

1 **The Intimin periplasmic domain mediates dimerisation and binding to peptidoglycan**

2

3 Jack C. Leo^{1,5}, Philipp Oberhettinger², Manish Chaubey¹, Monika Schütz², Daniel Kühner³, Ute Bertsche³,
4 Heinz Schwarz¹, Friedrich Götz³, Ingo B. Autenrieth², Murray Coles¹, Dirk Linke^{1,4*}

5

6 ¹Max Planck Institute for Developmental Biology, Department of Protein Evolution, 72076 Tübingen,
7 Germany

8 ²Interfaculty Institute for Microbiology and Infection Medicine, University Clinics Tübingen, 72076
9 Tübingen, Germany

10 ³Department of Microbial Genetics, University of Tübingen, 72076 Tübingen, Germany

11 ⁴Department of Biosciences, University of Oslo, 0316 Oslo, Norway

12 ⁵Current address: Department of Biosciences, University of Oslo, 0316 Oslo, Norway

13 *Corresponding author. E-mail: dirk.linke@ibv.uio.no

14

15 **Running title:** Functions of the Intimin periplasmic domain

16 **Keywords:** Intimin/Invasin/inverse autotransporter/lysin motif/peptidoglycan

17

18

19 **SUMMARY**

20 Intimin and Invasin are prototypical inverse (Type Ve) autotransporters and important virulence factors
21 of enteropathogenic *Escherichia coli* and *Yersinia spp.*, respectively. In addition to a C-terminal
22 extracellular domain and a β -barrel transmembrane domain, both proteins also contain a short N-
23 terminal periplasmic domain that, in Intimin, includes a lysin motif (LysM), which is thought to mediate
24 binding to peptidoglycan. We show that the periplasmic domain of Intimin – but not the shorter domain
25 of Invasin – does bind to peptidoglycan both *in vitro* and *in vivo*, but only under acidic conditions. We
26 present the solution structure of the Intimin LysM, which has an additional, potentially functionally
27 relevant α -helix compared to other LysMs. In contrast to previous reports, we demonstrate that the
28 periplasmic domain of Intimin mediates dimerisation. Our data suggests that the periplasmic domain
29 contains two dimerisation interfaces. We further show that dimerisation and peptidoglycan binding are
30 general features of LysM-containing inverse autotransporters. The periplasmic domain could be involved
31 in autotransport, and peptidoglycan binding may aid in resisting mechanical and chemical stress during
32 transit through the gastrointestinal tract.

33

34

35 INTRODUCTION

36 An often essential first step in host colonization by bacterial pathogens is the adherence of bacteria to
37 host cells and tissues. This binding is mediated by various adhesins, many of which are proteinaceous
38 molecules expressed on the cell surface. Intimin (Int) is a major adhesin of enteropathogenic and
39 enterohaemorrhagic *Escherichia coli* (EPEC and EHEC) and is instrumental in the formation of actin
40 pedestals leading to attaching and effacing (A/E) lesions on enterocytes (Schmidt, 2010). It is a
41 homologue of Invasin (Inv) from enteropathogenic *Yersinia spp.*, which mediates direct binding to host
42 cells *via* β_1 integrins (Leo and Skurnik, 2011). However, unlike Inv, Int does not bind primarily to a
43 cellular receptor; rather, the Int receptor Tir is produced by the bacteria themselves and transferred to
44 the host cell membrane through the type 3 secretion system (Schmidt, 2010).

45 The extracellular domain of Intimin consists of tandem immunoglobulin (Ig)-like domains capped by a C-
46 type lectin domain (Kelly *et al.*, 1999). The Tir-binding region is located in the C-terminal superdomain
47 consisting of the last Ig domain and the lectin domain (Luo *et al.*, 2000). The extracellular portion of
48 Invasin has a similar structure (Hamburger *et al.*, 1999), and the integrin-binding region is also located at
49 the C-terminal tip of the protein (Leong *et al.*, 1990). The extracellular or passenger domain of both
50 proteins is exported by a type V_e or inverse autotransport mechanism: a 12-stranded transmembrane β -
51 barrel N-terminal to the passenger domain acts as a translocation pore which facilitates the secretion of
52 the passenger domain across the outer membrane (Oberhettinger *et al.*, 2012; Leo *et al.*, 2012; Fairman
53 *et al.*, 2012).

54 In addition to the passenger and translocation domains, both proteins contain a small N-terminal
55 periplasmic domain, also referred to as the α -domain (Tsai *et al.*, 2010). In Int, the periplasmic domain
56 contains a lysin motif (LysM) found in many peptidoglycan-binding proteins (Buist *et al.*, 2008). The LysM
57 is a small domain consisting of a 2-stranded anti-parallel β -sheet packed against two α -helices with the
58 topology β - α - α - β (Bateman and Bycroft, 2000). LysMs usually bind to the N-acetylglucosamine moieties
59 in peptidoglycan (PGN) and chitin, or other carbohydrate structures (Buist *et al.*, 2008). In contrast to
60 Int, the Inv periplasmic region lacks a LysM.

61 Due to the presence of a LysM in the Int periplasmic domain, it has long been suspected that the
62 periplasmic domain would mediate binding to PGN (*e.g.* Bateman and Bycroft, 2000; Tsai *et al.*, 2010;
63 Pisano *et al.*, 2012). However, up to now, this hypothesis has not been experimentally validated. In this
64 study, we show that the periplasmic domain of Int does indeed bind PGN, but only under acidic

65 conditions. In addition, we show that the periplasmic domain mediates dimerisation of Int. In contrast,
66 the shorter periplasmic domain of Inv has neither of these functions. We also present a solution
67 structure of the Int LysM and show that dimerisation and PGN binding are conserved functions of LysMs
68 of inverse autotransporters. Dimerisation appears to be an important feature for the virulence function
69 of these proteins. Our results show that the periplasmic domain of type Ve autotransporters are
70 functional and may be involved in stabilising the cell envelope and virulence functions.

71 RESULTS

72 *Bioinformatic analysis of periplasmic domains from inverse autotransporters*

73 Secondary structure prediction of the Int sequence using software of the online Bioinformatics Toolkit of
74 the Max-Planck Institute for Developmental Biology (<http://toolkit.tuebingen.mpg.de/>) reveals that the
75 periplasmic domain contains the LysM motif, consisting of a small β -sheet packed against two α -helices,
76 followed C-terminally by two closely spaced α -helices (Fig. 1A). There is an unstructured loop of
77 approximately 30 residues linking the LysM domain to the first of the C-terminal helices, which we call
78 the spacer sequence. The periplasmic domain of Inv also contains the helices, but lacks the LysM motif
79 (Fig. 1A). An alignment of selected members of the Inv-Int family periplasmic domains shows that the C-
80 terminal helices are conserved, whereas the LysM motif is only present in some members of the family
81 (Fig. S1). Similar results were also obtained in an earlier study (Tsai *et al.*, 2010).

82 To investigate the distribution of different types of periplasmic domains within inverse autotransporters,
83 we performed clustering analysis of the periplasmic domain sequences using CLANS (Frickey and Lupas,
84 2004). A total of 172 sequences were included in the clustering (Table S1), and after singletons
85 (sequences with *P*-values for BLAST high-scoring segment pairs higher than 0.99) were removed 152
86 sequences remained. The results (Fig. 1B) show that the periplasmic domains cluster largely based on
87 taxonomic distribution rather than length or the presence of the LysM motif. Most of the sequences
88 were derived from the γ -Proteobacteria, especially the Enterobacteriaceae, which form a large cluster.
89 Int is located close to the centre of this cluster, whereas Inv is located more peripherally (Fig. 1B). Other
90 distinct clusters are formed by cyanobacterial sequences and sequences from the β -proteobacterial
91 genus *Bordetella*. Interestingly, periplasmic domains from the cyanobacterial genus *Synechococcus* form
92 two distinct clusters that are only distantly connected to the enterobacterial cluster; however, as the
93 synechococcal sequences are all very short, the separate clusters may be artefactual. The sizes of the
94 sequences do not have a large effect on the clustering, as both large and small sequences are found in

95 the same clusters (Fig. S2A). The LysM appears restricted to the large γ -proteobacterial cluster, as it was
96 not detected in any of the other groups (Fig. 1C). The presence of the LysM domain can to some degree
97 be predicted based on the length of the protein sequence; however, the correlation is not absolute and
98 there are several large members of the family lacking the LysM motif (Fig. S2B). The prevalence of LysMs
99 in large inverse autotransporters suggests that the LysM may be involved in the autotransport process,
100 possibly stabilising translocation intermediates during secretion.

101 ***The Int periplasmic domain mediates dimerisation***

102 Int has been reported to form dimers *via* its β -barrel domain (Touzé *et al.*, 2004). However, the Int β -
103 barrel crystallises as a monomer (Fairman *et al.*, 2012). This suggested to us that the dimerisation might
104 be mediated by the periplasmic domain, which was included in the construct used by Touzé *et al* (2004).
105 To test this, we produced the periplasmic domains of Int (from EPEC) and Inv (from *Y. enterocolitica*) as
106 maltose-binding protein (MBP) fusions, named IntPeri-MBP and InvPeri-MBP, with the MBP as a C-
107 terminal fusion and a hexahistidine tag on the N-termini of the periplasmic domains. The constructs are
108 depicted schematically in Fig. 2A, and all plasmids used in the study are summarised in Table 1. IntPeri-
109 MBP in particular is unstable, and degradation products were observed in cell lysates (data not shown);
110 however, we were able to purify the proteins to high purity by passing the proteins through both a
111 nickel and amylose column followed by size exclusion chromatography (SEC).

112 When we performed analytical SEC at physiological pH (7.4.), MBP and InvPeri-MBP run at the expected
113 sizes of the monomer, whereas IntPeri-MBP gives two peaks: the major peak migrates at an apparent
114 size slightly larger than the dimer and the smaller peak at the expected molecular weight of the
115 monomer (Fig. 2B). Both peaks contain Int-MBP, suggesting an equilibrium between monomeric and
116 dimeric forms. Indeed, when we reran the major peak through the SEC column, a similar distribution for
117 dimer and monomer was observed (Fig. S3A).

118 To further characterise the region in the periplasmic domain responsible for dimerisation, we fused
119 short fragments of the Int periplasmic domain to MBP (Fig. 2A). These fragments were the LysM domain
120 along with the spacer sequence (IntLysM-MBP), the two C-terminal helices (IntPeriHelix1-MBP) or just
121 the second, C-terminal helix (IntPeriHelix2-MBP). In SEC, IntLysM-MBP migrates as a dimer, whereas the
122 other two constructs run as a monomer (Fig. 2C). This suggests that the N-terminal region of the
123 periplasmic domain is the dimerisation interface.

124 Touzé and coworkers showed that the β -barrel and flanking regions (residues 189-550) contains a
125 dimerisation site (Touzé *et al.*, 2004). As the β -barrel itself is monomeric, Fairman *et al.* speculated that
126 the dimerisation interface would either be in the periplasmic region or the first extracellular Ig-like
127 domain (D00) (Fairman *et al.*, 2012). Our results above suggest a dimerisation site in or near the LysM.
128 We therefore produced and purified the D00 domain and tested its oligomerisation status by SEC. The
129 domain ran as a monomer (Fig. S4). Thus, the dimerisation observed by Touzé *et al* must be due to the
130 short stretch preceding the β -barrel domain (residues 189-209). This corresponds to C-terminal helix in
131 the periplasmic domain (Helix2). This helix failed to dimerise when fused to MBP, but it is possible that
132 the helix is not correctly folded outside its native context. We therefore submit that the Int periplasmic
133 domain contains two dimerisation interfaces, the LysM region and Helix2, though the latter only
134 mediates dimer formation when fused to the β -barrel.

135 ***The periplasmic domains of Int and Inv promote surface display of the passenger domain***

136 To investigate the relevance of the periplasmic domain under *in vivo* conditions, we produced deletion
137 mutants of Inv and Int in *E. coli*. We used full-length, wild-type (wt) Int and a construct missing the entire
138 periplasmic domain (Int Δ Peri). We did the same for Inv (Inv wt and Inv Δ Peri), and bacteria with the
139 empty vector (pASK-IBA2) acted as our negative control. In addition, we made two chimaeric constructs:
140 IntPeri-Inv, where the periplasmic domain of Inv was switched to that of Int, and InvPeri-Int, where the
141 Int periplasmic domain was replaced with the Inv periplasmic domain. These constructs are summarized
142 schematically in Fig. 3A.

143 To determine whether the truncations had a major effect on Int biogenesis, we examined the surface
144 display of the Int passenger by both fluorescence-activated cell sorting (FACS) and immunofluorescence
145 microscopy. Both Int wt and Inv wt were exported to the surface (Fig. 3B). The Δ Peri constructs were
146 also surface-exposed but displayed a bimodal distribution, with a part of the population exhibiting lower
147 fluorescence. This may be due to impaired secretion of the passenger domain. In contrast, the
148 chimaeras were more fluorescent than the corresponding wt constructs, suggesting higher expression
149 levels or more efficient surface display. The reason for this is unclear.

