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

Paula Corsini[1,2†], Sunjun Wang[1†], Saima Rehman[1†], Katherine Fenn[2†], Amin Sagar[3], Slobodan Sirovica[4], Leanne Cleaver[1], Ben Dorgan[1], Lee M. Sewell[1], Steven Lynham[5], James Jarvis[6], Guy Carpenter[1], Michael. A. Curtis[1], Pau Bernadó[3], Vidya C. Darbari[2*] & James A. Garnett[1,2*]

[1]*Centre for Host-Microbiome Interactions, Faculty of Dental, Oral & Craniofacial Sciences, King's College London, London, UK*

[2]*School of Biological and Chemical Sciences, Queen Mary University of London, London, UK*

[3]*Centre de Biologie Structurale. INSERM, CNRS, Université de Montpellier, France*

[4]*Centre for Oral Bioengineering, Institute of Dentistry, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK*

[5]*Proteomics Facility, Centre of Excellence for Mass Spectrometry, King's College London, London, UK*

[6]*Randall Division of Cell and Molecular Biophysics and Centre for Biomolecular Spectroscopy, King's College London, London, UK*

[†]*equal contributions*

**Correspondence to:*

James Garnett, Centre for Host Microbiome Interactions, Dental institute, King's College London, London, UK. email: james.garnett@kcl.ac.uk

Vidya C. Darbari, Department of Biochemistry, School of Biological and Chemical Sciences, Queen Mary University of London, UK. email: v.darbari@qmul.ac.uk

Abstract

***Escherichia coli* is a Gram-negative bacterium that colonizes the human intestine and virulent strains can cause severe diarrhoeal and extraintestinal diseases. The protein SslE is secreted by a range of pathogenic and some commensal *E. coli* strains. It can degrade mucins in the intestine, promotes biofilm maturation and in virulent strains, it is a major determinant of infection, although how it carries out these functions is not well understood. Here we examine SslE from the *E. coli* Waksman and H10407 strains and show that SslE has a highly dynamic structure in solution. We directly observe acidification within mature biofilms, describe a mechanism where SslE forms unique functional fibres under these conditions and determine that these SslE aggregates can bind cellulose, a major exopolysaccharide of many *E. coli* biofilms. Our data indicates that the spatial organization of SslE polymers and local pH are critical for biofilm maturation and SslE is a key factor that drives persistence of SslE-secreting bacteria during acidic stress.**

Introduction

Escherichia coli is a primary colonizer 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[1]. 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[2, 3, 4, 5, 6, 7]. 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[4] 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[3, 6, 8, 9].

59 SslE interacts with mucosal membranes in the host intestine where it can degrade mucins[6, 10, 11, 12], a
60 family of heavily *O*-linked glycosylated proteins and the primary constituents of mucus[13]. This provides
61 nutrients during bacterial growth but also enables these *E. coli* strains to penetrate the gut mucosa to
62 access host cells for efficient colonisation and targeting of toxins/effectors. However, SslE is also
63 important for mediating maturation of EPEC biofilms[4]; microbial aggregations encased within a self-
64 produced extracellular matrix comprised of exopolysaccharides, adhesive proteins and nucleic acids[14].
65 When *E. coli* is released into the environment through faeces or wastewater effluent it can survive for
66 long periods within complex biofilm communities[15, 16], and these are fundamental for both the
67 environmental ecology of *E. coli* but also for successful colonisation of the intestinal tract[17]. However,
68 the specific molecular mechanisms that SslE uses to promote ecology and disease are not well
69 understood.

70 SslE is a ~165 kDa lipoprotein composed of an N-terminal periplasmic localization sequence and
71 lipobox motif, an unstructured ~5 kDa region, a ~110 kDa residue region with no significant primary
72 sequence homology to any other known protein, and a ~50 kDa M60-like aminopeptidase domain at its
73 C-terminus[18] (**Fig. 1a**). M60-like domains are metalloproteases that contain a zinc binding HxxHE
74 motif and an additional conserved catalytic glutamate residue, which cleaves the peptide backbone of
75 mucin-like substrates. These and other related enzymes have been identified in both prokaryotic and
76 eukaryotic microbes that interact with host mucosal membranes[18] and structures of proteoglycan
77 complexes suggest that interactions with both the mucin peptide and *O*-linked carbohydrate sidechains
78 are important for specificity[19, 20]. The peptidase activity of the SslE M60 domain is effective for the
79 degradation of major mucins of the intestine (e.g. MUC2, MUC3, MUC5AC)[6, 10, 11, 12], however, we
80 know very little of how SslE interacts with these mucins or what function the remaining SslE sequence
81 has.

82 Using electron microscopy (EM), small angle X-ray scattering (SAXS), nuclear magnetic resonance
83 (NMR) spectroscopy and biochemical analyses we show that SslE from *E. coli* W is formed of three
84 defined regions (NT1, NT2 and NT3-M60), and is highly dynamic in solution. We also demonstrate

85 that SslE undergoes conformational changes under acidic conditions and this leads to the formation of
86 amyloid-like structures, which is mediated through its unique N-terminal (NT) region. We directly
87 observe acidification within mature *E. coli* W biofilms and determine that SslE fibres can bind cellulose,
88 a major exopolysaccharide of *E. coli* biofilms. We provide evidence that SslE aggregates are functional
89 for the maturation of both *E. coli* W and ETEC strain H10407 biofilms and this is likely a general
90 mechanism of all SslE-secreting bacteria.

91

92 **Results**

93 **SslE forms amyloid-like aggregates under acidic conditions.** Functional amyloids are important
94 proteinaceous structures in biofilms[21] and analysis of the SslE primary sequence had previously
95 highlighted several regions across the protein that may form amyloid-like structures[22]. Furthermore,
96 environmental pH plays a major role initiating the formation of amyloid-like structures in
97 *Staphylococcus aureus*, *Enterococcus faecalis* and *Streptococcus mutans* biofilms[23, 24, 25], and we
98 speculated that pH may also stimulate similar structural changes in SslE. SslE from *E. coli* W, minus
99 the N-terminal lipidation motif and adjacent disordered region (rSslE; residues 67 to 1497; UniProt
100 ID E0IW31) (**Fig. 1a**), was therefore produced in *E. coli* K-12 and purified by nickel-affinity and size
101 exclusion chromatography. We then incubated rSslE across the pH range 4.0 to 7.0 and observed a ring
102 of protein deposited on the walls of the tube at pH ≤ 4.4, which resisted solubilization up to pH 6.2 (**Fig.**
103 **1b**). Mass spectrometry analysis of the ring identified peptides that spanned the complete sequence of
104 rSslE and indicated that the aggregate contained intact protein, rather than a degradation product
105 (**Supplementary Fig. 1**). We next used the amyloid diagnostic dye Thioflavin-T (ThT) to assess the
106 formation of rSslE aggregates in solution over time. We observed large increases in ThT fluorescence
107 emission again across the pH range 4.0 to 4.4, but no change was detected when rSslE was incubated
108 above pH 4.4 (**Fig. 1c**). Likewise, when rSslE was boiled and treated with sodium dodecyl sulphate
109 (SDS) and then analysed by immunoblotting, stable high molecular weight species were observed over
110 the same pH range (**Fig. 1d**).

111 A main feature of amyloid is the presence of a cross-β-sheet structure[26] and so we first analysed the
112 secondary structure of soluble rSslE using far-UV circular dichroism (CD) spectroscopy. This
113 suggested that at pH 7.4 rSslE is composed of approximately 30% α-helix, 12% β-sheet and 45% coil,
114 while under acidic conditions and prior to aggregation, there is an ~8% loss of helical secondary
115 structure with a ~5% increase in β-sheet structure (**Fig. 1e; Supplementary Table 1**). We next used
116 attenuated total reflectance (ATR) Fourier transform mid-infrared (FT-MIR) spectroscopy to probe the
117 secondary structure composition of soluble rSslE at pH 7.4 and rSslE fibres at pH 4.4. Examination of
118 the amide I absorption band showed a peak at approximately 1656 cm^{-1} and 1624 cm^{-1}, respectively,
119 which corresponds to β-sheet structure and the shift from higher to lower frequency indicated that a
120 larger and more rigid structure had assembled under acid conditions (**Supplementary Fig. 2**). The
121 presence of both a high-intensity absorption band at ~1624 cm^{-1} and a low-intensity band at ~1695 cm^{-}
122 ^{1} is indicative of an anti-parallel β-sheet structure[27] and is consistent with previous FT-MIR based
123 characterisations of other anti-parallel β-sheet amyloid fibres[28] (**Fig. 1f**). Furthermore, deconvolution
124 of the amide I region suggested that rSslE fibres are predominantly composed of aggregated β-strands
125 (~21%) and other β-sheet structures (~54%) and with very little helical content (~2%) (**Supplementary**
126 **Table 2**).

127 **Biofilm maturation supports polymerisation of SslE.** Congo red is a dye which binds extracellular
128 polymeric substances (EPS) and we used this to assess the role of SslE during formation of macrocolony
129 biofilms. We created an isogenic Δ*sslE* mutant in *E. coli* W and ETEC strain H10407 and compared
130 these to their parental wild-type strains. When these mutants were cultured on Congo red agar plates
131 for 24 hrs, we observed no obvious differences, however, after 96 hrs the mutants exhibited an altered
132 morphology with the dye retained within the centre of the colony (**Fig. 2a**). Complementation of the
133 Δ*sslE* mutant in *E. coli* W with a plasmid containing intact *sslE* (Δ*sslE::sslE*) was able to restore
134 bacterial secretion of SslE and recover macrocolony morphology similar to that of the parental wild-
135 type (**Fig. 2a**; **Supplementary Fig. 3**). We also generated a plasmid expressing SslE with a truncated
136 C-terminus (Δ*sslE::sslEΔM60*; residues 1 to 1000) to examine the role of the M60-domain during
137 colony formation, and this again restored SslE secretion and wild-type colony morphology (**Fig. 2a**;

138 **Supplementary Fig. 3**). This led us to speculate that SslE experiences extracellular pH values <5.0
139 during the development of biofilms and this results in SslE being deposited as a functional aggregate.

