4 *Clostridium difficile spo0A* is a global regulator of virulence and transmission genes

4.1 Introduction

The *C. difficile spo0A* gene codes for a highly conserved transcriptional regulator of sporulation that is required, as shown in Chapter 3, for relapsing disease and transmission in mice. Defining the *spo0A* regulon using transcriptional and proteomic approaches should improve our understanding of the *C. difficile* sporulation programme at the whole genome level and could identify persistence and transmission factors controlled by Spo0A in *C. difficile*.

Furthermore, the profound morphological and physiological changes that occur during the course of spore formation in *C. difficile* will be accompanied by a shift in gene expression and proteome profiles. As such, the ability to investigate the global behaviour of the *C. difficile* transcriptome and proteome in a quantitative manner during sporulation is essential if we are to identify gene products, regulators and regulons associated with this phenomenon.

4.1.1 Transcriptomics

The genome sequence of *C. difficile* 630 was completed at the WTSI (http://www.sanger.ac.uk/Projects/C_difficile/) and was available to support this analysis. Additionally, the Illumina sequencing platform has been adapted to enable high-throughput screening of bacterial transcriptomes. One method, termed ssRNA-Seq, is a particularly

powerful and reproducible tool to quantitatively survey global transcript abundance at a single strand level (223).

4.1.2 Proteomics

Understanding the Spo0A regulon will clearly require a multi-layered understanding of the interactions between functional gene products, including proteins, and the mechanisms by which their expression is regulated. Proteomic analysis is particularly important since any phenotype observed should be a direct consequence of the activity of a cells protein composition. Recently, the liquid chromatography-tandem mass spectrometry (LC-MS/MS) platform at the WTSI has been adapted to allow accurate and comparative quantification of protein expression between biological samples.

4.1.3 An integrated approach to studying spore formation

Transcriptomic and proteomic approaches are mutually complementary (224) and an integrative approach enabling the parallel profiling of both RNA transcripts and proteins is clearly an advantage to gaining a comprehensive and integrated overview of the Spo0A regulon. Such a combined approach can be used to derive a holistic picture of *C. difficile* biology and physiology that could not be readily achieved using each approach independently

(224, 225). Taking advantage of these methods, we sought to apply mRNA and protein profiling to define the *C. difficile* Spo0A regulon.

4.2 Aims of the work described in this chapter

In this chapter, we aimed to provide a molecular description of the *C. difficile* Spo0A regulon at the whole genome level. Cataloguing transcripts and proteins should facilitate our understanding of the role of *C. difficile* Spo0A in disease and may lead to the identification of persistence and transmission factors controlled by Spo0A. Additionally, this approach will enable us to identify any pleiotropic effects of the *spo0A* mutation beyond sporulation and link these back to any role in disease and transmission.

4.3 Results

4.3.1 *C. difficile spo0A* is expressed during exponential growth

We previously demonstrated that Spo0A is essential for spore formation in *C. difficile* 630Δ*erm* (Chapter 3). To determine the effect of *spo0A* on *C. difficile* growth kinetics, we generated defined insertional mutations in the *spo0A* gene of *C. difficile* 630Δ*erm*, generating 630Δ*erm*Δ*spo0A*, and then performed growth curves in Wilson's broth. We found that *C.*

difficile 630Δ*erm* and 630Δ*erm*Δ*spo0A* displayed comparable growth kinetics, however the *spo0A* mutant failed to produce spores or spore-like elements as determined by anaerobic culturing after ethanol shock (spores but not vegetative cells of *C. difficile* are resistant to ethanol), which is in agreement with previously published data (Figure 4.1) (177).

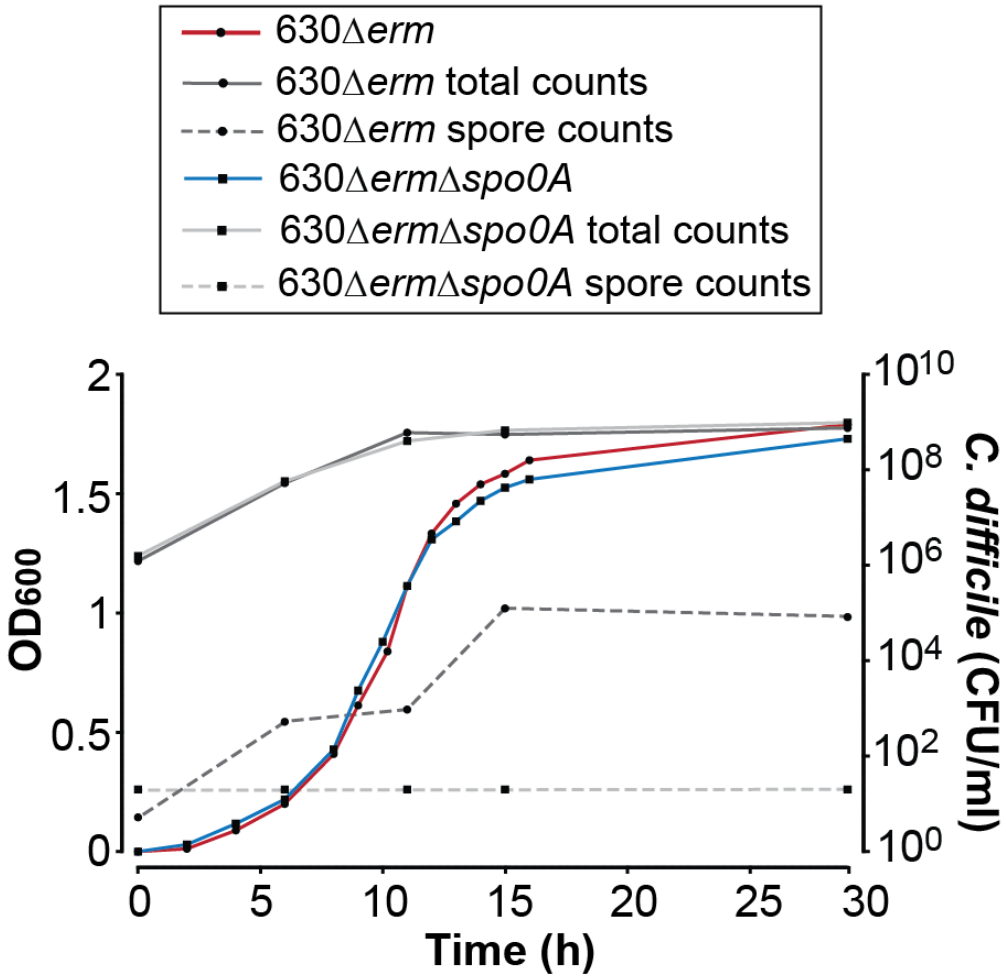

Figure 4.1. Growth and sporulation properties of *C. difficile* 630Δ*erm* derivatives. *C. difficile* cultures were grown in Wilson's broth under anaerobic conditions. Samples were taken at regular intervals throughout growth, during which the optical density, total cell and spore counts were determined. Optical density (OD_{600}) is plotted on the left *Y*-axis; bacterial counts (CFU/ml) are plotted on the right *Y*-axis.

Using *C. difficile* 630Δ*erm* cultures, we found that spores were produced during exponential phase, most notably at around 12 h of growth. This was surprising since spore formation is generally considered to be a stationary-phase phenomenon (associated with starvation or stress) and is somewhat conflicting with other models of bacterial sporulation (176, 177, 216). To demonstrate that the production of spores during exponential phase was directly attributable to the production of Spo0A, we performed a Western blot using *C. difficile* Spo0A-specific antibodies. We found that Spo0A was produced during both exponential and stationary growth in *C. difficile* 630Δ*erm* but was not produced by *C. difficile* 630Δ*erm*Δ*spo0A* at any time point (Figure 4.2). Given our interest in defining the *C. difficile spo0A* regulon and identifying genes involved in spore formation and sporulation, we chose to focus on exponentially growing (sporulating) *C. difficile*.

4.3.2 *C. difficile spo0A* is a global regulator of gene expression

Several studies in *Bacillus* have demonstrated that Spo0A is a global transcriptional regulator (184, 185, 226). In contrast, the role of Spo0A in *C. difficile* gene expression has not been well defined at the molecular level. For the first time, using high-density, strand-specific cDNA sequencing (ssRNA-seq), we surveyed gene expression profiles of exponentially growing *C. difficile* 630Δ*erm* and 630Δ*erm*Δ*spo0A* at the whole-genome level in order to identify global changes in gene expression profiles and to identify genes, regulators and regulons associated with this phenomenon.

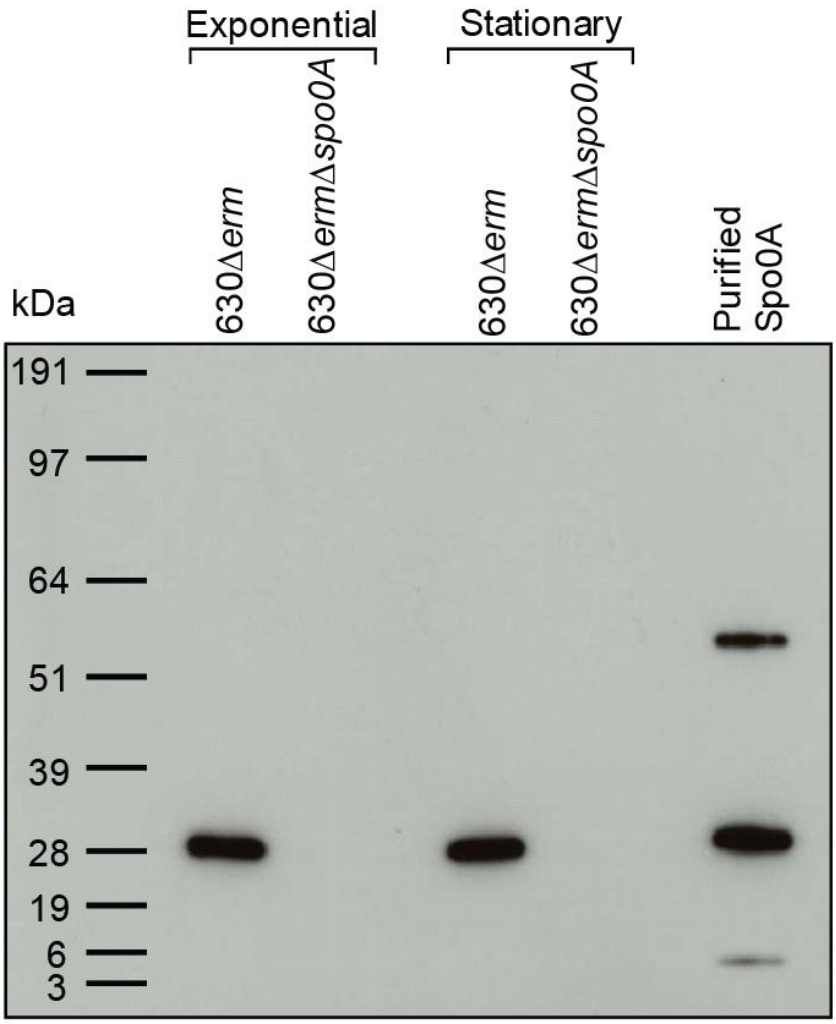

Figure 4.2. Western blot analysis of Spo0A expression in *C. difficile* 630Δ*erm* derivatives. *C. difficile* cultures were grown in Wilson's broth under anaerobic conditions, and samples were taken during mid-exponential and stationary growth. Extracted proteins with an equivalency to ~10^7 cells/lane were probed with anti-*C. difficile* Spo0A-specific antibodies. Purified Spo0A (~30 kDa) was included for reference. Lower and higher molecular weight bands in this lane likely represent degradation and Spo0A dimerisation, respectively.

4.3.2.1 Genome-wide identification of genes significantly regulated by Spo0A

We performed ssRNA-Seq on three biological replicates of both *C. difficile* 630Δ*erm* and *C. difficile* 630Δ*erm*Δ*spo0A*. However, since the total number of Ilumina read counts were not identical between the samples due to varying sequencing depths, the data were normalised using the DESeq package, as described in Materials and methods chapter 2.2.7.3. This reduces the systematic technical variation and thus enables a direct comparison between Spo0A mutant and parental *C. difficile* gene expression levels.

In order to identify and quantify genes that are controlled by Spo0A, normalised data were transformed on a variance stabilised scale resulting in homoscedastic data. This was additionally $\log_2$ transformed and filtered for a P value of ≤ 0.01. This stringent cutoff was chosen to reduce the likelihood of false positives and to generate a computationally manageable set of differentially expressed genes. Thus, it is likely that identified genes actually represent an underestimate of the true scale of Spo0A regulation in *C. difficile*.

Based on these parameters, we found that 261 genes were significantly differentially expressed in 630Δ*erm*Δ*spo0A* compared to the parental strain, of which 126 and 135 were up- and down-regulated, respectively (see Appendices 2 and 3) (Figure 4.3). We then mapped the relative gene expression values of all CDSs to the published *C. difficile* 630 genome, and overlaid the differentially expressed genes, as shown in Figure 4.4. In addition to mapping the differentially expressed genes, we also mapped an expanded list of 990 spore-associated polypeptides onto the annotated *C. difficile* 630 genome (Figure 4.4) in order to identify any

