Molecular Microbiology (2015) 95(1), 80–100 ■
doi:10.1111/mmi.12840
First published online 24 November 2014

The Intimin periplasmic domain mediates dimerisation and binding to peptidoglycan

Jack C. Leo,[1†] Philipp Oberhettinger,[2] Manish Chaubey,[1] Monika Schütz,[2] Daniel Kühner,[3] Ute Bertsche,[3] Heinz Schwarz,[1] Friedrich Götz,[3] Ingo B. Autenrieth,[2] Murray Coles[1] and Dirk Linke[1,4*]

[1]*Department of Protein Evolution, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany.*

[2]*Interfaculty Institute for Microbiology and Infection Medicine, University Clinics Tübingen, 72076 Tübingen, Germany.*

[3]*Department of Microbial Genetics, University of Tübingen, 72076 Tübingen, Germany.*

[4]*Department of Biosciences, University of Oslo, 0316 Oslo, Norway.*

Summary

Intimin and Invasin are prototypical inverse (Type Ve) autotransporters and important virulence factors of enteropathogenic *Escherichia coli* and *Yersinia spp.* respectively. In addition to a C-terminal extracellular domain and a β-barrel transmembrane domain, both proteins also contain a short N-terminal periplasmic domain that, in Intimin, includes a lysin motif (LysM), which is thought to mediate binding to peptidoglycan. We show that the periplasmic domain of Intimin does bind to peptidoglycan both *in vitro* and *in vivo*, but only under acidic conditions. We were able to determine a dissociation constant of 0.8 μM for this interaction, whereas the Invasin periplasmic domain, which lacks a LysM, bound only weakly *in vitro* and failed to bind peptidoglycan *in vivo*. We present the solution structure of the Intimin LysM, which has an additional α-helix conserved within inverse autotransporter LysMs but lacking in others. In contrast to previous reports, we demonstrate that the periplasmic domain of Intimin mediates dimerisation. We further show that dimerisation and peptidoglycan binding are general features of LysM-containing inverse autotransporters. Peptidoglycan binding by the periplasmic domain in the infection process may aid in resisting mechanical and chemical stress during transit through the gastrointestinal tract.

Introduction

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

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

In addition to the passenger and translocation domains, both proteins contain a small N-terminal periplasmic domain, also referred to as the α-domain (Tsai *et al.*, 2010). In Int, the periplasmic domain contains a lysin motif (LysM) found in many peptidoglycan-binding proteins (Buist *et al.*, 2008). The LysM is a small domain consisting of a 2-stranded anti-parallel β-sheet packed against two α-helices with the topology β-α-α-β (Bateman and Bycroft, 2000). LysMs usually bind to the N-acetylglucosamine

Accepted 26 October, 2014. *For correspondence. E-mail dirk.linke@ibv.uio.no; Tel. (+47) 228 57654; Fax (+47) 228 54726. †Present address: Department of Biosciences, University of Oslo, 0316 Oslo, Norway.

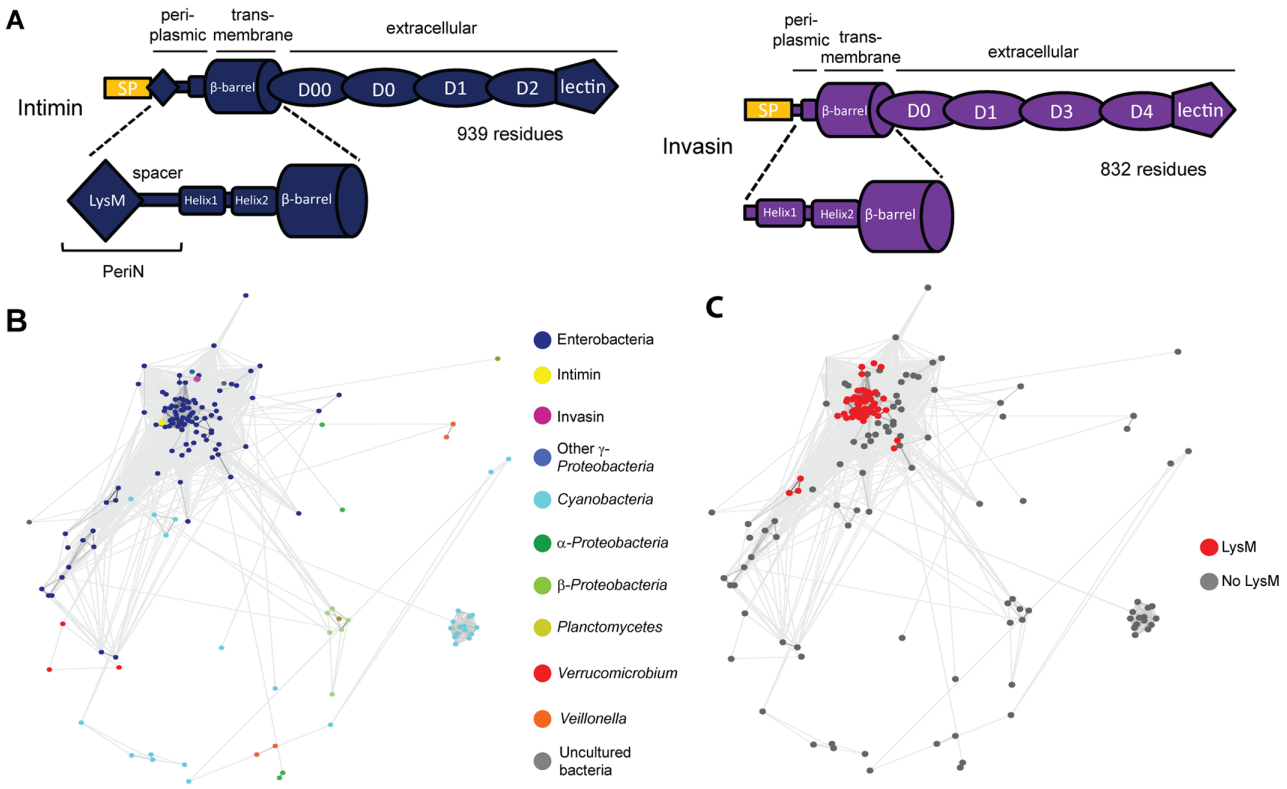


Fig. 1. Bioinformatic analyses of periplasmic domains of inverse autotransporters.
A. Schematic of structures of Int from EPEC (dark blue) and Inv from *Y. enterocolitica* (purple). Both proteins contain an N-terminal periplasmic domain, a transmembrane domain and an extracellular or passenger domain. The passenger domains contain tandemly repeated immunoglobulin-like domains [D00-D2 in Int and D0-4 in Inv; note that the D2 self-association domain (Dersch and Isberg, 2000) is not present in the *Y. enterocolitica* Inv] capped by a C-type lectin-like domain. The periplasmic domains of both (shown also in close-up) contain two conserved α-helices (Helix1 and Helix2) at the C-terminus. In addition, the Int periplasmic domain contains a LysM, which is connected to Helix1 by a spacer sequence. We refer to the N-terminal region of the Int periplasmic domain (containing the LysM and spacer) as IntPeriN, as shown in the close-up of the Int periplasmic domain. SP = signal peptide. The various domains are not to scale.
B. CLANS clustering of sequences of periplasmic domains from inverse autotransporters. One hundred fifty-two periplasmic domain sequences were clustered according to sequence similarity using the CLANS programme (Frickey and Lupas, 2004). Similar sequences cluster together; the grey lines represent BLAST connections below the *P* value cut-off (0.99). The dots representing individual sequences are coloured according to taxonomy. Int is indicated in yellow and Inv in magenta.
C. Highlighting of LysM-containing proteins in the cluster map from B. LysMs are restricted to sequences within the γ-proteobacterial clade. The sequences used for clustering are given is Table S1.

moieties in peptidoglycan (PGN) and chitin, or other carbohydrate structures (Buist *et al.*, 2008). In contrast to Int, the Inv periplasmic region lacks a LysM.

Due to the presence of a LysM in the Int periplasmic domain, it has long been suspected that the periplasmic domain would mediate binding to PGN (e.g. Bateman and Bycroft, 2000; Tsai *et al.*, 2010; Pisano *et al.*, 2012). However, up to now, this hypothesis has not been experimentally validated. In this study, we show that the periplasmic domain of Int does indeed bind PGN, but only under acidic conditions. In addition, we show that the periplasmic domain mediates dimerisation of Int. In contrast, the shorter periplasmic domain of Inv has neither of these functions. We also present a solution structure of the Int LysM and show that dimerisation and PGN binding are conserved functions of LysMs of inverse autotransporters.

Results

Bioinformatic analysis of periplasmic domains from inverse autotransporters

The Int periplasmic domain has previously been predicted to contain an N-terminal LysM and two conserved α-helices at the C-terminus (Tsai *et al.*, 2010). Using software of the online Bioinformatics Toolkit of the Max Planck Institute for Developmental Biology (http://toolkit.tuebingen.mpg.de/), we also confirmed the prediction of the LysM motif, both by homology searches using HHPred (Söding *et al.*, 2005) and secondary structure prediction with Quick2D and Ali2D. Our predictions also confirmed the two closely spaced C-terminal α-helices (Fig. 1A). The sequence connecting the LysM to the first of the C-terminal helices consists of approximately 30 residues and is pre-

dicted to be unstructured. We call this the spacer sequence. The periplasmic domain of Inv is also predicted to contain the α-helices, but to lack the LysM motif (Fig. 1A). An alignment of selected members of the Inv-Int family periplasmic domains shows that the C-terminal helices are conserved, whereas the LysM motif is only present in some members of the family (Fig. S1). Similar results were also obtained in an earlier study (Tsai *et al*., 2010).

To investigate the distribution of different types of periplasmic domains within inverse autotransporters, we performed clustering analysis of the periplasmic domain sequences using CLANS (Frickey and Lupas, 2004), which clusters sequences based on similarity from pairwise BLAST comparisons. A total of 172 sequences were included in the clustering (Table S1), and after singletons (sequences with *P* values for BLAST high-scoring segment pairs higher than 0.99) were removed, 152 sequences remained. The results (Fig. 1B) show that the periplasmic domains cluster largely based on taxonomic distribution rather than length or the presence of the LysM motif. Most of the sequences were derived from the *γ-Proteobacteria*, especially the *Enterobacteriaceae*, which form a large cluster. Int is located close to the centre of this cluster, whereas Inv is located more peripherally (Fig. 1B). Other distinct clusters are formed by cyanobacterial sequences and sequences from the β-proteobacterial genus *Bordetella*. Interestingly, periplasmic domains from the cyanobacterial genus *Synechococcus* form two distinct clusters that are only distantly connected to the enterobacterial cluster; however, as the synechococcal sequences are all very short, the separate clusters may be artefactual. The sizes of the sequences do not have a large effect on the clustering, as both large and small sequences are found in the same clusters (Fig. S2A). The LysM appears restricted to the large γ-proteobacterial cluster, as it was not detected in any of the other groups (Fig. 1C). The presence of a LysM can to some degree be predicted based on the length of the protein sequence; however, the correlation is not absolute and there are several large members of the family lacking a LysM (Fig. S2B).

The Int periplasmic domain mediates dimerisation

Int has been reported to form dimers *via* its β-barrel domain (Touzé *et al*., 2004). However, the Int β-barrel crystallises as a monomer (Fairman *et al*., 2012). This suggested to us that the dimerisation might be mediated by the periplasmic domain, part of which was included in the construct used by Touzé *et al*. (2004). To test this, we produced the periplasmic domains of Int (from EPEC) and Inv [from *Y. enterocolitica*, which does not dimerise (Dersch and Isberg, 2000)], as maltose-binding protein (MBP) fusions, named IntPeri-MBP and InvPeri-MBP, with the MBP as a C-terminal fusion and a hexahistidine tag on the N-termini of the periplasmic domains. The constructs are depicted schematically in Fig. 2A, and all plasmids used in the study are summarised in Table 1. The proteins were produced and purified from the cytoplasm of *E. coli*. IntPeri-MBP, in particular, is unstable, and degradation products were observed in cell lysates (data not shown); however, we were able to purify the proteins to high purity by passing the proteins through both a nickel and amylose column followed by size exclusion chromatography (SEC).

When we performed analytical SEC at physiological pH (7.4.), MBP and InvPeri-MBP run at the expected sizes of the monomer, whereas IntPeri-MBP gives two peaks: the major peak migrates at an apparent size of 152 kDa, which is somewhat higher than the 122 kDa expected for the dimer based on the amino acid sequence. The apparent size of the smaller peak (64 kD) is close to the expected molecular weight of the monomer (61 kDa) (Fig. 2B). Both peaks contain Int-MBP, suggesting an equilibrium between monomeric and dimeric forms. Indeed, when we reran the major peak through the SEC column, a similar distribution for dimer and monomer was observed (Fig. S3A).

