nature biotechnology

ARTICLES

https://doi.org/10.1038/s41587-020-0570-8

A dual-constriction biological nanopore resolves homonucleotide sequences with high fidelity

Sander E. Van der Verren[1,2], Nani Van Gerven[1,2], Wim Jonckheere[1,2], Richard Hambley[3], Pratik Singh[3], John Kilgour[3], Michael Jordan[3], E. Jayne Wallace[3], Lakmal Jayasinghe[3] and Han Remaut[1,2] ✉

Single-molecule long-read DNA sequencing with biological nanopores is fast and high-throughput but suffers reduced accuracy in homonucleotide stretches. We now combine the CsgG nanopore with the 35-residue N-terminal region of its extracellular interaction partner CsgF to produce a dual-constriction pore with improved signal and base-calling accuracy for homopolymer regions. The electron cryo-microscopy structure of CsgG in complex with full-length CsgF shows that the 33 N-terminal residues of CsgF bind inside the β-barrel of the pore, forming a defined second constriction. In complexes of CsgG bound to a 35-residue CsgF constriction peptide, the second constriction is separated from the primary constriction by ~25 Å. We find that both constrictions contribute to electrical signal modulation during single-stranded DNA translocation. DNA sequencing using a prototype CsgG–CsgF protein pore with two constrictions improved single-read accuracy by 25 to 70% in homopolymers up to 9 nucleotides long.

In nanopore sensing applications, the interaction of analyte molecules with a nanometer-scale pore that is inserted in an insulating membrane within an electrical field results in altered ion conductance levels. Together with the frequency and lifetime of the interaction, the modulated conductance levels can give information on the size, concentration and chemical and conformational nature of the analyte[1]. A practical application of this principle is nanopore sequencing, where a single-stranded polynucleotide is electrophoretically threaded through a protein or solid-state nanopore. The differential interaction of nucleotides with the pore's narrow constriction during translocation contributes to changes in the channel's ion current and can in principle be used to gain sequence information on the polynucleotide[2–4]. As the speed of unperturbed electrophoretic polynucleotide passage precludes the resolution of individual nucleotides, a DNA- or RNA-binding motor enzyme is added to the system to gain a processive and slowed down passage (millisecond scale per nucleotide) of the polynucleotide through the nanopore[5,6].

Biological nanopore systems offer the advantage over synthetic or solid-state nanopores of having a uniform pore size and conformation, and the possibility to make pore alterations through protein engineering and/or alternative combinations with processivity enzymes[7]. The use of protein nanopores for DNA sequencing was first introduced for α-hemolysin[8] and later for MspA (ref. [9]). Both pores were heavily optimized in the years that followed, notably by combining them with phi29 DNA polymerase to slow down DNA strand translocation[10–15]. Other pores that have been shown to thread single-stranded DNA with some level of nucleotide discrimination include aerolysin, FhuA, OmpF, OmpG, ClyA, PA63, SP1 and the viral connectors phi29, SPP1, T3 and T4 (ref. [7]). Together, the nature of the enzyme, the structure and chemical composition of the nanopore, and the configuration of the nanopore–enzyme complex contribute to the signal-to-noise ratio and therefore sequencing accuracy of the system. In 2016, Oxford Nanopore Technologies (ONT) shifted from R7.4 to R9 chemistry[16], implementing the *Escherichia coli* curli transport channel CsgG (ref. [17]) as the sequencing nanopore. CsgG is a 36-stranded β-barrel responsible for the secretion of the bacterial amyloid curli across the outer membrane of Gram-negative bacteria[17–19]. The CsgG pore has a nine-fold circular symmetry (C_9) and a well-defined ~1-nm-wide constriction centrally located in a 4-nm-wide channel[17,20]. During curli secretion, CsgG works in conjunction with the accessory proteins CsgE and CsgF (ref. [21]), which are located in the periplasm and on the cell surface, respectively, and give substrate selection to the channel and ensure the assembly of surface-associated curli fibers[22,23]. While CsgE was found to bind and gate the CsgG channel[17], whether and how CsgF alters the CsgG channel properties is currently unknown. Here, we report characterization of the interaction between CsgG and CsgF, including a single-particle electron cryo-microscopy (cryo-EM) structure and functioning of this complex in sequencing DNA.

Biochemical characterization of the CsgG–CsgF complex

Addition of CsgF to purified, detergent-solubilized CsgG produced a qualitative shift in CsgG retention time on size-exclusion chromatography, suggesting an interaction of both proteins (Fig. 1a). Wild-type (WT) *E. coli* CsgG elutes as two populations corresponding to D_9 (ninefold dihedral) octadecamers and C_9 nonamers, the former a known in vitro artifact resulting from tail-to-tail dimerization of concentrated, membrane-extracted CsgG pores[17]. Adding a fivefold molar excess of CsgF shifted both CsgG populations and showed the presence of CsgF in both CsgG elution fractions as visualized on SDS–PAGE (Fig. 1b). Excess CsgF eluted at the apparent molecular weight expected for the monomeric protein. Furthermore, tris-borate native PAGE showed a quasi-quantitative shift in D_9 and C_9 CsgG retention times, indicating stable and complete complex formation (Fig. 1c). Similarly, a tandem affinity pull-down of detergent-solubilized proteins from cultures co-expressing Strep II-tagged CsgG and His-tagged CsgF resulted in the isolation of pure CsgG–CsgF complex, indicating that also in vivo the two proteins form a complex in the bacterial outer membrane (Extended Data Fig. 1a).

[1]Structural Biology Brussels, Vrije Universiteit Brussel, Brussels, Belgium. [2]Structural and Molecular Microbiology, VIB-VUB Center for Structural Biology, VIB, Brussels, Belgium. [3]Oxford Nanopore Technologies Ltd, Oxford, UK. ✉e-mail: han.remaut@vub.be

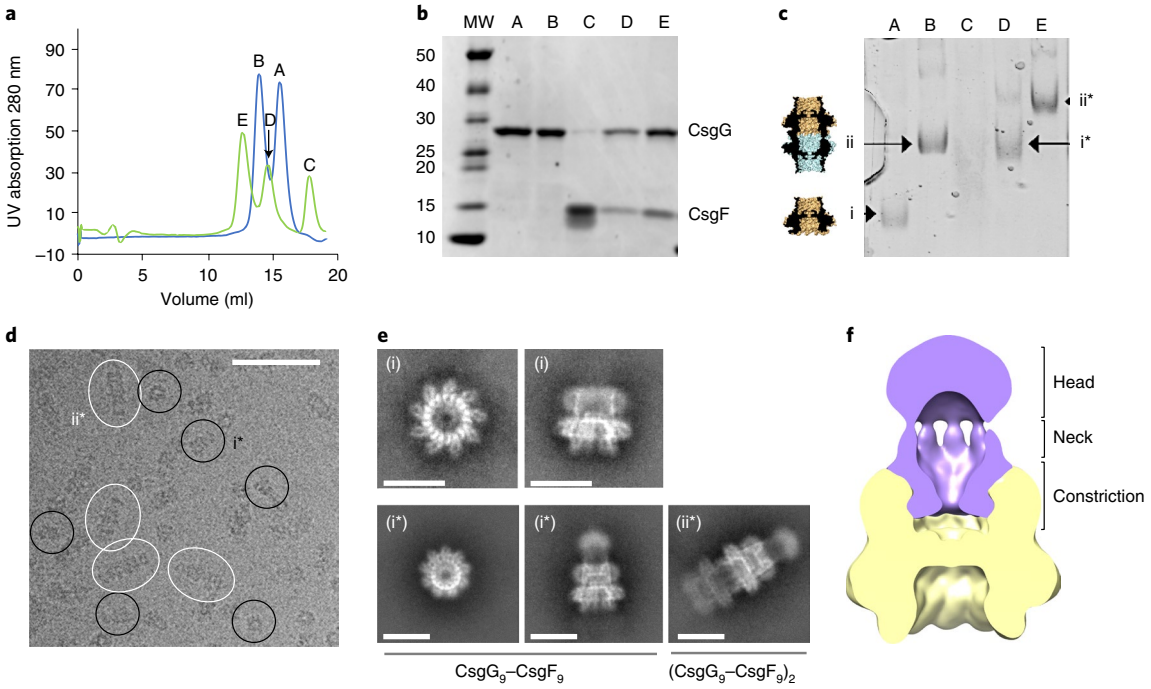


Fig. 1 | CsgG forms a stable complex with CsgF. a, Comparison of size-exclusion profiles of CsgG with (green) and without (blue) excess CsgF. **b**,**c**, 4–20% TGX stain-free SDS-PAGE (**b**) and tris-borate native PAGE (**c**) of the elution fractions labeled A–E in **a**, corresponding to, respectively, C_9 CsgG single channels (i), D_9 CsgG dimeric channels (ii), excess CsgF, and the CsgF complexes of C_9 CsgG (i*) and D_9 CsgG (ii*). The experiment was repeated three times. **d**, A cryo-EM image. Single particles of single (i*) and dimeric (ii*) CsgG–CsgF pores are circled black and white, respectively. Scale bar, 50 nm. **e**, Representative 2D class averages highlighting views along the C_9 symmetry axis (left) and side views of single (middle) and double (right) CsgG (upper row) and CsgG–CsgF (lower row) pores. Scale bars, 10 nm. **f**, Slice-through 3D volume of the CsgG–CsgF complex filtered down to 15 Å using EMAN2[29] and segmented and displayed at a contour of 0.0073 in UCSF Chimera[30]. Density corresponding to CsgG and CsgF is colored gold and purple, respectively.

Cryo-EM structure of the CsgG–CsgF curli assembly complex

To gain structural insights into the CsgG–CsgF channel, the complex was visualized by single-particle cryo-EM. Cryo-EM micrographs revealed discrete particles that split into two main populations corresponding to C_9 and D_9 CsgG channels. Top views show the expected nine-fold symmetry corresponding to the C_9/D_9 axis in the crystal structure[17] (Fig. 1e). In the CsgG–CsgF sample, both populations showed the presence of an additional protruding density along the nine-fold symmetry axis, not seen in micrographs of CsgG channels alone (Fig. 1d,e). Class averages show additional high-resolution features inside the main body of the channel as well as a diffuse protrusion from the extracellular side of the channel corresponding to CsgF (Fig. 1e). Based on class averages and a C_9-symmetrized low-resolution three-dimensional (3D) reconstruction, the CsgF density can be separated into a globular extracellular head domain that is linked by a thin neck region to a well-defined constriction region lining the inside of the CsgG barrel (Fig. 1f). Following computational enrichment for C_9 particles through multiple rounds of 2D and 3D classification in Relion (cfr class averages in Extended Data Fig. 1b), a 3D cryo-EM electron potential map was calculated to a resolution of 3.4 Å (Extended Data Fig. 1c,d). While the reconstruction showed unambiguous density for the first 35 residues of mature CsgF (Fig. 2 and Extended Data Fig. 1c,d), the head and neck density had poor definition and did not allow model building of the corresponding domains. The CsgF amino terminus reaches deep inside the CsgG β-barrel, where the first four residues (G1 to T4) make an intricate hydrogen-bond network with residues Q153, D155, T207 and N209 in the β-barrel lumen, which is followed by a stretch of more hydrophobic residues (Fig. 2a). A strictly conserved NPXFGG motif at residues 9–14 makes the CsgF N terminus kink into the lumen of the CsgG β-barrel, giving rise to a sharp, 15-Å-wide constriction of the channel, lined by residue N17 at the constriction apex (Figs. 2a,d and 3a). From there, the CsgF peptide runs back to the rim of the CsgG β-barrel via a 13-residue helix (helix 1), opening into a 38-Å-wide exit of the channel. Notably, the N terminus was found fully disordered in the solution structure of CsgF (ref. [24]), indicating that it adopts a stable conformation only upon interaction with its partner CsgG. In the CsgG–CsgF complex, the CsgF protomers form an interaction interface with CsgG corresponding to a 1,030 Å^2 solvent-accessible surface area (corresponding to a calculated solvation free energy of −8.1 kcal mol^{-1}) and incorporating ten and six putative hydrogen-bond and salt-bridge interactions, respectively, and have a 330 Å^2 solvent-accessible surface interface with an adjacent CsgF protomer, the latter lacking polar interactions. Together, these interactions result in a highly stable non-covalent pore complex that withstands heating up to 70 °C, the melting temperature of the CsgG channel (Extended Data Fig. 2a). Interestingly, the first 35 residues forming the CsgF constriction peptide (FCP) show a 48% average pairwise sequence identity amongst CsgF homologs, versus a mere 28% identity in the neck and head regions, suggesting that the CsgG contact and the conformation of the FCP are the most conserved features in the CsgF accessory protein (Extended Data Fig. 3).

