# Peptidoglycan remodelling during *Clostridioides difficile* sporulation.

Abigail Kelly


Thesis submitted in partial fulfilment of the requirements of the regulations for the degree of Doctor of Philosophy

September 2019

Institute for Cell and Molecular Bioscience

# Abstract

*Clostridioides difficile* is a Gram-positive enteric human pathogen capable of colonising the gut, and causing *C. difficile* infection (CDI), which can be life threatening. With a shifting CDI demographic, increasing virulence and high re-infection rates, an increased understanding of basic *C. difficile* physiology is required to identify new intervention targets.

As an anaerobic pathogen, *C. difficile* must be transmitted through the aerobic environment as specially adapted cells called endospores. The process of transitioning from vegetative cells, that are active in the gut, to the metabolically inactive endospores is known as sporulation and requires considerable remodelling of the mother cell. Firstly, the mother cell produces an asymmetric septum, designating a portion of its cytoplasm as the forespore, which will eventually become the mature endospore. The mother cell then engulfs the forespore to produce a cell-within-a-cell structure. The forespore matures before lysis of the mother cell releases the endospore into the environment. The sporulation process involves considerable changes in all cell components, including remodelling of the cell wall peptidoglycan. This work focuses on the remodelling of peptidoglycan throughout sporulation, with a particular focus on the engulfment mechanisms.

*C. difficile* peptidoglycan was characterised at various stages of engulfment. Surprisingly, the muropeptides identified in vegetative cells were detected throughout early stages of engulfment. Importantly, the proportions of each muropeptide differ throughout the process, indicating varying organisation of the peptidoglycan even at this early stage of sporulation.

In parallel, this work also aimed to identify and characterise peptidoglycan modifying enzymes involved in sporulation. SpoIID and SpoIIP are peptidoglycan hydrolases that have been implicated in *Bacillus subtilis* sporulation. These proteins were purified, characterised and their activities tested on various peptidoglycan types. SpoIIP is a bifunctional amidase and endopeptidase that produces the substrates for the lytic transglycosylase, SpoIID. SpoIID activity was shown to be impacted by zinc binding, and the catalytic residue was identified.

*C. difficile* peptidoglycan is unusual in that it is predominantly deacetylated. In order to further probe peptidoglycan remodelling during sporulation, polysaccharide deacetylase genes in *C. difficile* were investigated. Ten putative genes were identified and 2 were cloned and recombinantly expressed. Both putative deacetylases were characterised and were shown to be active on acetylated peptidoglycan, implicating them in peptidoglycan remodelling.

This work furthers current understanding of *C. difficile* peptidoglycan biology during sporulation, both in providing a detailed analysis of cell wall composition, and characterising key engulfasome enzymes. It further opens new research avenues by identifying potential peptidoglycan deacetylases in *C. difficile*. Together, this provides a deeper understanding of key mechanisms involved in the production of the infective agents in CDI that could be explored as novel therapeutic targets.

# Acknowledgments

Firstly, I would like to thank my supervisors Dr Paula Salgado and Professor Waldemar Vollmer, for the opportunity, mentorship and freedom to play around in the lab during my PhD. Thanks to Dr Joe Gray for all his help with HPLC and Mass Spectrometry, especially for explaining the same thing several times over.

Thanks to all past and present members of the Salgado and Waldron labs, especially to Dr Marcin Dembek for helping me to hit the ground running, and to Dr Anna Barwinska-Sendra for keeping me from falling over too often and helping with protein purifications and CD experiments especially. You've both helped me with bits of experiments and general lab-knowledge more times than I can count, and I really appreciate it. Thanks to everybody in the CBCB and in SBL who have made me welcome and my time as a PhD student more enjoyable for it.

To Anna and Louisa, thank you both for the regular group therapy sessions, long may they continue! Thanks to all the "Durham lot", without which I never would have made it out of undergrad in one piece. I can't thank Becky enough, not just for all the encouragement, advice and saving me from various killer arachnids, but for the unwavering friendship over the past eight years, even when I might have been slightly overreacting…

The biggest of thanks go to Mum, Dad and Hannah. All three of you have encouraged me from before I even applied for the PhD and have listened to me whinge regularly since. Thanks for your patience with me and the unconditional love. Thanks too to Grandad for the world's greatest piece of advice: "Try to think.". Last but not least, thank you Doug; for keeping me laughing, the constant support, and for reminding me that *C. difficile* peptidoglycan isn't the most important thing in the world.

I never would have been able to do this without you all.

# Table of contents

# List of figures

# List of tables

# Nomenclature and abbreviations

| | |
|---|---|
| AA | Amino acid |
| AIM | Auto Induction media |
| AMU | Atomic mass units |
| Amp | Ampicillin |
| ATc | Anhydrotetracycline |
| AUC | Area under curve |
| BHI | Brain heart infusion |
| BHIS | Brain heart infusion- supplemented |
| bp | Base pairs |
| BSA | Bovine serum albumin |
| CC | Correlation coefficient |
| CD | Circular dichroism |
| CDI | *Clostridioides difficile* infection |
| Cm | Chloramphenicol |
| CV | Column volumes |
| Da | Daltons |
| DMP | SpoIID SpoIIM SpoIIP complex |
| DMSO | Dimethylsulfoxide |
| dNTPs | Deoxyribonucleotide triphosphate |
| DUI | Dials user interface |
| EDTA | Ethylenediaminetetraacetic acid |
| GABA | γ-amino-butyric acid |
| gDNA | Genomic deoxyribonucleic acid |
| GFP | Green fluorescent protein |
| GlcN | Glucosamine |
| Glc*N*Ac | *N*-acetylglucosamine |
| h | Hours |
| HEPES | 4-(2-Hydrocyethyl)piperazine-1-ethanesulfonic acid, N-(2-Hydroxylexthyl)piperazine-N'-(2-ethanesulfonic acid) |
| HPLC | High pressure liquid chromatography |
| HRP | Horse radish peroxidase |
| ICP-MS | Inductively coupled plasma mass spectrometry |
| IMAC | Immobilised metal affinity chromatography |
| IPTG | Isopropyl-β-D-thiogalactoside |
| Km | Kanamycin |
| kDa | KiloDalton |
| kV | KiloVolts |
| LB | Luria broth |
| LC-MS | Liquid chromatography mass spectrometry |
| LTQ | Linear trap quadrupole |
| LTQ-FT | Linear trap quardrupole Fourrier Transform |
| Mdeg | Millidegrees |
| MRE | Mean residue elipticity |
| MW | Molecular weight |
| MS | Mass spectrometry |
| MS2 | Mass spectrum 2 |

| | |
|---|---|
| MS3 | Mass spectrum 3 |
| Mur*N*Ac | N-Acetylmuramic acid |
| MWCO | Molecular weight cut off |
| *m/z* | Mass:charge ratio |
| NCDAA | Non-canonical D-amino acids |
| NRMSD | Normalised root mean squared deviation |
| $OD_{600}$ | Optical density at 600nm |
| PCR | Polymerase chain reaction |
| PDB | Protein database |
| RFP | Red fluorescent protein |
| PG | Peptidoglycan |
| ppm | Parts per million |
| PVDF | Polyvinylidene fluoride |
| RMSD | Root mean square deviation |
| RP-HPLC | Reverse phase high pressure liquid chromatography |
| rpm | Revolutions per minute |
| RT | Room temperature |
| S200 | Superdex 200 |
| SDS | Sodium dodecyl sulphate |
| SDS-PAGE | Sodium dodecyl sulphate polyacrylamide gel electrophoresis |
| SEC | Size exclusion chromatography |
| SM | Sporulation media |
| TBS | Tris-buffered saline |
| TBST | Tris-buffered saline with Tween |
| TEMED | N, N, N', N'-tetramethylethane-1,2-diamine |
| TEV | Tobacco Edge Virus protease |
| TLC | Thin layer chromatography |
| $T_m$ | Melting temperature |
| UV | Ultraviolet |
| v/v | Volume/volume |
| WT | Wild type |
| W/V | Weight/ volume |
| W/W | Weight/ Weight |
| *g* | gravitational constant |

# Chapter 1. Introduction

## *1.1. Clostridioides difficile*

*Clostridium difficile*, recently renamed *Clostridioides difficile* (Lawson *et al.*, 2016; Oren and Rupnik, 2018), is a Gram-positive, anaerobic and opportunistic human pathogen. When the normal gut flora is disturbed, *C. difficile* proliferates in the bowel leading to inflammation and toxin release, resulting in the clinical presentation of *C. difficile* infection (CDI). In 2018, more than ten thousand CDI cases were reported in England alone, with almost 5,700 infections reported in the first 6 months of 2019 (Public Health England, 2019).

CDI can range from mild to severe diarrhoea and pseudomembranous colitis, to the potentially fatal toxic mega colon (Sayedy, Kothari and Richards, 2010), with systemic complications such as multiple organ failure associated with the most severe cases (Di Bella *et al.*, 2016). After an initial case of CDI, a patient is 15-30% more likely to have a second, recurrent infection (Tang-Feldman *et al.*, 2003), either with the same or a different strain of *C. difficile*. Following resolution of this first recurrence, a second recurrence will occur in 40% of patients, with further recurrences reported in 45-65 % of patients (Song and Kim, 2019). The financial cost of an initial CDI is estimated at $5,200-8,600 and for recurrent CDI is estimated to be ~$14,000 per recurrence (Ghantoji *et al.*, 2010).

Typically, CDI is encountered in elderly patients, those treated with long-term broad-spectrum antibiotics, those exposed to a healthcare setting, or the otherwise immunocompromised (Eze *et al.*, 2017). Disruption of the normal gut flora, normally due to broad spectrum antibiotic treatment, results in an empty niche for *C. difficile* to proliferate into, and consequently cause CDI. The role of the normal microbiota is becoming increasingly implicated in CDI; in terms of preventing germination of spores, from a general protective stance, and in terms of providing colonisation resistance, thus preventing *C. difficile* expansion (Bibbò *et al.*, 2014; Di Bella *et al.*, 2016; Ducarmon *et al.*, 2019; Leslie *et al.*, 2019). The roles of the normal flora and colonisation resistance in CDI are extensively reviewed in Ducarmon *et al* (2019).

Crucially, the demographic of CDI is shifting; younger and antibiotic naïve patients are presenting with CDI in the community (Garg *et al.*, 2013; Bloomfield and Riley, 2016); approximately 30% of CDI cases are not associated with canonical risk factors (Smits *et al.*, 2016). Furthermore, links between CDI and proton pump inhibitor usage (Janarthanan *et al.*, 2012) and irritable bowel syndrome (Reddy and Brandt, 2013) have recently been uncovered suggesting a larger proportion of the population is at risk of CDI.

The clinical symptoms of CDI have primarily been related to the activity of the toxins released by *C. difficile*, though toxin titre does not correlate with disease severity (Akerlund *et al.*, 2006). Three different *C. difficile* toxins have been described; TcdA, TcdB and CDT, though not all strains of *C. difficile* will produce all toxins (Sun *et al.*, 2010; Carman *et al.*, 2011; Eckert *et al.*, 2015; Di Bella *et al.*, 2016; Salazar *et al.*, 2017). TcdA and TcdB are the most thoroughly understood, and thought to contribute significantly to the aetiology of CDI (Di Bella *et al.*, 2016). TcdA/B inactivate Rho GTPases, a set of proteins integral to several host cell processes, with glucosylation of these Rho GTPases resulting in cytoskeletal breakdown, which in turn leads to loss of tight and adherent junctions between enterocytes

and increased epithelium permeability (Di Bella *et al.*, 2016). Furthermore, the loss of cell contacts leads to apoptosis and necrosis of tissue, activating the inflammasome. These changes thus lead to loss of intestinal barrier function and neutrophilic colitis (Rupnik, Wilcox and Gerding, 2009; Di Bella *et al.*, 2016). CDT similarly leads to loss of the actin cytoskeleton integrity in the host via actin ADP- ribosylation and may also be involved in *C. difficile* adherence; CDT is implicated in reorganisation of the host epithelial cell's microtubule cytoskeleton, leading to the projection of long cell extensions which wrap around bacterial cells (Schwan *et al.*, 2009, 2014).

In addition to the cytotoxic effects of the toxins, the immune response to CDI is becoming more heavily implicated in the clinical presentation of disease, with the intensity of the host immune response correlating with disease severity (Ng *et al.*, 2010; Cowardin *et al.*, 2015).

Treatments for CDI are relatively limited; only three antibiotics are currently recommended for use in the UK (vancomycin, metronidazole and fidaxomicin) (Public Health England, 2013), with resistance to vancomycin and metronidazole increasing (Peng *et al.*, 2017), and evidence of plasmid-mediated metronidazole resistance reported (Boekhoud *et al.*, 2019). Use of these drugs also exacerbates gut dysbiosis which can further increase the severity of disease and lead to recurrent infections. It is noteworthy that the CDI-associated antibiotics clindamycin and ciprofloxacin have recently been implicated in CDI in a microbiome-independent manner; administration of the antibiotics to cell culture resulted in reduced mucosal barrier function (Kester *et al.*, 2019).Consequently, several new avenues of intervention are being investigated to reduce CDI prevalence.

Faecal microbiota transplant (FMT), seeks to restore the gut flora of the patient and is reported to have a success in excess of 90% (Cammarota, Ianiro and Gasbarrini, 2014; Sbahi and Di Palma, 2016; Quraishi *et al.*, 2017; Ianiro *et al.*, 2018). This option is generally only used after several recurrent CDI episodes that do not respond to antibiotic regimens. Treatment with non-toxigenic *C. difficile* may also interrupt the infection-relapse cycle. In hamster models, treatment with spores of a non-toxigenic *C. difficile* strain protected the hamsters from subsequent CDI (Sambol *et al.*, 2002). Consequently, the possibility of a similar intervention is currently under investigation in human subjects, with a clinical trial suggesting reduced recurrence in treated patients (Gerding *et al.*, 2015). Several vaccines against *C. difficile* toxins have been produced and are under development (Bézay *et al.*, 2016; de Bruyn *et al.*, 2016; Sheldon *et al.*, 2016; Kitchin *et al.*, 2019), though these typically neutralise TcdA and/or TcdB rather than prevent colonisation. Bezlotoxumab is a monoclonal antibody therapy which is licensed for the prevention of CDI recurrence in high risk patients (Deeks, 2017).

As the demographic shift seen in CDI is combined with an aging population, increasing resistance and virulence, and the financial burden of CDI, new therapies and interventions are urgently required. To facilitate the discovery of new treatments and interventions more must be understood about the fundamental biology of *C. difficile*.

## 1.2. The life cycle of *C. difficile*

As an anaerobic organism, *C. difficile* must undergo drastic changes to its fundamental biology to persist in the aerobic environment, consequently allowing transmission between hosts. To this end, *C. difficile* produces endospores, henceforth referred to as spores (Figure 1-1). Indeed, *C. difficile* strains that cannot form spores are unable to persist or effectively transmit disease in a mouse model of infection (Deakin *et al.*, 2012). These metabolically dormant spores are released into the aerobic environment, where they can persist for extended periods of time, until favourable conditions are encountered.

During a CDI a patient will shed spores during, and after, the course of infection (Barra-Carrasco and Paredes-Sabja, 2014). These spores persist in the immediate environment, resistant to most hospital disinfection procedures (reviewed in Barra-Carrasco & Paredes-Sabja, 2014) and can also be transferred through the environment to multiple potential hosts (Barra-Carrasco and Paredes-Sabja, 2014). Spores are carried on the shoes of hospital workers (Janezic, Mlakar and Rupnik, 2018) and can survive disinfection of hospital gowns and surfaces (Dyer *et al.*, 2019), for example.

Once the spores encounter a favourable environment, for example if they are ingested by a susceptible individual, they germinate to produce vegetative cells. Spore germination is initiated upon the binding of germinants, such as bile salts and amino acids (Setlow, 2003; Setlow, Wang and Li, 2017), to, as yet not fully identified, germinant receptors on the spore (Kochan *et al.*, 2018; Rohlfing *et al.*, 2019). During this irreversible process, the spore rehydrates, the cortex is degraded, and a vegetative cell is eventually produced (Vincent and Manges, 2015). The primary bile salts cholate and chenodeoxycholate can be conjugated to taurine, to give taurocholate or taurochenodeoxycholate, or to glycine to produce glycocholate or glycochenodeoxycholate. These primary bile salts, found in the large intestine, can be metabolised by normal flora, producing the secondary bile salts deoxycholate and lithocholate respectively. Both chenodeoycholate and lithocholate prevent spore germination, with chenodeoycholate also preventing the vegetative growth of *C. difficile*. Taurocholate, cholate and deoxycholate promote spore germination (Sorg and Sonenshein, 2009, 2010; Vincent and Manges, 2015). When the normal flora is disturbed there is an increase in the ratio of cholate to chenodeoxycholate derivatives, leading to a relative increase in the pro-germination bile salts, thus promoting *C. difficile* outgrowth (Sorg and Sonenshein, 2009, 2010; Vincent and Manges, 2015). The resident microbiota of the host may have a further role to play; for example, sialic acid release from host glycans by *Bacteroides thetaiotaomicron*, a resident microbiota model organism, has been implicated in facilitating the expansion of *C. difficile* in the gut (Ng et al., 2013). Synthesis of secondary bile acids by *Clostridium scindens* (Kang *et al.*, 2008) has also been implicated in increased resistance to *C. difficile*; when *C. scindens* was introduced to a murine CDI model disease was ameloiorated (Buffie *et al.*, 2015).

Upon germination, vegetative cells proliferate and release toxins, form biofilms, or go on to produce more spores (Figure 1-1). It has been suggested that sporulating cells continue to produce toxins; Ransom *et al*. (2018) reported red fluorescence from *tcdA-rfp* (red fluorescent protein) fusion in cells that had formed phase bright, and in some cases mature, spores, though the authors acknowledge this could be residual fluorescence, where a cell

has produced toxin and then a spore (Ransom *et al.*, 2018). It is assumed some of the spores produced may remain in the host, increasing the likelihood of reinfection and persistence.

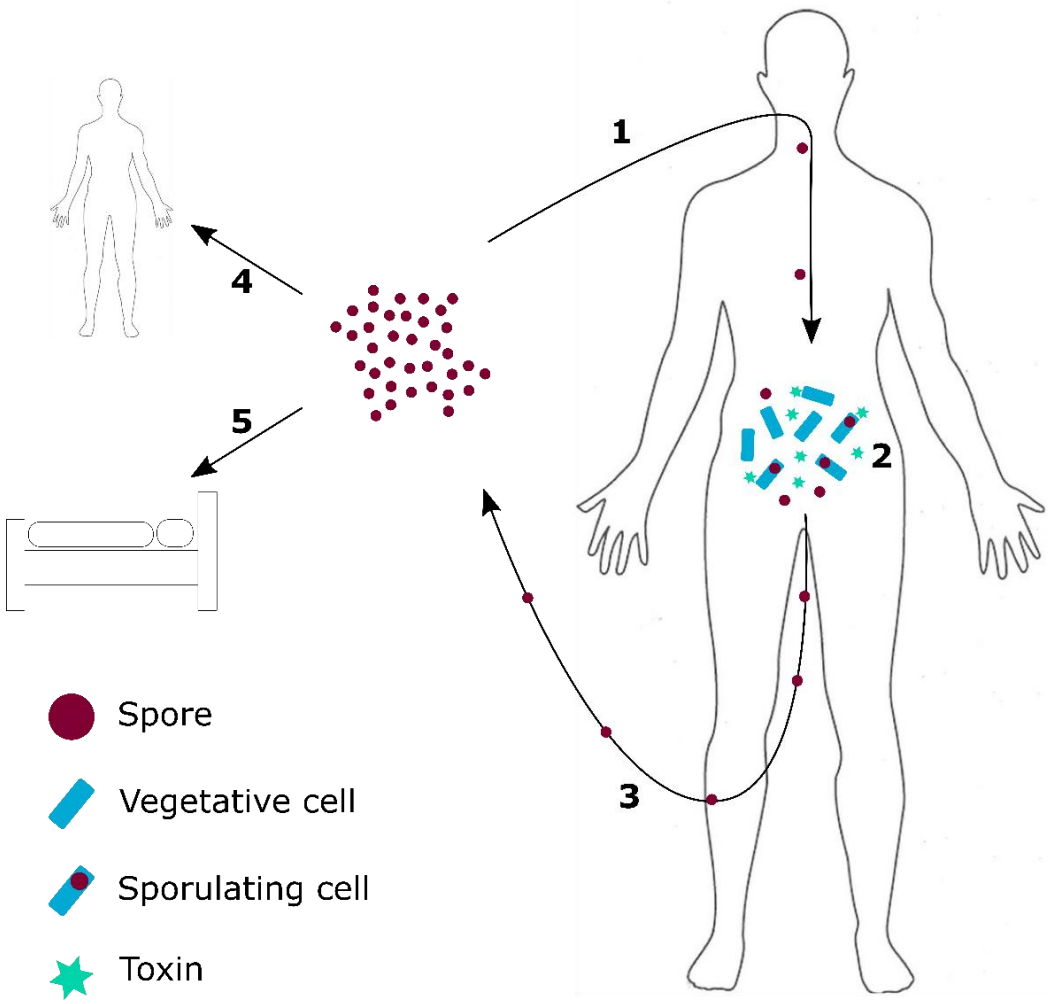


Figure 1-1: Infection cycle of *C. difficile*

*C. difficile* spores (purple circles) exist in the environment and are ingested (1). (2) In a susceptible individual, the spores then germinate, producing vegetative cells (blue rectangles) that are capable of replicating in the bowel, producing toxin, biofilms or spores. Vegetative cells release toxins (green stars) that significantly contribute to the symptoms of CDI. A subset of cells will undergo sporulation (blue and purple cells). The spores are then released into the environment (3), where they may go on to colonise/ infect other individuals (4), persist on hospital equipment such as beds (5), or re-infect the same host.

### 1.2.1. Spores as the transmissible agent

Spores consist of several layers (Figure 1-2) that protect the metabolically inactive core where genetic information is stored.

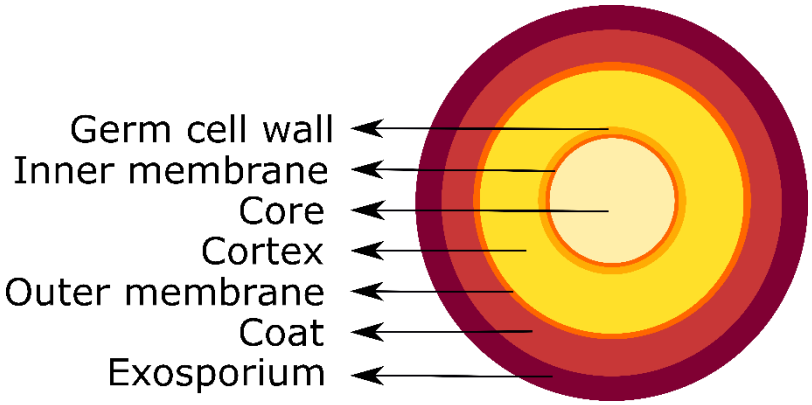


Figure 1-2: Spore architecture

*C. difficile* produces spores as the transmissible agent of CDI. Each layer is discussed below. Image adapted from Paredes-Sabja et al., 2014.

The spore core contains the genetic material transferred from the mother cell. The DNA is protected by small acid-soluble proteins, calcium-chelated dipicolonic acid and a reduced water content (Setlow, 2007).

Surrounding the spore core is an inner membrane, impenetrable to small molecules, including water (Paredes-Sabja, Shen and Sorg, 2014). A thick layer of highly modified peptidoglycan surrounds the germ cell wall and is discussed in detail in section 1.5.3. The germ cell wall is a layer of peptidoglycan derived from the mother cell, distinct from the cortex peptidoglycan. Whilst the germ cell wall peptidoglycan will become part of the cell wall of the germinating cell, the cortex peptidoglycan is specifically digested by cortex lytic enzymes during germination (Paredes-Sabja, Shen and Sorg, 2014).

The spore coat is a proteinaceous layer that is related to resistance to several disinfection methods including chemical and heat inactivation and irradiation (Setlow, 2006). Much of our understanding of coat proteins comes from *B. subtilis* (Eichenberger, Fawcett and Losick, 2002; Ramamurthi, Clapham and Losick, 2006; Wang *et al.*, 2009; Pereira *et al.*, 2018), and whilst many coat proteins are conserved, *C. difficile* has some notable deviations from the model organism. For example, SpoIVA is required for both *B. subtilis* and *C. difficile* spore formation, whereas SpoVM, which is largely dispensable in *C. difficile,* is essential in *B. subtilis* (Putnam *et al.*, 2013; Ribis *et al.*, 2017). Furthermore, SpoIVD, which recruits various proteins to the *B. subtilis* basement layer and is crucial to enabling SpoIVA to encase the maturing spore, is key to *B. subtilis* spore formation, but there is no obvious homologue in *C. difficile* (Wang *et al.*, 2009; Putnam *et al.*, 2013). Furthermore, the coat is increasingly being implicated in roles beyond resistance to external stresses. The coat protein CotE is implicated in binding to mucin, which in turn is related to increased colonisation and more severe symptoms in a mouse model of disease. This suggests CotE, and the coat in general, may have a more direct role in pathogenesis (Hong *et al.*, 2017).

The outermost layer, the exosporium, is a proteinaceous layer but its role is not thoroughly understood (Barra-Carrasco and Paredes-Sabja, 2014). Some studies suggest that it might be involved in spore adherence (Mora-Uribe *et al.*, 2016) and bacterial fitness and persistence (Calderón-Romero *et al.*, 2018).

Together, these adaptations protect the DNA from multiple stressors, making spores incredibly resilient. For example, it is thought that spores have the capability to survive the harsh environment of space and may even be able to survive re-entry into the atmosphere, thus leading to the hypothesis that life on earth could have originated from spores (Nicholson *et al.*, 2000; Horneck *et al.*, 2012).

## 1.3. Sporulation

The process of forming spores, known as sporulation, is schematised in Figure 1-3. After duplicating the chromosome, the vegetative cell forms an asymmetric septum, thus designating the smaller portion of the cytoplasm as the forespore, and the larger section of the cytoplasm as the mother cell. In *B. subtilis*, a portion of the copied chromosome is trapped in the forespore compartment, before being pumped into the forespore by SpoIIIE as soon as septation is completed (Lopez-Garrido *et al.*, 2018). After a copy of the chromosome is transferred, the mother cell surrounds the forespore to form a cell-within-a-cell structure in a process known as engulfment (Figure 1-3). The forespore is then matured by the formation of the different spore layers, before the mother cell lyses, releasing the mature spore into the environment. The spore then persists until favourable conditions are encountered, upon which the spore germinates into a vegetative cell, ready to produce toxins, biofilms or spores once again.

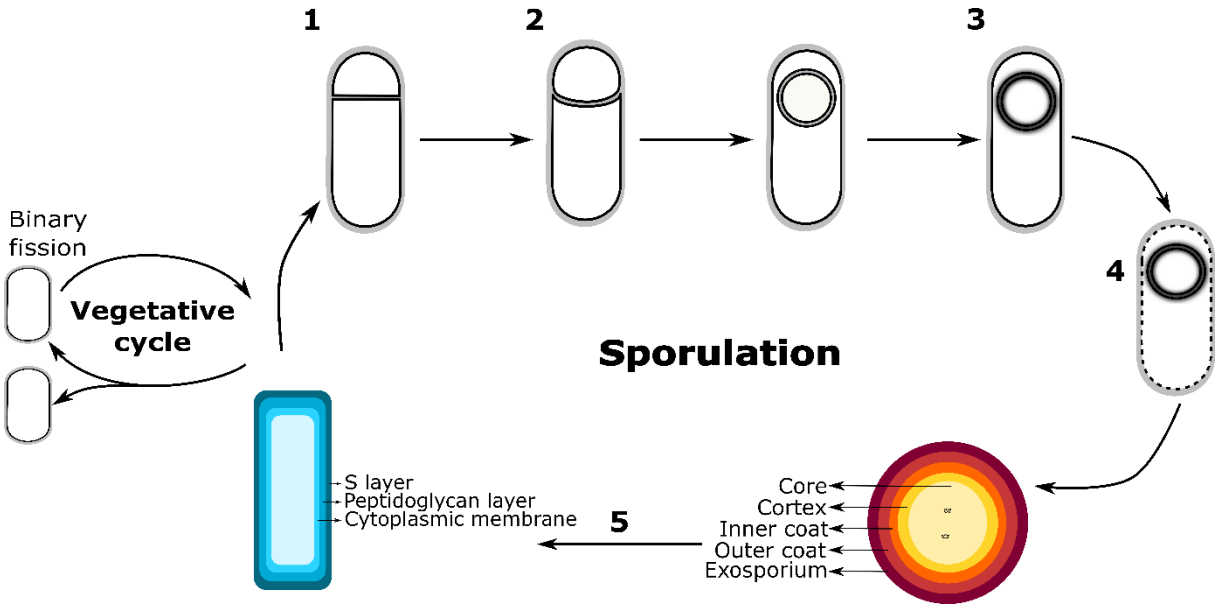


Figure 1-3: Sporulation cycle of *C. difficile*.

A proportion of the vegetative *C. difficile* cells in the bowel undergoes sporulation. Firstly, an asymmetric septum is formed (1) which then curves (2). The mother cell then engulfs the smaller compartment until a cell within-a-cell structure is formed (3). The forespore is matured, then the mother cell lyses releasing the mature spore into the environment (4). The spore persists until favourable conditions are encountered, at which point the spore germinates (5), forming a new vegetative cell, ready to begin the cycle once again.

### 1.3.1. Regulation of sporulation

Sporulation has been most extensively studied in *B. subtilis*, though more studies are being conducted in *C. difficile* as fundamental differences between the two bacteria are uncovered. The control of sporulation is one area with clear differences; whilst *B. subtilis* and *C. difficile* share many components of the sporulation machinery, they are often employed differently, as demonstrated in Figure 1-4.

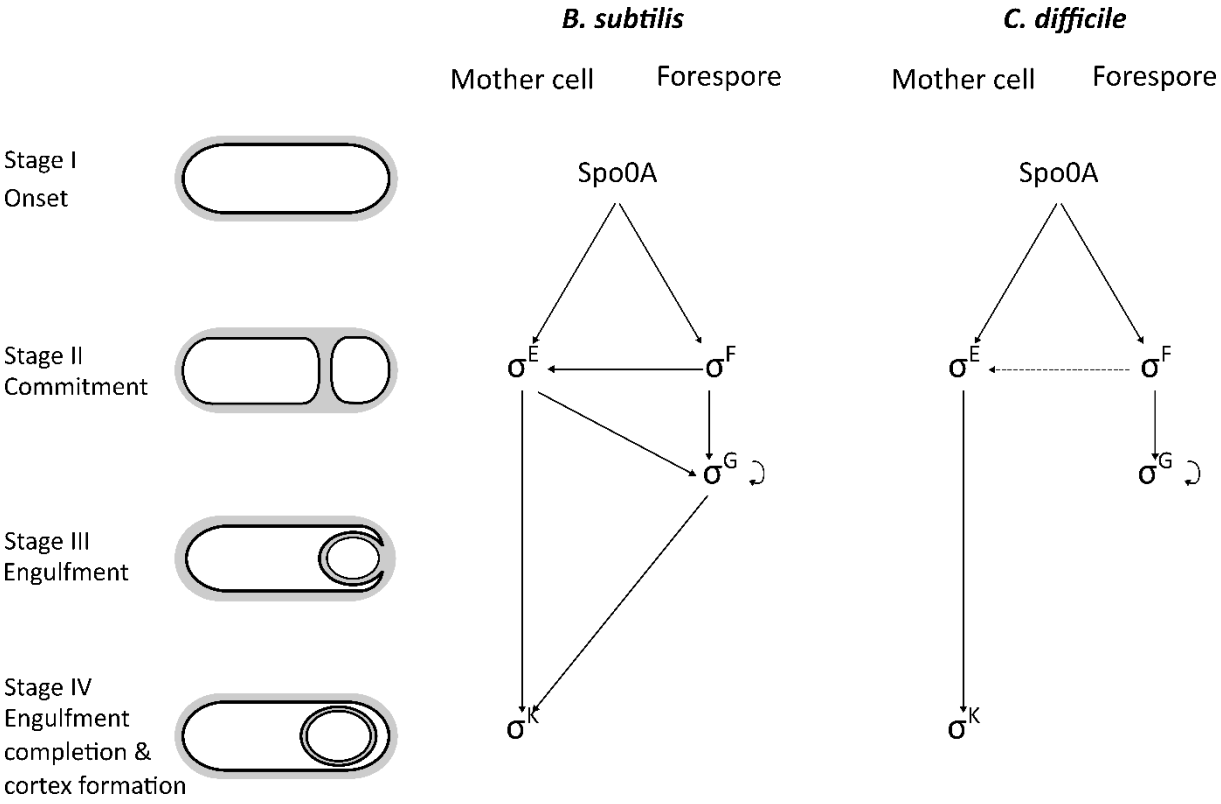


Figure 1-4: Regulation of sporulation

*B. subtilis* and *C. difficile* share many of the components of the control of sporulation pathways, although they interact differently and have different roles. Spo0A is the master regulator in both organisms. Key differences include; $\sigma^E$ is not required for $\sigma^G$ activation and $\sigma^G$ is not required for $\sigma^K$ activation in *C. difficile* as it is in *B. subtilis* (Zhu, Sorg and Sun, 2018). Dashed arrow relates to the dispensable regulation of $\sigma^E$ by $\sigma^F$ in *C. difficile*. Direction of the arrow indicates direction of regulation.

The master regulator of sporulation, in both *B. subtilis* and *C. difficile*, is the DNA binding protein Spo0A (Molle *et al.*, 2003; Deakin *et al.*, 2012; Rosenbusch *et al.*, 2012) (Figure 1-4). Disruption of *spo0A* in *C. difficile* leads to an inability to produce spores and an inability to persist or transfer between individuals in mouse models (Deakin *et al.*, 2012). Spo0A is also implicated in toxin production and regulation (Deakin *et al.*, 2012; Mackin *et al.*, 2013) and biofilm production (Dapa *et al.*, 2013). Indeed, in ribotype 027 strains, Spo0A may negatively regulate toxin production, as deletion of the gene leads to increased TcdA and TcdB levels in a mouse model of infection (Deakin *et al.*, 2012). The regulation of toxin production by Spo0A may be strain specific however, with conflicting results on the toxin production of 630 *Δerm Δspo0A* strains (Underwood *et al.*, 2009; Deakin *et al.*, 2012; Rosenbusch *et al.*, 2012).

In *B. subtilis*, Spo0A binding in the mother cell results in the activation of the RNA transcription factor $\sigma^F$ in the forespore, which in turn results in $\sigma^E$ activity in the mother cell (Hoch, 1993). This $\sigma^E$ activity causes $\sigma^G$ activation in late sporulation, which ultimately results in $\sigma^K$ activity in the mother cell (Fimlaid and Shen, 2015).

In contrast, there is much less intercompartmental communication between the sigma factors in *C. difficile* (Figure 1-4). In the mother cell, Spo0A activity leads to $\sigma^E$ activation, which in turn leads to $\sigma^K$ activation during late sporulation, potentially mediated via SpoIIID (Saujet *et al.*, 2014). In the forespore compartment, Spo0A activation leads to $\sigma^F$ activation, which in turn leads to $\sigma^G$ activation (Saujet *et al.*, 2014). $\sigma^F$ in the forespore contributes to $\sigma^E$ activation in the mother cell, but this interaction is dispensable in *C. difficile* (Fimlaid and Shen, 2015). There may also be some degree of auto-activation of $\sigma^G$ (Fimlaid and Shen, 2015).

The activity of these specific sigma factors at specific times allows the coordinated and controlled transcription and translation of various proteins thought to be involved in sporulation.

## 1.4. Peptidoglycan

Peptidoglycan is a macromolecule found encompassing the cell membrane of almost all bacteria, and is key in maintaining cell shape and viability (Vollmer, Blanot and de Pedro, 2008).

In order to analyse the composition of the peptidoglycan of a given species, the peptidoglycan is first digested into its building blocks, muropeptides, by a muramidase. Muropeptides are then separated by reverse phase high pressure liquid chromatography (HPLC) (Glauner, 1988) before identification by mass spectrometry (MS). Recent methods such as nuclear magnetic resonance (NMR) studies (Meroueh *et al.*, 2006; Kim, Chang and Singh, 2015), ultra high pressure liquid chromatography (UPLC) (Kühner *et al.*, 2014) and direct injection of the HPLC eluate directly into a MS (Bui *et al.*, 2009), have allowed for the peptidoglycan components of various species to be characterised, including, but not limited, to: *E. coli*, *B. subtilis*, *Enterococcus gallinarum*, *Borrelia burgdorferi*, *C. difficile*, *Campylobacter jejuni* , *Streptomyces coelicolor*, *Neisseria gonorrhoeae, Staphylococcus aureus* and *Caulobacter cresentus*) (Ghuysen and Strominger, 1963; Schleifer and Kandler, 1972; Blundell and Perkins, 1981; Glauner, Höltje and Schwarz, 1988; Atrih *et al.*, 1999; Grohs *et al.*, 2000; Bui *et al.*, 2009; Peltier *et al.*, 2011; van der Aart *et al.*, 2018; Billini *et al.*, 2019; Frirdich *et al.*, 2019; Jutras *et al.*, 2019).

### 1.4.1. Peptidoglycan composition

Peptidoglycan is typically formed by strands of alternating *N*-acetylglucosamine (Glc*N*Ac) and *N*-acetylmuramic acid (Mur*N*Ac) residues linked by β-1,4 glycosidic bonds. These glycan strands are then connected by short peptides, typically alanine-glutamine-*meso*diaminopimelate (*meso*DAP)-alanine bound to the Mur*N*Ac residues (Vollmer et al., 2008; Isidro et al., 2017; Peltier et al., 2013). In *E. coli*, the majority of peptides are joined between amino acids at the third and fourth positions on the peptide stems (Glauner, 1988; Glauner, Höltje and Schwarz, 1988) as schematised in Figure 1-5. The termini of glycan strands often contain a 1,6-anhydro-Mur*N*Ac residue (Vollmer, 2008; Vollmer *et al.*, 2008)

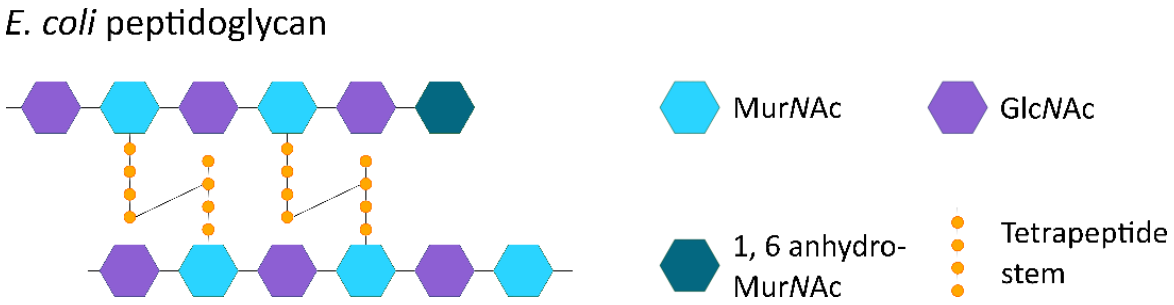


Figure 1-5: Peptidoglycan composition.

*E. coli* peptidoglycan is formed of Glc*N*Ac- Mur*N*Ac polymers, crosslinked by peptide stems, typically tetrapeptides crosslinked between amino acids at the third and fourth positions. The end of a glycan strand is modified slightly, to contain a 1, 6-anhydro-muramic acid residue.

Whilst the "general" composition of peptidoglycan is detailed above, modifications in the glycan strands, peptide stems and crosslinks are widespread (Schleifer and Kandler, 1972; Vollmer, 2008; Vollmer, Blanot and de Pedro, 2008). For example, the glycan strand can be

modified by O- acetylation (Blundell and Perkins, 1981), N- deacetylation (Peltier *et al.*, 2011; Moynihan, Sychantha and Clarke, 2014) or N-glycolation of the Mur*N*Ac residue (Raymond *et al.*, 2005). The terminal amino acid of a peptide stem is also known to be variable; *E. coli,* for example, has been shown to be able to incorporate any of the provided amino acids as long as the amine group was found in a D-stereocentre (Lupoli *et al.*, 2011). Amidation of the second and third position amino acid is not uncommon (Vollmer, Blanot and de Pedro, 2008; Figueiredo *et al.*, 2012); amidation of the third amino acid may aid in evasion of the host immune system (Vijayrajratnam *et al.*, 2016). Crosslinkages are not exclusively directly between peptide stems; *Staphylococcus aureus* peptidoglycan requires pentaglycine bridges between peptide stems to withstand the internal turgor pressure (Monteiro *et al.*, 2019).

### 1.4.2. Peptidoglycan synthesis

Peptidoglycan synthesis can be considered in three phases: the cytoplasmic (reviewed in Barreteau et al., 2008), the membrane-associated (Bouhss *et al.*, 2008) and the crosslinking stages (Vollmer, Blanot and de Pedro, 2008) (Figure 1-6).

Firstly, fructose-6-phosphate is metabolised to uridine diphosphate Glc*N*Ac (UDP-Glc*N*Ac), which is then used to produce uridine diphosphate MurNAc (UDP-Mur*N*Ac) (Barreteau *et al.*, 2008). Formation of UDP-Mur*N*Ac is the first committed step in peptidoglycan synthesis and requires two enzymes; MurA and MurB (Marquardt *et al.*, 1992; Benson *et al.*, 1993; Tayeh *et al.*, 1995). The peptide stem is then built upon UDP-Mur*N*Ac to produce UDP-Mur*N*Ac-pentapeptide. Typically, L-Ala, D-Glu and *meso*DAP are added by MurC (Liger *et al.*, 1995), MurD (Pratviel-Sosa *et al.*, 1994) and MurE (Abo-Ghalia *et al.*, 1985), respectively, with the dipeptide D-Ala-D-Ala added by MurF (Duncan, Van Heijenoort and Walsh, 1990; Yan *et al.*, 2000).

MraY then transfers the phospho-Mur*N*Ac-pentapeptide to the lipid carrier undecaprenyl phosphate, producing lipid I (Boyle and Donachie, 1998; Manat *et al.*, 2014). UDP-Glc*N*Ac is then transferred to lipid I by MurG to produce lipid II (Ha *et al.*, 2000). Lipid II is flipped across the membrane by MurJ and/ or FtsW (Mohammadi *et al.*, 2011; Sham *et al.*, 2014; Kumar *et al.*, 2019), which is then transported to the peptidoglycan layer of the bacterium, before crosslinking by glycosyltransferases and transpeptidases occurs.

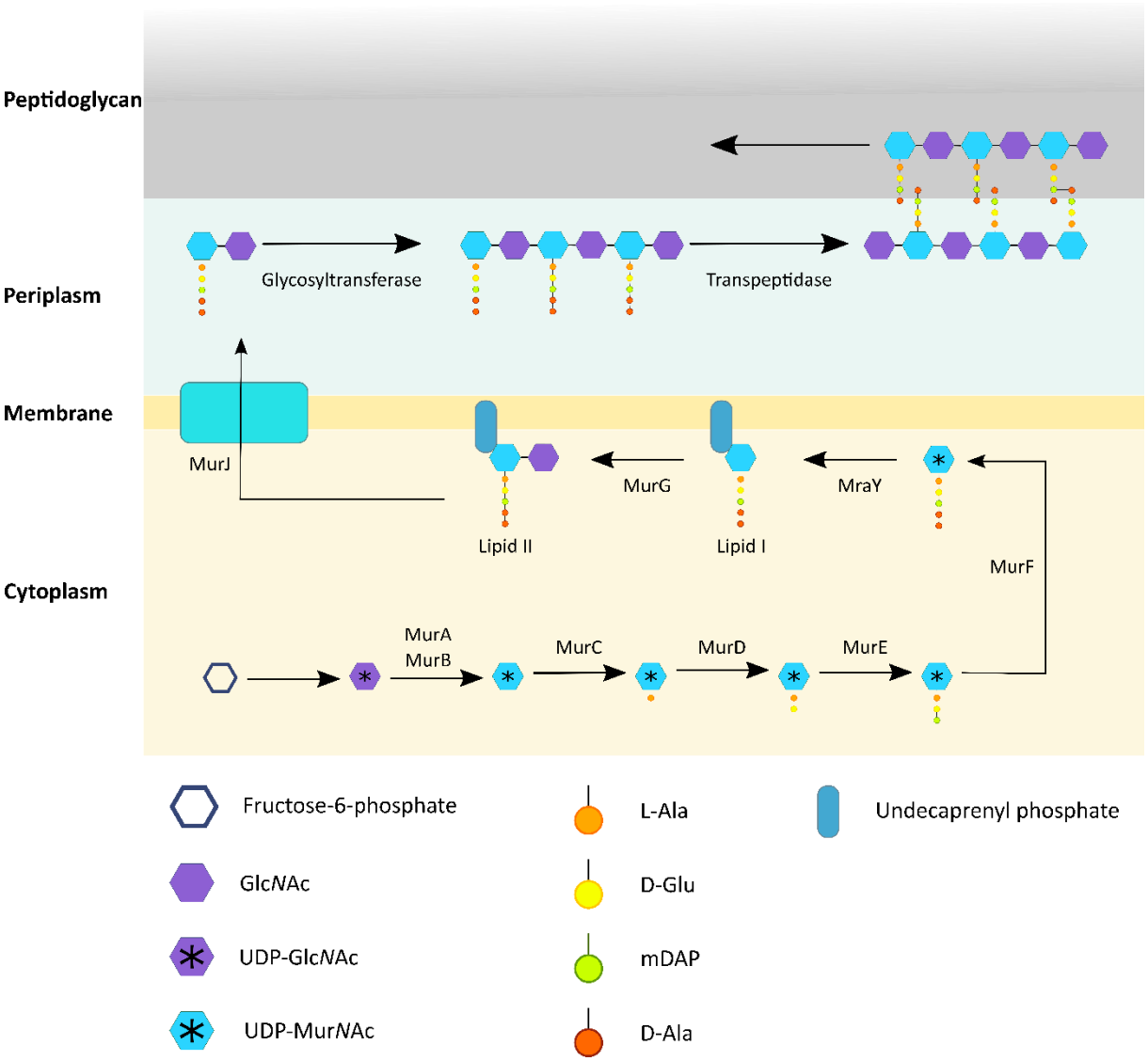


Figure 1-6: Schematic summary of peptidoglycan synthesis

Peptidoglycan synthesis begins in the cytoplasm with the synthesis of UDP-Glc*N*Ac from fructose-6-phosphate. This is used by MurA and MurC to synthesise UDP-Mur*N*Ac. The pentapeptide stem is added to UDP-Mur*N*Ac by MurC, MurD, MurE and MurF. MraY catalyses the reaction of the UDP-Mur*N*Ac pentapeptide and undecaprenyl phosphate to form lipid I. MurG then adds Glc*N*Ac, forming lipid II, which is then flipped across the membrane and into the periplasm by MurJ. Glycosyltransferases then form longer glycan strands which are crosslinked by transpeptidases and incorporated into the peptidoglycan sacculus.

## 1.5. *C. difficile* vegetative cell peptidoglycan

There has been little investigation into peptidoglycan synthesis in *C. difficile*, however homologues of several peptidoglycan synthesis enzymes have been identified in the genome (Isidro *et al.*, 2017).

Whilst the peptidoglycan composition of *C. difficile* resembles that of *E. coli*, there are some key differences, summarised in Figure 1-7.

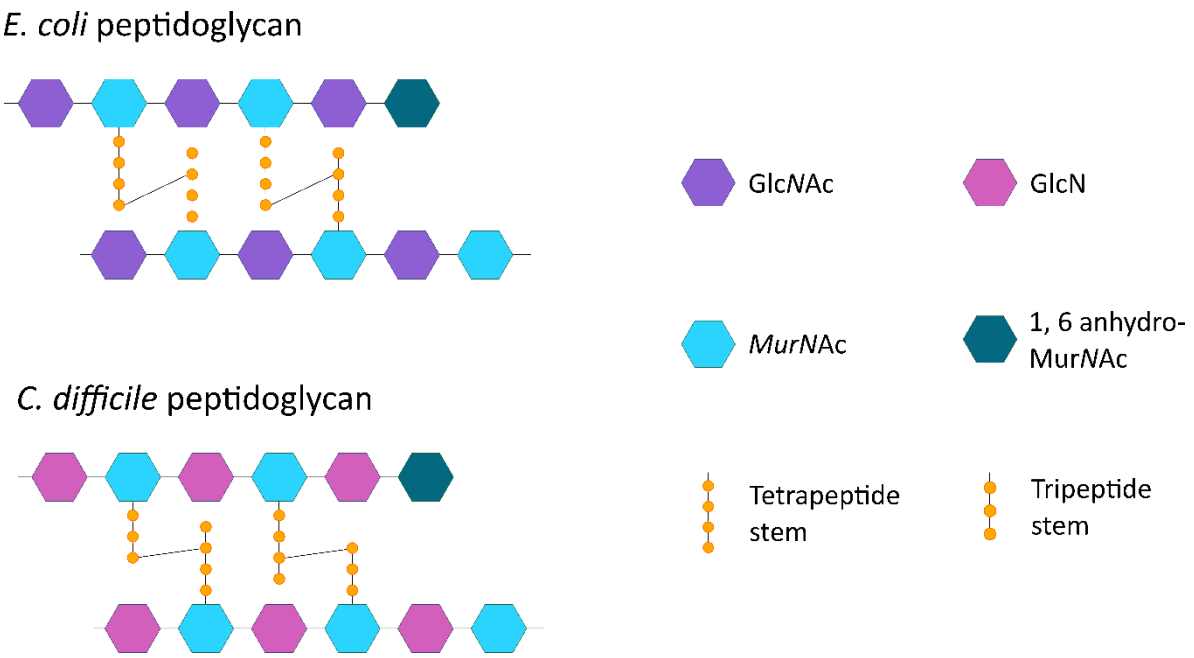


Figure 1-7: Schematic comparison of *E. coli* and *C. difficile* peptidoglycan composition

*E. coli* peptidoglycan glycan strands predominantly contain Glc*N*Ac (purple), whilst *C. difficile* is deacetylated and contains GlcN (pink). Crosslinks in *E. coli* are typically 4-3 between two tetrapeptides, whereas the most common dimer in *C. difficile* peptidoglycan contains a 3-3 crosslink between a tetra- and tripeptide.

### 1.5.1. The glycan backbone of *C. difficile* peptidoglycan

The sugar backbone of *C. difficile* is modified; 93% of Glc*N*Ac residues are deacetylated to glucosamine (GlcN) (Peltier *et al.*, 2011). This deacetylation has previously been described in *B. subtilis* though to a much lesser extent, where ~17% of Glc*N*Ac residues are deacetylated (Atrih *et al.*, 1999).

Peptidoglycan deacetylation is associated with increased resistance to lysozyme (Amano *et al.*, 1977) and increased virulence; when *pgdA* was deleted in *Streptococcus pneumoniae,* cells were more sensitive to lysozyme and mice were less susceptible to disease (Vollmer and Tomasz, 2000). Furthermore, in *Listeria monocytogenes* peptidoglycan deacetylation is key to evasion of the host immune system (Boneca *et al.*, 2007). Deletion of a key deacetylase gene, *pgdA*, results in reduced infectivity in mouse models of *S. pneumoniae* infection (Blair *et al.*, 2005) for example.

Deacetylation can have physiological functions, with the *Bacillus anthracis* deacetylases BA1961 and BA3679 implicated in vegetative cell elongation and division (Balomenou *et al.*, 2013). BA0330 and BA0331 are putative polysaccharide deacetylases implicated in the

adaptation of *B. anthracis* to high salt concentrations and maintenance of cell shape, respectively, though they do not alter the peptidoglycan structure *per se*, but may stabilise the cell wall (Arnaouteli *et al.*, 2015).

In *C. difficile*, $\sigma^V$ is critical to lysozyme resistance, increasing deacetylation of peptidoglycan. Removal of the $\sigma^V$ gene results in an attenuated virulence strain that is less able to cause CDI in a hamster model of infection (Ho *et al.*, 2014), highlighting the key role of lysozyme resistance, and therefore peptidoglycan deacetylation, in CDI.

Consequently, deacetylases are attractive targets in novel therapeutic development.

#### 1.5.1.1. Peptidoglycan deacetylases

Peptidoglycan deacetylases belong to the carbohydrate esterase 4 family of enzymes (http://www.cazy.org) (Lombard *et al.*, 2014) which also includes chitin and chitooligosaccharide deacetylases and acetylxylan esterases. Members typically adopt a distorted $(\alpha/\beta)_8$ barrel fold, use divalent metal ions in catalysis and all contain a NodB homology domain (Pfam domain PF01522), within which the catalytic motif is found (Caufrier *et al.*, 2003; Nakamura, Nascimento and Polikarpov, 2017).

Part of the NodB homology domain is the well conserved Asp-His-His triad, which is typically used to coordinate a metal ion, in conjunction with a water molecule (Nakamura, Nascimento and Polikarpov, 2017; Aragunde, Biarnés and Planas, 2018). This metal ion is often zinc (Andrés *et al.*, 2014), but cadmium, cobalt, iron and magnesium binding have all been described, with differing effects depending on the enzyme in question (Martinou, Koutsioulis and Bouriotis, 2002; Blair and van Aalten, 2004; Taylor *et al.*, 2006; Chibba *et al.*, 2012). Cobalt binding, for example, is associated with a ~30-fold increase in *S. pneumoniae* PdgA (*Sp*PdgA) activity (Blair *et al.*, 2005). Removal of the metal ion in *Sp*PgdA abolishes deacetylase activity (Blair *et al.*, 2005), treatment with 10µM EDTA is sufficient to reduce *Streptococcus mutans* PdgA activity by 97% (Deng *et al.*, 2009), and treatment with EDTA prevents the activity of *Vibrio cholerae* chitin deacetylase (Andrés *et al.*, 2014), demonstrating the conservation of metal binding and its key role in activity.

Characterised peptidoglycan deacetylases identified to date deacetylate Glc*N*Ac to GlcN, with the notable exception of PdaA in *B. subtilis,* which is implicated in Mur*N*Ac deacetylation to muramic-δ-lactam during spore formation (Blair and van Aalten, 2004; Blair *et al.*, 2005).

### 1.5.2. The peptide stems and crosslinks of *C. difficile*

Analysis of the peptidoglycan of *C. difficile* vegetative cells has revealed ~70% of crosslinks between peptides are between 2 amino acids at the third positions (Peltier *et al.*, 2011). This difference in crosslinkage highlights a difference in the synthesis pathways of *C. difficile* and *E. coli* peptidoglycan. The 4-3 crosslinks typically seen in *E. coli* peptidoglycan are formed by the D, D-transpeptidase activity of penicillin binding proteins (PBPs) (Waxman and Strominger, 1983; Pratt, 2008; Zapun, Contreras-Martel and Vernet, 2008). In contrast, the 3-3 crosslinks that predominate in *C. difficile* peptidoglycan are typically formed by L,D-transpeptidases (Sütterlin *et al.*, 2017).

Both alanine and glycine have been described as the terminal amino acid in *C. difficile* peptide stems (Peltier *et al.*, 2011). The physiological relevance of the terminal amino acids

is unknown in *C. difficile*, although, in *Vibrio cholerae,* incorporation of non-canonical amino acids has been implicated in determination of peptidoglycan mass, strength and the ability of the cells to resist osmotic stress (Cava *et al.*, 2011).

### 1.5.3. *C. difficile* spore cortex peptidoglycan

The spore cortex peptidoglycan of *C. difficile* is significantly different to that of the vegetative cell as demonstrated in Figure 1-8 and in preliminary work (Kelly, 2016).

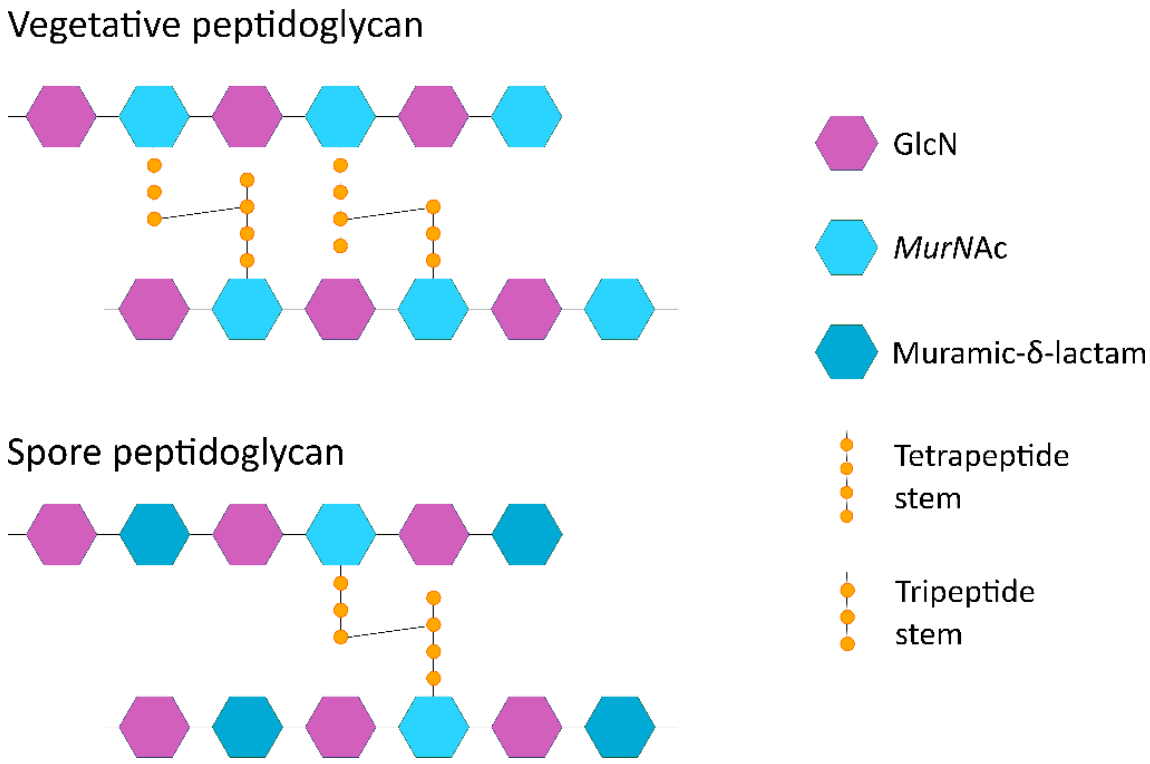


Figure 1-8: C. *difficile* vegetative cell *versus* spore peptidoglycan.

Vegetative cell peptidoglycan contains many more crosslinks than spore peptidoglycan. Muramic-δ-lactam residues have only ever been described in spore cortex peptidoglycan, and crosslinks between peptides are significantly reduced.

Glc*N*Ac deacetylation in spore cortex peptidoglycan is reduced to 54.7%, from the ~93% seen in vegetative cells (Peltier *et al.*, 2011; Coullon *et al.*, 2018). Although the mechanisms that lead to this reduction are unclear, one possibility could be due to reduced peptidoglycan deacetylase activity during sporulation, possibly during engulfment.

Monomers (disaccharides with peptide stems) are more prevalent in the spore peptidoglycan, accounting for up to 90% of all muropeptides detected, and dimers (two disachaarides joined by two crosslinked peptide stems) are reduced to about 10% of all detected muropeptides from ~56% in the vegetative cell (Peltier *et al.*, 2011; Coullon *et al.*, 2018).

Furthermore, 24% of all muropeptides contain a modified Mur*N*Ac residue: the peptide stem is removed from Mur*N*Ac, prior to Mur*N*Ac deacetylation and lactam ring formation to produce muramic-δ-lactam (MAL) (Gilmore *et al.*, 2004; Coullon *et al.*, 2018). MAL residues are unable to carry peptide stems, and consequently the crosslinking index of spore

peptidoglycan is significantly reduced to 4.9% (Coullon *et al.*, 2018). Similarly modified spore cortex peptidoglycan has been described in *B. subtilis* spores where crosslinking was reduced to ~3% of Mur*N*Ac residues, with ~50% of the Mur*N*Ac residues modified to MAL residues (Warth and Strominger, 1972; Atrih *et al.*, 1996).

It is thought that this highly modified spore peptidoglycan contributes to the hardiness of the spore, for example by maintaining spore core dehydration (Atrih *et al.*, 1996).

### 1.5.4. Peptidoglycan remodelling during engulfment

Engulfment refers to the surrounding of the smaller compartment by the larger, which is referred to as the mother cell. Again, most of our current understanding comes from work completed in the model Gram-positive model organism *B. subtilis.*

During engulfment, the asymmetric septum curves and the mother cell membrane moves around the forespore, until the advancing membrane meets and fuses, producing the cell-within-a-cell forespore structure.

For the mother cell membrane to advance, the steric hindrance of peptidoglycan must be removed without compromising cell integrity. Based largely on studies in *B. subtilis*, it has been shown that both peptidoglycan digestion and synthesis are required at the leading edge to allow engulfment.

Peptidoglycan synthesis has been localised to the advancing membrane using fluorescent microscopy (Meyer *et al.*, 2010). Blocking peptidoglycan synthesis with cephalexin, a PBP inhibitor, or fosfomycin, which blocks the initial step in peptidoglycan synthesis, prevents engulfment of the smaller compartment (Meyer *et al.*, 2010; Ojkic *et al.*, 2016), suggesting peptidoglycan synthesis is required for engulfment. Indeed, several peptidoglycan synthesis-related enzymes were specifically localised in the forespore and seen to track the leading edges of the advancing membrane during engulfment (Ojkic *et al.*, 2016). Furthermore, inhibition of the *C. difficile* sporulation-specific PBP SpoVD by cefotoxin prevents spore formation (Srikhanta *et al.*, 2019), demonstrating the requirement of continued peptidoglycan synthesis in spore formation.

Microscopy work in *B. subtilis* has demonstrated that peptidoglycan is digested uniformly across the septum, but that the septal peptidoglycan is never completely digested (Tocheva *et al.*, 2013; Ojkic *et al.*, 2016; Khanna *et al.*, 2019), contrary to previous “septal thinning hypotheses” (Chastanet and Losick, 2007).

The most recent model of engulfment, proposed whilst this work was ongoing, is that initially proposed by Ojkic *et al*., (2016), and recently strengthened based on observations in *B. subtilis* (Khanna *et al.*, 2019). Ojkic *et al*. (2016) proposed a model, termed the make-before-break model, in which new peptidoglycan is first inserted ahead of the advancing membrane by a forespore associated synthesis machinery. This new peptidoglycan is then targeted for digestion by peptidoglycan hydrolases, which results in peptidoglycan removal, allowing the mother cell membrane to move into this newly available space through entropic forces as illustrated in Figure 1-9.

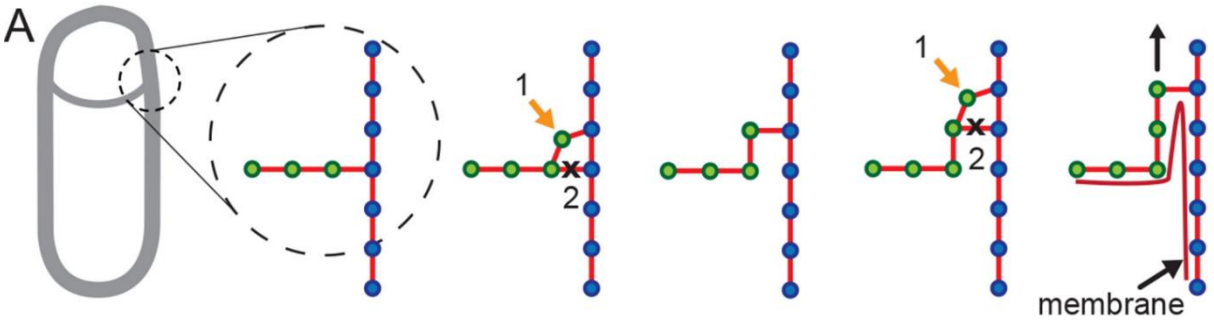


Figure 1-9: Make-before-break model of engulfment.

Glycan strands run parallel to the long axis of the cell. Blue circles denote old mother cell peptidoglycan, and green newly inserted peptidoglycan. Red lines are peptide bonds linking peptidoglycan types. New peptidoglycan is first inserted ahead of the leading edge, before the bond between the old and new peptidoglycan is digested by the DMP complex (black cross). This process is repeated, creating space for the mother cell membrane move into the newly created space (black arrow) (Ojkic *et al.*, 2016).

## 1.6. The DMP machinery

The DMP machinery consists of SpoIID, SpoIIM and SpoIIP and was first identified in *B. subtilis*, but has recently been described in *C. difficile* and other spore-formers (Morlot *et al.*, 2010; Kelly and Salgado, 2019). In *B. subtilis* SpoIIP is an amidase and transpeptidase, capable of removing the peptide stems from *E. coli* peptidoglycan, leaving "denuded" glycan strands that are the substrate for the lytic transglycosylase activity of SpoIID (Chastanet and Losick, 2007; Morlot *et al.*, 2010). SpoIID activity first requires naked glycan strands, and produces glycans with a terminal 1,6- anhydro-Mur*N*Ac residue (Morlot *et al.*, 2010) as shown in Figure 1-10.

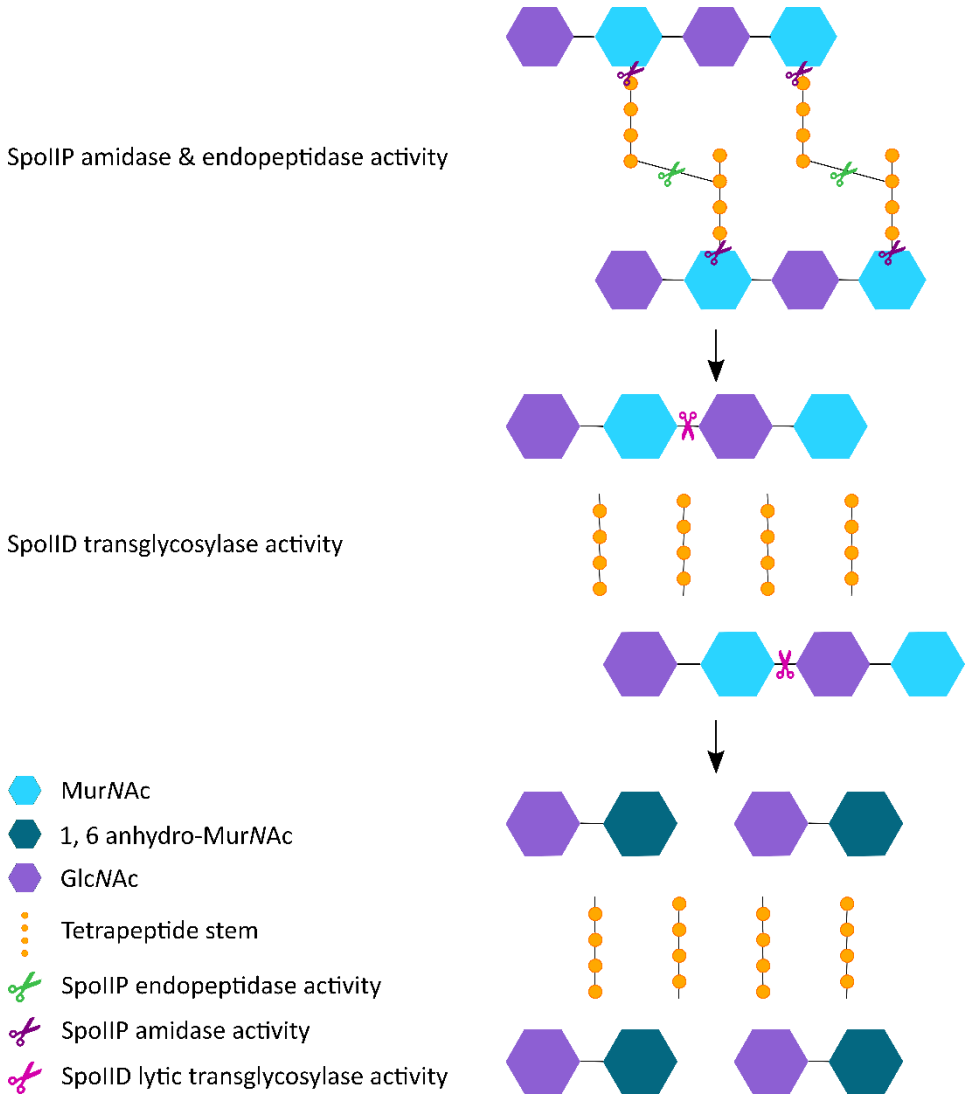


Figure 1-10: Peptidoglycan digestion by the DMP machinery.

Intact *E. coli* peptidoglycan can be digested by the amidase activity of SpoIIP producing naked glycans and peptide stems. Crosslinked peptides can be further digested by SpoIIP's endopeptidase activity to produce free tetrapeptides. The denuded glycans are further digested by SpoIID to produce disaccharides where 1,6-anhydro-Mur*N*Ac is found at the terminus.

### 1.6.1. SpoIIM

In *B. subtilis*, SpoIIM is thought to be the first protein of the DMP complex to localise to the asymmetric septum during engulfment and is under the control of the early sigma factor $\sigma^E$

(Smith, Bayer and Youngman, 1993; Chastanet and Losick, 2007). SpoIIM is unique in the DMP complex in that it is not predicted to have any enzymatic activity, it is thought to act as a scaffold for SpoIIP and SpoIID to localise to the membrane, first recruiting SpoIIP then SpoIID to the asymmetric septum (Aung *et al.*, 2007; Chastanet and Losick, 2007). When *spoIIM* is deleted, both SpoIIP and SpoIID localisation is disrupted (Chastanet and Losick, 2007).

Deletion of *spoIIM* prevents *B. subtilis* from forming spores; presumably because SpoIIM gene products are required for the downstream production of $\sigma^G$-associated gene products (Smith, Bayer and Youngman, 1993), and its role in recruiting SpoIIP and SpoIID to the advancing membrane (Chastanet and Losick, 2007).

### 1.6.2. SpoIIP

Transcription of *spoIIP* is controlled by $\sigma^E$ in *B. subtilis* (Piggot and Hilbert, 2004), and the protein localises to the midpoint of the mother cell membrane, before tracking the leading edges of the advancing membrane, as demonstrated by microscopy of green fluorescent protein (GFP) fusion proteins (Abanes-De Mello *et al.*, 2002; Chastanet and Losick, 2007).

Peptidoglycan digestion assays have demonstrated that *B. subtilis* SpoIIP is an amidase and an endopeptidase; cleaving the peptide stems from the glycan chains, and peptide stems from each other respectively (Aung *et al.*, 2007; Morlot *et al.*, 2010). Similarities between *B. subtilis* SpoIIP and CwlV lead to the identification of H189 and H278 as amidase catalytic residues, as mutation of these residues to arginine abolished SpoIIP activity (Chastanet and Losick, 2007).

Deletion of *spoIIP* in *B. subtilis* results in the formation of membrane bulges, thought to be the consequence of the uncoupling of peptidoglycan synthesis and degradation (Meyer *et al.*, 2010). Additionally, SpoIIP depleted cells are unable to complete engulfment to produce heat resistant spores (Eichenberger, Fawcett and Losick, 2002).

This inability of Δ*spoIIP* cells to produce heat resistant spores is not surprising given the enzymatic activity of SpoIIP, and the requirement of SpoIIP products for SpoIID activity.

### 1.6.3. SpoIID

SpoIID is under transcriptional control of the mother cell-produced $\sigma^E$. Like with SpoIIP, GFP-fusions in *B. subtilis* have demonstrated SpoIID initially localises to the septal midpoint then tracks the advancing edge, though some diffuse mother cell membrane localisation is observed (Chastanet and Losick, 2007). In *B. subtilis* Δ*spoIIP* strains, SpoIID localisation is lost, demonstrating the dependency of SpoIID on SpoIIP for localisation to the advancing mother cell membrane (Aung *et al.*, 2007; Chastanet and Losick, 2007; Gutierrez, Smith and Pogliano, 2010). Deletion of *spoIID* results in the inability to form spores, with membrane bulges observed in *B. subtilis* (Gutierrez, Smith and Pogliano, 2010).

