RTX Adhesins are key bacterial surface megaproteins in the formation of biofilms

Citation for published version (APA):

Guo, S., Vance, T. D. R., Stevens, C. A., Voets, I. K., & Davies, P. L. (2019). RTX Adhesins are key bacterial surface megaproteins in the formation of biofilms. *Trends in Microbiology*, 27(5), 453-467.
<https://doi.org/10.1016/j.tim.2018.12.003>

DOI:

[10.1016/j.tim.2018.12.003](https://doi.org/10.1016/j.tim.2018.12.003)

Document status and date:

Published: 01/05/2019

Document Version:

Author's version before peer-review

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

RTX adhesins are key bacterial surface megaproteins in the formation of biofilms

Shuaiqi Guo¹, Tyler D.R. Vance², Corey A. Stevens², Ilja Voets¹ and Peter L. Davies².

¹From the Institute for Complex Molecular Systems, and the Laboratory of Self-Organizing Soft Matter & the Laboratory of Macromolecular and Organic Chemistry of the Department of Chemical Engineering and Chemistry, Eindhoven University of Technology, Eindhoven, 5600MD the Netherlands

²From the Department of Biomedical and Molecular Sciences, Queen's University, Kingston, Ontario K7L 3N6, Canada

Many Gram-negative bacteria produce Repeats-In-Toxin adhesion proteins (RTX adhesins) to facilitate microbial adhesion to a variety of biotic and abiotic substrates. These large, multi-domain proteins share a common architecture comprised of four regions that have been deduced through a combination of bioinformatics and structure-function studies (Figure 1). First to emerge from the bacterium, C-terminal end leading, is the RTX export sequence that directs the protein through the Type I Secretion System. This is followed by the ligand-binding region responsible for host adhesion and cohesion, which contains diverse ligand-binding domains, both recognized and unrecognized. These serve a zip-code function in directing the host bacterium to a particular environmental niche. Thereafter is a large extension region consisting of tens to hundreds of tandem Bacterial Immunoglobulin-like (BIg) domains, whose function is to extend the reach of the ligand-binding domains away from the bacterial surface. Lastly, there is a conserved N-terminal cell-membrane-anchor region that retains the adhesin within the secretion system. This is also a site of *in situ* proteolysis when nutrients are scarce that enables the bacterium to leave the biofilm and return to a planktonic state in search of a better environment. In this review, the four regions of RTX adhesins will be presented in the order in which they emerge from the cell during synthesis and retention, beginning with the export signal.

Secretion of adhesins and retention in the outer membrane

Bacteria have evolved many different methods for the export of proteins into the extracellular space. While Gram-positive bacteria can use the Sec and Tat pathways to pass proteins through their single-layered membrane, Gram-negative bacteria are presented with the extra hurdle of a dual membrane, necessitating the evolution of a series of complex systems (Type 1 to 6) [7]. Aside from simple secretion, many of these systems are also able to localize proteins – such as adhesins – to the surface of the outer membrane. For most non-fimbrial adhesins (or adhesins made up of a single polypeptide chain) this localization is achieved via one of two methods [8]. The first is the Type 5 secretion system (T5SS), which uses a C-terminal autotransporter domain as a channel that self-inserts into the outer membrane and facilitates the threading of the rest of the protein through the pore. Examples of T5SS adhesins include invasins from *Yersinia ruckeri* and the AIDA-I from *Escherichia coli* [9]. The second – and more recently appreciated – method of non-fimbrial secretion is the Type 1 Secretion System (T1SS), which is used by the exceptionally large, calcium-dependent adhesins of the RTX family (Figure 1) [3,10–14].

RTX proteins and their Type 1 secretion signal

RTX adhesins are a sub-group of the RTX protein family that is mostly comprised of cytolytins and cytotoxins produced by many Gram-negative species as diffusible agents in the internecine struggle for survival that bacteria experience [15,16]. These RTX toxins characteristically contain an N-terminal catalytic domain, typically a lipase or protease, followed by blocks of RTX repeats. While the RTX adhesins have a different domain architecture, the RTX repeats are conserved between both the smaller toxins and the adhesins, giving the family its name. Conventional RTX repeats are tandemly repeated glycine- and aspartate-rich nonapeptides with a consensus sequence of GGxGxDxUx, where x indicates any amino acid and U is typically a large hydrophobic residue. Each pair of RTX repeats forms an 18-residue coil of β -roll structure with Ca^{2+} coordinated on the inside down both sides of the solenoid (Figure 2A&B).

At the C terminus of the RTX protein is a 50- to 100-residue non-cleavable T1SS signal sequence. It has been suggested that the T1SS signal sequences might have more than one function. They can initiate the secretion of substrates by interacting with the T1SS translocase complex, comprised of an inner-membrane ABC transporter and the periplasmic membrane-fusion protein (MFP). This interaction can subsequently trigger the recruitment of the outer membrane protein

(OMP) to form a contiguous tunnel that spans the bacterial envelope [17,18]. The T1SS signal sequence is the first part of the protein to be exported to the extracellular milieu during the translocation process [19], where it can act as the folding nucleus for the rest of the protein by folding in a Ca^{2+} -dependent manner. This would be a second function for the signal sequence. Indeed, deletion of the T1SS signal sequence not only stalls the secretion, but can severely affect the successive folding of other domains of the protein [20].

RTX toxins and RTX adhesins are similar in this export region. For example, following the classic RTX β -roll of the RTX toxin produced by *Bordetella pertussis* (CyaA) is the T1SS signal sequence localized to the C-terminal ~ 75 residues of the toxin [21]. This module is comprised of two pairs of anti-parallel β -strands and two α -helical elements (Figure 2A&2D) that are required for the correct folding and biological activity of CyaA [20]. The T1SS signal module of the toxin also serves as the capping structure for the RTX β -roll [22], which helps stabilize the solenoid by preventing it from forming amyloid-like aggregates [23]. Despite the absence of sequence homology, the T1SS signal sequence of the RTX adhesin, produced by *Marinomonas primoryensis* [1], referred to here as an ice-binding protein (*MpIBP*), also contains a capping moiety with similar anti-parallel β -strands and α -helices (Figure 2B&2E) that is thought to serve the same functions..

Although the RTX β -roll is not required for the substrate translocation by T1SS, it plays a crucial role in enhancing secretion efficiency [24,25]. Given its high content of acidic residues, the RTX module is overall negatively charged in the absence of Ca^{2+} . This might facilitate the extrusion of RTX proteins from the outer opening of T1SS by electrostatic repulsion from the negatively charged lipopolysaccharide [20]. Once the RTX domain enters the extracellular milieu, it folds upon binding Ca^{2+} to form a stable β -roll, which creates steric hindrance to prevent it from sliding back into the cell by Brownian motion.

