ACCEPTED VERSION

http://hdl.handle.net/2440/75237

TITLE

2   Residues located inside the *Escherichia coli* FepE protein oligomer are essential for lipopolysaccharide

3   O-antigen modal chain length regulation

4

5   **Running Title**

6   *Characterisation of LPS Oag chain length regulator FepE*

7

8   **Contents Category**

9   Cell and Molecular Biology of Microbes

10

11   **Authors**

12   Elizabeth Ngoc Hoa Tran and Renato Morona

13

14   **Address**

15   Discipline of Microbiology and Immunology, School of Molecular and Biomedical Science, University

16   of Adelaide, Adelaide 5005, Australia

17

18   **Correspondence**

19   Renato Morona; renato.morona@adelaide.edu.au.

20

21

22   **Abbreviations**

23   LPS, lipopolysaccharide; Oag, O-antigen; VL, very long; RUs, repeat units; TM, transmembrane;

24   PCP1a, polysaccharide co-polymerase class 1a; β-ME, beta-mercaptoethanol; WT, wild-type; Cys,

25   cysteine; MS, mass spectrometry; Und-PP, undecaprenyl pyrophosphate.

26

SUMMARY

The *E. coli* O157:H7 FepE protein regulates lipopolysaccharide (LPS) O-antigen (Oag) chain length to confer a very long (VL) modal chain length of >80 Oag repeat units (RUs). The mechanism by which FepE regulates Oag modal chain length and the regions within it that are important for its function remain unclear. Studies on the structure of FepE show that the protein oligomerises. However the exact size of the oligomer is in dispute, leading to a further unclear view of its mechanism. Guided by information previously obtained for regions known to be important for Oag modal chain length determination in the homologous *Shigella flexneri* $WzzB_{SF}$ protein, a set of FepE mutant constructs with single amino acid substitutions was created. Analysis of the resulting LPS conferred by these mutant His_6-FepE proteins showed that amino acid substitutions of leucine 168 (L168) and aspartic acid 268 (D268) resulted in LPS with consistently shortened Oag chain lengths of <80 Oag RUs. Substitution of FepE's transmembrane (TM) cysteine residues did not affect function. Chemical cross-linking experiments on mutant FepE proteins showed no consistent correlation between oligomer size and functional activity, and mass spectrometry analysis of FepE oligomers indicated that the *in vivo* size of FepE is consistent with a maximum size of a hexamer. Our findings suggest that different FepE residues, mainly located within the internal cavity of the oligomer, contribute to Oag modal chain length determination but not the oligomeric state of the protein.

46 **INTRODUCTION**

47 Lipopolysaccharide (LPS) is an important virulence factor of many Gram negative bacteria and

48 generally consists of 3 distinct regions: the membrane anchored lipid A domain, the core sugar

49 region and the O-antigen (Oag) polysaccharide chains (Raetz & Whitfield, 2002). As with other

50 members of the family *Enterobacteriaceae*, the genes encoding enzymes for Oag biosynthesis and

51 polymerization are mainly found in the bacterial chromosome, but can also be plasmid-encoded

52 (Raetz & Whitfield, 2002). Oag is a polymer of sugar repeat units (RUs) which define the Oag

53 serotype specificity. The basic Oag RU of *Shigella flexneri* is a tetrasaccharide made up of three

54 rhamnose sugars and one *N*-acetylglucosamine sugar. The following describes the Oag synthesis

55 strategy in *S. flexneri* and *E. coli* K-12 known as the Wzy-dependant polymerisation pathway

56 (Morona *et al.*, 2009; Raetz & Whitfield, 2002; Samuel & Reeves, 2003; Tocilj *et al.*, 2008).

57 Biosynthesis of the Oag is initiated on the cytoplasmic side of the inner membrane. After a series of

58 successive glycoyl transferase reactions, a RU is assembled on the membrane-bound carrier

59 undecaprenyl pyrophosphate (Und-PP) and is then transferred across to the periplasmic side of the

60 membrane by the Wzx flippase. Oag RUs then undergo polymerisation by the Wzy polymerase and

61 are attached onto the lipid A-core molecule by the WaaL ligase to form a chain of Oag RUs on a

62 complete LPS molecule. The number of Oag RUs can vary considerably from 0 to >100, are non-

63 randomly distributed into distinct modal lengths, and are controlled by Oag chain length regulators

64 belonging to the polysaccharide co-polymerase class 1a (PCP1a) protein family (Morona *et al.*,

65 2000). PCP1a proteins have N-terminal and C-terminal transmembrane helices (TM1 and TM2)

66 which flank a periplasmic polypeptide segment that contains predicted coiled-coil regions. LPS Oag

67 modal chain length regulation has an important role in the pathogenesis of different bacteria (al-

68 Hendy *et al.*, 1992; Crawford *et al.*, 2012; Hong & Payne, 1997; Kintz *et al.*, 2008; Murray *et al.*,

4

69 2003; Murray *et al.*, 2005; Najdenski *et al.*, 2003; Van Den Bosch *et al.*, 1997; Van Den Bosch &

70 Morona, 2003; Zhang *et al.*, 1997).

71

72 The *E. coli* O157:H7 PCP1a protein FepE regulates LPS Oag chain length to confer very long (VL)

73 modal chain lengths of >80 Oag RUs (Tocilj *et al.*, 2008). The mechanism by which FepE regulates

74 LPS Oag modal chain length and the regions within it that are important for its function remains

75 unclear. Studies on the structure of FepE show that the protein oligomerises, but the size of this

76 oligomer appears to be variable. Crystallization data on the periplasmic region of FepE and a FepE

77 mutant conferring a shorter LPS Oag chain length show that both form nonamer structures

78 (Kalynych *et al.*, 2012; Tocilj *et al.*, 2008). In contrast, cryo-electron microscopy studies on the

79 full-length FepE reconstituted into proteoliposomes suggests a preference for a hexameric state

80 (Larue *et al.*, 2009).

81

82 Several studies on FepE and other PCP1a proteins have been undertaken to define functional

83 regions that affect LPS Oag modal chain length determination (Daniels & Morona, 1999; Franco *et*

84 *al.*, 1998; Kalynych *et al.*, 2011; Kintz & Goldberg, 2011; Marolda *et al.*, 2008; Papadopoulos &

85 Morona, 2010; Purins *et al.*, 2008; Tocilj *et al.*, 2008). Franco *et al.* showed that site-directed

86 mutations made in the periplasmic region of *E. coli* O2 Wzz at residues D90 and L91 altered LPS

87 Oag modal chain length (Franco *et al.*, 1998). While in *S. flexneri* WzzB$_{SF,}$ mutagenesis of residue

88 K267 in the periplasmic domain and altering residues in the TM regions resulted in significant

89 changes (Daniels & Morona, 1999). A later study on WzzB$_{SF}$ also identified mutations made in the

90 coiled-coil regions of WzzB$_{SF}$ which conferred partial defects on LPS Oag modal chain length

91 (region I), eliminated WzzB$_{SF}$ function (region II) or had no effect on LPS Oag modal chain length

92 (region III) (Marolda *et al.*, 2008). Purins *et al*. (2008) also showed that mutations made in selected

93   coil-coiled domains region of *S. flexneri* Wzz$_{pHS2}$ could result in loss of modality (Purins *et al.*,

94   2008). In *E. coli* FepE, mutations in the periplasmic region at residues D95V, E133, K201, Q232,

95   R208 and D315 did not appear to effect LPS Oag modal chain length but a double mutation at

96   residues D225V and E297A conferred a slightly shorter LPS Oag modal chain length (Tocilj *et al.*,

97   2008). Interestingly, a recent study by Kintz *et al.* using *Pseudomonas aeruginosa* Wzz2 showed

98   that a mutation at residue 321 located within the second coil-coiled region of the protein resulted in

99   shortened LPS Oag modal chain length depending on the amino acid (aa) introduced (Kintz &

100   Goldberg, 2011). Since PCP proteins are predicted to share similar structural folds (Tocilj *et al.*,

101   2008), these studies suggest that specific aa rather than specific regions of PCP1a proteins affect

102   regulation of LPS Oag modal chain length. More recently, a study using in-frame linker insertion

103   mutagenesis on WzzB$_{SF}$ identified periplasmic sites which altered the LPS Oag modal chain length

104   to varying degrees (Papadopoulos & Morona, 2010), classifying them into 5 mutant classes: I

105   (function knockout), II & III (shorter that wild-type [WT] LPS Oag chains), IV (WT LPS Oag

106   chains) and V (longer than WT LPS Oag chains). A study on chimeric molecules containing

107   segments of WzzB$_{SF}$ and *Salmonella typhimurium* WzzB$_{ST}$ has also identified regions spanning

108   WzzB$_{SF}$ and FepE which conferred different LPS Oag modal chain lengths (Kalynych *et al.*, 2011).

109   To date, no additional site-directed mutagenesis has been undertaken to identify specific aa that

110   affect FepE regulation of LPS Oag modal chain length.

111

112   In this study, information previously obtained on residues important for WzzB$_{SF}$ function was used

113   to predict residues in FepE which may affect function and hence to create a set of mutant constructs

114   with single aa substitutions in FepE. LPS analysis on these constructs suggest that FepE residues

115   located inside the oligomer cavity contribute to LPS Oag modal chain length determination and

116   identified leucine 168 (L168) and aspartic acid 268 (D268) as residues important for VL Oag modal

6

117   chain length regulation. In addition to this, the role of TM region cysteine (Cys) residues was

118   investigated, and a second set of mutants with single Cys substitutions was also constructed and

119   characterised. The FepE mutants were used to investigate oligomer formation by cross-linking. The

120   data suggest that FepE can form a number of oligomeric structures although no consistent

121   correlation with an effect on function was found. Mass spectrometry (MS) of purified FepE

122   oligomers cross-linked *in situ* suggests the *in vivo* size of the FepE oligomer is consistent with a

123   maximum size of a hexamer.

124

METHODS

Bacterial strains and growth conditions. The bacterial strains and plasmids used in this study are listed in Table 1. Strains were routinely grown at 37°C in Luria-Bertani (LB) broth (10 g/liter Tryptone, 5 g/litre yeast extract, 5 g/litre NaCl) with aeration for 16 h, subcultured 1/20 into fresh broth and grown to an OD_{600} of ~0.8. Induction was carried out with 1 mM IPTG, followed by growth for another 3 h. Antibiotics were used at the following concentrations: 100 μg ampicillin ml^{-1}; 50 μg kanamycin ml^{-1}; and 100 μg streptomycin ml^{-1}.

Mutagenesis. Site-directed mutagenesis was carried out following the QuikChange Lightning® protocol (Stratagene) with complementary primers containing nucleotides encoding random aa (NNN) or specific nucleotide substitutions (Table S1). The latter were designed to alter charge or hydrophobicity of the residues at the mutated position. The expression plasmid pQE30-*wzz*FepE (encoding for His$_6$-FepE) (Tocilj *et al.*, 2008) was used as a template for the construction of mutant genes. Mutant constructs were confirmed by DNA sequencing with pQE primers ET11-13 (Table S1), prior to electroporation into *S. flexneri* RMA4053, a *S. flexneri* Y *wzz::kan* strain (lacking pHS2) and carrying pCDFDuet-1 (encoding *lacI^q*) (Table 1). Electroporation and preparation of electrocompetent cells were prepared as previously described (Purins *et al.*, 2008).

SDS-PAGE and Western immunoblotting. Bacteria were grown and induced as described above, harvested by centrifugation, and resuspended in 1x sample buffer (Lugtenberg *et al.*, 1975). Protein samples were heated at 100°C for 5 mins, except for *N*-[4-(p-azidosalicylamido)butyl]-3'-(2'-pyridyldithio)propionamide (APDP) treated samples which were heated at 60°C for 5 min, prior to SDS-PAGE on 12% or 15% gels. Protein gels were either stained with Coomassie R-250, or subjected to Western immunoblotting on nitrocellulose membrane (Medos) with polyclonal FepE

8

149 antibodies (obtained from a rabbit following immunisation with purified FepE$_{116-2193}$ protein

150 provided by Prof. M. Cygler) at 1/500 dilution. Detection was performed with goat anti-rabbit

151 horseradish-peroxidase-conjugated antibodies (KPL) and chemiluminescence reagent (Sigma). The

152 molecular mass standards used for Coomassie Blue stained gels were Low Molecular Weight-SDS

153 marker (Quantum Scientific) and Novex Sharp (Invitrogen) for MS analysis. BenchMark protein

154 ladder (Invitrogen) was used for Western immunoblots.

