Glycine acylation and trafficking of a new class of bacterial lipoprotein by a composite secretion system.

Christopher Icke[1], Freya J. Hodges[1], Karthik Puella[1], Samantha A. McKeand[2], Jack A. Bryant[2], Adam F. Cunningham[2,3], Jeff A. Cole[2], Ian R. Henderson[1*].

[1]Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia; [2]Institute of Microbiology and Infection and [3]Institute of Immunology and Immunotherapy, University of Birmingham, Birmingham, UK

[*]For correspondence: Ian Henderson (i.henderson@uq.edu.au)

Abstract

Protein acylation is critical for many cellular functions including signal transduction, cell division and development. In bacteria, such lipoproteins have important roles in virulence and are therefore potential targets for the development of novel antimicrobials and vaccines. To date, all known bacterial lipoproteins are secreted from the cytosol via the Sec pathway, acylated on an N-terminal cysteine residue through the action of Lgt, Lsp and Lnt, and then targeted to the appropriate cellular location. In the case of Gram-negative bacteria, the lipoprotein trafficking Lol pathway transports the lipoproteins to the outer membrane where most substrate molecules are retained within the cell. Here we identify a new secretion pathway that displays the substrate lipoprotein on the cell surface. We demonstrate that the previously identified *E. coli* Aat secretion system is a composite system that shares similarity with type I secretion systems and elements of the Lol pathway. Remarkably, during secretion by the Aat system, the AatD subunit acylates the substrate CexE on a highly conserved N-terminal glycine residue (rather than the canonical cysteine). Mutations in AatD or CexE that disrupt glycine acylation interfere with membrane incorporation and trafficking. Our data suggest that CexE is the first member of a new class of glycine-acylated bacterial lipoprotein, while Aat represents a new secretion system that we propose be defined as a lipoprotein secretion system (LSS).

Introduction

Protein acylation by the covalent attachment of fatty acids occurs for hundreds of proteins in eukaryotic and prokaryotic organisms. This event confers distinct biochemical properties, enabling acylation to regulate intracellular trafficking, subcellular localization, protein-protein and protein-lipid interactions and are of obvious importance to cell biology. As a consequence, lipoproteins are key components of bacterial pathogens and have been targeted for antibiotic and vaccine development (Dev et al., 1985). In Gram-negative bacteria, lipoproteins are transported across the inner membrane from the cytosol to the periplasm by the Sec pathway. However, in Gram-positive cells lipoproteins can also be secreted by the TAT pathway (Thompson et al., 2010; Widdick et al., 2011). The signal peptide inserts into the inner membrane and a diacylglycerol is attached by Lgt to the sulphur moiety of the invariant cysteine at the +1 position of the lipoprotein. The signal sequence is then cleaved by Lsp exposing the N-terminal amine group of the cysteine for monoacylation by Lnt (Gupta et al., 1993; Masao Tokunaga et al., 1984). Lnt mediated acylation occurs at the free amine moiety of the N-terminal cysteine after signal sequence cleavage by Lsp (Signal Peptidase II). The mature triacylated lipoprotein remains embedded in the inner membrane or is localised to the inner leaflet of the outer membrane by the essential Lol pathway. LolCE, in combination with the ATPase LolD, extracts the triacylated lipoprotein from the inner membrane for transport across the periplasm by LolA, where it is then incorporated into the inner leaflet of the outer membrane by LolB. While the majority remain periplasmically located, in recent years there have been a number of descriptions of surface localised outer membrane lipoproteins (Baldi et al., 2012; Cowles et al., 2012; Konovalova et al., 2014). However, the mechanism of translocation across the outer membrane to the cell surface remains poorly understood.

Previously, we described two outer membrane proteins, CexE and Aap, associated with enterotoxigenic (ETEC) and enteroaggregative *Escherichia coli* (EAEC), respectively (Crossman et al., 2010; Sheikh et al., 2002). CexE has been implicated in prolonged bacterial shedding and increased severity of infection (Rivas et al., 2020) whereas Aap influences biofilm formation and gut colonisation (Belmont-Monroy et al., 2020; Sheikh et al., 2002). Despite Aap and CexE sharing only 18% amino

acid identity, both proteins are secreted by the Aat system (Belmont-Monroy et al., 2020; Nishi et al., 2003; Rivas et al., 2020). The Aat system requires five proteins (AatPABCD) to facilitate protein secretion, two of which bear resemblance to components of the Type 1 secretion system (T1SS); an outer membrane protein (OMP), and a periplasmic adaptor protein (PAP). In contrast to the T1SS, the CexE and Aap substrate molecules are translocated using a two-step mechanism. First Aap/CexE is translocated across the inner membrane into the periplasm by the Sec pathway (Pilonieta et al., 2007). It then enters the Aat system to be secreted across the outer membrane (Belmont-Monroy et al., 2020; Nishi et al., 2003; Rivas et al., 2020).

While further characterising the two-step secretion mechanism of the Aat system, we noticed that during secretion CexE is post-translationally modified by the Aat system and this modification is required for the secretion. Here we reveal that following cleavage of the Sec-dependent signal sequence, the N-terminus of CexE is modified by the addition of an acyl chain. We demonstrate that AatD is a homolog of the apolipoprotein N-acyltransferase (Lnt). We reveal AatD is both necessary and sufficient for monoacylation of CexE and Aap. However, in contrast to Lol lipoprotein substrates, CexE lacks an N-terminal cysteine and instead an invariant glycine is the site of acylation. Furthermore, we demonstrate that the addition of an N-terminal glycine to the coding sequence of a heterologous protein was sufficient for this novel AatD-catalysed acylation event. We propose that Aap and CexE are members of a novel class of lipoprotein that are secreted through the Aat system, which is a conglomeration of the Lol pathway and a T1SS. Thus, we describe a new mechanism of N-palmitoylation. We reveal glycine as a new target of N-palmitoylation and AatD is a new acyltransferase. Consequently, we reveal a new function for acylation - protein secretion.

Results

Distribution of the Aat system

The Aat system was first identified in EAEC, where it corresponds to the molecular probe (CVD432) used to define this *E. coli* pathovar (Baudry et al., 1990; Nishi et al., 2003). In order to determine whether the Aat system was more widespread each of the Aat proteins from ETEC H0407 was used to search the non-redundant protein sequence database using repetitive iterations of the PSI-

BLAST algorithm. The full Aat system and an Aap or CexE homolog was identified in 826 separate nucleotide accessions (Supplementary Table S1). This revealed that the Aat system is distributed more widely than initially anticipated and is encoded in pathogens with diverse mechanisms of virulence such as ETEC, EAEC, enteropathogenic *E. coli*, Shiga-toxin producing *E. coli*, *Shigella* sp., *Salmonella enterica*, *Citrobacter rodentium*, *Providencia alcalifaciens* and *Yersinia entercolitica* (Crossman et al., 2010; Petty et al., 2010; Rivas et al., 2020).

To understand the conservation of the Aat system, the organisation of the *aat* operon was examined for each of the 826 genomes identified above. The relative genomic distance between each of the *aat* genes and *aap*/*cexE* was calculated from their boundaries on the nucleotide accession. From this analysis six separate organisations of the *aat* operon were identified (Fig 1). In the most common organisation *aap*/*cexE* are separated from the *aatPABCD* operon by at least 1 kb. This accounted for just over half of the *aat* systems identified. In just under a quarter of *aat* operons identified, *aap*/*cexE* are within 1 kb of the *aatPABCD* operon either within 400 bp upstream (18.4%) or 1 kb downstream (4.8%). In 25% of the *aat* systems identified *aatD* was separated from *aatPABC* by >1 kb. The most common of these organisations was *aatD* separated from the other members of the *aat* system and substrate protein (11.9%), closely followed by *aatD* and *aap*/*cexE* together but greater than 1 kb distance from for other *aat* genes (9.0%). Finally, in the least common example *aatD*, *aap*/*cexE* and *aatPABC* were all separated by greater than 1 kb (4.1%). We did not identify any examples of *aatD* encoded between 39 bp and 1 kb from the stop codon of *aatC*. However, despite these differences, in every example the *aatPABC* genes formed a single operonic unit located with *aatD* and *aap*/*cexE* on a chromosomal pathogenicity island or a large virulence plasmid. These data suggest that AatPABCD are sufficient and necessary to mediate Aap/CexE secretion.

Aat-dependent secretion of CexE

To test whether all of the *aat* genes are required for CexE secretion, we constructed single gene deletion mutants of *cexE* and each *aat* gene in ETEC H10407 pCfaD. The pCfaD plasmid encodes the CfaD transcriptional activator under the control of an arabinose inducible promoter allowing constitutive expression of the CfaD-dependent *cexE* and the *aat* genes in the presence of arabinose

(Hodson et al., 2017; Pilonieta et al., 2007). SDS-PAGE analysis of culture supernatant fractions collected from ETEC H10407 pCfaD revealed a protein with an apparent molecular mass of 11.8 kDa that could be detected by western blotting with CexE-specific polyclonal antibodies (Fig 2A). In contradistinction to the parent strain, no protein was detected in the culture supernatant fractions derived from cultures of the ETEC H10407 pCfaD *aat* or *cexE* mutants (Fig 2A). However, to ensure that the lack of CexE in the culture supernatant was not due to the lack of CexE production, but rather a result of an inability to secrete CexE, the presence of CexE in whole cell lysates was determined. CexE was detected by western blotting with CexE-specific antibodies in whole cell lysates of ETEC H10407 pCfaD and all *aat* mutants. In contrast, CexE could not be detected in the *cexE* null mutant (Fig. 2A). From these data we conclude that the Aat system is essential for CexE secretion.

Post-translational modification of Aat substrate molecules

Unexpectedly, when the whole cell lysates of ETEC H10407 pCfaD and its isogenic *aat* mutants were examined by western blotting, CexE appeared as two bands in the parental strain and each of the *aat* mutants except for *aatD* (Fig. 2B). The upper band had an apparent molecular mass of 12.7 kDa (uCexE) while the lower band (mCexE) had an apparent molecular mass of 11.8 kDa, which is consistent with the secreted form of CexE recovered from the supernatant fractions. As CexE is exported into the periplasm by the Sec pathway (Pilonieta et al., 2007), the inefficient cleavage of a signal peptide could be responsible for the difference in the molecular mass of uCexE compared to mCexE . To investigate this, ETEC H10407 pCfaD was grown in the presence of sodium azide. Sodium azide inhibits SecA to prevent translocation of proteins across the inner membrane and subsequent cleavage of the signal peptide (Huie & Silhavy, 1995; Oliver et al., 1990). In the absence of sodium azide the mCexE and uCexE forms could be detected in whole cell lysates as observed previously (Fig 2B and C). However, in the presence of sodium azide a protein (proCexE) with an apparent molecular mass of 13.2 kDa was detected (Fig. 2C). These observations suggest that the difference in molecular mass between uCexE and mCexE was not due to inefficient cleavage of the signal peptide but was due to a post translational event mediated by AatD.

A size change in the CexE homolog Aap has not been reported before despite numerous investigations of the Aat system. To determine whether the migration of Aap was altered by AatD, the effect of an *aatD* deletion on the apparent size of Aap was investigated in EAEC 042. EAEC 042, an *aap* mutant, an *aatD* mutant and an *aatD* mutant complemented with pJNW were grown in DMEM-HG; the pJNW plasmid encodes the complete *aat* operon from EAEC 042. Whole cell lysates of each strain were analysed by SDS-PAGE. As noted above for CexE, the deletion of *aatD* in EAEC 042 resulted in an apparent increase in size of Aap (Fig 2D). The size change of Aap in the *aatD* mutant could be restored to the wild-type size by complementation with the pJNW plasmid. From these data we conclude that AatD can modify Aap and CexE post-translationally and that this modification is required for secretion of the mature substrate molecule to the extracellular milieu.

Predicted functions of Aat components

To assist our understanding of the contribution of each Aat component to the secretion of CexE we constructed hidden Markov-models (HMM) for individual Aat proteins. These were used to search the Uniprot database for distant homologs. In agreement with previous publications, AatA is a homolog of the trimeric outer membrane protein TolC (Supplementary Fig S1) and AatB is a periplasmic adaptor protein (Supplementary Fig S2), components associated with T1SS and drug efflux pumps (Nishi et al., 2003). In contrast, AatD was not homologous to any T1SS or efflux-associated proteins. Instead, AatD was homologous to apolipoprotein N-acyltransferase (Lnt), which is a member of the carbon-nitrogen hydrolase (C-N hydrolase) family (PF00795) required for bacterial lipoprotein acylation and transport via the Lol system (Fig 3A). Comparison of the C-N hydrolase domain of AatD with other members of PF00795 revealed that AatD forms a clade with Lnt suggesting that they are functionally related. The ATPase domain of AatC is part of a clade containing the ATPase domains of LolD, the ATPase subunit of the LolCDE lipoprotein transporter (Fig 3B). Similarly, AatP is more closely related to LolC and LolE than typical ABC transporters of the T1SS, such as HlyB (Fig 3C). The homology of AatP, AatC, and AatD to proteins of the Lol pathway suggests that the Aat system is a composite system of a T1SS and the Lol lipoprotein trafficking system.

AatD-mediated acylation of mCexE

As AatD is a homolog of the Lnt acyltransferase, we hypothesised that mCexE represents an acylated form of CexE. To test this hypothesis, we used 17-ODYA, an 18-carbon alkyne fatty acid that can be conjugated to an azide linked fluorescent molecule by a copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC) reaction to investigate CexE lipidation. The *cexE* gene was introduced into a *cexE* mutant or a *cexE aatD* double mutant on a plasmid termed pACYC-*cexE*-6His, which encodes *cexE* under the transcriptional control of its native promoter. This allowed production of a C-terminally His-tagged variant of CexE. Both strains harbouring pACYC-*cexE*-6His were grown in the presence or absence of 17-ODYA. The bacterial lipoprotein YraP, was included as a positive control. CexE and YraP were isolated from each culture by virtue of their C-terminal 6xHis-tag by using nickel affinity chromatography. The purified proteins were then linked to an azide derivative of Alexa Fluor 488 using a CuAAC reaction. After the CuAAC reaction the CexE and YraP samples were separated by SDS-PAGE and fluorescence was measured (Fig 4). In the presence of AatD and 17-ODYA CexE was fluorescently labelled; in the absence of either 17-ODYA or AatD no fluorescence was observed. Moreover, only mCexE was labelled indicating that uCexE is an unacylated form of CexE; counter intuitively, the acylated form (mCexE) migrates faster on SDS-PAGE than the unacylated uCexE. As acylated proteins associate with membranes, we examined the cellular compartmentalisation of mCexE (Supplementary Fig S3). ETEC H10407 pCfaD and isogenic *aat* mutants were grown in LB supplemented with L-arabinose. Subsequently, cells were collected by centrifugation, lysed and the total membrane fraction was harvested by centrifugation. After separation by SDS-PAGE, only mCexE was detectable by western blotting in the membrane fraction of the parent strain and the *aatPABC* mutants; no mCexE could be detected in the membranes recovered from the *aatD* mutant (Supplementary Fig S3). In contrast, both mCexE and uCexE could be detected in whole cell lysates of ETEC H10407 pCfaD and only uCexE could be detected in whole cell lysates of the *aatD* mutant. These data suggest that AatD is an acyltransferase that mediates post-translational modification of CexE by the addition of one or more acyl chains.

