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

Paula M. Corsini, Sunjun Wang, Saima Rehman, Katherine Fenn, Amin Sagar, Slobodan Sirovica, Leanne Cleaver, Charlotte J. C. Edwards-Gayle, James Jarvis, Guy H. Carpenter, Michael A. Curtis, Pau Bernadó, Vidya C. Darbari and James A. Garnett

*Escherichia coli* is a Gram-negative bacterium that colonises the human intestine and virulent strains can cause severe diarrhoeal and extraintestinal diseases. The protein SslE is secreted by a range of pathogenic and commensal *E. coli* strains. It can degrade mucus 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 SslE from the commensal *E. coli* Waksman and BL21 (DE3) strains and the enterotoxigenic H10407 and enteropathogenic E2348/69 strains. We reveal that SslE 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 SslE monomers and aggregates bind DNA in vitro and co-localise with extracellular DNA (eDNA) in mature biofilms, and SslE aggregates may also associate with cellulose under certain conditions. Our results suggest that interactions between SslE and eDNA are important for biofilm maturation in many *E. coli* strains and SslE may also be a factor that drives biofilm formation in other SslE-secreting bacteria.

**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 pneumonia. A wide range of pathogenic *E. coli* strains, and some commensals, use a *Vibrio*-like type II secretion system (T2SS) to translocate the protein SslE across their outer membrane and onto their extracellular surface. These include the Waksman (W), enterotoxigenic (ETEC), enterohemorrhagic (EHEC), enteropathogenic (EPEC), enteroaggregative (EAEC), enteroinvasive (EIEC) and neonatal meningitis *E. coli* (NMEC) strains. SslE is required for full virulence in a rabbit model of EPEC infection and as a surface-exposed antigen, SslE has shown great promise as a broadly protective vaccine candidate against a wide range of *E. coli* pathotypes.

SslE interacts with mucosal membranes in the host intestine where it can degrade mucus, a family of heavily O-linked glycosylated proteins and the primary constituents of mucus. 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, SslE is important for mediating the maturation of EPEC biofilms; microbial aggregations encased within a self-produced extracellular matrix comprised of exopolysaccharides, adhesive proteins and nucleic acids. When *E. coli* is released into the environment through faeces or wastewater effluent it can survive for long periods within complex biofilm communities, and these are fundamental for both the environmental ecology of *E. coli* but also for successful colonisation of the intestinal tract. However, the specific molecular mechanisms that SslE uses to promote ecology and/or disease are not well understood.

SslE 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-terminus (Fig. 1a). M60-like domains are metallopeptidases 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 membranes and structures of proteoglycan complexes suggest that interactions with both the mucin peptide and O-linked carbohydrate side chains are important for specificity. The peptidase activity of the SslE M60 domain is effective for the degradation of major mucus of the intestine (e.g. MUC2, MUC3, MUC5AC), however, very little is known as to how SslE interacts with these mucus or the function of the remaining SslE sequence.

In this study, we provide new insight into the global structure of SslE and propose a molecular mechanism by which SslE 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...
reconstruction generated 20 reproducible models with an average solution (theoretical mass 160 kDa). Ab initio dummy residue molecular mass (151 kDa) was consistent with a monomer in Fig. 1 and Supplementary Table 1. Calculation of the particle bead model (teal) with the three defined regions in SslE highlighted. f Docking of an SslE M60 domain homology model (orange) into the TEM map. g TEM map of rSslE 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 1H-15N-HSQC spectrum of rSslE (black) overlaid with 1H-15N-HSQC spectra of the SslE NT1 (light green) and NT2 (olive green) domains.

that SslE is formed of three defined regions (NT1, NT2 and NT3-M60), and is dynamic in solution. We also demonstrate that SslE 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 SslE can bind extracellular DNA (eDNA) in E. coli W biofilms, and under some conditions SslE aggregates may also associate with extracellular cellulose. We propose that SslE acts like a hub and provides a cross-linking function in biofilms.

RESULTS

Overall structural features of SslE

To understand the role of SslE in E. coli biofilms, we first initiated a structural analysis of recombinant SslE from E. coli W, minus the N-terminal lipidation motif and adjacent disordered region (rSslE; 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 rSslE at pH 7.4. Guinier analysis provided a radius of gyration (Rg) of 4.03 ± 0.05 nm and examination of the distance distribution function (P(r)) suggested a maximum particle dimension (Dmax) 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 χ2 fit between calculated and experimental solution scattering of 1.2 (Fig. 1b, c). The averaged low-resolution bead model suggested that rSslE has a slightly elongated structure, however, examination of the Kratky plot indicated that rSslE is a dynamic particle (Supplementary Fig. 1) and no obvious domain features could be assigned.

