Structural and mechanistic insights in the bacterial amyloid secretion channel CsgG

Parveen Goyal[1,2], Petya V. Krasteva[3,4], Nani Van Gerven[1,2], Francesca Gubellini[3,4], Imke Van den Broeck[1,2], Anastassia Troupiotis-Tsaïlaki[5], Wim Jonckheere[1,2], Gérard Péhau-Arnaudet[4], Jerome S. Pinkner[6], Matthew R. Chapman[7], Scott J. Hultgren[6], Stefan Howorka[8], Rémi Fronzes[3,4]* and Han Remaut[1,2]*

[1] Structural and Molecular Microbiology, Structural Biology Research Center, VIB, Pleinlaan 2, 1050 Brussels, Belgium.

[2] Structural Biology Brussels, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium.

[3] Unité G5 Biologie structurale de la sécrétion bactérienne, Institut Pasteur, 25-28 rue du Docteur Roux, 75015 Paris, France

[4] UMR 3528, CNRS, Institut Pasteur, 25-28 rue du Docteur Roux, 75015 Paris, France

[5] Structure et Fonction des Membranes Biologiques (SFMB), Université Libre de Bruxelles, Brussels, Belgium

[6] Department of Molecular Microbiology and Microbial Pathogenesis, Washington University in Saint Louis School of Medicine, St. Louis, Missouri 63110-1010, USA.

[7] Department of Molecular, Cellular and Developmental Biology, University of Michigan, Ann Arbor, MI 48109-1048, USA.

[8] Department of Chemistry, Institute for Structural and Molecular Biology University College London, London WC1H 0AJ, UK

Curli are functional amyloid fibers that constitute the major protein component of the extracellular matrix in pellicle biofilms formed by Bacteroidetes and Proteobacteria (predominantly α- and γ-)[1-3]. They provide a fitness advantage in pathogenic strains and induce a strong pro-inflammatory response during bacteremia[1,4,5]. Curli formation requires a dedicated protein secretion machinery comprised of the outer membrane lipoprotein CsgG and two soluble accessory proteins, CsgE and CsgF [6,7]. Here we report the X-ray structure of *E. coli* CsgG in a non-lipidated, soluble form as well as in its native membrane-extracted conformation. CsgG forms an oligomeric transport complex composed of nine anticodon binding domain (ABD)-like units that give rise to a 36-stranded β-barrel that traverses the bilayer and is connected to a cage-like vestibule in the periplasm. The transmembrane and periplasmic domains are separated by a 0.9 nm channel constriction composed of three stacked concentric phenylalanine-, asparagine- and tyrosine-rings that may guide the extended polypeptide substrate through the secretion pore. The specificity factor CsgE forms a nonameric adaptor that binds and closes off the periplasmic face of the secretion channel, creating a 24.000 Å^3 pre-constriction chamber. Our structural, functional and electrophysiological analyses imply that CsgG is an ungated, non-selective protein secretion channel that likely employs a diffusion-based, entropy-driven transport mechanism.

Curli are bacterial surface appendages that have structural and physical characteristics of amyloid fibrils, best known from human degenerative diseases[7-9]. However, the role of bacterial amyloids such as curli are to facilitate biofilm formation [4,10]. Unlike pathogenic amyloids, which are the product of protein misfolding, curli

formation is coordinated by proteins encoded in two dedicated operons, *csgBAC* (*curli specific genes BAC*) and *csgDEFG* in *Escherichia coli* (Extended Data Fig. 1)[6,7]. Upon secretion, CsgB nucleates CsgA subunits into curli fibers [7,11,12]. Secretion and extracellular deposition of CsgA and CsgB is dependent on two soluble accessory factors, CsgE and CsgF, respectively, as well as CsgG, a 262-residue lipoprotein located in the outer membrane (OM)[13-16]. Due to the lack of hydrolyzable energy sources or ion gradients at the outer membrane, CsgG falls into a specialized class of protein translocators that must operate through an alternatively energized transport mechanism. In the absence of a structural model, the dynamic workings of how CsgG promotes secretion and assembly of highly stable amyloid-like fibers in a regulated fashion across a biological membrane has thus far remained enigmatic.

Prior to insertion into the OM, lipoproteins are piloted across the periplasm via the lipoprotein localization (Lol) pathway[17]. We observed that non-lipidated CsgG ($CsgG_{C1S}$) could be isolated as a soluble periplasmic intermediate, analogous to the pre-pore forms observed in pore-forming proteins and toxins[18]. $CsgG_{C1S}$ was found predominantly as monomers, in addition to a minor fraction of discrete oligomeric complexes (Extended Data Fig. 2)[19]. The soluble $CsgG_{C1S}$ oligomers were crystallized and their structure was determined to 2.8 Å, revealing a hexadecameric particle with 8-fold dihedral symmetry (D_8), consisting of two ring-shaped octameric complexes (C_8) that are joined in a tail-to-tail interaction (Extended Data Fig. 2 and Fig. 1). The $CsgG_{C1S}$ protomer shows an anticodon binding domain (ABD)-like fold that is extended with two α-helices at the N- and C-terminus (αN and αC, resp.; Fig. 1 and Extended Data Fig. 3a-c). Additional CsgG-specific elements are an extended loop linking β1 and α1, two insertions in the loops connecting β3-β4 and β5-α3 and an extended α2 helix that is implicated in CsgG oligomerization by packing between

adjacent monomers (Fig. 1b). Further inter protomer contacts are formed between the back of the β3-5 sheet and the extended β1-α1 loop (Extended Data Fig. 3d, e).

In the $CsgG_{C1S}$ structure, residues 1-17, which would link α1 to the N-terminal lipid anchor, are disordered and no obvious transmembrane (TM) domain can be discerned (Fig. 1). Attenuated Total Reflection Fourier Transform Infrared spectroscopy (ATR-FTIR) of $CsgG_{C1S}$ and native, membrane-extracted CsgG reveals the latter exhibits a higher absorption in the β-sheet region (1625-1630 cm^{-1}) and a concomitant reduction in the random coil and α-helical regions (1645-1650 cm^{-1} and 1656 cm^{-1}, resp., Fig. 2a), suggesting membrane associated CsgG contains a β-barrel domain. Candidate sequence stretches for β-strand formation are found in the poorly ordered, extended loops connecting β3-β4 (residues 134-154) and β5-α3 (residues 184-204) and deletion of these resulted in the loss of curli formation (Fig. 2b). The crystal structure of detergent extracted CsgG confirms a conformational rearrangement of both regions into two adjacent β-hairpins, extending the β-sheet formed by β3-β4 (TM1) and β5-α3 (TM2), respectively (Fig. 2c). Their juxtaposition in the CsgG oligomer gives rise to a composite 36-stranded β-barrel (Fig. 2d). Remarkably, whereas the crystallized $CsgC_{C1S}$ oligomers show a D_8 symmetry, the CsgG structure shows D_9 symmetry, with CsgG protomers retaining equivalent interprotomer contacts, except for a 5° rotation relative to the central axis and a 4 Å translation along the radial axes (Extended Data Fig. 2). This observation is reconciled in the *in solution* oligomeric states revealed by single-particle EM, which exclusively finds C_9 and D_9 symmetries for membrane-extracted CsgG (Extended Data Figs 2). The predominant presence of monomers in the non-lipidated sample and the symmetry mismatch with the membrane-bound protein argue that before membrane insertion, CsgG is targeted to the outer membrane in a monomeric, LolA-bound form and that

C8 and D8 particles are an artifact of highly concentrated solutions of $CsgG_{C1S}$. Furthermore, we show the C_9 nonamer rather than the D9 complex forms the physiologically relevant particle since in isolated *E. coli* outer membranes, cysteine substitutions in residues enclosed by the observed tail-to-tail dimerization are accessible to maleimide-polyethylene glycol (PEG – 5 kDa) labeling (Extended Data Fig. 4).

