

RESEARCH ARTICLE

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

Andreas R. Kiessling ¹ | Sarah A. Harris ² | Kathleen M. Weimer ^{1,3} |
Geoffrey Wells ⁴ | Adrian Goldman ^{1,5}

¹Astbury Centre for Structural Molecular Biology, School of Biomedical Sciences, University of Leeds, Leeds, UK

²Astbury Centre for Structural Molecular Biology, School of Physics and Astronomy, University of Leeds, Leeds, UK

³IGBMC, School of Life and Health Sciences, University of Strasbourg, Strasbourg, France

⁴UCL School of Pharmacy, University College London, London, UK

⁵MIBS, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

Correspondence

Adrian Goldman, Astbury Centre for Structural Molecular Biology, School of Biomedical Sciences, University of Leeds, Leeds, UK.
Email: adrian.goldman@helsinki.fi

Funding information

Biotechnology and Biological Sciences Research Council, Grant/Award Number: BBSRC (BB/T006048/1); Engineering and Physical Sciences Research Council, Grant/Award Number: EP/R029407/1; H2020 Marie Skłodowska-Curie Actions, Grant/Award Number: 765042

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

Sarah A. Harris and Kathleen M. Weimer contributed equally to this work.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Molecular Microbiology* published by John Wiley & Sons Ltd.

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 (TAAs) 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 TAAs 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* TAAs, 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 TAAs: adherence, invasion, serum resistance, and biofilm formation (for a review, see Kiessling et al. [2020]). In general, TAAs 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 carcinoembryonic 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

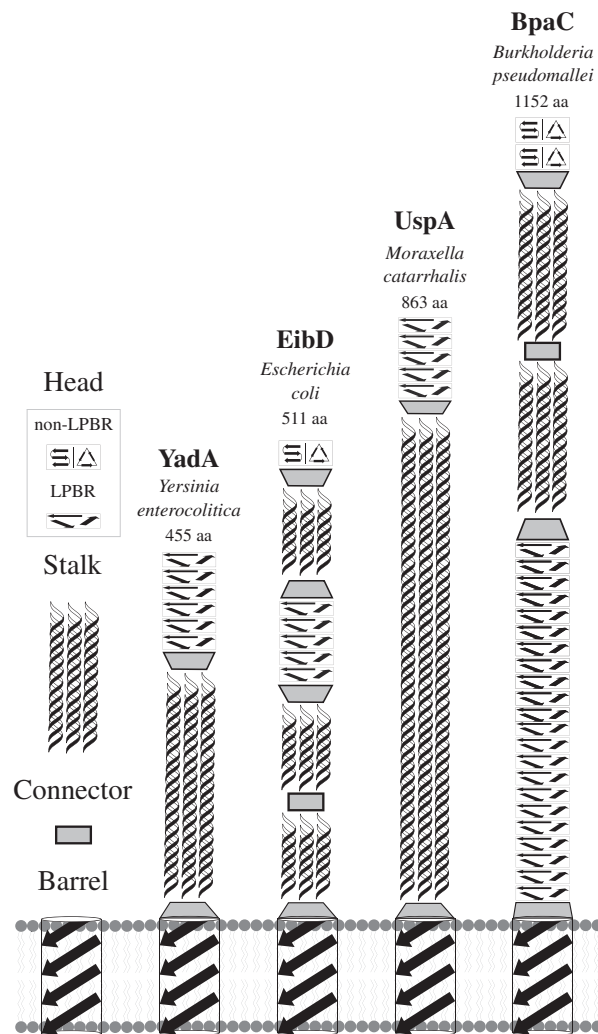


FIGURE 1 Schematic diagram of solvent-accessible passenger domains of four different TAAs. Left: Structural modules in TAAs consist of β -sheet rich head domains, coiled-coil stalk domains, connector neck motifs, and the membrane-anchoring β -barrel domain. Head domain folds can be identified as either left-handed parallel β -rolls (LPBR) or non-LPBR; with the latter predicted as either β -meander or β -prism motif (indicated by a split in the illustration). The different locations of the head domains are shown for the passenger domains of *Yersinia enterocolitica* YadA, *Escherichia coli* EibD, *Moraxella catarrhalis* UspA1, and *B. pseudomallei* BpaC. Connector domains show transition of coiled-coil stalk domains to head domains or from the β -barrel to the passenger domain. The length of each TAA is shown (aa).

(a collection of sequence-to-structure relations in TAAs) 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 GxNSVAIGAxSxSx (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

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% ($R_{\text{work}}/R_{\text{free}}$) 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	7O23
Data collection	
Space group	R32
Unit cell	
a, b, c (Å)	57.39, 57.39, 516.53
Resolution (Å)	44.79-1.40 (1.45-1.40) ^a
Reflections (observed/unique)	523,981/65682
R_{merge} (%)	5.43 (40.06)
Mean($I/\sigma[I]$)	7.13 (1.43)
Completeness (%)	99.53 (99.28)
CC _{1/2}	99.8 (68.9)
Multiplicity	2.0 (1.9)
Wilson B value (Å ²)	12.06
Refinement	
$R_{\text{work}}/R_{\text{free}}$ (%)	18.39/21.71
Protein/solvent/ligand atoms	2095/280/13
Average B (Å ²)	
Protein	22.42
Solvent	29.11
Ligand	34.61
R.m.s.d., bonds (Å)	0.006
R.m.s.d., angles (°)	0.82
Ramachandran plot (%)	
Most favoured regions	98.47
Additionally allowed regions	1.53
Rotamer outliers (%)	0.93
Clashscore	4.21
Molprobrity score	1.20

^aValues 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 $5.7^\circ \pm 0.6^\circ$ for layers 23-41 and a rise of $4.8 \text{ \AA} \pm 0.1 \text{ \AA}$ (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 \AA (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