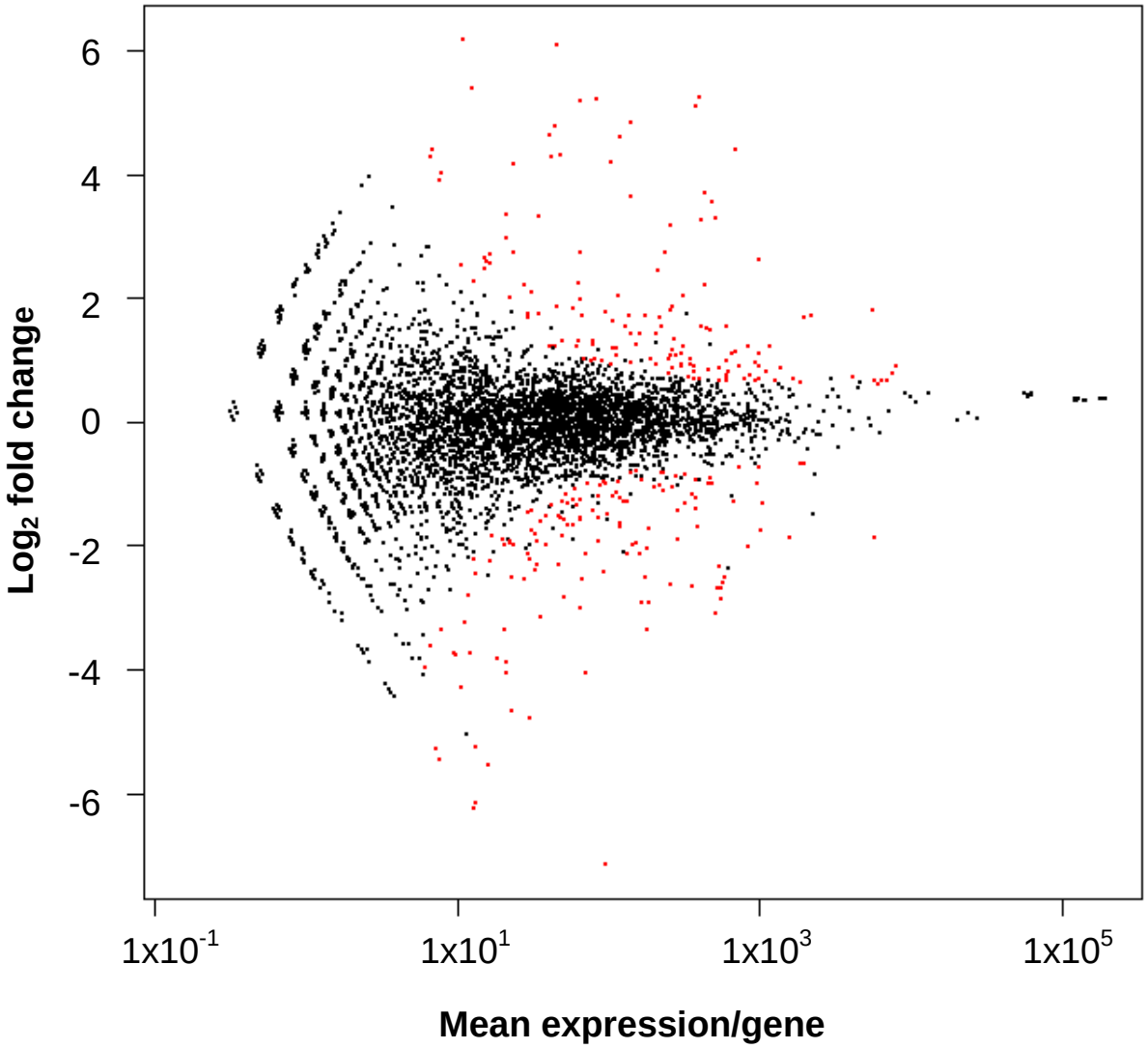

Figure 4.3. Identification of differentially expressed genes in *C. difficile* 630Δ*erm*Δ*spo0A* by transcriptional profiling. Scatter plot of the $\log_2$ fold changes against the normalised mean read abundance per gene (calculated at the base level). Red dots represent genes considered to be significantly differentially expressed using a *P* value of $P = \leq 0.01$. Black dots signify genes not deemed to be significantly differentially expressed according to these criteria.

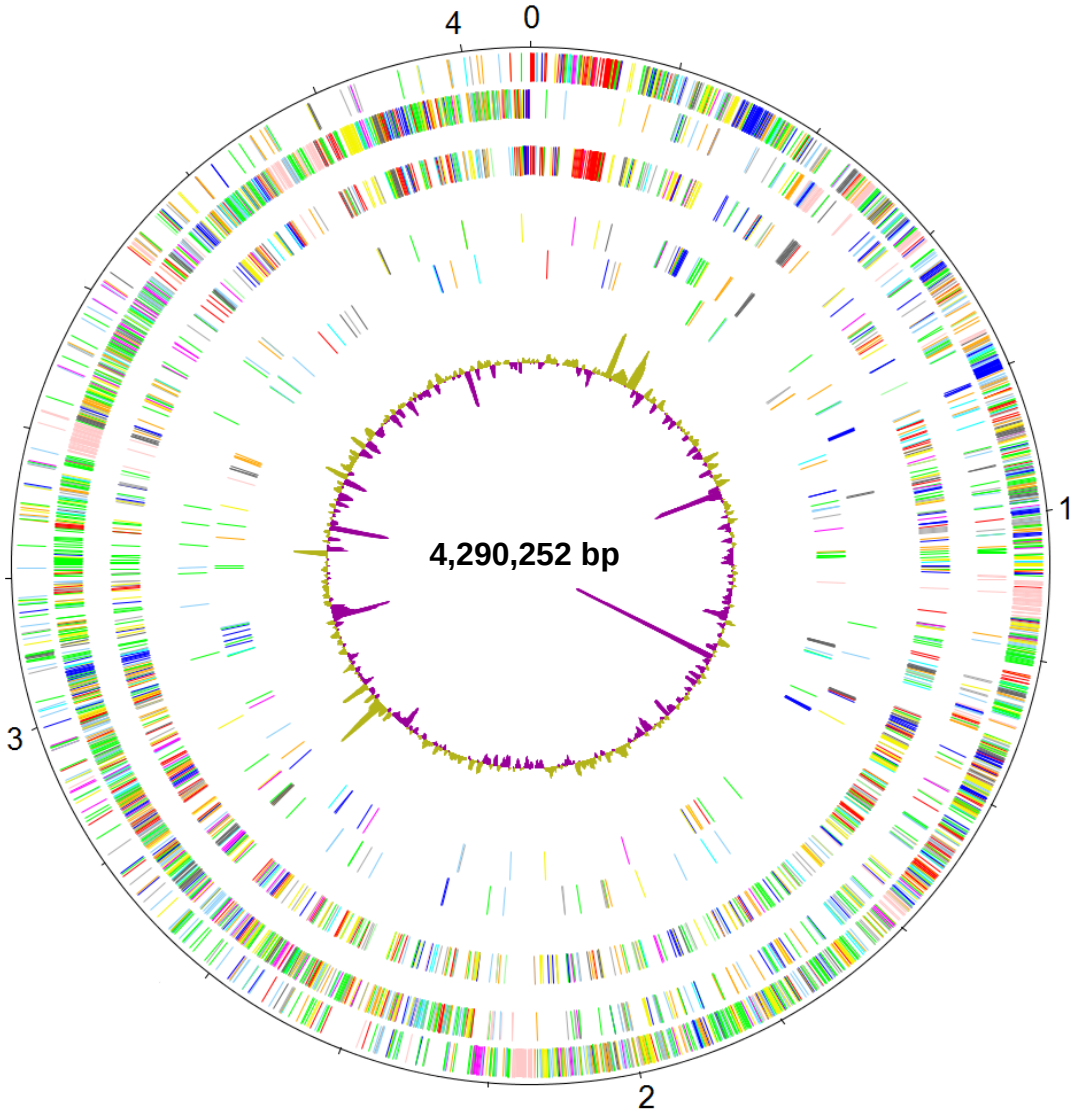

Figure 4.4. Global overview of the *C. difficile* 630Δ*erm* Spo0A-regulated transcriptome. The scale is marked in megabases (0-4) on the outermost circle. The concentric circles represent the following (from the outside in): 1+2, all CDS transcribed on the forward (clockwise) and reverse (counterclockwise) strands respectively; 3+4, CDS that are up and down regulated (P = 0.01) in 630Δ*erm*Δ*spo0A* compared to parental 630Δ*erm*, respectively; 5, plot of the $\log_2$ fold changes averaged over a 10-kbp window, in which green and purple shading corresponds to up and down regulation in 630Δ*erm*Δ*spo0A* compared to parental 630Δ*erm*, respectively. CDS are coloured according to predicted function as follows: dark blue, pathogenicity/adaptation; dark grey, energy metabolism; red, information transfer; dark green, surface-associated; cyan, degradation of large molecules; magenta, degradation of small molecules; yellow, central/intermediary metabolism; pale green, unknown; pale blue, regulators; orange, conserved hypothetical; brown, pseudogenes; pink, phage and insertion sequence (IS) elements; grey, miscellaneous.

genes that are critical for sporulation and/or germination. Genes regulated by Spo0A (either positively or negatively) were evenly distributed around the genome and were present on both the forward and reverse strands. A heat map of the top 40 most differentially expressed genes according to $\log_2$ fold-changes is shown in Figure 4.5.

In order to identify any global effects of the *spo0A* mutation, differentially expressed genes were assigned to a functional class, as designated in the annotated genome. These were then expressed as a percentage of the entire genome, which revealed a number of interesting trends, as shown in Figure 4.6, and discussed below. For example, of the 63 *in silico* predicted CDSs in the sporulation / germination class in the entire *C. difficile* 630 genome, almost a quarter (~22%) were found to be positively regulated by Spo0A, that is, their expression was attenuated in 630Δ*erm*Δ*spo0A* Figure 4.6A. None were found to be negatively regulated by Spo0A (Figure 4.6A). However, it is important to note that these data are based on the current annotation of the *C. difficile* 630 genome, which was published six years ago.

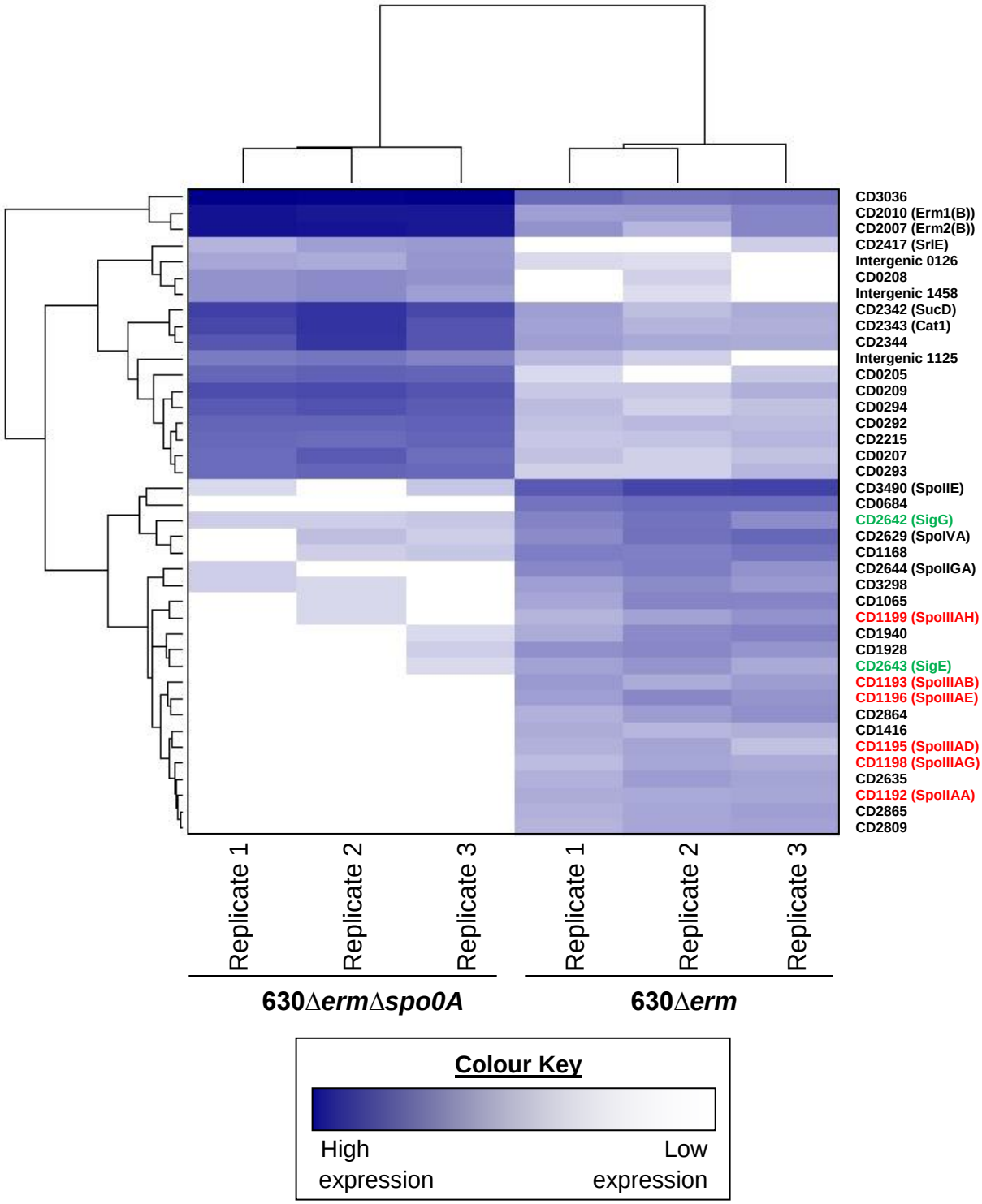

Figure 4.5. Heat map illustrating the expression data of the top 40 differentially expressed genes. Raw count data were normalised to enable a direct comparison and quantification of differentially expressed genes. Red, *spoIIIA* operon; green, *sigG* (CD2642) and *sigE* (CD2643) sigma factors.

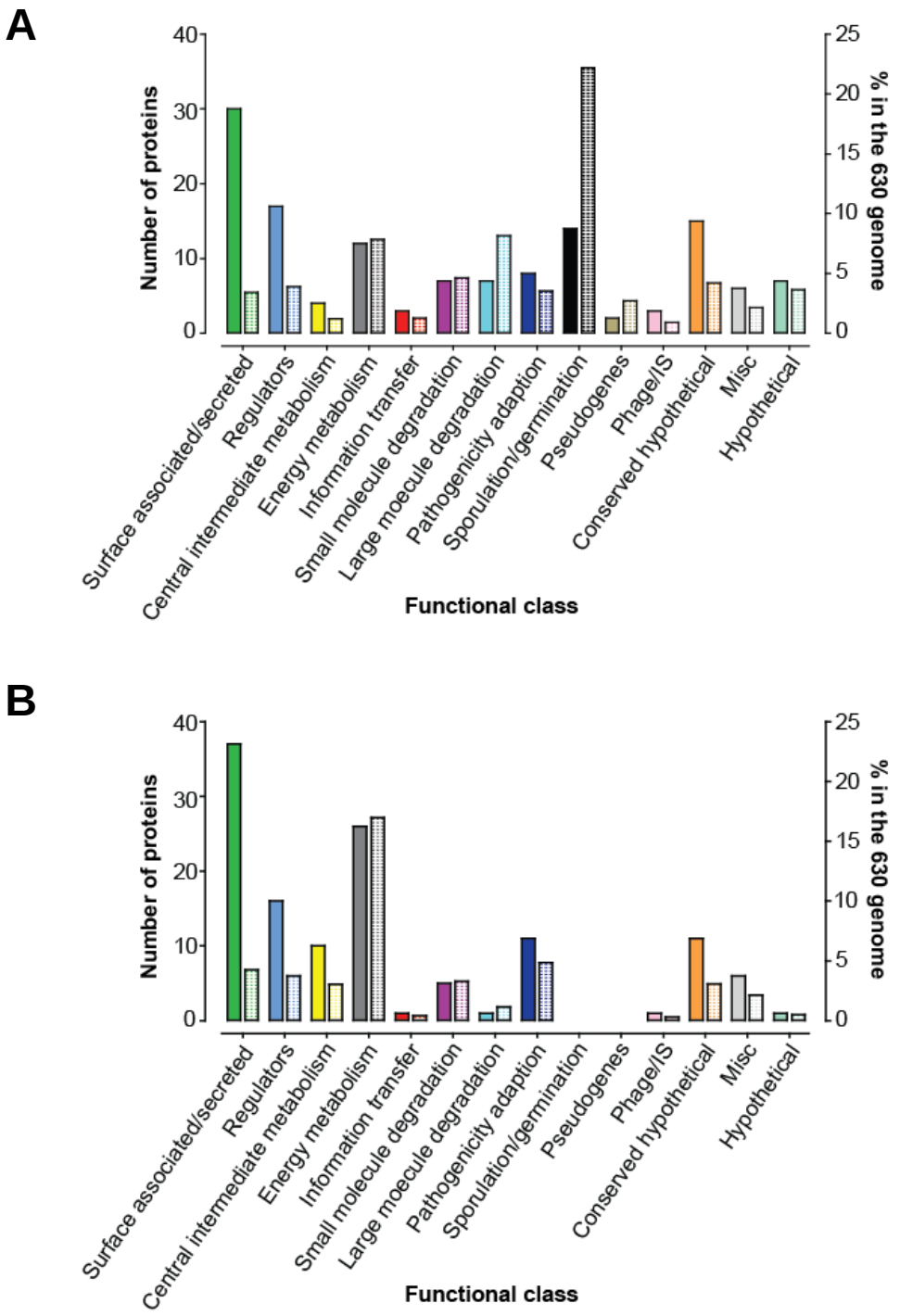

Figure 4.6. Distribution and relative abundance of *C. difficile* 630Δ*erm* Spo0A-regulated transcripts, according to functional classification. **A)** Transcripts down-regulated in 630Δ*erm*Δ*spo0A*, and **B)** transcripts up-regulated in 630Δ*erm*Δ*spo0A*. Transcripts were assigned to functional classes based on current *C. difficile* 630 annotation (10). Indicated are the absolute numbers of transcripts (filled bars, left *Y*-axis) and the percentage of the functional class relative to the total 630 genome (hashed bars, right *Y*-axis).

