotics: antibiotic enzymes as drugs and therapeutics*, Villa TG and Veiga-Crespo P, eds. John Wiley and Sons, 1st edition, 219-251.

Schmitz JE, Schuch R, and Fischetti VA. 2010b. Identifying active phage lysins through functional viral metagenomics. *Appl. Environ. Microbiol.* Submitted.

Schoenfeld T, Liles M, Wommack KE, Polson SW, Godiska R, and Mead D.

2010. Functional viral metagenomics and the next generation of molecular tools. Trends Microbiol. 18:20-9.

Scholl D, Kieleczawa J, Kemp P, Rush J, Richardson CC, Merrill C, Adhya S, and Molineux IJ. 2004. Genomic analysis of bacteriophages SP6 and K1-5, an estranged subgroup of the T7 supergroup. J. Mol. Biol. 335:1151-1171.

Schuch R, Nelson D, and Fischetti VA. 2002. A bacteriolytic agent that detects and kills *Bacillus anthracis*. Nature. 418:884-889.

Schuch R, Fazzini M, Russell R, Nelson D, and Fischetti V. 2004.

Characterization of the Bcp1 phage, encoding a potent lytic agent active against the causative agent of anthrax, *B. anthracis*. Abstracts of the General Meeting of the American Society for Microbiology. 104:395-396.

Schuch R and Fischetti VA. 2009. The secret life of the anthrax agent *Bacillus anthracis*: bacteriophage-mediated ecological adaptations. PLoS One. 4:e6532.

Schuch R, Fischetti VA, and Nelson D. 2009. A Genetic Screen to Identify Bacteriophage lysins. In: Bacteriophages: methods and protocols, Volume 2: Molecular and applied aspects, Clokie RJ and Kropinski, eds. Humana Press, 1st edition, 307-319.

Shearman C, Underwood H, Jury K, and Gasson M. 1989. Cloning and DNA sequence analysis of a *Lactococcus* bacteriophage lysin gene. *Mol. Gen. Genet.* 218:214-221.

Sheehan MM, García JL, and López R. 1997. The lytic enzyme of the pneumococcal phage Dp-1: a chimeric lysin of intergeneric origin. *Mol. Microbiol.* 25:717-725.

Sheehan MM, García JL, and López R. 1996. Analysis of the catalytic domain of the lysin of the lactococcal bacteriophage Tuc2009 by chimeric gene assembling. *FEMS Microbiol. Lett.* 140:23-28.

Shimizu T, Ohtani K, Hirakawa H, et al. 2002. Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. *Proc. Natl. Acad. Sci. USA.* 99:996-1001.

Shugar D. 1952. The measurement of lysozyme activity and the ultra-violet inactivation of lysozyme. *Biochim. Biophys ACTA.* 8:302-309.

Shuman S. 1994. Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase. *J. Biol. Chem.* 269:32678-32684.

Smedly III JG, Fisher DJ, Sayeed S, Chakrabarti G, and McClane BA. 2004. The enteric toxins of *Clostridium perfringens*. *Rev. Physiol. Biochem. Pharmacol.* 152:183-204.

Smith HW and Huggins MB. 1982. Successful treatment of experimental *Escherichia coli* infections in mice using phage: its general superiority over antibiotics. J. Gen. Microbiol. 128:307-318.

Smith TC, Capuano AW, Boese B, Myers KP, and Gray GC. 2008. Occupational exposure to *Streptococcus suis* among US swine workers. Emerging Infect. Dis. 14:1925-1927.

Smith HW and Huggins MB. 1983. Effectiveness of phages in treating experimental *Escherichia coli* diarrhoea in calves, piglets and lambs. J. Gen. Microbiol. 129:2659-2675.

Sonstein SA, M Hammel J, and Bondi A. 1971. Staphylococcal bacteriophage-associated lysin: a lytic agent active against *Staphylococcus aureus*. J. Bacteriol. 107:499-504.

Soothill JS. 1994. Bacteriophage prevents destruction of skin grafts by *Pseudomonas aeruginosa*. Burns. 20:209-211.

Staats JJ, Feder I, Okwumabua O, and Chengappa MM. 1997. *Streptococcus suis*: past and present. Vet. Res. Comm. 21:381-407.

Steffan RJ and Atlas RM. 1988. DNA amplification to enhance detection of genetically engineered bacteria in environmental samples. Appl. Environ. Microbiol. 54:2185-2191.

Stojković EA and Rothman-Denes LB. 2007. Coliphage N4 N-acetylmuramidase defines a new family of murein hydrolases. *J. Mol. Biol.* 366:406-419.

Strauch E, Kaspar H, Schaudinn C, Dersch P, Madela K, Gewinner C, Hertwig, Wecke SJ, and Appel B. 2001. Characterization of enterocolitacin, a phage tail-like bacteriocin, and its effect on pathogenic *Yersinia enterocolitica* strains. *Appl. Environ. Microbiol.* 67:5634-5642.

Strausberg RL, Levy S, and Rogers Y. 2008. Emerging DNA sequencing technologies for human genomic medicine. *Drug Discov. Today.* 13:569-577.

Sulakvelidze A, Alavidze Z, and Morris JG. 2001. Bacteriophage therapy. *Antimicrob. Agents Chemother.* 45:649-659.

Summer EJ, Berry J, Tran TA, Niu L, Struck DK, and Young R. 2007. Rz/Rz1 lysis gene equivalents in phages of Gram-negative hosts. *J. Mol. Biol.* 373:1098-1112.

Summer EJ, Gonzalez CF, Bomer M, et al. 2006. Divergence and mosaicism among virulent soil phages of the *Burkholderia cepacia* complex. *J. Bacteriol.* 188:255-268.

Summers WC. 1999. Felix d'Herelle and the origins of molecular biology. Yale University Press, 1st edition.

Summers WC. 2001. Bacteriophage therapy. *Annu. Rev. Microbiol.* 55:437-451.

Sutherland IW. 1967. Phage-induced fucosidases hydrolyzing the exopolysaccharide of *Klebsiella aerogenes* type 54. *Biochem. J.* 104:278–285.

Suttle CA. 2007. Marine viruses: major players in the global ecosystem. *Nat. Rev. Microbiol.* 5:801-812.

Swildens B, Nielen M, Wisselink HJ, Verheijden JH, and Stegeman JA. 2007. Elimination of strains of *Streptococcus suis* serotype 2 from the tonsils of carrier sows by combined medication and vaccination. *Vet. Rec.* 160:619-621.

Takác M, Witte A, and Bläsi U. 2005. Functional analysis of the lysis genes of *Staphylococcus aureus* phage P68 in *Escherichia coli*. *Microbiology.* 151:2331-2342.

Tan KS, Wee BY, and Song KP. 2001. Evidence for holin function of *tcdE* gene in the pathogenicity of *Clostridium difficile*. *J. Med. Microbiol.* 50:613-619.

Tang J, Wang C, Feng Y, et al. 2006. Streptococcal toxic shock syndrome caused by *Streptococcus suis* serotype 2. *PLoS Med.* 3:e151.

Taylor A, Das BC, and van Heijenoort J. 1975. Bacterial cell wall peptidoglycan fragments produced by phage λ or ViII endolysins containing 1,6-anhydro-N-acetylmuramic acid. *Eur. J. Biochem.* 63:47-54.

- Terzaghi E, Okada Y, Streisinger G, Emrich J, Inouye M, and Tsugita A.** 1966. Change of a sequence of amino acids in phage T4 lysozyme by acridine-induced mutations. *Proc. Natl. Acad. Sci. USA.* 56:500-507.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, and Higgins DG.** 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 24:4876-4882.
- Threlfall EJ and Frost JA.** 1990. The identification, typing and fingerprinting of *Salmonella*: laboratory aspects and epidemiological applications. *J. Appl. Bacteriol.* 68:5-16.
- Torsvik V and Ovreas L.** 2002. Microbial diversity and function in soil. *Curr. Opin. Microbiol.* 5:240-245.
- Tourville DR and Johnstone DB.** 1966. Lactic streptococcal phage-associated lysin. I. Lysis of heterologous lactic streptococci by a phage-induced lysine J. *Dairy Sci.* 49:158-162.
- Trautwetter A, Ritzenthaler P, Alatossava T, and Mata-Gilsinger M.** 1986. Physical and genetic characterization of the genome of *Lactobacillus lactis* bacteriophage LL-H. *J. Virol.* 59:551-555.
- Tringe SG and Rubin EM.** 2005. Metagenomics: DNA sequencing of environmental samples. *Nat. Rev. Genet.* 6:805-814.

- Twort FW.** 1915. An investigation on the nature of ultra-microscopic viruses. Lancet. 186:1241-1243.
- Twort FW.** 1925. The transmissible bacterial lysin and its action on dead bacteria. Lancet. 206:642-644.
- Tyson GW, Chapman J, Hugenholtz P, et al.** 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature. 428:37-43.
- Uchiyama T and Miyazaki K.** 2009. Functional metagenomics for enzyme discovery: challenges to efficient screening. Curr. Opin. Biotechnol. 20:616-622.
- United States Food and Drug Administration.** 2009. The *bad bug book*: foodborne pathogenic microorganisms and natural toxins handbook. (<http://www.fda.gov/Food/FoodSafety/FoodborneIllness/default.htm>).
- Ussery DW, Wassenaar TM, and Borini S.** 2009. Microbial genome sequences: a new era in microbiology. In: Computing for comparative microbial genomics: bioinformatics for microbiologists, Springer, 1st edition, 37-51.
- Uzal FA and Songer JG.** 2008. Diagnosis of *Clostridium perfringens* intestinal infections in sheep and goats. J. Vet. Diagn. Invest. 20:253-265.
- van der Ploeg JR.** 2007. Genome sequence of *Streptococcus mutans* bacteriophage M102. FEMS Microbiol. Lett. 275:130-138.

- van Immerseel F, Rood JI, Moore RJ, and Titball RW.** 2009. Rethinking our understanding of the pathogenesis of necrotic enteritis in chickens. *Trends Microbiol.* 17:32-36.
- van Duyne GD.** Lambda integrase: armed for recombination. *Curr Biol.* 15:R658-660 (2005).
- Vasala A, Välikillä M, Caldentey J, and Alatossava T.** 1995. Genetic and biochemical characterization of the *Lactobacillus delbrueckii* subsp. *lactis* bacteriophage LL-H lysin. *Appl. Environ. Microbiol.* 61:4004-4011.
- Veiga-Crespo P and Villa TG.** 2010. Phylogeny of enzybiotics. In: *Enzybiotics: antibiotic enzymes as drugs and therapeutics*, Villa TG and Veiga-Crespo P, eds. John Wiley and Sons, 1st edition, 75-105.
- Venter JC, Remington K, Heidelberg JF, et al.** 2004. Environmental genome shotgun sequencing of the Sargasso sea. *Science.* 304:66-74.
- Verheust C, Fornelos N, and Mahillon J.** 2004. The *Bacillus thuringiensis* phage GIL01 encodes two enzymes with peptidoglycan hydrolase activity. *FEMS Microbiol. Lett.* 237:289-295.
- Villa TG and Veiga-Crespo P, eds.** 2010. *Enzybiotics: antibiotic enzymes as drugs and therapeutics*. John Wiley and Sons, 1st edition.

Voget S, Steele H, and Streit WR. 2005. Metagenomes – an unlimited resource for novel genes, biocatalysts, and metabolites. *Minerva Biotec.* 17:47-53.

Vollmer W, Blanot D, and de Pedro MA. 2008a. Peptidoglycan structure and architecture. *FEMS Microbiol. Rev.* 32:149-167.

Vollmer W, Joris B, Charlier P, and Foster S. 2008b. Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiol. Rev.* 32:259-286.

Wang G, Clark CG, Liu C, Pucknell CK, Kruk TM, Cladiera R, Woodward DL, and Rodgers FG. 2003. Detection and characterization of the hemolysin genes in *Aeromonas hydrophila* and *Aeromonas sobria* by mutiplex PCR. *J. Clin. Microbiol.* 41:1048-1054.

Wang G, Tyler KD, Munro CK, and Johnson WM. 1996. Characterization of cytotoxic, hemolytic *Aeromonas caviae* clinical isolates and their identification by determining presence of a unique hemolysin gene. *J. Clin. Microbiol.* 34:3203-3205.

Wang IN, Smith DL, and Young R. 2000. Holins: the protein clocks of bacteriophage infections. *Annu. Rev. Microbiol.* 54:799-825.

Wang Y, Sun JH, and Lu CP. 2009. Purified recombinant phage lysin LySMP: an extensive spectrum of lytic activity for swine streptococci. *Curr. Microbiol.* 58:609-615.

Ward N. 2006. New directions and interactions in metagenomics research. *FEMS Microbiol. Ecol.* 55:331-338.

Ward LJ, Beresford TP, Lubbers MW, Jarvis BD, Jarvis AW. 1993. Sequence analysis of the lysin gene region of the prolate lactococcal bacteriophage c2. *Can. J. Microbiol.* 39:767-774.

Wax RG, Lewis K, Salyers AA, and Taber H, eds. 2008. Bacterial resistance to antimicrobials. CRC Press, 2nd edition.

Weaver LH and Matthews BW. 1987. Structure of bacteriophage T4 lysozyme refined at 1.7 Å resolution. *J. Mol. Biol.* 193:189-99

Welker NE. 1967. Purification and properties of a thermophilic bacteriophage lytic enzyme. *J. Virol.* 1:617-625.

Wiegand I, Hilpert K, and Hancock RE. 2008. Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat. Protoc.* 3:163-175.

Williams SR, Gebhart D, Martin DW, and Scholl D. 2008. Retargeting R-type pyocins to generate novel bactericidal protein complexes. *Appl. Environ. Microbiol.* 74:3868-3876.

Williamson SJ, Rusch DB, Yooseph S, et al. 2008. The Sorcerer II Global Ocean Sampling Expedition: Metagenomic Characterization of Viruses within Aquatic Microbial Samples. PLoS ONE 3:e1456.

Wolff A and Ionesco H. 1975. [Purification and characterization of *Clostridium perfringens* BP6K-N5 strain bacteriocin N5] Ann. Microbiol. (Paris) 126:343-356.

Wright A, Hawkins CH, Änggård EE, and Harper DR. 2009. A controlled clinical trial of a therapeutic bacteriophage preparation in chronic otitis due to antibiotic-resistant *Pseudomonas aeruginosa*; a preliminary report of efficacy. Clin. Otolaryngol. 34:349-357.

Xu Q, Sudek S, McMullen D, et al. 2009. Structural basis of murein peptide specificity of a γ -D-glutamyl-L-diamino acid endopeptidase. Structure. 17:303-313.

Ye T and Zhang X. Characterization of a lysin from deep-sea thermophilic bacteriophage GVE2. Appl. Microbiol. Biotechnol. 78:635-641.

Yokoi KJ, Sugahara K, Iguchi A, et al. 2008. Molecular properties of the putative autolysin Atl(WM) encoded by *Staphylococcus warneri* M: mutational and biochemical analyses of the amidase and glucosaminidase domains. Gene. 416:66-76.

Yokoi KJ, Shinohara M, Kawahigashi N, et al. 2005. Molecular properties of the two-component cell lysis system encoded by prophage phigaY of *Lactobacillus gasseri* JCM 1131T: cloning, sequencing, and expression in *Escherichia coli*. *Int. J. Food Microbiol.* 99:297-308.

Yokoi KJ, Kawahigashi N, Uchida M, Sugahara K, Shinohara M, Kawasaki K, Nakamura S, Taketo A, and Kodaira K. 2005. The two-component cell lysis genes holWMY and lysWMY of the *Staphylococcus warneri* M phage varphiWMY: cloning, sequencing, expression, and mutational analysis in *Escherichia coli*. *Gene.* 351:97-108.

Yoon SS, Kim JW, Breidt F, and Fleming HP. 2001. Characterization of a lytic *Lactobacillus plantarum* bacteriophage and molecular cloning of a lysin gene in *Escherichia coli*. *Int. J. Food Microbiol.* 65:63-74.

Yoong P, Schuch R, Nelson D, and Fischetti VA. 2004. Identification of a broadly active phage lytic enzyme with lethal activity against antibiotic-resistant *Enterococcus faecalis* and *Enterococcus faecium*. *J. Bacteriol.* 186:4808-4812.

Yoong P, Schuch R, Nelson D, and Fischetti VA. 2006. PlyPH, a bacteriolytic enzyme with a broad pH range of activity and lytic action against *Bacillus anthracis*. *J. Bacteriol.* 188:2711-2714.

- Yooseph S, Sutton G, Rusch DB, et al.** 2007. The *Sorcerer II* global ocean sampling expedition: expanding the universe of protein families. *PLOS Biol.* 5(3):e16.
- Young I, Wang I, and Roof WD.** 2000. Phages will out: strategies of host cell lysis. *Trends Microbiol.* 8:120-128.
- Young R.** 2005. Phage lysis. In: *Phages: Their role in bacterial pathogenesis and biotechnology*, Waldor MK, Freidman DI, and Adhya SL, eds. American Society for Microbiology Press, 1st edition, 92-127.
- Young R and Wang IN.** 2006. Phage lysis. In: *The bacteriophages*, Calendar R, ed. Oxford University Press, 2nd edition, 104-128.
- Young R, Way J, Way S, Yin J, and Syvanen M.** 1979. Transposition mutagenesis of bacteriophage lambda: a new gene affecting cell lysis. *J. Mol. Biol.* 132:307-322.
- Zeng L, Skinner SO, Zhong C, Jippy J, Feiss M, and Golding I.** 2010. Decision making at a subcellular level determines the outcome of bacteriophage infection. *Cell* 141:682-291.
- Zhang T, Breitbart M, Lee WH, Run J, Wei CL, Soh SW, Hibberd ML, Liu ET, Rohwer F, and Ruan Y.** 2006. RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLOS Biol.* 4(1):e3.

Zhang N and Young R. 1999. Complementation and characterization of the nested Rz and Rz1 reading frames in the genome of bacteriophage lambda. *Mol Gen Genet.* 262:659-667.

Zimmer M, Loessner M, Morgan AJ (2008) US Patent 7 371 375.

Zimmer M, Scherer S, and Loessner MJ. 2002a. Genomic analysis of *Clostridium perfringens* bacteriophage phi3626, which integrates into *guaA* and possibly affects sporulation. *J. Bacteriol.* 184:4359-4368.

Zimmer M, Vukov N, Scherer S, and Loessner MJ. 2002b. The murein hydrolase of the bacteriophage phi3626 dual lysis system is active against all tested *Clostridium perfringens* strains. *Appl. Environ. Microbiol.* 68:5311-5317.

