ENGINEERING BIOLOGICAL EXTREMES: GENETICALLY ENCODED IMMOBILIZATION OF PROTEINS THROUGH BACTERIAL SPORE-DISPLAY

A dissertation

submitted by

Trevor B. Nicks

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in

Biotechnology Engineering

TUFTS UNIVERSITY

February 2024

Advisor: Nikhil U. Nair, Ph.D.

Abstract

Recombinant proteins play critical roles in society as therapeutics, vaccines, sustainable catalysts, and more. The utility of proteins is intrinsically linked to their stability, and, while many proteins do their natural jobs very well, they often fall short in the extremes of medical and industrial settings. Consequently, numerous tools have been developed for engineering protein function and stability. Specifically, surface-display systems like phage-display, bacterial cell-surface display, and yeast-display enable powerful high-throughput engineering of proteins. But the use of these research tools often ends in the laboratory. Once identified, the proteins found with these tools are most often sent out into the world on their own. Bacillus subtilis spore-display is unique among surface-display technologies because of the innate durability of spores. Bacterial endospores (or just spores) are a dormant cell phenotype made in response to environmental stress; they are metabolically inactive, and their sole purpose is to protect a copy of the bacterial genome until it senses its environment is once again amenable to growth. They are one of the most, if not the most, stable forms of self-replicating life. The stability of spores comes from their unique structure; a dehydrated core surrounded by concentric shells of protective proteins called the spore coat.

Spore-display is the practice of engineering the spore coat to carry recombinant proteins. This has been demonstrated with antigens to make vaccines and with enzymes to functionalize spores as catalysts. Spore-display alone can improve a proteins stability and the unique qualities of spores could enable them to be more than a research tool. They could possibly be engineered alongside their cargo proteins to act as functional protein carriers, further stabilizing and augmenting protein function and enabling biology to function in new extremes. Towards this goal, we set out to advance our understanding of what protein components of the spore coat could be used as anchors for spore-display of enzymes and other proteins.

There are more than 60 known spore coat proteins, but only 11 had previously been used for spore-display of enzymes. We benchmarked a total of 33 spore coat proteins, testing their ability to act as anchors for the spore-display of β-glucuronidase (GUS) of *E. coli* and characterizing their surface availability. We found that a spore coat protein that had never been used before for spore-display SscA performed better than any of the previously known anchors. We also show that spore-display can protect the GUS cargo enzyme from protease degradation and that SscA can be used for spore-display of other enzymes. We also detail efforts towards engineering the natural competence machinery of *B. subtilis* that could enable enhanced genome engineering in future studies. Overall, this thesis describes tools and methods for studying and engineering bacterial spores as genetically-encoded protein carriers.

Dedicated to to my mom and grandmothers, each of whom raised me to love to cook. I cherish the fact that my first microbiology and chemistry experiments were in their kitchens – pickling fresh garden vegetables with my grandma Talitha and baking with my mom and grandma Henrietta.

Acknowledgments

Thank you to my father for showing me what hard work looks like. To my mother for doing the same while also being my biggest cheerleader. To my many wonderful teachers and coaches in Owensville who pushed me to do more. To my friends and teammates from high school and college who stuck by me after I came out and have continued to support me through this journey. To my chosen family that have surrounded me in times of joy and sorrow. To the many Cambridge and Somerville housemates turned lifelong friends.

Thank you to Elliot, Seki, Ian, and others at IndieBio EU 2016 for growing my love of biotechnology research. Thank you to Emily for supporting my love for it again now. Thank you to the faculty, coaches, staff, and my peers at William Jewell College who showed me the world is much bigger than I realized while growing up on a farm in Missouri.

A special thank you to Professor Lilah Rahn-Lee who in her first semester as a professor and in the subsequent two years as my research advisor first introduced me to – and got me hooked on – the wonderfully weird world of microbes.

Thank you to my lab mates, cohort, and friends at Tufts, who provided unfaltering support and guidance, and often made the late nights enjoyable: Dr. Sean Sullivan, Dr. Josef Bober, Dr. Vikas Trivedi, Dr. Karishma Mohan, Dr. Zachary Mays, Dr. Bruno Nascimento, Dr. Debika Choudhury, Dr. Damayanti Chakravarty, Dr. Ming Lei, Dr. Grace Yao, Dr. Ece Gulsan, Dr. Arlinda Rezhdo, Dr. Jessica Stieglitz, Dr. Sara Amin, Aaron Love, Brett Irwin, Karin Yanagi, Rana Said, Bec Condruti, Jessica Lee, Pomai Yamaguchi, Jack Cutting, and Arushi Kalia.

A special thank you to Dr. Todd C. Chappell whose mentorship, dedication, and creativity helped move this research forward.

A deep thank you to Professor Nikhil U. Nair who welcomed me as a mentee and enabled me to grow as an independent researcher while being incredibly gracious at times when I needed to support my family.

Thank you to my committee members, Prof. James Van Deventer, Prof. Andrew Camilli, and Prof. Neel Joshi, whose helpful conversations and insights provided not only context for this work but actionable feedback that helped make it possible.

To Eduardo, thank you for your love, your support, and for being my first editor.

Contents

List of Tables

List of Figures

Preface

My venture into sustainable technologies began when I was 14. My older brother had just dropped out of college to start his own sustainable construction company. Spurred by his interest in sustainable technology, and equipped with all the tools on our farm, my brothers, father, and I spent the entire summer of 2008 converting an old diesel truck to run on used frying oil from our local Chinese restaurant. I felt a great deal of joy and pride while riding in that truck on the dirt roads of the Ozarks, the smell of burnt Chinese food in the air as I imagined fleets of trucks and tractors running on recycled oil. It inspired me. My interest in sustainable carbon sources shaped my undergraduate research in biofuels, my first company, and now my Ph.D. research. Herein, I detail the development of a synthetic biology platform that could one day serve as the foundation for solutions ranging from new vaccines to sustainable chemical production. In her novel, Frankenstein, Mary Shelley tells the story of a scientist using biology, chemistry, and electrical engineering to create a new lifeform. As the first writer of science-fiction, she inspired countless other writers and created the genre of science fiction which has shaped the imagination of many inventors. While fiction, there is a truth to be found in her first novel interdisciplinary technologies that combine biological, chemical, electrical, and other engineering principles can be very powerful. We need new technologies to decarbonize the global economy and to pull carbon from our atmosphere. The entire lifecycle of these new technologies must be sustainable and non-polluting. And, importantly, they need to be economical.

Much like Dr. Frankenstein, the field of synthetic biology has repeatedly tried to create and engineer new organisms. These new organisms are meant to act as "cell factories" that create useful chemicals from sustainable carbon sources like CO2 and sugar. These cell factories are often proposed as the ideal solution for replacing the petroleum-based chemical industry with sustainable alternatives. However, as I learned with my first company, working with living cells is challenging and scaling them is often difficult. Their failure is often rooted in the fact that living cells are complex and fragile; they contain thousands of proteins and other molecules within thin membranes. Balancing product yield, titer, production host survival, and cost can become an impossible task. These systems are inherently inefficient because they are alive and prioritize their growth over production. For me, this raises the question: is the key to scaling biology using less of it? Rather than forcing living cells to act as miniature biochemical reactors, using simpler non-living enzymatic systems cell-free technologies for catalysis has many advantages over living systems (Rasor et al. 2021). These systems are made through the thoughtful curation and in vitro reconstitution of the enzymes, substrates, and other chemicals needed to create a product. Like Frankensteins monster, these systems can even be powered by electricity (Schlager et al. 2016). Cell-free systems are highly-customizable and enable the optimization of several essential production parameters: yield, concentration, and productivity. Cell-free systems can be made as simple as a single enzyme or contain complex enzyme cascades to convert sustainable substrates like glucose and CO2 to commodity or high-value chemicals. Intentional metabolic minimalism results in near

100% theoretical-yield for these systems. Importantly, because these systems are well-defined and highly-engineerable, if a biochemical pathway can be imagined it can be made. Cell-free technologies also liberate engineers to create pathways with increased concentrations of enzymes and chemical intermediates that would otherwise be unobtainable or kill a living cell. Finally, bioprocess design for cell-free systems could theoretically be simpler than for living cellular systems and further reduce operational cost. Cumulatively, these qualities could enable higher total productivity and efficiency than cellular systems on many relevant metrics. But there are two limiting constraints existing scientific hurdles that are preventing cell-free systems from being widely scaled and commercialized. The first is enzyme cost. The second is the catalytic energy source. Enzyme cost is shaped by three factors: production cost, enzyme lifetime, and enzyme recyclability. An existing low-cost method for producing enzymes for cell-free systems is to express the enzymes of thermophilic organisms in a mesophilic host. After cell lysis, heating of the cell lysate inactivates the mesophilic enzymes, leaving behind a defined set of thermostable catalysts. This strategy has even been used to produce myo-inositol in 20,000 L reactors (You et al. 2017). However, there arent always highly-active thermostable enzymes for a desired reaction and this solution doesnt solve the issue of enzyme lifetime and recyclability. Additional low-cost strategies for enzyme purification are highly desirable. Take for example the enzymatic digestion of lignocellulosic biomass, the breaking down of plant waste biomass in a process called saccharification. Biorefineries use saccharification as the first step

in turning plant waste into chemicals. These are mostly theoretical today, as many of the first saccharification biorefineries have shut down due to economic constraints, but there is renewed interest and vigor in this area because of climate change. To release the sugars from biomass so that microbes can ferment it and produce a desired product, bioprocess engineers treat it with a cocktail of different enzymes called cellulases and glucosidases. These enzymes are essential for products like lignocellulosic biofuels, but they are very expensive and can account for ≈20% of the total cost of biofuels (Aui, Wang, and Mba-Wright 2021). This cost is high because existing systems require new cellulases be used for every new batch of lignocellulose. If these enzymatic systems could be engineered to have improved stability and/or be recycled, the cost of biofuel production could be reduced. This is a topic of ongoing investigation at the National Renewable Energy Laboratory which has been designing membrane-based systems to recycle cellulase enzymes in a process called continuous enzymatic hydrolysis (Stickel et al. 2018). If these enzyme recycling systems will work at an industrial scale remains to be seen as membranes may foul and be damaged with higher levels of solids loading. Continuing with saccharification as an example, it can be broken down into three modules: pre-treatment, enzymatic conversion, and purification. Pretreatment of lignocellulosic biomass is essential because its crystalline structure would otherwise resist enzymatic breakdown. But pre-treatment itself often introduces organic solvents or acids that that can negatively affect enzyme function and introduces a requirement for biomass washing between pre-treatment and enzymatic digestion (Rodriguez-Ziga et al. 2015; Zhang, Tervo, and Reed

2016). If cellulase enzymes could be made resistant to these extremes, these additional wash steps could be avoided. This is just one instance in which added enzyme stability in a cell-free system would be advantageous. The second constraint in scaling cell-free systems in the catalytic energy source. Many, but not all, industrially relevant enzymes require an energy source to catalyze their reactions. In living systems, cellular metabolism continuously regenerates small molecules called cofactors that carry energy in the form of chemical bonds with which enzymes power their reactions. Adding fresh cofactors to a cell-free system would be prohibitively expensive at scale so they must be regenerated within the system. Careful planning of cofactor regeneration and pathway design can enable some cell-free processes. Ideally, however, traditional biological cofactors could be done away with entirely in favor of a less labile and expensive energy source. While this is not an area of research we pursued, there are several research groups actively working on this problem (Bowie et al. 2020). We are building a platform to improve enzyme stability and recyclability, and enable scaled cell-free technologies for economical, sustainable biocatalysis. Within my thesis work, I have characterized and developed new methods for engineering the dormant spores of the bacterium Bacillus subtilis. Bacterial endospores (hereafter referred to as spores) are the most robust form of life on the planet. Even though they are technically a cell, spores have no active metabolism and are relatively inert. Spores are made as a survival response to environmental stress and act as a protective cocoon to the dormant bacterias genome. Spores are extremely stable, surviving extremes in temperature, pH, solvent, UV radiation, and

other conditions that would kill a living cell. We use synthetic biology to engineer spores, turning them into microparticle protein carriers for recombinant proteins this is called spore-display. When proteins are incorporated onto the spores, they gain a level of stability as well and can have increased longevity. Moreover, spores of *B. subtilis*are large enough (1-2 microns in diameter) that they can be easily purified from a reaction mixture using centrifugation or easily embedded in materials. The technology and intellectual property (patent pending) I have synthesized during my Ph.D. is an expansion-pack for the synthetic biology toolkit for creating spore-display systems and has implications for sustainable catalysis, vaccines, and more. With spore-display, there is potential to re-design pre-processing and bioconversion modules with a "cell-free" bioprocess enabled by durable spore-displayed enzymes. Importantly, we will call these synthetic biochemistry systems, not cell-free, because they will contain spores. With continued research and development, spore-display has the potential to be an important tool in societys development of sustainable bioconversion technologies.

Chapter 1

Introduction and review of *B. subtilis* as a chassis organism: emphasis on spore-display and natural competence

1.1 History of *B. subtilis* as a model organism

In microbiology, the organisms *Escherichia coli*, *Bacillus subtilis*, and *Saccharomyces cerevisiae* have served as the model for Gram negative bacteria, Gram positive bacteria, and eukaryotes, respectively, for decades. The ease of cultivating and genetically manipulating all three has also led to their broad use in industrial biotechnology to produce chemicals and proteins that have uses ranging from therapeutic insulin to enzymes in laundry detergent. *B. subtilis*, specifically, is of interest to us because it is generally recognized as safe (GRAS) by the FDA, it is naturally competent, and is the most well-characterized of all endospore-forming bacteria (Sonenshein, Hoch, Losick, et al. 2002). *B. subtilis* has been used for centuries in Japan and other Asian countries to produce natto, a fermented soybean food product. During fermentation, *B. subtilis* produces a biofilm and poly-glutamic acid that leads to a tasty, sticky matrix between the soy beans. The isolation of *B. subtilis* strains for commercial sale and natto production was done by Professor J. Hanzawa of Hokkaido Imperial University in 1928. Because *B. subtilis* sporulates, spores of desirable natto-producing strains

Figure 1.1: Fermented soybeans (Natto) from Japan purchased at Ebisuya Japanese Market in Medford, MA. The grey-brown substance around the beans is the result of *B. subtilis* fermentation.

could be easily distributed for natto production without refrigeration (Hosoi, Kiuchi, et al. 2003). This is the first known commercial use of spores. But *B. subtilis* is not a model organism because it makes tasty beans. It was developed as a model organism because it sporulates. Our own human bodies go through extreme morphological changes as we progress from a single-celled zygote into a fully-developed multicellular organism. To understand the molecular underpinnings of these changes (cellular division, cell-signaling, control of gene expression, etc.), scientists needed a malleable, observable, and fast-growing biological model with which to tinker and experiment: sporulation became that model. And as microbiologists learned more about *B. subtilis*, it became used for even more.

1.2 The many uses of *B. subtilis*

Today, *B. subtilis* is the second-most well characterized and studied organism on Earth, surpassed only by *E. coli*. With a genome of 4.2 Mbp, *B. subtilis* is incredibly versatile and capable of several unique phenotypes that enables it to fill unique niches in the environment and science (Kunst, 1997; Earl, 2008).

1.2.1 Natural Competence

The first of these unique phenotypes, first documented over eighty years ago, is the ability of *B. subtilis* to take up DNA from its environment (Anagnostopoulos and Spizizen 1961). Now called Natural Competence, we know that when *B. subtilis* experiences stress/starvation it can turn on genes to create a protein apparatus and enzyme relay that binds external DNA, internalizes it, and then incorporate it into its genome (Maier et al. 2004). Microbiologists realized the potential utility of this phenomenon immediately and began using natural transformation to generate unique strains of *B. subtilis* to investigate the effects of different mutations on sporulation (Schaeffer 1969). Work in studying natural competence in *B. subtilis* shaped the entire fields perspective and methods for genome engineering of all kinds of organisms. Understanding these mechanisms for the transfer of genetic information between organisms has real world impacts, for example, how genes responsible for antibiotic resistance move from organism to organism and contribute to the spread of untreatable infections through genetic and even non-genetic inheritance (Dalia and Dalia 2019).

1.2.2 Biofilm Formation

In other instances, *B. subtilis* will form a biofilm like during natto production in which several different cells contribute to an external matrix of materials that enables them to stay in place create divisions in labor and morphologies, providing enhanced protection from invasion of other microbes. Learnings from studying *B. subtilis* biofilms have also contributed to our understanding of how pathogens can cause hard-to-treat drug-resistant infections (Morikawa 2006). Another interesting example of *B. subtilis* biofilm studies leading to interesting findings is the use potential use of a biofilm protein in ice cream! The BslA protein (biofilm surface layer protein A) has been shown to form an extremely hydrophobic surface, more hydrophobic than Teflon, on the top of *B. subtilis* biofilms. Researchers incorporated the protein into ice cream and found that it helped stabilize the emulsion of fat and air (Stanley-Wall and MacPhee 2015).

1.2.3 Plant-Bacteria Symbiosis

As a soil-dwelling microbe, *B. subtilis* often colonizes the roots of the plants in a form of symbiosis (Poole 2017). Our understanding of how *B. subtilis* interacts with plants has shaped the development of an entire synthetic biology industry for developing synthetic microbes that can act as fertilizer or aids in agriculture, increasing crop productivity. Companies like Joyn Bio and Robigo exist today because of the foundational work that started in studying *B. subtilis* as a model for plant-microbe interactions (Marrone 2023).

1.2.4 Protein Production Chassis

The first instance of recombinant human protein production was done in *E. coli*, and it is still commonly used today. However, *B. subtilis* has unique protein secretion machinery that can simplify protein production and purification. Additionally, *B. subtilis* does not produce endotoxins or other immunogenic molecules that pose a risk to human health. Accordingly, *B. subtilis* has seen wide adoption as an industrial workhorse to produce enzymes used in food, home, and therapeutic products at titers > 25 g/L (Cui et al. 2018).

1.2.5 Sporulation

Sporulation has also played roles other than as a bacterial storage system and model for development biology and genetics. In 2001, an Italian research group used recombinant DNA technologies to produce a spore-based tetanus vaccine (Isticato et al. 2001). This was the first instance of spore-display of recombinant proteins and their novel approach enabled the expression of the c-terminal fragment of the tetanus toxin to serve as a potential vaccine. Since this study, there have been numerous instances of spore-display including the production of a potential COVID-19 vaccine and the spore-display of enzymes for use in biocatalysis. In addition to spore-display, Aanika Bioscience, a biotechnology company in Brooklyn, New York, has been developing DNA-barcoded spores to track and verify the origin of produce and other goods as they move through the global supply chain. Our interest in *B. subtilis* is rooted in the many potential use cases of spore-display. Our aim was

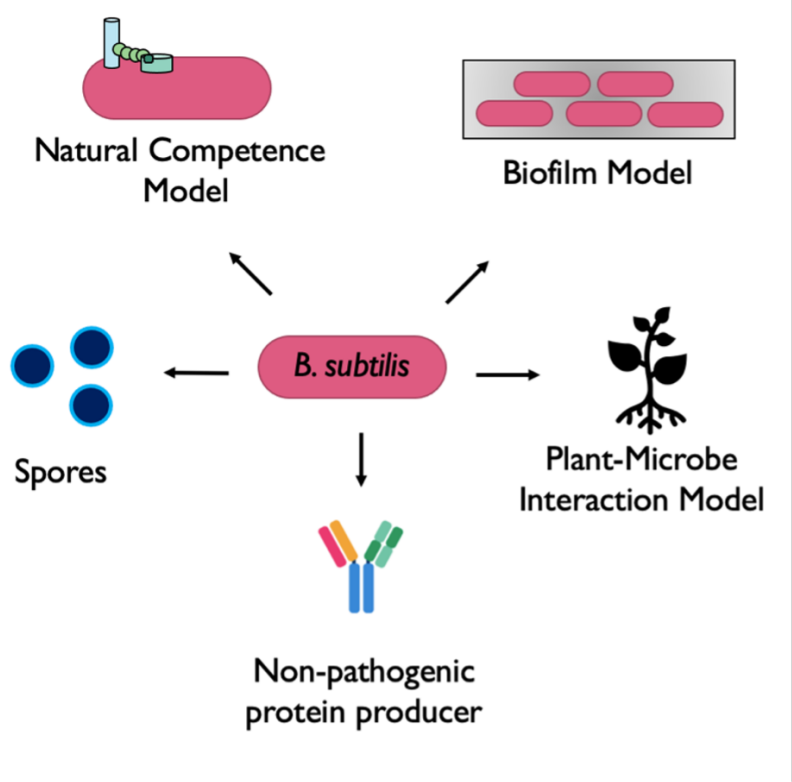


Figure 1.2: The many phenotypes of *B. subtilis*: While *E. coli* and *B. subtilis* have similarly sized genomes, *B. subtilis* is uniquely able to differentiate itself into spores, modify its own genome via natural competence, form non-pathogenic biofilms, form symbiotic relationships with plants, and efficiently secrete proteins.

to identify, characterize, and engineer additional genetic parts that can enable spore-display. These efforts, summarized in Chapter 2 and in an upcoming publication, resulted in a new framework for investigating spore-display of recombinant proteins, new and improved methods for evaluating and characterizing spores, and the identification of new genetic parts for use in spore-display. Additionally, in Chapter 3, because spore-display requires editing the *B. subtilis* genome, we completed research towards improving natural competence. While these efforts did not yield functionally useful results, we are sharing them publicly here so that others can learn from our efforts.

Chapter 2

A Thorough Characterization of Spore Coat Proteins as Anchors for Spore Display in *Bacillus subtilis*

2.1 Review of surface-display systems

2.1.1 Phage Display

The first microbial surface-display system was developed in 1985 by Professor George Smith at the University of Missouri. Now a Nobel Laureate, Prof. Smith developed what is called phage display (Smith 1985). He demonstrated that a foreign DNA fragment could be functionally inserted into the sequence of a bacteriophage coat protein. He and then post-doc Jamie Scott worked to identify fusion arrangements that did not inhibit the formation of the viral capsid and showed that (1) a recombinant protein could be displayed on the surface of a phage without inhibiting its ability to infect *E. coli* and (2) that millions of phage variants could be made. With these displayed proteins, they could use antibodies to bind specific protein-displaying phage from within a group and enrich a target recombinant phage by over 1000-fold within a population (Scott and Smith 1990). The essential function of phage-display as a tool is the establishing and maintenance of a genotype-phenotype link: the genetic sequence coding for the recombinant protein is physically associated with its protein product because it is encapsulated

within the phage on which the protein is surface-displayed. This link enables the enrichment or sorting of genetic sequences by the function of the proteins they encode. This foundational work in phage-display was an important precedent and laid the groundwork for several other microbial surface-display technologies. Phage-display was a powerful innovation, but it cannot be used with every protein. Some proteins proved to be too large for bacteriophage to display and others could not be folded properly by the bacterial translational machinery required for phage assembly. To increase the number of proteins that could be used with surface-display, researchers aimed to identify other systems in which a genotype-phenotype link could be established through surface-display of recombinant proteins. This led to innovations in several unique organisms and gave rise to three additional categories of surface-display platforms: bacterial cell-surface display, yeast-display, and spore-display. These systems have since been used for antibody engineering and as bioadsorbents, oral vaccines, biosensors, catalysts, and more. All these systems maintain the genotype-phenotype link by having the same basic architecture as phage-display: a "carrier" that simultaneously contains the genetic content and constitutes the surface on which the recombinant protein can be displayed. As such, the essential first step in developing surface-display systems has been in identifying what within the carriers structure can serve as an anchor for the recombinant protein. Anchor proteins and how they are assembled on the surface of the carrier play an important role in determining the functionality and potential use cases of a surface-display system and varies between different surface-display systems.

2.1.2 Bacterial Cell-Surface Display, emphasis on *E. coli*

In *E. coli* cell-surface display, one of the first anchors used was the Lpp-OmpA protein complex. This synthetic anchor was created by fusing the first nine amino acid residues of the *E. coli* lipoprotein (Lpp) with residues 46-159 of the *E. coli* OmpA porin protein which self-assembles into the Gram-negative bacterial cell wall envelope (Francisco, Earhart, and Georgiou 1992). When expressed with a recombinant protein fusion, the polypeptide is secreted into the periplasm and then assembles on the outer membrane with the cargo protein on the exterior of the *E. coli* cell, where it can be bound by antibodies or used for other application-specific assays. This was first documented with a beta-lactamase enzyme which was surface-displayed and then detected through an activity assay and immunostaining (Francisco, 1991). Since its inception, *E. coli* cell-surface display has been used for several protein engineering applications (Bloois et al. 2011). Unfortunately, there is one extreme weakness to *E. coli* cell surface display: it often dramatically weakens the cell envelope and reduces cell viability. While there has been work done to identify other anchor proteins that do not weaken the cell membrane, the potential for an increase in cell fragility is a limiting factor that could prevent *E. coli* cell-surface display systems from being grown at scale. Researchers have developed other bacterial cell-surface display systems in other Gram-negative and Gram-positive bacteria.