140 To test this, we first grew *E. coli* W biofilms using a microfluidic system under continuous flow and
141 assessed their overall morphology using confocal laser scanning microscopy (CLSM). While biofilms
142 grown for 48 hrs produced a relatively homogenous lawn of bacterial growth, after 96 hrs there were
143 clear signs of maturation, with a significant increase in biofilm mass, structural heterogeneity and the
144 presence of solvent channels (**Fig. 2b,c**). We then examined the *E. coli* W Δ*sslE* mutant under these
145 conditions, but in line with previous reports in EPEC strain E2348/69[4], it was unable to develop
146 structures beyond microcolonies (**Fig. 2c,d**). Furthermore, complementation of the mutant with either
147 intact *sslE* or *sslEΔM60* was again able to restore wild-type biofilm morphology. Together with our
148 previous observation, this indicated that the C-terminal M60-domain is not required for translocation
149 of SslE through its T2SS and it is dispensable for biofilm development, at least under these conditions.

150 We then examined pH distribution across *E. coli* W biofilms using CLSM coupled with the cell-
151 impermeant fluorescent ratiometric probe seminaphthorhodafluor-4F 5-(and-6) carboxylic acid (C-
152 SNARF-4). After growing *E. coli* W for 24 hrs under continuous flow, we observed pH values between
153 6.0 and 6.3 across biofilms (**Supplementary Fig. 4**), however, after 96 hrs, lower pH values were also
154 recorded (**Fig. 2e**). Although across the majority of biofilms examined, we detected pH values between
155 approximately 5.0 and 6.0, clearly defined microenvironments were also observed with pH values that
156 ranged from 4.2 and 4.8.

157 **Overall structural features of SslE.** We next initiated structural studies of SslE. Small angle X-ray
158 scattering (SAXS) coupled with size exclusion chromatography (SEC-SAXS) was first used to get
159 shape information for rSslE at pH 7.4. Guinier analysis provided a radius of gyration (R_g) of 4.03 nm
160 and examination of the distance distribution function (P(r)) suggested a maximum particle dimension
161 (D_{max}) of 14.1 nm and R_g of 4.07 nm (**Supplementary Fig. 5; Supplementary Table 3**). Calculation
162 of the particle molecular mass was consistent with a monomer in solution and *ab initio* dummy residue
163 reconstruction generated reproducible models with an average normalized spatial discrepancy (NSD)

164 score between reconstructions of 0.56 and a χ^2 fit between calculated and experimental solution
165 scattering of 1.2 (**Fig. 3a,b**). The low-resolution bead model suggested that rSslE has a slightly
166 elongated structure, however, examination of the Kratky plot indicated that rSslE is a dynamic particle
167 (**Supplementary Fig. 5**) and no obvious domain features could be assigned.

168 Negative-stain TEM data was next collected for rSslE at pH 7.4 and we performed single particle
169 analysis to generate an initial low-resolution structure, with overall dimensions of approximately
170 8×8×11.5 nm (**Fig. 3c; Supplementary Fig. 6**). Here, three well defined regions could be identified:
171 two adjacent globular domains (regions 1 and 2) with approximate diameters of 4 nm and 4.5 nm,
172 respectively, and a torus shaped structure (region 3) with a ~15 Å wide central channel and approximate
173 overall dimensions of 4.5×6.5×9 nm (**Fig. 3d**). Superposition of the SAXS bead model with the TEM
174 envelope supported SslE being dynamic in solution, with an apparent reduction of bead volume in all
175 three regions of the SAXS model. Docking analysis of a homology model of the M60 domain, based
176 on the IMPa M60 metalloprotease from *Pseudomonas aeruginosa* (PDB ID code 5kdv)[19], indicated that
177 the C-terminus of SslE is located at the base of region 3 (**Fig. 3e**). This suggested that the remaining
178 density in region 3 is composed of the SslE sequence directly N-terminal to the M60 domain and these
179 two regions pack against one another to form a stable structure around an inter-domain channel.
180 Furthermore, from this model the M60 HxxHE active site motif faces into the channel and this could
181 be an important feature for its mucolytic activity. The location of the M60 domain in the TEM map was
182 also supported by previous reports where it was not possible to purify a folded recombinant SslE M60
183 fragment due to sample instability[22].

184 Secondary structure prediction[29] of the SslE sequence indicated two potential inter-domain boundaries
185 within the N-terminal region and we predicted that these represented regions 1 and 2 located at the
186 extreme N-terminus (NT) of SslE. We therefore renamed these as the NT1 (residues 67-211) and NT2
187 (residues 230-425) domains, respectively, and renamed the remaining sequence as NT3 (residues 426-
188 1000) (**Fig. 3f**). The NT1 and NT2 domains, along with the NT1-NT2 region (residues 67-425) and
189 NT3-M60 core (region 3; residues 426-1497) were then produced in *E. coli* K-12 and purified by nickel-

190 affinity and size exclusion chromatography. Examination of these constructs using analytical gel
191 filtration and nuclear magnetic resonance (NMR) spectroscopy showed that they were well folded
192 (**Supplementary Fig. 7**), and supported NT1, NT2 and NT3-M60 being well-defined structural
193 boundaries within SslE. When we examined intact rSslE by NMR, a $^{1}H^{15}N$ transverse relaxation
194 optimized spectroscopy (TROSY) HSQC spectrum yielded only ~250 strong resonances with the
195 remaining ~1200 peaks being either very weak or completely absent (**Fig. 3g**). Furthermore, NMR
196 relaxation measurements of these intense peaks provided an estimated correlation time consistent with
197 an ~80 kDa domain (τ_c ~ 35 ns at 37°C). Comparison of rSslE with $^{1}H^{15}N$-HSQC spectra from SslE
198 subdomains also clearly showed that the rSslE NMR spectrum is dominated by the NT1 domain but
199 with minor contributions from the NT2 domain (**Fig. 3g; Supplementary Fig. 8**). This further
200 supported this region having significant independent motion with respect to the NT3-M60 core and we
201 considered whether the NT1-NT2 region alone could be responsible for mediating SslE aggregation.
202 We therefore repeated our ring assay with SslE NT1-NT2 and the NT3-M60 constructs incubated across
203 the pH range 4.0 to 7.0. This resulted in significantly less protein being deposited on the walls of the
204 tubes and for NT3-M60 this was over a wider pH range. This implied that both the NT1-NT2 and the
205 NT3 regions are necessary for polymerisation of intact SslE (**Supplementary Fig. 9**).

206 **Kinetics of SslE aggregation using SAXS.** We took advantage of SEC-SAXS to probe the global
207 shape of rSslE at pH 4.4 prior to fibre formation and both Guinier analysis (R_g 3.92 nm) and examination
208 of the distance distribution function (D_{max} 13.7 nm) produced very similar values to rSslE at pH 7.4
209 (**Supplementary Fig. 10; Supplementary Table 3)**. However, comparison of scattering profiles and
210 Kratky plots showed differences at q-values around 0.13 $Å^{-1}$, indicative of rSslE becoming more rigid
211 under acidic conditions (**Fig. 4a**; **Supplementary Fig. 11**). *Ab initio* dummy residue reconstruction
212 produced models with an average NSD score of 0.45 and a χ^2 fit between calculated and experimental
213 solution scattering of 1.0 (**Fig. 4b**; **Supplementary Fig. 12**). Comparison of the bead models indicated
214 an increase in bead volume within the NT1 and NT2 domains and adjacent NT3 region at pH 4.4, while
215 little change was observed in the NT3-M60 core, which supported an increase in rigidity under acidic
216 conditions that could trigger aggregation.

217 We then exploited the additive nature of SAXS along with its ability to examine structures over a large
218 size range to study the kinetics of fibre formation by rSslE. We acquired SAXS data at pH 4.4 during
219 the course of aggregation for 11 hrs and the resulting 23 profiles displayed a systematic upwards
220 intensity increase, indicating the presence of large aggregates in solution (**Fig. 4c**). The appearance of
221 these large species occurred relatively fast as the upwards intensity increase was already observed after
222 1 hr and the increase of aggregated species was concomitant to the loss of the SAXS features linked to
223 the monomeric species. Principal Component Analysis (PCA) indicated that the data set could be
224 described as a two-component system (**Supplementary Fig. 13**) and we decomposed the time-
225 dependent data set along with the SAXS profile of the monomer, obtained using SEC-SAXS, with
226 COSMiCS. COSMiCS uses a chemometric approach to decompose SAXS data sets in a model-free
227 manner to derive the pure SAXS curves of the co-existing species and their relative populations, which
228 we used here to report on the fibrillation kinetics[30].

229 The COSMiCS decomposition, assuming the co-existence of two components, was able to adequately
230 fit all the SAXS profiles with an average $<\chi^2>$ of 1.49, although some deviations from the perfect fit
231 were observed for the SAXS curves measured in the first 3 hrs (**Fig. 4c**). This observation suggested
232 the presence of a small population of a third species in the initial steps of fibrillation but this putative
233 third species could not be captured when considering a three-species decomposition with COSMiCS.
234 The extracted profile of the smallest species, and the subsequently derived R_g (3.93 nm) and D_{max} (13.3
235 nm) values, were almost identical to the SEC-SAXS profile for rSslE at pH 4.4 (**Fig. 4d,e;**
236 **Supplementary Table 3**), and this indicated that SslE is in a monomeric form at the beginning of the
237 fibrillation process. The extracted SAXS profile of the second component displayed the typical features
238 of a highly elongated amyloidogenic particle[30, 31, 32, 33] and the *P(r)* analysis of the profile indicated that
239 the repetitive unit of the fibre had an R_g of ~51.1 nm and D_{max} of ~140 nm (**Fig. 4d,e; Supplementary**
240 **Table 3**). Examination of the relative populations of the two species derived from the COSMiCS
241 decomposition over time showed that after 30 min from initiating the measurements, the amyloidogenic
242 species already represented 50% of the molar fraction (**Fig. 4f**). From this time point, this population
243 continued to grow and reached a plateau after ~10 hrs.

