Evolutionary plasticity of bacterial surface layer protein exoskeletons

Anna Barwinska-Sendra*[1] ✉, Paula S. Salgado[1] ✉, Kacper M. Sendra*[1] ✉

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

*These authors contributed equally to this work.

✉ Authors to whom correspondence should be addressed:

Kacper M. Sendra, email: kacper.sendra@newcastle.ac.uk.

Anna Barwinska-Sendra, email: anna.barwinska-sendra@newcastle.ac.uk.

Paula S. Salgado, email: paula.salgado@newcastle.ac.uk.

Abstract

All cellular life possesses environmental interfaces like cell membranes or cell walls, yet the compositional complexity of these major cell components limits structural and evolutionary studies. Prokaryotic surface layer (S-layer) exoskeletons, with their homopolymer paracrystalline architecture, offer a more tractable evolutionary model of an environmental interface. In this study, we reveal the functional, structural and evolutionary diversity of S-layers in Gram-positive *Peptostreptococcaceae*, including pathogens and cancer-promoting species. We uncover novel S-layer architectures with diverse biochemical and physiological properties, enabled by a modular design co-evolving with other cell envelope components. We elucidate the mechanisms and evolutionary pathways underpinning the emergence of novel S-layers and the diversification of existing ones. Our findings establish the S-layer as a paradigm of cellular and molecular modularity and evolutionary plasticity. The demonstrated adaptability of these biological exoskeletons enables rapid reconfiguration of bacterial cell surface architecture and physiology, facilitating immune evasion in pathogens.

Introduction

The cell wall and cell membrane usually constitute the core components of the bacterial cell envelope, the interface between a bacterial cell and its environment. The layered architecture of the cell envelope can be further augmented by the presence of an outer cell surface-layer (S-layer) exoskeleton. S-layers are found across Archaea, Gram-positive, and Gram-negative Bacteria[1-4], are visible using high-resolution imaging techniques, and their building blocks are some of the most abundant cellular proteins. The S-layers studied so far are generally composed of a major protein that assembles into the paracrystalline lattice, and S-layer-associated proteins that confer additional functionality[5], as observed in pathogens such as *Bacillus anthracis*[6] and *Clostridioides difficile*[7]. Proposed physiological roles of S-layers and S-layer proteins include protection against cell wall degradation and osmotic stress[8-11], biofilm formation[12-14], adhesion[15-17], phage recognition[18,19], and/or function as a molecular sieve[20,21]. Understanding the biology of S-layers is significant within biotechnology, with potential applications as scaffolds in bio-product design[22,23], drug nanocarrier[24] or vaccine delivery system[25], and within healthcare, since they have been implicated in immunomodulation by microbiota important for optimal health[26-29] and in the infection process of different pathogens[30-32], and are a target for the development of novel therapeutics[32,33].

Although the diversity of S-layers across different, distantly related bacterial taxa is well established[2,34,35], general evolutionary mechanisms shaping this diversity across the Tree of Life remain unexplored. S-layer proteins, including those sharing the same structural fold, often lack any identifiable sequence homology[2]. This complicates studies of S-layer origins and evolution and provides a major challenge for the identification of S-layer proteins and prediction of conserved S-layer functions such as protection against the human immune system. Recent accomplishments in structure-based predictive algorithms provide major advancements in our capacity to predict S-layer protein candidates and S-layer architectures[34,35]. However, due to their extreme sequence variation and low number of experimentally verified and studied S-layers, our understanding of functional diversity, ecological significance, and evolution of this important element of prokaryotic cell biology is still limited.

The evolutionary origins of these paracrystalline layers surrounding entire prokaryotic cells, including that of many pathogens, are also not clear. It is commonly hypothesised that S-layers have originated multiple independent times[34,36], mainly based on the sequence and structural diversity of S-layer proteins and lattices[34]. However, the observed variation of surface exoskeletons could, at least partially, also be explained by an alternative hypothesis of a common origin followed by an extreme divergence and lateral transfer of S-layer proteins. In pathogenic *C. difficile*, the S-layer protein (*Cd*SlpA), encoded within the genomic S-layer cassette, is shared across *C. difficile* isolates *via* homologous recombination[37], and its highly

variable sequence enables its use for typing (S-layer cassette type, SLCT) of different clinical isolates[37,38]. A homologue of *Cd*SlpA, *Peptostreptococcus anaerobius* surface protein PCWBR2 was recently implicated in promoting colorectal carcinogenesis[39]. Despite their medical significance, our understanding of the structural and functional landscape of S-layers formed by these diverse orthologues from closely related species is limited and current experimental structural information is available for only two *C. difficile* sequence types[40].

Here, to address these major gaps in our understanding of S-layer biology and to circumvent the inherent challenges in studying these cellular structures, we focus on narrower phylogenetic scales, where S-layer evolution is more tractable. We experimentally characterise novel S-layer types and present evidence of independent S-layer losses and gains across a bacterial family *Peptostreptococcaceae*. We reveal the role of modular evolution in shaping S-layer diversity and enabling anchoring to variable cell wall architectures. We present novel structures of *C. difficile* S-layers and uncover how, at the narrowest phylogenetic scales, extreme sequence divergence, recombination, and adaptive evolution drive extensive structural changes of the crystalline lattice surrounding the bacterial cells, without compromising its role in protection against the human immune system. Physiological tests and CRISPRi experiments demonstrate that the resistance to the innate immunity effectors emerged specifically in pathogenic lineages. Our results provide insights into the mechanisms of adaptation of host-exposed cell surfaces of S-layer-containing pathogens, and the evolution of the environmental interfaces across the bacterial Tree of Life.

Results

The evolutionary and structural diversity of S-layers across the Tree of Life

It is often hypothesized that S-layers are common in Prokaryotes[1-3]. To test this hypothesis, we compiled a list of previously studied examples of S-layer-positive species and S-layer proteins with strong experimental evidence identified in the available literature, including verification using high-resolution imaging and molecular techniques (Supplementary Data 1). Mapping this evidence onto the Tree of Life (Fig. 1a) revealed that, although verified S-layer-positive species are distributed across many lineages, they represent only a small subset of the analysed Bacteria and Archaea. This reflects the disparity between the number of known species of Prokaryotes and those that have been cultured[41], of which only a small number have been tested for the presence of S-layers. Our analysis highlights the extensive gaps in the current understanding of S-layer diversity and distribution.

To address these gaps, we investigated the distribution of homologues of the known full-length S-layer protein sequences and their constituent domains across the Tree of Life using protein sequence homology searches (Supplementary Fig. 1), hidden Markov model (HMM) profile searches (Supplementary Fig. 2), and protein similarity network analyses (Supplementary Fig. 3). The main building block of the paracrystalline S-layer array, here referred to as Surface Layer Lattice Protein (SLLP), usually displays a bipartite architecture[2]. It consists of a cell anchoring domain ($anchor_{DOM}$), which often represents a general anchoring mechanism of cell surface proteins, and an S-layer 'crystallisation' or surface-forming domain ($surface_{DOM}$), which plays a more specific role in forming the paracrystalline lattice layer exposed to the extracellular environment (Fig. 1b and Supplementary Fig. 1a). Anchoring domains such as SLH have additional physiological roles, including outer membrane tethering in diderm Negativicutes[42], and in our analysis, are found across a broad taxonomic range, often in multiple proteins other than SLLPs from the same genome (Supplementary Fig. 3g). In addition, multiple different types of $anchor_{DOM}$s are often found within a single cell-anchored protein family (Supplementary Fig. 3b, d), indicating that they can be swapped during cell-anchored protein evolution. Therefore, $anchor_{DOM}$s are not specific to SLLPs and constitute swappable protein modules that enable the binding of surface proteins to alternative cell envelope architectures. This likely reflects the structural co-evolution of different cell-envelope layers, such as the cell wall and the S-layer.

In contrast to the $anchor_{DOM}$s, specific surface domain sequence profiles are less common and display narrower taxonomic distribution, suggesting lineage-specific adaptations and more specific functions in S-layer formation (Fig. 1a and Supplementary Fig. 3c). The majority of the analysed $surface_{DOM}$s, including those sharing the common immunoglobulin-like domain architecture (Fig. 1b, and Supplementary Fig. 1a), display extensive sequence diversity. The substantial discontinuity in the structural and sequence landscape of bacterial S-layers is further apparent in SLLPs sharing similar domain folds (Fig. 1b, and Supplementary Fig. 1a) but no sequence homology (Supplementary Figs. 1c, 2a, 3b) like bacterial *B. anthracis* EA1 and *Deinococcus radiodurans* HPI, or archaeal *Haloferax volcanii* Csg, which assemble into extremely diverse paracrystalline layers (Fig. 1b) and utilise different cell surface anchoring mechanisms (Fig. 1b, and Supplementary Fig. 1a).

Combined with the identified gaps in the appreciation of S-layer diversity and distribution, and the possible existence of SLLPs with distinct $surface_{DOM}$ folds, such as *Tannerella forsythia* TfsA (Supplementary Fig. 1a), our analyses are consistent with the existence of undiscovered diversity of prokaryotic exoskeletons in nature. However, the lack of sequence homology often observed between SLLPs belonging to the same structural family and the utilisation of common structural folds complicate analyses of their evolution over wide phylogenetic scales. Given the structural and sequence diversity of surface-forming domains and the multitude of anchoring mechanisms, it is likely that the various S-layers have either

emerged multiple independent times from non-SLLPs and/or are a result of extreme surface$_{DOM}$ divergence and anchor$_{DOM}$ swapping among existing S-layer proteins.

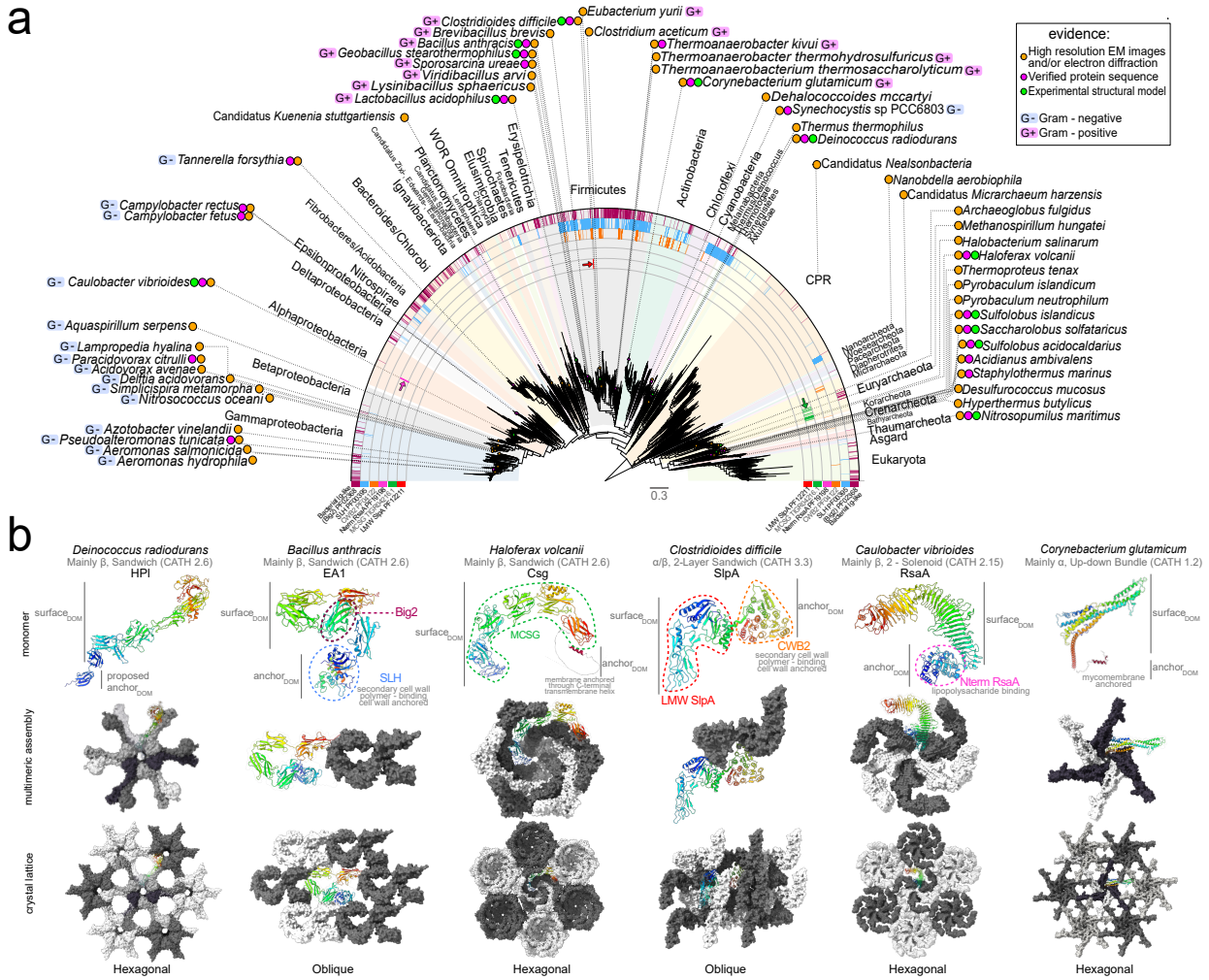

Fig. 1: Diversity of S-layer protein lattices across the Tree of Life.

a, Species names of representative 47 Bacteria and 21 Archaea with previously experimentally verified S-layers (Supplementary Data 1) were annotated onto a phylogeny of 2,705 Bacteria, 315 Archaea, and 149 Eukaryota. Coloured circles represent the available experimental evidence including verification *via* high-resolution imaging techniques (orange), structural biology (green), and protein sequence verification of SLLP (magenta). The coloured semi-circles represent hmmsearch results using HMM profiles of domains found in SLLPs with known structures presented in **b**, demonstrating the presence of both common (e.g. Big2 and SLH), and species- or lineage-specific domain profiles (coloured arrows, e.g. LMW SlpA, SLP$_L$).
b, Cartoon representations of structural models of SLLPs (monomer) and surface representations of their higher order organisation (multimeric assembly and crystal lattice) within the S-layer. The sequence regions (dashed outlines) identified with the available HMM profiles (**a** and Supplementary Fig. 2) can be found within two distinct domains, mainly involved in either cell surface anchoring (anchor$_{DOM}$) or paracrystalline S-layer lattice formation (surface$_{DOM}$). The experimental surface$_{DOM}$ structures (Supplementary Fig. 1a) belong to at least four types of structural folds identified using Foldseek (CATH: 2.6, 3.3, 2.15, 1.2). Multiple monomers (shades of grey) within the multimeric assemblies assume diverse higher-order geometries, which constitute a repeating unit within the crystal lattice with varying degrees of complexity and symmetry types (e.g. hexagonal and oblique). These diverse S-layer architectures are attached to the cell surfaces using a variety of mechanisms including membrane anchoring through transmembrane helices, and binding to different secondary cell wall polymers (SLH and CWB2) or lipopolysaccharides (Nterm RsaA).
Structures of the previously determined representative S-layer lattice proteins (8CKA, 8OPR, 7PTR, 7ACY, 6Z7P, 9GK2) were visualised using ChimeraX.

