The C-terminal head domain of *Burkholderia pseudomallei* BpaC has a striking hydrophilic core with an extensive solvent network

Andreas R. Kiessling | Sarah A. Harris | Kathleen M. Weimer | Geoffrey Wells | Adrian Goldman

**Abstract**

Gram-negative pathogens like *Burkholderia pseudomallei* use trimeric autotransporter adhesins such as BpaC as key molecules in their pathogenicity. Our 1.4 Å crystal structure of the membrane-proximal part of the BpaC head domain shows that the domain is exclusively made of left-handed parallel β-roll repeats. This, the largest such structure solved, has two unique features. First, the core, rather than being composed of the canonical hydrophobic Ile and Val, is made up primarily of the hydrophilic Thr and Asn, with two different solvent channels. Second, comparing BpaC to all other left-handed parallel β-roll structures showed that the position of the head domain in the protein correlates with the number and type of charged residues. In BpaC, only negatively charged residues face the solvent—in stark contrast to the primarily positive surface charge of the left-handed parallel β-roll "type" protein, YadA. We propose extending the definitions of these head domains to include the BpaC-like head domain as a separate subtype, based on its unusual sequence, position, and charge. We speculate that the function of left-handed parallel β-roll structures may differ depending on their position in the structure.

**KEYWORDS**

bacterial adhesin, bacterial outer membrane proteins, *Burkholderia pseudomallei*, melioidosis, protein conformation, Type V secretion systems, β-sheet

1 | **INTRODUCTION**

*Burkholderia pseudomallei*, an aerobic Gram-negative soil-dwelling bacterial pathogen endemic in areas of Southeast Asia and Northern Australia, causes a wide variety of acute and latent diseases in humans. Acute infections (melioidosis causing a septic shock) can have mortality rates as high as 50%, and the bacterium is resistant to many front-line antibiotics. The main route of infection is via aerosols after...
contact with an infected horse. Because of the low dose required for infection and the high propensity for aerosol formation, *B. pseudomallei* is considered a "class B" potential bioweapon (Wiersinga et al., 2006).

*B. pseudomallei* is an intracellular pathogen and escapes phagocytic digestion using a Type III secretion system (Gong et al., 2011). Once free, it spreads via intercellular fusion, thereby evading immune recognition (Burtnick et al., 2011). Type V secretion systems, including Type Vc trimeric autotransporter adhesins (TAA s) have been shown to be important virulence factors associated with adhesion and/or immune evasion. These include BoaA/B (Balder et al., 2010) and BpaA/C/D/E (Lazar Adler et al., 2015). *B. pseudomallei* can adapt to different ecological and host niches due to changes in gene and protein expression that alter factors like membrane composition, essential metabolism, and virulence (Duangurai et al., 2018); this may be important in understanding the different lengths and domain organizations of TAA s in *B. pseudomallei*. This also implies a complex interplay of different adhesins targeting a set of cells with high specificity: the adhesins expressed depend on signals from the environment, whether intracellular or extracellular.

The TAA BpaC was first extensively described in 2013 (Campos et al., 2013); of the nine predicted *B. pseudomallei* TAA s, BpaC is the only one associated with all of the three features involved in pathogenicity: macrophage survival, virulence, and serum survival (Lazar Adler et al., 2015). Studies have demonstrated BpaC adhesion to respiratory epithelial cell lines, establishing the importance of BpaC in the initial attachment and tropism of the pathogen (Lafontaine et al., 2014). BpaC thus represents a valid target for investigation as a potential drug target. The potential mechanisms of BpaC can be inferred from the roles reported for other TAA s: adherence, invasion, serum resistance, and biofilm formation (for a review, see Kiessling et al., 2020]). In general, TAA s bind many different partners such as parts of the extracellular matrix (Vaca et al., 2020), complement system down-regulators like C4b-binding protein (Hovingh et al., 2016), and specific receptors like the human carcinobryonic antigen-related cell adhesion molecule 1, which binds UspA1 (Conners et al., 2008).

The modularity of TAA domains in the solvent-accessible region of the protein (the passenger domain), and the structural constraints imposed by their trimERIC nature, enable the combination and diversification of protein function using a limited number of protein scaffolds of low sequence identity (Figure 1) with a general head-stalk-membrane β-barrel architecture (Kiessling et al., 2020). Two examples of structurally-diverse head domains are the left-handed parallel β-roll (LPBR) as in YadA (PDB: 1P9H; Nummelin et al., 2004) and the head domain of BadA (PDB: 3D9X; Szczesny et al., 2008) with two β-prism motifs. Such well-defined structural motifs and unique TAA sequence-to-structure connections have been used to predict at least parts of the structure of a TAA passenger domain through an expanding domain dictionary (a collection of sequence-to-structure relations in TAA s) and the bioinformatic program daTAA (Bassler et al., 2015; Hartmann et al., 2012; Szczesny & Lupas, 2008). For instance, the approximately 14-long YadA head repeat has a canonical sequence motif of GxNSVAIGAxSxAx (Nummelin et al., 2004).

We solved the structure of an unusual region of BpaC to test homology model predictions of the C-terminal region of the BpaC passenger domain. We show that it is an unusual LPBR with a

![Figure 1](https://example.com/figure1.png)
hydrophilic “hydrophobic core,” analogous to the “N@d” motif identified by Hartmann et al. in TAA stalk regions (Hartmann et al., 2009). Intriguingly, and like “N@d” in EibD (Hartmann et al., 2009; Leo et al., 2011), this occurs close to the membrane anchor. In addition, the charge distribution of the LPBR is very different than YadA: it is highly negatively charged. We have identified other, similar headgroups in other members of Burkholderia using this new structural classification and speculate that change in charge may explain how BpaC is involved in the infection process of B. pseudomallei.