155

156 **LPS PAGE and silver staining**. LPS samples and gels were prepared as described previously

157 (Murray *et al.*, 2003; Papadopoulos & Morona, 2010).

158

159 **FepE protein purification.** Cells were grown and induced as described above in 200 ml LB broth,

160 harvested by centrifugation (9,800x*g*, 10 min, 4°C, Beckman J2-21M Induction Drive Centrifuge)

161 and resuspended in 15 ml NaPO$_4$ buffer (50 mM NaPO$_4$, 500 mM NaCl, pH 7), sonicated and re-

162 centrifuged to remove cell debris. Whole membrane (WM) pellets were collected by

163 ultracentrifugation (126,000x*g*, 1 h, 4°C, Beckman Coulter Optima L-100 XP Untracentrifuge) and

164 solubilised in 1 ml NaPO$_4$ buffer containing 1% (w/v) SDS for 2 h at RT prior to re-

165 ultracentrifugation (as above). Solubilised FepE in the supernatant fraction was mixed with 100 μl

166 Profinity™ IMAC resin (BIO-RAD #156-0133) for 1 h at RT, washed three times with NaPO$_4$

167 buffer containing 0.008% (w/v) SDS, four times with NaPO$_4$ buffer containing 0.008% (w/v) SDS

168 and 20 mM imidazole, and then eluted with 200 μl NaPO$_4$ buffer containing 1% (w/v) SDS and 500

169 mM imidazole, pH 7. Eluted protein was mixed 1:1 with 2X sample buffer and ~5 μl samples were

170 analysed by SDS-PAGE. Gels were stained with Coomassie Blue to visualise protein bands.

171

Chemical cross-linking analysis. Cells grown and induced in 200 ml LB broth were harvested by centrifugation as above, resuspended in 15 ml NaPO$_4$ buffer, disrupted by sonication and re-centrifuged to remove cell debris. The sonicated cell extract containing WM vesicles was incubated with and without 500 μM APDP (Thermo Scientific #27720) for 1 h at RT (in the dark), and then exposed to UV light for 15 min with a 365 nm UV lamp (Thermo Scientific #95035) at a distance of 10 cm on ice, followed by quenching with 0.3 M Tris-HCl (pH 7) for 5 min. APDP reacts with the SH groups of Cys residues by a disulfide exchange reaction which is cleavable with β-ME. Following UV irradiation, its photoreactive azide group reacts with neighbouring molecules and establishes a cross-link. One ml volumes of treated and untreated samples were collected and mixed 1:1 with 2X sample buffer (with and without beta-mercaptoethanol [β-ME]) for analysis by SDS-PAGE. Gels were subjected to Western immunoblotting with FepE antibodies. Affinity purification of APDP treated and untreated proteins was carried with the remaining ~14 ml of samples by ultracentrifugation and solubilisation of WM pellets in NaPO$_4$ buffer containing 1% (w/v) SDS, followed by protein elution from Profinity™ IMAC resin as described above.

LC-eSI-IT mass spectrometry. Liquid chromatography-electrospray ionisation ion-trap (LC-eSI-IT) mass spectrometry (MS) was carried out by the Adelaide Proteomics Centre. Briefly, protein samples were electrophoresed on 4-12%-SDS polyacrylamide gels (Invitrogen #NPO322BOX) and Coomassie-stained bands of interest were excised, washed in 50 mM ammonium carbonate (NH$_4$HCO$_3$), destained with 30% acetonitrile (ACN) in 50 mM NH$_4$HCO$_3$, reduced with 0.5 μM dithiothretiol in 100 mM NH$_4$HCO$_3$, followed by alkylation with 2.75 μM iodoacetamide in 100 μM NH$_4$HCO$_3$. Samples were then digested with 100 ng trypsin in 5 mM NH$_4$HCO$_3$/10% ACN and extracted with 1% formic acid in water, 1% formic acid in 50% ACN and 100% ACN. The volumes of the resulting peptide extracts were reduced by vacuum centrifugation and resuspended with 0.1%

196   formic acid in 2% ACN prior to LC-eSI-IT MS analysis. LC-eSI-IT MS/MS was performed using

197   an online 1100 series HPLC system (Agilent Technologies) and a HCT Ultra 3D-Ion-Trap mass

198   spectrometer (Bruker Daltonics). MS and MS/MS spectra were subjected to peak detection and de-

199   convolution using DataAnalysis (Version 3.4, Bruker Daltonics), annotated using BioTools

200   (Version 3.1, Bruker Daltonics) and submitted to MASCOT (Version 2.2) for protein identification.

201

202 **RESULTS**

203 **Identification of putative sites that may affect FepE function**

204 To investigate FepE residues which may play a key role in its function, regions known to be

205 important for Oag modal chain length determination in the FepE homologue from *Shigella flexneri*,

206 WzzB$_{SF}$, were analysed and compared to FepE. Despite overall low sequence identity (Fig. 1), both

207 proteins are structurally similar (Kalynych *et al.*, 2011; Kalynych *et al.*, 2012; Tocilj *et al.*, 2008)

208 and belong to the same family of PCP1a proteins. Analysis of several WzzB$_{SF}$ mutants showed that

209 mutational alterations made between aa 102-107, 128-131 and 219-232 consistently had an effect

210 on Oag modal chain length determination (Daniels & Morona, 1999; Papadopoulos & Morona,

211 2010; Morona, unpublished data). Guided by sequence alignment of the two proteins, FepE regions

212 spanning aa 110-115, 168-172 and 259-274 were predicted to affect Oag modal chain length

213 determination (Fig. 1). Twenty mutant constructs containing different aa substitutions as a result of

214 either random or specific nucleotide mutagenesis were hence made at positions F111, V114, L168,

215 T170, D268 and G274 to give pRMET1 – 15, pRMET38 - 40 and pRMET42 - 43 (Table 1). The

216 locations of these targeted residues are mapped on the 3D structure of FepE (Tocilj *et al.*, 2008)

217 (Fig. 2a - c). None of the aa substitutions were predicted to have an effect on local secondary

218 structure as determined by the JPRED 3 secondary structure prediction server (Cole *et al.*, 2008).

219

220 **LPS Oag modal chain lengths conferred by FepE mutants**

221 Plasmids pQE30-*wzz*$_{FepE}$ (encoding for wild-type His$_6$-FepE), pRMET1 – 15, pRMET38 – 40,

222 pRMET42 - 43 (encoding for mutant His$_6$-FepE proteins), and pQE30 were transformed into

223 RMA4053 (Table 1) to investigate the LPS Oag modal chain length distribution mediated by each

224 mutant construct. Analysis of the resulting LPS by SDS-PAGE and silver staining showed FepE

225 conferred a VL Oag modal chain length of >80 Oag RUs (Fig. 3a, lane 1). No Oag modal chain

226   length was conferred by the control strain with pQE30 as expected (Fig. 3a, lane 2). FepE mutants

227   with different aa substitutions at phenylalanine 111 (F111) conferred either slightly shortened Oag

228   modal chain length (<VL Oag RUs) for F111V and F111G (Fig. 3a, lanes 3 & 5) or loss of Oag

229   modal chain length regulation for F111P (Fig. 3a, lane 4), suggesting that different aa have a

230   different effect at this site. F111P also showed less polymerization of LPS Oag than that seen in the

231   control strain with pQE30 only (Fig. 3a, lane 2), suggesting that the F111P substitution also

232   interferes with polymerization. V114P, a mutant with an aa substitution at nearby valine 114, had

233   no detectable change in LPS Oag modal chain length (Fig. 3a, lane 6).

234

235   Interestingly, FepE mutants with different aa substitutions at leucine 168 (L168) conferred LPS

236   with consistently shortened Oag modal chain lengths of 18 to <VL Oag RUs for L168D, L168Q,

237   L168A and L168P (Fig. 3a, lanes 7, 8, 9 & 11), with an arginine substitution conferring the shortest

238   Oag modal chain length of 14 – 28 Oag RUs for L168R (Fig. 3a, lane 10). These results suggest

239   that a positive charge at this site has a dramatic effect. Structurally, L168 is located at the bottom of

240   the 3D structure of FepE (Fig. 2d - f). Substitutions made at nearby threonine 170 (T170) showed

241   less of an effect with slightly shortened Oag modal chain lengths (<VL Oag RUs) conferred by

242   T170C and T170D (Fig. 3a, lanes 12 & 14), and no change in Oag modal chain length conferred by

243   T170R (Fig. 3a, lane 13).

244

245   Different aa substitutions of aspartic acid 268 (D268) showed varied shortening of the LPS Oag

246   modal chain length (Fig. 3a, lanes 15 - 22). D268Y, D268L and D268V conferred the shortest Oag

247   modal chain lengths of 3 – 14 or 7 – 15 Oag RUs (Fig. 3a, lanes 15, 18 & 19), while D268G,

248   D268N and D268R conferred shortened Oag modal chain lengths of 10 – 22, 9 – 18 and 10 to <VL

249   Oag RUs respectively (Fig. 3a, lanes 16, 18 & 19). Notably, substitution of D268 with neutral aa

250   residues tyrosine (Y), asparagine (N), leucine (L), valine (V) and glycine (G) appeared to confer

251   shorter Oag modal chain lengths than substitution of D268 with positively charged arginine (R).

252   When D268 was substituted with negatively charged glutamic acid (E), a WT Oag modal chain

253   length of >80 Oag RUs was conferred (Fig. 3a, lane 17), suggesting that conservation of a negative

254   charge at this position is essential. Structurally, D268 is located at the top of the 3D structure of

255   FepE (Fig. 2g - i). In contrast to all other mutational alterations investigated, a tryptophan (W) aa

256   substitution at nearby glycine 274 (G274) resulted in a broad lengthening of the Oag chain length

257   with loss of modality (Fig. 3a, lane 22).

258

259   Western immunoblotting performed on whole cell lysates from *S. flexneri* strains expressing WT

260   and mutant FepE proteins detected a band consistent with the size of the WT His$_6$-FepE protein

261   (~43 kDa) for all mutants except F111P which showed no expression (Fig. 3b, lane 4) and G274W

262   which had a band at 32 kDa only (Fig. 3b, lane 22). These results suggest a correlation between loss

263   of LPS Oag modal chain length regulation and absence of FepE production (Fig. 3b & Table 2).

264   The lower molecular mass ~32 kDa band detected in most samples with anti-FepE antibodies may

265   be either a degradation product or an altered conformation of His$_6$-FepE (Fig. 3b).

266

267   **Mutagenesis of Cys residues in the TM regions of FepE**

268   FepE has only two cysteine (Cys) residues located at aa position C52 (TM1) and C354 (TM2) (Fig.

269   1). To investigate if these Cys residues were required for function, mutants with single or double aa

270   substitutions at positions C52 and C354 were created. Analysis of the LPS profiles conferred by

271   different aa substitutions of C52 (C52Q, C52W, C52P and C52S) and C354 (C354S, C354D,

272   C354G and C354A) showed no effect on LPS Oag modal chain length (Fig. S1). Likewise, when

273   both C52 and C354 were mutated (C52S/C354A), no effect on WT Oag modal chain length was

14

274 observed (Fig. S1), suggesting that the TM region cysteines are not critical for LPS Oag modal

275 chain length regulation. Western immunoblotting performed on whole cell lysates detected a band

276 consistent with the size of the WT His$_6$-FepE protein for all mutants, with slightly less protein

277 detected for C52Q and C52P (Table 2).

278

279 **Introduction of Cys residues in aa positions not affecting FepE function**

280 In a new approach to investigating the structure of FepE in membranes, Cys directed cross-linking

281 was undertaken using APDP, a membrane permeable and heterobifunctional cross-linker with a

282 spacer arm of ~21 Å. In addition to using the native FepE TM region Cys residues, to assist in

283 cross-linking we also introduced Cys residues in the periplasmic domain of FepE at aa positions

284 that did not affect its LPS Oag modal chain length regulation (Fig. 1).

