
Article:
Banerji, O., Lanzoni-Mangutchi, P., Vaz, F. et al. (Submitted: 2020) Structure and assembly of the S-layer determine virulence in C. difficile. In Review. (Submitted)

https://doi.org/10.21203/rs.3.rs-79088/v1

Structure and assembly of the S-layer determine virulence in *C. difficile*

Oishik Banerji[2*#], Paola Lanzoni-Mangutchi[1*], Filipa Vaz[4*‡], Anna Barwinska-Sendra[1*], Jason Wilson[2*], Joseph A. Kirk[2,3], Shauna O'Beirne[2,3], Arnaud Baslé[1], Kamel El Omari[5], Armin Wagner[5], Neil F. Fairweather[6], Gillian R. Douce[4], Per A. Bullough[2], Robert P. Fagan[2,3§], Paula S. Salgado[1§]

[1] Biosciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK

[2] Krebs Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, UK

[3] Florey Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, UK

[4] Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK

[5] Diamond Light Source, Oxfordshire, UK

[6] Dept. Of Life Sciences, Imperial College London, London, UK

[#] Current address: Institute of Structural and Molecular Biology, Department of Biological Sciences, Birkbeck, University of London, London, UK

[‡] Current address: Department of Immunology, Oslo University Hospital, Oslo, Norway

* These authors contributed equally

[§] **Corresponding authors**: Address: [1] Biosciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK. Telephone: +44 (0)191 208 7432;

paula.salgado@newcastle.ac.uk or [2] Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, UK. Telephone: +44 (0)114 222 4182; r.fagan@sheffield.ac.uk

Many bacteria and archaea possess a cell surface layer – S-layer – made of a two-dimensional protein array that covers the entire cell. As the outermost component of the cell envelope, S-layers play crucial roles in many aspects of cell physiology[1]. Importantly, many clinically relevant bacterial pathogens possess a distinct S-layer that forms an initial interface with the host[2], making it a potential target for development of species-specific antimicrobials. Targeted therapeutics are particularly important for antibiotic resistant pathogens such as *Clostridioides difficile*, the most frequent cause of hospital acquired diarrhea, which relies on disruption of normal microbiota through antibiotic usage[3]. Despite the ubiquity of S-layers, only partial structural information from a very limited number of species is available[4-10] and their function and organization remains poorly understood.

Here we report the first complete atomic level structure and *in situ* assembly model of an S-layer from a bacterial pathogen and reveal its role in disease severity. SlpA, the main *C. difficile* S-layer protein, assembles through tiling of triangular prisms abutting the cell wall, interlocked by distinct ridges facing the environment. This forms a tightly packed array, unlike the more porous S-layer models previously described[4,5,8,11]. We report that removing one of the SlpA ridge features dramatically reduces disease severity, despite being dispensable for overall structure and S-layer assembly. Remarkably, the effect on disease severity is independent of toxin production and bacterial colonization within the mouse model of disease.

Our work combines X-ray and electron crystallography to reveal a novel S-layer organization in atomic detail, highlighting the need for multiple technical approaches to obtain structural information on these paracrystalline arrays. These data also establish a direct link between specific structural elements of the S-layer and virulence for the first

time, in a crucial paradigm shift in our understanding of *C. difficile* disease, currently largely attributed to the action of potent toxins[3].

This work highlights the crucial role of S-layers in pathogenicity and the importance of detailed structural information for providing new therapeutic avenues, targeting the S-layer. Understanding the interplay between S-layer and other virulence factors will further enhance our ability to tackle pathogens carrying an S-layer. We anticipate that this work provides a solid basis for development of new, *C. difficile*-specific therapeutics, targeting SlpA structure and S-layer assembly to reduce the healthcare burden of these infections.

Introduction

The surfaces of most bacteria and archaea are covered with a proteinaceous coat, the surface or S-layer, that is formed through the self-assembly of individual protein subunits into a regularly spaced, two-dimensional array[2]. The tendency of S-layer proteins to spontaneously form two-dimension (2D) assemblies has hampered structure determination and restricted understanding of both their function and architecture. Partial S-layer structures determined from a limited number of species[4,5,7,9,10] showcase the huge diversity in sequence and structure of S-layer proteins and their arrangements, providing some insights into the mechanisms of anchoring and structural organization. However, no complete X-ray structure has been presented for any major S-layer protein, despite the ubiquity of this type of array, particularly among the medically-important Firmicutes such as *Clostridioides difficile*. This Gram-positive opportunistic pathogen is the leading cause of hospital-acquired, antibiotic-associated diarrheal disease globally[3].

C. difficile infection (CDI) causes substantial morbidity and mortality with severe disease characterized by intestinal inflammation, resulting in extensive damage to the colon and even death[3]. This pathology has largely been attributed to the direct action of two potent toxins, that also initiate a proinflammatory response via activation of the inflammasome[12]. Other less studied factors, including the S-layer, contribute to the recruitment of neutrophils, via a TLR/Myd88 dependent signaling pathway[13,14]. However, the exact role of the S-layer remains unclear. In *C. difficile,* the S-layer largely consists of a major S-layer protein, SlpA (Fig. 1a), responsible for S-layer assembly into a paracrystalline array. Minor components of the S-layer, belonging to a family of 28 cell wall proteins (CWPs), are inserted in the array, comprise an estimated 5-10% of the S-layer and provide additional functions[2]. Recently, we reported that an S-layer-null mutant of *C. difficile* was avirulent in

the acute hamster model of disease, despite apparent normal colonization of the caecum and colon[15]. Notably, absence of a functional SlpA resulted in a range of pleiotropic effects, including reduced toxin production. Although this work suggested a role for S-layer in *C. difficile* disease, reduced toxin expression made it impossible to establish a direct effect.

To fully interrogate the role of the S-layer in pathogen survival and host disease severity, we need a complete, high resolution S-layer structure. Moreover, the uniqueness of each species' S-layer makes these arrays attractive targets for therapeutic interventions, provided sufficient structural and functional characterization is available. Here we present the first complete atomic level model of an S-layer, that of *C. difficile,* generated by combining high-resolution X-ray crystallography with electron microscopy.

Using this structural information, we uncover elements of SlpA that play an essential role in *C. difficile* pathogenicity. We demonstrate that an intact wild type S-layer is an absolute requirement for full disease severity, in a mechanism unrelated to toxin production or host colonization.

C. difficile major S-layer component forms an intricate complex

In order to address the knowledge gap on the *C. difficile* S-layer, we first sought to determine the structure of its main component, SlpA, which is post-translationally cleaved into two S-layer proteins (SLPs); the high molecular weight (HMW) and low molecular weight (LMW)[16], herein referred to as SLP_H and SLP_L, respectively. These subunits then form a complex (referred to as H/L) that is incorporated in the S-layer (Fig. 1a). The structure of the full-length H/L complex was determined by X-ray crystallography to an overall resolution of 2.55 Å (PDB ID: 7ACY) by combining single anomalous dispersion sulphur data (S-SAD) and molecular replacement using substructures of the interacting domains (LMW SLP interacting

domain, LID, and HMW SLP interacting domain, HID; PDB ID: 7ACW) and SLP_L/HID (PDB ID: 7ACV) (Fig. 1). Our H/L structural model reveals three distinct regions: the pseudo-threefold SLP_H tile, an intricate LID/HID interacting motif and a third region composed of two domains, D1 and D2, of SLP_L (Fig. 1b). These regions define two separate planes, with the SLP_L spanning ~35 Å above the SLP_H plane (Fig. 1b), linked by the LID/HID motif.

SLP_H is composed of the three conserved cell wall binding motifs - CWB2 - that define the *C. difficile* cell wall protein (CWP) family[2], and the HID. The three CWB2 motifs form a triangular prism and adopt an intertwined fold, with a β-strand from one CWB2 inserting into the neighbouring domain to complete a β-sheet, sandwiched between two α-helical regions (Extended Data Fig. 1a). At the core of the tile sits a helical bundle with each individual CWB2 contributing one α-helix, while two α-helices define the vertices of the pseudo-threefold arrangement (Extended Data Fig. 1a).

The HID motif interlocks with LID in an intricate arrangement, reminiscent of a paperclip; 2 α-helices from LID and one from HID pack against a β-sheet formed by insertion of one HID between two LID β-strands (Fig. 1c, Extended Data Fig. 2b, PDB ID: 7ACW). This novel structural motif locks SLP_L and SLP_H in a tight heterodimer, providing a structural basis for the stability of the H/L complex[6].

In order to identify residues essential for interaction of the two subunits, we analyzed H/L complex formation in an ELISA-based assay with a panel of individual point mutants (Fig. 1d, comprehensive list of tested mutants in Extended Data Fig. 2, see Methods for details). Mutations of a single amino acid within LID (F274A) or HID (Y27A) were sufficient to destabilize the H/L complex. Moreover, expression of either point mutant in an SlpA-null background resulted in SLP_L shedding from the cell surface of *C. difficile* and detection of a

fraction of SLP_H in the culture supernatant. Loss of SLP_L also resulted in partial degradation of SLP_H (Fig. 1e); N-terminal sequencing revealed truncation of the HID, indicating that this region is unstable in the absence of the LID/HID interaction (Fig. 1e).

Our structural model shows that SLP_L protrudes from the interacting motif, with D1 closest to the SLP_H plane and D2 extending outwards at an angle of ~120°, away from the long axis of D1. Whilst D1 is well ordered, formed by a 5-strand β-sheet packed against two α-helices, D2 is predominantly composed of long, flexible loops, particularly at the externally-exposed face (Fig. 1b); it is characterized by high B-factors (Extended data Fig. 1b) and weaker electron density. Structural flexibility accommodates high sequence variability observed across different strains of *C. difficile*, with 13 SlpA cassette types (SLCTs) currently identified[17]. Residues with high B-factor map almost identically to sequence variation hotspots in D2 (Extended Data Fig. 1b, 1c).

Conformational flexibility in the organization of SlpA is further demonstrated by different arrangements observed in the structure of a truncated derivative of SlpA (SLP_L/HID) and the H/L complex. In the structure of the truncated SLP_L/HID complex (PDB ID: 7ACV), the D1-D2 domains exhibit an orientation relative to the interacting domains different from that seen in the corresponding H/L complex (R7404 strain, SLCT-7b, PDB ID: 7ACX) (Extended Data Fig. 2a). Our models indicate the presence of a hinge, formed by the D1-LID linker (Extended Data Fig. 2a, SI). Analysis of interdomain dynamics also reveals increased mobility within D2. The fact that we observed this conformational flexibility in our crystal models, with no apparent effect on the fold of individual regions, suggests how the effector domains of other CWPs inserted in the functional S-layer can be accommodated by flexible rearrangement of D2 (see SI discussion).

158

159 **Crystal lattice reflects *in situ* S-layer assembly**

160 Due to the natural tendency of SlpA to form 2D crystal arrays, we hypothesized that the
161 packing of our crystal structures might reflect the *in situ* S-layer arrangement. Two H/L
162 complexes, related by pseudo-twofold symmetry, are present in the P1 asymmetric unit,
163 packed in a 2D planar array. The 2D lattices are then stacked to extend the crystal into the
164 third dimension (Extended Data Fig. 3a). The 2D lattice is achieved by tiling of SLP_H, with
165 interlocked ridges of SLP_L molecules covering gaps between the tiles, creating a tightly
166 packed layer (Fig. 2). Lattice contacts between CWB2 motifs of neighbouring SLP_H molecules
167 involve helix-helix interactions between the symmetry-related copies of helix 12 (see
168 topology in Extended Data Fig. 1), as well as electrostatic interactions generating a tightly
169 bonded network (Fig. 2a and Extended Data Fig. 4). The charge distribution generated by
170 the trimeric arrangement of the CWB2s provides complementary charges across the lateral
171 faces of the SLP_H triangular prism tile (Extended Data Fig. 3b and 5a), allowing these
172 interactions to be established (Extended Data Fig. 4). The pseudo-threefold organization of
173 the CWB2s that define the CWP family is also seen in other minor constituents of the S-layer
174 whose structures have been determined[18]. Analysis of the charge distribution in Cwp6 and
175 Cwp8 CWB2s (Extended Data Fig. 5) indicates that charge complementarity could play a role
176 in interaction between lateral faces of CWB2s triangular prisms from different CWPs and
177 SLP_H within a mature S-layer. The environment- and cell-facing sides of SLP_H exhibit
178 considerable charge differences, with a mostly negatively charged external surface and a
179 largely non-polar cell wall-facing base, decorated by positive patches (Extended Data Fig.
180 5a). The positive patches at the cell-wall base could provide the mechanism for anchoring

181 SlpA to the cell wall via interactions with the anionic secondary cell wall polymer PSII[19].

182

183 Other S-layer proteins have been proposed to create relatively permeable arrays[8,11], with
184 pores ranging from ~30 to up to 80 Å and possibly wider[5,10,20]. In contrast, tiling of SLP_H and
185 the SLP_L ridges generates a compact lattice apart from two distinct pores in the *C. difficile* S-
186 layer array. Pore 1 (Fig. 3b, between molecules 1 and 3) is approximately 20 Å across the
187 widest point at the environment-facing surface and is partially occluded by the LID/HID
188 motif narrowing it down to an 11 Å wide cavity. The interlocked D2 domains of adjacent
189 SLP_L molecules cap this pore, further reducing access from the external environment to the
190 cell wall (Fig. 3b, molecules 1 and 3, and Extended Data Fig. 3c). The second pore, formed
191 between two SLP_H (Fig. 3b, molecules 1 and 2) is fully accessible at both faces of the layer
192 and has a width of approximately 11 Å (pore 2, Fig. 2b), but is narrowed to 8 Å by two
193 pseudo-symmetry equivalent arginine residues within 10 Å of the pore outward side
194 (Extended data Fig. 3c).

195

196 We next investigated if the planar crystal packing observed in the X-ray structure reflects
197 the *in situ* packing of the native S-layer assembly. Intact S-layer extracted from *C. difficile*
198 vegetative cells (native S-layer ghosts) formed collapsed capsules. These double-layered 2D
199 crystals were interrogated by electron crystallography, with rotationally separated
200 diffraction patterns observed from images of the superimposed layers (Fig. 2c). As we had
201 hypothesized, the *p*2 symmetric 2D lattices of native S-layer ghosts (a = b = 85 Å, γ = 100°)
202 were consistent with unit cell parameters of the stacked lattices in the 2D plane of the X-ray
203 crystals (b = 78 Å, c = 80 Å, α = 100°), pointing towards a similar packing arrangement (Fig.
204 2c, Extended Data Fig. 6a). The 3D reconstruction from images of the native ghosts revealed

a molecular envelope with a staggered ridged surface on one face of the S-layer, with deep grooves between the parallel ridges, and an opposing surface defined by paired, globular domains arranged in rows (Fig. 2d). These features recapitulate the surface characteristics of the H/L array in the X-ray crystal structure, with the ridge-like arrangement of the SLP_L above the SLP_H globular rows. Indeed, manual fitting of the 2D X-ray lattice as a single rigid body into the EM density matches the ridged surface to SLP_L, with the paired globular domains on the opposite face corresponding to the SLP_H CWB2 motifs (Fig. 2d). The ridges are also observed in cryo-electron microscopy (cryo-EM) side views of intact cells (Fig. 2c). This confirms that the X-ray crystal lattice of the H/L complex has the same overall arrangement as the *in situ* lattice of a mature S-layer in intact cells, therefore establishing our crystallographic model as a template for interrogating S-layer assembly in *C. difficile*. To our knowledge, this is the first time that X-ray structural models of a full-length S-layer protein reflect the observed S-layer assembly *in situ*.

Probing the S-layer assembly model

In order to further test our S-layer assembly model, we sought to generate an altered, yet functional S-layer. The observed interactions for assembly involve mainly the SLP_H tiles and D1 in SLP_L. Moreover, these regions are conserved across different SLCTs (Extended data Fig. 4 and SI). We therefore hypothesized that the structurally flexible and less conserved D2 domain (Extended data Fig. 1b, 1c) might be dispensable for maintaining S-layer assembly and engineered a mutant strain devoid of D2 - RΔD2 (producing $SlpA_{R\Delta D2}$; see Methods). The X-ray crystal structure of the H/L complex from $SlpA_{R\Delta D2}$ (Fig. 3c and Extended Data Fig. 7, PDB ID: 7ACZ) superimposes readily onto the full-length model (backbone RMSD 1.09 Å), with the absence of D2 not perturbing the overall protein fold. Moreover, the crystal lattice

is similar to wild type, with equivalent interactions between SLP_H tiles and D1 domains (Extended Data Fig. 7). Importantly, the absence of D2 exposes pore 1 between SLP_H tiles (Fig. 3d and Extended Data Fig. 7c), which is occluded by interlocking D2 domains in the full-length structure. This creates two openings in the array of about 16 Å, with potential functional implications as it could indicate a more permeable S-layer than in the wild type structure, with twice as many pores, of slightly increased size.

Analysis of native S-layer ghosts from bacteria producing the $SlpA_{R\Delta D2}$ revealed that, despite lacking nearly half of SLP_L (145 of 318 amino acids), an S-layer with identical lattice parameters to the wild type is still formed. The calculated molecular envelope retains the paired globular domain organization (Fig. 3a) but the opposing face lacks the staggered ridge feature seen in the wild type EM reconstruction (Fig. 3a), confirming our assignment of this missing density to the D2 domain. This is further validated by difference maps of cryo-EM projections for full-length $SlpA_{R20291}$ and $SlpA_{R\Delta D2}$ showing a region of significant difference density matching the position of D2 in the complete structure (Extended Data Fig. 6).

Together, the structural models of $SlpA_{R\Delta D2}$ and the corresponding S-layer reconstruction confirm our model for S-layer assembly, where SLP_H tiling and SLP_L D1 domains provide the key contacts for paracrystalline array formation.

An altered S-layer surface reduces *C. difficile* virulence

The retention of S-layer integrity, despite loss of an exposed structural domain, provided an unprecedented opportunity to directly assess the role of SlpA in *C. difficile* infection. We employed the mouse model of acute disease, which allows a nuanced analysis of colonization and pathology, typified by weight loss and caecal and colon inflammation[21].

253 Animals infected with the wild type strain R20291 (producing $SlpA_{WT}$) lost on average 8% of
254 their body weight within 24 h of infection and 12% by the peak of infection at 48 h, before
255 returning to pre-infection weights approximately 4 days after infection. In contrast,
256 infection with RΔD2 (producing $SlpA_{RΔD2}$) resulted in little apparent disease; animals
257 displayed only a 1% weight loss after 24 h and 5% after 48 h, before a gradual return to pre-
258 infection weight (Fig. 4a). As RΔD2 was derived from our previously characterized *slpA*
259 mutant FM2.5[15], a control for any potential background genetic variation was needed. A
260 previously characterized strain FM2.5RW[15] contains a watermarked copy of the wild type
261 *slpA* gene (encoding $SlpA_{RW}$) and RΔD2 was constructed in a similar way but contains a
262 truncated version of *slpA* (see Methods for details). Animals infected with FM2.5RW strain
263 showed similar patterns of disease as those infected with wild type.
264 To determine if the surprising loss of virulence seen for RΔD2 could be attributed to
265 impaired colonization, we quantified *C. difficile* present in faeces collected each day post-
266 infection (Fig. 4b) and in intestinal contents post-mortem (Extended Data Fig. 8). No
267 statistically significant differences in bacterial recovery were observed, demonstrating that,
268 although D2 is surface-exposed in the S-layer, it is not required for efficient colonization.
269 RΔD2 sporulated normally *in vitro* (Extended Data Fig. 9a), but displayed an increased
270 sensitivity to lysozyme in comparison to the wild type strain R20291 (Extended Data Fig. 9b).
271 As the perceived dogma is that the intestinal pathology and symptoms associated with CDI
272 are largely toxin-driven, we next examined toxin expression and activity *in vitro* (Extended
273 Data Fig. 9c, 9d) and in intestinal contents (Fig. 4c and Extended Data Fig. 8f). Surprisingly,
274 no toxin production defect was observed, suggesting the decrease in disease severity seen
275 with RΔD2 is a direct consequence of the loss of the surface-exposed D2 domain. Strikingly,
276 although toxin activity was equivalent in all strains, a reduced level of epithelial damage was

observed in tissue from RΔD2-infected mice. Indeed, assessment of other markers including the extent of inflammatory infiltrate, tissue edema and crypt length measurements indicated that damage was reduced in RΔD2 infected mice when compared with animals infected with WT or control strain at the acute point of infection (48 h post-infection; Fig. 4d and Extended Data Fig. 8c).

Together, these observations demonstrate that the S-layer contributes directly to *C. difficile* disease severity, in a toxin-independent manner. Importantly, our data also reveal that the most surface-exposed domain of SlpA is dispensable for effective colonization, suggesting that other regions of the protein or other CWPs are likely to be involved in direct host-pathogen interactions required for colonization.

