Conserved structural features anchor biofilm-associated RTX-adhesins to the outer membrane of bacteria

The FEBS Journal

FEBS PRESS science publishing by scientists

Conserved structural features anchor biofilm-associated RTX–adhesins to the outer membrane of bacteria

Shuaiqi Guo[1,2,3], David N. Langelaan[1,*], Sean W. Phippen[1], Steven P. Smith[1], Ilja K. Voets[2,3] and Peter L. Davies[1]

1 Protein Function Discovery Group, Department of Biomedical and Molecular Sciences, Queen's University, Kingston, Canada
2 Institute for Complex Molecular Systems, Laboratory of Macromolecular and Organic Chemistry, Department of Chemical Engineering and Chemistry, Eindhoven University of Technology, the Netherlands
3 Laboratory of Physical Chemistry, Department of Chemical Engineering and Chemistry, Eindhoven University of Technology, the Netherlands

Keywords
bacterial adhesins; biofilms; structural biology; surface-retention domains

Correspondence
P. L. Davies, Department of Biomedical and Molecular Sciences, Queen's University, Kingston, ON, Canada K7L 3N6
Fax +1 613 533 2497
Tel. +1 613 533 2983
E-mail: peter.davies@queensu.ca

***Present address**
Department of Biochemistry & Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada

(Received 13 November 2017, revised 5 February 2018, accepted 19 March 2018)

doi:10.1111/febs.14441

Repeats-in-toxin (RTX) adhesins are present in many Gram-negative bacteria to facilitate biofilm formation. Previously, we reported that the 1.5-MDa RTX adhesin (*Mp*IBP) from the Antarctic bacterium, *Marinomonas primoryensis*, is tethered to the bacterial cell surface via its N-terminal Region I (RI). Here, we show the detailed structural features of RI. It has an N-terminal periplasmic retention domain (RIN), a central domain (RIM) that can insert into the β-barrel of an outer-membrane pore protein during *Mp*IBP secretion, and three extracellular domains at its C terminus (RIC) that transition the protein into the extender region (RII). RIN has a novel β-sandwich fold with a similar shape to βγ-crystallins and tryptophan RNA attenuation proteins. Because RIM undergoes fast and extensive degradation *in vitro*, its narrow cylindrical shape was rapidly measured by small-angle X-ray scattering before proteolysis could occur. The crystal structure of RIC comprises three tandem β-sandwich domains similar to those in RII, but increasing in their hydrophobicity with proximity to the outer membrane. In addition, the key Ca^{2+} ion that rigidifies the linkers between RII domains is not present between the first two of these RIC domains. This more flexible RI linker near the cell surface can act as a 'pivot' to help the 0.6-μm-long *Mp*IBP sweep over larger volumes to find its binding partners. Since the physical features of RI are well conserved in the RTX adhesins of many Gram-negative bacteria, our detailed structural and bioinformatic analyses serve as a model for investigating the surface retention of biofilm-forming bacteria, including human pathogens.

Introduction

Repeats-in-toxin (RTX) adhesins are a class of biofilm-associated proteins (BAPs) that help Gram-negative bacteria form biofilms [1–6]. Bacteria living within a biofilm are extremely difficult to eradicate as they become impervious to various bactericidal treatments including antibiotics and biocides [7,8]. As a result, greater than 80% of chronic bacterial infections in humans are caused by biofilms. During biofilm formation, RTX adhesins play a pioneering role in attaching their bacteria to a surface, and subsequently facilitate the development of micro-colonies, leading to the formation of a mature biofilm [1,2]. RTX adhesins

Abbreviation
BAPs, biofilm-associated proteins; *Mp*IBP, *Marinomonas primoryensis* ice-binding protein; OM, outer membrane; Phyre2, Protein Homology/analogy recognition Engine Server; RTX, repeats-in-toxin; SAXS, small-angle x-ray scattering.

are often encoded by the largest genes within their respective bacteria, and they share a common modular architecture. Starting from the C terminus, a noncleavable signal sequence of the Type I Secretion System (T1SS) is followed by several tandem Ca^{2+}-binding nona-peptide repeats (GGxGxDxUx; where x is any residue, while U is typically a large hydrophobic amino acid). Next are one or more ligand-binding domains that tether the bacteria to various surfaces, including other cells. The central region of the adhesin comprises numerous tandem Bacterial Immunoglobulin-like (BIg) domains that are thought to have an extension function. The N-terminal region is typically nonrepetitive and is hypothesized to anchor the large RTX adhesins to the bacterial outer membrane (OM). However, the mechanism for this membrane retention is unclear.

One of the most well-studied RTX adhesins is the 0.8-MDa Large Adhesion Protein A (LapA) of *Pseudomonas fluorescens* which is required for its biofilm formation [9–11]. The N-terminal region of LapA is localized in the periplasmic space, while the rest of the protein protrudes from the OM, despite the lack of predicted transmembrane domains. The periplasmic domain of LapA is cleaved under nutrient-depleted conditions to allow detachment of bacteria from biofilms, whereupon they become motile again and capable of searching for, and moving to, a more favorable location. Nevertheless, the molecular basis of the surface-anchoring of LapA is currently unknown, in part, due to a lack of detailed structural information.

One RTX adhesin that has been characterized in molecular detail is the 1.5-MDa *Mp*IBP found on the surface of an Antarctic Gram-negative marine bacterium, *Marinomonas primoryensis* [12,13]. *Mp*IBP is a homolog of LapA, and thus serves as model system to characterize membrane anchoring by these adhesins. *Mp*IBP has been divided into five distinct regions (Regions I–V). Over 90% of the adhesin's mass lies within its central Region II (RII), which contains approximately 120 identical tandem copies of a BIg domain [13–15]. In the presence of millimolar concentrations of Ca^{2+}, the BIg domains are extended into a long rod-like structure that is approximately 600 nm in length, thereby projecting the three substrate/ligand-binding domains in RIII and RIV well away from the cell surface. RIV binds ice [16–19], whereas RIII contains a sugar-binding domain and a peptide/protein-binding domain that can attach the bacterium to at least one species of Antarctic diatoms [2]. We have hypothesized that *Mp*IBP simultaneously links its bacterium to ice and diatoms to help form symbiotic biofilms on the underside of sea and lake ice in Antarctica [2].

Previously, we reported the overall structures of the RI domains solved by NMR spectroscopy (PDB: 5IX9), X-ray crystallography (PDB: 5IRB), and small-angle X-ray scattering (SAXS) [2]. Here, we describe the detailed structural features that enable the RI domains to indirectly span the OM through interacting with the T1SS β-barrel pore protein, and thereby anchor the large adhesin to the cell surface. On the extracellular side of the OM, RI has a flexible linker that allows the stiff adhesin to pivot. Using bioinformatic analyses, we show the structural features of RI are widespread in the RTX adhesins from many Gram-negative bacteria. Our work provides the structural basis for the cell-surface retention of RTX adhesion proteins in the T1SS system primarily designed for secretion.

Results and Discussion

Full-length RI undergoes proteolysis into a 35-kDa product at 4 °C

The recombinant RI construct, comprising residues 1–494 of *Mp*IBP (Fig. 1A) and a C-terminal His-tag, was produced in *Escherichia coli* and purified by Ni^{2+} affinity and size-exclusion chromatographies. RI migrated on SDS/PAGE with an apparent molecular weight of 55 kDa (Fig. 1B, lane 2). Crystallization trials of RI resulted in the growth of rectangular-prismatic crystals (Fig. 1C). SDS/PAGE analysis of these crystals revealed a band migrating to ~ 35 kDa and no trace of the 55-kDa starting material (Fig. 1D, lane 2). This was suggestive of rapid endogenous proteolysis of RI. Indeed, after an incubation period of ~ 2 days at 4 °C, the ~ 55-kDa RI showed several minor cleavage products down to ~ 40 kDa (Fig. 1B, lane 2). Bands corresponding to these cleavage products were more prominent after 5 days and were accompanied by smaller proteolytic fragments (Fig. 1B, lane 3). At 7 days, a single major proteolytic RI fragment of ~ 35 kDa was observed (Fig. 1B, lane 4), which is consistent with the size of the crystallized proteolytic RI fragment.