150 We also examined the effect of the truncations and domain switching on adhesion and induction of
151 interleukin (IL)-8. For Int, we performed qualitative adhesion assays, where we examined the adherence
152 of the omp2 strain containing the Int constructs to HeLa cells that had been preinfected with a Δ aeA
153 EPEC strain to prime the cells with Tir (Oberhettinger *et al.*, 2012). The wild-type construct adhered

154 strongly to preinfected cells, but not to uninfected cells (Fig. 3C). However, Int Δ Peri showed much lower
155 adhesion than the wt, probably because of reduced surface display, but possibly also because of
156 reduced avidity through the loss of dimer formation (Fig. 3C). The chimaeric construct InvPeri-Int
157 promoted intermediate adhesion; stronger than Int Δ Peri, but not as strong as the wt, despite high
158 surface exposure. We assume this is due to the lack of dimer formation and thus a loss of co-operativity
159 in the Int-Tir interaction (Touzé *et al.*, 2004). The empty vector served as a background control.

160 For the Inv constructs, we measured the induction of IL-8 production in infected cells with an enzyme-
161 linked immunosorbent assay (ELISA). Wild-type Inv induced IL-8 well over background levels, as did the
162 positive control tumor necrosis factor (TNF) (Fig. 3D). The vector control was barely above background
163 levels, and IL-8 production by Inv Δ Peri was reduced compared to the wt. However, IntPeri-Inv showed
164 significantly higher induction of IL-8 than even TNF. This increase is presumably largely due to the Int
165 periplasmic domain mediating dimerisation of Inv, but the higher levels of IntPeri-Inv on the surface may
166 also have an effect (Fig. 3B). In *Y. pseudotuberculosis*, Inv contains an additional extracellular domain
167 (D2) that mediates self-association and leads to more efficient bacterial uptake than in *Y. enterocolitica*
168 (Dersch and Isberg, 1999; Dersch and Isberg, 2000). In the chimaeric IntPeri-Inv, the Int periplasmic
169 domain could thus fulfil a role similar to D2, leading to dimerisation which in turn promotes receptor
170 clustering, more efficient cell invasion and higher IL-8 levels.

171 ***The periplasmic domain of Int, but not Inv, binds peptidoglycan in a pH-dependent manner***

172 We performed pull-down assays with our MBP fusions using purified PGN sacculi from *E. coli* to
173 determine whether the Int periplasmic domain binds PGN (Fig. 4). We fortuitously found that binding
174 was strongest at low pH: in assays performed at pH 5.0, approximately half the IntPeri-MBP fusion
175 precipitates with the sacculi, while the MBP control remain in the supernatant (Fig. 4A). In this assay,
176 Inv-MBP displayed only background-level binding to PGN (Fig. 4A). In the absence of PGN, all three
177 proteins remained soluble. At normal physiological pH (7.4), however, we saw no binding of IntPeri-MBP
178 to PGN, with most of the protein remaining in the supernatant (Fig. 4A).

179 To examine the effect of pH on the binding of IntPeri-MBP to PGN more systematically, we performed
180 binding at range of pH values between 8.0 and 3.0; strong binding first became evident at pH 6.0 and
181 the ratio of bound IntPeri-MBP to soluble protein remains constant even with reduction in pH value (Fig.
182 4B). In the control experiment without PGN sacculi, the protein remains soluble at all pH values tested
183 (Fig. 4B). We observed similar results with a glutathione-S-transferase (GST) fusion of IntPeri; however,

184 GST itself tends to precipitate at low pH, which made interpreting these results difficult (data not
185 shown). Int-MBP still dimerises at pH 4.0 (Fig. S3B), demonstrating that dimerisation is independent of
186 pH.

187 Physiologically abundant divalent cations (Ca^{2+} , Mg^{2+} , Mn^{2+} , Zn^{2+}) might have an effect on the binding of
188 IntPeri to PGN. To check this, we added these ions as chloride salts to the pull-down assay, and in
189 addition we tested the effect of the chelators ethylene diamine tetra-acetic acid (EDTA) and sodium
190 citrate. However, neither the divalent cations nor the chelators had an effect on binding, though some
191 of the ions, particularly Ca^{2+} , slightly modified the pattern of migration in SDS-PAGE (Fig. 4C). We
192 therefore conclude that the binding of IntPeri to peptidoglycan is pH-dependent but does not require
193 divalent ions.

194 We partially digested peptidoglycan sacculi with mutanolysin and performed a competition assay with
195 the soluble digestion products. Addition of the degradation products inhibited the binding of IntPeri-
196 MBP to the sacculi in a dose-dependent manner, though we did not observe complete inhibition of
197 binding to sacculi (Fig. 4D). This demonstrates that soluble muropeptides can compete for binding to
198 IntPeri-MBP. Many LysM-containing proteins also bind chitin (Buist *et al.*, 2008). We therefore ran
199 IntPeri-MBP, InvPeri-MBP and MBP as a control through a chitin column. However, we did not observe
200 binding for any of the constructs (Fig. 4E). This suggests that IntLysM binds to other determinants in PGN
201 than the N-acetylglucosamine – N-acetylmuramic acid backbone.

202 To find out which part of the periplasmic domain of Int is responsible for PGN binding, we performed
203 pull-down assays with IntLysM-MBP, IntPeriHelix1-MBP and IntPeriHelix2-MBP (Fig. 4F). In this assay,
204 only IntLysM-MBP coprecipitated with the sacculi in appreciable amounts, demonstrating that it is
205 indeed the LysM that mediates PGN binding.

206 As we saw some background levels of InvPeri-MBP bound to PGN, we wished to further investigate the
207 relative affinities of InvPeri-MBP and IntPeri-MBP. To this end, we performed pull-down assays using a
208 concentration series of IntPeri-MBP and InvPeri-MBP with a constant amount of PGN. At low
209 concentrations of IntPeri-MBP, almost all the protein is pulled down with the sacculi, whereas the
210 fraction in the pellet decreases steadily with increasing protein concentration (Fig. 5A). InvPeri-MBP
211 follows a similar trend, but even at low concentrations only approximately 50% is pulled down with the
212 sacculi, and at higher concentrations less than 20% remains in the pellet (Fig. 5A). In the reverse
213 experiment, where we varied the concentration of PGN but kept the protein concentration constant,

214 only a small fraction of IntPeri-MBP precipitates with low amounts of PGN, but as the amount of PGN
215 rises, increasingly more IntPeri-MBP is found in the pellet (Fig. 5B). The increase is roughly linear, and at
216 the highest concentration tested (50 μ g), 75% of the protein is in the pellet. Again, InvPeri-MBP follows a
217 similar trend, but the increase in the fraction of pelleted protein is not as large, and at most only 30% of
218 the protein was pulled down with the sacculi (Fig. 5B). It appears that IntPeri-MBP binds with high
219 affinity to PGN, as most of the protein precipitates at low protein concentration. However, the fraction
220 of bound protein declines more or less linearly with increasing concentration of IntPeri-MBP, which we
221 interpret to mean that there are a limited number of high-affinity binding sites in the sacculi and that
222 these are quickly saturated. The amount of IntPeri-MBP scales linearly with increasing PGN, which is
223 consistent with this interpretation. However, also InvPeri-MBP follows similar trends, though the
224 binding levels are much lower than for IntPeri-MBP. When we incubated sacculi with both IntPeri-MBP
225 and InvPeri-MBP, we did not see any effect of high concentrations of InvPeri-MBP on the binding levels
226 of IntPeri-MBP, suggesting that these proteins do not compete for the same binding site(s) and that the
227 binding of InvPeri-MBP is indeed unspecific (Fig. S5). The results obtained using concentration series of
228 either protein or PGN should be viewed with caution, however, as the assays used are rather crude and
229 include washing steps which we did not take into account when calculating the relative binding, and
230 even in the absence of PGN we often observe faint background bands in the pellets (Fig. 4). In addition,
231 the pull-down assays do not cover a wide enough range to estimate reliable dissociation constants. Our
232 tentative conclusions are therefore that IntPeri-MBP binds with high affinity to a finite number of
233 binding sites, whereas a small amount of InvPeri-MBP precipitates unspecifically with the sacculi, though
234 we cannot rule out some specific but low-affinity interaction. Unfortunately, we could not use more
235 sophisticated biophysical methods such as surface plasmon resonance to more accurately estimate the
236 dissociation constants due to the difficulty of immobilising sacculi on detector chips.

237 ***The Int periplasmic domain binds to PGN in vivo***

238 To determine whether the Int periplasmic domain can mediate PGN binding *in vivo*, we cloned the MBP
239 fusions into pIBA2C, which contains a signal sequence for periplasmic targeting. We used
240 chloramphenicol selection rather than ampicillin in order to avoid any structural changes to PGN. The
241 protein was expressed in medium buffered at either pH 7.4 or pH 5.5. To see if the MBP fusion proteins
242 bound to PGN, we extracted the periplasmic fraction by osmotic shock. All solutions were buffered at
243 either pH 7.4 or pH 5.5. We then detected the protein by Western blot using an anti-MBP antiserum. For
244 the whole-cell samples of IntPeri-MBP, a strong band is evident at \sim 65 kDa, as expected, along with a

245 small amount of apparent degradation products (Fig. 6A). We assume the 65 kDa band corresponds to
246 intact IntPeri-MBP. At pH 7.4, this band is also in the periplasmic fraction, but at pH 5.5 the 65 kDa band
247 is barely visible. InvPeri-MBP is less stable, with two stronger bands: ~50 kDa, which is presumably the
248 intact protein, and ~45 kDa, which we assume is a degradation product. Both bands are present at
249 similar levels in all samples. The vector control does not display any bands, demonstrating the specificity
250 of the antiserum, whereas in our control for periplasmic extraction (pIBA2C-MBP), the MBP band (41
251 kDa) is present in all the samples at an equal intensity. As the 65 kDa IntPeri-MBP band is present in the
252 periplasmic fraction at pH 7.4 but not at pH 5.5, we interpret this to mean that at pH 5.5, IntPeri-MBP is
253 retained in the periplasm due to PGN binding. This finding is consistent with our pull-down assays, as is
254 the observation that InvPeri-MBP is extracted at similar amounts at both pH values, which supports the
255 conclusion that it does not bind to PGN.

256 ***Structure of the Intimin LysM***

257 To date, there is no structural information about the inverse autotransporter periplasmic domains. To
258 structurally characterise the periplasmic domain, we produced and purified the N-terminal region of the
259 Int periplasmic domain (residues 40-153). We did this by producing the protein as a fusion with MBP,
260 followed by cleaving the MBP from the Int fragment with tomato etch virus (TEV) protease. This region
261 contains the predicted LysM domain (residues 63-114) and putative dimerisation interface (Fig. 2C). We
262 then proceeded to solve the solution structure of this region using nuclear magnetic resonance (NMR)
263 spectroscopy.

264 When measuring the NMR spectra, we noted the rapid decay of several signals with shifts suggesting a
265 structured environment and the appearance of others more consistent with unstructured protein chain.
266 The signals appearing were assigned to the C-terminus of the protein, *i.e.* in the spacer sequence.
267 Despite extensive efforts, we could not obtain high-resolution structural information from the transient
268 signals, although chemical shifts suggest that this region has considerable β -sheet content. We conclude
269 that this structure is unstable and unfolds once the C-terminal fusion partner is removed. At equilibrium,
270 both the N-terminus (residues 40-60) and the C-terminus (residues 115-143) are disordered (Fig. 7A).
271 However, the LysM itself was well defined (Fig. 7A). The NMR constraint and refinement statistics are
272 presented in Table S2. It is noteworthy that we could not detect the dimerisation interface in NMR
273 measurements on mixtures of differentially labelled protein. This is probably due to the unfolding of the
274 spacer sequence that presumably contains the dimerisation site.

275 The overall fold of the Int LysM is the typical β - α - α - β seen in other LysM structures (Fig. 7B). A α
276 superposition of the various LysMs shows that the main chain conformation is largely similar (Fig. 7C),
277 with the root mean square deviation (RMSD) between the different structures and the Int LysM varying
278 from 1.28 Å to 3.62 Å. Most of the LysM structures have an RMSD between 1.5 and 2.5 Å; the furthest
279 outliers are the hypothetical human protein SB145 (Protein database ID 2DJP) at 2.75 Å and the gpX
280 LysM from coliphage P2 (2LTF; (Maxwell *et al.*, 2013)) with an RMSD of 3.62 Å. The closest structure is
281 from the putative (trans)peptidase YkuD from *Bacillus subtilis* (Bielnicki *et al.*, 2006). Though there are
282 no major differences between the current LysM domain structures, the Int LysM does contain a short α -
283 helix connecting the structurally conserved C-terminal α -helix and the C-terminal β -strand (Fig. 7C).
284 Though gpX some other LysMs have a helical turn in this region, these are not as long and defined as the
285 α -helix in the Int LysM. This extra helix was predicted by bioinformatics (Fig. S1), and appears conserved
286 within the inverse autotransporter LysMs, suggesting it has some function. Future work will be needed
287 to uncover the relevance of this α -helix.

