

ARTICLE OPEN



Molecular and cellular insight into *Escherichia coli* SsIE and its role during biofilm maturation

Paula M. Corsini^{1,2,9}, Sunjun Wang^{1,9}, Saima Rehman^{1,9}, Katherine Fenn^{2,9}, Amin Sagar³, Slobodan Sirovica⁴, Leanne Cleaver¹, Charlotte J. C. Edwards-Gayle⁵, Giulia Mastroianni², Ben Dorgan¹, Lee M. Sewell¹, Steven Lynham⁶, Dinu Iuga⁷, W. Trent Franks^{1,7}, James Jarvis⁸, Guy H. Carpenter¹, Michael A. Curtis¹, Pau Bernadó³, Vidya C. Darbari^{1,2}✉ and James A. Garnett^{1,2}✉

Escherichia coli is a Gram-negative bacterium that colonises the human intestine and virulent strains can cause severe diarrhoeal and extraintestinal diseases. The protein SsIE is secreted by a range of pathogenic and commensal *E. coli* strains. It can degrade mucins in the intestine, promotes biofilm maturation and it is a major determinant of infection in virulent strains, although how it carries out these functions is not well understood. Here, we examine SsIE from the commensal *E. coli* Waksman and BL21 (DE3) strains and the enterotoxigenic H10407 and enteropathogenic E2348/69 strains. We reveal that SsIE has a unique and dynamic structure in solution and in response to acidification within mature biofilms it can form a unique aggregate with amyloid-like properties. Furthermore, we show that both SsIE monomers and aggregates bind DNA in vitro and co-localise with extracellular DNA (eDNA) in mature biofilms, and SsIE aggregates may also associate with cellulose under certain conditions. Our results suggest that interactions between SsIE and eDNA are important for biofilm maturation in many *E. coli* strains and SsIE may also be a factor that drives biofilm formation in other SsIE-secreting bacteria.

npj Biofilms and Microbiomes (2022)8:9; <https://doi.org/10.1038/s41522-022-00272-5>

INTRODUCTION

Escherichia coli is a primary coloniser of the lower intestinal tract of humans and other warm-blooded animals. While many strains are considered beneficial to the host and help to maintain a healthy immune system, virulent strains are the cause of severe diarrhoeal diseases, including haemorrhagic colitis, and extraintestinal diseases, such as neonatal meningitis, urinary tract infections, sepsis and pneumonia¹. A wide range of pathogenic *E. coli* strains, and some commensals, use a *Vibrio*-like type II secretion system (T2SS) to translocate the protein SsIE across their outer membrane and onto their extracellular surface^{2–7}. These include the Waksman (W), enterotoxigenic (ETEC), enterohemorrhagic (EHEC), enteropathogenic (EPEC), enteroaggregative (EAEC), enteroinvasive (EIEC) and neonatal meningitis *E. coli* (NMEC) strains. SsIE is required for full virulence in a rabbit model of EPEC infection⁴ and as a surface-exposed antigen, SsIE has shown great promise as a broadly protective vaccine candidate against a wide range of *E. coli* pathotypes^{3,6,8,9}.

SsIE interacts with mucosal membranes in the host intestine where it can degrade mucins^{6,10–12}, a family of heavily O-linked glycosylated proteins and the primary constituents of mucus¹³. This provides nutrients during bacterial growth but also facilitates the penetration of pathogenic *E. coli* strains through the gut mucosa to access host cells for efficient colonisation and targeting of toxins/effectors. Furthermore, SsIE is important for mediating the maturation of EPEC biofilms⁴; microbial aggregations encased within a self-produced extracellular matrix comprised of exopolysaccharides, adhesive proteins and nucleic acids¹⁴. When *E. coli* is released into the environment through faeces or wastewater

effluent it can survive for long periods within complex biofilm communities^{15,16}, and these are fundamental for both the environmental ecology of *E. coli* but also for successful colonisation of the intestinal tract¹⁷. However, the specific molecular mechanisms that SsIE uses to promote ecology and/or disease are not well understood.

SsIE is a ~165 kDa lipoprotein composed of an N-terminal periplasmic localisation sequence and lipobox motif, an unstructured ~5 kDa region, a ~110 kDa region with no significant primary sequence homology to any other known protein, and a ~50 kDa M60-like aminopeptidase domain at its C-terminus¹⁸ (Fig. 1a). M60-like domains are metalloproteases that contain a zinc-binding HExxH motif and an additional conserved catalytic glutamate residue, which cleaves the peptide backbone of mucin-like substrates. These and other related enzymes have been identified in both prokaryotic and eukaryotic microbes that interact with host mucosal membranes¹⁸ and structures of proteoglycan complexes suggest that interactions with both the mucin peptide and O-linked carbohydrate side chains are important for specificity^{19,20}. The peptidase activity of the SsIE M60 domain is effective for the degradation of major mucins of the intestine (e.g. MUC2, MUC3, MUC5AC)^{6,10–12}, however, very little is known as to how SsIE interacts with these mucins or the function of the remaining SsIE sequence.

In this study, we provide new insight into the global structure of SsIE and propose a molecular mechanism by which SsIE can promote biofilm maturation. Using transmission electron microscopy (TEM), small-angle X-ray scattering (SAXS), nuclear magnetic resonance (NMR) spectroscopy and biochemical analyses we show

¹Centre for Host-Microbiome Interactions, Faculty of Dental, Oral & Craniofacial Sciences, King's College London, London, UK. ²School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK. ³Centre de Biologie Structurale, Université de Montpellier, INSERM, CNRS, Montpellier, France. ⁴Centre for Oral Bioengineering, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. ⁵Diamond Light Source Ltd., Diamond House, Harwell Science and Innovation Campus, Didcot, Oxfordshire, UK. ⁶Proteomics Facility, Centre of Excellence for Mass Spectrometry, King's College London, London, UK. ⁷Department of Physics, University of Warwick, Coventry, UK. ⁸Randall Division of Cell and Molecular Biophysics and Centre for Biomolecular Spectroscopy, King's College London, London, UK. ⁹These authors contributed equally: Paula M. Corsini, Sunjun Wang, Saima Rehman, Katherine Fenn. ✉email: v.darbari@qmul.ac.uk; james.garnett@kcl.ac.uk

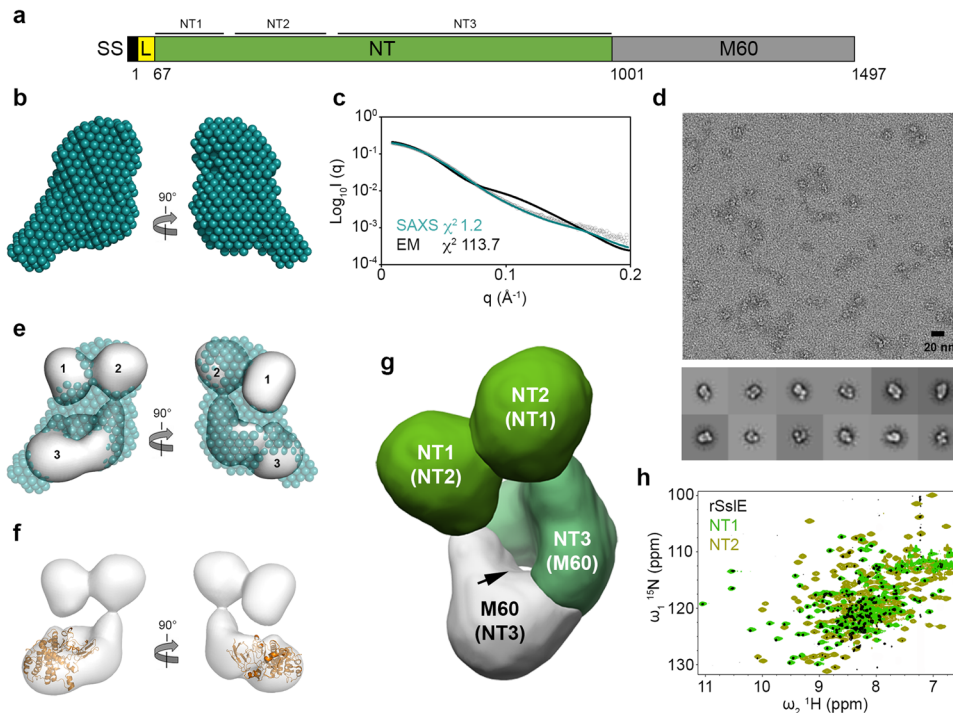


Fig. 1 Structural features of monomeric SsIE. **a** Schematic of SsIE from *E. coli* W with mature sequence numbers and structural features annotated. SS: periplasmic signal sequence; L: flexible linker; NT: unique N-terminal region; M60: peptidase/mucinase domain. **b** SAXS bead model of rSsIE at pH 7.4. **c** Fit of the SAXS bead model (teal line) and negative-stain TEM map (black line) of rSsIE to the rSsIE SAXS data (black open circles) with χ^2 of 1.2 and 113.7, respectively. **d** Representative negative-stain TEM micrograph of rSsIE (pH 7.4) at $\times 50,000$ nominal magnification with representative 2D classifications. Scale bar represents 20 nm. **e** Overlay of the 22 Å resolution TEM map (grey) and SAXS bead model (teal) with the three defined regions in SsIE highlighted. **f** Docking of an SsIE M60 domain homology model (orange) into the TEM map. **g** TEM map of rSsIE coloured based on domain organisation. Possible alternative assignments of the NT1 and NT2 domains in regions 1 and 2, and the NT3 and M60 domains in region 3 are shown in parentheses. The NT3-M60 interdomain channel is highlighted with an arrow. **h** TROSY $^1\text{H}^{15}\text{N}$ -HSQC spectrum of rSsIE (black) overlaid with $^1\text{H}^{15}\text{N}$ -HSQC spectra of the SsIE NT1 (light green) and NT2 (olive green) domains.

that SsIE is formed of three defined regions (NT1, NT2 and NT3-M60), and is dynamic in solution. We also demonstrate that SsIE undergoes conformational changes under acidic conditions, and this leads to the formation of higher-order structures with amyloid-like properties. We directly observe acidification within mature *E. coli* W biofilms and determine that both monomeric and aggregated forms of SsIE can bind extracellular DNA (eDNA) in *E. coli* W biofilms, and under some conditions SsIE aggregates may also associate with extracellular cellulose. We propose that SsIE acts like a hub and provides a cross-linking function in biofilms, and this is likely a general mechanism for SsIE in all SsIE-secreting bacteria.

RESULTS

Overall structural features of SsIE

To understand the role of SsIE in *E. coli* biofilms, we first initiated a structural analysis of recombinant SsIE from *E. coli* W, minus the N-terminal lipidation motif and adjacent disordered region (rSsIE; residues 67 to 1497; UniProt ID E0IW31) (Fig. 1a). This was produced in *E. coli* BL21 and purified by nickel-affinity and size-exclusion chromatography. Small-angle X-ray scattering (SAXS) coupled with size-exclusion chromatography (SEC-SAXS) was then used to get shape information for rSsIE at pH 7.4. Guinier analysis provided a radius of gyration (R_g) of 4.03 ± 0.05 nm and examination of the distance distribution function ($P(r)$) suggested a maximum particle dimension (D_{max}) of 14.1 nm (Supplementary Fig. 1 and Supplementary Table 1). Calculation of the particle molecular mass (151 kDa) was consistent with a monomer in solution (theoretical mass 160 kDa). Ab initio dummy residue reconstruction generated 20 reproducible models with an average

normalised spatial discrepancy (NSD) score between reconstructions of 0.56 and a χ^2 fit between calculated and experimental solution scattering of 1.2 (Fig. 1b, c). The averaged low-resolution bead model suggested that rSsIE has a slightly elongated structure, however, examination of the Kratky plot indicated that rSsIE is a dynamic particle (Supplementary Fig. 1) and no obvious domain features could be assigned.

Negative-stain TEM data were next collected for rSsIE at pH 7.4 and we performed single-particle analysis to generate an initial low-resolution structure, with overall dimensions of $\sim 8 \times 8 \times 11.5$ nm (Fig. 1d and Supplementary Fig. 2). Here, three well-defined regions could be identified: two adjacent globular domains (regions 1 and 2) with approximate diameters of 4 nm and 4.5 nm, respectively, and a torus-shaped structure (region 3) with a ~ 15 Å wide central channel and approximate overall dimensions of $4.5 \times 6.5 \times 9$ nm (Fig. 1e). The superposition of the SAXS bead model with the TEM envelope showed some disagreement, with an apparent reduction of bead volume in all three regions of the SAXS model. Although this observation suggests the presence of internal dynamics in rSsIE, in line with its Kratky plot, the low-resolution nature of SAXS data does not allow discarding a different origin. Rigid body docking of a homology model of the M60 domain, based on the IMPa M60 metalloprotease from *Pseudomonas aeruginosa* (PDB ID code 5kdv)¹⁹, into the TEM envelope indicated that the C-terminus of SsIE is located within region 3, most likely at the base (Fig. 1f), and would not fit the density within regions 1 and 2. This suggested that the remaining density in region 3 is composed of the SsIE sequence directly N-terminal to the M60 domain and these two regions pack against one another to form a stable structure around an interdomain channel. Furthermore, from this model the M60

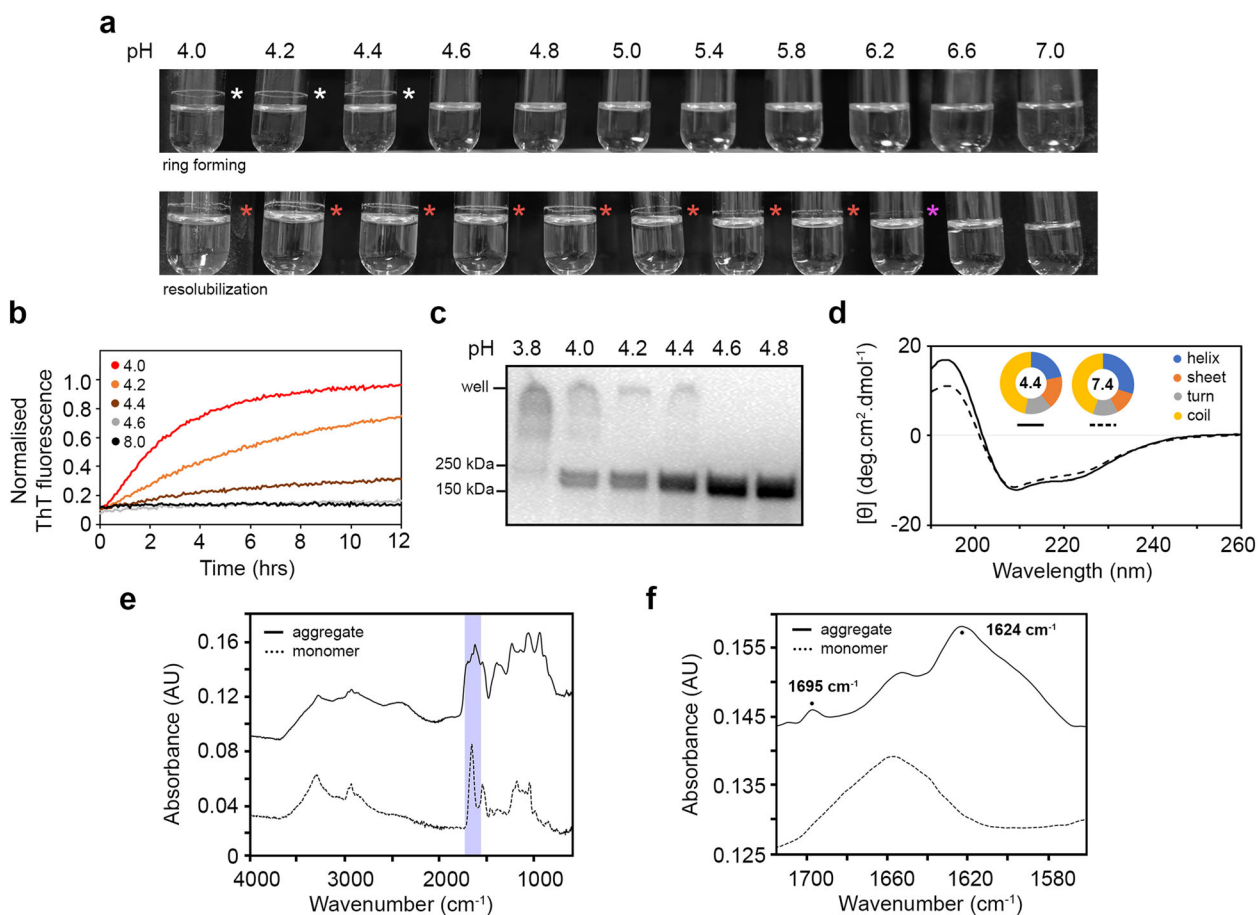


Fig. 2 **Characterisation of SsIE aggregates.** **a** Purified rSsIE forms a clear ring of aggregated protein when incubated in the citrate-phosphate buffer between pH 4.0 and 4.4 (upper panel; white asterisks). The ring formed above the liquid line is due to shaking of the sample and it resists solubilisation up to pH 6.2 once formed (lower panel; purple asterisk), although some loss of protein band is observed after pH 5.8 (lower panel; red asterisks). **b** Increase in ThT fluorescence emission on binding to rSsIE aggregates formed in citrate-phosphate buffer over pH 4.0–8.0. **c** Immunoblot of rSsIE incubated in citrate-phosphate buffer from pH 3.8 to 4.8 and detected with anti-His₆ antibody. rSsIE present in the well represents an aggregated protein. Blots were derived from the same experiment and processed in parallel. **d** Far-UV CD spectra of soluble rSsIE at pH 4.4 (solid line) and 7.4 (dashed line). Secondary structure analysis is shown in the doughnut chart. **e** ATR FT-MIR spectrum of monomeric rSsIE and aggregates with the amide I region shaded in violet. **f** Expanded ATR FT-MIR amide I region displaying the relatively high and low-intensity absorption bands in the aggregates but not monomeric spectrum at ~1624 and 1695 cm⁻¹, respectively, characteristic of an anti-parallel β -sheet structure. Spectra are offset in the absorbance axis for clarity.

