Towards a Structural Understanding of Spore Germination in Clostridium Difficile

Chloe M. Adams

University of Vermont

Follow this and additional works at: http://scholarworks.uvm.edu/graddis

Part of the Medical Cell Biology Commons, and the Microbiology Commons

Recommended Citation

This Thesis is brought to you for free and open access by the Dissertations and Theses at ScholarWorks @ UVM. It has been accepted for inclusion in Graduate College Dissertations and Theses by an authorized administrator of ScholarWorks @ UVM. For more information, please contact donna.omalley@uvm.edu.
TOWARDS A STRUCTURAL UNDERSTANDING OF SPORE GERMINATION IN CLOSTRIDIUM DIFFICILE

A Thesis Presented

by

Chloe Adams

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of Master of Science
Specializing in Cellular and Molecular Biology

January, 2015

Defense Date: September 30, 2014
Thesis Examination Committee:

Sylvie Doublé, Ph.D., Co-Advisor
Aimee Shen, Ph.D., Co-Advisor
Stephen Everse, Ph.D., Chairperson
Matthew Wargo, Ph.D.
Cynthia J. Forehand, Ph.D., Dean of the Graduate College
ABSTRACT

*Clostridium difficile* is a Gram-positive bacterium that causes a toxin-mediated disease, typically in individuals whose normal intestinal flora has been compromised by antibiotic therapy. *C. difficile* is naturally resistant to many antibiotics and produces spores that can withstand harsh environmental conditions and many disinfectants, making the infection difficult to clear and easy to spread. The infection begins when spores from the environment are ingested and germinate upon exposure to taurocholate and glycine in the digestive tract. This germination process is required to initiate infection and thus represents a good target for the development of novel therapeutics. Although spore germination is necessary for disease transmission, the molecular mechanisms regulating this process are poorly understood. Germination relies on sensing a germinant and triggering degradation of the cortex layer of the spore, which is important for spore resistance. Once the cortex is degraded, the spore can undergo outgrowth to a vegetative cell and secrete toxins to cause disease symptoms.

There are several discrete steps to the proteolytic cascade that ultimately lead to cortex hydrolysis. First, the pseudoprotease CspC acts as a germinant receptor for the bile salt taurocholate; CspC then relays this signal to the subtilisin-like serine protease, CspB. CspB is required for efficient cleavage and activation of the cortex hydrolase, SleC. Upon proteolytic activation of SleC, cortex hydrolysis can proceed, which allows subsequent outgrowth.

To better understand the mechanistic basis of the germination process, we solved the 1.6 Å structure of the required germination protease, CspB, from *C. perfringens* (a related pathogen). This structure revealed that CspB is comprised of three domains: an associated prodomain, a subtilase domain, and a jellyroll domain. Our work significantly advanced our understanding of the proteolytic cascade that leads to germination; in particular the structure and function of the CspB protease, and the role of its three domains. We have described the four domains of the cortex hydrolase, SleC, and how they contribute to the activity of SleC. We have recently obtained diffraction-quality crystals of the pseudoprotease, CspC, from an organism more closely related to *C. difficile*, *C. bifermentans*. Our latest work, focusing on the germination receptor, CspC, has brought us closer to a three-dimensional structure of this protein, which will likely reveal how it binds ligands and functions in germination.
CITATIONS

Material from this thesis has been published in the following form:

ACKNOWLEDGEMENTS

I owe thanks to a great many people for supporting me in pursuing studies related to my Masters degree. First and foremost, I would like to thank my co-advisors, Sylvie Doublié and Aimee Shen for their time and patience in mentoring me, and for giving me a project that is so interesting and important. It has been an honor to work on the germination of \textit{C. difficile} and I think in a lot of ways what I have studied has influenced my decision to ultimately pursue a career more closely related to patient care, for better or for worse, as during my studies I and both advisors noticed my tendency to get pulled away from basic science and towards the clinical aspects of infection. I am a very lucky girl to have been given the opportunity to spend so much time under the guidance of two very wise women. Thank you for your support and playing such an important part in my development as a person during these years and for sending me to conferences and learning opportunities. Had I not gone to Rapidata, Sylvie, I think my life during grad school may have unfolded in a different way, and I am so grateful for how things happened. And Aimee, Italy was beautiful, I very much enjoyed the conference and the people I met there, and I will never forget the experience.

To my other committee members: Stephen, thank you for time talking me through a very difficult decision, and Matt, thank you for your honest words and time serving on my committee.
Secondly, I wish to thank my friends and family. Karl and Brittany for their camaraderie and support in and out of lab, especially, and for feeding me dinner all the time. While these two have been my best support inside the lab, I have somehow managed to retain life-long friends outside of lab who were always eager to offer an ear, a shoulder, support, food, perspective, or just something else to do or think about when that’s just what I needed. Ellie, Emily, and Amanda especially, who have been my friends since age 3, 8, and 10 and know me as well as I know myself. I would like to thank my dragon boat teammates, Malia Paddling and Racing club, for their support and the opportunity to be part of a team made up of smart, wonderful people from diverse backgrounds, and, sometimes most important: our appreciation and love of things completely unrelated to *C. difficile* and my graduate school life in general. It was important to my sanity to have reasons to not spend all of my time in lab or thinking about science, and all of these people helped me with that.

And finally I would like to thank my boyfriend, Sorabh, whom I met at Rapidata in 2012 and who has been my rock and favorite person ever since. He has always been there to listen to me, to offer support and advice both related to my project and related to my life, and has been such a great partner to me. My parents and sister for their constant love and support, without which I wouldn’t have lasted this long, and for never letting me get any work done when I visited, despite my attempts. I think it’s better in the long run that way anyway. I owe these four people this degree, for reminding me that I am more than a graduate student when that is exactly what I needed to remember.
Table of Contents

CITATIONS .............................................................................................................................. II

ACKNOWLEDGEMENTS ........................................................................................................... III

LIST OF TABLES ......................................................................................................................... VIII

LIST OF FIGURES ..................................................................................................................... IX

CHAPTER 1: INTRODUCTION ..................................................................................................... 1

1.1. Clostridium difficile-associated disease ........................................................................... 1

1.2. The role of spores during C. difficile infections .............................................................. 3

1.3. The role of the microflora in regulating C. difficile spore germination .......................... 4

1.4. The molecular basis for spore germination in the Firmicutes ........................................... 5

1.5. The molecular basis for C. difficile spore germination ...................................................... 10

1.6. Figures .................................................................................................................................. 12

CHAPTER 2: STRUCTURAL AND FUNCTIONAL ANALYSIS OF THE CSPB PROTEASE
REQUIRED FOR CLOSTRIDIUM SPORE GERMINATION .................................................. 16

2.1. Introduction ......................................................................................................................... 16

2.2. Materials and Methods ....................................................................................................... 19

2.2.1. Bacterial growth conditions ......................................................................................... 19

2.2.2. Sporulation Assay ......................................................................................................... 19

2.2.3. Preparation of C. difficile samples for Western blot analysis .................................... 20

2.2.4. Western blot analysis .................................................................................................... 21

2.2.5. Purification of C. difficile spores .................................................................................. 22

2.2.6. Germination assay ......................................................................................................... 22

2.2.7. Crystallization ............................................................................................................... 23

2.2.8. Data Collection ............................................................................................................. 23

2.2.9. Data Processing ............................................................................................................ 24

2.2.10. Structure Solution and Refinement ............................................................................ 24

2.2.11. Limited Chymotrypsin Proteolysis ............................................................................. 25

2.2.12. Preparation of E. coli samples for Western blot analysis ............................................. 26

2.2.13. Activity-based probe labeling of CspB perfringens .................................................... 26

2.2.14. Accession codes .......................................................................................................... 27
4.1. Introduction .................................................................................................................................................. 109

4.2. Materials and methods .................................................................................................................................. 112
  4.2.1. C. difficile CspC protein expression .......................................................................................................... 112
  4.2.2. C. difficile CspC protein purification ........................................................................................................... 112
  4.2.3. CspC bifermentans protein purification ...................................................................................................... 113
  4.2.4. CspC crystallization trials ........................................................................................................................... 114
  4.2.5. CspC bifermentans crystallization optimization ......................................................................................... 114
  4.2.6. CspC bifermentans crystal cryoprotection ................................................................................................. 115
  4.2.7. CspC bifermentans data collection ........................................................................................................... 115

4.3. Results ........................................................................................................................................................... 115
  4.3.1. Optimization of CspC production in E. coli ................................................................................................. 115
  4.3.2. CspC bifermentans performs better in crystallization trials than CspC difficile ............................................. 116
  4.3.3. CspC bifermentans crystals nucleate readily and remain small ...................................................................... 117
  4.3.4. CspC bifermentans crystals require a gentle cryoprotection protocol .......................................................... 117
  4.3.5. Crystals contain multiple lattices, making data processing challenging ......................................................... 118

4.4. Discussion ......................................................................................................................................................... 118

4.5. Figures ............................................................................................................................................................. 121

CHAPTER 5: FUTURE DIRECTIONS ..................................................................................................................... 128

5.1. Figures ............................................................................................................................................................. 132
LIST OF TABLES

Table                                                                 Page

TABLE 2.9.S1.1.: PRODOMAIN INTERACTION INTERFACE WITH MATURE
SUBTILASE: HYDROGEN BONDS AND SALT BRIDGES. ........................................... 79
TABLE 2.9.S1.2.: JELLYROLL DOMAIN INTERACTION INTERFACE WITH MATURE
SUBTILASE: HYDROGEN BONDS AND SALT BRIDGES. ........................................... 80
TABLE 2.9.S1.3.: JELLYROLL DOMAIN INTERACTION INTERFACE WITH
PRODOMAIN: HYDROGEN BONDS. ........................................................................ 80
TABLE 2.9.S2.: STRAINS AND PLASMIDS USED IN THIS STUDY............................. 81
TABLE 2.9.S2.: PRIMERS USED IN THIS STUDY ..................................................... 83
TABLE 2.9.S3.: DATA COLLECTION AND REFINEMENT STATISTICS ....................... 85

TABLE 3.6.1. STRAINS AND PLASMIDS USED IN SLEC STUDIES ............................ 107
TABLE 3.6.2. PRIMERS USED IN SLEC STUDIES ...................................................... 108
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.6.1</td>
<td>SCHEMATIC OF COLONIC LUMEN DURING <em>CLOSTRIDIUM DIFFICILE</em> INFECTION.</td>
<td>12</td>
</tr>
<tr>
<td>1.6.2</td>
<td>SCANNING ELECTRON MICROGRAPHS OF DORMANT AND GERMINATING SPORES.</td>
<td>13</td>
</tr>
<tr>
<td>1.6.3</td>
<td>CARTOON REPRESENTATION SHOWING LAYERS OF THE SPORE.</td>
<td>14</td>
</tr>
<tr>
<td>1.6.4</td>
<td>STANDARD AND PROPOSED RECLASSIFICATION OF <em>C. DIFFICILE</em></td>
<td>15</td>
</tr>
<tr>
<td>2.7.1</td>
<td>THE CSPBA FUSION PROTEIN UNDERGOES PROCESSING DURING SPORULATION.</td>
<td>51</td>
</tr>
<tr>
<td>2.7.2</td>
<td>CSPB UNDERGOES AUTOPROCESSING IN A POSITION-DEPENDENT MANNER.</td>
<td>53</td>
</tr>
<tr>
<td>2.7.3</td>
<td>OVERALL STRUCTURE OF CSPB <em>PERFRINGENS</em>.</td>
<td>55</td>
</tr>
<tr>
<td>2.7.4</td>
<td>THE JELLYROLL DOMAIN CONFORMATIONALLY RIGIDIFIES CSPB <em>PERFRINGENS</em>.</td>
<td>57</td>
</tr>
<tr>
<td>2.7.5</td>
<td>DUAL SALT BRIDGES ARE REQUIRED FOR PRODOMAIN INTRAMOLECULAR CHAPERONE ACTIVITY.</td>
<td>59</td>
</tr>
<tr>
<td>2.7.6</td>
<td>C-TERMINAL PRODOMAIN RESIDUES STERICALLY OCCLUDE A CATALYTICALLY COMPETENT ACTIVE SITE.</td>
<td>61</td>
</tr>
<tr>
<td>2.7.7</td>
<td>THE JELLYROLL DOMAIN AND CATALYTIC SERINE OF CSPBA ARE REQUIRED FOR EFFICIENT GERMINATION.</td>
<td>63</td>
</tr>
<tr>
<td>2.7.8</td>
<td>CSPBA ACTIVITY DOWNSTREAM OF AUTOPROCESSING IS REQUIRED FOR EFFICIENT SLEC CLEAVAGE.</td>
<td>65</td>
</tr>
<tr>
<td>2.8.1</td>
<td>CSP PROTEASE AND SLEC ARE REQUIRED FOR SPORE GERMINATION IN <em>CLOSTRIDIUM</em> SP.</td>
<td>67</td>
</tr>
<tr>
<td>2.8.2</td>
<td>CLUSTALW SEQUENCE ALIGNMENT OF CSPBA PROTEINS.</td>
<td>69</td>
</tr>
<tr>
<td>2.8.3</td>
<td>CLUSTALW SEQUENCE ALIGNMENT OF MAJOR CLASSES OF BACTERIAL SUBTILISIN-LIKE PROTEASES.</td>
<td>71</td>
</tr>
<tr>
<td>2.8.4</td>
<td>CLUSTALW SEQUENCE ALIGNMENT OF DIVERSE CSP PROTEASES.</td>
<td>72</td>
</tr>
<tr>
<td>2.8.5</td>
<td>INTERACTION INTERFACES WITHIN CSPB <em>PERFRINGENS</em>.</td>
<td>74</td>
</tr>
<tr>
<td>2.8.6</td>
<td>CSP PRODOMAIN TRANSCOMPLEMENTATION.</td>
<td>76</td>
</tr>
<tr>
<td>2.8.7</td>
<td>ELECTRON DENSITY MAPS OF CSPB <em>PERFRINGENS</em>.</td>
<td>77</td>
</tr>
<tr>
<td>3.5.1</td>
<td>SCHEMATIC ILLUSTRATING THE SLEC CONSTRUCTS DESIGNED FOR CRYSTALLIZATION.</td>
<td>101</td>
</tr>
<tr>
<td>3.5.2</td>
<td>SCHEMATIC ILLUSTRATING THE THREE FORMS OF SLEC RELEVANT TO <em>C. DIFFICILE</em> SPORE GERMINATION.</td>
<td>103</td>
</tr>
<tr>
<td>3.5.3</td>
<td>POND PREDICTION OF ORDERED AND DISORDERED REGIONS FOR FULL-LENGTH <em>C. DIFFICILE</em> SLEC.</td>
<td>104</td>
</tr>
<tr>
<td>3.5.4</td>
<td>STRAIN #495 (PRE/PRO-SLEC-PG) COOMASSIE GEL ILLUSTRATING SLEC CONSTRUCT NINTA PURIFICATION PROCESS.</td>
<td>105</td>
</tr>
<tr>
<td>3.5.5</td>
<td>SLEC CRYSTALLIZATION TRIALS.</td>
<td>106</td>
</tr>
<tr>
<td>4.5.1</td>
<td>CSP EXPRESSION SCHEMATICS.</td>
<td>121</td>
</tr>
<tr>
<td>4.5.2</td>
<td>CSPC GERMINATION-NULL MUTANTS MAPPED TO CSPB STRUCTURE ILLUSTRATES CLUSTERING AROUND ACTIVE SITE.</td>
<td>122</td>
</tr>
<tr>
<td>4.5.3</td>
<td>PREDICTIONS OF NATURALLY DISORDERED REGIONS FOR CSPC FROM <em>CLOSTRIDIUM DIFFICILE</em>.</td>
<td>123</td>
</tr>
<tr>
<td>4.5.4</td>
<td>CSPC <em>BIFERMENTANS</em> PURIFICATION PROCESS.</td>
<td>124</td>
</tr>
</tbody>
</table>
FIGURE 4.5.5. CSPC BIFERMENTANS CRYSTAL PROGRESSION ........................................ 126
FIGURE 4.5.6. SIMPLIFIED VIEW OF CSPB CRYSTALLIZATION DIMER INTERFACE
WITH POTENTIAL CSPC LIGAND BINDING SITE SHOWN IN BLUE ................................. 127

FIGURE 5.1.1. DIAGRAM ILLUSTRATING THE PROPOSED MODEL FOR
GERMINATION INITIATION IN C. DIFFICILE, ILLUSTRATING PROTEIN
PARTNERS STUDIED IN PREVIOUS CHAPTERS OF THIS THESIS. ................................. 132
CHAPTER 1: INTRODUCTION

1.1. *Clostridium difficile*-associated disease

The Gram-positive bacterium *Clostridium difficile* causes about 300,000 infections annually in the US, costing the health care system an estimated $2-5 billion per year, with each patient costing between $3,000 and $15,000[1, 2-4]. *C. difficile* is the leading cause of both antibiotic-associated diarrhea and hospital-acquired diarrhea in part because *C. difficile* is naturally resistant to many first line broad-spectrum antibiotics. Indeed, infections are often only symptomatic in individuals who have recently undergone antibiotic therapy. Diarrheal symptoms will often develop during, or up to a few weeks following, a course of antibiotics (clindamycin, metronidazole, penicillins, fluoroquinolones and cephalosporins have all been associated with subsequent first-time *C. difficile* infections) and worsen from mild to severe diarrhea[5]. Disease symptoms are caused by two toxins secreted by vegetative cells in the colon. The first, cytotoxic Toxin B, TcdB, requires access to the basolateral layer as receptors for its activity are only located there. The second, TcdA, Toxin A, an enterotoxin, does not have this requirement and it is thought to be the predominant toxin causing early symptoms, until more severe infection has allowed Toxin B to have an effect, though this is still debated[6-8]. However, TcdB alone is sufficient to cause disease symptoms and frequently, infective strains of *C. difficile* are found to express TcdB but not TcdA, though previous results have identified TcdA as more capable of causing inflammation, adding to the confusion[9, 10]. The two toxins are the main virulence factors for *C. difficile*, but an additional toxin, CDT binary toxin, is one of several auxiliary virulence factors[6]. See Figure 1.6.1 for an illustration of a *C. difficile* infection in a human colon environment.
Infections are difficult to treat due to the natural resistance of \textit{C. difficile} to many antibiotics and also its tendency to recur after the initial infection is treated\cite{11}. The typical first choice in treatment of \textit{C. difficile} infection is metronidazole (especially for mild-to-moderate disease) or vancomycin (for more severe disease)\cite{12}. However, even with successful clearance of the infection following a course of broad-spectrum antibiotics, metronidazole or vancomycin will also disrupt the populations of many beneficial colonic commensals, leaving the individual with what is called a “permissive” colon, susceptible to recurrent infection, which occurs in 20-33\% of cases, with more virulent strains associated with greater risk of recurrence\cite{13}. Each re-infection leaves the patient more susceptible to future infections, often increasing in severity, and 3-6\% of \textit{C. difficile} infections will result in death\cite{14}. Fidaxomicin is a newer macrolactam drug with low absorption from the GI tract into the bloodstream (like vancomycin), high specificity for \textit{C. difficile}, and limited activity against other species, making it much less disruptive to the colonic population\cite{12}. Infections that are treated with fidaxomicin are significantly less likely to lead to recurrences\cite{15-17}. Severe \textit{C. difficile} infections (CDI) cause complications such as pseudomembranous colitis, toxic megacolon, and the most severe complication, fulminant colitis, which occurs in about 3\% of CDI and often requires colectomy to treat\cite{6, 18}. Recurrent \textit{C. difficile} infections require different treatment approaches, and for the third and subsequent infections the treatment is different than standard treatment used during first and second infection\cite{14, 19}. Fecal microbiota transplantation is now recognized as a treatment for recurrent \textit{C. difficile}, with impressive cure rates as high as 90\%\cite{20}. Stool from a healthy donor can provide the microorganisms necessary to help restore colonic balance and prevent reinfection\cite{6, 20-22}. 
1.2. The role of spores during *C. difficile* infections

The ability of *C. difficile* to initiate disease depends on its ability to form the metabolically dormant spore-form[23]. This is because *C. difficile* is an obligate anaerobe and thus cannot survive in the presence of oxygen. *C. difficile* infections begin when spores are ingested by a susceptible individual and germinate in response to a specific bile salt, taurocholate, found in the large intestine. During normal *C. difficile* growth in the colon, the bacterium strongly induces sporulation, and the host excretes large amounts of spores in the fecal matter[24, 25]. Though the sigma factor signaling cascade required to complete sporulation has recently been delineated, it remains poorly understood how the sporulation pathway is initiated in *C. difficile*[24, 26]. Excreted spores can contaminate environments such as hospital rooms and persist despite environmental conditions and disinfection protocols, thus serve as important reservoirs for *C. difficile*[25].

Spores can persist for long periods of time due to their intrinsic resistance to many physical insults such as extreme pH and temperature shifts, boiling, and desiccation. This intrinsic resistance is partially due to the thick peptidoglycan cortex layer of the spore and their metabolic dormancy[27-29]. In studies evaluating hospital disinfection protocols, ethanol, HiBiscrub product, Steri-7, and related products were not successful disinfectants; effective disinfection requires 1% sodium hypochlorite solutions or 10% hydrogen peroxide, or products such as Spor-Klenz[27, 30]. Spores of *Bacillus subtilis*, considered a model organism for spore germination, have even been exposed to outer space and still been able to effectively germinate[31]. *C. difficile* can also infect some livestock and human companion animals, including sheep, pigs, hamsters and dogs. Furthermore, while no reports have implicated it as a food poisoning species, it has been found in up to 42%
of meat products (beef, pork, and turkey) in the USA[32]. Thus, it is entirely possible that it could be introduced to the colon by way of the food we eat[30, 33, 34]. The food ingestion route is difficult to prove as *C. difficile* is frequently found in the environment, and recently it has been shown that non-susceptible individuals can be colonized asymptomatically and they may act as carriers of the bacteria. Estimates for the numbers of asymptomatic carriers vary, but rates of carriage could be around 2-7% of the healthy population[35], higher in those that are hospitalized (~35%) or in long-term care facilities or extended hospitalization (~50%)[36, 37]. Many newborns are briefly colonized by *C. difficile* in the process of establishing colonic flora and estimates vary widely as to what percentage of babies are colonized by *C. difficile*, ranging from 2% to 90%[38-40]. It is thought that a transient *C. difficile* infection is a normal part of acquiring gut commensals[41-43]. By age three, children typically have acquired balanced and diverse microbiota, similar to that of adults., which is protective against *C. difficile*-associated disease.

1.3. The role of the microflora in regulating *C. difficile* spore germination

Normal human digestion requires large and varied populations of gut commensals, which live in the large colon and help to complete the digestive process by further breaking down the compounds that enter the large bowel from the small intestine. The human diet selects for the species of bacteria that live in the colon and can help the digestive process[44]. A disruption of this balanced microcosm (dysbiosis) can be brought on by use of antibiotics, or found in people with disrupted colonic microbial populations, such as those with inflammatory bowel diseases[45]. Studies have shown that patients
with inflammatory bowel diseases are more likely to contract *C. difficile* infections due to the dysbiosis associated with the disease state and are more likely to have more severe outcomes during infection[45].