Variable ligand-binding region determines the biofilm niche and its cohabitants

The ligand-binding region near the C-terminal end of the RTX adhesin is the most varied part. Here the type and arrangement of the domains seem to have been shuffled over evolutionary time, making the adhesins distinct from one another. This is true of the six examples illustrated in

Figure 1. In many cases, the presence and identities of ligand-binding domains and/or their ligands have yet to be determined. As the adoption or incorporation of DNA for different domains can change the targeting of a bacterium, the ligand-binding region is presumably under strong selective pressure. In this way, the general RTX adhesin architecture (the other three regions) can be seen as a common scaffold to house and properly present varying ligand-binding modules, which can be swapped in and out as bacterial populations/species evolve to meet new environmental needs. Presented below are a few of the domains that recur frequently within the ligand-binding regions of RTX adhesins, possibly indicating a shared functionality amongst their bacterial hosts.

PA14 sugar-binding domain

Many RTX adhesins contain the PA14 sugar-binding domain, which might target cells to extracellular polysaccharides in biofilms, to carbohydrate substrates, or to surface glycolipids or glycoproteins on various cell types (Figure 1). Originally discovered as a domain of unknown function in the Protective Antigen (PA) toxin from *Bacillus anthracis*, the domain has been found in proteins from many different organisms, ranging from bacteria and yeast, to more complex eukaryotes like humans [37,38]. PA14's involvement in microorganism adhesion has been well documented in the yeasts *Saccharomyces cerevisiae* and *Candida glabrata*, where its sugar-binding function is used for self-adherence (flocculation) and human cell adhesion, respectively [39–41]. In fact, it is in yeast that the majority of PA14 structural information has been attained, namely in the flocculin proteins Flo1 and Flo5 (PDB: 4GQ7 & 2XJP) and the epithelial adhesins EpA1, EpA6, and EpA9 (PDB: 4ASL, 4COY, & 4CP0). Analysis of these yeast PA14 structures have revealed a highly-conserved double aspartic acid motif, which anchors a critical calcium ion that coordinates carbohydrates via two vicinal, equatorially-oriented hydroxyl groups. This same motif is present in all known RTX adhesin PA14 domains, including those from *Marinobacter hydrocarbonoclasticus*, certain *V. cholerae* strains, and *M. primoryensis*, the last of which is currently the only structurally characterized PA14 within an RTX adhesin [1]. The structure of this domain (PDB: 5J6Y) is shown in magenta in Figure 3 with a close-up view of where glucose is bound.

von Willebrand Factor A domains

Another commonly occurring domain in RTX adhesins (Figure 1) is the von Willebrand Factor A-like (vWFA) domain, a known peptide-binding fold found in many eukaryotic proteins. Several characterized RTX adhesins contain this domain – such as LapA from *P. fluorescens*, and the RtxA protein from *Legionella pneumophila* [42,43]. Although the role of the vWFA in RTX adhesins has yet to be identified, well-characterized vWFA proteins from eukaryotes, such as the titular von Willebrand Factor in blood [44] and the integrin family of cell-to-cell adhesion proteins [45], provide suggestions. In both these eukaryotic multi-domain proteins, the vWFA module encompasses a small portion of the total sequence, forming a Rossman fold (an antiparallel beta-sheet surrounded by alpha-helices) that can bind to extracellular matrix (ECM) proteins like collagen [46–49]. The vWFA domain in bacteria might be used to bind to similar proteins, which would be especially useful for bacterial pathogens that infect animals. Examples of bacterial vWFA domains capable of binding human ECM proteins have been observed [50], as has the conservation of the serine-coordinated calcium ion, known as the Metal-Ion Dependent Adhesion Site (MIDAS), which allows integrins to bind acidic residues on their target proteins [51].

An alternative role for vWFA domains in bacteria is inspired by the von Willebrand Factor protein, and its use as a mechano-sensor in blood clotting. The high shear forces associated with an open, bleeding wound leads the vWFA1 domain of the protein to become slightly unfolded, revealing a binding site for a glycoprotein on the outside of platelets [52–54]. The platelets can thereby cluster around the wound, facilitating clotting. Interestingly, a vWFA domain-containing protein – PilY – from *P. aeruginosa* has been proposed to function similarly, but as a detector for the heightened mechanical shear forces surrounding a surface [55,56]. Pre-exposing *P. aeruginosa* to a surface increases the bacteria's subsequent virulence during amoeba infection. Siryaporn *et al.* showed that activation of this virulent phenotype can be prevented through the deletion of the PilY protein: a surface-exposed protein required for pili biosynthesis that contains an N-terminal vWFA domain [55]. However, deletion of the vWFA, alone, led to a hyper-virulent strain that did not require pre-exposure to surfaces to increase pathogenicity. Taken together, the vWFA domain appears to act as a surface-sensing switch, keeping the virulence phenotype deactivated until a sufficient shear force is sensed in the presence of a surface.

The biophysical mechanisms at play in *Pseudomonas* are not yet explained, nor is it known how widespread this supposed vWFA-dependent strategy of surface-sensing is in bacteria. While

no other examples have been presented so far, *Caulobacter crescentus* relies on the resistance “felt” during pili retraction to sense and irreversibly adhere to a surface [57]. The protein that is predicted to anchor the pili to the peptidoglycan layer (TadG) also contains a predicted vWFA domain [58,59]. While the RTX adhesins have a very different architecture than both the PilY and TadG proteins, this interesting alternative function should be considered in future studies.

RTX-derived ice-binding domain and homologues of unknown function

The feature that led to the discovery of the massive RTX adhesion in the Antarctic bacterium *M. primoryensis* was the presence of a domain that avidly binds ice [60]. This seems to have been evolved from the nearby C-terminal β -roll, possibly after a gene duplication event (Figure 2C). While the RTX sequences in the β -roll aid secretion [20] (Figure 2B), RTX repeats in this larger more proximal domain have evolved the remarkable ability to bind ice, leading to the adhesin’s designation as the *M. primoryensis* ice-binding protein (*MpIBP*). In contrast to the conventional β -roll fold with two flanking rows of Ca^{2+} , each 19-residue coil of the ice-binding domain contains only one RTX repeat, resulting in a β -solenoid structure with 13 Ca^{2+} aligning down only one of its flanks (Figure 2C). The single RTX repeats in these coils have diverged to have a consensus of xGTGND (in contrast to the classic GGxGxD), where the Thr-Gly-Asn/Asp motif helps form a flat surface that runs the length of the β -solenoid [60]. The X-ray crystal structure of this ice-binding domain revealed that the array of outward-pointing Thr organizes cages of water molecules around the methyl groups that are then hydrogen bonded to nearby hydrophilic groups into an ice-like pattern that merges with and freezes to the quasi-liquid layer of water coating the ice lattice [61]. These “anchored clathrate waters” are able to freeze *MpIBP* to several planes of ice.

Homologues of this beta-solenoid ice-binding domain are present in the RTX adhesin of at least one other *Marinomonas* species, *M. ushuaiensis*, but without the regularly spaced Thr and Asn arrays that are essential for ice-binding. The same domain has popped up further afield in a magnetotactic bacterium RTX adhesin, but again without the ice-binding residues (GenBank: BAE50805). Other RTX domains with ligand-binding potential have also been identified. CyaA produced by *B. pertussis* has five distinct blocks of RTX repeats separated by linkers ranging between 23 to 49 residues in length. Two of the central blocks have evolved to be integrin-binding domains [62], while the C-terminal block facilitates the proper secretion and folding of the whole

protein [63]. Intriguingly, using ribosome display, Bulutoglu *et al.* showed the C-terminal block can be turned into a lysozyme-binding module with an apparent dissociation constant of 65 nM [64]. This *in vitro* evolution study further demonstrated the high degree of plasticity of these β -rolls to evolve adhesion capabilities for various ligands found on cells or other surfaces.

