DR. HAN REMAUT (Orcid ID : 0000-0002-9775-4102)

Article type : Research Article

Hop-family *Helicobacter* outer membrane adhesins form a novel class of Type 5-like secretion proteins with an interrupted β-barrel domain

Fanny Coppens*[1,2], Gaetano Castaldo*[1,2], Ayla Debraekeleer[1,2], Suresh Subedi[1,2], Kristof Moonens[1,2], Alvin Lo[1,2], and Han Remaut[1,2]

[1]*Structural and Molecular Microbiology, VIB-VUB Center for Structural Biology, VIB, Pleinlaan 2, 1050 Brussels, Belgium.*

[2]*Structural Biology Brussels, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium*

[*] *These authors contributed equally.*

Corresponding author: Han Remaut, Vrije Universiteit Brussel, Building E4, Pleinlaan 2, 1050 Brussels, Belgium. Tel: +32-2-6291923, Fax: +32-2-6291963, e-mail: han.remaut@vub.be

Running title: Novel class of Type V secretion proteins in H. pylori

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mmi.14075

Keywords: Helicobacter pylori, adhesin, autotransporter, protein secretion, outer membrane, Type V secretion.

Summary

The human stomach pathogen *Helicobacter pylori* attaches to healthy and inflamed gastric tissue through members of a paralogous family of "*Helicobacter* outer membrane proteins" (Hops), including adhesins BabA, SabA, HopQ, LabA and HopZ. Hops share a conserved 25 kDa C-terminal region that is thought to form an autotransporter-like transmembrane domain. Instead, our results show that Hops contain a non-continuous transmembrane domain, composed of seven predicted β-strands at the C-terminus and one at the N-terminus. Folding and outer membrane localization of the C-terminal β-domain critically depends on a predicted transmembrane β-strand within the first 16 N-terminal residues. The N-terminus is shown to reside in the periplasm, and our crystal and small angle X-ray scattering structures for the SabA extracellular domain reveal a conserved coiled-coil stem domain that connects to transmembrane β-strand 1 and 2. Taken together, our data show that Hop adhesins represent a novel outer membrane protein topology encompassing an OmpA-like 8-stranded β-barrel that is interrupted by a 15 to 108 kDa domain inserted inside the first extracellular loop. The insertion of large, folded domains in an extracellular loop is unprecedented in bacterial outer membrane proteins and is expected to have important consequences on how these proteins reach the cell surface.

Introduction

Helicobacter pylori is a major bacterial pathogen that infects approximately half of the world's population. The bacterium chronically infects the gastroduodenal tract as a non-invasive pathogen and is found in an intimate contact with the gastric mucosa through adherence to the gastric epithelial cells and the above mucus layer. Infection outcome ranges from mild asymptomatic inflammation to

chronic gastritis that can develop to peptic ulcers and/or gastric adenocarcinoma and mucosa associated lymphoid tissue (MALT) lymphomas in more virulent strains (Uemura *et al.*, 2001; Peek and Blaser, 2002). *H. pylori* is a predominantly primate tropic pathogen in which the blood group antigen binding adhesin BabA and the sialic acid binding adhesin SabA form two major virulence factors, binding respectively, Lewis b (Leb) and related ABO blood group antigens found in the gastric epithelium and secreted overlying mucus, or sialyl-Lewis x (sLex), sialyl-dimeric-Lewis x (sdiLex) and sialyl-Lewis a (sLea) glycosphingolipids more associated with inflamed gastric epithelia (Ilver *et al.*, 1998; Mahdavi *et al.*, 2002).

BabA (HopS) and SabA (HopP) are members of a paralogous family of 33 outer membrane proteins classified into Hops (*Helicobacter* outer membrane proteins) and Hors (Hop-related) on the basis of domain build-up and sequence similarity (Alm *et al.*, 2000) (Fig. 1A). The Hop and Hor family proteins range in size from 186 to 1237 residues and are characterized by a remarkable sequence conservation in the first ~20 residues of the mature proteins and in the C-terminal ~180 residues (Fig. 1) (Alm *et al.*, 2000). Both adhesins belong to the Hop family, which contains a A↓EX[D/N]G (where ↓ marks the signal peptidase cleavage site) signature sequence following the signal peptide and which shows large domain insertions between the conserved N- and C-terminal regions, ranging in size between ~150 and ~1000 residues. These insertions are variable in sequence and are predicted to contain the protein-specific functional features, such as the adhesive properties in case of BabA and SabA. At least three additional proteins in the phylogenetic cluster holding BabA and SabA have also been described as adhesins HopZ, LabA (HopD) and HopQ (Fig. 1B) (Peck *et al.*, 1999; Rossez *et al.*, 2014; Javaheri *et al.*, 2016; Königer *et al.*, 2016). For many Hop and Hor proteins the function is unknown and their number and distribution varies strongly amongst different *H. pylori* strains.

The BabA and SabA primary structures are reminiscent of the modular organization found in classical autotransporters (ATs) (Type 5a secretion proteins (SPs)) (Grijpstra *et al.*, 2013), a widespread group of outer membrane proteins that are composed of an N-terminal extracellular "passenger" domain of 20 to 400 kDa and an approximately 30 kDa C-terminal translocator domain (β-domain) with a 12-stranded β-barrel structure that is involved in passenger secretion, and anchors the protein in the outer membrane or releases it on the cell surface after autoproteolytic cleavage (Dautin and Bernstein, 2007; Leyton *et al.*, 2012). Hops, however, appear to deviate in multiple aspects. Secondary structure predictions and the recent structures of the BabA, SabA and HopQ ectodomains (Pang *et al.*, 2013; Hage *et al.*, 2015; Moonens *et al.*, 2016; Javaheri *et al.*, 2016) show the passenger domains to be predominantly α-helical in nature. Available structures for AT passengers, on the other hand, reveal that at least the regions proximal to the β-domain are high in β-strand content and usually fold in β-solenoids, shown to be important for the secretion efficiency of the passenger domain (Peterson *et al.*, 2010). Furthermore, the Hop extracellular domains are rich in disulfides (BabA and SabA contain 4 and 2 cystines, respectively), unlike most ATs, where the insertion of cysteines in the passenger can lead to a stalled transport due to the formation of disulfide-bound blocks (Leyton *et al.*, 2011). Finally, the apparent membrane topology of the Hop C-terminal domain deviates from that seen in AT structures. For Hops, multiple sequence alignments and transmembrane (TM) β-strand predictions indicate the presence of 7 β-strands in the C-terminal domain (Alm *et al.*, 2000). Instead, available AT structures share a common 12-stranded β-barrel architecture, composed of 12 consecutive β-strands in case of monomeric ATs or consisting of a composite β-barrel formed by three 4-stranded β-sheets in case of trimeric ATs (Oomen *et al.*, 2004; Meng *et al.*, 2006; Leyton *et al.*, 2012). AT extracellular passengers are connected to the downstream β-domain through a short linking α-helix that traverses the interior of the β-barrel domain.

To address these intriguing observations, we examined the structural buildup of Hop transmembrane domains through systematic mutational analyses of BabA and SabA and show these contain an unprecedented topology, composed of an interrupted 8-stranded β-barrel formed by the N-terminal 16 residues of the mature proteins and 7 predicted β-strands in the C-terminal domain.

Results

BabA does not contain an autonomous C-terminal β-domain.

For monomeric and trimeric ATs, N-terminal truncation mutants that lack the passenger domain are found to localize at the outer membrane as stable β-barrels (Oomen *et al.*, 2004; Meng *et al.*, 2006). Following an analogous approach, we constructed a series of BabA truncates (using *babA2* from *H. pylori* strain CCUG17875) that lack the region upstream of the apparent C-terminal β-domain in order to produce a stable fragment corresponding to the putative translocator domain. Full length BabA ($BabA^{1-721}$) as well as four BabA N-terminal truncates ($BabA^{440-721}$, $BabA^{479-721}$, $BabA^{495-721}$ and $BabA^{522-721}$) were cloned in the vector pASK-Iba12 downstream of the OmpA leader sequence and an N-terminal StrepII tag, and were recombinantly expressed in *E. coli* BL21 (DE3). For structural integrity of the truncates, the points of N-terminal deletion were chosen to fall in between predicted secondary structure elements in the passenger region proximal to the C-terminal β-domain. To monitor the correct sorting of $BabA^{1-721}$ and the N-terminal truncates to the *E. coli* cell envelope, inner and outer membrane protein fractions were separated by selective solubilization of the cytoplasmic membrane with 0.5% N-laurylsarcosine and subsequent extraction of the OMPs with 1% N-dodecyl-β-D-maltopyranoside (DDM). Most outer membrane β-barrel proteins remain folded in SDS-containing buffers at physiological temperatures (Reithmeier and Bragg, 1974). We therefore used heat modifiable migration of the various BabA constructs under standard denaturing and seminative conditions as a read-out for protein conformation. Recombinant $BabA^{1-721}$ was found to partially partition to the OM fraction, as well as to the insoluble fraction (Fig. 1C). When heat-treated, the protein extracted from the outer membrane ran at molecular mass (MW) as expected for monomeric unfolded BabA (78 kDa). The non-heated sample, however, shifted to an apparent MW of ~150 kDa, suggesting it ran as a folded, oligomeric species. An equivalent non-covalent oligomer is observed for the high affinity conformation of native BabA isolated from *H. pylori* outer membranes (Moonens *et al.*, 2016; Bugaytsova *et al.*, 2017). DDM-extracted BabA purified from the *E. coli* OM also shows a nanomolar avidity similar to the native protein (Moonens *et al.*, 2016). Intact cells expressing

BabA$^{1-721}$ were positive for binding with the BabA ligand Leb or a nanobody binding a conformational epitope in the BabA adhesin domain (Nb14; Moonens *et al.*, 2016), demonstrating that at least for part of OM-sorted BabA, the adhesin domain is folded, active and reached the extracellular surface (Supplementary Fig. 1).

