Structural basis for adaptation of lactobacilli to gastrointestinal mucus

Sabrina Etzold,1 Olivia I. Kober,1
Donald A. MacKenzie,1 Louise E. Tailford,1
A. Patrick Gunning,1 John Walshaw,1
Andrew M. Hemmings2* and Nathalie Juge1**

1Institute of Food Research, Gut Health and Food Safety Institute Strategic Programme, Norwich Research Park, Norwich NR4 7UA, UK.
2Schools of Biological Sciences and Chemistry, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK.

Summary

The mucus layer covering the gastrointestinal (GI) epithelium is critical in selecting and maintaining homeostatic interactions with our gut bacteria. However, the underpinning mechanisms of these interactions are not understood. Here, we provide structural and functional insights into the canonical mucus-binding protein (MUB), a multi-repeat cell-surface adhesin found in Lactobacillus inhabitants of the GI tract. X-ray crystallography together with small-angle X-ray scattering demonstrated a 'beads on a string' arrangement of repeats, generating 174 nm long protein fibrils, as shown by atomic force microscopy. Each repeat consists of tandemly arranged Ig- and mucin-binding protein (MucBP) modules. The binding of full-length MUB was confined to mucin via multiple interactions involving terminal sialylated mucin glycans. While individual MUB domains showed structural similarity to fimbrial proteins from Gram-positive pathogens, the particular organization of MUB provides a structural explanation for the mechanisms in which lactobacilli have adapted to their host niche by maximizing interactions with the mucus receptors, potentiating the retention of bacteria within the mucus layer. Together, this study reveals functional and structural features which may affect tropism of microbes across mucous and along the GI tract, providing unique insights into the mechanisms adopted by commensals and probiotics to adapt to the mucosal environment.

Introduction

The human gastrointestinal (GI) tract is colonized by a dense and diverse microbial community which is recognized as an important player in normal human physiology, metabolism, nutrition and immune function. Disruption of the gut microbiota (dysbiosis) has been linked with GI conditions such as inflammatory bowel disease (IBD) and obesity (Sommer and Backhed, 2013). In vivo, the intestinal epithelial cells are covered by a mucus layer, which is a biochemically complex medium, rich in glycoproteins (mucins), antimicrobial peptides, immunoglobulins and many other intestinal proteins as well as lipids and electrolytes (Johansson et al., 2011). In the colon, the mucus is organized in an outer loose layer which is home to gut bacteria and an inner layer which is firmly attached to the epithelium and protects the underlying mucosa from the bacterial load (Atuma et al., 2001; Johansson et al., 2008). Mucins are divided into gel-forming mucins (MUC2, MUC5AC, MUC5B and MUC6) and transmembrane mucins (MUC1, MUC3, MUC4, MUC12, MUC13, MUC16 and MUC17) (Johansson et al., 2011). The predominant mucin and major component of the mucus layer in the small intestine and colon is MUC2 (Johansson et al., 2011). In Muc2−/− deficient mice, bacteria come in direct contact with the epithelial cells, penetrate down into the normally sterile crypts and into the epithelial cells, resulting in severe inflammation and eventually colon cancer (Van der Sluis et al., 2006; Johansson et al., 2008), a condition that is comparable to the disease ulcerative colitis in humans. Mucins are rich in Ser/Thr residues that are targets for O-glycosylation primarily with N-acetylgalactosamine (GalNAc), creating the foundation upon which long and more complex oligosaccharide chains are built. The O-glycan chains are extended by the sequential addition of the monosaccharides, N-acetylgalactosamine (GlcNAc), galactose, GalNAc, fucose (Fuc) and sialic acid; the neutral sugars can be further modified by sulfation (Bergstrom and Xia, 2013). The diverse glycan structures along the GI tract are believed to provide binding sites for the gut bacteria which have adapted to the mucosal environment by expressing the right complement of adhesins (Johansson et al., 2008;
Juge, 2012; Ouwerkerk et al., 2013). Alterations in mucus structure, mucin expression and/or mucin glycosylation have been reported in intestinal inflammation and infection in humans and mouse models, and may be a cause or a consequence of an associated change in microbiota composition during disease development (Sheng et al., 2012). This could suggest that the mucin glycans are part of a mechanism for selecting our microbiota. Despite the importance of mucus in the maintenance of a homeostatic relationship with our gut microbiota, very little is known about the nature of the host components and bacterial effectors mediating this interaction.

Gram-positive lactobacilli are normal components of the intestinal microbiota and one of the first groups of bacteria to inhabit the human GI tract (Fan et al., 2013). Some strains (autochthonous) colonize the intestine stably throughout the lifetime of the host, whereas allochthonous strains persist only briefly in the intestine, many of these being probiotics (Reuter, 2001). Adhesion to the intestinal mucosa may prolong their persistence in the GI tract and their beneficial effects to the host, and is thus believed to be a requirement for the compliance of certain probiotic effects, such as immunomodulation and pathogen exclusion (Sanchez et al., 2010; van Baarlen et al., 2013). A growing number of reports indicate that Lactobacillus adherence to the intestinal mucosal layer is mediated by surface proteins with mucus-binding capacity (Juge, 2012; Turpin et al., 2012; Sengupta et al., 2013). Moreover, homology-driven genome mining in several Lactobacillus spp. identified the presence of various-sized putative mucus adhesins consisting of one or more copies of mucus-binding repeats, which is considered a unique functional feature for promoting host–microbe interplay in the GI tract (Boekhorst et al., 2006; Kleerebezem et al., 2010). The best characterized example is the canonical mucus-binding protein, MUB, produced by the Gram-positive bacterium Lactobacillus reuteri (Roos and Jonsson, 2002; MacKenzie et al., 2010). L. reuteri inhabits the GI tract of mammals as diverse as humans, pigs, mice and rats as well as different species of birds and has been used as a model organism to study the evolutionary strategy of a vertebrate gut symbiont (Oh et al., 2010; Frese et al., 2011). MUBs are multi-domain proteins for which the overall architecture is unknown. MUBs exhibit characteristics typical of Gram-positive cell surface proteins; a C-terminal sortase recognition motif (LPXTG) for covalently anchoring the protein to peptidoglycan, repeated sequence elements of 183–206 amino acids in length (Mub repeats), an N-terminal Gram-positive consensus secretion signal upstream of a domain of yet unknown function (Boekhorst et al., 2006). Mub repeats are individually classified as either type 1 or 2 depending on sequence conservation (Roos and Jonsson, 2002; Boekhorst et al., 2006). We recently reported the first X-ray crystal structure of a Mub type 2 repeat [Protein Data Bank (PDB) entry 3I57] (MacKenzie et al., 2009), composed of two discrete domains, an N-terminal B1 domain and a C-terminal B2 domain. The B1 domain displays structural and functional similarity to Protein L from Peptostreptococcus magnus (PpL), a Fab-dependent immunoglobulin-binding protein (Graille et al., 2001), and the B2 domain corresponds to MucBP (mucin-binding protein) in the Pfam database (PF06458). Mub repeats and MucBP domains have been shown to bind to mucins and glycans (Roos and Jonsson, 2002; Pretzer et al., 2005; Bumbaca et al., 2007; von Ossowski et al., 2011; Coic et al., 2012; Watanabe et al., 2012). Numerous MUB homologues and MucBP domain-containing proteins have been found almost exclusively in lactic acid bacteria and predominantly in lactobacilli naturally located in intestinal niches (Boekhorst et al., 2006; Kleerebezem et al., 2010). This suggests that MucBP domain-containing proteins play an important role in establishing host–microbial interactions in the gut and promoting the evolution of the species as primarily GI organisms.

