

28 **Abstract**

29 PapC ushers are outer-membrane proteins enabling assembly and secretion of P pili
30 in uropathogenic *E. coli*. Their translocation domain is a large β -barrel occluded by a
31 plug domain, which is displaced to allow the translocation of pilus subunits across the
32 membrane. Previous studies suggested that this gating mechanism is controlled by a
33 β -hairpin and an α -helix. To investigate the role of these elements in allosteric signal
34 communication we developed a method combining evolutionary and molecular
35 dynamics studies of the native translocation domain and mutants lacking the β -
36 hairpin and/or α -helix. Analysis of a hybrid residue interaction network suggests
37 distinct regions (residue “communities”) within the translocation domain (especially
38 around β 12- β 14) linking these elements, thereby modulating PapC gating. Antibiotic
39 sensitivity and electrophysiology experiments on a set of alanine-substitution mutants
40 confirmed functional roles for four of these communities. This study illuminates the
41 gating mechanism of PapC ushers and its importance in maintaining outer-
42 membrane permeability.

43 Introduction

44 Gram-negative pathogens commonly express a vast variety of complex surface
45 organelles that are involved in different cellular processes. One of these organelles,
46 known as pili (or fimbriae), forms a class of virulence factors involved in host cell
47 adhesion and recognition, invasion, cell mobility and biofilm formation. P pili from
48 uropathogenic *Escherichia coli* are specifically required for the colonization of the
49 human kidney epithelium, a critical event in the kidney infection process
50 (pyelonephritis) (Roberts et al., 1994). P pili are assembled on the bacterial outer
51 membrane (OM) via the chaperone/usher (CU) pathway (Thanassi et al., 1998),
52 which is often used as a model system to elucidate the mechanism of pilus
53 biogenesis (Waksman and Hultgren, 2009).

54 The biogenesis of pili via the CU pathway is a highly ordered process that comprises
55 sequential steps. The chaperone protein (PapD) brings the pilins to the bacterial OM
56 where they are assembled into a pilus at a transmembrane pore protein known as
57 the usher (PapC). The usher (~800 residues) is composed of 5 domains (Figure 1A):
58 a periplasmic N-terminal domain (NTD), an OM central translocation domain (TD)
59 that comprises a translocation pore domain (TP), interrupted by a conserved Ig-like
60 plug domain (PD), and two domains at the periplasmic C-terminal end (CTD1 and
61 CTD2) (Capitani et al., 2006, Geibel et al., 2013, Ng et al., 2004, Thanassi et al.,
62 2002, Phan et al., 2011). The structure of the apo TD (Figure 1B,C) consists of a 24-
63 stranded kidney-shaped β -barrel where the PD is inserted into the loop connecting
64 two β -strands ($\beta 6$ - $\beta 7$), occluding the luminal volume of the pore (Remaut et al., 2008,
65 Huang et al., 2009). In the activated form of another archetypal member of the usher
66 family, FimD, the PD is located outside the pore lumen in the periplasm, next to the
67 NTD (Phan et al., 2011, Geibel et al., 2013). In addition to the PD, there are two
68 secondary structure elements that uniquely characterize the large β -barrel structures
69 of the usher TD (Figure 1B). The first element is a β -hairpin that creates a large gap
70 in the side of the β -barrel, a feature unprecedented in previously known OM β -barrel
71 structures (Remaut et al., 2008). This element (located between strands $\beta 5$ and $\beta 6$ of
72 the barrel, Figure 1C) folds into the barrel lumen and constrains the PD laterally
73 inside the barrel pore. Mutants lacking the β -hairpin show an increased pore
74 permeability suggesting that the β -hairpin has a role in maintaining the PD in a
75 closed conformation (Volkan et al., 2013). The second element is an α -helix (located
76 on the loop between $\beta 13$ and $\beta 14$, Figure 1B), which caps the β -hairpin from the
77 extracellular side. Mutants lacking the α -helix, or in which the interface between the
78 helix and the PD is disrupted, present a remarkable increase in pore permeability,

79 comparable with that of the mutant lacking the PD, suggesting a role for the helix in
80 maintaining the PD in a closed state (Mapingire et al., 2009, Volkan et al., 2013).

81 The mutant lacking both the β -hairpin and the α -helix is defective for pilus biogenesis
82 (Mapingire et al., 2009). It has been observed in other OMP β -barrels that such
83 secondary structure elements (e.g., an α -helix that protrudes inside the barrel or
84 packs against the transmembrane strands) can use complex allosteric mechanisms
85 to mediate their function (Naveed et al., 2009). These are often combinations of large
86 conformational changes (“global motions”) dictated by the overall architecture
87 (including movement of secondary structure elements) and smaller changes (“local
88 motions”, such as the motion of recognition loops and side-chain fluctuations) (Liu
89 and Bahar, 2012). Additionally, it has been shown that important residues in terms of
90 evolution (highly-coevolved or conserved) could have a pivotal role in mediating such
91 allosteric communications (Suel et al., 2003, Tang et al., 2007).

92 Here, to understand the allosteric mechanism leading to the plug displacement in
93 PapC and the involvement of the α -helix and β -hairpin, we used a hybrid
94 computational approach and verified our results experimentally. By combining
95 sequence conservation analysis, mutual information-based coevolution analysis, and
96 all-atom molecular dynamics (AA-MD), we modelled the interaction network within
97 the native PapC TD as well as within different mutants lacking the α -helix, β -hairpin,
98 and both. This unique computational approach allowed us to identify residues that
99 are likely to be involved in the transmission of the allosteric signal between the α -
100 helix, β -hairpin elements and the plug. These residues were investigated by site-
101 directed mutagenesis, functional studies and planar lipid bilayer electrophysiology.
102 The results confirmed the involvement of 4 of the 5 distinct communities in
103 modulation of the usher’s channel activity and gating, suggesting that they all
104 participate in the allosteric mechanism controlling plug displacement.

105

106 **Results**

107

108 **To investigate if the β -hairpin or α -helix (or both) of the TD (residues 146-637 in the**
109 **full length PapC) have a role in the allosteric communication leading to the**
110 **displacement of the PD (residues 264-324) we performed four independent MD**
111 **simulations, corresponding to the PapC TD model (sim1, Table 1) and three mutants**
112 **(sim2-sim4, Table 1) embedded in a mixed lipid bilayer: sim2 where the region**
113 **corresponding to the hairpin between β 5 and β 6 (residues 233-240) was deleted,**

114 sim3 where the α -helix between β 13 and β 14 (residues 447-460) was removed, and
115 sim4 where both regions were removed. The last 50 ns of simulation were
116 considered for analysis, where the averaged root-mean-square deviation of C_{α} atoms
117 (C_{α} -RMSD) from the averaged structures stabilized around 2.00 ± 0.09 Å, 1.80 ± 0.09
118 Å, 1.86 ± 0.11 Å, and 2.03 ± 0.10 Å, for the *native* (sim1), *hairpin mutant* (sim2), *helix*
119 *mutant* (sim3), and *helix-hairpin mutant* (sim4), respectively (Table 1–figure
120 supplement 1). This timescale, although limited for a full exploration of the structural
121 changes induced by the mutations, was informative in revealing how local structural
122 perturbations may affect allosteric changes leading to the plug displacement in PapC
123 TD.

124

125 *Non-Covalent Interaction Network In the Native PapC Translocation Domain and Its*
126 *Perturbation in the Absence of the β -Hairpin, α -Helix or Both.*

127

128 The changes in the non-covalent interactions (hydrogen bonds and salt bridges)
129 between all residue pairs were analysed within the native TD by calculating their *non-*
130 *covalent interaction score* (NCI score) (see Methods). A non-covalent residue-
131 residue interaction network (RIN) comprising 492 nodes (residues) and 1350 edges
132 (interactions) was then constructed as a weighted undirected graph for the native TD
133 (Figure 2) and the three mutant systems (Figure 2- figure supplement 1A-C), with the
134 weight for each edge given by the corresponding NCI score (Table 2). All four RINs
135 have properties typical of small-world networks (Haiyan and Jihua, 2009, Atilgan et
136 al., 2004, Taylor, 2013), with significant higher clustering coefficient compared to a
137 corresponding random network and a higher mean short path length (Table 2).
138 Within the constructed non-covalent native RIN we identified 246 weak-to-strong
139 interactions (connecting 362 nodes) with an NCI score of at least 0.3. Among these,
140 231 nodes connected by 133 edges showed an NCI score greater than 0.6 (*i.e.*,
141 strong interaction) of which 78 involve residues that are part of the barrel strands
142 (58.6%).

143 Comparative analysis between the RINs of native and mutants systems revealed
144 slight changes, suggesting a rearrangement in the interaction network. To better
145 understand the mutation-induced changes in network components we calculated the
146 difference in non-covalent interaction score (Δ NCI score) between the native TD
147 system and each of the mutant systems (The weakened interaction are shown in
148 Figure 2- figure supplement 2A-C). This information was then added as a weighted
149 undirected edge to the pre-existing native non-covalent RIN (the Δ NCI edges are

150 shown in Figure 2- figure supplement 2D). Interestingly, 24% of the strong
151 interactions in the native RIN were weakened relative to the RIN of the mutant
152 lacking the β -hairpin, 22.6% relative to the mutant lacking the α -helix and 23.3%
153 relative to the mutant lacking both, suggesting that interactions between nodes that
154 are not part of the deleted secondary structure elements were consistently weakened
155 in the absence of these elements.