	GDN	GEN	GSN	Layer
	Position 123456789-1234	123456789-1234	123456789-1234	
Extended domain 434-740	GDN STASGD N ATAS	GEN STATGTD S TAS	GSN STANGT N STAS	1-2
	GDN STASG T NASAS	GEN STATGTD S TAS	GSN STANGT N STAS	3-5
	GDN STASG T NASAT	GEN STATGTD S TAS	GSN STANGT N STAS	6-8
	GDN STASG T NASAS	GEN STATGTD S TAS	GSN STANGT N STAS	9-11
	GDN STASG T NASAT	GEN STATGTD S TAS	GSN STANGT N STAS	12-14
	GDN STASG T NASAT	GEN STATGTD S TAS	GSN STANGT N STAS	15-17
	GDN STASG T NASAT	GEN STATGTD S TAS	GSN STANGT N STAS	18-20
Solved structure 741-1021	GDN STASG T NASAT	GEN STATGTD S TAS	GSN STANG A NSTAS	21-23
	GDN STASG T NASAT	GEN STATGTD S TAS	GSN STANGT N STAS	24-26
	G N N STASG T NASAT	GEN STATGTD S AAS	G T N STANGT N STAS	27-29
	GDN STASG T NASAT	GEN STATG T ASTAS	GSN STANG A NSTAS	30-32
	G AGATAT G ENAAAT	G AGATAT G NNASAS	G TSSTAG G ANAIAS	33-35
End of inner solvents		GEN ST T NG A NSTAS		36
	G NGSSAF G ESAAAA	G DG S TAL G ANAVAS	G VGS V AT G AGSVAS	37-39
	G ANSSAY G TGSNAT	G AG S VA I GOGATAS	G SN S VAL G TGSVAS	40-42

FIGURE 2 Alignment of the BpaC⁴³⁴⁻¹⁰²¹ region. 14-residue long LPBR layer repeats are aligned to emphasize similarities between GDN/GEN/GSN families. Bold: identical copies of the consensus sequence in each family. Residues included in the solved structure have a yellow background and D/E/S@2 highlighted by magenta boxes. Residues that do not match the consensus sequence are colored in red. Layers 37-42 deviate too much from the GDN/GEN/GSN family assignment and are excluded (grey). The end of the inner solvent channel is indicated, as this correlates with the end of the GDN/GEN/GSN family assignments.

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 γ OH 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 γ OH 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 γ OH 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 γ OH 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 δ NH 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

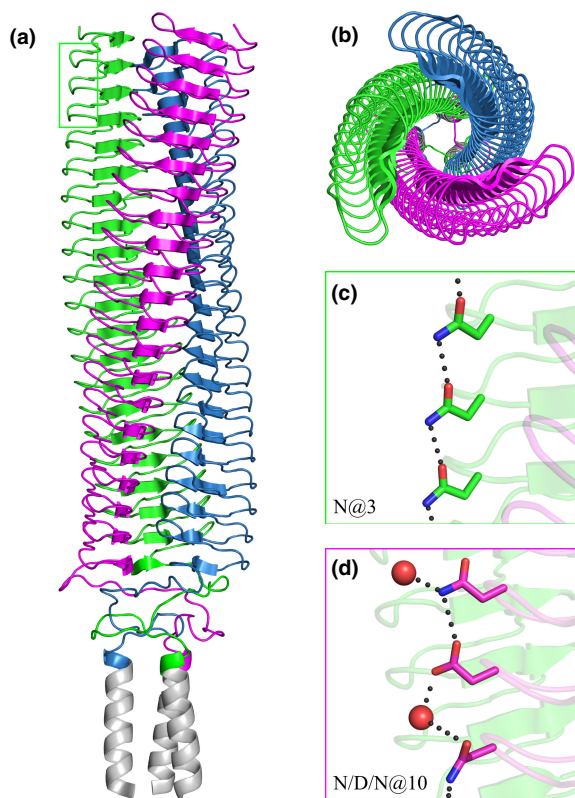


FIGURE 3 Overall view of the C-terminal head domain of BpaC^{741–1054}. The structure covers 20 of the predicted 42 LPBR head repeats, each layer consisting of 14 amino acids. The three monomers are colored magenta, blue, and green, apart from the added GCN4 leucine zipper (grey). (a) Side view. (b) Top view from the N-terminus. (c) N@3 in the first 10 repeats form a continuous network of stabilizing hydrogen bonds (only three shown). (d) N/D/N@10 also forms a network of stabilizing hydrogen bonds with interspersed solvent molecules (red spheres).

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 hydrophobic 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 S5, 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 “GEN” 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 LPBR 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 pIs of the solved LPBR structures demonstrate this clearly: the pI of the LPBR is clearly correlated with its position within the passenger domain, with BpaC having a pI of about 2.3 and YadA of 9.2 (Table 2). EibD, located centrally, is the only one with an almost neutral pI.