Thus, CsgG forms a nonameric transport complex of 120 Å width and 85 Å height. The complex traverses the outer membrane through a 36-stranded β-barrel of 40 Å inner diameter (Fig. 2e). The N-terminal lipid anchor is separated from the core domain by an 18-residue linker that wraps over the adjacent protomer (Extended Data Fig. 3d). The diacylglycerol- and amide-linked acyl chain on the N-terminal Cys are not resolved in the electron density maps, but based on the location of Leu 2 the lipid anchor is expected to flank the outer wall of the β-barrel. On the periplasmic side, the transporter forms a large solvent–accessible cavity of 35 Å inner diameter and 40 Å height that opens to the periplasm in a 50 Å mouth formed by helix 2 (Fig. 2e). At its apex, this periplasmic vestibule is separated from the TM channel by a conserved 12-residue loop connecting β1-α1 (C-loop, "CL"; Fig. 2e, Fig. 3a, b), which constricts the secretion conduit to a solvent-excluded diameter of 9.0 Å (Fig. 3a, c). These pore dimensions would be compatible with the residence of one to two (e.g. a looped structure) extended polypeptide segments, with 5 residues spanning the height of the constriction (Extended Data Fig. 5). The luminal lining of the constriction is composed of three stacked concentric rings formed by the side chains of residues Y51, N55 and F56 (Fig. 3a, b). Strikingly, in the anthrax PA_{63} toxin, a topologically equivalent concentric Phe ring (referred to as ϕ-clamp) lines the entry of the translocation channel and catalyzes polypeptide capture and passage [20-22]. Multiple sequence alignment of

CsgG-like translocators shows the absolute conservation of the F56 and the conservative variation of N55 to Ser or Thr (Extended Data Fig. 6). Mutation of F56 or N55 to Ala leads to a near loss in curli production (Fig. 3d), whereas a N55S substitution retains wild-type secretion levels, together alluding to the requirement of the stacked configuration of a ϕ-clamp followed by a H-bond donor/acceptor in the CsgG constriction (Extended Data Fig. 6, Fig. 3b).

Single-channel current recordings of CsgG reconstituted in planar phospholipid bilayers led to a steady current of 43 ± 5 pA (n = 32), - 45 ± 4 pA (n = 13) using standard electrolyte conditions and a potential of + 50 mV or – 50 mV, respectively (Fig. 3e, f, Extended Data Fig. 7). The observed current was in good agreement with the predicted value of 46 pA calculated based on a simple three-segment pore model and the dimensions of the central constriction seen in the X-ray structure (Fig. 3c). A second, low conductance conformation can also be observed under negative electrical field potential (-25 ± 4 pA; Extended Data Fig. 7). It is unclear, however, whether this species is present under physiological conditions.

Our structural data and single-channel recordings infer CsgG forms an ungated peptide diffusion channel. In PA_{63}, a model peptide diffusion channel, polypeptide passage depends on a ΔpH-driven Brownian ratchet that rectifies the diffusive steps in the translocation channel[20-22]. Such proton gradients, however, are not present at the outer membrane, requiring an alternative driving force. Whereas at elevated concentrations CsgG facilitates non-selective diffusive leakage of periplasmic polypeptides, secretion is specific for CsgA under native conditions and requires the periplasmic factor CsgE[16,23]. In presence of an excess CsgE, purified CsgG forms a slower migrating species on native PAGE (Fig. 4a). SDS-PAGE analysis shows this new species consists of a CsgG:CsgE complex that is present in a 1:1 stoichiometry

(Fig. 4b). CryoEM visualization of CsgG:CsgE isolated by pull-down affinity purification reveals a 9-fold symmetrical particle corresponding to the CsgG nonamer and an additional capping density at the entrance to the periplasmic vestibule, similar in size and shape to a C_9 CsgE oligomer also observed by single-particle EM and size exclusion chromatography (Fig. 4c-e, Extended Data Fig. 8). The location of the observed CsgG:CsgE contact interface is corroborated by blocking point mutations in CsgG helix 2 (Extended Data Fig. 8). In agreement with a capping function, single channel recordings show that CsgE binding to the translocator leads to the specific silencing of its ion conductance (Fig. 4f, Extended Data Fig. 7). This CsgE capping of the channel appears an all or none response in function of CsgE nomamer binding. At saturation, CsgE binding induces full blockage of the channel, whilst around 10 nM, an equilibrium between CsgE binding and dissociation events results in an intermittently blocked or fully open translocator. At 1 nM or below, transient (<1 millisecond) partial blockage events may stem from short-lived encounters with monomeric CsgE.

Thus, CsgG and CsgE appear to form an encaging complex enclosing a central cavity of ~24.000 $Å^3$, reminiscent in appearance to the substrate binding cavity and encapsulating lid structure seen in the GroEL chaperonin and GroES cochaperonin[24]. The CsgG:CsgE enclosure would be compatible with the full or partial entrapment of the 129-residue CsgA. Interestingly, the caging of a translocation substrate has recently been observed in ABC toxins[25]. Spatial confinement of an unfolded polypeptide leads to a decrease in its conformational space, creating an entropic potential that has been shown to favor polypeptide folding in case of chaperonins[24,26]. Similarly, we speculate that in curli transport the local high concentration and conformational confinement of curli subunints in the CsgG vestibule would generate

an entropic free energy gradient over the translocation channel (Fig. 4g). Upon capture into the constriction, the polypeptide chain is then expected to progressively move outwards by Brownian diffusion, rectified by the entropic potential generated from the CsgE-mediated confinement and/or substrate-trapping near the secretion channel. In case of a full confinement in the pre-constriction cavity, escape of an unfolded 129-residue polypeptide to the bulk solvent would correspond to an entropic free energy release of up to ~80 kcal/mole [27]. The initial entropic cost of substrate docking and confinement are likely to be at least partially compensated by binding energy released during assembly of the CsgG:CsgE:CsgA complex and an already lowered CsgA entropy in the periplasm. On theoretical grounds, three potential routes of CsgA recruitment to the secretion complex can be envisaged (Extended Data Fig. 9).

Curli-induced biofilms form a fitness and virulence factor in pathogenic Enterobacteriaceae [4,5]. Their unique secretion and assembly properties are also rapidly gaining interest for (bio)technological application[23,28,29]. Our structural characterization and biochemical study of two key secretion components provide a tentative model of an iterative mechanism for the membrane translocation of unfolded protein substrates in absence of a hydrolysable energy source, a membrane potential or ion gradient (Fig. 4e, Extended Data Fig. 9). The full validation and deconstruction of the contributing factors in the proposed secretion model will require the *in vitro* reconstitution of the translocator in order to accurately follow transport kinetics at the single molecule level.

References:

1 Olsen, A., Jonsson, A. & Normark, S. Fibronectin binding mediated by a novel class of surface organelles on Escherichia coli. *Nature* **338**, 652-655 (1989).

2 Collinson, S. K. *et al.* Thin, aggregative fimbriae mediate binding of Salmonella enteritidis to fibronectin. *J Bacteriol* **175**, 12-18 (1993).

3 Dueholm, M. S., Albertsen, M., Otzen, D. & Nielsen, P. H. Curli functional amyloid systems are phylogenetically widespread and display large diversity in operon and protein structure. *PLoS One* **7**, e51274 (2012).
4 Cegelski, L. *et al.* Small-molecule inhibitors target Escherichia coli amyloid biogenesis and biofilm formation. *Nat Chem Biol* **5**, 913-919 (2009).
5 Herwald, H. *et al.* Activation of the contact-phase system on bacterial surfaces--a clue to serious complications in infectious diseases. *Nat Med* **4**, 298-302 (1998).
6 Hammar, M., Arnqvist, A., Bian, Z., Olsen, A. & Normark, S. Expression of two csg operons is required for production of fibronectin- and congo red-binding curli polymers in Escherichia coli K-12. *Mol Microbiol* **18**, 661-670 (1995).
7 Chapman, M. R. *et al.* Role of Escherichia coli curli operons in directing amyloid fiber formation. *Science* **295**, 851-855 (2002).
8 Wang, X., Smith, D. R., Jones, J. W. & Chapman, M. R. In vitro polymerization of a functional Escherichia coli amyloid protein. *J Biol Chem* **282**, 3713-3719 (2007).
9 Dueholm, M. S. *et al.* Fibrillation of the major curli subunit CsgA under a wide range of conditions implies a robust design of aggregation. *Biochemistry* **50**, 8281-8290 (2011).
10 Hung, C. *et al.* Escherichia coli biofilms have an organized and complex extracellular matrix structure. *MBio* **4**, e00645-00613 (2013).
11 Hammar, M., Bian, Z. & Normark, S. Nucleator-dependent intercellular assembly of adhesive curli organelles in Escherichia coli. *Proc Natl Acad Sci U S A* **93**, 6562-6566 (1996).
12 Bian, Z. & Normark, S. Nucleator function of CsgB for the assembly of adhesive surface organelles in Escherichia coli. *EMBO J* **16**, 5827-5836 (1997).
13 Loferer, H., Hammar, M. & Normark, S. Availability of the fibre subunit CsgA and the nucleator protein CsgB during assembly of fibronectin-binding curli is limited by the intracellular concentration of the novel lipoprotein CsgG. *Mol Microbiol* **26**, 11-23 (1997).
14 Robinson, L. S., Ashman, E. M., Hultgren, S. J. & Chapman, M. R. Secretion of curli fibre subunits is mediated by the outer membrane-localized CsgG protein. *Mol Microbiol* **59**, 870-881 (2006).
15 Nenninger, A. A., Robinson, L. S. & Hultgren, S. J. Localized and efficient curli nucleation requires the chaperone-like amyloid assembly protein CsgF. *Proc Natl Acad Sci U S A* **106**, 900-905 (2009).
16 Nenninger, A. A. *et al.* CsgE is a curli secretion specificity factor that prevents amyloid fibre aggregation. *Mol Microbiol* **81**, 486-499 (2011).
17 Okuda, S. & Tokuda, H. Lipoprotein sorting in bacteria. *Annu Rev Microbiol* **65**, 239-259 (2011).
18 Iacovache, I., Bischofberger, M. & van der Goot, F. G. Structure and assembly of pore-forming proteins. *Curr Opin Struct Biol* **20**, 241-246 (2010).
19 Goyal, P., Van Gerven, N., Jonckheere, W. & Remaut, H. Crystallization and preliminary X-ray crystallographic analysis of the curli transporter CsgG. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **69**, 1349-1353 (2013).
20 Krantz, B. A. *et al.* A phenylalanine clamp catalyzes protein translocation through the anthrax toxin pore. *Science* **309**, 777-781 (2005).