Negative-stain TEM data were next collected for rSslE at pH 7.4 and we performed single-particle analysis to generate an initial low-resolution structure, with overall dimensions of ~8 × 8 × 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 × 6.5 × 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 rSslE, 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 5kd9) into the TEM envelope indicated that the C-terminus of SslE 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 SslE 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...
HExxH active site motif faces into the channel, and this could be an important feature for its mucolytic activity.

Secondary structure analysis of the SslE 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 SslE. 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 SslE.

We then examined intact rSslE by NMR, but only ~250 strong resonances were observed in a $^1$H$^{15}$N 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 rSslE with $^1$H$^{15}$N-HSQC spectra from SslE sub-domains also clearly showed that the rSslE 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 rSslE and further supported that this region has a significant independent motion with respect to the larger NT3-M60 core.

SslE forms aggregates with amyloid-like properties

Functional amyloids are important proteinaceous structures in biofilms and analysis of the SslE primary sequence had previously highlighted several regions across the protein that may form amyloid-like structures. Furthermore, environmental pH plays a major role in initiating the formation of amyloid-like structures in Staphylococcus aureus, Enterococcus faecalis and Streptococcus mutans biofilms, and we speculated that pH may also stimulate similar structural changes in SslE. We therefore incubated rSslE across the pH range 4.0–7.0 and observed a ring of protein deposited on the walls of the tube at pH ≤ 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 rSslE 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 rSslE 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 rSslE was incubated above pH 4.4 (Fig. 2b). Likewise, when rSslE was boiled and treated with sodium dodecyl sulphate (SDS) and then analysed by anti-His\(_6\) immunoblotting, stable high molecular weight species were observed over the same pH range (Fig. 2c). However, we also observed two distinct bands of monomeric SslE, which suggested that it had been proteolytically cleaved at its N-terminus. Together these data indicated that SslE 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 structure\(^2\) and so we first analysed the secondary structure of soluble rSslE using far-UV circular dichroism (CD) spectroscopy. This suggested that at pH 7.4 rSslE 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 rSslE at pH 7.4 and rSslE 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 rSslE aggregates\(^2\) and is consistent with previous FT-MIR based characterisations of anti-parallel β-sheet amyloid fibres\(^2\) (Fig. 2f).

SslE 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 SslE as the predominant protein (Supplementary Fig. 6). We then created an isogenic ΔsslE 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 ssslE (ΔsslE::ssslE) or SslE with a truncated C-terminus (ΔsslE:: sslEΔm60; residues 1–1000) restored both SslE 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-rSslE antibodies associated with the bacterial outer membrane, which was not seen in the ΔsslE 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 μm from each bacterium) was significantly higher in the wild-type strain compared with the ΔsslE mutant (Fig. 3b–e and Supplementary Fig. 8). Inspection of the E. coli W ΔsslE::ssslE complemented strain under the same conditions produced less clustering of anti-rSslE 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 SslE 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 SslE.

Negative-stain TEM was then used to visualise purified rSslE 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 rSslE 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 (±0.4 × 10\(^3\)) and an average \(R_g\) of 137.5 nm (±1.7), and a larger and more polydisperse species with an average molecular mass of 12 × 10\(^3\) kDa (±0.8 × 10\(^3\)) and an average \(R_g\) of 139.8 nm (±1.5). Examination of rSslE 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 SslE sequence.

Kinetics of SslE aggregation

We took advantage of SEC-SAXS to probe the global shape of rSslE at pH 4.4 prior to fibre formation and both Guinier analysis (\(R_g\) of 3.92 ± 0.05 nm) and examination of the distance distribution function \(P(q) (D_{\text{max}} = 13.7 \text{ nm})\) produced very similar values to rSslE 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 Kraty plots were indicative of rSslE 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 rSslE. 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...
The COSMiCS decomposition, assuming the co-existence of two components, was able to adequately fit all the SAXS profiles with an average $\chi^2$ 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 rSslE at pH 4.4 (Fig. 4d, e and Supplementary Table 1), and this indicated that SslE 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 $\sim$51.1 nm and $D_{max}$ of $\sim$140 nm (Fig. 4d, e and Supplementary Table 1). Furthermore, fractal fit analysis of the decomposed SAXS profiles also indicated that rSslE aggregates into a structure that is fractal in nature (Supplementary Fig. 16). The radius of the primary particle was determined to be $\sim$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 $\sim$10 h. As the SAXS fibrillation data at the 10 h time point had shown depletion of monomeric rSslE but also had reduced fractal properties compared with later time points.

