The structure of the OmpA/Pal protein of *Anaplasma phagocytophilum*

Cadby, I.T.

Address: Bristol Veterinary School, University of Bristol, Langford, Bristol, BS40 5DU, UK.

Correspondence: ian.cadby@bristol.ac.uk

ABSTRACT

Peptidoglycan associated lipoprotein (Pal) contributes to cell division in many Gram-negative bacteria via its interactions with the outer membrane, peptidoglycan, and other periplasmic proteins. Intriguingly, Pal and Outer Membrane Protein A (OmpA), which also interacts with the outer membrane and peptidoglycan, have moonlighting roles as virulence factors and facilitate infection and/or host-pathogen interactions in a range of bacteria. The Pal family protein of Anaplasma phagocytophilum, which is a tick-borne pathogen that infects a wide range of hosts, seems to function primarily as a virulence factor, since this bacterium lacks a peptidoglycan cell wall. Here we present crystal structures of the Pal protein of *A. phagocytophilum* (commonly referred to as an OmpA protein), demonstrating that this protein has amino acid insertions that confer flexibility and that this insertion is also found in the Pal proteins of some related bacteria, which are also pathogens. Whether this flexibility is reflective of any adaptations for host-pathogen interactions remains to be determined.

INTRODUCTION

Bacteria of genus *Anaplasma* lead obligate intracellular lifestyles and are entirely reliant on eukaryotic host cells for their survival and replication[1]. These Gram-negative bacteria infect ticks and a wide range of vertebrates including wild and domestic animals, livestock, and humans. Generally recognised as infecting the blood cells of their vertebrate hosts, different members of genus *Anaplasma* have varied host species and cell tropisms. For example, *Anaplasma marginale* and *Anaplasma ovis* preferentially infect the erythrocytes of ruminants whereas *Anaplasma phagocytophilum,* a zoonotic agent with a particularly wide host range, shows preference for neutrophils. *Anaplasma* infection can cause disease within their vertebrate hosts and tick-borne fever and bovine anaplasmosis cause significant losses to livestock industries across the globe[2,3]. Infection of humans with *A. phagocytophilum* causes human granulocytic anaplasmosis: an emerging and occasionally fatal disease[4]**.**

Peptidoglycan associated lipoprotein (Pal) and outer membrane protein A (OmpA) are conserved in most Gram-negative bacteria[5,6]. These proteins are integrated into or bound to the outer membrane and bind to peptidoglycan PG, tethering these major structural components to one another. Whilst OmpA bridging between the outer membrane and PG is regarded as serving a primarily structural function, Pal has roles in cell division known, such as invagination of the cell membrane and septum development, in addition to a structural role[7-9]. Both OmpA and Pal have moonlighting roles as virulence factors in some bacteria[10,11]. OmpA and Pal proteins are implicated in adhesion, biofilm formation, and immune modulation; and have been identified as vaccine candidates.

The genomes of *Ap*, and some other Rickettsiales (the bacterial Order to which *Ap* belongs) lack the full complement of genes necessary for PG biosynthesis[12-14]. Indeed, fluorescent labelling of PG structures in various Rickettsiales, supports that *Ap* does not synthesise PG[15]. Instead of binding PG, the OmpA protein of *Ap* ($OmpA^{Ap}$, APH_0338) functions as an invasin to facilitate infection of mammalian host cells[16-18]. Intriguingly, all evidence demonstrates that $OmpA^{Ap}$ recognises sialic acid and fucose moieties of mammalian cell surface glycoproteins, specifically sialyl-Lexis-X. To gain insights into how $OmpA^{Ap}$ binds its host receptors, and whether the ability to bind host glycoproteins is reliant on residues involved in its ancestral function of PG binding, we solved the crystal structure of $OmpA^{Ap}$ to a resolution of 1.9 Å and assessed its potential to bind sialyl-Lewis-X and other sugars. Our results demonstrate that $OmpA^{Ap}$ is a Pal protein and contains amino acid insertions, generating an extended loop region proximal to its PG binding site. Whilst *in vitro* binding of recombinant $OmpA^{Ap}$ to sialyl-Lewis-X could not be detected, this flexible loop region is also found in the Pal proteins of other pathogens which are related to *Ap*, suggesting it could represent an adaptation of this protein to its role as a virulence factor.