In contrast, although recombinant expression of the BabA truncates yielded products of the expected molecular mass, none of the truncates was found in the outer membrane extracts (Fig. 2A). Instead, expression products were found associated with the inner membrane fraction or in the insoluble fraction recovered after the extraction with DDM. Seminative SDS-PAGE analysis revealed these proteins were present in an unfolded conformation (Fig. 2A). These results suggest that, in absence of the N-terminal domain, BabA truncates corresponding to the presumed β-domain were not able to assemble into properly folded, outer membrane-embedded proteins. We next employed an alternative strategy to obtain the BabA C-terminal domain, by inserting a thrombin protease cleavage site followed by a 6xHis tag in a loop region between the predicted passenger and β-domains (Fig. 2B). Similar to native BabA, the resulting BabA mutant (BabA^{441TH}) could be extracted from the outer membrane as a folded β-barrel protein (data not shown). BabA^{441TH} was then subjected to thrombin digestion in order to achieve separation between the StrepII-tagged N-terminus and the now 6xHis-tagged C-terminal domain and was passed over a nickel affinity resin. However, SDS-PAGE and Western Blot analysis with anti-His or anti-StrepII antibodies revealed that the N- and C-terminal fragments in the thrombin-nicked material co-eluted from the affinity resin (Fig. 2B). Thus, following correct assembly at the outer membrane, a proteolytic nick in the interdomain region is not able to separate the N-terminal α-helical domain and the C-terminal β-domain, indicating that at least parts of both domains remain intimately connected into a stable folding unit.

Minimal transmembrane domain of BabA and SabA.

Based on the sequence alignment of the Hop and Hor families, Alm *et al.*, identified seven highly conserved blocks of amphiphatic sequence in the C-terminus that were hypothesized to form transmembrane β-strands (Alm *et al.*, 2000). This differs from known outer membrane β-barrel proteins, which contain an even number of transmembrane β-strands, ranging from 8 to 26 (Fairman *et al.*, 2011). The inability of the N-terminal BabA truncates to form an autonomous outer-membrane located β-barrel and the observation that upon proteolytic nicking the N- and C-terminal regions of BabA remain associated in a stable folding unit led us to hypothesize that the N-terminal region could contain additional transmembrane elements. This was further hinted by the topology prediction of full-length SabA from *H. pylori* J99 by the transmembrane β-barrel protein predictor BOCTOPUS, where the SVM (support vector machines) output indicates that the local structural preference for the first N-terminal residues could correspond to a transmembrane β-strand, in addition to seven transmembrane β-strands predicted at the C-terminus (Fig 3A). A closer examination of a multiple sequence alignment and associated secondary structure prediction of the 33 known Hops and Hors revealed that a highly conserved 16 amino acid stretch starting from the first residue of the mature proteins is predicted to form a β-strand with a motif of alternating hydrophobic and hydrophilic side chains (Fig 3A, B). In combination with the 7 predicted transmembrane elements at the protein C-terminus, this N-terminal region could thus give rise to an 8-stranded β-barrel. In order to verify this hypothesis, we made a minimal N-terminal truncation mutant that removes the conserved 16-terminal residues of the mature protein. Unlike the full-length protein, BabA$^{\Delta 1\text{-}16}$ was not found to be associated with the outer membrane, but instead accumulated in the inner membrane and insoluble fraction (Fig. 4A), similarly to what had been seen for the more extensive N-terminal deletion mutants. These experiments show a critical requirement for the predicted N-terminal β-strand for the overall stability and OM localization of BabA.

We next constructed an internal truncation mutant (BabA$^{\Delta 28\text{-}535}$) where the predicted α-helical region connecting the N-terminal 27 residues with the C-terminal β-strand region is replaced by a short

linker. Contrary to N-terminal deletion mutants that lack the first 16 residues, BabA$^{\Delta 28\text{-}535}$ was found in outer membrane extracts (Fig. 4B). On SDS-PAGE, OM-extracted BabA$^{\Delta 28\text{-}535}$ ran as a single band corresponding to the predicted molecular mass of approximately 28 kDa, and showed a heat mobility shift under seminative conditions, indicative of a folded β-barrel (Fig. 4C). This result showed that BabA$^{\Delta 28\text{-}535}$ could be productively targeted to the outer membrane of *E. coli*, where it folds into a stable β-barrel structure. Similarly, in SabA, deletion of the predicted adhesin domain and connection of the 18 N-terminal residues with the C-terminal β-strand region (residues 449-631) resulted in a SabA fragment that sorted to the OM as a folded beta-barrel (Fig. 4C).

Noteworthy, the faster migrating BabA$^{\Delta 28\text{-}535}$ and SabA$^{\Delta 19\text{-}448}$ are indicative of monomeric β-barrels, in agreement with previous observations of small Hops such as HopA (37 kDa) and HopB (38 kDa) purified from the *H. pylori* OM (Exner *et al.*, 1995). It thus appears that the formation of the oligomeric species observed as a slower running species in case of BabA$^{1\text{-}721}$ (Fig. 1C) depends on the presence of the BabA adhesin domain, and that the BabA oligomer contains individual β-barrels rather than a single composit TM domain. This is further corroborated by previous observations of a pH-sensitive monomer-oligomer equilibrium of native BabA in *H. pylori* outer membranes (Bugaytsova *et al.*, 2017).

Topology for Hop transmembrane domain.

Sequences for full length BabA and SabA or their minimal transmembrane constructs (BabA$^{\Delta 28\text{-}535}$, SabA$^{\Delta 19\text{-}448}$) were submitted to the protein homology/analogy recognition engine (Phyre^2) for fold recognition and template-based homology modelling (Kelley and Sternberg, 2009). The synthetic sequences BabA$^{\Delta 28\text{-}535}$ and SabA$^{\Delta 19\text{-}448}$ were predicted to share structural homology with OmpA-like outer membrane proteins at high confidence levels. Mapping of secondary structure predictions for BabA$^{\Delta 28\text{-}535}$ and SabA$^{\Delta 19\text{-}448}$ and known transmembrane β-strands in *E. coli* OmpA (PDB entry: 1QJP) onto a multiple sequence alignment reiterates a high degree of structural similarity (Fig. 3A). In addition, analysis of amino acid distribution in the OmpA template-based homology model of

BabA$^{\Delta 28\text{-}535}$ shows that lipid-facing regions in the model lack hydrophilic side chains and are lined with a "belt" of aromatic residues as is commonly found in transmembrane β-barrels (Fig. 4E). For the full-length BabA and SabA sequences, the Phyre fold recognition analysis picks up the structural homology of the C-terminal domain with OmpA-like proteins, but results in an open 7-stranded β-barrel. Thus, in agreement with the experimental data for the BabA truncation mutants, the BabA or SabA C-terminal β-barrel domain does not appear to form a stable transmembrane folding unit. Instead, the N-terminal peptide is required to complete an 8-stranded, OmpA-like β-barrel (Fig. 3A).

To experimentally verify the predicted topology of the transmembrane domain, we analysed the surface exposure of affinity tags at the SabA N-terminus or in the loop connecting the N-terminal 18 residues and the C-terminal β-strand region. To do so, we created a minimal transmembrane mutant of SabA by linking the predicted β-strand in the N-terminal 18 residues with the C-terminal β-rich domain by replacement of the α-helical adhesin domain (res 19-448) with a haemagglutinin (HA) tag linker sequence (SabA$^{\Delta 19\text{-}448HA}$). In the predicted topology for the Hop transmembrane domain, the N-terminus resides inside the periplasmic space rather than on the extracellular side as would be the case for Type 5a ATs, whilst the HA tag connector would represent the first extracellular loop (L1) (Fig 4D, E). In agreement with this, immunofluorescence microscopy (IF) on intact *E. coli* BL21 expressing SabA$^{\Delta 19\text{-}448HA}$ showed that the N-terminal StrepII tag is recognized only upon rupture of the OM with 0.5% Triton X-100 (Fig. 4D), whilst the HA epitope tag placed within the predicted extracellular loop L1, was readily visualized by IF on whole cells.