Uncharted adhesion domains

Bioinformatic analyses of C-terminal regions for both studied and hypothetical adhesins, reveal large stretches of sequence that do not belong to a known domain family (Figure 1). Interestingly, PSI-BLAST searches of these regions to probe for distantly-related sequences often show definitive breaks in homology, implying the amalgamation of multiple sequences from different sources. Indeed, such breaks fit with the concept of a modular ligand-binding region prone to switching domains as needed. To date, only a few identifiable domains have been structurally and functionally characterized (above). The mystery regions between known domains may be hiding previously unknown structures and binding-partners, and will require study if the role of individual RTX adhesins in the life cycle of their hosts are to be understood.

As an example, a different type of peptide-binding domain was discovered serendipitously when solving the structure of the *M. primoryensis* adhesin (Figure 1). This calcium-dependent beta-sandwich domain (PDB:5K8G) of previously unknown structure and function in the ligand-binding region of the adhesin was observed to have a binding pocket in which the C-terminal peptide of a second molecule in the crystal was bound (Figure 3) [1]. Tests with free peptides in solution have confirmed that this is a true binding interaction rather than a crystallization artifact (Guo *et al.* unpublished). The discovery and characterization of this domain – as can be said for the resolution of any mystery region – is applicable beyond one adhesin, as a domain homologue can be found in the RTX adhesins of several *Vibrio* species [65].

The ultrastructure of an RTX adhesin ligand-binding region

MpIBP is the first RTX adhesin in which the entirety of the ligand-binding region has been both structurally and functionally characterized [1]. In *M. primoryensis*, this ligand binding region serves at least two functions. It binds the bacteria to ice but also to the Antarctic diatom, *Chaetoceros neogracile* [4] (Figure 4). The latter function requires adhesin attachment to terminal sugars and peptide moieties on the surface of the diatom. Thus the sugar, peptide and ice-binding

domains near the C-terminal end of the adhesin facilitate the formation of a mixed species microcolony on the underside of sea or lake ice to the benefit of both contributing species. The diatoms occupy an optimal niche for photosynthesis thanks to the bacteria's ability to bind them and ice at the same time, and are treated to essential nutrients, such as vitamin B12, secreted by the bacteria [5,6]. Likewise, the aerobic bacteria potentially benefit from the oxygen and other waste products of the diatoms.

Six distinct domains are present in this region of *MpIBP*, three of which are ligand binding (Figure 3C). At the proximal end of this grouping are five β -sandwich-like domains (RIII_1 to RIII_5). They are not an extended 'beads-on-a-string' structure, but are instead organized into a T-shape, where structural domains RIII_1 and _2 make up the stalk of the T, while RIII_4 sits at the junction to project the two ligand-binding domains (RIII_3 and _5) away from each other (Figure 3). RIII_3 and _5 form the tips of the T-shape with RIII_3 being the peptide-binding domain, and RIII_5 the PA14 sugar-binding domain, both described earlier. The two adhesion domains require Ca^{2+} to coordinate their respective ligands, a common theme amongst RTX adhesin domains. The stringent way in which these domains are organized in 3-D space suggests the T-shape is designed to extend the ligand –binding domains away from each other to allow contact with one substrate or the other, but not both side by side (Figure 4). This organization is similar at one level to that of the IgG antibody where two antigen binding sites are projected away from each other. It is likely that multi-domain ultrastructures like this will be found in other adhesin ligand-binding regions. It may be necessary to crystalize the entire region as a unit, or, if that is not possible, to amalgamate individual domains into the whole using SAXS envelopes.

The sixth domain of *MpIBP*, the ice-binding, calcium-coordinated beta-solenoid, is closely linked by ~80 residues to the other five domains. It is not clear at this point if the ice-binding domain is part of a higher structure with the other five domains or is just connected to them by a linker. However, the sixth domain is too close to the T-shaped structure to permit simultaneous binding of one adhesin to both ice and a diatom. For this dual binding to occur, *M. primoryensis* must have at the very least two RTX-adhesins expressed as shown in Figure 4, with one directed to the ice and another to the diatom.

Extension Region: tandem B_Ig domains extend the reach of the adhesin

A striking feature of the RTX adhesin domain architecture is the central region, which often accounts for the majority of the proteins' molecular weight. This segment houses varying numbers of tandem repeats that most often fold as a 10- to 15-kDa beta-sandwich, similar to those found in immunoglobulins, which explains their general title of Bacterial Immunoglobulin-like (B_Ig) domains. B_Ig domains are prevalent in many species of Gram-negative bacteria, and are used in RTX adhesins to extend the substrate-binding C-terminal domains into the extracellular space, away from the cluttered cell membrane of the bacterium.

Calcium rigidifies and strengthens the extender region

To maintain effective extension from the cell surface, it stands to reason that extender regions require a relatively rigid, elongated conformation. Several structures of extender region fragments from RTX adhesins have been solved to date, including a construct containing three tandem B_Ig repeats from the SiiE adhesin of *Salmonella enterica* [66], and another containing four tandem B_Ig repeats from *MpIBP*, denoted as the *MpIBP* tetra-tandemer [67]. Both structures revealed the expected beta-sandwich domains to be aligned in extended rods, with calcium ions coordinated in the flexible linkers between domains. These ions were hypothesized to rigidify the tandem repeats, thereby facilitating an extended structure that prevents the adhesin from bending back on itself and interacting with its own cell surface. Indeed, molecular dynamic simulations and electron microscopy of the full-length SiiE showed an extended rod-like structure in solution, which was compromised by a loss of calcium [66]. Similarly, biophysical analysis of the *MpIBP* tetra-tandemer using size-exclusion chromatography, analytical ultracentrifugation (AUC) and small-angle X-ray scattering (SAXS) all demonstrated an apparent increase in rigidity upon calcium binding [67].

Aside from rigidifying the flexible linkers, calcium can also contribute to the strength of the B_Ig domain fold. Most domains in RTX adhesins require calcium to fold properly. Massive changes in secondary structure, from random coil to predominantly beta-strand, can be seen during the titration of *MpIBP* B_Ig domains with calcium monitored by circular dichroism spectroscopy [67,68]. The resulting beta-sandwich domains of Ca²⁺-bound *MpIBP* can withstand forces as high as 350 pN without unfolding, as seen through single-molecule force spectroscopy experiments of

an octa-tandemer of *MpIBP* under high calcium conditions [69]. Such a resistance to unfolding is remarkable, outclassing the folding strength of even the Ig-like domain I27 of the human muscle-protein titin by ~ 120 pN [70,71]. B1g domains from other RTX adhesins seem to share this folding strength, such as *MhLap* from the oil-degrading bacteria *M. hydrocarbonoclasticus* [69]. The structural element proposed to allow these beta-sandwich domains to attain such high stability in the face of intense shear forces is an altered form of the classic mechanical clamp: hydrogen-bonded parallel beta-strands found at the termini of the domain [72,73]. This motif is found in many stably folded beta-sandwich domains, including I27 from titin, and the extender domains of *MpIBP*. The latter structure also shows coordinated calcium ions flanking the clamps that make additional contacts. Further single-molecule force spectroscopy experiments showed that the *MpIBP* octa-tandemer in low calcium conditions unfolds under forces roughly similar to titin I27, supporting the added value of the calcium-mediated shear clamp motif. Additionally, the B1g domains of SiiE, which appear to lack the C-terminal calcium-mediated shear clamp, unfold at a much lower value of ~ 120 pN [66].

