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

Christopher Icke[1], Freya J. Hodges[1], Karthik Pullela[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)

Running title: Glycine acylation mediates lipoprotein secretion

Abstract

Cell envelope 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 the majority of 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).

26 **Introduction**

27 Diarrhoeal disease remains a major cause of death and morbidity across the globe. Persistent or
28 recurring infections of diarrhoeal pathogens can lead to malnutrition, developmental stunting, and
29 cognitive impairment (1). Members of the *Enterobacteriaceae*, such as *Escherichia coli* or *Salmonella*
30 *enterica*, are leading causes of diarrhoeal morbidity and mortality (1). Owing to increasing antibiotic
31 resistance within the *Enterobacteriaceae*, new methods are urgently required to treat and prevent such
32 infections (2).

33 Understanding the molecular mechanisms underpinning infection and virulence can provide
34 new targets for the design of antibiotics or vaccines that decrease pathology and prevent disease.
35 Lipoproteins are key components of bacterial pathogens and have been targeted for antibiotic and
36 vaccine development (3). In Gram-negative bacteria, lipoproteins are transported across the inner
37 membrane from the cytosol to the periplasm by the Sec pathway. The signal peptide inserts into the
38 inner membrane and a diacylglycerol is attached by Lgt to the sulphur moiety of the invariant cysteine
39 at the +1 position of the lipoprotein. The signal sequence is then cleaved by Lsp exposing the N-terminal
40 amine group of the cysteine for monoacylation by Lnt (4,5). Lnt mediated acylation occurs at the free
41 amine moiety of the N-terminal cysteine after signal sequence cleavage by Lsp (Signal Peptidase II).
42 The mature triacylated lipoprotein remains embedded in the inner membrane or is localised to the inner
43 leaflet of the outer membrane by the essential Lol pathway. LolCE, in combination with the ATPase
44 LolD, extracts the triacylated lipoprotein from the inner membrane for transport across the periplasm
45 by LolA, where it is then incorporated into the inner leaflet of the outer membrane by LolB. While the
46 majority remain periplasmically located, in recent years there have been a number of descriptions of
47 surface localised outer membrane lipoproteins (6–8). However, the mechanism of translocation across
48 the outer membrane to the cell surface remains poorly understood.

49 Previously, we described two outer membrane proteins, CexE and Aap, associated with
50 enterotoxigenic (ETEC) and enteroaggregative *Escherichia coli* (EAEC), respectively (9,10). CexE has
51 been implicated in prolonged bacterial shedding and increased severity of infection (11) whereas Aap
52 influences biofilm formation and gut colonisation (9,12). Despite Aap and CexE sharing only 18%

53 amino acid identity, both proteins are secreted by the Aat system (11–13). The Aat system requires five
54 proteins (AatPABCD) to facilitate protein secretion, two of which bear resemblance to components of
55 the Type 1 secretion system (T1SS); an outer membrane protein (OMP), and a periplasmic adaptor
56 protein (PAP). In contrast to the T1SS, the CexE and Aap substrate molecules are translocated using a
57 two-step mechanism. First Aap/CexE is translocated across the inner membrane into the periplasm by
58 the Sec pathway (14). It then enters the Aat system to be secreted across the outer membrane (11–13).

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

70 **Results**

71 **Distribution of the Aat system**

72 The Aat system was first identified in EAEC, where it corresponds to the molecular probe
73 (CVD432) used to define this *E. coli* pathovar (13,15). In order to determine whether the Aat system
74 was more widespread each of the Aat proteins from ETEC H0407 was used to search the non-redundant
75 protein sequence database using repetitive iterations of the PSI-BLAST algorithm. The full Aat system
76 and an Aap or CexE homolog was identified in 826 separate nucleotide accessions (Supplementary
77 Table S1). This revealed that the Aat system is distributed more widely than initially anticipated and is
78 encoded in pathogens with diverse mechanisms of virulence such as ETEC, EAEC, enteropathogenic

79 *E. coli*, Shiga-toxin producing *E. coli*, *Shigella* sp., *Salmonella enterica, Citrobacter rodentium*,
80 *Providencia alcalifaciens* and *Yersinia entercolitica* (10,11,16).

81 To understand the conservation of the Aat system, the organisation of the *aat* operon was
82 examined for each of the 826 genomes identified above. The relative genomic distance between each
83 of the *aat* genes and *aap*/*cexE* was calculated from their boundaries on the nucleotide accession. From
84 this analysis six separate organisations of the *aat* operon were identified (Fig 1). In the most common
85 organisation *aap*/*cexE* are separated from the *aatPABCD* operon by at least 1 kb. This accounted for
86 just over half of the *aat* systems identified. In just under a quarter of *aat* operons identified, *aap*/*cexE*
87 are within 1 kb of the *aatPABCD* operon either within 400 bp upstream (18.4%) or 1 kb downstream
88 (4.8%). In 25% of the *aat* systems identified *aatD* was separated from *aatPABC* by >1 kb. The most
89 common of these organisations was *aatD* separated from the other members of the *aat* system and
90 substrate protein (11.9%), closely followed by *aatD* and *aap*/*cexE* together but greater than 1 kb
91 distance from for other *aat* genes (9.0%). Finally, in the least common example *aatD*, *aap*/*cexE* and
92 *aatPABC* were all separated by greater than 1 kb (4.1%). We did not identify any examples of *aatD*
93 encoded between 39 bp and 1 kb from the stop codon of *aatC*. However, despite these differences, in
94 every example the *aatPABC* genes formed a single operonic unit located with *aatD* and *aap*/*cexE* on a
95 chromosomal pathogenicity island or a large virulence plasmid. These data suggest that AatPABCD are
96 sufficient and necessary to mediate Aap/CexE secretion.

97 **Aat-dependent secretion of CexE**

98 To test whether all of the *aat* genes are required for CexE secretion, we constructed single gene
99 deletion mutants of *cexE* and each *aat* gene in ETEC H10407 pCfaD. The pCfaD plasmid encodes the
100 CfaD transcriptional activator under the control of an arabinose inducible promoter allowing
101 constitutive expression of the CfaD-dependent *cexE* and the *aat* genes in the presence of arabinose
102 (14,17). SDS-PAGE analysis of culture supernatant fractions collected from ETEC H10407 pCfaD
103 revealed a protein with an apparent molecular mass of 11.8 kDa that could be detected by western
104 blotting with CexE-specific polyclonal antibodies (Fig 2A). In contradistinction to the parent strain, no
105 protein was detected in the culture supernatant fractions derived from cultures of the ETEC H10407

106 pCfaD *aat* or *cexE* mutants (Fig 2A). However, to ensure that the lack of CexE in the culture supernatant
107 was not due to the lack of CexE production, but rather a result of an inability to secrete CexE, the
108 presence of CexE in whole cell lysates was determined. CexE was detected by western blotting with
109 CexE-specific antibodies in whole cell lysates of ETEC H10407 pCfaD and all *aat* mutants. In contrast,
110 CexE could not be detected in the *cexE* null mutant (Fig. 2A). From these data we conclude that the Aat
111 system is essential for CexE secretion.

112 **Post-translational modification of Aat substrate molecules**

113 Unexpectedly, when the whole cell lysates of ETEC H10407 pCfaD and its isogenic *aat*
114 mutants were examined by western blotting, CexE appeared as two bands in the parental strain and each
115 of the *aat* mutants except for *aatD* (Fig. 2B). The upper band had an apparent molecular mass of 12.7
116 kDa (uCexE) while the lower band (mCexE) had an apparent molecular mass of 11.8 kDa, which is
117 consistent with the secreted form of CexE recovered from the supernatant fractions. As CexE is
118 exported into the periplasm by the Sec pathway (14), the inefficient cleavage of a signal peptide could
119 be responsible for the difference in the molecular mass of uCexE compared to mCexE . To investigate
120 this, ETEC H10407 pCfaD was grown in the presence of sodium azide. Sodium azide inhibits SecA to
121 prevent translocation of proteins across the inner membrane and subsequent cleavage of the signal
122 peptide (18,19). In the absence of sodium azide the mCexE and uCexE forms could be detected in whole
123 cell lysates as observed previously (Fig 2B and C). However, in the presence of sodium azide a protein
124 (proCexE) with an apparent molecular mass of 13.2 kDa was detected (Fig. 2C). These observations
125 suggest that the difference in molecular mass between uCexE and mCexE was not due to inefficient
126 cleavage of the signal peptide but was due to a post translational event mediated by AatD.

127 A size change in the CexE homolog Aap has not been reported before despite numerous
128 investigations of the Aat system. To determine whether the migration of Aap was altered by AatD, the
129 effect of an *aatD* deletion on the apparent size of Aap was investigated in EAEC 042. EAEC 042, an
130 *aap* mutant, an *aatD* mutant and an *aatD* mutant complemented with pJNW were grown in DMEM-
131 HG; the pJNW plasmid encodes the complete *aat* operon from EAEC 042. Whole cell lysates of each
132 strain were analysed by SDS-PAGE. As noted above for CexE, the deletion of *aatD* in EAEC 042

133 resulted in an apparent increase in size of Aap (Fig 2D). The size change of Aap in the *aatD* mutant
134 could be restored to the wild-type size by complementation with the pJNW plasmid. From these data
135 we conclude that AatD can modify Aap and CexE post-translationally and that this modification is
136 required for secretion of the mature substrate molecule to the extracellular milieu.

137 **Predicted functions of Aat components**

138 To assist our understanding of the contribution of each Aat component to the secretion of CexE
139 we constructed hidden Markov-models (HMM) for individual Aat proteins. These were used to search
140 the Uniprot database for distant homologs. In agreement with previous publications, AatA is a homolog
141 of the trimeric outer membrane protein TolC (Supplementary Fig S1) and AatB is a periplasmic adaptor
142 protein (Supplementary Fig S2), components associated with T1SS and drug efflux pumps (13). In
143 contrast, AatD was not homologous to any T1SS or efflux-associated proteins. Instead, AatD was
144 homologous to apolipoprotein N-acyltransferase (Lnt), which is a member of the carbon-nitrogen
145 hydrolase (C-N hydrolase) family (PF00795) required for bacterial lipoprotein acylation and transport
146 via the Lol system (Fig 3A). Comparison of the C-N hydrolase domain of AatD with other members of
147 PF00795 revealed that AatD forms a clade with Lnt suggesting that they are functionally related. The
148 ATPase domain of AatC is part of a clade containing the ATPase domains of LolD, the ATPase subunit
149 of the LolCDE lipoprotein transporter (Fig 3B). Similarly, AatP is more closely related to LolC and
150 LolE than typical ABC transporters of the T1SS, such as HlyB (Fig 3C). The homology of AatP, AatC,
151 and AatD to proteins of the Lol pathway suggests that the Aat system is a composite system of a T1SS
152 and the Lol lipoprotein trafficking system.