156

157 *Evolutionary analysis of PapC TD*

158

159 We first extracted evolutionary information from a multiple sequence alignment of the
160 PapC TD family. The patterns of conservation in the TD using ConSurf (Ashkenazy et
161 al., 2010) analysis suggested that the highly conserved residues (score 9) tend to be
162 clustered in two specific regions of the usher (Figure 3A). The first cluster mapped
163 onto the PD and the P-linkers (P-linker1 residues 248-263; P-linker2 residues 325-
164 335) connecting it to the TP. The second cluster (which included the majority of the
165 highly-conserved residues) mapped onto one side of the TP (strand β 1-14 and β 24).
166 It includes residues: (i) near the periplasmic side of the β -barrel within β 1-4 strands
167 and β 24 strand; (ii) on the extracellular side of the barrel (within β 5-10); (iii) in the β -
168 hairpin region (β -hairpin and β 7-9); and (iv) in the area of β 10-14 capped by the α -
169 helix region, which comprises the α -helix and its linkers: H-linker1 (residues 445-450)
170 and H-linker2 (461-468, respectively). Surface representation of the TD reveals a
171 continuous patch of conserved residues facing the lipid bilayer, including β 13, the
172 extracellular half of β 14 and the periplasmic half of β 12 (Figure 3B). Intriguingly, this
173 patch (β 13 conserved patch) reaches the full height of the pore from the α -helix
174 region to a functionally important loop located between β 12 and β 13 strands (Volkan
175 et al., 2013, Farabella, 2013).

176 In addition to investigating conservation we performed an analysis to identify the
177 coevolutionary relationships between residues in the structure. Using normalized
178 mutual information (NMI) analysis (Martin et al., 2005) with a Z-score cut-off = 4 (see
179 Methods) to detect the intra-molecular coevolved residues within PapC TD, a
180 coevolutionary RIN containing 100 coevolved residues (nodes) and 357 connections
181 (edges) was derived (Figure 3D). Mapping the network onto the PapC TD structure
182 showed that many of the residues involved are also connected spatially and are
183 clustered in the same regions where the highly-conserved residues were found (P-
184 linkers, the PD, and the barrel wall capped by the α -helix, in close proximity to the β -
185 hairpin) (Figure 3C and D). The obtained coevolutionary RIN showed a significant

186 clustering coefficient compared to a corresponding random network (of 0.493 vs.
187 0.187, respectively) and a comparable mean short path length (3.15 vs. 2.57,
188 respectively) (Daily et al., 2008).

189

190 *Identifying Allosteric 'Hot Spots' From a Hybrid Residue Interaction Network*

191

192 We constructed one hybrid RIN in which the attributes for the nodes and edges are
193 defined by the properties described above (non-covalent networks and evolutionary
194 analysis, see Methods). Starting from the secondary structure elements (that
195 uniquely characterise the barrel - the α -helix and β -hairpin) in **this hybrid** RIN, we
196 used a multi-step procedure to reconstruct a pathway of communication between
197 them (Figure 4).

198 **This initial large sub-network is formed by 208 nodes (residues) connected by 456**
199 **NCI edges (in the native RIN). Applying the dynamic filter (independently) on the**
200 **edges, based on the difference in non-covalent interaction score between the native**
201 **TD and each of the mutants ($\Delta\text{NCI}>0$), revealed that in each case a large part of**
202 **network have weakened interactions (hairpin mutant: 200 nodes, 438 NCI edges;**
203 **helix mutant: 199 nodes, 437 NCI edges; helix-hairpin: 202 nodes, 443 NCI edges)**
204 **(Figure 4- figure supplement 1A). The application of the evolutionary filter revealed**
205 **that only a small part of the sub-network is made of evolutionary important residues**
206 **(75 nodes connected by 104 native NCI edges). Combining the filters (Figure 4-**
207 **figure supplement 1B) resulted in 69 nodes connected by 100 NCI edges (thus**
208 **representing interacting residues in the native PapC network). The residues of this**
209 **sub-network (14% of all residues in the TD) were considered 'hot spots' in the**
210 **communication pathway of PapC TD ('hot spot' sub-network). Mapping the hot spot**
211 **residues onto the structure revealed that they are located close together in a**
212 **continuous area within PapC TD.**

213 We analysed the community structure of the hot spot sub-network using the edge-
214 betweenness clustering algorithm (Girvan and Newman, 2002, Morris et al., 2010).
215 This analysis shows that the sub-network has a modular structure, with a modularity
216 index of 0.73 (maximum value of the modularity index is 1), which is typical of 3D-
217 structure based RIN (Newman and Girvan, 2004, Sethi et al., 2009). Here, a total of
218 11 communities containing two or more residues were identified, from which only five
219 communities are composed by more than 5 residues. For further analysis, we chose
220 to consider only these five largest communities, which are located: between β 7-9 and
221 the P-linkers (C1); between the β -hairpin and the conserved region at the base of the

222 α -helix (β 12-14) (C2); between β 12-13 loop and the P-linker1 (C3); between the β -
223 hairpin, P-linker2 and the PD (β E-F) (C4); and on the tip of the PD (β E-F loop and
224 β A-B loop) (C5) (Table 3 and Figure 5).

225 We selected a number of key residues – core hot-spot residues – from the
226 communities that are linking different elements within each community for further
227 experimental investigations (Figure 5). These were found in communities C1-C4: in
228 C1, residues linking the P-linkers and the barrel wall that possibly help in maintaining
229 the P-linkers in a closed configuration (P-linker1:D249, P-linker2:Y329, P-
230 linker2:T331, β 7:R337, and β 8:S363); in C2, residues that bridge the base of the α -
231 helix (the extracellular end of the β 13 conserved patch) and β -hairpin (β -
232 hairpin:R237, β 12:S420, β 13:R442, β 13:S444); in C3, residues on the interface
233 between P-linker1 and β 12-13_loop (P-linker1:Y260, β 12- β 13 loop:K427, β 13:T437,
234 β 13:F438); and in C4, residues that are part of the interface with P-linker1 and the
235 periplasmic end of the β 13 conserved patch (P-linker2:A325, P-linker2:V327).

236

237 *Experimental Analysis of Residues In The Hot Spot Sub-Network*

238

239 To test experimentally if the key hot spot residues identified above (linking elements
240 within each community) contribute to allosteric signalling within PapC, we
241 constructed a set of single alanine substitution mutations (Table 4). Each of the
242 mutants was present at a similar level in the OM compared to the wild-type PapC
243 usher, and the mutations did not affect the ability of the usher to form a stable β -
244 barrel in the OM (data not shown). [1F2]The functionality of the PapC substitution
245 mutants was assessed by ability to assemble P pili on the bacterial surface. P pili
246 bind to receptors on human red blood cells, and assembly of functional P pili was
247 determined using a hemagglutination assay (HA). Seven (D249A, T331A, R442A,
248 S444A, Y260A, K427A, T437A) of the 14 tested mutants exhibited greater than 2-fold
249 defects in agglutination titers compared with wild-type PapC, with 4 of the mutants
250 (D249A, R442A, Y260A, and K427A) exhibiting no agglutination activity (HA titer = 0)
251 (Table 4). The defective mutants were in key residues from communities C1, C2 and
252 C3, confirming roles for these communities in proper usher function.

253 We next used an antibiotic sensitivity assay to screen the PapC substitution mutants
254 for effects on channel activity of the usher. The OM of Gram-negative bacteria has
255 low permeability to detergents such as SDS and to antibiotics such as erythromycin
256 and vancomycin, providing resistance to these molecules. In its resting state, the
257 usher TP is gated closed by the PD, preserving integrity of the OM. Mutations that
258 disrupt channel gating by the PD will result in opening of the large TD channel,

259 leading to increased sensitivity of the bacteria to antibiotics. Bacteria expressing 5 of
260 the PapC substitution mutants (T331A, R237A, Y260A, F438A and V327A) exhibited
261 increased sensitivity to one or more of the tested molecules (Table 4). Y329A^[IF3],
262 R337A, S363A and S420A did not appear to permutate the allosteric signalling within
263 PapC, showing the same antibiotic sensitivity phenotype and ability to assembly pili
264 of the native PapC (Table 4). However, the hemagglutination assay and antibiotic
265 sensitivity assay are screening tools, and as such, they lack in sensitivity to pick up
266 smaller changes in the ability to assemble pili or channel activity of the usher.