2 | RESULTS

2.1 | The structure of the C-terminal head domain of BpaC

In our attempts to obtain a structure of BpaC, we noticed that the C-terminal region was more stable and soluble than regions that included parts of either the N-terminal head domain or the stalk domain. C-terminal constructs stayed trimeric in SDS-PAGE even after heating to 95 °C for 10 min in loading buffer (Figure S1) and could be concentrated to >130 mg ml⁻¹. Having settled on BpaC₇₄₁⁻¹₀₅₄ as the region to express, we used the engineered trimeric GCN4 leucine zipper (Hernandez Alvarez et al., 2008) to preserve the DAVNxxQL neck motif at the C-terminal end by replacing the coiled-coil that would be in the membrane β-barrel in the wt structure. We solved the structure by molecular replacement using the C-terminal head domain of BoaA (PDB: 3S6L; Edwards et al., 2011) as a model. There was a single molecule in the asymmetric unit, and the TAA trimer formed around the crystallographic threefold axis. The final R-factors were 18.39%/21.71% (R_work/R_free) at a resolution of 1.4 Å and a total of 14 alternate conformations were built. The quality of the electron density is consistent with this resolution (Table 1) (Figure S2). The final model contains all of the residues in BpaC from position 14 in layer 22 to the end of layer 42 (residues 741–1021), followed by a long neck from E1022 to Q1054, the GCN4 anchor, with 282 water molecules, of which 42 are ordered in specific channels or the trimer core cavity.

2.2 | BpaC C-terminal LPBR has been formed by expansion of a three-layer motif

Sequence alignments show that BpaC₄₃₄⁻¹₀₂₁ contains 42 14-residue LPBR repeats, and intriguingly, it contains clear evidence (Figure 2) of a triplet-LPBR expansion. In this region, layers 3–26 and 30–32 have almost identical repeats of a “GDN”-“GEN”-“GSN” three-layer motif, while layers 1 and 2 are “GDN”-“GTN”, and layer 36 is a GEN-type. The structural consequences of this are discussed in what follows.

| Table 1 | Diffraction data and refinement statistics for BpaC₇₄₁⁻¹₀₅₄ |
|----------|-----------------------------------------------|
| Parameter | Value                                      |
| PDB accession code | 7023                                       |
| Space group | R32                                         |
| a, b, c (Å) | 57.39, 57.39, 516.53                         |
| Resolution (Å) | 44.79–1.40 (1.45–1.40)                      |
| Reflections (observed/unique) | 523,981/65682                              |
| R_merge (%) | 5.43 (40.06)                                |
| Mean(I/σ(I)) | 7.13 (1.43)                                 |
| Completeness (%) | 99.53 (99.28)                              |
| CC1/2 | 99.8 (68.9)                                |
| Multiplicity | 2.0 (1.9)                                  |
| Wilson B value (Å²) | 12.06                                    |

| Refinement |
|-----------------|
| R_work/R_free (%) | 18.39/21.71 |
| Protein/solvent/ligand atoms | 2095/280/13 |
| Average B (Å²) | 22.42        |
| Protein | 29.11       |
| Solvent | 34.61       |
| Ligand | 0.006       |
| R.m.s.d., bonds (Å) | 0.82     |
| Ramachandran plot (%) | 98.47       |
| Most favoured regions | 1.53       |
| Additional allowed regions | 0.93       |
| Clashscore | 4.21        |
| Molprobity score | 1.20      |

Values in parenthesis represent the highest-resolution bin.

2.3 | BpaC is an unusual left-handed β-roll with a primarily hydrophilic core

All three chains of the C-terminal head domain of BpaC thread around a central hydrophilic core in a helical fashion with an average turn of 5.7° ± 0.6° for layers 23–41 and a rise of 4.8 Å ± 0.1 Å (the last one, from 1008–1021, has a turn of just 2.5°), calculated per turn, typical of other LPBRs. From layer 23 to 42, the total twist is 103.8°, and the total rise is 81.5 Å (Table S1). In our numbering below, we refer to layer numbers from the full C-terminal head structure: The first layer in the solved structure is number 23.

Despite having a canonical LPBR structure (Figure 3a,b), there are a number of unique features. First, the presence of asparagine at position 3 within a 14-residue repeat (N@3) leads to a long chain of hydrogen bonds, completely conserved to the N-terminus of this domain, in the first 10 layers of our structure (Figure 3c). We use
X@m to denote single residue code X at position m (m = 1–14) in the 14-residue repeats. Next to this in the neighboring monomer is a three-layer repeat motif of N/D/N@10 (Figure 3d) also forming stabilizing hydrogen bonds down the spine of the structure.

Second, there are three identifiable solvent networks, which we have termed “outer,” “inner,” and “central” (Figure 4a–c). The location of these channels can be defined by their relative location within the LPBR repeats: The “outer” solvent molecules, present in all LPBRs (Figure S3), had not been identified by previous authors including ourselves. We describe them for the first time here: They form an ordered part of the structure between residues 3–4–5 within the 14-residue layers and are 4.8 ± 0.3 Å apart. The “inner” water molecules, situated between residues 5–6–7, are 4.6 ± 0.2 Å apart and completely buried; the “central” solvent molecules are closer to the threefold axis, near the sidechain of residue 7/14, and some, but not all, are hydrogen-bonded to the inner channel.

In BpaC, the outer solvent channel is typically formed by four H-bonds/dipole interactions: two conserved interactions, one to the monomer “A” backbone carbonyls of position 3/14 of layer “n” and the other from the monomer “B” backbone amide of position 9/14 of layer “n,” and the last two found most often in BpaC, from the sidechain γOH of monomer “A” and “B” position 5/14 in layer “n + 1” (Figure 4d). Position 5 is almost always threonine in BpaC (Figure 5), providing extra stability to this interaction in comparison with YadA. The interaction clearly provides stability between monomers, and between layers in an individual monomer.