Capture of ssDNA by CsgG channels

CsgG channels show stable constitutively open conductance levels in single-channel recordings corresponding to the 9 Å solvent-accessible diameter of the CsgG constriction loop[17] (Extended Data Fig. 4). During nanopore sequencing, the CsgG periplasmic vestibule resides on the *cis* side of the membrane, accepts the translocating polynucleotide and forms the contact face with the processivity enzyme (Extended Data Fig. 4a). Despite the

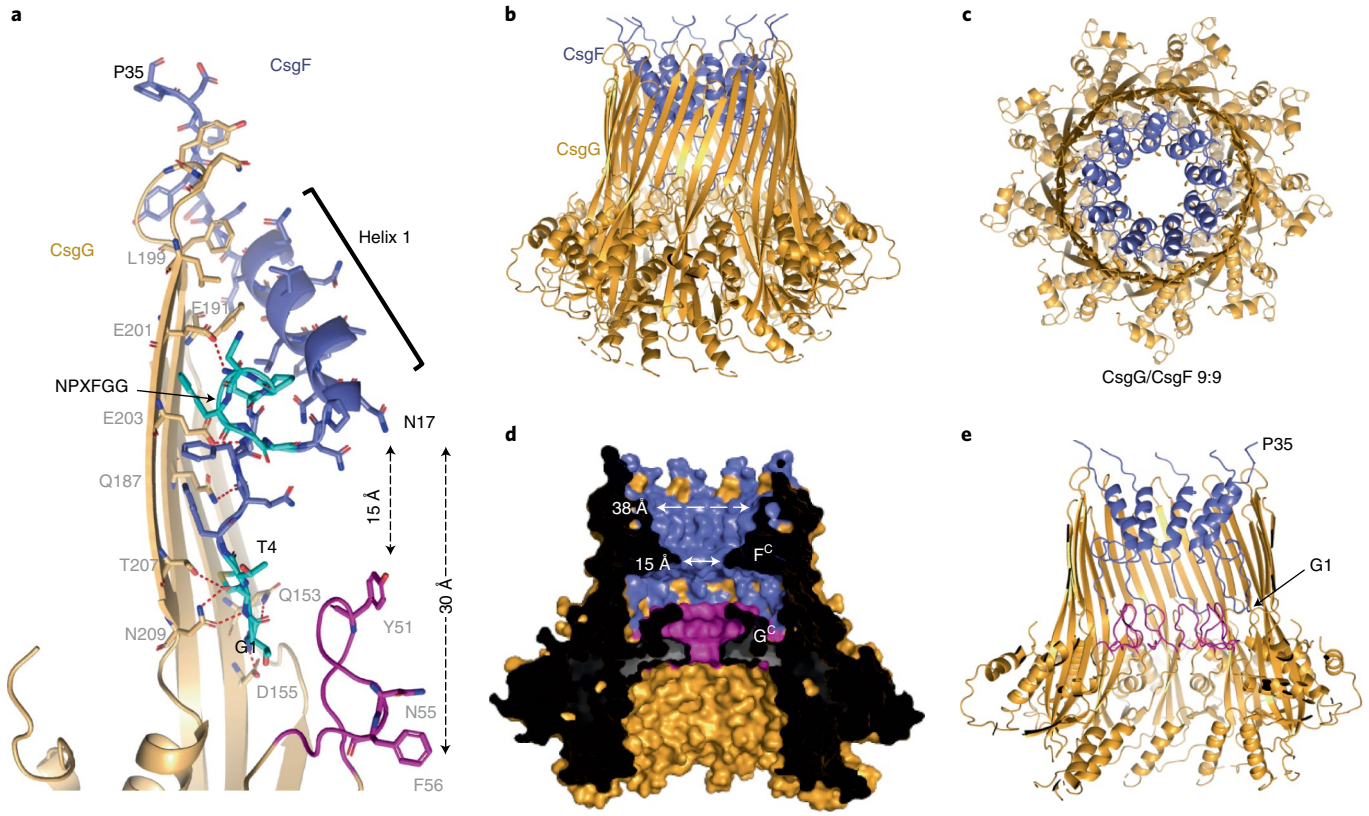


Fig. 2 | CsgG–CsgF cryo-EM structure reveals a dual-constriction pore. a, A close-up ribbon-and-sticks (CsgF) representation of a single CsgG (gold) and CsgF (purple) protomer of the CsgG–CsgF cryo-EM structure. The CsgG constriction formed by Y51, N55 and F56 is highlighted magenta, and the N-terminal four residues and the conserved NPXFGG motif in CsgF are highlighted in cyan. Oxygens are red; nitrogens are blue. Hydrogen bonds anchoring the CsgG–CsgF interaction are depicted as dashed red lines. **b**–**e**, The CsgG–CsgF pore shown in side (**b**), top (**c**) and cross-sectional views (**d**,**e**), depicted in ribbon (**b**,**c**,**e**) and solvent-accessible surface (**d**) representations, colored as in **a**. G^C, CsgG constriction; F^C, CsgF constriction. Only the CsgF N terminus (residue G1 to P35) forming the CsgF-constriction peptide (FCP) could be resolved in the cryo-EM map (Extended Data Fig. 2).

narrow constriction defined by Y51, N55 and F56, WT CsgG pores capture ssDNA, producing a squiggling signal centered around 30 pA with a range of ±10 pA as the DNA strand is translocated through the pore at −180 mV (Fig. 3b). WT CsgG pores have been modified extensively to improve their properties in DNA sequencing. One such modified CsgG pore that contains the F56Q alteration ($CsgG^{F56Q}$) produces a larger open-pore current level compared to the WT CsgG pore (Extended Data Fig. 4b) and produces a signal with a higher range when a DNA strand is translocated through the pore (Fig. 3b). Another modified CsgG channel referred to as R9, proprietary to ONT, with optimized signal-to-noise ratio and improved discrimination of passing nucleotides has formed the baseline for DNA and RNA sequencing nanopores used in ONT's sequencing platforms since 2016[16] (Fig. 3b).

Selected CsgG–FCP pores capture ssDNA

Our structural studies show that binding of CsgF results in the formation of a second constriction in the CsgG channel, spaced 15 to 30 Å above the exit and entrance of the CsgG constriction, respectively (Figs. 2d and 3a). With a solvent-excluded diameter of 15 Å, the constriction formed by the N-terminal peptide of CsgF is slightly wider than that of CsgG. This is in the range commonly regarded as being useful for nanopore sequencing, with described nanopores having dimensions ranging from 10 Å to 36 Å (ref. [7]). In contrast to the three-layered native CsgG constriction, CsgF forms a sharp, single-layered orifice made by N17 (Figs. 2 and 3a). The novel structural insights from the CsgG–CsgF complex were used to introduce a second constriction in the CsgG pore, made by the N-terminal peptide of CsgF, dubbed FCP. For production of the CsgG–FCP complex, TEV protease recognition sites were introduced in the CsgF neck region at positions 30, 35 or 40 of the mature protein (Extended Data Fig. 2b). Upon in vitro reconstitution of the CsgG–CsgF complex and proteolytic cleavage using TEV-protease, the CsgF neck and head region were removed, obtaining a stable CsgG–FCP channel complex (Extended Data Fig. 2c). Using this approach, we formed CsgG–FCP channels with WT CsgG as well as with the basic sequencing pore $CsgG^{F56Q}$ and the different proprietary $CsgG^{R9}$ pores developed by ONT. The resulting complexes could be reconstituted into artificial membranes of the MinION flow cells developed by ONT and showed stable single-channel conductance in agreement with the expected channel dimensions (Extended Data Fig. 4b,c). Next, addition of DNA–enzyme complexes to flow cells containing CsgG–FCP, $CsgG^{F56Q}$–FCP or $CsgG^{R9}$–FCP resulted in squiggling conductance profiles in agreement with the capture and passage of ssDNA (Fig. 3b). Remarkably, the conductivity profiles showed that the CsgG–FCP and $CsgG^{R9}$–FCP pores remain as stable complexes both in solution and in the polymeric membrane. Threading of DNA during the sequencing run had no apparent effect on $CsgG^{R9}$–FCP single-channel lifetime, where over 90% of the pore complex remained stable over the 24 h tested (Extended Data Fig. 5).