267

268 *Electrophysiological analysis of selected mutants*

269

270 In total, ten mutated PapC TDs were found to be affected either in their ability to
271 trigger hemagglutination or in their permeability to SDS or antibiotics (T331A, D249A
272 in C1; R237A, R442A, S444A in C2; Y260A, K427A, T437A, F438A in C3; and
273 V327A in C4). We attempted to purify those mutants in view of examining their
274 channel activity using planar lipid bilayer electrophysiology, which is a more sensitive
275 assay. Unfortunately, only seven of these 10 mutants yielded protein stable enough
276 (as wild-type) in detergent solutions to carry out the planned experiments (T331A in
277 C1; R237A, S444A in C2; K427A, T437A, F438A in C3; and V327A in C4). During
278 the OM extraction procedure, D249A, R442A, and Y260A were not stable enough
279 due to the loss of the membrane bilayer environment and inability to maintain their
280 native conformation in detergents. Insertion of PapC purified proteins (see Methods)
281 was promoted in planar lipid bilayers by clamping the membrane potential to – 90
282 mV. As soon as channel activity was observed, the potential was briefly returned to
283 zero and the chamber stirring stopped to minimize further insertions. Ten-minute long
284 recordings of channel activity at + and – 90 mV, and at + and – 50 mV were
285 performed.

286 The typical electrophysiological signature of the wild-type PapC usher is
287 characterized by prolonged dwell times at a low current level, representing the closed
288 state of the usher, and brief transitions of various current amplitudes. These
289 transitions represent short-lived openings of various conductance, ranging from 50 to
290 600 pS (“transient-mixed” behaviour, TM) (Figure 6A). Although it is not possible to
291 know exactly how many individual pores were inserted into the bilayer, the observed
292 fluctuations of various sizes are taken to represent various conformational states of a
293 single pore. As documented previously, the openings of the “transient-mixed”
294 behaviour appear rather small and may be due to the jiggling of the plug within the
295 TP and/or the thermally-induced mobility of various domains of the protein, such as

296 the NTD and CTDs or loops (Mapingire et al., 2009). Occasionally, and more so at
297 higher membrane potential, very large and sustained openings (“large-open”
298 behaviour, LO) are observed in wild-type PapC usher (Figure 6B). These openings
299 have a conductance of ~ 3-4 nS, which is similar to the monomeric conductance of
300 the mutant lacking the PD (Mapingire et al., 2009) and are interpreted as
301 representing a full displacement of the PD from a single monomer. Prolonged
302 opening of intermediate conductance (0.5-1 nS) can also be observed and may
303 represent partial PD displacement.

304 Because the electrophysiological behaviour of wild-type PapC usher is quite variable,
305 and in attempt to quantify the propensity at spontaneous PD displacement, we have
306 counted the number of 10-min long recordings (sweeps) that show “large-open”
307 behaviour, and we report the percent of such sweeps in various conditions. The
308 frequency of observing these large openings in wild-type PapC usher is ~ 20% at ±
309 50 mV, but increases to ~ 60% at ± 90 mV. The application of a larger
310 transmembrane voltage is likely to disrupt the interactions between key residues
311 involved in keeping the PD in place, leading to a more frequent spontaneous
312 displacement of the latter.

313 Three of the seven analysed mutants, V327A, T331A and F438A, showed an
314 increased propensity at displaying large openings, relative to the wild-type PapC
315 usher, as illustrated for V327A and T331A (Figure 6C, D). This was particularly true
316 at ±90 mV where the percent of sweeps with large openings reaches values of 75-
317 90% (Figure 6F). The T331A mutant was consistently more prone to open than WT
318 and any other mutants, which led to the occasional simultaneous opening of several
319 monomers (Figure 6D). Two of the seven mutants R237A and T437A still opened
320 occasionally to the 3 nS level, but the frequency of sweeps with such events was
321 slightly diminished relative to the wild-type PapC usher at ±90 mV (Figure 6F)
322 suggesting that these mutants are likely to be insignificantly different from the wild-
323 type PapC usher.

324 The K427A and S444A mutants showed a decreased frequency (or complete
325 absence) of large openings. The K427A mutant almost never showed “large-open”
326 behaviour at ±50 and ±90 mV, indicating an extremely closed channel (Figure 6E,
327 Figure 6E, F). The S444A mutant was even less prone to open, with 0% occurrence
328 of PD displacement at ±50 mV in the 11 bilayers that we investigated (Figure 6F).
329 However, increased activity with fast flickers and occasional more prolonged
330 openings could be seen for both mutants if the membrane potential was switched to
331 voltages in the ± 100-150 mV range, indicating that the channels are present in the
332 bilayer, but require higher voltages for activation.

333

334 **Discussion**

335

336 PapC usher catalyses the translocation across the outer membrane of P-pili, and its
337 gating mechanism is important for bacterial homeostasis and for catalysis of pilus
338 assembly. The TD of PapC is formed by the largest β -barrel pore known to be
339 formed by a single chain. The PD occludes the pore in an inactive state and
340 maintains the permeability of the channel. As previously documented, the native
341 PapC channel is highly dynamic and is characterized by spontaneous short-lived
342 openings of various conductance levels (Mapingire et al., 2009). Two distinct
343 structural elements, the β -hairpin and the α -helix, play an important role in
344 maintaining the PD in a closed conformation. In the absence of both elements the
345 usher is defective for pilus biogenesis. Our analysis of the non-covalent interaction
346 RIN in the native PapC TD shows that the interactions found between the TP and the
347 PD are mostly weak, possibly to allow an easy release of the PD. This finding
348 supports the idea that the highly dynamic behavior of the native PapC channel is
349 originated from the 'jiggling' of the PD within the TP (Mapingire et al., 2009). On the
350 other hand, we find that the interactions between the TP and the P-linkers, between
351 the TP and the β -hairpin, and between the TP and the α -helix region were mainly
352 stable, supporting their role in maintaining the PD in a closed conformation.

353

354 Analysis of the mutation-induced perturbation of the non-covalent interaction RIN of
355 the native PapC TD in absence of the β -hairpin, α -helix or both, shows that the
356 interactions between nodes that are not part of the deleted secondary structure
357 elements are consistently weakened (Figure 2- figure supplement 1), suggesting the
358 two elements are not independent. This feature can be interpreted as part of a
359 complex allosteric process regulating the PD gating mechanism. It has been
360 proposed that only a few residues play essential roles during allosteric
361 processes and that perturbing the interactions between these residues can facilitate
362 the population shift of the conformational ensembles (Tsai et al., 2009, del Sol et al.,
363 2009). Additionally, it has been shown that residues with a pivotal role in mediating
364 such allosteric communications are also important residues in terms of evolution
365 (both highly-coevolved residues and conserved residues) (Suel et al., 2003, Tang et
366 al., 2007, Ferguson et al., 2007). Remarkably, we show here that PapC is
367 characterised by an uneven distribution of the evolutionary important residues,
368 clustered in the P-linkers, the PD, and the barrel wall capped by the α -helix, in close
369 proximity to the β -hairpin. Another interesting finding is the presence of the ' β 13

370 conserved patch' that reaches the full height of the pore from the α -helix region to a
371 loop located between β 12 and β 13 strands (β 12-13_loop), which has been previously
372 identified as important (Volkan et al., 2013, Farabella, 2013).

373

374 To detect the allosteric network we implement a new method that integrates dynamic
375 and evolutionary information in a hybrid RIN and then apply network analysis. This
376 approach allows us to explore a large part of the protein, resulting in the detection of
377 only 14% of all residues in the TD as potential candidates. It has been shown that
378 detecting "residue communities" in protein structure networks leads to the
379 identification of key residues that are often part of a signal transduction pathway (del
380 Sol et al., 2007, Bode et al., 2007). The interconnection within and between the
381 communities is pivotal for the flow of allosteric signalling. Residues in the same
382 community are densely interconnected and have multiple routes to communicate with
383 each another. However, the interconnections between communities involve only a
384 few edges, which form the bottleneck for the flow of the signal in the network (del Sol
385 et al., 2007, Bode et al., 2007, Sethi et al., 2009). Here, all the identified communities
386 (C1-C5, Figure 5) comprise residues from multiple elements (e.g. β -hairpin, P-
387 linker1, P-linker2 and distinct part of the TP) of PapC TD except C5 that is composed
388 only by PD residues. Mutation of key residues linking elements within each of the
389 four communities C1-C4 showed an altered antibiotic sensitivity phenotype,
390 confirming a role of these communities in the pore gating mechanism. Additionally,
391 communities C1-C3 are required for proper usher function (as estimated by the
392 hemmoglutatione assay, suggesting dependency of the pore gating function and
393 pilus assembly (based on some of the mutations). For example, mutations in β 12-
394 13_loop:K427 in C3 and β 13:S444 in C2, which lead to a drastic decrease in the
395 frequency of plug displacement (stabilizing the closed state, based on channel
396 activity analysis) also show defective pilus assembly (possibly due to their deficiency
397 in relocating the plug in a functional conformation).