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

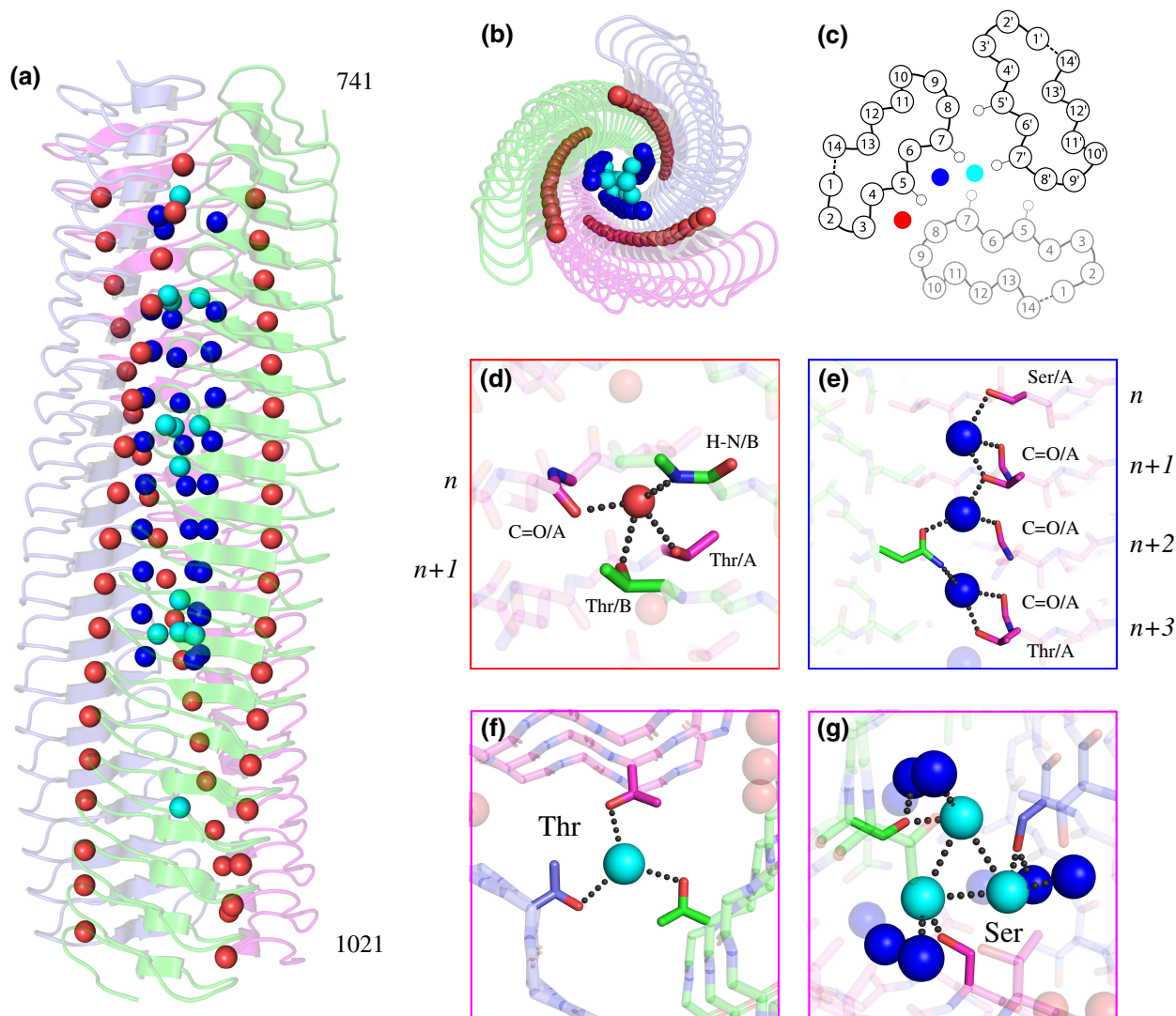


FIGURE 4 Solvent channels and interactions in BpaC⁷⁴¹⁻¹⁰²¹. Water molecules are color-coded: outer (red), inner (blue), and central (cyan). The three monomers are shown in magenta and blue, with hydrogen bonds in black dots. (a) Side view of the different solvent channels of BpaC⁷⁴¹⁻¹⁰²¹ from N- to C-terminus. (b) Top view of BpaC⁷⁴¹⁻¹⁰²¹ showing the different solvent channels. (c) Schematic of a 14-residue layer of BpaC⁷⁴¹⁻¹⁰²¹ indicating the location of the various solvent channels alongside the C β atoms of residues 5 and 7 to help distinguish the different channels. (d) Outer solvent channel: stabilizing chiefly by interactions between backbone atoms in layer “ n ”, and the γ OH of T@5 in layer “ $n+1$ ” between two monomers. (e) Inner solvent channel: stabilized by interactions of hydrophilic sidechains S/T/N@7, and supported by backbone carbonyls. The γ OH of S/T@7 connects to the inner solvent channel from monomer “A” (cyan spheres) while the interactions of N@7 contribute from the adjacent monomer “B” (green). (f) The central solvent molecule, with the only interaction from an adjacent T972@7 γ OH. (g) The central solvent channel stabilized by a S846@7 γ OH and further connections to the inner solvent channel.

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: UPI0005D74637)—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 Δ A761-N874 and Δ A920-A949) that together correspond to 10 \times 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