SpoIID itself has enzymatic activity as a lytic transglycosylase (CAZy family GH23); digestion of the glycosidic bond between Glc*N*Ac and Mur*N*Ac results in the production of 1,6-anhydro-Mur*N*Ac containing muropeptides (Abanes-De Mello *et al.*, 2002; Scheurwater, Reid and Clarke, 2008; Gutierrez, Smith and Pogliano, 2010; Morlot *et al.*, 2010; Pfeffer *et al.*, 2012). This reaction mechainism is distinct from that ustilised by lysozymes for example,

where the glyosidic bond is hydrolysed and does not result in the formation of 1,6-anhydro-Mur*N*Ac residues (Figure 1-11).

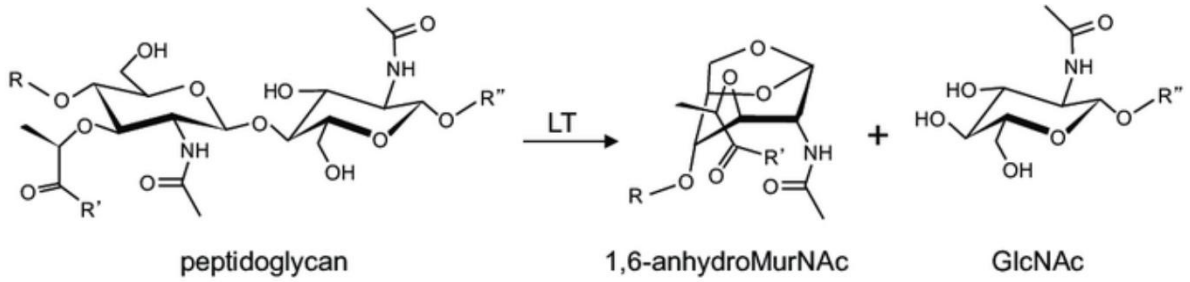


Figure 1-11: Lytic transglycosylases produce 1, 6-anhydro-Mur*N*Ac residues

The glycosidic bond between two sugar residues in peptidoglycan is digested by lytic transglycosylases producing products containing 1,6-anhydro-Mur*N*Ac residues. Figure modified from (Pfeffer *et al.*, 2012)

However, SpoIID is inactive on "intact" peptidoglycan and requires the peptide-free denuded glycan strands produced as a consequence of SpoIIP activity (Abanes-De Mello *et al.*, 2002; Gutierrez, Smith and Pogliano, 2010; Morlot *et al.*, 2010). This need for sequential SpoIIP-SpoIID activity is proposed to prevent the release of muropeptides that may trigger immediate spore germination (Shah *et al.*, 2008).

Point mutation of glutamate 88 in *B. subtilis* results in a loss of enzymatic activity (Morlot *et al.*, 2010) and prevents spore formation (Gutierrez, Smith and Pogliano, 2010), suggesting E88 is a key catalytic residue, and demonstrating that SpoIID activity is key in engulfment and, consequently, in sporulation.

During the course of this work, the structure of *C. difficile* SpoIID was determined to 1.95 Å by Nocadello *et al*., in 2016 and a mechanism for SpoIID activity based on the structure of SpoIID co-crystallised with the artificial substrate triacetylchiotriose ($NAG_3$) (PDB ID: 5I1T) was proposed (Nocadello *et al.*, 2016).

SpoIID consists of an α-helix rich α-domain, which forms the "hand", and a β-strand rich β-domain forming the "arm" of the protein (Nocadello *et al.*, 2016). Two $NAG_3$ substrate-analogue molecules sit across subsites -4 to +3, with the three sugar units of one molecule occupying positions -4, -3 and -2, and those in the second molecule occupying subsites +1, +2 and +3, with contacts between SpoIID and the sugars across all subsites.

By computationally fitting a Mur*N*Ac residue in the unoccupied -1 subsite, the authors propose an enzymatic mechanism (Nocadello *et al.*, 2016). Firstly, the carbonyl of E101, equivalent to E88 in *B. subtilis* SpoIID, protonates the oxygen of the β-1,4-Mur*N*Ac -Glc*N*Ac glycosidic bond. The C6 hydroxyl group of Mur*N*Ac is then deprotonated by O6 of E101, while Y194 facilitates nucleophilic attack of O1 by O6 within the Mur*N*Ac ring, to form the cyclical 1,6 -anhydro-Mur*N*Ac residue.

Several residues are described as interacting with the $NAG_3$ substrate: Y229, W325, S196, Q324, E101, E148, H318, C164, H145, Y213 and S210 (Nocadello *et al.*, 2016).

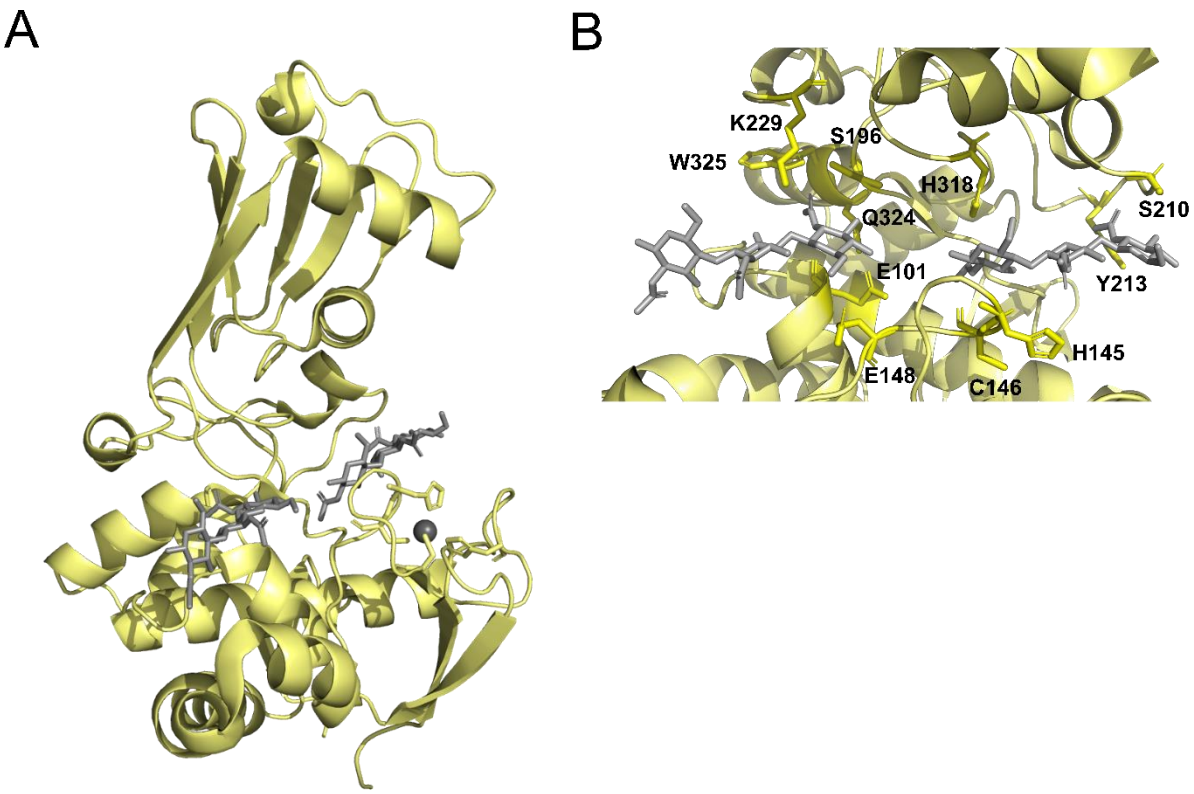


Figure 1-12: SpoIID model with artificial substrate.

Panel A shows the structure of SpoIID (pale yellow) co-crystallised with $NAG_3$, produced by Nocadello *et al*. (2016) (PDB ID: 5I1T), with the grey sphere representing the presumed zinc residue, and the grey sticks the $NAG_3$ substrate. Panel B shows the $NAG_3$ binding domains in greater detail.

Of note are the interactions of Y194, S196, Q324 and H318 which are proposed to interact with the acetyl groups of $NAG_3$. *C. difficile* peptidoglycan is known to be highly deacetylated (Peltier *et al.*, 2011; Coullon *et al.*, 2018) (section 1.5), therefore, the *in vivo* relevance of this interaction is questionable as one would expect SpoIID to interact primarily with GlcN, not Glc*N*Ac, containing peptidoglycan. Furthermore, $NAG_3$ lacks the Mur*N*Ac residue that would be present in the biological ligand, and the presence of this Mur*N*Ac may influence the entire network of proposed binding sites, though the authors propose there is room within subsite -1 for the expected Mur*N*Ac residue (Nocadello *et al.*, 2016). Mutation of S196 and Q324 to alanine reduced the ability of SpoIID to digest acetylated *E. coli* peptidoglycan (Nocadello *et al.*, 2016), and mutation of the equivalent H138 in *B. subtilis* prevented spore formation (Gutierrez, Smith and Pogliano, 2010; Morlot *et al.*, 2010), suggesting some of the SpoIID-$NAG_3$ interactions might be biologically relevant.

## 1.7. Aims

The overarching aim of this thesis was to further the understanding of *C. difficile* engulfment during sporulation, particularly with regards to the involvement of the DMP complex and peptidoglycan composition.

The first aim was to characterise the changes to peptidoglycan as *C. difficile* cells move away from a vegetative state, towards the highly specialised spore cortex peptidoglycan. Whilst *C. difficile* vegetative cell peptidoglycan and spore peptidoglycan was characterised immediately preceeding this work (Peltier *et al.*, 2011; Kelly, 2016), the actual process of the shift from vegetative to spore peptidoglycan had not been investigated. The results of these analyses are presented in Chapter 3.

Secondly, this work set out to investigate the DMP complex in *C. difficile*; specifically, its enzymatic activity and the crystal structure of the components. Prior to this only the DMP complex of *B. subtilis* had been assessed in any manner. The results of these investiagations are presented in Chapter 4.

Following the work on the DMP complex, I hypothesised that the acetylation state of peptidoglycan may impact, or even direct, the activity of SpoIID and SpoIIP during engulfment, which may have implications for our current understanding of engulfment. Therefore, the third aim of this work was to identify and characterise potential peptidoglycan deacetylases that may be active during sporulation in *C. difficile*. This work is presented in Chapter 5.

# Chapter 2. Materials and methods

## 2.1. Bacterial strains and growth conditions

### 2.1.1. *Escherichia coli*

*E. coli* NEB5-α cells [New England Biolabs (NEB)] incubated at 37 $^{o}$C were used for cloning throughout this work. For recombinant protein expression, Rosetta DE3 [Novagen] cells were used, in the media (Table 2-2) and at the temperature identified as optimal for the expression of that construct (section 2.4). A complete list of *E. coli* mutants used/produced in this thesis can be found in Table 2-1.

Table 2-1: *E. coli* strains used in this work

| Strain number | Plasmid name | Host strain | Plasmid backbone | Resistance | Description |
|---|---|---|---|---|---|
| PS042 | N/A | Rosetta | NF1329 | Amp, Cm | TEV expression |
| PS219 | pAXK001 | NEB5-α | pET-M11 | Km | Storage strain for plasmid containing soluble fragment of *C. difficile 630 spoIID* (AA 26-354) |
| PS220 | pAXK001 | Rosetta | pET-M11 | Km, Cm | Expression strain containing soluble fragment of *C. difficile 630 spoIID* (AA 26-354) |
| PS221 | pAXK002 | NEB5-α | pET-M11 | Km | Storage strain for plasmid containing soluble fragment of *C. difficile 630 spoIIP* (AA 27-399) |
| PS222 | pAXK002 | Rosetta | pET-M11 | Km, Cm | Expression strain containing soluble fragment of *C. difficile 630 spoIIP* (AA 27-399) |
| PS350 | pMLD179 | NEB5-α | pET-M11 | Km | Storage strain for plasmid containing soluble fragment of *C. difficile 630 spoIIP* (AA 27-399) with H142R mutation introduced via inverse PCR |
| PS351 | pMLD180 | NEB5-α | Pet-M11 | Km | Storage strain for plasmid containing soluble fragment of *C. difficile 630 spoIIP* (AA 27-399) with H222R mutation introduced via inverse PCR |
| PS359 | pAXK003 | NEB5-α | pET-M11 | Km | Storage strain for plasmid produced to introduce C140A mutation in soluble fragment of *C. difficile 630 spoIID* (AA 26-354) |
| PS360 | pAXK004 | NEB5-α | pET-M11 | Km | Storage strain for plasmid produced to introduce C146A mutation in soluble fragment of |

| Strain number | Plasmid name | Host strain | Plasmid backbone | Resistance | Description |
|---|---|---|---|---|---|
| | | | | | *C. difficile 630 spoIID* (AA 26-354) |
| PS361 | pAXK005 | NEB5-α | pET-M11 | Km | Storage strain for plasmid produced to introduce H145A mutation in soluble fragment of *C. difficile 630* SpoIID (AA 26-354) |
| PS362 | pAXK006 | NEB5-α | pET-M11 | Km | Storage strain for plasmid produced to introduce E101A mutation in soluble fragment of *C. difficile 630* SpoIID (AA 26-354) |
| PS383 | pAXK017 | NEB5-α | pET-M11 | Km | Storage strain for plasmid containing soluble fraction of *C. difficile 630 CD630_15520* (AA 29-293) |
| PS384 | pAXK018 | NEB5-α | pET-M11 | Km | Storage strain for plasmid containing soluble fraction of *C. difficile 630 CD_32570* (AA 30-275) |
| PS385 | pAXK019 | NEB5-α | pET-M11 | Km | Storage strain for plasmid containing soluble fraction of *C. difficile 630 CD_13190* (AA 36-247) |
| PS386 | pAXK020 | NEB5-α | pET-M11 | Km | Storage strain for plasmid containing soluble fraction of *C. difficile 630 CD_27240* (AA 53-232) |
| PS387 | pAXK021 | Rosetta | pET-M11 | Km, Cm | Expression strain for plasmid containing soluble fraction of *C. difficile 630 CD630_32570* (AA 30-275) |
| PS388 | pAXK022 | Rosetta | pET-M11 | Km, Cm | Expression strain for plasmid containing soluble fraction of *C. difficile 630 CD630_15520* (AA 29-293) |
| PS389 | pAXK023 | Rosetta | pET-M11 | Km, Cm | Expression strain for plasmid containing soluble fraction of *C. difficile 630 CD630_13190* (AA 36-247) |
| PS390 | pAXK024 | Rosetta | pET-M11 | Km, Cm | Expression strain for plasmid containing soluble fraction of *C. difficile 630 CD630_27240* (AA 53-232) |
| PS459 | pAXK025 | NEB5-α | pET-M11 | Km | Storage strain for plasmid containing soluble fraction of *C.* |

| Strain number | Plasmid name | Host strain | Plasmid backbone | Resistance | Description |
|---|---|---|---|---|---|
| | | | | | *difficile 630* CD_14440 (AA 45-249) |
| PS490 | pAXK026 | NEB5-α | pET-M11 | Km | Storage strain for plasmid produced to introduce H134A mutation in soluble fragment of *C. difficile 630 spoIID* (AA 26-354) |
| PS491 | pAXK027 | Rosetta | pET-M11 | Km, Cm | Expression strain for plasmid produced to introduce H134A to soluble fragment of *C. difficile 630 spoIID* (AA 26-354) |

*E. coli* strains used in this thesis. Strain numbers refer to the strain number in the Salgado lab strain collection. Amp= ampicillin, Cm= chloramphenicol, Km= kanamycin. All strains were made during this work, except strain number PS042 which was already in the Salgado lab strain collection, and strains PS350 and PS351 which were made by Dr M Dembek.

Table 2-2: Media compositions

| Media name | Abbreviation | Description | Quantity per litre | Manufacturer |
|---|---|---|---|---|
| Lysogeny Broth | LB | General purpose *E. coli* growth media | 20 g Lysogeny Broth pellets | Fisher |
| Brain heart infusion-supplemented | BHIS | General purpose C. difficile growth media | 37 g Brain heart infusion | Oxoid |
| | | | 5 g Yeast extract | Melford |
| | | | 1 g L-cysteine | Sigma |
| Terrific broth* | TB | Expression of sSpoIID, sSpoIIP and associated mutant proteins | 24 g Yeast extract | Melford |
| | | | 12 gTryptone | For Medium |
| | | | 4 ml Glycerol | Merck Millipore |
| TB supplement | N/A | Buffer autoclaved separately and added to TB post- autoclave | 23.1 g Potassium phosphate monobasic | Merck Millipore |
| | | | 125.4 g Potassium phosphate dibasic | Merck Millipore |
| Auto-induction media terrific broth | AIM TB | Expression of deacetylase constructs | 55.85 g AIM TB with trace elements | For Medium |
| Sporulation media | SM | For production of *C. difficile* spores | 90 g Bacto Peptone | BD Biosciences |
| | | | 5g Proteose peptone | Fulka |
| | | | 1g Ammonium sulphate | Sigma |
| | | | 1.5 g Tris Base | Sigma |

All media were made using MilliQ water and autoclaved before use. TB and AIM TB were autoclaved on a shorter cycle of 121 $^{o}$C for 20 minutes. *recipe for TB is in 900ml of MilliQ prior to autoclave. Once cooled 100ml of autoclaved TB supplement was added aseptically. For solid media, 15g/L agar was added before autoclaving.

### 2.1.1. *Clostridioidies difficile*

All *C. difficile* strains (Table 2-3) were grown anaerobically (10% $H_2$, 10% $CO_2$, 80% $N_2$) in a DG250 workstation [Don Whitley Scientific] at 37 $^oC$, in an appropriate medium, as described in Sections 2.9

Table 2-3: *C. difficile* strains used in this work

| Strain number | Strain name | Description | Source |
|---|---|---|---|
| PS0002 | 630 | Virulent clinical strain isolated in Zurich 1982 ((Hussain, Roberts and Mullany, 2005)) | Adriano Henriques |
| PS0003 | 630 Δ*erm* | Erythromycin sensitive derivative of *C. difficile* 630 | Adriano Henriques |
| PS0020 | 630 Δ*erm* Δ*spoIIQ* | Δ*pyrE* fixed via ACE, *spoIIQ* removed by ACE | Neil Fairweather |
| PS0022 | 630 Δerm Δ*spoIID* | Δ*pyrE* fixed via ACE, *spoIID* removed by ACE | Marcin Dembek |
| PS0026 | 630 Δ*erm* Δ*spoIIP* | Δ*pyrE* fixed via ACE, *spoIIP* removed by ACE | Marcin Dembek |
| PS0050 | 630 Δ*erm* Δ*spoIIIAH* | Δ*pyrE* fixed via ACE, *spoIIIAH* removed by ACE | Adriano Henriques |
| PS0094 | 630 Δ*erm* Δ*pyrE* $P_{tet}$ *spo0A* | Δ*pyrE* truncated by ACE and *spo0A* placed under control of ATc- inducible promoter. | Marcin Dembek |
| PS0097 | 630 Δ*erm* Δ*sigE* | Δ*pyrE* fixed via ACE, *sigE* removed by ACE | Marcin Dembek |
| PS0099 | 630 Δ*erm* Δ*sigF* | Δ*pyrE* fixed via ACE, *sigF* removed by ACE | Marcin Dembek |

No *C. difficile* strains were produced as part of this work; therefore, the creator/source laboratory of each strain is given. ACE refers to allele coupled exchange (Ng et al., 2013), ATc to anhydrotetracycline.

## 2.2. Bacterial storage

All strains produced as part of this thesis were catalogued in the Salgado lab strain collection. 1 ml of an overnight LB or BHIS (*E. coli* strains and *C. difficile* strains, respectively) culture of the sequencing-verified isolate was resuspended in 1ml sterile 50% glycerol and frozen at -80 $^{o}$C.

## 2.3. Molecular biology

All oligonucleotides (Table 2-4 and Table 2-5) were designed using Geneious 11.0.4 (Kearse *et al.*, 2012). Oligonucleotides were designed to be approximately 20-40 bp long, ending with a GC clamp, and to have an annealing temperature within 5 $^{o}$C of its primer pair. XhoI and NcoI restriction sites were incorporated into each primer to allow for a cut and stick approach to cloning. Additionally, all oligonucleotides contained a 5' GATC modification, to facilitate polymerase binding. All oligonucleotides were synthesised by Eurofins and dissolved in the recommended volume of PCR-grade water [Ambion] to 100 pg/µl.

All soluble constructs were cloned into the pET-M11 backbone. This introduces an N-terminal 6x His-tag linked to a Tobacco Edge Virus (TEV) cleavage site, allowing for simple removal of the His-tag during purification. All plasmids produced during this work are listed in Table 2-6.

Table 2-4: Oligonucleotides produced in this thesis

| Name | Oligonucleotide sequence (5'→ 3') | Description | $T_m$ ($^{o}$C) |
|---|---|---|---|
| 339 | GCAATGTCTTCAGAGTTTGATATAGAAGC | Forward primer to introduce E101A to SpoIID$_{35-354}$ *via* inverse PCR | 62.4 |
| 340 | ACCAGCTAGTACTCCACATAAATAGTTC | Reverse primer to introduce E101A to SpoIID$_{35-354}$ *via* inverse PCR | 62.2 |
| 341 | GCAACTGATTATAAACATTGTCAAG | Forward primer to introduce C140A to SpoIID$_{35-354}$ *via* inverse PCR | 56.4 |
| 342 | TACTACTGCATTTTTATGTTTACTAG | Reverse primer to introduce C140A to SpoIID$_{35-354}$ *via* inverse PCR | 55.3 |
| 343 | GCATGTCAAGAATATAAGAG | Forward primer to introduce H145A to SpoIID$_{35-354}$ *via* inverse PCR | 51.1 |
| 344 | TTTATAATCAGTACATACTACTG | Reverse primer to introduce H145A to SpoIID$_{35-354}$ *via* inverse PCR | 51.7 |
| 345 | GCACAAGAATATAAGAGTTATG | Forward primer to introduce C146A to SpoIID$_{35-354}$ *via* inverse PCR | 52.8 |
| 346 | TGTTTATAATCAGTACATACTAC | Reverse primer to introduce C146A to spoIID$_{35-354}$ *via* inverse PCR | 52.5 |
| 367 | GCAAAAAATGCAGTAGTATGTACTG | Forward primer to introduce H134A to sSpoIID$_{35-354}$ | 58.1 |
| 368 | TTTACTAGATTTACCATGTTCTTGTTTATATAC | Reverse primer to introduce H134A to sSpoIID$_{35-354}$ | 59.5 |
| 414 | GATCCCATGGTTAAATATGAAGGAAAATG | Forward sCD_1556 from gDNA and add restriction site, removing AA 1-28 | 51.1 |
| 416 | GATCCTCGAGTTATTTTATCAACTCAG | Reverse primer for sCD_14440 from gDNA and add restriction site, removing AA 1-35 | 50 |

| Name | Oligonucleotide sequence (5'→ 3') | Description | $T_m$ (°C) |
|---|---|---|---|
| 420 | GATCCCATGGAATAAACTATTAAATAAAGAGG | Forward primer to get sCD_27240 from gDNA and add restriction site. Paired with #437 to remove AAs 1-24 | 53 |
| 421 | GATCCCATGGCATAATACACATAATCAAG | Forward to get sCD_25980 from gDNA and add restriction site. To remove AAs 1-22. | 50 |
| 423 | GATCCCATGGGACATGTCTAGAAAAAAAG | Forward primer to get sCD_15220 from gDNA and add restriction site. To remove AAs 1-28 | 54 |
| 424 | GATCCTCGAGTTATATTAAAATTTTAAATTGATAAC | Reverse primers to get sCD_15220 from gDNA and add restriction site. To remove AAs 1-28 | 51.6 |
| 425 | CTCGAGCTACTTAAATACATCACTAAG | Forward primer to get sCD_13190 from gDNA and add restriction site, removing AA 1-35 | 58.9 |
| 426 | CTCGAGCTACTTAAATACATCACTAAG | Reverse primer to get sCD_13190 from gDNA and add restriction site, removing AA 1-35 | 55.2 |
| 427 | GATCCCATGGGAGGAAAAAATTAATAATG | Forward primer to get sCD_3257 from gDNA and add restriction site, removing AA 1-29 | 51.6 |
| 428 | GATCCTCGAGTTATAAGTTTTCTTGCC | Reverse primer to get sCD_32570 from gDNA and add restriction site, removing AA 1-29 | 51.7 |
| 429 | GATCCTCGAGCACCACCACCACCAC | Forward primer for inverse PCR of pET-M11 with XhoI and GATC 5' extensions | 65.7 |
| 430 | GATCCCATGGGCCCTGAAAATAAAGATTCTC | Reverse primer for inverse PCR of pET-M11 with XhoI and GATC 5' extensions | 63.4 |
| 452 | GATCCCATGGATGGATACTAATGGAAATGTTTTG | Amplify sCD_14440 from gDNA pair with primer 416 | 60.8 |
| 453 | GATCCTCGAGTTACTTTTGCACTTGCT | Reverse for sCD_25980 from gDNA without other binding sites. Pair with primer 421 | 53 |

Oligonucleotides produced in this thesis to amplify target genes from *C. difficile 630* or *C. difficile 630 Δerm* gDNA for production of expression vectors.

Table 2-5: Other oligonucleotides used in this thesis

| Name | Oligonucleotide sequence | Description | $T_m$ ($^oC$) |
|---|---|---|---|
| 211 | GATCCCATGGGATCTTATAAAAATGTAGAATTAACTGAAAAACC | Forward primer to clone spoIID$_{26-354}$ into pET-M11 with cleavable N-terminal 6xHis tag | 58 |
| 212 | GATCCTCGAGTTAGTATATATCTTTTATTTTTGTATCTGTGTAG | Reverse primer to clone spoIID$_{26-354}$ into pET-M11 with cleavable N-terminal 6xHis tag | 58.9 |
| 213 | GATCCTCGAGCTAATTTTTTTTGTTTAAAATATTCATCCAAAATC | Forward primer to clone spoIIP$_{27-339}$ into pET-M11 with cleavable N-terminal 6xHis tag | 57.4 |
| 214 | GATCCCATGGGAAATCAAGATGATTTTTTAAAGTTTTTAGTAAATTC | Reverse primer to clone spoIIP$_{27-339}$into pET-M11 with cleavable N-terminal 6xHis tag | 57.8 |
| 330 | AGAGGATGTGAGACTTATTCAAATTC | Forward primer to introduce H142R in spoIIP$_{27-339}$ | 58.5 |
| 331 | AGTATGATATATCAAAATTCTTGGATTTTC | Forward primer to introduce H142R in spoIIP$_{27-339}$ | 57.2 |
| 347 | AGAAGAGATGCTAGGGATTTAAC | Forward primer to introduce H222R in spoIIP$_{27-339}$ | 57.1 |
| 348 | TAAGTCTATCGCTATATCCACTG | Reverse primer to introduce H222R in spoIIP$_{27-339}$ | 57.1 |

These oligonucleotides had already been designed by Dr M Dembek before the start of this work and were used in the production of the associated strains.

Table 2-6: Plasmids constructed in this work

| Construct | Host plasmid | Description |
|---|---|---|
| sSpoIID | pET-M11 | Contains *SpoIID*$_{35\text{-}354}$ with a TEV-cleavable N-terminal 6xHis-tag |
| sSpoIID$_{E101A}$ | pET-M11 | Contains *SpoIID*$_{35\text{-}354}$ where glutamate 101 has been mutated to alanine via inverse PCR. Protein tagged with a TEV-cleavable N-terminal 6xHis-tag |
| sSpoIID$_{H134A}$ | pET-M11 | Contains *SpoIID*$_{35\text{-}354}$ where histidine 134 has been mutated to alanine via inverse PCR. Protein tagged with a TEV-cleavable N-terminal 6xHis-tag |
| sSpoIID$_{C140A}$ | pET-M11 | Contains *SpoIID*$_{35\text{-}354}$ where cysteine 140 has been mutated to alanine via inverse PCR. Protein tagged with a TEV-cleavable N-terminal 6xHis-tag |
| sSpoIID$_{H145A}$ | pET-M11 | Contains *SpoIID*$_{35\text{-}354}$ where histidine 145 has been mutated to alanine via inverse PCR. Protein tagged with a TEV-cleavable N-terminal 6xHis-tag |
| sSpoIID$_{C146A}$ | pET-M11 | Contains *SpoIID*$_{35\text{-}354}$ where cysteine 146 has been mutated to alanine via inverse PCR. Protein tagged with a TEV-cleavable N-terminal 6xHis-tag |
| sSpoIIP | pET-M11 | Contains a N-terminally 6xHis-tagged *spoIIP*$_{27\text{-}339.}$ Tag is cleavable by TEV |
| CD630_1319 | pET-M11 | Contains all of *CD630_13190* except the predicted transmembrane region (Amino acids 1-35). Protein produced is tagged with a TEV cleavable 6xHis-tag |
| CD630_1522 | pET-M11 | Contains all of *CD630_15220* except the predicted transmembrane region (Amino acids 1-28). Protein produced is tagged with a TEV cleavable 6xHis-tag |
| CD630_32570 | pET-M11 | Contains all of *CD630_32570* except the predicted transmembrane region (Amino acids 1-29). Protein produced is tagged with a TEV cleavable 6xHis-tag |
| CD630_14440 | pET-M11 | Contains all of *CD630_14440* except the predicted transmembrane region (Amino acids 1-35). Protein produced is tagged with a TEV cleavable 6xHis-tag |
| CD630_27240 | pET-M11 | Contains all of *CD630_27240* except the predicted transmembrane region (Amino acids 1-24). Protein produced is tagged with a TEV cleavable 6xHis-tag |

### 2.3.1. Production of expression vectors

The pET-M11 backbone was amplified from a lab stock by polymerase chain reaction (PCR) using oligonucleotides specific to pET-M11 plasmid, before 2 h digestion with DpnI [NEB] at 37 $^{o}$C to remove parent plasmid. The reaction was terminated by incubation at 80 $^{o}$C for 20 minutes before gel extraction after electrophoresis on 1 % agarose [Melford] supplemented with Sybr safe [Invitrogen] to 1X. The gel region to be excised was identified using a Safe Imager 2.0$^{TM}$ [ThermoFisher Scientific] and the plasmid extracted using a Gel extraction kit [Sigma]. The preparation was digested with XhoI and NcoI-HF in Cutsmart buffer [NEB], for 1-

2 h at 37 °C before supplementation with 1 µl Antarctic phosphatase [NEB], and Antarctic phosphatase buffer to 1X, to prevent plasmid re-circularisation. Linear plasmid was gel purified [Sigma], using the >2 % agarose gel protocol provided by the manufacturer.

Target protein sequences were submitted to TMHMM 2.0 (Moller, Croning and Apweiler, 2001), to identify possible transmembrane regions. Based on this, the target genes were amplified from *C. difficile* 630 or *C. difficile* 630 Δ*erm* genomic DNA without these predicted transmembrane regions.

For amplification of *spoIID* and *spoIIP*, reactions were set up as described in Table 2-7 using oligonucleotides designed by Dr M Dembek, as listed in Table 2-5 and thermocycling as described in Table 2-8 using KOD Hot-start polymerase [Merck]. Point mutations were introduced in plasmids encoding $SpoIID_{26-354}$ (sSpoIID) and $SpoIIP_{27-339}$ (sSpoIIP) *via* inverse PCR using the appropriate oligonucleotides (Table 2-4) and plasmid from strains PS219 or PS221, respectively (Table 2-1). All reactions were carried out as described for the production of pET-M11 *sSpoIID*.

Table 2-7: Composition of sSpoIID and sSpoIIP KOD Hot-start- based PCR reactions

| Component | Volume in reactions (µl) | Volume in negative control (µl) |
|---|---|---|
| PCR grade water | 21 | 22 |
| 10 µM forward primer | 1.5 | 1.5 |
| 10 µM reverse primer | 1.5 | 1.5 |
| pET M11 DNA | Sufficient for 10ng | 0 |
| KOD hot start master mix | 25 | 25 |

Table 2-8: KOD Hot-start- based PCR thermocycling conditions

| Stage | Time | Temperature (°C) |
|---|---|---|
| 1) Polymerase activation | 2 min | 95 |
| 2) DNA Denaturation | 20 s | 95 |
| 3) Annealing | 10 s | Primer pair dependent |
| 4) Extension | 25 or 20 s/Kb | 70 |
| 5) Termination | 10 min | 70 |

Thermocycling conditions for KOD-based PCR reactions were determined as per manufacturer's instructions. Stages 2 to 4 were repeated 30 times before progression onto step 5. An extension time of 25 s/Kb was used for templates >3000bp, and 20s/kb for templates 1000-3000bp as recommended by the manufacturer. The melting temperature was determined by the primer with the lowest annealing temperature in a primer pair.

For amplification of putative deacetylase genes, reactions were set up as described in Table 2-9 and thermocycling performed as described in Table 2-10. As Phusion polymerase [NEB] was used, the melting temperature used was adjusted by +3 °C to the lowest calculated annealing temperature of the primer pair, as suggested by the manufacturers.

Table 2-9: Composition of Phusion-based PCR reactions

| Component | Volume in reactions (µl) | Volume in negative control (µl) |
|---|---|---|
| PCR grade water | To 50 µl | To 50 µl |
| 10 µM forward primer | 2.5 | 2.5 |
| 10 µM reverse primer | 2.5 | 2.5 |
| gDNA | Sufficient for 250 ng | 0 |
| Phusion buffer | 10 | 10 |
| DMSO | 1 | 1 |
| dNTPs | 1 | 1 |
| Phusion | 0.5 | 0.5 |

Table 2-10: Phusion based PCR thermocycling conditions

| Stage | Time | Temperature ($^{o}$C) |
|---|---|---|
| 1) Denature | 30 s | 98 |
| 2) Melting | 10 s | 98 |
| 3) Anneal | 20 s | Lowest Tm of pair + 3 $^{o}$C |
| 4) Extension | 30s/Kb | 72 |
| 5) Final extension | 10 mins | 72 |

Thermocycling was undertaken as described above. Steps 2 to 4 were repeated 34 times before progressing to stage five. 30 s/Kb extensions time was recommended by the manufacturers for gDNA templates. The manufactures suggest increasing the Tm 3$^{o}$C above the lowest melting temperature of the primer pair.

The PCR products were then analysed by agarose gel electrophoresis as described above, at 100 V for approximately 40 minutes, depending on the length of the gel bed. Gels were then imaged [Bio-Rad], and the mass of the PCR product compared to that of a reference (Mass ruler [ThermoFisher scientific]), and to the expected size in base pairs of the product.

Target PCR products were combined with 2 µl each of XhoI and NcoI-HF and Cutsmart buffer [NEB] to 1X, diluted with PCR-grade water. These reactions were then incubated at 37 $^{o}$C for 1-2 h, before termination by incubation at 80 $^{o}$C for 2 minutes.

After PCR clean up [Sigma], digested gene and plasmid fragments were incubated with T4 QuickStick ligase [ThermoFisher Scientific] and incubated for 15 minutes at room temperature. NEB5-α cells were transformed with ligation reactionusing a standard transformation protocol. Briefly: NEB5α cells were aliquoted into pre-chilled Eppendorfs and incubated with the ligated vector for 30 minutes on ice at a ratio of 3:1, 5:1 or 10:1 as required to achieve transformants. Cells were heat shocked at 42$^{o}$C for 30 seconds and allowed to recover on ice for 2 minutes before addition of super optimal broth with catabolite repression (SOC) outgrowth medium [NEB]. Cells were incubated for an hour at 37 $^{o}$C with shaking before plating the entire volume on appropriately antibiotic-supplemented LB agar for overnight incubation at 37 $^{o}$C (Table 2-2).

Transformants were initially screened by colony PCR before verification by sequencing. For colony PCR, OneTaq [NEB] polymerase reactions were set up: a master mix of 2 µl T7 forward primer (Table 2-11), 2 µl of T7 reverse primer (Table 2-11), 50 µl of OneTaq 2X

Master Mix with standard buffer [NEB] and 46µl of water was split into 10 µl aliquots. Individual reactions were inoculated with one transformant, and the same clone was re-plated onto a fresh agar plate supplemented with the appropriate antibiotics. Thermocycling was undertaken as described in Table 2-12. PCR products were run on a 1 % agarose gel, as previously described. If PCR products were detected as expected, the same clone was used to inoculate an overnight culture. Plasmid DNA was extracted using a miniprep kit [Sigma], eluted in water, and sent for Sanger sequencing [Eurofins]. 1 ml of the same overnight culture was stored at -80 $^{o}$C and added to the Salgado strain collection upon confirmation by sequencing.

Table 2-11: Colony PCR oligonucleotides

| Oligonucleotide name | Sequence | Tm ($^{o}$C) |
|---|---|---|
| T7 forward | TAATACGACTCACTATAGGG | 54.3 |
| T7 reverse | GCTAGTTATTGCTCAGCGG | 57.5 |

These oligonucleotides were a kind gift of the Waldron lab.

Table 2-12: Colony PCR thermocycling conditions

| Stage | Time | Temperature ($^{o}$C) |
|---|---|---|
| 1) DNA Denaturation | 30 s | 94 |
| 2) Annealing | 30 s | 56 |
| 3) Extension | 1 min | 68 |
| 4) Final extension | 5 mins | 68 |

Colony PCR was undertaken using OneTaq polymerase [NEB] before sending plasmids for Sanger sequencing. Steps 1 to 3 were repeated 30 times before progression to step 4.

Once verified by sequencing, DNA obtained in the original miniprep was used to transform Rosetta cells [NEB], as described for NEB5-α cells. 1 ml aliquots of transformed cells were frozen in 1ml 50 % glycerol and stored in the Salgado strain collection.

### 2.3.2. Bioinformatic identification of putative deacetylases

Potential polysaccharide deacetylases in *C. difficile* were identified using the default settings on the online tool HMMER 2.3 (ebi.ac.uk/Tools/hmmer/) to search within Clostridia (tax ID: 186801) for orthologues to the known *S. pneumoniae* deacetylase *Sp*PdgA (Blair *et al.*, 2005) (PDB ID 2C1G). Results retrieved were filtered to those relating to *C. difficile* 630. This was repeated using *B. cereus* Bc1974 (Giastas *et al.*, 2018) (PDB ID 5N1J) and *B. subtilis* PdaA (Blair & van Aalten, 2004) (PDB ID 1W17), but no additional targets were identified. Using this approach, 10 putative deacetylases were identified.

From these 10 putative deacetylases, 5 were selected as potentially involved in *C. difficile* peptidoglycan deacetylation during sporulation based on known regulation by sporulation specific sigma factors (Saujet *et al.*, 2013; Dembek *et al.*, 2015), or by potential interactions with sporulation-associated proteins, as based on assessment using the STRING server (string-db.org/cgi/input.pl) (Szklarczyk *et al.*, 2019). Cloning of the genes associated with these proteins was then attempted as described in section 2.3.1.

## 2.4. Protein expression, purification and quantification

### 2.4.1. Protein quantification

Purified proteins were routinely quantified using absorbance at 280 nm as measured by Nanodrop [Thermo].

Where more accurate concentrations were required, for example in ICP-MS or CD, Bradford assays were used. 100 µl of 2-fold dilutions of samples to be tested were aliquoted into a 96-well plate [Corning] and incubated for 5 minutes with 100 µl of Pierce™ Coomassie Plus Bradford reagent [Thermo] before reading absorption at 595 nm in a spectrophotometer [BioTek]. A standard curve ranging from 0-20 µg/ml of bovine serum albumin (BSA) in diluent was used to calculate the sample protein concentration.

### 2.4.2. SDS-PAGE

Denaturing sodium dodecyl sulphate (SDS) polyacrylamide gel electrophoresis (PAGE) was used for the qualitative analysis of protein production and purity. As all proteins investigated in this study were expected to be between 10-200 kDa, 12% resolving and 5% stacking gels were used as a standard, with 15% gels occasionally used (Table 2-13). All gels were manually prepared using the Mini-PROTEAN Handcast system and ran in Mini-PROTEAN Tetra cell systems [Biorad] in Tris-Glycine running buffer (25 mM Tris base pH 8.3, 250 mM glycine, 0.1% (w/v) SDS).Samples were boiled in 1X loading buffer (100 mM Tris base pH 8.8, 4% SDS, 20% glycerol, 0.2% bromophenol blue, 200mM DTT) at 95°C for 10 minutes before loading, and ran at 200 V for 45-60 minutes, using PageRuler protein ladder [ThermoFisher Scientific] as protein molecular weight markers. Gels were stained with InstantBlue [Expedeon] before destaining in water, and imaging [Bio-Rad ChemiDoc].

Table 2-13: SDS-PAGE gel compositions

| Component | Final concentration | | |
|---|---|---|---|
| | Stacking gel (5%) | 12% gel | 15% gel |
| 30% Bis-acrylamide | 5% | 12% | 15% |
| 1.5 M Tris base pH 6.8 | 126 mM | - | - |
| 1.5 M Tris base pH 8.8 | - | 375 mM | 375 mM |
| 10% SDS | 0.1% | 0.1% | 0.1% |
| 10% APS | 0.1% | 0.1% | 0.1% |
| TEMED | 0.05% | 0.05% | 0.05% |

Bis-acrylamide, SDS, Ammonium persulphate (APS) and N,N, N', N'-tetramethylene diamine (TEMED) were used for protein separation.

### 2.4.3. Western blots

#### 2.4.3.1. sSpoIID, sSpoIIP and respective mutants

SDS-PAGE was undertaken as described above and protein transferred to Novex™ polyvinylidene difluoride (PVDF) membrane [ThermoFisher scientific], pre-equilibrated in Bjerrum-Schafer-Nielsen buffer (48 mM Tris base, 39 mM glycine pH 9.2 20% (v/v) methanol) using a semi-dry transfer method and a Bio-rad Trans-Blot Turbo system at 25V for 30 minutes. Transfer efficiency was assessed using Ponceau S [Sigma] before incubation in 1% (w/v) milk in Tris-buffered saline with Tween (TBST ) buffer (50 mM Tris base pH 7.6, 150 mM NaCl, 0.1% Tween 20) overnight at 4 °C to block non-specific binding. Membranes were then incubated in 1:15000 primary antibody raised against purified protein [Morivian Biotechnology] in 5% (w/v) milk for an hour with agitation, washed 3 times in TBST and once in Tris-buffered saline (TBS) buffer (50 mM Tris base pH 7.6, 150 mM NaCl) before incubation in 5% milk containing 1:2500 horse radish peroxidase (HRP) conjugated antibodies for 1 hour at 4 °C. Excess secondary antibody was removed by washing twice in TBST then once in TBS. Clarity Western ECL Blotting Substrate [Bio-Rad] was used to visualise antibody-bound protein as per manufacturer's instructions, before imaging using a Bio-Rad ChemiDoc XRS+ system.

#### 2.4.3.2. Deacetylase constructs

As above, 100ng of purified deacetylases were separated by SDS-PAGE on 15% gels, before transfer to a 0.45 µm pore size nitrocellulose membrane using a semi-dry transfer technique. The membranes were washed once in TBST before blocking in 5% milk (w/v) for 2 h at room temperature. Blots were incubated with primary sera (1:2500 in PBS) for 90 minutes at room temperature with agitation, before washing 3 times in TBST and once in TBS prior to incubation with secondary antibody and detection as described for sSpoIID/ sSpoIIP blots. Primary sera from rabbits inoculated with purified peptides specific to the protein of interest were produced by BioServ UK.

## 2.5. Purification of sSpoIID and sSpoIIP constructs

sSpoIID, sSpoIIP and their associated mutants, were expressed and purified routinely, for use in protein characterisation assays and X-ray crystallography trials.

### 2.5.1. Protein expression

Preliminary expression tests indicated the optimal expression conditions of SpoIID and SpoIIP. 1 L of antibiotic supplemented Terrific broth (TB) was inoculated with 10 ml overnight LB culture, grown at 37°C with agitation to $OD_{600}$ ~0.6-0.8 before supplementation with glucose to a final concentration of 1% (v/v) and induction of protein expression by the addition of isopropyl β-D-1- thiogalactopyranoside (IPTG) to 1 mM. Cultures were incubated with shaking overnight at 18 °C.

### 2.5.2. Protein purification

Cells were harvested by centrifugation at 4000 x*g* 20 minutes at 4 °C. The pellet was resuspended in lysis buffer (50 mM Tris base pH 8.0, 300 mM NaCl, 20 mM imidazole, 100 µg/ml lysozyme, 10 µg/ml DNase, 1 tablet of cOmplete EDTA-free protease inhibitor cocktail [Roche]) before sonication. Sonication was performed at 4 °C and consisted of 10 minutes of

5 seconds pulse at max speed and 5 seconds of rest on ice. The insoluble fraction was removed by centrifugation at 20,000 rpm for 30 minutes 4 °C (JA 25.50 rotor Avanti JX 26-P) and the supernatant filtered through a 0.45 µm syringe filter [Pall]. Immobilised metal ion affinity chromatography (IMAC) was used to purify His-tagged protein from the cell lysates; the filtered supernatant was applied to a 5ml HisTrap HP column [GE healthcare] pre-equilibrated in IMAC buffer A (Table 2-14) on the ÅKTA start [GE Healthcare]. His-tagged protein was eluted in 50 mM Tris base pH 8.0, 300 mM NaCl, 250 mM Imidazole in 10 ml fractions using the program in Table 2-15.

Table 2-14: Protein purification buffers

| Buffer name | Contains |
|---|---|
| IMAC A | 50 mM Tris base pH 8.0, 300 mM NaCl, 20 mM imidazole |
| IMAC B | 50 mM Tris base pH 8.0, 300 mM NaCl, 250 mM imidazole |
| SEC | 20 mM Tris base pH 8.0, 150 mM NaCl |
| Dialysis | 50 mM Tris base- pH 8.0, 300 mM NaCl |
| SEC+ EDTA | 20 mM Tris base pH 8.0, 150mM NaCl, 5 mM EDTA |

All buffers were filtered through 0.22 µm filters before use. EDTA; ethylenediamineteraacetic acid.

Table 2-15: sSpoIID/ sSpoIIP IMAC purification

| Step | Conditions |
|---|---|
| Prime and equilibration | 5 CV |
| Sample application | Sample dependent |
| Wash out unbound | 10 CV |
| Gradient elution | 0-100% IMAC B over 15 CV |

CV corresponds to column volumes, which in the case of IMAC would be 5ml. Sample application volume was determined by the volume of cell lysate applied to the column. IMAC B is 50 mM Tris base pH 8.0, 300 mM NaCl, 250 mM imidazole.

Fractions containing protein were pooled and dithiothreitol (DTT) added to a final concentration of 5mM before incubation with 1:100 Tobacco edge virus (TEV) protease to remove the 6xHis tag. Imidazole was removed by dialysing the protein against 5L of 50 mM Tris base pH8.0, 300 mM NaCl, overnight at 4 °C. Protein was recovered by reverse IMAC purification and collection of the flow through. Protein was concentrated using Amicon 30,000 Da molecular weight cut off (MWCO) spin concentrators before application to a HiLoad Superdex S200 16/600 size exclusion chromatography (SEC) column [GE Healthcare] using an ÅKTA Pure at 4 °C. The protein was eluted in SEC buffer (Table 2-14) and collected in 2ml fractions. Fractions containing protein were pooled and further concentrated and aliquoted before freezing at -80 °C.

### 2.5.3. Investigation of oligomeric state

sSpoIID and sSpoIIP were incubated on ice in a buffer containing a final concentration of 15 mM DTT. After 2 h incubation the protein was analysed by SEC using a Superdex 200 GL 10-

300 analytical column pre-equilibrated in SEC buffer (Table 2-14) supplemented with 15 mM DTT. The size of eluted protein was assessed by SDS-PAGE.

## 2.6. Deacetylase construct purification

### 2.6.1. Protein expression

One litre of auto induction media terrific broth base with trace elements (AIM TB) [For Medium] was inoculated to $OD_{600}$ 0.1 with an overnight culture of the expression strains containing pET-M11 vectors for production of CD630_3257$_{(30-275)}$, CD630_1522$_{(29-293)}$ or CD630_1319$_{(36-247)}$. Cultures were incubated with shaking (180 rpm) at 37 $^{o}$C for 5 h for expression of CD630_1522$_{(29-293)}$ and overnight CD630_3257$_{(30-275)}$, and CD630_1319$_{(36-247)}$, before centrifugation at 4,000 rpm (JA 25.50 rotor Avanti JX 26-P) for 30 minutes. Cells were resuspended in 1X PBS and centrifuged once more at 4,000 rpm for 25 minutes. The supernatant was removed, and the pellet frozen at -20 $^{o}$C.

### 2.6.2. Protein purification

Pellets were processed as described for sSpoIID/sSpoIIP purification although cells were lysed using 3 sets of 5 minutes sonication on cycle setting with power at 60-70% (Bandelin sonoplus), and the wash out unbound step of IMAC was reduced to 5 CV. SEC was undertaken at room temperature.

## 2.7. Characterisation of proteins

### 2.7.1. Circular dichroism

Samples for circular dichroism (CD) were transferred into a buffer appropriate for CD (50 mM sodium phosphate pH 8.0, 50 mM NaF for CD630_1522 constructs, 25 mM sodium phosphate for CD630_1319 constructs) by washing 3 times using 10,000 Da MWCO Vivaspin 500 concentrators [Sartorius], pre-equilibrated with the appropriate buffer.

When performing scans, buffer-exchanged proteins were diluted to approximately 0.025 mg/ml as measured by Bradford assay (section 2.4.1). A quartz cuvette of 1 mm path length [Hellma] was soaked in 1% hellmanex overnight, then rinsed in water and with the appropriate CD buffer. The cuvette was filled with buffer and scans undertaken on a Jasco J-810 spectrophotometer with a Jasco PTC-423S temperature controller. Scans configurations are detailed in Table 2-16.

Table 2-16: Circular dichroism scan parameters

| Parameter | Value for sSpoIID & mutants | Value for sSpoIIP & mutants | Value for CD630_1319 constructs | Value for CD630_1522 constructs |
|---|---|---|---|---|
| Sensitivity | Standard | Standard | Standard | Standard |
| Start | 260 nm | 260 nm | 260 nm | 260 nm |
| End | 180 nm | 180 nm | 190 nm | 185 nm |
| Pitch | 0.5 nm | 0.5 nm | 0.5 nm | 0.5 nm |
| Scan | Continuous | Continuous | Continuous | Continuous |
| Speed | 50 nm/min | 50 nm/min | 50 nm/min | 50 nm/min |
| Response | 4 s | 4 s | 4 s | 4 s |
| Bandwidth | 2 | 2 | 2 | 2 |
| Accumulation | 4 | 10 | 10 | 4 |

CD spectra were acquired as above at 20 $^{o}$C. Slightly different wavelength ranges were scanned to prevent the high tension (HT) value exceeding 600 V with certain proteins.

Protein samples were analysed, and the CD of the buffer alone subtracted from the CD of the proteins before submission to Dichroweb for analysis (Whitmore and Wallace, 2004, 2008), using the CDSSTR algorithm with reference set 4 (Sreerama and Woody, 2000). In parallel, the mean residue ellipticity was manually calculated using Equation 2-1.

Equation 2-1 Calculation of mean residue ellipticity in deg. cm$^2$.dmol$^{-1}$

$$[\Theta]_{\mathrm{mrw}} = \frac{(\mathrm{MRW} \times \Theta_{\mathrm{obs}})}{10 \times \mathrm{d} \times \mathrm{c}}$$

CD values in mdeg were retrieved from the JASCO instrument and used to calculate the mean residue ellipticity ($[\Theta]_{\mathrm{mrw}}$) in deg.cm$^2$.dmol$^{-1}$ where MRW is the mean residue weight , $\Theta_{\mathrm{obs}}$the observed ellipticity in degrees, d the path length of the cuvette in cm and c the concentration of protein in g/ml. MRW is calculated from the molecular mass/ (n-1) where N is the number of amino acids in the protein

For thermal stability analysis (thermal melts), the temperature was increased incrementally from 20 $^{o}$C to 95 $^{o}$C and the CD of the unfolding protein monitored as described in Table 2-17. Proteins were buffer exchanged and diluted to approximately 0.25 mg/ml prior to analysis. The melting temperature ($T_m$) of the proteins was calculated by calculation of the first derivative using the Savitzky-Golay filter within the Jasco software.

Table 2-17: Circular dichroism melt parameters

| Parameter | Value for sSpoIID, sSpoIIP and associated mutants | Value for CD630_1522 constructs and His1319 | Value for s1319 |
|---|---|---|---|
| Sensitivity | Standard | Standard | Standard |
| Start temperature | 20 $^{o}$C | 20 $^{o}$C | 20 $^{o}$C |
| End temperature | 95 $^{o}$C | 95 $^{o}$C | 90 $^{o}$C |
| Data pitch | 1 $^{o}$C | 1 $^{o}$C | 1 $^{o}$C |
| Bandwidth | 2 nm | 2 nm | 2 nm |
| Response | 4 s | 4 s | 4 s |
| Monitor wavelength | 222 nm | 222 nm | 222 nm |
| Temperature slope | 1 $^{o}$C/ min | 1 $^{o}$C/min | 1 $^{o}$C/min |

Protein melts were recorded at 222 nm. The melt of s1319 was undertaken between 20 $^{o}$C and 90 $^{o}$C due to technical restraints.

### 2.7.2. Inductively coupled plasma mass spectrometry

The metal content of all purified proteins was assessed by inductively coupled plasma- mass spectrometry (ICP-MS).

sSpoIID and sSpoIIP were diluted with SEC buffer to 10 µM (as determined by A280), and, where appropriate, 5 mM EDTA added. A Superdex 200 GL 10/300 column [GE] was pre-equilibrated in SEC buffer with and without 5 mM EDTA, and used to elute the proteins, ensuring that EDTA-containing samples were always separated before untreated samples, and that the column was always in the appropriate SEC buffer. 0.5 ml fractions were collected for metal content analysis.

To prepare samples for ICP-MS, 300 µl of each sample was combined with 2.7 ml of 2.5% nitric acid [Suprapur, Merck] containing the internal standards 20 ppb silver and platinum or 20 ppb indium and platinum. Samples were analysed by Drs Kevin Waldron/ Anna Barwinska-Sendra onto a Thermo x-series ICP-MS using a Cetac 900 autosampler. The ICP-MS was operating in collision cell mode using 3.0 ml/min flow of 8% hydrogen in helium as the collision gas. Samples were ionised within an argon plasma (99.99%) purity and analysed using the peak-jump method (100 sweeps, 20-30 MS dwell time n 3-5 channels per isotope, separated by 0.02 atomic mass units (AMU). Ions of Mg, Mn, Cu, Co, and Zn were quantified by comparison of the number of ions hitting the detector to a standard curve containing known numbers of the aforementioned ions. Two biological replicates per construct were analysed.

Soluble constructs of s1319, s1522, where the 6xHis tag had been removed, were also analysed as above. Samples were not treated with EDTA, but the column, injection port and loop were washed in SEC + EDTA buffer, before re-equilibration in SEC buffer.

### 2.7.3. Estimation of molecular weight

Analytical SEC allows the estimation of the molecular weight of a protein based on elution volume. A collection of protein standards with known molecular weights were used to determine the calibration curve of the size exclusion column. The equation describing the calibration curve, as per Equation 2-2 , is then used to calculate the mass of the protein of interest. Calibration curves at room temperature were performed by Dr Anna Barwinska-Sendra, and at 4°C by Paola Lanzoni-Manguchi.

Equation 2-2: Calculation of $K_{av}$

$$K_{av} = \frac{V_e - V_o}{v_c - v_o}$$

$K_{av}$; partition coeffiecient and can be related to the logMW of a protein via a calibration curve, $V_e$ is the elution volume of the protein of interest, $V_o$ is the void volume of the column, and $V_c$ is the volume of the column. All volumes are in ml.

### 2.7.4. Determination of accurate mass of purified putative deacetylases

The accurate masses of purified putative deacetylases were determined by Dr Joe Gray. Briefly; pure protein preparations were acidified to final concentration of 0.5% formic acid which was then injected into a CapTrap column [nanoLCMS Solutions] and eluted using a NanoAquity HPLC [Waters] at 0.05ml/min. Initial condition was 5% buffer B (0.1% formic acid in acetonitrile) in buffer A (0.1% formic acid in water) with elution over a gradient to 70% buffer B over 22 minutes followed by a step of 100% buffer B over 5 minutes before finally returning to 5% buffer B. Eluate was directed into an Impact II [Bruker] mass spectrometer using the standard Apollo electrospray ionisation ion source with a capillary voltage set at 2800V, temperature 180°C, 5L/min drying gas with the nebuliser spray gas pressure set to 0.8 bar. Data was collected over *m/z* 200-2500. Spectra were analysed using Compass DataAnalysis package [Bruker] with "Maximum entropy" charge deconvolution plug-in (parameters: adduct ions $H^+$, low mass 5000 *m/z*, high mass 100000 *m/z* with resolving power at 40,000 and resolution to high).

## 2.8. Structure determination

### 2.8.1. Protein crystallisation

Purified samples of sSpoIID, sSpoIIP and the respective mutated versions at a concentration between 8-15 mg/ml were used to set up commercially available crystallisation screens: Morpheus, JCSG+, Structure [Molecular Dimensions], PACT [Qiagen], and Index [Hampton Research]. Sitting drop screens were set up using a Mosquito dispensing robot [TTP Labtech] in 1:1 (100nl:100nl) and 2:1 (200nl:100nl) protein to reservoir ratios. Trays were stored at 20°C and checked for the presence of crystals using a 16X optical zoom light microscope [Leica].

$sSpoIID_{E101A}$ crystals appeared within 24-48 h in Index C6 (60% Tacsimate™ pH 7.0) and were harvested by Dr Arnaud Baslé and cryoprotected in paraffin oil before flash freezing in liquid nitrogen until diffraction experiments at Diamond Light Source (DLS) synchrotron.

#### 2.8.2. Data collection and processing for sSpoIID$_{E101A}$

Data were collected over 200° with images recorded every 0.1° with 0.4 second exposure time at beamline I04-1 at Diamond Light Source. Data collection was performed by Dr Arnaud Baslé as part of the Newcastle-Durham-York beamtime allocation group.

The DIALS user interface (DUI) was used for spot detection, indexing and integration and refinement. These steps were repeated until the signal to noise ratio (I/σI) was approximately 1.5 in the lowest resolution shell. Data was then scaled within DUI.

Scaling was carried out using the Aimless pipeline (Evans, Murshudov and IUCr, 2013), within CCP4i2 (Potterton *et al.*, 2018), followed by cell content estimation.

#### 2.8.3. Structure determination of sSpoIID$_{E101A}$

Molecular replacement was undertaken within CCP4i2 using Phaser (McCoy *et al.*, 2007) with the SpoIID structure (PDB ID: 5TXU) as a search model (Nocadello *et al.*, 2016).

The resulting maps and model were refined using REFMAC5 (Vagin *et al.*, 2004) followed by manual model building in Coot (Emsley *et al.*, 2010), in iterative cycles until $R_{free}$ and $R_{merge}$ converged. A final refinement step and model validation was undertaken in Phenix (Adams *et al.*, 2010).

#### 2.8.4. Thin layer chromatography

Thin layer chromatography (TLC) was undertaken on sSpoIID$_{E101A}$ by Dr Lucy Crouch according to standard protocols (Briliūtė *et al.*, 2019). Briefly, a sample was spotted onto silica plates before separation in a 2:1:1 butanol/ acetic acid/ water ratio, followed by staining with diphenylamine-aniline-phosphoric acid stain (1 ml of 37.5% HCl, 2 ml of aniline, 10 ml of 85% $H_3PO_3$, 100 ml of ethyl acetate and 2 g diphenylamine).

### 2.9. Peptidoglycan isolation and modification

#### 2.9.1. *E. coli* peptidoglycan

*E. coli* BW25113 Δ*lpp*, a strain of *E. coli* lacking the lipoprotein Lpp (Baba *et al.*, 2006) was used for peptidoglycan purification (Baba *et al.*, 2006). Isolation was largely carried out as previously described (Glauner, 1988; Dembek *et al.*, 2018). Briefly, cells were grown in LB to $OD_{600}$ ~0.8 before centrifugation at 5000 rpm 4 °C (JA 25.50 rotor Avanti JX 26-P). Pellets were resuspended in ice-cold MilliQ and added dropwise to boiling and stirring 8% HPLC-grade SDS and allowed to boil for 30 minutes. Warm water washes were used to remove SDS until the preparation was SDS-free as determined by the Hayashi test (Hayashi, 1975)., 335 µl of supernatant was combined with 7 µl 0.5% methylene blue, 170 µl 0.7 M sodium phosphate pH 7.2 and 1ml of chloroform before vortexting. A pink/ colourless choloroform layer indicates SDS removal, whereas a blue layer indicates SDS is still present in the supernatant. Pellets were then suspended in 10 mM Tris-Cl pH 7.0 10 mM NaCl, with amylase to 1 mg/ml final concentration. After 2 h incubation at 37 °C, pronase E was added to 0.1 mg/ml and incubated for a further hour. SDS was added to a final concentration of 2% and samples were boiled for a further 15 minutes. SDS was removed as above, and the resulting PG stored in 0.02% sodium azide at 4 °C.

### 2.9.2. *C. difficile* peptidoglycan

#### 2.9.2.1. Growth conditions

Peptidoglycan from *C. difficile* 630 Δ*erm*, *C. difficile* 630 Δ*erm* $P_{tet}$*spo0A*, *C. difficile* 630 Δ*erm* Δ*spoIID*, *C. difficile* 630 Δ*erm* Δ*spoIIP*, *C. difficile* 630 Δ*erm* Δ*spoIIQ*, *C. difficile* 630 Δ*erm* Δ*spoIIIAH*, *C. difficile* 630 Δ*erm* Δ*sigE* and *C. difficile* 630 Δ*erm* Δ*sigF* was extracted as below. Overnight BHIS cultures of each strain were used to inoculate pre-reduced BHIS 1:200, which were allowed to grow to early stationary phase ($OD_{600}$ ~0.7-1.6) before being used to inoculate 400 ml pre-reduced sporulation media (SM). SM cultures were incubated anaerobically at 37 ºC overnight.

#### 2.9.2.2. Cell wall purification

Cells were pelleted by centrifugation at 4000 ×*g*, at 4 ºC for 25 minutes and resuspended in 50mM Tris-Cl pH 7.0. Cell suspension was added dropwise to boiling 8% SDS and allowed to boil for 15 minutes. Once cooled, the mixture was washed twice in 1 M NaCl 12,000 ×*g* 30 minutes [JA25-50 Beckman coulter], and then with warm MilliQ until free of SDS, as judged by a lack of foaming during resuspension. Once free of SDS, cells were lysed using a Precellys cell disruptor [Bertin Instruments], in 2 ml tubes filled 1/3 with acid washed glass beads (150-212µm [Sigma]) on a cycle of 3 seconds pulse at max speed (setting 6.5) and 3 seconds of rest. Preparations were cooled on ice between sets. Disruption was repeated until cells were judged to be lysed by light microscopy (100X oil objective Nikon Ti). Beads were removed by filtering samples through a glass frit, and beads were washed with MilliQ. The filtrate was centrifuged 2000 ×*g* 5 minutes and the supernatant reserved before the pellet was washed in warm MilliQ water (2000 ×*g* 5 minutes). The supernatants containing broken cells was then centrifuged for 30 minutes at 25,000 ×*g* at room temperature. The pellet was resuspended in 100 mM Tris-Cl pH 7.5 20 mM $MgSO_4$ and incubated at 37 ºC with stirring for 2 h with DNaseI (10 µg/ml) and RNase (50 µg/ml). $CaCl_2$ was added to 10 mM and trypsin to 1 µg/ml and incubated overnight at 37 ºC with stirring. SDS was added to a final concentration of 1% and the preparation was incubated at 80ºC 15 minutes, then centrifuged 25,000×g 30 minutes and resuspended in 8 M $LiCl_2$ before incubation at 37 ºC 15 minutes. Pellets were then washed twice in MilliQ. Samples were frozen at -80 ºC overnight, before freeze drying [Christ freeze drier].

#### 2.9.2.3. Purification of peptidoglycan from cell walls

Freeze dried cell wall (5 mg) was incubated at 4 ºC for 48 h with 2700 µl hydrofluoric acid. The samples were then centrifuged, and waste HF removed before re-filling the tubes with ice water and centrifuging again. The pellet was then washed once with water, once with 100 mM Tris-Cl pH 7.0, and twice again with water. Recovered peptidoglycan was stored in 0.05% sodium azide at 4 ºC.

### 2.9.3. Chemical acetylation of peptidoglycan

Aliquots of purified *E. coli* and *C. difficile* vegetative cell peptidoglycan were chemically acetylated using a method based on that described previously (Vollmer and Tomasz, 2000). One volume of peptidoglycan was combined with 0.25 volumes saturated sodium bicarbonate and 0.25 volumes 5% acetic acid, freshly prepared from acetic anhydride. After 30 minutes incubation with stirring at 4 ºC, a further 0.25 volumes of acetic acid was added and incubated as described again, then at room temperature for one hour. Peptidoglycan was recovered and washed 3 times with water by centrifugation at 90,000 rpm for 30

minutes [Hitachi Himac CS150NX with TLA 100.3 rotor]. The resulting acetylated peptidoglycan was stored in 0.02% sodium azide at 4 °C until use in digestion experiments. Acetylation was confirmed by mass spectrometry; in *E. coli*, *meso*-diaminopimelic acid (*meso*DAP) was acetylated, and in *C. difficile*, both the *meso*DAP and glucosamine normally present were acetylated.

## 2.10. Peptidoglycan digestion assays

### 2.10.1. sSpoIID/ sSpoIIP digests of *E. coli* peptidoglycan

Peptidoglycan (10 µl) purified from *E. coli* BW25113 Δ*lpp* (as described in section 2.9.1), was incubated in peptidoglycan digestion buffer (10 mM Hepes pH 7.2, 50 mM NaCl, 0.05% Triton X-100, 1 mM $ZnCl_2$), with a final concentration of sSpoIID/ sSpoIIP (as purified in section 2.4.3.1) of 10 µM, for 24 h with stirring at 37 °C. Total reactions were supplemented with cellosyl buffer (80 mM sodium phosphate pH 4.8) and cellosyl [kindly provided by Hoechst, Frankfurt, Germany] to a final concentration 0.04 mg/ml and incubated overnight at 37 °C with stirring. Reactions were terminated by boiling at 100 °C for 10 minutes before drying (ScanVac). Dried samples were reconstituted in 0.25 M sodium borate buffer pH 9.0 and water before reduction with sodium borohydride. Samples were reduced by the addition of a small amount of solid sodium borohydride in sodium borate buffer pH 9.0. Reactions were terminated and acidified to pH 3-4 using 20 % HPLC-grade phosphoric acid. Samples were then briefly centrifuged, and the supernatant frozen at -20 °C until high pressure liquid chromatography (HPLC) and mass spectrometry (MS).

#### 2.10.1.1. Liquid chromatography mass spectrometry of sSpoIID/ sSpoIIP digestions

Samples prepared as above were analysed by liquid-chromatography mass spectrometry (LC-MS) aided by Dr Joe Gray. Samples were diluted 10-fold in 0.2 % formic acid before injection of 20µl into the HPLC-system (Agilent 1100, using an ACE3 C-18AQ column (2,1 x 150 mm)). Reverse phase HPLC (RP-HPLC) was undertaken at 0.2 ml/min in buffer A (0.1 % TFA in water), to a maximum of 85 % buffer B (0.1 % TFA in acetonitrile) over 80 minutes.

Eluate from RP-HPLC was directed to a LTQ-FT mass spectrometer [Thermo]. Electrospray ionisation voltage was set to 4.2 kV and the transfer capillary temperature to 250 °C. Mass spectra were collected between *m*/*z*= 150-2000 with MS2 fragmentation spectra triggered for all ion signals >$5x10^3$ intensity.

### 2.10.2. sSpoIID/ sSpoIIP digests on chemically acetylated peptidoglycan

Untreated and treated *E. coli* BW25113Δ*lpp* and *C. difficile 630* Δ*erm* peptidoglycan (as described in 2.9.3) were digested largely as described in section 2.10.1. Where *C. difficile* peptidoglycan was used, 30 µl of purified peptidoglycan was used.

#### 2.10.2.1. Liquid chromatography mass spectrometry of acetylated peptidoglycan digests

LC-MS was undertaken as described in section 2.10.1.1 with the following modifications. An Agilent 1100 HPLC system was used with an ACE3 C-18AQ 1X150mm column at 35 °C. After an 8 minute divert time, the muropeptide mixture was ionised using an IonMax ionisation source in positive ion mode, sheath gas set to 4, spray voltage 4.2 kV with capillary Temp 200 °C. Parent scans from the IonTrap were acquired between 300 and 2000 *m/z*, UltraZoom scans were undertaken on the most intense ion in the parent scan (±5 Da), MS2 was then

triggered within 2.5 *m/z*, normalised collision energy=35 within minimum signal threshold set to 5000 counts.

#### 2.10.3. Digests with putative deacetylases

Digests were performed largely as described for sSpoIID/sSpoIIP digestion assay with the following alterations. Samples were incubated with 30 µl of *E. coli* Δ*lpp* peptidoglycan and using the soluble proteins purified as described in section 2.6. Samples were reduced by the addition of 100ul 1M sodium borohydride in sodium borate pH 9.0 and acidified using 20 % formic acid.

LC-Ms was undertaken as described in section 2.10.2.1

### 2.11. Structural analysis of peptidoglycan

All mass-spectra analyses were aided by Dr Joe Gray.

#### 2.11.1. Liquid chromatography mass spectrometry - Linear trap quadrupole Fourier transform (LTQ-FT)

Peptidoglycan purified from *C. difficile* 630 strains: Δ*erm* Δ*spoIIQ,* Δ*erm* Δ*spoIID,* Δ*erm* Δ*spoIIP*¸ Δ*erm* Δ*spoIIIAH,* Δ*erm* Δ*pyrE* $P_{tet}$ *spo0A,* Δ*erm* Δ*sigE and* Δ*erm* Δ*sigF* (as described in section 2.9.2), was digested by the addition of cellosyl (0.01 mg ) in cellosyl buffer for 24 h with stirring at 37 $^{o}$C. After 24 h, a further aliquot of cellosyl (0.01 mg) was added and incubated for another 24 h. Cellosyl digestion was terminated by boiling for 10 minutes. After equilibration to room temperature, sodium borate buffer pH 9.0 with a few sodium borohydride crystals was added and the reaction was reduced for 30 minutes. Reduction reactions were terminated and acidified using 20% HPLC-grade phosphoric acid.

Reduced muropeptides were diluted 10-fold in 0.1% trifluoroacetic acid to give a final volume of 20 µl in total recovery autosampler vials [Waters]. After centrifugation at 15000 rpm for 5 minutes, 2 µl were injected onto the NanoAcquity ultra-performance liquid chromatography (UPLC) system [Waters] fitted with an ACE3 C18 0.5X150 mm column [HiChrom]. A gradient of 0.1% formic acid (buffer A) and 0.1% formic acid in acetonitrile (buffer B) with a flow rate of 50 µl/min and a column temperature of 35 $^{o}$C was used.

The initial 5 minutes of UPLC flow was diverted to waste, with the remainder of the eluted muropeptides subjected to LQT-FT Mass spectrometry [Thermo]. Muropeptides were ionised by electrospray ionisation at 1.4kV and analysed at a resolution of 100,000. A second fragmentation (MS2) was set to occur if any ion was at charge state 3 or lower using automated ion charge state screening. Data-dependent neutral loss (DDNL) was set so that if an ion corresponding to the parent ion without any of the fragments described in Table 2-18 was observed in the MS2, a further fragmentation (MS3) would be triggered. All spectra were analysed using the manufacturer's Qualbrowser software [Thermo].

Table 2-18: Data dependent neutral loss parameters

| Species | Charge state | m/z |
|---|---|---|
| Mur*N*Ac | 1+ | 277.116 |
| | 2+ | 137.550 |
| Mur*N*Ac (non reduced) | 1+ | 257.089 |
| Glc*N*Ac | 1+ | 203.039 |
| | 2+ | 101.540 |
| GlcN | 1+ | 161.068 |
| Glu | 1+ | 129.043 |
| Ala | 1+ | 71.037 |
| Gly | 1+ | 57.021 |

If the loss of one of the above species was observed in the MS2, MS3 would be triggered.

### 2.11.2. Development of an MZmine2 database

The masses and identities used to develop the MZmine2 (Pluskal *et al.*, 2010) database were ascertained by manually assessing every spectrum that triggered an MS3 scan.

Initially, the neutral mass of the parent ion was calculated, and ions corresponding to expected losses (Table 2-19) were followed, elucidating the structure of the muropeptide as it fragmented. When no expected masses could be observed, Chemdraw [Perkin Elmer] was used to build the remaining mass to account for the entire neutral mass. Where identities could not be solved, the entry was removed from analysis. Chemdraw was also used for chemical formula and expected neutral mass determination.