Discussion

Here we report the first experimentally determined structure of a complete S-layer protein from a medically important human pathogen, which allows us to observe the organization of the paracrystalline array at an atomic level. The tight packing of H/L complexes in the crystal replicates assembly into the functional S-layer which we observe *in situ* by electron microscopy. The repeating crystallographic array is created by tiling of one of the subunits (SLP_H), which also anchors the S-layer to the cell wall. Most other S-layers characterized to date are composed of distinct domains that contribute either to lattice formation (crystallization domain) or cell surface attachment (anchoring domain)[2]. In the *C. difficile* S-layer, the crystallization and anchoring functions are combined in the SLP_H, with assembly relying on contacts between adjacent tiles. Moreover, contacts between SLP_L and neighbouring H/L complexes further expand the S-layer assembly network. The SLP_L ridges are also important for generating a structure impermeable to the majority of folded

301 proteins such as lysozyme and other large molecules as it covers the pores present within
302 the packing of the triangular prism SLP_H tiles. This tight packing raises the question of how
303 large molecules such as the *C. difficile* toxins[22] are exported to the environment.
304 Furthermore, S-layers must be able to accommodate cell growth and division and this array
305 needs to adapt to the curvature of the cell poles. Having tightly packed core subunits or
306 domains, maintained by interchangeable electrostatic interactions, that are then decorated
307 with more flexible regions is a simple, yet seemingly effective, way to achieve both
308 requirements. Points of mismatched symmetry as observed by tomography (Extended Data
309 Fig. 10) could provide increased flexibility and permeability, creating discrete points for
310 export of larger molecules (see SI Discussion). Incorporation of other CWPs within the
311 crystalline array involves interactions of the conserved CWB2 domains that define *C. difficile*
312 cell wall proteins[2], while the flexibility of the D2 domain allows packing of other effector
313 CWP motifs present in the minor constituents of the S-layer (see SI discussion).
314 Our investigation of the functional role of the S-layer revealed that toxins are necessary but
315 not sufficient for full disease severity, a paradigm shift in our understanding of *C. difficile*
316 infections. Despite being dispensable for protein fold or even S-layer assembly, D2 seems to
317 confer a functional role to SLP_L and its absence leads to reduced disease severity.
318 Surprisingly, this is not due to changes in colonization of the gut, suggesting other domains
319 or S-layer proteins are involved in this type of interaction with the host. Instead, the
320 presence of D2 is associated with increased levels of inflammation when compared to the
321 full-length protein (Fig. 4). In *C. difficile* infections, tissue inflammation has been associated
322 with activation of additional immune pathways that results in enhanced disease[23]. The high
323 sequence variability and structural flexibility of the D2 domain, in contrast to the conserved
324 and relatively rigid SLP_H, could therefore confer an immune-evasion mechanism as a result

of the evolutionary pressure of the dynamic environment of the gut. It is therefore tempting to speculate that D2 is directly involved in the activation of the host immune response, however, the molecular mechanisms involved remain to be elucidated (see SI discussion). Importantly, we have established a direct link between the S-layer and disease severity and our characterization of S-layer assembly in *C. difficile* reveals new potential therapeutic avenues. The interacting SLP_H subunits and the flexible D2 domains present key targets for disruption of the S-layer, and molecules that affect S-layer assembly are attractive therapeutic agents.

334 **References**

335 1 Sleytr, U. B., Schuster, B., Egelseer, E. M. & Pum, D. S-layers: principles and
336 applications. *FEMS Microbiol Rev* **38**, 823-864, doi:10.1111/1574-6976.12063 (2014).
337 2 Fagan, R. P. & Fairweather, N. F. Biogenesis and functions of bacterial S-layers. *Nat*
338 *Rev Microbiol* **12**, 211-222, doi:10.1038/nrmicro3213 (2014).
339 3 Smits, W. K., Lyras, D., Lacy, D. B., Wilcox, M. H. & Kuijper, E. J. *Clostridium difficile*
340 infection. *Nat Rev Dis Primers* **2**, 16020, doi:10.1038/nrdp.2016.20 (2016).
341 4 Baranova, E. *et al.* SbsB structure and lattice reconstruction unveil Ca2+ triggered S-
342 layer assembly. *Nature* **487**, 119-122, doi:10.1038/nature11155 (2012).
343 5 Bharat, T. A. M. *et al.* Structure of the hexagonal surface layer on *Caulobacter*
344 *crescentus* cells. *Nat Microbiol* **2**, 17059, doi:10.1038/nmicrobiol.2017.59 (2017).
345 6 Fagan, R. P. *et al.* Structural insights into the molecular organization of the S-layer
346 from *Clostridium difficile*. *Mol Microbiol* **71**, 1308-1322, doi:10.1111/j.1365-
347 2958.2009.06603.x (2009).
348 7 Fioravanti, A. *et al.* Structure of S-layer protein Sap reveals a mechanism for
349 therapeutic intervention in anthrax. *Nat Microbiol* **4**, 1805-1814, doi:10.1038/s41564-019-
350 0499-1 (2019).
351 8 Gambelli, L. *et al.* Architecture and modular assembly of *Sulfolobus* S-layers revealed
352 by electron cryotomography. *Proc Natl Acad Sci U S A* **116**, 25278-25286,
353 doi:10.1073/pnas.1911262116 (2019).
354 9 Sychantha, D. *et al.* Molecular basis for the attachment of S-layer proteins to the cell
355 wall of *Bacillus anthracis*. *Biochemistry* **57**, 1949-1953, doi:10.1021/acs.biochem.8b00060
356 (2018).
357 10 von Kugelgen, A. *et al. In situ* structure of an intact lipopolysaccharide-bound

358 bacterial surface layer. *Cell* **180**, 348-358 e315, doi:10.1016/j.cell.2019.12.006 (2020).
359 11 Pum, D., Toca-Herrera, J. L. & Sleytr, U. B. S-layer protein self-assembly. *Int J Mol Sci*
360 **14**, 2484-2501, doi:10.3390/ijms14022484 (2013).
361 12 Ng, J. *et al. Clostridium difficile* toxin-induced inflammation and intestinal injury are
362 mediated by the inflammasome. *Gastroenterology* **139**, 542-552, 552 e541-543,
363 doi:10.1053/j.gastro.2010.04.005 (2010).
364 13 Cowardin, C. A. *et al.* Inflammasome activation contributes to interleukin-23
365 production in response to *Clostridium difficile*. *mBio* **6**, doi:10.1128/mBio.02386-14 (2015).
366 14 Ryan, A. *et al.* A role for TLR4 in *Clostridium difficile* infection and the recognition of
367 surface layer proteins. *PLoS Pathog* **7**, e1002076, doi:10.1371/journal.ppat.1002076 (2011).
368 15 Kirk, J. A. *et al.* New class of precision antimicrobials redefines role of C*lostridium*
369 *difficile* S-layer in virulence and viability. *Sci Transl Med* **9**,
370 doi:10.1126/scitranslmed.aah6813 (2017).
371 16 Calabi, E. *et al.* Molecular characterization of the surface layer proteins from
372 *Clostridium difficile*. *Mol Microbiol* **40**, 1187-1199 (2001).
373 17 Dingle, K. E. *et al.* Recombinational switching of the *Clostridium difficile* S-layer and a
374 novel glycosylation gene cluster revealed by large-scale whole-genome sequencing. *J Infect*
375 *Dis* **207**, 675-686, doi:10.1093/infdis/jis734 (2013).
376 18 Usenik, A. *et al.* The CWB2 cell wall-anchoring module Is revealed by the crystal
377 structures of the *Clostridium difficile* cell wall proteins Cwp8 and Cwp6. *Structure* **25**, 514-
378 521, doi:10.1016/j.str.2016.12.018 (2017).
379 19 Willing, S. E. *et al. Clostridium difficile* surface proteins are anchored to the cell wall
380 using CWB2 motifs that recognise the anionic polymer PSII. *Mol Microbiol* **96**, 596-608,
381 doi:10.1111/mmi.12958 (2015).

20 Farci, D. *et al.* Structural insights into the main S-layer unit of *Deinococcus radiodurans* reveal a massive protein complex with porin-like features. *J Biol Chem* **295**, 4224-4236, doi:10.1074/jbc.RA119.012174 (2020).

21 Jukes, C. A. *et al.* Bile salt metabolism is not the only factor contributing to *Clostridioides (Clostridium) difficile* disease severity in the murine model of disease. *Gut Microbes* **11**, 481-496, doi:10.1080/19490976.2019.1678996 (2020).

22 Pruitt, R. N. & Lacy, D. B. Toward a structural understanding of *Clostridium difficile* toxins A and B. *Front Cell Infect Microbiol* **2**, 28, doi:10.3389/fcimb.2012.00028 (2012).

23 Saleh, M. M. & Petri, W. A., Jr. Type 3 Immunity during *Clostridioides difficile* Infection: Too Much of a Good Thing? *Infect Immun* **88**, doi:10.1128/IAI.00306-19 (2019).

Acknowledgements

FV, ABS, JAK were supported by a Wellcome Trust Collaborative Award (204877/Z/16/Z) awarded to GRD, PSS, RPF. JW was supported by a BBSRC grant (BB/P02002X/1) awarded to PAB and RPF. PLM, OB and SOB were supported by an FMS Newcastle PhD studentship, a University of Sheffield Imagine: Imaging Life initiative PhD studentship and a University of Sheffield PhD studentship, respectively. We thank Diamond Light Source for access to beamlines I23, I04, and I24 (lb15523-6, mx18598-24, mx18598-14, through the "Macromolecular Crystallography at Newcastle, Durham, Lincoln and Durham" BAG) and eBIC (through the EM19832 "Northern England Cryo-EM consortium" BAG). We thank Dr. Svetomir Tzokov for supporting EM data collection in the University of Sheffield Electron Microscopy Facility and Dr. Giuseppe Cannone for help with EM data collection at the MRC Laboratory of Molecular Biology, Cambridge. The contents of this work are solely the responsibilities of the authors and do not reflect the official views of any of the funders, who had no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Author contributions

OB, JW carried out EM experiments, collected and analyzed data, revised the manuscript. PLM, ABS carried out X-ray crystallographic experiments, collected and analyzed X-ray data, determined and refined X-ray structural models, wrote and revised the manuscript. ABS carried out ELISA experiments and analyzed data. FV carried out *in vivo* experiments, analyzed data, wrote and revised manuscript. JAK, SOB carried out *in vitro* experiments, analyzed the data and revised the final manuscript. ABS, JAK, OB, RPF and SOB constructed plasmids and strains. AB collected, analyzed X-ray data and supervised. KEO, AW collected

and analyzed the S-SAD data and revised the final manuscript. NF designed and supervised initial structural studies, revised the manuscript. GRD designed, supervised and carried out *in vivo* experiments, analyzed the data, wrote and revised the manuscript. PAB designed, supervised and analyzed EM studies, wrote and revised the manuscript. RPF designed the study, prepared samples, collected X-ray data, analyzed data, supervised the study, wrote and revised the manuscript. PSS designed the study, prepared samples, collected, analyzed and determined X-ray structures, interpreted data, supervised the study, wrote and revised the manuscript.

Competing interest declaration

The authors have no competing interests as defined by Nature Research, or other interests that might be perceived to influence the interpretation of the article.

Additional information

Supplementary Information is available for this paper:

SI Discussion, SI Tables, SI Data, SI files (movie).

Correspondence and requests for materials should be addressed to:

Paula S. Salgado, Biosciences Institute, Faculty of Medical Sciences, Newcastle University, UK. paula.salgado@ncl.ac.uk

or

Robert P. Fagan, Department of Molecular Biology and Biotechnology, University of Sheffield, UK. r.fagan@sheffield.ac.uk

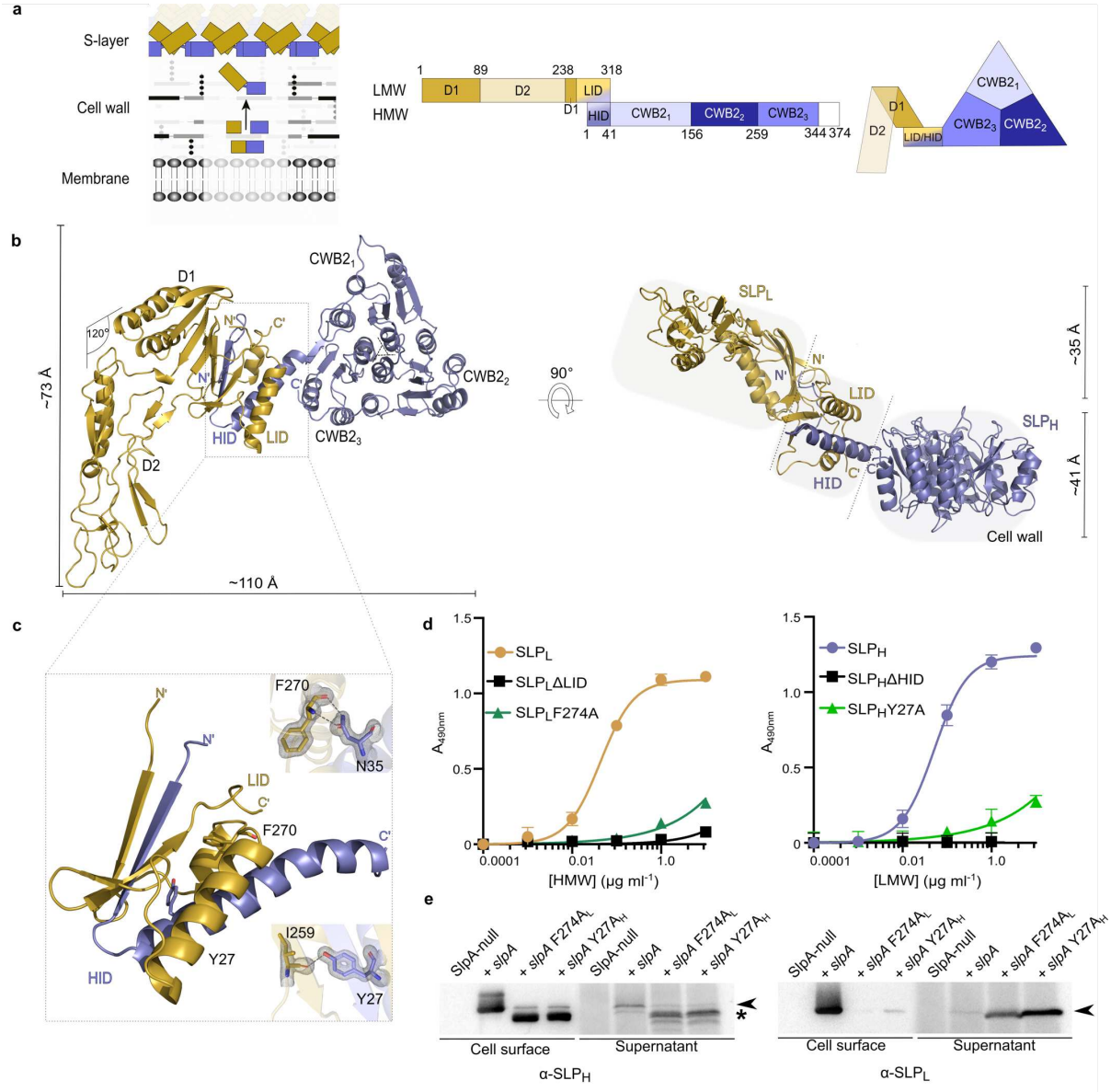


Fig. 1: Architecture and key interactions in *C. difficile* SLP$_H$/SLP$_L$ (H/L) complex

a, SlpA arrangement on the cell surface (left; SLP$_L$ colored in gold and SLP$_H$ in slate blue) with detailed organization of protein building blocks in its primary sequence (middle) and quaternary structure (right). Numbering based on the subunits of SlpA from strain CD630, S-layer cassette type 7 (SCLT-7), PBD ID: 7ACY.

b, Cartoon representation of H/L complex as viewed from the external environment (top view, left) and side (right). The SLP$_L$ protrudes above the SLP$_H$ subunit, creating a two-plane

arrangement. Three distinct structural features are observed: SLP_H, D1 and D2, and LID/HID (regions highlighted in gray).

c, 'Paperclip' organization of the interacting domains LID/HID is maintained by a range of interactions, with selected interface residues identified in strain R7404 (SCLT-7b, PDB ID: 7ACW) depicted as sticks. 2m*Fo*-D*Fc* electron density map is shown on the interacting amino acid pairs as a grey mesh contoured at 1.5 σ. Specific interatomic interactions identified with PDBePISA are represented as a dashed line (more details on Extended Data Fig. 2).

d, Probing of CD630 H/L complex interactions *in vitro* with ELISA, comparing effects of intact SLP_L (gold circles), SLP_H (slate blue circles), variants lacking interacting domains (black squares) and substitution mutants of F274A (structurally equivalent to F270 in R7404 LID/HID depicted in **c**, dark green triangles) and Y27A (light green triangles) on H/L complex formation. Graphs represent mean ± standard deviation (SD) of n = 3 experiments, with least squares curve fit of product formed upon interaction of the two subunits.

e, Western blot of cell surface extracts and culture supernatants, detecting (black arrowhead) SLP_H (left) and SLP_L (right) in strains devoid of endogenous *slpA* and expressing plasmid-borne $SlpA_{CD630}$ native protein or variants with either $F274A_L$ or $Y27A_H$ substitution mutants in SLP_L or SLP_H (denoted in subscript), respectively. Detected product of partial degradation of HMW indicated with an asterisk.

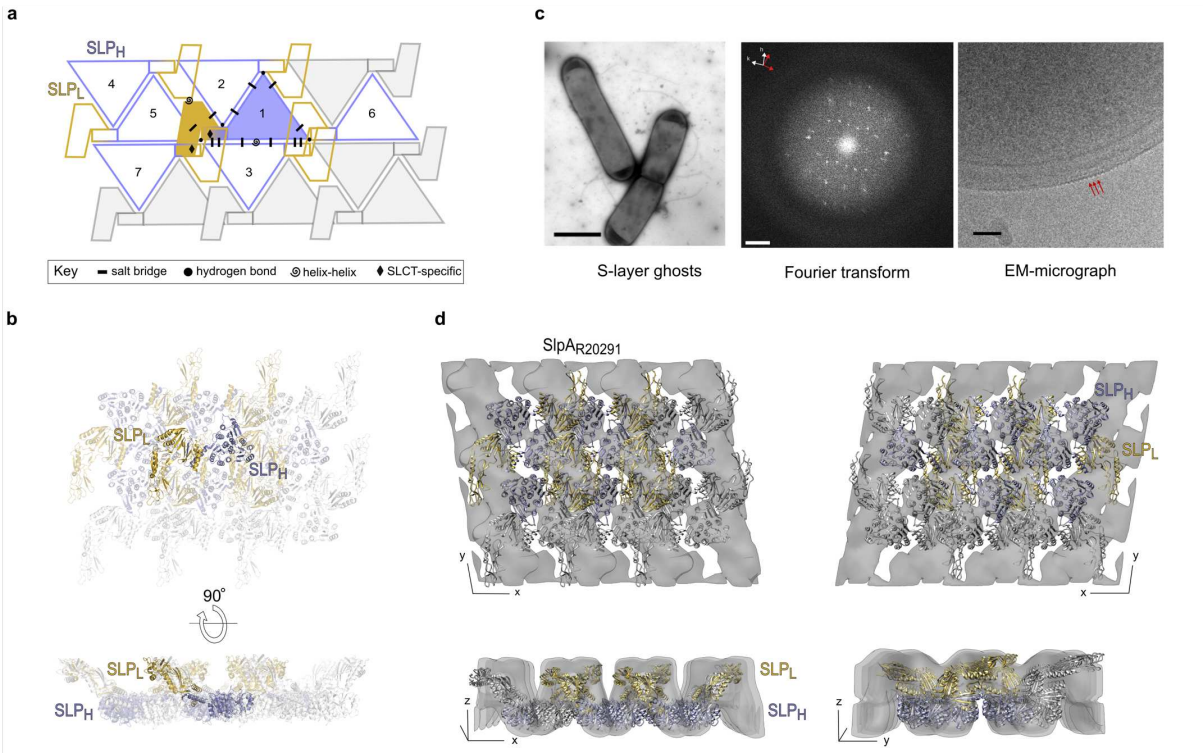


Fig. 2: Planar crystal packing in the X-ray structure fits the *in situ* packing of the native S-layer

a, 2D schematic of H/L complex crystal packing, indicating the interaction network linking a single SLP_L (gold)/ SLP_H (slate blue) complex with six other molecules in a planar arrangement generated by SLP_H tiling. Array is depicted as seen from the extracellular environment, with symbols representing key interaction types in the crystal lattice, shown in detail in Extended Data Fig. 4.

b, Cartoon representation of the H/L planar array (PDB ID 7ACY, colored as in **a**, views as defined in **Fig. 1b**).

c, Native *C. difficile* S-layer ghosts (electron micrograph, negatively stained, left. Scale bar: 2 µm) were used to compute Fourier transforms (middle). Typically spots from two or more lattices were observed. Reciprocal lattice axes (red and white axes) are indicated for two observed lattices (scale bar 0.0125 $Å^{-1}$). Intact frozen hydrated *C. difficile* cells, examined by

cryo-electron microscopy (right), show distinctive ridged surface indicated by red arrows (scale bar 50 nm).

d, Orthogonal views of the 3D reconstruction of negatively stained S-layer ghost indicating the overall envelope in the native lattice. A rigid body fit of the structure of H/L complex determined by X-ray crystallography (PDB ID: 7ACY, cartoon representation, SLP_L - gold, SLP_H - slate) indicates a similar arrangement in the native S-layer ghosts and crystal packing. Reconstruction is shown from the environment (top left) and cell wall (top right), and side views in the 2D plane (bottom panels).

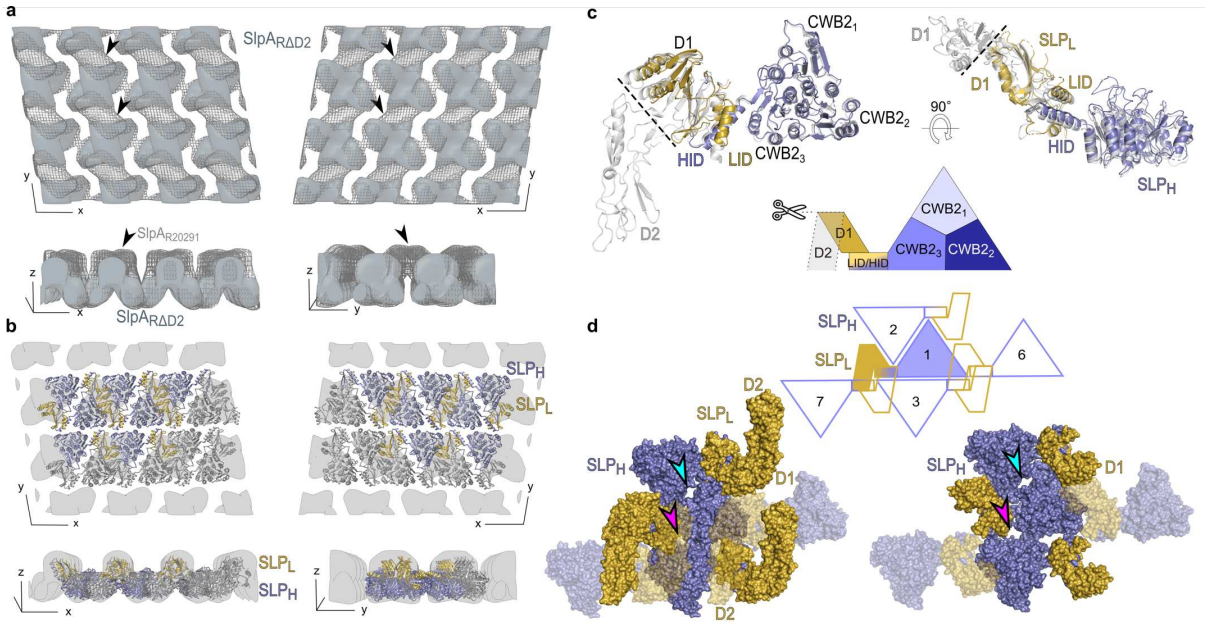


Fig. 3: The flexible D2 domain is dispensable for S-layer assembly

a, Superimposition of the 3D reconstruction of negatively stained S-layer ghost containing SlpA devoid of domain D2 (SlpA$_{RΔD2}$, light blue solid surface) on the reconstruction of native wild type S-layer ghost (SlpA$_{R20291}$, grey mesh). The missing density can be largely ascribed to that of the missing D2 domain (indicated with black arrowheads). Views are as described in **Fig. 2d**.

b, Fit of the SlpA$_{RΔD2}$ structure determined by X-ray crystallography (colored as in **c**) into the S-layer (gray) reconstruction indicates a similar arrangement in the crystal packing and the native array. Views as in **a.**

c, Cartoon representation of the SlpA$_{RΔD2}$ H/L complex crystal structure (slate blue and gold, PDB ID: 7ACZ), superimposed onto SlpA$_{CD630}$ H/L complex structure (PDB ID: 7ACY, gray). Deleted D2 region is marked with a dashed line on the CD630 structure and corresponding schematic representation of the complex. Views as in **Fig. 1b**.

d, Surface representation of wild type H/L (7ACY, left) and $SlpA_{R\Delta D2}$ H/L(7ACZ, right) crystal packing showing pores in the 3D crystal lattice (top). Positions of pores marked with arrowheads (pore 1 in magenta, pore 2 in cyan) are equivalent in both lattices.

