OPEN ACCESS

University of Dundee

University of Dundee

The majority of the matrix protein TapA is dispensable for Bacillus subtilis colony biofilm architecture

Earl, Chris; Arnaouteli, Sofia; Bamford, Natalie; Porter, Michael; Sukhodub, Tetyana; MacPhee, Cait E.

RESEARCH ARTICLE

WILEY

The majority of the matrix protein TapA is dispensable for *Bacillus subtilis* colony biofilm architecture

Chris Earl[1] | Sofia Arnaouteli[1] | Natalie C. Bamford[1] | Michael Porter [iD] [1] |
Tetyana Sukhodub[1] | Cait E. MacPhee[2] | Nicola R. Stanley-Wall [iD] [1]

[1]Division of Molecular Microbiology, School of Life Sciences, University of Dundee, Dundee, UK

[2]James Clerk Maxwell Building, School of Physics, University of Edinburgh, Edinburgh, UK

Correspondence
Nicola R. Stanley-Wall, Division of Molecular Microbiology, School of Life Sciences, University of Dundee, Dundee DD1 5EH, UK.
Email: n.r.stanleywall@dundee.ac.uk

Cait E. MacPhee, School of Physics and Astronomy, University of Edinburgh, James Clerk Maxwell Building, Edinburgh EH9 3JZ, UK.
Email: cait.macphee@ed.ac.uk

Funding information
Biotechnology and Biological Sciences Research Council, Grant/Award Number: BB/M013774/1, BB/N022254/1 and BB/R012415/1; Wellcome Trust, Grant/Award Number: 097818/Z/11

Abstract

Biofilm formation is a co-operative behaviour, where microbial cells become embedded in an extracellular matrix. This biomolecular matrix helps manifest the beneficial or detrimental outcome mediated by the collective of cells. *Bacillus subtilis* is an important bacterium for understanding the principles of biofilm formation. The protein components of the *B. subtilis* matrix include the secreted proteins BslA, which forms a hydrophobic coat over the biofilm, and TasA, which forms protease-resistant fibres needed for structuring. TapA is a secreted protein also needed for biofilm formation and helps in vivo TasA-fibre formation but is dispensable for in vitro TasA-fibre assembly. We show that TapA is subjected to proteolytic cleavage in the colony biofilm and that only the first 57 amino acids of the 253-amino acid protein are required for colony biofilm architecture. Through the construction of a strain which lacks all eight extracellular proteases, we show that proteolytic cleavage by these enzymes is not a prerequisite for TapA function. It remains unknown why TapA is synthesised at 253 amino acids when the first 57 are sufficient for colony biofilm structuring; the findings do not exclude the core conserved region of TapA having a second role beyond structuring the *B. subtilis* colony biofilm.

KEYWORDS
Bacillus subtilis, biofilm matrix, extracellular proteases, TapA, TasA

1 | INTRODUCTION

The predominant state in which bacteria and archaea live on Earth is in the form of biofilms (Flemming *et al*., 2016): aggregates of microorganisms embedded in a self-made extracellular matrix. Molecules that can be found in the biofilm matrix include extracellular DNA, exopolysaccharides, lipids and proteins, the latter of which can self-assemble into a variety of functional forms including filaments, films, or fibres (Erskine *et al*., 2018a). The biofilm matrix conveys 'emergent properties' to the cells in the biofilm (Dragos and Kovacs, 2017) including, but not limited to, providing structure and stability to the community, aiding the sequestration of nutrients and retaining extracellular enzymes which facilitates further processing of the biofilm matrix (Flemming *et al*., 2016).

Biofilm formation by the Gram-positive bacterium *Bacillus subtilis* has been intensively studied and the production of the molecules in the matrix is known to be highly controlled by a suite of transcription regulators (Cairns *et al*., 2014). The exopolymeric matrix of the

B. subtilis biofilm is composed of polymers that include an exopoly-saccharide (EPS) and an amphiphilic protein BslA which assembles to form a hydrophobic film on the biofilm surface (Kobayashi and Iwano, 2012; Hobley *et al.*, 2013). The major protein component of the *B. subtilis* matrix is protease-resistant fibres formed by the secreted protein TasA (Romero *et al.*, 2010; Erskine *et al.*, 2018b). These fibres provide structural integrity to the biofilm and are necessary for the characteristic wrinkled phenotype of biofilms and pellicles (Branda *et al.*, 2006). Experimental models to study biofilm formation include colony biofilms grown on a semi-solid agar surface and floating pellicles in which the biofilm forms at an air-liquid interface in standing liquid culture.

The *tasA* coding region is located within an operon alongside two other genes: namely *sipW* and *tapA* (formerly *yqxM*) (Zhu and Stülke, 2018). SipW is a specialised signal peptidase that is linked with removal of the signal peptide from both TasA (Stover and Driks, 1999b) and TapA (Stover and Driks, 1999a) during the secretion process. SipW also modulates expression of the *epsA-O* operon which encodes the proteins needed for the biofilm exopolysaccharide (Terra *et al.*, 2012). Thus, *sipW* is essential for biofilm formation. TapA is described as an accessory protein that is thought to be needed for the formation of TasA fibres (Romero *et al.*, 2010; Romero *et al.*, 2011), and more specifically, for the attachment of the TasA fibres to the cell surface. Secondary structure analysis revealed TapA to be a two-domain protein with significant regions of disorder (Abbasi *et al.*, 2019). Additionally TapA has recently been shown to form fibres (El Mammeri *et al.*, 2019). The absence of TapA is correlated with a reduction in the level of TasA in the biofilm matrix (Romero *et al.*, 2014). Evidence indicates however that, in the absence of TapA, recombinant TasA self-assembles into protease-resistant fibres that are structurally and functionally comparable to native fibres extracted from *B. subtilis* (Erskine *et al.*, 2018b; El Mammeri *et al.*, 2019). Moreover, when provided exogenously these self-assembled recombinant TasA fibres are also biologically active in vivo in the absence of TapA, suggesting that cell-surface attachment is not critical for biofilm architecture (Erskine *et al.*, 2018b). Thus, further evaluation of the function and activity of TapA is warranted.

Here we identify that amino acids 1-57 (inclusive) of TapA form a minimal functional unit of the protein that is required to give rise to the complex architecture of the *B. subtilis* colony and pellicle biofilm. Heterologous provision of the DNA encoding this truncated form of TapA is sufficient to restore rugose biofilm formation to the *tapA* deletion strain (the full-length protein is 253 amino acids in length). We identify, through site-directed mutagenesis, key amino acids in the minimal, functional TapA form that are required for bioactivity and, in doing so, uncover essential hydrophobic amino acids. We show that in vivo TapA is proteolytically cleaved to lower molecular weight forms by the native extracellular proteases secreted into the external environment. We demonstrate that Vpr, a serine protease, plays a specific role in cleavage of the TapA protein. Finally, we establish that proteolysis of TapA by the extracellular proteases is not a prerequisite to activity and that TapA can fulfil its role whether it is cleaved or not.

2 | RESULTS

2.1 | Identification of a TapA minimal functional unit

To investigate the functional region(s) of *tapA* an in-frame deletion was constructed in *B. subtilis* NCIB3610. As expected (Romero *et al.*, 2011), the rugose architecture exhibited by the colony and pellicle biofilms formed by the parental isolate was absent when *tapA* was deleted (Figures 1a and S1a). The Δ*tasA* strain is shown for reference (NRS5267) revealing the flat featureless biofilm that manifests when the TasA fibres are no longer synthesised. The architecture of the Δ*tapA* strain was fully reinstated when the *tapA* coding region was expressed from the heterologous *amyE* locus using an IPTG inducible promoter (Figures 1a and S1a). The *tapA* gene is present in the genome of a range of *Bacillus* species and analysis of the protein sequences reveals domains with a high degree of conservation and other regions of variability (Figure 1b). This includes a marked difference in the length of the *tapA* coding region (Figure 1b) (Romero *et al.*, 2014); experimental data showed that the coding region for amino acids 194-230 of TapA was dispensable (Romero *et al.*, 2014). Here, to determine the minimal *tapA* coding region needed for function, we systematically deleted *tapA* from the 3' end and tested the ability of the variant-length *tapA* constructs to genetically complement the *tapA* deletion strain. In total, 22 variants of the *tapA* coding region were assessed (Figures 1b,c and S1b). We concluded that the TapA$_{1-60}$ variant was fully capable of recovering architecture of the colony and pellicle biofilm to the *tapA* mutant, while the TapA$_{1-50}$ form lacked this ability (Figure 1c). We, therefore, made constructs with single codon deletions in the region encoding TapA amino acids between 60 and 50 (see Figure 1b). We established that amino acids 1-57 (inclusive) represented the minimal form of TapA that is capable of reinstating colony and pellicle biofilm architecture to the *tapA* deletion strain. This conclusion was reached through a visual analysis of colony biofilm and pellicle rugosity (Figures 1c and S1b,c). Additionally, as the level of TasA (calculated molecular mass 25.7 kDa) in the biofilm is substantially reduced in the absence of functional TapA (Romero *et al.*, 2014) (Figure S1d), the bioactivity of the *tapA* truncations was supported by immunoblot analysis which showed a recovery of TasA levels back to those seen for NCIB3610 (Figure S1d). The identification of this region of *tapA* as sufficient for TapA activity is consistent with, but significantly extends, the previous identification of amino acids 50-57 as being needed for TapA function in the context of the full-length protein (Romero *et al.*, 2014).