As the apparent molecular weight of the putative IntPeri-MBP dimer was between the expected sizes of the dimer and trimer, we wanted to confirm the oligomeric state of the protein by *in vitro* crosslinking. To this end, we used the amine crosslinker bis[sulfosuccinimidyl] suberate (BS_3). When BS_3 was added to IntPeri-MBP, a smeary crosslinked product appeared (Fig. 2C). Most of the crosslinked product migrated at approximately ~ 120 kDa, which is the expected size of the dimer (Fig. 2C). InvPeri-MBP and MBP did not form crosslinked multimers (Fig. 2C). We thus conclude that the major IntPeri-MBP peak seen in SEC is dimeric.

To further characterise the region in the periplasmic domain responsible for dimerisation, we fused short fragments of the Int periplasmic domain to MBP (Fig. 2A). These fragments were the N-terminal region of the periplasmic domain containing the LysM domain along with the spacer sequence (IntPeriN-MBP), the two C-terminal helices (IntPeriHelix1-MBP) or just the second, C-terminal helix (IntPeriHelix2-MBP). In SEC, IntPeriN-MBP migrates as a dimer, whereas the other two constructs run as a monomer (Fig. 2D). Upon addition of BS_3, a clear dimeric band (~ 110 kDa) appeared for IntPeriN-MBP, close to the expected size of 108 kDa (Fig. 2E). This shows that the N-terminal region of the periplasmic domain, containing the LysM and the spacer, contains the dimerisation interface. However, we also saw very faint dimeric bands for the other two constructs (asterisks in Fig. 2E).

Touzé and coworkers showed that the β-barrel and flanking regions (residues 189–550) contain a dimerisation site (Touzé *et al*., 2004). As the β-barrel itself is monomeric, Fairman *et al*. speculated that the dimerisation interface

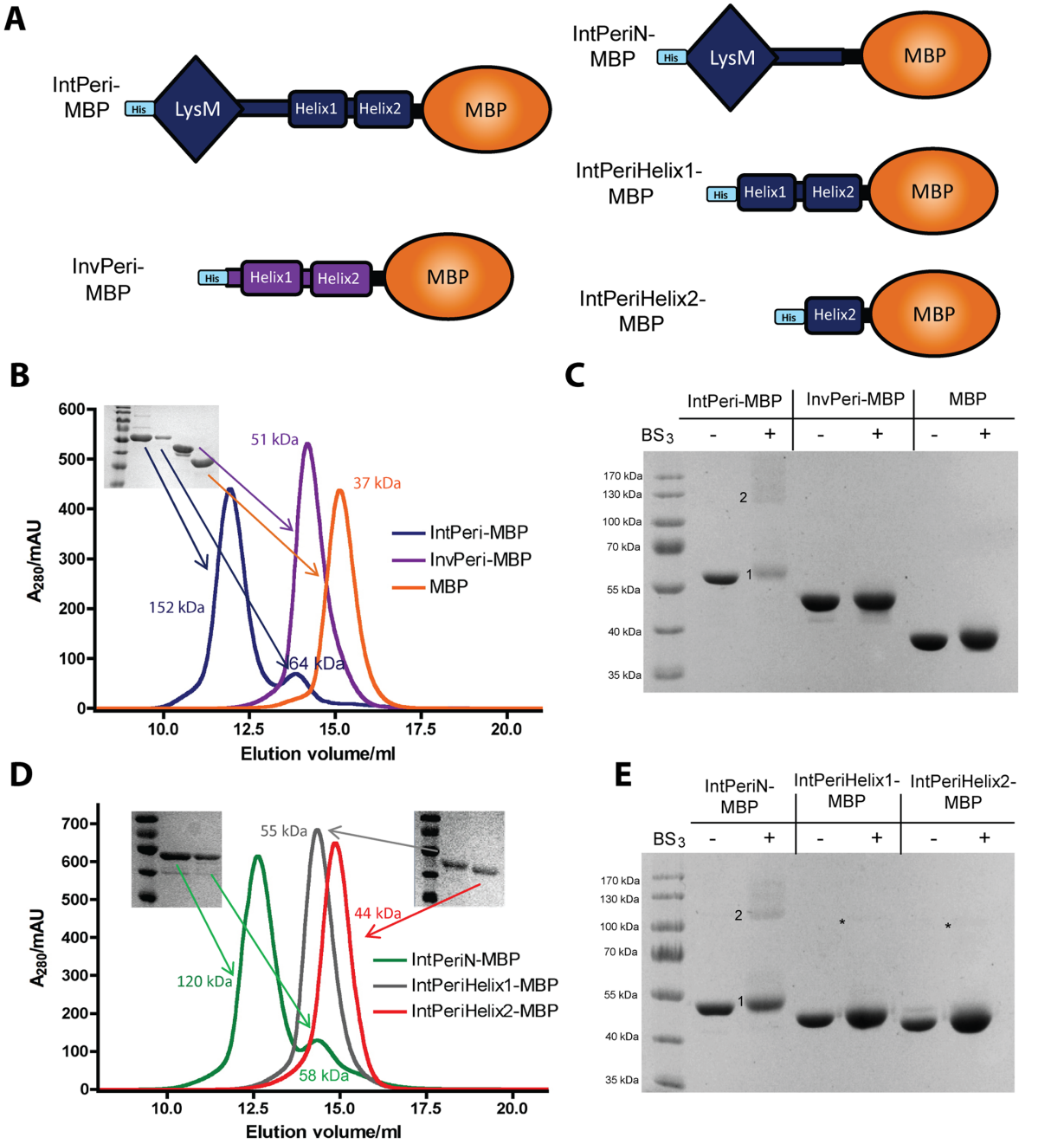


Fig. 2. The periplasmic domain of Int mediates dimerisation.
A. Schematic of periplasmic domain-MBP constructs. Fragments from Int are in blue, fragments from Inv in purple, MBP is in orange and the N-terminal hexahistidine tags in light blue.
B. SEC of Inv and Int periplasmic domains at pH 7.4. The curve for IntPeri-MBP is in blue, InvPeri-MBP in purple and MBP in orange. The inset shows an SDS-PAGE gel of the peaks. Apparent molecular weights of the major peaks are indicated. Expected molecular weights are 61 kDa for IntPeri-MBP, 49 kDa for InvPeri-MBP and 41 kDa for MBP.
C. *In vitro* crosslinking of MBP fusions. The proteins were treated with the amine crosslinker BS_3 and run in SDS-PAGE (10%). Samples without crosslinker serve as controls. The monomeric and assumed dimeric bands for IntPeri-MBP are indicated with the numbers 1 and 2 respectively.
D. SEC of Int periplasmic domain fragments. IntPeriN-MBP is in green, IntPeriHelix1-MBP in grey and IntPeriHelix2-MBP in red. The inset shows SDS-PAGE gels of the peaks. Apparent molecular weights of the major peaks are indicated. Expected molecular weights are 54 kDa for IntPeriN-MBP, 49 kDa for IntPeriHelix1-MBP and 46 kDa for IntPeriHelix2-MBP.
E. *In vitro* crosslinking of Int periplasmic domain fragments using BS_3 as above. The monomeric and dimeric bands for IntPeriN-MBP are indicated by the numbers 1 and 2 respectively. The asterisks denote faint, assumed dimeric bands seen when BS_3 was added to IntPeriHelix1-MBP and IntPeriHelix2-MBP.

Table 1. Plasmids used in this study.

Plasmid	Insert sequence	Vector	Comment	Source
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 for periplasmic transport	This study
pIBA33-PeriN-TEV-MBP	Int LysM domain and spacer sequence (residues 40–153)	pASK-IBA33	For production of Int LysM domain with C-terminal MBP fusion separated by a TEV protease site in the cytoplasm	This study
pIBA33-MBP	MBP	pASK-IBA33	Produces MBP with a C-terminal His-tag in the cytoplasm	This study
pIBA3-IntD00	Int D00 domain (residues 450–550)	pASK-IBA3	Produces Int N-terminal Ig-like domain in the cytoplasm	
pIBA3-IntPeriN-MBP	Int LysM domain and spacer sequence (residues 40–153)	pASK-IBA3	Produces N-terminal region of Int periplasmic domain (containing the LysM and spacer sequence) with a C-terminal MBP fusion in the cytoplasm	This study
pIBA3-IntPeriHelix1-MBP	2 C-terminal α-helices of Int periplasmic domain (residues 157–212)	pASK-IBA3	Produces conserved α-helices of Int periplasmic domain with a C-terminal MBP fusion in the cytoplasm	This study
pIBA3-IntPeriHelix2-MBP	C-terminal α-helix of Int periplasmic domain (residues 184–212)	pASK-IBA3	Produces second conserved α-helix of Int periplasmic domain with a C-terminal MBP fusion in the cytoplasm	This study
pIBA3-IntPeri-MBP	Int periplasmic domain (residues 40–212)	pASK-IBA3	Produces IntPeri with a C-terminal MBP fusion in the cytoplasm	This study
pIBA3-InvPeri-MBP	Inv periplasmic domain (residues 35–93)	pASK-IBA3	Produces Inv periplasmic domain with a C-terminal MBP fusion in the cytoplasm	This study
pIBA3-YrInvPeri-MBP	*Y. ruckeri* Invasin periplasmic domain (residues 49–195)		Produces YrInv periplasmic domain with a C-terminal MBP fusion in the cytoplasm	This study

would either be in the periplasmic region or the first extracellular Ig-like domain (D00) (Fairman *et al.*, 2012). Our results above suggest a dimerisation site in the N-terminal region of the periplasmic domain. We therefore produced and purified the D00 domain and tested its oligomerisation status by SEC and crosslinking. Both methods show the domain to be monomeric (Fig. S4).

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

We performed pull-down assays with our MBP fusion proteins using purified PGN sacculi from *E. coli* (EcPGN) to determine whether the Int periplasmic domain binds PGN (Fig. 3). We fortuitously found that binding was strongest at low pH: in assays performed at pH 5.0, approximately half the IntPeri-MBP fusion precipitates with the sacculi, while the MBP control remains in the supernatant (Fig. 3A). In this assay, Inv-MBP displayed only background-level binding to PGN (Fig. 3A). In the absence of EcPGN, all three proteins remained soluble. At normal physiological pH (7.4), however, we saw no binding of IntPeri-MBP to EcPGN, with most of the protein remaining in the supernatant (Fig. 3A).

To examine the effect of pH on the binding of IntPeri-MBP to EcPGN more systematically, we performed binding experiments at a range of pH values between 8.0 and 3.0; strong binding first became evident at pH 6.0 and the ratio of bound IntPeri-MBP to soluble protein remains constant even with reduction in pH value (Fig. 3B). In the control experiment without PGN sacculi, the protein remains soluble at all pH values tested (Fig. 3B). Int-MBP still dimerises at pH 4.0 (Fig. S3B), demonstrating that dimerisation is independent of pH.

The PGN composition of *Y. enterocolitica* and *E. coli* is very similar, though there is some difference in the relative abundance of particular muropeptides (Quintela *et al.*, 1995). However, to rule out that InvPeri-MBP might bind

with high affinity to *Y. enterocolitica* PGN (YePGN), we isolated PGN sacculi from *Y. enterocolitica*. Ultra performance liquid chromatography analysis of EcPGN and YePGN shows that the composition of the two is near-identical (Fig. S5, Table S3). In pull-down assays, we observed similar results to the assays performed with EcPGN: IntPeri-MBP precipitates in appreciable amounts with the sacculi, but InvPeri-MBP remains mostly and MBP completely soluble (Fig. 3C). As for EcPGN, the binding of IntPeri-MBP to YePGN is dependent on low pH (Fig. 3C). We verified the results seen in our pull-down assays using a solid-phase binding assay (SPBA). Here, IntPeri-MBP bound clearly above background levels to both EcPGN and YePGN, but InvPeri-MBP bound only marginally above background to both (Fig. 3F).