2.1.3 Yeast Display

An additional shortcoming of both phage and bacterial cell-surface display technologies is that they are often incompatible with eukaryotic proteins. As such, researchers aimed to develop a eukaryotic surface-display system that would have increased compatibility with human proteins such as antibodies. For this, a research group at the University of Amsterdam turned to the yeast S. cerevisiae for its qualities as the most-studied and well-characterized single-celled eukaryote. At the time, it was already known that the cell wall of S. cerevisiae was primarily composed of a macrostructure of glucan and mannoproteins. They chose a well-characterized mannoprotein, α-agglutinin (AGα1), to serve as the anchor for a yeast-display system. By fusing the membrane-anchoring C-terminal half of α-agglutinin to the α-galactosidase enzyme, they demonstrated that a recombinant protein could be expressed and surface-displayed on yeast (Schreuder et al. 1993). Since this first proof of concept, yeast-display has become increasingly important as a research tool and has been used to engineer both antibodies and enzymes through innovative approaches. A powerful example of yeast-display in enzyme engineering was developed by David Lius group at the Harvard University in which they created a general strategy for the evolution of bond-forming enzymes using yeast display (Chen, Dorr, and Liu 2011). Their system used the covalent co-loading of a bond-forming enzyme and its substrate to the yeast surface. Successful display of an active enzyme enabled ligation of a second labeled substrate to the surface-bound substrate: providing a measurable link between enzyme

activity and the encoding DNA sequence through cell self-labeling. Subsequent detection of the labeled substrate complex on the yeast surface using a fluorescent probe allowed them to use fluorescence-activated cell sorting to screen a large library of enzymes using yeast-display. This powerful approach enabled them to evolve a sortase variant with 140-fold more activity on a unique substrate (Chen, Dorr, and Liu 2011).

2.1.4 Spore Display

Finally, in 2001, a research group in Italy demonstrated the expression of recombinant proteins on the surface of *B. subtilis* spores (Isticato et al. 2001). They reasoned that spores would be able to act as a carrier for surface-display of proteins just like vegetative microbial cells and phage while bringing additional attributes that could be extremely advantageous: (1) spores are extremely stable and once functionalized could act as relatively inert carriers in comparison to more fragile and metabolically active cells, (2) the spores of many Bacillus species were already known to be safe because of their presence in food products, and (3) there were already simple and inexpensive methods for the production of large amounts of spores because of their use in commercial products. They chose to focus on the spores of *B. subtilis* because they were, and remain today, the most well-characterized bacterial spore. For their first application of *B. subtilis* spore-display, they focused engineering spores to act as a tetanus vaccine. The exterior of *B. subtilis* spores is not a membrane, however, so a membrane bound protein could not be used like in *E. coli* and yeast-surface display. Instead, the spore capsule structure is more akin to that of a bacteriophage in that the

outermost layers of spores are made of a complex and highly stable protein structure called the spore coat. Knowing this, and with phage, bacterial, and yeast-display as a precedent, (Isticato et al. 2001) used a portion of the spore coat protein CotB as a fusion partner and anchor for a fragment of the C-terminus of the tetanus toxin. They showed that recombinant fusion of the tetanus toxin to CotB enabled surface-display of the molecule as measured by immunostaining and flow cytometry. When mice were orally dosed with these spores, they formed a measurable immune response (Ciabattini et al. 2004). In theory, these modified spores could be used as vaccines with a higher shelf-life than traditional vaccines. Others took note, and several academic groups have made spore-based vaccines using different spore coat anchors like CotB, CotC, and CotG as anchor proteins (Isticato 2023). During the COVID-19 pandemic, spore-display was even assessed as a potential platform for a covid vaccine. (Katsande et al. 2022) demonstrated a mucosal immune response in mice and protection from SARS-CoV-2 in hamsters. Others have shown that spore-display based vaccines for other disease administered orally or nasally can ellicit an immune response and provide protection in mice (Lee et al. 2010; Zhao et al. 2014). In addition to antigens for vaccine development, spore-display has also been used to display enzymes as whole-cell biocatalysts. Importantly, because of the structural stability of spores, spore-displayed enzymes often have higher stability than their free (soluble) versions and can be used in more extreme conditions than free enzymes or enzymes surface-displayed on vegetative bacteria or yeast. For example, lipase enzymes,

which catalyze the transesterification of fatty acids and short chain alcohols are often deactivated by industrially relevant concentrations of their short chain alcohol substrates like methanol (Lotti et al. 2015). This unfortunate fact necessitates inefficient processes at suboptimal levels of methanol and can require more enzyme be used than if a more stable catalyst could be made. Knowing this, a team at the University of California Irvine and Caltech showed that a spore-displayed lipase could retain 10x more activity than the free version of the same lipase in 100% methanol. The spore-displayed lipase was also more thermostable, with the half-inactivation temperature increasing from 45 °C to 50 °C (Hui, Cui, and Sim 2022). An increase in stability of spore-displayed enzymes within thermal, solvent, pH, and other challenges has been demonstrated several other times with a variety of enzymes (Wang, Wang, and Yang 2017). Cumulatively, extensive research in microbial surface-display systems have created four categories of surface-display: phage display, bacterial cell-surface display, yeast display, and spore display. Each platform has its own qualities, use cases, and unique methods, but their fundamental architecture is the same: a genetically replicating carrier that can maintain a phenotype-genotype link. When comparing surface-display systems and selecting one for an application, it can be useful to consider the carriers size, maximum cargo protein size, how many of those cargo proteins it can carry, its stability, and the potential for genetic diversity within a population. It is important to note that these are generalizations. For example, we have listed the dimensions for a filamentous phage which are long and slender but T7 phage have an icosahedron capsid structure. Additionally, while

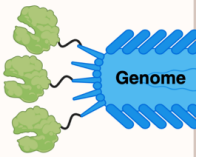
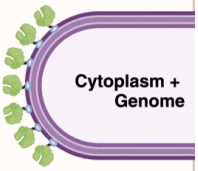
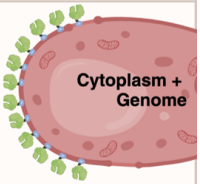
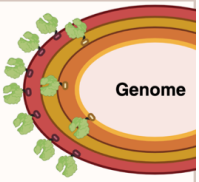
	Carrier Size	Cargo Size	Copy #	Stability	Library Size	
Phage Display	7 nm x 930 nm	110 kDa	3 - 3 x 10^3	65 °C	10^{9-12}	
Bacterial Cell-Surface Display	2 µm	119 kDa	10^3-10^6	70 °C	10^{8-10}	
Yeast Display	4 µm	80 kDa	10^5-10^6	60 °C	10^7	
Spore Display	1 µm	60 kDa	10^3-10^7	>100 °C	10^3	

Figure 2.1: Phage Display, bacterial cell-surface display, yeast display, and spore display are the four existing types of microbial surface-display platforms. Important difference between the platforms includes their size, maximum cargo size, the number of cargo proteins they can carry, stability, and potential library size. Other important qualities like protein translation and modification machinery are discussed elsewhere.

proteins as large as 110 kDa have been demonstrated with phage and bacterial cell-surface display, it is generally accepted that yeast display is better suited for the display of large proteins because of its secretion machinery. Moreover, while spore-display has not been demonstrated with a library greater than $3x10^3$ variants, larger libraries are certainly a possibility in future work. With these comparisons in mind, spore-display is the most compatible of these platforms with our goal to engineer more stable synthetic biochemistry systems for several reasons. (1) The metabolic dormancy of spores enables spore-displayed enzymes to maintain the metabolic minimalism of synthetic biochemistry systems, preventing the introduction of off-target enzymatic reactions, (2) the structural integrity of spores and their ability to stabilize displayed enzymes could lead to long-lasting synthetic biochemistry systems and (3) the size of spores enables them to be easily separated from liquid reaction products via centrifugation or filtration. However, a key limitation to spore-display for use in synthetic biochemistry systems is its anchor proteins. Several of the anchor proteins currently being used for spore-display are very large. Presumably because of their size and mishaps during protein folding, fusions to the anchor proteins sometimes fail to yield a functionalized spore. Accordingly, to make a spore-display system, one must first test several anchor proteins to identify a suitable partner. The existing "toolbox" of anchor proteins for spore-display of enzymes consists of 12 different anchors. Knowing there are more than 60 proteins that make up the structure of the spore coat (Peter T Mckenney, Driks, and Eichenberger 2012), we aimed to further characterize the proteins of the spore coat and assess their suitability as

anchor proteins.

2.2 Spore Structure, Spore Coat Proteins, and Instances of Spore-Display of Enzymes

Sporulation is a starvation stress response that *B. subtilis* and other endospore forming bacteria use to preserve themselves in low-nutrient environments. In the laboratory, sporulation occurs after the lag phase has begun in liquid cultures. It can also be induced using specific starvation liquid or solid medias. Key sporulation signals include cell-density, nutritional state, and DNA replication (Ireton et al. 1993). When specific stressor and environmental criteria are a met, a series of histidine sensor kinases are activated and lead to the phosphorylation of the sporulation transcription factor: Spo0A. This transcription factor activates a regulon that is responsible for the multi-step cellular differentiation process that is sporulation. The steps are as follows.

Sporulation as a step-wise process:

1. Starvation triggers the activation of a histidine sensor kinase phosphorelay consisting of KinA, KinB, and KinC. KinC phosphorylates Spo0A which activates the sporulation regulon through a series of sporulation-specific sigma factors.

2. The cell forms an inner membrane and bisects itself into two distinct chambers of unequal volume. A copy of the genome is then translocated across this membrane into the smaller chamber called the forespore. The larger chamber is called the mother cell. The two genomes then go through differential gene expression due to

the activation of the unique sigma factors.

3. The forespore then undergoes an engulfment process in which the bisecting membrane and a fraction of the cells original membrane undergo a process akin to phagocytosis and ultimately pull from the external membrane to form a membrane bound structure surrounding the forespores copy of the genome within the cytoplasm of the mother cell.

4. Sigma factors in the mother cell chamber lead to the expression of enzymes and spore coat proteins that begin to build the protective layers of the spore. The cortex, a layer of peptidoglycan, first begins to be synthesized on the exterior of the forespore membrane. Checkpoint proteins prevent full cortex synthesis until some spore coat proteins begin to assemble at the spores pole closest to the mother cell. Water begins to be expelled from the forespore and is replaced with Ca2+-dipicolinic acid; the cytoplasm becomes dehydrated and highly viscous.

5. The cortex and spore coat proteins then assemble and cover the entirety of the spore in distinct layers.

6. The mother cell then lyses, releasing the spore into the environment.

7. The GerQ transglutaminase and potentially other enzymes continue to work on final spore assembly by chemically cross-linking proteins within the spore coat.

Once assembled, spores can stay dormant and structurally intact for

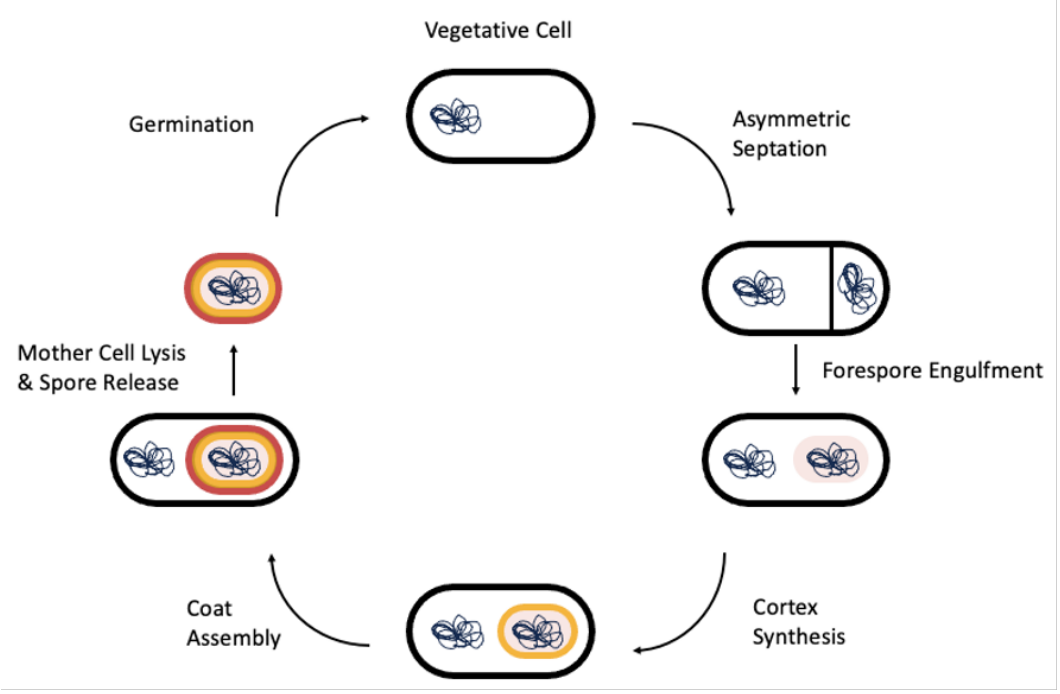


Figure 2.2: The Sporulation and Germination Cycle in *B. subtilis*

many years. A report in 1995 showed that spores trapped in amber within the digestive tract of a bee could still germinate after an estimated 25-40 million years in dormancy (Cano and Borucki 1995). These resurrected bacteria were shown to most resemble Bacillus sphaericus. The final structure of the *B. subtilis* spore is ellipsoid in shape and measures 1-2 ţm in length and 1 ţm in width. The outermost layer is the proteinaceous spore coat which itself consists of 4 unique layers: the basement layer, inner coat, outer coat, and crust. Each layer is known to be made of specific proteins thanks to a thorough set of experiments that measured the location of the spore coat proteins and characterized spore assembly using fluorescent protein fusions (Mckenney et al. 2010; Peter T Mckenney, Driks, and Eichenberger 2012). From these data, and other, a detailed map of the protein components of each spore coat layer can be made (Figure 2.3).

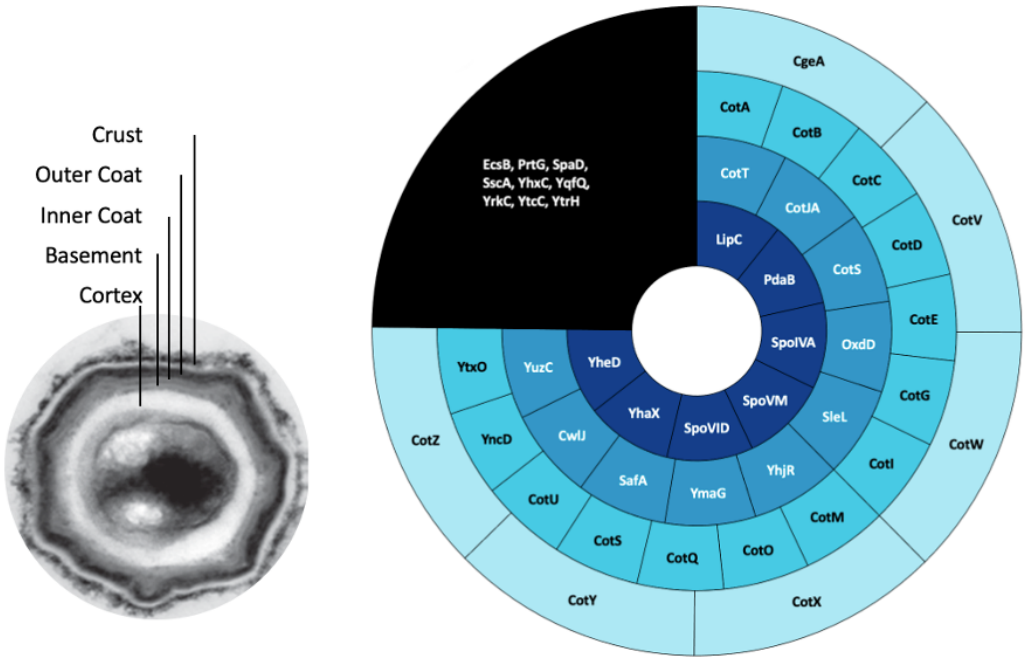


Figure 2.3: (A) Spore Coat Structure: Thin section transmission electron micrograph of a *B. subtilis* spore stained with ruthenium red (Adapted from (Mckenney et al. 2010)). (B) Map of the four spore coat layers and 47 constitutive proteins. The Basement layer: LipC, PdaB, SpoIVA, SpoVM, SpoVID, YhaX, YheD. The Inner Coat: CotT, CotJA, CotS, OxdD, SleL, YhjR, YmaG, SafA, CwlJ, YuzC. The Outer Coat: CotA, CotB, CotC, CotD, CotE, CotG, CotI, CotM, CotO, CotQ, CotS, CotU, YncD, YtxO. The Crust: CgeA, CotV, CotW, CotX, CotY, CotZ. And nine proteins with unknown location: EcsB, PrtG, SpaD, SscA, YhxC, YqfQ, YrkC, YtcC, YtrH.

While not every known spore coat protein was included in these fluorescent protein studies, those that were enable a detailed understanding of the *B. subtilis* spore coat structure. Of these spore coat proteins, only 11 have previously been used for spore-display of recombinant enzymes: CgeA, CotB, CotC, CotE, CotG, CotV, CotW, CotX, CotY, CotZ, and OxdD. A detailed summary of all published reports of spore-display of enzymes can be found in Table 2.1 below. We have also highlighted a few proteins of interest below. Of note, spore coat protein CotA is an enzyme itself and CotA-containing spores have garnered attention as potential biotechnology tools, so it is also discussed.

2.2.1 CgeA

CgeA is a member of the crust layer of the spore coat and was recently shown to play a role in the assembly of polysaccharides on the spore surface. Additionally, we know that the N-terminal half of CgeA is required for its localization to the spore while its C-terminal half contains an O-glycosylation site at T112 (Nakaya, 2023). Thus, CgeA is similar in its function and surface location as the AGα1 anchor protein used in yeast display. To our knowledge, there is only one report of using CgeA for spore-display of enzymes (Bartels et al. 2017) where N and C-terminal fusions were tested with laccases from both *B. subtilis* and *E. coli*. Only the *B. subtilis* laccase was functional when fused to CgeA.

2.2.2 CotA

CotA is one of the few protein members of the spore coat that display enzymatic activity (Table 2.2). CotA is a copper-dependent laccase (Hullo

et al. 2001) that localizes to the outer coat. While CotA hasnt served as an anchor for display of recombinant cargo proteins, the functionality of CotA as a spore-encased enzyme by itself has garnered it attention as a biotechnology tool. Indeed, the first instance of directed evolution of a spore-displayed enzyme was performed on CotA. The Farinas group at the New Jersey Institute of Technology has demonstrated directed evolution of CotA for substrate preference, solvent tolerance, and pH tolerance while encased within the spore coat a (Gupta, 2010; Jia, 2014; Sheng, 2017).

2.2.3 CotB

CotB was the first spore coat protein to be used for spore-display of any protein (Isticato, 2001). Importantly, spore-display does not require the full 42.8 kDA sequence of CotB. Instead, N or C-terminal fractions of 275 or 41 amino acids long, respectively, can be used for successful spore-display of proteins (Isticato, 2001). Since it was first used as an anchor for spore-display of the tetanus toxin, CotB has served successfully as an anchor for four enzymes including a lipase, esterase, amylase, and cellulase.

2.2.4 CotG

Of all the spore coat proteins, CotG has been the most widely used in spore-display of enzymes. CotG is an outer coat protein with a CDS coding for a 23 kDa protein but it undergoes a form of chemical cross-linking and is found in the spore coat as a 36 kDa protein. Its assembly is dependent on the heat-sensitive kinase CotH which phosphorylates

it as three unique sites (Isticato, 2020; Nguyen, 2016). It has been functionally fused to 15 unique enzymes.

2.2.5 CotY & CotZ

The crust proteins (CgeA, CotV, CotW, CotX, CotY, & CotZ) of *B. subtilis* were shown to form the thin outermost layer of *B. subtilis* spores for the first time in 2010 (McKenney, 2010). Since then, they have gained a great deal of attention for their use in spore-display. This is because they are presumably the most surface-available of all the spore coat proteins and fusion of cargo proteins to them would lead to greater immunogenicity for vaccines and potentially greater activity for enzymes. (Bartels et al. 2017) did a comparative analysis of all six crust proteins as spore-display anchors. They first tested the strength of the promoters controlling expression of these six proteins: PcotVWX, PcotX, PcotYZ, and PcgeA. They tested the strength of each promoter through the use of a luciferase reporter assay and found that PcotYZ gave the strongest signal. Afterward, they placed all six proteins under control of the PcotYZ promoter and tested their ability to act as N or C-terminal fusions to two GFP and two different laccases. CotY and CotZ consistently outperformed the other crust anchors in their tests. This may have been because CotY and CotZ were under control of their native promoters while the others were not and changes in expression dynamics may have affected assembly of the others A recent study also used CotY and CotZ for as the anchors of choice to increase the copy number of enzymes loaded to the spore. (Hui, Cui, and Sim 2022) built a T7 expression system for the synthetic induction of CotY and CotZ

fusions. Impressively, they report an increase of the loading density of spores by approximately two orders of magnitude to $10^6 - 10^7$ copies of cargo proteins with their synthetic circuits.

Table 2.1: Spore Coat Publications

Spore Coat Protein	**Molecular Weight (kDa)**	**Enzyme (Organism of Origin)**	**Publication**
CgeA	14.02		
		Laccase (*B. subtilis*)	Bartels, 2017
CotA	58.33		
		Laccase (*B. subtilis*)	Gupta, 2010
		Laccase (*B. subtilis*)	Gupta, 2010
CotB	42.8		
		Lipase	Chen, 2015a
		Esterase (*C. thermocellum*)	Chen, 2015b
		Amylase	Nguyen, 2011
		Cellulase	Nguyen, 2011
CotC	9.56		

		Alcohol Dehydrogenase (*B. mori*)	Wang, 2011
		Alcohol Dehydrogenase (*A. pasteruianus*)	Yuan, 2014
		Trehalose Synthase	Liu, 2018
		Cellulase	Nguyen, 2011
CotE	20.83		
		Lipase A (*B. subtilis*)	Kim, 2017
		Lipase B (*B. subtilis*)	Kim, 2017
		Tyrosinase (*B. megaterium*)	Hosseini-Abari, 2016
		Beta-galactosidase	Hwang, 2013
CotG	23.81		
		Haloalkane Dehalogenase	Wang, 2019
		Beta-galactosidase	Kwon, 2007; Tavassoli, 2012; Hwang, 2013
		Nitrilase (*C. thermocellum*)	Chen, 2016a; Chen, 2017
		Nitrilase (*C. maritima*)	Chen, 2016b
		Trehalose Synthase	Liu, 2019

		Transaminase (*V. fluvialis*)	Hwang, 2011b
		Phytase (*E. coli*)	Mingmongkolchai, 2019
		Chitinase (*B. pumilus*)	Rostami, 2017
		L-arabinose isomerase (*L. brevis*)	Guo 2018
		Aldolase (*E. coli*)	Xu, 2011
		Phosphotriesterase (*P. diminuta*)	Song, 2019
		Glucose 1-dehyderogenase (*L. sphaericus*)	Gao, 2016
		Triphenyl methane reductase (*Citrobacter sp.*)	Gao, 2016
		MCP Hydrolases (*D. ginsengisoli*)	Qu, 2014
		Cellulase	Nguyen, 2011
CotV	14.09		
		Laccase (*B. subtilis*)	Bartels, 2017

		Laccase (*E. coli*)	Bartels, 2017
CotX	18.46		
		Laccase (*B. subtilis*)	Bartels, 2017
		Laccase (*B. subtilis*)	Bartels, 2017
		Laccase (*E. coli*)	Bartels, 2017
		Beta-galactosidase	Wang, 2016
		Sucrose Isomerase	Zhan, 2020
CotW	12.2		
		Laccase (*B. subtilis*)	Bartels, 2017
		Laccase (*E. coli*)	Bartels, 2017
CotY	17.74		
		Laccase (*B. subtilis*)	Bartels, 2017
		Laccase (*E. coli*)	Bartels, 2017
		Lipase A (*B. subtilis*)	Hui, 2022

		Lipase B (*B. subtilis*)	Hui, 2022
		Beta-galactosidase	Wang, 2015
		Photodecarboxylase (*C. variabilis*)	Karava, 2021
CotZ	16		
		Laccase (BpuL)	Bartels, 2017
		Laccase (EcoL)	Bartels, 2017
		Lipase A (*B. subtilis*)	Hui, 2022
		Lipase B (*B. subtilis*)	Hui, 2022
		Beta-galactosidase	Wang, 2015
		D-psicose 3-epimerase (*C. scindens*)	He, 2016
OxdD	43.39		
		Beta-glucuronidase	Potot, 2010

The motivation for this work came from our curiosity and belief that other spore coat proteins may also be amenable to the spore-display of enzymes. We believe that by testing additional anchors, we may expand the toolbox for spore-display and increase the probability of finding a suitable spore coat protein when aiming to use spore-display for catalysis or any other application. Section 2.3 is an expanded version of the manuscript we are preparing based on our work to identify new

enzyme anchors.