285

286 FepE residues D95, E133, K201, R208, Q232 and D315 which did not appear to affect function in

287 Tocilj *et al.* (2008) were altered to Cys by site-directed mutagenesis on His$_6$-FepE. The locations of

288 these targeted residues are mapped on the 3D structure of FepE (Fig. 2j – k) with the exception of

289 E133; E133 resides in a disordered region not seen in the crystal structure, so its approximate

290 location is indicated in Fig. 2j – k. Analysis of the LPS conferred by these mutant His$_6$-FepE-Cys

291 proteins (Fig. 4a & Table 2) showed that E133C, Q232C and D315C conferred WT Oag modal

292 chain lengths of >80 Oag RUs and hence were not affected in function (Fig. 4a, lanes 4, 6 & 7).

293 D95C and K201C conferred slightly shortened LPS Oag modal chain lengths (<VL Oag RUs) (Fig.

294 4a, lanes 3 & 8), while R208C conferred a shortened Oag modal chain length of 20 to <VL Oag

295 RUs (Fig. 4a, lane 5). Western immunoblotting performed on whole cell lysates detected a band

296 consistent with the size of the WT His$_6$-FepE protein (~43 kDa) for all mutants (Fig. 4b, lanes 2 –

297 8). FepE mutants E133C, Q232C and D315C were hence chosen for subsequent experiments.

298

In situ cross-linking of FepE, C52S/C354A and FepE-Cys mutants

300   The ability of FepE and C52S/C354A to form higher order oligomers was then investigated by

301   APDP cross-linking using cell lysates from strains expressing the proteins. Western

302   immunoblotting analysis on FepE with 2 TM region Cys residues showed readily detectable ~43

303   kDa (monomeric) and ~86 kDa (dimeric) forms of FepE in the untreated sample (Fig. 5a, lanes 1 &

304   2). Treatment of FepE with UV only as a control showed similar results to the untreated control

305   (Fig. 5a, lanes 3 & 4). When FepE was treated with APDP, the ~43 kDa (monomeric) and ~86 kDa

306   (dimeric) forms were detected plus a new high molecular weight (HMW) band of >180 kDa,

307   indicating the presence of higher order FepE oligomerisation (Fig. 5a, lane 5). In the presence of β-

308   ME, this HMW band was greatly reduced, suggesting that it arose due to cross-linking of FepE

309   monomers by APDP (Fig. 5a, lane 6). Analysis of C52S/C354A with both TM region Cys residues

310   mutated detected only the ~43 kDa (monomeric) and ~86 kDa (dimeric) forms in untreated and

311   APDP treated samples (Fig. 5b, lanes 1-6), suggesting that the HMW band of >180 kDa detected in

312   FepE sample is due to cross-linking at the TM Cys residues and that the TM regions are close

313   together.

314

315   The ability of E133C, Q232C, and D315C to form oligomers was then investigated. Unexpectedly,

316   in both untreated and UV treated control samples of E133C, bands were detected at ~43 kDa

317   (monomeric), ~86 kDa (dimeric), between 115 - 182 kDa (intermediate species) and at >180 kDa

318   (HMW) (Fig. 5c, lanes 1- 4). These results suggest stable oligomer formation is exhibited by this

319   mutant in the absence of cross-linker. APDP treatment of E133C (Fig. 5c, lane 5) resulted in a

320   banding profile similar to that for likewise treated FepE (Fig. 5a, lane 5) but with the addition of the

16

321   intermediate band. Treatment with β-ME resulted in detectable ~43 kDa (monomeric) and ~86 kDa

322   (dimeric) species only (Fig. 5c, lanes 2, 3 & 6), as seen for FepE.

323

324   To investigate whether the native FepE TM Cys residues contributed to the E133C banding profile,

325   an E133C mutant with aa substitutions at both the TM region Cys residues was created

326   (C52S/C354A/E133C). Analysis of C52S/C354A/E133C treated with and without APDP (Fig. 5f,

327   lanes 1 - 6) showed the same banding profile detected for E133C (Fig. 5c, lanes 1 - 6), suggesting

328   that the intermediate and HMW bands observed for E133C are a consequence of the Cys at position

329   E133. Interestingly, the bands were also sensitive to β-ME (Fig. 5f, lane 2, 4 & 6).

330

331   When the Q232C mutant protein was investigated, bands were also detected at ~43 kDa

332   (monomeric), ~86 kDa (dimeric), between 115 - 182 kDa (faint intermediate) and at >180 kDa

333   (faint HMW) in the untreated sample (Fig. 5d, lanes 1 & 3), suggesting that this mutant also

334   exhibits stable oligomer formation. Like E133C, the intermediate and HMW bands were sensitive

335   to β-ME (Fig. 5d, lanes 2, 4 & 6). Following APDP treatment, Q232C showed an increase in the

336   HMW band (Fig. 5d, lane 5). In contrast, the D315C mutant protein showed a cross-linking profile

337   similar to FepE, with the ~43 kDa (monomeric) and ~86 kDa (dimeric) bands detected in untreated

338   and UV treated samples (Fig. 5e, lanes 1-4), and the additional HMW form detected in the APDP

339   treated sample (Fig. 5e, lane 5). Treatment with β-ME showed loss of the HMW form (Fig. 5e, lane

340   6) as seen for FepE (Fig. 5a, lane 6).

341

342   Hence, the higher oligomer forms of FepE were detectable either depending on the presence of an

343   additional Cys residue (in the case of E133C and Q232C) or the addition of APDP (in the case of

344 FepE and D315C). The apparent sizes of most bands detected were consistent with multiples of the

345 FepE monomer.

346

347 ***In situ* cross-linking of FepE mutants**

348 Having established a baseline for the cross-linking of FepE by APDP, we next investigated whether

349 mutations that affected LPS Oag modal chain length regulation had an effect on FepE oligomer

350 formation. Several FepE mutants that conferred a shortened LPS Oag modal chain length (L168R,

351 L168P, D268Y and D268G) were hence subjected to APDP cross-linking analysis (Fig. 5 & Table

352 2). Untreated samples containing L168R, L168P and D268Y had the ~43 kDa (monomeric) form

353 but showed loss of the ~86 kDa (dimeric) form (Fig. 5g - i, lanes 1 − 2). In the UV only and APDP

354 treated samples, the ~84 kDa form was detected for all 3 mutants (Fig. 5g - i, lanes 3 - 6),

355 suggesting that the L168R, L168P and D268Y mutations may either prevent or destabilise dimer

356 formation under the conditions used, but this can be restored by UV and/or APDP treatment. The

357 HMW band was detected in all APDP treated samples (Fig. 5g - i, lanes 5) as seen for FepE (Fig.

358 5a, lane 5). In contrast, the ~86 kDa (dimeric) form was detected in both untreated and APDP

359 treated samples of D268G (Fig. 5j, lanes 1 - 6), indicating that shortened LPS Oag modal chain

360 length due to the D268G mutation (Fig. 3a, lane 16) is not solely due to the absence of the ~86 kDa

361 form. APDP treated sample of D268G (Fig. 5j, lane 5) also showed a similar banding profile to

362 FepE (Fig. 5a, lane 5). Cross-linking analysis undertaken on all other FepE mutants (Fig. S2 &

363 Table 2) show that while specific substitutions in FepE can affect the observed banding profile

364 (F111P, F111G, T170C, D268V & G274W), no consistent correlation between banding profile and

365 the effect of the mutation on LPS Oag modal chain length could be made (Table 2). It is interesting

366 to note that the intermediate band previously observed for E133C and Q232C (Fig. 5c & d) is also

367 present for T170C having a Cys substitution (Fig. S2h).

18

368

Mass spectrometry analysis on purified FepE and E133C

369

370 Our cross-linking data suggested that *in situ* FepE can form oligomers but whether these oligomers

371 are solely composed of FepE or FepE associated with other proteins is unknown. To investigate the

372 identity of the proteins present in the higher oligomer structures of FepE as detected above, strains

373 expressing either His$_6$-FepE or His$_6$-FepE$_{E133C}$ were incubated with and without APDP, and

374 proteins were affinity purified via their His$_6$ tags. E133C was selected based on its ability to

375 oligomerise the presence and absence of APDP and still confer a WT Oag chain length distribution.

376 The profile of the affinity purified FepE proteins were first verified by Coomassie Blue staining on

377 standard SDS 12% gels (Fig. 6a), and then separated on SDS 4 - 12% gels for MS analysis (Fig.

378 6b). Unexpectedly, on the SDS 4 - 12% gels the ~86 kDa (dimeric) form migrated as a doublet and

379 the HMW band separated into several bands between 140 – 260 kDa (Fig. 6b). Interestingly, all

380 oligomers formed by His$_6$-FepE and His$_6$-FepE$_{E133C}$ did not exceed an apparent molecular mass of

381 260 kDa (Fig. 6b).

382

383 MS identification was carried out on the ~43 kDa (monomeric) and ~86 kDa doublet bands (from

384 both untreated and APDP treated samples), and on several of the separated HMW bands between

385 140 – 260 kDa (from APDP treated sample) for FepE (Fig. 6b, lanes 2 & 4). For E133C, the ~43

386 kDa (monomeric) band and the intermediate species (which now appeared to migrate at ~100 kDa)

387 were analysed (from untreated and APDP treated samples), as well as several of the separated

388 HMW bands between 140 – 260 kDa (from APDP treated sample) (Fig. 6b, lanes 6 & 8). MS

389 analysis of all the above mentioned protein bands cut from the SDS-PAGE gels showed the

390 presence of FepE as the major protein in all of the oligomeric FepE protein bands (Table S2 & Fig.

391 S3), suggesting that *in vivo*, FepE likely exists as a homogeneous oligomer of up to 6 protomers.

DISCUSSION

In this study we constructed and characterised a set of His$_6$-FepE mutant proteins to identify the regions that affect FepE function in conferring VL Oag modal chain length and to investigate FepE higher oligomer formation by cross-linking with APDP. Previous mutagenesis studies suggested that for WzzB$_{SF}$, different regions of the protein contribute to LPS Oag modal chain length regulation (Daniels & Morona, 1999; Franco *et al.*, 1998; Papadopoulos & Morona, 2010). Using this data (as well as unpublished data) and guided by secondary structure-based sequence alignments of WzzB$_{SF}$ and FepE, regions spanning aa 110 - 115, 168 - 172 and 259 - 274 in FepE were predicted to affect LPS Oag modal chain length conferred by FepE. It is interesting to note that these regions corresponded to intermediate regions between α helices and β sheets (Fig. 1) where a higher degree of structural flexibility is expected. His$_6$-FepE mutants with various substitutions at F111, V114, L168, T170, D268 and G274W within the three regions were found to have an impact on LPS Oag modal chain length (Fig. 3 & Table 2), suggesting that these residues are critical for FepE function. Based on sequence alignment (Tocilj *et al.*, 2008), residues F111, D268 and G274 are conserved across the FepE, WzzB and WzzB$_{pHS2}$ PCP1a proteins of *E. coli*, *S. typhimurium* and *S. flexneri,* while a neutral charge residue at positions 114, 168 and 170 appears to be conserved across the same PCP1a proteins (Tocilj *et al.*, 2008).

Several substitutions made at position L168, located between β2 and β3 sheets of FepE, conferred consistently shortened LPS Oag modal chain lengths of 18 to <VL Oag RUs (L168D, L168Q, L168A and L168P), with one mutant conferring 14 – 28 Oag RUs (L168R) (Fig. 3 & Table 2). Hence, L168 appears to be essential for conferring VL Oag modal chain length and the presence of a leucine residue at this site appears to be critical. Linker insertion (5aa) mutations made in the same region in WzzB$_{SF}$, which confers a short LPS Oag modal chain length of 10 – 17 Oag RUs,

20

416 were shown to increase the LPS Oag modal chain length to 16 – 25 Oag RUs (Papadopoulos &

417 Morona, 2010), suggesting that the same regions are important for Oag chain length control for

418 both PCP1a proteins.