HEXh active site motif faces into the channel, and this could be an important feature for its mucolytic activity.

Secondary structure analysis²¹ of the SsIE sequence indicated two potential interdomain boundaries within the N-terminal region and we predicted that these represented either region 1 or 2, located at the extreme N-terminus (NT) of SsIE. We therefore renamed these as the NT1 (residues 67–211) and NT2 (residues 230–425) domains, respectively, and renamed the remaining sequence as NT3 (residues 426–1000) (Fig. 1g). The NT1 and NT2 domains, along with the NT1-NT2 region (residues 67–425) and NT3-M60 core (region 3; residues 426–1497) were then produced in *E. coli* BL21 and purified by nickel-affinity and size-exclusion chromatography. Examination of these constructs using analytical gel filtration, nuclear magnetic resonance (NMR) spectroscopy and SEC-SAXS showed that they were well folded (Supplementary Figs. 3 and 4), and supported NT1, NT2 and NT3-M60 being defined structural boundaries within SsIE.

We then examined intact rSsIE by NMR, but only ~250 strong resonances were observed in a ¹H¹⁵N transverse relaxation optimised spectroscopy (TROSY) HSQC spectrum, with the remaining ~1200 peaks being either very weak or completely absent (Fig. 1h). However, NMR relaxation measurements of these intense peaks provided an estimated correlation time consistent

with an ~80 kDa domain ($\tau_c \sim 35$ ns at 37 °C). Comparison of rSsIE with ¹H¹⁵N-HSQC spectra from SsIE sub-domains also clearly showed that the rSsIE NMR spectrum is dominated by the NT1 domain but with minor contributions from the NT2 domain (Fig. 1h and Supplementary Fig. 5). This implied that the resonances from the NT1 and NT2 domains originated from intact rSsIE and further supported that this region has a significant independent motion with respect to the larger NT3-M60 core.

SsIE forms aggregates with amyloid-like properties

Functional amyloids are important proteinaceous structures in biofilms²² and analysis of the SsIE primary sequence had previously highlighted several regions across the protein that may form amyloid-like structures²³. Furthermore, environmental pH plays a major role in initiating the formation of amyloid-like structures in *Staphylococcus aureus*, *Enterococcus faecalis* and *Streptococcus mutans* biofilms^{24–26}, and we speculated that pH may also stimulate similar structural changes in SsIE. We therefore incubated rSsIE across the pH range 4.0–7.0 and observed a ring of protein deposited on the walls of the tube at pH \leq 4.4, which resisted solubilisation up to pH 6.2 (Fig. 2a). Examination of the ring by SDS-PAGE revealed a major high molecular weight species

that was retained in the well (Supplementary Fig. 6). Subsequent analysis of this by mass spectrometry identified peptides that spanned the complete sequence of rSsIE and indicated that the aggregate contained intact protein, rather than a degradation product (Supplementary Fig. 6). We next used the amyloid diagnostic dye Thioflavin-T (ThT) to assess the formation of rSsIE aggregates in solution over time. We observed large increases in ThT fluorescence emission again across the pH range 4.0–4.4, but no change was detected when rSsIE was incubated above pH 4.4 (Fig. 2b). Likewise, when rSsIE was boiled and treated with sodium dodecyl sulphate (SDS) and then analysed by anti-His₆ immunoblotting, stable high molecular weight species were observed over the same pH range (Fig. 2c). However, we also observed two distinct bands of monomeric SsIE, which suggested that it had been proteolytically cleaved at its N-terminus. Together these data indicated that SsIE may form amyloid-like structures at low pH and so this was further scrutinised with additional biophysical analysis.

A main feature of amyloids is the presence of a cross- β -sheet structure²⁷ and so we first analysed the secondary structure of soluble rSsIE using far-UV circular dichroism (CD) spectroscopy. This suggested that at pH 7.4 rSsIE is composed of ~30% α -helix, 12% β -sheet and 45% coil, while under acidic conditions and prior to aggregation, there is an ~8% loss of helical secondary structure with a ~5% increase in β -sheet structure (Fig. 2d and Supplementary Table 2). We next used attenuated total reflectance (ATR) Fourier transform mid-infrared (FT-MIR) spectroscopy to probe the secondary structure composition of soluble rSsIE at pH 7.4 and rSsIE fibres at pH 4.4. Examination of the amide I absorption band showed peaks at ~1656 cm⁻¹ and 1624 cm⁻¹, respectively (Fig. 2e). The shift in the amide I absorption band to a lower frequency indicated that a larger and more rigid structure had assembled under acid conditions. Furthermore, the presence of both a relatively high-intensity absorption band at ~1624 cm⁻¹ and a lower intensity band at ~1695 cm⁻¹ in the pH 4.4 sample was indicative of an increase in anti-parallel β -sheet structure in rSsIE aggregates²⁸ and is consistent with previous FT-MIR based characterisations of anti-parallel β -sheet amyloid fibres²⁹ (Fig. 2f).

SsIE aggregates display a fibrillar morphology

Congo red is a dye that can bind to amyloids and extracellular polymeric substances (EPS) in biofilms, including curli and cellulose. When we grew *E. coli* W strain overnight in liquid culture at pH 5.0, we observed a clear ring on the wall of the tube that retained Congo red, but this ring was less apparent when grown at pH 7.0 (Fig. 3a). Examination of this ring by SDS-PAGE again revealed a large molecular weight species retained in the wells and mass spectrometry analysis identified intact SsIE as the predominant protein (Supplementary Fig. 6). We then created an isogenic Δ sIE knockout mutant in *E. coli* W, but this strain was not able to form a substantial ring at either pH (Fig. 3a). However, complementation with a plasmid containing intact sIE (Δ sIE::sIE) or SsIE with a truncated C-terminus (Δ sIE::ssIE Δ M60; residues 1–1000) restored both SsIE secretion and ring formation at pH 5.0 (Fig. 3a and Supplementary Fig. 7).

Examination of wild-type *E. coli* W at pH 5.0 by immunoelectron microscopy revealed gold-labelled polyclonal anti-rSsIE antibodies associated with the bacterial outer membrane, which was not seen in the Δ sIE mutant (Fig. 3b, c). The fibrous material was also observed in the wild-type strain in the extracellular milieu and emanating away from the bacterial surface (Fig. 3c, d). Although antibody binding to these aggregates was less dramatic, the number of these single and clustered gold particles (measured up to 1.0 μ m from each bacterium) was significantly higher in the wild-type strain compared with the Δ sIE mutant (Fig. 3b–e and Supplementary Fig. 8). Inspection of the *E. coli* W Δ sIE::sIE complemented strain under the same conditions produced less clustering of anti-rSsIE antibodies, however, a wild-type

phenotype was observed when cells were incubated at pH 4.0 prior to fixation and staining (Fig. 3f, g). As a control, wild-type bacteria were also incubated with anti-CsgA (curli fibre) antibodies and there was no accumulation of gold particles on or away from the bacterial surface (Supplementary Fig. 9). Although these data indicated that SsIE may form fibres upon secretion from *E. coli* under acidic conditions, we were not able to differentiate between an accumulation of antibodies with aggregates and surface-associated or secreted monomeric SsIE.

Negative-stain TEM was then used to visualise purified rSsIE fibres and this revealed two common morphologies (Fig. 3h). The first form appeared as short single, flexible fibres with a core structure measuring ~4.5 nm wide by ~110 nm in length and decorated with globular structures ~4.5 nm in diameter (Fig. 3i, j). The second morphology again resembled a fibrous material but measured between ~20 and 40 nm in width by 200–300 nm in length (Fig. 3k) and appeared to be an aggregation of the smaller fibres. Further analysis of rSsIE aggregates in solution by real-time multi-angle light scattering (RT-MALS) supported these observations (Supplementary Fig. 10). We measured a smaller species with an average molecular mass of 5.4×10^3 kDa ($\pm 0.4 \times 10^3$) and an average R_g of 137.5 nm (± 1.7), and a larger more polydisperse species with an average molecular mass of 12×10^3 kDa ($\pm 0.8 \times 10^3$) and an average R_g of 139.8 nm (± 1.5). Examination of rSsIE fibres using solid-state NMR (ssNMR) showed individual narrow signals which confirmed that this was not an artefactual aggregation (Supplementary Fig. 11). We assigned ~40 ordered residues to specific residue types and this pattern suggested that these ordered regions were localised to more than one site across the SsIE sequence.

Kinetics of SsIE aggregation

We took advantage of SEC-SAXS to probe the global shape of rSsIE at pH 4.4 prior to fibre formation and both Guinier analysis (R_g 3.92 ± 0.05 nm) and examination of the distance distribution function ($P(r)$) (D_{max} 13.7 nm) produced very similar values to rSsIE at pH 7.4 (Supplementary Fig. 12 and Supplementary Table 1). However, the presence of better-defined features in the SAXS profile at acidic pH and the differences at larger q-values in the normalised Kratky plots were indicative of rSsIE becoming more rigid under acidic conditions (Fig. 4a and Supplementary Fig. 13). Ab initio dummy residue reconstruction produced 20 models with an average NSD score of 0.45 and a χ^2 fit between calculated and experimental solution scattering of 1.0 (Fig. 4b and Supplementary Fig. 14). Comparison of the averaged bead models indicated an increase in bead volume within the NT1 and NT2 domains and adjacent NT3 region at pH 4.4, while little change was observed in the NT3-M60 core, which supported an increase in rigidity under acidic conditions that could trigger aggregation.

We then exploited the additive nature of SAXS along with its ability to examine structures over a large size range to study the kinetics of fibre formation by rSsIE. We acquired SAXS data at pH 4.4 during the course of aggregation for 11 h and the resulting 23 profiles displayed a systematic upwards intensity increase, indicating the presence of large aggregates in solution (Fig. 4c). The appearance of these large species occurred relatively quickly as the upwards intensity increase was already observed after 1 h and the increase of aggregated species was concomitant to the loss of the SAXS features linked to the monomeric species. Principal Component Analysis (PCA) indicated that the dataset could be described as a two-component system (Supplementary Fig. 15) and we decomposed the time-dependent dataset along with the SAXS profile of the monomer, obtained using SEC-SAXS, with COSMICs. COSMICs uses a chemometric approach to decompose SAXS data sets in a model-free manner to derive the pure SAXS curves of the co-existing species and their relative populations, which we used here to report on the fibrillation

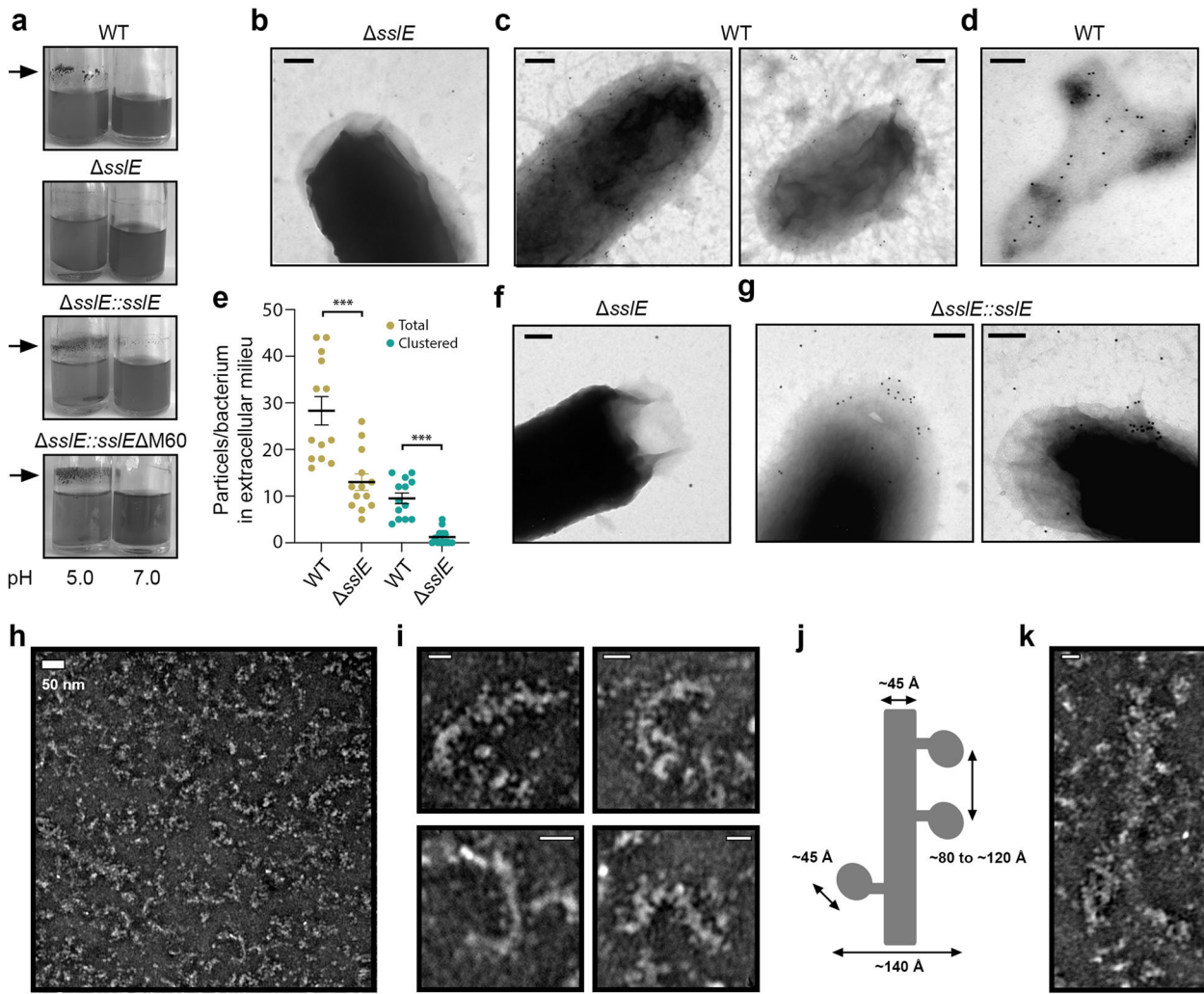


Fig. 3 TEM analysis of *E. coli* W SsIE aggregations. **a** Accumulation of SsIE aggregates on the side of glass tubes containing overnight cultures grown in LB media at pH 5.0 or 7.0. Arrows indicate Congo red-stained ring containing SsIE. **b–g** SsIE localisation upon secretion from *E. coli* W strains assessed by immunoelectron microscopy. Bacteria were reacted with primary α -rSsIE antibodies and secondary gold-labelled antibodies on carbon-coated nickel grids and then negatively stained with uranyl acetate. **b, f** *E. coli* W Δ ssIE mutant washed in citrate-phosphate buffer at pH 5.0 with α -rSsIE antibodies reacting with the bacterial surface and secreted fibrous material. **d** Image highlighting clustering of antibodies to an extracellular aggregate. **e** Analysis of total and clustered (≤ 25 nm between particles) α -rSsIE antibodies reacting with aggregates in the extracellular milieu of wild-type and Δ ssIE mutant bacteria at pH 5.0. *** $P < 0.001$; versus Δ ssIE mutant by two-tailed Student's test and error bars are the standard error of the mean. **g** Δ ssIE::ssIE W strain washed in citrate-phosphate buffer at pH 4.0, again with α -rSsIE antibodies reacting with the bacterial surface and secreted fibrous material. Scale bar is equivalent to 200 nm. **h** TEM analysis of negatively stained rSsIE aggregates. Two major species are observed. **i** Smaller filaments have an average width of ~ 4.5 nm and are coated with globular structures of ~ 4.5 nm in diameter. Scale bar, 20 nm. **j** Schematic describing the overall dimensions of the smaller SsIE fibre species. **k** Larger species of SsIE aggregates appear as ~ 20 – 40 nm wide structures with variable lengths. Scale bar, 20 nm.

kinetics^{30,31}. The COSMiCS decomposition, assuming the coexistence of two components, was able to adequately fit all the SAXS profiles with an average $\langle \chi^2 \rangle$ of 1.5, although some deviations from the perfect fit were observed for the SAXS curves measured in the first 3 h (Fig. 4c). This observation suggested the presence of a small population of a third species in the initial steps of fibrillation but this putative third species could not be captured when considering a three-species decomposition with COSMiCS.