21 Janowiak, B. E., Fischer, A. & Collier, R. J. Effects of introducing a single charged residue into the phenylalanine clamp of multimeric anthrax protective antigen. *J Biol Chem* **285**, 8130-8137 (2010).
22 Feld, G. K., Brown, M. J. & Krantz, B. A. Ratcheting up protein translocation with anthrax toxin. *Protein Sci* **21**, 606-624 (2012).
23 Van Gerven, N. *et al.* Secretion and functional display of fusion proteins through the curli biogenesis pathway. *Mol Microbiol* (2014).
24 Brinker, A. *et al.* Dual function of protein confinement in chaperonin-assisted protein folding. *Cell* **107**, 223-233 (2001).
25 Busby, J. N., Panjikar, S., Landsberg, M. J., Hurst, M. R. & Lott, J. S. The BC component of ABC toxins is an RHS-repeat-containing protein encapsulation device. *Nature* **501**, 547-550 (2013).
26 Takagi, F., Koga, N. & Takada, S. How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: molecular simulations. *Proc Natl Acad Sci U S A* **100**, 11367-11372 (2003).
27 Zhou, H. X. Protein folding in confined and crowded environments. *Arch Biochem Biophys* **469**, 76-82 (2008).
28 Chen, A. Y. *et al.* Synthesis and patterning of tunable multiscale materials with engineered cells. *Nat Mater* **13**, 515-523 (2014).
29 Sivanathan, V. & Hochschild, A. A bacterial export system for generating extracellular amyloid aggregates. *Nat Protoc* **8**, 1381-1390 (2013).

Extended Data is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements: This research was supported by VIB through project grant PRJ9 (PG, NVG, HR), by Hercules Foundation through equipment grant UABR/09/005, by NIH RO1 grants AI099099 and AI048689 (JP, SJH) and A1073847 (MRC), by Institut Pasteur and CNRS (FG, GPA, RF). SH is funded by the EPSRC (Institutional Sponsorship Award), National Physical Laboratory, and UCL Chemistry. FG is the recipient of a "Bourse Roux" from Institut Pasteur. We acknowledge Diamond Light Source for time on I02, I03, I04 and I24 under proposal mx7351 and the Soleil synchrotron for access to Proxima-1 and Proxima-2a under proposals 20100734, 20110924 and 20121253.

Author Contributions: PG produced, purified and crystallized CsgG and $CsgG_{C1S}$, and determined their X-ray structures. Single-particle EM was performed by PVK, FG

and GPH and supervised by RF. PVK, FG and WJ performed the *in vitro* characterization of the CsgG:CsgE complex. NVG and WJ performed mutagenesis and phenotyping experiments. IVDB carried out the single channel recordings, and SH supervised the acquisition and analysis of the recordings. ATT and PG recorded and analyzed FTIR spectra. JSP, MRC and SJH conceived the study and contributed expression constructs and protein. HR conceived and supervised the study, analyzed data and wrote the paper with contributions of all authors.

Author Information:

- Coordinates and structure factors for $CsgG_{C1S}$ and CsgG are deposited in the protein databank under accession codes: 4uv2 and 4uv3, respectively. CryoEM maps for CsgG:CsgE are deposited in the EMDataBank under accession code: EMDB-12772.
- Reprints and permissions information is available at www.nature.com/reprints.
- The authors have no competing interests.
- Correspondence and requests for materials should be addressed to Han Remaut at han.remaut@vib-vub.be.

Figure Legends

Figure 1. X-ray structure of $CsgG_{C1S}$ in pre-pore conformation. (a) Ribbon diagram of the $CsgG_{C1S}$ monomer colored as blue to red rainbow from the N- to C-terminus. Secondary structure elements are labeled according the ABD-like fold, with the additional N- and C-terminal α-helices, and extended loop connecting β1 and α1 labeled αN, αC and CL (C-Loop), respectively. **(b)** Side view of the $CsgG_{C1S}$ C8

octamer with subunits differentiated by color and one subunit oriented and colored as in (a).

Figure 2. Structure of CsgG in its channel conformation. (a) Amide I region (1700 – 1600 cm^{-1}) of ATR-FTIR spectra of $CsgG_{C1S}$ (blue) and membrane-extracted CsgG (red) **(b)** TM1 and TM2 sequence (bilayer facing residues in blue) and Congo Red binding of *E. coli* BW25141Δ*csgG* complemented with wild type *csgG* (WT), empty vector or *csgG* lacking underscored fragments of TM1 or TM2. (**c**) Overlay of CsgG monomer in pre-pore (light blue) and channel conformation (tan). (**d, e**) Side and cross-sectional view of CsgG nonamers in ribbon and surface representation, and Helix 2, the core domain and TM hairpins shown in blue, light blue and tan, respectively. A single protomer is colored according Fig. 1a. Magenta spheres show position of Leu 2. (OM: outer membrane, CL: C-loop).

Figure 3. CsgG channel constriction. (a) Cross-section of CsgG channel constriction and its solvent-excluded diameters. **(b)** The constriction is composed of 3 stacked concentric side chain layers: Y51, N55 and F56, preceded by F48 from the periplasmic side. **(c)** Congo Red binding of *E. coli* BW25141Δ*csgG* complemented with *csgG* (WT), empty vector or *csgG* carrying indicated constrictions mutants. **(c, d)** Representative single-channel current recordings and conductance histogram (d) of CsgG reconstituted in planar phopsholipid bilayers (PPB) and measured under an electrical field of + (*n*=32) or -50 mV (*n*=13).

Figure 4. Model of CsgG transport mechanism. (a) Native PAGE of CsgE (E), CsgG (G) and CsgG supplemented with an excess CsgE (E+G) shows the formation of

a CsgG:CsgE complex (E:G*). **(b)** SDS-PAGE of CsgE (E), CsgG (G) and the E:G* complex recovered from native PAGE. (MW: Molecular Weight markers in kDa). **(c)** Selected class averages of CsgG:CsgE particles. From left to right: top and side view visualized by cryo-EM; and comparison of negatively stained side views with CsgG nonamers. **(d)** CryoEM averages of top and tilted side viewed CsgE particles. Rotational autocorrelation shows 9-fold symmetry. **(e)** 3D reconstruction of CsgG:CsgE (24 Å resolution, 1221 single particles) shows a nonameric particle comprised of CsgG (blue) and an additional density assigned as a CsgE nonamer (orange). **(f)** Single channel current recordings of PPB-reconstituted CsgG at + or -50 mV and supplemented with incremental concentrations of CsgE. (horizontal scale bars lie at 0 pA) **(g)** Tentative model for CsgG-mediated protein secretion. CsgG and CsgE are proposed to form a secretion complex that entraps CsgA (discussed in Extended Data Figure 9), generating an entropic potential over the channel. Upon capture of CsgA in the channel constriction, a ΔS-rectified Brownian diffusion facilitates the progressive translocation of the polypeptide across the OM.