2.2 | Amino acids critical for function in the minimal functional region of TapA

We were interested in the features within the TapA minimal form that conferred activity. The in silico predicted TapA signal sequence comprises the first 43 amino acids (Petersen *et al.*, 2011) that is purported to be cleaved by a specialised signal peptidase, SipW
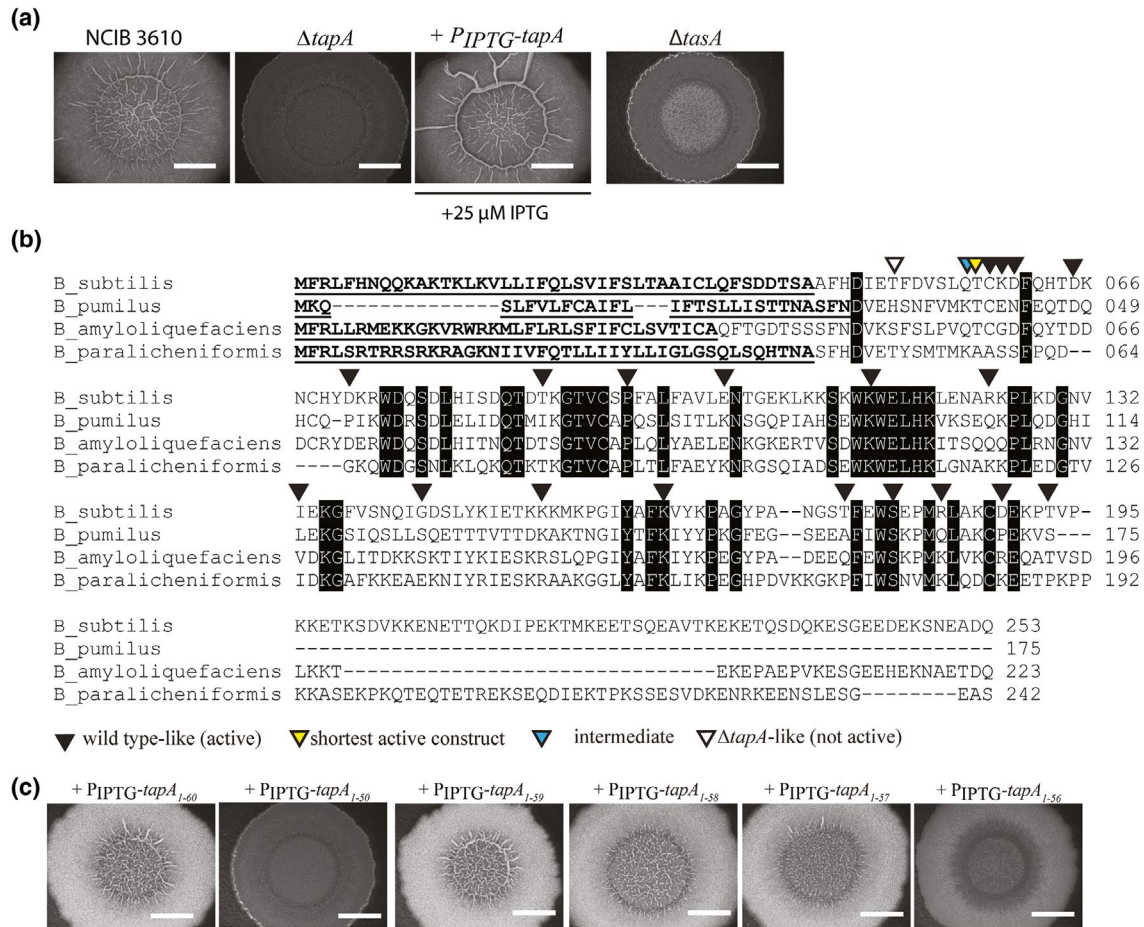
FIGURE 1 The *tapA* coding region is functional when truncated. (a) Colony biofilms formed by NCIB3610, Δ*tapA* (NRS3936), +P$_{IPTG}$-*tapA* (NRS5045) and Δ*tasA* (NRS5267); (b) Alignment of TapA protein sequences from *B. subtilis*, *B. pumilus*, *B. amyloliquefaciens* and *B. paralicheniformis*. The percentage amino acid sequence identity with regards the *B. subtilis* TapA sequence is as follows: *B. pumilus*- 42%, *B. amyloliquefaciens*- 49% and *B. paralicheniformis*- 38%. The bold underlined sequence represents the signal sequence, the black boxes indicate identical amino acids; The constructs generated and the ability to restore rugose biofilm architecture to the Δ*tapA* deletion strain when expressed from an ectopic position on the chromosome are indicated by the inverted triangles above the amino acid sequence; (c) Representative colony biofilms formed by + P$_{IPTG}$-*tapA*$_{1-60}$ (NRS6044), P$_{IPTG}$-*tapA*$_{1-50}$ (NRS6002), +P$_{IPTG}$-*tapA*$_{1-59}$ (NRS6043), +P$_{IPTG}$-*tapA*$_{1-58}$ (NRS6042), +P$_{IPTG}$-*tapA*$_{1-57}$ (NRS6041), and + P$_{IPTG}$-*tapA*$_{1-56}$ (NRS6025). In (A) and (C) biofilms were grown at 30°C for 48 hr in the presence of 25 μM IPTG. Biofilm images are representative of at least three independent biological replicates. The scale bars represent 1 cm

(Stover and Driks, 1999a). Here we demonstrate that the first 43 amino acids are not needed for activity, as the 43 amino acid sequence can be replaced with the 28 amino acid TasA signal sequence (which is also cleaved by SipW; Stover and Driks, 1999a; Stover and Driks, 1999b). More specifically, when the chimeric construct, P$_{IPTG}$-*tasAss-tapA*$_{44-253}$, was expressed in the *tapA* deletion strain rugose architecture was fully reinstated to the colony and pellicle biofilms (Figure S2a,b). Therefore, the influence of amino acids 1-43 of TapA with respect to protein function were excluded from further analysis.

Bioinformatics analysis of TapA from *B. subtilis*, and the sequences of orthologous proteins, reveal that each of the proteins are predicted to form a β-strand within the region needed for function, with conservation found on the level of amino acid properties rather than identity (Figure 2a). Consistent with these conclusions, orthologous coding regions could genetically substitute for the *B. subtilis tapA* coding region through heterologous expression of

the *tapA* coding region from *B. amyloliquefaciens*, *B. paralicheniformis* and *B. pumilus* in the *B. subtilis tapA* deletion strain. The resulting architecture of the colony biofilms upon expression of the *tapA* orthologues was indistinguishable from those formed by the wild-type (WT) NCIB3610 (Figure 2b). These findings are consistent with, but extend those previously published that revealed *tapA* from *B. amyloliquefaciens* was able to functionally replace TapA of *B. subtilis* (Romero *et al.*, 2014).

To identify amino acids critical to TapA function, site-directed mutagenesis was used to generate a series of constructs containing systematic substitutions in the *tapA*$_{44-57}$ coding region in the context of the minimal TapA$_{1-57}$ construct (Figure 2c). The variant constructs were introduced into the *tapA* deletion strain and the ability of the variant TapA forms to restore architecture to colony biofilms was assessed (Figure 2d-r). One key finding uncovered was that the length of the TapA$_{1-57}$ variant form is the important feature driving activity
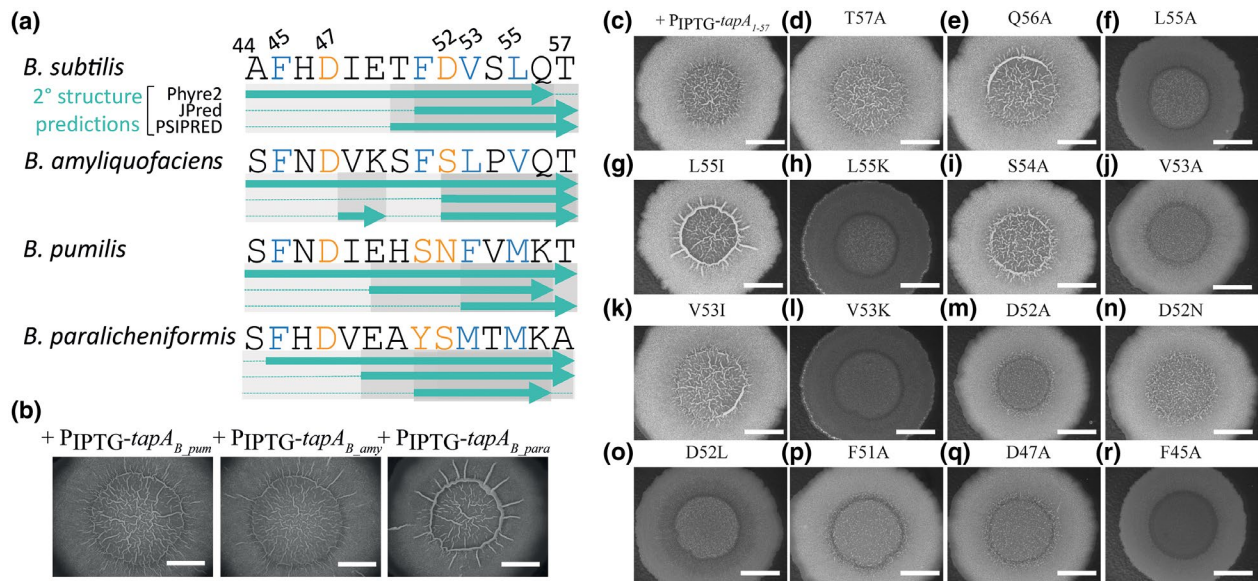
FIGURE 2 Identification of critical amino acids in the *tapA* coding region. (a) Schematic of the region of the *B. subtilis* TapA protein sequence from amino acids 44-57, with secondary structure predictions from Phyre2, JPred and PSIPRED servers, the equivalent regions from *B. pumilus*, *B. amyloliquefaciens* and *B. paralicheniformis* are shown for comparison. Green arrows represent predicted β-strands, residues critical for function in *B. subtilis* TapA$_{1-57}$ are shown in colour: hydrophobic residues are in blue; polar residues are in orange; (b) representative colony biofilms formed by + P$_{IPTG}$-*tapA*$_{B\_pum}$ (NRS5046), +P$_{IPTG}$-*tapA*$_{B\_amy}$ (NRS5047), and + P$_{IPTG}$-*tapA*$_{B\_para}$ (NRS5741). n > 3. Biofilms formed by (c) +P$_{IPTG}$-*tapA*$_{1-57}$ (NRS6041); (d) +P$_{IPTG}$-*tapA*$_{1-57}$ T57A (NRS6384); (e) +P$_{IPTG}$-*tapA*$_{1-57}$ Q56A (NRS6385); (f) +P$_{IPTG}$-*tapA*$_{1-57}$ L55A (NRS6472); (g) +P$_{IPTG}$-*tapA*$_{1-57}$ L55I (NRS6386); (h) +P$_{IPTG}$-*tapA*$_{1-57}$ L55K (NRS6387); (i) +P$_{IPTG}$-*tapA*$_{1-57}$ S54A (NRS6388); (j) +P$_{IPTG}$-*tapA*$_{1-57}$ V53A (NRS6473); (k) +P$_{IPTG}$-*tapA*$_{1-57}$ V53I (NRS6389); (l) +P$_{IPTG}$-*tapA*$_{1-57}$ V53K (NRS6502); (m) +P$_{IPTG}$-*tapA*$_{1-57}$ D52A (NRS6476); (n) +P$_{IPTG}$-*tapA*$_{1-57}$ D52N (NRS6516); (o) +P$_{IPTG}$-*tapA*$_{1-57}$ D52L (NRS6477); (p) +P$_{IPTG}$-*tapA*$_{1-57}$ F51A (NRS6390); (q) +P$_{IPTG}$-*tapA*$_{1-57}$ D47A(NRS6475); and (r) +P$_{IPTG}$-*tapA*$_{1-57}$ F45A (NRS6474). Biofilms were grown at 30°C for 48 hr in the presence of 25 μM IPTG. n = 3 biological replicates. The scale bars represent 1 cm

of TapA$_{1-57}$ bioactivity, as provision of the *tapA* coding region for amino acids 1-56 was unable to support biofilm recovery (Figure 1c), whereas that encoding TapA$_{1-57}$ or a variant of TapA$_{1-57,}$ where threonine 57 was replaced with alanine were biologically active (Figures 1c and 2c). We also identified amino acids F45, D47, F51, D52, V53 and L55 as critical for TapA function, where the hydrophobicity of the amino acids V53 and L55 was a key feature mediating biological activity. Further analysis will be required to elucidate the exact role the amino acids play in facilitating TapA function (Table 1).