The extracted profile of the smallest species, and the subsequently derived R_g (3.92 nm) and D_{max} (13.3 nm) values, were almost identical to the SEC-SAXS profile for rSsIE at pH 4.4 (Fig. 4d, e and Supplementary Table 1), and this indicated that SsIE is in a monomeric form at the beginning of the fibrillation process. The extracted SAXS profile of the second component displayed the typical features of a large particle and $P(r)$ analysis indicated it had

an R_g of ~ 51.1 nm and D_{max} of ~ 140 nm (Fig. 4d, e and Supplementary Table 1). Furthermore, fractal fit analysis^{32,33} of the decomposed SAXS profiles also indicated that rSsIE aggregates into a structure that is fractal in nature (Supplementary Fig. 16). The radius of the primary particle was determined to be ~ 3.5 nm while the mass fractal and surface fractal dimensions were determined to be 2.9 and 3.0, respectively.

Examination of the relative populations of the two species derived from the COSMiCS decomposition over time showed that after 30 min from initiating the measurements, the aggregated species already represented 50% of the molar fraction (Fig. 4f). From this time point, this population continued to grow and reached a plateau after ~ 10 h. As the SAXS fibrillation data at the 10 h time point had shown depletion of monomeric rSsIE but also had reduced fractal properties compared with later time points

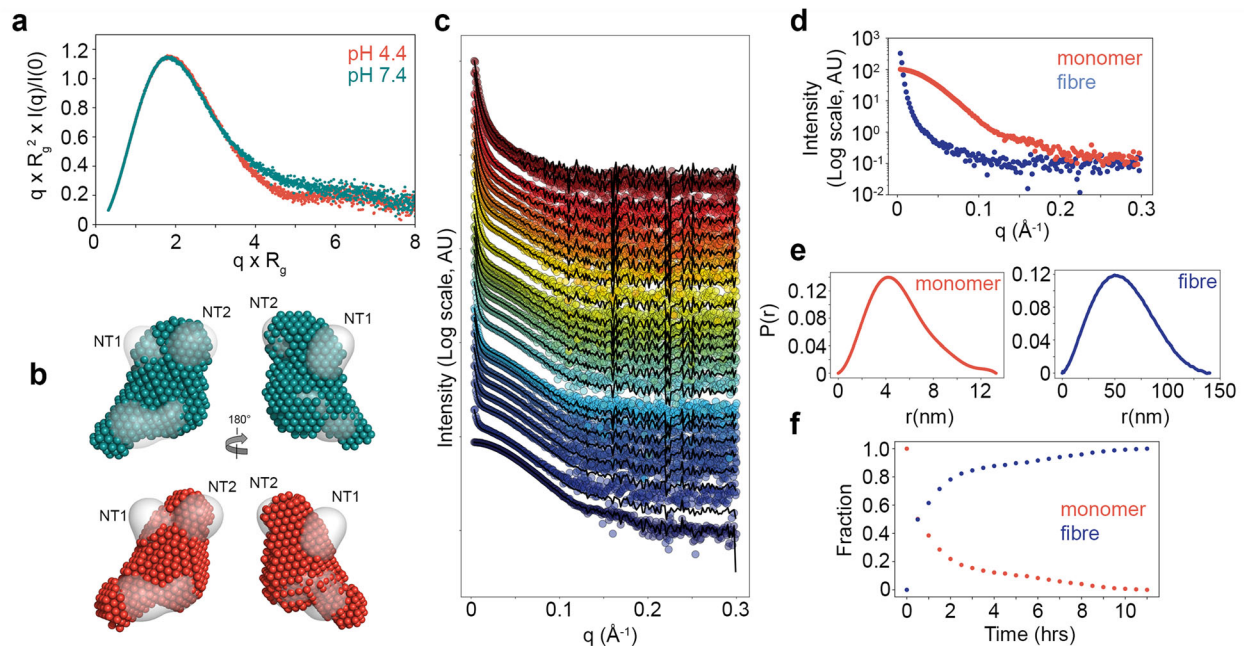


Fig. 4 Kinetics of SsIE aggregation. **a** SAXS normalised Kratky plots for rSsIE at pH 4.4 and 7.4. **b** Overlay of rSsIE SAXS bead models at pH 4.4 (red) and 7.4 (teal) with rSsIE TEM map (grey). NT1 and NT2 domains are highlighted. **c** COSMiCS decomposition of the time-dependent rSsIE SAXS fibrillation data, assuming that the whole dataset can be described with two co-existing components (monomer and fibre, panel **d**) with different relative populations (**f**). SEC-SAXS curve for rSsIE at pH 4.4 (bottom curve) is also shown for $t = 0$ and used to define the curve of the monomeric component. Experimental curves are displayed in colour and arbitrarily displaced along the y axis for a better inspection. Black curves are the result of the COSMiCS optimisation. The excellent agreement with the experimental data testifies the accuracy of COSMiCS decomposition. **e** Shape distribution ($P(r)$) functions derived from COSMiCS decomposed scattering curves for rSsIE monomer (red) and aggregates (blue). **f** Relative populations of the two species of rSsIE derived from the COSMiCS analysis over time.

(Fig. 4f), we discarded the low q region and used this data to carry out analysis of the fibre asymmetric unit. Geometrical shape analysis suggested that rSsIE fibres form a cylinder with a radius 3.2 nm and height 16.5 nm, and cross-sectional analysis indicated a repetitive unit with a cross-sectional R_g (R_C) of 4.78 nm. The cross-sectional $P(r)_C$ gave a cross-sectional D_{max} of 14.0 nm. Together these data supported our observations of single and clustered rSsIE fibres derived from negative-stain TEM and MALS.

Biofilm maturation supports polymerisation of SsIE

As SsIE aggregates had been shown to bind Congo red, we used this dye to assess the role of SsIE during the formation of *E. coli* W strain macrocolony biofilms. When we compared the Δ sse mutant cultured on LB Congo red agar at 37 °C for 24 h to its parental wild-type strain we observed no major differences (Fig. 5a). After 96 h both strains displayed a lobated morphology but in the mutant the dye was retained as a compact ring within the centre of the colony, while in the wild-type strain the dye radiated out from the middle of the colony towards the edge (Fig. 5a, b). However, complementation of the Δ sse mutant with intact *ssIE* or *ssIE Δ M60* was able to recover macrocolony morphology similar to that of the parental wild-type (Fig. 5a).

We reasoned that if SsIE aggregates are formed in macrocolonies and can bind Congo red dye, SsIE must experience extracellular pH values <5.0 during the development of these biofilms. To test this, we first grew *E. coli* W biofilms using a microfluidic system under continuous flow and assessed their overall morphology using confocal laser scanning microscopy (CLSM). While biofilms grown for 48 h produced a relatively homogenous lawn of bacterial growth, after 96 h there were clear signs of maturation, with a significant increase in biofilm mass, structural heterogeneity, and the presence of water channels (Fig. 5c and Supplementary Fig. 17). We then examined the *E. coli* W

Δ sse mutant under these conditions, but in line with previous reports in EPEC strain E2348/69⁴, it was unable to develop structures beyond microcolonies (Fig. 5c). Furthermore, complementation of the mutant with either intact *ssIE* or *ssIE Δ M60* was again able to restore wild-type biofilm morphology. Together with our previous observations, this indicated that the C-terminal M60 domain is not required for translocation of SsIE through its T2SS and it is dispensable for biofilm development, at least under these conditions.

We then examined pH distribution across *E. coli* W biofilms using CLSM coupled with the cell-impermeant fluorescent ratiometric probe seminaphthorhodafluor-4F 5-(and-6) carboxylic acid (C-SNARF-4). After growing *E. coli* W for 24 h under continuous flow, we observed pH values between 6.0 and 6.3 across biofilms (Supplementary Fig. 18), however, after 96 h, lower pH values were also recorded (Fig. 5d). Although across most biofilms examined, we detected pH values between approximately 5.0 and 6.0, clearly defined microenvironments were also observed with pH values that ranged from 4.2 and 4.8.

Analysis of EPS in *E. coli* macrocolony biofilms

We speculated that during the maturation of biofilms, SsIE would interact with other components of the biofilm matrix and may contribute to the structural integrity of these communities. Cellulose, curli, polymeric β -1,6-linked N-acetylglucosamine (PNAG) and colanic acid are major EPS in *E. coli* biofilms³⁴ and we examined which of these are produced in macrocolonies from several *E. coli* strains under conditions that support the production of SsIE. *E. coli* W, ETEC H10407 and EPEC E2348/69 strains were studied as they had previously been shown to secrete SsIE^{4,5,11}, while *E. coli* BL21 (DE3) strain was expected to secrete SsIE based on genomic analysis⁵. In addition, we also incorporated several other control strains. The *E. coli* K12 strain BW25113 was used as it

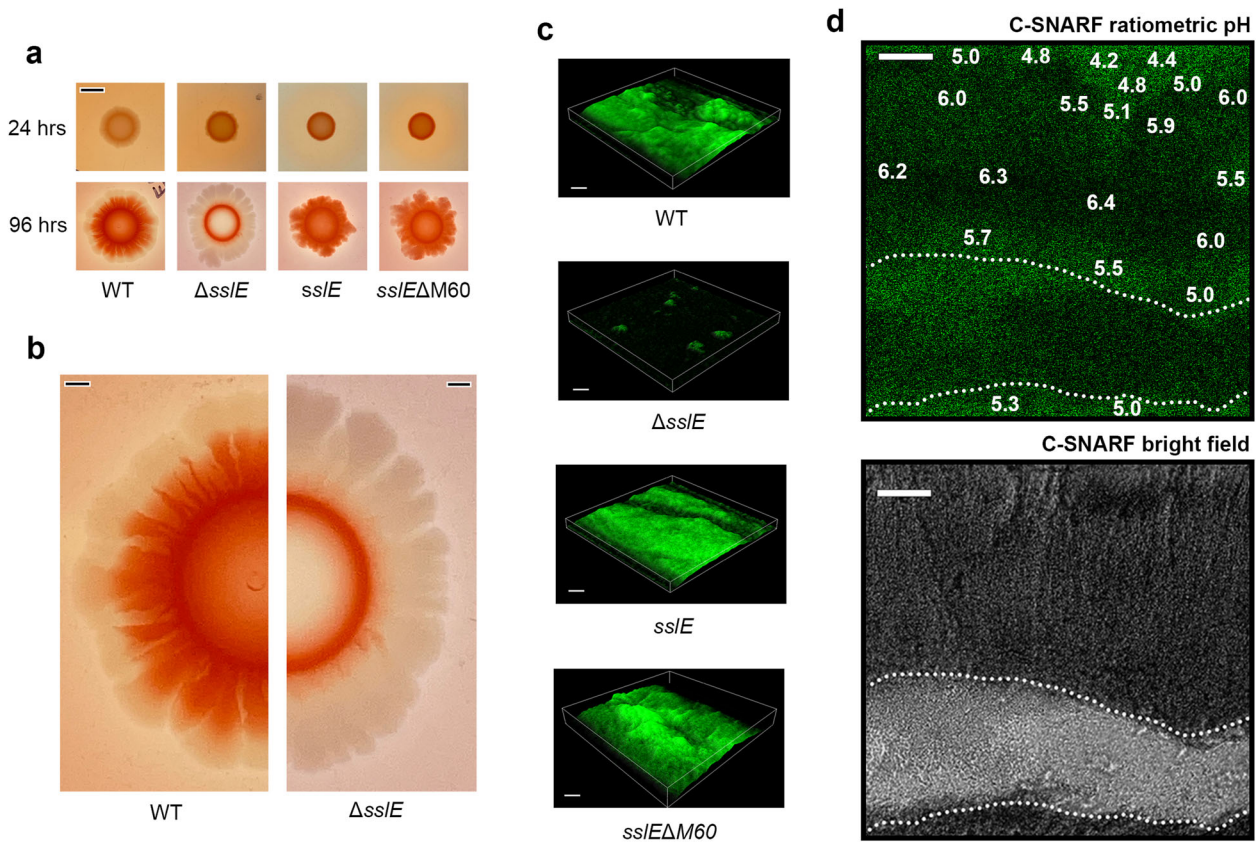


Fig. 5 Analysis of SslE-dependent biofilm formation by *E. coli* W strain. **a** Macrocolony biofilm phenotype of wild-type *E. coli* W and derivatives on LB Congo red agar medium after 24 and 96 h of growth. The derivatives *sslE* and *sslE* Δ M60 represent *trans*-complementation of intact *sslE* (Δ *sslE*::*sslE*) or *sslE* minus its M60 domain (Δ *sslE*::*sslE* Δ M60), respectively, into the *E. coli* W Δ *sslE* mutant. Scale bar = 5 mm. **b** Zoomed-in image of the wild-type *E. coli* W and Δ *sslE* mutant. Scale bar = 2 mm. **c** CLSM images of wild-type *E. coli* W biofilms and *E. coli* W Δ *sslE* mutant and *trans*-complementation with *sslE* and *sslE* Δ M60 stained with FM 1-43 (green). Scale bar represents 20 μ m. **d** The pH across wild-type *E. coli* W biofilms grown for 96 h was monitored ratiometrically using C-SNARF-4 (green). pH values were calculated over \sim 30 μ m² boxes and pH values for representative regions within the biofilm fringes and centres are annotated. Dotted line outlines water channel within the biofilm, where pH values are not shown. Scale bar represents 20 μ m. All data are representative of at least three independent experiments.

does not produce SslE or cellulose due to deletion of several *T2SS* genes and mutations in its cellulose biosynthesis genes^{5,35}, respectively, and we created an isogenic Δ *sslE* knockout mutant in ETEC H10407. We also examined the published EPEC E2348/69 Δ *sslE*, E2348/69 Δ *csgA* (curli defective) and E2348/69 Δ *csgA*/ Δ *bcsA* (curli and cellulose defective) knockout mutant strains^{4,36}, and the BW25113 Δ *wcaF* (colanic acid defective) and BW25113 Δ *pgaC* (PNAG defective) knockout mutant strains from the Keio collection³⁷.

