Fold and Function of the InlB B-repeat*[S]

Maria Ebbes[‡], Willem M. Bleymüller[‡], Mihaela Cernescu[§], Rolf Nölker[‡], Bernd Brutschy[§], and Hartmut H. Niemann[‡1]

From the [‡]Department of Chemistry, Bielefeld University, Universitätsstrasse 25, 33615 Bielefeld, Germany and the [§]Institute of Physical and Theoretical Chemistry, Goethe-University, Max-von-Laue-Strasse 9, 60438 Frankfurt, Germany

Host cell invasion by the facultative intracellular pathogen *Listeria monocytogenes* requires the invasion protein InlB in many cell types. InlB consists of an N-terminal internalin domain that binds the host cell receptor tyrosine kinase Met and C-terminal GW domains that bind to glycosaminoglycans (GAGs). Met binding and activation is required for host cell invasion, while the interaction between GW domains and GAGs enhances this effect. Soluble InlB elicits the same cellular phenotypes as the natural Met ligand hepatocyte growth factor/ scatter factor (HGF/SF), *e.g.* cell scatter. So far, little is known about the central part of InlB, the B-repeat. Here we present a structural and functional characterization of the InlB B-repeat. The crystal structure reveals a variation of the β-grasp fold that is most similar to small ubiquitin-like modifiers (SUMOs). However, structural similarity also suggests a potential evolutionary relation to bacterial mucin-binding proteins. The B-repeat defines the prototype structure of a hitherto uncharacterized domain present in over a thousand bacterial proteins. Generally, this domain probably acts as a spacer or a receptor-binding domain in extracellular multi-domain proteins. In cellular assays the B-repeat acts synergistically with the internalin domain conferring to it the ability to stimulate cell motility. Thus, the B-repeat probably binds a further host cell receptor and thereby enhances signaling downstream of Met.

Listeria monocytogenes is the causative agent of the rare but severe disease listeriosis, which occasionally kills dozens of people in outbreaks caused by consumption of contaminated food (1, 2). In addition, *Listeria* has become a model system in cellular microbiology because of its facultative intracellular life style (3). To induce its uptake into normally non-phagocytic cells, escape from the phagocytic vacuole, move inside of host cells, or spread from cell to cell, *L. monocytogenes* interferes with many endogenous cellular processes (4). Thus, investigation of host-pathogen interactions has also provided new insights into fundamental cell biology (5).

Uptake of *L. monocytogenes* into a variety of epithelial and endothelial cells requires activation of the receptor tyrosine kinase Met by the invasion protein InlB (6, 7). Normally Met acts as the sole receptor for hepatocyte growth factor/scatter factor (HGF/SF).[2] Met signaling is essential during embryonic development in vertebrates and has, among others, a mitogenic and a motogenic effect (8). Soluble InlB behaves like a growth factor and its effects are very similar to those of HGF/SF (9).

InlB belongs to the larger family of *Listeria* internalin proteins (10). Internalins are either secreted or cell surface-anchored proteins and all share common features in the N-terminal region, while the C terminus is more divergent and often contains different combinations of small domains (70–80 residues in size) like GW, PKD, or MucBP domains. The C terminus also determines whether the protein is covalently or noncovalently attached to the bacterial surface or secreted. The N terminus of the processed protein is characterized by an internalin domain that consists of a central leucine-rich repeat (LRR) region flanked by specialized capping structures (11).

LRR domains are typically involved in ligand-binding (12). The kidney-shaped internalin domain of InlB is necessary and sufficient for Met activation (7, 13) and binds to Met via its concave side (14, 15), a mode of interaction that is typical for the curved LRR proteins. Met activation most likely proceeds through ligand-mediated dimerization of the receptor, whereby the convex side of the InlB LRR mediates the low-affinity dimerization contact (16, 17). Stimulation of cells with the isolated, monomeric internalin domain leads to phosphorylation of Met and downstream signaling molecules like ERK or Akt, but does not elicit cellular phenotypes like scatter or division (13, 15, 16).

C-terminal to the internalin domain, InlB harbors a single B-repeat, and three GW domains (Fig. 1*A*). The GW domains do not interact with Met and, so far, there is no evidence that they can elicit any signaling event or cellular effect on their own (7, 13). However, they act synergistically with the internalin domain, when fused to it (13, 18), lowering the minimal concentration required for Met phosphorylation and conferring to the InlB internalin domain the ability to induce cellular phenotypes like cell scatter (15, 16). The GW domains are responsible for the attachment of InlB to the bacterial surface by non-covalent interaction with lipoteichoic acid (19, 20) and they bind to host cell glycosaminoglycans like heparan sulfate (18, 21). Most likely, they act by promoting the formation of higher-order receptor complexes through clustering (15).

[2] The abbreviations used are: HGF/SF, hepatocyte growth factor/scatter factor; GAG, glycosaminoglycan; LRR, leucine-rich repeat; LILBID, Laser-induced Liquid Bead Ion Desorption; rmsd, root mean square deviation; SUMO, small ubiquitin-like modifiers.

The middle part, the B-repeat, is the least characterized domain of InlB. Its name derives from the early notion that InlA contains two repeat regions. Region A comprises the leucine rich repeats and region B three repeats of 70 amino acids each (22). The presence of a single copy of region B in InlB was noticed only once the sequence of additional internalin family members became available for comparison (23). Shortly after, the term B-repeat was coined (24). Within the different internalins of *L. monocytogenes*, B-repeats are found, in InlB and InlE (1 copy) in InlC2, InlD, InlG, and InlH (2 copies) in InlA (3 copies) and in InlF (4 copies) (10). In the crystal structure of full-length InlB, the B-repeat could not be modeled due to poor electron density (21). No difference with respect to Met activation was found between the internalin domain alone and a construct comprising the internalin domain and the B-repeat, but the latter caused a stronger activation of ERK (25). Consequently, Ghosh and co-workers suggested that the InlB B-repeat may bind to a further, as yet unidentified host cell receptor. Within the last seven years, however, little progress has been made in understanding the function of the InlB B-repeat or of B-repeats from any other internalin.

Here we set out to investigate the structure and function of the InlB B-repeat. We determined the crystal structure of the B-repeat, revealing a well-folded stable domain with an ubiquitin-like fold. Cellular assays with the internalin domain (InlB$_{321}$; Fig. 1*A*) and the internalin domain plus the B-repeat (InlB$_{392}$, Fig. 1*A*) showed that only the latter was able to stimulate cell motility. *In vitro*, the B-repeat alone did not bind to the Met ectodomain. In solution, InlB$_{392}$, like InlB$_{321}$, did not show a propensity to dimerize the Met ectodomain, but only formed 1:1 complexes. These results suggest that the B-repeat contributes to stimulation of cellular phenotypes by binding to a further host cell receptor.

EXPERIMENTAL PROCEDURES

Cloning, Protein Expression, and Purification—The vector coding for InlB$_{392}$ was generated by mutating codon 393 into a stop codon in the previously published expression vector for full-length mature InlB (amino acids 36–630 fused to glutathione *S*-transferase (GST) in the pGEX-6P-1 backbone (15)). The DNA coding for the B-repeat (aa 322–392) and B-repeat+GW (aa 322–630) was amplified by PCR. The inserts were cloned into the NcoI and NotI sites of the pETM30 vector (Gunter Stier, EMBL Heidelberg). All constructs were verified by sequencing. Expression of GST fusion proteins was induced in *E. coli* BL21 CodonPlus-RIL (Invitrogen) with 1 mM IPTG at an A_{600} of 0.6–0.8 in LB medium at 20 °C overnight. For production of Se-Met-derivatized protein, bacteria were grown to an A_{600} of 0.8 in LB medium, pelleted, washed in water, and suspended in the same volume of SeMet minimal medium (26). Protein production was induced as above. After cell lysis in phosphate buffered saline (PBS) with Complete protease inhibitors (Roche) and centrifugation, the soluble supernatant was immobilized on glutathione-Sepharose (GE Healthcare). The elution took place through cleavage of the GST-tag by TEV protease (pETM30) or PreScission protease (pGEX-6P-1). The B-repeat and InlB$_{392}$ were purified further by anion exchange chromatography (SourceQ, GE Healthcare), dialyzed into 10

mM Tris pH 8.0 and 50 mM NaCl and concentrated to 17.3 mg/ml and 5 mg/ml, respectively, using VIVASPIN (Sartorius). Aliquots were frozen at −20 °C. Cation exchange chromatography (SourceS, GE Healthcare) was used for further purification of the B-repeat+GW, which was stored in the SourceS elution buffer (100 mM Hepes, pH 8.0 with more than 300 mM NaCl).