The 35-kDa proteolysis product of RI folds as three tandem β-sandwich domains bound with Ca^{2+} and Mg^{2+} ions

Homology searches of the full RI construct identified three putative domains in tandem at the C-terminal end that were 43%, 57% and 98% identical to the RII repeats of *Mp*IBP [14,15], which fold as Ca^{2+}-dependent β-sandwiches. However, the remaining amino-


A
M S D N Q F P F A T L G N A I G F I T K L D G S V T V Q S I D G Q E R V L K L G
β1
β2
β3
D P I F F G E T V L T G G S G S V T I A F V D G T D V V I G G D S I V E M T D E
β4
β5
β6
β7
β8
I Y N T G D N E D L V A D S S S E I D A L Q N A I L A G D D P T L I Q D A P A A
G N T L A D Q Q R V D V S I E R N D N S A Q A G F G V D T Q S S L P T Y G Y D T
D N G N G G Q A T E R E Y S A P S L S R T L N Q S P L L I N L D I D P V T G D S
V I N A A E A G G T V T L T G V V N G D V F S S G V V T L V I N G V T Y S T N V
N P N G T W S V S V A G S D L S A D S D R I V D A S V V V T N G A G Q Q G T A D
S T E S F I V K T S S R A T I R V N S I T S D D V V N A E E S N S T I T V S G R
V G L D A S A G D T V S M T I N G T L Y T T V V L A N K T W S V G V S G S D L A
Q D N S F Q V S V T G Q D S A G N P Y A G T T T S T H T V D T S A D A G T V T V
N A I T S D D V I N A S E A A G T V A V S G T A T G G D I A E G D T V T L E I N
G E T Y T T T V D A N G E W S V D V A G S D L A A D T A F D A V V T S S D A A G
N T V D T T G S S T H T V D
B
1 2 3 4
66 kDa
35 kDa
27 kDa
C
0.1 mm
D
1 2
66 kDa
35 kDa
27 kDa

Fig. 1. Amino-acid sequence of *Mp*AFP_RI and crystallization of its 35-kDa degradation product. (A) Secondary structures were obtained from the X-ray crystal and NMR structures of RIC and RIN (PDB codes: 5IRB and 5IX9), respectively. Horizontal red arrows are β-strands. Red corkscrews are α-helices. Vertical blue arrows mark the boundary residues of the RIN construct. The β-strands in RIN are numbered. 'X's indicate residues with no solved structure. Boxed residues indicate the N-terminal sequencing results of the endogenously proteolyzed RINM revealed by Edman degradation. (B) SDS/PAGE gel showing the progressive degradation of RI. Lane 1: protein marker with molecular masses indicated in kDa. Lanes 2–4 show the postpurification RI at days 2, 5, and 7. (C) Crystals obtained for the residual RI construct. (D) Crystals of residual RI were washed prior to being loaded onto a SDS/PAGE gel (lane2). Lane 1 contains protein markers used in B.

acid sequence leading to RI's N terminus showed no homology with known domains. Given the similar size of the stable proteolytic RI fragment and the predicted tandem domain architecture at its C terminus, the *Mp*IBP RII-monomer structure was used as a partial model to solve the trimodular X-ray structure using the molecular replacement method [14].

The resulting 2-Å resolution X-ray crystal structure of the 35-kDa proteolytic RI fragment comprises the C-terminal 309 amino-acid residues of *Mp*IBP RI (i.e., amino-acid residues 186–494; RIC; Fig. 1A) that form three Ig-like β-sandwich domains, which we have termed Ig1, Ig2, and Ig3, respectively (Fig. 2). Although no evidence of multimerization was observed in solution, the asymmetric unit of RIC contains two protein chains interacting in a parallel manner via a Ca^{2+} ion (not shown). Each chain is a slightly kinked rod with a length of roughly 140 Å and a cross-section of 27 Å × 30 Å. The tandem Ig-like β-sandwich domains are associated with seven Ca^{2+} and two Mg^{2+} ions (Fig. 2A).

Ig1 (residues 186–288) has 43% sequence identity to the RII repeat and is the one that differs the most, whereas Ig2 (residues 289–390) is 57% identical to the RII repeat. Ig3 matches the RII-monomer with the exception of two substituted residues close to the domain's N terminus (Ser instead Glu at residue 392, and Asp instead of Thr at residue 394). Given their high similarity to the RII repeat, it is not surprising that the RIC domains also require the presence of millimolar Ca^{2+} for proper folding [15]. The 104-aa RII repeat contains 18 acidic residues and no basic residues. In contrast, RI-Ig1 and RI-Ig2 contain 11 and 10 acidic residues, respectively, and two and four basic residues, respectively. There is also a clear trend toward increasing numbers of large aliphatic and aromatic residues (Ile, Leu, Val, Met, Phe and Trp) running from Ig3 to Ig1. Ig3 contains 21 of these hydrophobic residues, while the number increases slightly to 24 for Ig2 and increases significantly to 33 for Ig1 (Fig. 2B–D). The increasing divergence of these three Ig-like domains is in stark contrast to the absolute conservation of the following ~120 Ig-like domains that make up RII and are even identical at the DNA level.

Despite their differences in amino-acid composition, Ig1 and Ig2 adopt an identical three-dimensional fold to the RII tandem repeat (backbone RMSDs of 0.65 and 0.48 Å, respectively). The significant difference in hydrophobicity between the three RI β sandwich domains is suggestive of a change in RIC's environment toward its N terminus, or a change of the protein's binding partners in this region. Tandem Ig-like domains located closer to the bacterial OM are more likely to encounter other surface molecules. The exposed hydrophobic regions of RI_Ig1 and RI_Ig2 could be available to stabilize the *MpIBP*-anchoring point by interacting with other cell-surface entities such as lipopolysaccharides.

The linker region between RI_Ig2 and RI_Ig3 is rigidified by a bound Ca^{2+} ion similar to that previously observed between the RII repeats (Fig. 2A right and F) [14]. In the latter case, the inter-domain Ca^{2+} is required to rigidify the connection between RII repeats in order to project the ice- and sugar-binding domains away from the cell surface. In contrast, no bound Ca^{2+} is observed between Ig1 and Ig2 (Fig. 2A left and E), an observation consistent with the key Ca^{2+} coordinating aspartate residue (D390) (Fig. 2F) being substituted by a lysine residue (K288; Fig. 2E). The loss of Ca^{2+} is predicted to result in increased conformational flexibility between the Ig1 and Ig2 domains. Although K288 could form salt bridges with the two nearby aspartate residues, this interaction would be significantly weakened by the high level of salinity in the brackish environment from where *M. primoryensis* was isolated. Indeed, the two different conformations of the RIC chains in the crystal mainly originate in the loop between Ig1 and Ig2. When polypeptide backbones of the two RIC chains were aligned at Ig1 (residues: 183–288), the structures began to diverge at the linker regions between Ig1 and Ig2, clearly showing two distinct conformations (Fig. 2G). However, when polypeptide backbones of the two Ig3s (residues: 391–494) were aligned, the RIC fragment aligned well from Ig2-Ig3, and only began to diverge at the loop linking Ig1 and Ig2 (Fig. 2H). These observations demonstrated that the loss of Ca^{2+} coordination increases the flexibility between the Ig1 and Ig2 of *Mp*IBP_RI, which are located near the bacterial OM.