Levels of the bile salt germinant taurocholate are elevated in the dysbiotic colon. Although *C. difficile* germinates in response to the bile salt taurocholate, in healthy individuals, taurocholate is quickly cleaved into its constituents, cholate and taurine, by resident bacteria[46]. Thus the presence of uncleaved taurocholate in the large bowel is thought to be the signal for *C. difficile* of a dysbiotic colon and thus a suitable host[47-49]. Additionally, *C. difficile* spore germination is inhibited by the presence of chenodeoxycholate, a bile acid with structural similarity to taurocholate, but produced in complex with amino acids and only acts as an inhibitor upon cleavage by resident microbes, contrary to signaling by taurocholate[46, 50-52]. Together these bile acids form a two-way system by which *C. difficile* can detect colons without a robust microbe population. *C. difficile* appears to favor a colonic environment with reduced diversity, such as one that has been recently decimated by antibiotics, and into which only relatively few species have established themselves, since byproducts of limited bacterial metabolism seem to be associated with increased germination rates[53, 54]. A colon with a healthy microbiome will not only lack the germinant, it will also have the competitive inhibitor present in the same location. The process by which spores germinate is essential to infection[55] and thus a very important area for research in *C. difficile* pathogenesis.

1.4. The molecular basis for spore germination in the Firmicutes
Spores consist of several discrete layers (see Figure 1.6.2 for an electron micrograph of dormant and germinating *C. difficile* spores; see Figure 1.6.3 for a cartoon drawing of a dormant *C. difficile* spore). The innermost layer, the core, contains the dehydrated and densely packed components required to become a vegetative cell. Outside of the core is an inner membrane consisting of largely immobile lipids that is impermeable to most small molecules, even water[56, 57]. The germ cell wall surrounds this membrane and will eventually become the cell wall of the new vegetative cell. Outside of these structures is a thick cortex composed of modified peptidoglycan. The cortex is believed to be the most important spore structure conferring resistance to physical insults. Outside of the cortex is outer membrane, which is surrounded by a proteinaceous coat and the outermost layer, the exosporium. There are many good reviews that describe the structural features of *Clostridium difficile* spores and the similarities and differences between *C. difficile* and the model organism for the spore-forming bacteria, *Bacillus subtilis*[25, 28, 29, 58, 59]. Germination of spores can be thought of as two critical events: initiation of germination by sensing of a small molecule germinant, and hydrolysis of the cortex layer of the spore. Cortex hydrolysis must be tightly regulated to avoid loss of resistance in inhospitable environments as well as hydrolysis of vegetative cell peptidoglycan, which would cause lysis of the sporulating cell.

Spores will only germinate in response to precise signals. *Bacillus subtilis* can germinate in response pressure, dipocolinic acid, or with small nutritional germinants such as sugars and amino acids, sensed via germination receptors (Ger family) found in the inner membrane of the spore[25, 56, 60]. *Clostridium perfringens* spores also germinate in response to amino acids and ions, specifically L-asparagine and potassium chloride
binding to Ger family receptors or detection of calcium dipicolinic acid, which is secreted by germinating spores[61].

Germination initiation by Ger receptors involves release of calcium dipicolinic acid from the spore, and this release triggers cortex hydrolysis in both *C. perfringens* and *B. subtilis*[56]. It is interesting that calcium dipicolinic acid is released during germination, as both *C. perfringens* and *B. subtilis* spores will germinate in response to this, suggesting that germination of some spores will trigger germination of others. The Ger A and B receptors are located inside the spore along the inner membrane, with Ger C located more peripherally[56, 62]. However, no homologues to the Ger family of receptors are found in *C. difficile* and thus for a long time it was unknown how *C. difficile* could sense the germination signal[56, 63]. Activation of the cortex hydrolase must rely on some external signal, as discussed above, to avoid germination in inhospitable environments, but due to the lack of Ger receptors this initiation in *C. difficile* must happen differently than in *B. subtilis* or *C. perfringens*.

Despite the lack of Ger receptors, cortex hydrolysis machinery appears to be fairly well conserved between *C. perfringens* and *C. difficile*. *C. perfringens* has two cortex-lytic enzymes, SleC and SleM, though only SleC is required for complete hydrolysis of the cortex, while SleM degrades the leftover fragments[64]. SleC consists of four domains and exists in spores as an inactive zymogen, pro-SleC, requiring activation by a germination-specific protease to become an active hydrolyase[61].

Shio Makino’s group published most of the foundational work on spore cortex lytic enzymes in *Clostridium perfringens*, beginning in the mid 1990s, when they identified an enzyme that was released during germination of *Clostridium perfringens*.
spores[65]. They were the first to discover and report that SleC in *C. perfringens* was produced in a pre-pro-form and is inactive in the dormant spore due to the associated prodomain[66]. This was demonstrated by Western blotting using antibodies against SleC: in dormant spores, a 36-kD protein was present, and in germinating spores there were both the 36-kD protein and a smaller 31-kD protein[66]. A few years later they were able to demonstrate that, in addition to the pre-domain associated only in sporulating cells, SleC includes an extended C-terminal sequence of about 25 amino acids, which is also processed during sporulation, and they suggested that either or both of these could be required for proper translocation to the outer layers of the cortex during spore formation[67]. Thus, when spores of *C. perfringens* are completely formed but have not yet received signal to germinate, SleC is located in the outer region of the cortex (as shown by Miyata *et al.* using colloidal gold particles[68]) and exists in a form containing a 35-amino acid propeptide as well as the 264-amino acid hydrolase domain[67] but from which the 115-amino acid predomain, and the 25-amino acid C-terminal extended sequence (called a C-terminal prosequence in some literature), have both been cleaved[66]. From this stage, activation of SleC is required, by way of cleavage of the inhibitory prodomain.

Activation of SleC will result in cortex hydrolysis, core hydration and expansion, and eventually outgrowth to a vegetative cell. Makino’s group then demonstrated both that the N-terminal 115-amino acid predomain, though cleaved, remains associated with the pro-SleC under non-reducing conditions (predomain-pro-SleC), and that the predomain has both stabilization and chaperone effects, being required for proper folding and thermostability of the enzyme and preventing pro-SleC—pro-SleC dimerization[69]. At
this point, they hypothesized that the function of the predomain is in folding and stability, and the function of the propeptide is in enzyme inhibition, until an activation signal is received and cleavage of the prodomain occurs[69]. SleC expression, processing and regulation is in contrast to the cortex fragment-lytic enzyme, CFLE, called SleM in *C. perfringens*, which is produced and exists in dormant spores in its mature, active form[64, 67]. Though it exists in an active form in dormant spores, SleM’s cortex-fragment substrate is not present until after SleC has begun cortex hydrolysis[64, 67].

The germination-specific protease (GSP) implicated in germination was thought to be responsible for the cleavage and activation of SleC[67]. Most strains of *C. perfringens* encode three such germination-specific subtilisin-like serine proteases, CspA, CspB and CspC[70]. In 2001 it was shown that purified GSPs (containing CspA, CspB, and CspC), when incubated with the predomain-pro-SleC complex, were able to generate active SleC enzyme with a molecular weight of 31 kD by cleaving between Val-149 and Val-150 in the junction between the prodomain and the SleC hydrolase[71]. Thus, the Csps are required for activation of the SleC cortex hydrolase, downstream of detection of germinants, though it is not yet completely understood how the signal passes between sensing germinants and activation of Csps. In *C. perfringens*, only CspB is required for the activation of SleC, while CspA and CspC are thought to have the same, if redundant, function[56, 70, 72]. Until recently, very little was known about germination in *C. difficile* though the process was thought to be similar to germination in *C. perfringens* and other related species.
1.5. The molecular basis for *C. difficile* spore germination

Notably, *C. difficile* lacks *ger* family receptors, as previously discussed, making it unclear how the germination cascade begins in a new infection. Bile acids have been implicated for almost 40 years in causing germination of *C. difficile* isolates[49], and later it was discovered that taurocholate was the specific bile acid that spores responded to, in combination with glycine or other amino acids[73]. It was demonstrated a few years ago that germination in *C. difficile* proceeds at a rate suggestive of a specific germinant receptor, while a genetic screen enabled identification of CspC as the specific protein required[47, 74]. Several single amino acid substitutions within CspC rendered *C. difficile* unable to respond to taurocholate[47]. CspC is a subtilisin-like pseudoprotease, whose family members CspA and CspB are implicated in the germination cascade as activating the germination-specific cortex hydrolase SleC, required for germination. Unlike *C. perfringens* and other *Clostridium* spp. [75], *C. difficile* CspA and CspC each contain two single amino acid substitutions within their catalytic triads that render them inactive[76]. Only CspB retains its protease activity[76], but *C. difficile* CspB is expressed as a fusion to catalytically-inactive CspA, an arrangement only observed in other Peptostreptococcaceae family members, while organisms that possess three active CspS typically express them as separate polypeptides[75]. While CspC and CspA did not have any known function, they still were required for germination, as shown by our lab and others[47].

The concept of pseudoproteases has recently been found in many different biological pathways[77-79]. Often, these catalytically dead proteins are as essential as
their active counterparts, and often, a single amino acid substitution interrupts the activity, while retaining the overall structure of the protein, pseudoproteases could play a role in regulation—substrates still bind in the substrate binding cleft, but they cannot be cleaved[77-79].

It has not yet been observed whether germinants are able to directly bind CspC or if an intermediate is needed. Additionally, while it is thought that the germination signal is relayed through CspC’s sensing of germinant, activation of CspB, and CspB’s activation of inactive SleC and the initiation of cortex hydrolysis, none of these interactions have been structurally observed.

We sought to use X-ray crystallographic studies to determine the three dimensional structure of the key players in the germination cascade and use this information to guide future studies that could demonstrate their interaction or illustrate the signaling cascade at a molecular level. We were able to determine the crystal structure of CspB from *Clostridium perfringens* to 1.6 Angstroms (Å) and learn about function from form[76], and have made strides towards the structure of CspC from closely-related *Clostridium bifermentans* (unpublished data), which we hope to be able to crystallize in complex with associated germinant or inhibitor to observe how these interactions take place.

A close understanding of this process could aid in design of therapeutics, as the ability to influence germination could control infections, as germination is required to begin infection. Structure-based drug design is a burgeoning field and these methods could be applied for the treatment of *Clostridium difficile* infections, with control of germination being a suitable target for therapeutic development.
1.6. Figures

Figure 1.6.1. Schematic of colonic lumen during *Clostridium difficile* infection.

Illustration of the colonic epithelium in a *C. difficile* patient, displaying germination of spores, toxin production, effects of toxins and formation of pseudomembrane; sporulation and transmission of spores. This figure was adapted from Shen 2012[9].
Figure 1.6.2. Scanning electron micrographs of dormant and germinating spores.

(A): A scanning electron micrograph of a dormant *C. difficile* spore. Components of the spore as discussed in the text are labeled. (B) Scanning electron micrograph of a germinating *C. difficile* spore, fixed for imaging 10 minutes after germinant was added. Thinning of cortex layer and expansion of core are visible. Labeled components of the spore as above. Images credit: Emily Putnam.
Figure 1.6.3. Cartoon representation showing layers of the spore.

Cartoon representation illustrating the layers of a *C. difficile* spore, as discussed in the text. Spore core, central, shown as a deep gray, surrounded by the inner spore membrane and the germ cell wall, each shown as a black line. The cortex, shown in a pale tan color, is outside of the three inner structures, and inside of the outer spore membrane and the proteinaceous coat. The exosporium is not pictured, but is an additional loose-fitting layer outside of the spore coat.
### Figure 1.6.4. Standard and proposed reclassification of C. difficile

The standard classification of *C. difficile*, shown in the top panel, was order Clostridiales, family Clostridiaceae, genus *Clostridium* and many species were included within this genus, some quite distantly related. A proposed reclassification, shown in the lower panel, suggested by Yutin *et al.*[75], adds the family Peptostreptococcaceae, and the genus *Peptoclostridium* to include the species *difficile, bifermentans, sordellii, and bartlettii*, among others, while some *Clostridium* spp. remain as they were previously classified, including the species *perfringens, tetani, butyricum, and botulinum*, among others.

<table>
<thead>
<tr>
<th>ORDER</th>
<th>FAMILY</th>
<th>GENUS</th>
<th>SPECIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clostridiales</td>
<td>Clostridiaceae</td>
<td>Clostridium</td>
<td>tetani, perfringens, butyricum, butyricum, botulinum, difficile, bifermentans, sordellii, bartlettii</td>
</tr>
<tr>
<td>Clostridiales</td>
<td>Peptostreptococcaceae</td>
<td>Peptoclostridium</td>
<td>difficile, bifermentans, sordellii, bartlettii</td>
</tr>
</tbody>
</table>

The standard classification of *C. difficile*, shown in the top panel, was order Clostridiales, family Clostridiaceae, genus *Clostridium* and many species were included within this genus, some quite distantly related. A proposed reclassification, shown in the lower panel, suggested by Yutin *et al.*[75], adds the family Peptostreptococcaceae, and the genus *Peptoclostridium* to include the species *difficile, bifermentans, sordellii, and bartlettii*, among others, while some *Clostridium* spp. remain as they were previously classified, including the species *perfringens, tetani, butyricum, and botulinum*, among others.
CHAPTER 2: STRUCTURAL AND FUNCTIONAL ANALYSIS OF THE CspB PROTEASE REQUIRED FOR CLOSTRIDIUM SPORE GERMINATION

2.1. Introduction

The Gram-positive, spore-forming obligate anaerobe Clostridium difficile is the leading cause of nosocomial diarrhea worldwide [11, 80, 81]. The symptoms of C. difficile-associated disease (CDAD) range from mild diarrhea to pseudomembranous colitis and even death. Although CDAD is primarily a toxin-mediated disease [9, 81], the high cost and difficulty in treating C. difficile infections largely arises from its ability to form endospores [82, 83]. Because spores are metabolically dormant and intrinsically resistant to harsh physical insults [27, 28, 59, 81], they allow C. difficile to resist antibiotic treatment and persist in healthcare-associated settings. Thus, spores are the primary vectors for transmission [23] and the cause of recurrent infections, the latter of which occurs in ~25% of cases and can lead to severe CDAD [83, 84].

In order to initiate an infection, C. difficile spores ingested from the environment must germinate into toxin-producing vegetative cells in the intestinal tract [11, 81, 85]. Similar to other spore-forming bacteria, C. difficile spores germinate specifically in response to small molecules known as germinants [56, 86]. For C. difficile, these germinants are bile salts, which are abundant in the small intestine [48, 50, 55]. While germinants have been identified for a number of bacterial species, the molecular events that occur upon germinant sensing remain poorly characterized [56, 86, 87]. Shortly after germinant addition, cortex hydrolases become activated and degrade the spore cortex, a thick protective layer of modified peptidoglycan. Because the cortex maintains the spore in a dehydrated, metabolically dormant state, the removal of this physical constraint is
essential for germination to proceed and metabolism to resume in the spore core [56, 86, 87]. Nevertheless, despite the importance of cortex hydrolysis, little is known about the molecular mechanisms that regulate cortex hydrolase activity.

In the Clostridia, the primary cortex hydrolase appears to be SleC, since disruption of sleC in both C. difficile [88] and the related foodborne pathogen C. perfringens [89] results in a severe germination defect. In C. perfringens, SleC undergoes several processing events. During sporulation, the N-terminal prepeptide is removed to produce pro-SleC, which consists of an N-terminal propeptide attached to the hydrolase domain [65-67, 90]. During germination, the zymogen pro-SleC is cleaved at a conserved site to release the propeptide (Figure 1A); this event appears to activate its hydrolase activity [65, 91].

Biochemical analyses of C. perfringens germination exudates have shown that a fraction containing three serine proteases (CspA, CspB, and CspC) can proteolytically activate SleC hydrolase activity in vitro [91]. CspB alone appears sufficient to activate SleC, since the food-poisoning isolate SM101 encodes only the cspB gene, and disruption of this gene abrogates SleC cleavage and spore germination [70]. In the genome of C. difficile, three csp homologs are present in a bicistronic operon (cspBA-cspC, Figure S1), with cspB and cspA being present as a gene fusion [56]. Since disruption of the cspBA-cspC operon by transposon insertion results in a severe germination defect [92], cortex hydrolysis in C. difficile and C. perfringens would appear to be similarly regulated.

While studies have shown that SleC and CspB are key players during germination, the molecular mechanisms regulating their function are unknown. The sequence homology between Csp proteases (Csps) and the subtilase protease family [91] provides a starting
point for understanding how Csp transduce the germination signal and activate SleC. Subtilases are serine proteases that contain a catalytic triad in the order of Asp, His and Ser [93, 94], and most subtilases are produced as pro-enzymes that autoproteolytically remove their prodomain [93, 95, 96]. While Csp purifed from *C. perfringens* germination exudates are N-terminally processed [91], whether Csp are regulated in a manner analogous to other subtilases is unclear. Indeed, whether Csp actually have protease activity has not yet been directly demonstrated.

In this study, we investigated the protease activity of CspBA in *C. difficile*. By analyzing CspBA in sporulating *C. difficile* and purified spores, we demonstrate that CspBA is processed to CspB during spore assembly and that CspB undergoes autoprocessing. We also present the first crystal structure of the conserved Csp family of proteases at 1.6 Å resolution and define its key structural domains. These biochemical and mutational analyses reveal that, in contrast to previously characterized prokaryotic subtilases, wildtype CspB forms a stable complex with its prodomain. Similar to other subtilases [96], the prodomain acts as both an intramolecular chaperone and an inhibitor of CspB protease activity. These findings provide the first molecular insight into Csp function and may inform the development of strategies that can either prematurely activate *C. difficile* spore germination in the environment or prevent spore germination during disease transmission and recurrence.
2.2. Materials and Methods

2.2.1. Bacterial growth conditions

Bacterial strains and plasmids used in this study are listed in Table S2 in Text S1. The *C. difficile* strains are isogenic with the erythromycin-sensitive strain JIR8094 [97], a derivative of the sequenced clinical isolate 630 [63]. *C. difficile* strains from freezer stocks were grown on BHIS agar plates [98] supplemented with and 0.1% sodium taurocholate (BioSynth International). To induce sporulation, *C. difficile* strains were grown on 70:30 agar plates (63 g BactoPeptone, 3.5 g Protease Peptone, 11.1 g BHI, 1.5 g yeast extract, 1.1 g Tris base, 0.7 g NH₄SO₄ per liter). Media for *C. difficile* were supplemented with 10 µg thiamphenicol (Thi) mL⁻¹, 50 µg kanamycin (Kan) mL⁻¹, 8 µg mL⁻¹ cefoxitin (TKC); 10 µg thiamphenicol mL⁻¹; or 5 µg erythromycin mL⁻¹ (Erm) as needed. *C. difficile* strains were maintained at 37°C in an anaerobic chamber (Coy Laboratory Products) with an atmosphere of 10% H₂, 5% CO₂, and 85% N₂. *E. coli* strains were grown at 37ºC in Luria-Bertrani (LB) broth. Antibiotics were used at 100 µg mL⁻¹ carbenicillin for pET22b, 30 µg mL⁻¹ kanamycin for pET28a, and 20 µg mL⁻¹ chloramphenicol for pMTL83151 and pMTL84151 vectors in DH5α *E. coli* strains, and 100 µg mL⁻¹ and 20 µg mL⁻¹ in HB101 *E. coli* strains.

2.2.2. Sporulation Assay

*C. difficile* strains were inoculated from frozen stocks onto BHIS plates containing 0.1% taurocholate. After 24 hr growth, a heavy streak of the strain was transferred to a 70:30 plate and spread uniformly across the plate. Whereas <0.1% of cells are sporulating on BHIS plates [2], ~25% of cells undergo sporulation at the timepoints analyzed in this
study as determined by phase-contrast microscopy similar to Burns et al. [85] (Figure 1C). While the induction of sporulation occurs at different rates within the population, this assay allows us to produce relative high rates of sporulating cells. At the indicated timepoints, cells were scraped from the plate and resuspended in PBS. The cells were pelleted and then resuspended in PBS. For Western blot analysis, 50 μL of the cell resuspension was removed, and the sample was frozen at -80°C. The remainder of the sample was analyzed by phase contrast microscopy to assess the progression of sporulation.

2.2.3. Preparation of C. difficile samples for Western blot analysis

For sporulating C. difficile samples, cell pellets harvested from the sporulation assay were subject to three cycles of freeze-thaw. On the final thaw, 100 μL of EBB buffer (9 M urea, 2 M thiourea, 4% w/v SDS, 10% v/v β-mercaptoethanol) was added, and the sample was boiled with occasional vortexing for 20 min. The lysate was pelleted for 5 min at 13,000 x g and then resuspended; 7 μL of 4X loading buffer (40% v/v glycerol, 0.2 M Tris pH 6.8, 20% v/v β-mercaptoethanol, 12% SDS, 0.4 mg/mL bromophenol blue) was added. The sample was boiled again for a minimum of 5 min, pelleted at 13,000 x g, and 15 μL was resolved on a 7.5% (for analysis of CspB in sporulating cells) or an 11 or 12% SDS-polyacrylamide gel (for analyses of purified spores).

For analyses of purified spores, ~5 x 10^6 spores were pelleted at 15,000 x g for 5 min. The spore pellet was resuspended in 50 μL EBB buffer, boiled for 20 min with periodic vortexing, pelleted at 13,000 x g for 5 min, and resuspended to further solubilize proteins. Five μL of 4X loading buffer was added, and the sample was boiled for 5 min.
After pelleting at 13,000 x g, 10-15 µL of the sample was resolved on an 11% or 12% SDS-polyacrylamide gel.

2.2.4. Western blot analysis

All antibodies used in this study were raised in rabbits by CoCalico Biologicals, with the exception of the SleC antibody, which was raised in rabbits by Pacific Immunology. The antigens used were His<sub>6</sub>-tagged CspB(1-548 aa), full-length His<sub>6</sub>-tagged CspB <i>perfringens</i>, His<sub>6</sub>-tagged CspC, His<sub>6</sub>-tagged SleC and His<sub>6</sub>-tagged CD1433. Samples resolved by SDS-PAGE were transferred to Immobilon-FL PVDF membranes (Millipore). The membranes were blocked in 50:50 PBS:LiCOR blocking buffer (LiCOR) for 30 min, after which Tween 20 was added to 0.1% v/v, and polyclonal antisera was added at 1:1,000 for all antibodies with the exception of the anti-SleC antibody, which was used at a 1:5,000 dilution. After a minimum of 1 hr incubation with shaking, the membranes were washed a minimum of 3 times in PBS+0.01% v/v Tween. Anti-rabbit secondary antibodies conjugated to IR800 dye (LiCOR) were added at 1:30,000 dilution in 50:50 PBS:LiCOR blocking buffer containing 0.1% v/v Tween and 0.1% v/v SDS then incubated with shaking for 1 hr. The membranes were washed a minimum of 3 times in PBS+0.1% v/v Tween before imaging on an Odyssey Clx scanner (LiCOR). Western blot quantitation was performed using the indicated loading controls and LiCOR ImageStudio software.
2.2.5. Purification of *C. difficile* spores

Sporulation was induced for 3 - 4 days on five 70:30 plates. Spores and cell debris were scraped off the plate into 1 mL ice-cold sterile water and purified as previously described [98]. Briefly, the sample was subjected to 5 washes in ice-cold sterile water, followed by a HistoDenz gradient purification and 3-5 washes in ice-cold sterile water. Spores were stored at 4˚C in water.