Fig. 3 TEM analysis of E. coli W SslE aggregations. a Accumulation of SslE 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 SslE. b–g SslE localisation upon secretion from E. coli W strains assessed by immunoelectron microscopy. Bacteria were reacted with primary $\alpha$-rSslE antibodies and secondary gold-labelled antibodies on carbon-coated nickel grids and then negatively stained with uranyl acetate. h, i E. coli W ΔsslE mutant washed in citrate-phosphate buffer at pH 5.0 with $\alpha$-rSslE antibodies reacting with aggregates in the extracellular milieu of wild-type and ΔsslE mutant bacteria at pH 5.0. ***P < 0.001; verses ΔsslE mutant by two-tailed Student's t test and error bars are the standard error of the mean. g ΔsslE::sslE W strain washed in citrate-phosphate buffer at pH 4.0, again with $\alpha$-rSslE antibodies reacting with the bacterial surface and secreted fibrous material. Scale bar is equivalent to 200 nm. h TEM analysis of negatively stained rSslE aggregates. Two major species are observed. i Smaller filaments have an average width of $\sim$4.5 nm and are coated with globular structures of $\sim$4.5 nm in diameter. Scale bar, 20 nm. j Schematic describing the overall dimensions of the smaller SslE fibre species. k Larger species of SslE aggregates appear as $\sim$20–40 nm wide structures with variable lengths. Scale bar, 20 nm.
Biofilm maturation supports polymerisation of SslE

As SslE aggregates had been shown to bind Congo red, we used this dye to assess the role of SslE during the formation of *E. coli* W strain macrocolony biofilms. When we compared the ΔsslE 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 whereas 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 ΔsslE mutant with intact *sslE* or *sslEDΔM60* was able to recover macrocolony morphology similar to that of the parental wild-type (Fig. 5a).

We reasoned that if SslE aggregates are formed in macrocolonies and can bind Congo red dye, SslE 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 ΔsslE mutant under these conditions, but in line with previous reports in EPEC strain E2348/69, it was unable to develop structures beyond microcolonies (Fig. 5c). Furthermore, complementation of the mutant with either intact *sslE* or *sslEΔM60* 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 SslE 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 seminaphorhorodfluor-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, SslE 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* biofilms and we examined which of these are produced in macrocolonies from several *E. coli* strains under conditions that support the production of SslE. *E. coli* W, ETEC H10407 and EPEC E2348/69 strains were studied as they had previously been shown to secrete SslE<sup>13,15,11</sup>, while *E. coli* BL21 (DE3) strain was expected to secrete SslE based on genomic analysis<sup>8</sup>. In addition, we also incorporated several other control strains. The *E. coli* K12 strain BW25113 was used as it...
does not produce SslE or cellulose due to deletion of several T2SS genes and mutations in its cellulose biosynthesis genes, respectively, and we created an isogenic ΔsslE knockout mutant in ETEC H10407. We also examined the published EPEC E2348/69 ΔsslE, E2348/69 ΔcsgA (curli defective) and E2348/69 ΔcsgA/ΔbcsA (curli and cellulose defective) knockout mutant strains, and the BW25113 ΔwcaF (colanic acid defective) and BW25113 ΔpggC (PNAG defective) knockout mutant strains from the Keio collection.