To determine whether AatD was sufficient to mediate the acylation of mCexE, plasmids encoding AatD and CexE were introduced into the laboratory strain *E. coli* BL21(DE3), which does not encode any of the Aat proteins or Aap/CexE. *E. coli* BL21 was transformed with pET26b-*cexE* and

pACYC-*aatD* or their respective empty vector controls pET26b and pACYCDuet-1. These strains were grown in LB and the production of AatD and CexE was induced by IPTG. The production of CexE was monitored via western blotting (Fig 5). CexE was not detected in strains that contained pET26b. uCexE was produced in the absence of AatD at the size expected for the unmodified form after cleavage of the signal sequence. However, when CexE and AatD were produced together, mCexE migrated further (Fig 5). These data suggest AatD was solely responsible for the modification of CexE.

Catalytic residues of AatD

The C-N hydrolase family of proteins have a known conserved catalytic triad, which in the case of Lnt is E267, K335, and C387. Another Lnt residue E343 helps to stabilise the catalytic site (Vidal-Ingigliardi et al., 2007). Mutation of any of these residues to alanine results in a loss of Lnt function (Gélis-Jeanvoine et al., 2015). To assess the conservation of the catalytic residues the AatD sequence of EAEC 042 and ETEC H10407 and Lnt from *E. coli* MG1655 were aligned using Clustal Omega. Each of the residues associated with the catalytic site were conserved in both versions of AatD (Fig 6A). The C-N hydrolase catalytic triad is conserved in AatD: E217; K278; and C325. Also, Lnt E343 is conserved at position E286 in AatD. Further AatD sequences were identified using PSI-BLAST. These sequences were aligned and a WebLogo was produced for the regions flanking each of the catalytic residues; the four residues were 100% conserved in all AatD sequences identified. In addition, when a structure of AatD predicted by Phyre2 (Kelley et al., 2016) was compared to Lnt (PDB: 5N6L) (Wiktor et al., 2017) the four AatD residues (E217, K278, E286, and C325) were superimposed on the Lnt catalytic triad suggesting that these residues were required for acylation of CexE (Fig 6B and 6C). To test this hypothesis, pACYC-*aatD* derivatives encoding mutant derivatives of AatD with alanine substitutions at the conserved E217, K276, E286, or C325 were introduced into *E. coli* BL21(DE3) pET26b-*cexE*. Expression of AatD and CexE were induced by the addition of IPTG during growth in LB. As expected, mCexE was only present in the strain encoding the wild-type *aatD* sequence whereas only unmodified uCexE was detected in mutants harbouring the E217A, K276A, E286A, or C325A (Fig 6D). These observations suggest that the mechanism of protein acylation by AatD and Lnt are similar.

An N-terminal glycine is required for acylation

Due to the homology and functional similarity between AatD and Lnt, a likely site for acylation of CexE is the N-terminal amino acid immediately after the signal sequence. SignalP predictions suggested that CexE possesses a Sec-dependent signal sequence cleaved by Signal Peptidase I between alanine at position 19 and glycine at position 20 (Fig. 7A and Supplementary Table S2). N-terminal amino acid sequencing of uCexE revealed the amino acid sequence GGGNS confirming that glycine at position 20 formed the N-terminal amino acid of the processed protein. However, bioinformatic analyses of the amino acid sequences of 224 distinct CexE homologs failed to identify in their signal sequences the presence of a 'lipobox' (Supplementary Fig S4); lipoboxes are required for recognition of lipoproteins by the Lol system and thus acylation by Lnt (Babu et al., 2006). Further analyses revealed limited sequence identity between the signal sequences of the 224 CexE homologs indicating that an alternative lipobox is not present. Moreover, neither CexE nor any other homolog possessed a cysteine residue adjacent to the signal sequence; for bacterial lipoproteins the N-terminal cysteine of the mature protein is the target of Lnt mediated acylation. Since CexE does not contain features of a typical bacterial lipoprotein it must be recognised and acylated by AatD in a manner different to the acylation of lipoproteins by Lnt.

Further scrutiny of the amino acid sequences of all CexE homologs revealed that for all CexE and Aap proteins the first five amino acids immediately after the signal peptide cleavage site are a conserved mix of glycine and serine residues (Fig 7B). The presence of an invariant N-terminal glycine residue immediately downstream of the predicted signal sequence suggested that this might be the site of acylation. To test this hypothesis, we mutated each of the N-terminal residues of CexE to glutamic acid and observed the effect on acylation by western blotting. The amount of total fluorescence of the secondary antibody bound to primary polyclonal anti-CexE antibody was measured and compared for each mutant. The percentage of acylated mCexE was calculated for each mutant. The migration of the G20E mutant was similar to uCexE indicating that the change to glutamic acid completely abolished acylation (Fig 7C). In addition, the acylation of the G21E mutant was significantly reduced (p-value = 0.0013). However, there was no significant reduction in acylation of the G22E, N23E or S24E mutants

(Fig 7C). These data indicate that the acylation of CexE is highly dependent on the N-terminal glycine at position 20 and to a lesser extent on the second glycine at position 21.

Mass spectrometry analysis of CexE acylation

Mutation of the conserved N-terminal glycine can abolish acylation of CexE, and Lnt acylates the N-terminal amine of the signal peptidase processed lipoprotein, therefore we hypothesised that AatD would acylate the N-terminal amine of G20. To test this, we first purified his-tagged mCexE and uCexE by nickel affinity chromatography. Both forms of CexE were subjected to trypsin digestion followed by LC-MS/MS. A peak was observed in the HLPC trace of mCexE that was not present in uCexE (Fig 8A). The mass spectrometry of that peptide identified the five N-terminal residues of CexE with an addition of 238 Da, which is equal to that expected to a single 16 carbon fatty acid addition (Fig 8B) confirming the modification of CexE by the addition of an acyl chain onto the N-terminal glycine.

Heterologous acylation by AatD

Other domains of CexE might be responsible for interacting with AatD other than the N-terminus. Therefore, we fused the N-terminus of mCherry with the signal sequence of CexE and an increasing number of glycine residues to a maximum of three. These constructs were transformed into ETEC H10407Δ*cexE* pCfaD and the production of the mCherry fusions was induced in culture media supplemented with 17-ODYA. The mCherry constructs with N-terminal glycine residues were isolated using the incorporated C-terminal 8xHis tag. The purified proteins were subjected to CuAAC with azide-Cy5. The incorporation of 17-ODYA into the mCherry fusions was observed by detecting fluorescence of the Cy5 dye. The signal sequence alone did not cause an incorporation of 17-ODYA (Fig 9). However, a single glycine at the N-terminus was sufficient for mCherry acylation but the addition of further glycine residues to the N-terminus increased the fluorescence signal indicating an increase of AatD activity proportional with increase in the number of glycine residues present (Fig. 9). The minimum required for AatD-mediated acylation is a glycine at the N-terminus of the protein post signal sequence. However, an increase in the number of glycine residues appears to increase the efficiency of the reaction.

Discussion

Post-translational modification of proteins by the covalent attachment of fatty acids occurs for a myriad of proteins in eukaryotic and prokaryotic organisms. Such acylation events confer distinct biochemical properties on the proteins, enabling acylation to regulate intracellular trafficking, subcellular localization, and molecular interactions. The current dogma for Gram-negative bacterial lipoproteins purports that proteins are triacylated on an N-terminal cysteine through the action of three enzymes Lgt, Lsp and Lnt (Grabowicz, 2019; Nakayama et al., 2012; Zückert, 2014). To our knowledge there are no examples of bacterial proteins where acylation occurs on N-terminal glycine residues. While N-terminal glycine acylation by N-myristoyltransferase is present in life it is restricted to eukaryotes. N-myristoylation is one of the three major classes of fatty acylation reactions for eukaryotic proteins, which include: N-myristoylation of glycine/lysine in proteins such as c-Src or HIV Gag1 (Resh, 1994; Veronese et al., 1988); S-palmitoylation of cysteine in proteins such as Ras (Hancock et al., 1989); and N-palmitoylation of cysteine in proteins such as Hedgehog (Pepinsky et al., 1998). Examples of bacterial N-myristoylation rely on eukaryotic proteins produced in *E. coli* or the injection of myristoylatable proteins into eukaryotic cells by type 3 secretion systems (Duronio et al., 1990; Martin et al., 2011). To our knowledge this is the first report of enzyme mediated N-palmitoylation in nature. The recent publication by Belmont-Monroy *et al.* 2020 independently identified AatD as an acyltransferase of Aap and CexE.

The Aat system is, to our knowledge, the only secretion system that acylates and secretes a lipoprotein substrate to the cell surface. While surface exposed lipoproteins have been reported for *E.* e.g. Lpp, RcsF, and SslE (Baldi et al., 2012; Cowles et al., 2012; Konovalova et al., 2014), these lipoproteins are acylated by Lgt and Lnt on an N-terminal cysteine in the typical manner and translocated to the outer membrane by the Lol pathway. Similarly, *Neisseria* spp. decorate their cell surfaces with secreted lipoproteins that are acylated by Lgt and Lnt on an N-terminal cysteine and trafficked to the outer membrane by the Lol system. In the latter case an outer membrane protein termed Slam is required for translocation of the protein to the cell surface (Hooda et al., 2016). In contrast, the Aat system encodes a novel and specialised N-acyltransferase, AatD. To allow secretion of the novel lipoprotein CexE the Aat system has combined proteins of the T1SS and the Lol trafficking pathway.

Not only does the Aat system secrete a novel class of lipoproteins from the periplasm to the surface of the cell but also it acylates that protein itself. For this reason, we believe that reclassification of the Aat system as the first example of the Lipoprotein Secretion System (LSS) is warranted. We propose that the LSS would remain as a sub-class of the T1SS to emphasise its similarities to the T1SS.

In Fig 10, we propose our model for Aat mediated secretion. First, the Sec machinery secretes preCexE into the periplasm, resulting in cleavage of the signal peptide by Signal Peptidase I to form uCexE (Pilonieta et al., 2007). Once in the periplasm the N-terminal glycine of uCexE is acylated by AatD to create mCexE. Through this acylation event mCexE then associates with the inner membrane. We believe this to be first step for secretion by the Aat system since CexE acylation by AatD is independent of the other members of the Aat system. In addition, the deletion of any of the *aatPABC* genes does not result in a loss of CexE acylation. We propose that the acylation of CexE is required to enter the AatPABC tunnel. Due to the homology to LolCDE, we believe that mCexE is extracted from the inner membrane by AatPC. Finally, AatA, AatB and AatPC form a channel for mCexE secretion in a manner analogous to a typical T1SS. Subsequently, mCexE inserts into the outer leaflet of the outer membrane.

The Aat system is not unique in secreting a periplasmic substrate. Other T1SS employ a two-step secretion mechanism. One such example is the MacAB-TolC complex which secretes heat-stable enterotoxin from the periplasm (Yamanaka et al., 2008). We identified MacA and MacB as homologs of AatB and AatPC, respectively. Suggesting that AatPC may use a similar mechanism of mechanotransmission as MacB to secrete Aap/CexE (Crow et al., 2017). Furthermore AatB may form a similar gating ring to MacA (Fitzpatrick et al., 2017). However, how the Aat system is able to combine the mechanism of the MacAB-TolC T1SS and the LolCDE lipoprotein trafficking pathway has yet to be elucidated.

Belmont-Monroy *et al.* 2020 suggested that Lnt is able to acylate Aap. However, in our hands there was no evidence of CexE acylation in the absence of AatD. Given that Aap and CexE lack the traditional lipobox of bacterial lipoproteins that would be recognised by the Lol system, that the signal sequence of all Aap and CexE peptides were predicted to be cleaved by Signal Peptidase I, which has been confirmed experimentally for CexE (Pilonieta et al., 2007), and we were able to reconstruct AatD

acylation in *E. coli* BL21(DE3) which does not naturally encode AatD or CexE but does encode Lnt, we do not believe that Lnt plays a role in CexE acylation. Indeed, our demonstration of acylation of a heterologous protein with a single glycine residue (Fig 10) suggests that AatD is a specific N-acyltransferase that has the potential for exploitation for production of novel acylated peptides in *E. coli*. The addition of an acyl chain can improve the half-life of peptide drugs like insulin (Kurtzhals, 2007). Since we have shown that it is possible to reconstitute AatD acylation in a laboratory strain of *E. coli* there are further biotechnological applications to be explored.

Due to increasing levels of antibiotic resistance in pathogenic organisms there is a requirement for new protein targets. The conservation of this system within pathogenic bacterial strains represents a possible new drug target. Previous studies have shown that deletion of *aatD*, *aatC*, or *aap*/*cexE* reduces bacterial colonisation and disease (Belmont-Monroy et al., 2020; Rivas et al., 2020). Thus, chemicals targeting the Aat system have the potential to prevent or perhaps reduce the severity of bacterial disease. As the Aat system does not appear to be present in non-pathogenic strains this would be an ideal target to prevent off-target effects. As only pathogenic bacteria would be affected by the inhibition of the Aat system, there would be no risk of evolving resistant alleles in non-pathogenic populations.

In conclusion, we have identified a novel N-acyltransferase. In addition, we have characterised a novel class of bacterial lipoproteins that are acylated and secreted by a composite secretion system. There are still significant questions left to be answered on the mechanism of Aat secretion as well as the function of the secreted proteins. In addition, the ability to acylate heterologous proteins with a simple glycine addition is a valuable new tool for the biotechnology enabling new acylated peptides to be produced in *E. coli*.

.

Materials and methods

Bacterial cultivation

Supplementary Table S3 contains the bacterial strains used in this study. Bacterial strains were regularly cultivated in lysogeny broth (LB) which consisted of 10 g/L tryptone, 5 g/L yeast extract and 10 g/L NaCl. Bacterial strains were routinely grown on LB supplemented with 1.25% agar (LBA), and a single colony was used to inoculate liquid cultures for overnight growth at 37°C with aeration. Overnight cultures were used to inoculate Erlenmeyer flasks containing a fifth of their total volume of LB supplemented with the relevant antibiotics, and grown at 37°C with aeration. Antibiotics were used at the following final concentrations: kanamycin 100 µg/ml; carbenicillin 100 µg/ml; and chloramphenicol 35 µg/ml. For the induction of CexE in ETEC H10407, cultures harbouring pCfaD were grown for 90 minutes in LB, supplemented with carbenicillin, at 37°C with aeration. Expression of *cfaD* was induced with 0.2% L-arabinose for 2 hours at 37°C with aeration. Aap was induced by growing EAEC 042 in DMEM-HG with aeration at 37°C.