244 **SslE fibres bind cellulose.** We speculated that during the maturation of biofilms, SslE aggregates
245 would likely interact with other components of the biofilm matrix and may contribute to structural
246 integrity of these communities. We therefore examined whether rSslE could bind cellulose, a major
247 exopolysaccharide of many *E. coli* biofilms. We first produced rSslE fibres at pH 4.0 and then isolated
248 these aggregates in citrate-phosphate buffer at pH 6.0, a condition in which fibres would resist
249 solubilization (**Fig. 1b**). As controls, we used monomeric rSslE, and the C-terminal domain (CTD) of
250 the *Porphyromonas gingivalis* gingipain protein RgpB (**Supplementary Fig. 14**), both of which were
251 not expected to bind cellulose or form polymers at pH 6.0. When we assayed binding of rSslE fibres to
252 immobilised cellulose discs we observed a significant interaction, but as expected this was not detected
253 with monomeric rSslE or RgpB-CTD (**Fig. 5**). This indicated that under similar conditions SslE
254 aggregates may associate with cellulose and/or other exopolysaccharides during biofilm maturation.

255 **Discussion**

256 SslE is a relatively large protein and we have revealed that it is composed of four defined regions: two
257 globular domains at the N-terminus (NT1 and NT2) and a central torus shaped core, where a NT3
258 domain packs against an M60-like aminopeptidase. The N-terminal region (NT1, NT2 and NT3) is
259 composed of a unique primary sequence that has only been identified in SslE and its homologs[34], while
260 the C-terminal M60-like domain is found in a wide range of proteins secreted by both prokaryotic and
261 eukaryotic microorganisms that interact with mucosal membranes[18, 35]. The atomic structure for three
262 M60-like domains in complex with *O*-glycopeptides have been described: BT4244 from *Bacteroides*
263 *thetaiotaomicron* (PDB ID code: 5kd2), IMPa from *P. aeruginosa* (PDB ID code: 5kdw) and ZmpB
264 from *Clostridium perfringens* (PDB ID code: 5kdn)[19]. While the active site structures in these bacterial
265 enzymes are highly conserved, there are variations within the remaining M60 domain, which enables
266 them to recognize unique glycan sequences. Structural data is not available for the N-terminus of
267 BT4244, although it is thought to contain putative BACON and CBM32 carbohydrate binding
268 domains[18]. Likewise, structures for the N- and C-terminal regions of ZmpB do not exist but are
269 predicted to contain dicoidin, NPCBM and leucine rich repeat domains. It is likely that these M60

270 augmentations have a role in the recognition of *O*-glycopeptide substrates and it appears that the active
271 site in the M60 domain of SslE faces into the NT3 domain, which implies that both the NT3 and M60
272 domains are responsible for recognition of mucins in SslE. Furthermore, with significant flexibility of
273 the NT1/NT2 region, these domains might be functional in clamping mucin substrates in place once
274 bound at the NT3-M60 core.

275 Although SslE has a clear function in the processing of mucins, it is a unique member of the M60-like
276 aminopeptidase family as it is also fundamental for the maturation of EPEC biofilms grown on plastics.
277 Furthermore, we have now shown that it is also required for the maturation of *E. coli* W and ETEC
278 H10407 biofilms when grown on agar and silicone substrates. We have demonstrated that this unusual
279 function is due to SslE forming amyloid-like aggregates at pH ≤ 4.4; conditions which are also observed
280 as microenvironments across mature wild-type *E. coli* W biofilms. It is well established that
281 phenotypical heterogeneity develops during the maturation of biofilms and direct measurement of
282 extracellular pH within *E. coli* PHL628 biofilms has also shown a heterogenous distributions of pH,
283 with values ranging between 5.0 and 7.0[36]. Therefore, it is very likely that SslE can form functional
284 aggregates within transiently forming microenvironments across maturing and established biofilms, but
285 then resists solubilisation after any increase in pH.

286 Microbial amyloid structures are ubiquitous in biofilms and in gram negative bacteria most known
287 amyloid structures are synthesized by two distinct and dedicated systems[37, 38]. Curli fimbriae are formed
288 using the Csg secretion system and are the best characterized bacterial amyloid structure[39], while Fap
289 fimbriae in *Pseudomonas* are created by the Fap system[40]. Curli-like fibres are formed through
290 spontaneous aggregation of extended peptides and provide adhesion, cohesion and contribute to the
291 structural integrity of the biofilm matrix[39]. Environmental cues play a major role in the regulation of
292 curli expression [41], for example, in *E. coli* curli are generally expressed below 30°C, although clinical
293 isolates have been found that express at 37°C, and absence of curli is important for the survival of
294 EHEC strain O157:H7 in agricultural environments[42]. However, in recent years a growing family of
295 surface associated proteins have been identified in gram-positive bacteria that also display amyloid-like

296 properties[21]. These proteins have a high molecular weight and are functional as non-amyloidogenic
297 folded proteins; but in response to environmental stimuli, including changes in pH, they form functional
298 aggregates. Examples include the biofilm associated protein (Bap) from *S. aureus*[24], the related
299 enterococcal surface protein (Esp) from *E. faecalis*[23] and the *S. mutans* antigen I/II (AgI/II) protein[25].
300 Here we have shown that SslE is a new member of this family and it is the first example from a gram-
301 negative bacterium. However, it is not clear whether SslE works in synergy with curli or whether it
302 supports biofilm growth under those conditions where curli and other matrix components are not
303 expressed.

304 Polymerisation of SslE is mediated by its NT1 , NT2 and NT3 domains, which become more rigid prior
305 to SslE incorporation into growing fibres. There is also a ~5% increase in β-sheet structure within
306 monomeric SslE under acidic conditions and this suggests that up to ~70 residues could be directly
307 involved in aggregation. While at neutral pH SslE is formed of ~30% helix, SslE fibres display much
308 less helical content and the C-terminal M60-like domain is dispensable for the maturation of *E. coli*
309 biofilms when grown on agar and silicone. However, it could be that this domain while not involved in
310 polymerisation, may still decorate these fibres and be functional under different conditions, for example
311 through interacting with mucins during intestinal colonisation. We have shown that SslE fibres can
312 associate with cellulose, which is known to have an architectural role in many *E. coli* biofilms and
313 provides cohesion and elasticity[43]. Although the precise role of SslE during biofilm development is still
314 vague, its interaction with cellulose indicates that it could also have an architectural role and act as a
315 hub to promote binding between other EPS within the biofilm. Furthermore, as monomeric SslE does
316 not recognize cellulose, this suggests that new binding potentials in SslE become available after its
317 aggregation.

318 Outside of *E. coli*, high sequence identity to SslE has also been identified in enteric human pathogens
319 which include *Shigella flexneri* (93% identity), *Shigella boydii* (93% identity), *Shigella dysenteriae*
320 (92% identity), *Escherichia fergusonii* (90% identity) and *Escherichia albertii* (86% identity).
321 Furthermore, SslE homologs have been found in bacterial pathogens that infect aquatic organisms,

322 many of which also cause severe gastrointestinal diseases in humans. These are namely *Vibrio cholerae*
323 (52% identity), *Vibrio vulnificus* (51% identity), *Photobacterium piscicola* (50% identity), *Vibrio*
324 *harveyi* (39% identity), *Vibrio campbellii* (39% identity), *Vibrio crassostreae* (39% identity) and *Vibrio*
325 *parahaemolyticus* (40% identity). SslE was originally identified in *V. cholerae* (called AcfD)[44] and is
326 part of a gene cluster involved in bacterial colonisation and pathogenesis, however, outside of *E. coli*
327 little is known of its function. Based on sequence analysis[34] SslE homologues are expected to possess
328 mucolytic activity but it is unknown whether they can also regulate biofilm development.

329 Although both *E. coli* W and ETEC 10407 secrete SslE, the W strain is a harmless commensal, while
330 ETEC is a major etiological agent in developing countries[45]. SslE is also required for colonisation in a
331 rabbit model of EPEC infection[4] and is actively transcribed during ETEC infection of mice[6]. Biofilm
332 growth and the production of enterotoxins are essential for the pathogenesis of ETEC-induced disease[45]
333 and secretion of SslE and the ETEC labile toxin is via the same type II secretion system[4, 46]. This
334 suggests that SslE increases the virulence of *E. coli* pathotypes through its ability to promote biofilm
335 maturation and/or through its interactions with mucosal defences. Both enterotoxigenic and
336 enteropathogenic *E. coli* strains cause infection in the small intestine[1] where the intraluminal pH ranges
337 between approximately 6.0 to 7.5[47]. However, *E. coli* possess an acid-tolerance response that supports
338 exponential growth at pH values as low as pH 4.2[48] and due to the synthesis of organic acids by the
339 residing microbiota, during colonisation of the intestine *E. coli* will experience extracellular pH values
340 that range between 4.0 and 6.0[49]. Although this suggests that within the intestinal tract, conditions exist
341 where SslE could regulate biofilm development, it remains unclear whether this role is important for
342 infection or during extraintestinal survival. SslE represents a unique protein and further studies are now
343 required to understand its complete functions during ecology and disease.

344

345

346

347 **Methods**

348 **Bacterial strains and media**

349 Bacterial strains and plasmids used in this study are listed in Supplementary Table 4. Lysogeny broth
350 (LB) medium contained (per litre) 10 g tryptone, 5 g yeast extract and 10 g NaCl. Isotopically defined
351 M9 minimal medium (pH 7.4) contained (per litre) 6.0 g $Na_2HPO_4 \cdot 7H_2O$, 3 g KH_2PO_4, 0.5 g NaCl,
352 0.12 g $MgSO_4 \cdot 7H_2O$, 0.022 g $CaCl_2$, 0.04 g thiamine, 8.3 mg $FeCl_3 \cdot 6H_2O$, 0.5 mg $ZnCl_2$, 0.1 mg $CuCl_2$,
353 0.1 mg $CoCl_2 \cdot 6H_2O$, 0.1 mg H_3BO_3 and 13.5 mg $MnCl_2 \cdot 6H_2O$, supplemented with 2g [U-$^{13}C_6$]glucose
354 and/or 0.7 g $^{15}NH_4Cl$. M9 media was made up in deuterium oxide for the production of perdeuterated
355 protein samples and pH was adjusted using 1 M NaOH solution. All NMR isotopes were from Sigma.