Static DNA strand signal discrimination of hybrid CsgG–FCP pores

To evaluate whether both the CsgG and CsgF constrictions can modulate the global ion conductance during DNA strand passage,

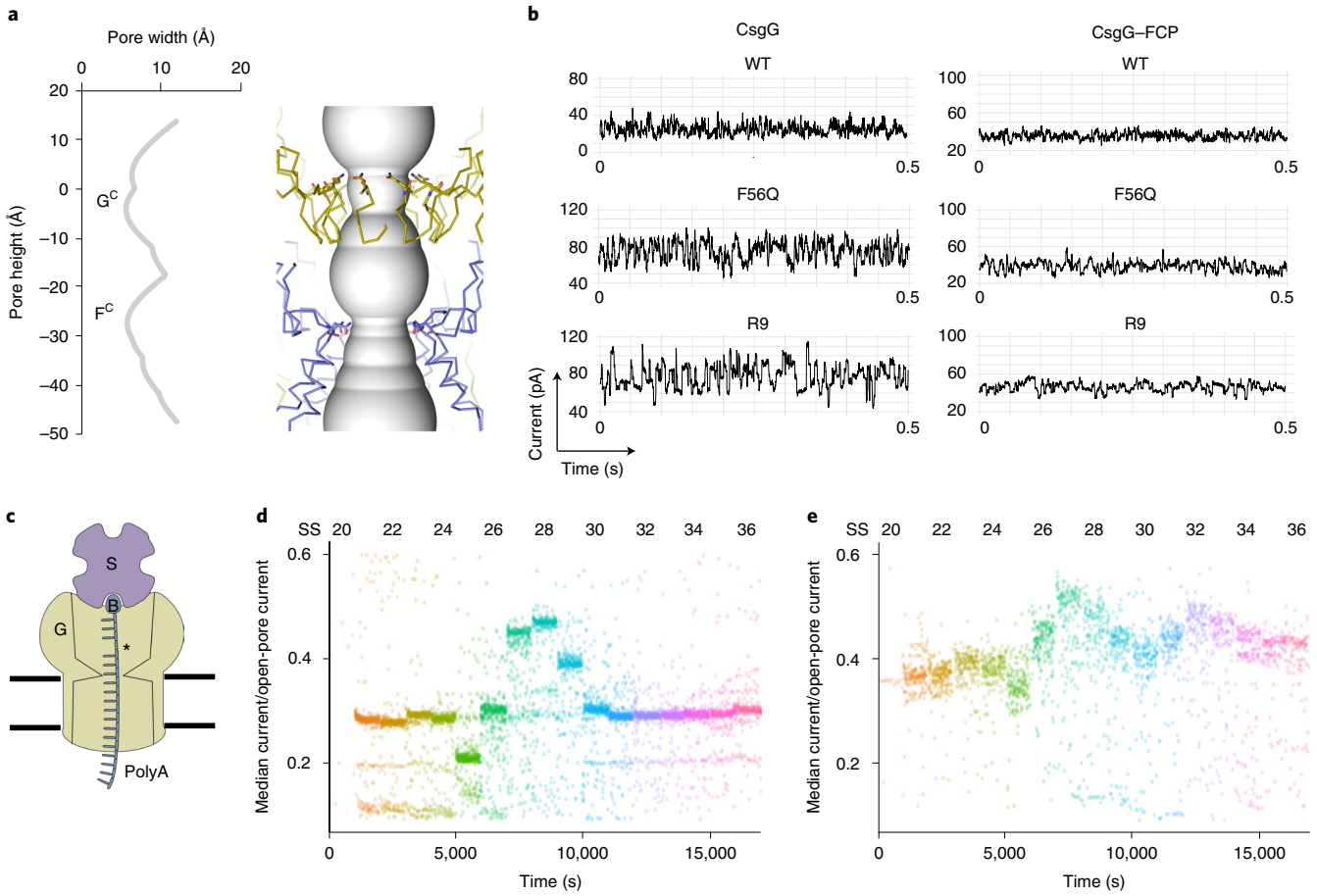


Fig. 3 | The CsgF constriction peptide creates a second constriction in the CsgG pore. a, Channel radius plotted against channel height (left) and its corresponding position in the CsgG^{F56Q}–FCP complex (right). Distances are in ångström. CsgG and FCP are depicted in gold and purple, respectively. Q56 in the CsgG constriction and N17 CsgF are shown in stick representation. **b**, Representative current signatures during passage of single DNA strands through the WT CsgG, CsgG^{F56Q} and CsgG^{R9} nanopores and their respective CsgG–FCP complexes. Data were measured at −180 mV and are representative of >100 capture events and >10 single channels. (The experiment shown is representative of at least three repeat experiments.) **c**, A schematic diagram of the pore read point detection assay. Pores are probed with oligonucleotides (SS20–SS36; Extended Data Fig. 6a) with an abasic nucleotide (asterisk) at a defined distance from a biotin (B)–streptavidin (S) blockage. When the abasic residue resides at the pore constriction, this results in increased conductance levels. **d**,**e**, Current levels for different static oligonucleotides (SS20–SS30) bound in CsgG^{F56Q} (**d**) or CsgG^{F56Q}–FCP (**e**). Each dot represents a single data point (at least 255 (**d**) and 76 (**e**) data points per oligonucleotide, measured from at least $n = 24$ pores).

we measured single-channel conductance of CsgG^{F56Q}–FCP channels exposed to a series of oligonucleotides (dubbed SS20 to SS38) that were modified to lack a single base (iSpc3 positions) at a discrete distance from a biotin–streptavidin complex (Extended Data Fig. 6a). The last procedure results in static pore–oligonucleotide complexes with single base aberrations at variable height from the channel entrance that result in an increase in conductance when residing at a channel constriction (Fig. 3c). Probing CsgG^{F56Q} channels with the different oligonucleotides showed a discrete rise in conductance with SS28 and extending into SS27 and SS29, demonstrating that a discrete modulation of channel conductance occurred at 9 ± 1 nucleotides from the channel entrance (Fig. 3d). When CsgG^{F56Q}–FCP channels were probed, an additional rise in conductance with SS31–SS33 was observed at 13 ± 1 nucleotides from the channel entrance (Fig. 3e). Assuming an extended ssDNA with a maximum length per base of 6.3 (±0.8) Å (ref. [25]), the nucleotide distance between these two regions causing discrete channel conductance modulations is in agreement with the ~25 Å spacing of the centers of the CsgG and CsgF constrictions measured from the cryo-EM structure (Fig. 3a).

Improved homopolymer resolution by a dual-constriction pore

By using immobilized synthetic oligonucleotides, it has been shown that additional sequence information can be gained when two constrictions, rather than one, are employed within a single nanopore[11]. Therefore, we analysed whether the observed dual-constriction nanopores offer opportunities to resolve homopolymer regions during polynucleotide sequencing, a known cause of inaccuracies in nanopore sequencing[26,27]. Translocation of homopolymer DNA through CsgG pores results in little to no modulation in conductance, leading to miscalling of the length of the homopolymer and may result in indel errors in base calling during nanopore sequencing (Fig. 4a,b,d), particularly for long homopolymers. To test the effect of a dual constriction, nanopore sequencing reads from oligonucleotides containing three regularly spaced stretches of ten deoxythymidines (T) were compared between the sequencing nanopore CsgG^{R9} and its complex with FCP (CsgG^{R9}–FCP) (Fig. 4a,b). Passage of the polyT section through CsgG^{R9}–FCP showed increased modulation in pore conductance demonstrating that the conductance signal during single strand passage is

likely a function of base interactions with both constrictions (Fig. 4b,c). Sequencing of a series of plasmids modified to contain three regularly spaced T homopolymers ranging in length from 3 to 9 showed that the increased signal complexity in the polynucleotide spanning regions observed for CsgG^{R9}–FCP squiggles results in a higher accuracy in homopolymer length calling compared to CsgG^{R9} (Fig. 4d). Whereas for CsgG^{R9} the unpolished homopolymer length calling started to go down in accuracy from pentameric polyT stretches onwards, CsgG^{R9}–FCP showed a 20 to 70% improved length calling up to at least octameric or nonameric polyT stretches. To evaluate whether the increased homopolymer accuracy extends to the four bases, *E. coli* genomic DNA was sequenced using both CsgG^{R9} and CsgG^{R9}–FCP pores. We found no substantial effects on average read length in CsgG^{R9}–FCP compared to CsgG^{R9}, a property that largely depends on library preparation rather than pore characteristics. When the single reads of the two pores were aligned and compared, more deletions can be seen at the edge of homopolymer regions read by CsgG^{R9} compared to the CsgG^{R9}–FCP pore (Extended Data Fig. 6b). As a result, the polished consensus accuracy of the homopolymer region drops in CsgG^{R9} pores when the length of the homopolymer is longer than five bases (Fig. 4e). Comparatively, CsgG^{R9}–FCP was able to read the homopolymeric region more accurately over a longer length of bases (Fig. 4e). Thus, the addition of a second constriction by means of the FCP increased accuracy in calling longer homopolymer regions. Further development of the neural network models used to base call the electrical signals[28] along with further modifications to the pore design may further improve base-calling accuracies of CsgG–FCP dual-constriction pores.

Discussion

Our data provide proof-of-principle for the construction of a dual-constriction nanopore using the curli secretion channel CsgG or its nanopore derivatives and the N-terminal 35 residues of its secretion partner CsgF, and demonstrate the combined contribution of the two constriction points to conductance signals during passage of a polymer through the nanopore. Using this concept, we built a prototype dual-constriction sequencing nanopore made of FCP-modified CsgG^{R9}, and show its principle advance in homopolymer calling. Thus, FCP peptides form a convenient tool to modify the channel properties of CsgG-based nanopores. Since FCP peptides bind the channel *trans* side, modifying CsgG-based channels with FCP does not affect *cis* side properties such as contact with nucleotide-binding enzymes and analyte delivery to the pore. Besides applications in polynucleotide sequencing as shown in this study, we anticipate that dual-constriction pores may have advantages in small-molecule analyte sensing due to the consecutive passage of analytes through two chemically distinct points of interaction.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-020-0570-8.