153 **AatD-mediated acylation of mCexE**

154 As AatD is a homolog of the Lnt acyltransferase, we hypothesised that mCexE represents an
155 acylated form of CexE. To test this hypothesis, we used 17-ODYA, an 18-carbon alkyne fatty acid that
156 can be conjugated to an azide linked fluorescent molecule by a copper(I)-catalyzed azide-alkyne
157 cycloaddition (CuAAC) reaction to investigate CexE lipidation. The *cexE* gene was introduced into a
158 *cexE* mutant or a *cexE aatD* double mutant on a plasmid termed pACYC-*cexE*-6His, which encodes
159 *cexE* under the transcriptional control of its native promoter. This allowed production of a C-terminally

160 His-tagged variant of CexE. Both strains harbouring pACYC-*cexE*-6His were grown in the presence or
161 absence of 17-ODYA. The bacterial lipoprotein YraP, was included as a positive control. CexE and
162 YraP were isolated from each culture by virtue of their C-terminal 6xHis-tag by using nickel affinity
163 chromatography. The purified proteins were then linked to an azide derivative of Alexa Fluor 488 using
164 a CuAAC reaction. After the CuAAC reaction the CexE and YraP samples were separated by SDS-
165 PAGE and fluorescence was measured (Fig 4). In the presence of AatD and 17-ODYA CexE was
166 fluorescently labelled; in the absence of either 17-ODYA or AatD no fluorescence was observed.
167 Moreover, only mCexE was labelled indicating that uCexE is an unacylated form of CexE; counter
168 intuitively, the acylated form (mCexE) migrates faster on SDS-PAGE than the unacylated uCexE. As
169 acylated proteins associate with membranes, we examined the cellular compartmentalisation of mCexE
170 (Supplementary Fig S3). ETEC H10407 pCfaD and isogenic *aat* mutants were grown in LB
171 supplemented with L-arabinose. Subsequently, cells were collected by centrifugation, lysed and the
172 total membrane fraction was harvested by centrifugation. After separation by SDS-PAGE, only mCexE
173 was detectable by western blotting in the membrane fraction of the parent strain and the *aatPABC*
174 mutants; no mCexE could be detected in the membranes recovered from the *aatD* mutant
175 (Supplementary Fig S3). In contrast, both mCexE and uCexE could be detected in whole cell lysates of
176 ETEC H10407 pCfaD and only uCexE could be detected in whole cell lysates of the *aatD* mutant. These
177 data suggest that AatD is an acyltransferase that mediates post-translational modification of CexE by
178 the addition of one or more acyl chains.

179 To determine whether AatD was sufficient to mediate the acylation of mCexE, plasmids
180 encoding AatD and CexE were introduced into the laboratory strain *E. coli* BL21(DE3), which does not
181 encode any of the Aat proteins or Aap/CexE. *E. coli* BL21 was transformed with pET26b-*cexE* and
182 pACYC-*aatD* or their respective empty vector controls pET26b and pACYCDuet-1. These strains were
183 grown in LB and the production of AatD and CexE was induced by IPTG. The production of CexE was
184 monitored via western blotting (Fig 5). CexE was not detected in strains that contained pET26b. uCexE
185 was produced in the absence of AatD at the size expected for the unmodified form after cleavage of the
186 signal sequence. However, when CexE and AatD were produced together, mCexE migrated further (Fig
187 5). These data suggest AatD was solely responsible for the modification of CexE.

188 **Catalytic residues of AatD**

189 The C-N hydrolase family of proteins have a known conserved catalytic triad, which in the case
190 of Lnt is E267, K335, and C387. Another Lnt residue E343 helps to stabilise the catalytic site (20).
191 Mutation of any of these residues to alanine results in a loss of Lnt function (21). To assess the
192 conservation of the catalytic residues the AatD sequence of EAEC 042 and ETEC H10407 and Lnt from
193 *E. coli* MG1655 were aligned using Clustal Omega. Each of the residues associated with the catalytic
194 site were conserved in both versions of AatD (Fig 6A). The C-N hydrolase catalytic triad is conserved
195 in AatD: E217; K278; and C325. Also, Lnt E343 is conserved at position E286 in AatD. Further AatD
196 sequences were identified using PSI-BLAST. These sequences were aligned and a WebLogo was
197 produced for the regions flanking each of the catalytic residues; the four residues were 100% conserved
198 in all AatD sequences identified. In addition, when a structure of AatD predicted by Phyre2 (22) was
199 compared to Lnt (PDB: 5N6L) (23) the four AatD residues (E217, K278, E286, and C325) were
200 superimposed on the Lnt catalytic triad suggesting that these residues were required for acylation of
201 CexE (Fig 6B and 6C). To test this hypothesis, pACYC-*aatD* derivatives encoding mutant derivatives
202 of AatD with alanine substitutions at the conserved E217, K276, E286, or C325 were introduced into
203 *E. coli* BL21(DE3) pET26b-*cexE*. Expression of AatD and CexE were induced by the addition of IPTG
204 during growth in LB. As expected, mCexE was only present in the strain encoding the wild-type *aatD*
205 sequence whereas only unmodified uCexE was detected in mutants harbouring the E217A, K276A,
206 E286A, or C325A (Fig 6D). These observations suggest that the mechanism of protein acylation by
207 AatD and Lnt are similar.

208 **An N-terminal glycine is required for acylation**

209 Due to the homology and functional similarity between AatD and Lnt, a likely site for acylation
210 of CexE is the N-terminal amino acid immediately after the signal sequence. SignalP predictions
211 suggested that CexE possesses a Sec-dependent signal sequence cleaved by Signal Peptidase I between
212 alanine at position 19 and glycine at position 20 (Fig. 7A and Supplementary Table S2). N-terminal
213 amino acid sequencing of uCexE revealed the amino acid sequence GGGNS confirming that glycine at
214 position 20 formed the N-terminal amino acid of the processed protein. However, bioinformatic

215 analyses of the amino acid sequences of 224 distinct CexE homologs failed to identify in their signal
216 sequences the presence of a 'lipobox' (Supplementary Fig S4); lipoboxes are required for recognition
217 of lipoproteins by the Lol system and thus acylation by Lnt (24). Further analyses revealed limited
218 sequence identity between the signal sequences of the 224 CexE homologs indicating that an alternative
219 lipobox is not present. Moreover, neither CexE nor any other homolog possessed a cysteine residue
220 adjacent to the signal sequence; for bacterial lipoproteins the N-terminal cysteine of the mature protein
221 is the target of Lnt mediated acylation. Since CexE does not contain features of a typical bacterial
222 lipoprotein it must be recognised and acylated by AatD in a manner different to the acylation of
223 lipoproteins by Lnt.

224 Further scrutiny of the amino acid sequences of all CexE homologs revealed that for all CexE
225 and Aap proteins the first five amino acids immediately after the signal peptide cleavage site are a
226 conserved mix of glycine and serine residues (Fig 7B). The presence of an invariant N-terminal glycine
227 residue immediately downstream of the predicted signal sequence suggested that this might be the site
228 of acylation. To test this hypothesis, we mutated each of the N-terminal residues of CexE to glutamic
229 acid and observed the effect on acylation by western blotting. The amount of total fluorescence of the
230 secondary antibody bound to primary polyclonal anti-CexE antibody was measured and compared for
231 each mutant. The percentage of acylated mCexE was calculated for each mutant. The migration of the
232 G20E mutant was similar to uCexE indicating that the change to glutamic acid completely abolished
233 acylation (Fig 7C). In addition, the acylation of the G21E mutant was significantly reduced (p-value =
234 0.0013). However, there was no significant reduction in acylation of the G22E, N23E or S24E mutants
235 (Fig 7C). These data indicate that the acylation of CexE is highly dependent on the N-terminal glycine
236 at position 20 and to a lesser extent on the second glycine at position 21.

237 **Mass spectrometry analysis of CexE acylation**

238 Mutation of the conserved N-terminal glycine can abolish acylation of CexE, and Lnt acylates
239 the N-terminal amine of the signal peptidase processed lipoprotein, therefore we hypothesised that AatD
240 would acylate the N-terminal amine of G20. To test this, we first purified his-tagged mCexE and uCexE
241 by nickel affinity chromatography. Both forms of CexE were subjected to trypsin digestion followed

242 by LC-MS/MS. A peak was observed in the HLPC trace of mCexE that was not present in uCexE (Fig
243 8A). The mass spectrometry of that peptide identified the five N-terminal residues of CexE with an
244 addition of 238 Da, which is equal to that expected to a single 16 carbon fatty acid addition (Fig 8B)
245 confirming the modification of CexE by the addition of an acyl chain onto the N-terminal glycine.

246 **Heterologous acylation by AatD**

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

261 **Discussion**

262 The current dogma for Gram-negative bacterial lipoproteins states that proteins are triacylated
263 on an N-terminal cysteine through the action of three enzymes Lgt, Lsp and Lnt (25–27). To our
264 knowledge there are no examples of bacterial proteins where acylation occurs on N-terminal glycine
265 residues. While N-terminal glycine acylation by N-myristoyltransferase is present in life it is restricted
266 to eukaryotes. Examples of bacterial N-myristoylation rely on eukaryotic proteins produced in *E. coli*
267 or the injection of myristoylatable proteins into eukaryotic cells by type 3 secretion systems (28,29).
268 We have not found any reports of bacterial N-terminal glycine palmitoylation. Indeed, to our knowledge

269 this is the first report of enzyme mediated N-palmitoylation in nature. Therefore, we believe our
270 demonstration that Aap and CexE are a new class of bacterial lipoproteins is novel. The recent
271 publication by Belmont-Monroy *et al.* 2020 independently identified AatD as an acyltransferase of Aap
272 and CexE.

273 The Aat system is, to our knowledge, the only secretion system that acylates and secretes a
274 lipoprotein substrate to the cell surface. While surface exposed lipoproteins have been reported for *E.*
275 e.g. Lpp, RcsF, and SslE (6–8), these lipoproteins are acylated by Lgt and Lnt on an N-terminal cysteine
276 in the typical manner and translocated to the outer membrane by the Lol pathway. Similarly, *Neisseria*
277 spp. decorate their cell surfaces with secreted lipoproteins that are acylated by Lgt and Lnt on an N-
278 terminal cysteine and trafficked to the outer membrane by the Lol system. In the latter case an outer
279 membrane protein termed Slam is required for translocation of the protein to the cell surface (30). In
280 contrast, the Aat system encodes a novel and specialised N-acyltransferase, AatD. To allow secretion
281 of the novel lipoprotein CexE the Aat system has combined proteins of the T1SS and the Lol trafficking
282 pathway. Not only does the Aat system secrete a novel class of lipoproteins from the periplasm to the
283 surface of the cell but also it acylates that protein itself. For this reason, we believe that reclassification
284 of the Aat system as the first example of the Lipoprotein Secretion System (LSS) is warranted. We
285 propose that the LSS would remain as a sub-class of the T1SS to emphasise its similarities to the T1SS.

286 In Fig 10, we propose our model for Aat mediated secretion. First, the Sec machinery secretes
287 preCexE into the periplasm, resulting in cleavage of the signal peptide by Signal Peptidase I to form
288 uCexE (14). Once in the periplasm the N-terminal glycine of uCexE is acylated by AatD to create
289 mCexE. Through this acylation event mCexE then associates with the inner membrane. We believe this
290 to be first step for secretion by the Aat system since CexE acylation by AatD is independent of the other
291 members of the Aat system. In addition, the deletion of any of the *aatPABC* genes does not result in a
292 loss of CexE acylation. We propose that the acylation of CexE is required to enter the AatPABC tunnel.
293 Due to the homology to LolCDE, we believe that mCexE is extracted from the inner membrane by
294 AatPC. Finally, AatA, AatB and AatPC form a channel for mCexE secretion in a manner analogous to
295 a typical T1SS. Subsequently, mCexE inserts into the outer leaflet of the outer membrane.