Crystallization—First crystals of the B-repeat were obtained using the PACT suite screen (Qiagen). This initial condition was optimized at 20 °C in hanging or sitting-drops with 2 μl of protein (17.3 mg/ml) + 1 μl of reservoir (0.1 M sodium acetate, pH 5.0, 0.2 M CaCl$_2$, 18% PEG6000). Crystals grew after 4–7 days. SeMet-crystals grew under the same condition. Crystals were cryoprotected in reservoir solution plus 30% PEG400 and flash-frozen in liquid nitrogen. Initial crystals of InlB$_{392}$ were grown from the JCSG Core I suite screen (Qiagen) and were optimized at 20 °C in hanging or sitting-drops with 1.5 μl of protein (5 mg/ml) + 1.5 μl of reservoir (0.1 M MES, pH 5.5, 14% MPD) microseeded from a similar condition (0.1 M MES, pH 6.0, 6.5% PEG6000, and 5 mM ZnCl$_2$). Crystals grew after 2–4 days and were cryoprotected by increasing the MPD concentration to 30%.

Data Collection, Structure Determination, and Refinement—Native data of the B-repeat were collected from a crystal of about 100 × 70 × 50 μm at beamline X11 (DESY Hamburg, Germany; λ = 0.81 Å) using a Mar555 flat panel detector. Further data were collected at 0.95 Å and 1.9 Å at the beamline X12 on a MAR225 CCD detector. A four-wavelength MAD data set was collected from a Se-Met crystal at beamline X12 (DESY Hamburg, Germany). Data of InlB$_{392}$ were collected at the beamline X12 (λ = 0.98 Å). All data were indexed and integrated with XDS and scaled with XSCALE (27). The heavy atom substructure was solved with SHELXD. Density modification and phase extension was carried out with SHELXE run through the interface HKL2MAP (28, 29). The structure of InlB$_{392}$ was solved by molecular replacement using Phaser (30). The model was built manually in COOT (31). All structures were refined with REFMAC5 (32). Difference density for InlB$_{392}$ was interpreted as three Zn^{2+} ions carried over from the seed solution and verified in an anomalous difference map. The coordinate and structure factor files of the InlB B-repeat and InlB$_{392}$ have been deposited with the Protein Data Bank (PDB) under accession codes 2y5p and 2y5q, respectively.

Database Searches, Alignments, and Figures—HHsenser was used to find near and remote homologs of the B-repeat in a non-redundant data base and returned 1825 sequences with a length of 36–119 amino acids (33). From an alignment, in which insertions with respect to the B-repeat were deleted ("masterslave" from HHsenser), we kept sequences with a length of more than 63 amino acids and removed sequences lacking homology in the region of strand β3 and β4. An HMM was generated from the resulting 370 sequences (34), which was turned into the HMM logo (35). Structure figures were prepared with PyMol (36). The SSM server (37) was run with default settings, *i.e.* requiring that at least 70% of the secondary structure are matched in both query and target and the results sorted by q-score. Multiple structure-based sequence align-

ments were generated with MAMMOTH (38). Alignments were visualized with Jalview (39).

Cell Lines and Media—Vero (African green monkey kidney), A549 (human lung carcinoma), and MDCK (Madin Darby canine kidney) cells were grown in Dulbecco's modified Eagle's medium (DMEM, PAA) (low glucose) containing 5% fetal bovine serum. HT29 (human colon adenocarcinoma) cells were grown in DMEM (high glucose) containing 10% fetal bovine serum.

Scatter Assays—HT29 cells were seeded at a density of 5×10^4 cells per well in a 12-well plate, grown for 48 h in DMEM with 10% fetal bovine serum and starved for 24 h in serum-free medium. Then the cells were incubated for 24 h with the ligand in serum-free medium. MDCK clone20 cells were grown in DMEM with 5% fetal bovine serum with an initial density of 3×10^4 cells per well for 24 h. Incubation with ligand was carried out in serum-containing medium for 24 h at 37 °C. Pictures were taken with a Leica DM IRB/Leica DC 300F. All scatter assays were repeated at least three times with three wells per ligand in each experiment and scored blindly by three researchers.

Wound-healing Assay—MDCK cells from ATCC, Vero or A549 cells were cultured in 12-well plates (2 x and 3×10^5 cells per well), grown to confluency and starved for 24 h. After wounding with a 200-μl pipette tip cells were incubated with the different ligands in serum free medium and images were taken at 0, 4, 6, 8, and 24 h with a Leica DM IRB microscope and the Leica DC 300F camera. The width of the wound was measured with ImageJ (40). All experiments were repeated at least three times with three wounds per ligand in each experiment. Wound closure was normalized separately for each experiment and then at least nine individual data points were averaged. Statistical analysis (one-way ANOVA) was carried out with Origin. The *asterisks* in Figs. 3 and 4 indicate significance at a level of 0.01 according to both the Bonferroni and the Tukey test.

Binding Assays—Enzyme linked immunosorbent assays (ELISAs) to test binding of InlB constructs to Met were carried out as described (15).

Surface Plasmon Resonance (SPR)—Interactions of the complete ectodomain of Met (Met$_{928}$) with InlB$_{321}$, InlB$_{392}$ and the B-repeat were analyzed using SPR spectroscopy on a Biacore 3000 device (GE Healthcare). The Met ectodomain was immobilized on a CM5 sensor chip (GE Healthcare) by EDC/NHS coupling. Flow cell one was used as reference cell activated with EDC/NHS and quenched with ethanolamine. Two different chips were used on which 6600 and 6800 RU of Met$_{928}$ were immobilized at a flow rate of 5 μl/min using a citrate buffer at pH 5.5. In total, three measurements were carried out for each InlB variant, one on the first and two on the second chip. The flow rate was 20 μl/min. All analytes were provided in PBS. Concentrations of InlB$_{321}$ and InlB$_{392}$, ranged from 16 μM to 7.8 nM. The B-repeat was measured between 55.8 μM and 0.6 nM. After an injection and a dissociation phase of 120 s each, the chip was regenerated with 1 M NaCl. Experiments with the B-repeat were carried out in between measurements of InlB$_{321}$ and InlB$_{392}$ to exclude a false negative result due to receptor degradation on the chip. For InlB$_{321}$ and InlB$_{392}$, the data were

TABLE 1

Refinement statistics of the InlB B-repeat

Crystallographic refinement statistics for the InlB B-repeat.

Resolution range (Å)	47.0–1.30 (1.33–1.3)
R_{cryst}	15.7 (26.3)
R_{free}	19.6 (27.5)
No. of reflections	
Working set	64294 (4576)
Test set	1989 (140)
No. of atoms	
Protein	2499
Water/ion	399/11
R.m.s. deviation from ideal geometry	
Bond lengths (Å)	0.024
Bond angles (°)	2.097
Ramachandran plot:	
Favored (%)	99.6
Disallowed (%)	0

fit kinetically with a 1:1 binding model with drifting baseline using the BIAevaluation 4.1 software (GE Healthcare). Between two experiments, the difference in K_d of the same protein was bigger than the difference in K_d of InlB$_{321}$ and InlB$_{392}$ within the same experiment. Hence, the K_d of InlB$_{321}$ and InlB$_{392}$ was the same within experimental error. The curve fits in Fig. 5 (Chi^2 of 0.25 and 0.76 for InlB$_{321}$ and InlB$_{392}$, respectively) resulted in a K_d of 178 nM for InlB$_{321}$ and 134 nM for InlB$_{392}$.

Laser-induced Liquid Bead Ion Desorption (LILBID) Mass Spectrometry—Details of the technique have been published elsewhere (41–43). Briefly, aqueous micro droplets containing the protein in the low micromolar range are transferred into vacuum where they are irradiated one by one by pulsed infrared laser radiation with a wavelength corresponding to the water absorption ($\lambda = 3$ μm), leading to stretching vibration of the water molecules and transfer of energy to the liquid droplets. Beyond a certain laser intensity threshold, the droplets "explode" and preformed ions are ejected from the liquid into the gas phase where they are analyzed by time-of-flight mass spectrometry (TOF-MS).

RESULTS

Structure Determination—The InlB B-repeat (residues 322–392) crystallized in space group P2$_1$2$_1$2$_1$ with four molecules in the asymmetric unit. Native crystals diffracted to a resolution of 1.3 Å (supplemental Table S1). The structure was solved by multiwavelength anomalous dispersion using seleno-methionine-labeled protein (supplemental Table S2). The experimental electron density after solvent flattening and phase extension was of excellent quality. The model was built manually and refined to a final R_{free} of 19.6%. (Table 1). There are no residues in the forbidden region of the Ramachandran plot. Anomalous differences from a long-wavelength (1.9 Å) data set of native crystals were interpreted as Ca^{2+} or Cl^{-} ions from the crystallization mixture. The four chains are virtually identical with a pairwise coordinate root mean square deviation (rmsd) of 0.4–0.6 Å for some 440 aligned atoms. Minor deviations are found in two loop regions (amino acids 347–355 and 363–368).