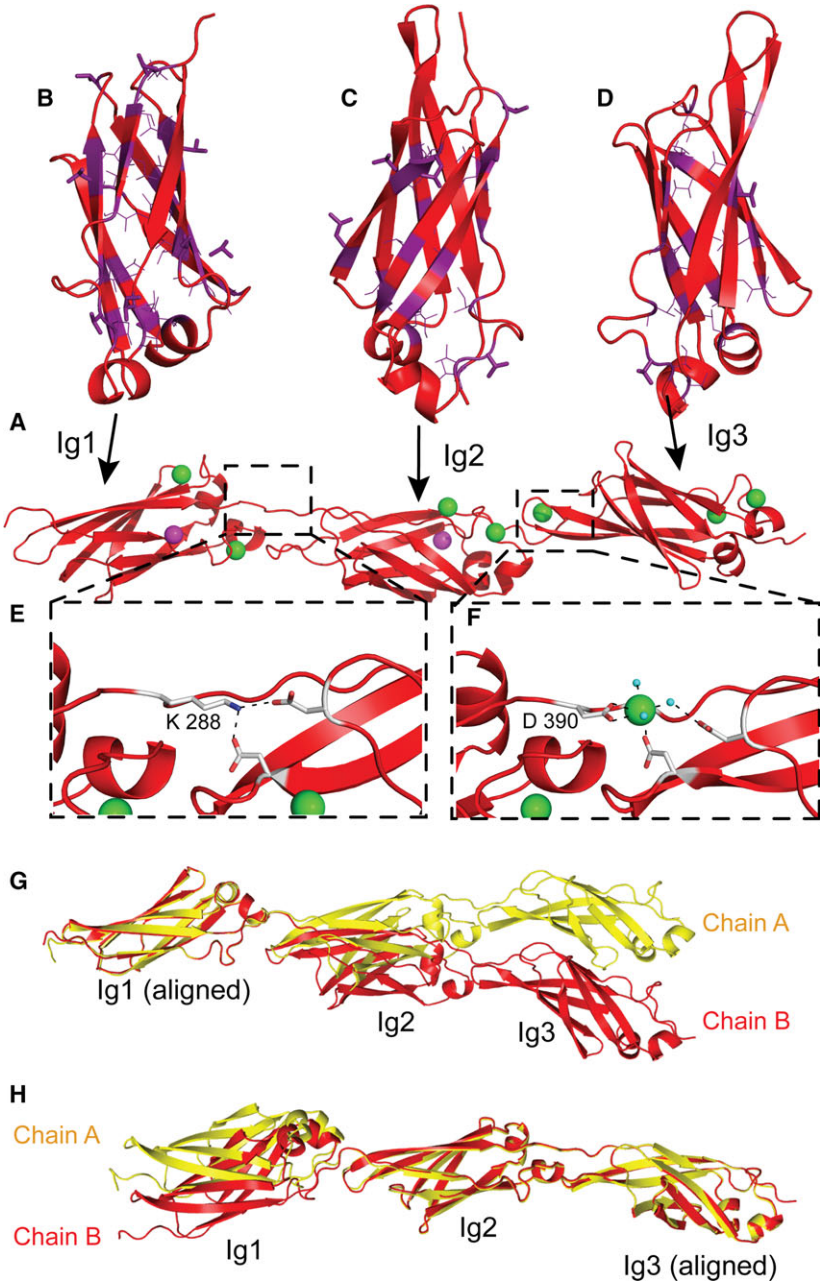

Fig. 2. Structure of three tandem BIg domains at the C-terminal end of RI. (A) Protein is colored red, while Ca^{2+} and Mg^{2+} ions are shown as green and magenta spheres, respectively. (B–D) Hydrophobic residues (Val, Leu, and Ile) are shown in purple. Inward-pointing hydrophobic residues are shown by thin lines, while outward-pointing hydrophobic residues are shown as thick sticks. (E, F) The inter-domain regions of Ig1-Ig2 and Ig2-Ig3, respectively. Ca^{2+}-binding waters are shown as small cyan spheres. (G) Structural alignment centered on Ig1 C_α of the two protein chains (chain A – yellow; chain B – red) found in the asymmetric unit of the RIC crystal. (H) Structural alignment centered on Ig3 C_α of the two protein chains found in the asymmetric unit of the RIC crystal.

Therefore, this flexible linker may act as a pivot to allow the rigid RII arm to move more freely around the cell surface, thereby enabling the ice-binding RIV at the C-terminal tip to sweep over a larger region to contact its main substrate, the ice surface.

The N-terminal 20 kDa of RI (residues 1–187) also undergoes rapid proteolysis

To derive the structural features near the N terminus of RI, a construct spanning the N-terminal 187-aa amino-acid residues of *MpIBP*_RI (i.e., the region

missing from the previously crystallized section of RI; Fig. 1A) was produced in *E. coli*. This construct spanned the N-terminal and middle regions of RI (RINM) and was purified in the same manner as the full-length RI construct (Fig. 3A). The ~ 18-kDa RINM migrated on SDS/PAGE as the major protein band with an apparent molecular weight of ~ 34 kDa, almost twice that of its calculated mass. Anomalies in SDS binding can largely account for these mobility differences [20]. Similar to the full-length RI construct, the RINM construct also underwent proteolysis at room temperature over a 3-week period. What remained was a stable fragment with an apparent molecular weight of 8 kDa (Fig. 3B lanes 2 and 3). Edman degradation identified the N-terminal six amino-acid residues of this fragment as A-T-L-G-N-A (Fig. 1A, boxed residues), which demonstrated that the residual protein represents the N-terminal domain of RI, referred to here as RIN.

The solution structure of RIN (residues 12–87)

Given the small size of the RINM proteolysis fragment, NMR structural analysis was pursued on a construct comprising amino-acid residues 12–87 (RINM 12-87). A $^{13}C/^{15}N$-labeled RIN sample was produced and purified to homogeneity by anion-exchange chromatography (Fig. 3B). The 2D ^{1}H-^{15}N HSQC spectrum of RINM 12–87 showed well-dispersed resonances of uniform intensity and numbering expected for the size of this construct (Fig. 3C), observations consistent with a stably folded protein. Steady-state ^{1}H-^{15}N NOE values were calculated for all residues except Pro42 (Fig. 3D). With the exception of 36V, the core residues (14–82) of RINM 12-87 showed little deviation from the calculated average NOE, suggesting the structure was well ordered. The two N-terminal and five C-terminal residues of RIN 12–87 showed significantly reduced NOE, indicating their increased mobility and flexibility due to being near the termini. Indeed, the backbone superposition of the 20-lowest-energy structural models showed excellent convergence in the core of RIN [2]. Superposition of all 20 structures (backbone atoms) within the ensemble to the lowest energy ensemble member produced a pairwise root-mean-square deviation (RMSD) of 0.33 ± 0.06 Å.

The 76-aa residual RIN protein had an overall β-sandwich fold with two β-sheets packed against each other to form a compact hydrophobic core with an overall triangular-shaped cross-section (Fig. 4E). The first β-sheet is composed of strands β1, β4, β6, and β7, while β2, β3, β5, and β8 make up the second β-sheet. The RIN β-sandwich is roughly 29 Å in length, with a width and a breadth about 22 Å each [2]. A point mutation was noted at residue 31 (31D-N). This residue is pointing outward on the short loop connecting β2 and β3 and is, therefore, unlikely to have an impact on the overall fold of RIN.

The RIN fold resembles structures from the βγ-crystallin superfamily and trp RNA attenuation proteins

The RIN 12–87 structure was submitted to the DALI server to identify similar folds. The top 84 structural homologs had *Z*-scores of 4–4.9 over 51–59 residues and sequence identities of 5–17% and belonged to the βγ-crystallin superfamily or to Trp RNA attenuation proteins (TRAPs). Despite sharing a similar β-sandwich fold with a triangular cross-section, the topology of the RIN β-strands is inherently different from those of the βγ-crystallin proteins. All βγ-crystallin folds are composed of two consecutive Greek-key motifs that each shares its third strand (β3 and β7) with the opposite motif (Fig. 4A,B). In addition, the first and second β-strands from each motif form highly ordered hairpin loops [21,22]. These hallmarks of βγ-crystallins are not present in RIN. In contrast, RIN shares the same β-strand topology as TRAPs, though the former comprises eight β-strands compared with seven for the latter (Fig. 4C–F).

Unusual structural features of RIN

RIN (12–87) contains 27 amino-acid residues with large hydrophobic side chains (Leu, Ile, Met, Val, Tyr and Phe), most of which are located in the core of the domain (Fig. 5A,B). But notably, nine of these hydrophobic residues are surface exposed giving rise to several distinct hydrophobic patches on the surface of RIN. For example, the outward-projecting F17, F44, P42, V62 and A14 form a protruding hydrophobic patch near the protein's N terminus (Fig. 5C,D).

Despite the presence of Ca^{2+} and Mg^{2+} in solution, these divalent cations had no obvious impact on the fold of RIN [2]. This is in sharp contrast to all other *Mp*IBP domains beyond RIM, which require the presence of millimolar Ca^{2+} for folding [2,14,15,17]. Furthermore, contrary to a conventional protein fold, several polar residues have their side chains pointing toward the interior of the protein. For example, R35, D41 E47 and E29 use their inward-pointing side chains