2.3 Advancing protein immobilization on bacterial spores through an extensive survey of coat proteins.

2.3.1 A survey of known spore coat proteins

Spore-display is finding increasing use in a variety of applications such as cell-free metabolic engineering, enzymatic biocatalysis, vaccine delivery, environmental remediation, and even directed evolution of enzymes. Spores are inherently highly robust to thermal, pH, proteinase, and solvent stresses, and they often confer some of these properties to anchored proteins as well. Their unique structure is responsible for their extreme durability: their cytoplasm is dehydrated and highly viscous and their membrane is protected by a multi-layered protein shell. Further, they are relatively unique among biological surface-display systems (e.g., yeast- or bacterial-display) since they are assembled within the cytoplasm of a mother cell prior to release, and thus, proteins do not need to be secreted to be loaded and surface available. Finally, spores derived from Bacillus subtilis can be rapidly engineered since the microbe is naturally competent and well-studied. The protein shell is called the spore coat and is highly organized, having four distinct layers known as the basement layer, inner coat, outer coat, and crust. Importantly, the spore coat is known to be porous, allowing small molecules to diffuse in and out but precluding larger macromolecules like lysozyme from interacting with the more fragile layers beneath. This porosity is an essential part of the spore structure, as it allows for germination

receptors in the lower layers of the spore coat access to molecules like amino acids that can signal for germination (Peter T. Mckenney, Driks, and Eichenberger 2013). Despite its promise, the potential of spore-display has not yet been fully realized. Specifically, spore-display lacks a robust set of well-characterized biological parts like other microbial surface display technologies. To date, publications on spore-display of enzymes have tested only a small subset of 1-6 spore coat proteins as fusion anchors. And while there are more than 60 known spore coat proteins (Table 2.2), only 11 have ever been used for spore-display of enzymes (CgeA, CotB, CotC, CotE, CotG, CotV, CotW, CotX, CotY, CotZ, and OxdD).

Table 2.2: Coat Protein and Spore-Display of Enzymes

Coat Protein	**Previously used as an Enzyme Anchor: N or C Fusions**	**Location**	**Regulates or Affects Incorporation of Others**	**Mutations Negatively Affect Sporulation Efficiency**	**Mutations Affect Structure or Function (germination, lysozyme sensitivity)**
CgeA	C	Crust			
CmpA					Yes
CotA		Outer Coat			
CotB	Both	Outer Coat			

CotC	Both	Outer Coat			
CotD		Outer Coat			
CotE	C	Outer Coat	Yes		
CotF		Inner Coat			
CotG	Both	Outer Coat			
CotH		Inner Coat	Yes		Yes
CotI		Outer Coat			
CotJA					
CotJB					
CotJC					
CotK					
CotM		Outer Coat			
CotN			Yes		
CotO			Yes		Yes
CotQ					
CotS		Outer Coat			

CotT		Inner Coat			Yes
CotU		Outer Coat			
CotV	C	Crust			
CotW		Crust			
CotX	C	Crust			
CotY	C	Crust			
CotZ	C	Crust			
CwIJ		Inner Coat			
GerQ		Inner Coat			
LipC		Basement			
OxdD	C	Inner Coat			
SafA		Inner Coat			Yes
SpoIVA		Basement	Yes		
SpoVID		Basement	Yes		
SpoVIF					Yes
SscA					Yes
Tgl		Inner and Outer Coat			

SleL		Inner Coat			
CotNE		Outer Membrane			
CotNH		Outer Membrane			
PdaB		Basement			
YeeK		Inner Coat			
YhaX		Basement			
YhbB					
YheD		Basement			
YhjR		Inner Coat			
YhxC					
YknT		Outer Coat			
YmaG		Inner Coat			
YppG		Basement			
YqfQ					
YrkC					

YsnD		Inner Coat			
YtcC					
YtrH				Yes	
YttB					
Ytrl		Membrane		Yes	
YtxO		Outer Coat			
YutH		Inner Coat			
YuzC		Inner Coat			
YxeE		Inner Coat			

2.3.2 Selection of Candidate Spore Coat Proteins and Construct Design

However, no study has systematically compared these 12 coat proteins to each other or to other proteins that are known to make up the spore coat. In this study, we sought to thoroughly characterize the potential of *B. subtilis* spore coat proteins as fusion partners for spore-displayed enzymes. In our selection of spore coat proteins to test as new enzyme anchors, we considered one important quality: do mutations to or deletions of the candidate spore coat protein negatively affect spore stability? If so, the protein was excluded from our list of final proteins.

We briefly considered excluding all spore coat proteins with known

Table 2.3: Spore Coat Protein and their Documented Enzyme Activity

Spore Coat Protein	Documented Enzyme Activity
CotA	Laccase
CotQ	Oxidoreductase
CwIJ	Cell Wall Hydrolase
LipC	Phospholipase
OxdD	Decarboxylase
SodA	Superoxidase Dismutase
Tgl	Transglutaminase
SleL	N-acetylglucosaminidase
YabG	Cysteine Protease
YbaN	Polysaccharide Deacetylase
YhbB	Amidase

enzyme activity (Table 2.3), because of our desire to make systems with minimal metabolic background activity but elected to keep those that had previously been used as anchors such as OxdD. Finally, due to DNA synthesis constraints, we additionally narrowed our list down to 44 unique spore coat proteins, prioritizing those that had previously been isolated from protein extracts of spores or were shown to localize to the spore through fluorescent protein fusion. With our 44 spore coat protein candidate anchors identified, we designed 88 DNA constructs that would enable integration and expression of spore coat protein (SCP) fusions to the Beta-glucuronidase enzyme of Escherichia coli (GUS). These fusions are hereafter referred to as SCP-GUS. The architecture of the DNA constructs is like that used in many surface-display systems with four unique parts:

1. Anchor Protein: The coding DNA sequence of each SCP under control of its native promoter.

2. Cargo Protein: The GUS coding DNA sequence without its stop codon at either the N or C-terminus of the SCP.

3. Linker Sequence: A peptide linker with a sequence of (SGGGG)X3-SSGG between the SCP and the GUS sequence.

4. Epitope Tag: a 6x-His Tag sequence on the GUS terminus opposite of the SCP.

To prevent competition between the native copy of the spore coat protein, we designed our constructs to integrate at the native locus of each SCP – enabling allelic exchange with the SCP-GUS fusion. We relied on the homologous recombination machinery of *B. subtilis* to integrate at each locus, enabled by the addition of 2 kbp of homology at the GUS insertion site to direct recombination. Accordingly, each SCP has its own unique homology arms that were also included in the construct design. For modularity, our constructs also included unique restriction enzyme sites at the junctions between the homology arms and epitope tag, the epitope tag and the GUS CDS, and the GUS CDS and the linker sequence, to enable swapping of each part of the construct (Figure 2.4). An example sequence of an SCP-GUS construct and a list of those constructs that deviated from this exact design can be found in the appendix to this chapter.

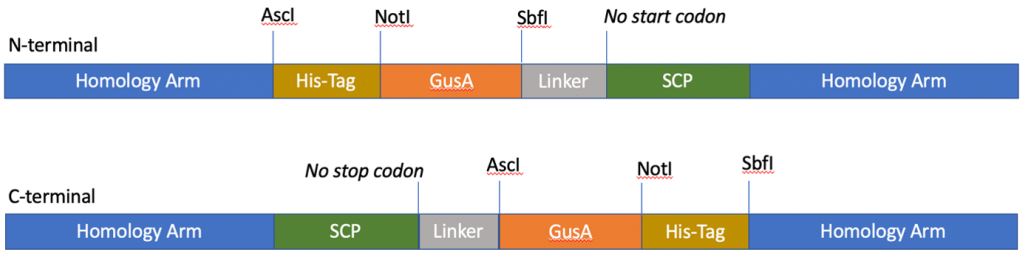


Figure 2.4: SCP-GUS Construct Design for Integration at Native Loci: 88 constructs representing N and C-terminal fusions to each of the 44 candidate SCPs were designed according to this standard with some modifications as outlined below.

2.3.3 Construct Synthesis

Synthesis of these constructs was performed by the Department of Energy Joint Genome Institute (DOE JGI). Of these, only 52 could be synthesized and verified, representing 33 of the original 44 spore coat proteins (Table 2.4, Figure 2.5).

Table 2.4: Synthesized Constructs

52 Synthesized Constructs	
CgeA-C-GusA-His	CotY-N-GusA-His
CgeA-N-GusA-His	EcsB-C-GusA-His
CotA-C-GusA-His	LipC-N-GusA-His
CotA-N-GusA-His	OxdD-C-GusA-His
CotB-C-GusA-His	OxdD-N-GusA-His
CotB-N-GusA-His	PdaB-C-GusA-His
CotC-C-GusA-His	PdaB-N-GusA-His
CotC-N-GusA-His	PrtG-C-GusA-His
CotD-C-GusA-His	PrtG-N-GusA-His
CotG-C-GusA-His	SafA-N-GusA-His
CotG-N-GusA-His	SleL-C-GusA-His
CotI-C-GusA-His	SpaD-C-GusA-His
CotI-N-GusA-His	SpaD-N-GusA-His
CotJA-C-GusA-His	SscA-C-GusA-His
CotM-C-GusA-His	SscA-N-GusA-His
CotQ-N-GusA-His	YhaX-C-GusA-His
CotS-C-GusA-His	YhaX-N-GusA-His
CotS-N-GusA-His	YhjR-N-GusA-His
CotU-C-GusA-His	YhxC-N-GusA-His
CotU-N-GusA-His	YmaG-N-GusA-His
CotV-C-GusA-His	YqfQ-C-GusA-His
CotV-N-GusA-His	YrkC-C-GusA-His
CotW-C-GusA-His	YrkC-N-GusA-His
CotW-N-GusA-His	YtcC-C-GusA-His
CotX-C-GusA-His	YtcC-N-GusA-His
CotY-C-GusA-His	YtrH-N-GusA-His

2.3.4 Co-transformation enables genome-wide editing of spore component genes

Traditional spore display approaches rely on construction of merodiploid strains where two copies of spore coat components are expressed – an unmodified copy from its native locus and a second recombinant copy from a neutral locus with the cargo protein fusion (Iwanicki et al. 2014; Bartels et al. 2017).To avoid any potential competition between two copies during spore assembly, we focused on constructing strains where the native copy of the spore coat protein is translationally fused to an enzyme reporter. We designed 88 fusion constructs to integrate GUS (*E. coli* β-glucuronidase) at either the amino- or carboxyl- terminus of each of 44 spore coat protein at their native locus 2.4. Each construct was ≈5.8 kb long – consisting of 2 kb homology arms at either end of the 1.8 kb GUS gene. We chose GUS as a reporter since it is a large homotetrameric enzyme (68.5 kDa Œ 4 = 260 kDa) and is representative of display partners that may be used for biocatalytic applications. Although there are >50 known spore coat protein components, we excluded proteins that have been documented to affect spore formation and function significantly and adversely when modified.

To create recombinant strains, we exploited the phenomenon of congression, whereby naturally competent Bacillus subtilis cells uptake and integrate multiple DNA fragments into their chromosome, at low frequency, in addition to an unlinked selectable marker. Our collection of 52 constructs, designed without selectable markers, represented an

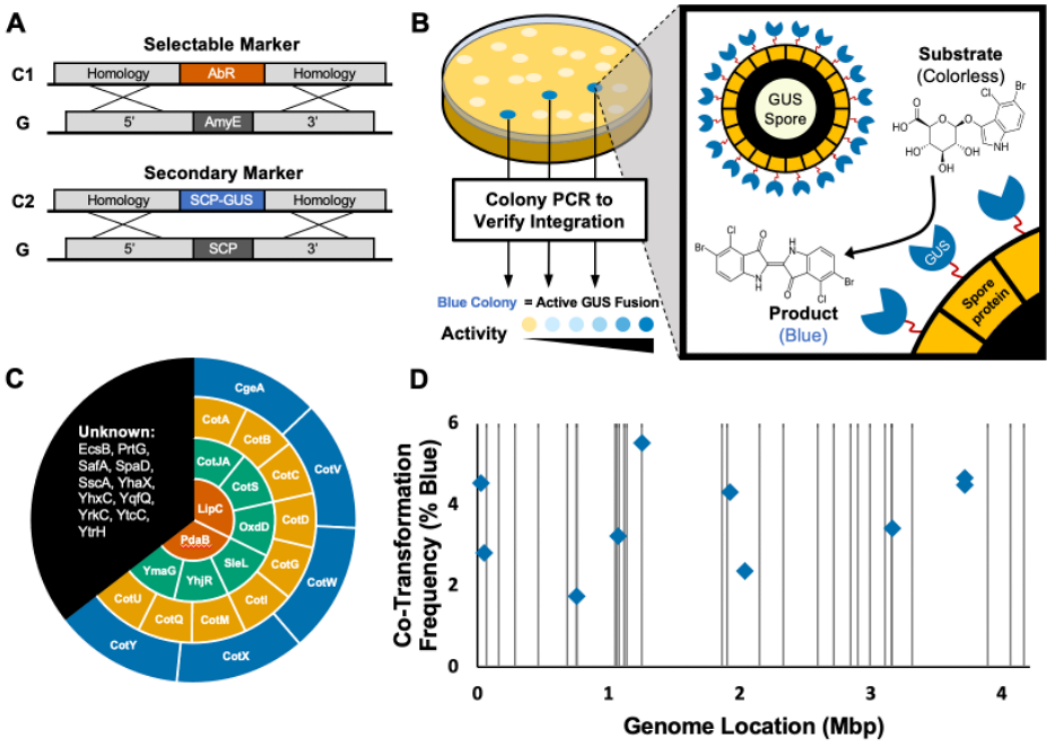


Figure 2.5: Co-Transformation for non-competitive spore-display: (A) The first cassette (C1) carries a selectable marker with homology arms for integration at the non-essential amyE locus. The second cassette (C2) carries an N- or C-terminal fusion of GUS to the spore coat protein (SCP) with 2 kb homology arms for integration at their respectively native loci. (B) On-plate sporulation and blue/white screen to identify GUS+ colonies. (C) 44 SCPs selected for screening as fusion anchors, including proteins from all layers of the spore coat crust (blue), outer coat (yellow), inner coat (green), and cortex (orange) and unknown location (black). (D) Initial co-transformation assessment showed co-transformation frequencies between 1.5 6% for the 10 test loci (blue diamonds). The distribution of the 34 other SCP loci are shown as vertical lines.

unprecedented opportunity to study congression. We first tested our co-selection methods for allelic exchange at the native SCP loci with a subset of 10 constructs spread across the genome (Figure 1.D). Accordingly, we co-transformed competent *B. subtilis* 168 cells with two linearized integrative plasmids: the first carrying an antibiotic marker and homology arms for the amyE site and the second carrying the SCP-GUS constructs with homology arms for the unique SCP loci. Antibiotic selection for the

primary marker on sporulation medium supplemented with 5-bromo-4-chloro-3-indolyl-beta-D-glucuronic acid (X-gluc) enabled identification of GUS+ clones through the formation of blue colonies between 24-72 hours after plating and incubating at 37 °C (Figures 2.6, 2.5). Notably,

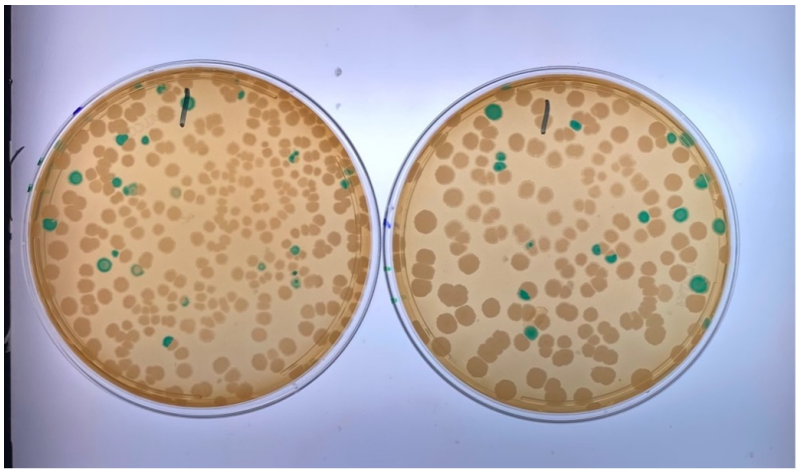


Figure 2.6: Sample Co-Transformation Results: Colonies from the co-transformation of pBS1K + CgeA-C-GUS and plating on 2xSG media + X-Gluc. Blue colonies are considered to be GUS+ and underwent colony purification and colony PCR to confirm integration

co-transformation frequency was reliably between 1.5 – 6% across all 10 test loci (Fig. 2.5). To us, this was a reasonable threshold to move forward with given co-transformation removes the need for additional unique constructs for locus editing like gRNA cassettes for CRISPR. We then moved to all 52 constructs and found that 37 yielded blue colonies upon co-transformation. If a SCP-GUS construct did not immediately yield blue colonies, we continued to do co-transformations and screened a minimum of 1,000 transformants across at least 3 independent co-transformations. The remaining 15 constructs were presumed to form non-expressive or non-functional GUS fusions. Of the 37 SCP-GUS constructs that did yield blue colonies, three could not be confirmed as having the integration at the correct locus by colony PCR and were

subsequently excluded from further analyses. The 34 confirmed strains represent 26 unique spore coat proteins across all four layers of the spore coat including eight with unknown localization. 16 among this set have not previously been used before for spore-display of enzymes. Notably, only 10 of the 34 constructs are N-terminal fusions. It may be that insertion of GUS at the N-terminus of the SCP CDS negatively affects transcription and prevented expression of the fusion constructs. While

Table 2.5: 34 cPCR Confirmed Strains

34 cPCR Confirmed Strains	
CgeA-C-GusA-His	CotX-C-GusA-His
CgeA-N-GusA-His	CotY-C-GusA-His
CotA-C-GusA-His	CotY-N-GusA-His
CotB-C-GusA-His	EcsB-C-GusA-His
CotC-C-GusA-His	OxdD-C-GusA-His
CotC-N-GusA-His	OxdD-N-GusA-His
CotG-C-GusA-His	PdaB-C-GusA-His
CotG-N-GusA-His	PdaB-N-GusA-His
CotI-C-GusA-His	PrtG-C-GusA-His
CotI-N-GusA-His	SpaD-C-GusA-His
CotJA-C-GusA-His	SscA-C-GusA-His
CotM-C-GusA-His	YhaX-C-GusA-His
CotS-C-GusA-His	YhjR-N-GusA-His
CotU-C-GusA-His	YqfQ-C-GusA-His
CotV-C-GusA-His	YrkC-C-GusA-His
CotW-C-GusA-His	YtcC-C-GusA-His
CotW-N-GusA-His	YtrH-N-GusA-His

the blue/white on-plate assay confirmed that the fusion proteins are enzymatically active, it does not indicate whether they can incorporate into spores. For the 34 correctly integrated strains, we assessed the biocatalytic proficiency of GUS+ spores using a colorimetric assay with p-Nitrophenyl-β-D-glucuronide (NPG) as substrate. We found that all but 2 of the 34 strains formed enzyme-functionalized spores (Figure 2.7). Mean activity levels varied across 4 orders of magnitude, ranging

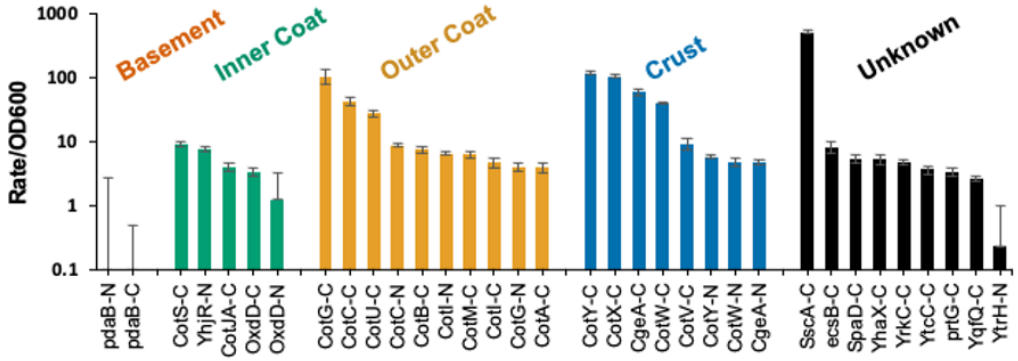


Figure 2.7: GUS Activity of Spore-Displayed Enzymes: Spore-displayed GUS activity across all 34 GUS+ strains. Only fusions to PdaB have no discernable activity over WT spores. Rate = Abs(405nm)/min

from <1 to >500 Rate/OD600. Basement fusions to PdaB were generally inactive whereas all inner coat proteins and most of those with unknown localization yielded low activity. Most of the highest activity fusion partners were located within the outer coat and crust of the spore, which is generally consistent with the literature. Specifically, CotY, CotG, CotX, and CgeA were 4 of the 5 best performing fusion partners, which is consistent with their general popularity as fusion partners in the published literature as highlighted in Table 2.1. Interestingly, of all the tested constructs, the most active spores were those with GUS fused to SscA-C, a small spore coat protein (28 amino acids and ≈3 kDa) with unknown location that had never been used for spore-display before. The SscA-GUS spores had an average Rate/OD600 = 516. Almost 5x more than the next best performer, CotY-C, with an average Rate/OD600 = 115. Structure prediction using AlphaFold2 revealed SscA forms a simple helical fold 2.8. The small size of SscA suggests that it would be minimally disruptive to assembly and activity of homotetrameric GUS and potentially other cargo proteins as well.

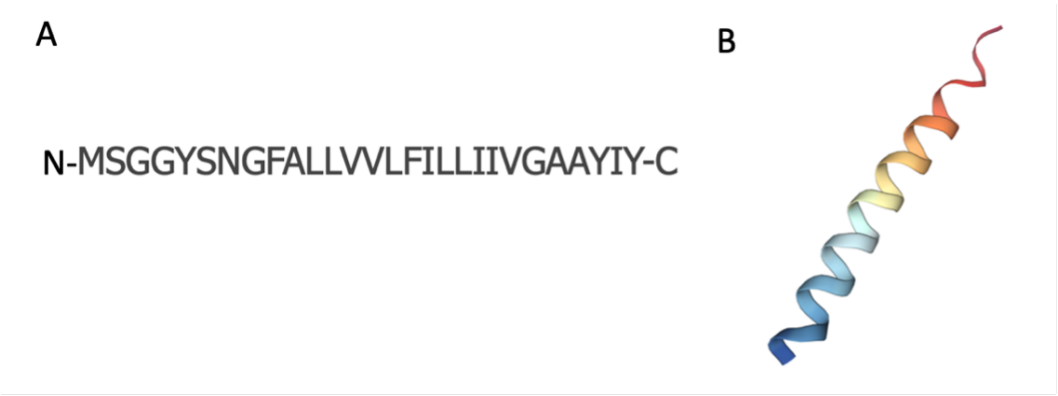


Figure 2.8: Sequence and AlphaFold-predicted structure of SscA

2.3.5 Several anchors support surface-display

We also characterized the spores of all 34 strains for the surface availability of GUS through immunostaining, flow cytometry, and microscopy. Using existing methods for spore purification, which often include a lysozyme purification step, we found that lysozyme treatment often produced small clumps of spores and protein aggregates. Clumps of multiple spores, some phase bright and some phase dark, and cell debris would form and following immunostaining would be highly fluorescent when checked with microscopy. All strains formed clumps, but not all were fluorescent. Some strains had extremely fluorescent clumps while their individual spores were not fluorescently labeled at all.

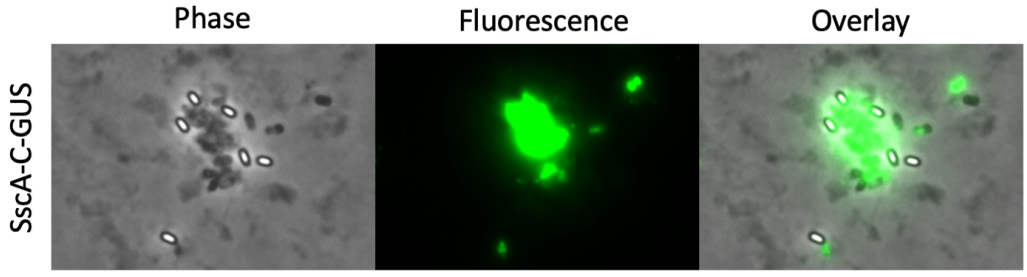


Figure 2.9: Clumping of spores following protease treatment: Phase-bright spores (white ovals with black outline) are not individually stained following immunostaining with anti-HIS antibody. Rather, fluorescence is localized to clumps of cell-debris.