2.3 | TapA is detected in vivo at a low molecular mass

As the truncated variant of TapA covering amino acids 1-57 (inclusive) is sufficient for the mature architecture of the colony biofilm to develop, we hypothesised that TapA could be processed in vivo to a smaller, potentially active, form. To probe the molecular weight of TapA in the colony biofilm we raised a custom TapA antibody that was able to detect recombinant TapA$_{44-253}$ protein as a single band at ~30 kDa by immunoblot (calculated molecular mass ~25 kDa) (Figure 3a). In contrast to purified recombinant protein, analysis using the same αTapA antibodies against proteins extracted from the WT NCIB3610 colony biofilm revealed three distinct bands with apparent molecular masses of approximately 30 kDa, 18 kDa,

and 16 kDa. Each band was absent from the protein sample derived from the *tapA* mutant colony (Figure 3b) and returned in an IPTG-dependent manner when proteins from the *tapA* complementation strain (NRS5045) was probed (Figure 3b).

The detected band profile was unexpected and did not reveal the anticipated single ~25 kDa band of the mature secreted protein. Thus, to test whether the banding profile of TapA was sustained amongst different *Bacillus* species, we analysed protein extracts by immunoblot from three other *B. subtilis* isolates, namely RO-FF-1, ATCC 9799 and B-14393T. After 48 hr incubation each of the isolates had formed a colony biofilm with a rugose architecture (Figure 3c). Using the αTapA antibody on protein extracts from the mature colony biofilms we revealed that the TapA banding pattern was replicated in each of the distinct isolates (Figure 3d).

2.4 | The C-terminus of TapA is processed in vivo

To understand in more detail the banding pattern detected when using the αTapA antibody to challenge protein extracts of the WT strain, we assessed the immunoblot banding profile for the *tapA* mutant that expressed either the *tapA*$_{1-193}$ (intermediate length, Mw 17.3 kDa), *tapA*$_{1-183}$ (intermediate length, Mw 16.2 kDa) or the *tapA*$_{1-57}$ (minimal length, Mw 1.6 kDa) coding region. We used the *tapA*$_{1-253}$ full-length coding sequence as a reference for the banding
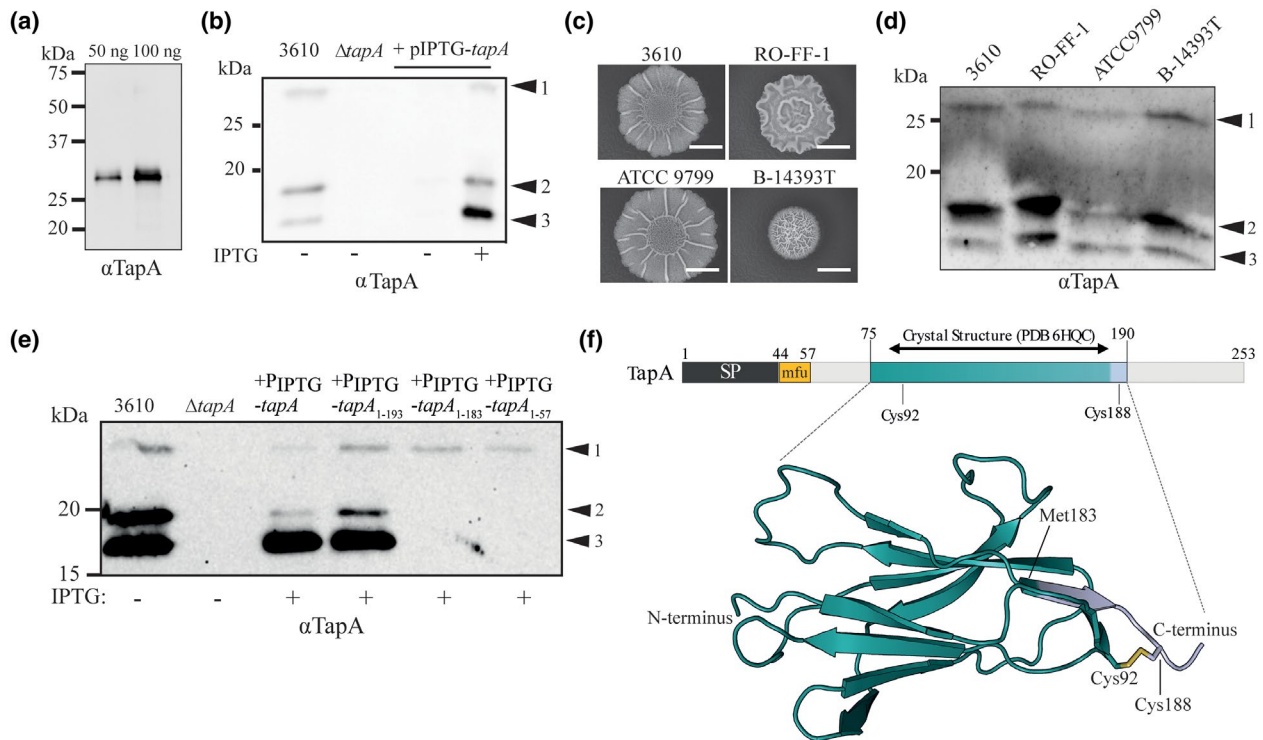
FIGURE 3 TapA is processed in vivo. (a) Immunoblot analysis of 50 and 100 ng of recombinant TapA$_{44-253}$ using αTapA antibodies; (b) Immunoblot analysis of proteins extracted from biofilms formed by NCIB3610, ΔtapA (NRS3936), +P$_{IPTG}$-tapA (NRS5045) (in the absence or presence of 25 μM IPTG as indicated) using αTapA antibodies. n = 2; (c) Biofilms formed by *B. subtilis* isolates NCIB3610, RO-FF-1, ATCC 9799, and B-14393T after growth at 30°C for 48 hr. Biofilm images are representative of at least three independent replicates. The scale bars represent 1 cm; (d) Immunoblot analysis of proteins extracted from biofilms formed by NCIB3610, RO-FF-1, ATCC 9799 and B-14393T using αTapA antibodies. n = 3; (e) Immunoblot analysis of proteins extracted from biofilms formed by NCIB3610, ΔtasA (NRS5267), ΔtapA (NRS3936), +P$_{IPTG}$-tapA$_{1-253}$ (NRS5045), +P$_{IPTG}$-tapA$_{1-193}$ (NRS5744), +P$_{IPTG}$-tapA$_{1-183}$ (NRS5790) and + P$_{IPTG}$-tapA$_{1-57}$ (NRS6044) using αTapA antibodies. About 25 μM IPTG was used as indicated, n = 2; Arrow 1 highlights the ~30 kDa band, arrow 2 the ~18 kDa band, and arrow 3 the ~16 kDa band. (f) Linear schematic of TapA outlining the predicted domains: signal peptide (SP) in grey; minimal functional unit (mfu) in yellow; and the crystal structure (residues 75-190, PDB 6HQC) in teal and light blue. The one disulphide bond, between Cys92 and Cys188, is shown in stick representation. The region of the folded domain that would be missing in the tapA$_{1-183}$ construct is in light blue, in both the linear and tertiary structure representations. The crystal structure was visualised using PyMOL 2.0

pattern of WT. Importantly, each of these variants of *tapA* can reinstate the WT architecture to the *tapA* deletion strain (see Figures 1c and S1) and produces variant TapA forms with different calculated molecular masses. We hypothesised that if the ~30 kDa band corresponded to the full-length secreted TapA protein (calculated Mw ~25 kDa), a size shift would be apparent in protein samples isolated after expression of the variant *tapA* coding regions. However, the protein samples extracted from cells expressing either the *tapA$_{1-253}$* or *tapA$_{1-193}$* coding region revealed identical banding patterns after probing with the αTapA antibody; a truncated form of TapA was not detected. Furthermore, the *tapA$_{1-183}$* and *tapA$_{1-57}$* forms only presented a single faint band that was detected at ~30 kDa (Figure 3e). As a difference in the apparent mass of at least one of the bands detected in the *tapA$_{1-253}$* or *tapA$_{1-193}$* samples was not observed. The simplest explanation is that the 30kDa band is a non-specific band, and therefore the only forms of TapA detected are the 18 kDa and 16 kDa forms. These findings allow us to deduce that TapA is likely to be processed in the WT strain such that the C-terminal domain is removed. These data also demonstrate that the ~30 kDa band is not

specific to TapA, but rather is a non-specific protein detected by the TapA antibody that is dependent on a functional form of TapA being made by the cell.