We first cultured these strains on LB Congo red agar grown at 37 °C for 96 h (Fig. 6a). Both wild-type BL21 (DE3) and H10407 strains produced smooth and round colonies with an even distribution of dye, but deletion of the *sslE* gene in H10407 resulted in Congo red being retained as a ring in the centre of the colony. Colonies of wild-type E2348/69 strain formed circular and smooth colonies with slight lobing and with dye radiating out from the middle of the colony. However, little difference in colony morphology was observed between the E2348/69 Δ *sslE* and Δ *csgA* mutants and their parental wild-type strain, although the Δ *csgA*/ Δ *bcsA* double mutant did show a reduction in Congo red intensity and localisation. Macrocolonies of wild-type BW25113 and its derivatives retained the dye as with the W and H10407 Δ *sslE* mutants, but while the wild-type and Δ *wcaF* mutant strains appeared similar in morphology, the Δ *pgaC* mutant showed significantly more lobation and grew approximately twice the size.

We next examined macrocolony morphologies after growth on LB Congo red agar and YESCA Congo red agar at 25 °C, conditions

known to repress and induce the production of SslE and curli/cellulose, respectively (Fig. 6b). All wild-type strains on LB displayed an overall colony morphology similar to growth at 37 °C, however, W, BL21 (DE3) and H10407 strains displayed distribution of Congo red like their corresponding Δ *sslE* mutant. Wild-type BW25113 and E2348/69 strains on the other hand showed increased Congo red retention, but this was reduced in the E2348/69 Δ *csgA* and Δ *csgA*/ Δ *bcsA* mutant strains. Examination of the W, BL21 (DE3), H10407 and E2348/69 Δ *csgA*/ Δ *bcsA* strains on YESCA all showed a smooth and white morphology and implied that these strains do not produce curli or cellulose³⁸. As anticipated, wild-type E2348/69 strain displayed a red, dry and rough morphology, indicative of both curli and cellulose production, while E2348/69 Δ *csgA* strain showed a pink, dry and rough morphology (cellulose only) and wild-type BW25113 appeared with a brown, dry and rough morphology (curli only)³⁸. This indicated that under conditions that support the expression of *sslE*, neither curli nor cellulose are produced.

Extracellular levels of cellulose were next quantified by measuring glucose release after digestion with cellulase (Fig. 6c). After treatment all wild-type strains showed levels of glucose consistent with the E2348/69 Δ *csgA*/ Δ *bcsA* double mutant strain. This was further supported through the examination of macrocolony growth in the presence of the cellulose stain calcofluor white, where only wild-type E2348/69 strain grown on YESCA agar at 25 °C displayed noticeable colony fluorescence with UV light (Supplementary Fig. 19). The relative expression of SslE and curli

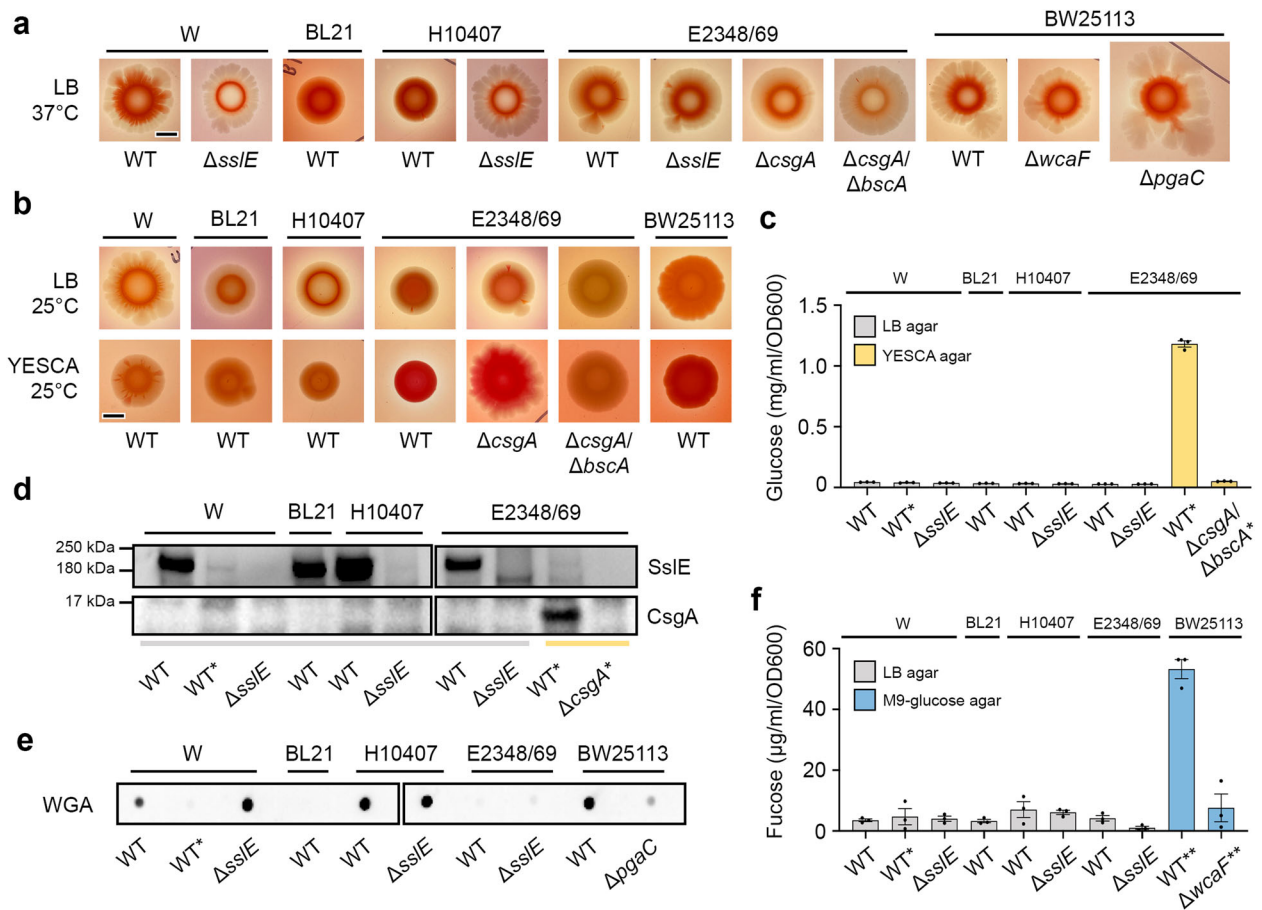


Fig. 6 EPS production in *E. coli* macrocolony biofilms. Analysis and quantification of EPS from macrocolony biofilms of wild-type *E. coli* W, BL21 (DE3), H10407 E2348/69 and BW25113 strains and their derivatives after 96 h growth. **a** Macrocolony phenotypes of strains grown at 37°C on LB Congo red agar medium, **b** 25°C on LB Congo red agar medium and 25°C on YESCA Congo red agar medium. Scale bar represents 5 mm. **c** Cellulose production after digestion with cellulase from strains grown on LB agar (grey) or YESCA agar (yellow) at 37°C (no asterisks) or 25°C (*), detected with either anti-rSsIE antibody or anti-CsgA antibody. **d** Immunoblot analysis of whole cells grown on LB agar (grey line) or YESCA agar (yellow line) at 37°C (no asterisks) or 25°C (*), detected with either anti-rSsIE antibody or anti-CsgA antibody. **e** Dot-blot analysis of PNAG content from whole cells grown on LB agar at 37°C (no asterisks) or 25°C (*), detected with wheat germ agglutinin (WGA). **f** Quantification of colanic acid from strains grown on LB agar (grey line) or M9-glucose agar (blue line) at 37°C (no asterisks), 25°C (*) or 18°C (**). Data are representative of at least three repeat experiments. All blots were derived from the same experiment and processed in parallel and error bars are the standard error of the mean.

were then assessed by immunoblot analysis of macrocolonies (Fig. 6d). Production of SsIE was observed in wild-type W, BL21 (DE3), H10407 and E2348/69 strains when grown at 37°C on LB but not when grown at 25°C on LB or YESCA agar. Conversely, the major curli subunit CsgA was only detected when E2348/69 strain was grown on YESCA agar at 25°C. We then assessed the production of PNAG by probing partially purified extracellular carbohydrates using wheat germ agglutinin (Fig. 6e). Our results showed significant amounts of PNAG could be detected in W, H10407 and BW25113 strains when they were grown at 37°C on LB agar, but no PNAG was detected in BL21 (DE3) or E2348/69 strains grown under the same conditions, or when the W strain grown at 25°C on LB. We finally examined the presence of colanic acid in colonies by measuring fucose evolution after treatment with sulphuric acid (Fig. 6f). However, we could only detect colanic acid production when wild-type BW25113 strain was grown under nutrient limiting conditions at low temperature³⁹ and not when strains were grown on LB agar at 37°C.

SsIE associates with eDNA in biofilms

To examine whether rSsIE can directly bind PNAG we first produced rSsIE fibres at pH 4.0 and then isolated them in citrate-

phosphate buffer at pH 6.0. At this pH, the structure of rSsIE fibres would be retained, while monomeric rSsIE would resist polymerisation (Fig. 2a). As a control, we used the C-terminal domain (CTD) of the *Porphyromonas gingivalis* gingipain protein RgpB, which was not expected to bind EPS and remained stable under these conditions (Supplementary Fig. 20). We then monitored the binding of rSsIE fibres, rSsIE monomers and RgpB-CTD to PNAG using a dot-blot overlay assay but could not detect any interaction (Supplementary Fig. 21). However, when we assayed the association of rSsIE to immobilised cellulose discs produced by the bacterium *Komagataeibacter rhaeticus*, we detected a significant interaction with fibres, but not monomeric rSsIE or RgpB-CTD (Supplementary Fig. 22). Nonetheless, the specificity of this binding was not investigated further as cellulose had not been detected in conditions that supported SsIE production (Fig. 6a–c).

Extracellular DNA (eDNA) is another key component of many biofilms and provides adhesion, structuring and chelation of cations⁴⁰. We considered whether SsIE could associate with eDNA and performed an electrophoretic mobility shift assay (EMSA) using the plasmid pUC19 as a DNA substrate. Plasmid DNA was incubated with increasing concentrations of either rSsIE fibres, rSsIE monomers or RgpB-CTD control at pH 6.0 and then analysed on agarose gels (Fig. 7a). Upon incubation of pUC19 with 100 μ M

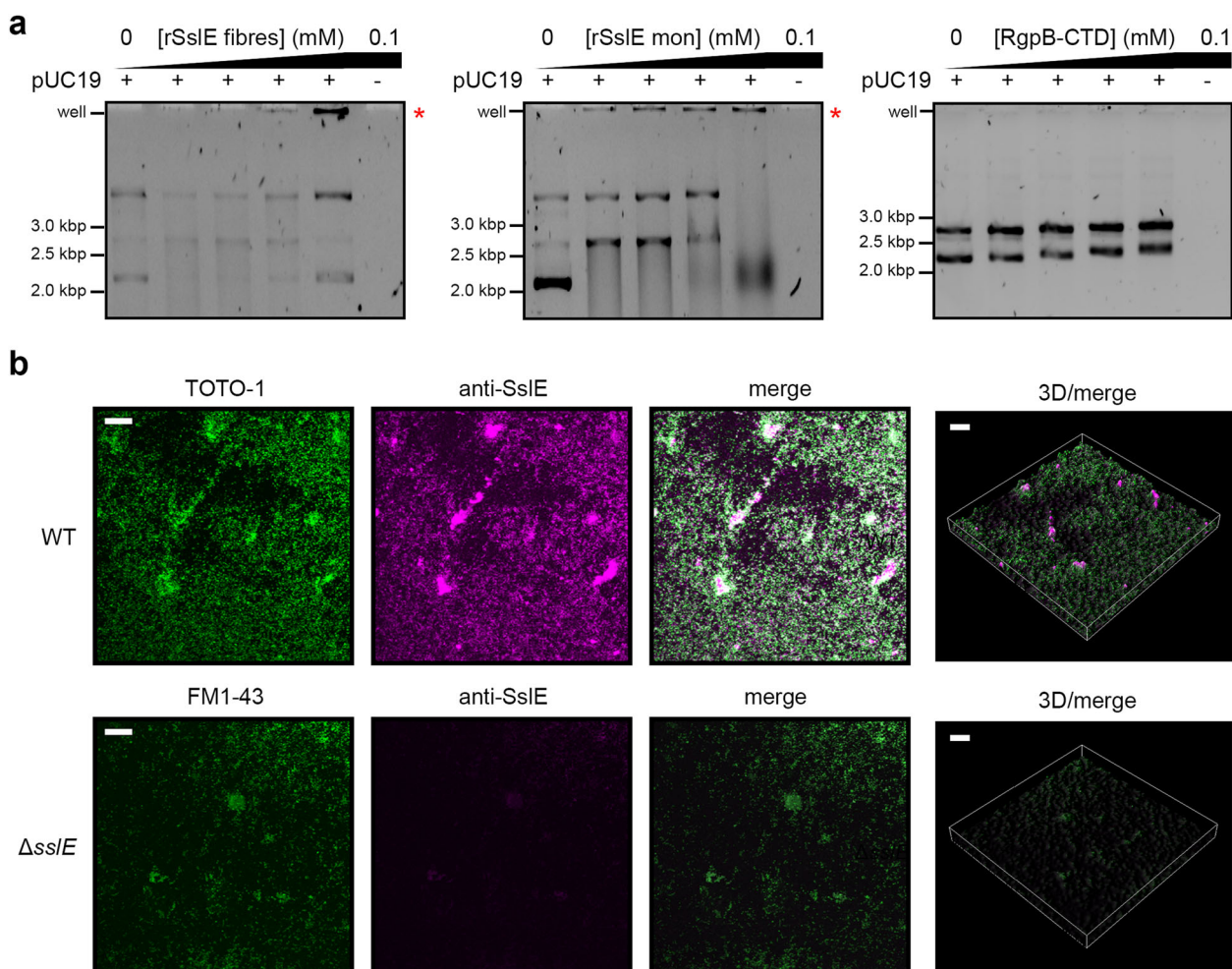


Fig. 7 SsIE associates with eDNA in *E. coli* W biofilms. **a** Electrophoretic mobility shift assay of rSsIE monomers and fibres with pUC19 plasmid DNA substrate. *P. gingivalis* RgpB-CTD was used as a negative control. Red asterisk (*) indicates the formation of a large protein–DNA complex retained in the wells. Gels were derived from the same experiment and processed in parallel. **b** Wild-type *E. coli* W biofilms were formed using a continuous flow cell system over 96 h. Immunofluorescence was performed where biofilms were incubated with α -rSsIE antibodies followed by goat anti-rabbit conjugated Alexa Fluor 633 (magenta). Extracellular DNA and dead cells were stained with TOTO-1 (green). *E. coli* W Δ ssIE strain was used as a control and again stained with α -rSsIE/Alexa Fluor 633 (magenta) and the membrane dye FM 1–43 (green). Biofilms were visualised by CSLM. Scale bar is equivalent to 50 μ m. All data are representative of at least three independent experiments.

rSsIE fibres, a large protein–DNA complex was detected in the wells of the gel, but not at lower concentrations or when incubated with the control. Likewise, when pUC19 was incubated with 12.5 μ M monomeric rSsIE there was again evidence of a protein–DNA complex in the well. We then assessed the distribution of SsIE and eDNA within established wild-type *E. coli* W biofilms stained with the rSsIE antibody and the cell-impermeant DNA-binding dye TOTO-1, respectively, using CSLM (Fig. 7b). Here, we observed TOTO-1 staining across the entire biofilm structure with eDNA localised to dead cells and distributed throughout the biofilm matrix. Labelling of SsIE was again seen across biofilms with SsIE associated with bacteria, likely through association with the bacterial surface, but also as large clearly defined clusters that co-localised with eDNA and were not visible in the Δ ssIE control.