This was especially true for SscA-C which had bright clumps of cell-debris and phase dark spores but little to no observable labeling on phase-bright spores (Figure 2.9). When measured with the flow cytometer, SscA-C appeared to have high levels of surface-labeling (data not shown). No labeling on phase bright spores but a strong fluorescent signal on the flow cytometer meant something in our methods was not working as planned. This mismatch meant two things were likely going wrong.

(1) Spores and cell debris were forming clumps. Some of these clumps either contained SCP-GUS fusions that were not localized to the spore during spore formation or were sticky enough to cause non-specific antibody binding during immunostaining despite blocking and washing steps performed during the immunostaining protocol. These clumps were giving false-positive fluorescence signals for some strains that had no apparent fluorescence when checked with microscopy. (2) The gating method we were using with our flow cytometry data was not precise enough to discriminate between individual spores and these clumps.

During our troubleshooting, we also realized that the FSC-H threshold on our flow cytometer was set high enough that by default it excluded many of the unique events observed during sample collection on spore samples. We adjusted the instruments FSC-H threshold setting (Lowering it from 5,000 to 1,000) to no longer exclude these events. Accordingly, we switched to a different spore purification method in which we grew larger 25 mL cultures of spores and relied on differential centrifugation to create a layered pellet of spores and the remaining vegetative cells which could be separated without lysozyme treatment

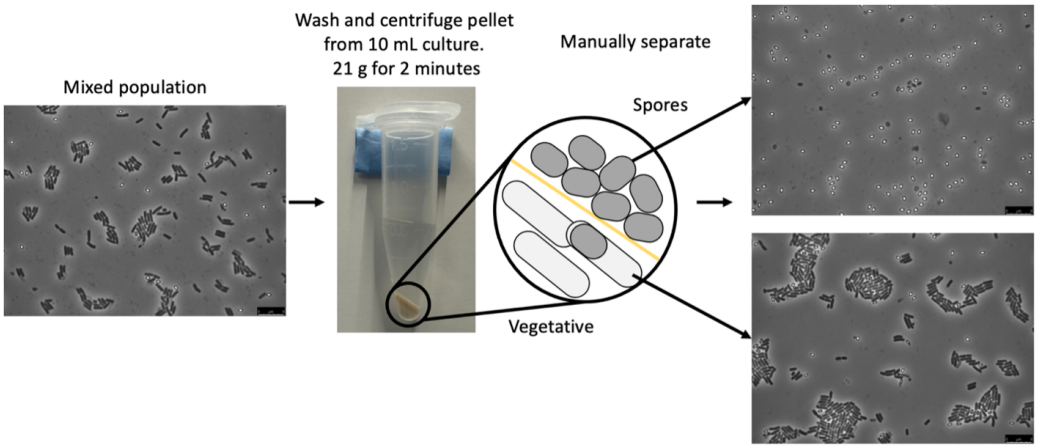


Figure 2.10: Spore Separation Method: Spores grown in cultures >10mL can be separated from vegetative cells through a centrifugal layering method detailed here. Cultures are grown and sporulated then pelleted. The pellet is moved to a 1.5 mL Eppendorf tube and washed twice. After a final centrifugation at 21k g for 2 minutes, a layered pellet forms. Spores reside in the upper layer and can be removed by gentle pipetting

(Figure 2.10). This prevented the formation of spore-debris aggregates. We also purchased flow cytometer calibration beads (ThermoFisher, Catalog#: F13838) and used them to create a gate based on the FSC-H signal of 1 and 2 μm beads (Figure 2.11). To verify our gates functionality, we used WT spores and spores carrying a copy of SscA fused to the fluorescent protein eYFP. eYFP is activated by the same fluorescent laser as our secondary antibody for the anti-HIS protocol and served as a positive control for spore fluorescence alongside the WT spores. With these two changes made, we found that we could readily isolate a population of both WT non-fluorescent spores and SscA-C-eYFP fluorescent spores, confirmed by microscopy (Figure 2.12). And, when we ran strains of both spores on the flow cytometer, they had similar FSC-H profiles which overlapped well with the FSC-H gate we had created using the calibrations beads. Finally, when we applied the

FSC-H gate to each of the spore populations, it excluded many of the non-fluorescent events that were present in the SscA-C-eYFP population (Figure 2.11). This was a lesson in instrument calibration, back testing, and iteration but we ultimately identified a suitably robust method for measuring individual spore fluorescence via flow cytometry. Cumulatively, these data gave us confidence in our ability to measure and quantify spore fluorescence via flow cytometry and we resumed immunostaining. With these new methods, we first tested the surface availability of the terminal 6x-His tag present in each construct. We were surprised at how few strains stained above the background WT signal with only three SCP anchors, CotY-C, CotX-C, and YtcC-C, having a fluorescence signal at least 2x > WT when evaluated by flow cytometry (Figures 2.13, 2.14). This was somewhat consistent with a previous study that showed only the crust proteins CgeA, CotY, and CotZ formed functional surface-available fusions to GFP in a panel of 18 SCPs (Imamura et al. 2011). However, considering that several other spore coat proteins had previously been used for surface display of antigens we decided to use a polyclonal antibody for GUS to try another method of surface detection in case conformational constraints occluded His6-presentation. This switch led to detection of SCP-GUS fusions on the surface of many more recombinant spores. With the anti-GUS staining, 11 SCP anchors exceeded a mean fluorescence signal of the WT by at least 2-fold: CotY-C, OxdD-N, CotX-C, CotY-N, CotC-C, CotC-N, EcsB-C, CotV-C, CotW-C, CotS-C, and CotI-C (Figure 2.13). Importantly, the polyclonal immunostaining with GUS yielded results more consistent with previously established surface availability

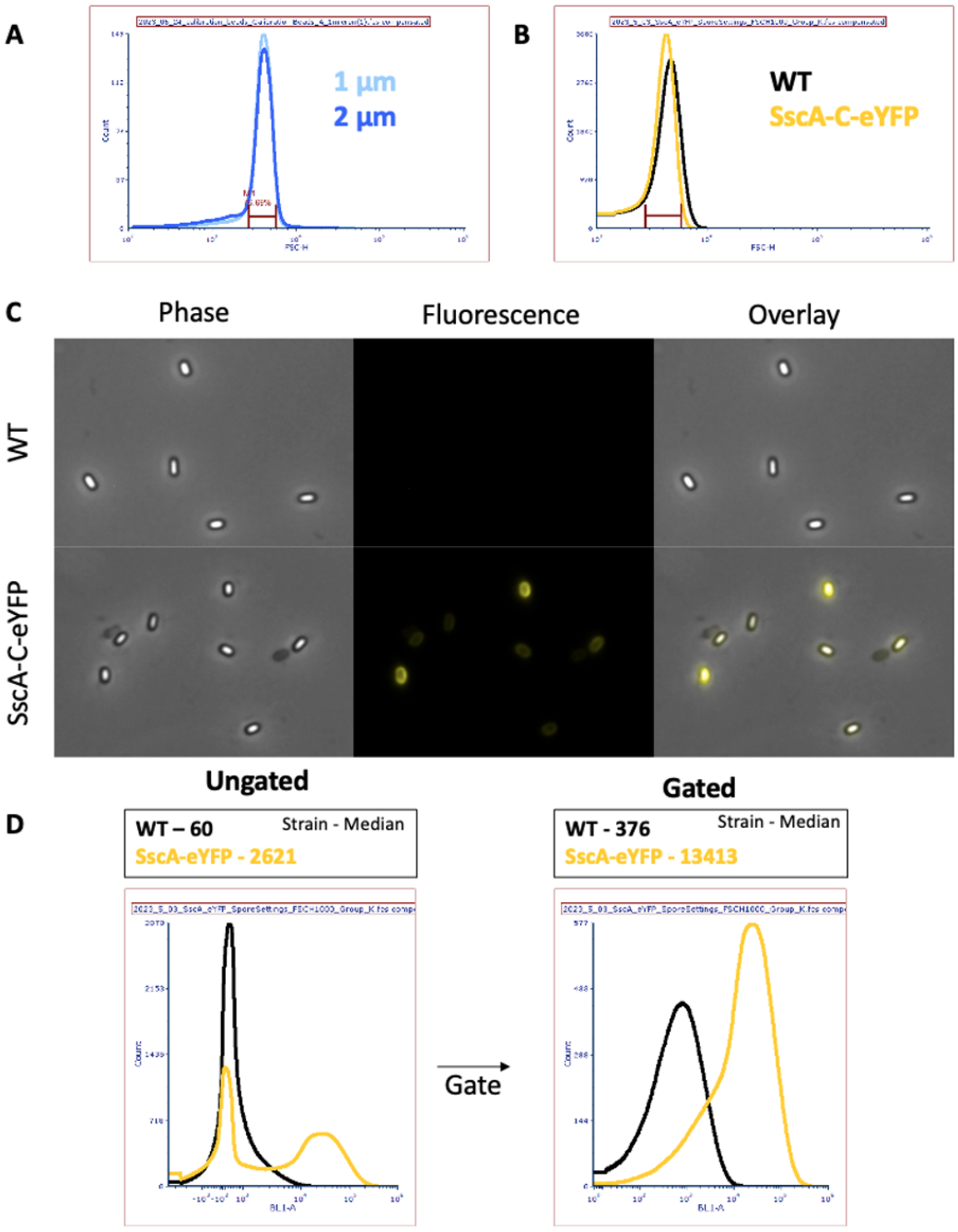


Figure 2.11: (A) FSC-H histograms of the 1 μm and 2 μm beads essentially overlapped. We created a gate (maroon) based on the peak of these distributions that encompassed ≈75% of all events between the two bead sizes. (B) FSC-H profiles of WT and SscA-C-eYFP spores purified through the method shown in Figure 10. The same gate overlapped with most of the events from both WT and SscA-C-eYFP spore populations. (C) Fluorescence microscopy images of the purified WT and SscA-C-eYFP spores show no clumping and loading of SscA-C-eYFP to the spore. (D) Applying the FSC-H gate to the WT and SscA-C-eYFP population yields two distinct peaks: a WT peak with a mean value of 376 and a SscA-eYFP peak with a mean value of 13413

of SCPs including GUS fused to OxdD-N, and other antigens fused to CotC (Potot et al. 2010; Isticato 2023).

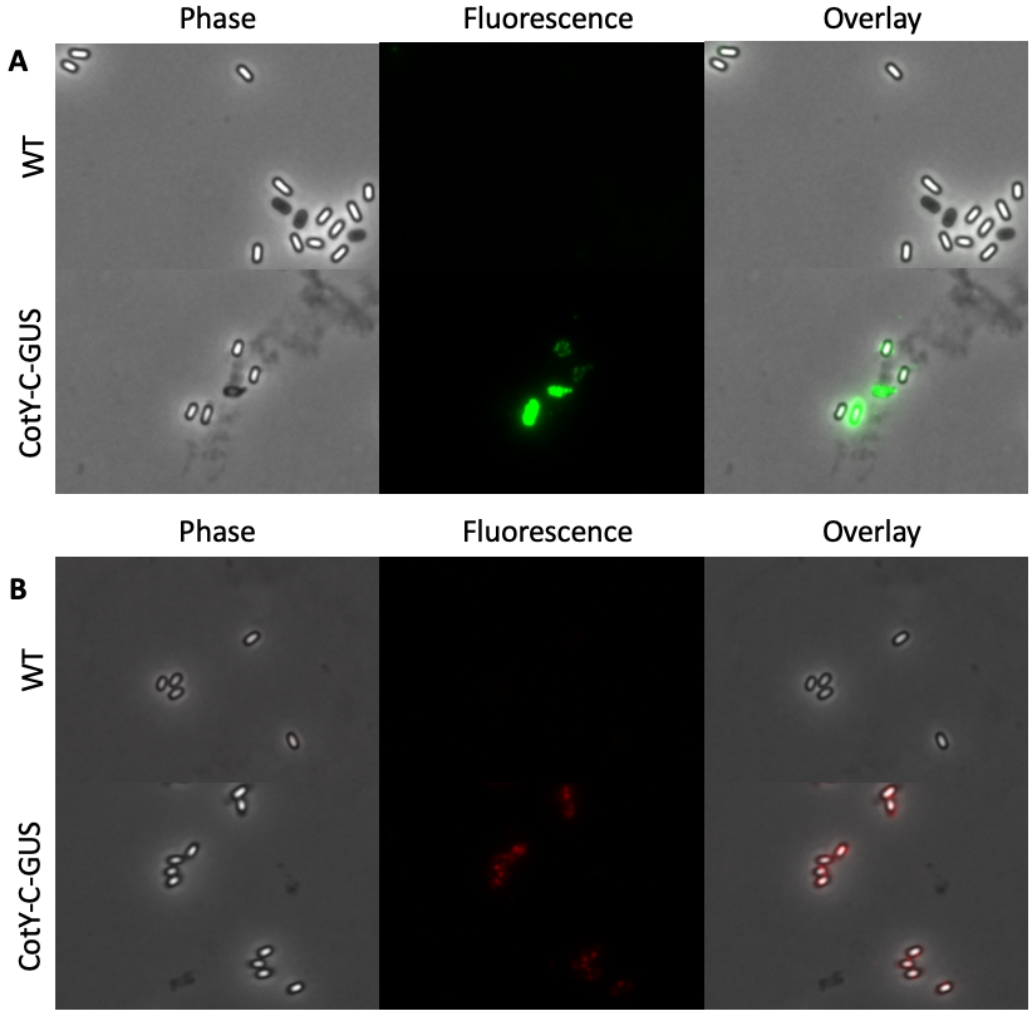


Figure 2.12: A) WT spores of *B. subtilis* 168 (top) and 168 + CotY-C-GUS (bottom) immunostained with an anti-6x-HIS antibody pair. (B) WT spores of *B. subtilis* 168 (top) and 168 + CotY-C-GUS (bottom) immunostained with a polyclonal anti-GUS antibody.

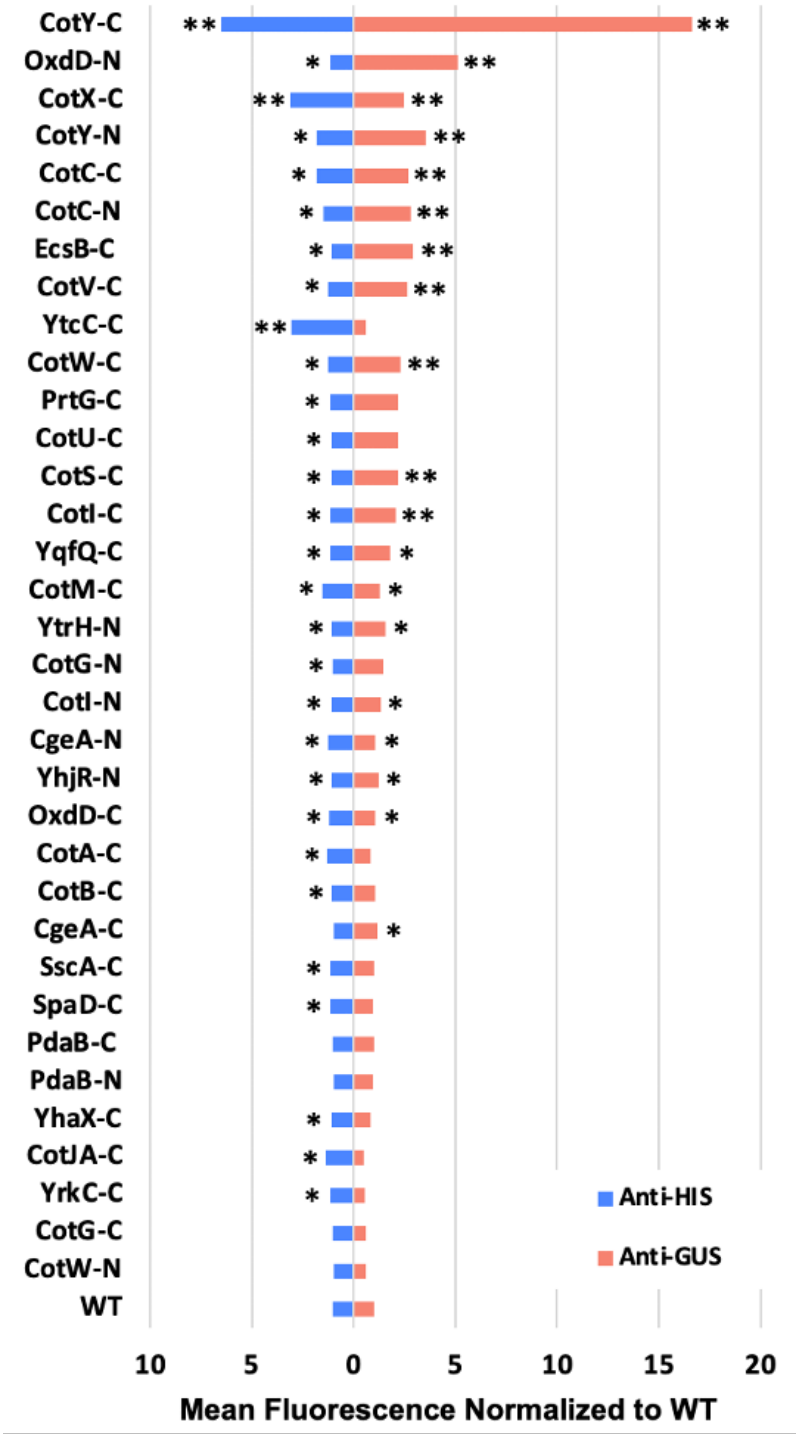


Figure 2.13: Immunostaining using two different antibodies (Anti-His and Anti-GUS) yielded different rankings of spore coat proteins, presumably due to differences in epitope availability. "*" indicates a mean population fluorescence significantly greater than WT, $p < 0.05$ (Krusal-Wallis one-way ANOVA). Large N (30,000 – 120,000 gated events) led to statistically significant differences that are unlikely to be functionally relevant. Accordingly, we have also labeled all strains that have a mean fluorescent signal at least 2x > WT with "**". Population distributions are in Figure 14

Notably, this is also the first time that the protein EcsB has been used for spore display. Interestingly, EcsB belongs to a three-cistron operon that encodes proteins that are like an ATPase and are thought to form an ABC transporter complex in the membrane of vegetative cells and play a role in secretion (**heinri**). EcsB, specifically, encodes a hydrophobic trans-membrane protein (Leskelä et al. 1999). The spore coat is known to have hydrophobic qualities, so it may be that EcsB also co-localizes to the spore coat during sporulation. In future spore-display efforts, it may be considered as a potential anchor protein for use when preparing vaccines due to the surface availability of GUS as a cargo. For example, if we look at the population distributions of the antibody-labeled spores 2.14 we can see that not all the strains have a log-normal population. CotY-N appears to have a small population of highly-labeled spores that are driving up its mean when treated with anti-HIS antibodies. In parallel, EcsB-C appears to have a small percentage of spore with high fluorescence when labeled with anti-GUS. In these populations, it is possible that there are differences in spore morphology that lead to these fusions being surface-available on some spores but not others. When engineering spores for applications where surface-display greatly matters, e.g., vaccines, it would be wise to test anchors with high-surface availability such as CotY and characterize the recombinant spores using flow cytometry as has been done previously (Isticato et al. 2001)

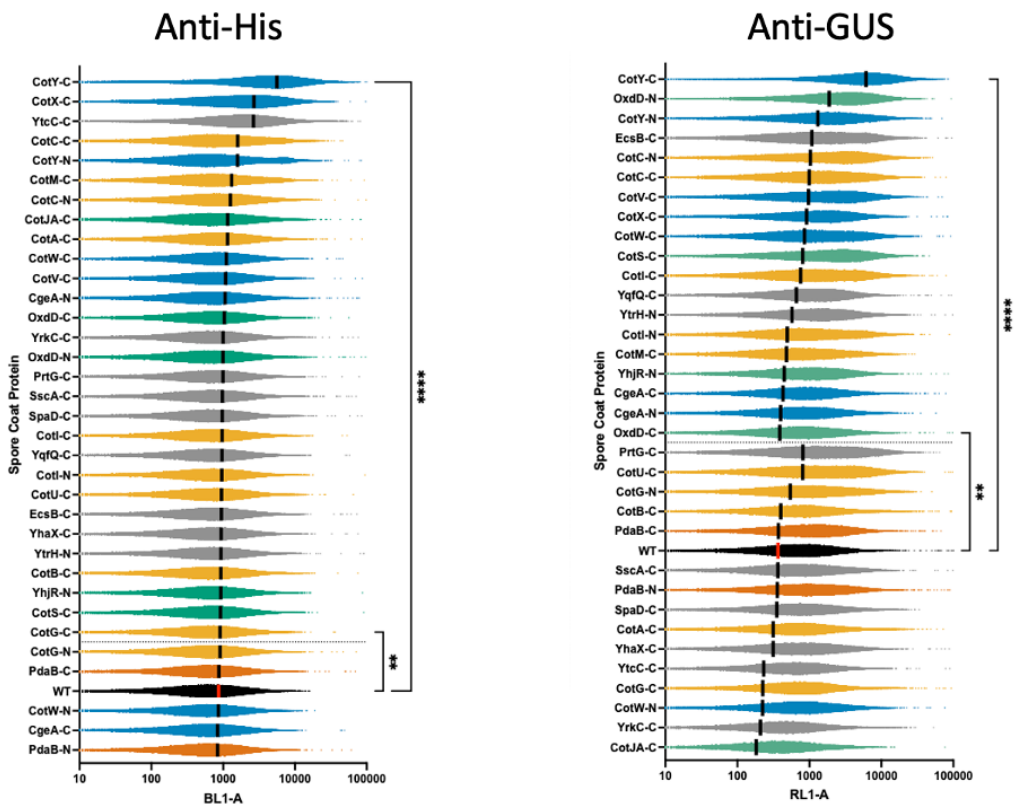


Figure 2.14: Gated Population histograms for each of the 34 strains and WT spores: Each strain is color-coded based on its location within the spore coat: blue = crust, yellow = outer coat, green = inner coat, orange = basement, and grey = unknown location. All strains above the dotted line have a mean average fluorescence that is significantly greater than WT spores (Kruskal-Wallis One-way ANOVA

2.3.6 SCP-fused GUS are stabilized against an inactivating protease

A motivating factor for this study is the ability of spore-display to increase the stability of cargo proteins. Previous studies have documented improved enzyme tolerance to thermal, protease, and solvent challenges. Importantly, the spore coat is a porous structure that precludes larger macromolecules like lysozyme from approaching the more fragile peptidoglycan and membrane layers beneath while still allowing small molecules to diffuse in and out so the dormant cell can sense when its environment is amenable to growth again. We tested the ability of

four anchors (CotG-C, CotW-C, CotX-C, and CotY-C) to protect GUS from 0.5 mg/mL proteinase-K. Samples were incubated for 30 minutes and their residual enzyme activity measured. Notably, free-enzyme and CotW-C-GUS lost 100% of enzyme activity while CotG-C, CotX-C, and CotY-C retained between 45 and 70% activity (Fig. 2.15). Others in our lab have now shown that of the 15 most active anchors, all provide some level of protection in thermal and protease challenges. This is detailed in our upcoming paper. We found that strains with significant GUS surface availability generally saw a decrease in total activity after a protease challenge whereas non-surface available strains such as SscA-C-GUS retained near-complete activity.

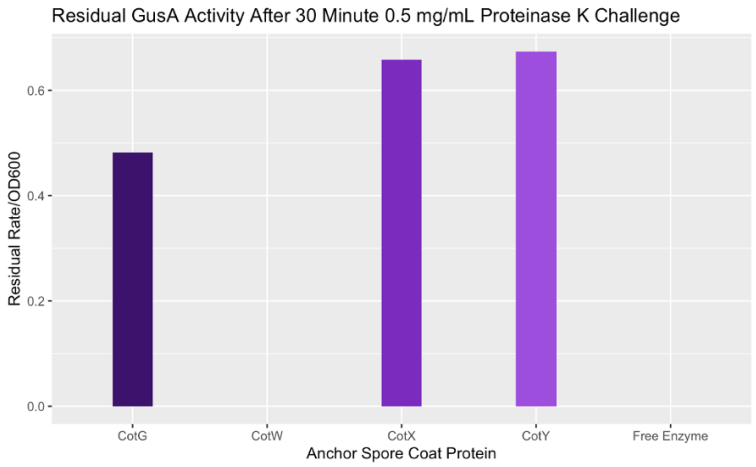


Figure 2.15: Spore-Display Protects Cargo from Proteinases: Shown are the residual activity of GUS fusions to CotG-C, CotW-C, CotX-C, CotY-C, and free-enzyme after being subjected to a proteinase challenge for 30 minutes. Values are the average of biological duplicates

2.3.7 Extending Spore-Display to other enzymes

Our direct comparison of 33 spore coat proteins as anchors for spore-display of enzymes represents the largest and most thorough characterization of spore-display to date. From our characterization of these

52 GUS constructs, we next wanted to test if we could extend spore-display to other enzymes with our new anchor, SscA, and see how it performs against other high-performing anchors. We chose the top five most active anchors CgeA-C, CotG-C, CotX-C, CotY-C, and SscA-C and fused them all to a new enzyme: BsCel5 from *B. subtilis*. BsCel5 is a non-processive endoglucanase, meaning it randomly cleaves cellulose polymers into smaller, low molecular weight oligomers (Fig. 2.16). We forwent integrating these fusions at the native locus of each SCP and instead created constructs bearing the same linker and his-tag structure as the original GUS constructs that would instead integrate at the non-essential amyE site. The amyE site encodes an amylase gene and this locus is commonly used in strain-engineering in *B. subtilis* because loss of AmyE activity can be easily screened for with iodine-stained starch plates.

