Allosteric signalling in the outer membrane translocation domain of PapC usher

Irene Farabella, Thieng Pham, Nadine S Henderson, Sebastian Geibel, Gilles Phan, David G Thanassi, Anne H Delcour, Gabriel Waksman, Maya Topf

DOI: http://dx.doi.org/10.7554/eLife.03532

Cite as: eLife 2014;10.7554/eLife.03532

Received: 30 May 2014
Accepted: 29 September 2014
Published: 1 October 2014

Allosteric Signalling in the Outer Membrane Translocation Domain of PapC Usher

Irene Farabella[a], Thieng Pham[b], Nadine S. Henderson[c], Sebastian Geibel[a,d], Gilles Phan[a,e], David G. Thanassi[c], Anne H. Delcour[b], Gabriel Waksman[a] and Maya Topf[a*]

[a]Institute of Structural and Molecular Biology, Birkbeck College and University College London, University of London, London WC1E 7HX

[b]Department of Biology & Biochemistry, University of Houston, Houston, TX 77204, USA

[c]Center for Infectious Diseases, Department of Molecular Genetics and Microbiology, Stony Brook University, Stony Brook, NY 11794-5120, USA

[d]Current address: Institut für Molekulare Infektionsbiologie, Josef-Schneider-Str. 2/D15, D-97080 Würzburg, Germany

[e]Current address: Université Paris Descartes, Faculté des Sciences, Pharmaceutiques et Biologiques, Laboratoire de Cristallographie et RMN Biologiques, 4 avenue de l'Observatoire, 75 270 Paris cedex 06 France

*Corresponding Author:
Maya Topf, +44 (0)20 7079 0886, m.topf@cryst.bbk.ac.uk

Competing interests

The authors declare that no competing interests exist.

Abstract

PapC ushers are outer-membrane proteins enabling assembly and secretion of P pili in uropathogenic *E. coli*. Their translocation domain is a large β-barrel occluded by a plug domain, which is displaced to allow the translocation of pilus subunits across the membrane. Previous studies suggested that this gating mechanism is controlled by a β-hairpin and an α-helix. To investigate the role of these elements in allosteric signal communication we developed a method combining evolutionary and molecular dynamics studies of the native translocation domain and mutants lacking the β-hairpin and/or α-helix. Analysis of a hybrid residue interaction network suggests distinct regions (residue "communities") within the translocation domain (especially around β12- β14) linking these elements, thereby modulating PapC gating**.** Antibiotic sensitivity and electrophysiology experiments on a set of alanine-substitution mutants confirmed functional roles for four of these communities. This study illuminates the gating mechanism of PapC ushers and its importance in maintaining outer-membrane permeability.

Introduction

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

The biogenesis of pili via the CU pathway is a highly ordered process that comprises sequential steps. The chaperone protein (PapD) brings the pilins to the bacterial OM where they are assembled into a pilus at a transmembrane pore protein known as the usher (PapC). The usher (~800 residues) is composed of 5 domains (Figure 1A): a periplasmic N-terminal domain (NTD), an OM central translocation domain (TD) that comprises a translocation pore domain (TP), interrupted by a conserved Ig-like plug domain (PD), and two domains at the periplasmic C-terminal end (CTD1 and CTD2) (Capitani et al., 2006, Geibel et al., 2013, Ng et al., 2004, Thanassi et al., 2002, Phan et al., 2011). The structure of the apo TD (Figure 1B,C) consists of a 24-stranded kidney-shaped β-barrel where the PD is inserted into the loop connecting two β-strands (β6-β7), occluding the luminal volume of the pore (Remaut et al., 2008, Huang et al., 2009). In the activated form of another archetypal member of the usher family, FimD, the PD is located outside the pore lumen in the periplasm, next to the NTD (Phan et al., 2011, Geibel et al., 2013). In addition to the PD, there are two secondary structure elements that uniquely characterize the large β-barrel structures of the usher TD (Figure 1B). The first element is a *β-hairpin* that creates a large gap in the side of the β-barrel, a feature unprecedented in previously known OM β-barrel structures (Remaut et al., 2008). This element (located between strands β5 and β6 of the barrel, Figure 1C) folds into the barrel lumen and constrains the PD laterally inside the barrel pore. Mutants lacking the β-hairpin show an increased pore permeability suggesting that the β-hairpin has a role in maintaining the PD in a closed conformation (Volkan et al., 2013). The second element is an *α-helix* (located on the loop between β13 and β14, Figure 1B), which caps the β-hairpin from the extracellular side. Mutants lacking the α-helix, or in which the interface between the helix and the PD is disrupted, present a remarkable increase in pore permeability,

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

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

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

Results

To investigate if the β-hairpin or α-helix (or both) of the TD (residues 146-637 in the full length PapC) have a role in the allosteric communication leading to the displacement of the PD (residues 264-324) we performed four independent MD simulations, corresponding to the PapC TD model (sim1, Table 1) and three mutants (sim2-sim4, Table 1) embedded in a mixed lipid bilayer: sim2 where the region corresponding to the hairpin between β5 and β6 (residues 233-240) was deleted,

sim3 where the α-helix between β13 and β14 (residues 447-460) was removed, and sim4 where both regions were removed. The last 50 ns of simulation were considered for analysis, where the averaged root-mean-square deviation of C_α atoms (C_α-RMSD) from the averaged structures stabilized around 2.00±0.09 Å, 1.80±0.09 Å, 1.86±0.11 Å, and 2.03±0.10 Å, for the *native* (sim1), *hairpin mutant* (sim2), *helix mutant* (sim3), and *helix-hairpin mutant* (sim4), respectively (Figure 1–figure supplement 1). This timescale, although limited for a full exploration of the structural changes induced by the mutations, was informative in revealing how local structural perturbations may affect allosteric changes leading to the plug displacement in PapC TD.

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

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

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

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

Evolutionary analysis of PapC TD

We first extracted evolutionary information from a multiple sequence alignment of the PapC TD family. The patterns of conservation in the TD using Consurf (Ashkenazy et al., 2010) analysis suggested that the highly conserved residues (score 9) tend to be clustered in two specific regions of the usher (Figure 3A). The first cluster mapped onto the PD and the P-linkers (P-linker1 residues 248-263; P-linker2 residues 325-335) connecting it to the TP. The second cluster (which included the majority of the highly-conserved residues) mapped onto one side of the TP (strand β1-14 and β24). It includes residues: (i) near the periplasmic side of the β-barrel within β1-4 strands and β24 strand; (ii) on the extracellular side of the barrel (within β5-10); (iii) in the β-hairpin region (β-hairpin and β7-9); and (iv) in the area of β10-14 capped by the α-helix region, which comprises the α-helix and its linkers: H-linker1 (residues 445-450) and H-linker2 (461-468, respectively). Surface representation of the TD reveals a continuous patch of conserved residues facing the lipid bilayer, including β13, the extracellular half of β14 and the periplasmic half of β12 (Figure 3B). Intriguingly, this patch ('β13 conserved patch') reaches the full height of the pore from the α-helix region to a functionally important loop located between β12 and β13 strands (Volkan et al., 2013, Farabella, 2013).