Zink RM, Loessner J, and Schere S. 1995. Characterization of cryptic prophages (monocins) in *Listeria* and sequence analysis of a holin/endolysin gene. *Microbiology.* 141:2577-2584.

APPENDIX OF DNA AND AMINO ACID SEQUENCES

A number of genes and their corresponding protein translations are referred to throughout the text of this thesis (primarily involving cloned phage lysins). This appendix reports their nucleotide and amino acid sequences. In a majority of cases, the sequences have already been submitted to the National Center for Biotechnology Information (NCBI) for inclusion in GenBank (www.ncbi.nlm.nih.gov/genbank/), in which case the corresponding accession numbers are provided.

Several basic parameters are also reported for each protein, including gene and protein length, protein molecular mass, and theoretical isoelectric point. For phage lysins, the predicted enzymatic and binding domains are also designated here. The position of these domains within the amino acid sequences are denoted with the following color scheme: **enzymatic domains are highlighted in blue** and **binding domains are highlighted in red**. The positions of all domains were predicted with the Pfam v24.0 algorithm (pfam.sanger.ac.uk). For several of the metagenomic lysins from Chapter 4, the genes were cloned with a truncated C-terminus (as described in the text); these cases are denoted as such in the appendix. No predicted isoelectric point is provided for the truncated lysins.

Chapter 2: PlyCM

- GenBank Accession Number: YP_695420 (from *C. perfringens* ATCC 13124)
- Originally annotated in GenBank as “Glycosyl hydrolase family protein”
- 1029 base pairs; 342 amino acids; 38.73 kDa; Theoretical pI = 6.04
- Glycosyl Hydrolase Type 25 (i.e. muramidase) enzymatic domain (PF01183)
- Dual SH3-Type 3 binding domain (PF08239)

```

ATGGAAAGTA GAAACAATAA TAATTTAAAA GGAATTGATG TATCAAACCTG
GAAAGGAAAT ATAAATTTTC AAAGTGTAAG AAATGATGGT GTAGAAGTAG
TTTATATTAA AGCTACAGAA GGTAATTACT TTAAGGATAA ATATGCTAAA
CAAAATTATG AGAGAGCGAA AGAACAAGGA TTAAGAGTAG GATTTTATCA
TTTCTTTAGA GCTAATAAAG GAGCCAAGGA TCAAGCAAAT TTTTTCGTAA
ATTATTTGAA TGAAATAGGA GCAGTTAATT ATGATTGTAA ATTAGCTTTA
GATATAGAAA CTAAGTAAGG GGTAGGAGCA AGAGATTTAA CTTCTATGTG
TATAGAATTC TTAGAAGAGG TAATAAAGAAT TACTGGAAAA GAAGTTGTTG
TATATACATA TACAAGCTTT GCAAATAATA ATTTAGATAG TAGATTATCC
AGTTATCCAG TTTGGATAGC TCATTATGGT GTAAACACTC CTGGAGCTAA
CAATATATGG AGTGAATGGG TTGGGTTCCA ATATTCAGAG AATGGAAGTG
TAGCTGGTGT AAGTGGTGGA TGTGATATGA ATGAGTTCAC TAATGGAATA
TTCATTGATT CAAATAATTT TACTTTAGAC AATGCTACTA CTAATAATGT
AAGTATTAAA TTAAATATAA GAGCTAAAGG AACTACTAAT TCTAAAGTAA
TTGGTTCAAT ACCAGCCAAT GAGAAGTTTA AAATAAAATG GGTTGATGAA
GATTATCTTG GTTGGTATTA CGTTGAGTAT AATGGAATAG TTGGCTATGT
AAATGCAGAT TATGTAGAAA AGCTACAAAT GGCTACTACT CATAATGTAA
GTACTTTTTT AAATGTAAGA GAAGAAGGAT CATTAAATTC TAGAATAGTA
GATAAGATAA ATACAGGTGA TATTTTTTGA ATAGATTGGG TGGATTCCGA
TTTTATAGGT TGGTATAGAG TAACAACATA AAATGGAAAA GTTGGATTG
TTAATGCTGA ATTTGTTAAG AAATTATAA

```

```

MQSRNNNNLK GIDVSNWKGN INFQSVKNDG VEVVYIKATE GNYFKDKYAK
QNYERAKEQG LRVGFYHFFR ANKGAKDQAN FFVNYLNEIG AVNYDCKLAL
DIETTEGVGA RDLTSMCIEF LEEVIRITGK EVVVYTYTSF ANNNLDSRLS
SYPVWIAHYG VNTPGANNIW SEWVGFQYSE NGSVAGVSGG CDMNEFTNGI
FIDSNNFTLD NATTKNVS IK LNIRAKGTTN SKVIGSIPAN EKFKIKWVDE
DYLGWYYVEY NGIVGYVNAD YVEKLQMAT HNVSTFLNVR EEGSLNSRIV
DKINTGDIFR IDWVDSDFIG WYRVTTKNGK VGFVNAEFVK KL

```


****NOTE:** In the preceding sequences for PlyCM, the underlined residues represent positions at which the experimentally-observed sequence differed from that in GenBank. 4G / 2E was intentionally changed from the wild-type 4C / 2Q to introduce an NcoI restriction site. 374T / 125I differed from the reported 374A / 125K, but the change was not intentional.

Chapter 3: BG-1 Phage Lysin (i.e. PlyBeta)

- GenBank Accession Number: EU258891
- 867 base pairs; 288 amino acids; 32.27 kDa; Theoretical pI = 5.96
- Glycosyl Hydrolase Type 25 (i.e. muramidase) enzymatic domain (PF01183)
- SH3-Type 5 binding domain (PF08460)

```
ATGGGTTATA TTGTTGATAT TTCTAAATGG AACGGTGACA TTAAGTGGGA
CGTTGCAGCA GGTCAATTGG ATTTAGCAAT TGCTCGTGTT CAAGACGGTT
CGAACTATGT TGACCCATATG TATAAGTCTT ACGTAGCATC TATGAAAGCT
CGTAACGTGC CATTTGGTAA CTACGCTTTC TGTCGTTTCG TATCTGTAGA
AGATGCAAAA GTAGAAGCTA GAGACTTTTG GGCTCGTGGA GACAAAGACG
CTTTATTTTG GGTAGCGGAC GTAGAAGTAA AAACATATGGG CGACATGCAA
GCGGGAACAC AAGCATTCAT TGACGAGCTA TATCGTTTAG GAGCTAAAAA
AGTAGGTCTA TATGTCGGGC ATCACACTTA TGTGGCTTTC GGTGCTAAAA
ACATTAGATG TGACTTCACT TGGATTCTTC GTTATGGTGG ACTAAAGCCT
GATTTCCCTT GTGATCTATG GCAGTACACT GAAACAGGTA ATGTTCTCTG
AATCGGTAAA TGTGACATTA ACAGCTTAAA CAGCGACAAG AACTAGAAAT
GGTTCACAGG TAAAGATTGC AACAACGGTA ACGTGACTCC GCCACCACAG
GGAGCTTACG ATTCTAGTTG GTTCACTAAG CAAACTGGTG TCTTCACTTT
AGATCGTACG ATTAACCTAC GTACTGCACC ATTCCCAAAT GCACCGTTAA
TTGCCCAATT AAACGCAGGA GATAACGTTA ACTACGAAGC GTACGGATAT
GAAAAAGACG GTTACGTTTG GTTACGTCAG CATCGTGGTA ATGGTAACTT
CGGTACATC GCATCAGGCG AAACATAAAA CGGTCAACGT ATCTCTACTT
GGGGAACCTT TAAATAA
```

```
MGYIVDISKW NGDINWDVAA GQLDLAIARV QDGSNYVDPM YKSYVASMKA
RNVFPGNYAF CRFVSVEDAK VEARDFWARG DKDALFWVAD VEVKTMGDMQ
AGTQAFIDEL YRLGAKKVGL YVGHHTYVAF GAKNIRCDFT WIPRYGGLKP
DFPCDLWQYT ETGNVPGIGK CDINSLNSDK TLEWFTGKDC NNGNVTPPPQ
GAYDSSWFTK QTGVFTLDRT INLRTAPFPN APLIAQLNAG DNVNIEAYGY
EKDGYVWLRQ HRGNNGFGYI ASGETKNGQR ISTWGTFK
```

Chapter 3: BG-2 Phage Lysin

- GenBank Accession Number: EU258892
- 936 base pairs; 311 amino acids; 34.31 kDa; Theoretical pI = 9.19
- N-acetylmuramoyl-L-alanine-amidase-Type 2 enzymatic domain (PF01510)
- Dual SH3-Type 3 binding domain (PF08239)

```

ATGGCTATTT CAGTAAGACA AAAAATGGTG GATTCAAGTA AGTATTCTTT
AAAATGTCCT TATGCTATGA CAGCAGAGTA CATTACAATC CACAACACGT
ACAATGATGC AAGTGCGAAC AATGAGGTTT AATACATGAT TACTAATAGT
AATGCAACTT CATTTCACTA TGCAATTGAT GATTTCGAAG TTGTACAAGG
TATTCCAACA AACCGTAATG CATGGCATTG TGGAGATGGA AACGGTAATG
GTAACCGTAA ATCTATCGGT GTTGAAATCT GTTACTCTAT GAGTGGTGGC
GATAGATATC GTAAAGCACA AGCTTTAGTT ATCAAATTCG TTGCACAAC
TTTAAGAGAA CGTGGATGGG GAATTGATAG AGTTAAGAAA CACCAAGATT
GGAGTGGAAA ATATTGTCCA CACCGTATCC TAGACGAAGG ACGTTGGCAA
TCAGTTCTAA ATGCTATTAA AGCAGAATTA AATGGTGGAG GTTCTACAGG
TGGAGGAACA ACTCAACCAC CAGTTGATAA CTCTACAGGT GTAGTTAGAG
TAACAGCAGA TGTTCTAAAC TTGCGTAATC AACCATCTAC TAACGGTTCT
ATCGTTGGTA AAATCTACAA AGGTCAAGAT TACAAGTTCT GGGCTATCTC
TAACGGATGG TACAACCTAG GTGGTAACCA ATGGGCTTCT GGTACTTATT
TACAAGTAAT TAGTGGAGGA ACGCCACAAC CACCAAAAGC AGTAACAGGT
ATTGCATATA TCACTGGATA CAATGTTAAT ATGCGTACAG GTGCAGGAAC
AGGTTACTCA GTAATCCGTC AATTAAATGC ACCAGAATCT TACAAAGTTT
GGGGAATGAA GGACGGTTGG TTAAACCTTG GTGGCGACCA ATGGATTAAG
AACGACTCTT CATTTGTTAG ATTTGTACAA GACTAA

```

```

MAISVRQKMV DSSKYSLKCP YAMTAEYITI HNTYNDASAN NEVQYMITNS
NATSFHYAID DFEVVQGIPT NRNAWHCGDG NGNGNRK SIG VEICYSMSGG
DRYRKAQALV IKFVAQLLRE RGWGIDRVKK HQDWSGKYCP HRI LDEGRWQ
SVLNAIKAEL NGGGSTGGGT TQPPVDNSTG VVRVTDV L NLRNPSTNGS
IVGKIYKGQD YKFWAISNGW YNLGGNQWAS GTYLQVISGG TPQPPKAVTG
IAYITGYNVN MRTGAGTGYS VIRQLNAPES YKVGWGMKDGW LNLG GDQWIK
NDSSFVRVQ D

```

Chapter 3: BG-3 Phage Lysin

•GenBank Accession Number: EU258893

•1056 base pairs; 351 amino acids; 39.35 kDa; Theoretical pI = 9.37

•N-acetylmuramoyl-L-alanine-amidase-Type 2 enzymatic domain (PF01510)

•SH3-Type 3 binding domain (PF08239)

```

ATGAAAAAAA CGTTAAACAA CATTCTTCT GTAGTCTTTG CAACTATTTT
AGCGTTATCC ATTACAACAA GTGCTGTTGC TGACAGAGTA TTGATCATTC
CTGATTTGCC GAAGCAAGGA TATAGAAATG GGGTAGGAGC TTATGAAGGT
GTAGTGGCGC ATAGTACAGC GACTCCAGAA GCTCCAGCAA TCAATATCCA
ACGTTATGAA ACGCGTACTT GGAGAAGTGC ATTTGTACAT TATGCAGTTG
ACTGGGATGA AGTTATTCAA ATTCAGATA CACGTTATAT TGCATATGGC
GCTGGACCAG CTGCTAACGC ACGTTTTGTA CACGTTGAAC TTTGTGAGAC
TTCAGATTAT AGCAAGTTTA AGCGAAGCTA TGACAAGTAT GTAAAATTAC
TAGCAAAAAT TTTACGTGAT CGTGGACTTT CAGTAGAAAA AGGATTATGG
ACACATGATG ACGTAAGGAA ATATCTTGGT GGAACAACTC ACACGGATCC
ACTGGATTAT CTAACAAAAC ATGGTATATC TGAAGCTCAA TTCCGAGCAG
ATGTGAAACG TGCTTATAAT AACACGGGTA TTTCTATTCC TGAACAACCT
TCTAAACCAG CGGAAAAACC AACAGCCAAT GTAGAAGGTG TAGCTTATAT
TGAAGGATAT AATGTAAATC TACGCAAAGG CCCAGATGCA AGCTATTCTG
TTATTCGTCA ATTAAATAAA CCAGAGGCTT ATAAGGTTTG GGGAGAAAAG
GGCGGATGGC TAAATTTAGG TTGGAATCAG TGGGTAAAT ATAATCAATC
GTATATCCGA TTTGAAAAGA AAGAAGCTGT AAGTCCAGTT GCTGGAAAAC
GTGTAGTGTC GAAAGTGAAT AATCTCAGAT TTTATAGCGC TCCATCGTGG
GAAGATAAGT ATGTTGCAGG TACTGTAGAT GTAGGATTAG GTTTTACGAT
TGACGCAACG GTAATGGTAA ATGGTTCCCC ACAATACAAA GTACACAATA
GTAAAGGAAC CACATACTAT ATCCCAGCAA GTGAAGCCTA TGTGTATGTG
AAGTAG

```

```

MKKTLKHISS VVFATILALS ITTSAVADRV LIIPDLPKQG YRNGVGAYEG
VVAHSTATPE APAINIQRYE TRTWRSFVH YAVDWDEVIQ IADTRYIAYG
AGPAANARFV HVELCETSDY SKFKRSYDKY VKLLAKILRD RGLSVEKGLW
THDDVRKYL GTHHTDPLDY LKKHGISEAQ FRADV KRAYN NTGISIPEQP
SKPAEKPTAN VEGVAYIEGY NVNLRKGPDA SYSVIRQLNK PEAYKVGWGEK
GGWLN LGWNQ WVKYNQSYIR FEKKEAVSPV AGKRVVSKVN NLRFY SAPSW
EDKYVAGTVD VGLGFTIDAT VMVNGSPQYK VHNSKGTYY IPASEAYVYV K

```

Chapter 3: BG-4 Phage Lysin

- GenBank Accession Number: Not yet submitted
- 1068 base pairs; 355 amino acids; 39.94 kDa; Theoretical pI = 9.11
- N-acetylmuramoyl-L-alanine-amidase-Type 2 enzymatic domain (PF01510)
- SH3-Type 3 binding domain (PF08239)

```

ATGAAAAAGA CAGTGAAACA TATTACCTCG TTCCTTATGA TTCTAGTACT
TGCCGTTTCG TTTGCTACAA GTGCTTTCGC GGATAGAACA CTTATTATTC
CTGATTTACC GAAACAACCA TACCGTAACG GTGTAGGTGC TTATGAGGGT
GTTGTAGCAC ATTCTACAGC AACACCAGAA GCGCCAGCTA TTAATATTCA
AAAATATGAG TCTCGTACAT GCGGTTTCAGC ATTCGTACAT TATGCAGTTG
ATTGGAATGA AACAATCCAA ATTGCTGATA CGAAATACAT TGCTTATGGT
GCTGGACCAG GAGCAAATAA ACGATTTGTT CATGTGGAAT TATGCGAAAC
AAAGGATTAT GAGAAATTCA AACGCAGCTA TGATAAATAC GTTAAGCTAT
TAGCTAAAAT TCTTCGTGAC CGTGGATTAT CTGTAGAAAA AGGATTATGG
ACTCACTATG ATGTTACAAA GTATCTTTTC GGTACAGATC ATGAAGATCC
ACTTGATTAC TTACGTAGTC ATGGAGTTTC AGAAGCGCAA TTTAGAACAG
ATGTACAACG AGCATACAAT AATTCTAATG TTGATGTTTC TGTACCGGAG
AAGCCATCTA AACCAGCGGA AGTTCCAATG GCTGTAACAG ACGGAATCGC
GTATATTGAA GGTTACAATG TTAACCTACG TAAAGGACCT GGTTCAAGTT
ATTCTAAGAT TCGTCAGTTA AACAAACCAG AAGCTTATGT TGTATGGGCT
GAAAAGGATG GTTGGTTAAA TCTTGGTGGC GAACAATGGA TTAAGAACGA
TCCATCTTAT GTAAAGTTTA GTAAGAAAAG CACTGTGGAT TCCTCTATTG
TAGGTAAACG CGTTGTTTCT AAAGTTAATA ATCTACGATT CTATGATGTT
CCATCTTGGC AGGATAAAGA TGTTGCTGGT TCTGTAGATG CAGGATTAGG
ATTTACAATT GATGCAAAAA TAAACGTCAA TGGATCACCA CAATACAAAG
TGCACAATAG TAAGGGTATA ACATATTATG TTACTGCAAA TGAAGCCTAT
GTGTATGTAG TAAAGTAA

```

```

MKKTVKHITS FLMILVLAVS FATSADFADRT LIIPDLPKQP YRNGVGAYEG
VVAHSTATPE APAINIQKYE SRTWRSFVH YAVDWNETIQ IADTKYIAYG
AGPGANKRFV HVELCETKDY EKFKRSYDKY VKLLAKILRD RGLSVEKGLW
THYDVTKYLF GTDHEDPLDY LRSHGVSEAE FRTDVQRAYN NSNVDVSVPE
KPSKPAEVPV AVTDGIAYIE GYNVNLRKGP GSSYSKIRQL NKPEAYVVAW
EKDGWLNLG EQWIKNDPSY VKFSKKSTVD SSIVGKRVVS KVNLRFYDV
PSWQDKDVAG SVDAGLGFTI DAKINVNGSP QYKVHNSKGI TYYVTANEAY
VYVVK