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 ΔsslE and ΔcsgA mutants and their parental wild-type strain, although the ΔcsgA/ΔbcsA 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 ΔsslE mutants, while the wild-type and ΔwcaF mutant strains appeared similar in morphology, the ΔpggC 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 ΔsslE 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 ΔcsgA and ΔcsgA/ΔbcsA mutant strains. Examination of the W, BL21 (DE3), H10407 and E2348/69 ΔcsgA/ΔbcsA strains on YESCA all showed a smooth and white morphology and implied that these strains do not produce curli or cellulose. As anticipated, wild-type E2348/69 strain displayed a red, dry and rough morphology, indicative of both curli and cellulose production, while E2348/69 ΔcsgA strain showed a pink, dry and rough morphology (cellulose only) and wild-type BW25113 appeared with a brown, dry and rough morphology (curli only). 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 ΔcsgA/ΔbcsA double mutant strain. This was further supported through the examination of macrocolony growth in the presence of the cellulose stain calcoflour 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. 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 sslEΔM60 represent trans-complementation of intact sslE (ΔsslE::sslE) or sslE minus its M60 domain (ΔsslE::sslEΔM60), respectively, into the E. coli W ΔsslE mutant. Scale bar = 5 mm. b Zoomed-in image of the wild-type E. coli W and ΔsslE mutant. Scale bar = 2 mm. c CLSM images of wild-type E. coli W biofilms and E. coli W ΔsslE mutant and trans-complementation with sslE and sslEΔM60 stained with FM 1-43 (green). Scale bar represents 20 μm. 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 ~30 μm² 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 μm. All data are representative of at least three independent experiments.
were then assessed by immunoblot analysis of macrocolonies (Fig. 6d). Production of SslE was observed in wild-type E. coli 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 (WGA) (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 temperature and not when strains were grown on LB agar at 37 °C.

**SslE associates with eDNA in biofilms**

To examine whether rSslE can directly bind PNAG we first produced rSslE fibres at pH 4.0 and then isolated them in citrate-phosphate buffer at pH 6.0. At this pH, the structure of rSslE fibres would be retained, while monomeric rSslE 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 rSslE fibres, rSslE monomers and RgpB-CTD control to PNAG using a dot-blot overlay assay but could not detect any interaction (Supplementary Fig. 21). However, when we assayed the association of rSslE to immobilised cellulose discs produced by the bacterium *Komagataeibacter rhaeticus*, we detected a significant interaction with fibres, but not monomeric rSslE 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 SslE production (Fig. 6a–c).

Extracellular DNA (eDNA) is another key component of many biofilms and provides adhesion, structuring and chelation of cations. We considered whether SslE 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 rSslE fibres, rSslE monomers or RgpB-CTD control at pH 6.0 and then analysed on agarose gels (Fig. 7a). Upon incubation of pUC19 with 100 μM
rSslE 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 μM monomeric rSslE there was again evidence of a protein–DNA complex in the well. We then assessed the distribution of SslE and eDNA within established wild-type E. coli W biofilms stained with the rSslE antibody and the cell-impermeant DNA-binding dye TOTO-1, respectively, using CLSM (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 SslE was again seen across biofilms with SslE 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 ΔsslE control.

**DISCUSSION**

In this study, we have revealed that SslE 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 SslE and its homologs, 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. 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). 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 domains 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 SslE faces into the NT3 domain and this implies that both the NT3 and M60
domains are responsible for the recognition of mucins in SslE. With significant flexibility of the NT1 and NT2 region, these domains may also interact with mucin substrates and may promote processivity.