356 **Gene deletion**

357 Plasmids and primers used in this study are listed in Supplementary Table 4. A non-polar deletion of
358 *sslE* was constructed in *E. coli* W and H10407 strains by allelic exchange with FLP recombination
359 target (FRT)-flanked Kan^r, using a modified protocol[50]. This fragment was amplified from *E. coli* strain
360 JW5925-1[51] using Platinum *Taq* DNA polymerase (Promega) and primers PC3 and PC4 with
361 specificity for the regions flanking *sslE*. These exchanges were facilitated by the λ Red recombinase
362 system carried on plasmid pKD46 and knockouts were confirmed by PCR and sequencing.

363 **Plasmid construction**

364 Plasmids and primers used in this study are listed in Supplementary Table 4. Synthetic genes used in
365 this study are listed in Supplementary Table 5. Complementation plasmids pCPC1 and pCPC2 were
366 generated by amplifying whole length *sslE*, or *sslE* minus the M60 domain, from *E. coli* W gDNA using
367 primer pairs PC7/PC8 or PC7/PC9, respectively. These were digested with HindIII/NheI (NEB), ligated
368 into HindIII/NheI-digested pBad-cm18 vector. Synthetic genes gPC4, gPC5, gPC6 and gPC7 were
369 synthesized by Synbio Technologies and cloned into pET28b vector using NcoI and XhoI restriction
370 sites to create plasmids pPC2, pPC3, pPC4 and pPC5, respectively. Plasmid pBD1 was created by

371 amplification of the RgpB-CTD from *P. gingivalis* W50 gDNA using primers BD1 and BD2. This was
372 then cloned into pET46 Ek/LIC vector through ligation independent cloning (Novagen).

373 **Microfluidic biofilm growth**

374 Biofilms were grown in a BioFlux 200 device (Labtech, UK) using a protocol adapted from the
375 manufacturer. Single colonies of *E. coli* W and its derivatives were resuspended in 10 ml of LB (with
376 additional 50 µg/ml kanamycin for *Δssle* mutant and 50 µg/ml kanamycin, 25 µg/ml chloramphenicol
377 *Δssle::pCPC1* (*sslE*) and *Δssle::pCPC2* (*sslEΔM60*)) and incubated at 37°C with shaking (200 rpm) for
378 16 hrs. All subsequent media contained appropriate antibiotics where required. The cultures were then
379 diluted (1:100) in 10 ml of LB to OD_{600nm} of 0.5. A 24 well Bioflux plate (0-20 dyn/cm^2) was pre-
380 warmed to 37°C on a heated stage and microfluidic channels were incubated with prewarmed LB. The
381 channels were then inoculated by injecting 20 µl of bacterial suspension into the output reservoir for 5
382 s at 2 dyne/cm^2. The microfluidic plate was incubated for 1 hr at 37°C to allow bacteria to bind to the
383 surface and then the flow reversed for 20 sec. Prewarmed LB was then added to the input reservoir, the
384 flow of media was initiated at 2 dyne/cm^2 for 5 min and then decreased to 1 dyne/cm^2 for 18 hrs. The
385 spent media was removed from the output reservoir, fresh prewarmed media was added to the input
386 reservoir and the flow was lowered to 0.5 dyne/cm^2 for up to 96 hrs.

387 **Biofilm pH measurement**

388 The ratiometric dye seminaphthorhodafluor-4F 5-(and-6) carboxylic acid (C-SNARF-4) was used to
389 directly measure pH across biofilms within the Bioflux channel, using a modified method[52]. HEPES
390 buffer was first adjusted between pH 3.2 to 7.8 in 0.2 increments using HCl, and 95 µl of each respective
391 solution was then mixed with 5 µl 1mM C-SNARF-4, resulting in a final 50 µM concentration of C-
392 SNARF-4. This solution was flushed into the Bioflux plate channel, pre-warmed at 37°C and the
393 channels were imaged at x64 magnification using a DM-IRE2 confocal laser scanning microscope
394 (Leica Microsystems Heidelberg GmbH, Germany) and the accompanying Leica Microsystems
395 Confocal Software (version 2.61 Build 1537). Simultaneous images were captured using an excitation

396 wavelength of 488 nm and emission wavelengths of 580 nm and 640 nm, in triplicate for each pH
397 increment with a background image obtained after each acquisition. Acquired images were analysed
398 using ImageJ Software (FIJI)[53] by determining the average intensity of fluorescence and standard
399 deviation of each channel, minus the background control. The intensity ratio for each pH was used to
400 make a standard curve by plotting known pH against green/red ratio. 50μM C-SNARF-4 was then
401 flushed into Bioflux plates containing *E. coli* W biofilm after either 24 or 96 hrs growth and five
402 separate images were acquired for each Bioflux channel to calculate the pH from different areas of the
403 biofilm. The fluorescence intensity and standard deviation were recorded, and the ratiometric value was
404 used to determine pH from the standard curve. A 12×12 grid (~30 μm^2 area) was applied to each image,
405 each box was used to define regions of interest (ROI) and an average measure of pH was calculated
406 within each region. The average pH for ROIs within the biofilm fringes and centres is presented.

407 **Bacterial culture ring assay**

408 10 ml of LB containing 0.8 mg/ml congo red dye, adjusted to either pH 5.0 or 7.0 with HCl, were
409 inoculated with wild-type *E. coli* W or H10407 strains and their derivatives (including 50 μg/ml
410 kanamycin for *Δssle* mutant and 50 μg/ml kanamycin, 25 μg/ml chloramphenicol for *Δssle::pCPC1*
411 (*sslE*) and *Δssle::pCPC2* (*sslEΔM60*)). These were incubated overnight at 37°C while shaking (200
412 rpm) and the presence of ring formation on the glass tube was visually assessed.

413 **Protein purification**

414 Recombinant SslE (residues 67-1497; numbered based on mature sequence) was expressed and
415 purified as described previously[12]. Likewise, SslE NT1 (residues 67-211), NT2 (residues 230-425),
416 NT1-NT2 (residues 67-425), NT3-M60 (residues 426-1497) and RgpB-CTD were transformed into *E.*
417 *coli* SHuffle cells (SslE; New England Biolabs) or BL21 (DE3) (RgpB-CTD; New England Biolabs),
418 grown at 37°C in LB media (rSslE, RgpB-CTD: 100 μg/ml ampicillin; NT1, NT2, NT1-NT2, NT3-
419 M60: 100 μg/ml kanamycin) and expression induced with 0.5 mM isopropyl-d-1-thiogalactopyranoside
420 (IPTG) at an OD_{600nm} of 0.6, followed by growth overnight at 18°C. Cells were resuspended in 20 mM

421 Tris–HCl pH 8, 200 mM NaCl, 5 mM $MgCl_2$, 1 mg/ml DNase I, 5 mg/ml lysozyme, lysed by sonication
422 and purified using nickel affinity chromatography followed by gel filtration using either a Superdex 75
423 (NT1, NT2, RgpB-CTD) or 200 (rSslE, NT1-NT2, NT3-M60) column (GE Healthcare).

424 **Recombinant protein ring assay**

425 20 µl of rSslE, NT1-NT2 or NT3-M60 (62.5 µM) in 10 mM Tris-HCl pH 8, 100 mM NaCl buffer were
426 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
427 and 7.0. These were transferred to glass tubes, incubated overnight at 37°C while shaking (200 rpm)
428 and the presence of ring formation on the glass tube was visually assessed. Buffer from tubes containing
429 rings formed at pH 4.0 was then removed without disturbing the rings and 3 ml of 100 mM citrate
430 phosphate buffer over the same pH range was added. These were left overnight with shaking at 37°C
431 and solubilization of the rings were visually assessed the next day.

432 **ThT binding**

433 15 µl of 50 µM rSslE in 10 mM Tris-HCl pH 8, 100 mM NaCl buffer was mixed with 1.5 µl of 1 mM
434 Thioflavin T (ThT) dye and transferred to a 96-well plate. 100 mM citrate phosphate buffer between
435 pH 4.0 and 8.0 was added to a final volume of 150 µl and fluorescence data was collected
436 (excitation/emission 438/480 nm) at 37°C with shaking every 15 min over 24 hrs using a BMG
437 CLARIOstar plate reader.

438 **SEC-SAXS**

439 SAXS data were collected on beamline B21 at the Diamond Light Source (DLS), United Kingdom at
440 25°C. 60 µl of rSslE (10 mg/ml) in 20 mM Tris–HCl pH 8, 200 mM NaCl was applied to a KW403-4F
441 column (Shodex) at 0.16 ml/min and SAXS data were measured over a momentum transfer range of
442 $0.003 < q < 0.44$ Å^{-1}. Peak integration and buffer subtraction were performed in CHROMIXS[54]. The
443 radius of gyration (R_g) and scattering at zero angle [$I(0)$] were calculated from the analysis of the Guinier
444 region by AUTORG[54]. The distance distribution function [$P(r)$] was subsequently obtained using

445 GNOM[54] yielding the maximum particle dimension (D_{max}). *Ab initio* low resolution shape restoration
446 was carried out by calculating 20 models using DAMMIF[54], which were subsequently averaged using
447 DAMAVER[54] and used as a staring model for refinement in DAMMIN[54]. An additional 20 models were
448 also calculated and averaged in DAMAVER[54]. CRYSOL[54] was used to compare final rSslE TEM
449 envelopes and models against the solution SAXS curve. Processing and refinement statistics can be
450 found in **Supplementary Table 2**.