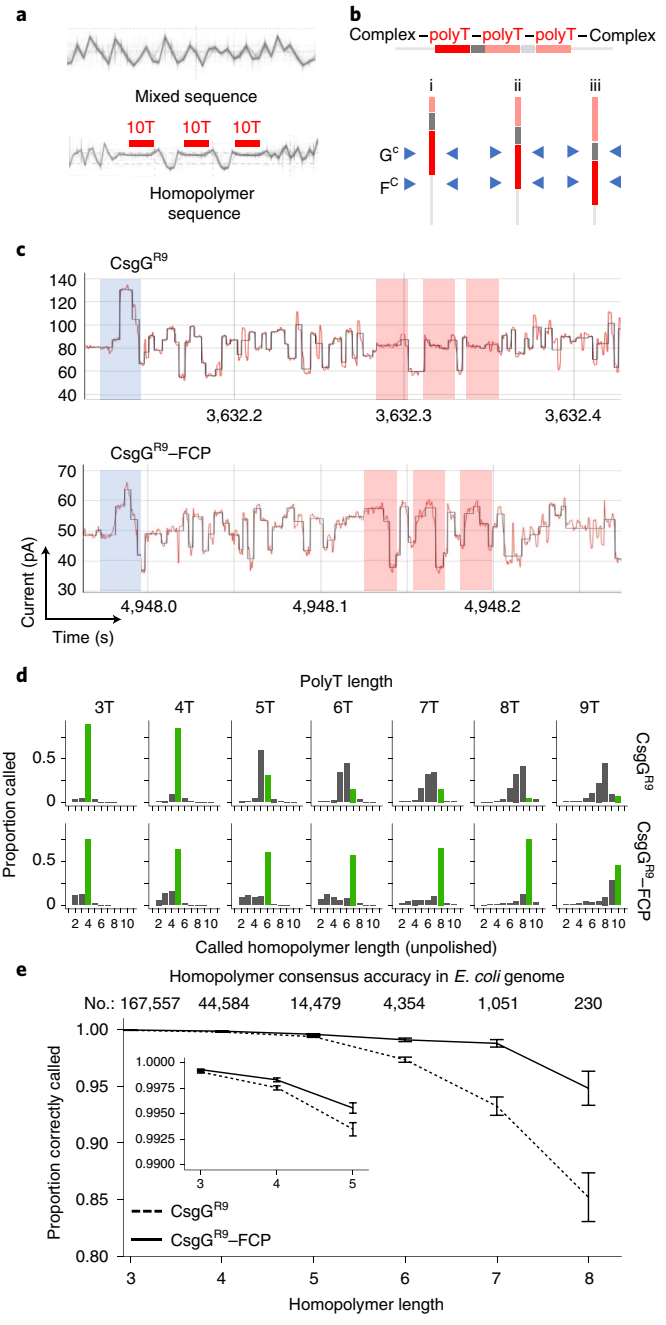


Fig. 4 | Homopolymer base calling by a prototype CsgG–FCP nanopore. **a**, Overlaid single-molecule conductance profiles of a mixed ssDNA sequence (top) and a trial ssDNA sequence (bottom) containing three consecutive homopolymers of ten deoxythymidines (10T) spaced by GGAA intervals as read using the CsgG^{R9} nanopore (>500 individual traces coming from at least 50 pores). **b**, A schematic representation of three interaction scenarios of the trial ssDNA sequence by the CsgG and FCP constrictions, labeled G^C and F^C, respectively. The dual constriction is expected to increase the inclusion of sequence outside the homonucleotide stretch during passage through the pore (scenarios i and iii). **c**, Single-molecule conductance signals of the 10T-homopolymer-containing trial sequence (shown in **a**) analysed by CsgG^{R9} (upper) or CsgG^{R9}–FCP (lower) pores. The shaded zones correspond to the adaptor (blue) and the 10T (red) regions. Traces are representative for >1,000 capture events and >50 pores. **d**, Histograms[31] of single-read homopolymer length calling of ssDNA containing polyT stretches ranging from 3 to 9 nucleotides in length, sequenced by CsgG^{R9} or CsgG^{R9}–FCP. Correctly called homopolymer lengths are shown in green (plots contain at least 166,000 single reads per oligonucleotide). **e**, Comparison of the proportion (±s.d.) of correctly called homopolymers versus homopolymer length for CsgG^{R9} and CsgG^{R9}–FCP pores. The plot shows consensus accuracies across the four bases, using data polished with the medaka software package developed at ONT[32], and is based on an *E. coli* assembly of depth 100×. Occurrences of the respective homopolymer lengths in the *E. coli* genome are indicated at the top. Nonameric ($n=22$) and longer ($n=2$) homopolymers become too rare to provide statistically relevant numbers.

Received: 13 August 2019; Accepted: 19 May 2020;
Published online: 06 July 2020

References

1. Bayley, H. & Cremer, P. S. Stochastic sensors inspired by biology. *Nature* **413**, 226–230 (2001).
2. Howorka, S., Cheley, S. & Bayley, H. Sequence-specific detection of individual DNA strands using engineered nanopores. *Nat. Biotechnol.* **19**, 636–639 (2001).
3. Meller, A., Nivon, L., Brandin, E., Golovchenko, J. & Branton, D. Rapid nanopore discrimination between single polynucleotide molecules. *Proc. Natl Acad. Sci. USA* **97**, 1079–1084 (2000).
4. Akeson, M., Branton, D., Kasianowicz, J. J., Brandin, E. & Deamer, D. W. Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules. *Biophys. J.* **77**, 3227–3233 (1999).
5. Benner, S. et al. Sequence-specific detection of individual DNA polymerase complexes in real time using a nanopore. *Nat. Nanotechnol.* **2**, 718–724 (2007).
6. Olasagasti, F. et al. Replication of individual DNA molecules under electronic control using a protein nanopore. *Nat. Nanotechnol.* **5**, 798–806 (2010).
7. Wang, S., Zhao, Z., Haque, F. & Guo, P. Engineering of protein nanopores for sequencing, chemical or protein sensing and disease diagnosis. *Curr. Opin. Biotechnol.* **51**, 80–89 (2018).
8. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl Acad. Sci. USA* **93**, 13770–13773 (1996).
9. Butler, T. Z., Pavlenok, M., Derrington, I. M., Niederweis, M. & Gundlach, J. H. Single-molecule DNA detection with an engineered MspA protein nanopore. *Proc. Natl Acad. Sci. USA* **105**, 20647–20652 (2008).
10. Stoddart, D., Franceschini, L., Heron, A., Bayley, H. & Maglia, G. DNA stretching and optimization of nucleobase recognition in enzymatic nanopore sequencing. *Nanotechnology* **26**, 084002 (2015).
11. Stoddart, D. et al. Nucleobase recognition in ssDNA at the central constriction of the alpha-hemolysin pore. *Nano Lett.* **10**, 3633–3637 (2010).
12. Stoddart, D., Heron, A. J., Mikhailova, E., Maglia, G. & Bayley, H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proc. Natl Acad. Sci. USA* **106**, 7702–7707 (2009).
13. Maglia, G., Heron, A. J., Stoddart, D., Japrung, D. & Bayley, H. Analysis of single nucleic acid molecules with protein nanopores. *Methods Enzymol.* **475**, 591–623 (2010).
14. Cherf, G. M. et al. Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. *Nat. Biotechnol.* **30**, 344–348 (2012).
15. Manrao, E. A. et al. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* **30**, 349–353 (2012).
16. Brown, C. G. No Thanks, I've already got one. *YouTube* https://www.youtube.com/watch?v=nizGyutn6v4 (2016).
17. Goyal, P. et al. Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. *Nature* **516**, 250–253 (2014).
18. 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).
19. Van Gerven, N., Van der Verren, S. E., Reiter, D. M. & Remaut, H. The role of functional amyloids in bacterial virulence. *J. Mol. Biol.* **420**, 3657–3684 (2018).
20. Cao, B. et al. Structure of the nonameric bacterial amyloid secretion channel. *Proc. Natl Acad. Sci. USA* **111**, E5439–E5444 (2014).
21. Chapman, M. R. et al. Role of *Escherichia coli* curli operons in directing amyloid fiber formation. *Science* **295**, 851–855 (2002).
22. 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. USA* **106**, 900–905 (2009).
23. Nenninger, A. A. et al. CsgE is a curli secretion specificity factor that prevents amyloid fibre aggregation. *Mol. Microbiol.* **81**, 486–499 (2011).
24. Schubeis, T. et al. Structural and functional characterization of the curli adaptor protein CsgF. *FEBS Lett.* **592**, 1020–1029 (2018).
25. Chi, Q., Wang, G. & Jian, J. The persistence length and length per base of single-stranded DNA obtained from fluorescence correlation spectroscopy measurements using mean field theory. *Phys. A* **392**, 1072–1079 (2013).
26. Jain, M., Olsen, H. E., Paten, B. & Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* **17**, 239 (2016).
27. Carter, J. M. & Hussain, S. Robust long-read native DNA sequencing using the ONT CsgG Nanopore system. *Wellcome Open Res.* **2**, 23 (2017).
28. Wick, R. R., Judd, L. M. & Holt, K. E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* **20**, 129 (2019).
29. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
30. Pettersen, E. F. et al. UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
31. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
32. Medaka v.0.8.1 (Oxford Nanopore Technologies, 2018); https://nanoporetech.github.io/medaka/

Methods

Strains and protein expression constructs. *E. coli* Top10 was used for all cloning procedures. *E. coli* C43(DE3)[33] and BL21(DE3) were used for protein production (Supplementary Table 1). Expression of carboxy-terminally Strep II–tagged *E. coli* CsgG as an outer-membrane-localized pore made use of plasmid pPG1 (ref. [17]). For the expression of C-terminally 6×His-tagged CsgF in the *E. coli* cytoplasm, a PCR product encompassing the coding sequence for mature *E. coli* CsgF (that is, CsgF without its signal sequence, Supplementary Table 1, sequence 1; primers 1 and 2) was cloned into pET22b via the NdeI and EcoRI sites, resulting in the CsgF–His expression plasmid pNA101. CsgG–Strep II and CsgF–His were co-expressed using plasmid pNA62. pNA62 is a pTrc99a-based vector encoding *csgF*–His and *csgG*–Strep II, and was created based on pNA152, a modified pTrc99a with the pDEST14 Gateway cassette and the *ampR* resistance gene replaced by the streptomycin/spectinomycin-resistance gene *aadA*. A PCR fragment encompassing part of the *E. coli* MC4100 (ref. [34]) *csgDEFG* operon corresponding to the coding sequences of *csgE*, *csgF* and *csgG* was generated with primers 3 and 4 (Supplementary Table 1; primer 4 adds the Strep II tag sequence SAWSHPQFEK to the CsgG C terminus) and was inserted into pDONR221 (ThermoFisher Scientific) via BP Gateway recombination to generate pNA41. The recombinant *csgEFG* operon was subsequently inserted into pNA152 via LR Gateway recombination to yield pNA43. The coding sequence for a 6×His-tag was added to the CsgF C terminus in pNA43 by Quick-change mutagenesis PCR using primers 5 and 6 (Supplementary Table 1), yielding pNA60b, followed by removal of *csgE* by outward PCR using primers 7 and 8 (Supplementary Table 1) to obtain pNA62. Coding sequences corresponding to the TEV recognition side (ELNYFQS) were introduced after positions 30, 35 and 45 of mature CsgF by Quick-change mutagenesis PCR.

Recombinant protein expression. CsgG–Strep II was expressed in *E. coli* BL21 (DE3) cells transformed with plasmid pPG1 (ref. [17]). The cells were grown at 37 °C to an optical density (OD) at 600 nm of 0.6 in Terrific broth medium. Recombinant protein production was induced with 0.0002% anhydrotetracyclin and the cells were grown at 25 °C for a further 16 h before being collected by centrifugation at 5,500*g*. CsgF–His was expressed in the cytoplasm of *E. coli* BL21(DE3) cells transformed with plasmid pNA101. Cells were grown at 37 °C to an OD at 600 nm of 0.6 followed by induction by 1 mM isopropyl-β-D-thiogalactoside and left to express protein for 15 h at 37 °C before being collected by centrifugation at 5,500*g*. Co-expression of CsgG–Strep II and CsgF–His was performed using *E. coli* C43(DE3) cells transformed with plasmid pNA62 and grown at 37 °C in Terrific Broth medium. When the cell culture reached an OD at 600 nm of 0.7, recombinant protein expression was induced with 0.5 mM isopropyl-β-D-thiogalactoside and cells were left to grow for 15 h at 28 °C, before being collected by centrifugation at 5,500*g*.