419

420 Substitutions at D268, located between the α6 and α7 helices of FepE, conferred the most dramatic

421 shortening of LPS Oag modal chain lengths in this study. Out of 7 mutants, 2 mutants conferred an

422 LPS Oag modal chain length of 3 – 14 Oag RUs (D268Y and D268L), and 4 mutants (D268V,

423 D268G, D268N and D268R) conferred varied LPS Oag modal chain length of 7 – 15, 10 – 22, 9 –

424 18 and 10 to <VL Oag RUs, respectively (Fig. 3 & Table 2). The remaining mutant FepE$_{D268E}$,

425 harbouring a conserved negatively charged residue substitution, conferred WT Oag modal chain

426 length (Fig. 3 & Table 2), suggesting that maintenance of the negative charge at this position is

427 critical. This correlates with data by Kalynch *et al.* (2011), whereby substitution or deletion of the

428 FepE aa region 256 – 273 (where D268 is located), resulted in shortened LPS Oag chain lengths.

429 We report for the first time 2 specific residues, L168 and D268, which are essential for FepE

430 function in LPS Oag modal chain length determination. Similar to data for WzzB$_{SF}$, our findings

431 support that more than one region of the protein is important for determining LPS Oag chain length

432 modal specificity (Daniels & Morona, 1999; Franco *et al.*, 1998; Papadopoulos & Morona, 2010).

433

434 Our data also support the previous suggestion that the internal cavity of the FepE crystal structure

435 oligomer may play a role in Oag polysaccharide length control (Tocilj *et al.*, 2008), as the majority

436 of His$_6$-FepE mutants shown to affect function had mutational alterations at positions located inside

437 the FepE oligomer (F111, V114, L168, T170, D268 and G274W) (Fig. 2 & Table 2). Notably, two

438 of these positions, D268 and G274, are located in the FepE region spanning aa residues 240-299

439 which correlates to the WzzB$_{SF}$ segment spanning 200-255 based on sequence alignment (Tocilj *et*

440   *al.*, 2008). Kalynych *et al.* (2011) showed that this functionally important WzzB$_{SF}$ segment was

441   predominately located on the external surface of the WzzB$_{SF}$ oligomer and hypothesised that the

442   external surface of chain length regulators might represent the principal site of modal length

443   control. Residues D268 and G274 located within this same region of FepE however, point towards

444   the internal cavity of the oligomer and still affect LPS Oag modal chain length (Fig. 2 & Fig. 3).

445   Similarly, residue L168 shown to consistently shorten the LPS Oag modal chain length as a result

446   of different aa substitutions, points towards the internal cavity of the FepE oligomer (Fig. 2). Based

447   on this, we propose that the nascent Oag linked Und-PP chains are threaded into the FepE oligomer

448   barrel during Wzy dependant polymerisation. Once filling has occurred, a conformation change in

449   the FepE oligomer releases the Oag linked Und-PP chains, and hence determines Oag modal length

450   (Fig. S4). Substitution with different aa inside the FepE oligomer may result in less filling of the

451   barrel and/or an earlier change in FepE oligomer conformation to release shorter model chain

452   lengths. In regards to other PCP1a proteins, their different barrel sizes and/or requirements for

453   different degrees of interaction with Oag chains may trigger the change in conformation that

454   releases the Oag linked Und-PP chains for ligation to the lipid A core by WaaL (Fig. S4).

455

456   In previous studies on WzzB$_{SF}$ it was found that introduction of mutations at the beginning of either

457   TM1 or within TM2 (Daniels & Morona, 1999; Papadopoulos & Morona, 2010) can affect function

458   and protein cross-linking profile. Mutational alterations made in the TM regions of FepE in this

459   study, specifically altering the only two Cys residues (C52 and C354) found in FepE, had no affect

460   on function (Table 2 & Fig. S1). However, other aa residues located in the TM region have not yet

461   been investigated. Analysis of a double TM Cys mutant C52S/C354A treated with APDP only

462   detected the ~43 kDa (monomeric) and ~86 kDa (dimeric) forms of the protein (Fig. 5b), whereas

463   an additional HMW band was detectable in the APDP treated sample of FepE (with TM cysteines)

464   (Fig. 5a). Since APDP is a membrane permeable probe these results suggest that at least some of the

465   HMW bands in FepE, are due to cross-linking at the TM Cys residues, and that the FepE TMs are

466   in close proximity to each other in the cytoplasmic membrane.

467

468   Substitution at position E133 had no effect on FepE function in this study (Fig. 3) and in a previous

469   study (Tocilj *et al.*, 2008). Structurally, E133 is located at the bottom of the crystal structure of the

470   FepE oligomer in an undefined region between α4 and α5 helices, which appears to protrude out

471   from the oligomer, close to the inner membrane. Unexpectedly, E133C showed a stable oligomeric

472   form between 115 - 182 kDa (intermediate form) (Fig. 5c) despite the absence of chemical cross-

473   linker, which might have arisen by oxidation either *in situ* or during sample preparation. Analysis of

474   APDP treated E133C showed both the intermediate form and the HMW band were detected and

475   that they were β-ME sensitive (Fig. 5c). These bands were also present in the C52S/C354A/E133C

476   mutant lacking the TM cysteines (Fig. 5f). This data suggests that E133C promotes stable FepE

477   interactions, possibly through other interactions promoted and stabilised as a result of the disulfide

478   bond formation at position E133. This is not unique to E133C as Q232C and T170C also had a

479   similar effect (Table 2).

480

481   Previous data for WzzB$_{SF}$ also suggested that the dimer form may be important in determining Oag

482   modal chain length (Papadopoulos & Morona, 2010). However, we were unable to find a

483   correlation between the dimer form of FepE and LPS Oag modal chain length regulation in this

484   study. L168R, L168P, D268Y and D268G all conferred shortened LPS Oag modal chain lengths

485   (Fig. 3) but only 3 mutants (L168R, L168P and D268Y) showed loss of the ~86 kDa (dimeric) form

486   in untreated cross-linking samples (Fig. 5g – j). Upon treatment with APDP, the dimeric form was

487   detected for all mutants (Fig. 5g – f), and the APDP cross-linking profile was no different to FepE

488   (Fig. 5a). Differences obtained between the two studies may be due to differences in the methods

489   used and the protein studied; the mutational alterations used are different, and subtle changes used

490   in this study may affect function but not have a detectable effect on oligomer formation.

491

492   No consistent relationship was observed between higher order oligomer formation and the observed

493   LPS profile of the mutants studied (Table 2, Fig. 4, 5 & S2) and hence these results suggest that the

494   oligomeric form of FepE may not necessarily correlate to function. Kalynych *et al*. (2012) observed

495   that a FepE mutant although conferring a shortened LPS Oag modal chain length had the same 3D

496   structure as FepE (Kalynych *et al.*, 2012; Tocilj *et al.*, 2008), suggesting that Oag modal chain

497   length regulation by FepE is likely controlled by aa residues along the structure of the oligomer,

498   rather than either the oligomer size, or conformational changes.

499

500   X-ray crystallographic studies of the periplasmic domain of FepE show that the protein forms a

501   nonamer (Tocilj *et al.*, 2008), while cryo-electron microscopy analysis on the full-length FepE

502   protein showed that FepE most likely exists as a hexamer state (Larue *et al.*, 2009). Cross-linking

503   analysis with formaldehyde suggests that detergent-solubilised FepE can form dimeric and

504   tetrameric complexes that can assemble into higher molecular weight oligomer forms (Larue *et al.*,

505   2009). In this study, *in situ* treatment of FepE and FepE mutants with APDP cross-linker showed

506   that FepE was able to form a number of higher order oligomers (Fig. 5 & S2) that did not appear to

507   be larger than 260 kDa (Fig. 6), suggesting that FepE may exist as a hexamer *in vivo*. Analysis of

508   FepE proteins affinity purified from strains expressing FepE and E133C by MS showed that all

509   Coomassie Blue stained bands investigated consisted of FepE derived peptides (Table S2),

510   confirming that FepE can form homogeneous oligomeric structures of varying sizes *in vivo*.

511

512   In summary, our results show that residues located inside the FepE oligomer within intermediate

513   regions between α helices and β sheets, in particular L168 and D268, are important for FepE

514   regulation of VL Oag modal chain length. FepE oligomer formation as detected by APDP cross-

515   linking does not appear to correlate with an effect of mutations on function, and our MS data

516   suggest that the *in vivo* size of FepE is consistent with a maximum size of a hexamer.

517

518   **Acknowledgements**

521

REFERENCES

al-Hendy, A., Toivanen, P. & Skurnik, M. (1992). Lipopolysaccharide O side chain of *Yersinia enterocolitica* O:3 is an essential virulence factor in an orally infected murine model. *Infect Immun* **60**, 870-875.

Cole, C., Barber, J. D. & Barton, G. J. (2008). The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* **36**, W197-201.

Crawford, R. W., Keestra, A. M., Winter, S. E., Xavier, M. N., Tsolis, R. M., Tolstikov, V. & Baumler, A. J. (2012). Very long O-antigen chains enhance fitness during *Salmonella*-induced colitis by increasing bile resistance. *PLoS Pathog* **8**, e1002918.

Daniels, C. & Morona, R. (1999). Analysis of *Shigella flexneri wzz* (Rol) function by mutagenesis and cross-linking: *wzz* is able to oligomerize. *Mol Microbiol* **34**, 181-194.

Franco, A. V., Liu, D. & Reeves, P. R. (1998). The *wzz* (*cld*) protein in *Escherichia coli*: amino acid sequence variation determines O-antigen chain length specificity. *J Bacteriol* **180**, 2670-2675.

Hong, M. & Payne, S. M. (1997). Effect of mutations in *Shigella flexneri* chromosomal and plasmid-encoded lipopolysaccharide genes on invasion and serum resistance. *Mol Microbiol* **24**, 779-791.

Kalynych, S., Ruan, X., Valvano, M. A. & Cygler, M. (2011). Structure-guided investigation of lipopolysaccharide O-antigen chain length regulators reveals regions critical for modal length control. *J Bacteriol* **193**, 3710-3721.

Kalynych, S., Yao, D., Magee, J. & Cygler, M. (2012). Structural characterization of closely related O-antigen lipopolysaccharide (LPS) chain length regulators. *J Biol Chem* **287**, 15696-15705.

Kintz, E., Scarff, J. M., DiGiandomenico, A. & Goldberg, J. B. (2008). Lipopolysaccharide O-antigen chain length regulation in *Pseudomonas aeruginosa* serogroup O11 strain PA103. *J Bacteriol* **190**, 2709-2716.

Kintz, E. N. & Goldberg, J. B. (2011). Site-directed mutagenesis reveals key residue for O antigen chain length regulation and protein stability in *Pseudomonas aeruginosa* Wzz2. *J Biol Chem* **286**, 44277-44284.

558 **Larue, K., Kimber, M. S., Ford, R. & Whitfield, C. (2009).** Biochemical and structural analysis of
559 bacterial O-antigen chain length regulator proteins reveals a conserved quaternary structure. *J Biol*
560 *Chem* **284**, 7395-7403.
561

562 **Lugtenberg, B., Meijers, J., Peters, R., van der Hoek, P. & van Alphen, L. (1975).** Electrophoretic
563 resolution of the "major outer membrane protein" of *Escherichia coli* K12 into four bands. *FEBS Lett*
564 **58**, 254-258.
565

566 **Marolda, C. L., Haggerty, E. R., Lung, M. & Valvano, M. A. (2008).** Functional analysis of
567 predicted coiled-coil regions in the Escherichia coli K-12 O-antigen polysaccharide chain length
568 determinant Wzz. *J Bacteriol* **190**, 2128-2137.
569

570 **Morona, R., Van Den Bosch, L. & Daniels, C. (2000).** Evaluation of Wzz/MPA1/MPA2 proteins
571 based on the presence of coiled-coil regions. *Microbiology* **146**, 1-4.
572

573 **Morona, R., Purins, L., Tocilj, A., Matte, A. & Cygler, M. (2009).** Sequence-structure relationships
574 in polysaccharide co-polymerase (PCP) proteins. *Trends Biochem Sci* **34**, 78-84.
575

576 **Murray, G. L., Attridge, S. R. & Morona, R. (2003).** Regulation of *Salmonella typhimurium*
577 lipopolysaccharide O antigen chain length is required for virulence; identification of FepE as a second
578 Wzz. *Mol Microbiol* **47**, 1395-1406.
579

580 **Murray, G. L., Attridge, S. R. & Morona, R. (2005).** Inducible serum resistance in *Salmonella*
581 *typhimurium* is dependent on *wzz*(*fepE*)-regulated very long O antigen chains. *Microbes Infect* **7**, 1296-
582 1304.
583

584 **Najdenski, H., Golkocheva, E., Vesselinova, A., Bengoechea, J. A. & Skurnik, M. (2003).** Proper
585 expression of the O-antigen of lipopolysaccharide is essential for the virulence of *Yersinia*
586 *enterocolitica* O:8 in experimental oral infection of rabbits. *FEMS Immunol Med Microbiol* **38**, 97-106.
587

588 **Papadopoulos, M. & Morona, R. (2010).** Mutagenesis and chemical cross-linking suggest that Wzz
589 dimer stability and oligomerization affect lipopolysaccharide O-antigen modal chain length control. *J*
590 *Bacteriol* **192**, 3385-3393.
591

592 **Purins, L., Van Den Bosch, L., Richardson, V. & Morona, R. (2008).** Coiled-coil regions play a role
593 in the function of the *Shigella flexneri* O-antigen chain length regulator WzzpHS2. *Microbiology* **154**,
594 1104-1116.
595

596 **Raetz, C. R. & Whitfield, C. (2002).** Lipopolysaccharide endotoxins. *Annu Rev Biochem* **71**, 635-700.