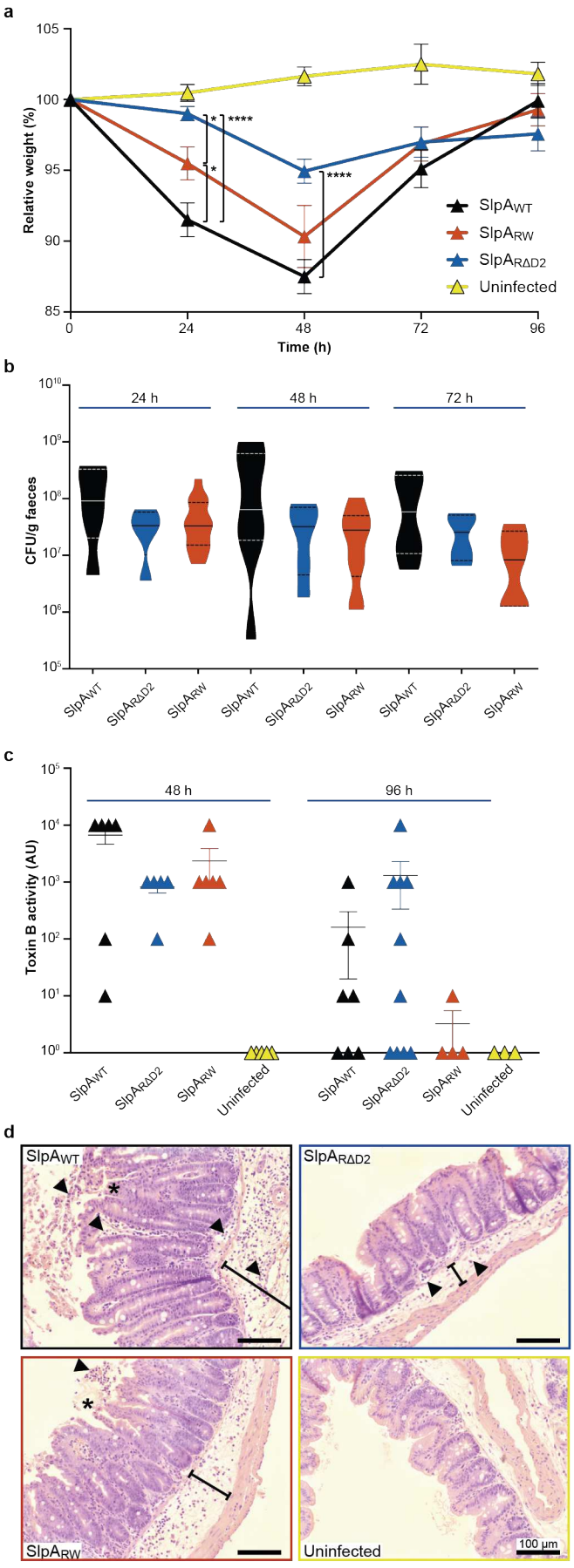


Fig. 4: ***In vivo*** **evaluation of** ***C. difficile*** **strains producing a modified S-layer**

a, Relative percentage weight loss of antibiotic-treated animals infected with R20291 (SlpA$_{WT}$, black triangle), RΔD2 (SlpA$_{RΔD2}$, blue triangle), FM2.5RW (SlpA$_{RW}$, orange triangle) and antibiotic-treated uninfected animals (Uninfected, yellow triangle).

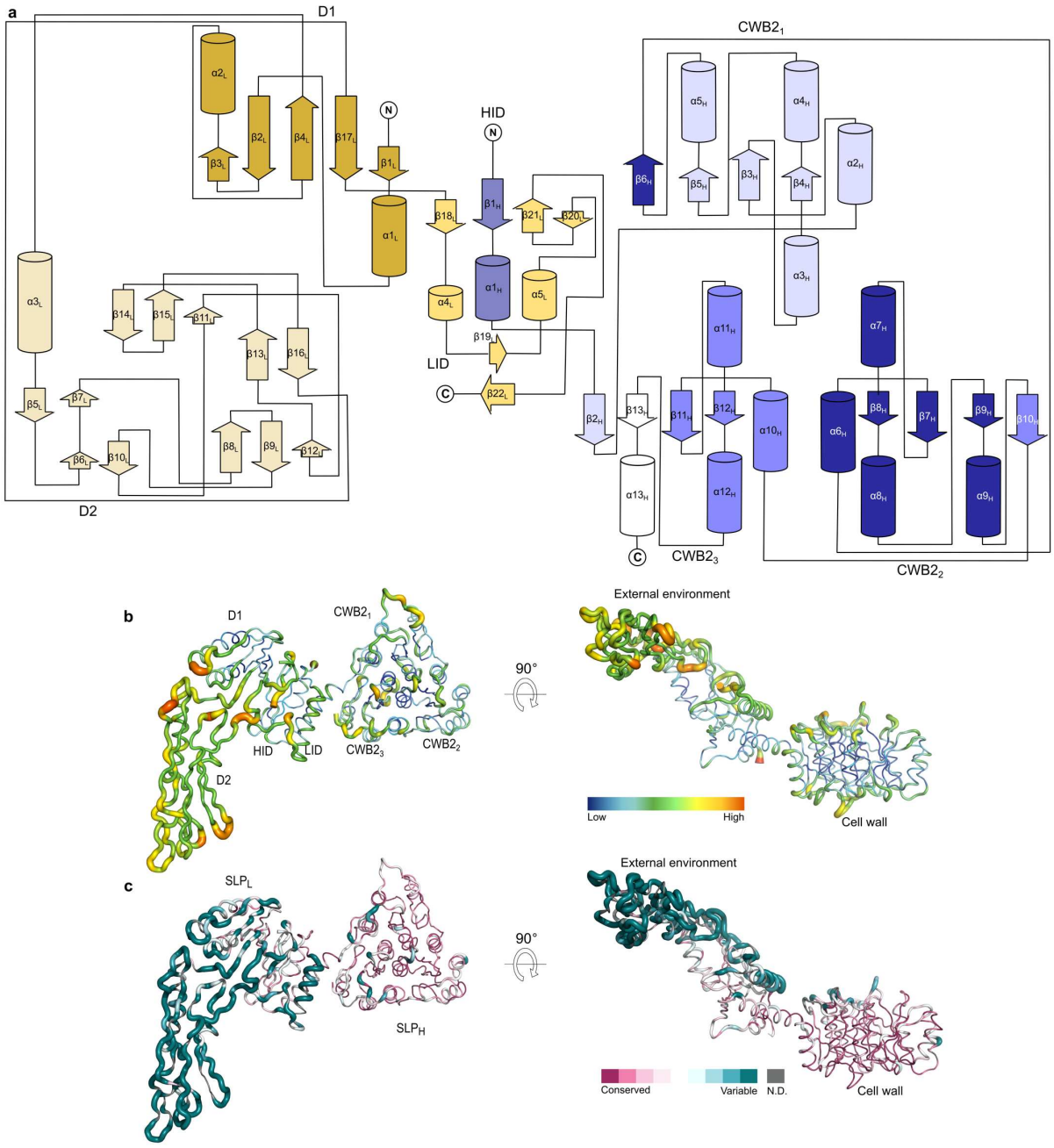
b, Total *C. difficile* counts (CFU ml^{-1}) recovered from faeces at 24, 48 and 72h post infection.

c, Measurement of toxin activity in filtered caecal extracts from individual animals. Annotations as in **a**.

d, Hematoxylin and eosin stained caecal histological sections from animals 48 h post infection with *C. difficile*, representing acute disease. Arrowheads highlight the margination and tissue infiltration of polymorphonuclear cells (PMNs) and the breached epithelial barrier (indicated with asterisk). The extent of

tissue edema between basal membrane and musculature is highlighted (line). Crypt hyperplasia is discernible in tissue from R20291 ($SlpA_{WT}$) and FM2.5RW ($SlpA_{RW}$) infected animals, compared to antibiotic-treated uninfected animals or those infected with RΔD2 ($SlpA_{R\Delta D2}$).

Data analysis represents a total of up to 15 mice per strain tested from 3 experiments expressed as mean ($\pm$ SEM) **a** (10-15 mice), **c** (5 mice); or as violin plots showing medians (solid lines), upper and lower quartile (dashed line) and max/min range (top and base of the plot) **b**, (min 5 mice). Non-parametric, non-paired Mann-Whitney statistical tests were performed with differences of *p = <0.05, ****p = <0.0001 indicated (**a**, **c**). Scale bars of 100mm are shown (**d**).

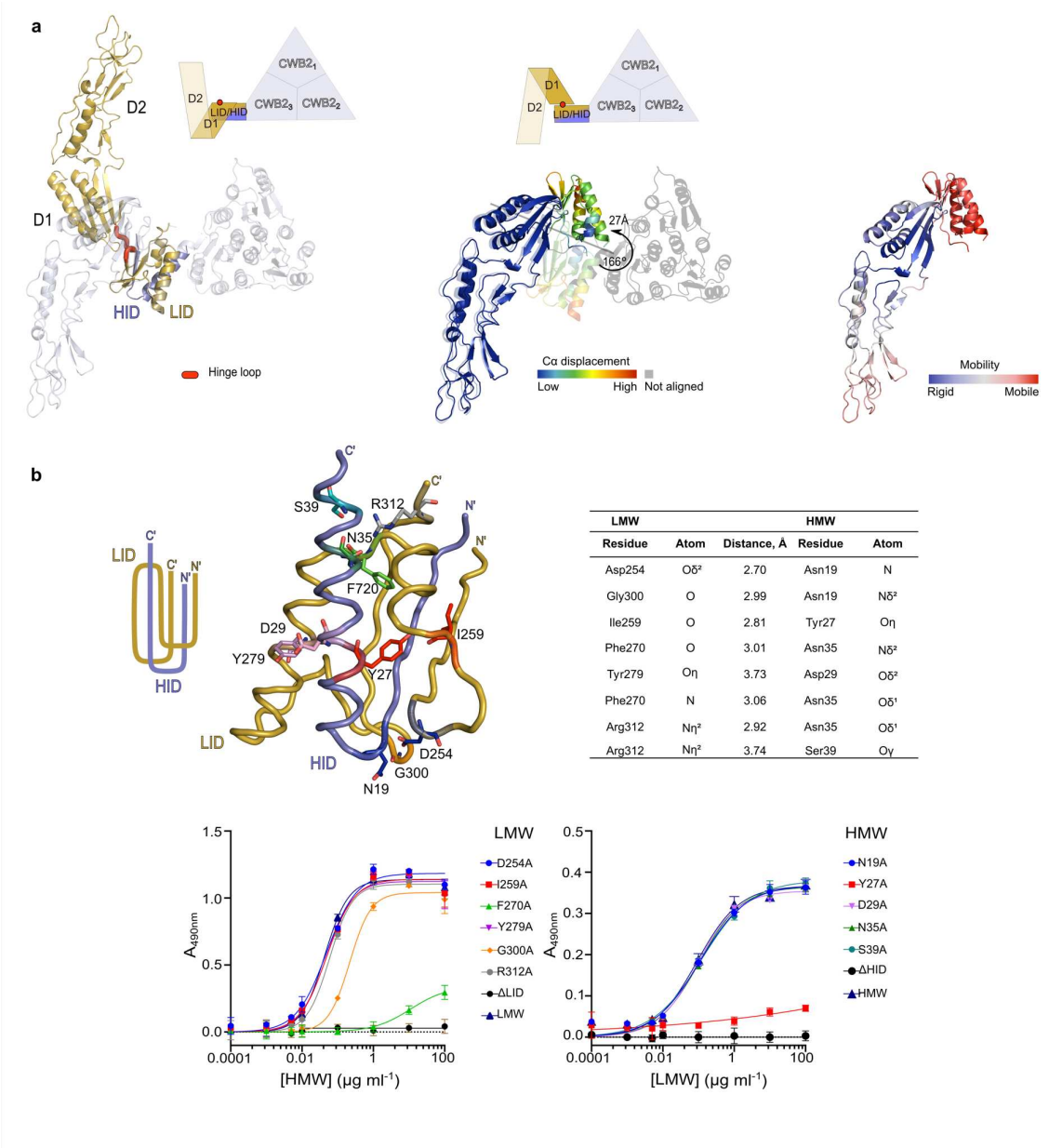


Extended Data Fig. 1: **H/L complex organization and conservation**

a, Topology of the mature $SlpA_{CD630}$ H/L complex, retrieved from structure analysis with PBDsum and Coot. SLP_L represented in gold and SLP_H colored in blue. Color shading represents different domains/motifs. Numbering of secondary structure components based on PDB ID 7ACY, subscripts indicate the relevant subunit.

b, Putty representations of $SlpA_{CD630}$ H/L complex showing B-factors ranging from low (blue and narrow) to high (red and wide). High B-factors are indicative of disorder/flexible regions.

c, Conservation of the SlpA sequence across annotated SlpA Cassette Types (SLCTs) depicted on putty representations of $SlpA_{CD630}$ H/L complex, colored from conserved (purple) to variable (cyan). Conservation was calculated using Consurf web server (see Methods for details).



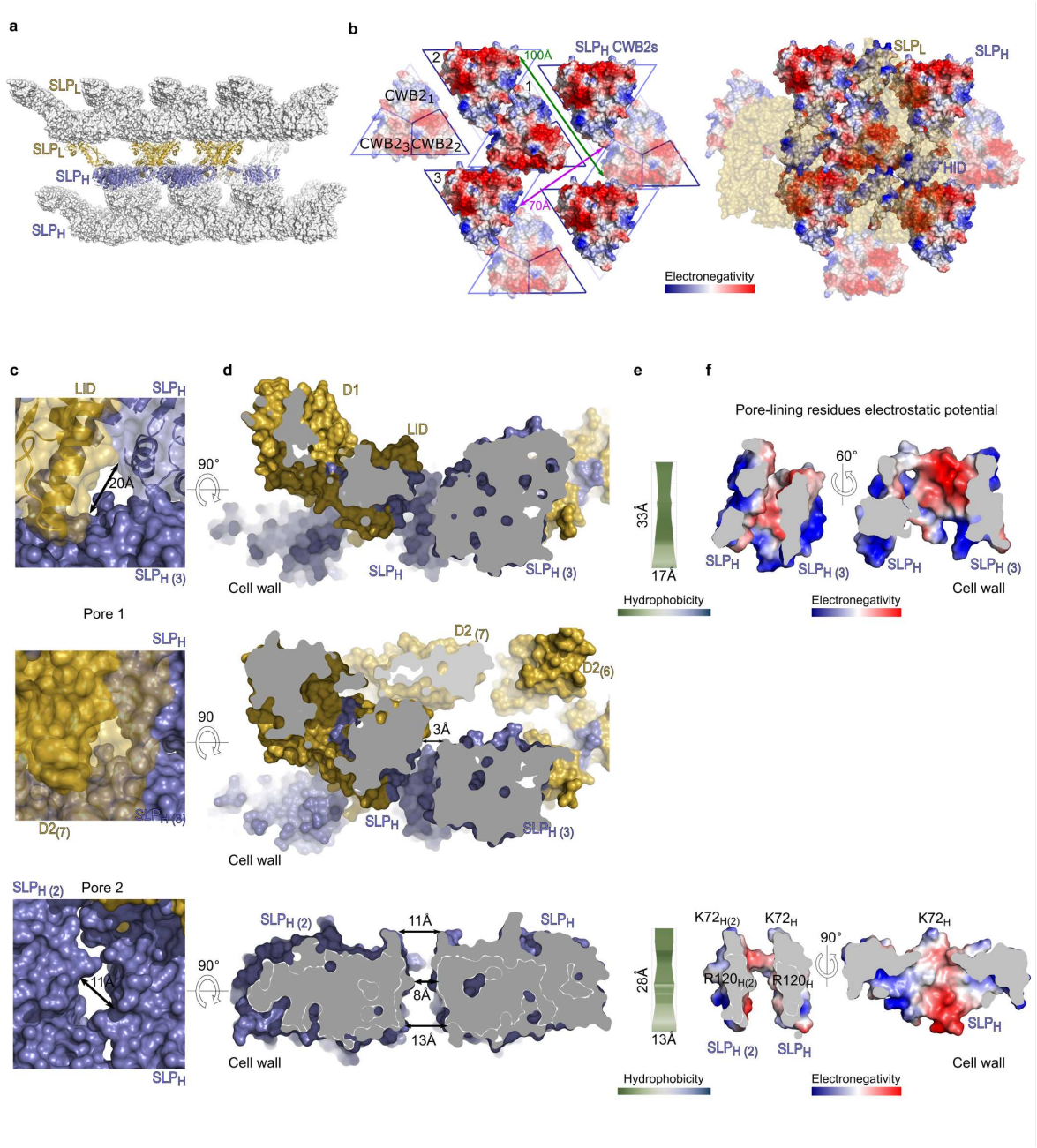
LMW			HMW	
Residue	Atom	Distance, Å	Residue	Atom
Asp254	Oδ²	2.70	Asn19	N
Gly300	O	2.99	Asn19	Nδ²
Ile259	O	2.81	Tyr27	Oη
Phe270	O	3.01	Asn35	Nδ²
Tyr279	Oη	3.73	Asp29	Oδ²
Phe270	N	3.06	Asn35	Oδ¹
Arg312	Nη²	2.92	Asn35	Oδ¹
Arg312	Nη²	3.74	Ser39	Oγ

Extended Data Fig. 2: **Interdomain interfaces of H/L complex**

a, Superimposing structures of SLP_L/HID (gold/slate blue, PDB ID: 7ACV) onto the native complex of $SlpA_{R7404}$ (SLCT-7b, PDB ID: 7ACX) (blue/white) reveals flexibility of the LID-D1 linker, as illustrated by rotation of D1-D2 domains in relation to fixed position of LID/HID motif (left). The hinge loop enabling this conformational flexibility (determined by DynDom6D) is colored in red. The backbone displacement (colored from blue – low, to red – high Cα displacement deviation) is shown on the alignment of D1-D2 region of both structures (middle; SLP_L/HID – opaque, H/L – semi-transparent) with the rotation angle of

the LID/HID motif indicated with an arrow. Structural dynamics (right) of the SLP_L/HID represented as increasing mobility (colored blue – rigid, to red - mobile) calculated based on elastic network models implemented in DynOmics ENM version 1.0 server.

b, Key interactions identified at the interface of the LID/HID complex from strain R7404 (SLCT-7b) (left; identified with PDBePISA in PDB ID: 7ACW) informed site directed mutagenesis for functional assessment by ELISA. Effects of point substitution mutations in SLP_L (middle) or SLP_H (right) on complex formation were tested. Graphs represent mean ± SD of n = 3 experiments, with least squares curve fit.



Extended Data Fig. 3: SLP_H tiling and SLP_L interactions create a tightly packed array

a, 3D crystal packing of SlpA showing the planar layer of a H/L array in slate blue and gold cartoon representation stacked between symmetry-related layers represented as white surface.

b, Tiling of SLP_H CWB2 motifs via charge complementarity across each triangular prism face. Poisson-Boltzmann electrostatic potential calculated for $SlpA_{CD630}$ SLP_H, represented as a charge distribution (positive in blue and negative in red, as per electronegativity gradient

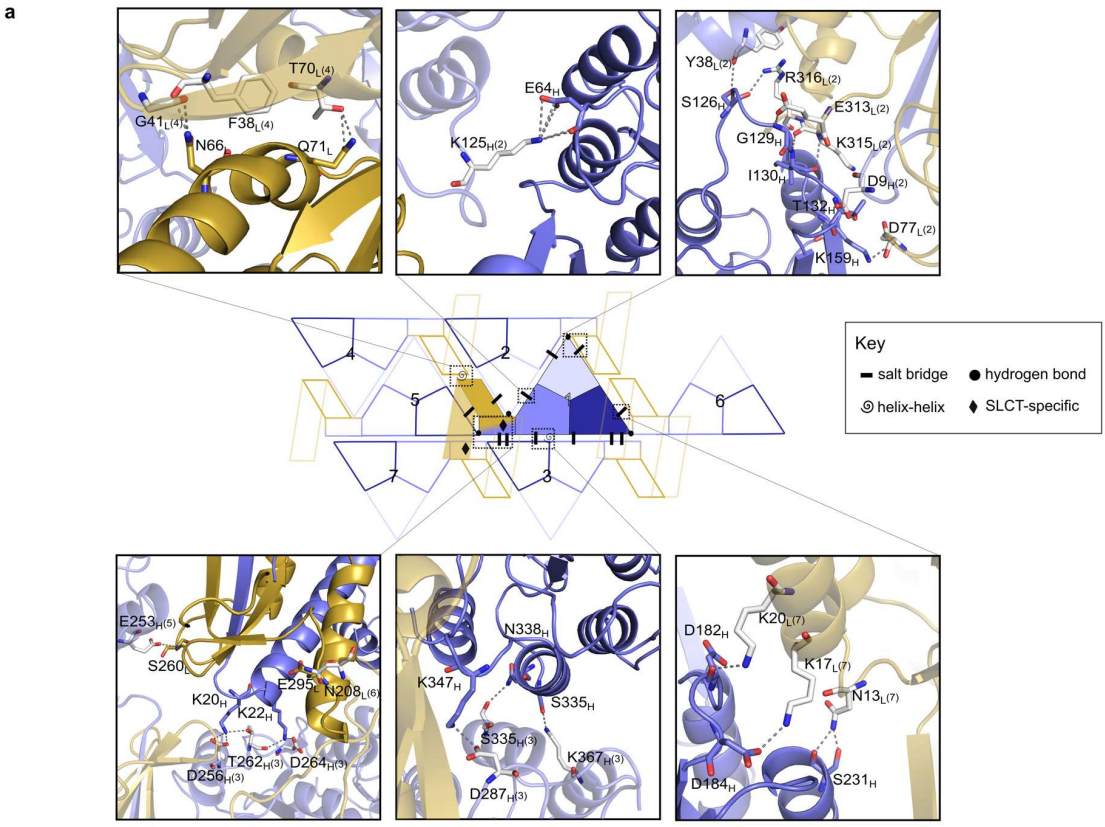
key) on the surface representation of SLP_H array, as defined in Fig. 2. Interacting surfaces between molecules 1-2, defined by pseudo-symmetry related $CWB2_3$-$CWB2_1$, and between molecules 1-3, defined symmetry-related $CWB2_3$ triangular prism faces, are labelled. Cavity between symmetry-related $CWB2_1$-$CWB2_2$ surfaces, represented by green arrows (left) is partially obstructed by HID domains (electrostatic potential surface representation, right) and completely occluded by SLP_L (gold) as shown on the right panel. A long cavity of ~70 Å at the $CWB2_2$ vertices represented by purple arrow (left) is also occluded by HID domains and interacting SLPL molecules (right).