4.3.2.2 Confirmation of toxin production and sporulation phenotypes by ssRNA-Seq

We noted that Spo0A regulates the transcription of genes with significant phenotypes, including genes for TcdA production and spore formation, which are discussed in detail in Chapter 3.

4.3.2.2.1 Spo0A is an indirect negative regulator of TcdA production

tcdA expression was significantly up-regulated in 630Δ*erm*Δ*spo0A* at the mRNA level (Figure 4.7). This was anticipated given the increased levels of TcdA produced *in vitro* by both *C. difficile* 630Δ*erm*Δ*spo0A* and *C. difficile* R20291Δ*spo0A* (discussed in Chapter 3). However, we failed to identify an '0A' box upstream of the *tcdA* start site, implying that regulation could be via indirect control. In addition, *in vitro* binding assays did not indicate that *tcdA* was a direct target of Spo0A (Wiep Klass, personal communication) and thus some *tcdA* regulators may be encoded by genes which are under direct Spo0A control. For example, Saujet *et al*. (2011) demonstrated that the deletion of *sigH* markedly reduced *spo0A* expression and concomitantly increased TcdA expression (182).

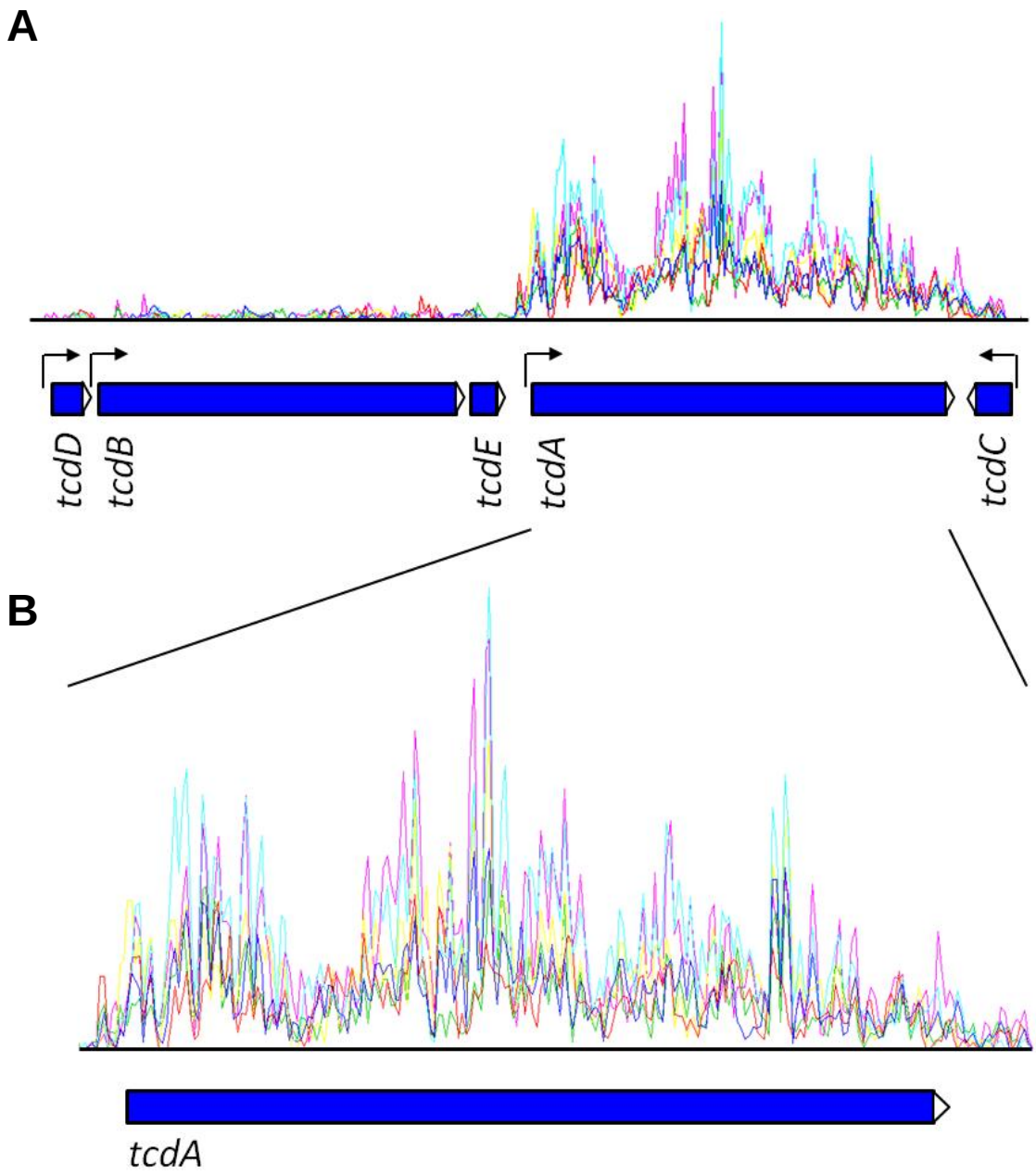

Figure 4.7. Linear representation of the *C. difficile* 630 genome with transcriptome reads mapping to the PaLoc region. A) Shown is the 19.6 kbp PaLoc region, which harbours five genes (*tcdDBEAC*) responsible for the synthesis and regulation of toxins A and B. **B)** Inset of the *tcdA* gene. Coloured lines represent transcriptome reads aligned to the 630 genomic sequence, and are coloured to represent biological replicates as follows: red, blue, green = *C. difficile* 630Δ*erm*; purple, turquoise, yellow = *C. difficile* 630Δ*erm*Δ*spo0A*. Arrows represent promoters (227). The expression of *tcdA* was significantly up-regulated at the mRNA level ($P = \leq 0.01$) in 630Δ*erm*Δ*spo0A*.

4.3.2.2.2 Spo0A is required for the expression of later stage sporulation genes

The mother-cell-specific *spoIIIA* polycistronic eight gene operon, which is involved in stage III spore formation, was significantly down-regulated in 630Δ*erm*Δ*spo0A* at the mRNA level (Figure 4.8). Many of the genes belonging to this operon (*CD1192-CD1199*) were also present in the heat map of the top 40 differentially regulated genes (Figure 4.5, highlighted in red). This was perhaps expected given the asporogenous phenotype observed in *C. difficile* 630Δ*erm*Δ*spo0A* (Chapter 3). Other sporulation genes including *spoIIIG*, *spoIIID*, *spoIVA*, *spoIIGA*, *spoVD*, *spoIIE* and the tricistronic *spoIIA* operon were also down-regulated in 630Δ*erm*Δ*spo0A*. Logically, *cspC* and *cspBA* encoding putative germination-specific proteases that are present in the spore proteome, and *CD3569* encoding a spore cortex-lytic enzyme were also down-regulated in the *spo0A* mutant. Interestingly, however, *ccpA*, a carbon catabolite repressor required for spore formation in *C. perfringens* was up-regulated in 630Δ*erm*Δ*spo0A*.

Sigma factors involved in the regulation of spore formation, such as *sigE* ($P = < 0.01$), *sigG* ($P = < 0.01$), *sigF* ($P = < 0.01$) and *sigH* ($P = 0.03$) were all significantly down-regulated in *C. difficile* 630Δ*erm*Δ*spo0A* compared to parental *C. difficile* 630Δ*erm*. In fact, *sigG* (*CD2642*) and *sigE* (*CD2643*) were present in the heat map of the top 40 differentially regulated genes (Figure 4.5, highlighted in green). *sigH* also has a consensus '0A' box located ~84 bp upstream of the start site, suggesting that regulation of *sigH* expression by Spo0A may be via a direct interaction (Figure 4.9A). This has been supported by evidence of direct binding using *in vitro* binding assays (Wiep Klass, personal communication).

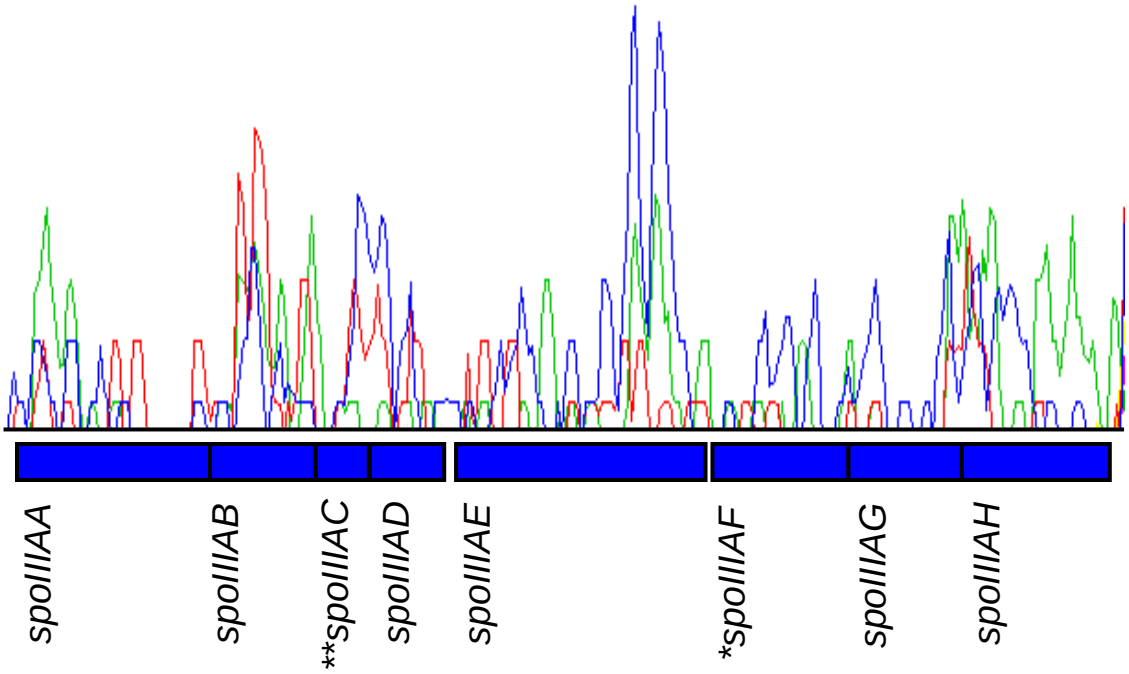

Figure 4.8. Linear representation of the *C. difficile* 630 genome with transcriptome reads mapping to the *spoIIIA* operon. Shown is the polycistronic eight gene s*poIIIA* operon, which is involved in stage III spore formation. Coloured lines represent transcriptome reads aligned to the 630 genomic sequence, and are coloured to represent biological replicates as follows: red, blue, green = *C. difficile* 630Δ*erm*; purple, turquoise, yellow = *C. difficile* 630Δ*erm*Δ*spo0A*. Expression of this operon was significantly attenuated in *C. difficile* 630Δ*erm*Δ*spo0A* ($P = \leq 0.01$) at the mRNA level. *$P = 0.04$; **$P = 0.07$.