Molecular techniques

Primers for PCR can be found in Table S4. Plasmids used in this study are detailed in Table S5. Unless otherwise stated all primers were used at a concertation of 10 µM. Q5® High-Fidelity 2X Master Mix (NEB) was used for cloning or mutagenesis. MyTaq™ Red Mix (Bioline) was used in all other cases. The *cexE* gene from ETEC H10407 was amplified using CcexE-F and CcexE-R primers. Both the vector, pET26b(+) (Novagen), and the insert were cut with NdeI (NEB) and XhoI (NEB) using the CutSmart® (NEB) protocol. The digested vector was treated with Antarctic Phosphatase (NEB). T4 DNA Ligase (NEB) was used to ligate the insert DNA to the vector DNA. The ligation mixture was transformed into NEB® 5-alpha Competent E. coli (High Efficiency). Recovered cells were plated on LBA supplemented with 100 µg/ml kanamycin and incubated at 37ºC overnight. The *cexE* gene and 573 bp upstream of the *cexE* gene, which included the two CfaD binding sites (Pilonieta et al., 2007), was amplified by PCR. The reverse primer included a 6His tag. The resulting amplified DNA was digested with SphI and BamHI and ligated into pACYC184. Successful ligations were selected for as

373 described above on LBA supplemented with 35 µg/ml chloramphenicol. Point mutations were
374 constructed using the QuickChange II method (Aglient). The *aat* genes and *cexE* were disrupted in
375 ETEC H10407 as previously described (Datsenko & Wanner, 2000). The plasmid pDOC-K was used
376 as the source of the kanamycin resistance cassette (Lee et al., 2009). The *cexE*-mCherry fusion and the
377 *aatD* gene from ETEC H10407 were synthesised by GenScript. Deletions of the *cexE* gene from the
378 plasmid encoding the CexE-mCherry fusion were constructed as previously described (Moore &
379 Prevelige, 2002). Plasmid sequences were confirmed by Sanger sequencing.

380 **Bioinformatic analysis of the Aat system**

381 The protein sequences of the Aat proteins from ETEC H10407 were used to search the NCBI non-
382 redundant protein sequences database using PSI-BLAST (Altschul et al., 1997). Strains encoding a
383 complete Aat system on the same nucleotide accession were used to assess the distance between *aat*
384 genes. Distant homologs of the Aat proteins were identified using HMMER (Finn et al., 2011). An
385 HMM was created for each Aat protein using the Aat protein sequences identified by PSI-BLAST.
386 These models were used to search the UniprotKB or Swissprot databases (The UniProt Consortium,
387 2019). Protein sequences were aligned using Clustal Omega (Madeira et al., 2019). RAxML was used
388 for the construction of phylogenetic trees (Stamatakis, 2014). Trees were drawn using iTOL (Letunic
389 & Bork, 2019).

390 **CexE purification for antibody production**

391 BL21(DE3) was transformed with pET26b-*cexE*. An overnight culture of BL21(DE3) pET26b-*cexE*
392 was used to inoculate 2 L of LB supplemented with 100 µg/ml kanamycin. The culture was grown at
393 37°C with aeration to an OD_{600} of 0.4. The culture was moved to 20°C incubator for 30 min prior to
394 induction. CexE production was induced with 50 µM Isopropyl β-D-1-thiogalactopyranoside (IPTG)
395 (Sigma) and incubated overnight at 20°C with aeration. After overnight growth, cells were harvested
396 by centrifugation at 6,000 x *g* for 10 min at 4ºC. The cell pellet was resuspended in ice-cold binding
397 buffer (50 mM NaP (77:33 ratio of Na_2HPO_4 to NaH_2PO_4), 500 mM NaCl, 20 mM imidazole, 0.5 mM
398 TCEP at pH 7.3). Cells were lysed using EmulsiFlex®-C3 (Avestin). Cellular debris and intact cells

were removed by centrifugation at 10,000 x *g* for 10 min at 4ºC. The cellular membrane components were removed by centrifugation at 75,000 x *g* for 1 h at 10ºC. The supernatant was applied to a HisTrap™ HP 5 ml column (GE Healthcare Life Sciences) overnight at 4°C. The column was washed with 5 column volumes of binding buffer then the protein was eluted in 5 ml fractions with elution buffer (50 mM NaP, 500 mM NaCl, 500 mM imidazole, 0.5 mM TCEP at pH 7.3). Samples containing purified protein were concentrated using Vivaspin® 6 5,000 MWCO columns (Sartorius Stedim). Protein was buffer exchanged using membrane filtration into 50 mM NaP, 250 mM NaCl and 0.5 mM TCEP at pH 7.3. Purified CexE protein was used to produce primary antibodies by Eurogentec using the 28-day speedy protocol. Polyclonal antibodies against CexE were concentrated prior to use for the detection of CexE.

Tris-Tricine SDS-PAGE and western blotting

Protein samples were separated using tris-tricine SDS-PAGE as previously described (Schägger, 2006). In brief, protein samples were mixed with sample buffer (4% SDS, 2.5% 2-mercaptoethanol, 7.5% glycerol, 0.01% bromophenol blue, 35 mM Tris (pH 7.0)) and separated on 10% tris-tricine SDS-PAGE gels. Gels were stained with Coomassie brilliant blue or transferred to nitrocellulose for western blotting. Nitrocellulose membranes were covered in 2% BSA (20 g BSA, 2.42 g Tris-base (pH 8.4), 8 g NaCl per L). The primary antibody was diluted in 2% BSA at the following concentrations: Aap 1 in 5,000 (Sheikh et al., 2002); CexE 1 in 2,000 (this study); and β subunit of RNAP 1 in 10,000 (*E. coli* RNA Polymerase beta Monoclonal Antibody, BioLegend). Primary antibodies were incubated overnight at 4°C with agitation. Membranes were washed three times in TBST (2.42 g Tris-base, 8 g NaCl, 0.1% Tween-20, pH 8.4 in 1 L) for 5 min at room temperature with agitation prior to incubation with secondary antibody. Anti-Rabbit (IRDye 800CW Goat anti-Rabbit IgG, Li-Cor) or anti-mouse (IRDye 680LT Goat anti-Mouse IgG, Li-Cor) secondary antibodies were used to detect primary antibody binding. Secondary antibodies were diluted 1 in 15,000 in 2% BSA and incubated for a minimum of 1 h at room temperature with agitation. Membranes were washed 4 times with TBST for 5 min at room temperature with agitation. Secondary antibody fluorescence was detected using the Odessy CLx imaging system.

Inhibition of SecA by sodium azide

Two cultures of ETEC H10407 pCfaD were inoculated from the same overnight culture. The cells were both grown at 37°C for 90 min with aeration. WCL samples of each culture were taken. L-arabinose was added to both cultures to a final concentration of 0.2%. Sodium azide was added to a final concentration of 3 mM to one of the cultures. Both cultures were grown for 2 h at 37°C with aeration. WCL samples of each culture were taken and separated by tris-tricine SDS-PAGE. CexE was detected by western blotting.

Proteomic analysis

Cultures were grown under the conditions previously described for CexE production in ETEC H10407 pCfaD. Cells were separated from the culture supernatant by centrifugation at 8,000 x *g* for 10 min at 4ºC. The culture supernatant was filtered using Millex-GP Syringe Filter Unit, 0.22 µm, polyethersulfone (Merck) then cooled on ice. Ice-cold 100% TCA was added to a final concentration of 20%. Samples were incubated on ice for 30 minutes. Proteins were pelleted by centrifugation. The supernatant was removed and discarded from the sample. The pellet was washed twice with 1 ml ice-cold 100% methanol, proteins were collected with centrifugation at 21,000 x *g* for 15 min at 4ºC between each wash. The supernatant was removed and residual methanol was evaporated by incubating the sample at 60ºC for 10 min. The pellet was resuspended in 50 µl tris-tricine sample buffer. If a colour change to yellow occurred saturated tris-base was added until the original colour returned. Samples were analysed by tris-tricine SDS-PAGE.

Membranes were extracted from 50 ml of cells grown under the conditions required for CexE production. Cells were collected by centrifugation and resuspended in 20 ml of 10 mM Tris (pH 8.0), 1 mM PMSF (Phenylmethylsulfonyl fluoride). Cells were lysed using EmulsiFlex®-C3 (Avestin). Unlysed cells were collected by centrifugation at 5,000 x g for 10 minutes at 4°C. The supernatant was separated. The membranes were collected by centrifugation at 50,000 x g for 60 minutes at 4°C. The resulting membrane pellet was washed twice with ice-cold 10 mM Tris (pH 8.0). The final pellet was resuspended in 100 µl 10 mM Tris (pH 8.0). The samples were normalised by protein concentration. The membrane samples were separated by tris-tricine SDS-PAGE.

CuAAC

Overnight cultures harbouring plasmids encoding his tagged proteins of interest to be labelled were used to inoculate 25 ml LB supplemented with the relevant antibiotics to an OD_{600} of 0.05. Cultures were grown for 1.5 h at 37°C. 17-ODYA dissolved in DMSO was added to a final concentration of 20 µM and protein production induced. An equal volume of DMSO was added as a negative control. Cells were collected by centrifugation and resuspended in 600 µl of 50 mM NaP (pH 7.0), 150 mM NaCl, 0.1% Triton X-100, and 1 mM PMSF. Cells were lysed by sonication for 15 min using the bioruptor on a 30 s on, 30 s off cycles. Insoluble material was removed by centrifugation at and the supernatant was retained. His tagged proteins were isolated from the supernatant using Dynabeads™ His-Tag Isolation and Pulldown (Invitrogen) as per the manufacturer's instructions. Imidazole was removed by dialysis. Protein sample concentrations were measured with Pierce™ BCA Protein Assay Kit (Thermo Scientific™). A master mix was prepared of 0.2 mM Cy5-azide (Sigma-Aldrich) or Alexa Fluor™ 488 azide (Invitrogen), 0.2 mM tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA), and 2 mM freshly prepared $CuSO_4$. This was made to 50 µl in H_2O and 5 µl of the master mix was added to 40 µl each normalised protein sample. Sodium ascorbate was added to a final concentration of 5 mM. Samples were incubated at 37ºC for 1 h with agitation. Protein was precipitated using chloroform-methanol extraction. To each sample 400 µl MeOH, 150 µl chloroform, and 300 µl H_2O were added and vortexed for 30 seconds. Samples were centrifuged at max speed for 2 min. The top layer was removed and 400 µl MeOH was added and samples were centrifuged again at max speed for 2 min. The supernatant was removed, and the pellet was washed twice with 400 µl MeOH then dried. Pellets were resuspended in 20 µl 1x tris-tricine sample buffer and separated by tris-tricine SDS-PAGE. Gels were incubated in fixative (10% acetic acid, 50% methanol) prior to fluorescence detection then stained with Coomassie Brilliant Blue.

Heterologous protein acylation

ETEC H10407 pCfaD was transformed with pRSF-GGG-mCherry, pRSF-GG-mCherry, pRSF-G-mCherry or pRSF-SS-mCherry. Cultures were grown as previously described for *cfaD* induction. At the same time as *cfaD* was induced 17-ODYA was added to a final concentration of 20 µM. Cells were

grown for 2 hours and harvested by centrifugation. 17-ODYA incorporation was detected as described above.

Mass spectrometry analysis of CexE acylation

CexE with a 6 His tag was isolated from a cexE mutant and a cexE aatD double mutant of ETEC H10407 pCfaD. The isolated CexE proteins were separated on a tris-tricine gel and stain with Coomassie brilliant blue. The band corresponding to uCexE and mCexE were excised and subjected to trypsin-LysC digestion. The tryptic peptide extracts were analyzed by nanoHPLC/MS MS/MS on an Eksigent, Ekspert nano LC400 uHPLC (SCIEX, Canada) coupled to a Triple TOF 6600 mass spectrometer (SCIEX, Canada) equipped with a PicoView nanoflow (New Objective, USA) ion source. 5 µl of each extract was injected onto a 5 mm x 300 µm, C18 3 µm trap column (SGE, Australia) for 5 min at 10uL/minute. The trapped tryptic peptide extracts were then washed onto the analytical 75 µm x 150 mm ChromXP C18 CL 3 µm column (SCIEX, Canada) at 400 nL/min and a column temperature of 45°C. Solvent A consisted of 0.1% formic acid in water and solvent B contained 0.1% formic acid in acetonitrile. Linear gradients of 2-40% solvent B over 60 min at 400 nL/minute flow rate, followed by a steeper gradient from 40% to 90% solvent B in 5 min, then 90% solvent B for 5 min, were used for peptide elution. The gradient was then returned to 2% solvent B for equilibration prior to the next sample injection. The ionspray voltage was set to 2600 V, declustering potential (DP) 80 V, curtain gas flow 30, nebuliser gas 1 (GS1) 30, interface heater at 150°C. The mass spectrometer acquired 50 ms full scan TOF-MS data followed by up to 30 100 ms full scan product ion data, with a rolling collision energy, in an Information Dependant Acquisition, IDA, mode for protein identification and peptide library production. Full scan TOF-MS data was acquired over the mass range 350-1800 and for product ion MS/MS 100-1500. Ions observed in the TOF-MS scan exceeding a threshold of 200 counts and a charge state of +2 to +5 were set to trigger the acquisition of product ion, MS/MS spectra of the resultant 30 most intense ions. The data was acquired and processed using Analyst TF 1.7 software (ABSCIEX, Canada). Protein identification was carried out using Protein Pilot 5.0 for database searching.

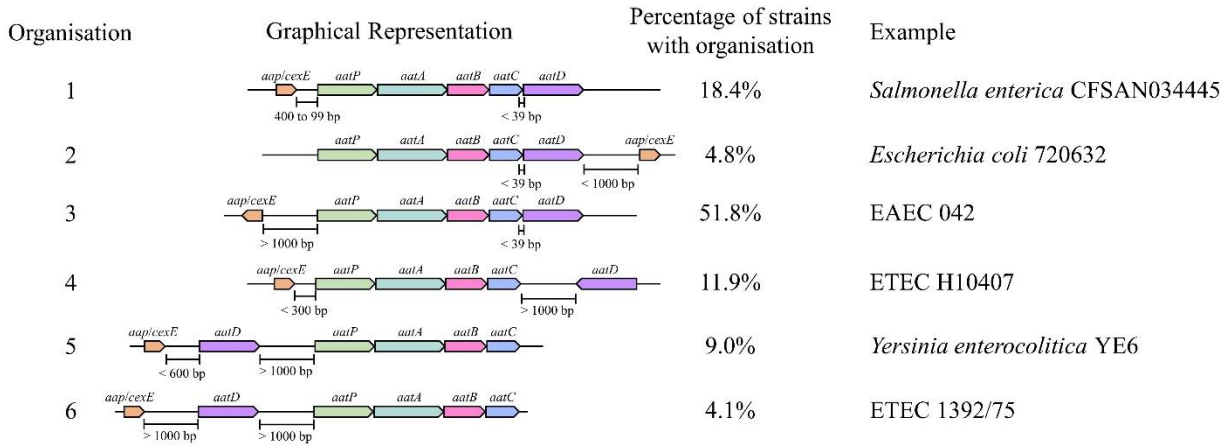


Fig 1. Organisation of the Aat operon. Unique Aat amino acid sequences were detected using PSI-BLAST. These sequences were using to identify the strains that encoded them in the NCBI identical protein groups. As the *aat* genes were present on contigs of whole genome sequencing projects it was not possible to assess if a strain encoded a gene on the chromosome or plasmid. Instead contigs were used to identify complete the complete Aat system. This analysis does not include strains that might encode the *aat* genes or *aap*/*cexE* on a separate genomic element. However, a total of 827 complete Aat systems were identified in the same nucleotide accession. The positions of these genes were used to assess the organisation of the *aat* operon. From this assessment five different classes of *aat* operon organisation were defined.