```

Chapter 3: A14 Phage Lysin

- GenBank Accession Number: Not yet submitted
- 954 base pairs; 317 amino acids; 34.84 kDa; Theoretical pI = 8.74
- N-acetylmuramoyl-L-alanine-amidase-Type 2 enzymatic domain (PF01510)
- Dual SH3-Type 3 binding domain (PF08239)

```
ATGACTATTT CAGTAAGACA AAAATTGGTG GATTCTAGTA AGTATGGTTT
AAAGTGTCCA AACGCTATGA CAGCAGAATA CATTACTATC CACAATACTT
ACAATGATGC AAGTGCAAAC AATGAGGTTT AGTACATGAT TACAAATGGT
AATGCTACTT CATTCCACTT TGCAATTGAT GATTTCGAGG TTGTCCAAGG
TATCCCTACA AACCGTAATG CTTGGCATTG TGGAGATGGA ACAGGTAGCG
GTAACATGAA GTCTATCGGA ATCGAAATCT GCTACTCTCT TTCTGGTGGA
GACAGATATC GTAAAGCAGA AGCTTTAGCT GTAAAATTTA CTGCTCAACT
TTTAAGAGAG CGTGGATGGG GAATTAGCCG AGTTAAGAAG CATCAAGATT
GGTCTGGTAA ATATTGTCCA CACCGTATCT TAGATGAAGG ACGTTGGCAG
TCATTCTTCA ATGCTGTTCA AACAGAACTT AATGGTGGAG GTTCTACACA
ACCACCAGTA AACAATACTA CAGGTGTTGT TCAAGTAATG GTTGCAGATT
TAAATCTACG TACACAACCA AGTGCAAGCG CTCCAATCAT CCGTAAATTA
GGTATCGGTG AAACCTACCA GTTCTGGGCA ATTTCTAATG GATGGTACAA
CTTAGGTGGA GACCAATGGG CTTATGGAGA CAATGGCAAC TACTTAAAAG
TTATTAGCGG TGGAAAGTCT GTAGCACCTG CTCCACAACC AAAACCAGAG
CCAAAACCAA TTACAGGTGT TGCTTATATC ACTGGATATA ATGTTAACAT
GCGTAAAGGT GCAGGAACAG GTTACGCTGT AATCCGTCAA TTGAATGCAC
CAGAATCATA TCAAGTGTGG GCAGTTAAAG ACGGTGGGTT GAATCTTGGT
GGAGACCAAT GGATTAAGAA TGACGCTTCA TTTGTCAGAT TCGTACAAGA
CTAA
```

```
MTISVRQKLV DSSKYGLKCP NAMTAEYITI HNTYNDASAN NEVQYMITNG
NATSFHFAID DFEVVQGIPT NRNAWHCGDG TGSGNMKSIG IEICYSLSGG
DRYRKAEEAL VKFTAQLLRE RGWGISRVKK HQDWSGKYCP HRI LDEGRWQ
SFLNAVQTEL NGGGSTQPPV NNTTGTVQVM VADLNLRTQP SASAPIIRKL
GIGETYQFWA ISNGWYNLGG DQWAYGDNGN YLKVISGGSS VAPAPQPKPE
PKPITGVAYI TGYNVNMKRG AGTGYAVIRQ LNAPESYQVW AVKDGWLNLG
GDQWIKNDAS FVRFVQD
```

Chapter 3: TSH Phage Lysin

- GenBank Accession Number: Not yet submitted
- 819 base pairs; 272 amino acids; 29.22 kDa; Theoretical pI = 6.65
- N-acetylmuramoyl-L-alanine-amidase-Type 3 enzymatic domain (PF01520)
- SH3-Type 5 binding domain (PF08460)

```
ATGGGAACAT ATAACGTACA CGGTGGTCAC AACTCGATCG TACAAGGTGC
TAACTGGGGT AACCGAAAAG AACACGTTAT GGATCGCCAG GTTAAAGACG
CTTTAATTAG CAAACTTCGT AGCCTTGGTC ACACAGTTTA TGACTGCACA
GACGAAACAG GTTCTACGCA AAGCGCTAAC TTACGTAACA TCGTAGCGAA
ATGTAATGCT CACCGAGTAG ACTTAGACAT TTCATTACAC TTAAATGCTT
ACAATGGTTC TGCTAGTGGC GTAGAGGTTT GTTACTACGA CCAACAAGCA
TTAGCTGCTA AAGTTTCTAA ACAGCTTTCA GACGACATCG GTTGGTCTAA
CCGTGGAGCT AAACCTCGTA CAGACCTTTA CGTATTAAAT AGCACGTCTG
CACCTGCTAT CCTAATCGAG CTTGGTTTCA TCGACAACGA GAGCGATATG
GCTAAATGGA ATGTAGACAA AATTGCAGAC TCTATCTGCT ATGCTATCAC
AGGGCAACGT ACAGGCTCTA CTGGTGGAAG CACAGGAGGT TCTACAGGCG
GAAGCACTGG TGGAGGTGGA TACGACTCTA GTTGGTTCAC ACCACAAAAC
GGTGTATTCA CAGCTAACAC TACAATCAAA GTAAGAAGCG AGCCAAGCGT
AAATGCAACT CACCTTCGTA CTCTGTACAG TGGTGGAACG TTCACGTATA
CTTCATTTCG AATGGAGAAA GACGGTTACG TTTGGATCAA AGGCGTAGAC
GGTACATACG TTGCAACAGG TGAAACTAGT AACGGAAAAC GTACCTCTTA
CTGGGGAACT TTCCAGTAA
```

```
MGTYNVHGGH NSIVQGANWG NRKEHVMDRQ VKDALISKLR SLGHTVYDCT
DETGSTQSAN LRNIVAKCNA HRVDLDISLH LNAYNGSASG VEVCYYDQQA
LAAKVSKQLS DDIGWSNRGA KPRTDLYVLN STSAPAILIE LGFDNESDM
AKWNVDKIAD SICYAITGQR TGSTGGSTGG STGGSTGGGG YDSSWFTPQN
GVFTANTTIK VRSEPSVNAT HLRTLSSGGT FTYTSFGMEK DGYVWIKGVD
GTYYVATGETS NGKRTSYWGT FQ
```

Chapter 3: Aer_M (Hemolysin, not a lysin)

•GenBank Accession Number: EU258894

```

ATGGATAAAG TAAAAATAAC GGGTTTGGCG TTGAGCATTT CGACGCTGTT
GATGAGCCAG GCTCATGGGG CAGAGCCCCT CTATCCTGAT CAGCTGAGAC
ACTTCAGCCT CGGTAGCCAG CGCTGTGGCG ATGATTATCG TGCCCTGACG
CGCAATGAGG CCATGAGTAT CCGCAGCGAT ATAGTCAGCA AGATGGGGCA
GTGGCAGATC ACCGGTCTGG CTGACAACCT GGTTCATCAT GGGTCAGGTT
ATAATGGCGA AATAAAGCAA GGCAGCGCGA GCGATACCTG GTGTTATCCC
ACGAAACCGG TAGCGGGGGA AATACCAGTT CTATCCGCAT GGAATATTCC
AGCTGGTGAC GAGATCGATG TGCAATGGCG AATGGTCCAT GATAATGATT
ATTTTATCAG ACCGATAAGT TATCTTGCGC ATAATTTGGG TTATGCCTGG
GTCAGTGGAA ACCATAGCCA ATATGTGGGT GAGGATATGG ACGTCACTCG
CGTCAGTGAT GGCTGGCTCA TTCAGGGAAT TAATGGGGGT GGCTGCAGTG
GTTATCGCTG TAGCGAGAAG AGCTCCATCA AGGTGAGCAA CTTCTCCTAC
ACCCTGGATC CCGGTTTCGT CAGCCATGGC CTGGTGACTG AAAGCGGCAA
GCAGCTGGTC AAGACCATCA CAGCTACGGC GACCAACTAT ACCGATTTGC
CCCAGCAGGT GGTGGTGACC CTCAAATATG ACAAGGCCAC CAACTGGTCC
AAGACCGACA CTTATGGTCT GAGTGAGAAG GTGGGCATCA AGAAGACCTT
CCAGATCCCA CAAGTATCCA GTACCGAATA CTCGGTGGAG ATATCGTCCA
GCCAGAGCTG GGCTCGCCAA GAAGGGGGGT CGGCAACCGA GACGGTATCG
ATAGAAGCCC GCCCGACGGT GCCGCCTCAC TCCAGCGTGC CGGTCCGGGT
TGCGCTCTAC AAGGCCAATG TATCTTATCC GTATGAGTTC AAAGCAGATA
TCAACTATGA CCTCACCTTG AATGGTTTCC TGCGCTGGGG CGGCAACGCC
TGGCACACCC ATCCGGACAA CCGGCCGACC TGGAATCACA CCTTCGTCAT
CGGTCCGTTT AAGGACAAGG CGAGCAGCAT CCGTTACCAG TGGGACAAGC
GCTACATCCC GGGTGAGGTG AAGTGGTGGG ACTGGAACCT GACCATACAA
CAGAACGGTG CCGACACCAT GAAGAATGCC CTGGCCAGGG TGCTGCGCCC
GGTGCGCGCC AGCATCACGG GGGACTTCCA CGCCGAGAGC CAGTTCGCCG
GCAATATCGA GATAGGTGCG GCCGTGCCCA TCGGGGGCGA CAGCAAGGTG
CGTCGTGCCC GCAGCGTCGA CAGCCCGGCA ACTGGCCTGC GCCTGGAGAT
CCCGCTCGAT GCCAGCGAGC TCTCGGCGCT GGGCTTTGAC AACGTCCAGC
TCACGCTGGA ACCCGCTACC GACAAATAA

```

```

MDKVKITGLA LSISTLLMSQ AHGAEPVYPD QLRHFSLSQ RCGDDYRALT
RNEAMSIRSD IVSKMGQWQI TGLADNWVIM GSGYNGEIKQ GSASDTWCYP
TKPVAGEIPV LSAWNIPAGD EIDVQWRMVH DNDYFIRPIS YLAHNLYGAW
VSGNHSQYVG EDMDVTRVSD GWLIQGNNGG GCSGYRCSEK SSIKVSNSFY
TLDPGSFSHG LVTESGKQLV KTITATATNY TDLPPQVVVT LKYDKATNWS
KTDYGLSEK VGIKKTQIP QVSSTEYSVE ISSSQSWARQ EGGSATETVS
IEARPTVPPH SSVPVRYALY KANVSYPYEF KADINYDLTL NGFLRWGGNA
WHTHPDNRPT WNHTFVIGPF KDKASSIRYQ WDKRYIPGEV KWWDNWNTIQ
QNGADTMKNA LARVLRPVRA SITGDFHAES QFAGNIEIGA AVPIGGDSKV
RRARSVDSPA TGLRLEIPLD ASELALGFD NVQLTLEPAT DK

```


****NOTE:** The above gene/protein is for a bacterial virulence factor of the aerolysin gene family. Aerolysin exotoxins are initially translated in the bacterial cytoplasm as preproaerolysin; this corresponds to the sequences listed above in their entirety. Preproaerolysin is recognized by the type II (i.e. Sec-dependent) secretion apparatus and translocated across the cytoplasmic membrane, with concomitant cleavage of the N-terminal signal peptide (designated with a single-underline), yielding proaerolysin. Following outer-membrane translocation, proaerolysin undergoes a C-terminal cleavage event (cleaved portion designated with a double-underline). This yields the final aerolysin protein, which undergoes subsequent insertion and homopentamerization in the target eukaryotic membrane. In the above amino-acid sequence, the position of the cleaved signal peptide was predicted with the SignalP v3.0 algorithm (www.cbs.dtu.dk/services/SignalP/) and verified experimentally by N-terminal sequencing. The position of the C-terminal cleavage point was predicted (but not experimentally confirmed) by homology analysis with previously-characterized aerolysin proteins.

- Gene length: 1479 bp
- Initial protein transcript (preproaerolysin): 492 amino acids; 54.45 kDa; pI = 6.14
- Translocated protein (proaerolysin): 469 amino acids; 52.05 kDa; pI = 5.97
- Final, active protein (aerolysin): 426 amino acids; 47.45 kDa; pI = 6.23

Chapter 4: PlyM1 (Truncated)

- GenBank Accession Number: HM011589
- 591+ base pairs; 197+ amino acids; 22.12+ kDa
- N-acetylmuramoyl-L-alanine-amidase-Type 2 enzymatic domain (PF01510)
- No C-terminal binding domain was recognized. However, given the fact that this lysin was cloned as a truncated construct – along with the fact that its N-terminal region demonstrated Blast homology to other Gram-positive enzymes – it is likely that a binding domain does exist but could not be recognized by Pfam due to the truncation.

```
ATGAGTATTA ATGTAGTTAA GAATTTGGTG TCTTCAAGTA AATATAGTGT
TAAGTGTCCT TATCCTATGA ATCCAGAAAT AATTGTCGTA CACAATACGG
CAAACGATGC ATCGGCTAAG TCAGAAATTT CTTATATGAT TAACAACAAT
AATGAAGTTT CCTACCATTT CGCTGTTGAT GACAAAGAAG TTGTTCAAGG
ATTGCCATTA AATCGAAATG GTTGGCATGC AGGTGACGGT GGTTCGGGAC
GTGGTAACAG AAAAGGTATT GGAGTTGAAA TTTGTTATTC TAAGTCTGGT
GGTATTAAAT ATAAAAAAGC AGAAGTTTTA GCTATCAAAT TTATTGCACA
ATTATTACAT GAAAGAAATT GGTCAATTGA TAGAGTAAAA ACTCATAATC
AGATGAATGG CAAGTATTGT CCCCATAGAA TCCTATCTGA AGGAAGATGG
GATGATGTAC TACGAGCAAT CCAAAAAGAG TTAGATAGAT TAAATAATGT
GGAAGTAAAA CCTGCTCCAG ATACTCAAGT GGATTCAGAA ATAACAATA
AACCAAATAA ACCAAAAGAA GAAAACAAAG TGAATCAAT T...
```

```
MSINVVKNLV SSSKYSVKCP YPMNPEIIVV HNTANDASAK SEISYMINNN
NEVSYHFAVD DKEVVQGLPL NRNGWHAGDG GSGRGNRKGI GVEICYSKSG
GIKYKKAIEVL AIKFIAQLLH ERNWSIDRVK THNQMGKYC PHRI LSEGRW
DDVLRAIQKE LDRLNNVEVK PAPDTQVDSE ITTKPNKPKE ENKVESI...
```

Chapter 4: PlyM2

- GenBank Accession Number: HM011590
- 750 base pairs; 249 amino acids; 28.23 kDa; Theoretical pI = 9.11
- N-acetylmuramoyl-L-alanine-amidase-Type 2 enzymatic domain (PF01510)
- Alanine-amidase-type-2-associated putative binding domain (PF12123)

```

ATGAAAATCC CCGTAAGACA AATGCTCGTT AATTCAAAGA ATCATAACAT
TAAATGTCCT TTCGACATGC AAGCAGAGTA TATTACCATC CACAATACCT
ATAATGATGC AAGCGCCAAT AACGAAATAC AATATATGAT TAACAATGCA
CACGAAGTAT CCTTTCACCT TGCTGTAGAT GACAATGAGG TTGTTCAAGG
ATTGCCATTA AACCGTAACG GATGGCACTG CGGAGATGGC GGAAAAGGTT
CTGGAAATCG AAAGTCTATT GGTGTTGAAA TCTGTTACTC CAAGTCAGGA
GGTACCAAAT ACTACAAAGC AGAAGGATTA GCTATCCAAT TTGTTGCTCA
ACTTTTGCAT GAAAGAGGAT GGGGGATTGA TCGAGTAAAA AAACATCAGG
ATTGGTCAAA GAAACATTGC CCACATCGTA TTTTAGATGA AGGTCGCTGG
CAGTCAGTGC TGGATGCAAT CTCAAAGGAA TTAGATGCAC TTAAAAAACC
TGTTAAAACT TCCGTTGTTG CTGCCGTTGA AAATTCCCCT TCTGTTAAAA
TCCAGACGGG TGGTTTAAAT CCTGAAATGG TAAAGGAAAT ATCTGATTTC
TTCATACAAA ACAAATGGTA TGCTGAAATC ACGTTCAATT TCAAAAAAGG
AAATCCAACA GCTGTGACAG GAGGTTTATC CGGGGCTACC AGGGATAAAT
TTGAATCCTG GCTAAAAGAA CGCGGCTGGT GGTATAAAGT AATTAAGTAA

```

```

MKIPVRQMLV NSKNHNIKCP FDMQAEYITI HNTYNDASAN NEIQYMINNA
HEVSFHFVAVD DNEVVQGLPL NRNGWHCGDG GKGSGNRKSI GVEICYSKSG
GTKYYKA EGL AIQFVAQLLH ERGWGIDRVK KHQDWSKKHC PHRILDEGRW
QSVLDAISKE LDALKKPKVT SVVAAVENSP SVKIQTGGLN PEMVKEISDF
FIQNKWYAEI TFNFKKGNPT AVTGGLSGAT RDKFESWLKE RGWWYKVIK

```

Chapter 4: PlyM3 (Truncated)