Here, we describe the structure of a type 1 Mub repeat solved by X-ray crystallography and the structural organization of Mub type 1 and type 2 repeats constituting the MUB molecule by small-angle X-ray scattering (SAXS). Full-length MUB was also purified to homogeneity and analysed by atomic force microscopy (AFM) and in cell and tissue-adhesion studies. The results reveal structural and functional commonality between MUB and fimbrial proteins from Gram-positive pathogens with specificity to sialylated mucin glycans. Furthermore, bioinformatics analyses indicated that the MUB family of adhesins is characterized by the type and number of discrete modules, with large numbers of Mub repeats occurring in proteins from probiotic, commensal or otherwise non-pathogenic species, enabling multiple interactions with mucins, consistent with the natural habitat of commensals within the outer mucus layer.

Results

Crystal structure of a Mub type 1 repeat

Lactobacillus reuteri MUB has 14 Mub repeats – six copies (RI through RVI) of a type 1 repeat (Mub1) and eight copies (R1 through R8) of a type 2 repeat (Mub2) (Supporting Information Table S1). These are tandemly arranged, with the Mub2 repeats inserted between Mub1 repeats RIV and RV (Fig. 1A). Mub type 2 repeats are highly similar (sequence identities of 84–100%), with Mub-R1 showing lowest similarity. In contrast, the Mub1 repeats are more diverse with sequence identities ranging from 29% for RIV and RV to 88% for RI and RII (MacKenzie et al., 2009). Here, the Mub-RV crystal structure (PDB entry 4MT5) was solved by molecular replace-
ment and refined at 2.6 Å with an overall crystallographic R-factor (Rcryst) of 21.1% (Rfree 26.9%) (Supporting Information Table S2). Two Mub-RV molecules are present in the crystallographic asymmetric unit and are highly similar having an root-mean-square deviation (rmsd) of 0.9 Å over their Cα-atoms. Mub-RV folds to form an elongated structure 110 Å in length and 24 Å in diameter, comprising an N-terminal B1 domain and a C-terminal B2 domain with a small β-sheet inter-domain region (IR) (Fig. 1B). The structure is similar to that reported for the Mub2 repeat, Mub-R5 (MacKenzie et al., 2009) with a structural alignment Z-score of 15.1 over 176 aligned residues (rmsd 4.1 Å) (Fig. 1C) while sharing 42% sequence identity. The Mub-RV B1 domain has an ubiquitin-like β-grasp fold containing two pairs of antiparallel β-strands in a four-stranded sheet connected by an α-helix, which is similar to that found in members of the immunoglobulin-binding (Ig) superfamily (MacKenzie et al., 2009). However, in contrast to Mub-R5, there was no evidence for a Ca2+ ion coordinated by residues of the loop connecting strands β3 and β4. The Mub-RV repeat lacks the three amino acids (Asp60, Asp62 and Asn65) present in Mub-R5 which mediate ion binding. The Mub-RV B2 domain is classed as MucBP in the Pfam database, with a modified ubiquitin-like β-grasp fold, in which the outer strands of the four-stranded β-sheet are connected by a β-strand (β3′) instead of an α-helix as in the B1 domain. This connecting strand, together with an additional β-strand (β5′) located between β4′ and β6′, forms a third antiparallel β-sheet (Fig. 1B). The B2 domain of Mub-RV shows high structural homology to the MucBP domain of the Gram-positive cell surface protein Spr1345 of Streptococcus pneumoniae (PDB entry 3NZ3) (Du et al., 2011) with a structural alignment Z-score of 11.6 at 37% sequence identity (rmsd 1.4 Å). The MucBP domain of Spr1345 has been shown to be essential for mucin-binding (Du et al., 2011).

**Structural organization of Mub type 1 and type 2 repeats**

SAXS analysis of single Mub repeat proteins (Mub-R5, Mub-RV and Mub-RI), multiple Mub repeats of one Mub type or of mixed repeats (Mub-R8-V, Mub-RV-VI and Mub-RI-II-III), and the MUB N-terminal domain (Nterm) was carried out to gain insights into MUB architecture, tandem-repeat organization and behaviour in solution. The composition of secondary structural elements (1% α-helix, 64–71% β-sheet and β-turn) as determined by far UV circular dichroism was shared by all recombinant Mub repeats (Supporting Information Table S2), suggesting a similar fold of individual Mub domains present in single or multiple repeat proteins. In contrast, the MUB Nterm domain showed a distinct secondary structure composition (6% α-helix, 56% β-sheet and β-turn), likely indicating an alternative protein fold (Supporting Information Fig. S1). SASS profiles of single Mub repeats (Supporting Information Fig. S2A), double and triple repeats (Fig. 2A and C) as well as the Nterm domain (Fig. 2C) revealed these proteins to be monomeric in solution (Supporting Information Table S2). The pair distribution P(r) curves of all Mub repeats (Supporting Information Fig. S2B, Fig. 2B
Fig. 2. (A–H) SAXS analysis and modelling of multiple Mub-repeat proteins and the Nterm domain. (A + C) Scattering profiles of Mub-R8-V (green), Mub-RV-VI (dark blue), Mub-RI-II-III (light blue) and Nterm (red) with the scattering intensity I (black dots) on a logarithmic scale plotted against the reverse momentum transfer s are represented offset for better visualisation. Fits of the reconstructed averaged models for Mub-R8-V, Mub-RV-VI and Nterm generated by GASPOR, and Mub-RI-II-III by DAMMIF, are overlaying the experimental scattering curves. (B + D) Pair distribution functions P(r) were generated from the experimental scattering using GNOM. High-resolution structures of Mub-RV (light blue) and Mub-R5 (green) were manually docked into the GASPOR low-resolution shape reconstructions of Mub-R8-V (E), Mub-RV-VI (F) and Mub-RI-II-III (G) and refined using Sculptor and SITUS. (H) Reconstructed protein envelope for the Nterm domain. (I–J) Atomic force microscopy (AFM) measurements of native full-length MUB. (I) Representative AFM image of native full-length MUB deposited onto mica from ultrapure water and (J) distribution of observed MUB length over all collected images.
and D) showed a single peak with an elongated tail, which is characteristic of extended rod-shaped proteins (Mertens and Svergun, 2010). The maximal particle diameters (D_{max}) for the single repeats Mub-RV, -RI and -R5, were found to be 110 Å, 106 Å and 105 Å (Supporting Information Table S3), respectively, and in good agreement with Mub-RV and -R5 X-ray crystal structures. The D_{max} values for the double repeats Mub-R8-V and Mub-RV-VI were calculated to be 205 Å and 206 Å, respectively, whereas the triple Mub-RI-II-III protein showed a maximal particle diameter of 292 Å (Supporting Information Table S3). Solution envelopes were reconstructed from P(r) functions (Supporting Information Fig. S2C–E; Fig. 2E–G), revealing an extended boomerang-like solution structure of Mub repeats. A good fit was observed with Mub-RV and -R5 crystal structures manually docked and refined into these envelopes (Supporting Information Table S3). Taken together, these results indicate an elongated conformation of single and multiple Mub-repeat proteins in solution and suggest an arrangement of Mub repeats akin to ‘beads on a string’ within the full-length MUB surface protein. In contrast, the D_{max} for the Nterm protein was found to be 159 Å, significantly lower than that of Mub-RI-II-III (a protein of similar molecular weight), suggesting a different solution structure for the Nterm domain compared to the Mub repeats. The solution envelope of Nterm demonstrated a more globular protein shape with an elongated tail (Fig. 2H), which may be indicative of a different function.