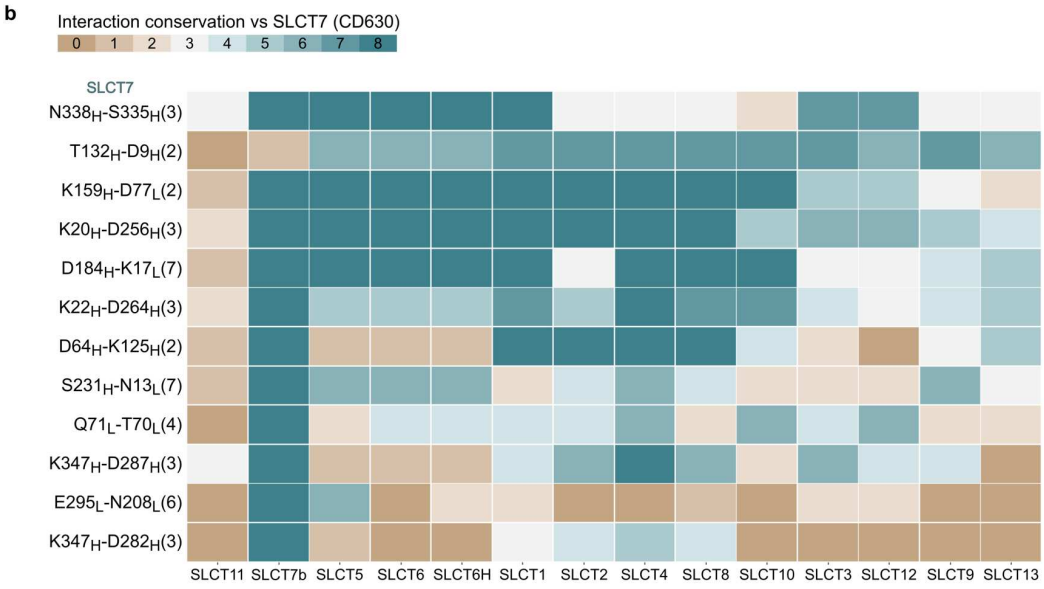
c, Zoomed in view of the pores generated by $SlpA_{CD630}$ multimerization. Pore 1, uncovered top view, as defined in Fig. 1 (top), top view covered by D2 (middle, gold). Pore 2 – top view (bottom). Widest openings are labelled for each pore.

d, Cross section views of pore 1 (top, uncovered by removing domain D2 from structural model; middle, covered by D2 in crystal structure, gold) and pore 2 (bottom). Neighbouring SLP_H (slate blue) and SLP_L (gold) molecules that create the pores are shown in surface representation.

e, Hydrophobicity characteristics of the residues lining pore 1 (top) and 2 (bottom) calculated in ChexVis (see Methods for details) according to Kyte-Doolittle scale, ranging from hydrophilic (green) to hydrophobic (blue), as per hydrophobicity gradient key (middle).

f, Poisson-Boltzmann electrostatic potential calculated for residues lining pore 1 (top) and 2 (bottom) represented as a charge distribution (positive in blue and negative in red, as per electronegativity gradient key). Views are as in d, (left) and as a slice across the largest pore surface (right). Pseudo-symmetry related lysine residues at the top and arginine residues at the bottleneck of pore 2 are highlighted.



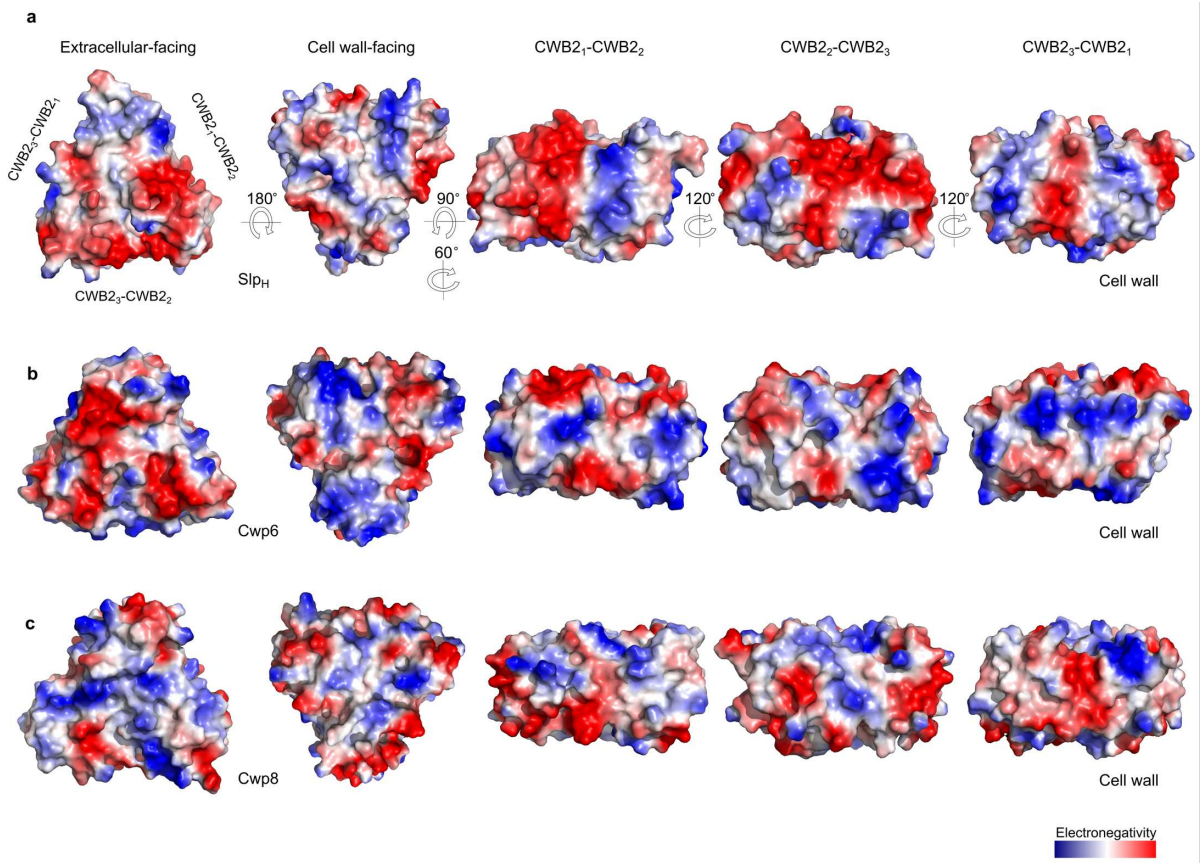


Extended Data Fig. 4: Interactions between neighbouring molecules in S-layer packing

a, Details of the interaction network between 7 $SlpA_{CD630}$ H/L complexes within the 2D array, with each molecule represented as cartoon, interacting residues as sticks (colored as in schematic for molecule 1, white for neighbouring residues, with molecule number identifier

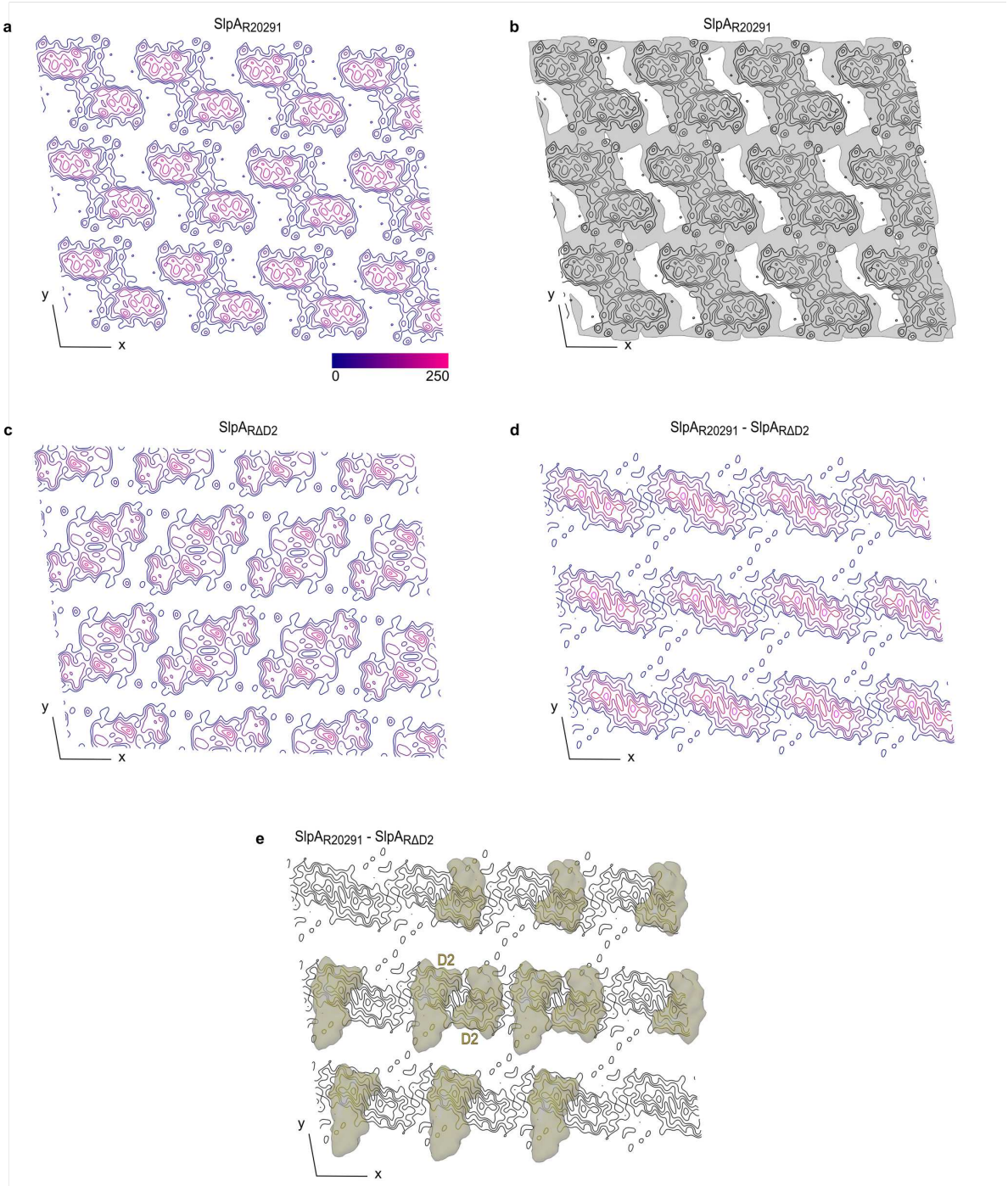
in parenthesis) and interactions as dashed lines. The interface depicted in each panel is marked by a corresponding box within the array representation. Interaction types are shown as bars (salt bridges), circles (hydrogen bonds) and diamonds (between D2 domains) in the central schematic.

b, Clustermap of predicted conservation across known SLCTs for sidechain-sidechain interactions found in $SlpA_{CD630}$ H/L. Representatives of each SLCT were aligned (SI data), SWISS-MODEL structural homology models were generated and superimposed. The residues corresponding to interactions identified in $SlpA_{CD630}$ H/L were analyzed and interaction conservation compared to SLCT-7 is depicted based on residue conservation and prediction of similar or different type of possible interaction. Key: 0 – no residue conservation, unstructured region; 1 – one conserved residue, unstructured region; 2 – no residue conservation, no interaction; 3 – one conserved residue, no predicted interaction; 4 – no residue conservation, different interaction type; 5 – one conserved residue, different interaction type; 6 – no residue conservation, same interaction type; 7 – one conserved residue, same interaction type; 8 – residues and interaction conserved.



Extended Data Fig. 5: Charge distribution across CWB2 motifs in SLP_H and two minor components of the *C. difficile* S-layer

Comparison of the Poisson-Boltzmann electrostatic potential calculated for CWB2 motifs from $SlpA_{CD630}$ (**a**), Cwp6 (**b**) and Cwp8 (**c**). The triangular CWB2 motifs of each CWP were superimposed onto the $SlpA_{CD630}$ CWB2s to determine orientation of Cwp6 and Cwp8. Views are shown from the extracellular and cell wall surfaces, followed by side views of the lateral faces defined by two interacting CWB2s, as per $SlpA_{CD630}$ H/L complex orientation at the cell surface.



Extended Data Fig. 6: Cryo-electron microscopy of wild type and SlpA$_{RΔD2}$ S-layer ghosts.

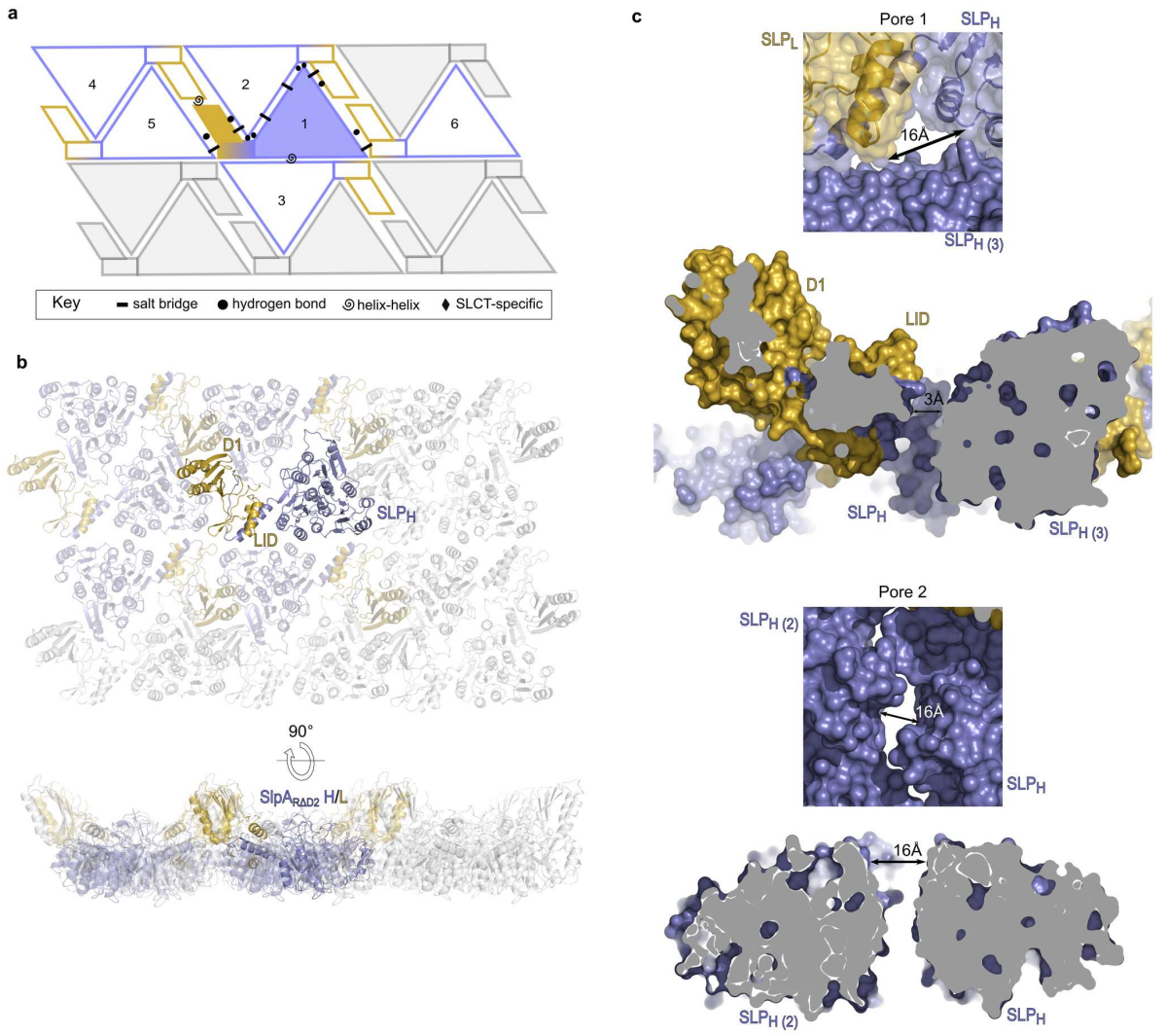
a, Projection map of frozen hydrated native S-layer ghost from strain R20291 at 8.7 Å resolution. Contours represent density greater than mean density, contour interval 0.5 RMS density, as per gradient.

b, Superimposition of the reconstruction of negatively stained native S-layer (gray surface) on the projection in **a.**

c, Projection map at 8.7 Å resolution of frozen hydrated *C. difficile* S-layer ghost containing $SlpA_{R\Delta D2}$, depicted as in **a**.

d, Wild type minus $SlpA_{R\Delta D2}$ difference projection map. Positive difference density is seen to correspond with the projection of the ridge-like density in the 3D reconstruction depicted in **b.**

e, Superposition of isolated domain D2 crystal structure (gold surface) on the difference projection map in **d**. Pseudo-symmetrically related structures are shown together in the central density.

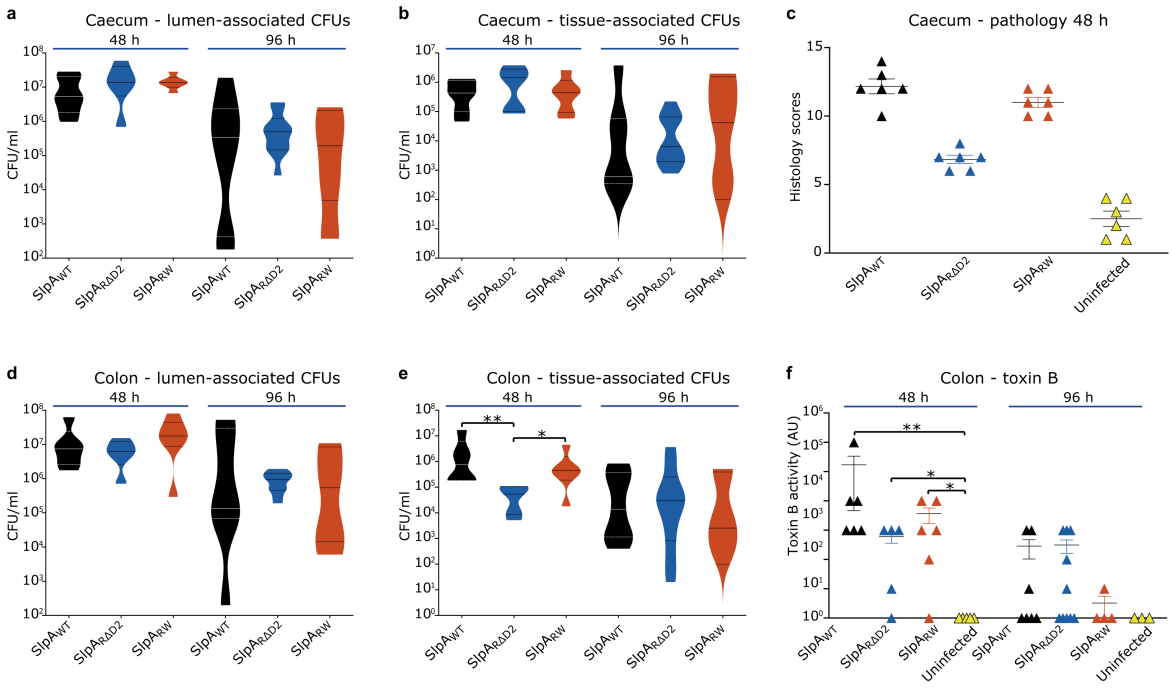


Extended Data Fig. 7: Absence of D2 generates a more porous H/L packing

a, 2D tiling representation of $SlpA_{RΔD2}$ assembly in crystal packing (PDB ID: 7ACZ), with identified interactions represented as symbols defined in the key (as in **Fig. 2**).

b, Cartoon representation of the $SlpA_{RΔD2}$ array (SLP_L colored in gold and SLP_H in slate blue) in top and a side view (as defined in **Fig. 2**).

c, Zoomed in view of the pores generated by $SlpA_{RΔD2}$ multimerization. From top to bottom: pore 1 – top view and cross-section, pore 2 – top view and cross-section. Widest points are marked in each view (black arrows).