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

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

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

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

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

We analysed the community structure of the hot spot sub-network using the edge-betweenness clustering algorithm (Girvan and Newman, 2002, Morris et al., 2010). This analysis shows that the sub-network has a modular structure, with a modularity index of 0.73 (maximum value of the modularity index is 1), which is typical of 3D-structure based RIN (Newman and Girvan, 2004, Sethi et al., 2009). Here, a total of 11 communities containing two or more residues were identified, from which only five communities are composed by more than 5 residues. For further analysis, we chose to consider only these five largest communities, which are located: between β7-9 and the P-linkers (C1); between the β-hairpin and the conserved region at the base of the

α-helix (β12-14) (C2); between β12-13 loop and the P-linker1 (C3); between the β-hairpin, P-linker2 and the PD (βE-F) (C4); and on the tip of the PD (βE-F loop and βA-B loop) (C5) (Table 3 and Figure 5).

We selected a number of key residues – core hot-spot residues – from the communities that are linking different elements within each community for further experimental investigations (Figure 5). These were found in communities C1-C4: in C1, residues linking the P-linkers and the barrel wall that possibly help in maintaining the P-linkers in a closed configuration (P-linker1:D249, P-linker2:Y329, P-linker2:T331, β7:R337, and β8:S363); in C2, residues that bridge the base of the α-helix (the extracellular end of the β13 conserved patch) and β-hairpin (β-hairpin:R237, β12:S420, β13:R442, β13:S444); in C3, residues on the interface between P-linker1 and β12-13_loop (P-linker1:Y260, β12-β13 loop:K427, β13:T437, β13:F438); and in C4, residues that are part of the interface with P-linker1 and the periplasmic end of the β13 conserved patch (P-linker2:A325, P-linker2:V327).

Experimental Analysis of Residues In The Hot Spot Sub-Network

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

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

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

Electrophysiological analysis of selected mutants

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

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

Discussion

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

Analysis of the mutation-induced perturbation of the non-covalent interaction RIN of the native PapC TD in absence of the β-hairpin, α-helix or both, shows that the interactions between nodes that are not part of the deleted secondary structure elements are consistently weakened (Figure 2- figure supplement 1), suggesting the two elements are not independent. This feature can be interpreted as part of a complex allosteric process regulating the PD gating mechanism. It has been proposed that only a few residues play essential roles during allosteric processes and that perturbing the interactions between these residues can facilitate the population shift of the conformational ensembles (Tsai et al., 2009, del Sol et al., 2009). Additionally, it has been shown that residues with a pivotal role in mediating such allosteric communications are also important residues in terms of evolution (both highly-coevolved residues and conserved residues) (Suel et al., 2003, Tang et al., 2007, Ferguson et al., 2007). Remarkably, we show here that PapC is characterised by an uneven distribution of the evolutionary important residues, clustered in the P-linkers, the PD, and the barrel wall capped by the α-helix, in close proximity to the β-hairpin. Another interesting finding is the presence of the 'β13

conserved patch' that reaches the full height of the pore from the α-helix region to a loop located between β12 and β13 strands (β12-13_loop), which has been previously identified as important (Volkan et al., 2013, Farabella, 2013).

To detect the allosteric network we implement a new method that integrates dynamic and evolutionary information in a hybrid RIN and then apply network analysis. This approach allows us to explore a large part of the protein, resulting in the detection of only 14% of all residues in the TD as potential candidates. It has been shown that detecting "residue communities" in protein structure networks leads to the identification of key residues that are often part of a signal transduction pathway (del Sol et al., 2007, Bode et al., 2007). The interconnection within and between the communities is pivotal for the flow of allosteric signalling. Residues in the same community are densely interconnected and have multiple routes to communicate with each another. However, the interconnections between communities involve only a few edges, which form the bottleneck for the flow of the signal in the network (del Sol et al., 2007, Bode et al., 2007, Sethi et al., 2009). Here, all the identified communities (C1-C5, Figure 5) comprise residues from multiple elements (e.g. β-hairpin, P-linker1, P-linker2 and distinct part of the TP) of PapC TD except C5 that is composed only by PD residues. Mutation of key residues linking elements within each of the four communities C1-C4 showed an altered antibiotic sensitivity phenotype, confirming a role of these communities in the pore gating mechanism. Additionally, communities C1-C3 are required for proper usher function (as estimated by the hemagglutination assay, suggesting dependency of the pore gating function and pilus assembly (based on some of the mutations). For example, mutations in β12-13_loop:K427 in C3 and β13:S444 in C2, which lead to a drastic decrease in the frequency of plug displacement (stabilizing the closed state, based on channel activity analysis) also show defective pilus assembly (possibly due to their deficiency in relocating the plug in a functional conformation).

Additionally, our study identifies two residues with opposite effects on plug displacement in the same community: C3: β12-13_loop:K427, which is located on β12-13_loop, and β13:F438, which is located at the periplasmic end of the β13 conserved patch (linking β12-13_loop with the α-helix region). This observation suggests that β12-13_loop has a pivotal role in modulating plug displacement (acting as a 'latch') as proposed previously (Farabella, 2013), and supported by additional mutagenesis studies (Volkan et al., 2013). Intriguingly, at the extracellular end of the β13 conserved patch there is another mutation (β13:S444) leading to a channel more

reluctant to open, suggesting a regulatory role for the β13 conserved patch (Figure 7) in modulating the 'latch' (β12-13_loop).

Interestingly, our community residues in some cases are found to have very different patterns of interactions in an alternative conformation. For example, the functionally important 'latch' (β12-13_loop) is shown to be in a different conformation in the open state (based on FimD:FimC:FimH structure (Phan et al., 2011)) as well as other residues in C3 and C4 communities that change their interaction pattern in the open state. As a result, some residues that are identified by our method, such as V327, could in principle be selected also by visual inspection for further experimental investigation. V327 (located on P-linker2 in C4 at interface with P-linker1 and the barrel wall) is found to have a different interaction pattern in the open state of the usher and is also shown to promote plug displacement. However, more importantly, the method is able to predict residues in regions that would not attract our attention at all but may contain important information in respect to the allosteric pathway. (This aspect can become even more significant in the absent of an alternative conformation). For example, we predict two such residues to be functionally important –T331 and S444. T331 is located on P-linker2 at the interface between P-linker2 and the barrel wall, which stays intact in the open conformation of the usher; S444 is located on β13 at the base of the α-helix, i.e. on the barrel itself, and it lacks any direct contact with the plug domain. However, both residues were shown to have an effect on plug displacement. Thus, the main strength of our method is that the knowledge of the structure of the protein in one conformation only and enough sequence information to extract evolutionary information are sufficient for the detection of functionally important residues that can be pivotal for transferring the regulatory information within the protein.