More surprising, however, are the “inner” and “central” chains of buried water molecules, which form because the BpaC core is uniquely hydrophilic in 15/20 of the LPBR layers, composed of repeating units of N/S/T@7 (Figure 2). H-bonds from the γOH of S/T@7 to the inner water serve as a bridge between layers, which are supported by intramonomer H-bonds from the carbonyl oxygen on (conserved) A@6 in layers “n + 1” to “n + 3” (Figure 4e). The asparagine layers in addition bridge between monomers, so the sidechain amide of N@7 in layer “n + 2” of monomer “B” hydrogen bonds to the inner water, which in turn hydrogen bonds to monomer “A” A@6 backbone carbonyl and in some cases also to monomer “A,” level “n + 3” of the γOH of T@7. The chain of hydrophilic interactions spans across layers and between monomers. With the exception of the special solvent network in layers 35 and 36 with G@7 and N@7, central solvent molecules are H-bonded to the γOH of S/T@7, either sharing a central molecule or having one solvent molecule per residue (Figure 4f,g).

Finally, the core of the molecule around the threefold axis is completely hydrophilic in 15 of the 20 layers in the solved structure (23–36 and 39), as well as all of the layers 1–22 due to N/S/T@7, supported by hydrophilic residues at T/S@5 (Figure S5). Layers 23, 26, 29, and 32 (N@7) have, we believe, an unusual arrangement in which, stochastically, the δΝΗ of one of the three asparagines points into the center, enabling a hydrogen-bond network around the three-fold (Figure S5, T N). At layers 35–36 (G@7, N@7), a tetrahedral arrangement of water molecules forms, with the increased
space at G916 allowing one water molecule on the three-fold axis hydrogen-bonded to a water molecule attached to each N930 in the layer below (Figure 6d). Only the very C-terminal end of the LPBR (layers 37–38 and 40–42) have the hydrophilic residues usually found at positions 5 and 7 of an LPBR (Figure 5, YadA). The hydrophilic core includes 14 “central” buried water molecules distributed over 7 layers (Figure 6a): five on the threefold axis (Figure 6b,d) associated with layers 25 (T776), 31 (T860), 34 (T902), 36 (N930), and 39 (T972) and nine (one per monomer per layer) = 2 Å from the threefold axis associated with layers 27 (S804), 30 (S846), and 35 (G916) (Figure 6c,d). This hydrophilic core coupled with very high stability (Figure S1) is one of the unique structural features of the BpaC head. We performed molecular dynamics (MD) simulations to verify our observations of the described novel solvent patterns in BpaC—specifically the inner solvent channel and the central solvent molecules. Hydration densities approximately 7 times that of bulk water were found to correlate with most of the experimentally observed solvent molecules (Figure S6). Our hydration density maps confirm the presence of the inner solvent channel from layers 23–24 to 35–36, even filling in solvent molecules not visible in our X-ray structure model between layers 23–24, 24–25, and 26–27 (Figure 4a for comparison). Additionally, the layers containing S@7 and T@7 that do not have central solvent molecules in the X-ray structure model (Figure 5, layers 24, 28, and 33) have a visible density in the hydration density maps. On this basis, a reasonable assumption would be that the solvation trend for the central and inner solvent channel in the crystal structure can be extrapolated to the rest of the C-terminal head domain (residues 434 to 740). The MD simulations also support our final refinement strategy of not building any possible ions for locations that would have had a viable amount of coordination sites for them. In essence, no significant replacement of solvent molecules by potassium ions occurred during the simulation.

2.4 The electrostatic charge on an LPBR domain reveals a new head domain subcategory

The other unique structural feature is the negative electrostatic charge of the BpaC head (Figure 7). This is due to the presence of D/E@2 on the outside of the trimer in the “GDN” and “GSN” layers, and the aspartate in N/D/N@10 (Figure 2). The D/E@2 within a single monomer are next to the chain of hydrogen bonds running down the spine of BpaC created by N@3 (Figure 3c) from layer 1–31 (see above), while the D@10 is relatively close (Figure 3d) in a neighboring monomer in the “GEN” layers (Figure 2) and held in position by N@10 in the “GDN” and “GSN” layers. Unsurprisingly, the full-length BpaC head (Figure 7, transparent model), has rows of negatively-charged residues forming a distinct helical pattern on the surface of BpaC, possibly of functional significance.

Unlike all other YadA-like head domain structures (Figure 7), BpaC only has negatively charged residues on its surface. Furthermore, the electrostatic surface charge in YadA-like head domains seems to correlate with the relative position of the head in the passenger domain (Figure 7): the closer the head domain is to the N-terminus (i.e., membrane distal) the higher the ratio of positively-to-negatively charged residues became (Table 2). BpaC is thus the archetype for a new subcategory of negatively charged LPBR membrane-proximal head domains. These differences can be seen clearly in the LBPR frequency plots (Figure 5). Focussing on the charged solvent-accessible residues (Table 2) shows that BpaC and BoaA exclusively have negatively charged residues, while the others have a mix of positively and negatively charged residues, and this is especially so for LPBRs toward the more N-terminal end of the protein (Figure 7). The pls of the solved LPBR structures demonstrate this clearly: the pls of the LPBR is clearly correlated with its position within the passenger domain, with BpaC having a pls of about 2.3 and YadA of 9.2 (Table 2). ElbD, located centrally, is the only one with an almost neutral pls.

2.5 Identification of a BpaC homolog from Burkholderia oklahomensis and the evolutionary relationship of TAA head domains

Similar domains found in multiple TAAs (e.g. LPBRs) can display high structural conservation despite variations in sequence
identity and length. This has an impact on the accuracy of phylogenetic analysis and prompted us to refine our analysis to only include LPBR head domains with the fully conserved G@8. Fifteen sequences from well-characterized TAAs across nine bacterial species were identified as possessing domains with the fully conserved G@8 required for LPBR heads and selected for subsequent phylogenetic analysis. All sequences in the initial analysis were aligned by the G@8 and trimmed using neck motifs as a domain boundary. The alignment of head domains was then used to build an HMM profile and perform an HMM search to identify novel head domains by their conserved G@8 for further analysis (Figure 8a).