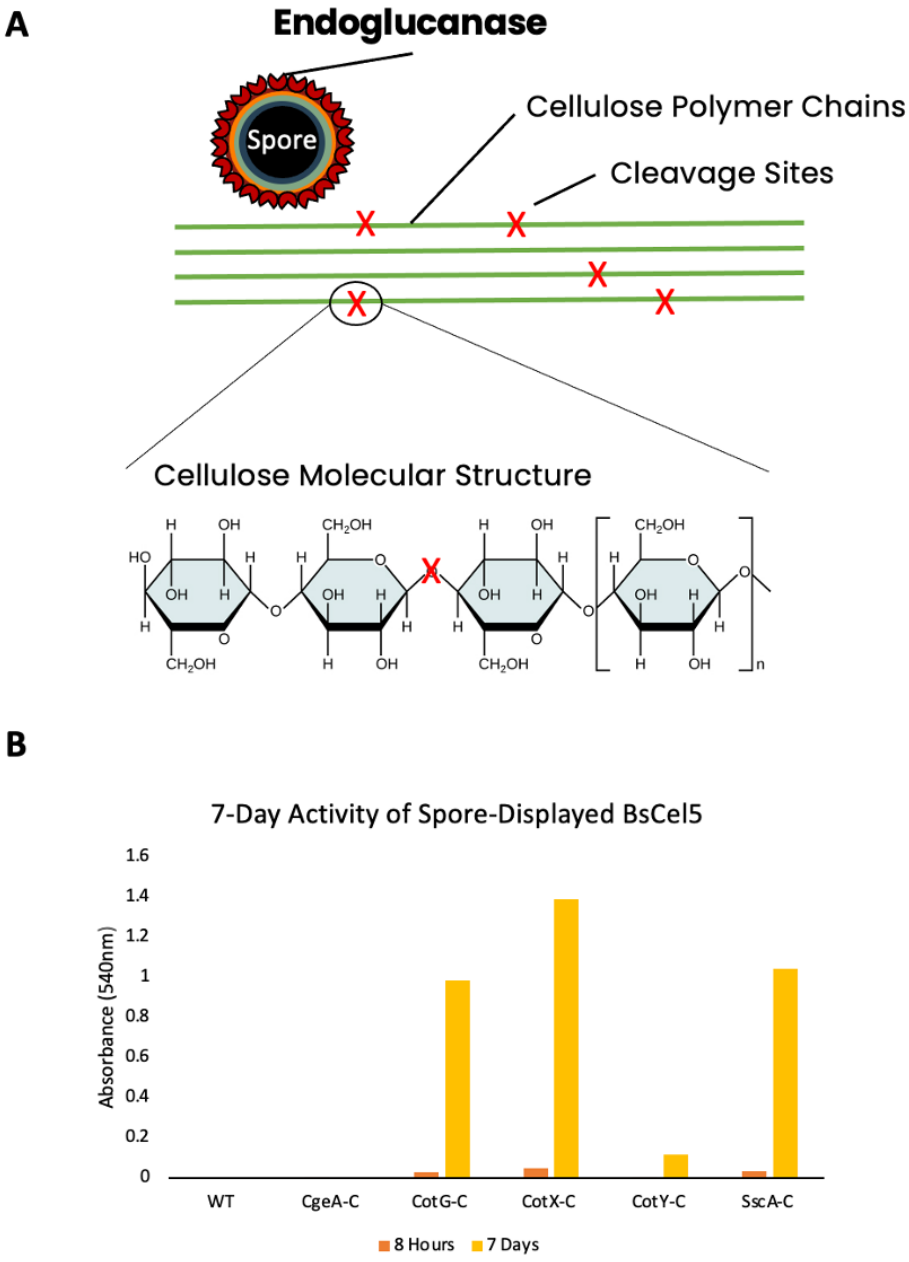


Figure 2.16: (A) Diagram for the enzymatic activity of spore-displayed BsCel5, an endoglucanase that randomly cleaves long cellulose polymers into smaller pieces. (B) Five SCP-BsCel5 fusions were tested after at 8 hours and 7 days of incubation with carboxymethylcellulose. Values are the mean of two biological replicates.

We tested the activity of spore-displayed BsCel5 on carboxymethylcellulose (CMC). We found that four of the five fusions had some level

of activity with BsCel5 (Fig. 2.16). CgeA-C had no detectable cellulase activity. CotY-C did not perform well in comparison to CotG, CotX, and SscA. These mixed results – CgeA having no activity and CotY having comparatively less activity with BsCel5 than with GUS – demonstrate that there is still a need to test a set of candidate anchor proteins when developing spore-displayed biocatalysts to find the best possible solution.

2.3.8 Discussion

The unique production, structure, and stability of bacterial endospores positions them as a useful tool for biotechnologists with applications ranging from vaccines to catalysis and, more recently, even as a genetic device to track products as they move through the global supply chain (Isticato 2023); aanikabio.com. When thinking about spore-display, we find it useful to conceive of spores as genetically encoded particles that could be modified to carry anything that can be genetically encoded or added via chemistry. When designing spore-display systems, like in other microbial surface-display technologies, it is important to understand what anchor proteins can be used to functionalize spores through the fusion of cargo proteins to members of the spore coat. With the large number of known spore coat proteins, and only a handful having been previously fused to enzymes, we aimed characterize which SCPs could be added to the synthetic biology toolbox for spore-display. Importantly, there are many ways in which spore-display could fail to yield a functionalized spore: (1) the SCP-cargo protein fusion could fail to express or fold properly and yield a non-functional protein, (2) the

cargo protein could preclude the anchor from loading properly to the spore, (3) a functional protein could load to the spore in a functional state but be positioned in way that it is not accessible to its substrate or binding target, or (4) the correctly loaded protein could be too unstable for the necessary reaction or binding conditions. We purposefully designed our spore-characterization constructs so we could assess spore-display modes of failure for each SCP. Using GUS as a model enzyme, with its multiple substrates, enabled us to use simple screens for correct GUS translation on plates and in simple assays with purified spores. Additionally, it is emblematic of many industrially relevant enzymes as its active form is a 260 kDa homotetramer. We tested a total of 52 unique SCP-GUS fusions covering 33 unique SCP. Of these, we confirmed spore-display of enzymes with 10 previously used SCP and identified 16 novel anchor proteins. Of great interest in this set of novel anchors is SscA (Small spore coat protein A) which yielded the most activity for any single fusion enzyme partner. With the little we do know about SscA from previous studies, its high activity and limited surface availability does make sense. In an analysis of mRNA within the spore, it was the only spore coat protein in the top 46 of most abundant spore mRNA suggesting it is highly expressed (Korza, 2019). Additionally, CotB, CotG, and CotH were significantly reduced in ΔsscA spores suggesting it may play a role in assembly of the outer coat (Odama, 2011). We have shown that SscA can serve as a functional anchor for GUS, eYFP, and BsCel5. Its small size makes it more akin to a signal sequence or epitope tag than a large fusion protein anchor like CotY and other spore coat proteins that are more than twice its size. This suggests

it may serve as a robust anchor for future work in spore-display. We also tested the ability of 3 surface-available proteins to protect GUS from protease activity while still being surface-available. Our results suggest that spores carrying enzymes or other proteins on their surface can provide a level of protection against proteolytic activity and speaks to the potential for spores to act as potential therapeutics. For example, could spores carrying enzymes or nanobodies one day play a role in the delivery of biologics to the gut, where protease resistance while surface-displayed would be highly advantageous. Or could the enzyme phenylalanine ammonia lyase be delivered to the gut for the treatment of phenylketonuria via spore-display? Previous work has shown that the residency time of spores in the gut can be increased through the spore-display of bacterial adhesins. Combinations of adhesins, therapeutic proteins, or other combinator spore-display approaches may lead to powerful and resilient therapeutics and catalysts in the future. Based on activity alone, we defined a set top five anchor proteins and tested their ability to act as anchors for an additional enzyme: BsCel5. The comparative activity of the top 5 anchor proteins differed between the two enzymes. With GUS, the top 5 ranking was SscA > CotY > CotX > CotG > CgeA while for BsCel 5 it was CotX > CotG > SscA > CgeA > CotY. that chose to test for their ability to stabilize GUS against thermal and protease challenges. In my opinion, this strategy of testing a few potential SCP anchors with a new enzyme is likely to continue to be necessary when aiming to build a high-performing spore-based vaccine, catalyst, or spore-display system. In the future, when selecting anchor proteins for spore-display, the data presented here can

serve as a starting point for selecting and testing a subset of anchor proteins, decreasing the intensity of the design-build-test cycle needed to create a spore-displayed protein. Finally, the strategy we developed for evaluating spore coat proteins could be easily applied to sporulating bacteria beyond *B. subtilis* that may have unique features that would benefit society in its search for biotechnology solutions in healthcare, sustainability, and more.

2.4 Developing SporeCatcher

With all surface-display technologies, there is risk that fusion and expression of cargo proteins in recombinant hosts could prevent proper folding and cause reduced or even complete loss of enzyme activity. There are strategies for avoiding this issue including codon optimization, testing sets of anchor proteins to find a functional fusion, and testing different amino acid linkers between the fusion partners (Bartels et al. 2017; Ullah et al. 2017). These prototyping strategies can lead to functional spore-display systems to great effect as evidenced in our identification of SscA-C-GUS as a highly-functional anchor-cargo fusion.

An interesting approach to spore-display was developed by Xin Ge's group at the University of California Riverside (Chen, Mulchandani, and Ge 2017). Taking inspiration from extracellular enzyme complexes called cellulosomes, they fused the cohesin protein (CtCoh) of Clostridium thermocellum to spore coat protein CotG. This resulted in surface-available cohesin. Cohesin is one half of the two-part cohesin-dockerin

system. Once the cohesin+ spores were produced, they could be incubated with dockerin protein which would then strongly bind the spores. By making enzyme-dockerin fusions in other expression hosts, they could stoichiometrically load cargo enzymes to the spores. They demonstrated cohesin+ spores could be functionalized with two enzymes at once with xylose reductase and phosphite dehydrogenase. This enabled the reduction of xylose to xylitol while regenerating the cofactor NADPH for continued use. There could be many instances where having the ability to selectively load proteins to spores after they have been made like this could be advantageous. This system is useful because it could enable spore-display of proteins after they are produced in another host that may be more amenable to their expression. Thus, the cohesin-dockerin system could make spore-display simpler and more tunable while also increasing the number of potential cargo proteins.

In other applications, there is risk associated with the escape of recombinant genetic material into the environment. Recombinant genetic material can provide new functions to organisms, which could have unintended and potentially disastrous consequences (e.g., the spread of antibiotic resistance genes). Thus, for increased safety, proteins are often used after they have been purified away from the genetic material that encodes them (e.g., in a cell-free system). However, in such methods, protein instability leads to challenges for both protein storage and delivery. Spore display is known to confer stability to loaded proteins, increasing their storage lifetime and their robustness in the face of stability challenges (e.g., pH extremes, temperature swings, solvents, and detergents). Spore display also simplifies the protein purification

process because spores can be easily isolated and concentrated by centrifugation and/or filtration. The cohesin-dockerin spores could be used to produce agenetic spores which could safely deliver cargo without the risk of horizontal gene transfer. By trans-loading cargo proteins produced by another cell onto the spore surface, once could eliminate the risk associated with the escape of potentially harmful recombinant genetic material. However, while cohesin-dockerin bind each other very tightly they do not bind covalently (Stahl et al. 2012). Consequently, this system could have limited utility where stability of the spore-cargo complex is paramount.

We see a great deal of potential in the idea of trans-loading cargo proteins to form agenetic spores. We wanted to engineer a covalent trans-loading system that could instill the highest amount of stability possible by loading the agenetic cargo proteins with covalent bonds. For start, we turned to SpyCatcher-SpyTag. The SpyCatcher-SpyTag system was developed at the University of Oxford by Mark Howarths group. SpyCatcher-SpyTag was built by splitting the fibronectin-binding protein FbaB of Streptococcus pyogenes (Zakeri et al. 2012). The protein natively form an intrachain isopeptide bond between Asp and Lys residues. By splitting FbaB into two smaller fragments they created SpyCatcher and SpyTag which, when in solution together, will bind each other and form a covalent bond ≈10x stronger than that made by cohesin-dockerin (≈1 nanonewton versus ≈120 piconewtons) (Zakeri et al. 2012; Stahl et al. 2012). Since its inception, SpyCatcher-SpyTag has become an extremely important biotechnology tool. Notably, the

Howarth group has continued to refine SpyCatcher-SpyTag using various strategies, including a phage-display library, which have today led to SpyCatcher003-SpyTag003 which rapidly find and bind each other even at low concentrations (Keeble et al. 2019).

This system has successfully been used to covalently decorate the surface of phage and cells (Kellmann et al. 2023). We aimed to engineer spores to display SpyCatcher so we could trans-load SpyTag fusions to the spore as cargo (Figure 2.17).

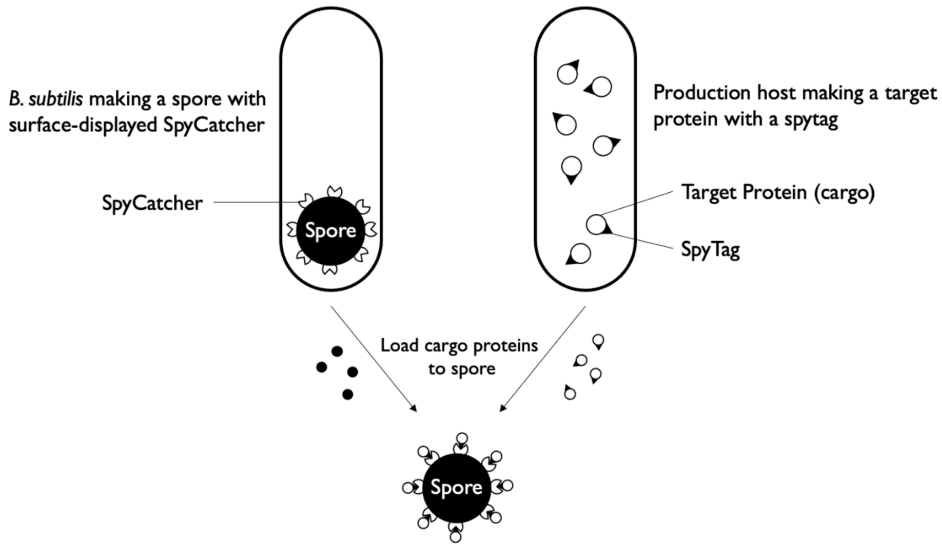


Figure 2.17: Proposed method for Covalent Trans-Loading of Spores as Agenetic Protein Carriers using SpyCatcher-SpyTag

To first test this, we built strains of *B. subtilis* carrying fusions of the engineered SpyCatcher variant SpyCatcher003 (SC3) on the N-terminus of spore coat proteins CotY and CotZ. We integrated these fusion proteins at the amyE site and placed them under control of their native promoters using the suite of sporovectors developed by the Mascher group (Bartels et al. 2017). These constructs were sequence-confirmed

before transformation into *B. subtilis* and integration at amyE was confirmed by loss of amylase activity on starch plates. Spores were made, purified, and then incubated with a SpyTagged red fluorescent protein made in *E. coli* (SpyTag003-mKate2) and then were observed by fluorescent microscopy. We did not see any significant loading of SpyTag003-mKate2 to the CotY-N-SC3 or CotZ-N-SC3 spores (data not shown). To test if the N termini of CotY and CotZ can be fused to SpyCatcher as a cargo proteins that is then surface-available upon spore formation (a requirement for our SpyCatcher-SpyTag system to work) we made fusions that combine the SCP, eYFP, and SpyCatcher simultaneously. Both CotY-N-eYFP-SC3 and CotZ-N-eYFP-SC3 spores could form fluorescent spores (Figure 2.18).

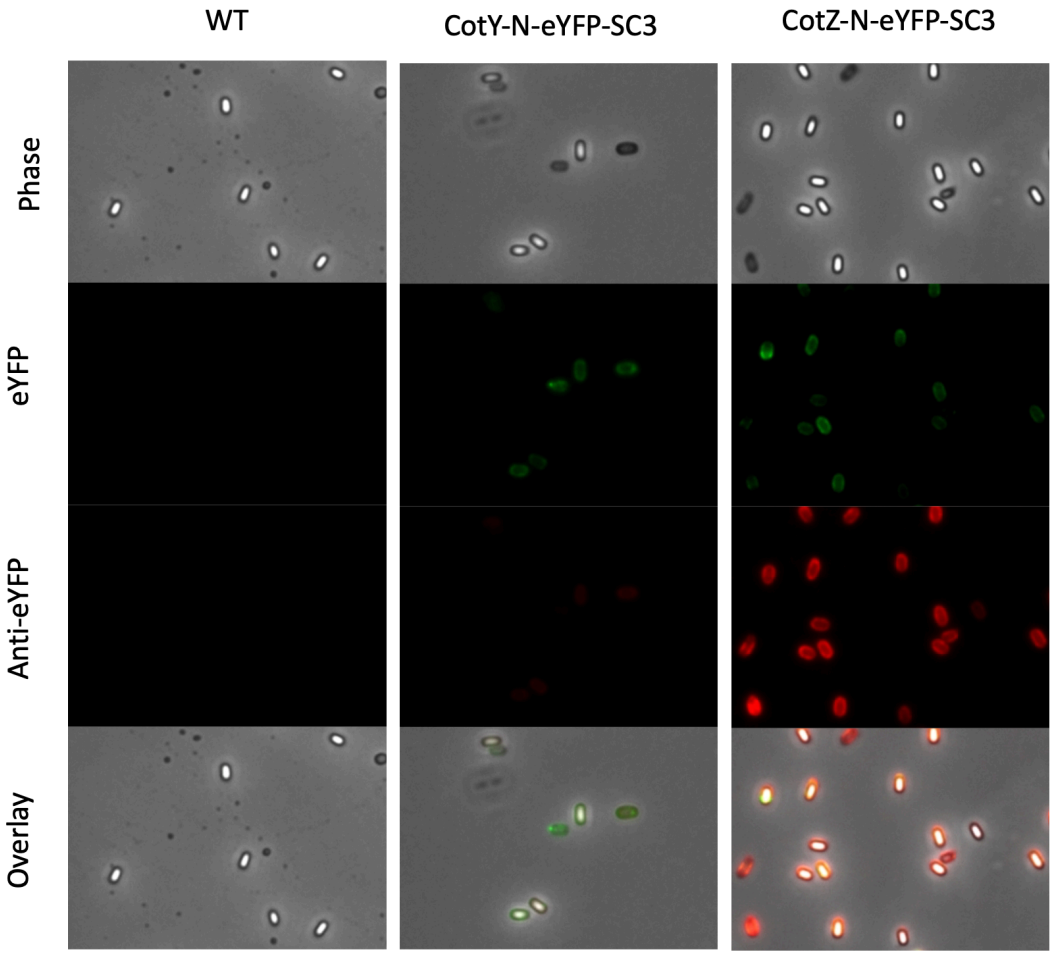


Figure 2.18: Surface Availability of CotY and CotZ SC003-eYFP Fusions: Images of Wildtype (WT), CotY-N-eYFP-SC3, and CotY-Z-eYFP-SC3 spores. Phase images show phase bright and phase dark spores. eYFP images show expression and incorporation of the eYFP into the spore coat (green), immunostaining with anti-eYFP antibody shows labeling of the eYFP positive spores (red) indicating that eYFP is surfac availabile, overaly shows both signals are coming from both phase bright and phase dark spores.

When we incubated these spores with SpyTag003-mKate2 we still didn't see any noticeable red fluorescence. This could have been because either (1) the SpyCatcher003 unit isn't surface avilable on these spores and/or (2) it is mis-folded or truncated in some way. To check for surface-avilability we used an anti-eYFP antibody to check for the

middle third of our fusion protein. We founds lableing of both CotY-N-eYFP-SC3 and CotZ-N-eYFP-SC3 spores with the anti-eYFP antibody. This suggests that on these spores SpyCatcher003 may be mis-folded or truncated in some way. This is an active area of investigation for our lab. In future work, we will be testing additional spore coat anchor proteins for SpyCatcher and adding a terminal epitope tag to the eYFP-SC3 constructs to test for full translation of the fusion protein.

2.4.1 Future Perspective on Spore-Display

Self-Assembling, genetically-encoded, evolvable immobilized systems.

Enzymes are powerful, sustainable catalysts. Directed evolution and computer-aided enzyme design have led to increases in enzyme utility through improvements in catalytic rate, substrate profiles, and stability. The value of these enzymes can be further increased through adoption of immobilization technologies to create heterogenous biocatalysts – increasing enzyme stability, extending enzyme lifetime, and enabling enzyme recycling. This was first demonstrated at an industrial level with glucose isomerase to produce high-fructose corn syrup in the 1950s (Di Cosimo, 2013). However, existing abiotic immobilization methods like gel matrices or magnetic beads require additional purification and processing steps after enzyme production and have relatively small, low-throughput design spaces. Spore-displayed enzymes can also act as heterogenous biocatalysts and spore-display has the potential to serve as a platform technology for creating self-assembling, evolvable, immobilized enzyme systems. In their 2018 review, Bernal, et al. note that enzyme immobilization and enzyme engineering have

traditionally been disparate disciplines. Immobilization engineers focus on support materials and cross-linking chemistries with little ability to understand how an enzymes sequence and structure might affect its ability to be immobilized and retain its activity. Meanwhile, enzyme engineers almost exclusively engineer enzymes in their soluble forms or in temporary systems like yeast-display that, more often than not, do not resemble the environment they will need to operate in at scale. Consequently, improvements in free enzyme activity do not always translate to improvements in performance of immobilized catalysts (Bernal, Rodríguez, and Martínez 2018). Spore-display presents an opportunity to truly unite these two fields into one discipline. The Farinas group has demonstrated directed evolution of CotA on *B. subtilis* spores (Gupta and Farinas 2010). I see no reason why spore-display could not be extended to the directed evolution of recombinant proteins with the number of potential anchor proteins now available. Because spores are genetically encoded and self-assembling, millions of immobilized protein variants could theoretically be screened using spore-display to identify catalysts with improved activity. Moreover, the spores could be evolved alongside the enzyme to improve their functionality as enzyme carriers; enhancing important qualities like enzyme loading or surface qualities to maximize substrate interactions. With advancements in microfluidics and cell-sorting technologies, and numerous precedents set by other library technologies, establishing spore-display as a high-throughput engineering platform appears feasible. Enormous sequence-structure-function data sets could be generated through the combined application of technologies like deep sequencing, cryo-electron tomography, and

cell-sorting. In comparison to surface-display technologies like phage, *E. coli*, or yeast, spores can maintain their genotype-phenotype link in industrially relevant extremes of pH, heat, and solvents. This could enable rapid exploration of new areas of protein sequence space that were previously limited to low-throughput methods, enabling the generation of enormous amounts of data as we repeatedly push enzymes to their breaking point in the search for more stable and long-lasting catalysts to enable the sustainable chemical economy.

Spore-displayed biocatalysts could then be used in packed-bed reactors common to traditional immobilized enzyme systems or, because of their small size and low cost of production, could enable other reactor designs like bubble columns or continuous stirred tank reactors. For example, others in the Nair lab have now shown that spore-displayed esterases can degrade plastic films. If we could further improve the activity of these enzmes and demonstrate recyclability with bioprocess innovations, spore-displayed enzymes could one day be used to recycle plastics or enable other industrial reactions at scale.

2.4.2 Methods & Materials

Bacterial strains and culture conditions

B. subtilis 168 (ΔtrpC) and its derivatives were used for all experiments unless noted otherwise. Strains were regularly cultured in Luria Bertani (LB) media in aerobic conditions using 24 deepwell plates, test tubes, or shake flasks; temperature set to 37 °C with shaking at 225 rpm. Antibiotics used for selection were Chloramphenicol (5 μg ml-1) and Kanamycin (10 μg ml-1). For solid agar plates, 1.5% w/v agar was

added.