Extended Data Fig. 8: *In vivo* phenotypic characterization of RΔD2

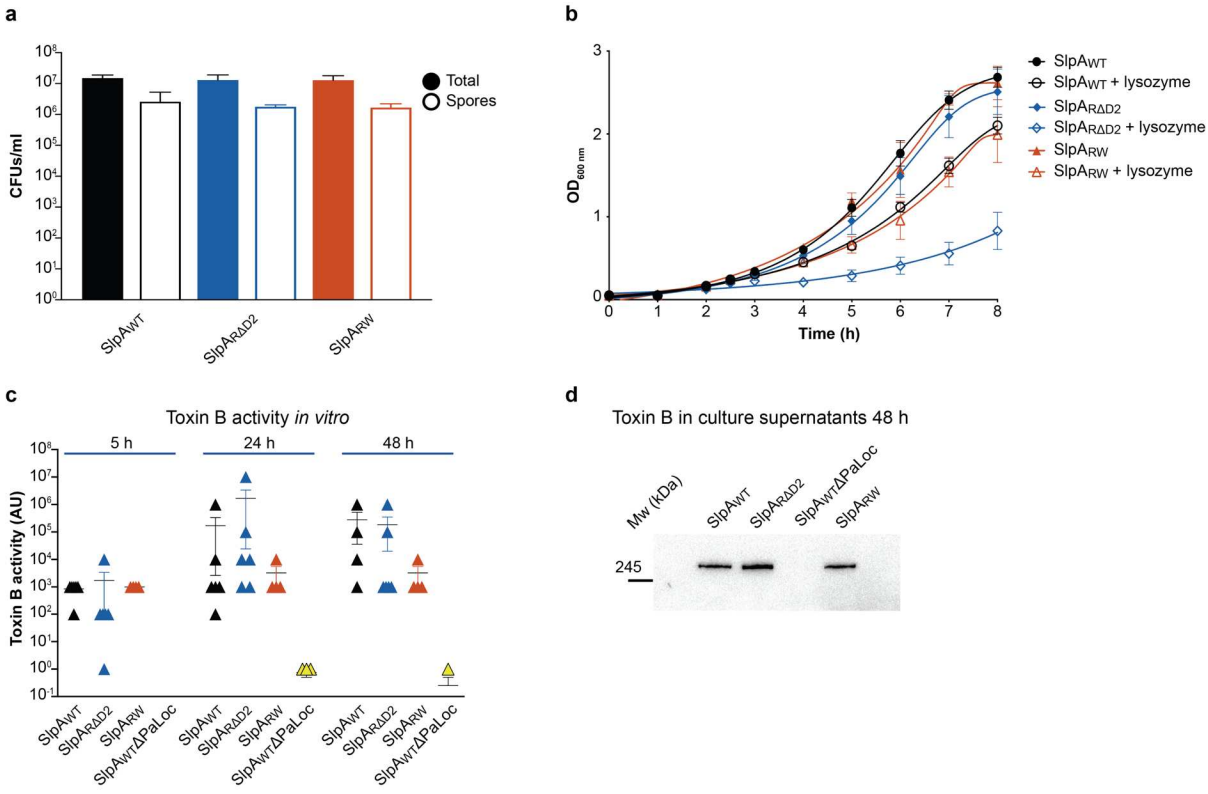
a, and **b,** Total *C. difficile* CFUs recovered from the lumen (**a**) and tissue (**b**) of dissected caeca of animals euthanized 48 and 96 h post-infection.

c, Quantification of histological sections harvested 48 h post-infection. Three independent sections from two animals from each infection group were single blind scored on four independent criteria (epithelial damage, neutrophil margination and tissue infiltration, haemorrhagic congestion and tissue edema, crypt hyperplasia). Scores were assigned for each feature ranging from 1 for no change to 4 indicating substantial change, and cumulative scores calculated.

d, and **e,** Total *C. difficile* CFUs in the lumen (**d**) and tissue (**e**) from dissected colons of animals euthanized 48 and 96 h post-infection.

f, Toxin B activity in filtered luminal contents of the colon. Monolayers of Vero cells were incubated with serial dilutions of filtered luminal contents from individual animals. Toxicity

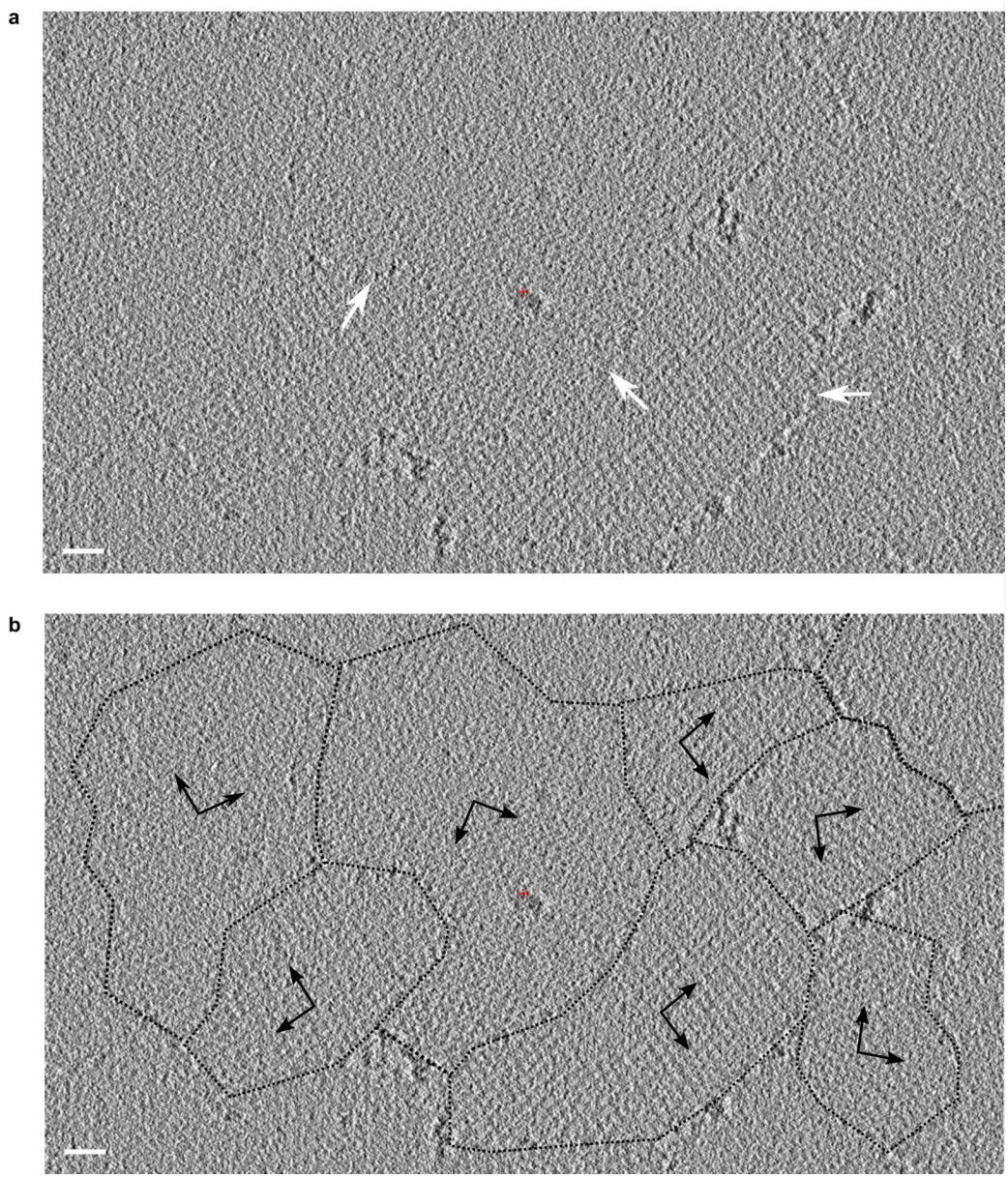
is represented as the maximum fold dilution at which toxicity was observed using samples from animals during the acute phase of infection (48 h) and during recovery (96 h). Data analysis represents a total of 10 mice per strain from 2 independent experiments as violin plots showing medians (solid lines), upper and lower quartile (dashed line) and max/min range (top and base of the plot) for panels **a**-**b**, **d**-**e** (min 5 mice per group), calculated means (± SEM) from histology scoring (2 mice/ 3 sections) in panel **c** or mean toxin activity (±SEM, min 5 mice per group tested in duplicate) for panel **f**. Non-parametric, non-paired Mann-Whitney statistical tests were performed with differences of *p = <0.05, ** p=<0.005 indicated (**e, f**).



Extended Data Fig. 9: *In vitro* phenotypic characterization of RΔD2

a, Sporulation. Stationary phase cultures were incubated for 5 days anaerobically at 37°C. Total CFUs were enumerated on BHI-S agar supplemented with taurocholate (0.1% v/v), while spores were enumerated on the same solid medium following incubation at 65°C for 30 min to kill vegetative cells. **b**, Lysozyme resistance. Cultures were inoculated at an OD_{600nm} of 0.05 and grown anaerobically at 37°C with hourly OD_{600nm} measurements. Where indicated, lysozyme (500 μg ml^{-1}) was added after 2.5 h growth. **c**, Toxin B activity *in vitro*. *C. difficile* strains were grown for the indicated time in TY broth. Monolayers of Vero cells were incubated with serial dilutions of culture supernatants. Toxicity was determined by observation of the integrity of the monolayers through Giemsa staining. **d**, Detection of Toxin B in culture supernatant by western immunoblot. *C. difficile* strains were grown in TY broth for 48 h and secreted toxin B was detected using a monoclonal antibody following SDS-PAGE and electrotransfer to PVDF membrane. Data are presented as mean values (±

SD) from 3 biological replicates, assayed in duplicate **(a-b)** or the mean maximum dilution (± SEM) at which toxicity was observed from six independent experiments **(c)**.



Extended Data Fig. 10: Patches of S-layer present in cryo-electron tomographic slice of extracted S-layer ghosts

a, Distinct S-layer lattice patches can be seen in tomographic slices of S-layer ghosts, with 'fault lines' present where patches intersect (white arrows).

b, Annotated patches (dotted lines) have distinct orientations on the surface, with unit cell axes of the different lattices highlighted (black arrows). Scale bar: 25nm.

Methods

Strains and growth conditions

C. difficile and *E. coli* strains are described in Table 1. *E. coli* strains were routinely grown at 37 °C in LB broth and on LB agar (VWR or Fisher Scientific). *C. difficile* strains were routinely grown under anaerobic conditions at 37 °C on BHI (Oxoid) or BHI-S[1] agar and in TY broth[2]. Growth media were supplemented with chloramphenicol (15 µg ml^{-1}), thiamphenicol (15 µg ml^{-1}) or kanamycin (50 µg ml^{-1}) as required. For detection of *C. difficile* in mouse faeces and tissues, samples were cultured on Brazier's agar supplemented with 5% (v/v) egg yolk (LABM, Neogen), 10% (v/v) defibrinated horse blood (TCS Biosciences) and cycloserine/cefoxitin (LABM, Neogen). For detection of RΔD2, samples were additionally plated on ChromID® *C. difficile* (Biomerieux) chromogenic agar plates.

Construction of RΔD2

DNA oligonucleotides are described in Table 2. Plasmid pRPF233, containing a copy of the complete *slpA* gene from *C. difficile* strain R20291 was modified by inverse PCR using oligonucleotides RF102 and RF103 to delete the coding sequence of SlpA residues 115-259 and replace with GGAGGT, encoding two glycine residues. The resulting plasmid, pOB001, was transferred to the *C. difficile* S-layer mutant strain FM2.5[3] by conjugation[4]. FM2.5 displays an aberrant colony morphology that is easily distinguished from wild type *C. difficile*. Recombination between the plasmid-borne *slpA* gene and the mutated copy on the chromosome was detected by reversion to normal colony morphology. Plasmid curing was confirmed by loss of thiamphenicol resistance, the chromosomal location of the engineered *slpA* gene was confirmed by PCR and the resulting protein profile was determined by SDS-PAGE of S-layer proteins isolated using standard methods (see below).

Plasmid construction

For crystallization studies, fragments of R7404 *slpA*, encoding mature HID (residues 1-41) and SLP_L (residues 1-316) or LID (residues 240-316) were amplified from genomic DNA and cloned into pACYC-Duet1 yielding plasmids pJAK149 and pJAK147, respectively. C-terminally 6x His-tagged HID was amplified using RF1396 and RF1397 and cloned into pACYC-Duet (MCS1) linearized using RF1398 and RF1400 by Gibson assembly, and SLP_L or LID were amplified using RF1394 and RF1395 or RF1395 and RF1396, respectively, and cloned into MCS2 using *Nde*I-*Kpn*I restriction cloning.

To study protein-protein interactions *in vitro*, DNA encoding mature SLP_L or SLP_H of CD630 and R7404 was amplified using Q5 (NEB) PCR and cloned into pET28a using *Nco*I-*Xho*I restriction cloning, in frame with a C-terminal 6x His-tag. Deletion variants (see Table 1 for construct details) lacking HID or LID, or point mutants within HID and LID were constructed by inverse PCR, using primers listed in Table 2, as previously described[5].

To study the impact of individual LID and HID point mutations on H/L complex assembly in *C. difficile*, codons for SLP_L F274 or SLP_H Y27 in pRPF170 were mutated to GCA (Ala) by inverse PCR cloning, yielding plasmids pRPF209 and pJAK186, respectively.

Protein expression and purification

S-layer was extracted as previously described[6]. Briefly, 400 ml of *C. difficile* CD630, R7404, RΔD2 16-hour culture were harvested by centrifugation at 4,696 × *g* for 30 min at room temperature. Cells were washed with 20 ml of phosphate buffered saline (PBS) pH 7.4, centrifuged for 10 min at 4,696 × *g*, and resuspended in 5 ml of 0.2 M glycine-HCl pH 2.2. Cell suspension was centrifuged at 21,100 × *g* for 10 min, and recovered supernatant was neutralized with 2 M Tris-base. S-layer extract was then filtered and resolved onto a

Superdex 200 26/600 column using an ÄKTA Pure FPLC system (GE Healthcare) in 10 mM Tris-HCl pH 7.5, 100 mM NaCl, 5 mM EGTA buffer. Protein eluate was analyzed by 12% SDS-PAGE.

BL21 (DE3) cells transformed with plasmids pJAK149 or pJAK147 were used to co-express HID-6x His-tag and SLP_L or HID-6x His-tag and LID, in Auto Induction Media TB (Formedium) supplemented with 30 µg ml^{-1} chloramphenicol. Cells from 1 L of culture grown for 18h at 37 °C were harvested by centrifugation at 4,000 × *g*, 4 °C for 30 min. Pellets were washed with PBS and frozen at -20 °C.

Variants of SlpA subunits (Table 1) for interaction studies were expressed in Rosetta (DE3) cells, in 50 ml of Merck Novagen Overnight Express Instant TB Medium, supplemented with 50 µg ml^{-1} kanamycin and 15 µg ml^{-1} chloramphenicol for 18 hours at 37 °C. Harvested cells were washed with PBS and frozen at -20 °C.

Cell lysis for crystallography of recombinant proteins and interaction studies was performed using BugBuster Protein Extraction Reagent (Novagen). Pellets were resuspended to homogeneity in lysis buffer (50 mM Tris-HCl pH 8.0, 250 mM NaCl, 1x cOmplete EDTA-free protease inhibitors (Roche), 100 µg ml^{-1} lysozyme, 10 µg ml^{-1} DNase I, 1x BugBuster) and incubated at room temperature for 30 min. Extracts were centrifuged at 20,000 × *g* for 30 min, supernatant was filtered and separated on a 5 ml HisTrap (GE Healthcare) column in 50 mM Tris-HCl pH 8.0, 250 mM NaCl, 10 mM imidazole, with a linear gradient of imidazole (10-500 mM). Eluate fractions were further purified using size exclusion chromatography on Superdex 75 (GE Healthcare) in 50 mM Tris-HCl pH 7.5, 150 mM NaCl.

Variants of SLP_H were recovered by affinity chromatography from isolated purified inclusion bodies, solubilized in 50 mM Tris-HCl pH 7.5, 500 mM NaCl and 8 M urea for 20 min at room temperature. Solubilized, cleared supernatant was loaded onto 5 ml HisTrap (GE Healthcare)

column in solubilization buffer, and column-bound protein was refolded in 50 ml of 50 mM Tris-HCl pH 7.5, 500 mM NaCl, 0.1% Triton X-100, followed by 50 ml of 50 mM Tris-HCl pH 7.5, 250 mM NaCl, 5 mM β-cyclodextrin. Affinity purification was performed in 50 mM Tris pH 8.0, 250 mM NaCl, 10 mM imidazole, with protein elution by a linear gradient of imidazole (10-500 mM). Each chromatographic step was followed by analysis of the eluate by 12% SDS-PAGE.

Protein analysis by western immunoblotting

For analysis of H/L complex interactions on the surface of *C. difficile*, plasmids carrying a tetracycline-inducible copy of CD630 *slpA* (pRPF170) or derivatives with a point mutation in the LID (F274A, pRPF209) or the HID (Y27A, pJAK186) were transferred into the *slpA*-null strain FM2.5 by conjugation[4]. Strains were grown to an OD_{600nm} of ~0.4 in TY broth and induced with anhydrotetracycline (20 ng ml^{-1}). Surface-localized H/L subunits were extracted using low pH glycine as described above and normalized to an equivalent OD_{600nm} of 25. Culture supernatants were filtered, concentrated to an equivalent OD_{600nm} of 50 using a Vivaspin column with a 10 kDa MWCO. Samples were then subjected to SDS-PAGE and western immunoblotting using polyclonal antibodies specific for the CD630 SLP_H or SLP_L[5].

Analysis of protein-protein interactions by enzyme-linked immunosorbent assay (ELISA)

The assays were performed as previously described[5]. Briefly, Maxisorp microtiter plates (Nunc) were coated with 10 µg ml^{-1} of SLP_L or SLP_H and their variants (Table 1), blocked with 3% (w/v) milk in PBS with 0.05% (v/v) Tween-20, and overlaid with respective interacting partner, SLP_L or SLP_H, across 0.0001-100 µg ml^{-1} range. Binding was assessed with polyclonal rabbit primary antibodies against the overlay protein. Detection was carried out

spectrophotometrically by monitoring formation of product of horse radish peroxidase (HRP-conjugated secondary anti-rabbit antibody) with *o*-phenylenediamine dihydrochloride (OPD) upon addition of hydrogen peroxide to the reaction mix, with maximum of absorbance at 490 nm using Biotek ELx800 plate reader.

X-ray crystallography

Purified and concentrated proteins (recombinant LID/HID-6x His-tag at 38 mg ml^{-1} and SLP_L/HID-6x His-tag at 20.9 mg ml^{-1}, CD630, R7404 and RΔD2 H/L at 10 mg ml^{-1}) were subjected to crystallization using a Mosquito liquid handling robot (TTP Labtech), with the sitting drop vapor-diffusion method at 20 °C. Native H/L complex crystallized in 0.1 MES pH 6.0, 1.25 M lithium chloride, 16 PEG 6,000 and 10 % glycerol (v/v). Recombinant SLP_L/HID-6x His-tag produced diffraction quality crystals in 0.2 M ammonium sulphate, 0.1 M MES pH 6.5, 35% (v/v) MPD, while LID/HID-6x His-tag was crystallized in 1.6 M sodium citrate tribasic dihydrate pH 6.5. Data were collected on the I04 (λ= 0.98 Å), I23 (λ = 2.75 Å) and I24 (λ = 0.93 Å) beamlines at the Diamond Light Source Synchrotron (Didcot, UK) at 100 K. The data were acquired from the automatic software pipeline xia2 within the Information System for Protein Crystallography Beamline (ISPyB), processed with XDS[7], iMosflm[8] or DIALS[9] and scaled with Aimless[10] within CCP4i[11] or CCP4i2[12] software suites. When needed, density modification was performed with PARROT[13].

The structure of LID/HID-6x His was solved *de novo* using Arcimboldo_lite[14] within CCP4i, starting from several 10-14 residues-long polyalanine models of α-helices. Automatic model building was performed with Buccaneer[15], followed by manual building with Coot[16] and refinement with Phenix_refine[17].

The structure of SLP_L/HID-6x His-tag was determined by sequential molecular replacement in Phaser[18] searching first for SLP_L D1-D2 domains model (PDB ID: 3CVZ[5]), followed by the search with LID/HID structure into a fixed SLP_L solution, and subsequent manual building (COOT) and refinement (Phenix_refine).