- GenBank Accession Number: HM011591
- 921+ base pairs; 307+ amino acids; 33.21+ kDa
- N-acetylmuramoyl-L-alanine-amidase-Type 2 enzymatic domain (PF01510)
- PG-1 binding domain (PF01471)

```
ATGGCATACA CAATTGTTAA CAAATATATC CCTACATCTA AATACGGTAG
AAAAGCGCCA TACTCGATGA ATGCAGAAAC TATCACCTTC CACAATACAG
CAAATGACGC AACTGCGTTA GGCGAGATTT CTTACATGAC GAATAACAAT
AATCAAACCTT CCTATCACGT CGCTATTGAT GATAAACACG CTGTGCAGGC
GATTCCGTTT AACCGGTCAG CTTGGCACGC AGGCGATGGT CAGGGAGCAG
GAAATCGCAA GTCTATTGGT ATCGAGGTGT GCTATAGCAA AAGCGGCGGC
GTGAAATATG CGGCGGCAGA AGAAAACGCC ATTGAATATA TCGCACACAT
TCTCAAAGAT AAAGGTTGGG GAATTGATCG CGTGAAATGG CATCGTGAAT
GGTCGGGTAA AAATTGCCCG CATCGCGTAT TAGATGAAGG AAGAGCAACG
AGTGTGCGTA ACCGAATCAG CGCTAAATTA GCAGAGTTGA AAGGTGAGAA
GGTCGTTGAA GTTGCATCGC CTTCCACAGT GAAGAGAGAC TACCTATTAG
ATGGCGATAC AGGCGCAGCG GTTAAAACGC TACAAAGCGA ATTGAAGCAA
GCAGGTTTCT TACTATCTGT GGATGGCATA TTCGGAAAAG GCACAGAGAC
AGCGGTAAAG GCGTTCCAGC GTGCGAATGG ATTAGCCGTT GATGGCGTGT
TCGGAACAGG CTCACAGGCT AAACCTCAACG CAATACTGGC TAATTTGAAT
AAGAAGCCGG TAGTTAAGCC TGCAGCGCCA ACAAACCAA AGGAGGAATC
CACAGTGGAG AAAACAAACG AACCCTCAA ATGGGCAGAA GCGACGATTA
AAGAAGCGGT GAAGATTGGC GTTACCGATG GCAGCAACTT GCACGATCCA
GTCACAAGAC AGGAAGCAAT T...
```

```
MAYTIVNKYI PTSKYGRKAP YSMNAETITF HNTANDATAL GEISYMTNNN
NQTSYHVAID DKHAVQAIPF NRSAWHAGDG QGAGNRKSIG IEVCYSKSGG
VKYAAAEENA IEYIAHILKD KGWGIDRVKW HRDWSGKNCP HRVLDEGRAT
SVRNRISAKL AELKGEKVVE VASPSTVKRD YLLDGDTGAA VKTLQSELKQ
AGFLLSVDGI FGKGTETAVK AFQRANGLAV DGVFGTGSQA KLNAILANLN
KKPVVKPAAP TKPKEESTVE KTNEPSKWAE ATIKEAVKIG VTDGSNLHDP
VTRQEAI...
```

Chapter 4: PlyM4

- GenBank Accession Number: HM011592
- 990 base pairs; 329 amino acids; 35.61 kDa; Theoretical pI = 9.69
- N-acetylmuramoyl-L-alanine-amidase-Type 2 enzymatic domain (PF01510)
- PG-1 binding domain (PF01471)

```

ATGTTTACAA TCAAAAACA GATAGTTCCG GAATCGAACG TTAGAACGCG
CACGTTTGGC AGGTTGAACA AGAAGCGGTA CATCACAATC CACGAAACGG
GTAACGCCAG TAGAGGTGCA GGCGCTCAAA GCCACGCCAA TCTACAAAGC
AACACAAATC CACGCGCAGC TTCCTGGCAT TGGCAAGTGG ATGAAAAGGT
AGCGATTCAA TCATACGATC ATTCGGTATC GTGTTGGGCC GGTGGCGATG
GGCGAGGGAA TGGCAACATG AATTTCGATC ACATTGAAAT TTGTATCAAC
AGTGATGGCG ACTATTTGAA AACGATAGAA AACGCAGCGC AATTGGTCAA
GTATATAATG AAGGAAGAGA ATATTCCGTT GGCTAATGTG GTGCAGCATA
ACAAGTGGTC TGGCAAGAAT TGCCCTACGT TGTTACGCGC AGGCAATCGC
GGCATTAAAT GGGGCGGCTT CCTTAGCAAA GTTAACGGTG CGGTTGTATC
GGCTACGCCT AAGCCTTCCA CAGTGAAGAG AGACTACCTA TTAGATGGCG
ATACAGGCGA TAACGTCAGA GCGCTACAGA CCGGATTAAA GCAAGCAGGT
TTCTTACTAT CTGTGGATGG CATATTCGGT AAAGGCACAG AGACAGCGGT
TAAGGCGTTC CAACGTGCGA ATGGATTAGC AGTTGATGGC GTGTTCGGAA
CAGGCTCACA GGCTAAACTC AACGCAATAC TGGCGAATTT GAATAAGAAG
CCGGTAGTTA AGCCTGCAGC GCCATCAAAA CCAAAGGAGG AATCCACAGT
GGAGAAAACA AACCACCCT CAAAATGGGC AGAAGCGACG ATTAAAGAAG
CGGTGAAGAT TGGCGTTACG GATGGCAGCA ACTTGCACGA TCCAGTTACA
AGACAGGAGG CAATTGTGCT GCGGATGCGT GCCGCAGGAC TTGCGCCAAG
ACTTAAACAA CACACTTATA AATATTGTAA GAATTGTAA

```

```

MVTIKKQIVP ESNVRTRTFG RLNKKRYITI HETGNASRGA GAQSHANLQS
NTNPRAASWH WQVDEKVAIQ SYDHSVSCWA GGDGRGNGNM NSIHIEICIN
SDGDYLTIE NAAQLVKYIM KEENIPLANV VQHNKWSGKN CPTLLRAGNR
GINWGGFLSK VNGAVVSATP KPSTVKRDYL LDGDTGDNVR ALQTGLKQAG
FLLSVDGIFG KGTETAVKAF QRANGLAVDG VFGTGSQAKL NAILANLNKK
PVVKPAAPSK PKEESTVEKT NQPSKWAEAT IKEAVKIGVT DGSNLHDPVT
RQEAIVLAMR AAGLAPRLKQ HTYKYCKNC

```

Chapter 4: PlyM5 (Truncated)

- GenBank Accession Number: HM011593
- 853+ base pairs; 284+ amino acids; 30.86+ kDa
- N-acetylmuramoyl-L-alanine-amidase-Type 3 enzymatic domain (PF01520)
- PG-1 binding domain (PF01471)

```
ATGGCAAAAT TATATAACGA ACCAGGACAC GGCGTAAATA CATGGCCGCC
AAGCAAAGGG ACGCGCGCGA GTGGTGGCGT TCCTGAAATG GCAGAACATG
ATTTTAATGC AGCTGTAGCA GATGAAGTGA ATCGCCTGTT AAGCGGCAAG
CTAACGACTT ATAGCGCGCA GCCGAGTCGC GGCAAGGATG TTTCGCTAAC
GACGCGCACG CGCCTATATA ATGAAGAGTT TCGCAAAGAT AAGAGCGCAA
TTGGCTTCTC GCATCACGGA AACGCCAACG CTAACAAAGC GACAAAAGGA
TTCGGTGTAT TCTACTGGGG TGGATCTGCA ACGGGTAAGA AGTTGGCGCA
AATGCTACTC GCGGCGTATA AAAAAGAATT CCCTGGCTAT CCAATTTGGG
GCAGCGGGAT ATTCGAAAGC AAGCGCGGAG ACTGGACTAA CTTGCGCAATC
CTGCGCGATA CATCAGCGCC TTTCGTGTTG ATCGAGTGGG ATTTCTTTAC
GAATGATGAA GCGCGGAAAC GTATGCTATC CACAGATTAT AGAAAGCGCT
GCGGTAAGGT GCGGCAAGT GTCGCGTGTG ATTGGTACGG GATTCCGTTT
ACTGATTTCA CAGCAGCGAA ACCTACGCCT TCCACAGTGA AGAGAGACTA
CCTGTTAGAT GCGGATACAG GCGCAGCGGT TAAAACGCTA CAAAGCGAAT
TGAAACAAGC AGGTTTCTTA CTATCTGTGG ATGGCATATT CGGAAAAGGC
ACAGAGACAG CGGTGAAGGC GTTTCAGCGC GGCAACGGAC TAGTTGTTGA
TGGCGTGTTT GGAACAGGCT CACAGGCTAA ACTCAACGCA ATACTGGCTA ATT...
```

```
MAKLYNEPGH GVNTWPFSKG TRASGGVPEM AEHDFNAAVA DEVNRLLSGK
LTTYSAQPSR GKDVSLTTRT RLYNEEFRKD KSAIGFSHHG NANANKATKG
FGVIFYWGSA TGKKLAQMLL AAYKKEFPY PIWGSIFES KRGDWTNFAI
LRDTSAPFVL IEWDDFTNDE ARKRMLSTDY RKRCGKVAAS VACDWYGIPF
TDFTAAKPTP STVKRDYLLD GDTGAAVKTL QSELKQAGFL LSVDGIFGKG
TETAVKAFQR GNGLVVDGVF GTGSQAKLNA ILAN...
```

Chapter 4: PlyM6

- GenBank Accession Number: HM011594
- 1074 base pairs; 357 amino acids; 38.63 kDa; Theoretical pI = 9.62
- N-acetylmuramoyl-L-alanine-amidase-Type 3 enzymatic domain (PF01520)
- PG-1 binding domain (PF01471)

```

GTGGCGAAGT TATATAACGA ACCAGGACAC GGCGTAAACA CATGGCCGCC
AAGCAAAGGG ACTCGCGCAA GTGGCGGCGT TCCCGAAATG GCGGAACATG
ATTTTAACGC AGCGGTAGCA GATGAAGTGA ATCGCCTATT AAGCGGCAAG
CTAACGACTT ATAGCGCGCA GCCTAGTCGC GGCAAGGATG TTTCGCTAAC
GACGCGCACA CGCCTATATA ATGAAGAGTT TAAGAAGGAC AGAAGCGCAA
TCGGATTTTC ACACCACGGA AACGCCAATG CCAACAAAGC AACTAAAGGA
TTCGGTGTAT TCTACTGGGG CGGATCTGCA ACAGGCAAAA AGTTGGCGCA
AATGCTACTT GCTGCATACA AGAAAGAATT CCCTGGCTAT CCGATTTGGG
GCAGCGGGAT ATTTGAAAGC AAGCGCAGCG ACTGGACCAA CTTTGCTATC
TTGCGTGATA CGTCAGCGCC TTTCGTGTTG ATCGAGTGGG ATTTCTTTAC
GAATGATGAA GCGCGCAAAC GTATGCTATC CACAGATTAT AGAAAGCGCT
GCGGTAAGGT GCGGCAAGT GTCGCGTGTG ATTGGTACGG CATTCCGTTT
ACTGATTTTA CAGCAGCGAA GCCTACGCCT TCCACAGTGA AGAGAGACTA
CCTATTAGAT GGCGTTACAG GCGCAGCGGT TAAAACGCTA CAAAGCGAAT
TGAAGCAAGC AGGTTTCTTA CTATCTGTGG ATGGCGTATT CGGAAAAGGC
ACAGAGACAG CGGTTAAGGC GTTCCAACGT GCGAATGGAT TAGCCGTTGA
TGGCGTGTTT GGAACAGGCT CACAGGCTAA ACTCAACGCA ATACTGGCGA
ATCTGAATAA GAAGCCGGCA GTTAAGCCTG CAGCGCCAAC AAAACCAAGG
GAGGAATCCA CAGTGGAGAA AACAAACCAA CCCTCAAAAT GGGCAGAAGC
GACGATTAAA GAAGCAGTGA AGATTGGCGT TACGGATGGC AGCAGCTTAC
ACGATCCAGT CACAAGACAG GAAGCAATTG TGCTGGCGAT GCGTGCTGCA
GGACTTGCGC CAAAACTTAA GTAG

```

```

MAKLYNEPGH GVNTWPPSKG TRASGGVPEM AEHDFNAAVA DEVNRLLSGK
LTTYSAQPSR GKDVSLTTRT RLYNEEFKGD RSAIGFSHHG NANANKATKG
FGVFWGGS TGKKLAQMLL AAYKKEFPY PIWGSIFES KRSDWTNFAI
LRDTSAPFVL IEWDFFTND ARKRMLSTDY RKRCGKVAAS VACDWYGIFP
TDFTAAPTP STVKRDYLLD GVTGAAVKTL QSELKQAGFL LSVDGVFGK
TETAVKAFQR ANGLAVDGVF GTGSQAKLNA ILANLNKKPA VKPAAPTKPR
EESTVEKTNQ PSKWAEATIK EAVKIGVTDG SSLHDPVTRQ EAIVLAMRAA
GLAPKLK

```

Chapter 4: PlyM7

- GenBank Accession Number: HM011595
- 1071 base pairs; 356 amino acids; 38.53 kDa; Theoretical pI = 9.41
- N-acetylmuramoyl-L-alanine-amidase-Type 3 enzymatic domain (PF01520)
- PG-1 binding domain (PF01471)

```

ATGGCAAAAT TATATAACGA ACCAGGACAC GGCGTAAACA CATGGCCGCC
AAGCAAAGGG ACGCGCGCAA GTGGTGGCGT TCCCGAAATG GCGGAACATG
ATTTTAATTC AGCTGTAGCA GATGAAGTGA ATCGCCTGTT AAGTGGCAAG
CTAACGACTT ATAGCGCGCA GCCTAGTCGC GGCAAGGATG TTTCGCTAAC
GACGCGCACA CGCCTATATA ATGAAGAGTT TCGCAAAGAT AAGAGCGCAA
TCGGCTTCTC GCATCACGGA AATGCCAACG CTAACAAAGC GACAAAAGGA
TTCGGCGTAT TCTACTGGGG CGGATCGGCA ACGGGTAAGA AGTTAGCGCA
AATGCTACTC GCGGCGTATA AAAAAGAACT CCCTGGCTAT CCGATTTGGG
GCAGCGGGAT TTTTGAAAGC AAACGCGGCG ACTGGACCAA CTTGCTATC
TTGCGTGATA CGTCAGCGCC TTTTGTATTA ATCGAGTGGG ATTTCTTTAC
GAACGATGAA GCGCGCAAAC GGATGCTATC CACAGATTAT AGAAAGCGTT
GCGGTAAAGT GGCAGCGAGT GTTGCGTGTG ATTGGTACGG CATTCCGTTT
ACTGATTTCA CTGCAGCTAA ACCTTCGCCG TCCACAGAGA GAGACTACCT
ATTAGATGGC GATACAGGCG CAGCGGTAA AACGCTACAA AGCGAATTGA
AGCAAGCGGG TTTCTTACTA TCTGTGGATG GCATATTCGG AAAAGGCACA
GAGACAGCGG TTAAGGCGTT CCAACGTGCG AATGGATTAG CAGTTGATGG
TGTTTTTCGA ACAGGCTCAC AGGCTAAACT CGACGCAATA CTGGCTAATC
TGAATAAGAA GCCGGTAGTT AAGCCTGCAG CGCCAACAAA ACCAAAGGAG
GAATCCACAG TGGAGAAAAC AAACCAACCC TCAAAATGGG CAGAGGCGAC
GATTAAAGAA GCGGTGAAGA TTGGCGTTAC GGATGGCAGC AACTTACACG
ATCCAGTTAC ACGACAGGAG GCAATTGTGC TGGCGATGCG TGCTGCAGGA
CTTGCGCCAA AACTTAAGTA A

```

```

MAKLYNEPGH GVNTWPPSKG TRASGGVPEM AEHDFNSAVA DEVNRLLSGK
LTTYSAQPSR GKDVSLTTRT RLYNEEFRKD KSAIGFSHHG NANANKATKG
FGVFWGGS TGKKLAQMLL AAYKKELPGY PIWGS GIFES KRGDWTNFAI
LRDTSAPFVL IEWDFFTND ARKRMLSTDY RKRCGKVAAS VACDWYGIPF
TDFTAAPSP STERDYLLDG DTGAAVKTLQ SELKQAGFLL SVDGIFGKGT
ETAVKAFQRA NGLAVDGVFG TGSQAKLDAI LANLNKKPVV KPAAPTKEPKE
ESTVEKTNQP SKWAEATIKE AVKIGVTDGS NLHDPVTRQE AIVLAMRAAG
LAPKLK

```


Chapter 4: PlyM8

- GenBank Accession Number: HM011596
- 1074 base pairs; 357 amino acids; 38.72 kDa; Theoretical pI = 9.62
- N-acetylmuramoyl-L-alanine-amidase-Type 3 enzymatic domain (PF01520)
- PG-1 binding domain (PF01471)

```

ATGGCGAAAT TATATAACGA ACCAGGGCAC GCGGTAAACA CATGGCCGCC
AAGCAAAGGG ACGCGCGCAA GTGGCGGCGT TCCCGAAATG GCGGAACATG
ATTTTAACGC AGCGGTAACA GATGAAGTGA ATCGCCTGTT AAGCGGCAAG
CTAACGACTT ATAGCGCGCA GCCGAGTCGC GGCAAGGATG TTTTCGCTAAC
GACGCGTACA CGCCTATATA ATGAAGAGTT TCGCAAAGAT AGAAGCGCAA
TCGGTTTTTC TCACCACGGA AACGCCCATG CCAACAAAGC GACAAAAGGA
TTCGGCGTAT TCTACTGGGG CGGATCTGCA ACGGGTAAGA AGTTGGCGCA
AATGCTACTC GCGGCGTATA AAAAAGAATT CCCTGGTTAT CCAATTTGGG
GCAGTGGGAT ATTCGAAAGC AAGCGCGGCG ACTGGACAAA CTTTGCAATT
CTGCGCGACA CATCAGCGCC TTTCGTATTA ATCGAGTGGG ATTTCTTTAC
GAATGATGAA GCGCGCAAAC GGATGCTATC CACAGATTAC AGAAAGCGTT
GCGGTAAAGT GGCAGCGAGT GTCGCGTGCG ATTGGTACGG TATTCGTTTT
ACTGATTTTA CAGCAGCGAA ACCTACGCCT TCCACAGTGA AGAGAGATTA
CCTACTACTT GACGATACCG GCGCAGCGGT TAAAACGCTA CAAAGCGAAT
TGAAACAAGC AGGTTTCTTA CTATCTGTGG GTGGCATATT TGGTAAAGGT
ACAGAGACAG CGGTTAAGGC GTTCCAACGT GCGAATGGAT TAGCAGTTGA
TGGCGTGTTT GGAACAGGCT CACAGGCTAA ACTCAACGCA ATACTGGCGA
ATCTGAATAA GAAGCCGGTA GTTAAGCCTG CAGCGCCAAC AAAACCAAAG
GAGGAATCCA CAGTGGAGAA AACAAACCTA CCCTCAAAAT GGGCAGAAGC
TACGATTAAA GAAGCGGTGA AGATTGGCGT TACCGATGGC AGTAACTTAC
ACGATCCGGT CACAAGACAG GAGGCAATTG TGCTGGCGAT GCGTGCTGCG
GGACTCGCGC CAAAACTCAA GTAG