398

399 Additionally, our study identifies two residues with opposite effects on plug
400 displacement in the same community: C3: β 12-13_loop:K427, which is located on
401 β 12-13_loop, and β 13:F438, which is located at the periplasmic end of the β 13
402 conserved patch (linking β 12-13_loop with the α -helix region). This observation
403 suggests that β 12-13_loop has a pivotal role in modulating plug displacement (acting
404 as a 'latch') as proposed previously (Farabella, 2013), and supported by additional
405 mutagenesis studies (Volkan et al., 2013). Intriguingly, at the extracellular end of the
406 β 13 conserved patch there is another mutation (β 13:S444) leading to a channel more

407 reluctant to open, suggesting a regulatory role for the β 13 conserved patch (Figure 7)
408 in modulating the 'latch' (β 12-13_loop).

409
410 Interestingly, our community residues in some cases are found to have very different
411 patterns of interactions in an alternative conformation. For example, the functionally
412 important 'latch' (β 12-13_loop) is shown to be in a different conformation in the open
413 state (based on FimD:FimC:FimH structure (Phan et al., 2011)) as well as other
414 residues in C3 and C4 communities that change their interaction pattern in the open
415 state. As a result, some residues that are identified by our method, such as V327,
416 could in principle be selected also by visual inspection for further experimental
417 investigation. V327 (located on P-linker2 in C4 at interface with P-linker1 and the
418 barrel wall) is found to have a different interaction pattern in the open state of the
419 usher and is also shown to promote plug displacement. However, more importantly,
420 the method is able to predict residues in regions that would not attract our attention
421 at all but may contain important information in respect to the allosteric pathway. (This
422 aspect can become even more significant in the absent of an alternative
423 conformation). For example, we predict two such residues to be functionally
424 important –T331 and S444. T331 is located on P-linker2 at the interface between P-
425 linker2 and the barrel wall, which stays intact in the open conformation of the usher;
426 S444 is located on β 13 at the base of the α -helix, i.e. on the barrel itself, and it lacks
427 any direct contact with the plug domain. However, both residues were shown to have
428 an effect on plug displacement. Thus, the main strength of our method is that the
429 knowledge of the structure of the protein in one conformation only and enough
430 sequence information to extract evolutionary information are sufficient for the
431 detection of functionally important residues that can be pivotal for transferring the
432 regulatory information within the protein.

433

434 **Conclusion**

435 In this study, we provide a first deep insight into the allostery regulating the gating
436 mechanism of the usher family. Using PapC TD as model system, we developed an
437 integrative approach, combining computational modelling, sequence conservation
438 analysis, mutual information-based coevolution analysis and information from AA-MD
439 simulations, to study the potential involvement of particular secondary structure
440 elements (the α -helix and β -hairpin) in the allosteric communication. The construction
441 of a hybrid interaction network and the use of network analysis allowed us to identify
442 communities of residues within the TD that potentially mediate this process. Antibiotic
443 sensitivity and electrophysiology experiments on a set of alanine-substitution mutants

444 confirmed that residues located in the P-linkers, the β -hairpin, and β 13 conserved
445 patch (part of four communities) alter channel gating and that residues located in P-
446 linker2, β 12- β 13_loop and β 13 conserved patch (both periplasmic end and
447 extracellular end) are sensitive to plug displacement. Therefore, we suggest that the
448 β 13 conserved patch acts as a regulator of the 'latch' (β 12- β 13_loop) mediating
449 channel opening. Furthermore, our study shows how the integration of different
450 computational approaches based on evolution, structure and dynamics of proteins,
451 into a hybrid network can unveil communication pathways within proteins. Such an
452 integrative network modelling approach can guide the experimental investigation by
453 pinpointing key candidates involved in the transmission of the allosteric signal.

454

455 **Methods**

456 *Systems Modelling*

457 We built four model systems based on the X-ray structure of the TD (residues 1-492)
458 of PapC at 3.2 Å resolution (PDB ID: 2vqi (Remaut et al., 2008)). The starting model
459 for the simulation of the native TD (sim1, native) was generated by adding the
460 missing loops to the X-ray structure using the *dope_loopmodel* method (Shen and
461 Sali, 2006) in MODELLER-9v7 (Sali and Blundell, 1993). Additionally, three mutant
462 model systems were constructed based on the native model: a mutant that lacking
463 the β -hairpin (Δ 233-240) (sim2, hairpin mutant); a mutant lacking the α -helix (Δ 447-
464 460) (sim3, helix mutant) and a mutant lacking both (Δ 233-240 and Δ 447-460) (sim4,
465 helix-hairpin mutant).

466 Each of the systems was oriented with respect to the membrane normal (the Z axis
467 by definition) using the database (Lomize et al., 2006). For the native model (sim1) a
468 mixed lipid bilayer (POPE/POPG 3:1) was generated around the protein using the
469 replacement method (Jo et al., 2007). To obtain a mixed lipid bilayer that reproduces
470 an estimated surface area per lipid of $61.5 \pm 0.2 \text{ \AA}^2$ (Murzyn et al., 2005) we used the
471 InflateGRO method (Kandt et al., 2007).

472

473 *Systems Set Up For MD Simulations*

474 All MD simulations were performed using Gromacs 4.0.5 (Van Der Spoel et al.,
475 2005). TIP3P parameters were used for water molecules (Jorgensen et al., 1983),
476 the OPLSA-AA force-field (Kaminski et al., 2001) was applied to the protein and ions,
477 and the Berger force-field (Berger et al., 1997) to the lipids. All four systems were
478 solvated in water and ions were added to neutralize the total charge (0.15 M NaCl),
479 resulting in more than 75,000 atoms in total. Next, each system was energy-
480 minimized using a steepest descent algorithm in the presence of different position

481 restraints on the protein and the lipid bilayer head-groups, which were gradually
482 removed. Note that in the simulations of the mutant systems (sim2 to sim4), the
483 protein models were embedded in the pre-equilibrated membrane obtained after 15
484 ns of unrestrained equilibration of the native TD (sim1).

485

486 *Equilibration Procedure and Production Run*

487 The assembled systems were equilibrated in a multistage process using periodic
488 boundary conditions and a 2 fs time step. Short-range interactions were used with a
489 cut-off of 1 nm. The PME algorithm (Darden et al., 1993) was used for long-range
490 electrostatic interactions. All bonds were constrained using the LINCS algorithm
491 (Hess et al., 1997, Hess, 2008). The first equilibration step was performed in the NVT
492 ensemble, using a restraining force of 1000 kJ/(mol nm²) for 0.1 ns on the protein
493 and lipids. The V-rescale thermostat (Bussi et al., 2007) was employed to couple the
494 temperature of the system to 310K with a time constant of $t_{\tau} = 0.1$ ps. All the
495 following equilibration steps were performed in the NPT ensemble. During the next 3
496 steps different parts of the system were restrained using a force constant of 1000
497 kJ/(mol nm²): the protein and lipids, the protein atoms only, and the protein backbone
498 atoms.

499 The resulting model of each system was then simulated without restraints. Constant
500 temperature of 310K was maintained using the Nose-Hoover thermostat (Hoover,
501 1985) (Nose, 1984.) with a time constant of $t_{\tau} = 0.1$ ps. Using semi-isotropic coupling
502 with a Parrinello-Rahman barostat (Parrinello, 1981), a constant pressure of 1 bar
503 was applied with a coupling constant (t_{P}) of 1 ps and a compressibility 4.5×10^{-5} bar⁻¹.
504 Each unrestrained simulation was performed for ~70-72 ns. The last 50 ns of
505 simulation were used for analysis.

506

507 *Non-covalent residue interaction network*

508 Hydrogen bonds were defined using a cut-off of 30° for the acceptor–donor–
509 hydrogen angle and a cut-off of 3.5 Å for the hydrogen–acceptor distance. The
510 definition of salt-bridges was based on a 4 Å distance cut-off between any oxygen
511 atoms of acidic residues and nitrogen atoms of basic residues. The *non-covalent*
512 *interaction score* (NCI score) of the identified bonds was defined as the percentage
513 of simulation time during which a bond occurs between two amino acids normalised
514 by the number of bonds. Using the normalized score, a non-covalent residue
515 interaction network (RIN) was built for each of the simulated systems as a weighted
516 undirected graph, in which each node represents a residue and each edge is

517 weighted by the normalized score. The difference in non-covalent interaction score
518 (Δ NCI score) between the native TD system and each of the mutant systems was
519 then calculated and added as a weighted undirected edge to the pre-existing non-
520 covalent RIN.

521

522 *Sequence Conservation Analysis*

523 The sequence corresponding to the X-ray structures of PapC TD (PDB id: 2vqi;
524 Uniprot id: P07110) was used as input to psiBlast resulting in a set of unique related
525 sequences from the non-redundant NCBI dataset (Altschul et al., 1997). The E-value
526 threshold was set as 10^{-3} and sequences with id > 90% and < 30% sequence identity
527 were excluded. The structure-based multiple sequence alignment was calculated
528 using Expresso (3DCoffee) (Armougom et al., 2006). Finally, an evolutionary
529 conservation score was calculated for each residue an empirical Bayesian inference
530 method (Mayrose et al., 2004) as implemented in the ConSurf web server
531 (Ashkenazy et al., 2010).