597

598 **Samuel, G. & Reeves, P. (2003).** Biosynthesis of O-antigens: genes and pathways involved in
599 nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr Res* **338**, 2503-2519.
600

601 **Tocilj, A., Munger, C., Proteau, A. & other authors (2008).** Bacterial polysaccharide co-
602 polymerases share a common framework for control of polymer length. *Nat Struct Mol Biol* **15**, 130-
603 138.
604

605 **Van Den Bosch, L., Manning, P. A. & Morona, R. (1997).** Regulation of O-antigen chain length is
606 required for *Shigella flexneri* virulence. *Mol Microbiol* **23**, 765-775.
607

608 **Van Den Bosch, L. & Morona, R. (2003).** The actin-based motility defect of a *Shigella flexneri rmlD*
609 rough LPS mutant is not due to loss of IcsA polarity. *Microb Pathog* **35**, 11-18.
610

611 **Zhang, L., Radziejewska-Lebrecht, J., Krajewska-Pietrasik, D., Toivanen, P. & Skurnik, M.**
612 **(1997).** Molecular and chemical characterization of the lipopolysaccharide O-antigen and its role in the
613 virulence of *Yersinia enterocolitica* serotype O:8. *Mol Microbiol* **23**, 63-76.
614
615
616

617

TABLE 1. Bacterial strains and plasmids

Strain/plasmid	Description	Source/Reference
XL10 Gold	*end*A1 *gln*V44 *rec*A1 *thi*-1 *gyr*A96 *rel*A1 *lac* Hte Δ(*mcr*A)183 Δ(*mcr*CB-*hsd*SMR-*mrr*)173 *tet*[R] F'[*pro*AB *lac*I[q]ZΔM15 Tn10(Tet[R] Amy Cm[R])]	Stratagene
RMA4053	*S. flexneri* PE860 serotype Y *wzz::kan*[r], cured of virulence plasmid and pHS-2, carrying pCDFDuet-1	Laboratory stock
Plasmids		
pCDFDuet-1	expression vector carrying *lacI* [q], Sm[R]	Novagen
pQE30	IPTG inducible, expression vector, Amp[R]	Qiagen
pQE30-*wzz*FepE	pQE30 with *E. coli* O157:H7 *fepE* gene, Amp[R]	Tocilj *et al.* (2008)
pRMET1	pQE30::*fepE* (F111 --> P) encoding F111P	This study
pRMET2	pQE30::*fepE* (F111 --> G) encoding F111G	This study
pRMET3	pQE30::*fepE* (V114 --> P) encoding V114P	This study
pRMET4	pQE30::*fepE* (L168 --> D) encoding L168D	This study
pRMET5	pQE30::*fepE* (L168 --> Q) encoding L168Q	This study
pRMET6	pQE30::*fepE* (T170 --> D) encoding T170D	This study
pRMET7	pQE30::*fepE* (D268 --> R) encoding D268R	This study
pRMET8	pQE30::*fepE* (D268 --> L) encoding D268L	This study
pRMET9	pQE30::*fepE* (F111 --> V) encoding F111V	This study
pRMET10	pQE30::*fepE* (L168 --> A) encoding L168A	This study
pRMET11	pQE30::*fepE* (L168 --> R) encoding L168R	This study
pRMET12	pQE30::*fepE* (T170 --> C) encoding T170C	This study
pRMET13	pQE30::*fepE* (T170 --> R) encoding T170R	This study
pRMET14	pQE30::*fepE* (D268 --> V) encoding D268V	This study
pRMET15	pQE30::*fepE* (G274 --> W) encoding G274W	This study
pRMET16	pQE30::*fepE* (D95 --> C) encoding D95C	This study
pRMET17	pQE30::*fepE* (E133 --> C) encoding E133C	This study
pRMET18	pQE30::*fepE* (R208 --> C) encoding R208C	This study
pRMET19	pQE30::*fepE* (Q232 --> C) encoding Q232C	This study
pRMET20	pQE30::*fepE* (D315 --> C) encoding D315C	This study
pRMET21	pQE30::*fepE* (K201 --> C) encoding K201C	This study
pRMET22	pQE30::*fepE* (C52 --> Q) encoding C52Q	This study
pRMET23	pQE30::*fepE* (C52 --> W) encoding C52W	This study
pRMET24	pQE30::*fepE* (C52 --> P) encoding C52P	This study
pRMET25	pQE30::*fepE* (C52 --> S) encoding C52S	This study
pRMET26	pQE30::*fepE* (C354 --> S) encoding C354S	This study
pRMET27	pQE30::*fepE* (C354 --> D) encoding C354D	This study
pRMET28	pQE30::*fepE* (C354 --> G) encoding C354G	This study
pRMET29	pQE30::*fepE* (C354 --> A) encoding C354A	This study
pRMET30	pQE30::*fepE* (C52 --> S, C354 --> A) encoding C52S/C354A	This study
pRMET38	pQE30::*fepE* (D268 --> Y) encoding D268Y	This study
pRMET39	pQE30::*fepE* (D268 --> G) encoding D268G	This study

pRMET40	pQE30::*fepE* (L168 --> P) encoding L168P	This study
pRMET41	pQE30::*fepE* (C52 --> S, C354 --> A, E133 --> C) encoding C52S/C354A/E133C	This study
pRMET42	pQE30::*fepE* (D268 --> E) encoding D268E	This study
pRMET43	pQE30::*fepE* (D268 --> N) encoding D268N	This study

620

Table 2

**TABLE 2. Summary of His$_6$-FepE mutants LPS phenotypes and protein detection**

His$_6$-FepE mutant	Location on 3D crystal structure*	LPS modal length†	Protein detection‡	Apparent oligomer size (kDa)§	
				Before	After
FepE		VL	+++	43, 86	43, 86, >180
pQE30		Non-modal	-	nd	nd
F111V	between α2 & α3 (in)	<VL	+++	43	43, 86, >180
F111P	between α2 & α3 (in)	Non-modal	-	43	43
F111G	between α2 & α3 (in)	<VL	+	43	43
V114P	beginning of α3 (in)	VL	+++	43	43, 86, >180
L168D	between β2 & β3 (in)	18 - <VL	+++	43, 86	43, 86, >180
L168Q	between β2 & β3 (in)	18 - <VL	+++	43	43, 86, >180
L168A	between β2 & β3 (in)	18 - <VL	+++	43, 86	43, 86, >180
L168R	between β2 & β3 (in)	14 - 28	+++	43	43, 86, >180
L168P	between β2 & β3 (in)	18 - <VL	+++	43	43, 86, >180
T170C	between β2 & β3 (in)	<VL	+++	43, 86, 115-182	43, 86, 115-182, >180
T170R	between β2 & β3 (in)	VL	+++	43	43, 86, >180
T170D	between β2 & β3 (in)	<VL	+++	43, 86	43, 86, >180
D268Y	between α6 & α7 (in)	3 - 14	+++	43	43, 86, >180
D268G	between α6 & α7 (in)	10 - 22	+++	30, 43, 86	43, 86, >180
D268E	between α6 & α7 (in)	VL	+++	43	43, 86, >180
D268N	between α6 & α7 (in)	9 - 18	+++	43	43, 86, >180
D268R	between α6 & α7 (in)	10 - <VL	+++	30, 43	30, 43, 86, >180
D268L	between α6 & α7 (in)	7 - 15	+++	43	43, 86, >180
D268V	between α6 & α7 (in)	7 - 15	+++	43	43
G274W	beginning of α7 (in)	Non-modal	-	30, 43	43
His6-FepE-TM mutants					
C52Q	undefined (TM1)	VL	++	nd	nd
C52W	undefined (TM1)	VL	+++	nd	nd
C52P	undefined (TM1)	VL	++	nd	nd
C52S	undefined (TM1)	VL	+++	nd	nd
C354S	undefined (TM2)	VL	+++	nd	nd
C354D	undefined (TM2)	VL	+++	nd	nd
C354G	undefined (TM2)	VL	+++	nd	nd
C354A	undefined (TM2)	VL	+++	nd	nd
C52S/C354A	refer above	VL	+++	43, 86	43, 86
C52S/C354A/E133C		VL	+++	43, 86, 115-182, >180	43, 86, 115-182, >180
His6-FepE-Cys mutants					
D95C	between α1 & α2 (in)	<VL	+++	nd	nd
E133C	undefined (out)	VL	+++	43, 86, 115-182, >180	43, 86, 115-182, >180
K201C	α6 helix (out)	<VL	+++	nd	nd
R208C	α6 helix (out)	20 - <VL	+++	nd	nd
Q232C	α6 helix (in)	VL	+++	43, 86, 115-182, >180	43, 86, 115-182, >180
D315C	between α8 & β4 (out)	VL	+++	43, 86	43, 86, >180