Native full-length MUB forms elongated fibre-like structures in solution

MUB is normally covalently attached to peptidoglycan in the bacterial cell wall by a sortase-anchoring motif but can be shed into the culture medium in vitro (MacKenzie et al., 2010). In order to gain further insights into the overall structure of the full-length MUB protein (353 kDa), native MUB was purified to homogeneity from the spent culture medium of L. reuteri ATCC 53608 by size-exclusion chromatography (Supporting Information Fig. S3). Analytical ultracentrifugation sedimentation velocity experiments (data not shown) showed a sedimentation coefficient 5.7S for native MUB; such low sedimentation values are frequently observed for elongated proteins (Erickson, 2009; Kollman et al., 2009). Images of native MUB were acquired by AFM, revealing the presence of monomeric extended fibres (Fig. 2I). A total of 80 MUB molecules were observed, and the measured contour lengths exhibit a distribution of 120 to 192 nm (Fig. 2J). Histogramic analysis of the length data yields a modal value of 174 nm and a mean value of 163 ± 15.5 nm. These values are in good agreement with the sum of a total of 14 Mub repeats (each of about 11 nm in length as observed in X-ray crystal structures and SAXS studies) plus the N-terminal domain of MUB with a D_{max} of approximately 16 nm preceding the first Mub repeat.

Mub repeats share structural similarity with Gram-positive pathogenic adhesins

As with the corresponding domain of the type 2 repeat, Mub-R5, the B1 domain of the type 1 repeat, Mub-RV, resembles the immunoglobulin (Ig)-binding Protein L (PpL) (Z-score 5.4, rmsd 2.9 Å, 15% sequence identity). The B2 domain is structurally similar to the B-repeat of the Listeria invasion protein Internalin B (InIB) (PDB entry 2Y5P) (Z-score 5.9, rmsd 2.6 Å, 13% sequence identity), a protein with as yet unknown receptor specificity (Ebbes et al., 2011). More remarkable is the structural similarity of B2 to a number of pilins such as GBS52 of Streptococcus agalactiae (PDB entry 3PHS) (Z-score 5.4), BcpA of Bacillus cereus (PDB entry 3KPT) (Z-score 4.7), RrgB of S. pneumoniae (PDB entry 3RPK) (Z-score 4.6), Spy0128 of Streptococcus pyogenes (PDB entry 3B2M) (Z-score 4.2) and SpaA of Corynebacterium diphtheriae (PDB entry 3HR6) (Z-score 4.0) as well as the microbial surface components recognizing adhesive matrix molecules (MSCRAMM) CnaB of Staphylococcus aureus (PDB entry 1D2P) (Z-score 2.5) (Deivanayagam et al., 2000; Kang et al., 2007; 2009; Krishnan et al., 2007; Budzik et al., 2009; Paterson and Baker, 2011). These proteins contain a similar domain organization of two to four IgG-like domains of at least seven β-strands with a maximum dimension between 85 Å and 134 Å. The pilin GBS52 is the closest structural homologue of Mub-RV and folds into two IgG-like domains, N1 and N2 (Krishnan et al., 2007). The GBS52 N2 domain shows a structural alignment Z-score of 5.4 and an rmsd of 2.4 Å over 64 aligned Ca-atoms (Fig. 3A). The IgG-like domains of GBS52, BcpA and SpaA share an IgG-rev fold with two sheets of three and four β-strands in a DAG-CBEF topology, while RrgB and Spy0128 display the closely related canonical DAG-CBE topology first observed in the structure of the CnaB D1 domain (Bork et al., 1994; Deivanayagam et al., 2000). The core of the β-sandwich of the Mub-RV B2 domain comprises two sheets, three-stranded and two-stranded, with a modified IgG-rev fold generated from that observed in GBS52 by a simple deletion of the C and B strands from the edge of the larger four-stranded sheet of the latter (Fig. 3B and C). Hence, the conserved strands from the IgG-rev fold observed in GBS52, BcpA and SpaA pilins are A,D,E,F and G. The question of whether this structural homology has arisen by a process of convergent or divergent evolution is an intriguing one. It is interesting to note in this respect that the GY sequence motif noted as being conserved in the
InlB B-repeat, Mub repeats and in other mucin-binding MucBP domains (Ebbes et al., 2011) is also conserved in the GBS52 (Krishnan et al., 2007) and BcpA pilin proteins (Budzik et al., 2009). This motif may play an important structural role as the side chain of the tyrosine (Y) residue of the motif forms a cap for the hydrophobic core of the domain, stabilizing its fold.

L. reuteri native MUB binds to mucus via terminal sialylated structures

Mub repeats and MucBP domains have been shown to bind to mucins and Igs in vitro (Roos and Jonsson, 2002; Pretzer et al., 2005; Bumbaca et al., 2007; MacKenzie et al., 2009; von Ossowski et al., 2011; Coic et al., 2012; Watanabe et al., 2012). To gain insights into the binding of full-length MUB to mucus, adhesion assays were carried out using a human mucus-secreting intestinal cell line and mammalian tissue sections. Purified full-length MUB bound to mouse gastric tissue with staining localized above the gut epithelium (Fig. 4A). A reduction of MUB adhesion was observed after tissue desialylation with sodium periodate at pH 5.5 [as confirmed by reduction of sialic acid-binding lectin *Sambucus nigra* (SNA-I) staining] (Fig. 4A). Upon a more drastic treatment at pH 4.5 leading to a reduction in general glycosylation [as confirmed by staining with Fuc-binding lectin, peanut agglutinin (PNA) and GlcNAc-binding lectin, wheat germ agglutinin (WGA)], MUB binding appeared to be redistributed within the intestinal crypts (Fig. 4A). In order to further explore the binding specificity of full-length MUB to mucus glycan epitopes, we tested binding of native MUB to HT29-MTX cell monolayers, which secrete both MUC5AC and MUC2 (Lesuffleur et al., 1993). MUB adhered to mucus droplets overlying the HT29-MTX cell monolayer, as confirmed by MUC5AC co-staining (Supporting Information Fig. S4). Addition of benzyl-N-acetyl-α-galactosaminide (benzyl-α-GalNAc), an inhibitor of sialylation of glycoproteins and mucin secretion (Delannoy et al., 1996; Prescher and Bertozzi, 2006), caused a reduction of MUC5AC immunostaining and mucus sialylation (SNA-I staining) but not of general glycosylation (WGA and PNA staining) (Fig. 4B). There was a marked reduction in MUB adhesion to benzyl-α-GalNAc treated HT29-MTX cells after 24 h, which was reversible after a further 24 h (Fig. 4B). These functional
studies indicate that terminal sialylated glycan structures in mucus contribute to the binding of MUB to mucus above the epithelium surface.