451 **SAXS fibrillation analysis**

452 SAXS data were collected on beamline B21 at the DLS, UK at 20°C. Immediately prior to data
453 collection, 10 mg/ml rSslE in 20 mM Tris–HCl pH 8, 200 mM NaCl was buffer exchanged into 20 mM
454 citrate-phosphate buffer, pH 4.4 using a PD10 column (GE Healthcare) and the flow through was used
455 as a buffer reference. Using a peristaltic pump and while constantly stirring, 5 ml of rSslE (0.7 mg/ml)
456 was circulated through the SAXS imaging cuvette and data were collected every 30 min over a
457 momentum transfer range of $0.004 < q < 0.4$ Å^{-1}, with the initial scattering data captured at 30 min after
458 initiating fibre growth. Data collected over the course of 11 hrs, consisting of 22 profiles, was
459 decomposed using COSMiCS[30], which utilizes MCR-ALS[55] to perform model free decomposition of
460 the entire SAXS data set. The SEC-SAXS curve collected at pH 4.4 was also included in the dataset as
461 the representative state of the protein at time 0 s, *i.e.* before initiating fibrillation. The time 0 s curve
462 was selected as one of the initial estimates and the selectivity restraint was used to ascertain that the
463 curve had no contribution from the other species. In addition, non-negativity restraint was imposed for
464 both the SAXS profiles and concentrations using FNNLS algorithm[56]. Before COSMiCS analysis,
465 SAXS data were scaled according to the large angle data to enhance the decomposition capacity of the
466 approach. Although COSMiCS was run assuming a two-component system as suggested by Principal
467 Component Analysis (PCA), a three-species run was also performed. The convergence criterion of <
468 0.01 % change in lack of fit was used with 1,000 maximum allowed iterations. The analysis of the
469 resulting COSMiCS curves was performed with the ATSAS suit of programs[54].

470

471 **TEM single particle analysis**

472 4 µl of rSslE (625 nM) in 50 mM Tris-HCl pH 8.0, 150 mM NaCl was applied to previously glow-
473 discharged 300 mesh continuous carbon-coated copper grids (Agar Scientific Ltd) for 1 min and blotted
474 for excess liquid. 4 µl of 2% (v/v) uranyl acetate was applied for staining for 1 min. The excess liquid
475 was blotted and left to dry. Data was acquired using a JEOL JEM-2100 plus TEM operating at 200 kV
476 equipped with a OneView 16 Megapixel camera (Gatan). 50 micrographs were collected at a nominal
477 magnification of 50,000x with a pixel size of 2.1 Å/pixel and a range of defocus from 1 to 3 µm. Data
478 were processed using Relion 3.1[57]. Defocus and astigmatism parameters were estimated using
479 CTFFIND4[58] in Relion 3.1. An initial dataset of 41,752 particles were autopicked using 2D class
480 averages generated using manually picked particles (approx. 1500 particles) as reference. After a few
481 rounds of 2D classification ignoring CTF until the first peak, 10,988 particles were taken forward for
482 an initial model generation using 3D initial model in Relion 3.1[57]. Following initial model generation,
483 a few rounds of 3D classification followed by 3D refinement was carried out. The final model was
484 refined to 22 Å using the gold standard FSC (0.143 criterion). PHYRE2[59] was used to analyse the SslE
485 sequence (residues 67-1497; UniProt ID E0IW31) and generate a homology model for the C-terminal
486 M60 domain (residues 962-1415) based on residues 421-894 of *P. aueriginosa* IMPa (PDB ID code
487 5kdv; 100% confidence, 21% identity)[19]. This was then docked into the rSslE TEM envelope using
488 UCSF CHIMERA[60].

489 **Solution NMR spectroscopy**

490 NMR measurements for rSslE were performed at 25°C on a 100 µM $^{2}H^{15}N$-labelled sample in 50 mM
491 $NaPO_4$ pH 7.4, 100 mM NaCl, 10 % D_2O. NMR measurements for SslE NT1 (0.6 mM), SslE NT2 (1.3
492 mM) and SslE NT1-NT2 (0.8 mM) were performed at 25°C on ^{15}N-labelled samples in 50 mM $NaPO_4$
493 pH 7.0, 100 mM NaCl, 10 % D_2O. NMR measurements for RgpB-CTD (0.3 mM) were performed at
494 37°C on a ^{15}N-labelled sample in 20 mM $NaPO_4$ pH 6.0, 100 mM NaCl, 10 % D_2O. Transverse
495 relaxation optimized spectroscopy (TROSY) based $^{1}H^{15}N$-HSQC experiment and T_1 and T_2 relaxation
496 times for rSslE were recorded on a Bruker Avance III HD 950, equipped with a TXI cryoprobe. TROSY

497 $^{1}H^{15}N$-HSQC experiment for SslE NT1-NT2 was recorded on a on a Bruker Avance III HD 800,
498 equipped with a TCI cryoprobe. Standard $^{1}H^{15}N$-HSQC spectra of SslE NT1, SslE NT2 and RgpB-CTD
499 were recorded on a Bruker Avance III HD 700, equipped with a TCI cryoprobe. Data were processed
500 in NMRPIPE[61] and analysed/visualized with ANALYSIS[62] and NMRVIEW[63].

501 **Cellulose binding assay**

502 An Immulon 2-HB 96-well plate (VWR) was blocked for 1 hr at 25°C with 300 µl of 0.1 % (w/v) BSA
503 in PBS-Tween and then washed once with 300 µl of incubation buffer (100 mM citrate-phosphate buffer
504 pH 6.0, 0.05 % (w/v) BSA, 0.05 % Tween-20). One cellulose disc[64] was added to each well and then
505 200 µl of incubation buffer was added to cover the discs, followed by incubation for 5 min. The discs
506 were washed twice in incubation buffer and then 200 µl of either 100 µM rSslE monomer, rSslE fibre
507 (produced as described for immunoblotting) or RgpB-CTD control, all in incubation buffer, were added
508 to the cellulose discs. The plate was incubated for 3 hrs at 25°C and then the discs were transferred
509 carefully to new pre-blocked wells and washed with 200 µl incubation buffer. After repeating three
510 times, the discs were incubated with anti-His_6 HRP conjugated antibody (ThermoFisher) diluted 1:2000
511 with incubation buffer for 1 hr at 24°C. The discs were removed again, and the washing step was
512 repeated 3 times. Finally, the solution was removed from the wells and the discs were incubated with
513 150 µl of *o*-Phenylenediamine dihydrochloride (Sigma) for 30 min. The solution in the wells was stirred
514 thoroughly, the discs were removed and then data was recorded at 450 nm.

515

516 **Acknowledgements**

517 PC was supported by the Academy of Medical Sciences and Wellcome Trust (SBF002/1150) and SR
518 was supported by the Medical Research Council (MR/R017662/1), awarded to JAG. SW and BD were
519 supported by China Scholarship Council and Biotechnology and Biological Sciences Research Council
520 studentships, respectively, awarded to JAG and MAC. LS was supported by the Leverhulme Trust
521 (RPG-2017-222), awarded to JAG. KF was supported by a PhD studentship from Queen Mary

522 University of London and VCD was supported by a start-up grant from Queen Mary University of
523 London. This work was also supported by the Labex EpiGenMed, an « Investissements d'avenir »
524 program (ANR-10-LABX-12-01) awarded to PB. The CBS is a member of France-BioImaging (FBI)
525 and the French Infrastructure for Integrated Structural Biology (FRISBI), 2 national infrastructures
526 supported by the French National Research Agency (ANR-10-INBS-04-01 and ANR-10-INBS-05,
527 respectively). We thank the beamline scientists at B21 of the Diamond Light Source, United Kingdom.
528 This work was also supported by the Francis Crick Institute through provision of access to the MRC
529 Biomedical NMR centre. The Francis Crick Institute receives its core funding from Cancer Research
530 United Kingdom (FC001029), the United Kingdom Medical Research Council (FC001029), and the
531 Wellcome Trust (FC001029). We also thank the Centre for Biomolecular Spectroscopy at King's
532 College London for additional NMR access, funded by the Wellcome Trust and British Heart
533 Foundation (ref. 202767/Z/16/Z and IG/16/2/32273 respectively). Cellulose discs were gifted by Prof.
534 Tom Ellis and Kenneth T. Walker (Imperial College London).

535

536 **Author Contributions**

537 Conceived and designed the experiments: PC, SW, SR, KF, AS, SS, BD, LC, LMS, SL, JJ, GC, MAC,
538 PB, VCD, JAG. Performed the experiments: PC, SW, SR, KF, AS, SS, BD, LC, LMS, SL, JJ, VCD,
539 JAG. Analyzed the data: PC, SW, SR, KF, PB, VCD, JAG. Contributed reagents/materials/analysis
540 tools: SS, SL, JJ, PB, VCD, JAG. Wrote the paper: PC, SW, SR, KF, AS, SS, BD, LC, LMS, SL, JJ,
541 GC, MAC, PB, VCD, JAG.