148 **Novel SLLPs form structurally diverse S-layers within a single family of Bacteria**

149 To experimentally investigate the diversity of the S-layer exoskeletons at narrower phylogenetic scales,
150 we first identified the lineages with a high degree of variation in the known SLLP domains. Our analyses
151 across the Tree of Life (Fig. 1a) revealed an interspersed distribution of different anchoring domains – SLH
152 and CWB2 – in Firmicutes. This bacterial phylum includes two important pathogens*, Bacillus anthracis* and
153 *Clostridioides difficile*, with two distinct S-layer surface domain folds (Fig. 1b). Within Firmicutes,
154 *Peptostreptococcaceae* represent a particular 'hotspot' of SLH/CWB2 anchoring domain repertoire
155 variation. Therefore, this lineage was selected for combinatorial bioinformatic, cell biology, microbiological,
156 and structural biology analyses to characterise their S-layers and SLLPs and unravel evolutionary
157 mechanisms shaping their diversity.

158 The experimental investigations (Fig. 2a and d, and Supplementary Fig. 4a) revealed four
159 *Peptostreptococcaceae* (29%, 4/14 species) with no apparent S-layer in whole cell electron micrographs, no
160 'ghosts' in S-layer electron microscopy (EM) preparations, and no abundant protein bands (Supplementary
161 Fig. 4a), typically observed for SLLPs, in surface protein preparations from S-layer-positive species. In three
162 out of four S-layer-negative species, the identified instances of apparent S-layer absence had scattered
163 distribution across the phylogeny, and no identifiable genomic SLLP homologue sequences, consistent with
164 multiple independent losses of S-layer (Fig. 2d). The existence of multiple S-layer-negative lineages of
165 Bacteria could lend some support to the hypothesis that S-layers have evolved multiple independent
166 times[2,34]. However, the presence of life cycle-specific or environmentally induced S-layers, or S-layer loss in
167 laboratory conditions, cannot be fully excluded. This is apparent in previously described difficulties in the
168 detection of S-layers in some bacteria[43] and, in our dataset, in the single case of *Romboutsia hominis* RFIFI,
169 which contains an SLLP homologue but displays an S-layer-negative morphotype (Fig. 2a and d, and
170 Supplementary Fig. 4). Therefore, currently available data suggests the possibility of independent S-layer
171 losses within *Peptostreptococcaceae*, although these represented the minority of the studied species.

172 In the majority (71%, 10/14 species) of the studied *Peptostreptococcaceae*, we found evidence of S-
173 layers and SLLPs (Fig. 2d; and Supplementary Fig. 4a). AlphaFold2 structural predictions and Foldseek
174 analyses revealed that six of the identified SLLPs (60%, 6/10 SLLPs) contained surface domains homologous
175 to the α/β-sandwich structural fold (Fig. 2b, c), previously only found in *C. difficile* SLLP (*Cd*SlpA). This type
176 of SLLP, with only two or three surface-exposed subdomains (Fig. 2c), features a distinctively shorter
177 $surface_{DOM}$ than other identified SLLP types and can potentially assemble into a more compact S-layer
178 lattice, as observed in *C. difficile*. Another distinction of this SLLP type is the higher proportion of α-helices
179 within its $surface_{DOM}$[40], as opposed to predominantly β-strand $surface_{DOM}$s of S-layers commonly found
180 across the Tree of Life (Fig. 1b). Four of these shorter SLLPs were found in close relatives of *C. difficile*,
181 including *Asaccharospora irregularis* and *Clostridioides mangenotii*, suggesting that this type of S-layer was
182 present in their last common ancestor. Interestingly, the SLLP from *Peptostreptococcus anaerobius*
183 contained the same fold, and our experiments identified it as PCWBR2 (Supplementary Fig. 4a), a protein
184 implicated in promoting colorectal carcinogenesis[39].

185 Three other identified *Peptostreptococcaceae* SLLPs (30%, 3/10) contained β-sandwich $surface_{DOM}$s,
186 with two from the most basal *Peptostreptococcaceae* outgroup lineages (Fig. 2a-d and Supplementary Fig.
187 4). Interestingly, while α/β-sandwich $surface_{DOM}$ shared enough sequence homology to group weakly within
188 a single protein similarity network, the β-sandwich fold SLLPs grouped within two distinct networks
189 (Supplementary Fig. 4b). The more broadly distributed β-sandwich S-layers were found within the outgroup
190 lineages including *Acetoanaerobium sticklandii* and *Peptoanaerobacter stomatis*. Their broader distribution
191 across the species tree suggested more ancient S-layer acquisition, contrasted by the second type of β-
192 sandwich S-layers, found only in a narrow subsample of closely related *Intestinibacter bartlettii* genome
193 assemblies (Fig. 2e). Inspection of all available *I. bartlettii* genomes revealed a recent SLLP acquisition likely

194 *via* transposition, indicated by its co-distribution with an adjacent IS256 gene (Fig. 2e) encoding a
195 transposase homologue. Furthermore, structural and sequence comparisons of the *I. bartelettii* SLLP
196 revealed that it has evolved *via* multiple tandem duplications of a single common β-sandwich
197 Immunoglobulin-like (PFAM ID, PF16403) surface subdomain (Fig. 2f). At this stage, we were able to
198 sequence-verify SLLP from the single *I. bartelettii* isolate available in culture collections. The significance of
199 subdomain number variation observed across available *I. bartlettii* genome assemblies remains unclear, as
200 SLLP-coding sequences are often incomplete and found at the end of a contig. Nevertheless, all six
201 $surface_{DOM}$ subdomains of the verified sequence shared, on average, 96 % of pairwise nucleotide sequence
202 identity, consistent with the recent emergence of this β-sandwich SLLP type.
203 The single SLLP from *Peptoclostridium acidaminophilum* contained the third distinct type of S-layer
204 surface-forming domain, β-barrel (10%, 1/10 SLLPs). Its similarity to the β-sandwich fold suggests that it
205 may have evolved either from a shared β-sandwich SLLP ancestor or originated independently from a non-
206 SLLP β-barrel-containing protein. The β-barrel SLLP contained the largest $surface_{DOM}$ among the identified
207 *Peptrostreptococcaece* SLLPs, with an extended 10-subdomain architecture, previously also found in the β-
208 sandwich S-layer of ammonia-oxidizing marine archaea[21]. The available crystal structures or calculated
209 SymProFold S-layer lattice models indicate that the β-barrel SLLP can assemble into an hexagonal lattice, in
210 contrast to the oblique lattice determined in *C. difficile* α/β-sandwich S-layer or the predicted square
211 symmetry of the *A. sticklandii* β-sandwich S-layer model (Fig. 2c).
212 By combining bioinformatics and experimental investigations, we uncovered novel types of SLLPs and S-
213 layers, recent SLLP emergence and S-layer acquisition, and multiple independent SLLP losses in
214 *Peptostreptococcaceae*. The revealed structural diversity of *Peptostreptococcaceae* SLLPs, including
215 multiple distinct S-layer surface-forming folds assembled into multiple distinct S-layer architectures, reflects
216 what we observed in our analyses across the Tree of Life (Fig. 1). Therefore, the presence of diverse S-layer
217 architectures utilising different surface domain folds does not only represent distinct features of distantly
218 related bacterial phyla but can be generated within much shorter evolutionary time frames. The emergence
219 of a novel SLLP-type *via* tandem subdomain duplication in *I. bartelettii* provides an insight into the
220 mechanism likely involved in the origins of novel S-layer architectures.
221

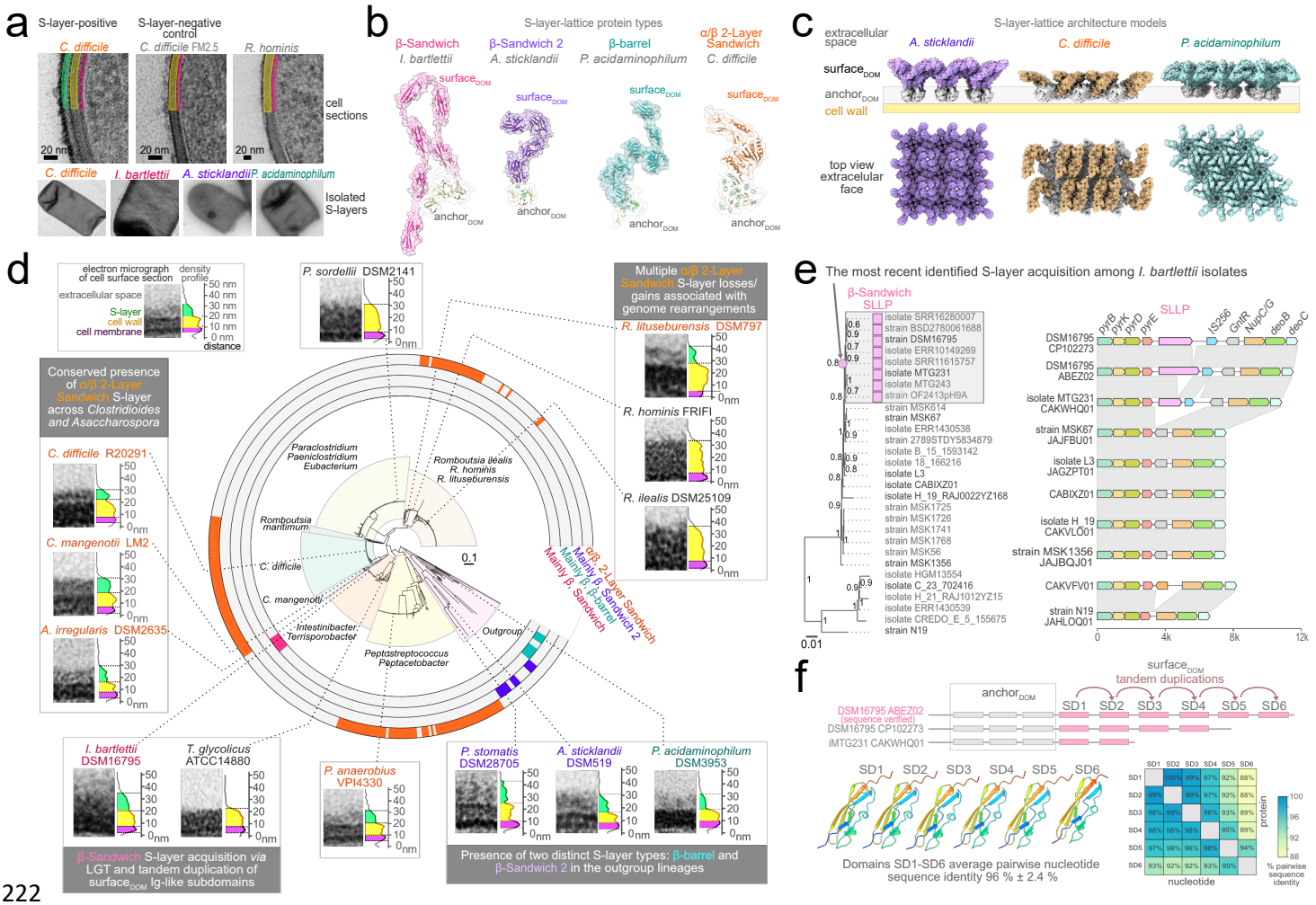

Fig. 2: Identification of new types of S-layer and mechanism of emergence of novel S-layer protein.

a, Representative electron micrographs of uranyl acetate and ruthenium red stained vegetative cells (top panel) of S-layer-positive *C. difficile,* S-layer-negative control of *C. difficile* FM2.5, and *R. hominis* with no apparent S-layer. The coloured regions correspond to S-layer (green), cell wall (yellow), and cell membrane (pink). S-layers (bottom panel) can be stripped from the vegetative cells due to the structural stability of their paracrystalline lattice.

b, Determined crystal structure (*C. difficile)* and AlphaFold2 predicted structural models of the S-layer lattice proteins (SLLPs) identified by mass spectrometry. The experimentally verified *Peptostreptococcaceae* SLLPs belonged to one of the four distinct surface domain ($surface_{DOM}$) folds.

c, Surface representations of S-layer models, including top-scoring SymProFold predictions (*A. sticklandii, P. acidaminophilum*), demonstrate that different *Peptostreptococcaceae* SLLP $surface_{DOM}$s (coloured) can assemble into distinct lattice architectures on the bacterial cell surface.

d, Species tree mapping of the *Peptostreptococcaceae* experimentally tested for the presence of S-layers and SLLPs. Density profiles next to the electron micrographs of the cell envelopes were colour-coded to indicate the location of the S-layer (green), cell wall (yellow), and cell membrane (pink). The scale bars measure distance (in nm) from the bottom of the micrograph. Species names were coloured based on the structural classification (**b**) of the verified SLLP $surface_{DOM}$ of *Peptostreptococcaceae* with apparent S-layer or those with no evidence of S-layer (black). Coloured semi-circles represent the distribution of the protein homologues of each SLLP-type identified in 453 *Peptostreptococcaceae* genomes.

e, The presence of the β-sandwich SLLP (pink rectangles) was mapped onto the species tree of *I. bartlettii* isolates (left panel). The recent acquisition of the SLLP (pink circle) is apparent in genomic region alignment (right panel), indicating ancestral insertion of the SLLP (pink) and IS256 transposase (blue) coding genes within a broadly conserved nucleotide metabolism locus.

f, Domain architecture of the *I. bartlettii* SLLPs (top panel) demonstrates a possible variation in the number of the β-sandwich $surface_{DOM}$ subdomains (SD1-6) among different isolates. All six SD1-6 (bottom panel, left) share high levels of pairwise protein and nucleotide sequence identity (bottom panel, right) consistent with SLLP origin *via* multiple tandem subdomain duplications.

247 **Homologous S-layers can impart different resistance to antimicrobial agents**

248 Although the structural and sequence diversity of S-layers at both wider and narrower taxonomic scales
249 is clear, the conservation of their physiological roles remains elusive. Furthermore, the function of S-layers
250 in many organisms has not been experimentally investigated[2]. S-layers of bacteria such as that of gut
251 pathogen *C. difficile* protect cells from natural antimicrobial agents produced by the host immune system
252 such as lysozyme or human cathelicidin peptide LL-37[10]. However, it is not known if this is also true for
253 other S-layers we identified across *Peptostreptococcaceae*.