Conclusion

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

confirmed that residues located in the P-linkers, the β-hairpin, and β13 conserved patch (part of four communities) alter channel gating and that residues located in P-linker2, β12-β13_loop and β13 conserved patch (both periplasmic end and extracellular end) are sensitive to plug displacement. Therefore, we suggest that the β13 conserved patch acts as a regulator of the 'latch' (β12-β13_loop) mediating channel opening. Furthermore, our study shows how the integration of different computational approaches based on evolution, structure and dynamics of proteins, into a hybrid network can unveil communication pathways within proteins. Such an integrative approach can guide the experimental investigation by pinpointing key candidates involved in the transmission of the allosteric signal.

Methods

Systems Modelling

We built four model systems based on the X-ray structure of the TD (residues 1-492) of PapC at 3.2 Å resolution (PDB ID: 2vqi (Remaut et al., 2008)). The starting model for the simulation of the native TD (sim1, native) was generated by adding the missing loops to the X-ray structure using the *dope_loopmodel* method (Shen and Sali, 2006) in MODELLER-9v7 (Sali and Blundell, 1993). Additionally, three mutant model systems were constructed based on the native model: a mutant that lacking the β-hairpin (Δ233-240) (sim2, hairpin mutant); a mutant lacking the α-helix (Δ447-460) (sim3, helix mutant) and a mutant lacking both (Δ233-240 and Δ447-460) (sim4, helix-hairpin mutant).

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

Systems Set Up For MD Simulations

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

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

Equilibration Procedure and Production Run

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

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

Non-covalent residue interaction network

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

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

Sequence Conservation Analysis

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

Sequence Coevolution Analysis

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

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

Hybrid RIN

To store the entire information we combined the coevolutionary RIN, the non-covalent RIN, and the conservation analysis into one network. In this hybrid network 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 ΔNCI).
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 (ΔNCI>0) in each of the mutant systems. Intersecting the identified sets
567 resulted in a 'hot spots' sub-network. The 'hot spots' sub-network was then
568 decomposed into communities using the edge-betweenness clustering algorithm
569 (GLay) as implemented in clusterMaker (Girvan and Newman, 2002, Morris et al.,
570 2010). Cytoscape 2.8.2 (Smoot et al., 2011), RINalyzer plug-in (Doncheva et al.,
571 2011), and Chimera (Pettersen et al., 2004) were used for mapping the network on
572 the PapC structure.
573
574 *PapC substitution mutants*
575 The PapC alanine substitution mutants were derived from plasmid pDG2 using the
576 QuikChange Site-Directed Mutagenesis Kit (Stratagene) and the primers listed in
577 Supplementary file 1. Plasmid pDG2 encodes wild-type *papC* with a C-terminal,
578 thrombin-cleavable His-tag (Li et al., 2004). All mutants were sequenced to verify the
579 intended mutation.
580
581 *Expression and folding of the PapC substitution mutants in the outer membrane*
582 Each of the PapC mutants was compared with wild-type PapC for expression levels
583 and ability to fold into a stable β-barrel in the OM. OM isolation, analysis of usher
584 protein levels, and the heat-modifiable mobility assay for β-barrel stability were done
585 as previously described (Henderson et al., 2011).
586
587 *Hemagglutination assay*

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

Top soft agar assay for antibiotic sensitivity

Bacteria were grown in LB medium supplemented with 100 µg/ml ampicillin (Amp) to an OD_{600} of 0.6 and then induced for PapC expression with 0.1% arabinose for 1 h. Aliquots of 0.1 ml bacteria were added to 3 ml melted soft top agar (0.75% LB agar) cooled to 45ºC and supplemented with 100 µg/ml Amp and 0.1% arabinose. The bacteria and melted agar were mixed well and poured on top of 1.5% solid LB agar plates containing 100 µg/ml Amp and 0.1% arabinose. Once the top agar solidified, sterile 6 mm filter discs were placed on top and 10 µl of the following antibiotics were added: 75 mg/ml SDS, 2 mg/ml vancomycin, or 1.5 mg/ml erythromycin. The diameter of the growth inhibition zone around the antibiotics, including the filter disc, was measured after overnight growth at 37°C. Each PapC mutant strain was tested twice and the values averaged.

Electrophysiological analysis of selected PapC substitution mutants

PapC mutants (R237A, F438A, V327A, K427A, T437A, T331A and S444A) were purified according to published protocols (Henderson and Thanassi, 2013) and investigated by planar lipid bilayer electrophysiology. Planar bilayers were made from a preparation of L-α-phosphatidylcholine Type II-S from Sigma (also known as asolectin) according to the Montal and Mueller technique (Montal and Mueller, 1972) following a published protocol (Mapingire et al., 2013). Protein aliquots were diluted 1:1 in buffer T (1M KCl, 5 mM Hepes, pH 7.2) containing either 1 or 2% *N*□Octyl□oligo□oxyethylene (octyl-POE, Axxora). Eight micrograms of protein from the diluted sample was added to the *cis* side of a planar lipid bilayer chamber containing ~ 1.5 mL of buffer T. Gentle stirring was applied to promote spontaneous insertions of the protein into the bilayer. Channel activity was monitored by measuring current under voltage clamp conditions using an Axopatch 1D amplifier with a CV4B headstage or an Axopatch 200B amplifier (Axon Instruments). The current was digitized (ITC-18, Instrutech), and stored on a PC computer using the Acquire software (Bruxton). Ten-minute long traces were sampled at 1.25 ms intervals and filtered at 500 Hz. Both chambers contained buffer T and Ag/AgCl

electrodes with pellet. The *trans* side of the bilayer was set as ground. Insertions were typically performed at -90 mV. Data display and analysis were done with pCLAMP software (Axon Instruments).

Acknowledgments

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

References

ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res,* 25**,** 3389-402.

ARMOUGOM, F., MORETTI, S., POIROT, O., AUDIC, S., DUMAS, P., SCHAELI, B., KEDUAS, V. & NOTREDAME, C. 2006. Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res,* 34**,** W604-8.

ASHKENAZY, H., EREZ, E., MARTZ, E., PUPKO, T. & BEN-TAL, N. 2010. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res,* 38**,** W529-33.

ASSENOV, Y., RAMIREZ, F., SCHELHORN, S. E., LENGAUER, T. & ALBRECHT, M. 2008. Computing topological parameters of biological networks. *Bioinformatics,* 24**,** 282-4.

