Supplementary Information

SI Text

Supplementary Experimental Procedures

Cloning, expression and purification of chaperone-subunit complexes

PapD\textsuperscript{his}PapG\textsubscript{II} (1-3), PapG\textsubscript{ad} (4), PapDPapF\textsubscript{G8N} (3), PapD\textsuperscript{his}PapE\textsubscript{Ntd} (5), PapD\textsuperscript{his}PapK (5), PapD\textsuperscript{his}PapA\textsubscript{Ntd1G15N} (6) and PapD\textsuperscript{his}PapH\textsubscript{Ntd} (6) were cloned, expressed and purified as previously reported. A plasmid encoding PapD\textsuperscript{his}PapG\textsubscript{p} (PapG\textsubscript{p}; residues 214-336) was transformed into \textit{E.coli} C600 cells and incubated in Luria-Bertani medium at 37 °C with antibiotic (100 μg.ml\textsuperscript{-1} ampicillin) selection. At an OD\textsubscript{600} of 0.6, protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and the cells left to grow for a further 3.5 h before harvesting. The periplasm was then extracted using a sucrose/lysozyme method (7). PapD\textsuperscript{his}PapG\textsubscript{p} was purified using nickel and hydrophobic interaction chromatography (HisTrap HP and Phenyl HP, GE Healthcare, Buckinghamshire, UK), stored in 20 mM TrisHCl pH 8.0, 150 mM NaCl, 20% (v/v) glycerol at -20 °C and dialysed into 5 mM ammonium acetate pH 6.8 immediately before use.

Cloning, expression and purification of PapC usher domains

The PapC usher domains were cloned in pASK-IBA vector series (IBATagnology). The PapC UsherN domain (residues 1-131) was purified from the full-length PapC sequence, yet degrades rapidly to UsherN\textsubscript{4-131}, the latter which was used in this study. UsherN\textsubscript{24-131} and UsherP were cloned into vector pASK-IBA2 (IBATagnology). By design, this vector adds the signal peptide of the OmpA protein at the N-terminus of the construct ensuring its delivery to the periplasm, while adding a Strep-Tactin tag at the C-terminus of the constructs.
The PapC UsherC2 domain (residues 714-809) was cloned, expressed and purified as previously reported (8).

All proteins were expressed from the clones described above introduced into C600 cells. 6 L of Luria Broth media with 100 μg.mL⁻¹ ampicillin were used to express the domains and cells were grown at 37 °C until the OD reached 0.7 when protein expression was induced with 2 μg.mL⁻¹ anhydrotetracycline (AHT). Cells were then cultured for a further 3 h at 37° C. Purification from the periplasm proceeded as previously described (9). UsherN, UsherN_{24-131} and UsherP were purified on a Strep-Tactin Sepharose column, followed by an anionic exchange chromatography (HiTrap Q HP; GE Healthcare). For the UsherC2 domain a C-terminal His-tag was added to aid purification (8).

**Analysis of different usher domain constructs by ESI-MS**

The molecular masses of the soluble usher domains were measured to be 15,487 Da (15,487 Da), 8,665 Da (8,664 Da) and 10,704 Da (10,704 Da) for UsherN, UsherP and UsherC2, respectively (Fig. S2), which are within 0.01% error of the masses expected (shown in parenthesis). In addition, a truncated version of UsherN (residues 24-131) was created (with measured mass, 12,827 Da (expected mass, 12,830 Da) in which the N-terminal 23 residues shown previously to be important for chaperone-subunit binding (10) were deleted. The ESI-MS charge state distribution of UsherN reveals the presence of at least two distinct conformations, (Fig. S2A) but mainly one conformation is observed for the truncated form of this domain (UsherN_{24-131}, Fig. S2B). It has been suggested the N-terminal residues of UsherN are unstructured when not bound to a chaperone:subunit complex, possibly explaining the multiple conformations accessible to the full-length UsherN domain (11).
single, compact charge state distribution is dominant for the UsherP and UsherC2 domains, consistent with the known globular structures of these domains (Fig. S2C, D) (8, 12).

**Binding of chaperone-subunits to usher domains is specific.**

A series of control experiments was performed to determine whether the interaction between UsherN, UsherP and UsherC2 and the different PapD-PapX complexes is specific. For these experiments each usher domain was mixed with an equimolar concentration of two other domains with an Ig fold (I27(C47A C63A)) or β2-microglobulin (expressed and purified as described in (13) and (14)). Neither of these two proteins revealed evidence for binding to the soluble usher domains under the experimental conditions used (Fig. S4). In addition, no binding was observed between the soluble three usher domains and free PapD or free PapX, which are also present in the samples. Further control experiments revealed the specificity of binding of PapD-PapG to UsherN. In these experiments either PapD-PapG alone or PapD-PapGUsherN was formed and then analyzed by ESI-MS either alone or after addition of an equimolar concentration of the C-terminal domain 2 of FimD (FimD UsherC2; residues 750-833) or hen egg lysozyme as a decoy protein. The relative contribution of ions arising from PapD-PapGUsherN was then determined. The results revealed no dissociation of the PapD-PapGUsherN complex occurred when either decoy protein was added (Fig. S9).
**Supplementary Figures and Legends**