254 Our initial experimental investigations demonstrated that *Peptostreptococcaceae* S-layers display
255 variable susceptibility to extraction using different chemical methods (Fig. 3a), indicating that at least some
256 of the properties of the outermost cell envelope layer of these bacteria are not conserved. Surprisingly, this
257 variation was not only observed for the S-layers made of different types of SLLPs but also across
258 homologous S-layers formed by the α/β-sandwich SLLPs, including that of *C. difficile* (Fig. 3a). Protein
259 similarity network analyses (Fig. 3b) and manual inspection of multiple sequence alignments revealed
260 extensive sequence divergence within this SLLP-fold. This is particularly apparent in the case of S-layer
261 $surface_{DOM}$s from different isolates of *C. mangenotii* (LM2 and DMS29133), which are not directly
262 connected with each other or with sequences from their closest relatives (*C. difficile* and *A. irregularis*)
263 within the network (Fig. 3b).

264 Structural comparisons of the experimentally verified α/β-sandwich SLLPs (Fig. 3c) and homologous S-
265 layer-associated proteins Cwp2 and Cwp8, revealed that, despite their extensive sequence variation, the
266 overall fold architecture remains conserved. The general α/β-sandwich SLLP $surface_{DOM}$ consists of three
267 subdomains, equivalent to *C. difficile* SlpA D1, D2 and ID[40] and S-layer associated Cwp8 D1, D2, D3[44], which
268 are connected via flexible linkers, providing further structural variability as rotation of the relative
269 orientation of the subdomains leads to varying surfaces (Fig. 3b,c). The AlphaFold3[45] predictions for the
270 experimentally verified SLLPs fold with three α/β-sandwich subdomains follows the topology seen in
271 *Cdl*SlpA, where D1 leads to the loop-rich D2. Insertion of a β-strand into D1 then links to the third
272 subdomain (ID), which connects to the CWB2 $anchor_{DOM}$. The most structurally conserved subdomain in
273 $surface_{DOM}$ is D1, with an average RMSD of 1.7 Å across all SLLPs, while the least conserved is D2 (average
274 RMSD 3.7 Å). The most distinct modification to the general α/β-sandwich architecture was observed in *A.*
275 *irregularis* DSM2635 where an additional, elongated α-helical region (~80 Å long) has been acquired and
276 inserted within D2 (Fig. 3c). These additional helices likely correspond to the protrusions extending out of
277 the lattice towards the extracellular environment observed in the cryo-electron micrographs of the *A.*
278 *irregularis* S-layer preparations (Fig. 3b). Importantly, acquisition of novel types of $surface_{DOM}$ subdomains
279 provides another mechanism of S-layer lattice evolution, alongside variation in the number of subdomains
280 as identified in *I. bartlettii* (Fig. 2f).

281 Next, we tested if S-layer dependent resistance to innate immune effectors observed in *C. difficile* is
282 conserved in other *Peptostreptococcaceae* containing α/β-sandwich S-layers. Physiological growth tests in
283 the presence of lysozyme and LL-37 revealed variable resistance to these natural antimicrobials (Fig. 3d).
284 Consistent with previous investigation[10], S-layer-positive *C. difficile* R20291, but not S-layer-negative *C.*
285 *difficile* FM2.5, displayed resistance against LL-37 and lysozyme. Interestingly, similar resistance was also
286 observed in cancer-promoting pathogen[39] *P. anaerobius* but not in any of the other α/β-sandwich S-layer
287 *Peptostreptococcaceae*, including closest *C. difficile* relatives *A. irregularis* and *C. mangenotii* (Fig. 3d). The
288 presented data is consistent with resistance to these antimicrobials representing specific and potentially
289 independent adaptations of *C. difficile* and *P. anaerobius* cell envelopes to the gut environment.

290 To further test the variable resistance of α/β-sandwich S-layers to antimicrobial agents, and to
291 experimentally verify that the identified SLLPs are essential for S-layer formation, we performed CRISPRi
292 gene knock-down experiments in *C. difficile* (Fig. 3e) and *R. lituseburensis* (Fig. 3f), two more distantly

293 related *Peptostreptococcaceae* species but with SLLPs sharing a close structural relationship, according to
294 Dali analyses (Fig. 3c). Consistent with the experiments in the wild type-strains (Fig. 3d), knock-down of the
295 SLLP gene abolishes *C. difficile* resistance to lysozyme and LL-37 (Fig. 3e), as opposed to *R. lituseburensis*,
296 where the presence of SLLP provided no apparent protection against either (Fig. 3f).
297 Previous experiments have also demonstrated that the presence of S-layer plays a role in *C. difficile*
298 sporulation[10]. In our experiments, CRISPRi S-layer knock-down strains of *C. difficile* and *R. lituseburensis*
299 exhibited sporulation deficiencies in the tested conditions (Figure 3e, f, Supplementary Table 1), suggesting
300 that S-layer integrity plays a conserved role necessary for sporulation of these two α/β-sandwich S-layer
301 containing *Peptostreptococcaceae*. In addition, both mutants displayed a propensity to aggregation in liquid
302 culture, highlighting the conserved role of their S-layers in its prevention.
303 In summary, despite their structural conservation, homologous α/β-sandwich S-layers from closely
304 related species display a high degree of biochemical and sequence diversity and differ in their contribution
305 to the protection of the cells they surround against natural antimicrobials lysozyme and LL-37. Despite
306 these differences, they also play conserved roles in cellular processes such as preventing cell aggregation
307 and sporulation, important in bacterial survival and infection. Therefore, not all phenotypic effects
308 associated with the presence of homologous S-layers are conserved. Certain functional features could
309 reflect specific ecological niches, such as the gut environment for *C. difficile* and *P. anaerobius*, while more
310 general life cycle adaptations like sporulation and cell-cell interactions might be under broader selective
311 pressure.

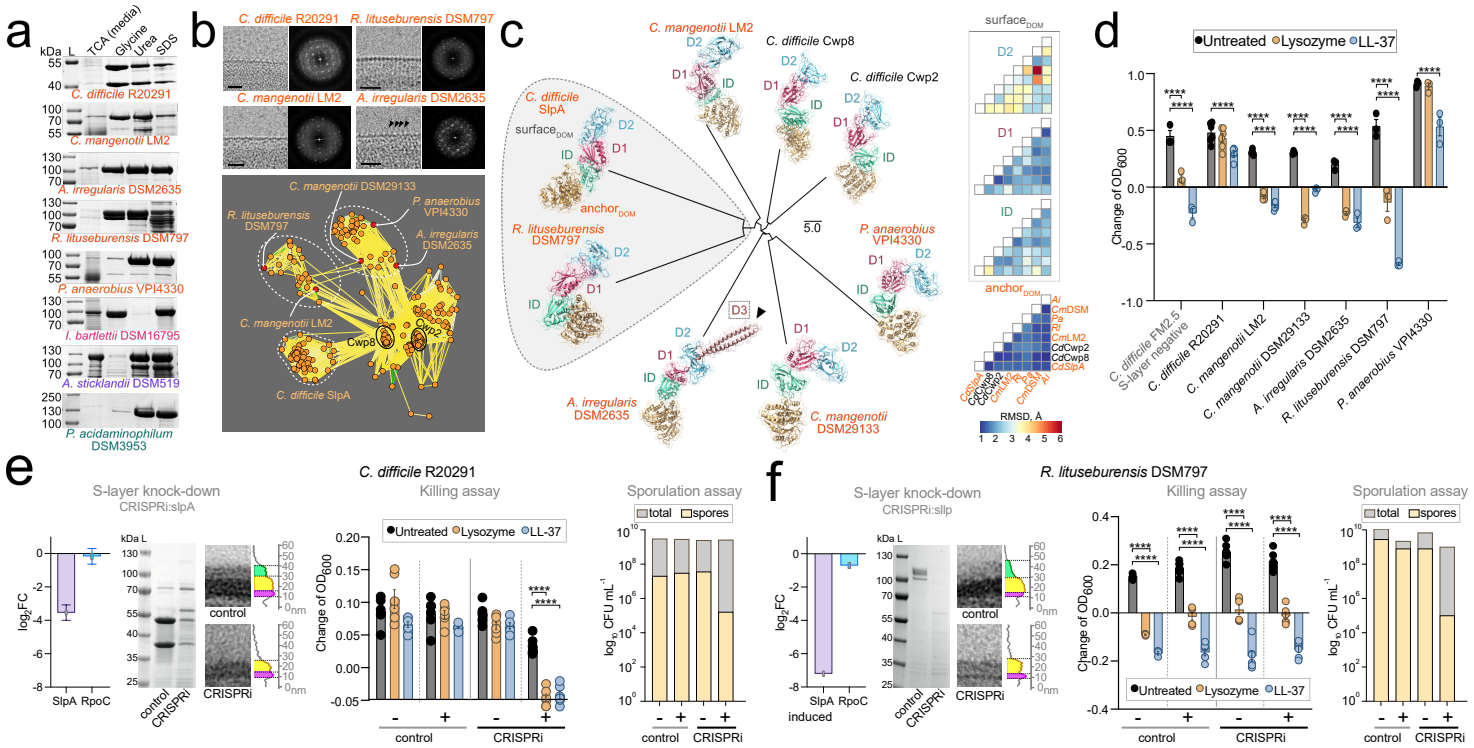

Fig. 3: Variable biochemical and physiological properties within a single structural type of S-layer.

a, SDS-polyacrylamide gels of *Peptostreptococcaceae* low pH glycine, urea, and SDS S-layer preparations. The α/β-sandwich S-layers (orange) displayed contrasting sensitivity to extraction in SDS and glycine. β-sandwich S-layers (pink or purple) varied when extracted in glycine and urea. Trichloroacetic acid (TCA) precipitation determined the SLLP presence in the medium. Molecular weight ladder (L).

b, Cryo-EM micrographs and Fast Fourier Transform (FFT) diffraction patterns of the α/β-sandwich S-layers (top panel). Additional protrusions were apparent at the extracellular face of *A. irregularis* S-layer (black arrowheads).

Protein sequence similarity network (bottom panel) of the α/β-sandwich SLLP surface$_{DOM}$s homologues (orange circles) including S-layer-associated Cwp2 and Cwp8 from the available *Peptostreptococcaceae* genomes. Highly divergent sequences of the verified S-layer surface$_{DOM}$s (red circles) formed three distinct groups (dashed outlines).

c, The Alphafold3 structural models of identified *Peptostreptococcaceae* α/β-sandwich SLLPs and Cwp2/8 were mapped onto their Dali surface$_{DOM}$s structural comparison dendrogram. Two additional α-helices of *A. irregularis* (dashed rectangle), likely correspond to the S-layer extracellular protrusions (**b**). Heatmap matrices for RMSD values across subdomains D1, D2, ID and anchor$_{DOM}$ show the highest structural divergence within the D2 (average RMSD 3.7 Å). Conversely, the anchor$_{DOM}$ CWB2 fold is highly conserved (average RMSD 1.2 Å).

d, Broth-based killing assay in the presence of immune effectors (Lysozyme, orange; LL-37, blue) relative to the non-treated control (black). Resistance to growth impairment and cell death was observed only for two pathogenic S-layer-positive species (*C. difficile* and *P. anaerobius*).

e, f, *C. difficile* CRISPRi SlpA knock-down (**e**) resulted in the sensitivity to both antimicrobial agents in broth-based killing assay (as in **d**), as observed in the S-layer-negative strain FM2.5 (**d**), consistent with S-layer protecting against the host's innate immunity. The S-layer knock-down in *R. lituseburensis* (**f**) did not change its natural sensitivity (**d**) to lysozyme and LL-37, yet it impacted sporulation in plate-based assay for both species.

The gene knock-down was induced with xylose (+) relative to negative plasmid controls (control). CRISPRi SLLP gene knock-downs were validated with RT-qPCR (left panel), SDS-polyacrylamide gels (middle panel) and electron microscopy (right panel). All experiments were performed in biological triplicate, with a standard error of the mean and statistical significance (****, P<0.0001) shown.

340 **S-layer anchoring domains belong to a broader system of swappable cell attachment modules**

341 The observed diversity of *Peptostreptococcaceae* S-layer surface-forming domains results in multiple S-
342 layer architectures (Fig. 2b-d) with distinct biochemical properties (Fig. 3a) and variable physiological
343 functions (Fig. 3d-f). In addition to surface$_{DOM}$s, SLLPs also contain various anchoring domains that attach
344 the S-layer to the cell. Our analysis across the Tree of Life demonstrated that anchor$_{DOM}$s are not unique to
345 SLLPs, can be found in multiple other proteins encoded in bacterial genomes, and that alternative
346 anchor$_{DOM}$s can be swapped during cell-anchored protein evolution (Supplementary Fig. 3). In pathogenic
347 Firmicutes, S-layers utilise at least two different cell-wall anchoring domains, CWB2 in *C. difficile* and SLH in
348 *B. anthracis*, binding to different secondary cell wall polymers (SCWPs)[6,46,47]. The capacity to utilise
349 swappable cell wall anchors could enable attachment of S-layers to different cell wall components, enabling
350 lateral transfer of S-layers between bacteria utilising different cell wall architectures and facilitating co-
351 evolution of different cell envelope layers. Therefore, we investigated if the *Peptostreptococcaceae* S-layers
352 utilise variable cell anchoring mechanisms and if their anchor$_{DOM}$s evolve as part of the wider cell surface
353 proteome.

354 The four types of SLLPs we identified in *Peptostreptococcaceae* (Fig. 2b,c) utilise four distinct cell-
355 anchoring domains (Fig. 4a and Supplementary Fig. 5a,b). All the SLLPs with α/β-sandwich surface$_{DOM}$s,
356 including *C. difficile* SlpA, contained CWB2 anchor$_{DOM}$, while the single β-barrel SLLP contained the SLH
357 anchor$_{DOM}$ (Fig. 3a), also found in β-sandwich surface$_{DOM}$-containing S-layers of *Bacillaceae,* including
358 pathogenic *B. anthracis* (Fig. 1b). However, *Peptostreptococcaceae* S-layers with β-sandwich surface$_{DOM}$s
359 contained one of two distinct anchors: SH3 or Cu-oxidase-like N-terminal domain[48-50] (Cu-OX, Fig. 3a). This
360 demonstrated that, within the single bacterial family of *Peptostreptococcaceae,* a variety of anchor$_{DOM}$s are
361 utilised by different S-layer lattice-forming surface$_{DOM}$s, reminiscent of the patterns observed across the
362 Tree of Life (Fig. 1 and Supplementary Fig. 1a).