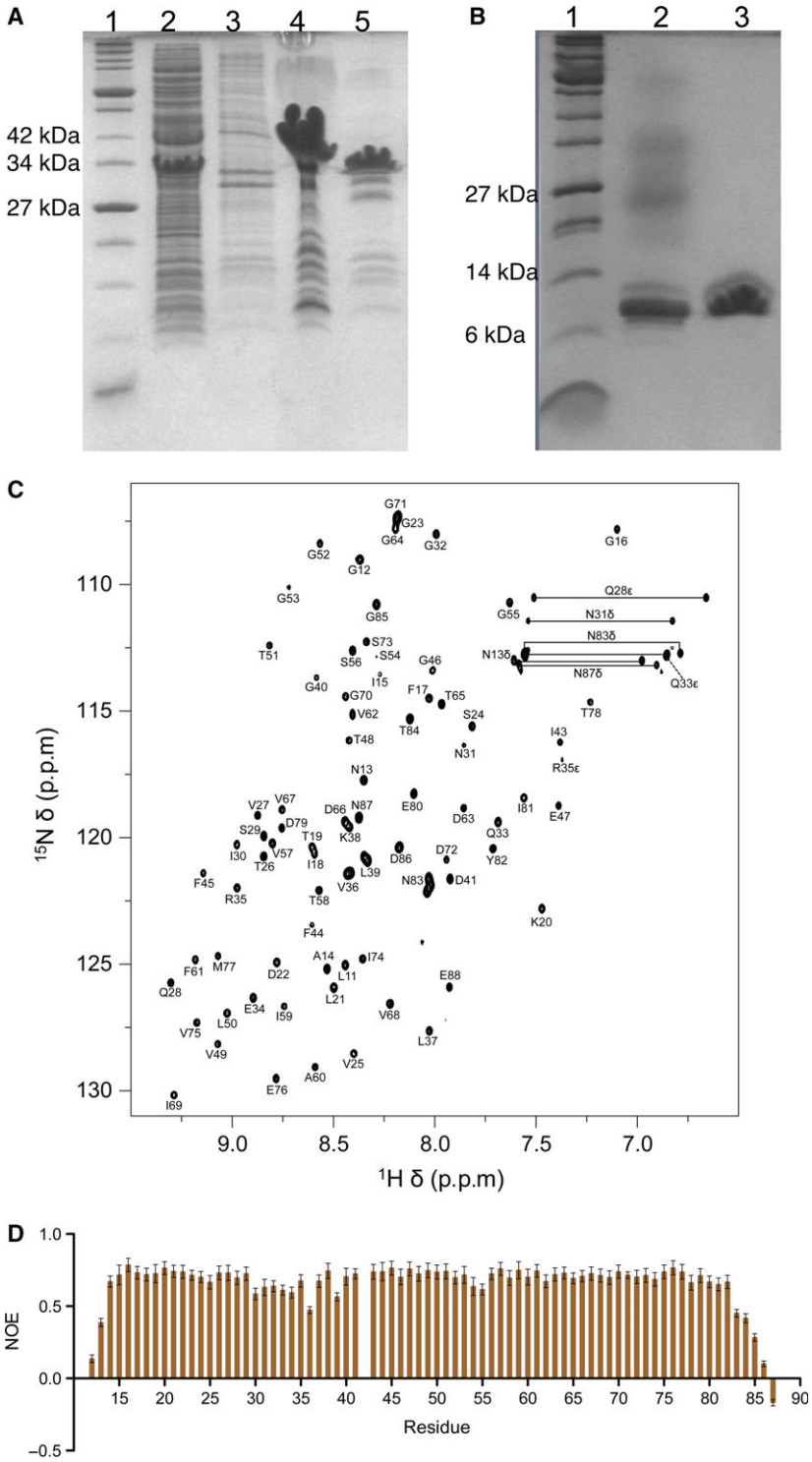

Fig. 3. NMR spectroscopy of the RIN construct obtained by proteolysis of RINM. (A) SDS/PAGE gel showing the purification of RIN. Lane 1: protein marker with molecular masses indicated in kDa. Lane 2: supernatant fraction from the *E. coli* lysate. Lane 3: pellet fraction from the *E. coli* lysate. Lane 4: pooled fractions from the Ni-NTA–agarose affinity-chromatographic step. Lane 5: pooled fractions from the S200 size-exclusion-chromatographic step. (B) SDS/PAGE gel showing degradation of RIN. Lane 1: protein marker. Lane 2: Degradation product of RIN 3 weeks after purification (room temperature). Lane 3: pooled fractions from mono-Q-chromatography used for structure determination. (C) 2D ^{1}H-^{15}N HSQC spectrum of the residual RIN at 600 MHz. Each peak is labeled with its corresponding amino acid. (D) Backbone dynamics of RIN shown by a plot of ^{1}H -^{15}N NOE as a function of residue number.

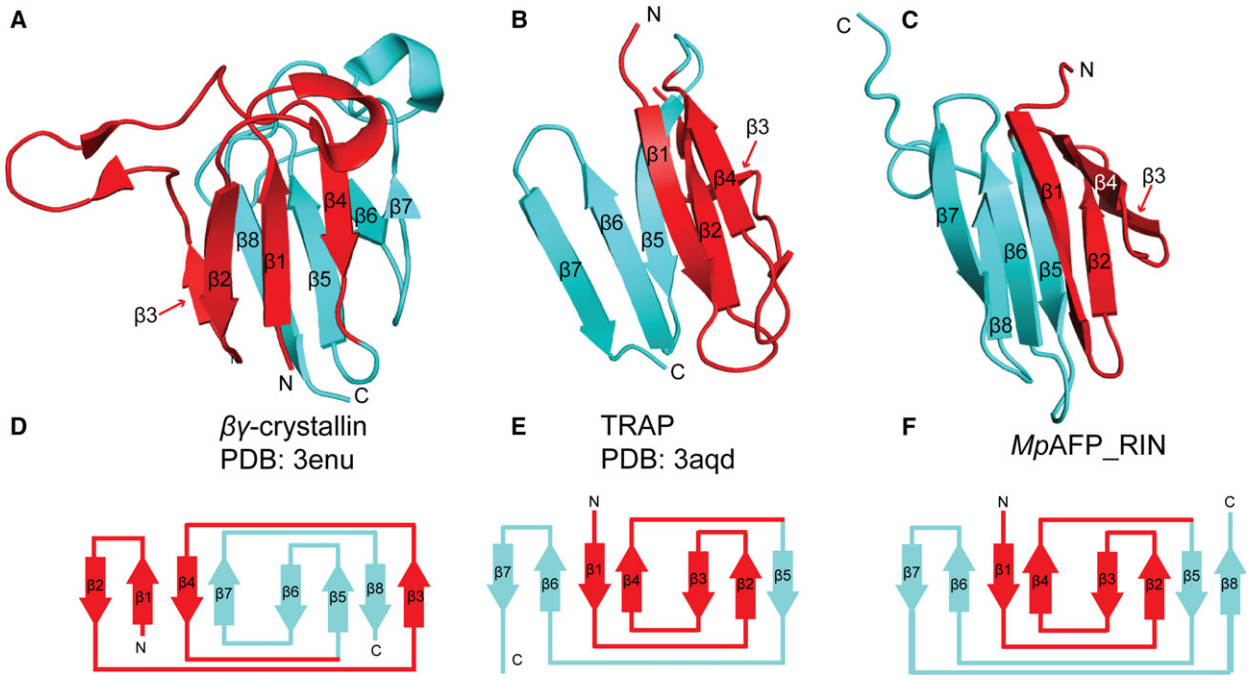

Fig. 4. Structural comparison of RIN with βγ-crystallin and TRAP folds. (A) Structure of a βγ-crystallin fold (PDB: 3enu) in cartoon representation. First four β-strands (arrows) are colored in red, while β5–β8 are colored in cyan. (B) Structure of a TRAP domain (PDB: 3aqd) in cartoon representation. First four β-strands are colored in red, while β5–β7 are colored in cyan. (C) Structure of RIN in cartoon representation. The same color scheme is used as in (A). (D–F) Schematics showing the topology and connectivity of β-strands in βγ-crystallin, TRAP and RIN from (A), (C) and (E), respectively.

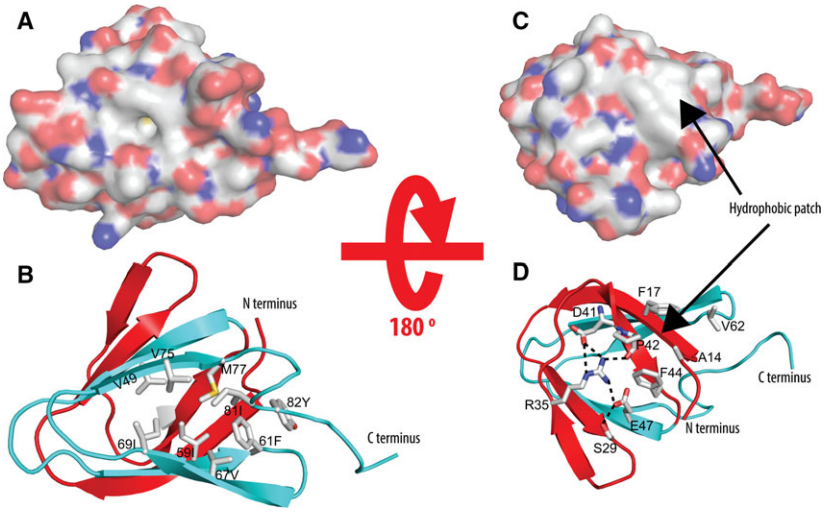

Fig. 5. Unusual structural features of RIN. (A) Surface representation of RIN with O atoms in red and N atoms in blue. (B) Cartoon representation of RIN showing its stabilization by a conventional hydrophobic core comprised the gray residue side chains indicated by their residue numbers. (C, D) RIN is rotated 180° around a horizontal axis from (A) and (B) to show the face that is stabilized by a cluster of inward-pointing polar residues. Several residues (two Phe, one Pro, one Val, and one Ala) are outward-pointing to form a hydrophobic patch as indicated by the arrows. The color scheme of (B) and (D) are the same as that used in Fig. 4.

to form an extensive hydrogen bond network that acts as a capping motif near the bottom of the protein (Fig. 5D). Intriguingly, these inward-pointing polar residues are well conserved among the RTX adhesins/BAPs of Gram-negative bacteria as described in detail below.