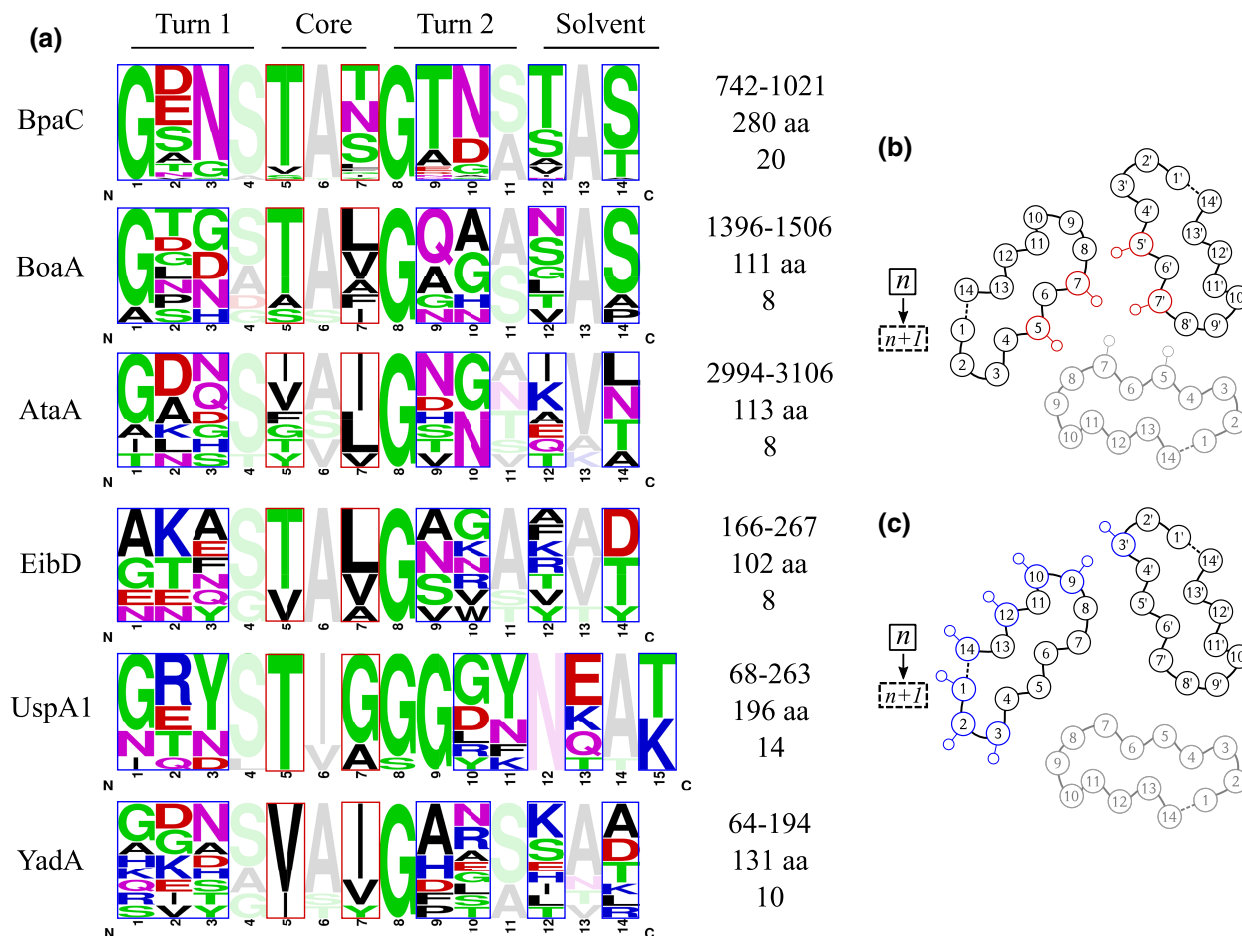


FIGURE 5 Residue frequencies for TAA left-handed parallel β -roll repeats, in which the height represents the frequency of residues in each layer. The G@8 is completely conserved. Each layer consists of 14 residues except for the mostly 15-residue repeats in UspA1, and rare loop insertions or deletions that were excluded prior to logo creation (see Figure S4). These occur between Pos14/14 of layer “ n ” and Pos1/14 of layer “ $n + 1$ ”. (a) Sequence logos of residue frequencies. Amino acids are colored by side-chain properties. Residues that face into the chain interior are shown transparently for better visibility. (b) C β positions of amino acids contributing to core interactions are boxed in both frequency plot and sketch (red). (c) Solvent-facing C β positions are highlighted in both plot and sketch (blue). There is one exception to the completely conserved G@8: UspA1 S105 (PDB: 3PR7; Agnew et al., 2011). We ascribe this to an error in the sequence or the structure of the protein, as the ϕ and φ angles are disallowed for serine and there is no sidechain density even at C β in the $2F_o - F_c$ map (data not shown).

Δ S266-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 TAAs included in this study: downstream of a chaperone-usher pili 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 | BpaC⁷⁴¹⁻¹⁰⁵⁴, the largest LPBR head structure so far, has an atypical surface charge and core

The structure reported here of BpaC⁷⁴¹⁻¹⁰⁵⁴ 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

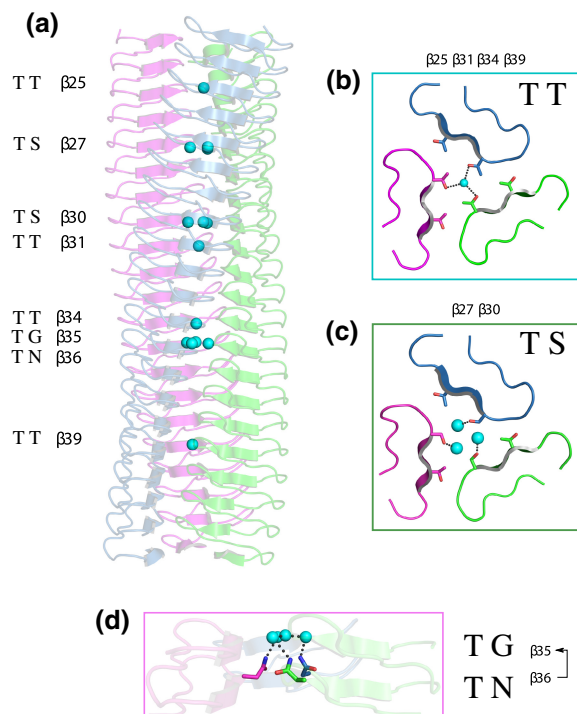


FIGURE 6 Varying arrangement of central solvent molecules in BpaC⁷⁴¹⁻¹⁰²¹. (a) Side view of BpaC⁷⁴¹⁻¹⁰²¹ showing LPBR layers containing solvent molecules (cyan spheres) within the trimer core. Layers with shown motifs are numbered; motifs without solvent molecules are not included in this Figure. (b) Top view of a T@7 layer. Hydrogen bonds (black) are indicated between three γ OH sharing a single solvent molecule. (c) Top view of hydrogen bonds is shown between three S@7 γ OH and three solvent molecules. (d) Side view of hydrogen bonds displayed between three N@7 δ NH and four solvent molecules arranged in a tetragonal arrangement in the layer above. The missing sidechain of the G@7 creates a large cavity for these solvent molecules.

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 pI of the BpaC C-terminal head, as estimated by IPC 2.0 (Kozłowski, 2021), is around 2.3, whereas the pI 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

FIGURE 7 Surface charge distribution of selected LPBR head structures. The structures were trimmed to contain just the LPBR layers to show the distribution of solvent-accessible side chain charges (APBS plugin, PyMOL). Structures are sorted by surface charge (negative to positive, red to blue). Structures are labeled with name, PDB ID, residues used for APBS map, and overall size of the whole TAA. The positions of the head domains relative to the membrane and TAA termini are indicated. Models not to scale. BpaC is divided into a known structure (741–1021, grey) and model inferred by the identity of repeats (431–740, transparent). LPBR structures are divided into BpaC-like and YadA-like depending on their relative position and their surface charge profile.

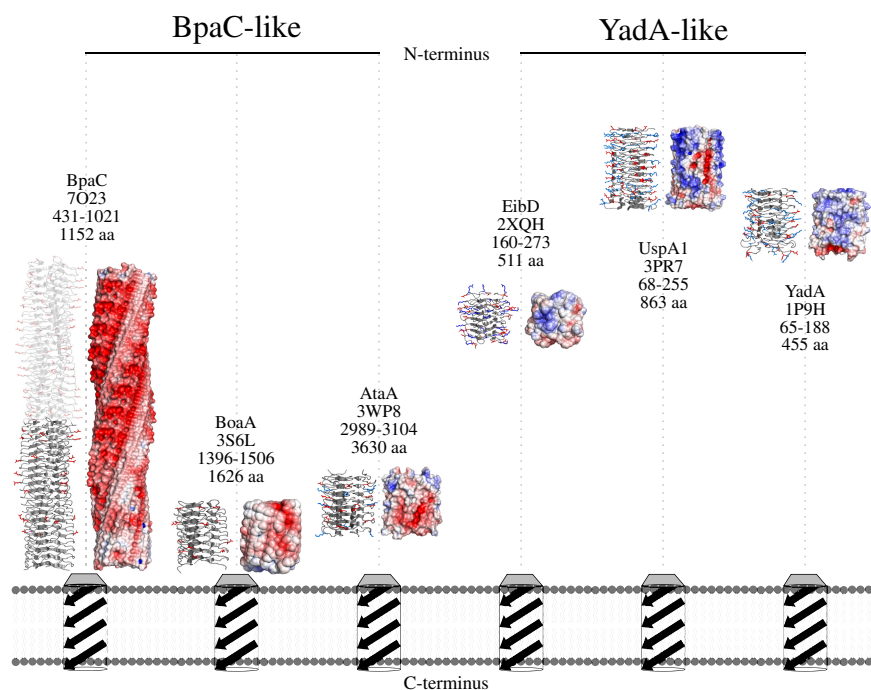


TABLE 2 Number of charged residues in selected LPBRs

	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 pI	2.32	3.93	4.71	6.44	8.90	9.17

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. pI value was estimated in IPC 2.0 (Kozłowski, 2021) using the residues of the head position as input. Reported is the average pI value for all results out of IPC 2.0.