* Based on FepE crystal structure PDB 3b8n & Fig. 1; in, inside; out, outside; TM1/TM2, predicted TM regions 1 or 2
† Average length of Oag RUs; VL, very long LPS (>80 Oag RUs); <VL, less than very long LPS; >VL, more than very long LPS
‡ +++, wild type; +, less than wild type; -, not detected.
§ Bands detected before and after cross-linking with 500μM APDP; 115-182, intermediate band in text; >180, HMW band in text; nd, not done
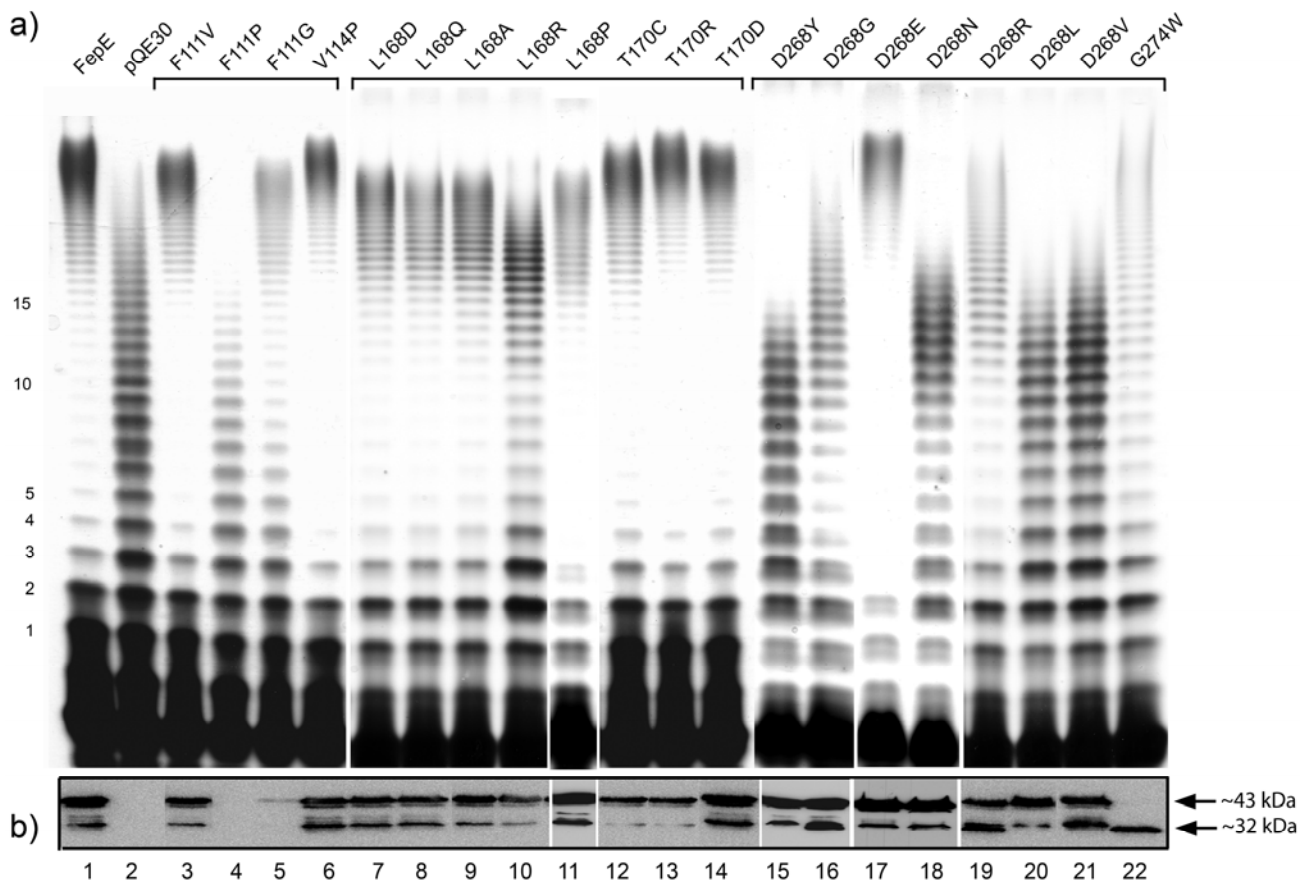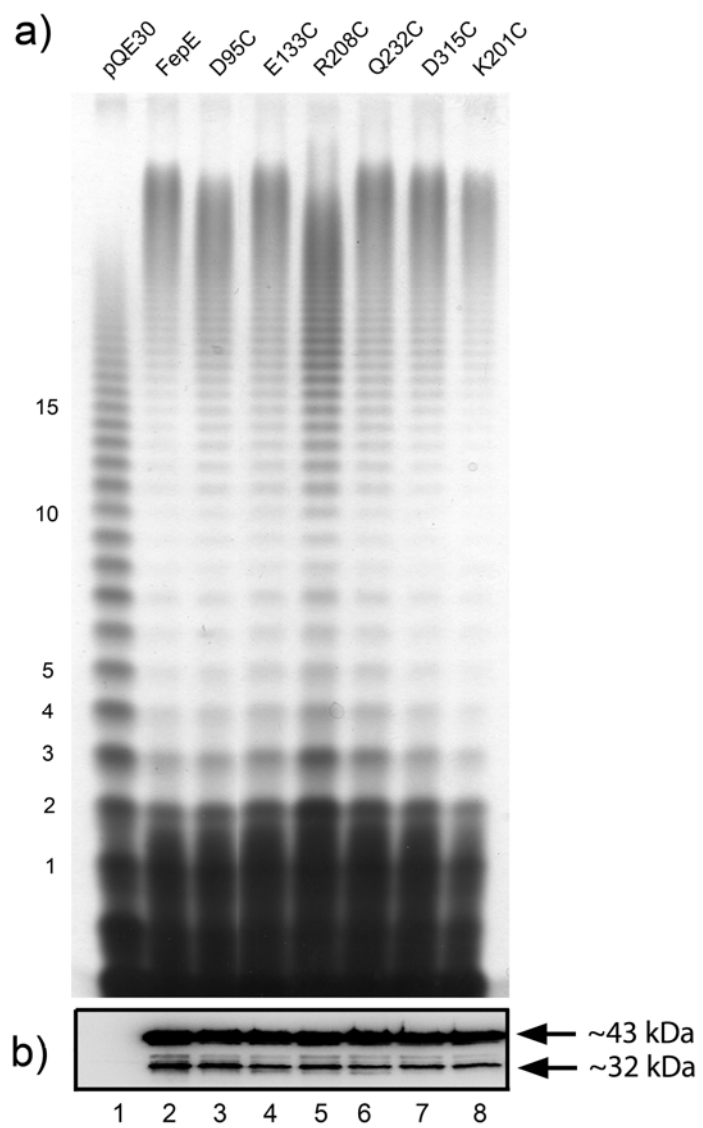
623

624

Figure 1

629   Figure 2
630

631
632

34

a)
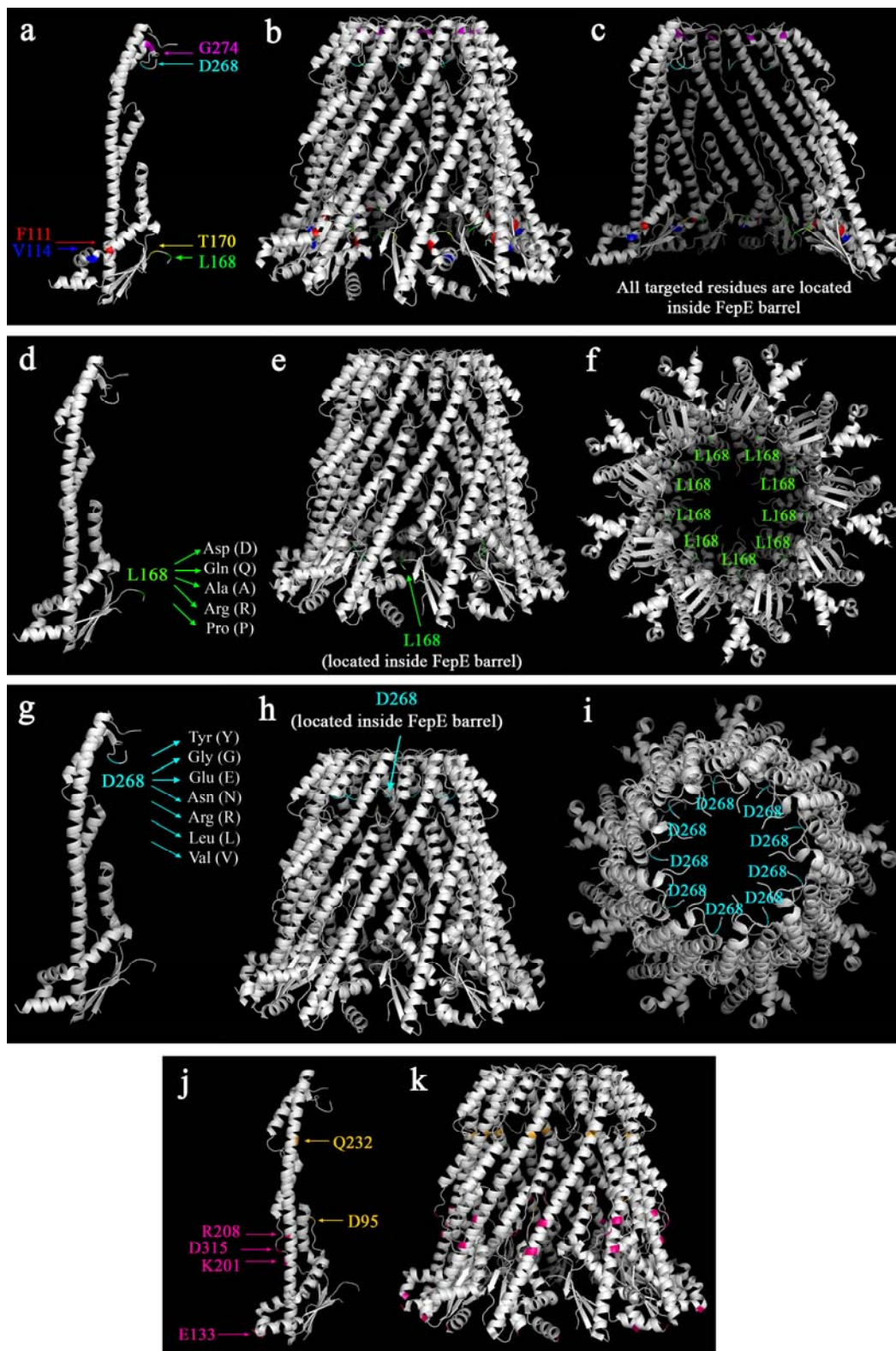
Figure 4
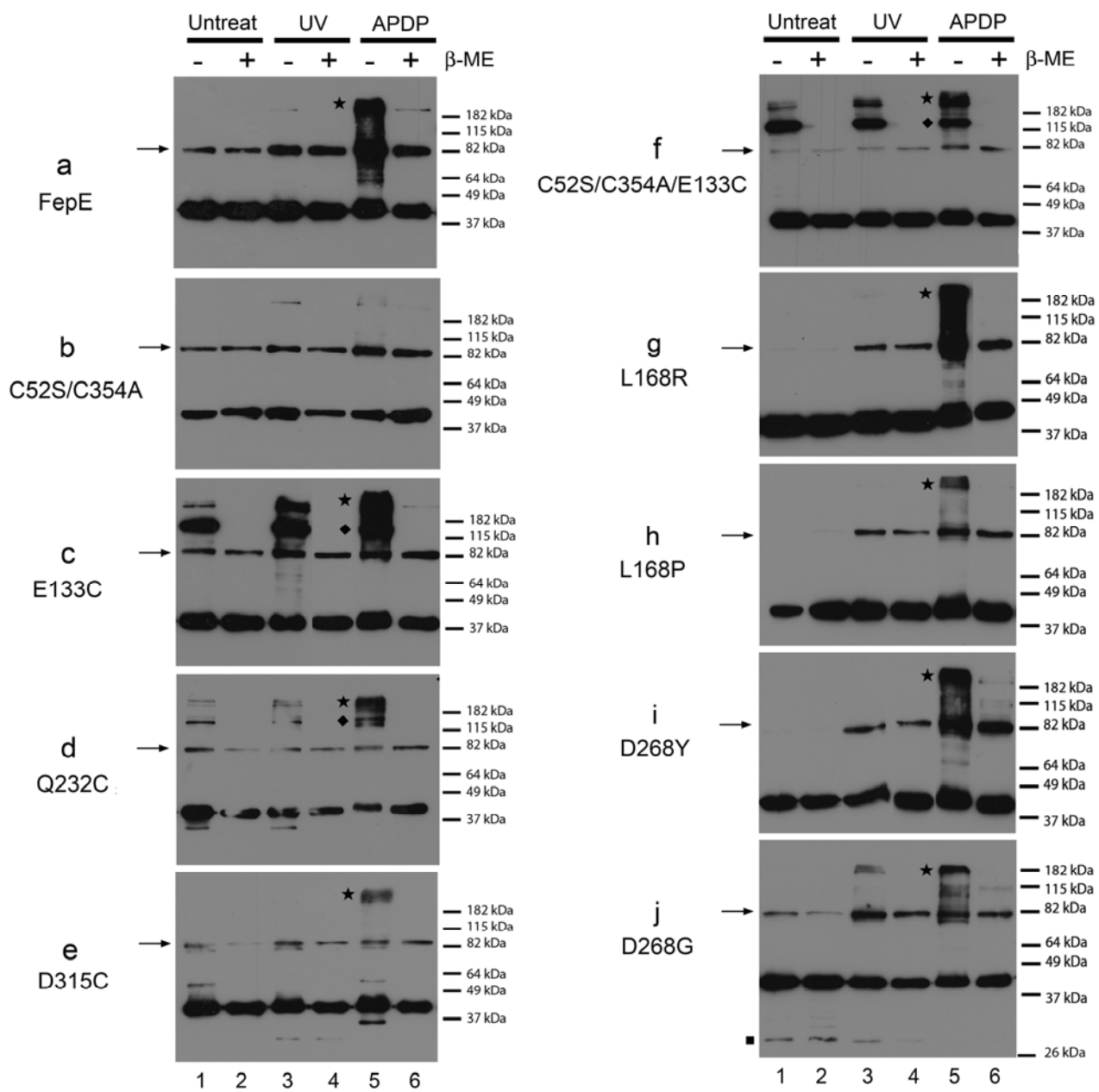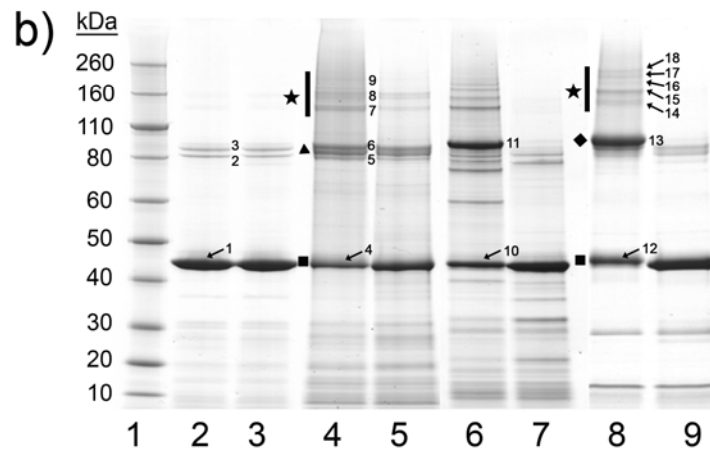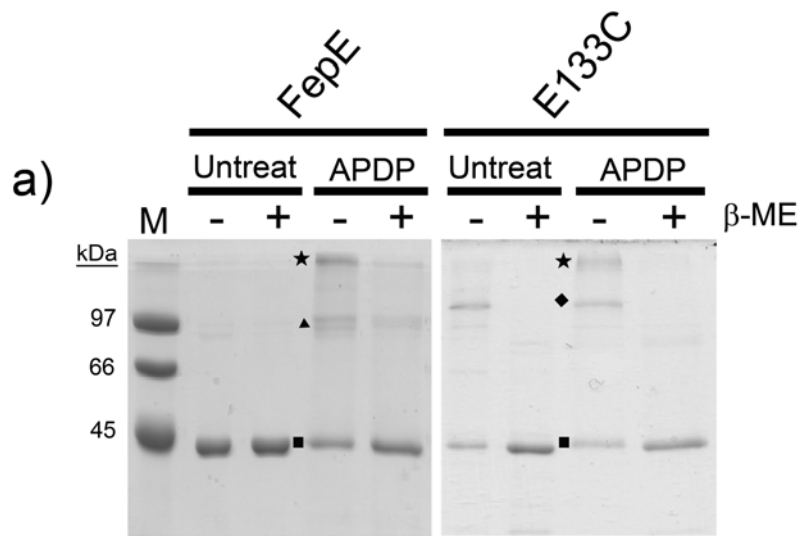
641   Figure 5