Structure of the Hop ectodomain.

In Type 5a SPs, the passenger and TM domains are connected by an α-helix that traverses the lumen of the β-barrel, and the membrane-proximal region of the extracellular domain is frequently composed of a β-solenoid structure that has been shown to be important for surface translocation of the passenger domain (Ieva *et al.*, 2011). In Hops, the extracellular domains are predominantly α-

helical in nature. The recent X-ray structures of the BabA, SabA and HopQ ectodomains (Pang *et al.*, 2013; Hage *et al.*, 2015; Moonens *et al.*, 2016; Javaheri *et al.*, 2016) show that the extracellular domains of these adhesins share a common buildup composed of a 3+4 helix bundle "head" domain connected to a coiled-coil "stem" domain (Supplementary Fig 2). Mapping the secondary structure elements on a multiple sequence alignment of Hops shows that despite a low sequence identity, this topology is conserved within a monophyletic group of the Hops that encompasses the known adhesins (Supplementary Fig 2). The adhesive properties are formed by insertion segments in the head domain unique to the individual adhesins (Moonens *et al.*, 2016; Javaheri *et al.*, 2016).

To gain further insight in the stem domain connecting the Hop extracellular and TM domains, we determined the X-ray structure of the full extracellular domain of SabA as predicted by our topology assignment above. A gene fragment of *H. pylori* J99 *sabA* corresponding to residues 22-455 ($SabA^{22-455}$) was crystallized and its structure was determined to 2.52 Å resolution. The resolved structure comprises residues 22 to 438 and reveals an L-shaped molecule with the "head" domain spanning residues 56 to 368 (H2-H10), connected by a right angle to a coiled-coil "stem" domain corresponding to residues 22-55 (H1) and 369-438 (H11-βs1-βs2-H12) (Fig. 5A, B). Comparison of the experimental small angle X-ray scattering curves and theoretical scattering data of the $SabA^{22-455}$ crystal structure, shows the X-ray structure closely matches the particle found in solution (Fig. 2A). The $SabA^{22-455}$ structure is in good agreement with the recently reported structure of refolded SabA from *H. pylori* strain 26695 (PDB entry: 4O5J, encompassing residues 26-399; 0.97 Å root mean square displacement for equivalent Cα atoms) (Pang *et al.*, 2013). Compared to the latter, however, our X-ray data for the periplasmic $SabA^{22-455}$ shows additional density corresponding to residues 393-438, which form a β-hairpin that packs against the H1-H11 coiled-coil and ends at its C-terminus in a short helix (H12) that is juxtaposed to helix 1 (H1) (Fig. 5B). Superimposition of the SabA and BabA ectodomains and multiple sequence alignement show that the topology of the stem domain is conserved in the monophyletic group encompassing the Hop adhesins, with the notable exception of HopQ, which lacks the equivalent of H11 and H12 (Fig. 5C, Supplementary Fig. 2). The stem

domains of BabA and SabA show the juxtaposition of the N- and C-termini of the adhesin domains, which would be in agreement with them connecting to the transembrane strands 1 and 2 of the discontinuous β-barrel domain identified above (Fig 5C, D). *In silico* assembly of the SabA$^{22\text{-}455}$ X-ray structures and OmpA-template-based homology model for the SabA minimal transmembrane domain (Phyre, see above) provides a plausible connectivity with one another (Supplementary Fig. 2), bridged by a 10 and 17-amino acid linker at the N- and C-terminal boundary of the ectodomain, respectively.

Discussion

The gastric pathogen *H. pylori* contains a paralogous family of "*Helicobacter* outer membrane proteins" or Hops, including a set of important glycan- and protein-binding adhesins with unique receptor binding profiles, each implicated in various stages of infection. The Hop family is unique to *H. pylori* and shows a remarkable sequence conservation in the C-terminal ~180 residues, encompassing the presumed transmembrane domain. In addition, the Hop N-terminus is composed of a ~20 residues stretch of high sequence conservation. The intermittent variable region (200-1000 residues) contains the adhesive properties that are unique amongst the different Hop members. The conserved N- and C-termini have previously been seen to promote homologous recombination between different adhesins (Bäckström *et al.*, 2004; Colbeck *et al.*, 2006).Here we demonstrate that together, the conserved N- and C-termini come together to form a discontinuous 1+7-stranded transmembrane region that in Hops is disrupted by a large functional α-helical domain at the position of the first extracellular loop. Structural and phylogenetic analysis of Hops shows the presence of three monophyletic groups with distinct architecture in the intermittent domain. A first monophyletic group formed by SabA*, SabB, HopZ*, LabA*, HopQ*, HopM, HopN, BabA*, BabB, BabC and HopA constitutes of the described adhesins (indicated with an asterisk), hereafter referred to as Hop-I subfamily (Fig 1 B, Supplementary Fig. 2). The adhesive properties are located in a variable extracellular domain of 290 – 540 residues that shares a common topology formed of a head and stem domain. A second and third monophyletic group is formed by AlpA, AlpB, HopG and HopF (Hop-II);

and HopK and HopJ (Hop-III), respectively (Supplementary Fig. 2). Our structure-based sequence alignment shows that the putative extracellular region connecting TM1 and TM2 in the Hop-II (250 – 500 residues) and Hop-III (~150 residues) subfamilies is highly deviant from the conserved topology seen in the Hop-I subfamily, lacking both the Hop-I head domain and coiled-coil stem domain. In the Hop-II subfamily, AlpA and AlpB have been shown important for *H. pylori* adhesion to laminin (Odenbreit *et al.*, 1999; Senkovich *et al.*, 2011), but no information is currently available to localize the adhesive properties on the protein. No function has been ascribed to Hop-III members HopK and HopJ. Compared to Hops, Hor proteins share the N-terminal and 7 C-terminal stretches of high sequence conservation, here predicted to form transmembrane elements TM1 and TM2-8, respectively (Fig 1A, Fig 2B). However, Hor proteins lack the extensive sequence insertions in the region connecting TM1 and TM2 that form the functional extracellular domains of Hop-I proteins and possible Hop-II and Hop-III family proteins. Fold recognition and template-based homology modelling using Phyre2 (Kelley and Sternberg, 2009) shows that Hor proteins, as well as Hop sequences lacking the TM1-TM2 spanning region, are readily recognized as OmpA-like 8-stranded β-barrels with high confidence levels. The high level of sequence conservation in Hop and Hor N- and C-termini points to a common origin of the transmembrane domain. A phylogenetic analysis of the Hor and Hop transmembrane regions (ie. lacking the variable region spanning TM1 – TM2) shows a higher degree of sequence diversity in Hor proteins compared to Hops (Fig 1B), suggesting a more recent origin for Hops, or a stronger negative selection on TM diversification in Hops. Additionally, whilst Hor proteins are found in different *Helicobacter* species, Hops are restricted to *H. pylori* and its closest relatives *H. cetorum* and *H. acinonychis* (Bauwens *et al*, unpublished data). Together, these observations suggest that the Hops originated from the insertion of a soluble α-helical domain into the first extracellular region of a Hor. Divergent evolution of the insertion domain then lead to the diversification of the adhesive properties seen amongst Hop-I family members. Future structural analysis of Hop-II and Hop-III family members may reveal whether these families result from a separate insertion event or reflect a reduction of a Hop-I type insertion domain. The adhesive properties in the Hop-I type proteins reside in a 3+4 helix head domain (Pang *et al.*, 2013; Hage *et al.*,

2015; Moonens *et al.*, 2016; Javaheri *et al.*, 2016), that is separated from the TM domain by a conserved stem domain (Fig. 5A, B, Supplementary Fig. 2), except in HopQ, which lacks the stem domain (Supplementary Fig. 2; Javaheri *et al.*, 2016). The stem domain may play a role in preferential orientation of the head domain to optimise its adhesive properties, and may provide a contact zone in Hop oligemerization. In BabA, mutants that alter the proteins' pH responsive monomer-oligomer equilibrium map to the hinge of the head and stem domain (Bugaytsova *et al.*, 2017).