363 The utilisation of the four distinct cell wall anchoring domains in *Peptostreptococcaceae* S-layers,
364 including those sharing similar structural folds (Fig. 4a), suggested that these domains may represent
365 swappable modules. Our analyses revealed multiple independent swaps, gains, and/or losses of the entire
366 SLH/CWB2 cell wall anchoring systems within *Peptostreptococcaceae* (Fig. 4c), consistent with the patterns
367 we observed within the broader sampling of Firmicutes (Supplementary Fig. 5). These changes usually
368 affect multiple proteins, as SLH and CWB2 anchor$_{DOM}$ are found on average in 12.5 and 9.8 proteins per
369 anchor$_{DOM}$-containing Firmicute genome, respectively (Supplementary Fig. 5b). Based on protein similarity
370 network analyses, the identified anchoring domain swaps were often observed with retention of the
371 attached functional domain homology (Fig. 4e,f, and Supplementary Fig. 3a-d), suggesting that anchoring
372 system swap does not necessarily reflect changes in the surface proteome functionality. Although at least
373 some of the inferred swaps involved rearrangement of entire genomic regions encoding multiple surface
374 proteins (Supplementary Fig. 6), individual surface proteins are often scattered across the chromosome
375 (Supplementary Fig. 5c). Phylogenetic, structural, and sequence homology analyses are consistent with
376 multiple independent anchor replacements among closely related proteins (Supplementary Fig. 6)
377 indicating that, in addition to the swaps of gene cassettes encoding multiple surface proteins, surface
378 proteomes evolve by replacing anchor$_{DOM}$s in individual proteins. Multiple anchor$_{DOM}$ swaps were also found
379 in species with no evidence of S-layer (Supplementary Fig. 6). Therefore, anchoring domain variation is not
380 specific to SLLPs and S-layers, but it rather reflects a more general trend affecting entire cell wall-anchored
381 protein repertoires.

382 Neither CWB2 nor SLH systems were found in any of the analysed *Terrisporobacter/Intestinibacter*
383 genomes. Instead, they contained an expanded repertoire of SH3 domain-containing proteins compared to
384 other *Peptostreptococcaceae* (Fig. 4c), and other Firmicutes (Supplementary Fig. 5a), suggesting CWB2/SLH-
385 loss and SH3-expansion represents their lineage-specific adaptation. Of the two species we studied

386 experimentally, *Terrisporobacter glycolicus* and *Intestinibacter bartlettii,* only *I. bartlettii* exhibited evidence
387 of S-layer presence (Fig. 2d, and Supplementary Fig. 4) and its SLLP contained the SH3 anchoring domain
388 (Fig. 4a). Although the presence of SH3 domains is conserved across all *Terrisporobacter/Intestinibacter*, the
389 SH3-anchored β-sandwich SLLP homologues were specific to a distinct group of *I. bartlettii* isolates (Fig. 2e)
390 suggesting that this novel S-layer has evolved subsequently to the major changes to the cell envelope
391 architecture in the *Terrisporobacter/Intestinibacter* ancestor.
392 For the anchoring domains to perform their function in the attachment of S-layers and other surface
393 proteins to the cell, specific secondary cell wall polymers (SCWP) need to be synthesised and incorporated
394 into the cell wall[51]. In protein similarity network analyses of all proteins encoded in the 117 *Romboutsia*-
395 and *Paeniclostridium*-genomes which have undergone multiple anchoring system swaps/gains/losses, we
396 identified only a single protein without anchor enriched in CWB2-positive or SLH-positive species (Fig. 4d).
397 The CWB2 system-associated protein was a homologue of a polysaccharide synthase (CD2777) involved in
398 the synthesis of SCWP for CWB2 anchoring in *C. difficile*[47]*.* The SLH system-associated protein was a
399 homologue of TagA, involved in SLH-specific SCWP biosynthesis in Firmicutes, including *B. anthracis*[52] and
400 *Paenibacillus alvei*[51]. Therefore, enzymes required for the biosynthesis of SCWP anchors co-distribute with
401 their cognate anchoring domains. The presence of these secondary cell wall polymer biosynthetic enzymes
402 was also detected in organisms where both SLH and CWB2 proteins were encoded in the same genome
403 (Fig. 4d). However, in those cases, the inferred SCWP type reflected that of the more abundant anchoring
404 domain, while the second domain was found in only a few proteins (Fig. 4d), suggesting residual presence
405 from incomplete anchoring domain swap, an indication that this can be a stepwise process.
406 For S-layers to assemble on the cell surface, in addition to the presence of specific cell surface anchors, a
407 protein secretion system is needed to transport SLLP monomers from the cytosol across the cell
408 membrane. An SLLP-specific SecA2 secretion system is utilised in the assembly of α/β-sandwich *C. difficile*
409 S-layer[53]. However, our analyses revealed that neither of the *Peptostreptococcaceae* genomes encoding
410 SH3- or CuOX-anchored β-sandwich SLLPs encoded any identifiable orthologues of SecA2. This suggests
411 that, in *Peptostreptococcaceae*, SecA2 is associated predominantly with α/β-sandwich CWB2-anchored S-
412 layers and some homologues of β-barrel SLH-anchored SLLPs*,* and that β-sandwich SH3- and CuOX-
413 anchored SLLPs can utilise a more broadly distributed canonical SecA or a not yet identified alternative
414 secretion system. Therefore, S-layers within a single bacterial family are not only assembled from different
415 building blocks but also utilise different mechanisms for their assembly and the evolution of S-layer needs
416 to be coordinated with associated cell wall modification and protein secretion machineries.
417 Together, our analyses revealed that S-layer anchor$_{DOM}$s are part of broader cell wall anchoring systems
418 shaped by modular evolution. Entire systems and anchor$_{DOM}$ of individual proteins were swapped, acquired,
419 or lost on multiple independent occasions, resulting in a mosaic of different anchors and functional
420 domains often found in the genomes of closely related species. Homologues of enzymes known to be
421 involved in the biosynthesis of CWB2- and SLH-specific secondary cell wall polymer anchors are commonly
422 swapped alongside the SLH/CWB2 anchoring systems. This indicates that, at least in
423 *Peptostreptococcaceae*, evolution of S-layer anchoring mechanisms is coordinated with that of cell wall
424 chemistry. The capacity for utilisation of pre-existing swappable anchor$_{DOM}$ modules can also provide an
425 explanation for the diversity of anchors observed in S-layers sampled from across the Tree of Life, including
426 those utilising the same surface$_{DOM}$-fold (Fig. 1b).

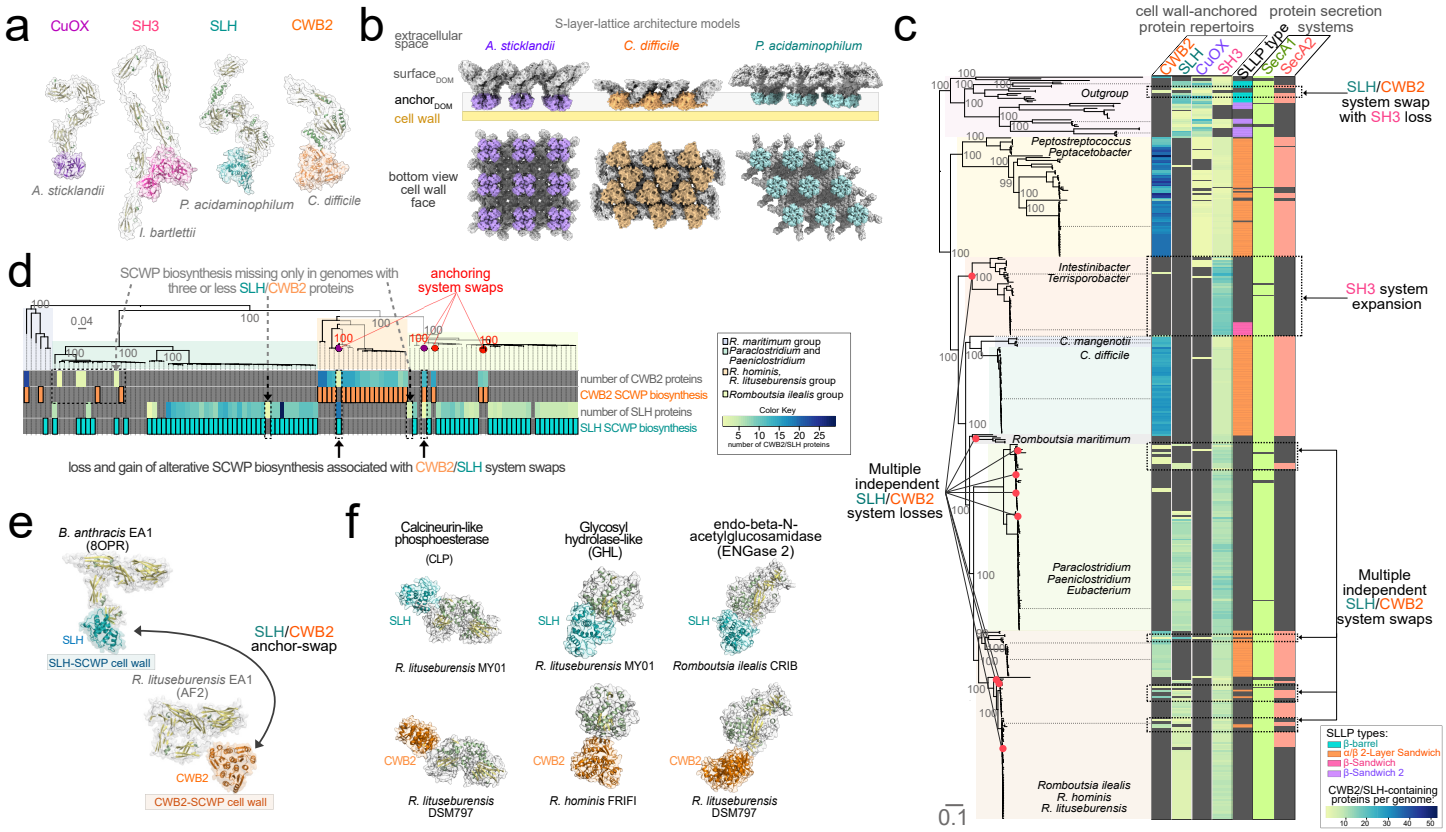

Fig. 4: Modular architecture and evolution of the S-layer and other cell envelope systems.

a, The four experimentally identified *Peptostreptococcaceae* S-layer surface forming fold-types (Fig. 2b) are anchored to the cell using four distinct anchoring domains (orange, CWB2; cyan, SLH; magenta, CuOX; pink, SH3).
b, Surface representations of S-layer lattice models (as in Fig. 2c) with anchor$_{DOM}$s coloured as in **a**.
c, Species tree mapping of the total number (yellow-blue heatmap) of cell wall anchored proteins identified in genomes of 453 *Peptostreptococcaceae*. SH3 anchoring is more broadly conserved, while relatively less common CWB2 and SLH systems are interchangeable and were lost (red circles) or swapped (dashed rectangles) multiple independent times. The more general SecA1 (green) secretion system is present in all lineages, whereas SecA2 (red) with an SLLP-specific role in in *C. difficile*, is found primarily with CWB2-anchored S-layers (SLLP type).
d, Numbers of CWB2 and SLH anchor-containing proteins (yellow-blue heatmaps), and the presence of the secondary cell wall polymer (SCWP) biosynthesis enzymes specific to SLH (cyan) or CWB2 (orange) were mapped onto the *Romboutsia/Paraclostridium/Paeniclostridium* species tree. Enzymes involved in the biosynthesis of SLH- and CWB2-specific SCWPs co-distribute with their respective anchor$_{DOM}$s and are missing only in the genomes containing a low number of anchored proteins (dashed arrows). Multiple independent SLH/CWB2 system swaps (red arrows) and possible complete (red circles) and incomplete (purple circles) system losses (dashed arrows) are apparent.
e, AlphaFold2 structure predictions of the *R. lituseburensis* homologue of *B. anthracis* SLH-anchored S-layer protein EA1 contains a CWB2 anchor$_{DOM}$ consistent with anchoring domain swap and lateral gene transfer between *Peptostreptococcaceae* and *Bacillaceae*.
f, AlphaFold2 structure prediction of individual non-SLLP cell wall-anchored proteins such calcineurin-like phosphoesterase (CLP) and peptidoglycan hydrolases (endo-β-*N*-acetylglucosamidase, ENGase; and glycosyl hydrolases, GHL), with swapped SLH/CWB2 anchor$_{DOM}$s.

450 **Extreme divergence, LGT, and recombination shape diversity of homologous S-layers**