```

```

MAKLYNEPGH GVNTWPPSKG TRASGGVPEM AEHDFNAAVT DEVNRLLSGK
LTTYSAQPSR GKDVSLTTRT RLYNEEFRKD RSAIGFSHHG NAHANKATKG
FGVFWGGSA TGKKLAQMLL AAYKKEFPGY PIWGSGIFES KRGDWTNFAI
LRDTSAPFVL IEWDFFTNDE ARKRMLSTDY RKRCGKVAAS VACDWYGIPF
TDFTAAKPTP STVKRDYLLL DDTGAAVKTL QSELKQAGFL LSVGGIFGKG
TETAVKAFQR ANGLAVDGVF GTGSQAKLNA ILANLNKKPV VKPAAPTKPK
EESTVEKTNL PSKWAEATIK EAVKIGVTDG SNLHDPVTRQ EAIVLAMRAA
GLAPKLK

```

Chapter 4: PlyM9 (Truncated)

- GenBank Accession Number: HM011597
- 1029+ base pairs; 343+ amino acids; 37.2+ kDa
- N-acetylmuramoyl-L-alanine-amidase-Type 3 enzymatic domain (PF01520)
- PG-1 binding domain (PF01471)

```
ATGGCAAAAT TATATAACGA ACCAGGACAC GGCGTAAATA CATGGCCGCC
AAGCAAAGGG ACGCGCGCAA GTGGTGGCGT TCCTGAAATG GCGGAACATG
ATTTTAACGC AGCGGTAGCA GATGAAGTGA ATCGCCTTTT AAGCGGCAAG
CTAACGACTT ATAGCGCACA GCCGAGTCGC GGCAAGGATG TTTCGCTAAC
GACGCGCACA CGCCTTTACA ACGCTGAATA CAGCAAAGAC AGAAGCGCAA
TCGGTATGTC ACATCACGGA AACGCCACG CCCACAAAGC AACAAAAGGT
TTCGGCGTAT TCTATTGGGG CGGATCTACC ACAGGTAAGA AGTTGGCGCA
AATGCTACTC GCGGCGTATA AGAAAGAATT CCCTGGCTAT CCGATTTGGG
GCAGCGGGAT ATTCGAAAGT AAGCGCGGAG ACTGGACTAA CTTTGCAATC
CTGCGCGATA CATCAGCGCC TTTTGTATTA ATTGAGTGGG ATTTCTTTAC
GAATGATGAA GCGCGGAAAC GTATGCTGTC CACAGATTAT AGAAAGCGTT
GCGGTAAAGT GGCAGCGAGT GTCGCGTGTG ATTGGTACGG TATTCGTTT
ACTGATTTTA CAGCAGCGAA ACCTACGCCT TCCACAGAGA AGAGAAACTA
CCTATTAGAT GGCGATACAG GCGCAGCGGT TAAAACGCTA CAAAGCGAGT
TGAAGCAAGC AGGTTTCTTA CTATCTGTGG ATGGCATATT CGGAAAAGGC
ACAGAGACAG CGGTTAAGGC GTTCCAACGT GCGAATGGAT TAGCGGTTGA
TGGCGTGTTT GGAACAGGCT CACAGGCTAA ACTCAACGCA ATACTGGCTA
ATTTGAATAA GAAGCCGGTA GTTAATCCTG CAGCGCCAAC GAAACCAAAG
GAGGAATCCA CAGTGGAGAA AACAAACCAA CCCTCAAAAT GGGCAGAAGC
GACGATTAAA GAAGCAGTTA AGATTGGCGT TACCGATGGC AGCAGCTTAC
ACGATCCAGT CACAAGACAG GAAGCAATT...
```

```
MAKLYNEPGH GVNTWPPSKG TRASGGVPEM AEHDFNAAVA DEVNRLLSGK
LTTYSAQPSR GKDVSLTTRT RLYNAEYSKD RSAIGMSHHG NAHAHKATKG
FGVFWGGST TGKKLAQMLL AAYKKEFPY PIWGSIFES KRGDWTNFAI
LRDTSAPFVL IEWDFFTND ARKRMLSTDY RKRCGKVAAS VACDWYGIPF
TDFTAAKPTP STEKRNLLD GDTGAAVKTL QSELKQAGFL LSVDGIFGKG
TETAVKAFQR ANGLAVDGVF GTGSQAKLNA ILANLNKKPV VNPAAPTKPK
EESTVEKTNQ PSKWAEATIK EAVKIGVTDG SSLHDPVTRQ EAI...
```

Chapter 4: PlyM10 (Truncated)

- GenBank Accession Number: HM011598
- 848+ base pairs; 282+ amino acids; 30.7+ kDa
- N-acetylmuramoyl-L-alanine-amidase-Type 3 enzymatic domain (PF01520)
- PG-1 binding domain (PF01471)

```
ATGGCAAAAT TATATAACGA ACCAGGACAC GGCGTAAACA CATGGCCGCC
AAGCAAAGGG ACGCGCGCAA GTGGTGGCGT TCCCGAAATG GCGGAACATG
ATTTTAACGC AGCGGTAGCA GATGAAGTAA ATCGCCTGTT AAGCGGCAAG
CTAACGACTT ATAGCGCACA GCCTAGTCGC GGCAAGGATG TTTCGCTAAC
GACGCGCACA CGCCTATATA ATGAAGAGTT TCGCAAAGAT AAGAGCGCAA
TTGGCTTCTC ACATCACGGA AACGCTAATT CCAACAAAGC GACAAAAGGT
TTCGGCGTAT TCTATTGGGG TGGATCTGCA ACGGGTAAGA AGTTGGCGCA
AATGCTACTC GCGGCGTATA AGAAAGAATT CCCTGGCTAT CCGATTTGGG
GCAGCGGGAT ATTCGAAAGT AAGCGCGGCG ACTGGACCAA CTTCGCTATC
TTGCGTGATA CGTCAGCGCC TTTCGTGTTA ATCGAGTGGG ATTTCTTTAC
GAATGACGAA GCGCGCAAAC GGATGCTATC CACAGATTAT AGAAAGCGTT
GCGGTAAAGT GGCAGCGAGT GTCGCGTGTG ATTGGTACGG GATTCCGTTT
ACTGATTTCA CAGCAGCGAA ACCTTCGCCG TCCACAGAGA GAGACTACCT
ATTAGATGGC GATACAGGCG CAGCGGTAA AACGCTACAA AGCGAATTGA
AACAAGCGGG TTTCTTACTA TCTGTGGATG GCATATTCGA TAAAGGTACA
GAGACAGCGG TTAAGGCGTT CCAACGTGCG AATGGATTAG CCGTTGATGG
CGTGTTTCGA ACAGGCTCAC AGGCTAAACT CAACGCAATA CTGGCTTA...
```

```
MAKLYNEPGH GVNTWPPSKG TRASGGVPEM AEHDFNAAVA DEVNRLLSGK
LTTYSAQPSR GKDVSLTTRT RLYNEEFRKD KSAIGFSHHG NANSNKATKG
FGVFIWGGSA TGKKLAQMLL AAYKKEFPY PIWGSIFES KRGDWTNFAI
LRDTSAPFVL IEWDDFTNDE ARKRMLSTDY RKRCGKVAAS VACDWYGIFP
TDFTAAKPSP STERDYLLDG DTGAAVKTLQ SELKQAGFLL SVDGIFDKGT
ETAVKAFQRA NGLAVDGVFG TGSQAKLNAI LA...
```

Chapter 4: PlyM11

- GenBank Accession Number: HM011599
- 1071 base pairs; 356 amino acids; 38.52 kDa; Theoretical pI = 9.45
- N-acetylmuramoyl-L-alanine-amidase-Type 3 enzymatic domain (PF01520)
- PG-1 binding domain (PF01471)

```

ATGGCGAAGT TATATAACGA ACCAGGACAC GGCGCTAACA CATGGCCGCC
AAGTAAAGGG ACGCGCGCAA GTGGTGGCGT TCCTGAAATG GCGGAGCATG
ATTTTAACGC AGCGGTAGCA GATGAAGTTA ATCGCCTGTT AAGCGGAAAG
CTAACGACTT ATAGCGCACA GCCTAGTCGC GGCAAGGATG TTTGCTGAC
GACGCGCACA CGCCTATATA ATGAAGAGTT TCGCAAAGAT AAGAGCGCAA
TTGGCTTCTC ACATCACGGA AATGCTAATG CTAACAAAGC GACAAAAGGA
TTGCGCGTAT TTTATTGGGG CGGATCGGCA ACGGGTAAGA AGTTAGCGCA
AATGCTACTC GCGGCGTATA AAAAAGAATA CCCTGGCTAT CCGATTTGGG
GCAGCGGAAT ATTCGAAAGT AAACGCGGCG ACTGGACAAA CTTTGCAATT
CTGCGCGACA CATCAGCGCC ATTTGTATTA ATCGAGTGGG ATTTCTTTAC
GAATGATGAA GCGCGTAAAC GGATGCTATC CACAGATTAT AGAAAGCGTT
GCGGTAAAGT GGCAGCGAGT GTCGCGTGTG ATTGGTACGG AATTCGTTT
ACTGATTTTA CAGCAGCGAA ACCTACGCCT TCCACAGAGA GAGACTACCT
ATTAGATGGC GATACCGGCG CAGCGGTAA AACGCTACAA AGCGAATTGA
AACAAGCGGG TTTCTTACTA TCTGTGGATG GCATATTCGG AAAAGGCACA
GAGACAGCGG TTAAGGCGTT CCAACGTGCG AATGGATTAG CCGTTGATGG
CGTGTTTCGA ACAGGCTCAC AGGCTAAACT CAACGCAATA CTGGCTAATT
TGAATAAGAA GCCGGTAGTT AAGTCGGCAG CGCCAACAAA ACCAAAGGAG
GAATCCACAG TGGAGAAAAC AAACCAACCA TCAAAATGGG CAGAAGCGAC
GATTAAAGAA GCCGTTAAGA TTGGCGTTAC GGATGGCAGC AGCTTACACG
ATCCAATTAC ACGACAGGAA GCAATTGTGC TGGCGATGCG TGCTGCAGGA
CTTGCACCAA AACTTAAGTA A

```

```

MAKLYNEPGH GANTWPPSKG TRASGGVPEM AEHDFNAAVA DEVNRLLSGK
LTTYSAQPSR GKDVSLTTRT RLYNEEFRKD KSAIGFSHHG NANANKATKG
FGVIFYWGSA TGKKLAQMLL AAYKKEYPGY PIWGSIFES KRGDWTFNFAI
LRDTSAPFVL IEWDFFTNDE ARKRMLSTDY RKRCGKVAAS VACDWYGIPF
TDFTAAPKPTP STERDYLLDG DTGAAVKTLQ SELKQAGFLL SVDGIFGKGT
ETAVKAFQRA NGLAVDGVFG TGSQAKLNAI LANLNKKPVV KSAAPTCKPKE
ESTVEKTNQP SKWAEATIKE AVKIGVTDGS SLHDPITRQE AIVLAMRAAG
LAPKLK

```

Chapter 4: PlyM12

- GenBank Accession Number: HM011600
- 1017 base pairs; 338 amino acids; 37.39 kDa; Theoretical pI = 9.36
- [N-acetylmuramoyl-L-alanine-amidase-Type 2 enzymatic domain \(PF01510\)](#)
- [Additional conserved domain of unknown function DUF 3957 \(PF12200\)](#)
- [PG-1 binding domain \(PF01471\)](#); [PG-3 binding domain \(PF09374\)](#)

****NOTE:** The above binding domains show only low homology to the consensus sequences for PG-1 and PG-3, with respective E-value scores of 0.0045 and 0.0056.

ATGGTAGCAT	TAAATTATAA	GAAAGATTAC	GTAATTAAAA	ATAAATACTC
ACGCCCTGGG	TTGAAGCTAC	TTGGCGTAAG	GGCTATCGTT	CTACACTACA
CGGCTAGTCC	GGGAGGTTCC	GCGCAAAACC	ACCAAGACTT	CTTCGATGGG
GCTGACGGAG	GCGCCTACCG	TTACGCAGGG	GCGCACATAT	TCGTAGATAA
AAAGGAAGCA	ATCGAGATCA	TCCCGTTAAA	CGAGGTTGCA	TATCACGCCA
ACGAGAGGTC	GCCTAAACTG	TCGTACCTAC	GGGCGTCAAC	GTCTTCTTAC
AACGGAAATG	CGAACCTAAC	AACGATCGGC	ATCGAGATGT	GTATCGAAAA
AGACGGATCC	TTCCACCCGG	ACACTGTCTGA	GCGTACAGCG	CTAGTAGTCG
AGAAGCTACG	TAAACAGTTC	GGCAAAGTCC	CTGTAATCCG	TCACTACGAT
GTAACAGGCA	AGATCTGCCC	GAAACCTTTC	GTAGAGGACG	TATCTGCGTG
GAAGGCGTTC	TTAAAGCGCT	TAGACGGTGG	CGTTACGGTC	AAGCCTGCGC
CTAAAACCGA	AGTTAAGGCC	GTGTCCACAG	AAAAAGAATA	CACGAGCATC
GTCGAGTATC	TGAAAGATCT	CGGAAAACCT	TACTCGTTCT	CTTACCGTAA
AGAACTAGCG	AAGGATTACG	GGATCACTAA	TTACGAAGGT	TCAGCAGCGC
AAAACCTAGA	GTTGCTCGAT	CTTCTCCAAA	ACGGAAGACC	GAAAGCCAAC
CTGACGGTAG	ACGGTTACTT	CGGACCGGCA	ACGATAAAAG	CATTACAGCG
TTACTTCGGT	ACCCCGGTTG	ACGGCGTAAT	CAGCGAACCT	TCGTTGGTTG
TAAAAGCCCT	GCAGGGGTTA	CTAGACGTGA	GACAAGACGG	TTACATGGGT
CCAATCACGA	TCAAGGCGTT	ACAGAAACGT	TTCGGCACTC	CGGTTGACGG
CGTAATTAGC	AAACCGTCTC	TCGTCATCAA	AGAGTTACAA	AGACGTTTAA
ATAACGGTAA	ACTGTAA			

MVALNYKKDY	VIKNKYSRPG	LKLLGVRAIV	LHYTASPGGS	AQNHQDFFDG
ADGGAYRYAG	AHIFVDKKEA	IEIIPLNEVA	YHANERSPKL	SYLRASTSSY
NGNANLTTIG	IEMCIEKDGS	FHPDVERTA	LVVEKLRKQF	GKVPVIRHYD
VTGKICPKPF	VEDVSAWKAF	LKRLDGGVTV	KPAPKTEVKA	VSTEKE <u>EYTSI</u>
<u>VEYLKDLGKP</u>	<u>YSFSYRKELA</u>	<u>KDYGITNYEG</u>	<u>SAAQNLEL</u>	LD LLQNGRPKAN
<u>LTVDGYFGPA</u>	<u>TIKALQRYFG</u>	<u>TPVDGVI</u>	SEP SLV <u>VKALQGL</u>	<u>LDVRQDGYMG</u>
<u>PITIKALQKR</u>	FGTPVDGVIS	KPSLVIKELQ	RRLNNGKL	

Chapter 4: PlyM13

- GenBank Accession Number: HM011601
- 699 base pairs; 232 amino acids; 25.07 kDa; Theoretical pI = 9.52
- N-acetylmuramoyl-L-alanine-amidase-Type 3 enzymatic domain (PF01520)
- SPOR binding domain (PF05036)

```
ATGGCGAAAA GGAAAGTATA TATTGATTTA GGGCACGAAG GCGACGGAAA
GGGAATGGAT CCTGGAGCAG TTGCCAATGG GTTGAAAGAA GCTAATGTTG
TTTTAGAGAT TGGGAAGTAT ATGAAAGATA TGTTTGCCAA CTATGAGAAT
GTAGAAGTGA AGTTTTCAAG ATTGGCAAAT AAAAATCTTT CTTTAAATCA
GCGCACGAAT GAAGCTAATG CCTGGGGTGC TGATGTTCTC TGTTCAATCC
ATATTAATGC AGGCGGCGGT AAAGGGTTTG AATCCTTCAT ATATCCAGGT
GCAGGATCAG CAACGCAGGC TTTCCAAAAT GCTGTTCATG CCAAGATTAT
GGGAACAGGG GTATTCACGA CAGACAGAGG GAAGAAAAAG GCGAATTTCC
ATATGTTAAG AGAATCTAAA ATGACCGCAA TCTTAACCGA AAATGGGTTC
ATTGATAACG GTACAGACGC AGCAACTTTG AAAAACAGAG CAAAGTTAAT
CGCTATAGCT TCAGGTCATG TGGAGGGAGT AGCTTCTTTC CTTAACCTTA
AAAAGAAACC TGTTTCTAAA CCAGAGCCGA AACCATCTGG TAAGTTATAC
AAAGTTCAAG TAGGTGCTTT CTCTGACCGA AAGAACGCTG ACAACTTAGC
TGCTGAACTA AAGAAAATAG GCTATTCAAC TTATATTGTT CACGAATAA
```

```
MAKRKVYIDL GHEGDGKGMD PGAVANGLKE ANVVLEIGKY MKDMFANYEN
VEVKFSRLAN KNLSLNQRTN EANAWGADVL CSIHINAGGG KGFESFIYPG
AGSATQAFQN AVHAKIMGTG VFTTDRGKKK ANFHMLRESK MTAILTENGF
IDNGTDAATL KNRAKLIAIA SGHVEGVASF LNLKKKPVSK PEPKPSKLY
KVQVGAFSDR KNADNLAAEL KKIGYSTYIV HE
```