2.2.6. Germination assay

Purified spores were enumerated using disposable semen test counting chambers (InCyto C-Chip). Approximately 5 x 10^7 spores were resuspended in a total volume of 100 μL sterile H_2O. The spores were heat activated at 60˚C for 30 min, cooled for 2 min on ice, then 100 μL of 2X BHIS was added. 100 μL of the spores were removed to a tube containing 2 μL of 10% sodium taurocholate to induce germination. Both samples were incubated at 37˚C for 20 min after which spores were serially diluted 10-fold into PBS. 10 μL of the dilutions was spotted onto either BHIS or BHIS + 0.1% taurocholate agar plates in triplicate and incubated anaerobically at 37˚C for ~24 hr before assessing spore viability. Equivalent numbers of viable spores were recovered on untreated spores plated on BHIS + 0.1% taurocholate plates and taurocholate-treated spores plated on BHIS or BHIS + 0.1% taurocholate plates. Because spore clumping increased the variability in counting spores, CD1433 [99] was used as a loading control in some Western blot analyses.
2.2.7. Crystallization

Appropriate protein concentrations for crystallization were determined using Pre-Crystallization Test (Hampton Research, Aliso Viejo, CA). Hanging drop crystallization experiments were conducted with CspB (11 mg/mL) in 150 mM NaCl, 10 mM Tris-HCl pH 7.5 and Crystal Screen 2 (Hampton Research). Crystal trays were incubated at 12°C and initial crystal hits in 25% (v/v) ethylene glycol (Condition 3) were discovered within 24 hours. After refinement of crystallization conditions, crystals grew reproducibly to about 100*250*60 µm³ in 27-30% (v/v) ethylene glycol buffered to pH 5 with 50 mM sodium acetate. Crystals grew in space group P2₁2₁2₁, with unit cell dimensions a = 73.87, b = 138.17, and c = 140.08 Å and two molecules in the asymmetric unit for an estimated 57% solvent content [100-102]. As crystallization conditions contained sufficient ethylene glycol to serve as a cryoprotectant, crystals were flash cooled in liquid nitrogen directly from the crystallization drop.

2.2.8. Data Collection

A complete 1.6 Å single-wavelength data set of a representative selenomethionyl-CspB crystal was collected at the selenium edge (0.9794 Å) at 100 K at the General Medical Sciences and Cancer Institutes Structural Biology Facility (GM/CA @ APS) beamline 23ID-B at the Advanced Photon Source, Argonne National Laboratory (Chicago, IL, Table S4 in Text S1).
2.2.9. Data Processing

Data were processed using Denzo and Scalepack [103]. Twelve selenium sites were expected, from 6 methionines in the protein sequence and two predicted molecules in the asymmetric unit, using the Matthews Coefficient program [100, 101] in the CCP4 Program Suite [102]. ShelXC/D/E, also part of the CCP4 Suite, was used to identify the selenium sites and gain initial phase information [102, 104, 105]. The 12 selenium sites and phase information were used in ShelX/E for density modification and generation of the initial phased map (Fig. S7) [104, 105].

2.2.10. Structure Solution and Refinement

The initial model was produced by Phenix.AutoBuild using input phases from ShelX/E [104, 106]. Manual building was performed into the original phased map to reduce model bias. Refinement of the structure was done with manual building and adjustment in COOT [107] and refinement of the latest iteration of the model using Phenix.Refine [106]. All protein and ligand (non-water) B-factors were refined anisotropically. Phenix.AutoBuild with simulated annealing was used after multiple rounds of refinement to gain density for some poorly-resolved loops in the structure, resulting in the placement of several previously missing residues [106]. Ten percent of reflections were set aside for $R_{\text{free}}$ calculation. Model was refined to an $R_{\text{work}}/R_{\text{free}}$ of 0.15/0.18 and Ramachandran statistics were 97.9% in favored regions and 2.1% in allowed regions, with no Ramachandran outliers. 957 water molecules were placed by Phenix.Refine and checked with the Check Waters feature in COOT [106, 107].
Although CspB is a dimer in the asymmetric unit, gel filtration chromatography experiments (see Text S1) indicate that CspB is a monomer in solution. In monomer 1, five residues were not built due to disorder in the electron density map (residues 411-415); in monomer 2, three residues were not built due to disorder (residues 411-413). These residues are part of a small loop located between two strands of the jellyroll domain. Additionally, the first four residues (residues 1-4) of the prodomain in each monomer were disordered and not built. Electron density for the C-terminal His$_6$-tag used for protein purification (see Text S1) was seen in the second monomer only; these residues were stabilized by a crystal-packing interface, thus enabling residues 566-573 to be built in this monomer. Although the presence of calcium in the model was expected because this metal is present in many subtilisin family members [96], elemental analysis did not detect Ca$^{2+}$ in our enzyme preparation (Dartmouth Trace Elemental Analysis Lab, data not shown). Two putative Na$^+$ and three Cl$^-$ atoms (confirmed by sodium iodide soaks) were placed in the model, in addition to ethylene glycol, a crystallization reagent.

2.2.11. Limited Chymotrypsin Proteolysis

Wildtype CspB and its mutant variants were diluted to 15 µM in 10 mM Tris pH 7.5 buffer in a total volume of 150 µL. Twenty-four microliters of the mixture were transferred into 8 well strip tubes. One microliter of chymotrypsin (Sigma, 25-fold concentrate relative to indicated concentration) was added, and the mixture was mixed then incubated for 60 min at 37°C. Chymotrypsin activity was quenched by the addition of 8 µL of 4X loading buffer. The samples were boiled for 3 min at 95°C and then 7 µL was resolved on a 15% SDS-PAGE gel and visualized by Coomassie staining.
2.2.12. Preparation of *E. coli* samples for Western blot analysis

*E. coli* cultures were grown as described for protein purification. One hour after IPTG induction, a 1 mL sample was removed, the OD600 measured, and the sample pelleted at 13,000 x g for 2 min. Cells were lysed in 1X loading buffer (10 OD600/mL). To obtain cleared lysate samples, 30 µL of the supernatant produced upon high-speed centrifugation of sonicated lysates was added to 10 µL of 4X loading buffer. For eluate samples, 30 µL of the eluate was added to 10 µL of 4X loading buffer. All samples were boiled at 95˚C for 5 min, pelleted at 13,000 x g for 5 min, then 2.5 µL of induced and cleared lysate samples or 5 µL of eluate samples were resolved on a 12% SDS-PAGE gel and analyzed by Western blotting.

2.2.13. Activity-based probe labeling of CspB *perfringens*

Wildtype CspB and its mutant variants were diluted to 10 µM in 10 mM Tris-HCl pH 7.5 buffer in a total volume of 155 µL. Twenty-five microliters were aliquoted in triplicate into strip tubes. 0.25 µL of 100 µM FP-Rh (fluorophosphanate-rhodamine probe) was added to CspB and incubated at RT for 10 min. Labeling was quenched by adding 8 µL 4X loading buffer to the sample and boiling at 95˚C for 3 min. Six microliters of the labeling reaction was resolved on a 15% SDS-PAGE gel, and fluorescence was imaged using a Biorad PharosFX scanner.
2.2.14. Accession codes
Coordinates and structure factors have been deposited in the Protein Data Bank (www.rcsb.org) under the accession number 4I0W.

2.3. Supporting Materials and Methods

2.3.1. Strain and plasmid construction for genetic manipulation of C. difficile.
Oligonucleotides used in this study are listed in Supporting Table 3. C. difficile strain 630 DNA was used as a template for PCR amplification. Sequencing was performed at the University of Vermont DNA Analysis Core Facility. To construct the null mutation in the cspBAC locus, the ClosTron method for targeted gene disruption in C. difficile [108, 109] was used. A modified plasmid containing the retargeting group II intron, pCE240 (a gift from C. Ellermeier, University of Iowa), was used as the template with primers #533, 534, and 535 and the EBS Universal primer (#532), as outlined in the TargeTron users’ manual (Sigma Aldrich). The resulting retargeting sequence was digested with BsrGI/HindIII and cloned into pJS107 (a gift from J. Sorg, University of Texas A&M), which is a derivative of pJIR750ai (Sigma Aldrich) with similarity to pMTL007 [108, 109]. The resulting plasmid pJS107-cspBAC-81 was transformed into HB101/pRK24 [97] for conjugation into JIR8094 to generate cspBAC−. Specifically, the E. coli donor strain was grown aerobically in 2 mL of LB supplemented with ampicillin and chloramphenicol until early stationary phase. The culture was pelleted gently at 2,500 x g for 5 min, and the resulting pellet was transferred into the anaerobic chamber. The pellet was resuspended in 1 mL of a C. difficile recipient strain grown anaerobically in BHIS at 37°C with slow shaking until early stationary phase. 100 µL of the cell mixture was spotted onto a single
BHIS agar plate and incubated overnight for 14 – 18 hr after which the cells were scraped into 2.5 mL BHIS. Seven 100 µL drops of the mixture were spotted onto three BHIS/TKC plates. Thiamphenicol-resistant colonies typically appeared with 30 – 48 hr, after which they were streaked onto BHIS plates containing Thi, Kan and 50 µM FeSO₄. This step induces the ferredoxin promoter that controls the expression of the group II intron. The resulting streaks were transferred to BHIS-Erm plates; Erm-resistant colonies typically appeared after 48 hr, after which they were struck to single colonies then screened by PCR using primers #457 and 538, the latter of which binds within the pJS107 vector.

To complement the targetron disruption of cspBAC, a 5364 bp fragment containing the cspBAC operon and its upstream region was amplified using primers #691 and 665. The resulting PCR product was digested with NotI-HF and XhoI (New England Biolabs) and ligated into pMTL83151 [110]. The S461A mutation was introduced using PCR splicing by overlap extension (SOE) [111]. Primer pair #691 and 464 were used to amplify the 5’ SOE product, while primer pair #665 and 465 were used to amplify the 3’ SOE product. The resulting fragments were mixed together, and the flanking #691 and #665 primers were used to amplify the 5364 bp S461A fragment. The Δjelly mutation (corresponds to deletion of aa 260-392) was constructed similar to the S461A mutation, with the exception that the internal SOE primers were #746 (used with #691) and #745 (used with #665). All complementation plasmids were transformed into HB101/pRK24, and the conjugation was performed as described except that after 14 hr incubation on non-selective media, the bacterial mixture was scraped into 1 mL PBS. 100 µL of the mixture was spotted onto a total of five BHIS-TKC plates. Thiamphenicol-resistant colonies
typically appeared after 48 hr incubation, after which they were re-streaked to single colonies. At least two independent clones of each strain constructed were analyzed.

The *sleC* mutant was constructed as previously described by Burns *et al.* [88] using primers #641, 642, and 643, with the exception that pJS107 was used for the targeting vector. Primer pair #461 and 463 were used to screen colonies for clones containing the targetron insertion.

To construct *cspBAC* pMTL84151 complementation constructs, PCR SOE was used to introduce the following sequence, TAGGAGGGATTTATG between the codons for Gln66 and Asp70. This sequence is derived from the end of the *cspBA* gene, including the stop codon, and the start of the *cspC* gene, including the start codon. Insertion of this sequence results in the introduction of a stop codon, ribosome binding site and start codon in between the codons for Gln66 and Asp70. In particular, primers #691 and #951 were used to amplify the 5’ SOE product, while primers #864 and #950 were used to amplify the 3’ SOE product using either pMTL83151-*cspBAC* or pMTL83151-*cspBAC* S461A as the PCR template. The resulting fragments were mixed together, and the flanking #691 and #864 primers were used to amplify a 2790 bp fragment. The fragment was gel-purified, digested with NotI and KpnI and ligated to gel-purified pMTL83151-*cspBAC* digested with the same enzymes. The resulting pMTL83151-Q66 and Q66/S461A plasmids were used as PCR templates for a second PCR reaction, this time using primers #691 and #665. The resulting PCR fragment was gel purified, digested with NotI and XhoI, and ligated to pMTL84151 digested with the same enzymes. The pMTL84151 [110] multicopy vector was used for the transcomplementation because expression of the split prodomain constructs from the pMTL83151 multicopy vector resulted in lower levels of
CspBA relative to wildtype (data not shown). This observation likely reflects a reduced efficiency in prodomain intermolecular chaperone activity relative to intramolecular chaperone activity. pMTL84151 plasmids were conjugated into *C. difficile* using the same method as conjugating pJS107 plasmid constructs.

### 2.3.2. Construction of CspB constructs for recombinant protein purification.

Primers used are listed in Supporting Table 1; strains and plasmids constructed are listed in Supporting Table 2. To construct a strain producing recombinant CspBA (1-548 aa, expression construct used for antibody production), primers #457 and #459 were used to amplify the sequence encoding amino acids 1-548 of CspBA from *C. difficile* 630 genomic DNA. The amplified DNA was digested with NcoI and XhoI, ligated to pET28a, and transformed into DH5α. The resulting pET28a-cspB(1-548 aa) plasmid was transformed into BL21(DE3). To mutate the catalytic Ser of CspBA, PCR SOE was used to introduce the S461A catalytic Ser point mutation into pET28a-cspB(1-548 aa). Primers #457 and 465 were used to amplify a 5’ fragment, and primers #459 and #464 were used to amplify a 3’ fragment. The resulting fragments were used as the templates for a second PCR reaction using the flanking primers, #457 and #459, to amplify *cspB*(1-548 aa) carrying the S461A mutation. A similar strategy was used to mutate the P3-P1 residues of CspBA to produce plasmid pET28a-cspB(1-548)-QTQ/AAA, with the exception that the SOE primers were #502 and #503.

To construct a *cd1433* expression construct for antibody production, primers #522 and #523 were used to amplify the *cd1433* gene lacking the stop codon. The resulting PCR product was digested with NcoI and XhoI, ligated to pET28a, and transformed into
DH5α. The resulting pET28a-cd1433 plasmid was transformed into BL21(DE3). To construct a sleC expression construct for antibody production, primers #461 and #463 were used to amplify the sleC gene lacking the stop codon. The resulting PCR product was digested with NdeI and XhoI, ligated to pET22b, and transformed into DH5α. The resulting pET22b-sleC plasmid was transformed into BL21(DE3). To construct a cspC expression construct for antibody production, primers #524 and #525 were used to amplify the cspC gene lacking the stop codon. The resulting PCR product was digested with NdeI and XhoI, ligated to pET22b, and transformed into DH5α. The resulting pET22b-cspC plasmid was transformed into BL21(DE3).

To construct a strain producing recombinant CspB perfringens, primers #512 and #513 were used to amplify cspB lacking the stop codon from from C. perfringens ATCC 13124 genomic DNA (a kind gift of Jimmy Ballard). The PCR product was digested with NdeI and XhoI, ligated to pET22b, and transformed into DH5α. The resulting pET22b-cspB perfringens plasmid was transformed into BL21(DE3). To construct a strain expressing the mature form of CspB perfringens carrying a C-terminal His₆-tag, primers #670 and #671 were used to amplify cspB missing the N-terminal 96 aa and carrying a C-terminal His₆-tag. The PCR product was digested with NdeI and XhoI and ligated to pRSFduet1 digested with the same enzymes. The resulting pRSFduet1-cspB Δ96 (C. perfringens) plasmid was transformed into BL21(DE3). This plasmid was also used to clone a series of transcomplementation constructs. Specifically, primers #668 and #701 were used to amplify the prodomain region of CspB perfringens (1-96 aa); primers #668 and #771 were used to amplify aa 1-92 of the CspB perfringens prodomain region; primers #668 and #772 were used to amplify aa 1-93 of the CspB perfringens prodomain.
region; primers #703 and #704 were used to amplify the prodomain region of CspC *perfringens*; and primers #457 and #702 were used to amplify the prodomain region of CspBA *difficile* (1-66 aa). A stop codon was added to the 3’ primer of all these prodomain constructs. The resulting PCR products were digested with NcoI and SalI and ligated to pRSFduet1- *cspB* Δ96 plasmid that had been digested with the same enzymes. The resulting plasmid constructs were transformed into BL21(DE3).

To construct a strain expressing full-length CspB *perfringens* carrying a C-terminal His<sub>6</sub>-tag, primers #512 and #671 were used to amplify *cspB* carrying a C-terminal His<sub>6</sub>-tag. The PCR product was digested with NdeI and XhoI and ligated to pRSFduet1 digested with the same enzymes. The resulting pRSFduet1- *cspB* (*C. perfringens*) plasmid was transformed into BL21(DE3). This plasmid was also used to introduce a series of point mutations using site-directed mutagenesis. Primer pair #516 and 517 was used to introduce the catalytic S494A mutation; primer pair #868 and 869 were used to introduce the S96R prodomain mutation; primer pair #709 and 710 was used to introduce the K91D prodomain mutation; primer pair #818 and 819 and primer pair #765 and 766 were used to introduce the R231Q and R231E salt bridge mutations, respectively; primer pairs #820 and 821 and #952 and 953 were used to introduce the E35R and E35Q salt bridge mutations, respectively; and primer pair #822 and 823 was used to introduce the E59A salt bridge mutation. To construct the E35R-R231E salt bridge swap construct, plasmid pRSFduet1- *cspB* R231E was used as the template in a site-directed mutagenesis PCR reaction using primer pair #820 and 821. The resulting plasmids were transformed into BL21(DE3).
PCR SOE was used to construct the YTS_AAA cleavage site mutation construct by using primer pair #512 and 741 to amplify the 5’ fragment and primer pair #740 and 671 to amplify the 3’ fragment. The resulting PCR products were purified and used in a second PCR reaction containing flanking primer pair #512 and 671. The PCR SOE product was digested with NdeI and XhoI and ligated to pRSFduet1 digested with the same enzymes to produce the pRSFduet1-cspB YTS/AAA plasmid. A similar strategy was used to delete the P3-P1 residues of CspB *perfringens* to produce plasmid pRSFduet1-cspB ΔYTS, with the exception that the SOE primers were #774 and #773. To delete the jellyroll domain of CspB *perfringens*, the same PCR SOE strategy was used except that the SOE primers were #718 and #717. The resulting plasmids were transformed into BL21(DE3).

To construct an expression construct producing the isolated jellyroll domain, primer pair #753 and 754 were used to amplify the region corresponding to aa 293-424 of *cspB* *perfringens*. The resulting PCR product was digested with NdeI and XhoI and ligated into pET22b digested with the same enzymes to produce plasmid pET22b-cspB jelly (293-424). This construct was transformed into BL21(DE3).

### 2.3.3. Protein Sequencing.

The N-termini of CspB variants were mapped using Edman sequencing and performed by the Protein and Nucleic Acid Facility at Stanford University. Briefly, 10 µg of each CspB variant was resolved by SDS-PAGE and transferred to a PVDF membrane. The membrane was stained with Ponceau S (0.1% w/v, 5% acetic acid), and CspB was excised from the membrane and sequenced using Edman degradation.
2.3.4. **Protein Purification.**

For purification of His$_6$-tagged proteins, overnight cultures of the appropriate BL21(DE3) strain were diluted 1:500 in 2L 2YT (5 g NaCl, 10 g yeast extract, 15 g tryptone/L) media and grown shaking (225 rpm) at 37ºC. When an OD$_{600}$ of 0.6-0.9 was reached, IPTG was added to 250 µM, and cultures were grown for 12-16 hr at 19ºC. Cultures were pelleted, resuspended in 25 mL lysis buffer [500 mM NaCl, 50 mM Tris-HCl, pH 7.5, 15 mM imidazole, 10% v/v glycerol] and flash frozen in liquid nitrogen. Lysates were thawed, then lysed by sonication and cleared by centrifugation at 15,000 x g for 30 minutes. His$_6$-tagged proteins were affinity purified by incubating the lysates in batch with 1.0 mL Ni-NTA Agarose beads (Qiagen) with shaking for 3 hr at 4ºC. The binding reaction was pelleted at 1,500 x g, the supernatant was set aside, and the pelleted Ni-NTA agarose beads were washed 3 x with lysis buffer. His$_6$-tagged proteins were eluted from the beads by the addition of 350 µL high imidazole elution buffer [500 mM NaCl, 50 mM Tris-HCl, pH 7.5, 175 mM imidazole, 10% v/v glycerol]. The elution was repeated four times; the eluate was pooled, buffer exchanged in gel filtration buffer [200 mM NaCl, 10 mM Tris pH 7.5, 5% v/v glycerol], and concentrated to 750 µL. The concentrated prep was pelleted at 13,000 x g for 10 min at 4ºC prior to loading on a Superdex 200 10/30 column (GE Healthcare).

For crystallization studies, C-terminally His$_6$-tagged CspB *perfringens* was affinity purified as described above then gel purified using a HiPrep S200 16/60 Sephacryl column (GE Healthcare); the gel filtration buffer was 150 mM NaCl, 10 mM Tris pH 7.5. The purified protein was concentrated to 10 mg/mL. A single peak was observed by gel
filtration, with the elution volume indicating that CspB likely exists as a monomer based on a calibration curve determined using gel filtration standards (data not shown). Consistent with this observation, the elution volume was independent of protein concentration. Protein purity was analyzed by SDS-PAGE followed by Coomassie staining (GelCode Blue, Pierce).

To prepare seleno-methionine (SeMet) substituted CspB for crystallization studies, an overnight culture of B384(DE3) harboring pET22b-CspB *perfringens* was grown in 60 mL minimal media [7.5 mM (NH₄)₂SO₄, 30 mM KH₂PO₄, 60 mM K₂HPO₄, 800 mg/L 19 amino acids excluding methionine, 5% w/v glucose, 1 mM MgSO₄, 4 mg/L thiamine, 4 mg/L D-biotin] supplemented with 5% LB media. The overnight culture was diluted 1:100 in 6 L minimal media supplemented with 100 mg/L SeMet (Sigma) and grown for 6 hr until an OD₆₀₀ of 0.6 was reached. Cultures were induced with 250 μM IPTG and grown at 30ºC for an additional 3 hr. SeMet-substituted CspB was purified using a similar protocol to native CspB with the exception that SeMet-substituted CspB was first purified on a 5 mL HiTrap Q anion-exchange column before gel filtration. Affinity-tagged SeMet-substituted CspB was buffer-exchanged into 10 mM Tris-HCl pH 7.5 (buffer A) using an Amicon Ultra-10 (Millipore) before loading it onto the HiTrapQ column; it was eluted over 20 column volumes using 0.75 M NaCl, 10 mM Tris-HCl pH 7.5 as buffer B. Fractions containing SeMet-labeled CspB *perfringens* were pooled, buffer-exchanged into the 150 mM NaCl, 10 mM Tris-HCl pH 7.5 gel filtration buffer, and purified by gel filtration.
2.4. Results

2.4.1. CspBA is processed during incorporation into C. difficile spores

The cspBA fusion gene is encoded in the genomes of only five clostridial species (Figure S2). C. difficile is unique among these species in that the CspA portion of CspBA (CD2247) lacks an intact catalytic triad (Figures 1A and S2). In order to determine whether CspBA is produced as a fusion protein, we raised antibodies against the CspB portion of CspBA and analyzed CspB production in both sporulating cells and purified spores by Western blotting. As a control, we constructed a targeted gene disruption of the cspBA-cspC (cd2247-cd2246) operon (Figure S1). In sporulating C. difficile cells, the anti-CspB antibody detected two polypeptides of ~130 kDa and ~55 kDa (Figure 1B). The former corresponds to the predicted MW of CspBA of 125 kDa, while the latter corresponds to the size of Csp proteases detected in C. perfringens spores (~60 kDa) [91]. Notably, the ~55 kDa protein was enriched in purified spores, suggesting that interdomain cleavage of CspBA occurs during spore formation and that CspB may be preferentially incorporated into the developing spore. Although the mutant strains exhibited similar levels of sporulation (Figure 1C), CspB levels in sleC~ mutants spores were consistently ~3-fold lower than in wildtype spores (Figure 1B). Nevertheless, these results indicate that CspBA is processed to CspB during C. difficile spore assembly.