451 The observed diversity of S-layer cell anchoring mechanisms and variation in the number and type of
452 individual subdomains within a single surface$_{DOM}$ structural fold can be explained by modular evolution.
453 However, at this point, it is unclear what mechanisms shape the structural and functional diversity of
454 homologous S-layers utilising the same anchor$_{DOM}$- and surface$_{DOM}$-type. The extremely high level of
455 sequence diversity of homologous S-layer surface-forming domains such as those of α/β-sandwich fold-
456 type (Fig. 3b), complicates evolutionary analyses across the broader taxonomic scales. To circumvent this
457 difficulty, and to better understand the mechanisms of the SLLP-surface$_{DOM}$ evolution, we leveraged the
458 established model of *C. difficile,* an important human pathogen where the S-layer provides a selective
459 advantage within the host during infection[30].
460 Our bioinformatic analyses of available *C. difficile* SLLP (*Cd*SlpA) sequences (Fig. 5a) using Duplication-
461 Transfer-Loss reconciliation identified signatures of at least 15 unambiguous recent LGT events (Fig. 5b),
462 consistent with an active process, likely involved in immune evasion[54,55]. Of all 18,035 non-atypical *C.*
463 *difficile* genome assemblies available in NCBI, 99.7% contained *Cd*SlpA coding genes. This suggests selection
464 pressure for retention of a single copy of an *slpA* gene, in contrast to *Bacillus anthracis* with its two SLLPs[6],
465 for example. Recombination analyses provided strong evidence for at least a single recombination event
466 between parental SlpA$_{SLCT4}$ and SlpA$_{SLCT6}$, resulting in a recombinant divergent SlpA$_{SLCT4}$ (Fig. 5c). Thus,
467 recombination not only plays a part in exchanging different SLCT cassettes between different genome types
468 but can also contribute to generating the diversity observed within a single species, such as among *Cd*SlpAs.
469 To investigate the level of SlpA divergence relative to other protein families encoded within *C. difficile*
470 genomes, we identified 2,950 orthogroups containing a single copy orthologue in at least 90% (SCO90) of
471 265 non-redundant representative *C. difficile* genome assemblies. Within this dataset, SlpA was the least
472 conserved of all identified SCO90s (48.1% average sequence identity, Fig. 5d). Thus, *C. difficile* SlpA
473 represents the most rapidly evolving of all identifiable single-copy orthologue proteins found across all
474 analysed *C. difficile* genomes. It is therefore likely that the high level of SlpA divergence reflects an
475 adaptation of *C. difficile* to selective pressures inside the mammalian gut[30]. This hypothesis is further
476 supported by the observation that another major cell surface antigen, S-layer-associated protein Cwp66,
477 previously implicated in adhesion[56,57] and survival in the human gut[57], is the second least conserved SCO90
478 (64.3 % average sequence identity, Fig. 5d). The fourth least conserved SCO90 is a homologue of the cell
479 wall teichoic acid glycosylation protein GtrA[7]. *Listeria monocytogenes* homologue (GtcA) plays a role in cell
480 wall glycosylation[58,59], likely as a flippase[60], and the *B. anthracis* homologue is involved in the biosynthesis
481 of SCWPs essential for S-layer surface attachment[59]. The sixth SCO90 protein is SecA2, the secondary
482 secretion system ATPase required for specific translocation of *Cd*SlpA[53], which co-distributes with CWB2-
483 anchored SLLPs in *Peptostreptococcaceae* (Fig. 4c).
484 It is not clear if both LGT and high levels of sequence divergence can also contribute to the diversity of
485 SLLPs in other bacterial lineages. Our bioinformatic analysis of *Bacillus* genomes revealed patterns
486 consistent with both mechanisms in closely related *B. cereus/anthracis* isolates (Supplementary Fig. 7a-d),
487 suggesting that these mechanisms are not limited to *C. difficile* SlpA and may be more common in other S-
488 layer containing species. Interestingly, SAP/EA1 homologues were identified within the *B. cereus/anthracis*-
489 group and not in their closest relatives from SAP/EA1-negative *B. cereus*-lineage (Supplementary Fig. 7d),
490 suggesting differences in their cell envelope architectures. It is important to note that extreme levels of
491 sequence divergence such as those observed in *Cd*SlpA (Fig. 5d) could mask more ancient LGTs, leading to
492 underestimation of S-layer protein exchange between more distantly related taxa. In our analysis, we
493 identified a single clear example of a β-sandwich EA1 surface-forming domain homologue being shared *via*
494 LGT between *Bacillus* and *Peptostreptococcaceae* (Supplementary Fig. 7f,g). As *Peptostreptococcaceae* EA1
495 contained CWB2 and not SLH anchoring domains, it provided the first clear example of anchoring domain

496 swap in SLLP homologues. We were able to experimentally verify the presence of an EA1 homologue in *C.*
497 *mangenotii* LM2 surface protein preparations, where it constituted a minor S-layer component, alongside
498 its dominant CWB2-α/β-sandwich SLLP (Supplementary Fig. 7h,i).
499 Together, our data demonstrate that the observed diversity of S-layers utilising the same anchor$_{DOM}$-
500 and surface$_{DOM}$-types can be generated through extreme sequence divergence, homologous recombination,
501 and LGT between close and more distant relatives. Our analyses also revealed patterns of accelerated
502 evolution in S-layer-associated proteins and proteins involved in S-layer assembly.
503

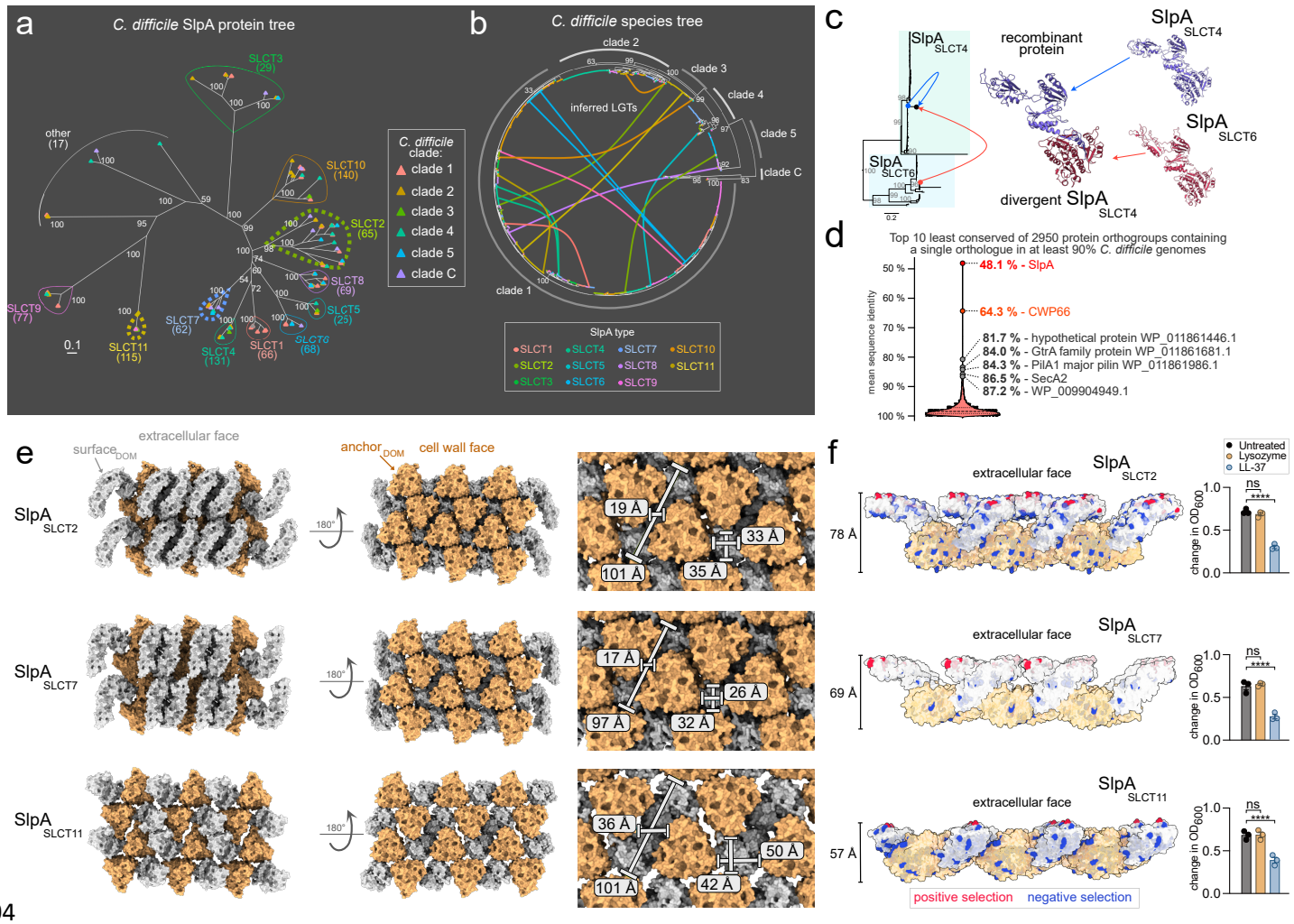

Fig. 5: Evolutionary mechanisms shaping the diversity of S-layers across clinical isolates of a pathogen.

a, Phylogeny of SlpA SLLP colour-coded based on the clade-membership of its source *C. difficile* genome assembly (tree tip colours). Diverse SlpA sequences belong to multiple distinct S-layer cassette types ($SlpA_{SLCT1-11}$).
b, The 14 unambiguous LGT events (coloured lines), identified using Duplication-Transfer-Loss reconciliation analysis, were mapped onto the core genome species tree of 866 *C. difficile* isolates including all six clades (clades 1-5 and C) and coloured based on the SlpA-type membership.
c, An unambiguous recombination event (red and blue arrows) between SLCT4 and SLCT6 identified in recombination detection program RDP4 analyses was mapped onto a $SlpA_{SLCT4}$ and $SlpA_{SLCT6}$ protein phylogeny (left) and onto AlphaFold2 structure models (right) of the two inferred parent proteins (blue or red) and the recombinant sequence (red/blue).
d, Violin plot represents the distribution of mean pairwise sequence identity within 2,950 orthogroups containing a single copy orthologue (SCO) in at least 90 % (240 of 267) analysed representative non-redundant *C. difficile* genome assemblies. Scale bars represent the number of substitutions per site.
e, Representation of the SLLP-formed lattice derived from crystal packing of the three determined SlpA structures. The three SlpAs arrange in different lattice topologies, represented here at different tilt angles: extracellular (left) and cell wall-facing orientations (middle). Distances measured between neighbouring $anchor_{DOM}$s (right, white lines) show that $anchor_{DOM}$s are packed more closely in $SlpA_{SLCT2}$ and $SlpA_{SLCT7}$ lattices and are more separated in the $SlpA_{SLCT11}$ lattice.
f, Mapping of negatively (blue) and positively (red) selected residues onto a side view of the S-layer lattice shown in surface representation for each $SlpA_{SLCT}$ ($surface_{DOM}$ – white, $anchor_{DOM}$ – orange, left). The three lattices differ in thickness reflecting the differences in structural organisation of the SlpA monomers. Despite their extreme sequence divergence (**d**) and structural differences (**e**), all analysed S-layers display conserved patterns of differentially distributed sites under adaptive evolution. While negatively selected residues are found in $surface_{DOM}$ and $anchor_{DOM}$ and on all SlpA surfaces, the sites under positive selection are exclusively found on the environment-exposed extracellular surface of the $surface_{DOM}$. Broth-based killing assay (right) of different $SlpA_{SLCT}$-carrying strains in the presence of lysozyme (orange) and LL-37 (blue) relative to the non-treated control (black) confirms conserved resistance to these innate immune effectors.

530 **Extensive S-layer rearrangements and positive selection across clinical isolates of a pathogen**

531 Over the broad phylogenetic scales, S-layer building blocks from the same structural family can
532 assemble into diverse cell surface lattice architectures (Fig. 1b). To better understand the structural effects
533 of the extreme SLLP sequence variation at the narrowest phylogenetic scales (Fig. 5d), we determined the
534 crystal structure of *C. difficile* $SlpA_{SLCT11}$ and $SlpA_{SLCT2}$, representing distinct phylogenetic SlpA-types from
535 multiple clinical isolates of a single species (Fig. 5a), and compared them to that of the previously studied
536 $SlpA_{SLCT7}$[40].
537 All tested SlpA homologues are proteolytically processed following secretion, resulting in two separate
538 peptides: low molecular weight surface layer protein (LMW SLP or SLP_L), containing the S-layer surface-
539 forming domain, and high molecular weight (HMW SLP or SLP_H), containing the CWB2 anchoring domain.
540 Regardless of SLCT membership, the overall topology of *Cd*SlpA structures (Supplementary Fig. 8a) is
541 maintained and consists of a triangular prism arrangement of the CWB2 anchoring domain, an interacting
542 domain (ID) linking $anchor_{DOM}$ and $surface_{DOM}$, followed by one or two domains (D1-only, or D1/D2) of
543 $surface_{DOM}$, protruding the plane of the array[40]. The structure of the $anchor_{DOM}$ is highly conserved among
544 all determined *Cd*SlpA structures, with core root mean square deviation (RMSD) of ~1 Å (Supplementary
545 Fig. 8a). This is consistent with the higher degree of $anchor_{DOM}$ sequence conservation relative to that of
546 $surface_{DOM}$ in SLLPs from other organisms[2] and apparent in the analyses across the Tree of Life presented
547 here (Supplementary Fig. 1-3). Within the more conserved *C. difficile* CWB2 $anchor_{DOM}$, the highest degree
548 of structure variability was observed in loops projecting from the $anchor_{DOM}$ plane towards the extracellular
549 face of the S-layer in $SlpA_{SLCT11}$ or the $surface_{DOM}$/$anchor_{DOM}$ interspace in $SlpA_{SLCT2}$ and $SlpA_{SLCT7}$
550 (Supplementary Fig. 8a). Consistent with the more general pattern observed across the Bacterial Tree of
551 Life (Supplementary Fig. 1 and 2), the most prominent differences across the SlpA structures were observed
552 within their environment exposed S-layer surface-forming domain (Fig. 5e-f, Supplementary Fig. 8). Unlike
553 the $SlpA_{SLCT7}$ and $SlpA_{SLCT2}$, which share a 2-domain (D1/D2) arrangement, $SlpA_{SLCT11}$ lacks the entire D2
554 domain, which accounts for over 20% of total solvent accessible area of the S-layer unit, resulting in a
555 smoother (Fig. 5e) and thinner array (Fig. 5f). The stark contrast between the $surface_{DOM}$s of $SlpA_{SLCT2}$,
556 $SlpA_{SLCT7}$, and $SlpA_{SLCT11}$ provided evidence of how the entire fold that determines S-layer architecture can
557 be reshaped within the relatively short phylogenetic timescales of emergence of different isolates of the
558 same species. The *Cd*SlpA $surface_{DOM}$ amino acid sequence regions align poorly with each other, further
559 demonstrating their extreme sequence divergence (Fig. 3b and 5d) and providing a likely explanation for
560 difficulties in identifying homologous SLLPs from more distantly related species. Interestingly, $surface_{DOM}$-
561 $anchor_{DOM}$ interactions form the majority of the lattice interfaces in all studied *Cd*SlpAs (Supplementary Fig.
562 9), suggesting that the main role of the $surface_{DOM}$ may be as a crystallisation enhancer module which
563 forms an S-layer lattice only in conjunction with the more common CWB2 anchoring domain. This mode of
564 S-layer lattice assembly, involving both the anchoring and surface domains, could explain the shorter
565 $surface_{DOM}$ in *Cd*SlpA relative to those identified in other *Peptostreptococcaceae* S-layer-types (Fig. 2c) and
566 possibly the lack of any identifiable $anchor_{DOM}$ swaps involving α/β-sandwich $surface_{DOM}$.
567 Despite the observed differences in their lattice architecture, all the studied *C. difficile* isolates utilising
568 different S-layer variants displayed conserved resistance to natural antimicrobial agents LL-37 and lysozyme
569 produced by the human immune system (Fig. 5f). This demonstrates that considerable structural changes
570 of lattice can evolve without compromising the physiological functions of the S-layer. It also raises
571 important questions about the reasons behind the apparent sensitivity to the antimicrobials observed in
572 other *Peptostreptococcaceae* utilising homologous S-layers (Fig 3d-f).
573 Finally, it remained unclear if the observed structural variation in *Cd*SlpAs is mainly shaped by genetic
574 drift and recombination or if any of the sites are under positive selection. To test this, we analysed non-
575 redundant nucleotide sequences encoding SlpA identified in all 18,035 analysed *C. difficile* genomes. All