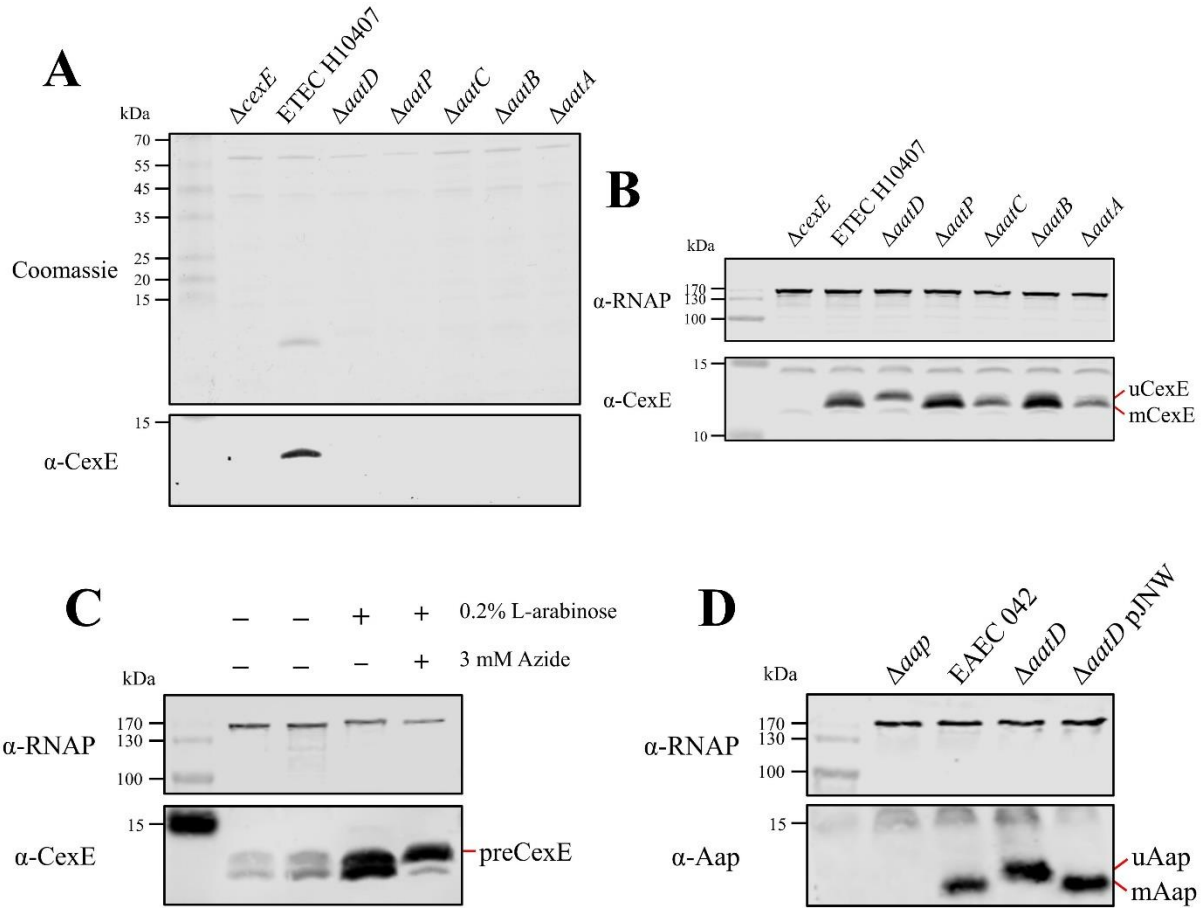
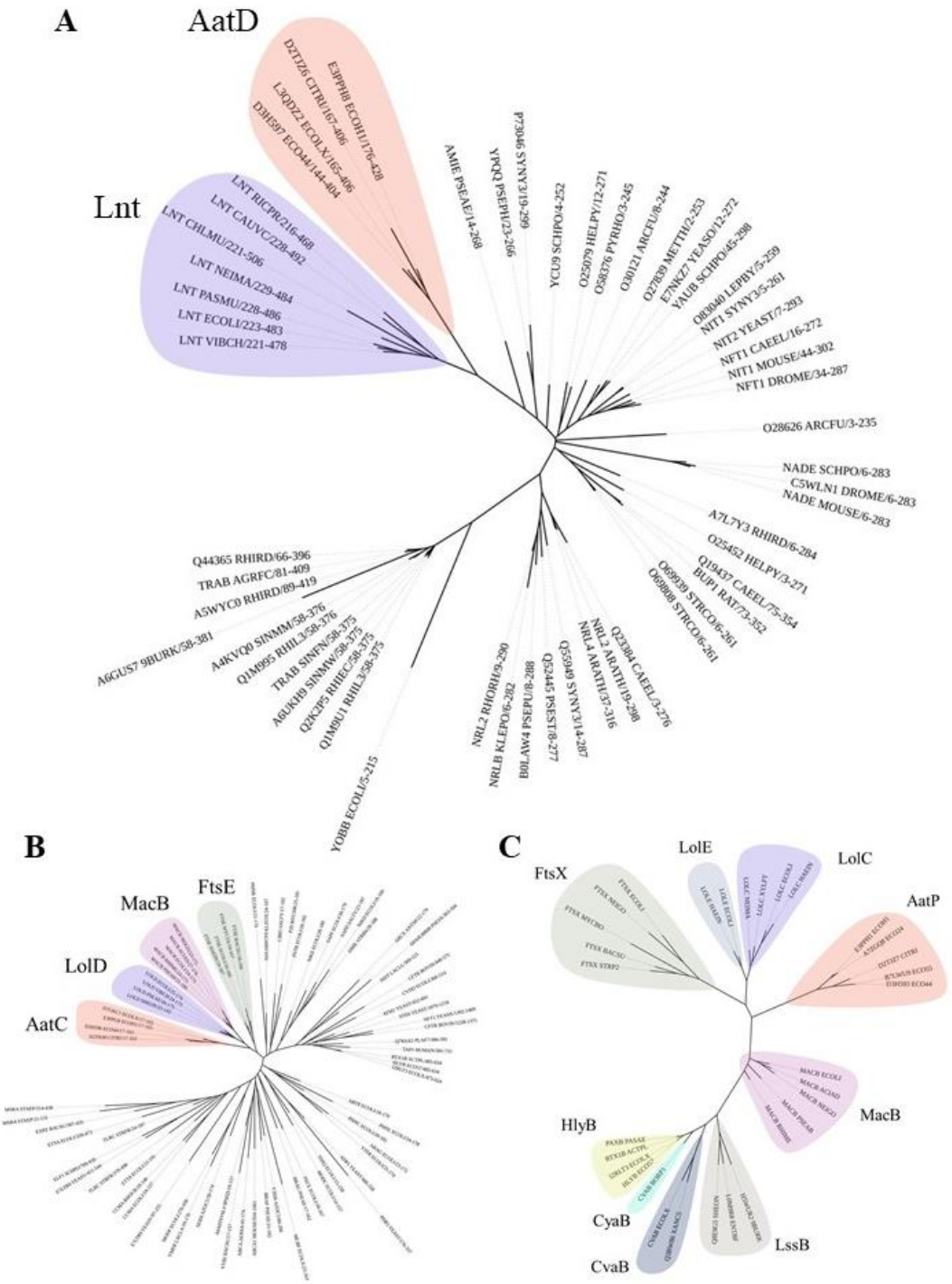


Fig 2. AatD post-translational modification of Aap and CexE. (A) Culture supernatant of ETEC H10407 and *aat* mutants harbouring pCfaD grown in LB supplemented with L-arabinose. Cells were removed from the culture supernatant and remaining protein precipitated with trichloroacetic acid. Protein samples were separated by tris-tricine SDS-PAGE and stained with Coomassie or transferred to nitrocellulose for western blotting with polyclonal antibodies against CexE. **(B)** Whole cell lysates of the *aat* mutants and parental strain separated by tris-tricine SDS-PAGE. **(C)** Whole cell lysates of ETEC H10407 grown in LB with or without azide. **(D)** Whole cell lysates of EAEC 042, *aap*, *aatD* and *aatD* complemented strains grown in DMEM high glucose. The positions of uCexE, mCexE, uAap, mAap, and preCexE are indicated as appropriate. CexE and Aap were detected by western blotting (α-CexE and α-Aap, respectively) and RNA polymerase (α-RNAP) was included as a loading control where appropriate.



529

530 **Fig 3. Phylogenetic analysis of AatD, AatP and AatC.** Sequences were aligned using Clustal omega
531 and the tree was calculated using RAxML. **(A)** The C-N hydrolase domains of C-N hydrolase family
532 seed sequences of the were aligned to the C-N hydrolase domains of AatD sequences from ETEC
533 H10407, EAEC 042, *C. rodentium* ICC168, and *E. coli* KTE75. **(B)** Phylogram of the ATPase domains
534 of the pfam ABC transporters (PF00005) and AatC sequences. (C) Phylogram of T1SS ABC
535 transporters, Lol ABC transporters and AatP sequences.

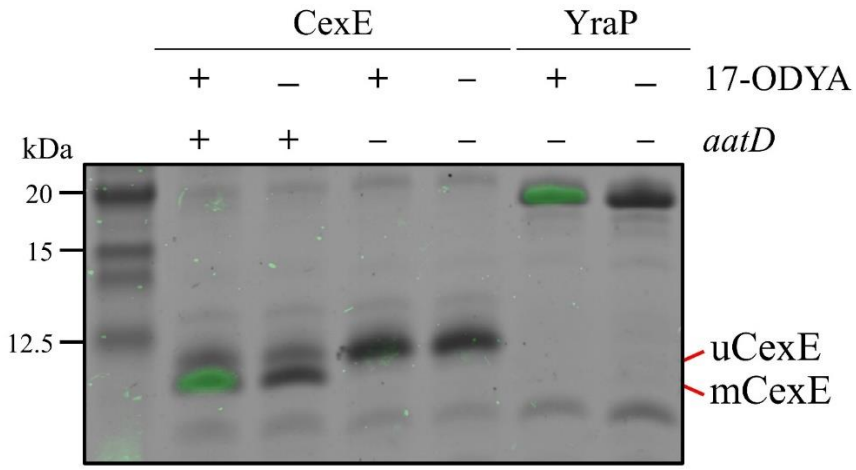


Fig 4. Incorporation of 17-ODYA into CexE in the presence of AatD. His tagged CexE was isolated from *cexE* or *cexE aatD* double mutants harbouring pACYC-*cexE*-6His grown in the presence or absence of 17-ODYA and separated by SDS-PAGE. An azide linked Alexa Fluor 488 was conjugated to the alkyne moiety present in 17-ODYA by CuAAC. The incorporation of 17-ODYA into the target protein was detected by fluorescence (green bands) and the image was overlaid on the image of the SDS-PAGE gel. A known lipoprotein YraP was used as a positive control.

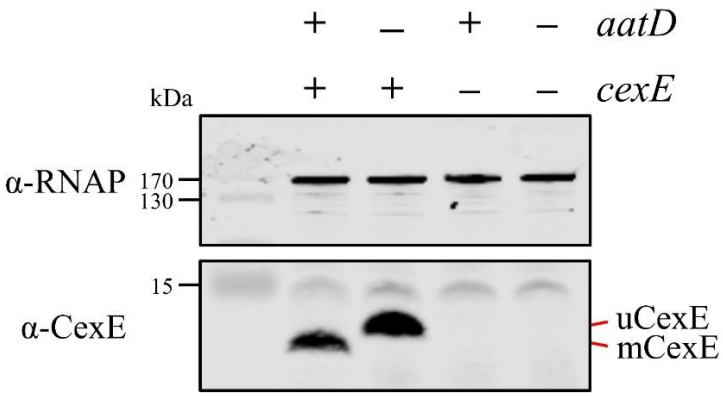
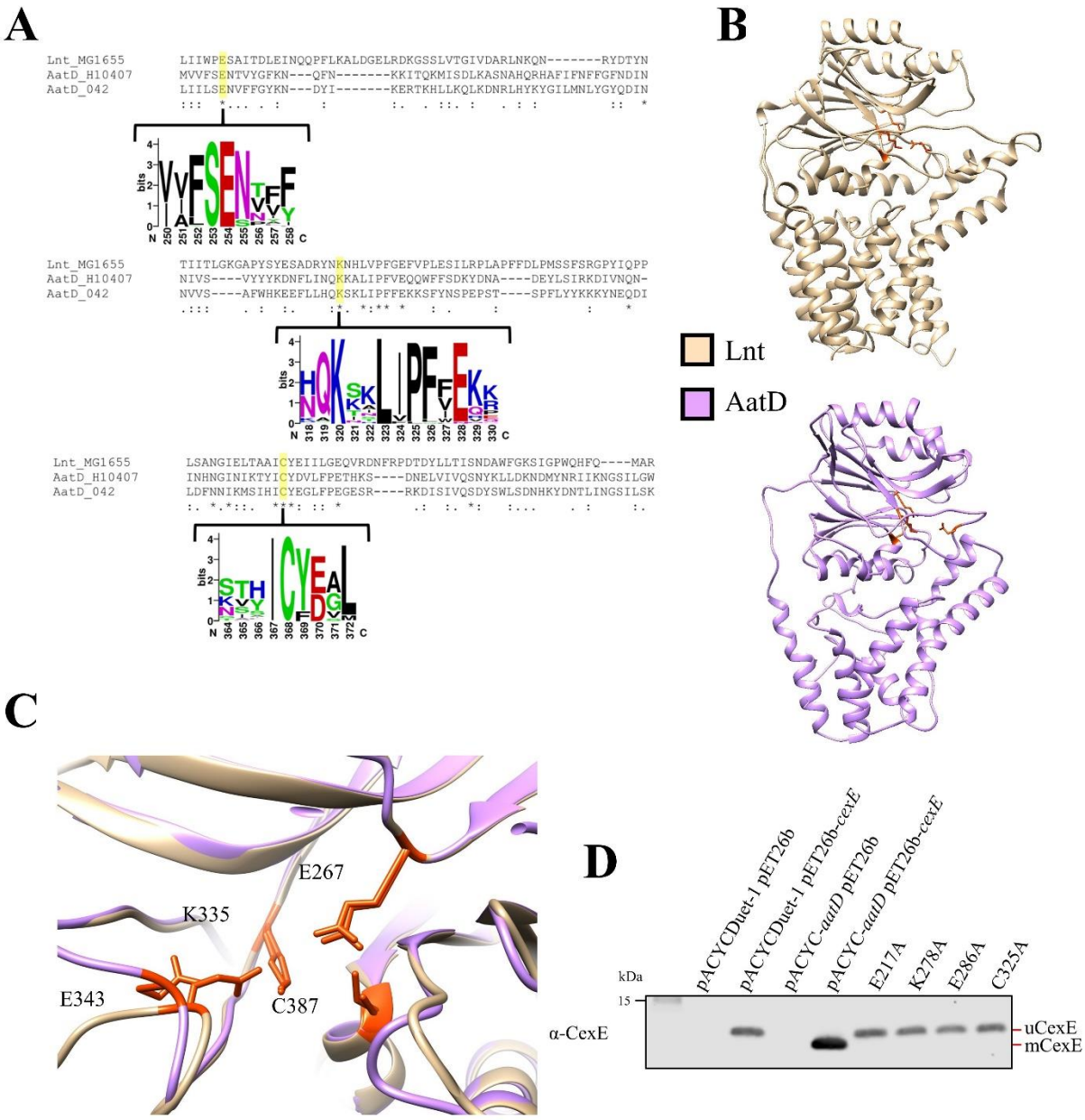


Fig 5. Recapitulation of CexE acylation in *E. coli* BL21(DE3). *E. coli* BL21(DE3) was transformed with pET26b, or pET26b-*cexE* and pACYCDuet-1 or pACYC-*aatD*. Cultures were grown in LB and protein production was stimulated with IPTG. Whole cell lysate samples were taken and separated by tris-tricine SDS-PAGE. CexE was detected using a polyclonal antibody. RNAP was used as a loading control. The position of mCexE and uCexE are indicated.