Protein purification of the CsgG–CsgF complex, CsgG and CsgF. For CsgG–CsgF tandem affinity purification, *E. coli* cells transformed with pNA62 were resuspended in 50 mM Tris–HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 5 mM $MgCl_2$, 0.4 mM AEBSF, 1 μg ml^{-1} leupeptin, 0.5 mg ml^{-1} DNase I and 0.1 mg ml^{-1} lysozyme. The cells were disrupted at 20 kPsi using a TS series cell disruptor (Constant Systems) and the lysed cell suspension was incubated for 30 min with 1% *n*-dodecyl-β-D-maltopyranoside (DDM; Inalco) for extraction of outer membrane components. Next, remaining cell debris and membranes were spun down by ultracentrifugation at 100,000*g* for 40 min. Supernatant was loaded onto a 5-ml HisTrap column (GE Healthcare) equilibrated in buffer A (25 mM Tris pH 8, 200 mM NaCl, 10 mM imidazole, 10% sucrose and 0.06% DDM). The column was washed with >10 column volumes of 5% buffer B (25 mM Tris pH 8, 200 mM NaCl, 500 mM imidazole, 10% sucrose and 0.06% DDM) in buffer A, and eluted with a gradient of 5–100% buffer B over 60 ml. The eluate was diluted twofold before loading overnight on a 5-ml Strep-Tactin column (IBA GmbH) equilibrated with buffer C (25 mM Tris pH 8, 200 mM NaCl, 10% sucrose and 0.06% DDM). The column was washed with >10 column volumes of buffer C and the bound protein was eluted in buffer C complemented with 2.5 mM desthiobiotin. CsgG–Strep II purification for in vitro reconstitution follows the protocol for CsgG–CsgF, except omitting sucrose in the buffers and bypassing the immobilized metal affinity chromatography step. CsgF–His purification for in vitro reconstitution was performed by resuspension of the cell mass in 50 mM Tris–HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 5 mM $MgCl_2$, 0.4 mM AEBSF, 1 μg ml^{-1} leupeptin, 0.5 mg ml^{-1} DNase I and 0.1 mg ml^{-1} lysozyme. The cells were disrupted at 20 kPsi using a TS series cell disruptor (Constant Systems) and the lysed cell suspension was centrifuged at 10,000*g* for 30 min to remove intact cells and cell debris. Supernatant was added to 5 ml of Ni immobilized metal affinity chromatography beads (Workbeads 40 IDA, Bio-Works Technologies AB) equilibrated with buffer D (25 mM Tris pH 8, 200 mM NaCl, 10 mM imidazole) and left incubating for 1 h at 4 °C. Ni-NTA beads were pooled in a gravity flow column and washed with 100 ml of 5% buffer E (25 mM Tris pH 8, 200 mM NaCl, 500 mM imidazole) diluted in buffer D. Bound protein was eluted by stepwise increase of buffer E (10% steps of each 5 ml). For EM and electrophysiology, CsgG–CsgF and CsgG–FCP protein complexes were obtained by either in vitro reconstitution or co-expression. For reconstitution, purified CsgG and CsgF were mixed at a molar ratio of 1:5 CsgG/CsgF to saturate the CsgG barrel with CsgF. Next, the reconstituted or co-expressed complex was injected on a Superose 6 10/30 column (GE Healthcare) equilibrated with buffer F (25 mM Tris pH 8, 200 mM NaCl and 0.03% DDM) and run at 0.5 ml min^{-1}. For CsgG–FCP, the CsgF forms correspond to CsgF^{35TEV} and reconstituted CsgG–CsgF complexes were digested at room temperature overnight with TEV protease in buffer F. The mixture was then run back through a 5-ml HisTrap (GE Healthcare) column and the flow-through was collected, heated at 60 °C for 15 min and centrifuged at 21,000*g* for 10 min before use in electrophysiology. Protein concentrations were determined on the basis of calculated absorbance at 280 nm and assuming 1/1 stoichiometry.

SDS and native PAGE. Fractions from size exclusion were boiled for 5 min in loading buffer (final concentrations of 60 mM Tris-HCl pH 6.8, 2% SDS, 10% v/v glycerol and 0.01% bromophenol blue) and run on Mini-Protean 4–15% TGX SDS–PAGE in Laemmli running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS pH 8.3) at 300 V for 20 min followed by staining with InstantBlue (Expedeon). The molecular weight ladder used is PageRuler Unstained (ThermoFisher). Samples for native PAGE were mixed 3:1 with sample buffer (1× TBE, 20% v/v glycerol, 0.04% bromophenol blue) and run at 90 V for 2 h on a 4.5% native PAGE in 0.5× TBE buffer (1× TBE is 90 mM Tris-borate pH 8.3; 1 mM EDTA).

Structural analysis using cryo-EM. Sample behavior of the size-exclusion fractions was probed using negative-stain EM. Samples are stained with 1% uranyl formate and imaged using an in-house 120-kV JEM 1400 (JEOL) microscope equipped with a LaB_6 filament. For high-resolution cryo-EM analysis, CsgG–CsgF samples were prepared by spotting 3 μl of sample on R2/1 holey grids (Quantifoil) coated with graphene oxide (Sigma Aldrich), manually blotted and plunged in liquid ethane using a CP3 plunger (Gatan). Sample quality was screened on the in-house JEOL JEM 1400 before collecting a dataset on a 300-kV Titan Krios (FEI, Thermo-Scientific) microscope equipped with a K2 Summit direct electron detector (Gatan) at the Astbury Centre in Leeds, UK (CsgG–CsgF) or at Diamond – eBIC, Harwell Science and Innovation Campus, UK (CsgG). The detector was used in counting mode with a cumulative electron dose of 56 electrons per square ångström spread over 50 frames. For CsgG–CsgF, 2,045 images were collected with a pixel size of 1.07 Å. Images were motion-corrected with MotionCor2.1[35] and defocus values were determined using ctffind4[36]. No dose-weighting scheme was applied and the first two frames were discarded as the result of excessive motion. Data were further analysed using a combination of EMAN2[29], RELION2.0[37] and SIMPLE[38]. Fifty micrographs were manually picked using e2boxer.py from the EMAN2 package resulting in 4,000 particles for which 2D templates were generated using Relion 2D classification. Next, particles were picked automatically using Gautomatch (Kai Zhang, https://www2.mrc-lmb.cam.ac.uk/research/locally-developed-software/zhang-software/#gauto), yielding a total of 750,000 particles. Subsequent 2D stack cleaning, discarding graphene oxide lines and the dimeric barrels (Fig. 1e), yielded a stack of 135,000 particles for 3D refinement. C_9 symmetry was imposed during 3D model generation and refinement. An initial model was computed using e2initialmodel.py from the EMAN2 package. 3D classification in Relion was used to enrich for head-containing particles, indicative for the presence of CsgF. This yielded a stack of 62,000 particles used to calculate the final map at 3.4 Å resolution (Supplementary Table 2). De novo model building of CsgF was performed with COOT[39] and iterative cycles of model building, and refinement of the full complex comprising nine CsgG copies ranging from residue 10 to 260 (the loop spanning residues 102–111 was disordered and is absent from the model) and nine copies of CsgF ranging from residue 1 to 35 (mature protein numbering) was performed with PHENIX real-space refinement[40] in combination with COOT (data and model statistics are provided in Supplementary Table 2). Surface and cartoon representations were generated with The PyMOL Molecular Graphics System, Version 2.0 Schrödinger. The pore plots in Fig. 3a were generated with HOLE[41], using the endrad 10 parameter, plotted with MS Excel 2013 and dumbbell representation was visualized in PyMol (https://pymol.org/2/).

Testing CsgG and CsgG–FCP pores on MinIONs. All CsgG or CsgG–FCP samples were incubated with Brij58 (final concentration of 0.1%) for 10 min at room temperature before diluting CsgG (1 mg ml^{-1}) samples 1:100,000 and CsgG–FCP (1 mg ml^{-1}) samples 1:5,000 in MinION flow cell buffer (25 mM potassium phosphate, 150 mM potassium ferrocyanide, 150 mM potassium ferricyanide, pH 8.0) for pore insertion. All pore experiments were performed on MinION flow cells (lacking any pre-inserted nanopores) using MinION devices developed by ONT. MinKNOW core 1.11.5 version software developed and provided by ONT was used to control scripts during all experiments. For insertion of CsgG or CsgG–FCP pores, 300 μl of diluted pore samples was loaded into the priming port of the flow cell. The pore insertion script of MinKNOW was used to apply voltage starting from −150 mV, increasing 10 mV every 15 s until −450 mV. A 1-ml volume of flow cell buffer was perfused through the priming port to remove any excess pores. Groups and positions with single pores were evaluated using the standard flow cell check protocol using MinKNOW. Open-pore currents of CsgG or CsgG–FCP pores were recorded at −180 mV before adding any DNA (Extended Data Fig. 4b).

Generation of *I–V* curves on MinION. To obtain the current–voltage (*I–V*) curves of CsgG or CsgG–FCP pores in Extended Data Fig. 4c, flow cells with different pores were prepared as shown above. Different electric potentials were applied

using MinKNOW, ranging from −200 mV to +200 mV, where every 30 s the potential was increased by 25 mV. For both CsgG and CsgG–FCP pores, data were analysed over multiple pores (as indicated) within a flow cell by measuring the mean open-pore current at each potential. The mean open-pore current and the 95% confidence level were calculated for each potential and plotted using R[42].

DNA sequencing experiments on MinION. To evaluate the performance of CsgG and CsgG–FCP pores, DNA experiments were performed using a 3.6-kilobase (kb) ssDNA section from the 3′ end of the lambda genome, linearized pTrc99a vectors with synthetic polyT inserts (see below) or genomic DNA from *E. coli*. For nanopore sequencing, DNA strands need to be ligated to the ONT adapter mix to optimize the capture and threading of the DNA strands into the pore. Unless stated otherwise, all components used for ligation and sequencing are provided in the ONT-SQK-LSK109 kit developed by ONT. Ligation was performed as per the instructions provided with the kit. Briefly, 1 µg of dA-tailed 3.6-kb lambda DNA was mixed with 40 nM adapter mix in the presence of 50 µl of NEB Blunt/TA Ligase Master Mix (NEB). The reaction was incubated for 10 min at room temperature. The ligation mixture was purified from unligated/free adapter using Agencourt AMPure XP Spri beads (Beckman Coulter). 0.4× (v/v) Spri beads were added to the ligation mixture and incubated for a further 10 min. The beads were then washed twice using 250 µl of short fragment buffer for 10 min each. The DNA was then eluted from the beads using 25 µl of elution buffer. The final sequencing library mixture for each flow cell was prepared by adding 37.5 µl of sequencing buffer, 25.5 µl of library loading beads and 12 µl of eluted DNA. Before loading the sequencing mix, the flow cell was initially flushed with 800 µl of priming buffer through the priming port. The SpotON port cover was opened and an additional 200 µl of priming buffer was flushed through the priming port. Finally, 75 µl of sequencing library mix was added to the flow cell using the SpotON port. All recordings were conducted at −180 mV and data acquisition and analysis were performed using MinKNOW.

The DNA squiggles shown in Fig. 3b were generated using *E. coli* genomic DNA and data were plotted using ONT in-house software. Figure 4a,b was generated with a 3.6-kb ssDNA section from the 3′ end of the lambda genome. This strand contains three stretches of ten deoxythymidine nucleotides, spaced by GGAA and flanked by mixed sequence preceding and following the homopolymer regions. To test homopolymer calling accuracy (Fig. 4d), a series of oligonucleotides with three stretches of deoxythymidine nucleotides, ranging from 3 to 9 in length and spaced by GGAA, were cloned into the pTrc99a vector using XmaI and HindIII restriction (called 3x_T3 to 3x_T9, Supplementary Table 1). Before sequencing, the modified plasmids were linearized by PCR using primers 9 and 10 (Supplementary Table 1). The different constructs were confirmed by Sanger sequencing and run in different parallel MinION sequencing runs. Data acquisition and analysis were performed using MinKNOW and base calling was performed locally using Guppy v2.3.5 software. Per construct, the length of the polyT insert as called in the unpolished single reads was plotted in histograms, using at least 166,000 single reads. *E. coli* genomic DNA was used as a reference for comparing the homopolymer base calling of $CsgG^{R9}$ and $CsgG^{R9}$–FCP across the four bases and in the context of random flanking sequence (Fig. 4e and Extended Data Fig. 4b). Data acquisition and analysis were performed as described above. The pile up of base calls in Fig. 4d was plotted using the Integrative Genomics Viewer software[31].