ATILGAN, A. R., AKAN, P. & BAYSAL, C. 2004. Small-world communication of residues and significance for protein dynamics. *Biophys J,* 86**,** 85-91.

BERGER, O., EDHOLM, O. & JAHNIG, F. 1997. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys J,* 72**,** 2002-13.

BODE, C., KOVACS, I. A., SZALAY, M. S., PALOTAI, R., KORCSMAROS, T. & CSERMELY, P. 2007. Network analysis of protein dynamics. *FEBS Lett,* 581**,** 2776-82.

BUSSI, G., DONADIO, D. & PARRINELLO, M. 2007. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics,* 126.

CAPITANI, G., EIDAM, O., GLOCKSHUBER, R. & GRUTTER, M. G. 2006. Structural and functional insights into the assembly of type 1 pili from Escherichia coli. *Microbes Infect,* 8**,** 2284-90.

CAPORASO, J. G., SMIT, S., EASTON, B. C., HUNTER, L., HUTTLEY, G. A. & KNIGHT, R. 2008. Detecting coevolution without phylogenetic trees? Tree-ignorant metrics of coevolution perform as well as tree-aware metrics. *BMC Evol Biol,* 8**,** 327.

DAILY, M. D., UPADHYAYA, T. J. & GRAY, J. J. 2008. Contact rearrangements form coupled networks from local motions in allosteric proteins. *Proteins,* 71**,** 455-66.

DARDEN, T., YORK, D. & PEDERSEN, L. 1993. Particle mesh Ewald: An N□log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics,* 98.

DEL SOL, A., ARAUZO-BRAVO, M. J., AMOROS, D. & NUSSINOV, R. 2007. Modular architecture of protein structures and allosteric communications: potential implications for signaling proteins and regulatory linkages. *Genome Biol,* 8**,** R92.

DEL SOL, A., TSAI, C. J., MA, B. & NUSSINOV, R. 2009. The origin of allosteric functional modulation: multiple pre-existing pathways. *Structure,* 17**,** 1042-50.

DONCHEVA, N. T., KLEIN, K., DOMINGUES, F. S. & ALBRECHT, M. 2011. Analyzing and visualizing residue networks of protein structures. *Trends Biochem Sci,* 36**,** 179-82.

FARABELLA, I. Feb 2013. *Computational studies of the usher outer-membrane assembly platform involved in pilus biogenesis.* Doctor of Philosophy, Birkbeck College, University of London.

FERGUSON, A. D., AMEZCUA, C. A., HALABI, N. M., CHELLIAH, Y., ROSEN, M. K., RANGANATHAN, R. & DEISENHOFER, J. 2007. Signal transduction

pathway of TonB-dependent transporters. *Proc Natl Acad Sci U S A,* 104**,** 513-8.
GEIBEL, S., PROCKO, E., HULTGREN, S. J., BAKER, D. & WAKSMAN, G. 2013. Structural and energetic basis of folded-protein transport by the FimD usher. *Nature,* 496**,** 243-6.
GIRVAN, M. & NEWMAN, M. E. 2002. Community structure in social and biological networks. *Proc Natl Acad Sci U S A,* 99**,** 7821-6.
GLOOR, G. B., MARTIN, L. C., WAHL, L. M. & DUNN, S. D. 2005. Mutual information in protein multiple sequence alignments reveals two classes of coevolving positions. *Biochemistry,* 44**,** 7156-65.
HAIYAN, L. & JIHUA, W. Network Properties of Protein Structures at Three Different Length Scales. Proceedings of the 2009 Second Pacific-Asia Conference on Web Mining and Web-based Application %@ 978-0-7695-3646-0, 2009. IEEE Computer Society, 161-164.
HENDERSON, N. S., NG, T. W., TALUKDER, I. & THANASSI, D. G. 2011. Function of the usher N-terminus in catalysing pilus assembly. *Mol. Microbiol.,* 79**,** 954-67.
HENDERSON, N. S. & THANASSI, D. G. 2013. Purification of the outer membrane usher protein and periplasmic chaperone-subunit complexes from the P and type 1 pilus systems. *Methods Mol Biol,* 966**,** 37-52.
HESS, B. 2008. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.,* 4**,** 116-122.
HESS, B., BEKKER, H., BERENDSEN, H. J. C. & FRAAIJE, J. G. E. M. 1997. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.,* 18**,** 1463-1472.
HOOVER, W. G. 1985. Canonical dynamics: Equilibrium phase-space distributions. *Phys Rev A,* 31**,** 1695-1697.
HUANG, Y., SMITH, B. S., CHEN, L. X., BAXTER, R. H. & DEISENHOFER, J. 2009. Insights into pilus assembly and secretion from the structure and functional characterization of usher PapC. *Proc Natl Acad Sci U S A,* 106**,** 7403-7.
JO, S., KIM, T. & IM, W. 2007. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PloS one,* 2**,** e880.
JORGENSEN, W. L., CHANDRASEKHAR, J., MADURA, J. D., IMPEY, R. W. & KLEIN, M. L. 1983. Comparison of simple potential functions for simulating liquid water. *J Chem Phys,* 79**,** 926.
KAMINSKI, G. A., FRIESNER, R. A., TIRADO-RIVES, J. & JORGENSEN, W. L. 2001. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B. ,* 105**,** 6474–6487.
KANDT, C., ASH, W. L. & TIELEMAN, D. P. 2007. Setting up and running molecular dynamics simulations of membrane proteins. *Methods,* 41**,** 475-88.
KNIGHT, R., MAXWELL, P., BIRMINGHAM, A., CARNES, J., CAPORASO, J. G., EASTON, B. C., EATON, M., HAMADY, M., LINDSAY, H., LIU, Z., LOZUPONE, C., MCDONALD, D., ROBESON, M., SAMMUT, R., SMIT, S., WAKEFIELD, M. J., WIDMANN, J., WIKMAN, S., WILSON, S., YING, H. & HUTTLEY, G. A. 2007. PyCogent: a toolkit for making sense from sequence. *Genome Biol,* 8**,** R171.
LI, H., QIAN, L., CHEN, Z., THAHBOT, D., LIU, G., LIU, T. & THANASSI, D. G. 2004. The outer membrane usher forms a twin-pore secretion complex. *J Mol Biol,* 344**,** 1397-1407.
LIU, Y. & BAHAR, I. 2012. Sequence evolution correlates with structural dynamics. *Mol Biol Evol,* 29**,** 2253-63.
LOMIZE, M. A., LOMIZE, A. L., POGOZHEVA, I. D. & MOSBERG, H. I. 2006. OPM: orientations of proteins in membranes database. *Bioinformatics,* 22**,** 623-5.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

SETHI, A., EARGLE, J., BLACK, A. A. & LUTHEY-SCHULTEN, Z. 2009. Dynamical networks in tRNA:protein complexes. *Proc Natl Acad Sci U S A,* 106**,** 6620-5.