288 ***PGN binding and dimerisation are general features of LysM-containing inverse autotransporters***

289 Several inverse autotransporters from the Enterobacteriaceae contain a periplasmic LysM domain (Fig.
290 1). To test whether binding to PGN and dimerisation are general features of inverse autotransporter
291 LysMs, we cloned the LysM-containing periplasmic domain from an Inv homologue from the fish
292 pathogen *Yersinia ruckeri* (GI: 238705545), which we refer to as YrInv for *Y. ruckeri* Invasin. We
293 produced the periplasmic domain as an MBP fusion (YrInvPeri-MBP). During later purification steps, we
294 noticed that in SDS-PAGE, the band of the expected size (58 kDa) began to lose intensity and a new band
295 migrating at ~200 kDa appeared, and almost all the protein was in this higher molecular-weight band
296 after the SEC step. This suggested to us that YrInv was forming multimers stabilised by disulphide
297 bridges, which were not disrupted in our SDS-PAGE experiments because we do not routinely add
298 reducing agent to our sample buffer. When we used reducing sample buffer, all the protein previously in
299 the ~200 kDa band ran at the expected size of the monomer, showing that the multimer resulted from
300 the formation of disulphides by oxidation during purification (Figure 8A). Indeed, there is a single
301 cysteine in the periplasmic domain in the spacer sequence between the LysM and Helix1 which
302 presumably mediates the disulphide formation (Fig. 8D). When we performed analytical SEC, YrInvPeri-
303 MBP migrated with an apparent size of ~240 kDa (Fig. 8B). However, if the buffer was supplemented
304 with dithiothreitol (DTT), the protein migrated with an apparent molecular weight of 91 kDa (Fig. 8B). As
305 the expected size of the monomer is 58 kDa, we interpret the 91 kDa peak to represent the dimeric form

306 of the molecule. The 240 kDa peak seen under non-reducing conditions would therefore correspond to a
307 tetrameric molecule, *i.e.* a dimer of dimers stabilised by disulphides. We are not sure whether this
308 tetramer represents the physiological quaternary structure of YrInv, as we were unable to express the
309 full-length protein in *E. coli*. Though this tetramer could be an artefact of the purification procedure, it
310 seems more likely that the tetrameric form is the physiologically relevant one, as the tetramer of
311 YrInvPeri-MBP forms quantitatively in the presence of oxygen, and disulphides would be formed in the
312 periplasm. Interestingly, in addition to dimerising, Inv from *Y. pseudotuberculosis* also forms tetramers
313 (Dersch and Isberg, 1999), further pointing towards the interpretation of the tetramer being the
314 physiologically relevant form of YrInv.

315 We also performed pull-down assays with PGN sacculi. In the pull-down assay, approximately half the
316 protein precipitated with the sacculi at pH 5.0, similar to Int, whereas in the control reaction the protein
317 remained soluble (Fig. 8C). Addition of DTT did not have a significant effect on the amount of protein in
318 the pellet fraction, suggesting that the tetrameric form is not required for PGN binding (Fig. 8C).
319 Interestingly, similar to Int, YrInvPeri-MBP did not bind to PGN at pH 7.4 (Fig. 8C). These results show
320 that also YrInv-MBP dimerises, and probably further tetramerises, and binds to PGN. This suggests that
321 dimerisation and PGN binding are general properties of LysM-containing periplasmic domains of inverse
322 autotransporters, and that the pH dependence of the PGN interaction may also be a general feature.

323 **DISCUSSION**

324 ***PGN binding by the Intimin LysM***

325 Here, we provide the first experimental evidence to show that the Int periplasmic region has affinity for
326 PGN. In addition, we present the NMR structure of the Int LysM, giving the first structural information
327 on an inverse autotransporter periplasmic domain. Using pull-down assays with purified PGN sacculi, we
328 were able to experimentally show, for the first time, that the LysM of the Int periplasmic domain
329 mediates binding to PGN, whereas the periplasmic domain from Inv did not bind with appreciable
330 affinity. In our *in vivo* experiments, InvPeri-MBP was efficiently extracted from the periplasm at pH 5.5.,
331 whereas IntPeri-MBP was retained. These data all point to IntPeri binding strongly to PGN, whereas
332 InvPeri binds only at low levels, which we contend represent unspecific background binding. An arguable
333 shortcoming of our experiments is that we were not able to test the binding of InvPeri to PGN from *Y.*
334 *enterocolitica*. However, the composition of PGN from *E. coli* and *Y. enterocolitica* is very similar, though
335 there is some variation in the relative abundance of specific muropeptides (Quintela *et al.*, 1995). This

336 suggests that if Inv did bind with high affinity to PGN, we would see stronger binding in our experiments.
337 This view is supported by the fact that YrInvPeri clearly does bind to PGN from *E. coli* (Fig. 7). Our
338 conclusion is therefore that only LysM-containing inverse autotransporter periplasmic domains mediate
339 PGN binding.

340 Most LysM-containing proteins that have been studied bind to the carbohydrate backbone of PGN, and
341 many also bind to the N-acetylglucosamine backbone of chitin. However, we did not detect any binding
342 to chitin beads. Furthermore, although we observed strong binding by IntPeri-MBP to purified sacculi,
343 the binding saturates at a rather low concentration of protein in relation to PGN (Fig. 5). This would
344 suggest that either the affinity is low, or the affinity is high and there are only a limited number of
345 binding sites available for Int. Our data points to the latter explanation; IntPeri-MBP does not bind to
346 chitin, which suggests to us that the PGN carbohydrate backbone alone is not the target of the Int LysM
347 domain, but that some particular muropeptide or combination of crosslinked PGN strands is required for
348 binding. The dimerisation of the LysM may have something to do with the specific binding: it could
349 either position the two PGN binding sites of the dimer in such a way that the protein can optimally
350 interact with the specifically crosslinked strands, or dimerisation increases the affinity of the dimer for
351 PGN. In contrast to inverse autotransporters, many PGN-binding proteins contain several LysMs in
352 tandem (Buist *et al.*, 2008). Dimerisation may thus be a mechanism to increase the number of PGN
353 binding units and thereby the strength of the interaction. We are now pursuing further characterisation
354 of the binding determinants for the Int periplasmic domain in PGN.

355 ***The dimerisation interface(s)***

356 We demonstrate here that the Int periplasmic domain is a dimerisation interface for Int. This is in
357 contrast to a previous report implicating the β -barrel domain as the dimerisation interface (Touzé *et al.*,
358 2004). However, Fairman *et al.* have shown that the β -barrel alone does not dimerise (Fairman *et al.*,
359 2012), which suggests that the dimerisation is mediated by the regions immediately upstream or
360 downstream of the β -barrel domain. We tested whether the N-terminal Ig-like D00 domain forms
361 dimers, but our SEC experiments show that it does not. Therefore, the dimerisation observed by Touzé
362 *et al.* is probably due to the C-terminal helix (Helix2) of the periplasmic domain. However, this region of
363 the periplasmic domain does not mediate dimerisation when fused N-terminally to MBP. A likely
364 explanation for this discrepancy is that Helix2 is intimately connected to the β -barrel domain, and
365 therefore protease-resistant, as shown by Touzé *et al.* (2004). When fused to MBP, these connections

366 are lost and the helix remains either unfolded or misfolded and thus cannot mediate dimerisation. This
367 finding is supported by the observation that both IntPeri-MBP and IntHelix2-MBP were prone to
368 degradation.

369 We also observed clear dimerisation by the LysM region alone, a finding has not been reported before.
370 Both the full periplasmic domain and the N-terminal, LysM-containing region mediated dimerisation in
371 SEC experiments when fused to MBP. A notable feature of these experiments was that the dimeric form
372 was in equilibrium with the monomeric form. This is in contrast to the findings of Touzé *et al.*, who only
373 observed obligate dimerisation of Int.

374 Though our SEC experiments show clear evidence for the N-terminal region of IntPeri mediating
375 dimerisation, unfortunately we did not observe dimers when solving the solution structure. This is most
376 likely due to the unfolding of the C-terminal spacer element upstream of the LysM. This region may
377 require a C-terminal anchor (Helix1 under native condition, or MBP in our constructs) to fold stably. The
378 instability of this region outside its native context could explain why we saw an equilibrium between the
379 dimeric and monomeric states in SEC. We thus conclude that it is not the LysM itself that is the
380 dimerisation interface, but the upstream sequence. This conclusion is supported by the fact that
381 YrInvPeri forms disulphide-bonded tetramers. The single cysteine in the periplasmic domain is located in
382 the spacer sequence, suggesting that this region is in fact the dimerisation (or oligomerisation) interface.
383 That the periplasmic domain of YrInv also mediates oligomerisation suggests that oligomer formation is
384 a general feature of LysM-containing inverse autotransporters. In the case of YrInv, the physiological
385 relevance of the disulphide-bonded tetramer remains to be determined.

386 In conclusion, we present a new model for Int dimerisation, where the LysM of the periplasmic domain
387 is the dimerisation interface, with two sites for dimer formation: Helix2 and the spacer sequence
388 between the LysM and Helix1 (Fig. 9). This is a modification of the earlier model of Touzé *et al.* (2004),
389 which suggested that a single Int dimer interacts with two separate Tir dimers, leading to a reticular
390 array of Int-Tir interactions and receptor clustering. This is consistent with the crystal structure of Int-Tir,
391 where the Int monomers jut out from the Tir dimer in opposite directions (Fig. 8) (Luo *et al.*, 2000).

392 ***Biological implications***

393 What might be the function of PGN binding in a subset of inverse autotransporters? One possibility
394 might be to anchor the protein to specific sites in PGN and thus prevent lateral diffusion in the outer
395 membrane. This could be useful for A/E pathogens, such as EPEC, for pedestal formation and in

396 maintaining intimate attachment to host cells. Another interesting facet of the Int-PGN interaction is the
397 dependence on low pH, which appears to be a conserved feature based on our results with YrInv. The
398 pH dependence of inverse autotransporter LysM binding suggests that charged residues are involved in
399 binding. IntPeri has a calculated isoelectric point (pI) of 8.7, and YrInvPeri has a pI of 6.1. Thus, at pH 5.0,
400 both proteins carry a net positive charge which may play a role in PGN binding. A similar result was
401 obtained for AcmD, an autolysin from *Lactococcus lactis* containing three LysMs, where binding to the
402 cell wall was only detected below the pI of the protein (Visweswaran *et al.*, 2013). Interestingly, Int is
403 upregulated under acid stress (House *et al.*, 2009). As EPEC and other A/E pathogens must travel
404 through the stomach to reach the intestine, and the pH of the periplasm closely follows the pH of the
405 extracellular medium (Wilks and Slonczewski, 2007), the periplasm will experience a significant drop in
406 pH during the pathogen's journey through the stomach. The binding to PGN may help in stabilising the
407 cell envelope and aid in acid resistance; however, our attempts to probe the acid resistance of Int-
408 expressing bacteria using survival assays did not yield any conclusive results compared to control cells
409 (data not shown).

410 The function of PGN binding and dimerisation remain unclear. The model proposed by Touzé *et al.*
411 (2004) suggests that Int dimerisation leads to more efficient Tir clustering on the host cell membrane
412 and the initiation of downstream effects. Purified Int passenger domain can also lead to actin
413 rearrangements (Liu *et al.*, 1999), but these experiments were carried out using a GST fusion, and GST
414 forms dimers, which could explain this observation. Bacteria expressing periplasmically truncated Int or
415 InvPeri-Int did not bind as strongly to HeLa cells as wt Int, possibly due to less surface exposure of the
416 protein in the case of Int Δ Peri, but potentially as a result of the loss of dimer formation in the case of
417 the chimaeric protein. Interestingly, a recent study showed that an Inv-Int fusion that binds to Tir failed
418 to form actin pedestals when expressed in *Citrobacter rodentium*, and that this strain was defective in
419 colonising a mouse model (Mallick *et al.*, 2012). As this fusion lacks a dimerisation interface, it is
420 tempting to speculate that the defects are at least partially due to the protein not forming a functional
421 dimer. Dimerisation seems to be important for inverse autotransporter function, as it is widely
422 conserved within the protein family. In contrast to LysM-containing inverse autotransporters, Inv from *Y.*
423 *pseudotuberculosis* forms dimers and tetramers, but in the absence of a LysM (Dersch and Isberg, 1999).
424 The multimerisation interface is in the D2 Ig-like domain in the passenger region, which is missing in the
425 *Y. enterocolitica* orthologue (Dersch and Isberg, 2000). The *Y. pseudotuberculosis* protein is the more
426 potent invasin and promotes a stronger cellular response compared to *Y. enterocolitica* Inv, presumably

427 due to receptor clustering (Dersch and Isberg, 2000). Dimerisation of inverse autotransporters thus
428 appears functionally important and can occur through at least two distinct mechanisms: through
429 dimerisation of the periplasmic domain or by self-association of domains in the passenger region.

430 Apart from adhesion, another aspect of the biology of type Ve-secreted adhesins in which the
431 periplasmic domain might play a role is autotransport. Some of our data supports this hypothesis:
432 though deletion of the entire periplasmic domain does not abolish surface display, it does reduce it. A
433 reasonable assumption would therefore be that the conserved C-terminal helices (Helix1 and Helix2)
434 would play a role in passenger secretion. Additionally, longer inverse autotransporters are more likely to
435 include a LysM, which might also have a function in the autotransport process. Autotransporters of
436 some other classes contain periplasmic domains: a few trimeric (type Vc) autotransporters contain
437 coiled-coil periplasmic extensions, and type Vd-secreted proteins have periplasmic POTRA (polypeptide
438 transport-associated) domains (Szczesny and Lupas, 2008; Salacha *et al.*, 2010; Arnold *et al.*, 2010). We
439 are currently investigating the involvement of autotransporter periplasmic domains in passenger
440 secretion.