576 *Cd*SlpA types displayed patterns of positive selection and all identified sites under positive selection were
577 within the $surface_{DOM}$ and were surface exposed (Fig. 5f and Supplementary Fig. 10). This pattern was
578 contrasted by the negatively selected residues found within $surface_{DOM}$ and $anchor_{DOM}$, and at all lattice-
579 forming faces of the S-layer array (Fig. 5f). This is consistent with previous analyses indicating that surface
580 antigens of pathogenic bacteria and viruses, and especially their exposed epitopes, are often under positive
581 selection[61,62].
582 The positive selection patterns detected only at the S-layer face exposed to the extracellular
583 environment are consistent with its proposed role in phage and immune evasion. Extreme sequence and
584 structural variation can facilitate the pathogen's arms race with human adaptive immunity without
585 compromising its conserved role in protection against lysozyme and LL-37, two components of the innate
586 immune response. This model is consistent with the recently described capacity to generate monoclonal
587 antibodies selective towards different SlpA variants[63].
588 In summary, the evolution of a single S-layer-forming protein can have a dramatic impact on the
589 architecture of the bacterial cell surface at the narrowest phylogenetic scales which, in pathogens such as
590 *C. difficile*, contributes to the protection against the human immune system.
591
592
593 **Discussion**

594 As opposed to the archetypical elements of bacterial cell envelope architecture - lipid cell membrane
595 and peptidoglycan cell wall - the main building block of a specific S-layer lattice is commonly encoded by a
596 single gene and therefore can provide a substantial level of evolutionary plasticity to the microbial cell
597 surface. This plasticity is of particular significance in bacterial pathogens such as *B. anthracis* or *C. difficile*,
598 where the S-layer plays a role in disease[10,16,30,32].
599 Our bioinformatic analyses demonstrate that the observed diversity of *C. difficile* S-layers is shaped by
600 lateral gene transfer, homologous recombination, adaptive evolution, and extreme sequence divergence.
601 Similar mechanisms can also generate the diversity of SLLP folds, and S-layer crystal lattice architectures
602 across the bacterial Tree of Life (Fig. 1 and references[34,35]), as we observed in wider sampling of
603 *Peptostreptococcaceae* and *Bacillus*. Comparative analyses of the two crystal structures of *Cd*SlpA
604 homologues experimentally determined in this study show that extreme $surface_{DOM}$ sequence divergence
605 can result in significant changes to the S-layer architecture including lattice thickness, $anchor_{DOM}$ packing,
606 and extensive rearrangements of its environment-exposed face. The observed architectural variation of *C.*
607 *difficile* S-layers does not perturb its physiological functions in spore formation and resistance against
608 effectors of the human innate immunity such as lysozyme and LL-37.
609 In the case of pathogenic bacteria, the extensive plasticity of their environment exposed surfaces may
610 reflect an arms race with external threats to the pathogen's survival such as bacteriophages[18,19] and human
611 adaptive immunity[64]. Since the S-layer is an assembly of multiple copies of *Cd*SlpA, a single mutation of an
612 environment-exposed positively selected residue can correspond to changes in up to an estimated 0.1 - 1%
613 of the entire solvent accessible outer surface of *C. difficile* SlpA crystal lattice (Supplementary Fig. 9d). The
614 observed structural plasticity, particularly at the more externally exposed domains (e.g. sites under positive
615 selection), shows how a single mutation can affect a significant proportion of the host-pathogen interface.
616 This includes altering potential surface epitopes, and therefore contribute to immune evasion, as *Cd*SlpA
617 has been directly implicated in interactions with the host and eliciting immune response[64]. Understanding
618 the evolutionary mechanisms shaping *Cd*SlpA sequence diversity and the resulting structural plasticity of
619 the S-layer lattice is important if we are to therapeutically target the S-layer[10,19,33]. To avoid mutation
620 and/or recombination-driven resistance mechanisms arising against any drugs designed to disrupt the S-
621 layer, we need to consider the short-time scale evolution processes identified here. Furthermore, the

622 structural landscape of S-layers needs to be considered as recent immunological studies identified that
623 high-affinity anti-LMW monoclonal antibodies binding distinct epitopes of a single SlpA have different
624 effects on *C. difficile*'s growth, toxin secretion, and biofilm formation[33].
625 The high level of S-layer diversity is not limited to clinical isolates of *C. difficile* and was observed in
626 different strains of *Bacillus stearothermophilus*[65] and on a broader taxonomic scale in divergent SLLP
627 sequences found across multiple species of *Bacillus* and *Lactobacillus*[66]. At an even broader phylogenetic
628 scale, among fourteen tested representatives of a single family of Bacteria, *Peptostreptococcaceae*, we
629 detected and experimentally verified nine new SLLPs with one novel surface$_{DOM}$ fold-type (β-barrel, CATH
630 2.4) and two anchoring domains (Cu-OX and SH3), not observed in any of the previously characterised
631 SLLPs. Our results demonstrate that the diversity of S-layers in *Peptostreptococaceae* is shaped by modular
632 evolution[67,68]. This is apparent when the mosaic of different genes, protein folds, secretion systems, and
633 surface anchoring systems, consisting of different anchoring domains and different chemical pathways
634 generating different secondary cell wall polymer modules, are mapped onto their species tree (Fig. 6a).
635 However, this modularity is not restricted to a single mode of evolution but rather encompasses multiple
636 'complexity' levels. At the genome level, entire multigene S-layer cassettes can be swapped *via* LGT and
637 recombination, and alternative multicomponent anchoring systems (e.g. CWB2 and SLH), including the
638 surface-anchor biosynthesis machinery and its cognate anchoring domains of multiple proteins, can be
639 replaced. At the individual protein level, alternative anchoring domain types can be swapped or gained,
640 enabling otherwise homologous proteins to utilise one of the multiple cell-anchoring systems that coexist
641 within a single bacterial genome (e.g. choline-binding and CWB2). Similar patterns were observed at the
642 subdomain level in lattice-forming surface$_{DOM}$s, as demonstrated by variation in the number of β-sandwich
643 (sub)domains within Ig-like arrays[69], the most common fold of known SLLP architectures, acquisition of a
644 novel α-helical surface subdomain in α/β-sandwich S-layer of *A. irregularis*, or as the capacity to lose the
645 entire D2 subdomain in the α/β-sandwich fold of *C. difficile* SlpA$_{SLCT11}$. In addition, the reorganisation of
646 subdomains may also play roles in the incorporation of divergent S-layer proteins into an existing lattice
647 such as that proposed for the minor S-layer component SlpX in *Lactobacillus*[70] or anchor-swapped EA1
648 homologue in *C. mangenotii.* At the cell biology level, the entire S-layer can be considered as a cell
649 envelope module which can be swapped, remodelled, or lost. This multilevel modular architecture provides
650 a flexible framework for various lineage- and species-specific cell surface adaptations (Fig. 6b). In
651 *Romboutsia*, CWB2 and SLH secondary cell wall polymers were swapped, gained, or lost multiple
652 independent times, often as part of major genome rearrangements leading to multiple inferred S-layer
653 losses. In *Clostridioides/Asaccharospora*, a single core CWB2-anchored α/β-sandwich S-layer scaffold is
654 retained, and the variation of the cell envelope is mostly due to the modification to the pre-existing lattice
655 or addition of S-layer associated proteins. In *Terrisporobacter/Intestinibacter*, major ancestral changes to
656 the cell wall chemistry have likely caused a loss or swap of the ancestral S-layer and driven the emergence
657 of a novel SH3-anchored β-sandwich SLLP.

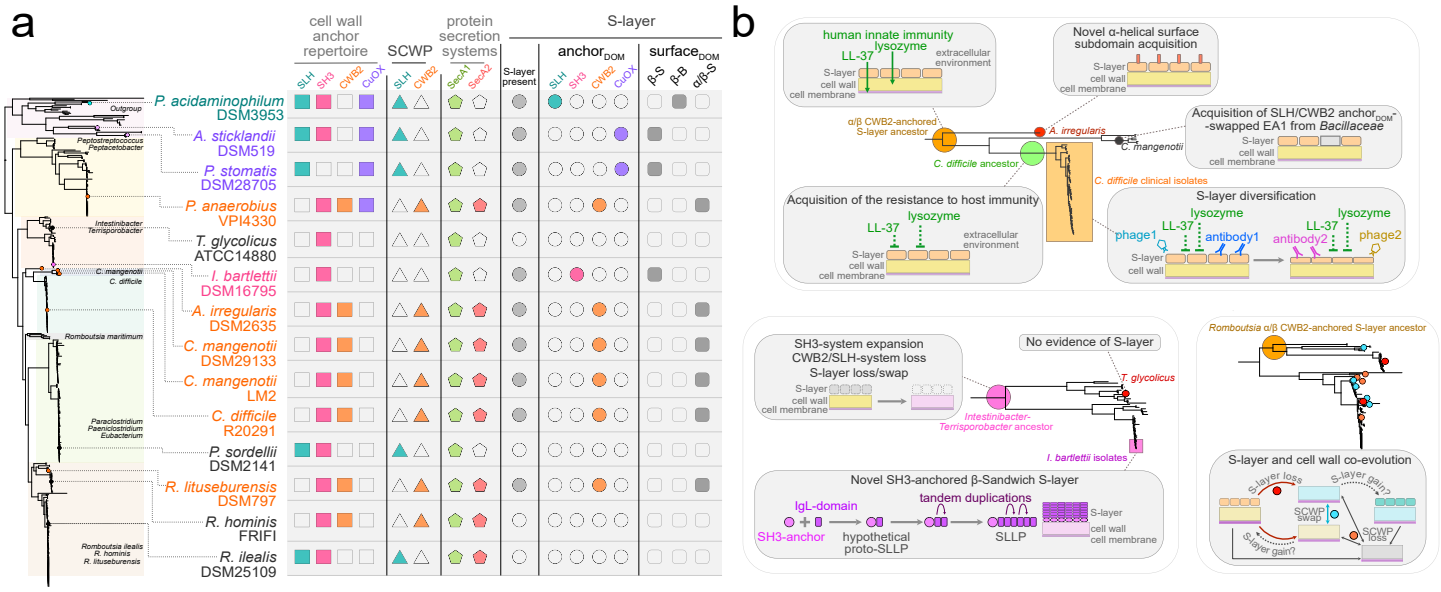

Fig. 6: Multiple evolutionary pathways to S-layer exoskeleton gain, loss and diversification in a single family of Bacteria.

a, An array representing presence (solid shapes) or absence (empty shapes) of cell wall anchoring domains found in at least a single protein per genome (rectangles), enzymes involved in biosynthesis of SLH- or CWB2-specific SCWP cell anchors (triangle), SecA protein secretion systems, including SecA2 involved in S-layer assembly in *C. difficile* (pentagons), experimentally verified presence of S-layer (grey circles), and anchor$_{DOM}$-types (coloured circles) and surface$_{DOM}$ folds (rounded rectangles) of experimentally identified SLLPs mapped onto a species tree of *Peptostreptococcaceae*. The scattered distribution of a mosaic of different cell envelope components including systems involved in S-layer assembly and anchoring to the cell is consistent with their diversity being shaped by modular evolution.

b, Three distinct models of S-layer evolution based on the results presented in this study demonstrate that the observed diversity of S-layers is generated by multiple mechanisms underlying the extreme evolutionary plasticity of this major component of bacterial cell architecture: (i) structural and functional diversification of the conserved α/β-sandwich CWB2-anchored S-layers in *Clostridioides-Asaccharospora* lineage (top panel); (ii) recent strain-specific emergence of a novel SH3-anchored β-sandwich S-layer protein in *Intestinibacter bartlettii* (bottom left panel) and (iii) evidence of a specific loss of S-layer within *Romboutsia*, a lineage displaying a high frequency of cell envelope remodelling (bottom right panel).

674 Modular domain organisation was recently observed in the structure of the proposed S-layer-
675 associated protein, extracellular Fe(III) reductase from the Gram-positive bacterium *Carboxydothermus*
676 *ferrireducens*, containing a domain homologous to *B. anthracis* SLH[71]. This suggests that modular evolution
677 can play a role in additional functions provided by S-layer-associated proteins, in this case in the Fe(III)
678 respiratory process, as well as in shaping S-layer architecture. Furthermore, modular architecture was
679 observed in many cell wall binding proteins[72], including lytic enzymes containing choline-binding modules[73],
680 consistent with our detection of cell wall homeostasis proteins being subject to frequent cell wall binding
681 domain swaps. Therefore, the observed evolutionary mechanisms shaping the S-layer likely represent a
682 more general phenomenon of bacterial cell envelope modularity that also includes cell wall and cell
683 membranes. Although the gene clusters encoding proteins involved in peptidoglycan biosynthesis are
684 highly conserved across Bacteria, rare insertions of outlier genes were observed in Firmicutes and
685 Candidate Phyla Radiation (CPR)[74] and our results demonstrate that secondary cell wall polymer
686 biosynthesis machinery was changed on multiple independent occasions, at least in Firmicutes. It was also
687 hypothesized that another cell envelope module, the outer membrane, was lost in Gram-negative (diderm)
688 bacteria multiple times, resulting in the independent emergence of Gram-positive (monoderm)
689 lineages[42,75,76]. However, as opposed to more complex cell membranes and cell walls, the main S-layer
690 building block SLLP is encoded by a single gene and therefore it can provide an extreme level of
691 evolutionary plasticity to microbial cell surfaces.
692 The diversification of S-layer cell surfaces can result not only in diverse architectures, symmetry types,
693 and biochemical properties but also in variable physiological functions. Probably the most striking example
694 is that of *Peptostreptococcaceae* S-layer utilising the same α/β-sandwich surface$_{DOM}$-fold and the same
695 CWB2 anchor$_{DOM}$ where only human gut pathogens *P. anaerobius* and *C. difficile* isolates (with their diverse
696 S-layer variants) are resistant to natural antimicrobials lysozyme and LL-37. In contrast, resistance to these
697 innate immunity effectors was not observed in α/β-sandwich surface$_{DOM}$-containing species unlikely to
698 cause human disease including *A. irregularis* DSM2635 and *R. lituseburensis* DSM797 (isolated from mud),
699 and *C. mangenotii* DSM29133 (isolated from intestine of a healthy mouse[77]). These results indicate that the
700 presence of divergent versions of the same type of S-layer can contribute to the protection against the
701 innate immunity effectors in pathogens but not in non-pathogenic species of the studied
702 *Peptostreptococcaceae*. Therefore, S-layer evolution can play a role in adaptation to the pathogenic
703 lifestyle. Furthermore, recent data suggests that distinct S-layers of *Lactobacillus* also play an important
704 role in probiotic bacteria associated with optimal health, where it can act as a mediator of the host
705 inflammatory response and prevent activation of pro-inflammatory signalling[26,28,29]. It is currently unclear if
706 these properties are conserved across diverse *Lactobacillus* S-layers[66,70].
707 Our data also suggests that the adoption of certain S-layer architectures can enforce some structural
708 and evolutionary constraints. While surface domains of some SLLPs appear to be found mostly with a
709 certain type of anchoring domain, others have undergone multiple independent anchor$_{DOM}$ swaps. For
710 example, in our analyses, all the α/β-sandwich SLLPs, including *C. difficile* SlpA, contained only CWB2
711 anchoring domains. This could be due to structural or functional constraints such as the importance of the
712 surface$_{DOM}$-anchor$_{DOM}$ interactions within their S-layer lattices, or it may merely represent a limitation of
713 currently available sampling. One potential explanation for these observations is that the emergence of S-
714 layers where surface$_{DOM}$-anchor$_{DOM}$ interactions form most of the lattice interfaces results in increased
715 structural dependence between the specific surface$_{DOM}$ and anchor$_{DOM}$, leading to the entrenchment of this
716 architecture. Such evolutionary ratchet mechanism[78,79] could lead to a decreased capacity to accept
717 alternative anchor$_{DOM}$s. No such restrictions were observed in the more common β-sandwich surface$_{DOM}$
718 homologues, which can function with multiple different anchor$_{DOM}$s and can be transferred *via* LGT
719 between more distantly related species.

720 Elucidating the origins and evolution of S-layers is complicated by the major gaps in our knowledge,
721 including (1) insufficient number of experimentally verified S-layer-positive and -negative species from
722 across the Tree of Life, (2) limited understanding of the conserved and lineage-specific S-layer functions
723 outside of a few model organisms, (3) unclear capacity for the *de novo* origins of S-layers from non-S-layer
724 proteins, (4) limited understanding of the physiological selection pressures and mechanisms underlying
725 possible S-layer gains and losses. The limited experimental evidence for S-layer loss includes the emergence
726 of *Cd*slpA mutants in response to the experimental exposure to specific antimicrobials (diffocins)[10], the
727 absence of S-layer in some strains of *Corynebacterium glutamicum*[36], and interspersed phylogenetic
728 distribution of S-layer positive *Peptostreptococcaceae* and their relatives with no evidence of S-layer
729 presented in this study. Interestingly, pathogenic *Corynebacterium* species lack identifiable homologous of
730 canonical S-layer protein PS2 found in the non-pathogenic representatives of the genus[80] and the presence
731 of the S-layer in the pathogenic species awaits verification. If absent, this would suggest that, at least in
732 some lineages, loss of the S-layer cell envelope module could also play a role in the emergence and
733 evolution of pathogens. More indirect evidence of the possibility of S-layer loss comes from the absence of
734 S-layers in model bacteria Gram-negative *Escherichia coli* and Gram-positive *Bacillus subtilis*[2]. It is also
735 unclear if any of the extant Eukaryotes contain cellular outer shell homologous to the prokaryotic S-layer.
736 Although the presence of S-layers *sensu stricto* to our knowledge has not been demonstrated in
737 Eukaryotes, paracrystalline glycoprotein cell wall layers can be found in some species such as green algae *C.*
738 *reinhardtii*[81,82]. The available evidence suggests that the S-layer is present in most *Archaea*[2], raising
739 questions about whether it was also present in the Last Common Eukaryotic ancestor and subsequently lost
740 or whether its loss was an important step in early Eukaryogenesis. Interestingly, the latest computational[34]
741 and experimental[83] approaches did not identify S-layers and SLLP candidates in Asgardarchaeota,
742 suggesting that the closest archaeal relatives of eukaryotes can be S-layer negative. However, this may not
743 be a conserved feature of all Asgardarchaeota[84] and further experimental studies are required to robustly
744 verify this hypothesis.
745 To our knowledge, there are no previously verified examples of S-layer gain by an S-layer-negative
746 ancestor in the natural environment. However, further studies in *Peptostreptococcaceae,* including
747 *Romboutsia,* or different species of *Corynebacterium*[36,80] could provide a much-needed model for studying
748 possible S-layer acquisitions. The presence of multiple bacterial species with extremely diverse S-layers and
749 species with no evidence of S-layer scattered across *Peptostreptococcaceae* phylogeny presented in this
750 study could lend credibility to the hypothesis that S-layers have evolved multiple independent times[2].
751 Perhaps the most striking example is that of *I. bartlettii* where a lineage-specific SLLP has emerged within a
752 subset of isolates of a single species, likely providing an excellent model for studying S-layer acquisitions. It
753 remains to be tested if and how a non-S-layer protein can evolve *de novo* to form a paracrystalline cellular
754 exoskeleton in nature. One captivating example of a possible evolutionary intermediate in the S-layer
755 evolution is that of a toga, a proteinaceous surface layer with membrane patches, found in *Thermotoga*
756 *maritima*[85]. The evolutionary model of S-layer origin from an outer membrane or a toga-like cell envelope
757 layer could be a reverse of that proposed for outer membrane biogenesis from the S-layer[86], i.e. a decrease
758 in the proportion of lipids and an increase in the proportion of proteins capable of self-assembly within the
759 outer cell envelope layer. This hypothesis, however, does not account for the possibility of the independent
760 S-layer emergence in monoderm Bacteria, where, in the absence of the outer lipid-containing layer, *de*
761 *novo* S-layer would have to evolve *via* a different mechanism. An alternative mechanism can be postulated
762 based on the inferred origin of the S-layer in *I. bartlettii*. Its SLLP, containing six nearly identical β-sandwich
763 Immunoglobulin-like surface$_{DOM}$ subdomains, has emerged *via* multiple tandem duplications of a single
764 ancestral subdomain of a type commonly found across bacteria. If this mechanism indeed plays a role in *de*
765 *novo* evolution of S-layers, it could provide one explanation for the common use of β-sandwich fold in S-
766 layers observed across the Tree of Life.

767 Our findings demonstrate that bacteria can rapidly gain, lose, replace, or modify the S-layer cell surface
768 exoskeleton through the utilisation of a multilevel modular architecture. Thanks to this scalable framework,
769 commonly composed of a single repeating SLLP building block, a single genetic change can significantly alter
770 the entire exoskeletal lattice. The segmented design of the SLLP adds another layer of flexibility, allowing it
771 to attach to various cell surface configurations through swappable cell anchoring domains. The ability to
772 remove, add, or remodel surface domain segments promotes rapid adaptation of the S-layer layout. This
773 versatile framework underlies the diversity of S-layer architectures observed across the prokaryotic Tree of
774 Life and offers opportunities for bioprospecting new natural lattice designs for nanotechnology, protein
775 engineering, and bio-product development. Our study uncovered the molecular, structural, and
776 evolutionary processes behind multiple independent origins, losses, and the remarkable structural and
777 functional divergence of this major component of bacterial cell biology. This extraordinary evolutionary
778 plasticity underscores the frontline role of these biological exoskeletons in adapting to extracellular
779 environments. In the case of pathogenic *Clostridioides difficile*, the S-layer has played a pivotal role in the
780 emergence of ancestral resistance to innate immunity effectors and its evolvability contributes to the
781 ongoing thug-of-war with adaptive immunity at the host-pathogen interface.
782
783
784
785 **Data availability**
786 The authors declare that the main data supporting the findings of this study are available within the article
787 and Supplementary information. Data sets generated and analysed in this study are provided as
788 Supplementary Data Files. The atomic coordinate files for X-ray structures have been deposited in the
789 Protein Data Bank (PDB) with the accession codes 9F8E, 9F8F.
790
791
792 **Acknowledgements**
793 This work was supported by a Medical Research Council grant [MR/V032151/1] awarded to P.S.S and A.B.-
794 S. and a Research Excellence Development Award [NU-018171] from the Faculty of Medical Sciences,
795 Newcastle University awarded to A.B.-S. We thank Dr Joseph Gray for expert mass spectrometry analysis
796 and protein identification, and Prof Craig Ellermeir for providing plasmids used in CRISPRi. Our thanks to
797 Tracey Davey and Ross Laws from Newcastle University Electron Microscopy Research Services, supported
798 by BB/R013942/1 for TEM sample processing; York Structural Biology Laboratory, University of York,
799 supported by the Wellcome Trust (206161/Z/17/Z) for supporting cryo-EM data collection; Newcastle
800 University Structural Biology Laboratory for facilitating crystallisation and data collection. We thank
801 Diamond Light Source for access to beamline I24 (mx24948-136), through the "Macromolecular
802 Crystallography at Newcastle, Durham, Lincoln and Durham" BAG). The contents of this work are solely the
803 responsibilities of the authors and do not reflect the official views of any of the funders, who had no role in
804 study design, data collection, analysis, decision to publish, or preparation of the manuscript.
805
806
807 **Author contributions**
808 K.M.S., A.B.-S. and P.S.S. are the corresponding authors. K.M.S. and A.B.-S conceptualised and planned the
809 study. K.M.S. wrote the manuscript with contributions from A.B.-S. and P.S.S. K.M.S. performed all
810 bioinformatic analyses. A.B.-S and K.M.S. performed experimental work. A.B.-S crystallized *Cd*SlpA. A.B.-S.

and P.S.S. determined and analysed protein structures. P.S.S. and A.B.-S. secured funding. K.M.S. and A.B.-S. managed the project.

Competing financial interests

The authors declare no competing interests.

References

1 Messner, P. & Sleytr, U. B. Crystalline bacterial cell-surface layers. *Adv Microb Physiol* **33**, 213-275, doi:10.1016/s0065-2911(08)60218-0 (1992).

2 Bharat, T. A. M., von Kugelgen, A. & Alva, V. Molecular Logic of Prokaryotic Surface Layer Structures. *Trends Microbiol* **29**, 405-415, doi:10.1016/j.tim.2020.09.009 (2021).

3 Pum, D., Toca-Herrera, J. L. & Sleytr, U. B. S-layer protein self-assembly. *Int J Mol Sci* **14**, 2484-2501, doi:10.3390/ijms14022484 (2013).

4 Messner, P., Schäffer, C., Egelseer, E.-M. & Sleytr, U. B. in *Prokaryotic Cell Wall Compounds: Structure and Biochemistry* (eds Helmut König, Harald Claus, & Ajit Varma) 53-109 (Springer Berlin Heidelberg, 2010).

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

6 Missiakas, D. & Schneewind, O. Assembly and Function of the Bacillus anthracis S-Layer. *Annu Rev Microbiol* **71**, 79-98, doi:10.1146/annurev-micro-090816-093512 (2017).

7 Bradshaw, W. J., Roberts, A. K., Shone, C. C. & Acharya, K. R. The structure of the S-layer of Clostridium difficile. *J Cell Commun Signal* **12**, 319-331, doi:10.1007/s12079-017-0429-z (2018).

8 Trautner, C. & Vermaas, W. F. The sll1951 gene encodes the surface layer protein of Synechocystis sp. strain PCC 6803. *J Bacteriol* **195**, 5370-5380, doi:10.1128/JB.00615-13 (2013).

9 Lasa, I., Caston, J. R., Fernandez-Herrero, L. A., de Pedro, M. A. & Berenguer, J. Insertional mutagenesis in the extreme thermophilic eubacteria Thermus thermophilus HB8. *Mol Microbiol* **6**, 1555-1564, doi:10.1111/j.1365-2958.1992.tb00877.x (1992).

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

11 Fioravanti, A., Mathelie-Guinlet, M., Dufrene, Y. F. & Remaut, H. The Bacillus anthracis S-layer is an exoskeleton-like structure that imparts mechanical and osmotic stabilization to the cell wall. *PNAS Nexus* **1**, pgac121, doi:10.1093/pnasnexus/pgac121 (2022).

12 Ali, S. *et al.* Slr4, a newly identified S-layer protein from marine Gammaproteobacteria, is a major biofilm matrix component. *Mol Microbiol* **114**, 979-990, doi:10.1111/mmi.14588 (2020).

13 Wong, L. L. *et al.* Surface-layer protein is a public-good matrix exopolymer for microbial community organisation in environmental anammox biofilms. *ISME J* **17**, 803-812, doi:10.1038/s41396-023-01388-y (2023).

14 Shimotahira, N. *et al.* The surface layer of Tannerella forsythia contributes to serum resistance and oral bacterial coaggregation. *Infect Immun* **81**, 1198-1206, doi:10.1128/IAI.00983-12 (2013).

15 Kern, J. W. & Schneewind, O. BslA, a pXO1-encoded adhesin of Bacillus anthracis. *Mol Microbiol* **68**, 504-515, doi:10.1111/j.1365-2958.2008.06169.x (2008).

16 Merrigan, M. M. *et al.* Surface-layer protein A (SlpA) is a major contributor to host-cell adherence of Clostridium difficile. *PLoS One* **8**, e78404, doi:10.1371/journal.pone.0078404 (2013).

17 Sakakibara, J. *et al.* Loss of adherence ability to human gingival epithelial cells in S-layer protein-deficient mutants of Tannerella forsythensis. *Microbiology (Reading)* **153**, 866-876, doi:10.1099/mic.0.29275-0 (2007).

18 Plaut, R. D. *et al.* Genetic evidence for the involvement of the S-layer protein gene sap and the sporulation genes spo0A, spo0B, and spo0F in Phage AP50c infection of Bacillus anthracis. *J Bacteriol* **196**, 1143-1154, doi:10.1128/JB.00739-13 (2014).

19 Royer, A. L. M. *et al.* Clostridioides difficile S-Layer Protein A (SlpA) Serves as a General Phage Receptor. *Microbiol Spectr* **11**, e0389422, doi:10.1128/spectrum.03894-22 (2023).

20 Sara, M. & Sleytr, U. B. Molecular sieving through S layers of Bacillus stearothermophilus strains. *J Bacteriol* **169**, 4092-4098, doi:10.1128/jb.169.9.4092-4098.1987 (1987).

21 von Kugelgen, A. *et al.* Membraneless channels sieve cations in ammonia-oxidizing marine archaea. *Nature*, doi:10.1038/s41586-024-07462-5 (2024).

22 Sleytr, U. B., Egelseer, E. M., Ilk, N., Pum, D. & Schuster, B. S-Layers as a basic building block in a molecular construction kit. *FEBS J* **274**, 323-334, doi:10.1111/j.1742-4658.2006.05606.x (2007).