296 Belmont-Monroy *et al.* 2020 suggested that Lnt is able to acylate Aap. However, in our hands
297 there was no evidence of CexE acylation in the absence of AatD. Given that Aap and CexE lack the
298 traditional lipobox of bacterial lipoproteins that would be recognised by the Lol system, that the signal
299 sequence of all Aap and CexE peptides were predicted to be cleaved by Signal Peptidase I, which has
300 been confirmed experimentally for CexE (14), and we were able to reconstruct AatD acylation in *E.*
301 *coli* BL21(DE3) which does not naturally encode AatD or CexE but does encode Lnt, we do not believe
302 that Lnt plays a role in CexE acylation. Indeed, our demonstration of acylation of a heterologous protein
303 with a single glycine residue (Fig 10) suggests that AatD is a specific N-acyltransferase that has the
304 potential for exploitation for production of novel acylated peptides in *E. coli*. The addition of an acyl
305 chain can improve the half-life of peptide drugs like insulin (31). Since we have shown that it is possible
306 to reconstitute AatD acylation in a laboratory strain of *E. coli* there are further biotechnological
307 applications to be explored.

308 Due to increasing levels of antibiotic resistance in pathogenic organisms there is a requirement
309 for new protein targets. The conservation of this system within pathogenic bacterial strains represents
310 a possible new drug target. Previous studies have shown that deletion of *aatD*, *aatC*, or *aap*/*cexE*
311 reduces bacterial colonisation and disease (11,12). Thus, chemicals targeting the Aat system have the
312 potential to prevent or perhaps reduce the severity of bacterial disease. As the Aat system does not
313 appear to be present in non-pathogenic strains this would be an ideal target to prevent off-target effects.
314 As only pathogenic bacteria would be affected by the inhibition of the Aat system, there would be no
315 risk of evolving resistant alleles in non-pathogenic populations.

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

322 .

323 **Materials and methods**

324 **Bacterial cultivation**

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

336 **Molecular techniques**

337 Primers for PCR can be found in Table S4. Plasmids used in this study are detailed in Table S5. Unless
338 otherwise stated all primers were used at a concertation of 10 µM. Q5® High-Fidelity 2X Master Mix
339 (NEB) was used for cloning or mutagenesis. MyTaq™ Red Mix (Bioline) was used in all other cases.
340 The *cexE* gene from ETEC H10407 was amplified using CcexE-F and CcexE-R primers. Both the
341 vector, pET26b(+) (Novagen), and the insert were cut with NdeI (NEB) and XhoI (NEB) using the
342 CutSmart® (NEB) protocol. The digested vector was treated with Antarctic Phosphatase (NEB). T4
343 DNA Ligase (NEB) was used to ligate the insert DNA to the vector DNA. The ligation mixture was
344 transformed into NEB® 5-alpha Competent E. coli (High Efficiency). Recovered cells were plated on
345 LBA supplemented with 100 µg/ml kanamycin and incubated at 37ºC overnight. The *cexE* gene and
346 573 bp upstream of the *cexE* gene, which included the two CfaD binding sites (14), was amplified by
347 PCR. The reverse primer included a 6His tag. The resulting amplified DNA was digested with SphI and
348 BamHI and ligated into pACYC184. Successful ligations were selected for as described above on LBA

349 supplemented with 35 µg/ml chloramphenicol. Point mutations were constructed using the
350 QuickChange II method (Aglient). The *aat* genes and *cexE* were disrupted in ETEC H10407 as
351 previously described (32). The plasmid pDOC-K was used as the source of the kanamycin resistance
352 cassette (33). The *cexE*-mCherry fusion and the *aatD* gene from ETEC H10407 were synthesised by
353 GenScript. Deletions of the *cexE* gene from the plasmid encoding the CexE-mCherry fusion were
354 constructed as previously described (34). Plasmid sequences were confirmed by Sanger sequencing.

355 **Bioinformatic analysis of the Aat system**

356 The protein sequences of the Aat proteins from ETEC H10407 were used to search the NCBI non-
357 redundant protein sequences database using PSI-BLAST (35). Strains encoding a complete Aat system
358 on the same nucleotide accession were used to assess the distance between *aat* genes. Distant homologs
359 of the Aat proteins were identified using HMMER (36). An HMM was created for each Aat protein
360 using the Aat protein sequences identified by PSI-BLAST. These models were used to search the
361 UniprotKB or Swissprot databases (37). Protein sequences were aligned using Clustal Omega (38).
362 RAxML was used for the construction of phylogenetic trees (39). Trees were drawn using iTOL (40).

363 **CexE purification for antibody production**

364 BL21(DE3) was transformed with pET26b-*cexE*. An overnight culture of BL21(DE3) pET26b-*cexE*
365 was used to inoculate 2 L of LB supplemented with 100 µg/ml kanamycin. The culture was grown at
366 37°C with aeration to an OD_{600} of 0.4. The culture was moved to 20°C incubator for 30 min prior to
367 induction. CexE production was induced with 50 µM Isopropyl β-D-1-thiogalactopyranoside (IPTG)
368 (Sigma) and incubated overnight at 20°C with aeration. After overnight growth, cells were harvested
369 by centrifugation at 6,000 x *g* for 10 min at 4ºC. The cell pellet was resuspended in ice-cold binding
370 buffer (50 mM NaP (77:33 ratio of Na_2HPO_4 to NaH_2PO_4), 500 mM NaCl, 20 mM imidazole, 0.5 mM
371 TCEP at pH 7.3). Cells were lysed using EmulsiFlex®-C3 (Avestin). Cellular debris and intact cells
372 were removed by centrifugation at 10,000 x *g* for 10 min at 4ºC. The cellular membrane components
373 were removed by centrifugation at 75,000 x *g* for 1 h at 10ºC. The supernatant was applied to a
374 HisTrap™ HP 5 ml column (GE Healthcare Life Sciences) overnight at 4°C. The column was washed

375 with 5 column volumes of binding buffer then the protein was eluted in 5 ml fractions with elution
376 buffer (50 mM NaP, 500 mM NaCl, 500 mM imidazole, 0.5 mM TCEP at pH 7.3). Samples containing
377 purified protein were concentrated using Vivaspin® 6 5,000 MWCO columns (Sartorius Stedim).
378 Protein was buffer exchanged using membrane filtration into 50 mM NaP, 250 mM NaCl and 0.5 mM
379 TCEP at pH 7.3. Purified CexE protein was used to produce primary antibodies by Eurogentec using
380 the 28-day speedy protocol. Polyclonal antibodies against CexE were concentrated prior to use for the
381 detection of CexE.

382 **Tris-Tricine SDS-PAGE and western blotting**

383 Protein samples were separated using tris-tricine SDS-PAGE as previously described (41). In brief,
384 protein samples were mixed with sample buffer (4% SDS, 2.5% 2-mercaptoethanol, 7.5% glycerol,
385 0.01% bromophenol blue, 35 mM Tris (pH 7.0)) and separated on 10% tris-tricine SDS-PAGE gels.
386 Gels were stained with Coomassie brilliant blue or transferred to nitrocellulose for western blotting.
387 Nitrocellulose membranes were covered in 2% BSA (20 g BSA, 2.42 g Tris-base (pH 8.4), 8 g NaCl
388 per L). The primary antibody was diluted in 2% BSA at the following concentrations: Aap 1 in 5,000
389 (Sheikh et al., 2002); CexE 1 in 2,000 (this study); and β subunit of RNAP 1 in 10,000 (*E. coli* RNA
390 Polymerase beta Monoclonal Antibody, BioLegend). Primary antibodies were incubated overnight at
391 4°C with agitation. Membranes were washed three times in TBST (2.42 g Tris-base, 8 g NaCl, 0.1%
392 Tween-20, pH 8.4 in 1 L) for 5 min at room temperature with agitation prior to incubation with
393 secondary antibody. Anti-Rabbit (IRDye 800CW Goat anti-Rabbit IgG, Li-Cor) or anti-mouse (IRDye
394 680LT Goat anti-Mouse IgG, Li-Cor) secondary antibodies were used to detect primary antibody
395 binding. Secondary antibodies were diluted 1 in 15,000 in 2% BSA and incubated for a minimum of 1
396 h at room temperature with agitation. Membranes were washed 4 times with TBST for 5 min at room
397 temperature with agitation. Secondary antibody fluorescence was detected using the Odessy CLx
398 imaging system.

399 **Inhibition of SecA by sodium azide**

400 Two cultures of ETEC H10407 pCfaD were inoculated from the same overnight culture. The cells were
401 both grown at 37°C for 90 min with aeration. WCL samples of each culture were taken. L-arabinose

402 was added to both cultures to a final concentration of 0.2%. Sodium azide was added to a final
403 concentration of 3 mM to one of the cultures. Both cultures were grown for 2 h at 37°C with aeration.
404 WCL samples of each culture were taken and separated by tris-tricine SDS-PAGE. CexE was detected
405 by western blotting.

406 **Proteomic analysis**

407 Cultures were grown under the conditions previously described for CexE production in ETEC H10407
408 pCfaD. Cells were separated from the culture supernatant by centrifugation at 8,000 x *g* for 10 min at
409 4ºC. The culture supernatant was filtered using Millex-GP Syringe Filter Unit, 0.22 µm,
410 polyethersulfone (Merck) then cooled on ice. Ice-cold 100% TCA was added to a final concentration
411 of 20%. Samples were incubated on ice for 30 minutes. Proteins were pelleted by centrifugation. The
412 supernatant was removed and discarded from the sample. The pellet was washed twice with 1 ml ice-
413 cold 100% methanol, proteins were collected with centrifugation at 21,000 x *g* for 15 min at 4ºC
414 between each wash. The supernatant was removed and residual methanol was evaporated by incubating
415 the sample at 60ºC for 10 min. The pellet was resuspended in 50 µl tris-tricine sample buffer. If a colour
416 change to yellow occurred saturated tris-base was added until the original colour returned. Samples
417 were analysed by tris-tricine SDS-PAGE.

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

426 **CuAAC**

427 Overnight cultures harbouring plasmids encoding his tagged proteins of interest to be labelled were
428 used to inoculate 25 ml LB supplemented with the relevant antibiotics to an OD_{600} of 0.05. Cultures
429 were grown for 1.5 h at 37°C. 17-ODYA dissolved in DMSO was added to a final concentration of 20
430 µM and protein production induced. An equal volume of DMSO was added as a negative control. Cells
431 were collected by centrifugation and resuspended in 600 µl of 50 mM NaP (pH 7.0), 150 mM NaCl,
432 0.1% Triton X-100, and 1 mM PMSF. Cells were lysed by sonication for 15 min using the bioruptor on
433 a 30 s on, 30 s off cycles. Insoluble material was removed by centrifugation at and the supernatant was
434 retained. His tagged proteins were isolated from the supernatant using Dynabeads™ His-Tag Isolation
435 and Pulldown (Invitrogen) as per the manufacturer's instructions. Imidazole was removed by dialysis.
436 Protein sample concentrations were measured with Pierce™ BCA Protein Assay Kit (Thermo
437 Scientific™). A master mix was prepared of 0.2 mM Cy5-azide (Sigma-Aldrich) or Alexa Fluor™ 488
438 azide (Invitrogen), 0.2 mM tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA), and 2 mM
439 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
440 each normalised protein sample. Sodium ascorbate was added to a final concentration of 5 mM. Samples
441 were incubated at 37ºC for 1 h with agitation. Protein was precipitated using chloroform-methanol
442 extraction. To each sample 400 µl MeOH, 150 µl chloroform, and 300 µl H_2O were added and vortexed
443 for 30 seconds. Samples were centrifuged at max speed for 2 min. The top layer was removed and 400
444 µl MeOH was added and samples were centrifuged again at max speed for 2 min. The supernatant was
445 removed, and the pellet was washed twice with 400 µl MeOH then dried. Pellets were resuspended in
446 20 µl 1x tris-tricine sample buffer and separated by tris-tricine SDS-PAGE. Gels were incubated in
447 fixative (10% acetic acid, 50% methanol) prior to fluorescence detection then stained with Coomassie
448 Brilliant Blue.