Variability amongst adhesin B1g domains

Though the presence and proposed biological role of the extender region appears universal to all RTX adhesins, the number of repeats present can vary five-fold (Figure 1). It is not clear why certain adhesins contain tens of repeats (ex. *MhLap* has ~ 25 repeats), while others have over a hundred (ex. *MpIBP* is estimated to contain ~ 120). Several hypotheses can be put forward to explain the varied lengths of the extension region in RTX adhesins. 1) Different bacteria-substrate interactions may require different degrees of proximity. For instance, the smaller non-RTX adhesion protein intimin maintains few repeat domains to facilitate close contact between cells, which may not be necessary (or advantageous) for adhesion to other substrates [74]. 2) Different bacteria express macromolecules on their surfaces of varying lengths (ex. lipopolysaccharide (LPS), other adhesins, fimbriae etc.), thereby changing the extension required for an adhesin to out-reach potentially interfering surface components. 3) Certain B1g domains may have additional functions beyond extension. The SiiE adhesin B1g domains are capable of adhering to sugar-containing ligands in a lectin-like manner, as shown via lectin blockade experiments [34]. This gives an additional adhesion role to the region, and may partially explain the requirement of many

SiiE repeats for effective invasion. Which (if any) of these hypotheses answers the question of repeat number will require further study.

It is also possible that the vast difference in repeat number is indicative of a lack of evolutionary selection, implying that the exact number of repeats is unimportant to adhesin function and therefore left to the whims of DNA duplication and recombination. That said, it seems unlikely that organisms like *M. primoryensis*, which produce upwards of a hundred repeats, would invest energy in producing a 1.5-MDa protein if it was not beneficial. Indeed, it has been observed that the ability of SiiE – which contains 56 repeats – to facilitate invasion of *S. enterica* into epithelial cells declines if even five repeats are removed [75].

The sequence identity shared between repeats within the same adhesin is almost as variable as the number of repeats, ranging from low values of 40 to 60 % at the amino-acid level (ex. LapF), to 100 %, even at the *nucleotide* level (ex. *MpIBP*, *V. cholerae* str. FORC_055 adhesin GENEbank: WP_096070493). For the adhesins with many nearly-identical repeats, the inability of most DNA sequencing methods to acquire sequencing reads long enough to cover the whole extender region can stymie attempts at assembling the reads into one contiguous sequence. Wrobel *et al.* found that the number of Ig-like repeats present in the invasin-like autotransporter adhesin from *Y. ruckeri* was underestimated by the short sequencing reads of Illumina sequencing, requiring the exceptionally long reads of Pacific Bioscience sequencing to attain a correct number [76]. *MpIBP* serves as a notable case study for this problem, as the total number of identical 312-bp repeats in its long extender region could not be determined, even by PacBio sequencing [1,77]. Interestingly, while the older sequencing methods were simply unable to connect the two ends of *MpIBP*, the genome analysis software used to assemble the PacBio sequencing data of *M. primoryensis* provided a supposed “complete” *MpIBP*, with the N and C termini connected by 75 Blg repeats. However, upon closer examination, this assembly simply used the maximum number of repeats through which the sequencing was able to read, and is therefore not representative of the actual number of Blg domains in *MpIBP*. This false assembly would be hard to detect if one was not looking for it, potentially calling into question other such “completed” adhesin sequences. It is unclear how often difficult cases like the *MpIBP* gene sequence occur; but we have noted examples of sequencing maladies caused by highly-identical sequences in other adhesin genes. The N and C termini of *MpIBP*-homologs from *Marinomonas* species *M. profundimaris* and *M.*

blondensis str. MED121 are split into two contigs, leaving gaps in the genomic sequence [78,79]. Similarly, an RTX adhesin in *Magnetospirillum magneticum* has a sequencing gap in its extender region sequence [80].

The inability to trust the assembled sequence of some adhesin genes may require in-depth scrutiny of an individual extender region. To estimate the number of repeats in the extender region of *MpIBP*, the gene was restriction digested using four-cutter enzymes that cut immediately outside the extender region, but not within. This digest yielded the full-length extender region to be resolved and measured by pulsed-field gel electrophoresis. By frequently changing the direction of the current during electrophoresis, this method enables large DNA fragments to pass through the gel and be accurately sized alongside large DNA standards like the genome of lambda phage. Using pulsed-field gel electrophoresis, *MpIBP* was found to contain ~ 120 repeats [77].

Cell-surface-retention region: a calcium-independent plug anchors the adhesin in the T1SS outer-membrane protein

Cell fractionation analyses suggested that the RTX adhesin LapA produced by *Pseudomonas fluorescens* is retained on the outer membrane [26,27]. However, the mechanism of how these adhesion structures are localized to the cell surface remained a mystery until recently. Given RTX adhesins require T1SS for translocation to the cell surface; Guo *et al.*, reasoned these giant proteins might be anchored there by interacting with a component of the secretion machinery [1].

This hypothesis has been supported by structural studies on *MpIBP*. T1SS secretes its substrates with a C-to-N-terminal directionality [19], thus it is likely the domains responsible for cell surface retention are near the protein's N terminus after most of the adhesin has been extruded into the medium (Figure 4). Using a combination of NMR and small-angle X-ray scattering, Guo *et al.* showed the N-terminal domain of *MpIBP* folds as a stable β -sandwich with a triangular cross-section (30 Å x 29 Å); whereas the adjoining domain on the C-terminal side has a slim cylindrical neck shape with a diameter of roughly 18 Å [1]. This cylinder can snugly fit into the TolC, the outer-membrane β -barrel of the T1SS apparatus with an internal diameter of ~ 20 Å [28,29]. But the β -sandwich plug cannot pass through the TolC pore because of steric hindrance, thereby

blocking the total release of the giant adhesin to the medium (Figure 5). In contrast to all the other >120 domains of *MpIBP* that are unstructured during secretion and require Ca^{2+} for folding, the N-terminal plug and neck domains are folded in the absence of Ca^{2+} before entering the T1SS tunnel. This finding provided the structural basis for what seems to be a general mechanism widely used by Gram-negative bacteria to retain large adhesins on their cell surface [30]. In this paradigm, the many distal parts of the adhesins protrude from the cell surface to interact with the environment, while the N-terminal plug domain is anchored on the periplasmic side of the outer-membrane via the T1SS- β -barrel-spanning module.