The B-repeat Has a Ubiquitin-like Fold—The B-repeat shows a variation of the β-grasp fold (44) with a single β-sheet consisting of four β-strands in the order 2143. Strands β1 and β4
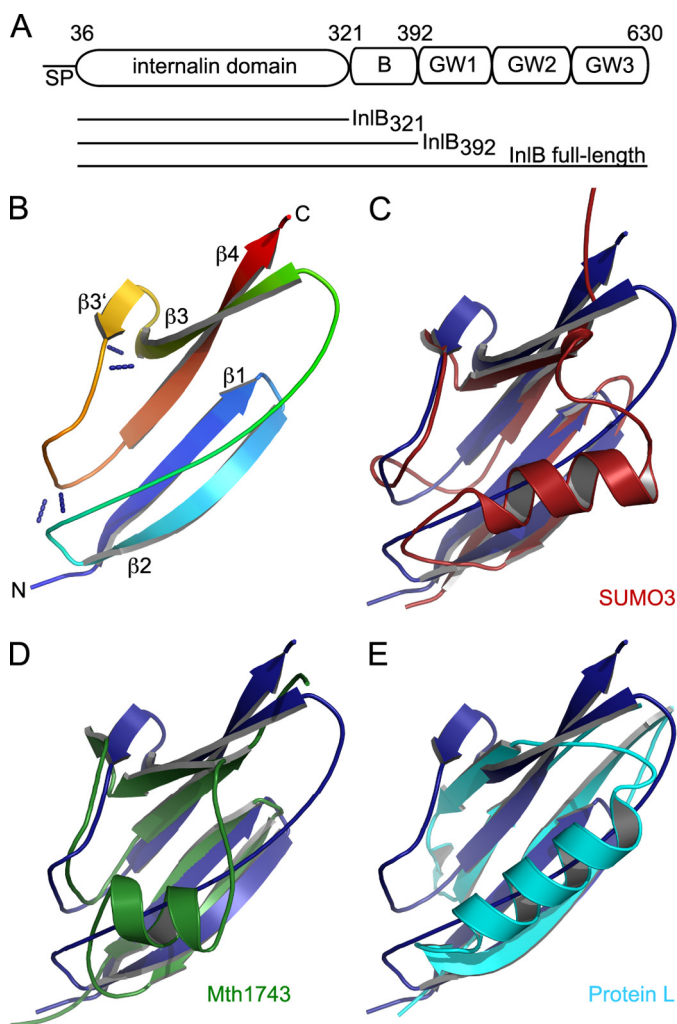
FIGURE 1. **Domain structure of InlB and crystal structure of the B-repeat overlaid with structurally similar proteins.** *A*, domain structure of InlB and the constructs employed in this study. *SP*: signal peptide. *B*, schematic representation of the B-repeat with backbone H-bonds at the connector-arm/shelf and between β3 and β3' shown as *blue dashes*. *C–E*, overlay of the B-repeat (*blue*) with *C*, SUMO3 (*red*; PDB ID 2io1); *D*, Mth1743 from *Methanobacterium thermoautotrophicum* (*green*, PDB ID 1ryj); *E*, a variant of protein L (*cyan*, PDB ID 1kh0).

are parallel to each other and lie at the center, while the edge strands β2 and β3 are anti-parallel to β1 and β4, respectively (Fig. 1*B*). An extended loop forms a right-handed crossover between strands β2 and β3. This essentially results in a two-layered structure with one side of the β-sheet exposed and the other one covered by the loop connecting strands β2 and β3. Structurally the B-repeat is related to ubiquitin and particularly to ubiquitin-like proteins. A search for structurally similar proteins with the SSM server returned as top matches various structures of small ubiquitin-like modifiers (SUMO) with an rmsd of around 2 Å for some 55 structurally aligned residues (Fig. 1*C*). Ubiquitin itself and other ubiquitin-like proteins, *e.g.* NEDD8, also matched well, as did Mth1743 (Fig. 1*D*), a protein of unknown function from *Methanobacterium thermoautotrophicum* (45). Further, similarity exists to two bacterial immunoglobulin-binding proteins, namely *Streptococcus sp.* protein G and protein L (Fig. 1*E*) from *Peptostreptococcus magnus* with an rmsd of around 2.5 Å for some 40 aligned residues.

The B-repeat shares with these proteins the topology of the β-sheet. A major difference between the B-repeat and other β-grasp fold proteins is the connection between strands β2 and β3, which is helical in the classical β-grasp fold (44), while it is extended in the B-repeat. Thus the B-repeat represents the founding member of a new domain superfamily with an ubiquitin-like fold.

The B-repeat shares two additional characteristic features with ubiquitin-like proteins (46). First, there are two backbone hydrogen bonds between strand β3 and a very short additional antiparallel β-strand, which we term β3', at the edge of the sheet (residues 366–369) (Fig. 1, *B* and *C*, Fig. 2*B*). Although the hydrogen bonding pattern is the same in all four crystallographically independent molecules of the B-repeat, DSSP (47) assigns a β-conformation to the additional strand only in chain D, due to considerable structural variation between the four chains. Second, the presence of the additional strand requires a longer loop for the connection between strands β3 (or β3') and β4. This loop, which has also been termed connector arm (46), pairs through main chain hydrogen bonds with the loop connecting strands β2 and β3, which has also been referred to as lateral shelf (Fig. 1*B*).

The Internalin B-repeat Defines the Structure of a Novel Bacterial Domain—A sensitive intermediate profile search of a non-redundant sequence database using HMM-HMM comparison (33) found over 1800 related bacterial sequences with lengths of 37–120 amino acids. Our structure is the first representative of this novel superfamily of bacterial domains. We generated an HMM logo from aligned sequences with a length of 63–73 amino acids, which can be used to explore the structural and sequence determinants of this fold (Fig. 2, *A–C*). Sequence conservation is highest in strands β3 (and its preceding loop) and β4 followed by strand β1. Together with the somewhat less conserved extended connection between strands β2 and β3, these three strands form the hydrophobic core of the protein. The hydrophobic core is packed around three highly conserved residues from strand β4, namely Leu-384, Ala-386, and Phe-388. On one side, Trp-360 from the GW signature motif and the preceding Phe-357 flank those residues. On the other side, Val-325 and Tyr-327 from strand β1 pack against strand β4. The hydrophobic core is completed by conserved residues located in the extended region between strand β2 and β3, among them Ile-344, Pro-347, Pro-350, and Lys-352, which all pack against the hydrophobic face of the β-sheet. Gly-354 and Tyr-355 at the start of strand β3 are also conserved and the phenolic side chain caps the hydrophobic core at the C-terminal end. Further conserved residues are Gly-331 in the β-turn between stands β1 and β2 and Asp-381 located just before strand β4, the side-chain of which forms a hydrogen bond to the backbone of strand β1, stabilizing its β-conformation.