449 **Heterologous protein acylation**

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

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

455 **Mass spectrometry analysis of CexE acylation**

456 CexE with a 6 His tag was isolated from a cexE mutant and a cexE aatD double mutant of ETEC H10407
457 pCfaD. The isolated CexE proteins were separated on a tris-tricine gel and stain with Coomassie
458 brilliant blue. The band corresponding to uCexE and mCexE were excised and subjected to trypsin-
459 LysC digestion. The tryptic peptide extracts were analyzed by nanoHPLC/MS MS/MS on an Eksigent,
460 Ekspert nano LC400 uHPLC (SCIEX, Canada) coupled to a Triple TOF 6600 mass spectrometer
461 (SCIEX, Canada) equipped with a PicoView nanoflow (New Objective, USA) ion source. 5 µl of each
462 extract was injected onto a 5 mm x 300 µm, C18 3 µm trap column (SGE, Australia) for 5 min at
463 10uL/minute. The trapped tryptic peptide extracts were then washed onto the analytical 75 µm x 150
464 mm ChromXP C18 CL 3 µm column (SCIEX, Canada) at 400 nL/min and a column temperature of
465 45°C. Solvent A consisted of 0.1% formic acid in water and solvent B contained 0.1% formic acid in
466 acetonitrile. Linear gradients of 2-40% solvent B over 60 min at 400 nL/minute flow rate, followed by
467 a steeper gradient from 40% to 90% solvent B in 5 min, then 90% solvent B for 5 min, were used for
468 peptide elution. The gradient was then returned to 2% solvent B for equilibration prior to the next
469 sample injection. The ionspray voltage was set to 2600 V, declustering potential (DP) 80 V, curtain gas
470 flow 30, nebuliser gas 1 (GS1) 30, interface heater at 150°C. The mass spectrometer acquired 50 ms
471 full scan TOF-MS data followed by up to 30 100 ms full scan product ion data, with a rolling collision
472 energy, in an Information Dependant Acquisition, IDA, mode for protein identification and peptide
473 library production. Full scan TOF-MS data was acquired over the mass range 350-1800 and for product
474 ion MS/MS 100-1500. Ions observed in the TOF-MS scan exceeding a threshold of 200 counts and a
475 charge state of +2 to +5 were set to trigger the acquisition of product ion, MS/MS spectra of the resultant
476 30 most intense ions. The data was acquired and processed using Analyst TF 1.7 software (ABSCIEX,
477 Canada). Protein identification was carried out using Protein Pilot 5.0 for database searching.

478

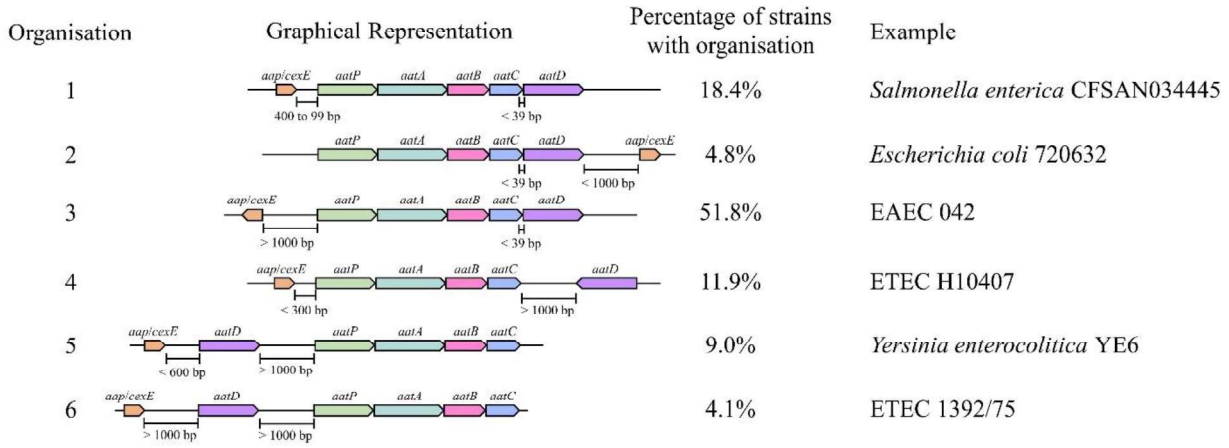


Fig 1. Organisation of the Aat operon. Unique Aat amino acid sequences were detected using PSI-BLAST. These sequences were using to identify the strains that encoded them in the NCBI identical protein groups. As the *aat* genes were present on contigs of whole genome sequencing projects it was not possible 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.

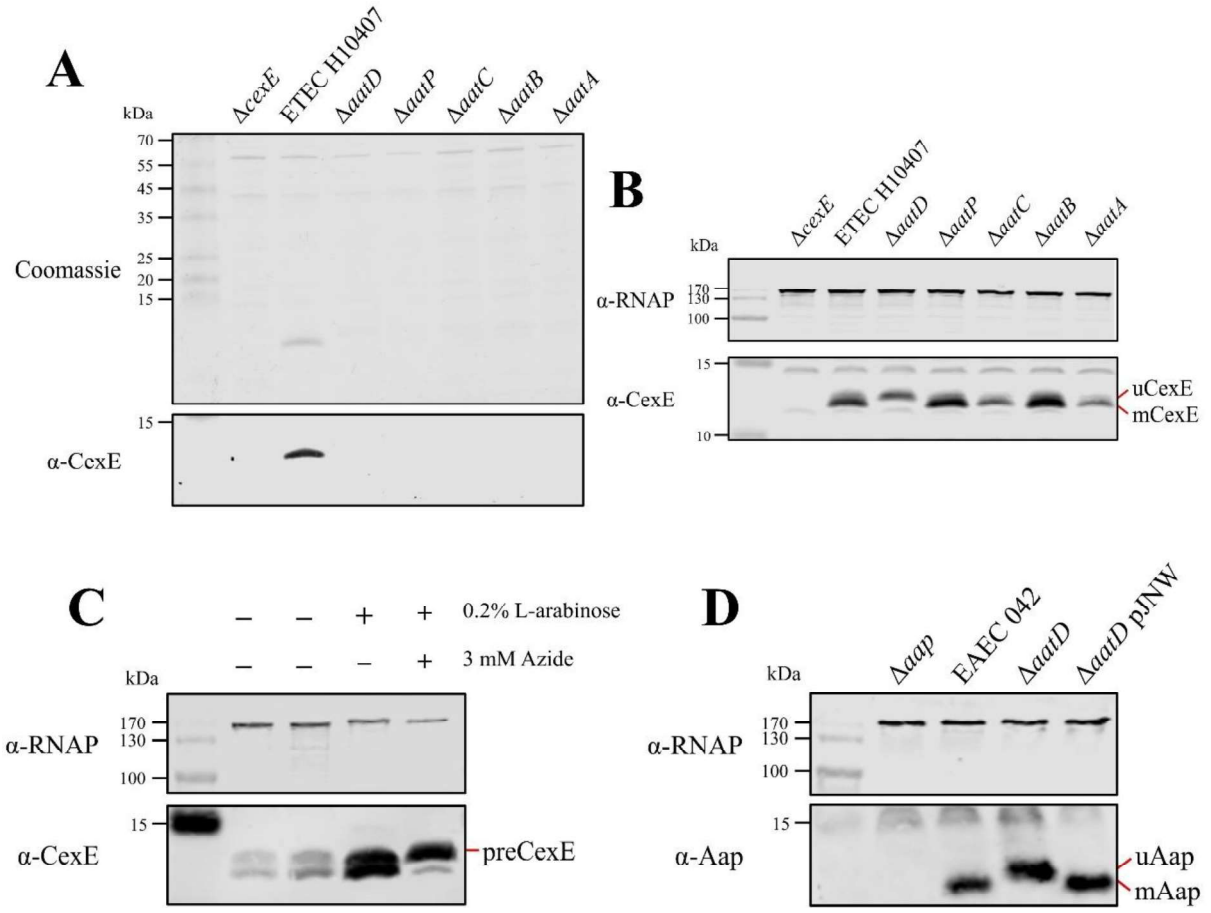


Fig 2. AatD post-translational modification of Aap and CexE. (A) Culture supernatant of ETEC H10407 and *aat* mutants harbouring pCfaD grown in LB supplemented with L-arabinose. Cells were removed from the culture supernatant and remaining protein precipitated with trichloroacetic acid. Protein samples 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.

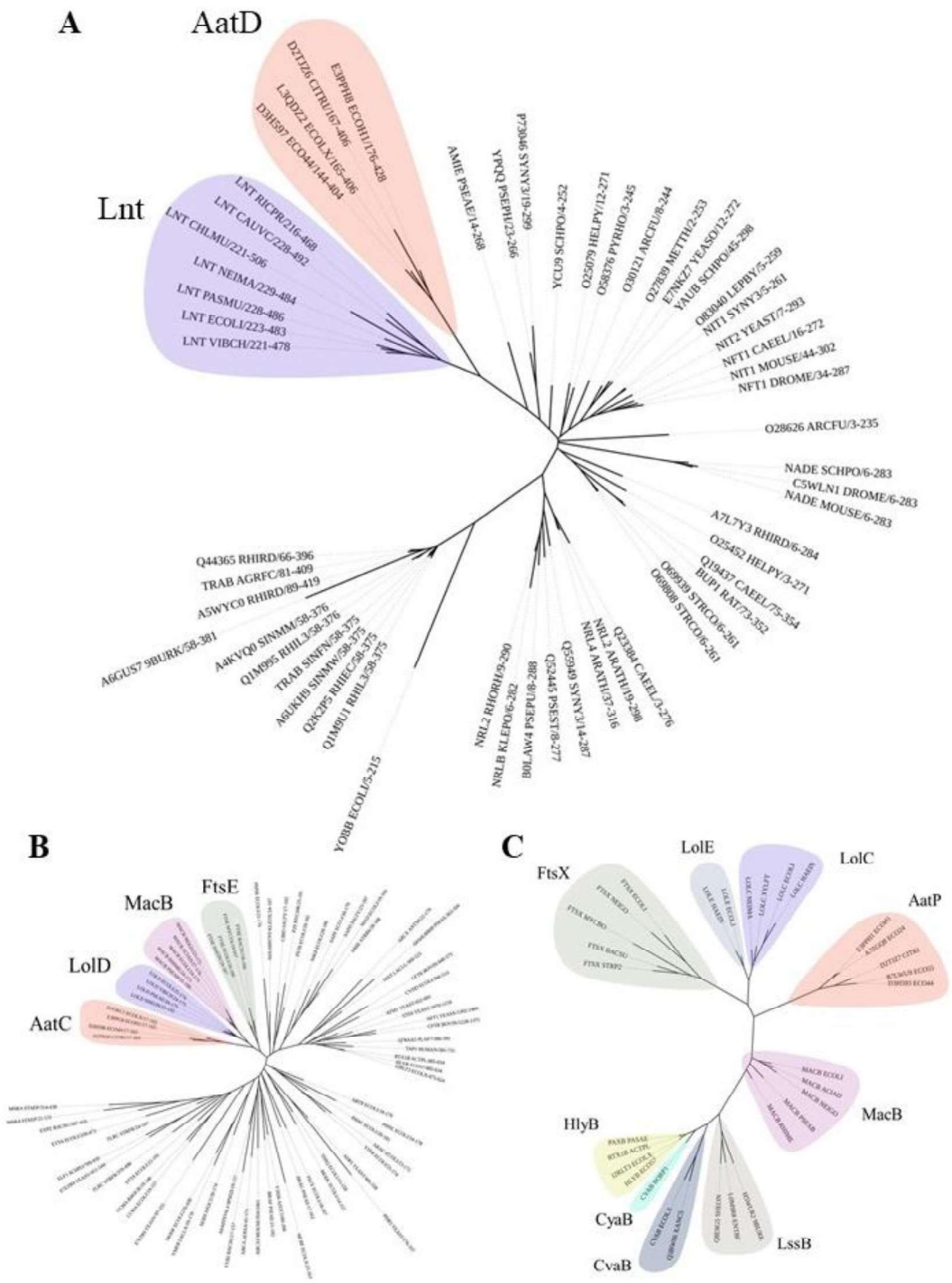
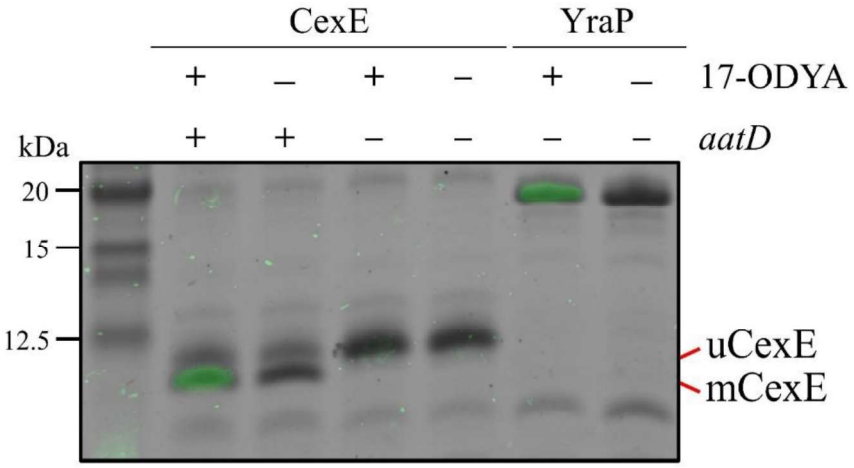