4.3.2.2.3 Spo0A may be auto-regulated in C. difficile via a feedback loop

In *Bacillus*, *spo0A* transcription is auto-regulated via the direct binding of Spo0A to an '0A' box upstream of *spo0A*. This binding in turn represses *abrB* (a transition state regulator),

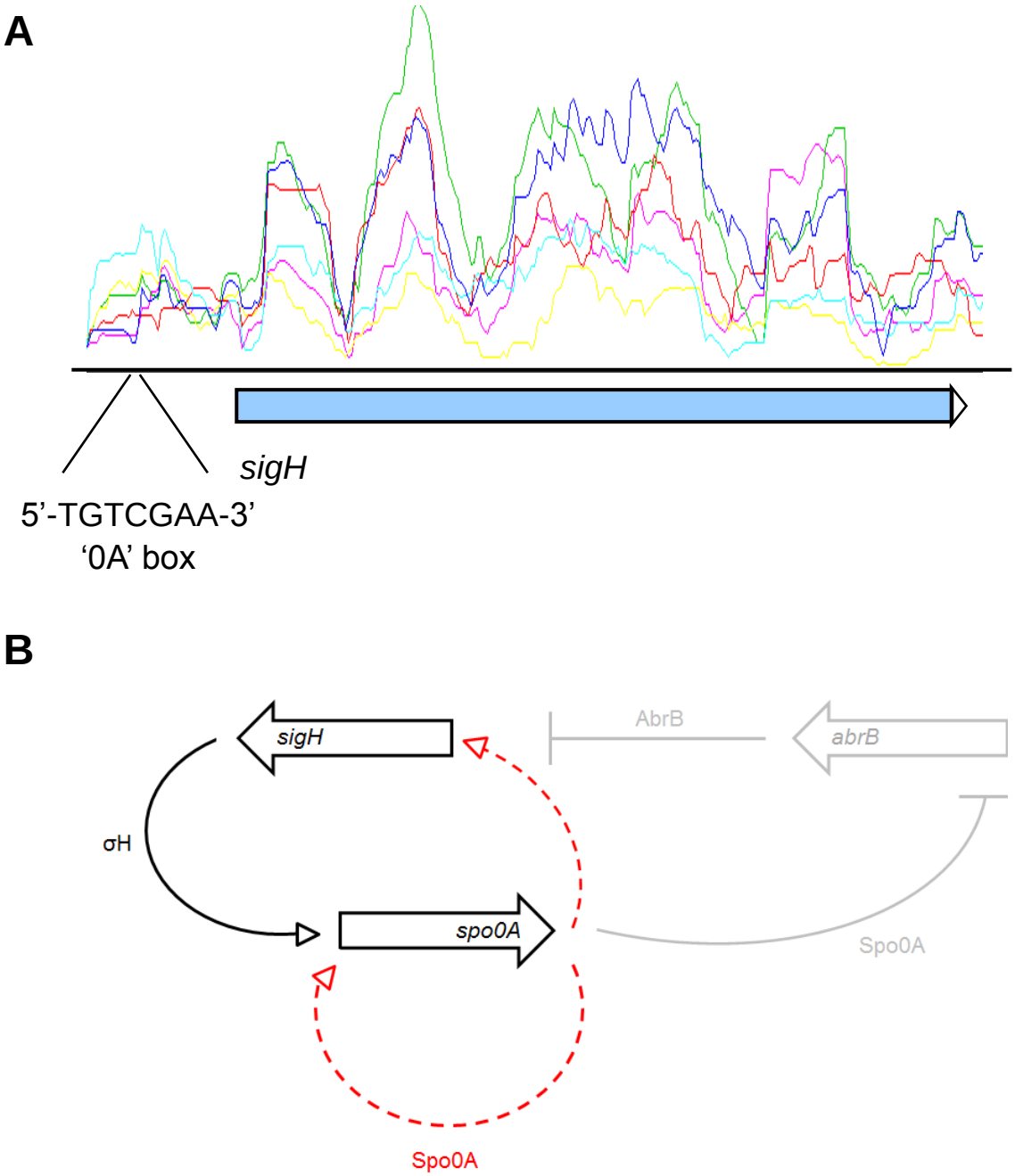

Figure 4.9. Spo0A may regulate *sigH* expression via a direct interaction. A) Expression of *sigH* was significantly attenuated in *C. difficile* 630Δ*erm*Δ*spo0A* (*P* = 0.03) at the mRNA level. Coloured lines represent transcriptome reads aligned to the 630 genomic sequence, and are coloured to represent biological replicates as follows: red, blue, green = *C. difficile* 630Δ*erm*; purple, turquoise, yellow = *C. difficile* 630Δ*erm*Δ*spo0A*. *sigH* also has a consensus '0A' box located directly upstream of its start site suggestive of direct Spo0A regulation. **B)** Schematic representation of a predicted network of *spo0A* and *sigH* regulation. Grey shading indicates regulation networks in *B. subtilis* that are absent in *C. difficile*. Red dashed lines indicate potential regulatory links in *C. difficile*.

resulting in *sigH* transcription and a concomitant activation of *spo0A*, as described in Figure 4.9B. However, based on sequence identity there is no *abrB* orthologue in *C. difficile*, indicating that the regulation loop may occur via a different mechanism.

It has previously been demonstrated that *C. difficile spo0A* is transcribed from a SigH-dependent promoter (182). Similarly, in *C. difficile* there is a consensus '0A' box directly upstream of *spo0A*, indicating direct auto-regulation by Spo0A. This is also supported by *spo0A* gel shift binding assays (Wiep Klass, personal communication). It is thus possible that *spo0A* transcription may be auto-regulated in a way comparable with *Bacillus*, but that *sigH* expression is directly driven by Spo0A rather than indirectly via AbrB (Figure 4.9B). Over-expression or ChIP-Seq studies could be used to test this hypothesis.

4.3.2.3 Spo0A is a regulator of *C. difficile* metabolism

We noted that Spo0A regulates the transcription of many genes of multiple functions. However, a large set of genes that were differentially expressed appeared to function in metabolism. For example, several genes involved in glycolysis were down-regulated in 630Δ*erm*Δ*spo0A*. These included the glycolytic enzymes glyceraldehyde-3-phosphate dehydrogenase (*gapB*) and glucose-6-phosphate isomerase (*pgi*), a central glycolytic genes regulator (*cggR*), beta-glucoside utilisation genes (*bglP*, *bglG*, *bglF*), and putative 6-phospho-alpha- and beta-glucosidases (*bglA3* and *bglA2*, respectively).

Additionally, components of the phosphotransferase (PTS) system (*CD0816*, *crr*, *ptsG* and *CD3137*) were down-regulated in 630Δ*erm*Δ*spo0A*. Similarly, two genes encoding a putative bi-functional glycine dehydrogenase (*CD1657*) and glycine cleavage system P-protein (*gcvPB*) were down-regulated in 630Δ*erm*Δ*spo0A*. These are thought to be components of the glycine cleavage system, whereby certain strict anaerobes are able to oxidise glycine to serve as an electron donor in the Stickland reaction (228-230).

In contrast, several mediators of amino group metabolism were up-regulated in 630Δ*erm*Δ*spo0A*. These included putative amino acid aminotransferases (*aspC*, *CD3664* and *CD2382*), a histidol-phosphate aminotransferase (*CD2502*) a glutamate dehydrogenase (*gluD*), a phosphoserine phosphatase (*CD0241*), an acetylornithine deacetylase (*argE*), a pyrroline-5-carboxylate reductase (*proC1*) and a D-alanine phosphoribitol ligase (*dltA*). The *hadAIBC* operon putatively involved in L-leucine reduction (231) was also found to be negatively regulated by Spo0A.

Genes involved in acidogenesis such as butyrate kinase (*buk*), phosphate butyryltransferases (*ptb*, *CD0715*), and the butyrate metabolism genes NAD-dependent 4-hydroxybutyrate dehydrogenase (*abfH*), 4-hydroxybutyrate CoA transferase (*abfT*), gamma-aminobutyrate dehydratase (*abfD*), succinate-semialdehyde dehydrogenase (*sucD*), succinyl-CoA transferase (*cat1*), acyl-CoA dehydrogenase (*bcd1*) and electron transport flavoprotein (*etfB1*) were also up-regulated in 630Δ*erm*Δ*spo0A*. Interestingly, all of these metabolic genes (excluding *CD2502*, *CD0241*, *argE* and *cat1*) encode proteins that are also present in the spore proteome.

4.3.2.3.1 Spo0A regulation of energy metabolism

Of the 153 *in silico* predicted CDSs belonging to the energy metabolism class in *C. difficile* 630, ~17% were found to be up-regulated in 630Δ*erm*Δ*spo0A* based on ssRNA-seq analyses (Figure 4.6B). These included the *rnf* complex (*rnfCDGEAB*), which is purported to be a membrane-bound multi-component system involved in NADH-dependent electron transport. This operon has been identified in many clostridial genomes, and was found to be negatively regulated by Spo0A at the mRNA level (excluding *rnfE* which was not differentially expressed). The complex is thought to oxidise NADH to NAD+ and reduce ferredoxin, however it has also been suggested that electrons can be transferred in the reverse direction (232). Previous publications have suggested the the *rnf* complex may have a role in generating NADH for butyrate fermentation (232).

Additionally, genes encoding metabolic enzymes such as oxidoreductases (including two indolepyruvate oxidoreductases, *iorA* and *iorB*), hydrogenases (*hymB* and *hymC*), flavoproteins and a cardiolipin synthetase (*cls*) were also up-regulated in 630Δ*erm*Δ*spo0A*, suggesting a possible role of Spo0A as a negative regulator of energy metabolism. Indeed, *spo0A* mutations have been shown to increase the metabolic maintenance demands of *B. subtilis* (233). However, *ntpB*, *ntpA* and *ntpI* which are members of the 8 gene ATPase synthase operon (*ntpIKECGABD*) were significantly down-regulated in 630Δ*erm*Δ*spo0A*.

4.3.2.4 Spo0A regulates the expression of surface-associated transcripts

Secreted/surface associated transcripts were well represented in our dataset (Figure 4.6). In fact, these were the most common functional class detected, and constituted ~22% and ~29% of detected transcripts which were positively and negatively regulated by Spo0A, respectively. Such proteins could be implicated in intestinal adherence, colonisation, or subversion of the hosts' immune system.

4.3.2.4.1 Surface-associated transcripts positively regulated by Spo0A

In silico predicted transcripts belonging to the secreted/surface-associated class whose expression was down-regulated in 630Δ*erm*Δ*spo0A* included some encoding ABC transporters such as ATP-binding proteins and permeases. *ssuA*, a spore-associated polypeptide, which encodes a putative periplasmic ABC transporter of aliphatic sulfonates, was also positively regulated by Spo0A. Sulfonates may have a role as electron acceptors in anaerobic respiration (234). Additionally, transcripts involved in cobalt transport (*cbiM* and *cbiQ*) were down-regulated in 630Δ*erm*Δ*spo0A*.

Adhesive cell surface proteins were also found to be potentially positively regulated by Spo0A. For example, *CD0440* and *CD2518*, which encode SLP-like proteins, are both similar to the *C. difficile* Cwp66 cell surface protein, suggesting adhesive activity. The down

regulation of such transcripts may explain the inability of *C. difficile* to cause a long-term persistent infection in our mouse model of infection (discussed in Chapter 3).

4.3.2.4.2 Surface-associated transcripts negatively regulated by Spo0A

Secreted/surface-associated transcripts whose expression was up-regulated in 630Δ*erm*Δ*spo0A* also included ABC transporters and permeases. Additionally, many transcripts encoding components of the phosphotransferase (PTS) carbohydrate transport system were also up-regulated. This system is important for anaerobic glycolysis (235) and up-regulated transcripts included sugar transporters and sugar-specific enzymes. One transcript (*CD0244*) encoding a CDP-glycerophosphotransferase involved in the biosysthesis of teichoic acids in *B. subtilis* was also negatively regulated by Spo0A in *C. difficile* 630Δ*erm*.

4.3.3 Global relationship between the transcriptome and proteome of exponentially growing *C. difficile*

According to the central dogma of molecular biology, the transcriptome is viewed as a precursor for the proteome. Whilst this is true, indeed the flow of information is from DNA to RNA to protein, it is being increasingly recognised that the concordance between transcript and protein expression is not always intuitive (224). In order to gain an accurate and thorough representation of the role of Spo0A in *C. difficile* biology and physiology, we integrated both

transcriptome and proteome data from exponentially growing *C. difficile* 630Δ*erm* and 630Δ*erm*Δ*spo0A*.

4.3.3.1 Analysis of proteins from exponentially growing *C. difficile*

4.3.3.1.1 SDS-PAGE

To ensure that proteomic data were comparable with transcriptome analyses, proteins were harvested from exponentially growing *C. difficile* 630Δ*erm* and 630Δ*erm*Δ*spo0A*, which were then electrophoretically separated by SDS-PAGE, as shown in Figure 4.10. Many polypeptide bands were resolved, with molecular weights ranging from 6 to > 100 kDa. Visual observations indicated that there was a common backbone of proteins shared between *C. difficile* 630Δ*erm* and 630Δ*erm*Δ*spo0A*, with both having similar mobilities and distribution profiles.

4.3.4.1.2 Mass-spectrometry

We then performed LC-MS/MS on three biological replicates and three technical replicates of both *C. difficile* 630Δ*erm* and *C. difficile* 630Δ*erm*Δ*spo0A*. In order for proteins to be quantified and to thus enable direct comparison of protein expression between mutant and parental *C. difficile*, samples were also dimethyl labelled prior to mass-spectrometry, as

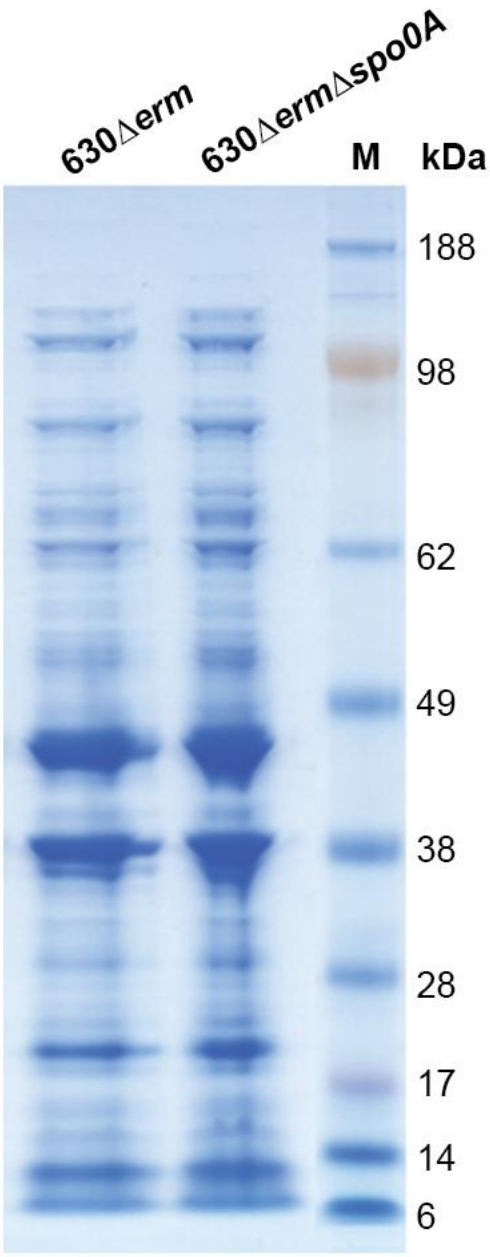

Figure 4.10. SDS-PAGE of cellular proteins of *C. difficile* 630Δ*erm* and 630Δ*erm*Δ*spo0A*. Secreted and cellular proteins with an equivalency to ~2x10^7 cells per lane were harvested from exponentially growing cultures, resolved by 12% SDS-PAGE and visualised by Coomassie-blue staining. Lane: 630, proteins extracted from *C. difficile* 630Δ*erm*; Δ*spo0A*, proteins extracted from *C. difficile* 630Δ*erm*Δ*spo0A*; M, protein molecular weight markers.

described in Materials and methods chapter 2.2.8.7. Based on a FDR of 1% and a minimum positive match of two peptides, we identified a total of 1,276 proteins by LC-MS/MS analysis. To identify a subset of these proteins that were considered to be significantly differentially regulated by Spo0A, all 1,276 proteins were $\log_2$ transformed and ranked based on relative expression between *C. difficile* 630Δ*erm* and 630Δ*erm*Δ*spo0A*, as shown in Figure 4.11. Based on these parameters, 150 proteins considered to be significantly differentially expressed were identified, of which 86 and 64 were up- and down-regulated in *C. difficile* 630Δ*erm*Δ*spo0A*, respectively (see Appendices 4 and 5, respectively).