Initial attempts to solve the substructure of the complete H/L complex by S-SAD provided only weak phases, and were improved by combining molecular replacement in Phaser using the CWB2 domain core of *C. difficile* Cwp8 (PDB ID: 5J6Q[19]) with sulphur anomalous difference Fourier maps using Anode[20]. This solution was used for MR-SSAD in Phenix.autosol and cycles of manual building in COOT and density improvement were used to improve the electron density maps and the model of the core SLP_H. A complete H/L model was obtained by successive molecular replacement runs using Phaser in CCP4i2 combining the SLP_H model with the obtained LID/HID (PDB ID: 7ACW), D2 from SLP_L/HID structure (PDB ID: 7ACV) and D1 (loops removed) from CD630 SLP_L (PDB ID: 3CVZ). Presence of non-water solvent molecules was investigated with CheckMyBlob[21]. Final models were obtained after iterative cycles of manual model building with Coot and refinement in Phenix_refine and REFMAC5[22], with a final optimization step using PDB-REDO[23], where relevant. Applied strategies included refinement of XYZ coordinates, real space, individual B-factors, TLS parameters and occupancies. Structure of LID/HID consisted of 100% Ramachandran favored rotamers, SLP_L/HID consisted of 98% Ramachandran favored and 2% Ramachandran allowed rotamers. Native protein structures of CD630, R7404 and RΔD2 H/L were modelled to 98%, 96%, 96% of Ramachandran favored and 0.2%, 0.2%, 0.1% of Ramachandran outlier rotamers, respectively. Full data collection and refinement statistics are summarized in SI Table 1. Validation of final models was performed using COOT and Phenix internal tools, as well as MOLPROBITY [24] web server. All structural models were

144 validated using wwPDB validation server prior to deposition of files. Data collection and
145 refinement details are summarized in SI Table 1. Structural representations were generated
146 using PyMOL Molecular Graphics System (Schrödinger, LLC) or Chimera[25].
147
148 **Electron crystallography data collection**
149 To allow visualization by electron microscopy, S-layers were either removed from *C. difficile*
150 cells in a single piece following peptidoglycan digestion (S-layer ghosts) or cells were
151 mechanically fragmented (S-layer/cell wall fragments). *C. difficile* cells were harvested by
152 centrifugation and resuspended to an OD_{600} of 10 in 20 mM HEPES pH 7.5, 150 mM NaCl,
153 500 mM sucrose. For S-layer ghosts, cell walls were digested using purified φCD27 endolysin
154 for 30 min at 37 °C[26]. The resulting membrane-bound spheroplasts were removed from the
155 sample by centrifugation at 2,000 × *g* for 2 min and the supernatant, containing S-layer
156 ghosts, was retained for imaging. 5 µl of S-layer ghosts were loaded on glow-discharged,
157 amorphous carbon-coated 300 mesh copper EM grids and stained with 2% uranyl acetate,
158 as previously described[27]. Samples were examined on a Phillips CM200 FEG transmission
159 electron microscope at 200 kV. Images were collected on a 4096 x 4096 pixel Gatan
160 UltraScan 4000SP Model 890 CCD camera (Gatan Inc.), with 15 µm pixel size. A total of 36
161 micrographs of R20291 and 29 micrographs for RΔD2 S-layer extracts were collected at a
162 magnification of 82351 x and defocus range from -800 to -2200 nm. The specimen tilt angle
163 ranged from −55° to +55° in increments of 10°.
164
165 For cryo-EM, fragments of S-layers were generated by mechanical disruption. Briefly, 60 ml
166 of *C. difficile* cells at OD_{600nm} 0.6 - 0.8 were centrifuged at 2,000 × *g* for 15 min at 4 °C. The
167 cell pellet was washed twice in cold deionized water and combined with an equal volume of

pre-cooled acid-washed glass beads (Sigma) and homogenized in a Braun MSK homogenizer for 30 s. The homogenate was cooled and centrifuged at 800 × *g* for 10 min to remove glass beads and unbroken cells. S-layer fragments were then harvested at 3000 × *g* for 10 min, washed with cold 1 M NaCl and resuspended in cold 2% (v/v) Triton X-100. 2.5 µl of the S-layer fragments were added to glow-discharged Quantifoil® 2/2 grids. The grids were then blotted for 30 s and plunged into liquid ethane, using a Vitrobot Mark III (FEI). The frozen grids were stored in liquid nitrogen for later observation. Micrographs, at 68,000 x magnification and defocus range of -2000 to -3000 nm, were obtained on a Falcon II direct electron detector (FEI) using a Tecnai F20 microscope (FEI) operating at 200 keV.

Electron crystallography data processing

Images were initially processed using the *2dx* suite[28-30]. Most micrographs of S-layer ghosts showed two rotationally separated lattices in Fourier transforms and these were indexed independently. Images were masked based on crystal size and good crystalline order and subjected to two cycles of unbending using the programs *QUADSEARCH* and *CCUNBEND*. The symmetry was determined from images of untilted crystals using *ALLSPACE*[31]. Phase origins for individual images were refined against each other using *ORIGTILTK*, sequentially adding images of higher tilt to the refinement. Crystal tilt angles were estimated from lattice distortion. *LATLINE*[32] was used to determine interpolated amplitudes and phases on a regular lattice of 1/160 $Å^{-1}$ in the *z** direction. A Gaussian tapered real-space envelope of width slightly larger than that of the H/L complex estimated from the X-ray crystal structure (70 Å for wild type and 60 Å for RΔD2) was applied. The phase origin and tilt parameters were further refined using the output interpolated lattice lines as reference. The variation of amplitude and phase along *0,0,l* was estimated by examining a plot of maximum contrast on

each Z-section in real space[33]. The final structure factors were sampled from the interpolated lattice lines[32] and a 3D map generated within the CCP4 suite of programs[11,28]. Cryo-EM micrographs of untilted R20291 and RΔD2 samples were processed similarly to generate 2D projection maps. B-factors were calculated using SCALIMAMP3D with bacteriorhodopsin diffraction amplitudes as reference[34]. Data collection, processing and analysis details are summarized in SI Tables 2 and 3.

Fitting X-ray structures to EM density

The coordinates of R20291 and RΔD2 H/L complex X-ray structural models were fitted using Chimera 1.13.1[25] into the wild type and mutant electron microscopy reconstructions. The extended lattice of each H/L complex was generated by calculating symmetry-related molecules from the crystal packing in PyMOL, and then manually orienting them in the EM density, based on the known surface orientation *i.e*. SLP_L facing the environment, and SLP_H facing the cell wall. The 'Fit in map' function was then used to calculate the highest correlation to a map simulated from the X-ray structure coordinates at 20 Å resolution.

Tomography

For cryo-electron tomography (cryo-ET), the homogenized S-layer ghost sample used in electron crystallography was mixed with an equal volume of 10 nm BSA-treated nanogold beads, and 3 µl of this mixture was applied to a glow discharged lacey carbon with ultra-thin carbon 300 mesh grid, blotted for 3 s and plunged into liquid ethane, using a Leica EM GP. The frozen grids were stored in liquid nitrogen temperature for later observation. Tilt series were collected on a Titan Krios microscope operating at 300 keV with a GIF Quantum energy filter, Volta phase plate, and K2 camera operating in super-resolution mode. Micrographs

were collected using SerialEM, at a pixel size of 5.47 Å, with each tilt series covering ± 60° with a tilt increment of 3°, and collected with a grouped dose-symmetric acquisition scheme and group sizes of 4. Samples from each tilt series received 100 e/Å^2 total dose with 20 frames per tilt. Tomograms were constructed using the IMOD package[35]. Tilt series were tracked and aligned based on fiducial markers, and then tomograms were reconstructed by weighted back projection with 1x binning.

Analysis of sporulation and resistance to lysozyme

Quantitative analysis of sporulation efficiency was performed as described previously[36]. Overnight stationary phase cultures were first diluted to an OD_{600nm} of 0.01, grown for 8 hours before a second dilution to an OD_{600nm} of 0.0001. Following overnight growth, the resulting stationary phase culture (T=0) was then incubated. After 5 days, the proportion of vegetative cells and spores was determined; total and heat-resistant (65 °C for 30 min) colony forming units (CFUs) were enumerated in BHI-S agar supplemented with 0.1% (w/v) taurocholate. Assays were repeated in triplicate with biological triplicates.

To assess resistance to lysozyme, overnight *C. difficile* cultures were grown in TY broth, subcultured to an OD_{600nm} of 0.05 in 1 ml fresh TY in a 1.5 ml cuvette and then grown for 8 h with hourly OD_{600nm} measurements. Where appropriate, lysozyme (500 µg ml^{-1}) was added after 2.5 h growth. Experiments were performed in triplicate on biological duplicates and data expressed as the mean and standard deviation.

Animal experiments

All procedures were performed in strict accordance with the Animals (Scientific Procedures) Act 1986 with specific approval granted by the Home Office, UK (PPL60/8797). C57/Bl6

specific pathogen free female mice aged 6 - 8 weeks were supplied by Charles River (Edinburgh). Animals were housed within individual sterilized ventilated cages (IVCs) in groups of five. Sterilized food and water were provided *ad libitum* and animals kept at a constant room temperature of 20 - 22 °C with a 12 hour light/dark cycle. To limit the impact of cage effects, experiments (n = 5 mice) were performed at least in duplicate using animals delivered on different dates. Animals were prepared for *C. difficile* challenge by provision of an antibiotic cocktail (0.035 mg ml^{-1} gentamycin, 0.4 mg ml^{-1} kanamycin, 850 U ml^{-1} colistin, 0.215 mg ml^{-1} metronidazole, 0.045 mg ml^{-1} vancomycin) administered *ad libitum* in the drinking water for 3 days[37]. Clindamycin (150 mg kg^{-1}) was then given by oral gavage the following day. Mice were rested for two days without antibiotic administration before challenge of each animal with 10^6 spores of *C. difficile* delivered by oral gavage. Daily relative weight loss was determined by dividing individual daily weights by the weight of each mouse prior to challenge. Animals with a weight loss greater than 15% were culled. *C. difficile* colonization was quantified through the serial dilution and culturing of fresh faecal material collected from individual animals or from caecal and colonic samples prepared at the time of cull. These were generated by opening the tissue longitudinally and recovering the contents by washing the tissue in 5 ml or 2 ml of PBS. respectively. Colony forming units were calculated from counts generated from diluted samples from individual animals and results reflect the median of a minimum of n = 5 mice at each time point.

Tissue histology

Histological samples were harvested from the cecum at 48 h (peak of infection) and immediately fixed in 10% formalin. Embedded tissue sections were cut and stained with hematoxylin and eosin. The histological severity was graded using an adapted scoring

system accounting for: a. epithelial damage, b. neutrophil margination and tissue infiltration, c. hemorrhagic congestion and tissue edema, and d. crypt hyperplasia. At least six independent fields of view (a maximum of three taken from any individual animal) were examined. Scores for each parameter were assigned for each feature (1, mild; 2, moderate; 3, significant; 4, severe). Cumulative scores reflect the sum of each scored feature.

Toxin activity

Toxin activity in caecal and colonic contents collected at post-mortem was determined by measurement of cytotoxic activity on cultured cells. Activity of the toxins produced *in vitro* was also analyzed from spent culture supernatant taken from TY culture at 5, 24 and 48 h. Samples from both *in vivo* and *in vitro* experiments were filtered (0.2 μm), serially diluted and applied to monolayers of Vero cells (Toxin B), as described previously[38]. Each sample was analyzed minimally in two independent experiments, with samples collected from a minimum of n = 5 mice per time point.

For direct detection of toxin *in vitro*, 48 h culture supernatants were filtered using a 0.44 μm filter and concentrated to an equivalent OD_{600nm} of 50 using a Vivaspin column with a 10 kDa MWCO. Concentrated samples were separated by SDS-PAGE, electroblotted to PVDF membrane and probed using a specific mouse monoclonal antibody (Toxin B: MA1-7413, Thermo Fisher).

Other methods

PDBeFold[39] was used to search for homologous structures and to compare the similarity between models determined in this study. PISA[40], PDBSum[41] and LigPlot+[42] were used to investigate interdomain and protein-protein interfaces. Structural flexibility of models was

assessed by DynDom6D (v1.0 with default settings[43]), HingeProt webserver[44] and the components of DynOmics webserver.

Evolutionary conservation of SlpA across the unique annotated SLCTs retrieved from the *Clostridioides difficile* Multi Locus Sequence Typing database (https://pubmlst.org/bigsdb?db=pubmlst_cdifficile_seqdef&page=alleleQuery&locus=slpA&submit=1) was performed in the ConSurf version 3.0 server[45], using the structural model of the CD630 H/L complex (PDB ID: 7ACY) to map the conservation scores onto 3D crystal structure. Conservation scores were calculated with the Bayesian method, using the WAG model of amino acid substitution (selected based on ProtTest 3.4.1).

Representatives of each SLCT, as defined previously[3], were used to analyze conservation of structural features and H/L interactions. Sequences were aligned using Clustal Omega, with secondary structure assignment from the CD630 H/L structural model as defined in PDBSUM. Structural models for SLCT representatives were generated with SWISS-MODEL webserver using the H/L complex crystal model as a user provided template. Structural alignments between homology model and template were performed using SSM algorithm in Coot. Data was mined using Numpy (v1.16.6)[46] Pandas (v0.24.2)[47] and the heatmap generated with Seaborn (v0.10.1)[48].

For SI SLCT sequence alignment, protein sequences for SlpA representatives used to generate homology models were aligned in Clustal Omega[49] with default settings. Secondary structure information from CD630 H/L complex was added to alignment by ESPript3[50]. Manual annotations were added to the figure for clarification on subunits, domains and position of interacting residues.

Analysis of pores in the H/L array was carried out using ChexVis[51] and hydrophobicity patterns for residues lining each pore calculated using the Kyte-Doolittle scale[52].

312

313 **Data availability**

314 Crystal structures were deposited in Protein Data Bank with PDB IDs 7ACW (LID/HID), 7ACV

315 (SLP_L/HID), 7ACX (H/L, R7404), 7ACY (H/L, CD630) and 7ACZ (H/L, RΔD2).

316

317 **Statistical analysis**

318 Statistical analysis was carried out in GraphPad Prism 8. Non-parametric, non-paired Mann-

319 Whitney test results were considered.

320 References

321

322 1 Sorg, J. A. & Dineen, S. S. Laboratory maintenance of *Clostridium difficile*. *Curr Protoc*
323 *Microbiol* **Chapter 9**, Unit9A 1, doi:10.1002/9780471729259.mc09a01s12 (2009).

324 2 Dupuy, B. & Sonenshein, A. L. Regulated transcription of *Clostridium difficile* toxin
325 genes. *Mol Microbiol* **27**, 107-120, doi:10.1046/j.1365-2958.1998.00663.x (1998).

326 3 Kirk, J. A. *et al.* New class of precision antimicrobials redefines role of *Clostridium*
327 *difficile* S-layer in virulence and viability. *Sci Transl Med* **9**,
328 doi:10.1126/scitranslmed.aah6813 (2017).

329 4 Kirk, J. A. & Fagan, R. P. Heat shock increases conjugation efficiency in *Clostridium*
330 *difficile*. *Anaerobe* **42**, 1-5, doi:10.1016/j.anaerobe.2016.06.009 (2016).

331 5 Fagan, R. P. *et al.* Structural insights into the molecular organization of the S-layer
332 from *Clostridium difficile*. *Mol Microbiol* **71**, 1308-1322, doi:10.1111/j.1365-
333 2958.2009.06603.x (2009).

334 6 Calabi, E., Calabi, F., Phillips, A. D. & Fairweather, N. F. Binding of *Clostridium difficile*
335 surface layer proteins to gastrointestinal tissues. *Infect Immun* **70**, 5770-5778,
336 doi:10.1128/iai.70.10.5770-5778.2002 (2002).

337 7 Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta*
338 *Crystallogr D Biol Crystallogr* **66**, 133-144, doi:10.1107/S0907444909047374 (2010).

339 8 Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. iMOSFLM: a
340 new graphical interface for diffraction-image processing with MOSFLM. *Acta*
341 *Crystallogr D Biol Crystallogr* **67**, 271-281, doi:10.1107/S0907444910048675 (2011).

342 9 Winter, G. *et al.* DIALS: implementation and evaluation of a new integration package.
343 *Acta Crystallogr D Struct Biol* **74**, 85-97, doi:10.1107/S2059798317017235 (2018).

10 Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204-1214, doi:10.1107/S0907444913000061 (2013).

11 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242, doi:10.1107/S0907444910045749 (2011).

12 Potterton, L. *et al.* CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta Crystallogr D Struct Biol* **74**, 68-84, doi:10.1107/S2059798317016035 (2018).

13 Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr D Biol Crystallogr* **66**, 470-478, doi:10.1107/S090744490903947X (2010).

14 Rodriguez, D. D. *et al.* Crystallographic *ab initio* protein structure solution below atomic resolution. *Nat Methods* **6**, 651-653, doi:10.1038/nmeth.1365 (2009).

15 Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **62**, 1002-1011, doi:10.1107/S0907444906022116 (2006).

16 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).

17 Headd, J. J. *et al.* Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution. *Acta Crystallogr D Biol Crystallogr* **68**, 381-390, doi:10.1107/S0907444911047834 (2012).

18 McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).

19 Usenik, A. *et al.* The CWB2 cell wall-anchoring module Is revealed by the crystal structures of the *Clostridium difficile* cell wall proteins Cwp8 and Cwp6. *Structure* **25**, 514-521, doi:10.1016/j.str.2016.12.018 (2017).

20 Thorn, A. & Sheldrick, G. M. ANODE: anomalous and heavy-atom density calculation. *J Appl Crystallogr* **44**, 1285-1287, doi:10.1107/S0021889811041768 (2011).

21 Kowiel, M. *et al.* Automatic recognition of ligands in electron density by machine learning. *Bioinformatics* **35**, 452-461, doi:10.1093/bioinformatics/bty626 (2019).

22 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355-367, doi:10.1107/S0907444911001314 (2011).

23 Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **1**, 213-220, doi:10.1107/S2052252514009324 (2014).

24 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21, doi:10.1107/S0907444909042073 (2010).

25 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).

26 Peltier, J. *et al. Clostridium difficile* has an original peptidoglycan structure with a high level of N-acetylglucosamine deacetylation and mainly 3-3 cross-links. *J Biol Chem* **286**, 29053-29062, doi:10.1074/jbc.M111.259150 (2011).

27 Ball, D. A. *et al.* Structure of the exosporium and sublayers of spores of the *Bacillus cereus* family revealed by electron crystallography. *Mol Microbiol* **68**, 947-958, doi:10.1111/j.1365-2958.2008.06206.x (2008).

28 Crowther, R. A., Henderson, R. & Smith, J. M. MRC image processing programs. *J Struct Biol* **116**, 9-16, doi:10.1006/jsbi.1996.0003 (1996).

29 Gipson, B., Zeng, X. & Stahlberg, H. 2dx_merge: data management and merging for 2D crystal images. *J Struct Biol* **160**, 375-384, doi:10.1016/j.jsb.2007.09.011 (2007).

30 Henderson, R., Baldwin, J. M., Downing, K. H., Lepault, J. & Zemlin, F. Structure of purple membrane from *Halobacterium halobium*: recording, measurement and evaluation of electron micrographs at 3.5 Å resolution. *Ultramicroscopy* **19**, 147-178, doi:10.1016/0304-3991(86)90203-2 (1986).

31 Valpuesta, J. M. a., Carrascosa, J. L. & Henderson, R. Analysis of electron microscope images and electron diffraction patterns of thin crystals of Ø29 connectors in ice. *Journal of Molecular Biology* **240**, 281-287, doi:10.1006/jmbi.1994.1445 (1994).

32 Agard, D. A. & Moody, M. F. A least-squares method for determining structure factors in three-dimensional tilted-view reconstructions. *Journal of Molecular Biology* **167**, 849-852, doi:10.1016/S0022-2836(83)80114-4 (1983).

33 Amos, L. A., Henderson, R. & Unwin, P. N. Three-dimensional structure determination by electron microscopy of two-dimensional crystals. *Prog Biophys Mol Biol* **39**, 183-231, doi:10.1016/0079-6107(83)90017-2 (1982).

34 Havelka, W. A., Henderson, R. & Oesterhelt, D. Three-dimensional structure of halorhodopsin at 7 A resolution. *J Mol Biol* **247**, 726-738, doi:10.1006/jmbi.1995.0176 (1995).

35 Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J Struct Biol* **116**, 71-76, doi:10.1006/jsbi.1996.0013 (1996).

414 36 Dembek, M. *et al.* High-throughput analysis of gene essentiality and sporulation in
415 *Clostridium difficile*. *MBio* **6**, e02383, doi:10.1128/mBio.02383-14 (2015).
416 37 Theriot, C. M. & Young, V. B. Interactions between the gastrointestinal microbiome
417 and *Clostridium difficile*. *Annu Rev Microbiol* **69**, 445-461, doi:10.1146/annurev-
418 micro-091014-104115 (2015).
419 38 Buckley, A. M., Spencer, J., Candlish, D., Irvine, J. J. & Douce, G. R. Infection of
420 hamsters with the UK *Clostridium difficile* ribotype 027 outbreak strain R20291. *J*
421 *Med Microbiol* **60**, 1174-1180, doi:10.1099/jmm.0.028514-0 (2011).
422 39 Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast
423 protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr*
424 **60**, 2256-2268, doi:10.1107/S0907444904026460 (2004).
425 40 Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline
426 state. *J Mol Biol* **372**, 774-797, doi:10.1016/j.jmb.2007.05.022 (2007).
427 41 Laskowski, R. A., Jablonska, J., Pravda, L., Varekova, R. S. & Thornton, J. M. PDBsum:
428 Structural summaries of PDB entries. *Protein Sci* **27**, 129-134, doi:10.1002/pro.3289
429 (2018).
430 42 Laskowski, R. A. & Swindells, M. B. LigPlot+: multiple ligand-protein interaction
431 diagrams for drug discovery. *J Chem Inf Model* **51**, 2778-2786,
432 doi:10.1021/ci200227u (2011).
433 43 Veevers, R. & Hayward, S. Methodological improvements for the analysis of domain
434 movements in large biomolecular complexes. *Biophys Physicobiol* **16**, 328-336,
435 doi:10.2142/biophysico.16.0_328 (2019).

44 Emekli, U., Schneidman-Duhovny, D., Wolfson, H. J., Nussinov, R. & Haliloglu, T. HingeProt: automated prediction of hinges in protein structures. *Proteins* **70**, 1219-1227, doi:10.1002/prot.21613 (2008).

45 Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* **44**, W344-350, doi:10.1093/nar/gkw408 (2016).

46 Walt, S. v. d., Colbert, S. C. & Varoquaux, G. The NumPy array: A structure for efficient numerical computation. *Computing in Science & Engineering* **13**, 22-30, doi:10.1109/MCSE.2011.37 (2011).

47 McKinney, W. Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference* (2010).

48 Waskom, M. *et al.* mwaskom/seaborn: v0.11.0 (September 2020). doi:10.5281/ZENODO.4019146 (2020).

49 Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* **47**, W636-W641, doi:10.1093/nar/gkz268 (2019).

50 Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* **42**, W320-W324, doi:10.1093/nar/gku316 (2014).

51 Masood, T. B., Sandhya, S., Chandra, N. & Natarajan, V. CHEXVIS: a tool for molecular channel extraction and visualization. *BMC Bioinformatics* **16**, 119, doi:10.1186/s12859-015-0545-9 (2015).

52 Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105-132, doi:10.1016/0022-2836(82)90515-0 (1982).

53 Wust, J., Sullivan, N. M., Hardegger, U. & Wilkins, T. D. Investigation of an outbreak of antibiotic-associated colitis by various typing methods. *J Clin Microbiol* **16**, 1096-1101, doi:10.1128/JCM.16.6.1096-1101.1982 (1982).

54 Stabler, R. A. *et al.* Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome Biol* **10**, R102, doi:10.1186/gb-2009-10-9-r102 (2009).

55 Stubbs, S. L., Brazier, J. S., O'Neill, G. L. & Duerden, B. I. PCR targeted to the 16S-23S rRNA gene intergenic spacer region of *Clostridium difficile* and construction of a library consisting of 116 different PCR ribotypes. *J Clin Microbiol* **37**, 461-463, doi:10.1128/JCM.37.2.461-463.1999 (1999).

Table 1. Bacterial strains and plasmids used in x-ray crystallography and protein-protein interaction studies.