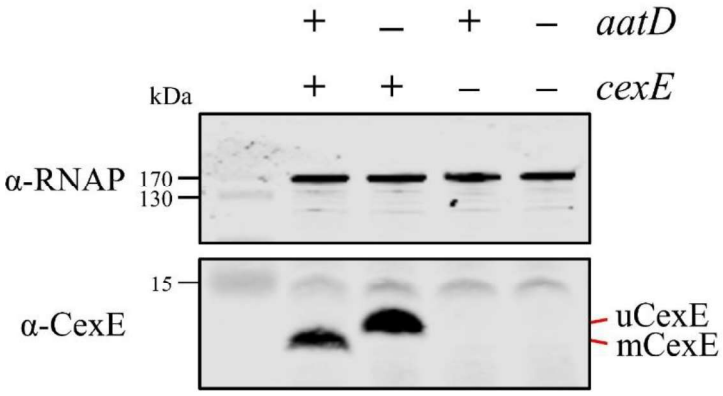
Fig 3. Phylogenetic analysis of AatD, AatP and AatC. Sequences were aligned using Clustal omega and the tree was calculated using RAxML. **(A)** The C-N hydrolase domains of C-N hydrolase family seed sequences of the were aligned to the C-N hydrolase domains of AatD sequences from ETEC H10407, EAEC 042,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



509

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

516



517

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

523

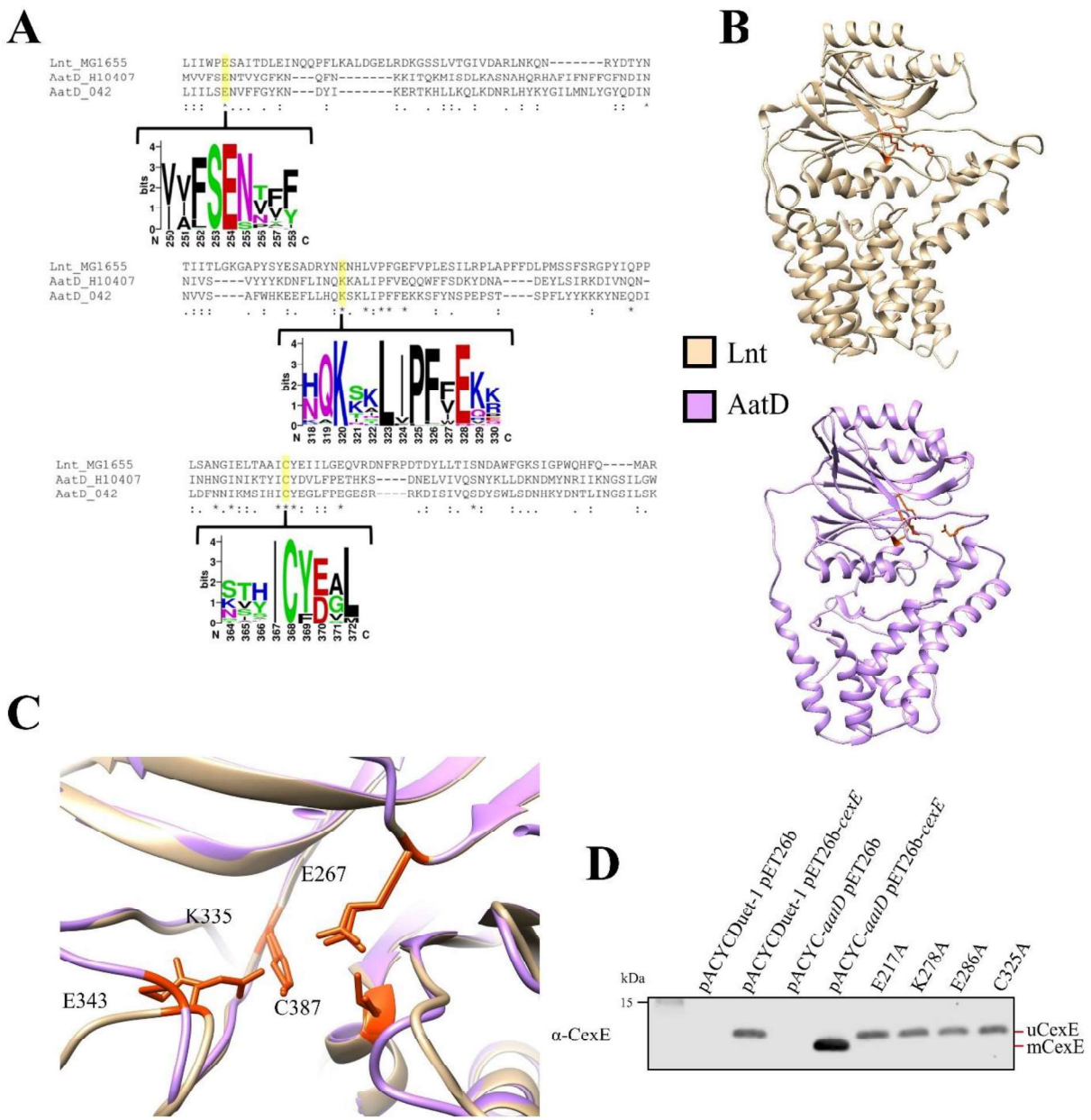


Fig 6. Catalytic residues of AatD. (A) AatD from EAEC 042 and ETEC H10407 were aligned with Lnt from MG1655. AatD sequences identified by PSI-BLAST were aligned and used to create a Weblogo. The catalytic residues of the C-N hydrolase family are highlighted. **(B)** The structure of Lnt compared to the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

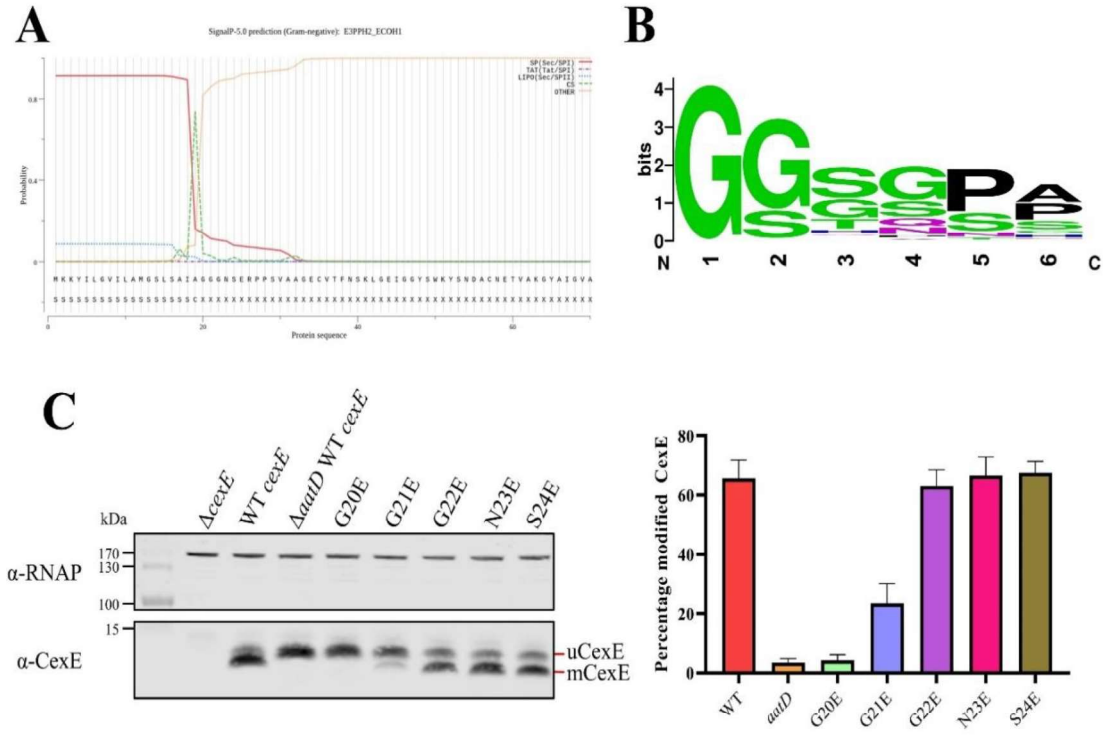
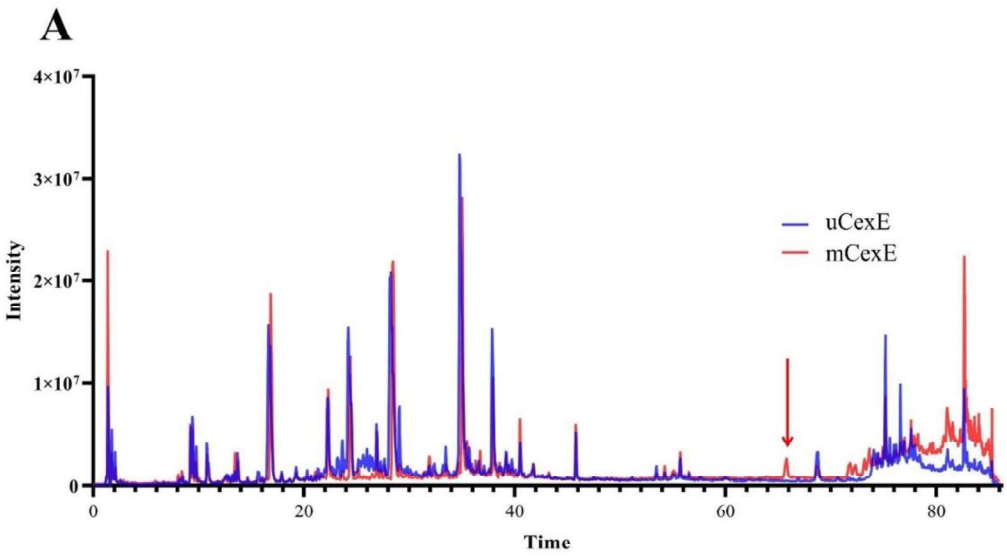
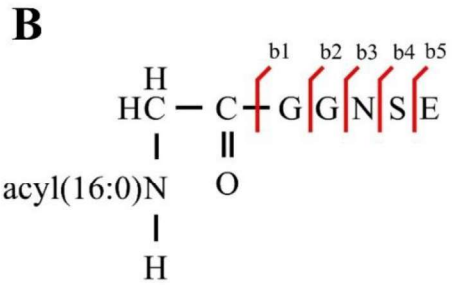
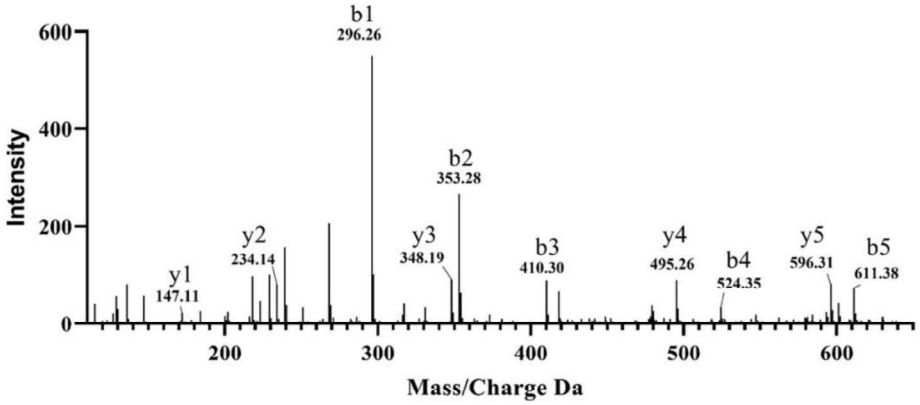