4.3.3.2 Correlation between transcriptome and proteome expression

The level of concordance between the Spo0A-regulated transcriptome and proteome of *C. difficile* 630Δ*erm* is shown in Figure 4.12. Interestingly, only a small subset of CDSs were found to be differentially regulated by Spo0A at both the transcriptional and translational levels, indicating that transcript quantification alone may be insufficient to generate a comprehensive overview of this biological system. Furthermore, one CDS, *CD0865*, encoding a conserved hypothetical protein was down-regulated as a transcript, but up-regulated as a protein in 630Δ*erm*Δ*spo0A*. Thus, the key assumption that mRNA expression accurately informs protein expression may not hold true in *C. difficile*. Additionally, molecular events such as the efficiency of translation or protein degradation can notably alter protein levels independently of mRNA expression (236). However, it is possible that some of these observed differences are due to the technical limitations of our assays.

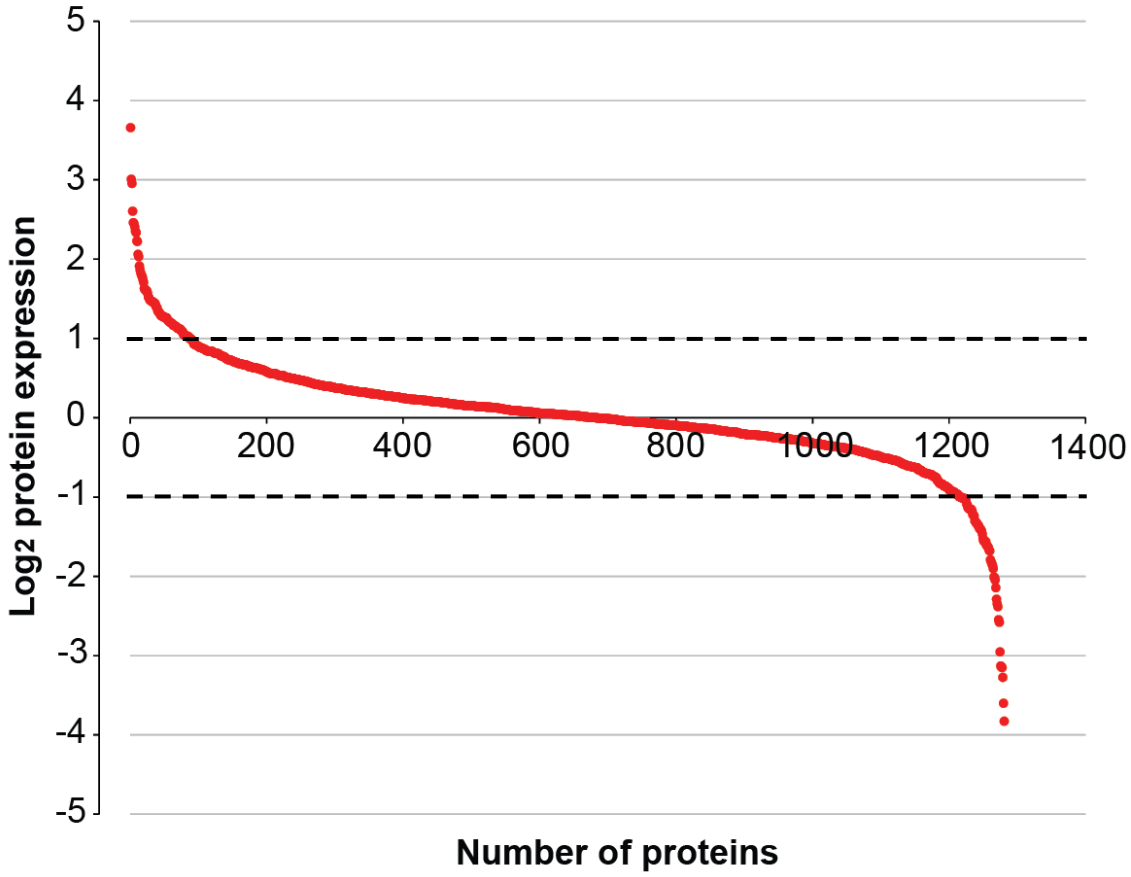

Figure 4.11. Total proteins identified by LC-MS/MS in *C. difficile* 630Δ*erm* and 630Δ*erm*Δ*spo0A*. A total of 1,276 proteins were identified by dimethyl labelled LC-MS/MS analysis, and are shown ranked based on relative expression between *C. difficile* 630Δ*erm* and 630Δ*erm*Δ*spo0A*. 86 proteins were considered to be significantly up-regulated in *C. difficile* 630Δ*erm*Δ*spo0A* and 64 proteins were considered to be significantly down-regulated in *C. difficile* 630Δ*erm*Δ*spo0A*, according to our parameters of significance (represented by a dashed line).

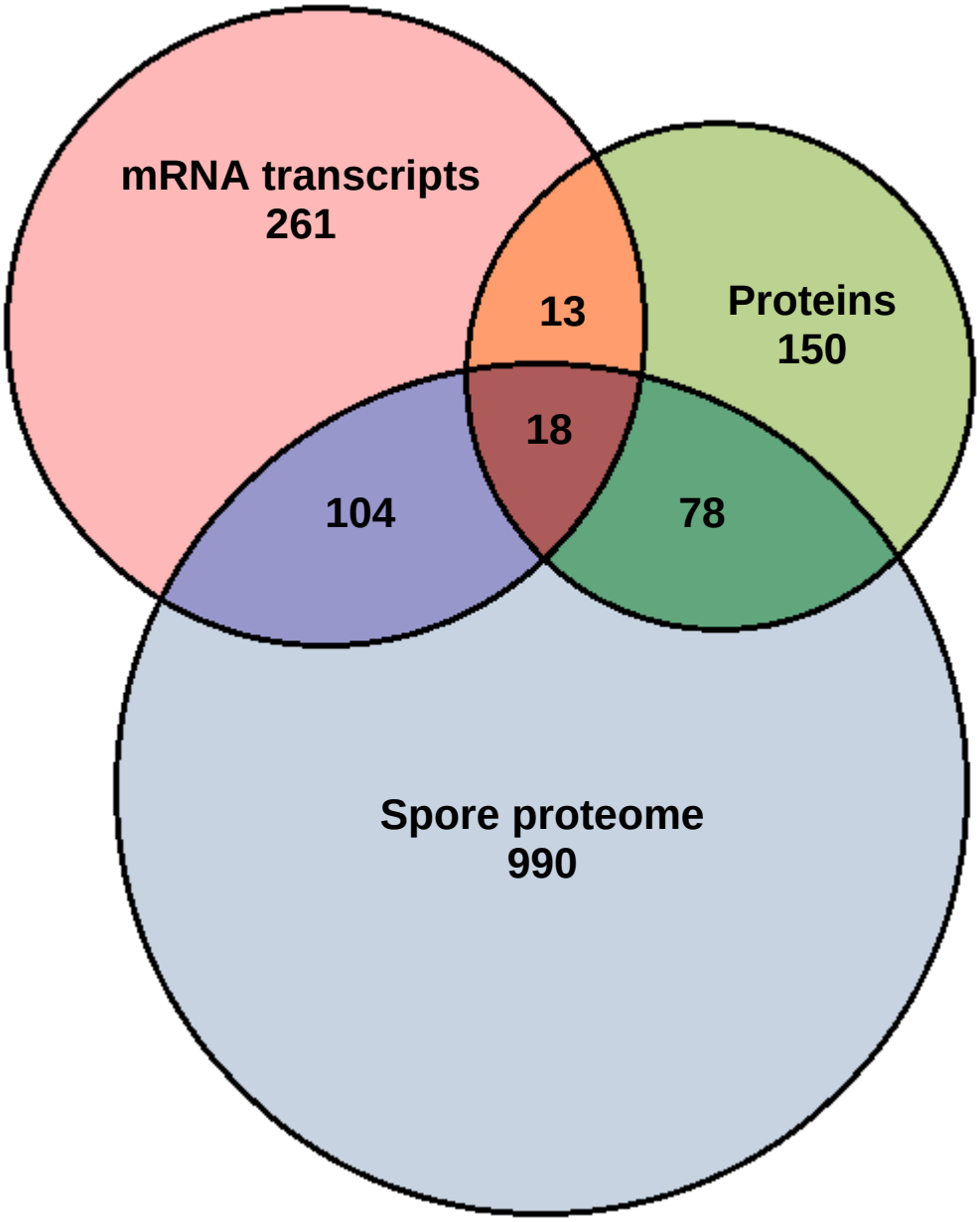

Figure 4.12. Venn diagram demonstrating the correlation between the Spo0A-regulated transcriptome and proteome of *C. difficile* 630Δ*erm*. 261 mRNA transcripts and 150 proteins under Spo0A regulation were identified, of which 31 "core" overlapping genes were detected. 18 of these "core" genes were also found to be present in the spore proteome, which is shown for reference.

*4.3.3.2.1 Proteins up-regulated in C. difficile 630Δ*erm*Δ*spo0A*

86 proteins were found to be negatively regulated by Spo0A, that is, their expression was increased in 630Δ*erm*Δ*spo0A*. Of these 86, 17 (~20%) were predicted to belong to the central/intermediate metabolism class of proteins. These include numerous metabolic enzymes such as AroK (shikimate kinase), CoaD (phosphopantetheine adenylyltransferase), Cmk (cytidylate kinase), Gmk (guanylate kinase) and LplA (lipoate-protein ligase), of which the latter was also significantly up-regulated as mRNA.

In addition, several amino group metabolism proteins were up-regulated in 630Δ*erm*Δ*spo0A*, such as GabT (4-aminobutyrate aminotransferase), NanA (N-acetylneuraminate lyase), ProC1 and ProC2 (pyrroline-5-carboxylate reductases). ProC1 was also negatively regulated by Spo0A at the transcript level, and is present in the spore proteome. The glycolytic enzyme GapA (glyceraldehyde-3-phosphate dehydrogenase) was also up-regulated in 630Δ*erm*Δ*spo0A*. This was unexpected since several genes involved in glycolysis were down-regulated in 630Δ*erm*Δ*spo0A* as transcripts (discussed earlier). In addition, FliM (flagellar motor switch protein) and FleN (flagellar number regulator) were up-regulated.

In a similar trend to the transcriptome data, enzymes which function in energy metabolism such as Hcp (hydroxylamine reductase), GrdA and GrdD (glycine/sarcosine/betaine reductases) HymA (iron-only hydrogenase) and oxidoreductases (CD1797 and CD2100) were also up-regulated in 630Δ*erm*Δ*spo0A*. Interestingly, the expression of *CD0065* encoding a NADP-dependent 7-alpha-hydroxysteroid dehydrogenase was significantly increased in

630Δ*erm*Δ*spo0A* as both mRNA and protein, and was also present in the spore proteome. This is thought to function in the catabolism of bile acids (237, 238). Thus although there is a low concordance between the transcriptome and proteome data, both appear to point towards an altered metabolic capacity.

*4.3.3.2.2 Proteins down-regulated in C. difficile 630Δ*erm*Δ*spo0A*

64 proteins were down-regulated in *C. difficile* 630Δ*erm*Δ*spo0A*. Proteins assigned to the surface-associated functional class were well represented and accounted for ~27% of polypeptides positively regulated by Spo0A. They included ABC transporters (CD0873, CD2311, CD2818, CD2989) and cell wall hydrolases (CD0183, CD3567). Several components of the PTS system were also represented (Crr, CD0492, CD3127) including BglF, which was also down-regulated as a transcript in 630Δ*erm*Δ*spo0A*.

Although *sigE*, *sigG* and *sigH* were down-regulated in 630Δ*erm*Δ*spo0A* as transcripts, they were not significantly differentially expressed as proteins. However, SigB (RNA polymerase sigma-B factor), which was not represented as a differentially regulated transcript was down-regulated in 630Δ*erm*Δ*spo0A* as a protein. SpoVG (stage V sporulation protein G) was the only known sporulation gene found to be potentially positively regulated by Spo0A at the protein level, despite several others being positively regulated as transcripts (see section *4.3.3.2.2*). SspA (small acid-soluble spore protein A) was also down-regulated in 630Δ*erm*Δ*spo0A*. This was logical since 630Δ*erm*Δ*spo0A* does not make spores and SASPs

are abundant in the spore core and function to protect DNA during periods of dormancy (157, 160, 161) (see Chapter 1.4.2).

4.3.3.2.3 Identification of "core" Spo0A-regulated genes

Despite the modest correlation between our transcriptome and proteome datasets, we were able to identify a set of 31 "core" genes that are differentially regulated by Spo0A at both the RNA and protein levels (Figure 4.12). These "core" genes are listed in Table 4.1. Based on our extensive transcriptome and proteome data, we decided to focus on two phenotypes associated with Spo0A:

1. Butyrate production
2. Flagellar assembly

Table 4.1. "Core" *C. difficile* 630Δ*erm* genes regulated by Spo0A

Identifier (gene name)	Product	Regulation in 630Δ*erm*Δ*spo0A*
CD0065*	NADP-dependent 7-alpha-hydroxysteroid dehydrogenase	Up
CD0142*	Putative RNA-binding protein	Down
CD0208*	PTS system, IIb component	Up
CD0209*	Putative sugar-phosphate kinase	Up
CD0242	Conserved hypothetical protein	Up
CD0440	Cell surface protein	Down
CD0865*	Conserved hypothetical protein	Down as mRNA but up as protein
CD1054 (Bcd2)*	Butyryl-CoA dehydrogenase	Down
CD1055 (EtfB2)*	Electron transfer flavoprotein beta-subunit	Down
CD1056 (EtfA2)*	Electron transfer flavoprotein alpha-subunit	Down
CD1057 (Crt2)*	3-hydroxybutyryl-CoA dehydratase	Down
CD1058 (Hbd)*	3-hydroxybutyryl-CoA dehydrogenase	Down
CD1059 (ThlA1)*	Acetyl-CoA acetyltransferase	Down
CD1484 (SsuA)*	Putative aliphatic sulfonates ABC transporter, substrate-binding lipoprotein	Down
CD1494	Putative transcriptional regulator	Up
CD1495 (ProC1)*	Pyrroline-5-carboxylate reductase	Up

CD1522	Putative polysaccharide deacetylase	Up
CD1654 (LplA)	Putative lipoate-protein ligase	Up
CD1717	Conserved hypothetical protein	Up
CD1797*	Putative pyridine nucleotide-disulfide oxidoreductase	Up
CD1930	Putative exported protein	Up
CD2010 (Erm1(B))	rRNA adenine N-6-methyltransferase (erythromycin resistance)	Up
CD2127	Putative membrane protein	Down
CD2181*	Putative aromatic compounds hydrolase	Down
CD2195*	Ferritin	Down
CD2765	Putative penicillin-binding protein repressor	Up
CD2796	Cell surface protein	Up
CD2797	Putative exported protein	Up
CD3116 (BglF)	PTS system, beta-glucoside-specific IIabc component	Down
CD3137*	PTS system, IIabc component	Down
CD3555*	Conserved hypothetical protein	Up

* CDS that is also present in the spore proteome.

Note that CD0865* is up-regulated as a protein, but down-regulated as a transcript. Grey shaded area represents an entire operon (CD1054-CD1059) involved in butyrate production that is down-regulated in 630Δ*erm*Δ*spo0A*.