Preparation of Naturally Competent Cells & Transformations

Competent cells were prepared as described by (Bennallack et al. 2014)with slight modifications. Overnight cultures of 4 mL in plain LB were grown in test tubes for 14 hours at 37 °C at 225 rpm. Subsequently, a volume of overnight culture was transferred to a volume of pre-warmed SM1 media to an OD600 of 0.5 and a final volume of 15 mL in a 125 mL shake flask. This culture was then incubated at 37 °C and 225 rpm until it reached an OD600 of 0.5-0.6, ≈4 hours. 15 mL of pre-warmed SM2 media was then added and the culture incubated for an additional 90 minutes. Cells were then portioned into 0.5 mL aliquots and used for transformation or, after the addition of glycerol to 10% v/v, frozen to -80 °C for storage up to 60 days. Single transformations were regularly done with 100 ng of linearized vector DNA prepared from Escherichia coli mini-preps and co-transformation with 100 ng of the selectable construct and 1.5 μg of the SCP-GUS construct.

Spore Coat Protein Beta-Glucuronidase Fusion Construct Design

Constructs for making fusions of Spore Coat Proteins to GUS (SCP-GUS) were designed as follows. For N-terminal fusions of GUS to SCPs, 2 Kb of genomic homology were amplified from the +3-nucleotide position within the coding DNA sequence (CDS) of the SCP to insert the GUS fusion at the native locus. The upstream homology arm was amplified with a C-terminal side primer overhang to introduce a flanking AscI cut site between the homology arm and a 6xHis-Tag sequence.

The 6xHis-Tag sequence preceded a 3x(4GS) linker sequence flanked by a NotI site and the GUS sequence. The GUS sequence was flanked by a coding SbfI site and 3x(4GS) linker sequence (no stop codon) that joined its downstream homology arm which includes the SCP CDS and additional locus homology. For C-terminal constructs, 2 Kb of genomic homology were amplified from the nucleotide preceding the stop codon in each SCP. A 3x(4GS) linker was added to the SCP CDS which was flanked by an AscI site preceding the GUS CDS. The GUS CDS was flanked by an NotI site, and 3x(4GS) linker to 6xHis-Tag sequence followed by a SbfI site and the downstream homology arm 2.19. For sequences which contained AscI, NotI, or SbfI sites within the CDS or homology arms, other restriction enzyme cut sites were used on a case-by-case basis as detailed in the appendix. These constructs were then cloned into the pENTR vector backbone and propagated in *E. coli* through culturing in LB + Kan (50 μg ml-1).

Co-Transformation for in-situ Expression of SCP-GUS Fusions

Freshly prepared competent cells were portioned into 0.5 mL aliquots within 15 mL test tubes or 24 deepwell plates to which 100 ng of linearized selection construct (pBS1K or PBS1C) and 3 μg of SmaI-linearized pENTR-SCP-GUS construct was added and gently mixed by pipetting followed by incubation at 37 °C and 225 rpm for 1 hour. Subsequently, 0.5 mL of LB was added to each transformation mixture and incubated for an additional hour. 0.2 mL fractions of the culture were then plated on individual 2xSG sporulation plates containing an antibiotic for the primary marker (pBS1K = 10 μg mL-1 Kan, pBS1C = 5 μg

mL-1 Cm) and X-Gluc (5-bromo-4-chloro-3-indolyl-beta-D-glucuronic acid, 100 μg mL-1). Plates were then incubated at 37 °C for 96 hours, during which they were observed for the formation of blue colonies. Blue colonies were colony purified on 2xSG plates and GUS-SCP integration was confirmed by colony PCR. 2xSG sporulation plates were composed of 1.6% Difco Nutrient Broth, 0.05% MgS04-7H20, 0.2% KCl, 1 mM CaN03, 1 mM MnCl2, 0.1 ţM FeS04, 0.1 % Glucose, and 1.5% w/v agar (Leighton and Doi, 1971). X-Gluc was freshly prepared in a 100 mg mL-1 solution in dimethylformamide which was then added to the autoclaved unset agar media alongside the antibiotic before plates were poured.

Sporulation & Spore Purification

Sporulation was done via exhaustion in Difco Sporulation Media (DSM, (DSM, 0.8% w/v Tryptone, 0.1% w/v KCl, 1 mM MgSO4,10 μM MnCl2,1 μM FeSO4, 0.5 mM CaCl2). For enzyme assays, strains were struck on fresh LB plates from -80 °C stocks and grown overnight at 37 °C, which were then used to inoculate 2 mL of LB in 24 deepwell plates and incubated for 24 hours at 37 °C and 225 rpm. At 24 hours, 100 μL of the LB culture was passaged into 2 mL of DSM also in 24 deepwell plates. Spores were harvested after 24 hours in DSM through centrifugation at 21 g for 2 minutes and washed 3 times with 500 μL water. For flow cytometry and microscopy, strains were struck on fresh LB plates from -80 °C stocks and grown overnight at 37 °C, which were then used to inoculate 2 mL of LB in 15 mL plastic test tubes and gown for 24 hours at at 37 °C and 225 rpm. 500 μL of the LB culture was passaged into

25 mL of DSM in 250 mL flasks and grown for 28 hours. Cultures were ten transferred to 50 mL conical tubes and centrifuged at 2,400g for 10 minutes. The pellet was then moved to a 1.5 mL Eppendorf tube and washed twice in sterile water. After a final centrifugation at 21,000 g for 2 minutes, a layered pellet forms. Spores resided in the upper layer and could be removed by gentle pipetting. Spore purity was assessed by microscopy.

GUS Assay

Spore-displayed GUS activity was quantified using an assay with the substrate 4-Nitrophenyl β-glucopyranoside (NPG). Reactions of 200 μL were composed of 140 μL reaction buffer (0.1% w/v β-mercaptoethanol in PBS, pH = 7), 30 μL of substrate solution (4 mg/mL NPG in PBS, pH = 7), and 30 μL of spores with a final OD600 of 0.1-0.3. Reactions were performed at room temperature in a 96-well flat bottom plate, initiated by the addition of the substrate solution, and the reaction monitored by the increase in absorbance at 405 nm caused by product formation using a SpectraMax M3 plate reader (Molecular Devices, San Jose, CA).

Flow Cytometry

Detection of the Spore-Displayed 6x-His tag and GUS was done using immunohistochemistry and flow cytometry. Strains were plated fresh from -80 °C on LB and incubated overnight at 37 °C before being inoculated into 5 mL of LB in 15 mL aerobic test tubes and incubated for 16 hours at 37 °C and 225 rpm. 1 mL was then subcultured into 25

mL of DSM in 250 mL fluted Erlenmeyer flasks and incubated for 28 hours at 37 °C and 225 before being harvested. Following sporulation, cells were transferred to 50 mL falcon tubes and centrifuged at 2,400g for 10 minutes. The pellet was then transferred to a 1.5 mL Eppendorf and washed 3x with PBS (pH = 7) with centrifugation done at 21,000g for 2 minutes. After the final centrifugation, the upper layer of the cell pellet 2.10 was isolated by gentle pipetting and resuspended in PBSA (2% w/v Bovine Serum Albumin in PBS, pH = 7) at an OD600 = 1 for blocking. Following blocking, 50 uL fractions were pelleted and re-suspended in 200 μL of chilled PBSA containing 1 μg mL-1 primary antibody (Anti-His Tag Antibody, Invitrogen, Catalog # MA1-21315) and incubated for 1 hour. This was followed by four washes with 200 μL of chilled PBSA and incubation with the secondary antibody in 200 μL of PBSA containing 1 μg mL-1 (AlexaFluor 467 goat anti-mouse, Invitrogen, Catalog #: A32728) for 1 hour at 4 °C. Samples were then washed four additional times in chilled 200 μL PBSA before being analyzed using an Attune NxT Flow Cytometer (ThermoFisher, Catalog # A24858). 300,000 events were collected for each sample using the lowest possible collection speed, 12.5 μL min-1, and five rinse cycles were used between each sample to prevent carryover. Data files were exported in .FCS format. Gating for single spore events was based on the FSC-H profile of 1 and 2 micron calibration beads. 2.11. Gating and analysis was done using FCS Express Software (De Novo Software, Pasadena, California).

Cellulase Assay

BsCel5 activity was measured on carboxymethylcellulose (CMC) using a dinitrosalicylic acid assay. Spores were prepared and washed twice in water. Reactions were run in 10 mg/mL CMC in 0.1 M Na-acetate buffer (pH =5). Following the reaction incubation time (8 hours or 7 days) a volume of 300 μL was removed and combined with 1.2 mL DNSA10 reagent (1% Dinitrosalicylic Acid, 0.05% Sodium sulfite, 1% sodium hydroxide, and 0.0001% glucose) in a 1.5 mL Eppendorf tube and incubated in boiling water for 5 minutes. The samples were then cooled for two minutes by sitting at room temperature and then aliquoted into 96-well plates for absorbance measurements. Absorbance was measured at 540 nm.

2.5 Spore-Display Appendix

2.5.1 Unique Changes to SCP-GUS Constructs

Some of the SCP loci contained restriction sites that made them incompatible with the standard modular construct design. Accordingly, some were altered as described below.

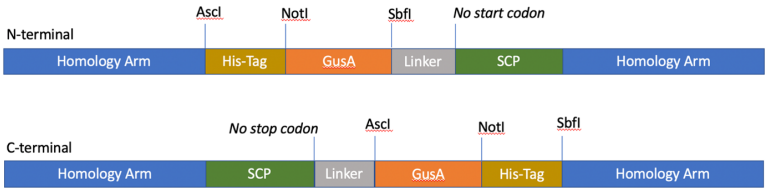


Figure 2.19: Standard SCP-GUS Design

Change	Construct
AscI to PacI	
	CotNH-N
	CotNH-C
NotI to PacI	
	CotK-N
	CotK-C
	SpaD-C
SbfI to PacI	
	CotW-C
	CotX-C
	CotY-C
	CotZ-C
	CotM-N
	CotQ-N
	CotX-N
	CotY-N
	CotZ-N

2.5.2 Sample SCP-GUS Sequence, CgeA-C-GUS

CCTTCCGGAAGCGTCTCCATCTCTTCTTTTCCAATTGTCACT
ATGTCTTTTCTTTGAAATTGATAAAATTGAAACGCATACTG
GTCCACAATTCGCCTTTTTTTTCCGGCTTCGATATAAAATA
TTGCAGGATGGTCAGCTTCCTTTACTAAAATATAATTCGGA
ATGTTTAAAATTTTGCCAATCCTATATTGGAATAAATATGG
GTCTGGAACAGCGACAGTCCGATTTAGATACAAAAAACCCC
ACTTTTCTCCAACAGGATGCCTTACACCCTGATCAAAAAAA

TACACATTTTGATCTGAACCTTTTACAAAAGAGCCGTCAAT
CAAATCGGCGGGCAGAAACCGATATGCATTTTGAAATGAAT
CTGGGGTTTTATAGGTTGTGTCTAATACTTCATTTATCGGC
AGAAATGGGGTAAAGTTGTTGAGTCTCGCCCAAAAAATTGC
GTCGCCATGATTCCAATGCTTCATATCATCATCCCAGTAGC
CTCCAAATTTATGGTGAATGAGATTAAGTAACGAACGGCGA
TGCATAACTGAGCAATGATCAACTTGAAAAGCGGCCTGATC
AAGCACTGCATTGGCATTTCTGTAGAAATGTGAGATTTCCT
CTCCCCTTTCGTTTACATGTACAACTTTTTGTTTTGAATAGA
CCGCCTGCGCTTCAGGCTTATGTGAAAACTCTTGCACCATA
CGGCTCAGCCGCTCAGGATGATAAACAGTGTCGTCAGTAAG
ATAGGAAATGTAATCGCCGTCTGCAAATGGGAGCGCACTGT
TAATTAATGTGGCATAACGGGCTGTTTTTAATCGATCTGCA
GGATGCACGAAACTGTTATGATATTGAATGCGCCGGTCATG
GAGATACTTGTGAATCACCGCAGTTGTTTCTGCATTCGAAT
GATCATCCATGATAAAAAGCTCCCAAAGATCATGCGTTTGC
TGTATCACGCTTTCGATCGCTTTTTGCAAATAATCGGGTTT
GTTGTAACTTGTTAAAATAATAGACACTTTCTCTCCCATCT
ATACCGCCTCCCGCCGTCAAACATAAAAACGGCTCCGCAGC
AAGCAGAGCCTCTGTCATCATTTAAAAAGCACCCCAGCTTA
CAACACTTGAGAGTGAAACATGAGATCTCGTGTTATTTAAC
GGGTTTTGGAAAATAACCAAATCAGTCAGAACTTCTTTTAC
GATGACATTTTGAAGTGTATCTCCGTTGATCAAATTGACTT
TTGAAACCGGTACGCCCAGCTCTTTTGCAAAAACAAGCTGG
CGTCTGAGGTCAGATGGATTTCCTTCAGAAGACGATTGACG
TATCAGCAGAATCAGCAAAAGCAAAATAAAATGCAGATCAT

CTTGGTTTTCTTCTCGATCCACTTGTTTCTCTAAATCCAAAT

CATCCCTCTCCCTTATCCTTATCTCTGTTTCACAATATGTGC

GACTTCGATTTATGGTATAGGCTATGTCCACTCCATTAAAA

AATCAGCATTTTCCTTAAATAGATGAAAAATAGATCAGGTA

CGGCGTTCGACTCATACCAAATAACAGCCGGAAGAATATGA

ATAACGTGAGTTCACTATCGAATAGGAGGTGTGTGTAATGA

GCTCTGAAAATGCACAGTTAAAAAAGGATTTAATTAAAGCC

GTTTTAAGCCCTTTATTTCCAACTGCAACAGAAGGAGGAGA

AAATATGGATAGTAATCTTAAAGCCTTGCTTGATGCTGCCA

TCGATCAAAAAGTAGATGAAAGTGAAACGGTTACGGCAGAA

TCTATTTTAGACCCTTCTCTTCCGGCAAGATGGATTTTTGC

CAGAATTACGCCAGGCACTACAATCAGCATTGTGACTGATT

CAGGTGACATGATCGGACCGGTTGTTTTCGTTGCTTTCGAT

CAGGTTCACGGGATCGTATTTGTAACACAGGAAAGCTCCGT

CACTCCGGCAGGCCAAGCTACAACATTAATTGATGTAGACA

AAGTAGAAAGCGTTACGTTCTTTTCAggtggcggtggctctggtggcggt

ggctctggtggcggtggctctagcggc**GGCGCGCCgATGTTACGTCCTGTA**

GAAACCCCAACCCGTGAAATCAAAAAACTCGACGGCCTGT

GGGCATTCAGTCTGGATCGCGAAAACTGTGGAATTGATCA

GCGTTGGTGGGAAAGCGCGTTACAAGAAAGCCGGGCAATT

GCTGTGCCAGGCAGTTTTAACGATCAGTTCGCCGATGCAG

ATATTCGTAATTATGCGGGCAACGTCTGGTATCAGCGCGA

AGTCTTTATACCGAAAGGTTGGGCAGGCCAGCGTATCGTG

CTGCGTTTCGATGCGGTCACTCATTACGGCAAAGTGTGGG

TCAATAATCAGGAAGTGATGGAGCATCAGGGCGGCTATAC

GCCATTTGAAGCCGATGTCACGCCGTATGTTATTGCCGGG

AAAAGTGTACGTATCACCGTTTGTGTGAACAACGAACTGA

ACTGGCAGACTATCCCGCCGGGAATGGTGATTACCGACGA

AAACGGCAAGAAAAAGCAGTCTTACTTCCATGATTTCTTTA

ACTATGCCGGGATCCATCGCAGCGTAATGCTCTACACCAC

GCCGAACACCTGGGTGGACGATATCACCGTGGTGACGCAT

GTCGCGCAAGACTGTAACCACGCGTCTGTTGACTGGCAGG

TGGTGGCCAATGGTGATGTCAGCGTTGAACTGCGTGATGC

GGATCAACAGGTGGTTGCAACTGGACAAGGCACTAGCGGG

ACTTTGCAAGTGGTGAATCCGCACCTCTGGCAACCGGGTG

AAGGTTATCTCTATGAACTGTGCGTCACAGCCAAAAGCCA

GACAGAGTGTGATATCTACCCGCTTCGCGTCGGCATCCGG

TCAGTGGCAGTGAAGGGCGAACAGTTCCTGATTAACCACA

AACCGTTCTACTTTACTGGCTTTGGTCGTCATGAAGATGCG

GACTTGCGTGGCAAAGGATTCGATAACGTGCTGATGGTGC

ACGACCACGCATTAATGGACTGGATTGGGGCCAACTCCTA

CCGTACCTCGCATTACCCTTACGCTGAAGAGATGCTCGACT

GGGCAGATGAACATGGCATCGTGGTGATTGATGAAACTGC

TGCTGTCGGCTTTAACCTCTCTTTAGGCATTGGTTTCGAAG

CGGGCAACAAGCCGAAAGAACTGTACAGCGAAGAGGCAGT

CAACGGGGAAACTCAGCAAGCGCACTTACAGGCGATTAAA

GAGCTGATAGCGCGTGACAAAAACCACCCAAGCGTGGTGA

TGTGGAGTATTGCCAACGAACCGGATACCCGTCCGCAAGG

TGCACGGGAATATTTCGCGCCACTGGCGGAAGCAACGCGT

AAACTCGACCCGACGCGTCCGATCACCTGCGTCAATGTAA

TGTTCTGCGACGCTCACACCGATACCATCAGCGATCTCTTT

GATGTGCTGTGCCTGAACCGTTATTACGGATGGTATGTCC

AAAGCGGCGATTTGGAAACGGCAGAGAAGGTACTGGAAAA
AGAACTTCTGGCCTGGCAGGAGAAACTGCATCAGCCGATT
ATCATCACCGAATACGGCGTGGATACGTTAGCCGGGCTGC
ACTCAATGTACACCGACATGTGGAGTGAAGAGTATCAGTG
TGCATGGCTGGATATGTATCACCGCGTCTTTGATCGCGTCA
GCGCCGTCGTCGGTGAACAGGTATGGAATTTCGCCGATTT
TGCGACCTCGCAAGGCATATTGCGCGTTGGCGGTAACAAG
AAAGGGATCTTCACTCGCGACCGCAAACCGAAGTCGGCGG
CTTTTCTGCTGCAAAAACGCTGGACTGGCATGAACTTCGGT
GAAAAACCGCAGCAGGGAGGCAAACAAGCGGCCGCgcatcacc
accatcaccacCCTGCAGGgTAACATCCGATGAAAGTCTTGTATA
TCCAGTCGGGATACGGAGGAATATACAGTTATTTTGATCGC
TGGGCTGAAGAATGTTTTCAGAACACTCATACAGAGTATAT
GATTGCAGATAAACCTGAAGCAGAATCTTTGATGAAGATCG
AAGCGTTTCAGCCGGATTTCACACTTATGATGGTTGGAGAC
CGCGTTCCTCACGACTGGCTGACCTGGTTAAAGGGTAAGGA
TATCCCCGTGTATGTTTGGCTGACCGAAGACCCATTTTATA
TGGATATCAGCCTTCAGGTAATTAAGCTTGCTGACGCCATA
TTAACGATAGAACAAAATGCAGCTCTCTACTATCAAGAGCT
CGGCTACCAGAACGTCTATTATGTTCCGATACCAGTTAACC
ATCGGCTGTTTAAAAAAATGGGCACCGAACATTCCTATCAT
TCAAATCTGTTAATCATCGGCTATCCTTACCCTAATCGGGT
TCAGCTTATGAAAGAAGCGGTCCATCTGCCATTCACGGTTC
GGGTAATCGGCAAGGAATGGGGAAAATATCTGCCTAAAAAA
GTGCTCAAACAGCCGCATATTGATGTCGTCAGCACATGGGT
TCCACCCGAACAGGCCGTTCATTATTATAATGGGGCAGACA

TCGTAATTAATGCGCATCGTCCCTATCATTTTGCTTTCAACC
AAAATACTATGCGCATCAAAAATGCCAGTTTTAATAACCGA
ACGTTTGATATCGCTGCATGCGAAGCATTTCAATTAACAGA
TTTACCCGCGGCGCATCCTTTTTCATCCATTATTTCTTATCA
CGGCATGAACGACTTTAAGGAAAAAGCCGCTTTCTATATCA
ACCACCCTGAAGAACGGCAAAAAGCTGCAGCAGCCAATTAT
AAAGAGACCGTACCCGCATTTACGTTTGATGAGCTTCCTGC
AAAACTAAAAGCGATTCATCTGGCGCTTTCTTGAAAAATCG
TTCATTGCTATGAACGATTTTTTTATTCATAGAAAAAAGCA
GCTGCACAAGCTGCTTTCTATTCTAGCCGTCAGAACGGTCT
TTCAGCTTCCTCGGATTTACCTGTTTATGAAGATCAGGCAT
TTCGCCGAGATGCTGTGCAATTTGTTCCCACGATACAATTT
TGAAATTTTGATTGACGGCTTGTCCGTTATCAATATTTTCG
CCGTCCTGCGCCACAAAAATCCCGTACGGATATTTTGGGCC
AAGTCCGAAACCGAGAACATCAATACCATCCGTGTCACTAG
TACCGTCTATCTTCTCGCCATCTGTAATCTCAAAGTTGGCT
ACATAGCGATTTTTCCCCTGCCGTTCATACATTGCATAGCT
GTTATTTCCTTGACTTGAAGCCATGAGATATCCTTTGCCAT
TTGGTGCATAATAGATTGTCAGTCCTTCAATATCAGCTGTC
AAATGATCTCCTGTCGCACGGTCAACAACCTGCCCCTTTGA
CCCTCCGCCGGGCTCAGCGTTAAATTTCCAGATGGCCTCAT
CTTCCTCTGCTATGTATAGGTTTCCGTACTCATCATCCGCA
ACAAGGCCTTCGGTCTGAGAATTCAACTTAAATTCACGCAC
CTTTTTCCCTGTTACATAACCCTTTCCACCATCAACAATTTC
ATACTGCTCAAATTCCCCTTGTTTGCCTGTCACTAATGCGT
AAAATGCTCCTGTTTTCTGGCTGTGATACAAGCTGAATCCA

TAAACCTCAGAAATATTGGTGGAAATAGGATGGTTCGGATC
TGTAATGCTTTTCAATTTTCCTTTATCCCCGTCTATTGCATA
TACTTCAATTGTATTTTTTCCTTCGGACCGGTTGGATGCGG
CAGCAATATCAATTTTTTCGCCGTTCAATGGAAAATCATAG
CGCAGATCGACATTATTGAGCTTGCCAAACTCATAAGAATG
AAGCTGTTTTCCGTCTAAATCATACACAACGAGCCCTGACT
TCTTATTTGTTGTAATCAACTTGCTTTTTTCCGGGTGTTTT
TCATG

(CgeA-C-GUS Construct Sequence: 2kb homology arms, Native CgeA promoter (yellow), CgeA coding sequence minus the stop codon (green), linker (lowercase), GUS (bold), and 6xHis (lowercase and bold).)