The reason why TapA$_{1-183}$ (and other forms with more extreme truncations) cannot be detected by immunoblot can be informed by analysis of the recently released crystal structure of TapA$_{75-190}$ (PDB 6HQC; Roske *et al.*, 2018). TapA$_{75-190}$ forms a β-sandwich most structurally similar to the macroglobulin folds found in bacterial lipoproteins from *Escherichia coli* (PDB 4ZIQ; Garcia-Ferrer *et al.*, 2015) and *Salmonella typhimurium* (PDB 4U4J; Wong and Dessen, 2014) α-2-macroglobulins (Figure 3f). The structure of TapA contains a disulphide bond between Cys-92 and Cys-188. Therefore, the TapA$_{1-183}$ variant protein would lack the disulphide bond and would lose the C-terminal β-strand that would disrupt one of the β-sheets. Thus, the TapA$_{1-183}$ construct is highly likely to render the usually folded domain unstable, and therefore more prone to degradation. This would result in TapA being undetectable by immunoblot. We reason that as TapA is active as a minimal unit of 57 amino acids, this N-terminal portion of the protein must remain present, albeit not detectable, in

TABLE 1 Qualitative description of biofilm phenotype for variant *tapA* minimal functional unit forms

Variant *tapA* region	Biofilm phenotype[a]	Strain number
+ P_{IPTG}-$tapA_{1-57}$	Wild type-like	NRS6041
+ P_{IPTG}-$tapA_{1-57(T57A)}$	Wild type-like	NRS6384
+ P_{IPTG}-$tapA_{1-57(Q56A)}$	Wild type-like	NRS6385
+ P_{IPTG}-$tapA_{1-57(L55I)}$	Wild type-like	NRS6386
+ P_{IPTG}-$tapA_{1-57(L55K)}$	Δ*tapA*-like	NRS6387
+ P_{IPTG}-$tapA_{1-57(L55A)}$	Δ*tapA*-like	NRS6472
+ P_{IPTG}-$tapA_{1-57(S54A)}$	Wild type-like	NRS6388
+ P_{IPTG}-$tapA_{1-57(V53I)}$	Wild type-like	NRS6389
+ P_{IPTG}-$tapA_{1-57(V53K)}$	Δ*tapA*-like	NRS6502
+ P_{IPTG}-$tapA_{1-57(V53A)}$	Intermediate	NRS6473
+ P_{IPTG}-$tapA_{1-57(D52A)}$	Intermediate	NRS6476
+ P_{IPTG}-$tapA_{1-57(D52L)}$	Δ*tapA*-like	NRS6477
+ P_{IPTG}-$tapA_{1-57(D52N)}$	Intermediate	NRS6516
+ P_{IPTG}-$tapA_{1-57(F51A)}$	Intermediate	NRS6390
+ P_{IPTG}-$tapA_{1-57(D47A)}$	Intermediate	NRS6475
+ P_{IPTG}-$tapA_{1-57(F45A)}$	Δ*tapA*-like	NRS6474

[a]Biofilm phenotypes are described qualitatively as: 1) 'wild type-like', displayed prominent corrugations resembling NCIB3610 *B. subtilis* biofilms; 2) 'intermediate', failed to form prominent wrinkles but more complex architecture than displayed by the Δ*tapA* strain; and 3) 'Δ*tapA*-like', retained flat, featureless biofilms shown by the Δ*tapA* mutant.

the colony biofilm formed by strains that express either $tapA_{1-183}$ or $tapA_{1-57}$. The development of a rugose architecture of a colony biofilm when TapA bands cannot be detected using the αTapA antibody indicates that the bands are not representative of the region of the protein responsible for conferring rugose architecture.

2.5 | Recombinant TapA is cleaved by extracellular proteases

The consistent detection of TapA at low molecular mass forms in the immunoblots is suggestive that in vivo TapA is subjected to processing. These findings led us to the hypothesis that TapA is cleaved in the extracellular environment to release a functionally active form of the protein. *B. subtilis* encodes eight major secreted extracellular proteases (Bpr, Vpr, NprB, Mpr, Epr, AprE, NprE and WprA) (Zhu and Stülke, 2018) that might cleave TapA. We revealed that TapA is sensitive to proteolysis as incubation of recombinant $TapA_{44-253}$ with spent culture supernatant, containing extracellular proteases secreted by NCIB3610, resulted in full size $TapA_{44-253}$ no longer being detected by SDS-PAGE analysis. In a control reaction where the spent culture supernatant was heat treated for 10 min at 100°C, to denature the native exoproteases, the full-length $TapA_{44-253}$ was detected (Figure 4a). Taken together with data indicating that TapA is a secreted protein, these findings are consistent with the hypothesis that the extracellular proteases could cleave TapA in vivo.
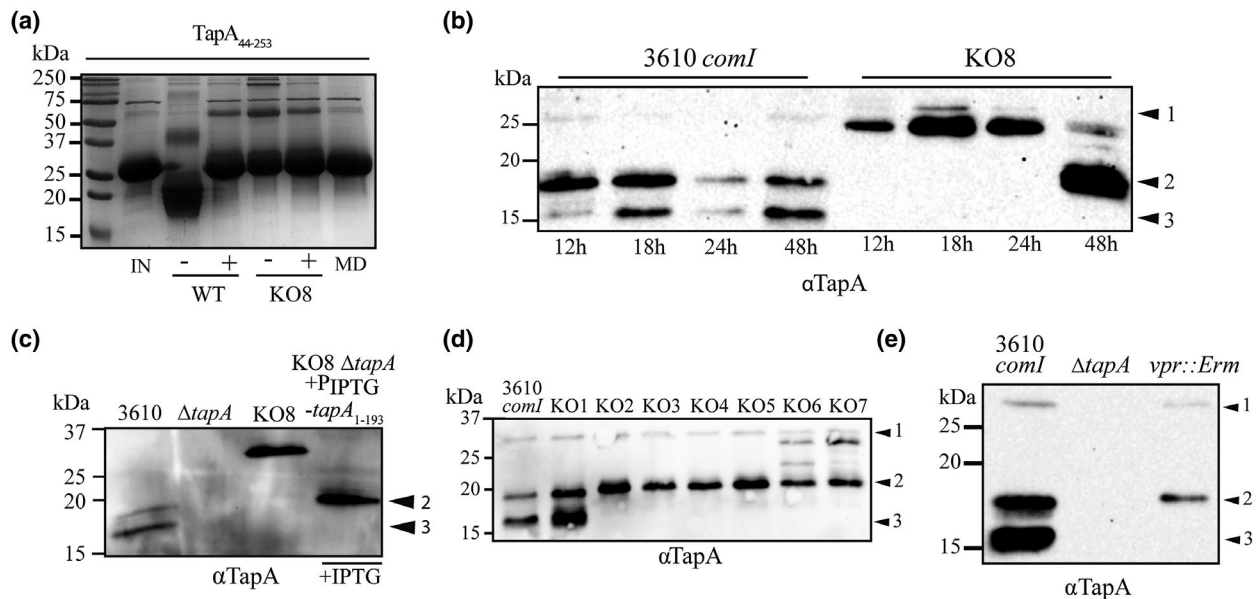

FIGURE 4 TapA is cleaved by the extracellular proteases secreted by *B. subtilis*. (a) Integrity of 28 μg of recombinant $TapA_{44-253}$ incubated for 8 hr at 37°C analysed by SDS-PAGE. The protein (IN) was incubated with filtered spent supernatants, or MSgg medium (MD), collected from NCIB3610 *comI* (NRS6017) and KO8 (NRS5645). The (−) and (+) indicate if the supernatant had been heat inactivated at 100°C prior to incubation with the recombinant protein. n = 2; (b) Immunoblot analysis of proteins extracted from biofilms formed by NCIB3610 and KO8 (NRS5645), at 12, 18, 24 and 48 hr, using αTapA antibodies. n = 2. (c) Immunoblot analysis of NCIB3610, Δ*tapA* (NRS3936), KO8 (NRS5645) and KO8 + P_{IPTG}-$tapA_{1-193}$ (NRS7011) biofilm extracts at 18 hr with αTapA antibodies. n = 2; (d) Immunoblot of 48 hr biofilms of 3610 *comI* and KO1 to KO7 using αTapA antibodies. n = 2. (e) Immunoblot of biofilms harvested at 48 hr of 3610 *comI* (NRS6017), Δ*tapA* (NRS3936) and *vpr::erm* (NRS7010) using αTapA antibodies. n = 2. Arrow 1 highlights the ~30 kDa band, arrow 2 the ~18 kDa band, and arrow 3 the ~16 kDa band

2.6 | Removing the extracellular proteases impacts TapA processing

To test the potential role of the extracellular proteases on TapA cleavage in vivo, we constructed a strain that lacked eight secreted exoproteases (hereafter KO8: NRS5645 *bpr, vpr, nprB, mpr, epr, aprE, nprE* and *wprA*). The KO8 strain showed no evidence of proteolytic activity when grown on LB agar for 20 hr that was supplemented with 1.5% (w/v) milk, similar to the *degU* deletion strain which was used as a negative control (Figure S3a) (Msadek *et al.*, 1990). Moreover, recombinant TapA$_{44-253}$ protein was stable when incubated in spent culture supernatant isolated from the KO8 strain (Figure 4a). The impact of removing the extracellular proteases on the size of TapA in vivo was assessed by immunoblot analysis of proteins extracted from the WT and KO8 strain after growth for 12, 18, 24 or 48 hr under colony biofilm formation conditions. The ~30, ~18 and ~16 kDa bands were detected in the WT strain at each time point. In contrast, the ~16 kDa form was not detected in the KO8 strain at any of the time points tested and the ~18 kDa band was only detected after 48 hr. An additional band with an apparent molecular mass of ~25 kDa was detected in the KO8 samples at 12, 18, 24 and 48 hr (Figure 4b). A molecular mass of ~25 kDa band is broadly consistent with the calculated mass of mature secreted TapA. When the *tapA*$_{1-193}$ coding region was expressed in a KO8 *tapA* strain, the ~25 kDa band was no longer detected. Instead, a band of ~18 kDa was present (Figure 4c) [17.3 kDa is the calculated molecular weight of the TapA$_{1-193}$ variant minus the 43 amino acid signal sequence]. Taken together these data reveal that TapA is cleaved in vivo at two distinct positions, one position which generates the ~18 kDa form and another position which yields the ~16 kDa form.

2.7 | Vpr cleaves TapA in a specific manner

Cleavage of TapA into the ~16 kDa form was completely blocked in the KO8 strain. To elucidate which exoprotease was responsible we examined the TapA banding profile in the suite of strains built during the construction of the NCIB3610 KO8 strain. These are called strains KO1 to KO7 (see Table S1). We noted that the ~16 kDa TapA band was not detected in the protein samples extracted from colony biofilms once the coding region for *vpr* was deleted (Figure 4d). Consistent with this, in a single *vpr::erm* NCIB3610 strain (NRS7010) only one αTapA antibody reactive band was detected after 48 hr growth under biofilm formation conditions, the band at ~18 kDa (Figure 4e). We, therefore, conclude that Vpr is needed for the generation of the ~16 kDa form of TapA.