METHODS

Cloning, protein expression and purification

The APH_0338 gene, corresponding to amino acids 10-205, was amplified from Anaplasma phagocytophilum HZ gDNA using Q5 DNA polymerase (New England Biolabs). Cloning into pET-28a and a modified pET41 was performed using restriction free PCR-based methods[19]. The cloning yielded two T7 inducible overexpression vectors: pET_APH0338N, encoding

APH_0338 lacking its N-terminal signal peptide and with an N-terminal 6 x histidine tag and pET_APH0338C, encoding the same gene sequence but with a C-terminal 8 x histidine tag. The constructs were transformed into *E. coli* DE3 and then used for protein over-expression. Briefly, 10-20 ml cultures of *Escherichia coli* strains were grown in LB broth at 37 °C overnight and used as the inoculum for large-scale (1-2 L) cultures in auto-induction medium. Cultures were grown at 37 °C with shaking for approximately 3 hours until an OD_{600} reading of 0.6-0.8 had been reached, at which point the temperature was reduced to 18 °C. After an additional 18 hours of growth the cultures were harvested by centrifugation and cell pellets were then frozen for later use in protein purification.

To purify OmpA proteins, cell pellets (10-25 g) were resuspended in Buffer A (20 mM imidazole-HCl pH 8.0; 50 mM Tris-HCl pH 8.0; 400 mM NaCl; 2 mM BME) supplemented with 10 ug/ml hen egg lysozyme and 0.02 % Tween 20 and lysed by sonication. Lysates were clarified by centrifugation at 20,000 rpm and then passed over nickel agarose beads in gravity flow columns. Beads were washed extensively with Buffer A, then with 15 column volumes of Buffer A supplemented with 10 % Buffer B (as Buffer A but with 400 mM imidazole, pH 8.0), then with 15 column volumes of Buffer A supplemented with 20 % Buffer B, before final washes with Buffer B. The eluates were analysed by SDS-PAGE, indicating that high purity OmpA was collected in the Buffer B washes. Proteins were dialysed into Buffer C (Tris-HCl, pH 8.0; 250 mM NaCl; 1 mM BME) and then further purified by gel filtration in Buffer C using a Superdex 200 26/60 column. OmpA proteins typically eluted at a volume consistent with a protein with a MW of 20 kDa. This monodisperse peak was pooled and concentrated to 20 mg/ml with a 10 kDa MWCO ultrafiltration device (Sartorius).

Protein crystallography

Crystallization of OmpA proteins was achieved by mixing equal volumes of concentrated purified protein and various Molecular Dimensions crystallization screens in sitting drop trays and then incubation at 18 °C. Crystals of C-terminally tagged OmpA grew after four weeks in condition A10 of screen JCSG. These small protein crystals were briefly transferred into a drop of mother liquor supplemented with 25 % ethylene glycol prior to snap freezing in liquid nitrogen. X-ray diffraction data was collected at the Diamond Light Source Synchrotron and processed using the CCP4 and Phenix software suites. The structure of OmpA was initially solved using by molecular replacement using the structure of Pal from *Burkholderia cepacian* (PDB:5LKW) as the search query. This model was refined using Phenix, manual adjustments using Coot software, and the PDB-redo server. Attempts were also made to crystallize OmpA in the presence of sialyl Lewis X, fucose, and sialic acid. Trials involved protein supplemented with a 2-fold molar excess of the proposed ligand but yielded no crystals. Attempts were also made to soak these ligands into apo-OmpA crystals: mother liquor was supplemented with 2 mM of the ligand and added to OmpA protein crystal containing drops. Crystals were harvested at various time points, from 5 minutes after addition up until 24 hours after addition before harvesting as described above. None of these strategies yielded electron density maps consistent with ligand bound protein.