551

552 **Fig 6. Catalytic residues of AatD. (A)** AatD from EAEC 042 and ETEC H10407 were aligned with
553 Lnt from MG1655. AatD sequences identified by PSI-BLAST were aligned and used to create a
554 Weblogo. The catalytic residues of the C-N hydrolase family are highlighted. **(B)** The structure of Lnt
555 compared to the predicted structure of ETEC H10407 AatD. The catalytic residues of Lnt and AatD are
556 highlighted in orange. **(C)** Magnified view of the catalytic site of Lnt with the predicted structure of
557 AatD superimposed. **(D)** Residues are numbered as they appear in Lnt. *E. coli* BL21(DE3) was
558 producing CexE in the presence of each of the four single site substitution mutants of AatD. CexE was
559 detected using a polyclonal antibody.

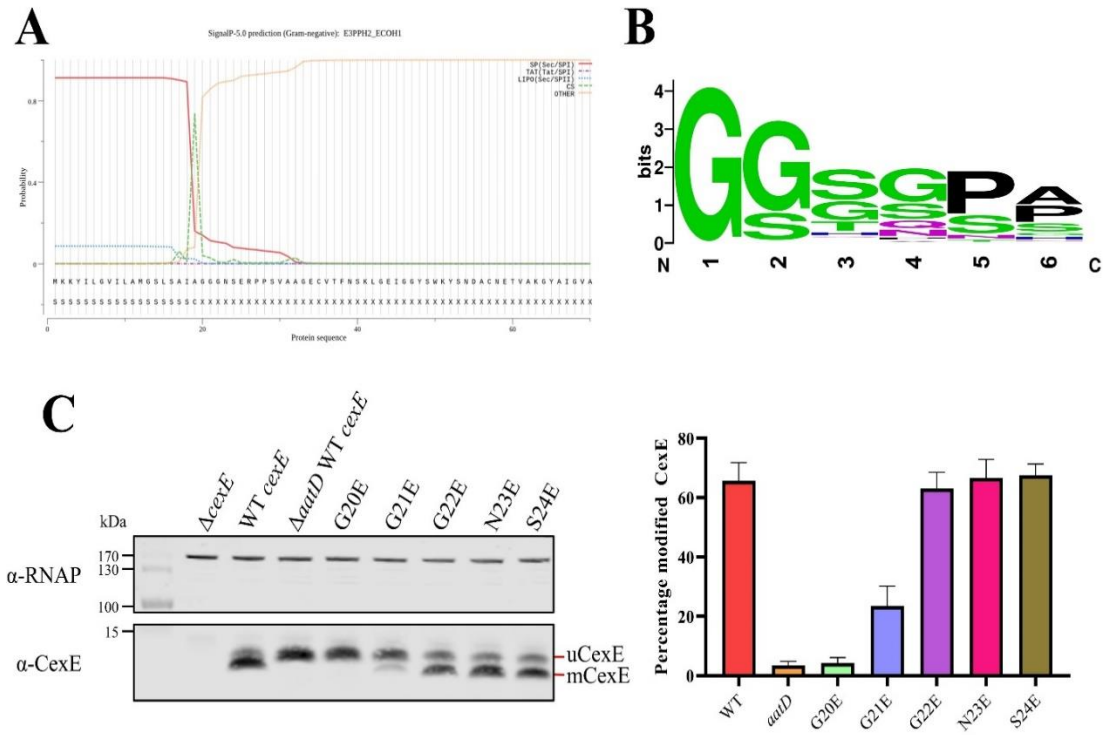
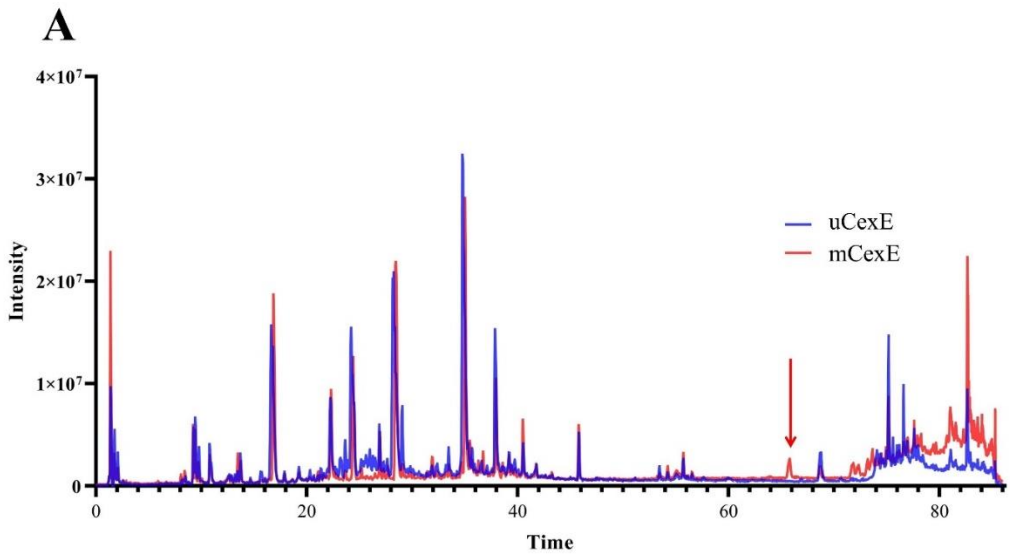
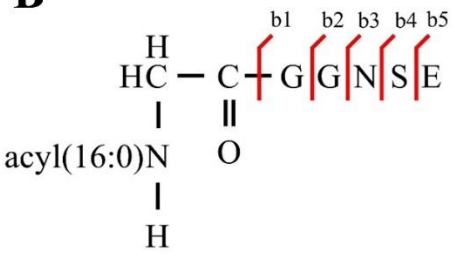
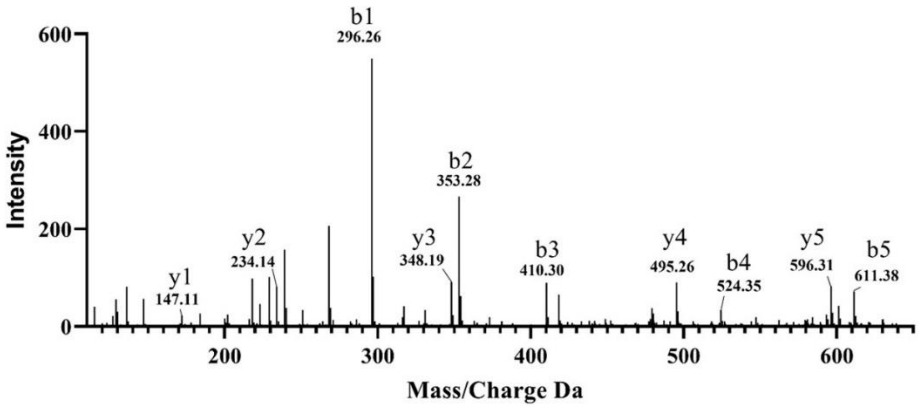


Fig 7. Single site substitution of the 5 N-terminal amino acids of CexE. (A) SignalP result of CexE sequence from ETEC H10407 (E3PPH2_ECOH1) **(B)** Weblogo of the 5 N-terminal residues of Aap and CexE sequences post Sec signal sequence cleavage. **(C)** ETEC H10407 *cexE* mutants transformed with pACYC184 (Δ*cexE*) or pACYC-*cexE*-6His with either the wild-type sequence (WT *cexE*) or one of the first five amino acids mutated to glutamic acid. CexE was detected by polyclonal antibodies and RNAP was used as a loading control. The percentage of mCexE was determined from 3 biological replicates.





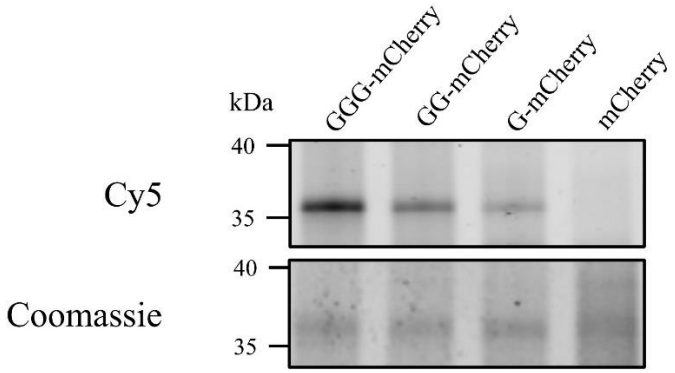
Position after signal sequence	Residue	Theoretical weight (Da) b	y
1	G, Gly	296.26 (b1)	
2	G, Gly	353.28 (b2)	
3	G, Gly	410.30 (b3)	
4	N, Asn	524.34 (b4)	
5	S, Ser	611.38 (b5)	
	//		
18	T, Thr		596.30 (y5)
19	F, Phe		495.26 (y4)
20	N, Asn		348.19 (y3)
21	S, Ser		234.14 (y2)
22	K, Lys		147.11 (y1)



568

569 **Fig 8. Mass spectrometry analysis of modified CexE. (A)** CexE-6His was isolated from a *cexE* mutant
570 and a *cexE aatD* double mutant and subjected to LC-MS/MS. CexE and pro-CexE were trypsinated and
571 separated by HPLC. **(B)** The indicated peak was subjected to MS/MS to identify the amino acid
572 sequence.

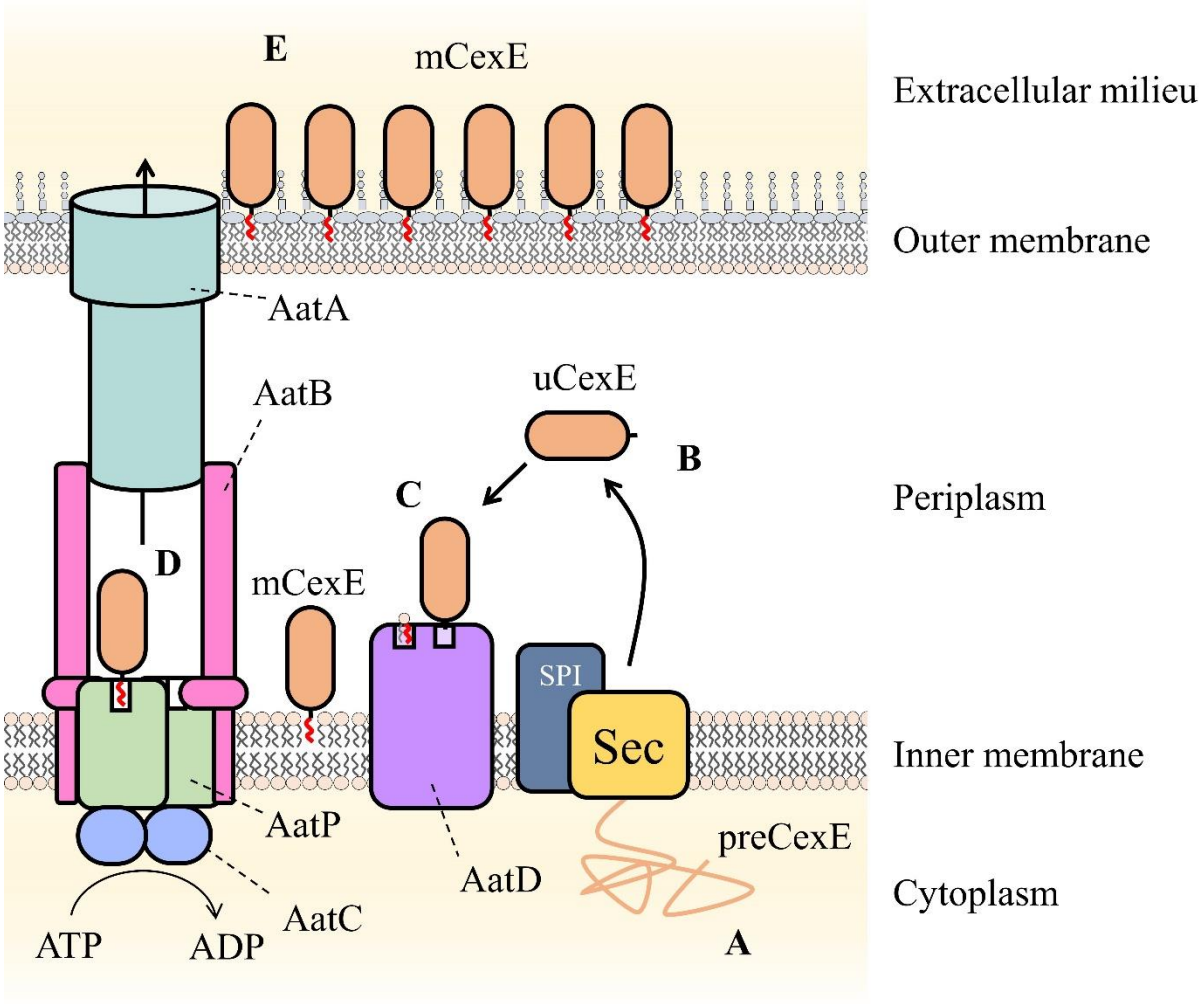
573



574

575 **Fig 9. Heterologous acylation by AatD** plasmid encoding mCherry with the CexE signal sequence
576 followed by none, one (G), two (GG) or three (GGG) glycine residues post signal sequence was
577 produced in an ETEC H10407 *cexE* mutant grown in the presence of 17-ODYA. The mCherry proteins
578 were isolated using a C-terminal 8 His tag and azide linked Cy5 was incorporated into mCherry proteins
579 using CuAAC. The acylation of mCherry was detected using fluorescence.