While the cspBAC locus was previously identified by transposon mutagenesis as being essential for C. difficile spore germination [92], the effect of CspBA on SleC cortex hydrolase processing was not tested. To determine whether loss of CspBA prevents SleC processing, we analyzed SleC cleavage in response to a bile salt germinant [55] by Western blotting. As predicted, disruption of cspBAC in C. difficile prevented SleC
cleavage during germination (Figure 1B), and this defect could be complemented by ectopic expression of the \textit{cspBAC} locus from a multicopy plasmid [112] (Figure S1). Thus, the \textit{cspBAC} locus appears to regulate SleC activity in a manner similar to \textit{C. perfringens} [65, 70, 89, 91].

2.4.2. CspB undergoes autoprocessing in a position-dependent manner

In order to gain insight into the mechanism by which CspBA activates SleC during germination, we conducted structure-function analyses of the CspB domain of CspBA, since CspB is the only Csp protease encoded by \textit{C. difficile} with an intact catalytic triad (Figure 1A). Based on its homology to subtilases [91] (Figure S3), we hypothesized that CspB is synthesized as a pro-enzyme that undergoes autoprocessing. To test this hypothesis, we recombinantly produced wildtype CspB \textit{difficile} (residues 1-548 of CspBA) and CspB \textit{perfringens}, along with mutants with the catalytic serine inactivated, and compared their apparent MW by SDS-PAGE. Whereas mutation of the catalytic serine caused both CspB \textit{difficile} and CspB \textit{perfringens} to run at their expected MWs of ~60 kDa, the wildtype CspB proteins migrated with MWs of ~55 kDa (Figure 2A). Thus, Csps autocatalytically remove their prodomain in a manner similar to other subtilases.

Using Edman degradation, we mapped the autoprocessing site of CspB \textit{difficile} (1-548 aa) to Qln66 (data not shown). Alignment of this autocleavage site with previously mapped processing sites for CspA, CspB, and CspC of \textit{C. perfringens} [91] revealed that Csps cleave at a similar position relative to their mature domains (Figure 2B and S4). Given the limited conservation in amino acid sequence around the Csp autoprocessing sites (Figures 2B and S4), we tested whether CspB recognizes specific amino acid
residues upstream of the cleavage site. Mutation of the CspB *perfringens* P1 serine to a
bulky, charged Arg did not affect autoprocessing (P1 refers to the residue N-terminal to
the scissile bond based on the Schecter and Berger convention [113], Figure 2A); similarly, mutation of the P3-P1 residues to alanine did not affect CspB *perfringens* and
CspB *difficile* autoprocessing (Figure 2) or the position of cleavage (data not shown). In
contrast, deletion of the P3-P1 residues (∆YTS) of CspB *perfringens* markedly reduced
prodomain cleavage (Figure 2A), suggesting that the length of the prodomain affects
substrate recognition or binding.

### 2.4.3. Overall structure of CspB

While these findings highlighted similarities between Csps and other subtilases, all
CspB proteins capable of undergoing autoprocessing unexpectedly remained in complex
with their prodomain following multiple rounds of purification (Figure 2A). In contrast,
all previously characterized prokaryotic subtilases degrade their prodomain shortly after
autoprocessing [96]. To gain insight into the interaction between the prodomain and
subtilase domain, we determined the crystal structure of the CspB homolog from *C.
perfringens*. CspB contains a subtilase domain that is similar to other subtilisin-like
proteases [96], with the active site tucked within a conserved fold comprised of a six-
stranded antiparallel β-sheet that is sandwiched between four conserved α-helices (Figure
3A). The catalytic triad of the active site of CspB superimposes directly with other active
subtilisin-like proteases (Figure 3C), with an RMSD over the Cα atoms of only 0.11 Å
between the catalytic triads of CspB and Tk-SP, the most structurally related enzyme from
*Thermococcus kodakaraensis* as determined by the Dali server [114, 115].
In contrast with all previously solved prokaryotic subtilase structures, the autoprocessed prodomain stays bound to the wildtype, mature enzyme in our CspB structure. Notably, structures of prokaryotic subtilases in complex with their prodomain exist only for active site mutants [116-120]. The prodomain of CspB exhibits a similar structural organization to these subtilases, consisting of a 4-stranded antiparallel β-sheet and 3 α-helices (Figure 3A), with an additional β-strand extending into the catalytic cleft. The C-terminus of the CspB prodomain also extends directly into the oxyanion hole, with 19 hydrogen bonds stabilizing the intimate interaction between the C-terminal P6-P1 residues of the prodomain and the catalytic cleft (Figure S5 and Table S1). The prodomain-subtilase domain interface buries 1,472 Å² of accessible surface area (Figure 3C).

A second major feature that distinguishes the structure of CspB from other subtilases is the interruption of the protease domain by an ~130 aa insertion (Figure S3). This insertion assumes a β-barrel jellyroll fold, consisting of nine antiparallel β-strands that pack in a small hydrophobic core. The jellyroll domain interacts with both the prodomain and subtilase domain (Figure 3 and S5, Table S1). Although a similar jellyroll fold is present in the archaeal subtilisin Tk-SP [114] (Figures 3D and 4A, RMSD of 2.1 Å over 81 Ca atoms), the Tk-SP jellyroll domain is a C-terminal extension that interacts exclusively with the subtilase domain (Figure 3D). Nevertheless, both Tk-SP and CspB hold their jellyroll domains tightly in place with 22 and 19 bonds (primarily hydrogen bonds, Table S1), with a buried surface area between the jellyroll domain and subtilase domain of 1,115 Å² and 1,018 Å², respectively.
2.4.4. The jellyroll domain rigidifies CspB

In Tk-SP, the jellyroll domain has been shown to stabilize enzyme activity at high temperatures (>90°C) [114]. To test whether the jellyroll domain might similarly stabilize CspB, we compared the susceptibility of wildtype CspB and a mutant lacking the jellyroll domain (CspB Δjelly) to limited proteolysis. *In vitro* structure-function analyses were done on CspB *perfringens* rather than *C. difficile* because the structure was solved for CspB *perfringens*. In the presence of increasing concentrations of chymotrypsin, wildtype CspB exhibited remarkable resistance to degradation even when chymotrypsin levels were approximately equimolar to CspB (0.5 mg/mL or 20 µM chymotrypsin, Figure 4B). While mutation of the catalytic serine had little effect on CspB degradation, deletion of the jellyroll domain sensitized the mutant to chymotrypsin digestion at 50 ng/mL (Figure 4B). Loss of the jellyroll domain also reduced the efficiency of CspB autoprocessing, since both uncleaved and mature CspB Δjelly were observed following purification. In contrast, only uncleaved CspB Δjelly was observed upon mutation of the catalytic serine (Figure 4B). Taken together, these results implicate the jellyroll domain in (1) positioning the prodomain to undergo autocleavage and (2) markedly restraining the conformational flexibility of CspB.

2.4.5. The prodomain functions as an intramolecular chaperone

Having identified functions for the jellyroll domain, we next investigated the role of the prodomain in regulating CspB activity. For many subtilases, the prodomain acts as an intramolecular chaperone that catalyzes proper folding of the subtilase domain; once folding is complete, the mature enzyme autocatalytically separates the prodomain from its
subtilase domain [95, 96]. In most subtilases, the prodomain acts as a temporary inhibitor until it is autoproteolytically removed [96, 121]. To determine the extent to which CspB follow this model of subtilase maturation, we examined the chaperone activity of the CspB prodomain. Similar to other subtilases, deletion of the prodomain dramatically reduced the solubility and yield of mature CspB, while co-expression of the prodomain in trans restored folding to the subtilase domain (Figure S6). The chaperone activity of the prodomain was highly specific for CspB perfringens, since co-expression of the prodomains of CspB difficile and CspC perfringens in trans only marginally restored folding to the subtilase domain (Figure S6).

Indeed, the CspB subtilase domain recognizes its prodomain with an extensive network of interactions, consisting of 27 hydrogen bonds and three salt bridges (Figure S5 and Table S1 in Text S1). The prodomain adopts a similar fold to the prodomains of related subtilisin-like proteases (Figure 5A), with the C-terminal region extending deep into the catalytic cleft (Figure 3A). The 94 Cα atoms of the prodomain align with an RMSD of 2.4 Å compared to the Tk-SP prodomain and 2.5 Å when compared to the mammalian proprotein convertase subtilisin kexin type 9 (PCSK9), respectively [43].

We compared the CspB prodomain to the PCSK9 prodomain because PCSK9 is the only other example of a wild-type subtilase that remains bound to its prodomain [122-124], whereas the prodomain of Tk-SP only stays bound if the catalytic serine of Tk-SP is mutated to cysteine [119, 120]. Since we did not observe any obvious structural differences to account for the difference in prodomain retention, we examined the free energy of dissociation of prodomains from their cognate subtilase domains using PDBe PISA, which is a computational server for examining interaction interfaces on proteins.
This analysis revealed that CspB and PCSK9 have the highest energy barriers to prodomain dissociation relative to other subtilases bound to their cognate prodomain or inhibitor ($\Delta G = -19.2$ and $-17.7$ kcal/mol, respectively, Figure 5B). Interestingly, while most of the interactions holding the prodomain to the subtilase domain are not sequence specific (Table S1), with 15 bonds directed at backbone atoms, there are a few salt bridges that mediate specific recognition of the prodomain (Figure 5C). These salt bridges occur between Glu35/Glu59 of the prodomain and Arg231 of the subtilase domain and between Lys91 of the prodomain and Asp257 of the subtilase domain (Figure 5C). To determine the contribution of these salt bridges to CspB folding, we mutated each salt bridge residue and analyzed the effect on CspB solubility. Mutations of Glu35 to glutamine and Glu59 to alanine slightly reduced yields relative to wildtype, whereas mutation of Arg231 to glutamine strongly decreased recovery of CspB (Figure 5D), presumably because it disrupts both potential salt bridge interactions. Flipping the charges on Glu35 and Arg231 (E35R or R231E, respectively) also significantly reduced CspB yields, while swapping the Glu35-Arg231 salt bridge (E35R-R231E) failed to rescue CspB solubility. In contrast, flipping the charge on Lys91 to aspartate (K91D), which forms a salt bridge with subtilase domain residue Asp257, had little effect on K91D solubility relative to wildtype CspB (Figure 5C and 5D, Table S1). Taken together, these results highlight the importance of the Glu35-Arg231 and Glu59-Arg231 prodomain-subtilase domain salt bridges in promoting subtilase domain folding.
2.4.6. The prodomain C-terminus sterically occludes a catalytically competent active site

Having demonstrated the intramolecular chaperone activity of the prodomain, we next tested whether the prodomain functions as an inhibitor similar to other subtilases. Consistent with this hypothesis, the C-terminal P3-P1 residues of the prodomain bind the catalytic site in a manner analogous to an inhibitory peptide, fitting snugly within the catalytic cleft and presumably occluding access to the active site residues (Figure 6A). The S1 and S2 binding pockets perfectly accommodate the P1 serine and P2 threonine (P1 refers to the residue N-terminal to the cleavage site; S1 refers to the P1 substrate binding pocket). The bulky P3 tyrosine residue is wedged between Arg222 and Ser254 of the subtilase (Figure S5 and Table S1). The C-terminal P1-P3 prodomain residues form a total of 13 bonds to the S1-S3 regions of the subtilase domain. The P1 Ser96 forms seven hydrogen bonds with NE2 of catalytic His183, Ser252, Asn287, Thr493 and catalytic Ser494; P2 Thr95 forms four hydrogen bonds to different atoms of Arg222; and P3 Tyr94 forms hydrogen bonds to both the backbone amide and carbonyl of Ser254.

To test whether these residues block substrate access to the CspB active site, we used a small activity-based probe (FP-Rh, Figure 6B) to detect CspB catalytic activity. The fluorophosphonate electrophilic group of the probe reacts exclusively with catalytically competent serine hydrolases such as the subtilisins, which are a subfamily of the subtilases [126]. Nucleophilic attack by the catalytic serine results in the probe becoming covalently bound to the catalytic serine, while the rhodamine tag allows for detection of the covalently labeled enzyme by fluorescent gel scanning. Incubation of either wildtype or catalytically inactive S461A CspB with FP-Rh failed to produce detectable fluorescence, implying that the active site is inaccessible in the wildtype
enzyme (Figure 6C). In contrast, mutation of the P3-P1 residues (YTS/AAA) produced a CspB variant that could be labeled on its catalytic serine, suggesting that the C-terminal prodomain residues act as gatekeepers to a catalytically competent active site. Accordingly, truncation of the C-terminal gatekeeper of the prodomain expressed in trans of residues YTS (P3-P1) or LYTS (P4-P1) permitted labeling of the CspB active site, whereas the full-length prodomain expressed in trans prevented labeling (Figure 6C). Taken together, these results indicate that the C-terminal YTS prodomain residues inhibit CspB activity.

2.4.7. The CspB jellyroll domain stabilizes CspBA

Having identified key structural features of CspB perfringens in vitro, we next tested their functional significance in regulating CspBA activity in C. difficile. To this end, we cloned cspBAC complementation constructs in which the jellyroll domain was deleted (Δjelly, Figure 7A) or the active site serine was mutated (S461A). The cspBAC constructs were expressed from their native cspBA promoter on a multicopy plasmid (pMTL83151) [112]. Deletion of the jellyroll domain appeared to destabilize CspBA, since CspBA Δjelly levels were markedly reduced relative to wildtype and the cspBAC complementation strain and degradation products were apparent (Figure 7B). In contrast, mutation of the catalytic serine (S461A) did not affect CspBA levels relative to the cspBAC complementation strain, although CspBA S461A failed to undergo autoprocessing (Figure 7B). In purified spores, the predominant form of CspB was autoprocessed (m-CspB) in wildtype and cspBAC-complemented spores, whereas the predominant form of CspB in S461A spores was not autoprocessed (Figure 7C). Given that CspBA S461A was still processed at the
CspB-CspA junction, an as-yet-unidentified protease apparently separates CspB from CspA.

To determine the role of CspBA autoprocessing in *C. difficile* spore germination, we examined the ability of S461A mutant spores to germinate in response to bile salts. Relative to wildtype and *cspBAC*-complemented spores, S461A mutant spores exhibited an ~20-fold defect in germination and SleC cleavage (Figure 7C), while loss of the jellyroll domain (Δjelly) reduced spore germination by ~70-fold (Figure 7C). Nevertheless, loss of CspBA and CspC production in the *cspBAC*– mutant produced a more severe phenotype than loss of the catalytic activity (S461A) or jellyroll domain (Δjelly) of CspBA. Taken together, these results indicate that CspB catalytic activity and its jellyroll domain are required for efficient *C. difficile* spore germination.

2.4.8. The protease activity of CspBA is required for germination downstream of autoprocessing

The observation that ~5% of pro-SleC undergoes cleavage during germination of S461A mutant spores (Figure 7C) raised the question as to how SleC was being activated in the absence of CspB protease activity. One possibility is that a redundant protease cleaves SleC during germination of S461A mutant spores. Another possibility is that CspB activates a second protease that directly cleaves SleC. While this latter model is more complicated, it reflects how the subtilisin-like proprotein convertase PCSK9 indirectly regulates low-density lipoprotein receptor (LDLR) levels. Rather than enzymatically degrading LDLR, PCSK9 binds and targets LDLR to the lysosome [121, 127]. However, in order to bind LDLR, PCSK9 must undergo autoprocessing to form a non-covalent
complex with its prodomain; only after autoprocessing can PCSK9 recognize LDLR [121, 127]. As a result, PCSK9 is the only other wildtype subtilisin-like protease that retains its prodomain in its crystal structure following autoprocessing [122-124].

If CspB activity is regulated similarly to PCSK9, CspB protease activity should be dispensable once autoprocessing has occurred. To test this hypothesis, we co-expressed the CspBA prodomain with a CspBA variant lacking its prodomain such that the CspBA produced is identical to wildtype CspBA after autoprocessing (Q66, Figure 8A). The prodomain was also co-expressed with a catalytically inactive CspBA variant lacking its prodomain (Q66/S461A, Figure 8A). As predicted, Q66 and Q66/S461A transcomplementation mutants produced CspBA variants that were indistinguishable in size from wildtype in sporulating cells (Figure 8B) and purified spores (Figure 8C), although more CspBA fusion protein was observed in the transcomplementation mutant spores relative to wild type (Figure 8C). Nevertheless, Q66/S461A mutant spores exhibited a 10-fold defect in both germination and SleC cleavage relative to wildtype and Q66 mutant spores. This result indicates that the catalytic activity of CspBA is required for efficient SleC cleavage downstream of CspBA autoprocessing.

2.5. Discussion

Spore germination is essential for Clostridium sp. pathogens such as C. perfringens and C. difficile to initiate infection [56, 85]. A critical step during germination is the degradation of the thick, protective cortex layer surrounding the spore core by cortex hydrolases [56, 86, 87]. However, despite their functional importance, little is known about the molecular mechanisms that control cortex hydrolase activity. In this study, we
provide the first molecular insight into cortex hydrolase regulation by solving the structure of CspB, a protease required for cortex hydrolase activation. Combined with our functional analyses of CspB in vitro and in vivo, the structure reveals that Csps are subtilisin-like proteases with two distinctive functional features: a central jellyroll domain and a retained prodomain.

The central β-barrel jellyroll domain of CspB interrupts the subtilase domain and wedges itself tightly between the subtilase domain and prodomain in three-dimensional space (Figure 3C). This unique position is likely critical for CspB function, since the jellyroll domain markedly restrains the conformational mobility of CspB through extensive and specific interactions at the subtilase-jellyroll domain interface (Figure 4B and S5). The rigidity conferred by the jellyroll domain presumably helps CspB survive the environmental extremes that spores can encounter, such as freeze-thaw cycles and boiling temperatures [28]. The jellyroll domain also facilitates CspB autoprocessing in vitro (Figure 4B), indicative of a role in helping CspB adopt the correct subtilase fold. Consistent with this proposal, deletion of the jellyroll domain in C. difficile markedly reduced CspBA levels relative to wild type (Figure 7B).

In these respects, the jellyroll domain is more functionally analogous to the β-barrel P-domains of kexin-like subtilisins than to the jellyroll domain of prokaryotic Tk-SP subtilisin. Like the CspB jellyroll domain, the P-domain of kexin-like proteases, such as the mammalian enzyme furin, is important for autoprocessing, folding, stability, and activity of the subtilase domain [96, 121, 128-130]. In contrast, the jellyroll domain of Tk-SP is dispensable for autoprocessing, protein folding and activity in vitro, despite being important for Tk-SP thermostability [114].
The retention of the CspB prodomain is another unique feature identified by our study. Unlike the majority of subtilisin-like proteases, the prodomain stays bound to the wildtype subtilase domain via a network of interactions that result in tighter prodomain binding relative to other subtilases (Figure 5B and S5). Prodomain binding to its cognate protease appears highly specific, since prodomain swapping does not result in efficient folding of CspB (Figure S6). This conclusion is consistent with the limited sequence conservation of prodromains across Csps (Figure S3); indeed, even the salt bridges critical for prodomain chaperone activity (Figure 5D) are not conserved. Despite the low level of sequence conservation, the position of prodomain autoprocessing is highly conserved (Figure 2B), and a small internal deletion of the prodomain disrupts autocleavage even though diverse residues are tolerated at the P1 position (Figure 2A).

Mechanistically, Csps exhibit less specificity in P1 substrate recognition than most subtilases [131-135]. Nevertheless, while residues around the prodomain cleavage site do not affect autocleavage efficiency, they do control active site accessibility after autoprocessing, excluding even a small, highly reactive, serine protease probe in vitro (Figure 6C). Taken together, Csps appear more functionally similar to the site-specific kexin-like protease subfamily than to the highly processive subtilisin subfamily [93, 96]. Similar to kexin-like proteases, Csps cleave their putative substrate, SleC, at a single site during germination [65] (Figure 1B) and remain more closely associated with their prodomain following autoprocessing [96, 121]. By contrast, subtilisin subfamily members such as Tk-SP function as major degradative enzymes that rapidly degrade their prodomain following autoprocessing [96].
While these observations provide new insight into the structure and function of Csp proteases, they raise a number of questions for future study. Does the prodomain remain associated with autoprocessed CspB in dormant spores as it does in vitro? If the prodomain stays bound to mature CspB in dormant spores, what happens to the prodomain during germination? Given that chymotrypsin cannot access numerous prodomain cleavage sites during extended incubation in vitro (Figure 4), a significant change in CspB conformational mobility would appear to be required for the prodomain to be degraded and its putative substrate SleC to gain access to the CspB substrate binding pocket.

Another question raised by our study is the role of CspC in regulating germination in C. difficile. Given that Δjelly, S461A, and Q66/S461A mutant spores exhibit germination defects that are >100-fold less severe than cspBAC− spores and that a major difference between these mutant spores is the absence of CspC in cspBAC− mutant spores (Figures 7 and 8), catalytically inactive CspC (Figure 1A) may play a role in SleC activation. Recent data suggests that CspC helps transduce the germination signal to CspB (J. Sorg, personal communication). In addition, it is unclear what fraction of pro-SleC must be proteolytically activated to induce successful spore germination. Approximately 5% of spores of the CspBA catalytic mutant S461A successfully germinate, which correlates with a small fraction of pro-SleC undergoing processing in the mutant strain (Figure 8). This result suggests that only a small fraction of SleC must be proteolytically activated in order to mediate spore germination in some cells; alternatively, a small fraction of S461A spores could efficiently cleave pro-SleC and thus germinate successfully.
While further experimentation is needed, the work presented here provides the first structure-function analyses of Csp proteases \textit{in vitro} and \textit{in vivo} and lays the groundwork for mechanistically addressing how the germination pathway senses and integrates the germination signal. Furthermore, this study may provide the structural basis for designing therapeutics that either block prodomain and/or jellyroll domain binding to the CspBA subtilase domain during spore formation or prematurely activate CspBA to induce cortex hydrolysis. These CspBA agonists or antagonists could prevent \textit{C. difficile} transmission and disease recurrence.