Methods

Cloning and strains

Expression constructs for production of outer membrane localized C-terminally StrepII-tagged CsgG (pPG1) and periplasmic C-terminally StrepII-tagged $CsgG_{C1S}$ (pPG2) have been described in Goyal *et al.* 2013 [19]. For selenomethionine labeling, StrepII-tagged $CsgG_{C1S}$ was expressed in the cytoplasm because of increased yields. Therefore, pPG2 was altered to remove the N-terminal signal peptide using inverse PCR with primers forward: 5'- TCT TTA AC CGC CCC GCC TAA AG -3' and reverse: 5'- CAT TTT TTG CCC TCG TTA TC -3' (pPG3). For phenotypic assays, a *csgG* deletion mutant of *E. coli* BW25141 (*E. coli* NVG2) was constructed by the

method described by Datsenko and Wanner [30] (with primers 5'- AAT AAC TCA ACC GAT TTT TAA GCC CCA GCT TCA TAA GGA AAA TAA TCG TGT AGG CTG GAG CTG CTT C-3' and 5'- CGC TTA AAC AGT AAA ATG CCG GAT GAT AAT TCC GGC TTT TTT ATC TGC ATA TGA ATA TCC TCC TTA G-3'). The various CsgG substitution mutants used for Cys accessibility assays and for phenotypic probing of the channel constriction were constructed by site-directed mutagenesis (QuikChange protocol, Stratagene) starting from pMC2, a pTRC99a vector containing *csgG* under control of the *trc* promotor [14].

Protein expression and purification

CsgG and $CsgG_{C1S}$ were expressed and purified as described [19]. Briefly, CsgG was recombinantly produced in *E. coli* BL 21 (DE3) transformed with pPG1 and extracted from isolated outer membranes using 1 % n-dodecyl-β-D-maltoside (DDM) in buffer A: 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM ethylene diamine tetraacetic acid (EDTA), 1 mM dithiothreitol (DTT). Strep-II tagged CsgG was loaded onto a 5 mL Strep-Tactin sepharose column (Iba GmbH) and detergent-exchanged by washing with 20 column volumes buffer A supplemented with 0.5% tetraethylene glycol monooctyl ether (C8E4, Affymetrix) and 4 mM lauryldimethylamine-N-oxide (LDAO, Affymetrix). The protein was eluted by addition of 2.5 mM D-desthiobiotin and concentrated to 5 mg mL^{-1} for crystallization experiments. For selenomethionine labeling, $CsgG_{C1S}$ was produced in the Met auxotrophic strain B834 (DE3) transformed with pPG3 and grown on M9 minimal medium supplemented with 40 mg L^{-1} L-selenomethionine. Cell pellets were resuspended in 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 5 mM DTT, supplemented with cOmplete Protease Inhibitor Cocktail (Roche) and disrupted by passage through a TS series cell disruptor (Constant Systems Ltd) operated at 20 kPsi. Labeled $CsgG_{C1S}$ was purified as described [19]. 5 mM dithiothreitol (DTT) was added throughout the purification

procedure to avoid selenomethionine oxidation.

CsgE was produced in *E. coli* NEBC2566 cells harboring pNH27 [16]. Cell lysates in 25 mM Tris-HCl pH 8.0, 150 mM NaCl, 25 mM Imidazol, 5 % (v/v) glycerol were loaded on a HisTrap™ FF (GE Healthcare). CsgE-his was eluted with a linear gradient to 500 mM Imidazol in 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 (v/v) % glycerol buffer. Fractions containing CsgE were supplemented with 250 mM $(NH_4)_2SO_4$ and applied on a 5 ml HiTrap™ Phenyl HP column (GE Healthcare) equilibrated with 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 250 mM $(NH_4)_2SO_4$, 5 (v/v) % glycerol. A linear gradient to 20 mM Tris-HCl pH 8.0, 10 mM NaCl, 5 (v/v) % glycerol was applied for elution. CsgE containing fractions were loaded on Superose 6 PREP Grade 10/600 (GE Healthcare) column equilibrated in 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 (v/v) % glycerol.

In solution oligomeric state assessment

Approximately 0.5 mg each of detergent-solubilized CsgG (0.5% C8E4, 4 mM LDAO) and $CsgG_{C1S}$ were applied to a Superdex 200 10/300 GL analytical gel filtration column (GE Healthcare) equilibrated with 25 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM DTT, 4 mM LDAO, and 0.5% C8E4 (CsgG) or 25 mM Tris-HCl pH 8.0, 200 mM NaCl ($CsgG_{C1S}$), and run at 0.7 mL min^{-1}. The column elution volumes were calibrated using bovine thyroglobulin, bovine γ-globulin, chicken ovalbumin, horse myoglobulin and vitamin B12 (Bio-Rad) (Extended Data Fig. 2). Membrane-extraceted CsgG, 20 μg of the detergent-solubilized protein was also run on a 3-10% blue native PAGE using the procedure described by Swamy *et al.* [31] (Extended Data Fig. 2). NativeMark (Life Technologies) unstained protein standard (7 μL) was used

for molecular weight estimation.

Crystallization, data collection and structure determination

Selenomethionine labeled $CsgG_{C1S}$ was concentrated to 3.8 mg mL^{-1} and crystallized by sitting drop vapor diffusion against a solution containing 100 mM sodium acetate pH 4.2, 8% PEG 4000 and 100 mM sodium malonate pH 7.0. Crystals were incubated in crystallization buffer supplemented with 15 % glycerol and flash-frozen in liquid nitrogen. Detergent-solubilized CsgG was concentrated to 5 mg mL^{-1} and crystallized by hanging drop vapor diffusion against a solution containing 100 mM Tris-HCl pH 8.0, 8% PEG 4000, 100 mM NaCl and 500 mM $MgCl_2$. Crystals were flash-frozen in liquid nitrogen and cryoprotected by the detergent present in the crystallization solution. For optimization of crystal conditions and screening for crystals with good diffraction quality, crystals were analyzed on beamlines Proxima-1, Proxima-2a (Soleil, France), PX-I (Swiss Light Source, Swiss), I02, I03, I04, I24 (Diamond Light Source, UK) and ID14eh2, ID23eh1, ID23eh2 (ESRF, France). Final diffraction data used for structure determination of $CsgG_{C1S}$ and CsgG were collected at beamlines I04 and I03, respectively (Diamond Light Source, UK) (see Extended Data Table 1 for data collection and refinement statistics). Diffraction data for $CsgG_{C1S}$ were processed using the Xia2 and the XDS package[32,33]. Crystals of $CsgG_{C1S}$ belonged to space group P1 with unit cell dimensions of a = 101.3 Å, b = 103.6 Å, c = 141.7 Å, α = 111.3°, β = 90.5°, γ = 118.2°, containing 16 protein copies in the asymmetric unit (AU). For structure determination and refinement, date collected at 0.9795 Å wavelength were truncated at 2.8 Å based on an I/σI cutoff of 2 in the highest resolution shell. The structure was solved using experimental phases calculated from a Single Anomalous Dispersion (SAD) experiment. 92 selenium sites were located in the AU using ShelxC and ShelxD[34], and refined and used for phase calculation with Sharp[35] (phasing power:

0.79, figure of merit (FOM): 0.25). Experimental phases were density modified and non-crystallographic symmetry (NCS) -averaged using Parrot[36] (Extended Data Fig. 10; FOM: 0.85). An initial model was built with Buccaneer[37] and refined by iterative rounds of maximum likelihood refinement with Phenix refine[38] and manual inspection and model (re)building in Coot[39]. The final structure contains 28853 atoms in 3700 residues belonging to 16 $CsgG_{C1S}$ chains (Extended Data Figs. 2), with a molprobity[40] score of 1.34, and 98 % of the residues lying in favored regions of the Ramachandran plot (99.7 % in allowed regions). Electron density maps showed no unambiguous density corresponding to possible solvent molecules and no water molecules or ions were, therefore, built in. 16-fold NCS averaging was maintained throughout refinement, using strict and local NCS restraints in early and late stages of refinement, respectively.