Chapter 4: PlyM14

- GenBank Accession Number: HM011602
- 792 base pairs; 263 amino acids; 28.8 kDa; Theoretical pI = 9.78
- N-acetylmuramoyl-L-alanine-amidase-Type 3 enzymatic domain (PF01520)
- SPOR binding domain (PF05036)

```

ATGACGAAGA TAATCGCTAT CGACGCTGGC CACGGACTTA ATACTCCGGG
CAAACGTACG CCGGATAATG AGCGCGAATG GTCGTTTAAT AATAAGGTCA
CACTAGCCGC GATTAAGTAT TTGAACGATT ATGAAGGCGT TAAAATCGTA
AGACTCGACG ATCCGACCGG TAAATCGGAC GTACCTTTAA AAGCGCGCAC
CGACAAGGCG AATAAGGCGA AGGCTGACGT ACTAGTGTCG ATTCATCATA
ATGCGCTCAC GGGCAAGTGG GGAACGCACG GCGGTACGGA AGTATTTACG
TATCTCGGAA ACTGGCCGGA CGCGGAAAGG CTGGCGAAGT TAGTACTCGA
TCGTATCCTA AAAGCTTACG GATTAAAGAG TCGCGGACTA AAGAAGGCTA
ACTTCCACAT GGTACGCGAG AGCGCTATGC CTGCGATATT AATCGAAGGA
GGCTTCATGG ACTCGACCGT CGATATTAAG AAGATGCGCG ACGATAAAGT
CCTCGATGAA GCCGGCAAAG CTATCGCGGA AGCACTGGCG GTCTACTTCG
GATTAAAGAA GAAGAAACCC GCAGCACCTT CGAAACTGTA TCGCGTACAA
ATCGGAGCTT ATTCGGTAAA AGCGAACGCT GACGCACAAG CTGCGAAAGC
TAAACGCGCA GGCTATTTCGC CATATATCGC ACGCGAAGGC GGACTTTACA
AAGTTCAAAT CGGAGCTTAT TCCGTCAAAG CGAACGCCGA TAAAATGGCA
TCGGAATTGA AGCGCAAAGG CTTCAACGTA TATATTGCGT AA

```

```

MTKI IAIDAG HGLNTPGKRT PDNEREWSFN NKVTLAAIKY LNDYEGVKIV
RLDDPTGKSD VPLKARTDKA NKAkadVLVS IHhNALTGKW GTHGGTEVFT
YLGnWpDAER LAKLVLDRIl KAYGLKSRGL KKANFhMVRE SAMPAILIEG
GFMDSTVDIK KMRDDKVLDE AGKAIAEALA VYFGLKkkkP AAPSKLYRVQ
IGAYSVKANA DAQAaKAKRA GYSPYIAREG GLYKVQIGAY SVKANADKMA
SELKRKGfNV YIA

```


Chapter 4: PlyM15

- GenBank Accession Number: HM011603
- 693 base pairs; 230 amino acids; 25.71 kDa; Theoretical pI = 9.17
- N-acetylmuramoyl-L-alanine-amidase-Type 3 enzymatic domain (PF01520)
- SPOR binding domain (PF05036)

```
ATGAGCAAAT GGATTCAAGA TGCAGGACAC GGAGGAAAAG ACCCTGGCGC
AGTTACGAAA GGAAATACCG AAAAAGTTTA CACTCTTGAA GCTGCACTTT
ATGTAGACAA GCGTCTGGAT GAATTAGGGA TTAAAAGTGA TGTGACTCGT
TCAAGTGATG TTACGTTAGA TCAAGGCCCG CGTACTGGAA AAGTTAAAGC
TTTCGACAAG TGTATTTCTC ACCACTTTAA CGCAGGGGGC GGTAGTGGTT
TTGAAGCAAT TCATTCTATC TACTCAAACG GGAAGTTTGA ACACCTAATT
GCAGAAGAAT TTAAAAAAGC GGGATATCCA GTGCGACCTC GTTCAGTGTA
TTTTAAAAAA TACGGAAACA ACGATTATTA CTACATGCAC CGCCAAACAG
GTAAATGTCTG GACTACGATT GTAGAATATG ACTTTGTGGA CGGTCCGCAA
TCTGAAAAGA TTAAGGATAA AGCCTATCGT GAAGGCATGT ATGAATGTGT
AGTGCGAGCC ATTTGCCGTG ATGAAGGGGT AACATATAAA GCGCCTAATC
AACCTAAACC ACAATCTAAA CCATCTAATA AAGGCGGATT ATACAAGGTG
CAAGTAGGTG CATTTTTCGA TAAATCAAAC GCAGATAAAC TTGCAGCAGA
TCTAAAGAAA AAAGGCTACA GCACCTATAT TGTACAGGAA TAA
```

```
MSKWIQDAGH GGKDPGAVTK GNTEKVYMLE AALYVDKRLD ELGIKSDVTR
SSDVTLDDQP RTGKVKAFDK CISHHFNAGG GSGFEAIHSI YSNGKFEHLI
AEEFKKAGYP VRPRSVYFKK YGNNDYYMH RQTGKCRTTI VEYDFVDGPQ
SEKIKDKAYR EGMYESVVRV ICRDEGVTYK APNQPKPQSK PSNKGLLYKV
QVGAFSDKSN ADKLAADLKK KGYSTYIVQE
```

Chapter 4: PlyM16

- GenBank Accession Number: HM011604
- 702 base pairs; 233 amino acids; 26.15 kDa; Theoretical pI = 8.98
- N-acetylmuramoyl-L-alanine-amidase-Type 3 enzymatic domain (PF01520)
- SPOR binding domain (PF05036)

```
ATGAGTGAAA AGTGGAAGAA CGATGGTGGT CATGGCGGTA CAGATCCCGG
AGCGGTTGCC AATGGTATCA AGGAAAAGGA ATACACGCTT GAAGCGGCTT
TATATGTAGA CAAGCGTTTA AAGGAGCATG GTATTGACTC AGGACTCACA
AGAGATAAAG ACATCACCTT TGAACCAAAC GCGCGTACTG CGGTTGTCAG
AGCTTCTAAA GCACCATTCG GATTAAAGTCA CCATTACAAT GCCGCGGGTG
GTGCCGGTGC AGAGTTCATC CATTCCATTC ATTCTGATGG CAAGTTTGAA
AACATGCTTG CTGATGAATT TAAAAGAGCC GGTATCCTG TTCGTAGAGT
ATTCTGCAAA GCTGGCAAAA ACCCGGCAAA GGACTATTAC TACATGCACA
GGGAAACGGG CTTTTCGAGA ATGACGATTG TGGAATATGA TTTTGTTGAC
GGCCCGAACG CTGAAAAGCT GAAAGACCGT AAATATCGCG AGTGTATGTA
TGAGTGCGTA GTGAGGGCTG TATGTCGTCA GGAAGGGAAA GCCTATAAAC
CTGTGCGAGCA ACCGAAACTA AAGCCACAGG AAGCCCCTAA GAAGGGATTA
TACAAGGTTT AAGTGGGCGC ATTCGGCGAT AAATCAAATG CTGACCGACT
CGCCAAAGAA TTAGAGGGCA AAGGGTATAA AACGTATATT GTGCAGGAGT AA
```

```
MSEKWKNDGG HGGTDPGAVA NGIKEKEYTL EAALYVDKRL KEHGIDSGLT
RDKDITLEPN ARTAVVRASK APFGLSHHYN AGGGAGAEFI HSIHSDGKFE
NMLADEFKRA GYPVRRVFCK AGKNPAKDYY YMHRETGFRCR MTIVEYDFVD
GPNAEKLKDR KYRECMYECV VRAVCRQEGK AYKPVEQPKL KPQEAPKKGL
YKVQVGAFGD KSNADRLAKE LEGKGYKTYI VQE
```

Chapter 4: PlyM17

- GenBank Accession Number: HM011605
- 771 base pairs; 256 amino acids; 28.98 kDa; Theoretical pI = 8.86
- N-acetylmuramoyl-L-alanine-amidase-Type 2 enzymatic domain (PF01510)
- No recognized binding domain

```
ATGTACCAAA TCACAAGGGA TTACATTAAA TTCGGTAATT CTCGATGCGG
GCAACAAATT AAGAAAGTGT TGTTTATCGT AAGCCATGAC ACAGGGAATC
CTGGTAGTAC AGCATACGGG AATCGAAATT ACTTTAATAA TCAACAACCT
AGTGCGTCAG CCCACACGTT CATTGATGAT AAATACATTT TGGAGATCAT
CCCTATCTAT GAGAAAGCAT GGCACGTACA GTATCAAAAA CCAAAGATA
ATCAAATGTT TGGTGATGAT GCGAATGATG CAGCTATCGG GGTGAACTT
TGTTGGGGGA ATGGTATCAA CTTCAATGAA GCATACAAGC GTTTTGTGTG
GTATCATGCT CATCTTTGTA AGACATTTAA ACTCAATCCT AGAAAGCACA
TCGTATCTCA TAAGACATTG GATCCAGAAC GTAAAGTAGA CCCGATAGAC
TGCTTTAAAC GTCATGGTAT TACATGGGAA CAATTCATAA ATGATGTAGA
AAATGTGTTT GTAGGGAAAA AGGAGGAAAA TGATATGTTA GAAAAAGCGA
TTCTAATAGG TGGATTCCCT GATTTTGCAG TAGCGGAGAT ATTAGCAGCA
CGATTAAAAG CACCTATCTA CACTCGTGAT GCTTATCCGG GTGGAAAGGT
TACAAAAGAA TTATTTGTAG TCGGTGGTTC TACCGCAGGC CTACAAGCAG
ACAAAGTAAT CAATCTATCC GAAAAGACC GATTTGCTGT TGCTGCAGCA
GTTAAAAAGT TTATCGGATA A
```

```
MYQITRDYIK FGNSRCGQQI KKVLFIVSHD TGNPGSTAYG NRNYFNNQQP
SASAHTFIDD KYILEIPIY EKAWHVQYQK PKDNQMFQDD ANDAAIGVEL
CWGNGINFNE AYKRFVWYHA HLCKTFKLNP RKHIVSHKTL DPERKVDPID
CFKRHGITWE QFINDVENVF VGKKEENDML EKAILIGGFP DFAVAEILAA
RLKAPIYTRD AYPGGKVTKE LFVVGGSTAG LQADKVINLS GKDRFAVAAA
VKKFIG
```

Chapter 4: PlyM18

- GenBank Accession Number: HM011606
- 762 base pairs; 253 amino acids; 28.71 kDa; Theoretical pI = 7.77
- N-acetylmuramoyl-L-alanine-amidase-Type 2 enzymatic domain (PF01510)
- LysM binding domain (PF01476)

```
ATGTTGACGG TCCAGCAGAT CATCCGGCAG GGTGGCAAGA GATTGATTAG
CCACCCCAGC GACTATCAGG TCGCTTTTTT AATGGAGATT TTTATGGCAT
ACGAATTGAG AGACAATAGA CGCTCTACCT ACCTGGCTAA AGGCAGAGCC
AATACAGCGA TTGACGTCAT TGTTATCCAT CACTGGGGTG TTGACGGTCA
AAATTGGGAG AACTTAACCA CCTACACGGC CAATAACCGC AATATGAGTA
CCCATTACGT TGCCATGGCT GGCAAGGTCG AGCGACAGGT GGACGAGGAG
GATACGGCTT ACCATGGTGG CAATCCACCG ATCAATCAAC GCTCTATCGG
CATTGAGTGC CGTCCAGAAG CCACAGACGG CGATTACGAC ACAGTGGCTG
AGCTGGTGGC TGATATCTGG AACAGGCACG GCAAGCTACC CTTGGTCGGC
CATAAGCAAG TGCCATCCGT CAGGCCGGGT CAGCAATACG TGGCGACAAG
CTGCCCCGGC AGATACGATG TGGAGCGGAT CCGTAAAGAA GCGGAGGGTT
GGTACACGAA GAAGTACGCC AAGAAGGACA GCGACGCACG ACCTAAGACG
CACACGGTCG TGAGTGGCGA ATCCTACTGG AGTATTGCAG CCAAGTACCT
AGGCGATGGC ATGAGGTACA CAGAAATATT AGACCTGAAC AAGGTCAAAG
ATCCGAGTGC CTTATTAGTC GGACAGGTCT TGGAACTGCC TGAAAAGTAC
GCTAAGCACT GA
```

```
MLTVQQIIRQ G GKRLISHPS DYQVAFLMEI FMAYELRDNR RSTYLAKGRA
NTAIDVIVIH H HWGVDGQNWE NLTTYTANNR NMSTHYVAMA GKVERQVDEE
DTAYHGGNPP I INQRSIGIEC RPEATDGDYD TVAELVADIW NRHGKLPLVG
HKQVPSVRPG Q QYVATSCPG RYDVERIRKE AEGWYTKKYA KKDSARPKT
HTVVSAGESYW S SIAAKYLGDG MRYTEILDIN KVKDPSALLV GQVLELPEKY AKH
```

Chapter 4: PlyM19

- GenBank Accession Number: HM011607
- 819 base pairs; 272 amino acids; 29.1 kDa; Theoretical pI = 7.05
- N-acetylmuramoyl-L-alanine-amidase-Type 3 enzymatic domain (PF01520)
- SH3-Type 5 binding domain (PF08460)

```

ATGGGTACAT ATAACGTTCA TGGTGGTCAC AACTCTATCG TACAAGGTGC
TAACTGGGGT GCTCGTAAAG AACACGTTAT GGATCGCCAA GTAAAAGATG
CTTTAATTAG CAGGCTACGT AGCCTTGGTC ACACAGTTTA TGACTGCACA
GACGAAACAG GTTCTACGCA ATCTGCTAAC TTACGTAACA TCGTAGCTAA
ATGTAATGCT CACCGAGTAG ACTTAGGCAT TTCGTTACAC TTAAACGCAT
ACAATGGTTC TGCTAACGGA GTAGAGGTTT GCTATTCCGA CCAACAAGCG
TTAGCTGCTA AAGTTTCTAA GCAATTGTCT GACGACATCG GATGGTCTAA
CCGAGGAGCT AAACCTCGTA CAGACCTTTA CGTACTAAAC AGCACATCTG
CACCTGCTAT CCTAATCGAG TTAGGTTTCA TCGATAACGA GGGCGATATG
GGTAAATGGA ACGTGGATAA AATCGCAGAC TCCATCTGCT ATGCTATCAC
GGGTCAACGT GCAGGTTCTA CAGGTGAAA CACAGGTGGA GGTTCTACAG
GCGGAAGCAC TGGTGGTGGA TACGACTCTA GTTGGTTCAC TCCACAAAAT
GGTGTATTCA CAGCTAACAC AGCAATCAAA GTTAGAAGCG AACCAAGTGT
AAATGCAACT CATCTTCGTA CTCTGTACAG TGGCGGCACA TACAAGTACA
CTTCATTTCG AATGGAGAAA GACGGTTACG TTTGGATCAA GGGAGCAGAC
GGTACATACG TTGCAACAGG GGAAACTCGT GACGGTAAAC GTATCTCTTA
CTGGGGTTCT TTCGAGTAA

```

```

MGTYNVHGGH NSIVQGANWG ARKEHVMDRQ VKDALISRLR SLGHTVYDCT
DETGSTQSAN LRNIVAKCNA HRVDLGISLH LNAYNGSANG VEV CYSDQQA
LAAKVSKQLS DDIGWSNRGA KPRTDLYVLN STSAPAILIE LGFIDNEGDM
GKWNVDKIAD SICYAITGQR AGSTGGNTGG GSTGGSTGGG YDSSWFTPQN
GVFTANTAIK VRSEPSVNAT HLRTLYSGGT YKYTSFGMEK DGYVWIKGAD
GTYVATGETR DGKRISYWGS FE

```

Chapter 4: PlyM20

- GenBank Accession Number: HM011608
- 909 base pairs; 302 amino acids; 33.34 kDa; Theoretical pI = 8.81
- N-acetylmuramoyl-L-alanine-amidase-Type 2 enzymatic domain (PF01510)
- SPOR binding domain (PF05036)

```

ATGGCTTATA CGAACAGTCC TTTGGTGGAC TACACGAAAC TGAGTCCCAA
TCATTCGGGA CAGAGAAACC ACGCTATTGA CACCATCACA ATCCATTGTG
TAGTAGGTCA ATGCTCGGTA GAGACTCTTG GAGAGATTTT CGCACCTACT
TCGAGACAGG CATCTTCCAA CTACGGCGTA GCGGTGGATG GTCGAATCGG
AATGTATGTT GAAGAGAAGA ATCGTTCCTG GTGTACTTCT TCTGCATCCA
ACGACCACAG AGCAATCACC ATTGAGGTTG CAAGTGATAC CACCGAACCT
TATGCGGTAA ACGACAAAGC GTATGCCGCA ATGCTCGACC TTGTAACCGA
TATTTGCAAG AGAAATGGTA TCAAGAAGTT GGTATGGTCT ACCAACAAAT
CCGACAGAGT GAATCACAAA AACGGATGCA ATATGACCGT TCACAGAGAC
TATGCGAATA AGTCCTGTCC CGGCAAGTAT CTGTATGACC GCCATCAGCA
GATTGCAGAT GAGGTCAACA AGAGATTGAA TGGTGCAAAT GAGGAGGGAA
GCAAGGTGCT ATACAAGGTA CAAATCGGTG CTTACTCTAA GGTAGAGAAC
GCCGAAAAGC AGTTAGAGAA AGCAAAGGCA GCAGGTTTCA CCGATGCTTT
TATCGTTAAG ACTGTCGTAG AAAATGAACC TGCGAAGCAA GAACCCCGA
AAGAACCCGA AGTGGTTTTA GCCGTTGGCG ATAAGGTCAA GATGGCTAAG
GATGCTCCTG TTTATGGAAA GACCACTAAG TTTCAGTCTT GGGTTTATGA
CTCTGTGCTT TATGTTCTGT AAATCAACGG CTCTCGTGTG GTTATTTCCA
CCTTGAAGAC AGGTGCAGTA ACGGGTGCAG TTGACAAGAA ATATCTTACC
AAAGTATAA

```

```

MAYTNSPLVD YTKLSPNHSG QRNHAIDTIT IHCVVGQCSV ETLGEIFAPT
SRQASSNYGV GVDGRIGMYV EEKNRSWCTS SASNDHRAIT IEVASDTTEP
YAVNDKAYAA MLDLVTDICK RNGIKKLWVS TNKSDRVNHK NGCNMTVHRD
YANKSCPGKY LYDRHQIAD EVNKRLNGAN EEGSKVLYKV QIGAYSKVEN
AEQLEKAKA AGFTDAFIVK TVVENEPKQ EPPKEPEVVL AVGDKVKMAK
DAPVYGKTTK FQSWVYDSVL YVREINGSRV VISTLKTGAV TGAVDKKYLT KV

```

Chapter 4: PlyM21

•GenBank Accession Number: HM011609

•1092 base pairs; 363 amino acids; 41.07 kDa; Theoretical pI = 9.26

•N-terminal M23 peptidase enzymatic domain (PF01551)

Central phage lysozyme (i.e. muramidase) enzymatic domain (PF00959)