### 2.6. Acknowledgements

We would like to thank B. Cravatt and A. Speers for the FP-Rhodamine probe, N. Minton for pMTL83151 and pMTL84151, J. Ballard for \textit{C. perfringens} ATCC 13124 genomic DNA, J. Sorg and C. Ellermeier for plasmid reagents and important technical advice, J. Sorg for sharing unpublished data, T. Lawley and N. Villanueva for helpful advice, and members of the Doublié and Shen lab for helpful discussion and advice. We would like to thank K. Zahn for collecting diffraction datasets at the Advanced Photon Source Synchrotron. The beamline 23ID-B and the Advanced Photon Source are supported by National Cancer Institute Grant Y1-CO-1020, National Institute of General Medical Sciences Grant Y1-GM-1104, and U.S. Dept. of Energy, Basic Energy Sciences, Office of Science Contract DE-AC02-06CH11357.
2.7. Figures

Figure 2.7.1. The CspBA fusion protein undergoes processing during sporulation.
(a) Schematic of Csps and SleC in *C. perfringens* and *C. difficile*. Intact catalytic residues are black, while catalytic mutations are grey. The prodomain of *C. perfringens* Csps are shown in light grey, with their lengths indicated. The predicted prodomain of CspBA is also indicated. SleC is outlined in black, with the prepeptide (Pre), propeptide (Pro), and Csp cleavage site indicated for *C. perfringens* SleC [66, 90] (b) Western blot analysis of sporulating *C. difficile* and purified spores. Purified spores of the indicated strain were either untreated (–) or exposed to 0.2% w/v sodium taurocholate [55] (+, germinant) for 15 min at 37°C and analyzed by Western blotting and for germination efficiency via colony forming unit (cfu) determination. The processing products of CspB and SleC are indicated. CD1433 was previously shown to be a component of *C. difficile* spores and is used as a loading control [99]; the anti-CD1433 antiserum primarily recognizes the chitinase domain of CD1433. CspB levels were 3.5-fold lower in *sleC*– spores relative to wildtype spores, despite containing similar amounts of CD1433. (c) Phase-contrast microscopy of sporulating *C. difficile* strains used in (b) showing equivalent levels of sporulation as measured by particle counting. The white triangles indicate mature phase-bright spores that have been released from the mother cell; the black triangles highlight immature forespores in the mother cell.
Figure 2.7.2. CspB undergoes autoprocessing in a position-dependent manner

(a) Coomassie staining of recombinant \textit{C. perfringens} and \textit{C. difficile} CspB variants. 7.5 \textmu g of each purified CspB variant was resolved by SDS-PAGE on a 4-12\% Bis-Tris gel and visualized by Coomassie staining. The P3-P1 residues of the prodomain were mutated to Ala for the YTS/AAA and QTQ/AAA mutants, while the P3-P1 residues
were deleted from CspB *perfringens* in the ΔYTS mutant. The products resulting from autoprocessing are indicated. (b) Sequence alignment of Csp prodomain cleavage sites mapped by Edman sequencing; the Csp *perfringens* cleavage sites were mapped in a previous study [91]. Completely conserved identical residues are blocked in black with white text, conserved identical residues in grey with white text, and conserved similar residues in light grey.
Figure 2.7.3. Overall structure of CspB *perfringens*

(a) Ribbon representation showing subtilase domain in purple, jellyroll domain in green, and prodomain in teal extending into the active site. Catalytic residues are shown as stick models with yellow carbons. (b) Close-up view of catalytic site. An overlay of CspB (purple) and Tk-SP (grey). The three catalytic residues are shown. Tk-SP and CspB catalytic residues are labeled in black and purple, respectively. (c) Space-filling model of CspB with same orientation and color scheme as (a). (d) Overlay of CspB (colors, same as
(a)) and Tk-SP (shown in grey), showing similar overall structures with the exception of the position of the jellyroll domain. The jellyroll domains of CspB and Tk-SP are shown in green and grey, respectively. Note that only the regions with conserved secondary structure in the prodomain and subtilase domain are shown.
Figure 2.7.4. The jellyroll domain conformationally rigidifies CspB *perfringens*

(a) Overlay of jellyroll domain of CspB *perfringens* (green) and Tk-SP (grey). (b) Limited proteolysis profile of CspB and its variants. 15 µM of CspB and its variants were incubated with increasing concentrations of chymotrypsin for 60 min at 37°C. Reactions were resolved by SDS-PAGE and visualized by Coomassie staining. Schematic of CspB variants is shown below the Coomassie stained gel. “Pro” refers to the prodomain; black rectangle demarcates the jellyroll domain; thin white rectangle represents the jellyroll
deletion; and white star denotes the S494A mutation. m-CspB refers to mature CspB, which is produced after autoprocessing.
Figure 2.7.5. Dual salt bridges are required for prodomain intramolecular chaperone activity

(a) Overlay of prodomains from CspB perfringens (teal), Tk-SP (grey), and PCSK9 (pink). (b) PDBe PISA analyses of free energy of prodomain dissociation from...
mature subtilase, with CspB in teal, PCSK9 in pink, and others in grey. (c) Close-up view of dual salt-bridge interaction at prodomain-subtilase interface. The C-terminus of the prodomain (C, teal) extends toward the substrate-binding pocket. Prodomain Glu35, Glu59 and Arg91 residues are shown in teal; subtilase domain Arg231 and D257 residues are shown in magenta. (d) Analysis of CspB prodomain mutant solubility using Western blotting and Coomassie staining. Cultures expressing cspB variants were induced with IPTG, and aliquots were removed 30 minutes later (“induced-IPTG” sample). Cells were lysed by sonication and centrifuged at high speed; the “cleared lysate” sample represents the soluble fraction. CspB variants were purified by affinity chromatography. Equivalent amounts of samples were resolved by SDS-PAGE and analyzed either by Western blotting using anti-CspB perfringens antisera or by Coomassie staining (bottom gel, affinity-purified CspB).
Figure 2.7.6. C-terminal prodomain residues sterically occlude a catalytically competent active site

(a) Close-up of interaction between prodomain C-terminus and substrate binding pocket. Subtilase, jellyroll and prodomains are shown in semi-transparent surface representation (purple, green, and teal, respectively). Residues 89-96 of prodomain are shown in yellow. (b) Structure of fluorophosphonate-rhodamine (FP-Rh) activity-based probe. Rhodamine dye is shown in red. (c) Schematic of CspB variants. “Pro” refers to the
prodomain; “+” reflects co-expression of the prodomain in trans, with the number reflecting the prodomain length. (d) Labeling of CspB variants by FP-Rh. CspB variants (10 μM) were incubated with 1 μM FP-Rh probe for 20 min at RT in triplicate. The labeling reactions were resolved by SDS-PAGE on a 15% gel and visualized by fluorescent scanning followed by Coomassie staining. A single representative replicate is shown. m-CspB refers to mature CspB lacking its prodomain.
Figure 2.7.7. The jellyroll domain and catalytic serine of CspBA are required for efficient germination

(a) Schematic of CspBA variants produced by cspBAC complementation constructs. “Pro” denotes the prodomain; black rectangle demarcates the jellyroll domain;
a thin white rectangle represents the jellyroll deletion; and white star indicates S461A mutation. (b) Western blot analyses of sporulating cells expressing *cspBAC* complementation constructs and (c) germinating spores expressing *cspBAC* complementation constructs. Purified spores of the indicated strain were either untreated (–) or exposed to 0.2% w/v sodium taurocholate (+, germinant) for 15 min at 37˚C and analyzed by Western blotting with the indicted antibodies. Germination efficiency was determined via colony forming unit (cfu) determination. Representative clones of each construct are shown, but more than two clones of each complementation construct were tested. m-CspBA reflects the mature form of CspBA following autoprocessing, and m-CspB reflects the mature form of CspB following autoprocessing. The different mutant CspB variants are indicated.
Figure 2.7.8. CspBA activity downstream of autoprocessing is required for efficient SleC cleavage

(a) Schematic of CspBA variants produced by cspBAC transcomplementation constructs. “Pro” denotes the prodomain; black rectangle demarcates the jellyroll domain; a thin white rectangle represents the jellyroll deletion; and white star indicates S461A mutation. (b) Western blot analyses of sporulating cells expressing cspBAC transcomplementation constructs and (c) germinating spores expressing transcomplementation constructs. Purified spores of the indicated strain were either untreated (−) or exposed to 0.2% w/v sodium taurocholate (+, germinant) for 15 min at
37°C and analyzed by Western blotting with the indicated antibody. Germination efficiency was determined via colony forming unit (cfu) determination. Representative clones of each construct are shown, but more than two clones of each complementation construct were tested. m-CspBA reflects the mature form of CspBA following autoprocessing, and m-CspB reflects the mature form of CspB following autoprocessing.
2.8. Supporting Information

Supplemental Figures

Figure 2.8.S1. Csp protease and SleC are required for spore germination in Clostridium sp.

(a) Schematic of sleC and csp genes in C. perfringens ATCC 13124 (gas gangrene isolate) [136], C. perfringens SM101 (food poisoning isolate) [136], and C. difficile 630.

(b) Western blot analyses of sporulating cells and (c) germinating spores for cspBAC− complementation strains. Sodium taurocholate was used to stimulate germination for 20
min at 37°C. The number of viable spores obtained upon plating on BHIS plates containing 0.2% w/v taurocholate is given as colony forming units (cfus).
Figure 2.8.S2. ClustalW sequence alignment of CspBA proteins

Completely conserved identical residues are blocked in blue, conserved identical residues in green, and conserved similar residues in yellow. A red triangle indicates...
catalytic triad residues (also boxed in red). Note that the catalytic His of CspB *difficile* did not align with the other CspBA homologs, despite being conserved in position in alignments with isolated CspB proteins (*Figures S3* and *S4*). Because of this discrepancy, the ClustalW alignment was altered to reflect the conservation of the catalytic His. CspBAs from *C. butyricum* (ZP_045298777), *C. botulinum* E3 (ZP_04529497), *C. acetobutycicum* (AE007820_5), *Candidatus arhtomitus* (EGX28514), and *C. difficile* 630 (YP_001088762.1). We also note that *C. tetani* E88 encodes an N-terminally truncated CspBA homolog lacking the first Asp in the catalytic triad (AAO36820), but this was not included in the alignment.
<table>
<thead>
<tr>
<th>Major ISP</th>
<th>B. subtilis</th>
<th>C. difficile</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ClustalW sequence alignment of major classes of bacterial subtilisin-like proteases</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Major ISP</th>
<th>B. subtilis</th>
<th>C. difficile</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Completely conserved identical residues (blue)</strong>, conserved identical residues (green), and conserved similar residues (yellow). A red triangle indicates catalytic triad residues (boxed in red). The central insertion corresponds to the jellyroll domain. Major intracellular serine proteases (ISP): B. subtilis str. 168 (NP_389202.1) and C. difficile (YP_0010888508.1); extracellular serine proteases: AprE from B. amyloliquefaciens (YP_003919715.1) and subtilisin E from B. subtilis str. 168 (NP_388911.2); CspB: C. perfringens ATCC 13124 (YP_697251.1) and C. difficile 630, 1-548 aa (YP_001088762.1).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 2.8.4. ClustalW sequence alignment of diverse Csp proteases

Completely conserved identical residues are blocked in blue, conserved identical residues in green, and conserved similar residues in yellow. A red triangle indicates
catalytic triad residues (also boxed in red). *C. perfringens* Csp prodomain cleavage sites are boxed in pink [91], while *C. difficile* CspB (1-548 aa) autoprocessing site is boxed in orange. α-helices and β-sheets in the CspB *perfringens* structure are indicated as a helix or a bracket above the sequence alignment, respectively. CspA *Roseburia intestinalis* (CBL08898), CspB *Blautia hansenii* (ZP_05853381), CspA *Ruminococcus gnarus* (ZP_02042115), CspA *Coprococcus catus* (CBK81001), CspC *C. perfringens* ATCC 13124 (YP_697250), CspA *C. perfringens* ATCC 13124 (YP_697252), CspB *C. perfringens* ATCC 13124 (YP_697251), and CspB *C. difficile* 630, 1-548 aa (YP_001088762.1).
The prodomain is shown in teal, the subtilase domain in purple, and the jellyroll domain in green. Each residue involved in a predicted hydrogen bond is shown as a stick model. Bonds predicted by PDBePISA [125] are shown as dashed grey lines. All bonds
predicted by PISA have been drawn, but not all are visible. (a) Prodomain interaction with mature subtilase domain, with the prodomain C-terminus extending into the active site. The Glu35/Glu59/Arg231 salt bridge interactions (Fig. 5) are shown, and selected residues are labeled. (b) Jellyroll domain interaction with prodomain and subtilase domain.
Figure 2.8.S6. Csp prodomain transcomplementation

(a) Schematic of transcomplementation constructs. The source of the prodomain is indicated. (b) Western blot and Coomassie stain showing the purification of CspB transcomplementation mutants. Cultures expressing cspB variants were induced with IPTG, and aliquots were removed 30 minutes later (“induced” sample). During the purification process, a sample of the soluble fraction was removed (“cleared lysate” sample). These samples were resolved by SDS-PAGE and analyzed by Western blotting using an anti-CspB perfringens antibody. Following affinity purification of the His$_6$-tagged CspB variants, equivalent amounts of the “eluate” were loaded and analyzed by SDS-PAGE and Coomassie staining.
(a) Stereo view of bias-free, density-modified experimental map produced from SHELX/C/D/E [104] by SAD phasing using 12 selenium sites and prior to model building (map shown as dark blue mesh). The 1.6 Å map is contoured at 1 σ and shown over the C-terminal residues (92-96) of the prodomain, the catalytic triad (Asp126, His183, and Ser494), and within a 3 Å radius of each atom. Prodomain residue carbons are shown in cyan and catalytic residue carbons in yellow. OXT indicates the prodomain C-terminus resulting from proteolytic cleavage. (b) 1.6 Å resolution anomalous electron density map
from SHELX/C/D/E showing selenium anomalous signal (orange mesh) in selenomethionine (MSe) residues. Map is contoured at 3 $\sigma$ and shown over MSe237 and MSe558.
2.9. Supporting Tables

Table 2.9.S1. Interactions between prodomain and mature CspB and jellyroll domain and CspB subtilase domain as determined by PDBe PISA [125].

<table>
<thead>
<tr>
<th>Hydrogen Bonds</th>
<th>Dist. [Å]</th>
<th>Subtilase</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>##</strong></td>
<td><strong>Prodomain</strong></td>
<td><strong>Dist. [Å]</strong></td>
<td><strong>Subtilase</strong></td>
</tr>
<tr>
<td>1</td>
<td>10 ASP OD2</td>
<td>2.76</td>
<td>362 TYR OH</td>
</tr>
<tr>
<td>2</td>
<td>61 LEU O</td>
<td>2.77</td>
<td>269 TYR OH</td>
</tr>
<tr>
<td>3</td>
<td>88 GLU OE1</td>
<td>2.73</td>
<td>265 LEU N</td>
</tr>
<tr>
<td>4</td>
<td>88 GLU OE2</td>
<td>3.24</td>
<td>266 LEU N</td>
</tr>
<tr>
<td>5</td>
<td>91 LYS O</td>
<td>2.84</td>
<td>226 SER N</td>
</tr>
<tr>
<td>6</td>
<td>91 LYS N</td>
<td>2.88</td>
<td>226 SER OG</td>
</tr>
<tr>
<td>7</td>
<td>91 LYS NZ</td>
<td>2.86</td>
<td>257 ASP O</td>
</tr>
<tr>
<td>8</td>
<td>92 ILE O</td>
<td>2.9</td>
<td>255 THR OG1</td>
</tr>
<tr>
<td>9</td>
<td>93 LEU O</td>
<td>2.87</td>
<td>224 ALA N</td>
</tr>
<tr>
<td>10</td>
<td>93 LEU N</td>
<td>2.74</td>
<td>224 ALA O</td>
</tr>
<tr>
<td>11</td>
<td>94 TYR O</td>
<td>2.97</td>
<td>254 SER N</td>
</tr>
<tr>
<td>12</td>
<td>94 TYR N</td>
<td>2.88</td>
<td>254 SER O</td>
</tr>
<tr>
<td>13</td>
<td>95 THR OG1</td>
<td>3.47</td>
<td>222 ARG NE</td>
</tr>
<tr>
<td>14</td>
<td>95 THR O</td>
<td>2.82</td>
<td>222 ARG NH2</td>
</tr>
<tr>
<td>15</td>
<td>95 THR N</td>
<td>3.01</td>
<td>222 ARG O</td>
</tr>
<tr>
<td>16</td>
<td>95 THR OG1</td>
<td>3.7</td>
<td>222 ARG O</td>
</tr>
<tr>
<td>17</td>
<td>96 SER O</td>
<td>3.24</td>
<td>183 HIS NE2</td>
</tr>
<tr>
<td>18</td>
<td>96 SER OG</td>
<td>2.61</td>
<td>494 SER OG</td>
</tr>
<tr>
<td>19</td>
<td>96 SER OXT</td>
<td>2.97</td>
<td>494 SER N</td>
</tr>
<tr>
<td>20</td>
<td>96 SER OXT</td>
<td>2.82</td>
<td>287 ASN ND2</td>
</tr>
<tr>
<td>21</td>
<td>96 SER OXT</td>
<td>3.38</td>
<td>493 THR N</td>
</tr>
<tr>
<td>22</td>
<td>96 SER N</td>
<td>3.35</td>
<td>252 SER O</td>
</tr>
<tr>
<td>23</td>
<td>96 SER OG</td>
<td>3.45</td>
<td>252 SER O</td>
</tr>
</tbody>
</table>

Salt Bridges

<table>
<thead>
<tr>
<th><strong>##</strong></th>
<th><strong>Prodomain</strong></th>
<th><strong>Dist. [Å]</strong></th>
<th><strong>Subtilase</strong></th>
<th><strong>Protein</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>35 GLU OE1</td>
<td>3.51</td>
<td>231 ARG NH2</td>
<td>Prodomain</td>
</tr>
<tr>
<td>2</td>
<td>35 GLU OE1</td>
<td>2.96</td>
<td>231 ARG NH1</td>
<td>Prodomain</td>
</tr>
<tr>
<td>3</td>
<td>35 GLU OE2</td>
<td>2.97</td>
<td>231 ARG NH2</td>
<td>Prodomain</td>
</tr>
<tr>
<td>4</td>
<td>35 GLU OE2</td>
<td>3.57</td>
<td>231 ARG NH1</td>
<td>Prodomain</td>
</tr>
<tr>
<td>5</td>
<td>59 GLU OE1</td>
<td>2.83</td>
<td>231 ARG NH2</td>
<td>Prodomain</td>
</tr>
<tr>
<td>6</td>
<td>59 GLU OE1</td>
<td>2.89</td>
<td>231 ARG NE</td>
<td>Prodomain</td>
</tr>
<tr>
<td>7</td>
<td>59 GLU OE2</td>
<td>3.18</td>
<td>231 ARG NE</td>
<td>Prodomain</td>
</tr>
<tr>
<td>8</td>
<td>91 LYS NZ</td>
<td>3.89</td>
<td>257 ASP OD2</td>
<td>Prodomain</td>
</tr>
</tbody>
</table>

The distance between two atoms is shown for each bond; residue number, residue 3-letter code, and atom name specify each atom involved.
Table 2.9.S1.2.: Jellyroll domain interaction interface with mature subtilase: hydrogen bonds and salt bridges.

### Hydrogen Bonds

<table>
<thead>
<tr>
<th>##</th>
<th>Jellyroll</th>
<th>Dist. [Å]</th>
<th>Subtilase</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>291 SER N</td>
<td>3.26</td>
<td>288 GLU O</td>
</tr>
<tr>
<td>2</td>
<td>291 SER N</td>
<td>3.49</td>
<td>290 ASN OD1</td>
</tr>
<tr>
<td>3</td>
<td>291 SER OG</td>
<td>2.92</td>
<td>288 GLU OE2</td>
</tr>
<tr>
<td>4</td>
<td>291 SER OG</td>
<td>3.21</td>
<td>287 ASN O</td>
</tr>
<tr>
<td>5</td>
<td>293 HIS ND1</td>
<td>2.76</td>
<td>288 GLU OE1</td>
</tr>
<tr>
<td>6</td>
<td>294 HIS ND1</td>
<td>3.18</td>
<td>425 SER O</td>
</tr>
<tr>
<td>7</td>
<td>294 HIS O</td>
<td>2.82</td>
<td>425 SER N</td>
</tr>
<tr>
<td>8</td>
<td>319 ASP OD2</td>
<td>3.28</td>
<td>258 GLY N</td>
</tr>
<tr>
<td>9</td>
<td>362 TYR OH</td>
<td>2.76</td>
<td>10 ASP OD2</td>
</tr>
<tr>
<td>10</td>
<td>370 ASP OD2</td>
<td>3.29</td>
<td>261 ASN ND2</td>
</tr>
<tr>
<td>11</td>
<td>372 GLN OE1</td>
<td>3.55</td>
<td>261 ASN ND2</td>
</tr>
<tr>
<td>12</td>
<td>406 ASP OD2</td>
<td>3.03</td>
<td>428 ASN ND2</td>
</tr>
<tr>
<td>13</td>
<td>408 TRP NE1</td>
<td>2.89</td>
<td>428 ASN O</td>
</tr>
<tr>
<td>14</td>
<td>409 LEU O</td>
<td>3.46</td>
<td>256 ASN ND2</td>
</tr>
<tr>
<td>15</td>
<td>420 ARG NH2</td>
<td>2.84</td>
<td>290 ASN O</td>
</tr>
</tbody>
</table>

### Salt Bridges

<table>
<thead>
<tr>
<th>##</th>
<th>Jellyroll</th>
<th>Dist. [Å]</th>
<th>Subtilase</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>406 ASP OD1</td>
<td>2.74</td>
<td>462 ARG NH1</td>
</tr>
<tr>
<td>2</td>
<td>406 ASP OD1</td>
<td>3.46</td>
<td>462 ARG NH2</td>
</tr>
<tr>
<td>3</td>
<td>406 ASP OD2</td>
<td>3.77</td>
<td>462 ARG NH1</td>
</tr>
<tr>
<td>4</td>
<td>406 ASP OD2</td>
<td>3.01</td>
<td>462 ARG NH2</td>
</tr>
</tbody>
</table>

Table 2.9.S1.3.: Jellyroll domain interaction interface with prodomain: hydrogen bonds.

### Hydrogen Bonds

<table>
<thead>
<tr>
<th>##</th>
<th>Jellyroll</th>
<th>Dist. [Å]</th>
<th>Prodomain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>362 TYR O</td>
<td>2.76</td>
<td>10 ASP OD2</td>
</tr>
</tbody>
</table>

The distance between two atoms is shown for each bond; residue number, residue 3-letter code, and atom name specify each atom involved.
Table 2.9.S2.: Strains and plasmids used in this study.