Preparation and testing of DNA–biotin–streptavidin static strand complexes on MinION. A set of polyA DNA strands (SS20 to SS38; Fig. 3c) in which one base is missing from the DNA backbone (iSpc3) was obtained by Integrated DNA Technologies. The 3′ end of each of these strands comprises a biotin modification. The DNA strands were incubated with monovalent streptavidin in a 1:1 ratio (9.9 µM final concentration) at room temperature for 20 min, resulting in the DNA–biotin–streptavidin static strand complexes for each polyA DNA strand. The complexes were diluted to 2 µM with priming buffer. MinION flow cells with CsgG or CsgG–FCP pores were prepared as described above. An 800-µl volume of priming buffer was flushed into the flow cell via the priming port in preparation for the static strands. The script used for the static strand experiment was run at −200 mV for CsgG–FCP pores and −180 mV for CsgG pores with a reverse potential flick every minute (0 mV for 2 s, 100 mV for 2 s and then 0 mV for 2 s). A 75-µl volume of each DNA–biotin–streptavidin complex was added to the flow cell sequentially via the SpotON port and data were recorded for 15 min for each complex. An 800-µl volume of priming buffer was flushed via the priming port in between each addition of the complex to make sure the first DNA–biotin–streptavidin complex is removed before the addition of the next. This process was repeated for all static strands. Once the final static strand complex had been incubated on the flow cell, 800 µl of priming buffer was flushed via the priming port and 10 min of open pore recording was generated before finishing the experiment.

The median block current level for each static strand complex, SS20–SS38, was measured by filtering the data to show only 'block events' over time. Block events are defined as being longer than 5 s, less than 60 s in duration and less than 150 pA median current for $CsgG^{F56Q}$ or less than 80 pA for $CsgG^{F56Q}$–FCP. A scatter plot was generated for both $CsgG^{F56Q}$ and $CsgG^{F56Q}$–FCP by plotting each data point over time (seconds), where each data point represents a single block event normalized against the open-pore current of that event.

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

Data availability

Coordinates and the electron potential maps for the CsgG–CsgF cryo-EM structure have been deposited in the PDB and EMDB under accession codes 6SI7 and EMD-10206, respectively. R9 pores are proprietary mutants of *E. coli* CsgG developed by ONT and are available as membrane-embedded single pores incorporated in Flongle, MinION, GridION and PromethION flow cells.

Code availability

The ONT software packages MinKNOW and Guppy are available through https://community.nanoporetech.com/downloads, and medaka is available through https://github.com/nanoporetech/medaka.

References

33. Miroux, B. & Walker, J. E. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289–298 (1996).
34. Casadaban, M. J. Transposition and fusion of the *lac* genes to selected promoters in *Escherichia coli* using bacteriophage lambda and Mu. *J. Mol. Biol.* **104**, 541–555 (1976).
35. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
36. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
37. Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *Elife* **5**, e18722 (2016).
38. Reboul, C. F., Eager, M., Elmlund, D. & Elmlund, H. Single-particle cryo-EM-improved ab initio 3D reconstruction with SIMPLE/PRIME. *Protein Sci.* **27**, 51–61 (2018).
39. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
40. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018).
41. Smart, O. S., Neduvelil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **14**, 354–360 (1996).
42. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2017); https://www.R-project.org/

Acknowledgements

We are grateful to R. Thompson and J. van Rooyen for assistance during cryo-EM data collection on Titan Krios 1 at the Astbury Biostructure Laboratory, Leeds and Krios m02 at Diamond - eBIC, Harwell Science and Innovation Campus, UK, respectively. We thank R. Efremov for advice on cryo-EM image processing, and are grateful to S. Young at ONT for helpful discussion and advice on MinION data analysis. This work received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme under grant agreement no. 649082 (BAS-SBBT). S.E.V.d.V. is a recipient of a PhD fellowship of the Flanders Research Foundation (FWO).

Author contributions

S.E.V.d.V. produced and characterized CsgG–CsgF complexes and determined their cryo-EM structure, supervised by H.R. N.V.G. and W.J. produced and analysed CsgG, CsgF and CsgG–FCP constructs and protein. P.S., R.H., J.K., M.J. and E.J.W. produced CsgG–FCP pores, and recorded and analysed electrophysiology data. L.J. and H.R. supervised the study and analysed data. S.E.V.d.V., E.J.W., L.J. and H.R. wrote the paper, with contributions of all authors.

Competing interests

VIB and ONT have jointly filed two provisional patent applications on the construction and use of dual-constriction pores in nanopore sensing applications (PCT/GB2018/051858 and PCT/GB2018/051191). VIB has a funded research collaboration agreement with ONT related to CsgG-derived nanopores. ONT uses CsgG-derived nanopores in its MinION, GridION and PromethION nanopore sequencing devices. As inventors on VIB intellectual property, S.E.V.d.V., N.V.G. and H.R. receive a share in royalty payments. R.H., P.S., J.K., M.J., E.J.W. and L.J. are employees of ONT and own company share options.

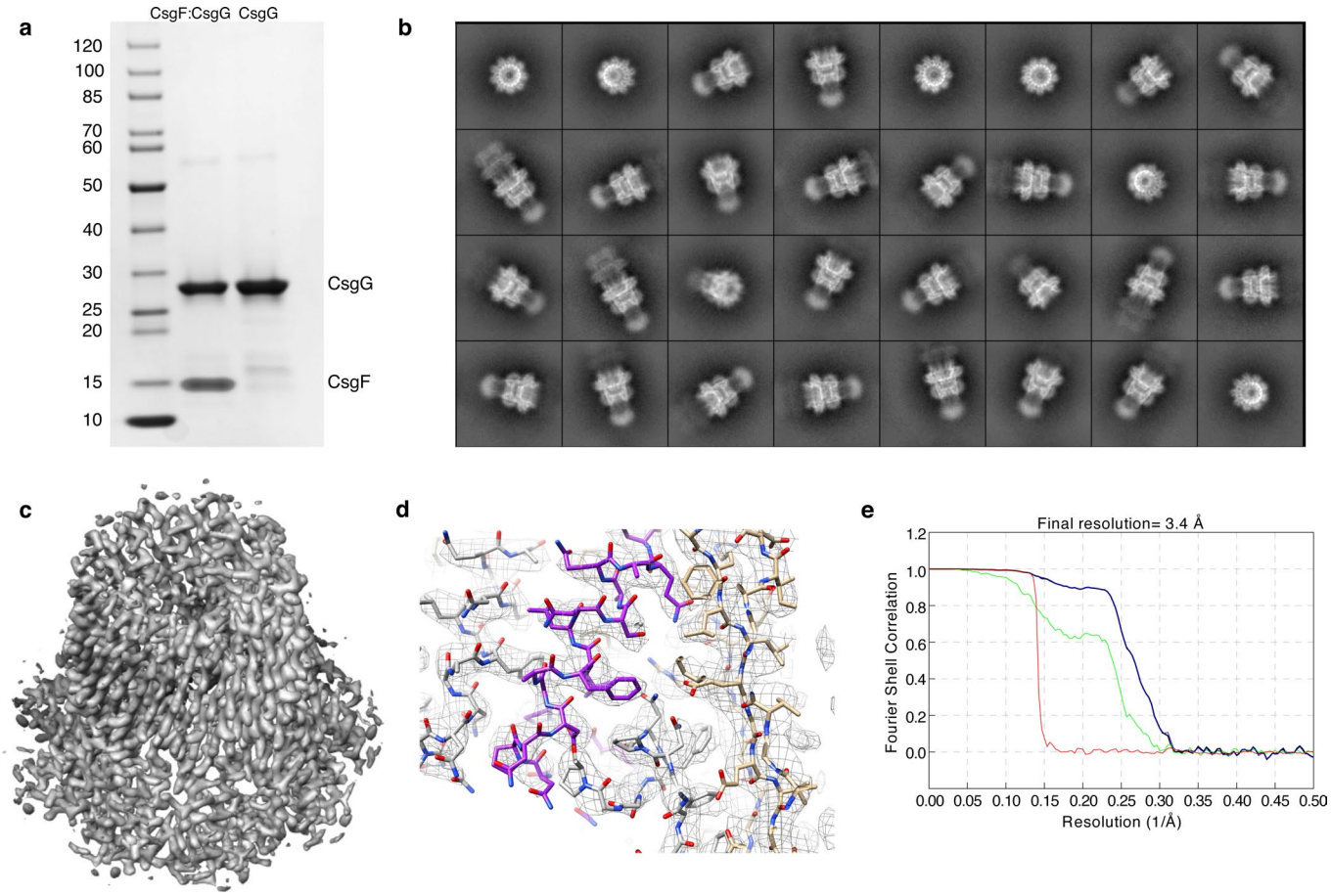
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41587-020-0570-8.

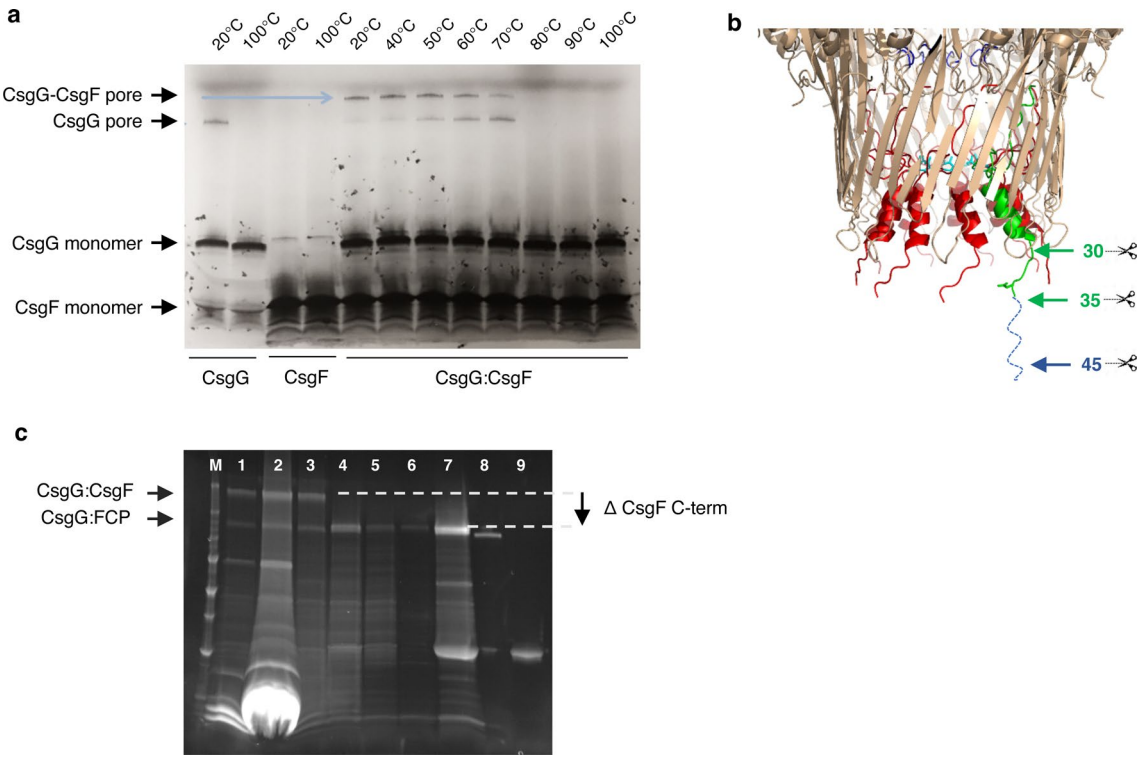
Supplementary information is available for this paper at https://doi.org/10.1038/s41587-020-0570-8.