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 (AOA0H3HIJ5) has an AlphaFold model (AF-AOA0H3HIJ5-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.

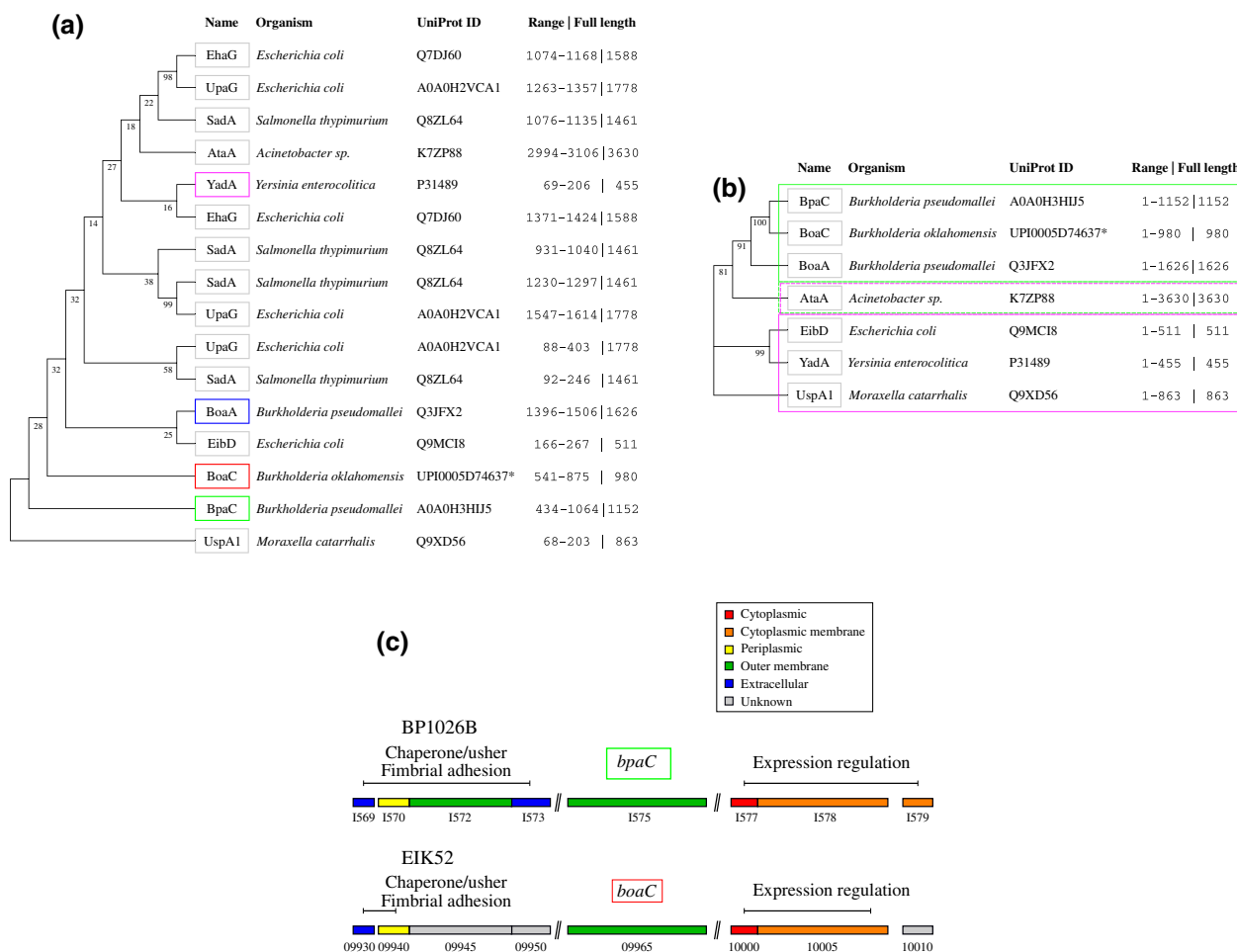


FIGURE 8 Evolutionary relationship of TAA LPBRs and genetic environment of *bpaC* and *boaC*. (a) Alignments of LPBRs with the BpaC (green) are produced based on the conserved G@8. A phylogenetic tree is shown with TAA names, organism name, UniProt ID, the range of residues included in the final alignment, and the total length of the TAA. “Adhesin YadA-like” or BoaC (red) from *B. oklahomensis* is given with UniParc ID (*). The BoaA head domain which was used for molecular replacement for the structure in this publication is highlighted (blue). UspA1 is a separate branch, probably due to the unusual 15-residue repeats. YadA as the canonical LPBR is highlighted to show the distance to BpaC (magenta). (b) Alignment of full-length sequences focussing 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.

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 segments, even in the middle of domains: their modular nature predefines logical construct start and endpoints. By making a series of constructs, we identified that the part of the gene encoding a fraction of the C-terminal head domain of BpaC ranging from S741 to Q1054 (BpaCCHead, 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

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 BpaC^{741–1054} fragment. RearGCN4 extends the coiled-coil segments and increases 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 calculated 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

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 kit, Eurofins Genomics). pET28a-BpaC⁷⁴¹⁻¹⁰⁵⁴ was then transformed into *Escherichia coli* BL21(DE3) cells and plated onto an LB agar plate with Kanamycin for selection.