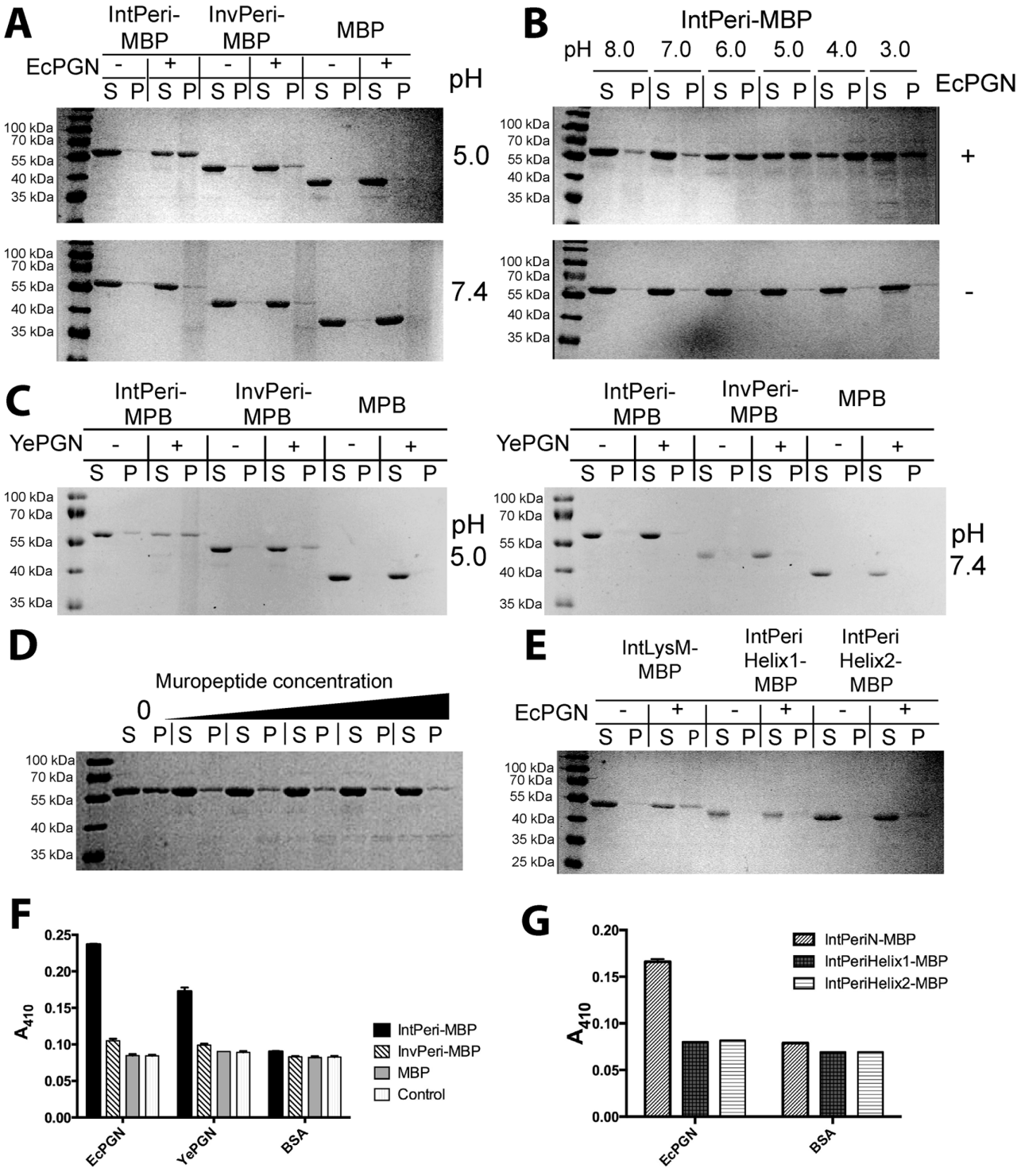
We partially digested EcPGN sacculi with mutanolysin and performed a competition assay with the soluble digestion products. The addition of the degradation products inhibited the binding of IntPeri-MBP to the sacculi in a dose-dependent manner, though we did not observe complete inhibition of binding to sacculi (Fig. 3D). This demonstrates that soluble muropeptides can compete for binding to IntPeri-MBP.

We also tested the effect of physiologically abundant divalent cations (Ca^{2+}, Mg^{2+}, Mn^{2+}, Zn^{2+}) on the binding of IntPeri to EcPGN, but we did not observe any difference to controls (Fig. S6A). To examine the stability of IntPeri-MBP binding to EcPGN, we performed a pull-down assay as above, and then washed the pellet multiple times with ABS at pH 5.0. The amount of bound EcPGN was only slightly diminished after five consecutive washes (Fig. S6B). This shows that the Int-PGN interaction is stable at low pH. However, even a single wash at pH 7.4 reduced the amount of bound IntPeri-MBP and virtually no IntPeri-MBP remained bound after three washes at pH 7.4 (Fig. S6B).

To find out which part of the periplasmic domain of Int is responsible for PGN binding, we performed pull-down assays with IntPeriN-MBP, IntPeriHelix1-MBP and IntPeriHelix2-MBP (Fig. 3E). In this assay, only IntPeriN-MBP coprecipitated with the sacculi in appreciable amounts. Similarly, only IntPeriN-MBP bound strongly to EcPGN in SPBA (Fig. 3G). This demonstrates that the N-terminal region of IntPeri mediates PGN binding.

Quantitative analysis of PGN binding

As we saw some background levels of InvPeri-MBP bound to PGN, we wished to further investigate the relative affinities of InvPeri-MBP and IntPeri-MBP. To this end, we performed pull-down assays using a concentration series of IntPeri-MBP and InvPeri-MBP with a constant amount of EcPGN. At low concentrations of IntPeri-MBP, almost all the protein is pulled down with the sacculi, whereas the fraction in the pellet decreases steadily with increasing protein concentration (Fig. 4A). InvPeri-MBP follows a similar trend, but even at low concentrations only approximately 50% are pulled down with the sacculi, and at higher concentrations less than 20% remain in the pellet (Fig. 4A). In the reverse experiment, where we varied the concentration of EcPGN but kept the protein concentration constant, only a small fraction of IntPeri-MBP precipitates with low amounts of EcPGN, but as the amount of EcPGN rises, increasingly more IntPeri-MBP is found in the pellet (Fig. 4B). The increase in roughly linear, and at the highest concentration tested (50 μg), 75% of the protein are in the pellet. Again, InvPeri-MBP follows a similar trend, but the increase in the fraction of pelleted protein is not as large, and at most only 30% of the protein were pulled down with the sacculi (Fig. 4B). It appears that IntPeri-MBP binds with high affinity to EcPGN, as most of the protein precipitates at low protein concentration. However, the fraction of bound protein declines more or less linearly with increasing concentration of IntPeri-MBP, which we interpret to mean that there are a limited number of high-affinity binding sites in the sacculi and that these are quickly saturated. The amount of IntPeri-MBP scales linearly with increasing EcPGN, which is consistent with this interpretation. However, also InvPeri-MBP follows similar trends, though the binding levels are much lower than for IntPeri-MBP. When we incubated sacculi with both IntPeri-MBP and InvPeri-MBP, we did not see any effect of high concentrations of InvPeri-MBP on the binding levels of IntPeri-MBP, suggesting that these proteins do not compete for the same binding site(s) and that the binding of InvPeri-MBP is indeed unspecific (Fig. S6C and D). The results obtained using a concentration series of either protein or EcPGN should be viewed with caution, however, as the assays used are rather crude and include washing steps which we did not take into account when calculating the relative binding, and even in the absence of PGN we often observe faint background bands in the pellets (Fig. 3). In addition, the pull-down assays do not cover a wide-enough range to estimate reliable dissociation constants. We therefore measured the binding of a wider concentration range of IntPeri-MBP, InvPeri-MBP and MBP to EcPGN with the SPBA. Using this method, we probed a concentration series between 0.01 μM and 10 μM. When a hyperbolic binding curve was fitted to the resulting absorbance values, an apparent dissociation constant (K_d) of 0.8 ± 0.1 μM was obtained for IntPeri-MBP (Fig. 4C). For InvPeri-MBP, the results gave a K_d of 6.3 ± 0.9 μM, almost an order of magnitude larger than for IntPeri-MBP (Fig. 4C). No K_d value could be estimated for MBP. We thus conclude that IntPeri-MBP binds with moderately high affinity to PGN, and InvPeri-MBP binds with significantly lower affinity, which we assume is due to unspecific binding, though we cannot rule out a specific but



low-affinity interaction between InvPeri-MBP and the PGN sacculi.

The Int periplasmic domain binds to PGN in vivo

To determine whether the Int periplasmic domain can mediate PGN binding *in vivo*, we cloned the MBP fusions into pIBA2C, which contains a signal sequence for periplasmic targeting. We used chloramphenicol selection rather than ampicillin to avoid any structural changes to PGN. The protein was expressed in medium buffered at either pH 7.4 or pH 5.5. To see if the MBP fusion proteins bound to PGN, we extracted the periplasmic fraction by osmotic shock. All solutions were

Fig. 3. The Int periplasmic domain binds peptidoglycan under acidic conditions.
A. Pull-down assays using purified peptidoglycan sacculi from *E. coli* (EcPGN). IntPeri-MBP, InvPeri-MBP and MBP were mixed with peptidoglycan sacculi at pH 5.0 (upper gel) or pH 7.4 (lower gel) followed by centrifugation. Samples were taken from the supernatant (S), and after washing, the resuspended pellet (P) and separated in SDS-PAGE. Control reactions were without PGN.
B. Pull-down assays of IntPeri-MBP using PGN sacculi in solutions at different pH (upper gel). The lower gel is a control gel without PGN. S = supernatant, P = pellet.
C. Pull down assays using peptidoglycan sacculi from *Yersinia enterocolitica* (YePGN). The assays were performed as in A. The assays shown in the left-hand gel were performed at pH 5.0, and at pH 7.4 for the right-hand gel. S = supernatant, P = pellet.
D. Effect of soluble muropeptides on the binding of IntPeri-MBP to PGN sacculi. Pull-down assays were performed with increasing amounts (5–25 μl) of mutanolysin-digested PGN at pH 5.0. S = supernatant, P = pellet.
E. Binding of IntPeriN-MBP, IntHelix1-MBP and IntHelix2-MBP to PGN sacculi. Pull-down assays were performed as in A at pH 5.0. S = supernatant, P = pellet.
F. Binding of IntPeri-MBP, InvPeri-MBP and MBP to PGN by solid-phase binding assay. Peptidoglycan sacculi from either *E. coli* (EcPGN) or *Y. enterocolitica* (YePGN) were adsorbed onto wells of a microtitre plate. After blocking with BSA, the ligand proteins were added at 1 μM. Bound proteins were detected using an anti-MBP antiserum. BSA-coated wells served as the background control. The control samples lacked the ligand protein, i.e. they show the background levels generated by the antibodies alone. The bars show the mean and error bars denote the standard error of the mean (SEM) from three replicate wells.
G. Binding of truncated IntPeri constructs to EcPGN by solid-phase binding assay. The assay was performed as in F. The bars show the mean and error bars denote the SEM from three replicate wells.
The results shown are representative of at least two repeated experiments.

buffered at either pH 7.4 or pH 5.5. We then detected the protein by Western blot using an anti-MBP antiserum. For the whole-cell samples of IntPeri-MBP, a strong band is evident at ~ 60 kDa, similar to the expected size of 61 kDa, along with a small amount of apparent degradation products (Fig. 5). We assume that the 60 kDa band corresponds to intact IntPeri-MBP. At pH 7.4, this band is also in the periplasmic fraction, but at pH 5.5 the 60 kDa band is barely visible. InvPeri-MBP is less stable, with two stronger bands: ~ 50 kDa, which is presumably the intact protein (expected size 49 kDa), and ~ 45 kDa, which we assume is a degradation product. Both bands are present at similar levels in all samples. The vector control does not display any bands, demonstrating the specificity of the antiserum, whereas in our control for periplasmic extraction (pIBA2C-MBP), the MBP band (expected size 41 kDa) is present in all the samples at equal intensity (Fig. 5). As the 60 kDa IntPeri-MBP band is present in the 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 retained in the periplasm due to PGN binding. This finding is consistent with our pull-down assays, as is the observation that InvPeri-MBP is extracted at similar amounts at both pH values, which supports the conclusion that it does not bind to PGN.

We also attempted to repeat the assay using *Y. enterocolitica* cells. However, in the strain we used, periplasmic extraction was very inefficient at pH 5.5 (data not shown). The reason for this is unclear. Even at pH 6.0, periplasmic extraction was not efficient, though we were able to extract some MBP (Fig. S6). However, we did not see extraction for any of the bands in the IntPeri-MBP or InvPeri-MBP samples at pH 6.0 (Fig. S7). Therefore, we cannot be certain whether IntPeri-MBP or InvPeri-MBP bind to PGN in *Y. enterocolitica* at pH 6.0.

Preliminary investigation into determinants in EcPGN for Int binding

Many LysM-containing proteins also bind chitin (Buist *et al.*, 2008). We therefore ran IntPeri-MBP, InvPeri-MBP and MBP as a control through a chitin column. However, we did not observe binding for any of the constructs (Fig. 6A). This suggests that IntLysM binds to other determinants in PGN than just the N-acetylglucosamine – N-acetylmuramic acid backbone. However, the disaccharide may still form part of the binding interface.

To gain further insight into the determinants in PGN for Int binding, we employed the mutant strain D456 (Potluri *et al.*, 2010). This strain lacks penicillin-binding proteins (PBPs) 4, 5 and 6. PBP5 has D,D-carboxypeptidase activity that, in wild-type *E. coli*, cleaves the terminal D-Ala residue from the pentapeptide attached to the PGN disaccharide. The D456 strain, which lacks PBP5, is thus enriched for pentapeptides (Kraus and Höltje, 1987; Fig. S5). We investigated the effect of increased pentapeptide content of PGN on IntPeri-MBP binding by pull-down assay, where IntPeri-MPB clearly bound to PGN sacculi isolated from D456 (D456 PGN), though apparently not as strongly as to EcPGN (Fig. 6B). However, in SPBA, the binding of IntPeri-MBP to D456 PGN appeared slightly tighter than for EcPGN (Fig. 6C). This finding was corroborated when we estimated the K_d for the IntPeri-MBP – D456 PGN interaction by SPBA: the apparent affinity of IntPeri-MBP for D456 PGN was 0.45 ± 0.04 μM, i.e. a factor of two tighter than for EcPGN (Fig. 6D). We also tested *in vivo* binding in the strain D456. This strain behaved essentially like BL21Gold, and IntPeri-MBP was retained in the periplasm at pH 5.5 (Fig. 6E). These data show that IntPeri-MBP binds also to PGN with increased pentapeptide content, and the binding is slightly stronger than for wild-type PGN.