23 An, B. *et al.* Engineered Living Materials For Sustainability. *Chem Rev* **123**, 2349-2419, doi:10.1021/acs.chemrev.2c00512 (2023).
24 Yuliawati, Y. *et al.* Potency of surface layer protein from Lactobacillus sp. as drug nanocarriers. *Journal of Applied Pharmaceutical Science*, doi:10.7324/JAPS.2024.199203 (2024).
25 Zhang, B. *et al.* Display of porcine epidemic diarrhea virus spike protein B-cell linear epitope on Lactobacillus mucosae G01 S-layer surface induce a robust immunogenicity in mice. *Microb Cell Fact* **23**, 142, doi:10.1186/s12934-024-02409-x (2024).
26 Decout, A. *et al.* Lactobacillus crispatus S-layer proteins modulate innate immune response and inflammation in the lower female reproductive tract. *Nat Commun* **15**, 10879, doi:10.1038/s41467-024-55233-7 (2024).
27 Li, J. *et al.* Exopolysaccharides and Surface-Layer Proteins Expressed by Biofilm-State Lactiplantibacillus plantarum Y42 Play Crucial Role in Preventing Intestinal Barrier and Immunity Dysfunction of Balb/C Mice Infected by Listeria monocytogenes ATCC 19115. *J Agric Food Chem* **72**, 8581-8594, doi:10.1021/acs.jafc.4c00460 (2024).
28 Konstantinov, S. R. *et al.* S layer protein A of Lactobacillus acidophilus NCFM regulates immature dendritic cell and T cell functions. *Proc Natl Acad Sci U S A* **105**, 19474-19479, doi:10.1073/pnas.0810305105 (2008).
29 Wang, H., Zhang, L., Li, Q., Xu, S. & Lu, R. Surface-layer protein produced by Lactobacillus crispatus JCM 2009 ameliorates lipopolysaccharide-induced inflammation through autophagy cross-talk with the NF-kappaB signaling pathway. *Int J Biol Macromol* **166**, 633-640, doi:10.1016/j.ijbiomac.2020.10.221 (2021).
30 Ormsby, M. J. *et al.* An intact S-layer is advantageous to Clostridioides difficile within the host. *PLoS Pathog* **19**, e1011015, doi:10.1371/journal.ppat.1011015 (2023).
31 Wang, Y. T., Oh, S. Y., Hendrickx, A. P., Lunderberg, J. M. & Schneewind, O. Bacillus cereus G9241 S-layer assembly contributes to the pathogenesis of anthrax-like disease in mice. *J Bacteriol* **195**, 596-605, doi:10.1128/JB.02005-12 (2013).
32 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).
33 Hunault, L. *et al.* Anti-S-layer monoclonal antibodies impact Clostridioides difficile physiology. *Gut Microbes* **16**, 2301147, doi:10.1080/19490976.2023.2301147 (2024).
34 Johnston, E., Isbilir, B., Alva, V., Bharat, T. A. M. & Doye, J. P. K. Punctuated and continuous structural diversity of S-layers across the prokaryotic tree of life. PREPRINT (Version 1) available at. *bioRxiv*, 2024.2005.2028.596244, doi:10.1101/2024.05.28.596244 (2024).
35 Buhlheller, C. *et al.* SymProFold: Structural prediction of symmetrical biological assemblies. *Nat Commun* **15**, 8152, doi:10.1038/s41467-024-52138-3 (2024).
36 Sogues, A. *et al.* Cryo-EM structure and polar assembly of the PS2 S-layer of Corynebacterium glutamicum. *bioRxiv*, doi:10.1101/2024.09.05.611363 (2024).
37 Dingle, K. E. *et al.* Recombinational switching of the Clostridium difficile S-layer and a novel glycosylation gene cluster revealed by large-scale whole-genome sequencing. *J Infect Dis* **207**, 675-686, doi:10.1093/infdis/jis734 (2013).
38 Ni Eidhin, D., Ryan, A. W., Doyle, R. M., Walsh, J. B. & Kelleher, D. Sequence and phylogenetic analysis of the gene for surface layer protein, slpA, from 14 PCR ribotypes of Clostridium difficile. *J Med Microbiol* **55**, 69-83, doi:10.1099/jmm.0.46204-0 (2006).
39 Long, X. *et al.* Peptostreptococcus anaerobius promotes colorectal carcinogenesis and modulates tumour immunity. *Nat Microbiol* **4**, 2319-2330, doi:10.1038/s41564-019-0541-3 (2019).
40 Lanzoni-Mangutchi, P. *et al.* Structure and assembly of the S-layer in C. difficile. *Nat Commun* **13**, 970, doi:10.1038/s41467-022-28196-w (2022).
41 Lewis, W. H., Tahon, G., Geesink, P., Sousa, D. Z. & Ettema, T. J. G. Innovations to culturing the uncultured microbial majority. *Nat Rev Microbiol* **19**, 225-240, doi:10.1038/s41579-020-00458-8 (2021).
42 Silale, A. *et al.* Dual function of OmpM as outer membrane tether and nutrient uptake channel in diderm Firmicutes. *Nat Commun* **14**, 7152, doi:10.1038/s41467-023-42601-y (2023).

43 Messner, P., Hollaus, F. & Sleytr, U. B. Paracrystalline Cell Wall Surface Layers of Different Bacillus stearothermophilus Strains. *International Journal of Systematic Bacteriology* **34**, 202-210, doi:10.1099/00207713-34-2-202 (1984).
44 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).
45 Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493-500, doi:10.1038/s41586-024-07487-w (2024).
46 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).
47 Chu, M. *et al.* A Clostridium difficile Cell Wall Glycopolymer Locus Influences Bacterial Shape, Polysaccharide Production and Virulence. *PLoS Pathog* **12**, e1005946, doi:10.1371/journal.ppat.1005946 (2016).
48 Schmidt, A., Muller, N., Schink, B. & Schleheck, D. A proteomic view at the biochemistry of syntrophic butyrate oxidation in Syntrophomonas wolfei. *PLoS One* **8**, e56905, doi:10.1371/journal.pone.0056905 (2013).
49 Pinheiro, B. A. *et al.* Functional insights into the role of novel type I cohesin and dockerin domains from Clostridium thermocellum. *Biochem J* **424**, 375-384, doi:10.1042/BJ20091152 (2009).
50 Artzi, L. *et al.* Cellulosomics of the cellulolytic thermophile Clostridium clariflavum. *Biotechnol Biofuels* **7**, 100, doi:10.1186/1754-6834-7-100 (2014).
51 Hager, F. F. *et al.* Functional Characterization of Enzymatic Steps Involved in Pyruvylation of Bacterial Secondary Cell Wall Polymer Fragments. *Front Microbiol* **9**, 1356, doi:10.3389/fmicb.2018.01356 (2018).
52 Oh, S. Y., Lunderberg, J. M., Chateau, A., Schneewind, O. & Missiakas, D. Genes Required for Bacillus anthracis Secondary Cell Wall Polysaccharide Synthesis. *J Bacteriol* **199**, doi:10.1128/JB.00613-16 (2017).
53 Fagan, R. P. & Fairweather, N. F. Clostridium difficile has two parallel and essential Sec secretion systems. *J Biol Chem* **286**, 27483-27493, doi:10.1074/jbc.M111.263889 (2011).
54 Ausiello, C. M. *et al.* Surface layer proteins from Clostridium difficile induce inflammatory and regulatory cytokines in human monocytes and dendritic cells. *Microbes Infect* **8**, 2640-2646, doi:10.1016/j.micinf.2006.07.009 (2006).
55 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).
56 Waligora, A. J. *et al.* Characterization of a cell surface protein of Clostridium difficile with adhesive properties. *Infect Immun* **69**, 2144-2153, doi:10.1128/IAI.69.4.2144-2153.2001 (2001).
57 Zhou, Q. *et al.* The cwp66 Gene Affects Cell Adhesion, Stress Tolerance, and Antibiotic Resistance in Clostridioides difficile. *Microbiol Spectr* **10**, e0270421, doi:10.1128/spectrum.02704-21 (2022).
58 Promadej, N., Fiedler, F., Cossart, P., Dramsi, S. & Kathariou, S. Cell wall teichoic acid glycosylation in Listeria monocytogenes serotype 4b requires gtcA, a novel, serogroup-specific gene. *J Bacteriol* **181**, 418-425, doi:10.1128/JB.181.2.418-425.1999 (1999).
59 Chateau, A. *et al.* Distinct Pathways Carry Out alpha and beta Galactosylation of Secondary Cell Wall Polysaccharide in Bacillus anthracis. *J Bacteriol* **202**, doi:10.1128/JB.00191-20 (2020).
60 Rismondo, J., Percy, M. G. & Grundling, A. Discovery of genes required for lipoteichoic acid glycosylation predicts two distinct mechanisms for wall teichoic acid glycosylation. *J Biol Chem* **293**, 3293-3306, doi:10.1074/jbc.RA117.001614 (2018).
61 Endo, T., Ikeo, K. & Gojobori, T. Large-scale search for genes on which positive selection may operate. *Mol Biol Evol* **13**, 685-690, doi:10.1093/oxfordjournals.molbev.a025629 (1996).
62 Blanc, G. *et al.* Molecular evolution of rickettsia surface antigens: evidence of positive selection. *Mol Biol Evol* **22**, 2073-2083, doi:10.1093/molbev/msi199 (2005).
63 Hunault, L. *et al.* A monoclonal antibody collection for C. difficile typing ? *Gut Pathog* **16**, 4, doi:10.1186/s13099-023-00592-7 (2024).
64 Chandra, H., Kovall, R. A., Yadav, J. S. & Sun, X. Host Immune Responses to Surface S-Layer Proteins (SLPs) of Clostridioides difficile. *Microorganisms* **11**, doi:10.3390/microorganisms11020380 (2023).

65 Sara, M. & Sleytr, U. B. Comparative studies of S-layer proteins from Bacillus stearothermophilus strains expressed during growth in continuous culture under oxygen-limited and non-oxygen-limited conditions. *J Bacteriol* **176**, 7182-7189, doi:10.1128/jb.176.23.7182-7189.1994 (1994).
66 Avall-Jaaskelainen, S. & Palva, A. Lactobacillus surface layers and their applications. *FEMS Microbiol Rev* **29**, 511-529, doi:10.1016/j.femsre.2005.04.003 (2005).
67 Moore, A. D., Bjorklund, A. K., Ekman, D., Bornberg-Bauer, E. & Elofsson, A. Arrangements in the modular evolution of proteins. *Trends Biochem Sci* **33**, 444-451, doi:10.1016/j.tibs.2008.05.008 (2008).
68 Pereira-Leal, J. B., Levy, E. D. & Teichmann, S. A. The origins and evolution of functional modules: lessons from protein complexes. *Philos Trans R Soc Lond B Biol Sci* **361**, 507-517, doi:10.1098/rstb.2005.1807 (2006).
69 von Kugelgen, A. *et al.* Interdigitated immunoglobulin arrays form the hyperstable surface layer of the extremophilic bacterium Deinococcus radiodurans. *Proc Natl Acad Sci U S A* **120**, e2215808120, doi:10.1073/pnas.2215808120 (2023).
70 Sagmeister, T. *et al.* The molecular architecture of Lactobacillus S-layer: Assembly and attachment to teichoic acids. *Proc Natl Acad Sci U S A* **121**, e2401686121, doi:10.1073/pnas.2401686121 (2024).
71 Tikhonova, T. V. *et al.* Extracellular Fe(III) reductase structure reveals a modular organization enabling S-layer insertion and electron transfer to insoluble substrates. *Structure* **31**, 174-184 e173, doi:10.1016/j.str.2022.12.010 (2023).
72 Navarre, W. W. & Schneewind, O. Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiol Mol Biol Rev* **63**, 174-229, doi:10.1128/MMBR.63.1.174-229.1999 (1999).
73 Lopez, R., Garcia, E., Garcia, P. & Garcia, J. L. The pneumococcal cell wall degrading enzymes: a modular design to create new lysins? *Microb Drug Resist* **3**, 199-211, doi:10.1089/mdr.1997.3.199 (1997).
74 Megrian, D., Taib, N., Jaffe, A. L., Banfield, J. F. & Gribaldo, S. Ancient origin and constrained evolution of the division and cell wall gene cluster in Bacteria. *Nat Microbiol* **7**, 2114-2127, doi:10.1038/s41564-022-01257-y (2022).
75 Megrian, D., Taib, N., Witwinowski, J., Beloin, C. & Gribaldo, S. One or two membranes? Diderm Firmicutes challenge the Gram-positive/Gram-negative divide. *Mol Microbiol* **113**, 659-671, doi:10.1111/mmi.14469 (2020).
76 Witwinowski, J. *et al.* An ancient divide in outer membrane tethering systems in bacteria suggests a mechanism for the diderm-to-monoderm transition. *Nat Microbiol* **7**, 411-422, doi:10.1038/s41564-022-01066-3 (2022).
77 Lagkouvardos, I. *et al.* The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. *Nat Microbiol* **1**, 16131, doi:10.1038/nmicrobiol.2016.131 (2016).
78 Hochberg, G. K. A. *et al.* A hydrophobic ratchet entrenches molecular complexes. *Nature* **588**, 503-508, doi:10.1038/s41586-020-3021-2 (2020).
79 Lukes, J., Archibald, J. M., Keeling, P. J., Doolittle, W. F. & Gray, M. W. How a neutral evolutionary ratchet can build cellular complexity. *IUBMB Life* **63**, 528-537, doi:10.1002/iub.489 (2011).
80 Isbilir, B., Yeates, A., Alva, V. & Bharat, T. A. M. Mapping the ultrastructural topology of the corynebacterial cell surface. *bioRxiv*, doi:10.1101/2024.09.05.611374 (2024).
81 Roberts, K. Crystalline glycoprotein cell walls of algae: their stucture, composition and assembly. *Philos Trans R Soc Lond B Biol Sci* **268**, 129-146, doi:10.1098/rstb.1974.0021 (1974).
82 Baudelet, P.-H., Ricochon, G., Linder, M. & Muniglia, L. A new insight into cell walls of Chlorophyta. *Algal Research* **25**, 333-371, doi:10.1016/j.algal.2017.04.008 (2017).
83 Rodrigues-Oliveira, T. *et al.* Actin cytoskeleton and complex cell architecture in an Asgard archaeon. *Nature* **613**, 332-339, doi:10.1038/s41586-022-05550-y (2023).
84 Imachi, H. *et al.* Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* **577**, 519-525, doi:10.1038/s41586-019-1916-6 (2020).
85 Sexton, D. L. *et al.* The cell envelope of Thermotogae suggests a mechanism for outer membrane biogenesis. *Proc Natl Acad Sci U S A* **120**, e2303275120, doi:10.1073/pnas.2303275120 (2023).

1029 86 Hashimi, A. & Tocheva, E. I. Cell envelope diversity and evolution across the bacterial tree of life.
1030 *Nat Microbiol* **9**, 2475-2487, doi:10.1038/s41564-024-01812-9 (2024).

1031