441 **EXPERIMENTAL PROCEDURES**

442 ***Bioinformatic analyses***

443 To identify periplasmic domains of inverse autotransporters, we performed a PSI-BLAST (Altschul and
444 Koonin, 1998) search with 5 iterations against the non-redundant database using either the Int or Inv
445 periplasmic domain and β -barrel domain sequence. We included the latter to correctly identify type Ve-
446 secreted proteins, and only hits covering over 80% of the query (so as to contain sequence from both
447 the β -barrel and periplasmic domain) were included. The search results were pooled and duplicate hits
448 removed. The sequences were then checked manually and any clearly false positive results (*i.e.* those
449 lacking the β -barrel domain) were removed. The remaining sequences (many of which appear to
450 misannotated in the database) were aligned and the β -barrel domains removed based on the alignment.
451 To identify signal peptide cleavage sites, we submitted the sequences to SignalP 4.0 (Petersen *et al.*,
452 2011) or Phobius (Käll *et al.*, 2004), and the sequence N-terminal to the consensus cleavage site were
453 removed. The remaining sequences were then checked for the presence of a LysM motif using HHPred
454 (Söding *et al.*, 2005). For clustering, we used CLANS (Frickey and Lupas, 2004) with default parameter
455 values. Alignments and secondary structure analysis were performed using programmes in the MPI
456 Bioinformatics Toolkit (Biegert *et al.*, 2006)(<http://toolkit.tuebingen.mpg.de/>).

457 ***Cloning***

458 We amplified the regions of interest from genomic DNA of *E. coli* O127:H6 strain 2348/69 (for Int), *Y.*
459 *enterocolitica* O:8 strain 8081 (for Inv) or *Y. ruckeri* strain CECT 4319 (for YrInv) using polymerase chain
460 reaction (PCR) with Phusion polymerase (Thermo Scientific). Primers were constructed to include N- and
461 C-terminal BsaI sites for cloning into vectors of the pASK-IBA series (IBA GmbH). For the periplasmic
462 domains, primers were designed to introduce an N-terminal hexahistidine tag for protein purification.
463 Primer sequences are available on request. Constructs used in this study are summarised in Table 1. For
464 producing MBP fusions, we introduced in-frame BamHI and NheI restriction sites after the 3' end of the
465 insert by PCR-based site-directed mutagenesis (Byrappa *et al.*, 1995). The MBP coding sequence was
466 amplified from the plasmid pMal-c2 (New England Biolabs) with primers that introduced BamHI and
467 NheI sites to the 5' and 3'-ends, respectively. This product was then cloned into the corresponding sites
468 in the modified pIBA-ASK plasmids to produce the final fusion construct. For fusion controls, we
469 amplified MBP with primers introducing BsaI sites and cloned the PCR product into pASK-IBA33 to
470 include a C-terminal hexahistidine tag for efficient purification. To purify Int LysM alone, we amplified
471 the IntLysM-MBP insert from pIBA2-IntPeriLysM-MBP with the forward primer lacking the His tag
472 sequence. This product was then cloned into pASK-IBA33 to introduce a C-terminal His tag on the MBP
473 moiety, and a TEV site was introduced between the LysM and MBP by site-directed mutagenesis. To
474 produce domain-swapped Inv and Int constructs, we performed a two-step PCR reaction, where the
475 primers at the fusion junction were designed to overlap (Barik, 1997). Restriction enzymes were from
476 New England Biolabs, and T4 DNA ligase from Fermentas.

477 ***Bacterial strains and growth media***

478 For cloning, all ligation reactions were transformed into chemically competent *E. coli* TOP10 (Invitrogen).
479 For protein overproduction in the cytoplasm, we used the expression strain BL21Gold(DE3) (Novagen).
480 For surface production, the BL21 derivative strain omp2 was used, which is optimised for outer
481 membrane protein expression (Prilipov *et al.*, 1998). Bacteria were usually grown in lysogeny broth
482 medium (LB) (Bertani, 1951) supplemented with ampicillin (100 µg/ml) or chloramphenicol (25 µg/ml).
483 For protein production, we used the buffered, rich medium ZYP (Studier, 2005) supplemented with
484 ampicillin as above.

485 ***Protein production and purification***

486 For production of MBP and MBP fusions, an overnight culture of BL21Gold transformed with the
487 required plasmid was diluted 1:100 in 1 l ZYP medium and grown to mid-log phase ($OD_{600} \sim 0.5$) at 37 °C,
488 at which time protein production was induced with anhydrotetracycline (200 ng/ml). The cultures were
489 then grown for another 3 hours at 37 °C after which the cells were harvested by centrifuging (10
490 minutes $5000 \times g$) and then resuspended in Ni binding buffer (20 mM sodium phosphate, 20 mM
491 imidazole, 500 mM sodium chloride, pH 7.4). We added $MgCl_2$ and $MnCl_2$ to 1 mM, lysozyme to 0.1
492 mg/ml, EDTA-free complete protease inhibitor cocktail (Roche) and a pinch of DNase I (Applichem) and
493 proceeded to rupture the cells using a French pressure cell with two passes at 18 000 p.s.i. Cellular
494 debris were pelleted for 1 hour at $100\ 000 \times g$, the supernatant was passed through a 0.22 μm filter and
495 applied to a nickel iminodiacetic acid column (PrepEase; USB). Bound proteins were eluted with a step
496 gradient of imidazole (250 mM) in Ni binding buffer. We pooled the fractions containing the periplasmic
497 domain-MBP fusions, added protease inhibitor cocktail and applied these to an amylose column
498 (MBPTrap; GE Healthcare) equilibrated with 20 mM Tris, 200 mM NaCl, 1 mM EDTA at pH 7.4. Bound
499 protein was eluted using the same buffer supplemented with 10 mM maltose. Proteins were further
500 purified by size exclusion chromatography (Superdex200™ 26/60 column; GE Healthcare) in Tris-
501 buffered saline (TBS; 20 mM Tris pH 7.4 with 150 mM NaCl and 0.02% NaN_3).

502 For producing isotope-labeled protein for NMR experiments, we grew cells overnight in minimal
503 medium M9 (Hochuli *et al.*, 2000) supplemented with 1% LB and ampicillin at 100 $\mu g/ml$. The cultures
504 were then diluted 1:200 in 2 l of M9 + 1% LB + ampicillin containing $^{15}NH_3Cl$ and ^{13}C -glucose (Sigma).
505 Once the cultures reached mid-log phase, protein production was induced and the cultures were
506 harvested after 3 hours at 37 °C. The cells were resuspended in Ni binding buffer and lysed as above.
507 After passing over a nickel column, the eluted protein was dialysed overnight against phosphate-
508 buffered saline (PBS; 20 mM phosphate pH 7.4, 150 mM NaCl). The following morning, TEV protease
509 (produced according to (Tropea *et al.*, 2009) was added to 1/10 of the concentration of sample protein
510 as estimated by absorbance at 280 nm. The digestion was allowed to proceed for 4 hours at room
511 temperature, after which the digested protein was passed over the nickel column again. The flow
512 through contained the digested, labeled protein. This was further purified by size exclusion
513 chromatography using a Superdex75™ 16/60 column (GE Healthcare) equilibrated with PBS.

514 For outer membrane expression, overnight cultures were diluted 1:100 in 20 ml ZYP medium and grown
515 till mid-log phase at 37 °C. The temperature was then adjusted to 27 °C and protein production was

516 induced with anhydrotetracycline (50 ng/ml). The cells were grown for approximately two hours at 27 °C
517 before harvesting.

518 ***NMR structure determination***

519 All spectra were recorded at 298 K on Bruker AVIII-600 and AVIII-800 spectrometers. Backbone
520 sequential assignments were completed using a strategy based on a 3D-HN(CA)NNH experiment
521 (Weisemann *et al.*, 1993). Aliphatic sidechain assignments were completed with standard TOCSY-based
522 experiments, while aromatic assignments were made by linking aromatic spin systems to the respective
523 C^βH₂ protons in a 2D-NOESY spectrum. Stereospecific assignments and the resulting χ 1 rotamer
524 assignments were determined from a combination of HNHB and HA[HBHN](CACO)NH experiments (Löhr
525 *et al.*, 1999).

526 Distance data were derived from a set of five 3D-NOESY spectra, including the heteronuclear edited
527 NNH-, CCH-, and CNH-NOESY spectra (Diercks *et al.*, 1999) in addition to conventional ¹⁵N- and ¹³C-
528 HSQC-NOESY spectra. A ¹²C-filtered 2D-NOESY spectrum was recorded for the observation of contacts to
529 aromatic groups. Backbone dihedral angle restraints were derived using the TALOS+ server (Shen *et al.*,
530 2009). Generic backbone dihedral restraints designed to restrict residues to allowed regions of the
531 Ramachandran map and well-populated sidechain rotamers were applied for unstructured residues.
532 Hydrogen bond restraints were applied as pseudo-covalent bonds, as outlined in (Truffault *et al.*, 2001).
533 Refinement was carried out by comparing experimental and back-calculated NOESY spectra using in-
534 house software. Strips were back calculated for the amide protons of all ordered atoms, plus selected
535 sidechain groups. These were compared to the experimental spectra to confirm backbone and sidechain
536 dihedral angles and to extract additional distance restraints.

537 Structures were calculated with XPLOR (NIH version 2.9.4) using standard protocols with modifications
538 for the inclusion of H-bonds as pseudo-covalent bonds. For the final set, 100 structures were calculated
539 and 18 chosen on the basis of lowest restraint violations. An average structure was calculated and
540 regularized to give a structure representative of the ensemble. Details of the input data and the final
541 ensemble are given in Table S2.

542 ***Surface display and functional controls***

543 Immunofluorescence microscopy, FACS, pre-infection adhesion assays and ELISA for IL-8 production
544 were all performed as described (Oberhettinger *et al.*, 2012).

545 **Analytical size exclusion chromatography**

546 Gel sizing was done using a Superdex™200 10/300 column (GE Healthcare) equilibrated either with TBS
547 or acetate-buffered saline (ABS; 20 mM sodium acetate pH 4.0 with 150 mM NaCl). Approximately 1 mg
548 of protein was applied to the column, and elution was monitored by absorbance at 280 nm. For sizing, a
549 preparation of standard proteins (Ribonuclease A (13.7 kDa), Conalbumin (75 kDa), Aldolase (158 kDa),
550 Ferritin (440 kDa) from GE Healthcare) was passed through the column. We then compared the position
551 of the sample proteins' peaks to a standard curve drawn based on the elution profile of the standard
552 mix to obtain apparent molecular weights.

553 **Peptidoglycan and chitin binding**

554 Peptidoglycan sacculi from *E. coli* Nissle were purified according to the method of Glauner with some
555 modifications (Glauner, 1988). The lyophilized sacculi were resuspended in ultrapure water (with
556 sodium azide added to 0.02% w/v) to a concentration of 10 mg/ml. For pull-down assays, 2 µl of this
557 suspension was mixed with 5 µg of protein in a total volume of 30 µl and incubated for 15 minutes at
558 room temperature. The buffer used depended on the pH: we used TBS (pH 8.0 or pH 7.4), MOPS-
559 buffered saline (pH 7.0), MES-buffered saline (pH 6.0), ABS (pH 5.0 and pH 4.0) or glycine-buffered saline
560 (pH 3.0). The sacculi were then pelleted by centrifuging 30 minutes at ~20 000 x *g* (full speed using a
561 tabletop centrifuge). The supernatant was carefully removed and the pellet was washed once with the
562 corresponding buffer. After a second centrifugation step the supernatant was removed and the pellet
563 was resuspended in 30 µl TBS (pH 7.4). For analysis, we added 10 µl of 4X non-reducing sample buffer to
564 the first supernatant fraction and the pellet fraction, boiled the samples for 5 minutes and then loaded
565 10 µl onto a 12% polyacrylamide gel for SDS-PAGE.

566 For competition experiments, 20 mg of *E. coli* PGN sacculi were digested with mutanolysin (4000 U;
567 from Sigma) in a total volume of 800 µl PBS for 16 hours at 37 °C. After heat inactivation of the enzyme
568 (100 °C, 2 min), the undigested PGN was pelleted (30 min at ~20 000 x *g*) and the supernatant was
569 transferred to a new tube. The pH of the solution was changed to 5.0 by the addition of 50 µl 1 M NaAc
570 at this pH; the pH of the solution was checked after mixing. An estimated 30-40% of the sacculi were
571 digested, giving a concentration of ~7 mg/ml of various muropeptides. We used 5, 10, 15, 20 or 25 µl of
572 the muropeptide solution in our competition experiment. The muropeptides and IntPeri-MBP (5 µg as
573 above) were mixed first, and peptidoglycan sacculi were subsequently added. The total volume was
574 adjusted to 30 µl with ABS pH 5.0. The procedure from this point on was as above.

575 For the concentration series, we varied the amount of protein between 0.5 μg and 10 μg . The amount of
576 peptidoglycan was varied between 2 μg to 50 μg . The binding buffer was ABS (pH 5.0). The samples
577 were otherwise treated as above.

578 Chitin binding was assayed using chitin beads (New England Biolabs). We prepared columns using 2 ml
579 of the bead slurry and equilibrated with ABS (pH 5.0). We added 1 mg of protein diluted in 1 ml of ABS
580 and allowed the protein to enter the column by gravity flow. After washing with 10 ml ABS, bound
581 protein was eluted with 10 ml TBS (pH 7.4), followed by a second elution step with 0.3 M NaOH. A
582 sample was taken from each step for SDS-PAGE analysis.

