Functional interaction and structural characteristics of unique components of *Helicobacter pylori* T4SS

Enrique Merino¹, Marcos Flores-Encarnación² and Germán Rubén Aguilar-Gutiérrez³

1 Enrique Merino, Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México
2 Marcos Flores-Encarnación, Laboratorio de Microbiología Molecular y Celular, Facultad de Medicina, Benemérita Universidad Autónoma de Puebla, Puebla, México
3 Germán Rubén Aguilar-Gutiérrez, Centro de Investigación Sobre Enfermedades Infecciosas, Instituto Nacional de Salud Pública, Cuernavaca, Morelos, México

**Keywords**

Cag3; CagF; CagG; CagH; CagI; CagM; CagU; CagZ; *Helicobacter pylori*; pathogenicity island; type IV secretion system

**Correspondence**

G. R. Aguilar-Gutiérrez, Centro de Investigación Sobre Enfermedades Infecciosas, Instituto Nacional de Salud Pública, Cuernavaca, Morelos 62100, México
Fax: +52-777-3175485
Tel: +52-777-3293000
E-mail: graguila@insp.mx
Website: http://www.insp.mx/

(Received 7 November 2016, revised 14 March 2017, accepted 26 April 2017)
doi:10.1111/febs.14092

Introduction

The *Helicobacter pylori* infection of the human gastric mucosa causes chronic active gastritis and peptic ulcers and is associated with the development of gastric cancer. Epidemiological studies show that these gastric diseases are related to virulent *H. pylori* strains that harbor the cytotoxin-associated gene pathogenicity island (*cag* PAI). The *cag* PAI is a DNA insertion in the *H. pylori* chromosome that encodes ~27 proteins, including the oncoprotein CagA. Approximately 20 of these proteins have been designated as *cag* type IV secretion system (T4SS) components. However, only 11 of these proteins share function, structure, and/or sequence similarities with the prototypical VirB/VirD4 T4SS of *Agrobacterium tumefaciens*. The VirB/VirD4 orthologs of the *cag* T4SS of *H. pylori* are required for CagA translocation and stimulate the gastric epithelial cells to produce and secrete interleukin-8 (IL-8). The *cag* PAI encodes eight additional proteins, such as Cag3 (Cag6/HP0522), CagM (Cag16/HP0537), CagU (Cag11/HP0531), CagI (Cag19/HP0540), and CagH (Cag20/HP0541), which are also required for the translocation of CagA and IL-8 secretion, meanwhile CagF (Cag22/HP0543), CagG (Cag21/HP0542), and CagZ (Cag6/HP0526) are just required for the translocation of CagA. However, relatively little is known about their functions and structural organization because they exhibit a nondetectable sequence similarity with T4SS components in the current databases. In this review, we conducted an exhaustive analysis of the literature to present the biochemistry, putative role, localization, and interactions of each of these eight additional *cag* T4SS components.

Abbreviations

cag PAI, cytotoxin-associated gene pathogenicity island; IL-8, interleukin 8; pI, isoelectric point; amino-terminal SS, amino-terminal signal sequence; T4SS, type IV secretion system.
The T4SS protein cag A, CagA (HP0547/Cag26), which is regarded as a bacterial oncoprotein [3]. The cag T4SS delivers the CagA protein that promotes chemokine release and causes alterations in the cell signaling once it is within the gastric epithelial cells [4–6]. The cag T4S system also induces the production and secretion of the proinflammatory cytokine interleukin-8 (IL-8) in gastric epithelial cells in a process associated with the delivery of peptidoglycan [7]. For this reason, the presence of the cag PAI is a hallmark of virulent H. pylori strains because it causes chronic inflammation of the gastric mucosa and alterations in the cell signaling associated with gastric carcinogenesis [8,9].

**Type IV secretion system**

Type IV secretion systems are versatile macromolecular machines that are widely distributed among several bacteria species and are classified in three main functional categories as follows: (a) conjugation systems, such as those translocating DNA substrates to recipient cells by direct contact; (b) DNA release or uptake systems that transport DNA to and from the extracellular space [10,11]; and (c) translocation machines that transport substrates, ranging from small proteins to large nucleoprotein complexes, to the extracellular space or into target cells. Nucleoprotein complexes or effector proteins can be transported into many types of recipient cells, including eukaryotic and fungi cells [12].

The prototypical T4SS is the *Agrobacterium tumefaciens* VirB/VirD4 system, which is the most well studied [13]. This T4SS transfers virulent proteins and the T-DNA segment from the resident tumor-inducing plasmid into a wide variety of recipient plant cells. Once integrated into the host’s nuclear genome, T-DNA is expressed and causes cell transformation resulting in the formation of crown-gall disease [14]. This transport system is a syringe-like structure comprising 12 subunits, including 11 essential proteins (VirB1–VirB11) encoded by the operon virB and a coupling protein (VirD4) that mediates substrate recognition [12]. Most of these 12 proteins are conserved among other known bacterial T4SSs. However, some systems are adapted for a particular function and include either more or <12 structural components [13].