Fig 7. Single site substitution of the 5 N-terminal amino acids of CexE. (A) SignalP result of CexE sequence from ETEC H10407 (E3PPH2_ECOH1) **(B)** Weblogo of the 5 N-terminal residues of Aap and CexE sequences post Sec signal sequence cleavage. **(C)** ETEC H10407 *cexE* mutants transformed with pACYC184 (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)



541

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

546

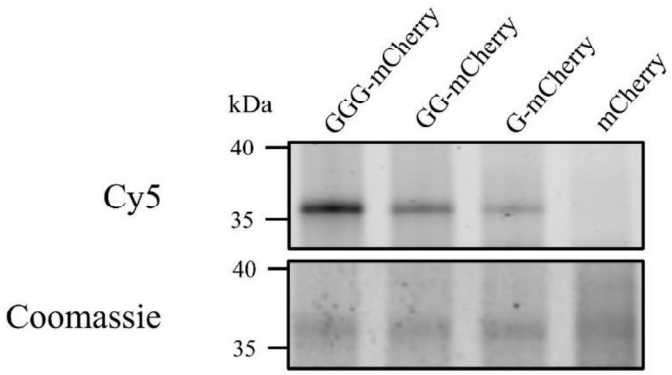
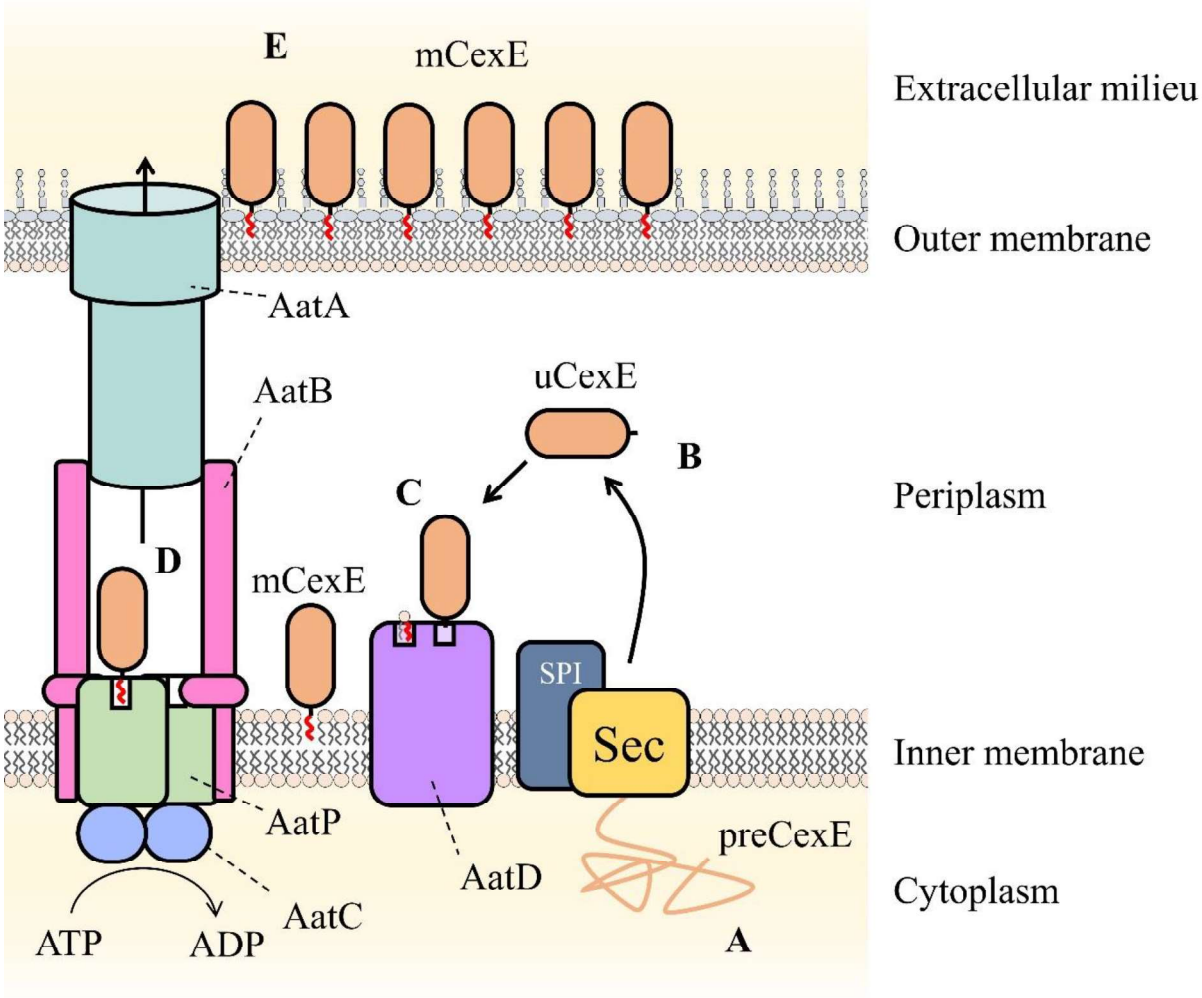


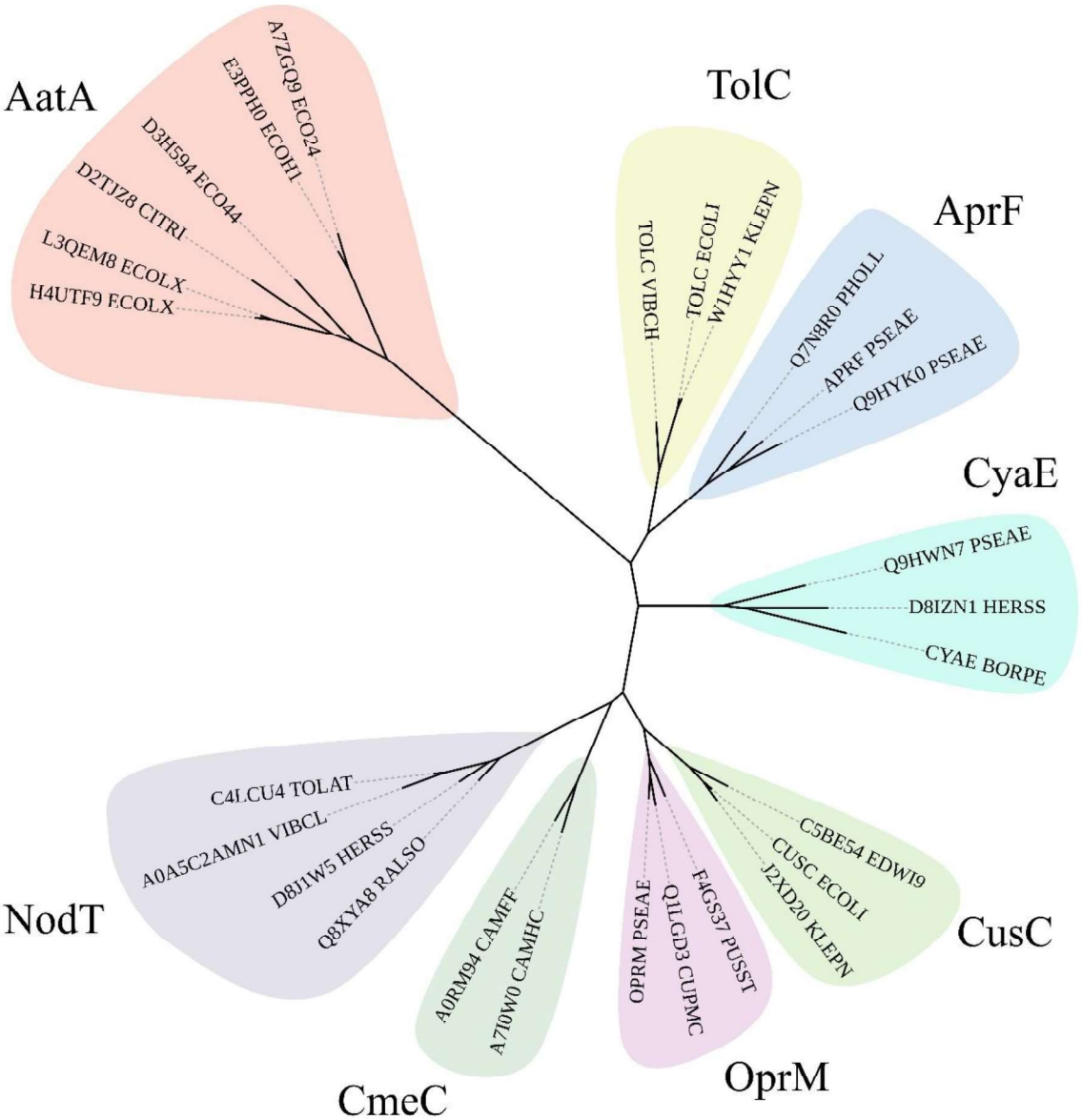
Fig 9. Heterologous acylation by AatD plasmid encoding mCherry with the CexE signal sequence followed by none, one (G), two (GG) or three (GGG) glycine residues post signal sequence was produced in an ETEC H10407 *cexE* mutant grown in the presence of 17-ODYA. The mCherry proteins were isolated using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