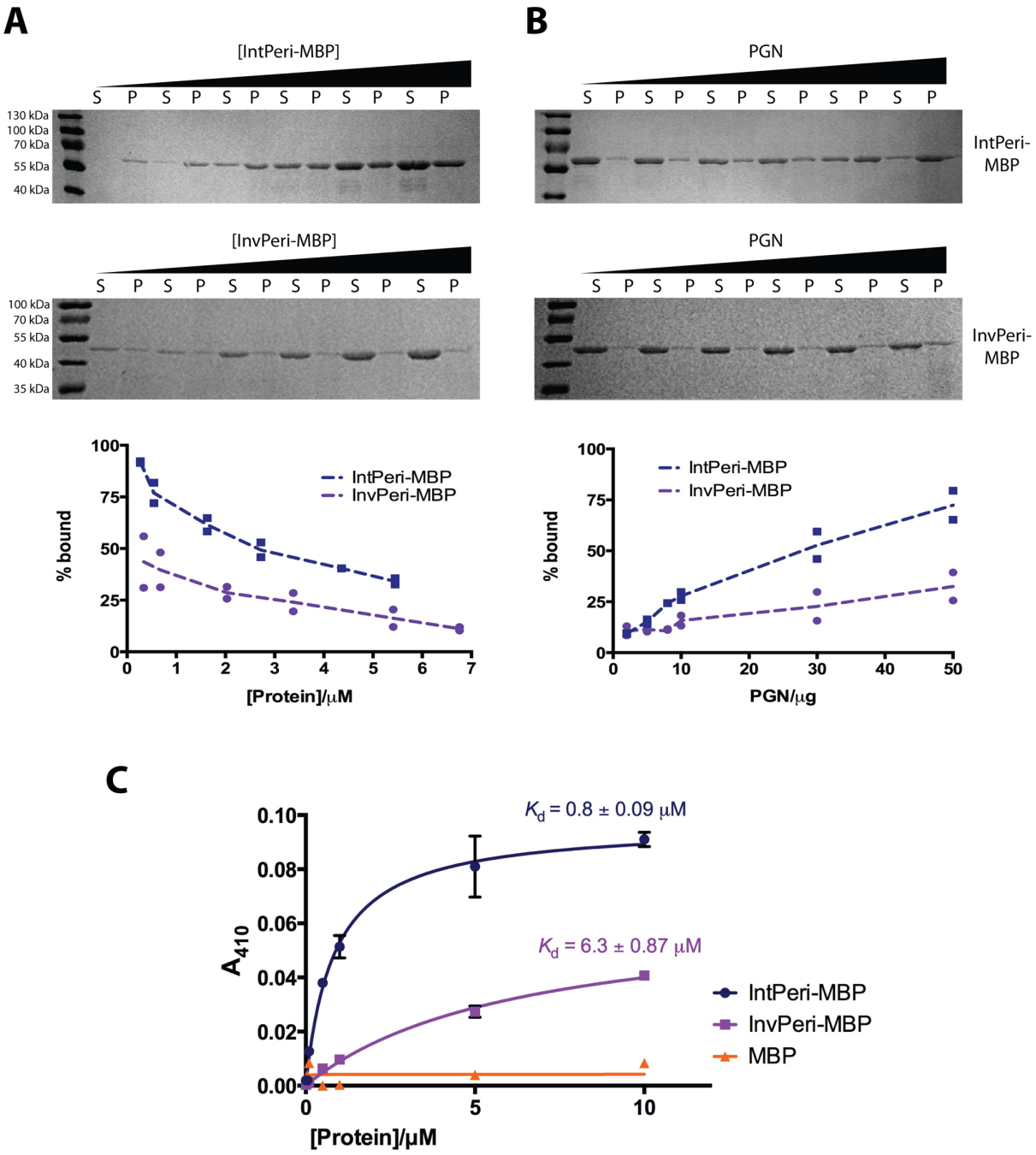


Fig. 4. Quantitative comparison of IntPeri-MBP and InvPeri-MBP binding to PGN.
A. Pull-down assays using a concentration series of IntPeri-MBP (upper gel) and InvPeri-MBP (lower gel) against a constant amount of PGN at pH 5.0. The concentrations of IntPeri-MBP and InvPeri-MBP were varied between 0.5 μg and 10 μg per reaction. The graph at the bottom shows relative binding of the proteins to PGN, with the protein concentration expressed as micromoles per litre. Data points from two replicate experiments represent the band intensity of the pellet divided by the sum of the supernatant (S) and pellet (P) intensities for that concentration, multiplied by 100. The curves are plotted based on the average of the two measurements. The bands intensities were quantified using the ImageJ software (Schneider *et al*., 2012).
B. Pull-down assays using a varying amount of PGN sacculi against a constant amount of either IntPeri-MBP (upper gel) or InvPeri-MBP (lower gel). The graph depicts relative binding for each PGN amount (in micrograms). The binding percentage was calculated as for A.
S = supernatant, P = pellet.
C. Estimation of apparent dissociation constants by solid-phase binding assay. The assay was performed as in Fig. 3F, with a series of 7 concentrations (0.01–10 μM) for each protein. Data points show the mean and error bars denote SEM from three replicate wells. After background subtraction, a hyperbolic binding curve was fitted to the data points by non-linear regression. Curve fitting and analysis was done using GraphPad Prism®.

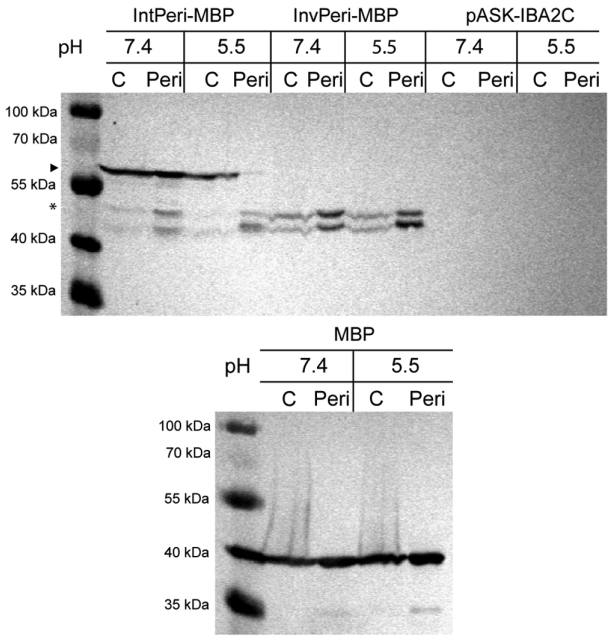


Fig. 5. The *Int* periplasmic domain binds to PGN *in vivo*. *In vivo* binding to EcPGN. *E. coli* BL21Gold expressing IntPeri-MBP or InvPeri-MBP periplasmically were grown at pH 7.4 or pH 5.5. The periplasmic fraction was extracted using a modified osmotic shock protocol. Samples from whole cells (C) or the periplasmic fraction (Peri) were separated by SDS-PAGE and subjected to Western blotting using an anti-MBP antibody. pIBA2C is the vector control, and MBP acts as a control for periplasmic extraction. The arrowhead marks the position of the ~ 60 kDa intact IntPeri-MBP band (expected 61 kDa) and the asterisk the ~ 50 kDa band corresponding to intact InvPeri-MBP (expected 49 kDa).

Structure of the Intimin LysM

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

When measuring the NMR spectra, we noted the rapid decay of several signals, with shifts suggesting a structured environment and the appearance of others more consistent with unstructured protein chain. The signals appearing were assigned to the C-terminus of the protein, i.e. in the spacer sequence. Despite extensive efforts, we could not obtain high-resolution structural information from the transient signals, although chemical shifts suggest that this region has considerable β-sheet content. We conclude that this structure is unstable and unfolds once the C-terminal fusion partner is removed. At equilibrium, both the N-terminus (residues 40–60) and the C-terminus (residues 115–143) are disordered (Fig. 7A). This is consistent with bioinformatics predictions (Fig. S1). However, the LysM itself was well defined (Fig. 7A). The NMR constraint and refinement statistics are presented in Supporting Information Table S2. It is noteworthy that we could not detect the dimerisation interface in NMR measurements on mixtures of differentially labelled protein. This is probably due to the unfolding of the spacer sequence that presumably contains the dimerisation site.

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

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

Several inverse autotransporters from the *Enterobacteriaceae* contain a periplasmic LysM domain (Fig. 1). To test whether binding to PGN and dimerisation are general features of inverse autotransporter LysMs, we cloned the LysM-containing periplasmic domain of an Inv homologue from the fish pathogen *Yersinia ruckeri* (GI: 238705545), which we refer to as YrInv for *Y. ruckeri* Invasin. The periplasmic domain of YrInv (YrInvPeri) has a similar architecture to IntPeri (Fig. 8A). We produced the YrInvPeri as an MBP fusion (YrInvPeri-MBP) (Fig. 8B). During later purification steps, we noticed that in sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS-PAGE), the band of the expected size (58 kDa) began to lose intensity and a new band migrating at ~ 200 kDa appeared, and almost all the protein was in this higher molecular-weight band after the SEC step. This sug-