RESULTS

Architecture of *A. phagocytophilum* OmpA

The OmpA^{Ap} protein is comprised of an N-terminal signal peptide and lipobox, a central predicted peptidoglycan associated lipodomain (PAL) domain, and a disordered C-terminal stretch of approximately 40 amino acids **(Figure 1A)**. Comparison of the predicted domains of OmpA^{Ap} with the known domain architecture of OmpA and Pal proteins indicates that OmpA^{Ap} lacks the N-terminal beta-barrel domain. The beta-barrel domain of *E. coli* OmpA (OmpA^{Ec}) is inserted into the bacterial outer-membrane whilst the C-terminal PAL domain is located in the periplasm where it binds to peptidoglycan. In terms of protein architecture, OmpA^{Ap} more closely resembles Pal.

Overall structure of Omp^{Ap}

Crystals of Omp^{Ap} were grown by sitting drop vapour diffusion and took approximately 60 days to appear. X-ray diffraction data was collected on these crystals, which belong to the orthorhombic C222 space group, and the Omp^{Ap} structure was solved by molecular replacement and refined to a resolution of 1.9 Å. Data processing and refinement statistics are displayed in Table 1. Two copies of Omp^{Ap} are present in the symmetric unit. The overall fold of Omp^{Ap} is similar to that of other Pal/OmpA domains and is comprised of an alpha-beta sandwich (**Figure 1B)**. No density was observed for amino acids 1-32 and 165-205, possibly reflecting the prediction that these regions are disordered. Comparison with other Pal/OmpA proteins suggests that the amino acids necessary for binding PG are conserved in Omp^{Ap}, despite the lack of PG in *Ap.*

An extended loop of Omp^{Ap} forms a domain able to form multiple conformations

The two protomers of Omp^{Ap} in the asymmetric unit are mostly identical with the exception of amino acids 136-150, which bridges β3- β4 and forms are region we term the 'lid' (Figure 1B). Superposition of these two copies of Omp^{Ap} reveals that the lid adopts two conformational states. In one conformation, chain A, residues 142-153 adopt an α-helical conformation (**Figure 2A)**. However, this helix in the other copy, chain B, is shortened to residues 149-153, with residues 142-147 adopting a coil conformation. The sum of this restructuring results in the lid region turning moving by 14.94, relative to the Cα of residue 144. These two alternative conformations of Omp^{Ap} are each stabilized by crystal contacts so it is not possible to determine how readily the lid forms these two different states. Superposition of Pal from *E. coli* (Pal^{Ec}, PDB: 1OAP) with these structures indicates that the lid region of Omp^{Ap} is extended compared to canonical members of this protein family (**Figure 2A).** A multiple alignment of Omp^{Ap} with the Pal family proteins of other Rickettsiales and representative bacteria Gram negative bacteria demonstrates that this extension is due to the insertion of amino acids, relative to most other Pal proteins **(Figure 2B).**. This insertion is seen in other Rickettsiales Pal proteins but varies in length. For example; in *Ap, A. marginale, Ehrlichia chaffeensis,* and *Wolbachia* species; this insertion is six amino acids in length, whereas in *Orientia tsutsugamushi* it is two amino acids and selected *Rickettsia* species do not have any amino acid insertion to the lid region of their Pal proteins.