Expression of BpaC⁷⁴¹⁻¹⁰⁵⁴ in LB media was induced at an OD₆₀₀ value of 0.6 at a final concentration of 750 μM

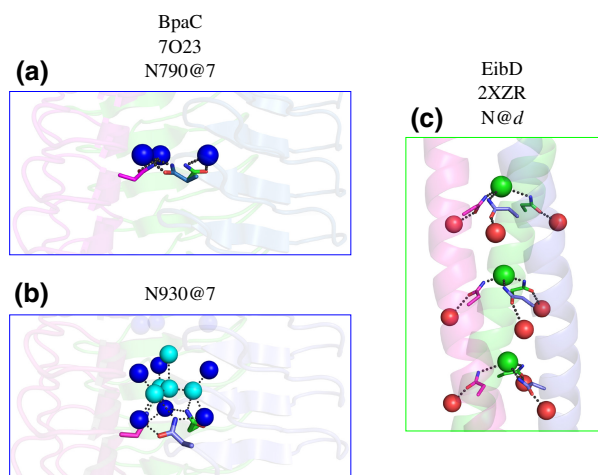


FIGURE 9 Comparison of N@d motif in EibD with BpaC N790 and N930 interactions—(a) Solvent interactions of the inner channel (blue spheres) are displayed with N790@7 as the example for other layers like this in BpaC. (b) A complex solvent network involving central solvent molecules (cyan spheres) and the inner channel (blue spheres) are being supported by N930@7 pointing toward the layer above which has G@7 that creates a cavity for these solvent molecules to fill in. (c) The N@d motif in the coiled-coil of EibD (2XZR; Leo et al., 2011) sequesters chloride ions (green spheres) and interacts with adjacent solvent molecules (red spheres).

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 (I1577-1579).

BpaC in bold as the gene around which we searched.

Isopropyl-β-D-thiogalactopyranoside for 4 h. Cells were harvested by centrifugation (4000×g for 20min at RT) and stored at -20 °C until further use. Frozen cells were thawed on ice in 20ml buffer A consisting of 50mM NaH₂PO₄ pH 8, 300mM NaCl, and 10mM Imidazole pH 8. Cells were lysed using sonication and a further 15ml of buffer A was added before centrifugation (40,000×g for 40min at 7 °C). The supernatant was transferred to a 50ml 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, 30mM Imidazole pH 8) and 10 CV buffer C (buffer A, 50mM Imidazole pH 8). Elution of target protein was carried out using 8 CV buffer D (buffer A, 250mM Imidazole pH 8). The purity of the sample was assessed by SDS-PAGE. The protein was dialyzed into 2 L of 20mM Tris-HCl pH 8, 150mM NaCl using SnakeSkin™ Dialysis Tubing (10 K MWCO, Thermo Fisher Scientific) overnight at 7 °C. Protein was concentrated using a 15ml Amicon® Ultra centrifugal filter unit (50kDa MWCO, Merck Life Science) to about 130mgml⁻¹ from an initial concentration of 2.3 mgml⁻¹. Protein concentration was estimated by UV absorption at 280nm using the theoretical extinction coefficient calculated by ExPASy (Gasteiger et al., 2003).

4.2 | Crystallization, data collection, and processing

As BpaC⁷⁴¹⁻¹⁰⁵⁴ is extremely soluble, initial crystallization screens for BpaC⁷⁴¹⁻¹⁰⁵⁴ were set up using about 130mgml⁻¹ of protein using the sitting-drop vapor-diffusion method with a drop volume of 200nl, the protein-to-reservoir ratio of 1:1 and a reservoir volume of 25 μl 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 (NH₄)₂SO₄. Here, 400nl of cryo-protectant buffer (0.15M HEPES pH 6.5, 1.2 M (NH₄)₂SO₄, 35% Glycerol) was added directly to the drop and incubated for 1 min before transferring to a liquid nitrogen container for data collection.

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 BpaC⁷⁴¹⁻¹⁰⁵⁴ in the Advanced Search of the Protein Data Bank (PDB; Burley et al., 2019) and using the top hit (3S6L) 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, BpaC⁴³¹⁻¹⁰⁵⁴, 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 BpaC⁷⁴¹⁻¹⁰⁵⁴. 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://pymol>

[wiki.org/index.php/Psico](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/A0A0H3HIJ5>) was annotated using a combination of results from PSIPRED (Buchan & Jones, 2019), daTAA (Szczeny & 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 BpaC⁷⁴¹⁻¹⁰⁵⁴ 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 | Phylogenetic 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 μs 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 ccpraj module within Amber20 using a grid spacing of 1 Å and a grid box size of 200 Å³ 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 (PDB); accession code 7O23.

ORCID

Andreas R. Kiessling  <https://orcid.org/0000-0001-9575-8153>
 Sarah A. Harris  <https://orcid.org/0000-0002-2812-1651>
 Kathleen M. Weimer  <https://orcid.org/0000-0003-2937-2994>
 Geoffrey Wells  <https://orcid.org/0000-0002-0253-911X>
 Adrian Goldman  <https://orcid.org/0000-0001-8032-9700>

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](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.