Strain or plasmid	**Description**	**Reference/ Source Application**
***C. difficile* strains**		
CD630	Ribotype 012, SLCT-7	[53]
R20291	Ribotype 027, SLCT-4	[54]
FM2.5	R20291 *slpA* 282_283insA	[3]
RΔD2	FM2.5 *slpA*ΔD2	This study
R7404	Ribotype 017, SLCT-7b	[55]
***E. coli* strains**		
NEB5a	fhuA2 Δ(*argF-lacZ*)U169 *phoA glnV44* Φ80 Δ(*lacZ*)M15 *gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	New England Biolabs
BL21 (DE3)	E. *coli* str. B F– *ompT gal dcm lon hsdSB(rB–mB–)* λ(DE3 [*lacI lacU*V5-T7p07 *ind1 sam7 nin5*]) [*mal*B+]K-12(λS)	Novagen
Rosetta (DE3)	F- *ompT hsdSB*(*rB- mB-*) *gal dcm* (DE3) pRARE (CamR)	Novagen
Plasmids		
pOB001	pMTL960-Ptet-slpAΔD2 (R20291)	This study. RF102/ RF103
pRPF170	pMTL960-Ptet-slpA (CD630)	This study. NF1414/ NF1415
pRPF233	pMTL960-Ptet-slpA (R20291)	[3]
Plasmids for recombinant expression of mature protein		
pABS17	pET28a-LMW-6xHis-tag (R20291)	This study. LMW of R20291; oABS46/ oABS47
pABS18	pET28a-HMW-6xHis-tag (R20291)	This study. HMW of R20291; oABS44/ oABS45
pABS19	pET28a-LMW-6xHis-tag (R7404)	This study. LMW of R7404; oABS1/ oABS2
pABS20	pET28a-HMW-6xHis-tag (R7404)	This study. HMW of R7404; *Nco*I/*Xho*I subcloning from pJAK148 into pET28a

pABS21	pET28a-LMWDLID-6xHis-tag (R20291)	This study. LMW of R20291; oABS31/oABS48
pABS22	pET28a-HMWDHID-6xHis-tag (R20291)	This study. HMW of R20291; oABS15/oABS16
pABS23	pET28a-LMWDLID-6xHis-tag (R7404)	This study. LMW of R7404; oABS31/oABS32
pABS24	pET28a-HMWDHID-6xHis-tag (R7404)	This study. HMW of R7404; oABS39/oABS16
pHMW630	pET28a-HMW-6xHis-tag (CD630)	[5] HMW of CD630
pHMWΔ1-40	pET28a-HMWΔLID-6xHis-tag (CD630)	[5] HMW of CD630 lacking the N terminal HID
pLMW630	pET28a-LMW-6xHis-tag (CD630)	[5] LMW of CD630
pLMWΔ260-321	pET28a-LMWΔHID-6xHis-tag (CD630)	[5] LMW of CD630 lacking the C terminal LID
pJAK149	pETDuet-1-HID-6xHis-tag – LMW (R7404)	This study. Recombinant co-expression of mature LMW/HID of R7404; RF1396/ RF1397, RF1398/ RF1400, RF1394/ RF1395
pJAK147	pETDuet-1-HID-6xHis-tag – LID (R7404)	This study. Recombinant co-expression of mature LID/HID of R7404; RF1396/ RF1397, RF1398/ RF1400, RF1395/ RF1396
Plasmids for expression of the mature protein with a point mutation in the interaction domain, as specified in subscript		
pABS1	pET28a-LMW$_{F274A}$-6xHis-tag (CD630)	This study. oABS33/ oABS34
pABS2	pET28a-HMW$_{Y27A}$-6xHis-tag (R20291)	This study. oABS37/ oABS38
pABS3	pET28a-LMW$_{F273A}$-6xHis-tag (R20291)	This study. oABS35/ oABS36
pABS4	pET28a-HMW$_{Y27A}$-6xHis-tag (R7404)	This study. oABS5/ oABS6
pABS5	pET28a-LMW$_{F270A}$-6xHis-tag (R7404)	This study. oABS21/ oABS22
pABS6	pET28a-HMW$_{N19A}$-6xHis-tag (R7404)	This study. oABS3/ oABS4
pABS7	pET28a-HMW$_{D29A}$-6xHis-tag (R7404)	This study. oABS7/ oABS8

pABS8	pET28a-HMW$_{L31A}$-6xHis-tag (R7404)	This study. oABS9/ oABS10
pABS9	pET28a-HMW$_{N35A}$-6xHis-tag (R7404)	This study. oABS11/ oABS12
pABS10	pET28a-HMW$_{S39A}$-6xHis-tag (R7404)	This study. oABS13/ oABS14
pABS11	pET28a-LMW$_{D254A}$-6xHis-tag (R7404)	This study. oABS17/ oABS18
pABS12	pET28a-LMW$_{I259A}$-6xHis-tag (R7404)	This study. oABS19/ oABS20
pABS13	pET28a-LMW$_{Y279A}$-6xHis-tag (R7404)	This study. oABS23/ oABS24
pABS14	pET28a-LMW$_{G300A}$-6xHis-tag (R7404)	This study. oABS25/ oABS26
pABS15	pET28a-LMW$_{F306A}$-6xHis-tag (R7404)	This study. oABS27/ oABS28
pABS16	pET28a-LMW$_{R309A}$-6xHis-tag (R7404)	This study. oABS29/ oABS30
pHMW_Y27A	pET28a-HMW$_{Y27A}$-6xHis-tag (CD630)	This study. NF1386/ NF1387
pJAK186	pMTL960 Ptet-slpA HMW$_{Y27A}$ (CD630)	This study. NF1386/ NF1387
pRPF209	pMTL960-Ptet-slpA LMW$_{F274A}$ (CD630)	This study. NF1189/ NF1190

472

473

474

475

476 Table 2. Oligonucleotides used in this study.

Name	Sequence	Application
oABS1	GATCCCATGGCAGATAGTAC	Amplification of LMW R7404 with *Nco*I site forward primer
oABS2	GATCCTCGAGAGATTTAGTTTC	Amplification of LMW R7404 with *Xho*I site reverse primer
oABS3	GCTAAATTAAAAGATTTAAAAGATTATGTAG	Introduction of N19A point mutation in R7404 HMW
oABS4	AGCTTTTATAGTTATTTTAGCTGG	Introduction of N19A point mutation in R7404 HMW
oABS5	GCTGTAGATGATTTAAAAACATAC	Introduction of Y27A point mutation in R7404 HMW
oABS6	ATCTTTTAAATCTTTTAATTTATTAGC	Introduction of Y27A point mutation in R7404 HMW
oABS7	GCTGATTTAAAAACATACAATAATAC	Introduction of D29A point mutation in R7404 HMW
oABS8	TACATAATCTTTTAAATCTTTTAATTTATTAG	Introduction of D29A point mutation in R7404 HMW
oABS9	GCTAAAACATACAATAATACTTACTCAAATG	Introduction of L31A point mutation in R7404 HMW
oABS10	ATCATCTACATAATCTTTTAAATCTTTTAATTTATTAG	Introduction of L31A point mutation in R7404 HMW
oABS11	GCTAATACTTACTCAAATGTTGTAAC	Introduction of N35A point mutation in R7404 HMW
oABS12	GTATGTTTTTAAATCATCTACATAATC	Introduction of N35A point mutation in R7404 HMW
oABS13	GCAAATGTTGTAACAGTAGCAG	Introduction of S39A point mutation in R7404 HMW
oABS14	GTAAGTATTATTGTATGTTTTTAAATCATC	Introduction of S39A point mutation in R7404 HMW
oABS15	GCGCGCACAGTAGCAGGAGAAGATAGAATAG	Deletion of HID in R7404 HMW
oABS16	CATGGTATATCTCCTTCTTAAAGTTAAAC	Deletion of HID in HMW (R7404, R20291)
oABS17	GCTTCAAGTTCATATATTAGTGC	Introduction of D254A point mutation in R7404 LMW
oABS18	CACATCAATAGATTCTTCTTTTGC	Introduction of D254A point mutation in R7404 LMW
oABS19	GCTAGTGCTGAAAATTTAGC	Introduction of I259A point mutation in R7404 LMW
oABS20	ATATGAACTTGAATCCACATC	Introduction of I259A point mutation in R7404 LMW

oABS21	GCTAATCCTAAAGAGGTTTCTG	Introduction of F270A point mutation in R7404 LMW
oABS22	TACATATTTTTTAGCTAAATTTTCAGCAC	Introduction of F270A point mutation in R7404 LMW
oABS23	GCTAATGCAATAGTTGCATTAC	Introduction of Y279A point mutation in R7404 LMW
oABS24	AGCTTCAGAAACCTCTTTAGG	Introduction of Y279A point mutation in R7404 LMW
oABS25	GCAAAATATCAAGTTATTTTCTATCC	Introduction of G300A point mutation in R7404 LMW
oABS25	ATTAACTAATTGTACTAAATCAGATTC	Introduction of G300A point mutation in R7404 LMW
oABS27	GCTTATCCAGAAGGAAAAAGATTAG	Introduction of F306A point mutation in R7404 LMW
oABS28	AATAACTTGATATTTTCCATTAACTAATTGTAC	Introduction of F306A point mutation in R7404 LMW
oABS29	GCATTAGAAACTAAATCTCTCG	Introduction of R309A point mutation in R7404 LMW
oABS30	TTTTCCTTCTGGATAGAAAATAAC	Introduction of R309A point mutation in R7404 LMW
oABS31	GCGCGCCTCGAGCACCACCACCAC	Deletion of LID in LMW (R7404, R20291)
oABS32	TTTTATAGTACCTGTTGCAGCCATATC	Deletion of LID in R7404 LMW
oABS33	GCTGATCCAGATGAAATTTCTG	Introduction of F274A point mutation in CD630 LMW
oABS34	TACATATCTTTTAGCTAAATTTTCAGC	Introduction of F274A point mutation in CD630 LMW
oABS44	GATCCCATGGCTGCAAAGGCTTCAATTGCTG	Amplification of HMW R20291 with *Nco*I site forward primer
oABS45	GATCCTCGAGCATACTTAATAAATCTTTT AATTTATTTATAACTG	Amplification of HMW R20291 with *Xho*I site reverse primer
oABS46	GATCCATGGCAGAAGATATGTCGAAAGTTG	Amplification of LMW R20291 with 5′ *Nco*I site
oABS47	GATCCTCGAGACTCTTAGTTGTAACTCTTTTTCC	Amplification of LMW R20291 with 3′ *Xho*I site
oABS48	AGTTATTACTGGGCTTCCAGATTGTG	Introduction of deletion of LID in R20291 LMW
oABS35	GCTAATAAAACAGATTTAAATACTCTTTAC	Introduction of F273A point mutation in R20291 LMW
oABS36	TACATATTTTTTAGCTAAATCTTGTGCTG	Introduction of F273A point mutation in R20291 LMW
oABS37	GCTGTGGATGATTTAAGAACATATAATAATG	Introduction of Y27A point mutation in R20291 HMW

oABS38	ATCTTTTAAGTCTTTCTTCTTATCTGAC	Introduction of Y27A point mutation in R20291 HMW
oABS39	GCGCGCGAAGTAGCAGGAGAAGATAG	Deletion of HID in R20291 HMW
oABS40	GCTAGAACATATAATAATGGATATTCAAATG	Introduction of L31A point mutation in R20291 HMW
oABS41	ATCATCCACATAATCTTTTAAGTCTTTC	Introduction of L31A point mutation in R20291 HMW
oABS42	GCTAAAACATATAATAATACTTATTC	Introduction of L31A point mutation in CD630 HMW
oABS43	ATCATCTACATAATCTTTTAAATC	Introduction of L31A point mutation in CD630 HMW
NF1189	GCAGATCCAGATGAAATTTCTGAAGC	Introduction of F274A point mutation in CD630 LMW
NF1190	TACATATCTTTTAGCTAAATTTTCAGCTG	Introduction of F274A point mutation in CD630 LMW
NF1386	GCAGTAGATGATTTAAAAACATATAATAATACTTATTC	Introduction of Y27A point mutation in CD630 HMW
NF1387	ATCTTTTAAATCTTTTAATTTATTAGCTTTTATAAC	Introduction of Y27A point mutation in CD630 HMW
NF1414	GATCGAGCTCTATAATGTTGGGAGGAATTTAAGAAATG	Amplification of *slpA* from CD630 with 5′ *Sac*I site
NF1415	GATCGGATCCTTACATATCTAATAAATCTTTCATTTTG	Amplification of *slpA* from CD630 with 3′ *Bam*HI site
RF102	GGTTCTGGAAGCCCAGTAATAACTAAAC	Replacement of coding sequence of LMW domain 2 with GGA GGT
RF103	TCCAGAGCTTATTAAGAAATCTACATAATCC	Replacement of coding sequence of LMW domain 2 with GGA GGT
RF1393	GATCCATATGGCAGATAGTACTACGCCAGG	Amplification of LMW for insertion into pACYC-Duet1
RF1394	GATCGGTACCTTAAGATTTAGTTTCTAATCTTTTTCCTTCTG	Amplification of LID and LMW for insertion into pACYC-Duet1
RF1395	GATCCATATGGTTAGAGTTACAAGTGCAAAAGAAG	Amplification of LID for insertion into pACYC-Duet1
RF1396	CATGGTATATCTCCTTATTAAAGTTAAAC	Linearization of pACYC-Duet1 for insertion of HID
RF1397	CTCGAGCACCACCACCAC	Linearization of pACYC-Duet1 for insertion of HID
RF1398	TGTTTAACTTTAATAAGGAGATATACCATGGCAGATATAATAGCTGATGCAG	Amplification of HID for insertion into pACYC-Duet1
RF1400	ATCTCAGTGGTGGTGGTGGTGCTCGAGTACAACATTTGAGTAAGTATTATTGTATG	Amplification of HID for insertion into pACYC-Duet1

477

478

Supplementary Information – Discussion

Our structural models and functional analysis of the S-layer of *C. difficile* provide the first detailed insights of this important layer in a human pathogenic bacterium. This work allows us to explore this array in unprecedented detail, both elucidating key features and raising new questions requiring further investigation.

S-layer assembly: how can a 2D crystal array remain flexible?

A degree of conformational flexibility is required to accommodate the wrapping of the 2D S-layer lattice around the curved surface of the *C. difficile* cell. Indeed, dynamic flexibility between S-layer protein domains has been shown to promote efficient crystal nucleation on the curved cellular surface in *Caulobacter crescentus*[1]. Our recent work showed formation of *C. difficile* S-layer at specific sites coinciding with cell wall synthesis[2], suggesting discrete S-layer assembly points. Furthermore, Fourier analysis of S-layer ghosts and tomographic imaging (Extended Data Fig. 10) indicates a highly mosaic surface, with many crystal defects, particularly at the cell poles, where the paracrystalline array must be disrupted to allow for cell curvature. The pattern of crystalline patches with grain boundaries observed is consistent with the random secretion of S-layer protein monomers and self-assembly of 2D crystals occurring at gaps and grain boundaries within the curved S-layer, as proposed for other organisms[3-5].

While increasing numbers of S-layer structural models are available[6-11], to our knowledge, this is the first report of a complete X-ray structure of a major S-layer protein where the crystal lattice mimics S-layer assembly in the cell. This indicates that S-layer assembly in *C.*

difficile does not require an underlying ordered polysaccharide array, unlike the apparent organization observed in LPS-mediated S-layer anchoring in the Gram-negative *C. crescentus*[10]. As the *C. difficile* S-layer is anchored via interactions of the CWB2 motifs with PSII[12], a much simpler glycan that is unlikely to be ordered at the cell surface, it is not surprising that the protein can assemble independently. In order to elucidate the anchoring mechanisms of *C. difficile* S-layer, we are investigating the interactions of SPL_H and the H/L complex with PSII using a combination of biochemical, biophysical and structural methods.

SlpA is the main component of the S-layer in *C. difficile* but additional proteins, which together correspond to an estimated 10% of the protein molecules forming the array, must be accommodated within the layer. Our assembly model suggests that tiling of the CWB2 triangular prism present in SLP_H and in all minor cell wall proteins (CWPs)[13], is a mechanism that allows insertion of these proteins while maintaining the crystalline arrangement. In our crystal structure, SLP_H tiling is maintained by charge complementarity of interacting triangular prism surfaces (Extended Data Fig. 3b). Our structural analysis of homology models of other SLCTs suggests that most of the interactions between neighbouring H/L complexes which define those interfaces are conserved across different SLCTs (Extended Data Fig. 4), indicating that they are likely to be important for *C. difficile* S-layer assembly. It is worth noting that the most conserved interactions are at the interface of neighbouring $CWB2_3$-$CWB2_1$ motifs and around pore 2, and involve residues from both SLP_H and SLP_L (Extended Data Fig. 4b, top 6 rows). This conservation across SLCTs suggests that SLP_L is also important to maintain the S-layer assembly and that these interactions could be potential targets for disrupting the array.

Our analysis of the charge distribution of the CWB2 motifs in Cwp6 and Cwp8, the only other CWB2-containing proteins whose structures have been determined to date[14], indicates that the charge complementarity between H/L complex and these minor S-layer components would also be possible, supporting our proposed global assembly model. The specific insertion points of additional CWPs are yet to be determined and it is possible that these regions could create other mismatch points that would allow growth, accommodate curvature and confer flexibility to the S-layer.

This proposed model for insertion of additional proteins via tiling of CWB2 motifs raises the question of how the accessory domains in CWPs are incorporated in the SLP_L ridge features. Homology between CWPs and SlpA is restricted to the CWB2 trimeric motif as CWPs have distinct accessory domains replacing the SLP_L[13]. These structurally diverse domains have to be accommodated in the S-layer, in order to maintain the integrity of the crystalline array. Based on the observation that most of the X-ray crystallographic model of the H/L complex fits well into the envelope defined by the EM reconstruction, with the exception of the D2 domain, we speculate that this region of SLP_L might confer further flexibility to the assembled S-layer. It is also possible that D2 adopts a slightly different position relative to SLP_H in the mature S-layer as it could be better accommodated into the EM envelope, in the region corresponding to missing density in the $SlpA_{R\Delta D2}$ reconstruction, by a shift of 10° relative to D1 (Extended Data Fig. 6). Indeed, strikingly different conformations of the D1-D2 domains in SLP_L are observed in the H/L and SLP_L/HID structural models (Extended Data Fig. 2). Analysis of the architecture and simulated motions (DynDom6D v1.0) in these models identified regions of D1-D2 and LID/HID as two dynamic domains, with the linker between the D1 C-terminal β-strand and LID N-terminal β-strand acting as interdomain hinge residues (Extended Data Fig. 2). The calculated rotation angle of the centers of mass of the

D1-D2 region relative to LID/HID of 166° suggests a high degree of flexibility of these regions, at least in the absence of the CWB2-containing SLP_H core, with minimal effects on the fold of each individual domain. Together, these observations suggest that D2 could adopt a number of conformations in the mature S-layer, therefore facilitating the insertion of other S-layer proteins.

S-layer as a molecular sieve: how do molecules go in or out?

S-layers have been proposed to act as a molecular filter[15] but the bulk of the *C. difficile* cell surface seems virtually impenetrable to large molecules, even though cells appear to be able to secrete toxin even in the absence of cell lysis[16]. The pores we observe *in situ* would not allow for most proteins to directly diffuse in or out. Whether mismatch regions in the 2D paracystalline array (Extended Data Fig. 10) are sufficient for access and how these processes can be controlled are key questions to pursue to further our understanding of S-layer in *C. difficile*.

The pores observed in our crystal lattice (Fig. 3) are highly hydrophilic (Extended Data Fig. 3e), suggesting most small charged molecules could diffuse into the cell. Pore 1, mostly occluded by D2 in the full H/L complex, has a mixed charged distribution, with patches of both positive and negative charges throughout (Extended Data Fig. 3f), suggesting important metabolites such as ATP (negatively charged) or metal ions (positively charged) could have access. In contrast, the fully exposed pore 2, is mostly negatively charged, indicating that positively charged small metabolites could preferably be diffused via this pore. The presence of a positive patch formed by two pseudo symmetry-related lysines covering the outermost opening of this pore (Extended Data Fig. 3f) could provide a gating mechanism for these metabolites.

It is worth noting that interacting D1 domains from neighbouring SLP_L molecules completely cover the widest cavity in the SLP_H CWB2s tiling. This interface, defined by neighbouring $CWB2_1$-$CWB2_2$ motifs, at around 20 Å wide but spanning over 100 Å across the triangular prism tiles, is also hydrophilic, with complimentary charges (Extended Data Fig. 3b). The SLP_H CWB2 motifs tiling also creates a cavity of approximately 70 Å between symmetry-related molecules which is partly occluded by the HID and LID domains, with the interlocking D1 domain ridges covering this gap (Extended Data Fig. 3b). If the interacting D1 domains are flexible and can at least partially expose these cavities, it could potentially allow diffusion of larger molecules through the S-layer.

Absence of D2 creates a more permeable S-layer as it exposes pore 1 in the SLP_H tiling (Fig. 3, Extended Data Fig. 3). Moreover, many of the residues lining the two exposed pores in this lattice are not resolved in the electron density of the $SlpA_{R\Delta D2}$ and could not be modelled, suggesting weaker interactions. A scenario where D1-LID/HID interactions with the CWB2s are less stable and could allow access to the wider openings in the CWB2s tiles would possibly explain the susceptibility to lysozyme seen in the RΔD2 strain (Extended Data Fig. 8). It would also point to a role of the D2 domain in preventing access of at least some antimicrobials and this could relate to the reduced disease severity observed for RΔD2.

S-layer role in disease: potential mechanism

Our work with RΔD2 strain, producing a modified S-layer is, as far as we are aware, the first example of a *C. difficile* strain exhibiting both similar toxin and colonization levels but reduced disease severity in a mouse model. The reduction in weight loss upon infection with a strain of *C. difficile* lacking the surface-exposed D2 domain of SLP_L suggests that the S-layer and toxins act synergistically to mediate epithelial damage, highlighting the possible

120 contribution of the S-layer in delivery of toxin and enhancement of cellular responses
121 important for disease severity. Toxins have been shown to be essential for *C. difficile*
122 disease[17,18] and their role in glucosylation of small GTPases and key structural features have
123 been elucidated[19]. Toxin activity has also been linked to inflammatory influx through
124 activation of the inflammasome[20]. While the specific contribution of the D2 domain and S-
125 layer in disease remains unclear, several possibilities exist to explain their functional role.
126 One explanation is that modification to the S-layer alters the capacity of the bacteria to
127 adhere to the epithelial barrier, limiting proximity and impact of localized toxin on release.
128 This would reduce the overall impact of toxins on both barrier integrity and inflammasome
129 activation and hence limit tissue damage and disease. While this is feasible, the equivalence
130 of toxins detected in filtered gut washes from RΔD2 versus WT infected animals and the
131 epithelial cellular susceptibility to purified toxins, suggests this explanation is unlikely.
132 Alternatively, reduction in disease severity seen in mice infected with RΔD2 may be linked to
133 failure of host cells to effectively signal an immune response. The evidence that TLR (Toll-
134 like receptor) signaling is important in *C. difficile* disease is now significant, with several
135 studies showing that signaling through a MyD88 dependent pathway is essential in immune
136 protection[21,22]. The S-layer has been implicated in activation of this pathway through
137 activation of TLR4[23] and it is tempting to speculate that the D2 domain is important for this
138 function. *In vitro*, S-layer signaling in both mouse and human dendritic cells induces the
139 production of proinflammatory cytokines including IL-23, and enhances the activity of the
140 toxin-activated inflammasome[23]. However, as MyD88 and TLR4 -/- mice appear more
141 susceptible to disease, the failure of the modified S-layer ($SlpA_{R\Delta D2}$) to signal through this
142 pathway should lead to enhanced disease. This is in contrast to our results, in which animals
143 infected with RΔD2 showed reduced pathology.

144 Another possibility is that modifications in S-layer-mediated signaling influences the
145 downstream generation of specific proinflammatory cytokines such as IL-23. Infection of IL-
146 23 -/- mice with *C. difficile* resulted in limited tissue edema, reduced inflammatory influx
147 and less epithelial damage[24]; mirroring the observations in animals infected with the
148 modified RΔD2 strain in this study. Further, *in vitro*, a combination of cell filtrates and toxin
149 but not toxin alone, was required to stimulate the expression of IL-23 in both mouse and
150 human bone marrow-derived macrophages[25]. As these filtrates are likely to contain high
151 levels of SlpA, it could be speculated that IL-23 expression relies on two independent
152 signals, the first provided by the toxin (through inflammasome activation) and the second, a
153 MyD88-dependent event implicating TLR signaling, possibly involving SlpA. Therefore, if the
154 D2 domain is essential for TLR signaling, reduction of disease could be linked to prevention
155 of downstream IL-23 mediated events such as enhancement of inflammation and sustained
156 barrier damage.
157 While the specific mechanism by which S-layer is able to modify the severity of toxin
158 mediated-disease is currently unclear, we now have the tools and structural knowledge to
159 allow us to dissect and determine the relevance of the S-layer structure in *C. difficile*
160 disease.
161

References

1 Herrmann, J. *et al.* A bacterial surface layer protein exploits multistep crystallization for rapid self-assembly. *Proc Natl Acad Sci U S A* **117**, 388-394, doi:10.1073/pnas.1909798116 (2020).

2 Oatley, P., Kirk, J. A., Ma, S., Jones, S. & Fagan, R. P. Spatial organization of *Clostridium difficile* S-layer biogenesis. *Sci Rep* **10**, 14089, doi:10.1038/s41598-020-71059-x (2020).

3 Comerci, C. J. *et al.* Topologically-guided continuous protein crystallization controls bacterial surface layer self-assembly. *Nat Commun* **10**, 2731, doi:10.1038/s41467-019-10650-x (2019).

4 Garcia, N. A., Pezzutti, A. D., Register, R. A., Vega, D. A. & Gomez, L. R. Defect formation and coarsening in hexagonal 2D curved crystals. *Soft Matter* **11**, 898-907, doi:10.1039/c4sm02234c (2015).

5 Garcia, N. A., Register, R. A., Vega, D. A. & Gomez, L. R. Crystallization dynamics on curved surfaces. *Phys Rev E Stat Nonlin Soft Matter Phys* **88**, 012306, doi:10.1103/PhysRevE.88.012306 (2013).

6 Baranova, E. *et al.* SbsB structure and lattice reconstruction unveil Ca2+ triggered S-layer assembly. *Nature* **487**, 119-122, doi:10.1038/nature11155 (2012).

7 Bharat, T. A. M. *et al.* Structure of the hexagonal surface layer on *Caulobacter crescentus* cells. *Nat Microbiol* **2**, 17059, doi:10.1038/nmicrobiol.2017.59 (2017).

8 Fioravanti, A. *et al.* Structure of S-layer protein Sap reveals a mechanism for therapeutic intervention in anthrax. *Nat Microbiol* **4**, 1805-1814, doi:10.1038/s41564-019-0499-1 (2019).

9 Sychantha, D. *et al.* Molecular basis for the attachment of S-layer proteins to the cell wall of *Bacillus anthracis*. *Biochemistry* **57**, 1949-1953, doi:10.1021/acs.biochem.8b00060 (2018).

10 von Kugelgen, A. *et al. In situ* structure of an intact lipopolysaccharide-bound bacterial surface layer. *Cell* **180**, 348-358 e315, doi:10.1016/j.cell.2019.12.006 (2020).

11 Farci, D. *et al.* Structural insights into the main S-layer unit of *Deinococcus radiodurans* reveal a massive protein complex with porin-like features. *J Biol Chem* **295**, 4224-4236, doi:10.1074/jbc.RA119.012174 (2020).

12 Willing, S. E. *et al. Clostridium difficile* surface proteins are anchored to the cell wall using CWB2 motifs that recognise the anionic polymer PSII. *Mol Microbiol* **96**, 596-608, doi:10.1111/mmi.12958 (2015).

13 Fagan, R. P. & Fairweather, N. F. Biogenesis and functions of bacterial S-layers. *Nat Rev Microbiol* **12**, 211-222, doi:10.1038/nrmicro3213 (2014).

14 Usenik, A. *et al.* The CWB2 cell wall-anchoring module Is revealed by the crystal structures of the *Clostridium difficile* cell wall proteins Cwp8 and Cwp6. *Structure* **25**, 514-521, doi:10.1016/j.str.2016.12.018 (2017).

15 Sara, M. & Sleytr, U. B. S-Layer proteins. *J Bacteriol* **182**, 859-868, doi:10.1128/jb.182.4.859-868.2000 (2000).

16 Wydau-Dematteis, S. *et al.* Cwp19 Is a novel lytic transglycosylase involved in stationary-phase autolysis resulting in toxin release in *Clostridium difficile*. *mBio* **9**, doi:10.1128/mBio.00648-18 (2018).

17 Kuehne, S. A. *et al.* The role of toxin A and toxin B in *Clostridium difficile* infection. *Nature* **467**, 711-713, doi:10.1038/nature09397 (2010).

18 Lyras, D. *et al.* Toxin B is essential for virulence of *Clostridium difficile*. *Nature* **458**, 1176-1179, doi:10.1038/nature07822 (2009).

19 Pruitt, R. N. & Lacy, D. B. Toward a structural understanding of *Clostridium difficile* toxins A and B. *Front Cell Infect Microbiol* **2**, 28, doi:10.3389/fcimb.2012.00028 (2012).

20 Ng, J. *et al. Clostridium difficile* toxin-induced inflammation and intestinal injury are mediated by the inflammasome. *Gastroenterology* **139**, 542-552, 552 e541-543, doi:10.1053/j.gastro.2010.04.005 (2010).

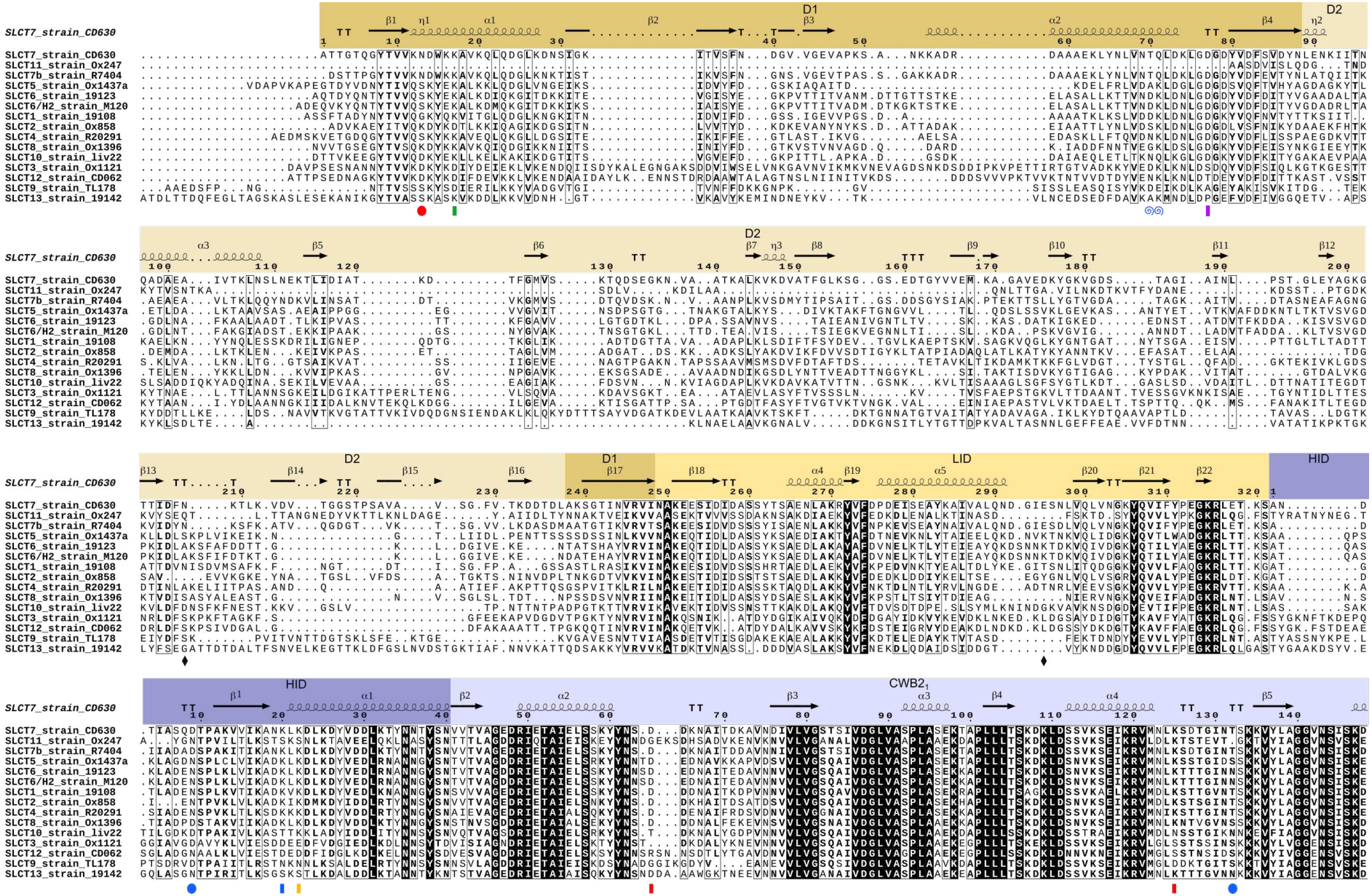
21 Jarchum, I., Liu, M., Shi, C., Equinda, M. & Pamer, E. G. Critical role for MyD88-mediated neutrophil recruitment during *Clostridium difficile* colitis. *Infect Immun* **80**, 2989-2996, doi:10.1128/IAI.00448-12 (2012).

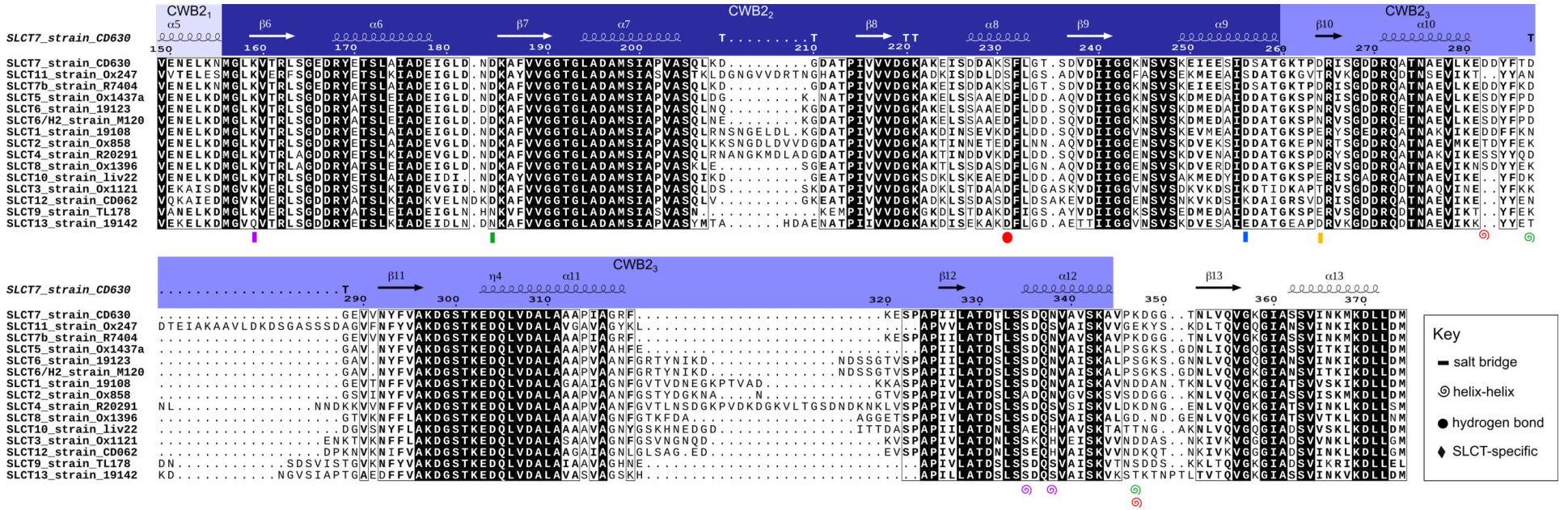
22 Mamareli, P. *et al.* Epithelium-specific MyD88 signaling, but not DCs or macrophages, control acute intestinal infection with *Clostridium difficile*. *Eur J Immunol* **49**, 747-757, doi:10.1002/eji.201848022 (2019).

23 Ryan, A. *et al.* A role for TLR4 in *Clostridium difficile* infection and the recognition of surface layer proteins. *PLoS Pathog* **7**, e1002076, doi:10.1371/journal.ppat.1002076 (2011).

24 McDermott, A. J. *et al.* 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**, 114-124, doi:10.1111/imm.12545 (2016).

25 Cowardin, C. A. *et al.* Inflammasome activation contributes to interleukin-23 production in response to *Clostridium difficile*. *mBio* **6**, doi:10.1128/mBio.02386-14 (2015).


SLCT7_strain_CD630
SLCT7_strain_CD630
SLCT11_strain_Ox247
SLCT7b_strain_R7404
SLCT5_strain_Ox1437a
SLCT6_strain_19123
SLCT6/H2_strain_M120
SLCT1_strain_19108
SLCT2_strain_Ox858
SLCT4_strain_R20291
SLCT8_strain_Ox1396
SLCT10_strain_liv22
SLCT3_strain_Ox1121
SLCT12_strain_CD062
SLCT9_strain_TL178
SLCT13_strain_19142
D1
D2
HID
LID
CWB2$_1$



Sequence alignment of representatives of each S-layer Cassete Type (SLCT) using CD630 as a reference. Secondary structural elements identified in the H/L CD630 complex (PDB ID: 7ACY) are indicated by looped lines – α-helices and arrows – β-sheets. SLP_L D1, D2 and LID domains and SLP_H HID domain and CWB2 motifs are highlighted, colored as in Fig. 1. Residues involved in the interactions depicted in Extended Data Fig. 4a are marked as per key, with interacting residues marked with the same color. Analysis of conservation of both residue and potential interaction across these SlpA variants is summarised in the clustermap in Extended Data Fig. 4b. Strictly conserved residues across all SLCTs are highlighted in black background, similar (partially conserved) groups are delimited by a box, with residues conserved within each group highlighted in bold, as per default in ESPript3[1] (http://espript.ibcp.fr). Reference **SLCT7 - strain CD630**; SLCT1 - strain 1912; SLCT2 - strain Ox858; SLCT3 - Ox1121; **SLCT4 - strain R20291**; SLCT5 - Ox1437a; SLCT6 - strain 19123; SLCT6/H2 - strain M120; **SLCT7b - R7404**; SLCT8 - Ox1396; SLCT9 - strain TL178; SLCT10 - strain Liv22; SLCT11 - strain Ox247; SLCT 12 - strain CD062; SLCT13 - strain 19142. Bold indicates strains with structural models included in this work.

1 Robert *et al. Nucleic Acids Res.* 42, W320–W324, 2014

Table 1 - **Data collection and refinement statistics**

	LID/HID	SLP_L/HID	H/L R7404 S-SAD	H/L R7404[1]	H/L CD630	H/L RΔD2
Data collection						
Space group	*C*2	*C*2	$P2_1$	$P2_1$	*P*1	*P*1
Cell dimensions						
a, *b*, *c* (Å)	73.3, 56.7, 61.8	172.9, 29.5, 144.3	76.7, 134.7, 83.9	78.1, 137.9, 84.7	72.7, 78.3, 81.6	52.8, 80.4, 81.9
α, β, *γ* (°)	90.0, 122.7, 90.0	90.0, 94.2, 90.0	90.0, 100.8, 90.0	90.0, 100.7, 90.0	81.9, 67.0, 65.3	97.0, 90.2, 90.2
Wavelength (Å)	0.975	0.969	2.755	0.928	0.969	0.969
Resolution (Å)	41.74- 1.50 (1.55-1.50)	86.20-2.40 (2.49-2.40)	47.70-3.00 (3.16-3.00)	83.26-2.65 (2.75-2.65)	52.30-2.55 (2.65-2.55)	52.74-3.50 (3.63-3.50)
I/σI	8.1 (2.3)	5.5 (1.9)	16.5 (1.2)	12.0 (1.4)	7.0 (1.5)	10.0 (2.5)
CC1/2	0.998 (0.975)	0.979 (0.713)	0.998 (0.707)	0.712 (0.360)	0.992 (0.732)	0.616 (0.438)
Completeness (%)	99 (97)	100 (100)	91 (80)	100 (100)	96 (97)	99 (97)
Redundancy	3.4 (2.6)	5.1 (5.2)	6.4 (6.1)	48.0 (43.0)	3.3 (3.4)	3.4 (3.5)
Anomalous completeness			91.2 (79.7)			
Anomalous multiplicity			3.3 (3.3)			
Refinement						
Resolution (Å)	41.74- 1.50	86.20-2.40		83.26-2.65	52.50-2.55	52.74-3.50
No. reflections	33987	29330		51027	47183	16607
R_{work} / R_{free}	18.1/21.0	25.2/30.1		22.6/27.8	23.0/25.7	28.4/31.4
No. atoms						
Protein	1709	3477		9945	10306	7890
Ligand/ion	-	-		40	15	5
Water	107	217		111	97	7
B-factors						
Protein	33.11	42.01		71.87	60.21	38.01
Ligand/ion	-	-		71.79	123.30	94.83
Water	37.55	39.32		59.43	45.98	8.48
Ramachandran %						
favoured	100.0	98.2		97.2	98.5	96.4
allowed (	0.0	1.8		2.6	1.5	3.6
outliers (%)	0.0	0.00		0.4	0.00	0.0
R.m.s. deviations						
Bond lengths (Å)	0.009	0.005		0.007	0.007	0.006
Bond angles (°)	1.14	0.98		1.19	1.11	0.97
PDB ID	7ACW	7ACV		7ACX	7ACY	7ACZ

*Values in parentheses are for highest-resolution shell.

[1]Two crystals were used to determine the structure of H/L complex R7404. All others required one crystal only.

Table 2 - EM data and reconstruction statistics

	$SlpA_{R20291}$	$SlpA_{R\Delta D2}$
3D merging statistics (EM reconstructions; negative staining)		
Resolution limit (Å)	20	20
No. structure factors	1085	667
Overall R-factor	0.33	0.33
Overall phase residual (°)	22.3	13.9
Phase residuals in CryoEM projections (*p2*-averaged Fourier terms)		
No. independent phases		
Resolution shell (Å)		
∞ – 15	43	43
15 - 11	42	42
11 - 8.7	47	45
8.7 – 7.5	44	-
Mean value phase error*		
Resolution shell (Å)		
∞ – 15	15.3	21.5
15 - 11	22.9	28.7
11 - 8.7	32.2	33.8
8.7 – 7.5	36.8	-
Standard error (°)		
Resolution shell (Å)		
∞ – 15	2.4	3.5
15 - 11	2.8	4.1
11 - 8.7	3.6	3.6
8.7 – 7.5	3.9	-

*Mean value phase error against symmetry-imposed phase of 0° or 180° (45° is expected for random phases [1]).

Table 3 – Symmetry table for cryo-EM reconstruction of $SlpA_{R20291}$

Plane Group [1]	**Phase residual** (°) **(random = 90°)**	**Target residual** [2] (°)
p1	27.0	-
p2*	37.5	39.7
p12b	77.6	29.8
p12a	55.1	29.8
p121b	57.2	29.8
p121a	51.0	29.8
c12b	77.6	29.8
c12a	55.1	29.8
p222	67.7	33.6
p2221b	52.7	33.6
p2221a	47.5	33.6
p2221a	62.2	33.6
c222	67.7	33.6
p4	46.8	33.6
p422	61.9	30.1
p4212	63.5	30.1
p3	52.5	27.0
p312	60.4	27.7
p321	63.0	28.5
p6	50.4	31.6
p622	58.1	29.3

*Represents most likely plane group

[1]a and b represent the respective symmetry axis for the plane group

[2]Target residual indicates the expected phase residual of each symmetry group based on the signal-to-noise ratio of the respective reflections [2].

1 Bullough, P. A. & Henderson, R. Phase accuracy in high-resolution electron microscopy of trigonal and orthorhombic purple membrane. *Biophys J* **58**, 705-711, doi:10.1016/S0006-3495(90)82413-9 (1990).

2 Valpuesta, J. M. a., Carrascosa, J. L. & Henderson, R. Analysis of electron microscope images and electron diffraction patterns of thin crystals of Ø29 connectors in ice. *Journal of Molecular Biology* **240**, 281-287, doi:10.1006/jmbi.1994.1445 (1994).