DISCUSSION

In this study, we have revealed that SsIE is composed of four defined regions: two globular domains at the N-terminus (NT1 and

NT2) and a central torus-shaped core, where a NT3 domain packs against an M60-like aminopeptidase. The N-terminal region (NT1, NT2 and NT3) is composed of a unique primary sequence that has only been identified in SsIE and its homologs⁴¹, while the C-terminal M60-like domain is found in a wide range of proteins secreted by both prokaryotic and eukaryotic microorganisms that interact with mucosal membranes^{18,42}. The atomic structures for three M60-like domains in complex with O-glycopeptides have been described: BT4244 from *Bacteroides thetaiotaomicron* (PDB ID code: 5kd2), IMPa from *P. aeruginosa* (PDB ID code: 5kdw) and ZmpB from *Clostridium perfringens* (PDB ID code: 5kdn)¹⁹. While the active site structures in these bacterial enzymes are highly conserved, there are variations within the remaining M60 domain, which enables them to recognise unique glycan sequences. Structural data is not available for intact BT4244 or ZmpB, although they are thought to contain additional carbohydrate-binding domains¹⁸ and it is likely that these M60 augmentations have a role in the recognition of O-glycopeptide substrates. It also appears that the active site in the M60 domain of SsIE faces into the NT3 domain and this implies that both the NT3 and M60

domains are responsible for the recognition of mucins in SsIE. With significant flexibility of the NT1 and NT2 region, these domains may also interact with mucin substrates and may promote processivity.

Although SsIE has a clear function in the processing of mucins, it is a unique member of the M60-like aminopeptidase family as it is also fundamental for the maturation of EPEC biofilms grown under flow. We have also now shown that SsIE is required for the maturation of *E. coli* W biofilms, where it is widely distributed throughout the extracellular matrix. We have demonstrated that SsIE can form aggregates that display amyloid-like properties at $\text{pH} \leq 4.4$; conditions which are also observed as microenvironments across mature wild-type *E. coli* W biofilms. It is well established that phenotypical heterogeneity develops during the maturation of biofilms and direct measurement of extracellular pH within *E. coli* PHL628 biofilms has also shown a heterogenous distributions of pH, with values ranging between 5.0 and 7.0⁴³. Therefore, it is very likely that SsIE can form aggregates within these transiently forming microenvironments and after increases in the local pH they then resist solubilisation. To support this, we have demonstrated that aggregated SsIE is deposited throughout macrocolony biofilms and deletion of the *sslE* gene in *E. coli* W and ETEC H10407 resulted in altered colony morphologies.

Microbial amyloid structures are ubiquitous in biofilms and in Gram-negative bacteria most known amyloid structures are synthesised by either the Curli or Fap secretion systems^{44–47}. These form through the spontaneous aggregation of extended peptides and provide adhesion, cohesion and contribute to the structural integrity of the biofilm matrix. Furthermore, there are many examples where reduction in pH increases the rate of amyloid formation, through protonation of aspartate and glutamate residues and subsequent structural rearrangements^{48,49}. Although our biochemical and biophysical analyses suggest that under acidic condition rSsIE displays some amyloid-like features, examination of aggregates using negative-stain TEM has revealed short fibres associated with globular structures on their surface. FTIR analysis of rSsIE aggregates provided an amide I maxima at $\sim 1624 \text{ cm}^{-1}$, which is consistent with an increase in anti-parallel β -sheet conformation and also lies within the range of amide I maxima for typical amyloid structures (e.g. $1611\text{--}1630 \text{ cm}^{-1}$)⁵⁰. However, larger, and more rigid amyloid structures generally have peaks at lower frequencies, while native β -sheets show absorption bands with maxima between $1630\text{--}1643 \text{ cm}^{-1}$. This suggests that the rSsIE amide I band peak represents an amyloid structure which is relatively small and is not overly rigid, although it could alternatively originate from fibre stacking/clumping effects. Binding of ThT to rSsIE at acidic pH also supports a general increase in β -sheet content in fibres but could also indicate the formation of new cavities upon fibrillation⁵¹. It is therefore still unclear whether these changes reflect the formation of a continuous amyloid-like structure that runs the length of the fibre core or instead represent β -sheet augmentation between adjacent SsIE molecules along the fibre, upon fibrillation of SsIE. Nonetheless, based on our biophysical data, we propose a tentative model for the fibrillation of SsIE into functional aggregates via end-to-end stacking. In this model, the asymmetric unit of the fibre core is composed of the NT2, NT3 and M60 domains and the NT1 domain is flexible and can move freely around the surface.

Moreover, we have demonstrated that rSsIE fibres and monomers can directly bind DNA in vitro and SsIE antibodies co-localise with eDNA and bacteria in *E. coli* W biofilms. After secretion SsIE can associate with the bacterial membrane but is also present in a soluble form in the extracellular milieu (Supplementary Fig. 7)⁴. Therefore, cell surface-associated SsIE could presumably interact with extracellular fibres or may help to nucleate the aggregation of soluble SsIE. This could cross-link bacteria and act as a hub to link with eDNA in the matrix.

However, it is still unclear which form of SsIE is functional in binding eDNA within biofilms and alternatively, monomeric SsIE may directly bind eDNA without the need for SsIE aggregates. As the nature and consequence of this interaction is not yet clear this now needs more investigation. Nonetheless, we have also presented some evidence that SsIE fibres can also associate with cellulose, which is known to have a structural function in biofilms and provides cohesion and elasticity⁵². Although we did not detect cellulose under conditions of SsIE production, cellulose secreted by other bacteria in multispecies communities could provide a source for this interaction. As a monomer, SsIE does not recognise cellulose, and this suggests that new binding sites are formed after its aggregation. The presence of a surface-exposed domain also implies that this domain may be functional in binding other yet to be identified ligands, or is functional under different conditions, for example through interacting with mucins during intestinal colonisation. Likewise, the fibre core may still retain its mucolytic activity and could play a role in mucin restructuring during biofilm growth in the gut.

Although *E. coli* W, BL21 (DE3), ETEC 10407 and EPEC E2348/69 strains all secrete SsIE, the first two strains are harmless commensals, while ETEC and EPEC are major aetiological agents in developing countries^{53,54}. SsIE is also required for colonisation in a rabbit model of EPEC infection⁴ and is actively transcribed during ETEC infection of mice⁶. This appears to be a common feature of SsIE production as expression of *sslE* was previously shown to be repressed at low temperatures in *E. coli* W⁵ and ETEC⁷ and here we have shown that this is also the case in BL21 (DE3) and EPEC E2348/69 strains. This suggests that SsIE increases the virulence of *E. coli* pathotypes through its ability to promote biofilm maturation and/or through its interactions with mucosal defences. Both enterotoxigenic and enteropathogenic *E. coli* strains cause infection in the small intestine¹ where the intraluminal pH ranges between ~ 6.0 and 7.5 ⁵⁵. However, *E. coli* possess an acid-tolerance response that supports exponential growth at pH values as low as $\text{pH} 4.2$ ⁵⁶ and due to the synthesis of organic acids by the residing microbiota, during colonisation of the intestine *E. coli* will experience extracellular pH values that range between 4.0 and 6.0⁵⁷. Together this suggests that SsIE is important for infection within the intestinal tract rather than extraintestinal survival, where conditions exist that support its secretion and ability to function as both a monomer and aggregate during biofilm formation. SsIE represents a unique protein and further studies are now required to understand its complete functions during ecology and disease.

METHODS

Media

Lysogeny broth (LB) medium contained (per litre) 10 g tryptone, 5 g yeast extract and 10 g NaCl. LB agar contained (per litre) 10 g tryptone, 5 g yeast extract, 10 g NaCl and 15 g agar. Super optimal broth (SOB) medium essentially contained (per litre) 20 g tryptone, 5 g yeast extract and 0.5 g NaCl. Super Optimal broth with Catabolite repression (SOC) medium was SOC medium with additional 20 mM glucose. YESCA agar contained (per litre) 10 g casamino acids, 1 g yeast extract and 15 g agar. M9-glucose agar contained (per litre) 5.3 g K_2HPO_4 , 2 g KH_2PO_4 , 1 g $(\text{NH}_4)_2\text{SO}_4$, 0.5 g sodium citrate, 16 g agar, 4 g glucose, 0.1 g MgSO_4 and 2 μg thiamine. Isotopically defined M9 minimal medium (pH 7.4) contained (per litre) 6.0 g $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$, 3 g KH_2PO_4 , 0.5 g NaCl, 0.12 g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 22 μg CaCl_2 , 40 μg thiamine, 8.3 mg $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$, 0.5 mg ZnCl_2 , 0.1 mg CuCl_2 , 0.1 mg $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$, 0.1 mg H_3BO_3 and 13.5 mg $\text{MnCl}_2 \cdot 6\text{H}_2\text{O}$, supplemented with 2 g $[\text{U}\text{-}^{13}\text{C}_6]\text{glucose}$ and/or 0.7 g $^{15}\text{NH}_4\text{Cl}$. M9 media was made up in deuterium oxide for the production of perdeuterated protein samples and pH was adjusted using 1 M NaOH solution. All NMR isotopes were from Sigma (UK).

Bacterial strains

All bacterial strains, plasmids and primers used in this study are listed in Supplementary Table 3.

Gene deletion

A non-polar deletion of *sslE* was constructed in *E. coli* W and H10407 strains by allelic exchange with a FLP recombination target (FRT)-flanked Kan^r cassette, using a modified protocol⁵⁸. The fragment was amplified from *E. coli* strain JW5925-1³⁷ using Platinum Taq DNA polymerase (Promega) and primers PC3 and PC4 and then exchange into *E. coli* W and H10407 strains were facilitated by the λ Red recombinase system carried on plasmid pKD46. *E. coli* strains carrying the pKD46 plasmid were grown in SOB media with 100 μ g/ml ampicillin and 1 mM L-arabinose at 30 °C to an OD of 0.6. Cells were made electrocompetent, electroporated with 10–100 ng of PCR fragment and then selected on LB agar at 37 °C containing 25 μ g/ml kanamycin, followed by transfer and growth on medium containing no antibiotics. Loss of the pKD46 plasmid was tested by ampicillin sensitivity. Knockouts were confirmed by PCR and sequencing.

Plasmid construction

Complementation plasmids pCPC1 and pCPC2 were generated by amplifying whole length *sslE*, or *sslE* minus the M60 domain, from *E. coli* W gDNA using primer pairs PC7/PC8 or PC7/PC9, respectively. These were digested with HindIII/NheI (NEB), ligated into HindIII/NheI-digested pBad-cm18 vector. Synthetic genes gPC4, gPC5, gPC6 and gPC7 were synthesised by Synbio Technologies (USA) and cloned into pET28b vector using NcoI and XhoI restriction sites to create plasmids pPC2, pPC3, pPC4 and pPC5, respectively (Supplementary Table 4). Plasmid pBD1 was created by amplification of the RgpB-CTD from *P. gingivalis* W50 gDNA using primers BD1 and BD2. This was then cloned into pET46 Ek/LIC vector through ligation-independent cloning (Novagen).

Protein purification

Recombinant SsIE (rSsIE: residues 67–1497; numbered based on mature sequence)¹², SsIE NT1 (residues 67–211), NT2 (residues 230–425), NT1-NT2 (residues 67–425), NT3-M60 (residues 426–1497) and RgpB-CTD were transformed into *E. coli* SHuffle cells (SsIE; New England Biolabs) or BL21 (DE3) (RgpB-CTD; New England Biolabs), grown at 37 °C in LB media (rSsIE, RgpB-CTD: 100 μ g/ml ampicillin; NT1, NT2, NT1-NT2, NT3-M60: 100 μ g/ml kanamycin) and expression induced with 0.5 mM isopropyl-d-1-thiogalactopyranoside (IPTG) at an OD_{600nm} of 0.6, followed by growth overnight at 18 °C. Cells were resuspended in 20 mM Tris–HCl pH 8, 200 mM NaCl, lysed by sonication and purified using nickel-affinity chromatography followed by gel filtration with either a Superdex 75 (NT1, NT2, RgpB-CTD) or 200 (rSsIE, NT1-NT2, NT3-M60) column (GE Healthcare).

SEC-SAXS

SAXS data were collected on beamline B21⁵⁹ at Diamond Light Source (DLS, Oxford, UK) at 25 °C. 60 μ l of rSsIE (10 mg/ml) in 20 mM Tris–HCl pH 8, 200 mM NaCl was applied to a KW403-4F column (Shodex) at 0.16 ml/min, pre-equilibrated in 20 mM citrate-phosphate buffer, 200 mM NaCl at either pH 4.4 or 7.4. SsIE N1, N2 and N3-M60 sub-domains were applied to a KW403-4F column (Shodex) at 0.16 ml/min, pre-equilibrated in 20 mM Tris–HCl pH 8, 200 mM NaCl. SAXS data were measured over a momentum transfer range of $0.003 < q < 0.44 \text{ \AA}^{-1}$. Peak integration and buffer subtraction were performed in CHROMIXS⁶⁰. The radius of gyration (R_g) and scattering at zero angle $I(0)$ were calculated from the analysis of the Guinier region by AUTORG⁶⁰. The distance distribution function $[P(r)]$ was subsequently obtained using GNOM⁶⁰ yielding the maximum particle dimension (D_{max}). Ab initio low-resolution shape restoration was carried out by calculating 20 models in DAMMIF⁶⁰, which were subsequently averaged using DAMAVER⁶⁰ and used as a starting model for refinement in DAMMIN⁶⁰. An additional 20 models were also calculated and averaged in DAMAVER⁶⁰. CRYSOLO⁶⁰ was used to compare final rSsIE TEM envelopes and models against the solution SAXS curve. Processing and refinement statistics can be found in Supplementary Table 1.

TEM single-particle analysis

In all, 4 μ l of rSsIE (625 nM) in 50 mM Tris–HCl pH 8.0, 150 mM NaCl was applied to previously glow-discharged 300 mesh continuous carbon-

coated copper grids (Agar Scientific Ltd, UK) for 1 min and blotted for excess liquid. 4 μ l of 2% (v/v) uranyl acetate was applied for staining for 1 min. The excess liquid was blotted and left to dry. Data were acquired using a JEOL JEM-2100 plus TEM operating at 200 kV equipped with a OneView 16 Megapixel camera (Gatan). 50 micrographs were collected at a nominal magnification of $\times 50,000$ with a pixel size of 2.1 $\text{\AA}/\text{pixel}$ and a range of defocus from 1 to 3 μ m. Data were processed using Relion 3.1⁶¹. Defocus and astigmatism parameters were estimated using CTFIND4⁶² in Relion 3.1. An initial dataset of 41,752 particles were autopicked using 2D class averages generated using approx. 1,500 manually picked particles as reference. After a few rounds of 2D classification ignoring CTF until the first peak, 10,988 particles were taken forward for an initial model generation using 3D initial model in Relion 3.1⁶¹. Following initial model generation, a few rounds of 3D classification followed by 3D refinement was carried out. The final model was refined to 2.2 \AA using the gold standard FSC (0.143 criterion). PHYRE2⁶³ was used to analyse the SsIE sequence (residues 67–1497; UniProt ID E0IW31) and generate a homology model for the C-terminal M60 domain (residues 962–1415) based on residues 421–894 of *P. aeruginosa* IMPa (PDB ID code 5kdv; 100% confidence, 21% identity)¹⁹. This was then docked into the rSsIE TEM envelope using UCSF CHIMERA⁶⁴.

Solution NMR spectroscopy

NMR measurements for rSsIE were performed at 25 °C on a 100 μ M ²H¹⁵N-labelled sample in 50 mM NaPO₄ pH 7.4, 100 mM NaCl, 10 % D₂O. NMR measurements for SsIE NT1 (0.6 mM), SsIE NT2 (1.3 mM) and SsIE NT1-NT2 (0.8 mM) were performed at 25 °C on ¹⁵N-labelled samples in 50 mM NaPO₄ pH 7.0, 100 mM NaCl, 10 % D₂O. NMR measurements for RgpB-CTD (0.3 mM) were performed at 37 °C on a ¹⁵N-labelled sample in 20 mM NaPO₄ pH 6.0, 100 mM NaCl, 10% D₂O. Transverse relaxation optimised spectroscopy (TROSY) based ¹H¹⁵N-HSQC experiment and T_1 and T_2 relaxation times for rSsIE were recorded on a Bruker Avance III HD 950, equipped with a TXI cryoprobe. TROSY ¹H¹⁵N-HSQC experiment for SsIE NT1-NT2 was recorded on a Bruker Avance III HD 800, equipped with a TCI cryoprobe. Standard ¹H¹⁵N-HSQC spectra of SsIE NT1, SsIE NT2 and RgpB-CTD were recorded on a Bruker Avance III HD 700, equipped with a TCI cryoprobe. Data were processed in NMRPIPE⁶⁵ and analysed/visualised with ANALYSIS⁶⁶ and NMRVIEW⁶⁷.