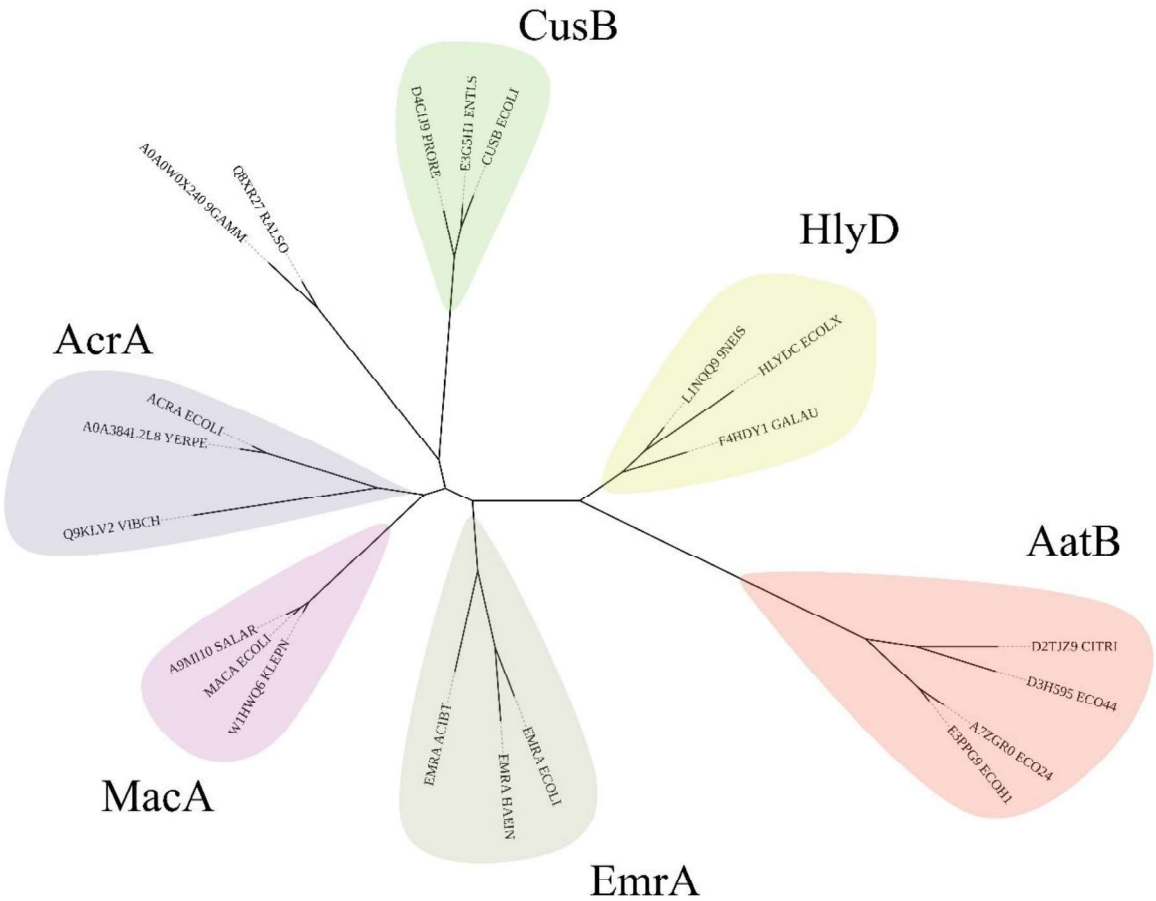
554

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

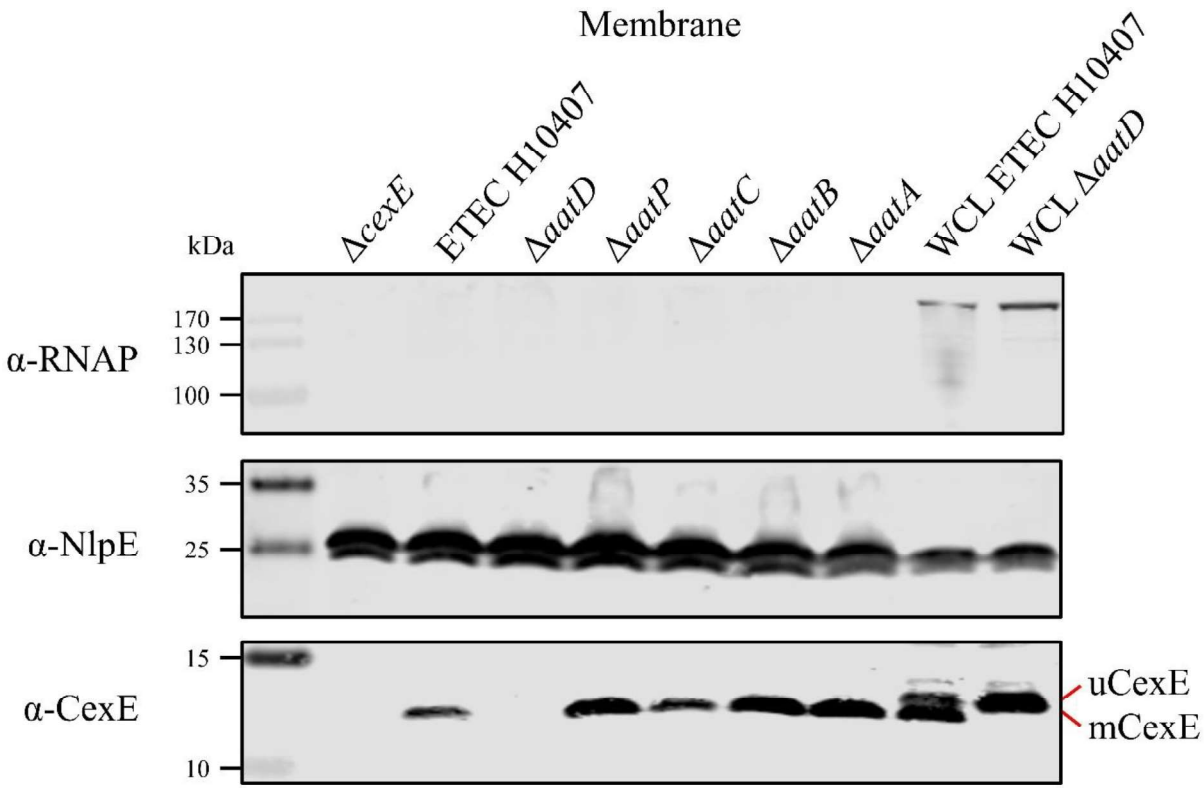
563



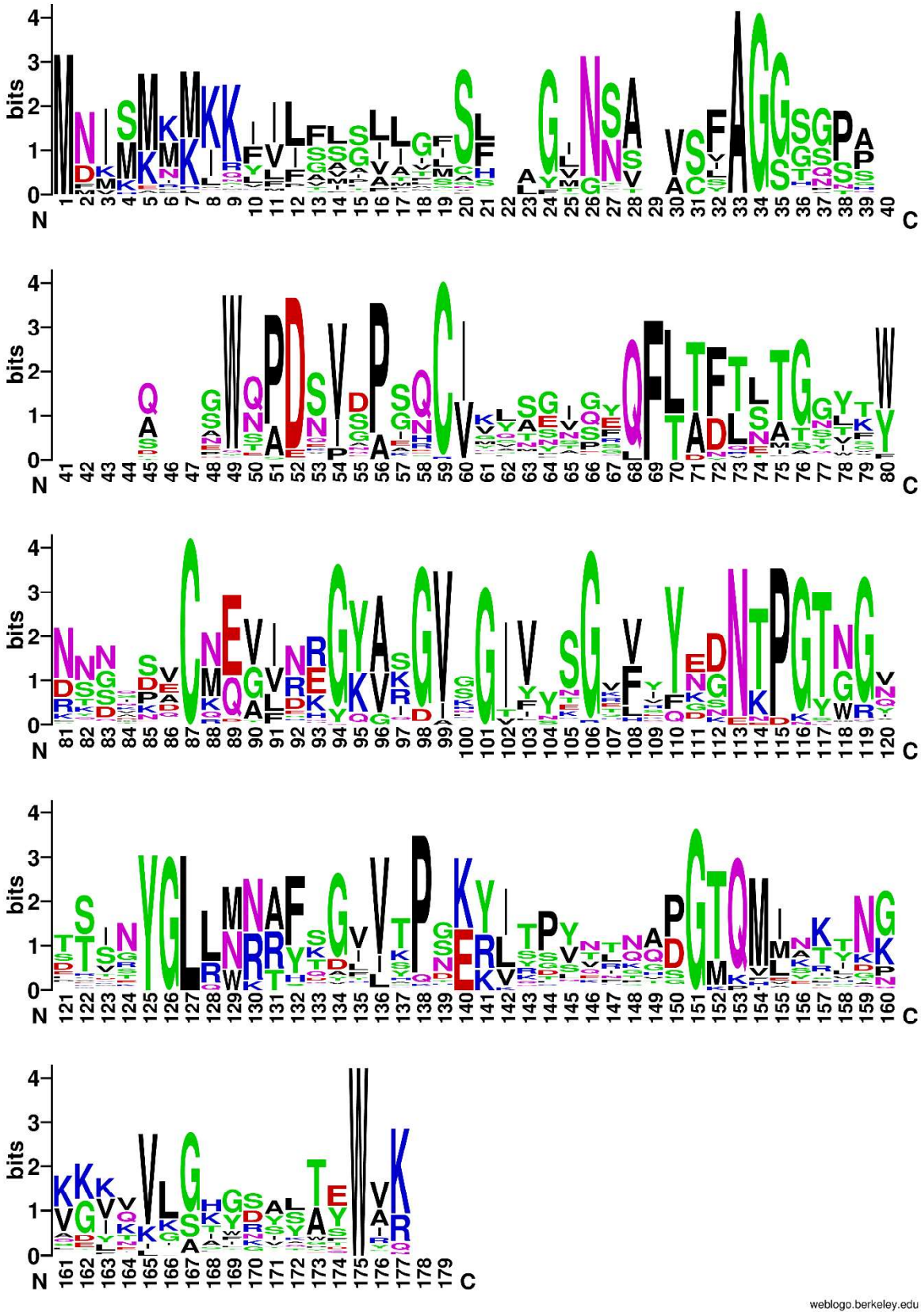
Supplementary Fig S1. **Phylogram of AatA, TolC and other OMP associated with T1SS and efflux pumps.** The HMM of AatA proteins was used to confirm AatA as a homolog of TolC and OMP proteins. Representative sequences of secretion associated OMPs (AprF, CyaE and TolC) and efflux associated OMPs (CusC,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



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



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



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

587 **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)