583 ***In vivo PGN binding***

584 To test for PGN binding in whole cells, we transformed the plasmids pIBA2C-IntPeri-MBP, pIBA2C-
585 InvPeri-MBP or pASK-IBA2C into BL21omp2 cells. The cells were then grown at 37 °C in LB +
586 chloramphenicol (25 $\mu\text{g}/\text{ml}$) till mid-log, at which time the temperature was changed to 27 °C. Growth at
587 this temperature in the presence of maltose leads to expression of endogenous MBP in BL21 and its
588 derivatives (REF). After 30 minutes, recombinant protein production was induced with AHTC (50 ng/ml).
589 After 1 hour at 27 °C, we added buffer, either Tris pH 7.4 or sodium acetate pH 5.5, to 100 mM. After a
590 further hour of growth the turbidity of the cultures (OD_{600}) was measured and an amount of cells
591 corresponding to 10 ml at an OD_{600} value of 1.0 was harvested by centrifugation (10 minutes 3500 x g).
592 To examine the soluble periplasmic fraction, we used a modified osmotic shock protocol. The pelleted
593 cells were resuspended in 400 μl 5 mM CaCl_2 with 5 mM buffer (Tris pH 7.4 or sodium acetate pH 5.5)
594 and incubated on ice for 10 minutes. This step improves the yield of the periplasmic extraction (Chen *et*
595 *al.*, 2004). The cells were pelleted (5 minutes at 8000 x g) and then resuspended in 400 μl ice-cold
596 osmotic shock solution (33 mM buffer, 20 % sucrose, 5 mM EDTA) buffered at either pH 7.4 or pH 5.5, as
597 above. After 10 minutes incubation at 8 °C with shaking, the cells were centrifuged as above and then
598 resuspended in 400 μl 5 mM buffer (either Tris pH 7.4 or sodium acetate pH 5.5). The cells were
599 incubated for 10 minutes with shaking at 8 °C and then centrifuged as above. A sample (120 μl) was
600 taken from the supernatant, and 40 μl 4 x SDS-PAGE sample buffer was added. This represented the
601 periplasmic fraction. For the whole-cell sample, we pelleted an amount of cells from the induced culture
602 corresponding to 120 μl at $\text{OD}_{600} = 25$, resuspended these in 120 μl PBS and then added 40 μl sample
603 buffer.

604 To probe for the recombinant proteins, we performed a Western blot. The proteins were separated in a
605 10% polyacrylamide gel and then transferred to a nitrocellulose membrane (Protran BA 85, GE
606 Healthcare) using a semi-dry apparatus. The membrane was blocked for 30 minutes with 5% fat-free
607 milk powder in TBS. The primary antibody was a rabbit anti-MBP (anti-MaE) antiserum diluted in
608 blocking buffer 1:5000. After 1 hour of blocking, the membrane was washed twice 10 minutes with TBS
609 + 0.05% Tween20 (TBS-T), and then the secondary antibody (goat anti-rabbit-alkaline phosphatase (AP),
610 Jackson Immunoresearch) was added at a dilution of 1:10 000 in blocking buffer. The membrane was
611 washed twice with TBS-T as above, and a final time in AP buffer (100 mM Tris pH 9.5, 100 mM NaCl, 5
612 mM MgCl₂). Chromogenic detection was performed using the AP substrate nitro blue tetrazolium/5-
613 bromo-4-chloro-3-indolyl phosphate (NBT/BCIP), diluted to 33 ng/ml and 17 ng/ml, respectively, in AP
614 buffer. The reaction was stopped with deionized water once colour had developed.

615 The anti-MaE antiserum was obtained by immunizing a rabbit with MBP (MaE). The vaccine was
616 prepared by emulsifying 1 mg MaE protein (denaturated with 1% SDS) with Freund's adjuvant in the
617 cold and injected subcutaneously and intramuscularly (hind legs); after four weeks the rabbit obtained a
618 booster injection and blood was collected with 5-6 days intervals.

619 ***Accession numbers***

620 The NMR structure of the Int LysM from this publication has been submitted to the Protein Data Bank
621 (<http://www.rcsb.org/pdb>) and assigned the identifier 2MPW.

622 **Author contributions**

623 J.C.L., P.O., M.S., Mu.Co., I.B.A. and D.L. designed and J.C.L., P.O., Ma.Ch., and Mu.Co. performed the
624 experiments and analysed the data. D.K., U.B., H.S. and F.G. provided essential materials. J.C.L., P.O. and
625 D.L. wrote the paper.

626 **Acknowledgments**

627 We would like to thank M.Sc. Iwan Grin (MPI for Developmental Biology, Tübingen) for assistance with
628 parsing the sequences for bioinformatic analyses. We are also grateful to Dr. Jörg Rau (Chemisches und
629 Veterinäruntersuchungsamt, Stuttgart) for providing the *Y. ruckeri* DNA. We thank Prof. Andrei Lupas
630 (Protein Evolution, MPI for Developmental Biology, Tübingen) for continued support. This study was

631 funded by the Sonderforschungsbereich 766 of the DFG (German Science Foundation) to D.L., I.B.A., U.B.
632 and F.G., UKT fortune F1433253 to P.O., and a FEMS Advanced Fellowship to J.C.L.

633 REFERENCES

634 Altschul, S.F., and Koonin, E.V. (1998) Iterated profile searches with PSI-BLAST--a tool for discovery in
635 protein databases. *Trends Biochem Scis* **23**: 444-7.

636 Arnold, T., Zeth, K., and Linke, D. (2010) Omp85 from the thermophilic cyanobacterium
637 *Thermosynechococcus elongatus* differs from proteobacterial Omp85 in structure and domain
638 composition. *J Biol Chem* **285**: 18003-15.

639 Barik, S. (1997) Mutagenesis and gene fusion by megaprimer PCR. *Methods Mol Biol (Clifton, N.J.)* **67**:
640 173-82.

641 Bateman, A., and Bycroft, M. (2000) The structure of a LysM domain from *E. coli* membrane-bound lytic
642 murein transglycosylase D (MltD). *J Mol Biol* **299**: 1113-9.

643 Bertani, G. (1951) Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J*
644 *Bacteriol* **62**: 293-300.

645 Biegert, A., Mayer, C., Remmert, M., Söding, J., and Lupas, A.N. (2006) The MPI Bioinformatics Toolkit for
646 protein sequence analysis. *Nucleic Acids Res* **34**: W335-9.

647 Bielnicki, J., Devedjiev, Y., Derewenda, U., Dauter, Z., Joachimiak, A., and Derewenda, Z.S. (2006) *B.*
648 *subtilis* ykuD protein at 2.0 Å resolution: insights into the structure and function of a novel, ubiquitous
649 family of bacterial enzymes. *Proteins* **62**: 144-51.

650 Buist, G., Steen, A., Kok, J., and Kuipers, O. (2008) LysM, a widely distributed protein motif for binding to
651 (peptido)glycans. *Mol Microbiol* **68**: 838-47.

652 Byrappa, S., Gavin, D.K., and Gupta, K.C. (1995) A highly efficient procedure for site-specific mutagenesis
653 of full-length plasmids using Vent DNA polymerase. *Genome Res* **5**: 1-5.

654 Chen, Y., Chen, L., Chen, S., Chang, M., and Chen, T. (2004) A modified osmotic shock for periplasmic
655 release of a recombinant creatinase from *Escherichia coli*. *Biochem Eng J* **19**: 211-215.

656 Dersch, P., and Isberg, R.R. (1999) A region of the *Yersinia pseudotuberculosis* invasin protein enhances
657 integrin-mediated uptake into mammalian cells and promotes self-association. *EMBO J* **18**: 1199-213.

658 Dersch, P., and Isberg, R.R. (2000) An immunoglobulin superfamily-like domain unique to the *Yersinia*
659 *pseudotuberculosis* invasin protein is required for stimulation of bacterial uptake via integrin receptors.
660 *Infect Immun* **68**: 2930-8.

661 Diercks, T., Coles, M., and Kessler, H. (1999) An efficient strategy for assignment of cross-peaks in 3D
662 heteronuclear NOESY experiments. *J Biomol NMR* **15**: 177-80.

663 Fairman, J.W., Dautin, N., Wojtowicz, D., Liu, W., Noinaj, N., Barnard, T.J., *et al.* (2012) Crystal structures
664 of the outer membrane domain of intimin and invasin from enterohemorrhagic *E. coli* and
665 enteropathogenic *Y. pseudotuberculosis*. *Structure* **20**: 1233-43.

666 Frickey, T., and Lupas, A. (2004) CLANS: a Java application for visualizing protein families based on
667 pairwise similarity. *Bioinformatics (Oxford, England)* **20**: 3702-4.

668 Glauner, B. (1988) Separation and quantification of mucopeptides with high-performance liquid
669 chromatography. *Anal Biochem* **172**: 451-64.

670 Hamburger, Z.A., Brown, M.S., Isberg, R.R., and Bjorkman, P.J. (1999) Crystal structure of invasin: a
671 bacterial integrin-binding protein. *Science (New York, N.Y.)* **286**: 291-5.

672 Hochuli, M., Szyperski, T., and Wüthrich, K. (2000) Deuterium isotope effects on the central carbon
673 metabolism of *Escherichia coli* cells grown on a D₂O-containing minimal medium. *J Biomol NMR* **17**: 33-
674 42.

675 House, B., Kus, J.V., Prayitno, N., Mair, R., Que, L., Chingcuanco, F., *et al.* (2009) Acid-stress-induced
676 changes in enterohaemorrhagic *Escherichia coli* O157:H7 virulence. *Microbiology (Reading, England)*
677 **155**: 2907-18.

678 Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J Molecular Graph*
679 **14**: 33-8, 27-8.

680 Kelly, G., Prasannan, S., Daniell, S., Fleming, K., Frankel, G., Dougan, G., *et al.* (1999) Structure of the cell-
681 adhesion fragment of intimin from enteropathogenic *Escherichia coli*. *Nat Struct Biol* **6**: 313-8.

682 Käll, L., Krogh, A., and Sonnhammer, E.L. (2004) A combined transmembrane topology and signal
683 peptide prediction method. *J Mol Biol* **338**: 1027-36.

684 Leo, J.C., Grin, I., and Linke, D. (2012) Type V secretion: mechanism(s) of autotransport through the
685 bacterial outer membrane. *Philos Trans R Soc Lond, B, Biol Sci* **367**: 1088-1101.

686 Leo, J.C., and Skurnik, M. (2011) Adhesins of human pathogens from the genus *Yersinia*. *Adv Exp Med*
687 *Biol* **715**: 1-15.

688 Leong, J.M., Fournier, R.S., and Isberg, R.R. (1990) Identification of the integrin binding domain of the
689 *Yersinia pseudotuberculosis* invasin protein. *EMBO J* **9**: 1979-89.

690 Liu, H., Magoun, L., Luperchio, S., Schauer, D., and Leong, J. (1999) The Tir-binding region of
691 enterohaemorrhagic *Escherichia coli* intimin is sufficient to trigger actin condensation after bacterial-
692 induced host cell signalling. *Mol Microbiol* **34**: 67-81.

693 Luo, Y., Frey, E.A., Pfuetzner, R.A., Creagh, A.L., Knoechel, D.G., Haynes, C.A., *et al.* (2000) Crystal
694 structure of enteropathogenic *Escherichia coli* intimin-receptor complex. *Nature* **405**: 1073-7.

695 Löhr, F., Schmidt, J.M., and Rüterjans, H. (1999) Simultaneous Measurement of $^3J_{\text{HN,H}\alpha}$ and $^3J_{\text{H}\alpha,\text{H}\beta}$ Coupling
696 Constants in ^{13}C , ^{15}N -Labeled Proteins. *J Am Chem Soc* **121**: 11821-11826.

697 Mallick, E.M., Brady, M.J., Luperchio, S.A., Vanguri, V.K., Magoun, L., Liu, H., *et al.* (2012) Allele- and Tir-
698 independent functions of intimin in diverse animal infection models. *Front Microbiol* **3**: 11.

699 Maxwell, K.L., Hassanabad, M.F., Chang, T., Pirani, N., Bona, D., Edwards, A.M., and Davidson, A.R.
700 (2013) Structural and functional studies of gpX of *Escherichia coli* phage P2 reveal a widespread role for
701 LysM domains in the baseplates of contractile-tailed phages. *J Bacteriol* **195**: 5461-8.

702 Oberhettinger, P., Schütz, M., Leo, J.C., Heinz, N., Berger, J., Autenrieth, I.B., and Linke, D. (2012) Intimin
703 and invasin export their C-terminus to the bacterial cell surface using an inverse mechanism compared
704 to classical autotransport. *PLoS ONE* **7**: e47069.

705 Petersen, T.N., Brunak, S., Heijne, von, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal
706 peptides from transmembrane regions. *Nat Methods* **8**: 785-6.

707 Pisano, F., Kochut, A., Uliczka, F., Geyer, R., Stolz, T., Thiermann, T., *et al.* (2012) *In vivo*-induced Inva-like
708 autotransporters Ifp and InvC of *Yersinia pseudotuberculosis* promote interactions with intestinal
709 epithelial cells and contribute to virulence. *Infect Immun* **80**: 1050-64.

710 Prilipov, A., Phale, P.S., van Gelder, P., Rosenbusch, J.P., and Koebnik, R. (1998) Coupling site-directed
711 mutagenesis with high-level expression: large scale production of mutant porins from *E. coli*. *FEMS*
712 *Microbiol Lett* **163**: 65-72.

713 Quintela, J.C., Caparrós, M., and de Pedro, M.A. (1995) Variability of peptidoglycan structural
714 parameters in gram-negative bacteria. *FEMS Microbiol Lett* **125**: 95-100.