This structural model has been experimentally validated by Smith *et al.* who performed *in vivo* secretion competition experiments between the 0.8-MDa RTX adhesin LapA produced by *Pseudomonas fluorescens* and the C-terminal secretion domain of LapA tagged with a three-hemagglutinin epitope (HA-C235) [30,31]. The rationale for this experiment was that if the surface-retention of LapA is achieved by plugging the T1SS duct, the secretion of HA-C235 may be affected as the two substrates share the same secretion machinery. Indeed, when the native LapA was held to the cell surface, the secretion level of HA-C235 was greatly reduced. However, HA-C235 secretion was resumed when the N terminal plug domain of LapA was cleaved off by a periplasmic protease (LapG). This is because proteolysis released LapA to the medium and freed up the T1SS channels for the export of HA-C235. Moreover, it was demonstrated that the secretion of HA-C235 can be restored when the T1SS machinery proteins are overexpressed in *P. fluorescens*, further consolidating the proposed model that the N-terminal domain retains RTX adhesins by plugging and occupying the T1SS channel.

Detailed structural analysis of the plug domain of *MpIBP* revealed that it contains several outward-pointing phenyl-alanine residues, which might directly interact with the outer-membrane to help anchor the adhesin. With relatively low sequence identities (~ 20%), the 3-D fold of the plug domain is conserved in the RTX adhesin from many bacteria, including pathogens such as *V. cholerae*, *P. aeruginosa* and *Shewanella oneidensis* [1,30]. In fact, plug domains from the *V. cholerae* or *P. aeruginosa* adhesins are capable of anchoring chimeric LapAs to the cell surface of *P. fluorescens*. In contrast to the stable β -sandwich plug, its neighboring cylindrical β -barrel-spanning module is flexible and labile to proteolysis [1,30]. Indeed, a specific proteolytic site between the 2nd and 3rd positions of the conserved sequence of Thr(Pro)-Ala-Ala-Gly can be activated to release the bacterium from its biofilm when growth conditions are unfavorable, which

allows it to become planktonic in the search for a better niche [26,32]. Although the above model represents a widely-used mechanism for surface retention, it should be noted that some RTX adhesins have evolved modified surface release strategies. For example, the giant adhesin SiiE from the human pathogen *Salmonella enterica* uses a predicted coiled-coil moiety in place of the β -sandwich plug to transiently hold the protein to its T1SS apparatus [33,34]. With no predicted proteolytic site, it remains unknown what triggers the release of SiiE into the medium. Intriguingly, although the β -sandwich plug is conserved in the non-RTX adhesin of *P. aeruginosa* (CdrA), it lacks the T(P)-A-A-G site for proteolysis. Instead, CdrA is retained by a “cysteine hook” that restricts its secretion through the outer-membrane pore [35,36].

Strategies to block adhesion/infection

Adhesion is a critical step in biofilm formation, which for pathogenic bacteria can lead to infection. Adhesion allows targeting of a given bacterium to a specific surface, and – in flow environments – it enables bacteria to resist physical removal by shear forces. The target specificity of the ligand-binding domains within an RTX-adhesin is instrumental in guiding the bacterium to its optimal environmental niche. Blocking the ligand-binding domains of RTX-adhesins is an attractive strategy to prevent surface adhesion and subsequent biofilm formation. Previously, multiple strategies have been proposed to inhibit adhesion by spoiling ligand-binding domain interactions; these include: antibodies, substrate analogs, metal-ion interference, and surface modification.

By incubating *M. primoryensis* with antibodies raised to the ice-binding domain of *MpIBP*, which makes up only ~2% of the whole adhesin, Bar-Dolev *et al.* were able to completely prevent bacteria from adhering to ice [81]. In principle, antibodies to the key surface-binding domains of other RTX adhesins could similarly prevent colonization by their bacterial hosts. The adhesins must be studied for the structure-function relationships to determine their key domains that could be used as vaccines in the prevention of infection.

Using *MpIBP* as an exemplar, there are additional ligand-binding domains nearby in the hyper-variable ligand-binding region of this RTX-adhesin that contribute to biofilm cohesion. In the case of *MpIBP* these are sugar- and peptide-binding domains that link the bacteria to diatoms

to form a mixed – species biofilm (Figure 4) [1]. It is hypothesized that the two domains bind to as yet unidentified polysaccharides and proteins on the diatom surface. Therefore, it may be possible to prevent adhesion or cohesion by blocking the ligand-binding domain from associating with its appropriate substrate through competition with an analog. A precedent for this approach has been the use of non-metabolisable mannose mimics to prevent the ligand-binding domain of a fimbrial adhesion (FimH) of uropathic *Escherichia coli* (UPEC), from binding urothelial cells leading to a drastic reduction in urinary tract infections [82,83].

Acknowledgements

This work was funded by a Natural Sciences and Engineering Research Council of Canada discovery grant to PLD (RGPIN 2016-04810), and by the European Union through an ERC grant (ERC-2014-StG Contract No. 635928) to I. Voets. P. Davies holds the Canada Research Chair in Protein Engineering, and TDRV is funded by a Canadian Graduate Scholarship from the National Science and Engineering Research Council (NSERC).

References

- 1 Guo, S. *et al.* (2017) Structure of a 1.5-MDa adhesin that binds its Antarctic bacterium to diatoms and ice. *Sci. Adv.* 3, e1701440
- 2 Chou, L.-F. *et al.* (2009) Localization and characterization of VVA0331, a 489-kDa RTX-like protein, in *Vibrio vulnificus* YJ016. *Arch. Microbiol.* 191, 441–450
- 3 Satchell, K.J.F. (2011) Structure and function of MARTX toxins and other large repetitive RTX proteins. *Annu. Rev. Microbiol.* 65, 71–90
- 4 Jung, G. *et al.* (2007) Annotation and expression profile analysis of cDNAs from the Antarctic diatom *Chaetoceros neogracile*. *J. Microbiol. Biotechnol.* 17, 1330–1337
- 5 Amin, S.A. *et al.* (2012) Interactions between Diatoms and Bacteria. *Microbiol. Mol. Biol. Rev. MMBR* 76, 667–684