542

543

544

545

546 **References**

547 1. Kaper JB, Nataro JP, Mobley HL. Pathogenic Escherichia coli. *Nat Rev Microbiol* **2**, 123-140
548 (2004).
549

550 2. Gu S, Shevchik VE, Shaw R, Pickersgill RW, Garnett JA. The role of intrinsic disorder and
551 dynamics in the assembly and function of the type II secretion system. *Biochim Biophys Acta*
552 *Proteins Proteom* **1865**, 1255-1266 (2017).
553

554 3. Moriel DG*, et al.* Identification of protective and broadly conserved vaccine antigens from
555 the genome of extraintestinal pathogenic Escherichia coli. *Proc Natl Acad Sci U S A* **107**,
556 9072-9077 (2010).
557

558 4. Baldi DL*, et al.* The type II secretion system and its ubiquitous lipoprotein substrate, SslE, are
559 required for biofilm formation and virulence of enteropathogenic Escherichia coli. *Infect*
560 *Immun* **80**, 2042-2052 (2012).
561

562 5. Decanio MS, Landick R, Haft RJ. The non-pathogenic Escherichia coli strain W secretes SslE
563 via the virulence-associated type II secretion system beta. *BMC Microbiol* **13**, 130 (2013).
564

565 6. Nesta B*, et al.* SslE elicits functional antibodies that impair in vitro mucinase activity and in
566 vivo colonization by both intestinal and extraintestinal Escherichia coli strains. *PLoS Pathog*
567 **10**, e1004124 (2014).
568

569 7. Yang J, Baldi DL, Tauschek M, Strugnell RA, Robins-Browne RM. Transcriptional
570 regulation of the yghJ-pppA-yghG-gspCDEFGHIJKLM cluster, encoding the type II
571 secretion pathway in enterotoxigenic Escherichia coli. *J Bacteriol* **189**, 142-150 (2007).
572

573 8. Naili I*, et al.* Mixed mucosal-parenteral immunizations with the broadly conserved
574 pathogenic Escherichia coli antigen SslE induce a robust mucosal and systemic immunity
575 without affecting the murine intestinal microbiota. *Vaccine* **37**, 314-324 (2019).
576

577 9. Luo Q, Qadri F, Kansal R, Rasko DA, Sheikh A, Fleckenstein JM. Conservation and
578 immunogenicity of novel antigens in diverse isolates of enterotoxigenic Escherichia coli.
579 *PLoS Negl Trop Dis* **9**, e0003446 (2015).
580

581 10. Valeri M*, et al.* Pathogenic E. coli exploits SslE mucinase activity to translocate through the
582 mucosal barrier and get access to host cells. *PLoS One* **10**, e0117486 (2015).
583

584 11. Luo Q*, et al.* Enterotoxigenic Escherichia coli secretes a highly conserved mucin-degrading
585 metalloprotease to effectively engage intestinal epithelial cells. *Infect Immun* **82**, 509-521
586 (2014).
587

588 12. Rehman S, *et al.* Structure and functional analysis of the Legionella pneumophila chitinase
589 ChiA reveals a novel mechanism of metal-dependent mucin degradation. *PLoS Pathog* **16**,
590 e1008342 (2020).
591

592 13. Corfield AP. Mucins: a biologically relevant glycan barrier in mucosal protection. *Biochim*
593 *Biophys Acta* **1850**, 236-252 (2015).
594

595 14. Garnett JA, Matthews S. Interactions in bacterial biofilm development: a structural
596 perspective. *Curr Protein Pept Sci* **13**, 739-755 (2012).
597

598 15. Vogeleer P, Tremblay YD, Mafu AA, Jacques M, Harel J. Life on the outside: role of
599 biofilms in environmental persistence of Shiga-toxin producing Escherichia coli. *Front*
600 *Microbiol* **5**, 317 (2014).
601

602 16. Chitlapilly Dass S, *et al.* Impact of mixed biofilm formation with environmental
603 microorganisms on E. coli O157:H7 survival against sanitization. *NPJ Sci Food* **4**, 16 (2020).
604

605 17. Rossi E, *et al.* "It's a gut feeling" - Escherichia coli biofilm formation in the gastrointestinal
606 tract environment. *Crit Rev Microbiol* **44**, 1-30 (2018).
607

608 18. Nakjang S, Ndeh DA, Wipat A, Bolam DN, Hirt RP. A novel extracellular metallopeptidase
609 domain shared by animal host-associated mutualistic and pathogenic microbes. *PLoS One* **7**,
610 e30287 (2012).
611

612 19. Noach I, *et al.* Recognition of protein-linked glycans as a determinant of peptidase activity.
613 *Proc Natl Acad Sci U S A* **114**, E679-E688 (2017).
614

615 20. Trastoy B, Naegeli A, Anso I, Sjogren J, Guerin ME. Structural basis of mammalian mucin
616 processing by the human gut O-glycopeptidase OgpA from Akkermansia muciniphila. *Nat*
617 *Commun* **11**, 4844 (2020).
618

619 21. Taglialegna A, Lasa I, Valle J. Amyloid Structures as Biofilm Matrix Scaffolds. *J Bacteriol*
620 **198**, 2579-2588 (2016).
621

622 22. Belousov MV, *et al.* M60-like metalloprotease domain of the Escherichia coli YghJ protein
623 forms amyloid fibrils. *PLoS One* **13**, e0191317 (2018).
624

625 23. Taglialegna A, *et al.* The biofilm-associated surface protein Esp of Enterococcus faecalis
626 forms amyloid-like fibers. *NPJ Biofilms Microbiomes* **6**, 15 (2020).
627

628 24. Taglialegna A, *et al.* Staphylococcal Bap Proteins Build Amyloid Scaffold Biofilm Matrices
629 in Response to Environmental Signals. *PLoS Pathog* **12**, e1005711 (2016).
630

631 25. Chen D*, et al.* Characteristics and influencing factors of amyloid fibers in S. mutans biofilm.
632 *AMB Express* **9**, 31 (2019).
633

634 26. Gallardo R, Ranson NA, Radford SE. Amyloid structures: much more than just a cross-beta
635 fold. *Curr Opin Struct Biol* **60**, 7-16 (2020).
636

637 27. Moran SD, Zanni MT. How to Get Insight into Amyloid Structure and Formation from
638 Infrared Spectroscopy. *J Phys Chem Lett* **5**, 1984-1993 (2014).
639

640 28. Conway KA, Harper JD, Lansbury PT, Jr. Fibrils formed in vitro from alpha-synuclein and
641 two mutant forms linked to Parkinson's disease are typical amyloid. *Biochemistry* **39**, 2552-
642 2563 (2000).
643

644 29. Buchan DWA, Jones DT. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic*
645 *Acids Res* **47**, W402-W407 (2019).
646

647 30. Herranz-Trillo F, Groenning M, van Maarschalkerweerd A, Tauler R, Vestergaard B,
648 Bernado P. Structural Analysis of Multi-component Amyloid Systems by Chemometric
649 SAXS Data Decomposition. *Structure* **25**, 5-15 (2017).
650

651 31. Giehm L, Svergun DI, Otzen DE, Vestergaard B. Low-resolution structure of a vesicle
652 disrupting α-synuclein oligomer that accumulates during fibrillation. *Proc Natl Acad*
653 *Sci U S A* **108**, 3246-3251 (2011).
654

655 32. Groenning M, Campos RI, Hirschberg D, Hammarstrom P, Vestergaard B. Considerably
656 Unfolded Transthyretin Monomers Preceed and Exchange with Dynamically Structured
657 Amyloid Protofibrils. *Sci Rep* **5**, 11443 (2015).
658

659 33. Vestergaard B*, et al.* A helical structural nucleus is the primary elongating unit of insulin
660 amyloid fibrils. *PLoS Biol* **5**, e134 (2007).
661

662 34. Gish W, States DJ. Identification of protein coding regions by database similarity search. *Nat*
663 *Genet* **3**, 266-272 (1993).
664

665 35. Ta BTT*, et al.* Identification, recombinant protein production, and functional analysis of a
666 M60-like metallopeptidase, secreted by the liver fluke Opisthorchis viverrini. *Parasitol Int*
667 **75**, 102050 (2020).
668

669 36. Hidalgo G*, et al.* Functional tomographic fluorescence imaging of pH microenvironments in
670 microbial biofilms by use of silica nanoparticle sensors. *Appl Environ Microbiol* **75**, 7426-
671 7435 (2009).
672

673 37. Van Gerven N, Van der Verren SE, Reiter DM, Remaut H. The Role of Functional Amyloids
674 in Bacterial Virulence. *J Mol Biol* **430**, 3657-3684 (2018).
675

676 38. Shanmugam N, Baker M, Ball SR, Steain M, Pham CLL, Sunde M. Microbial functional
677 amyloids serve diverse purposes for structure, adhesion and defence. *Biophys Rev* **11**, 287-
678 302 (2019).
679

680 39. Evans ML, Chapman MR. Curli biogenesis: order out of disorder. *Biochim Biophys Acta*
681 **1843**, 1551-1558 (2014).
682

683 40. Dueholm MS*, et al.* Functional amyloid in Pseudomonas. *Mol Microbiol* **77**, 1009-1020
684 (2010).
685

686 41. Bokranz W, Wang X, Tschape H, Romling U. Expression of cellulose and curli fimbriae by
687 Escherichia coli isolated from the gastrointestinal tract. *J Med Microbiol* **54**, 1171-1182
688 (2005).
689

690 42. Ravva SV, Sarreal CZ, Cooley MB. Expression of Curli by Escherichia coli O157:H7 Strains
691 Isolated from Patients during Outbreaks Is Different from Similar Strains Isolated from Leafy
692 Green Production Environments. *Front Cell Infect Microbiol* **6**, 189 (2016).
693

694 43. Serra DO, Richter AM, Hengge R. Cellulose as an architectural element in spatially
695 structured Escherichia coli biofilms. *J Bacteriol* **195**, 5540-5554 (2013).
696

697 44. Parsot C, Mekalanos JJ. Structural analysis of the acfA and acfD genes of Vibrio cholerae:
698 effects of DNA topology and transcriptional activators on expression. *J Bacteriol* **174**, 5211-
699 5218 (1992).
700

701 45. Qadri F, Svennerholm AM, Faruque AS, Sack RB. Enterotoxigenic Escherichia coli in
702 developing countries: epidemiology, microbiology, clinical features, treatment, and
703 prevention. *Clin Microbiol Rev* **18**, 465-483 (2005).
704

705 46. Sjoling A, von Mentzer A, Svennerholm AM. Implications of enterotoxigenic Escherichia
706 coli genomics for vaccine development. *Expert Rev Vaccines* **14**, 551-560 (2015).
707

708 47. Fallingborg J. Intraluminal pH of the human gastrointestinal tract. *Dan Med Bull* **46**, 183-196
709 (1999).
710