- International Journal of Medical Microbiology*, 305(2), 265–275. <https://doi.org/10.1016/j.ijmm.2014.12.010>
- Bertelli, C., Laird, M.R., Williams, K.P., Simon Fraser University Research Computing Group, Lau, B.Y., Hoad, G. et al. (2017) IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Research*, 45(W1), W30–W35. <https://doi.org/10.1093/nar/gkx343>
- Blum, M., Chang, H.Y., Chuguransky, S., Grego, T., Kandasamy, S., Mitchell, A. et al. (2021) The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*, 49(D1), D344–D354. <https://doi.org/10.1093/nar/gkaa977>
- Buchan, D.W.A. & Jones, D.T. (2019) The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Research*, 47(W1), W402–W407. <https://doi.org/10.1093/nar/gkz297>
- Burley, S.K., Berman, H.M., Bhikadiya, C., Bi, C., Chen, L., di Costanzo, L. et al. (2019) RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Research*, 47(D1), D464–D474. <https://doi.org/10.1093/nar/gky1004>
- Burntack, M.N., Brett, P.J., Harding, S.V., Ngugi, S.A., Ribot, W.J., Chantratita, N. et al. (2011) The cluster 1 type VI secretion system is a major virulence determinant in *Burkholderia pseudomallei*. *Infection and Immunity*, 79(4), 1512–1525. <https://doi.org/10.1128/IAI.01218-10>
- Campos, C.G., Byrd, M.S. & Cotter, P.A. (2013) Functional characterization of *Burkholderia pseudomallei* trimeric autotransporters. *Infection and Immunity*, 81(8), 2788–2799. <https://doi.org/10.1128/IAI.00526-13>
- Case, D.A. et al. (2021) *Amber 2021*. San Francisco: University of California.
- Cheng, A.C. (2010) Melioidosis: advances in diagnosis and treatment. *Current Opinion in Infectious Diseases*, 23(6), 554–559. <https://doi.org/10.1097/QCO.0b013e32833fb88c>
- Connors, R., Hill, D.J., Borodina, E., Agnew, C., Daniell, S.J., Burton, N.M. et al. (2008) The *Moraxella* adhesin UspA1 binds to its human CEACAM1 receptor by a deformable trimeric coiled-coil. *The EMBO Journal*, 27(12), 1779–1789. <https://doi.org/10.1038/emboj.2008.101>
- Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica. Section D, Biological Crystallography*, 62(Pt 9), 1002–1011. <https://doi.org/10.1107/S0907444906022116>
- Cowtan, K. (2012) Completion of autobuilt protein models using a database of protein fragments. *Acta Crystallographica. Section D, Biological Crystallography*, 68(Pt 4), 328–335. <https://doi.org/10.1107/S0907444911039655>
- Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Research*, 14(6), 1188–1190. <https://doi.org/10.1101/gr.849004>
- Duangurui, T., Indrawattana, N. & Pumirat, P. (2018) *Burkholderia pseudomallei* adaptation for survival in stressful conditions. *BioMed Research International*, 2018, 3039106. <https://doi.org/10.1155/2018/3039106>
- Edwards, T. E., Gardberg, A. S., Lafontaine, E. R. & Seattle Structural Genomics Center for Infectious Disease (SSGICD), "Crystal structure of a YadA-like head domain of the trimeric autotransporter adhesin BoaA from *Burkholderia pseudomallei*," 2011, doi: <https://doi.org/10.2210/pdb3S6L/pdb>.
- El Tahir, Y. & Skurnik, M. (2001) YadA, the multifaceted *Yersinia* adhesin. *International Journal of Medical Microbiology*, 291(3), 209–218. <https://doi.org/10.1078/1438-4221-00119>
- Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. (2010) Features and development of Coot. *Acta Crystallographica. Section D, Biological Crystallography*, 66(Pt 4), 486–501. <https://doi.org/10.1107/S0907444910007493>
- Evans, P.R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallographica. Section D, Biological Crystallography*, 67(Pt 4), 282–292. <https://doi.org/10.1107/S090744491003982X>
- Evans, P.R. & Murshudov, G.N. (2013) How good are my data and what is the resolution? *Acta Crystallographica. Section D, Biological Crystallography*, 69(Pt 7), 1204–1214. <https://doi.org/10.1107/S0907444913000061>
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. & Bairoch, A. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, 31(13), 3784–3788. <https://doi.org/10.1093/nar/gkg563>
- Gong, L., Cullinane, M., Treerat, P., Ramm, G., Prescott, M., Adler, B. et al. (2011) The *Burkholderia pseudomallei* type III secretion system and BopA are required for evasion of LC3-associated phagocytosis. *PLoS One*, 6(3), e17852. <https://doi.org/10.1371/journal.pone.0017852>
- Hartmann, M.D., Grin, I., Dunin-Horkawicz, S., Deiss, S., Linke, D., Lupas, A.N. et al. (2012) Complete fiber structures of complex trimeric autotransporter adhesins conserved in enterobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 109(51), 20907–20912. <https://doi.org/10.1073/pnas.1211872110>
- Hartmann, M.D., Ridderbusch, O., Zeth, K., Albrecht, R., Testa, O., Woolfson, D.N. et al. (2009) A coiled-coil motif that sequesters ions to the hydrophobic core. *Proceedings of the National Academy of Sciences of the United States of America*, 106(40), 16950–16955. <https://doi.org/10.1073/pnas.0907256106>
- Hernandez Alvarez, B., Hartmann, M.D., Albrecht, R., Lupas, A.N., Zeth, K. & Linke, D. (2008) A new expression system for protein crystallization using trimeric coiled-coil adaptors. *Protein Engineering, Design & Selection*, 21(1), 11–18. <https://doi.org/10.1093/protein/gzm071>
- Hoiczky, E., Roggenkamp, A., Reichenbecher, M., Lupas, A. & Heesemann, J. (2000) Structure and sequence analysis of *Yersinia* YadA and *Moraxella* UspAs reveal a novel class of adhesins. *The EMBO Journal*, 19(22), 5989–5999. <https://doi.org/10.1093/emboj/19.22.5989>
- Holm, L. & Laakso, L.M. (2016) Dali server update. *Nucleic Acids Research*, 44(W1), W351–W355. <https://doi.org/10.1093/nar/gkw357>
- Hovingh, E.S., van den Broek, B. & Jongerius, I. (2016) Hijacking complement regulatory proteins for bacterial immune evasion. *Frontiers in Microbiology*, 7, 2004. <https://doi.org/10.3389/fmicb.2016.02004>
- Ishikawa, M., Nakatani, H. & Hori, K. (2012) AtaA, a new member of the trimeric autotransporter adhesins from *Acinetobacter* sp. Tol 5 mediating high adhesiveness to various abiotic surfaces. *PLoS One*, 7(11), e48830. <https://doi.org/10.1371/journal.pone.0048830>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O. et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Kabsch, W. (2010) XDS. *Acta Crystallographica. Section D, Biological Crystallography*, 66(Pt 2), 125–132. <https://doi.org/10.1107/S0907444909047337>
- Kaiser, P.O., Riess, T., Wagner, C.L., Linke, D., Lupas, A.N., Schwarz, H. et al. (2008) The head of *Bartonella* adhesin A is crucial for host cell interaction of *Bartonella henselae*. *Cellular Microbiology*, 10(11), 2223–2234. <https://doi.org/10.1111/j.1462-5822.2008.01201.x>
- Kanehisa, M. & Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1), 27–30. <https://doi.org/10.1093/nar/28.1.27>
- Kibbe, W.A. (2007) OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Research*, 35, W43–W46. <https://doi.org/10.1093/nar/gkm234>
- KieSSLing, A.R., Malik, A. & Goldman, A. (2020) Recent advances in the understanding of trimeric autotransporter adhesins. *Medical Microbiology and Immunology*, 209, 233–242. <https://doi.org/10.1007/s00430-019-00652-3>