<table>
<thead>
<tr>
<th>Strain #</th>
<th>C. difficile strain</th>
<th>Relevant genotype or features</th>
<th>Source/reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>JIR8094</td>
<td>Erm-sensitive derivative of 630</td>
<td>C. Ellermeier [97]</td>
</tr>
<tr>
<td>13</td>
<td>630</td>
<td>Clinical isolate 630</td>
<td>T. Lawley [63]</td>
</tr>
<tr>
<td>30</td>
<td>cspBAC</td>
<td>JIR8094 cspBAC::ermB</td>
<td>This study</td>
</tr>
<tr>
<td>47</td>
<td>sleC</td>
<td>JIR8094 sleC::ermB</td>
<td>This study</td>
</tr>
<tr>
<td>84</td>
<td>JIR8094/pMTL84151</td>
<td>JIR8094/pMTL84151</td>
<td>This study</td>
</tr>
<tr>
<td>88</td>
<td>cspBAC/pMTL84151</td>
<td>JIR8094 cspBAC::ermB/pMTL84151</td>
<td>This study</td>
</tr>
<tr>
<td>111</td>
<td>cspBAC/pMTL83151</td>
<td>JIR8094 cspBAC::ermB/pMTL83151- cspBAC</td>
<td>This study</td>
</tr>
<tr>
<td>113</td>
<td>cspBAC/pMTL83151- cspBAC S461A</td>
<td>JIR8094 cspBAC::ermB/pMTL83151- cspBAC S461A</td>
<td>This study</td>
</tr>
<tr>
<td>193</td>
<td>cspBAC/pMTL83151-cspBAC Δjelly</td>
<td>JIR8094 cspBAC::ermB/pMTL83151-cspBAC Δjelly</td>
<td>This study</td>
</tr>
<tr>
<td>197</td>
<td>cspBAC/pMTL84151- cspBAC Q66/S461A</td>
<td>JIR8094 cspBAC::ermB/pMTL84151-cspBAC Q66/S461A</td>
<td>This study</td>
</tr>
<tr>
<td>228</td>
<td>cspBAC/pMTL84151- cspBAC Q66</td>
<td>JIR8094 cspBAC::ermB/pMTL84151-cspBAC Q66</td>
<td>This study</td>
</tr>
<tr>
<td>234</td>
<td>cspBAC/pMTL84151- cspBAC Q66</td>
<td>JIR8094 cspBAC::ermB/pMTL84151-cspBAC Q66</td>
<td>This study</td>
</tr>
</tbody>
</table>

E. coli strains with C. difficile constructs

<table>
<thead>
<tr>
<th>Strain (#)</th>
<th>Relevant genotype or features</th>
<th>Source/reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>DH5a</td>
<td>F– Φ80lacZΔM15 Δ(lacZYA argF) U169 recA1 endA1 hsdR17 (rK−, mK+) phoA supE44 λ– thi-1 gyrA96 relA1</td>
<td>D. Cameron</td>
</tr>
<tr>
<td>BL21(DE3)</td>
<td>F– ompT hsdSB(rB−, mB+) gal dcm (DE3)</td>
<td>Novagen</td>
</tr>
<tr>
<td></td>
<td>F– mcrB mrr hsdS20(rB+, mB+) recA13 leuB6 ara-13 proA2 lacYI galK2 xyl-6 mtl-1 rpsL20</td>
<td>C. Ellermeier</td>
</tr>
<tr>
<td>HB101</td>
<td>pK424 in HB101</td>
<td>C. Ellermeier</td>
</tr>
<tr>
<td>7</td>
<td>pET22b in DH5a</td>
<td>D. Higgins</td>
</tr>
<tr>
<td>269</td>
<td>pET28a in DH5a</td>
<td>M. Bogyo</td>
</tr>
<tr>
<td>548</td>
<td>pRSFDuet1 in DH5a</td>
<td>Novagen</td>
</tr>
<tr>
<td>556</td>
<td>pJS107 in DH5a</td>
<td>J. Sorg</td>
</tr>
<tr>
<td>686</td>
<td>pMTL83151 in HB101/pK424</td>
<td>This study</td>
</tr>
<tr>
<td>655</td>
<td>pMTL83151 in DH5a</td>
<td>This study</td>
</tr>
<tr>
<td>455</td>
<td>pET28a-cspB(548aa) in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>471</td>
<td>pET28a-cspB(548aa)-S461A in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>493</td>
<td>pET28a-cspB(548aa)-QTQ/AAA in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>514</td>
<td>pET28a-cd1433 in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>516</td>
<td>pET22b-cspC (CD2246) in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>533</td>
<td>pET22b-sleC in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>604</td>
<td>pJS107 cspBA 81 in HB101/pK424</td>
<td>This study</td>
</tr>
<tr>
<td>646</td>
<td>pJS107-sleC 128 in HB101/pK424</td>
<td>This study</td>
</tr>
<tr>
<td>667</td>
<td>pMTL83151-cspBAC in HB101/pK424</td>
<td>This study</td>
</tr>
<tr>
<td>799</td>
<td>pMTL83151-cspBAC S461A in HB101/pK424</td>
<td>This study</td>
</tr>
<tr>
<td>800</td>
<td>pMTL83151-cspBAC Δjelly in HB101/pK424</td>
<td>This study</td>
</tr>
</tbody>
</table>
This study

**E. coli strains with C. perfringens constructs**

<table>
<thead>
<tr>
<th>Strain #</th>
<th>Relevant genotype or features</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>506</td>
<td>pET22b-cspB in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>673</td>
<td>pRSFduet1-cspBΔ96 in DH5a</td>
<td>This study</td>
</tr>
<tr>
<td>674</td>
<td>pRSFduet1-cspBΔ96 in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>711</td>
<td>pRSFduet1-cspB in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>712</td>
<td>pRSFduet1-cspB(1-96)/cspBΔ96 in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>713</td>
<td>pRSFduet1-cspB Δjelly (1-66)/cspBΔ96 in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>714</td>
<td>pRSFduet1-cspC perfringens (1-78)/cspBΔ96 in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>742</td>
<td>pRSFduet1-cspB Δjelly in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>744</td>
<td>pRSFduet1-cspB Δjelly/S461A in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>745</td>
<td>pRSFduet1-cspB K91D in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>747</td>
<td>pRSFduet1-cspB S494A in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>753</td>
<td>pRSFduet1-cspB YTS/AAA in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>786</td>
<td>pET22b-cspB jelly (293-424 aa) in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>788</td>
<td>pRSFduet1-cspB(1-92)/cspBΔ96 in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>789</td>
<td>pRSFduet1-cspB(1-93)/cspBΔ96 in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>791</td>
<td>pRSFduet1-cspB R231E in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>792</td>
<td>pRSFduet1-cspB ΔYTS in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>793</td>
<td>pRSFduet1-cspB R231Q in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>820</td>
<td>pRSFduet1-cspB E35R in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>821</td>
<td>pRSFduet1-cspB E59A in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>822</td>
<td>pRSFduet1-cspB E35R-R231E in BL21(DE3)</td>
<td>This study</td>
</tr>
<tr>
<td>823</td>
<td>pRSFduet1-cspB S96R in BL21(DE3)</td>
<td>This study</td>
</tr>
</tbody>
</table>

**Plasmids**

<table>
<thead>
<tr>
<th>Strain #</th>
<th>Relevant genotype or features</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pET22b</td>
<td>bla</td>
<td>Novagen</td>
</tr>
<tr>
<td>pET28a</td>
<td>kan</td>
<td>Novagen</td>
</tr>
<tr>
<td>pRSFduet1</td>
<td>kan</td>
<td>Novagen</td>
</tr>
<tr>
<td>pK424</td>
<td>Tra Mob⁺, bla, tet</td>
<td>C. Ellermeier</td>
</tr>
<tr>
<td>pJS107</td>
<td>C. difficile Targetron construct based on pJIR750ai (group II intron ermB::RAM, ltrA); catP</td>
<td>J. Sorg</td>
</tr>
<tr>
<td>pCE245</td>
<td>C. difficile Targetron construct based on pJIR750ai (group II intron ermB::RAM, ltrA); catP</td>
<td>C. Ellermeier</td>
</tr>
<tr>
<td>pMTL83151</td>
<td>pCB102, Tra⁺; catP</td>
<td>N. Minton [110]</td>
</tr>
<tr>
<td>pMTL84151</td>
<td>pCD6, Tra⁺; catP</td>
<td>N. Minton [110]</td>
</tr>
</tbody>
</table>
Table 2.9.S2.: Primers used in this study.

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>457</td>
<td>5' Neol cspBA start</td>
<td>ATACCATGGCTATTATAATAATATATGAAATTATTTGTA</td>
</tr>
<tr>
<td>459</td>
<td>3' Xhol cspBA 548 aa</td>
<td>GCCACTCGAGTGAATTTAATATACCTTATATAC</td>
</tr>
<tr>
<td>461</td>
<td>5' NdeI sleC</td>
<td>AGCATATGGAAGATGTGTTCTAAAAAGATTAATTTTCTTGA</td>
</tr>
<tr>
<td>463</td>
<td>3' Xhol sleC no stop corr</td>
<td>AGTGGGCACTTAAACTGGAACCTCCATGGCTACACCT CATGTTACA</td>
</tr>
<tr>
<td>464</td>
<td>5' S461A SOE</td>
<td>TGTAACATGAGGTAGTGCATTCCAGTCTAAGAGTTC</td>
</tr>
<tr>
<td>465</td>
<td>3' S461A Roes</td>
<td>TGCTCCACT</td>
</tr>
<tr>
<td>502</td>
<td>5' QTQ/AAA cspBA SOE</td>
<td>GAATTGCAAAAGTTTTTTGA</td>
</tr>
<tr>
<td>503</td>
<td>3' QTQ/AAA cspBA Roes</td>
<td>TGAAAAAACCTTGGCAACCTCCAGGCAGCCGAATTCATAA</td>
</tr>
<tr>
<td>512</td>
<td>5' NdeI cspB perf ATCC 13124</td>
<td>AGCCATATGGAAATAAAGCTAAGGTAGTC</td>
</tr>
<tr>
<td>513</td>
<td>3' Xhol cspB perf ATCC 13124</td>
<td>AGCCATCGATGTCTCTTATTAATATAATTCCATCA</td>
</tr>
<tr>
<td>516</td>
<td>5' S494A cspB perf SOE</td>
<td>CCACAA</td>
</tr>
<tr>
<td>517</td>
<td>3' S494A cspB perf Roes</td>
<td>TTGGCCGCAGCCATTGTGACCTACACTTTTTTAGTACA AAATCC</td>
</tr>
<tr>
<td>522</td>
<td>5' Ncol cd1433 start</td>
<td>AGGCATATGGCAATGTTACATGCCAAATTTGCCCA</td>
</tr>
<tr>
<td>523</td>
<td>3' Xhol cd1433 no stop</td>
<td>AGCTCTCGAGGAAATGCCTCAGTAAAC</td>
</tr>
<tr>
<td>524</td>
<td>5' NdeI cspC</td>
<td>AGGCATATGGAAATAAATTTGTGATAATTTCAATCAA</td>
</tr>
<tr>
<td>525</td>
<td>3' Xhol cspC no stop</td>
<td>TTGAACCGAAGTTTTCTATAATTCGTTGCAATCCATGAG</td>
</tr>
<tr>
<td>533</td>
<td>5' EBS2 cspBA 81</td>
<td>AGGAAAAATGTCT</td>
</tr>
<tr>
<td>534</td>
<td>3' EBS1d cspBA 81</td>
<td>CAGATTGTGACAAATAATGTTGATAACAGATAATGCTA</td>
</tr>
<tr>
<td>535</td>
<td>5' IBS1 cspBA 81</td>
<td>GGCACCCAGATAGGTG</td>
</tr>
<tr>
<td>536</td>
<td>5' CDEP692</td>
<td>GGAATTTCAGATTCCTGAG</td>
</tr>
<tr>
<td>537</td>
<td>5' CDEP1070a</td>
<td>AAAACGTATAAAATAGGAGGG</td>
</tr>
<tr>
<td>538</td>
<td>3' CDEP1138</td>
<td>TTTCCGCTGCGAGCTTAAAGCA</td>
</tr>
<tr>
<td>641</td>
<td>5' IBS1 sleC 128a</td>
<td>AAAAAGCTTATAATATTATCTTACATTACTTCTTGA</td>
</tr>
<tr>
<td>642</td>
<td>3' EBS1d sleC 128a</td>
<td>CTATTGTAACCTAATTTTTGT</td>
</tr>
<tr>
<td>643</td>
<td>5' EBS2 sleC 128a</td>
<td>TGAACGCAAGTTTCTAATTTTCCGTTTAAATGCTGAG</td>
</tr>
<tr>
<td>665</td>
<td>3' Xhol cspC cd2246 + TAA</td>
<td>AAAACGTGACCAATAAAATGTTGATACAGATAATGCTT</td>
</tr>
<tr>
<td>668</td>
<td>5' Ncol cspB perf ATG</td>
<td>AAATGTCCATGGAAATAAAGCTAAGGTTGAGC</td>
</tr>
<tr>
<td>670</td>
<td>5' NdeI cspB perf Δ96</td>
<td>AAATGTCCATGGACCTTATGATGATATAAAATTTTTGGTA GCTC</td>
</tr>
<tr>
<td>671</td>
<td>3' Xhol 6His cspB perf</td>
<td>AAATGTCCATGACCTTACGAGCTGTCGATTTACATTATATACGAGCATCATGC</td>
</tr>
<tr>
<td>691</td>
<td>3' Nol cspB perf cd2247 upstream</td>
<td>AAATGTCCATGACCTTATGATGATATAAAATTTTTGGTA GCTC</td>
</tr>
<tr>
<td>701</td>
<td>3' Sall cspB perf pro TAA</td>
<td>AAATGTCCGCGCTATTAGGTTGATGATATAAAATTTTTGGTA GCTC</td>
</tr>
<tr>
<td>702</td>
<td>3' Sall cspBA diff pro TAA</td>
<td>AAATGTCCGCGCTATTAGGTTGATGATATAAAATTTTTGGTA GCTC</td>
</tr>
<tr>
<td>703</td>
<td>5' Ncol cspC perf</td>
<td>TTACATGGAACACTGTTACGAGCTGTCGATTTACATTATATACGAGCATCATGC</td>
</tr>
<tr>
<td>704</td>
<td>3' Sall cspC perf pro TAA</td>
<td>AAATGTCCGCGCTATTAGGTTGATGATATAAAATTTTTGGTA GCTC</td>
</tr>
<tr>
<td>709</td>
<td>5' cspB perf K91D SOE</td>
<td>AAATGTCCGCGCTATTAGGTTGATGATATAAAATTTTTGGTA GCTC</td>
</tr>
<tr>
<td>83</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Restriction enzyme sequences are underlined; point mutations are in bold italics; deletion sites are indicated by |.
Table 2.9.S3.: Data collection and refinement statistics.

<table>
<thead>
<tr>
<th>Data collection</th>
<th>SeMet CspB</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space group</strong></td>
<td>P2₁2₁2₁</td>
</tr>
<tr>
<td><strong>Cell dimensions</strong></td>
<td></td>
</tr>
<tr>
<td>(a, b, c) (Å)</td>
<td>73.87, 138.17, 140.08</td>
</tr>
<tr>
<td>(\alpha, \beta, \gamma) (°)</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td><strong>Wavelength (Å)</strong></td>
<td>0.9794</td>
</tr>
<tr>
<td><strong>Resolution (Å)</strong></td>
<td>50-1.6 (1.618-1.6)*</td>
</tr>
<tr>
<td>(R_{\text{merge}}) (%)</td>
<td>7.8 (45.1)*</td>
</tr>
<tr>
<td>(I/\sigma I)</td>
<td>18.43 (2.05)*</td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
<td>94.66 (70.0)*</td>
</tr>
<tr>
<td><strong>Redundancy</strong></td>
<td>6.6 (2.8)*</td>
</tr>
</tbody>
</table>

**SAD Phasing Statistics**

| Selenium sites (ShelX/D) | 12 |
| PATFOM (/D) | 24.47 |
| Overall CC (ShelX/E) (%) | 46.27 |
| Pseudo-Free CC (/E) (%) | 81.61 |
| Contrast/Connectivity (/E) | 0.803/0.957 |
| Final Map CC (/E) | 0.936 (0.936)* |

**Refinement**

| Resolution (Å) | 43-1.6 |
| No. reflections | 1,332,184 total |
| 202,733 unique |
| \(R_{\text{work}} / R_{\text{free}}\) | 0.15/0.18 |
| No. atoms | 9766 |
| Protein | 8772 |
| Ligand/ion | 37 |
| Water | 957 |
| \(B\)-factors (Å²) | 28.6 |
| Wilson B | 19.1 |
| Protein | 27.5 |
| Ligand/ion | 33.5 |
| Water | 38.2 |
| R.m.s. deviations |  |
| Bond lengths (Å) | 0.0065 |
| Bond angles (°) | 0.844 |

*Values in parentheses are for highest-resolution shell.

*Statistics for phasing as defined by ShelX/C/D/E program authors.⁴,¹².
CHAPTER 3: TOWARDS AN UNDERSTANDING OF THE SPORE CORTEX

LYTIC ENZYME IN *C. DIFFICILE*, SleC

3.1. Introduction

*Clostridium difficile* is a costly and common hospital-acquired infection largely due to its ability to form metabolically dormant and resistant spores. These spores are able to survive harsh cleaning protocols in place at hospitals and yet remain poised to germinate as soon as the proper environment is detected in the form of specific bile salts[55].

The spore consists of the spore core, the innermost layer, which is composed of the dehydrated components necessary to form a vegetative cell- the genetic material, ribosomes, enzymes, etc. For a figure illustrating the components of *C. difficile* spores, turn to Figure 1.6.2 for electron micrographs, or 1.6.3 for a cartoon depiction. Outside of the spore core is a membrane, called the germ cell wall, which will become part of the cell wall of the new vegetative cell formed of the spore core when germination is complete. The spore cortex, outside of this inner membrane, consists of loosely cross-linked peptidoglycan residues[137]. It is the spore cortex that maintains metabolic dormancy by acting as a mechanical barrier to water and other small molecules, and keeping the core dehydrated; thus, the cortex is critical for maintaining spore resistance[64]. Outside of the cortex there is an additional membrane and a proteinaceous spore coat. In some *Clostridiales* and *Bacillales* species, including *C. difficile*, there is a loose-fitting exosporium, outside of the coat, thought to be involved in resistance, adhesion to surfaces, and interactions with the host[58, 138].
As the spore cortex physically enforces spore dormancy in every spore-forming bacterium, there are necessarily enzymes commonly called Spore Cortex Lytic Enzymes (SCLEs), which are able to digest the spore cortex peptidoglycan during germination. Without degradation of the cortex, there is no germination or hydration, and therefore no vegetative cell. Conversely, if cortex degradation is initiated at the wrong time, the spore will lose the cortex and the resistance it confers. Finally, if SCLEs were active against vegetative cell wall peptidoglycan, they could cause lysis of the outgrowing or vegetative cells. Therefore, these enzymes need to be tightly regulated, act only on specific cortex peptidoglycan residues, and only upon response to a germination signal.

In *Bacillus subtilis*, the model organism for spore formation and germination, there are three types of enzyme activity that must be accounted for: *N*-acetylglucosaminidases, lytic transglycosylases, and amidases[139]. Each of these has specificity for a particular bond found uniquely within the spore cortex peptidoglycan structure (see Figure 3.5.1.). SCLEs from both *Bacillus* and *Clostridium* require the presence of the muramyl-δ-lactam residue binding within the catalytic cleft in a specific position relative to the scissile bond[140]. This requirement provides specificity for spore cortex peptidoglycan and prevents SCLEs from hydrolyzing vegetative cell wall peptidoglycan, which does not contain muramyl-δ-lactam[141]. In *Bacillus* sp., two such SCLEs are SleB and CwlJ, and either is sufficient for complete spore cortex degradation[56]. Though *C. difficile* appears to have homologues to SleB and CwlJ, they are not necessary for germination, and little is known about how they function[88]. In *C. perfringens*, there are only two known SCLEs: SleC and SleM, and only SleC is required for germination[89]. In *C. perfringens*, SleC is expressed in the sporulating cell and remains as an inactive zymogen until needed (see
SleC, which is released into the medium by germinating spores, has two pre-sequences that are cleaved during sporulation, both a 125 amino-acid N-terminal sequence called the predomain, and a much smaller C-terminal sequence, flanking 3 interior divisions: the propeptide (approximately, amino acids 125-155), the catalytic SleC domain (amino acids 156-336), and the peptidoglycan binding domain (amino acids 337-423) (see Figure 3.5.1). Although the significance of the processing of the C-terminal sequence is not known, the N-terminal predomain functions as an intramolecular chaperone and is essential for proper folding of SleC. After cleavage of the N-terminal predomain, the inactive zymogen, pro-SleC, remains inactive in the dormant spore until a germination-specific subtilisin-like serine protease, CspA, CspB, and/or CspC, removes the propeptide sequence in response to germinant. Although many clostridial species encode multiple Csps, we showed in Chapter 2 that CspB alone is sufficient for proteolytically activating pro-SleC.

The spore cortex in C. difficile is made up of three types of peptidoglycan residues: N-acetylglucosamine (GlcNAc), N-acetylmuramic acid (MurNAc), and muramic acid δ-lactam. SleC is predicted to be a monofunctional and processive lytic transglycosylase, acting on the bond between MurNAc and GlcNAc, in contrast to the C. perfringens SleC, which is bifunctional as both a lytic transglycosylase and an amidase. Spore cortex peptidoglycan is composed of alternating GlcNAc and MurNAc residues, where every alternate MurNAc is substituted with muramic acid δ-lactam. Crosslinking occurs between peptide chains off of muramic acid δ-lactam residues on adjacent strands to a total of only about 3% (in contrast to vegetative cell wall, which is about 20-30% cross-linked). Aside from a paper by Christopher Reid’s
group that found that *C. difficile* SleC does not require proteolytic activation for activity against *B. subtilis* cortex peptidoglycan *in vitro*[139], little is known of the differences in the processing and function of SleC in *C. difficile* compared to *C. perfringens*. A recent paper suggests that *C. perfringens* is not a good choice of model for understanding *C. difficile* germination, since *C. difficile* should be re-classified as *Peptoclostridium difficile*, while *C. perfringens* is correctly classified as a *Clostridium* sp. The two have more divergently-evolved germination pathways than previously thought (see Figure 1.6.4 for the previous classification (*C. difficile*) and proposed reclassification (*P. difficile*))[75]. Additionally, unpublished evidence from our own lab suggests that *C. difficile* SleC does not remain associated with the cleaved predomain or require cleavage of the predomain prior to activation of the zymogen (as *C. perfringens* SleC does[69]). However, the predomain is still required for proper folding of SleC in our hands, similar to *C. perfringens*[69]. In contrast with the *in vitro* analyses of Gutelius *et al.*, we presented in Chapter 2 that pro-SleC cleavage in *C. difficile* is positively correlated with germination efficiency[76]. In this same chapter we demonstrated that the predomain processing happens in SleC *difficile* as it does in *C. perfringens*, prior to incorporation into spores (see Figure 2.7.1).