715 Salacha, R., Kovacic, F., Brochier-Armanet, C., Wilhelm, S., Tommassen, J., Filloux, A., *et al.* (2010) The
716 *Pseudomonas aeruginosa* patatin-like protein PlpD is the archetype of a novel Type V secretion system.
717 *Environ Microbiol* **12**: 1498-512.

718 Schmidt, M.A. (2010) LEEways: tales of EPEC, ATEC and EHEC. *Cell Microbiol* **12**: 1544-52.

719 Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012) NIH Image to ImageJ: 25 years of image analysis.
720 *Nat Methods* **9**: 671-5.

721 Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS+: a hybrid method for predicting protein
722 backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44**: 213-23.

723 Studier, F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr*
724 *Purif* **41**: 207-34.

725 Szczesny, P., and Lupas, A. (2008) Domain annotation of trimeric autotransporter adhesins--daTAA.
726 *Bioinformatics (Oxford, England)* **24**: 1251-6.

727 Söding, J., Biegert, A., and Lupas, A.N. (2005) The HHpred interactive server for protein homology
728 detection and structure prediction. *Nucleic Acids Res* **33**: W244-8.

729 Touzé, T., Hayward, R.D., Eswaran, J., Leong, J.M., and Koronakis, V. (2004) Self-association of EPEC
730 intimin mediated by the β -barrel-containing anchor domain: a role in clustering of the Tir receptor. *Mol*
731 *Microbiol* **51**: 73-87.

732 Tropea, J.E., Cherry, S., and Waugh, D.S. (2009) Expression and purification of soluble His(6)-tagged TEV
733 protease. *Methods Mol Biol (Clifton, N.J.)* **498**: 297-307.

734 Truffault, V., Coles, M., Diercks, T., Abelmann, K., Eberhardt, S., Lüttgen, H., *et al.* (2001) The solution
735 structure of the N-terminal domain of riboflavin synthase. *J Mol Biol* **309**: 949-60.

736 Tsai, J.C., Yen, M., Castillo, R., Leyton, D.L., Henderson, I.R., and Saier, M.H. (2010) The bacterial intimins
737 and invasins: a large and novel family of secreted proteins. *PLoS ONE* **5**: e14403-14.

738 Visweswaran, G.R.R., Steen, A., Leenhouts, K., Szeliga, M., Ruban, B., Hesselting-Meinders, A., *et al.*
739 (2013) AcnD, a homolog of the major autolysin AcnA of *Lactococcus lactis*, binds to the cell wall and
740 contributes to cell separation and autolysis. *PLoS ONE* **8**: e72167.

741 Weisemann, R., Rüterjans, H., and Bermel, W. (1993) 3D triple-resonance NMR techniques for the
742 sequential assignment of NH and ¹⁵N resonances in ¹⁵N- and ¹³C-labelled proteins. *J Biomol NMR* **3**:
743 113-20.

744 Wilks, J.C., and Slonczewski, J.L. (2007) pH of the cytoplasm and periplasm of *Escherichia coli*: rapid
745 measurement by green fluorescent protein fluorimetry. *J Bacteriol* **189**: 5601-7.

746

747 **TABLES**748 **Table 1.** Plasmids used in this study.

Plasmid	Insert sequence	Vector	Comment	Reference/source
pASK-IBA2	-	pASK-IBA2	Expression vector with N-terminal OmpA signal peptide for periplasmic targeting, with ampicillin resistance.	IBA GmbH
pASK-IBA2C	-	pASK-IBA2C	Expression vector with N-terminal OmpA signal peptide for periplasmic targeting, with chloramphenicol resistance.	IBA GmbH
pASK-IBA3	-	pASK-IBA3	Expression vector for cytoplasmic expression.	IBA GmbH
pASK-IBA33	-	pASK-IBA33	Expression vector for cytoplasmic expression, with C-terminal His tag.	IBA GmbH
pIBA2C-IntPeri-MPB	Int periplasmic domain (residues 40-212)	pASK-IBA2C	Produces IntPeri with a C-terminal MBP fusion; includes N-terminal signal peptide for periplasmic transport.	This study
pIBA2C-InvPeri-MBP	Inv periplasmic domain (residues 35-93)	pASK-IBA2C	Produces InvPeri with a C-terminal MBP fusion; includes N-terminal signal peptide for periplasmic transport.	This study
pIBA2C-MBP	MBP	pASK-IBA2C	Produces MBP with N-terminal signal peptide	This study

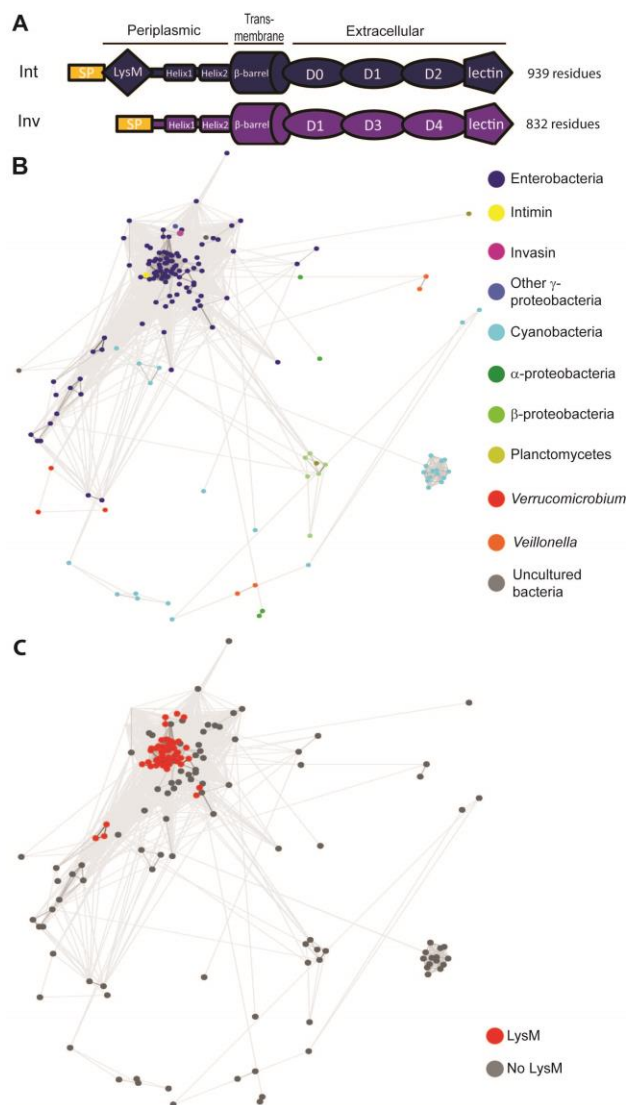
				for periplasmic transport.	
pIBA2-Int Δ LysM	Int lacking LysM domain (contains residues 144-939)	pASK-IBA2		For production of N-terminally truncated Int at cell surface.	This study
pIBA2-Int Δ Peri	Int lacking entire periplasmic domain (contains residues 213-939)	pASK-IBA2		For production of N-terminally truncated Int at cell surface.	This study
pIBA2-IntFull	Full-length Int (without native signal peptide; residues 40-939)	pASK-IBA2		For production of full-length Int at cell surface.	This study
pIBA2-IntPeriHelix1	Int lacking LysM domain and linker to first conserved helix (contains residues 157-939)	pASK-IBA2		For production of N-terminally truncated Int at cell surface.	This study
pIBA2-IntPeriHelix2	Int lacking LysM domain and first conserved helix (residues 184-939)	pASK-IBA2		For production of N-terminally truncated Int at cell surface	This study
pIBA2-Inv Δ Peri	Inv lacking entire periplasmic domain (residues 94-835)	pASK-IBA2		For production of N-terminally truncated Int at cell surface.	This study
pIBA2-InvFull	Full-length Inv (without native signal peptide; residues 35-835)	pASK-IBA2		For production of full-length Inv at cell surface.	This study
pIBA33-IntLysM-TEV-MBP	Int LysM domain (residues 40-153)	pASK-IBA33		For production of Int LysM domain with C-	This study

			terminal MBP fusion separated by TEV site.	
pIBA33-MBP	MBP	pASK-IBA33	Produces MBP with a C-terminal His-tag.	This study
pIBA3-IntD00	Int D00 domain (residues 450-550)	pASK-IBA3	Produces Int N-terminal Ig-like domain	
pIBA3-IntLysM-MBP	Int LysM domain (residues 40-153)	pASK-IBA3	Produces Int LysM with a C-terminal MBP fusion.	This study
pIBA3-IntPeriHelix1-MBP	2 C-terminal helices of Int periplasmic domain (residues 157-212)	pASK-IBA3	Produces conserved helices of Int periplasmic domain with a C-terminal MBP fusion.	This study
pIBA3-IntPeriHelix2-MBP	C-terminal helix of Int periplasmic domain (residues 184-212)	pASK-IBA3	Produces second conserved helix of Int periplasmic domain with a C-terminal MBP fusion.	This study
pIBA3-IntPeri-MBP	Int periplasmic domain (residues 40-212)	pASK-IBA3	Produces IntPeri with a C-terminal MBP fusion.	This study
pIBA3-InvPeri-MBP	Inv periplasmic domain (residues 35-93)	pASK-IBA3	Produces InvPeri with a C-terminal MBP fusion.	This study

749

750

751 **FIGURES AND LEGENDS**



752

753 **Fig. 1.** Bioinformatic analyses of periplasmic domains of inverse autotransporters.

754 A. Schematic of structures of Int from EPEC (dark blue) and Inv from *Y. enterocolitica* (purple). The

755 periplasmic domains of both contain two conserved helices (Helix1 and Helix2) at the C-terminus. In

756 addition, the Int periplasmic domain contains a LysM, which is connected to Helix1 by a spacer

757 sequence. SP = signal peptide. The various domains are not to scale.

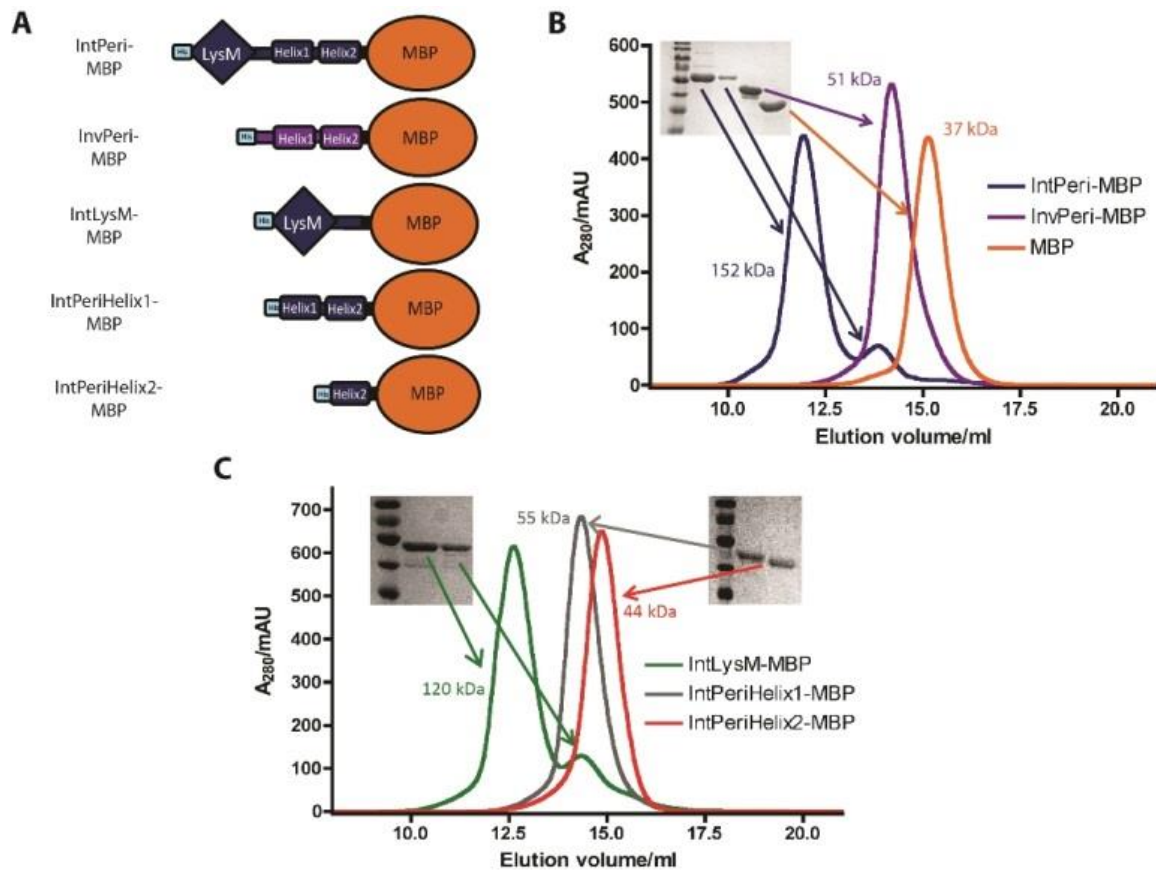
758 B. CLANS clustering of sequences of periplasmic domains from inverse autotransporters. The sequence

759 dots are coloured according to taxonomy. Int is indicated in yellow and Inv in magenta.

760 C. Highlighting of LysM-containing proteins in the cluster map from panel B. The sequences used for

761 clustering are given in Table S1.

762



763

764 **Fig. 2.** The periplasmic domain of Int mediates dimerisation.

765 A. Schematic of periplasmic domain-MBP constructs. Fragments from Int are in blue, fragments from Inv
 766 in purple, MBP is in orange and the N-terminal hexahistidine tags in light blue.