Analysis of the head domains revealed a clade linking the BpaC head with a sequence from *Burkholderia oklahomensis*, “Adhesin YadA-like” (UniParc: UP10005D74637)—which we have named BoaC (o for *oklahomensis*)—like the one most closely related to BpaC. The “Hep-Hag Family Protein” BoaA from *B. pseudomallei*, used for molecular replacement (PDB: 3S6L), is a more distant relative of BpaC. Full-length BpaC and BoaC possess 78% identity, with 86% identity over the C-terminal head domain. BoaC has a much shorter C-terminal head domain with two separate deleted segments (BpaC equivalent of ΔA761-N874 and ΔA920-A949) that together correspond to 10×14-residue LPBR layers (Figure S7). A further deleted region, likely a coiled-coil segment, can be found at the N-terminal end of the protein (BpaC equivalent of...
ΔS266-A299). Using this information we extended the analysis to full TAA sequences but only included those which belong to the six LPBR heads included in the previous structural comparison (Figure 7). The resulting phylogenetic tree (Figure 8b) revealed clades that confirm our proposed differentiation into BpaC-like head domains (BpaC, BoaC, BoaA) and YadA-like head domains (EibD, YadA, and UspA1). AtaA is unusual because the full-length sequence includes both an N-terminal LPBR head domain (residues 110–265) and a C-terminal head domain (residues 2989–3104), which is the likely reason for the branch position in between both categories. UspA1 has a separate clade in both trees as it contains an irregular 15-residue repeat with an additional G@9 that likely impacted the alignment generation.

Further proof that bpaC and boaC are closely related is that they are both part of very similar pathogenic islands, unlike other TAs included in this study: downstream of a chaperone-usher pilus assembly gene cluster, and upstream of a two-component response regulator gene cluster, suggesting that both the pili and BpaC are involved in initial adhesion steps of the infection process in both organisms (Figure 8c).

3 | DISCUSSION

3.1 | BpaC$^{741-1054}$, the largest LPBR head structure so far, has an atypical surface charge and core

The structure reported here of BpaC$^{741-1054}$ contains the highest number of LPBR repeats in an experimental structure and is the longest, at 89.5 Å. Its behavior, in crystallizing with the trimer arranged around a crystallographic 3-fold axis, indicates that LPBRs are essentially rigid over at least 10 nm length. We estimate the size of the full-length C-terminal head domain to be about 20 nm, with a twist of about 260°. Using the domain annotations and our predicted homology models, the overall length of BpaC is about
55 nm including the anchor domain. BpaC (1152 residues) is thus at the lower medium-sized end of the TAA spectrum compared to YadA (422 residues, ~20 nm; Hoiczky et al., 2000) and BadA (3973 residues; Thibau et al., 2022); ~240 nm; Müller et al., 2011).

The core of BpaC is very different from YadA (Figure 5). Instead of predominantly β-branched hydrophobic residues (Nummelin et al., 2004) at positions 5 and 7, BpaC has T@5 almost exclusively as well as a number of hydrophilic residues with T/N/S/G@7. T@5 is common in other LPBRs, particularly BoaA, EibD, and UspA, but BpaC is the only example of an LPBR with hydrophilic residues at position 7. This leads to the hydrophilic core, complete with “inner” and “central” water molecules, and may explain why the 42-residue BpaC repeat is maintained throughout the domain: all three layers are needed to stabilize the hydrophilic core. The arrangement is analogous to, but much longer than, the hydrophilic “hydrophobic core” found in the coiled-coil region of EibD (Leo et al., 2011) and described as “N@d” by Hartmann et al. (Figure 9) (Hartmann et al., 2009). We speculate that these structural repeats can be used to adapt to environmental changes by addition or subtraction of individual 14-residue long layers within the head domain. This allows modulation of the overall length of the TAA and changes in the binding surface of the head domain. Because of the strong protein sequence similarity, the DNA tandem repeats are very similar, possibly providing a mechanism for rapid headgroup expansion and bacterial adaptation (Zhou et al., 2014).

In addition, there is a distinct trend in the charge on LPBRs: those that are N-terminal, distal from the membrane, like YadA, are more positively charged, and those that are C-terminal, close to the membrane, are negatively charged (Figure 7). Indeed, the pl of the BpaC C-terminal head, as estimated by IPC 2.0 (Kozlowski, 2021), is around 2.3, whereas the pl of the N-terminal head domain of YadA is around 9.2. We, therefore, propose that YadA-like head domains can be subdivided into two categories: YadA-like head domains (more N-terminal, positively charged, hydrophobic core) and BpaC-like head domains (more C-terminal, negatively charged) as shown in Figure 7.

3.2 Potential functional differences between LPBRs

What is the biological significance of this large difference in charge? We suggest that it enables LPBR head domains to bind different cellular and extracellular matrix components. The solvent-facing side chains are modulated to accommodate changing environmental conditions with a highly-conserved fold. YadA, for instance, binds collagen (El Tahir & Skurnik, 2001), and positively-charged residues have been shown to be important in this interaction (Nummelin et al., 2004), while UspA1/2 binds laminin (Tan et al., 2006). We speculate that BpaC C-terminal LPBR, close to the membrane surface, may interact with the positive ions trapped in the lipopolysaccharide layer. Conversely, the BpaC N-terminal β-prism head due to its hypothesized similarity to the BadA head, which binds to fibronectin (Kaiser et al., 2008), would interact with similar intracellular ligands as B. pseudomallei is an intracellular pathogen. There are other examples of LPBRs with different charge properties: the predicted model of the N-terminal LPBR of AtaA (Ishikawa et al., 2012) (residues 110 to 265, UniProt K7ZP88) has a positive electrostatic surface and mostly positively charged surface-facing residues (Figure S8) as opposed to the structure of the C-terminal LPBR, which is clearly negative (Figure 7). The N-terminal part of AtaA plays a major role in the adhesion properties of this protein while the C-terminal part contributes to the flexibility and toughness of the overall structure (Koiwai et al., 2016). A more mixed charge profile can be seen for the predicted model of the N-terminal LPBR of BoaA (residues 180 to 437, UniProt Q3JFX2, Figure S8) which still differs from the exclusively negative surface charge of the C-terminal head domain (Figure 7). The diversity of LPBRs is much greater than expected, and the presence of a hydrophilic core that is greater than 20 nm in length may have potential synthetic biology applications.