Correspondence and requests for materials should be addressed to H.R.

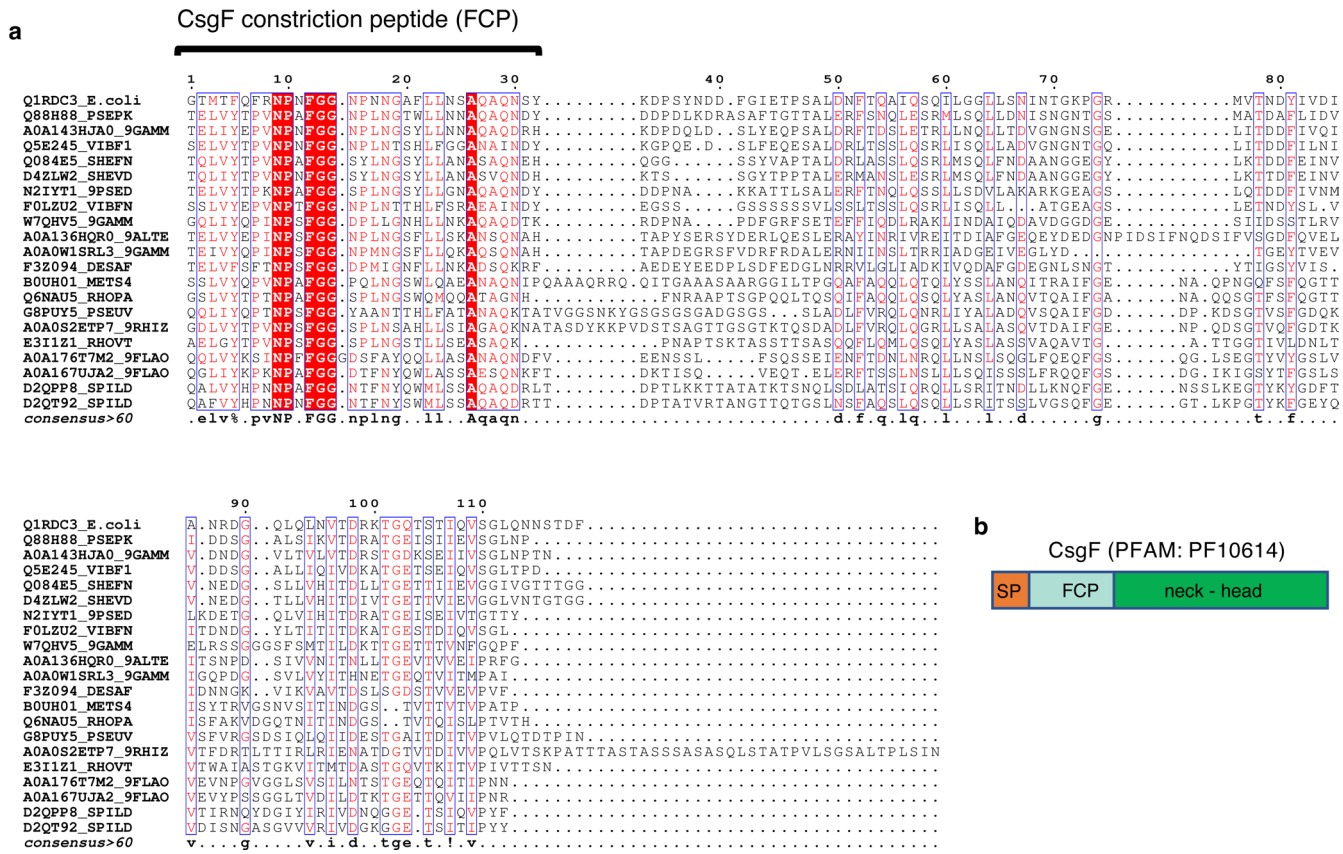
Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | Electron cryo-microscopy of the CsgG:CsgF complex. a, SDS PAGE of CsgG or CsgG:CsgF complex obtained by tandem affinity purification of the outer membrane proteins extracted from cells expressing CsgG-Strep II (pPG1) or CsgG-Strep II and CsgF-His (pNA62), respectively. Gel representative for n>10 experiments. **b**, Representative 2D class averages for the CsgG:CsgF dataset enriched for single pores (that is C_9 CsgG:CsgF complexes), generated using SIMPLE and used for 3D reconstruction using Relion-2.0. **c**, Off-axis top view and cross-sectional side view of the CsgG:CsgF cryo-EM 3D electron potential map reconstructed to 3.4 Å. **d**, Representative region of electron potential map of the CsgG:CsgF complex. Region of focus is the constriction helix of FCP, stacking against the lumen of the CsgG β-barrel. One CsgF protomer is highlighted in purple, the others in grey; CsgG is depicted in gold. Heteroatoms are in blue (nitrogens) and red (oxygens). The electron potential map is cut-off at a contour of 0.5, shown in stick and mesh representation, rendered using UCSF Chimera 1.10.2. (**e**) Fourier Shell Correlation (FSC) curves of the final 3D reconstruction (black: FSC corrected map, green: FSC unmasked map, blue: FSC masked map, red: FSC phase randomized unmasked map).

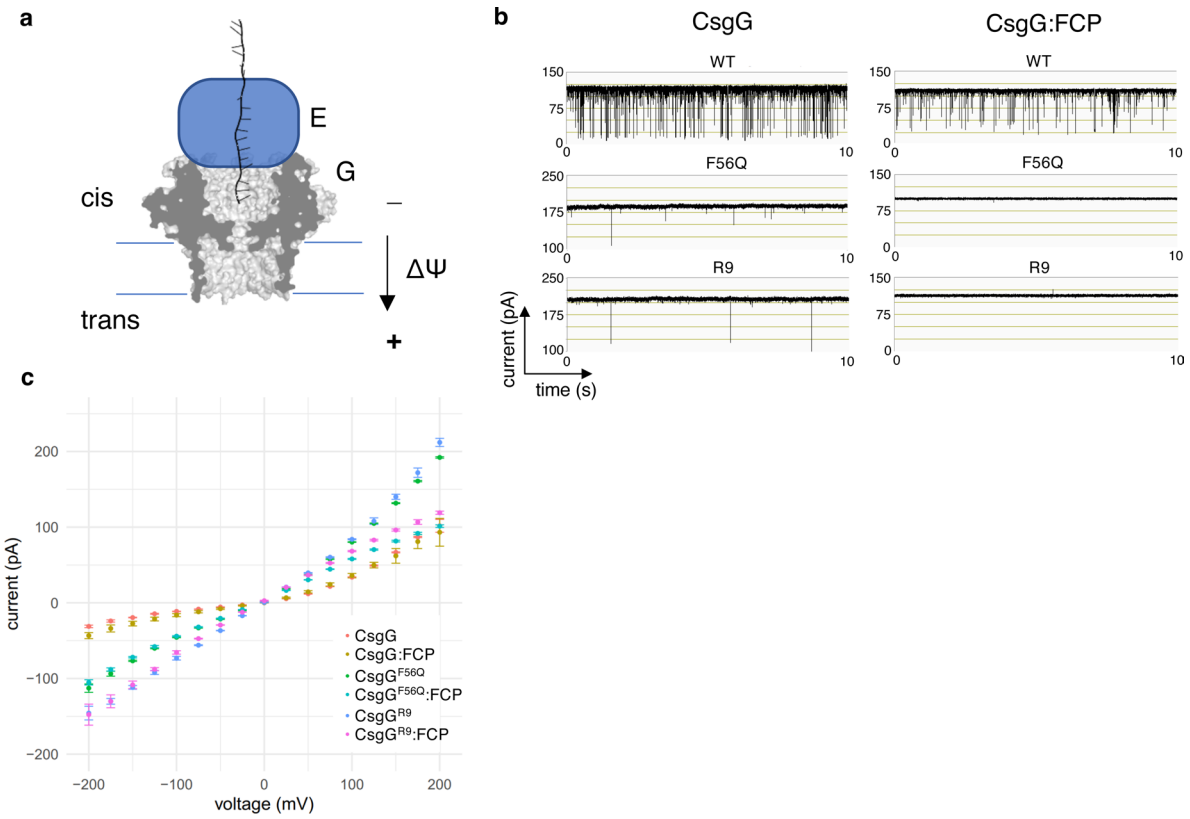


Extended Data Fig. 2 | Production and thermal stability of CsgG:CsgF and CsgG:FCP pores. a, Production and temperature stability assessment of the CsgG:CsgF pore complex. Incubation of purified CsgG and CsgF in a 1:1 ratio results in the formation of a SDS stable CsgG:CsgF pore complex that is heat stable up to 70 °C. **b**, The N-terminal residues of CsgF insert into the CsgG channel and form a second region of constriction, whilst the remaining ~100 residues form a cap like head structure (Fig. 1e, f). For nanopore sensing purposes, we sought to produce a complex of CsgG with the CsgF constriction peptide (FCP), lacking the neck and head domains. To do so, CsgG was complexed with CsgF mutants modified to insert a TEV cleavage site at position 30, 35 or 45. The reconstituted CsgG:CsgF pore complexes were digested with TEV protease and analysed by SDS-PAGE (**c**). M: molecular mass marker, Lane 1, 2: Strep II-tag affinity purified CsgG:CsgF complex and excess CsgG, Lane 3: isolation of CsgG:CsgF complex by size exclusion chromatography, Lane 4: CsgG:CsgF$^{35\text{-TEV}}$ cleaved with TEV protease to generate CsgG:FCP complex, Lane 5: flow through of CsgG:FCP after Strep purification, Lane 6: CsgG:FCP heated to 60 °C for 10 minutes. Lane 7: Eluted CsgG:FCP complex from Strep column, Lane 8: CsgG pore as the control, Lane 9: TEV protease as the control.