532

533 *Sequence Coevolution Analysis*

534 To estimate the coevolution within the residues in the usher TD we used Normalized
535 Mutual Information (NMI) (Martin et al., 2005) over all position pairs in the multiple
536 sequence alignment obtained as described above. NMI calculations were performed
537 using PyCogent (Knight et al., 2007, Caporaso et al., 2008) and a Z-score was
538 calculated for each residue pair based on the standard deviation from the mean NMI
539 values. Only residue pairs that had Z-score > 4 were identified as coevolved
540 pairs(Martin et al., 2005, Gloor et al., 2005).

541 Next, a coevolutionary RIN was built as a weighted undirected graph where each
542 node represents a residue (as in the non-covalent RIN) and an edge connecting two
543 nodes is the NMI score. The network was visualized and analyzed in Cytoscape
544 2.8.2 (Smoot et al., 2011) using NetworkAnalyzer plug-in (Assenov et al., 2008) for
545 calculating degrees of connectivity and RINalyzer plug-in (Doncheva et al., 2011) for
546 mapping the network on the PapC structure.

547

548 *Hybrid RIN*

549 To store the entire information we combined the coevolutionary RIN, the non-
550 covalent RIN, and the conservation analysis into one network. In this hybrid network
551 each node represents each PapC TD residue and is associated with the

552 corresponding conservation score. Two nodes can have multiple edges, each
553 weighted according to the information it carries (NMI score, NCI score, or Δ NCI).

554

555 *Allosteric 'hot spots' sub-network reconstruction*

556 Using the hybrid network and starting from the α -helix region (residues 445-468) and
557 β -hairpin residues (residues 230-240) we generated a sub-network of first
558 neighbours residues based only on the NCI score higher than 0.3 in the native RIN.

559 This sub-network was expanded by again adding only neighbouring residues
560 connected by NCI score higher than 0.3. The procedure was repeated until no more

561 new residues could be added to the sub-network. A first set was generated by

562 filtering the sub-network based on the evolutionary information. The filtering was

563 done by selecting nodes with a conservation score of 9 (i.e., highly conserved) or

564 nodes that are connected by an NMI edge (i.e., coevolved significantly). A second

565 set was generated by filtering out the nodes that are not connected by a weakened

566 interaction (Δ NCI $>$ 0) in each of the mutant systems. Intersecting the identified sets

567 resulted in a 'hot spots' sub-network. ~~TA first set was generated by filtering the sub-~~

568 ~~network based on the evolutionary information. The filtering was done by only leaving~~

569 ~~nodes with a conservation score of 9 (i.e., highly conserved) or nodes that are~~

570 ~~connected by an NMI edge (i.e., coevolved significantly). A second set was~~

571 ~~generated by filtering out the nodes that are not connected by a weakened~~

572 ~~interaction (Δ NCI $>$ 0) in each of the mutant systems. Intersecting the identified sets~~

573 ~~resulted in a 'hot spots' sub-network.~~ The 'hot spots' sub-network was then

574 decomposed into communities using the edge-betweenness clustering algorithm

575 (GLay) as implemented in clusterMaker (Girvan and Newman, 2002, Morris et al.,

576 2010). Cytoscape 2.8.2 (Smoot et al., 2011), RINalyzer plug-in (Doncheva et al.,

577 2011), and Chimera (Pettersen et al., 2004) were used for mapping the network on

578 the PapC structure.

579

580 *PapC substitution mutants*

581 The PapC alanine substitution mutants were derived from plasmid pDG2 using the

582 QuikChange Site-Directed Mutagenesis Kit (Stratagene) and the primers listed in

583 Supplementary file 1. Plasmid pDG2 encodes wild-type *papC* with a C-terminal,

584 thrombin-cleavable His-tag (Li et al., 2004). All mutants were sequenced to verify the

585 intended mutation.

586

587 *Expression and folding of the PapC substitution mutants in the outer membrane*

588 Each of the PapC mutants was compared with wild-type PapC for expression levels
589 and ability to fold into a stable β -barrel in the OM. OM isolation, analysis of usher
590 protein levels, and the heat-modifiable mobility assay for β -barrel stability were done
591 as previously described (Henderson et al., 2011).

592

593 *Hemagglutination assay*

594 HA assays were performed to test the ability of each of the PapC substitution
595 mutants to assemble functional P pili on the bacterial surface. HA assays were
596 performed by serial dilution in microtiter plates as previously described (Henderson
597 et al., 2011). HA titers were determined visually as the highest fold dilution of
598 bacteria still able to agglutinate human red blood cells. Each assay was performed in
599 triplicate; each mutant was analyzed twice and the values averaged.

600

601 *Top soft agar assay for antibiotic sensitivity*

602 Bacteria were grown in LB medium supplemented with 100 μ g/ml ampicillin (Amp) to
603 an OD₆₀₀ of 0.6 and then induced for PapC expression with 0.1% arabinose for 1 h.
604 Aliquots of 0.1 ml bacteria were added to 3 ml melted soft top agar (0.75% LB agar)
605 cooled to 45°C and supplemented with 100 μ g/ml Amp and 0.1% arabinose. The
606 bacteria and melted agar were mixed well and poured on top of 1.5% solid LB agar
607 plates containing 100 μ g/ml Amp and 0.1% arabinose. Once the top agar solidified,
608 sterile 6 mm filter discs were placed on top and 10 μ l of the following antibiotics were
609 added: 75 mg/ml SDS, 2 mg/ml vancomycin, or 1.5 mg/ml erythromycin. The
610 diameter of the growth inhibition zone around the antibiotics, including the filter disc,
611 was measured after overnight growth at 37°C. Each PapC mutant strain was tested
612 twice and the values averaged.

613

614 *Electrophysiological analysis of selected PapC substitution mutants*

615 PapC mutants (R237A, F438A, V327A, K427A, T437A, T331A and S444A) were
616 purified according to published protocols (Henderson and Thanassi, 2013) and
617 investigated by planar lipid bilayer electrophysiology. Planar bilayers were made from
618 a preparation of L- α -phosphatidylcholine Type II-S from Sigma (also known as
619 asolectin) according to the Montal and Mueller technique (Montal and Mueller, 1972)
620 following a published protocol (Mapingire et al., 2013). Protein aliquots were diluted
621 1:1 in buffer T (1M KCl, 5 mM HEPES, pH 7.2) containing either 1 or 2%
622 *N*-Octyl- oligo- oxyethylene (octyl-POE, Axxora). Eight micrograms of protein from
623 the diluted sample was added to the *cis* side of a planar lipid bilayer chamber
624 containing ~ 1.5 mL of buffer T. Gentle stirring was applied to promote spontaneous

625 insertions of the protein into the bilayer. Channel activity was monitored by
626 measuring current under voltage clamp conditions using an Axopatch 1D amplifier
627 with a CV4B headstage or an Axopatch 200B amplifier (Axon Instruments). The
628 current was digitized (ITC-18, Instrutech), and stored on a PC computer using the
629 Acquire software (Bruyton). Ten-minute long traces were sampled at 1.25 ms
630 intervals and filtered at 500 Hz. Both chambers contained buffer T and Ag/AgCl
631 electrodes with pellet. The *trans* side of the bilayer was set as ground. Insertions
632 were typically performed at -90 mV. Data display and analysis were done with
633 pCLAMP software (Axon Instruments).

634

635 **Acknowledgments**

636 The authors are grateful to Dr. David Houldershaw for computing support, and Drs.
637 Tsjerk Wassenaar and Paula Petrone for useful discussions. The authors
638 acknowledge the use of the UCL Legion High Performance Computing facility, and
639 associated services, in the completion of this work. The work was supported by a
640 Wellcome Trust studentship (to IF), MRC (G0600084) and BBSRC (BB/K01692X/1)
641 (to MT) as well as National Institutes of Health grant GM062987 (to DGT).