- Koiwai, K., Hartmann, M.D., Linke, D., Lupas, A.N. & Hori, K. (2016) Structural basis for toughness and flexibility in the c-terminal passenger domain of an *Acinetobacter* trimeric autotransporter adhesin. *The Journal of Biological Chemistry*, 291(8), 3705–3724. <https://doi.org/10.1074/jbc.M115.701698>
- Kozłowski, L.P. (2021) IPC 2.0: prediction of isoelectric point and pKa dissociation constants. *Nucleic Acids Research*, 49(W1), W285–W292. <https://doi.org/10.1093/nar/gkab295>
- Kumar, S., Stecher, G., Li, M., Nknyaz, C. & Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35(6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Lafontaine, E.R., Balder, R., Michel, F. & Hogan, R.J. (2014) Characterization of an autotransporter adhesin protein shared by *Burkholderia mallei* and *Burkholderia pseudomallei*. *BMC Microbiology*, 14(1), 92. <https://doi.org/10.1186/1471-2180-14-92>
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P., McWilliam, H. et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), 2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>
- Lazar Adler, N.R., Stevens, M.P., Dean, R.E., Saint, R.J., Pankhania, D., Prior, J.L. et al. (2015) Systematic mutagenesis of genes encoding predicted autotransported proteins of *Burkholderia pseudomallei* identifies factors mediating virulence in mice, net intracellular replication and a novel protein conferring serum resistance. *PLoS One*, 10(4), e0121271. <https://doi.org/10.1371/journal.pone.0121271>
- Leo, J.C., Lyskowski, A., Hattula, K., Hartmann, M.D., Schwarz, H., Butcher, S.J. et al. (2011) The structure of *E. coli* IgG-binding protein D suggests a general model for bending and binding in trimeric autotransporter adhesins. *Structure*, 19(7), 1021–1030. <https://doi.org/10.1016/j.str.2011.03.021>
- Letunic, I. & Bork, P. (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research*, 47(W1), W256–W259. <https://doi.org/10.1093/nar/gkz239>
- Ludwiczak, J., Winski, A., Szczepaniak, K., Alva, V. & Dunin-Horkawicz, S. (2019) DeepCoil-a fast and accurate prediction of coiled-coil domains in protein sequences. *Bioinformatics*, 35(16), 2790–2795. <https://doi.org/10.1093/bioinformatics/bty1062>
- Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N. et al. (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, 47(W1), W636–W641. <https://doi.org/10.1093/nar/gkz268>
- Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E. & Simmerling, C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8), 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. & Read, R.J. (2007) Phaser crystallographic software. *Journal of Applied Crystallography*, 40(Pt 4), 658–674. <https://doi.org/10.1107/S0021889807021206>
- Müller, N.F., Kaiser, P.O., Linke, D., Schwarz, H., Riess, T., Schäfer, A. et al. (2011) Trimeric autotransporter adhesin-dependent adherence of *Bartonella henselae*, *Bartonella quintana*, and *Yersinia enterocolitica* to matrix components and endothelial cells under static and dynamic flow conditions. *Infection and Immunity*, 79(7), 2544–2553. <https://doi.org/10.1128/IAI.01309-10>
- Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A. et al. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica. Section D, Biological Crystallography*, 67(Pt 4), 355–367. <https://doi.org/10.1107/S0907444911001314>
- Nummelin, H., Merckel, M.C., Leo, J.C., Lankinen, H., Skurnik, M. & Goldman, A. (2004) The *Yersinia* adhesin YadA collagen-binding domain structure is a novel left-handed parallel beta-roll. *The EMBO Journal*, 23(4), 701–711. <https://doi.org/10.1038/sj.emboj.7600100>
- Schrodinger, L. (2017) *The PyMOL molecular graphics system, version 2.0*. Oxford: Oxford University Press.
- Stein, N. (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *Journal of Applied Crystallography*, 41(3), 641–643. <https://doi.org/10.1107/S0021889808006985>
- Szczesny, P. & Lupas, A. (2008) Domain annotation of trimeric autotransporter adhesins—daTAA. *Bioinformatics*, 24(10), 1251–1256. <https://doi.org/10.1093/bioinformatics/btn118>
- Szczesny, P. et al. (2008) Structure of the head of the *Bartonella* adhesin BadA. *PLoS Pathogens*, 4(8), e1000119. <https://doi.org/10.1371/journal.ppat.1000119>
- Tan, T.T., Forsgren, A. & Riesbeck, K. (2006) The respiratory pathogen *Moraxella catarrhalis* binds to laminin via ubiquitous surface proteins A1 and A2. *The Journal of Infectious Diseases*, 194(4), 493–497. <https://doi.org/10.1086/505581>
- Thibau, A., Hipp, K., Vaca, D.J., Chowdhury, S., Malmström, J., Saragliadis, A. et al. (2022) Long-read sequencing reveals genetic adaptation of *Bartonella* adhesin A among different *Bartonella henselae* isolates. *Frontiers in Microbiology*, 13, 838267. <https://doi.org/10.3389/fmicb.2022.838267>
- Vaca, D.J., Thibau, A., Schütz, M., Kraiczky, P., Happonen, L., Malmström, J. et al. (2020) Interaction with the host: the role of fibronectin and extracellular matrix proteins in the adhesion of Gram-negative bacteria. *Medical Microbiology and Immunology*, 209(3), 277–299. <https://doi.org/10.1007/s00430-019-00644-3>
- Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. & Barton, G.J. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), 1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>
- Wiersinga, W.J., van der Poll, T., White, N.J., Day, N.P. & Peacock, S.J. (2006) Melioidosis: insights into the pathogenicity of *Burkholderia pseudomallei*. *Nature Reviews. Microbiology*, 4(4), 272–282. <https://doi.org/10.1038/nrmicro1385>
- Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N. et al. (2018) MolProbity: more and better reference data for improved all-atom structure validation. *Protein Science: A Publication of the Protein Society*, 27(1), 293–315. <https://doi.org/10.1002/pro.3330>
- Winsor, G.L., Khaira, B., Van Rossum, T., Lo, R., Whiteside, M.D. & Brinkman, F.S. (2008) The *Burkholderia* Genome Database: facilitating flexible queries and comparative analyses. *Bioinformatics*, 24(23), 2803–2804. <https://doi.org/10.1093/bioinformatics/btn524>
- Zhou, K., Aertsen, A. & Michiels, C.W. (2014) The role of variable DNA tandem repeats in bacterial adaptation. *FEMS Microbiology Reviews*, 38(1), 119–141. <https://doi.org/10.1111/1574-6976.12036>

SUPPORTING INFORMATION

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

How to cite this article: Kiessling, A. R., Harris, S. A., Weimer, K. M., Wells, G. & Goldman, A. (2022). The C-terminal head domain of *Burkholderia pseudomallei* BpaC has a striking hydrophilic core with an extensive solvent network. *Molecular Microbiology*, 118, 77–91. <https://doi.org/10.1111/mmi.14953>