Recombinant protein ring assay

In all, 20 μ l of rSsIE, NT1-NT2 or NT3-M60 (62.5 μ M) in 10 mM Tris–HCl pH 8, 100 mM NaCl buffer were diluted to 2 ml in 100 mM citrate-phosphate buffer at pH 4.0, 4.2, 4.4, 4.6, 4.8, 5.0, 5.4, 5.8, 6.2, 6.6 and 7.0. These were transferred to glass tubes, incubated overnight at 37 °C while shaking (200 rpm) and the presence of ring formation on the glass tube was visually assessed. Buffer from tubes containing rings formed at pH 4.0 was then removed without disturbing the rings and 3 ml of 100 mM citrate-phosphate buffer over the same pH range was added. These were left overnight with shaking at 37 °C and solubilisation of the rings were visually assessed the next day.

Mass spectrometry

Ring aggregates formed from either bacterial cultures or recombinant SsIE were scraped from glass tubes and resuspended in 0.5 ml 100 mM citrate-phosphate buffer, pH 4.0. Samples were resolved using SDS PAGE and after staining, SsIE fibre bands retained in the wells were excised and incubated with 10 mM dithiothreitol at 56 °C and then alkylated with 55 mM iodoacetamide at room temperature. Samples were digested using 1:20 dilution of bovine trypsin incubated in a shaking heat block at 37 °C for 16 h. Peptides were extracted with aqueous dehydration/hydration using acetonitrile and 50 mM triethylammonium bicarbonate, pooled and dried. Samples were resuspended in 2% (v/v) acetonitrile, 0.05% (v/v) formic acid and peptides were resolved by reversed-phase chromatography on a 75 μ m C18 Pepmap column (50 cm length) using a three-step linear gradient of 80% acetonitrile in 0.1% formic acid (U3000 UHPLC NanoLC system; ThermoFisherScientific, UK). The gradient was delivered to elute the peptides at a flow rate of 250 nl/min over 60 min starting at 5% B (0–5 min) and increasing solvent to 40% B (5–40 min) prior to a wash step at 99% B (40–45 min) followed by an equilibration step at 5% B (45–60 min). The eluate was ionised by electrospray ionisation using an Orbitrap Fusion Lumos (ThermoFisherScientific, UK) operating under Xcalibur v4.1.5. The instrument was first programmed to acquire using an Orbitrap-Ion Trap method by defining a 3 s cycle time between a full MS scan and MS/MS fragmentation. Orbitrap spectra (FTMS1) were collected at

a resolution of 120,000 over a scan range of m/z 375–1500 with an automatic gain control (AGC) setting of 4×10^5 with a maximum injection time of 35 ms. Monoisotopic precursor ions were filtered using charge state (+2 to +7) with an intensity threshold set between 5×10^3 and 1×10^{20} and a dynamic exclusion window of 35 secs \pm 10 ppm. MS2 precursor ions were isolated in the quadrupole set to a mass width filter of 1.6 m/z . Ion trap fragmentation spectra (ITMS2) were collected with an AGC target setting of 1×10^4 with a maximum injection time of 35 ms with CID collision energy set at 35%. Data were processed using Proteome Discoverer (v2.2; ThermoFisher) to search against Uniprot Swissprot All Taxonomy (561,911 entries) and the sequence of SsIE (Uniprot Accession number—E3PJ90) with Mascot search algorithm (v2.6.0; www.matrixscience.com) and the Sequest search algorithm⁶⁸. Precursor mass tolerance was set to 20 ppm with fragment mass tolerance set to 0.8 Da with a maximum of two missed cleavages. Variable modifications included: Carbamidomethylation (Cys) and Oxidation (Met). Database generated files (.msf) were uploaded into Scaffold software (v 4.10.0; www.proteomesoftware.com) for visualisation of fragmentation spectra.

ThT binding

In all, 15 μ l of 50 μ M rSsIE in 10 mM Tris–HCl pH 8, 100 mM NaCl buffer was mixed with 1.5 μ l of 1 mM Thioflavin-T (ThT) dye and transferred to a 96-well plate. 100 mM citrate-phosphate buffer between pH 4.0 and 8.0 was added to a final volume of 150 μ l and fluorescence data were collected (excitation/emission 438/480 nm) at 37 °C with shaking every 15 min over 24 h using a BMG CLARIOstar plate reader.

SsIE fibre immunoblot

Ring aggregates formed from rSsIE were scraped from glass tubes and resuspended in 0.5 ml 100 mM citrate-phosphate buffer, pH 4.0, centrifuged at 15,000 $\times g$ and then the top 950 μ l solution was carefully removed and discarded. This was followed by three rounds of addition of 950 μ l 100 mM citrate-phosphate buffer at pH 4.0, centrifugation at 15,000 $\times g$ and the top 950 μ l discarded. The final 50 μ l sample was mixed with 1 \times NuPAGE LDS Sample Buffer (ThermoFisher), 5% (v/v) β -mercaptoethanol and incubated at 100 °C for 5 min prior loading. This was run on a Criterion 4–20% SDS-PAGE gel (Bio-Rad), followed by transfer onto a PVDF membrane using the semi-dry Invitrogen Power-Blotter and Power-Blotter transfer blotting solution. The membrane was blocked in 1% (w/v) BSA, PBS-Tween for 1 h at room temperature followed by the addition of 1:2000 dilution mouse anti-His₆ antibody (Sigma) in 0.5% (w/v) BSA, PBS-Tween incubation buffer for 2 h. After five rounds of washing with incubation buffer, the membrane was incubated with 1:2000 anti-mouse HRP-conjugated antibody (Sigma) for 1 h, followed by five further washes and then treatment with ELC substrate (Pierce) before detection. Raw immunoblots are provided in Supplementary Fig. 23a.

Far-UV CD spectroscopy

Far-UV CD spectra were measured on a Chirascan (Applied Photophysics) spectropolarimeter thermostated at 25 °C. rSsIE (10 mg/ml) in 20 mM Tris–HCl pH 8, 50 mM NaCl was diluted 200-fold to 0.05 mg/ml in 10 mM MOPS at either pH 4.4 or 7.4. Spectra were recorded from 260 to 190 nm, at 0.2 nm intervals, 1 nm bandwidth, and a scan speed of 50 nm/min. Three accumulations were averaged for each spectrum. Deconvolution of data was performed using the BESTSEL server⁶⁹.

ATR FT-MIR spectroscopy

FT-MIR spectroscopy measurements were acquired using a Perkin Elmer Frontier spectrometer (Perkin Elmer, Buckinghamshire, UK) in attenuated total reflectance (ATR) mode, fitted with a Golden Gate ATR accessory (Specac Ltd., Kent, UK) with a diamond crystal. Samples were enclosed in a sealed Plexiglas container within a nitrogen-filled atmosphere at 23 ± 1 °C with additional calcium chloride dihydrate pellets (Sigma). Spectra were acquired using an infrared light source and a DTGS (deuterated triglycine sulphate) detector over a spectral range of 500–4000 cm^{-1} with a 4 cm^{-1} spectral resolution (filling factor = 2) and 64 co-added scans. 5 μ l rSsIE solution, as either soluble monomer in PBS buffer (pH 7.4) or aggregates in 100 mM phosphate-citrate buffer (pH 4.4), was deposited and dried onto the ATR crystal ($n = 3$ per protein solution). Spectral data were background-subtracted and baseline corrected using the Spectrum One software package (version 6, Perkin Elmer, Buckinghamshire, UK).

Bacterial culture ring assay

In total, 10 ml of LB containing 0.8 mg/ml Congo red dye, adjusted to either pH 5.0 or 7.0 with HCl, were inoculated with wild-type *E. coli* W and derivatives (including 50 μ g/ml kanamycin for Δ sse mutant and 50 μ g/ml kanamycin, 25 μ g/ml chloramphenicol for Δ sse::pCPC1 (sseI) and Δ sse::pCPC2 (sseI Δ M60)). These were incubated overnight at 37 °C while shaking (200 rpm) and the presence of ring formation on the glass tube was visually assessed.

SsIE secretion

E. coli W strains were grown on LB agar with 50 μ g/ml kanamycin for Δ sse mutants and 50 μ g/ml kanamycin, 25 μ g/ml chloramphenicol for Δ sse::pCPC1 (sseI) and Δ sse::pCPC2 (sseI Δ M60). Single colonies were resuspended in 10 ml LB (with appropriate antibiotics) and incubated overnight at 37 °C with shaking (200 rpm). Cells were centrifuged to separate out the media (supernatant) and pellet (whole cells) and 3 μ l of samples were resuspended in 17 μ l of 1 \times NuPAGE LDS Sample Buffer (ThermoFisher), 5% (v/v) β -mercaptoethanol. Samples were run on a Criterion 4–20% SDS-PAGE gel (Bio-Rad), followed by transfer onto a PVDF membrane using the semi-dry Invitrogen Power-Blotter and Power-Blotter transfer blotting solution. The membrane was blocked with 1% (w/v) BSA, PBS-Tween for 1 h at room temperature, and then incubated overnight at 4 °C with polyclonal anti-rSsIE antibody (rabbit; Invitrogen) or monoclonal anti-DsbA (mouse; Invitrogen), diluted 1:1000 using 0.5% (w/v) BSA, PBS-Tween incubation buffer. After three, 5 min washes with incubation buffer, membranes were incubated for 1 h at 37 °C with either anti-rabbit or anti-mouse secondary antibody conjugated to HRP (1:2000 dilution; Invitrogen) for 1 h at 25 °C and then treated with enhanced chemiluminescence substrate (ECL; Pierce) before detection.

Immunolectron microscopy

Overnight cultures were incubated for 30 min in 20 mM citrate-phosphate buffer at either pH 4.0 or 5.0 and then washed in the same buffer before fixing with the 3% (w/v) paraformaldehyde for 1 h. Cells were then loaded onto a Glow-discharged carbon-coated Ni grid (Agar Scientific) for 10 min, washed with 50 mM glycine and then blocked with 1% (v/v) Natural Donkey Serum (Jackson ImmunoResearch) for 30 min. This was then incubated for 1 h with either primary polyclonal rSsIE antibody (rabbit; Invitrogen) or primary polyclonal CsgA antibody (guinea pig; Invitrogen) diluted 1:100 in blocking buffer, washed five times for 3 min each with 0.05% Natural Donkey Serum and then incubated for 1 h with gold-conjugated secondary antibody (donkey; Jackson ImmunoResearch) diluted 1:100 in blocking buffer. Washing was repeated and the samples were negatively stained with 2% (w/v) uranyl acetate for 30 s, followed by two quick washes with ddH₂O. The grid was air-dried, and images were recorded on a JEM-1230 (JEOL-Japan) at 80 kV with a Morada CCD camera, iTEM software (EMSIS). Images of 13 *E. coli* W wild-type and 13 Δ sse mutant bacteria were then masked and gold particles were counted between \sim 0.2 and 1.0 μ m from the bacterium.

TEM fibre analysis

rSsIE at 1 mg/ml was incubated in 100 mM citrate-phosphate pH 3.8 overnight at room temperature while shaking at 180 rpm. This was centrifuged at 15,000 $\times g$ for 10 min, the top 80% of buffer discarded and then 4 μ l of the remaining sample was applied to a previously glow-discharged 300 mesh continuous carbon-coated copper grids (Agar Scientific Ltd) and immediately blotted for excess liquid. In all, 4 μ l of 2% (v/v) uranyl acetate was applied for staining for 10 s. The excess liquid was blotted and left to dry. Data were acquired using a JEOL JEM-1230 TEM operating at 80 kV equipped with a Morada 2k CCD camera system and its iTEM software package (Olympus Europa, UK). Micrographs were collected at a nominal magnification of 80,000 \times with a pixel size of 5.96 \AA /pixel.

RT-MALS

In total, 1 ml of rSsIE (1 mg/ml) in 25 mM citrate-phosphate buffer pH 4.0 was incubated overnight at room temperature while shaking at 180 rpm. The sample was then centrifuged at 15,000 $\times g$ and the top 850 μ l of buffer was discarded. MALS experiments were then performed on beamline B21 at the DLS (Oxford, UK). In total, 20 μ l of SsIE fibres were directly injected into the RT-MALS multiple times at a flow rate of 0.05 ml/min. Detectors were standardised through a direct injection using BSA. Data were analysed using ASTRA version 6.1.7. (Wyatt).

Solid-state NMR spectroscopy

rSsIE at 20 mg/ml was buffer exchanged into 10 mM citrate-phosphate pH 4.0 using a PD10 column (Sigma) and then incubated overnight at room temperature while shaking at 180 rpm. This was centrifuged at $15,000 \times g$ for 10 min, the top 80% of buffer was discarded and then the remaining sample was used for solid-state NMR analysis. Experiments were performed using a Bruker Neo Console operating at 850 MHz ^1H frequency with a 3.2 mm E-Free probe in HC mode spinning at a rate $\omega_r = 15$ kHz. A standard CP excitation ^{13}C - ^{13}C DARR experiment was acquired. The direct dimension was acquired for 16.4 ms with ~ 55 kHz SPINAL-64 decoupling. 256 rows with 256 co-added transients and a recovery delay of 2.5 s were acquired using Time-Proportional Phase Increment (TPPI) with a 22.2 μs dwell (45 kHz sweep width, 2.84 ms total evolution) for a total of 45.5 h total acquisition time. The applied power was adjusted so that ^1H and ^{13}C hard pulses were both 4 μs ($\omega_{1(\text{H,C})} = 62.5$ kHz). The initial carbon excitation was achieved with 1.5 ms of ramped Hartmann–Hahn CP, where the optimal polarisation transfer was found at $\omega_{1\text{C}} = \sim 70$ kHz and $\omega_{1\text{H}} = \sim 55$ kHz with an upwards linear ramp from 70–100% on the ^1H channel. Homonuclear Carbon mixing was achieved with 50 ms of DARR mixing ($\omega_{1\text{H}} = \omega_r = 15$ kHz).

SAXS fibrillation analysis

SAXS data were collected on beamline B21⁵⁹ at DLS (Oxford, UK) at 25 °C. Immediately prior to data collection, 10 mg/ml rSsIE in 20 mM Tris–HCl pH 8, 200 mM NaCl was buffer exchanged into 20 mM citrate-phosphate buffer, pH 4.4 using a PD10 column (GE Healthcare) and the flow-through was used as a buffer reference. Using a peristaltic pump and while constantly stirring, 5 ml of rSsIE (0.7 mg/ml) was circulated through the SAXS imaging cuvette and data were collected every 30 min over a momentum transfer range of $0.004 < q < 0.4 \text{ \AA}^{-1}$, with the initial scattering data captured at 30 min after initiating fibre growth. Data collected over the course of 11 h, consisting of 22 profiles, was decomposed using COSMiCS³⁰, which utilises MCR-ALS⁷⁰ to perform model-free decomposition of the entire SAXS dataset. The SEC-SAXS curve collected at pH 4.4 was also included in the dataset as the representative state of the protein at time 0 s, i.e. before initiating fibrillation. The time 0 s curve was selected as one of the initial estimates and the selectivity restraint was used to ascertain that the curve had no contribution from the other species. In addition, non-negativity restraint was imposed for both the SAXS profiles and concentrations using FNNLS algorithm⁷¹. Before COSMiCS analysis, SAXS data were scaled according to the large angle data to enhance the decomposition capacity of the approach. Although COSMiCS was run assuming a two-component system as suggested by Principal Component Analysis (PCA), a three-species run was also performed. The convergence criterion of $< 0.01\%$ change in lack of fit was used with 1000 maximum allowed iterations. The analysis of the resulting COSMiCS curves was performed with the ATSAS suite of programmes⁶⁰. Fractal analysis was carried out using SASVIEW version 4.2.2 (<http://www.sasview.org/>). The programmes ATSAS⁶⁰ and Scatter were also used to obtain a cross-sectional radius of gyration R_c from the 10 h post-fibre induction scattering profile and from this the cross-sectional $P(r)_c$ was calculated. BODIES⁶⁰ was then used to approximate the geometric shape of rSsIE fibres, which suggested a cylindrical shape with radius of 3.93 nm and a height of 16.5 nm (χ^2 1.6). Processing and refinement statistics can be found in Supplementary Table 1.