3.3 The homolog of BpaC in Burkholderia oklahomensis enables microbiological studies in a BSL-2 environment

Elucidating the function of B. pseudomallei proteins is challenging, as it is a biosafety level 3 (BSL-3) pathogen, which—appropriately—limits
research access and the ability to make mutations to study function (Cheng, 2010). Using our G@8 alignment method for classifying LPBR repeats into N-terminal YadA-like head domains and C-terminal BpaC-like head domains, we identified a homolog, BoaC in *B. oklahomensis*, a BSL-2 pathogen with 86% identity in the head sequence and 78% identity over the full-length protein. As TAAs often have low sequence similarity despite high structural conservation, these proteins are likely true functional homologs despite the presence of some insertions and deletions. *B. oklahomensis* should thus provide an effective way to study the function of BpaC in a near-native host environment, especially as both the upstream and downstream gene environments for *bpaC* and *boaC* are very similar (Table 3, Figure 8b). This strongly supports the functional connection and transferability of scientific findings between these two proteins.

In conclusion, we have identified the first hydrophilic LPBR core: N/S/T@7 in the C-terminal LPBR head domain of BpaC. Its primarily negative surface charge (D/E@2 and D@10) distinguishes it from the mainly positive charged N-terminal head domain of YadA and so these represent two different classes of LPBR heads with different functional roles. We have also identified a close homolog of BpaC in *B. oklahomensis*, a BSL-2 strain, which we propose to be named BoaC to reflect the close relationship between these two proteins. This homolog may help future investigations of the function of BpaC, identification of its binding partners, and the studies of the role of different charges on head domain function.

Finally, we note that, at the time of writing, the UniProt entry for BpaC (A0A0H3HIJ5) has an AlphaFold model (AF-A0A0H3HIJ5-F1; Jumper et al., 2021) that describes the full-length trimeric BpaC as a monomer that folds back into the membrane anchor. TAAs are obligate trimers and essentially linear over all distance scales examined, so this monomeric model is misleading and cannot be used to inform structural or functional studies.

![Figure 7: Surface charge distribution of selected LPBR head structures.](image)

**Table 2: Number of charged residues in selected LPBRs**

|              | BpaC | BoaA | AtaA | EibD | UspA1 | YadA |
|--------------|------|------|------|------|-------|------|
| Head position| 742-1021 | 1396-1506 | 2994-3106 | 124-266 | 166-267 | 64-194 |
| Length (aa)  | 1152 | 1626 | 3630 | 511 | 863 | 455 |
| Positively charged | 0 | 0 | 3 | 8 | 17 | 11 |
| Negatively charged | 34 | 3 | 6 | 7 | 14 | 10 |
| Positive (%)  | 0 | 0 | 33 | 53 | 55 | 52 |
| Negative (%)  | 100 | 100 | 67 | 47 | 45 | 48 |
| Average pl    | 2.32 | 3.93 | 4.71 | 4.71 | 6.44 | 8.90 |

Note: Solvent-accessible residues were counted and only the ones carrying a charge at pH 7 are listed. The overall size of the whole protein is given alongside the position of the head domain. The amount of positively and negatively charged residues per LPBR is given in % of total charged residues. Average pl value was estimated in IPC 2.0 (Kozlowski, 2021) using the residues of the head position as input. Reported is the average pl value for all results out of IPC 2.0.
with a one-sided GCN4 anchor fragment (denoted RearGCN4 in
Hernandez Alvarez et al. [2008]) from the pIBA-GCN4tri vector (gift
from Dirk Linke, University of Oslo, Norway) and the backbone of
the pET28-a vector (Novagen) (Table S2) to create the BpaC241-1054
fragment. RearGCN4 extends the coiled-coil segments and in-
creases domain stability (Hernandez Alvarez et al., 2008) and, in
our case, replaces the coiled-coil inside the β-barrel transporter
domain. Fragments were amplified using Q5® High-Fidelity 2X Master
Mix (New England Biolabs) with the annealing temperature calcu-
lated from OligoCalc (Kibbe, 2007). Analysis and purification of
each PCR reaction product were performed using a 1% agarose gel
stained with SYBR® Safe DNA gel stain (Invitrogen) followed by the
Macherey-Nagel™ NucleoSpin™ Gel and PCR Clean-up Kit (Fisher
Scientific) according to the manufacturer's protocols. Purified PCR
products were then combined by a 3-part Gibson assembly with

4 | EXPERIMENTAL PROCEDURES

4.1 | Expression and Purification of the BpaC C-
terminal head domain

We used a "divide-and-conquer" approach to split BpaC up. Because
TAAs are repetitive and modular, they can be split into smaller seg-
ments, even in the middle of domains: their modular nature prede-
fin es logical construct start and endpoints. By making a series of
constructs, we identified that the part of the gene encoding a frac-
tion of the C-terminal head domain of BpaC ranging from S741
to Q1054 (BpaCHead, UniProt: A0A0H3HIJ5) was expressed in high
yield and soluble. We amplified this region from synthetic DNA along
with a one-sided GCN4 anchor fragment (denoted RearGCN4 in

| Name | Organism | UniProt ID | Range | Full length |
|------|----------|-----------|-------|-------------|
| BpaC | Escherichia coli | Q9XD56 | 741–1054 | 314 |
| UspA | E. coli | Q9MCI8 | 411–585 | 175 |
| SadA | Staphylococcus aureus | Q7DJ60 | 88–265 | 178 |
| YadA | Vibrio cholerae | Q3JFX2 | 88–265 | 178 |
| SadA | Staphylococcus aureus | Q7DJ60 | 88–265 | 178 |
| BpaC | Burkholderia pseudomallei | Q9XD56 | 68–203 | 135 |
| UspA | Moraxella catarrhalis | Q9XD56 | 68–203 | 135 |