**Bioinformatics analysis supports a commensal-specific role for many multi-Mub proteins**

These novel structural and functional results prompted us to reanalyse MUB proteins across bacterial species, based on the precisely defined boundaries of the regions seen in the crystal structures of type 1 (this study) and type 2 (PDB entry 3I57) repeats. Profile Hidden Markov Models for each of the four domains (B1 and B2 of each of types 1 and 2) were used to query the full complement of predicted proteins from bacterial genomes (ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria). Ninety per cent of the positive sequences were from *Lactobacillus* or *Streptococcus* (136 and 38 proteins respectively). The type 1 B1 and type 2 B2 domains are specific to *Lactobacillus*, with a single exception (type 1 B1) in *Streptococcus mitis* B6 (Denapaite et al., 2010). The type 2 B1 and type 1 B2 are more widespread and numerous.

---

**Fig. 4.** (A) MUB binding to gastric tissue sections following sodium periodate treatment. Mouse gastric tissue sections (n = 3) were treated with sodium periodate (SP) at pH 5.5 or 4.5. Tissue was incubated with MUB for 2 h, and stained with anti-Mub-R5 and anti-rabbit Alexa Flou 488 secondary antibody (top panels). Tissue was stained with SNA-I-fluorescein isothiocyanate (FITC) (middle panels) and a mixture of PNA-rhodamine (Rh) and WGA-Rh lectins (lower panels). All sections were counterstained with DAPI. (B) MUB binding to HT29-MTX cell monolayers following benzyl-α-GalNAc treatment. HT29-MTX monolayers (n = 9) were incubated with 5 mM benzyl-α-GalNAc for 24 h and stained with anti-MUC5AC (top left panel), SNA-I-FITC (top right panel) or PNA-Rh and WGA-Rh (bottom left panel). HT29-MTX monolayers incubated with MUB for 2 h, were stained with rabbit anti-Mub-R5 and goat anti-rabbit Alexa Flou 488 (bottom right panel). Staining was repeated at 48 h (24 h treatment followed by 24 h culture in normal culture medium). Magnification ×400, scale bars 50 μm; magnification ×100, scale bars 200 μm (B top right panel). PBS, control; pH 5.5, SP treatment pH 5.5; pH 4.5, SP treatment pH 4.5.
Fig. 5. Numbers of each type of Mub domain. Each protein domain was classified as a single unique type. For each genus, the total numbers of Mub-positive species (s), strains (t) and proteins (p) are indicated; species and strain counts are shown as fractions of the total numbers of species and strains in the genomes database. Circles: light grey (B1 domains) and dark grey (B2 domains) segments show numbers of predicted Type 1 (top), Type 2 (middle) and ambiguous (bottom). Black: total Mub domains. White: maximum numbers of Mub domains in one protein.

(Fig. 5), and often occur together in alternating repeats; both types of B2 also occur in tandem self-repeats (Supporting Information Table S4). The number of Mub domains per protein ranges widely from 1 to 28, with the most common frequencies being 1–4. Mub-containing proteins among the non-opportunistic pathogens are generally limited to only one or two domains (usually type 1 B2). A sole exception is a Streptococcus suis D12 protein (Zhang et al., 2011) containing two type 1 B2 and five type 2 B1 domains, although these are not in a close tandem arrangement. Supporting Information Table S5 lists the proteins predicted to contain at least eight Mub domains, constituting the top 15% Mub-abundant proteins. Notably, the multi-Mub Streptococcus proteins occur only in S. mitis, Streptococcus parasanguinis and Streptococcus salivarius, inhabiting the human oral cavity and/or upper respiratory tract as commensals. In contrast the incidence in S. pneumoniae is limited to a single domain (type 1 B2) in a single protein in each of 17 positive strains. In Listeria, all proteins were found to be negative for the specific Mub domains, apart from two Listeria monocytogenes strains (Holch et al., 2013) where Internalin I contains a single domain (MucBP). Only one quarter of all Mub-containing proteins, and 13 of the 38 multi-Mub proteins (Supporting Information Table S5) were predicted positive for both secretory signal and anchor motif. The signal/anchor-
positive group contains more Mub domains per MUB protein (mean 6.3) than the signal/anchor-negative group (3.3), consistent with the notion that Mub domains functioning in an extracellular binding role are present in multiple copies, while single or oligo-Mub repeats may be associated with other functions.

Discussion

Adhesion to host tissues is a necessary first step of bacterial colonization and is mediated by cell surface adhesion proteins. Typical adhesins such as pili and fimbriae or other cell surface proteins have been extensively studied in enteropathogens (Kline et al., 2009). In the past decade, structural analysis using X-ray crystallography has enhanced our understanding of the interactions between MSCRAMM and the host extracellular matrix proteins, i.e. collagen, fibrogen and fibronectin (Joh et al., 1999) by revealing several novel structural features that dictate surface protein assembly and the mode of their adhesion to host tissue (Vengadesan and Narayana, 2011). Although the adhesive abilities of enteropathogenic bacteria have been extensively studied, the systems responsible for intestinal adhesion of gut commensals and probiotic bacteria to mucus are poorly understood.

High-throughput sequencing of Lactobacillus genomes provides a platform for functional analysis of genes that may contribute to adhesion and probiotic function (Pridmore et al., 2004; Pfeifer and Klaenhammer, 2007; Ventura et al., 2009; Turpin et al., 2012; Meyrand et al., 2013). A growing number of reports indicate that LPXTG-anchored cell wall proteins with modular domain organization, such as L. reuteri MUB, function in the adherence of probiotic strains to the host intestinal mucosa (Kleerebezem et al., 2010; Walter et al., 2011; Call and Klaenhammer, 2013). In addition, mucus-binding pili have recently been reported in Lactobacillus rhamnosus strains (Douillard et al., 2013). The mucus-binding ability of L. rhamnosus GG is conferred by the pilus-associated SpaC, present along the length of the pilus shaft potentially resulting in the prolonged residency of L. rhamnosus GG in the GI tract (Kankainen et al., 2009; Reunanen et al., 2012). More recently pili have also been implicated in GI colonization and persistence of Bifidobacteria in the murine gut (O’Connell Motherway et al., 2011; Turroni et al., 2013) and in bacterial autoaggregation and biofilm formation of Lactococcus lactis in vitro (Oxaran et al., 2012; Meyrand et al., 2013). These appendages that decorate the bacterial surface are increasingly considered as key molecules in mediating bacterial adherence to the host epithelium and in influencing mucosal immune responses. Unfortunately, no structural information is available on commensal pilins. Filamentous pili structures are more common and better studied in pathogenic bacteria, and can reach lengths of 70–200 or 300–3000 nm depending on pilus type (Telford et al., 2006).