Bioinformatic analyses indicate RIN is a periplasmic domain conserved in many BAPs found in Gram-negative bacteria

The cell-surface retention of BAPs is needed for many Gram-negative bacteria to form biofilms, but a lack of detailed structural information about the adhesins' membrane anchoring points has impeded a comprehensive understanding of this key mechanism. For example, biofilm formation by *P. fluorescens* is regulated by the retention and proteolytic release of its 0.8-MDa BAP, LapA. In high nutrient environments (indicated by high inorganic phosphate levels), LapA is retained at the bacterial cell surface to aid in biofilm formation. However, it is known that low inorganic phosphate levels trigger the protease LapG to cleave off the N-terminal domain of LapA in the periplasmic space, thereby freeing this large adhesin from the cell surface. The release of LapA results in the dispersal of *P. fluorescens* from biofilms back to the free-swimming, planktonic life style. To date, no three-dimensional structures have been published for any domains of LapA [9–11].

An amino-acid sequence alignment between RINM and the 210-aa N-terminal residues of LapA shows the two regions share approximately 30% identity (Fig. 6A). Residues 15–121 of RINM and 12–117 of LapA are well aligned (36% identity). This region spans the β-sandwich core of RIN, followed by a conserved putative α-helix (Fig. 6A, red text underlined), and the LapG cleavage site of LapA. Although it is not known if inorganic phosphate is needed for *M. primoryensis* biofilm retention, a putative LapG cleavage site is also found in *Mp*IBP (Fig. 6A, arrowhead). The Thr residue at the first position of the 'TAAG' site in LapA is substituted by a Pro in *Mp*IBP. Sequence alignment beyond G121 of RINM showed much higher variation between the two proteins, with gaps and fewer conserved residues. The unusual feature of inward-pointing polar residues of *Mp*IBP_RIN is conserved in LapA, suggesting its importance in stabilizing the fold. We then used the Phyre2 server for structural modeling of the LapA N-terminal residues using RIN (12–87) as the template structure. The resulting three-dimensional model has an identical fold to RIN with an exceptionally high confidence score of 99% (Fig. 6B,C). LapG cleaves LapA in the periplasmic space; therefore, residues preceding the proteolytic site (residues 1–116) are localized in the periplasm. As RIN precedes the putative LapG cleavage site in *Mp*IBP, it is also likely to be localized to the periplasmic space.

Protein Psi-BLAST identified over 300 distant homologs of RINM from the N-terminal domains of putative RTX adhesins found in many different Gram-negative bacteria. A Weblogo plot was used to show conserved features of RINM among 10 homologs from bacteria including plant and human pathogens, such as *Pectobacterium atrosepticum*, *Vibrio cholerae*, and *P. fluorescens* [23] (Fig. 6D; Table 1). Despite aligning to RINM with relatively low sequence identities of 20.7–30%, the secondary structural elements within the RIN β-sandwich core are well conserved in all 10 sequences [24]. The capping motif of inward-pointing polar residues (positions 36, 42, and 48 of the Weblogo) that is present in *Mp*IBP and LapA is also conserved (Fig. 6D). Indeed, structural models of these sequences all have an identical fold to RIN and produced confidence scores of > 97% (Table 1). This observation suggests that the RIN fold is widespread among RTX adhesins of Gram-negative species. Beyond RIN, more variation was found in what would be classed as RIM in *Mp*IBP (gaps in the alignment are indicated by blank space in the Weblogo plot), and the sequences are predominantly predicated as random coil [24]. However, an α-helical element (xxIQxAIAA; where x can be any residues) with a short coil (GxDPT) followed soon after by the putative LapG cleavage site (uAAG from position 139–142; where u is either a threonine or proline) is exceptionally well aligned (Fig. 6D). This is consistent with reports that periplasmic proteolysis could be a general mechanism for bacteria to break free from biofilms in nutrient-depleted conditions [9–11].

Fig. 6. Conserved features of RINM in RTX adhesins of other Gram-negative bacteria. (A) Amino-acid sequence alignment between RINM (1–187) and the N-terminal part of *Pseudomonas fluorescens* LapA (1–218). Conserved residues that make up β-strands are colored in blue, while those that form putative α-helices are colored in red with an underline. The red arrowhead points to the LapG cleavage site. Consensus amino acids are indicated in the third line. (B) Side-view of RIN structure (red) and the LapA structural model by Phyre2 (green). The β-strand numbers are indicated and the N and C termini are marked. (C) Top-down view of the RIN structure (red) and the LapA structural model (green). (D) Weblogo plot produced from an amino-acid sequence alignment between RINM and the N-terminal sequences of 10 different putative RTX adhesins (listed in Table 1) from Gram-negative species such as *Shewanella oneidensis* (GenBank Identifier: WP_011073976), *Pseudomonas fluorescens* (GenBank Identifier: ABA71877) and *Vibrio cholerae* (GenBank Identifier: ZP_04408076.1). Conserved secondary structural elements are indicated by red horizontal arrows (β-strands) or a red corkscrew (α-helix). The putative LapG cleavage site is indicated by a red arrowhead. Consensus sequences are shown in large, tall letters, while gaps in the alignment are indicated by blank space.