2.8 | Cleavage of TapA by the extracellular proteases is not needed for biofilm architecture

Having demonstrated that TapA is processed in the extracellular environment, we asked if processing was a prerequisite for the architecture of the *B. subtilis* colony and pellicle biofilms to develop. We found that the KO8 strain displayed a minor defect in colony biofilm architecture, with the exoprotease-free strain showing fewer of the large corrugations seen in the colony biofilm formed by the parental strain (Figure 5a). The macroscale images of the pellicle biofilms formed by the two strains were largely comparable (Figure 5a). Imaging of the colony biofilms at the microscale using confocal microscopy revealed very little difference in comparison to the WT strain in terms of wrinkle formation at the microscale, however, there was a small, but consistent, difference in the overall height of structures in the biofilm (Figures 5b and S3b). Images of the biofilms were acquired from the central region of the colony after 12, 18, 24 and 48 hr incubation. Analysis of the extent to which the features in the biomass extended in the Z direction revealed that from 12 to 24 hr there was no substantial difference between WT and KO8 strains, with the volume in μm^3 for the samples being comparable in each of the biofilms (Figure 5b). Although both WT and KO8 strains showed significant increase in volume between 24 and 48 h, the change was slightly greater for WT, at 2.6-fold, compared to 2.4-fold for KO8 (*p* < .005) (Figure 5b). The alteration in biofilm architecture displayed by the KO8 strain is unlikely to be a consequence of a generalised growth defect as both the WT and KO8 strain exhibit comparable growth rates and yields when grown under shaking culture conditions in MSgg (Figure 5c).

The NCIB3610 KO8 strain shows altered cleavage of TapA and a minor alteration in the rugosity of the colony biofilm architecture. We reasoned that if the differences in the rugosity displayed by the KO8 strain compared with WT were due to a lack of TapA cleavage, introduction of the minimal functional unit coding region of *tapA*, TapA$_{1-57}$, at the heterologous *amyE* location would recover biofilm architecture. However, we found that TapA$_{1-57}$ was unable to reinstate the larger architectural features to the KO8 biofilms (Figure 5d). We confirmed that the TapA$_{1-57}$ variant form was functional in the KO8 strain as both the full-length and minimal *tapA* coding region could restore the colony biofilm morphology displayed by the KO8 Δ*tapA* biofilms, such that the phenotype mimicked the parental KO8 strain (Figure S3c). In addition, we assessed the direct impact of lacking Vpr using a *vpr* single deletion strain. This strain is unable to cleave TapA to yield the ~16 kDa form. The *vpr* deletion isolate retained a colony biofilm morphology that could not be distinguished from the WT NCIB3610 (Figure S3c). Therefore, we conclude that cleavage of TapA, at least to a 16 kDa form, is not a prerequisite to generate functionally active TapA, as assessed by structure of the colony and pellicle biofilm. Thus, the subtle impact on biofilm architecture which manifests upon removal of the eight extracellular protease genes in the KO8 strain must be TapA independent.

3 | DISCUSSION

An extracellular matrix composed of polymers is critical to structured sessile communities of bacterial cells called biofilms. In *B. subtilis*, this adhesive matrix is needed for assembly of both
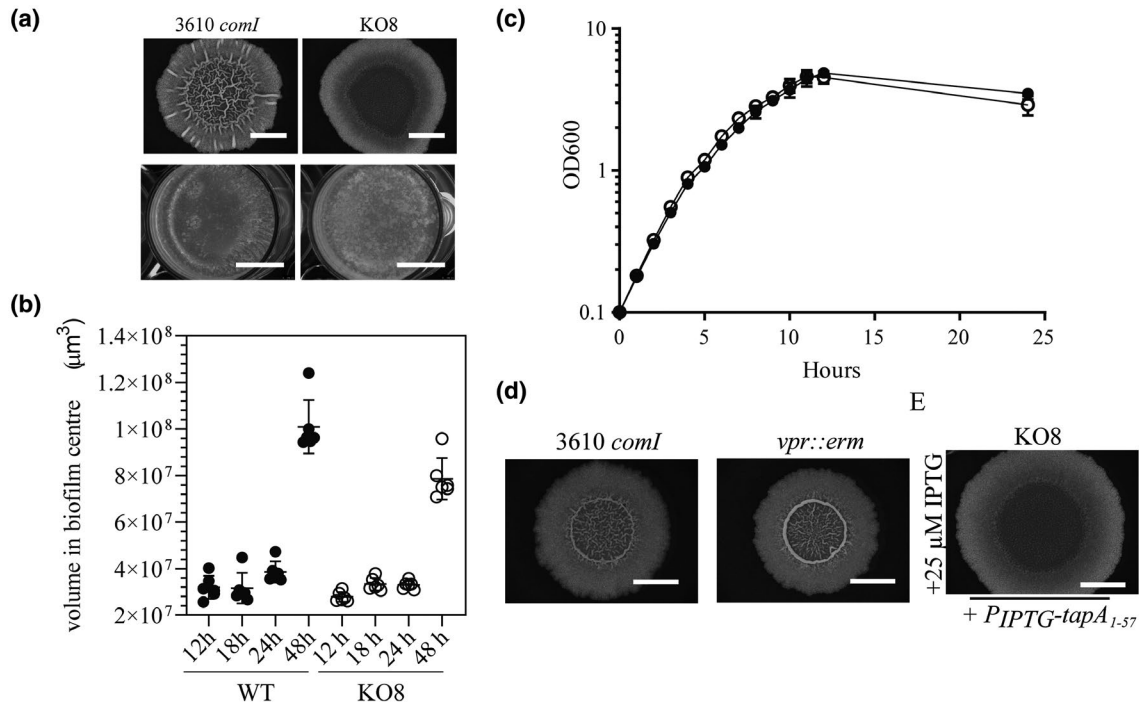
FIGURE 5 Cleavage of TapA is not needed for biofilm architecture. (a) Colony and pellicle biofilms formed by *B. subtilis* isolates 3610 *comI* (NRS6017) and KO8 (NRS5645); (b) Volume of features in the middle of colony biofilms of NCIB3610 *comI* constitutively expressing GFP (NRS5634), filled circles and NCIB3610 comI KO8 constitutively expressing GFP (NRS6991), empty circles. Images acquired at 12h, 18h, 24h and 48 hr of biofilm growth at 30°C. Shown are six individual points per strain per time point comprising two biological replicates with three technical repeats each. Lines represent mean and standard deviation; (c) Growth of NCIB3610 *comI* (closed circles) and NCIB3610 *comI* KO8 (NRS5645) (open circles) in MSgg medium with shaking at 30°C measured by OD_{600}. An average of two independent experiments are shown with the error bars being the standard deviation; (d) Biofilms formed by 3610 *comI* (NRS6017) and *vpr::erm* (NRS7010); (e) Biofilms of KO8 P_{IPTG}-*tapA*$_{1-57}$ (NRS6960). The scale bars represent 1 cm and the biofilms were incubated for 48 hr at 30°C

pellicle and colony biofilms in addition to the attachment to plant roots (Beauregard *et al.*, 2013). Biofilm matrix production by *B. subtilis* depends on the protein TapA, which has a role in promoting TasA stability and fibre formation in vivo (Branda *et al.*, 2006; Romero *et al.*, 2010; Romero *et al.*, 2014; Abbasi *et al.*, 2019; El Mammeri *et al.*, 2019). Here we show that orthologous *tapA* genes, originating from other *Bacillus* species, are functional in the WT *B. subtilis* NCIB 3610 strain. This is in agreement with previous work which found that *tapA* from *B. amyloliquefaciens* can substitute for the *B. subtilis tapA* coding region (Romero *et al.*, 2014). The *B. pumilis* TapA protein is 61 amino acids shorter at the C-terminus, compared with *B. subtilis* TapA, indicating that the extreme C-terminal amino acids are dispensable for the rugose biofilm architecture. Consistent with this, when a truncated form of the *B. subtilis tapA* open reading frame was used to encode a variant form of TapA that lacked the C-terminus, it was functional. Further investigations presented here determined that the minimal functional unit of TapA consists of amino acids TapA$_{1-57}$. This was unexpected as the region of TapA with the highest conservation amongst the orthologues (amino acids 74-190 inclusive) was not needed to restore rugose biofilms to the *tapA* deletion strain. We have also shown that the predicted TapA signal sequence, which consists of amino acids 1-43, can be replaced with the predicted TasA 28 amino acid signal sequence. This, therefore, leaves a functional

component of TapA of only 14 amino acids (amino acids 44-57 inclusive). Intriguingly, there are some shared sequence features between the N-terminal amino acids of the mature TapA and TasA proteins, but the significance of this shared similarity is unknown. Finally, we cannot rule out the possibility that amino acids 58-253 of TapA serve a distinct function beyond promoting rugose biofilm formation, but what role it may play is undefined.