•No recognized binding domain

```

ATGATTTTTA AACCGCCAGT ACAGAACATG AAGCTAACTA GCGAATACGA
TCCTTTTAAGG TTGCATCCAG TATTGAAGAT TGTACGCAA CATGCAGGTG
TTGATTTAAT TAACACCAAA TTAGGAAAAG CTCCGATATT TGCAACTGCT
AATGGTAAAG TACGCTTAGT TAAAACAACA GTAGATGGTT ATGGTAAGCA
CGTCATCATA ACACATAAAG TAGGTGGACA AGTGTACGAG AGTGTATACG
CACATTTAGA TTCATACGAG GTAAAAGTTG GTCAAGAAGT TTTACAAGGC
CAGAAAATAG GCGTAATGGG TAATACAGGT ATAGGAACGG GAATCCACTT
GCATTTTGAG TTACATAGGA ATCAGTGGGA ATTGATAAC TATAATTATC
CTAATAGCTT CAATCCATGG CCTTGGATAA ACGAAGGGGA GTTTAAATG
AAAGTATCTA AAGAAGGATT GGATTTAATT AAGTTTACG AAGGTTTTTA
CGATAAACT TACTTGGATC CTATTGGACT TCCTACTATT GGCTACGGTA
CAACTAAGTG GCCTAATGGT AATTCTGTAA AAATGGGTGA AAAGATTAGT
AAGGTTGAAG CTGATATTTT ACTAGAGCAA CAGGTTAATG AACACGCTAA
GACAATATTC AATTACGTTA AGGTTGATTT AACTCAAAAT CAATTTCGATT
CTTTAGCTTC TTTCCAATAC AACTTAGGTA GTGGAATACT TAAAAAGGAC
CCATCTATTG CTGCATATAT TAATAAAAAG GATTGGGCTA ACGCTACTCG
CGTTATGAAG CTATACAATA AGGCTGGTGG TAAAGTATTA GCTGGATTGG
ATAAACGTAG GATTGCTGAA GCTGAATTGT TTATGAAGCA ATCTGATAAG
AATAAGGAGG ATGACGATAC ATTGAAATTC ACTAACGAAG CCACAAAGAG
CATTGTTAAG CAATTCTTAG AACAATCAAT TAAAGCGGGT TACTTCGATA
AGTCATGGTT AACTAAGTTT AATGAAGGAA CTATTACTAA TGGAGACATT
ACAGGTTTAC AAATGATTGC TAACAACAAA AAGAATAAAT AA

```

```

MIFKPPVQNM KLTSEYDPLR LHPVLKIVRK HAGVDLINTK LGKAPIFATA
NGKVRLVKTT VDGYGKHVII THKVGQVYE SVYAHLDSEY VKVGQEVLLQG
QKIGVMGNTG IGTGIHLHFE LHRNQWFDN YNYPNSFNPW PWINEGEFKM
KVSKEGLDLI KFYEGFYDKT YLDPIGLPTI GYGTTKWPNG NSVKMGEEKIS
KVEADILLEQ QVNEHAKTIF NYVKVDLTQN QFDSLASFQY NLGSGILKKD
PSIAAYINKK DWANATRVMK LYNKAGGKVL AGLDKRRIAE AELFMKQSDK
NKEDDDTLKF TNEATKSIVK QFLEQSIKAG YFDKSWLTKF NEGITITNGDI
TGLQMIANNK KNK

```

Chapter 4: PlyM22

- GenBank Accession Number: HM011610
- 480 base pairs; 159 amino acids; 17.38 kDa; Theoretical pI = 8.49
- Phage lysozyme (i.e. muramidase) enzymatic domain (PF00959)
- No recognized binding domain

```
ATGGCGCGCG CAAAAGGCTT GAGTCGCAAC ATGGCTGCAT TCTTGGATAT
GATTGGTTGG TCTGAGATTG GCGATAAATT GTTGGCAGTG TCCGACGATG
GGTATAACGT TATCGTTGGC TCCACCCCCA CCAAACCAA GCTGTTTGAC
GACTACGCCG CTCATCCGCA AATTTATGTG AAGTCGGTTA ACTCGACCGC
TGCGGGCAGA TATCAGATTC TCGGTAAGTA CGCAACTCAT TACATGGCCC
AACTCAAGTT GCCTGATTTT GCCCCTGCAT CACAAGACAA AATTGCAATT
CAATTGATCC GCGAATGCAA GCGGGTCCAG TTGATTGAAG ATGGCCACAT
TGGTCGAGCA ATCACTGCTT GCAAAAGCCG CTGGGCGAGT TTTGAGGGTG
CCGGGTACGG CCAACGTGAG CACTCTATTG ATGATTTGAT TTCCCATTTT
ATTGCGTATG GTGGAGTTCT TGCAAAATGA
```

```
MARAKGLSRN MAAFLDMIGW SEIGDKLLAV SDDGYNVIVG STPTKPKLFD
DYAAHPQIYV KSVNSTAAGR YQILGKYATH YMAQLKLPDF APASQDKIAI
QLIRECKAVQ LIEDGHIGRA ITACKSRWAS FEGAGYGQRE HSIDDLISHF
IAYGGVLAK
```


Chapter 4: PlyM23

- GenBank Accession Number: HM011611
- 450 base pairs; 149 amino acids; 16.39 kDa; Theoretical pI = 9.38
- Phage lysozyme (i.e. muramidase) enzymatic domain (PF00959)
- No recognized binding domain

```
ATGCAAGTTA GTGATGCAGG TATTGAGTTA ATCAAGTCAT TCGAGGGATT
CCGCGCTAAC GCATATCCCG ACCCGAAATC AGGTGGCGAC CCGTGGACCG
TTGGCTACGG CACCACCAA TTCCCATCCG GCAGACCAGT CAAGCAAGGC
GATAAGGTAA CGCCTGGCCA GGCTGAGCTA TATCTTCGCG AGGATGTGAA
GAAGTTTGCA AACTCAGTTG ACGCTCTTGT CACCGCCCCCT CTGAAGCAAT
GCCAGTATGA TGCCTGGTG TCATTCGTTT ACAACCTTGG CGCAACCAAT
TTCCGCACGT CCACCCTACT GAAGAAATTG AACGCCAAAG ACTACAATGG
TGCTGCTGAT GAATTTCTCC GCTGGGTATC TCCCGGCTCA TCCGTTGAGG
CTGGATTACG CCGCCGTCGC ACGGCAGAGC GCGCTATGTT CCTTTCGTAA
```

```
MQVSDAGIEL IKSFEGRFRAN AYPDPKSGGD PWTVGYGTTK FPSGRPVKQG
DKVTPGQAEI YLREDVKKFA NSVDALVTAP LKQCQYDALV SFVYNLGATN
FRTSTLLKKL NAKDYNGAAD EFLRWVSPGS SVEAGLRRRR TAERAMFLS
```

Chapter 4: PlyM24

- GenBank Accession Number: HM011612
- 546 base pairs; 181 amino acids; 20.15 kDa; Theoretical pI = 9.59
- Phage lysozyme (i.e. muramidase) enzymatic domain (PF00959)
- No recognized binding domain

```
ATGAAACTCA TATTCGACGA GTTCCGAAAA CTGGCAGGCG GTAAATTAAC
GCAAGCGCAA GTGGATAAAA TCAATGCTCT GATTGATGAA ATCCAGGTTA
AGTCTATGAA GGTTCGATGCG GCAGGGATTG ACTTGATTGG TCAGTTTGAA
GGGCTGCGCC TAAATTCCTA TGATGATGGC GTTGGTGTGT GGACTATCGG
ATGGGGAACA ACGGTTTACC CAAACGGTCA AAAGGTCAAG AAAGGCGATA
AAATCACACT AGAGCAGGCG AAGCAATACA AGGCGCATGA TTTAGCTAAG
TTTGAAAAAG CTGTCAATGA TGCGGTTAAA GTGCCGCTGA ATCAAAACCA
GTTTAACGCT TTGGTGTCAC TGGCCTATAA TATTGGGGTA TCTGCTTTCT
CTAACAGCAC ATTAGTCAAG CGGTAAACG AAGGCAACTA TAAGGCTGCT
GCTGATCAGT TTTTAGTGTG GGTCAATGCA GGCGGTAAAC GTATGCAGGG
CTTAGTCAAT CGACGTAATA AAGAACGGGA GTTATTTTTA AAATGA
```

```
MKLIFDEFRK LAGGKLTQAQ VDKINALIDE IQVKSMKVDA AGIDLIGQFE
GLRLNSYDDG VGVWTIGWGT TVYPNGQKVK KGDKITLEQA KQYKAHDLAK
FEKAVNDAVK VPLNQNQFNA LVSLAYNIGV SAFSNSTLVK RLNEGNYKAA
ADQFLVWVNA GGKRMQGLVN RRNKEERELFL K
```

Chapter 4: PlyM25

•GenBank Accession Number: HM011613

•771 base pairs; 256 amino acids; 28.17 kDa; Theoretical pI = 8.76

•N-terminal PG-1 binding domain (PF01471)

•Domain of unknown function DUF3380 (PF11860), likely enzymatic domain

**NOTE: The above binding domain shows only low homology to the consensus sequences for PG-1, with an E-value score of 0.0007.

```
ATGGCGAAAA CACACATAGT ACAGGCACAA ACAGCACTGG CGGCGGCAGG
TTATTATAAC GGTAAGATTG ACGGCGATTT TGGCGGCGGT TCATTGCGAG
CAGTGCAAAA TTTAATCGAC AATACAGATA AGGCGGTTGA TATTTTGAGC
AAGCCTGATA TTGCTGAGCC AGTAGATAAA AAACATCTA AAGCTGATAT
TATCGCAGTG GCTAATGATT TAAACATTGA GCCAGCAGCA CTTAAAGCAG
TCATTGATGT TGAGGCAGCC GGCAATGGTT TTGATAATCA AGGTCGTCCG
ACAATACTTT TTGAGCCTCA TGTGTTTTTG GATGAATTAG GTAAAATACA
TTACTACACT AAACGCGCCG AACTAGCTAA AAAGCATAAC GGGTTGTTGT
CACCAAAGTG GGATAGGTCG CTATATCGCA TTGGTGGTAG CTCGCATGAT
AAGTTAAAGA TTGCTGCTGA TCTACACTGG GAGGCGGCGC ACAAATCTGC
ATCATGGGGT TTGGGGCAAA TCATGGGGTT TAACGCACAA AAGATAGGCT
ATGCAACGCT TAAAGAGTTT ATTGATGATA TGTATGAGAG CGAAGCTAAG
CAACTTAAAG CGATGGGAAT GTTTTTTAAA GCTAACGGCT TAATCAGTAA
GCTACAGCGC CATGATTGGG CGGGGTTTGC ACGAGGTTAT AACGGCTCGG
CTTATGCCAA AAATCAGTAT GATGTAAAAC TCGCTGATGC CTACAATCTG
GCTAAAAAAC AAGGTTGGTA G
```

```
MAKTHIVQAQ TALAAAGYYN GKIDGDFGGG SLRAVQNLID NTDKAVDILS
KPDIAEPVDK KLSKADI IAV ANDLNIEPAA LKAVIDVEAA GNGFDNQGRP
TILFEPHFVW DELGKIHYYT KRAELAKKHN GLLSPKWDRS LYRIGGSSHD
KLKIAADLHW EAAHKSASWG LGQIMGFNAQ KIGYATLKEF IDDMYESEAK
QLKAMGMFLK ANGLISKLQR HDWAGFARGY NGSAYAKNQY DVKLADAYNL
AKKQGW
```

Chapter 4: PlyM26

- GenBank Accession Number: HM011614
- 567 base pairs; 188 amino acids; 21.16 kDa; Theoretical pI = 9.56
- Glycosyl Hydrolase Type 19 (i.e. chitinase class I) enzymatic domain (PF00182),
not likely a dedicated chitinase
- No recognized binding domain

```
ATGATCCTGA CTGCCGGTGG TTTTAATATT CTTCGCAATG GCTTGGGTCG
ATTAAACCAG AGCCAGGTTA ATGAGATTAA TTTTCTAGTA TCCCAATTCTG
ACAAAGACAA ATCTATTTCC TACCCACAAG CTGCCTATAT GTTGGCGACA
ACATGGCATG AAAGTGCAGC AACCATGCTG CCAATTGAGG AATATGGAAA
GGGCAAAGGT CGCACCTATG GCAAAAAAAT TGATAGTAAT GGCTCTGCCT
ATACTGGCTT GGATCATATT TATTATGGTC GTGGTTATGT TCAGCTCACC
TGGCTTAGCA ACTACGTTCT AGCCAAAAG AACTTGGCA TTGATTTTGT
GAACAAGCCT GAATTGGCTC TGGTTCCTGA GCATGCAGTC AAGATTCTGA
TTACCGGAAT GAAGGAAGGT TGGTTTACCG GCAAGAACT TTCCGACTAT
ATCCATCAAT CCAAAAAGGA CTACATTAAC GCTCGACGTA TTATCAATGG
AACTGATAAA GCAAACTTA TCGCCGGCTA TGCCGAAATA TTTGAACGTG
CTTTGAGGTC CTTATAA
```

```
MILTAGGFNI LRNGLGRLNQ SQVNEINFLV SQFDKDKSIS YPQAAYMLAT
TWHETAATML PIEEYKGKKG RTYGKKIDSN GSAYTGLDHI YYGRGYVQLT
WLSNYVLAKK KLGIDFVNKP ELALVPEHAV KILITGMKEG WFTGKKLSDY
IHQSKKDYIN ARRIINGTDK AKLIAGYAEI FERALRSL
```

Chapter 4: PlyM27 (Not actively expressed)

- GenBank Accession Number: HM011615
- 534 base pairs; 177 amino acids; 19.93 kDa; Theoretical pI = 9.7
- N-acetylmuramoyl-L-alanine-amidase-Type 2 enzymatic domain (PF01510)
- SH3-Type 3 binding domain (PF08239)

```
ATGCAGAGCA ACAATAAGCA GACATCATTC CATTACGCTG TTGATGACAA
GGAAGTCATC CAAGGTATCC CCCTTGATAG GAATGCCTGG CATGCGGGGG
ATGGCAGAAA CGGAAAAGGG AATAGAGAGG GAATAAGTAT CGAGATCTGT
TACTCCAAGA GTGGCGGGTC TCGTTTTATT TTGGCTGAAA AGAGGGCTGC
AAAGCTTATC GCTGATATTC TGGACGAGAA AGGATGGGAT ATCTCAAAAG
TAACTAAGCA TAGGGATTAT TCAGGCAAGT ATTGCCCACA TCGTACACTG
GATATGGGAT GGGACCGATT CCTTAAGCTG GTTGAATCAG AACGAAAGAC
ATCTTTTAAG GTTAGAGTAA CTGCAGACTC CCTGAACTAC AGGGCAGGAG
CTGGGATCAA ATACAAGATA AATGGAGCGA TTACGGATAA GGGAGTATAT
ACAATCGTTG ATCAGACTAA TGGCTGGGGC AAACCTAAGA GCGGGGCAGG
CTGGATAAAT TTGAAATATA CGAAAAGAGT ATAA
```

```
MQSNNKQTSF HYAVDDKEVI QGIPLDRNAW HAGDGRNGKG NREGISIEIC
YSKSGGSRFI LAEKRAAKLI ADILDEKGWD ISKVTKHRDY SGKYCPHRTL
DMGWDRFLKL VESERKTSFK VRVTADSLNY RAGAGIKYKI NGAITDKGVY
TIVDQTNGWG KLKSGAGWIN LKYTKRV
```

Chapter 4: PlyM28 (Not actively expressed)

- GenBank Accession Number: HM011616
- 711 base pairs; 236 amino acids; 26.42 kDa; Theoretical pI = 9.62
- N-acetylmuramoyl-L-alanine-amidase-Type 3 enzymatic domain (PF01520)
- SPOR binding domain (PF05036)

```
ATGAAAATTA TATTAGACGC AGGTCACGGA CCAAACACGC CAGGCAAACG
GAGTCCTGAC GGAATGCGCG AATTTTCAGTT TAATAGTCGA GTTGCTGACG
TCATGAAAGC GGAGCTTGAG AAATACGAAG GCGTGACGGT CTACTTTGCG
CACGATCCCA AGCGCGACGT TCCGTTGAAG GAGCGGACTG ACAAGGCGAA
TAAGCTTTGC GTCGACCTAT TCTTCAGTAT TCACGCGAAC GCTAACACCG
GCAAGATGGG CGATTGGGGC GGAATCGATT CGTTCGTCTA CACGTCTAAT
CCGAAGGAAG CGCGCAAGCT GGCCGACATT GTGCAACGCA ATCTGATTGC
GGCCACTAAA CTGCGTAATC GCGGCGTCAA GACGGCCGAC TTCCTCGTAC
TGCGCGAAAC ACATATGACA GCAATCCTCG TCGAGCACGG ATTTATGGAT
TCGAAGACCG ACTTGCCGTA CTTGAAATCG GACGCCTATC GCAAGCTTTG
CGCAGTAACT AACGTAAAGT CAATCGCGCA AATGTACGGA TTGAAGCCAA
AGAAGGTTGC GCCAAAGCCT GCGGTCAAAC CTGCCGTCAG CTCTGACGTA
TTTTATCGCG TTGTGACCGG CTCATTGGGC GACAGAGAAA ACGCCGAAAA
GCGTATGGCT CAGTTGAAGA AAGCCGATT CGAATCGTTC ATTGACGCTT
ATAAAAAATA A
```

```
MKIILDAGHG PNTPGKRSPD GMREFQFNSR VADVMKAELE KYEGVTVYFA
HDPKRDVPLK ERTDKANKLC VDLFFSIHAN ANTGKMGDWG GIDSFVYTSN
PKEARKLADI VQRNLIAATK LRNRGVKTAD FLVLRETHMT AILVEHGFMD
SKTDLPYLKS DAYRKLCAVT NVKSIAQMYG LKPKKVAPKP AVKPAVSSDV
FYRVVTGSFG DRENAEKRMA QLKKAGFESF IDAYKK
```

Chapter 4: PlyM29 (Not actively expressed)