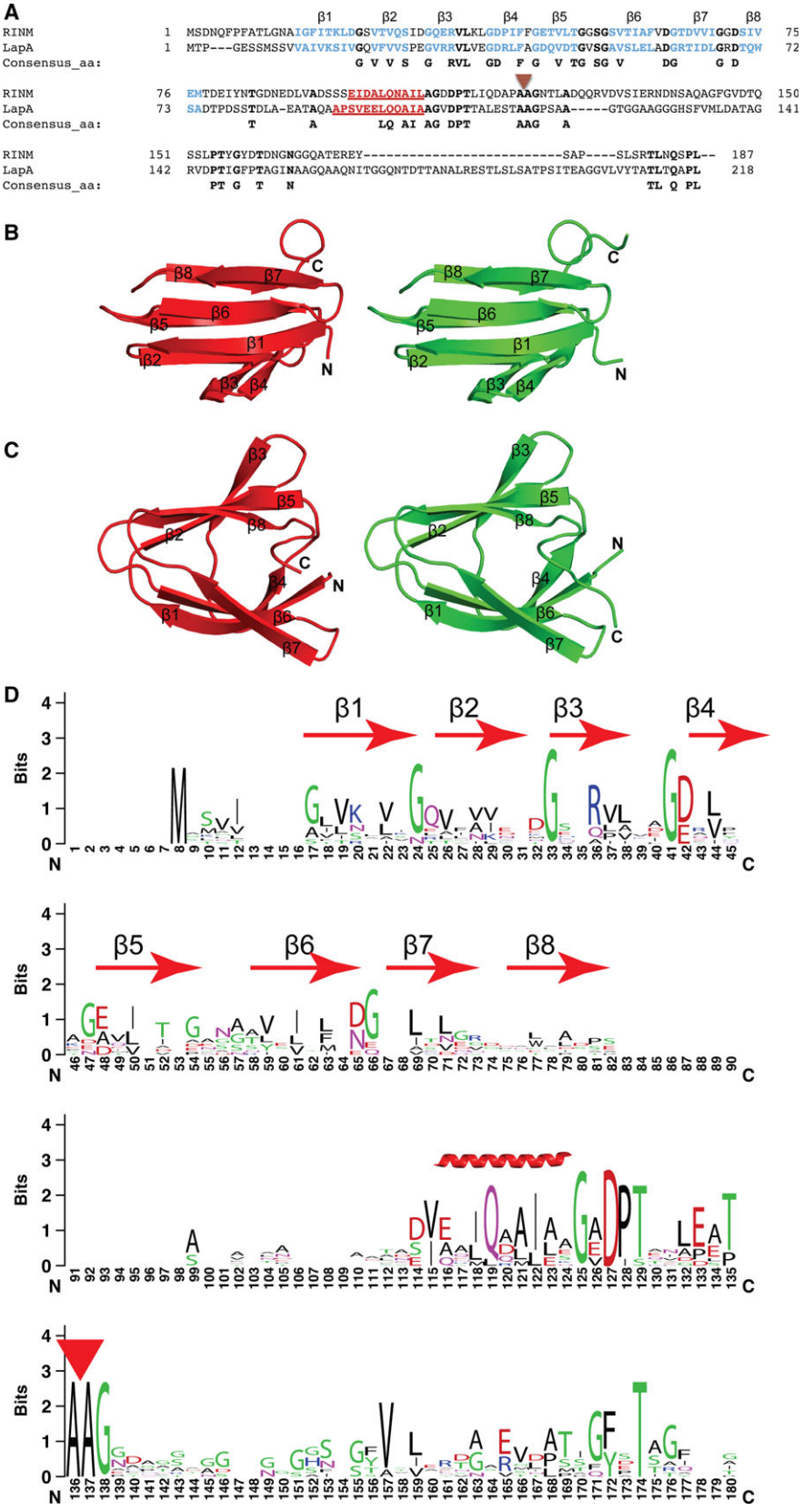
A
β1 β2 β3 β4 β5 β6 β7 β8
RINM 1 MSDNQFPFATLGNAIGFITKLDGSVTVQSIDGQERVLKLGDPIFFGETVLTGGSGSVTIAFVDGTDVVIGGDSIV 75
LapA 1 MTP---GESSMSSVVAIVKSIVGQVFVVSPEGVRRVLVEGDRLFAGDQVDTGVSGAVSLELADGRTIDLGRDIQW 72
Consensus_aa: G V V S G RVL GD F G V TG SG V DG G D
RINM 76 EMTDEIYNTGDNEDLVADSSSEIDALQNAILAGDDPTLIQDAPAAGNTLADQQRVDVSIERNDNSAQAGFGVDTQ 150
LapA 73 SADTPDSSTDLA-EATAQAAPSVEELQQAIAAGVDPTTALESTAAGPSAA-----GTGGAAGGGHSFVMLDATAG 141
Consensus_aa: T A LQ AI AG DPT AAG A
RINM 151 SSLPTYGYDTDNGNGGQATEREY----------SAP----SLSRTLNQSPL-- 187
LapA 142 RVDPTIGFPTAGINAAGQAAQNITGGQNTDTTANALRESTLSLSATPSITEAGGVLVYTATLTQAPL 218
Consensus_aa: PT G T N TL Q PL
B
C
D
Bits
N
C
β1 β2 β3 β4
β5 β6 β7 β8

Table 1. RI homologs from representative RTX adhesins produced by 10 different Gram-negative bacteria.

Bacteria	GenBank Identifier	Length (aa)	Identity to *Mp*IBP_RINM (%)	Confidence score of Phyre2 model using RINM as template (%)
Pseudomonas fluorescens	ABA71877	5218	28.07	99.5
Shewanella baltica	WP_011848055	2239	24.71	98.4
Shewanella oneidensis	NP_719842	2768	25.44	98
Pectobacterium atrosepticum	CAG76164	4558	30.06	99.5
Oxalobacter formigenes	ZP_04578419	2801	26.59	99.5
Sulfurospirillum deleyianum	WP_01285659	1538	24.85	99.6
Neptuniibacter caesariensis	ZP_01166963	2550	28.30	99.6
Pseudomonas entomophila	WP_01153500	2350	28.47	99.6
Vibrio parahaemolyticus	WP_08667751	5227	20.71	98.9
Vibrio cholerae	ZP_04408076	2821	25	97

Mechanism of cell-surface retention by the RI domains

As described above, extensive proteolysis occurred in the region of RIM (ca. 90–180 aa) in both the full-length RI and RINM constructs. Therefore, it was not possible to solve the structure of RIM by NMR or X-ray crystallography because of the long time-frame of these experiments. Nevertheless, structural information on RIM was crucial for understanding how the RI domains anchor the giant adhesin in the bacterial membrane. The low-resolution solution structure of RI was measured rapidly with SAXS before proteolysis occurred. We have previously shown that the SAXS envelope of RIM is a narrow cylinder in solution with a diameter of approximately 18 Å and a length of 40 Å [2], which matches the thickness of the OM of Gram-negative bacteria. Given that RIN resides in the periplasmic space while RIC is extracellular, we proposed that RIM spans the OM. RIM has a high content of charged residues and is not predicted to be a transmembrane domain. It is likely that RIM interacts with an outer-membrane-anchoring component such as a pore protein.

*Mp*IBP is transported to the bacterial surface by T1SS, which is composed of three machinery proteins [25–30]. The membrane fusion protein and an ATP-binding cassette (ABC) transporter form a stable inner-membrane (IM) translocase complex that makes the initial contact with a T1SS substrate. This subsequently triggers the recruitment of the T1SS outer-membrane protein (OMP), TolC. The three proteins form a contiguous channel across the cell envelope to directly translocate the substrate from the cytosol to the extracellular milieu (Fig. 7A–D) [31,32]. *Mp*IBP needs to pass through the constriction point of T1SS, which is the outer-membrane β-barrel pore of TolC with an internal diameter of ~ 20 Å [26,27] (Fig. 7E). While *Mp*IBP domains on the C-terminal side of RIM are threaded through TolC in their unfolded states due to the lack of Ca^{2+}, RIM fits snugly into the interior of the β-barrel (Fig. 7C,D), which is also composed of predominantly polar residues (Fig. 7E). In contrast, steric hindrance prevents the passage of RIN, which folds in the absence of Ca^{2+} and, with dimensions of 30 Å × 29 Å, is too large to pass through the β-barrel pore (Fig. 7E,F). Secondary structure analysis suggests that the first 11 residues of RIN (MSDNQ**F**P**F**ATL, Fig. 7F; not observed in the HSQC spectrum) fold as a random coil [24]. Given that both termini of RIN point toward RIM near the OM [2], it is possible that the two N-terminal phenylalanine residues directly interact with the membrane. Interestingly, the hydrophobic patch comprised outward-pointing phenylalanine and proline residues of RIN is near the protein's N terminus, and thus could also play a role in stabilizing the anchoring. After *Mp*IBP passes the IM translocase and plugs TolC with RINM, the T1SS channel will disassemble (Fig. 7D). This will make the putative LapG cleavage site in the periplasmic space more accessible for the proteolytic processing as seen with LapA.

It should be noted that RIN-like domains and the putative LapG cleavage site are absent from several known surface-associated RTX adhesins that transit via T1SS [4,33]. This suggests that these adhesins are regulated by a different release/biofilm-dispersal mechanism. For example, the RTX adhesin SiiE is loosely associated with the cell surface of *Salmonella enterica*; however, the type of proteolytic processing that occurs in LapA was not observed in SiiE, and its release mechanism is currently unknown [33–35]. It is predicted that a N-terminal coiled-coil domain anchors