Although SslE 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 SslE is required for the maturation of E. coli W biofilms, where it is widely distributed throughout the extracellular matrix. We have demonstrated that SslE can form aggregates that display amyloid-like properties at pH ≤ 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 heterogeneous distributions of pH, with values ranging between 5.0 and 7.0. Therefore, it is very likely that SslE 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 SslE 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. Gram-negative bacteria most known amyloid structures are mate residues and subsequent structural rearrangements. FTIR analysis of rSslE aggregates provided an amide I maxima at peaks at lower frequencies, while native examination of aggregates using negative-stain TEM has revealed structural integrity of the biopeptides and provide adhesion, cohesion and contribute to the formation of new cavities upon fibrillation. In this model, the asymmetric core or instead represent β-bre stacking/clumping effects. 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 SslE molecules along the fibre, upon fibrillation of SslE. Nonetheless, based on our biophysical data, we propose a tentative model for the fibrillation of SslE 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 rSslE fibres and monomers can directly bind DNA in vitro and SslE antibodies co-localise with eDNA and bacteria in E. coli W biofilms. After secretion SslE can associate with the bacterial membrane but is also present in a soluble form in the extracellular milieu (Supplementary Fig. 7). Therefore, cell surface-associated SslE could presumably interact with extracellular fibres or may help to nucleate the aggregation of soluble SslE. 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 SslE is functional in binding eDNA within biofilms and alternatively, monomeric SslE may directly bind eDNA without the need for SslE 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 SslE fibres can also associate with cellulose, which is known to have a structural function in biofilms and provides cohesion and elasticity. Although we did not detect cellulose under conditions of SslE production, cellulose secreted by other bacteria in multispecies communities could provide a source for this interaction. As a monomer, SslE 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 SslE, the first two strains are harmless commensals, while ETEC and EPEC are major aetiological agents in developing countries. SslE is also required for colonisation in a rabbit model of EPEC infection and is actively transcribed during EPEC infection of mice. This appears to be a common feature of SslE production as expression of sslE was previously shown to be repressed at low temperatures in E. coli W5 and ETEC and here we have shown that this is also the case in BL21 (DE3) and EPEC E2348/69 strains. This suggests that SslE 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 intestine where the intraluminal pH ranges between ~6.0 and 7.5. However, E. coli possess an acid-tolerance response that supports exponential growth at pH values as low as pH 4.256 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 6.0. Together this suggests that SslE 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. SslE 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 K2HPO4, 2 g KH2PO4, 1 g (NH4)2SO4, 0.5 g sodium citrate, 16 g agar, 4 g glucose, 0.1 g MgSO4 and 2 μg thiamine. Isotopically defined M9 minimal medium (pH 7.4) contained (per litre) 6.0 g Na2HPO4·7H2O, 3 g KH2PO4, 0.5 g NaCl, 0.12 g MgSO4·7H2O, 22 μg CaCl2·2H2O, 0.5 μg thymine, 8.3 mg FeCl3·6H2O, 0.5 mg ZnCl2, 0.1 mg CuCl2, 0.1 mg CoCl2·6H2O, 0.1 mg H3BO3 and 13.5 mg MnCl2·6H2O, supplemented with 2 g [U-13C6]glucose and/or 0.7 g [15N]NH4Cl. 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 sssl was constructed in E. coli W and H10407 strains by allelic exchange with a FLP recombination target (FRT)-ranked Kan' cassette, using a modified protocol38. The fragment was amplified from E. coli strain JW5925,17 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 μg/ml ampicillin and 1 M L-arabinose at 30 °C to OD of 0.6. Cells were made electrocompectent, electroporated with 10–100 ng of PCR fragment and then selected on LB agar at 37 °C containing 25 μg/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 pPC1 and pPC2 were generated by amplifying whole length sssl, or sssl minus the M60 domain, from E. coli W gDNA using primer pairs PC7/PC8 or PC7/PC9, respectively. These were digested with HindIII/Nhel (NEB), ligated into HindIII/Nhel-digested pBADcm18 vector. Synthetic genes gPC4, gPC5, gPC6 and gPC7 were synthesised by Synbio Synthesis (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 Sssl (rSssl: residues 67–1497; numbered based on mature sequence)13, Sssl 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 (Sssl; New England Biolabs) or BL21 (DE3) (RgpB-CTD; New England Biolabs), grown at 37 °C in LB media (rSssl, RgpB-CTD: 100 μg/ml ampicillin; NT1, NT1-NT2, NT3-M60: 100 μg/ml kanamycin) and expression induced with 0.5 mM isopropyl-d-1-thiogalactopyranoside (IPTG) at an OD600nm 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 (rSssl, NT1-NT2, NT3-M60) column (GE Healthcare).

SEC-SAXS
SAXS data were collected on beamline B2190 at Diamond Light Source (DLS, Oxford, UK) at 25 °C over 0.16 μl/min, pre-equilibrated in 20 mM citrate-phosphate buffer, 200 mM NaCl at either pH 4.4 or 7.4. Sssl NT1, 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 was applied to a KW403-4F column (Shodex) at 0.16 ml/min, pre-equilibrated in 20 mM Tris–HCl pH 8, 200 mM NaCl. Sssl data were measured over a momentum transfer range of 0.003–q < 0.44 Å⁻¹. Peak integration and buffer subtraction were performed in CHROMIX36. The radius of gyration (Rg) and scattering at zero angle (I0) were calculated from the analysis of the Guinier region by AUTORG36. The distance distribution function [P(r)] was subsequently obtained using GNOM36 yielding the maximum particle dimension (Dmax). Ab initio low-resolution shape restoration was carried out by calculating 20 models in DAMMIF30, which were subsequently averaged using DAMAVER30 and used as a starting model for refinement in DAMMIN30. An additional 20 models were also calculated and averaged in DAMAVER30. CRYSOL was used to compare final rSssl 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 μl of rSssl (625 nM) in 50 mM Tris–HCl pH 8, 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 μl 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 ×50,000 with a pixel size of 2.1 Å/pixel and a range of defocus from 1 to 3 μm. Data were processed using Relion 3.161. Defocus and astigmatism parameters were estimated using CTFIND42 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 3.1. For obtaining initial model generation, a few rounds of 3D classification followed by 3D refinement was carried out. The final model was refined to 22 Å using the gold standard FSC (0.143 criterion), PHYRE235 was used to analyse the sssl sequence (residues 67–1497; UniProt ID E0IW31) and generate a homology model for the C-terminal 60 domain (residues 962–1415) based on residues 421–894 of P. aeruginosa IMPA (PDB ID code Skdv; 100% confidence, 21% identity). This was then docked into the rSssl TEM envelope using UCSF CHIMERAS64.