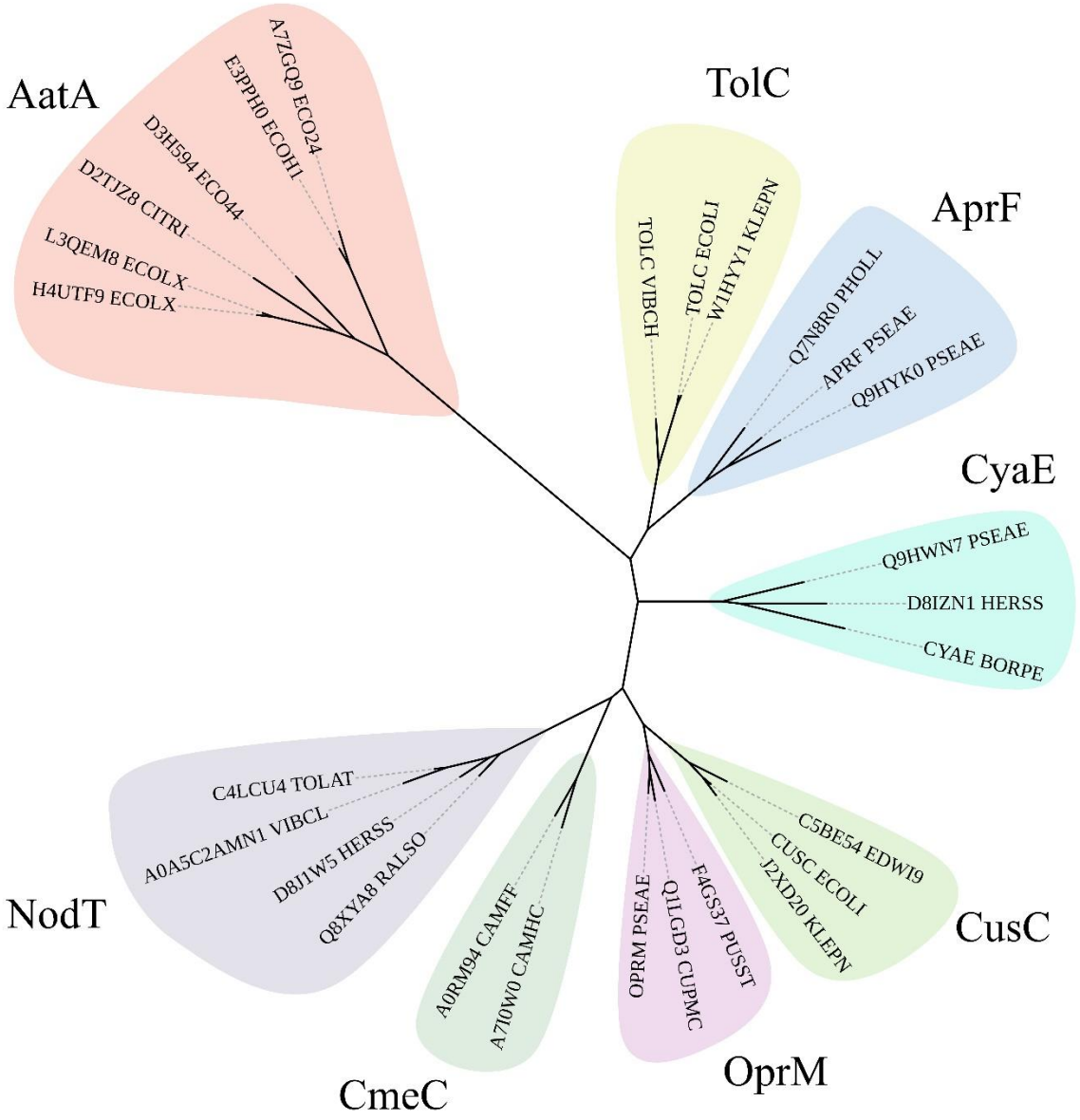
580



581

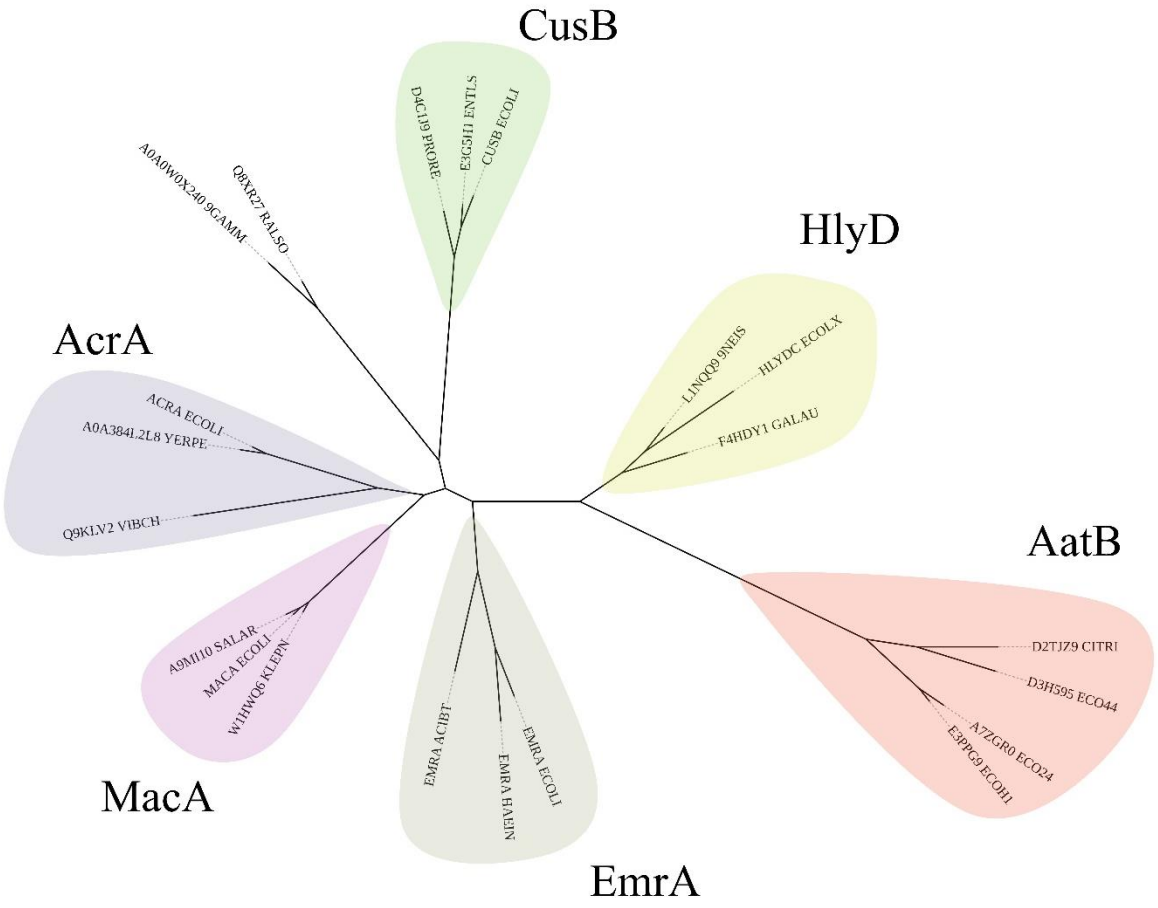
582 **Fig 10**. **Schematic of proposed Aat system mechanism**. **A**, preCexE produced in the cytoplasm is
583 secreted into the periplasm by the Sec pathway (Sec). **B**, The signal sequence is cut by signal peptidase
584 I (SPI) resulting in soluble uCexE. **C**, An acyl chain is added to the N-terminal glycine of uCexE by
585 AatD. Accordingly, mCexE associates with the membrane. **D**, mCexE is extracted out of the inner
586 membrane by the AatP and AatC complex. A single channel is formed comprised of AatP, AatA, AatB,
587 and AatC that allows the secretion of mCexE. **E**, mCexE is inserted into the outer leaflet of the outer
588 membrane. mCexE remains associated with the outer membrane by the single acyl chain incorporated
589 onto the N-terminal glycine.

590



591

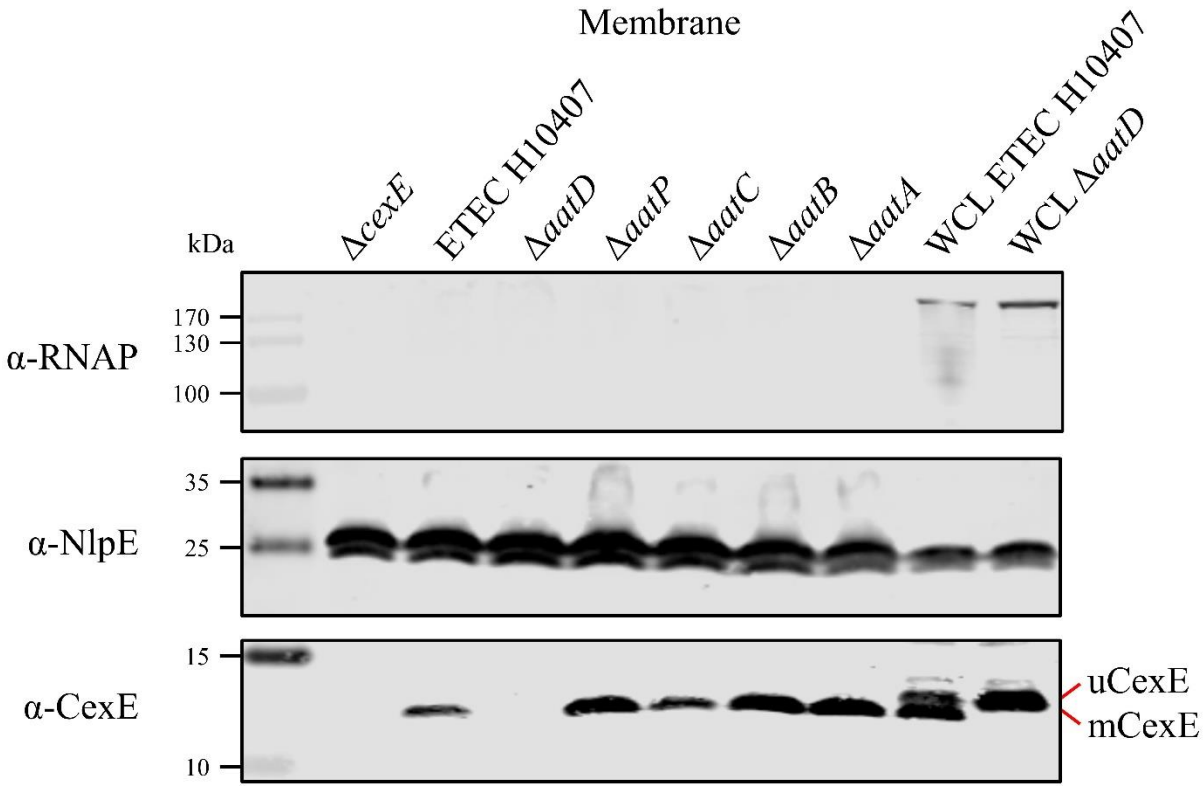
592 **Supplementary Fig S1**. **Phylogram of AatA, TolC and other OMP associated with T1SS and efflux**
593 **pumps.** The HMM of AatA proteins was used to confirm AatA as a homolog of TolC and OMP
594 proteins. Representative sequences of secretion associated OMPs (AprF, CyaE and TolC) and efflux
595 associated OMPs (CusC, OprM, CmeC, and NodT) were used to construct a phylogram. Although AatA
596 is more closely related to the secretion associated OMPs AprF and CyaE than the efflux associated
597 OMPs. TolC is a promiscuous protein that functions as the OMP for secretion and efflux systems. All
598 Phylograms were constructed using RAxML and drawn using iTOL.



599

600 **Supplementary Fig S2**. **Phylogram of AatB and PAPs involved in T1SS and efflux pumps.** The
601 AatB HMM identified it as a homolog of the periplasmic adaptor proteins. Representative sequences of
602 PAPs involved in T1SS and efflux were used to construct a phylogenetic tree using RAxML. The
603 resulting tree was drawn in iTOL. AatB clades with the T1SS PAP HlyD which is involved in
604 haemolysin secretion instead of the efflux associated PAPs of EmrA, MacA, AcrA, and CusB.

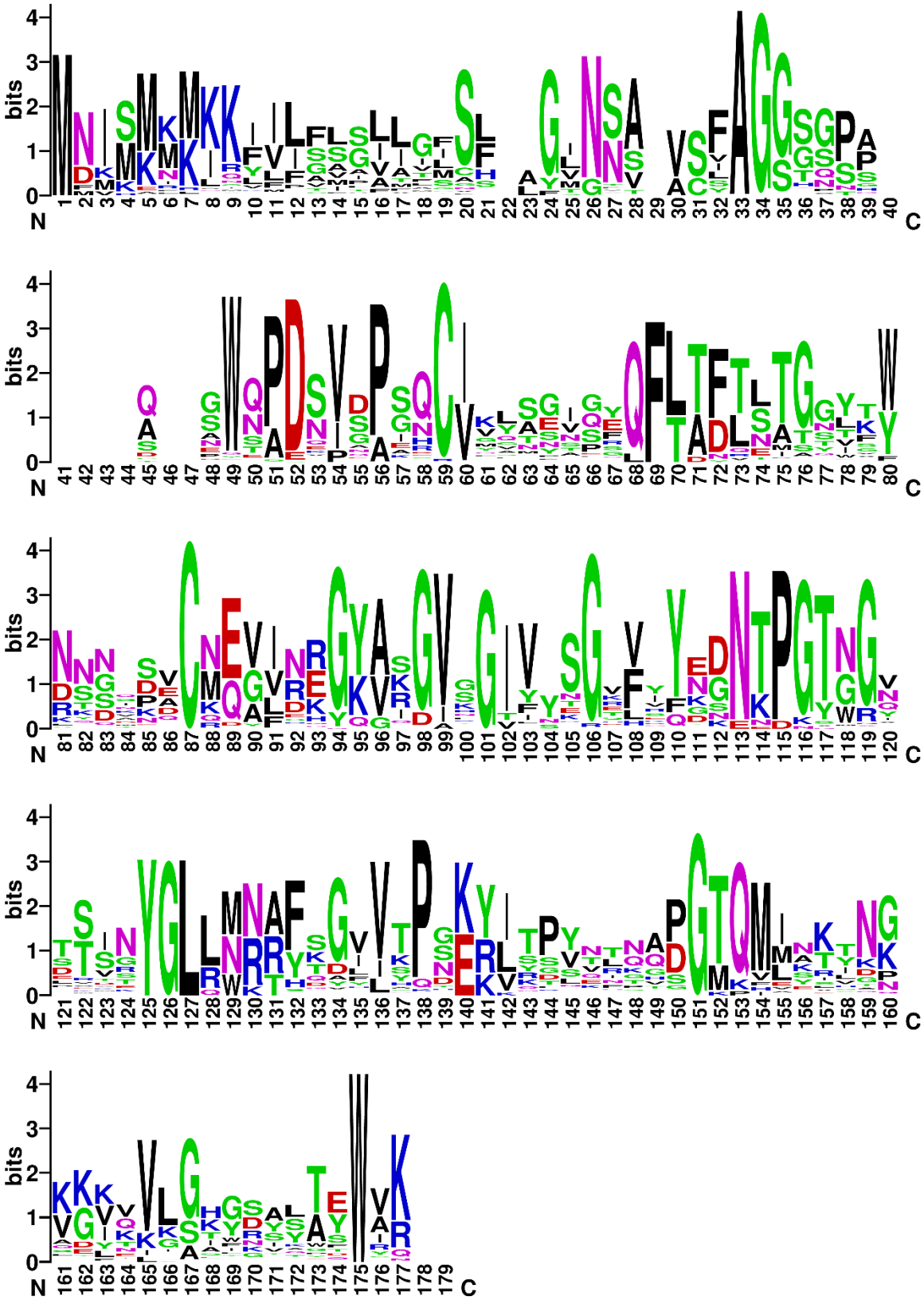
605



606

607 **Supplementary Fig S3. Membrane localisation of CexE.** The membrane fraction of ETEC H10407
608 and *aat* mutant harbouring pCfaD were separated by SDS-PAGE. Transferred to nitrocellulose and
609 probed for the cytoplasmic protein RNAP, lipoprotein NlpE and CexE using protein specific antibodies.