The cleavage of TapA to lower molecular weight forms is conserved amongst other *B. subtilis* isolates and depends on self-produced extracellular proteases. Exactly how the minimal functional unit leads to a structured colony and pellicle biofilm remains unknown. Structural details for the region of TapA needed for biofilm architecture are limited, however, the crystal structure of the core of TapA$_{75-190}$ (PDB 6HQC) reveals a β-sandwich (Figure 3f) and recent NMR analysis has suggested that the C-terminus of TapA is a disordered protein, while the N-terminus is a structured domain, with the two domains interacting with lipid vesicles in a co-operative manner (Abbasi *et al.*, 2019). Collectively, our immunoblot and genetic data support the conclusion that TapA is cleaved at two positions. Once in the C-terminal region, yielding a ~18 kDa form and additionally in the N-terminal region by Vpr, a serine protease, yielding a ~16 kDa form likely to comprise the stable core of the protein (Figure 6). Our data also indicate that cleavage at the C-terminal region of TapA can either use a protease that remains active in the KO8 strain, albeit not
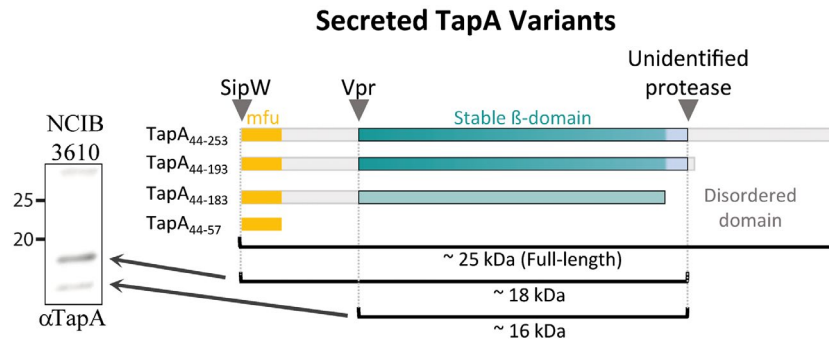
Secreted TapA Variants

FIGURE 6 Schematic of TapA processing in *B. subtilis*. The full-length secreted form of TapA is predicted to have a molecular weight of 24.18 kDa. The 18 kDa band observed by αTapA immunoblot is generated by the cleavage of the C-terminus either by an as yet unidentified protease or by a non-enzymatic process. Processing at the N-terminus by the exoprotease Vpr releases a ~16 kDa band leaving the structured, β-sandwich, core of the protein intact. The minimal functional unit (mfu) is shown in pale orange. The region for which there is structural information is shown in green. The approximate location of protease cutting sites are shown as triangles. The minimal functional form of the protein still conferred structure to biofilms, confirming that the bands detected by immunoblot do not represent the important functional unit of the protein in this context

as efficiently as seen in the WT strain (Figure 6), or that this cleavage can occur via a non-enzymatic reaction over an extended timeframe.

Vpr plays a specific role in the cleavage of TapA and is a serine protease that is part of the family of subtilisin-like proteases first noted using an unbiased screen seeking to identify extracellular proteases produced by *B. subtilis* (Sloma *et al.*, 1991). Production of Vpr is controlled at the level of transcription by the regulator activator CodY (Barbieri *et al.*, 2015), and repressor DnaA (Smith and Grossman, 2015). Transcription of *vpr* is also promoted under conditions of phosphate starvation (Allenby *et al.*, 2005). Vpr has not been specifically linked with growth or sporulation under nutrient-rich condition and is part of a family of proteins that are broadly classified as being used for nutrition acquisition (Rawlings *et al.*, 2010). However, Vpr is needed to produce quorum sensing signalling molecules, with a direct role in the generation of the five amino acid signal CSF from proCSF in the extracellular environment (Lanigan-Gerdes *et al.*, 2007). Therefore, our findings broaden the social biology behaviours exhibited by *B. subtilis* that Vpr is associated with.

The activity of extracellular proteases is linked with biofilm matrix formation in other species of bacteria. For example, the matrix protein RmbA of *Vibrio cholerae* specifically binds the biofilm exopolysaccharide in a manner dependent on its structural configuration which promotes biofilm formation (Fong *et al.*, 2017). However, RmbA is cleaved during biofilm formation by the HapA, PrtV and IvaP proteases (Berk *et al.*, 2012; Hatzios *et al.*, 2016), which releases a form of RmbA that allows for recruitment of both exopolysaccharide-producing and exopolysaccharide-nonproducing cells to the growing community (Smith *et al.*, 2015). The exact consequences for the processing of RmbA on biofilm formation in natural communities remains to be addressed. Based on this precedent, we speculated that TapA processing may generate an active variant of TapA and tested the impact of deleting the genes encoding the extracellular proteases on biofilm formation. We found that TapA processing to lower molecular weight forms was altered in the KO8 strain lacking Bpr, Vpr, NprB, Mpr, Epr, AprE, NprE and WprA. However, in

contradiction to our hypothesis, the colony biofilm architecture differences exhibited by the KO8 strain could not be mitigated by expression of the minimal functional coding region of *tapA*. We, therefore, conclude that the difference in KO8 colony biofilm architecture alteration is likely to be due to pleiotropic effects of deletion of the exoprotease genes, rather than being a specific impact of altered TapA processing. It is possible that the lack of extracellular proteases could impact the abundance of quorum sensing peptides in the extracellular environment. This is in line with evidence demonstrating that exoproteases control both the production and degradation of the quorum sensing signalling molecule ComX (Spacapan *et al.*, 2018). An accumulation of quorum sensing peptides in the biofilm microenvironment may have a global impact on the physiology of the KO8 strain due to an alteration in cell signalling pathways (Miller and Bassler, 2001; Kalamara *et al.*, 2018). The susceptibility of TapA to cleavage by the exoproteases raises the possibility that TapA is degraded (recycled) after it has fulfilled its role in biofilm formation. The alternative notion is that TapA plays a second role in *B. subtilis* physiology that is yet to be elucidated. This hypothesis is strengthened if you take into account that this is the conserved and stable core of the protein.

4 | CONCLUDING REMARKS

Through this analysis we have expanded our knowledge of the protein TapA which is needed for biofilm formation by *B. subtilis*. We have revealed that TapA$_{1-57}$ is a key component of the functional form of TapA in vivo, allowing rugose biofilm architecture to manifest. In contrast, the main body of the protein is entirely dispensable in this experimental set up, despite it containing the most conserved amino acid sequences. The exact mechanism by which TapA is needed to stimulate TasA stability in vivo requires more elucidation, but the amino acids that were found to be critical for function between amino acids 45-57 provide a starting point for analysis.

We have shown that TapA is processed in two distinct positions in vivo but that cleavage by the self-produced exoproteases is not an essential step to generate a functional protein form. We have also identified Vpr as a protease with a specific role in cleavage. Despite this, exactly how TapA functions to promote biofilm formation and the physiological role of the core of TapA remains to be elucidated.

5 | MATERIALS AND METHODS

5.1 | Growth media and additives

Lysogeny broth (LB) was prepared using 10 g of tryptone, 10 g of NaCl and 5 g of yeast extract for 1 litre. LB agar was prepared by solidifying with 15 g of select agar for growth of *B. subtilis* and *E. coli*. For antibiotic selection with *B. subtilis* strains antibiotics were used at the following final concentrations: erythromycin (0.5 μg/ml), spectinomycin 100 μg/ml and MLS, erythromycin (0.5 μg/ml) together with lincomycin (12.5 μg/ml). For antibiotic selection of plasmids in *E. coli* ampicillin was used at a concentration of 100 μg/ml.

To make Minimal Salts glycerol glutamate (MSgg) plates 5 mM potassium phosphate, 100 mM MOPS at pH 7.0 with agar to a final concentration of 1.5% (w/v) was autoclaved and cooled to 55°C before being supplemented with 2 mM $MgCl_2$, 700 μM $CaCl_2$, 50 μM $FeCl_3$, 50 μM $MnCl_2$, 1 μM $ZnCl_2$, 2 μM thiamine, 0.5% (v/v) glycerol and 0.5% (w/v) glutamic acid. For induction of gene expression from the P_{spank} promoter (P_{IPTG}) isopropyl β-D-1-thiogalactopyranoside (IPTG) was included at a final concentration of 25 μM. For MSgg broth the same recipe was followed without the addition of agar.

5.2 | Strain construction

Strains used in this study are detailed in Table S1. Complementation alleles and antibiotic resistance cassette marked gene deletions were moved between strains using either SSP1 mediated phage transduction (Verhamme *et al.*, 2007) or genetic competence with genomic DNA. The following strains were used for amplification of coding regions: *B. subtilis* strain NCIB3610 (GenBank Accession number: CP020102.1); *B. amyloliquefaciens* FZB42 (GenBank Accession number: CP000560.1); *B. paralicheniformis* (kindly provided by Dr. Nijland) and *B. pumilis* SAFR-032 (GenBank Accession number: CP000813.4).

The Δ*tapA* deletion strain was generated by allelic exchange in a method similar to that previously published using the pMAD plasmid (Arnaud *et al.*, 2004). Briefly, a 395 bp upstream region was amplified by PCR with primers NSW1308 and NSW1332 and a 641 bp downstream region was amplified using primers NSW1333 and NSW1334. Both fragments were cloned into the pMini-MAD vector (Patrick and Kearns, 2008) to generate plasmid pNW685.

To construct the in-frame deletions in all eight genes encoding the secreted proteases *B. subtilis* NCIB3610 *comI* (Patrick and Kearns, 2008), the BKE collection was utilised (Koo *et al.*, 2017) where single gene deletions have been replaced with a cassette providing resistance to erythromycin. Genomic DNA was extracted from strains in the BKE collection and used to transform competent *B. subtilis* 3610 *comI* (Konkol *et al.*, 2013) before selection on LB erythromycin plates. The erythromycin cassette contains *lox* sites and was subsequently removed leaving a 150 base pair scar by the action of a Cre recombinase, which was expressed on the heat-sensitive plasmid pDR244. In cases in which transformation with genomic DNA proved unsuccessful then the mutation was introduced by phage transduction with SPP1 phage. All the strains were examined using PCR and DNA sequencing to ensure the specificity in the region deleted from the chromosome. The intermediate strains are fully detailed in Table S1. The in-frame *tapA* deletion was introduced using plasmid pNW685 described above.

The variant *tapA* coding regions were introduced into *B. subtilis* chromosome at the *amyE* locus. Double recombination events were identified by assessing the production of α-amylase on LB growth medium supplemented with 1% (w/v) soluble starch.

5.3 | Plasmid construction

Plasmids (Table S2) were constructed using standard methods using the primers detailed in Table S3.

5.4 | Growth analysis

Lawn plates were set up by suspending a single colony in 100 μl of LB medium and plating the suspension onto LB agar and incubated overnight at 25°C. After ~24 hr the cells were washed from the plate in 5 ml of MSgg broth and the absorbance at 600 nm measured. The volume of cell suspension used to inoculate was calculated based on a desired starting OD_{600} of 0.1. Cultures were grown in 25 ml of MSgg broth in a 250 ml conical flask in a water bath set to 30°C with shaking at 200 rpm.

5.5 | In vivo analysis of exoprotease production

Detection of exoprotease production was conducted as previously described (Verhamme *et al.*, 2007) using LB agar plates supplemented with 1.5% (w/v) dried milk powder. Strains were grown in 3 ml of LB broth at 37°C to an OD_{600} of ~1. The cultures were normalised and 10 μl of the prepared cell culture spotted onto the plate. The samples were then grown for 20 hr at 37°C prior to photography.