588

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

Name	Sequence	Description
C*cexE*-F	GGCGGCCATATGAAAAAATATATATTAGGTGT	Cloning *cexE* from ETEC H10407 into pET26(b)
C*cexE*-R	CGAAGCCTCGAGTTTATACCAATAAGGGGTGT	Cloning *cexE* from ETEC H10407 into pET26(b)
GD*aatP*-F	GAATTCAAGCTTGCAAAGCGTATTGTTGGTGCAGGCTTGTATAAGTTATATGGACCGGTCAATTGGCTGGAG	Forward primer for generating kanamycin cassette and disruption of *aatP* gene in ETEC H10407
GD*aatP*-R	GAGCTCCATATGAATATAAATATAATCTTCATGAAAATCTTTCTTTTATTAAAATATCCTCCTTAGTTCC	Reverse primer for generating kanamycin cassette and disruption of *aatP* gene in ETEC H10407
GD*aatA*-F	GAATTCAAGCTTGATTTATCAATCTTAATAAAAGAAAGATTTTCATGAAGAGACCGGTCAATTGGCTGGAG	Forward primer for generating kanamycin cassette and disruption of *aatA* gene in ETEC H10407
GD*aatA*-R	GAGCTCCATATGTATATAATTCAATTTCATTTTCCTTTTTTATTAACTCTCAATATCCTCCTTAGTTCC	Reverse primer for generating kanamycin cassette and disruption of *aatA* gene in ETEC H10407
GD*aatB*-F	GAATTCAAGCTTGAATTATATAAAATGCATATGTTTTTTATTGGTTAGTTTGACCGGTCAATTGGCTGGAG	Forward primer for generating kanamycin cassette and disruption of *aatB* gene in ETEC H10407
GD*aatB*-R	GAGCTCCATATGATCAATACTTAATTTAATCATGTAGTGTTATCTCAAATGCTGAATATCCTCCTTAGTTCC	Reverse primer for generating kanamycin cassette and disruption of *aatB* gene in ETEC H10407
GD*aatC*-F	GAATTCAAGCTTCTTGGATACAGCATTTGAGATAACACTACATGATTAAATTAGACCGGTCAATTGGCTGGAG	Forward primer for generating kanamycin cassette and disruption of aatC gene in ETEC H10407
GD*aatC*-R	GAGCTCCATATGAAATCACATAAAATTTTTATTAAAAGGATATAACCTTCATAATATCCTCCTTAGTTCC	Reverse primer for generating kanamycin cassette and disruption of *aatC* gene in ETEC H10407
GD*aatD*-F	GAATTCAAGCTTTCGTATATTGGCATTAAATTATCTTTTGAATTGATTAATGGACCGGTCAATTGGCTGGAG	Forward primer for generating kanamycin cassette and disruption of aatD gene in ETEC H10407
GD*aatD*-R	GAGCTCCATATGACCAAATGATGTATATGGTTTATACATCAATGACAAAAAAAATATCCTCCTTAGTTCC	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

590

591 **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

592

593 **References**

594

595 1. Troeger C, Forouzanfar M, Rao PC, Khalil I, Brown A, Reiner RC, et al. Estimates of global,
596 regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic
597 analysis for the Global Burden of Disease Study 2015. Lancet Infect Dis. 2017;17(9):909–48.
598 2. Ericsson CD. Traveler's Diarrhea: Prevention and Self-Treatment. Int J Antimicrob Agents.
599 2017;21(2):116–24.
600 3. Dev IK, Harvey RJ, Ray PH. Inhibition of prolipoprotein signal peptidase by globomycin. J
601 Biol Chem. 1985;260(10):5891–4.
602 4. Masao Tokunaga, Hiroko Tokunaga, Wu HC. Post-translational modification and processing
603 of outer membrane prolipoproteins in Escherichia coli. Mol Cell Biochem. 1984;60(1):5–15.
604 5. Gupta SD, Gan K, Schmid MB, Wu HC. Characterization of a temperature-sensitive mutant of
605 Salmonella typhimurium defective in apolipoprotein N-acyltransferase. J Biol Chem.
606 1993;268(22):16551–6.
607 6. Cowles CE, Li Y, Semmelhack MF, Cristea IM, Silhavy TJ. The free and bound forms of Lpp
608 occupy distinct subcellular locations in Escherichia coli. Mol Microbiol. 2012;79(5):1168–81.
609 7. Konovalova A, Perlman DH, Cowles CE, Silhavy TJ. Transmembrane domain of surface-
610 exposed outer membrane lipoprotein RcsF is threaded through the lumen of β -barrel proteins.
611 Proc Natl Acad Sci USA. 2014;111(41):E4350–8.
612 8. Baldi DL, Higginson EE, Hocking DM, Praszkier J, Cavaliere R, James CE, et al. The type II
613 secretion system and its ubiquitous lipoprotein substrate, SslE, are required for biofilm
614 formation and virulence of enteropathogenic Escherichia coli. Infect Immun.
615 2012;80(6):2042–52.
616 9. Sheikh J, Czeczulin JR, Harrington S, Hicks S, Henderson IR, Le Bouguénec C, et al. A novel
617 dispersin protein in enteroaggregative Escherichia coli. J Clin Invest. 2002;110(9):1329–37.
618 10. Crossman LC, Chaudhuri RR, Beatson SA, Wells TJ, Desvaux M, Cunningham AF, et al. A
619 commensal gone bad: complete genome sequence of the prototypical enterotoxigenic
620 Escherichia coli strain H10407. J Bacteriol. 2010;192(21):5822–31.
621 11. Rivas ZP, Talbot KM, Merselis LC, McCormack RM, Adkins B, Munson GP. CexE Is a Coat
622 Protein and Virulence Factor of Diarrheagenic Pathogens. Front Microbiol. 2020;11(1374):1–
623 13.
624 12. Belmont-Monroy L, Saitz-rojas W, Soria-bustos J, Mickey S, Sherman NE, Orsburn BC, et al.
625 Characterization of a novel AraC/XylS- regulated family of N-acyltransferases in pathogens of
626 the order Enterobacterales. PLoS Pathog. 2020;16(8):e1008776.
627 13. Nishi J, Sheikh J, Mizuguchi K, Luisi B, Burland V, Boutin A, et al. The export of coat protein
628 from enteroaggregative Escherichia coli by a specific ATP-binding cassette transporter system.
629 J Biol Chem. 2003;278(46):45680–9.
630 14. Pilonieta MC, Bodero MD, Munson GP. CfaD-dependent expression of a novel
631 extracytoplasmic protein from enterotoxigenic Escherichia coli. J Bacteriol.
632 2007;189(14):5060–7.
633 15. Baudry B, Savarino SJ, Vial P, Kaper JB, Levine MM. A Sensitive and Specific DNA Probe to
634 Identify Enteroaggregative Escherichia coli, a Recently Discovered Diarrheal Pathogen.
635 1990;1249–51.
636 16. Petty NK, Bulgin R, Crepin VF, Cerdeño-Tárraga AM, Schroeder GN, Quail M a, et al. The
637 Citrobacter rodentium genome sequence reveals convergent evolution with human pathogenic
638 Escherichia coli. J Bacteriol. 2010;192(2):525–38.
639 17. Hodson C, Yang J, Hocking DM, Azzopardi K, Chen Q, Holien JK, et al. Control of Virulence
640 Gene Expression by the Master Regulator, CfaD, in the Prototypical Enterotoxigenic
641 Escherichia coli Strain , H10407. Front Microbiol. 2017;11(8):1525.
642 18. Oliver DB, Cabelli RJ, Dolan KM, Jarosik GP. Azide-resistant mutants of Escherichia coli
643 alter the SecA protein, an azide-sensitive component of the protein export machinery.
644 1990;87(21):8227–31.
645 19. Huie JL, Silhavy TJ. Suppression of signal sequence defects and azide resistance in

Escherichia coli commonly result from the same mutations in secA. 1995;177(12):3518–26.
20. Vidal-Ingigliardi D, Lewenza S, Buddelmeijer N. Identification of Essential Residues in Apolipoprotein N-Acyl Transferase, a Member of the CN Hydrolase Family. J Bacteriol. 2007;189(12):4456–64.
21. Gélis-Jeanvoine S, Lory S, Oberto J, Buddelmeijer N. Residues located on membrane-embedded flexible loops are essential for the second step of the apolipoprotein N-acyltransferase reaction. Mol Microbiol. 2015;95(4):692–705.
22. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web protal for protein modeling, prediction and analysis. Nat Protoc. 2016;10(6):845–58.
23. Wiktor M, Weichert D, Howe N, Huang CY, Olieric V, Boland C, et al. Structural insights into the mechanism of the membrane integral N-acyltransferase step in bacterial lipoprotein synthesis. Nat Commun. 2017;8:15952.
24. Babu MM, Priya ML, Selvan AT, Madera M, Gough J, Aravind L, et al. A database of bacterial lipoproteins (DOLOP) with functional assignments to predicted lipoproteins. J Bacteriol. 2006;188(8):2761–73.
25. Zückert WR. Secretion of bacterial lipoproteins: through the cytoplasmic membrane, the periplasm and beyond. Biochim Biophys Acta. 2014;1843(8):1509–16.
26. Grabowicz M. Lipoproteins and Their Trafficking to the Outer Membrane. EcoSal Plus. 2019;8(2):1–8.
27. Nakayama H, Kurokawa K, Lee BL. Lipoproteins in bacteria: structures and biosynthetic pathways. FEBS J. 2012;279(23):4247–68.
28. Duronio RJ, Jackson-Machelski E, Heuckeroth RO, Olins PO, Devine CS, Yonemoto W, et al. Protein N-myristoylation in Escherichia coli: Reconstitution of a eukaryotic protein modification in bacteria. Proc Natl Acad Sci USA. 1990;87(4):1506–10.
29. Martin DDO, Beauchamp E, Berthiaume LG. Post-translational myristoylation: Fat matters in cellular life and death. Biochimie. 2011;93(1):18–31.
30. Hooda Y, Lai CCL, Judd A, Buckwalter CM, Shin HE, Gray-Owen SD, et al. Slam is an outer membrane protein that is required for the surface display of lipidated virulence factors in Neisseria. Nat Microbiol. 2016;1(4):16009.
31. Kurtzhals P. Pharmacology of Insulin Detemir. Endocrinol Metab Clin North Am. 2007;36(1):14–20.
32. Datsenko K a, Wanner BL. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci USA. 2000 Jun 6;97(12):6640–5.
33. Lee DJ, Bingle LEH, Heurlier K, Pallen MJ, Penn CW, Busby SJW, et al. Gene doctoring: a method for recombineering in laboratory and pathogenic Escherichia coli strains. BMC Microbiol. 2009;9:252.
34. Moore SD, Prevelige, PE. A P22 Scaffold Protein Mutation Increases the Robustness of Head Assembly in the Presence of Excess Portal Protein. J Virol. 2002;76(20):10245–55.
35. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST : a new generation of protein database search programs. 1997;25(17):3389–402.
36. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011;39:W29-37.
37. The UniProt Consortium. UniProt: A worldwide hub of protein knowledge. Nucleic Acids Res. 2019;47:D506–15.
38. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 2019;47:W636–41.
39. Stamatakis A. RAxML version 8 : a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3.
40. Letunic I, Bork P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. Nucleic Acids Res. 2019;47(W1):256–9.
41. Schägger H. Tricine-SDS-PAGE. Nat Protoc. 2006;1(1):16–22.
42. Evans DJ, Evans DG. Three Characteristics Associated with Enterotoxigenic Escherichia coli Isolated from Man. Infect Immun. 1973;8(3):322–8.
43. Nataro J, Baldini M, Kaper J, Black R, Bravo N, Levine M. Detection of an adherence factor

701 of enteropathogenic Escherichia coli with a DNA probe. J Infect Dis. 1985;152(3):560–5.
702 44. Chang ACY, Cohen SN. Construction and characterization of amplifiable multicopy DNA
703 cloning vehicles derived from the P15A cryptic miniplasmid. J Bacteriol. 1978;134(3):1141–
704 56.
705
706