767 B. SEC of Inv and Int periplasmic domains at pH 7.4. The curve for IntPeri-MBP is in blue, InvPeri-MBP in
 768 purple and MBP in orange. The inset shows an SDS-PAGE gel of the peaks. Apparent molecular weights
 769 of the major peaks are indicated. Expected molecular weights are 61 kDa for IntPeri-MBP, 49 kDa for
 770 InvPeri-MBP, and 41 kDa for MBP.

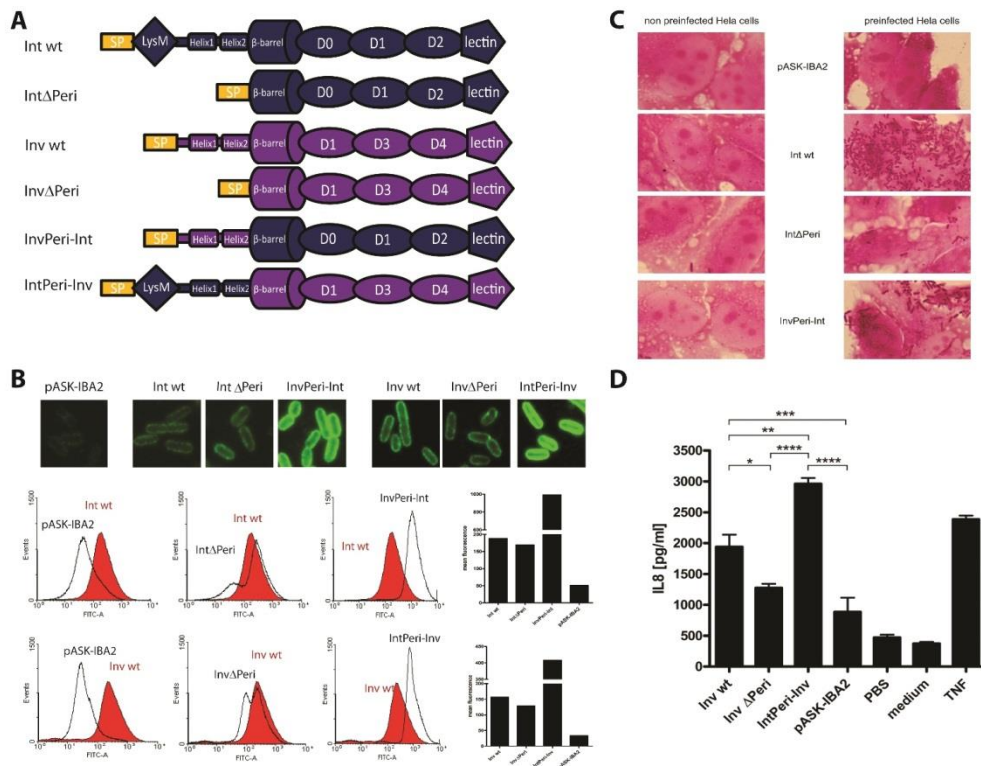
771 C: SEC of Int periplasmic domain fragments. IntLysM-MBP is in green, IntPeriHelix1-MBP in grey and
 772 IntPeriHelix2-MBP in red. The insets show SDS-PAGE gels of the peaks. Apparent molecular weights of
 773 the major peaks are indicated. Expected molecular weights are 54 kDa for IntLysM-MBP, 49 kDa for
 774 IntPeriHelix1-MBP, and 46 kDa for IntPeriHelix2-MBP.

775

776

777

778



779

780 **Fig. 3.** Effect of Int and Inv periplasmic domain deletions.

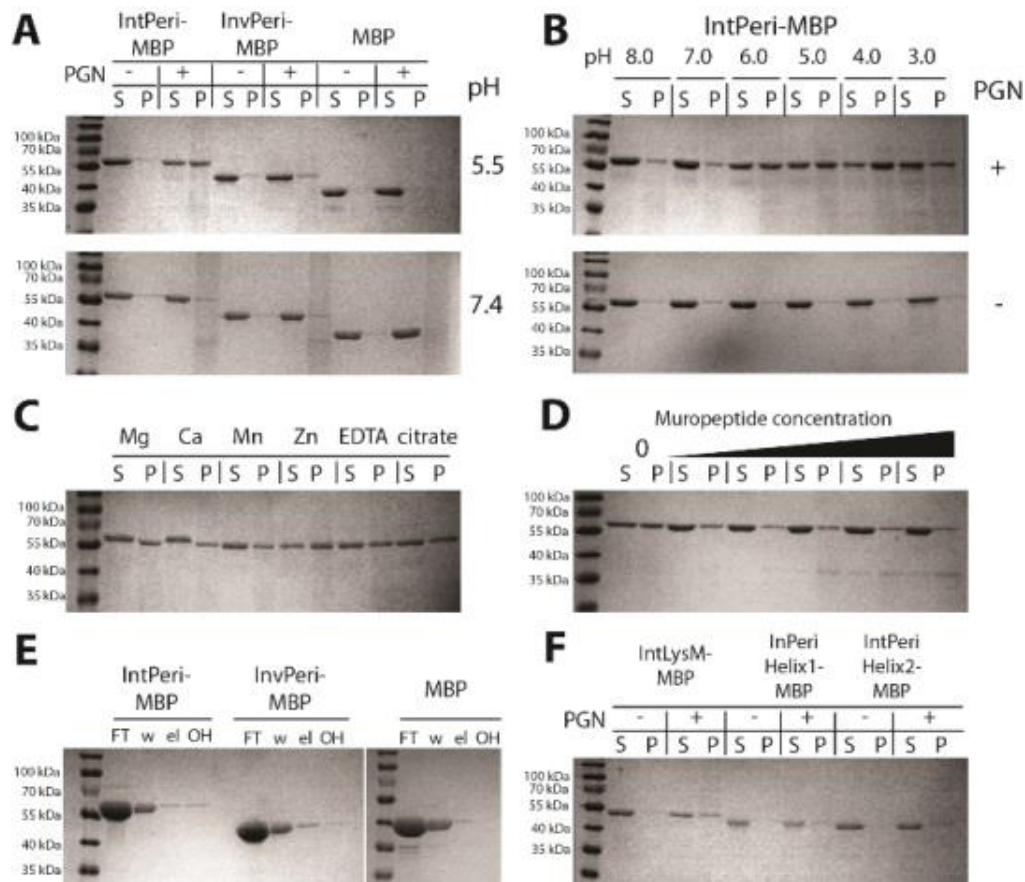
781 A. Schematic of constructs used for cell surface expression experiments. In the chimaeric constructs
782 IntPeri-Inv and InvPeri-Int, the periplasmic domains of Inv and Int have been switched.

783 B. Immunofluorescence and FACS measurements for Int and Inv constructs. For the FACS
784 measurements, mean fluorescence intensity for the constructs is shown as a bar graph on the right.

785 C. Preinfection adhesion assay for Int constructs. HeLa cells were preinfected with EPEC Δ leaeA and then
786 incubated with *E. coli* omp2 expressing Int constructs (right-hand micrographs). Non-preinfected cells
787 and the vector pASK-IBA2 acted as negative controls.

788 D. IL-8 induction by Inv constructs. HeLa cells were incubated with *E. coli* expressing the Inv constructs
789 from panel A and IL-8 production was measured by ELISA. TNF (0.1 μ g) was the positive control, PBS and
790 cell culture medium were negative controls. Error bars (S.E.M.) are from 3 replicates, stars denote
791 statistical significance as determined by ANOVA. $P < 0.05$ (*), $P < 0.01$ (**), $P < 0.001$ (***), $P < 0.0001$
792 (****).

793



794

795 **Fig. 4.** The Int periplasmic domain binds peptidoglycan under acidic conditions

796 A. Pull-down assays using purified peptidoglycan (PGN) sacculi from *E. coli*. IntPeri-MBP, InvPeri-MBP
 797 and MBP were mixed with peptidoglycan sacculi at pH 5.0 (upper gel) or pH 7.4 (lower gel) followed by
 798 centrifugation. Samples were taken from the supernatant (S), and after washing, the resuspended pellet
 799 (P) and separated in SDS-PAGE. Control reactions were without PGN.

800 B. Pull-down assays of IntPeri-MBP using PGN sacculi in solutions at different pH (upper gel). The lower
 801 gel is a control gel without PGN. S = supernatant, P = pellet.

802 C. Effect of divalent cations and chelators on PGN binding by IntPeri-MBP. Pull-down assays were
 803 performed in the presence of either divalent cations or the chelators EDTA or sodium citrate at pH 5.0. S
 804 = supernatant, P = pellet.

805 D. Effect of soluble muropeptides on the binding of IntPeri-MBP to PGN sacculi. Pull-down assays were
 806 performed with increasing amounts (5-25 μ l) of mutanolysin-digested PGN at pH 5.0. S = supernatant, P
 807 = pellet.

808 E. Binding of IntPeri-MBP, InvPeri-MBP and MBP to a chitin column. 1 mg of protein was loaded onto a
 809 chitin column equilibrated at pH 5.0. The flow through (FT) was collected, the column washed with

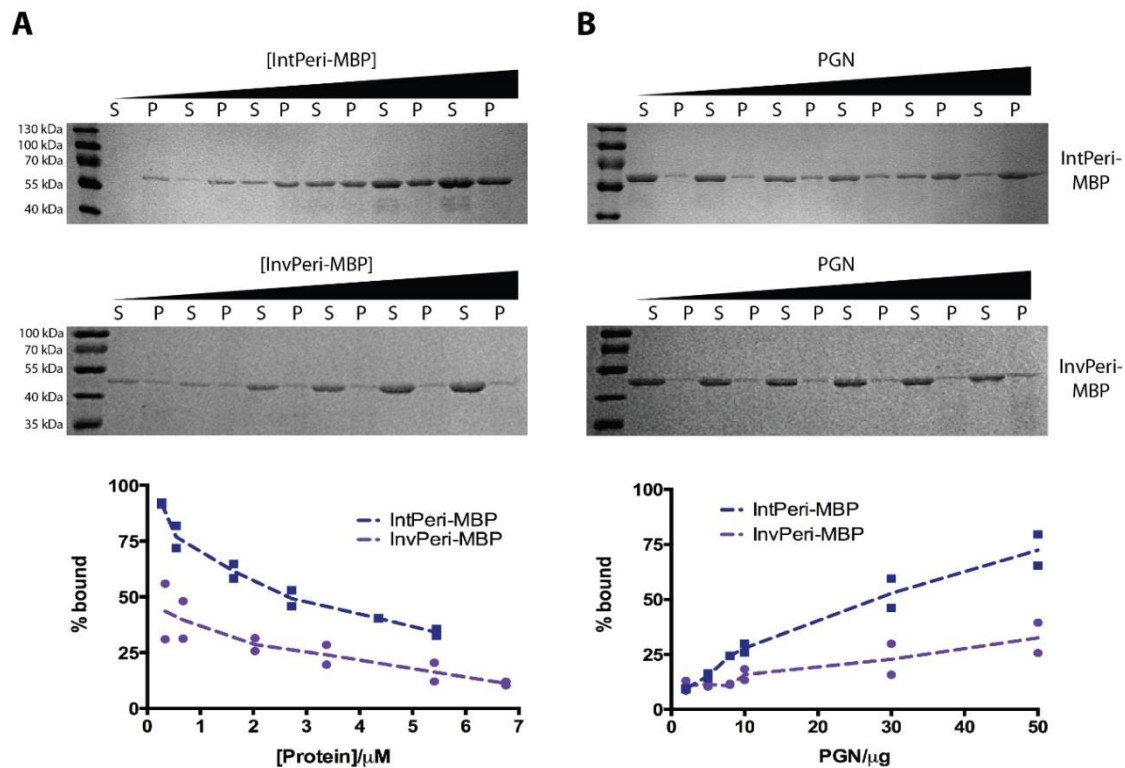
810 running buffer (w) and bound protein was eluted first with a buffer at pH 7.4 followed by elution with
811 0.3 M NaOH (OH).

812 F. Binding of IntLysM-MBP, IntHelix1-MBP and IntHelix2-MBP to PGN sacculi. Pull-down assays were
813 performed as in panel A at pH 5.0. S = supernatant, P = pellet.

814 The results shown are representative of at least two repeated experiments.

815

816



817

818 **Fig. 5.** Quantitative comparison of IntPeri-MBP and InvPeri-MBP binding to PGN.

819 A. Pull-down assays using a concentration series of IntPeri-MBP (upper gel) and InvPeri-MBP (lower gel)

820 against a constant amount of PGN at pH 5.0. The concentrations of IntPeri-MBP and InvPeri-MBP were

821 varied between 0.5 μ g and 10 μ g per reaction. The graph at the bottom shows relative binding of the

822 proteins to PGN, with the protein concentration expressed as micromoles per litre. Data points from two

823 replicate experiments represent the band intensity of the pellet divided by the sum of the supernatant

824 (S) and pellet (P) intensities for that concentration, multiplied by 100. The curves are plotted based on

825 the average of the two measurements. The bands intensities were quantified using the ImageJ software

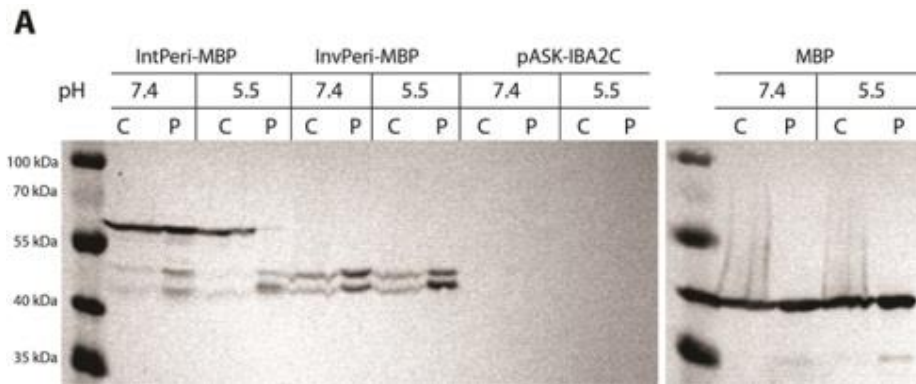
826 (Schneider *et al.*, 2012).