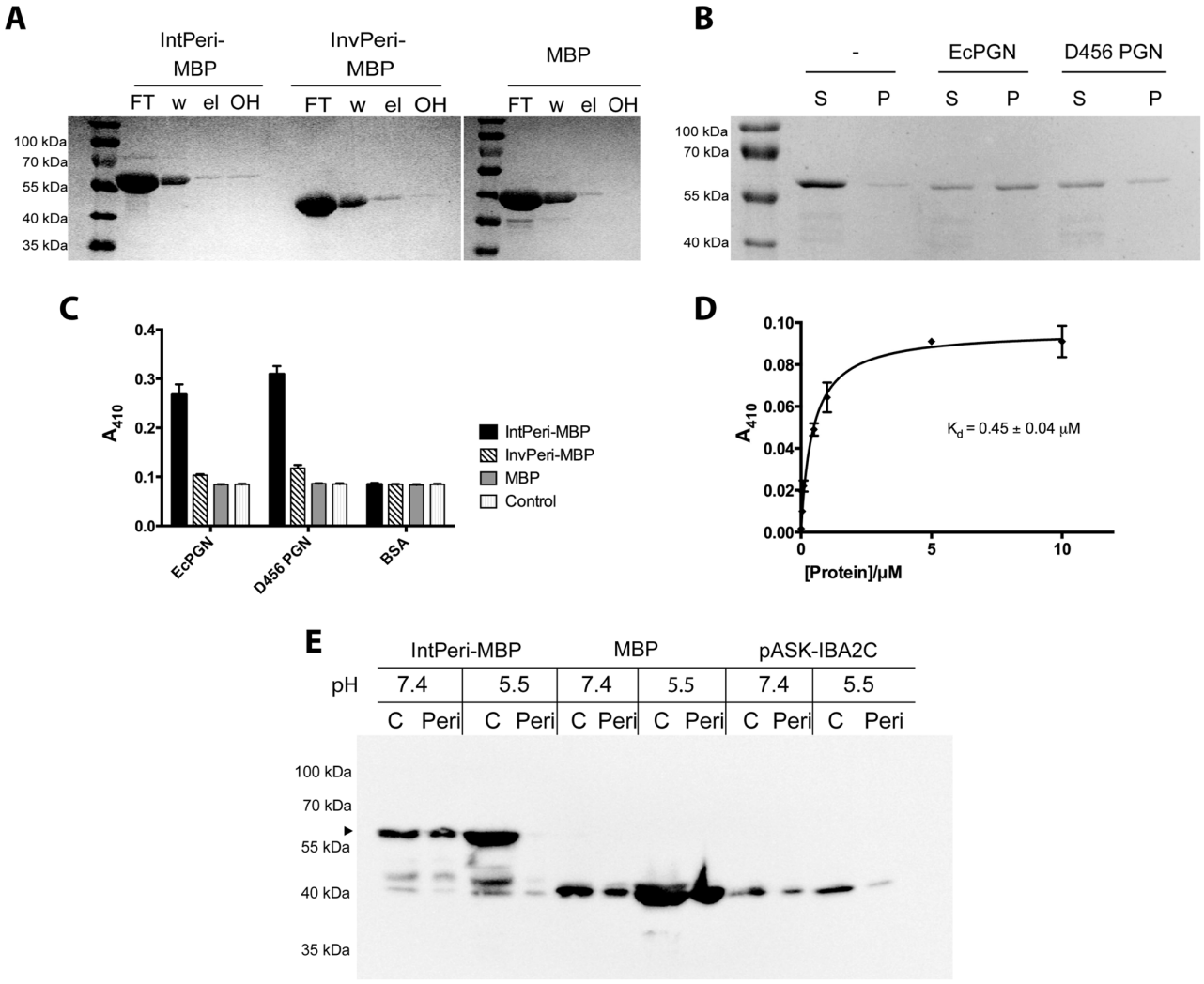


Fig. 6. Binding of IntPeri to chitin and PGN enriched with pentapeptides.
A. Binding of IntPeri-MBP, InvPeri-MBP and MBP to a chitin column. One milligram of protein was loaded onto a chitin column equilibrated at pH 5.0. The flow through (FT) was collected, the column washed with running buffer (w) and bound protein was eluted first with a buffer at pH 7.4 (el) followed by elution with 0.3 M NaOH (OH).
B. Pull-down assays with IntPeri-MBP using purified peptidoglycan sacculi from *E. coli* D456 (D456 PGN). As controls, IntPeri-MBP with EcPGN and IntPeri-MBP without PGN were included. S = supernatant, P = pellet.
C. Binding of IntPeri-MBP, InvPeri-MBP and MBP to D456 by solid-phase binding assay. The assay was performed as in Fig. 3F. The bars show the mean and error bars denote the SEM from three replicate wells.
D. Estimation of apparent dissociation constant of IntPeri-MBP for D456 PGN. The assay was performed as in Fig. 4C. Error bars denote SEM from three replicate wells for each data point. Curve fitting and analysis was done using GraphPad Prism®.
E. *In vivo* binding to D456 PGN. The assay was performed in *E. coli* D456 as in Fig. 5. The ~ 60 kDa IntPeri-MBP band is indicated with an arrowhead; the approximate positions of the markers are shown. The band seen in the vector control is probably due to endogenous MBP. C = whole cells, Peri = periplasmic fraction.

gested to us that YrInv was forming multimers stabilised by disulphide bridges, which were not disrupted in our SDS-PAGE experiments because we do not routinely add reducing agent to our sample buffer. When we added reducing agent to the sample buffer, all the protein previously in the ~ 200 kDa band ran at the expected size of the monomer, showing that the multimer resulted from the formation of disulphides by oxidation during purification (Fig. 8B). Indeed, there is a single cysteine in the periplasmic domain in the spacer sequence between the LysM and Helix1, which presumably mediates the disulphide formation (Fig. 8A). When we performed analytical SEC, YrInvPeri-MBP migrated with an apparent size of ~ 240 kDa (Fig. 8C). However, if the buffer was supplemented with dithiothreitol (DTT), the protein migrated with an apparent molecular weight of 91 kDa (Fig. 8C). As the expected size of the monomer is 58 kDa, we interpret the 91 kDa peak to represent the dimeric form of the mol-

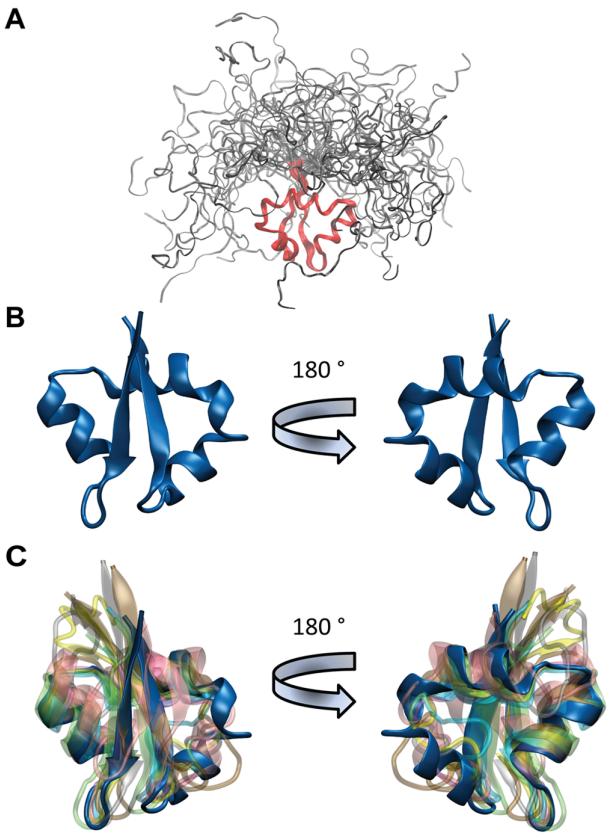


Fig. 7. Solution structure of Int LysM.
A. Ensemble of 20 acceptable NMR structures for the Int LysM region, showing the LysM itself (highlighted in red) is well defined whereas the N-terminus and C-terminal spacer sequences are disordered.
B. The Int LysM structure in cartoon representation.
C. Superposition of of other LysM structures onto the Int LysM NMR structure (in blue). The structures include bacterial and bacteriophage LysMs [MltD from *E. coli* in green (PDB 1E0G); YkuD from *B. subtilis* in mauve (1Y7M); gpX from coliphage P2 in salmon (2LTF)], LysMs from fungal protein [LysM1 from Ecp6 of *Cladosporium fulvum* in yellow (4B8V); LysM of CVNH lectin from *Magnaporthe oryzae* in cyan (2L9Y)], a plant protein [LysM1 from AtCERK1 of *A. thaliana* in grey (4EBY)] and a human hypothetical protein [SB145 (2DJP)]. The major difference in Int LysM compared with the others is the α-helical region between the C-terminal conserved α-helix and the C-terminal β-strand. The figures were prepared using VMD (Humphrey *et al.*, 1996, http://www.ks.uiuc.edu/Research/vmd/)

ecule. The 240 kDa peak seen under non-reducing conditions would therefore correspond to a tetrameric molecule, i.e. a dimer of dimers stabilised by disulphides. Crosslinking experiments with BS_3 confirmed this: when DTT was added to the BS_3-treated sample, bands corresponding to the monomer, dimer, trimer and tetramer were observed (Fig. 8D). We are not sure whether the tetramer represents the physiological quaternary structure of YrInv, as we were unable to express the full-length protein in *E. coli* (data not shown). Though this tetramer could be an artefact of the purification procedure, it seems more likely that the tetrameric form is the physiologically relevant one, as the tetramer of YrInvPeri-MBP forms quantitatively in the presence of oxygen, and disulphides would be formed in the periplasm. Interestingly, in addition to dimerising, Inv from *Y. pseudotuberculosis* also forms tetramers (Dersch and Isberg, 1999), further pointing towards the interpretation of the tetramer being the physiologically relevant form of YrInv.

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

Discussion

PGN binding by the Intimin LysM

Here, we provide the first experimental evidence to show that the Int periplasmic region has affinity for PGN. In addition, we present the NMR structure of the Int LysM, giving the first structural information on an inverse autotransporter periplasmic domain. Using binding assays with purified PGN sacculi, we were able to experimentally show, for the first time, that the Int periplasmic domain mediates binding to PGN, whereas the periplasmic domain from Inv bound only with low affinity. The binding of IntPeri to PGN is most probably mediated by the LysM, a well-known PGN-binding motif, rather than the spacer sequence, though both regions are present in the constructs that bound to PGN. In our *in vivo* experiments, InvPeri-MBP was efficiently extracted from the periplasm of *E. coli* at pH 5.5, whereas IntPeri-MBP was retained. We were unable to determine *in vivo* binding in *Y. enterocolitica*, but all our *in vitro* data point to InvPeri-MBP binding to YePGN at low levels, similar to EcPGN. Therefore, it is highly unlikely that InvPeri-MBP would behave differently in *Y. enterocolitica*. We were able to determine an apparent affinity of 0.8 μM for the Int-PGN interaction, whereas for Inv, the affinity was almost an order of magnitude lower. These data all point to IntPeri binding strongly to PGN, whereas InvPeri binds only at low levels, which we contend represent unspecific background binding. In addition, we

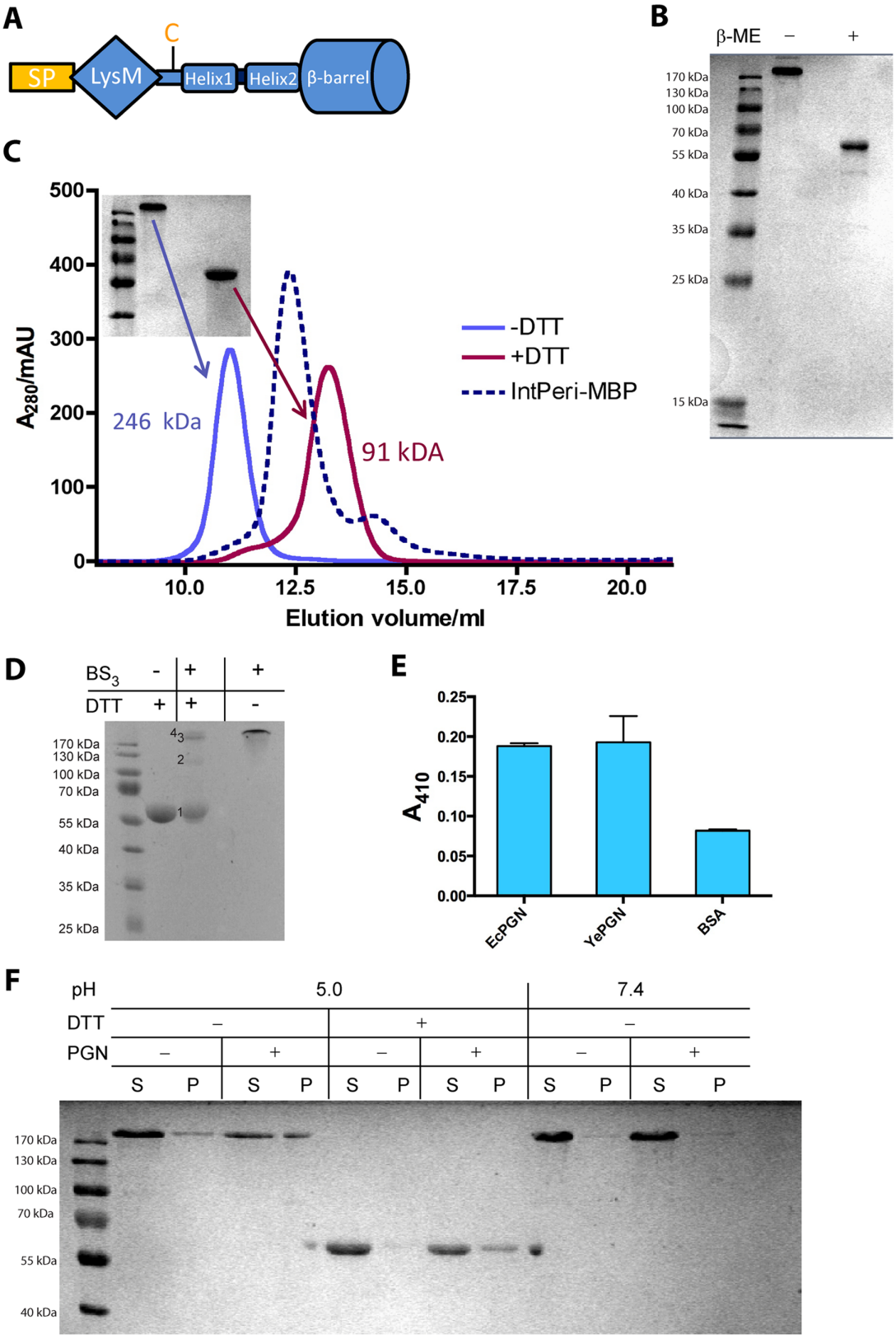

A
C
SP
LysM
Helix1
Helix2
β-barrel
B
β-ME
–
+
170 kDa
130 kDa
100 kDa
70 kDa
55 kDa
40 kDa
35 kDa
25 kDa
15 kDa
C
500
400
300
200
100
0
A280/mAU
246 kDa
91 kDA
-DTT
+DTT
IntPeri-MBP
10.0
12.5
15.0
17.5
20.0
Elution volume/ml
D
BS3
DTT
-
+
+
+
+
-
170 kDa
130 kDa
100 kDa
70 kDa
55 kDa
40 kDa
35 kDa
25 kDa
4
3
2
1
E
A410
0.25
0.20
0.15
0.10
0.05
0.00
EcPGN
YePGN
BSA
F
pH
5.0
7.4
DTT
PGN
S
P
170 kDa
130 kDa
100 kDa
70 kDa
55 kDa
40 kDa

Fig. 8. The periplasmic domain of YrInv forms multimers and binds peptidoglycan.
A. Schematic of YrInvPeri periplasmic domain and barrel. The position of the single cysteine in the spacer sequence is indicated.
B. Effect of reducing agent on YrInvPeri-MBP. β-ME = β-mercaptoethanol.
C. SEC of YrInvPeri-MBP at pH 7.4 with (magenta) and without DTT (light blue). The inset shows an SDS-PAGE gel of the peaks (no additional reducing agent was added). The expected size for the monomer is 58 kDa. Apparent molecular weights are indicated. The SEC curve for IntPeri-MBP from Fig. 2B (dashed blue line) is shown for comparison.
D. Crosslinking of YrInvPeri-MBP. The protein was crosslinked using the amine crosslinker BS_3. After crosslinking, half the sample was treated with DTT. Bands corresponding to the monomer (expected size 58 kDa), dimer (116 kDa), trimer (174 kDa) and tetramer (232 kDa) are indicated in the figure by the numbers 1, 2, 3 and 4, respectively.
E. Binding of YrInvPeri-MBP to EcPGN and YePGN by SPBA. BSA served as the background control. The bars show the mean and error bars denote the SEM from three replicate wells.
F. PGN binding of YrInvPeri-MBP. Pull-downs with or without PGN sacculi were performed at pH 5.0 or pH 7.4. At pH 5.0, pull-down assays were performed either with or without the addition of 10 mM DTT. The small amount of monomeric YrInv seen at the edges of lanes 5 and 10 can be explained by diffusion of the DTT from the neighbouring wells. S = supernatant, P = pellet.

could show that the LysM-containing periplasmic domain from YrInv bound to both EcPGN and YePGN (Fig. 8). Our conclusion is therefore that LysM-containing inverse autotransporter periplasmic domains bind to PGN and those lacking a LysM probably do not, though we cannot rule out that some of the latter type of inverse autotransporters might bind to PGN by another mechanism.