Hop adhesins comprise the first example of outer membrane proteins that contain a folded domain inside an extracellular loop of the β-barrel. Although many OMPs have been found to contain folded domains in periplasmic loops, no example is known of a folded domain in an extracellular loop. This notion brings with it important considerations on how we think of the processes that facilitate the insertion in the outer membrane and in particular the translocation of the extracellular domains across the OM bilayer. The membrane insertion and passenger secretion processes of OMPs and Type 5 SPs are now known to depend on the Bam or Tam outer membrane protein assembly machineries comprising an Omp85-family protein (Voulhoux *et al.*, 2003; Leyton *et al.*, 2012). It is presently unclear whether Hops rely on the Bam machinery for OM insertion and/or passenger translocation, or whether *H. pylori* posseses a cognate assembly machinery for this protein family. It is noteworthy that although full-length BabA could be functionally expressed in the *E. coli* OM, this constituted a minor fraction of the expressed material, most of which still accumulated into the insoluble fraction (Fig. 1 C). We found that compared to the full-length proteins, the BabA and SabA minimal transmembrane domains, in which the α-helical passengers were deleted, showed an increased efficiency of OM insertion, suggesting that the observed difficulty in OM insertion of full-length Hop adhesins in *E. coli* may at least in part result from their unusual topology and the presence of a large cysteine-rich α-helical passenger domain in an extracellular loop. Compared to the *E. coli* Bam machinery that is composed of five components (BamABCDE), *H. pylori* posseses a minimal Bam machinery that consists of the BamA and BamD components only (Liechti and Goldberg, 2012; Anwari *et al.*, 2012; Webb *et al.*, 2017). Unfortunately, our attempts to test whether recombinant

expression of *H. pylori* BamA and BamD in *E. coli* would increase the OM sorting efficiency of full length Hops have failed due to an inability to express *Hp*BamA in *E. coli*, as also reported by Browning *et al.* (Browning *et al*., 2015). Whether *H. pylori*'s rudimentary Bam machinery and/or the high degree of sequence conservation in the Hop C-terminal β-domains are an adaptation to the secretion of these unusual OMPs, or whether instead there is a dedicated assembly machinery for these proteins, will require future characterization of the Hops assembly process. Given the high sequence conservation in the TM domain, a further understanding of the Hop biogenesis pathway holds promise for the future development of broad-acting assembly inhibitors for this large group of important *H. pylori* colonization factors.

Finally, a systematic nomenclature has been proposed to classify Type 5 secretion system proteins based on their topological features (Fig. 5D) (Grijpstra *et al.*, 2013). It is unclear whether these various Type5SS proteins share a common phylogeny or originated from unrelated OMPs. Our phylogenetic and structural analysis suggest *H. pylori* Hop proteins originated from the insertion of a folded domain into an extracellular loop of Hor proteins, OmpA-like 8-stranded β-barrel proteins unique to the genus *Helicobacter*. Based on the new insights in the structural build-up of Hops, we propose this family of proteins to form a novel family of Type 5-like outer membrane proteins, refered to as Type 5f secretion proteins.

Experimental Procedures

Bacterial strains and culture conditions. Bacterial laboratory strains used in these studies include . *coli* TOP10 and *E. coli* BL21 (DE3). For localization experiments, 200 ml cultures were grown in LB broth supplemented with 100 µg ml^{-1}ampicillin at 26 °C and induced with 0.2-µg ml^{-1} anhydrotetracyclin at an OD_{600} value of 1.0. After 6 hours of induction, the cells were recovered, washed in TBS (20 mM Tris pH 8.0, 150 mM NaCl), weighed and stored at -80 °C. For the crystals

and solution X-ray structures of the SabA adhesin domain, a construct comprising residues 22 to 455 of SabA from *H. pylori* J99 was cloned in pASK-HR3 (pASK-Iba12 modified to replace the OmpA leader with the DsbA leader sequence) and expressed in *E. coli* BL21(DE3) cells. For $SabA^{22\text{-}455}$ expression was induced by addition of 0.2 µg ml^{-1} anhydrotetracyclin and grown overnight at 18°C.

BabA and SabA constructs. Mature BabA coding region (*babA2*) and truncates from *H. pylori* strain CCUG-17875 and derivative constructs were amplified by PCR and cloned into pASK-Iba12 (OmpA leader sequence, N-terminal *StrepII*-tag® and thrombin cleavage site). Sequences corresponding to the transmembrane β-barrel domain of SabA, with ($SabA^{\Delta 19\text{-}448HA}$) or without ($SabA^{\Delta 19\text{-}448}$) insertion of the HA-tag, were cloned from *H. pylori* strain J99 *sabA* into pASK-Iba12. The plasmid constructs generated in this study and the oligonucleotide primers used to do so are listed in the Supplementary Tables S1 and S2, respectively. BabA, SabA or their derivative constructs were expressed and purified from the outer membrane of *E. coli* strain BL21 (DE3) as detailed in the Supplementary methods. All presented residue numbers start from the N-terminus of the mature protein.

Outer membrane protein purification. Cells were resuspended in TBS supplemented with 5 µM leupeptin and 1 mM AEBSF, 20 µg ml^{-1} deoxyribonuclease and 100 µg ml^{-1} lysozyme, lysed by a single passage in a Constant System Cell Cracker at 20 kPsi at 4 °C and debris was removed by centrifugation at 10,000 *g* for 10 minutes. The supernatant underwent ultracentrifugation at 100,000 *g* for 45 minutes. Inner membranes were selectively solubilized by resuspension of total membrane pellets in TBS/0.5% N-lauroylsarcosine supplemented with proteases inhibitors, 5 µM Leupeptin and 1 mM AEBSF, by stirring for 1-2 hours at 4 °C. Outer membranes were pelleted from the suspension by ultracentrifugation at 100,000 *g* for 45 minutes and subsequently with 100 ml of 1 % N-dodecyl- β -D-maltopyranoside (DDM) in 20 mM Tris pH 8.0, 500 mM NaCl, 5% glycerol, 100 µg ml^{-1} lysozyme and 5 µM leupeptin and 1 mM AEBSF. Following a final ultracentrifugation step, the outer membrane extract was loaded at 4 °C on a 5 mL Ni-Sepharose 6 Fast Flow resin (GE

Healthcare) previously equilibrated with 5 column volumes of Buffer A (20 mM Tris pH 8.0, 500 mM NaCl, 5% glycerol, 0.03 % DDM) supplemented with 20 mM imidazole. The column was washed with 20 column volumes of Buffer A and a linear gradient was applied with Buffer A and 500 mM imidazole. Eluate fractions were further purified by affinity chromatography by using 5 ml Strepactin column (Iba GmbH) equilibrated with 20 mM Tris pH 8.0, 500 mM NaCl, 5% Glycerol, 0.03 % DDM, 1 mM EDTA and eluted with buffer A supplemented with 2.5 mM of desthibiotin. For $SabA^{\Delta 19\text{-}448}$, the outer membrane purification consisted only of the Strep affinity chromatography.

Electrophoretic and Immunoblot Analysis of BabA. Seminative SDS-PAGE was performed as previously described (Jansen *et al.*, 2000). Proteins were separated by SDS-PAGE and seminative PAGE, transferred to PVDF membranes and blocked with a 5 % w/v solution of bovine serum albumin (BSA) in T-TBS (TBS supplemented with 0.5 % v/v Tween 20). The blotted membrane was then incubated with an appropriate dilution of primary antibody (anti-StrepII (Quiagen), anti-His (Sigma) or rabbit anti-BabA polyclonal) for 2 hours, washed three times with T-TBS, incubated with alkaline phosphatase-conjugated secondary antibody (Sigma) for 2 hours and washed three times with T-TBS. The membrane was developed with Western Blue® Stabilized (Promega) substrate for alkaline phosphatase.

Immunofluorescence assays. *E. coli BL21 (pFC100)* was grown in LB medium at 26 °C and induced with 0.2 µg ml^{-1} anhydrotetracycline at an OD_{600} of 0.8. After 30 minutes expression, cells of 2 ml of culture were harvested by centrifugation for 3 min at 4500 *g*, washed with cold PBS (phosphate buffered saline) and resuspended in 1 ml PBS. Following fixation in PBS and 4 % v/v para-Formaldehyde for 20 minutes at 4 °C, the cells were first washed with PBS and successively incubated for 30 minutes at room temperature in PBS/5 % w/v BSA (bovine serum albumin). The cells were washed with PBS and resuspended in PBS/0.5 % w/v BSA supplemented with a 1000-fold dilution of either anti-HA (Sigma, H3663) or anti-StrepII during 1 hour, washed 3 times with PBS/0.5

% w/v BSA and incubated with Alexa fluor goat anti-mouse 488 (AF488; Invitrogen). Part of the induced cells were incubated with PBS/ 0.5% v/v Triton X-100 before incubation with anti-StrepII antibody.

To test extracellular exposure and conformational state of BabA$^{1-721}$, nanobody 14 (Nb14; Moonens *et al.*, 2016) or a Leb-human serum albumin glycoconjugate (Isosep) was labelled with a three-fold molar excess AF488 (ThermoFisher) at pH 9.0 during 30 min at room temperature using N-hydroxysuccinimide (NHS) chemistry. The labelling was quenched with 50 mM Tris and the excess AF488 was removed by Zeba spin column (ThermoFisher). *E. coli BL21* transformed either with the empty pASK12 vector or pHRAFT containing *babA*$^{1-721}$, were grown in LB and induced with 0.2 µg/ml anhydrotetracycline at OD 0.8. After 1 h, 1 mL was pelleted and washed in PBS. When dry, 10 µL of cells resuspended in PBS was added to the poly-L-lysine coated glass slides for 10 min, rinsed with PBS and consequently blocked with 10 µL of 5 % bovine serum albumin (BSA). 10 µL of 30 µM AF488-labelled Nb14 was added and left incubating in the dark for 10 min, before rinsing with PBS. For the assay following Leb binding, cells in PBS were incubated with 0.2 µg/mL AF488-labelled Leb-HSA for 10 min, were pelleted at 5000*g* and washed twice with PBS before administering to a poly-L-lysine coated glass slide. The samples were studied at 100x magnification on a Leica Dmi8 microscope.