Solution NMR spectroscopy
NMR measurements for rSssl were performed at 25 °C on a 100 μM 2H14N-labelled sample in 50 mM NaPO4 pH 7.4, 100 mM NaCl, 10 % D2O. NMR measurements for Sssl NT1 (0.6 mM), Sssl NT2 (1.3 mM) and Sssl NT1-NT2 (0.8 mM) were performed at 25 °C on 14N-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-HSOQC experiment and TR2, TR2 relaxation times for rSssl were recorded on a Bruker Avance III HD 950, equipped with a T1I cryo probe. TROSY 1H15N-HSOQC experiment for sssl NT1-NT2 was recorded on a Bruker Avance III HD 800, equipped with a TCI cryo probe. Standard 1H15N-HSOQC spectra of Sssl NT1, Sssl NT2 and RgpB-CTD were recorded on a Bruker Avance III HD 700, equipped with a TCI cryo probe. Data were processed in NMRPIPE65 and analysed/visualised with ANALYSIS66 and NMRVIEW67.

Recombinant protein ring assay
In all, 20 μl of rSssl, NT1-NT2 or NT3-M60 (62.5 μM) 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 Sssl 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, Sssl fibre bands retained in the wells were excised and incubated with 10 mM dithiothreitol at 56 °C. 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 μm i. D. 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 × 106 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 × 105 and 1 × 106 and a dynamic exclusion window of 35 secs ± 10 ppm. MS2 precursor ions were isolated in the quadrupole stage to a mass width filter of 1.6 m/z. Ion trap fragmentation spectra (ITMS2) were collected with an AGC target setting of 1 × 106 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 Ssle (Uniprot Accession number—E3P90J) with Mascot search algorithm (v2.6.0; www. matricscience.com) and the Sequest search algorithm69. 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 µl of 50 µM rSsIE in 10 mM Tris–HCl pH 8, 100 mM NaCl buffer was mixed with 1.5 µl 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 µl 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 scrapped from glass tubes and resuspended in 0.5 ml 100 mM citrate-phosphate buffer, pH 4.0, centrifuged at 15,000 × g and then the top 950 µl solution was carefully removed and discarded. This was followed by three rounds of addition of 950 µl 100 mM citrate-phosphate buffer at pH 4.0, centrifugation at 15,000 × g and the top 950 µl discarded. The final 50 µl sample was mixed with 1 × 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, and then incubated overnight at 4 °C with polyclonal anti-rSsIE antibody (rabbit; Invitrogen) or monoclonal anti-DsbA antibody (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). 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 antibody (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.

Immunoelectron 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 Grid-discharged carbon-coated 25 µm grid (Agar Scientific) for 10 min, washed with 50 mM glycine and then blocked with 1% (w/v) Natural Donkey Serum (Jackson Immuno Research) for 30 min. This was then incubated for 1 h with either primary polyclonal rSsIE antibody (rabbit; Invitrogen) or primary polyclonal CsgA Antibody (guineaigp; Invitrogen) diluted 1:100 in blocking buffer, washed five times for 3 min each with 0.05% (v/v) 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 package (Olympus Europe, UK). Images of 13 E. coli W wild-type and 13 ΔsslE mutant bacteria were then masked and gold particles were counted between 0.2 and 1.0 µm 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 × g for 10 min, the top 80% of buffer discarded and then 4 µl 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 µl 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 the iTEM software package (Olympus Europa, UK). Micrographs were collected at a nominal magnification of 80,000x with a pixel size of 5.96 Å/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 × g and the top 850 µl of buffer was discarded. MALD experiments were then performed on baseline B21 at the DLS (Oxford, UK). In total, 20 µl 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
rsSE 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 × 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 w1 = 15 kHz. A standard CP excitation 10-12C-13C 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 45.3 s total acquisition time. The applied power was adjusted so that 1H and 13C hard pulses were both 4 μs (w1, wCP = 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 wCP = ~70 kHz and w1H = ~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 (w1H = 15 kHz).