Our study provides the first insights to the architecture of an exemplar of Gram-positive commensal adhesins. Following an N-terminal capping domain, MUB from L. reuteri has 14 tandemly arranged Mub repeats and an LPXTG cell-wall anchor motif. We showed that L. reuteri MUB forms 174 nm extended protein fibrils, potentially enhancing ligand-accessibility to mucus. A similar protein shape was observed for the Gram-positive serine-rich pilus of S. parasanguinis with an estimated length of several hundred nanometres by electron microscopy (Ramboarina et al., 2010). Our AFM analysis of MUB architecture was supported by SAXS experiments, providing structural information on fragments consisting of single, tandem and triple Mub repeats in solution. Single-Mub repeats display ‘boomerang-shaped’ solution envelopes about 105 to 110 Å (~10 nm) in length, fitted well by the elongated Mub type 1 and the type 2 X-ray crystal structures. In addition, the low resolution X-ray scattering envelope reconstruction of double and triple Mub repeats gave elongated protein shapes of around 20 and 30 nm in length respectively. These findings suggest a ‘beads on a string’ arrangement of Mub repeats each consisting of two (B1 and B2) domains capped by the more globular N-terminal domain. The close structural similarity between the B1 and B2 domains and their co-occurrence in multiple proteins in many species, as indicated by genome-wide sequence analysis, suggest that these domains may have evolved from a common ancestor by duplication. Our crystal structure data suggest only limited flexibility between the B1 and B2 domains about the inter-domain region, and the SAXS data are consistent with only limited flexibility about the B2-B1 domain junctions. Taken together, the picture that evolves of MUB modular organization and dynamics is then of a relatively rigid chain of tubular Mub repeats protruding from the bacterial cell wall terminated by a more globular N-terminal domain. The close structural similarity between the B1 and B2 domains and their co-occurrence in multiple proteins in many species, as indicated by genome-wide sequence analysis, suggest that these domains may have evolved from a common ancestor by duplication. Our crystal structure data suggest only limited flexibility between the B1 and B2 domains about the inter-domain region, and the SAXS data are consistent with only limited flexibility about the B2-B1 domain junctions. Taken together, the picture that evolves of MUB modular organization and dynamics is then of a relatively rigid chain of tubular Mub repeats protruding from the bacterial cell wall terminated by a more globular N-terminal capping domain with B1 (Ig binding) and B2 (MucBP: mucus binding) domains having evolved to fulfill different functions. The function of the MUB N-terminal domain, located at the tip of the protein distal from the bacterial membrane, has yet to be established.

Recent structural studies of the component pilins from pathogenic bacteria have revealed a common pattern of tandem immunoglobulin (Ig)-like domains, joined end-on-end (Kang and Baker, 2012). Here, we described extensive structural similarity between the newly determined crystal structure of the Mub-RV B2 domain and a number of pilin and MSCRAMM from pathogens displaying the IgG-rev fold. Those structures include the CnaB structures of S. aureus (Deivanayagam et al., 2000), the N2 domain from GBSS2 of S. agalactiae (Krishnan et al., 2007), the C-terminal D4-domain of RrgB.
of *S. pneumoniae* (Paterson and Baker, 2011), the major Spy0128 pilin of *S. pyogenes* (Kang *et al.*, 2007) and the SpaA shaft pilin of *C. diphtheriae* (Kang *et al.*, 2009).

Hence, Mub-RV shows structural similarity to members of the four major groups of invasive Gram-positive pathogens. This close structural similarity suggests an evolutionary relatedness. However, whereas in pathogens the primary site of adhesion is usually associated at the tip of the shaft with other domains providing a scaffold for optimum presentation of the glycan recognition site (Imberty and Varrot, 2008; Juge, 2012), the picture emerging from the structural characterization of MUB is one where multiple binding domains are presented along the length of the protein (Fig. 6). A similar role has been suggested for *L. rhamnosus* GC SpaC pilin (Reunanen *et al.*, 2012) and, although no structural information is available for this protein, sequence analysis suggests the presence of multiple IgG-like domains.

Bioinformatics analyses indicated that the MUB family of adhesins is characterized by the type, number and permutation of discrete modules, with extracellular B2 domain-containing multi-Mub repeats associated with commensals, whereas proteins from pathogens were more frequently limited to single or double copies only. This structural arrangement is consistent with the niche adaptation of bacteria to the gut. Commensal bacteria are believed to inhabit and be retained within the mucus layer through interactions with mucins and, in normal physiological conditions, do not reach the epithelium (Johansson *et al.*, 2011; Moran *et al.*, 2011; Ouwerkerk *et al.*, 2013). The presence of multiple Mub repeats and their distribution along the length of the fibre may be necessary to ensure binding to these complex and highly glycosylated proteins by enhancing multivalency to mucin glycans (Fig. 6). Possible functions of Mub repeats other than glycan binding may include their involvement in cis or trans cross-linking with other MUB proteins protruding from the bacterial cell surface, as suggested by autoaggregation phenotypes of MUB-expressing strains (MacKenzie *et al.*, 2010). Additionally, interaction with glycan-independent ligands, in particular proteins (such as IgA) present in large amounts in mucus (Rodriguez-Pineiro *et al.*, 2013), is suggested by adhesion assays (MacKenzie *et al.*, 2009). These multiple interactions will potentiate the retention of commensals within the outer mucus layer (Fig. 6). Here, adhesion assays of native MUB to mammalian tissue sections or epithelial cell cultures showed evidence for binding to terminal sialylated mucin structures within mucus, whereas alteration of glycosylation resulted in MUB penetrating further down towards the epithelium and into the intestinal crypts. Sialic acid is highly abundant in mucus (Hamer *et al.*, 2009) and displays regio-specific distribution along the mammalian GI tract (Robbe *et al.*, 2003; Holmen-Larsson *et al.*, 2013). In addition to playing a role in the tropism of bacteria across mucus, these interactions may also explain the regio-specific location of bacteria along the human GI tract. A different situation occurs with enteric pathogens, which have evolved strategies to circumvent the mucus barrier, through flagella-mediated motility or through

---

**Fig. 6.** Schematic representation of MUB binding to mucus. MUB forms fibre-like structures with multiple binding repeats distributed along the fibre potentiating multiple interactions with terminal sialic acid structures and mucus receptors including mucin glycans and proteins. These multiple and intimate associations will enhance retention of the bacteria within the outer mucus layer. (Not drawn to scale.)
enzymatic degradation of the mucus, and can reach the epithelium surface where they bind to specific oligosaccharide structures via individual lectins or lectin N-terminal domains (Imberty and Varrot, 2008; McGuckin et al., 2011; Juge, 2012).