SHEN, M. Y. & SALI, A. 2006. Statistical potential for assessment and prediction of protein structures. *Protein Sci,* 15**,** 2507-24.

SMOOT, M. E., ONO, K., RUSCHEINSKI, J., WANG, P. L. & IDEKER, T. 2011. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics,* 27**,** 431-2.

SUEL, G. M., LOCKLESS, S. W., WALL, M. A. & RANGANATHAN, R. 2003. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Biol,* 10**,** 59-69.

TANG, S., LIAO, J. C., DUNN, A. R., ALTMAN, R. B., SPUDICH, J. A. & SCHMIDT, J. P. 2007. Predicting allosteric communication in myosin via a pathway of conserved residues. *J Mol Biol,* 373**,** 1361-73.

TAYLOR, N. R. 2013. Small world network strategies for studying protein structures and binding. *Comput Struct Biotechnol J,* 5**,** e201302006.

THANASSI, D. G., SAULINO, E. T. & HULTGREN, S. J. 1998. The chaperone/usher pathway: a major terminal branch of the general secretory pathway. *Curr Opin Microbiol,* 1**,** 223-31.

THANASSI, D. G., STATHOPOULOS, C., DODSON, K., GEIGER, D. & HULTGREN, S. J. 2002. Bacterial outer membrane ushers contain distinct targeting and assembly domains for pilus biogenesis. *J Bacteriol,* 184**,** 6260-9.

TSAI, C. J., DEL SOL, A. & NUSSINOV, R. 2009. Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. *Mol Biosyst,* 5**,** 207-16.

VAN DER SPOEL, D., LINDAHL, E., HESS, B., GROENHOF, G., MARK, A. E. & BERENDSEN, H. J. 2005. GROMACS: fast, flexible, and free. *J Comput Chem,* 26**,** 1701-18.

VOLKAN, E., KALAS, V., PINKNER, J. S., DODSON, K. W., HENDERSON, N. S., PHAM, T., WAKSMAN, G., DELCOUR, A. H., THANASSI, D. G. & HULTGREN, S. J. 2013. Molecular basis of usher pore gating in Escherichia coli pilus biogenesis. *Proc Natl Acad Sci U S A,* 110**,** 20741-6.

WAKSMAN, G. & HULTGREN, S. J. 2009. Structural biology of the chaperone-usher pathway of pilus biogenesis. *Nat Rev Microbiol,* 7**,** 765-74.

Figure Legend

Figure 1. PapC usher organization and detail of its translocation domain. A. A diagram of the domain organization of PapC usher. NTD (dark-blue) represents the N-terminal domain, CTD1 (light-violet) and CTD2 (dark-violet) represent the C-terminal domains; TD represents the translocation domain, comprising the TP (translocation pore, light-blue) and the PD (plug domain, magenta). B and C: Ribbon representation of the starting model of the native translocation domain (TD) of PapC with the labels 'N' and 'C' indicating the N and C termini of the translocation channel. The β-barrel, PD (including the P-linkers), β-hairpin, and α-helix (including the H-linkers) are coloured blue, magenta, orange, and yellow, respectively. The outer membrane position is represented schematically with the labels 'E', 'M' and 'P' indicating the extracellular side, the membrane and the periplasmic side, respectively. Side view of the TD (B) is shown with the α-helix, β-hairpin, H-linker1, H-linker2, P-linker1, P-linker2, and PD, labelled. Extracellular top view of the TD (C) is shown with the barrel β strands labelled β1 through β24 and with the PD strands labelled βA through βF. The figures were created with Chimera (Pettersen et al., 2004).

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

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

light yellow. A. Ribbon representation of the model with the highly conserved residues (grade9) shown as spheres and key elements labelled. B. Molecular surface of the model with β12-β14 labelled. C-D Sequence co-evolution calculated with PyCogent (Knight et al., 2007, Caporaso et al., 2008). C The co-evolving residues are mapped onto the initial model of the native PapC TD (sim1, t=0). D. The co-evolution network as visualized with Cytoscape 2.8.2 Cytoscape 2.8.2 (Smoot et al., 2011) based on RINalyzer plug-in analysis (Doncheva et al., 2011). Edges (connecting two co-evolved residues) are shown in blue, and nodes (representing coevolved residues) are coloured by structural element. The PD, P-linker1, P-linker2, β-hairpin and α-helix are indicated schematically and coloured as in Figure 2. The node size is proportional to its degree of connectivity.

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

Figure 5. PapC TD communities. The hot-spot network communities are shown as surface by colors and indicated schematically (C1 to C5). The inset shows a close up of the identified core residues located in β7, β8, the P-linkers, the β-hairpin, the conserved region at the base of the α-helix, in the junction between β12-β13_loop. The core residues are labelled in bold and numbered according to the X-ray structure of the apo PapC TD (PDB id: 2vqi).

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

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

Figure 7. Residues involved in the allosteric signalling to control PapC gating. A schematic model summarizing the location of the detected hot spots involved in the gating mechanism. The β12-β13_loop (the ‘latch’) and the β13 conserved patch are colored in dark grey and light gray respectively. The PD, P-linker1, P-linker2, β-hairpin and α-helix are coloured as in Figure 2. Hot spot residues are colour-coded based on their communities (C1-C4, as in Figure 5). The different symbols indicate the mutant’s electrophysiological behavior (‘X’ where no data were available). Mutants that show a pilus assembly defect or an increased antibiotic sensitivity, or both, are represented by a triangle, pentagon and rectangle, respectively.

Tables

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

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

929 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

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

938 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

939

940 Descriptions of the items are: Community, the name of the community; Residues,
941 residues that are part of the community.

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

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).

Figure Supplement Titles and Legends

Figure 1–figure supplement 1. (A) Cutaway view across the membrane plane of the native PapC TD starting model in a POPE/POPG lipid bilayer (sim1, t=0). Molecular surface of PapC TD is coloured as in Figure 1, the lipid are shown in grey with the lipid head group coloured by element, the water is coloured by element and the ions (the Na^+ in blue and the CL^- in yellow) are represented as sphere. The Cα-RMSD values for each system from the starting structure (t=0) for the native TD (B), the hairpin mutant (C), helix mutant (D), and helix-hairpin mutant (E) are plotted as a function of time.