- 6 Croft, M.T. *et al.* (2005) Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* 438, 90–93
- 7 Green, E.R. and Mecsas, J. (2016) Bacterial Secretion Systems – An overview. *Microbiol. Spectr.* 4, doi: 10.1128/microbiolspec.VMBF-0012-2015
- 8 Gerlach, R.G. and Hensel, M. (2007) Protein secretion systems and adhesins: The molecular armory of Gram-negative pathogens. *Int. J. Med. Microbiol.* 297, 401–415
- 9 Leo, J.C. *et al.* (2012) Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. *Philos. Trans. R. Soc. B Biol. Sci.* 367, 1088–1101
- 10 De Gregorio, E. *et al.* (2015) Biofilm-associated proteins: news from *Acinetobacter*. *BMC Genomics* 16, 933
- 11 Hinsa, S.M. *et al.* (2003) Transition from reversible to irreversible attachment during biofilm formation by *Pseudomonas fluorescens* WCS365 requires an ABC transporter and a large secreted protein. *Mol. Microbiol.* 49, 905–918
- 12 Martínez-Gil, M. *et al.* (2012) Calcium Causes Multimerization of the Large Adhesin LapF and Modulates Biofilm Formation by *Pseudomonas putida*. *J. Bacteriol.* 194, 6782–6789
- 13 Martínez-Gil, M. *et al.* (2010) LapF, the second largest *Pseudomonas putida* protein, contributes to plant root colonization and determines biofilm architecture. *Mol. Microbiol.* 77, 549–561
- 14 Yoshida, K. *et al.* (2017) Biofilm formation by *Paracoccus denitrificans* requires a type I secretion system-dependent adhesin BapA. *FEMS Microbiol. Lett.* 364,
- 15 Lally, E.T. *et al.* (1999) The interaction between RTX toxins and target cells. *Trends Microbiol.* 7, 356–361
- 16 Linhartová, I. *et al.* (2010) RTX proteins: a highly diverse family secreted by a common mechanism. *Fems Microbiol. Rev.* 34, 1076–1112
- 17 Balakrishnan, L. *et al.* (2001) Substrate-triggered recruitment of the TolC channel-tunnel during type I export of hemolysin by *Escherichia coli*. *J. Mol. Biol.* 313, 501–510
- 18 Thanabalu, T. *et al.* (1998) Substrate-induced assembly of a contiguous channel for protein export from *E.coli*: reversible bridging of an inner-membrane translocase to an outer membrane exit pore. *EMBO J.* 17, 6487–6496
- 19 Lenders, M.H.H. *et al.* (2015) Directionality of substrate translocation of the hemolysin A Type I secretion system. *Sci. Rep.* 5, doi: 10.1038/srep12470

- 20 Bumba, L. *et al.* (2016) Calcium-Driven Folding of RTX Domain β -Rolls Ratchets Translocation of RTX Proteins through Type I Secretion Ducts. *Mol. Cell* 62, 47–62
- 21 Mitchell, A. *et al.* (2015) The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* 43, D213–221
- 22 Bryan, A.W. *et al.* (2011) Structure-based prediction reveals capping motifs that inhibit β -helix aggregation. *Proc. Natl. Acad. Sci. U. S. A.* 108, 11099–11104
- 23 Peralta, M.D.R. *et al.* (2015) Engineering Amyloid Fibrils from β -Solenoid Proteins for Biomaterials Applications. *ACS Nano* 9, 449–463
- 24 Holland, I.B. *et al.* (2016) Type I Protein Secretion-Deceptively Simple yet with a Wide Range of Mechanistic Variability across the Family. *EcoSal Plus* 7, doi: 10.1128/ecosalplus.ESP-0019-2015
- 25 Thomas, S. *et al.* (2014) The Type 1 secretion pathway - the hemolysin system and beyond. *Biochim. Biophys. Acta* 1843, 1629–1641
- 26 Navarro, M.V.A.S. *et al.* (2011) Structural basis for c-di-GMP-mediated inside-out signaling controlling periplasmic proteolysis. *PLoS Biol.* 9, e1000588
- 27 Newell, P.D. *et al.* (2009) LapD is a bis-(3',5')-cyclic dimeric GMP-binding protein that regulates surface attachment by *Pseudomonas fluorescens* Pf0–1. *Proc. Natl. Acad. Sci.* 106, 3461–3466
- 28 Koronakis, V. *et al.* (2000) Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* 405, 914–919
- 29 Koronakis, V. *et al.* (2004) Structure and function of TolC: the bacterial exit duct for proteins and drugs. *Annu. Rev. Biochem.* 73, 467–489
- 30 Smith, T.J. *et al.* (2018) An N-terminal Retention Module Anchors the Giant Adhesin LapA of *Pseudomonas fluorescens* at the Cell Surface: A Novel Sub-family of Type I Secretion Systems. *J. Bacteriol.* DOI: 10.1128/JB.00734-17
- 31 Smith, T.J. *et al.* (2018) Type 1 Does The Two-Step: Type 1 Secretion Substrates With A Functional Periplasmic Intermediate. *J. Bacteriol.* DOI: 10.1128/JB.00168-18
- 32 Newell, P.D. *et al.* (2011) A c-di-GMP Effector System Controls Cell Adhesion by Inside-Out Signaling and Surface Protein Cleavage. *PLOS Biol.* 9, e1000587
- 33 Barlag, B. and Hensel, M. (2015) The giant adhesin SiiE of *Salmonella enterica*. *Mol. Basel Switz.* 20, 1134–1150

- 34 Wagner, C. *et al.* (2014) The *Salmonella enterica* giant adhesin SiiE binds to polarized epithelial cells in a lectin-like manner. *Cell. Microbiol.* 16, 962–975
- 35 Borlee, B.R. *et al.* (2010) *Pseudomonas aeruginosa* uses a cyclic-di-GMP-regulated adhesin to reinforce the biofilm extracellular matrix. *Mol. Microbiol.* 75, 827–842
- 36 Cooley, R.B. *et al.* (2016) Cyclic Di-GMP-Regulated Periplasmic Proteolysis of a *Pseudomonas aeruginosa* Type Vb Secretion System Substrate. *J. Bacteriol.* 198, 66–76
- 37 Petosa, C. *et al.* (1997) Crystal structure of the anthrax toxin protective antigen. *Nature* 385, 833–838
- 38 Rigden, D.J. *et al.* (2004) The PA14 domain, a conserved all- β domain in bacterial toxins, enzymes, adhesins and signaling molecules. *Trends Biochem. Sci.* 29, 335–339
- 39 de Groot, P.W.J. and Klis, F.M. (2008) The conserved PA14 domain of cell wall-associated fungal adhesins governs their glycan-binding specificity. *Mol. Microbiol.* 68, 535–537
- 40 Maestre-Reyna, M. *et al.* (2012) Structural basis for promiscuity and specificity during *Candida glabrata* invasion of host epithelia. *Proc. Natl. Acad. Sci.* 109, 16864–16869
- 41 Veelders, M. *et al.* (2010) Structural basis of flocculin-mediated social behavior in yeast. *Proc. Natl. Acad. Sci.* 107, 22511–22516
- 42 Boyd, C.D. *et al.* (2014) Structural Features of the *Pseudomonas fluorescens* Biofilm Adhesin LapA Required for LapG-Dependent Cleavage, Biofilm Formation, and Cell Surface Localization. *J. Bacteriol.* 196, 2775–2788
- 43 Cirillo, S.L.G. *et al.* (2001) *Legionella pneumophila* Entry Gene rtxA Is Involved in Virulence. *Infect. Immun.* 69, 508–517
- 44 Springer, T.A. (2014) von Willebrand factor, Jedi knight of the bloodstream. *Blood* 124, 1412–1425
- 45 Barczyk, M. *et al.* (2010) Integrins. *Cell Tissue Res.* 339, 269–280
- 46 Bienkowska, J. *et al.* (1997) The von willebrand factor A3 domain does not contain a metal ion-dependent adhesion site motif. *J. Biol. Chem.* 272, 25162–25167
- 47 Emsley, J. *et al.* (1997) Crystal structure of the I domain from integrin $\alpha 2\beta 1$. *J. Biol. Chem.* 272, 28512–28517
- 48 Emsley, J. *et al.* (1998) Crystal structure of the von Willebrand Factor A1 domain and implications for the binding of platelet glycoprotein Ib. *J. Biol. Chem.* 273, 10396–10401