610



Supplementary Fig S4. WebLogo of Aap and CexE sequences. Aap and CexE sequences were aligned using ClustalO. The alignment was then placed into WebLogo.

614 **Supplementary Table S3. Bacterial strains used in this study**

Strain	Description	Source
ETEC H10407	Prototypical enterotoxigenic strain. CFA/1$^+$, LT$^+$, ST$^+$	Evans and Evans (1973)
ETEC H10407Δ*aatP*	*aatP* deletion mutant	This study
ETEC H10407Δ*aatA*	*aatA* deletion mutant	This study
ETEC H10407Δ*aatB*	*aatB* deletion mutant	This study
ETEC H10407Δ*aatC*	*aatC* deletion mutant	This study
ETEC H10407Δ*aatD*	*aatD* deletion mutant	This study
ETEC H10407Δ*cexE*	*cexE* deletion mutant	This study
E. coli DH5α	Cloning strain	NEB
E. coli BL21(DE3)	Protein production strain	Invitrogen
EAEC 042	Prototypical enteroaggregative strain. AAF/1+, EAST1+, Pet+	Nataro et al. (1985)
EAEC 042Δ*aap*	EAEC 042 with pJB5603 inserted into *aap*	Sheikh et al. (2002)
EAEC 042Δ*aatD*	EAEC 042 with pJB5603 inserted into *aatD*	Nishi et al. (2003)

615

616 **Supplementary Table S4: Primers used in this study.**

Name	Sequence	Description
C*cexE*-F	GGCGGCCATATGAAAAAATATATATTAG GTGT	Cloning *cexE* from ETEC H10407 into pET26(b)
C*cexE*-R	CGAAGCCTCGAGTTTATACCAATAAGGG GTGT	Cloning *cexE* from ETEC H10407 into pET26(b)
GD*aatP*-F	GAATTCAAGCTTGCAAAGCGTATTGTTG GTGCAGGCTTGTATAAGTTATATGGACC GGTCAATTGGCTGGAG	Forward primer for generating kanamycin cassette and disruption of *aatP* gene in ETEC H10407
GD*aatP*-R	GAGCTCCATATGAATATAAATATAATCT TCATGAAAATCTTTCTTTTATTAAAATA TCCTCCTTAGTTCC	Reverse primer for generating kanamycin cassette and disruption of *aatP* gene in ETEC H10407
GD*aatA*-F	GAATTCAAGCTTGATTTATCAATCTTAA TAAAAGAAAGATTTTCATGAAGAGACCG GTCAATTGGCTGGAG	Forward primer for generating kanamycin cassette and disruption of *aatA* gene in ETEC H10407
GD*aatA*-R	GAGCTCCATATGTATATAATTCAATTTC ATTTTCCTTTTTTATTAACTCTCAATAT CCTCCTTAGTTCC	Reverse primer for generating kanamycin cassette and disruption of *aatA* gene in ETEC H10407
GD*aatB*-F	GAATTCAAGCTTGAATTATATAAAATGC ATATGTTTTTTATTGGTTAGTTTGACCG GTCAATTGGCTGGAG	Forward primer for generating kanamycin cassette and disruption of *aatB* gene in ETEC H10407
GD*aatB*-R	GAGCTCCATATGATCAATACTTAATTTA ATCATGTAGTGTTATCTCAAATGCTGAA TATCCTCCTTAGTTCC	Reverse primer for generating kanamycin cassette and disruption of *aatB* gene in ETEC H10407
GD*aatC*-F	GAATTCAAGCTTCTTGGATACAGCATTT GAGATAACACTACATGATTAAATTAGAC CGGTCAATTGGCTGGAG	Forward primer for generating kanamycin cassette and disruption of aatC gene in ETEC H10407
GD*aatC*-R	GAGCTCCATATGAAATCACATAAAATTT TTATTAAAAGGATATAACCTTCATAATA TCCTCCTTAGTTCC	Reverse primer for generating kanamycin cassette and disruption of *aatC* gene in ETEC H10407
GD*aatD*-F	GAATTCAAGCTTTCGTATATTGGCATTA AATTATCTTTTGAATTGATTAATGGACC GGTCAATTGGCTGGAG	Forward primer for generating kanamycin cassette and disruption of aatD gene in ETEC H10407
GD*aatD*-R	GAGCTCCATATGACCAAATGATGTATAT GGTTTATACATCAATGACAAAAAAAATA TCCTCCTTAGTTCC	Reverse primer for generating kanamycin cassette and disruption of *aatD* gene in ETEC H10407
cexE-F1	GGGTTAAGTGATAACAGGCG	Forward flanking primer of *cexE* gene in ETEC H10407
cexE-R1	GCACCAACAATACGCTTTGC	Reverse flanking primer of *cexE* gene in ETEC H10407
aatP-F1	TACAATGTCGGGACTCAACC	Forward flanking primer of *aatP* gene in ETEC H10407
aatP-R1	ACATACGCAAATGCGGATGG	Reverse flanking primer of *aatP* gene in ETEC H10407
aatA-F1	CATCAGGTTGAGTTACTCGC	Forward flanking primer of *aatA* gene in ETEC H10407
aatA-R1	CATTGTAAGCATTGCTGGCG	Reverse flanking primer of *aatA* gene in ETEC H10407
aatB-F1	CATTCCGCTAAGTGGATTGC	Forward flanking primer of *aatB* gene in ETEC H10407
aatB-R1	ATATCCGACACATCACTCCC	Reverse flanking primer of *aatB* gene in ETEC H10407
aatC-F1	GTAATCCACATGTGCAAGGG	Forward flanking primer of *aatC* gene in ETEC H10407
aatC-R1	TGCCGTAGTTATTCGTGAGG	Reverse flanking primer of *aatC* gene in ETEC H10407
aatD-F1	CTCCCTCTATAGTGTGTAGC	Forward flanking primer of *aatD* gene in ETEC H10407
aatD-R1	TCTGGCGTCCATTCATTTCC	Reverse flanking primer of *aatD* gene in ETEC H10407
RM_*cexE*_3-RV	ACCACCACCCGCAATCG	Remove all of *cexE* apart from the last 3 N-terminal glycine residues
RM_*cexE*_2-RV	ACCACCCGCAATCGCGC	Remove all of *cexE* apart from the last 2 N-terminal glycine residues
RM_*cexE*_1-RV	ACCCGCAATCGCGCTCAGG	Remove all of *cexE* apart from the last 1 N-terminal glycine residues

RM_*cexE*-RV	CGCAATCGCGCTCAGGCTACC	Remove all of *cexE* apart from the signal sequence.
RM_cexE-FW	GTTAGCAAGGGCGAGGAAGATAAC	Forward primer for removing *cexE* from CexE-mCherry fusion.
G20A-F	TCCGCGATAGCTGCAGGCGGTAATTC	Substitution mutant G20A of *cexE*
G20A-R	GAATTACCGCCTGCAGCTATCGCGGA	Substitution mutant G20A of *cexE*
G20E-F	TCCGCGATAGCTGAAGGCGGTAATTC	Substitution mutant G20E of *cexE*
G20E-R	GAATTACCGCCTTCAGCTATCGCGGA	Substitution mutant G20E of *cexE*
G21A-F	GCGATAGCTGGAGCCGGTAATTCTGAACG	Substitution mutant G21A of *cexE*
G21A-R	CGTTCAGAATTACCGGCTCCAGCTATCGC	Substitution mutant G21A of *cexE*
G21E-F	GCGATAGCTGGAGAAGGTAATTCTGAACG	Substitution mutant G21E of *cexE*
G21E_R	CGTTCAGAATTACCTTCTCCAGCTATCGC	Substitution mutant G21E of *cexE*
G22A-F	GATAGCTGGAGGCGCTAATTCTGAACG	Substitution mutant G22A of *cexE*
G22A-R	CGTTCAGAATTAGCGCCTCCAGCTATC	Substitution mutant G22A of *cexE*
G22E-F	GATAGCTGGAGGCGAGAATTCTGAACG	Substitution mutant G22E of *cexE*
G22E-R	CGTTCAGAATTCTCGCCTCCAGCTATC	Substitution mutant G22E of *cexE*
N23A-RH-F	GCGTCTGAACGACCGCCTTCCG	Substitution mutant N23A of *cexE*
N23E-RH-F	GAATCTGAACGACCGCCTTCCG	Substitution mutant N23E of *cexE*
N23-RH-R	ACCGCCTCCAGCTATCGCG	Reverse primer for both N23A and N23E substitution mutants
S24A-F	GGAGGCGGTAATGCTGAACGACCGCC	Substitution mutant S24A of *cexE*
S24A-R	GGAGGCGGTAATGCTGAACGACCGCC	Substitution mutant S24A of *cexE*
S24E-F	GGAGGCGGTAATGAAGAACGACCGCC	Substitution mutant S24E of *cexE*
S24E-R	GGCGGTCGTTCTTCATTACCGCCTCC	Substitution mutant S24E of *cexE*
H10407_E217A-FW	CGACATCGATATGGTGGTTTTTAGCGCGAACACCGTGTACGGTTTCAAGAACCAG	Substitution mutant E217A of *aatD*
H10407_E217A-RV	CTGGTTCTTGAAACCGTACACGGTGTTCGCGCTAAAAACCACCATATCGATGTCG	Substitution mutant E217A of *aatD*
H10407_K278A-FW	CTATTACAAAGACAACTTCCTGATCAACCAAGCGAAAGCGCTGATTCCGTTTGTTGAGCAGC	Substitution mutant K278A of *aatD*
H10407_K278A-RV	GCTGCTCAACAAACGGAATCAGCGCTTTCGCTTGGTTGATCAGGAAGTTGTCTTTGTAATAG	Substitution mutant K278A of *aatD*
H10407_E286A-FW	GCGCTGATTCCGTTTGTTGCGCAGCAATGGTTCTTTAGC	Substitution mutant E286A of *aatD*
H10407_E286A-RV	GCTAAAGAACCATTGCTGCGCAACAAACGGAATCAGCGC	Substitution mutant E286A of *aatD*
H10407_C325A-FW	CAACCACAACGGTATCAACATTAAGACCTATATTGCGTACGATGTTCTGTTCCCGGAAACCCATAAAAGC	Substitution mutant C325A of *aatD*
H10407_C325A-RV	GCTTTTATGGGTTTCCGGGAACAGAACATCGTACGCAATATAGGTCTTAATGTTGATACCGTTGTGGTTG	Substitution mutant C325A of *aatD*
T7F	TAATACGACTCACTATAGGG	Forward primer for amplification of pET26b(+) multiple cloning site
T7R	GCTAGTTATTGCTCAGCGG	Reverse primer for amplification of pET26b(+) multiple cloning site

617

618 **Supplementary Table S5. Plasmids used in this study**

Plasmid	Description	Source
pDOC-K	contains *bla* and *kanR* with FRT sites	Lee et al. (2009)
pKD46	Heat sensitive plasmid with λ Red recombinase genes expressed in response to L-arabinose	Datsenko and Wanner (2000)
pCP20	Heat sensitive plasmid encoding Flp recombinase gene	Datsenko and Wanner (2000)
pET26b(+)	T7 expression vector with C-terminal 6His tag, IPTG inducible, *kanR*	Novagen
pET26b-*cexE*	pET26b with *cexE* from ETEC H10407	This study
pBAD/myc-HisA	L-arabinose expression vector with C-terminal Myc and 6His tag	Invitrogen
pJNW	Complete *aat* gene cluster from EAEC 042 cloned into pZC320	Nishi et al. (2003)
pCfaD	pBAD/myc-HisA with *cfaD* under PBAD from ETEC H10407	This study
pACYC184	Plasmid with p15A origin of replication	Chang and Cohen (1978)
pACYC-*cexE*-6His	pACYC184 with *cexE* under the control of the native *cexE* promoter with a C-terminal 6 His tag	This study
pACYCDuet-1	Plasmid with p15A origin of replication and two T7 promoters.	Novagen
pACYC-*aatD*	The ETEC H10407 *aatD* gene was synthesised by Genscript and transformed in pACYCDuet-1	This study
pRSF-GGG-mCherry	mCherry with three glycine residues at the N-terminus after the CexE signal sequence	This study
pRSF-GG-mCherry	mCherry with two glycine residues at the N-terminus after the CexE signal sequence	This study
pRSF-G-mCherry	mCherry with one glycine residue at the N-terminus after the CexE signal sequence	This study
pRSF-SS-mCherry	mCherry with the CexE signal sequence	This study