To understand more about SleC from *C. difficile*, we sought to determine its crystal structure. To this end, we generated multiple expression constructs to identify the best construct for crystallization as well as to assess the role of the pro-peptide in regulating SleC folding. PONDR disorder-prediction software suggests that SleC is moderately disordered, especially in the predomain, and the transition between SleC hydrolase domain and substrate binding domain (see Figure 3.5.2.)[144-146]. POODLE,
another disorder predicting site, gave similar results[147]. PONDR’s algorithm considers the amino acid sequence in a 9-amino-acid window around a particular residue to predict order or disorder, and other predictors of order and disorder operate in a similar way, so it is possible that the predomain is much more ordered in the context of the rest of the sequence. The other large spike of disorder is predicted to occur around residue 350, near the N-terminus of the peptidoglycan-binding domain. It is possible that a flexible loop separates the peptidoglycan-binding domain from the hydrolase domain; this loop could function to optimize binding of the PGB domain to its substrate. The presence of a flexible loop would likely result in this prediction of disorder. Structures of similar enzymes from Bacillus have not included the peptidoglycan-binding domain; indeed, only the catalytic domain of SleB has been crystallized[141, 148].

3.2. Materials and Methods

3.2.1. Constructing SleC constructs for expression in E. coli.
Strains generated are illustrated in Figure 4.5.1, and listed in Table 3.6.1. The generation of a SleC construct containing a TEV cleavage site would allow us to carry out protein expression and folding using the complete sequence and the predomain cleaved off after protein purification with the TEV protease. The TEV cleavage site consists of the amino acid sequence Glu-X-X-Tyr-X-Gln-Gly/Ser, where the X’s can represent a range of different amino acids and cleavage occurs between Gln and Gly/Ser[149]. We used ENLYFQG, inserted in between the pre- and pro- domains.

We used strain #965 to explore the role of the propeptide of SleC in expression and solubility. To generate strain #965, we used splicing by overlap extension (SOE) PCR
to remove the sequence encoding the 33 amino acids (122-156) of the propeptide. First, the N terminal region was amplified using primers #461 and #1090 (see Table 3.6.2 for a list of primers and sequences), and the C terminal portion was amplified using primers #1089 and #463. The primers were constructed such that the 3’ primer of the N-terminal amplicon primer contained a 30 amino acid overlap with the 5’ primer for the C-terminal region- the overlap was designed so that the propeptide was missing and the overlap occurred on the domain immediately N or C terminal to it. Then, the SOE PCR was carried out using the amplified DNA from the first two reactions, and the 5’ and 3’ primers of the first reaction and second reaction, respectively, #461 and #463. The resulting amplified DNA encoded amino acids 1-122 ligated to 156-423. This was digested with NdeI and XhoI, ligated into pET22b, and transformed into DH5α, the resulting plasmid was transformed into BL21(DE3) for protein expression.

We used strain #933 to explore the ‘active’ SleC construct, that is, the form of SleC that exists in germinating spores and carries out the hydrolase activity. To create strain #933, carrying a construct encoding amino acids 156-423, containing the hydrolase domain as well as the peptidoglycan binding domain, primers #636 and #463 were used. Strain #933 was subsequently used to provide plasmid for multiple strains, see below.

To learn more about the hydrolase domain in the absence of the other domains, we used strain #967. This strain, carrying a construct with amino acids 156-336 (the hydrolase domain only), was generated as above, using primers #636 and #477.

To explore constructs that could be purified away from the predomain and propeptide, we created the following strains. To create strain #1003, carrying the region encoding amino acids 1-156 co-expressed with amino acids 156-423, the plasmid isolated
from strain #986 (primers #469 and #1076, ligated into pET28a) was used to express the predomain and propeptide, and the plasmid isolated from strain #933 (primers #636 and #463, ligated into pET22b) was used to express the hydrolase and substrate binding domains. Strain #1004 was created in the same manner, carrying the hydrolase and substrate binding domains, using plasmid isolated from strain #933, co-expressed with only empty vector pET28a as a control.

To create strain #1159, carrying the region encoding amino acids 1-123 co-expressed with amino acids 156-423 (with the propeptide missing), the plasmid contained in strain #485 was used to express the predomain, and the plasmid contained in strain #933 for the hydrolase and substrate binding domains.

3.2.2. Protein Purification of His$_6$-tagged proteins.

For IPTG induction, overnight cultures of the appropriate BL21(DE3) strain were diluted approximately 1:500 in 2xYT (5 g NaCl, 10 g yeast extract, 15 g tryptone/L) media and grown shaking (225 rpm) at 37°C. When an OD$_{600}$ of 0.6-0.9 was reached, IPTG was added to 250 µM, and cultures were grown for 12-16 hr at 19°C.

We found autoinduction improved the amount of protein isolated from cells, and thus this protocol was used in nearly all protein expression experiments[150]. For autoinduction, the overnight culture of BL21(DE3) was diluted approximately 1:500–1:1000-fold in 1 liter terrific broth supplemented with 5052 sugar mix to 5 g glycerol, .5 g glucose and 2 g lactose per liter of medium. The cultures were incubated at 20 °C with shaking (at 225 rpm) for the ~60-70 hour growth period.
Cells were collected by centrifugation, transferred to a falcon tube for freezing, and stored at -20 °C until ready for use. Frozen pellets were allowed to thaw in ice water and resuspended in lysis buffer [500 mM NaCl, 50 mM Tris-HCl, pH 7.5, 15 mM imidazole, 10% v/v glycerol]. Protease inhibitor (PMSF, phenylmethylsulfonyl fluoride, Pierce) was added to resuspended lysates to a final concentration of approximately 0.1mM. Cells were then lysed by sonication and clarified by centrifugation at 14,000 x g for ~50 minutes.

Cleared lysates were incubated with Ni-NTA agarose beads (Qiagen) at about 4 mL bead-slurry volume per ~45 g cell pellet for affinity purification with gentle rocking overnight at 4 °C. The binding reaction was pelleted by gentle centrifugation, the supernatant set aside, and the pelleted Ni-NTA agarose beads were divided into ~8-10 Eppendorf tubes and washed three times with lysis buffer. His6-tagged proteins were eluted from the beads by the addition of 350 µL high imidazole elution buffer [500 mM NaCl, 50 mM Tris-HCl, pH 7.5, 175 mM imidazole, 10% v/v glycerol]. The elution was repeated a second and third time; the eluate was pooled, buffer exchanged in anion exchange buffer A [150 mM NaCl, 10 mM Tris pH 7.5, 5% v/v glycerol], and concentrated to 750 µL. The concentrated sample was pelleted at 13,000 x g for 10 min at 4°C prior to loading on an anion exchange column (GE Healthcare). Buffer B (750mM NaCl, otherwise as buffer A) concentration was gradually increased throughout the protocol from 100% A, 0% B to 0% A, 100% B.
3.2.3. Crystallization studies

For crystallization studies, His\textsubscript{6}-tagged SleC \textit{difficile} (for location of His-tag, see Figure 3.5.1) constructs were affinity purified as described above, then further purified using an anion exchange column (Hi-Trap Q HP, GE Healthcare); buffer A was 150 mM NaCl, 10 mM Tris-HCl pH 7.5, 10% v/v glycerol and buffer B was 750mM NaCl, 10mM Tris-HCl pH 7.5, 10% v/v glycerol. The resulting single peak fractions were pooled and purified protein was concentrated to between 3 and 25 mg/mL depending on solubility of purified protein and yields. Protein purity was analyzed by SDS-PAGE followed by Coomassie staining (GelCode Blue, Pierce) or a western blot visualized with our own \(\alpha\)-SleC antibody.

Preliminary crystallization trials were carried out using Crystal Screen (Hampton Research) and SleC #464 at 8 mg/ml in hanging-drop vapor diffusion trays. Trays were incubated at 12°C and checked daily for signs of crystal growth for the first week and less frequently after that. We adjusted the ratio of protein to buffer in the drops, tried both sitting drops and hanging drops, changed the size of the drops, and adjusted the pH or ionic strength of the mother liquor. These attempts to further optimize crystallization results did not demonstrate improvement in crystal growth.

3.3. Results

3.3.1. Some SleC constructs were more soluble than others.

Several SleC expression plasmids were constructed (See Figure 3.5.1 for schematics of the constructs). The rationale for different constructs was to enable us to solve the structures of different forms of SleC. Specifically, comparing the structures of
pro-SleC with active SleC (which is missing the propeptide), would allow us to answer questions such as: does the propeptide rest in the catalytic cleft to block access by substrates when it is present? Another motivation was to visualize the relationship between the peptidoglycan binding domain and the hydrolase domain, and to have the ability to express, purify, and crystallize a form of SleC without it, especially knowing that all of the deposited structures of similar enzymes lack the peptidoglycan binding domain. Constructs were expressed in BL21(DE3) cells and were induced with IPTG or autoinduction[150]. Autoinduction cultures produced far more protein than did IPTG induction and was used more often. The His-tagged constructs were purified on Ni-NTA beads followed by an anion exchange column (Hi-Trap Q HP, GE Healthcare). The average yield varied widely, as some constructs could not be purified in any appreciable amount while others expressed and purified with good yields (see Figure 3.5.1 for relative solubility of each construct). A good yield for C. difficile SleC was about 20 mg of protein for 1 liter of autoinduction culture. Soluble SleC was concentrated to about 25 mg/mL prior to crystallization trials.

3.3.2. **The predomain is essential for proper folding of SleC**

In our studies, we found that the predomain was essential for proper folding and solubility. Without the predomain, relative expression was poor, suggesting that the protein was improperly folded and degraded or in the insoluble fraction. Similar results were found with SleC from C. perfringens by Okamura and colleagues in 2000[69]. While *Clostridium perfringens* and *Peptoclostridium difficile* are actually quite distantly
related[75], our results suggest that the regulation of the SCLE may be similar, with the predomain functioning as an intramolecular chaperone necessary for folding.

We were able to determine that the predomain does not have to belong to the same peptide chain in order to carry out its chaperone function. Separate peptides containing the predomain and the pro-SleC domains were still able to produce reliable protein expression levels. See Figure 3.5.1 to see the co-expressed strains #464 (pre/pro-SleC-PG), #1003 (pre-pro/SleC-PG), #1159 (pre/SleC-PG), and #1038 (pre-pro/SleC), #1160 (pre/SleC), and #1161 (pre/pro-SleC) featuring the co-expressed predomain. We found a slight increase in protein yield when the propeptide was associated with the hydrolase instead of the predomain (compare #464, #495 (both are pre/pro-SleC-PG), and #1161 (pre/pro-SleC) versus #1038 (pre-pro/SleC) and #1003 (pre-pro/SleC-PG) in Figure 3.5.1). Because activation of SleC requires cleavage of the propeptide, perhaps the orientation of the propeptide when it is associated with the hydrolase serves to stabilize the enzyme. Indeed, the best expression in trans constructs was observed in those in which the predomain was co-expressed and the propeptide was associated with the hydrolase domain.

3.3.3. TEV construct to allow post-purification cleavage of the SleC predomain

In the case of strain #626, we created a recombinant SleC in which the predomain was separated from pro-SleC by a TEV cleavage site insert. Thus, we could purify full length protein and separate the predomain (predicted to be disordered) before attempting to crystallize the protein, while having the benefit of the predomain for expression and preliminary purification steps. We were unable to see protease cleavage of this construct, although the TEV protease was active. We suspect that either a) the location of the
recognition sequence was buried and inaccessible or b) the absence of a linker region flanking either side of the sequence led to improper secondary structure, influenced by the local structure of SleC, and thus a lack of recognition by TEV or steric occlusion of the scissile bond.

3.3.4. Presence of the peptidoglycan binding domain has little impact on expression levels

We found that the presence or absence of the peptidoglycan binding domain did not impact the expression levels of our recombinant SleC constructs. In Figure 3.5.1, the relative expression levels of pre/pro-SleC constructs such as #1161, and pre/pro-SleC-PG constructs such as #464 and #495 were approximately the same. The substrate binding domain serves to position the enzyme in relative proximity to the substrate, but could be located distant from the catalytic domain, so perhaps the expression of SleC is independent of the presence or absence of the peptidoglycan binding domain.

3.3.5. Crystallization attempts were unsuccessful

The recombinant constructs that were promising in expression and purification steps (full-length wild-type SleC, #495, #464 and #1003) were used in crystallization trays using pre-formulated kits of reagents available from Hampton Research. We found that all of the constructs gave very similar results and formed shiny clusters in a great variety of conditions (Figure 3.5.5). However, despite repeated attempts to improve on these initial hits, we were never able to progress past this stage or obtain any different crystalline morphology. Crystal clusters were present in all of the following conditions: 0.2M
ammonium sulfate, 30% w/v PEG 8000; 4M sodium formate; 0.1M sodium acetate trihydrate pH 4.6, 2M sodium formate; 0.5M sodium chloride, 0.01M magnesium chloride hexahydrate, 0.01M hexadecyltrimethylammonium bromide; 35% v/v 1,4-dioxane; 0.2M potassium thiocyanate, 20% PEG 3350; and 0.2M ammonium sulfate, 20% PEG 3350.

3.4. Discussion

The expression data suggest that the predomain serves as an intramolecular chaperone, essential for proper folding, in *C. difficile* as has been shown in *C. perfringens*[69]. The data we have gathered so far on the propeptide could support a role in stabilizing the catalytic domain of SleC, in addition to the evidence that this propeptide needs to be removed in *C. perfringens* in order to activate the enzyme. It is possible that the propeptide sterically occludes access to the catalytic cleft from possible substrates.

A recent paper by Christopher Reid’s group represents the best data besides our own data (presented in Chapter 2) about SleC regulation in *C. difficile*[139]. In this paper, Gutelius and colleagues referred to SleC with different descriptions for the domains that we have used in our own data. The domains they included consisted of an N-terminal region comprising amino acids 1-292 with no known function, a SpoIID/LytB domain between amino acids 293-341 containing the catalytic cleft, and the peptidoglycan binding domain between residues 342 to 423, similar to the boundaries of our own peptidoglycan binding domain. Contrary to our data showing that CspB processing of SleC is required for SleC activity *in vivo*, Gutelius and colleagues found that *in vitro* the activity of SleC does not depend on previous processing by germination-specific proteases such as CspB[139]. However they did not determine if this processing would increase SleC’s
catalytic efficiency. It is possible that this discrepancy between our results and theirs is entirely due to the differences between in vitro and in vivo processes. Gutelius et al. also observed that the absence of the peptidoglycan binding domain did not prevent activity, but rather slowed it down, suggesting that this domain functions as a processivity factor, but is not required for catalytic activity or binding of SleC to substrate[139].

We found SleC to be difficult in all aspects of experimentation—recombinant constructs often were poorly expressed and difficult to purify in sufficient amounts needed for crystallographic assays. Additionally, we were unable to obtain crystals of SleC despite repeated attempts to optimize plentiful preliminary hits. Perhaps the PONDR plot says it all—SleC is disordered and not a good target for crystallization studies (Figure 3.5.3 for disorder plot, Figure 3.5.5 for crystallization results). However, there is still much to learn about how the only known SCLE functions in C. difficile and structural data could play an important role in our understanding of the function of the individual domains of SleC. If the full-length enzyme could be crystallized, would we see the predomain as a sort of scaffold upon which the rest of the enzyme is built, similar to the propeptide intramolecular chaperone in CspB perfringens? Could we see the mechanism by which the propeptide prevents catalytic activity indicated by the structure, and find the propeptide draped across the catalytic cleft, blocking access to would-be substrates? Does the peptidoglycan domain rest apart from the catalytic core of the enzyme as a processivity factor, or is it positioned near the catalytic cleft to directly link substrate with active site?

There is much more to be learned about SleC, and perhaps the way forward consists of more patience in crystallization trials or identification of more unique
crystallization conditions (high salt or chaotropic agents). Another option, as we found helpful with CspB, would be to switch organisms to find a protein that will be more amenable to crystallization. We have previously tried SleC (full length) from *C. perfringens* in crystallization attempts without success. If that method is to be used, we have more guidance on the best options for related species due to proposed reclassification of some *Clostridium* spp. as *Peptoclostridium* spp., including *Clostridium bifermentans* or *sordellii*[75].
3.5. Figures

![Schematic illustrating the SleC constructs designed for crystallization.](image)

Figure 3.5.1. Schematic illustrating the SleC constructs designed for crystallization. Schematic showing different *C. difficile* SleC constructs designed for expression in *E. coli*. Predomain is drawn in blue, propeptide in purple, SleC hydrolase domain in light green, and peptidoglycan binding domain in dark green. A gap between domains indicates that these were co-expressed from separate plasmids. The 6-His tag for purification purposes is shown as a red line, and the TEV protease cleavage site is denoted by a lightning bolt. Relative expression levels to full length SleC is indicated with + or – signs.
Strains #464 and #495 (pre/pro-SleC-PG) are similar to the native form of SleC found in
*C. difficile* spores, with the understanding that the predomain does not appear to be
associated. Strains #464 and #495 closely represent the form of SleC observed by others in
*C. perfringens*, with the predomain cleaved but remaining noncovalently associated with
the pro-SleC form.
Figure 3.5.2. Schematic illustrating the three forms of SleC relevant to *C. difficile* spore germination.

The illustration shows the three relevant forms of SleC, as discussed in the text. Full length SleC is only present in sporulating cells, but in mature spores, the second form is present, having had the predomain processed away. This form is autoinhibited by the presence on the propeptide. The third and final form, active SleC, has had the propeptide cleaved off and is an active hydrolase.
Figure 3.5.3. PONDR prediction of ordered and disordered regions for full-length *C. difficile* SleC

Prediction of ordered and disordered regions in full length *C. difficile* SleC. A segment of protein with a PONDR score greater than 0.5 is expected to be disordered; with a score less than 0.5 it is more likely to be ordered. The black bar at the midline in the prediction for the predomain illustrates that this region is predicted to be highly disordered for about 50 residues. A disordered protein is less likely to be amenable to crystallization.
Strain #495 (pre/pro-SleC-PG) coomassie gel illustrating SleC construct NiNTA purification process

Strain #495 (pre/pro-SleC-PG) coomassie gel illustrating SleC construct purification. Construct #495 includes a 6His tag on both the pro-SleC-PG and the predomain, and leads to a high yield of the predomain copurifying with pro-SleC-PG. #495 pro-SleC-PG was very soluble, and purified nicely.
Figure 3.5.5. SleC crystallization trials.

Representative results of the most promising hits obtained for the SleC crystallization. These results occurred in a wide variety of crystallization solutions but could not be further improved. Condition shown is 400mM KCl, 12% PEG 3350. Upper right panel shows a zoomed-in version of the same image. Drops were 1 µL protein and 1 µL reservoir, for a starting size of 2 µL, and filled with soft crystalline phasing.
### 3.6. Tables

Table 3.6.1. Strains and plasmids used in SleC studies.

<table>
<thead>
<tr>
<th>Strain #</th>
<th>Relevant features</th>
<th>Plasmids used</th>
</tr>
</thead>
<tbody>
<tr>
<td>464</td>
<td>Pre/ pro-SleC-PG-6His</td>
<td>pET22b/pET28a</td>
</tr>
<tr>
<td>626</td>
<td>Pre-TEV-pro-SleC-PG-6His</td>
<td>pET21a</td>
</tr>
<tr>
<td>965</td>
<td>Pre-SleC-PG-6His</td>
<td>pET22b</td>
</tr>
<tr>
<td>933</td>
<td>SleC-PG-6His</td>
<td>pET22b</td>
</tr>
<tr>
<td>967</td>
<td>SleC-6His</td>
<td>pET22b</td>
</tr>
<tr>
<td>1003</td>
<td>Pre-pro/ SleC-PG-6His</td>
<td>pET28a/pET22b</td>
</tr>
<tr>
<td>1004</td>
<td>SleC-PG-6His with empty vector</td>
<td>pET22b/empty pET28a</td>
</tr>
<tr>
<td>1159</td>
<td>Pre/ SleC-PG-6His</td>
<td>pET28a/pET22b</td>
</tr>
<tr>
<td>495</td>
<td>6His-Pre/ pro-SleC-PG-6His</td>
<td>pET28a/pET22b</td>
</tr>
<tr>
<td>1038</td>
<td>Pre-pro/ SleC-6His</td>
<td>pET28a/pET22b</td>
</tr>
<tr>
<td>1160</td>
<td>Pre/ SleC-6His</td>
<td>pET28a/pET22b</td>
</tr>
<tr>
<td>1161</td>
<td>Pre/ pro-SleC-6His</td>
<td>pET28a/pET22b</td>
</tr>
</tbody>
</table>
Table 3.6.2. Primers used in SleC studies.

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>461</td>
<td>5’ NdeI sleC</td>
<td>AGCCATATGCAAGATGGTTTCTTAAACAGT AAGC</td>
</tr>
<tr>
<td>463</td>
<td>3’ Xhol sleC no stop corr</td>
<td>GCACTCGAGAATTTAAGGATTTAAAGAAN</td>
</tr>
<tr>
<td>469</td>
<td>5’ NcoI sleC GGA start</td>
<td>AGCCCATGGGACAAGATGTTTCTTAAACA GTAAGC</td>
</tr>
<tr>
<td>477</td>
<td>3’ Xhol sleC 334 aa no PG</td>
<td>GCACTCGAGTGCACACTGTGTCACCACCACACTC</td>
</tr>
<tr>
<td>636</td>
<td>5’ NdeI sleC activated 156 aa</td>
<td>AAATCATATGGTTAGATAAACCCTGAG TG</td>
</tr>
<tr>
<td>1076</td>
<td>3’ Xhol sleC REV</td>
<td>AAGCCTCGAGTTAAGCAAGCGAGCCAGTCC</td>
</tr>
<tr>
<td>1089</td>
<td>5’ sleC Δ122-155aa SOE</td>
<td>GCGTTCTTTCTCAAGCAAAAGTGTATTAG ATAAACCTGTAGTGCC</td>
</tr>
<tr>
<td>1090</td>
<td>3’ sleC Δ122-155aa rev oes</td>
<td>GGCACTACAGGGTTATCTAATACACTTTG TCTTGAAGGAAGACGC</td>
</tr>
</tbody>
</table>
4.1. Introduction

Like all spore-forming bacteria, *C. difficile* requires a signal to initiate germination. Spore forming bacteria germinate by sensing small molecule germinants, which vary with strain and species, using specific germinant receptors. Much of what is known about spore germination is from *Bacillus subtilis*, the model organism for spore forming bacteria, while other studies have been done with *C. perfringens*. In most cases, germination begins with a *ger* receptor sensing its cognate germinant, leading to activation of the cortex lytic enzymes[29, 56, 87]. For *C. difficile*, it has been known that bile salts induce germination; however, it was not known how *C. difficile* sensed these germinants, as it lacks the *ger* family of receptors that are used in other species[56, 85]. Additionally, PrkC has been shown to initiate germination of *B. subtilis* in response to cortex peptidoglycan fragments, and while the *C. difficile* genome contains a homologue of *prkC*, it seemed unlikely that the same receptor could respond specifically to such different substrates as peptidoglycan and bile salts[74, 151].