Diffraction data for CsgG were collected from a single crystal at 0.9763 Å wavelength and were indexed and scaled, using the XDS package[32,33], in space group C2 with unit cell dimensions a = 161.7 Å, b = 372.3 Å, c = 161.8 Å and β = 92.9°, encompassing 18 CsgG copies in the AU and a 72 % solvent content. Diffraction data for structure determination and refinement were elliptically truncated to resolution limits of 3.6 Å, 3.7 Å and 3.8 Å along reciprocal cell directions a*, b* and c* and anisotropically scaled using the Diffraction Anisotropy Server[41]. Molecular replacement using the $CsgG_{C1S}$ monomer proved unsuccessful. Analysis of the self rotation function revealed D_9 symmetry in the asymmetric unit (not shown). Based on the $CsgG_{C1S}$ structure, a nonameric search model was generated in the assumption that upon going from a C_8 to C_9 oligomer, the interprotomer arc at the particle circumference would stay approximately the same as the interprotomer angle changes form 45° to 40°, giving a calculated increase in radius of approximately 4 Å. Using the calculated nonamer as search model, a molecular replacement solution containing to copies was found with

Phaser[42]. Inspection of density-modified and NCS-averaged electron density maps (Parrot[36]; Extended Data Fig. 10) allowed manual building of the TM1 and TM2 and remodeling of adjacent residues in the protein core, as well as the building of residues 2 to 18, which were missing from the $CsgG_{C1S}$ model and link the α1 helix to the N-terminal lipid anchor. Refinement of the CsgG model was performed with Buster-TNT[43] and Refmac5[44] for initial and final refinement rounds, respectively. 18-fold local NCS restraints were applied throughout refinement and Refmac5 was run employing a jelly-body refinement with sigma 0.01 and H-bond restraints generated by Prosmart[45]. The final structure contains 34165 atoms in 4451 residues belonging to 18 CsgG chains (Extended Data Fig. 2), with a molprobity score of 2.79 and 93.0 % of the residues lying in favored regions of the Ramachandran plot (99.3 % in allowed regions). No unambiguous electron density corresponding the N-terminal lipid anchor could be discerned.

Congo Red assay

For analysis of Congo Red binding, a bacterial overnight culture grown at 37°C in LB medium was diluted in Lysogeny Broth (LB) till an OD_{600nm} of 0.5. Five µl was spotted on LB agar plates supplemented with ampicillin (100 mg L^{-1}), Congo Red (100 mg l^{-1}) and 0.1% (w/v) isopropyl β-D-1-thiogalactopyranoside (IPTG). Plates were incubated at room temperature for 48 hours to induce curli expression. The development of the colony morphology and dye binding was observed at 48 hours.

Cysteine accessibility assays

Cysteine mutants were generated in pMC2 using site directed mutagenesis and expressed in *E. coli* LSR12 [7]. Bacterial cultures grown overnight were spotted onto LB

agar plates containing 1 mM IPTG and 100 mg L^{-1} ampicillin. Plates were incubated at room temperature and cells were scrapped after 48 hours, resuspended in 1 mL Phosphate Buffered Saline (PBS) and normalized using OD_{600nm}. The cells were lysed by sonication and centrifuged for 20 sec at 3000 RCF at 4°C to remove unbroken cells from cell lysate and suspended membranes. Proteins in the supernatant were labeled using 15 mM methoxypolyethylene glycol-maleimide (MAL-PEG 5 kDa) for one hour at room temperature. The reaction was stopped using 100 mM DTT and centrifuged at 40, 000 rpm in 50.4 Ti rotor for 20 min at 4°C to pellet total membranes. The pellet was washed with 1% sodium lauroyl sarcosinate to solubilize cytoplasmic membranes and again centrifuged. The resulting outer membranes were resuspended and solubilized using PBS containing 1% DDM. Metal affinity pull downs done using nickel beads were used for SDS-PAGE and anti his western blots. *E. coli* LSR12 cells with empty pMC2 vector were used as negative control.

Attenuated Total Reflection Fourier Transform Infrared (ATR-FTIR) spectroscopy

ATR-FTIR measurements were performed on an Equinox 55 infrared spectrophotometer (Bruker), continuously purged with dried air, equipped with a liquid nitrogen-refrigerated Mercury Cadmium Telluride detector and a Golden Gate reflectance accessory (Specac). The internal reflection element was a diamond crystal (2 mm x 2 mm) and the beam incidence angle was of 45°. 1 µL of each purified protein sample was spread at the surface of the crystal and dried under a gaseous nitrogen flow to form a film. Each spectrum, recorded at a 2 cm^{-1} resolution, was an average of 128 accumulations for improved signal-to-noise ratio. All the spectra were treated with water vapor contribution subtraction, smoothed at a final resolution of 4

cm^{-1} by apodization and normalized on the area of Amide I band (1700 – 1600 cm^{-1}) in order to allow their comparison[46].

Negative stain EM and symmetry determination

Negative stain EM was used to monitor *in solution* oligomerization states of CsgG, $CsgG_{C1S}$ and CsgE. CsgE, $CsgG_{C1S}$ and amphipol-bound CsgG were adjusted to a concentration of 0.05 mg mL^{-1} and applied to glow-discharged carbon-coated copper grids (CF-400, Electron Microscopy Sciences). After 1 minute incubation, samples were blotted, then washed and stained in 2 % uranyl acetate. Images were collected on a Tecnai T12 BioTWIN LaB6 microscope operating at a voltage of 120 kV, at magnification of 49,000 and defocus between 800 - 2000 nm. CTF, phase flipping and particle selection was performed as described for cryo-EM. For membrane-extracted CsgG, octadecameric particles (1780 total) were analyzed separately from nonamers and top views. For purified CsgE a total of 2452 particles were analyzed. 3D models were obtained as described for the CsgG:CsgE cryo-EM analysis below and refined by several rounds of MRA, MSA and anchor set refinement. In all cases, after normalization and centering, images were classified using IMAGIC-4D as described in the cryo-EM section The best classes corresponding to characteristic views were selected for each set of particles. Symmetry determination of CsgG top views was performed using the best class averages with approximately 20 images per class. The rotational auto-correlation function was calculated using IMAGIC and plotted.

CsgG:CsgE complex formation

For CsgG:CsgE complex formation, the solubilizing detergents in purified CsgG were exchanged for Amphipols A8-35 (Anatrace) by adding 120 μL CsgG (24 mG mL^{-1} protein in 0.5% C8E4, 4 mM LDAO, 25 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM

DTT) to 300 μL detergent-destabilized liposomes (1mg mL^{-1} DMPC and 0.4% LDAO) and incubating for 5 minutes on ice before the addition of 90 μL of A8-35 amphipols at 100 mg mL^{-1}stock. After additional 15 min incubation on ice, the sample was loaded on a Superose 6 10/300 GL (GE Healthcare) column and gel filtration was performed in 200 mM NaCl, 2.5% xylitol, 25 mM Tris pH 8, and 0.2 mM DTT. Equal volumes of purified monomeric CsgE in 200mM NaCl, 2.5% xylitol, 25 mM Tris-HCl pH 8, and 0.2mM DTT was added to the amphipol-solubilized CsgG at final protein concentrations of 15 and 5 μM for CsgE and CsgG, respectively, and the sample was run at 125 V at 18 °C on a 4.5 % native PAGE in 0.5x TBE (Tris/Borate/EDTA) buffer. For the second, denaturing dimension, the band corresponding to the CsgG:CsgE complex was cut out of unstained lanes ran in parallel on the same gel, boiled for 5 minutes in Laemmli buffer (60 mM Tris-HCl pH 6.8, 2% SDS, 10% glycerol, 5% β-mercaptoethanol, 0.01% bromophenol blue) and run in a 4-20% SDS-PAGE. Purified CsgE and CsgG were run alongside the complex as control samples. Gels were stained with InstantBlue™ coomassie for visual inspection or SYPRO orange for stoichiometry assessment of the CsgG:CsgE complex by fluorescence detection (Typhoon FLA 9000) of the CsgE and CsgG bands on SDS-PAGE, yielding a CsgG/CsgE ratio of 0.97.