DISCUSSION

The Omp^{Ap} structures presented here demonstrate that the Pal family proteins of some Rickettsiales contain inserted sequences and are divergent from typical members of this protein family. This insertion provides structural flexibility to Omp^{Ap}, proximal to its PG binding

site, which is conserved despite the lack of PG in this bacterium. It is unknown whether this switching of conformations occurs in solution, but is unlikely to impact the potential for Omp^{Ap} to bind its ancestral ligand, PG, since similar insertions are seen in Pal proteins found in Rickettsiales that do synthesise a PG cell wall, such as *A. marginale.*

Attempts to measure binding of Omp^{Ap} to its host targets, such as sialyl-Lewis-X, by various approaches such as microscale thermophoresis or isothermal calorimetry were unsuccessful, as were attempts to co-crystallize Omp^{Ap} with this sugar (data not shown). This is consistent with the proposed hypothesis that Omp^{Ap} binds to host surfaces co-operatively with other *Ap* adhesins and/or invasins. The potential role that insertions in Pal proteins might play in mediating host-pathogen interactions, which seems to be a conserved function for the Pal proteins of *Anaplasma*, is unknown.

REFERENCES

1. O'Neal, A. J., Singh, N., Mendes, M. T., & Pedra, J. H. (2021). The genus Anaplasma: drawing back the curtain on tick–pathogen interactions. Pathogens and Disease, 79(5), ftab022.
2. Kocan, K. M., De la Fuente, J., Guglielmone, A. A., & Meléndez, R. D. (2003). Antigens and alternatives for control of Anaplasma marginale infection in cattle. Clinical microbiology reviews, 16(4), 698-712.
3. Ogden, N. H., & Lindsay, L. R. (2016). Effects of climate and climate change on vectors and vector-borne diseases: ticks are different. Trends in parasitology, 32(8), 646-656.

4. Dumler, J. S., Choi, K. S., Garcia-Garcia, J. C., Barat, N. S., Scorpio, D. G., Garyu, J. W., Grab, D.J., & Bakken, J. S. (2005). Human granulocytic anaplasmosis and Anaplasma phagocytophilum. *Emerging infectious diseases*, *11*(12), 1828.
5. Wang, Y. (2002). The function of OmpA in Escherichia coli. Biochemical and biophysical research communications, 292(2), 396-401.
6. Parsons, L. M., Lin, F., & Orban, J. (2006). Peptidoglycan recognition by Pal, an outer membrane lipoprotein. Biochemistry, 45(7), 2122-2128.
7. Gerding, M. A., Ogata, Y., Pecora, N. D., Niki, H., & De Boer, P. A. (2007). The trans-envelope Tol–Pal complex is part of the cell division machinery and required for proper outer-membrane invagination during cell constriction in E. coli. Molecular microbiology, 63(4), 1008-1025.
8. Yakhnina, A. A., & Bernhardt, T. G. (2020). The Tol-Pal system is required for peptidoglycan-cleaving enzymes to complete bacterial cell division. Proceedings of the National Academy of Sciences, 117(12), 6777-6783.
9. Baccelli, P., Rachedi, R., Serrano, B., Petiti, M., Bernard, C. S., Houot, L., & Duche, D. (2022). Timing of TolA and TolQ recruitment at the septum depends on the functionality of the Tol-Pal system. Journal of Molecular Biology, 434(7), 167519.
10. Weiser, J. N., & Gotschlich, E. C. (1991). Outer membrane protein A (OmpA) contributes to serum resistance and pathogenicity of Escherichia coli K-1. Infection and Immunity, 59(7), 2252-2258.