The germination process in *C. difficile* has remained relatively poorly understood in contrast with the processes in *C. perfringens* and *B. subtilis*. In the 1980s it was discovered that bile salts played an important role in *C. difficile* spore germination[49], and by 1987, evidence suggested that taurocholate was the preferred bile acid[73]. Still, it was not known if there were specific receptors for the germinants until 2010, when Ramirez and colleagues demonstrated that the kinetic profile of germination of *C. difficile*
spores strongly suggested the use of a receptor-mediated mechanism as opposed to physical membrane-permeabilization[74].

Joseph Sorg and colleagues published in 2008 that taurocholate and glycine were co-germinants[55], and discovered that chenodeoxycholate was a competitive inhibitor of the taurocholate germinant[50, 51, 152]. Having championed the effort to discover the germinant receptor for \textit{C. difficile}, in 2013 Sorg and coworkers identified the taurocholate-binding germinant receptor in \textit{C. difficile}, CspC, a subtilisin-like serine pseudoprotease whose function was previously unknown[47]. The Csp proteases are conserved in the Clostridia and function in germination. In \textit{C. perfringens}, it has been shown that CspB is necessary for the cleavage and activation of SleC, the germination hydrolase[70]. However, only one of the CspCs, CspB, has an intact catalytic triad in \textit{C. difficile}, and this is also found in other members of the Peptostreptococcaceae family to which \textit{C. difficile} belongs, such as \textit{C. sordellii} and \textit{C. bifermentans}.

CspB from \textit{C. perfringens}, as discussed in chapter two, is a subtilisin-like serine protease. The subtilisins are a large family of processive and nonspecific proteases, that require cleavage of the prodomain, which functions as an intramolecular chaperone, in order to gain activity[153]. In CspB \textit{perfringens}, the prodomain remained associated with the catalytic domain in our structure, which could be a way for CspB’s activity to be limited to germination[76]. CspA and CspC are also expected to adopt subtilisin folds, though in \textit{C. difficile} their catalytic triads are not intact and as such they would be unable to function as proteases[76]. Interestingly, a random mutagenesis screen carried out by Francis and colleagues revealed that germination-null mutants primarily had point mutations in CspC[47]. When these point mutations were mapped to our structure of CspB
(see Figure 4.5.2), we found that most cluster right around the active site and are in highly or strictly conserved residues. In most cases, the presence of the mutation would cause steric clashing with nearby chains of the protein and disrupt the fold (see Figure 4.5.2). Only on two occasions, with a stop codon generated at the C-terminus of the CspA domain of CspBA, and one generated in the middle of the CspB domain, did the inactivating point mutation fall outside of CspC. Given the importance of CspB protease function in SleC activation (Chapter 2), the identification of null mutations in CspBA was not as surprising as the mutations in CspC, which lacks catalytic function. Another interesting result from this group identified a gain-of-function mutant that germinated in response to chenodeoxycholate, which functions as an inhibitor for wild type C. difficile[47]. The single residue mutation that enabled the gain of function mapped to the dimer interface we observed in our CspB crystal structure (see Figure 4.5.6). As CspC is expected to have the same overall fold as CspB, it is possible that the CspB/CspB dimer we observed in the crystals could translate to a CspC/CspB dimer in vivo, perhaps explaining how CspC, sensing germinant, relays this signal to CspB, though we do not yet have any evidence for this idea.

The function of CspA during germination is still not known, but preliminary data from Joseph Sorg’s group, as mentioned above[47], and our own unpublished observations indicate that it is also required during germination. We sought to determine the structure of CspC to learn more about how it binds germinants and how it may act to signal downstream constituents such as CspB.
4.2. Materials and methods

4.2.1. C. difficile CspC protein expression

CspC from C. difficile was expressed in E. coli BL21(DE3) cells from a pET22b plasmid using the autoinduction method in Terrific Broth supplemented with 5052 sugar mix and ampicillin or carbenicillin with ~60 hour incubation at 20°C[150]. The CspC construct was codon-optimized for expression in E. coli and was fused C-terminally to the cysteine protease domain (CPD) from Vibrio cholerae MARTX toxin[154].

4.2.2. C. difficile CspC protein purification

Purification was carried out in three steps: immobilized metal affinity chromatography (IMAC), induced cleavage of CPD, and a Superdex 200 gel filtration column. IMAC was carried out with Ni-NTA agarose beads (5 Prime). The C-terminal 6-His tag on the CPD binds the nickel beads. Washes were carried out three times with low imidazole buffer (MLIB) (500mM NaCL, 10mM Tris-HCl pH 7.5, 10% (v/v) glycerol, 15 mM imidazole) to decrease non-specific binding to the beads.

The cysteine protease domain contains auto-cleavage activity in the presence of inositol hexakis phosphate (InsP₆), which is present in mammalian cells but not found in prokaryotes[154]. Therefore, InsP₆ was supplied to the culture, inducing cleavage between CspC and CPD-6-His and leaving behind a four amino acid VDAL linker C-terminal to CspC. InsP₆ was added to about 200µM and incubated with gentle shaking at 4°C for 4-6 hours or overnight. Without the addition of imidazole, the His-tag remained bound to the beads, and the CPD with it, leaving untagged CspC released to the supernatant. Supernatant was collected and bead washes were carried out three times using MLIB
(without supplemented InsP₆), and then the cleavage step was repeated, with more InsP₆ added, incubation, and bead washes in MLIB. Supernatants and washes, containing untagged CspC, were pooled and gently dialyzed into gel filtration buffer (150mM NaCl, 10mM Tris-HCl pH 7.5, 10% (v/v) glycerol) overnight. The buffer-exchanged protein was concentrated to 20 mg/mL or less, and gel filtration chromatography was carried out using a Superdex 200 (GE Healthcare) column, 500 µL injection volume and less than 10 mg of protein per run. A typical yield for *difficile* CspC purification was about 4 mg of purified protein per liter of autoinduction culture using the protocol described above[154]. Protein was concentrated to about 30mg/mL in gel filtration buffer (described above— shorthand 150/10/10) or a modified buffer (200mM NaCl, 10mM Tris-HCl pH 7.5, 5% (v/v) glycerol—shorthand 200/10/5), separated into small aliquots, flash cooled in liquid nitrogen and stored at -80 °C.

4.2.3. CspC *bifermentans* protein purification

The *cspC* gene from a related species, *C. bifermentans*, was codon-optimized for expression in *E. coli* and synthesized by Genscript. The codon-optimized *cspC* from *C. bifermentans* fragment was inserted N-terminal to the CPD sequence for expression and purification similar to CspC *C. difficile*. Purification of CspC *bifermentans* proceeded as described above, although the yields were much greater and often required dividing purification experiments up into several tubes at once to keep volumes and concentrations lower. A typical yield for CspC *bifermentans* was about 20 mg of purified protein per liter of autoinduction culture, about 5 fold greater than the yield of a typical CspC *difficile*
purification. For a figure showing the gels and traces obtained during a purification process, see Figure 4.5.4.

4.2.4. CspC crystallization trials

Purified and concentrated CspC from both *C. difficile* and *C. bifermentans* was set in preliminary crystallization screens at about 15 mg/mL to assess a wide variety of possible crystallization conditions. We used a variety of sparse-matrix crystal screens: Crystal Screen, Index, and Peg/Ion primarily (Hampton Research), and Wizard screens I and II (Emerald Biosystems). We explored several tray incubation temperatures, 12 °C, 18 °C, and 24 °C.

4.2.5. CspC *bifermentans* crystallization optimization

Crystals were discovered a mixture of ammonium acetate and polyethylene glycol. Optimization for initial crystal hits required adjusting the pH: with the combination of buffer in the protein solution and buffer in the crystallization solution, the final pH was about 6, and we found that a pH closer to 8 improved crystal growth. We achieved the higher pH by the addition of 1 M BIS-TRIS propane at pH 8.8 in addition to the sodium citrate pH 5.6. The ideal concentrations of these buffers seemed to be around 75 µM and 25 µM respectively, but any combination of the two with a 1:1 ratio or better for BIS-TRIS propane would reliably produced improved crystals.
4.2.6. CspC *bifermentans* crystal cryoprotection

0.5 µL (1 µL could also be used without consequence) of cryoprotectant solution (26% (w/v) PEG 3350, 18% (v/v) glycerol, 200 mM ammonium acetate, 60 mM BIS-TRIS propane, pH 8.8, 40 mM sodium citrate, pH 5.6) was added directly to a 2 µl drop containing crystals to be scooped and frozen, let to rest for at least one minute, and then a crystal scooped using a nylon loop and flash cooled in liquid nitrogen.

4.2.7. CspC *bifermentans* data collection

Data sets for CspC *bifermentans* were collected at 100K and \( \lambda = 0.979 \) Å at beamline 23ID-B the Advanced Photon Source synchrotron in Chicago, IL.

4.3. Results

4.3.1. Optimization of CspC production in *E. coli*

We found that using codon-optimized constructs helped with *E. coli* expressions, but the largest difference we found was the addition of the CPD. While it had been shown that CPD fusion proteins are sometimes more soluble than the target protein alone[154], for CspC it improved our yields (unpublished data). We found that during purification, CspC *difficile* tended to precipitate prior to gel filtration if it exceeded concentrations of \(~2\) mg/mL. The gel filtration protocol requires that small volumes of protein be loaded at the start of a run, and, due to low concentration, CspC *difficile* therefore often required multiple runs. This issue was not as prevalent with CspC *bifermentans*, which behaved
better during purifications, with better expression levels, a more complete cleavage of the CPD tag used during purification, and tolerant of higher concentrations prior to gel filtration, in the 15-20 mg/mL range.

CspC *C. bifermentans* was a good option for a related species to use for protein, because it, like *C. difficile*, should likely be reclassified as a *Peptoclostridium* sp.[75], see Figure 1.6.4. It contains the same catalytic-triad disruptions as *C. difficile* CspC and CspA, and preliminary data from our lab has also indicated that *cspC* from *C. bifermentans* can complement a *cspC*– mutant in *C. difficile* and induce germination in response to taurocholate.

4.3.2. **CspC *bifermentans* performs better in crystallization trials than CspC *difficile*.**

While we did not find any crystalline-like material using CspC from *C. difficile*, we found that CspC from *C. bifermentans* crystallized more readily, and were able to identify spherulites and crystal clusters forming in Crystal Screen HT from Hampton Research (See Figure 4.5.3 for PONDR disorder prediction for CspC *difficile*). The initial crystallization conditions for CspC *bifermentans* were a mixture of ammonium acetate and polyethylene glycol. Very small crystals nucleated readily, nucleating off the same point and overlapping each other. Crystals grew very slowly, taking three weeks to a month to grow large enough for X-ray data collection (see Figure 4.5.5 for initial crystal hits). There were frequently multiple crystals fused into a single crystal cluster. If there were crystals that appeared to be single crystals, they were smaller and more difficult to tease apart from
their multi-crystal counterparts. Most clusters were only ~30 microns prior to optimization efforts.

4.3.3. *CspC* bifermentans crystals nucleate readily and remain small.

We found that checking the trays frequently worsened nucleation. We therefore set most trays in duplicate- one to be checked regularly, one to rest undisturbed for at least the first week, after which point it was checked only infrequently. This strategy helped us to obtain fewer and slightly larger crystals per drop. The best crystals we obtained (Figure 4.5.5) grew after about 4 months of infrequent checking.

4.3.4. *CspC* bifermentans crystals require a gentle cryoprotection protocol.

Cryoprotection was challenging for these crystals. We found that the cryoprotecting solution had to be added directly to the drop containing the crystals as they could not be scooped from their mother liquor and put into another drop containing artificial mother liquor supplemented with a cryoprotectant. Glycerol was far superior to sucrose, glucose and PEG 400 as a cryoprotective agent for these crystals, but as most of the crystals were grown in the presence of very little glycerol (10% present in the storage buffer, diluted in the drop to 5%), they required a gentle introduction to the 15% (v/v) required for sufficient cryoprotection. Optimization of the cryoprotection protocol was an important step for obtaining the best possible data from the CspC crystals, though the crystals still contained multiple lattices.
4.3.5. Crystals contain multiple lattices, making data processing challenging.

The CspC crystals are small and require long exposures on our copper Kα X-ray generator (Rigaku HR300) equipped with a MAR345 image plate detector. Typical exposures were at least 20 minutes with 4 times 0.5° oscillations during this time. Data could be collected to about 3 Å resolution. The resulting data proved to be highly mosaic and frequently contained more than one lattice, making it difficult to process.

The data acquired at APS beamline 23ID-B extends to 2 Å resolution. The space group is most likely C-centered monoclinic with unit cell dimensions a = 169.72 Å, b = 163.58 Å, c = 102.84 Å. Unit cell angles are α= 86.89°, β= 113.69°, γ= 90.49°. We expect four monomers per asymmetric unit, with a Matthews coefficient of 2.6 and a solvent content of 53%. While the resolution is improved over that obtained using our home X-ray source, these crystals have multiple lattices, as observed earlier, making data processing problematic.

4.4. Discussion

Germination of C. difficile spores takes place in the large intestine[85]. However, taurocholate is generally cleaved to its substituents, taurine and cholate, by resident bacteria in the large bowel[46, 48]. These two products do not reliably induce germination either individually or in combination, though there is some debate on this point [51, 55, 74, 155]. Taurocholate would be present in higher concentrations in the large bowel of individuals with decreased numbers of resident bacteria, such as individuals who have
recently undergone antibiotic therapy[48, 156]. The presence of taurocholate in the large bowel, therefore, indicates a smaller competing population and a chance for *C. difficile* vegetative cells to find the nutrients they require.

Additionally, chenodeoxycholate is a known competitive inhibitor of taurocholate-induced germination[50, 51], and chenodeoxycholate follows an inverse concentration relationship to taurocholate: in healthy patients, taurocholate is present in the small bowel, but undetectable in the large bowel, having been hydrolyzed into taurine and cholate by resident bacteria[46, 48]. Chenodeoxycholate, conversely, is undetectable in the small bowel, instead appearing only as a moiety of other compounds such as glycochenodeoxycholate and taurochenodeoxycholate[156, 157]. As taurocholate is hydrolyzed in the large bowel so, too, is chenodeoxycholate cleaved from conjugated amino acids and can then function to inhibit germination[155].

Therefore, in the large bowel of a healthy individual with many commensal bacteria, there would be little to no taurocholate, thus the absence of a germination signal, and there would be ample chenodeoxycholate to inhibit germination. In an antibiotic treated individual, however, the relative amounts would be reversed and would drive germination rather than prevent it.

Ramirez and colleagues propose that dormant spores may only bind taurocholate, the limiting step, at which point subsequent taurocholate moieties would bind more readily via cooperativity, though there is no evidence that more than one taurocholate germinant may bind the same receptor[74]. Only when activated by the binding of taurocholate, would the spores become able to bind glycine, cooperatively until all binding sites are full, and begin the germination process[74].
A structure of CspC from *C. difficile* or the closely related *C. bifermentans*[75] would add invaluable information to the current understanding of how the germination process happens in *C. difficile* and other *Peptoclostridium* spp, and how it differs from the mechanism utilized in *C. perfringens* and other related *Clostridia*. A structure in complex with taurocholate, or chenodeoxycholate, and glycine would add the valuable understanding of exactly how the interaction between receptor and substrates takes place and would represent the first crystal structure of a germinant receptor bound to its ligand. Accordingly, we have made substantial progress in expressing, purifying, and crystallizing CspC and have obtained preliminary crystallographic data towards the structure determination of unliganded CspC from *C. bifermentans*. Further work is needed to improve the crystals that grow readily, and obtain larger, single crystals for data collection. There are multiple avenues to accomplish this, including the use of detergents or additives, use of micro- or macro-seeding, and varying incubation temperatures. The addition of a ligand or inhibitor could also influence crystal packing and thus improve the crystals.
4.5. Figures

Figure 4.5.1. Csp expression schematics

Schematics illustrating (A) the arrangement of Csp proteases in *C. difficile* and (B) the CspC-CPD fusion generated for purposes of expression and purification. The predicted prodomain of CspBA is indicated “Pro?” . Location of catalytic residues and cleavage sites are marked.
Figure 4.5.2. CspC germination-null mutants mapped to CspB structure illustrates clustering around active site.

Single amino-acid substitutions within CspC that rendered *Clostridium difficile* germination null, were aligned with CspB structure from *C. perfringens*. Almost all of the mutations were located near the active site in such a way that they would disrupt the fold due to steric clash (mutations shown as black sticks, active site residues of CspB shown as yellow sticks; right panel), suggesting that CspC retains the shape of a subtilisin serine protease for function, despite a lack of catalytic activity. Figure was rendered using PyMOL.
Figure 4.5.3. Predictions of Naturally Disordered Regions for CspC from *Clostridium difficile*

PONDR plot generated for full-length CspC from *C. difficile* illustrated more prediction of order than for SleC (Chapter 3), but containing several short segments predicted to be disordered.
Figure 4.5.4. CspC \textit{bifermentans} purification process.

CspC \textit{bifermentans} purification process. Top panel, NiNTA beads purification illustrating cleared lysates, flow through and washes of beads, addition of IP6 (IP61), second IP6 cleavage (IP62), and washes of IP6 cleavages, and finally the eluate showing uncleaved CspC-CPD still stuck to beads. Pooled fractions from IP6 cleavages and washes.
were concentrated and run on Superdex 200 gel filtration column, illustrating trace of pure CspC *bifermentans*, bottom panel.
Figure 4.5.5. CspC *bifermentans* crystal progression

CspC crystallization optimization process, showing progression of CspC crystals. All were grown at 18°C. Bottom right panel took four months to grow in a tray that was only checked infrequently.
CspB was found to dimerize in our crystals, despite evidence it functions as a monomer \textit{in vitro}. It is possible that the dimer interaction observed in the CspB crystals represents a dimerization of CspB with CspC in germinating spores. The proposed CspC taurocholate binding site as aligned with the CspB structure, highlighted in blue, and pointed to by the arrow, suggests a direct communication between CspC + germinant and CspB. One CspB monomer is shown in tan, and the other in pink.
CHAPTER 5: FUTURE DIRECTIONS

The germination cascade in *C. difficile*, based on our current understanding, begins with the germination receptor pseudoprotease CspC binding to co-germinants taurocholate and glycine, passing the activation signal on to germination protease CspB in a way that is not yet understood, resulting in CspB activation of SleC, the cortex hydrolase, which begins breaking down the cortex to allow for core hydration and expansion and eventually outgrowth to a vegetative, toxin-producing cell. This germination step is common and critical to each infection. This area of study is thus of paramount importance, especially given the persistence and tenacity of *C. difficile* infections. See Figure 5.1.1 for a schematic of the Csp/SleC cascade required to initiate germination, as we currently understand it.

For a more complete understanding of the germination process, several additional studies are necessary. A structure demonstrating the interaction of CspC with germinant or inhibitors bound would be illustrative of how this binding may be disrupted. The gain-of-function point mutation identified by Francis and colleagues which enabled CspC to respond to its inhibitor added some insight to this effort[47]. The specific mutation required an arginine in place of the glycine in position 457, and in our CspB structure this corresponded to a small loop located in the dimer interface. It is possible that this location is where taurocholate or chenodeoxycholate would bind, thus potentially identifying the method by which CspC and CspB communicate with each other.

We also need to understand exactly how CspC and CspB interact, and if this interaction is transient or involves binding and catalysis. Additionally, it is still unknown...
if CspC and CspB interact directly or if there are other proteins involved. The dimer interface observed in crystallized CspB *C. perfringens* calls to mind an exciting idea for a CspC/CspB dimer, but we do not know if it is relevant and indicative of observations to come in *C. difficile*, or simply a crystal packing coincidence. How does CspC activate CspB? We showed in Chapter 2 that CspB from *C. perfringens* had cleaved but not dissociated from its inhibitory prodomain. It is possible that CspB in *C. difficile* could exist in this same inhibited state in spores until activation by CspC, which could trigger protease activity against the prodomain, thus exposing the competent active site.

Similarly, can CspB and SleC be observed binding together? SleC is activated by CspB’s cleavage of its inhibitory propeptide, but to see a structure of the two in complex would answer so many questions about how this process takes place and help us to continue to expand our understanding of the germination process. A SleC mutant that is unable to be cleaved (or a CspB mutant that is unable to hydrolyze SleC) would be a good place to start to look for these interaction, using pull downs, western blots, dynamic light scattering experiments and gel-shift assays, with the addition of a cross-linking compound if necessary to try to observe binding between both CspB and SleC, and CspC and CspB.

An antibody to observe the location of CspA within both sporulating cells and the mature spore would be helpful in understanding the role it plays in germination. Similarly, pull-down and interaction studies could be carried out using CspA, CspB, and CspC as perhaps CspA helps mediate this interaction by acting as a scaffold, and stabilizing the complex. We now have evidence that CspA has an important role in germination[47], as suggested by its evolutionary conservation, but we do not yet know what that role might be.
Crystal structures of all complexes would be extremely helpful in furthering this field of study. As crystals of protein complexes are often difficult to obtain, several truncation constructs could be made of target proteins to observe fragment-binding if expression and purification of full-length proteins proves difficult. Design of these fragments would be greatly aided by the structures of each protein, so this is of top importance. The use of nanobodies to stabilize crystallization-resistant proteins could be employed, as this method has proved beneficial in some cases[158, 159].

Understanding the molecular basis underlying spore germination in \textit{C. difficile} could lead to novel treatments that could block the cycle of infection and reinfection and serve to greatly aid in our treatments of this disease. A particularly challenging aspect with \textit{C. difficile} infections relates to its ability to cause recurring infections that increase in severity, and current treatments leave a patient more susceptible to reinfection by way of a disrupted microbiota, and because the patients typically need to be hospitalized during the treatment for the \textit{C. difficile} infection, thus exposing them to more spores.

There are two ways to interfere with \textit{C. difficile} spread and infection from a germination perspective. The first way is to block germination, despite the presence of germinants and a suitable environment[155]. This method is somewhat complicated, requiring either infection by a mutated form of \textit{C. difficile} or taking by mouth doses of a compound that could interfere with the germinant taurocholate or the inhibitor, chenodeoxycholate in such a way that favorable germination conditions would go undetected by the spores. The second way is to force germination in the absence of the suitable environment. Vegetative cells are much more susceptible to cleaning agents and disinfection protocols than are spores, as well as being unable to live in the presence of
oxygen if one can force outgrowth to vegetative cells in a hospital room, the cells will quickly die. This has been done with some success recently[160], and is simpler than attempting to interfere with germination in the colon, as *C. difficile* and the digestive process have co-evolved to work together.
Proposed model for activation of germination in *Clostridium difficile* spores. CspC (pink rectangle) within dormant spore binds to the germinant (yellow circle). CspC relays this signal to CspB (blue rectangle), which becomes active upon receiving signal from upstream CspC. Active CspB then cleaves pro-SleC (orange circle) into active SleC (orange pac-man). SleC hydrolyzes the peptidoglycan cortex of the spore, allowing hydration and expansion of the core.