Strand β2 hardly contributes to the hydrophobic core and is poorly conserved. Even higher sequence variability is found in the long loop between strands β3 and β4. In this region the InlB B-repeat contains four aromatic side chains (Trp-370, Phe-372, Tyr-376, and Phe-382). With the exception of Tyr-376, these are mostly buried. However, these aromatic residues are not
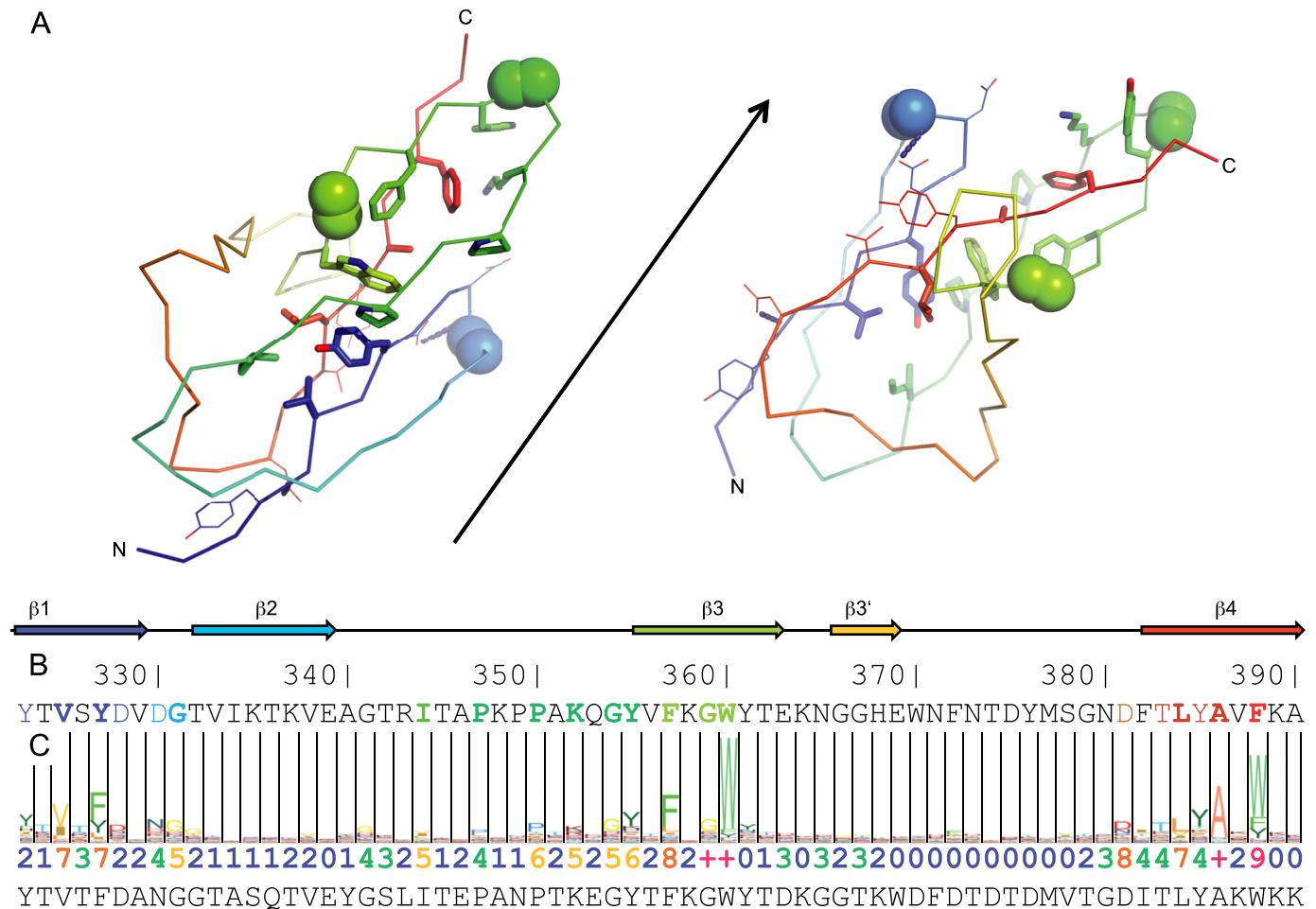
FIGURE 2. **Conserved residues and hydrophobic core of the B-repeat.** *A*, ribbon representation of the B-repeat in two orientations rotated by 180° (axis indicated as *arrow*). Coloring as in Fig. 1*B*. Residues that form the hydrophobic core and are conserved among homolgous bacterial proteins are shown as *sticks*. Conserved glycines are shown as *spheres*. Further conserved residues are shown as *thin lines*. *B*, sequence of the B-repeat and secondary structure elements indicated above. Colored residues are those that are shown in *A* as *lines* or, if in *bold face type*, as *sticks* and *spheres*. *C*, HMM logo without gaps generated from an alignment of over 300 homologous sequences with a length between 63 and 73 residues. The conservation score and the consensus sequence calculated with Jalview from the same alignment are shown below.

conserved, suggesting that they are less crucial to the fold than Tyr-327, Tyr-355, Phe-357, Trp-360, and Phe-388. In homologous proteins from other species insertions are found at various sites but cluster mainly in two positions, namely in the long connections between strands $\beta 2$ and $\beta 3$ and between strands $\beta 3$ and $\beta 4$.

The B-repeat Is Flexibly Attached to the Internalin Domain—In the published structure of full-length InlB, the electron density of the B-repeat was too weak to reliably model it *de novo* (21). Nevertheless, there is some positive difference density in the region of the B-repeat, especially close to the internalin domain. The length of the B-repeat (some 40 Å from Val-322 to Thr-392) is similar to the gap size in the structure of full-length InlB (some 48 Å between Leu-319 and Thr-392), suggesting that the structure of the B-repeat in the complete protein is the same as the structure of the isolated domain. Therefore, we tried to position our structure of the B-repeat in the data of the full-length molecule by molecular replacement, but these attempts failed. To determine the relative orientation of the internalin domain and the B-repeat we crystallized InlB$_{392}$ and solved its structure at a resolution of 3.2 Å ($R_{\text{free}} = 22.5\%$; [supplemental Tables S1 and S3](#)). The internalin domain could be

located easily by molecular replacement. However, attempts to locate the B-repeat failed. Moreover, there is no difference density for the B-repeat. The crystals of InlB$_{392}$ have a high solvent content of 80% (Matthews coefficient of 6.4) and the molecules arrange such that the C terminus of the internalin domain points into large solvent channels ([supplemental Fig. S1](#)). The internalin domain forms all contacts, while the B-repeat dangles freely into the solvent channel. Apparently, the relative domain orientation of the internalin domain and the B-repeat is not fixed but highly flexible.

InlB$_{392}$ Stimulates Wound Healing in Primate Cells—We compared the ability of InlB$_{321}$ and InlB$_{392}$ (Fig. 1*A*) to induce cellular phenotypes. HGF/SF and full-length InlB were used as positive controls. First, we used an *in vitro* scratch wound assay to assess stimulation of cell motility in A549 cells. A549 cells showed a basal level of wound closure in the absence of ligand, which was subtracted from all experiments with ligand for normalization. InlB$_{321}$ at 1 nM did not increase wound closure beyond the basal level. In contrast, 1 nM InlB$_{392}$ stimulated wound closure to an extent similar to that of full-length InlB or HGF/SF at the same concentration (Fig. 3*A*). We repeated the experiment with Vero cells (Fig. 3, *B* and *C*), as Vero cells have
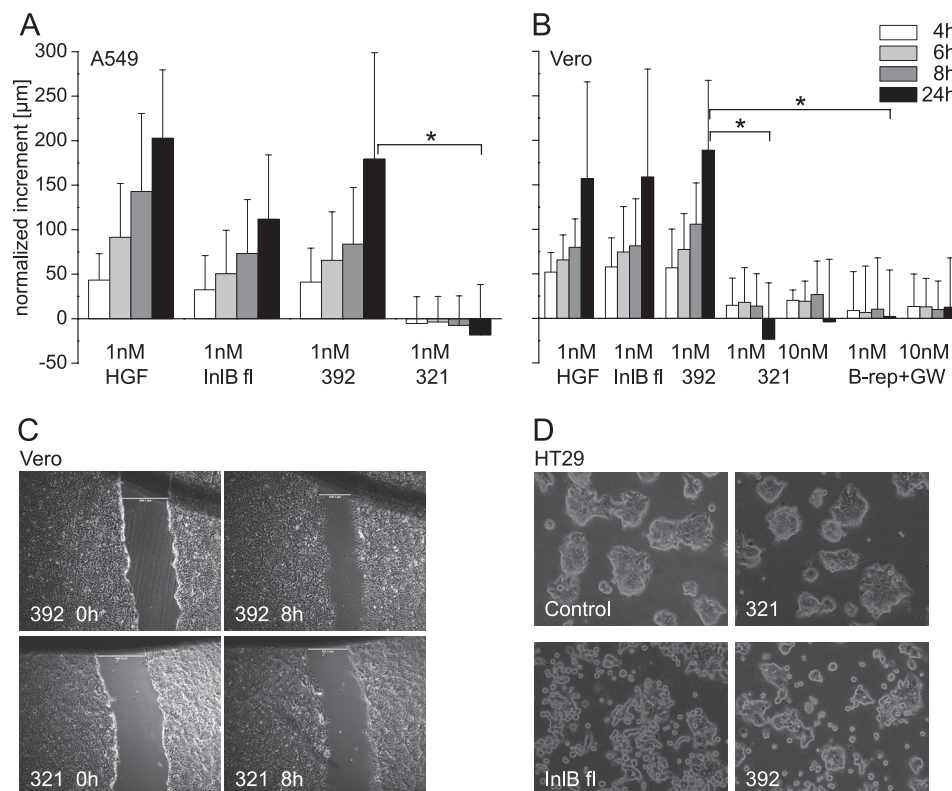
FIGURE 3. **InlB$_{392}$ stimulates wound healing and scatter in primate cells.** *A* and *B*, closure of scratch wounds in A549 or Vero cells upon incubation with various ligands. For normalization, wound closure of cells incubated without ligand was subtracted. Full-length InlB and InlB$_{392}$ stimulated wound healing comparable to HGF/SF, while all other InlB constructs were inactive. The *asterisk* marks statistical significance at a level of 0.01. Note that for the sake of clarity, statistical significance is only shown for the two most relevant pairs of proteins (InlB$_{392}$ *versus* InlB$_{321}$ and InlB$_{392}$ *versus* B-rep+GW) at 24 h and 1 nM. *C*, closure of scratch wounds in Vero cells upon incubation with 1 nM InlB$_{392}$ (*top*) and 1 nM InlB$_{321}$ (*bottom*). *D*, full-length InlB and InlB$_{392}$ at 1 nM stimulate scattering of HT29 cells, while 1 nM InlB$_{321}$ shows little or no activity.

frequently been used in biochemical experiments with InlB. InlB$_{392}$ at 1 nM showed an activity comparable to that of HGF/SF and full-length InlB. InlB$_{321}$, in contrast, was inactive. In Vero cells, we also tested InlB$_{321}$ at 10 nM, but again found no activity. We also tested, whether a construct comprising the B-repeat and the GW domains can stimulate cell motility independent of the internalin domain. This construct showed no activity at 1 and 10 nM.