Table 2-19: Approximate expected losses in MS analysis

| Ion | +1 charge (*m/z*) |
|---|---|
| Glc*N*Ac | 202 |
| GlcN | 160 |
| Mur*N*Ac(reduced) | 277 |
| Mur*N*Ac(non-reduced) | 275 |
| 1,6 anhydro-Mur*N*Ac | 185 |
| Ala | 71 |
| Ala(OH) | 89 |
| Glu | 129 |
| *meso*DAP | 172 |
| Gly | 57 |
| Gly(OH) | 75 |

Losses expected when analysing mass spectra. Ala(OH) and Gly(OH) correspond to Ala and Gly residues found on the termini of peptide stems.

### 2.11.3. Mzmine2-powered analysis of LTQ-FT data

MZmine2 version 2.37 (Katajamaa, Miettinen and Oresic, 2006; Pluskal *et al.*, 2010) was used in conjunction with the aforementioned custom database to analyse the LTQ-FT data. Raw data files were imported into MZmine2, MS1 and MS2 scans were detected, a chromatograph built from this data, and the chromatograph deconvoluted. Isotope peaks

were identified, chromatograph of independent samples aligned, and peaks identified using the custom database. Results were then filtered to include only rows with at least one peak and two isotopes. Parameters for all steps are available in Table 2-20.

The 10 most abundant peaks in *C. difficile* 630 Δ*pyrE* Δ*erm* $P_{tet}$ *spo0A* (referred to as wild type in these experiments) were identified as peaks of interest, and the area under the curve (AUC) for these peaks extracted for each sample and normalised to the sporulation efficiency of that strain. The AUC of all 10 muropeptides of interest was summed to give the total AUC for each sample. The AUC accounted for by a given muropeptide was then expressed as a percentage of the AUC accounted for by all 10 muropeptides.

Table 2-20: MZmine2 processing parameters

| Step | Parameter | Value |
|---|---|---|
| Peak detection 1 | Retention time | 0.00-82.02 |
| | MS level | 1 |
| | Polarity | + |
| | Spectrum type | Profile |
| | Mass detector | Exact mass |
| | Noise | 1E3 |
| Peak detection 2 | Retention time | 0.00-82.02 |
| | MS level | 2 |
| | Polarity | + |
| | Spectrum type | Centroided |
| | Mass detector | Centroid |
| | Noise | 1E1 |
| Chromatogram builder | MS level | 1 |
| | Polarity | + |
| | Spectrum type | Any |
| | Time span | 0.2 min |
| | Minimum height | 2.5E3 |
| | *m/z* tolerance | 0.0 *m/z* or 20.0 ppm |
| Deconvolution | Local minimum search | 15.0% chromatographic threshold<br>Search minimum in RT range (min) 0.05<br>Minimum relative height (5.0%)<br>Minimum absolute height (1E3)<br>Minimum ratio of peak top/edge 1.5<br>Peak duration range (min) 0.00-2.00 |
| | *m/z* centre calculation | Median |
| | *m/z* range for MS2 scan pairing | 0.02 Da |
| | RT range for MS2 scan pairing | 0.1 min |
| Isotopic peak grouper | *m/z* tolerance | 0.0 *m/z* or 20 ppm |

| Step | Parameter | Value |
|---|---|---|
| | RT tolerance | 0.2 absolute min |
| | Monotonic shape | Unchecked |
| | Maximum charge | 4 |
| | Representative isotope | Lowest *m/z* |
| Alignment- join aligner | *m/z* tolerance | 0.05 *m/z* or 0 ppm |
| | Weight for *m/*z | 75 |
| | RT tolerance | 1.0 absolute minutes |
| | Weight for RT | 25 |
| | Require same charge state | No |
| | Requires same ID | No |
| | Compare isotope pattern | No |
| Identification- custom database | Database file | |
| | *m/z* tolerance | 0.05 *m/z* or 0 ppm |
| | RT tolerance | 1.0 absolute minutes |
| Filter- peak list rows | Minimum number of peaks in row | 1 |
| | Minimum peaks in isotope pattern | 2 |
| | *m/z* | Unchecked |
| | RT | Unchecked |
| | Peak duration | Unchecked |
| | Chromatographic FWHM | Unchecked |
| | Parameter | No parameter identified |
| | Only identified? | Unchecked |
| | Text in identity | Unchecked |
| | Text in comment | Unchecked |
| | Keep or remove rows | Keep rows that match criteria |
| | Keep only peaks with MS2 scan (GNPS) | Unchecked |
| | Reset the peak number ID | Unchecked |
| | Remove source peak list after filtering | Unchecked |

MZmine2 parameters used during automated processing of the peptidoglycan of various *C. difficile* mutants.

# Chapter 3. Characterisation of *C. difficile* peptidoglycan throughout engulfment

## 3.1. Introduction

As *C. difficile* is an anaerobic organism it must undergo radical changes to its physiology whilst forming spores to permit transmission through the aerobic environment. Once the asymmetric septum is formed and the cell is committed to sporulation, the mother cell must engulf the forespore to produce a cell-within-a-cell structure. This process requires remodelling of the cell wall in such a manner that the integrity of the cell is not prematurely compromised.

The cell wall of Gram-positive bacteria contains of a thick layer of peptidoglycan, that must be constantly synthesised and digested to allow vegetative growth and cell division (Dik *et al.*, 2017). Peptidoglycan is comprised of long glycan strands, canonically alternating *N*-acetylglucosamine (Glc*N*Ac) and *N*-acetyl-muramic acid (Mur*N*Ac) residues. These long glycan strands are then connected by short peptides, typically tetra or tripeptides, that are crosslinked between the amino acids at the third and fourth position, or between two third position amino acids (Sütterlin *et al.*, 2017).

*C. difficile* vegetative cell peptidoglycan characterisation revealed several differences when compared to the Gram-positive model organism *B. subtilis* and the Gram-negative model organism *E. coli*. Whereas *E. coli* peptidoglycan exclusively contains Glc*N*Ac, both *B. subtilis* and, to a greater extent, *C. difficile* consist of deacetylated Glc*N*Ac (GlcN): 17% of *B. subtilis* Glc*N*Ac residues are deacetylated, where 93% are deacetylated in vegetative *C. difficile* peptidoglycan (Atrih *et al.*, 1999; Peltier *et al.*, 2011). Glc*N*Ac deacetylation has been observed in the peptidoglycan of other species such as *Bacillus cereus* (Araki *et al.*, 1971), *Streptococcus pneumoniae* (Ohno, Yadomae and Miyazaki, 1982; Vollmer and Tomasz, 2000) and *Lactobacillus fermentum* (Logardt and Neujahr, 1975). The degree of deacetylation has been related to both physiological conditions (Balomenou *et al.*, 2013) and as a response to suboptimal conditions, for example oxidative stress induced deacetylation in *Helicobacter pylori* (Wang *et al.*, 2010).

In *C. difficile*, 3-3 crosslinks account for 73% of crosslinks, in contrast to the predominance of 4-3 crosslinks in *E. coli* (Peltier *et al.*, 2011; Sütterlin *et al.*, 2017).

The peptidoglycan found in the spore cortex is substantially different to that of the vegetative cell. Every second Mur*N*Ac residue is replaced by a MAL residue, which is unable to carry stem peptides. Consequently, crosslinking in C*. difficile* spore peptidoglycan is reduced from 33.8% (Peltier *et al.*, 2011) in vegetative cells to 4.9% (Coullon *et al.*, 2018). The degree of deacetylation is also altered, with a reduction from ~93% (Peltier *et al.*, 2011) in vegetative cells to ~55% (Coullon *et al.*, 2018) in the spore cortex peptidoglycan.

*C. difficile* sporulation is controlled by a series of sigma factors under the control of the master regulator Spo0A (Saujet *et al.*, 2013). The mother cell sigma factor $\sigma^E$ and the forespore sigma factor $\sigma^F$ are under the direct control of *spo0A* (Piggot and Hilbert, 2004; Saujet *et al.*, 2013; Pettit *et al.*, 2014; Ribis, Fimlaid and Shen, 2018), with evidence that $\sigma^F$ does stimulate $\sigma^E$ in *C. difficile*, but is not strictly necessary (Pishdadian, Fimlaid and Shen,

2015). SpoIID and SpoIIIAH are produced in the mother cell in response to $\sigma^E$, with $\sigma^F$ controlling the expression of *spoIIP* and *spoIIQ* in the forespore (Saujet *et al.*, 2013). Of note is the differential expression of SpoIIP; whilst thought to be confined to the forespore in *C. difficile*, SpoIIP is expressed in the mother cell in *B. subtilis* (Ribis, Fimlaid and Shen, 2018), with so far unknown implications to function and activity.

We sought to characterise the changes to peptidoglycan throughout engulfment; from the crosslink rich vegetative cell, to the MAL containing spore peptidoglycan. Mutant strains carrying deletions $\sigma^E$ and $\sigma^F$, as well as each of the known engulfment proteins (DMP and SpoIIQ and SpoIIIAH (Q:AH)) were used. These strains allow a degree of control of sporulation; each strain is sporulation defective and halted at a different stage of sporulation due to the absence of either the sigma factor or specific proteins. The wild type strain has the master regulator of sporulation, *spo0A*, under the control of an inducible promoter, therefore, in the absence of inducer, no cells will sporulate (Dembek *et al.*, 2017). The peptidoglycan of these "arrested", sporulation-stalled mutants was characterised by high resolution HPLC-MS.

### 3.1.1. Identification of muropeptides

All *C. difficile* strains used in this analysis were produced and characterised by Dr Marcin Dembek. For analysis of the composition of peptidoglycan in vegetative cells, we used a strain with *spo0A* under an inducible promoter (*C. difficile* 630 Δ*pyrE*Δ*erm* $P_{tet}$-*spo0A*) in the absence of the inducer anhydrotetracycline to ensure cells do not initiate sporulation. As such, the peptidoglycan analysed will originate exclusively from vegetative cells, unlike that in previous studies (Peltier *et al.*, 2011; Bern, Beniston and Mesnage, 2016) where it is possible some cells had already initiated sporulation.

Peptidoglycan was purified as described in section 2.9.2 and digested with the muramidase cellosyl, generating muropeptides for separation by HPLC and analysis by Fourier transform mass spectrometry (FT-MS) (section 2.11). HPLC separates the muropeptides (MP) based on their hydrophobicity, with monomers generally eluting before crosslinked species. After the parent scan, a second level of mass spectrometry was completed on the ion of interest (MS2), and, if particular ions were observed, a further level of MS (MS3) was undertaken. All ions triggering MS3 were manually analysed. Those that could be identified as muropeptides formed a database consisting of 49 unique muropeptides in several charge states with various adducts. This database was then used for automatic analysis (section 3.2) of muropeptides originating from sporulation-stalled *C. difficile* mutants. Muropeptides determined to be formed due to in-source decay during ionisation (i.e. those missing only GlcN or Mur*N*Ac residues) were excluded from the database.

To identify a muropeptide, the neutral mass of the ion of interest was calculated and any possible adducts were identified using the parent scan (MS1) (panel B Figure 3-1). For example, the ion 866.8774 triggered MS2 and MS3 scans (black filled triangle in hashed box of Figure 3-1), but there is a more dominant ion in the spectra at 855.8856. These two ions are related by a mass:charge (*m/z*) difference of approximately 10.99, suggesting that 866.8774 is the doubly charged sodium adduct of 855.8856. With this in mind, the neutral

mass of both the sodium adduct (the mass spectra of which we used to identify the muropeptide), and the protonated ion (855.8856) were calculated.

The first component of the muropeptide is deduced by subtracting the first major ion seen in the MS2, in this case 1571.3248, from the calculated neutral mass of the ion (1731.74024 Da), giving a difference of 160.4, approximately the predicted mass of a GlcN residue (~161). From there, the next ion observed (marked II in panel B, Figure 3-1) also corresponds to a GlcN loss. This is followed by the ion marked III which corresponds to the loss of a Mur*N*Ac residue. At this point we can now see this ion (~1133 *m/z*) in the MS3 (Figure 3-1D), and therefore move to using the MS3 for structure determination, as this often provides more detail. Ion IV corresponds to an in-line alanine (that is, an alanine that is not at the end of a peptide stem represented by the ion 1066 *m/z*), ion V to a glutamic acid, VI a *meso*DAP, VII a terminal alanine, VIII a second *meso*DAP residue, IX a second glutamic acid residue, and finally ion X, an in-line alanine residue. This then leaves a deficit that can be accounted for by a Mur*N*Ac residue and the mass of the sodium adduct. The accurate mass of this structure can be calculated at 1709.75684 Da, with the theoretical mass 1709.7466 Da, demonstrating the accuracy and power of FT-MS.

A similar strategy to the one described above was used for all muropeptides identified in this work. However, in some cases it was necessary to use the mass deficit characteristic profiles in terms of *C. difficile* peptidoglycan in order to assign a putative identity to these muropeptides.

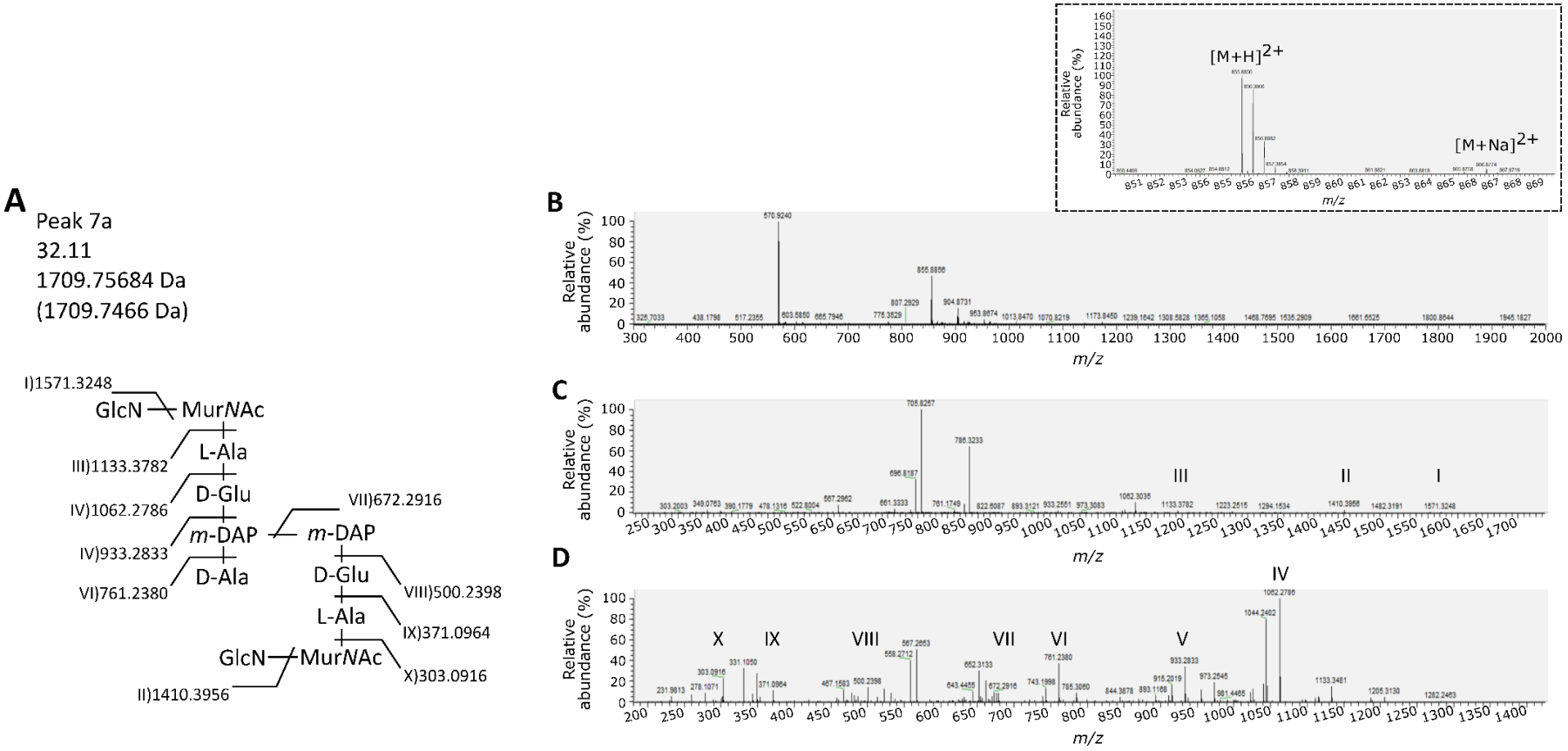


Figure 3-1: Identification of muropeptides from mass spectra.

Muropeptides were separated by HPLC with the eluate injected into the FT-MS. Parent scans (panel B) are used to calculate the neutral mass of the ion of interest. The MS2 (panel C) and MS3 (panel D) are used in conjunction with expected ions to determine the identity of the muropeptide of interest. In this example, the parent ion at ~866 *m/z* (marked $[M+Na]^{2+}$), dashed box) is a sodium adduct of the ion at ~855 *m/z* ($[M+H]^{2+}$), therefore the neutral mass of the 855 ion is calculated in addition to that of the sodium adduct. An ion corresponding to a GlcN residue is observed as at ~1571 in the MS2. Ion II corresponds to a second GlcN and ion III to a Mur*N*Ac residue. At this point the MS3 is used. Ions VI-X correspond to Ala, Glu, *meso*DAP, terminal Ala, *meso*DAP, Glu and a final Ala residue. The remaining mass can be accounted for by the presence of a Mur*N*Ac residue, and a sodium ion. Thus, the identity is determined as GlcN Mur*N*Ac AEmA AEm GlcN Mur*N*Ac.

### 3.1.2. Major muropeptides identified in *C. difficile* peptidoglycan

The ten most abundant muropeptides were designated as the major muropeptides and identified as described in section 3.1.1. The observed neutral masses, identities and the percentage of the area under the curve (AUC) each muropeptide accounts for are shown in Table 3-1, with the spectra and muropeptide structures shown in Figure 3-2. Muropeptides 1 and 2 are considered monomers, muropeptides 3-9 dimers, and muropeptide 10 a trimer.

Table 3-1: Major muropeptides of interest in *C. difficile* 630 Δ*pyrE* Δ*erm* $P_{tet}$ -*spo0A*

| MP | Retention time (mins) | Observed neutral mass (Da) | Identity | Cross link type | AUC (%) |
|---|---|---|---|---|---|
| 1 | 18.67 | 885.3815 | GlcN Mur*N*Ac AEmG | N/A | 5.6 |
| 2 | 21.94 | 899.3971 | GlcN Mur*N*Ac AEmA | N/A | 6.6 |
| 3 | 25.43 | 1271.5616 | GlcN Mur*N*Ac AEmA AEm | 3-3 | 1.8 |
| 4 | 29.15 | 1695.7309 | GlcN Mur*N*Ac AEmG AEm | 3-3 | 10.7 |
| 5 | 30.26 | 1638.7095 | GlcN Mur*N*Ac AEm AEm GlcN Mur*N*Ac | 3-3 | 5.0 |
| 6 | 31.11 | 1766.7681 | GlcN Mur*N*Ac AEmA AEmG GlcN Mur*N*Ac | 3-3 | 2.2 |
| 7a/b | 32.11/34.45 | 1709.7466 | GlcN Mur*N*Ac AEmA AEm GlcN Mur*N*Ac | 3-3 | 53.9 |
| 8 | 34.05 | 1751.7572 | Glc*N*Ac Mur*N*Ac AEmA AEm GlcN Mur*N*Ac | 3-3 | 3.1 |
| 9 | 35.78 | 1780.7837 | GlcN Mur*N*Ac AEmA AEmA GlcN Mur*N*Ac | 4-3 | 10.7 |
| 10 | 38.62 | 2520.0961 | GlcN Mur*N*Ac AEmA AEm AEm GlcN Mur*N*Ac GlcN Mur*N*Ac | 3-3-3 | 0.4 |

Muropeptides derived from *C. difficile* 630 Δ*pyrE* Δ*erm* $P_{tet}$-*spo0A* were separated by HPLC and analysed by FT-MS. The top ten muropeptides (MP) were identified and are listed in terms of retention time during HPLC. The observed mass was calculated and the MS2 and MS3 used to assign an identity to the muropeptide, as described in section 3.1.1. For dimers and trimers, crosslinks were predominantly seen between amino acids at the third position, though muropeptide 9 was linked between amino acids at the third and fourth positions. The area under the curve (AUC) each muropeptide accounts for is given as a percentage of the total area under the curve accounted for by all 10 muropeptides of interest. GlcN: glucosamine, Mur*N*Ac: *N*-acetylmuramic acid, A: alanine, E: glutamic acid, m: *meso*DAP, G: glycine, Glc*N*Ac: N-acetylglucosamine.

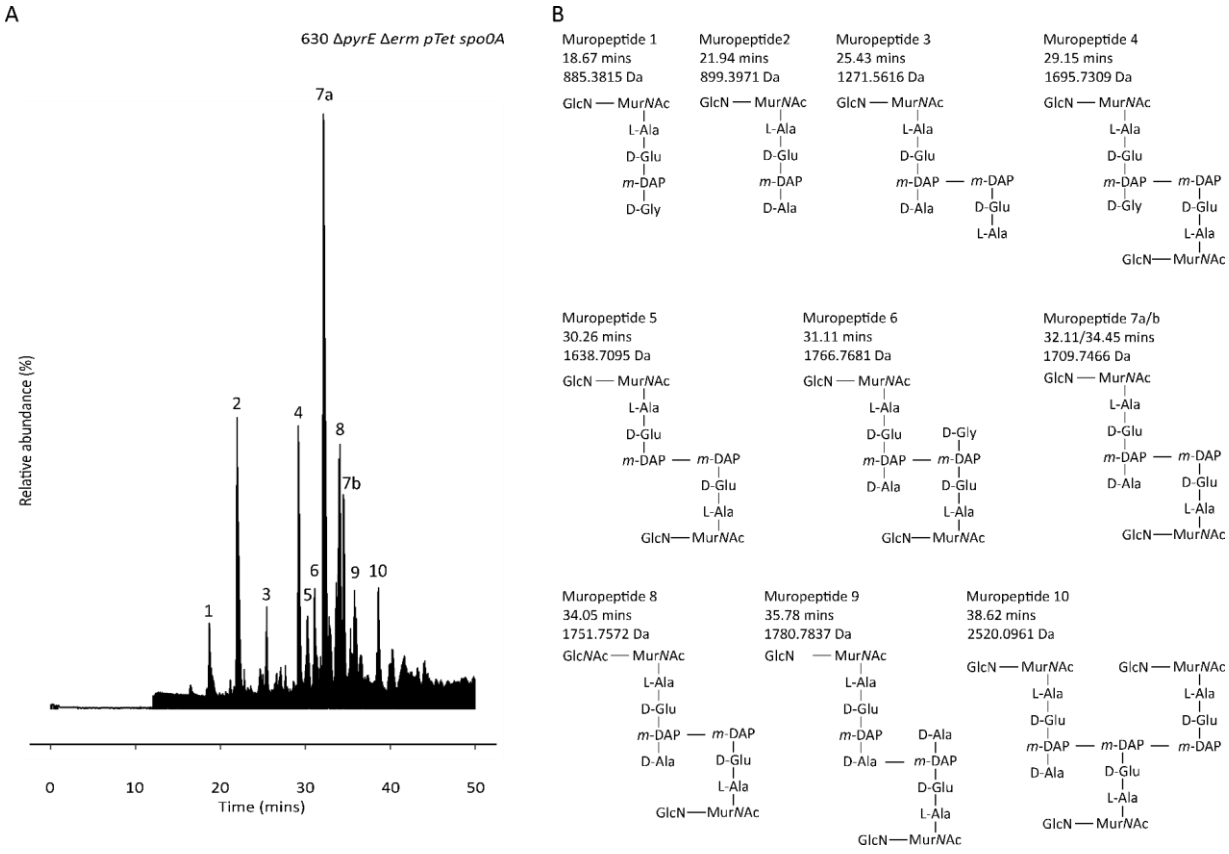


Figure 3-2: HPLC FT-MS analysis of *C. difficile* 630 Δ*pyrE* Δ*erm* $P_{tet}$ *-spo0A* peptidoglycan.

Muropeptides were separated by HPLC before analysis by FT-MS. Panel A shows the total ion chromatograph. The numbered peaks in panel A correspond to the identities in Table 3-1, and the structures presented in panel B. Retention times are given in minutes, along with the theoretical mass of each muropeptide. All Mur*N*Ac residues are reduced.

The monomer containing the 3-3 cross linked AEm AEmA peptide (muropeptide 2) was the most common monomer, accounting for 6.64% of the total area of the 10 majormuropeptides. This muropeptide was also the most common monomer in the analysis by Peltier *et al*., (2011) and Bern *et al.,* (2016).

The most common dimer was muropeptide 7, which eluted as two peaks. Again, this identification is consistent with previous reports (Peltier *et al.*, 2011; Bern, Beniston and Mesnage, 2016).

Of the top ten muropeptides, only one was a trimer. This is consistent with the lower occurrence of trimers (~30% of the area attributed to all identified peaks) reported elsewhere (Peltier *et al.*, 2011; Bern, Beniston and Mesnage, 2016). Differences in the number of trimers observed may be due to differences in HPLC setup; different buffer compositions, gradients, column sizes and types will affect the elution profiles achieved. Future work may involve trialling gradients that allow the resolution of more hydrophobic, and larger muropeptides, as these tend to elute from the column later, and therefore may not have high quality mass spectra associated.

Of all the muropeptides identified in this experiment, 67% were fully deacetylated. Peltier *et al*., (2013) reported deacetylation at 93%. This discrepancy is likely because we identified several low abundance Glc*N*Ac containing muropeptides, which would result in a lower proportion of the total number of unique muropeptides identified being fully deacetylated. However, 90% of the top ten muropeptides are fully deacetylated.

Of the 26 unique dimers identified during this work, 53.1% were 3-3 crosslinked. This is consistent with previous characterisations of *C. difficile* peptidoglycan (Peltier *et al.*, 2011), and the analysis of Bern *et al*. (2016), which was published during this work, where 52.9% and 54% of dimers were found to be 3-3 crosslinked, respectively. The proportions of monomeric peptide stems to crosslinked is shown in Table 3-2, and compared with previous data (Peltier *et al.*, 2011; Bern, Beniston and Mesnage, 2016).

Table 3-2: Monomers and crosslinked species in *C. difficile* peptidoglycan

| Muropeptide size | Percentage of all identified muropeptides (%) | | |
|---|---|---|---|
| | This experiment<br>n MPs=49 | Peltier *et al*. (2011)<br>n MPs=39 | Bern *et al*. (2016)<br>n MPs=65 |
| Monomer | 34.69 | 19.44 | 36.92 |
| Dimer | 53.06 | 52.78 | 33.85 |
| Trimer | 12.24 | 27.78 | 29.23 |

Monomers, dimers and trimers as a percentage of all identified muropeptides in this study compared with those of two previous studies (Peltier *et al.*, 2011; Bern, Beniston and Mesnage, 2016). As represented by n MPs, this study identified 49 unique muropeptides (not including muropeptides that were identified and determined to be the products of in source decay), Peltier *et al.* (2011) identified 39 muropeptides and Bern *et al.* (2016) 65.

Whilst our values for monomers and dimers are similar to either Peltier *et al* (2011) or Bern *et al*., (2016), we identified fewer trimers. This may be due to several factors including differences in growth media, phase of *C. difficile* 630 Δ*pyrE* Δ*erm* $P_{tet}$-*spo0A* growth, or differences in HPLC gradient and acidification prior to analysis.

### 3.1.3. Muropeptides with alternative terminal amino acids

As part of this analysis, several muropeptides with termini other than alanine and glycine were identified (Table 3-3), largely as described in section 3.1.1. Whilst some of these had previously been identified (Bern, Beniston and Mesnage, 2016), seven are described for the first time.

Table 3-3: Muropeptides identified in *C. difficile* 630 Δ*pyrE* Δ*erm* $P_{tet}$ -*spo0A* with unusual terminal amino acids

| RT (min) | Muropeptide | Observed neutral mass (Da) | Theoretical neutral mass (Da) | Crosslink type | Reported by Bern *et al* (2016)? |
|---|---|---|---|---|---|
| 17.14 | GlcN Mur*N*Ac AEmS | 915.3991 | 915.3921 | N/A | Yes |
| 19.96 | GlcNAc Mur*N*Ac AEmR | 1026.476 | 1026.472 | N/A | No |
| 26.49 | GlcN Mur*N*Ac AEmB | 913.4207 | 913.4128 | N/A | No |
| 27.54 | GlcN Mur*N*Ac AEmR AEm GlcN Mur*N*Ac | 1794.825 | 1794.811 | 3-3 | Yes |
| 28.50 | GlcN Mur*N*Ac AEmS AEm GlcN Mur*N*Ac | 1725.754 | 1725.742 | 3-3 | No |
| 33.00 | Glc*N*Ac Mur*N*Ac$_{anhydro}$ AEm | 850.3505 | 850.3444 | N/A | No |
| 35.17 | GlcNAc Mur*N*Ac$_{anhydro}$ AEmG | 907.3788 | 907.3658 | N/A | No |
| 42.21 | GlcN Mur*N*Ac AEmI/L | 941.4588 | 941.4441 | N/A | Yes |
| 43.77 / 43.85 | GlcN Mur*N*Ac Aem AEmV GlcN Mur*N*Ac | 1737.796 | 1737.778 | 3-3 | Yes |
| 45.26 | GlcN Mur*N*Ac AEmY AEm GlcN Mur*N*Ac | 1801.783 | 1801.773 | 3-3 | Yes |
| 46.64 | GlcN Mur*N*Ac AEmA AEmV GlcN Mur*N*Ac | 1808.842 | 1808.815 | 3-3 | Yes |
| 47.28 | Gln*N*Ac Mur*N*Ac AEm AEmP/V GlcN Mur*N*Ac | 1779.811 | 1779.789 | 3-3 | No |
| 49.93 | GlcN Mur*N*Ac AEmF | 975.4398 | 975.4284 | N/A | Yes |
| 53.33 | GlcN Mur*N*Ac AEmI/L GlcN Mur*N*Ac | 1751.814 | 1751.794 | 3-3 | Yes |
| 57.54 | Glc*N*Ac Mur*N*Ac AEmI/L GlcN Mur*N*Ac | 1793.827 | 1793.804 | 3-3 | No |

Observed and theoretical masses are given. Where muropeptides have not been previously identified (Bern, Beniston and Mesnage, 2016), the determination of their structure is described below. Mur*N*Ac$_{anhydro}$ refers to 1,6-anhydro-Mur*N*Ac, B refers to γ-amino butyric acid. All Mur*N*Ac residues were reduced.

#### 3.1.3.1. Glc*N*Ac Mur*N*Ac AEmR

A monomer was identified with a neutral mass of 1026.48 Da. The presence of Glc*N*Ac can be confirmed as the difference between the neutral mass and the ion at 824.29 is 202.2, which indicates the loss of a Glc*N*Ac residue. Following this are ions corresponding to

Mur*N*Ac, L-Ala and D-Glu (II, III and IV respectively Figure 3-3). This results in ~348.98 Da to account for which can be attributed to a terminal arginine residue, assuming the presence of a *meso*DAP residue. Often the loss of a terminal amino acid can be seen from the sugar residues, though this is not the case here. However, the software ChemDraw [PerkinElmer] returns a theoretical mass of 1026.4720 Da for this muropeptide, suggesting this is indeed a valid identification.

This terminal amino acid has been previously identified, but with a GlcN-Mur*N*Ac sugar backbone (Bern, Beniston and Mesnage, 2016) with a neutral mass of 984.4611 Da, approximately 42 Da smaller than the muropeptide described here. This further strengthens our identification, as a difference of 42 Da corresponds to the addition/removal of an acetyl group. It can be expected that this will likely occur on the Glc*N*Ac residue as deacetylation of Glc*N*Ac in *C. difficile* is high, thus suggesting the muropeptide eluting at 19.96 minutes is Glc*N*Ac Mur*N*Ac AEmR.

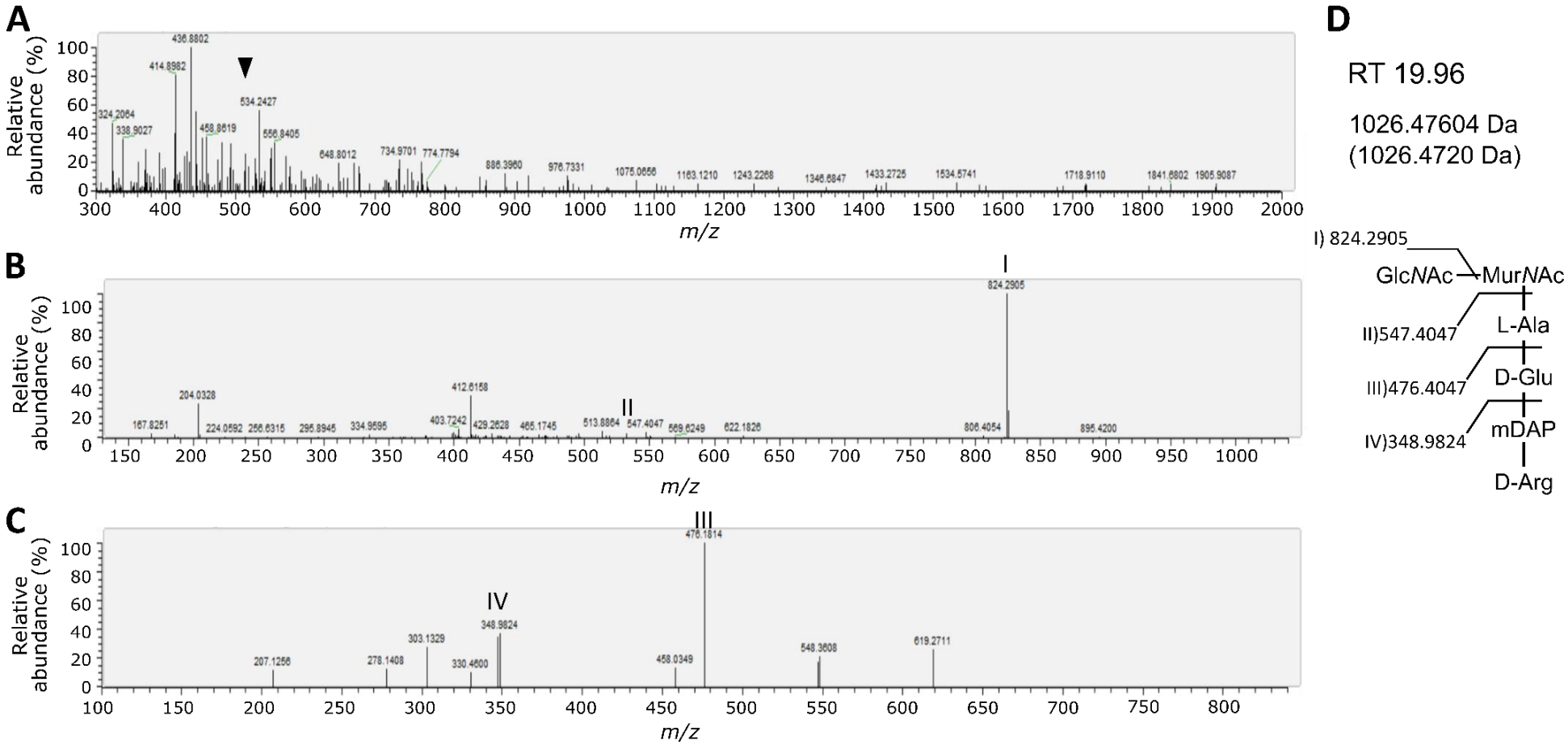


Figure 3-3: Muropeptide eluting at 19.96 minutes.

The ion marked I corresponds to a loss of a Glc*N*Ac residue from the parent ion. Ions corresponding to Mur*N*Ac, L-Ala, and D-Glu are then seen (marked II, III, IV respectively). The remaining mass is then attributed to *meso*DAP and D-Arg. Panel A is the parent scan, B the MS2 and C the MS3. Ions marked in panels B and C correspond to the ion fragments marked in panel D. Panel D provides a schematic of the muropeptide identified with the retention time (RT) given in minutes. The oserved and calculated (in parentheses) neutral masses are provided. The ion marked with a triangle in panel A is the parent ion. All Mur*N*Ac residues were reduced..

#### 3.1.3.2. GlcN Mur*N*Ac AEmB

The muropeptide eluting at 26.49 minutes was identified as GlcN Mur*N*Ac AEmB, with B representing γ-aminobutyric acid (GABA). Similarly to the muropeptide eluting at 19.96 minutes (section 3.1.3.1), the MS2 and MS3 only provided information as far as the glutamic acid residue in the peptide stem (Figure 3-4). Again, assuming the loss of a *meso*DAP residue from the remaining mass of ~276 Da, 104 Da remains to be attributed. It can be assumed that this 104 Da will correspond to a terminal amino acid, as the neutral mass is too small to account for a dimer structure. A candidate identity for this terminal residue is GABA.

This muropeptide has a theoretical neutral mass of 913.4128 Da. With a loss of ~103 from Mur*N*Ac in the MS2 and MS3, observed as the ion present at ~373 *m/z*, followed by the expected ion corresponding to alanine, it can be concluded that this peptide stem likely contains a terminal GABA residue, the first time that this has been described in peptidoglycan. This muropeptide should be enriched and further characterised to confirm the presence of GABA at position 4 of the peptide stem. GABA has been implicated in resistance to acidic pH in *Clostridium perfringens* and is synthesised *in vivo* via the decarboxylation of glutamate by glutamte decarboxylase (Li and Cao, 2010).

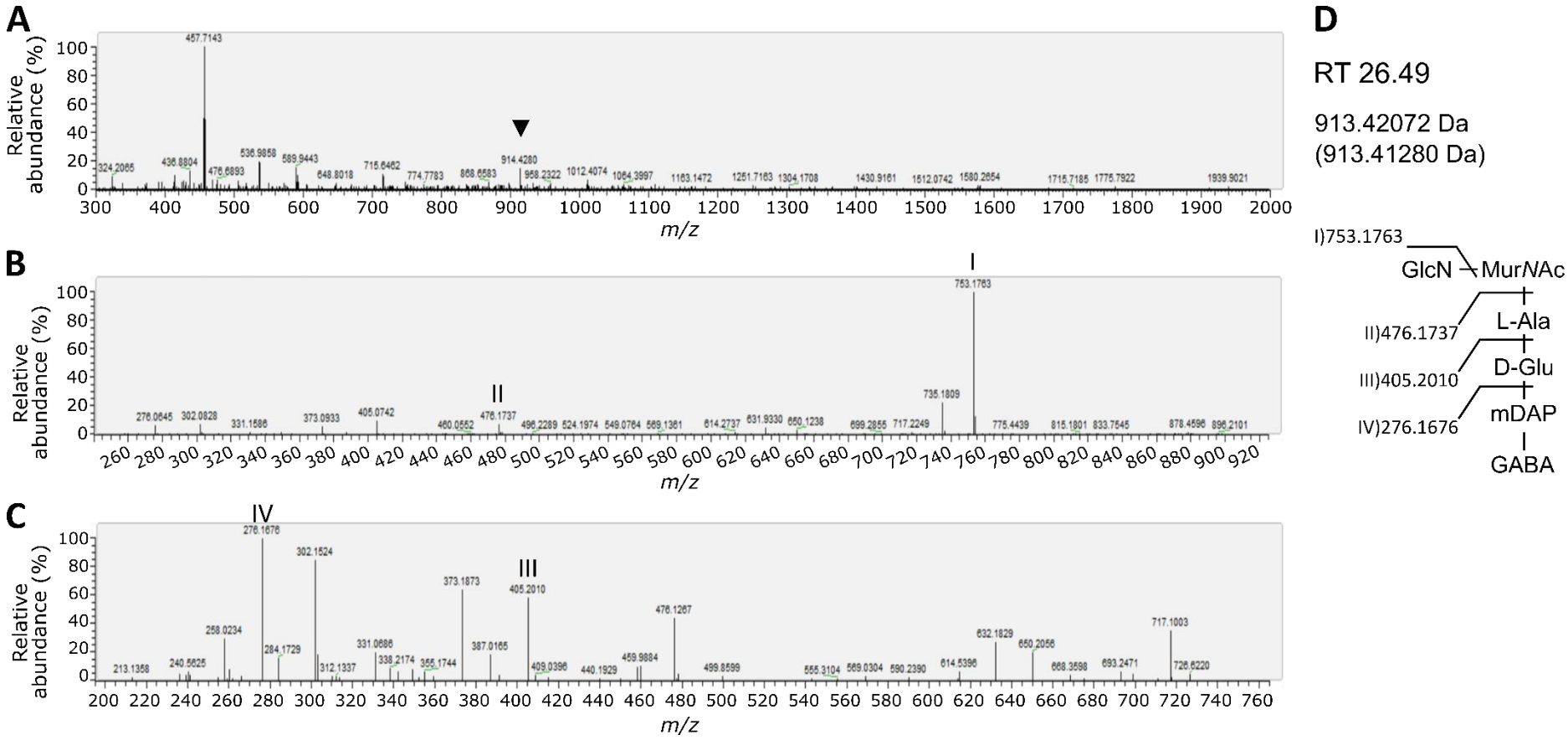


Figure 3-4: Muropeptide eluting at 26.49 minutes.

The muropeptide eluting at 26.49 was determined to have a neutral mass of 913.42072. GlcN is observed (I), followed by Mur*N*Ac, (marked II), L-Ala (ion III) and D-Glu (IV). Not marked, but used in the confirmation of the identity, was the ion seen in the MS2 at 650, which corresponds to the terminal GABA residue. Panel A is the parent scan, B the MS2 and C the MS3. Ions marked in panels B and C correspond to the ion fragments marked in panel D where the retention time (RT) is given in minutes with the observed and calculated (in parentheses) neutral mass. The ion marked with a triangle in panel A is the parent ion. The Mur*N*Ac residue is reduced. GABA; γ-amino butyric acid.

#### 3.1.3.3. GlcN Mur*N*Ac AEmA AEmS GlcN Mur*N*Ac

A muropeptide with a D-Ala-*meso*DAP 4-3 crosslink and a terminal serine residue was identified eluting at 28.5 minutes.

Whilst a monomer containing a terminal serine residue has been previously described (Bern, Beniston and Mesnage, 2016), and is also seen in this experiment (retention time 17.14 minutes), this is the first time the dimer has been characterised.

Incorporation of serine to the terminal position of the peptide stem is associated with resistance to vancomycin in other bacteria (Ammam *et al.*, 2013). Whilst *C. difficile* contains vancomycin-resistance associated genes, it is not yet resistant to vancomycin, which may be a consequence of the high proportion of D-Ala terminating muropeptides (Ammam *et al.*, 2013). Enzymes required for serine integration into the peptide stem have been identified in *C. difficile* (Ammam *et al.*, 2013), further validating this identification.

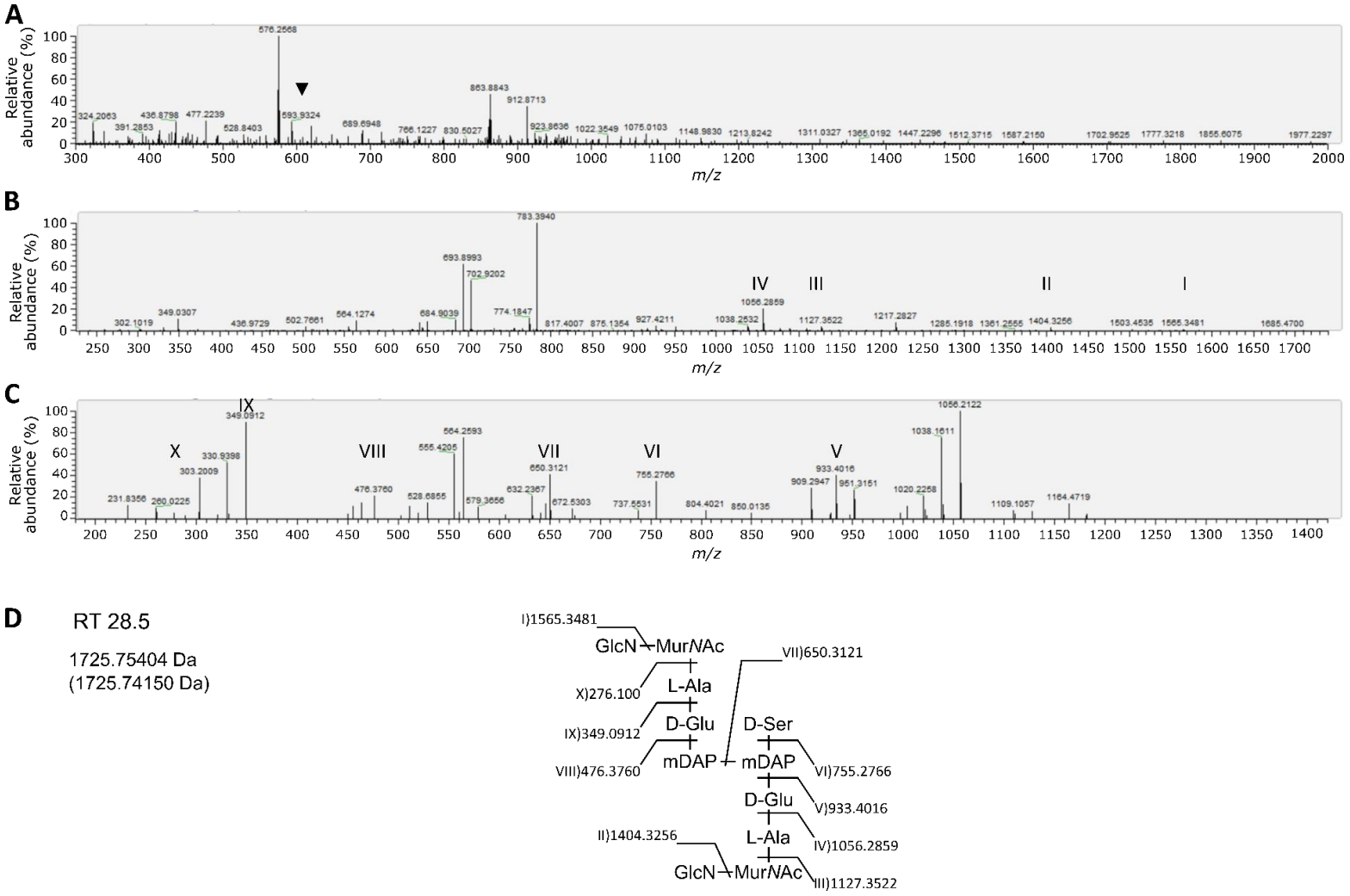


Figure 3-5: Identification of muropeptide eluting at 28.5 minutes

Panel A shows the parent scan, panel B the MS2 and panel C the MS3. Ions marked on panels B and C correspond to the ion fragments marked in panel D. Panel D also provides the retention time (RT) in minutes, the observed and theoretical masses (theoretical mass in parentheses). Both Mur*N*Ac residues are reduced.

#### 3.1.3.4. Glc*N*Ac Mur*N*Ac$_{anhydro}$ AEm

A muropeptide containing a Glc*N*Ac residue and a 1,6-anhydro-Mur*N*Ac residue was observed to elute at 33 minutes. The tripeptide monomer with this sugar combination has not been reported before, though 1,6-anhydro-Mur*N*Ac containing muropeptides have been previously reported (Peltier *et al.*, 2011).

The presence of a 1,6-anhydro-Mur*N*Ac residue is usually associated with either the activity of a lytic transglycosylase, or corresponding to the terminal sugar in the glycan chain (Vollmer, 2008).

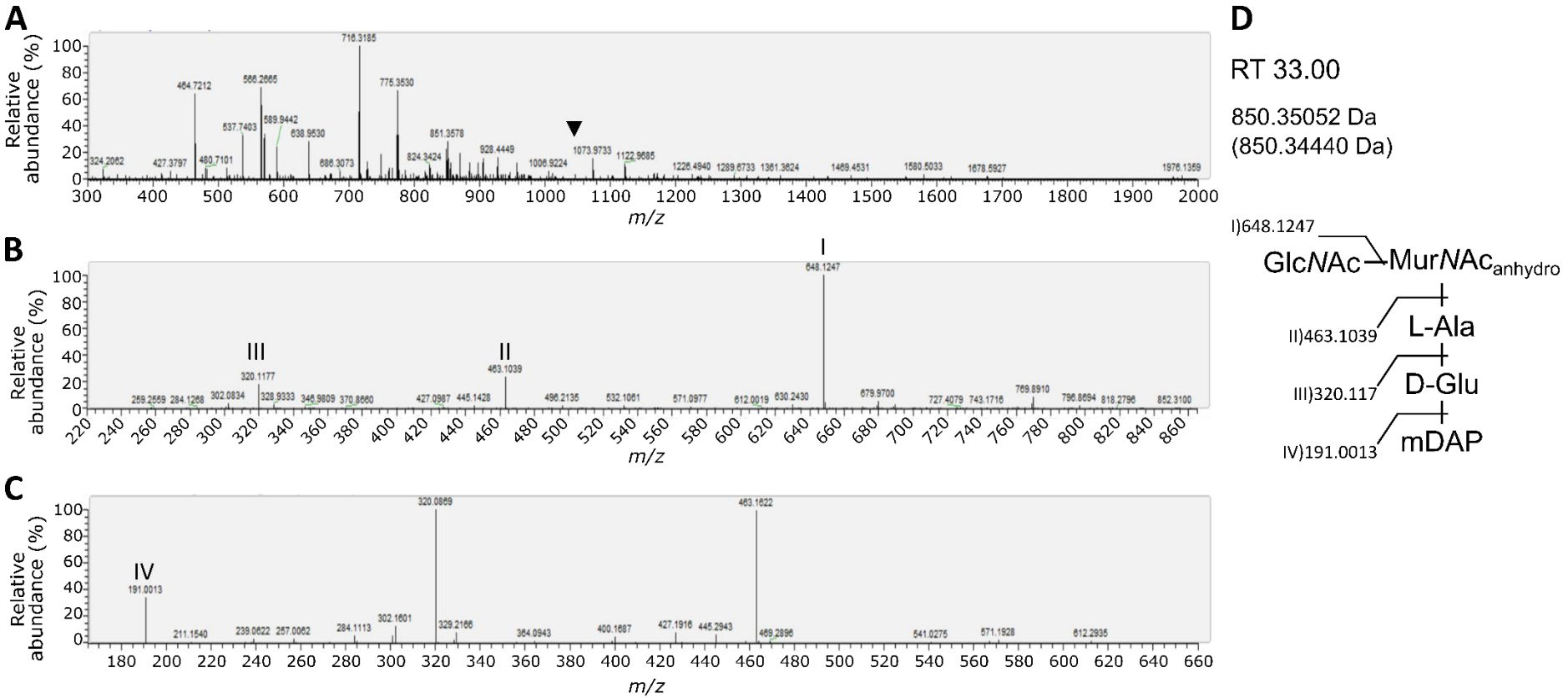


Figure 3-6: Identification of muropeptide eluting at 33.00 minutes.

The presence of the Glc*N*Ac is clear as a fragment ion (at ~648) from the parent ion. Ions II and III in the MS2 correspond to the 1,6-anhydro-Mur*N*Ac residue, and the alanine, though the fragmentation pattern is slightly different as the loss of lactyl-alanine is observed. Using the MS3, an ion containing glutamic acid is seen. The remaining mass can then be accounted for by the presence of a terminal *meso*DAP residue. Panel A is the parent scan, B the MS2 and C the MS3. Ions marked in panels B and C correspond to the ion fragments marked in panel D where the retention time (RT) is provided in minutes, the observed and theoretical masses (in parentheses) provided. The ion marked with a triangle in panel A is the parent ion. The Mur*N*Ac residue is reduced.

#### 3.1.3.5. Glc*N*Ac Mur*N*Ac$_{anhydro}$ AEmG

Similarly to the muropeptide eluting at 33.00 minutes, the muropeptide observed eluting at 35.17 minutes was deduced to contain the Glc*N*Ac-1,6-anhydro-Mur*N*Ac sugar backbone, with the loss of alanine (Figure 3-7 ion III) and a terminal glycine residue (IV) leaving a deficit that can be accounted for by the presence of glutamic acid and *meso*DAP, as would be expected in a peptide stem.

The presence of a terminal glycine has been reported previously, though never in conjunction with this sugar backbone (Peltier *et al.*, 2011; Bern, Beniston and Mesnage, 2016).

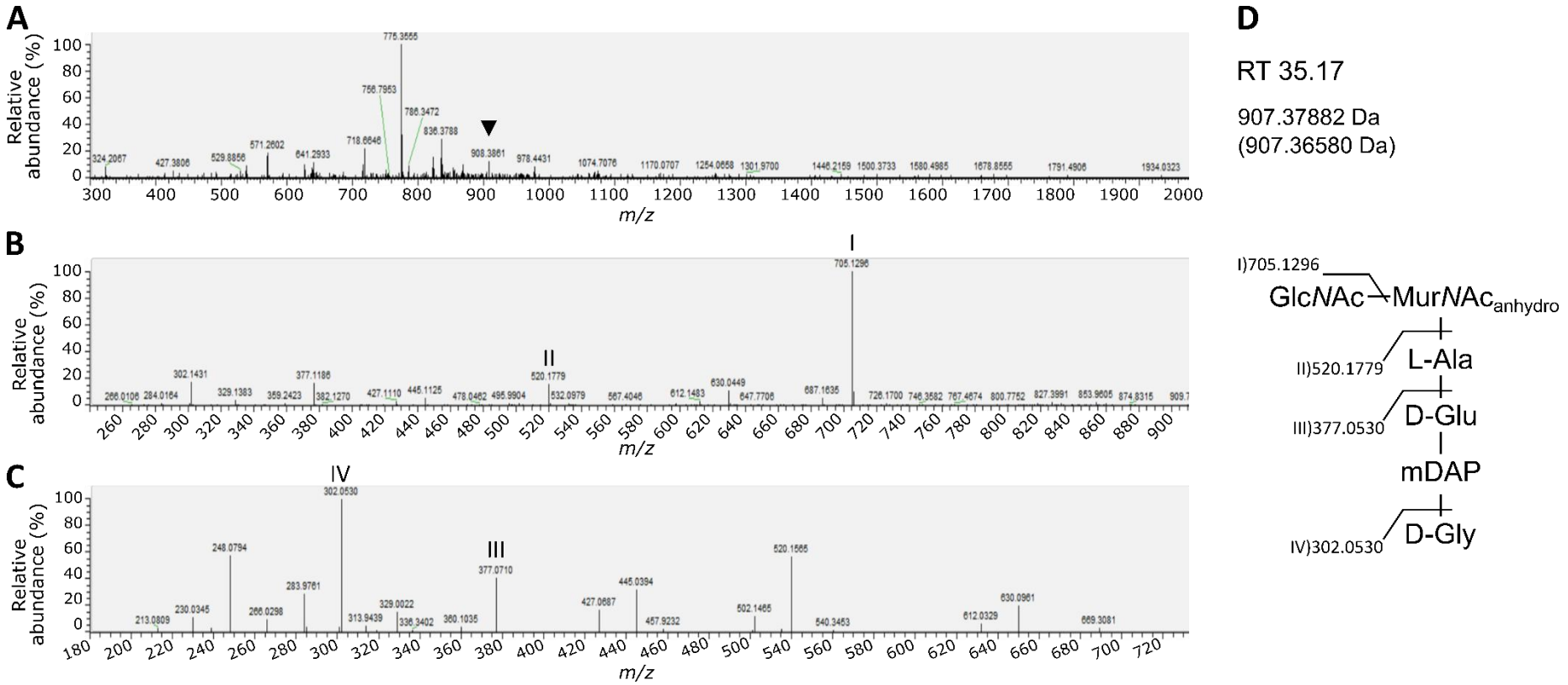


Figure 3-7: Identification of muropeptide eluting at 35.17 minutes

The muropeptide eluting at 35.17 minutes was inferred to be a monomer. Panel A is the parent scan, B the MS2 and C the MS3. Ions marked in panels B and C correspond to the ion fragments marked in panel D where the retention time (RT) is provided in minutes, the observed and theoretical masses (in parentheses) provided. The ion marked with a triangle in panel A is the parent ion. Ions corresponding to Glc*N*Ac 1,6-anhydro-Mur*N*Ac and alanine (ions I II and III respectively) were observed, followed by and ion corresponding to a terminal glycine residue (ion IV), the remaining unaccounted mass was accounted for by the presence of the usual D-Glu and *meso*DAP residues.

#### 3.1.3.6. Glc*N*Ac Mur*N*Ac AEmV AEm GlcN Mur*N*Ac

A dimer containing a terminal valine residue was observed eluting at 47.28 minutes (Figure 3-8). A loss of 117 can be seen following the ion corresponding to the loss of Glc*N*Ac or Ala, which may suggest a terminal valine residue, though no further ions that would correspond to the remaining structure are seen. Using ChemDraw, the neutral mass of this muropeptide was calculated to be 1779.7885 Da. A version of this muropeptide was previously reported (Bern, Beniston and Mesnage, 2016), but only with all Glc*N*Ac residues deacetylated to GlcN.

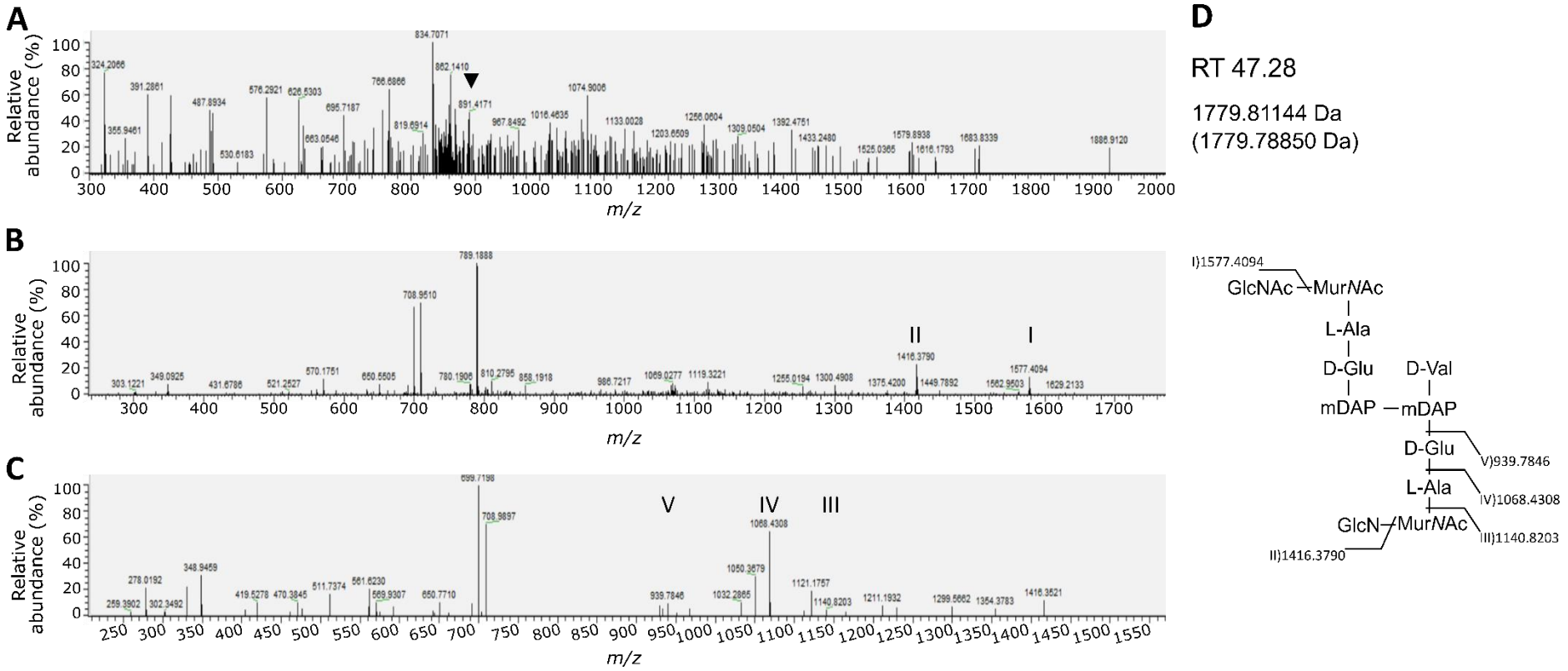


Figure 3-8: Identification of muropeptide eluting at 47.28 minutes

The muropeptide eluting at 47.28 minutes was determined to contain a terminal valine residue, based on the fragment ions observed, and the parent mass. Ions I, II, III, IV and V correspond to identification of Glc*N*Ac, GlcN, Mur*N*Ac alanine, and glutamic acid respectively. Panel A is the parent scan, B the MS2 and C the MS3. Ions marked in panels B and C correspond to the ion fragments marked in panel D where the retention time (RT) is provided in minutes, the observed and theoretical masses (in parentheses) provided. The ion marked with a triangle in panel A is the parent ion. Both Mur*N*Ac residues are reduced.

#### 3.1.3.7. Glc*N*Ac Mur*N*Ac AEmI/L AEm GlcN Mur*N*Ac

The muropeptide eluting at 57.54 minutes was determined to contain leucine or isoleucine at the terminal position. A dimer with this terminal amino acid has been previously reported, but was associated with fully deacetylated Glc*N*Ac residues, whereas here, we see the partially deacetylated backbone as the loss of Glc*N*Ac from the parent ion is present as the strong 1591 ion in the MS2 (Figure 3-9).

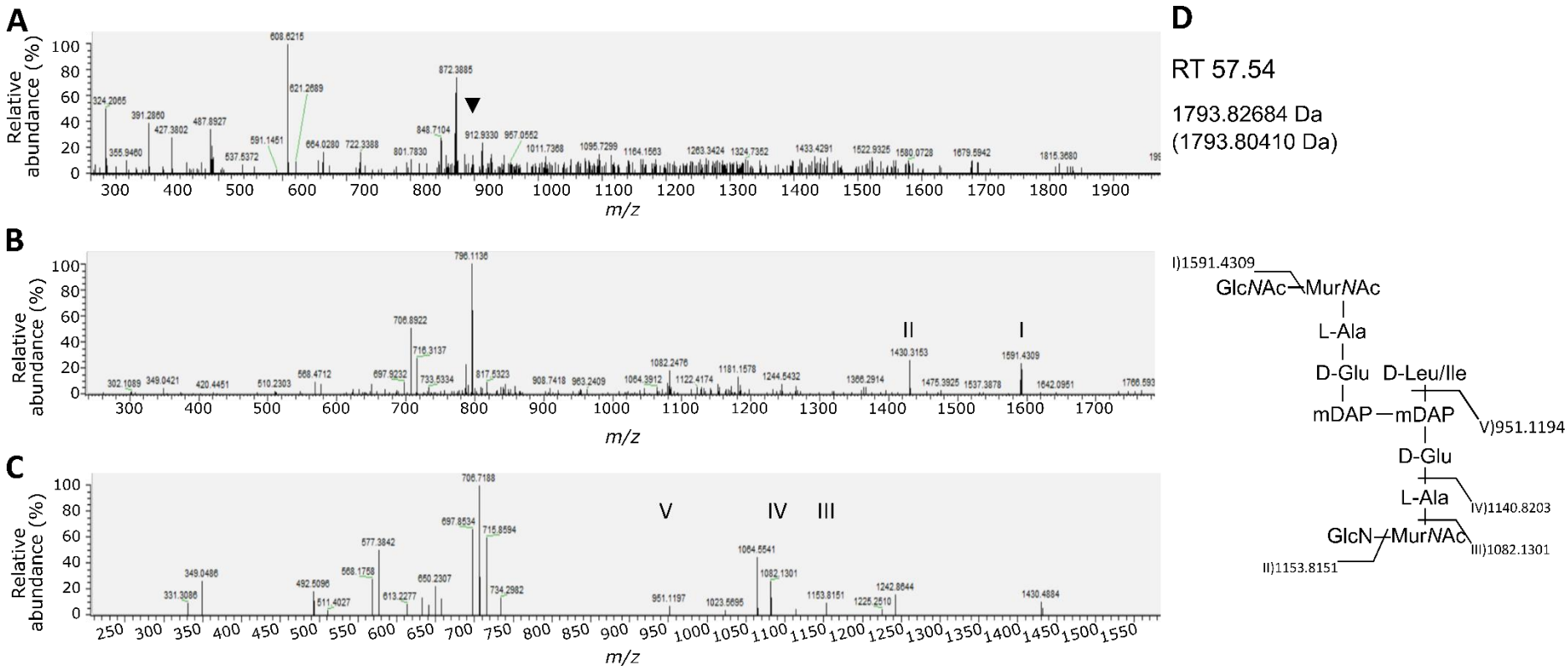


Figure 3-9: Identification of muropeptide eluting at 57.54

A muropeptide containing a terminal leucine or isoleucine residue was observe eluting at 57.54 minutes. The presence of the leucine/isoleucine can be observed as the ion marked V. One Glc*N*Ac residue, and one GlcN residue (ions I and II respectively) are observed, indicating this muropeptide differs from that previously identified (Bern, Beniston and Mesnage, 2016).All muropeptides are reduced. Panel A is the parent scan, B the MS2 and C the MS3. Ions marked in panels B and C correspond to the ion fragments marked in panel D where the retention time (RT) is provided in minutes, the observed and theoretical masses (in parentheses) provided. The ion marked with a triangle in panel A is the parent ion. Both Mur*N*Ac residues are reduced.

The incorporation of non-canonical D-amino acids (NCDAAs) has been previously described in *C. difficile* peptidoglycan (Bern, Beniston and Mesnage, 2016). This work expands on this; verifying those NCDAA-termini and identifying novel termini (Table 3-3).

NCDAAs are also observed in other bacteria (Caparrós, Pisabarro and de Pedro, 1992; Cava *et al.*, 2011) and are incorporated post-peptidoglycan synthesis. This is yet to be investigated fully in *C. difficile*, though enzymes that may be involved in terminal amino acid switching have been identified (Cava *et al.*, 2011; Ammam *et al.*, 2013).

### 3.1.4. Amidated muropeptides

One muropeptide, eluting at approximately 31 minutes, was determined to have an observed neutral mass of 1412.64784 Da.

Ions correspondinf to the loss of GlcN, Mur*N*Ac and alanine were easily identified. The alanine was followed by an ion that could be attributed to either glutamic acid, which would be seen as a 129.04312 ion in the MS2/3, or a glutamine residue, which would be seen as a 128.05912 ion. Therefore, the remainder of the structure was determined using both the mass deficit and the masses of known amino acids.

Using this approach and ChemDraw, two possible structures were identified: i) GlcN Mur*N*Ac AQmA AQmAlactyl or ii) GlcN Mur*N*Ac AQmA AEmAlactyl where one of the mesoDAP residues is amidated (Figure 3-10). Both muropeptides are feasible as neither are definitively observed in the spectra. However, as losses that could correspond to glutamate and *meso*DAP (an ion of 128.8611 from alanine followed by an ion the is 171.1995 smaller, the amidation on this muropeptide was assigned to the glutamic acid residues, though this identification comes with the caveat that the losses observed on the spectra are not definitive.

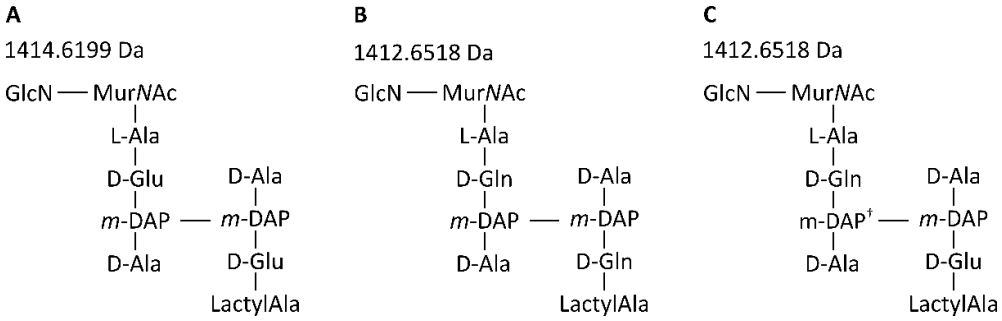


Figure 3-10: Various amidation options of muropeptide eluting at 31.59 minutes

Panel A shows the expected non-amidated muropeptide and its theoretical neutral mass. Panel B shows the configuration of the muropeptide where both expected glutamic acid residues are amidated to glutamate (Gln), and its associated theoretical neutral mass. Panel C shows a second possibility, one of the two glutamic acid residues are amidated to glutamate, and one of the *meso*DAP (m-DAP) residues is amidated to yield m-DAP$^{+}$. Muropeptides B and C have equivalent calculated neutral masses, therefore distinguishing between the two is difficult.

The glutamic acid residues may have been amidated *in vivo*. Analysis by Ammam *et al*. (2013) identified peptidoglycan precursors that may have been amidated either on the Glu

or *meso*DAP residues, though similar amidation was not identified in analyses of mature *C. difficile* peptidoglycan (Peltier *et al.*, 2011; Bern, Beniston and Mesnage, 2016). Amidation of *iso*Glu has recently been implicated in control of crosslinking in *Enterococcus faecium, Mycobacterium tuberculosis* and *Mycobacterium smegmatis:* a lack of *iso*Glu amidation lead to a reduction in crosslinking of peptidoglycan by PBPs (Ngadjeua *et al.*, 2018; Pidgeon *et al.*, 2019).

Whilst one alanine residue is bound to Mur*N*Ac, the second alanine is bound only to a lactyl group. This may indicate that, *in vivo,* there is a Mur*N*Ac residue bound here, however this is not seen in this analysis. This may be because the Mur*N*Ac-alanine bond has been broken i*n vivo*, leaving lactyl-alanine, or that the Mur*N*Ac has been lost during ionisation.

## 3.2. Peptidoglycan composition during sporulation in *C. difficile*

### 3.2.1. Stalling sporulation at different stages

The rationale of this experiment is based on the control of sporulation in *C. difficile* (Figure 1-4). We hypothesised that by characterising the peptidoglycan of sporulation-stalled strains, we would achieve a "step-by-step" view of sporulation in *C. difficile*. Deletion of *sigE* and *sigF* would arrest the sporulation process early in engulfment and would prevent the transcription of several genes, including *spoIID* and *spoIIP*, respectively. Deletion of *spoIID*, *spoIIP, spoIIQ* and *spoIIIAH* would allow the direct impact of these proteins on peptidoglycan remodelling to be identified.

All mutants were able to undergo asymmetric division to varying degrees, but none of these strains were able to produce mature spores. Sporulation efficiency was considered to be the ability to form asymmetric septa, which was determined to be; 78% for Δ*sigE* mutants, approximately 30% in Δ*spoIID/P* and Δ*sigF* mutants, around 45% in Δ*spoIIIAH,* and as low as 22% in Δ*spoIIQ*. It is worth noting that sporulation efficiency in the wildtype strain is between 30-50%, depending on growth media, and that all mutants are stalled soon after asymmetric septa formation (Dembek *et al.*, 2018). This varying degree of sporulation has implications in the peptidoglycan composition analysis, as samples contain a mixture of sporulating and vegetative cells. The analysis presented here considers this caveat and all calculations were normalised according to sporulation efficiency of the specific strain.

### 3.2.2. Using MZmine2

MZmine2 is an open source software that allows automated identification of compounds in a mass spectrum, given a database (Pluskal *et al.*, 2010). An imported mass spectrum must undergo various steps of processing within MZmine2, such as peak detection, peak building and chromatograph deconvolution, for example, before ion identification. Following this, a custom database can be provided, which MZmine2 then uses to assign identities to the various detected ions. The database used in this work was produced by the manual evaluation of all MS3 spectra produced as part of the analysis of 630 Δ*erm* $P_{tet}$ *-spo0A* cells peptidoglycan and is provided in 6.7.Appendix A.

### 3.2.3. Major muropeptides identified throughout sporulation

The same 10 major muropeptides identified in the parental strain are observed in the peptidoglycan from the different mutants, as seen in Figure 3-11, Figure 3-12 and Figure 3-13. This was unexpected, as we had hypothesised that deletion of key genes would prevent peptidoglycan remodelling to varying degrees, but that distinct changes would be observed.