- 49 Lee, J.O. *et al.* (1995) Two conformations of the integrin A-domain (I-domain): a pathway for activation? *Struct. Lond. Engl.* 1993 3, 1333–1340
- 50 Krishnan, V. *et al.* (2013) Structure of *Streptococcus agalactiae* tip pilin GBS104: a model for GBS pili assembly and host interactions. *Acta Crystallogr. D Biol. Crystallogr.* 69, 1073–1089
- 51 Emsley, J. *et al.* (2000) Structural basis of collagen recognition by integrin $\alpha 2\beta 1$. *Cell* 101, 47–56
- 52 Huizinga, E.G. *et al.* (2002) Structures of Glycoprotein Iba and Its Complex with von Willebrand Factor A1 Domain. *Science* 297, 1176–1179
- 53 Peyvandi, F. *et al.* (2011) Role of von Willebrand factor in the haemostasis. *Blood Transfus.* 9, s3–s8
- 54 Schneider, S.W. *et al.* (2007) Shear-induced unfolding triggers adhesion of von Willebrand factor fibers. *Proc. Natl. Acad. Sci.* 104, 7899–7903
- 55 Ellison, C. and Brun, Y.V. (2015) Mechanosensing: A Regulation Sensation. *Curr. Biol.* 25, R113–R115
- 56 Siryaporn, A. *et al.* (2014) Surface attachment induces *Pseudomonas aeruginosa* virulence. *Proc. Natl. Acad. Sci.* 111, 16860–16865
- 57 Ellison, C.K. *et al.* (2017) Obstruction of pilus retraction stimulates bacterial surface sensing. *Science* 358, 535–538
- 58 Mignolet, J. *et al.* (2018) More than a Tad: spatiotemporal control of *Caulobacter* pili. *Curr. Opin. Microbiol.* 42, 79–86
- 59 Wang, Y. and Chen, C. (2005) Mutation analysis of the flp operon in *Actinobacillus actinomycetemcomitans*. *Gene* 351, 61–71
- 60 Garnham, C.P. *et al.* (2011) Anchored clathrate waters bind antifreeze proteins to ice. *Proc. Natl. Acad. Sci. U. S. A.* 108, 7363–7367
- 61 Nutt, D.R. and Smith, J.C. (2008) Dual function of the hydration layer around an antifreeze protein revealed by atomistic molecular dynamics simulations. *J. Am. Chem. Soc.* 130, 13066–13073
- 62 El-Azami-El-Idrissi, M. *et al.* (2003) Interaction of *Bordetella pertussis* adenylate cyclase with CD11b/CD18: Role of toxin acylation and identification of the main integrin interaction domain. *J. Biol. Chem.* 278, 38514–38521

- 63 Masin, J. *et al.* (2015) Bordetella adenylate cyclase toxin: a unique combination of a pore-forming moiety with a cell-invading adenylate cyclase enzyme. *Pathog. Dis.* 73, ftv075
- 64 Bulutoglu, B. *et al.* (2017) Catch and Release: Engineered Allosterically Regulated β -Roll Peptides Enable On/Off Biomolecular Recognition. *ACS Synth. Biol.* 6, 1732–1741
- 65 Syed, K.A. *et al.* (2009) The Vibrio cholerae Flagellar Regulatory Hierarchy Controls Expression of Virulence Factors. *J. Bacteriol.* 191, 6555–6570
- 66 Griessl, M.H. *et al.* (2013) Structural Insight into the Giant Ca^{2+} -Binding Adhesin SiiE: Implications for the Adhesion of Salmonella enterica to Polarized Epithelial Cells. *Structure* 21, 741–752
- 67 Vance, T.D.R. *et al.* (2014) Ca^{2+} -stabilized adhesin helps an Antarctic bacterium reach out and bind ice. *Biosci. Rep.* 34, doi: 10.1042/BSR20140083
- 68 Guo, S. *et al.* (2013) Role of Ca^{2+} in folding the tandem β -sandwich extender domains of a bacterial ice-binding adhesin. *FEBS J.* 280, 5919–5932
- 69 Vrieling, A.S.O. *et al.* (2017) Unusually high mechanical stability of bacterial adhesin extender domains having calcium clamps. *PLOS ONE* 12, e0174682
- 70 Carrion-Vazquez, M. *et al.* (1999) Mechanical and chemical unfolding of a single protein: a comparison. *Proc. Natl. Acad. Sci. U. S. A.* 96, 3694–3699
- 71 Li, H. *et al.* (2002) Reverse engineering of the giant muscle protein titin. *Nature* 418, 998–1002
- 72 Lu, H. *et al.* (1998) Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.* 75, 662–671
- 73 Oberhauser, A.F. and Carrión-Vázquez, M. (2008) Mechanical Biochemistry of Proteins One Molecule at a Time. *J. Biol. Chem.* 283, 6617–6621
- 74 Luo, Y. *et al.* (2000) Crystal structure of enteropathogenic Escherichia coli intimin-receptor complex. *Nature* 405, 1073–1077
- 75 Gerlach, R.G. *et al.* (2008) Cooperation of Salmonella pathogenicity islands 1 and 4 is required to breach epithelial barriers. *Cell. Microbiol.* 10, 2364–2376
- 76 Wrobel, A. *et al.* (2018) The repeat structure of two paralogous genes, Yersinia ruckeri invasin (yrInv) and a “Y. ruckeri invasin-like molecule”, (yrIlm) sheds light on the evolution of adhesive capacities of a fish pathogen. *J. Struct. Biol.* 201, 171–183

- 77 Guo, S. *et al.* (2012) Re-Evaluation of a Bacterial Antifreeze Protein as an Adhesin with Ice-Binding Activity. *PLOS ONE* 7, e48805
- 78 Arahall, D.R. *et al.* (2016) *Marinomonas blandensis* sp. nov., a novel marine gammaproteobacterium. *Int. J. Syst. Evol. Microbiol.* 66, 5544–5549
- 79 Bai, X. *et al.* (2014) *Marinomonas profundimaris* sp. nov., isolated from deep-sea sediment sample of the Arctic Ocean. *Antonie Van Leeuwenhoek* 106, 449–455
- 80 Matsunaga, T. *et al.* (2005) Complete genome sequence of the facultative anaerobic magnetotactic bacterium *Magnetospirillum* sp. strain AMB-1. *DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes* 12, 157–166
- 81 Bar Dolev, M. *et al.* (2016) Putting life on ice: bacteria that bind to frozen water. *J. R. Soc. Interface* 13, doi: 10.1098/rsif.2016.0210
- 82 Spaulding, C.N. *et al.* (2017) Selective depletion of uropathogenic *E. coli* from the gut by a FimH antagonist. *Nature* 546, 528–532
- 83 Spaulding, C.N. *et al.* (2018) Precision antimicrobial therapeutics: the path of least resistance? *Npj Biofilms Microbiomes* 4, 4
- 84 Bonham, K.S. *et al.* (2017) Extensive horizontal gene transfer in cheese-associated bacteria. *eLife* 6, doi: 10.7554/eLife.22144
- 85 Pham, N.-P. *et al.* (2017) Comparative genomic analysis of *Brevibacterium* strains: insights into key genetic determinants involved in adaptation to the cheese habitat. *BMC Genomics* 18, 955
- 86 Raymond, J.A. and Kim, H.J. (2012) Possible role of horizontal gene transfer in the colonization of sea ice by algae. *PloS One* 7, e35968
- 87 Vance, T.D.R. *et al.* (2018) An ice-binding and tandem beta-sandwich domain-containing protein in *Shewanella frigidimarina* is a potential new type of ice adhesin. *FEBS J.* 285, 1511–1527
- 88 Guo, S. *et al.* (2018) Conserved structural features anchor biofilm-associated RTX-adhesins to the outer membrane of bacteria. *FEBS J.* 285, 1812–1826