4.3.4 Spo0A positively regulates butyrate production in *C. difficile* 630Δ*erm*

Butyrate, a short-chain fatty acid (SCFA) is the product of the fermentative metabolism by anaerobic bacteria in the colon (239). The role of butyrate in colonic health is established, and it is well recognised that butyrate serves as an important substrate for colonocyte metabolism (239). The production of butyrate is a common feature of anaerobic, Gram positive, colon-dwelling bacteria, but is perhaps best studied in the solventogenic clostridia, such as *C. acetobutylicum* and *C. butyricum* (240, 241).

In *C. difficile* 630, butyrate production is encoded by a 6 kbp polycistronic operon (Figure 4.13). This encodes butyryl-CoA dehydrogenase (*bcd2*), electron transfer flavoprotein β-subunit (*etfB2*), electron transfer flavoprotein α-subunit (*etfA2*), 3-hydroxybutyryl-CoA dehydratase (*crt2)*, 3-hydroxybutyryl-CoA dehydrogenase (*hbd)* and acetyl-CoA acetyltransferase (*thlA1)*. Our analyses have indicated that this entire operon (*bcd2, etfB2, etfA2, crt2, hbd, thlA1*) was significantly down-regulated in 630Δ*erm*Δ*spo0A*, at both the RNA and protein levels (Table 4.1). Figure 4.13 describes this operon and its positive regulation by Spo0A. We were unable to identify an '0A' box upstream of the butyrate operon, implying that its regulation is most likely indirect via an intermediate mediator. In addition, *in vitro* binding assays did not indicate that any genes in this operon were direct targets of Spo0A (Wiep Klass, personal communication). Figure 4.14 illustrates the principle components involved in butyrate metabolism, overlaid with the *C. difficile* Spo0A regulated

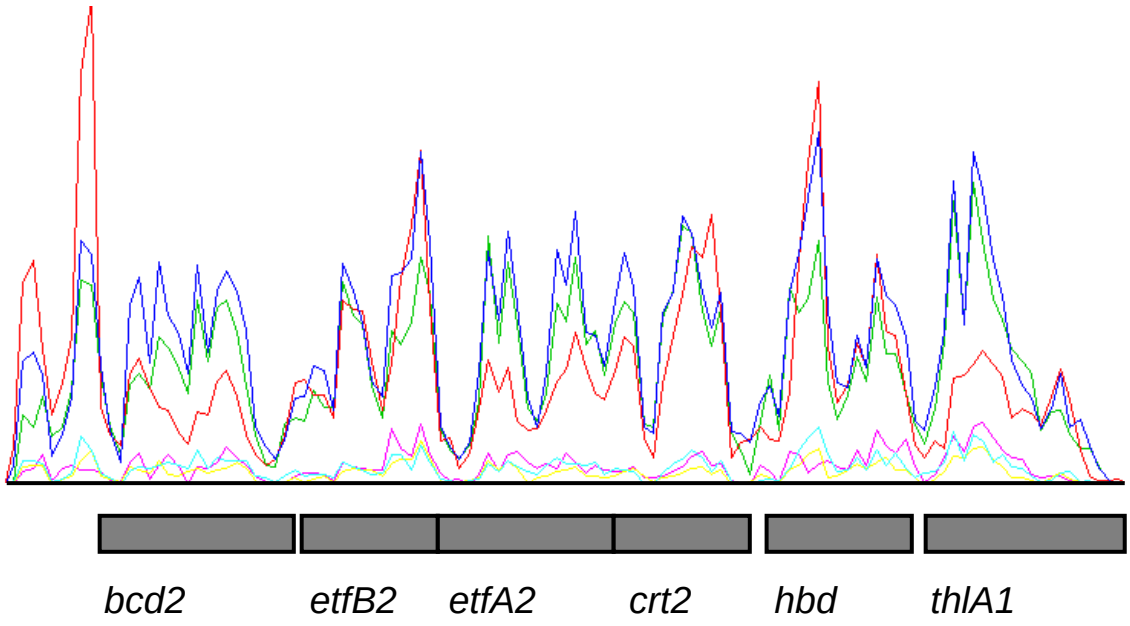

Figure 4.13. Linear representation of the *C. difficile* 630 genome with transcriptome reads mapping to a region involved in butyrate production. The 6 kbp region comprises a polycistronic, six gene operon (*CD1054-CD1059*) involved in butyrate metabolism in *C. difficile* 630. *bcd2*, butyryl-CoA dehydrogenase; *etfB2*, electron transfer flavoprotein beta-subunit; *etfA2*, electron transfer flavoprotein alpha-subunit; *crt2*, 3-hydroxybutyryl-CoA dehydratase; *hbd*, 3-hydroxybutyryl-CoA dehydrogenase; *thlA1*, acetyl-CoA acetyltransferase. Coloured lines represent transcriptome reads aligned to the 630 genomic sequence, and are coloured to represent biological replicates as follows: red, blue, green = *C. difficile* 630Δ*erm*; purple, turquoise, yellow = *C. difficile* 630Δ*erm*Δ*spo0A*. Expression of this operon was significantly attenuated in *C. difficile* 630Δ*erm*Δ*spo0A* at both the mRNA and protein levels. All proteins were also identified in the *C. difficile* 630Δ*erm* proteome.

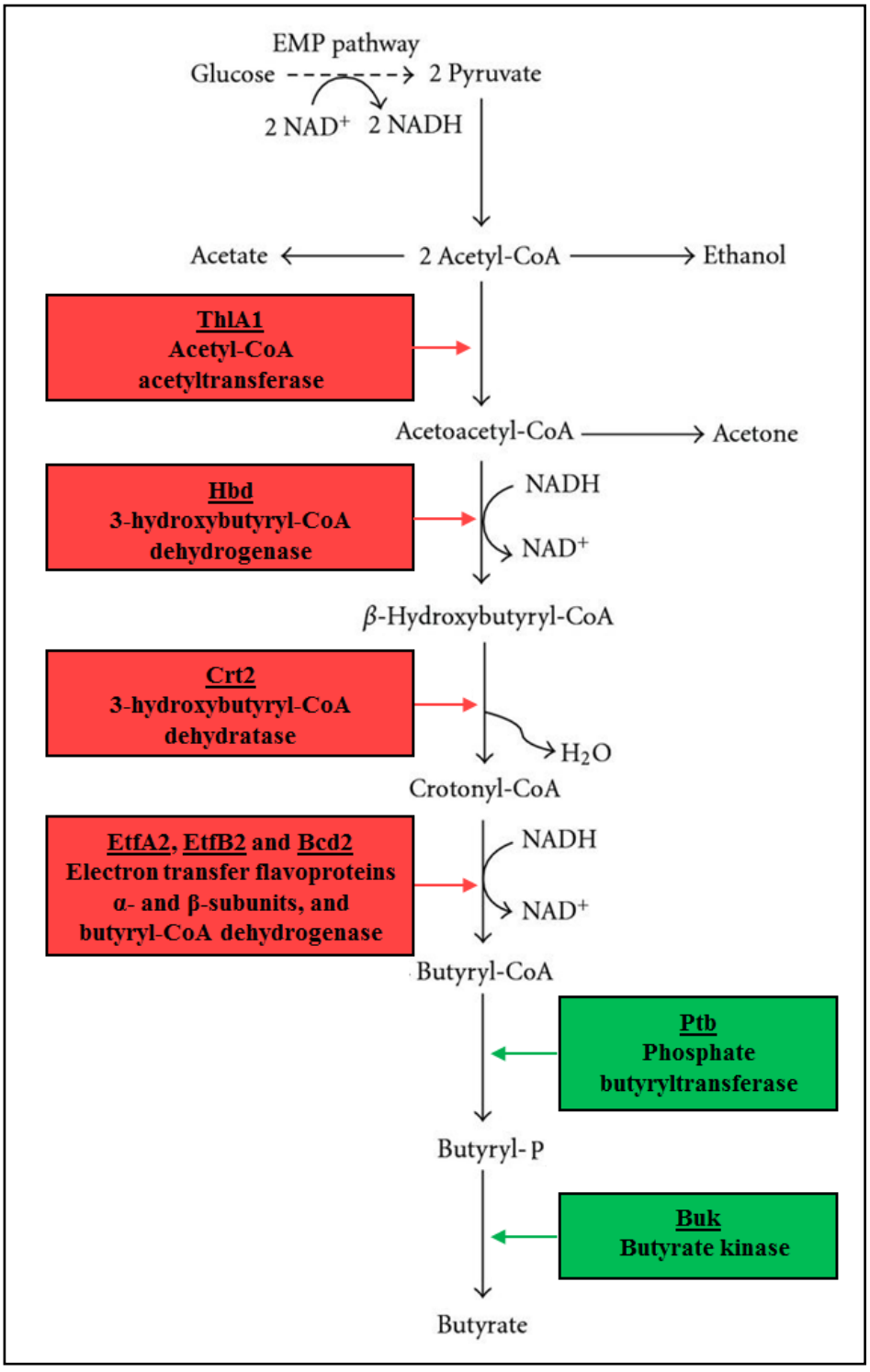

Figure 4.14. Metabolic pathway of butyrate biosynthesis. The pathway is based on data from *C. acetobutylicum* and *C. butyricum.* Components are coloured as follows: red, enzymes that are positively regulated by *C. difficile* Spo0A at both the transcriptional and translational levels; green, enzymes negatively regulated by *C. difficile* Spo0A at the transcriptional level. Adapted from: (240, 241).

enzymes. Interestingly, *ptb* and *buk*, encoding a phosphate butyryltransferase and a butyrate kinase, respectively, were up-regulated in 630Δ*erm*Δ*spo0A* at the mRNA level.

Proteins encoded by this operon are also present in the *C. difficile* 630Δ*erm* spore proteome, perhaps suggesting a possible association between butyrate and spore formation and/or germination in *C. difficile* 630Δ*erm*. Indeed, it has been previously reported that butyrate is a spore lipid precursor and that it's fermentation is linked to anaerobic sporulation (242). Similarly, the incorporation of butyrate into germinating spores of *B. thuringiensis* has been described (243).

4.3.4.1 Spo0A increases butyrate production in *C. difficile*

Given that our molecular data appeared to suggest that Spo0A positively regulates butyrate biosynthesis in *C. difficile* 630Δ*erm*, we decided to quantify butyrate taken from exponentially growing *C. difficile* 630Δ*erm* and 630Δ*erm*Δ*spo0A*.

We found that *C. difficile* 630Δ*erm*Δ*spo0A* produced significantly less butyrate than the parental 630Δ*erm* strain (P = 0.0005; Figure. 4.15). Moreover, the 630Δ*erm*Δ*spo0A* complemented mutant produced levels of butyrate that were statistically comparable with the 630Δ*erm* parental strain (P = 0.6; Figure 4.15). Our phenotypic and genotypic data are concordant with previous studies, which have demonstrated that asporogenic *C. botulinum* do not produce butyrate (242).

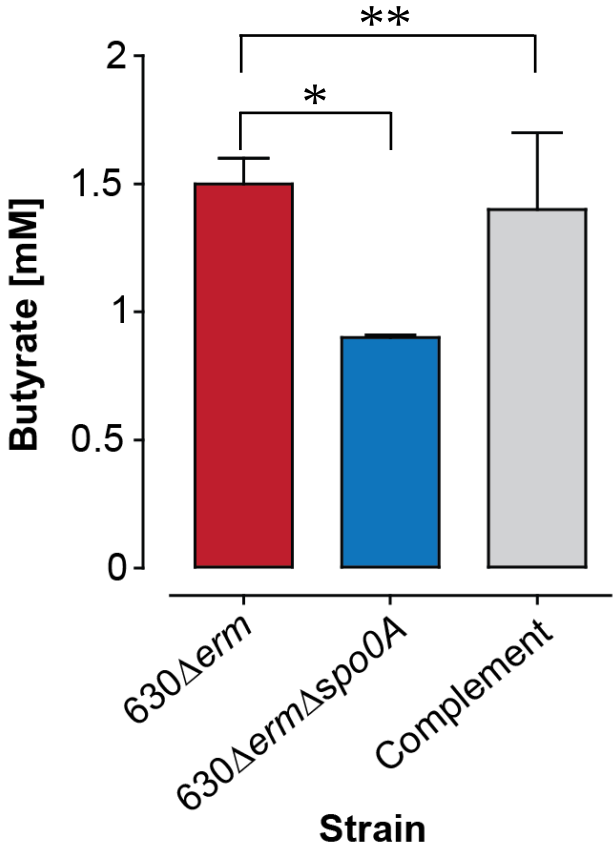

Figure 4.15. Butyrate quantification formed by *C. difficile* 630Δ*erm* derivatives. Cultures of *C. difficile* 630Δ*erm* and 630Δ*erm*Δ*spo0A* were grown in Wilson's broth under anaerobic conditions, and samples were taken during mid-exponential growth. The level of butyrate (mM) in the culture supernatants was then quantified via capillary gas chromatography. **P* = 0.0005; ***P* = 0.6.

4.3.5 Spo0A negatively regulates components of the *C. difficile* flagellar assembly apparatus

Although components of the flagellar assembly apparatus did not appear in our list of "core" genes, we did identify several differentially expressed flagella-associated transcripts and proteins, suggesting that this phenotype may be negatively regulated by Spo0A.

In *C. difficile* 630, components of the flagellar assembly apparatus are encoded by two loci that are divided by an inter-flagellar locus that may be involved in flagellin glycosylation (11). In *Bacillus*, motility and sporulation show mutual regulation (176). Furthermore, in *C. acetobutylicum* motility genes were shown to be inhibited by Spo0A (244). Our analyses indicated that seven genes encoding flagellar proteins (*fliC, fliF, fliG, fliH, fliJ, fliK* and *CD0255A*) were up-regulated in 630Δ*erm*Δ*spo0A*. Interestingly, *CD0744* and *CD0745*, both encoding putative chemotaxis proteins, were down-regulated in 630Δ*erm*Δ*spo0A*.