SabA$^{22-455}$ purification and structure determination. Periplasmic extract containing SabA$^{22-455}$ was prepared by resuspending the cells in 100 ml sucrose-buffer (40% sucrose, 20 mM Tris pH 8.5, 2 mM EDTA), allowing them to stir gently for 20 min at 4°C, then adding 120 ml 20 mM Tris pH 8.5 buffer and allowing another 30 min of stirring at 4°C. Centrifugation at 13,000 *g* for 45 min yielded the periplasmic extract. SabA$^{22-455}$ was purified by StrepII-affinity chromatography using Strep-Tactin Superflow beads (Iba). The beads were stirred at 4°C in the periplasmic extract, then collected in a gravity-flow column and washed with Strep-buffer (20 mM Tris pH 8.5, 150 mM NaCl, 1 mM EDTA). SabA$^{22-455}$ was eluted with Strep-buffer supplemented with 2.5 mM D-desthiobiotin,

concentrated and injected on a 60 ml Superose 6 column for size exclusion chromatography in buffer 20 mM acetate pH 5.0, 150 mM NaCl, 5% glycerol. The eluted protein was concentrated to 5-10 mg ml^{-1} and crystallization trials were set up using the sitting drop vapour diffusion method against salt- and polyethylene glycol (PEG)-based sparse matrix screens. The best crystals were obtained with a buffer containing 15% PEG 3350, 0.2 M KF, 0.1 M citrate pH 6.5. Crystals were cryoprotected by addition of 15% glycerol, mounted on a cryoloop and flash-cooled in liquid nitrogen. X-ray diffraction data were collected on the Proxima-1 beamline at SOLEIL, St-Aubin, France, tuned to 12.65 keV and equipped with a Pilatus 6M detector. The data was collected in rotation wedges of 0.2 degrees and processed and integrated using the XDS package to 2.52 Å (Kabsch, 2010) and found to be in the $P22_12_1$ space group, containing two $SabA^{22\text{-}455}$ copies. Phases were obtained through molecular replacement with PDB entry 4O5J (Pang *et al.*, 2013), using Phaser_MR (McCoy *et al.*, 2007). The model was refined by iterative rounds of maximum likelihood refinement with Refmac 5 (Collaborative Computational Project, Number 4, 1994) using 16 TLS groups identified with TLSMD (Painter and Merritt, 2006) and manual inspection and rebuilding in Coot (Emsley *et al.*, 2010). Data collection and refinement statistics are summarized in Table S3.

SAXS analysis of $SabA^{22\text{-}455}$. SAXS datasets were collected of $SabA^{22\text{-}455}$ using the batch method, on five different protein concentrations ranging between 1 and 10 mg ml^{-1}. Protein in buffer pH 5.0 (20 mM acetate pH 5.0, 150 mM NaCl, 5% glycerol) was measured on beamline BM29 at ESRF (Grenoble, France), equipped with a Pilatus 1M detector. Ten frames were collected with 2 s exposure time at 15°C at a wavelength of 1.0 Å. Data processing and analysis was performed using PRIMUS (Konarev *et al.*, 2003) and the radius of gyration and distance distribution function were calculated using GNOM (Svergun, 1992). The dimensionless Kratky-plot was generated as per Durand et al. 2010 (Durand *et al.*, 2010). *Ab initio* models were generated with GASBORi (Svergun *et al.*, 2001) and the theoretical scattering curve from the high-resolution X-ray structure was calculated by CRYSOL (Svergun *et al.*, 1995), run in default mode. Rigid-body fitting of the crystal structure of

SabA$^{22\text{-}455}$ in a typical SAXS *ab initio* model was performed manually in PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC).

Bioinformatics tools. BOCTOPUS, a tool for the prediction of transmembrane β-barrel protein domains (Hayat and Elofsson, 2012), was used with the sequence of full-length mature SabA from *H. pylori* strain J99. TCOFFEE (Notredame *et al.*, 2000) was used for multiple sequence alignment and visualisation of sequence identity of several members of the Hop and Hor protein families. The structure of SabA$^{\Delta 19\text{-}448}$ (predicted transmembrane β-barrel domain) was predicted by the PHYRE2 server (Kelley and Sternberg, 2009). Phylogenetic trees were made with FigTree 1.4. Figures of protein structure models were prepared with MacPyMOL (PyMOL v.1.3; Schrödinger).

Data availability

Coordinates and structure factors for the SabA$^{22\text{-}455}$ have been deposited in the PDB under accession code 6GW5.

Funding information

This work is supported by the Fonds Wetenschappelijk Onderzoek (FWO) through grants G.0902.09 and G033717N to Han Remaut, by Flanders Institute for Biotechnology (VIB) through project grant PRJ9 to Han Remaut, by the Hercules Foundation through equipment grant UABR/09/005. KM is an FWO post-doctoral fellow. The funders had no role in the study design, data collection and interpretation or writing and submission of the publication of the work.

Acknowledgements

We thank Soleil synchrotron, Biostruct-X and beamline staff at Proxima-1 and Proxima-2a for access and support under proposal 20121253, and Pierre Lebrun for help with the SAXS data analysis. We thank Thomas Borén and Zhanna Bugaytsova for discussions on the study and manuscript. The authors do not have competing financial interests.

Author contributions

FC solved the SabA SAXS and X-ray structures, made SabA constructs. GC made BabA constructs and performed BabA sorting experiments. FC and ADB performed the localization imaging and functional assays. AL, SB and KM assistant in construct and study design. HR designed and supervised the study, and assisted in data interpretation. HR, GC and FC wrote the manuscript, with input from all authors.

Figure legends

Figure 1. **Phylogeny and domain architecture of *H. pylori* Hop and Hor proteins. (A)** Schematic representation of Type 5a secretion proteins and *Helicobacter* Hops and Hors, which show two regions of high sequence conservation at the N- and C-termini (blue boxes) that in Hops are spaced by an intersecting region with high sequence and length variation and that corresponds to the surface exposed adhesin domain (red box). In Hors the intersecting variable region is lacking. ID: pairwise sequence identity **(B)** Phylogenetic tree of Hop and Hor protein sequences corresponding to the conserved N- and C-terminal region (known adhesins are marked with an asterisk). **(C)** Subcellular localization of BabA$^{1\text{-}721}$ recombinantly expressed in *E. coli* BL21 as assessed by anti-StrepII immunoblot of the fractions: IM (inner membrane, N-lauroylsarcosine soluble fraction), OM (outer membrane DDM extract) and P (pellet), heat-treated ("+" 95 °C) or not ("-" 95°C) prior to separation

by to SDS-PAGE. Bands corresponding to the folded (F) and/or unfolded (U) form of the constructs are indicated.

Figure 2. *H. pylori* BabA does not have an independent C-terminal β-barrel domain. (A) Subcellular localization of BabA N-terminal truncates (BabA$^{441\text{-}721}$, BabA$^{479\text{-}721}$, BabA495$^{\text{-}721}$ and BabA$^{522\text{-}721}$) recombinantly expressed in *E. coli* BL21 as assessed by anti-StrepII immunoblot of the fractions: IM (inner membrane, N-lauroylsarcosine soluble fraction), OM (outer membrane DDM extract) and P (pellet), heat-treated ("+" 95 °C) or not ("-" 95°C) prior to separation by to SDS-PAGE. Bands corresponding to the folded (F) and/or unfolded (U) form of the constructs are indicated. **(B)** Anti-His and anti-BabA immunoblots of OM DDM extract of mutant BabA^{441TH} treated with thrombin protease. The N-terminal (NT) and 6His-tagged C-terminal cleavage products (CT) remained associated before (Load, labelled "L") and after (Eluate, labelled "E") passage over a nickel affinity column. Schematic drawings show canonical domain organization the various BabA constructs, where the conserved regions at the N- and C-termini are coloured blue, and the predicted adhesin domain is shown in red.

Figure 3. *In silico* analysis of the Hop transmembrane elements. (A) Transmembrane β-strand prediction of SabA by the support vector machines (SVM) and Hidden Markov model (HMM) employed by BOCTOPUS (Hayat and Elofsson, 2012). Whilst the SVM algorithm predicts the presence of a candidate transmembrane strand in the first 16 residues, it is missed by the Hidden Markov model possibly due to the absence of singleton β-strands in the training set. Additionally, the HMM predicts an eight TM strand in the C-terminal region that has very poor TM strand propensity according the SVM, likely due to the bias of even-numbered, and 8-stranded OMPs in the HMM training set. **(B)** Multiple sequence alignment of the transmembrane regions of representative members of the Hop and Hor subfamilies and OmpA. For Hop and Hor proteins, the shown sequences exclude the predicted α-helical insertion domains. Arrows indicate the experimentally determined

(OmpA) and the consensus predicted β-strands (Hop and Hor) as identified by Phyre2. Shaded residues indicated bilayer facing side chains in OmpA (PDB entry: 1QJP) and OmpA-based homology model of the BabA or SabA β-barrel. A full sequence alignment of Hops is shown in Supplementary Fig. 2.