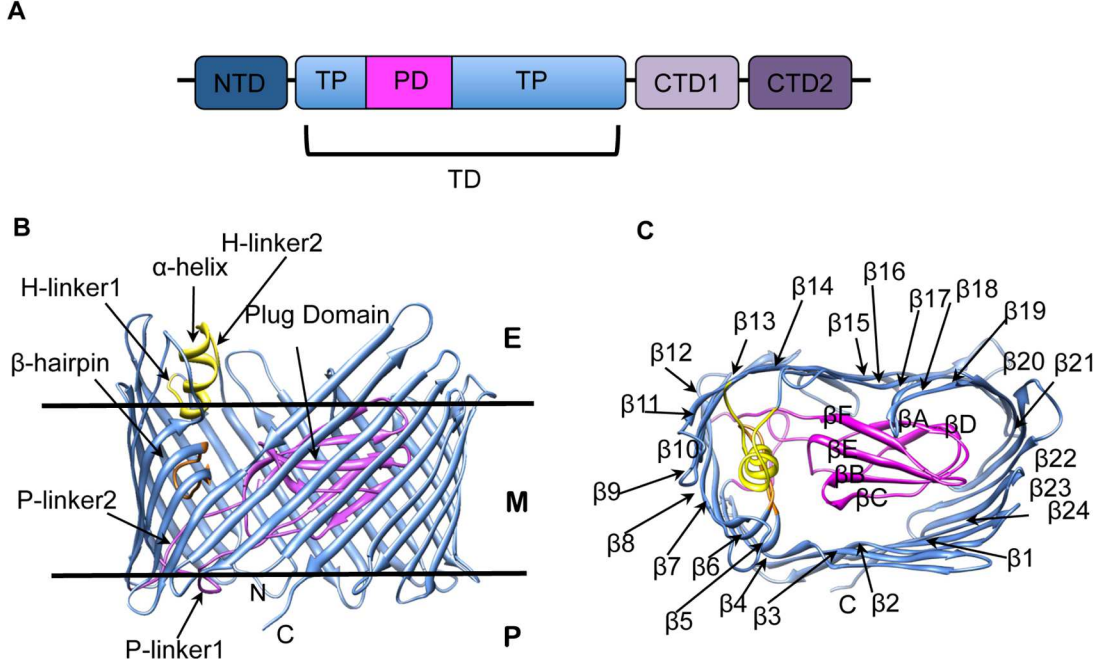
Figure 2–figure supplement 1. Non-covalent interaction network (non-covalent RIN). Protein structure network representation of the hairpin mutant (A), helix mutant (B), and helix-hairpin mutant (C) translocation domain (TD) of PapC visualized as in figure 2.

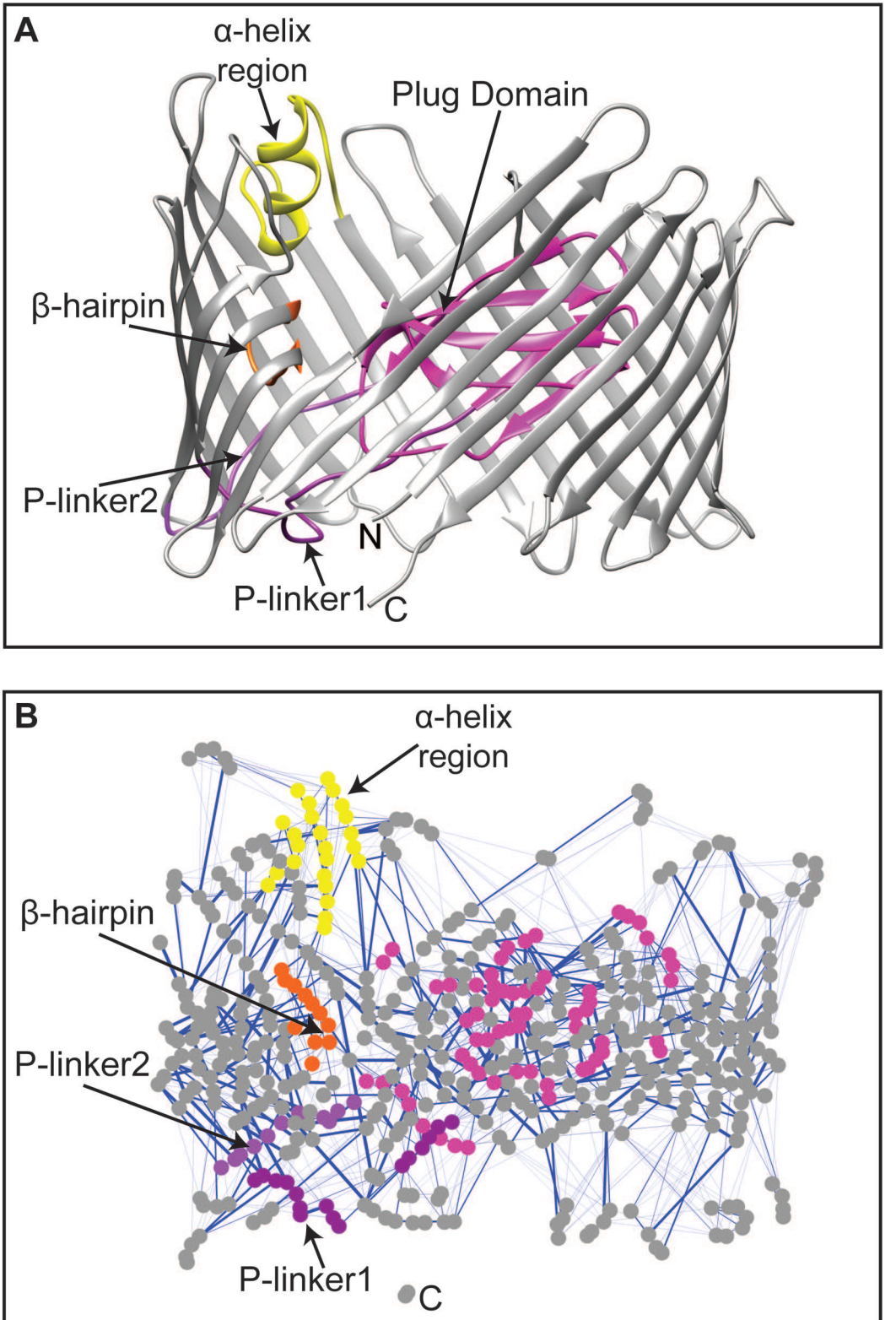
Figure 2–figure supplement 2. Non-covalent interaction network (non-covalent RIN). RINs showing the difference in non-covalent interaction score (ΔNCI score) between the native TD system and the hairpin mutant (A), helix mutant (B), and helix-hairpin mutant (C). The combined RIN (D) was created by merging the three RINs. The nodes (representing residues) are coloured as in figure 2A. Edges (connecting two residues) are shown in blue, with edge width proportional to its corresponding ΔNCI score (from lower to higher values).