Surprisingly, this was not in concordance with our proteome dataset and none of the differentially regulated transcripts were considered to be differentially regulated proteins, according to our parameters of significance. Nonetheless, all of these transcripts (excluding *fliJ*) were represented in the proteome dataset alongside other flagellar proteins including FliA, FliS1, FliD, FliK and FliN, but at a statistically comparable level to parental *C. difficile* 630Δ*erm*. In addition, FliM, a flagellar motor switch protein and FleN, a flagellar number regulator were both found to be up-regulated in 630Δ*erm*Δ*spo0A* at the protein level, but were not considered to be significantly differentially regulated as transcripts. Figure 4.16 indicates

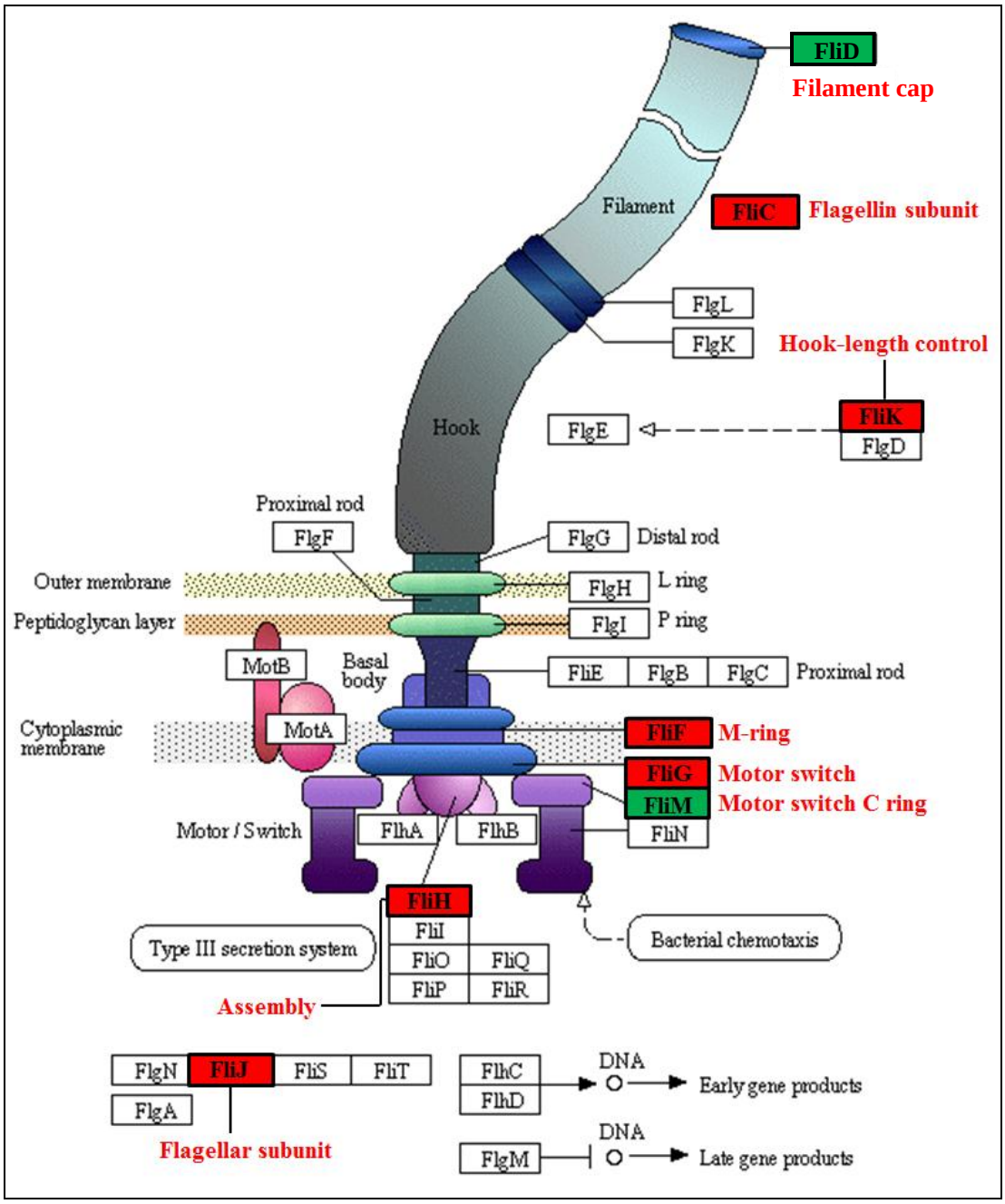

Figure 4.16. Schematic model of the bacterial flagellar assembly apparatus. Overview of the components involved in flagellar production, based on KEGG pathway analysis. CDS highlighted in red indicate those which are significantly up-regulated in *C. difficile* 630Δ*erm* Δ*spo0A* at the mRNA level, those in green indicate significant up-regulation in 630Δ*erm*Δ*spo0A* at the protein level. Note that CD0255A encoding a putative flagellar protein, and FleN encoding a flagellar number regulator are not represented in this model, although they are differentially expressed at the mRNA and protein levels, respectively, in our datasets. Courtesy: http://www.genome.jp/kegg/

the principle components of flagellar assembly, with differentially expressed genes given in red, and differentially expressed proteins given in green.

The expression of *CD0226*, encoding a putative transglycosylase, was up-regulated in 630Δ*erm* Δ*spo0A*, suggesting a possible link between Spo0A and flagellar glycosylation. This idea is further supported by the protein-protein interactions predicted by the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) 9.0 database, as shown in Figure 4.17.

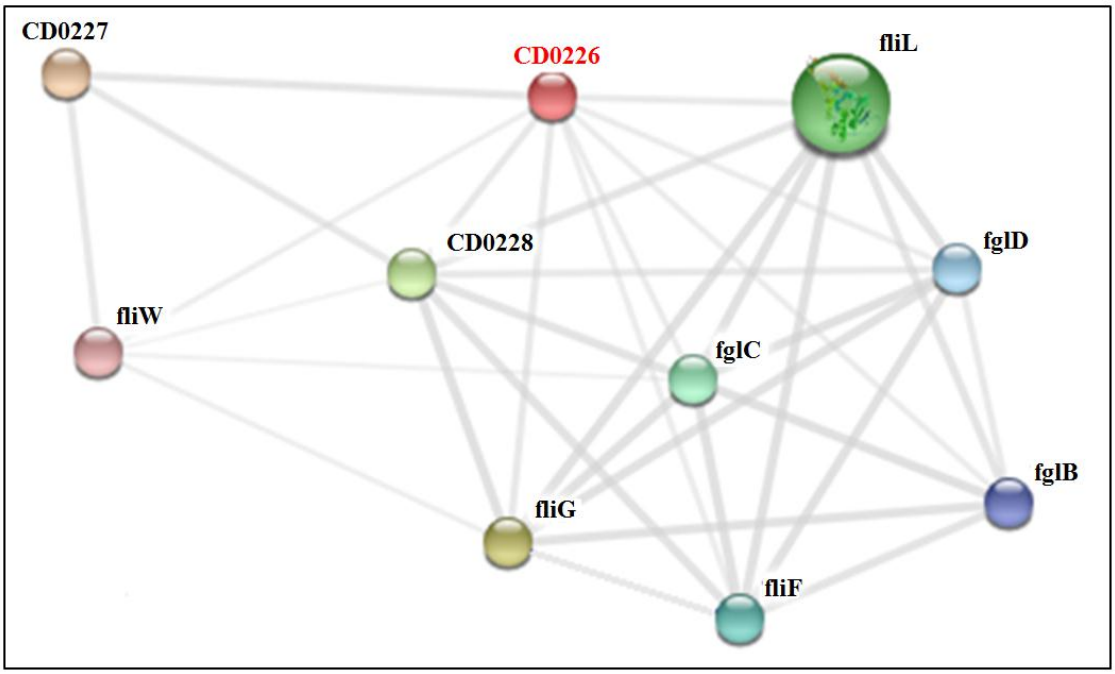

Figure 4.17. Schematic representation of the protein-protein interactions among components of the flagellar apparatus. Interactions are predicted by the STRING 9.0 database. CD0226 (shown in red) encodes a putative transglycosylase predicted to interact with components of the flagellar assembly apparatus that are significantly up-regulated in 630Δ*erm*Δ*spo0A*. Courtesy: http://string-db.org/

Furthermore, the four transcripts (*CD0241-CD0244*) that constitute the inter-flagellar glycosylation locus (11, 245) were also up-regulated in 630Δ*erm*Δ*spo0A* at the mRNA level. Three of these (CD0241, CD0243 and CD0244) were present in the proteome but were not considered to be significantly differentially regulated by Spo0A, whereas CD0242 was deemed be negatively regulated by Spo0A at both the mRNA *and* protein levels. This again indicates some negative association between Spo0A and flagellar glycosylation.

4.3.5.1 Spo0A represses *C. difficile* flagellum assembly

Given the debate surrounding the role of Spo0A in the regulation of motility, we analysed *C. difficile* 630Δ*erm* and 630Δ*erm*Δ*spo0A* for the presence of flagella by negative staining and TEM. We also examined *C. difficile* R20291 and R20291Δ*spo0A* in the same manner, since previous publications have suggested that 630 derivatives are inherently less motile than ribotype 027 strains (11).

We found that *C. difficile* has peritrichous flagella, which is consistent with previous reports. However, there were notable differences between the parent and *spo0A* mutant derivatives, in both 012 (630Δ*erm*) and 027 (R20291) ribotypes. TEM negative stains are shown in Figure 4.18. Most striking was the increased number of flagella in the 630Δ*erm*Δ*spo0A* and R20291Δ*spo0A* mutant derivatives in comparison with their respective parental strains (Table 4.2). In fact, in *C. difficile* 630Δ*erm* images, no cell anchored flagella were visible at all (this observation was consistent over three biological replicates).

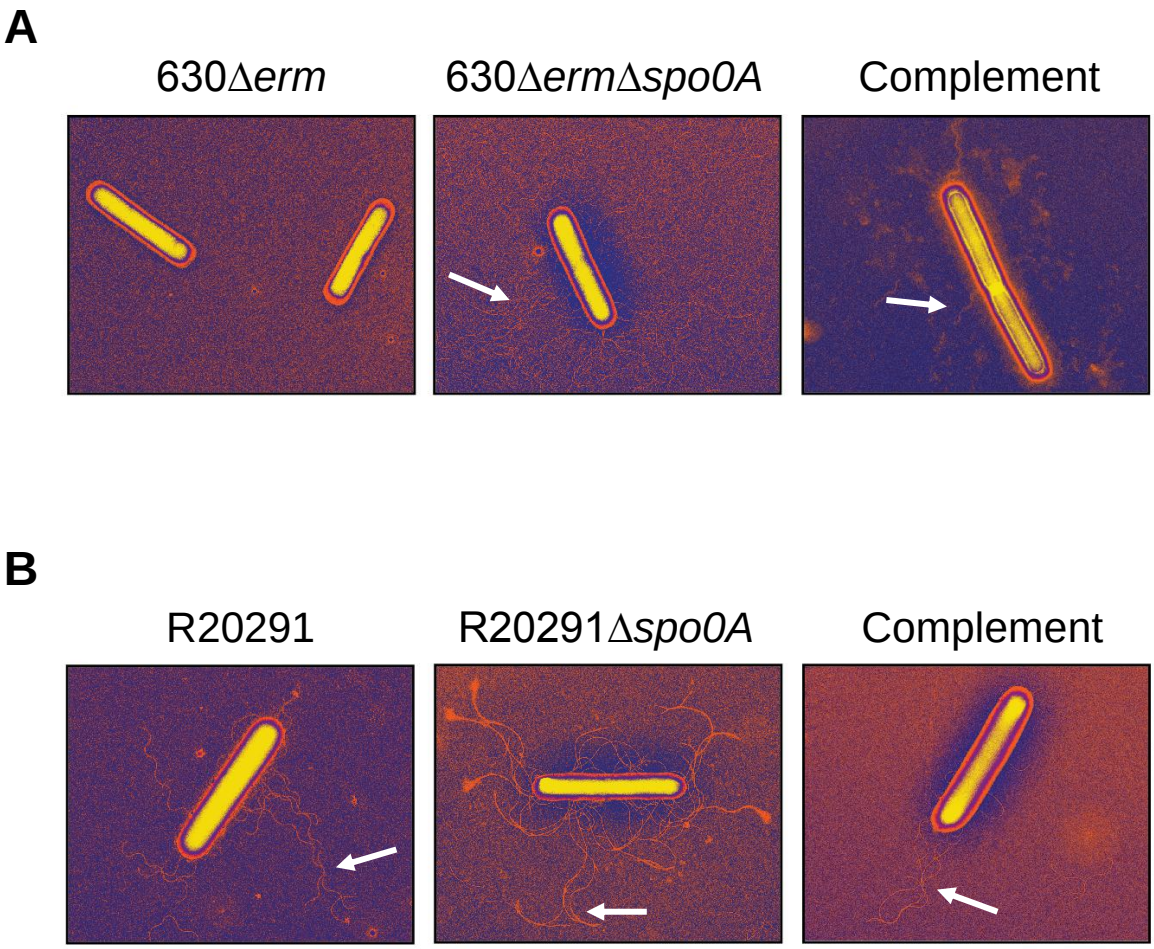

Figure 4.18. Spo0A is a negative regulator of flagellar synthesis. Representative Transmission Electron Micrographs of negatively stained *C. difficile* 630Δ*erm* **(A)** and R20291 **(B)** derivatives. Genetic complementation of the *spo0A* mutation restored flagella synthesis to levels comparable with the parental strains. Note that we were consistently unable to detect any cell anchored flagella in *C. difficile* 630Δ*erm*. Images taken by David Goulding, WTSI.

Table 4.2. Summary of flagellar composition in *C. difficile* ribotypes 012 (630Δ*erm*) and 027 (R20291)

	Flagellar abundance* (average number/cell)	**Flagellar length (average µm)**	**Flagellar diameter (average nm)**
Ribotype 012			
630Δ*erm*	0	Not determined	Not determined
630Δ*erm*Δ*spo0A*	8.1	8.8	15.2
630Δ*erm* complement	4.8	3.5	14.6
Ribotype 027			
R20291	4.6	5.3	15.1
R20291Δ*spo0A*	12.6	5.6	13.2
R20291 complement	6.2	6.8	12.0

* Note that only flagella which appeared to be anchored to the cell surface were recorded.

Increased levels of the flagella number regulator, FleN, at the protein level may account for the overproduction of flagella in 630Δ*erm*Δ*spo0A*. Since we could not detect any flagella on

the parental 630Δ*erm* strain, we were unable to draw conclusions regarding the role of Spo0A in flagellar length and diameter regulation in this genetic background.

In the R20291 background, there were almost three times as many flagella in the Spo0A mutant (average of 12.6) compared to its parent R20291 strain (average of 4.6). However, there were no notable differences in flagellar length or diameter, as described in Table 4.2. In both *C difficile* 630Δ*erm* and *C. difficile* R20291, the complemented *spo0A* mutants appeared to have a flagellar phenotype that most closely resembled its respective parental strain (Figure 4.18 and Table 4.2).

In addition, we performed a Western blot with FliC (significantly regulated by Spo0A at the mRNA level) and FliD (significantly regulated by Spo0A at the protein level) antibodies. Samples were taken at mid-exponential phase, in order to give the highest level of concordance between the transcriptome and proteome datasets. We found that *C. difficile* 630Δ*erm* produced lower levels of both FliC and FliD compared to the 630Δ*erm*Δ*spo0A* mutant derivative (Figure 4.19). Somewhat paradoxically, the 630Δ*erm*Δ*spo0A* complemented strain produced levels of FliC and FliD that were more comparable with the 630Δ*erm*Δ*spo0A* mutant (Figure 4.19), despite producing a reduced number of flagella as observed via TEM (Figure 4.18). However, given that the Western blot was performed on whole cell lysates of the *C. difficile* cultures, it is possible that these proteins may have been present intracellularly in the complemented strain, rather than being expressed as functional components of the *C. difficile* flagellar apparatus on the cell surface. Furthermore, it may be that glycosylation plays an important role in the activity of these proteins.

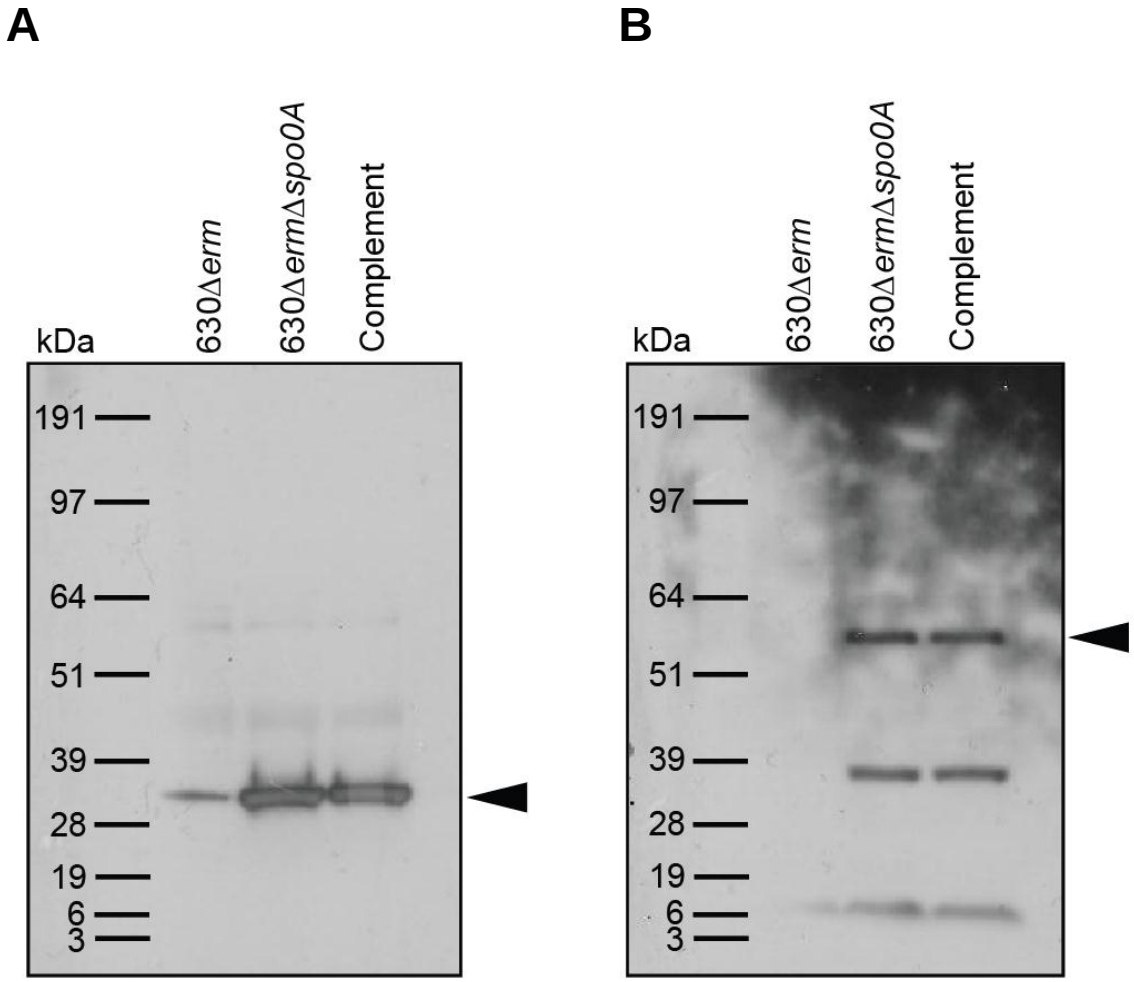

Figure 4.19. Western Blot analysis of flagellar gene expression in *C. difficile* 630Δ*erm* derivatives. *C. difficile* cultures were grown in Wilson's broth under anaerobic conditions, and samples were taken during mid-exponential growth. Extracted proteins with an equivalency to ~10^7 cells/lane were probed with **(A)** anti-*C. difficile* FliC and **(B)** anti-*C. difficile* FliD antibodies. Arrow heads indicate the approximate predicted molecular masses of FliC (30.9 kDa) and FliD (56 kDa).

4.4 Discussion

We previously demonstrated that the *C. difficile spo0A* gene is a persistence and transmission factor *in vivo* (Chapter 3). The availability of a defined *spo0A* mutant facilitated studies on the Spo0A regulon and its role in growth, sporulation and pathogenesis, at the whole genome level. Here, using a combination of transcriptomic and proteomic approaches, we demonstrate that the *spo0A* gene of *C. difficile* 630Δ*erm* plays a highly pleiotropic role and is a global transcriptional regulator that coordinates multiple virulence, sporulation and metabolic phenotypes.

Importantly, we also validated Spo0A as a transcriptional regulator of a number of sporulation genes, which are often demarked by canonical Spo0A binding sites, as shown in Figure 4.20. Furthermore, we confirmed that Spo0A negatively regulates toxin production in *C. difficile* 630Δ*erm* at the transcriptional level. In addition, we identified novel phenotypes associated with the Spo0A regulon. For example, we demonstrated that Spo0A negatively regulates key components of the *C. difficile* flagellar assembly apparatus and positively regulates several metabolic pathways, including the fermentation of carbohydrates leading to the production of butyrate. These phenotypes may have implications relating to the colonisation and persistence of *C. difficile* 630Δ*erm in vivo*.

The role of flagella in the colonisation and pathogenesis of several enteropathogens, including *V. cholera* (246), *Campylobacter jejuni* (247) and *Helicobacter pylori* (248) is well recognised. Additionally, recent work in *C. difficile* has described the requirement of

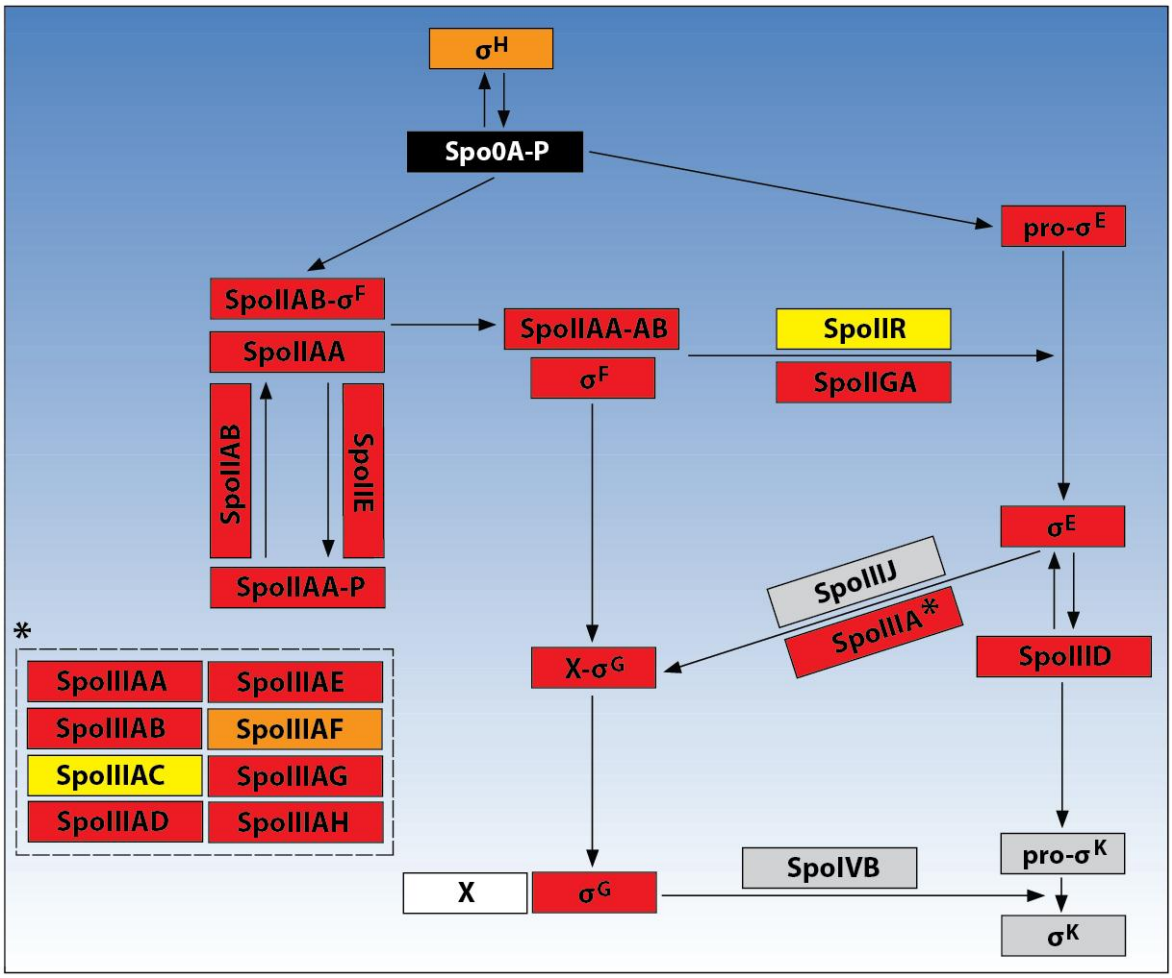

Figure 4.20. Schematic representation of the *C. difficile*-specific sporulation cascade, indicating genes belonging to the Spo0A regulon at the transcriptional level. Sporulation-associated genes belonging to the Spo0A regulon are colour coded as follows: red, positively regulated by Spo0A at $P = \leq 0.01$; orange, positively regulated by Spo0A at $P = \leq 0.05$; yellow, positively regulated by Spo0A at $P = \leq 0.08$); grey, not significantly regulated by Spo0A according to our parameters of significance; white, unknown factor. Generated from: (176, 249)

glycosylation in flagellar assembly and function (245). We found that Spo0A represses flagellum assembly and that *C. difficile* 630Δ*erm*Δ*spo0A* derivatives produce increased levels of FliC, FliD and genes encoding putative transglycosylases compared to parental *C. difficile* 630Δ*erm*, suggesting a possible link between Spo0A and flagellar assembly and glycosylation in *C. difficile* 630Δ*erm*.

It is possible that the enhanced flagellar phenotype of *C. difficile* 630Δ*erm*Δ*spo0A* may contribute to the inability of this strain to cause a long-term persistent infection in our murine model of infection (as discussed in Chapter 3). For example, it has been demonstrated that for *V. cholerae* to cause a persistent, long-term infection, the *loss* of motility is important (though flagella are important for the initial infection) (250, 251). Additionally, flagellin proteins are recognised by Toll-like receptor (TLR) 5, which acts as a receptor that can stimulate induction of host innate immunity (252, 253). Thus, the loss of flagella may be linked to a mechanism of immune subversion. Given that the role of flagella in *C. difficile* pathogenesis is currently unclear, we are attempting to make mutations in the *fliC* (flagellin subunit) and *fliD* (filament cap) genes of *C. difficile*. Using our mouse model of infection, we will then determine whether the repression of flagellar biosynthesis by *C. difficile* is required for long-term persistence, and whether this is associated with virulence *in vivo*.

The role of butyrate in human colonic health is well established (239). However, the role of butyrate production in the pathogenesis of human-virulent *C. difficile* is currently unknown. We noted that *C. difficile* 630Δ*erm*Δ*spo0A* produced significantly less butyrate than its parental strain, that is, Spo0A positively regulates butyrate production in 630Δ*erm.* Given the

role of butyrate in human colonic health, we hypothesise that the production of this SCFA by 630Δ*erm* may facilitate the establishment of a favourable environment to initiate an infection. For example, there is increasing evidence that butyrate has anti-inflammatory properties on the host (254, 255). Following colonisation, Spo0A may then function to increase the fermentation of glucose into SCFAs (such as butyrate) which concomitantly results in a de-repression of TcdA and TcdB production (glucose represses toxin synthesis). Clearly, however, this hypothesis would need to be explored further. In addition, we are currently profiling other SCFAs and alcohols from exponentially growing *C. difficile* to determine whether 630Δ*erm*Δ*spo0A* shifts its fermentative metabolism towards the production of other end-products (such as ethanol or acetate), whilst limiting its net carbon flow to butyrate. The generation of mutants in the *thlA1* and *hbd* genes of the butyrate biosynthesis operon is currently underway, and should facilitate such studies.

The degree of concordance between the Spo0A-regulated transcriptome and proteome of *C. difficile* 630Δ*erm* was low. Indeed, only a small subset of CDSs were found to be differentially regulated by Spo0A at both the transcriptional *and* translational levels. This disconnect is likely the result of an interplay between 2 major factors: (i) inherent biological variation in mRNA to protein correlation, and (ii) the technical limitations of our assay. Firstly, previous publications have reported that transcripts and proteins only demonstrate a weak positive correlation, and that the underlying principle that mRNA abundance accurately informs protein expression may not hold true in all biological systems (256, 257). Moreover, molecular events such as the efficiency of translation or protein degradation can notably alter protein levels independently of mRNA expression (236).

Secondly, the lack of correlation between the profiles may be explained somewhat by the technical limitations of our assay. For example, the excision of protein bands is a potential source of error prior to LC-MS/MS analysis. Thus, the use of a 2D gel electrophoresis approach may have increased the number of proteins recovered. In addition, the proteomic data presented in this thesis does not include complementary data on the proteins secreted by *C. difficile*. An attempt was made to catalogue the secreted proteins (secretome) of exponentially growing *C. difficile* 630Δ*erm* and 630Δ*erm*Δ*spo0A*, though this fraction was regrettably contaminated during LC- MS/MS analysis and was unavailable for comparative analyses. This may explain, for example, why TcdA (a secreted toxin) was significantly up-regulated as a transcript, but not as a protein according to LC-MS/MS analysis. Furthermore, given that secreted proteins can mediate important host - pathogen interactions *in vivo*, and may also represent novel candidate virulence factors, it would be advantageous to have a full proteomic (cellular *and* secreted) description of *C. difficile* during exponential growth.

In conclusion, we have demonstrated that Spo0A regulates multiple virulence, sporulation and metabolic phenotypes in *C. difficile* 630Δ*erm*, which informs on the intricate pleiotropy of a *spo0A* mutation. The phenotypes described in this chapter (butyrate and flagellar biosynthesis) may have important implications relating to *C. difficile* colonisation and persistence *in vivo*, and will inform the direction of future research. In addition, future experiments that involve blocking spore formation at a later stage in the sporulation cascade may reduce the number of pleiotropic effects in comparison to those observed for the *spo0A* mutation, which may further define the events that occur during the course of spore formation in *C. difficile*.