642

643
644

37

645 Figure 6

646

647
648

FIGURE LEGENDS

650 **Fig. 1. Regions in FepE targeted for mutation.** (a) The aa sequence alignment of FepE (NP_286314)

651 and WzzB$_{SF}$ (X71970) adapted from Tocilj *et al*. (2008) showing conserved residues (black shading),

652 residues with strongly similar properties (yellow shading), TM regions (grey boxed) and the location of

653 selected α helices and β strands near putative areas predicted to affect FepE function in determining

654 Oag modal chain length (circled in green). The aa positions predicted to affect function (F111, V114,

655 L168, T170, D268 and G274) are indicated by red arrows, the positions targeted for directed Cys

656 substitutions (D95, E133, K201, R208, Q232, D315) by purple arrows, and the TM region Cys are

657 indicated by blue arrows; (b) Topology diagram of FepE monomer (PDB 3b8n) showing the same

658 putative areas of interest (circled in green) (adapted from Tocilj *et al.* (2008)).

659

660 **Fig. 2. Location of residues targeted for mutation on the 3D structures of FepE.** The location of

661 the targeted residues for mutation are mapped on the 3D structures of FepE (PDB 3b8n). Residues

662 F111 (red), V114 (blue), L168 (green), T170 (yellow), D268 (cyan) and G274 (magenta) mapped on

663 (a) FepE monomer and (b & c) FepE nonamer (side view & mid-section view, respectively). L168

664 (green) mapped on (d) FepE monomer (with its aa substitutions highlighted in white), and (e & f) FepE

665 nonamer (side view & bottom view, respectively). D268 (cyan) mapped on (g) FepE monomer (and its

666 aa substitutions highlighted in white), and (h & i) FepE nonamer (side view & bottom view,

667 respectively). Residues targeted for Cys substitution inside the FepE oligomer (D95 and Q232 in

668 orange) and outside the FepE oligomer (E133, K201, R208 and D315 in pink) are mapped on (j) FepE

669 monomer, and (k) FepE nonamer (side view).

670

671 **Fig. 3. Analysis of LPS profile conferred by FepE mutants.** (a) LPS was isolated and detected from

672 whole cell lysates of *S. flexneri* strains expressing FepE, pQE30 and His$_6$-FepE mutant proteins

673   (indicated above). The first 15 Oag RUs are indicated on the side of the gel. Mutants containing aa

674   substitutions in similar putative regions are grouped. Each lane contains ~2 x 10^8 bacterial cells of each

675   strain; (b) Western blots on whole cell lysates expressing the above proteins were probed with rabbit

676   anti-FepE antiserum. The size of the full length His_6-FepE protein (~43 kDa) and the degraded/altered

677   His_6-FepE (~32 kDa) are indicated. Each lane contains 5 x 10^7 bacterial cells of each strain.

678

679   **Fig. 4. Analysis of LPS profile conferred by His_6-FepE-Cys mutants.** (a & c) LPS was isolated and

680   detected from whole cell lysates of *S. flexneri* strains expressing FepE, pQE30 and mutant His_6-FepE-

681   Cys proteins (indicated above). The first 15 Oag RUs are indicated on the side of the gel. Each lane

682   contains ~2 x 10^8 bacterial cells of each strain; (b & d) Western blots on whole cell lysates expressing

683   the above proteins were probed with rabbit anti-FepE antiserum. The size of the full length His_6-FepE

684   protein (~43 kDa) and the degraded/altered His_6-FepE protein (~32 kDa) are indicated. Each lane

685   contains 5 x 10^7 bacterial cells of each strain.

686

687   **Fig. 5. Chemical cross-linking of His_6-FepE-Cys mutants.** *S. flexneri* strains expressing FepE and

688   His_6-FepE-Cys mutant proteins as indicated were prepared and incubated with 500 μM APDP and

689   exposure to UV (APDP); no APDP and exposure to UV only (UV); and no APDP and UV exposure

690   (Untreat). Samples were mixed in sample buffer with (+) and without (-) β-ME. Western blots were

691   probed with rabbit anti-FepE antiserum. The HMW band (black star) and intermediate species between

692   115 - 182 kDa (♦) are indicated in APDP treated samples (lane 5). The position of the ~86 kDa

693   (dimeric) form is indicated by black arrows on the side of the gels (lane 1). The ~30 kDa band in

694   untreated sample of D268G is indicated by a black square. Each lane contains ~5 x 10^7 bacterial cells.

695

696   **Fig. 6. Analysis of purified FepE and E133C proteins by MS.** Protein were affinity purified from

697   untreated and APDP treated samples of *S. flexneri* strains expressing His$_6$-FepE and His$_6$-FepE$_{E133C}$.

698   All samples were mixed in sample buffer with (+) and without (-) β-ME, heated at 60°C for 5 mins, and

699   electrophoresed on SDS 12% polyacrylamide gels (a) and a SDS 4-12% polyacrylamide gel (b),

700   followed by Coomassie staining. The ~43 kDa (monomeric) form (■), ~86 kDa (dimeric) form (▲),

701   intermediate species between 115 - 182 kDa (♦), and separated HMW bands between 140 – 260 kDa

702   (black star) are indicated in APDP treated samples (lanes 4 & 8). Protein bands excised for MS analysis

703   are numbered as indicated and summarized in Table S2.

41

SUPPLEMENTARY FIGURE LEGENDS

Fig. S1. Analysis of LPS profile conferred by His$_6$-FepE TM mutants. Whole cell lysates of *S. flexneri* strains expressing FepE, pQE30 and mutant His$_6$-FepE TM proteins as indicated were proteinase-K treated and electrophoresed on a SDS 15% polyacrylamide gel, followed by detection of LPS by silver-staining (Murray *et al.*, 2003). The first 15 Oag RUs are indicated on the side of the gel. Each lane contains ~2 x 10^8 bacterial cells of each strain.

Fig. S2. **Chemical cross-linking of His$_6$-FepE mutants.** *S. flexneri* strains expressing FepE and His$_6$-FepE mutant proteins as indicated were harvested, resuspended in NaPO$_4$ buffer and disrupted by sonication, then incubated with 500 μM APDP and exposure to UV (APDP); no APDP and exposure to UV only (UV); and no APDP and UV exposure (Untreat). Samples were mixed in sample buffer with (+) and without (-) β-ME, heated at 60°C for 5 mins, and electrophoresed on a SDS 12% polyacrylamide gel followed by Western immunoblotting with anti-FepE antibodies. The HMW band (black star) and the intermediate species between 115 - 182 kDa (♦) are indicated in APDP treated samples (lane 5). The position of the ~86 kDa (dimeric) form is indicated by black arrows on the side of the gels (lane 1). The ~30 kDa band in untreated samples of D268R and G274W is indicated by a black square. Each lane contains ~5 x 10^7 bacterial cells.

Fig. S3. FepE and E133C protein sequence showing identified peptides. The aa sequences of FepE and E133C correspond to MASCOT entries FEPE_ENHT and FEPE_ENHTC (modified to contain modified Cys at position E133), respectively. The peptides identified from MS for the

monomeric species in untreated samples of FepE and E133C, Bands 1 and 10 respectively, are

mapped on the aa sequence as indicated in red.

Fig. S4. Proposed model of Oag chain length regulation by FepE.

Supplementary Table S1

Table S1. Oligonucleotides used in this study

Primer name	Oligonucleotide sequence (5' - 3')*	Target†	AA (and nt) change‡
ETR1	CCTGTTTATCAAGAAG**NNN**CAGTCGGTTAGCTTGCTGG	F111	V(GTA), P(CCC), G(GGT)
ETR2	Reverse complement of ETR1		
ETR3	CAAGAAGTTTCAGTCG**CCG**AGCTTGCTGGAAGAG	V114	P(CCG)
ETR4	Reverse complement of ETR3		
ETR5	AAGATGAACCGTCA**NNN**TATACCTCCTGGACGC	L168	D(GAC), Q(CAG), A(GCG), R(CGG), P(CCA)
ETR6	Reverse complement of ETR5		
ETR7	GATGAACCGTCACTGTAT**NNN**TCCTGGACGCTAAG	T170	C(TGT), R(CGT), D(GAC)
ETR8	Reverse complement of ETR7		
ETR9	CGTTAAAGATGACCCC**NNN**TTCTCTATTTCTCTCGGC	D268	Y(TAC), G(GGA), E(GAA), N(AAT), R(CGT), L(CTC), V(GTG)
ETR10	Reverse complement of ETR9		
ETR11	GATTTCTCTATTTCTCTC**TGG**GCAGACGGTATTGAACGC	G274	W(TGG)
ETR12	Reverse complement of ETR11		
ETR13	CTTCGTGTGCTGGATCTG**TGT**ATCAAAATTGATCGTACA	D95	C(TGT)
ETR14	Reverse complement of ETR13		
ETR15	GTGATGGACCAATTAAAA**TGT**GCGAAAATCGACGAACTG	E133	C(TGT)
ETR16	Reverse complement of ETR15		
ETR17	CTCTGCGTTGGTGGTG**TGT**GAGTCGATAGAAAACG	K201	C(TGT)
ETR18	Reverse complement of ETR17		
ETR19	AGAGTCGATAGAAAACGTC**TGT**AATAAACTGGAGATCAAA	R208	C(TGT)
ETR20	Reverse complement of ETR19		
ETR21	CCGCATTAAAATGAAAAT**TGT**CTTGATGCAAACATTCAGCGC	Q232	C(TGT)
ETR22	Reverse complement of ETR21		
ETR23	CAGTTAACAAAAGCAAATATCAAC**TGT**GTGAATTTTACGCCG	D315	C(TGT)
ETR24	Reverse complement of ET23		
ETR25	CGTTTTTGCGTTTGCC**NNN**GCAGGCTTGCTGATCT	C52	Q(CAA), W(TGG), P(CCG), S(TCT)
ETR26	Reverse complement of ET25		
ETR27	CGTTTTTGCGTTTGCC**TCT**GCAGGCTTGCTGATCT	C52	S(TCT)
ETR28	Reverse complement of ET27		
ETR29	CGGGATGGTGGCT**NNN**GGTAGCGTGTTATTGCG	C354	S(AGC),