CsgG:E CryoEM

Cryo electron microscopy was used to determine the *in solution* structure of the C_9 CsgG:CsgE complex. CsgG:CsgE complex prepared as described above was bound and eluted from a HisTrap™ FF (GE Healthcare) to remove unbound CsgG and upon elution immediately applied on a Quantifoil R2/2 carbon coated grids (Quantifoil Micro Tools GmbH) that had been glow discharged at 20mA for 30 seconds. Samples

were plunge-frozen in liquid nitrogen using an automated system (Leica) and visualized on a FEI F20 microscope operating at a voltage of 200 kV, a nominal magnification of 50,000 under low dose conditions and a defocus range of 1.4-3 μm. Image frames were recorded on a Falcon II detector. The pixel size at the specimen level was 1.9 Å per pixel. The CTF (contrast transfer function) parameters were assessed using CTFFIND3 [47], and the phase flipping was done in SPIDER [48]. Particles were automatically selected from CTF-corrected micrographs using BOXER (EMAN2 [49]). Images with an astigmatism of more than 10% were discarded. 4881 particle images were selected from 145 frames and windowed into 128x128 pixel boxes. Images were normalized to the same mean and standard deviation and high-pass filtered at a low-resolution cut-off of ~200 Å. They were centered, and subjected to a first round of multi-statistical analysis (MSA). An initial reference set was obtained using reference free classification in IMAGIC-4D (Image Science Software, GmbH). The best classes corresponding to characteristic side views of the C_9 cylindrical particles were used as references for the multireference alignment (MRA). For CsgG:E complex, the first 3D model was calculated from the best 125 characteristic views (with good contrast and well defined features) encompassing 1221 particles of the complex with orientations determined by angular reconstitution (Image Science Software, GmbH). The 3D map was refined by iterative rounds of MRA, MSA and anchor set refinement. The resolution was estimated by Fourier Shell Correlation (FSC) accordingly to the 0.5 criteria level as 24 Å (Extended Data Fig. 7). Visualization of the map and figures were done in UCSF Chimera [50].

Bile salt toxicity assay

Outer membrane permeability was investigated by reduced growth on agar plates containing bile salts. Five μl of tenfold serial dilutions of *E. coli* LSR12 [7] cells harbouring both pLR42 [16] and pMC2 [14] (or derived Helix 2 mutants) were spotted on McConkey agar plates containing 100 μg L^{-1} ampicillin, 25 μg L^{-1} chloramphenicol, 1 mM IPTG with or without 0.2% (w/v) L-arabinose. After overnight incubation at 37°C colony growth was examined.

Single-channel current recordings

Single-channel current recordings were performed using parallel high-resolution electrical recording with the Orbit 16 kit from Nanion (Munich, Germany). Briefly, horizontal bilayers of 1,2-diphytanoyl-sn-glycero-3-phosphocholine (Avanti Polar Lipids) were formed over microcavities (of sub-picoliter volume) in a 16-channel multielectrode cavity array (MECA) chip (Ionera, Freiburg, Germany) [51]. Both the *cis* and *trans* cavities above and below the bilayer contained 1.0 M KCl, 25 mM Tris-HCl, pH 8.0. To insert channels into the membrane, CsgG dissolved in 25 mM Tris pH 8.0, 500 mM NaCl, 1 mM DTT, 0.5% C8E4, 5 mM LDAO was added to the *cis* compartment to a final concentration of 90 – 300 nM. To test the interaction of the CsgG channel with CsgE, a solution of the latter protein dissolved in 25 mM Tris pH 8.0 and 150 mM NaCl was added to the *cis* compartment to a final concentration of 0.1, 1, 10 and 100 nM. Transmembrane currents were recorded at a holding potential of +50 mV and -50 mV (with the *cis* side grounded) using a Tecella Triton 16 channel amplifier at a low-pass filtering frequency of 3 kHz and a sampling frequency of 10 kHz. Current traces were analyzed using the Clampfit of the pClamp suite (Molecular Devices, USA). Plots were generated using Origin 8.6 (Microcal, USA) [52].

Measured currents were compared with those calculated based on the pore dimensions of the CsgG X-ray structure, modeled to be composed of three segments: (1) the transmembrane section, (2) the periplasmic vestibule, and (3) the inner channel constriction connecting the two. The corresponding resistances were calculated as:

$R_1 = L_1/(\pi D_1 d_1 \kappa)$

$R_2 = L_2 /(\pi D_2 d_2 \kappa)$

$R_3 = L_3 /(\pi d_1 d_2 \kappa)$

where L_1, L_2, and L_3 are the axial lengths of the segments, measuring 3.5 nm, 4.0 nm, and 2.0 nm, respectively, and D_1, d_1, D_2 and d_2 are the maximum and minimum diameters of segments 1 and 2, measuring 4.0 nm, 0.8 nm, 3.4 nm, and 0.8 nm, respectively. The conductivity κ has a value of 10.6 S m^{-1}. The current was calculated by inserting R_1, R_2 and R_3 and voltage V = 50 mV into:

$I = V/(R_1+R_2+R_3)$

Access resistance was not found to significantly alter the predicted current.

Extended Data:

Extended Data Figure 1 | Curli biosynthesis pathway in *E. coli*. The major curli subunit CsgA (light green) is secreted from the cell as a soluble monomeric protein. The minor curli subunit CsgB (dark green) is associated with the outer membrane (OM) and acts as a nucleator for the conversion of CsgA from a soluble protein to amyloid deposit. CsgG (orange) assembles into an oligomeric curli-specific translocation channel in the OM. CsgE (purple) and CsgF (light blue) form soluble accessory proteins required for productive CsgA and CsgB transport and deposition. CsgC forms a putative oxido-reductase of unknown function. All curli proteins have putative Sec signal sequences for transport across the cytoplasmic (inner) membrane (IM).

Extended Data Figure 2 | *In solution* oligomerization states of CsgG and $CsgG_{C1S}$ analyzed by size exclusion chromatography and negative-stain electron microscopy. (a) Raw negative-stain EM image of C8E4/LDAO-solubilized CsgG. Arrows indicate the different particle populations as labeled in the size exclusion profile shown in panel (g), being (a) aggregates of CsgG nonamers (b) CsgG octadecamers and (c) CsgG nonamers. Scale bar: 20 nm. **(b)** Representative class-average for top- and side-views of the indicated oligomeric states. **(c)** Rotational auto-correlation function graph of LDAO-solubilized CsgG in top view, showing 9-fold symmetry. **(d)** Raw negative-stain EM image of $CsgG_{C1S}$. Arrows indicate the hexadecameric (d) and octameric (e) particles observed by size exclusion chromatography in panel (g). **(e)** Representative class-average for side-views of $CsgG_{C1S}$ oligomers. No top views were observed for this construct. **(f)** Table of elution volumes (EV) of $CsgG_{C1S}$ and CsgG particles observed by size exclusion chromatography shown in panel (g), calculated MW (MW_{calc}), expected MW (MW_{CsgG}) according CsgG oligomerization state ($CsgG_n$) and the particles symmetry

as observed by negative-stain EM and X-ray crystallography. **(g)** Size exclusion chromatogram of $CsgG_{C1S}$ (black) and C8E4/LDAO-solubilized CsgG (grey) ran on Superdex 200 10/300 GL (GE Healthcare). **(h, i)** Ribbon representation of crystallized oligomers in top and side view show the D_8 hexadecamers for $CsgG_{C1S}$ (h) and D_9 octadecamers for membrane-extracted CsgG (i). One protomer is colored in rainbow from N-terminus (blue) to C-terminus (red). The two C_8 octamers ($CsgG_{C1S}$) or C_9 nonamers (CsgG) that form the tail-to-tail dimers captured in the crystals are colored blue and tan. 'r' and 'θ' give radius and interprotomer rotation.

Extended Data Figure 3 | Comparison of CsgG with structural homologues and interprotomer contacts in CsgG. (a) Ribbon diagram for the $CsgG_{C1S}$ monomer (eg. CsgG in pre-pore conformation) and **(b)** the nucleotide binding domain-like domain of TolB (PDB entry: 2hqs), respectively, colored both in rainbow from N-terminus (blue) to C-terminus (red). Common secondary structure elements are labeled equivalently. **(c)** Panels show $CsgG_{C1S}$ (grey) in superimposition with, from left to right, *Xanthomonas campestris* rare lipoprotein B (PDB entry: 2r76, colored pink), *Shewanella oneidensis* hypothetical lipoprotein – DUF330 (PDB entry: 2iqi, colored pink) and *Escherichia coli* TolB (PDB entry 2hqs, colored pink and yellow for the NTD and beta-propeller domain, resp.). CsgG-specific structural elements are labeled and colored as in upper left panel. **(d, e)** Ribbon diagram of two adjacent protomers as found in the CsgG structure, viewed along the plane of the bilayer, either from out- (c) or inside (d) the oligomer. One protomer is shown in rainbow (dark blue to red) from N- to C-terminus, a second protomer is shown in light blue (core domain), blue (helix 2 and tan (TM domain). Four main oligomerization interfaces are apparent: (1) β6-β3' main chain interactions inside the β-barrel, (2) the constriction loop (CL), (3)

side chain packing of helix 1 (α1) against β1-β3-β4-β5 and (4) helix-helix packing of helix 2 (α2). The 18-residue N-terminal loop connecting the lipid anchor (a magenta sphere shows Cα position of Leu 2) and N-terminal helix (αN) is also seen to wrap over the adjacent two protomer. The projected position of the lipid anchor is expected to lie against the TM1 and TM2 hairpins of the +2 protomer (not shown for clarity).