Macrocolony biofilm growth

In all, 5 μl of wild-type *E. coli* W, BL21 (DE3), H10407 and E2348/69 strains and their derivatives were grown on LB agar, YESCA agar or M9-glucose agar containing either 0.8 mg/ml Congo red dye or 50 μM Fluorescent Brightener 28 (Sigma, UK) at 37 °C, 25 °C or 18 °C for either 24 or 96 h. In total, 50 $\mu\text{g}/\text{ml}$ kanamycin was included for W and H10407 *ΔssIE*, E2348/69 *ΔcsgA*, BW25113 *pgaC* and BW25113 *wcaF* mutant strains, and 50 $\mu\text{g}/\text{ml}$ kanamycin, 25 $\mu\text{g}/\text{ml}$ chloramphenicol was included for the E2348/69 *ΔcsgA/ΔbcsA* mutant and W *ΔssIE::pCPC1 (ssIE)* and *ΔssIE::pCPC2 (ssIEΔM60)*.

Microfluidic biofilm growth

Biofilms were grown in a BioFlux 200 device (Labtech, UK) using a protocol adapted from the manufacturer. Single colonies of *E. coli* W and its derivatives were resuspended in 10 ml of LB (with additional 50 $\mu\text{g}/\text{ml}$ kanamycin for *ΔssIE* mutant and 50 $\mu\text{g}/\text{ml}$ kanamycin, 25 $\mu\text{g}/\text{ml}$

chloramphenicol *ΔssIE::pCPC1 (ssIE)* and *ΔssIE::pCPC2 (ssIEΔM60)*) and incubated at 37 °C with shaking (200 rpm) for 16 h. All subsequent media contained appropriate antibiotics where required. The cultures were then diluted (1:100) in 10 ml of LB to $\text{OD}_{600\text{nm}}$ of 0.05. A 24-well Bioflux plate (0–20 dyn/cm^2) was prewarmed to 37 °C on a heated stage and microfluidic channels were incubated with prewarmed LB. The channels were then inoculated by injecting 20 μl of bacterial suspension into the output reservoir for 5 s at 2 dyne/cm^2 . The microfluidic plate was incubated for 1 h at 37 °C to allow bacteria to bind to the surface and then the flow reversed for 20 s. Prewarmed LB was then added to the input reservoir, the flow of media was initiated at 2 dyne/cm^2 for 5 min and then decreased to 1 dyne/cm^2 for 18 h. The spent media was removed from the output reservoir, fresh prewarmed media was added to the input reservoir and the flow was lowered to 0.5 dyne/cm^2 for up to 96 h.

Assessment of biofilm growth

FilmTracer FM 1–43 dye (Life Technologies) was added to the BioFlux 200 input reservoir mixed with prewarmed LB to a final concentration of 0.3 μM . The heated stage holding the BioFlux plate was then moved to a DM-IRE2 confocal laser scanning microscope (Leica Microsystems Heidelberg GmbH, Germany) for image acquisition with the Leica Microsystems Confocal Software (version 2.61 Build 1537). The fluorophore was excited at 488 nm and emission collected at 580 nm as recommended by the manufacturer.

Biofilm pH measurement

The ratiometric dye seminaphthorhodafluor-4F 5-(and-6) carboxylic acid (C-SNARF-4) was used to directly measure pH across biofilms within the Bioflux channel, using a modified method⁷². HEPES buffer was first adjusted between pH 3.2 and 7.8 in 0.2 increments using HCl, and 95 μl of each respective solution was then mixed with 5 μl 1 mM C-SNARF-4, resulting in a final 50 μM concentration of C-SNARF-4. This solution was flushed into the Bioflux plate channel, prewarmed at 37 °C and the channels were imaged at $\times 64$ magnification using a DM-IRE2 confocal laser scanning microscope (Leica Microsystems Heidelberg GmbH, Germany) and the accompanying Leica Microsystems Confocal Software (version 2.61 Build 1537). Simultaneous images were captured using an excitation wavelength of 488 nm and emission wavelengths of 580 nm and 640 nm, in triplicate for each pH increment with a background image obtained after each acquisition. Acquired images were analysed using ImageJ Software (FIJI)⁷³ by determining the average intensity of fluorescence and standard deviation of each channel, minus the background control. The intensity ratio for each pH was used to make a standard curve by plotting known pH against green/red ratio. 50 μM C-SNARF-4 was then flushed into Bioflux plates containing *E. coli* W biofilm after either 24 or 96 h growth and five separate images were acquired for each Bioflux channel to calculate the pH from different areas of the biofilm. The fluorescence intensity and standard deviation were recorded, and the ratiometric value was used to determine pH from the standard curve. A 12×12 grid ($\sim 30 \mu\text{m}^2$ area) was applied to each image; each box was used to define regions of interest (ROI) and an average measure of pH was calculated within each region. The average pH for ROIs within the biofilm fringes and centres is presented.

Cellulose quantification

A lawn of wild-type *E. coli* W, BL21 (DE3), H10407 and E2348/69 strains or their derivatives were scraped from the surface of LB agar or YESCA agar plates after 96 h at 37 °C or 25 °C. For each strain, ~ 30 mg of cells was transferred into 400 μl of 2.5 M 2-(N-Morpholino)-ethanesulfonic acid (MES) buffer pH 5.5 with or without 6 U/ml cellulase (Sigma, UK) and incubated at 37 °C overnight. Each sample was adjusted to a final turbidity of 40 at 600 nm with 2.5 mM MES pH 5.5 and then centrifuged at $5000 \times g$ for 20 min. The glucose concentration in each sample was measured using the glucose (HK) assay kit (Sigma, UK) and compared with data from none cellulase treated samples.

Curli and SsIE immunoblot

A lawn of wild-type *E. coli* W, BL21 (DE3), H10407 and E2348/69 strains or their derivatives were scraped from the surface of LB agar or YESCA agar plates after 96 h at 37 °C or 25 °C and transferred to 1 ml PBS and the volume adjusted to an $\text{OD}_{600\text{nm}} = 1.0$. Formic acid was added to 70% (v/v) and then lyophilised overnight. Samples were resuspended in 200 μl 1 \times

NuPAGE LDS loading buffer (ThermoFisher) and run on a Criterion 4–20% SDS-PAGE gel (Bio-Rad), followed by transfer onto a PVDF membrane using the semi-dry Invitrogen Power-Blotter and Power-Blotter transfer blotting solution. The membrane was blocked with 3% (w/v) BSA, PBS-Tween for 1 h at room temperature followed by the addition of 1:1000 dilution polyclonal primary anti-rSsIE antibody (rabbit; Invitrogen) or 1:5000 dilution polyclonal anti-CsgA antibody (guinea pig; Invitrogen) in 0.5% (w/v) BSA, PBS-Tween incubation buffer for 1 h. After 3 rounds of washing with incubation buffer, the membrane was incubated with 1:2000 HRP-conjugated anti-rabbit or anti-guinea pig antibody (Sigma), respectively, for 1 h, followed by three further washes and then treatment with ELC substrate (Peirce) before detection. Raw immunoblots are provided in Supplementary Fig. 23b, c.

PNAG dot-blot and overlay assay

A lawn of wild-type *E. coli* W, BL21 (DE3), H10407, E2348/69 and BW25113 strains or their derivatives were scraped from the surface of LB agar plates after 96 h at 37 °C or 25 °C. For each strain, bacteria were transferred into 3 ml of 0.5 M EDTA (pH 8.0) per g of cells, followed by 5 min incubation at 100 °C. Samples were centrifuged at 10,500 × g for 6 min and then 100 µl of supernatant was incubated with 10 µl of proteinase K (20 mg/ml; NEB, UK) for 60 min at 60 °C, followed by 80 °C for 30 min. This was diluted threefold with ddH₂O and then spotted onto a PVDF membrane. The membrane was dried, blocked with 3% (w/v) BSA, PBS-Tween for 1 h at room temperature and then incubated with 70 µg/ml wheat germ agglutinin Biotin-conjugated (WGA-biotin; Sigma, UK) in 0.5% (w/v) BSA, PBS-Tween for 1 h at room temperature. After four rounds of washing with incubation buffer, the membrane was incubated with 1:2000 HRP-conjugated streptavidin (Sigma, UK) for 1 h, followed by 3 further washes and then treatment with ELC substrate (Peirce) before detection. For overlay assays membranes were blocked with 3% (w/v) BSA in 100 mM citrate-phosphate buffer pH 6.0, 0.05% (w/v) Tween-20 and then incubated with either 30 µM rSsIE monomer, rSsIE fibre (produced as described for immunoblotting) or RgpB-CTD control, all in 0.5% (w/v) BSA, 100 mM citrate-phosphate buffer pH 6.0, 0.05% (w/v) Tween-20, for 1 h at room temperature. After four rounds of washing with 0.5% (w/v) BSA, 100 mM citrate-phosphate buffer pH 6.0, 0.05% (w/v) Tween-20, the membrane was incubated with 1:2000 anti-His₆ HRP-conjugated antibody (ThermoFisher, UK) in the same buffer for 1 h, followed by 4 further washes and treatment with ELC substrate (Peirce) before detection.

Colanic acid quantification

A lawn of wild-type *E. coli* W, BL21 (DE3), H10407, E2348/69 and BW25113 strains or their derivatives were scraped from the surface of LB agar or M9-glucose agar plates after 96 h at 37 °C, 25 °C or 18 °C and resuspended in ddH₂O to a final OD_{600nm} = 1.0. After incubating at 100 °C for 10 min and centrifuging at 16,000 × g for 10 min, 100 µl of supernatant was then diluted to 1 ml with ddH₂O and mixed with 4.5 ml H₂SO₄/H₂O (6:1; v/v). This was boiled at 100 °C for 20 min, cooled to room temperature, divided in two and then 50 µl ddH₂O (control) was added to one sample and 50 µl 1.0 M cysteine hydrochloride (Sigma, UK) added to the other sample. Absorbance was measured at 396 nm and 427 nm and then control subtracted from the cysteine hydrochloride values. A standard curve of L-fucose (Sigma, UK) was produced using concentrations ranging from 5 to 100 µg/ml plotted against absorbance at 396 nm and 427 nm. This was used to determine the amount of fucose released from colanic acid samples, which were then scaled and normalised based on the cell density measurements.

Cellulose-binding assay

An Immulon 2-HB 96-well plate (VWR) was blocked for 1 h at 25 °C with 300 µl of 0.1% (w/v) BSA in PBS-Tween and then washed once with 300 µl of incubation buffer (100 mM citrate-phosphate buffer pH 6.0, 0.05% (w/v) BSA, 0.05% Tween-20). One *Komagataeibacter rhaeticus* cellulose disc⁷⁴ was added to each well and then 200 µl of incubation buffer was added to cover the discs, followed by incubation for 5 min. The discs were washed twice in incubation buffer and then 200 µl of either 100 µM rSsIE monomer, rSsIE fibre (produced as described for immunoblotting) or RgpB-CTD control, all in incubation buffer, were added to the cellulose discs. The plate was incubated for 3 h at 25 °C and then the discs were transferred carefully to new pre-blocked wells and washed with 200 µl incubation buffer. After repeating three times, the discs were incubated with anti-His₆ HRP-conjugated antibody (ThermoFisher) diluted 1:2000 with incubation

buffer for 1 h at 24 °C. The discs were removed again, and the washing step was repeated three times. Finally, the solution was removed from the wells and the discs were incubated with 150 µl of *o*-Phenylenediamine dihydrochloride (Sigma) for 30 min. The solution in the wells was stirred thoroughly, the discs were removed and then data were recorded at 450 nm.

Electrophoretic mobility shift assay

In all, 10 nM of pUC19 plasmid DNA (NEB, UK) was mixed with 0, 12.5, 25, 50 and 100 µM of either rSsIE monomer, rSsIE fibre (produced as described for immunoblotting) or RgpB-CTD control in 10 mM 2-(N-morpholino) ethanesulfonic acid (MES) pH 6.0, 50 mM NaCl, 10% (v/v) glycerol, and incubated for 1 h at room temperature. The DNA-fibre sample was then mixed with 60% glycerol, 0.01% (v/v) bromophenol blue in MES buffer pH 6.0, while DNA-monomeric rSsIE and RgpB-CTD samples were mixed with “purple no SDS DNA loading dye” at pH 8.0 (NEB, UK). Samples were loaded onto a 1% (w/v) agarose gel containing GelRed (Biotium, UK) in TBE buffer and run at 70 V for 45 min at 4 °C and visualised using UV light.

Biofilm immunofluorescence

Single colonies of wild-type *E. coli* W and *ΔssIE* mutant control were resuspended in 10 ml of LB and incubated at 37 °C with shaking (200 rpm) for 16 h. Cultures were diluted (1:100) in prewarmed LB and 1 ml was injected onto borosilicate glass coupons held within FC310 flow cells (Biosurface Technology) and incubated for 1 h. The flow of LB media was initiated at 0.25 ml/min and maintained for 96 h. Coupons were removed and washed three times for 5 min each in PBS. Coupons were then blocked for 1 h in 2% (w/v) BSA, PBS at room temperature and incubated overnight at 4 °C with 10 µg/ml polyclonal anti-rSsIE antibody (rabbit; Invitrogen) in 0.1% (w/v) BSA, PBS. After three 5 min washes in PBS, coupons were incubated in the dark for 1 h with either 1 µM TOTO-1 (Invitrogen) or 1 µg/ml FM1-43 (Invitrogen), and anti-rabbit Alexa Fluor 633 secondary antibody (goat; Invitrogen) diluted 1:500 in 0.1% (w/v) BSA, PBS. This was followed by three 5 min washes in the dark in PBS and then images were captured at ×20 magnification using an excitation wavelength of 488 nm (TOTO-1), 488 nm (FM1-43), and 633 nm (Alexa Fluor 633), and emission wavelengths of 534 nm, 581 nm and 650 nm, respectively, with a DM-IRE2 confocal laser scanning microscope (Leica Microsystems Heidelberg GmbH, Germany). Additional negative controls consisted of *E. coli* W biofilms being incubated with either no primary anti-rSsIE antibody or no Alexa Fluor 633 secondary antibody.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

The data that support the findings of this study are included in the article, its supplementary information files, or are available from the corresponding author upon reasonable request. The plasmids pPC1, pPC2, pPC3, pPC4, pPC5, pBD1, pCPC1 and pCPC2 have been deposited in the Addgene plasmid repository (<https://www.addgene.org>) with IDs 182013, 182014, 182015, 182016, 182017, 182018, 182019 and 182020, respectively. The negative-stain TEM electron density map for SsIE pH 7.4 has been deposited in the Electron Microscope Data Bank (<https://www.ebi.ac.uk/emdb>) with accession code EMD-14170. The buffer subtracted SAXS curves and DAMMIN models with the lowest NSD score for monomeric SsIE at pH 4.4 and 7.4, and the SsIE N1, N2 and N3-M60 domains, have been deposited in the Small-Angle Scattering Biological Data Bank (<https://www.sasbdb.org/>) with accession ID SASDLW2, SASDLV2, SASDMU6, SASDMV6 and SASDMW6, respectively.