619

620 **References**

621

622 Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J.
623 (1997). Gapped BLAST and PSI-BLAST : a new generation of protein database search
624 programs. *Nucleic Acids Res*, *25*(17), 3389–3402.
625 Babu, M. M., Priya, M. L., Selvan, A. T., Madera, M., Gough, J., Aravind, L., & Sankaran, K. (2006).
626 A database of bacterial lipoproteins (DOLOP) with functional assignments to predicted
627 lipoproteins. *J Bacteriol*, *188*(8), 2761–2773. https://doi.org/10.1128/JB.188.8.2761-2773.2006
628 Baldi, D. L., Higginson, E. E., Hocking, D. M., Praszkier, J., Cavaliere, R., James, C. E., Bennett-
629 Wood, V., Azzopardi, K. I., Turnbull, L., Lithgow, T., Robins-Browne, R. M., Whitchurch, C.
630 B., & Tauschek, M. (2012). The type II secretion system and its ubiquitous lipoprotein substrate,
631 SslE, are required for biofilm formation and virulence of enteropathogenic Escherichia coli.
632 *Infect Immun*, *80*(6), 2042–2052. https://doi.org/10.1128/IAI.06160-11
633 Baudry, B., Savarino, S. J., Vial, P., Kaper, J. B., & Levine, M. M. (1990). A Sensitive and Specific
634 DNA Probe to Identify Enteroaggregative Escherichia coli, a Recently Discovered Diarrheal
635 Pathogen. *J Infect Dis*, *161*(6), 1249–1251.
636 Belmont-Monroy, L., Saitz-rojas, W., Soria-bustos, J., Mickey, S., Sherman, N. E., Orsburn, B. C.,
637 Ruiz-perez, F., & Santiago, A. E. (2020). Characterization of a novel AraC/XylS- regulated
638 family of N-acyltransferases in pathogens of the order Enterobacterales. *PLoS Pathog*, *16*(8),
639 e1008776. https://doi.org/10.1371/journal.ppat.1008776
640 Chang, A. C. Y., & Cohen, S. N. (1978). Construction and characterization of amplifiable multicopy
641 DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J Bacteriol*, *134*(3), 1141–
642 1156. https://doi.org/10.1128/jb.134.3.1141-1156.1978
643 Cowles, C. E., Li, Y., Semmelhack, M. F., Cristea, I. M., & Silhavy, T. J. (2012). The free and bound
644 forms of Lpp occupy distinct subcellular locations in Escherichia coli. *Mol Microbiol*, *79*(5),
645 1168–1181. https://doi.org/10.1111/j.1365-2958.2011.07539.x.The
646 Crossman, L. C., Chaudhuri, R. R., Beatson, S. A., Wells, T. J., Desvaux, M., Cunningham, A. F.,
647 Petty, N. K., Mahon, V., Brinkley, C., Hobman, J. L., Savarino, S. J., Turner, S. M., Pallen, M.
648 J., Penn, C. W., Parkhill, J., Turner, A. K., Johnson, T. J., Thomson, N. R., Smith, S. G. J., &
649 Henderson, I. R. (2010). A commensal gone bad: complete genome sequence of the prototypical
650 enterotoxigenic Escherichia coli strain H10407. *J Bacteriol*, *192*(21), 5822–5831.
651 https://doi.org/10.1128/JB.00710-10
652 Crow, A., Greene, N. P., Kaplan, E., & Koronakis, V. (2017). Structure and mechanotransmission
653 mechanism of the MacB ABC transporter superfamily. *Proc Natl Acad Sci USA*, *114*(47),
654 12572–12577. https://doi.org/10.1073/pnas.1712153114
655 Datsenko, K. a, & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in Escherichia
656 coli K-12 using PCR products. *Proc Natl Acad Sci USA*, *97*(12), 6640–6645.
657 https://doi.org/10.1073/pnas.120163297
658 Dev, I. K., Harvey, R. J., & Ray, P. H. (1985). Inhibition of prolipoprotein signal peptidase by
659 globomycin. *J Biol Chem*, *260*(10), 5891–5894.
660 Duronio, R. J., Jackson-Machelski, E., Heuckeroth, R. O., Olins, P. O., Devine, C. S., Yonemoto, W.,
661 Slice, L. W., Taylor, S. S., & Gordon, J. I. (1990). Protein N-myristoylation in Escherichia coli:
662 Reconstitution of a eukaryotic protein modification in bacteria. *Proc Natl Acad Sci USA*, *87*(4),
663 1506–1510. https://doi.org/10.1073/pnas.87.4.1506
664 Evans, D. J., & Evans, D. G. (1973). Three Characteristics Associated with Enterotoxigenic
665 Escherichia coli Isolated from Man. *Infect Immun*, *8*(3), 322–328.
666 Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: interactive sequence similarity
667 searching. *Nucleic Acids Res*, *39*(W29), W29-37. https://doi.org/10.1093/nar/gkr367
668 Fitzpatrick, A. W. P., Llabrés, S., Neuberger, A., Blaza, J. N., Bai, X., Okada, U., Murakami, S.,
669 Veen, H. W. Van, Zachariae, U., Scheres, S. H. W., Luisi, B. F., & Du, D. (2017). Structure of
670 the MacAB–TolC ABC-type tripartite multidrug efflux pump. *Nat Microbiol*, *2*, 17070.
671 https://doi.org/10.1038/nmicrobiol.2017.70
672 Gélis-Jeanvoine, S., Lory, S., Oberto, J., & Buddelmeijer, N. (2015). Residues located on membrane-

673 embedded flexible loops are essential for the second step of the apolipoprotein N-acyltransferase
674 reaction. *Mol Microbiol*, *95*(4), 692–705. https://doi.org/10.1111/mmi.12897
675 Grabowicz, M. (2019). Lipoproteins and Their Trafficking to the Outer Membrane. *EcoSal Plus*, *8*(2),
676 1–8. https://doi.org/10.1128/ecosalplus.esp-0038-2018
677 Gupta, S. D., Gan, K., Schmid, M. B., & Wu, H. C. (1993). Characterization of a temperature-
678 sensitive mutant of Salmonella typhimurium defective in apolipoprotein N-acyltransferase. *J*
679 *Biol Chem*, *268*(22), 16551–16556.
680 Hancock, J. F., Magee, A. I., Childs, J. E., & Marshall, C. J. (1989). All ras proteins are
681 polyisoprenylated but only some are palmitoylated. *Cell*, *57*(7), 1167–1177.
682 https://doi.org/10.1016/0092-8674(89)90054-8
683 Hodson, C., Yang, J., Hocking, D. M., Azzopardi, K., Chen, Q., Holien, J. K., Parker, M. W.,
684 Tauschek, M., & Robins-browne, R. M. (2017). Control of Virulence Gene Expression by the
685 Master Regulator, CfaD, in the Prototypical Enterotoxigenic Escherichia coli Strain, H10407.
686 *Front Microbiol*, *11*(8), 1525. https://doi.org/10.3389/fmicb.2017.01525
687 Hooda, Y., Lai, C. C. L., Judd, A., Buckwalter, C. M., Shin, H. E., Gray-Owen, S. D., & Moraes, T.
688 F. (2016). Slam is an outer membrane protein that is required for the surface display of lipidated
689 virulence factors in Neisseria. *Nat Microbiol*, *1*(4), 16009.
690 https://doi.org/10.1038/nmicrobiol.2016.9
691 Huie, J. L., & Silhavy, T. J. (1995). Suppression of signal sequence defects and azide resistance in
692 Escherichia coli commonly result from the same mutations in secA. *J Bacteriol*, *177*(12), 3518–
693 3526.
694 Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2016). The Phyre2 web
695 protal for protein modeling, prediction and analysis. *Nat Protoc*, *10*(6), 845–858.
696 https://doi.org/10.1038/nprot.2015-053
697 Konovalova, A., Perlman, D. H., Cowles, C. E., & Silhavy, T. J. (2014). Transmembrane domain of
698 surface-exposed outer membrane lipoprotein RcsF is threaded through the lumen of β -barrel
699 proteins. *Proc Natl Acad Sci USA*, *111*(41), E4350–E4358.
700 https://doi.org/10.1073/pnas.1417138111
701 Kurtzhals, P. (2007). Pharmacology of Insulin Detemir. *Endocrinol Metab Clin North Am*, *36*(1), 14–
702 20. https://doi.org/10.1016/S0889-8529(07)80004-1
703 Lee, D. J., Bingle, L. E. H., Heurlier, K., Pallen, M. J., Penn, C. W., Busby, S. J. W., & Hobman, J. L.
704 (2009). Gene doctoring: a method for recombineering in laboratory and pathogenic Escherichia
705 coli strains. *BMC Microbiol*, *9*, 252. https://doi.org/10.1186/1471-2180-9-252
706 Letunic, I., & Bork, P. (2019). Interactive Tree of Life (iTOL) v4: Recent updates and new
707 developments. *Nucleic Acids Res*, *47*(W1), 256–259. https://doi.org/10.1093/nar/gkz239
708 Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R.
709 N., Potter, S. C., Finn, R. D., & Lopez, R. (2019). The EMBL-EBI search and sequence analysis
710 tools APIs in 2019. *Nucleic Acids Res*, *47*(W1), W636–W641.
711 https://doi.org/10.1093/nar/gkz268
712 Martin, D. D. O., Beauchamp, E., & Berthiaume, L. G. (2011). Post-translational myristoylation: Fat
713 matters in cellular life and death. *Biochimie*, *93*(1), 18–31.
714 https://doi.org/10.1016/j.biochi.2010.10.018
715 Masao Tokunaga, Hiroko Tokunaga, & Wu, H. C. (1984). Post-translational modification and
716 processing of outer membrane prolipoproteins in Escherichia coli. *Mol Cell Biochem*, *60*(1), 5–
717 15. https://doi.org/10.1007/BF00226297
718 Moore, S. D., & Prevelige, P. E. (2002). A P22 Scaffold Protein Mutation Increases the Robustness of
719 Head Assembly in the Presence of Excess Portal Protein. *J Virol*, *76*(20), 10245–10255.
720 https://doi.org/10.1128/jvi.76.20.10245-10255.2002
721 Nakayama, H., Kurokawa, K., & Lee, B. L. (2012). Lipoproteins in bacteria: structures and
722 biosynthetic pathways. *FEBS J*, *279*(23), 4247–4268. https://doi.org/10.1111/febs.12041
723 Nataro, J., Baldini, M., Kaper, J., Black, R., Bravo, N., & Levine, M. (1985). Detection of an
724 adherence factor of enteropathogenic Escherichia coli with a DNA probe. *J Infect Dis*, *152*(3),
725 560–565.
726 Nishi, J., Sheikh, J., Mizuguchi, K., Luisi, B., Burland, V., Boutin, A., Rose, D. J., Blattner, F. R., &
727 Nataro, J. P. (2003). The export of coat protein from enteroaggregative Escherichia coli by a

specific ATP-binding cassette transporter system. *J Biol Chem*, *278*(46), 45680–45689. https://doi.org/10.1074/jbc.M306413200

Oliver, D. B., Cabelli, R. J., Dolan, K. M., & Jarosik, G. P. (1990). Azide-resistant mutants of Escherichia coli alter the SecA protein, an azide-sensitive component of the protein export machinery. *Proc Natl Acad Sci USA*, *87*(21), 8227–8231.

Pepinsky, R. B., Zeng, C., Went, D., Rayhorn, P., Baker, D. P., Williams, K. P., Bixler, S. A., Ambrose, C. M., Garber, E. A., Miatkowski, K., Taylor, F. R., Wang, E. A., & Galdes, A. (1998). Identification of a palmitic acid-modified form of human Sonic hedgehog. *J Biol Chem*, *273*(22), 14037–14045. https://doi.org/10.1074/jbc.273.22.14037

Petty, N. K., Bulgin, R., Crepin, V. F., Cerdeño-Tárraga, A. M., Schroeder, G. N., Quail, M. a, Lennard, N., Corton, C., Barron, A., Clark, L., Toribio, A. L., Parkhill, J., Dougan, G., Frankel, G., & Thomson, N. R. (2010). The Citrobacter rodentium genome sequence reveals convergent evolution with human pathogenic Escherichia coli. *J Bacteriol*, *192*(2), 525–538. https://doi.org/10.1128/JB.01144-09

Pilonieta, M. C., Bodero, M. D., & Munson, G. P. (2007). CfaD-dependent expression of a novel extracytoplasmic protein from enterotoxigenic Escherichia coli. *J Bacteriol*, *189*(14), 5060–5067. https://doi.org/10.1128/JB.00131-07

Resh, M. D. (1994). Myristylation and palmitylation of Src family members: The fats of the matter. *Cell*, *76*(3), 411–413. https://doi.org/10.1016/0092-8674(94)90104-X

Rivas, Z. P., Talbot, K. M., Merselis, L. C., McCormack, R. M., Adkins, B., & Munson, G. P. (2020). CexE Is a Coat Protein and Virulence Factor of Diarrheagenic Pathogens. *Front Microbiol*, *11*(1374), 1–13. https://doi.org/10.3389/fmicb.2020.01374

Schägger, H. (2006). Tricine-SDS-PAGE. *Nat Protoc*, *1*(1), 16–22. https://doi.org/10.1038/nprot.2006.4

Sheikh, J., Czeczulin, J. R., Harrington, S., Hicks, S., Henderson, I. R., Le Bouguénec, C., Gounon, P., Phillips, A., & Nataro, J. P. (2002). A novel dispersin protein in enteroaggregative Escherichia coli. *J Clin Invest*, *110*(9), 1329–1337. https://doi.org/10.1172/JCI16172

Stamatakis, A. (2014). RAxML version 8 : a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

The UniProt Consortium. (2019). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res*, *47*, D506–D515. https://doi.org/10.1093/nar/gky1049

Thompson, B. J., Widdick, D. A., Hicks, M. G., Chandra, G., Sutcliffe, I. C., Palmer, T., & Hutchings, M. I. (2010). Investigating lipoprotein biogenesis and function in the model Gram-positive bacterium Streptomyces coelicolor. *Mol Microbiol*, *77*(4), 943–957. https://doi.org/10.1111/j.1365-2958.2010.07261.x

Veronese, F. D., Copeland, T. D., Oroszlan, S., Gallo, R. C., & Sarngadharan, M. G. (1988). Biochemical and immunological analysis of human immunodeficiency virus gag gene products p17 and p24. *J Virol*, *62*(3), 795–801. https://doi.org/10.1128/jvi.62.3.795-801.1988

Vidal-Ingigliardi, D., Lewenza, S., & Buddelmeijer, N. (2007). Identification of Essential Residues in Apolipoprotein N-Acyl Transferase, a Member of the CN Hydrolase Family. *J Bacteriol*, *189*(12), 4456–4464. https://doi.org/10.1128/JB.00099-07

Widdick, D. A., Hicks, M. G., Thompson, B. J., Tschumi, A., Chandra, G., Sutcliffe, I. C., Brülle, J. K., Sander, P., Palmer, T., & Hutchings, M. I. (2011). Dissecting the complete lipoprotein biogenesis pathway in Streptomyces scabies. *Mol Microbiol*, *80*(5), 1395–1412. https://doi.org/10.1111/j.1365-2958.2011.07656.x

Wiktor, M., Weichert, D., Howe, N., Huang, C. Y., Olieric, V., Boland, C., Bailey, J., Vogeley, L., Stansfeld, P. J., Buddelmeijer, N., Wang, M., & Caffrey, M. (2017). Structural insights into the mechanism of the membrane integral N-acyltransferase step in bacterial lipoprotein synthesis. *Nat Commun*, *8*, 15952. https://doi.org/10.1038/ncomms15952

Yamanaka, H., Kobayashi, H., Takahashi, E., & Okamoto, K. (2008). MacAB is involved in the secretion of Escherichia coli heat-stable enterotoxin II. *J Bacteriol*, *190*(23), 7693–7698. https://doi.org/10.1128/JB.00853-08

Zückert, W. R. (2014). Secretion of bacterial lipoproteins: through the cytoplasmic membrane, the periplasm and beyond. *Biochim Biophys Acta*, *1843*(8), 1509–1516. https://doi.org/10.1016/j.bbamcr.2014.04.022

783
784