11. Zeng, X., Wang, N., Xiang, C., Liu, Q., Li, D., Zhou, Y., ... & Shi, Y. (2023). Peptidoglycan-associated lipoprotein contributes to the virulence of Acinetobacter baumannii and serves as a vaccine candidate. Genomics, 115(2), 110590.
12. Lin, M., & Rikihisa, Y. (2003). Ehrlichia chaffeensis and Anaplasma phagocytophilum lack genes for lipid A biosynthesis and incorporate cholesterol for their survival. Infection and immunity, 71(9), 5324-5331.
13. Otten, C., Brilli, M., Vollmer, W., Viollier, P. H., & Salje, J. (2018). Peptidoglycan in obligate intracellular bacteria. Molecular microbiology, 107(2), 142-163.
14. Dunning Hotopp, J. C., Lin, M., Madupu, R., Crabtree, J., Angiuoli, S. V., Eisen, J., ... & Tettelin, H. (2006). Comparative genomics of emerging human ehrlichiosis agents. PLoS genetics, 2(2), e21.
15. Atwal, S., Chuenklin, S., Bonder, E. M., Flores, J., Gillespie, J. J., Driscoll, T. P., & Salje, J. (2021). Discovery of a diverse set of bacteria that build their cell walls without the canonical peptidoglycan polymerase aPBP. *MBio*, *12*(4), e01342-21.
16. Ojogun, N., Kahlon, A., Ragland, S. A., Troese, M. J., Mastronunzio, J. E., Walker, N. J., ieBrock, L., Thomas, R.J., Borjesson, D.L., Fikrig, E., & Carlyon, J. A. (2012). Anaplasma phagocytophilum outer membrane protein A interacts with sialylated glycoproteins to promote infection of mammalian host cells. Infection and immunity, 80(11), 3748-3760.
17. Seidman, D., Hebert, K. S., Truchan, H. K., Miller, D. P., Tegels, B. K., Marconi, R. T., & Carlyon, J. A. (2015). Essential domains of Anaplasma phagocytophilum invasins utilized to infect mammalian host cells. PLoS pathogens, 11(2), e1004669.

18. Hebert, K. S., Seidman, D., Oki, A. T., Izac, J., Emani, S., Oliver Jr, L. D., Miller, D.P., Tegels, B.K., Kannagi, R., Marconi, R.T. & Carlyon, J. A. (2017). Anaplasma marginale outer membrane protein A is an adhesin that recognizes sialylated and fucosylated glycans and functionally depends on an essential binding domain. Infection and Immunity, 85(3), e00968-16.
19. Van Den Ent, F., & Löwe, J. (2006). RF cloning: a restriction-free method for inserting target genes into plasmids. Journal of biochemical and biophysical methods, 67(1), 67-74.