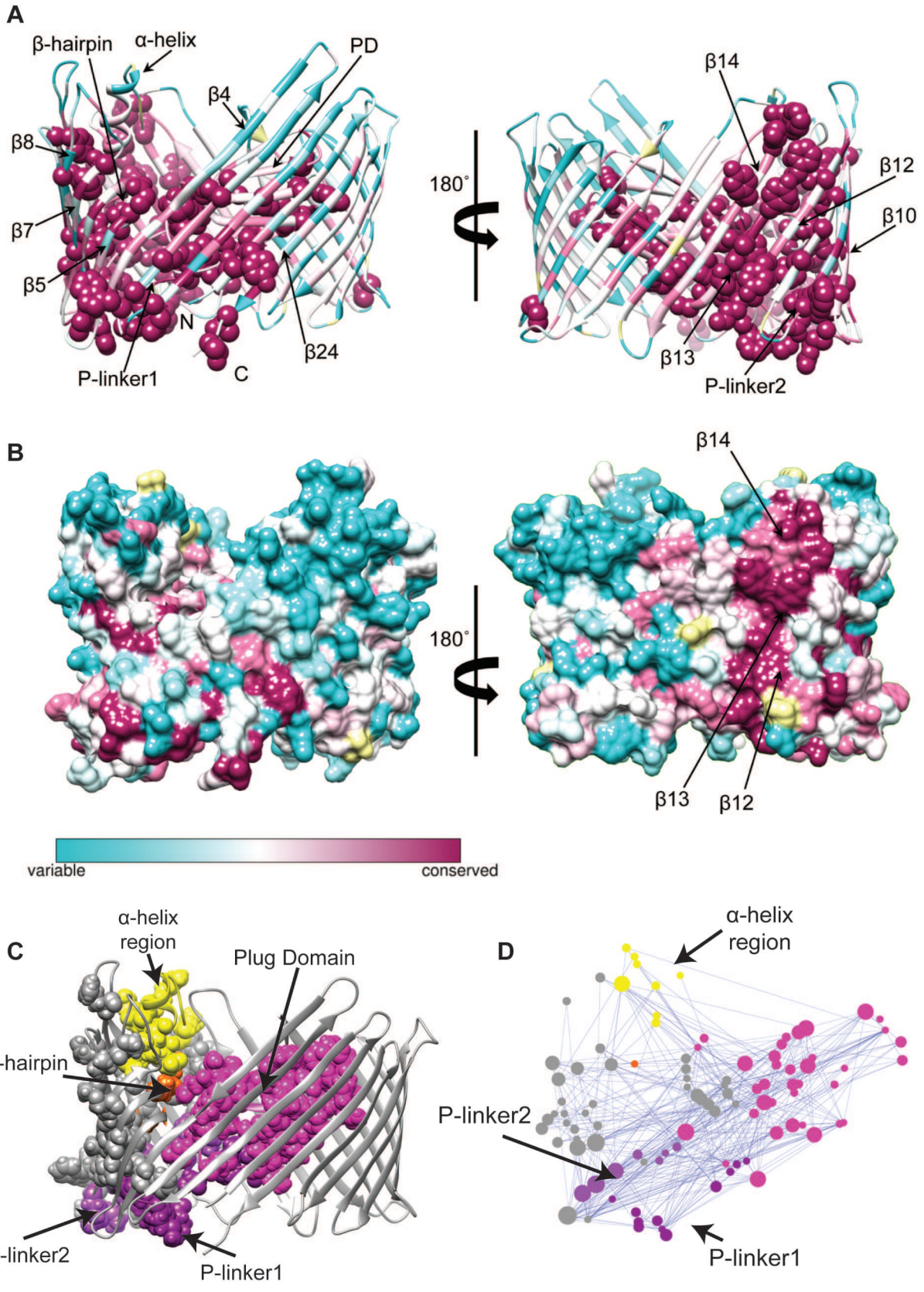
Figure 4–figure supplement 1. Contribution of each filter in the detection of allosteric hot spots. Venn diagrams illustrating the contribution of each filter in the set definition. (A) The dynamic filter resulting from the intersection of the difference in non-covalent interaction score between the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

Additional File

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


A
NTD
TP
PD
TP
CTD1
CTD2
TD
B
H-linker2
α-helix
H-linker1
Plug Domain
β-hairpin
E
M
P
P-linker2
N
C
P-linker1
C
β16
β14
β17
β18
β19
β13
β15
β12
β20
β21
β11
βF
βA
βD
β10
βE
β22
βB
β9
βC
β23
β24
β8
β1
β7
β6
β5
β4
β3
C
β2

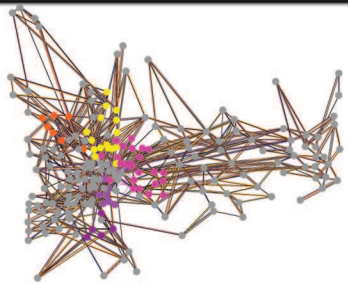

A
α-helix region
Plug Domain
β-hairpin
P-linker2
N
P-linker1
C
B
α-helix region
β-hairpin
P-linker2
P-linker1
C


A
β-hairpin
α-helix
PD
β4
β8
β7
β5
N
C
β24
P-linker1
180˚
β14
β12
β10
β13
P-linker2
B
180˚
β14
β13
β12
variable
conserved
C
α-helix region
Plug Domain
β-hairpin
P-linker2
P-linker1
D
α-helix region
P-linker2
P-linker1