711 48. Xu Y*, et al.* An acid-tolerance response system protecting exponentially growing Escherichia
712 coli. *Nat Commun* **11**, 1496 (2020).
713

714 49. Lin J, Smith MP, Chapin KC, Baik HS, Bennett GN, Foster JW. Mechanisms of acid
715 resistance in enterohemorrhagic Escherichia coli. *Appl Environ Microbiol* **62**, 3094-3100
716 (1996).
717

718 50. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in Escherichia coli
719 K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640-6645 (2000).
720

721 51. Baba T*, et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants:
722 the Keio collection. *Mol Syst Biol* **2**, 2006 0008 (2006).
723

724 52. Schlafer S, Dige I. Ratiometric Imaging of Extracellular pH in Dental Biofilms. *J Vis Exp*,
725 (2016).
726

727 53. Schindelin J*, et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods*
728 **9**, 676-682 (2012).
729

730 54. Franke D*, et al.* ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering
731 from macromolecular solutions. *J Appl Crystallogr* **50**, 1212-1225 (2017).
732

733 55. Jaumot J, Gargallo R, de Juan A, Tauler R. A graphical user-friendly interface for MCR-ALS:
734 a new tool for multivariate curve resolution in MATLAB. *Chemometrics and Intelligent*
735 *Laboratory Systems* **76**, 101-110 (2005).
736

737 56. Bro R, De Jong S. A fast non-negativity-constrained least squares algorithm. *Journal of*
738 *Chemometrics* **11**, 393-401 (1997).
739

740 57. Zivanov J*, et al.* New tools for automated high-resolution cryo-EM structure determination in
741 RELION-3. *Elife* **7**, (2018).
742

743 58. Rohou A, Grigorieff N. CTFFIND4: Fast and accurate defocus estimation from electron
744 micrographs. *J Struct Biol* **192**, 216-221 (2015).
745

746 59. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for
747 protein modeling, prediction and analysis. *Nat Protoc* **10**, 845-858 (2015).
748

749 60. Pettersen EF*, et al.* UCSF Chimera--a visualization system for exploratory research and
750 analysis. *J Comput Chem* **25**, 1605-1612 (2004).
751

752 61. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional
753 spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-293 (1995).
754

755 62. Vranken WF, *et al.* The CCPN data model for NMR spectroscopy: development of a software
756 pipeline. *Proteins* **59**, 687-696 (2005).
757

758 63. Johnson BA, Blevins RA. NMR View: A computer program for the visualization and analysis
759 of NMR data. *J Biomol NMR* **4**, 603-614 (1994).
760

761 64. Florea M, *et al.* Engineering control of bacterial cellulose production using a genetic toolkit
762 and a new cellulose-producing strain. *Proc Natl Acad Sci U S A* **113**, E3431-3440 (2016).
763

764

Figures

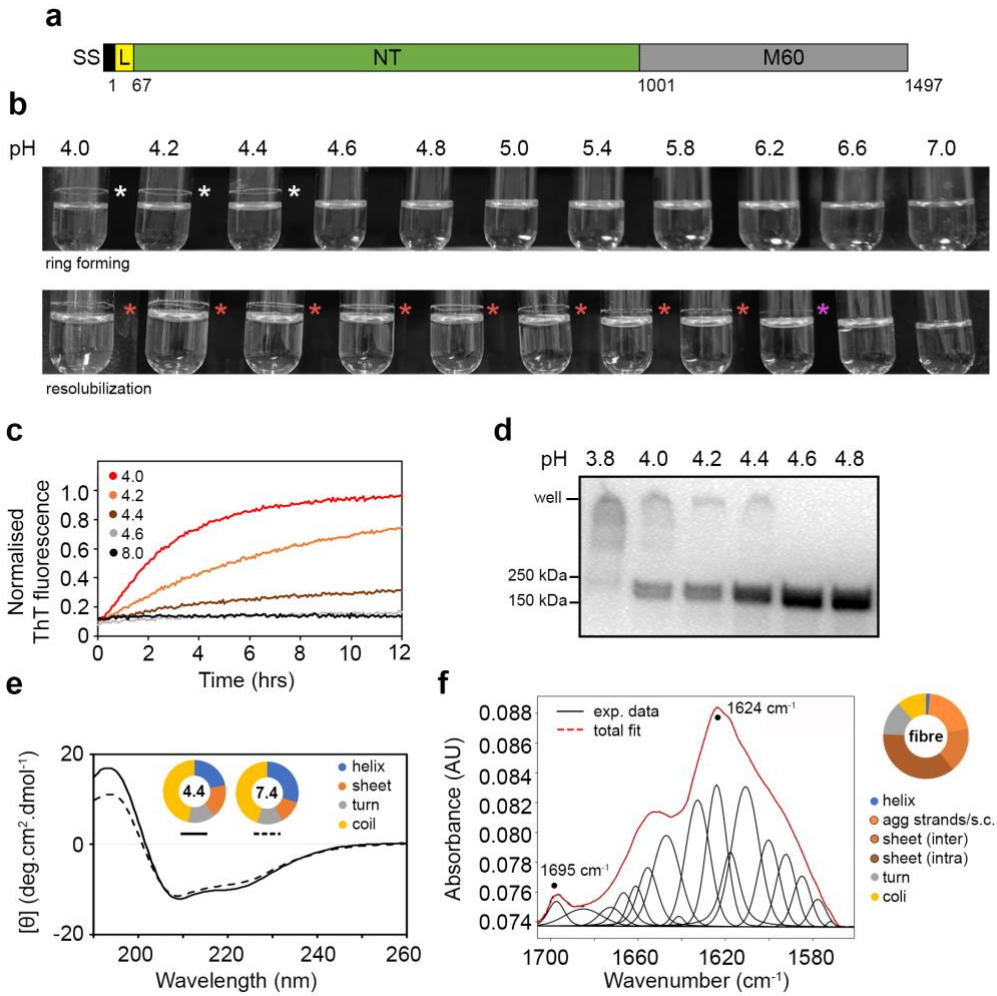

Fig. 1 | Characterization of SslE aggregates. a, Schematic of SslE from *E. coli* W with mature sequence numbers and structural features annotated. SS: periplasmic signal sequence; L: flexible linker; NT: unique N-terminal region; M60: peptidase/mucinase domain. **b,** Purified rSslE forms a clear ring of aggregated protein when incubated in citrate-phosphate buffer between pH 4.0 and 4.4 (upper panel; white asterisks). Rings resist solubilization up to pH 6.2 (lower panel; purple asterisk), although some loss of protein band is observed after pH 5.8 (lower panel; red asterisks). **c,** Increase in ThT fluorescence emission on binding to rSslE fibres formed in citrate-phosphate buffer over pH 4.0 to 8.0. **d,** Immunoblot of rSslE incubated in citrate-phosphate buffer from pH 3.8 to 4.8 and detected with anti-His_6 antibody. rSslE present in the well represents aggregated protein. **e,** Far-UV CD spectra of soluble rSslE 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 the amide I region of rSslE fibres. Deconvolution is shown for the underlying peaks, which correspond to aggregated β-strands, β-sheets, random coil, α-helical and β-turn structures. All data are representative of three independent experiments.