InlB$_{392}$ Stimulates Cell Scatter of Human Cells—In initial assays with MDCK cells 1 nM InlB$_{392}$ did not stimulate cell scatter. This apparent discrepancy with the wound healing assays may be due to the different assay format or to the different cell lines. To resolve this issue, we performed scatter assays with primate cells and wound healing assays with MDCK cells. For scatter assays, we used HT29 cells (Fig. 3*D*). Similar to MDCK cells, these cells showed no or only a very weak response to InlB$_{321}$. In contrast, InlB$_{392}$ clearly induced cell scatter at 1 nM, although somewhat less pronounced than HGF/SF or full-length InlB (Fig. 3*D*).

InlB$_{392}$ Stimulates Cell Motility of Canine Cells Only at Higher Concentration—In scatter assays, we used a MDCK cell isolate specifically selected to form tight colonies in the absence of HGF/SF (clone 20 from E. Gherardi, MRC Cambridge). Due to the particularly strong intercellular junctions, it was not possible to generate scratch wounds without completely destroying the confluent monolayer of these cells. Therefore we switched to another isolate of MDCK cells obtained from

ATCC that allowed formation of scratch wounds. In both scatter and wound healing assays InlB$_{392}$ was inactive at 1 nM, while cells showed a clear response to full-length InlB at that concentration (Fig. 4, *A* and *B*). This showed that the difference in response to 1 nM InlB$_{392}$ between wound healing of A549 and scatter of MDCK cells is due to the different cell types and not the different assay formats. Next, we repeated the assays with MDCK cells with a higher concentration of InlB$_{392}$. At 10 nM of InlB$_{392}$ MDCK cells showed a clear response in the scatter assay and increased wound healing, although wound closure was less pronounced than with full-length InlB (Fig. 4, *A* and *B*). InlB$_{321}$ at 1 nM and 10 nM was completely inactive in both assays and we had shown previously that it stays inactive in scatter assays up to 1 μM (16). The construct consisting of the B-repeat and the GW domains did not stimulate cell scatter or wound healing at a concentration of 10 nM (Fig. 4, *A* and *B*). Taken together the cellular assays showed that InlB$_{392}$ could stimulate cell motility, whereas both InlB$_{321}$ and the construct B-repeat+GW were inactive. Thus, the presence of the B-repeat in InlB$_{392}$ conferred to this protein the ability to elicit cellular phenotypes that the isolated internalin domain cannot induce.

The InlB B-repeat Does Not Bind to Met—The observed cellular effect could be explained if the B-repeat directly interacted with Met, thereby increasing the affinity of InlB$_{392}$ for Met. Therefore, we tested in an ELISA whether the B-repeat could bind to Met. ELISA plates were coated with the complete, recombinantly produced ectodomain of Met and incubated
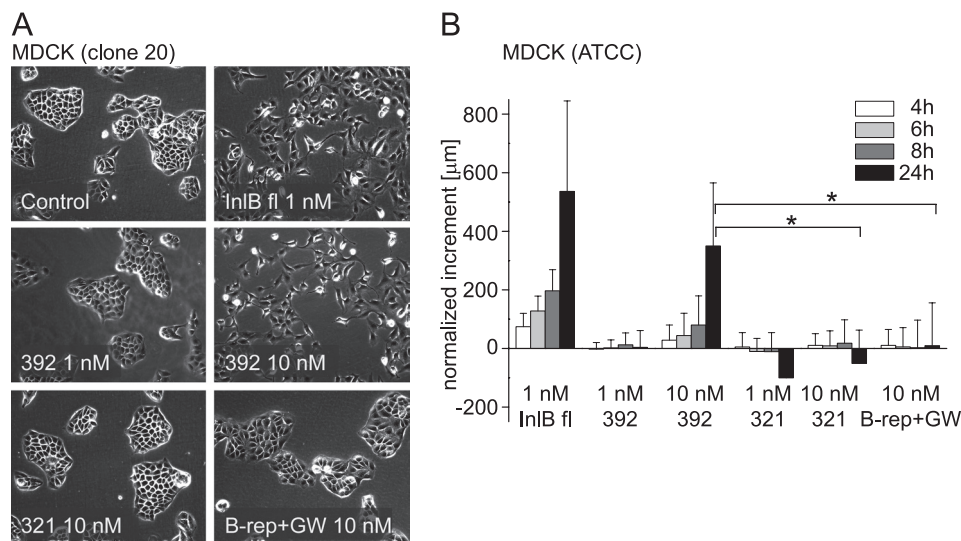
FIGURE 4. **InlB$_{392}$ stimulates wound healing and scatter in canine cells.** *A*, full-length InlB stimulates scattering of MDCK cells (clone 20) at 1 nM. InlB$_{392}$ is inactive at 1 nM but induces scattering at 10 nM. InlB$_{321}$ and a construct consisting of the InlB B-repeat and the GW domains are inactive even at 10 nM. *B*, closure of scratch wounds in MDCK cells upon incubation with various ligands. InlB$_{392}$ is inactive at 1 nM but shows activity at 10 nM. InlB$_{321}$ and the construct B-repeat+GW domains are both inactive at 1 nM and 10 nM. The *asterisk* marks statistical significance at a level of 0.01. Note that for the sake of clarity, statistical significance is only shown for the two most relevant pairs of proteins (InlB$_{392}$ *versus* InlB$_{321}$ and InlB$_{392}$ *versus* B-rep+GW) at 24 h and 10 nM.

with a GST-fusion of the B-repeat. We did not detect binding of the B-repeat, whereas a GST-fusion of the internalin domain used as positive control clearly bound to Met (supplemental Fig. S2). Using the same assay format there was no difference in binding affinity to Met between InlB$_{321}$ and InlB$_{392}$ (supplemental Fig. S2). To test for weak binding of the B-repeat to Met, we also performed surface plasmon resonance experiments, in which the complete Met ectodomain was coupled to a CM5 sensor chip. InlB$_{321}$, InlB$_{392}$, and the B-repeat were used as analyte in the mobile phase. There was no difference in binding affinity between InlB$_{321}$ and InlB$_{392}$ and the isolated B-repeat did not bind to Met at all (Fig. 5).

The B-repeat Does Not Increase the Propensity of InlB or InlB/ Met Complexes to Dimerize—InlB promotes dimerization of the Met receptor by a 2-fold symmetric contact on the convex face of the LRR (16, 17). The Ig2 domains of the Met stalks form an additional contact. The forces stabilizing this 2:2 assembly are very weak because it is observed in crystals, but not in solution (15, 16, 48). The 2-fold symmetric assembly of InlB from these 2:2 InlB/Met complexes is also present in the structure of full-length InlB (21). The existing difference density for the B-repeat in the structure of full-length InlB indicates that also the B-repeat makes contact with a symmetry mate related by the crystallographic 2-fold axis. Therefore, we initially hypothesized that this additional contact of the B-repeat may stabilize the dimeric assembly. However, in contrast to our expectations, the InlB$_{392}$ crystals described above did not show this 2-fold symmetric packing.

We also analyzed the packing of molecules in the crystals of the isolated B-repeat. Chains A and B pack against each other in a way virtually identical to the packing of chains C and D, thus forming two very similar pairs of molecules (supplemental Fig. S3*A*). Analysis of this contact with the PISA server (49) reports a buried surface area of some 1200 Å^2 and five or six hydrogen bonds, putting it into the gray area of assemblies that may or may not be stable in solution. These potential dimers are not

2-fold symmetric. Instead, monomers are related by a rotation of 176° and a translational component (supplemental Fig. S3, *B and C*). In contrast, the vast majority of biological dimers are 2-fold symmetric (50, 51) and also the 2:2 InlB/Met complex shows 2-fold symmetry. If this contact between two B-repeats were to stabilize a 2:2 InlB/Met complex, asymmetry would need to be introduced in the short linker between the end of the internalin domain and the start of the B-repeat, which is very unlikely.