α-helix

β-hairpin

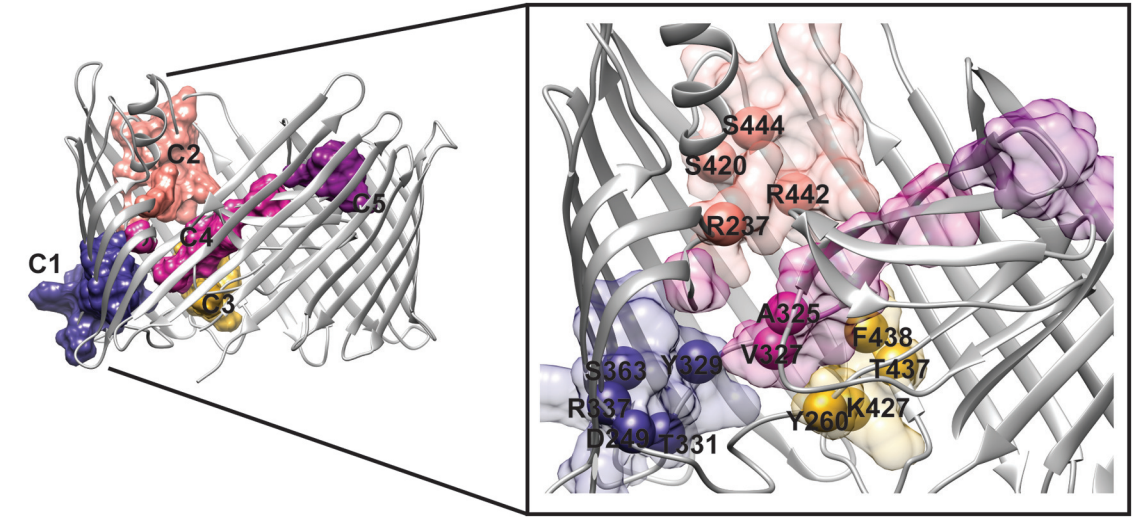
Hybrid sub-network reconstruction

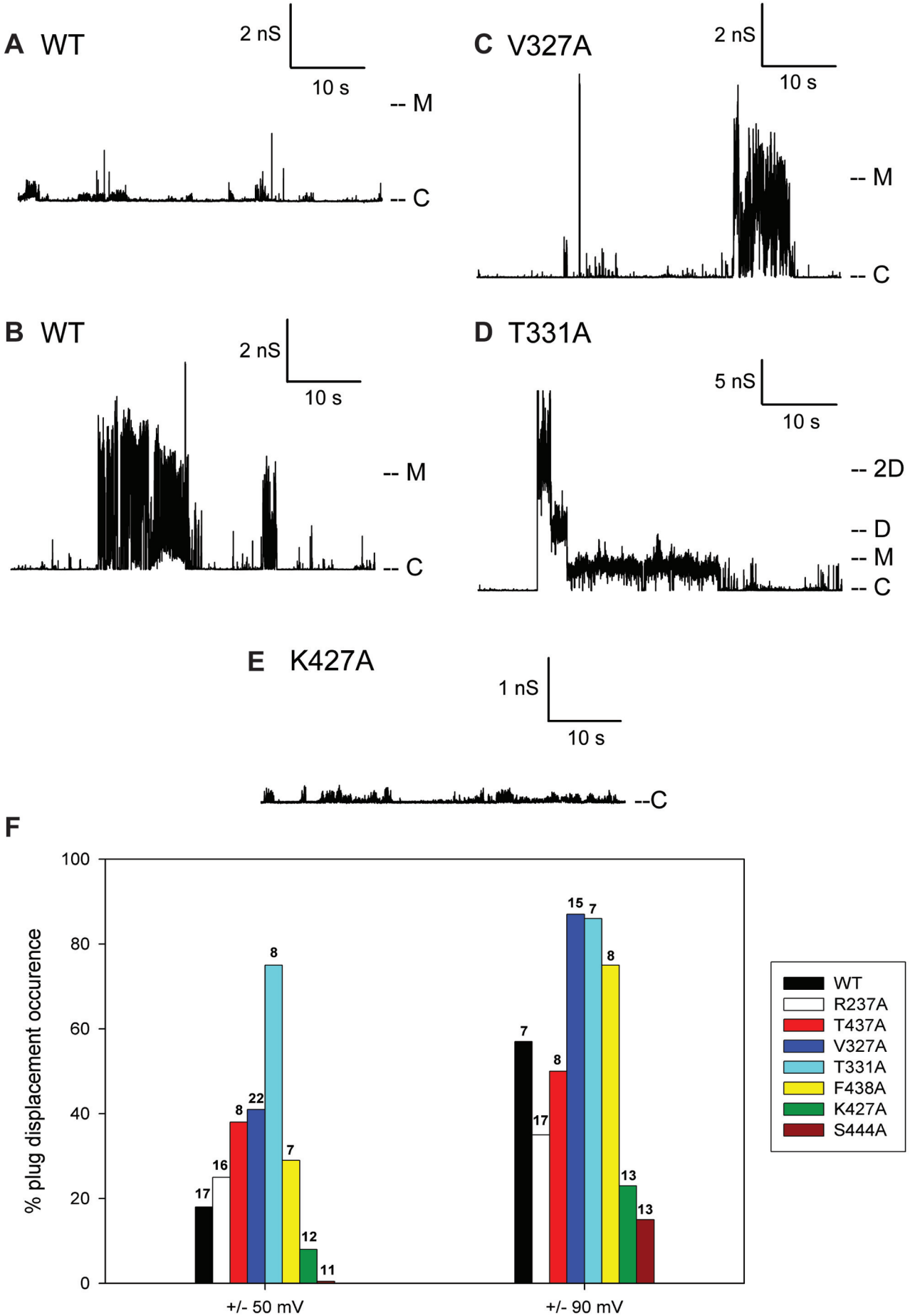


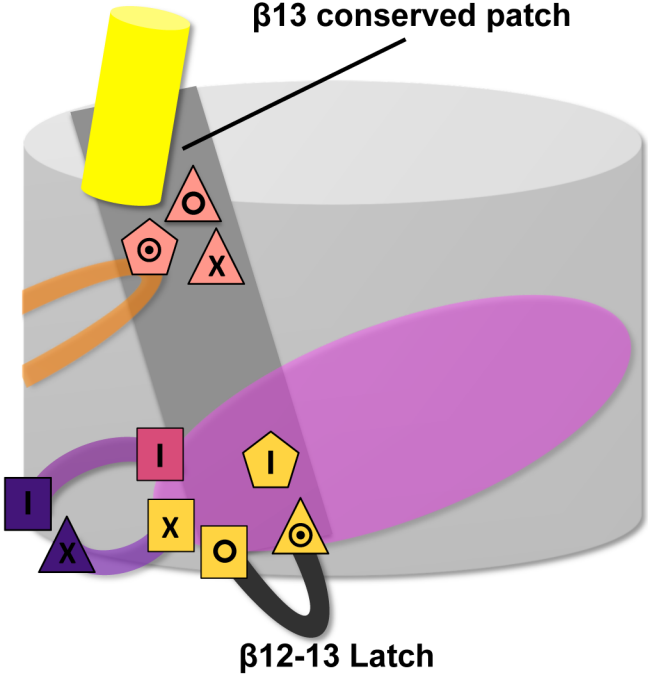
Evolutionary information

Weakened interaction

'Hot Spot' sub-network


C1
C2
C3
C4
C5
S444
S420
R442
R237
A325
F438
V327
T437
S363
Y329
R337
K427
Y260
D249
T331


A WT
2 nS
10 s
-- M
-- C
C V327A
2 nS
10 s
-- M
-- C
B WT
2 nS
10 s
-- M
-- C
D T331A
5 nS
10 s
-- 2D
-- D
-- M
-- C
E K427A
1 nS
10 s
--C
F
% plug displacement occurence
100
80
60
40
20
0
17
16
8
22
8
7
12
11
7
17
8
15
7
8
13
13
+/- 50 mV
+/- 90 mV
WT
R237A
T437A
V327A
T331A
F438A
K427A
S444A



I Large opening

O Absence of large opening

⊙ Transient mixed

X No electrophysiology data

Pilus assembly defect

Increase in antibiotic sensitivity

Pilus assembly defect and increase in antibiotic sensitivity

Community 1

Community 2

Community 3

Community 4