The dimerisation of the LysM may have something to do with the specific binding: it could either position the two PGN binding sites of the dimer in such a way that the protein can optimally interact with the specifically crosslinked strands, or dimerisation increases the affinity of the protein for PGN. In contrast to inverse autotransporters, many PGN-binding proteins contain several LysMs in tandem (Buist *et al*., 2008). Dimerisation may thus be a mechanism to increase the number of PGN binding units and thereby the strength of the interaction.

Most LysM-containing proteins that have been studied bind to the carbohydrate backbone of PGN, and many also bind to the N-acetylglucosamine backbone of chitin. However, we did not detect any binding to chitin beads. Furthermore, although we observed strong binding by IntPeri-MBP to purified sacculi in pull-down assays, the binding saturates at a rather low concentration of protein in relation to PGN (Fig. 4). The moderately high apparent affinity of IntPeri for PGN (0.8 μM) suggests that there are only a limited number of binding sites available for Int. IntPeri-MBP does not bind to chitin, which is often used as a PGN analogue in binding assays. This suggests to us that the PGN disaccharide backbone alone is not the target of the Int LysM domain, though of chitin, of course, does not contain the N-acetylmuramic acid moiety of PGN. However, it seems more likely that some particular muropeptide or combination of crosslinked PGN strands is required for binding, rather than the carbohydrate backbone of PGN. Using the mutant strain D456, we could show that IntPeri-MBP bound with slightly higher affinity to PGN enriched with pentapeptides. In wild-type *E. coli*, PBP5 removes the terminal D-Ala residue from the pentapeptides, the major species thus being the tetrapeptide. In the D456 mutant, the tetrapeptidic muropeptides are still the dominant species, and the pentapeptides are only somewhat enriched (Fig. S5). Thus, it is possible that the pentapeptide or its crosslinked form could be the ligand for the Int LysM. However, based on our current data, we cannot conclusively state that the pentapeptidic muropeptides bind to the Int LysM. We are now pursuing further characterisation of the binding determinants for the Int periplasmic domain in PGN.

The dimerisation interface(s)

We demonstrate here that the Int periplasmic domain is a dimerisation interface for Int. This is in contrast to a previous report implicating the β-barrel domain as the dimerisation interface (Touzé *et al*., 2004). However, Fairman *et al*. have shown that the β-barrel alone does not dimerise (Fairman *et al*., 2012), which suggests that the dimerisation is mediated by the regions immediately upstream or downstream of the β-barrel domain. We tested whether the D00 Ig-like domain forms dimers, but our SEC and crosslinking experiments show that it does not. Thus, the dimerisation observed by Touzé *et al*. must be due to the 129 short stretch preceding the β-barrel domain (residues 189–209). This corresponds to the C-terminal helix in the periplasmic domain (Helix2). However, this α-helix failed to dimerise when fused to MBP. A likely explanation for this discrepancy is that Helix2 is intimately connected to the β-barrel domain, and therefore protease resistant, as shown by Touzé *et al*. (2004). When fused to MBP, these connections are lost and Helix2 remains either unfolded or misfolded, and thus cannot mediate dimerisation. Consistent with this, we could observe faint dimer bands for both IntPeriHelix1-MBP and IntPeriHelix2-MBP in crosslinking experiments, suggesting that this region might be able to dimerise, but only weakly and transiently, such that the dimers are not observed in SEC.

We also observed clear dimerisation by the IntPeriN region alone, a finding that has not been reported before. Both the full periplasmic domain and the N-terminal region

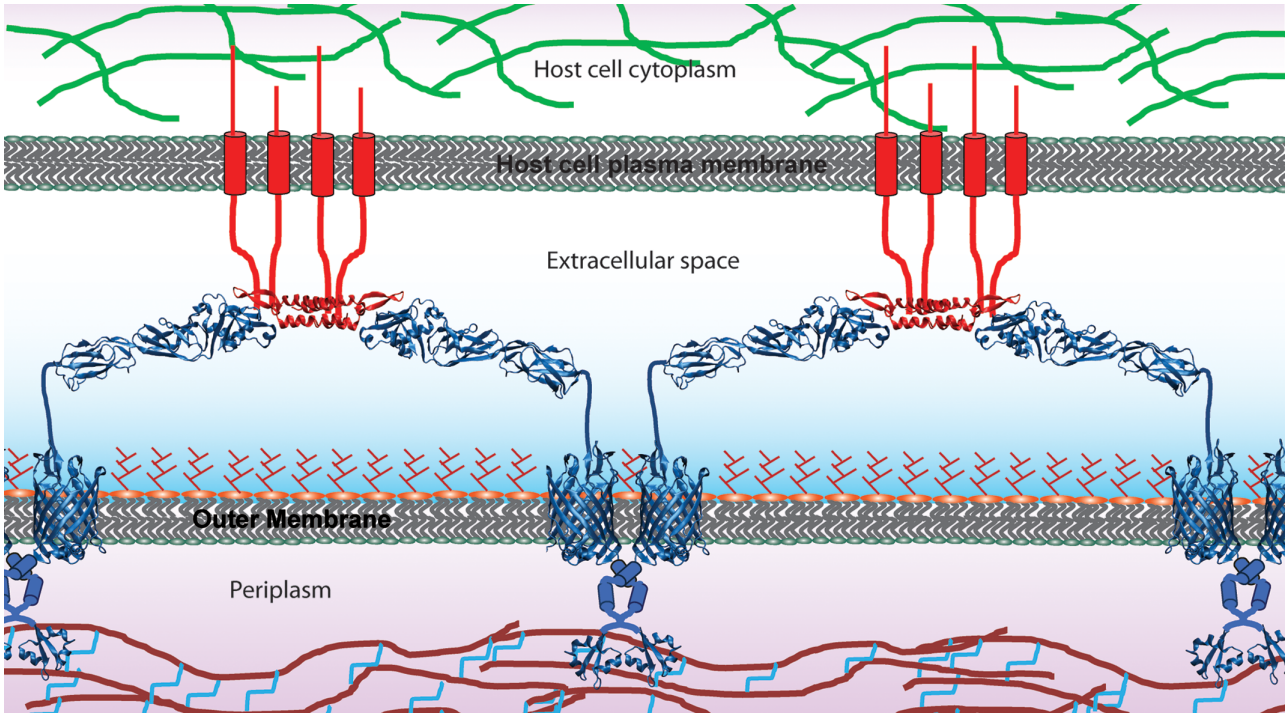


Fig. 9. Model for Intimin dimerisation and Tir clustering.
Schematic of Int (in blue) bound to Tir (red). Dimerisation of Int through the periplasmic domain in the periplasmic domain leads to a reticular array which promotes Tir clustering and pedestal formation. Dimerisation takes place at two sites within the periplasmic domain: at the C-terminal α-helix and in the spacer sequence between the LysM and the N-terminal α-helix. Where structural information is available, cartoon structures are depicted. Regions of unknown structure are depicted as connecting lines, or cylinders in the case of α-helices. The actin cortex of the host cell is in green, the peptidoglycan backbone in brown and peptide crosslinks in light blue. PDB IDs of the structures used for the figure are 1F02, 4E1S and 2MPW. The structural figures were prepared using VMD (Humphrey *et al.*, 1996, http://www.ks.uiuc.edu/Research/vmd/)

containing the LysM and spacer mediated dimerisation in SEC experiments when fused to MBP. A notable feature of these experiments was that the dimeric form was in equilibrium with the monomeric form. This is in contrast to the findings of Touzé *et al.*, who only observed obligate dimerisation of Int.

Though our SEC experiments show clear evidence for IntPeriN mediating dimerisation, unfortunately we did not observe dimers when solving the solution structure. This is most likely due to the unfolding of the C-terminal spacer element upstream of the LysM. This region may require a C-terminal anchor (Helix1 under native condition, or MBP in our constructs) to fold stably. The instability of this region outside its native context could explain why we saw an equilibrium between the dimeric and monomeric states in SEC. We thus conclude that it is not the LysM itself that is the dimerisation interface, but the upstream spacer sequence. This conclusion is supported by the fact that YrInvPeri forms disulphide-bonded tetramers. The single cysteine in YrInvPeri is located in the spacer sequence, suggesting that this region is in fact the dimerisation (or oligomerisation) interface. That the periplasmic domain of YrInv also mediates oligomerisation suggests that oligomer formation is a general feature of LysM-containing inverse autotransporters. In the case of YrInv, the physiological relevance of the disulphide-bonded tetramer remains to be determined. We therefore submit that the Int periplasmic domain contains two dimerisation interfaces, the spacer sequence and Helix2, though the latter only mediates dimer formation when fused to the β-barrel.

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

Biological implications

What might be the function of PGN binding in a subset of inverse autotransporters? One possibility might be an

involvement in the autotransport process itself, direct or indirect. Another might be to anchor the protein to specific sites in PGN and thus prevent lateral diffusion within the outer membrane. This could be useful for A/E pathogens, such as EPEC, for pedestal formation and in maintaining intimate attachment to host cells. Another interesting facet of the Int-PGN interaction is the dependence on low pH, which appears to be a conserved feature based on our results with YrInv. The pH dependence of inverse autotransporter LysM binding suggests that charged residues are involved in binding. IntPeri has a calculated isoelectric point (pI) of 8.7, and YrInvPeri has a pI of 6.1. Thus, at pH 5.0, both proteins carry a net positive charge. A similar result was obtained for AcmD, an autolysin from *Lactococcus lactis* containing three LysMs, where binding to the cell wall was only detected below the pI of the protein (Visweswaran *et al*., 2013). Interestingly, Int is upregulated under acid stress (House *et al*., 2009). As EPEC and other A/E pathogens must travel through the stomach to reach the intestine, and the pH of the periplasm closely follows the pH of the extracellular medium (Wilks and Slonczewski, 2007), the periplasm will experience a significant drop in pH during the pathogen's journey through the stomach and proximal small intestine. The binding to PGN may help in stabilising the cell envelope and aid in acid resistance during transit through stomach and duodenum; however, our attempts to probe the acid resistance of Int-expressing bacteria using survival assays did not yield any conclusive results compared with control cells (data not shown).

The function of PGN binding and dimerisation remain unclear. The model proposed by Touzé *et al*. (2004) suggests that Int dimerisation leads to more efficient Tir clustering on the host cell membrane and the initiation of downstream effects. Interestingly, a recent study showed that an Inv-Int fusion that binds to Tir failed to form actin pedestals when expressed in *Citrobacter rodentium*, and that this strain was defective in colonising a mouse model (Mallick *et al*., 2012). As this fusion lacks the periplasmic dimerisation interface(s) of Int, it is tempting to speculate that the defects are at least partially due to the protein not forming a functional dimer. Dimerisation seems to be important for inverse autotransporter function, as it is widely conserved within the protein family. In contrast to LysM-containing inverse autotransporters, Inv from *Y. pseudotuberculosis* forms dimers and tetramers, but this is not mediated by the periplasmic domain (Dersch and Isberg, 1999). Rather, the multimerisation interface is in the D2 Ig-like domain in the passenger region, which is missing in the monomeric *Y. enterocolitica* orthologue (Dersch and Isberg, 2000). The *Y. pseudotuberculosis* protein is the more potent invasin and promotes a stronger cellular response compared with *Y. enterocolitica* Inv, presumably due to receptor clustering (Dersch and Isberg, 2000). Dimerisation of inverse autotransporters can thus occur through at least two distinct mechanisms: through dimerisation of the periplasmic domain or by self-association of domains in the passenger region.