**Figure S1:** Topology representations of Donor Strand Complementation (DSC) and Donor Strand Exchange (DSE). (A) In DSC the chaperone (brown) donates its G₁ strand to the subunit (green) forming a chaperone-subunit complex. (B) DSE occurs when the N-terminal extension (Nte) of the subunit next in assembly displaces this donated G₁ chaperone strand. The resulting complex is shown where the Nte of one subunit (blue) is inserted to displace the chaperone and complete the Ig fold of an adjacent subunit (green). (C) Schematic of the 7 chaperone-subunit complexes used in this study and the PapG_adhesin domain (Ga) (adapted from (15)). The chaperone, PapD (brown), with oval subunit pilin domains PapG (red), PapF (orange), PapE (yellow), PapK (green), PapA (cyan) and PapH (dark blue) are displayed.

**Figure S2:** ESI-MS spectra of different usher domain constructs. Spectra are shown of (A) UsherN, 15,487 Da; (B) UsherN<sub>24-131</sub>, 12,827Da (C) UsherP, residues 253-335, 8,665 Da and (D), UsherC2, residues 723-809, 10,704 Da.

**Figure S3:** ESI-MS spectra of the different chaperone-subunit-usher complexes. Spectra are shown for complexes of different chaperone subunits with UsherN (left); UsherP (middle) or UsherC2 (right). In each spectrum PapDPapX is shown in blue and PapDPapXUsherY ternary complexes are shown in red. In each row the results of mixing different chaperone-subunit complexes with each usher domain are shown: PapDPapG (A); PapDPapF (B); PapDPapE (C); PapDPapK (D); PapDPapA (E); and PapDPapH (F). Note that any unbound UsherN, UsherC2 and UsherP present are not shown on this m/z scale.

**Figure S4:** ESI-MS spectra of Usher domains mixed at equimolar concentrations. (A) UsherP (red) + Usher C (blue); (B) UsherN (green) + Usher P (red); (C) UsherN (green) +
Usher C (blue). Insets: bar charts representing the relative total ion intensity of each usher domain within the mixture. Error bars represent the SEM of several replicate measurements.

**Figure S5:** ESI-MS spectra of UsherN domains mixed in a 2-fold molar excess over the concentration of the chaperone-subunit complex. (Upper) PapDPapA (DA, green), PapDPapAUsherN (DAN, purple); (Lower) PapDPapG (DG, cyan), PapDPapGUAsherN (DGN, red). Bar charts represent the relative ion intensity of the chaperone-subunit and chaperone-subunit-usher complex. Error bars represent the SEM of several replicate measurements.

**Figure S6:** ESI-MS spectra of control experiments adding other proteins to different soluble usher domains. (A) β2m or (B) I27 (B) (blue peaks) were each mixed separately with the three soluble usher domains, UsherN (upper), UsherP (middle) or UsherC2 (lower) (red peaks). Peaks not labelled represent fragmented usher domains and control proteins. No UsherY binding was observed to either of the control proteins.

**Figure S7:** Binding of different chaperone-subunit complexes to the truncated domain UsherN_{24-131}. (A) ESI-MS spectra of the chaperone-subunit-usher ternary complexes (red). Unbound PapDPapX is shown in blue. (B) Bar graphs of the normalized collision energy required to fragment each ternary ion to 50% of its initial ion count. Collision energies for complexes with full length UsherN are shown in solid fill and those with UsherN_{24-131} are shown in hatched fill. Error bars represent the standard error of the mean over four replicate measurements. Peaks not labelled represent fragmented usher domains and protein complexes.
**Figure S8:** Binding of the adhesion domain of PapG to UsherN. ESI-MS spectra of the adhesion domain of PapG bound to UsherN (red), together with free PapG$_{ad}$ ions (blue peaks) and free UsherN ions (green peaks). Peaks not labelled represent fragmented usher domains and protein complexes.

**Figure S9:** ESI-MS spectra demonstrate the specificity of the interaction of PapDPapG with UsherN. (A) ESI-MS spectra of PapDPapGUsherN incubated without (upper) or with (lower) a two-fold molar excess of the C-terminal domain 2 of the FimD usher. Peaks arising from PapDPapGUsherN (red), PapDPapG (cyan) and the dimer of PapDPapG (dark blue) are highlighted. The C-terminal domain of Fim usher is not detected within this $m/z$ range. No evidence of PapDPapG binding to the C-terminal domain 2 of the FimD usher was observed. (B) Histogram showing the relative ion intensities of the PapDPapGUsherN ternary complex (red) compared with PapDPapG alone (blue) before (Ø) or after addition of a two-fold molar excess of either the Fim Usher C-terminal domain 2, or hen egg white lysozyme. Errors are the standard error of the mean over four replicate measurements.