642 **References**

643

- 644 ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., ZHANG, Z.,
645 MILLER, W. & LIPMAN, D. J. 1997. Gapped BLAST and PSI-BLAST: a new
646 generation of protein database search programs. *Nucleic Acids Res*, 25,
647 3389-402.
- 648 ARMOUGOM, F., MORETTI, S., POIROT, O., AUDIC, S., DUMAS, P., SCHAELEI, B.,
649 KEDUAS, V. & NOTREDAME, C. 2006. Espresso: automatic incorporation of
650 structural information in multiple sequence alignments using 3D-Coffee.
651 *Nucleic Acids Res*, 34, W604-8.
- 652 ASHKENAZY, H., EREZ, E., MARTZ, E., PUPKO, T. & BEN-TAL, N. 2010. ConSurf
653 2010: calculating evolutionary conservation in sequence and structure of
654 proteins and nucleic acids. *Nucleic Acids Res*, 38, W529-33.
- 655 ASSENOV, Y., RAMIREZ, F., SCHELHORN, S. E., LENGAUER, T. & ALBRECHT,
656 M. 2008. Computing topological parameters of biological networks.
657 *Bioinformatics*, 24, 282-4.
- 658 ATILGAN, A. R., AKAN, P. & BAYSAL, C. 2004. Small-world communication of
659 residues and significance for protein dynamics. *Biophys J*, 86, 85-91.
- 660 BERGER, O., EDHOLM, O. & JAHNIG, F. 1997. Molecular dynamics simulations of
661 a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant
662 pressure, and constant temperature. *Biophys J*, 72, 2002-13.
- 663 BODE, C., KOVACS, I. A., SZALAY, M. S., PALOTAI, R., KORCSMAROS, T. &
664 CSERMELY, P. 2007. Network analysis of protein dynamics. *FEBS Lett*, 581,
665 2776-82.
- 666 BUSSI, G., DONADIO, D. & PARRINELLO, M. 2007. Canonical sampling through
667 velocity rescaling. *The Journal of Chemical Physics*, 126.
- 668 CAPITANI, G., EIDAM, O., GLOCKSHUBER, R. & GRUTTER, M. G. 2006.
669 Structural and functional insights into the assembly of type 1 pili from
670 *Escherichia coli*. *Microbes Infect*, 8, 2284-90.
- 671 CAPORASO, J. G., SMIT, S., EASTON, B. C., HUNTER, L., HUTTLEY, G. A. &
672 KNIGHT, R. 2008. Detecting coevolution without phylogenetic trees? Tree-
673 ignorant metrics of coevolution perform as well as tree-aware metrics. *BMC*
674 *Evol Biol*, 8, 327.
- 675 DAILY, M. D., UPADHYAYA, T. J. & GRAY, J. J. 2008. Contact rearrangements form
676 coupled networks from local motions in allosteric proteins. *Proteins*, 71, 455-
677 66.
- 678 DARDEN, T., YORK, D. & PEDERSEN, L. 1993. Particle mesh Ewald: An $N \cdot \log(N)$
679 method for Ewald sums in large systems. *The Journal of Chemical Physics*,
680 98.
- 681 DEL SOL, A., ARAUZO-BRAVO, M. J., AMOROS, D. & NUSSINOV, R. 2007.
682 Modular architecture of protein structures and allosteric communications:
683 potential implications for signaling proteins and regulatory linkages. *Genome*
684 *Biol*, 8, R92.
- 685 DEL SOL, A., TSAI, C. J., MA, B. & NUSSINOV, R. 2009. The origin of allosteric
686 functional modulation: multiple pre-existing pathways. *Structure*, 17, 1042-50.
- 687 DONCHEVA, N. T., KLEIN, K., DOMINGUES, F. S. & ALBRECHT, M. 2011.
688 Analyzing and visualizing residue networks of protein structures. *Trends*
689 *Biochem Sci*, 36, 179-82.
- 690 FARABELLA, I. 2013. *Computational studies of the usher outer-membrane assembly*
691 *platform involved in pilus biogenesis*. Doctor of Philosophy, Birkbeck College,
692 University of London.
- 693 FERGUSON, A. D., AMEZCUA, C. A., HALABI, N. M., CHELLIAH, Y., ROSEN, M.
694 K., RANGANATHAN, R. & DEISENHOFER, J. 2007. Signal transduction

695 pathway of TonB-dependent transporters. *Proc Natl Acad Sci U S A*, 104,
696 513-8.

697 GEIBEL, S., PROCKO, E., HULTGREN, S. J., BAKER, D. & WAKSMAN, G. 2013.
698 Structural and energetic basis of folded-protein transport by the FimD usher.
699 *Nature*, 496, 243-6.

700 GIRVAN, M. & NEWMAN, M. E. 2002. Community structure in social and biological
701 networks. *Proc Natl Acad Sci U S A*, 99, 7821-6.

702 GLOOR, G. B., MARTIN, L. C., WAHL, L. M. & DUNN, S. D. 2005. Mutual
703 information in protein multiple sequence alignments reveals two classes of
704 coevolving positions. *Biochemistry*, 44, 7156-65.

705 HAIYAN, L. & JIHUA, W. Network Properties of Protein Structures at Three Different
706 Length Scales. Proceedings of the 2009 Second Pacific-Asia Conference on
707 Web Mining and Web-based Application %@ 978-0-7695-3646-0, 2009. IEEE
708 Computer Society, 161-164.

709 HENDERSON, N. S., NG, T. W., TALUKDER, I. & THANASSI, D. G. 2011. Function
710 of the usher N-terminus in catalysing pilus assembly. *Mol. Microbiol.*, 79, 954-
711 67.

712 HENDERSON, N. S. & THANASSI, D. G. 2013. Purification of the outer membrane
713 usher protein and periplasmic chaperone-subunit complexes from the P and
714 type 1 pilus systems. *Methods Mol Biol*, 966, 37-52.

715 HESS, B. 2008. P-LINCS: A Parallel Linear Constraint Solver for Molecular
716 Simulation. *J. Chem. Theory Comput.*, 4, 116-122.

717 HESS, B., BEKKER, H., BERENDSEN, H. J. C. & FRAAIJE, J. G. E. M. 1997.
718 LINCS: A linear constraint solver for molecular simulations. *J. Comput.*
719 *Chem.*, 18, 1463-1472.

720 HOOVER, W. G. 1985. Canonical dynamics: Equilibrium phase-space distributions.
721 *Phys Rev A*, 31, 1695-1697.

722 HUANG, Y., SMITH, B. S., CHEN, L. X., BAXTER, R. H. & DEISENHOFER, J. 2009.
723 Insights into pilus assembly and secretion from the structure and functional
724 characterization of usher PapC. *Proc Natl Acad Sci U S A*, 106, 7403-7.

725 JO, S., KIM, T. & IM, W. 2007. Automated builder and database of protein/membrane
726 complexes for molecular dynamics simulations. *PloS one*, 2, e880.

727 JORGENSEN, W. L., CHANDRASEKHAR, J., MADURA, J. D., IMPEY, R. W. &
728 KLEIN, M. L. 1983. Comparison of simple potential functions for simulating
729 liquid water. *J Chem Phys*, 79, 926.

730 KAMINSKI, G. A., FRIESNER, R. A., TIRADO-RIVES, J. & JORGENSEN, W. L.
731 2001. Evaluation and reparametrization of the OPLS-AA force field for
732 proteins via comparison with accurate quantum chemical calculations on
733 peptides. *J Phys Chem B.* , 105, 6474–6487.

734 KANDT, C., ASH, W. L. & TIELEMAN, D. P. 2007. Setting up and running molecular
735 dynamics simulations of membrane proteins. *Methods*, 41, 475-88.

736 KNIGHT, R., MAXWELL, P., BIRMINGHAM, A., CARNES, J., CAPORASO, J. G.,
737 EASTON, B. C., EATON, M., HAMADY, M., LINDSAY, H., LIU, Z.,
738 LOZUPONE, C., MCDONALD, D., ROBESON, M., SAMMUT, R., SMIT, S.,
739 WAKEFIELD, M. J., WIDMANN, J., WIKMAN, S., WILSON, S., YING, H. &
740 HUTTLEY, G. A. 2007. PyCogent: a toolkit for making sense from sequence.
741 *Genome Biol*, 8, R171.

742 LI, H., QIAN, L., CHEN, Z., THAHBOT, D., LIU, G., LIU, T. & THANASSI, D. G. 2004.
743 The outer membrane usher forms a twin-pore secretion complex. *J Mol Biol*,
744 344, 1397-1407.

745 LIU, Y. & BAHAR, I. 2012. Sequence evolution correlates with structural dynamics.
746 *Mol Biol Evol*, 29, 2253-63.

747 LOMIZE, M. A., LOMIZE, A. L., POGOZHEVA, I. D. & MOSBERG, H. I. 2006. OPM:
748 orientations of proteins in membranes database. *Bioinformatics*, 22, 623-5.

749 MAPINGIRE, O. S., HENDERSON, N. S., DURET, G., THANASSI, D. G. &
750 DELCOUR, A. H. 2009. Modulating effects of the plug, helix, and N- and C-
751 terminal domains on channel properties of the PapC usher. *J Biol Chem*, 284,
752 36324-33.

753 MAPINGIRE, O. S., WAGER, B. & DELCOUR, A. H. 2013. Electrophysiological
754 characterization of bacterial pore-forming proteins in planar lipid bilayers.
755 *Methods Mol Biol*, 966, 381-96.

756 MARTIN, L. C., GLOOR, G. B., DUNN, S. D. & WAHL, L. M. 2005. Using information
757 theory to search for co-evolving residues in proteins. *Bioinformatics*, 21,
758 4116-24.

759 MAYROSE, I., GRAUR, D., BEN-TAL, N. & PUPKO, T. 2004. Comparison of site-
760 specific rate-inference methods for protein sequences: empirical Bayesian
761 methods are superior. *Mol Biol Evol*, 21, 1781-91.

762 MONTAL, M. & MUELLER, P. 1972. Formation of bimolecular membranes from lipid
763 monolayers and a study of their electrical properties. *Proc Natl Acad Sci U S A*,
764 69, 3561-6.