Finally, we assayed the oligomeric state of InlB$_{392}$ alone and in complex with the Met ectodomain by gel filtration (data not shown) and by LILBID mass spectrometry, which is more sensitive for low affinity complexes. Judged by these methods, InlB$_{392}$ is monomeric on its own and forms a 1:1 complex with the Met ectodomain (Fig. 6 and supplemental Fig. S4). Hence, both the arrangement of molecules in the crystalline state and their behavior in solution argued against a role of the B-repeat in stabilizing 2:2 InlB/Met complexes.

DISCUSSION

The B-repeat May Be Related to Mucin-binding Protein Repeats—Is the structural similarity to SUMO the remnant of a common ancestor or is it caused by convergent evolution? The ubiquitin fold has been classified as a super-fold that is found in several different protein families that appear evolutionarily unrelated (52) and a monophyletic origin of all proteins with β-grasp fold has been questioned (44). A structure based sequence alignment revealed limited sequence conservation between the B-repeat and SUMO in residues that determine the fold (compare our HMM logo for the B-repeat with the Pfam HMM logo for SUMO and see supplemental Fig. S5). Thus, it is conceivable that this structural similarity is due to convergent rather than divergent evolution. Instead, we noticed some similarity of the B-repeat to the structure of the fifth repeat (R5) of mucus-binding protein (Mub) from *Lactobacillus reuteri* (53). The Mub-R5
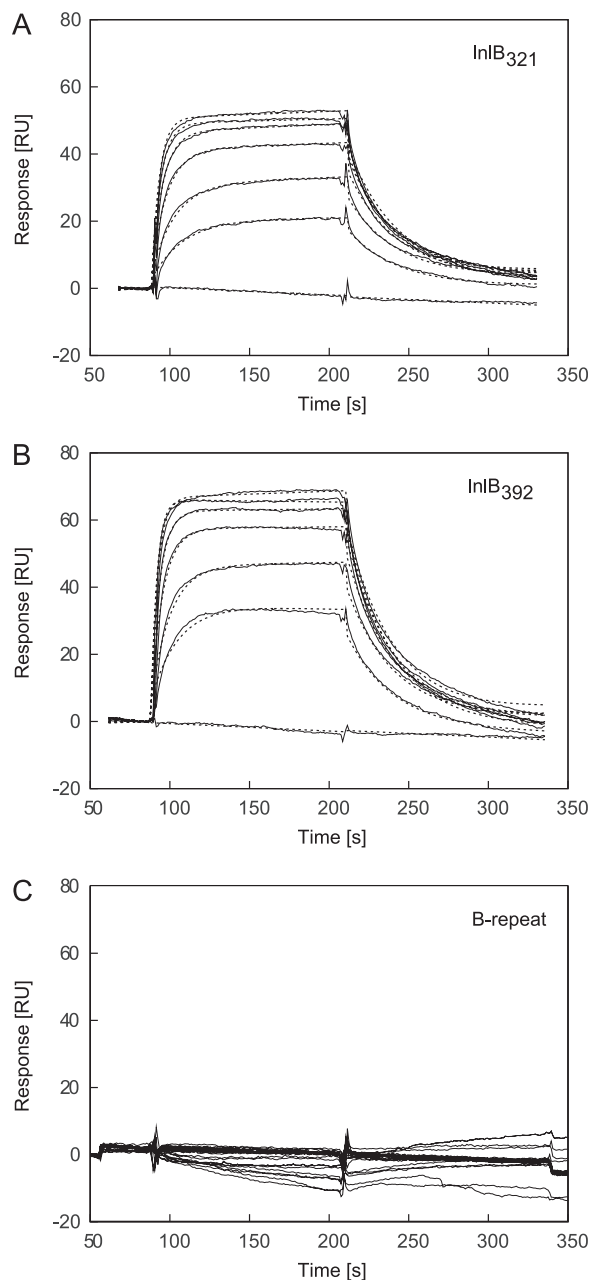
FIGURE 5. **Surface plasmon resonance shows no binding of the InlB B-re-peat to Met.** Sensorgrams depicting the binding of different InlB constructs to the complete extracellular domain of Met are shown. Varying concentrations of InlB were injected over a chip surface coated with the Met ectodomain. The *solid lines* represent the measured response in resonance units (*RU*). *A* and *B*, $InlB_{321}$ and $InlB_{392}$ were injected at concentrations of 0, 100, 200, 400, 600, 800, and 1000 nM. *Dashed lines* represent fits using a 1:1 Langmuir interaction with drifting baseline. *C*, overlay of 24 sensorgrams of the B-repeat in the concentration range of 0.6 nM to 55.8 μM.

FIGURE 6. **Native mass spectrometry shows that $InlB_{392}$, just like $InlB_{321}$, forms 1:1 complexes with Met_{928}.** LILBID mass spectrum of Met_{928} and of the Met_{928} + $InlB_{321}$ and Met_{928} + $InlB_{392}$ complexes. Under soft desorption conditions the intact Met_{928} + $InlB_{321}$ and Met_{928} + $InlB_{392}$ complexes appear at four different overall charge states (the *black sticks* indicate the theoretical mass/charge positions).

repeat comprises 184 amino acids, which form two separate domains of 75 (called B1) and 109 residues (B2) (53). The two domains in Mub-R5 are structurally related and both show structural similarity to the B-repeat (Fig. 7, *A* and *B* and supplemental Table S4). Mub-R5 B1 has a canonical β-grasp fold with a four-stranded β-sheet and a single helix (Fig. 7*A*). The short extra strand following β3 is missing and protein L is its closest structural homolog. The larger C-terminal B2 domain lacks the helix, and its extra residues form an addi-
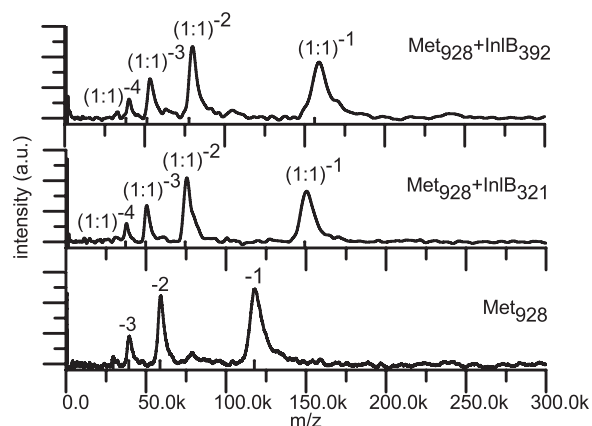
tional three-stranded β-sheet (Fig. 7*B*). Our SSM search had missed the similarity between the B-repeat and the Mub-R5 domains B1 and B2. As an alternative option to search for structurally similar proteins, we used the Dali server (54). In a search limited to the PDB90, a representative subset of PDB chains that share less than 90% sequence identity, SUMO was again found to be most similar with a Z-score of 5.5 and an rmsd of 2.3 Å for 58 aligned residues. Interestingly, the Mub-R5 B1 repeat scored second (Z-score of 5.2; rmsd of 2.4 Å for 61 aligned residues). However, this similarity was only found for chain B, while for the other three chains of the B-repeat (A, C, & D) Mub-R5 was not listed among the proteins with Z-scores above 2. Therefore, we performed a pairwise comparison between chains A, C, or D and the Mub-R5 B1 domain using DaliLite (55), which gave Z-scores above 5 for all. The DaliLite Z-scores for pairwise comparison of the Mub-R5 B2 domain with the B-repeat are lower although the rmsd is similar to that of the B1 domain, probably because the B2 domain is bigger.