FIGURE 8 Evolutionary relationship of TAA LPBRs and genetic environment of bpaC and boaC. (a) Alignments of LPBRs with the BpaC
gene (green) are produced based on the conserved G@8. A phylogenetic tree is shown with TAA names, organism name, UniProt ID, the length of the TAA, and the total length of the TAA.

(b) Alignment of full-length sequences focusing on TAAs that include the LPBR domains analyzed in the previous structural comparison. Clade assignment was split into BpaC-like (green box with solid line) and YadA-like (magenta box with solid line) supporting our previous subclassification evolving out of the structural comparison of LPBRs. AtaA is assigned to BpaC-like and YadA-like (dashed lines), as the full sequence contains both an N-terminal and a C-terminal LPBR head domain. (c) Genome island surrounding bpaC (green) and boaC (blue). Locus tag of genes adjacent to bpaC of B. pseudomallei 1026b and boaC of B. oklahomensis LMG 23618 is displayed in shortened form (BP1026B_X and EIK52_X). Localization of gene product is shown.
NEBuilder® HiFi DNA Assembly. Heat-shock transformation of a ligated vector into chemically competent One Shot® OmniMAX™ Escherichia coli cells (Invitrogen™, Thermo Fisher Scientific) was performed according to the manufacturer’s protocol and plated onto LB agar plates with Kanamycin antibiotic selection. Amplification and purification of plasmid DNA using NucleoSpin Plasmid Mini Kit (Macherey-Nagel) were done following the manufacturer’s instructions. The target sequence was confirmed by sequencing (Mix2Seq (Macherey-Nagel) were done following the manufacturer’s instructions. The target sequence was confirmed by sequencing (Mix2Seq kit, Eurofins Genomics). pET28a-BpaC was then transformed into Escherichia coli BL21(DE3) cells and plated onto an LB agar plate with Kanamycin for selection.

Expression of BpaC\textsuperscript{741-1054} in LB media was induced at an OD600 value of 0.6 at a final concentration of 750\textmu M Isopropyl-\textbeta-D-thiogalactopyranoside for 4 h. Cells were harvested by centrifugation (4000 x g for 20 min at RT) and stored at −20 °C until further use. Frozen cells were thawed on ice in 20 ml buffer A consisting of 50 mM Na\textsubscript{2}HPO\textsubscript{4} pH 8, 300 mM NaCl, and 10 mM Imidazole pH 8. Cells were lysed using sonication and a further 15 ml of buffer A was added before centrifugation (40,000 x g for 40 min at 7 °C). The supernatant was transferred to a 50 ml Falcon tube containing Ni Sepharose® 6 Fast Flow resin (GE Healthcare Life Sciences) preincubated with buffer A and left on a tube roller for 15 min at RT. Resin and supernatant were then transferred to an Econo-Pac® Chromatography column (Bio-Rad). The flowthrough was collected and 20 column volume (CV) buffer A added to the column. Further wash steps consisted of 5 CV buffer B (buffer A, 30 mM Imidazole pH 8) and 10 CV buffer C (buffer A, 50 mM Imidazole pH 8). Elution of target protein was carried out using 8 CV buffer D (buffer A, 250 mM Imidazole pH 8). The purity of the sample was assessed by SDS-PAGE. The protein was dialyzed into 2 L of 20 mM Tris–HCl pH 8, 150 mM NaCl using SnakeSkin™ Dialysis Tubing (10 K MWCO, Thermo Fisher Scientific) overnight at 7 °C. Protein was concentrated using a 15 ml Amicon© Ultra centrifugal filter unit (50 kDa MWCO, Merck Life Science) to about 130 mg ml\textsuperscript{−1} from an initial concentration of 2.3 mg ml\textsuperscript{−1}. Protein concentration was estimated by UV absorption at 280 nm using the theoretical extinction coefficient calculated by ExPASy (Gasteiger et al., 2003).

### 4.2 | Crystallization, data collection, and processing

As BpaC\textsuperscript{741-1054} is extremely soluble, initial crystallization screens for BpaC\textsuperscript{741-1054} were set up using about 130 mg ml\textsuperscript{−1} of protein using the sitting-drop vapor-diffusion method with a drop volume of 200 nl, the protein-to-reservoir ratio of 1:1 and a reservoir volume of 25 \textmu l at 20 °C. Crystals grew readily in several conditions after 1 day in the JCSG screens I-IV (NeXtal Biotechnologies, USA). We harvested crystals from 0.1 M HEPES pH 6.5, 0.8 M (NH\textsubscript{4})\textsubscript{2}SO\textsubscript{4}. Here, 400 nl of cryo-protectant buffer (0.15 M HEPES pH 6.5, 1.2 M (NH\textsubscript{4})\textsubscript{2}SO\textsubscript{4}, 35% Glycerol) was added directly to the drop and incubated for 1 min before transferring to a liquid nitrogen container for data collection.