765 MORRIS, J. H., MENG, E. C. & FERRIN, T. E. 2010. Computational tools for the
766 interactive exploration of proteomic and structural data. *Mol Cell Proteomics*,
767 9, 1703-15.

768 MURZYN, K., ROG, T. & PASENKIEWICZ-GIERULA, M. 2005.
769 Phosphatidylethanolamine-phosphatidylglycerol bilayer as a model of the
770 inner bacterial membrane. *Biophys J*, 88, 1091-103.

771 NAVEED, H., JACKUPS, R., JR. & LIANG, J. 2009. Predicting weakly stable regions,
772 oligomerization state, and protein-protein interfaces in transmembrane
773 domains of outer membrane proteins. *Proc Natl Acad Sci U S A*, 106, 12735-
774 40.

775 NEWMAN, M. E. & GIRVAN, M. 2004. Finding and evaluating community structure in
776 networks. *Phys Rev E Stat Nonlin Soft Matter Phys*, 69, 026113.

777 NG, T. W., AKMAN, L., OSISAMI, M. & THANASSI, D. G. 2004. The usher N
778 terminus is the initial targeting site for chaperone-subunit complexes and
779 participates in subsequent pilus biogenesis events. *J Bacteriol*, 186, 5321-31.

780 NOSE, S. 1984. A molecular dynamics method for simulations in the canonical
781 ensemble. *Mol. Phys*, 255-268.

782 PARRINELLO, M. 1981. Polymorphic transitions in single crystals: A new molecular
783 dynamics method. *Journal of Applied Physics*, 52.

784 PETERSEN, E. F., GODDARD, T. D., HUANG, C. C., COUCH, G. S.,
785 GREENBLATT, D. M., MENG, E. C. & FERRIN, T. E. 2004. UCSF Chimera--
786 a visualization system for exploratory research and analysis. *J Comput*
787 *Chem*, 25, 1605-12.

788 PHAN, G., REMAUT, H., WANG, T., ALLEN, W. J., PIRKER, K. F., LEBEDEV, A.,
789 HENDERSON, N. S., GEIBEL, S., VOLKAN, E., YAN, J., KUNZE, M. B.,
790 PINKNER, J. S., FORD, B., KAY, C. W., LI, H., HULTGREN, S. J.,
791 THANASSI, D. G. & WAKSMAN, G. 2011. Crystal structure of the FimD usher
792 bound to its cognate FimC-FimH substrate. *Nature*, 474, 49-53.

793 REMAUT, H., TANG, C., HENDERSON, N. S., PINKNER, J. S., WANG, T.,
794 HULTGREN, S. J., THANASSI, D. G., WAKSMAN, G. & LI, H. 2008. Fiber
795 formation across the bacterial outer membrane by the chaperone/usher
796 pathway. *Cell*, 133, 640-52.

797 ROBERTS, J. A., MARKLUND, B. I., ILVER, D., HASLAM, D., KAACK, M. B.,
798 BASKIN, G., LOUIS, M., MOLLBY, R., WINBERG, J. & NORMARK, S. 1994.
799 The Gal(alpha 1-4)Gal-specific tip adhesin of Escherichia coli P-fimbriae is
800 needed for pyelonephritis to occur in the normal urinary tract. *Proc Natl Acad*
801 *Sci U S A*, 91, 11889-93.

802 SALI, A. & BLUNDELL, T. L. 1993. Comparative protein modelling by satisfaction of
803 spatial restraints. *J Mol Biol*, 234, 779-815.

804 SETHI, A., EARGLE, J., BLACK, A. A. & LUTHEY-SCHULTEN, Z. 2009. Dynamical
805 networks in tRNA:protein complexes. *Proc Natl Acad Sci U S A*, 106, 6620-5.
806 SHEN, M. Y. & SALI, A. 2006. Statistical potential for assessment and prediction of
807 protein structures. *Protein Sci*, 15, 2507-24.
808 SMOOT, M. E., ONO, K., RUSCHEINSKI, J., WANG, P. L. & IDEKER, T. 2011.
809 Cytoscape 2.8: new features for data integration and network visualization.
810 *Bioinformatics*, 27, 431-2.
811 SUEL, G. M., LOCKLESS, S. W., WALL, M. A. & RANGANATHAN, R. 2003.
812 Evolutionarily conserved networks of residues mediate allosteric
813 communication in proteins. *Nat Struct Biol*, 10, 59-69.
814 TANG, S., LIAO, J. C., DUNN, A. R., ALTMAN, R. B., SPUDICH, J. A. & SCHMIDT,
815 J. P. 2007. Predicting allosteric communication in myosin via a pathway of
816 conserved residues. *J Mol Biol*, 373, 1361-73.
817 TAYLOR, N. R. 2013. Small world network strategies for studying protein structures
818 and binding. *Comput Struct Biotechnol J*, 5, e201302006.
819 THANASSI, D. G., SAULINO, E. T. & HULTGREN, S. J. 1998. The chaperone/usher
820 pathway: a major terminal branch of the general secretory pathway. *Curr
821 Opin Microbiol*, 1, 223-31.
822 THANASSI, D. G., STATHOPOULOS, C., DODSON, K., GEIGER, D. & HULTGREN,
823 S. J. 2002. Bacterial outer membrane ushers contain distinct targeting and
824 assembly domains for pilus biogenesis. *J Bacteriol*, 184, 6260-9.
825 TSAI, C. J., DEL SOL, A. & NUSSINOV, R. 2009. Protein allostery, signal
826 transmission and dynamics: a classification scheme of allosteric mechanisms.
827 *Mol Biosyst*, 5, 207-16.
828 VAN DER SPOEL, D., LINDAHL, E., HESS, B., GROENHOF, G., MARK, A. E. &
829 BERENDSEN, H. J. 2005. GROMACS: fast, flexible, and free. *J Comput
830 Chem*, 26, 1701-18.
831 VOLKAN, E., KALAS, V., PINKNER, J. S., DODSON, K. W., HENDERSON, N. S.,
832 PHAM, T., WAKSMAN, G., DELCOUR, A. H., THANASSI, D. G. &
833 HULTGREN, S. J. 2013. Molecular basis of usher pore gating in Escherichia
834 coli pilus biogenesis. *Proc Natl Acad Sci U S A*, 110, 20741-6.
835 WAKSMAN, G. & HULTGREN, S. J. 2009. Structural biology of the chaperone-usher
836 pathway of pilus biogenesis. *Nat Rev Microbiol*, 7, 765-74.
837
838
839

840 **Figure Legend**

841

842 **Figure 1. PapC usher organization and detail of its translocation domain.** A. A
843 diagram of the domain organization of PapC usher. NTD (dark-blue) represents the
844 N-terminal domain, CTD1 (light-violet) and CTD2 (dark-violet) represent the C-
845 terminal domains; TD represents the translocation domain, comprising the TP
846 (translocation pore, light-blue) and the PD (plug domain, magenta). B and C: Ribbon
847 representation of the starting model of the native translocation domain (TD) of PapC
848 with the labels 'N' and 'C' indicating the N and C termini of the translocation channel.
849 The β -barrel, PD (including the P-linkers), β -hairpin, and α -helix (including the H-
850 linkers) are coloured blue, magenta, orange, and yellow, respectively. The outer
851 membrane position is represented schematically with the labels 'E', 'M' and 'P'
852 indicating the extracellular side, the membrane and the periplasmic side,
853 respectively. Side view of the TD (B) is shown with the α -helix, β -hairpin, H-linker1,
854 H-linker2, P-linker1, P-linker2, and PD, labelled. Extracellular top view of the TD (C)
855 is shown with the barrel β strands labelled β 1 through β 24 and with the PD strands
856 labelled β A through β F. The figures were created with Chimera (Pettersen et al.,
857 2004).

858

859 **Figure 2. Non-covalent interaction network (non-covalent RIN).** A. Ribbon
860 representation of the starting model of the native translocation domain (TD) of PapC
861 with the labels 'N' and 'C' indicating the N and C termini of the translocation channel.
862 The β -barrel, PD, P-linker1, P-linker2, β -hairpin, and α -helix (including the H-linkers)
863 are coloured grey, magenta, light purple, dark purple, orange, and yellow,
864 respectively. The α -helix, β -hairpin, P-linker1, P-linker2, and PD are labelled. B.
865 Protein structure network representation of the native translocation domain (TD) of
866 PapC visualized (see Figure 2–figure supplement 1 for the RINs of the mutant TD)
867 with Cytoscape 2.8.2 (Smoot et al., 2011) based on RINalyzer plug-in analysis
868 (Doncheva et al., 2011). The nodes (representing residues) are coloured by
869 structural element as in A. Edges (connecting two residues) are shown in blue, the
870 edge width is proportional to its NCI score from lower to higher values.