5.6 | Biofilm growth and analysis

Biofilm colonies were prepared by growing *B. subtilis* in 3 ml of LB broth at 37°C with aeration for ~3.5 hr. After which, 10 μl of the culture was spotted onto an MSgg agar plate that was incubated at

30°C for 48 hr. Biofilm pellicles were prepared by growing *B. subtilis* in 3 ml of LB broth at 37°C with aeration for ~3.5 hr. After which, 4 µl of the culture were inoculated into 2 ml of MSgg liquid in a 24-well plate that was incubated at 25°C for 72 hr. Imaging used a Leica MZ16 stereoscope (Leica Microsystems). About 25 µM IPTG was included in the growth medium to induce expression from the P_{spank} promoter as indicated.

5.7 | Protein extraction from biofilms

Biofilms were isolated from MSgg plates with a sterile loop and suspended in 250 µl of BugBuster solution (Millipore) using a syringe with a 23 x 1-gauge needle until dispersed. The samples were sonicated at an amplitude of 20% power for 5 s. Sonicated biofilms were incubated at 26°C for 20 min with shaking at 1,400 rpm before centrifugation for 10 min in a benchtop centrifuge at 13,000 rpm. The liquid phase was retained for further analysis by SDS-PAGE and/or immunoblot. Protein concentration of biofilm lysates was determined by measuring absorbance at 280 nm (NanoDrop spectrophotometer) or using the DC protein assay (BIO-RAD) which is based on the Lowry assay (Lowry *et al.*, 1951).

5.8 | Protein purification

Recombinant *B. subtilis* $TapA_{44-253}$, mTasA (Erskine *et al.*, 2018b) and fTasA (Erskine *et al.*, 2018b) proteins were produced and separated from a Glutathione S-transferase-tag with a tobacco etch virus (TEV) protease-cleavage site using the pGEX-6P-1 system. The pGEX-6P-1 plasmid carrying the gene encoding the protein was introduced into *E. coli* strain BL21 (DE3) pLysS. The cells were grown overnight in 5 ml of LB broth and used to inoculate auto-induction media (Studier, 2005) supplemented with ampicillin (100 µg/ml) at a ratio of 1:1,000 (vol:vol) in a total volume of 1 litre. The cultures were incubated at 30°C with shaking for approximately 6-7 hr at which point the temperature was reduced to 18°C for overnight incubation. The cell culture was pelleted by centrifugation for 45 min at 5,020 g and re-suspended in 25 ml of purification buffer (Tris 25 mM and NaCl 250 mM [pH 7.6]) supplemented with Complete EDTA-free proteinase inhibitors mixture (Roche). Cell lysis was carried out by sonication at an amplitude of 20% for a total of 6 min. Unlysed cells and cell debris were removed by centrifugation at 27,000 × g for 20 min. The cleared lysate was mixed with 750 µl of (per litre of culture) Glutathione Sepharose 4B (GE Healthcare) and gently agitated at 4°C for at least 3 hr to allow binding of GST to the beads. The lysate/bead mixture was loaded onto a single-use, 25 ml gravity flow column (Bio-Rad). The beads were washed using 50 ml of purification buffer, collected and incubated overnight at 4°C with agitation in 25 ml of purification buffer supplemented with 1 mM DTT and 0.5 mg of TEV protease to release the protein from the GST-tag. The solution containing target protein, TEV protease, free GST and the beads was loaded onto the gravity flow column. This removed the used beads which stay on the column. The flow-through was added to 250 µl of Ni-nitrilotriacetic acid agarose (Qiagen) slurry to remove the TEV protease and 750 µl of glutathione sepharose 4B to remove the free GST-tag. The mixture was incubated at 4°C with agitation overnight, and then passed through a gravity flow column. The purified protein was concentrated using a Vivaspin 20 concentrator (with a MW cut-off of 5,000 or 10,000 Sartorius). The protein concentration was determined by measuring absorbance at 280 nm (NanoDrop spectrophotometer) and then analysed by separating ~30 µg by SDS-PAGE.

5.9 | Protein stability in culture supernatant

To collect spent culture supernatant the following process was used. Initially, lawn plates were set up by suspending a single colony in 100 µl of LB medium and plating the suspension onto LB agar. Lawn plates were incubated overnight at 25°C. After ~24 hr the cells were washed from the plate in 5 ml of MSgg broth and the absorbance at 600 nm measured. The volume of cell suspension used to inoculate was calculated based on a desired starting OD_{600} of 0.1. Cultures were grown in 25 ml of MSgg broth in a 250 ml conical flask in a water bath set to 30°C with shaking at 200 rpm. Early stationary phase cultures were harvested at an OD_{600} of approximately 4.0 and normalised. The harvested culture was then pelleted by centrifugation at 3,220 × g for 10 min and the full volume of supernatant was filter-sterilised to remove bacterial cells. The supernatant protease activity assay was set up by mixing supernatant 1:1 (vol:vol) with purified TapA, mTasA or fTasA protein (to give a total volume of 40 µl) and incubating the mixture at 37°C for 8 hr. This gave a protein concentration in the assay of 3 µg/µl, and a total recombinant protein amount of 120 µg. Heat-inactivation of the supernatant was carried out by incubating the supernatant samples at 100°C for 10 min prior to use. The integrity of the recombinant protein was then assessed by SDS-PAGE, stained with InstantBlue™ (Sigma-Aldrich), Coomassie based stain. About 28 µg of protein was loaded onto the gel, this was calculated based on the starting assay concentration of 3 µg/µl.

5.10 | Antibody production

A custom antibody that could be used to detect TapA from *B. subtilis* was raised in a rabbit using purified recombinant $TapA_{34-253}$ as the antigen (Eurogentec). The antibodies specific to recombinant $TapA_{34-253}$ were purified from the serum using standard methods by the MRC Protein reagents and services team (https://mrcppureagents.dundee.ac.uk/our-services/custom-antibody-production).

5.11 | Immunoblot analysis

Samples to be analysed by immunoblotting were separated by SDS-PAGE under denaturing conditions. The proteins were transferred

to hydrophobic polyvinylidene difluoride (PVDF) membranes (Immobilon-P [Millipore]) by electroblotting. The membranes were first blocked with a 5% (w/v) semi-skimmed dry milk solution in TBS-tween 0.2% (v/v) for at least 1 hr at room temperature or overnight at 4°C. After which the membrane was incubated with the primary antibody overnight at 4°C (dilution as follows: 1:5,000 αTapA, 1:25,000 αTasA). After incubation with the primary antibody, the membrane was washed three times with TBS-tween 0.2% (v/v) to the remove unbound primary antibody and incubated with the species-specific secondary HRP-conjugated antibody (Goat α Rabbit, dilution 1:5,000) for 1 hr at room temperature in TBS-tween 0.2% (v/v). The wash steps were repeated before development was induced with Enhanced Chemi-Luminescence reagents (ECL; BioRad Clarity). An electronic imaged was captured using the GeneGnome (SynGene) system.

5.12 | Bioinformatics analysis

Protein sequences of TapA homologues were aligned using Clustal Omega using default settings (Sievers *et al.*, 2011). Percentage identity between TapA orthologues was calculated with reference to *B. subtilis* TapA using the pairwise alignment function on the Jalview 2 workbench (Waterhouse *et al.*, 2009). For the prediction of signal peptides for all TapA variants then the SignalP 4.1 server was used and set to the organism group 'Gram-positive' (Petersen *et al.*, 2011). Secondary structure predictions were calculated using the Phyre2 (Kelley *et al.*, 2015), JPred (Drozdetskiy *et al.*, 2015; MacGowan *et al.*, 2020), and PSIPRED (Buchan and Jones, 2019) webservers.

The crystal structure of $TapA_{75-190}$ was retrieved from the Protein Data Bank (PDB 6HQC) (Roske *et al.*, 2018). The DALI server (Holm, 2019) was used for structural comparison to all structures in the Protein Data Bank. Structural visualisation was done using PyMOL 2.0 (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC).

5.13 | Confocal microscopy

For confocal microscopy, two biological replicates of WT and KO8 strains (NRS5634 and NRS6991 respectively) constitutively expressing the coding region for the Green Fluorescent Protein (GFP) were grown in liquid LB at 37°C and 200 rpm for approximately 4 hr. The cell density was then normalised to an OD_{600} of 0.9. From each of these biological repeats three technical repeats were prepared by depositing 1 μl of cells onto MSgg agar 1.5% (w/v) plates and incubated at 30°C. Imaging was performed on a Leica SP8 upright confocal with a 488 nm excitation laser at 2% power and a 10x 0.3NA dry objective, detecting emission wavelengths from 490 to 600 nm with the pinhole set to 1AU for 525 nm. A system-optimised z-step of 3.88 μm was used to gather a z-stack encompassing the structures of a field of view in the central region of each biofilm at 12, 18, 24 and 48 hr after the cells were initially deposited.

5.14 | Image analysis

Image data were stored in an OMERO server (Allan *et al.*, 2012) and analysed in Matlab 2016b. For each image, the z-stack was downloaded into Matlab using the OMERO.matlab toolkit (Allan *et al.*, 2012) and segmented using Otsu thresholding (Otsu, 1979). The xz orthogonal plane at each y position of the segmented image was then analysed for the 'top-most' segmented pixel in z, and these values were used to build up a single xy matrix of z positions representing the structure height at every pixel location. To calculate the volume of structures in the centre of biofilms at each time point the product of height matrix and the voxel size was taken. The code is available: https://github.com/mporter-gre/Calculation-of-Feature-Volumes-of-Biofilms. Data were categorised by strain and time point and represented as scatterplot drawn in GraphPad Prism 8.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID
Michael Porter https://orcid.org/0000-0001-9573-5066
Nicola R. Stanley-Wall https://orcid.org/0000-0002-5936-9721

REFERENCES
Abbasi, R., Mousa, R., Dekel, N., Amartely, H., Danieli, T., Lebendiker, M., *et al.* (2019) The bacterial extracellular matrix protein tapa is a two-domain partially disordered protein. *ChemBioChem*, 20, 355–359.

Allan, C., Burel, J.M., Moore, J., Blackburn, C., Linkert, M., Loynton, S., *et al.* (2012) OMERO: flexible, model-driven data management for experimental biology. *Nature Methods*, 9, 245–253.