Figure 4. Topology of Hop transmembrane domain. (A, B) Subcellular localization of BabA$^{17\text{-}721}$ (A) or BabA$^{\Delta 28\text{-}535}$ (B) recombinantly expressed in *E. coli* BL21 and assessed by anti-StrepII immunoblot of the inner membrane (IM), outer membrane (OM) and pellet (P) fractions. Samples heat-treated ("+" 95 °C) or not ("-" 95 °C) prior to separation by to SDS-PAGE. Bands corresponding to the folded (F) and/or unfolded (U) form of the constructs are indicated. **(C)** Coomassie-stained SDS-PAGE of BabA$^{\Delta 28\text{-}535}$ or SabA$^{\Delta 19\text{-}448}$ purified from the outer membrane fraction shows the presence of a fast migrating, folded monomeric β-barrel in the non-boiled samples. **(D)** Bright field (1, 3, 5, 7) and immunofluorescence microscopy images (2, 4, 6, 8) of *E. coli* BL21(DE3) transformed with pFC100, non-induced (1, 2) or induced (3 to 8) for SabA$^{\Delta 19\text{-}448HA}$ expression. IF uses extracellularly added anti-StrepII (6, 8) or anti-HA (2, 4) tag antibodies and an Alexa fluor 488-conjugated secondary antibody. In (8) cells were permeabilized by addition of 0.5 % Triton X-100. **(E)** Ribbon and stick representation of the OmpA-based homology-model of the minimal BabA β-barrel domain, comprising residues 1-17 and 536-721 of the mature BabA (ie. removing the adhesin domain). The OmpA-based homology model is produced by the protein homology/analogy recognition engine Phyre (Kelley and Sternberg, 2009). The distribution of hydrophobic (orange) and aromatic side chains (magenta) in the expected transmembrane region is in general agreement with that seen in outer membrane β-barrels.

Figure 5. Domain organization in *H. pylori* Hops. (A) Experimental (orange) and calculated (blue) small angle X-ray scattering curves for SabA$^{25\text{-}455}$. The 2 to 6 q*Rg region of the dimensionless Kratky plot of the data (inset) is in good agreement with a structure consisting of two domains, whilst

the Kratky analysis at higher q*Rg reveals the presence of some flexibility in parts of the particle. This conformational flexibility may correspond to the extended loops in the SabA head domain and possibly to the C-terminal 17 residues (439-455) not resolved in the crystal structure. **(B)** Rigid-body fitting of the SabA$^{22\text{-}455}$ crystal structure (red) into a typical *ab initio* volume generated by GASBORi from the experimental SAXS-data. Superimposition of the published structure of the refolded SabA ectodomain (PDB 4O5J, green (Pang *et al.*, 2013)) shows that an additional subdomain is seen at the C-terminus of the soluble SabA$^{22\text{-}455}$ fragment purified from the *E. coli* periplasm, as well as a ~5 degree rotation of the head and stem domain relative to each other. **(C)** Superimposition of the BabA (PDB ID 4ZH0) and SabA$^{25\text{-}455}$ coiled-coil stem domains. **(D)** Schematic representation of proposed classes of Type 5 secretion proteins based on domain organization (Grijpstra *et al.*, 2013), with the numbers in de strands representing their order from N- to C-terminus. Classical monomeric (5a) and trimeric ATs (5c) share a modular structure, with an extracellular N-terminal passenger (red) and a C-terminal β-barrel domain embedded in the outer membrane (blue). In two-partner proteins secretion system (TPPS) (5b), the passenger and the translocator domains reside on two different polypeptides respectively termed TpsA and TpsB. TpsA is recognized by two periplasmic polypeptide transport-associated (POTRA; labeled P) domains of its cognate translocator TpsB. Type 5d represents ATs in which the passenger and the translocator domain are interconnected by a POTRA domain, whilst Type 5e indicates ATs with inverted topology, in which the β-domain is located at the N-terminus, as in the Intimin/Invasin family. *H. pylori* Hop proteins represent the first reported member of a novel class of Type 5 secretion-like proteins, refered to as Type 5f, with discontinuous β-barrel domain and insertion of the passenger domain in the extracellular loop connecting β-strands 1 and 2. (OM: outer membrane).

References

Alm, R.A., Bina, J., Andrews, B.M., Doig, P., Hancock, R.E., and Trust, T.J. (2000) Comparative genomics of Helicobacter pylori: analysis of the outer membrane protein families. *Infect Immun* **68**: 4155–4168.

Anwari, K., Webb, C.T., Poggio, S., Perry, A.J., Belousoff, M., Celik, N., Ramm, G., Lovering, A., Sockett, R.E., Smit, J., Jacobs-Wagner C., and Lithgow, T. The evolution of new lipoprotein subunits of the bacterial outer membrane BAM complex. (2012) *Mol Microbiol.* 84:832-844.

Bäckström, A., Lundberg, C., Kersulyte, D., Berg, D.E., Borén, T., and Arnqvist, A. (2004) Metastability of Helicobacter pylori bab adhesin genes and dynamics in Lewis b antigen binding. *Proc Natl Acad Sci USA* **101**: 16923–16928.

Browning, D.F., Bavro, V.N., Mason, J.L., Sevastsyanovich, Y.R., Rossiter, A.E., Jeeves, M., Wells, T.J., Knowles, T.J., Cunningham, A.F., Donald, J.W., Palmer, T., Overduin, M., and Henderson, I.R. (2015) Cros-species chimeras reveal BamA POTRA and β-barrel domains must be fine-tuned for efficient OMP insertion. *Mol Microbiol* **97**:646-659.

Bugaytsova, J.A., Björnham, O., Chernov, Y.A., Gideonsson, P., Henriksson, S., Mendez, M., *et al.* (2017) Helicobacter pylori Adapts to Chronic Infection and Gastric Disease via pH-Responsive BabA-Mediated Adherence. *Cell Host Microbe* **21**: 376–389.

Colbeck, J.C., Hansen, L.M., Fong, J.M., and Solnick, J.V. (2006) Genotypic profile of the outer membrane proteins BabA and BabB in clinical isolates of Helicobacter pylori. *Infect Immun* **74**: 4375–4378.

Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**: 760–763.

Dautin, N., and Bernstein, H.D. (2007) Protein secretion in gram-negative bacteria via the autotransporter pathway. *Annu Rev Microbiol* **61**: 89–112.

Durand, D., Vivès, C., Cannella, D., Pérez, J., Pebay-Peyroula, E., Vachette, P., and Fieschi, F. (2010) NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. *Journal of Structural Biology* **169**: 45–53.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**: 486–501.

Exner, M.M., Doig, P., Trust, T.J., and Hancock, R.E. (1995) Isolation and characterization of a family of porin proteins from Helicobacter pylori. *Infect Immun* **63**: 1567–1572.

Fairman, J.W., Noinaj, N., and Buchanan, S.K. (2011) The structural biology of β-barrel membrane proteins: a summary of recent reports. *Curr Opin Struct Biol* **21**: 523–531.

Grijpstra, J., Arenas, J., Rutten, L., and Tommassen, J. (2013) Autotransporter secretion: varying on a theme. *Res Microbiol* **164**: 562–582.

Hage, N., Howard, T., Phillips, C., Brassington, C., Overman, R., Debreczeni, J., *et al.* (2015) Structural basis of Lewisb antigen binding by the Helicobacter pylori adhesin BabA. *Science Advances* **1**: e1500315–e1500315.

Hayat, S., and Elofsson, A. (2012) BOCTOPUS: improved topology prediction of transmembrane β

barrel proteins. *Bioinformatics* **28**: 516–522.

Ieva, R., Tian, P., Peterson, J.H., and Bernstein, H.D. (2011) Sequential and spatially restricted interactions of assembly factors with an autotransporter beta domain. *Proceedings of the National Academy of Sciences* **108**: E383–91.

Ilver, D., Arnqvist, A., Ogren, J., Frick, I.M., Kersulyte, D., Incecik, E.T., *et al.* (1998) Helicobacter pylori adhesin binding fucosylated histo-blood group antigens revealed by retagging. *Science* **279**: 373–377.

Jansen, C., Heutink, M., Tommassen, J., and de Cock, H. (2000) The assembly pathway of outer membrane protein PhoE of Escherichia coli. *Eur J Biochem* **267**: 3792–3800.

Javaheri, A., Kruse, T., Moonens, K., Mejías-Luque, R., Debraekeleer, A., Asche, C.I., *et al.* (2016) Helicobacter pylori adhesin HopQ engages in a virulence-enhancing interaction with human CEACAMs. *Nat Microbiol* **2**: 16189.