Legends

Figure 1. General domain architecture of RTX adhesins. The RTX adhesins share a common domain architecture: a C-terminal secretion signal and accompanying RTX repeats for transport through the T1SS (green); followed by the ligand-binding region, which is the most variable part among different RTX adhesins; a central extender region composed of many repeating units (cyan); and an N-terminal cell-anchoring region (grey). Representative ligand-binding domains with known or predicted structures in these six RTX adhesin examples are the sugar-binding PA14 domain (magenta), the von Willebrand Factor (vWFA) domain (yellow), and, within *MpIBP*, the peptide-binding domain (PBD) (blue), and the ice-binding domain (orange). Parts of the ligand-binding region in white may contain domains of unknown structure and functionality.

Figure 2. T1SS domain comparisons. Structures of the C-terminal domains of CyaA and *MpIBP*. A) Structure of the C-terminal domain of CyaA of *B. pertussis* (PDB: 5CVW). The N-terminal RTX β -roll is colored cyan, whereas the C-terminal T1SS signal sequence is colored magenta. Calcium ions are represented by gold spheres. B) Structure of C-terminal RTX repeats and T1SS sequence of *MpIBP* (PDB: 5JUH). The same color scheme is used as in (A). C) Structure of the ice-binding domain of *MpIBP*. The structures are rotated 90° around a horizontal axis to show the face of the capping motifs in A), B) and C), respectively.

Figure 3. Road map of the *MpIBP* ice-adhesin showing four functional regions. A) Linear domain map of *MpIBP* drawn to scale. The membrane-anchoring region at the N terminus is coloured grey. The ~120 repeats that span the central extender region are colored cyan. In the ligand-binding region, the peptide-binding region is coloured dark blue, the sugar-binding domain is coloured magenta, while the ice-binding domain is coloured orange. The T1SS sequence at the C terminus is coloured green. The small white regions near the C terminus are linkers with no known three-dimensional structures. B) NMR and X-ray crystal structures of linked *MpIBP* domains from N to C termini are shown in cartoon representation. The same color scheme is applied as in A). Near the N terminus, the SAXS structure of the T1SS- β -barrel-spanning module of *MpIBP* is illustrated as a gray cylinder. Outer membrane (OM) is indicated by horizontal lines on either side of the *MpIBP* T1SS- β -barrel-spanning module. EC and PM refer to extra-cellular and periplasmic locations. Hatched lines indicate the ~ 108 central repeats that are not shown in the figure. The linker regions near the C terminus are indicated by wavy lines. C) Zoomed-in view of the ligand-binding region of *MpIBP*. The same color scheme is used as in B). Ca^{2+} ions are indicated as yellow spheres. The insets show enlarged views of the peptide (left) and sugar (right)

molecules anchored to the ligand-binding domains and shown in stick representation (carbon atoms in white, oxygen atoms in red, and nitrogen atoms in blue) with $2F_o - F_c$ map counteracted at 1σ (blue mesh). This figure is adapted from [1].

Figure 4. Schematic of *MpIBP* binding *M. primoryensis* and diatoms to ice. (A) Ice/snow that covers the water surface in Antarctica (eg. Ace Lake) is represented by a grey rectangle with three internal brine channels of irregular shape. Lake water is colored blue with a light to dark gradient from top to bottom signifying the increased availability of light and oxygen towards the top of the water column as indicated by the grey arrow. Bacteria and photosynthetic microorganisms such as diatoms within the brine pits and underneath the ice are drawn as small white ovals and large green ovals, respectively. The phototrophic and anoxic zones are indicated on the right. (B) Expanded view of (A) showing two linked bacterial cells bound to ice and a diatom. Cell-surface proteins and carbohydrates are drawn as fuzzy black hairs and the polar flagella are drawn as squiggles. *MpIBPs* protrude from cell surfaces. The extender region, peptide-binding region, sugar-binding domain, ice-binding domain and T1SS sequence are drawn as cyan rods, blue ovals, magenta hexagons, orange rectangles, and dark green triangles, respectively. (C) Enlarged view of *MpIBP* peptide-binding domain (PBD) interacting with a protein on the cell surface of a diatom. Ligand-binding Ca^{2+} are drawn as yellow spheres. Surface protein is indicated by a wavy line from the cell surface. (D) Enlarged view of *MpIBP* sugar-binding domain (SBD) interacting with a sugar molecule on the cell surface of a diatom. Ligand-binding Ca^{2+} are drawn as yellow spheres. (E) Expanded view of the membrane anchoring domains of *MpIBP* (grey triangle, cylinder and ovals) near the outer membrane (OM). RIN is the N-terminal domain of the Retention region that cannot pass through TolC. RIM is the middle domain of this region that becomes lodged within the tunnel of TolC. Surface glycans are drawn as connected brown hexagons. The hollow T1SS β -barrel pore through the OM is outlined in black. Arrow with a broken line indicates the protein continues to the C terminus of *MpIBP*. This figure is adapted from [1].

Figure 5. Retention of RTX adhesin in the bacterial outer membrane. A) In T1SS, a membrane fusion protein (MFP) and an ABC transporter form a stable translocase complex anchored to the inner-membrane. (B) Upon contact with the unstructured C-terminal sequence of *MpIBP* or

another T1SS substrate, this translocase complex undergoes conformational changes to recruit the T1SS outer-membrane protein, TolC, and form a continuous channel across the cell envelope. (C) Most of the *MpIBP* sequence remains unstructured in the secretion channel, but folds upon entering the Ca^{2+} -rich extracellular environment. In contrast, the N-terminal plug (grey triangle) and β -barrel-spanning (grey cylinder) domains are structured in the absence of Ca^{2+} during secretion. (D) When the entire T1SS substrate has passed the IM-translocase complex, the T1SS machinery can disassemble, leaving the β -barrel-spanning module inserted in the TolC pore, while N-terminal plug module prevents the total release of *MpIBP* by steric hindrance. This schematic is adapted from [88].