### TABLE 3 Overview of genes adjacent to bpaC

| Locus tag          | Refseq ID     | Gene product                        | Function/GO term                         |
|--------------------|---------------|-------------------------------------|------------------------------------------|
| BP1026b_I1569      | WP_004521426  | Fimbrial subunit                     | Fimbrial-type adhesion domain            |
| BP1026b_I1570      | WP_004527078  | Fimbrial chaperone protein           | PapD-like                                 |
| BP1026b_I1572      | WP_004527078  | Usher protein                        | PapC-like                                 |
| BP1026b_I1573      | WP_004554117  | Type-1 fimbrial protein              | Pilin (type 1 fimbria component)         |
| BP1026b_I1575      | WP_014696818  | BpaC                                 | Pathogenesis                              |
| BP1026b_I1577      | WP_004193126  | DNA-binding response regulator       | DNA-binding response regulator            |
| BP1026b_I1578      | WP_004531338  | Two-component regulatory system, sensor kinase protein | Phosphorylation, signal transduction |
| BP1026b_I1579      | WP_004550403  | EAL domain-containing protein        | EAL domain, signalling protein            |

Note: Gene annotation of adjacent functional clusters of bpaC that are either selected as part of the predicted pathogenicity island (I1569-I1575) or the possible association is inferred by literature reference (I1577-I1579).

BpaC in bold as the gene around which we searched.
A total of 3600 images were collected by oscillation method with a range of 0.1 ° per image on a Dectris Eiger2 XE 16M detector using single-wavelength synchrotron radiation on beamline I04 at Diamond Light Source (Didcot, UK). Image processing was performed using XDS (Kabsch, 2010). Only images in the range of 800 to 2300 were used for data processing due to large changes in the c axis during the first 800 images, and the final dataset had a completeness of 99.5% to a resolution of 1.4 Å (Table 1).

### 4.3 Structure solution and refinement

Data reduction was performed with AIMLESS (Evans, 2011; Evans & Murshudov, 2013) and the model for molecular replacement (MR) was selected by using the sequence of BpaC\textsuperscript{741–1054} in the Advanced Search of the Protein Data Bank (PDB; Burley et al., 2019) and using the top hit (356L) for molecular replacement. The model was prepared using CLUSTALW (Larkin et al., 2007) to obtain a sequence alignment and CHAINSAW (Stein, 2008) to retain the conserved residues and truncate non-conserved residues to alanine. The output model was used in a molecular replacement search using PHASER (McCoy et al., 2007) and then passed on to Buccaneer (Cowtan, 2006, 2012) for automated model building. This initial model was improved with several rounds of REFMAC5 (Murshudov et al., 2011) and manual model building in Coot (Vers. 0.8.9.3; Emsley et al., 2010), and the quality of refinement was checked against the MolProbity online server (Williams et al., 2018). Final refinement steps (Table 1) were done in PHENIX (Vers. 1.18-3845-000; Adams et al., 2011), and the final R-factors were \( R_{\text{work}} / R_{\text{free}} \) 18.39%/21.71%.

### 4.4 Structure analysis and creation of a full C-terminal head domain model

The structure was analyzed and images were created using PyMOL (Vers. 2.4.1; Schrodinger, 2017) and Inkscape (Version 1.0.1, https://inkscape.org). A full C-terminal head domain, BpaC\textsuperscript{421–1054}, was created by aligning layers in the solved structures in PyMOL and stacking them on top of each other. The term “layer” refers to a single 14-long residue repeat within the C-terminal head domain of BpaC with the conserved glycine at position 8 of the 14 residues that are present in all LPBRs. This was possible because the sequences between layers in the remaining C-terminal head domain are identical to the first two layers of BpaC\textsuperscript{741–1054}. This process is repeated until T431 and then merged in Coot. Varying side chains are replaced using the actual sequence and the side-chain geometry information is taken from the corresponding layers in the actual structure which had the same residue at that position in the repeat. Size estimation of homology models and the extended head domain model (T731-Q1054) was performed in PyMOL. APBS Electrostatics plugin in PyMOL was used for electrostatic surface visualization (https://pymolwiki.org/index.php/APBS). The angle between sheets was calculated using the psico plugin (https://pymolwiki.org/index.php/Psico) and angle_between_domains, using each layer as a single domain.

### 4.5 Sequence analysis and comparison with other trimeric autotransporters

The sequence of BpaC (https://www.uniprot.org/uniprot/A0A0H3HJ5) was annotated using a combination of results from PSIPRED (Buchan & Jones, 2019), daTAA (Szczesny & Lupas, 2008), DeepCoil (Ludwiczak et al., 2019), Clustal Omega (Madeira et al., 2019), and well-defined TAA structure motifs (Bassler et al., 2015; Kiessling et al., 2020). Alignment of LPBR head repeats (G742-S1021) was performed by designation of the start and end point of each individual repeat, allowing comparison of residues within the repeats and the creation of a frequency plot using the WebLogo server (Crooks et al., 2004). Similar TAA structures were identified using a shortened structure model (S741-A782, three repeats) as input for the Dali server (Holm & Laakso, 2016). Top hits that belonged to the TAA protein class were assessed and structurally compared using PyMOL. Frequency plots were created by following a similar pattern as for the BpaC C-terminal head domain logo motif. First, the start and the end of the domain were identified in PyMOL; second, the glycine at position 8 of the usually 14-residue long repeats was aligned by calculating the sequence length between glycines and identifying the correctly spaced ones (usually 14 residues), and breaking the sequence between residue 14 and 1 of the next layer. Alignments were cross-checked with the actual structure. In the special case of the UspA1 LPBR (3PR7; Agnew et al., 2011) most layers consist of a 15-residue repeat with an additional conserved glycine at position 9, which led to a different logo than for the other LPBRs. Residues in loops outside the 14-residue core motif were not included in the frequency plots for the remainder of the LPBRs. These are still included in later calculations of solvent-accessible residues and in the electrostatic surface presentation.