Chapter 3

Towards the Engineering of Natural Competence

3.1 Design of an RBS Library for Enhanced Multiplex Genome Engineering in *B. subtilis*

As demonstrated by our use of co-transformation for editing the native loci of spore coat proteins in *B. subtilis*, natural competence and recombination enables precise genome editing with relatively simple methods. Our co-transformation methods for spore-display were inspired by previous work. First termed congression, the simultaneous uptake and integration of separate pieces of DNA by the same clone, the phenomenon was used in foundational studies of *B. subtilis* that shed light on the local organization of genes into ordered clusters called operons (Nester, Schafer, and Lederberg 1963). More recently, exciting work has demonstrated that natural competence can be enhanced to increase the transformation efficiency of naturally competent organisms. In bacteria including *B. subtilis*, Vibrio cholerae, and Acinetobacter baylyi, rational engineering approaches have yielded increases in transformation efficiency by as much as 30-fold (Dalia, McDonough, and Camilli 2014; Rahmer, Heravi, and Altenbuchner 2015; Dalia et al. 2018). Many factors can affect a cells ability to be transformed. This includes but is not limited to the cells ability to 1) uptake DNA,

2) at least temporarily house the DNA stably in its cytoplasm, and 3) integrate that DNA into its genome using homologous recombination. In *B. subtilis*, the Altenbuchner group at the University of Stuttgart showed that synthetically-induced expression of the competence regulator ComK obviated the need for two-media starvation methods and yielded transformation efficiencies ≈6-fold higher (Rahmer, Heravi, and Altenbuchner 2015). Meanwhile, in Vibrio cholerae, the Camilli group demonstrated Multiplex Genome Engineering by Natural Transformation (MuGENT) for the first time (Dalia, McDonough, and Camilli 2014). Using the principals of congression, they showed that not only two but as many as four independent pieces of DNA could simultaneously be taken up and integrated by a single cell of V. cholerae. They then used natural competence to engineer natural competence. Knowing they could edit multiple loci simultaneously; they used co-transformation to modify the expression of 5 genes known to play a part in DNA uptake and recombination using an RBS and promoter library. After successive rounds of co-transformation and selection, they identified strains with as many as five mutations. One strain, with mutations in just two genes effecting the expression of TfoX (involved in DNA uptake) and RecA (involved in recombination) had a transformation efficiency >30x the parent strain (Dalia, McDonough, and Camilli 2014). Moreover, Ankur Dalia and his research group at Indiana University went on to show that deleting single-stranded DNA (ssDNA) exonucleases could further improve MuGENT by enabling genome editing with smaller, easier to synthesize DNA constructs. In Vibrio natriegens, a

fast-growing Gram-negative bacteria, they showed they could even create synthetic competence through the inducible expression of the tfoX gene (Dalia et al. 2017). Inspired by this body of work, we set out

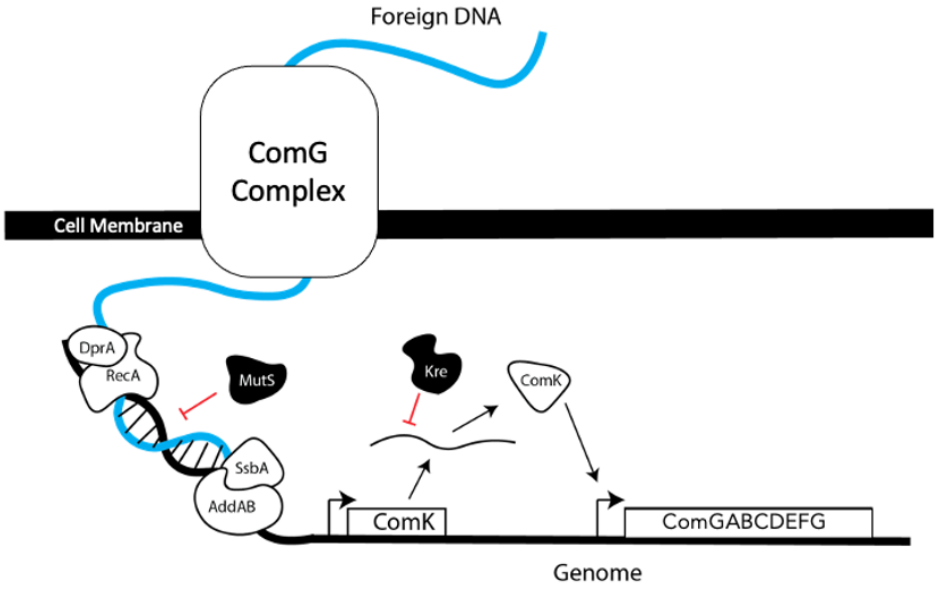


Figure 3.1: Natural Competence in *B. subtilis*: A simplified diagram of the 15 genes we included in our initial RBS library designs and the roles they play in natural competence

to further improve natural competence in *B. subtilis* by modifying the expression of genes that control natural competence. To do so, we aimed to design an RBS and promoter library like had been done for V. cholerae in the original MuGENT strategy. A detailed overview of the natural competence machinery of *B. subtilis* is presented in (Rahmer, Heravi, and Altenbuchner 2015). We identified 15 genes that play a role in the competence of *B. subtilis* that we included in our library design. This includes genes responsible for single-stranded DNA (ssDNA) uptake, ssDNA stability, DNA-integration, DNA repair, and two genes that inhibit transformation. For example, we anticipated that

an increase in expression of RecA, which integrates new DNA into the genome, could yield improvements in transformation efficiency as it had in Vibrio cholerae. However, we also expected a decrease in the expression of Kre, which destabilizes the mRNA of the competence regulator ComK, to improve transformation efficiency. We applied similar logic to each gene of interest and designed our library accordingly. Thus, for genes which are assigned (+), we designed our RBS library to increase their expression as much as possible-while also allowing for WT and lower levels of expression. For genes assigned (-) we have aimed to decrease their expression level as much as possible – while still allowing for WT expression as well. These genes, their functional category, and their desired change in expression levels are outlined in Table 3.1. Accordingly, we used degenerate nucleotides to create sequences that 1) retained the wild-type RBS sequence as a member of the library, 2) covered the highest possible range of predicted expression levels and 3) if necessary, due to overlap of a genes RBS and the preceding CDS, encoded for synonymous codons. Gene sequences were gathered from the NCBI database of the *B. subtilis* 168 genome. The 35 nucleotides preceding the start codon of every gene were selected for modification as shown in Figure 3.2. Sequences were generated using two methods. In

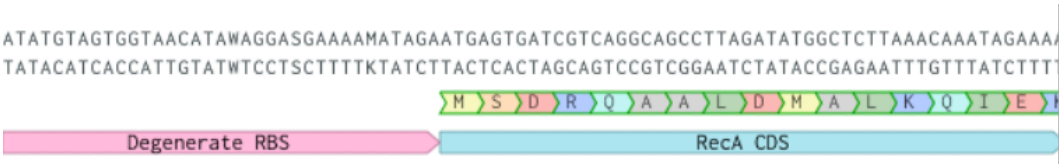


Figure 3.2: Example of the RBS and CDS structure of a gene of interest. The 35 nucleotides upstream of the CDS were selected for modification. Those and the first 50 nucleotides of each genes CDS were used to predict expression levels of library members.

the first, RBS sequences were simply randomly modified and evaluated using RBS calculators "Predict: RBS Library" mode. In the second, the RBS calculators "Optimize: Search Mode" was used to identify mutational hot spots in the RBS – hot spots being defined as nucleotide positions in which modification yielded the greatest change in predicted expression levels. From those hot spots, nucleotides were selected for degeneracy and evaluated using the RBS calculator. We found that pairing a triple-degenerate and double-degenerate nucleotide in the RBS to create a 6-membered library was often sufficient for creating large changes in predicted expression levels. However, for those genes that did have RBS-CDS overlap, we found that it was easier to gain diversity in expression level while preserving AA sequence when using three double-degenerate nucleotides to create an 8-member library. We were further limited by the fact that most synonymous mutations can only occur in the third nucleotide of each codon. Thus, members of the library encode for either 6 or 8 sequences. Additionally, the start codon of one gene, comGF, was changed from TTG to ATG to achieve the desired increase in predicted expression levels. The RBS calculator predicts protein expression levels on a scale of 0 100,000+ arbitrary units (Salis 2011). Of the 15 genes in our library, 4 have WT expression levels over 100,000 and we were unable to increase their predicted expression levels further, even after using the RBS calculators "Optimize: Search Mode" function. This includes genes addA, comK, recA, and ssbA. Each gene, its WT expression level, the min and max expression levels of its library members, the number of library members for each gene, and the corresponding degenerate sequences are featured in Table 3.2. Total

Table 3.1: Gene Function Changes

Gene Name	Function	Desired Direction of Change
RecA	Homologous Recombination Machinery	+
ComGA	DNA Uptake	+
ComGB	DNA Uptake	+
ComGC	DNA Uptake	+
ComGD	DNA Uptake	+
ComGE	DNA Uptake	+
ComGF	DNA Uptake	+
ComGG	DNA Uptake	+
AddA	DNA Helicase	+
SsbA	ssDNA Protector Protein	+
ComK	Competence Regulation	+
Kre	ComK mRNA Destabilizer	-
MutS	Mismatch Repair Protein	-
AddB	DNA Helicase	+
DprA	ssDNA Protector Protein	+

library diversity is 1.486x1012 members. These constructs were to be synthesized by the Department of Energys Joint Genome Institute. We only received correctly assembled RBS libraries for seven members of the RBS library: AddA, ComGB, ComGC, ComGD, ComGEComGF, MutS. Libraries for four others were also delivered but carried errors in their sequences that rendered them useless. Finally, in January of 2022, we were informed that work on trying to synthesize these constructs had stopped and we did not receive anymore. It is possible that a combination within the library of seven genes that we did receive could produce a strain of *B. subtilis* with enhanced competence, but I never started working on this strategy because I focused solely on the spore-display work in my final year at Tufts.

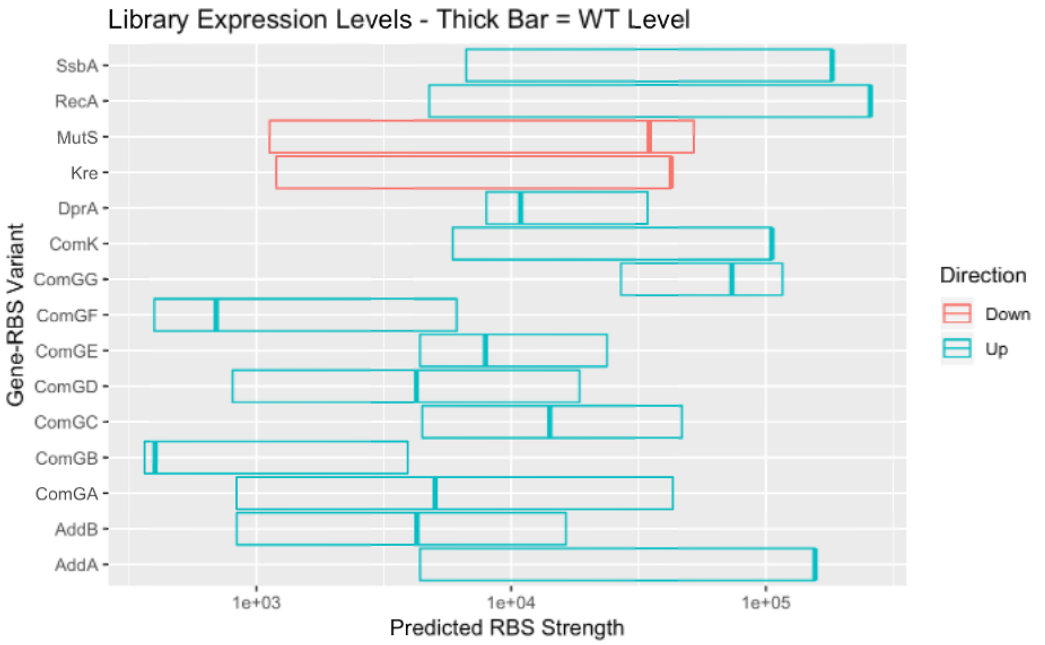


Figure 3.3: The minimum and maximum expression levels of each gene are displayed by the ends of each bar. WT expression levels are displayed as thick vertical lines. Genes colored blue are those in which an increase from WT levels is desired, while a decrease is desired in those colored red.

Table 3.2: RBS Sequences

RBS	WT	Min	Max	%WT	%WT	M#	Degenerate Sequences
AddA	155936	4387	155936	3	100	6	AATACTTGAGTGGATAAAAAHGGRGGCGGATGGCA
AddB	4265	837	16410	20	385	6	AATATAGAGAGATAAAAAGAGVGGGRTCTTCTAAT
ComGA	5032	835	43156	17	858	6	TTCTACAATATATGTTGAAAGGMGVCGGAATCAAA
ComGB	400	364	3923	91	981	6	AGGATCAAGCCAGGTTATTHARGAGGCTCGGTGAA
ComGC	14173	4478	46829	32	330	6	TCAGATGATGAATCAAATGTGAADGGAAGAGGCKG
ComGD	4239	806	18554	19	438	8	TAAGCGCATTATCATCACCGGMGGWGARGTTAAGG
ComGE	7927	4260	29335	54	370	6	TGACATMACAGTTTATCTAGGGAGCGGGAGAGTVA
ComGF	695	397	6111	57	879	8	CGCTTCTTAACGTATTATTTTCGCKRTRAGTCTTT
ComGG	73596	26977	116105	37	158	8	CTTTTCCAGTCTATTCRTATTTAGGRGGGGGKTGA
ComK	105595	5899	105595	6	100	6	TATAAATTTTGCAGAAAARGGATGVAGGCCATAAT
DprA	10896	1900	34340	17	315	6	GTTCCTGMTTTTTCTAACAVAAGGAGGTCAATCTA
RecA	257315	4768	257315	2	100	8	ATATGTAGTGGTAACATAWAGGASGAAAAMATAGA
SsbA	182000	6662	182062	4	100	6	TTTGAAATATATAATGGTWAAAGVTGGTCTTTCTT
Kre	42494	1197	42494	3	100	6	CATATAGAAATTGATAAGAGCARAGVAGAGAAATA
MutS	34859	1127	52161	3	150	6	AATAGAAACGTTGATACATAGTGRGVGATTAAACA

3.2 Screening Competence Effector Deletion Strains for Enhanced Competence

While we were waiting for the RBS library members to be synthesized, we chose to screen additional mutations that could potentially lead to enhanced natural competence in *B. subtilis*. We were curious if we could find single gene deletions that could have a positive impact on natural competence. Previous studies had shown that single deletions of the genes *kre* and *rok* could improve transformation efficiencies and we reasoned that (1) there may be other single deletions that could improve competence and (2) that attaining these mutants would be much simpler than RBS library generation because a library of non-essential gene knockout strains already exists for *B. subtilis* (Gamba, Jonker, and Hamoen 2015; Serrano, Torres, and Alonso 2021; Koo et al. 2017). Competence in *B. subtilis* is bi-modal, with ≈10-20% of a

population gaining competence when induced. This competent subpopulation is in the "K-state," meaning these cells have highly elevated levels of ComK, the competence master-regulator protein (Leisner et al. 2007). Competence is bimodal and once a threshold of ComK is achieved it essentially becomes locked into competence for a short period of time (Mirouze et al. 2012). During this time, genome replication and cell division slows or stops entirely. Instead, cellular resources are directed toward DNA uptake and recombination. This is a product of the positive-feedback loop in which ComK contributes to its own expression by binding its own promoter, pComK, and activation of the competence regulon consisting of >100 genes. The higher the concentration of ComK in a cell, the more likely it is to enter the K-state. As such, it makes sense that Δ*kre* and Δ*rok* strains have higher transformation efficiencies. Kre actively targets and destabilizes ComK mRNA and Rok actively binds and represses the ComK promoter, pComK (Gamba, Jonker, and Hamoen 2015; Hoa et al. 2002). Both lead to lower levels of ComK when present in *B. subtilis*. We wanted to test if deletion of other ComK repressors could lead to enhanced competence like in Δ*kre* and Δ*rok* strains. pComK is one of the most complex bacterial promoters that has been characterized to date (Maier 2020). In addition to ComKs self-activation, there are five additional transcription factors that are known to directly bind pComK and effect ComK expression. These include SpoA, Rok, CodY, AbrB, and DegU. SpoA, Rok, CodY, and AbrB all bind pComK and contribute to ComK repression (Mirouze et al. 2012; Hoa et al. 2002; Hamoen et al. 2003;

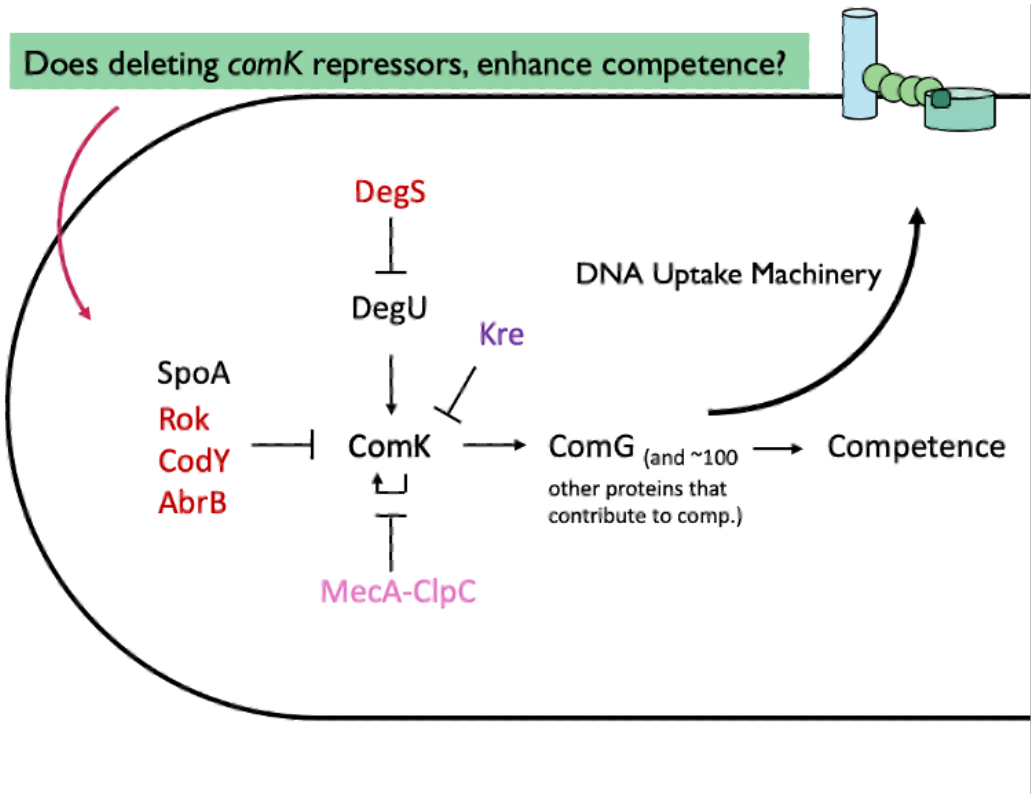


Figure 3.4: ComK Regulation in *B. subtilis*

Miras and Dubnau 2016). DegU is a ComK activator and promotes entrance into the K-state but it itself is down-regulated by DegS (Ogura and Tanaka 1996). We chose not to study a SpoA knockout because SpoA is also involved in sporulation signaling and if we were able to identify a strain with enhanced competence we wanted it to be able to sporulate too. So, for transcriptional repressors of ComK, we chose to study *codY*, *abrB*, and *degS* knockout strains in addition to Δ*rok*. We also included a DegU knockout strain as a negative control since DegU deletion is already known to negatively affect *comK* expression (Ogura and Tanaka 1996). We ordered single knockouts of each of these strains from the Bacillus Genomic Stock Center (BGSC) and quantified their transformation efficiency 3.4. In our study, we found that the Δ*codY*

and Δ*degU* strains had completely lost their ability to be transformed while Δ*abrB* and Δ*degS* strains both had significantly reduced transformation efficiencies. On the other hand, Δ*rok* had a slightly higher mean transformation efficiency but we were unable to replicate the 6x-higher transformation efficiency presented in Serrano, 2021. This may have been due to (1) differences in our transforming DNA and (2) generally high variability in our transformation efficiency results. In addition to controlling ComK expression through transcription, like Rok, and through mRNA destabilization, Kre, the MecA-ClpC protease complex is known to selectively target and degrade ComK. Previous studies had shown that a Δ*mecA* strain could not exit the competence state after entering it and would not form colonies following transformation (Hahn, 1995). We made plans to test if an inducible *mecA* construct could be made to allow the Δ*mecA* strain to enter a state of ultra-competence, be transformed, and then induce degradation of ComK and rescue the cells from their arrest. So we added Δ*mecA* to our list of strains to test as well. Finally, because of the positive results the Camilli group saw in co-transformation of Streptococcus pneumoniae in a mismatch repair deficient strain we added four additional *B. subtilis* knockouts to our list: Δ*mutL*, Δ*mutS*, Δ*ydiR*, and Δ*ydiS*. Inactivation of the mismatch repair genes *mutL* and *mutS* have been used extensively in multiplex genome engineering in *E. coli* in strategies like MAGE and PORTMAGE (Wannier et al. 2021). These proteins help proofread the genome as it is being replicated and repressing their activity can increase the occurrence of desired mutations in some bacteria. YdiR and YdiS are a part of the *B. subtilis* immune system as they check DNA

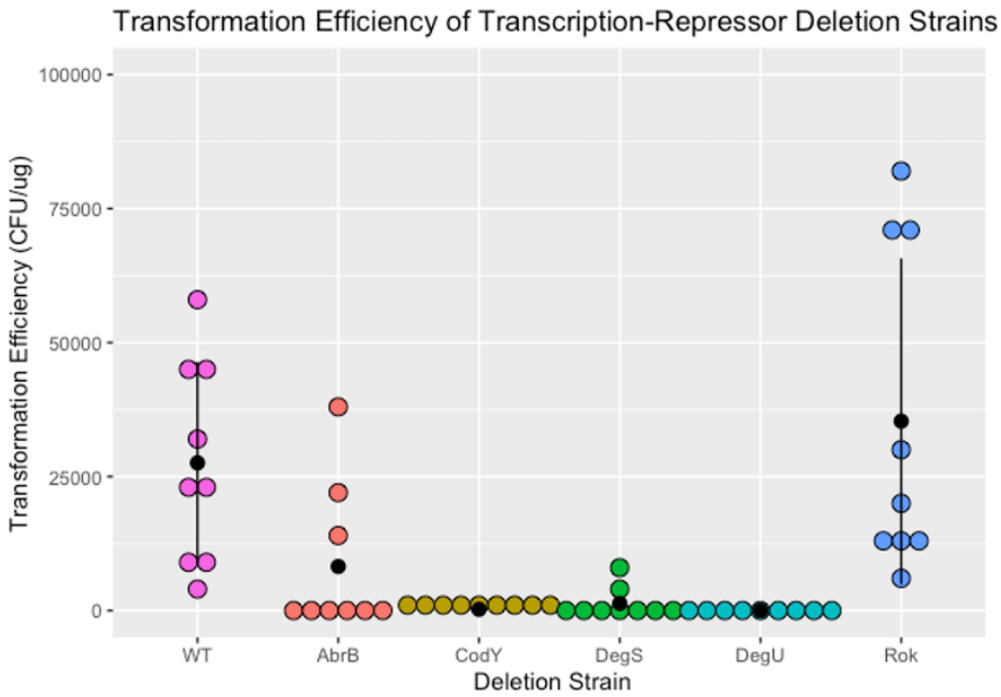


Figure 3.5: Transformation Efficiency of Transcription Repressor Knockout Strains: WT, Δ*abrB*, Δ*codY*, Δ*degS*, Δ*degU*, and Δ*rok* strains transformation efficiency. Each strain was tested in three sets of three biological replicates (n = 9)

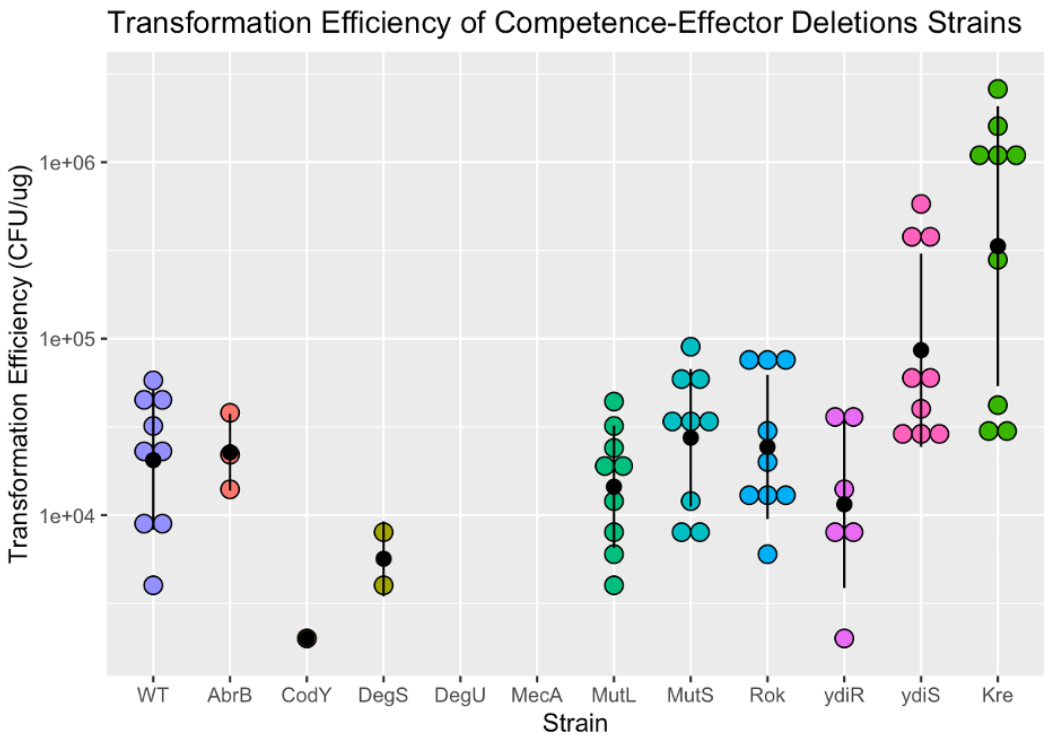


Figure 3.6: Transformation efficiencies of single KO strains: Comparison of the measured transformation efficiencies of all 11 tested single KO strains, done using freshly prepared competent cells

for specific methylation patterns. If they find DNA that is not methylated as if though it was made by *B. subtilis* it is cut and degraded by nucleases to prevent viral infection. We tested all of these mutations in addition to Δ*kre* but found that only the Δ*kre* strain had an average transformation efficiency significantly greater than WT *B. subtilis* 168. Importantly, our measured transformation efficiency for Δ*kre* was ≈36x that of the WT which was in good agreement with the findings of (Gamba, Jonker, and Hamoen 2015).