871

872 **Figure 3. Evolutionary analysis of PapC TD.** A-B. Sequence conservation
873 calculated with ConSurf (Ashkenazy et al., 2010) and mapped onto the initial model
874 of the native PapC TD (sim1, t=0). Amino acid conservation scores are classified into
875 9 levels. The colour scale for residue conservation goes from cyan (non-conserved:
876 grade 1) to maroon (highly conserved: grade 9), unreliable positions are coloured

877 light yellow. A. Ribbon representation of the model with the highly conserved
878 residues (grade9) shown as spheres and key elements labelled. B. Molecular surface
879 of the model with β 12- β 14 labelled. C-D Sequence co-evolution calculated with
880 PyCogent (Knight et al., 2007, Caporaso et al., 2008). C The co-evolving residues
881 are mapped onto the initial model of the native PapC TD (sim1, t=0). D. The co-
882 evolution network as visualized with Cytoscape 2.8.2 Cytoscape 2.8.2 (Smoot et al.,
883 2011) based on RINalyzer plug-in analysis (Doncheva et al., 2011). Edges
884 (connecting two co-evolved residues) are shown in blue, and nodes (representing
885 coevolved residues) are coloured by structural element. The PD, P-linker1, P-linker2,
886 β -hairpin and α -helix are indicated schematically and coloured as in Figure 2. The
887 node size is proportional to its degree of connectivity.

888

889 **Figure 4. Detection of allosteric hot spots.** A flowchart representing the protocol
890 to identify allosteric hot-spots. First a sub-network of the protein hybrid RIN was
891 generated starting from the α -helix and β -hairpin. Then filters based on the
892 evolutionary information and on the interactions analysis were applied (see Figure 4–
893 figure supplement 1).

894

895 **Figure 5. PapC TD communities.** The hot-spot network communities are shown as
896 surface by colors and indicated schematically (C1 to C5). The inset shows a close up
897 of the identified core residues located in β 7, β 8, the P-linkers, the β -hairpin, the
898 conserved region at the base of the α -helix, in the junction between β 12- β 13_loop.
899 The core residues are labelled in bold and numbered according to the X-ray structure
900 of the apo PapC TD (PDB id: 2vqi).

901

902 **Figure 6. Kinetic signatures of channel activity in wildtype and mutant PapC**
903 **ushers and frequency of PD displacement.** Fifty-second segments of recordings
904 obtained in planar lipid bilayers were selected to illustrate the behaviour of the
905 different proteins. (A) Recording from the wild-type PapC usher showing the
906 characteristic “transient-mixed” behaviour. (B) Recording from the wild-type PapC
907 usher showing an example of spontaneous large openings due to plug displacement.
908 Note the large amount of current fluctuations during the openings, and the “transient-
909 mixed” behaviour in between such events. Examples of similar large openings (C, D)
910 are shown for the V327A, and T331A mutants, respectively. (E) A recording from the
911 K427A shows that the channel barely displays any activity at this voltage. The
912 voltage was +90 mV for all panels. The current level for the closed channels is
913 marked as “C”, and openings are seen as upward deflections of the traces; current

914 levels corresponding to fully open monomeric or dimeric forms are denoted by “M”
915 and “D”, respectively. Note that the traces are plotted as conductance, rather than
916 current, vs. time and the scale bars are given in nS. (F) The percent of sweeps
917 displaying “large-open” behaviour” (LO) indicative of PD displacement is shown for
918 WT and each mutant at the indicated voltages. The number of individual bilayers
919 investigated in each case is given above the bars.

920

921 **Figure 7. Residues involved in the allosteric signalling to control PapC gating.**

922 A schematic model summarizing the location of the detected hot spots involved in the
923 gating mechanism. The β 12- β 13_loop (the ‘latch’) and the β 13 conserved patch are
924 colored in dark grey and light gray respectively. The PD, P-linker1, P-linker2, β -
925 hairpin and α -helix are coloured as in Figure 2. Hot spot residues are colour-coded
926 based on their communities (C1-C4, as in Figure 5). The different symbols indicate
927 the mutant’s electrophysiological behavior (‘X’ where no data were available).
928 Mutants that show a pilus assembly defect or an increased antibiotic sensitivity, or
929 both, are represented by a triangle, pentagon and rectangle, respectively.

930 **Tables**

931 Table 1 Summary of the simulations.

Simulation	Model Systems	Length (ns)
Sim1	Native PapC TD	72
Sim2	Hairpin Mutant	70
Sim3	Helix Mutant	70
Sim4	Helix-Hairpin Mutant	70

932

933 Descriptions of the items are: Simulation, the name of the simulation; model systems,
934 PapC TD model systems simulated; and Length, the length of the simulation.

935 Table 2. Summary of the residue-residue interaction networks (RINs) parameter.

RIN	Full RIN	C	Cr	C/Cr	L	Lr	L/Lr
Native PapC TD	1350 (492)	0.384	0.012	32.00	6.67	3.78	1.76
Hairpin Mutant	1196 (485)	0.368	0.011	33.45	7.20	3.90	1.84
Helix Mutant	1225 (476)	0.362	0.011	32.90	6.67	3.90	1.71
Helix-Hairpin Mutant	854 (466)	0.262	0.008	32.75	8.10	4.70	1.72

936 Descriptions of the items are: RIN residue-residue interaction networks of the
 937 different model systems; Full RIN, number of edges in the RIN, in parenthesis the
 938 number of node; C, average clustering coefficient; L, average shortest path length;
 939 Cr, average clustering coefficient for the random networks with the same size; Lr,
 940 average shortest path length for the random networks with the same size; C/Cr is the
 941 average clustering coefficient ratio (as used in (Atilgan et al, 2004)); L/Lr is the
 942 average shortest path length ratio (as used in (Atilgan et al, 2004)).
 943

944 Table 3 Communities in the hot-spot sub-network

Community	Residues
C1	E247, D249, Y329, L330, T331, G334, Q335, R337, K339, E361, S363, W364, G365, L366, S371, L372
C2	R237, D402, S420, Y441, R442, F443, S444, K468, E469, M470, E475, W496
C3	Y260, Y425, S426, K427, T437, F438, A439
C4	S233, R303, G304, L306, V308, F320, T324, A325, V327
C5	E269, E312, N314, G315, R316, K318

945

946 Descriptions of the items are: Community, the name of the community; Residues,

947 residues that are part of the community.

948
949

Table 4 Analysis of PapC substitution mutants

PapC	Community	HA Titer	Antibiotic Sensitivity		
			SDS	Erythromycin	Vancomycin
WT		64	15	6	6
D249A	C1	0	15	6	6
Y329A	C1	32	15	6	6
T331A	C1	24	15	15	10
R337A	C1	64	15	6	6
S363A	C1	64	15	6	6
R237A	C2	64	16	6	15
S420A	C2	32	15	6	6
R442A	C2	0	15	6	6
S444A	C2	24	14	6	6
Y260A	C3	0	15	14	6
K427A	C3	0	14	6	6
T437A	C3	24	14	6	6
F438A	C3	32	14	12	6
V327A	C4	64	20	14	16

950
951
952
953
954
955
956
957

Descriptions of the items are: PapC, the PapC construct tested; Community, the name of the community to which the mutated residue belongs; HA (hemagglutination assay) Titer, the maximum fold dilution of bacteria able to agglutinate human red blood cells; Antibiotic Sensitivity, the diameter of zone of inhibition (mm) around filter disc impregnated with SDS (750 µg), erythromycin (15 µg), or vancomycin (20 µg). The antibiotic sensitivity measurement includes the filter disc (6 mm diameter).

958 **Figure Supplement Titles and Legends**

959

960 **Table 1–figure supplement 1.** C α -RMSD values for each system throughout the
961 simulations. C α -RMSD from the starting structure (t=0) for the native TD (A), the
962 hairpin mutant (B), helix mutant (C), and helix-hairpin mutant (D).

963

964 **Figure 2–figure supplement 1. Non-covalent interaction network (non-covalent
965 RIN).** Protein structure network representation of the hairpin mutant (A), helix mutant
966 (B), and helix-hairpin mutant (C) translocation domain (TD) of PapC visualized as
967 before.

968

969 **Figure 2–figure supplement 2. Non-covalent interaction network (non-covalent
970 RIN).** RINs showing the difference in non-covalent interaction score (Δ NCl score)
971 between the native TD system and the hairpin mutant (A), helix mutant (B), and
972 helix-hairpin mutant (C). The combined RIN (D) is created by merging the three
973 RINs. The nodes (representing residues) are coloured as in figure 2A. Edges
974 (connecting two residues) are shown in blue, note that in this RIN the edge width is
975 proportional to its Δ NCl score from lower to higher values.

976

977 **Figure 4–figure supplement 1. Contribution of each filter in the detection of
978 allosteric hot spots.** Venn diagrams illustrating the contribution of each filters in the
979 set definition. (A) The dynamic filter raises from the intersection of the difference in
980 non-covalent interaction score between the native TD and each of the mutants. (B)
981 The relative combination of the set from dynamic filter and the set from the
982 evolutionary filter to the final hot spots sub-network.

983

984 **Additional File**

985

986 **Supplementary file 1.** Primers used in this study to generate PapC substitution
987 mutations