### 4.6 Homology analysis

We identified potential BpaC homologs using BpaC\textsuperscript{741–1054} in a PSI-BLAST search against the NCBI database (Altschul et al., 1997) and a pBLAST (Altschul et al., 1990) against the Burkholderia genome database (Winsor et al., 2008). Prior to alignment, we removed all hits that did not provide at least 50% coverage of the query, were partial sequences, or were “obsolete” entries. The remaining sequences were aligned initially using MAFFT (Madeira et al., 2019) followed by manual curation in Jalview (Waterhouse et al., 2009) of the TAA head domain repeats. We deemed sequences lacking 14-residue periodicity, G@8, and an additional fully conserved G@1 not BpaC-like and discarded them from the analysis. The final sequences in alignment were used to construct a BpaC phylogenetic tree in MEGA X (Kumar et al., 2018). Default parameters for maximum-likelihood methods in
MEGA X were applied (Kumar et al., 2018) and the result was viewed using the interactive tree of life (iTOL) (Letunic & Bork, 2019).

4.7 | Phyllogenetic analysis and identification of pathogenicity islands

Sequences for BoaA, AtaA, EibD, UspA1, YadA, and “Adhesin YadA-like (BoaC),” a presumed BpaC homolog from B. oklahomensis, were trimmed and manually aligned to BpaC \(^{741-1054}\). Alignments were generated in Jalview (Waterhouse et al., 2009) by designating G434 of BpaC as the start point for periodic repeats, as it contains a completely conserved Glycine at position 8. All other sequences were aligned to this one and gaps were introduced for repeats that deviated from the 14-residue motif. Genomic islands for bpaC and boaC were predicted using IslandViewer4 (Bertelli et al., 2017). Functional annotation was acquired through InterPro (Blum et al., 2021). For the remaining TAA genes, adjacent genes were compared using a BLAST search against the KEGG database (Kanehisa & Goto, 2000).

4.8 | Molecular dynamic simulations of solvent hydration maps

Atomistic MD simulations were performed using the crystal structure of BpaC \(^{741-1054}\) as input. All atomistic MD simulations used the AmberTools21 and Amber20 suite of programs (Case et al., 2021) with the FF14SB forcefield (Maier et al., 2015) used to describe the protein. The experimentally derived structures were protonated according to the Amber residue templates and then solvated with TIP3P water molecules in an octahedral box that extended 12 Å from the protein. Potassium ions were added to neutralize the system, then potassium chloride was added to a final concentration of 150 mM. After an initial energy minimization the system was heated to 300 K as positional restraints were decreased from 100 to 0 kcal mol\(^{-1}\) Å\(^{-2}\). Two unrestrained MD simulations were each performed for 1 \(\mu s\) starting from different arbitrary initial velocity distributions. The MD simulations used the pmemd.cuda module from Amber20 and were run on V100 GPUs. Hydration density maps were calculated using the ccptraj module within Amber20 using a grid spacing of 1 Å and a grid box size of 200 Å\(^3\) and saved relative to a time-averaged pdb file of the protein calculated over the trajectory. The hydration densities were normalized relative to bulk water (1 g cm\(^{-3}\)) and saved as XPLOR files for visualization in PyMOL. We used a density cutoff of seven times that of bulk water as it provides a good tradeoff between signal and background noise.

AUTHOR CONTRIBUTIONS
Andreas R. Kiessling designed and carried out all wet-lab experiments, structure building/analysis, and gene cluster analysis. Kathleen M. Weimer performed homology and phylogenetic analysis. Sarah A. Harris and Geoffrey Wells performed molecular dynamic simulations. Andreas R. Kiessling, Adrian Goldman, Sarah A. Harris, Kathleen M. Weimer and Geoffrey Wells wrote the manuscript. Adrian Goldman provided scientific input. All authors read and approved the manuscript.

ACKNOWLEDGMENTS
This project has received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 765042, and from the BBSRC (BB/T006048/1), both to AG. We acknowledge access to the N8 Bede supercomputer through the HecBioSim EPSRC grant (EP/R029407/1).

CONFLICT OF INTEREST
The authors declare that they have no conflicts of interest.

ETHICAL STATEMENT
All GMO work was done in BSL 1 bacteria, according to best practice at the University of Leeds, and the University and laboratory have all appropriate licenses for this kind of work.

DATA AVAILABILITY STATEMENT
Model coordinates and associated structure factors of BpaC \(^{741-1054}\) were deposited at the Protein Data Bank in Europe (PDBe); accession code 7O23.

REFERENCES
Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Echols, N., Headd, J.J. et al. (2011) The Phenix software for automated determination of macromolecular structures. Methods. 55(1), 94–106. https://doi.org/10.1016/j.ymeth.2011.07.005
Agnew, C., Borodina, E., Zaccai, N.R., Conners, R., Burton, N.M., Vicary, J.A. et al. (2011) Correlation of in situ mechanosensitive responses of the Moraxella catarrhalis adhesin UspA1 with fibronectin and receptor CEACAM1 binding. Proceedings of the National Academy of Sciences of the United States of America, 108(37), 15174–15178. https://doi.org/10.1073/pnas.1106341108
Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. Journal of Molecular Biology, 215(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research, 25(17), 3389–3402. https://doi.org/10.1093/nar/25.17.3389
Balder, R., Lipski, S., Lazarus, J.J., Grose, W., Wooten, R.M., Hogan, R.J. et al. (2010) Identification of Burkholderia mallei and Burkholderia pseudomallei adhesins for human respiratory epithelial cells. BMC Microbiology, 10(1), 250. https://doi.org/10.1186/1471-2180-10-250
Bassler, J., Hernandez Alvarez, B., Hartmann, M.D. & Lupas, A.N. (2015) A domain dictionary of trimeric autotransporter adhesins.
Kiessling, A. R., Harris, S. A., Weimer, K. M., Wells, G. & Goldman, A. (2022). The C-terminal passenger domain of \textit{Burkholderia pseudomallei} BpaC has a striking hydrophilic core with an extensive solvent network. \textit{Molecular Microbiology}, 118, 77–91. https://doi.org/10.1111/mmi.14953

**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.