Allenby, N.E., O'Connor, N., Pragai, Z., Ward, A.C., Wipat, A. and Harwood, C.R. (2005) Genome-wide transcriptional analysis of the phosphate starvation stimulon of *Bacillus subtilis*. *Journal of Bacteriology*, 187, 8063–8080.

Arnaud, M., Chastanet, A. and Débarbouillé, M. (2004) New vector for efficient allelic replacement in naturally nontransformable, low-GC-content, Gram-positive Bacteria. *Applied and Environmental Microbiology*, 70, 6887–6891.

Barbieri, G., Voigt, B., Albrecht, D., Hecker, M., Albertini, A.M., Sonenshein, A.L., *et al.* (2015) CodY regulates expression of the *Bacillus subtilis* extracellular proteases Vpr and Mpr. *Journal of Bacteriology*, 197, 1423–1432.

Beauregard, P.B., Chai, Y., Vlamakis, H., Losick, R. and Kolter, R. (2013) *Bacillus subtilis* biofilm induction by plant polysaccharides. *Proceedings of the National Academy of Sciences of the United States of America*, 110, E1621–E1630.

Berk, V., Fong, J.C., Dempsey, G.T., Develioglu, O.N., Zhuang, X., Liphardt, J., *et al.* (2012) Molecular architecture and assembly principles of *Vibrio cholerae* biofilms. *Science*, 337, 236–239.

Branda, S.S., Chu, F., Kearns, D.B., Losick, R. and Kolter, R. (2006) A major protein component of the *Bacillus subtilis* biofilm matrix. *Molecular Microbiology*, 59, 1229–1238.

Buchan, D.W.A. and Jones, D.T. (2019) The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Research*, 47, W402–W407.

Cairns, L.S., Hobley, L. and Stanley-Wall, N.R. (2014) Biofilm formation by *Bacillus subtilis*: new insights into regulatory strategies and assembly mechanisms. *Molecular Microbiology*, 93, 587–598.

Dragos, A. and Kovacs, A.T. (2017) The peculiar functions of the bacterial extracellular matrix. *Trends in Microbiology*, 25, 257–266.

Drozdetskiy, A., Cole, C., Procter, J. and Barton, G.J. (2015) JPred4: a protein secondary structure prediction server. *Nucleic Acids Research*, 43, W389–394.

El Mammeri, N., Hierrezuelo, J., Tolchard, J., Camara-Almiron, J., Caro-Astorga, J., Alvarez-Mena, A., *et al.* (2019) Molecular architecture of bacterial amyloids in *Bacillus* biofilms. *The FASEB Journal*, 33(11), 12146–12163.

Erskine, E., MacPhee, C.E. and Stanley-Wall, N.R. (2018a) Functional amyloid and other protein fibers in the biofilm matrix. *Journal of Molecular Biology*, 430, 3642–3656.

Erskine, E., Morris, R.J., Schor, M., Earl, C., Gillespie, R.M.C., Bromley, K.M., *et al.* (2018b) Formation of functional, non-amyloidogenic fibres by recombinant *Bacillus subtilis* TasA. *Molecular Microbiology*, 110, 897–913.

Flemming, H.C., Wingender, J., Szewzyk, U., Steinberg, P., Rice, S.A. and Kjelleberg, S. (2016) Biofilms: an emergent form of bacterial life. *Nature Reviews Microbiology*, 14, 563–575.

Fong, J.C., Rogers, A., Michael, A.K., Parsley, N.C., Cornell, W.C., Lin, Y.C., *et al.* (2017) Structural dynamics of RbmA governs plasticity of Vibrio cholerae biofilms. *eLife*, 6, e26163.

Garcia-Ferrer, I., Arede, P., Gomez-Blanco, J., Luque, D., Duquerroy, S., Caston, J.R., *et al.* (2015) Structural and functional insights into *Escherichia coli* alpha(2)-macroglobulin endopeptidase snap-trap inhibition. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 8290–8295.

Hatzios, S.K., Abel, S., Martell, J., Hubbard, T., Sasabe, J., Munera, D., *et al.* (2016) Chemoproteomic profiling of host and pathogen enzymes active in cholera. *Nature Chemical Biology*, 12, 268–274.

Hobley, L., Ostrowski, A., Rao, F.V., Bromley, K.M., Porter, M., Prescott, A.R., *et al.* (2013) BslA is a self-assembling bacterial hydrophobin that coats the *Bacillus subtilis* biofilm. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 13600–13605.

Holm, L. (2019) Benchmarking fold detection by DaliLite v.5. *Bioinformatics*, 35, 5326–5327.

Kalamara, M., Spacapan, M., Mandic-Mulec, I. and Stanley-Wall, N.R. (2018) Social behaviours by *Bacillus subtilis*: quorum sensing, kin discrimination and beyond. *Molecular Microbiology*, 110, 863–878.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. and Sternberg, M.J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10, 845–858.

Kobayashi, K. and Iwano, M. (2012) BslA (YuaB) forms a hydrophobic layer on the surface of *Bacillus subtilis* biofilms. *Molecular Microbiology*, 85, 51–66.

Konkol, M.A., Blair, K.M. and Kearns, D.B. (2013) Plasmid-encoded ComI inhibits competence in the ancestral 3610 strain of *Bacillus subtilis*. *Journal of Bacteriology*, 195, 4085–4093.

Koo, B.M., Kritikos, G., Farrelli, J.D., Todor, H., Tong, K., Kimsey, H., *et al.* (2017) Construction and analysis of two genome-scale deletion libraries for *Bacillus subtilis*. *Cell systems*, 4, 291–305.e297.

Lanigan-Gerdes, S., Dooley, A.N., Faull, K.F. and Lazazzera, B.A. (2007) Identification of subtilisin, Epr and Vpr as enzymes that produce CSF, an extracellular signalling peptide of *Bacillus subtilis*. *Molecular Microbiology*, 65, 1321–1333.

Lowry, O.H., Rosebrough, N.J., Farr, A.L. and Randall, R.J. (1951) Protein measurement with the Folin phenol reagent. *Journal of Biological Chemistry*, 193, 265–275.

MacGowan, S.A., Madeira, F., Britto-Borges, T., Warowny, M., Drozdetskiy, A., Procter, J.B., *et al.* (2020) The dundee resource for sequence analysis and structure prediction. *Protein Science*, 29, 277–297.

Miller, M.B. and Bassler, B.L. (2001) Quorum sensing in bacteria. *Annual Review of Microbiology*, 55, 165–199.

Msadek, T., Kunst, F., Henner, D., Klier, A., Rapoport, G. and Dedonder, R. (1990) Signal transduction pathway controlling synthesis of a class of degradative enzymes in *Bacillus subtilis*: expression of the regulatory genes and analysis of mutations in *degS* and *degU*. *Journal of Bacteriology*, 172, 824–834.

Otsu, N. (1979) Threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9, 62–66.

Patrick, J.E. and Kearns, D.B. (2008) MinJ (YvjD) is a topological determinant of cell division in *Bacillus subtilis*. *Molecular Microbiology*, 70, 1166–1179.

Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, 8, 785–786.

Rawlings, N.D., Barrett, A.J. and Bateman, A. (2010) MEROPS: the peptidase database. *Nucleic Acids Research*, 38, D227–233.

Romero, D., Aguilar, C., Losick, R. and Kolter, R. (2010) Amyloid fibers provide structural integrity to *Bacillus subtilis* biofilms. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 2230–2234.

Romero, D., Vlamakis, H., Losick, R. and Kolter, R. (2011) An accessory protein required for anchoring and assembly of amyloid fibres in *B. subtilis* biofilms. *Molecular Microbiology*, 80, 1155–1168.

Romero, D., Vlamakis, H., Losick, R. and Kolter, R. (2014) Functional analysis of the accessory protein TapA in *Bacillus subtilis* amyloid fiber assembly. *Journal of Bacteriology*, 196, 1505–1513.

Roske, Y., Heinemann, U., Oschkinat, H. and Schmieder, P. (2018) Structural investigation of the TasA anchoring protein TapA from Bacillus subtilis.https://doi.org/10.2210/pdb6HQC/pdb

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Molecular Systems Biology*, 7, 539.

Sloma, A., Rufo, G.A. Jr, Theriault, K.A., Dwywe, M., Wilson, S.W. and Pero, J. (1991) Cloning and characterisation of the gene for an additional extracellular serine protease of *Bacillus subtilis*. *Journal of Bacteriology*, 173, 6889–6895.

Smith, D.R., Maestre-Reyna, M., Lee, G., Gerard, H., Wang, A.H. and Watnick, P.I. (2015) In situ proteolysis of the *Vibrio cholerae* matrix protein RbmA promotes biofilm recruitment. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 10491–10496.

Smith, J.L. and Grossman, A.D. (2015) In vitro whole genome DNA binding analysis of the bacterial replication initiator and transcription factor DnaA. *PLoS Genetics*, 11, e1005258.

Spacapan, M., Danevcic, T. and Mandic-Mulec, I. (2018) ComX-induced exoproteases degrade ComX in *Bacillus subtilis* PS-216. *Frontiers in microbiology*, 9, 105.

Stover, A.G. and Driks, A. (1999a) Control of synthesis and secretion of the *Bacillus subtilis* protein YqxM. *Journal of Bacteriology*, *181*, 7065–7069.

Stover, A.G. and Driks, A. (1999b) Secretion. localization and antibacterial activity of TasA, a *Bacillus subtilis* spore-associated protein. *Journal of Bacteriology*, *181*, 1664–1672.

Studier, F.W. (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, *41*, 207–234.

Terra, R., Stanley-Wall, N.R., Cao, G. and Lazazzera, B.A. (2012) Identification of *Bacillus subtilis* SipW as a bifunctional signal peptidase that controls surface-adhered biofilm formation. *Journal of Bacteriology*, *194*, 2781–2790.

Verhamme, D.T., Kiley, T.B. and Stanley-Wall, N.R. (2007) DegU coordinates multicellular behaviour exhibited by *Bacillus subtilis*. *Molecular Microbiology*, *65*, 554–568.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. and Barton, G.J. (2009) Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*, 1189–1191.

Wong, S.G. and Dessen, A. (2014) Structure of a bacterial alpha(2)-macroglobulin reveals mimicry of eukaryotic innate immunity. *Nature Communications*, *5*, 4917.

Zhu, B. and Stülke, J. (2018) SubtiWiki in 2018: from genes and proteins to functional network annotation of the model organism *Bacillus subtilis*. *Nucleic Acids Research*, *46*, D743–D748.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.