**Figure S10:** Schematic of the comparison of subunits comprising a P (left) and Type 1 (right) pilus. Organization of the pili, which includes one copy each of PapG/FimH (red), PapF/FimG (orange), PapK/FimF (green) and PapH (dark blue), five to ten copies of PapE (yellow) and up to one thousand copies of Pap/FimA (cyan). The chaperone, PapD/FimC (brown), PapC/FimD usher (pink), periplasm (P), extracellular space (Ex) and bacterial outer membrane (OM) are indicated.
Table S1: The expected and observed masses (Da) of the binary PapDPapX and ternary PapDPapXUsherY complexes shown in Figure S3. All masses observed are within 0.03 % error of the calculated masses.
Supplementary References


Figure S2

A

100%

% 12+ 11+ 10+ 9+ 8+ 7+ 6+

UsherN 15487 Da

B

100%

% 10+ 9+ 8+ 7+ 6+ 5+

UsherN_{24-131} 12827 Da

C

100%

% 7+ 6+ 5+ 4+

UsherP 8665 Da

D

100%

% 9+ 8+ 7+ 6+ 5+

UsherC2 10704 Da

m/z 1200 1900 2700
Figure S5

[Graph showing mass spectrometry data with m/z values ranging from 2900 to 4800. Peaks are labeled with relative intensity (%) and charge states (+). Molecules are identified with molecular weights (Da).]
Figure S9

(A) Mass spectra of PapDPapG+UsherN and PapDPapG+UsherN+FimUsherC2. Mass values:
- PapDPapG 61013 Da
- PapDPapG\textsubscript{dimer} 122021 Da
- PapDPapGUsherN 76510 Da

(B) Bar chart showing relative total ion intensity (%) for +Ø, +FimDUsherC2, and +Lysozyme conditions.
<table>
<thead>
<tr>
<th>Complex</th>
<th>Estimated Mass (Da)</th>
<th>Observed Mass (Da)</th>
<th>% difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>UsherN</td>
<td>15487</td>
<td>15487 ± 2</td>
<td>0</td>
</tr>
<tr>
<td>UsherC2</td>
<td>10704</td>
<td>10704 ± 1</td>
<td>0</td>
</tr>
<tr>
<td>UsherP</td>
<td>8664</td>
<td>8665 ± 1</td>
<td>0.03</td>
</tr>
<tr>
<td>PapDPapG</td>
<td>60994</td>
<td>61013 ± 4</td>
<td>0.01</td>
</tr>
<tr>
<td>PapDPapGUsherN</td>
<td>76481</td>
<td>76500 ± 10</td>
<td>0.02</td>
</tr>
<tr>
<td>PapDPapGUsherC2</td>
<td>71698</td>
<td>71711 ± 3</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>PapDPapGUsherP</td>
<td>69685</td>
<td>69666 ± 4</td>
<td>0.02</td>
</tr>
<tr>
<td>PapDPapF</td>
<td>40314</td>
<td>40311 ± 2</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>PapDPapFUsherN</td>
<td>55801</td>
<td>55797 ± 5</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>PapDPapFUsherC2</td>
<td>51018</td>
<td>51017 ± 2</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>PapDPapFUsherP</td>
<td>48978</td>
<td>48984 ± 1</td>
<td>0.01</td>
</tr>
<tr>
<td>PapDPapE</td>
<td>40063</td>
<td>40063 ± 3</td>
<td>0</td>
</tr>
<tr>
<td>PapDPapEUsherN</td>
<td>55550</td>
<td>55549 ± 1</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>PapDPapEUsherC2</td>
<td>50767</td>
<td>50776 ± 9</td>
<td>0.02</td>
</tr>
<tr>
<td>PapDPapEUsherP</td>
<td>48728</td>
<td>48737 ± 1</td>
<td>0.02</td>
</tr>
<tr>
<td>PapDPapK</td>
<td>42222</td>
<td>42222 ± 3</td>
<td>0</td>
</tr>
<tr>
<td>PapDPapKUsherN</td>
<td>57710</td>
<td>57709 ± 2</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>PapDPapKUsherC2</td>
<td>52927</td>
<td>52928 ± 3</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>PapDPapKUsherP</td>
<td>50886</td>
<td>50888 ± 5</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>PapDPapA</td>
<td>41288</td>
<td>41285 ± 3</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>PapDPapAUsherN</td>
<td>56776</td>
<td>56784 ± 4</td>
<td>0.01</td>
</tr>
<tr>
<td>PapDPapAUsherC2</td>
<td>51992</td>
<td>51983 ± 3</td>
<td>0.02</td>
</tr>
<tr>
<td>PapDPapAUsherP</td>
<td>49952</td>
<td>49947 ± 5</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>PapDPapH</td>
<td>42262</td>
<td>42260 ± 7</td>
<td>0.01</td>
</tr>
<tr>
<td>PapDPapHUsherN</td>
<td>57749</td>
<td>57750 ± 4</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>PapDPapHUsherC2</td>
<td>52966</td>
<td>52964 ± 7</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>PapDPapHUsherP</td>
<td>50926</td>
<td>50924 ± 3</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>