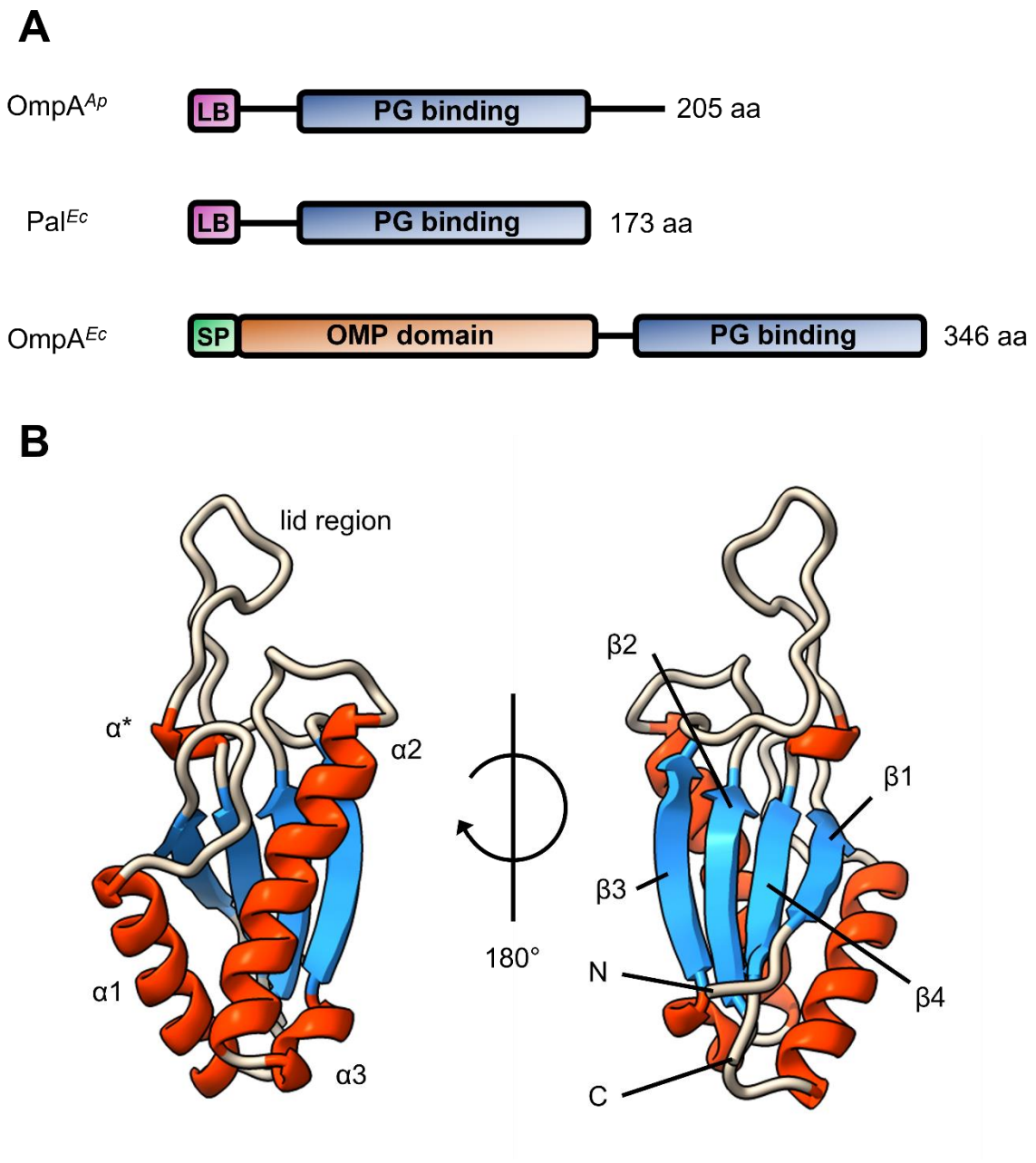

Figure 1: Structural characterisation of OmpA^{Ap}. A. Schematic representation of the protein architecture of OmpA and Pal proteins from *E. coli* and *Anaplasma phagocytophilum.* Peptidoglycan binding (PAL) domains (PG binding), Lipoboxes (LB),Signal peptides (SP), and Outer Membrane Protein (OMP) domains are labelled. B. Cartoon representation of the crystal structure of OmpA^{Ap}, solved to 1.9 Å. α* marks a variable length helix.

Table 1: Data collection and refinement statistics

	OmpA^{Ap}
Data collection	
Spacegroup	C 2 2 2
Cell dimensions	
a,b,c (Å)	133.4 138.5 34.7
α, β, γ (°)	90 90 90
Resolution	69.3 – 1.90
I/σI	12.11 2.45)
Completeness (%)	99.65 (97.71)
Redundancy	2.0 (2.0)
Refinement	
Resolution (Å)	1.9
No. of reflections	26011
R_{work}/R_{free}	0.1797 / 0.2175
No. of atoms	
Protein	1908
Ligand/ion	6
Water	340
B factors	
Protein	21.54
Ligand/ion	15.54
Water	33.29
RMSD	
Bond lengths (Å)	0.007
Bond angles (°)	1.15