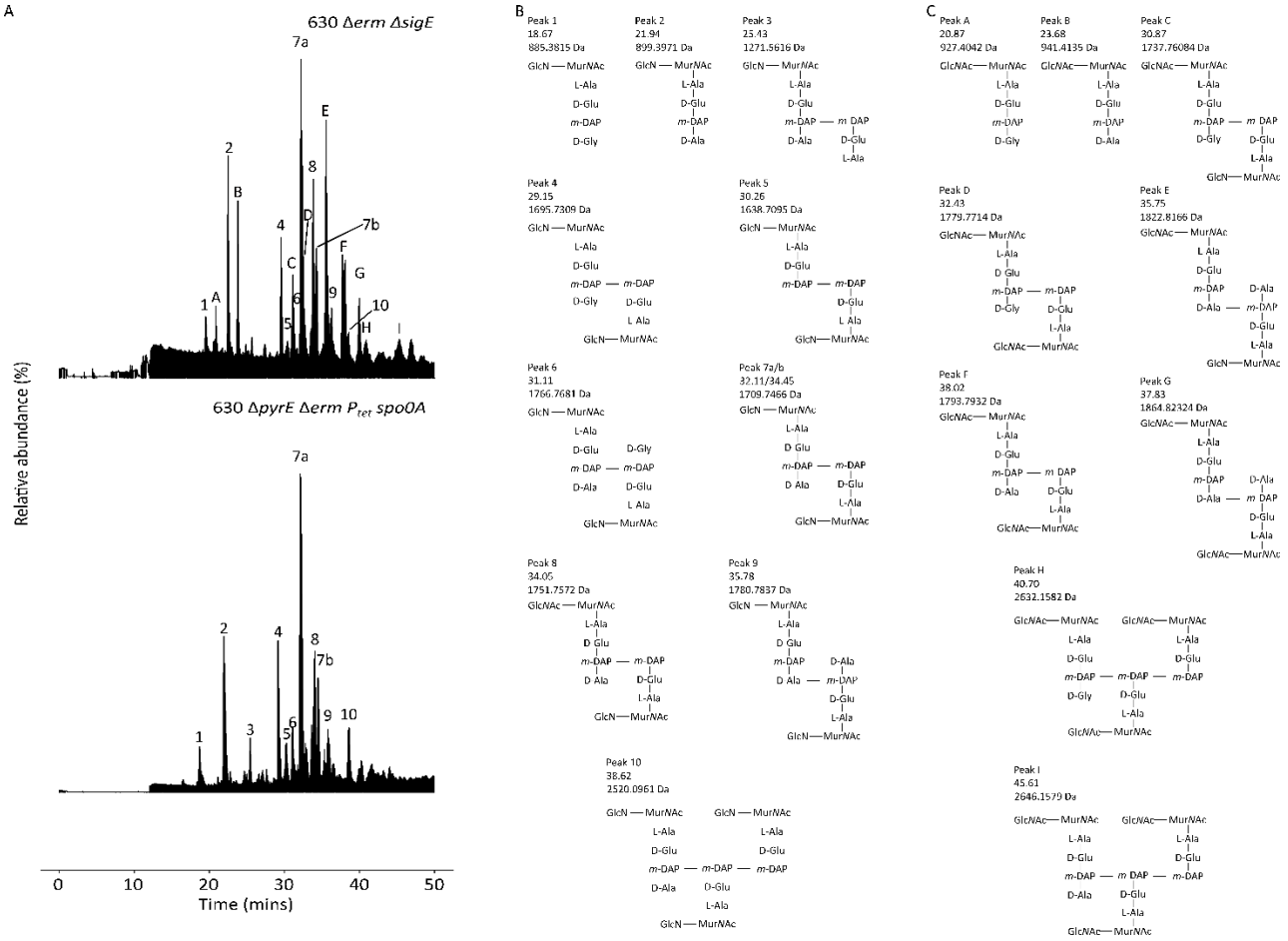


Figure 3-11: HPLC FT-MS analysis of peptidoglycan from *C. difficile 630* Δ*erm* Δ*sigE*.

Muropeptides from *C. difficile 630* Δ*pyrE* Δ*erm* $P_{tet}$ *-spo0A* and *C. difficile 630* Δ*erm* Δ*sigE*. Panel A shows the total ion chromatographs, with the peaks labelled with numbers corresponding to muropeptides identified in analysis of wildtype peptidoglycan (panel B) and the peaks labelled with letters corresponding to those identified in the *C. difficile 630* Δ*erm* Δ*sigE* peptidoglycan analysis (panel C). Each muropeptide is accompanied by the retention time in minutes and the theoretical neutral mass. All Mur*N*Ac residues are reduced.

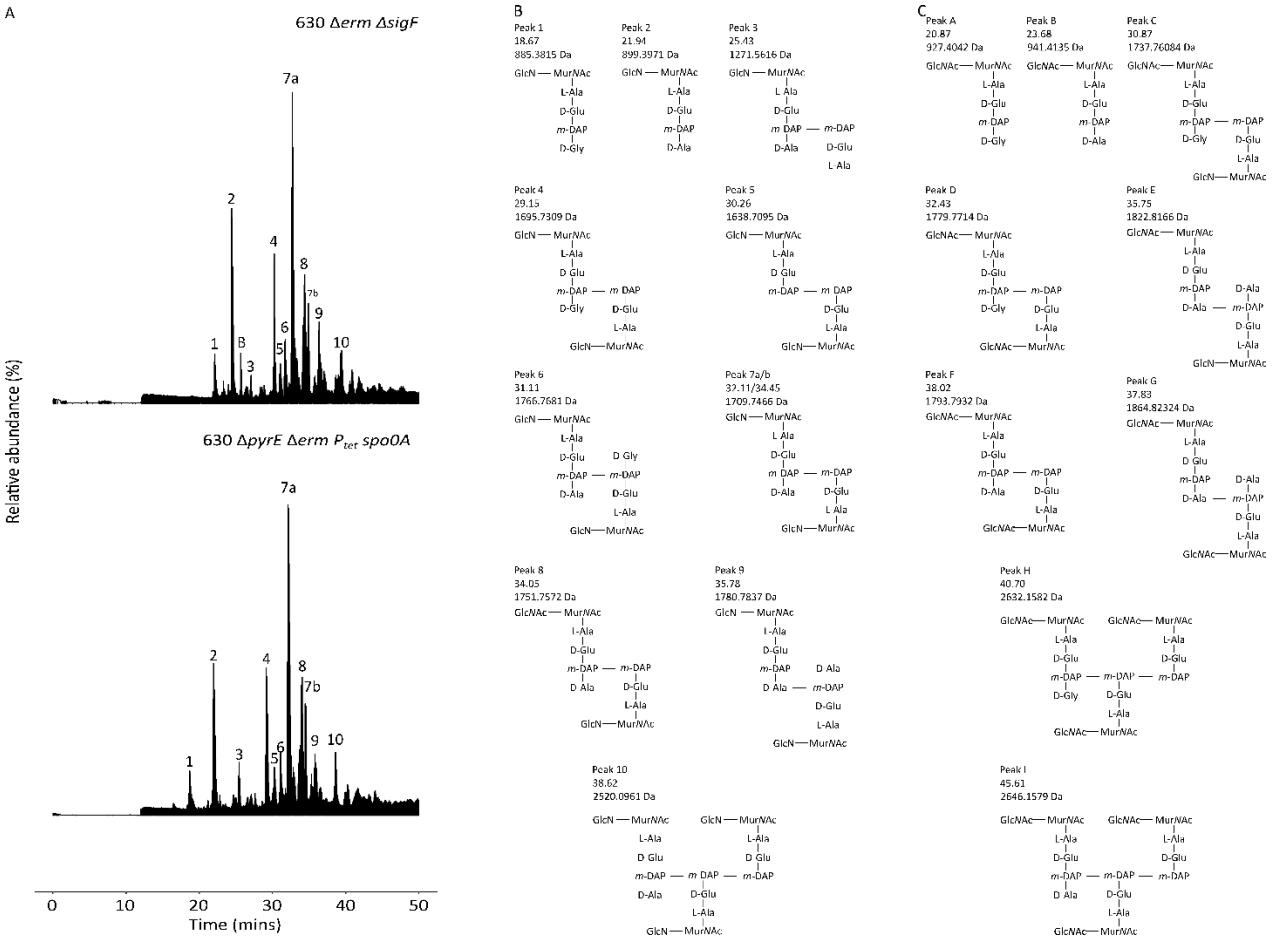


Figure 3-12: Peptidoglycan analysis of *C. difficile 630* Δ*erm* Δ*sigF*

Muropeptides *from C. difficile 630* Δ*pyrE* Δ*erm* $P_{tet}$ *-spo0A* and *C. difficile 630* Δ*erm* Δ*sigF*. Panel A shows the total ion chromatograph, with the peaks labelled with numbers corresponding to muropeptides identified in analysis of wildtype peptidoglycan (panel B) and the peaks labelled with letters corresponding to those identified in the *C. difficile 630* Δ*erm* Δ*sigF* peptidoglycan analysis (panel C). Each muropeptide is accompanied by the retention time in minutes and the theoretical neutral mass. All Mur*N*Ac residues are reduced.

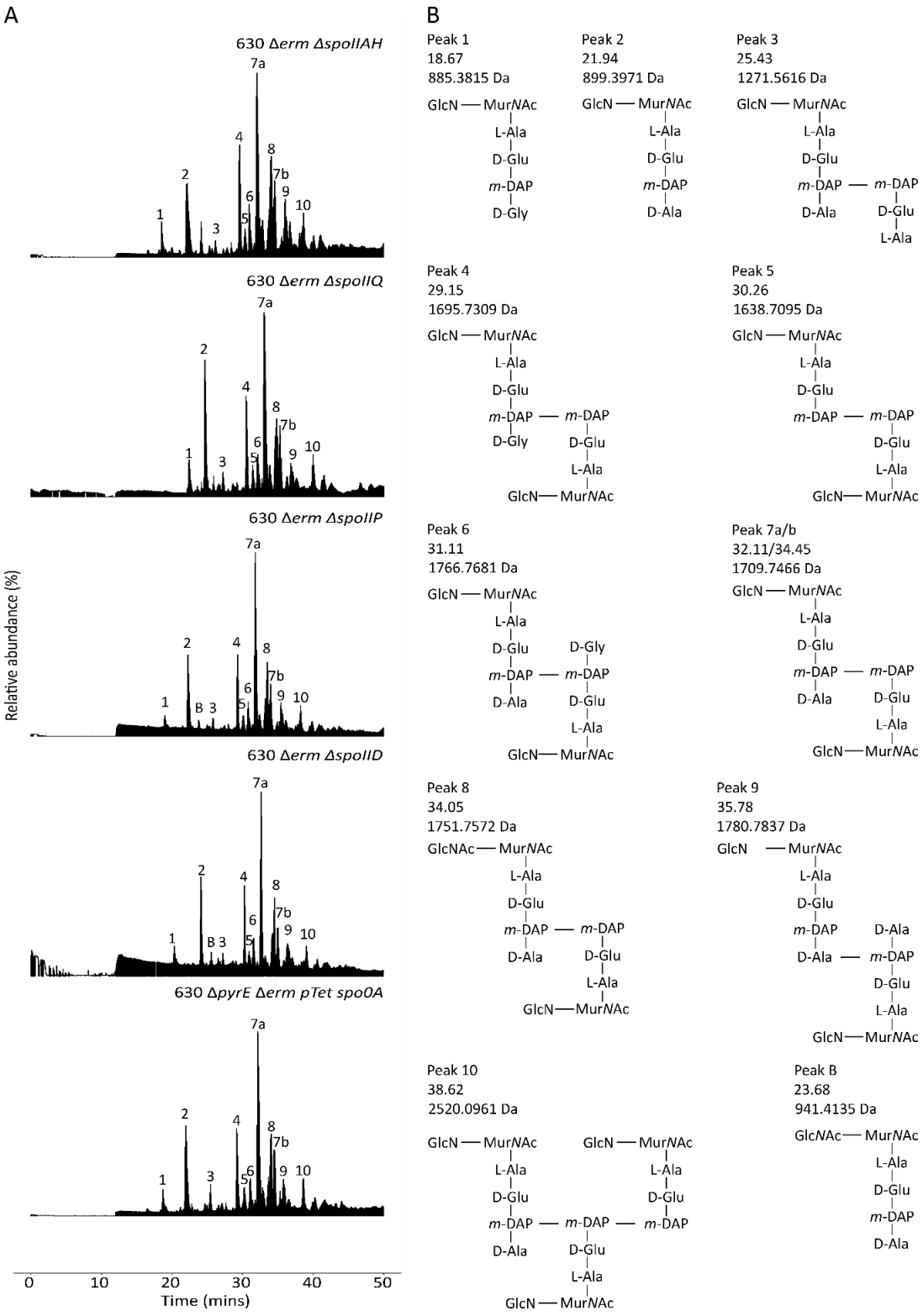


Figure 3-13: HPLC FT-MS analysis of peptidoglycan isolated from *C. difficile* mutant strains. Muropeptides from *C. difficile* 630 Δ*pyrE* Δ*erm* $P_{tet}$ -*spo0A*, *C. difficile* 630 Δ*erm* Δ*spoIID*, *C. difficile* 630 Δ*erm* Δ*spoIIP*, *C. difficile* 630 Δ*erm* Δ*spoIIQ* and *C. difficile* 630 Δ*erm* Δ*spoIIIAH*, were separated by HPLC before analysis by FT-MS. Panel A shows the total ion chromatograph. Numbered peaks correspond to the identities in panel B. Muropeptide B is as identified in Figure 3-11 and Figure 3-12. Retention times are given in minutes, with the theoretical mass of each muropeptide. All Mur*N*Ac residues are reduced.

The area under the curve (AUC) accounted for by each major muropeptide in the peptidoglycan from each mutant strain was calculated and is provided in 6.7.Appendix A, with Figure 3-14 demonstrating the differences between strains.

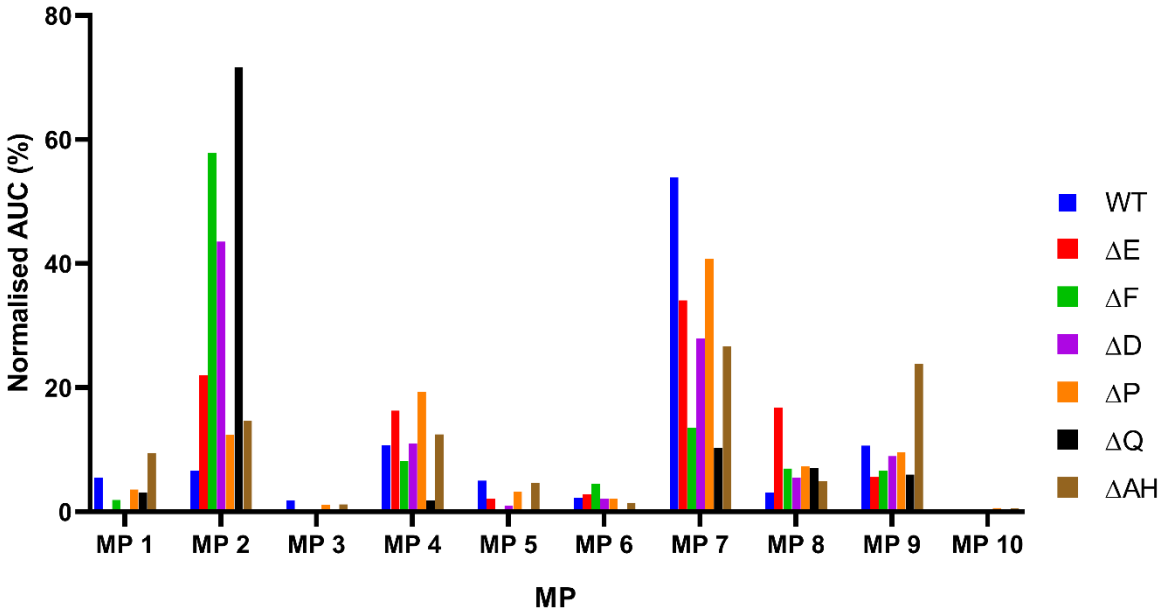


Figure 3-14: Variation of each of the major muropeptides across strains

The area accounted for by each muropeptide and the total area accounted for by the 10 muropeptides of interest were normalised to the sporulation efficiency of the strain in question. The area under the curve (AUC) refers to the normalised area accounted for by a given muropeptide in a given strain, values for which are in 6.7.Appendix A. Muropeptide (MP) numbers refer to those identified in Table 3-1. WT*; C. difficile 630* Δ*pyrE* Δ*erm* $P_{tet}$ *-spo0A,* ΔE*; C. difficile 630* Δ*erm* Δ*sigE,* ΔF*; C. difficile 630* Δ*erm* Δ*sigF,* ΔD*; C. difficile 630* Δ*erm* Δ*spoIID,* ΔP*; C. difficile 630* Δ*erm* Δ*spoIIP,* ΔQ*; C. difficile 630* Δ*erm* Δ*spoIIQ and* ΔAH*; C. difficile 630* Δ*erm* Δ*spoIIIAH.*

Upon visual analysis of Figure 3-11 to Figure 3-13, the elution profiles of each strain look remarkably similar. However, when we use MZmine2 to quantify these changes we see there are some differences between strains. For example, muropeptide 3 is not detected by MZmine2 in Δ*sigE*, Δ*sigF* or Δ*spoIID* mutant profiles; however, there are peaks in the chromatograph that, when checked manually, contain this muropeptide. This phenomenon is likely a result of the parameters used during peak detection, peak joining and peak identification during the MZmine2 automated analysis. Several parameters may be responsible for this discrepancy, and this should be investigated further in the future ensuring that all peaks present in the chromatograph, that can be manually identified, are detected and identified by MZmine2. The discussion of the results presented in this thesis is done with this caveat in mind. Furthermore, the peptidoglycan of *C. difficile 630* Δ*erm* Δ*sigE* contains several partially or fully acetylated muropeptides (see section 3.2.4.1).

### 3.2.4. Changes to peptidoglycan throughout sporulation

To better interpret these results, the data were processed so that a comparison of the area accounted for by each of the muropeptides to the same muropeptide in the wild type strain could be undertaken. These results are provided in 6.7.Appendix A and Figure 3-15.

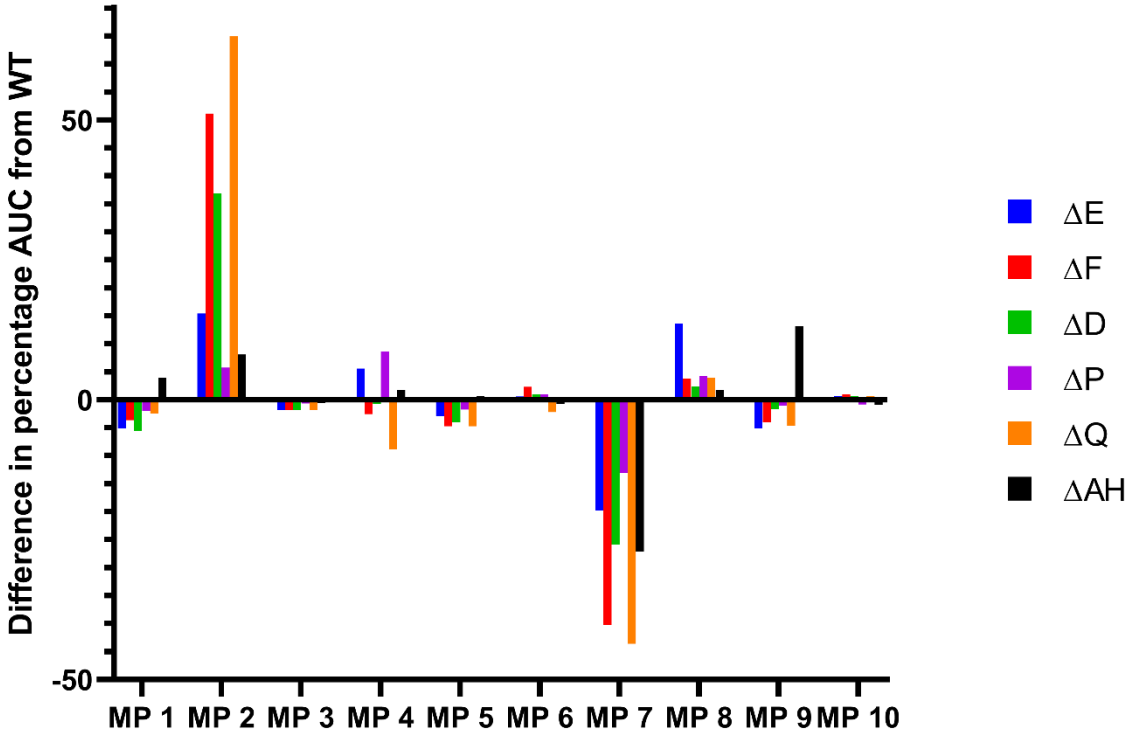


Figure 3-15: Comparison of each major muropeptide peak in mutant *vs* wildtype strains.

The difference between the normalised area under the curve (AUC) attributed to a given muropeptide (MP) in peptidoglycan isolated from mutant strains compared to that from wild type. Data is provided in 6.7.Appendix A. ΔE*; C. difficile* 630 Δ*erm* Δ*sigE,* ΔF*; C. difficile 630* Δ*erm* Δ*sigF,* ΔD*; C. difficile 630* Δ*erm* Δ*spoIID,* ΔP*; C. difficile 630* Δ*erm* ΔspoIIP*,* ΔQ*; C. difficile 630* Δ*erm* Δ*spoIIQ* and ΔAH*; C. difficile 630* Δ*erm* Δ*spoIIIAH.* MP numbers are as in Table 3-1..

Further to this, the proportions of monomers (muropeptides 1 and 2) to crosslinked muropeptides was assessed (Figure 3-16), to investigate if net crosslinking was altered throughout sporulation.

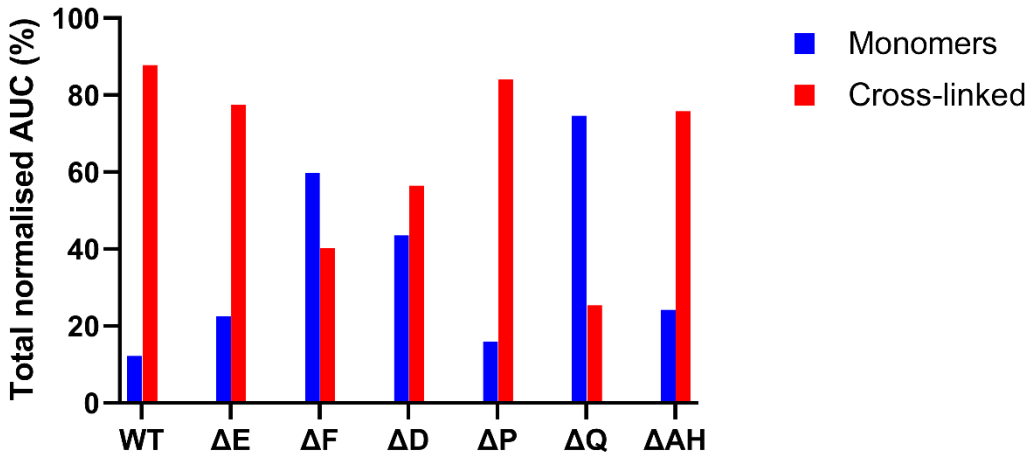


Figure 3-16: Changes to crosslinking across mutants

The proportion of the normalised area under the curve (AUC) accounted for by monomers (muropeptides 1 and 2) and crosslinked muropeptides (muropeptides 3-10) was calculated and expressed as a percentage. Muropeptide (MP) numbers refer to those identified in Table 3-1. WT*; C. difficile 630* Δ*pyrE* Δ*erm* $P_{tet}$ *-spo0A,* ΔE*; C. difficile 630* Δ*erm* Δ*sigE,* ΔF*; C. difficile 630* Δ*erm* Δ*sigF,* ΔD*; C. difficile 630* Δ*erm* Δ*spoIID,* ΔP*; C. difficile 630* Δ*erm* ΔspoIIP*,* ΔQ*; C. difficile 630* Δ*erm* Δ*spoIIQ and* ΔAH*; C. difficile 630* Δ*erm* Δ*spoIIIAH*. Data is provided in 6.7.Appendix A.

In all the mutant strains, the greatest changes seen to the muropeptide profile are in muropeptides 2 (GlcN Mur*N*Ac AEmA) and 7 (GlcN Mur*N*Ac AEmA AEm GlcN Mur*N*Ac). An increase in muropeptide 2 dominates over the reduction in muropeptide 7 in Δ*sigF*, Δ*spoIID* and Δ*spoIIQ,* perhaps suggesting that deletion of these genes results in either an increase in endopeptidase activity, cleaving the bonds between peptide stems of peptidoglycan strands, or, conversely, a reduction in crosslinking between these stem peptides during peptidoglycan synthesis. The opposite is true in Δ*sigE*, Δ*spoIIP* and Δ*spoIIIAH;* there is a net increase in muropeptide 7.

#### 3.2.4.1. *C. difficile 630* Δ*erm* Δ*sigE* peptidoglycan

As $\sigma^E$ controls the early steps in sporulation, it is expected that there will be little difference from the vegetative cell peptidoglycan.

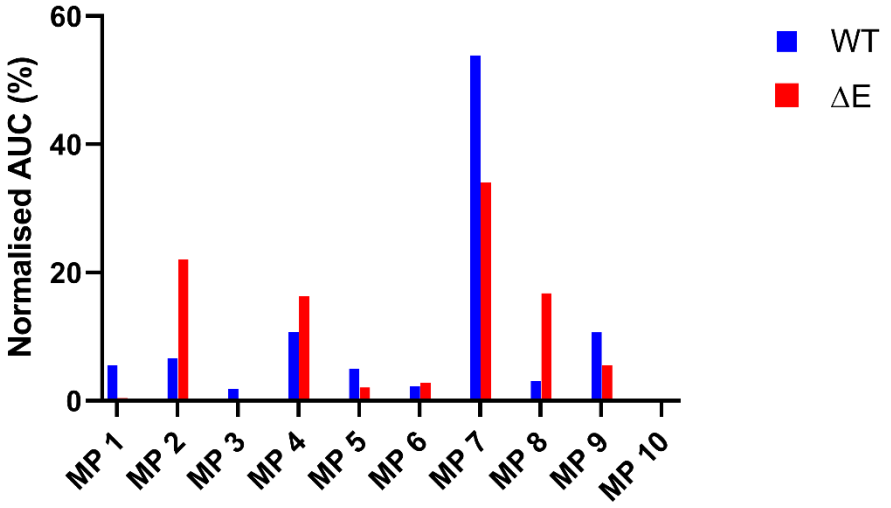


Figure 3-17: Changes to the 10 most abundant muropeptides in *C. difficile* 630 Δ*pyrE Δerm* $P_{tet}$ *spo0A* peptidoglycan in *C. difficile* 630 Δ*erm* Δ*sigE* peptidoglycan.

The percentage of the normalised area under the curve (AUC) attributed to each muropeptide is visualised with the same muropeptide in the wild type (WT) for comparison. Muropeptide identities correspond to those in Figure 3-11. WT*; C. difficile 630* Δ*pyrE* Δ*erm* $P_{tet}$ *-spo0A,* ΔE*; C. difficile 630* Δ*erm* Δ*sigE*. Data is provided in 6.7.Appendix A.

*C. difficile 630* Δ*erm* Δ*sigE* peptidoglycan differs from vegetative cells with respect to muropeptides 2, 7 and 8 (Figure 3-15). The reduction in muropeptide 7 and concurrent increase in muropeptide 2 may suggest that, by disrupting $\sigma^E$, peptidoglycan synthesis or digestion has been affected, leading to a decrease in the crosslinked muropeptide 7. Considering that several peptidoglycan remodelling enzymes are regulated by $\sigma^E$ and the downstream $\sigma^K$ (Saujet *et al.*, 2013), for example alanine racemase 2, N-acetylmuramoyl-L-alanine amidase and UDP-*N*-acetylmuramyl-tripeptide synthetase, it could be that crosslinking is reduced in the *sigE* mutant during peptidoglycan synthesis. Alternatively, peptidoglycan digestion by endopeptidases may be ongoing, digesting muropeptide 7 into muropeptide 2. Where $\sigma^E$ is inactive, one would still expect SpoIIP activity, as it is under control of the forespore-specific $\sigma^F$ (Ribis, Fimlaid and Shen, 2018). However, this is partially inconsistent with the observation that crosslinked muropeptides as a whole (muropeptides 3-10) predominate in the Δ*sigE* mutant (Figure 3-16). It is possible that SpoIIP, or other endopeptidases potentially involved in engulfment and not regulated by $\sigma^E$, could have some level of specificity that would result in a decrease in muropeptide 7 but be less effective in cleaving either Gly termini (muropeptide 4) or the acetylated form of the dimer (muropeptide 8). Alternatively, peptidoglycan synthesis and crosslinking enzymes might be more active in this mutant.

Muropeptide 8 (Glc*N*Ac Mur*N*Ac AEmA AEm GlcN Mur*N*Ac) is much more predominant in this mutant peptidoglycan than is seen in wild type vegetative cells. This muropeptide is only partially deacetylated containing both GlcN and Glc*N*Ac. Analysis of the chromotographs in Figure 3-11 reveals the presence of acetylated versions of the 10 major peaks seen in peptidoglycan in this strain: peak A is the acetylated version of muropeptide 1, peak B the acetylated version of muropeptide 2, peak C is a partially deacetylated and peak D the fully acetylated versions of muropeptide 4, peaks E and G the partially and fully acetylated forms muropeptide 9, peak F the fully acetylated muropeptide 7, peak H a fully acetylated trimer with a terminal glycine residue, and peak I a fully acetylated version of muropeptide 10. This is highly suggestive of a disruption to peptidoglycan deacetylation in the $\sigma^E$ mutant. Interestingly, $\sigma^E$ regulates expression of a deacetylase encoded by *CD630_13190* (Saujet *et al.*, 2013) and absence of transcription of this enzyme could result in a net reduction in deacetylation. That fact that deacetylated muropeptides are still present suggests there is some redundancy in deacetylation and indicates that more than one deacetylase is active during sporulation.

#### 3.2.4.1. *C. difficile 630* Δ*erm* Δ*sigF* peptidoglycan

Unlike $\sigma^E$, which is produced in the mother cell, $\sigma^F$ is produced in the forespore and controls the activity of, amongst many other genes, *spoIIP*, *spoIIQ* and *sigG* (Saujet *et al.*, 2013). As $\sigma^F$ controls *spoIIP* expression (Saujet *et al.*, 2013; Ribis, Fimlaid and Shen, 2018), it is surprising that there is a substantial increase in monomers in this mutant peptidoglycan (Figure 3-18) . This suggests that SpoIIP is not the only endopeptidase involved in removing the crosslinks. Alternatively, the increase in the monomeric muropeptide 2 and drastic reduction in the dimer MP 7 could be a result of disrupting peptidoglycan synthesis, particularly transpeptidase-related genes controlled by $\sigma^F$ and $\sigma^G$ that have yet to be fully characterised and investigated in *C. difficile*.

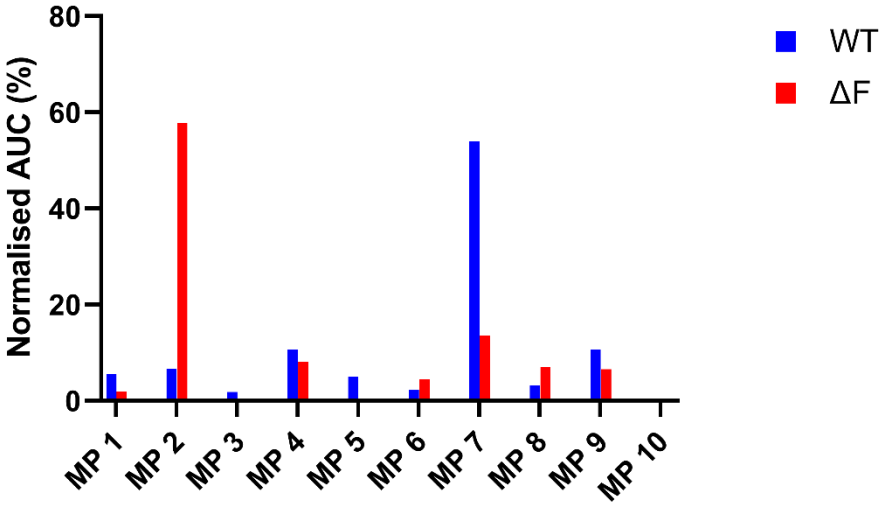


Figure 3-18: Changes to the 10 most abundant muropeptides in *C. difficile* 630 Δ*pyrE Δerm* $P_{tet}$ *spo0A* peptidoglycan in *C. difficile* 630 Δ*erm* Δ*sigF* peptidoglycan The percentage of the normalised area under the curve (AUC) attributed to each muropeptide (MP) is visualised with the same muropeptide in the wild type (WT) for comparison. Muropeptide identities correspond to those in Figure 3-11. WT*; C. difficile 630* Δ*pyrE* Δ*erm* $P_{tet}$ *-spo0A, ΔE; C. difficile 630* Δ*erm* Δ*sigF*. Normalised data is provided in 6.7.Appendix A.

The intricate nature of the $\sigma^E$ and $\sigma^F$ regulons means that interpretation of these results and elucidation of the role of particular enzymes is challenging. Despite this difficulty, analysis of the composition of peptidoglycan of cells stalled at these particular stages provides important insight into how and when peptidoglycan is remodelled, not only to allow engulfment, but also towards preparing cortex formation.

#### 3.2.4.2. *C. difficile 630* Δ*erm* Δ*spoIID* peptidoglycan

Unlike $\sigma^E$, SpoIID is not thought to control the expression of any other genes, though it may have implications for SpoIIQ and SpoIIIAH localisation, as SpoIID interacts with SpoIIQ and SpoIIIAH in a BACTH system (Dembek *et al.*, 2018). Moreover, SpoIID has been shown to at least partially compensate for loss of SpoIIQ (Ribis, Fimlaid and Shen, 2018) which may indicate a yet unidentified role of both proteins.

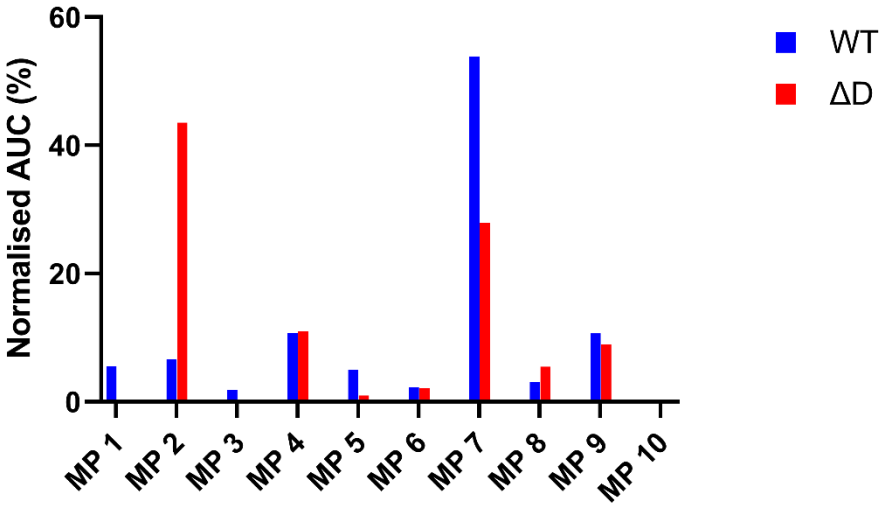


Figure 3-19: Changes to the 10 most abundant muropeptides in *C. difficile* 630 Δ*pyrE* Δ*erm* $P_{tet}$ *spo0A* peptidoglycan in *C. difficile* 630 Δ*erm* Δ*spoIID* peptidoglycan. The percentage of the normalised area under the curve (AUC) attributed to each muropeptide (MP) is visualised with the same muropeptide in the wild type (WT) for comparison. Muropeptide identities correspond to those in Figure 3-11. WT; *C. difficile 630* Δ*pyrE* Δ*erm* $P_{tet}$ *-spo0A*, ΔE; C*. difficile 630* Δ*erm* Δ*spoIID*. Normalised data is provided in 6.7.Appendix A.

It is noteworthy that muropeptides 1, 3 and 10 are not detected by MZmine2 in this mutant strain peptidoglycan (Figure 3-19) despite being identified during manual inspection of the LC-MS data. This highlights the limitations of using an automated analysis software as it will only detect peaks that fit the input parameters. There is a balance to be sought between including peaks that are significant and returning false-positive results. Here, it would appear the parameters in MZmine2 are too stringent. When testing the software with different parameters in order to identify muropeptides 1 ,3 and 10, other artefacts were introduced, including in the analysis of vegetative cells, so the original settings (Table 2-20) were used throughout.

Like peptidoglycan from *sigE* mutants, muropeptide 2 is significantly increased in the Δ*spoIID* mutant. This could be attributed to the continued endopeptidase activity of SpoIIP, for example, and other endopeptidases which may also account for the increased proportion of monomers in this mutant peptidoglycan compared to the wild type (Figure 3-13).

### 3.2.4.3. *C. difficile 630* Δ*erm* Δ*spoIIP* peptidoglycan

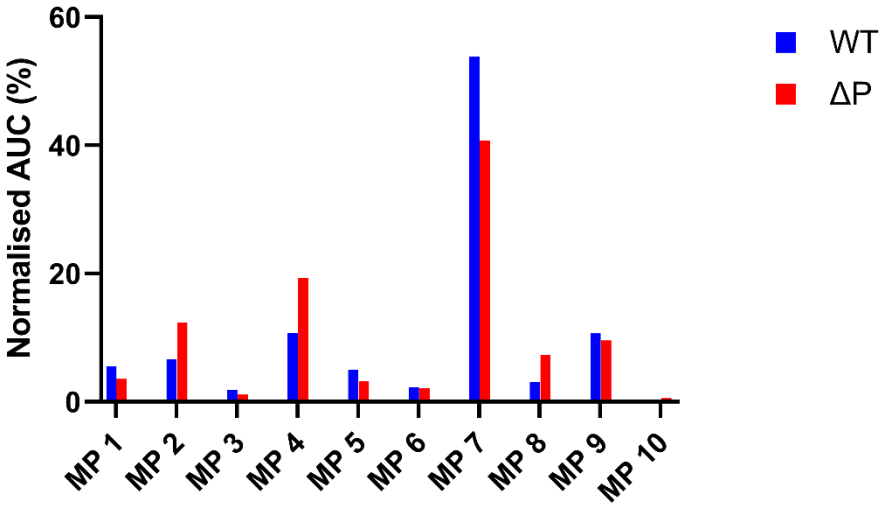


Figure 3-20: Changes to the 10 most abundant muropeptides in *C. difficile* 630 Δ*pyrE Δerm* $P_{tet}$ *spo0A* peptidoglycan in C*. difficile* 630 Δ*erm* Δ*spoIIP* peptidoglycan The percentage of the normalised area under the curve (AUC) attributed to each muropeptide (MP) is visualised with the same muropeptide in the wild type (WT) for comparison. Muropeptide identities correspond to those in Figure 3-11. WT; *C. difficile 630* Δ*pyrE* Δ*erm* $P_{tet}$ *-spo0A*, ΔE; C*. difficile 630* Δ*erm* Δ*spoIIP*. Normalised data is provided in 6.7.Appendix A.

As demonstrated in Figure 3-20 and Figure 3-16, the lack of SpoIIP leads to a peptidoglycan very similar to that of seen in vegetative cells. Indeed, the most significant difference is an increased proportion of muropeptide 4 (GlcN Mur*N*Ac AEmG AEm GlcN Mur*N*Ac). It may be that SpoIIP is the only endopeptidase able to digest muropeptides with Gly termini for example, thus leading to an accumulation of muropeptide 4.

The similarity with vegetative cell peptidoglycan composition is consistent with what has been reported in Δ*spoIIP* mutants; sporulation is arrested very early and cells cannot progress beyond forming asymmetric septa (Dembek *et al.*, 2018). Together, these observations seem to indicate that lack of amidase and endopeptidase activity of SpoIIP prevents engulfment and the accompanying peptidoglycan remodelling. Indeed, it has been proposed that one of the main drivers of engulfment is the cleavage of new peptidoglycan by SpoIIP and SpoIID (Ojkic *et al.*, 2016). If so, and as SpoIID activity depends on SpoIIP, absence of this protein halts sporulation before engulfment, and possibly the related peptidoglycan synthesis, is initiated.

### 3.2.4.4. *C. difficile 630* Δ*erm* Δ*spoIIQ* peptidoglycan

*C. difficile* SpoIIQ contains a LytM endopeptidase domain, a domain that is present in many peptidoglycan digestion enzymes (Meisner et al., 2008; Camp & Losick, 2008). It has been hypothesised therefore, that SpoIIQ may act as an endopeptidase during engulfment (Crawshaw *et al.*, 2014; Serrano *et al.*, 2016). Whilst the enzymatic activity, if any, of SpoIIQ in *C. difficile* is yet to be established, it is known that SpoIIQ forms a complex with SpoIIIAH

(Meisner et al., 2008), and deletion of *spoIIQ* in *C. difficile* results in arrest of engulfment and the formation of membrane bulges (Serrano *et al.*, 2016; Ribis, Fimlaid and Shen, 2018).

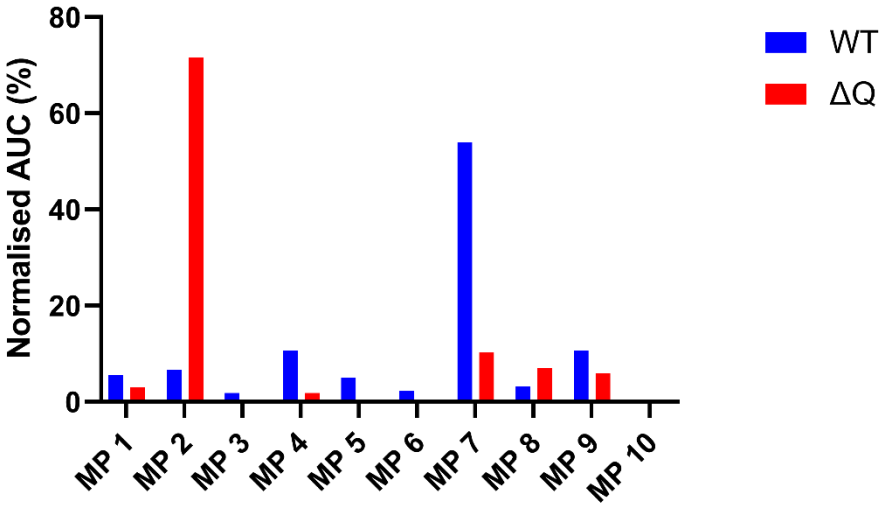


Figure 3-21: Changes to the 10 most abundant muropeptides in *C. difficile* 630 Δ*pyrE* Δ*erm* $P_{tet}$ *spo0A* peptidoglycan in *C. difficile* 630 Δ*erm* Δ*spoIIQ* peptidoglycanThe percentage of the normalised area under the curve (AUC) attributed to each muropeptide (MP) is visualised with the same muropeptide in the wild type (WT) for comparison. Muropeptide identities correspond to those in Figure 3-13. WT; *C. difficile 630* Δ*pyrE* Δ*erm* $P_{tet}$ *-spo0A*, ΔE; C*. difficile 630* Δ*erm* Δ*spoIIQ*. Normalised data is provided in 6.7.Appendix A.

Crosslinking is much reduced in peptidoglycan isolated from this mutant and is accompanied by an increase in monomers, as demonstrated by the increase in muropeptide 2 at the presumed expense of muropeptide 7 for example, as demonstrated in Figure 3-16 and Figure 3-21. This is likely the result of continued SpoIIP, and potentially other unidentified endopeptidases, activity. If this result were solely due to SpoIIP activity, one would expect to see a peak corresponding to the AEmA peptide on the chromatograph, though this is absent. This could suggest several things: digestion is incomplete, and these peptide stems are not released, that they are released but are not observed in these HPLC-MS conditions, or that they are released but quickly recycled and therefore don't accumulate.

As both SpoIID and SpoIIP are expected to be active in a Δ*spoIIQ* mutant, it is surprising that there is not a large rise in the final SpoIID and SpoIIP products (disaccharides terminating in a 1,6-anhydro-Mur*N*Ac residue, and small peptide stems respectively (Abanes-De Mello *et al.*, 2002; Morlot *et al.*, 2010; Dembek *et al.*, 2018). However, the formation of membrane bulges in SpoIIQ mutants seems to indicate that peptidoglycan degradation is affected whilst peptidoglycan synthesis continues. Moreover, SpoIIQ has been shown to interact with SpoIID and SpoIIP (Dembek *et al.*, 2018), suggesting that lack of SpoIIQ could at least partially affect the activity of these peptidoglycan hydrolases. Therefore, the differences seen in the peptidoglycan of these mutants, particularly the reduction of crosslinked muropeptide 7 and increase muropeptide 2 could be a result of a combination of only partial SpoIIP activity and continued peptidoglycan synthesis.

#### 3.2.4.5. *C. difficile 630* Δ*erm* Δ*spoIIIAH* peptidoglycan

SpoIIIAH is not thought to have any enzymatic activity, but forms part of the Q:AH complex, proposed to enable mother cell-forespore communication and stabilise the double membrane system surrounding the forespore (Morlot and Rodrigues, 2018).

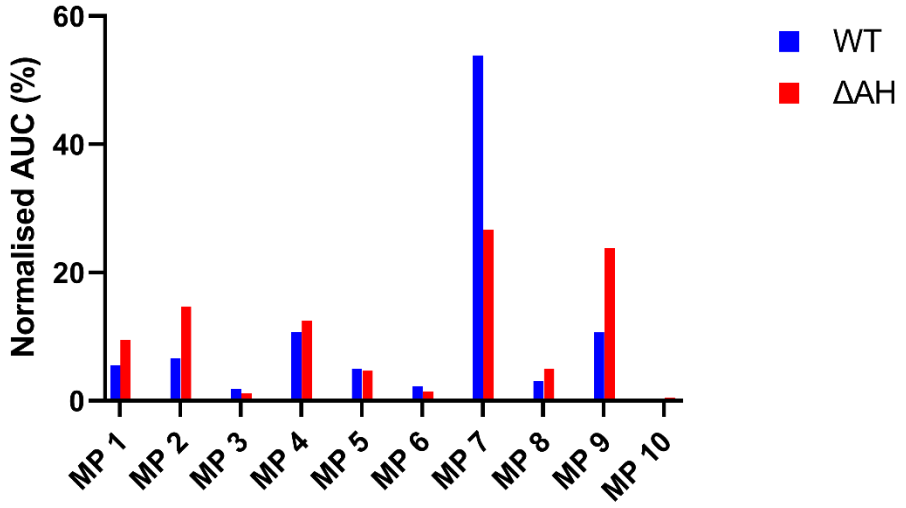


Figure 3-22: Changes to the 10 most abundant muropeptides in *C. difficile* 630 Δ*pyrE* Δ*erm* $P_{tet}$ *spo0A* peptidoglycan in *C. difficile* 630 Δ*erm* Δ*spoIIIAH* peptidoglycan The percentage of the normalised area under the curve (AUC) attributed to each muropeptide (MP) is visualised with the same muropeptide in the wild type (WT) for comparison. Muropeptide identities correspond to those in Figure 3-13. WT; *C. difficile 630* Δ*pyrE* Δ*erm* $P_{tet}$ *-spo0A*, ΔE; C*. difficile 630* Δ*erm* Δ*spoIIIAH*. Normalised data is provided in 6.7.Appendix A.

Again, muropeptides 2 and 7 are the most altered; muropeptide 2 is increased whilst 7 is reduced compared to the wildtype (Figure 3-22). This could also be attributed to continued SpoIIP activity. That the overall degree of crosslinking is largely unchanged in comparison to the wildtype suggests the changes to muropeptides 2 and 7 are masked by changes to the other muropeptides. Deletion of *spoIIIAH* leads to inverted septa (Serrano *et al.*, 2016), preventing sporulation progressing at a very early stage, therefore it is not surprising that the *spoIIIAH* mutant is similar to the vegetative cells, as discussed for the SpoIIQ mutant above.

## 3.3. Discussion

### 3.3.1. Vegetative *C. difficile 630* Δ*pyrE* Δ*erm* $P_{tet}$-*spo0A* peptidoglycan

The major components of exponentially growing *C. difficile* peptidoglycan had been described prior to this work commencing (Peltier *et al.*, 2011). One of the aims of this work was to improve on this characterisation, by improving the resolution of the HPLC to allow a more complete understanding of peptidoglycan composition in *C. difficile*.

The muropeptides identified in this experiment corroborate those previously described (Peltier *et al.*, 2011; Bern, Beniston and Mesnage, 2016), and extend the list with novel muropeptides (Table 3-3). Muropeptides containing glutamine in place of glutamic acid are described for the first time, as is the use of GABA as a terminal amino acid. This suggests yet another layer of heterogeneity to peptidoglycan. The significance of these termini *in vivo* is unclear however, due to the low abundance of these muropeptides. Future work should involve confirming the prevalence of these novel muropeptides, both in different strains and at different phases of growth. Investigation into the impact of the growth media on muropeptide termini should also be investigated.

Whilst the analysis of Bern *et al.* (2016) did not identify any muropeptides containing 1,6-anhydro-Mur*N*Ac, this analysis confirms previous observations where 1,6-anhyrdo-Mur*N*Ac was observed (Peltier *et al.*, 2011). A previous analysis identified 2 muropeptides, both dimers, containing 1,6-anhyrdo-Mur*N*Ac; one with AEmA AEm stem peptides and the second with an additional terminal alanine (Peltier *et al.*, 2011). Both muropeptides contained GlcN in their backbones. In the current analysis, 3 muropeptides, two monomers and one dimer, were identified as containing 1,6-anhydro-Mur*N*Ac. In contrast to the dimers seen by Peltier et al. (2016), the 1,6-anhyrdo-Mur*N*Ac-containing muropeptides identified in this experiment contained at least one Glc*N*Ac residue; the monomer contained only Glc*N*Ac, whereas the dimer contains one Glc*N*Ac residue and one GlcN residue. This difference may be due to the peptidoglycan itself; Peltier et al. (2016) used exponentially growing *C. difficile* 630 cultures, whereas we used stationary phase *C. difficile* 630 Δ*pyrE* Δ*erm* $P_{tet}$-*spo0A* cultures. On the one hand it may be the case that the culture used in the Peltier *et al*. (2016) analysis contained a proportion sporulating cells, which would be expected at this growth phase. Our experiment was designed to ensure no sporulating cells would be present by using a strain where the master regulator of sporulation *spo0A* is not active as anhydrotetracycline was not included in the growth conditions. This observed difference could also be a consequence of the growth phase at harvesting or different growth conditions between solid, nutrient rich BHI media (Peltier *et al.*, 2011), liquid, nutrient rich TY media (Bern, Beniston and Mesnage, 2016) and liquid, nutrient-limited SM used here. If this experiment were to be repeated, it would be interesting to harvest the peptidoglycan from exponential phase *C. difficile* 630 Δ*pyrE* Δ*erm* $P_{tet}$-*spo0A* cultures grown on solid rich media.

During this work, a further study of *C. difficile* vegetative cell peptidoglycan was completed where novel peptide stem termini were identified (Bern, Beniston and Mesnage, 2016). Work presented here offers improved HPLC resolution, confirms some of the termini previously identified (Bern, Beniston and Mesnage, 2016) and identifies novel muropeptides, therefore furthers our understanding of *C. difficile* vegetative cell peptidoglycan.

### 3.3.2. Peptidoglycan changes during early engulfment

The profiles of peptidoglycan composition of all mutant strains analysed were very similar to the vegetative cell peptidoglycan and the changes to the proportions of muropeptides were more modest than anticipated. As $\sigma^E$ and $\sigma^F$ are responsible for the transcription of multiple genes, the effect their absence has on peptidoglycan remodelling would be intricate. In any case, the small differences observed indicate that some peptidoglycan remodelling is required, even at early stages of sporulation. It also provides some indication of other potential enzymatic activities involved in remodelling such as specific deacetylases and synthases. Most notable were the changes in *C. difficile* 630 Δ*erm* Δ*sigE*; several new peaks containing partially or fully acetylated muropeptides were identified; this suggests that at least one peptidoglycan Glc*N*Ac deacetylase is under the control of $\sigma^E$.

As SpoIIP and SpoIID are proposed to be responsible for peptidoglycan hydrolysis required for engulfment (Meyer *et al.*, 2010; Ojkic *et al.*, 2016), it was surprising to see that, overall, changes were not very pronounced. Moreover, as we also know SpoIIQ and SpoIIIAH are required for engulfment and lead to membrane bulging (thought to be related to continued peptidoglycan synthesis in the absence of peptidoglycan digestion), we also expected a more pronounced effect in these mutants. It should be noted that all the mutants are arrested at early stages of engulfment and there could be other factors affecting peptidoglycan composition that are altered in the absence of these proteins.

Taken together, these results suggest that we have yet to identify all contributors to peptidoglycan remodelling during engulfment.

### 3.3.3. Future work

In the future, this experiment could be repeated with several improvements. One confounding aspect of this analysis is the presence of vegetative cells in the samples analysed. Moreover, the low efficiency of sporulation of the mutants (Table 3-4) means that, in some cases, vegetative cell peptidoglycan will dominate the analysis.

Table 3-4: Asymmetric septa formation by various *C. difficile* strains

| Strain | Cells that have produced asymmetric septa (%) |
|---|---|
| 630 Δ*erm* Δ*pyrE* $P_{tet}$ *-spo0A* | 0 |
| 630 Δ*erm* Δ*sigE* | 79 |
| 630 Δ*erm* Δ*sigF* | 35 |
| 630 Δ*erm* Δ*spoIID* | 28 |
| 630 Δ*erm* Δ*spoIIP* | 28 |
| 630 Δ*erm* Δ*spoIIQ* | 22 |
| 630 Δ*erm* Δ*spoIIIAH* | 47 |

The percentage of cells that had undergone asymmetric septa was assessed by fluorescence microscopy after 16 hours of growth in SM. Percentages were calculated by dividing the number of cells producing asymmetric septa in a field of view by the total number of cells in that field of view and multiplying by 100. All microscopy and data processing were undertaken by Dr M Dembek.

Separating vegetative from sporulating cells prior to peptidoglycan isolation and analysis would allow a more definite characterisation of the changes to peptidoglycan composition. Taking advantage of well-established fluorescence systems (Serrano *et al.*, 2016) to label sporulation-specific proteins would allow separation of vegetative cells and those that have undergone asymmetric septation by flow cytometry. Sorting the cells will enrich the samples, reducing the contribution of vegetative cell peptidoglycan to the total extracted peptidoglycan. The mass spectra of the vegetative cells could then be subtracted from the spectra of the mutant strains, leaving only those peaks that represent the muropeptides observed in the mutant peptidoglycan.

It would be interesting to carry out similar experiments with strains lacking $\sigma^G$ and $\sigma^K$, which control later sporulation events (Saujet *et al.*, 2014). Although engulfment has been completed at this point, cortex formation happens at this stage, so more pronounced differences to vegetative cells are expected. Comparing the composition at these stages with spore cortex peptidoglycan will also provide interesting insights into peptidoglycan remodelling during spore formation.

# Chapter 4. Characterisation of *C. difficile* sSpoIID and sSpoIIP

## 4.1. Introduction

Both *spoIID* and *spoIIP* are found across endospore formers (Galperin *et al.*, 2012; Abecasis *et al.*, 2013; Kelly and Salgado, 2019) and are key in engulfment; *B. subtilis* mutants lacking either gene cannot form heat resistant spores (Eichenberger, Fawcett and Losick, 2002) and phenotypically form characteristic membrane bulges (Abanes-De Mello *et al.*, 2002; Meyer *et al.*, 2010).

Investigations into the enzymatic activities of SpoIID and SpoIIP have been previously conducted using *B. subtilis* enzymes. *B. subtilis* and *C. difficile* SpoIID share 52% sequence similarity (36% identity) with the SpoIIP from the 2 organisms sharing 40% similarity (27% identity) as determined using EMBOSS Matcher (Waterman and Eggert, 1987). SpoIIP is predicted to vary between Bacilli and Clostridia; for example, *B. subtilis* SpoIIP is predicted to contain a N-terminal transmembrane anchor, whereas *C. difficile* SpoIIP is predicted to contain a signal peptide (Kelly and Salgado, 2019). Furthermore, conservation of the proposed zinc binding residues of SpoIID is low, even amongst Clostridia (Kelly and Salgado, 2019).

*B. subtilis* SpoIIP acts as an amidase and endopeptidase, removing the peptide stems from peptidoglycan to produce the long glycan strands that are the substrate for SpoIID transglycosylase activity (Chastanet and Losick, 2007; Morlot *et al.*, 2010). This sequential activity has been proposed to prevent the release of free muropeptides, which have been implicated in spore germination (Shah *et al.*, 2008). However, the activities of *C. difficile* SpoIID and SpoIIP, and their catalytic residues, had yet to be confirmed.

Work in this chapter aimed to purify and characterise *C. difficile* SpoIID and SpoIIP, verify their activities *in vitro*, and to further the understanding of the enzymatic mechanisms employed by these proteins. Concurrently with this work, the role of *spoIID* and *spoIIP* were demonstrated *in vivo* in *C. difficile*; deletion of either gene leads to the arrest of sporulation and prevents the formation of mature spores (Dembek *et al.*, 2018; Ribis, Fimlaid and Shen, 2018).

## 4.2. Purification and biophysical characterisation

TMHMM 2.0 (Moller, Croning and Apweiler, 2001) was used to identify predicted transmembrane regions of SpoIID and SpoIIP. Cloning of both genes was undertaken in such a manner that the identified potential membrane regions were omitted, resulting in $SpoIID_{26\text{-}354}$ and $SpoIIP_{27\text{-}399}$, referred to as sSpoIID and sSpoIIP, respectively, throughout the rest of this thesis. The theoretical masses of these constructs were calculated using the ProtParam server (https:/web.expasy.org/protparam). The resulting strains were used as a template to produce point mutants in sSpoIID and sSpoIIP. The constructed plasmids were verified by sequencing before transformation into expression strains.

### 4.2.1. Purification

All sSpoIID and sSpoIIP constructs were purified in the same manner, as described in section 2.5. Once the protocol was established protein purifications were also routinely undertaken by Drs Anna Barwinska-Sendra and Will Stanley.

Following the initial affinity sSpoIID purification, fractions eluting during a peak in UV measurement were assessed by SDS-PAGE (Figure 4-1A).The ProtParam server estimated His-tagged sSpoIID to be 41.682 kDa, therefore the fractions containing a protein of ~40 kDa (fractions 4-6) were pooled and incubated overnight with TEV to remove the His-tag. Tag-free sSpoIID was recovered from the flow through of a second IMAC purification before concentration and application to a pre-equilibrated SEC column (Figure 4-1B). The theoretical size of tag-free sSpoIID, calculated using Protparam, is 37.684 kDa, therefore fractions corresponding to an increase in absorbance, which contained a protein ~38 kDa in size, were pooled and concentrated. Using calibration curves (produced by Paola Lanzoni-Manguchi) and the elution profile of sSpoIID, the estimated size of sSpoIID is approximately 40 kDa. This suggests sSpoIID was purified as a monomer.

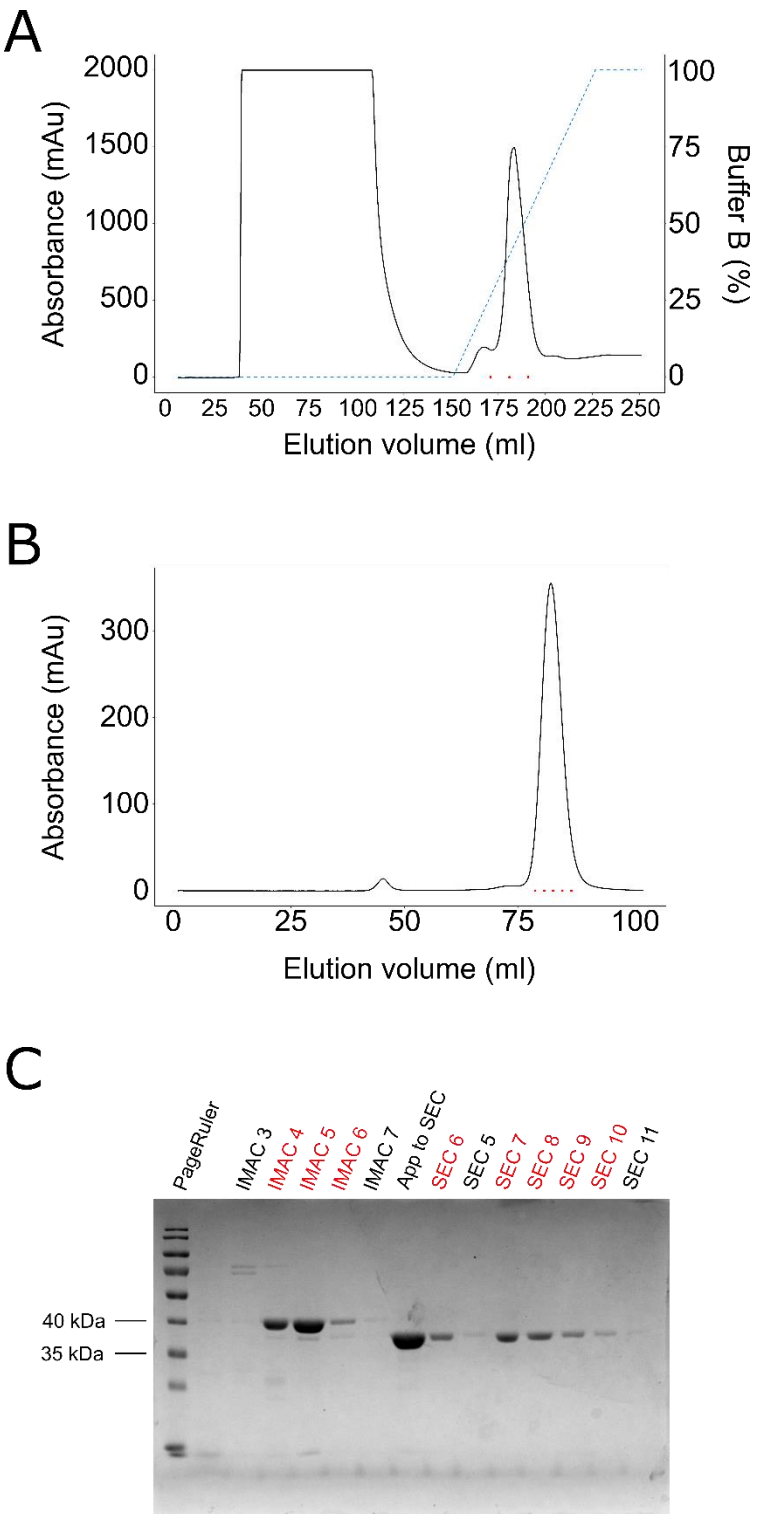


Figure 4-1: Purification of sSpoIID.

sSpoIID was purified as described in section 2.5; by IMAC (A), where absorbance is shown in black, buffer B concentration as a dashed blue line, and the fractions pooled and carried onto SEC as red dots. Buffer B is 50 mM Tris base pH 8.0, 300 mM NaCl, 250 mM imidazole. The SEC chromatograph (B) is labelled as the IMAC chromatograph, with fractions marked with red dots analysed by SDS-PAGE (C), demonstrating successful His-tag removal, that the protein is pure, and is not subjected to degradation during purification. The fractions pooled at each stage are marked in red in panel C.

The purification of sSpoIID$_{E101A}$, sSpoIID$_{C140A}$, sSpoIID$_{C146A}$, sSpoIID$_{H134A}$ and sSpoIID$_{H145A}$ was carried out following the same protocol as used for sSpoIID (Figure 4-2). All constructs behaved in the same manner, eluting as monomers to homogeneity. The results of sSpoIID$_{E101A}$ purification is provided as an example in Figure 4-2.

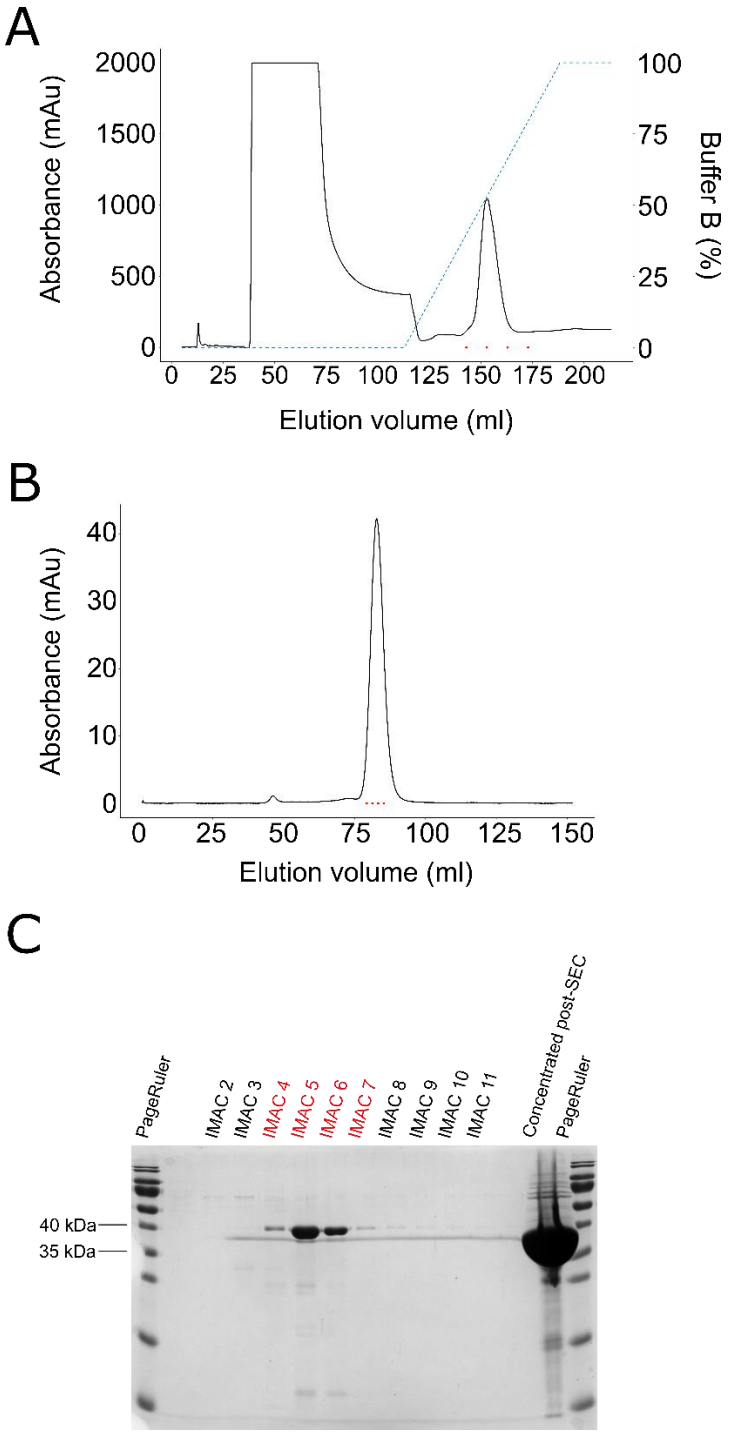


Figure 4-2: sSpoIID$_{E101A}$ purification.

sSpoIID$_{E101A}$ was purified as described in section 2.5. (A) IMAC purification, (absorbance black line, buffer B concentration dashed blue line and fractions pooled and carried forward are marked as red dots) (B) the SEC chromatograph and is labelled as panel A. IMAC fractions marked in red in (C) were pooled and incubated with TEV prior to application to SEC. Panel C shows the SDS-PAGE analysis of IMAC fractions and the protein post-SEC. Buffer B is 50 mM Tris base pH 8.0, 300 mM NaCl, 250 mM imidazole.

sSpoIIP was purified as described in section 2.5. ProtParam estimated His-tagged sSpoIIP to be 38.951 kDa and the tag-free sSpoIIP to be 35.953 kDa. However, both forms of sSpoIIP migrate less than expected on SDS-PAGE, closer to the 40 kDa than the 35 kDa marker (Figure 4-3A). Despite this, the difference between SEC and IMAC fractions, when analysed by SDS-PAGE, indicates that TEV cleavage was successful, so fractions containing protein were pooled and concentrated before storage at -80°C.

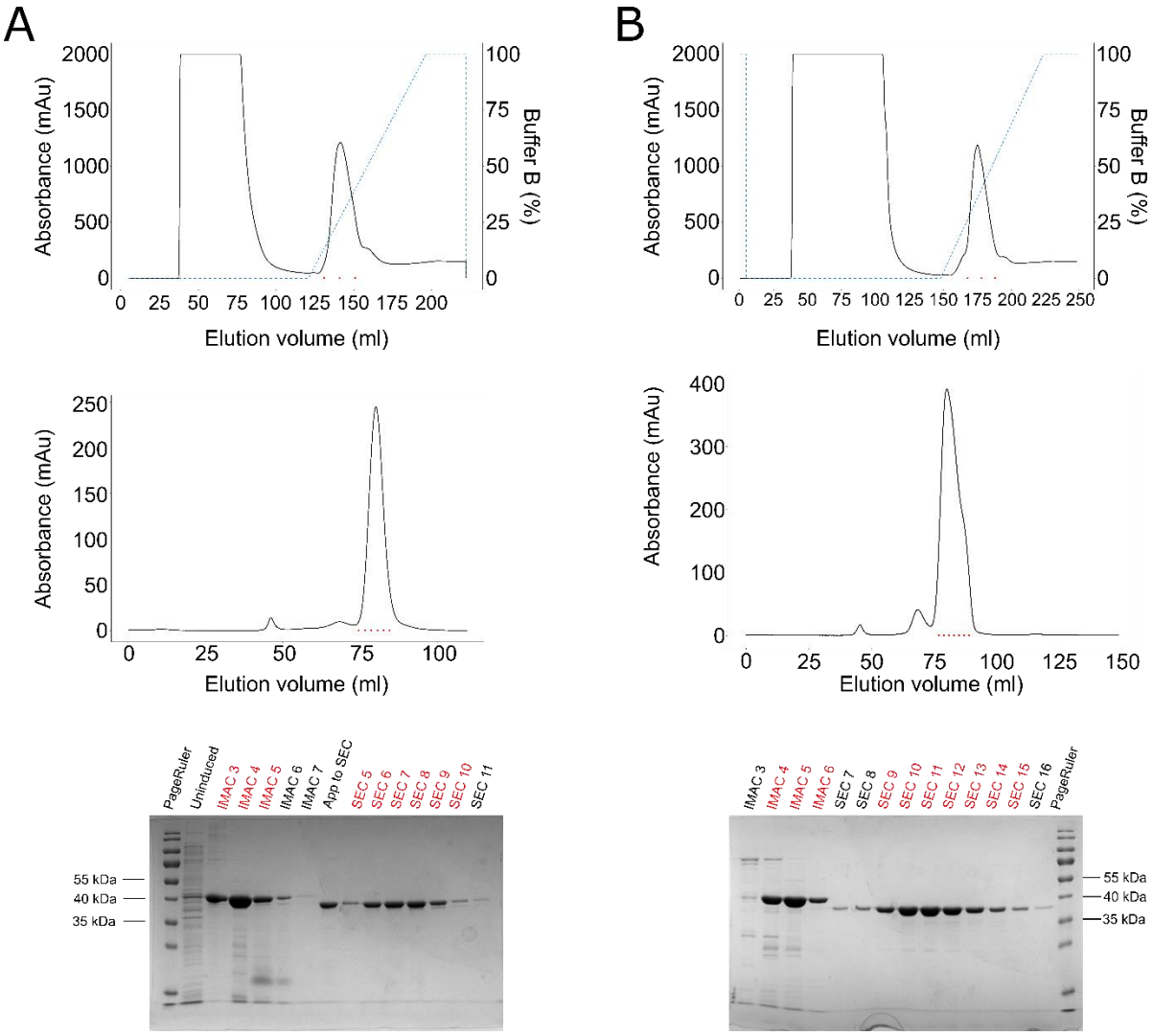


Figure 4-3: sSpoIIP and sSpoIIP$_{H142R}$ purifications

sSpoIIP (panel A), sSpoIIP$_{H142R}$ (panel B) were purified as described in section 2.4.3.1. Top panels show the IMAC purification (absorbance blank line, buffer B concentration dashed blue line and fractions pooled and carried forward to marked as red dots). Buffer B is 50 mM Tris base pH 8.0, 300 mM NaCl, 250 mM imidazole. The middle panels show the SEC chromatograph and are labelled as the IMAC chromatograph. The bottom panel shows the SDS-PAGE analysis of the labelled fractions, those labelled in red were pooled before continuing with the purification. sSpoIIP$_{H222R}$ was purified as sSpoIIP$_{H142R}$.

Western blot analysis, using either anti-SpoIID or anti-SpoIIP antibodies as appropriate (Figure 4-4), suggests that preparations of all proteins were not subject to excessive degradation, and thus suitable for biophysical and enzymatic activity analysis.

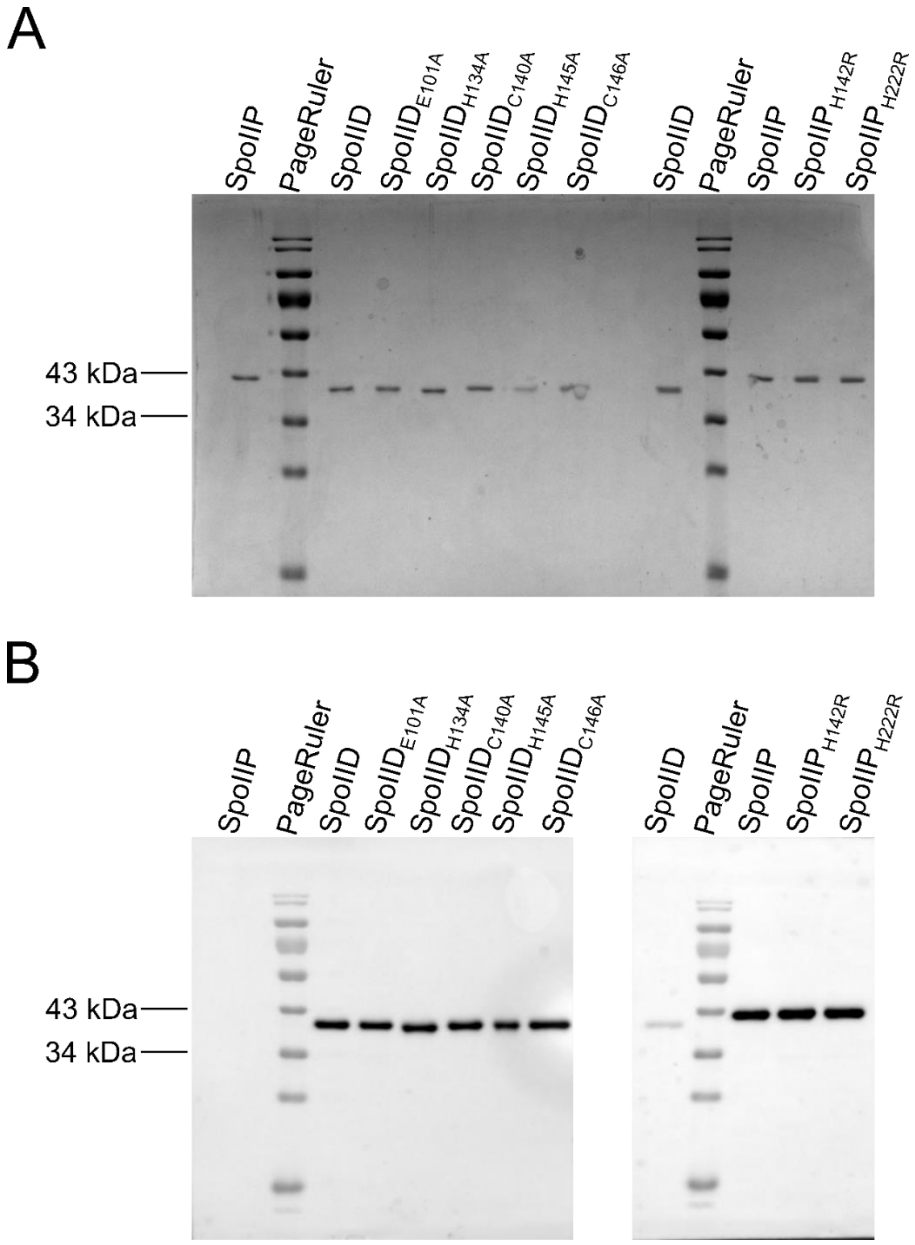


Figure 4-4: SDS-PAGE and western blot analysis of purified proteins.

Two 12% SDS-PAGE gels, each loaded with 100ng of each protein, were run in parallel, and one was further analysed by immunoblot. Panel A shows the SDS-PAGE gel used to resolve sSpoIID and its mutants (right) and sSpoIIP and its mutants (left). The second gel was then split and used for western blots (B). Blots were probed with rabbit anti-sSpoIID (left) or anti-sSpoIIP (right) antibodies at a 1:15,000 dilution, before conjugation to HRP-conjugated secondary antibodies. The sSpoIID antibody is specific to sSpoIID, showing no cross-reactivity to sSpoIIP. The sSpoIIP antibody, however, cross-reacts partially with sSpoIID, though the size difference between sSpoIID and sSpoIIP demonstrates that the antibody preferentially binds sSpoIIP. Both western blots demonstrate the proteins are pure and are not subject to degradation or aggregation.

To verify the oligomeric state of sSpoIID and sSpoIIP, purified protein was further analysed using the analytical Superdex 200 GL 10/300 Increase column in the presence of 15mM DTT. DTT is a reducing agent that prevents the formation of disulphide bonds and thus reduces the likelihood of intermolecular interactions. For sSpoIID, comparison of the DTT untreated and treated chromatographs (Figure 4-5A) suggests there is little effect of DTT. In the case of sSpoIIP, the small peak eluting at approximately 12 ml (Figure 4-5B upper panel) is much reduced when the sample was treated with DTT (Figure 4-5 lower panel), suggesting that it may correspond to an oligomeric form of sSpoIIP mediated by disulphide bonds. Using elution volumes and calibration curves produced by Dr Anna Barwinska-Sendra and Paola Lanzoni-Manguchi, sSpoIIP is estimated to be ~52.6 kDa (significantly larger than estimated by Protparam), and the small peak eluting at ~12 minutes is estimated to be 147.53 kDa, suggesting this small bump may represent either dimerisation or trimerisation of sSpoIIP. Estimation of protein mass using gel filtration relies on the assumption that a protein is globular in shape; a non-globular protein of a given size will travel through the column differently to a globular protein of the same size and elute at a different volume, and thus the estimation of their size will be affected (Sahin and Roberts, 2012). In this case, sSpoIIP is estimated to be ~53 kDa by gel filtration methods, whilst the predicted mass by Protparam is ~36kDa. Protparam predicts the mass of a protein by summing the mass of each amino acid minus the mass of a water molecule per peptide bond formed (Gasteiger *et al.*, 2005).

As the majority of purified protein appeared to be in a monomeric state, the purification protocol was considered to be appropriate., and the Protparam derived estimated size of ~36 kDa used throughout.

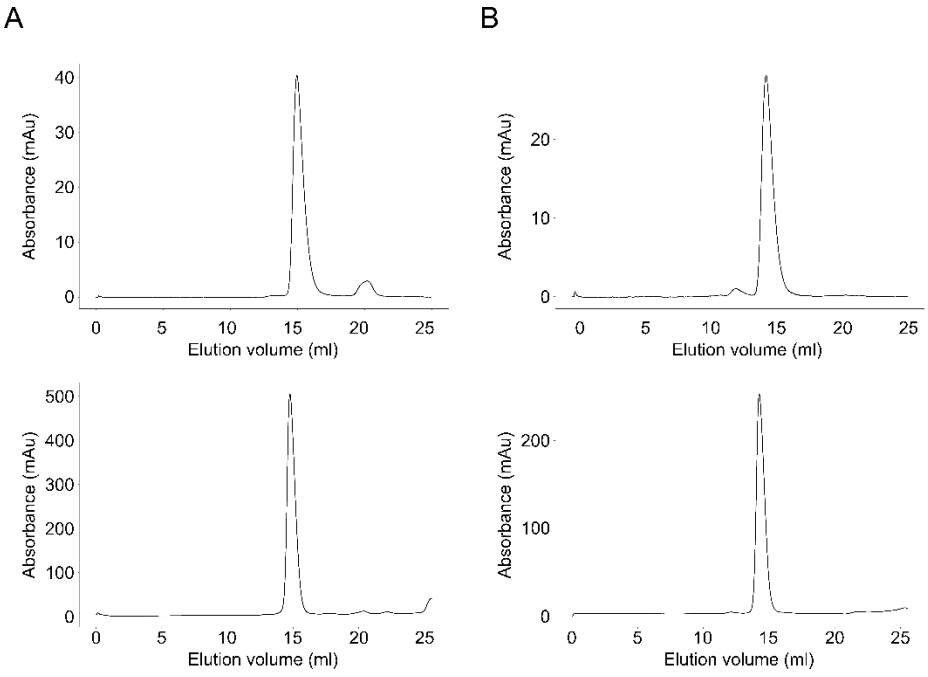


Figure 4-5: Purification of sSpoIID and sSpoIIP in the presence of 15mM DTT

sSpoIID (A) and sSpoIIP (B) were further purified using an analytical grade SEC column (top panels). sSpoIID and sSpoIIP were also incubated and eluted in buffer supplemented with 15mM DTT (bottom panels). Addition of DTT removes the smaller, earlier eluting peak seen in the sSpoIIP chromatographs (panel B).

### 4.2.2. Characterisation of *C. difficile* sSpoIID and sSpoIIP

Before the enzymatic activities of sSpoIID and sSpoIIP were investigated, the stability of the proteins was assessed by circular dichroism (CD), and the presence/absence of divalent metal ions assessed by inductively coupled plasma mass spectrometry (ICP-MS).

#### 4.2.2.1. Circular dichroism

The CD spectra of sSpoIID, sSpoIIP and all associated mutants were measured and processed as described in section 2.7.1 and are presented in Figure 4-6. The predicted secondary structure of each wild type protein was also assessed using PSIPRED, an online server which predicts the secondary structure of a provided protein sequence (bioinf.cs.ucl.ac.uk/psipred/) (Jones, 1999; Buchan and Jones, 2019), (Table 4-1).

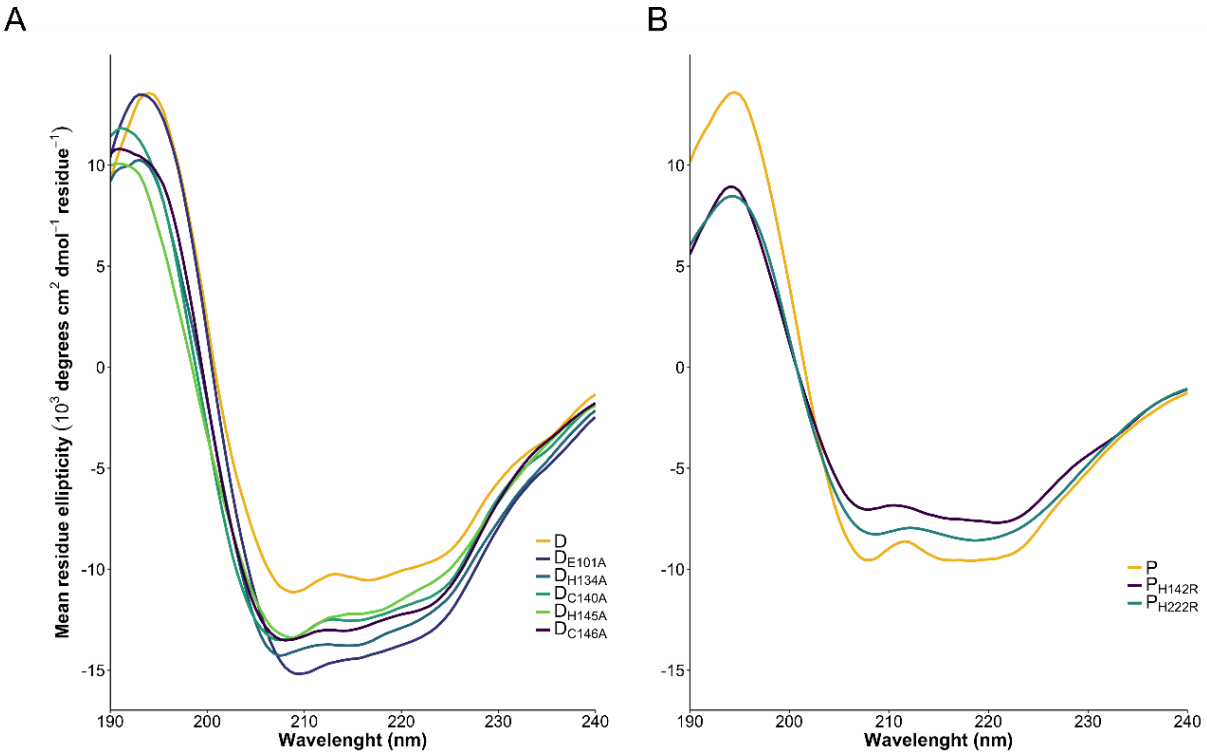


Figure 4-6: Circular dichroism spectra of sSpoIID, sSpoIIP and mutants

Panel A shows the CD spectra of sSpoIID and related mutants after deconvolution by Dichroweb as described in section 2.7.1 and Dembek *et al*., 2018. Panel B shows the spectra associated with sSpoIIP and its mutants. Figure modified from Dembek *et al*., 2018. CD experiments were aided by Dr Anna Barwinska-Sendra. D; sSpoIID, $D_{E101A}$; sSpoIID$_{E101A}$, sSpoIID$_{H134A}$; sSpoIID$_{H134A}$, $D_{C140A}$; sSpoIID$_{C140A}$, $D_{H145A}$; sSpoIID$_{H145A}$, $D_{C146A}$; sSpoIID$_{C146A}$, P; sSpoIIP, $P_{H142R}$; sSpoIIP$_{H142R}$, $P_{H222R}$; sSpoIIP$_{H222R}$.

Table 4-1: Secondary structure prediction and melting temperature of sSpoIID, sSpoIIP and associated mutants

| Protein | Method | Helix (%) | Strand (%) | Coil (%) | Disordered (%) | $T_m$ (°C) |
|---|---|---|---|---|---|---|
| sSpoIID | PSIPRED | 24 | 21 | - | 55 | |
| | CDSSTR | 33 | 17 | 20 | 29 | 60.5 |
| $sSpoIID_{E101A}$ | CDSSTR | 36 | 18 | 20 | 26 | 63.4 |
| $sSpoIID_{C140A}$ | CDSSTR | 33 | 17 | 21 | 29 | 54.5 |
| $sSpoIID_{C146A}$ | CDSSTR | 33 | 17 | 21 | 29 | 57.1 |
| $sSpoIID_{H134A}$ | CDSSTR | 35 | 17 | 22 | 29 | 55.3 |
| $sSpoIID_{H145A}$ | CDSSTR | 28 | 19 | 22 | 31 | 54.5 |
| sSpoIIP | PSIPRED | 25 | 11 | - | 64 | |
| | CDSSTR | 30 | 20 | 20 | 29 | 53.7 |
| $sSpoIIP_{H142R}$ | CDSSTR | 17 | 18 | 26 | 30 | 54.2 |
| $sSpoIIP_{H222R}$ | CDSSTR | 16 | 18 | 25 | 31 | 57.8 |

PSIPRED predictions of secondary structure composition and CD spectra derived values using the CDSSTR algorithm provided by DiChroweb. Melting temperatures, derived using JASCO software, are also provided for each construct.

Visual assessment of the spectra of sSpoIID, its associated mutants (Figure 4-6), and reference spectra suggest the proteins are folded and predominantly α-helical (Greenfield, 2006). This is reflected in the CDSSTR outputs, where 24-36% of each protein is designated as α-helical.

The change in CD of each protein was assessed at 222nm as the temperature was increased from 20 °C to 95 °C to determine the melting temperature of the protein, results are reported in Table 4-1. Mutation of any of the sSpoIID zinc binding sites (H134, C140, H145 or C146) results in a reduction in both ordered domains and in melting temperature, suggesting that zinc binding may play a role in the stability of the enzyme. Interestingly, mutation of the catalytic glutamate residue increases the melting temperature, presumably as the protein becomes more ordered. Similarly, mutation of either of the catalytic residues of sSpoIIP increases the melting temperature.

#### 4.2.2.2. ICP-MS

The cobalt, copper, manganese and zinc occupancy of purified proteins was assessed by ICP-MS as a published SpoIID crystal structure was proposed to coordinate zinc (Nocadello *et al.*, 2016), and SpoIIP may bind zinc based on similarity to CwlV, which can use zinc, cobalt or manganese for activity (Shida *et al.*, 2001). Experiments were performed by either Dr Kevin Waldron or Dr Anna Barwinska-Sendra. Results are summarised in Figure 4-7 and Figure 4-8.

Data collected for sSpoIID and $sSpoIID_{E101A}$ demonstrate that sSpoIID binds zinc in a 1:1 ratio and that glutamate 101 is not involved in metal binding. Mutation of the proposed zinc binding residues C140 or H145 to alanine abolishes zinc binding, whereas C146A and H134A attenuates zinc binding to varying degrees. One would expect that mutation of any of these residues would abolish zinc binding, however, it appears H134 and C146 are less critical for metal coordination. It may be that other nearby residues or ions can compensate for the loss of either of these residues and thus allow some degree of zinc binding. These data confirm

that sSpoIID uses C140, C146, H134 and H145 to coordinate $Zn^{2+,}$ as suggested by Nocadello *et al*. (2016) based on their crystal structures (PDB accession identities: 5l1T and 5TXU). As a result, subsequent peptidoglycan digestion experiments (section 4.4) were performed in the presence of 10mM $ZnCl_2$.

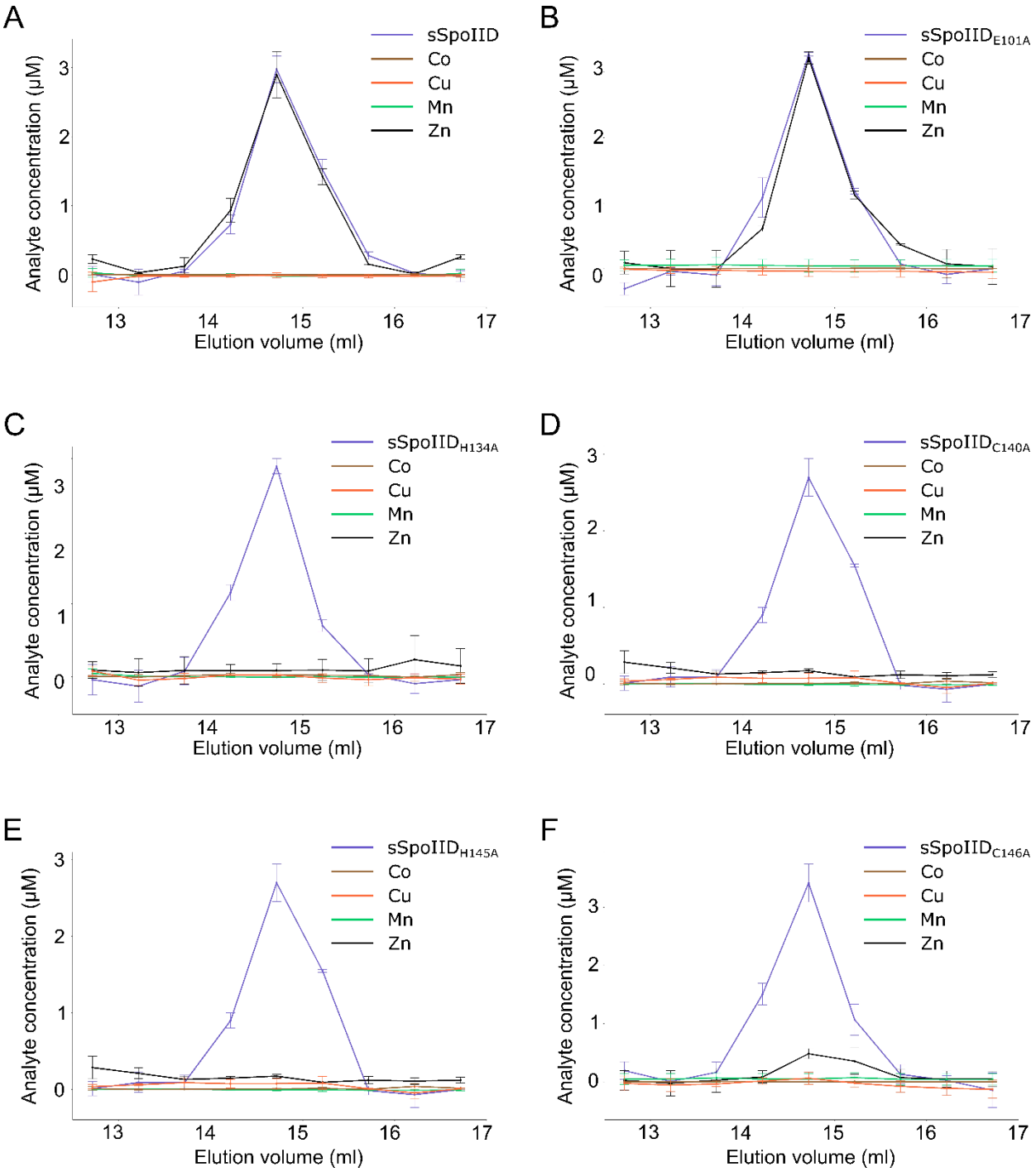


Figure 4-7: ICP-MS analysis of sSpoIID and sSpoIID mutants

sSpoIID (A), sSpoIID$_{E101A}$ (B), sSpoIID$_{H134A}$ (C), sSpoIID$_{C140A}$ (D), sSpoIID$_{H145A}$ (E) and sSpoIID$_{C146A}$ (F) were analysed by ICP-MS. Standard deviations from 2 biological replicates are shown. ICP-MS analysis was undertaken by Dr Anna Barwinska-Sendra and Dr Kevin Waldron.

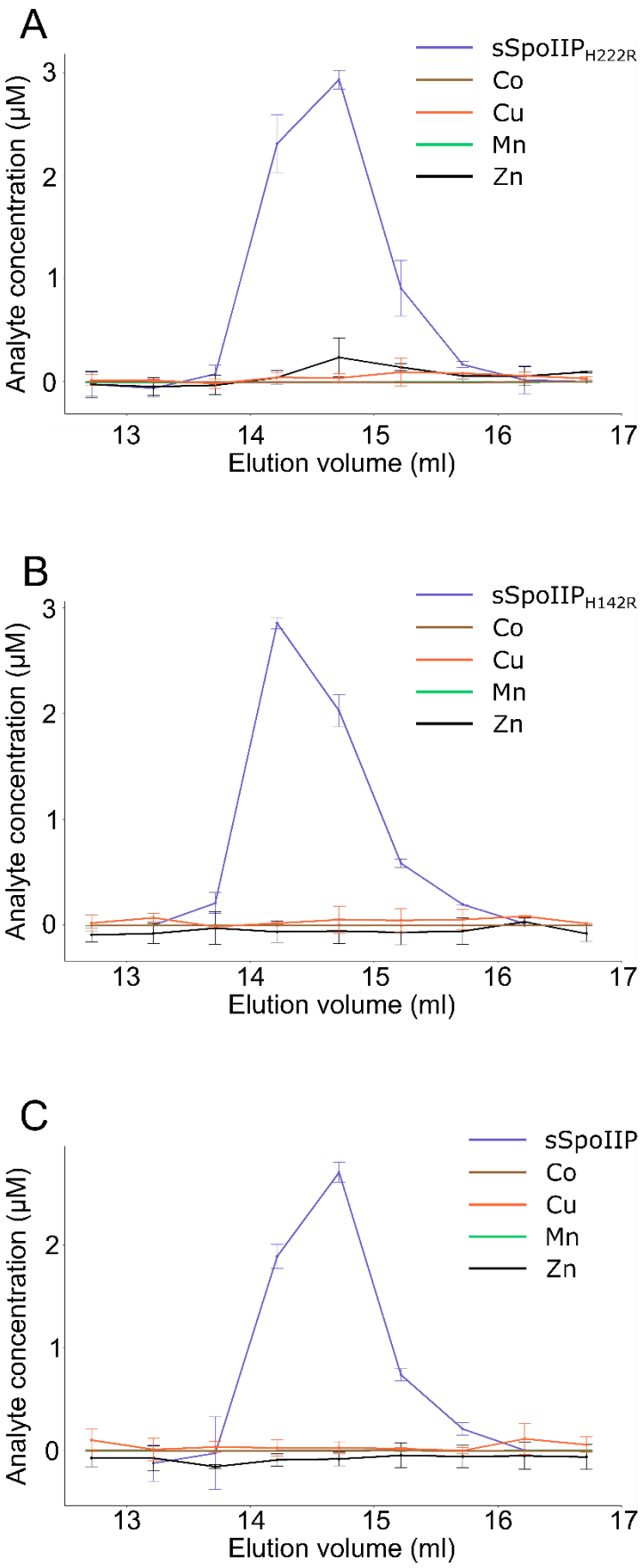


Figure 4-8: ICP-MS analysis of sSpoIIP and sSpoIIP mutants.

sSpoIIP (A), sSpoIIP$_{H142R}$ (B), and sSpoIIP$_{H222R}$ (C) were analysed by ICP-MS. Standard deviations from 2 biological replicates are shown. ICP-MS analysis was undertaken by Dr Anna Barwinska-Sendra.

## 4.3. Crystallisation of sSpoIID and its point mutants

sSpoIID and the E101A variant produced diffraction-quality crystals within 24 hours in Index condition C6: 60% Tascimate™ pH 7.0. E101A was the only sSpoIID mutant to produce crystals, which strengthens the hypothesis that the zinc ion has a structural role, stabilising the protein. Diffraction experiments were undertaken by Dr Arnaud Baslé.

### 4.3.1. Data processing

Data collected was indexed, refined, integrated and scaled using Dials (Gildea *et al.*, 2014). Resulting data collection parameters are detailed in Table 4-2.

Table 4-2: Data collection statistics

| Parameter | $sSpoIID_{E101A}$ |
|---|---|
| Resolution (Å) | 66.1- 2.3 (2.38-2.30) |
| Unit cell dimensions | |
| a, b, c (Å) | 97.6046, 97.6046, 106.054 |
| α, β, γ (°) | 90, 90, 120 |
| Space group | P $3_1$ 2 1 |
| $R_{merge}$ | 0.01947 (0.1471) |
| Total reflections | 52848 (5236) |
| Unique reflections | 26429 (2618) |
| Mean I/σI | 16.42 (3.09) |
| Completeness (%) | 99.96 (100.00) |

Values in parentheses relate to the highest resolution shell.

Data scaled to 2.3 Å was input into CCP4i2 (Potterton *et al.*, 2018), where unit cell content was estimated using the Matthews coefficient (Kantardjieff and Rupp, 2003), which suggested either 1 or 2 molecules of $sSpoIID_{E101A}$ could be present in the asymmetric unit (Table 4-3).

Table 4-3: Cell content of $sSpoIID_{E101A}$ crystal

| Nmol | Solvent volume (%) | Matthews coefficient | Probability |
|---|---|---|---|
| 1 | 66 | 3.65 | 0.55 |
| 2 | 33 | 1.82 | 0.45 |

CCP4i2 was used to estimate the cell contents of the crystal. The probabilities of one or two molecules (Nmol) in the asymmetric unit were similar.

### 4.3.2. Molecular replacement, refinement and validation

Within CCP4i2, Phaser (McCoy *et al.*, 2007) was provided with chain A of a previously solved SpoIID structure (PDB ID: 5TXU) (Nocadello *et al.*, 2016) and used for molecular replacement. Phaser was run twice; once with 1 molecule per asymmetric unit, and once with 2 molecules, based on the analysis of the cell content (Table 4-3).

Using Phaser to search for 2 molecules per asymmetric unit results in a clash within the unit cell. Molecular replacement searches with one molecule per asymmetric unit resulted in a

good solution, as indicated by the absence of clashes as well as a TFZ score of 16.8 and initial and final log-likelihood gain scores of 317 and 6622 respectively.

The resulting map and model were then refined using REFMAC5 (Vagin *et al.*, 2004) in CCP4i2. Iterative cycles of manual model building in Coot (Emsley *et al.*, 2010) and refinement were carried out until $R_{work}$ and $R_{free}$ converged. MolProbity (Chen *et al.*, 2010) was used to assess the quality of the resulting model. Final refinement and model validation where carried out using Phenix (Adams *et al.*, 2010, 2011). The final model statistics are show in Table 4-4.

Table 4-4: sSpoIID$_{E101A}$ refinement statistics

| Parameter | sSpoIID$_{E101A}$ |
|---|---|
| $R_{work}$ (%) | 0.18 (0.22) |
| $R_{free}$ (%) | 0.21 (0.28) |
| No. of non-H atoms | 2372 |
| No. of macromolecules | 2258 |
| No. of solvent atoms | 113 |
| Protein residues | 283 |
| Ligands | 1 |
| RMSD | |
| Bond angle ($^{O}$) | 0.81 |
| Bond length (Å) | 0.007 |
| Average B factor (Å$^2$) | 58.03 |
| Ramachandran plot, residues in | |
| Most favoured regions (%) | 96.80 |
| Allowed (%) | 3.20 |
| Outliers (%) | 0.00 |

Refinement statistics of the final SpoIID$_{E101A}$ model, provided by Phenix after multiple rounds of refinement using REFMAC5 within CCP4i2 and a final refinement cycle within Phenix.

### 4.3.3. Model of sSpoIID$_{E101A}$

The refinement statistics (Table 4-4) are summarised by Phenix during model validation into a Polygon graph (Urzhumtseva *et al.*, 2009) (Figure 4-9). The Polygon graph allows comparison between our model and a predetermined set of models, of similar resolution, in the PDB. A good model will have a small, roughly equal polygon shape and be within the bounds of the "spokes".

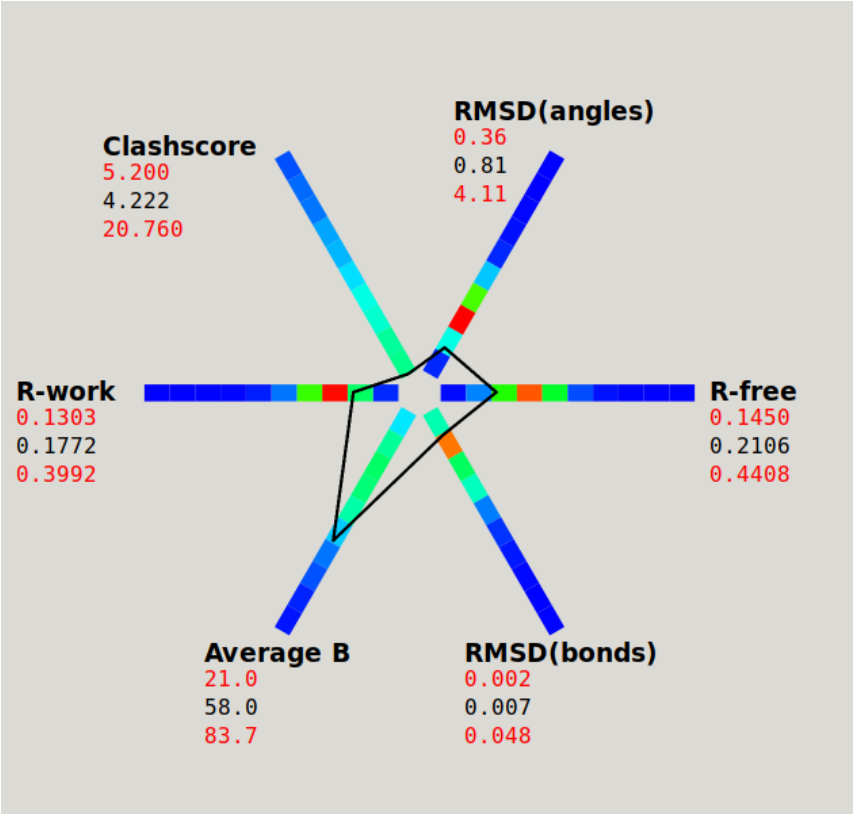


Figure 4-9: Polygon of sSpoIID$_{E101A}$

Within Phenix, a Polygon graph was produced to allow rapid assessment of key model parameters. Root mean square deviation (RMSD) of both the angles and bond lengths are given, along with $R_{free}$ and $R_{work}$, the average B factor and the clash score. Colours correspond to the frequency with which the value of a given parameter is observed within the predefined PDB set with red being a rare value, green usual and blue often seen.

As seen in Figure 4-9, the polygon generated from our data suggests a good solution, with the only outlier being the average B factor, though this is still within acceptable ranges.

$R_{work}$ and $R_{free}$ (calculated from 5% non-refined reflections) differ by 0.03%, which is well within the ±0.05 difference considered acceptable for a good model (Kleywegt and Jones, 1997). The root mean square deviation (RMSD) of the bond angles and lengths suggest the mean bond angle and length in this model are within the ranges of those published in the PDB, which reflects a favourable geometry. Clashes refers atoms being too close which would indicate poor positioning, this model has no such clashes. Furthermore, the model has no Ramachandran outliers, with the majority of residues in favoured conformations.

The model is presented in Figure 4-10.

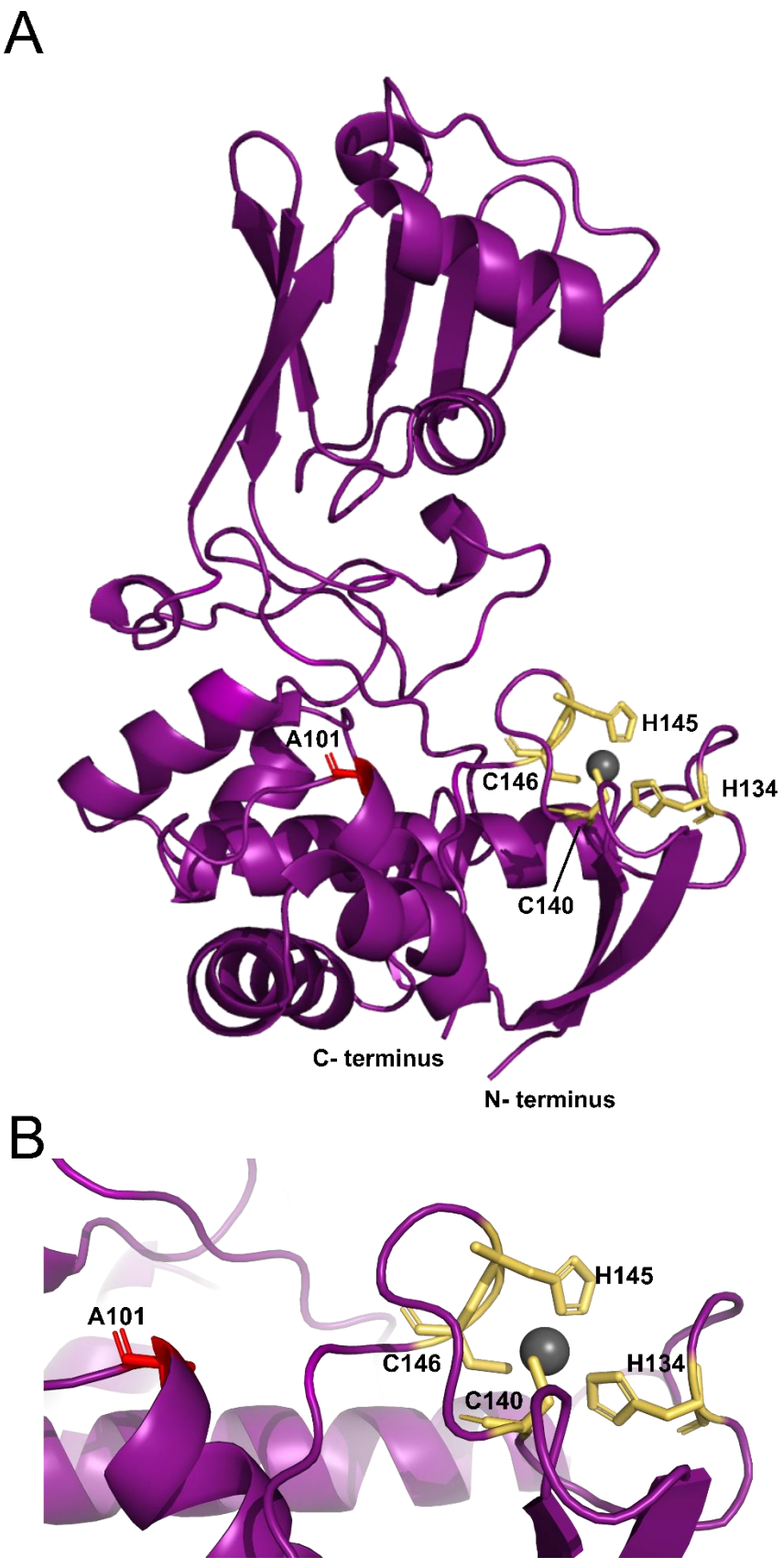


Figure 4-10: Crystal structure of sSpoIID$_{E101A}$

Cartoon representation of the SpoIID$_{E101A}$ structure is show in purple (A). (B) shows the zinc-binding site, and relative position of A101 in more detail, with the residues Zn-binding shown as yellow sticks. The zinc atom is shown as a grey sphere. A101 is shown as red sticks.

## 4.4. Investigation of sSpoIID and sSpoIIP enzymatic activities

### 4.4.1. sSpoIIP produces the substrates for sSpoIID activity

To characterise the activities of sSpoIID and sSpoIIP, the purified proteins, alone and in combination, were incubated with purified *E. coli* peptidoglycan. The products of this reaction were then further digested with the muramidase cellosyl. Cellosyl digests the glycosidic bond between Glc*N*Ac and Mur*N*Ac and therefore allows for soluble muropeptides to be resolved and identified by HPLC-MS.

Digestion of *E. coli* peptidoglycan with cellosyl alone results in 2 major peaks eluting at approximately 20 and 40 minutes. These peaks were identified to contain Glc*N*Ac Mur*N*Ac L-Ala D-Glu *meso*DAP D-Ala muropeptide (muropeptide 4 Figure 4-11 A and B) and a 4-3 crosslinked Glc*N*Ac Mur*N*Ac L-Ala D-Glu *meso*DAP D-Ala dimer, respectively (muropeptide 6 Figure 4-11 A and B).

Digestion with sSpoIID followed by cellosyl results in an almost identical chromatograph, with the only significant peaks containing muropeptides 4 and 6 (Figure 4-11). This demonstrates that sSpoIID has no enzymatic activity on intact *E. coli* peptidoglycan, as has been previously reported for *B. subtilis* SpoIID (Morlot *et al.*, 2010). If sSpoIID were active, one would expect the release of 1,6-anhydro-Mur*N*Ac residues, the product of lytic transglycosylases (Morlot *et al.*, 2010; Vermassen *et al.*, 2019).

Incubation with sSpoIIP alone results in a chromatograph with a different elution profile (Figure 4-11). Peak 3 is the product of amidase activity on peak 6; as sSpoIIP digests both Mur*N*Ac-L-Ala bonds in the dimer, it releases the crosslinked peptides. muropeptide 1 (D-Ala D-Glu *meso*DAP D-Ala) could be produced in two ways; either by amidase activity on peak 4, or by endopeptidase activity on peak 3 (L-Ala D-Glu *meso*DAP D-Ala crosslinked to L-Ala D-Glu *meso*DAP D-Ala), and is the final product of sSpoIIP digestion. These results confirm that sSpoIIP is both an amidase and an endopeptidase and demonstrate that *C. difficile* sSpoIIP functions as *B. subtilis* SpoIIP (Morlot *et al.*, 2010).

When sSpoIID and sSpoIIP are co-incubated with peptidoglycan before further digestion with cellosyl, the chromatograph changes again; muropeptides 1 and 3 are still present, showing that sSpoIIP is active (Figure 4-11). The AUC of peak 2 increases. Peaks 2 and 5 both contain an expected product of transglycosylase activity; a muropeptide terminating in a 1, 6-anhydro-Mur*N*Ac residue. Peak 2 is the further digestion of peak 5, producing the smallest product of sSpoIID activity. This result verifies previous observations from experiments using *B. subtilis* proteins; that SpoIID requires the products of SpoIIP activity, the denuded glycan strands, as a substrate (Figure 4-11C) (Morlot *et al.*, 2010). Dye release assays conducted using *B. subtilis* SpoIID and SpoIIP suggested that SpoIID may have a stimulatory effect on SpoIIP activity (Morlot *et al.*, 2010), however, this was not observed in our semi-quantitative HPLC-MS assays. This may be due to differences in analysis technique rather than a difference in SpoIID/SpoIIP activity between *B. subtilis* and *C. difficile*.

Interestingly, we see 1,6 anhydro-Mur*N*Ac containing muropeptides (peaks 2 and 5 Figure 4-11) in the SpoIIP alone digest. This is surprising considering SpoIIP is not predicted to have any transglycosylase activity (Gutierrez, Smith and Pogliano, 2010; Morlot *et al.*, 2010). Therefore, we hypothesised that these 1,6-anhydro-Mur*N*Ac-containing residues were the

consequence of SpoIIP releasing the peptide stems from the naturally occurring 1,6-anhydro-Mur*N*Ac-containing termini peptidoglycan strands. To test this hypothesis, we digested peptidoglycan both in the presence and absence of cellosyl. As can be seen in Figure 4-12, only when cellosyl is present is peak 5 observed, confirming our hypothesis.

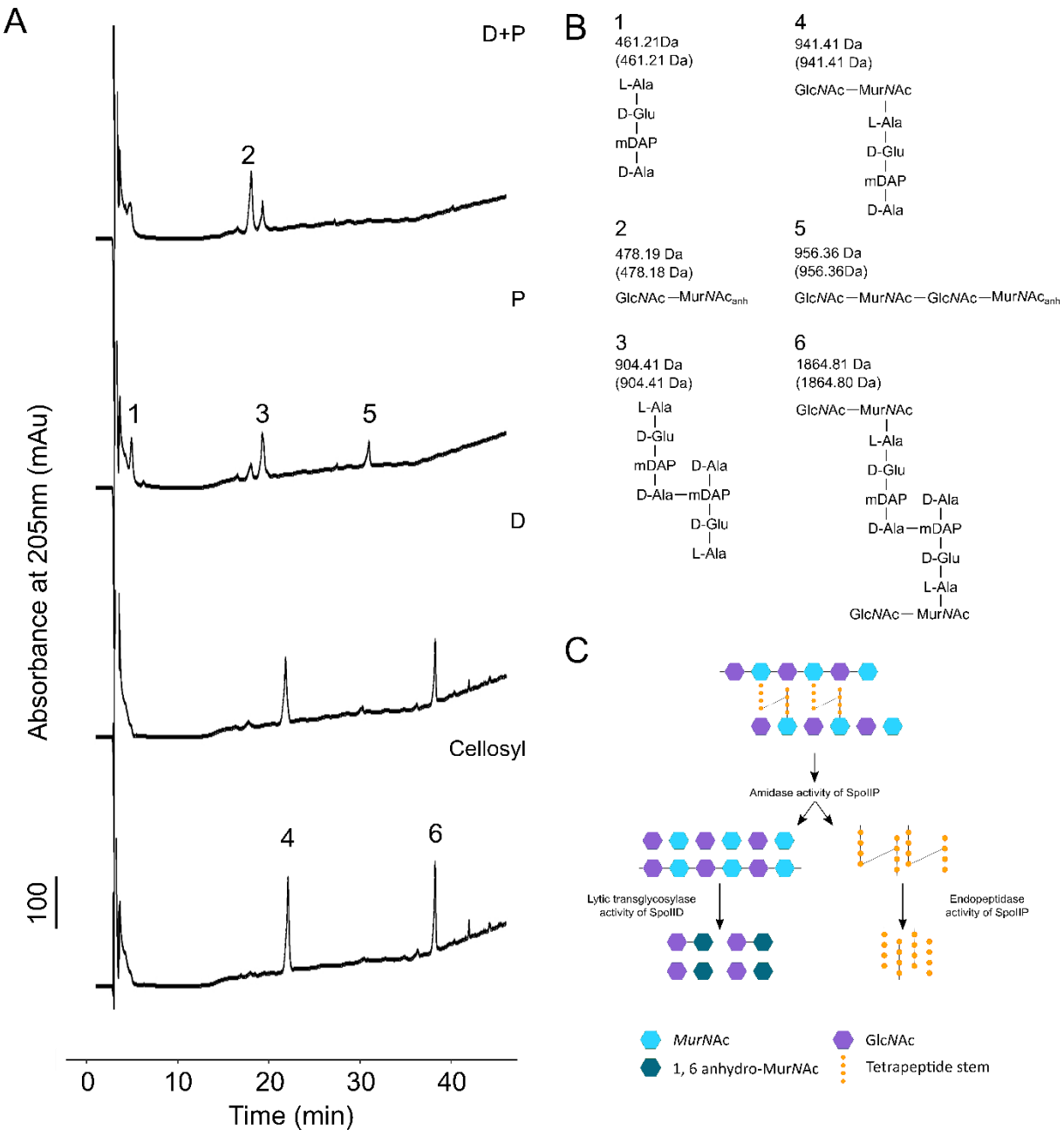


Figure 4-11: sSpoIIP and sSpoIID digest peptidoglycan.

sSpoIIP produces the substrates for sSpoIID activity. (A) cellosyl, sSpoIID (D), sSpoIIP (P) and sSpoIID sSpoIIP (D+P) co-incubations (bottom to top) of *E. coli* peptidoglycan. Numbers above peaks refer to panel (B). Observed mass and structures of muropeptides identified in each peak. Theoretical mass is shown in brackets. All Mur*N*Ac residues are reduced. Panel C is a schematic of SpoIIP and SpoIID activity, SpoIIP creates denuded sugars for SpoIID activity and breaks crosslinks between peptides.

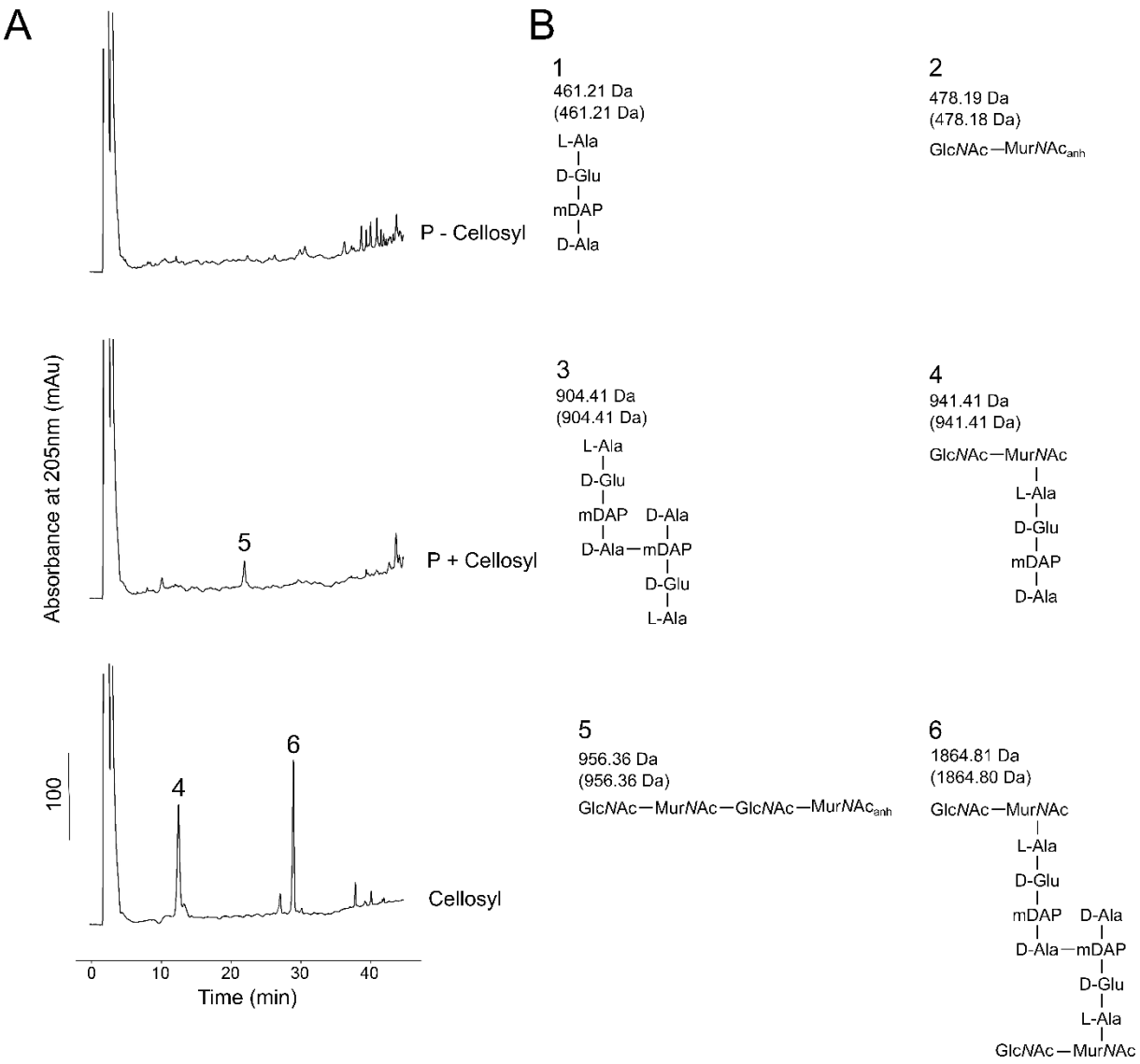


Figure 4-12**:** *E. coli* peptidoglycan digestion by sSpoIIP in the presence and absence of cellosyl.

Due to the unexpected presence of muropeptide 5 when digesting *E. coli* peptidoglycan with sSpoIIP (P + Cellosyl), we digested the same peptidoglycan with sSpoIIP but omitted the cellosyl digestion step (P- Cellosyl). When cellosyl is absent, muropeptide 5 is not observed, suggesting sSpoIIP digests the peptide stems from peptidoglycan with naturally occurring 1,6-anhydro-Mur*N*Ac residues present. Panel A shows the HPLC chromatograph with the numbered peaks corresponding to the muropeptides identified in panel B. Observed and theoretical masses (brackets) are given. All Mur*N*Ac residues are reduced.

### 4.4.2. sSpoIIP H142 and H222 are required for enzymatic activity

H142 and H222 of *C. difficile* sSpoIIP were mutated to arginine, as they are the equivalent of H189 and H278 residues identified as potentially responsible for amidase activity in *B. subtilis* (Chastanet and Losick, 2007).

When peptidoglycan is incubated with either mutant, the chromatographs are equivalent to that seen after digestion with cellosyl alone (Figure 4-13). If H142 and H222 were involved solely in amidase activity, one would expect an increase in muropeptide 4 and a concurrent reduction in muropeptide 6; as the endopeptidase activity of sSpoIIP followed by the muramidase activity of cellosyl would digest peak 6 into peak 4. This is not observed,

suggesting that H142 and H222 are involved either directly or indirectly in endopeptidase activity. Furthermore, these proteins are unable to produce the substrates required for sSpoIID activity as no sSpoIID products are detected when co-incubating peptidoglycan with sSpoIIP$_{H142R}$ or sSpoIIP$_{H222R}$ and sSpoIID.

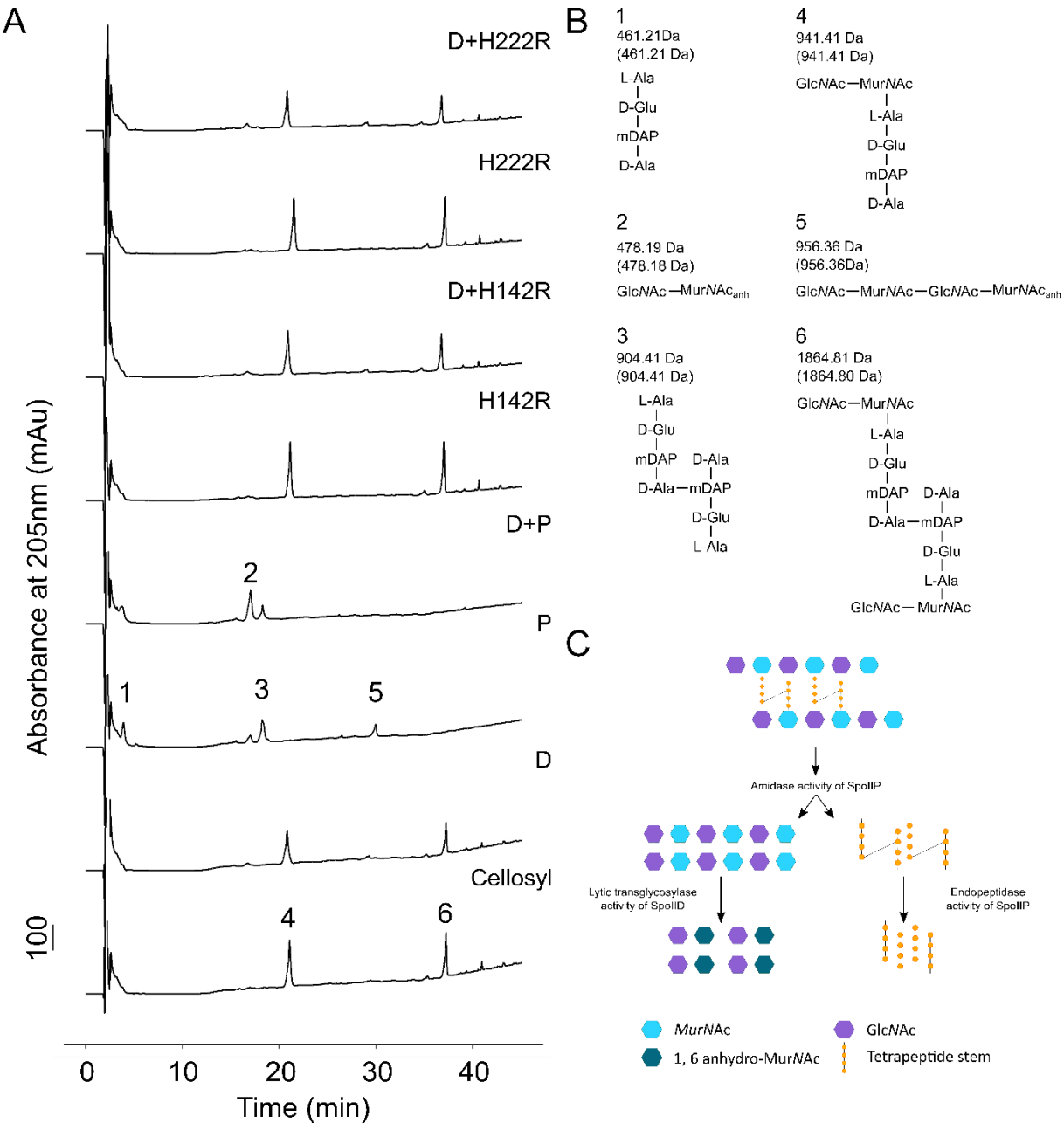


Figure 4-13: sSpoIIP H142 and H222 are required for enzymatic activity.

sSpoIIP H142R and H222R mutants were incubated alone and with sSpoIID. Both mutants appear to be enzymatically inactive, as digests resemble those of sSpoIID alone and no 1,6-anhydro-Mur*N*Ac residues are observed (A and B). Panel C contains a schematic of sSpoIID and sSpoIIP activity. Numbered peaks in panel A refer to the muropeptides in panel B. Observed and theoretical masses are provided (theoretical masses are given in parentheses). All Mur*N*Ac residues are reduced D; sSpoIID, P; sSpoIIP, H142R; sSpoIIP$_{H142R}$, H222R; sSpoIIP$_{H222R}$.

### 4.4.3. Glutamate 101 is responsible for sSpoIID catalysis.

Alignment of *B. subtilis* and *C. difficile* sSpoIID identifies E101 as the putative enzymatic residue in *C. difficile* (Morlot *et al.*, 2010; Nocadello *et al.*, 2016). Co-incubation of $sSpoIID_{E101A}$ and sSpoIIP results in a chromatograph like that seen when only sSpoIIP is used (Figure 4-14 and Figure 4-11 respectively); amidase and endopeptidase products are observed, but muropeptide 2 is not enriched, suggesting no lytic transglycosylase activity. This confirms the results of Nocadello *et al.* (2016) which identified E101 as required for catalysis.

### 4.4.4. Mutation of sSpoIID zinc coordinating residues

Based on the crystal structure determined by Nocadello *et al.*, one would expect that mutating any of the four metal binding residues would abolish zinc binding and enzymatic activity. However, this does not appear to be the case, as seen in our ICP-MS analysis and confirmed by enzymatic activity assays.

Incubation of peptidoglycan with sSpoIIP and either $sSpoIID_{C140A}$ or $sSpoIID_{H145A}$ results in a much-reduced production of muropeptide 2 when compared to co-incubation with wild type sSpoIID (Figure 4-14). This suggests that the enzymatic activity of these zinc-binding mutants is reduced. Interestingly, incubation with $sSpoIID_{C146A}$ and $sSpoIID_{H134A}$ (Figure 4-14 and Figure 4-15 respectively) reduces sSpoIID activity to a lesser extent; 1,6-anhydro-Mur*N*Ac containing muropeptides are identified in both analyses, suggesting that not all zinc binding residues are essential for enzymatic activity. These results differ from previous investigations; mutation of any zinc binding residues to alanine abolished SpoIID activity in dye-release assays (Nocadello *et al.*, 2016). We suggest this is a result of our more sensitive methodology, LC-MS, rather than relying on spectroscopic measurement of a colour change.

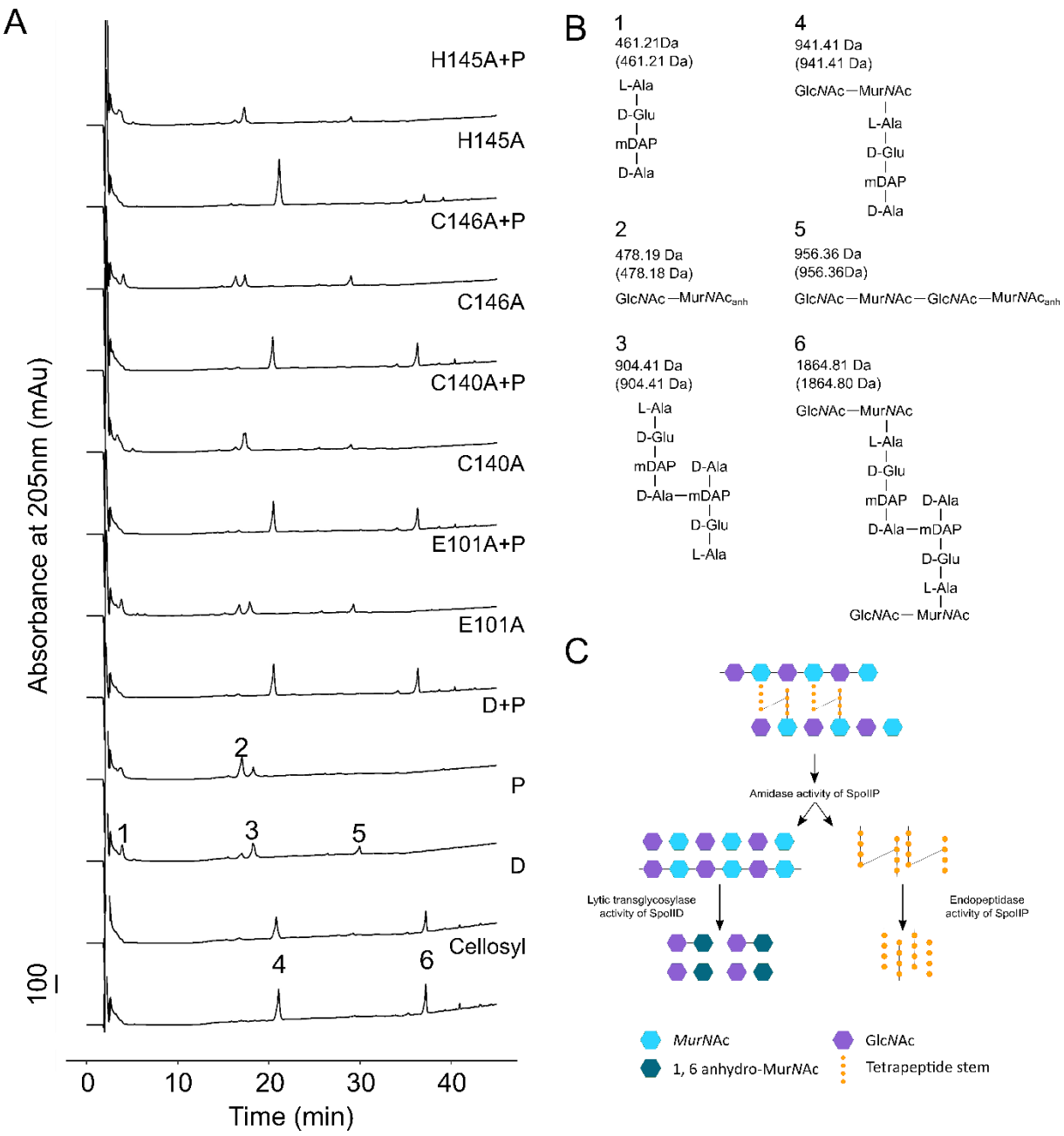


Figure 4-14: Residues required for sSpoIID activity.

Whilst E101 is required for activity, mutation of sSpoIID zinc binding residues produces different effects depending on the residue mutated. (A) sSpoIID$_{E101A}$ is inactive; no additional 1,6-anhydro-Mur*N*Ac containing muropeptides are produced when co-incubated with sSpoIIP alone. Mutation of C140 or H145 reduces sSpoIID activity to a greater extent than mutation of C146 or H134. (B) shows schematics of the muropeptides corresponding to the peaks labelled in (A) with observed masses provided. Theoretical masses are provided in parentheses. All Mur*N*Ac residues are reduced. (C) is a schematic of sSpoIID and sSpoIIP activity on peptidoglycan. D; sSpoIID, P; sSpoIIP, E101A; sSpoIID$_{E101A}$, C140A; sSpoIID$_{C140A}$, C146A; sSpoIID$_{C146A}$ and H145A; sSpoIID$_{H145A}$.

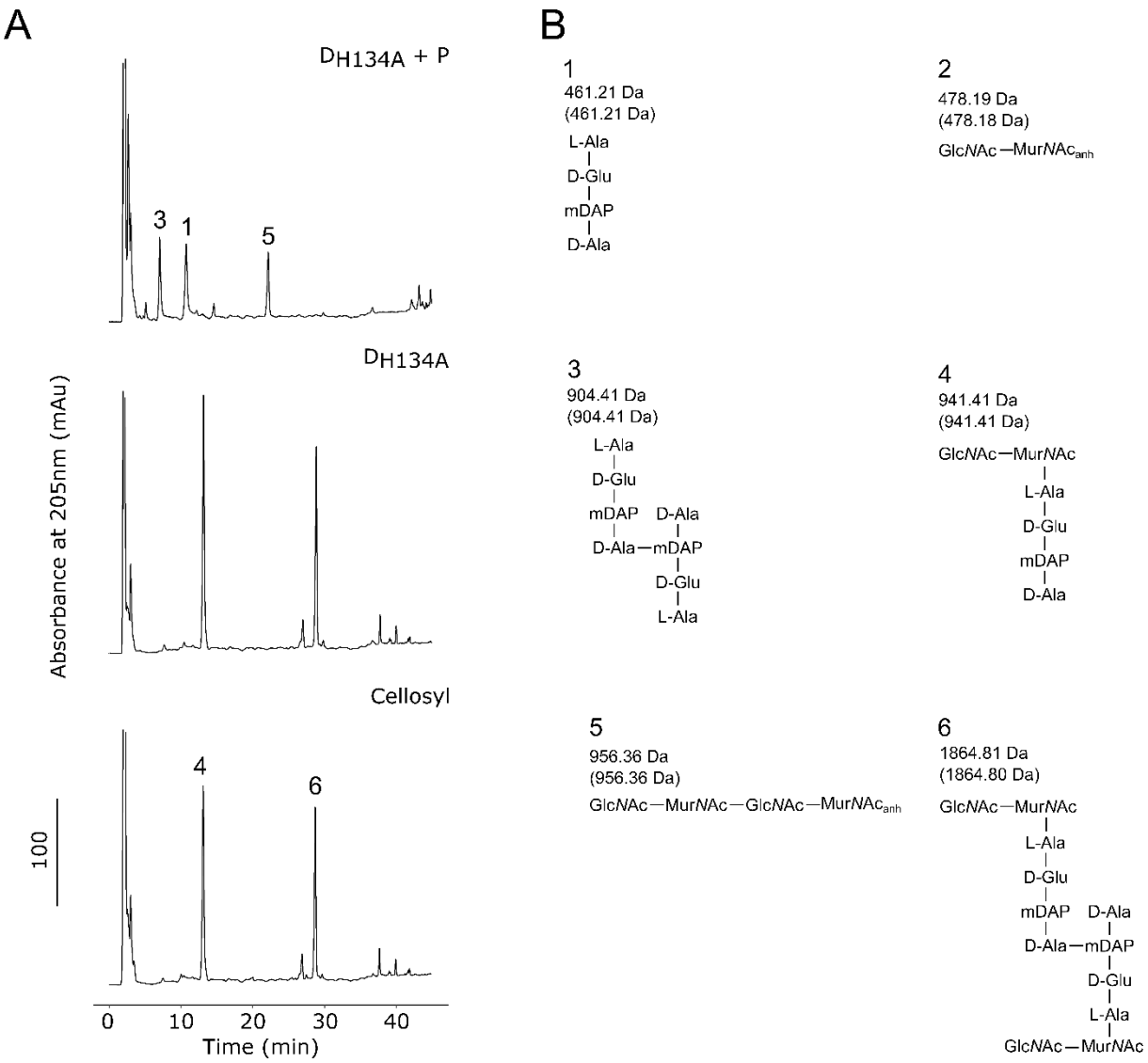


Figure 4-15: Digestion of *E. coli* peptidoglycan by sSpoIID$_{H134A}$.

(A) HPLC trace of *E. coli* Δ*lpp* peptidoglycan digested by cellosyl, sSpoIID$_{H134A}$, and sSpoIID$_{H134A}$ with sSpoIIP. Numbered peaks refer to the structures presented in panel B, where observed (top) and theoretical masses (in parentheses) are provided. All Mur*N*Ac are reduced. D$_{H134A}$; sSpoIID$_{H134A}$, P; sSpoIIP

## 4.5. Acetylation state of peptidoglycan influences sSpoIID and sSpoIIP activity

Both in this work (section 4.4), and in previous literature (Gutierrez, Smith and Pogliano, 2010; Morlot *et al.*, 2010; Nocadello *et al.*, 2016), all investigations on sSpoIID and sSpoIIP activity were performed using *E. coli* peptidoglycan, even though the natural SpoIID and SpoIIP substrate is *C. difficile* peptidoglycan. Although the overall architecture of long glycan strands made of alternating sugar residues crosslinked by short peptide stems is shared, there are some key differences between the Gram-positive *C. difficile* and Gram-negative *E. coli* peptidoglycan. Whilst 4-3 cross links predominate in *E. coli* peptidoglycan (Glauner, Höltje and Schwarz, 1988), 3-3 crosslinks between *meso*DAP residues are the norm in *C. difficile* vegetative cell peptidoglycan (Peltier *et al.*, 2011; Bern, Beniston and Mesnage, 2016). Furthermore, the sugar backbone of the two peptidoglycans are subtly different; where *E. coli* peptidoglycan is formed of alternating Glc*N*Ac and Mur*N*Ac residues, *C. difficile* peptidoglycan is ~90% deacetylated on the Glc*N*Ac residue, resulting in strands of alternating glucosamine (GlcN) and Mur*N*Ac residues (Peltier *et al.*, 2011).

To investigate the impact of these differences, if any, on sSpoIID and sSpoIIP activity, both *E. coli* and *C. difficile* peptidoglycan were chemically acetylated, with the position of any additional acetyl groups assessed by mass spectrometry. As can be seen in Figure 4-16, untreated *E. coli* peptidoglycan is predominantly formed of Glc*N*Ac-Mur*N*Ac-AEmA monomers (peak 1) and 4-3 crosslinked dimers (peak 2). When chemically acetylated, an additional acetyl group is positioned on the *meso*DAP residue of the monomer, forming peak 6 and on one of the two *meso*DAP residues of the dimer, (to give peak 7 (Figure 4-16). This can be seen in the shift in retention time and the addition of ~42Da to the neutral mass of the major peaks in the chromatograph.

In untreated *C. difficile* peptidoglycan, GlcN-Mur*N*Ac-AEmA monomers (Figure 4-17 peak 1) and dimers (Figure 4-17 peak 2) predominate. When this peptidoglycan is chemically acetylated, acetyl groups are added to available GlcN residues to form Glc*N*Ac and to *meso*DAP residues as described for *E. coli* (Figure 4-17 peaks 4 and 5). By comparing the digestion profiles of treated and untreated *E. coli* and *C. difficile* peptidoglycan, the effect of acetyl group and crosslink position on sSpoIID and sSpoIIP enzymatic activities was probed.

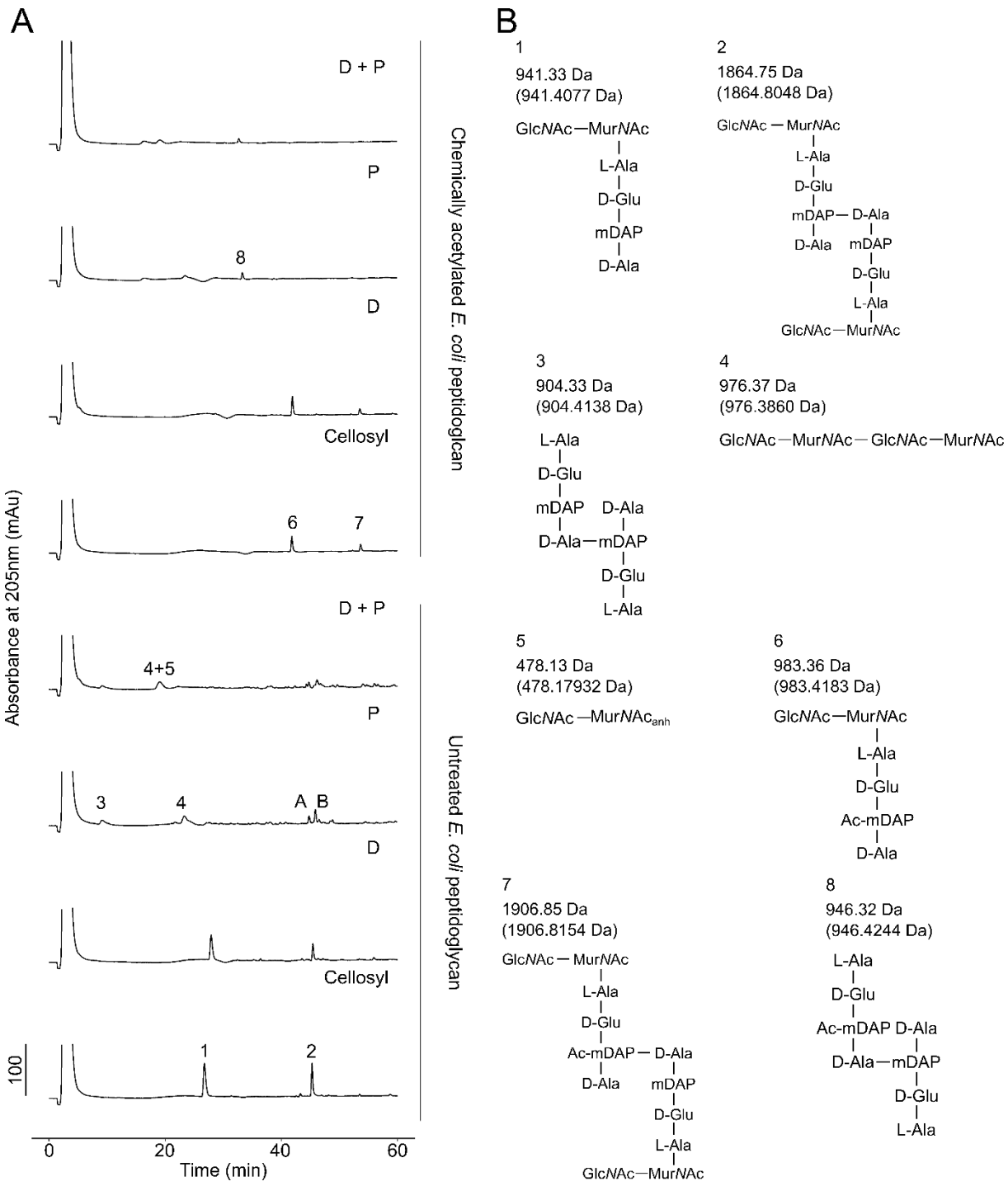


Figure 4-16: Digestion of chemically acetylated peptidoglycan by sSpoIID and sSpoIIP

*E. coli* Δ*lpp* peptidoglycan was digested with cellosyl, sSpoIID (D), SSpoIIP (P) or both sSpoIID and sSpoIIP (D + P) and analysed by HPLC-MS. Chemically acetylated *E. coli* Δ*lpp* peptidoglycan was similarly digested. Numbered peaks contain the corresponding muropeptides schematised in panel B. Peaks A and B could not be identified. Ac-mDAP corresponds to acetylated *meso*DAP. Observed and theoretical masses (in brackets) are provided. All Mur*N*Ac residues are reduced.

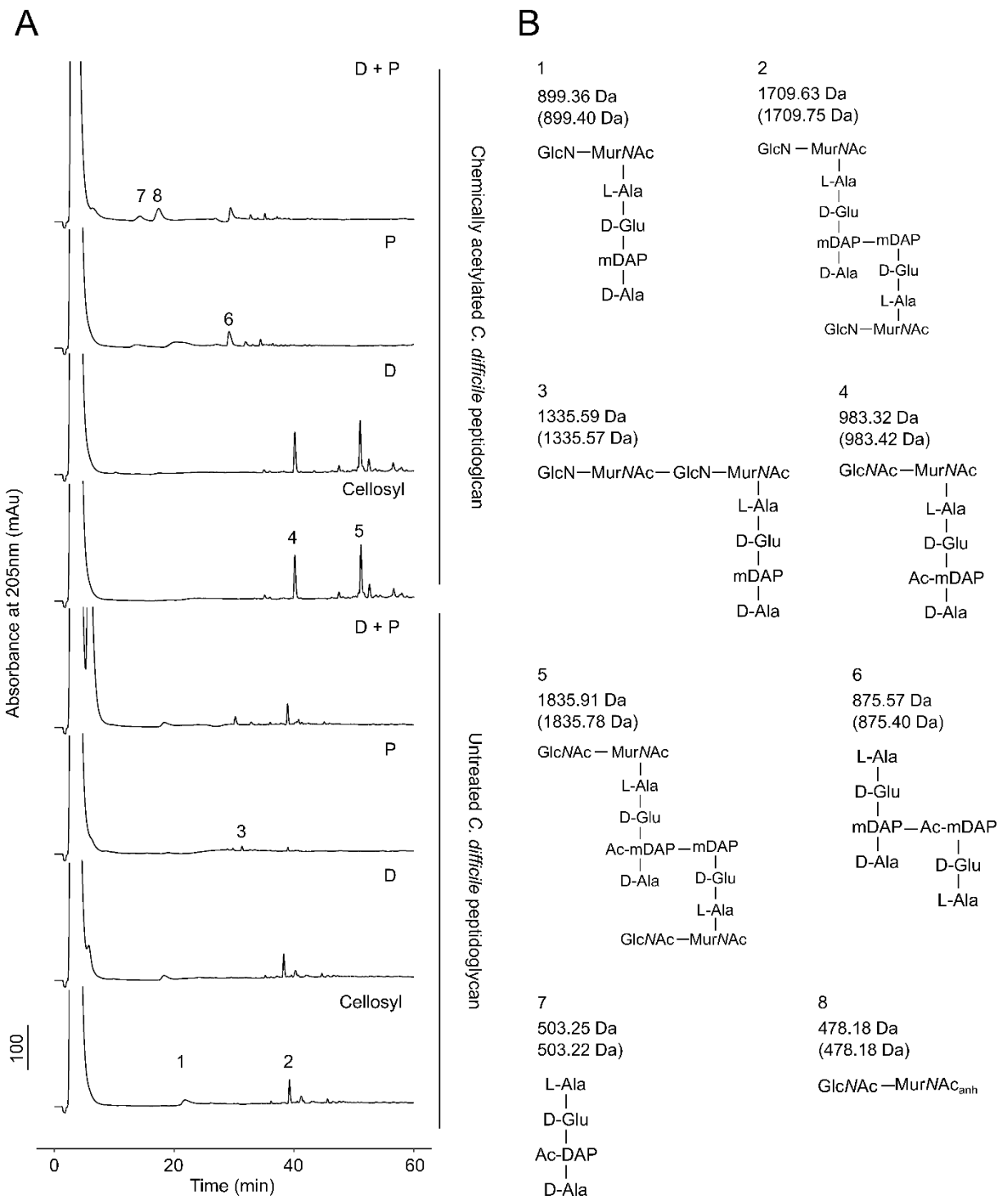


Figure 4-17: Digestion of chemically acetylated *C. difficile* peptidoglycan

Cellosyl, sSpoIID (D), sSpoIIP (P) or sSpoIID and sSpoIIP (D + P) were incubated either with untreated *C. difficile* peptidoglycan (lower half of panel A), or with chemically acetylated *C. difficile* peptidoglycan (upper half panel A) and the resulting muropeptides resolved and identified using LC-MS. Numbers in panel A correspond to the muropeptides in panel B, where observed masses and theoretical masses (in brackets) are provided. Ac-mDAP; acetylated *meso*DAP. All Mur*N*Ac residues are reduced.

Digestion of untreated *E. coli* peptidoglycan with sSpoIID and sSpoIIP produces the peaks already described; sSpoIIP acts as an amidase and endopeptidase to produce long glycan stands and free peptides, sSpoIID then cleaves these long glycans into short disaccharides that terminate in 1,6-anhydro-Mur*N*Ac residues (Figure 4-16). Incubation of sSpoIID and sSpoIIP with chemically acetylated *E. coli* peptidoglycan produces similar muropeptides: free peptide stems where the *meso*DAP has an additional acetyl group (peak 8 Figure 4-16) and the Glc*N*Ac-1,6- anhydro-Mur*N*Ac disaccharide (peak 5, Figure 4-16). Whilst the UV profiles for these peaks are low, inspection of the MS data confirms the presence and identification of these peaks.

Repeating these digests with *C. difficile* peptidoglycan produces different results, shown in Figure 4-17. Incubation of untreated *C. difficile* peptidoglycan with sSpoIIP produces a tetra-saccharide strand with a single peptide stem (muropeptide 3); a muropeptide not identified in the *E. coli* digests. Incubation with both sSpoIID and sSpoIIP produces muropeptides 1 and 2, which are cellosyl digestion products, and muropeptide 3, the product of sSpoIIP digestion, with no evidence of any 1,6-anhydro-Mur*N*Ac containing muropeptides either in the UV chromatographs or in a manual search of the mass spectrum. This suggests sSpoIID is incapable of digesting untreated, i.e. deacetylated *C. difficile* peptidoglycan.

Incubation of chemically acetylated *C. difficile* peptidoglycan with sSpoIIP leads to the production of muropeptide 6: the amidase product of sSpoIIP with an additional acetyl group on one of the *meso*DAP residues, as would be expected based on previous digestions of *E. coli* peptidoglycan (section 4.4). Incubation with both sSpoIID and sSpoIIP leads to the production of muropeptides 6, 7 and 8. Muropeptides 6 and 7 are the acetylated amidase and endopeptidase products of sSpoIIP, with peak 8 being the Glc*N*Ac-1,6-anhydro-Mur*N*Ac product of sSpoIID activity on the sSpoIIP-produced denuded strands.

## 4.6. Discussion

The work presented in this chapter was undertaken with the aim of characterising sSpoIID and sSpoIIP and their activities in *C. difficile*. This work demonstrates that, as has been described for *B. subtilis* (Chastanet and Losick, 2007; Morlot *et al.*, 2010), SpoIID acts as lytic transglycosylase and SpoIIP as a dual activity amidase and endopeptidase (Figure 4-11). Furthermore, zinc binding by sSpoIID is confirmed (Figure 4-7), with the effects of zinc-binding residue mutations investigated (Figure 4-14).

### 4.6.1. sSpoIIP activty

sSpoIIP has amidase and endopeptidase activity that cleaves peptide cross links and peptide stems from Mur*N*Ac residues. The denuded glycan strands are then digested further by sSpoIID, ultimately to produce disaccharides that terminate in a 1,6-anhydro-Mur*N*Ac residue, the product of lytic transglycosylases (Figure 4-11)(Höltje *et al.*, 1975; Herlihey and Clarke, 2016).

### 4.6.2. Zinc binding by sSpoIID

*C. difficile* sSpoIID binds zinc, in contrast to *B. subtilis* SpoIID. Peptidoglycan digestion experiments, CD (Table 4-1) and ICP-MS data(Figure 4-7), combined with microscopic analysis of progression through sporulation (Dembek *et al.*, 2018), have demonstrated that not all zinc binding residues are equally important for *C. difficile* sSpoIID function.

Deletion of *spoIID* in *C. difficile* prevents spore formation, with most cells arrested just after the asymmetric septum begins to curve (Dembek *et al.*, 2018; Ribis, Fimlaid and Shen, 2018). Complementation of Δ*spoIID* with *spoIID*$_{C140A}$ fails to restore spore formation, with no spores detectable in this mutant. ICP-MS analysis shows that zinc binding is completely abolished in sSpoIID$_{C140A}$ and thermal stability CD experiments demonstrate a reduction in the melting temperature of sSpoIID$_{C140A}$. *In vitro* peptidoglycan digests with sSpoIID$_{C140A}$+sSpoIIP show similar levels of 1,6-anhydro-Mur*N*Ac products as seen with sSpoIIP alone (Figure 4-14). Taken together, this may indicate that sSpoIID$_{C140A}$ is less stable and therefore less active, and that this instability and inactivity may be due to the lack of zinc coordination. This would result in reduced *in vivo* activity of SpoIID$_{C140A}$, thus leading to an inability of *spoIID*$_{C140A}$ to restore normal spore formation in a *C. difficile spoIID* deletion mutant.

sSpoIID$_{H145A}$ also showed a reduction in melting temperature, abolishment of zinc binding and a substantial reduction in 1,6-anhydro-Mur*N*Ac production in *in vitro* peptidoglycan digestion experiments. Surprisingly, *spoIID*$_{H145A}$ was able to partially complement a Δ*spoIID* mutant; sporulation was reduced only 5 fold, with cells at all stages of sporulation visible in microscopy images (Dembek *et al.*, 2018). If zinc binding were directly involved in enzymatic activity, one would expect that both SpoIID$_{C140A}$ and SpoIID$_{H145A}$, as both are devoid of zinc, would be completely inactive *in vitro* and *in vivo.* Whilst this appears to be the case *in vitro*, *spoIID*$_{H145A}$ can partially restore sporulation to a Δ*spoIID* mutant, suggesting that there is perhaps some degree of compensation *in vivo* that allows protein stabilisation and some degree of activity. Perhaps other residues can compensate for the loss of H145, by completing the tetrahedral zinc binding motif.

sSpoIID$_{H134A}$, and to a lesser extent sSpoIID$_{C146A}$, maintained some degree of zinc binding and *in vitro* peptidoglycan digestion activity, and were both able to restore spore formation in a Δ*spoIID* mutant to near wild type levels (Dembek *et al.*, 2018). This suggests that not all zinc

binding residues are important for catalysis, and that *in vivo* activity may be greater than *in vitro* activity due to alternative compensatory mechanisms.

Nocadello *et al.* (2016) suggested that zinc binding may have a structural role in SpoIID, increasing rigidity around the substrate binding groove, which in turn may affect enzymatic activity indirectly (Nocadello *et al.*, 2016). Interactions observed between the artificial substrate and SpoIID are mediated by main-chain atoms, and therefore one would not expect mutation to alanine to completely abolish these interactions. The work conducted in this chapter, and in the accompanying paper (Dembek *et al.*, 2018), suggest that zinc does indeed play a structural role, which in turn may affect enzymatic activity.

### 4.6.3. Crystal structure

During the course of this work, the apo structure of *C. difficile* SpoIID was determined by Nocadello *et al.*,in 2016 to 1.95 Å. This structure (PDB ID 5TXU) was used in molecular replacement during model building of $sSpoIID_{E101A}$. As can be seen in Figure 4-18, our model model can be superimposed well, as would be expected, with a RMSD of 0.164 Å. All four proposed zinc binding residues are found in the same region, though their exact positioning is slightly shifted.

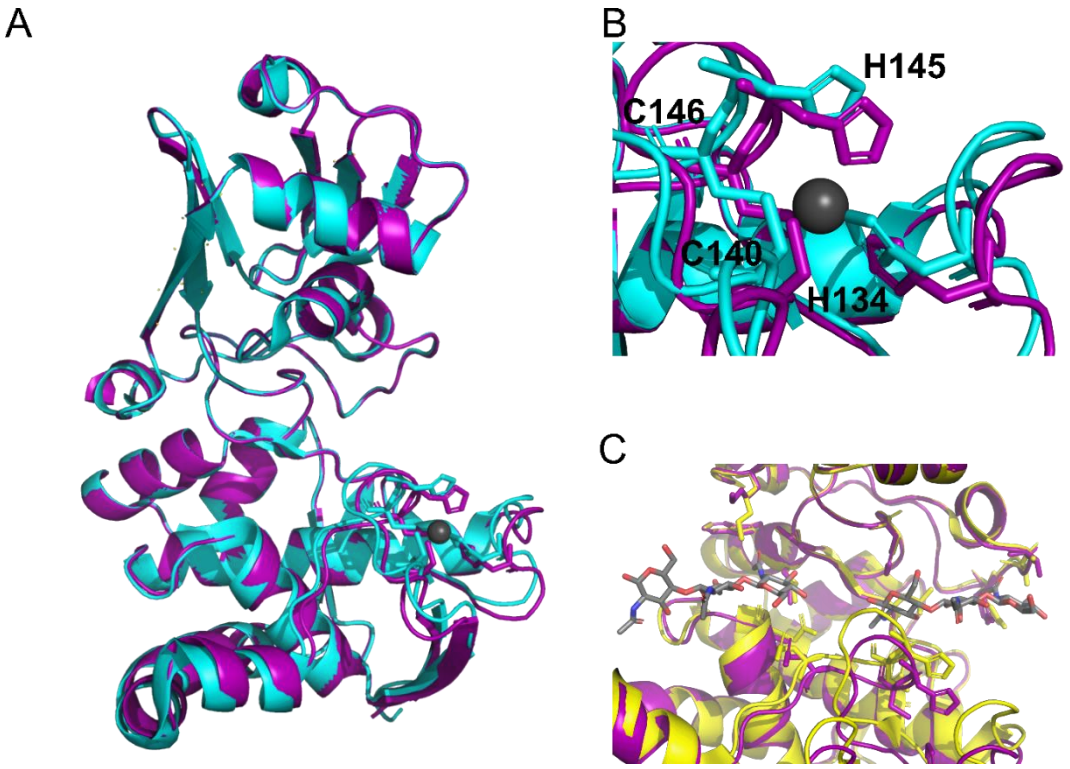


Figure 4-18: Overlay of apo SpoIID and $sSpoIID_{E101A}$.

(A) The apo SpoIID structure (cyan) published by Nocadello et al (2016) (PDB ID: 5TXU) is overlaid onto $sSpoIID_{E101A}$ (purple). (B) Shows the zinc binding site of both molecules in more detail. Cyan residues are those of 5TXU and purple those of $sSpoIID_{E101A}$. (C) $sSpoIID_{E101A}$ (purple) overlaid onto 5I1T (yellow), the $NAG_3$ bound substrate (gray sticks) of SpoIID as determined by Nocadello *et al* (2016), with the proposed residues of significance shown as yellow sticks.

As can be seen in Figure 4-18 H145 is shifted out of position in our model, though this may just be due to a lack of substrate moving the loop structure. The orientation of all other

residues proposed to interact with the artificial ligand $NAG_3$ is conserved in our $sSpoIID_{E101A}$ model.

One notable feature of our model is a large unmodelled electron density within the proposed substrate binding groove (Figure 4-19) As the catalytic glutamate is mutated to alanine, any potential substrate bound to the protein would not be cleavable, so it is possible that this density correspond to a natural substrate of SpoIID. However, this density could not be unequivocally interpreted as a glycan strand as may be expected. Thin layer chromatography of the purified protein did not indicate the presence of any Mur*N*Ac residues, suggesting that the bound molecule might not be a natural peptidoglycan. It should be noted that the crystallisation process can enrich for a certain species in solution so it is possible that ligand-bound sSpoIID dominate the crystal, but the sugars would be in a concentration lower than the limit of detection of TLC. Future experiments will aim to characterise this density; the same preparation of protein will be crystallised and the crystals themselves used in both TLC, which will be used to detect a range of sugars not just Mur*N*Ac, combined with MS analysis, in the hope of identifying the ligand.

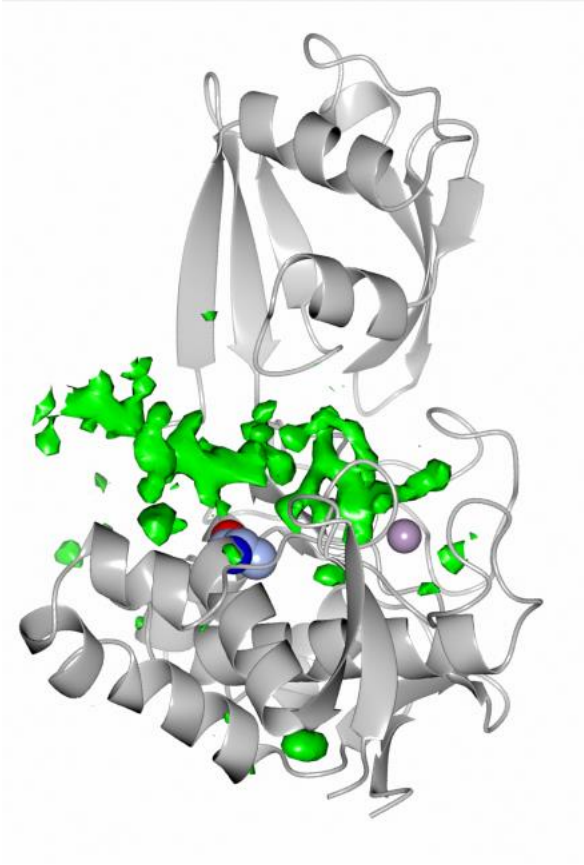

Figure 4-19: sSpoIID with unidentified density

The model of sSpoIID is shown as a grey cartoon, with the unidentified density shown in green. A101 is shown as spheres coloured by atom type: carbon, grey; oxygen, red; nitrogen, blue, and the zinc as a violet sphere. The density lays across the “hand” and “arm” domains and may represent a substrate or intermediate for/of sSpoIID activity. Figure made by Dr Paula Salgado.

Although crystallisation trials with commercially available screens were set up for all sSpoIID mutants, only $sSpoIID_{E101A}$ produced crystals. Thermostability experiments conducted on all

constructs revealed a lower melting temperature of all zinc-binding mutants in comparison to sSpoIID and sSpoIID$_{E101A}$ (Table 4-1) which suggests a reduction in protein stability, that could prevent crystal formation in the conditions tested to date.

In order to try to obtain protein crystals for the remaining mutants, the following parameters could be optimised: a) protein concentration b) other commercially available crystal screens could be tested or c) the condition that produced sSpoIID$_{E101A}$ crystals could be further optimised by screening pH and concentration of Tacsimate™. Microseeding methods using sSpoIID or the E101A crystals could also allow the formation of crystals of the other variants.

### 4.6.4. sSpoIID does not stimulate sSpoIIP

LytC is a major *B. subtilis* autolysin, the activity of which is enhanced by the enzymatically inactive LytB (Chastanet and Losick, 2007). It has been proposed that SpoIID may enhance the activity of SpoIIP in a manner similar to that seen on LytC by LytB. This enhancement activity has been demonstrated through the use of dye release assays and *B. subtilis* SpoIID catalytically inactive mutants (Morlot *et al.*, 2010). However, no enhancement of activity was observed in the semi-quantitative peptidoglycan digests presented in this work; co-incubation of sSpoIIP and sSpoIID$_{E101A}$ did not lead to a significant increase in the proportion of spoIIP products (namely peaks 1 and 3 Figure 4-11) (Dembek *et al.*, 2018). We hypothesise that this difference does not reflect a fundamental difference in the activities of sSpoIID between the two species, rather that there is a difference in the accuracy and sensitivity of the two methods. Dye release assays rely on the release of a dye that is bound to the sugar backbone of peptidoglycan (Zhou, Chen and Recsei, 1988). Therefore, one would not necessarily expect an increase in dye release when acted upon by an amidase or endopeptidase, as dye would presumably remain bound to the unaffected sugar backbone. In contrast, peptidoglycan digestion experiments separate products by HPLC, the eluate of which is then assessed by mass spectrometry allowing precise identification of products. However, whilst more precise, these HPLC-MS assays are only semi-quantitative in comparison to the fully quantitative dye release assays. Ideally, future work would involve the development of a quantitative HPLC-MS method, combining both the quantitative nature of the dye release assay with the precision of HPLC-MS. One possibility would be the injection of a known “standard” peptide that could be used to normalise all data within a digest, thus, allowing some degree of quantitative analysis.

### 4.6.5. Acetylation state of peptidoglycan affects sSpoIID and sSpoIIP activity

As demonstrated in section 4.4, sSpoIID and sSpoIIP appear to digest acetylated peptidoglycan, either that of *E. coli* (Figure 4-16) or chemically acetylated *C. difficile* peptidoglycan (Figure 4-17), to a greater extent than they can digest deacetylated peptidoglycan. Indeed, sSpoIID appears incapable of digesting native *C. difficile* peptidoglycan. This suggests that sSpoIID and sSpoIIP have a preference for acetylated peptidoglycan and that the type of cross link (i.e. 4-3 as seen in *E. coli* or 3-3 as in *C. difficile*) has little, if any, impact.

This may hint to a possible method of distinguishing the newly synthesised, presumably Glc*N*Ac-containing, septal peptidoglycan and the deacetylated peptidoglycan found in the mother cell wall, which in turn may advance our present model of peptidoglycan

remodelling during engulfment (Ojkic *et al.*, 2016). In the current model (Figure 1-9), newly synthesised peptidoglycan is inserted at the septum and the peptide bonds between new and old peptidoglycan digested (presumably by sSpoIID and sSpoIIP). This creates space for the advancing mother cell membrane to expand into.

Ojkic *et al* suggest there is some method of distinguishing the peptide bonds linking the old and new peptidoglycan from the peptide bonds keeping the layers of mother cell peptidoglycan together. Our data suggests that the acetylation state of the peptidoglycan itself may influence this specific peptide bond digestion. sSpoIID and sSpoIIP digest mature GlcN-containing peptidoglycan to a lesser extent than they digest Glc*N*Ac-containing peptidoglycan, which would be found in the newly synthesised peptidoglycan. This suggests sSpoIID and sSpoIIP would be inherently less able to digest the mother cell wall peptidoglycan than the Glc*N*Ac-containing septal peptidoglycan, which supports the hypothesis of Ojkic *et al* that newly synthesised peptidoglycan is preferentially digested by the DMP complex.

Whilst the role of sSpoIIP as an amidase and endopeptidase is easily accounted for in this model, it is harder to account for the lytic transglycosylase activity of sSpoIID. If glycan strands are arranged perpendicular to the long axis of the cell as suggested (Ojkic *et al.*, 2016), peptide bond removal as in Figure 1-9 would be required to free up space for the advancement of the mother cell membrane. However, why digestion of denuded strands would be required is less obvious and should be investigated.

## 4.7. Conclusions and future work

Work presented in this chapter confirms that sSpoIID and sSpoIIP share similar enzymatic activities with their *B. subtilis* counterparts; sSpoIIP is an amidase and endopeptidase that cleaves peptides stems from peptidoglycan to leave denuded glycan strands, the substrate of sSpoIID. sSpoIID then cleaves long glycan strands into GlcN(Ac)-1,6 -anhydro-Mur*N*Ac disaccharides.

The enzymatic activities of sSpoIIP and sSpoIID were probed further; digestion of acetylated peptidoglycan appears to be more complete than deacetylated peptidoglycan, with the suggestion that sSpoIID cannot digest deacetylated mature *C. difficile* peptidoglycan. This result may relate to the interactions proposed between SpoIID and the acetyl groups of $NAG_3$ by Nocadello *et al*. Presumably, the newly synthesised septal peptidoglycan is acetylated, therefore preferentially digested by SpoIIP and SpoIID, resulting in specific digestion of only new peptidoglycan, and not the deacetylated mother cell wall peptidoglycan, thus preventing premature cell lysis during engulfment.

Future work in this area should include co-crystallisation of SpoIID and $SpoIID_{E101A}$ with other peptidoglycan mimics, such as a GlcN trimer and with purified acetylated and deacetylated peptidoglycan. Furthermore, some degree of quantitative kinetic analysis of SpoIID/SpoIIP digestion of both acetylated and deacetylated peptidoglycan should be performed, to determine if the differences observed in section 4.5 are relevant in terms of the engulfment time span. A combination of these kinetic analyses and crystallisation trials may result in the ability to propose an informed enzymatic mechanism, with regards to substrate specificity of SpoIID, which may elucidate any preferential digestion of acetylated peptidoglycan.

Furthermore, co-crystallisation could be used to identify potential inhibitors of sSpoIID, which in turn may have practical implications in terms of preventing *C. difficile* sporulation. For the above reasons, crystallisation of sSpoIIP should also be pursued, with alternative strategies explored.

Future work should focus on further characterising the changes to peptidoglycan during engulfment: how are the activities of sSpoIID and sSpoIIP regulated? Is the "preference" of sSpoIID and sSpoIIP for acetylated peptidoglycan biologically relevant? Answering these, and similar, questions will advance our knowledge of *C. difficile* spore formation, and consequently potentially provide methods to interrupt this process, which in turn would prevent *C. difficile* persisting in the environment or the gut.

# Chapter 5. Identification of putative *C. difficile* polysaccharide deacetylases

## 5.1. Introduction

Whilst the earliest steps in *C. difficile* peptidoglycan synthesis have not been experimentally explored, it is expected that *C. difficile* peptidoglycan is synthesised containing Glc*N*Ac, and that the acetyl group is removed during peptidoglycan maturation to produce the GlcN-containing peptidoglycan found in the mother cell wall (Peltier *et al.*, 2011; Bern, Beniston and Mesnage, 2016), or in the spore cortex (Coullon *et al.*, 2018).

*C. difficile* has orthologues (inferred from sequence identity) of all six Mur enzymes which are involved in the initial cytoplasmic steps of peptidoglycan synthesis (Heijenoort, 2001). MurA is responsible for the first committed step of peptidoglycan synthesis, and whilst the homologue in *C. difficile* has never been experimentally characterised, it seems logical to assume it would also catalyse the transfer of phosphoenolpyruvate onto UDP-*N*-Acetylglucosamine, as is seen in *B. subtilis* (Kock, Gerth and Hecker, 2004). If this is the case, then *C. difficile* peptidoglycan would be produced containing Glc*N*Ac. Detailed MS analysis has demonstrated that the vegetative cell (section 3.1.1) (Peltier *et al.*, 2011; Coullon *et al.*, 2018) and spore (Coullon *et al.*, 2018) peptidoglycan is predominantly deacetylated to glucosamine (GlcN). Therefore, it is assumed that deacetylation of Glc*N*Ac occurs during a peptidoglycan maturation phase.

Based on this assumption, the potential preferential digestion of acetylated peptidoglycan by SpoIID and SpoIIP (section 4.5), the detection of acetylated muropeptides in peptidoglycan from a *sigE* mutant strain (section 3.2.4.1), and the hypothesis of Ojkic *et al* (2016) (that newly inserted peptidoglycan is speficially digested by the DMP machinery (Figure 1-9)), a new hypothesis was produced. I hypothesised that the new “immature” peptidoglycan is acetylated and contains Glc*N*Ac, whereas the “mature” mother cell wall peptidoglycan is deacetylated to contain GlcN, and that this difference in acetylation state contributes to the targeted degradation of only the newly synthesised peptidoglycan during engulfment. Therefore, we sought to investigate potential polysaccharide deacetylases in *C. difficile* that may significant during engulfment.

## *5.2.* Identification of putative polysaccharide deacetylases in *C. difficile*

HMMER is an online tool that searches against sequence databases and HMM libraries for proteins similar to the input query protein (Potter *et al.*, 2018). We limited our search to Clostridia (taxID: 186801) and used the *S. pneumoniae* PgdA sequence as a query. This identified 10 potential *C. difficile* 630 proteins as potential polysaccharide deacetylases identified by our bioinformatics search, 5 were selected for cloning in such a manner that soluble protein would be produced by recombinant expression. As described in section 2.6, predicted transmembrane regions were not included in the amplification of target genes from gDNA. As summarised in Table 5-1, only 2 constructs reliably produced soluble protein: CD630_1319 and CD630_1522, which were further characterised.

Table 5-1: Summary of identified potential polysaccharide deacetylases in *C. difficile*

| Putative deacetylase gene | Notes | Cloning | Expression test |
|---|---|---|---|
| *CD*630_*32480* | *sigE* dependent (Saujet *et al.*, 2013) | - | - |
| *CD*630_*32570* | *sigE* dependent (Fimlaid *et al.*, 2013; Saujet *et al.*, 2013; Dembek *et al.*, 2015) | Strains PS384 & PS387 | Potentially produced in inclusion bodies |
| *CD*630_*25980* | - | Unsuccessful | - |
| *CD*360_*27190* | Minor Mur*N*Ac deacetylase involved in muramic-δ-lactam synthesis, renamed *pdaA2* (Coullon *et al.*, 2018) | - | - |
| *CD*630_*27240* | - | Strains PS386 & PS390 | Not reliable |
| *CD*630_*13190* | *sigE* dependent (Fimlaid *et al.*, 2013; Saujet *et al.*, 2013; Dembek *et al.*, 2015).<br>Gene product observed in spores (Lawley *et al.*, 2009)<br>Transcriptome and proteome unaffected in 6*30 Δerm Δspo0A vs 630* Δerm (Pettit *et al.*, 2014)<br>Possible link to CD630_15220* | Strains PS385 & PS389 | Yes |
| *CD*630_*14300* | Major Mur*N*Ac deacetylase involved in muramic-δ-lactam synthesis, renamed *pdaA1* (Coullon *et al.*, 2018). | - | - |
| *CD*630_*14440* | - | Strain PS459 | Not tested |
| *CD*630_*15220* | Possible link to CD630_13190*<br>Upregulated in *630* Δ*erm* Δ*spo0A* vegetative cell transcriptome and *630Δerm* proteome (Pettit *et al.*, 2014) | Strains PS383 & PS388 | Yes |
| *CD*630_*15560* | Possible link to *rsiV**.<br>In same operon as *csfV* which is *sigV* regulated and upregulated in response to lysozyme stress, equivalent of *pdaV* and Glc*N*Ac deacetylases (Ho *et al.*, 2014) | - | - |

Potential *C. difficile* peptidoglycan deacetylases were identified based on homology to known peptidoglycan deacetylases. Potential interactions marked * identified using the STRING server (Szklarczyk *et al.*, 2019). – indicates nothing to report.

One Glc*N*Ac deacetylase, PdaV (identified as CD630_15560 in Table 5-1), has been previously identified in *C. difficile*, and was demonstrated to be upregulated in response to lysozyme treatment in a $\sigma^V$ dependent manner (Ho *et al.*, 2014). However, there was no investigation

into when PdaV was natively active, *i.e.* was PdaV responsible for the baseline deacetylation of *C. difficile* peptidoglycan, or only enhanced deacetylation seen in response to lysozyme stress?

In many other bacteria, peptidoglycan deacetylation is associated with increased resistance to lysozyme (Psylinakis *et al.*, 2005; Ho *et al.*, 2014), reduced infectivity (Blair *et al.*, 2005) and is implicated in evasion of the host NOD1 and NOD2 immune receptors (Girardin *et al.*, 2003; Boneca *et al.*, 2007). It is noteworthy that knock out of the *Streptococcus mutans* PdgA did not affect lysozyme sensitivity (Deng *et al.*, 2009).

Two of the identified *C. difficile* putative deacetylases, CD630_1430 and CD630_2719, are confirmed Mur*N*Ac deacetylases and were renamed PdaA1 and PdaA2, respectively (Coullon *et al.*, 2018), and therefore were not investigated further in this work.

Deacetylases that are under the control of sporulation related sigma factors (CD630_1319 and CD630_3257) (Fimlaid *et al.*, 2013; Saujet *et al.*, 2013; Dembek *et al.*, 2015), were cloned into pET-M11 vectors, however soluble protein was only produced with the CD630_1319 construct.

CD630_1522 and CD630_2724 are not known to be under the control of a sporulation-associated sigma factor. It is possible that these deacetylases are active during vegetative growth, but inhibited during engulfment, in order to ensure only newly synthesised peptidoglycan is degraded at the leading edge. Although both were cloned for recombinant expression, only the CD630_1552 construct reliably produced soluble protein for characterisation.

## 5.3. Purification of putative polysaccharide deacetylases

### 5.3.1. CD_1319

For expression of CD630_1319, Rosetta cells transformed with pET M11 CD630_1319$_{36-247}$ were grown, protein was expressed in AIM TB, and the protein purified as described in section 2.6.2. Filtered cell lysate was initially purified by IMAC and fractions containing protein pooled (Figure 5-1A). At this point the preparation was split, with half of the sample immediately purified further by SEC to produce purified CD630_1319 containing an intact 6xHis-tag, herein denoted His1319 (Figure 5-1B) and the other half incubated with TEV overnight prior to SEC to remove the 6xHis-tag, producing untagged protein (s1319 Figure 5-1C). SDS-PAGE analysis (Figure 5-1B and C) and western blot analysis (Figure 5-2) shows the proteins produced were sufficiently pure for subsequent investigations.

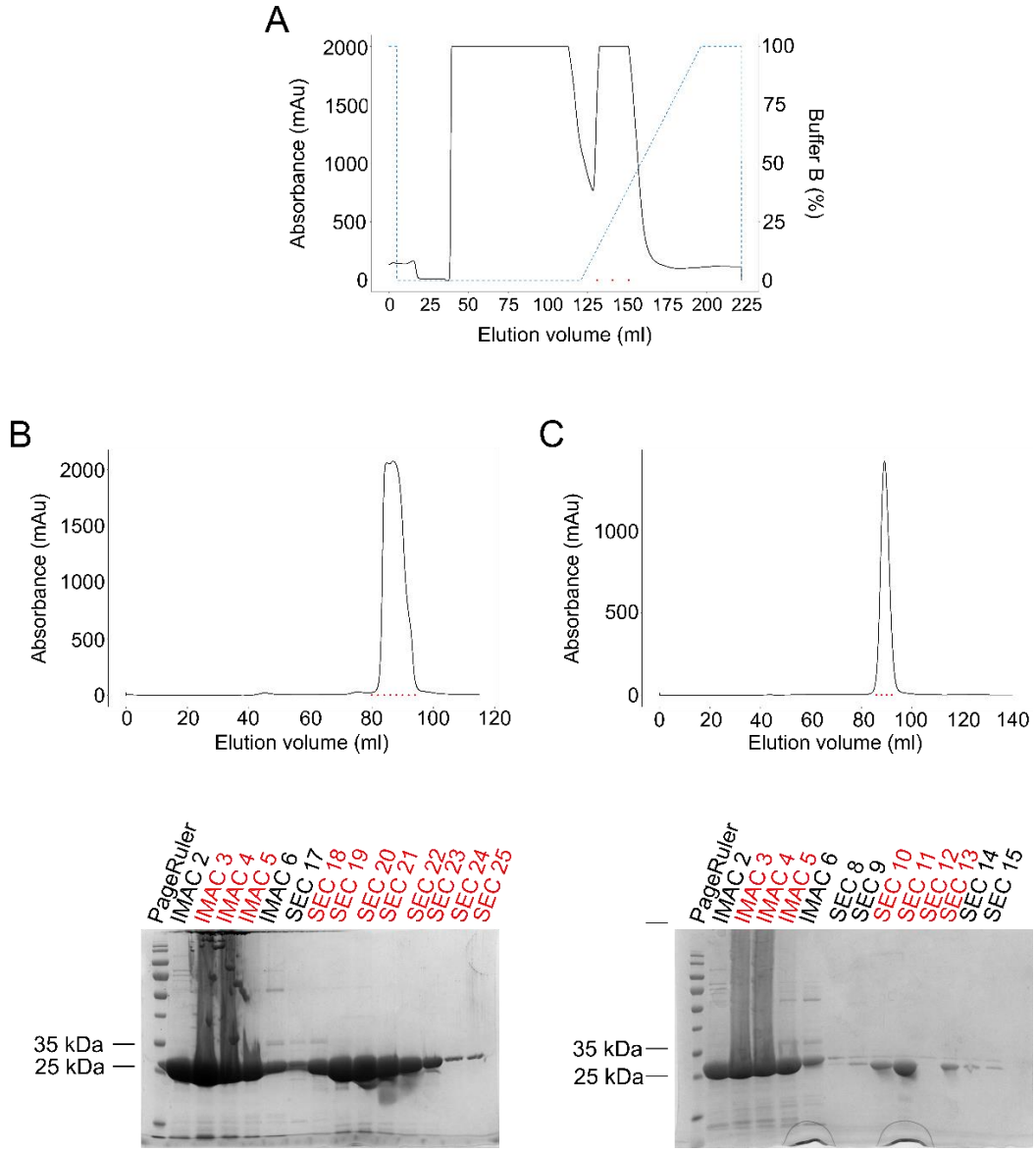


Figure 5-1: Purification of CD_1319.

(A) IMAC purification of CD_1319, the marked fractions (red) eluted in the presence of buffer B. Buffer B is 50 mM Tris base pH 8.0, 300 mM NaCl, 250 mM imidazole. (blue dashed line) were pooled. Half of the preparation was immediately purified by SEC (B). The other half was incubated with TEV to remove the 6xHis-tag prior to further SEC (C). Fractions marked in red in (B & C) were pooled before the next step of purifcation.

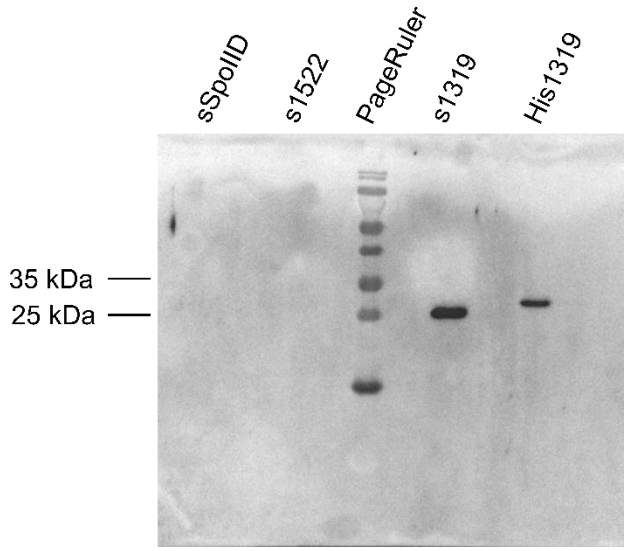


Figure 5-2: CD630_1319 western blot

100ng of each protein was resolved on a 15% SDS-PAGE gel, before semi-dry transfer onto a nitrocellulose membrane. The membrane was blocked before treatment with the primary antibody, which was raised against a CD630_1319-specific peptide, washed, then incubated with the HRP-conjugated secondary antibody. sSpoIID and s1522 were included as negative controls. Both s1319 and His1319 were detected by the antibody.

### 5.3.2. CD_1522

CD630_1522 was purified as described for CD630_1319, with two samples obtained; the 6xHis tagged His1522 and the tag-free s1522. Chromatographs, SDS-PAGE (Figure 5-3B and C) and western blots (Figure 5-4) all suggest the proteins produced were of sufficient quality for further characterisation.

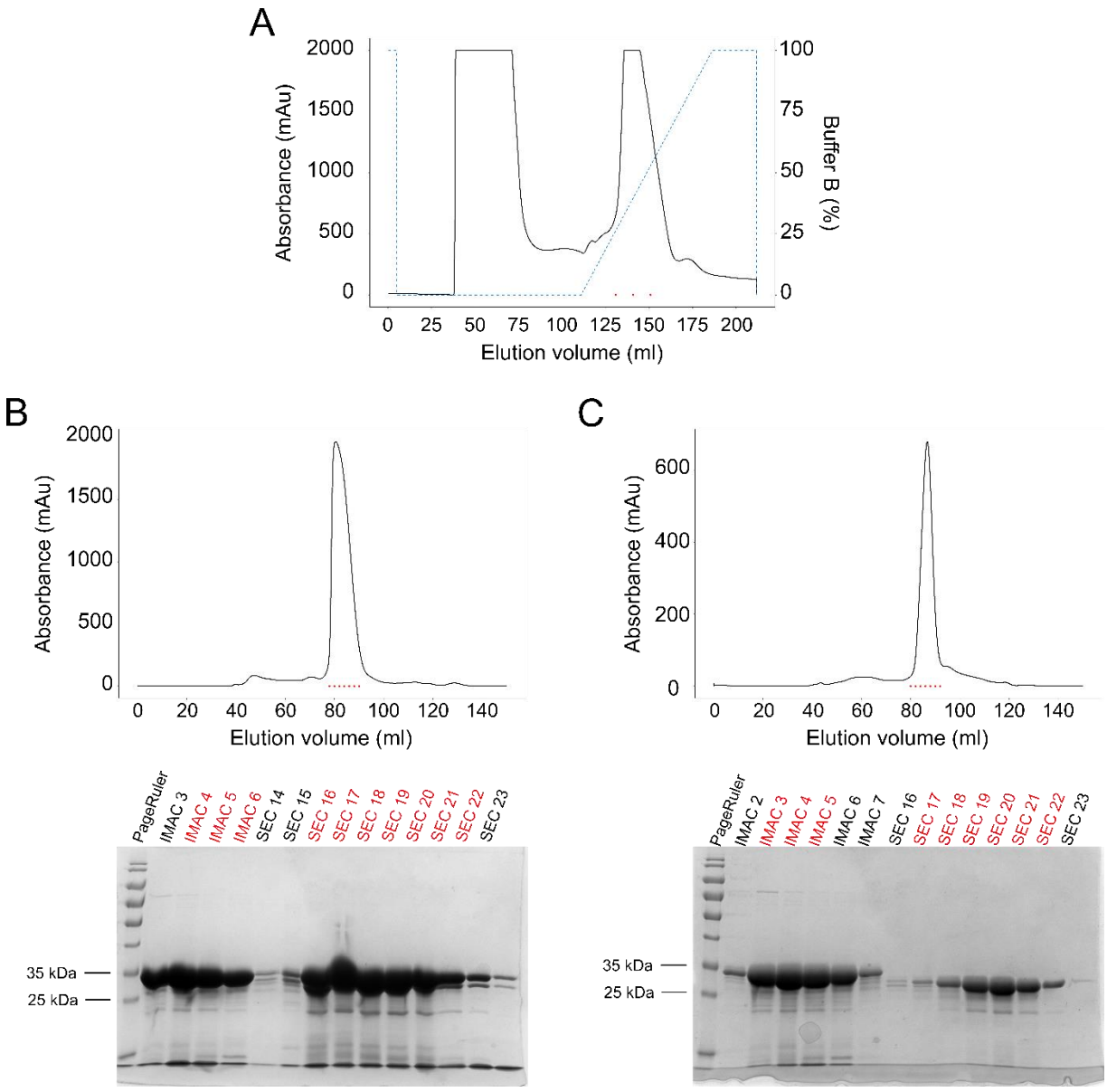


Figure 5-3: Purification of CD630_1522.

Panel A shows the IMAC purification of CD630_1522, the marked fractions (red dots) that eluted in the presence of buffer B (blue dashed line) were pooled. Buffer B is 50 mM Tris base pH 8.0, 300 mM NaCl, 250 mM imidazole. Half of the protein preparation was immediately further purified by SEC (panel B), whilst the other half was incubated overnight with TEV to remove the 6xHis-tag prior to further SEC (panel C). In panels B and C, marked fractions (red) were assessed by SDS-PAGE (lower panel) before pooling and concentration of the fractions marked red in the SDS-PAGE gel.

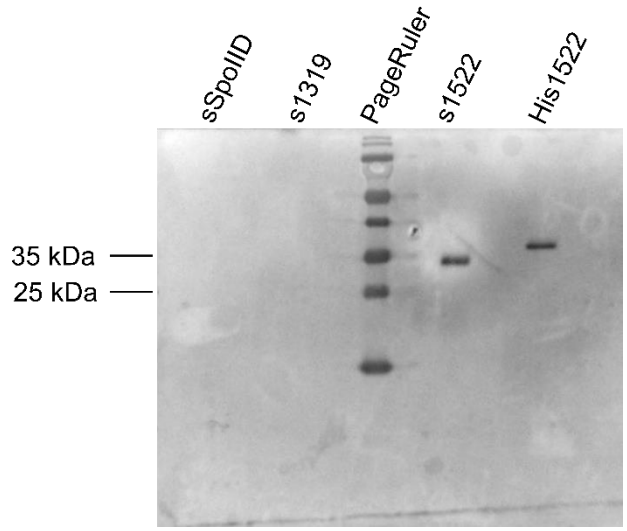


Figure 5-4: CD630_1522 western blot

100ng of each protein was resolved on a 15% SDS-PAGE gel before semi-dry transfer onto a nitrocellulose membrane. The membrane was blocked, probed with primary antibody, which was raised in rabbits inoculated with a CD630_1522-specific peptide, and secondary antibody as described in section 2.4.3.2. Only s1522 and His1522 were detected. sSpoIID and s1319 were included as negative controls.

## 5.4. Protein characterisation

### 5.4.1. Determination of accurate mass

Quadrupole time of flight (qTOF) MS was undertaken to determine the accurate masses (±1Da) of the purified proteins, both with and without 6xHis-tags (Table 5-2). Calculated masses were obtained using the ProtParam tool (web.expasy.org/protparam/).

Table 5-2: Calculated and observed masses of purified putative deacetylases

| Protein | Calculated mass (Da) | Observed mass (Da) |
|---|---|---|
| His1319 | 27521.32 | 27520.73 |
| S1319 | 24523.01 | 24522.36 |
| His1522 | 33974.48 | 33973.85 |
| S1522 | 30976.18 | 30975.97 |

Calculated masses were obtained using the amino acid sequence of the gene fragment expressed and the ProtPram tool. Observed masses are the neutral masses derived from qTOF data. This data was acquired by Dr Joe Gray at Pinn@cle.

It is of note that the calculated mass using analytical SEC is significantly different, both for s1319 (12.61 kDa) and s1522 (15.5 kDa). These estimates suggest significantly smaller proteins than are observed by both qTOF and SDS-PAGE analysis. Analytical SEC methods of estimating protein size assume that the proteins are globular, which may not be the case here. Currently available peptidoglycan deacetylase structures suggest a “head” domain, where the catalytic metal is found, and a longer “tail” as described in the structure of *S. pneumoniae* PgdA (PDB ID: 2C1G) (Blair *et al.*, 2005). If this is the case with s1319 and s1522,

this would affect the hydrodynamic radius of the protein, their elution from analytical SEC, and thus their estimated masses by this method.

### 5.4.2. ICP-MS

The metal content of both s1319 and s1522 was assessed by ICP-MS, as many identified NodB homology family members use metal ions in their catalysis (Blair *et al.*, 2005; Andrés *et al.*, 2014; Bhattacharjee *et al.*, 2017).

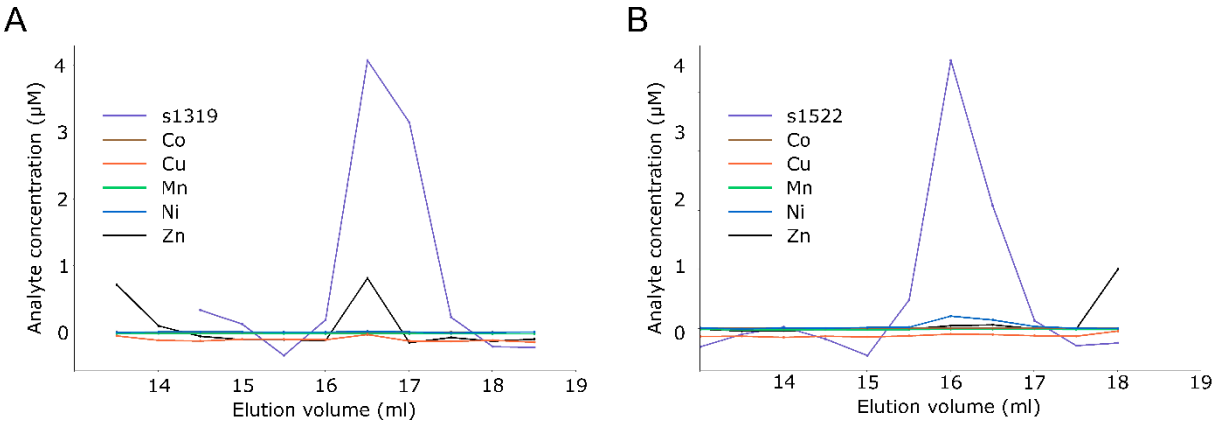


Figure 5-5: Metal content analysis of putative polysaccharide deacetylases.

Both s1319 (panel A) and s1522 (panel B) were further purified using an analytical SEC column, before analysis of metal content by ICP-MS. Protein concentration is shown in purple, zinc in black, manganese in green, nickel in blue, copper in orange and cobalt in gold. Repeats were not possible and therefore no standard deviations are shown.

As can be seen in Figure 5-5, neither protein appears to obviously coordinate any of the metal ions tested. However, there is some indication that s1319 might bind zinc to a small degree, as demonstrated by the slight increase in zinc eluting concurrently with s1319 (Figure 5-5A).

This experiment was only carried out once due to a breakdown of the ICP-MS instrument, and therefore biological and technical replicates will be required before any definitive conclusions are drawn from this data. Additionally, this experiment should be repeated with increasing concentrations of EDTA, to ensure that any metal is truly bound to the protein.

### 5.4.3. Circular dichroism and computational modelling

Before any investigations into the enzymatic capacities of CD630_1319 and CD630_1522 were carried out, CD was used to ascertain if the proteins were folded, and to determine their thermal stability. Crystallisation screens were performed as described in section 2.8.1. To date, no protein crystals have been produced. Therefore, the data retrieved from the CD experiments of the tag-free proteins was compared to computational modelling using PSIPRED (Buchan and Jones, 2019), I-TASSER (Yang *et al.*, 2015) and Phyre2 (Kelley *et al.*, 2015).

#### 5.4.3.1. Circular dichroism

Scans and thermal melts were performed as described in section 2.7.1, with the resulting spectra provided in Figure 5-6. Melting temperatures and secondary structure details are provided in Table 5-3. Secondary structure predictions from PSIPRED are provided for all constructs, with I-TASSER and Phyre2 predictions provided for soluble proteins.

Inspection of the spectra (Figure 5-6), and comparison to spectra of well characterised proteins, suggest that the proteins are composed of both α- helical and β-strand elements (Greenfield, 2006). The normalised root mean square deviation (NRMSD) value compares the observed and expected values of ellipticity at each wavelength, as calculated by DiChroweb (Whitmore and Wallace, 2004, 2008). Agreement between the theoretical and the observed scan is reflected in a lower NRMSD value, with NRMSD values lower than 0.05 considered acceptable. Here, all NRMSD are below 0.05 and therefore the data can be considered reliable. Low NRMSD values and the shape of the spectra suggests that all proteins are stable and folded at pH 8.0.

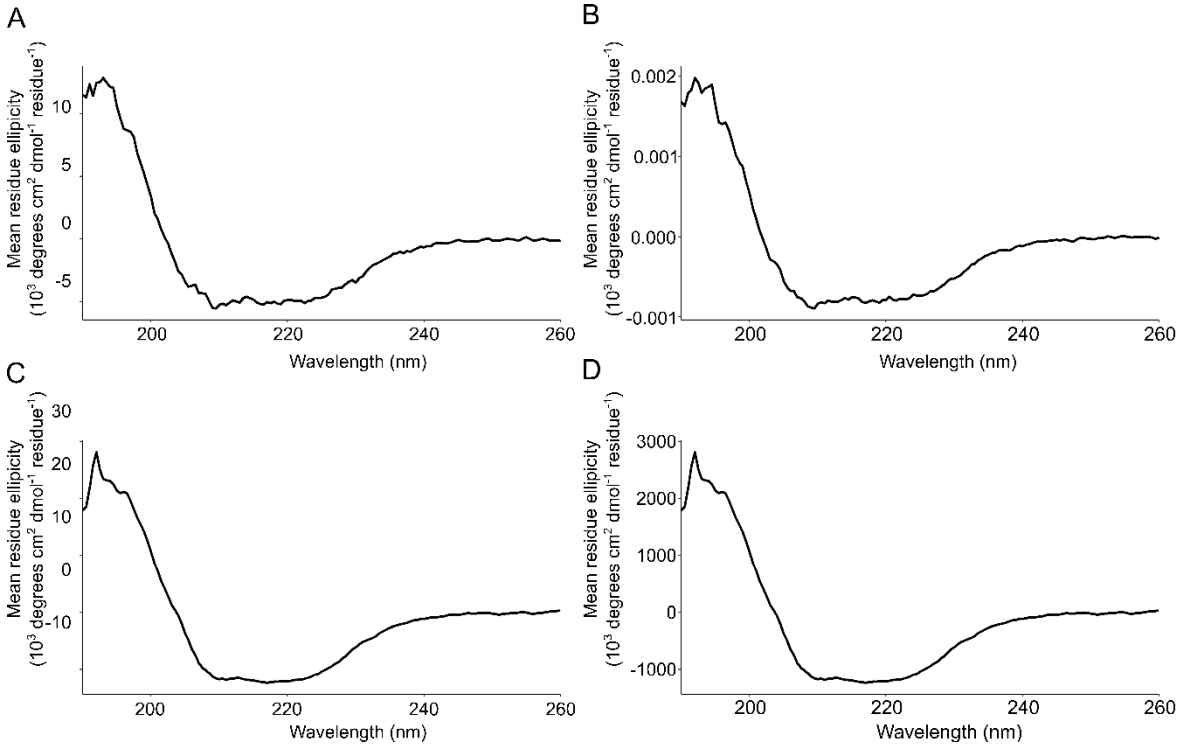


Figure 5-6: Circular dichroism of two putative deacetylases

s1319 (panel A), His1319 (panel B), s1522 (panel C) and His1522 (panel D) were analysed by CD, with the mean residue ellipticity (MRE) reported. The shapes of the graphs and MRE minima suggest a mixture of α-helical and β-strand structural elements.

Table 5-3: Circular dichroism of two putative polysaccharide deacetylases

| Protein | Method | Helix (%) | Strand (%) | Coil (%) | Disordered (%) | NRMSD | $T_m$ (°C) |
|---|---|---|---|---|---|---|---|
| His1319 | PsiPred | 39 | 9 | - | 49 | - | |
| | CDSSTR | 5 | 35 | 23 | 34 | 0.034 | 65 |
| S1319 | PsiPred | 41 | 15 | - | 43 | - | |
| | I-TASSER | 38 | 17 | 45 | - | - | |
| | Phyre2 | 43 | 17 | - | 9 | - | |
| | CDSSTR | 12 | 32 | 22 | 34 | 0.030 | 64 |
| His1522 | PsiPred | 45 | 10 | - | 46 | - | |
| | CDSSTR | 17 | 32 | 12 | 28 | 0.033 | 64 |
| S1522 | PsiPred | 49 | 10 | - | 41 | - | |
| | I-TASSER | 33 | 12 | 55 | - | - | |
| | Phyre2 | 42 | 12 | - | 26 | - | |
| | CDSSTR | 20 | 27 | 18 | 27 | 0.031 | 60 |

Results of CD scans processed using DiChroweb and the CDSSTR algorithm are summarised and compared with the predictions provided by the PSIPRED server. Melting temperatures derived from the first derivative using the JASCO software are also provided. CDSSTR α-helical and β-strand element values are the sum of the ordered and disordered values reported. Normalised root mean square deviation (NRMSD) values are provided for CDSSTR outputs. The disordered % for PSIPRED data includes all residues not accounted for by either α-helices or β-sheets.

#### 5.4.3.2. Computational modelling

Phyre2 (Kelley *et al.*, 2015) and I-TASSER (Yang *et al.*, 2015) are online tools able to predict protein structure based on a given amino acid sequence and structural models. Phyre2 is based on remote homology detection whereas I-TASSER is based on fold recognition and *ab initio* modelling.

A common enzymatic mechanism (Figure 5-7) has been proposed for peptidoglycan deacetylase activity, in which the bound metal and the associated Asp-His-His triad is central to catalytic activity (Bhattacharjee *et al.*, 2017). The metal ion is thought to interact directly with the acetate group of Glc*N*Ac (Bhattacharjee *et al.*, 2017), with substrate specificity determined by the presence and rearrangement of various loops around the substrate binding groove in a "substrate capping model" (Andrés *et al.*, 2014). A conserved proline residue (Figure 5-10), which undergoes hydroxylation of the α-carbon to produce 2-hydroxyproline, is also thought to be key to catalysis (Fadouloglou *et al.*, 2017). This post-translational modification increases the activity of a *V. cholerae* chitin deacetylase tenfold, and is thought to stabilise the oxyanion intermediate produced during catalysis (Fadouloglou *et al.*, 2017).

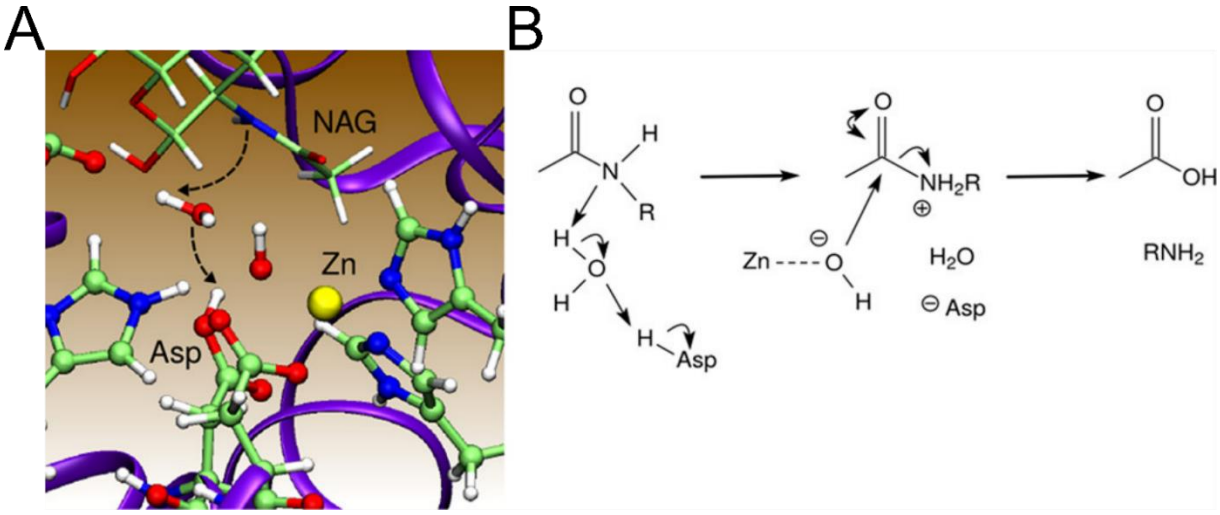


Figure 5-7: Proposed peptidoglycan deacetylase mechanism.

Panel A demonstrates summarises the enzymatic mechanism proposed by Bhattacharjee *et al* (2017). Panel B demonstrates how a conserved aspartate residue acts as a base, activating a water molecule, which then attacks the acetate group of Glc*N*Ac, breacking the carbon-nitrogen bond, thus deacetylating Glc*N*Ac. NAG, Glc*N*Ac. Figure modified from Bhattacharjee et al (2017).

##### 5.4.3.2.1. Modelling s1319 structure

When submitting the sequence of s1319, the highest scoring I-TASSER result is a model highly similar to *Bacillus cereus* Bc1960 Glc*N*Ac deacetylase (PDB ID 4L1G) (Fadouloglou *et al.*, 2017). The TM-score is 0.902 (two identical proteins will have a score of 1), with an RMSD of aligned residues of 1.28 Å and 0.944 coverage. Superimposition of the returned I-TASSER model and Bc1960 demonstrates the two structures are highly similar (Figure 5-8B). Phyre2 returns several possible models with 100% confidence of the query being a true structural homologue, therefore the model with the highest percentage sequence identity (28%) was analysed in more detail (Figure 5-8) .This structural homologue is *S. pneumoniae*

peptidoglycan deacetylase (*Sp*PgdA), a metal dependent Glc*N*Ac deacetylase (PDB ID: 2C1G) (Blair *et al.*, 2005). The coverage of this model is relatively limited (87%), with a significant proportion of *Sp*PgdA absent in the aligned s1319 section, however the catalytic region is well conserved. Both Bc1960 and *Sp*PgdA contain metal binding motifs (Blair *et al.*, 2005; Fadouloglou *et al.*, 2017) which are partially conserved in s1319 (Figure 5-8C and F, respectively); only the aspartate residue is imperfectly conserved in the sequence. However, a nearby aspartate residue may be able to contribute to metal coordination. Perhaps this is why we see ~25% Zn occupancy in the ICP-MS analysis (Figure 5-5A); in some situations, the aspartate residue may be in the correct conformation to permit zinc binding by s1319.

The I-TASSER and Phyre2 models, and the catalytic domains of 4L1G and 2C1G are highly similar; suggesting that s1319 adopts a NodB-like fold.

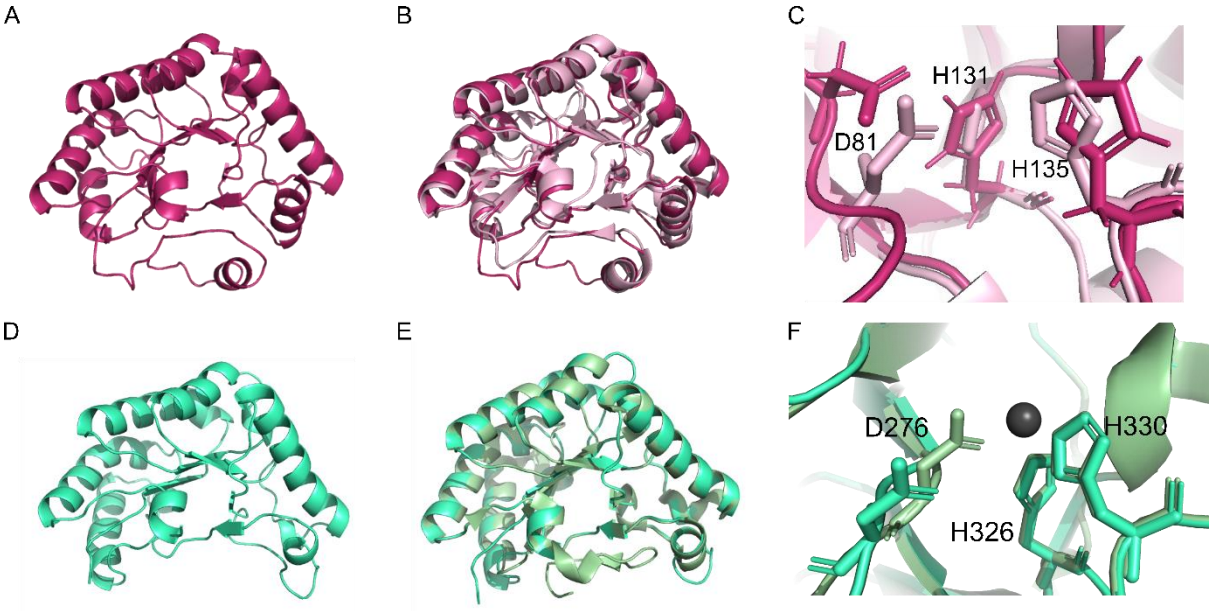


Figure 5-8: I-TASSER and Phyre2 modelling of s1319

Both I-TASSER (A) and Phyre2 (D) were used to model the folds adopted by s1319. Panel B shows the I-TASSER model of s1319 (dark pink) superimposed on the catalytic domain of the top structural homologue identified by I-TASSER, Bc1960 (light pink). Panel C shows Bc1960 metal binding residues Asp81, His 131 and His135. Panel E shows the superimposition of the Phyre2 model of s1319 on the template based on *Sp*PgdA. Panel F shows conservation of the zinc binding residues of *Sp*PdgA; Asp276, His326 and His330.

#### 5.4.3.2.2. Modelling s1522 structure

When submitting s1522 to I-TASSER, the model returned is most similar to a metal dependent *Streptococcus mutans* PgdA (PDB ID 2W3Z), with a TM-score of 0.804 and an RMSD of 1.97 Å with coverage at 85% (Figure 5-9B). The model provided by Phyre2 is based on *Vibrio cholerae* chitin deacetylase (Andrés *et al.*, 2014) (PDB ID 4NZ3) with 100% confidence, 25% identity and 74% coverage (Figure 5-9E).

Both structural homologues identified by I-TASSER and Phyre2 demonstrate metal binding through a conserved Asp-His-His triad (Deng *et al.*, 2009; Andrés *et al.*, 2014). Whilst s1522 appears to contain this triad, it may be that the Asp residue is not correctly positioned to

bind a metal ion. This would explain why no metal binding was observed during ICP-MS analysis (Figure 5-5B).

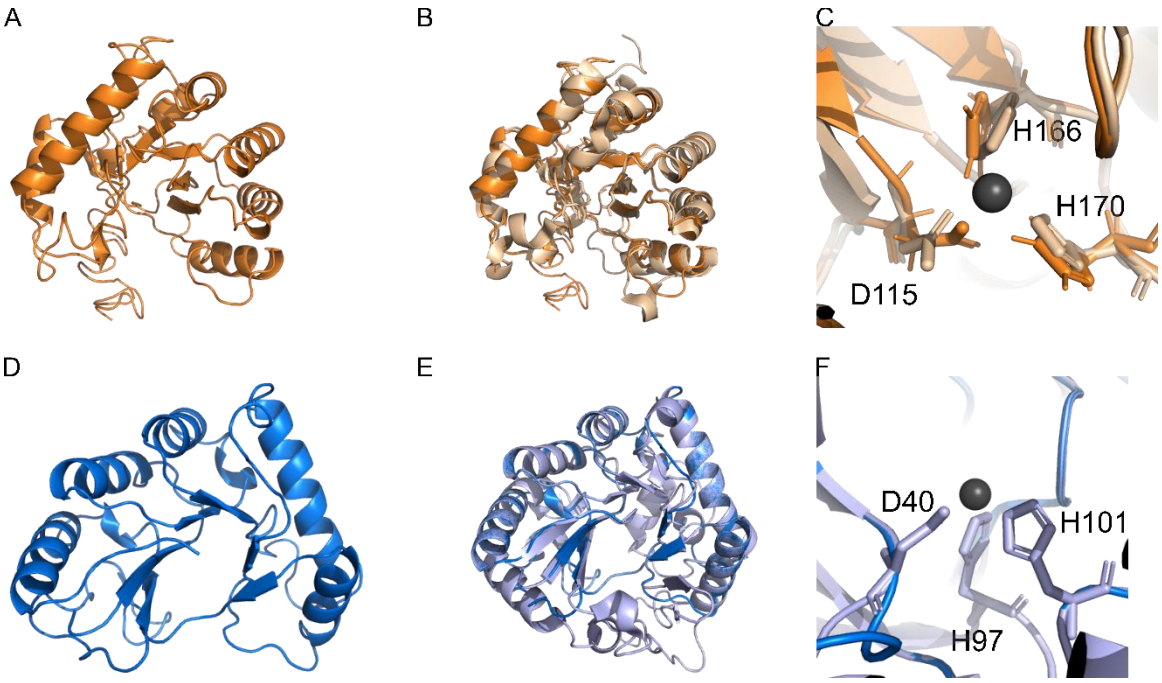


Figure 5-9: I-TASSER and Phyre2 models of s1522.

Both I-TASSER (A) and Phyre2 (D) were used to model the folds adopted by s1522. Panel B shows the I-TASSER model of s1319 (dark orange) superimposed on the catalytic domain of *S. mutans* PdgA, as identified by I-TASSER (lighter orange). Panel C shows the metal binding residues Asp155, His 166 and His170 of *S. mutans* PgdA. Panel E shows the Phyre2 model (dark blue) superimposed on the template based on a *Vibrio cholerae* CE4 family member (light blue). Panel F shows the almost identical conservation of the zinc binding residues of the confirmed CE4 family member: Asp97, His97 and His101.

Both s1319 and s1522 are predicted to fold into a NodB-like homology domain, by both Phyre2 and I-TASSER, and all highlighted structures were related to Carbohydrate Esterase family 4 (CE4) members (as per the CaZY database (www.cazy.org)). Therefore, s1319 and s1522 are likely to be true polysaccharide deacetylases and, potentially, peptidoglycan deacetylases.

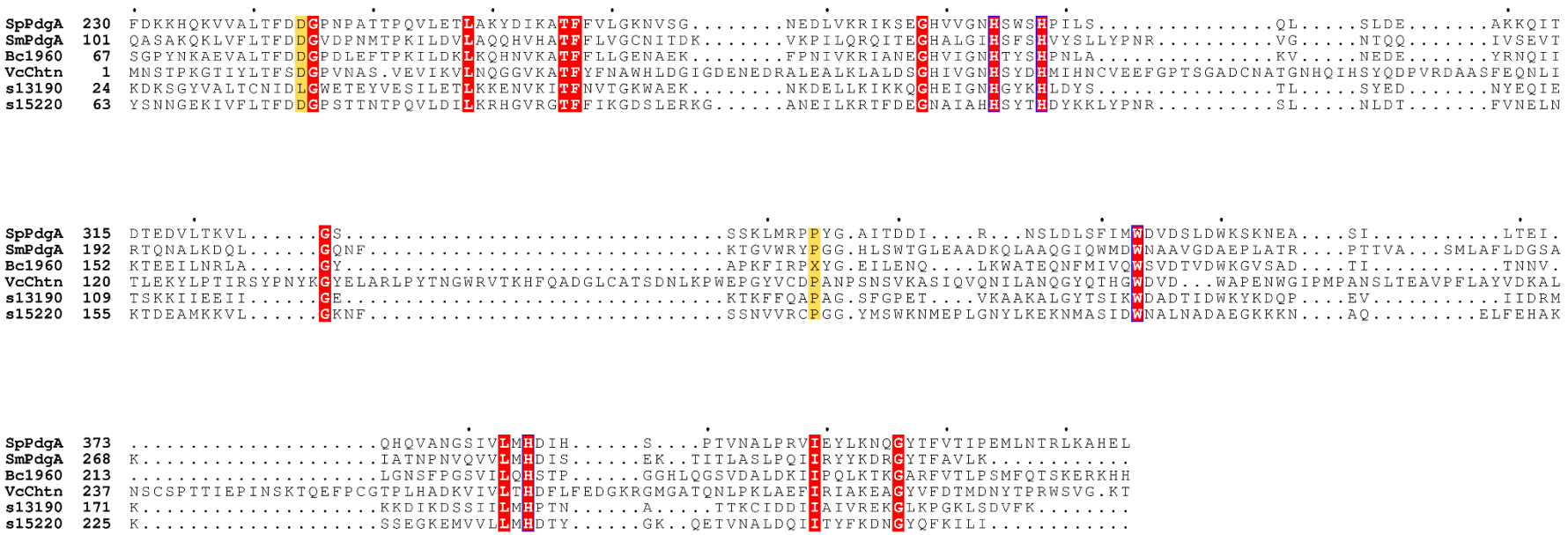


Figure 5-10: Multiple sequence alignment of CE4 family members

MUSCLE (Madeira *et al.*, 2019) was used to align the sequences of *S. pneumoniae* PdgA (*Sp*Pgda), *S. mutans* PdgA (*Sm*PdgA), *B. cereus* 1960 (Bc1960), *V. cholerae* Chitin de-N-acetylase (*Vc*Chtn), s1319 and s1522. Alignment figure was generated using ESPript 3.0 (Robert and Gouet, 2014) aligning residues 230-430 with the Blosum62 colouring system. Residues highlighted in red background are strictly conserved. Those highlighted in yellow were not identified as being strictly conserved in the MUSCLE but are conserved. The D residue (*Sp*PgdA residue 241) is conserved across all 6 proteins aligned expect for s1319, where the D is replaced by an L, however there is a D residue in the position immediately preceding this residue. The P (*Sp*PgdA position 334) is conserved across all 6 proteins but is presumably identified as X in Bc1960 as it is known to be subject to α-carbon hydroxylation (Fadouloglou *et al.*, 2017).

### 5.4.4. CD630_1319 and CD630_1522 encode peptidoglycan Glc*N*Ac deacetylases

Based on the predicted CE4 family membership of both CD630_1319 and CD630_1522, the peptidoglycan digestion activities for both the 6xHis-tagged and tag-free versions were investigated. *E. coli* Δ*lpp* peptidoglycan was digested with the protein (10µM) followed by overnight digestion with cellosyl, and the resulting muropeptides separated and identified using LC-MS as described in section 2.10.3.

Figure 5-11 demonstrates that *E. coli* peptidoglycan is acetylated; Glc*N*Ac containing muropeptides (muropeptides 4 and 7) are the norm when digested with cellosyl (lower panel Figure 5-11A). Digestion with SpoIID and SpoIIP was included as an additional control.

Digestion with either s1522 or H1522 produces largely similar chromatographs (Figure 5-11), with muropeptides 3 (GlcN Mur*N*Ac AEmA) and 5 (GlcN Mur*N*Ac AEmA AemA GlcN Mur*N*Ac) dominating the chromatograph. Analysis of the mass spectra produced from the eluate of these peaks demonstrate that all Glc*N*Ac residues have been deacetylated to GlcN by s1522 or H1522: muropeptides 4 (Glc*N*Ac Mur*N*Ac AEmA) and 7 (Glc*N*Ac Mur*N*Ac AEmA AEmA Glc*N*Ac Mur*N*Ac) are completely deacetylated to muropeptides 3 and 5. This demonstrates CD630_1522 encodes for an active Glc*N*Ac deacetylase and that the 6xHis tag does not impede activity.

Digestion of peptidoglycan with either s1319 or His1319 results in a more complex chromatograph (Figure 5-11). After digestion with these proteins, muropeptide 4 is fully deacetylated to produce muropeptide 3 but the deacetylation of muropeptide 7 is incomplete; producing both the partially deacetylated muropeptide 6, where Glc*N*Ac is present in one disaccharide and GlcN in the other, and the fully deacetylated muropeptide 5.

Neither s1319 nor His1319 deacetylated all the peptidoglycan available, perhaps suggesting the digestion conditions were sub-optimal. Alternatively, it may be the Glc*N*Ac residues were not oriented in a manner to allow digestion by His/s1319, or that only certain sites in the peptidoglycan are substrates for His/s1319 (i.e. a certain number of sugar residues may be needed for substrate recognition, which may not be possible at all locations on the glycan strand). Future work should involve optimising buffer conditions, changing pH and metal content as a priority to determine if the reaction can proceed further under different conditions.

The LC-MS analysis undertaken in this work demonstrates that both CD630_1522 and CD630_1319 encode for active Glc*N*Ac deacetylases (Figure 5-11), capable of deacetylating Glc*N*Ac residues on intact peptidoglycan. Purified s1522 was fully active under the conditions tested, whereas conditions for optimal s1319 activity should be investigated further, though the presence/absence of the 6xHis-tag does not appear to affect activity. Future work should involve optimisation of buffer composition and pH.

Further investigation into the substrate specificities of s1319 and s1522 should be undertaken, as some polysaccharide deacetylases are also capable of deacetylating a variety of substrates. Furthermore, kinetic experiments to determine the rate of reaction should be coupled with experiments to determine at what point in the cell cycle these deacetylases are expressed, as this may have implications for any future therapeutic interventions.

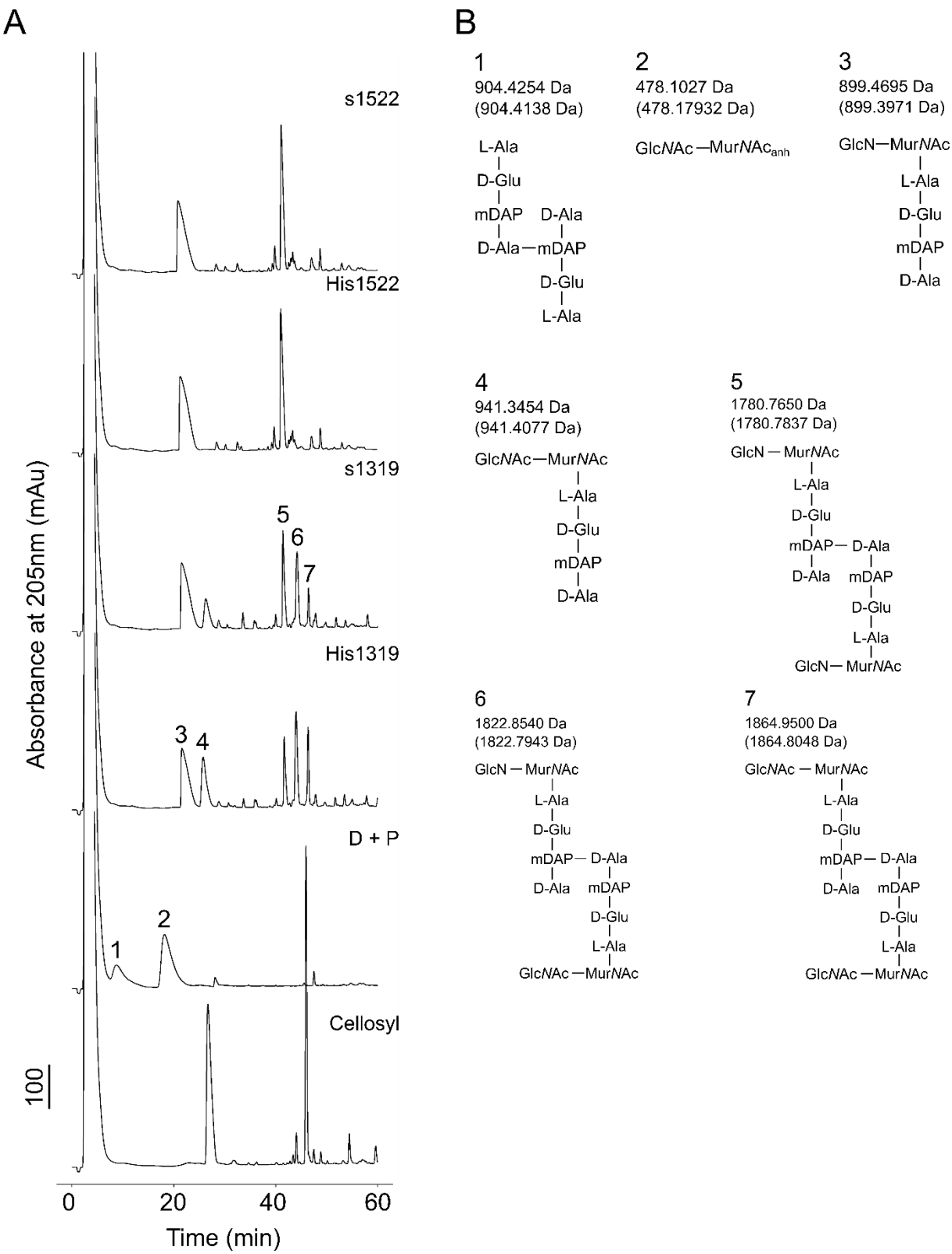


Figure 5-11: Peptidoglycan digestion by CD630_1319 and CD630_1522.

Panel A shows chromatographs resulting from *E. coli* Δ*lpp* peptidoglycan digestions. Numbered peaks correspond to the muropeptides illustrated in Panel B, where observed masses (top) and theoretical masses (bottom) are given in Daltons. D + P; SpoIID SpoIIP co-incubation. All muropeptides were reduced.

## 5.5. Discussion

Work in this chapter has furthered understanding of deacetylases in *C. difficile*, identifying 10 putative deacetylases and confirming that CD630_13190 and CD630_15220 encode for deacetylases capable of deacetylating Glc*N*Ac residues in intact peptidoglycan.

### 5.5.1. Identification of polysaccharide deacetylases in *C. difficile*

By searching the *C. difficile* 630 genome for homologues to SpPgdA, we identified 10 putative polysaccharide deacetylases (Table 5-1).

Based on the hypothesis that newly inserted peptidoglycan is preferentially digested by SpoIID and SpoIIP at the leading membrane edge during engulfment, we assessed both putative deacetylases that may be associated with the deacetylation of newly inserted peptidoglycan during sporulation, as well as putative deacetylases that may not be sporulation related as they may be down regulated during sporulation. *CD*630*_32570* has been identified as *sigE* dependent, and therefore also under the control of the master regulator *spo0A*. Whilst a construct lacking the predicted transmembrane region could be reliably expressed, the protein was insoluble, and therefore not pursued at this point.

*CD*630*_13190* has been identified as $\sigma^E$ controlled in multiple studies (Saujet *et al.*, 2013; Dembek *et al.*, 2015; Fimlaid and Shen, 2015) with the gene product being observed in the spore proteome (Lawley *et al.*, 2009). However, no significant change in translation was observed in the proteome of 630Δ*erm* Δ*spo0A* compared to 630Δ*erm* (Pettit *et al.*, 2014), perhaps indicating that *CD*630*_13190* control is multifactorial, or that *CD*630*_13190* is constitutively active and upregulated by *spo0A*. Due to the proposed $\sigma^E$ control, this protein may be expressed in early sporulation, and therefore may be involved in the deacetylation of newly synthesised peptidoglycan during spore formation.

*CD*630*_15220* was upregulated in the 630Δ*erm* Δ*spo0A* vegetative cell transcriptome and proteome (Pettit *et al.*, 2014), suggesting CD630_15220 may be suppressed either directly or indirectly by Spo0A. CD630_1522 is also proposed to interact with the product of *CD*630*_13190*, therefore this protein was also characterised.

### 5.5.2. Modelling CD630_1319 and CD630_1522

Phyre2 and I-TASSER are online tools that enable structure prediction based on similarities to structures already deposited in the PDB. Phyre2 is based on similarities to deposited structural models, I-TASSER employs iterative threading and *de novo* building to produce a model. Whilst both have inherent biases towards modelling the provided sequence onto models already in the PDB, that both servers consistently predict a NodB-like homology domain suggests it is highly likely that the two proteins contain this fold and therefore belong to the CE4 protein family (http://www.cazy.org/). Furthermore, the predicted percentage of α-helices for both I-TASSER and Phyre2 concur with the values obtained experimentally by CD, when one combines the helix and coil values provided by CDDSTR (Table 5-3).

Considering the that both proteins are likely members of the CE4 family, the presence of the conserved Asp-His-His metal binding triad (Figure 5-10), and the conservation of the oxyanion stabilising proline residue (Bhattacharjee *et al.*, 2017), it was surprising that our ICP-MS analysis suggested little to no metal bound to the purified proteins (Figure 5-5).

There was some indication that s1319 may bind zinc at 25% occupancy, but no metal tested was detected in the s1522 analysis. It may be that both s1319 and s1522 have a catalysis mechanism like that of *Colletotrichum lindemuthianum*, where no metal ion is required (Hekmat, Tokuyasu and Withers, 2003).

However, in the majority of CE4 family members, including proteins identified in the I-TASSER and Phyre2 modelling, the bound metal ion is central to catalysis (Blair *et al.*, 2005; Andrés *et al.*, 2014; Bhattacharjee *et al.*, 2017). It is expected that, *in vivo*, the majority of the CE4 family members bind a divalent metal ion. Zinc occupancy is preferred by some family members (Andrés *et al.*, 2014) whereas cobalt has been implicated in increasing the enzymatic activity of *Sp*PdgA by 30-fold, with provision of additional zinc increasing activity 5.5-fold (Blair *et al.*, 2005). Perhaps, if ICP-MS analysis was undertaken on *Sp*PdgA under native levels of zinc, it too would demonstrate only 25% occupancy, and that the enzymatic activity seen in Figure 5-11 is attributable to only 25% of the protein used containing zinc.

Whilst non-metal dependent CE4 family members have been described (Hekmat et al., 2003; Blair & van Aalten, 2004), zinc coordination is more common (Blair *et al.*, 2005; Deng *et al.*, 2009; Andrés *et al.*, 2014). Therefore, the metal binding capacity, if any, of s1319 and s1522 should be validated and, further ICP-MS analysis should be carried out to determine the metal occupancy of s1319 and 1522. Alternatively, if protein crystallisation screens prove successful, the identity of any metal present may be confirmed.

### 5.5.3. Future work

Future work would involve determining at what point during the cell cycle are the 10 identified deacetylases are expressed, potentially combining the sigma factor mutants used in Chapter 3, with RNA sequencing, microarray analysis and/or quantitative PCR. Localisation during cell cycle should also be probed; do deacetylases co-localise with other peptidoglycan synthesis/modification enzymesthe elongasome or divisome? This in turn may reveal more about the physiological function of these deacetylases, i.e. are they involved in vegetative growth and cell division, engulfment or the maturation of spore cortex peptidoglycan?

Determining the crystal structure of these two proteins will enable further understanding of the catalytic mechanism and may help determine the presence/absence of a metal ion in the conserved Asp-His-His triad. Investigation of the presence, and if applicable role, of the unusual 2-hydroxyproline PTM should also be investigated. Co-crystallisation with a panel of small molecules or with targeted molecules may also lead to the identification of novel inhibitors which may have therapeutic potential. Indeed, a histone deacetylase inhibitor has already been demonstrated to inhibit the deacetylase activity of 2 Glc*N*Ac deacetylases in *B. cereus* (Balomenou *et al.*, 2018), and several other deacetylase inhibitors have been developed (Chibba *et al.*, 2012; Ariyakumaran *et al.*, 2015; DiFrancesco, Morrison and Nitz, 2018).

To complement this potential drug discovery-based avenue of investigation, *C. difficile* deletion mutants of both deacetylases should be used to investigate the sporulation efficiency, lysozyme resistance, infectivity and persistence, and toxicity in disease models.

Additionally, the uncharacterised deacetylases identified should be investigated further; are they active against intact peptidoglycan? When are they expressed? Where are they expressed and are they essential for sporulation?

Whilst many questions remain about the two characterised deacetylases, and the remainder of the 10 proteins identified, the potential for therapeutic intervention suggests this could be a very interesting avenue of research in *C. difficile.*

# Chapter 6. Discussion

*C. difficile* continues to be a substantial healthcare burden, with the Center for Disease Control (USA) recognising *C. difficile* as an urgent threat. With the increasing resistance to the limited antibiotics available to treat CDI (Peng *et al.*, 2017; Boekhoud *et al.*, 2019) and a widening patient demographic (Garg *et al.*, 2013; Bloomfield and Riley, 2016), novel interventions are required to prevent and treat CDI. To this end, our knowledge of the fundamental biology of *C. difficile* must be advanced.

Spores contribute significantly to the pathogenesis of *C. difficile* by allowing transmission and persistence in the aerobic environment, due to their resistance to various commonly used decontamination methods (Deakin *et al.*, 2012; Barra-Carrasco and Paredes-Sabja, 2014; Janezic, Mlakar and Rupnik, 2018; Dyer *et al.*, 2019). As spores are so crucial to CDI it is tempting to identify interventions that could disrupt either their formation or persistence. Due to the possibility of spores that are resistant to germination, known as "super-dormant" spores (Zhang and Mathys, 2019), an intervention that could prevent *C. difficile* forming them in the first place is an attractive alternative. This requires a greater understanding of the *C. difficile* sporulation process, which this work sought to contribute to.

## 6.1. Peptidoglycan in *C. difficile*

The peptidoglycan of *C. difficile* vegetative cells is remarkable due to both the prevalence of 3-3 crosslinks, in comparison to the more canonical 4-3 linkage, and the high level of Glc*N*Ac deacetylation (Peltier *et al.*, 2011; Bern, Beniston and Mesnage, 2016). In addition to verifying the composition of *C. difficile* peptidoglycan (Chapter 3), the work presented here has identified several additional NCDAA termini present in vegetative cells peptidoglycan (section 3.1.3).

By analysing the peptidoglycan of several sporulation-defective mutants by FT-MS, we have been able to probe the changes to peptidoglycan throughout sporulation. Whilst the observed changes were not as dramatic as expected, they suggest both that peptidoglycan remodelling begins early in sporulation and that these changes may be associated with altering the crosslinking and deacetylation levels rather than the presence of novel spore cortex-related muropeptides. This analysis relies on a mixture of peptidoglycan isolated from vegetative and sporulating cells. It would be interesting to repeat this experiment in order to purify peptidoglycan from only those cells that have formed asymmetric septa to allow a more specific insight into the changes observed throughout sporulation.

## 6.2. *B. subtilis* and *C. difficile* - similar but different

*B. subtilis* is the model Gram-positive spore-former, with the majority of our understanding of sporulation originating from this organism (Illing and Errington, 1991; Popham *et al.*, 1996; Popham, Gilmore and Setlow, 1999; Molle *et al.*, 2003; Eichenberger *et al.*, 2004; Wang *et al.*, 2006; Camp and Losick, 2008; de Hoon, Eichenberger and Vitkup, 2010). However, as *C. difficile* has become more clinically relevant, and more research has been conducted, an increasing number of differences have been discovered between seemingly conserved processes between the two organisms (Pereira *et al.*, 2013; Saujet *et al.*, 2013, 2014).

Perhaps the most noticeable of these differences are those at the level of control of sporulation; despite the conservation of the sigma factors regulating sporulation, the *C. difficile* activation cascade has limited intercompartmental communication, whilst *B. subtilis* requires extensive communication between the mother cell and forespore compartments (Fimlaid *et al.*, 2013; Pereira *et al.*, 2013; Saujet *et al.*, 2013) (reviewed in (Paredes-Sabja, Shen and Sorg, 2014)).

This conservation of components but not necessarily of function is further reflected in the DMP machinery. SpoIIM is essential in *B. subtilis*, but dispensable for the formation of mature spores in *C. difficile* (Dembek *et al.*, 2018; Ribis, Fimlaid and Shen, 2018). Our investigation on the activities of SpoIID and SpoIIP (section 4.4), demonstrates that SpoIID does not enhance the activity of SpoIIP as is reported in *B. subtilis* (Chastanet and Losick, 2007). Furthermore, our biophysical studies (section 4.2.2.1) corroborate the evidence of Nocadello *et al* (2016); *C. difficile* SpoIID contains a tetrahedral zinc binding site, that is required for full activity, whereas this zinc site is absent in *B. subtilis* SpoIID (Nocadello *et al.*, 2016; Kelly and Salgado, 2019).

*C. difficile* SpoIIP was confirmed as a bifunctional endopeptidase and amidase (section 4.4). Interestingly, recent studies suggest that *C. difficile spoIIP* is under the control of $\sigma^F$ whereas in *B. subtilis spoIIP* is under the control of $\sigma^E$ (Piggot and Hilbert, 2004; Saujet *et al.*, 2013; Ribis, Fimlaid and Shen, 2018). This divergent regulation of expression also suggests divergent localisation, with $\sigma^F$ controlled genes expressed in the forespore and $\sigma^E$ in the mother cell; SpoIIP likely localises to the forespore during engulfment (Ribis, Fimlaid and Shen, 2018). Western blot analysis by Ribis *et al* (2018), demonstrated the presence of 3 isoforms of SpoIIP in *C. difficile* cell lysates, leading the authors to hypothesise that SpoIIP may exist as a free, membrane-unbound isoform that is produced in the forespore.

The DMP complex may have roles beyond that of peptidoglycan digestion. Peptidoglycan digestion by the DMP machinery is rate-limiting to engulfment, even when the septal barrier has been bypassed (Abanes-De Mello *et al.*, 2002; Gutierrez, Smith and Pogliano, 2010; Khanna *et al.*, 2019). The DMP complex has also been implicated in localising spore coat proteins; in *C. difficile* strains lacking DMP, the spore coat proteins SpoIVA and SipL mislocalised (Ribis, Fimlaid and Shen, 2018). Together, this suggests the DMP complex may have a role beyond peptidoglycan digestion in *C. difficile*.

### 6.3. Current model of engulfment

For the mother cell to engulf the forespore, it must overcome the barrier formed by the septal peptidoglycan; this lead to the septal thinning model of engulfment, where, starting at the middle of the septa and progressing outwards, the septal peptidoglycan is completely removed (Chastanet and Losick, 2007). However, the septal peptidoglycan is present as a thin layer throughout engulfment (Tocheva *et al.*, 2013; Lopez-Garrido *et al.*, 2018). Furthermore, peptidoglycan synthesis is required throughout engulfment; disruption of peptidoglycan synthesis prevents the mother cell from engulfing the forespore, though permits the asymmetric septa to curve into the mother cell (Meyer *et al.*, 2010; Ojkic *et al.*, 2016). This suggests that the septal thinning model does not completely explain the mechanisms of engulfment.

The Make-Before-Break model of engulfment was therefore proposed (Ojkic *et al.*, 2016); new peptidoglycan is synthesised ahead of the engulfing membrane, using one strand from the forespore and one from the lateral mother cell wall peptidoglycan as a template, before the newly inserted peptidoglycan is specifically degraded (potentially by the DMP complex). This creates space for the engulfing membrane to move into without complete dissolution of the septal peptidoglycan or compromising the integrity of the mother cell wall.

Recent investigations have developed this model (Khanna *et al.*, 2019). Using advanced electron microscopy techniques, the authors demonstrate that the septal peptidoglycan found between the mother cell and forespore compartments is not completely degraded at the start of engulfment from the centre out towards the advancing membrane, but is degraded uniformly across the entire septum with a layer of peptidoglycan present throughout, contrary to the septal thinning model. This slight thinning requires the DMP machinery, and is required to maintain a flexible septum (Khanna *et al.*, 2019).

Further investigations revealed that the newly synthesised peptidoglycan is inserted ahead of the advancing edge and the DMP machineries, and that insertion of this peptidoglycan results in deformation of the forespore membrane, presumably as peptidoglycan accumulates (Khanna *et al.*, 2019). As the mother cell moves around the forespore, finger-like projections are observed, which the authors suggest are a result of the newly inserted peptidoglycan that is yet to be digested by the DMP complex due to the limited number of DMP complexes available in the cell at any given time.

Furthermore, Khanna *et al*., (2019) suggest that DMP acts both to tether the engulfing membrane to peptidoglycan ahead of the leading edge, and to digest the newly inserted peptidoglycan to make room for the engulfing membrane to expand into. However, as the digestion of peptidoglycan by the DMP complex is rate limiting, peptidoglycan may be digested unevenly along the advancing edge, leading to the production of the finger-like projections.  This model is summarised in Figure 6-1.

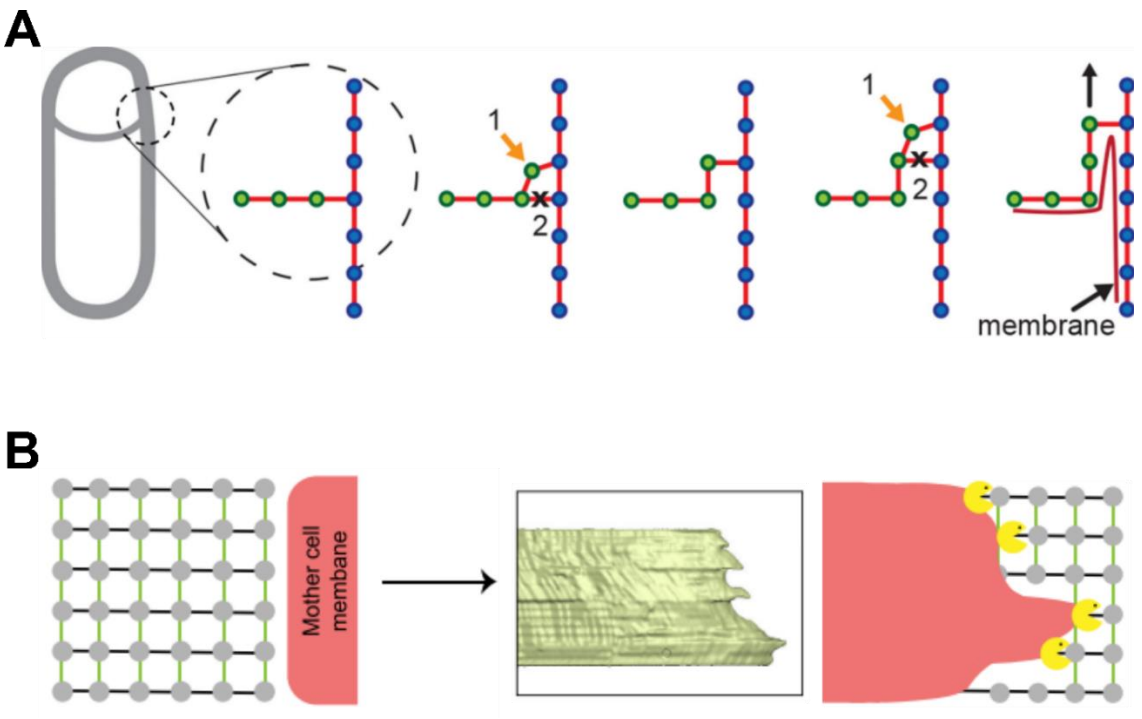


Figure 6-1: Current understanding of peptidoglycan remodelling during engulfment

Panel A shows the model proposed by Ojkic *et al*., (2016). New peptidoglycan (green) is inserted and bound to the mother cell peptidoglycan (blue) (marked 1). The peptidoglycan between the new and old strands is specifically digested (cross) before the process is repeated, allowing the membrane to advance. Panel B is modified from Khanna *et al*., (2019); the mother cell membrane (red) is tethered to the peptidoglycan (gray) that is inserted ahead of the membrane. The DMP complexes (yellow pacman) tether the mother cell membrane to the peptidoglycan in addition to their peptidoglycan digestion role, thus the mother cell may move ahead unevenly, resulting in finger-like projections. Figures modified from Ojkic e*t al.*, (2016) and Khanna *et al.*, (2019)

Interestingly, when peptidoglycan synthesis is inhibited, SpoIIP localisation to the leading edge is reduced, but still possible, suggesting that SpoIIP preferentially binds newly synthesised peptidoglycan (Ojkic *et al.*, 2016). Assuming this newly synthesised peptidoglycan is acetylated, this may explain our data suggesting that acetylated peptidoglycan is preferentially digested by SpoIID and SpoIIP (section 4.5). This in turn bolsters the hypothesis that the DMP complex can target the newly synthesised peptidoglycan due to the difference in acetylation seen between the “mature” deacetylated mother cell, and the newly synthesised, acetylated, septal peptidoglycan. Furthermore, the published crystal structure of SpoIID (Nocadello *et al.*, 2016) has interactions between the acetyl groups of the $NAG_3$ substrate and SpoIID throughout the proposed substrate binding groove, perhaps hinting at the possible preference of SpoIID, and the DMP complex as a whole, for acetylated peptidoglycan.

## 6.4. Peptidoglycan deacetylases

Whilst peptidoglycan deacetylases have been described in various organisms, only recently have *C. difficile* peptidoglycan deacetylases been identified and characterised (Ho *et al.*, 2014; Coullon *et al.*, 2018).

CD630_14300 and CD630_27190 were determined to encode Mur*N*Ac deacetylases and renamed *pdaA1* and *pdaA2* respectively. Deletion of *pdaA1* dramatically affects the peptidoglycan structure of *C. difficile* spores; muramic-δ-lactam residue-containing muropeptides represented only 0.4% of the spore cortex muropeptides, whereas in the parental strain, muramic-δ-lactam is present in ~24% of muropeptides. In a Δ*pdaA1*Δ*pdaA2* double mutant, muramic-δ-lactam-containing muropeptides are abolished completely; thus, PdaA1 and PdaA2 have been designated as the major and minor Mur*N*Ac deacetylases in *C. difficile*. Mutation of *pdaA1* resulted in reduced sporulation rate, delayed germination, and delayed virulence, implicating deacetylation of peptidoglycan in *C. difficile* fitness and pathogenesis (Coullon *et al*., 2018).

Here, we characterised 2 of 10 identified putative peptidoglycan deacetylases; CD630_1319 and CD630_1522. The non-membrane associated domains of both proteins were soluble when expressed in *E. coli*, and both proteins were active against *E. coli* peptidoglycan in peptidoglycan digestion experiments (section 5.4.4). Further characterisation of s1319 and s1522 should include verification of metal binding capacity, extended crystal screening and further investigation into their enzymatic activities. Of particular interest is investigation of any possible interactions between the DMP complex and the identified deacetylases. It may be that the deacetylases are differentially expressed during sporulation, leading to the newly synthesised septal peptidoglycan remaining acetylated long enough to be specifically targeted by the DMP complex for degradation during engulfment. This would allow the targeted peptidoglycan digestion described (Ojkic *et al.*, 2016; Khanna *et al.*, 2019), protecting the mother cell from premature lysis. Even if the deacetylases and the DMP complex do not directly interact, the impact of the deacetylases on the rate of SpoIIP and SpoIID activity should be investigated - do SpoIID and SpoIIP really prefer deacetylated peptidoglycan?

The deacetylases pose an interesting new avenue of research. Not only is little known about them in *C. difficile*, but deacetylases in other organisms have direct links to pathogenesis (Vollmer and Tomasz, 2000; Boneca *et al.*, 2007). Furthermore, peptidoglycan deacetylases have been identified as drug targets in *B. cereus* and *B. anthracis* (Balomenou *et al.*, 2018; Giastas *et al.*, 2018).

## 6.5. The engulfasome

The most recent models of engulfment (Ojkic *et al.*, 2016; Khanna *et al.*, 2019), combined with observations that SpoIIP may exist as a free isoform in *C. difficile* (Ribis, Fimlaid and Shen, 2018), that peptidoglycan synthesis is required for engulfment to continue (Meyer *et al.*, 2010) and that SpoIID and SpoIIP may preferentially digest acetylated peptidoglycan, together with observed interactions between DMP and the SpoIIQ:SpoIIIAH complex (Serrano *et al.*, 2016; Dembek *et al.*, 2018), lead to the development of the concept of the engulfasome (Kelly and Salgado, 2019).

Like the elongasome and divisome (Den Blaauwen *et al.*, 2008; Adams and Errington, 2009; Lutkenhaus, Pichoff and Du, 2012; Szwedziak and Löwe, 2013), the engulfasome would contain several proteins localising around a site of activity at a given time, in this case at the engulfing membrane. The engulfasome would span the distance between the mother cell and forespore allowing for both peptidoglycan digestion and synthesis in a controlled

manner. It seems likely that all the components of the system have yet to be identified, but we hypothesise that the DMP complex, Q:AH complex and a peptidoglycan synthesis and/or modifying machinery, at least, are present. The engulfasome does not refer to a static complex but rather to a dynamic association of various components and regulators throughout the engulfment process.

Furthermore, we hypothesise that whilst critical components may be conserved between spore-formers, the specific function of each component may deviate between species (Figure 6-2). For example, SpoIIM is essential for sporulation in *B. subtilis* (Chastanet and Losick, 2007), whilst it is largely dispensable in *C. difficile* (Dembek *et al.*, 2018; Ribis, Fimlaid and Shen, 2018).

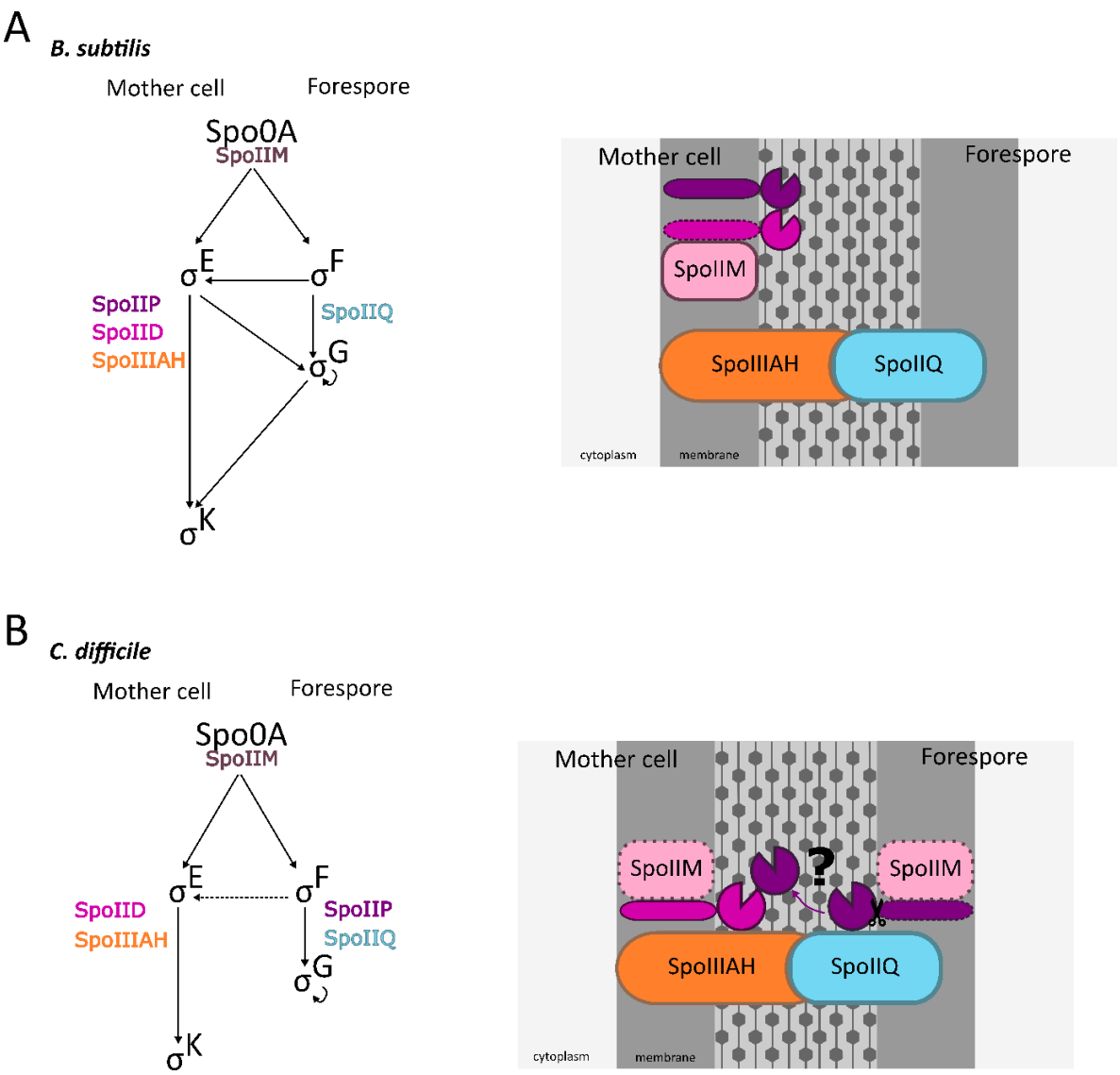


Figure 6-2: The engulfasome may differ between classes and species Panel A shows the sigma factor control of proposed engulfasome components in *B. subtilis*, with their arrangement at the engulfing membrane demonstrated in the righthand panel. Panel B demonstrates how the different sigma factor control of the proposed components may lead to an alternative organisation at the engulfing membrane. The proteins controlled by each sigma factor are coloured in the sigma factor cascade as they are in the schematic. Solid lines denote the dependency of one sigma factor upon another, with the dashed arrow depiciting a dispensible interaction. The scissors and question mark in the schematic of the *C. difficile* engulfasome refer to the possibiiltiy that *C. difficile* SpoIIP is processed to release a non-membranebound isoform (Ribis, Fimlaid and Shen, 2018). The dashed lines arouns SpoIIM demonstrate that SpoIIM is known to be dispensible for successful engulfment in *C. difficile* (Dembek *et al.*, 2018). Figure modified from Kelly and Salgado (2019).

## 6.6. Future directions

Several lines of enquiry are available to further characterise the engulfasome and its regulation.

The identification of 3 isoforms of SpoIIP, expected to be produced in the forespore (Ribis, Fimlaid and Shen, 2018), should be investigated. Are all 3 isoforms relevant *in vivo*?

Identifying whether SpoIIP is active when membrane-associated or free in the intermembrane space during engulfment is key in furthering our model of engulfment. To this end, super-resolution microscopy imaging using fluorescent tags such as SNAP or mCherry could be carried out.

The apparent preferential digestion of acetylated peptidoglycan by SpoIID and SpoIIP requires further investigation. To this end, time resolved peptidoglycan digestions with internal, quantifiable, controls should be developed. In these experiments, two parallel peptidoglycan digests, one containing chemically acetylated peptidoglycan and the other native *C. difficile* peptidoglycan, would be sampled at discrete time points, and the peptidoglycan characterised at various time points. This may highlight any differences in the extent of peptidoglycan digestion at a given timepoint depending on the peptidoglycan digested. In a similar manner, s1522 could be used as a tool to deacetylate *E. coli* peptidoglycan. The product of this reaction could then be used to investigate the impact of the type of crosslink present in the peptidoglycan, as *E. coli* peptidoglycan is primarily 4-3 crosslinked in comparison to the 3-3 crosslinks seen in *C. difficile*.

The identification and partial characterisation of s1319 and s1522 raises many questions. To investigate when in the cell cycle each putative deacetylase is expressed, a PCR-based approach, such as reverse transcriptase PCR or quantitative PCR could be employed. Alternatively luciferase reporter strains (Oliveira Paiva *et al.*, 2016) could be used in conjunction with the *C. difficile* 630 Δ*pyrE* Δ*erm* $P_{tet}$-*spo0A* strain. Similar to previous experiments (Oliveira Paiva *et al.*, 2016) cells would initially be grown in the absence of anhydrotetracycline, to ensure no cells were sporulating, the whole cell lysate of a sample of this culture would be taken to use as time point zero, sporulation would be induced, and samples regularly taken to assay the whole cell lysates for luciferase activity as the cells sporulated. Split-SNAP tagged proteins combined with super-resolution microscopy could allow for *in vivo* investigation of localisation and possible interactions (Cassona *et al.*, 2016; Serrano *et al.*, 2016). Interactions between deacetylases, and between deacetylases and other components of the engulfasome, could be investigated *in vitro* using BACTH or ELISA systems, or pull-down assays.

## 6.7. Final comments

This work set out to further the understanding of *C. difficile* engulfment. Here, we have characterised the peptidoglycan of vegetative cells and investigated the peptidoglycan of several engulfment-defective mutants to begin understanding peptidoglycan remodelling during sporulation. We have also characterised the peptidoglycan digestion activities of SpoIID and SpoIIP, generated a new hypothesis on how the DMP complex specifically targets the septal peptidoglycan for digestion during engulfment, and begun to test it by investigating a novel group of proteins potentially important for engulfment - the deacetylases.

Whilst many questions remain to be investigated, this work has contributed to the knowledge in the field and acts as the basis for further work.

# References

van der Aart, L. T. *et al.* (2018) 'High-Resolution Analysis of the Peptidoglycan Composition in *Streptomyces coelicolor*.', *Journal of bacteriology*. American Society for Microbiology Journals, 200(20), pp. e00290-18. doi: 10.1128/JB.00290-18.

Abanes-De Mello, A. *et al.* (2002) 'A cytoskeleton-like role for the bacterial cell wall during engulfment of the *Bacillus subtilis* forespore.', *Genes & development*, 16(24), pp. 3253–64. doi: 10.1101/gad.1039902.

Abecasis, A. B. *et al.* (2013) 'A genomic signature and the identification of new sporulation genes.', *Journal of bacteriology*. American Society for Microbiology Journals, 195(9), pp. 2101–15. doi: 10.1128/JB.02110-12.

Abo-Ghalia, M. *et al.* (1985) 'Specificity of the uridine-diphosphate-N-acetylmuramyl-l-alanyl-d-glutamate: meso-2,6-diaminopimelate synthetase from *Escherichia coli*', *European Journal of Biochemistry*, 153(1), pp. 81–87. doi: 10.1111/j.1432-1033.1985.tb09269.x.

Adams, D. W. and Errington, J. (2009) 'Bacterial cell division: assembly, maintenance and disassembly of the Z ring', *Nature Reviews Microbiology*, 7(9), pp. 642–653. doi: 10.1038/nrmicro2198.

Adams, P. D. *et al.* (2010) '*PHENIX* : a comprehensive Python-based system for macromolecular structure solution', *Acta Crystallographica Section D Biological Crystallography*. International Union of Crystallography, 66(2), pp. 213–221. doi: 10.1107/S0907444909052925.

Adams, P. D. *et al.* (2011) 'The Phenix software for automated determination of macromolecular structures', *Methods*, 55(1), pp. 94–106. doi: 10.1016/j.ymeth.2011.07.005.

Akerlund, T. *et al.* (2006) 'Correlation of disease severity with fecal toxin levels in patients with *Clostridium difficile*-associated diarrhea and distribution of PCR ribotypes and toxin yields in vitro of corresponding isolates.', *Journal of clinical microbiology*. American Society for Microbiology (ASM), 44(2), pp. 353–8. doi: 10.1128/JCM.44.2.353-358.2006.

Amano, K. *et al.* (1977) 'The Action of Lysozyme on Peptidoglycan with N-Unsubstituted Glucosamine Residues. Isolation of Glycan Fragments and Their Susceptibility to Lysozyme', *European Journal of Biochemistry*. John Wiley & Sons, Ltd (10.1111), 76(1), pp. 299–307. doi: 10.1111/j.1432-1033.1977.tb11596.x.

Ammam, F. *et al.* (2013) 'The functional *vanG* $_{Cd}$ cluster of *Clostridium difficile* does not confer vancomycin resistance', *Molecular Microbiology*. John Wiley & Sons, Ltd (10.1111), 89(4), pp. 612–625. doi: 10.1111/mmi.12299.

Andrés, E. *et al.* (2014) 'Structural basis of chitin oligosaccharide deacetylation.', *Angewandte Chemie (International ed. in English)*, 53(27), pp. 6882–7. doi: 10.1002/anie.201400220.

Aragunde, H., Biarnés, X. and Planas, A. (2018) 'Substrate Recognition and Specificity of Chitin Deacetylases and Related Family 4 Carbohydrate Esterases.', *International journal of molecular sciences*. Multidisciplinary Digital Publishing Institute (MDPI), 19(2). doi: 10.3390/ijms19020412.

Araki, Y. *et al.* (1971) 'Occurrence of non-N-substituted glucosamine residues in lysozyme-

resistant peptidoglycan from Bacillus cereus cell walls.', *Biochemical and biophysical research communications*, 42(4), pp. 691–7. doi: 10.1016/0006-291x(71)90543-2.

Ariyakumaran, R. *et al.* (2015) 'Direct Staudinger-Phosphonite Reaction Provides Methylphosphonamidates as Inhibitors of CE4 De-N-acetylases', *ChemBioChem*. John Wiley & Sons, Ltd, 16(9), pp. 1350–1356. doi: 10.1002/cbic.201500091.

Arnaouteli, S. *et al.* (2015) 'Two Putative Polysaccharide Deacetylases Are Required for Osmotic Stability and Cell Shape Maintenance in *Bacillus anthracis*.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 290(21), pp. 13465–78. doi: 10.1074/jbc.M115.640029.

Atrih, A. *et al.* (1996) 'Structural analysis of *Bacillus subtilis* 168 endospore peptidoglycan and its role during differentiation.', *Journal of bacteriology*, 178(21), pp. 6173–83. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=178487&tool=pmcentrez&rendertype=abstract (Accessed: 26 May 2016).

Atrih, A. *et al.* (1999) 'Analysis of Peptidoglycan Structure from Vegetative Cells of *Bacillus subtilis* 168 and Role of PBP 5 in Peptidoglycan Maturation', *J. Bacteriol.*, 181(13), pp. 3956–3966. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10383963 (Accessed: 25 May 2016).

Aung, S. *et al.* (2007) 'Dual localization pathways for the engulfment proteins during *Bacillus subtilis* sporulation.', *Molecular microbiology*, 65(6), pp. 1534–46. doi: 10.1111/j.1365-2958.2007.05887.x.

Baba, T. *et al.* (2006) 'Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection.', *Molecular systems biology*. European Molecular Biology Organization, 2, p. 2006.0008. doi: 10.1038/msb4100050.

Balomenou, S. *et al.* (2013) 'Distinct functions of polysaccharide deacetylases in cell shape, neutral polysaccharide synthesis and virulence of *Bacillus anthracis*', *Molecular Microbiology*. Wiley/Blackwell (10.1111), 87(4), pp. 867–883. doi: 10.1111/mmi.12137.

Balomenou, S. *et al.* (2018) 'Polysaccharide deacetylases serve as new targets for the design of inhibitors against *Bacillus anthracis* and *Bacillus cereus*', *Bioorganic & Medicinal Chemistry*. Pergamon, 26(13), pp. 3845–3851. doi: 10.1016/J.BMC.2018.06.045.

Barra-Carrasco, J. and Paredes-Sabja, D. (2014) '*Clostridium difficile* spores: a major threat to the hospital environment', *Future Microbiology*. Future Medicine Ltd London, UK , 9(4), pp. 475–486. doi: 10.2217/fmb.14.2.

Barreteau, H. *et al.* (2008) 'Cytoplasmic steps of peptidoglycan biosynthesis', *FEMS Microbiology Reviews*. Narnia, 32(2), pp. 168–207. doi: 10.1111/j.1574-6976.2008.00104.x.

Di Bella, S. *et al.* (2016) '*Clostridium difficile* Toxins A and B: Insights into Pathogenic Properties and Extraintestinal Effects.', *Toxins*. Multidisciplinary Digital Publishing Institute, 8(5), p. 134. doi: 10.3390/toxins8050134.

Benson, T. E. *et al.* (1993) 'Overexpression, purification, and mechanistic study of UDP-N-acetylenolpyruvylglucosamine reductase', *Biochemistry*. American Chemical Society, 32(8), pp. 2024–2030. doi: 10.1021/bi00059a019.

Bern, M., Beniston, R. and Mesnage, S. (2016) 'Towards an automated analysis of bacterial

peptidoglycan structure', *Analytical and Bioanalytical Chemistry*. Springer Berlin Heidelberg, pp. 1–10. doi: 10.1007/s00216-016-9857-5.

Bézay, N. *et al.* (2016) 'Safety, immunogenicity and dose response of VLA84, a new vaccine candidate against *Clostridium difficile*, in healthy volunteers.', *Vaccine*, 34(23), pp. 2585–2592. doi: 10.1016/j.vaccine.2016.03.098.

Bhattacharjee, N. *et al.* (2017) 'Catalytic Mechanism of Peptidoglycan Deacetylase: A Computational Study', *The Journal of Physical Chemistry B*. American Chemical Society, 121(1), pp. 89–99. doi: 10.1021/acs.jpcb.6b10625.

Bibbò, S. *et al.* (2014) 'Role of microbiota and innate immunity in recurrent *Clostridium difficile* infection.', *Journal of immunology research*. Hindawi, 2014, p. 462740. doi: 10.1155/2014/462740.

Billini, M. *et al.* (2019) 'A specialized MreB-dependent cell wall biosynthetic complex mediates the formation of stalk-specific peptidoglycan in *Caulobacter crescentus*', *PLOS Genetics*. Edited by J. Casadesús. Public Library of Science, 15(2), p. e1007897. doi: 10.1371/journal.pgen.1007897.

Den Blaauwen, T. *et al.* (2008) 'Morphogenesis of rod-shaped sacculi', *FEMS Microbiology Reviews*. John Wiley & Sons, Ltd (10.1111), 32(2), pp. 321–344. doi: 10.1111/j.1574-6976.2007.00090.x.

Blair, D. E. *et al.* (2005) 'Structure and metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal virulence factor.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 102(43), pp. 15429–34. doi: 10.1073/pnas.0504339102.

Blair, D. E. and van Aalten, D. M. F. (2004) 'Structures of Bacillus subtilis PdaA, a family 4 carbohydrate esterase, and a complex with N-acetyl-glucosamine.', *FEBS letters*, 570(1–3), pp. 13–9. doi: 10.1016/j.febslet.2004.06.013.

Bloomfield, L. E. and Riley, T. V (2016) 'Epidemiology and Risk Factors for Community-Associated *Clostridium difficile* Infection: A Narrative Review.', *Infectious diseases and therapy*. Springer, 5(3), pp. 231–51. doi: 10.1007/s40121-016-0117-y.

Blundell, J. K. and Perkins, H. R. (1981) 'Effects of beta-lactam antibiotics on peptidoglycan synthesis in growing *Neisseria gonorrhoeae*, including changes in the degree of O-acetylation.', *Journal of bacteriology*, 147(2), pp. 633–41. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6790518 (Accessed: 11 September 2019).

Boekhoud, I. M. *et al.* (2019) 'Plasmid-mediated metronidazole resistance in *Clostridioides difficile*', *bioRxiv*. Cold Spring Harbor Laboratory, p. 643775. doi: 10.1101/643775.

Boneca, I. G. *et al.* (2007) 'A critical role for peptidoglycan N-deacetylation in Listeria evasion from the host innate immune system.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 104(3), pp. 997–1002. doi: 10.1073/pnas.0609672104.

Bouhss, A. *et al.* (2008) 'The biosynthesis of peptidoglycan lipid-linked intermediates', *FEMS Microbiology Reviews*, 32(2), pp. 208–233. doi: 10.1111/j.1574-6976.2007.00089.x.

Boyle, D. S. and Donachie, W. D. (1998) 'mraY is an essential gene for cell growth in *Escherichia coli*.', *Journal of bacteriology*, 180(23), pp. 6429–32. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/9829961 (Accessed: 10 September 2019).

Briliūtė, J. *et al.* (2019) 'Complex N-glycan breakdown by gut Bacteroides involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci', *Nature Microbiology*. Nature Publishing Group, p. 1. doi: 10.1038/s41564-019-0466-x.

de Bruyn, G. *et al.* (2016) 'Defining the optimal formulation and schedule of a candidate toxoid vaccine against *Clostridium difficile* infection: A randomized Phase 2 clinical trial.', *Vaccine*, 34(19), pp. 2170–2178. doi: 10.1016/j.vaccine.2016.03.028.

Buchan, D. W. A. and Jones, D. T. (2019) 'The PSIPRED Protein Analysis Workbench: 20 years on', *Nucleic Acids Research*. doi: 10.1093/nar/gkz297.

Buffie, C. G. *et al.* (2015) 'Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile.', *Nature*. Nature Publishing Group, 517(7533), pp. 205–8. doi: 10.1038/nature13828.

Bui, N. K. *et al.* (2009) 'The peptidoglycan sacculus of *Myxococcus xanthus* has unusual structural features and is degraded during glycerol-induced myxospore development.', *Journal of bacteriology*. American Society for Microbiology Journals, 191(2), pp. 494–505. doi: 10.1128/JB.00608-08.

Calderón-Romero, P. *et al.* (2018) '*Clostridium difficile* exosporium cysteine-rich proteins are essential for the morphogenesis of the exosporium layer, spore resistance, and affect C. difficile pathogenesis', *PLOS Pathogens*. Edited by T. M. Koehler. Public Library of Science, 14(8), p. e1007199. doi: 10.1371/journal.ppat.1007199.

Cammarota, G., Ianiro, G. and Gasbarrini, A. (2014) 'Fecal Microbiota Transplantation for the Treatment of *Clostridium difficile Infection*', *Journal of Clinical Gastroenterology*, 48(8), pp. 693–702. doi: 10.1097/MCG.0000000000000046.

Camp, A. H. and Losick, R. (2008) 'A novel pathway of intercellular signalling in *Bacillus subtilis* involves a protein with similarity to a component of type III secretion channels', *Molecular Microbiology*, 69(2), pp. 402–417. doi: 10.1111/j.1365-2958.2008.06289.x.

Caparrós, M., Pisabarro, A. G. and de Pedro, M. A. (1992) 'Effect of D-amino acids on structure and synthesis of peptidoglycan in *Escherichia coli*.', *Journal of bacteriology*. American Society for Microbiology Journals, 174(17), pp. 5549–59. doi: 10.1128/jb.174.17.5549-5559.1992.

Carman, R. J. *et al.* (2011) '*Clostridium difficile* binary toxin (CDT) and diarrhea', *Anaerobe*, 17(4), pp. 161–165. doi: 10.1016/j.anaerobe.2011.02.005.

Cassona, C. P. *et al.* (2016) 'A Fluorescent Reporter for Single Cell Analysis of Gene Expression in *Clostridium difficile*', in, pp. 69–90. doi: 10.1007/978-1-4939-6361-4_6.

Caufrier, F. *et al.* (2003) 'Carbohydrate esterase family 4 enzymes: substrate specificity', *Carbohydrate Research*. Elsevier, 338(7), pp. 687–692. doi: 10.1016/S0008-6215(03)00002-8.

Cava, F. *et al.* (2011) 'Distinct pathways for modification of the bacterial cell wall by non-canonical D -amino acids', *The EMBO Journal*, 30(16), pp. 3442–3453. doi: 10.1038/emboj.2011.246.

Chastanet, A. and Losick, R. (2007) 'Engulfment during sporulation in *Bacillus subtilis* is governed by a multi-protein complex containing tandemly acting autolysins.', *Molecular*

*microbiology*, 64(1), pp. 139–52. doi: 10.1111/j.1365-2958.2007.05652.x.

Chen, V. B. *et al.* (2010) '*MolProbity* : all-atom structure validation for macromolecular crystallography', *Acta Crystallographica Section D Biological Crystallography*, 66(1), pp. 12–21. doi: 10.1107/S0907444909042073.

Chibba, A. *et al.* (2012) 'Synthesis and evaluation of inhibitors of *E. coli* PgaB, a polysaccharide de-N-acetylase involved in biofilm formation', *Organic & Biomolecular Chemistry*. The Royal Society of Chemistry, 10(35), p. 7103. doi: 10.1039/c2ob26105g.

Coullon, H. *et al.* (2018) '*N* -Deacetylases required for muramic-δ-lactam production are involved in *Clostridium difficile* sporulation, germination, and heat resistance', *Journal of Biological Chemistry*, 293(47), pp. 18040–18054. doi: 10.1074/jbc.RA118.004273.

Cowardin, C. A. *et al.* (2015) 'Inflammasome Activation Contributes to Interleukin-23 Production in Response to *Clostridium difficile*', *mBio*. Edited by G. B. Huffnagle, 6(1). doi: 10.1128/mBio.02386-14.

Crawshaw, A. D. *et al.* (2014) 'A mother cell-to-forespore channel: current understanding and future challenges.', *FEMS microbiology letters*. The Oxford University Press, 358(2), pp. 129–36. doi: 10.1111/1574-6968.12554.

Dapa, T. *et al.* (2013) 'Multiple factors modulate biofilm formation by the anaerobic pathogen *Clostridium difficile*', *Journal of Bacteriology*, 195(3). doi: 10.1128/JB.01980-12.

Deakin, L. J. *et al.* (2012) 'The *Clostridium difficile spo0A* Gene Is a Persistence and Transmission Factor', *Infection and Immunity*. Edited by A. J. Bäumler. American Society for Microbiology, 80(8), pp. 2704–2711. doi: 10.1128/IAI.00147-12.

Deeks, E. D. (2017) 'Bezlotoxumab: A Review in Preventing *Clostridium difficile* Infection Recurrence', *Drugs*, 77(15), pp. 1657–1663. doi: 10.1007/s40265-017-0809-y.

Dembek, M. *et al.* (2015) 'High-Throughput Analysis of Gene Essentiality and Sporulation in *Clostridium difficile*', *mBio*. Edited by R. J. Collier, 6(2), p. e02383. doi: 10.1128/mBio.02383-14.

Dembek, M. *et al.* (2017) 'Inducible Expression of spo0A as a Universal Tool for Studying Sporulation in *Clostridium difficile*', *Frontiers in Microbiology*. Frontiers, 8, p. 1793. doi: 10.3389/fmicb.2017.01793.

Dembek, M. *et al.* (2018) 'Peptidoglycan degradation machinery in *Clostridium difficile* forespore engulfment', *Molecular Microbiology*. John Wiley & Sons, Ltd (10.1111), 110(3), pp. 390–410. doi: 10.1111/mmi.14091.

Deng, D. M. *et al.* (2009) '*Streptococcus mutans SMU.623c* codes for a functional, metal-dependent polysaccharide deacetylase that modulates interactions with salivary agglutinin.', *Journal of bacteriology*, 191(1), pp. 394–402. doi: 10.1128/JB.00838-08.

DiFrancesco, B. R., Morrison, Z. A. and Nitz, M. (2018) 'Monosaccharide inhibitors targeting carbohydrate esterase family 4 de-N-acetylases', *Bioorganic & Medicinal Chemistry*. Pergamon, 26(21), pp. 5631–5643. doi: 10.1016/J.BMC.2018.10.008.

Dik, D. A. *et al.* (2017) 'Lytic transglycosylases: concinnity in concision of the bacterial cell wall', *Critical Reviews in Biochemistry and Molecular Biology*, 52(5), pp. 503–542. doi: 10.1080/10409238.2017.1337705.

Ducarmon, Q. R. *et al.* (2019) 'Gut Microbiota and Colonization Resistance against Bacterial Enteric Infection.', *Microbiology and molecular biology reviews : MMBR*. American Society for Microbiology, 83(3), pp. e00007-19. doi: 10.1128/MMBR.00007-19.

Duncan, K., Van Heijenoort, J. and Walsh, C. T. (1990) 'Purification and characterization of the D-alanyl-D-alanine-adding enzyme from *Escherichia coli*', *Biochemistry*, 29(9), pp. 2379–2386. doi: 10.1021/bi00461a023.

Dyer, C. *et al.* (2019) 'Biocide resistance and transmission of *Clostridium difficile* spores spiked onto clinical surfaces from an American healthcare facility.', *Applied and environmental microbiology*. American Society for Microbiology, p. AEM.01090-19. doi: 10.1128/AEM.01090-19.

Eckert, C. *et al.* (2015) 'Prevalence and pathogenicity of binary toxin–positive *Clostridium difficile* strains that do not produce toxins A and B', *New Microbes and New Infections*. Elsevier, 3, pp. 12–17. doi: 10.1016/J.NMNI.2014.10.003.

Eichenberger, P. *et al.* (2004) 'The Program of Gene Transcription for a Single Differentiating Cell Type during Sporulation in Bacillus subtilis', *PLoS Biology*. Edited by Jonathan A. Eisen, 2(10), p. e328. doi: 10.1371/journal.pbio.0020328.

Eichenberger, P., Fawcett, P. and Losick, R. (2002) 'A three-protein inhibitor of polar septation during sporulation in *Bacillus subtilis*', *Molecular Microbiology*. John Wiley & Sons, Ltd (10.1111), 42(5), pp. 1147–1162. doi: 10.1046/j.1365-2958.2001.02660.x.

Emsley, P. *et al.* (2010) 'Features and development of *Coot*', *Acta Crystallographica Section D Biological Crystallography*, 66(4), pp. 486–501. doi: 10.1107/S0907444910007493.

Evans, P. R., Murshudov, G. N. and IUCr (2013) 'How good are my data and what is the resolution?', *Acta Crystallographica Section D Biological Crystallography*. International Union of Crystallography, 69(7), pp. 1204–1214. doi: 10.1107/S0907444913000061.

Eze, P. *et al.* (2017) 'Risk factors for *Clostridium difficile* infections - an overview of the evidence base and challenges in data synthesis.', *Journal of global health*. Edinburgh University Global Health Society, 7(1), p. 010417. doi: 10.7189/jogh.07.010417.

Fadouloglou, V. E. *et al.* (2017) 'Unusual α-Carbon Hydroxylation of Proline Promotes Active-Site Maturation.', *Journal of the American Chemical Society*, 139(15), pp. 5330–5337. doi: 10.1021/jacs.6b12209.

Figueiredo, T. A. *et al.* (2012) 'Identification of Genetic Determinants and Enzymes Involved with the Amidation of Glutamic Acid Residues in the Peptidoglycan of *Staphylococcus aureus*', *PLoS Pathogens*. Edited by A. Cheung. Public Library of Science, 8(1), p. e1002508. doi: 10.1371/journal.ppat.1002508.

Fimlaid, K. A. *et al.* (2013) 'Global Analysis of the Sporulation Pathway of Clostridium difficile', *PLoS Genetics*. Edited by P. H. Viollier. Public Library of Science, 9(8), p. e1003660. doi: 10.1371/journal.pgen.1003660.

Fimlaid, K. A. and Shen, A. (2015) 'Diverse mechanisms regulate sporulation sigma factor activity in the Firmicutes.', *Current opinion in microbiology*. Elsevier Current Trends, 24, pp. 88–95. doi: 10.1016/j.mib.2015.01.006.

Frirdich, E. *et al.* (2019) 'The *Campylobacter jejuni* helical to coccoid transition involves changes to peptidoglycan and the ability to elicit an immune response', *Molecular*

*Microbiology*. John Wiley & Sons, Ltd (10.1111), 112(1), pp. 280–301. doi: 10.1111/mmi.14269.

Galperin, M. Y. *et al.* (2012) 'Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes.', *Environmental microbiology*, 14(11), pp. 2870–90. doi: 10.1111/j.1462-2920.2012.02841.x.

Garg, S. *et al.* (2013) 'Epidemiology of *Clostridium difficile*-Associated Disease (CDAD): A Shift from Hospital-Acquired Infection to Long-Term Care Facility-Based Infection', *Digestive Diseases and Sciences*, 58(12), pp. 3407–3412. doi: 10.1007/s10620-013-2848-x.

Gasteiger, E. *et al.* (2005) 'Protein Identification and Analysis Tools on the ExPASy Server', in *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press, pp. 571–607. doi: 10.1385/1-59259-890-0:571.

Gerding, D. N. *et al.* (2015) 'Administration of Spores of Nontoxigenic *Clostridium difficile* Strain M3 for Prevention of Recurrent C difficile Infection: A Randomized Clinical Trial.', *Jama*, 313(17). doi: 10.1001/jama.2015.3725.

Ghantoji, S. S. *et al.* (2010) 'Economic healthcare costs of *Clostridium difficile* infection: a systematic review', *Journal of Hospital Infection*. W.B. Saunders, 74(4), pp. 309–318. doi: 10.1016/J.JHIN.2009.10.016.

Ghuysen, J.-M. and Strominger, J. L. (1963) 'Structure of the Cell Wall of *Staphylococcus aureus*, Strain Copenhagen. II. Separation and Structure of Disaccharides', *Biochemistry*. American Chemical Society, 2(5), pp. 1119–1125. doi: 10.1021/bi00905a036.

Giastas, P. *et al.* (2018) 'Structures of the Peptidoglycan *N* -Acetylglucosamine Deacetylase Bc1974 and Its Complexes with Zinc Metalloenzyme Inhibitors', *Biochemistry*. American Chemical Society, 57(5), pp. 753–763. doi: 10.1021/acs.biochem.7b00919.

Gildea, R. J. *et al.* (2014) 'New methods for indexing multi-lattice diffraction data', *Acta Crystallographica Section D Biological Crystallography*. International Union of Crystallography, 70(10), pp. 2652–2666. doi: 10.1107/S1399004714017039.

Gilmore, M. E. *et al.* (2004) 'Production of muramic delta-lactam in *Bacillus subtilis* spore peptidoglycan.', *Journal of bacteriology*, 186(1), pp. 80–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14679227 (Accessed: 9 January 2019).

Girardin, S. E. *et al.* (2003) 'Nod2 Is a General Sensor of Peptidoglycan through Muramyl Dipeptide (MDP) Detection', *Journal of Biological Chemistry*, 278(11), pp. 8869–8872. doi: 10.1074/jbc.C200651200.

Glauner, B. (1988) 'Separation and quantification of muropeptides with high-performance liquid chromatography.', *Analytical biochemistry*, 172(2), pp. 451–64. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3056100 (Accessed: 4 July 2016).

Glauner, B., Höltje, J. V and Schwarz, U. (1988) 'The composition of the murein of *Escherichia coli*.', *The Journal of biological chemistry*, 263(21), pp. 10088–95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3292521 (Accessed: 4 September 2019).

Greenfield, N. J. (2006) 'Using circular dichroism spectra to estimate protein secondary structure.', *Nature protocols*. NIH Public Access, 1(6), pp. 2876–90. doi: 10.1038/nprot.2006.202.

Grohs, P. *et al.* (2000) 'Vancomycin resistance is associated with serine-containing peptidoglycan in *Enterococcus gallinarum*.', *Journal of bacteriology*. American Society for Microbiology Journals, 182(21), pp. 6228–32. doi: 10.1128/jb.182.21.6228-6232.2000.

Gutierrez, J., Smith, R. and Pogliano, K. (2010) 'SpoIID-mediated peptidoglycan degradation is required throughout engulfment during *Bacillus subtilis* sporulation.', *Journal of bacteriology*, 192(12), pp. 3174–86. doi: 10.1128/JB.00127-10.

Ha, S. *et al.* (2000) 'The 1.9 A crystal structure of *Escherichia coli* MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis.', *Protein science : a publication of the Protein Society*, 9(6), pp. 1045–52. doi: 10.1110/ps.9.6.1045.

Hayashi, K. (1975) 'A rapid determination of sodium dodecyl sulfate with methylene blue', *Analytical Biochemistry*, 67(2), pp. 503–506. doi: 10.1016/0003-2697(75)90324-3.

Heijenoort, J. van (2001) 'Recent advances in the formation of the bacterial peptidoglycan monomer unit (1985 to 2000)', *Natural Product Reports*. The Royal Society of Chemistry, 18(5), pp. 503–519. doi: 10.1039/a804532a.

Hekmat, O., Tokuyasu, K. and Withers, S. G. (2003) 'Subsite structure of the endo-type chitin deacetylase from a deuteromycete, *Colletotrichum lindemuthianum*: an investigation using steady-state kinetic analysis and MS.', *The Biochemical journal*. Portland Press Ltd, 374(Pt 2), pp. 369–80. doi: 10.1042/BJ20030204.

Herlihey, F. A. and Clarke, A. J. (2016) 'Controlling Autolysis During Flagella Insertion in Gram-Negative Bacteria', in. Springer, Singapore, pp. 41–56. doi: 10.1007/5584_2016_52.

Ho, T. D. *et al.* (2014) '*Clostridium difficile* extracytoplasmic function σ factor σV regulates lysozyme resistance and is necessary for pathogenesis in the hamster model of infection.', *Infection and immunity*. American Society for Microbiology, 82(6), pp. 2345–55. doi: 10.1128/IAI.01483-13.

Hoch, J. A. (1993) 'Regulation of the Phosphorelay and the Initiation of Sporulation in Bacillus Subtilis', *Annual Review of Microbiology*, 47(1), pp. 441–465. doi: 10.1146/annurev.mi.47.100193.002301.

Höltje, J. V *et al.* (1975) 'Novel type of murein transglycosylase in *Escherichia coli*.', *Journal of bacteriology*. American Society for Microbiology (ASM), 124(3), pp. 1067–76. Available at: http://www.ncbi.nlm.nih.gov/pubmed/357 (Accessed: 18 September 2017).

Hong, H. A. *et al.* (2017) 'The Spore Coat Protein CotE Facilitates Host Colonization by *Clostridium difficile*', *The Journal of Infectious Diseases*. Narnia, 216(11), pp. 1452–1459. doi: 10.1093/infdis/jix488.

de Hoon, M. J. L., Eichenberger, P. and Vitkup, D. (2010) 'Hierarchical evolution of the bacterial sporulation network.', *Current biology : CB*. Elsevier Ltd, 20(17), pp. R735-45. doi: 10.1016/j.cub.2010.06.031.

Horneck, G. *et al.* (2012) 'Resistance of bacterial endospores to outer space for planetary protection purposes--experiment PROTECT of the EXPOSE-E mission.', *Astrobiology*. Mary Ann Liebert, Inc., 12(5), pp. 445–56. doi: 10.1089/ast.2011.0737.

Hussain, H. A., Roberts, A. P. and Mullany, P. (2005) 'Generation of an erythromycin-sensitive derivative of Clostridium difficile strain 630 (630 erm) and demonstration that the conjugative transposon Tn916 E enters the genome of this strain at multiple sites', *Journal of*

*Medical Microbiology*. Microbiology Society, 54(2), pp. 137–141. doi: 10.1099/jmm.0.45790-0.

Ianiro, G. *et al.* (2018) 'Efficacy of different faecal microbiota transplantation protocols for Clostridium difficile infection: A systematic review and meta-analysis.', *United European gastroenterology journal*. SAGE Publications, 6(8), pp. 1232–1244. doi: 10.1177/2050640618780762.

Illing, N. and Errington, J. (1991) 'The spoIIIA operon of *Bacillus subtilis* defines a new temporal class of mother-cell-specific sporulation genes under the control of the ? E form of RNA polymerase', *Molecular Microbiology*. John Wiley & Sons, Ltd (10.1111), 5(8), pp. 1927–1940. doi: 10.1111/j.1365-2958.1991.tb00816.x.

Isidro, J. *et al.* (2017) 'Overview of *Clostridium difficile* Infection: Life Cycle, Epidemiology, Antimicrobial Resistance and Treatment', in *Clostridium Difficile - A Comprehensive Overview*. InTech. doi: 10.5772/intechopen.69053.

Janarthanan, S. *et al.* (2012) '*Clostridium difficile*-Associated Diarrhea and Proton Pump Inhibitor Therapy: A Meta-Analysis', *American Journal of Gastroenterology*, 107(7), pp. 1001–1010. doi: 10.1038/ajg.2012.179.

Janezic, S., Mlakar, S. and Rupnik, M. (2018) 'Dissemination of *Clostridium difficile* spores between environment and households: Dog paws and shoes', *Zoonoses and Public Health*, 65(6), pp. 669–674. doi: 10.1111/zph.12475.

Jones, D. T. (1999) 'Protein secondary structure prediction based on position-specific scoring matrices 1 1Edited by G. Von Heijne', *Journal of Molecular Biology*, 292(2), pp. 195–202. doi: 10.1006/jmbi.1999.3091.

Jutras, B. L. *et al.* (2019) '*Borrelia burgdorferi* peptidoglycan is a persistent antigen in patients with Lyme arthritis.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 116(27), pp. 13498–13507. doi: 10.1073/pnas.1904170116.

Kang, D. *et al.* (2008) '*Clostridium scindens baiCD* and *baiH* genes encode stereo-specific 7α/7β-hydroxy-3-oxo-Δ4-cholenoic acid oxidoreductases', *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1781(1–2), pp. 16–25. doi: 10.1016/j.bbalip.2007.10.008.

Kantardjieff, K. A. and Rupp, B. (2003) 'Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals.', *Protein science : a publication of the Protein Society*. Wiley-Blackwell, 12(9), pp. 1865–71. doi: 10.1110/ps.0350503.

Katajamaa, M., Miettinen, J. and Oresic, M. (2006) 'MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data', *Bioinformatics*. Oxford University Press, 22(5), pp. 634–636. doi: 10.1093/bioinformatics/btk039.

Kearse, M. *et al.* (2012) 'Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data', *Bioinformatics*, 28(12), pp. 1647–1649. doi: 10.1093/bioinformatics/bts199.

Kelley, L. A. *et al.* (2015) 'The Phyre2 web portal for protein modeling, prediction and analysis', *Nature Protocols*. Nature Publishing Group, 10(6), pp. 845–858. doi:

10.1038/nprot.2015.053.

Kelly, A. (2016) *MRes dissertation: Analysis of peptidoglycan composition during sporulation in* Clostridium difficile.

Kelly, A. and Salgado, P. S. (2019) 'The engulfasome in C. difficile: Variations on protein machineries', *Anaerobe*. Academic Press, p. 102091. doi: 10.1016/J.ANAEROBE.2019.102091.

Kester, J. C. *et al.* (2019) '*C. difficile*-associated antibiotics prime the host for infection by a microbiome-independent mechanism.', *bioRxiv*. Cold Spring Harbor Laboratory, p. 728170. doi: 10.1101/728170.

Khanna, K. *et al.* (2019) 'The molecular architecture of engulfment during *Bacillus subtilis* sporulation', *eLife*, 8. doi: 10.7554/eLife.45257.

Kim, S. J., Chang, J. and Singh, M. (2015) 'Peptidoglycan architecture of Gram-positive bacteria by solid-state NMR', *Biochimica et Biophysica Acta - Biomembranes*. doi: 10.1016/j.bbamem.2014.05.031 Review.

Kitchin, N. *et al.* (2019) 'A Phase 2 Study Evaluating the Safety, Tolerability, and Immunogenicity of Two 3-Dose Regimens of a *Clostridium difficile* Vaccine in Healthy US Adults Aged 65 to 85 Years', *Clinical Infectious Diseases*. doi: 10.1093/cid/ciz153.

Kleywegt, G. J. and Jones, T. A. (1997) 'Model building and refinement practice.', *Methods in enzymology*, 277, pp. 208–30. doi: 10.1016/s0076-6879(97)77013-7.