SAXS fibration analysis
SAXS data were collected on beamline B2159 at DLS (Oxford, UK) at 25 °C. Immediately prior to data collection, 10 mg/ml rsSE in 20 mM Tris–HCl pH 8, 200 mM NaCl was buffer exchanged into 20 mM citrate-phosphate buffer, pH 4.0 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 rsSE (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 Å⁻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 COSMICS 86, which utilises MCR-ALS 87 to perform model-free decomposi- tion 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 fibration. 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 FNLLS algorithm 88. 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 ATLAS suite of programmes 89. Fractal analysis was carried out using SASVIEW version 4.2.2 (http://www.sasview.org/). The programmes ATLAS 89 and Scatter were also used to obtain a cross-sectional radius of gyration Rg from the 10 h post-fibre induction scattering profile and from this the cross-sectional Price calculated. BODIES 90 was then used to approximate the geometric shape of rsSE fibres, which suggested a cylindrical shape with radius of 3.93 nm and a height of 16.5 nm (χ² 1.6). Processing and refinement statistics can be found in Supplementary Table 1.

Macrolony 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 volume adjusted to an OD600nm ~0.7. A lawn of W and H10407 strains or derivatives was 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 x 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 cellu- lose treated samples.

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 x 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 cellu- lose 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 OD600nm = 1.0. Formic acid was added to 70% (v/v) and then lyophilised overnight. Samples were resuspended in 200 µl 1×
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-Blottter and Power-Blottter 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-rSslE antibody (rabbit; Invitrogen) or 1:50 diluted polyclonal (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 ECL 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 protease K (20 mg/ml; NEB, UK) for 60 min at 60 °C, followed by 80 °C for 6 min and then 100 µl of supernatant was 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 and then incubated with 20 µg/ml of *E. coli* BW25113 strain or its 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 protease K (20 mg/ml; NEB, UK) for 60 min at 60 °C, followed by 80 °C for 6 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) in the same buffer for 1 h, followed by three further washes and then treatment with ECL 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 rSslE monomer, rSslE 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-His, HRP-conjugated antibody (ThermoFisher, UK) in the same buffer for 1 h, followed by 4 further washes and treatment with ECL 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 OD600nm=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. 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-His, HRP-conjugated antibody (ThermoFisher, UK) in the same buffer for 1 h, followed by 4 further washes and treatment with ECL substrate (Peirce) before detection.

**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 rSslE monomer, rSslE 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% (w/v) bromophenol blue in MES buffer pH 6.0, while DNA-monomeric rSslE and RgpB-CTD samples were mixed with “purple no SDS DNA loading dye” at pH 8.0 (NEB, UK). Samples were run on 7% polyacrylamide gel containing GelRed (Biotium, UK) 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 α-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.

**Biofilm immunofluorescence**
Single colonies of wild-type *E. coli* W and Δsae 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-rSslE 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 FITC (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 PBS and then images were captured at ×20 magnification using an excitation wavelength of 488 nm (TOTO-1), 488 nm (FITC-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-rSslE 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 SslE 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 SslE at pH 4.4 and 7.4, and the SslE N1, N2 and N3-M60 domains, have been deposited in the Small-Angle Scattering Biological Data Bank (https://www.sasdbd.org/) with accession ID SASDLS2, SASDLSV, SASDSMU6, SASDSMV6 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
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/R017562/1), 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. S.W. 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-BioImaging (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. 202276/20Z/16/Z and IG16/2/32273, 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 ΔΔE strain and Prof. Jorge Girón (Benemérita Universidad Autónoma de Puebla, Mexico) for providing the EPEC E2348/69 wild-type, ΔcsgA and ΔcsgA/ΔclusA 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., J.J., G.H.C., 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., 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., L.C., C.E.-G., G.M., B.D., L.M.S., J.J., G.H.C., V.C.D. and J.A.G. Contributed reagents/materials/analysis tools: S.S., L.M.S., 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., J.J., G.H.C., V.C.D. and J.A.G. The authors declare no competing interests.

Additional 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/.