A Dali search for proteins structurally similar to Mub-R5 found as top hits the MucBP (mucin-binding protein) domain of the adhesion protein PEPE_0118 from *Pediococcus pentosaceus* (PDB ID 3lyy) and two domains of a putative peptidoglycan bound protein lmo0835 from *L. monocytogenes* (residues 34–128; PDB ID 2kt7; and residues 161–235 2kvz). The B-repeat can be aligned to all of these structures with reasonable Z-scores and rmsd values (Fig. 7 and supplemental Table S4). All except the B1 domain of Mub-R5 share the β-grasp fold lacking the helix (Fig. 7). Back-bone hydrogen bonds between the connector arm and the lateral shelf are also found in all cases, although only the B-repeat has the extra strand next to β3 (Fig. 7) and the GW signature motif (supplemental Fig. S6). A structure-based sequence alignment of these five domains with the B-repeat shows conservation of several residues that are conserved in the B-repeat and among these four bacterial protein domains (supplemental Fig. S6). Most notably, the GY motif is found in all domains and is structurally equivalent forming the start of strand β3 (supplemental Fig. S7). All domains are packed
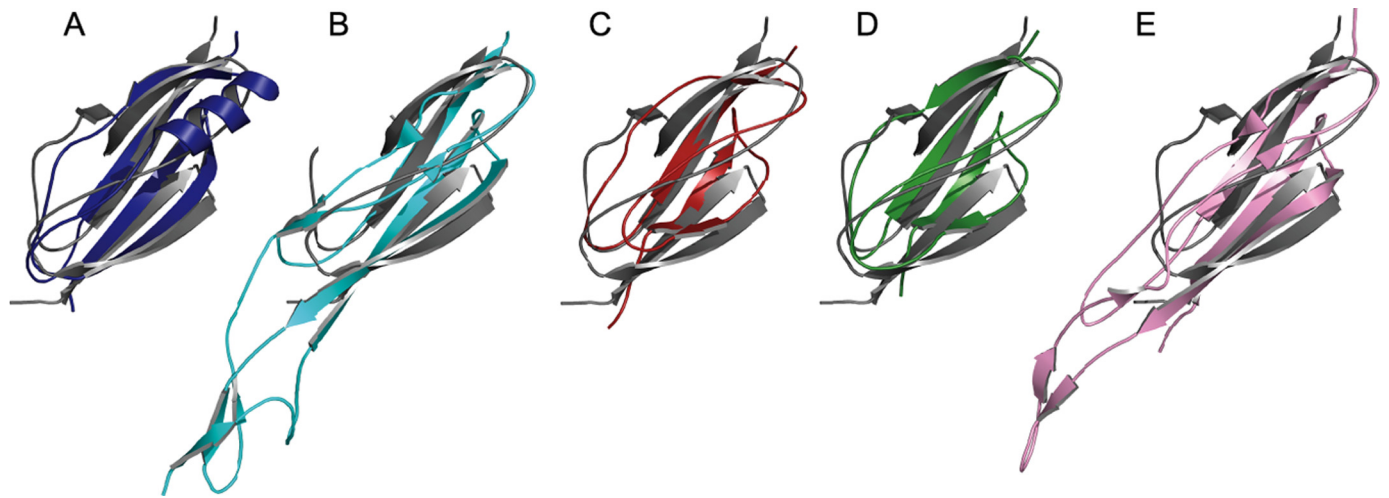
FIGURE 7. **Pairwise overlay of the InlB B-repeat and structurally related domains from other bacterial surface proteins.** All structures are shown in schematic representation. The InlB B-repeat (*gray*) is overlaid with *A*, the B1 (*dark blue*) and *B*, the B2 (*cyan*) domain of repeat 5 of mucus-binding protein (*Mub-R5*) from *Lactobacillus reuteri* (PDB ID 3i57); *C*, residues 34–128 (*red*, PDB ID 2kt7) and *D*, residues 161–235 (green, PDB ID 2kvz) of lmo0835, a putative peptidoglycan-bound protein from *Listeria monocytogenes*; *E*, the adhesion protein PEPE_0118 from *Pediococcus pentosaceus* (*pink*, PDB ID 3lyy).

around hydrophobic residues from strand $\beta4$, of which the last is aromatic (supplemental Fig. S7). Other residues forming the hydrophobic core or turns are also conserved, although with a lower level of identity. In addition to the structural and sequence similarity, there are functional similarities. First, all of these proteins are extracellular bacterial proteins, probably involved in binding of host cell proteins. Second, these domains are frequently arranged in tandem and mucin-binding repeats are even present in several internalins. Although one cannot rule out that all this is mere coincidence and the result of convergent evolution, it seems well possible that the B-repeat is evolutionarily related to these other bacterial domains, sharing a common ancestor.

Potential Functions of the B-repeat/Flg_New Superfamily— Sequences similar to the B-repeat are confined to bacteria, as our search with HHSenser found no homologs in archaea or eukaryotes, but they are not limited to the *Listeria* internalins. Pfam (56) defines a "Flg_new" family (PF09479) that is present in at least 138 different proteins from 50 bacterial species including both Gram-positive and Gram-negative bacteria. We inspected the domain architecture and annotation of proteins containing the B-repeat/Flg_new domain in Pfam. The vast majority, if not all of the proteins, are secreted or cell surface proteins. So far, no function has been attributed to most of these proteins and many are only annotated as hypothetical protein. Generally, the proteins containing B-repeat/Flg_new domains are multi-domain proteins with either a single or several, often tandemly repeated, copies of the B-repeat. Some proteins contain no recognizable domains in addition to the B-repeat/Flg_new domains, *e.g.* a cell wall attached protein from *Listeria inoccua* and a putative uncharacterized protein from *Coprococcus eutactus* with 11 tandem repeats. Even more repeats are present in predicted proteins from *Clostridium hylemonae* (13 repeats) and from *Mollicutes bacterium D7* (26 repeats). This domain organization is reminiscent of classical repeat domains like Ig and Ig-like domains, cadherin domains, and fibronectin

type III (FN-III) repeats, all of which are found in cell surface molecules as well. The B-repeat is structurally analogous to these domains. It has about the same size, a roughly oval shape, and N and C termini are located on opposite sides of the long axis, allowing the domains to be arranged like beads on a string. The analogies in terms of structure, domain architecture and context suggest that also some functional analogy may exist between B-repeat/Flg_new domains and Ig, FN-III and cadherin domains. Two potential functions seem likely. One is a structural role where the domain acts as a spacer between a functionally important domain and the bacterial cell surface. The other potential function is the binding of receptor molecules, *e.g.* on eukaryotic host cells.

*Function of the B-repeat in InlB—*Ghosh and co-workers (25) first demonstrated the functional relevance of the B-repeat and suggested that it may bind to a receptor other than Met. Our cellular assays also show that the InlB B-repeat clearly has a function beyond that of a mere spacer between the internalin and the GW domains. A direct interaction with Met seems unlikely. Our binding assay showed no binding of the B-repeat to Met. Consistent with this observation, the structure of the complex between Met and the InlB internalin domain suggests that the B-repeat will point away from Met (15). Likewise, there is no experimental support for the idea that the B-repeat may promote dimerization of the Met ectodomain by dimerizing itself. Instead it seems most likely that the InlB B-repeat will bind another host cell receptor, an idea that is also supported by the different response to InlB$_{392}$ in primate and canine cells. At present, however, the identity of such a receptor is unclear. The known InlB receptor gC1q-R, was shown to interact with the GW domains, not the B-repeat (21). CD44v6 is a candidate (57), but its importance for Met signaling has recently been questioned (58). The similarity to repeats from mucin-binding proteins is suggestive. Interestingly, several internalins including InlB were reported to bind MUC2 from human intestinal mucin, but in InlJ the internalin domain was found

to be sufficient for binding of MUC2 (59). Moreover, the Mub-R5 domain was shown to bind human immunoglobulins (53). Thus, not all mucin-binding repeats seem to interact (exclusively) with mucins and the potential receptor for the B-repeat still needs to be identified.

Potential Binding Site—An obvious question is whether one can use the structural similarity to ubiquitin-like proteins to exploit the extensive biochemical and structural data about their protein-protein interactions to predict a potential receptor binding site in the InlB B-repeat. A recent comprehensive survey showed that binding sites of proteins with a β-grasp fold vary widely (46). Basically any part of the molecular surface is used for binding purposes by at least some of these proteins. A comparison of 16 ubiquitin complexes showed that ubiquitin alone uses three quarters of its accessible surface area to make contact with different binding proteins (60). However, two hot-spots exist in β-grasp fold proteins that mediate most interactions. One is the exposed face of the β-sheet, the second is a groove between the helix and strand β2 on the obscured face of the β-sheet (46, 60). In some cases, the latter interaction also involves backbone hydrogen bonds between strand β2 of the β-grasp fold and an edge β-strand of the binding partner. Such β-sheet extension is also employed by protein L and protein G to bind antibodies (61, 62). Thus, both the exposed face of the β-sheet and the poorly conserved edge strand β2 are good candidates for receptor binding sites in the B-repeat.