Kabsch, W. (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66**: 125–132.

Kelley, L.A., and Sternberg, M.J.E. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**: 363–371.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003) *PRIMUS*: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography* **36**: 1277–1282.

Königer, V., Holsten, L., Harrison, U., Busch, B., Loell, E., Zhao, Q., *et al.* (2016) Helicobacter pylori exploits human CEACAMs via HopQ for adherence and translocation of CagA. *Nat Microbiol* **2**: 16188.

Leyton, D.L., Rossiter, A.E., and Henderson, I.R. (2012) From self sufficiency to dependence: mechanisms and factors important for autotransporter biogenesis. *Nat Rev Micro* **10**: 213–225.

Leyton, D.L., Sevastsyanovich, Y.R., Browning, D.F., Rossiter, A.E., Wells, T.J., Fitzpatrick, R.E., *et al.* (2011) Size and conformation limits to secretion of disulfide-bonded loops in autotransporter proteins. *J Biol Chem* **286**: 42283–42291.

Liechti, G., and Goldberg, J.B. (2012) Outer membrane biogenesis in Escherichia coli, Neisseria meningitidis, and Helicobacter pylori: paradigm deviations in H. pylori. *Front Cell Infect Microbiol* **2**: 29.

Mahdavi, J., Sondén, B., Hurtig, M., Olfat, F.O., Forsberg, L., Roche, N., *et al.* (2002) Helicobacter pylori SabA Adhesin in Persistent Infection and Chronic Inflammation. *Science* **297**: 573–578.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007) Phaser crystallographic software. *Journal of Applied Crystallography* **40**: 658–674.

Meng, G., Surana, N.K., St Geme, J.W., and Waksman, G. (2006) Structure of the outer membrane translocator domain of the Haemophilus influenzae Hia trimeric autotransporter. *EMBO J* **25**: 2297–2304.

Moonens, K., Gideonsson, P., Subedi, S., Bugaytsova, J., Romaõ, E., Mendez, M., *et al.* (2016) Structural Insights into Polymorphic ABO Glycan Binding by Helicobacter pylori. *Cell Host Microbe* **19**: 55–66.

Notredame, C., Higgins, D.G., and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* **302**: 205–217.

Odenbreit, S., Till, M., Hofreuter, D., Faller, G., and Haas, R. (1999) Genetic and functional characterization of the alpAB gene locus essential for the adhesion of Helicobacter pylori to human gastric tissue. *Mol Microbiol* **31**: 1537–1548.

Oomen, C.J., van Ulsen, P., van Gelder, P., Feijen, M., Tommassen, J., and Gros, P. (2004) Structure of the translocator domain of a bacterial autotransporter. *EMBO J* **23**: 1257–1266.

Painter, J., and Merritt, E.A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* **62**: 439–450.

Pang, S.S., Nguyen, S.T.S., Perry, A.J., Day, C.J., Panjikar, S., Tiralongo, J., *et al.* (2013) The three-dimensional structure of the extracellular adhesion domain of the sialic acid-binding adhesin SabA from Helicobacter pylori. *J Biol Chem*.

Peck, B., Ortkamp, M., Diehl, K.D., Hundt, E., and Knapp, B. (1999) Conservation, localization and expression of HopZ, a protein involved in adhesion of Helicobacter pylori. *Nucleic Acids Res* **27**: 3325–3333.

Peek, R.M., and Blaser, M.J. (2002) Helicobacter pylori and gastrointestinal tract adenocarcinomas. *Nat Rev Cancer* **2**: 28–37.

Peterson, J.H., Tian, P., Ieva, R., Dautin, N., and Bernstein, H.D. (2010) Secretion of a bacterial virulence factor is driven by the folding of a C-terminal segment. *Proceedings of the National Academy of Sciences* **107**: 17739–17744.

Reithmeier, R.A., and Bragg, P.D. (1974) Purification and characterization of heat-modifiable protein from the outer membrane of Escherichia coli. *FEBS Lett* **41**: 195–198.

Rossez, Y., Gosset, P., Boneca, I.G., Magalhães, A., Ecobichon, C., Reis, C.A., *et al.* (2014) The lacdiNAc-specific adhesin LabA mediates adhesion of Helicobacter pylori to human gastric mucosa. *J Infect Dis* **210**: 1286–1295.

Senkovich, O.A., Yin, J., Ekshyyan, V., Conant, C., Traylor, J., Adegboyega, P., *et al.* (2011) Helicobacter pylori AlpA and AlpB bind host laminin and influence gastric inflammation in gerbils. *Infect Immun* **79**: 3106–3116.

Svergun, D., Barberato, C., and Koch, M. (1995) CRYSOL-a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography* **28**: 768–773.

Svergun, D.I. (1992) Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *Journal of Applied Crystallography* **25**: 495–503.

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

Uemura, N., Okamoto, S., Yamamoto, S., Matsumura, N., Yamaguchi, S., Yamakido, M., *et al.* (2001) Helicobacter pylori infection and the development of gastric cancer. *N Engl J Med* **345**: 784–789.

Voulhoux, R., Bos, M.P., Geurtsen, J., Mols, M., and Tommassen, J. (2003) Role of a highly conserved bacterial protein in outer membrane protein assembly. *Science* **299**: 262–265.

Webb, C.T., Chandrapala, D., Oslan, S.N., Bamert, R.S., Grinter, R.D., Dunstan, R.A., Gorrell, R.J., Song, J., Strugnell, R.A., Lithgow, T., and Kwok, T. (2017) Reductive evolution in outer membrane protein biogenesis has not compromised cell surface complexity in *Helicobacter pylori*. *Microbiologyopen*. doi: 10.1002/mbo3.513.