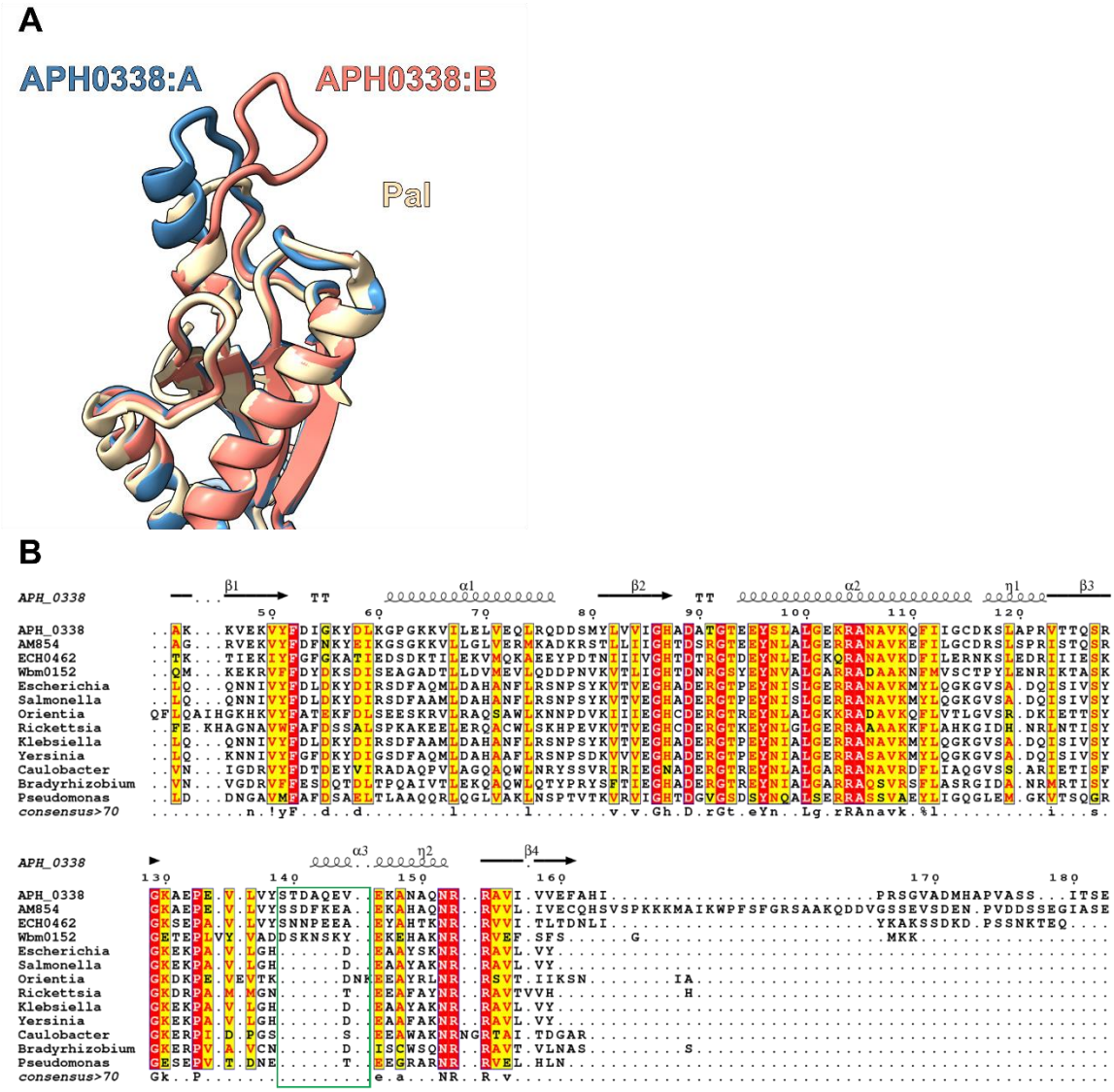

Figure 2: Structure and sequence comparisons between Omp^{Ap} and representative Pal family proteins of diverse Gram-negative bacteria. A. Superposition of two conformations of Omp^{Ap} and Pal^{Ec}. The lid region of these Pal family proteins is shown, detailing the amino acid insertion and apparent flexibility of Omp^{Ap}. B. Multiple sequence alignment of Pal family proteins. The inserted amino acids of the lid region are marked with a green box.