3.2.1 Discussion & Future Directions

Cumulatively, our efforts in enhancing natural competence in *B. subtilis* have yet to yield any functionally useful results. The correctly synthesized RBS and promoter library parts that I designed are still in cold storage in the Nair lab and could one day be used to screen for functional improvements. Since we first designed these libraries, additional light has been shed on the mechanisms of natural competence and I would likely add additional genes to the list of candidate proteins. For example, more is now known about ComEA and how it plays a critical role in the binding and internalization of transforming DNA (Hahn, Desantis, and Dubnau 2021). Moreover, during this time, an exciting publication from a Chinese team demonstrated multiplex genome engineering in *B. subtilis* using a unique combination of induced natural competence and recombineering (Deng et al. 2021). One drawback to their system is the need to assemble large, complex vectors containing several regions of homology with the *B. subtilis* genome. We have learned that large regions of *B. subtilis* genomic DNA can be difficult to work with, even for even well-funded commercial DNA synthesis labs. Strategies that enable high-frequency MuGENT in *B. subtilis* with smaller, independent editing templates could make engineering *B. subtilis* simpler, and less prone to failure during DNA synthesis, and should be pursued.

Bibliography

Anagnostopoulos, C., and John Spizizen. 1961. " Requirements for Transformation in Bacillus Subtilis." *Journal of Bacteriology* 81 (5): 741–746. ISSN: 0021-9193. https://doi.org/10.1128/jb.81.5.741-746.1961.

Aui, A, Y Wang, and M Mba-Wright. 2021. "Evaluating the economic feasibility of cellulosic ethanol: A meta-analysis of techno-economic analysis studies." *Renewable and Sustainable Energy Reviews* 145:111098.

Bartels, Julia, Sebastián López Castellanos, Jara Radeck, and Thorsten Mascher. 2017. "Sporobeads: The Utilization of the Bacillus subtilis Endospore Crust as a Protein Display Platform." *ACS Synthetic Biology,* acssynbio.7b00285. ISSN: 2161-5063. https://doi.org/10.1021/acssynbio.7b00285. http://pubs.acs.org/doi/10.1021/acssynbio.7b00285.

Bennallack, Philip R, Scott R Burt, Michael J Heder, Richard A Robison, and Joel S Griffitts. 2014. "Characterization of a novel plasmid-borne thiopeptide gene cluster in Staphylococcus epidermidis strain 115." *Journal of bacteriology* 196 (24): 4344–4350.

Bernal, Claudia, Karen Rodríguez, and Ronny Martínez. 2018. "Integrating enzyme immobilization and protein engineering: An alternative path for the development of novel and improved industrial biocatalysts." *Biotechnology Advances* 36 (5): 1470–1480. ISSN: 07349750. https://doi.org/10.1016/j.biotechadv.2018.06.002. https://doi.org/10.1016/j.biotechadv.2018.06.002.

Bloois, Edwin van, Remko T. Winter, Harald Kolmar, and Marco W. Fraaije. 2011. "Decorating microbes: Surface display of proteins on Escherichia coli." *Trends in Biotechnology* 29 (2): 79–86. ISSN: 01677799. https://doi.org/10.1016/j.tibtech.2010.11.003. http://dx.doi.org/10.1016/j.tibtech.2010.11.003.

Bowie, James U., Saken Sherkhanov, Tyler P. Korman, Meaghan A. Valliere, Paul H. Opgenorth, and Hongjiang Liu. 2020. "Synthetic Biochemistry: The Bio-inspired Cell-Free Approach to Commodity Chemical Production." *Trends in Biotechnology* 38 (7): 766–778. ISSN: 18793096. https://doi.org/10.1016/j.tibtech.2019.12.024. https://doi.org/10.1016/j.tibtech.2019.12.024.

Cano, Raul J, and Monica K Borucki. 1995. "Revival and identification of bacterial spores in 25-to 40-million-year-old Dominican amber." *Science* 268 (5213): 1060–1064.

Chen, Irwin, Brent M Dorr, and David R Liu. 2011. "A general strategy for the evolution of bond-forming enzymes using yeast display." 2011. https://doi.org/10.1073/pnas.1101046108.

Chen, Long, Ashok Mulchandani, and Xin Ge. 2017. "Spore-displayed enzyme cascade with tunable stoichiometry." *Biotechnology progress* 33 (2): 383–389.

Ciabattini, Annalisa, Riccardo Parigi, Rachele Isticato, Marco R Oggioni, and Gianni Pozzi. 2004. "Oral priming of mice by recombinant spores of Bacillus subtilis." *Vaccine* 22 (31-32): 4139–4143.

Cui, Wenjing, Laichuang Han, Feiya Suo, Zhongmei Liu, Li Zhou, and Zhemin Zhou. 2018. "Exploitation of Bacillus subtilis as a robust workhorse for production of heterologous proteins and beyond." *World Journal of Microbiology and Biotechnology* 34 (10): 1–19. ISSN: 15730972. https://doi.org/10.1007/s11274-018-2531-7. http://dx.doi.org/10.1007/s11274-018-2531-7.

Dalia, A. B., E. McDonough, and A. Camilli. 2014. "Multiplex genome editing by natural transformation." *Proceedings of the National Academy of Sciences* 111 (24): 8937–8942. ISSN: 0027-8424. https://doi.org/10.1073/pnas.1406478111. http://www.pnas.org/cgi/doi/10.1073/pnas.1406478111.

Dalia, Ankur B., and Triana N. Dalia. 2019. "Spatiotemporal Analysis of DNA Integration during Natural Transformation Reveals a Mode of Nongenetic Inheritance in Bacteria." *Cell* 179 (7): 1499–1511. ISSN: 10974172. https://doi.org/10.1016/j.cell.2019.11.021. https://doi.org/10.1016/j.cell.2019.11.021.

Dalia, Triana N, Chelsea A Hayes, Sergey Stolyar, Christopher J Marx, James B Mckinlay, and Ankur B Dalia. 2017. "Multiplex Genome Editing by Natural Transformation (MuGENT) for Synthetic Biology in Vibrio natriegens," https://doi.org/10.1021/acssynbio.7b00116.

Dalia, Triana N, Soo Hun Yoon, Elisa Galli, Francois-xavier Barre, M Waters, and Ankur B Dalia. 2018. "Enhancing multiplex genome editing by natural transformation (MuGENT) via inactivation of ssDNA exonucleases." 45 (12): 7527–7537. https://doi.org/10.1093/nar/gkx496.

Deng, Aihua, Zhaopeng Sun, Tiantian Wang, Di Cui, Lai Li, Shuwen Liu, Fei Huang, and Tingyi Wen. 2021. "Simultaneous Multiplex Genome Engineering via Accelerated Natural Transformation in Bacillus subtilis." *Frontiers in Microbiology* 12 (August). ISSN: 1664302X. https://doi.org/10.3389/fmicb.2021.714449.

Francisco, Joseph A, Charles F Earhart, and George Georgiou. 1992. "Transport and anchoring of beta-lactamase to the external surface of Escherichia coli." *Proceedings of the National Academy of Sciences* 89 (7): 2713–2717.

Gamba, Pamela, Martijs J. Jonker, and Leendert W. Hamoen. 2015. "A Novel Feedback Loop That Controls Bimodal Expression of Genetic Competence." *PLoS Genetics* 11 (6): 1–32. ISSN: 15537404. https://doi.org/10.1371/journal.pgen.1005047.

Gupta, Nirupama, and Edgardo T. Farinas. 2010. "Directed evolution of CotA laccase for increased substrate specificity using Bacillus subtilis spores." *Protein Engineering, Design and Selection* 23 (8): 679–682. ISSN: 17410134. https://doi.org/10.1093/protein/gzq036.

Hahn, Jeanette, Micaela Desantis, and David Dubnau. 2021. "Mechanisms of Transforming DNA Uptake to the Periplasm of Bacillus subtilis," https://doi.org/10.1128/mBio. https://doi.org/10.1128/mBio.

Hamoen, Leendert W., Daisy Kausche, Mohamed A. Marahiel, Douwe Van Sinderen, Gerard Venema, and Pascale Serror. 2003. "The Bacillus subtilis transition state regulator AbrB binds to the -35 promoter region of comK." *FEMS Microbiology Letters* 218 (2): 299–304. ISSN: 03781097. https://doi.org/10.1016/S0378-1097(02)01172-2.

Hoa, Tran Thu, Pablo Tortosa, Mark Albano, and David Dubnau. 2002. "Rok (YkuW) regulates genetic competence in Bacillus subtilis by directly repressing comK." *Molecular Microbiology* 43 (1): 15–26. ISSN: 0950382X. https://doi.org/10.1046/j.1365-2958.2002.02727.x.

Hosoi, Tomohiro, Kan Kiuchi, et al. 2003. "Natto–a food made by fermenting cooked soybeans with Bacillus subtilis (natto)." *Handbook of fermented functional foods* 20034675:227–250.

Hui, Yue, Ziyu Cui, and Seunghyun Sim. 2022. "Stress-tolerant, recyclable, and autonomously renewable biocatalyst platform enabled by engineered bacterial spores." *bioRxiv,* 2022.03.16.484680. https://doi.org/10.1021/acssynbio.2c00256. http://biorxiv.org/content/early/2022/03/17/2022.03.16.484680.abstract.

Hullo, Marie-Francoise, Ivan Moszer, Antoine Danchin, and Isabelle Martin-Verstraete. 2001. "CotA of Bacillus subtilis is a copper-dependent laccase." *Journal of bacteriology* 183 (18): 5426–5430.

Imamura, Daisuke, Ritsuko Kuwana, Hiromu Takamatsu, and Kazuhito Watabe. 2011. "Proteins Involved in Formation of the Outermost Layer of Bacillus subtilis Spores." 193 (16): 4075–4080. https://doi.org/10.1128/JB.05310-11.

Ireton, Keith, David Z. Rudner, Kathryn Jaacks Siranosian, and Alan D. Grossman. 1993. "Integration of multiple developmental signals in Bacillus subtilis through the Spo0A transcription factor." *Genes and Development* 7 (2): 283–294. ISSN: 08909369. https://doi.org/10.1101/gad.7.2.283.

Isticato, Rachele. 2023. "Bacterial Spore-Based Delivery System: 20 Years of a Versatile Approach for Innovative Vaccines." *Biomolecules* 13 (6): 947.

Isticato, Rachele, Giuseppina Cangiano, H T Tran, Annalisa Ciabattini, Donata Medaglini, Marco R Oggioni, M De Felice, Gianni Pozzi, and Ezio Ricca. 2001. "Surface display of recombinant proteins on Bacillus subtilis spores." *Journal of bacteriology* 183 (21): 6294–301. ISSN: 0021-9193. https://doi.org/10.1128/JB.183.21.6294. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=100119&tool=pmcentrez&rendertype=abstract.

Iwanicki, Adam, Iwona Piatek, Magorzata Stasioj, Anna Grela, Tomasz ga, Micha Obuchowski, and Krzysztof Hinc. 2014. "A system of vectors for Bacillus subtilis spore surface display." *Microbial Cell Factories* 13 (1): 30. ISSN: 1475-2859. https://doi.org/10.1186/1475-2859-13-30. http://microbialcellfactories.biomedcentral.com/articles/10.1186/1475-2859-13-30.

Katsande, Paidamoyo M., Leira Fernández-Bastit, William T. Ferreira, Júlia Vergara-Alert, Mateusz Hess, Katie Lloyd-Jones, Huynh A. Hong, Joaquim Segales, and Simon M. Cutting. 2022. "Heterologous Systemic PrimeIntranasal Boosting Using a Spore SARS-CoV-2 Vaccine Confers Mucosal Immunity and Cross-Reactive Antibodies in Mice as well as Protection in Hamsters." *Vaccines* 10 (11). ISSN: 2076393X. https://doi.org/10.3390/vaccines10111900.

Keeble, Anthony H., Paula Turkki, Samuel Stokes, Irsyad N.A.Khairil Anuar, Rolle Rahikainen, Vesa P. Hytönen, and Mark Howarth. 2019. "Approaching infinite affinity through engineering of peptide-protein interaction." *Proceedings of the National Academy of Sciences of the United States of America* 116 (52): 26523–26533. ISSN: 10916490. https://doi.org/10.1073/pnas.1909653116.

Kellmann, Sarah-Jane, Christian Hentrich, Mateusz Putyrski, Hanh Hanuschka, Manuel Cavada, Achim Knappik, and Francisco Ylera. 2023. "SpyDisplay: A versatile phage display selection system using SpyTag/SpyCatcher technology." In *Mabs,* 15:2177978. 1. Taylor & Francis.

Koo, Byoung-Mo, George Kritikos, Jeremiah D Farelli, Horia Todor, Kenneth Tong, Harvey Kimsey, Ilan Wapinski, Marco Galardini, Angelo Cabal, Jason M Peters, et al. 2017. "Construction and analysis of two genome-scale deletion libraries for Bacillus subtilis." *Cell systems* 4 (3): 291–305.

Lee, Su-jin, Jae-gu Pan, Seung-hwan Park, and Soo-keun Choi. 2010. "Development of a stationary phase-specific autoinducible expression system in Bacillus subtilis." *Journal of Biotechnology* 149 (1-2): 16–20. ISSN: 0168-1656. https://doi.org/10.1016/j.jbiotec.2010.06.021. http://dx.doi.org/10.1016/j.jbiotec.2010.06.021.

Leisner, Madeleine, Kerstin Stingl, Joachim O. Rädler, and Berenike Maier. 2007. "Basal expression rate of comK sets a 'switching-window' into the K-state of Bacillus subtilis." *Molecular Microbiology* 63 (6): 1806–1816. ISSN: 0950382X. https://doi.org/10.1111/j.1365-2958.2007.05628.x.

Leskelä, Soile, Eva Wahlström, Hanne Leen Hyyryläinen, Myra Jacobs, Airi Palva, Matti Sarvas, and Vesa P. Kontinen. 1999. "Ecs, an ABC transporter of Bacillus subtilis: Dual signal transduction functions affecting expression of secreted proteins as well as their secretion." *Molecular Microbiology* 31 (2): 533–543. ISSN: 0950382X. https://doi.org/10.1046/j.1365-2958.1999.01194.x.

Lotti, Marina, Jürgen Pleiss, Francisco Valero, and Pau Ferrer. 2015. "Effects of methanol on lipases: Molecular, kinetic and process issues in the production of biodiesel." *Biotechnology journal* 10 (1): 22–30.

Maier, Berenike. 2020. "Competence and transformation in bacillus subtilis." *Current Issues in Molecular Biology* 37:57–76. ISSN: 14673045. https://doi.org/10.21775/CIMB.037.057.

Maier, Berenike, Ines Chen, David Dubnau, and Michael P. Sheetz. 2004. "DNA transport into Bacillus subtilis requires proton motive force to generate large molecular forces." *Nature Structural and Molecular Biology* 11 (7): 643–649. ISSN: 15459993. https://doi.org/10.1038/nsmb783.

Marrone, Pamela G. 2023. "Biopesticide commercialization in North America: state of the art and future opportunities." *Development and Commercialization of Biopesticides,* 173–202.

Mckenney, Peter T, Adam Driks, and Patrick Eichenberger. 2012. "The Bacillus subtilis endospore : assembly and functions of the multilayered coat." *Nature Reviews Microbiology* 11 (1): 33–44. ISSN: 1740-1526. https://doi.org/10.1038/nrmicro2921. http://dx.doi.org/10.1038/nrmicro2921.

Mckenney, Peter T, Adam Driks, Haig A Eskandarian, Paul Grabowski, Jonathan Guberman, Katherine H Wang, and Zemer Gitai. 2010. "A Distance-Weighted Interaction Map Reveals a Previously Uncharacterized Layer of the Bacillus subtilis Spore Coat." *Current Biology* 20 (10): 934–938. ISSN: 0960-9822. https://doi.org/10.1016/j.cub.2010.03.060. http://dx.doi.org/10.1016/j.cub.2010.03.060.

Mckenney, Peter T., Adam Driks, and Patrick Eichenberger. 2013. "The Bacillus subtilis endospore: Assembly and functions of the multilayered coat." *Nature Reviews Microbiology* 11 (1): 33–44. ISSN: 17401526. https://doi.org/10.1038/nrmicro2921. http://dx.doi.org/10.1038/nrmicro2921.

Miras, Mathieu, and David Dubnau. 2016. "A DegU-P and DegQ-Dependent regulatory pathway for the K-state in Bacillus subtilis." *Frontiers in Microbiology* 7 (NOV): 1–14. ISSN: 1664302X. https://doi.org/10.3389/fmicb.2016.01868.

Mirouze, Nicolas, Yaanik Desai, Arjun Raj, and David Dubnau. 2012. "Spo0Ap imposes a temporal gate for the bimodal expression of competence in bacillus subtilis." *PLoS Genetics* 8 (3). ISSN: 15537390. https://doi.org/10.1371/journal.pgen.1002586.

Morikawa, Masaaki. 2006. "Beneficial biofilm formation by industrial bacteria Bacillus subtilis and related species." *Journal of Bioscience and Bioengineering* 101 (1): 1–8. ISSN: 13891723. https://doi.org/10.1263/jbb.101.1.

Nester, Eugene W, Marion Schafer, and Joshua Lederberg. 1963. "Gene Linkage in Dna Transfer: a Cluster of Genes Concerned With Aromatic Biosynthesis in Bacillus Subtilis." *Genetics* 48 (4): 529–551. https://doi.org/10.1093/genetics/48.4.529.

Ogura, Mitsuo, and Teruo Tanaka. 1996. "Bacillus subtilis DegU acts as a positive regulator for comK expression." *FEBS letters* 397 (2-3): 173–176.

Poole, Philip. 2017. "Shining a light on the dark world of plant root–microbe interactions." *Proceedings of the National Academy of Sciences* 114 (17): 4281–4283.

Potot, Sebastien, Cláudia R. Serra, Adriano O. Henriques, and Ghislain Schyns. 2010. "Display of recombinant proteins on Bacillus subtilis spores, using a coat-associated enzyme as the carrier." *Applied and Environmental Microbiology* 76 (17): 5926–5933. ISSN: 00992240. https://doi.org/10.1128/AEM.01103-10.

Rahmer, Regine, Kambiz Morabbi Heravi, and Josef Altenbuchner. 2015. "Construction of a super-competent Bacillus subtilis 168 using the PmtlA-comKS inducible cassette." *Frontiers in Microbiology* 6 (DEC): 1–11. ISSN: 1664302X. https://doi.org/10.3389/fmicb.2015.01431.

Rasor, Blake J., Bastian Vögeli, Grant M. Landwehr, Jonathan W. Bogart, Ashty S. Karim, and Michael C. Jewett. 2021. "Toward sustainable, cell-free biomanufacturing." *Current Opinion in Biotechnology* 69:136–144. ISSN: 18790429. https://doi.org/10.1016/j.copbio.2020.12.012.

Rodriguez-Ziga, Ursula Fabiola, David Cannella, Roberto de Campos Giordano, Raquel de Lima Camargo Giordano, Henning Jorgensen, and Claus Felby. 2015. "Lignocellulose pretreatment technologies affect the level of enzymatic cellulose oxidation by LPMO." *Green chemistry* 17 (5): 2896–2903.

Salis, Howard M. 2011. "The ribosome binding site calculator." *Methods in Enzymology* 498:19–42. ISSN: 00766879. https://doi.org/10.1016/B978-0-12-385120-8.00002-4.

Schaeffer, P. 1969. "Sporulation and the production of antibiotics, exoenzymes, and exotonins." *Bacteriological reviews* 33 (1): 48–71.

Schlager, Stefanie, Liviu Mihai Dumitru, Marianne Haberbauer, Anita Fuchsbauer, Helmut Neugebauer, Daniela Hiemetsberger, Annika Wagner, Engelbert Portenkirchner, and Niyazi Serdar Sariciftci. 2016. "Electrochemical Reduction of Carbon Dioxide to Methanol by Direct Injection of Electrons into Immobilized Enzymes on a Modified Electrode." *ChemSusChem* 9 (6): 631–635. ISSN: 1864564X. https://doi.org/10.1002/cssc.201501496.

Schreuder, Maarten P., Stephan Brekelmans, Herman Van Den Ende, and Frans M. Klis. 1993. "Targeting of a heterologous protein to the cell wall of Saccharomyces cerevisiae." *Yeast* 9 (4): 399–409. ISSN: 10970061. https://doi.org/10.1002/yea.320090410.

Scott, Jamie K, and George P Smith. 1990. "Searching for peptide ligands with an epitope library." *Science* 249 (4967): 386–390.

Serrano, Ester, Rubén Torres, and Juan C. Alonso. 2021. "Nucleoid-associated Rok differentially affects chromosomal transformation on Bacillus subtilis recombination-deficient cells." *Environmental Microbiology* 23 (6): 3318–3331. ISSN: 14622920. https://doi.org/10.1111/1462-2920.15562.

Smith, George P. 1985. “Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface.” *Science* 228 (4705): 1315–1317.

Sonenshein, Abraham L, James A Hoch, Richard Losick, et al. 2002. “Bacillus subtilis and its closest relatives: from genes to cells.”

Stahl, Stefan W, Michael A Nash, Daniel B Fried, Michal Slutzki, Yoav Barak, Edward A Bayer, and Hermann E Gaub. 2012. “Single-molecule dissection of the high-affinity cohesin–dockerin complex.” *Proceedings of the National Academy of Sciences* 109 (50): 20431–20436.

Stanley-Wall, Nicola R, and Cait E MacPhee. 2015. “Connecting the dots between bacterial biofilms and ice cream.” *Physical biology* 12 (6): 063001.

Stickel, Jonathan J, Birendra Adhikari, David A Sievers, and John Pellegrino. 2018. “Continuous enzymatic hydrolysis of lignocellulosic biomass in a membrane-reactor system.” *Journal of Chemical Technology & Biotechnology* 93 (8): 2181–2190.

Ullah, Jawad, Huayou Chen, Ake Vastermark, Jinru Jia, Bangguo Wu, Zhong Ni, Yilin Le, and Hongcheng Wang. 2017. “Impact of orientation and flexibility of peptide linkers on T. maritima lipase Tm1350 displayed on Bacillus subtilis spores surface using CotB as fusion partner.” *World Journal of Microbiology and Biotechnology* 33 (9): 166. ISSN: 0959-3993. https://doi.org/10.1007/s11274-017-2327-1. http://link.springer.com/10.1007/s11274-017-2327-1.

Wang, He, Yunxiang Wang, and Ruijin Yang. 2017. “Recent progress in Bacillus subtilis spore-surface display: concept, progress, and future.” *Applied Microbiology and Biotechnology* 101, no. 3 (February): 933–949. ISSN: 0175-7598. https://doi.org/10.1007/s00253-016-8080-9. http://link.springer.com/10.1007/s00253-016-8080-9.

Wannier, Timothy M., Peter N. Ciaccia, Andrew D. Ellington, Gabriel T. Filsinger, Farren J. Isaacs, Kamyab Javanmardi, Michaela A. Jones, et al. 2021. “Recombineering and MAGE.” *Nature Reviews Methods Primers* 1 (1). ISSN: 2662-84492662-8449. https://doi.org/10.1038/s43586-020-00006-x. http://dx.doi.org/10.1038/s43586-020-00006-x.

You, Chun, Ting Shi, Yunjie Li, Pingping Han, Xigui Zhou, and Yi Heng Percival Zhang. 2017. “An in vitro synthetic biology platform for the industrial biomanufacturing of myo-inositol from starch.” *Biotechnology and Bioengineering* 114 (8): 1855–1864. ISSN: 10970290. https://doi.org/10.1002/bit.26314.

Zakeri, Bijan, Jacob O Fierer, Emrah Celik, Emily C Chittock, Ulrich Schwarz-Linek, Vincent T Moy, and Mark Howarth. 2012. “Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin.” *Proceedings of the National Academy of Sciences* 109 (12): E690–E697.

Zhang, Xiaolin, Christopher J. Tervo, and Jennifer L. Reed. 2016. "Metabolic assessment of E. coli as a Biofactory for commercial products." *Metabolic Engineering* 35:64–74. ISSN: 10967184. https://doi.org/10.1016/j.ymben.2016.01.007. http://dx.doi.org/10.1016/j.ymben.2016.01.007.

Zhao, Guangyu, Yu Miao, Yan Guo, Hongjie Qiu, Shihui Sun, Zhihua Kou, Hong Yu, Junfeng Li, Yue Chen, Shibo Jiang, et al. 2014. "Development of a heat-stable and orally delivered recombinant M2e-expressing B. subtilis spore-based influenza vaccine." *Human vaccines & immunotherapeutics* 10 (12): 3649–3658.