Figure 1

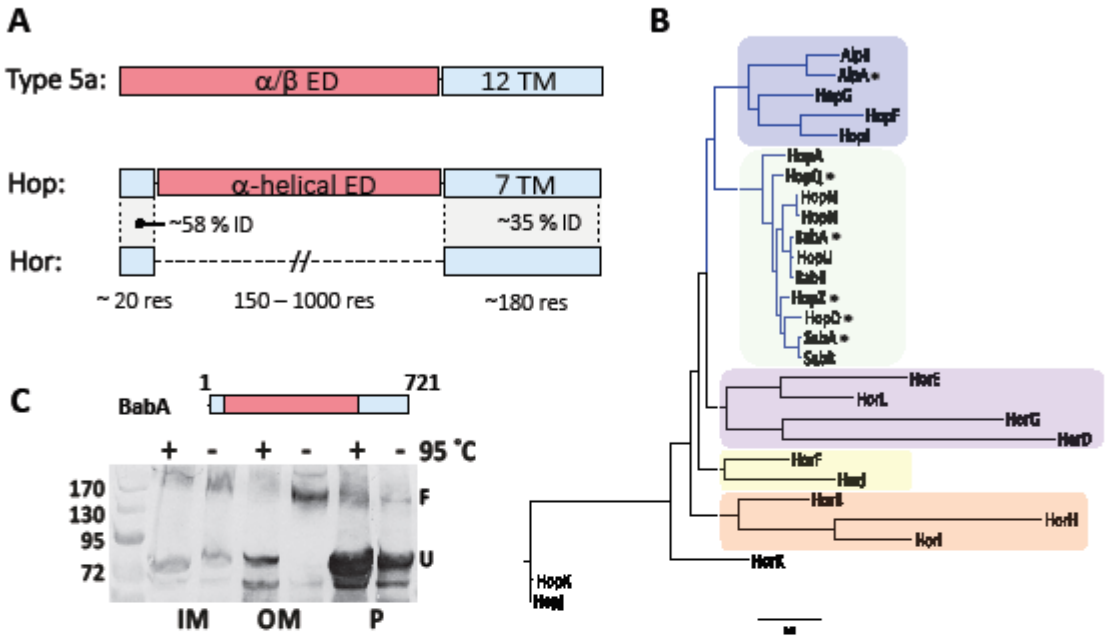

Figure 2

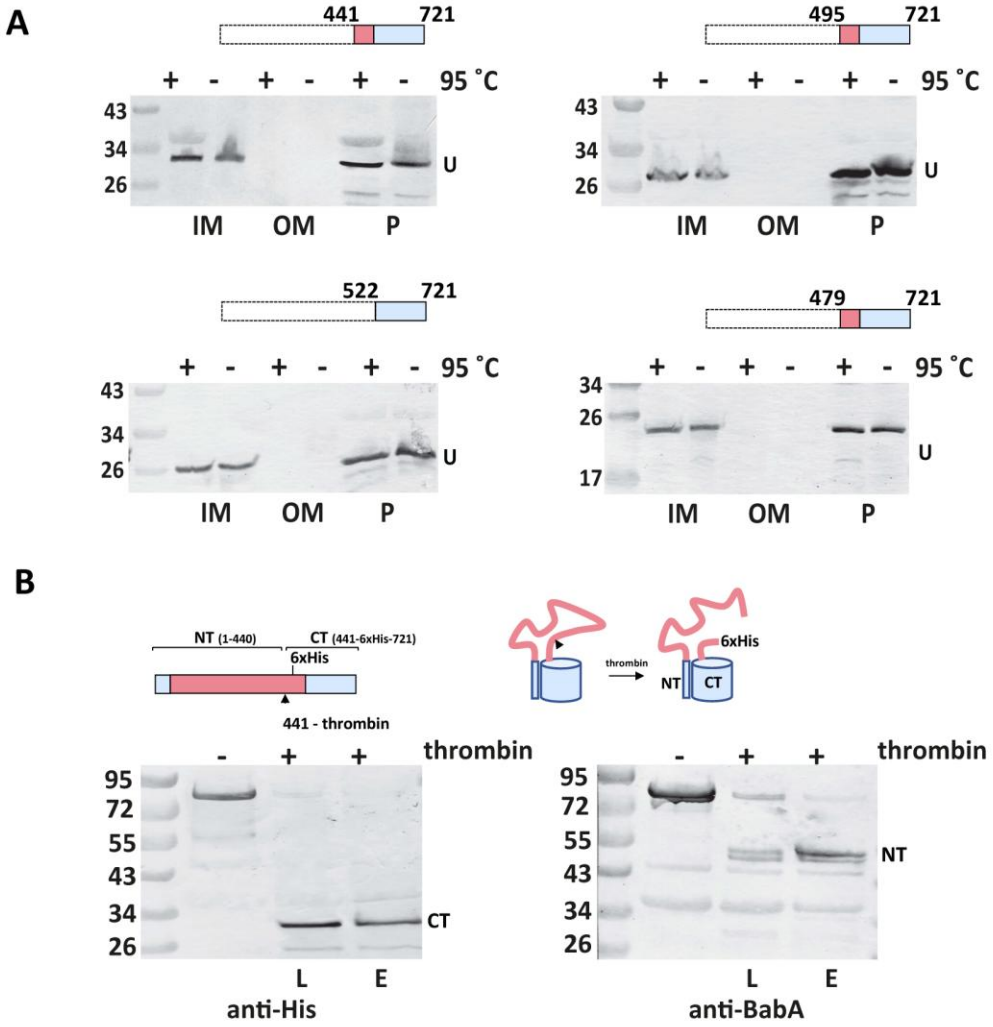

Figure 3

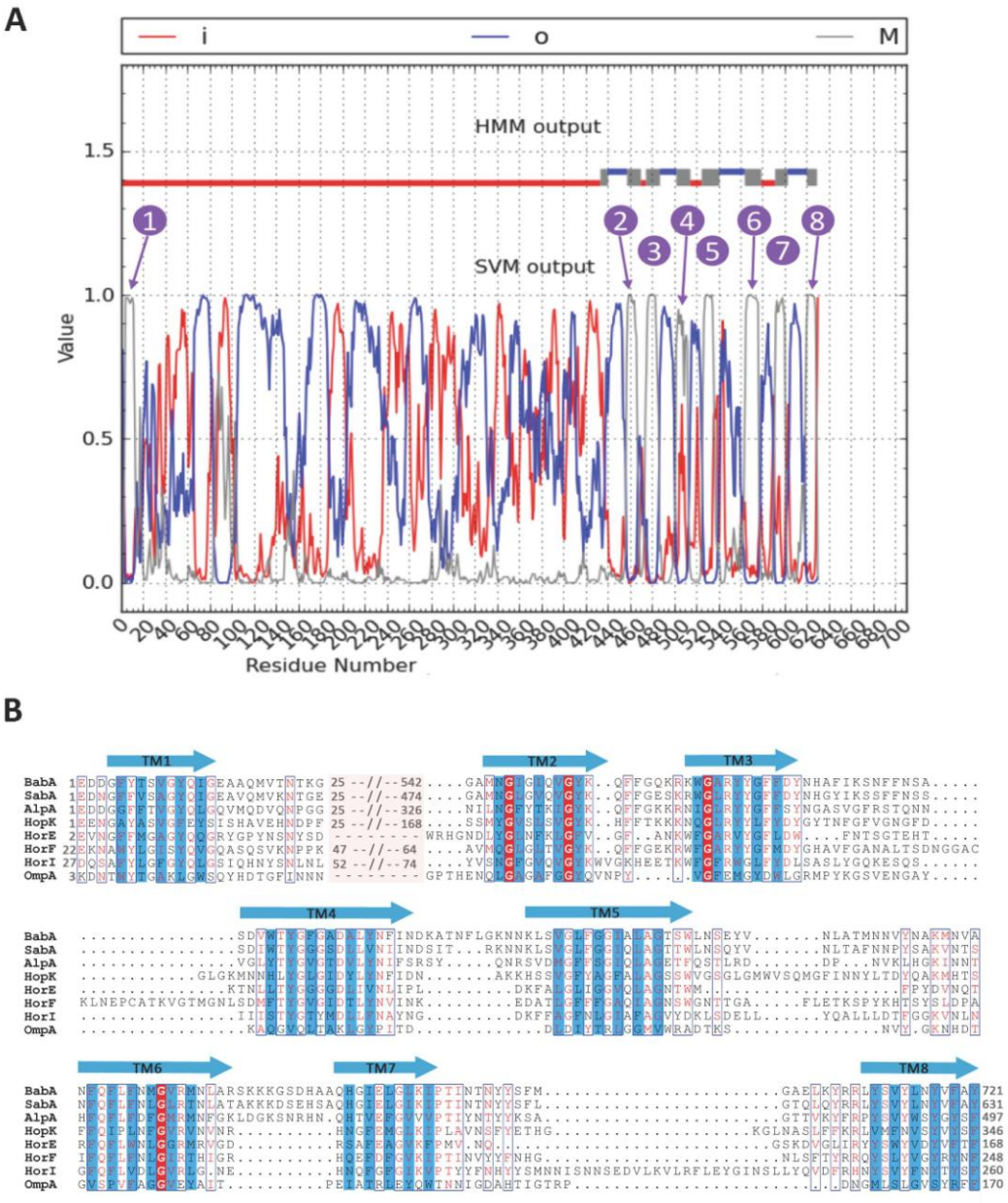

Figure 4

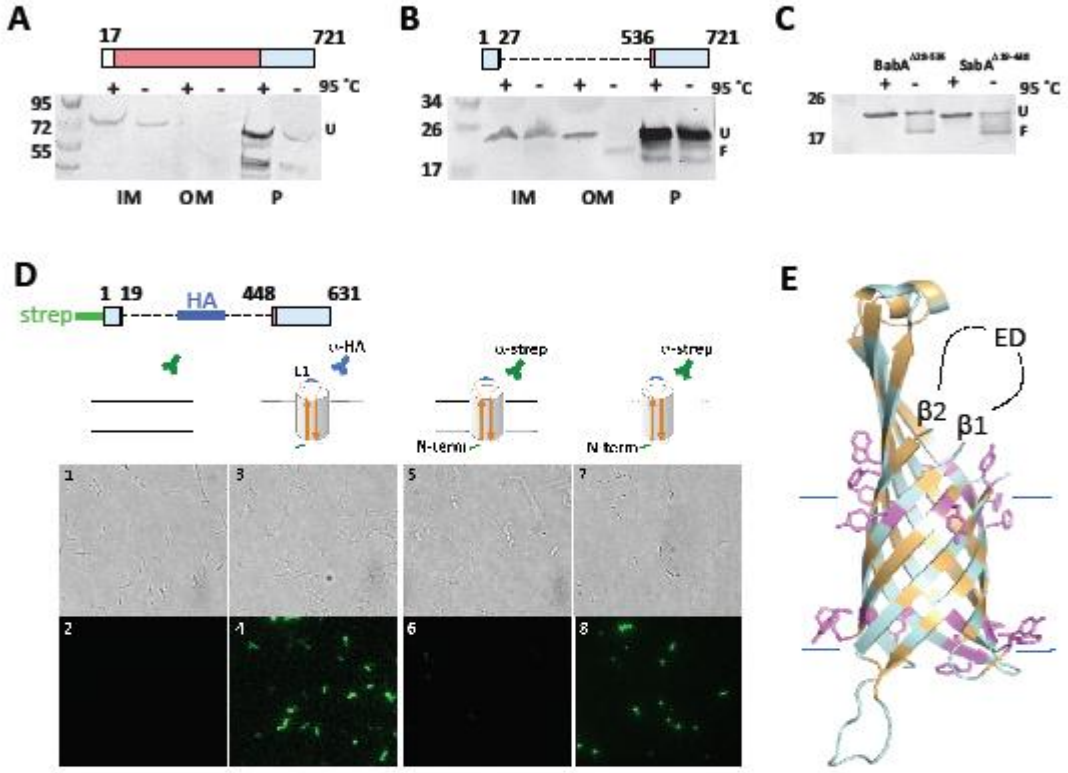

Figure 5