Extended Data Figure 4 | Cys accessibility assays for selected surface residues in the CsgG oligomers. Ribbon representation of CsgG nonamers shown in periplasmic **(a)**, side **(b)** or extracellular view **(c)**. One protomer is colored in rainbow from N-terminus (blue) to C-terminus (red). Cysteine substitutions are labeled and the equivalent locations of the S atoms are shown as spheres, colored according accessibility to maileimide-polyethyleneglycol (MAL-PEG, 5000 Da) labeling in *E. coli* outer membranes **(d)**. Western blot of MAL-PEG reacted samples analyzed on SDS-PAGE show 5 kDa increase upon MAL-PEG binding of the introduced cysteine. Accessible (++ and +++), medium accessible (+) and inaccessible (-) sites are colored green, orange or red in the panels a, b, c and e. For R97 and R110 a second species at 44 kDa is present, corresponding to a fraction of protein where both the introduced and native cysteine got labeled. **(e)** Side view of the dimerization interface in the D_9 octadecamer as present in the X-ray structure. Introduced cysteines in the dimerization interface or inside the lumen of the D_9 particle are labeled. In membrane-bound CsgG, these residues are accessibility to MAL-PEG, demonstrating that the D_9 particles are an artifact of concentrated solutions of membrane-extracted CsgG and that the C_9 complex forms the physiologically relevant species. Residues in the C-terminal helix (αC; K242, D248 and H255) are found to be non- to poorly accessible, indicating αC

may form additional contacts with the *E. coli* cell envelope, possibly the peptidoglycan layer.

Extended Data Figure 5 | Molecular dynamics simulation of CsgG constriction with model poly-alanine chain. Top **(a)** and side **(b)** view of the CsgG constriction modeled with a poly-alanine chain threaded through the channel in an extended conformation, here shown in C- to N-terminal direction. Substrate passage through the CsgG transporter is itself non sequence-specific[16,23]. For clarity, a poly-alanine chain was used for modeling the putative interactions of a passing polypeptide chain. The modeled area is composed of 9 concentric CsgG C-loops, each comprising residues 47 to 58. Side chains lining the constriction are shown in stick representation, with F51 colored slate blue, N55 (amide-clamp) colored cyan, F48 and F56 (ϕ-clamp) in light and dark orange, resp. N, O and H atoms (only hydroxyl or side chain amide H's are shown) are colored blue, red and white, resp. The poly-alanine chain is colored green, blue, red and white for C, N, O and H atoms respectively. Solvent molecules (water) within 10 Å from the poly-alanine residues inside the constriction (residues labeled +1 to +5) are shown as red dots. Panel **(c)** shows modeled solvation of the poly-alanine chain, position as in panel (b) and with C-loops removed for clarity (shown solvent molecules are those within 10 Å of the full poly-alanine chain). The height of the amide- and ϕ-clamps, the solvation of the poly-alanine chain is reduced to a single water shell that bridges the peptide backbone and amide-clamp side chains. Most side chains in the Y51 ring have rotated towards the solvent as compared to their inward, center pointing position observed in the CsgG (as well as $CsgG_{C1S}$) X-ray structure. The model is the result off a 40 ns all atom explicit solvent molecular dynamics

simulation with GROMACS[53] using the AMBER99SB-ILDN[54] force field and with the Cα atoms of the residues at the extremity of the C-loop (Q47 and T58) positionally restricted.

Extended Data Figure 6 | Sequence conservation in CsgG homologues. (a) Surface representation of the CsgG nonamer colored according to sequence similarity (colored yellow to blue from low to high conservation score)[55] and viewed from the periplasm (far left), the side (middle left), the extracellular (middle right) or as a cross-sectional side view (far right). The figures show that the regions of highest sequence conservation map to the entry of the periplasmic vestibule, the vestibular side of the constriction loop and the luminal surface of the TM domain. **(b)** Multiple sequence alignment (MST) of CsgG-like lipoproteins. The selected sequences were chosen from monophyletic clades across the phylogenetic three of CsgG-like sequences (not shown) in order to give a representative view of sequence diversity. Secondary structure elements are shown as arrows or bars for β-strands and α-helices, respectively, and are based on the *E. coli* CsgG crystals structure. **(c, d)** CsgG protomer in secondary structure representation (c) and a cross-sectional side view (d) of the CsgG nonamer in surface representation, both colored grey and with three continuous blocks of high sequence conservation colored red (HCR1), blue (HCR2) and yellow (HCR3). HCR1 and HCR2 shape the vestibular side of the constriction loop, HCR3 corresponds to helix 2, lying at the entry of the periplasmic vestibule. Inside the constriction, F_{56} is 100% conserved, whilst N_{55} can be conservatively substituted by Ser or Thr, eg. by a small polar side chain that can act as H-bond donor / acceptor. The concentric side chain ring at the exit of the constriction (Y_{51}) is not

conserved. Strikingly, the presence of the Phe-ring at the entrance of the constriction is topologically similar the Phe_{427}-ring (referred to as ϕ-clamp) in the Anthrax protective antigen PA_{63}, where it was shown to catalyzes polypeptide capture and passage[20]. MST of *toxB* superfamily proteins reveals a conserved motif D(D/Q)**F**(S/N)S at the height of the Phe-ring. This is similar to the S(Q/N/T)**F**ST motif seen in curli-like transporters. Although an atomic resolution structure of PA_{63} in pore conformation is not yet available, available structures suggest the Phe-ring may similarly be followed by a conserved H-bond donor / acceptor (S/N_{428}) as a subsequent concentric ring in the translocation channel (note that the orientation of element is inversed in both transporters).

Extended Data Figure 7 | CsgE oligomer and CsgG:CsgE complex. (a) Size exclusion chromatography of CsgE (Superose 6, 16/600; running buffer 20 mM Tris-HCl pH8, 100 mM NaCl, 2.5% glycerol) shows an equilibrium of two oligomeric states, (1) and (2), with an apparent MW ratio of 9.16. Negative stain EM inspection of peak (1) shows discrete CsgE particles (5 representative class averages show inset, ordered by increasing tilt angles) compatible in size with 9 CsgE copies **(b)** Selected class average of CsgE oligomer observed in top view by cryo-EM and its rotational autocorrelation show the presence of C_9 symmetry. **(c)** Fourier Shell Correlation (FSC) analysis of CsgG:CsgE cryo-EM model. 3D reconstruction achieved a resolution of 24 Å as determined by Fourier shell correlation at a threshold of 0.5 correlation using 125 classes corresponding to 1221 particles. **(d)** Overlay of CsgG:CsgE cryo-EM density and the CsgG nonamer observed in the X-ray structure. The overlays are shown viewed from the side as semi-transparent density (left) or as a cross-sectional view. **(e)** Congo Red binding of *E. coli* BW25141D*csgG* complemented with wild type *csgG*

(WT), empty vector (*DcsgG*) or *csgG* helix 2 mutants (single amino acid replacements labeled in single letter code). **(f)** Effect of bile salt toxicity on *E. coli* LSR12 complemented with *csgG* (WT), or *csgG* carrying different helix 2 mutations, complemented with (+) or without (-) *csgE*. Tenfold serial dilution starting from 10^7 bacteria were spotted on McConkey agar plates. Expression of the CsgG pore in the OM leads to an increased bile salt sensitivity that can be blocked by co-expression of CsgE . **(g)** Cross-sectional view of CsgG X-ray structure in molecular surface representation. CsgG mutants without an effect on CR binding or toxicity are shown in blue, mutants that interfere with CsgE-mediated rescue of bile salt sensitivity are indicated in red.