Kochan, T. J. *et al.* (2018) 'Updates to *Clostridium difficile* Spore Germination.', *Journal of bacteriology*. American Society for Microbiology Journals, 200(16), pp. e00218-18. doi: 10.1128/JB.00218-18.

Kock, H., Gerth, U. and Hecker, M. (2004) 'MurAA, catalysing the first committed step in peptidoglycan biosynthesis, is a target of Clp-dependent proteolysis in *Bacillus subtilis*', *Molecular Microbiology*. John Wiley & Sons, Ltd (10.1111), 51(4), pp. 1087–1102. doi: 10.1046/j.1365-2958.2003.03875.x.

Kühner, D. *et al.* (2014) 'From cells to muropeptide structures in 24 h: peptidoglycan mapping by UPLC-MS.', *Scientific reports*. Nature Publishing Group, 4, p. 7494. doi: 10.1038/srep07494.

Kumar, S. *et al.* (2019) 'The bacterial lipid II flippase MurJ functions by an alternating-access mechanism.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 294(3), pp. 981–990. doi: 10.1074/jbc.RA118.006099.

Lawley, T. D. *et al.* (2009) 'Proteomic and genomic characterization of highly infectious *Clostridium difficile* 630 spores.', *Journal of bacteriology*, 191(17), pp. 5377–86. doi: 10.1128/JB.00597-09.

Lawson, P. A. *et al.* (2016) 'Reclassification of *Clostridium difficile* as *Clostridioides difficile* (Hall and O'Toole 1935) Prévot 1938', *Anaerobe*, 40, pp. 95–99. doi: 10.1016/j.anaerobe.2016.06.008.

Leslie, J. L. *et al.* (2019) 'The Gut Microbiota Is Associated with Clearance of Clostridium difficile Infection Independent of Adaptive Immunity', *mSphere*. American Society for Microbiology (ASM), 4(1). doi: 10.1128/MSPHEREDIRECT.00698-18.

Li, H. and Cao, Y. (2010) 'Lactic acid bacterial cell factories for gamma-aminobutyric acid', *Amino Acids*, 39(5), pp. 1107–1116. doi: 10.1007/s00726-010-0582-7.

Liger, D. *et al.* (1995) 'Over-production, purification and properties of the uridine-diphosphate-N-acetylmuramate:L-alanine ligase from *Escherichia coli*.', *European journal of biochemistry*, 230(1), pp. 80–7. doi: 10.1111/j.1432-1033.1995.0080i.x.

Logardt, I. M. and Neujahr, H. Y. (1975) 'Lysis of modified walls from Lactobacillus fermentum.', *Journal of bacteriology*, 124(1), pp. 73–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1176437 (Accessed: 31 October 2019).

Lombard, V. *et al.* (2014) 'The carbohydrate-active enzymes database (CAZy) in 2013', *Nucleic Acids Research*, 42(D1), pp. D490–D495. doi: 10.1093/nar/gkt1178.

Lopez-Garrido, J. *et al.* (2018) 'Chromosome Translocation Inflates Bacillus Forespores and Impacts Cellular Morphology.', *Cell*. NIH Public Access, 172(4), pp. 758-770.e14. doi: 10.1016/j.cell.2018.01.027.

Lupoli, T. J. *et al.* (2011) 'Transpeptidase-mediated incorporation of D-amino acids into bacterial peptidoglycan.', *Journal of the American Chemical Society*. NIH Public Access, 133(28), pp. 10748–51. doi: 10.1021/ja2040656.

Lutkenhaus, J., Pichoff, S. and Du, S. (2012) 'Bacterial cytokinesis: From Z ring to divisome', *Cytoskeleton*, 69(10), pp. 778–790. doi: 10.1002/cm.21054.

Mackin, K. E. *et al.* (2013) 'Spo0A Differentially Regulates Toxin Production in Evolutionarily Diverse Strains of *Clostridium difficile*', *PLoS ONE*. Edited by M. R. Popoff. Public Library of Science, 8(11), p. e79666. doi: 10.1371/journal.pone.0079666.

Madeira, F. *et al.* (2019) 'The EMBL-EBI search and sequence analysis tools APIs in 2019.', *Nucleic acids research*, 47(W1), pp. W636–W641. doi: 10.1093/nar/gkz268.

Manat, G. *et al.* (2014) 'Deciphering the metabolism of undecaprenyl-phosphate: the bacterial cell-wall unit carrier at the membrane frontier.', *Microbial drug resistance (Larchmont, N.Y.)*. Mary Ann Liebert, Inc., 20(3), pp. 199–214. doi: 10.1089/mdr.2014.0035.

Marquardt, J. L. *et al.* (1992) 'Cloning and sequencing of *Escherichia coli* murZ and purification of its product, a UDP-N-acetylglucosamine enolpyruvyl transferase.', *Journal of Bacteriology*, 174(17), pp. 5748–5752. doi: 10.1128/jb.174.17.5748-5752.1992.

Martinou, A., Koutsioulis, D. and Bouriotis, V. (2002) 'Expression, purification, and characterization of a cobalt-activated chitin deacetylase (Cda2p) from *Saccharomyces cerevisiae*.', *Protein expression and purification*, 24(1), pp. 111–6. doi: 10.1006/prep.2001.1547.

McCoy, A. J. *et al.* (2007) '*Phaser* crystallographic software', *Journal of Applied Crystallography*. International Union of Crystallography, 40(4), pp. 658–674. doi: 10.1107/S0021889807021206.

Meisner, J. *et al.* (2008) 'A channel connecting the mother cell and forespore during bacterial endospore formation.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 105(39), pp. 15100–5. doi: 10.1073/pnas.0806301105.

Meroueh, S. O. *et al.* (2006) 'Three-dimensional structure of the bacterial cell wall

peptidoglycan.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 103(12), pp. 4404–9. doi: 10.1073/pnas.0510182103.

Meyer, P. *et al.* (2010) 'Cell wall synthesis is necessary for membrane dynamics during sporulation of *Bacillus subtilis*.', *Molecular microbiology*, 76(4), pp. 956–970. doi: 10.1111/j.1365-2958.2010.07155.x.

Mohammadi, T. *et al.* (2011) 'Identification of FtsW as a transporter of lipid-linked cell wall precursors across the membrane', *The EMBO Journal*, 30(8), pp. 1425–1432. doi: 10.1038/emboj.2011.61.

Molle, V. *et al.* (2003) 'The Spo0A regulon of *Bacillus subtilis*', *Molecular Microbiology*, 50(5), pp. 1683–1701. doi: 10.1046/j.1365-2958.2003.03818.x.

Moller, S., Croning, M. D. R. and Apweiler, R. (2001) 'Evaluation of methods for the prediction of membrane spanning regions', *Bioinformatics*. Narnia, 17(7), pp. 646–653. doi: 10.1093/bioinformatics/17.7.646.

Monteiro, J. M. *et al.* (2019) 'The pentaglycine bridges of *Staphylococcus aureus* peptidoglycan are essential for cell integrity', *Scientific Reports*. Nature Publishing Group, 9(1), p. 5010. doi: 10.1038/s41598-019-41461-1.

Mora-Uribe, P. *et al.* (2016) 'Characterization of the Adherence of *Clostridium difficile* Spores: The Integrity of the Outermost Layer Affects Adherence Properties of Spores of the Epidemic Strain R20291 to Components of the Intestinal Mucosa', *Frontiers in Cellular and Infection Microbiology*, 6, p. 99. doi: 10.3389/fcimb.2016.00099.

Morlot, C. *et al.* (2010) 'A highly coordinated cell wall degradation machine governs spore morphogenesis in *Bacillus subtilis*.', *Genes & development*. Cold Spring Harbor Laboratory Press, 24(4), pp. 411–22. doi: 10.1101/gad.1878110.

Morlot, C. and Rodrigues, C. D. A. (2018) 'The New Kid on the Block: A Specialized Secretion System during Bacterial Sporulation', *Trends in Microbiology*. Elsevier Current Trends, 26(8), pp. 663–676. doi: 10.1016/J.TIM.2018.01.001.

Moynihan, P. J., Sychantha, D. and Clarke, A. J. (2014) 'Chemical biology of peptidoglycan acetylation and deacetylation', *Bioorganic Chemistry*, 54, pp. 44–50. doi: 10.1016/j.bioorg.2014.03.010.

Nakamura, A. M., Nascimento, A. S. and Polikarpov, I. (2017) 'Structural diversity of carbohydrate esterases', *Biotechnology Research and Innovation*. Elsevier, 1(1), pp. 35–51. doi: 10.1016/J.BIORI.2017.02.001.

Ng, J. *et al.* (2010) '*Clostridium difficile* Toxin–Induced Inflammation and Intestinal Injury Are Mediated by the Inflammasome', *Gastroenterology*, 139(2), pp. 542-552.e3. doi: 10.1053/j.gastro.2010.04.005.

Ng, K. M. *et al.* (2013) 'Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens', *Nature*. Nature Publishing Group, 502(7469), pp. 96–99. doi: 10.1038/nature12503.

Ng, Y. K. *et al.* (2013) 'Expanding the Repertoire of Gene Tools for Precise Manipulation of the *Clostridium difficile* Genome: Allelic Exchange Using pyrE Alleles', *PLoS ONE*. Edited by M. R. Popoff. Public Library of Science, 8(2), p. e56051. doi: 10.1371/journal.pone.0056051.

Ngadjeua, F. *et al.* (2018) 'Critical Impact of Peptidoglycan Precursor Amidation on the Activity of l,d -Transpeptidases from *Enterococcus faecium* and *Mycobacterium tuberculosis*', *Chemistry - A European Journal*. John Wiley & Sons, Ltd, 24(22), pp. 5743–5747. doi: 10.1002/chem.201706082.

Nicholson, W. L. *et al.* (2000) 'Resistance of Bacillus endospores to extreme terrestrial and extraterrestrial environments.', *Microbiology and molecular biology reviews : MMBR*. American Society for Microbiology, 64(3), pp. 548–72. doi: 10.1128/MMBR.64.3.548-572.2000.

Nocadello, S. *et al.* (2016) 'Crystal structures of the SpoIID lytic transglycosylases essential for bacterial sporulation.', *The Journal of biological chemistry*. doi: 10.1074/jbc.M116.729749.

Ohno, N., Yadomae, T. and Miyazaki, T. (1982) 'Identification of 2-amino-2-deoxyglucose residues in the peptidoglucan of Streptococcus pneumoniae.', *Carbohydrate research*, 107(1), pp. 152–5. doi: 10.1016/s0008-6215(00)80785-5.

Ojkic, N. *et al.* (2016) 'Cell wall remodeling drives engulfment during *Bacillus subtilis* sporulation', *eLife*. eLife Sciences Publications Limited, 5, p. e18657. doi: 10.7554/eLife.18657.

Oliveira Paiva, A. M. *et al.* (2016) 'The Signal Sequence of the Abundant Extracellular Metalloprotease PPEP-1 Can Be Used to Secrete Synthetic Reporter Proteins in Clostridium difficile', *ACS Synthetic Biology*, 5(12), pp. 1376–1382. doi: 10.1021/acssynbio.6b00104.

Oren, A. and Rupnik, M. (2018) '*Clostridium difficile* and *Clostridioides difficile*: Two validly published and correct names', *Anaerobe*. Academic Press, 52, pp. 125–126. doi: 10.1016/J.ANAEROBE.2018.07.005.

Paredes-Sabja, D., Shen, A. and Sorg, J. A. (2014) '*Clostridium difficile* spore biology: sporulation, germination, and spore structural proteins.', *Trends in microbiology*, 22(7), pp. 406–16. doi: 10.1016/j.tim.2014.04.003.

Peltier, J. *et al.* (2011) '*Clostridium difficile* Has an Original Peptidoglycan Structure with a High Level of N-Acetylglucosamine Deacetylation and Mainly 3-3 Cross-links', *Journal of Biological Chemistry*, 286(33), pp. 29053–29062. doi: 10.1074/jbc.M111.259150.

Peltier, J. *et al.* (2013) 'Genomic and expression analysis of the vanG-like gene cluster of *Clostridium difficile*', *Microbiology*, (159), pp. 1510–1520. doi: 10.1099/mic.0.065060-0.

Peng, Z. *et al.* (2017) 'Update on Antimicrobial Resistance in *Clostridium difficile*: Resistance Mechanisms and Antimicrobial Susceptibility Testing', *Journal of Clinical Microbiology*. American Society for Microbiology (ASM), 55(7), p. 1998. doi: 10.1128/JCM.02250-16.

Pereira, F. C. *et al.* (2013) 'The Spore Differentiation Pathway in the Enteric Pathogen Clostridium difficile', *PLoS Genetics*. Edited by P. H. Viollier. Public Library of Science, 9(10), p. e1003782. doi: 10.1371/journal.pgen.1003782.

Pereira, F. C. *et al.* (2018) 'A LysM domain intervenes in sequential protein-protein and protein-peptidoglycan interactions important for spore coat assembly in *Bacillus subtilis*', *Journal of Bacteriology*. doi: 10.1128/JB.00642-18.

Pettit, L. J. *et al.* (2014) 'Functional genomics reveals that *Clostridium difficile Spo0A* coordinates sporulation, virulence and metabolism', *BMC Genomics*. BioMed Central, 15(1),

p. 160. doi: 10.1186/1471-2164-15-160.

Pfeffer, J. *et al.* (2012) 'Control of Lytic Transglycosylase Activity Within Bacterial Cell Walls', in *Bacterial Glycomics: Current Research, Technology and Applications*, pp. 55–68. Available at: https://www.researchgate.net/profile/John_Pfeffer/publication/235964384_Control_of_Lytic_Transglycosylase_Activity_Within_Bacterial_Cell_Walls/links/0deec514dcb2b57c41000000/Control-of-Lytic-Transglycosylase-Activity-Within-Bacterial-Cell-Walls.pdf.

Pidgeon, S. E. *et al.* (2019) 'L,D-transpeptidase Specific Probe Reveals Spatial Activity of Peptidoglycan Crosslinking', *ACS Chemical Biology*, p. acschembio.9b00427. doi: 10.1021/acschembio.9b00427.

Piggot, P. J. and Hilbert, D. W. (2004) 'Sporulation of *Bacillus subtilis*', *Current Opinion in Microbiology*. Elsevier Current Trends, 7(6), pp. 579–586. doi: 10.1016/J.MIB.2004.10.001.

Pishdadian, K., Fimlaid, K. A. and Shen, A. (2015) 'SpoIIID-mediated regulation of σK function during *Clostridium difficile* sporulation', *Molecular Microbiology*, 95(2). doi: 10.1111/mmi.12856.

Pluskal, T. *et al.* (2010) 'MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data', *BMC Bioinformatics*. BioMed Central, 11(1), p. 395. doi: 10.1186/1471-2105-11-395.

Popham, D. *et al.* (1996) 'Analysis of the peptidoglycan structure of Bacillus subtilis endospores', *J. Bacteriol.*, 178(22), pp. 6451–6458. Available at: http://jb.asm.org/content/178/22/6451.abstract (Accessed: 26 February 2016).

Popham, D. L., Gilmore, M. E. and Setlow, P. (1999) 'Roles of low-molecular-weight penicillin-binding proteins in *Bacillus subtilis* spore peptidoglycan synthesis and spore properties.', *Journal of bacteriology*, 181(1), pp. 126–32. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=103540&tool=pmcentrez&rendertype=abstract (Accessed: 24 April 2016).

Potter, S. C. *et al.* (2018) 'HMMER web server: 2018 update', *Nucleic Acids Research*. Narnia, 46(W1), pp. W200–W204. doi: 10.1093/nar/gky448.

Potterton, L. *et al.* (2018) '*CCP* 4 *i* 2: the new graphical user interface to the *CCP* 4 program suite', *Acta Crystallographica Section D Structural Biology*. International Union of Crystallography, 74(2), pp. 68–84. doi: 10.1107/S2059798317016035.

Pratt, R. F. (2008) 'Substrate specificity of bacterial DD-peptidases (penicillin-binding proteins)', *Cellular and Molecular Life Sciences*, 65(14), pp. 2138–2155. doi: 10.1007/s00018-008-7591-7.

Pratviel-Sosa, F. *et al.* (1994) 'Effect of various analogues of D-glutamic acid on the D-glutamate-adding enzyme from *Escherichia coli.*', *FEMS microbiology letters*, 115(2–3), pp. 223–8. doi: 10.1111/j.1574-6968.1994.tb06642.x.

Psylinakis, E. *et al.* (2005) 'Peptidoglycan N-acetylglucosamine deacetylases from *Bacillus cereus*, highly conserved proteins in *Bacillus anthracis*.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 280(35), pp. 30856–63. doi: 10.1074/jbc.M407426200.

Public Health England (2013) 'Updated guidance on the management and treatment of

*Clostridium difficile* infection'. Available at: http://www.gov.uk/phe (Accessed: 12 May 2016).

Public Health England (2019) C. difficile *infection: monthly data by prior trust exposure - GOV.UK*. Available at: https://www.gov.uk/government/statistics/c-difficile-infection-monthly-data-by-prior-trust-exposure (Accessed: 7 August 2019).

Putnam, E. E. *et al.* (2013) 'SpoIVA and SipL are *Clostridium difficile* spore morphogenetic proteins.', *Journal of bacteriology*. American Society for Microbiology Journals, 195(6), pp. 1214–25. doi: 10.1128/JB.02181-12.

Quraishi, M. N. *et al.* (2017) 'Systematic review with meta-analysis: the efficacy of faecal microbiota transplantation for the treatment of recurrent and refractory *Clostridium difficile* infection', *Alimentary Pharmacology & Therapeutics*, 46(5), pp. 479–493. doi: 10.1111/apt.14201.

Ramamurthi, K. S., Clapham, K. R. and Losick, R. (2006) 'Peptide anchoring spore coat assembly to the outer forespore membrane in *Bacillus subtilis*', *Molecular Microbiology*. John Wiley & Sons, Ltd (10.1111), 62(6), pp. 1547–1557. doi: 10.1111/j.1365-2958.2006.05468.x.

Ransom, E. M. *et al.* (2018) 'Multiple factors contribute to bimodal toxin gene expression in *Clostridioides ( Clostridium) difficile*', *Molecular Microbiology*. John Wiley & Sons, Ltd (10.1111), 110(4), pp. 533–549. doi: 10.1111/mmi.14107.

Raymond, J. B. *et al.* (2005) 'Identification of the *namH* Gene, Encoding the Hydroxylase Responsible for the *N* -Glycolylation of the Mycobacterial Peptidoglycan', *Journal of Biological Chemistry*, 280(1), pp. 326–333. doi: 10.1074/jbc.M411006200.

Reddy, S. S. and Brandt, L. J. (2013) '*Clostridium difficile* Infection and Inflammatory Bowel Disease', *Journal of Clinical Gastroenterology*, 47(8), pp. 666–671. doi: 10.1097/MCG.0b013e31828b288a.

Ribis, J. W. *et al.* (2017) 'The Conserved Spore Coat Protein SpoVM Is Largely Dispensable in *Clostridium difficile* Spore Formation.', *mSphere*. American Society for Microbiology Journals, 2(5), pp. e00315-17. doi: 10.1128/mSphere.00315-17.

Ribis, J. W., Fimlaid, K. A. and Shen, A. (2018) 'Differential requirements for conserved peptidoglycan remodeling enzymes during *Clostridioides difficile* spore formation', *Molecular Microbiology*. John Wiley & Sons, Ltd (10.1111), 110(3), pp. 370–389. doi: 10.1111/mmi.14090.

Robert, X. and Gouet, P. (2014) 'Deciphering key features in protein structures with the new ENDscript server', *Nucleic Acids Research*. Narnia, 42(W1), pp. W320–W324. doi: 10.1093/nar/gku316.

Rohlfing, A. E. *et al.* (2019) 'The CspC pseudoprotease regulates germination of Clostridioides difficile spores in response to multiple environmental signals.', *PLoS genetics*. Edited by D. A. Garsin, 15(7), p. e1008224. doi: 10.1371/journal.pgen.1008224.

Rosenbusch, K. E. *et al.* (2012) 'C. difficile 630Δerm Spo0A Regulates Sporulation, but Does Not Contribute to Toxin Production, by Direct High-Affinity Binding to Target DNA', *PLoS ONE*. Edited by Y.-F. Chang. Public Library of Science, 7(10), p. e48608. doi: 10.1371/journal.pone.0048608.

Rupnik, M., Wilcox, M. H. and Gerding, D. N. (2009) '*Clostridium difficile* infection: new developments in epidemiology and pathogenesis', *Nature Reviews Microbiology*, 7(7), pp. 526–536. doi: 10.1038/nrmicro2164.

Sahin, E. and Roberts, C. J. (2012) 'Size-Exclusion Chromatography with Multi-angle Light Scattering for Elucidating Protein Aggregation Mechanisms', in *Methods in molecular biology (Clifton, N.J.)*, pp. 403–423. doi: 10.1007/978-1-61779-921-1_25.

Salazar, C. L. *et al.* (2017) 'Molecular, microbiological and clinical characterization of *Clostridium difficile* isolates from tertiary care hospitals in Colombia', *PLOS ONE*. Edited by G. Vedantam. Public Library of Science, 12(9), p. e0184689. doi: 10.1371/journal.pone.0184689.

Sambol, S. P. *et al.* (2002) 'Colonization for the Prevention of *Clostridium difficile* Disease in Hamsters', *The Journal of Infectious Diseases*. Narnia, 186(12), pp. 1781–1789. doi: 10.1086/345676.

Saujet, L. *et al.* (2013) 'Genome-Wide Analysis of Cell Type-Specific Gene Transcription during Spore Formation in *Clostridium difficile*', *PLoS Genetics*. Edited by P. H. Viollier. Public Library of Science, 9(10), p. e1003756. doi: 10.1371/journal.pgen.1003756.

Saujet, L. *et al.* (2014) 'The regulatory network controlling spore formation in *Clostridium difficile*.', *FEMS microbiology letters*, 358(1), pp. 1–10. doi: 10.1111/1574-6968.12540.

Sayedy, L., Kothari, D. and Richards, R. J. (2010) 'Toxic megacolon associated *Clostridium difficile* colitis.', *World journal of gastrointestinal endoscopy*. Baishideng Publishing Group Inc, 2(8), pp. 293–7. doi: 10.4253/wjge.v2.i8.293.

Sbahi, H. and Di Palma, J. A. (2016) 'Faecal microbiota transplantation: applications and limitations in treating gastrointestinal disorders', *BMJ Open Gastroenterology*. BMJ Specialist Journals, 3(1), p. e000087. doi: 10.1136/bmjgast-2016-000087.

Scheurwater, E., Reid, C. W. and Clarke, A. J. (2008) 'Lytic transglycosylases: Bacterial space-making autolysins', *The International Journal of Biochemistry & Cell Biology*. Pergamon, 40(4), pp. 586–591. doi: 10.1016/J.BIOCEL.2007.03.018.

Schleifer, K. H. and Kandler, O. (1972) 'Peptidoglycan types of bacterial cell walls and their taxonomic implications.', *Bacteriological reviews*, 36(4), pp. 407–77. Available at: http://www.ncbi.nlm.nih.gov/pubmed/4568761 (Accessed: 13 February 2017).

Schwan, C. *et al.* (2009) '*Clostridium difficile* Toxin CDT Induces Formation of Microtubule-Based Protrusions and Increases Adherence of Bacteria', *PLoS Pathogens*. Edited by S. R. Blanke. Public Library of Science, 5(10), p. e1000626. doi: 10.1371/journal.ppat.1000626.

Schwan, C. *et al.* (2014) '*Clostridium difficile* toxin CDT hijacks microtubule organization and reroutes vesicle traffic to increase pathogen adherence.', *Proceedings of the National Academy of Sciences of the United States of America*, 111(6), pp. 2313–8. doi: 10.1073/pnas.1311589111.

Serrano, M. *et al.* (2016) 'The SpoIIQ-SpoIIIAH complex of *Clostridium difficile* controls forespore engulfment and late stages of gene expression and spore morphogenesis.', *Molecular microbiology*, 100(1), pp. 204–28. doi: 10.1111/mmi.13311.

Setlow, P. (2003) 'Spore germination', *Current Opinion in Microbiology*. Elsevier Current Trends, 6(6), pp. 550–556. doi: 10.1016/J.MIB.2003.10.001.

Setlow, P. (2006) 'Spores of *Bacillus subtilis*: Their resistance to and killing by radiation, heat and chemicals', *Journal of Applied Microbiology*, 101(3), pp. 514–525. doi: 10.1111/j.1365-2672.2005.02736.x.

Setlow, P. (2007) 'I will survive: DNA protection in bacterial spores.', *Trends in microbiology*, 15(4), pp. 172–80. doi: 10.1016/j.tim.2007.02.004.

Setlow, P., Wang, S. and Li, Y.-Q. (2017) 'Germination of Spores of the Orders *Bacillales* and *Clostridiales*', *Annual Review of Microbiology*. Annual Reviews , 71(1), pp. 459–477. doi: 10.1146/annurev-micro-090816-093558.

Shah, I. M. *et al.* (2008) 'A Eukaryotic-like Ser/Thr Kinase Signals Bacteria to Exit Dormancy in Response to Peptidoglycan Fragments', *Cell*. Cell Press, 135(3), pp. 486–496. doi: 10.1016/J.CELL.2008.08.039.

Sham, L.-T. *et al.* (2014) 'MurJ is the flippase of lipid-linked precursors for peptidoglycan biogenesis', *Science*, 345(6193), pp. 220–222. doi: 10.1126/science.1254522.

Sheldon, E. *et al.* (2016) 'A phase 1, placebo-controlled, randomized study of the safety, tolerability, and immunogenicity of a *Clostridium difficile* vaccine administered with or without aluminum hydroxide in healthy adults.', *Vaccine*, 34(18), pp. 2082–2091. doi: 10.1016/j.vaccine.2016.03.010.

Shida, T. *et al.* (2001) 'Mutational analysis of catalytic sites of the cell wall lytic N-acetylmuramoyl-L-alanine amidases CwlC and CwlV.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 276(30), pp. 28140–6. doi: 10.1074/jbc.M103903200.

Smith, K., Bayer, M. E. and Youngman, P. (1993) 'Physical and functional characterization of the Bacillus subtilis spoIIM gene.', *Journal of bacteriology*. American Society for Microbiology (ASM), 175(11), pp. 3607–17.

Smits, W. K. *et al.* (2016) '*Clostridium difficile* infection', *Nature Reviews Disease Primers*. Nature Publishing Group, 2, p. 16020. doi: 10.1038/nrdp.2016.20.

Song, J. H. and Kim, Y. S. (2019) 'Recurrent *Clostridium difficile* Infection: Risk Factors, Treatment, and Prevention', *Gut and Liver*. Editorial Office of Gut and Liver, 13(1), pp. 16–24. doi: 10.5009/gnl18071.

Sorg, J. A. and Sonenshein, A. L. (2009) 'Chenodeoxycholate is an inhibitor of *Clostridium difficile* spore germination.', *Journal of bacteriology*. American Society for Microbiology Journals, 191(3), pp. 1115–7. doi: 10.1128/JB.01260-08.

Sorg, J. A. and Sonenshein, A. L. (2010) 'Inhibiting the initiation of *Clostridium difficile* spore germination using analogs of chenodeoxycholic acid, a bile acid.', *Journal of bacteriology*. American Society for Microbiology Journals, 192(19), pp. 4983–90. doi: 10.1128/JB.00610-10.

Sreerama, N. and Woody, R. W. (2000) 'Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Comparison of CONTIN, SELCON, and CDSSTR Methods with an Expanded Reference Set', *Analytical Biochemistry*, 287(2), pp. 252–260. doi: 10.1006/abio.2000.4880.

Srikhanta, Y. N. *et al.* (2019) 'Cephamycins inhibit pathogen sporulation and effectively treat recurrent Clostridioides difficile infection', *Nature Microbiology*. Nature Publishing Group,

pp. 1–9. doi: 10.1038/s41564-019-0519-1.

Sun, X. *et al.* (2010) 'The Enterotoxicity of *Clostridium difficile* Toxins', *Toxins*. Molecular Diversity Preservation International, 2(7), pp. 1848–1880. doi: 10.3390/toxins2071848.

Sütterlin, L. *et al.* (2017) 'Peptidoglycan Cross-Linking Activity of L,D-Transpeptidases from *Clostridium difficile* and Inactivation of These Enzymes by β-Lactams', *Antimicrobial Agents and Chemotherapy*, 62(1). doi: 10.1128/AAC.01607-17.

Szklarczyk, D. *et al.* (2019) 'STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets', *Nucleic Acids Research*, 47(D1), pp. D607–D613. doi: 10.1093/nar/gky1131.

Szwedziak, P. and Löwe, J. (2013) 'Do the divisome and elongasome share a common evolutionary past?', *Current Opinion in Microbiology*. Elsevier Current Trends, 16(6), pp. 745–751. doi: 10.1016/J.MIB.2013.09.003.

Tang-Feldman, Y. *et al.* (2003) 'Molecular analysis of *Clostridium difficile* strains isolated from 18 cases of recurrent clostridium difficile-associated diarrhea.', *Journal of clinical microbiology*, 41(7), pp. 3413–4. doi: 10.1128/jcm.41.7.3413-3414.2003.

Tayeh, M. A. *et al.* (1995) 'Overproduction and One-Step Purification of *Escherichia coli* UDP-N-Acetylglucosamine Enolpyruvyl Reductase', *Protein Expression and Purification*, 6(6), pp. 757–762. doi: 10.1006/prep.1995.0006.

Taylor, E. J. *et al.* (2006) 'Structure and Activity of Two Metal Ion-dependent Acetylxylan Esterases Involved in Plant Cell Wall Degradation Reveals a Close Similarity to Peptidoglycan Deacetylases', *Journal of Biological Chemistry*, 281(16), pp. 10968–10975. doi: 10.1074/jbc.M513066200.

Tocheva, E. I. *et al.* (2013) 'Peptidoglycan transformations during *Bacillus subtilis* sporulation', *Molecular Microbiology*, 88(4), pp. 673–686. doi: 10.1111/mmi.12201.

Underwood, S. *et al.* (2009) 'Characterization of the sporulation initiation pathway of *Clostridium difficile*and its role in toxin production.', *Journal of bacteriology*. American Society for Microbiology (ASM), 191(23), pp. 7296–305. doi: 10.1128/JB.00882-09.

Urzhumtseva, L. *et al.* (2009) 'Crystallographic model quality at a glance.', *Acta crystallographica. Section D, Biological crystallography*. International Union of Crystallography, 65(Pt 3), pp. 297–300. doi: 10.1107/S0907444908044296.

Vagin, A. A. *et al.* (2004) '*REFMAC* 5 dictionary: organization of prior chemical knowledge and guidelines for its use', *Acta Crystallographica Section D Biological Crystallography*, 60(12), pp. 2184–2195. doi: 10.1107/S0907444904023510.

Vermassen, A. *et al.* (2019) 'Cell Wall Hydrolases in Bacteria: Insight on the Diversity of Cell Wall Amidases, Glycosidases and Peptidases Toward Peptidoglycan.', *Frontiers in microbiology*. Frontiers Media SA, 10, p. 331. doi: 10.3389/fmicb.2019.00331.

Vijayrajratnam, S. *et al.* (2016) 'Bacterial peptidoglycan with amidated meso-diaminopimelic acid evades NOD1 recognition: an insight into NOD1 structure-recognition', *Biochemical Journal*, 473(24), pp. 4573–4592. doi: 10.1042/BCJ20160817.

Vincent, C. and Manges, A. R. (2015) 'Antimicrobial Use, Human Gut Microbiota and *Clostridium difficile* Colonization and Infection.', *Antibiotics (Basel, Switzerland)*.

Multidisciplinary Digital Publishing Institute (MDPI), 4(3), pp. 230–53. doi: 10.3390/antibiotics4030230.

Vollmer, W. *et al.* (2008) 'Bacterial peptidoglycan (murein) hydrolases', *FEMS Microbiology Reviews*, 32(2), pp. 259–286. doi: 10.1111/j.1574-6976.2007.00099.x.

Vollmer, W. (2008) 'Structural variation in the glycan strands of bacterial peptidoglycan', *FEMS Microbiology Reviews*, 32(2), pp. 287–306. doi: 10.1111/j.1574-6976.2007.00088.x.

Vollmer, W., Blanot, D. and de Pedro, M. A. (2008) 'Peptidoglycan structure and architecture.', *FEMS microbiology reviews*. The Oxford University Press, 32(2), pp. 149–67. doi: 10.1111/j.1574-6976.2007.00094.x.

Vollmer, W. and Tomasz, A. (2000) 'The pgdA Gene Encodes for a peptidoglycan N-Acetylglucosamine Deacetylase in *Streptococcus pneumoniae*', *The Journal of biological chemistry*. in Press, 275(27), pp. 20496–20501. doi: 10.1074/jbc.M910189199.

Wang, G. *et al.* (2010) 'Peptidoglycan deacetylation in Helicobacter pylori contributes to bacterial survival by mitigating host immune responses.', *Infection and immunity*. American Society for Microbiology Journals, 78(11), pp. 4660–6. doi: 10.1128/IAI.00307-10.

Wang, K. H. *et al.* (2009) 'The coat morphogenetic protein SpoVID is necessary for spore encasement in *Bacillus subtilis*', *Molecular Microbiology*, 74(3), pp. 634–649. doi: 10.1111/j.1365-2958.2009.06886.x.

Wang, S. T. *et al.* (2006) 'The Forespore Line of Gene Expression in *Bacillus subtilis*', *Journal of Molecular Biology*, 358(1), pp. 16–37. doi: 10.1016/j.jmb.2006.01.059.

Warth, A. D. and Strominger, J. L. (1972) 'Structure of the peptidoglycan from spores of *Bacillus subtilis*', *Biochemistry*. American Chemical Society, 11(8), pp. 1389–1396. doi: 10.1021/bi00758a010.

Waterman, M. S. and Eggert, M. (1987) 'A new algorithm for best subsequence alignments with application to tRNA-rRNA comparisons.', *Journal of molecular biology*, 197(4), pp. 723–8. doi: 10.1016/0022-2836(87)90478-5.

Waxman, D. J. and Strominger, J. L. (1983) 'Penicillin-Binding Proteins and the Mechanism of Action of Beta-Lactam Antibiotics1', *Annual Review of Biochemistry*, 52(1), pp. 825–869. doi: 10.1146/annurev.bi.52.070183.004141.

Whitmore, L. and Wallace, B. A. (2004) 'DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data', *Nucleic Acids Research*, 32(WEB SERVER ISS.). doi: 10.1093/nar/gkh371.

Whitmore, L. and Wallace, B. A. (2008) 'Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases', *Biopolymers*, 89(5), pp. 392–400. doi: 10.1002/bip.20853.

Yan, Y. *et al.* (2000) 'Crystal structure of *Escherichia coli* UDPMurNAc-tripeptide d-alanyl-d-alanine-adding enzyme (MurF) at 2.3 A resolution.', *Journal of molecular biology*, 304(3), pp. 435–45. doi: 10.1006/jmbi.2000.4215.

Yang, J. *et al.* (2015) 'The I-TASSER Suite: protein structure and function prediction', *Nature Methods*, 12(1), pp. 7–8. doi: 10.1038/nmeth.3213.

Zapun, A., Contreras-Martel, C. and Vernet, T. (2008) 'Penicillin-binding proteins and β-lactam resistance', *FEMS Microbiology Reviews*. Narnia, 32(2), pp. 361–385. doi: 10.1111/j.1574-6976.2007.00095.x.

Zhang, Y. and Mathys, A. (2019) 'Superdormant spores as a hurdle for gentle germination-inactivation based spore control strategies', *Frontiers in Microbiology*. Frontiers Media SA, 10(JAN), p. 3163. doi: 10.3389/fmicb.2018.03163.

Zhou, R., Chen, S. and Recsei, P. (1988) 'A dye release assay for determination of lysostaphin activity', *Analytical Biochemistry*. Academic Press, 171(1), pp. 141–144. doi: 10.1016/0003-2697(88)90134-0.

Zhu, D., Sorg, J. A. and Sun, X. (2018) '*Clostridioides difficile* Biology: Sporulation, Germination, and Corresponding Therapies for *C. difficile* Infection', *Frontiers in Cellular and Infection Microbiology*. Frontiers, 8, p. 29. doi: 10.3389/fcimb.2018.00029.

# *Appendix A*

Table A-1: Custom MZmine2 database

| ID | m/z | Retention time (min) | Identity | Formula |
|---|---|---|---|---|
| 7 | 851.3578 | 33 | GlcNAc MurNAc anhydro AEm 850 | C34H54N6O19 |
| 8 | 871.3837 | 18.26 | GlcNAc MurNAc Aem 870 | C34H59N7O19 |
| 10 | 916.4064 | 17.14 | GlcN MurNAc AEmS 915 | C35H61N7O21 |
| 11 | 886.3948 | 18.6 | GlcN MurNAc AEmG 885 | C34H59N7O20 |
| 12 | 886.4306 | 21.02 | GlcN MurNAc AEmG 885 | C34H59N7O20 |
| 13 | 886.3948 | 18.6 | GlcN MurNAc AEmG885 | C34H59N7O20 |
| 14 | 907.3656 | 18.75 | GlcN MurNAc AEmG Na adduct 885 | C34H59N7O20 |
| 15 | 984.3948 | 18.6 | GlcN MurNAc AEmG phosph adduct 885 | C34H59N7O20 |
| 16 | 454.69 | 18.6 | GlcN MurNAc AEmG Na adduct 885 | C34H59N7O20 |
| 17 | 462.67 | 18.6 | GlcN MurNAc AEmG K adduct 885 | C34H59N7O20 |
| 18 | 908.3861 | 35.17 | GlcNAc MurNAcanhydro AEmG 907 | C36H57N7O20 |
| 19 | 725.3215 | 18.67 | MurNAc AEmG 724 | C28H48N6O16 |
| 20 | 514.2453 | 19.96 | GlcNAc MurNAc AEmR CD only 1026 | C40H70N10O21 |
| 21 | 643.305 | 21.29 | GlcN MurNAc AD 642 | C24H42N4O16 |
| 23 | 900.4016 | 22.16 | GlcN MurNAc AEmA 899 | C35H61N7O20 |
| 24 | 900.4005 | 22.23 | GlcN MurNAc AEmA 899 | C35H61N7O20 |
| 25 | 900.4003 | 24.51 | GlcN MurNAc AEmA 899 | C35H61N7O20 |
| 26 | 450.7063 | 24.78 | GlcN MurNAc AEmA 2+ 899 | C35H61N7O20 |
| 27 | 998.382 | 24.66 | GlcN MurNAc AEmA 2H3PO4+ 899 | C35H61N7O20 |
| 28 | 922.3918 | 21.94 | GlcN MurNAc AEmA 2Na+ 899 | C35H61N7O20 |
| 29 | 469.6812 | 24.66 | GlcN MurNAc AEmA 2K+ 899 | C35H61N7O20 |
| 30 | 471.7107 | 23.45 | GlcN MurNAc AEmA 2H3PO4 2K+ 899 | C35H61N7O20 |
| 31 | 482.7036 | 23.45 | GlcNAc MurNAc AemA 2Na 941 | C37H36N7O21 |

| ID | m/z | Retention time (min) | Identity | Formula |
|---|---|---|---|---|
| 32 | 490.6873 | 23.65 | GlcNAc MurNAc AemA 2K 941 | C37H36N7O21 |
| 33 | 739.3388 | 21.94 | MurNAc AEmA 738 | C29H50N6O16 |
| 34 | 739.3395 | 24.66 | MurNAc AEmA 738 | C29H50N6O16 |
| 35 | 914.428 | 26.49 | GlcN MurNAc AEmB 913 | C36H63N7O20 |
| 36 | 476.6896 | 26.64 | GlcN MurNAc AEmB k2+ 913 | C36H63N7O20 |
| 37 | 928.4447 | 32.74 | GlcN MurNAc AEm927 | C37H65N7O20 |
| 38 | 942.4661 | 42.21 | GlcN MurNAc AEmI/L 941 | C38H67N7O20 |
| 39 | 942.4632 | 42.35 | GlcN MurNAc AEmI/L 941 | C38H67N7O20 |
| 40 | 490.71 | 42.15 | GlcN MurNAc AEmI/L K2+ 941 | C38H67N7O20 |
| 41 | 976.4471 | 49.93 | GlcN MurNAc AEmF 975 | C14H65N7O20 |
| 44 | 815.3733 | 50.15 | MurNAc AEmF CD based termini 814 | C35H54N6O16 |
| 45 | 820.3662 | 30.18 | GlcN MurNAc AEm AEm GlcN MurNAc 3-3 2+ 1638 | C64H110N12O37 |
| 47 | 839.3592 | 30.26 | GlcN MurNAc AEm AEm GlcN MurNAc 3-3 2xH3PO4+ 1638 | C64H110N12O37 |
| 48 | 869.36 | 30.18 | GlcN MurNAc AEm AEm GlcN MurNAc 3-3 H3PO42+ 1638 | C64H110N12O37 |
| 49 | 839.33 | 30.26 | GlcN MurNAc AEm AEm GlcN MurNAc 3-3 2K+ 1638 | C64H110N12O37 |
| 50 | 599.8934 | 30.18 | GlcN MurNAc AEm AEm GlcN MurNAc 3-3 3K+ 1638 | C64H110N12O37 |
| 51 | 739.8326 | 30.26 | GlcN MurNAc AEm AEm MurNAc 1477 | C58H99N11O33 |
| 52 | 841.3748 | 31.67 | GlcNAc MurNAc AEm AEm GlcN MurNAc 1680 | C66H112N12O38 |
| 54 | 928.4078 | 20.05 | GlcNAc MurNAc AEmG 927 | C36H61N7O21 |
| 55 | 629.7791 | 22.68 | GlcN MurNAc AEmG Aem 3-3 1257 | C49H83N11O27 |
| 56 | 848.8812 | 28.96 | GlcN MurNAc AEmG AEm GlcN MurNAc 3-3 1695 | C66H113N13O38 |
| 57 | 848.878 | 28.96 | GlcN MurNAc AEmG AEm GlcN MurNAc 3-3 isotope 1695 | C66H113N13O38 |
| 58 | 848.8799 | 31.01 | GlcN MurNAc AEmG AEm GlcN MurNAc 3-3 1695 | C66H113N13O38 |
| 59 | 848.9003 | 33.08 | GlcN MurNAc AEmG AEm GlcN MurNAc 3-3 1695 | C66H113N13O38 |

| ID | m/z | Retention time (min) | Identity | Formula |
|---|---|---|---|---|
| 60 | 848.9011 | 33.17 | GlcN MurNAc AEmG AEm GlcN MurNAc 3-3 1695 | C66H113N13O38 |
| 62 | 897.8682 | 29.15 | GlcN MurNAc AEmG AEm GlcN MurNAc 3-3 2H3PO4+ 1695 | C66H113N13O38 |
| 63 | 946.8576 | 29.15 | GlcN MurNAc AEmG AEm GlcN MurNAc 3-3 2H3PO4+ 1695 | C66H113N13O38 |
| 64 | 946.8574 | 29.23 | GlcN MurNAc AEmG AEm GlcN MurNAc 3-3 2H3PO4+ 1695 | C66H113N13O38 |
| 65 | 860.391 | 31.38 | GlcN MurNAc AEmG AEm GlcN MurNAc 3-3 2Na+ 1695 | C66H113N13O38 |
| 66 | 869.8853 | 30.81 | GlcNAc MurNAc AEmG AEm GlcN MurNAc 3-3 1737 | C68H115N13O39 |
| 67 | 869.9047 | 34.94 | GlcNAc MurNAc AEmG AEm GlcN MurNAc 3-3 1737 | C68H115N13O39 |
| 68 | 650.7891 | 24.27 | GlcNAc MurNAc AEmG Aem 3-3 1299 | C51H85N11O28 |
| 70 | 870.8896 | 30.89 | GlcNAc MurNAc AEmG AEm GlcN MurNAc 3-3 isotope 1737 | C68H115N13O39 |
| 71 | 855.8871 | 31.77 | GlcN MurNAc AEmA AEm MurNAc GlcN 3-3 1709 | C67H115N13O38 |
| 72 | 855.8863 | 32.02 | GlcN MurNAc AEmA AEm MurNAc GlcN 3-3 1709 | C67H115N13O38 |
| 73 | 855.8898 | 33.17 | GlcN MurNAc AEmA AEm MurNAc GlcN 3-3 1709 | C67H115N13O38 |
| 74 | 855.8869 | 3445 | GlcN MurNAc AEmA AEm MurNAc GlcN 3-3 1709 | C67H115N13O38 |
| 75 | 855.8869 | 34.45 | GlcN MurNAc AEmA AEm MurNAc GlcN 3-3 1709 | C67H115N13O38 |
| 76 | 855.8882 | 35.88 | GlcN MurNAc AEmA AEm MurNAc GlcN 3-3 1709 | C67H115N13O38 |
| 77 | 855.8875 | 35.98 | GlcN MurNAc AEmA AEm MurNAc GlcN 3-3 1709 | C67H115N13O38 |
| 79 | 916.3721 | 32.11 | GlcN MurNAc AEmA AEm MurNAc GlcN 3-3 2Na 2H3PO4 1709 | C67H115N13O38 |
| 80 | 866.8743 | 32.25 | GlcN MurNAc AEmA AEm MurNAc GlcN 3-3 2Na 1709 | C67H115N13O38 |
| 81 | 866.8774 | 32.17 | GlcN MurNAc AEmA AEm MurNAc GlcN 3-3 2Na 1709 | C67H115N13O38 |
| 82 | 866.886 | 34.61 | GlcN MurNAc AEmA AEm MurNAc GlcN 3-3 2Na 1709 | C67H115N13O38 |
| 83 | 953.8677 | 32.11 | GlcN MurNAc AEmA AEm MurNAc GlcN 3-3 2H3PO4x2 1709 | C67H115N13O38 |
| 84 | 875.3648 | 32.45 | GlcN MurNAc AEmA AEm MurNAc GlcN 3-3 2K isotopes 1709 | C67H115N13O38 |
| 85 | 583.5688 | 32.11 | GlcN MurNAc AEmA AEm MurNAc GlcN 3-3 3K+ 1709 | C67H115N13O38 |

| ID | m/z | Retention time (min) | Identity | Formula |
|---|---|---|---|---|
| 86 | 583.9047 | 34.45 | GlcN MurNAc AEmA AEm MurNAc GlcN 3-3 3K+ 1709 | C67H115N13O38 |
| 87 | 775.3529 | 32.74 | GlcN MurNAc AEmA AEm MurNAc 3-3 1548 | C61H104N12O34 |
| 88 | 876.8952 | 33.96 | GlcNAc MurNAc AEmA AEm MurNAc GlcN 3-3 2+ 1571 | C69H117N13O39 |
| 89 | 876.8945 | 34.05 | GlcNAc MurNAc AEmA AEm MurNAc GlcN 3-3 2+ 1751 | C69H117N13O39 |
| 91 | 876.8976 | 36.19 | GlcNAc MurNAc AEmA AEm MurNAc GlcN 3-3 2+ 1751 | C69H117N13O39 |
| 92 | 876.8951 | 38.16 | GlcNAc MurNAc AEmA AEm MurNAc GlcN 3-3 2+ 1751 | C69H117N13O39 |
| 93 | 887.8832 | 34.19 | GlcNAc MurNAc AEmA AEm MurNAc GlcN 3-3 2Na+ 1751 | C69H117N13O39 |
| 94 | 887.8806 | 36.65 | GlcNAc MurNAc AEmA AEm MurNAc GlcN 3-3 2Na+ 1751 | C69H117N13O39 |
| 95 | 867.3929 | 49.11 | GlcNAc MurNAc GlcN MurNAcanhydro AEmA AEm 4-3 CD only 1731 | C68H109N13O39 |
| 96 | 868.8809 | 49.79 | GlcNAc MurNAc AEmA AEm GlcN MurNAc anhydro 3-3 Some evi 1731 | C69H113N13O38 |
| 97 | 866.8782 | 50.15 | GlcNAc MurNAc AEmA AEm GlcN MurNAc anhydro 3-3 Some evi 1731 | C69H113N13O38 |
| 98 | 866.8813 | 53.56 | GlcNAc MurNAc AEmA AEm GlcN MurNAc anhydro 3-3 Some evi 1731 | C69H113N13O38 |
| 99 | 866.8809 | 53.75 | GlcNAc MurNAc AEmA AEm GlcN MurNAc anhydro 3-3 Some evi 1731 | C69H113N13O38 |
| 100 | 453.7062 | 25 | lactyl AEmA EmA 3-3 905 | C36H59N9O18 |
| 101 | 683.7904 | 33.63 | GlcNAc MurNAc AEmA AEmAlactyl 3-3 2Na 1343 | C53H89N11O29 |
| 102 | 672.805 | 33.36 | GlcNAc MurNAc AEmA AEmAlactyl 3-3 2+ 1343 | C53H89N11O29 |
| 103 | 647.7823 | 25.43 | GlcN MurNAc AEmA AEm 3-3 1271 | C50H85N11O27 |
| 104 | 696.7698 | 25.43 | GlcN MurNAc AEmA AEm 3-3 2Na 2H3PO4 1271 | C50H85N11O27 |
| 105 | 636.7873 | 29.92 | GlcN MurNAc AEmA AEm 3-3 2Na 1271 | C50H85N11O27 |
| 106 | 734.7682 | 25.43 | GlcN MurNAc AEmA AEm 3-3 2H3PO4 1271 | C50H85N11O27 |
| 107 | 708.3232 | 35.52 | GlcN MurNAc AEmA lactylAEm unkwn Xlnk 1414 | C56H94N12O30 |
| 108 | 708.3214 | 37.62 | GlcN MurNAc AEmA lactylAEm unkwn Xlnk 1414 | C56H94N12O30 |
| 109 | 891.4082 | 33.96 | 2 GlcN MurNAc AEmA AEmA 4-3 1780 | C70H120N14O39 |

| ID | m/z | Retention time (min) | Identity | Formula |
|---|---|---|---|---|
| 110 | 891.4068 | 35.6 | 2 GlcN MurNAc AEmA AEmA 4-3 1780 | C70H120N14O39 |
| 111 | 940.3981 | 33.96 | 2 GlcN MurNAc AEmA AEmA 4-3 H3PO42+ 1780 | C70H120N14O39 |
| 112 | 940.3963 | 35.78 | 2 GlcN MurNAc AEmA AEmA 4-3 H3PO42+ 1780 | C70H120N14O39 |
| 113 | 607.2519 | 34.05 | 2 GlcN MurNAc AEmA AEmA 4-3 K3+ 1780 | C70H120N14O39 |
| 115 | 951.3901 | 34.05 | 2 GlcN MurNAc AEmA AEmA 4-3 Na H3PO42+ 1780 | C70H120N14O39 |
| 116 | 890.4208 | 30.47 | 2 GlcN MurNAc AEmA AEmA 4-3 1 MurNac unred 1778 | C70H118N14O39 |
| 117 | 939.413 | 30.55 | 2 GlcN MurNAc AEmA AEmA 4-3 1 MurNac unred H3PO4 1778 | C70H118N14O39 |
| 118 | 911.9313 | 32.53 | GlcNAc MurNAc AEmA GlcN MurNAc AEmA 1 MurNAc unred Unkwn Xlink CD mostly 1820 | C72H120N14O40 |
| 119 | 912.4122 | 35.78 | GlcNAc MurNAc AEmA GlcN MurNAc AEmA 4-3 1822 | C72H122N14O40 |
| 121 | 913.9188 | 38.08 | GlcNAc MurNAc AEmA GlcN MurNAc AEmA 4-3 1822 | C72H122N14O40 |
| 122 | 923.92 | 38.16 | GlcNAc MurNAc AEmA AEmA GlcN MurNAc 2Na+ obs and exp -1Da 1822 | C72H112N14O40 |
| 123 | 884.3935 | 31.01 | GlcN MurNAc AEmG AEmA GlcN MurNAc 3-3 2+ 1766 | C69H118N14O39 |
| 124 | 884.3942 | 31.11 | GlcN MurNAc AEmG AEmA GlcN MurNAc 3-3 2+ 1766 | C69H118N14O39 |
| 125 | 589.9319 | 31.01 | GlcN MurNAc AEmG AEmA GlcN MurNAc 3-3 3+ 1766 | C69H118N14O39 |
| 126 | 589.9316 | 31.11 | GlcN MurNAc AEmG AEmA GlcN MurNAc 3-3 3+ 1766 | C69H118N14O39 |
| 127 | 933.3864 | 31.11 | GlcN MurNAc AEmG AEmA GlcN MurNAc 3-3 H3PO4 2+ 1766 | C69H118N14O39 |
| 128 | 978.4423 | 35.03 | LactylAEmA LactylAEmA 3-3 see notes977 | C39H64N10O19 |
| 129 | 489.2268 | 27.2 | LactylAEm AEmA 4-3 see notes 976 | C39H64N10O19 |
| 130 | 869.9054 | 43.77 | GlcN MurNac Aem AEmV GlcN MurNAc 3-3 1737 | C69H129N13O38 |
| 131 | 919.3948 | 43.85 | GlcNAc MurNAc AEmA AEmB GlcN MurNAc 3-3 speculative 1836 | C73H124N14O40 |
| 132 | 888.8782 | 44.03 | GlcNAc MurNac Aem AEmV GlcN MurNAc 3-3 2K+ 1737 | C69H119N13O38 |
| 133 | 891.4171 | 47.28 | GlnNAc MurNAc AEM AEMP/V GlcN MurNAc 1779 | C71H121N13O39 |

| ID | m/z | Retention time (min) | Identity | Formula |
|---|---|---|---|---|
| 134 | 890.9128 | 47.36 | GlnNAc MurNAc AEM AEMP/V GlcN MurNAc 1779 | C71H121N13O39 |
| 135 | 905.9291 | 46.64 | GlcN MurNAc AEmA AEmV GlcN MurNAc 3-3 1808 | C72H124N14O39 |
| 136 | 905.9285 | 50.15 | GlcN MurNAc AEmA AEmV GlcN MurNAc 3-3 1808 | C72N124N14O39 |
| 137 | 926.4342 | 50.65 | GlcNAC MurNAc AEmA AEmV GlcN MurNAc 4-3 1850 | C74H126N14O40 |
| 138 | 686.8207 | 48.12 | GlcN MurNAc AEMV lactylAEM 3-3 CD 1371 | C55H93N11O29 |
| 139 | 898.9222 | 27.54 | GlcN MurNAc AEmR AEm GlcN MurNAc 3-3 1794 | C70H122N16O38 |
| 140 | 898.4171 | 40.85 | GlcN MurNAc AEmR AEm GlcN MurNAc 3-3 1794 | C70H122N16O38 |
| 141 | 898.9189 | 38.79 | GlcN MurNAc AEmR AEm GlcN MurNAc 3-3 1794 | C70H122N16O38 |
| 142 | 876.9145 | 53.33 | GlcN MurNAc AEmI/L GlcN MurNAc 3-3 1751 | C70H122N13O38 |
| 143 | 884.4099 | 26.28 | GlcN MurNAc AEmE AEm GlcN MurNAc 3-3 exp and obsv -1Da 1767 | C69H117N13O40 |
| 144 | 897.9207 | 57.54 | GlcNAc MUrNAc AEmI/L GlcN MurNAc 3-3 1793 | C72H123N13O39 |
| 145 | 877.3898 | 27.77 | GlcN MurNAc AEmI/L GlcN MurNAc 3-3 1752 | C67H115N15O39 |
| 146 | 863.8843 | 28.5 | GlcN MurNAc AEmS AEm GlcN MurNAc 3-3 1725 | C67H115N15O39 |
| 147 | 883.8996 | 39.21 | GlcNAc MurNAc(nonred) AEmS AEm GlcN MurNAc 3-3 1765 | C69H115N13O40 |
| 148 | 707.8405 | 31.59 | GlcN MurNAc AQmA AQmlactylA 3-3 mostly CD inferred 1412 | C56H96N14O28 |
| 149 | 902.407 | 45.26 | GlcN MurNAc AEmY AEm GlcN MurNAc 3-3 1801 | C73H119N13O39 |
| 150 | 1254.066 | 35.31 | GlcN MurNAc x3 AEm AEm AEmG ambiguous Xlinks 3+ 2506 | C98H167N19O56 |
| 151 | 849.358 | 35.31 | GlcN MurNAc x3 AEm AEm AEmG ambiguous Xlinks 3K+ 2506 | C98H167N19O56 |
| 152 | 849.3624 | 35.44 | GlcN MurNAc x3 AEm AEm AEmG ambiguous Xlinks 3K+ 2506 | C98H167N19O56 |
| 153 | 893.0408 | 36.52 | GlcN MurNAc x3 AEm AEmA AEmG 3 H3PO4+ 2577 | C101H172N20O57 |
| 154 | 1261.574 | 38.41 | GlcN MurNAc x3 AEm AEm AEmA 3+ 2520 | C99H169N19O56 |
| 155 | 1261.575 | 38.51 | GlcN MurNAc x3 AEm AEm AEmA 3+ 2520 | C99H169N19O56 |
| 156 | 854.3674 | 38.62 | GlcN MurNAc x3 AEm AEm AEmA 3K+ 2520 | C99H169N19O56 |

| ID | m/z | Retention time (min) | Identity | Formula |
|---|---|---|---|---|
| 157 | 848.7076 | 38.41 | GlcN MurNAc x3 AEm AEm AEmA 3Na+ 2520 | C99H169N19O56 |
| 158 | 848.707 | 38.51 | GlcN MurNAc x3 AEm AEm AEmA 3Na+ 2520 | C99H169N19O56 |
| 159 | 886.6891 | 38.62 | GlcN MurNAc x3 AEm AEm AEmA 3K 3H3PO4+ 2520 | C99H169N19O56 |
| 160 | 906.37 | 38.62 | GlcN MurNAc x3 AEm AEm AEmA 2XH3PO4+ 2520 | C99H169N19O56 |
| 161 | 874.0353 | 40.22 | GlcN MurNAc x3 AEm AEm AEmA 3H3PO4+ 2520 | C99H169N19O56 |
| 162 | 854.0323 | 38.7 | GlcN MurNAc x3 AEm AEm AEmA 3K+ 2520 | C99H169N19O56 |
| 163 | 1282.579 | 41.02 | GlcNAc 2GlcN AEmA 2AEm 2562 | C101H171N19O57 |
| 164 | 1297.099 | 40.22 | 3GlcN 3MurNAc 2AEmA AEmA 2591 | C102H174N20O57 |
| 165 | 877.7065 | 39.85 | 3GlcN 3MurNAc 2AEmA AEmA 3K+ 2591 | C102H174N20O57 |
| 166 | 877.7065 | 39.96 | 3GlcN 3MurNAc 2AEmA AEmA 3K+ 2591 | C102H174N20O57 |
| 167 | 872.387 | 39.85 | 3GlcN 3MurNAc 2AEmA AEmA 3K+ 2591 | C102H174N20O57 |
| 168 | 930.3663 | 40.22 | 3GlcN 3MurNAc 2AEmA AEmA 3H3PO4x2+ 2591 | C102H174N20O57 |
| 169 | 901.3875 | 41.7 | 3GlcN 3MurNAc 3AEmA 2662 3K+ | C105H179N21O58 |
| 170 | 902.7454 | 44.03 | GlcNAc 2GlcN 3MurNAc 3AEmA 2704 | C107H181N21O59 |
| 171 | 1098.99 | 44.34 | GlcNAc GlcN 2MurNAc lactylAEm AEmA AEm 2194 | C87H145N17O48 |
| 172 | 1070.975 | 37.2 | 2GlcN 2MurNAc lactylAEm AEm AEmG 2139 | C84H141N17O47 |
| 175 | 708.3217 | 37.7 | GlcN MurNAc AEmA AEmAlactyl 1414 | C56H94N12O30 |
| 178 | 869.9037 | 43.85 | GlcN MurNAc AEmV Aem GlcN MurNAc 3-3 1737 | C69H119N13O38 |

All the muropeptides manually identified in the analysis of *C. difficile 630* Δ*pyrE* Δ*erm* $P_{tet}$-*spo0A*, their adducts and alternative charge states were collated, along with their mass:charge ratio (*m/z*), retention time (RT), deduced identity (identity) and empirical formula (formula). The identity field also contains a rough estimate of the neutral mass, and in some cases, notes on the muropeptide identification. The fields, and their arrangement, are as dictated by the MZmine2 software. GlcN; glucosamine, GlcNAc; N-acetylglucosamine, MurNAc; N-acetylmuramic acid, A; Alanine, E, Glutamic acid, m; mesoDAP, G; glycine, S; serine, R; arginine, B; γ-aminobutyric acid, I/L; Isoleucine/ leucine, V; valine,P; proline, Y; tyrosine, F; phenylalanine.

Table A-2: Area under the curve of 10 muropeptides of interest - raw data

| | Area under curve (total ion counts) | | | | | | |
|---|---|---|---|---|---|---|---|
| | WT | ΔE | ΔF | ΔD | ΔP | ΔQ | ΔAH |
| Total | 4.3E+07 | 1.1E+07 | 5.4E+07 | 1.5E+07 | 2.8E+07 | 6.8E+07 | 3.9E+07 |
| MP 1 | 2.4E+06 | 4.3E+04 | 1.0E+06 | 0.0E+00 | 9.8E+05 | 2.1E+06 | 3.7E+06 |
| MP 2 | 2.8E+06 | 2.4E+06 | 3.1E+07 | 6.6E+06 | 3.4E+06 | 4.8E+07 | 5.8E+06 |
| MP 3 | 7.7E+05 | 0.0E+00 | 0.0E+00 | 0.0E+00 | 3.1E+05 | 0.0E+00 | 4.7E+05 |
| MP 4 | 4.6E+06 | 1.8E+06 | 4.4E+06 | 1.7E+06 | 5.3E+06 | 1.2E+06 | 4.9E+06 |
| MP 5 | 2.1E+06 | 2.2E+05 | 1.2E+05 | 1.5E+05 | 8.9E+05 | 1.5E+05 | 1.8E+06 |
| MP 6 | 9.4E+05 | 3.0E+05 | 2.4E+06 | 3.2E+05 | 5.8E+05 | 0.0E+00 | 5.6E+05 |
| MP 7 | 2.3E+07 | 3.7E+06 | 7.4E+06 | 4.2E+06 | 1.1E+07 | 6.9E+06 | 1.0E+07 |
| MP 8 | 1.3E+06 | 1.8E+06 | 3.8E+06 | 8.3E+05 | 2.0E+06 | 4.8E+06 | 1.9E+06 |
| MP 9 | 4.5E+06 | 6.0E+05 | 3.6E+06 | 1.4E+06 | 2.6E+06 | 4.0E+06 | 9.3E+06 |
| MP 10 | 1.7E+05 | 0.0E+00 | 1.8E+05 | 0.0E+00 | 1.6E+05 | 0.0E+00 | 2.1E+05 |

The 10 muropeptides of interest were extracted from the MZmine2 output for each strain with the raw values provided above. MP; muropeptide, WT*; C. difficile* 630 Δ*pyrE* Δ*erm* $P_{tet}$ *-spo0A,* ΔE*; C. difficile* 630 Δ*erm* Δ*sigE, i; C. difficile* 630 Δ*erm* Δ*sigF,* ΔD*; C. difficile* 630 Δ*erm* Δ*spoIID,* ΔP*; C. difficile* 630 Δ*erm* Δ*spoIIP,* ΔQ*; C. difficile* 630 Δ*erm* Δ*spoIIQ and* ΔAH *C. difficile* 630 Δ*erm* Δ*spoIIIAH* MP numbers refer to those identified in the wild type peptidoglycan analysis Table 3-1.

**Table A-3:** Area under the curve of 10 muropeptides of interest - normalised data

| | Normalised area under curve (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | WT | ΔE | ΔF | ΔD | ΔP | ΔQ | ΔAH |
| Sporulation efficiency (%) | 0.0E+00 | 79 | 35 | 28 | 28 | 22 | 47 |
| Peak area | | | | | | | |
| Total area | 4.3E+07 | 8.6E+06 | 1.9E+07 | 4.2E+06 | 7.7E+06 | 1.5E+07 | 1.8E+07 |
| Peak 1 area | 2.4E+06 | 3.4E+04 | 3.6E+05 | 0.0E+00 | 2.8E+05 | 4.5E+05 | 1.7E+06 |
| Peak 2 area | 2.8E+06 | 1.9E+06 | 1.1E+07 | 1.8E+06 | 9.6E+05 | 1.1E+07 | 2.7E+06 |
| Peak 3 area | 7.7E+05 | 0.0E+00 | 0.0E+00 | 0.0E+00 | 8.7E+04 | 0.0E+00 | 2.2E+05 |
| Peak 4 area | 4.6E+06 | 1.4E+06 | 1.6E+06 | 4.7E+05 | 1.5E+06 | 2.7E+05 | 2.3E+06 |
| Peak 5 area | 2.1E+06 | 1.8E+05 | 4.1E+04 | 4.1E+04 | 2.5E+05 | 3.3E+04 | 8.6E+05 |
| Peak 6 area | 9.4E+05 | 2.4E+05 | 8.6E+05 | 9.0E+04 | 1.6E+05 | 0.0E+00 | 2.6E+05 |
| Peak 7 area | 2.3E+07 | 2.9E+06 | 2.6E+06 | 1.2E+06 | 3.2E+06 | 1.5E+06 | 4.9E+06 |
| Peak 8 area | 1.3E+06 | 1.4E+06 | 1.3E+06 | 2.3E+05 | 5.7E+05 | 1.1E+06 | 9.1E+05 |
| Peak 9 area | 4.5E+06 | 4.8E+05 | 1.2E+06 | 3.8E+05 | 7.4E+05 | 8.9E+05 | 4.4E+06 |
| Peak 10 area | 1.7E+05 | 0.0E+00 | 6.2E+04 | 0.0E+00 | 4.5E+04 | 0.0E+00 | 1.0E+05 |

The 10 muropeptides of interest were extracted from the MZmine2 output for each strain with the normalised values provided above. Data were normalised to the sporulation efficiency of each strain. MP; muropeptide, *WT; C. difficile* 630 Δ*pyrE* Δ*erm* $P_{tet}$ *-spo0A,* ΔE*; C. difficile* 630 Δ*erm* Δ*sigE, i; C. difficile* 630 Δ*erm* Δ*sigF,* ΔD*; C. difficile* 630 Δ*erm* Δ*spoIID,* ΔP*; C. difficile* 630 Δ*erm* Δ*spoIIP,* ΔQ*; C. difficile* 630 Δ*erm* Δ*spoIIQ and* ΔAH *C. difficile* 630 Δ*erm* Δ*spoIIIAH* MP numbers refer to those identified in the wild type peptidoglycan analysis Table 3-1. Sporulation efficiencies were calculated by Dr Marcin Dembek.

**Table A-4:** Normalised percentage area under the curve accounted for by each muropeptide in each strain

| MP | Normalised percentage area under curve (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | WT | ΔE | ΔF | ΔD | ΔP | ΔQ | ΔAH |
| MP 1 | 5.6 | 0.4 | 1.9 | 0.0 | 3.6 | 3.0 | 9.5 |
| MP 2 | 6.6 | 22.0 | 57.8 | 43.5 | 12.4 | 71.6 | 14.7 |
| MP 3 | 1.8 | 0.0 | 0.0 | 0.0 | 1.1 | 0.0 | 1.2 |
| MP 4 | 10.7 | 16.3 | 8.2 | 11.0 | 19.3 | 1.8 | 12.5 |
| MP 5 | 5.0 | 2.1 | 0.2 | 1.0 | 3.2 | 0.2 | 4.7 |
| MP 6 | 2.2 | 2.8 | 4.5 | 2.1 | 2.1 | 0.0 | 1.4 |
| MP 7 | 53.9 | 34.1 | 13.6 | 27.9 | 40.8 | 10.2 | 26.7 |
| MP 8 | 3.1 | 16.8 | 7.0 | 5.5 | 7.3 | 7.1 | 5.0 |
| MP 9 | 10.7 | 5.6 | 6.6 | 9.0 | 9.6 | 6.0 | 23.8 |
| MP 10 | 0.4 | 0.0 | 0.3 | 0.0 | 0.6 | 0.0 | 0.5 |

Sporulation efficiency- normalised areas under the curve are expressed as a percentage of the total normalised area under the curve of all 10 muropeptides of interest. MP; muropeptide, *WT; C. difficile* 630 Δ*pyrE* Δ*erm* $P_{tet}$ *-spo0A,* ΔE*; C. difficile* 630 Δ*erm* Δ*sigE, i; C. difficile* 630 Δ*erm* Δ*sigF,* ΔD*; C. difficile* 630 Δ*erm* Δ*spoIID,* ΔP*; C. difficile* 630 Δ*erm* Δ*spoIIP,* ΔQ*; C. difficile* 630 Δ*erm* Δ*spoIIQ and* ΔAH *C. difficile* 630 Δ*erm* Δ*spoIIIAH* MP numbers refer to those identified in the wild type peptidoglycan analysis Table 3-1.

**Table A-5:** Difference between area under curve attributed to a given muropeptide in a given mutant cell peptidoglycan compared to the wild type

| MP | Change to normalised percentage area under curve | | | | | |
|---|---|---|---|---|---|---|
| | ΔE | ΔF | ΔD | ΔP | ΔQ | ΔAH |
| MP 1 | -5.1 | -3.7 | -5.6 | -2.0 | -2.5 | 3.9 |
| MP 2 | 15.4 | 51.2 | 36.9 | 5.8 | 65.0 | 8.1 |
| MP 3 | -1.8 | -1.8 | -1.8 | -0.7 | -1.8 | -0.6 |
| MP 4 | 5.6 | -2.5 | 0.3 | 8.6 | -8.9 | 1.8 |
| MP 5 | -2.9 | -4.8 | -4.0 | -1.8 | -4.8 | -0.3 |
| MP 6 | 0.6 | 2.3 | -0.1 | -0.1 | -2.2 | -0.8 |
| MP 7 | -19.8 | -40.3 | -25.9 | -13.1 | -43.6 | -27.2 |
| MP 8 | 13.6 | 3.8 | 2.3 | 4.2 | 3.9 | 1.8 |
| MP 9 | -5.1 | -4.1 | -1.7 | -1.1 | -4.7 | 13.2 |
| MP 10 | -0.4 | -0.1 | -0.4 | 0.2 | -0.4 | 0.1 |

The difference in the normalised percentage area under the curve for a given muropeptide in a mutant cell peptidoglycan compared to that seen in the wild type peptidoglycan. MP; muropeptide, WT*; C. difficile* 630 Δ*pyrE* Δ*erm* $P_{tet}$ *-spo0A,* ΔE*; C. difficile* 630 Δ*erm* Δ*sigE, i; C. difficile* 630 Δ*erm* Δ*sigF,* ΔD*; C. difficile* 630 Δ*erm* Δ*spoIID,* ΔP*; C. difficile* 630 Δ*erm* Δ*spoIIP,* ΔQ*; C. difficile* 630 Δ*erm* Δ*spoIIQ and* ΔAH *C. difficile* 630 Δ*erm* Δ*spoIIIAH* MP numbers refer to those identified in the wild type peptidoglycan analysis Table 3-1.

**Table A-6**: Comparison of crosslinking between mutant cell peptidoglycans and the wild type

| Strain | Total normalised percentage area under curve (%) | |
|---|---|---|
| | Monomers | Crosslinked |
| WT | 12.2 | 87.8 |
| ΔE | 22.4 | 77.6 |
| ΔF | 59.7 | 40.3 |
| ΔD | 43.5 | 56.5 |
| ΔP | 15.9 | 84.1 |
| ΔQ | 74.7 | 25.3 |
| ΔAH | 24.2 | 75.8 |

Normalised percentage areas under the curve were summed for monomers and crosslinked muropeptides. Muropeptides 1 and 2 were considered monomers, with muropeptides 3-10 considered crosslinked. MP; muropeptide, WT*; C. difficile* 630 Δ*pyrE* Δ*erm* $P_{tet}$ *-spo0A,* ΔE*; C. difficile* 630 Δ*erm* Δ*sigE, i; C. difficile* 630 Δ*erm* Δ*sigF,* ΔD*; C. difficile* 630 Δ*erm* Δ*spoIID,* ΔP*; C. difficile* 630 Δ*erm* Δ*spoIIP,* ΔQ*; C. difficile* 630 Δ*erm* Δ*spoIIQ and* ΔAH *C. difficile* 630 Δ*erm* Δ*spoIIIAH* MP numbers refer to those identified in the wild type peptidoglycan analysis Table 3-1.