The structural and functional analysis reported herein provides new insights into how lactobacilli adapt to the mucosal environment in the human GI tract thus to exert beneficial effects. The adhesion mechanisms adopted by Gram-positive bacteria revealed intriguing structural similarities and singularities between commensal and pathogenic bacteria. These findings are important to refine targets and probiotic strategies for inhibiting interactions of pathogenic bacteria with the host.

Experimental procedures

Expression and purification of Mub repeats

The Mub-repeat genes coding for the Mub-R5, -RI, -RV, -R8-V, -RV-VI and –RI-II-III proteins or the N-terminal domain protein (Nterm) of the mucus-binding protein of *Lactobacillus reuteri* ATCC 53608 (American Type Culture Collection) were cloned into the pETBlue-1 Accpetor vector or the pOPINF vector and expressed in *Escherichia coli*. Recombinant Mub-repeat proteins were purified to homogeneity via ion-exchange chromatography followed by size exclusion chromatography (SEC). The Nterm protein was purified via immobilized metal ion affinity chromatography followed by SEC. Proteins were dialysed in 10 mM Tris-HCl (pH 7.5) or 10 mM sodium phosphate buffer (Mub-RV for crystallization purposes, Mub-RI-II-III) and concentrated to about 10 mg ml$^{-1}$.

Protein crystallization, data collection and structure determination

Purified Mub-RV was crystallized at a protein concentration of 10 mg ml$^{-1}$ by sitting drop vapour diffusion at 16°C with a precipitant solution of 0.2 M ammonium acetate and 24% (w/v) PEG 3350. Crystals were cryo-protected by adding 25% precipitant solution of 0.2 M ammonium acetate and 24% Coot (Emsley and Cowtan, 2004) model building and was generated after several alternating rounds of manual the N-terminal domain structure of Mub-R5 and the mucin-Molecular Replacement (autoMR) and autobuilding using and refinement. An initial model was obtained via automated PHENIX program suit was used for phasing, model building performed by the Xia2 automated data reduction system. The integration and reduction of X-ray diffraction data was calculated by Guinier approximation with R,$s$ < 0.8 for elongated proteins, demonstrating the scattering profiles to be free from aggregation. The distance distribution function [P(r)] was generated by GNOM computing the maximum particle diameter (D$_{max}$) and an R$_g$ value calculated for the whole scattering range. Ten ab initio shapes were reconstructed for Mub-R5, -RI, -RV, -R8-V, -RV-VI and Nterm by GASPOR (Svergun et al., 2001) or for Mub-RI-II-III by DAMMIF and averaged by the DAMAVER program package (Volkov and Svergun, 2003), generating a X (Chi) value, a measure for the fit of the experimental data and shape reconstruction, and a normalized spatial discrepancy, a measure for the agreement between computed shape models. Manual docking of the high-resolution X-ray structures of Mub-R5 (PDB entry ID 3I57) and Mub-RV (PDB entry 4MT5), representatives for Mub2 and Mub1 repeats, into the low-resolution reconstructions was performed using SCULPTOR and SITUS (http://situs.biomachina.org/). The refinement of the docking solution by SITUS calculated a cross correlation coefficient R allowing quantitative evaluation of volumetric map and docked structure. The solution scattering of Mub-R5 and Mub-RV CRYSOL were computed from their atomic structures and fitted to the collected experimental scattering curves using CRYSOL (Svergun et al., 1995). The molecular weight of the particles was calculated from the scattering intensity value at zero angle I(0) after scaling against protein concentration with bovine serum albumin (BSA) as standard, providing information on the oligomeric state of the proteins in solution.

Native MUB purification

The full-length native MUB was purified from spent media of a *L. reuteri* ATCC 53608 culture grown in LDMII [Lactobacillus defined media type II, (Roos and Jonsson, 2002)] with 2% (w/v) sucrose at 37°C for 24 h. Bacterial cells were discarded after centrifugation at 7500 × g (15 min, 4°C) and spent media further clarified by vacuum-filtration using 0.45 μm and 0.2 μm filters. The clarified media extract was concentrated by tangential flow filtration using Vivaflow 200 cassette (100 000 MWCO) (Vivascience AG, Hanover, Germany). The sample solution was then dialysed in 10 mM phosphate buffered saline (PBS) (pH 7.4), filtered using...
HT29-MTX monolayers or tissue sections were fixed with 4% paraformaldehyde (PFA) for 10 min and blocked for unspecific binding. Primary antibody (MUC5AC 1:100, Santa Cruz Biotechnology) was applied overnight at 4°C. Secondary antibody (goat anti-rabbit Alexa Fluor 488, Invitrogen) was applied (1:200) for 1 h at room temperature. Lectins (PNA 1:250, Vector Laboratories; WGA 1:250, Vector Laboratories; SNA 1:30, Vector Laboratories) were applied for 2 h at room temperature. Nuclei were counterstained with DAPI (1:200, Invitrogen). Sections and coverslips were mounted in Hydromount mounting medium (National Diagnostics, UK).

MUB adhesion assay
HT29-MTX monolayers or tissue sections were incubated with 4 μg ml⁻¹ MUB diluted in PBS for 2 h at room temperature or overnight at 4°C. Monolayers or tissues were incubated with antiserum of rabbit anti-Mub-R5 (1:100) overnight at 4°C or for 3 h at room temperature respectively. Secondary antibody (1:200 goat anti-rabbit Alexa Fluor 488, Invitrogen) was applied for 1 h at room temperature. Coverslips were washed three times in PBS and mounted in Hydromount mounting medium (National Diagnostics, UK).

Bioinformatics analyses
Four seed alignments of each of the four domains (type 1 B1, type 1 B2, type 2 B1, type 2 B2) were used as queries of the UniProt database (UniProt, 2013) using PSI-BLAST (Altschul et al., 1997). Each of the resulting four sets of hit domain sequences was then used to build a profile HMM using HMMER3 (http://hmmer.org). Each profile HMM, along with a fifth [MucBP from Pfam (Punta et al., 2012)] was used to search all the complete NCBI bacterial genomes (ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/) supplemented by five L. reuteri genomes in the NCBI drafts database (http://www.ncbi.nlm.nih.gov/genome/genomes/438). Processing with SignalP (Petersen et al., 2011) and presence of the Pfam domains YSIRK_signal, Gram_pos_anchor and C-term_anchor were used to assess the presence or absence or secretory signals and anchor domains. Refer to the Supplementary Material for full details.

Acknowledgements
The authors gratefully acknowledge the support of the Biotechnology and Biological Sciences Research Council (BBSRC). This research was partly funded by the BBSRC Institute Strategic Programme BBI/J004529/1: The Gut Health and Food Safety ISP and BBSRC responsive grant.
References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402.