Experimental procedures

Bioinformatic analyses

Bioinformatics was performed largely using programmes in the MPI Bioinformatics Toolkit (Biegert *et al*., 2006) (http://toolkit.tuebingen.mpg.de/). To identify periplasmic domains of inverse autotransporters, we performed a PSI-BLAST (Altschul and Koonin, 1998) search with five iterations against the non-redundant database using either the Int or Inv periplasmic domain and β-barrel domain sequence. We included the latter to correctly identify type Ve-secreted proteins, and only hits covering over 80% of the query (so as to contain sequence from both the β-barrel and periplasmic domain) were included. The search results were pooled and duplicate hits removed. The sequences were then checked manually and any clearly false positive results (i.e. those lacking the β-barrel domain) were removed. The remaining sequences (many of which appear to be misannotated in the database) were aligned using Clustal Omega (Thompson *et al*., 1994) and Kalign (Lassmann and Sonnhammer, 2005), and the β-barrel domains removed based on the alignment. To identify signal peptide cleavage sites, we submitted the sequences to SignalP 4.0 (Petersen *et al*., 2011) or Phobius (Käll *et al*., 2004), and the sequence N-terminal to the consensus cleavage site was removed. The remaining sequences were then checked for the presence of a LysM motif using HHPred (Söding *et al*., 2005). For clustering, we used CLANS (Frickey and Lupas, 2004) with default parameter values.

Cloning

We amplified the regions of interest from genomic DNA of *E. coli* O127:H6 strain 2348/69 (for Int), *Y. enterocolitica* O:8 strain 8081 (for Inv) or *Y. ruckeri* strain CECT 4319 (for YrInv) using polymerase chain reaction (PCR) with Phusion polymerase (Thermo Scientific). Primers were constructed to include N- and C-terminal BsaI sites for cloning into vectors of the pASK-IBA series (IBA GmbH). For the periplasmic domains, primers were designed to introduce an N-terminal hexahistidine tag for protein purification. Primer sequences are available on request. Constructs used in this study are summarised in Table 1. For producing MBP fusions, we introduced in-frame BamHI and NheI restriction sites after the 3′ end of the insert by PCR-based site-directed mutagenesis (Byrappa *et al*., 1995). The MBP coding sequence was amplified from the plasmid pMal-c2 (New England Biolabs) with primers that

introduced BamHI and NheI sites to the 5′ and 3′-ends respectively. This product was then cloned into the corresponding sites in the modified pIBA-ASK plasmids to produce the final fusion construct. For fusion controls, we amplified MBP with primers introducing BsaI sites and cloned the PCR product into pASK-IBA33 to include a C-terminal hexahistidine tag for efficient purification. To purify IntPeriN alone, we amplified the IntPeriN-MBP insert from pIBA3-IntPeriN-MBP with the forward primer lacking the His tag sequence. This product was then cloned into pASK-IBA33 to introduce a C-terminal His tag on the MBP moiety, and a TEV site was introduced between the LysM and MBP by site-directed mutagenesis. Restriction enzymes were from New England Biolabs, and T4 DNA ligase from Fermentas.

Bacterial strains and growth media

For cloning, all ligation reactions were transformed into chemically competent *E. coli* TOP10 (Invitrogen). For protein overproduction in the cytoplasm, we used the expression strain BL21Gold(DE3) (Novagen). For *in vivo* expression with modified PGN, we used the *E. coli* strain D456 (Edwards and Donachie, 1993). For expression in *Yersinia enterocolitica*, we used the pYV plasmid-cured O:3 strain 6471/67-c (Skurnik, 1984). Bacteria were usually grown in lysogeny broth medium (LB) (Bertani, 1951) supplemented with ampicillin (100 μg ml^{-1}) or chloramphenicol (25 μg ml^{-1}). For protein production, we used the buffered, rich medium ZYP (Studier, 2005) supplemented with ampicillin as above.

Protein production and purification

For production of MBP and MBP fusions, an overnight culture of BL21Gold transformed with the required plasmid (based on either pASK-IBA3 or pASK-IBA33 for production in the cytoplasm) was diluted 1:100 in 1 l ZYP medium and grown to mid-log phase (OD_{600} ~ 0.5) at 37°C, at which time protein production was induced with anhydrotetracycline (200 ng ml^{-1}). The cultures were then grown for another 3 h at 37°C, after which the cells were harvested by centrifuging (10 min, 5000 × *g*) and then resuspended in Ni binding buffer (20 mM sodium phosphate, 20 mM imidazole, 500 mM sodium chloride, pH 7.4). We added $MgCl_2$ and $MnCl_2$ to 1 mM, lysozyme to 0.1 mg ml^{-1}, EDTA-free complete protease inhibitor cocktail (Roche) and a pinch of DNase I (Applichem), and proceeded to rupture the cells using a French pressure cell with two passes at 18 000 p.s.i. Cellular debris were pelleted for 1 h at 100 000 × *g*; the supernatant was passed through a 0.22 μm filter and applied to a nickel iminodiacetic acid column (PrepEase; USB). Bound proteins were eluted with a step gradient of imidazole (250 mM) in Ni binding buffer. We pooled the fractions containing the periplasmic domain-MBP fusions, added protease inhibitor cocktail and applied these to an amylose column (MBPTrap; GE Healthcare) equilibrated with 20 mM Tris, 200 mM NaCl, 1 mM EDTA at pH 7.4. Bound protein was eluted using the same buffer supplemented with 10 mM maltose. Proteins were further purified by size exclusion chromatography (Superdex200™ 26/60 column; GE Healthcare) in Tris-buffered saline (TBS; 20 mM Tris pH 7.4 with 150 mM NaCl and 0.02% NaN_3).

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

NMR structure determination

All spectra were recorded at 298 K on Bruker AVIII-600 and AVIII-800 spectrometers. Backbone sequential assignments were completed using a strategy based on a 3D-HN(CA)NNH experiment (Weisemann *et al.*, 1993). Aliphatic sidechain assignments were completed with standard TOCSY-based experiments, while aromatic assignments were made by linking aromatic spin systems to the respective $C^{\beta}H_2$ protons in a 2D-NOESY spectrum. Stereospecific assignments and the resulting χ1 rotamer assignments were determined from a combination of HNHB and HA[HBHN](CACO)NH experiments (Löhr *et al.*, 1999).

Distance data were derived from a set of five 3D-NOESY spectra, including the heteronuclear edited NNH-, CCH- and CNH-NOESY spectra (Diercks *et al.*, 1999) in addition to conventional ^{15}N- and ^{13}C-HSQC-NOESY spectra. A ^{12}C-filtered 2D-NOESY spectrum was recorded for the observation of contacts to aromatic groups. Backbone dihedral angle restraints were derived using the TALOS+ server (Shen *et al.*, 2009). Generic backbone dihedral restraints designed to restrict residues to allowed regions

of the Ramachandran map and well-populated sidechain rotamers were applied for unstructured residues. Hydrogen bond restraints were applied as pseudo-covalent bonds, as outlined in Truffault *et al*. (2001). Refinement was carried out by comparing experimental and back-calculated NOESY spectra using in-house software. Strips were back-calculated for the amide protons of all ordered atoms, plus selected sidechain groups. These were compared with the experimental spectra to confirm backbone and sidechain dihedral angles and to extract additional distance restraints.

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

Analytical size exclusion chromatography

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

Crosslinking

For *in vitro* crosslinking, the buffer of purified proteins was exchanged to 10 mM HEPES (4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid) at pH 7.4 by diluting the proteins to 2 ml with the HEPES buffer and then concentrating the proteins using a Centricon 10 kDa molecular weight cut-off concentrater to a small volume (< 500 µl), then diluting a second time with HEPES and re-concentrating. The crosslinking was performed in a total volume of 30 µl of 10 mM HEPES pH 7.4, with the buffer-exchanged proteins diluted to 1 mg ml^{-1}. For the crosslinking reactions, the amine crosslinker BS_3 (Thermo Scientific) was first dissolved to 50 mM in water, and 0.5 µl was added to the protein solution. The reaction was allowed to proceed for 5 min at room temperature (RT), after which the reaction was stopped by the addition of 3 µl 1 M Tris pH 7.5. After 15 min incubation at RT, 10 µl of 4 × SDS-PAGE sample buffer was added, the samples were heated for 5 min at 95°C and then subjected to SDS-PAGE.

Peptidoglycan and chitin binding

Peptidoglycan sacculi from *E. coli* Nissle, *E. coli* D456 and *Y. enterocolitica* O:8 8081 were purified according to the method of Glauner, with some modifications (Glauner, 1988). The lyophilised sacculi were resuspended in ultrapure water (with sodium azide added to 0.02% w/v) to a concentration of 10 mg ml^{-1}. For pull-down assays, 2 µl of this suspension was mixed with 5 µg of protein in a total volume of 30 µl and incubated for 15 min at RT. The buffer used depended on the pH: we used TBS (pH 8.0 or pH 7.4), MOPS-buffered saline (pH 7.0), MES-buffered saline (pH 6.0), ABS (pH 5.0 and pH 4.0) or glycine-buffered saline (pH 3.0). The sacculi were then pelleted by centrifuging 30 min at ~ 20 000 × *g* (full speed using a tabletop centrifuge). The supernatant was carefully removed and the pellet was washed once with the corresponding buffer. After a second centrifugation step, the supernatant was removed and the pellet was resuspended in 30 µl TBS (pH 7.4). For analysis, we added 10 µl of 4× non-reducing sample buffer to the first supernatant fraction and the pellet fraction, boiled the samples for 5 min and then loaded 10 µl onto a 12% polyacrylamide gel for SDS-PAGE.

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

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

Chitin binding was assayed using chitin beads (New England Biolabs). We prepared columns using 2 ml of the bead slurry and equilibrated with ABS (pH 5.0). We added 1 mg of protein diluted in 1 ml of ABS and allowed the protein to enter the column by gravity flow. After washing

with 10 ml ABS, bound protein was eluted with 10 ml TBS (pH 7.4), followed by a second elution step with 0.3 M NaOH. A sample was taken from each step for SDS-PAGE analysis.

Solid-phase binding assay

Wells of a polystyrene microtitre plate (Sarstedt microtest plate) were coated with 100 µl of a PGN suspension (100 µg ml^{-1}) in coating buffer (100 mM sodium carbonate pH 9.6) overnight at 4°C (Petrović *et al.*, 2012). The following day, the wells were emptied and blocked with 150 µl of 5% bovine serum albumin (BSA) in ABS pH 5.0 for 1 h at RT. The wells were then washed twice with washing buffer (ABS with 0.1% BSA and 0.05% Tween20). Proteins were diluted to indicated concentrations in blocking buffer and 100 µl of the dilutions were added to the wells (in triplicate for each sample). As controls, we included wells coated only with BSA. After incubating for 1 h at RT, the wells were washed three times as above, and the primary antibody, a rabbit anti-MBP (anti-MalE) antiserum, was diluted 1:5000 in blocking buffer and 100 µl was added to the wells. After 1 h, the wells were washed three times as above, and 100 µl of goat-anti-rabbit IgG-horseradish peroxidase (HRP) conjugate (from Santa Cruz Biotechnology) solution (1:10 000 in blocking buffer) was added. After 1 h, the wells were washed four times as above. We then added 150 µl of HRP substrate solution (from Pierce ABTS tablets, prepared according to the manufacturer's instructions). The reaction was allowed to proceed for 1 h at RT. The reaction was stopped by the addition of 100 µl of 1% SDS. Absorbance at 410 nm was measured with a Biotek Synergy Mx plate reader.

The anti-MalE antiserum was obtained by immunising a rabbit with MBP (MalE). The vaccine was prepared by emulsifying 1 mg MalE protein (denaturated with 1% SDS) with Freund's adjuvant in the cold, and injected subcutaneously and intramuscularly (hind legs); after 4 weeks the rabbit obtained a booster injection and blood was collected with 5–6 days intervals.