REFERENCES

1. Allerberger, F., and Wagner, M. (2010) *Clin. Microbiol. Infect.* **16,** 16–23
2. Gilmour, M. W., Graham, M., Van Domselaar, G., Tyler, S., Kent, H., Trout-Yakel, K. M., Larios, O., Allen, V., Lee, B., and Nadon, C. (2010) *BMC. Genomics* **11,** 120
3. Hamon, M., Bierne, H., and Cossart, P. (2006) *Nat. Rev. Microbiol.* **4,** 423–434
4. Cossart, P., and Toledo-Arana, A. (2008) *Microbes. Infect.* **10,** 1041–1050
5. Mostowy, S., and Cossart, P. (2009) *Cell. Host Microbe* **5,** 510–513
6. Dramsi, S., Biswas, I., Maguin, E., Braun, L., Mastroeni, P., and Cossart, P. (1995) *Mol. Microbiol.* **16,** 251–261
7. Shen, Y., Naujokas, M., Park, M., and Ireton, K. (2000) *Cell* **103,** 501–510
8. Birchmeier, C., Birchmeier, W., Gherardi, E., and Vande Woude, G. F. (2003) *Nat. Rev. Mol. Cell. Biol.* **4,** 915–925
9. Bierne, H., and Cossart, P. (2002) *J. Cell. Sci.* **115,** 3357–3367
10. Bierne, H., Sabet, C., Personnic, N., and Cossart, P. (2007) *Microbes Infect.* **9,** 1156–1166
11. Schubert, W. D., Göbel, G., Diepholz, M., Darji, A., Kloer, D., Hain, T., Chakraborty, T., Wehland, J., Domann, E., and Heinz, D. W. (2001) *J. Mol. Biol.* **312,** 783–794
12. Kobe, B., and Kajava, A. V. (2001) *Curr. Opin. Struct. Biol.* **11,** 725–732
13. Banerjee, M., Copp, J., Vuga, D., Marino, M., Chapman, T., van der Geer, P., and Ghosh, P. (2004) *Mol. Microbiol.* **52,** 257–271
14. Machner, M. P., Frese, S., Schubert, W. D., Orian-Rousseau, V., Gherardi, E., Wehland, J., Niemann, H. H., and Heinz, D. W. (2003) *Mol. Microbiol.* **48,** 1525–1536
15. Niemann, H. H., Jäger, V., Butler, P. J., van den Heuvel, J., Schmidt, S., Ferraris, D., Gherardi, E., and Heinz, D. W. (2007) *Cell* **130,** 235–246
16. Ferraris, D. M., Gherardi, E., Di, Y., Heinz, D. W., and Niemann, H. H. (2010) *J. Mol. Biol.* **395,** 522–532
17. Niemann, H. H. (2011) *Eur. J. Cell. Biol.* doi 10.1016/j.ejcb.2010.11.014
18. Jonquières, R., Pizarro-Cerdá, J., and Cossart, P. (2001) *Mol. Microbiol.* **42,** 955–965
19. Braun, L., Dramsi, S., Dehoux, P., Bierne, H., Lindahl, G., and Cossart, P. (1997) *Mol. Microbiol.* **25,** 285–294
20. Jonquières, R., Bierne, H., Fiedler, F., Gounon, P., and Cossart, P. (1999) *Mol. Microbiol.* **34,** 902–914
21. Marino, M., Banerjee, M., Jonquières, R., Cossart, P., and Ghosh, P. (2002) *EMBO. J.* **21,** 5623–5634
22. Gaillard, J. L., Berche, P., Frehel, C., Gouin, E., and Cossart, P. (1991) *Cell* **65,** 1127–1141
23. Dramsi, S., Dehoux, P., Lebrun, M., Goossens, P. L., and Cossart, P. (1997) *Infect. Immun.* **65,** 1615–1625
24. Braun, L., Nato, F., Payrastre, B., Mazié, J. C., and Cossart, P. (1999) *Mol. Microbiol.* **34,** 10–23
25. Copp, J., Marino, M., Banerjee, M., Ghosh, P., and van der Geer, P. (2003) *J. Biol. Chem.* **278,** 7783–7789
26. Guerrero, S. A., Hecht, H. J., Hofmann, B., Biebl, H., and Singh, M. (2001) *Appl. Microbiol. Biotechnol.* **56,** 718–723
27. Kabsch, W. (2010) *Acta. Crystallogr. D Biol. Crystallogr.* **66,** 125–132
28. Pape, T., and Schneider, T. R. (2004) *J. Appl. Crystallogr.* **37,** 843–844
29. Sheldrick, G. M. (2010) *Acta. Crystallogr. D Biol. Crystallogr.* **66,** 479–485
30. McCoy, A. J. (2007) *Acta. Crystallogr. D Biol. Crystallogr.* **63,** 32–41
31. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) *Acta. Crystallogr. D Biol. Crystallogr.* **66,** 486–501
32. Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F., and Murshudov, G. N. (2004) *Acta. Crystallogr. D Biol. Crystallogr.* **60,** 2184–2195
33. Söding, J., Remmert, M., Biegert, A., and Lupas, A. N. (2006) *Nucleic Acids Res.* **34,** W374–378
34. Eddy, S. R. (2008) *PLoS. Comput. Biol.* **4,** e1000069
35. Schuster-Böckler, B., Schultz, J., and Rahmann, S. (2004) *BMC. Bioinformatics.* **5,** 7
36. DeLano, W. L. (2002) *World Wide Web,* www.pymol.org
37. Krissinel, E., and Henrick, K. (2004) *Acta. Crystallogr. D Biol. Crystallogr.* **60,** 2256–2268
38. Ortiz, A. R., Strauss, C. E., and Olmea, O. (2002) *Protein. Sci.* **11,** 2606–2621
39. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) *Bioinformatics* **25,** 1189–1191
40. Abramoff, M. D., Magelhaes, P. J., and Ram, S. J. (2004) *Biophotonics Int.* **11,** 36–42
41. Morgner, N., Barth, H. D., and Brutschy, B. (2006) *Austral. J. Chem.* **59,** 109–114
42. Morgner, N., Kleinschroth, T., Barth, H. D., Ludwig, B., and Brutschy, B. (2007) *J. Am. Soc. Mass Spectrom.* **18,** 1429–1438
43. Morgner, N., Zickermann, V., Kerscher, S., Wittig, I., Abdrakhmanova, A., Barth, H. D., Brutschy, B., and Brandt, U. (2008) *Biochim. Biophys. Acta.* **1777,** 1384–1391
44. Overington, J. P. (1992) *Curr. Opin. Struct. Biol.* **2,** 394–401
45. Yee, A., Chang, X., Pineda-Lucena, A., Wu, B., Semesi, A., Le, B., Ramelot, T., Lee, G. M., Bhattacharyya, S., Gutierrez, P., Denisov, A., Lee, C. H., Cort, J. R., Kozlov, G., Liao, J., Finak, G., Chen, L., Wishart, D., Lee, W., McIntosh, L. P., Gehring, K., Kennedy, M. A., Edwards, A. M., and Arrowsmith, C. H. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99,** 1825–1830
46. Burroughs, A. M., Balaji, S., Iyer, L. M., and Aravind, L. (2007) *Biol. Direct* **2,** 18
47. Kabsch, W., and Sander, C. (1983) *Biopolymers* **22,** 2577–2637
48. Niemann, H. H., Petoukhov, M. V., Härtlein, M., Moulin, M., Gherardi, E., Timmins, P., Heinz, D. W., and Svergun, D. I. (2008) *J. Mol. Biol.* **377,** 489–500
49. Krissinel, E., and Henrick, K. (2005) in *CompLife* (Berthold, M. R., ed), Springer-Verlag, Berlin Heidelberg

50. Levy, E. D., Boeri Erba, E., Robinson, C. V., and Teichmann, S. A. (2008) *Nature* **453,** 1262–1265
51. Schulz, G. E. (2010) *J. Mol. Biol.* **395,** 834–843
52. Orengo, C. A., Jones, D. T., and Thornton, J. M. (1994) *Nature* **372,** 631–634
53. MacKenzie, D. A., Tailford, L. E., Hemmings, A. M., and Juge, N. (2009) *J. Biol. Chem.* **284,** 32444–32453
54. Holm, L., and Rosenström, P. (2010) *Nucleic. Acids. Res.* **38,** suppl., W545–W549
55. Hasegawa, H., and Holm, L. (2009) *Curr. Opin. Struct. Biol.* **19,** 341–348
56. Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L., Eddy, S. R., and Bateman, A. (2010) *Nucleic. Acids. Res.* **38,** D211–D222
57. Jung, C., Matzke, A., Niemann, H. H., Schwerk, C., Tenenbaum, T., and Orian-Rousseau, V. (2009) *Mol. Microbiol.* **72,** 1196–1207
58. Dortet, L., Veiga, E., Bonazzi, M., and Cossart, P. (2010) *Microbes Infect.* **12,** 919–927
59. Lindén, S. K., Bierne, H., Sabet, C., Png, C. W., Florin, T. H., McGuckin, M. A., and Cossart, P. (2008) *Arch. Microbiol.* **190,** 101–104
60. Perica, T., and Chothia, C. (2010) *Curr. Opin. Struct. Biol.* **20,** 367–376
61. Derrick, J. P., and Wigley, D. B. (1992) *Nature* **359,** 752–754
62. Graille, M., Stura, E. A., Housden, N. G., Beckingham, J. A., Bottomley, S. P., Beale, D., Taussig, M. J., Sutton, B. J., Gore, M. G., and Charbonnier, J. B. (2001) *Structure* **9,** 679–687