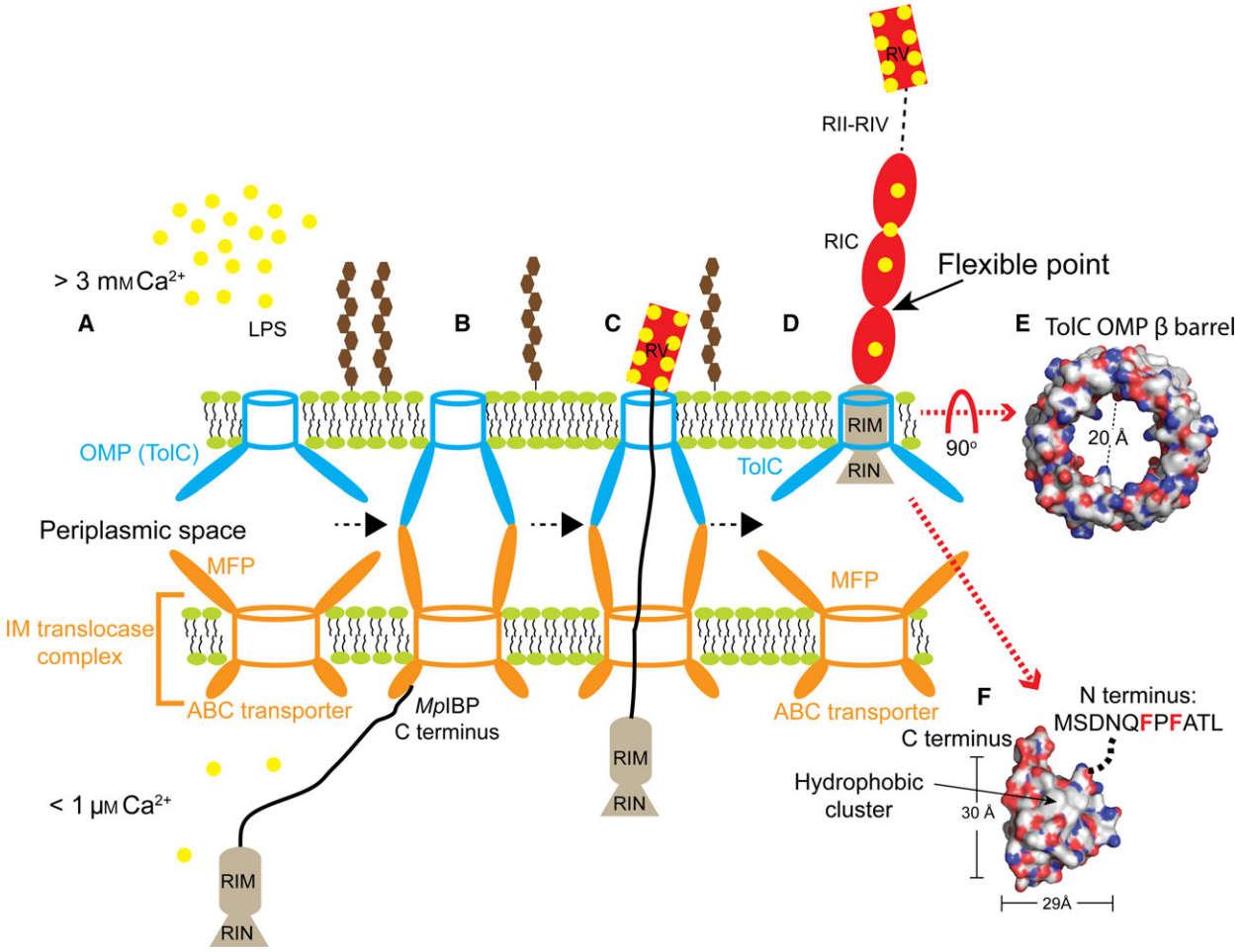

Fig. 7. Proposed mechanism of *Mp*IBP cell-surface retention, and the assembly/disassembly processes of the T1SS channel. (A) In T1SS, a membrane fusion protein (MFP) and an ABC transporter form a stable translocase complex anchored to the inner-membrane. (B) Upon contact with the unstructured C-terminal sequence of *Mp*IBP or another T1SS substrate, this translocase complex undergoes conformational changes to recruit the T1SS OMP, TolC, and form a continuous channel across the cell envelope. (C) Most of the *Mp*IBP sequence remains unstructured in the secretion channel, but folds upon entering the Ca^{2+}-rich extracellular environment. In contrast, the N-terminal RIN and RIM 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 RIM inserted in the TolC pore, while RIN prevents the total release of *Mp*IBP by steric hindrance. (E) Modeled *Marinomonas primoryensis* T1SS OMP is rotated 90° around a horizontal axis to show the size of its polar interior. The structure is shown in surface representation with O atoms in red and N atoms in blue. (F) RIN shown in surface representation. The positions of the termini and the hydrophobic patch are marked. The N-terminal residues not observed in the NMR structure are represented by a dashed line on the structure, and the amino-acid sequence is shown with the two Phe residues indicated by red bold letters. Ca^{2+} ions are indicated by small yellow spheres.

SiiE by interacting with a T1SS subunit [31]. Although SiiE has no obvious homology to RINM (sequence identity: 20.8%; Phyre2 confidence score: 6.4%), this interaction with the T1SS protein would also result in the plugging of the secretion channel in a similarly manner as that of RINM. In Gram-negative bacteria, T1SS is not only an important pathway for protein export, but also is a crucial defense mechanism by helping bacteria expel bactericidal small molecules [28]. It seems counterintuitive that RTX adhesins would hijack T1SS channels for their cell-surface presentation. Nevertheless, biofilms are often the preferred life mode of many bacteria, and the selective benefit offered by having RTX adhesins on the cell surface could compensate for the loss of a few functional T1SS. On the other hand, given that only a small number of TolC are blocked, the remaining T1SS IM translocase complexes may still function with additional copies of TolC.

Conclusions and outlook

In conclusion, our detailed structural analyses of the *Mp*IBP_RI domains reveal a simple yet elegant mechanism for the surface retention of the 1.5-MDa *Mp*IBP. This proposed mechanism could be widely used by other T1SS BAPs found in many different Gram-negative bacteria, including those of human and animal pathogens. While most parts of BAPs require Ca^{2+} to fold, their N-terminal domains (RINM in the case of *Mp*IBP) fold independently of the divalent ions and acts as a 'stopper' during the secretion, to hold the large protein in the OM of the bacteria. In our model, *Mp*IBP_RIM interacts with the OMP of T1SS, and it is likely cleaved under nutrient-depleted conditions to help free bacteria from biofilms when that location becomes unfavorable. The molecular characterization of extremely large and repetitive RTX adhesins is challenging, and many of its key members remain to be studied. Taken together with our characterization on *Mp*IBP_RIN, future structural studies of SiiE and other RTX adhesins will help answer key questions about the regulation of their surface retention to help bacteria form biofilms, and their release that triggers bacterial dispersal from the biofilms.

Materials and methods

Expression, purification, and proteolysis of the RI constructs

Genes encoding the RI and RINM proteins were ligated between the *NdeI*/*XhoI* sites of the pET24a vector, placing a C-terminal 6X His-tag on each protein. Expression and purification procedures were followed as previously described [2,14,17]. After its production in BL21 cells, the RI protein was purified with one round of Ni^{2+}-NTA affinity chromatography followed by one round of size-exclusion chromatography in the presence of 5 mM $CaCl_2$. The full-length RI protein was used in crystallization trials, where it degraded into RIC and crystallized in ~ 2 weeks. Sample crystals were harvested, washed with the mother liquor, and analyzed by SDS/PAGE.

To prevent the rapid degradation of full-length RI before SAXS measurements, the purified protein was flash-frozen in liquid nitrogen and subsequently stored in a freezer at −80 °C until the day of the experiment. The protein remained frozen on dry ice during transportation to the European Synchrotron Radiation Facility (Grenoble, France), where SAXS data were rapidly acquired (for each sample, 10 frames of 0.1 s were collected). Detailed procedures were reported in previous protocols [2,36,37].

Purification of RINM followed the same protocols used for RI. Purified RINM was degraded into RIN in approximately 15 days at room temperature using endogenous proteolysis. One round of anion-exchange chromatography was used to purify RIN from degradation products. The structure of RIN was determined by NMR following the previously published protocol [2,38] N-terminal sequencing of RIN was performed by Edman degradation analysis at the Hospital for Sick Children in Toronto, Canada. Detailed procedures for structure determination by X-ray Crystallography, NMR, and SAXS have been previously described [2].

Bioinformatics and structural homology analyses

Illustrations of protein structure were generated by PyMOL [39]. Bioinformatics analyses were performed using on-line tools: amino-acid sequence alignments were generated using the PROMALS and the Weblogo servers [40]. Secondary structural prediction was performed using Psipred. Structural homology analyses were done using the Dali and Phyre2 servers [41,42].

Acknowledgements

We thank Mrs Sherry Gauthier and Dr Qilu Ye for their assistance with cloning and crystallization trials. We are grateful to staff at the 08ID-1 beamline of Canadian Light Source (Saskatoon, Canada), the X6A beamline at the National Synchrotron Light Source (Brookhaven National Laboratory, NY, USA), and the BM29 beamline at the European Synchrotron Radiation Facilities (Grenoble, France) for access to the synchrotron facilities, and for their help with acquiring X-ray crystallographic and SAXS data. This work was funded by Natural Sciences and Engineering Research Council of Canada discovery grants to SPS (RGPIN 2015-06667) and PLD (RGPIN 2016-04810), by the European Union through an ERC grant (ERC-2014-StG Contract No. 635928) to IKV, and by a Canadian Institutes of Health Research operating grant to PLD (106612). PLD holds the Canada Research Chair in Protein Engineering.

Author contributions

SG and PLD conceived the study. SG designed the experiments and wrote the manuscript with editorial input from PLD. SG and SWP performed crystallization, data collection, and structure determination of the X-ray crystal structures. SG and DNL performed NMR data collection and structure determination. SPS contributed to the NMR data interpretation. SG and IKV performed SAXS data collection and analyses. All authors contributed to writing and editing the drafts of the manuscript.