Atuma, C., Strugala, V., Allen, A., and Holm, L. (2001) The adherent gastrointestinal mucus gel layer: thickness and physical state in vivo. *Am J Physiol Gastrointest Liver Physiol* 280: G922–G929.

van Baarlen, P., Wells, J.M., and Kleerebezem, M. (2013) Regulation of intestinal homeostasis and immunity with probiotic lactobacilli. *Trends Immunol* 34: 208–215.

Bergstrom, K.S.B., and Xia, L.J. (2013) Mucin-type O-glycans and their roles in intestinal homeostasis. *Glycobiology* 23: 1026–1037.

Boekhorst, J., Helmer, Q., Kleerebezem, M., and Siezen, R.J. (2010) Microarray analysis of the surface proteome and structural biochemistry (CMSB) for access to facilities. *Bioinformatics* 26: i125–133.

Booher, J.W., and Cliver, D.O. (2000) A method for predicting the complete sequence and function of all genes in a given bacterial genome. *J Bacteriol* 182: 1147–1155.

Bumbaca, D., Littlejohn, J.E., Nayakanti, H., Lucas, A.H., Coic, Y.M., Baleux, F., Poyraz, O., Thibeaux, R., Labruyere, H., Henrich, B., Maurer, P., et al. (2011) Crystal structure of the mucin-binding domain of Spr1345 from *Streptococcus pneumoniae*. *J Struct Biol* 174: 252–257.

Carr, M.C., and Sinko, P.J. (2009) Absorption: what is a commensal? *PLoS ONE* 4: e7176.

Chen, J. (2013) Diversity of the intestinal microbiota in children during the first two years of life. *Environ Microbiol* 15: 3023–3037.

Deivanayagam, C.C., Rich, R.L., Carson, M., Owens, R.T., Dantuluri, S., Bice, T., et al. (2000) Novel fold and assembly of the repetitive B region of the *Staphylococcus aureus* collagen-binding surface protein. *Structure* 8: 67–78.

Delannoy, P., Kim, I., Emery, N., De Bolos, C., Verbert, A., Degand, P., and Huet, G. (1996) Benzyl-N-acetyl-alpha-D-galactosaminide inhibits the sialylation and the secretion of mucins by a mucin secreting HT-29 cell population. *Glycobiology* 13: 717–726.

Denapaite, D., Bruckner, R., Nuhn, M., Reichmann, P., Henrich, B., Maurer, P., et al. (2010) The genome of *Streptococcus mitis* B6 – what is a commensal? *PLoS ONE* 5: e9426.

Dourlaid, F.P., Ribbera, A., Kant, R., Pietila, T.E., Jarvinen, H.M., Messing, M., et al. (2013) Comparative genomic and functional analysis of 100 *Lactobacillus rhamnosus* strains and their comparison with strain GG. *PLoS Genet* 9: e1003683.

Du, Y., He, Y.X., Zhang, Z.Y., Yang, Y.H., Shi, W.W., Frolet, C., et al. (2011) Crystal structure of the mucin-binding domain of Spr1345 from *Streptococcus pneumoniae*. *J Struct Biol* 174: 252–257.

Ebbes, M., Bleymuller, W.M., Cernescu, M., Nolker, R., Brutschy, B., and Niemann, H.H. (2013) Fold and function of the InIB B-repeat. *J Biol Chem* 288: 15496–15506.

Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60: 2126–2132.

Erickson, H.P. (2009) Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biol Proced Online* 11: 32–51.

Fan, W., Luo, G., Li, X., Yang, L., Duan, C., Wang, T., and Chen, J. (2013) Diversity of the intestinal microbiota in different patterns of feeding infants by Illumina high-throughput sequencing. *World J Microbiol Biotechnol* 29: 2365–2372.

Frese, S.A., Benson, A.K., Tannock, G.W., Loach, D.M., Kim, J., Zhang, M., et al. (2011) The evolution of host specialization in the vertebrate gut symbiont *Lactobacillus reuteri*. *PLoS Genet* 7: e1001314.

Graille, M., Stura, E.A., Houssen, N.G., Beckingham, J.A., Bottomley, S.P., Beale, D., et al. (2001) Complex between Peptostreptococcus magnus protein L and a human antibody reveals structural convergence in the interaction modes of Fab binding proteins. *Structure* 9: 679–687.

Hamer, H.M., Jonkers, D.M., Loot, A., Vanhoutvin, S.A., Troost, F.J., Venema, K., et al. (2009) Analyses of human colonic mucus obtained by an in vivo sampling technique. *Dig Liver Dis* 41: 559–564.

Holch, A., Webb, K., Lukjanenko, O., Userry, D., Rosenthal, B.M., and Gram, L. (2013) Genome sequencing identifies two nearly unchanged strains of persistent *Listeria monocytogenes* isolated at two different fish processing plants sampled 6 years apart. *Appl Environ Microbiol* 79: 2944–2951.

Holmen-Larsson, J.M., Thomsson, K.A., Rodriguez-Pineiro, A.O., Imberty, A., and Varrot, A. (2008) Microbial recognition of human cell surface glycoconjugates. *Curr Opin Struct Biol* 18: 567–576.
Hook, M. (1999) Role of fibronectin-binding MSCRAMMs in bacterial adherence and entry into mammalian cells. *Matrix Biol* **18**: 211–223.

Johansson, M.E., Phillipson, M., Petersson, J., Velchich, A., Holm, L., and Hansson, G.C. (2008) The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proc Natl Acad Sci USA* **105**: 15064–15069.

Johansson, M.E., Larsson, J.M., and Hansson, G.C. (2011) The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proc Natl Acad Sci USA* **108** (Suppl. 1): 4659–4665.

Juge, N. (2012) Microbial adhesions to gastrointestinal mucus. *Trends Microbiol* **20**: 30–39.

Kang, H.J., and Baker, E.N. (2012) Structure and assembly of Gram-positive bacterial pilus: unique coherent polymers. *Curr Opin Struct Biol* **22**: 200–207.

Kang, H.J., Coulibaly, F., Clow, F., Proft, T., and Baker, E.N. (2007) Stabilizing isopeptide bonds revealed in Gram-positive bacterial pilus structure. *Science* **318**: 1625–1628.

Kang, H.J., Paterson, N.G., Gaspar, A.H., Ton-That, H., and Baker, E.N. (2009) The Corynebacterium diphtheriae shaft pilin SpaA is built of tandem Ig-like modules with stabilizing isopeptide and disulfide bonds. *Proc Natl Acad Sci USA* **106**: 16967–16971.

Kankanainen, M., Paulin, L., Tynkkynen, S., von Ossowski, I., Reunanen, J., Partanen, P., et al. (2009) Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pilus containing a human-mucus binding protein. *Proc Natl Acad Sci USA* **106**: 17193–17198.