Received: 31 March 2021; Accepted: 31 January 2022;

Published online: 25 February 2022

REFERENCES

1. Kaper, J. B., Nataro, J. P. & Mobley, H. L. Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* **2**, 123–140 (2004).
2. Gu, S., Shevchik, V. E., Shaw, R., Pickersgill, R. W. & Garnett, J. A. The role of intrinsic disorder and dynamics in the assembly and function of the type II

- secretion system. *Biochim Biophys. Acta Proteins Proteom.* **1865**, 1255–1266 (2017).
3. Moriel, D. G. et al. Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **107**, 9072–9077 (2010).
 4. Baldi, D. L. et al. The type II secretion system and its ubiquitous lipoprotein substrate, SsIE, are required for biofilm formation and virulence of enteropathogenic *Escherichia coli*. *Infect. Immun.* **80**, 2042–2052 (2012).
 5. Decanio, M. S., Landick, R. & Haft, R. J. The non-pathogenic *Escherichia coli* strain W secretes SsIE via the virulence-associated type II secretion system beta. *BMC Microbiol.* **13**, 130 (2013).
 6. Nesta, B. et al. SsIE elicits functional antibodies that impair in vitro mucinase activity and in vivo colonization by both intestinal and extraintestinal *Escherichia coli* strains. *PLoS Pathog.* **10**, e1004124 (2014).
 7. Yang, J., Baldi, D. L., Tauschek, M., Strugnell, R. A. & Robins-Browne, R. M. Transcriptional regulation of the yghJ-pppA-yghG-gspCDEFGHIJKLM cluster, encoding the type II secretion pathway in enterotoxigenic *Escherichia coli*. *J. Bacteriol.* **189**, 142–150 (2007).
 8. Naili, I. et al. Mixed mucosal-parenteral immunizations with the broadly conserved pathogenic *Escherichia coli* antigen SsIE induce a robust mucosal and systemic immunity without affecting the murine intestinal microbiota. *Vaccine* **37**, 314–324 (2019).
 9. Luo, Q. et al. Conservation and immunogenicity of novel antigens in diverse isolates of enterotoxigenic *Escherichia coli*. *PLoS Negl. Trop. Dis.* **9**, e0003446 (2015).
 10. Valeri, M. et al. Pathogenic *E. coli* exploits SsIE mucinase activity to translocate through the mucosal barrier and get access to host cells. *PLoS ONE* **10**, e0117486 (2015).
 11. Luo, Q. et al. Enterotoxigenic *Escherichia coli* secretes a highly conserved mucin-degrading metalloprotease to effectively engage intestinal epithelial cells. *Infect. Immun.* **82**, 509–521 (2014).
 12. Rehman, S. et al. Structure and functional analysis of the *Legionella pneumophila* chitinase ChiA reveals a novel mechanism of metal-dependent mucin degradation. *PLoS Pathog.* **16**, e1008342 (2020).
 13. Corfield, A. P. Mucins: a biologically relevant glycan barrier in mucosal protection. *Biochim Biophys. Acta* **1850**, 236–252 (2015).
 14. Garnett, J. A. & Matthews, S. Interactions in bacterial biofilm development: a structural perspective. *Curr. Protein Pept. Sci.* **13**, 739–755 (2012).
 15. Voegelé, P., Tremblay, Y. D., Mafu, A. A., Jacques, M. & Harel, J. Life on the outside: role of biofilms in environmental persistence of Shiga-toxin producing *Escherichia coli*. *Front. Microbiol.* **5**, 317 (2014).
 16. Chitlapilly Dass, S. et al. Impact of mixed biofilm formation with environmental microorganisms on *E. coli* O157:H7 survival against sanitization. *npj Sci. Food* **4**, 16 (2020).
 17. Rossi, E. et al. “It’s a gut feeling”—*Escherichia coli* biofilm formation in the gastrointestinal tract environment. *Crit. Rev. Microbiol.* **44**, 1–30 (2018).
 18. Nakjang, S., Ndeh, D. A., Wipat, A., Bolam, D. N. & Hirt, R. P. A novel extracellular metalloprotease domain shared by animal host-associated mutualistic and pathogenic microbes. *PLoS ONE* **7**, e30287 (2012).
 19. Noach, I. et al. Recognition of protein-linked glycans as a determinant of peptidase activity. *Proc. Natl Acad. Sci. USA* **114**, E679–E688 (2017).
 20. Trastoy, B., Naegeli, A., Anso, I., Sjogren, J. & Guerin, M. E. Structural basis of mammalian mucin processing by the human gut O-glycopeptidase OgpA from *Akkermansia muciniphila*. *Nat. Commun.* **11**, 4844 (2020).
 21. Buchan, D. W. A. & Jones, D. T. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407 (2019).
 22. Taglialegna, A., Lasa, I. & Valle, J. Amyloid structures as biofilm matrix scaffolds. *J. Bacteriol.* **198**, 2579–2588 (2016).
 23. Belousov, M. V. et al. M60-like metalloprotease domain of the *Escherichia coli* YghJ protein forms amyloid fibrils. *PLoS ONE* **13**, e0191317 (2018).
 24. Taglialegna, A. et al. The biofilm-associated surface protein Esp of *Enterococcus faecalis* forms amyloid-like fibers. *npj Biofilms Microbiomes* **6**, 15 (2020).
 25. Taglialegna, A. et al. Staphylococcal Bap proteins build amyloid scaffold biofilm matrices in response to environmental signals. *PLoS Pathog.* **12**, e1005711 (2016).
 26. Chen, D. et al. Characteristics and influencing factors of amyloid fibers in *S. mutans* biofilm. *AMB Express* **9**, 31 (2019).
 27. Gallardo, R., Ranson, N. A. & Radford, S. E. Amyloid structures: much more than just a cross-beta fold. *Curr. Opin. Struct. Biol.* **60**, 7–16 (2020).
 28. Moran, S. D. & Zanni, M. T. How to get insight into amyloid structure and formation from infrared spectroscopy. *J. Phys. Chem. Lett.* **5**, 1984–1993 (2014).
 29. Conway, K. A., Harper, J. D. & Lansbury, P. T. Jr. Fibrils formed in vitro from alpha-synuclein and two mutant forms linked to Parkinson’s disease are typical amyloid. *Biochemistry* **39**, 2552–2563 (2000).
 30. Herranz-Trillo, F. et al. Structural analysis of multi-component amyloid systems by chemometric SAXS data decomposition. *Structure* **25**, 5–15 (2017).
 31. Sagar, A., Herranz-Trillo, F., Langkilde, A. E., Vestergaard, B. & Bernado, P. Structure and thermodynamics of transient protein-protein complexes by chemometric decomposition of SAXS datasets. *Structure* **29**, 1074–1090 e1074 (2021).
 32. Schmidt, P. Small-angle scattering studies of disordered, porous and fractal systems. *J. Appl. Crystallogr.* **24**, 414–435 (1991).
 33. Hurd, A. J., Schaefer, D. W. & Martin, J. E. Surface and mass fractals in vapor-phase aggregates. *Phys. Rev. A Gen. Phys.* **35**, 2361–2364 (1987).
 34. Sharma, G. et al. *Escherichia coli* biofilm: development and therapeutic strategies. *J. Appl. Microbiol.* **121**, 309–319 (2016).
 35. Ferenci, T. et al. Genomic sequencing reveals regulatory mutations and recombinational events in the widely used MC4100 lineage of *Escherichia coli* K-12. *J. Bacteriol.* **191**, 4025–4029 (2009).
 36. Saldana, Z. et al. Synergistic role of curli and cellulose in cell adherence and biofilm formation of attaching and effacing *Escherichia coli* and identification of Fis as a negative regulator of curli. *Environ. Microbiol.* **11**, 992–1006 (2009).
 37. Baba, T. et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006 0008 (2006).
 38. Barnhart, M. M. & Chapman, M. R. Curli biogenesis and function. *Annu. Rev. Microbiol.* **60**, 131–147 (2006).
 39. Wu, H. et al. Activation of colanic acid biosynthesis linked to heterologous expression of the polyhydroxybutyrate pathway in *Escherichia coli*. *Int. J. Biol. Macromol.* **128**, 752–760 (2019).
 40. Okshesky, M. & Meyer, R. L. The role of extracellular DNA in the establishment, maintenance and perpetuation of bacterial biofilms. *Crit. Rev. Microbiol.* **41**, 341–352 (2015).
 41. Gish, W. & States, D. J. Identification of protein coding regions by database similarity search. *Nat. Genet.* **3**, 266–272 (1993).
 42. Ta, B. T. T. et al. Identification, recombinant protein production, and functional analysis of a M60-like metalloprotease, secreted by the liver fluke *Opisthorchis viverrini*. *Parasitol. Int.* **75**, 102050 (2020).
 43. Hidalgo, G. et al. Functional tomographic fluorescence imaging of pH micro-environments in microbial biofilms by use of silica nanoparticle sensors. *Appl. Environ. Microbiol.* **75**, 7426–7435 (2009).
 44. Van Gerven, N., Van der Verren, S. E., Reiter, D. M. & Remaut, H. The role of functional amyloids in bacterial virulence. *J. Mol. Biol.* **430**, 3657–3684 (2018).
 45. Shanmugam, N. et al. Microbial functional amyloids serve diverse purposes for structure, adhesion and defence. *Biophys. Rev.* **11**, 287–302 (2019).
 46. Dueholm, M. S. et al. Functional amyloid in *Pseudomonas*. *Mol. Microbiol.* **77**, 1009–1020 (2010).
 47. Evans, M. L. & Chapman, M. R. Curli biogenesis: order out of disorder. *Biochim Biophys. Acta* **1843**, 1551–1558 (2014).
 48. Pfefferkorn, C. M., McGlinchey, R. P. & Lee, J. C. Effects of pH on aggregation kinetics of the repeat domain of a functional amyloid, Pmel17. *Proc. Natl Acad. Sci. USA* **107**, 21447–21452 (2010).
 49. Morel, B., Varela, L., Azuaga, A. I. & Conejero-Lara, F. Environmental conditions affect the kinetics of nucleation of amyloid fibrils and determine their morphology. *Biophys. J.* **99**, 3801–3810 (2010).
 50. Zandomeneghi, G., Krebs, M. R., McCammon, M. G. & Fandrich, M. FTIR reveals structural differences between native beta-sheet proteins and amyloid fibrils. *Protein Sci.* **13**, 3314–3321 (2004).
 51. Groenning, M. et al. Study on the binding of Thioflavin T to beta-sheet-rich and non-beta-sheet cavities. *J. Struct. Biol.* **158**, 358–369 (2007).
 52. Serra, D. O., Richter, A. M. & Hengge, R. Cellulose as an architectural element in spatially structured *Escherichia coli* biofilms. *J. Bacteriol.* **195**, 5540–5554 (2013).
 53. Qadri, F., Svennerholm, A. M., Faruque, A. S. & Sack, R. B. Enterotoxigenic *Escherichia coli* in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. *Clin. Microbiol. Rev.* **18**, 465–483 (2005).
 54. Chen, H. D. & Frankel, G. Enteropathogenic *Escherichia coli*: unravelling pathogenesis. *FEMS Microbiol. Rev.* **29**, 83–98 (2005).
 55. Fallingborg, J. Intraluminal pH of the human gastrointestinal tract. *Dan. Med Bull.* **46**, 183–196 (1999).
 56. Xu, Y. et al. An acid-tolerance response system protecting exponentially growing *Escherichia coli*. *Nat. Commun.* **11**, 1496 (2020).
 57. Lin, J. et al. Mechanisms of acid resistance in enterohemorrhagic *Escherichia coli*. *Appl. Environ. Microbiol.* **62**, 3094–3100 (1996).
 58. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
 59. Cowieson, N. P. et al. Beamline B21: high-throughput small-angle X-ray scattering at diamond light source. *J. Synchrotron Radiat.* **27**, 1438–1446 (2020).
 60. Franke, D. et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).
 61. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).

62. Rohou, A. & Grigorieff, N. CTFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
63. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
64. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
65. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
66. Vranken, W. F. et al. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696 (2005).
67. Johnson, B. A. & Blevins, R. A. NMR view: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–614 (1994).
68. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
69. Micsonai, A. et al. BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res.* **46**, W315–W322 (2018).
70. Jaumot, J., Gargallo, R., de Juan, A. & Tauler, R. A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemometrics Intell. Lab. Syst.* **76**, 101–110 (2005).
71. Bro, R., De & Jong, S. A fast non-negativity-constrained least squares algorithm. *J. Chemometrics* **11**, 393–401 (1997).
72. Schlafer, S. & Dige, I. Ratiometric imaging of extracellular pH in dental biofilms. *J. Vis. Exp.* **109**, e53622 (2016).
73. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
74. Florea, M. et al. Engineering control of bacterial cellulose production using a genetic toolkit and a new cellulose-producing strain. *Proc. Natl Acad. Sci. USA* **113**, E3431–E3440 (2016).

ACKNOWLEDGEMENTS

P.C. was supported by the Academy of Medical Sciences and Wellcome Trust (SBF002/1150) and S.R. was supported by the Medical Research Council (MR/R017662/1), awarded to J.A.G. S.W. and B.D. were supported by China Scholarship Council and Biotechnology and Biological Sciences Research Council studentships, respectively, awarded to J.A.G. and M.A.C. L.S. was supported by the Leverhulme Trust (RPG-2017-222), awarded to J.A.G. K.F. was supported by a PhD studentship from Queen Mary University of London and V.C.D. was supported by a start-up grant from Queen Mary University of London. This work was also supported by the Labex EpiGenMed, an « Investissements d’avenir » program (ANR-10-LABX-12-01) awarded to P.B. The C.B.S. is a member of France-Biologymaging (FBI) and the French Infrastructure for Integrated Structural Biology (FRISBI), 2 national infrastructures supported by the French National Research Agency (ANR-10-INBS-04-01 and ANR-10-INBS-05, respectively). We thank the beamline scientists at B21 of the Diamond Light Source, United Kingdom. This work was also supported by the Francis Crick Institute through provision of access to the MRC Biomedical NMR centre. The Francis Crick Institute receives its core funding from Cancer Research United Kingdom (FC001029), the United Kingdom Medical Research Council (FC001029), and the Wellcome Trust (FC001029). As this research was funded in part by the Wellcome Trust, for the purpose of open access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. We also thank the Centre for Biomolecular Spectroscopy at King’s College London for additional NMR access, funded by the Wellcome Trust and British Heart Foundation (ref. 202767/

Z/16/Z and IG/16/2/32273, respectively). The UK 850 MHz solid-state NMR Facility used in this research was funded by EPSRC and BBSRC (contract reference EP/T015063/1), as well as the University of Warwick including via part funding through Birmingham Science City Advanced Materials Projects 1 and 2 supported by Advantage West Midlands (AWM) and the European Regional Development Fund (ERDF). We would like to thank Prof. Cynthia Whitchurch (Quadram Institute, UK) for providing the EPEC E2348/69 Δ sslE strain and Prof. Jorge Girón (Benemérita Universidad Autónoma de Puebla, Mexico) for providing the EPEC E2348/69 wild-type, Δ csgA and Δ csgA/ Δ bcsA strains. Cellulose discs were gifted by Prof. Tom Ellis and Kenneth T. Walker (Imperial College London, UK).

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: P.C., S.W., S.R., K.F., S.S., C.E.-G., A.S., G.M., B. D., L.M.S., S.L., D.I., P.B., V.C.D. and J.A.G. Performed the experiments: P.C., S.W., S.R., K. F., A.S., S.S., L.C., C.E.-G., G.M., B.D., L.M.S., S.L., D.I., J.J., G.H.C., V.C.D. and J.A.G. Analysed the data: P.C., S.W., S.R., K.F., A.S., S.S., C.E.-G., G.M., B.D., S.L., D.I., T.F., P.B., V.C.D. and J. A.G. Contributed reagents/materials/analysis tools: S.S., S.L., P.B., V.C.D. and J.A.G. Wrote the paper: P.C., S.W., S.R., K.F., A.S., S.S., L.C., C.E.-G., G.M., B.D., L.M.S., S.L., D.I., T. F., J.J., G.H.C., M.A.C., P.B., V.C.D. and J.A.G. P.C., S.W., S.R. and K.F. contributed equally to this work.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41522-022-00272-5>.

Correspondence and requests for materials should be addressed to Vidya C. Darbari or James A. Garnett.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022