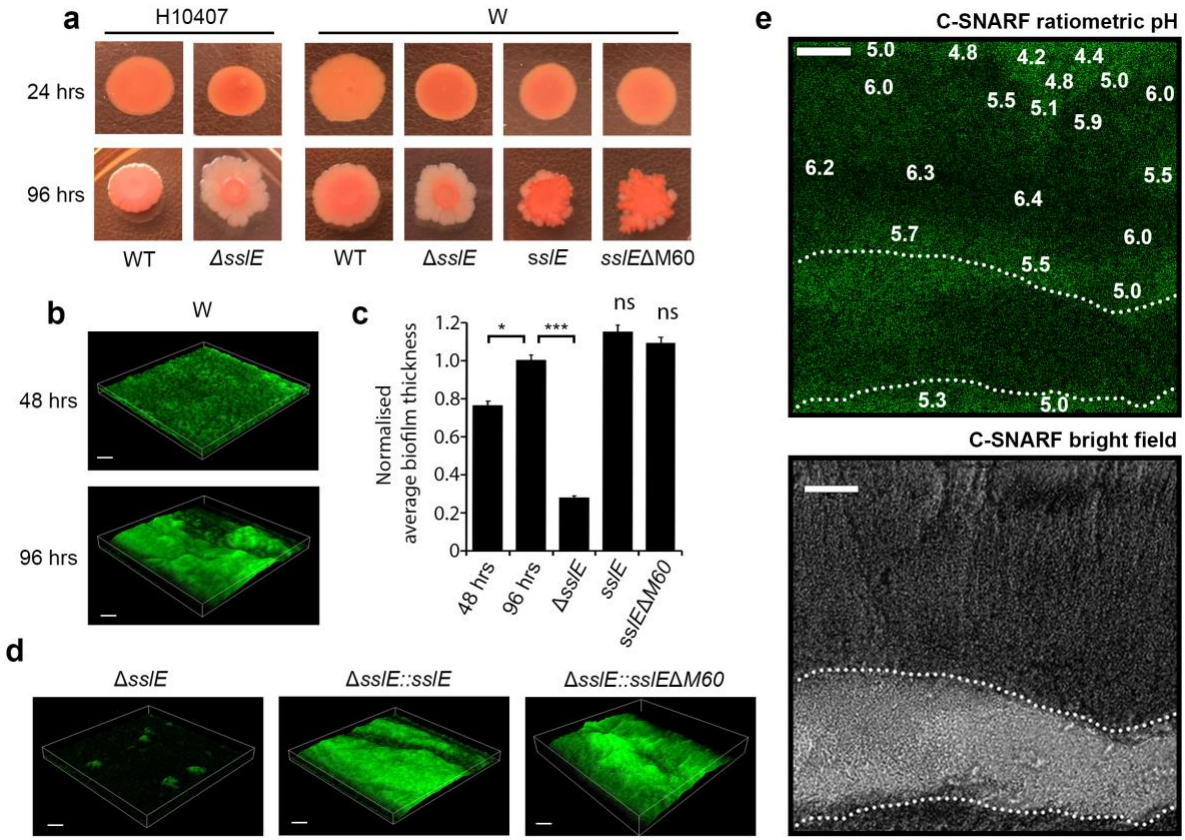

Fig. 2 | Analysis of SslE-dependent biofilm formation by *E. coli* W and H10407 strains. a, Macrocolony biofilm phenotype of wild-type *E. coli* W and H10407 strains and their derivatives on Congo red agar medium after 24 and 96 hrs of growth. The derivatives *sslE* and *sslE*ΔM60 represent *trans*-complementation of intact *sslE* or *sslE* minus its M60 domain, respectively, into the W Δ*sslE* mutant. **b,** Confocal laser scanning images of wild-type *E. coli* W biofilms grown for 28 and 96 hrs. **c,** Quantification of 3-dimensional biofilm intensity of wild-type *E. coli* W biofilms and its derivatives carried out in ImageJ[40]. $^{*}P < 0.05$; $^{***}P < 0.001$; verses W 96 hrs growth by two-tailed Student's test. **d,** Confocal laser scanning images of *E. coli* W Δ*sslE* mutant and *trans*-complementation with *sslE* (Δ*sslE::sslE*) and *sslE*ΔM60 (Δ*sslE::sslE*ΔM60). **e,** The pH across wild-type *E. coli* W biofilms grown for 96 hrs was monitored ratiometrically using C-SNARF-4. pH values were calculated over ~30 μm^2 boxes and pH values for representative regions within the biofilm fringes and centres are annotated. Dotted line outlines solvent 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

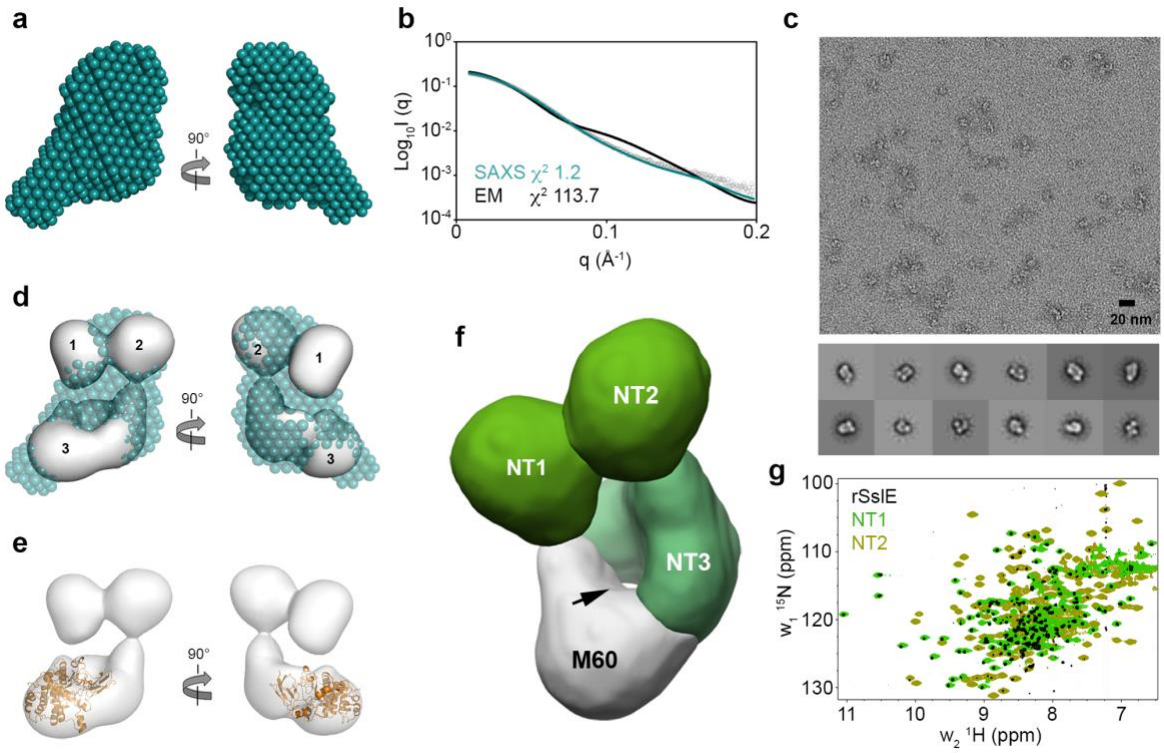

Fig. 3 | Structural features of monomeric SslE. a, SAXS bead model of rSslE at pH 7.4. **b,** Fit of the SAXS bead model (teal line) and negative-stain TEM map (black line) of rSslE to the rSslE SAXS data (black open circles) with χ^2 of 1.2 and 113.7, respectively. **c,** Representative negative-stain TEM micrograph of rSslE (pH 7.4) at 50,000× nominal magnification with representative 2D classifications. Scale bar represents 20 nm. **d,** Overlay of the 22 Å resolution TEM map (grey) and SAXS bead model (teal) with the three defined regions in SslE highlighted. **e,** Docking of an SslE M60 domain homology model (orange) into the TEM map. **f,** TEM map of rSslE colored based on domain organization. The NT3-M60 interdomain channel is highlighted with an arrow. **g,** TROSY $^{1}H^{15}N$-HSQC spectrum of rSslE (black) overlaid with $^{1}H^{15}N$-HSQC spectra of the SslE NT1 (light green) and NT2 (olive green) domains.

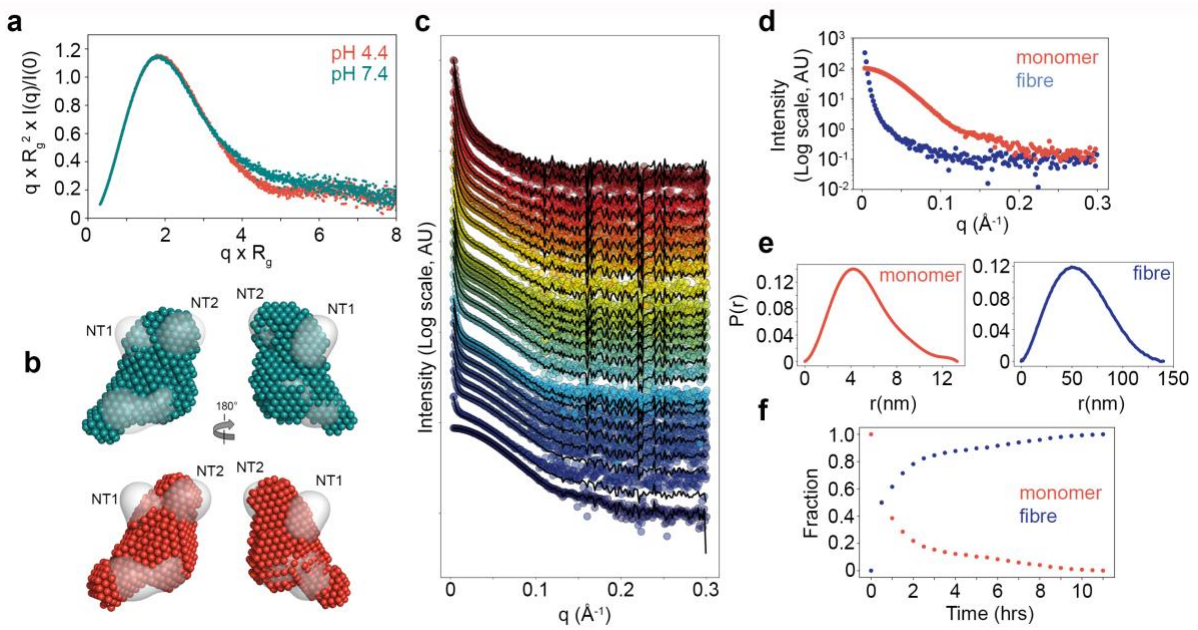

Fig. 4 | Kinetics of SslE polymerization. a, SAXS normalized Kratky plots for rSslE at pH 4.4 and 7.4. **b,** Overlay of rSslE SAXS bead models at pH 4.4 (red) and 7.4 (teal) with rSslE TEM map (grey). NT1 and NT2 domains are highlighted. COSMiCS decomposition of rSslE SAXS fibrillation data, assuming the co-existence of two components. SEC-SAXS curve for rSslE at pH 4.4 is also shown for t=0. **c,** COSMiCS decomposed scattering curves for rSslE monomer (red) and fibres (blue). **d,** Shape distribution [P(r)] functions derived from COSMiCS decomposed scattering curves for rSslE monomer (red) and fibres (blue). **f,** Relative populations of the two species of rSslE derived from the COSMiCS analysis over time.

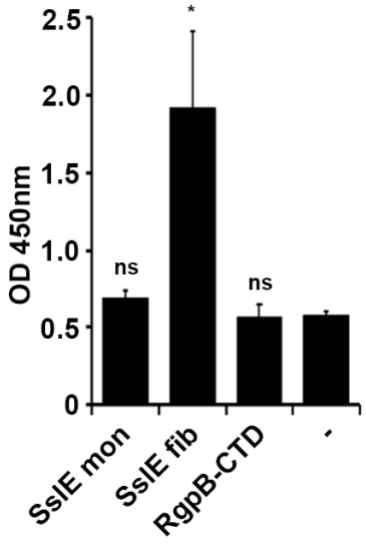

Fig. 5 | Binding properties of SslE aggregates. a, ELISA analysis of binding between cellulose discs and His-tagged monomeric rSslE, His-tagged polymeric rSslE and His-tagged *P. gingivalis* RgpB-CTD control, at pH 6.0 and detected with anti-His_6 antibody. BSA-coated wells were used as controls. $^{*}P$ <0.05; verses control empty well (-) by two-tailed Student's test.