Extended Data Figure 8 | Single-channel current analysis of CsgG and CsgG + CsgE pores. (a) Under negative field potential, CsgG pores show two conductance states. Upper left and right panels show a representative single-channel current trace of, respectively, the regular (measured at +50 mV, 0 mV and -50 mV) and low conductance forms (measured at 0 mV, + 50 mV, and - 50 mV. No conversions between both states were observed during the total observation time (n=22), indicating the conductance states have long life times (second to minute timescale)**.** Lower left panel shows current histogram of regular and low conductance CsgG pores acquired at + 50 mV and - 50 mV (*n*=32). IV curves for CsgG pores with regular and low conductance are shown in the lower right panel. The data represent the averages and standard deviations from at least 4 independent recordings. The nature or physiological existence of the low conductance form are unknown. **(b)** Electrophysiology of CsgG channels titrated with the accessory factor CsgE. The plots display the fraction of open, intermediate, and closed channel as a function of CsgE

concentration. Open and closed states of CsgG are illustrated in Fig. 4f of the main manuscript. Increasing the concentration of CsgE to over 10 nM leads to the closure of CsgG pores. The effect occurs at both + 50 mV (left) and – 50 mV (right) ruling out the possibility that the pore blockade can be explained by electrophoresis of CsgE (calculated pI 4.7) into the CsgG pore. An infrequent (<5%) intermediate state has approximately half the conductance of the open channel. It may represent CsgE-induced incomplete closures of the CsgG channel, or alternatively, it could represent the temporary formation of a CsgG dimer caused by the binding of residual CsgG monomer from the electrolyte solution to the membrane-embedded pore. The fraction for the three states was obtained from all-point histogram analysis of single-channel current traces. The histograms yielded peak areas for up to three states, and the fraction for a given state was obtained by dividing the corresponding peak area by the sum of all other states in the recording. Under negative field potential two open conductance states are discerned, similar to the observations for CsgG (see (a)). As both open channel variations were blocked by higher CsgE concentrations, the "open" traces in panel b combine both conductance forms. The data in the plot represent the averages and standard deviations from three independent recordings and channels. **(c)** The crystal structure, size exclusion chromatography and EM show that detergent extracted CsgG pores forms non-native tail-to-tail stacked dimers (eg. two 9-mers as D_9 particle, Extended Data Fig. 2) at higher protein concentration. These dimers can also be observed in single channel recordings. Upper panel shows the single-channel current trace of a stacked CsgG pore at + 50 mV, 0 mV, and - 50 mV (left to right). Lower left panel shows current histogram of dimeric CsgG pores recorded at + 50 mV and - 50 mV. The experimental conductances of +16 ±2 and –16 ±3 pA (n=29) at + and – 50mV, respectively, are near the theoretically calculated value of 23 pA. Lower right panels shows IV curves for stacked CsgG pores. The data represent the averages

and standard deviations from six independent recordings. **(d)** The possibility of CsgE to bind and block stacked CsgG pores was tested by electrophysiology. Single-channel current traces of stacked CsgG pore in the presence of 10 nM or 100 nM CsgE at + 50 mV (upper), and – 50 mV (lower). The current traces show that otherwise saturating concentrations of CsgE do not lead to pore closure in case of stacked CsgG dimers. These observations are in good agreement with the mapping of the CsgG:CsgE contact zone to helix 2 and the mouth of the CsgG periplasmic cavity as discerned by EM and site-directed mutagenesis (Fig. 4 and Extended Data Fig. 7).

Extended Data Figure 9 | Assembly and substrate recruitment of the CsgG secretion complex. The curli transporter CsgG and the soluble secretion co-factor CsgE form a secretion complex with 9:9 stoichiometry that encloses a ~24.000 $Å^3$ chamber that is proposed to entrap the CsgA substrate and facilitate its entropy-driven diffusion across the outer membrane (OM; Main text and Fig. 4). On theoretical grounds, three putative pathways for substrate recruitment and assembly of the secretion complex can be envisaged: **(a)** A **catch-and-cap** mechanism that entails the binding of CsgA to the apo CsgG translocation channel (1), leading to a conformational change in the latter that exposes a high affinity binding platform for CsgE binding (2). CsgE binding leads to capping of the substrate cage. Upon secretion of CsgA, CsgG would fall back into its low affinity conformation, leading to CsgE dissociation and liberation of the secretion channel for a new secretion cycle. **(b)** A **dock-and-trap** mechanism where periplasmic CsgA is first captured by CsgE (1), causing the latter to adopt a high affinity complex that docks onto the CsgG translocation pore (2), enclosing CsgA into the secretion complex. CsgA binding could be directly to CsgE oligomers or CsgE monomers, the latter leading to

subsequent oligomerization and CsgG binding. Secretion of CsgA leads CsgE to fall back into its low affinity conformation and to dissociate from the secretion channel. **(c)** In a third model, CsgG and CsgE form a constitutive complex, where CsgE conformational dynamics cycle between an open and closed form in function of CsgA binding.

Currently published or available data do not allow us to discriminate or put forward one of the putative recruitment modes or derivatives thereof.

Extended Data Figure 10 | Data collection statistics and electron density maps of $CsgG_{C1S}$ and CsgG. (a) Data collection statistics for $CsgG_{C1S}$ and CsgG X-ray structures **(b)** 2.8 Å electron density map for $CsgG_{C1S}$ calculated using NCS-averaged and density-modified experimental SAD phases, and contoured at 1.5 σ. The map shows the region of the channel construction (CL; a single protomor is labeled) and is overlaid on top of the final refined model. **(c)** Electron density map (resolution: 3.6 Å, 3.7 Å and 3.8 Å along reciprocal vectors a*, b* and c*, resp.) in the CsgG TM domain region, calculated from NCS-averaged and density modified molecular replacement phases (TM loops were absent from the input model), B-factor sharpened by -20 $Å^2$ and contoured at 1.0 σ. The shows the TM1 (K135-L154) and TM2 (L182-N209) region of a single CsgG protomer, overlaid on the final refined model.

References:

30 Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640-6645 (2000).

31 Swamy, M., Siegers, G. M., Minguet, S., Wollscheid, B. & Schamel, W. W. A. Blue native polyacrylamide gel electrophoresis (BN-PAGE) for the identification and analysis of multiprotein complexes. *Science's STKE : signal transduction knowledge environment* **2006**, pl4 (2006).

32 Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* **43**, 186-190 (2010).

33 Kabsch, W. Xds. *Acta Crystallogr D* **66**, 125-132 (2010).

34 Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta crystallographica. Section D, Biological crystallography* **66**, 479-485 (2010).

35 Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta crystallographica. Section D, Biological crystallography* **59**, 2023-2030 (2003).

36 Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr D Biol Crystallogr* **66**, 470-478 (2010).

37 Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta crystallographica. Section D, Biological crystallography* **62**, 1002-1011 (2006).

38 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* **66**, 213-221 (2010).

39 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* **66**, 486-501 (2010).

40 Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic acids research* **35**, W375-383 (2007).

41 Strong, M. *et al.* Toward the structural genomics of complexes: Crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *P Natl Acad Sci USA* **103**, 8060-8065 (2006).

42 McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658-674 (2007).

43 Smart, O. S. *et al.* Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta crystallographica. Section D, Biological crystallography* **68**, 368-380 (2012).

44 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological crystallography* **67**, 355-367 (2011).

45 Nicholls, R. A., Long, F. & Murshudov, G. N. Low-resolution refinement tools in REFMAC5. *Acta crystallographica. Section D, Biological crystallography* **68**, 404-417 (2012).

46 Goormaghtigh, E. & Ruysschaert, J. M. Subtraction of atmospheric water contribution in Fourier transform infrared spectroscopy of biological membranes and proteins. *Spectrochimica Acta* **50A**, 2137-2144 (1994).

47 Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* **142**, 334-347 (2003).
48 Shaikh, T. R. *et al.* SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nat Protoc* **3**, 1941-1974 (2008).
49 Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* **157**, 38-46 (2007).
50 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).
51 Baaken, G., Ankri, N., Schuler, A. K., Ruhe, J. & Behrends, J. C. Nanopore-based single-molecule mass spectrometry on a lipid membrane microarray. *ACS Nano* **5**, 8080-8088 (2011).
52 Movileanu, L., Howorka, S., Braha, O. & Bayley, H. Detecting protein analytes that modulate transmembrane movement of a polymer chain within a single protein pore. *Nat Biotechnol* **18**, 1091-1095 (2000).
53 Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845-854 (2013).
54 Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950-1958 (2010).
55 Capra, J. A. & Singh, M. Predicting functionally important residues from sequence conservation. *Bioinformatics* **23**, 1875-1882 (2007).