In vivo *PGN binding*

To test for PGN binding in whole cells, we transformed the plasmids pIBA2C-IntPeri-MBP, pIBA2C-InvPeri-MBP or pASK-IBA2C into *E. coli* BL21Gold(DE3), *E. coli* D456 or *Y. enterocolitica* 6471/76-c cells. The cells were then grown at 37°C in LB + chloramphenicol (25 µg ml^{-1}) until mid-log, at which time the temperature was changed to 27°C. After 30 min, recombinant protein production was induced with AHTC (50 ng ml^{-1}). After 1 h at 27°C, we added buffer, either Tris pH 7.4, MES pH 6.0 or sodium acetate pH 5.5, to 100 mM. After a further hour of growth, the turbidity of the cultures (OD_{600}) was measured and an amount of cells corresponding to 10 ml at an OD_{600} value of 1.0 was harvested by centrifugation (10 min 3500 × *g*). To examine the soluble periplasmic fraction, we used a modified osmotic shock protocol. The pelleted cells were resuspended in 400 µl 5 mM $CaCl_2$ with 5 mM buffer (Tris pH 7.4, MES pH 6.0 or sodium acetate pH 5.5) and incubated on ice for 10 min. This step improves the yield of the periplasmic extraction (Chen *et al.*, 2004). The cells were pelleted (5 min at 8000 × *g*) and then resuspended in 400 µl ice-cold osmotic shock solution (33 mM buffer, 20% sucrose, 5 mM EDTA) buffered at either pH 7.4, 6.0 or pH 5.5, as above. After 10 min incubation at 8°C with shaking, the cells were centrifuged as above and then resuspended in 400 µl 5 mM buffer (either Tris pH 7.4, MES pH 6.0 or sodium acetate pH 5.5). The cells were incubated for 10 min with shaking at 8°C and then centrifuged as above. A sample (120 µl) was taken from the supernatant, and 40 µl 4 × SDS-PAGE sample buffer was added. This represented the periplasmic fraction. For the whole-cell sample, we pelleted an amount of cells from the induced culture corresponding to 120 µl at $OD_{600} = 25$, resuspended these in 120 µl PBS and then added 40 µl sample buffer.

To probe for the recombinant proteins, we performed a Western blot. The proteins were separated in a 10% polyacrylamide gel and then transferred to a nitrocellulose (Protran BA 85, GE Healthcare) or polyvinylidene fluoride (PVDF) membrane (Thermo Scientific) using a semi-dry apparatus. The membrane was blocked for 30 min at RT or overnight at 4°C, with 5% fat-free milk powder in TBS. The primary antibody was the anti-MBP antiserum described above, diluted in blocking buffer 1:5000. After 1 h of blocking, the membrane was washed twice 10 min with TBS + 0.05% Tween20 (TBS-T), and then the secondary antibody [goat anti-rabbit-alkaline phosphatase (AP), Jackson Immunoresearch or goat anti-rabbit-HRP, Santa Cruz Biotechnology] was added at a dilution of 1:10 000 in blocking buffer. The membrane was washed twice with TBS-T as above, and a final time in AP buffer (100 mM Tris pH 9.5, 100 mM NaCl, 5 mM $MgCl_2$). Chromogenic detection was performed using the AP substrate nitro blue tetrazolium/5-bromo-4-chloro-3-indolyl phosphate (NBT/BCIP), diluted to 33 ng ml^{-1} and 17 ng ml^{-1}, respectively, in AP buffer. The reaction was stopped with deionised water once colour had developed. Alternatively, detection was performed using enhanced chemiluminescence (Pierce ECL substrate) with a CCD camera (Kodak Image Station 4000R).

Accession numbers

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

Author contributions

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 experiments and analysed the data. D.K., U.B., H.S. and F.G. provided essential materials. J.C.L., P.O. and D.L. wrote the paper.

Acknowledgements

We would like to thank M.Sc. Iwan Grin (MPI for Developmental Biology, Tübingen) for assistance with parsing the sequences for bioinformatic analyses. We are also grateful to Dr Jörg Rau (Chemisches und Veterinäruntersuchungsamt, Stuttgart) for providing the *Y. ruckeri* DNA and to Professor Mikael Skurnik (Haartman Institute, University of Helsinki) for furnishing us with the *Y. enterocolitica* strain 6471/76-c. We thank Professor Andrei Lupas (Protein Evolution, MPI for Developmental Biology, Tübingen) for continued support. This study was funded by the Sonderforschungsbereich 766 of the DFG (German Science Foundation) to D.L., I.B.A., U.B. and F.G., UKT fortüne grant F1433253 to P.O., and a FEMS Advanced Fellowship to J.C.L.

References

Altschul, S.F., and Koonin, E.V. (1998) Iterated profile searches with PSI-BLAST – a tool for discovery in protein databases. *Trends Biochem Sci* **23:** 444–447.

Bateman, A., and Bycroft, M. (2000) The structure of a LysM domain from *E. coli* membrane-bound lytic murein transglycosylase D (MltD). *J Mol Biol* **299:** 1113–1119.

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

Biegert, A., Mayer, C., Remmert, M., Söding, J., and Lupas, A.N. (2006) The MPI Bioinformatics Toolkit for protein sequence analysis. *Nucleic Acids Res* **34:** W335–W339.

Bielnicki, J., Devedjiev, Y., Derewenda, U., Dauter, Z., Joachimiak, A., and Derewenda, Z.S. (2006) *B. subtilis* ykuD protein at 2.0 Å resolution: insights into the structure and function of a novel, ubiquitous family of bacterial enzymes. *Proteins* **62:** 144–151.

Buist, G., Steen, A., Kok, J., and Kuipers, O. (2008) LysM, a widely distributed protein motif for binding to (peptido)glycans. *Mol Microbiol* **68:** 838–847.

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

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

Dersch, P., and Isberg, R.R. (1999) A region of the *Yersinia pseudotuberculosis* invasin protein enhances integrin-mediated uptake into mammalian cells and promotes self-association. *EMBO J* **18:** 1199–1213.

Dersch, P., and Isberg, R.R. (2000) An immunoglobulin superfamily-like domain unique to the *Yersinia pseudotuberculosis* invasin protein is required for stimulation of bacterial uptake *via* integrin receptors. *Infect Immun* **68:** 2930–2938.

Diercks, T., Coles, M., and Kessler, H. (1999) An efficient strategy for assignment of cross-peaks in 3D heteronuclear NOESY experiments. *J Biomol NMR* **15:** 177–180.

Edwards, D., and Donachie, W.D. (1993) Construction of a triple deletion of Penicillin-binding proteins 4, 5 and 6 in *Escherichia coli*. In *Bacterial Growth and Lysis*. de Pedro, M., Höltje, J.-V., and Löffelhardt, W. (eds). New York and London: Plenum Press, pp. 369–374.

Fairman, J.W., Dautin, N., Wojtowicz, D., Liu, W., Noinaj, N., Barnard, T.J., *et al*. (2012) Crystal structures of the outer membrane domain of intimin and invasin from enterohemorrhagic *E. coli* and enteropathogenic *Y. pseudotuberculosis*. *Structure* **20:** 1233–1243.

Frickey, T., and Lupas, A. (2004) CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* **20:** 3702–3704.

Glauner, B. (1988) Separation and quantification of muropeptides with high-performance liquid chromatography. *Anal Biochem* **172:** 451–464.

Hamburger, Z.A., Brown, M.S., Isberg, R.R., and Bjorkman, P.J. (1999) Crystal structure of invasin: a bacterial integrin-binding protein. *Science* **286:** 291–295.

Hochuli, M., Szyperski, T., and Wüthrich, K. (2000) Deuterium isotope effects on the central carbon metabolism of *Escherichia coli* cells grown on a D_2O-containing minimal medium. *J Biomol NMR* **17:** 33–42.

House, B., Kus, J.V., Prayitno, N., Mair, R., Que, L., Chingcuanco, F., *et al*. (2009) Acid-stress-induced changes in enterohaemorrhagic *Escherichia coli* O157:H7 virulence. *Microbiology* **155:** 2907–2918.

Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J Mol GraphH* **14:** 33–38, 27–28.

Käll, L., Krogh, A., and Sonnhammer, E.L. (2004) A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* **338:** 1027–1036.

Kelly, G., Prasannan, S., Daniell, S., Fleming, K., Frankel, G., Dougan, G., *et al*. (1999) Structure of the cell-adhesion fragment of intimin from enteropathogenic *Escherichia coli*. *Nat Struct Biol* **6:** 313–318.

Kraus, W., and Höltje, J.V. (1987) Two distinct transpeptidation reactions during murein synthesis in *Escherichia coli*. *J Bacteriol* **169:** 3099–3103.

Lassmann, T., and Sonnhammer, E.L.L. (2005) Kalign – an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics* **6:** 298.

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

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

Leong, J.M., Fournier, R.S., and Isberg, R.R. (1990) Identification of the integrin binding domain of the *Yersinia pseudotuberculosis* invasin protein. *EMBO J* **9:** 1979–1989.

Löhr, F., Schmidt, J.M., and Rüterjans, H. (1999) Simultaneous measurement of 3JHN,Hα and 3JHα,Hβ coupling con-

stants in 13C,15N-labeled proteins. *J Am Chem Soc* **121:** 11821–11826.
Luo, Y., Frey, E.A., Pfuetzner, R.A., Creagh, A.L., Knoechel, D.G., Haynes, C.A., *et al*. (2000) Crystal structure of enteropathogenic *Escherichia coli* intimin-receptor complex. *Nature* **405:** 1073–1077.
Mallick, E.M., Brady, M.J., Luperchio, S.A., Vanguri, V.K., Magoun, L., Liu, H., *et al*. (2012) Allele- and Tir-independent functions of intimin in diverse animal infection models. *Front Microbiol* **3:** 11.
Maxwell, K.L., Hassanabad, M.F., Chang, T., Pirani, N., Bona, D., Edwards, A.M., and Davidson, A.R. (2013) Structural and functional studies of gpX of *Escherichia coli* phage P2 reveal a widespread role for LysM domains in the baseplates of contractile-tailed phages. *J Bacteriol* **195:** 5461–5468.
Oberhettinger, P., Schütz, M., Leo, J.C., Heinz, N., Berger, J., Autenrieth, I.B., and Linke, D. (2012) Intimin and invasin export their C-terminus to the bacterial cell surface using an inverse mechanism compared to classical autotransport. *PLoS ONE* **7:** e47069.
Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8:** 785–786.
Petrović, D.M., Leenhouts, K., van Roosmalen, M.L., Kleinjan, F., and Broos, J. (2012) Monitoring lysin motif-ligand interactions *via* tryptophan analog fluorescence spectroscopy. *Anal Biochem* **428:** 111–118.
Pisano, F., Kochut, A., Uliczka, F., Geyer, R., Stolz, T., Thiermann, T., *et al*. (2012) *In vivo*-induced InvA-like autotransporters Ifp and InvC of *Yersinia pseudotuberculosis* promote interactions with intestinal epithelial cells and contribute to virulence. *Infect Immun* **80:** 1050–1064.
Potluri, L., Karczmarek, A., Verheul, J., Piette, A., Wilkin, J., Werth, N., *et al*. (2010) Septal and lateral wall localization of PBP5, the major D,D-carboxypeptidase of *Escherichia coli*, requires substrate recognition and membrane attachment. *Mol Microbiol* **77:** 300–323.
Quintela, J.C., Caparrós, M., and de Pedro, M.A. (1995) Variability of peptidoglycan structural parameters in gram-negative bacteria. *FEMS Microbiol Lett* **125:** 95–100.
Schmidt, M.A. (2010) LEEways: tales of EPEC, ATEC and EHEC. *Cell Microbiol* **12:** 1544–1552.
Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9:** 671–675.
Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44:** 213–223.
Skurnik, M. (1984) Lack of correlation between the presence of plasmids and fimbriae in *Yersinia enterocolitica* and *Yersinia pseudotuberculosis*. *J Appl Bacteriol* **56:** 355–363.
Söding, J., Biegert, A., and Lupas, A.N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33:** W244–W248.
Studier, F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41:** 207–234.
Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22:** 4673–4680.
Touzé, T., Hayward, R.D., Eswaran, J., Leong, J.M., and Koronakis, V. (2004) Self-association of EPEC intimin mediated by the β-barrel-containing anchor domain: a role in clustering of the Tir receptor. *Mol Microbiol* **51:** 73–87.
Tropea, J.E., Cherry, S., and Waugh, D.S. (2009) Expression and purification of soluble His(6)-tagged TEV protease. *Methods Mol Biol* **498:** 297–307.
Truffault, V., Coles, M., Diercks, T., Abelmann, K., Eberhardt, S., Lüttgen, H., *et al*. (2001) The solution structure of the N-terminal domain of riboflavin synthase. *J Mol Biol* **309:** 949–960.
Tsai, J.C., Yen, M., Castillo, R., Leyton, D.L., Henderson, I.R., and Saier, M.H. (2010) The bacterial intimins and invasins: a large and novel family of secreted proteins. *PLoS ONE* **5:** e14403–e14414.
Visweswaran, G.R.R., Steen, A., Leenhouts, K., Szeliga, M., Ruban, B., Hesseling-Meinders, A., *et al*. (2013) AcmD, a homolog of the major autolysin AcmA of *Lactococcus lactis*, binds to the cell wall and contributes to cell separation and autolysis. *PLoS ONE* **8:** e72167.
Weisemann, R., Rüterjans, H., and Bermel, W. (1993) 3D triple-resonance NMR techniques for the sequential assignment of NH and 15N resonances in ^{15}N- and ^{13}C-labelled proteins. *J Biomol NMR* **3:** 113–120.
Wilks, J.C., and Slonczewski, J.L. (2007) pH of the cytoplasm and periplasm of *Escherichia coli*: rapid measurement by green fluorescent protein fluorimetry. *J Bacteriol* **189:** 5601–5607.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site.