Kleerebezem, M., Hols, P., Bernard, E., Rolain, T., Zhou, M., Kankainen, M., Paulin, L., Tynkkynen, S., von Ossowski, I., Satokari, R., Reunanen, J., Lebeer, S., et al. (2011) Functional characterization of a mucus-specific LPXTG surface adhesin from probiotic *Lactobacillus rhamnosus* GG. *Appl Environ Microbiol* **77**: 4465–4472.

Klopper, J.P., de Vos, W.M., and Belzer, C. (2013) Glycomobiome: bacteria and mucus at the epithelial interface. *Best Pract Res Clin Gastroenterol* **27**: 25–38.

Kolarevic, B., Pretzer, G., Snel, J., Molenaar, D., Wiersma, A., Bron, P.A., et al. (2011) Strain-specific pilus expression is a hallmark of *Boswellia carterii* plants. *PLoS ONE* **6**: e22095.

Klopper, J.P., de Vos, W.M., and Belzer, C. (2013) Glycomobiome: bacteria and mucus at the epithelial interface. *Best Pract Res Clin Gastroenterol* **27**: 25–38.
Semenyuk, A.V., and Svergun, D.I. (1991) Gnom – a program package for small-angle scattering data-processing. J Appl Crystallogr 24: 2337–2344.

Robbe, C., Capon, C., Maes, E., Roussel, M., Zweibaum, A., Zanet, J.P., and Michalski, J.C. (2003) Evidence of a specific glycosylation in human intestinal mucus: presence of an acidic gradient along the intestinal tract. J Biol Chem 278: 46337–46348.

Rodriguez-Pineiro, A.M., Bergstrom, J.H., Ermund, A., Gustafsson, J.K., Schuette, A., Johansson, M.E., and Hansson, G.C. (2011) Studies of mucus in mouse stomach, small intestine, and colon. II. Gastrointestinal mucus proteome reveals Muc2 and Muc5ac accompanied by a set of core proteins. Am J Physiol Gastrointest Liver Physiol 305: G348–G356.

Roos, S., and Jonsson, H. (2002) A high-molecular-mass cell-surface protein from Lactobacillus reuteri 1063 adheres to mucus components. Microbiol-Sgm 148: 433–442.

Sanchez, B., Urdaci, M.C., and Margolles, A. (2010) Extracellular proteins secreted by probiotic bacteria as mediators of effects that promote mucosa-bacteria interactions. Microbiology 156: 3232–3242.

Semenyuk, A.V., and Svergun, D.I. (1991) Gnom – a program package for small-angle scattering data-processing. J Appl Crystallogr 24: 537–540.

Sengupta, R., Altermann, E., Anderson, R.C., McNabb, W.C., Moughan, P.J., and Roy, N.C. (2013) The role of cell surface architecture of lactobacilli in host-microbe interactions in the gastrointestinal tract. Mediators Inflamm 2013: 237921.

Sheng, Y.H., Hasnain, S.Z., Florin, T.H., and McGuckin, M.A. (2012) Mucins in inflammatory bowel diseases and colorectal cancer. J Gastroenterol Hepatol 27: 28–38.

Sommer, F., and Backhed, F. (2013) The gut microbiota – masters of host development and physiology. Nat Rev Microbiol 11: 227–238.

Svergun, D., Barberato, C., and Koch, M.H.J. (1995) CRYSOl – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J Appl Crystallogr 28: 768–773.

Svergun, D.I., Petoukhov, M.V., and Koch, M.H. (2001) Determination of domain structure of proteins from X-ray solution scattering. Biophys J 80: 2946–2953.

Telford, J.L., Barocchi, M.A., Margarit, I., Rappuoli, R., and Grandi, G. (2006) Pili in gram-positive pathogens. Nat Rev Microbiol 4: 509–519.

Turpin, W., Humblot, C., Noordine, M.L., Thomas, M., and Guyot, J.P. (2012) Lactobacillaceae and cell adhesion: genomic and functional screening. PLoS ONE 7: e38034.

Turroni, F., Serafini, F., Foroni, E., Duranti, S., O’Connell Motherway, M., Taverniti, V., et al. (2013) Role of sortase-dependent pili of Bifidobacterium bifidum PRL2010 in modulating bacterium-host interactions. Proc Natl Acad Sci USA 110: 11151–11156.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher’s web-site:

Fig. S1. Secondary structure determination of Mub-proteins by circular dichroism (CD).

(A) Far UV CD spectra of purified Mub-repeat and Nterm proteins collected over scan range 180–260 nm [see also MacKenzie et al., 2009]. (B) The proportion of secondary structural elements was calculated using the CONTIN analysis program, revealing a 64–71% content of β-sheet and β-turns and a low percentage of α-helices for all tested Mub-proteins. Nterm showed an α-helix content of 6.2% and a 56.2% content of β-sheet and β-turns.

Fig. S2. SAXS analysis of single Mub-repeat proteins and solution shape reconstruction. SAXS data for the type 1 Mub-RV (blue) and Mub-R (red) repeats and the type 2 Mub-R5 repeat (green) are presented. (A) The experimental scattering curves are shown as the logarithm of the scattering intensity I (black dots) as a function of the reverse momentum transfer s and presented offset for better visualisation. Overlaying the scattering profiles, are fits of the reconstructed averaged models for Mub-RV, Mub-R and Mub-R5.
generated by GASPOR. (B) Pair distribution functions \( P(r) \) were generated from the experimental scattering using GNOM. Low resolution shape reconstructions of Mub-RV (C), Mub-RI (D) and Mub-R5 (E) were achieved by GASPOR and high resolution structures of Mub-RV (blue) and Mub-R5 (green) were manually docked and refined using Sculptor and SITUS.

**Fig. S3. Purification of native full-length MUB protein.** Native MUB protein was purified from *Lactobacillus reuteri* ATCC 53608 spent media in a multi-step process. (A) In a final size exclusion chromatography (SEC) MUB elutes with the void volume of the column in elution fractions of high protein homogeneity verified by SDS-PAGE gel (B) and specific protein detection via anti-Mub-R5 and –Mub-RI antibodies after Western blotting (C) 1: cell pellet, 2: extracted protein before SEC, 3–11: MUB-containing elution fractions (red box).

**Fig. S4. MUB protein binding to HT29-MTX cell monolayers.** HT29-MTX cell monolayers (\( n = 12 \)) were incubated with MUB for 2 h at 37°C, followed by staining with rabbit anti-Mub-R5 and goat anti-rabbit Alexa Fluor 488 secondary antibody (A). HT29-MTX cell monolayers were stained with polyclonal anti-MUC5AC followed by goat anti-rabbit Alexa Fluor 488 secondary antibody (B). Bright field images are shown adjacent to fluorescent images to demonstrate the correlation of fluorescence staining with mucus droplet structures. Magnification \( \times 400 \); scale bars, 50 \( \mu \)m.

**Table S1. Size of Mub repeats and the N-terminal domain.**

**Table S2. Mub-RV data collection and refinement parameters.**

**Table S3. SAXS data statistics.**

**Table S4. Frequencies of Mub domain pairs in close association.**

**Table S5. Predicted protein sequences with at least 8 Mub domains.**