3

ETR30	Reverse complement of ET29		D(GAT), G(GGG), A(GCT)
ETR31	CGGGATGGTGGCT**GCT**GGTAGCGTGTTATTGCG	C354	A(GCT)
ETR32	Reverse complement of ETR31		
ET11	CCCGAAAAGTGCCACCTG	pQE30	
ET12	GGTCATTACTGGAGTCTTG	pQE30	
ET13	CGGATAACAATTTCACACAG	pQE30	

* Underlined sequences indicate the nucleotides that undergo site directed mutagenesis

† Residues targeted for mutation in FepE

‡ Amino acid (AA) and nucleotide (nt) changes made at targeted residues

Supplementary Table S2

TABLE S2. Mass Spectrometry analysis of FepE & E133C

Band number*	Sample	Description*	Gene†	No. of peptides identified	% FepE sequence coverage‡
FepE					
Band 1	Untreated	Monomeric species	FepE	217	45%
Band 2	Untreated	Dimeric species	FepE	20	31%
Band 3	Untreated	Dimeric species	FepE	26	31%
Band 4	APDP	Monomeric species	FepE	56	44%
Band 5	APDP	Dimeric species - lower band	FepE	108	41%
Band 6	APDP	Dimeric species - upper band	FepE	102	48%
Band 7	APDP	HMW species	FepE	32	28%
Band 8	APDP	HMW species	FepE	41	31%
Band 9	APDP	HMW species	FepE	23	28%
E133C					
Band 10	Untreated	Monomeric species	E133C	244	48%
Band 11	Untreated	Intermediate species	E133C	305	51%
Band 12	APDP	Monomeric species	E133C	285	45%
Band 13	APDP	Intermediate species	E133C	56	45%
Band 14	APDP	HMW species	E133C	59	60%
Band 15	APDP	HMW species	E133C	57	60%
Band 16	APDP	HMW species	E133C	54	50%
Band 17	APDP	HMW species	E133C	46	51%
Band 18	APDP	HMW species	E133C	48	50%

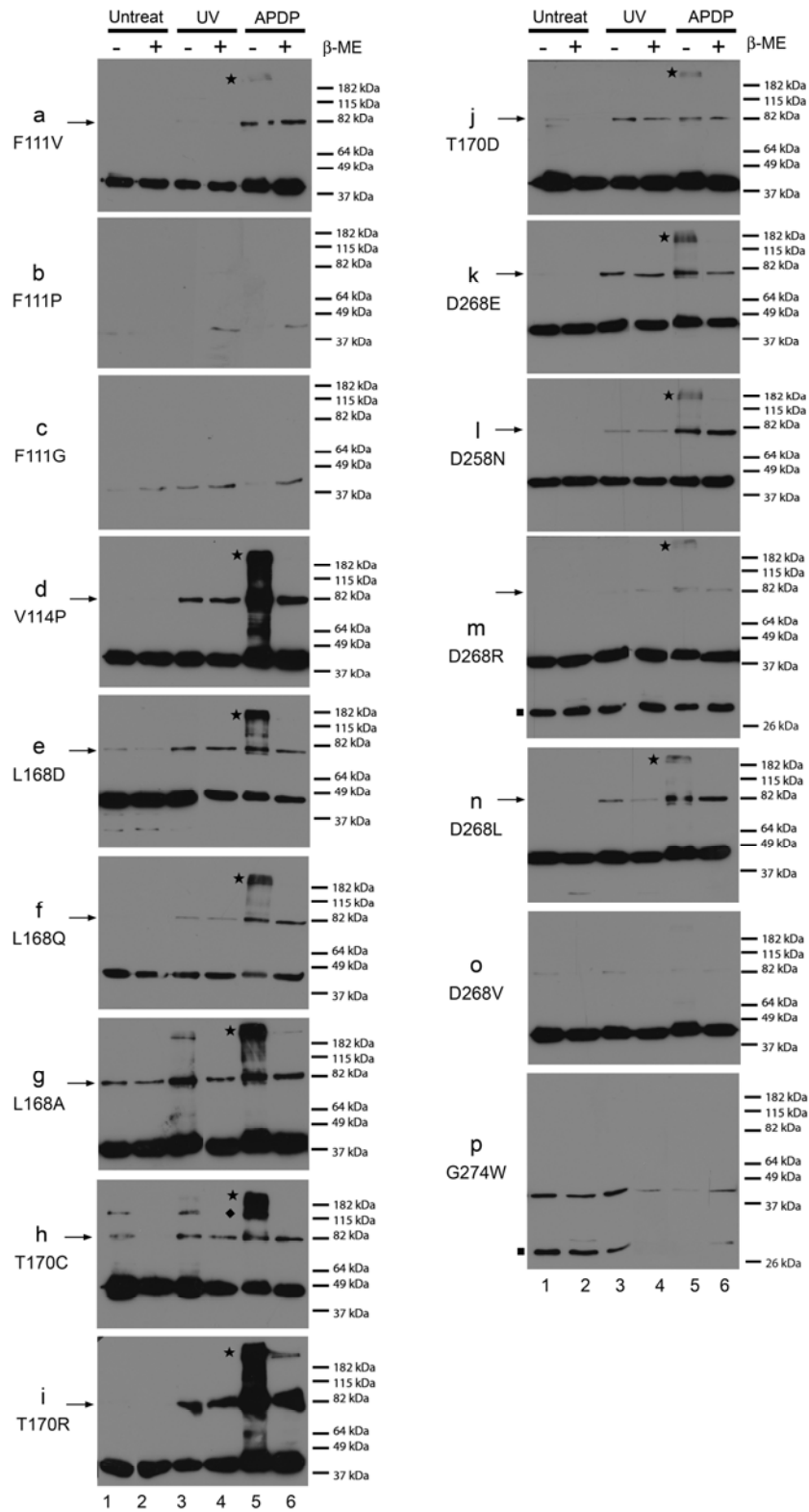* Bands excised for MS analysis are numbered as shown in Fig. 6

† aa sequences FepE and E133C corresponds to MASCOT entries FEPE_ENHT and FEPE_ENHTC (modified to contain modified Cys at position E133), respectively

‡ % sequence coverage of identified peptides on FepE or E133C aa sequence

Supplementary Figure S1
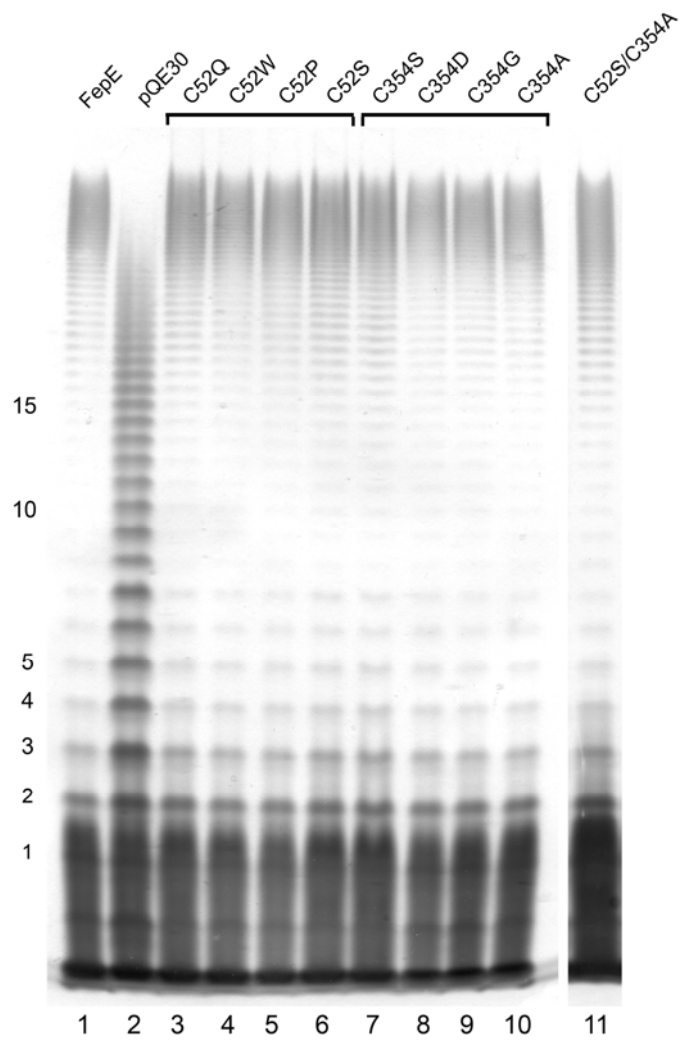
Supplementary Figure S2

Supplementary Figure S3

Band 1

Match to: FEPE_ENHT Score: 1192

Elizabeth Tran Translation of Escherichia coli O157:H7 FepE

Nominal mass (M_r): 42222; Calculated pI value: 6.08

Fixed modifications: Carbamidomethyl (C)

Variable modifications: Oxidation (M)

Cleavage by Trypsin: cuts C-term side of KR unless next residue is P

Sequence Coverage: 45%

Matched peptides shown in Red

```
  1 MSSLNIKQGS DAHFPDYPLA SPSNNEIDLL NLISVLWRAK KTVMAVVFAF
 51 ACAGLLISFI LPQKWTSAAV VTPPEPVQWQ ELEKTFTKLR VLDLDIKIDR
101 TEAFNLFIKK FQSVSLLEEY LRSSPYVMDQ LKEAKIDELD LHRAIVALSE
151 KMKAVDDNAS KKKDEPSLYT SWTLSFTAPT SEEAQTVLSG YIDYISALVV
201 KESIENVRNK LEIKTQFEKE KLAQDRIKMK NQLDANIQRL NYSLDIANAA
251 GIKKPVYSNG QAVKDDPDFS ISLGADGIER KLEIEKAVTD VAELNGELRN
301 RQYLVEQLTK ANINDVNFTP FKYQLSPSLP VKKDGPGKAI IVILSALIGG
351 MVACGSVLLR YAMASRKQDA MMADHLV
```

Band 10

Match to: FEPE_ENHTC Score: 5169

Cys-Modified Elizabeth Tran translation E133C

Nominal mass (M_r): 42285; Calculated pI value: 6.39

Fixed modifications: Carbamidomethyl (C)

Variable modifications: Oxidation (M)

Cleavage by Trypsin: cuts C-term side of KR unless next residue is P

Sequence Coverage: 48%

Matched peptides shown in Red

```
  1 MSSLNIKQGS DAHFPDYPLA SPSNNEIDLL NLISVLWRAK KTVMAVVFAF
 51 ACAGLLISFI LPQKWTSAAV VTPPEPVQWQ ELEKTFTKLR VLDLDIKIDR
101 TEAFNLFIKK FQSVSLLEEY LRSSPYVMDQ LKCAKIDELD LHRAIVALSE
151 KMKAVDDNAS KKKDEPSLYT SWTLSFTAPT SEEAQTVLSG YIDYISALVV
201 KESIENVRNK LEIKTQFEKE KLAQDRIKMK NCLDANIQRL NYSLDIANAA
251 GIKKPVYSNG QAVKDDPDFS ISLGADGIER KLEIEKAVTD VAELNGELRN
301 RQYLVEQLTK ANINDVNFTP FKYQLSPSLP VKKDGPGKAI IVILSALIGG
351 MVACGSVLLR YAMASRKQDA MMADHLV
```

Supplementary Figure S4

Oag RUs undergo polymerisation by Wzy

The resulting nascent Oag-linked undecaprenyl pyrophosphate chain is threaded into FepE

Once filling has occurred, a conformation change in the FepE oligomer releases the Oag modal length chain

The Oag polymer is then attached to the pre-formed lipid A-core molecule by WaaL and transported to the OM

Wzy

WaaL

IM