References

1 Satchell KJ (2011) Structure and function of MARTX toxins and other large repetitive RTX proteins. *Annu Rev Microbiol* **65**, 71–90.

2 Guo S, Stevens CA, Vance TDR, Olijve LLC, Graham LA, Campbell RL, Yazdi SR, Escobedo C, Bar-Dolev M, Yashunsky V *et al.* (2017) Structure of a 1.5-MDa adhesin that binds its Antarctic bacterium to diatoms and ice. *Sci Adv* **3**, e1701440.

3 Yoshida K, Toyofuku M, Obana N & Nomura N (2017) Biofilm formation by *Paracoccus denitrificans* requires a type I secretion system-dependent adhesin BapA. *FEMS Microbiol Lett* **364**, fnx029.

4 Martinez-Gil M, Yousef-Coronado F & Espinosa-Urgel M (2010) LapF, the second largest Pseudomonas putida protein, contributes to plant root colonization and determines biofilm architecture. *Mol Microbiol* **77**, 549–561.

5 De Gregorio E, Del Franco M, Martinucci M, Roscetto E, Zarrilli R & Di Nocera PP (2015) Biofilm-associated proteins: news from Acinetobacter. *BMC Genom* **16**, 933.

6 Syed KA, Beyhan S, Correa N, Queen J, Liu J, Peng F, Satchell KJ, Yildiz F & Klose KE (2009) The Vibrio cholerae flagellar regulatory hierarchy controls expression of virulence factors. *J Bacteriol* **191**, 6555–6570.

7 Romling U & Balsalobre C (2012) Biofilm infections, their resilience to therapy and innovative treatment strategies. *J Intern Med* **272**, 541–561.

8 Hall-Stoodley L, Costerton JW & Stoodley P (2004) Bacterial biofilms: from the natural environment to infectious diseases. *Nat Rev Microbiol* **2**, 95–108.

9 Boyd CD, Smith TJ, El-Kirat-Chatel S, Newell PD, Dufrene YF & O'Toole GA (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.

10 Newell PD, Boyd CD, Sondermann H & O'Toole GA (2011) A c-di-GMP effector system controls cell adhesion by inside-out signaling and surface protein cleavage. *PLoS Biol* **9**, e1000587.

11 Navarro MV, Newell PD, Krasteva PV, Chatterjee D, Madden DR, O'Toole GA & Sondermann H (2011) Structural basis for c-di-GMP-mediated inside-out signaling controlling periplasmic proteolysis. *PLoS Biol* **9**, e1000588.

12 Gilbert JA, Davies PL & Laybourn-Parry J (2005) A hyperactive, Ca^{2+}-dependent antifreeze protein in an Antarctic bacterium. *FEMS Microbiol Lett* **245**, 67–72.

13 Guo SQ, Garnham CP, Whitney JC, Graham LA & Davies PL (2012) Re-evaluation of a bacterial antifreeze protein as an adhesin with ice-binding activity. *PLoS One* **7**, e48805.

14 Guo SQ, Garnham CP, Partha SK, Campbell RL, Allingham JS & Davies PL (2013) Role of Ca^{2+} in folding the tandem beta-sandwich extender domains of a bacterial ice-binding adhesin. *FEBS J* **280**, 5919–5932.

15 Vance TDR, Olijve LLC, Campbell RL, Voets IK, Davies PL & Guo SQ (2014) Ca^{2+}-stabilized adhesin helps an Antarctic bacterium reach out and bind ice. *Biosci Rep* **34**, 357-U218.

16 Garnham CP, Campbell RL & Davies PL (2011) Anchored clathrate waters bind antifreeze proteins to ice. *Proc Natl Acad Sci U S A* **108**, 7363–7367.

17 Garnham CP, Gilbert JA, Hartman CP, Campbell RL, Laybourn-Parry J & Davies PL (2008) A Ca^{2+}-dependent bacterial antifreeze protein domain has a novel beta-helical ice-binding fold. *Biochem J* **411**, 171–180.

18 Bar Dolev M, Bernheim R, Guo S, Davies PL & Braslavsky I (2016) Putting life on ice: bacteria that bind to frozen water. *J R Soc Interface* **13**, 20160210.

19 Davies PL (2014) Ice-binding proteins: a remarkable diversity of structures for stopping and starting ice growth. *Trends Biochem Sci* **39**, 548–555.

20 Rath A, Glibowicka M, Nadeau VG, Chen G & Deber CM (2009) Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proc Natl Acad Sci U S A* **106**, 1760–1765.

21 Aravind P, Mishra A, Suman SK, Jobby MK, Sankaranarayanan R & Sharma Y (2009) The betagamma-crystallin superfamily contains a universal motif for binding calcium. *Biochemistry* **48**, 12180–12190.

22 Shimeld SM, Purkiss AG, Dirks RP, Bateman OA, Slingsby C & Lubsen NH (2005) Urochordate betagamma-crystallin and the evolutionary origin of the vertebrate eye lens. *Curr Biol* **15**, 1684–1689.

23 Crooks GE, Hon G, Chandonia JM & Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* **14**, 1188–1190.

24 Buchan DW, Minneci F, Nugent TC, Bryson K & Jones DT (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res* **41**, W349–W357.

25 Delepelaire P (2004) Type I secretion in gram-negative bacteria. *Biochim Biophys Acta* **1694**, 149–161.

26 Koronakis V, Eswaran J & Hughes C (2004) Structure and function of TolC: the bacterial exit duct for proteins and drugs. *Annu Rev Biochem* **73**, 467–489.

27 Koronakis V, Sharff A, Koronakis E, Luisi B & Hughes C (2000) Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* **405**, 914–919.

28 Costa TR, Felisberto-Rodrigues C, Meir A, Prevost MS, Redzej A, Trokter M & Waksman G (2015) Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat Rev Microbiol* **13**, 343–359.

29 Kanonenberg K, Schwarz CKW & Schmitt L (2013) Type I secretion systems – a story of appendices. *Res Microbiol* **164**, 596–604.
30 Lenders MHH, Reimann S, Smits SHJ & Schmitt L (2013) Molecular insights into type I secretion systems. *Biol Chem* **394**, 1371–1384.
31 Balakrishnan L, Hughes C & Koronakis V (2001) Substrate-triggered recruitment of the TolC channel-tunnel during type I export of hemolysin by *Escherichia coli*. *J Mol Biol* **313**, 501–510.
32 Thanabalu T, Koronakis E, Hughes C & Koronakis V (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.
33 Wagner C, Polke M, Gerlach RG, Linke D, Stierhof YD, Schwarz H & Hensel M (2011) Functional dissection of SiiE, a giant non-fimbrial adhesin of *Salmonella enterica*. *Cell Microbiol* **13**, 1286–1301.
34 Barlag B & Hensel M (2015) The giant adhesin SiiE of *Salmonella enterica*. *Molecules* **20**, 1134–1150.
35 Peters B, Stein J, Klingl S, Sander N, Sandmann A, Taccardi N, Sticht H, Gerlach RG, Muller YA & Hensel M (2017) Structural and functional dissection reveals distinct roles of Ca^{2+}-binding sites in the giant adhesin SiiE of *Salmonella enterica*. *PLoS Pathog* **13**, e1006418.
36 Olijve L, Sun TJ, Narayanan T, Jud C, Davies PL, Voets IK & Meijer EW (2013) Solution structure of hyperactive type I antifreeze protein. Abstracts of Papers of the American Chemical Society. 245.
37 Pernot P, Round A, Barrett R, Antolinos AD, Gobbo A, Gordon E, Huet J, Kieffer J, Lentini M, Mattenet M *et al*. (2013) Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *J Synchrotron Radiat* **20**, 660–664.
38 Grondin JM, Chitayat S, Ficko-Blean E, Boraston AB & Smith SP (2012) 1H, 15N and 13C backbone and side-chain resonance assignments of a family 32 carbohydrate-binding module from the *Clostridium perfringens* NagH. *Biomol NMR Assign* **6**, 139–142.
39 Alexander N, Woetzel N & Meiler J (2011) Bcl:Cluster: a method for clustering biological molecules coupled with visualization in the Pymol Molecular Graphics System. *IEEE Int Conf Comput Adv Bio Med Sci* **2011**, 13–18.
40 Pei J, Kim BH & Grishin NV (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res* **36**, 2295–2300.
41 Holm L & Rosenstrom P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545–W549.
42 Kelley LA, Mezulis S, Yates CM, Wass MN & Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845–858.