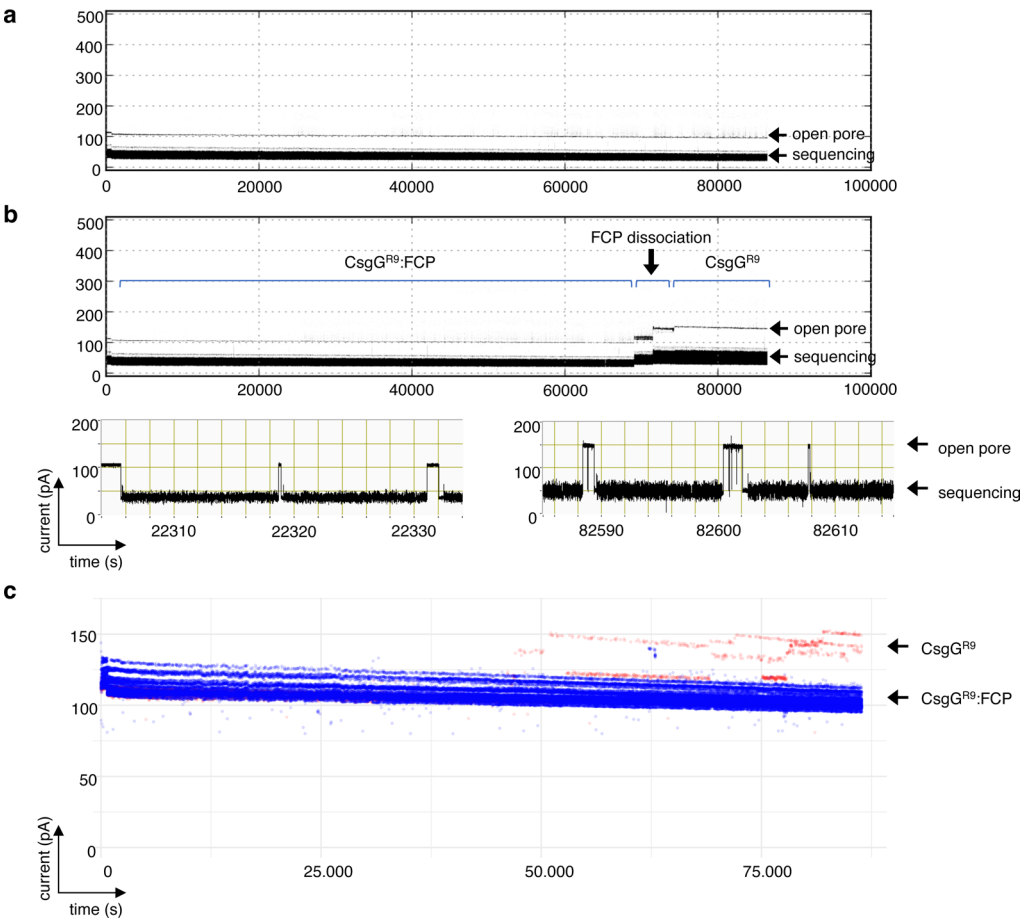


Extended Data Fig. 3 | Multiple sequence alignment of CsgF-homologues. a, Multiple sequence alignment (Multalin[43]) of 22 representative CsgF sequences. Aligned sequences are shown as mature proteins (that is lacking their N-terminal signal peptide). The N-terminal 33 residues of the mature protein form a continuous stretch of high sequence conservation (48% average pairwise sequence identity) encompassing the region interacting with CsgG and forming the CsgF constriction peptide. CsgF homologues included in the multiple sequence alignment are UniProt entries Q88H88; A0A143HJA0; Q5E245; Q084E5; F0LZU2; A0A136HQR0; A0A0W1SRL3; B0UH01; Q6NAU5; G8PUY5; A0A0S2ETP7; E3I1Z1; F3Z094; A0A176T7M2; D2QPP8; N2IYT1; W7QHV5; D4ZLW2; D2QT92; A0A167UJA2. **b**, Schematic diagram of CsgF protein architecture. (SP) signal peptide, cleaved upon secretion; (FCP) CsgF constriction peptide, CsgF neck and head region are coloured green.

43. Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16**, 10881–10890 (1988).



Extended Data Fig. 4 | Sequencing setup and channel characteristics of CsgG and CsgG:FCP nanopores. a, Schematic representation of the electrophysiology setup of CsgG-based nanopores as used for polynucleotide sequencing. CsgG-based channels (G) are reconstituted into artificial membranes with the periplasmic vestibule and β-barrel exposed to the cis and trans sides, respectively. Polynucleotide – enzyme (E) complexes are added to the cis side and current reads are recorded under an electric potential (Δψ) of 100 to 300 mV. **b**, **c**, Representative single channel traces (b) and current - voltage (IV) curves (c) for wildtype CsgG, CsgG^{F56Q} and CsgG^{R9} and their FCP complexes: CsgG:FCP, CsgG^{F56Q}:FCP and CsgG^{R9}:FCP. I-V curves show mean ± 95% confidence interval of at least 60 single channels per pore, with the exception of wildtype CsgG (36 single channels) and CsgG:FCP (14 single channels).

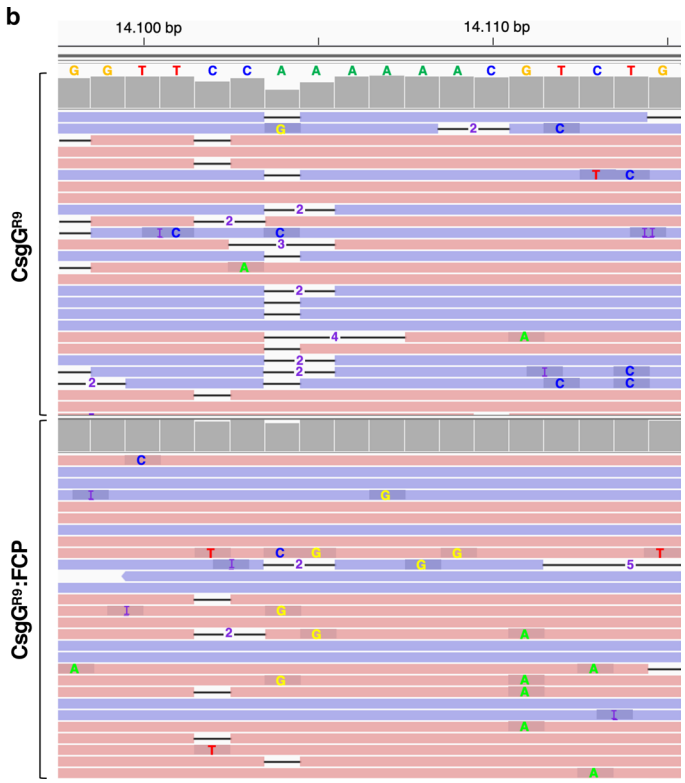


Extended Data Fig. 5 | Single channel stability of CsgG^{R9}:FCP complexes. a, **b**, Single channel conductance trace of two representative CsgG^{R9}:FCP nanopores during a 24 hour sequencing run, recorded at -180 mV. The data show both CsgG^{R9} and CsgG^{R9}:FCP are predominantly in a sequencing, DNA-occupied state, with apo pores capturing new DNA strands within seconds. The two traces show a CsgG^{R9}:FCP pore complex that stays intact of the 24h sequencing run (a), as well as a pore complex that shows dissociation of the FCP peptides during the sequencing run (at ~ 19h; b). Upon FCP dissociation, the channel continues sequencing now as a CsgG^{R9} apo pore (labeled CsgG^{R9}). Arrows indicate the average conductance levels of the open pore and the DNA-occupied pore during sequencing intervals. The zoomed in panels show two representative 30s time windows of the sequencing run of the intact CsgG^{R9}:FCP channel (left) and the CsgG^{R9} channel following dissociation of FCP (right). The full and zoomed in sequencing runs show high DNA capture rates for CsgG^{R9}:FCP channels throughout the 24h sequencing run. **c**, Scatter plot of the open pore current of 25 CsgG^{R9}:FCP channels during 24h sequencing runs, recorded at -180 mV. Open pore plots for CsgG^{R9}:FCP pores that stay intact throughout the 24h run (n=22), and pores that lose FCP (n=3) are coloured blue and red, respectively.

a

Strand	Sequence	C3 Location
SS20	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA/3Bio/	0
SS21	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA/iSpC3/A/3Bio/	2
SS22	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA/iSpC3/AA/3Bio/	3
SS23	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA/iSpC3/AAA/3Bio/	4
SS24	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA/iSpC3/AAAA/3Bio/	5
SS25	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA/iSpC3/AAAAA/3Bio/	6
SS26	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA/iSpC3/AAAAAA/3Bio/	7
SS27	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA/iSpC3/AAAAAAA/3Bio/	8
SS28	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA/iSpC3/AAAAAAAA/3Bio/	9
SS29	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA/iSpC3/AAAAAAAAA/3Bio/	10
SS30	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA/iSpC3/AAAAAAAAAA/3Bio/	11
SS31	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA/iSpC3/AAAAAAAAAAA/3Bio/	12
SS32	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA/iSpC3/AAAAAAAAAAA/3Bio/	13
SS33	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA/iSpC3/AAAAAAAAAAAA/3Bio/	14
SS34	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA/iSpC3/AAAAAAAAAAAA/3Bio/	15
SS35	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA/iSpC3/AAAAAAAAAAAAA/3Bio/	16
SS36	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA/iSpC3/AAAAAAAAAAAAAA/3Bio/	17
SS37	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA/iSpC3/AAAAAAAAAAAAAA/3Bio/	18
SS38	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA/iSpC3/AAAAAAAAAAAAAAA/3Bio/	19

b



Extended Data Fig. 6 | Constriction mapping oligos and single read basecalls for CsgG^{R9} and CsgG^{R9}:FCP nanopores. a, Set of static polyA ssDNA oligonucleotides in which one base is missing from the DNA backbone (iSpc3). These oligos that have differing location of the abasic nucleotide, dubbed SS20 to SS38, were used to map the constriction position in CsgG^{F56Q} or CsgG^{F56Q}:FCP (Fig. 3d). Biotin modification at the 3′ end of each strand is complexed with monovalent streptavidin to block translocation of the oligo and give a defined distance marker between the pore entrance (block site) and pore constriction (site of increased conductance when occupied by the abasic nucleotide; Fig. 3c). SS27-SS28 and SS32 (highlighted red) have their abasic nucleotide located at the CsgG and FCP constriction, respectively (Fig. 3d, e). **b**, Comparison of errors in single read (n=26) basecalls from CsgG^{R9} and CsgG^{R9}:FCP pores that have been aligned to a representative region of the *E. coli* reference genome sequence. The region displayed corresponds to the locus 14,098 to 14,115. The figure is plotted using the Integrative Genomics Viewer software[31]. Pink/purple bars correspond to single reads in the forward and reverse directions respectively. Black horizontal bars correspond to deletions in the basecalls, where the number corresponds to the number of deletions at the specific loci. Individual substitutions are labeled with the miscalled nucleotide (C in blue, T in red, G in orange and A in green). Insertions are labeled 'I' (purple). Grey bars on top of the list of single reads of the CsgG^{R9} and CsgG^{R9}:FCP pores correspond to the consensus accuracy per position.

nature research

Corresponding author(s): Remaut Han

Last updated by author(s): Apr 29, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	MinKNOW core 1.11.5
Data analysis	Gautomatch, CtfFind 4.1.8, Relion2.0, Simple3.0, Eman2, UCSF Chimera 1.10.2, Pymol 2.0, Coot 0.8.9, Phenix 1.14, MotionCorr 2.1, MS Excell 2013, HOLE, R, Integrative Genomics Viewer, Guppy v2.3.5

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Coordinates and the electron density maps for the CsgG:CsgF cryo-EM structure have been deposited in the PDB and EMDB under accession codes 6SI7 and EMD-10206, respectively. R9 pores are proprietary mutants of E. coli CsgG developed by ONT and are available as membrane embedded single pores incorporated in Flongle, MinION, GridION and PromethION flowcells.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All current recordings are taken from multiple pores, and represent multiple events per pore. Numbers are reported in figure legends and methods.
Data exclusions	MinION asic chip groups and positions with single pores were evaluated using the standard flow cell check protocol using MinKNOW. Only positions with stable single pore currents are selected for data acquisition.
Replication	Pore conductivity profiles are representative multiple single pores for the different mutants and complexes tested, which each pore recording multiple DNA capturing events.
Randomization	not applicable
Blinding	not applicable

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging