An intact S-layer is advantageous to *Clostridioides difficile* within the host.

Short Title: *C. difficile* requires the S-layer *in vivo*

Authors:

Michael J. Ormsby[*1,$], Filipa Vaz[*1,#] Joseph A. Kirk[2], Anna Barwinska-Sendra[3], Jennifer C. Hallam[1], Paola Lanzoni-Mangutchi[3±], John Cole[1], Roy R. Chaudhuri[2], Paula S. Salgado[3], Robert P. Fagan[2] and Gillian R Douce[1&]

Affiliations:

[1]Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary & Life Sciences, University of Glasgow, Scotland, UK.

[2]Molecular Microbiology, School of Biosciences, University of Sheffield, England, UK.

[3]Biosciences Institute, Faculty of Medical Sciences, Newcastle University, England, UK.

Current addresses: [$]Biological and Environmental Sciences, Faculty of Natural Sciences, University of Stirling, Stirling, UK, [#]Department of Immunology, University of Oslo and Oslo University Hospital, Oslo, Norway, [±]Université Grenoble Alpes, CNRS, CEA, IBS, Grenoble, France

@BugS_layers

[*]These authors contributed equally to the manuscript.

&Author responsible for submission

Abstract

Clostridioides difficile is responsible for substantial morbidity and mortality in antibiotically-treated, hospitalised, elderly patients, in which toxin production correlates with diarrhoeal disease. While the function of these toxins has been studied in detail, the contribution of other factors, including the paracrystalline surface layer (S-layer), to disease is less well known. Here, we highlight the essentiality of the S-layer *in vivo* by reporting the recovery of S-layer revertants, following infection with the S-layer-null strain, FM2.5. Sequencing of the *slp*A gene revealed either correction of the original point mutation or modification of the sequence upstream of the mutation, which restored the reading frame, and translation of *slpA*. Selection of these strains was rapid, with up to 90% of isolates identified as revertants 24 h post infection.

Two revertant isolates, RvA and RvB, showed modification of 3 and 13 amino acids respectively, compared to wild type sequence. Structural determination of SlpA from RvB revealed a different orientation of its domains, resulting in a reorganisation of the lattice assembly and changes in interacting interfaces which might result in functional differences. These revertants showed differing patterns of disease i*n vivo*; RvA causing equivalent severity to R20291 and RvB an attenuated FM2.5-like phenotype. Comparative RNA sequencing (RNA-Seq) analysis of *in vitro* grown isolates showed large changes in differentially expressed genes (DEGs)

between R20291 and FM2.5 namely in TcdA/TcdB expression, in transcripts associated with sporulation and those linked to cell wall integrity, which may account for attenuation observed *in vivo*. In comparison, smaller differences were observed between RvA/R20291, and RvB/FM2.5 respectively, which correlated with observed disease severity *in vivo*. Cumulatively, these data highlight that the S-layer plays a role in *C. difficile* disease.

Author Summary.

The S-layer of *C. difficile* is a paracrystalline array that covers the outer surface of the bacterial cell but its contribution to overall disease remains unclear. A previously described, spontaneous *slpA*-null mutant, FM2.5, with a point mutation in *slp*A offered an opportunity to study the role of the S-layer *in vivo*. Here, we confirm that this strain is less virulent *in vivo* despite effectively colonising the host and producing toxin. We also show *in vivo* selection for sequence modifications that restore *slp*A translation and produce an S-layer. While such modifications do not affect the overall 3D structure of individual SlpA (sub)domains, they can lead to altered orientation of the structural domains and subsequent S-layer assembly. Importantly, RNA-Seq analysis *in vitro* showed large differences in gene expression between FM2.5 and R20291. Detected differences in transcription of genes involved in toxin expression and sporulation suggests that the S-layer provides a selective survival advantage within the host, which contributes to disease severity.

67

68

69 **Keywords**

70 S-layer, *C. difficile*, virulence, SlpA mutant, revertants

Introduction

Clostridioides difficile is the most common cause of hospital acquired diarrhoea globally, with disease linked to disruption of the intestinal microbiota through antibiotic use (Smits et al. 2016). The virulence of *C. difficile* has widely been attributed to the production of two toxins; toxins A (TcdA, enterotoxin) and B (TcdB, cytotoxin), responsible for cytoskeletal modifications, epithelial damage, inflammation, and fluid loss (Braun et al. 1996; Chandrasekaran and Lacy 2017). A third toxin, the binary *C. difficile* toxin (CDT), expressed by only a subset of strains, has been linked to enhanced disease severity (Chandrasekaran and Lacy 2017). Consequently, *C. difficile* colitis has widely been considered as a toxin-mediated disease. However, the availability of tools to analyse gene expression and improved methods of mutagenesis (Cartman et al. 2012a), together with the availability of an accessible murine animal model (Chen et al. 2008a; Theriot et al. 2011), have offered new opportunities to identify other traits, both bacterial and host-associated, that impact disease severity (McDermott et al. 2017; Fletcher et al. 2021). The recent use of such approaches has provided clearer understanding of the metabolic flexibility of these organisms, the role of the microbiome in disease progression (Buffie et al. 2015; Fletcher et al. 2021; Girinathan et al. 2021) and established several bacterial factors that influence the host response (Maldarelli et al. 2014; Batah et al. 2017; Arato et al. 2019). Of particular interest in this context, is the role of the S-layer in disease. This paracrystalline protein array is the outermost layer of the *C. difficile*

cell envelope, with similar structures found in many bacteria and virtually all archaea (Fagan and Fairweather 2014).

The S-layer has been shown to perform multiple and vital roles including providing protection from environmental factors such as variations in pH, mechanical and osmotic stresses (Engelhardt and Peters 1998; Claus et al. 2002; Engelhardt 2007). *In vivo*, it is proposed to play a role in molecular sieving (Sleytr and Beveridge 1999) and ion trapping, protecting the organism from antimicrobial peptides and bacteriolytic enzymes produced in response to infection (Lortal et al. 1992; Kirk et al. 2017). The S-layer has also been shown to be a key target in bacteriophage predation (Callegari et al. 1998; Kirk et al. 2017; Royer et al. 2022).

In *C. difficile*, the main component of the S-layer is SlpA, which is post-translationally cleaved by a cell wall protein (CWP), Cwp84, into two functional S-layer proteins (SLPs), SLP_L and SLP_H (Kirby et al. 2009; Fagan and Fairweather 2014). The proteinaceous array is further decorated by other CWPs, which provide additional functionality (Fagan and Fairweather 2014). Assembly of the paracrystalline array relies on tiling of SLP_H triangular prisms on the cell wall, interlocked by SLP_L ridges facing the environment (Lanzoni-Mangutchi et al. 2022). Exposure of SLP_L to the environment is consistent with its high sequence variability observed between different *C. difficile* strains, with 13 different S-layer cassette types (SLCTs) identified to date (Dingle et al. 2013; Kirk et al. 2017). Strikingly, the lattice is very compact compared to other studied S-layers, which have pores of between 30 – 100 Å compared with only ~10 Å in diameter in *C. difficile* (Lanzoni-

Mangutchi et al. 2022). This tight packing correlates well with the hypothesis that S-layer acts as a molecular sieve (Sleytr and Beveridge 1999), as deletion of the most exposed regions of SLP_L results in a strain with increased sensitive to lysozyme, in comparison to the parent strain, R20291 (Lanzoni-Mangutchi et al. 2022).

In *C. difficile*, the S-layer has also been implicated in host cell adhesion (Merrigan et al., 2013), biofilm formation (Kirby et al. 2009; Dapa et al. 2013; Richards et al. 2018) and immunomodulation through cell signalling of the host response (Ausiello et al. 2006; Sakakibara et al. 2007; Sekot et al. 2011). SlpA has been shown to induce innate and adaptive immune responses through activation of TLR4 (Ryan et al., 2011). However, the role of the S-layer in *C. difficile* pathogenesis and in immune evasion remains poorly understood.

Previously, we reported the isolation and characterization of a spontaneous *C. difficile* strain lacking an S-layer, FM2.5 (Kirk et al. 2017). In initial studies using the Golden Syrian hamster as the infection model, FM2.5 caused no symptoms of disease, despite effectively colonising infected animals (Kirk et al. 2017). However, the acute sensitivity of hamsters to *C. difficile* toxins and lack of readily available immunological tools, limits their usefulness in studying the more nuanced facets of this infection. In contrast, mice are naturally less susceptible to CDI, requiring more extensive antibiotic treatment to suppress the flora, and higher challenge doses to achieve colonisation (Chen et al. 2008b; Winston et al. 2016). However, mice offer greater opportunities to determine the contributions of other virulence-associated traits on disease outcome, including long-term persistence associated with relapsing

disease (Best et al. 2012).

Here, we sought to elucidate the role of the S-layer as a major virulence determinant in a murine model of infection, to determine whether the loss of virulence observed in the hamster model is reciprocal in other hosts. Our results suggest that the S-layer offers a competitive colonisation advantage within the mouse intestine and is important for *in vivo* disease severity.

Results

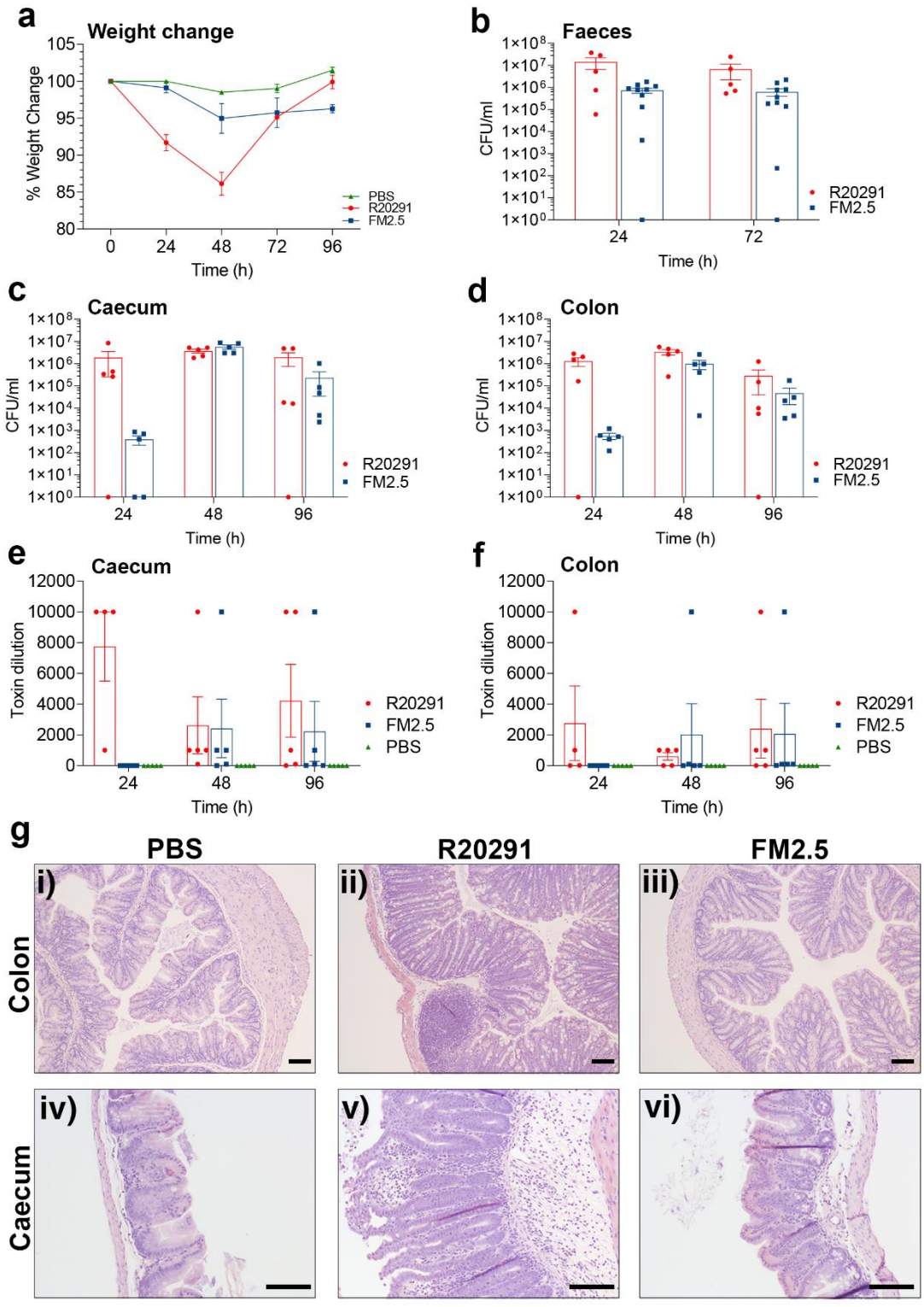
The S-layer contributes to severe disease in the murine model of *C. difficile*

In a murine model of infection, loss of body weight offers a strong correlative measure of *C. difficile* disease severity (Jukes et al., 2020). Infection of antibiotic pre-treated mice with strain R20291 resulted in significant weight loss of up to 15%; peaking between 24 and 48 h post-infection (hpi). In contrast, mice infected with the S-layer deficient derivative strain FM2.5 showed consistently less weight loss (average 6%; Fig. 1a). Measurement of total *C. difficile* in faecal material showed comparable levels of shedding at 24 and 72 hpi (Fig. 1b), while analysis of total *C. difficile* present in caecal (Fig. 1c) and colonic (Fig. 1d) material taken at post-mortem indicated a trend for less recovery of FM2.5 24 hpi. However, comparable levels of FM2.5 and R20291 were recovered from these tissues at 48 and 96 hpi. Analysis of *C. difficile* spores within the total faecal, caecal and colonic material showed a similar trend, with less FM2.5 spores recovered at 24 hpi, and comparable numbers of spores at 72 and 96 hpi (Fig. S1).

158 Interestingly, mice infected with R20291 that survived infection showed full
159 signs of recovery by 96 hpi, returning to pre-infection weights, equivalent to those
160 of non-infected mice. In contrast, FM2.5-infected mice failed to return to their pre-
161 infection weight even when animals were monitored for a further five days (9 days
162 pi; typical profile Fig. 4c) despite the animals remaining asymptomatic, with no
163 evidence of loose faeces.

164 Assessment of *in vivo* toxin production by both R20291 and FM2.5 in the
165 caecum (Fig. 1e) and colon (Fig. 1f) revealed that, at 24 hpi, less toxin was recovered
166 from mice infected with FM2.5. However, at 48 and 96 hpi, comparable levels of
167 toxin were recovered. Histological examination (at 48 hpi) showed that mice
168 infected with R20291 displayed significantly greater pathology than those infected
169 with FM2.5 and a PBS challenged control group (Fig. 1g i-vi. Cumulative scoring of
170 tissue damage is reported in Fig. S1).

171



172

173 **Fig 1. SlpA deficient *C. difficile* is less pathogenic in a murine model of infection.**

174 Female C57/Bl6 mice were challenged with spores of R20291 or FM2.5, or mock

175 infected with sterile PBS. (**a**) Weight loss was monitored every 24h for four

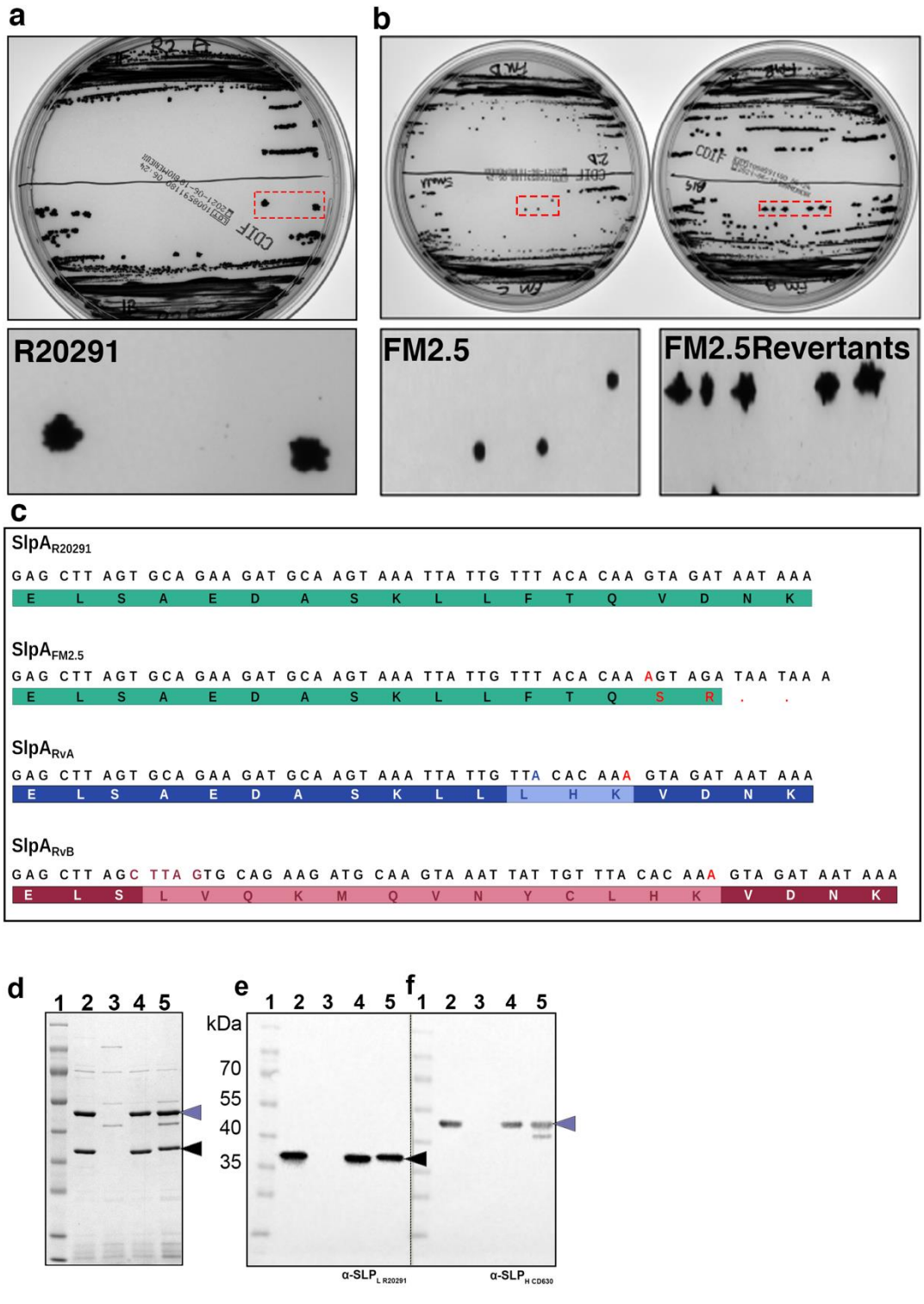
176 consecutive days following infection. Each point is the average of several replicate
177 experiments (n>3), with at least 5 animals per time point. (**b**) CFU/ml of faecal
178 material collected at 24 and 72 hpi. (**c**) CFU/ml of caecal content at 24, 48 and 96
179 hpi. (**d**) CFU/ml of colon content at 24, 48 and 96 hpi. (**e**) Toxin activity of caecal
180 content and (**f**) colonic content was determined at 24, 48 and 96 hpi; through
181 challenge of Vero cells *in vitro*. Results displayed indicate the reciprocal of lowest
182 dilution at which toxin activity could be measured. (**g**) Histopathological sections
183 representing colon (i, ii and iii) and caecal (iv, v and vi) sections following challenge
184 with PBS (i and iv); R20291 (ii and v); or FM2.5 (iii and vi). Scale bars represent 100
185 μm. Results displayed are the mean ± SEM of at least three independent biological
186 replicates. Statistical tests were conducted using GraphPad Prism software v.12.
187 Statistical tests include one-way ANOVA with Tukeys post-test; or a student's t-test
188 with Welch's correction. Statistical significance is indicated: ns – not significant; *p
189 < 0.05; **p < 0.01; and ***p < 0.001.
190
191 **_In vivo_ pressure drives selection for S-layer revertants**
192 When grown on selective chromogenic agar (ChromID®; BioMerieux), the
193 morphology of R20291 presents with the typical 'fried egg' *C. difficile* colony,
194 visible after approximately 16 h of incubation (Fig. 2a). In contrast, FM2.5 produces
195 smaller and smoother colonies, which take ~24 h to emerge (Fig. 2b). Following
196 infection of mice with R20291 and FM2.5, faecal material was recovered and plated
197 daily. During examination of resultant colonies, it was noted that, while colonies

198 from R20291 infected mice showed the expected morphology, material retrieved
199 from FM2.5-infected mice showed a mixture of both large ($FM2.5_{large}$) and small
200 colony types ($FM2.5_{small}$) (Fig. 2b). Additionally, $FM2.5_{large,}$ were countable after 16
201 h incubation, while the expected FM2.5-like colonies, $FM2.5_{small}$, were only
202 observable from 24 h. Several colonies from both types, $FM2.5_{small}$ and $FM2.5_{large}$,
203 were streaked from the original plates and were sub-cultured twice to ensure
204 clonality. Individual clones were then stored at -80°C.

205 Amplification of the *slpA* sequence from these clones revealed that the
206 $FM2.5_{large}$ colonies contained modifications in the genomic sequence upstream of the
207 FM2.5 mutation site (single nucleotide insertion, Fig. 2c). Several sequence variants
208 were identified, with two most common mutants named Revertant A (RvA) and
209 Revertant B (RvB). In RvA, a single nucleotide deletion (246delT) restored the
210 original reading frame, rescuing translation of the full SlpA*;* modifying three amino
211 acid residues in the translated protein. In, RvB, an insertion of five-nucleotides
212 (249_253insCTTAG), which again restored the reading frame and resulted in
213 modification of 13 amino acids within the mature protein (Fig 2c Interestingly,
214 revertants were identified in several *in vivo* experiments, using batches of
215 independently prepared spores.

216 To confirm expression of SlpA, low pH cell surface extracts of strains RvA
217 and RvB were analysed. SDS-PAGE showed that both SLP_H and SLP_L proteins were
218 present in R20291, absent in FM2.5 but restored in RvA and RvB (Fig. 2d). This was
219 confirmed through western immunoblot analysis using anti-SLP_H and anti-SLP_L

220 antibodies (Fig. 2e and f).



221

222 **Fig. 2 Recovery of S-layer revertants following *in vivo* challenge.**

223 Following challenge with spores of R20291 and FM2.5, faecal material was

recovered and plated on *C. difficile* selective chromogenic agar (Biomerieux). (**a**) Colonies of R20291. The dashed red boxed area is enlarged and shown below. (**b**) The two colony types of FM2.5. The dashed red boxed area is enlarged and shown below. $FM2.5_{large}$ is representative of the large colony phenotype recovered, while $FM2.5_{small}$ is indicative of the typical FM2.5 colony morphology. (**c**) Sequencing of a region of *slpA* shows the sequence of R20291; the insertion (red) in FM2.5 responsible for the truncation of the SlpA protein; a single nucleotide deletion (blue) in the FM2.5 sequence, resulting in RvA; a five-nucleotide insertion (crimson) in FM2.5 sequence resulting in RvB. (**d**) SDS-PAGE analysis of cell wall proteins extracted by low pH preparation. Lane 1: MW Marker; Lane 2: R20291; Lane 3: FM2.5; Lane 4: RvA; Lane 5: RvB. (**e**) Western immunoblot analysis using an anti-SLP_H antibody. (**f**) Western immunoblot analysis using an anti-SLP_L antibody.

Reversion can affect SlpA structure and assembly

To understand the effects of the detected reversions on SlpA structure and S-layer assembly, crystallisation of $SlpA_{R20291}$, $SlpA_{RvA}$ and $SlpA_{RvB}$ was carried out. Although crystals were obtained for all three variants, only $SlpA_{RvB}$ crystals were of sufficient quality for x-ray diffraction data collection and structural determination by molecular replacement, using previous SlpA structures as models, including a variant of $SlpA_{R20291}$ lacking the most exposed region of SLP_L - $SlpA_{R\Delta D2}$ - (PDB ID: 7ACZ) (Lanzoni-Mangutchi et al. 2022). SLP_H and the interacting domains were easily traceable in the electron density but D1 was only partially built, whilst

density for domain D2 was very poor and this region could not be traced in the final $SlpA_{RvB}$ model (Fig. 3a, PDB ID: 8BBY, Table S1). This implies that D2 is flexible and/or unstructured, while the structure of the core domains required for S-layer assembly - SLP_H, and, to a lesser extent, D1 and LID/HID (Lanzoni-Mangutchi et al. 2022) - seems to be generally maintained. However, the relative orientation of these domains in the SlpA molecule is altered (Fig. 3a), with D1 and the interacting domains rotated towards the SLP_H plane by ~30 ° (Fig. 3a). The 13 altered residues in $\alpha2_L$ in $SlpA_{RvB}$ result in disruption of the α-helix secondary structure and introduce disorder in the upstream loop that links the preceding β-strand ($\beta3_L$) and $\alpha2_L$. It is worth noting that SLP_H in R20291, which belongs to SLCT 4, has several insertions within the cell wall binding 2 (CWB2) sequence motifs that define CWPs in *C. difficile*, when compared to other SlpA types. These insertions could not be traced in our $SlpA_{R\Delta D2}$ model (Lanzoni-Mangutchi et al. 2022) but were traceable in the $SlpA_{RvB}$ and result in several loops protruding above the SLP_H plane, towards the environment, partially occluding the CWB2 motifs (Fig. 3a, right). Together with the movement of the interacting domains and D1 towards the SLP_H tiles, this creates a more compressed arrangement (~66 Å compared to ~76 Å in $SlpA_{CD630}$, PDB ID: 7ACY, Fig. 3b, bottom).

In the crystallographic models of $SlpA_{CD630}$ (PDB ID: 7ACY), $SlpA_{R7404}$ (PDB ID: 7ACX) and $SlpA_{R\Delta D2}$ (PDB ID: 7ACZ), $\alpha2_L$ was responsible for closing a gap between neighbouring SlpA molecules via D1-D1 interactions (Lanzoni-Mangutchi et al. 2022). In the crystallographic model of the R20291-derived $SlpA_{R\Delta D2}$ variant,

D1-D1 interactions are mediated by hydrogen bonds between $S50_L$-$S50_L$ and $Q70_L$-$A49_L$ from neighbouring molecules.

In the $SlpA_{RvB}$ structure, disruption of $\alpha2_L$ and reorientation of D1 and LID/HID relative to SLP_H leads to changes in the interactions between neighbouring molecules and, consequently, a rearrangement of the S-layer array. Strikingly, the D1-D1 interactions seen in previous models were not observed here, possibly due to the flexibility of $\alpha2_L$ and preceding loop caused by the changes in the reverted sequence leading to a different orientation of D1 domains. Unlike in the previously determined structures, neighbouring D1 domains in the $SlpA_{RvB}$ structure are too far apart to mediate contacts (> 12 Å). In the previous models, SLP_H tiling creates two wide channels, which are stabilised by interactions mediated by the interacting domains and D1 (Lanzoni-Manguthi, 2022). A different mode of stabilising the SLP_H tiling is observed in $SlpA_{RvB}$, with the interacting domains now partially inserted in those cavities (Fig. 3b, Fig. S2a). A new interacting interface between HID from one molecule and $CWB2_3$ occludes these gaps and stabilises the S-layer lattice (Fig. 3c, Fig. S2a). This new arrangement of the crystal lattice is in line with our proposed assembly model, where the S-layer 2D array is maintained mostly by hydrogen bonds and salt bridges across surfaces with complementary charges (Lanzoni-Mangutchi et al., 2022), largely dependent on SLP_H-SLP_H interactions and stabilised by varying degree of interactions involving SLP_L (Fig. 3c and Table S2). The structural model of RvB confirms that changes in SLP_L can be accommodated with minor structural changes to the (sub)domains, by exploring flexible loops and hinges

to provide a stable S-layer.

As no crystal data was obtainable, we also calculated models for $SlpA_{R2021}$ and $SlpA_{RvA}$ using the SWISS-MODEL server, based on previous models (Lanzoni-Mangutchi et al. 2022) and the $SlpA_{RvB}$ structure determined here. Depending on which template was used ($SlpA_{R\Delta D2}$ or $SlpA_{RvB}$), different predicted structures of $SlpA_{RvA}$ were obtained, varying mostly in the orientation of D1 and interacting domains relative to SLP_H (Fig. S2b). Interestingly, one common feature was that the changes resulting from the revertant sequence seem to be accommodated not by altering the α-helix but by varying the length of the upstream loop that links the preceding β-strand ($\beta3_L$) and $\alpha2_L$ (Fig. S2b). It is therefore unclear if $SlpA_{RvA}$ is more likely to adopt a R20291-like as observed in the $SlpA_{R\Delta D2}$ model or RvB-like S-layer assembly, as both can accommodate the modified sequence.

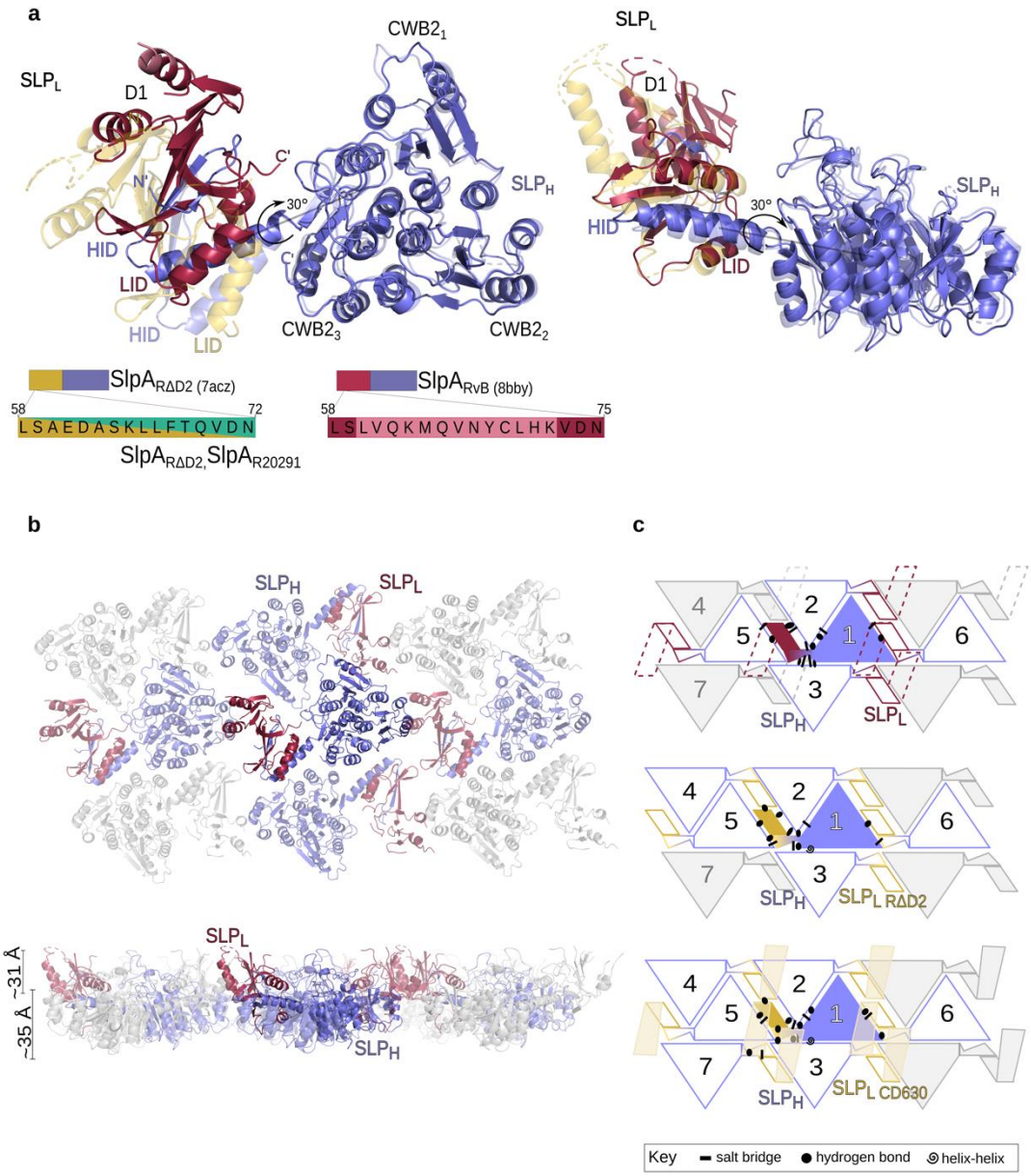


Fig. 3 Structure of SlpA$_{RvB}$ shows a different assembly arrangement

(**a**) Structural model of SlpA$_{RvB}$ (SLP$_L$ – pale red, SLP$_H$ – slate blue, PDB ID: 8BBY), superimposed on SlpA$_{RΔD2}$ model (SLP$_L$ – gold, SLP$_H$ – slate blue, semi-transparent), with the rotation angle of the D1 and LID/HID domains shown by an arrow. Three distinct structural features are observed: SLP$_H$, LID/HID and D1. Cartoon

representation of the SLP_H/SLP_L (H/L) complex, as seen from the environmental side (left) and side view (right). Sequence of $\alpha2_L$, with paler colours indicating differences, is shown schematically. (**b**) Cartoon representation of the H/L planar array (PDB ID 8BBY, interacting molecules coloured and viewed as in **a**). (**c**) 2D schematic of H/L complex crystal packing in $SlpA_{RvB}$ (top), $SlpA_{R\Delta D2}$ (centre) and $SlpA_{CD630}$ (bottom), indicating the interaction network linking a single H/L (slate blue/crimson or slate/blue/gold) complex with neighbouring molecules in a planar arrangement generated by SLP_H tiling. The missing D2 in the $SlpA_{RvB}$ model is represented as dashed lines. Notably, D1-D1 interactions seen in other models are missing in RvB and the Slp_H tiles are shifted, with new HID-$CWB2_3$ interactions stabilising the lattice. Array is depicted as seen from the extracellular environment, with symbols representing key interaction types in the crystal lattice, detailed in Table S2.

In vivo S-layer selection is independent of toxin expression

FM2.5 has previously been observed to show a delay in toxin production (Kirk et al. 2017), consequently we chose to investigate whether *slpA* reversion was accompanied by a potential restoration of toxin production *in vivo*. Mice were infected with FM2.5*ΔPaLoc*, in which the **Pa**thogenicity **Loc**us (PaLoc), encoding toxins A and B, had been deleted. In contrast to mice infected with FM2.5, animals challenged with FM2.5*ΔPaLoc* showed no weight loss over the 96 hours of infection (Fig. 4a) and as expected, no toxin was observed in the caecum or colon (Fig. S3c).

Total bacteria (Fig. S3a) and spores (Fig. S3b) recovered from the caecum and colon of these mice were comparable at 96 hpi to that observed in animals infected with FM2.5.

Interestingly, S-layer revertants were also recovered from these mice, as identified by sequence modifications in the same region of *slpA* compared to the R20291 sequence (RvD, Fig. 4a). These changes also facilitated restoration of an intact S-layer, indicating that any potential selection advantage is independent of toxin production.

Reproducible recovery of S-layer variants *in vivo* raised the possibility that low numbers of genetic variants exist within the FM2.5 population, which are amplified by the *in vivo* environment. To test this hypothesis, we undertook amplicon sequencing of *slpA* in the spore preparations used for mouse infections, and in bacteria recovered from faecal material from mice infected with FM2.5 at 24, 48, 72 and 96 hpi (Fig. 4c). This analysis revealed that revertants in which the original frameshift mutation found in FM2.5 was corrected by deletion of the extra nucleotide (252delA, RvC) were present in the initial spore preparation, albeit as a low proportion of the population (<5%, Fig. 4d). Isolation of revertants as early as 24h (>94% of the population) suggests that expressing an intact S-layer provides a competitive advantage *in vivo* over the S-layer deficient strain.

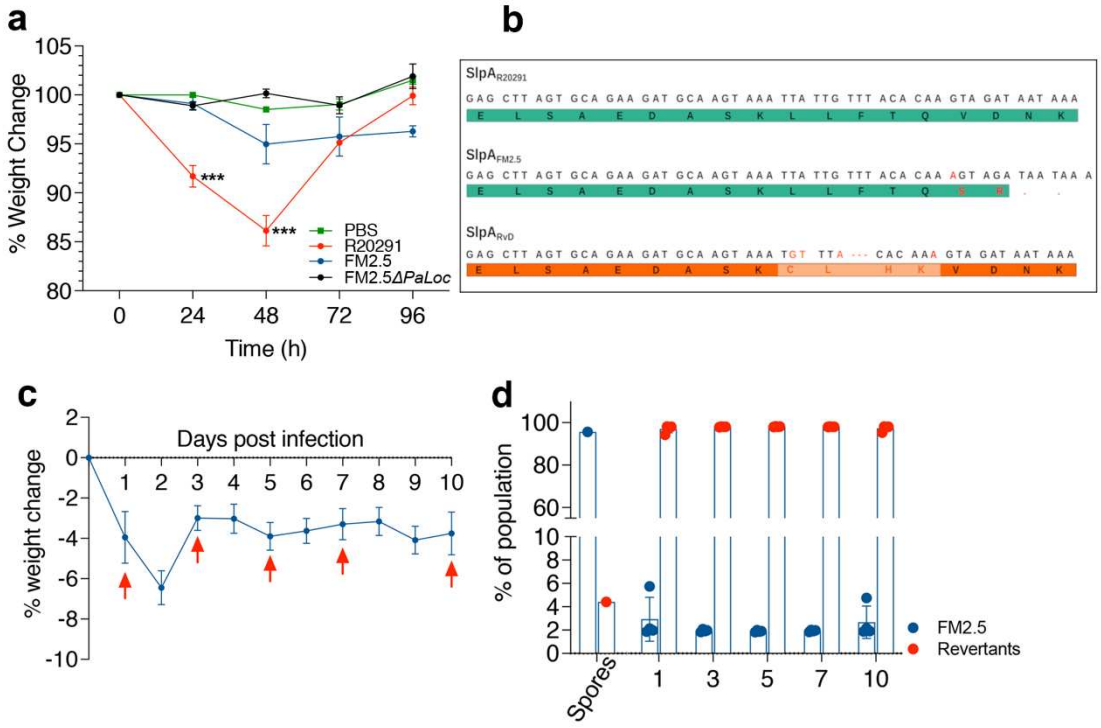


Fig 4. ***In vivo*** **challenge of mice with FM2.5*ΔPaLoc*.**

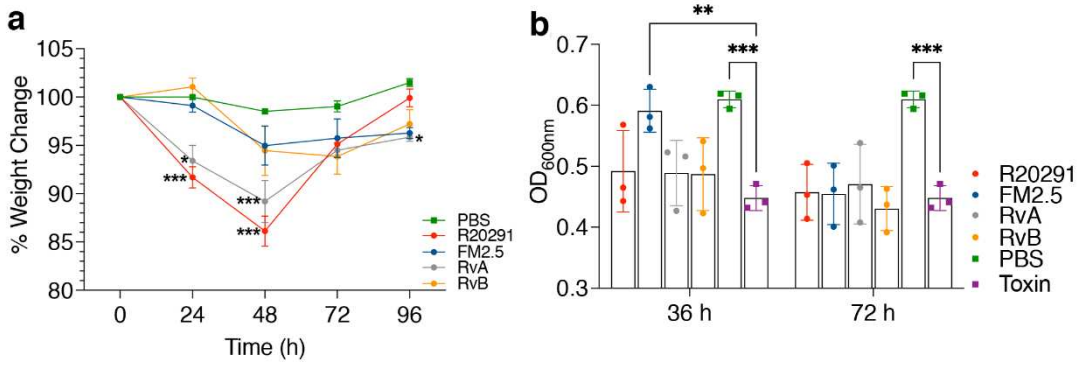
Female C57/Bl6 mice were challenged with spores of R20291, FM2.5 and FM2.5*ΔPaLoc*. (**a**) Weight loss was monitored every 24h for four consecutive days following infection. Each point is the average of multiple mice, where n≥ 5. (**b**) Sequencing of a region of *slpA* shows the sequence of R20291; the insertion in FM2.5 (red) responsible for the truncation of the SlpA protein; with an additional revertant named RvD, which shows a complex array of sequence insertions and deletions (orange) in this region of the *slpA*. (**c**) Time course showing infection and weight loss in nine mice infected with FM2.5. Arrows indicate times of faecal sample collection used in amplicon sequencing of the variable region of *slpA*. (**d**) Description of relative proportion of FM2.5 sequences in samples analyzed prior to and post infection.

363

364 **The revertant strains display differing levels of virulence in mice**

365 To assess whether recovery of SlpA by revertants correlated with rescued
366 virulence, mice were infected with spore preparations of RvA and RvB, alongside
367 R20291 and FM2.5 (Fig. 5a). Interestingly, infection with RvA resulted in
368 significant weight loss within the first 48h of infection, which was similar to mice
369 infected with R20291. RvB, in contrast, showed a similar limited pattern of weight
370 loss to animals infected with FM2.5, which stabilized from 48 hpi.

371 To determine whether these differences were associated with changes in
372 toxin production, the revertants were cultured *in vitro* and filtered spent growth
373 medium from 36 and 72h growth was used to determine the level of toxin B activity.
374 Both RvA and RvB produced comparable levels of toxin to R20291 at these time
375 points (Fig. 5b), with toxin-mediated damage to Vero cells, resulting in cell
376 rounding, cellular loss and reduced levels of staining with Giemsa. In agreement
377 with previous reports (Kirk et al. 2017), FM2.5 produced less toxin than R20291 at
378 36 h, although toxin production levels were comparable in all strains by 72 h.



379

380 **Fig 5. Functional analysis of RvA and RvB *in vivo* and *in vitro*.**

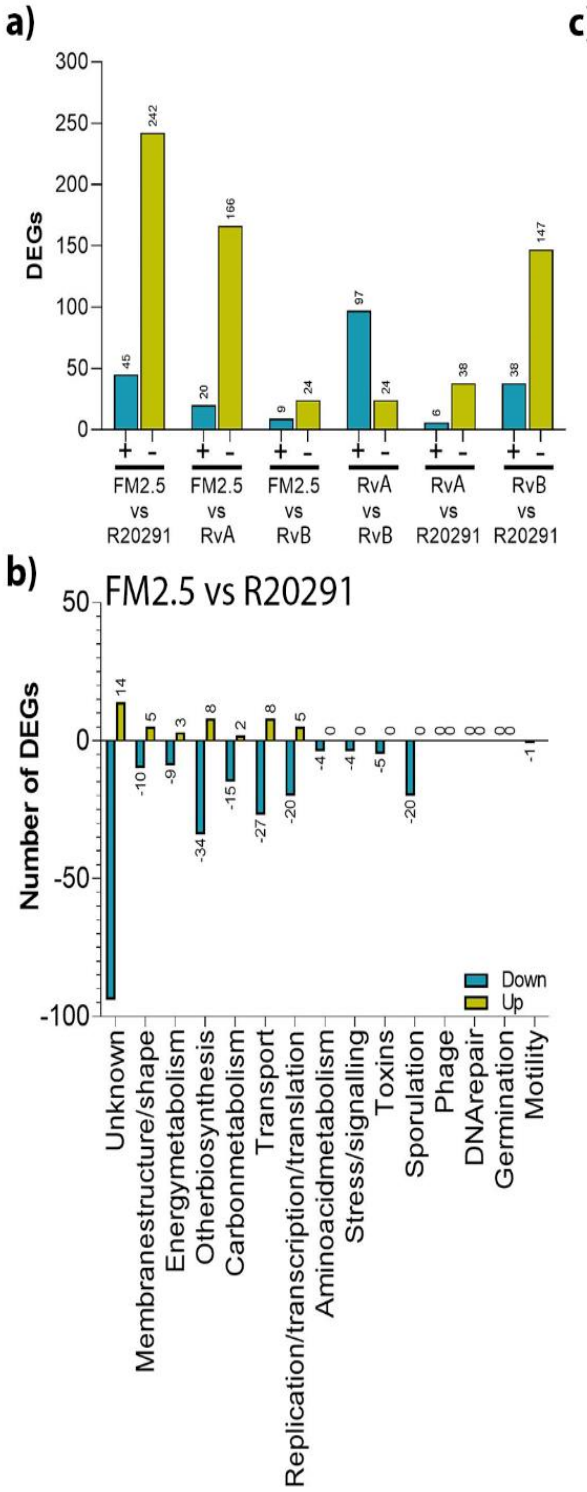
(**a**) Female C57/Bl6 mice were challenged with spores of R20291, FM2.5, RvA, RvB or mock infected with sterile PBS. Weight loss was monitored at the same timepoint each day for four consecutive days following infection. Each point is the average of multiple individuals, with at least 5 animals. (**b**) *In vitro* toxin activity as measured through challenge of Vero cells. Samples were prepared by filtering supernatant following *C. difficile* growth for 36 or 72 h and activity was measured through challenge of Vero cells. Supernatants were harvested at the same phase of growth for each strain. OD_{600} represents the optical density of Giemsa stain incorporated and released from intact Vero cells, hence high OD represents limited toxicity. Results displayed are the mean ± SEM of at least three independent replicates. Statistical tests were conducted using GraphPad Prism software v.12. Statistical tests include one-way ANOVA with Tukeys post-test; or a student's t-test with Welch's correction. Statistical significance is indicated: ns – not significant; $^{*}p < 0.05$; $^{**}p < 0.01$; and $^{***}p < 0.001$.

Modification of the S-layer results in large changes in gene expression

To gain a greater understanding of the differences in gene expression between R20291, FM2.5 and revertant strains, comparative RNA-Seq analysis was conducted following *in vitro* growth. Analysis revealed differences in gene expression between R20291 and FM2.5, with over 287 differentially expressed genes (DEGs) (Fig. 6a), linked to alterations in metabolism, transport, membrane integrity and sporulation (Fig. 6b). In contrast, less differences were observed when R20291

was compared to RvA (44 DEGs), than RvB (185 DEGs), which showed similar numbers of DEGs to FM2.5. This correlates well with the observed behaviour of these strains within animals (RvA associated with WT-like disease and RvB with FM2.5-like attenuation). Analysis of these data suggest that differences in observed disease severity could be linked to changes in transcription of several virulence-associated traits, including toxin A and B, and genes associated with sporulation. While the recovery of an intact S-layer would appear sufficient in the case of RvA, to restore wild type gene transcription, only partial transcription profile restoration, including toxin expression, is observed in RvB. However, as toxic activity was observed at 36h in culture, the alterations in transcription control seen in FM2.5 and RvB would appear to be limited to timing rather than absolute prevention of toxin production.

Taken together, these data suggest that expression of the S-layer plays a key role in *C. difficile* disease within the host.



c)

	Product	Gene ID	Log2 fold-change in FM2.5 vs. R20291	RvA	RvB
Toxin	CdtA	RS13505	-4.00	-4.22	-4.13
	CdtB	RS13510	-3.77	-3.93	-3.91
	TcdA	RS03460	-2.24	-2.97	-
	TcdB	RS03440	-4.14	-5.53	-
	TcdC	RS03465	-1.65	-1.93	-
	TcdE	RS03445	-	-2.21	-
Sporulation	BclA2	RS16650	-	-2.05	-
	BclA3	RS17180	-3.06	-4.89	-
	CdeC	RS05225	-4.66	-6.34	-
	CdeM	RS08125	-5.63	-7.84	-6.96
	CotA	RS08300	-3.15	-4.92	-
	CotJA	RS03120	-5.02	-6.34	-
	CotJA	RS12470	-4.72	-5.65	-
	CotJB	RS03125	-4.9	-6.46	-
	CotJB	RS12475	-4.6	-5.82	-
	SleC	RS02870	-6.24	-7.16	-
	SpoIIIAH-like family protein	RS05790	-3.46	-4.22	-
	Spore coat protein	RS01475	-	-3.26	-
	SrtB-anchored collagen-binding adhesin	RS14700	-	-1.9	-
	StaF	RS11245	-	-3.53	-
	Stage II sporulation protein P	RS12845	-	-2.39	-
	Stage II sporulation protein R	RS18290	-2.69	-2.62	-
	Stage III sporulation protein AA	RS05755	-3.64	-3.87	-
	Stage III sporulation protein AB	RS05760	-3.85	-3.82	-
	Stage III sporulation protein AE	RS05775	-	-2.18	-
	Stage III sporulation protein AF	RS05780	-2.13	-2.14	-
	Stage III sporulation protein AG	RS05785	-3.83	-4.29	-
	Stage IV sporulation protein A	RS13615	-3.6	-4.11	-
	Stage V sporulation protein	RS04075	-2.49	-	-
	Stage V sporulation protein AC	RS04070	-	-3.17	-
	Stage V sporulation protein AE	RS04080	-	-2.38	-
	Stage V sporulation protein T	RS17945	-3.58	-3.33	-
	TauE/SafE family	RS04170	-4.53	-	-
	TauE/SafE family	RS04175	-4.18	-	-
Cell structure	3D domain-containing protein	RS08095	-5.16	-4.23	-
	3D domain-containing protein	RS12905	3.36	3.29	-
	4Fe-4S binding protein	RS03270	-	-5.44	-
	Amidase domain-containing protein	RS06595	-	-3.77	-
	CD1845 family protein	RS09490	-	-4.73	-
	Cell wall hydrolase	RS18285	-2.78	-	-
	Cell wall-binding repeat-containing protein	RS11495	2.04	1.5	-
	C-GCAxxG-C-C family protein	RS09215	-	-5.79	-
	Cwp19	RS14350	2.1	1.99	-
	Cwp25	RS04420	-1.24	-1.09	-
	Cwp5	RS14445	-1.37	-	-
	Cwp9	RS14505	2.2	1.89	-
	CwpV	RS02685	4.27	5.44	5.85
	Delta-lactam-biosynthetic de-N-acetylase	RS07010	-	-5.85	-
	DMT family transporter	RS06925	-	-2.94	-
	Membrane protein	RS05630	-	-3.3	-
	OmpA family protein	RS10165	-3.2	-2.96	-
	OmpA family protein	RS03920	-1.14	-	-
	Rod shape-determining protein	RS01045	-4.88	-4.84	-
	Trypsin-like peptidase domain-containing protein	RS16935	-	-	-1.78
	VanW family protein	RS11265	-2.52	-1.86	-
	YhgE/Pip domain-containing protein	RS07215	-1.73	-1.88	-
	YlEGIA family protein	RS13665	-2.61	-2.64	-
	YtxH domain-containing protein	RS18635	-	-1.52	-

417

418 **Fig. 6 Global transcriptional differences between isolates of *C. difficile* following *in***

419 ***vitro* growth.**

420 Analysis of mRNA recovered from *in vitro* grown cultures of *C. difficile* isolates
421 R20291, FM2.5, RvA and RvB. (**a**) Total number of differentially expressed genes
422 (DEGs) between experimental groups are highlighted in blue (upregulated) and gold
423 (downregulated). (**b**) DEGs from experimental comparison of R20291 and FM2.5
424 were categorised based on function. (**c**) Transcriptional differences in select genes
425 of FM2.5 relative to experimental comparison with R20291, RvA and RvB.
426

Discussion

The S-layer of *C. difficile* has long been considered as integral to its physiology and pathogenesis, with several roles reported, including adherence to the epithelial barrier (Merrigan et al., 2013), immune cell signalling (Ryan et al., 2011; Chen et al., 2020) , sensitivity to antimicrobial peptides (Kirk et al. 2017) and sporulation efficiency (Kirk et al. 2017). Here, we describe the pathogenesis of the S-layer-null mutant FM2.5 within the mouse model of disease and report the recovery of toxin-independent, spontaneous S-layer variants in which SlpA expression is restored. This unexpected but reproducible phenomenon supports the growing evidence that this structure plays a key role in adaptation and survival within the host.

FM2.5, a strain originally selected through its resistance to the R-type bacteriocin Av-CD291.2, was previously reported to be attenuated in the Syrian hamster model of *C. difficile* (Kirk et al. 2017). These studies indicated that SlpA was essential for disease, with no diarrhoeal symptoms observed in infected animals, despite the recovery of FM2.5 from the caecum and colon of infected hamsters 14 days pi. While here we confirmed that the attenuated phenotype was reproducible in mice, we also detected SlpA revertant clones from infected animals. Interestingly, SlpA revertants were not observed during hamster infections, despite using the same chromogenic agar for recovery of FM2.5 isolates from infected animals.

Isolation of FM2.5 SlpA revertants in mice but not in hamsters suggests that differences in the local environmental conditions between the hamster and mouse may influence the amplification and outgrowth of these S-layer variant strains.

449 Indeed, it has been suggested that *C. difficile* colonisation efficiency may reflect
450 variation in expression of cathelicidins (such as LL-37) within these hosts (Woods
451 et al. 2018), which would help to explain why revertants were not amplified in the
452 hamster gut. This observation also lends support to the hypothesis that an intact S-
453 layer confers resistance to the antimicrobial activity of enzymes (such as lysozyme)
454 and antimicrobial peptides (LL-37) (Kirk et al. 2017). The observation that FM2.5
455 is acutely sensitive to LL-37 and lysozyme (Kirk et al. 2017) and becomes resistant
456 following S-layer restoration (Lanzoni-Mangutchi, 2022) further supports the
457 premise that revertants with an intact S-layer have a competitive advantage *in vivo*.
458 Selection and amplification of the revertants in this study further highlights the
459 competitive advantage offered by the S-layer in *C. difficile* intestinal survival. Poor
460 recovery of *C. difficile* at 24 hpi in FM2.5-infected mice could also be linked to
461 lower or less efficient rates of germination by FM2.5 compared to R20291. However,
462 *in vitro* studies using taurocholic acid as a germinant, indicate that these strains
463 show equivalent rates of germination (Kirk et al. 2017). Alternatively, lower rates
464 of FM2.5 at 24h hours could be a result of an increased susceptibility of the SlpA-
465 null mutant to anti-microbial peptides within the gut. The equivalent recovered
466 numbers at 48 and 72h for infections with FM2.5 when compared to R20291 may
467 correspond to an increased population of revertant clones, considering the high
468 level of reversion, as quantified by genetic analysis.
469 Although SlpA variants have not been observed during sequential growth of
470 the FM2.5 isolate *in vitro*, several factors may influence their presence, albeit at low

numbers, within the inoculum used to infect mice. As a largely obligate anaerobe, the sensitivity of *C. difficile* vegetative cells to oxygen complicates their use in animal models of infection. In contrast, the preparation and use of spores correlates with the natural infection and avoids co-administration of toxins expressed during *in vitro* growth. However, as FM2.5 has a known reduced sporulation efficiency (Kirk et al. 2017), it is possible that any variants expressing an intact S-layer present would sporulate more efficiently and therefore represent a higher proportion of the inoculum used for infection. Sequence analysis of *slpA* of several different batches of FM2.5 spores supported this hypothesis, with a small (>2.5%) proportion of the population displaying variations upstream of the original *slpA* mutation.

A delay in toxin production by FM2.5 has been reported previously (Kirk et al. 2017) which, coupled with the apparent delayed growth *in vivo*, could account for the difference in weight loss between R20291 and FM2.5 infected mice after 24 hpi. Using our FM2.5*ΔPaLoc* strain, we were able to demonstrate that the weight loss is entirely dependent on toxin production. However, this raises the question as to why animals infected with FM2.5, that show high and equivalent levels of tissue colonisation by 48 hpi, do not show equivalent levels of weight loss and tissue inflammation as R20291 infected animals. Several studies have shown that the S-layer is essential in immune activation (Jarchum et al. 2012; Mamareli et al. 2019), driving the production of proinflammatory cytokines via TLR4/MyD88 dependent pathways and enhancing the toxin-activated inflammasome (Ryan et al., 2011; Cowardin et al., 2015; McDermott et al., 2016). Together, this implies that the

493 timing or spatial localisation of toxins and the S-layer relative to the epithelial
494 barrier could be crucial to immune activation. This is further supported by the
495 observation that RvB showed an equivalent reduction in toxin expression to FM2.5
496 in the RNA-Seq analysis and a reduction in disease severity in the mouse.

497 Alternatively, as a feedback mechanism between sporulation and the complex
498 regulatory network controlling toxin production has been proposed (Deakin et al.
499 2012), it is possible that a defect in FM2.5 sporulation may result in a delay in toxin
500 production. The observed reduction in toxin activity at 36h in the filtered culture
501 supernatant from FM2.5 correlated well with the log2 fold reduction in *tcdB*
502 transcripts (-4.14) and *tcdA* (-2.24) in mRNA recovered from FM2.5 cultures grown
503 for 6h, when compared to transcripts recovered from R20291 cultures, at the
504 equivalent timepoint. In contrast, RvB which also presented a reduction in toxin
505 gene expression compared to R20291 by RNA-Seq, demonstrated equivalent toxin
506 functional activity to R20291 when grown *in vitro* for 36h; RvA and R20291 showed
507 comparable toxin expression in both systems. This suggests that whilst
508 modifications in S-layer can result in delays to toxin production, these changes do
509 not prevent or limit final activity supporting the hypothesis that disease severity
510 might be linked to the timing and co-ordination of S-layer and toxin by the host.

511 Mice infected with either RvA or RvB showed different disease severity, as
512 indicated by differences in weight loss. The low virulence of an RvB infection may
513 be, in part, explained by the structural differences. Indeed, structural analysis of
514 $SlpA_{RvB}$ revealed a new packing of SlpA molecules in the array, with a

515 rearrangement of the position of SLP_H and both interacting domains, now involved
516 in tiling of the S-layer (Fig. 3). This suggests a considerable degree of adaptability of
517 both SLP_L, where the reversion is located, and SLP_H, to accommodate varying
518 interactions between neighbouring molecules. The absence of density to model D2
519 in $SlpA_{RvB}$ further illustrates that this domain is dispensable for S-layer assembly, as
520 previously reported (Lanzoni-Mangutchi et al., 2022). As S-layer assembly is
521 maintained mostly by hydrogen bonds and salt bridges (Lanzoni-Mangutchi et al.,
522 2022), rearrangement of the subdomains to create structures with complementary
523 surface charges seems to enable the different assemblies observed so far. These
524 changes in quaternary structure, with minor changes of secondary and tertiary
525 structure of the subdomains, suggests that the ability to form a paracrystalline array
526 is central to S-layer function and can be achieved in different arrangements. The S-
527 layer must retain a certain degree of flexibility, not only to account for the cell pole
528 curvature and allow cell division, but also for incorporation of minor cell wall
529 proteins that enhance functionality. The presence of a more intricate network of
530 interactions and more extensive interface areas when compared to the R20291-
531 related $SlpA_{R\Delta D2}$ structure (Table S2) between neighbouring molecules seen in the
532 $SlpA_{RvB}$ structure suggests a potentially less flexible paracrystalline array, with less
533 ability to incorporate specific functions of minor CWPs, which may help to explain
534 differences in disease patterns observed between RvB and the other revertant strain,
535 RvA; the structure of which needs to be studied in more detail. Further structural
536 studies of $SlpA_{RvA}$ revertant and $SlpA_{R20291}$ as well as detailed analysis of S-layer

assembly and composition, including the capacity to incorporate other minor cell wall proteins, will help elucidate the role of specific aspects of the S-layer.

While identification of SlpA revertants was unexpected and adds complexity to the interpretation of the data from the mouse disease model, the rapid recovery of these strains highlights the key contribution that the expression of an intact S-layer offers to *C. difficile* infection *in vivo*. This work supports previous observations that strains lacking the S-layer are less virulent *in vivo*, although it remains difficult to identify the specific contribution of the S-layer in the infection process. Instead, this work highlights the potential multifunctional contribution that the S-layer plays in disease as, despite the number of differentially expressed genes observed between R20291, FM2.5 and the revertants *in vitro,* recovery of the intact S-layer was sufficient to restore virulence, at least in one revertant. Importantly, isolation and characterisation of these variants, together with greater knowledge of gene regulation and metabolic pathways impacted, offers a new opportunity to better understand the role of the S-layer in *C. difficile* pathogenesis.

553 **Materials and Methods**

554 **Bacterial strains and growth conditions**

555 The bacterial strains used in this study include *C. difficile* strain R20291, its
556 derivative FM2.5 (Kirk et al. 2017), RvA, RvB, RvC, RvD and FM2.5Δ*PaLoc* (this
557 study). Strains were routinely grown under anaerobic conditions on Braziers
558 cycloserine, cefoxitin egg yolk (CCEY) agar (Oxoid, UK); CHROMID® *C. difficile*
559 Chromogenic medium (bioMérieux); or in Tryptone yeast (TY) broth (Oxoid, UK).

560

561 **Generation of FM2.5Δ*PaLoc***

562 Homologous recombination was used to generate a derivative of strain
563 FM2.5 that lacked the entire pathogenicity locus (PaLoc). Briefly, 1.2 kb up and
564 downstream of the PaLoc was amplified by PCR using RF920
565 (cgtagaaatacggtgttttttgttaccctaTGGAATTTAGATATAAAAACCAATTC) and
566 RF921 (atttattttggtgtgGACAACATTGGAATTAAATCAG), and RF922
567 (aattccaatgttgtcCACACCAAAATAAATGCC) and RF923
568 (gggattttggtcatgagattatcaaaaaggCCCAACTATGGAAAAACC), respectively, and
569 cloned by Gibson assembly into plasmid pJAK112 (Fuchs et al. 2021) that had been
570 linearised by PCR using RF311 (TAGGGTAACAAAAAACACCG) and RF312
571 (CCTTTTTGATAATCTCATGACC). The resulting plasmid, pJAK143, was then
572 conjugated into *C. difficile* (Kirk and Fagan 2016) and mutagenesis to knock out the
573 PaLoc was carried out using standard allele exchange (Cartman et al. 2012b).

574

Murine model of infection

All procedures were performed in strict accordance with the Animals (Scientific Procedures) Act 1986 with specific approval granted by the Home Office, UK (PPL 60/8797 and PPL PI440270). Food and water were provided *ad libitum* and animals kept at a constant room temperature of 20–22 °C with a 12 h light/dark cycle. Groups of up to six C57/bl6 mice aged 6–8 weeks supplied by Charles River (Edinburgh) were used in each treatment group. An antibiotic cocktail (kanamycin [0.40 mg ml^{-1}]; metronidazole [0.215 mg ml^{-1}]; colistin [850 U ml^{-1}]; gentamicin [0.035 mg ml^{-1}]; and vancomycin [0.045 mg ml^{-1}] [all Sigma Aldrich, UK]) was administered *ad libitum* in the drinking water as previously described (Jukes et al., 2020) with clindamycin sulphate (150 mg Kg^{-1}), administered by oral gavage following cessation of the antibiotic cocktail. Animals were each challenged with approximately 10^5 spores of *C. difficile* 72 h after clindamycin treatment. Mice were monitored closely post-infection and weighed daily to determine the severity of the disease. Animals with a weight loss greater than 10 % of pre-challenge weight were given soft food and were culled if weight loss reached 20 %.

C. difficile shedding and organ colonization

Fresh faecal samples collected daily were weighed, serially diluted in phosphate buffered saline (PBS) and cultured on CCEY agar at 37 °C for 48 h. At the experimental endpoint, animals were culled, and the caecum and colon harvested. Enumeration of total counts and spore-specific counts in lumen associated material

597 were performed as previously described (Jukes et al., 2020) . In brief, total viable
598 counts were determined by plating serial dilutions on ChromID selective media
599 (Biomeuriex). Spores were enumerated following heat treatment at 56 °C for 20 min.
600
601 **Quantification of toxin expression**
602 Quantification of toxin activity was performed using monolayers of Vero
603 cells (kidney epithelial cells) as described previously (Buckley et al. 2011). Briefly,
604 toxin was recovered from the spent filtered TY medium used to support bacterial
605 growth for 36-72 h. Spent medium was recovered at the same stage of the growth
606 cycle and at the same OD_{600nm}. *In vivo* toxin activity was measured by filtering
607 luminal content collected from the caecum and colon of infected mice. Luminal
608 content collected from the caecum and colon of uninfected mice was used as a
609 control. Samples for toxin measurement were tested by the addition of serial
610 dilutions to confluent monolayers within 72 h of collection. Cells and toxin were
611 co-cultured for 24 h before cells were washed with phosphate buffered saline (PBS),
612 fixed with 5 % formal saline (Fisher), and stained with Giemsa for 30 min, before
613 thorough washing to remove excessive stain. For data presented in Figures 1d, and
614 S3, toxin activity was determined as the reciprocal of last dilution in which toxin
615 activity was observed, i.e. showing cell destruction. In Figure 5, in an attempt to
616 quantify the toxin activity more precisely, excess stain was removed by washing
617 with PBS before cells were permeabilised to release internalised stain using 200 µl
618 1% SDS. 100 µl of the supernatant was transferred to a fresh plate and the OD_{620nm}

values determined. In this context higher values indicate Vero cells are intact and unaffected by the toxin, lower values indicate that toxin mediated damage prevents uptake and retention of the dye.

Histology and immunohistochemistry

Tissue samples were harvested from the caecum and colon of antibiotically susceptible animals infected with either R20291 or FM2.5 at post-mortem, 48 hpi. These tissues were gently washed in sterile PBS and immediately fixed in 10% formalin. Embedded tissue sections were cut and stained with Hematoxylin and Eosin (Jukes et al., 2020). Blind histological scoring of tissue was performed on 3 independent sections of caecal and colonic tissue. Each section was scored out of a total of 20, with a score of 1 indicating no change, 2 mild change, 3 moderate change and 4 severe changes, for the following categories: epithelial damage, neutrophil migration, haemorrhagic congestion, tissue oedema and crypt hyperplasia. Data presented represents the mean scores for 3 mice for each treatment.

slpA sequencing from isolated revertant clones

Individual clones of bacteria, recovered from faecal or tissue associated material that showed different morphology on ChromID plates, where subject to at least two rounds of clonal selection. Genomic DNA was isolated from a 20 ml culture grown anaerobically for 18 h in tryptic soya broth (TSB). Bacterial cells were initially disrupted enzymatically by resuspending the pellet in lysis buffer (20 mM

641 Tris-Cl, pH8.0, 2 mM Na EDTA, 1.2 % Triton X-100, lysozyme 200 mg ml^{-1}), and
642 incubated at 56 °C for 90 min. The DNA was recovered using the DNeasy Blood and
643 Tissue Kit (Qiagen), following manufacturer's instructions. A 478 bp fragment of
644 *slpA*, centred on the FM2.5 point mutation (Kirk et al. 2017), was amplified by using
645 oligonucelotides RF110 (GACATAACTGCAGCACTACTTG) and RF111
646 (CAGGATTAACAGTATTAGCTTCTGC). The resulting fragments were subjected
647 to Sanger sequencing and compared to wild type and FM2.5 sequences.
648
649 **Isolation and sequencing of *slpA* from faecal extracts**
650 Faecal samples were also collected for sequencing of *slp*A, by directly
651 extracting DNA from faecal samples using the FastDNA SPIN kit for soil (MP
652 Biomedicals). Briefly, approximately 200-600 mg of faeces sample was suspended in
653 978 µl sodium phosphate buffer with 122 µl MT™ buffer lysis solution. Samples
654 were then homogenised in a FastPrep instrument using two 30 second pulses, at
655 speed setting 6.5, in lysing matrix E. Samples were centrifuged for 10 min at 14,000
656 *x g* to remove debris. 250 µl protein precipitation solution was added to the lysate
657 supernatant, and the precipitant formed removed by centrifugation at 14,000 *x g* for
658 5 min. DNA was then bound to a silica matrix, washed using the kit wash buffer,
659 and eluted with water.
660 DNA extracted from faeces was used as a template for PCR amplification of
661 a 330 bp fragment of *slpA* using Phusion polymerase (NEB) and RF2193
662 (ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTACTTGTAGCTACTTTTA

663 TTGCAC) and RF2194 (GACTGGAGTTCAGACGTGTGCTCTTCCGATCT
664 CAAGGATATACAGTAGTACAGAGC) oligonucleotides. Resulting DNA
665 fragments were purified and sequenced using the amplicon-EZ service offered by
666 GENEWIZ (Azenta Life Sciences).
667
668 **Extraction and western immunoblot analysis of S-layer and associated proteins**
669 Surface layer proteins were extracted using low pH glycine as previously
670 described (Fagan et al. 2009) and analysed by SDS-PAGE using standard methods
671 (Laemmli 1970). Proteins were transferred to nitrocellulose membranes via semi-
672 dry transfer (Bio-Rad Trans Blot Turbo; 25 V, 30 min) for western immunoblot
673 analysis. Transfer efficiency was confirmed by PonceauS staining of membrane post
674 transfer, and Coomassie staining of the polyacrylamide gel following transfer (Fig.
675 S5). Membranes were blocked for 1 h in Phosphate-buffered Saline containing 0.1
676 % Tween20 (PBS-T) with 5 % milk powder. Blots were subsequently incubated in
677 primary antibody (rabbit anti-SLP_H raised against *C. difficile* 630 1:100,000 dilution;
678 rabbit anti-SLP_L raised against *C. difficile* R20291 1:200,000 dilution) in PBS-T
679 containing 3 % milk powder, for 1 h at room temperature. Membranes were washed
680 thoroughly in PBS-T before incubation with secondary antibodies (anti-rabbit
681 horseradish peroxidase, Promega WB401B 1:2,500 dilution) for 1 h at room
682 temperature. Blots were washed in PBS-T before detection by chemiluminescence
683 (Bio-Rad). Molecular weight (MW) markers (Thermo Scientific™ 26616) were

684 imaged (Bio-Rad ChemiDoc XRS+) simultaneously and overlaid onto the blots to
685 aid visualisation.
686
687 **Protein purification and X-ray crystallography**
688 *C. difficile* revertant strains were cultured in 400 ml of TYG broth for 16 h.
689 Cultures were then centrifuged at room temperature at 4,696 x *g* and resulting
690 pellets were washed with 40 ml of 0.01 M HEPES pH 7.4 and 0.15 M sodium
691 chloride (HBS) buffer. S-layer extraction was performed by resuspending the
692 washed pellet in 4 ml of 0.2 M glycine-HCl pH 2.2 and centrifugation for 5 min at
693 21,100 x *g*. Collected supernatant was then neutralized with 2 M Tris-base. S-layer
694 extract was filtered and resolved onto a Superdex 200 26/600 column using an ÄKTA
695 Pure FPLC system (Cytiva) in 50 mM Tris-HCl pH 7.5, 150 mM NaCl buffer.
696 Purified $SlpA_{RvB}$ at 10 mg ml^{-1} was subjected to crystallization using a
697 Mosquito liquid handling robot (TTP Labtech), with the sitting drop vapor-
698 diffusion method, at 20 °C. Crystals were obtained in 0.03 M magnesium chloride
699 hexahydrate; 0.03 M calcium chloride dihydrate, 0.12 M ethyleneglycol, 0.05 M Tris
700 (base); 0.05 M bicine pH 8.5, 20% v/v glycerol; 10% w/v PEG 4,000.
701 Data were collected on the I24 (λ = 0.71 Å) beamline at the Diamond Light
702 Source Synchrotron (Didcot, UK; mx24948-136) at 100 K. The data were acquired
703 from the automatic multi-crystal data-analysis software pipeline xia2.multiplex
704 (Gildea et al. 2022) within the Information System for Protein Crystallography
705 Beamline (ISPyB), re-processed using Automatic Image Processing with Xia-2

706 (DIALS [Winter et al. 2018] and Aimless 3d [Evans and Murshudov 2013]) and
707 scaled with Aimless within ccp4.cloud of CCP4 (Winn et al. 2011) software suit.
708 The initial model of the core SLP_H was obtained by molecular replacement
709 in Phaser (McCoy et al. 2007), using an $SlpA_{RvB}$ model of CWB2 domains, derived
710 from the $SlpA_{R\Delta D2}$ model (PDB ID: 7ACZ) and calculated using SWISS-MODEL
711 (Waterhouse et al. 2018). The generated solution model was then subjected to
712 automatic model building with Modelcraft (Cowtan et al. 2020), followed by
713 manual building with Coot (Emsley and Cowtan 2004) and refinement in Refmac5
714 (Murshudov et al. 2011).
715 Final models were obtained after iterative cycles of manual model building
716 with Coot and refinement in phenix_refine (Liebschner et al. 2019). Data collection
717 and refinement statistics are summarized in Table S1.
718 PDBePISA (Krissinel and Henrick 2007) was used to investigate interdomain
719 and protein-protein interfaces in the crystallographic lattice to identify interacting
720 residues, which were confirmed by manual inspection within COOT.
721 Structural representations were generated using PyMOL Molecular Graphics
722 System (Schrödinger, LLC).
723

724 **Protein structure prediction**

725 Homology models for $SlpA_{R20291}$ and $SlpA_{RvA}$ were generated by providing
726 SWISS-MODEL webserver with the $SlpA_{RvB}$ (PDB ID: 8BBY) or $SlpA_{R\Delta D2}$ (PDB ID:
727 7ACZ) as user templates, as well as without a template. Structural alignments

728 between predicted models and templates were performed using COOT (Emsley and
729 Cowtan, 2004). As $SlpA_{R\Delta D2}$ lacks the D2 domain, predicted models based on this
730 experimental model have a disordered D2 domain. Therefore, overall comparison
731 of the three predicted models was based on models calculated in the default mode,
732 which uses $SlpA_{R7404}$ (PDB ID: 7ACX) as a template, while analysis of the reversion-
733 containing region in the D1 domain was done using $SlpA_{R\Delta D2}$ or $SlpA_{RvB}$ derived
734 models.
735
736 **Recovery of mRNA for RNA sequence analysis**
737 RNA was recovered from *C. difficile* strains R20291, FM2.5, RvA and RvB
738 which had been cultured *in vitro* in TY broth. Briefly, bacterial cells reaching an
739 OD_{620nm} = 0.6 were pelleted (5000 x *g*, 15 min) and immediately fixed in 1.5 ml RNA-
740 protect (Qiagen) for 10 min before being processed using a PureLink RNA mini kit
741 (Ambion) to extract total RNA. To ensure maximal lysis of bacteria and recovery of
742 RNA, the bacterial pellet was additionally subject to treatment with 100 ml
743 lysozyme solution (10 mg ml^{-1} in 10 mM Tris-HCl [pH8.0] 0.1 mM EDTA), 0.5 ml
744 10% SDS solution and 350 ml of Lysis Buffer (Invitrogen PureLink RNA Mini Kit)
745 containing 2-mercaptoethanol. Cells were then homogenised using MP Biomedical
746 beads (0.1 mm) and bead beater (MP Biomedical FastPrep24) with cells subject to 2
747 cycles of 60 s beating, followed by incubation on ice for 2 min. Total RNA was then
748 extracted using the standard PureLink RNA mini kit protocol, according to the
749 manufacturer's specifications. Genomic DNA was removed using a TURBO DNase

kit (Ambion) and samples were tested for efficient removal of DNA by conventional PCR. Samples for RNA-Seq were prepared in triplicate on two separate occasions (6 samples for each) for all four bacterial strains.

Illumina library preparation of mRNA samples for RNA-Seq was prepared using a TruSeq Stranded mRNA library prep kit (Illumina) according to the manufacturer's instructions. Sequencing was performed on the Illumina NextSeq 500 platform (75 bp length; single-end). Library generation, optimisation of amplification and sequencing were performed at the University of Glasgow Polyomics facility. Quality control of sequencing data was performed using FastQC (Babraham Bioinformatics) to assess the minimum Phred threshold of 20 and potential data contamination. The raw data will be deposited to the Gene Expression Omnibus reference ID GSE205747.

RNA analysis and identification of differentially expressed genes

Raw RNA-Seq datasets were subject to the following pipeline. Firstly, fastQ files were assessed using FastP (Chen et al. 2018) and then were aligned to the *C. difficile* R20291 (accession number NC_013316) reference genome using STAR (Dobin et al. 2013) (v2.6) with –quantMode GeneCounts –outFilterMultimapNmax 1 and –outFilterMatchNmin 35. We used a Star index with a –sjdbOverhang of the maximum read length – 1. Next, read count files were merged and genes with mean of < 1 read per sample were excluded. Finally, the expression and differential expression values were generated using DESeq2 (Love et al. 2014) (v1.24). For

772 differential comparisons, we used an A versus B model with no additional
773 covariates. All other parameters were left to default.
774 The processed data was then visualised using Searchlight (Cole et al.
775 2021), specifying one differential expression workflow for each comparison, an
776 absolute $\log_2$-fold cut-off of 1 and adjusted *p* of 0.05. All other parameters were left
777 to default.
778

779 **Pathway Analysis Methods**

780 Functional and metabolic pathways were implied by interrogation of the
781 WP numbers assigned using the *C. difficile* annotated genome (NC_013316.1) and
782 entered into Uniprot or NCBI Blastp databases. Gene ontology (GO) was assigned
783 based on Biological Process assignment within Uniprot.
784

785 **Statistical analysis**

786 Statistical analysis was carried out in GraphPad Prism v.9. The tests and
787 parameters used are detailed in the figure legends throughout. Tests used included
788 *t*-test with Welch correction, ANOVA with Tukey's post-test and Kruskal–Wallis
789 with Dunn's multiple comparisons.
790

791 **Acknowledgments**

792 This work was supported by the Wellcome Trust [204877/Z/16/Z]. FV, MO,
793 ABS, JAK and JH were supported by this Wellcome Trust Collaborative Award,

794 awarded to GRD, PSS, RPF. The authors would like to thank Dr Arnaud Baslé,
795 facility manager of the Newcastle Structural Biology Lab, and beamline staff at I24,
796 Diamond Light Source (BAG mx24948) for support on structural data collection.
797 The contents of this work are solely the responsibilities of the authors and do not
798 reflect the official views of any of the funders, who had no role in study design, data
799 collection, analysis, decision to publish, or preparation of the manuscript.

800

801 **Contributions**

802 MO, FV, GRD carried out animal experiments, collected and analyzed data, wrote
803 and revised the manuscript. JAK designed and performed experiments to generate
804 the FM2.5Δ*PaLoc* strain. JAK, RRC and RFP carried out analysis on the sequential
805 samples from the mouse experiments, wrote and revised the manuscript. ABS, PLM
806 and PSS carried out crystallization, structural determination, modelling and analysis
807 of the revertants, wrote and revised the manuscript. MO, JH and JC carried out the
808 transcriptomics analysis, MO and JH processed the samples and JC undertook
809 bioinformatic analysis. GRD, PSS and RPF designed experiments, analyzed the data,
810 supervised the study, wrote and revised the manuscript.

811

812 **Conflict of Interest:** The authors have no competing interests that might be
813 perceived to influence the interpretation of the article.

814

815

816

References

Arato, V., Gasperini, G., Giusti, F., Ferlenghi, I., Scarselli, M. and Leuzzi, R. 2019. Dual role of the colonization factor CD2831 in *Clostridium difficile* pathogenesis. *Scientific Reports* 9(5554).

Ausiello, C.M. et al. 2006. Surface layer proteins from *Clostridium difficile* induce inflammatory and regulatory cytokines in human monocytes and dendritic cells. *Microbes and Infection* 8(11), pp. 2640–2646.

Batah, J. et al. 2017. *Clostridium difficile* flagella induce a pro-inflammatory response in intestinal epithelium of mice in cooperation with toxins. *Scientific Reports* 7(3256).

Best, E.L., Freeman, J. and Wilcox, M.H. 2012. Models for the study of *Clostridium difficile* infection. *Gut Microbes* 3(2), pp. 145–167.

Braun, V., Hundsberger, T., Leukel, P., Sauerborn, M. and Eichel-Streiber, C. von 1996. Definition of the single integration site of the pathogenicity locus in *Clostridium difficile*. *Gene* 181(1–2), pp. 29–38.

Buckley, A.M., Spencer, J., Candlish, D., Irvine, J.J. and Douce, G.R. 2011. Infection of hamsters with the UK *Clostridium difficile* ribotype 027 outbreak strain R20291. *Journal of Medical Microbiology* 60(8), pp. 1174–1180.

Buffie, C.G. et al. 2015. Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature* 517(7533), pp. 205–208.

Callegari, M.L., Riboli, B., Sanders, J.W., Cocconcelli, P.S., Kok, J., Venema, G. and Morelli, L. 1998. The Slayer gene of Lactobacillus helveticus CNRZ 892 : cloning, sequence and heterologous expression. *Microbiology* 144, pp. 719–726.

Cartman, S.T., Kelly, M.L., Heeg, D., Heap, J.T. and Minton, N.P. 2012a. Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between the tcdC genotype and toxin production. *Applied and environmental microbiology* 78(13), pp. 4683–4690.

Chandrasekaran, R. and Lacy, D.B. 2017. The role of toxins in *Clostridium difficile* infection. *FEMS Microbiology Reviews* 41(6), pp. 723–750.

Chen, S., Zhou, Y., Chen, Y. and Gu, J. 2018. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34(17), pp. i884–i890.

Chen, X., Katchar, K., Goldsmith, J.D., Nanthakumar, N., Cheknis, A., Gerding, D.N. and Kelly, C.P. 2008a. A mouse model of *Clostridium difficile*-associated disease. *Gastroenterology* 135(6), pp. 1984–1992.

Chen, X., Katchar, K., Goldsmith, J.D., Nanthakumar, N., Cheknis, A., Gerding, D.N. and Kelly, C.P. 2008b. A Mouse Model of *Clostridium difficile*-Associated Disease. *Gastroenterology* 135(6), pp. 1984–1992.

Chen, Y. et al. 2020. Membrane Cholesterol Is Crucial for *Clostridium difficile* Surface Layer Protein Binding and Triggering Inflammasome Activation. *Frontiers in Immunology* 11, p. 1675.

857 Claus, H., Aka, E., Debaerdemaeker, T., Evrard, C., Declercq, J.P. and König, H.
858 2002. Primary structure of selected archaeal mesophilic and extremely
859 thermophilic outer surface layer proteins. *Systematic and Applied Microbiology*
860 25(1), pp. 3–12.
861 Cole, J.J., Faydaci, B.A., McGuinness, D., Shaw, R., Maciewicz, R.A., Robertson,
862 N.A. and Goodyear, C.S. 2021. Searchlight: automated bulk RNA-seq exploration
863 and visualisation using dynamically generated R scripts. *BMC Bioinformatics*
864 22(411).
865 Cowardin, C.A., Kuehne, S.A., Buonomo, E.L., Marie, C.S., Minton, N.P. and Petri,
866 W.A. 2015. Inflammasome activation contributes to interleukin-23 production in
867 response to *Clostridium difficile*. *mBio* 6(1):e02386-14.
868 Dapa, T. et al. 2013. Multiple factors modulate biofilm formation by the anaerobic
869 pathogen *Clostridium difficile*. *Journal of Bacteriology* 195(3), pp. 545–555.
870 Deakin, L.J. et al. 2012. The *Clostridium difficile* spo0A gene is a persistence and
871 transmission factor. *Infection and Immunity* 80(8), pp. 2704–2711.
872 Dingle, K.E. et al. 2013. Recombinational switching of the *Clostridium difficile* S-
873 layer and a novel glycosylation gene cluster revealed by large-scale whole-genome
874 sequencing. *The Journal of infectious diseases* 207(4), pp. 675–686.
875 Dobin, A. et al. 2013. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*
876 29(1), pp. 15–21.
877 Emsley, P. and Cowtan, K. 2004. Coot: model-building tools for molecular
878 graphics. *Acta crystallographica. Section D, Biological crystallography* 60(Pt 12 Pt
879 1), pp. 2126–2132.
880 Engelhardt, H. 2007. Are S-layers exoskeletons? The basic function of protein
881 surface layers revisited. *Journal of Structural Biology* 160(2), pp. 115–124.
882 Engelhardt, H. and Peters, J. 1998. Structural research on surface layers: A focus
883 on stability, surface layer homology domains, and surface layer-cell wall
884 interactions. *Journal of Structural Biology* 124(2–3), pp. 276–302.
885 Fagan, R.P., Albesa-Jové, D., Qazi, O., Svergun, D.I., Brown, K.A. and
886 Fairweather, N.F. 2009. Structural insights into the molecular organization of the
887 S-layer from *Clostridium difficile*. *Molecular Microbiology* 71(5), pp. 1308–1322.
888 Fagan, R.P. and Fairweather, N.F. 2014. Biogenesis and functions of bacterial S-
889 layers. *Nature Reviews Microbiology* 12(3), pp. 211–222.
890 Fletcher, J.R. et al. 2021. *Clostridioides difficile* exploits toxin-mediated
891 inflammation to alter the host nutritional landscape and exclude competitors from
892 the gut microbiota. *Nature Communications* 12(462).
893 Fuchs, M. et al. 2021. An RNA-centric global view of *Clostridioides difficile*
894 reveals broad activity of Hfq in a clinically important gram-positive bacterium.
895 *Proceedings of the National Academy of Sciences of the United States of America*
896 118(25).
897 Girinathan, B.P. et al. 2021. In vivo commensal control of *Clostridioides difficile*
898 virulence. *Cell Host & Microbe* 29(11), pp. 1693-1708.e7.

899 Jarchum, I., Liu, M., Shi, C., Equinda, M. and Pamer, E.G. 2012. Critical role for
900 myd88-Mediated Neutrophil recruitment during *Clostridium difficile* colitis.
901 *Infection and Immunity* 80(9), pp. 2989–2996.
902 Jukes, C.A. et al. 2020a. Bile salt metabolism is not the only factor contributing to
903 *Clostridioides* (*Clostridium*) *difficile* disease severity in the murine model of
904 disease. *Gut Microbes* 11(3), pp. 481–496.
905 Jukes, C.A. et al. 2020b. Bile salt metabolism is not the only factor contributing to
906 *Clostridioides* (*Clostridium*) *difficile* disease severity in the murine model of
907 disease. *Gut Microbes* 11(3), pp. 481–496.
908 K. Laemmli, U. 1970. Cleavage of structural proteins during the assembly of the
909 head of bacteriophage T4. *Nature* 227(15), pp. 680–685.
910 Kirby, J.M., Ahern, H., Roberts, A.K., Kumar, V., Freeman, Z., Acharya, K.R. and
911 Shone, C.C. 2009. Cwp84, a surface-associated cysteine protease, plays a role in the
912 maturation of the surface layer of *Clostridium difficile*. *Journal of Biological*
913 *Chemistry* 284(50), pp. 34666–34673.
914 Kirk, J.A. et al. 2017. New class of precision antimicrobials redefines role of
915 *Clostridium difficile* S-layer in virulence and viability. *Science Translational*
916 *Medicine* 9(406).
917 Kirk, J.A. and Fagan, R.P. 2016. Heat shock increases conjugation efficiency in
918 *Clostridium difficile*. *Anaerobe* 42, pp. 1–5.
919 Krissinel, E. and Henrick, K. 2007. Inference of macromolecular assemblies from
920 crystalline state. *Journal of molecular biology* 372(3), pp. 774–797.
921 Lanzoni-Mangutchi, P. et al. 2022. Structure and assembly of the S-layer in *C.*
922 *difficile*. *Nature Communications 2022 13:1* 13(1), pp. 1–13.
923 Lortal, S., Heijenoort, J. van, Gruber, K. and Sleytr, U.B. 1992. S-layer of
924 *Lactobacillus helveticus* ATCC 12046: Isolation, chemical characterization and re-
925 formation after extraction with lithium chloride. *Journal of General Microbiology*
926 138(3), pp. 611–618.
927 Love, M.I., Huber, W. and Anders, S. 2014. Moderated estimation of fold change
928 and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15(12).
929 Maldarelli, G.A., Masi, L. de, von Rosenvinge, E.C., Carter, M. and Donnenberg,
930 M.S. 2014. Identification, immunogenicity, and cross-reactivity of type IV pilin
931 and pilin-like proteins from *Clostridium difficile*. *Pathogens and Disease* 71(3), pp.
932 302–314.
933 Mamareli, P., Kruse, F., Friedrich, C., Smit, N., Strowig, T., Sparwasser, T. and
934 Lochner, M. 2019. Epithelium-specific MyD88 signaling, but not DCs or
935 macrophages, control acute intestinal infection with *Clostridium difficile*.
936 *European journal of immunology* 49(5), pp. 747–757.
937 McDermott, A.J., Falkowski, N.R., McDonald, R.A., Frank, C.R., Pandit, C.R.,
938 Young, V.B. and Huffnagle, G.B. 2017. Role of interferon-γ and inflammatory
939 monocytes in driving colonic inflammation during acute *Clostridium difficile*
940 infection in mice. *Immunology* 150(4), pp. 468–477.

McDermott, A.J., Falkowski, N.R., McDonald, R.A., Pandit, C.R., Young, V.B. and Huffnagle, G.B. 2016. Interleukin-23 (IL-23), independent of IL-17 and IL-22, drives neutrophil recruitment and innate inflammation during *Clostridium difficile* colitis in mice. *Immunology* 147(1), pp. 114–124.

Merrigan, M.M. et al. 2013. Surface-Layer Protein A (SlpA) is a major contributor to host-cell adherence of *Clostridium difficile*. *PLoS ONE* 8(11), p. e78404.

Richards, E. et al. 2018. The S-layer protein of a *Clostridium difficile* SLCT-11 strain displays a complex glycan required for normal cell growth and morphology. *Journal of Biological Chemistry* 293(47), pp. 18123–18137.

Ryan, A. et al. 2011. A Role for TLR4 in *Clostridium difficile* Infection and the Recognition of Surface Layer Proteins. *PLOS Pathogens* 7(6), p. e1002076.

Sakakibara, J., Nagano, K., Murakami, Y., Higuchi, N., Nakamura, H., Shimozato, K. and Yoshimura, F. 2007. Loss of adherence ability to human gingival epithelial cells in S-layer protein-deficient mutants of *Tannerella forsythensis*. *Microbiology* 153(3), pp. 866–876.

Sekot, G., Posch, G., Messner, P., Matejka, M., Rausch-Fan, X., Andrukhov, O. and Schäffer, C. 2011. Potential of the *Tannerella forsythia* S-layer to delay the immune response. *Journal of Dental Research* 90(1), pp. 109–114.

Sleytr, U.B. and Beveridge, T.J. 1999. Bacterial S-layers. *Trends in Microbiology* 7(6), pp. 253–260.

Smits, W.K., Lyras, D., Lacy, D.B., Wilcox, M.H. and Kuijper, E.J. 2016. Clostridium difficile infection. *Nature reviews. Disease primers* 2, pp. 1–20.

Theriot, C.M., Koumpouras, C.C., Carlson, P.E., Bergin, I.I., Aronoff, D.M. and Young, V.B. 2011. Cefoperazone-treated mice as an experimental platform to assess differential virulence of *Clostridium difficile* strains. *Gut microbes* 2(6), pp. 326–334.

Winston, J.A., Thanissery, R., Montgomery, S.A. and Theriot, C.M. 2016. Cefoperazone-treated Mouse Model of Clinically-relevant *Clostridium difficile* Strain R20291. *J. Vis. Exp* 118, p. 54850.

Woods, E.C., Edwards, A.N., Childress, K.O., Jones, J.B. and McBride, S.M. 2018. The *C. difficile* clnRAB operon initiates adaptations to the host environment in response to LL-37. *PLoS Pathogens* 14(8):e1007153.