827 B. Pull-down assays using a varying amount of PGN sacculi against a constant amount of either IntPeri-

828 MBP (upper gel) or InvPeri-MBP (lower gel). The graph depicts relative binding for each PGN amount (in

829 micrograms). The binding percentage was calculated as for panel A. S = supernatant, P = pellet.

830

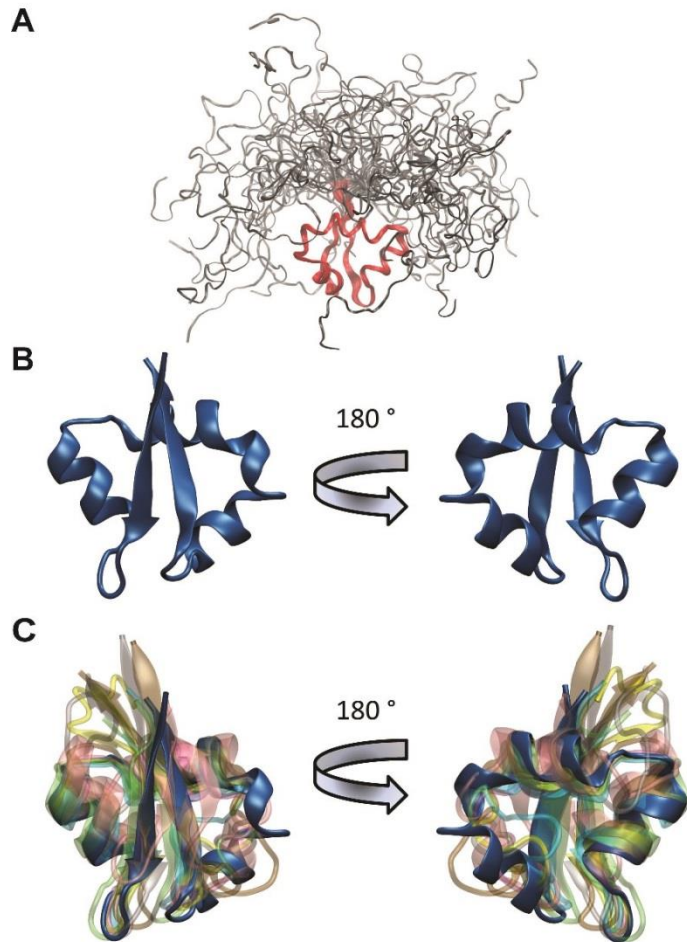


831

832 **Fig. 6.** The *Int* periplasmic domain binds to PGN *in vivo*.

833 *In vivo* binding to peptidoglycan. Cells expressing IntPeri-MBP or InvPeri-MBP periplasmically were
 834 grown at pH 7.4 or pH 5.5. The periplasmic fraction was extraction using a modified osmotic shock
 835 protocol. Samples from whole cells (C) or the periplasmic fraction (P) were separated by SDS-PAGE and
 836 subjected to Western blotting using an anti-MBP antibody. pIBA2C is the vector control, and MBP acts as
 837 a control for periplasmic extraction.

838



839

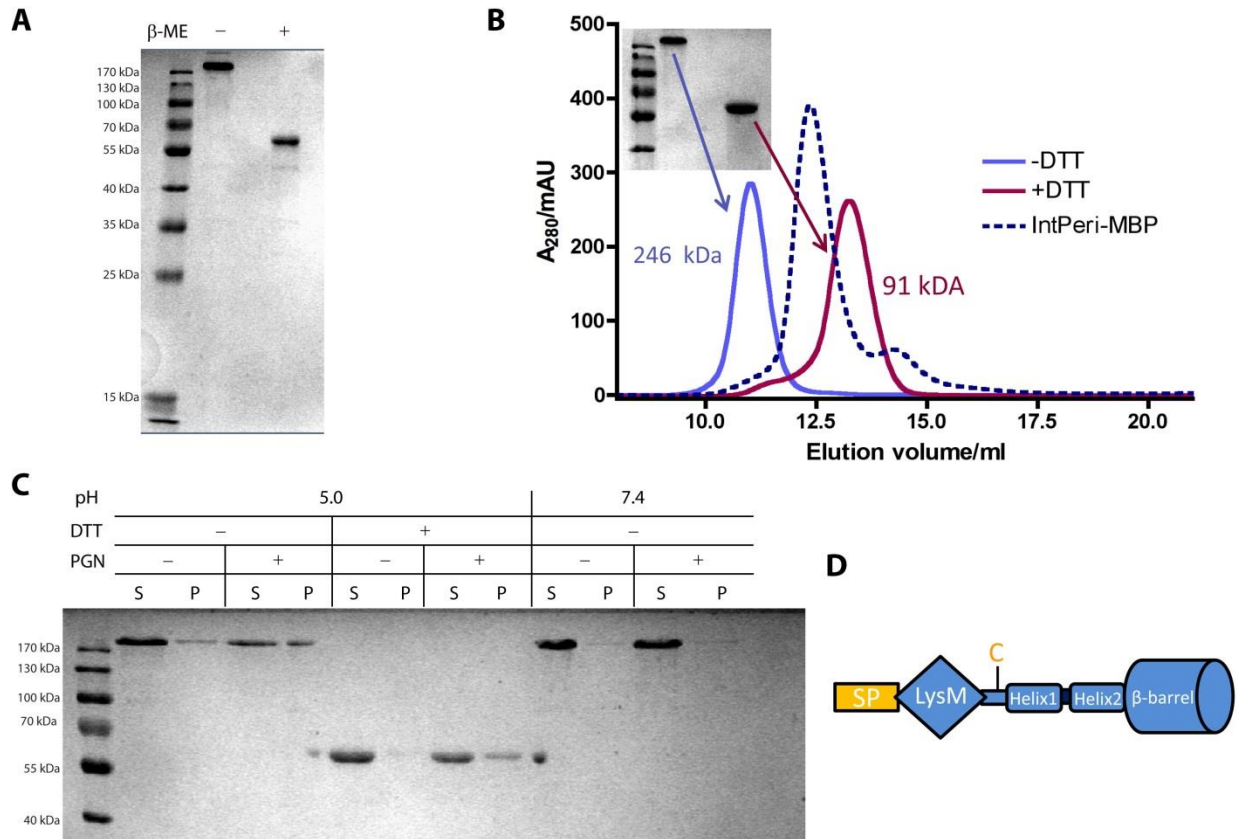
840 **Fig. 7.** Solution structure of Int LysM

841 A. Ensemble of 20 acceptable NMR structures for the Int LysM region, showing the LysM itself
 842 (highlighted in red) is well defined whereas the N-terminus and C-terminal spacer sequence are
 843 disordered.

844 B. The Int LysM structure in cartoon representation.

845 C. Superposition of of other LysM structures onto the Int LysM NMR structure (in blue). The structures
 846 include bacterial and bacteriophage LysMs (MltD from *E. coli* in green (PDB 1E0G)); YkuD from *B. subtilis*
 847 in mauve (1Y7M); gpX from coliphage P2 in salmon (2LTF)), LysMs from fungal protein (LysM1 from Ecp6
 848 of *Cladosporium fulvum* in yellow (4B8V); LysM of CVNH lectin from *Magnaporthe oryzae* in cyan
 849 (2L9Y)), a plant protein (LysM1 from AtCERK1 of *A. thaliana* in grey (4EBY)) and a human hypothetical
 850 protein (SB145 (2DJP)). The major difference in Int LysM compared to the others is the α -helical region
 851 between the C-terminal conserved α -helix and the C-terminal β -strand. The figures were prepared using
 852 VMD (Humphrey *et al.*, 1996, <http://www.ks.uiuc.edu/Research/vmd/>)

853



854

855 **Fig. 8.** The periplasmic domain of YrInv forms multimers and binds peptidoglycan.

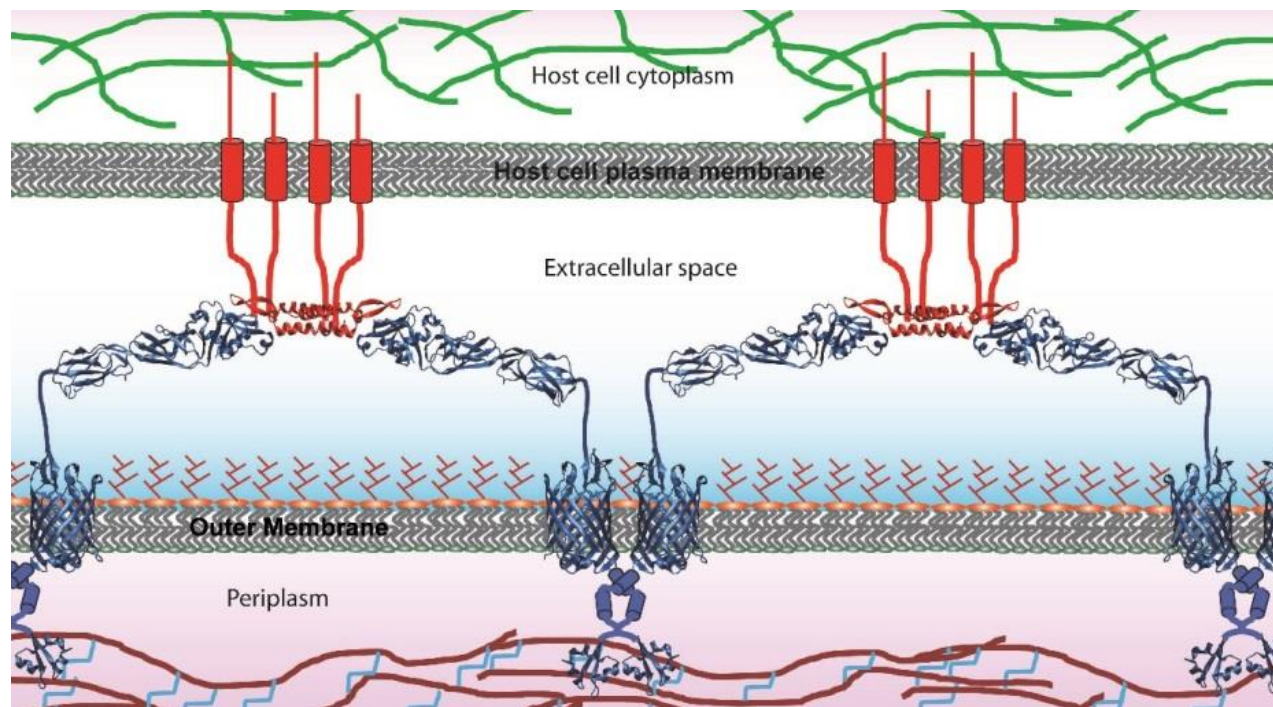
856 A. Effect of reducing agent on YrInvPeri-MBP. β -ME = β -mercaptoethanol.

857 B. SEC of YrInvPeri-MBP at pH 7.4 with (magenta) and without DTT(light blue). The inset shows an SDS-
 858 PAGE gel of the peaks (no additional reducing agent was added). The expected size for the monomer is
 859 58 kDa. Apparent molecular weights are indicated. The SEC curve for IntPeri-MBP (dashed blue line) is
 860 shown for comparison.

861 C: PGN binding of YrInvPeri-MBP. Pull-downs with or without PGN sacculi was performed at pH 5.0 or pH
 862 7.4. At pH 5.0, pull-down assays were performed either with or without the addition of 10 mM DTT. The
 863 small amount of monomeric YrInv seen at the edges of lanes 5 and 10 can be explained by diffusion of
 864 the DTT from the neighbouring wells. S = supernatant, P = pellet.

865 D. Schematic of YrInvPeri periplasmic domain and barrel. The position of the single cysteine in the
 866 spacer sequence is indicated.

867



868

869 **Fig. 9.** Model for Intimin dimerisation and Tir clustering

870 Schematic of Int (in blue) bound to Tir (red). Dimerisation of Int through the periplasmic domain in the
 871 periplasmic domain leads to a reticular array which promotes Tir clustering and pedestal formation.

872 Dimerisation takes place at two sites within the periplasmic domain: at the C-terminal α -helix and in the
 873 spacer sequence between the LysM and the N-terminal α -helix. Where structural information is

874 available, cartoon structures are depicted. Regions of unknown structure are depicted as connecting
 875 lines, or cylinders in the case of α -helices. The actin cortex of the host cell is in green, the peptidoglycan

876 backbone in brown and peptide crosslinks in light blue. PDB IDs of the structures used for the figure are

877 1F02, 4E1S and 2MPW. The structural figures were prepared using VMD (Humphrey *et al.*, 1996,

878 <http://www.ks.uiuc.edu/Research/vmd/>)