- GenBank Accession Number: HM011617
- 708 base pairs; 235 amino acids; 26.33 kDa; Theoretical pI = 9.34
- N-acetylmuramoyl-L-alanine-amidase-Type 3 enzymatic domain (PF01520)
- SPOR binding domain (PF05036)

```

ATGAAAATAT TAATCGATGC CGGCCACGGA CCAAACACGC CAGGCAAACG
GAGTCCTGAC GGAATGCGCG AATTTCGAATT TAATAGCCGA GCTGCTGACG
TCATGAAGGC GGAGCTTGAA GAATACGAGG GTGTGACGGT CTACTTTGCG
CATGAGCCTA AGCGCGATGT GCCGTTGAAA GAGCGGACTG ACAACGCGAA
CAAACCTTTGC GTCGACCTAT TCTTCAGTAT TCACGCGAAC GCTAACACCG
GCAAGATGGG CGACTGGGGC GGAATTGATT CGTTCGTCTA TACGTCCAAT
CCGAAGGAAG CACGGAAGCT GGCCGACATT GTGCAGCGCA ATCTAATCGC
GGCTACTAAA CTTCGTAATC GCGGCGTCAA GACGGCCGAC TTTCACGTAC
TGCGCGAAAC GCACATGACG GCAATCCTCG TCGAGCACGG CTTTATGGAT
TCGAAGACAG ACTTGCCGTA CTTGAAGTCG GATGCCTATC GCAAGCTTTG
CGCTGAGTCT AACGTAAAGT CAATCGCGCA AATGTACGGA CTCAAGCGCA
AGCCTGCGCC GAAGCCCGTG GTCAAGCCAG CCGTCAGCTC CGACGTATTC
TATCGCGTTG TGACCGGTTT ATTTGGCGAC AGGGATAACG CCGGACGACG
CCTGGCTGAG CTGAAAAAGG CCGGCTTTGA TTCGTTTCATT GACGTTTATA
AAAAGTAA

```

```

MKILIDAGHG PNTPGKRSPD GMREFEFNSR AADVMKAELE EYEGVTYVFA
HEPKRDVPLK ERTDNANKLC VDLFFSIHAN ANTGKMGDWG GIDSFVYTSN
PKEARKLADI VQRNLIAATK LRNRGVKTAD FHVLRETHMT AILVEHGFM
SKTDLPYLKS DAYRKLCAES NVKSIAQMYG LKRKPAPKPV VKPAVSSDVF
YRVVTGSFGD RDNAGRRLAE LKKGAFDSFI DVIKK

```

Chapter 4: PlyM30 (Not actively expressed)

- GenBank Accession Number: HM011618
- 1017 base pairs; 338 amino acids; 36.89 kDa; Theoretical pI = 9.61
- N-acetylmuramoyl-L-alanine-amidase-Type 2 enzymatic domain (PF01510)
- PG-1 binding domain (PF01471)

```

ATGCGCTTAC TAAAAAATGG CACAAAAATA GGAAACGCCA CAGTAATTGT
TGATATTGTG GACGGCGCAA ACAGAGAGAT TCGACCGCGC AAAAAAATGA
ATCCGCGTTT CTTGACGGAC CACGACACGG GTAAC TCGGG TAATGGCGCA
GACGCGAAAA GTCATAATCG CTACATCCAC AACTTGGGCG ATAAGCTACC
ACGCGACACA AGCCATATTT CCTGGCATGT AACTGTGGAT GAAAAC TTCA
TCATTCAACA TATCCCATTT GACGAGTGTG CTTATCATTG CGGCGACGGT
TGGGCGACGA CTAGCGGGAA TCGCACAAGT ATCGGCATAG AGAAGTGCAT
GCACAGAGGC GCGGATCGAA ATAAGATTGA GGCCAACGCG ATTGCGCTTT
ATGCGTATCT AATGAAGGAA CTCAAAATAC CGATTACATC TGTACGCCCT
CATCAGCATT GGAGCGGCAA ATACTGTCCA CAGTTGATAT TGAATCGGTA
CGGTTTCGTTT ACGCCATTCC GCAATAAGAT TGAAGCGGCG TTTAAAGGTG
GCGCAGTAAC GGCAGCGAAG CCTTCCACAG TGAAGAGAGA CTACCTATTA
GATGGCGATA CAGGCGCGGC GGTAAAACG CTACAAAGCG AATTGAAACA
AGCAGGTTTC TTACTATCTG TGGATGGCAT ATTTCGAAAA GGTACAGAGA
CAGCGGTTAA GCGGTTTCAA CGTGCGAATG GATTAGCAGT TGATGGCGTG
TTCGGAACGG GCTCACAGGC TAAACTCAAC GCAATACTGG CGAATCTGAA
TAAGAAGCCG GTAGTTAAGC CGGCAGCGCC AACAAAACCA AAGGAGGAAT
CCACAGTGGA GAAAACAAAA CAACCCTCAA AATGGGCAGA AGTGACGATT
AAAGAAGCGG TGAAGATTGG TGTATCGGAC GGATCTCGCT TGCATGATAC
AGTGACTCGT GAAGAATCAA TTGTAATCGC CATGCGCGCT GCAGGACTTG
CGCCGAGACT TAAATAA

```

```

MRLKNGTKI GNATVIIVDIV DGANREIRPR TKMNPRLTD HDTGNSGNGA
DAKSHNRYIH NLGDKLPRDT SHISWHVTVD ENFIIQHIF DECAHYHCGDG
WATTSGNRTS IGIEKCMHRG ADRNKIEANA IALYAYLMKE LKIPITSVRP
HQHWSGKYCP QLILNRYGSF TPFNRKIEAA FKGGAVTAAK PSTVKRDYLL
DGD TGA AVKT LQSELKQAGF LLSVDGIFGK GTETAVKAFQ RANGLAVDGV
FGTGSQAKLN AILANLNKKP VVKPAAPT KP KEESTVEKTK QPSKWAEVTI
KEAVKIGVSD GSRLHDTVTR EESIVIAMRA AGLAPRLK

```


Chapter 4: PlyM31 (Not actively expressed)

- GenBank Accession Number: HM011619
- 1074 base pairs; 357 amino acids; 38.61 kDa; Theoretical pI = 9.57
- N-acetylmuramoyl-L-alanine-amidase-Type 3 enzymatic domain (PF01520)
- PG-1 binding domain (PF01471)

```

ATGGCAAAGA TTATGAATGA GCCGGGGCAC GGCAACAACA CATGGCCGCC
AAGCAAAGGG GTGCGCGCAA GTGGTGGCGT TCCCGAAATG GCGGAGCATG
ATTTTAACGC AGCGGTAGCA GATGAAGTGA ATCGCCTATT AAGCGGCAAG
CTAACGACTT ATAGCGCACA GCCTAGTCGC GGCAAGGATG TTTCGCTAAC
GACGCGCACG CGCCTATATA ATGAAGAGTT TCGCAAAGAT AAGAGCGCAA
TTGGCTTCTC ACATCACGGA AACGCCAACG CTAACAAAGC GACAAAAGGT
TTCGGTGTAT TCTACTGGGG CGGATCTGCA ACGGGTAAGA AGTTGGCGCA
AATGCTTCTC GCGGCGTATA AAAAAGAATT CCCTGGCTAT CCGATTTGGG
GCAGCGGGAT ATTCGAAAGT AAACGCGGAG ACTGGACAAA CTTTGCAATC
CTGCGCGACA CGTCAGCGCC TTTCGTGTTA ATCGAGTGGG ATTTCTTTAC
GAATGATGAA GCGCGTAAAC GGATGCTATC TACAGATTAT AGAAAGCGTT
GCGGTAAAGT GGCAGCGAGT GTCGCGTGTG ATTGGTACGG TATTCGGTTT
ACTGATTTC AAGCAGCTAA ACCTGCGCCT TCCACAGTGA AGAGAGACTA
CCTATTAGAT GGCGATACTG GCGCAGCGGT TAAAACGCTA CAAAGCGAAT
TGAAACAAGC GGGTTTCTTA CTATCTGTGG ATGGCATATT CGGAAAAGGC
ACAGAGACAG CGGTTAAGGC GTTCCAACGT GCGAATGGAT TAGCAGTTGA
TGGCGTGTT CAAACAGGCT CACAGGCTAA GCTCAACGCA ATACTGGCTA
ATTTGAATAA GAAGCCGGTA GTTAAGCCTG CAGCGCCATC AAAACCAAAG
GAGGAGTCCA CAGTGGAGAA AACAAACCAA CCCTCAAAAT GGGCAGAAGC
GACGATTAAA GAAGCGGTGA AAATTGGTGT TACGGATGGC AGCAACTTAC
ACGATCCAAT TACACGACAG GAAGCAATTG TGCTGGCGAT GCGTGCTGCA
GGGCTTGCGC CAAAACTTAA GTAA

```

```

MAKIMNEPGH GNNTWPPSKG VRASGGVPEM AEHDFNAAVA DEVNRLLSGK
LTTYSAQPSR GKDVSLTTRT RLYNEEFRKD KSAIGFSHHG NANANKATKG
FGVFWGGS TGKKLAQMLL AAYKKEFPY PIWGSIFES KRGDWTNFAI
LRDTSAPFVL IEWDFFTND ARKRMLSTDY RKRCCKVAAS VACDWYGIFP
TDFTAAPAP STVKRDYLLD GDTGAAVKTL QSELKQAGFL LSVDGIFGKG
TETAVKAFQR ANGLAVDGVF GTGSQAKLNA ILANLNKKPV VKPAAPSKPK
EESTVEKTNQ PSKWAEATIK EAVKIGVTDG SNLHDPITRQ EAIVLAMRAA
GLAPKLK

```

Chapter 4: PlyM32 (Not actively expressed)

- GenBank Accession Number: HM011620
- 423 base pairs; 140 amino acids; 15.78 kDa; Theoretical pI = 9.16
- N-acetylmuramoyl-L-alanine-amidase-Type 2 enzymatic domain (PF01510)
- No recognized binding domain

```
ATGAGAAAAA TCAAGGAAAT AATAGTCCAT TGCTCTGCAA CGAGAGAAGG
CAAGCCCTTC ACCGTTGCCG ATATTGAGCG TTGGCATCGG GAAAGAGGAT
GGAAGGGCTG TGGCTATCAC TACGTTATCA CCCTTGACGG AAAGGTGGAG
CAGGGCAGAC CAGAGCAGAT TGCTGGCGCA CATTGCTCTG GCAGAAACAG
CGAGAGCATA GGCATATGCT ACATCGGTGG ATGCGACCTC GTGGGCAAGA
GCAAGGACAC ACGCACACCA GAGCAGAAGA AGGCAATGGT AACACTCATT
CGTCAGCTCA TGCAGACACA CGGCATCAGC ATCAAGGATG TCCGCTGCCA
TAATGAGTTC AGCACAAGGG CTTGCCCCAG TTTCTCTGTC CAGAAGTTGC
ACAAGGAAAT ACTGGAAGGA TGA
```

```
MRKIKEIIVH CSATREGKPF TVADIERWHR ERGWKCGYH YVITLDGKVE
QGRPEQIAGA HCSGRNSES I GVCYIGGCDL VGKSKDTRTP EQKKAMVTLI
RQLMQTHGIS IKDVRCHNEF STRACPSFSV QKLHKEILEG
```

Chapter 5: PlySs1

- GenBank Accession Number: Not yet submitted
- Full-length construct: 1359 bp; 452 amino acids; 49.67 kDa; Theoretical pI = 6.87
- Truncated construct: 765 bp; 254 amino acids; 28.09 kDa; Theoretical pI = 7.7
- N-terminus: Pfam predicts an N-acetylmuramoyl-L-alanine-amidase-Type 5 enzymatic domain (PF05382); it was experimentally verified, however, that this region actually possesses γ -D-glutaminyll-L-lysine endopeptidase activity against *S. suis* peptidoglycan.
- Dual CPL-7 binding domain (PF08230)
- C-terminus: Endo-beta-N-acetylglucosaminidase enzymatic domain (PF01832)

```
ATGACAATCA ATCTTGAAAC ATCCATTCGT TGGATGAGCG ACCGTGTCGG
CAAAGTCTCT TACTCAATGG ACTATCGTAA CGGTCCGAAT AGTTATGACT
GCTCTAGTGC TGTATATTAT GCGCTAATGG CGGGTGGTGC AATTTCTGCA
GGTTGGGCGG TTAACACTGA GTATATGCAT GACTGGTTGA TACGTAACGG
ATATGTTTTG GTAGCTGAAA ATAAACCATT TAACGCTCAA AGACATGACG
TTTGTATTTT GGGTAAACGT GGCTATTCTGA GCGGAGCAGG TGGTCACGTC
GTTATCTTTG TGGATAATGT TAATGTGATA CATTGTAAC TATGCACGTAA
CGGAATTTCC ATTGATAATT ATAATCAAGT GCATCGTGGT ATGTATTACT
ATCTATATCG CCCAGCAAAT CAACCCAGCA TCAGCAACAA ATCACTGGAT
CAGCTTGTTA AGGAGACTTT GGCTGGGGTA CATGGCAACG GGGACACCCG
TAAGGCAAGT CTTGGCAGTC AATACGAGGC TGTCATGGCG GTTATCAATG
GCAAAGCTTC GGCAAGCGAG AAATCTGATG AGGAACTTGC TAGGGAAGTC
TTAGCAGGTA AGCACGGGGC TGGAGAGGAC CGAAAACGGT CATTAGGACC
ACGCTATGAG CCTGTTCAAG CCAAGGTCAA CGAATTGCTC AAGGCTAAGG
AAAAACCGTC TGAGACGGCC AAAAATGAAC CACAGACGGT GCAATTCAAG
GAGGACGGGG ACTTGCTCTT CAATGGTGCC ATTCTTAAGA AGTCTGTCTT
CGAAATTATC CTGAAAAAGT GTAAAGAACA TGACATCTTA CCAAGCTATG
CCCTAACTAT CCTACACTAT GAAGGGCTTT GGGGCACTTC TGCTGTCTGGT
AAGGCCGACA ACAACTGGGG CGGTATGACC TGGACTGGCC AAGGCAACCG
TCCGAGCGGA GTAATTGTGA CTCAAGGTTT GGCTCGGCCA TCGAACGAGG
GAGGCCACTA CATGCACTAT GCCACCGTGG ATGATTTCCT GACGGACTGG
```

TTCTACCTGC TTCGCAAGGA CGGGTCTTAC AAGGTATCTG GTGCATTGAC
 CTTCAGCGAG TCCATTAAGG GCATGTTCCA GGTTGGCGGA GCTAAATACG
 ACTATGCAGC CGCCGGCTAC GATAGTTACC TGGTCGGCGC CACTAGCAGG
 CTAAAAGCTA TCGAGTCCGA AAATGGCAGT CTGACACGGT TTGATGCCAC
 ATCAAATAAT GTCCATTCCG TTGACCCTGA TAAAATCTCT GTTGATATTG
 ACGGCATTGA AGTTACGATC AATGGTGTTG TCTACAAGCT GGAAAAGAAA
 CCAGTCTAA

MTINLETSIR WMSDRVGVKS YSMYRNGPN SYDCSSAVYY ALMAGGAISA
GWAVNTEYMH DWLIRNGYVL VAENKPFNAQ RHDVCILGKR GYSSGAGGHV
VIFVDNVNVI HCNYARNGIS IDNYNQVHRG MYYYLYRPAN QPSISNKSLD
QLVKETLAGV HGNGDTRKAS LGSQYEAVMA VINGKASASE KSD EELAREV
LAGKHGAGED RKRS LGPRYE PVQAKVNELL KAKEKPSETA KNEPQTVQFK
EDGDLSEFNGA ILKKS VLEII LKKCKEHDIL PSYALTILHY EGLWGTSAVG
KADNNWGGMT WTGQGNRPSG VIVTQGLARP SNEG GHYMHY ATVDDFLTDW
FYLLRKDGSY KVS GALTFSE SIKGMFQVGG AKYDYAAAGY DSYLVGATSR
 LKAIESENGS LTRFDATSNN VHSVDPDKIS VDIDGIEVTI NGVVYKLEKK PV

**NOTE: The full-length gene/protein shown here is what was cloned during the initial functional screen of the *S. suis* 7711 genome. For large-scale expression, purification, and functional analysis, a truncated construct was sub-cloned that omitted the C-terminal enzymatic domain. The residues included in this construct are underlined above.

Chapter 5: PlySs2

- GenBank Accession Number: ZP_03625529 (from *S. suis* strain 89/1591)
- Originally annotated in GenBank as “SH3-type 5 domain protein”
- 738 base pairs; 245 amino acids; 26.06 kDa; Theoretical pI = 9.06
- CHAP enzymatic domain (PF05257)
- SH3-Type 5 binding domain (PF08460)

```
ATGACAACAG TAAATGAAGC ATTAAATAAT GTAAGAGCTC AGGTTGGGTC
CGGTGTGTCT GTTGGCAACG GCGAATGCTA CGCTTTGGCT AGTTGGTACG
AGCGCATGAT TAGTCCGGAT GCAACTGTCT GACTTGGCGC TGGTGTGGGC
TGGGTCAGCG GTGCAATCGG CGATACAATC TCTGCCAAAA ACATCGGCTC
ATCATACAAC TGGCAAGCTA ACGGCTGGAC AGTTTCCACA TCTGGTCCAT
TTAAAGCAGG TCAGATTGTG ACGCTTGGGG CAACACCAGG AAACCCTTAC
GGACATGTGG TAATCGTCGA AGCAGTGGAC GGCGATAGAT TGACTATTTT
GGAGCAAAAC TACGGCGGGA AACGTTATCC CGTCCGTAAT TATTACAGCG
CTGCAAGCTA TCGTCAACAG GTCGTGCATT ACATCACACC GCCTGGCACG
GTCGCACAGT CAGCACCCAA CCTTGCAGGC TCTCGTTCCT ATCGCGAGAC
GGGCACTATG ACTGTCACGG TCGATGCTCT CAATGTTTCGC AGGGCGCCAA
ATACTTCAGG CGAGATTGTA GCAGTATACA AGCGTGGTGA ATCATTTGAC
TATGATACTG TCATCATCGA TGTC AATGGC TATGTCTGGG TGTCTTACAT
AGGCGGCAGC GGCAAACGTA ACTACGTTGC GACGGGCGCT ACCAAAGACG
GTAAGCGTTT CCGCAATGCT TGGGGTACAT TTAAATAA
```

```
MTTVNEALNN VRAQVGSGVS VGNGECYALA SWYERMISPD ATVGLGAGVG
WVSGAIGDTI SAKNIGSSYN WQANGWTVST SGPFKAGQIV TLGATPGNPY
GHVVIVEAVD GDLRTILEQN YGGKRYPV RN YYS AASYRQQ VVHYITPPGT
VAQSAPNLAG SRSYRETGTM TTVTDALNVR RAPNTSGEIV AVYKRGESEFD
YDTVIIDVNG YVWVSYIGGS GK RNYVATGA TKDGKRFGNA WGTFK
```