The *A. tumefaciens* VirB/VirD4 T4SS proteins form a macromolecular machine comprising a T-pilus complex, which is an extracellular appendage component extending from the cell surface, the inner membrane complex, and a transmembrane channel known as the core complex that is responsible for transporting substrates across bacterial membranes [13]. The T-pilus protein complex is thought to initiate cell–cell contact with the plant target cells before the T-DNA is transferred. This structure is composed of a main VirB2 subunit and an accessory VirB5 component, functioning as a specialized adhesin that bridges T4SS to the target cells [15–18]. The inner membrane complex contains the energetic components VirB4, VirB11, and VirD4, which have conserved Walker A and B box motifs that are required for nucleotide binding and hydrolysis, respectively. These energetic components are conserved among other Gram-negative bacteria and provide energy for T4SS assembly and/or other functions [19,20]. The VirB3 and VirB6–VirB10 are components of the core complex, which structure the substrate translocation channel that span the two membranes with VirB6, VirB8, and VirB10 attached to the inner membrane with domains spanning the periplasm, and VirB7 and VirB9 form the outer membrane portion [12,19]. The functional T4SS also includes VirB1, a protein whose transglycosylase activity is important for assembly efficiency [15,18].

**cag pathogenicity island**

The cag PAI of *H. pylori* is a ~40 kb region of chromosomal DNA that encodes ~27 Cag proteins [21–23]. Approximately 20 of these proteins have been designated as cag T4SS components. However, only 11 of these proteins share function, structure, and/or sequence similarities with the prototypical VirB/VirD4 T4SS of *A. tumefaciens*. The VirB/VirD4 orthologs of the cag T4SS of *H. pylori* are required for CagA translocation and to stimulate the gastric epithelial cells to produce and secrete IL-8 [20,24,25]. However, only 5 of these 11 orthologs of the cag T4SS show significant sequence similarity to proteins of the prototypical T4SS of *A. tumefaciens* [20,24–27]. For example, CagE (Cag23/HP0544), CagX (Cag8/HP0528), CagY (Cag7/HP0527), Cagx (HP0525), and Cag5 (Cagβ/HP0524) show important sequence similarities with VirB3/VirB4, VirB9, VirB10, VirB11, and VirD4 of *A. tumefaciens*, respectively [26]. The remaining six proteins, Cag4 (Cag/HP0523), CagC (Cag25/HP0546), CagL (Cag18/HP0539), CagW (Cag9/HP0529), CagT (Cag12/HP0532), and CagV (Cag10/HP0530), do not share sequence similarity with VirB/VirD4 T4SS components of *A. tumefaciens*. Nevertheless, protein localization, protein–protein interaction, and functional analysis suggested that they correspond to VirB1, VirB2, VirB5, VirB6, VirB7, and VirB8 components of cag T4SS, respectively [26–33]. For example, in many T4SSs, VirB1 and its orthologs share a conserved SLT
domain in amino acid sequence and may function as lytic transglycosylase. CagY is the only protein that contains a conserved SLT sequence motif in cag PAI, and it has a demonstrated lytic activity against bacterial cell walls [28]. Therefore, CagY corresponds to the VirB1 ortholog that contributes to the assembly of cag T4SS by digesting the peptidoglycan meshwork of the H. pylori cell wall [26,28,34].

Similarly, the T-pilus is an external structural part of the A. tumefaciens T4SS composed mostly of the processed VirB2 component. CagC has been identified, using a motif-based search, as the VirB2 ortholog. It is partially exposed on the surface, possesses predicted structural similarities when is compared with VirB2-like pilins of other T4SSs. Nevertheless, in contrast with A. tumefaciens T4SS, in cag T4SS, the VirB10 (CagY) ortholog is the main component of the needle-like structure (Fig. 1) [27,29].

Additionally, when the crystal structure of TraC, the VirB5 homolog from the pKM101 T4SS, was used as a template to perform a detailed protein modeling analysis, many sequences of VirB5 proteins returned a high score, indicating that the structure of TraC can be considered a paradigm for most VirB5 orthologs [30]. A structural modeling of CagL using the TraC structure revealed a reasonable fit of the CagL sequence within the TraC structure [30]. Functional studies of TraC have suggested that this protein plays a role in adhesion, mediating cell–cell interaction during conjugation, which is consistent with role of CagL in the interaction with host cell receptors and its location at the tip of cag T4SS pilus (Fig. 1) [27,30,34–38]. Accordingly, CagL corresponds to the VirB5 ortholog of the cag T4SS.

Similarly, among cag PAI-encoded proteins, CagW has been proposed as the ViB6 component of the cag T4SS [27]. The proteins of the VirB6 family share a limited similarity in amino acid sequence, but they have 5–7 predicted transmembrane helices. CagW has six potential transmembrane helices, but has no similarity on the primary sequence level. Additionally, the last predicted transmembrane helix in CagW is rich in

Fig. 1. Schematic representation cag T4SS of Helicobacter pylori. The cag T4SS components are shown in a simple manner; they are not drawn to scale or exact number of subunits. The 11 proteins of cag T4SS that present orthology with those components of the prototypical Agrobacterium tumefaciens VirB/VirD4 T4SS are described as B1-2, B3/4, B5-11, and D4, respectively. The eight additional components of cag T4SS required for CagA translocation and/or to induce IL-8 are represented by Cag3, CagM, CagF, CagH, CagI, CagZ, CagU, and CagG. These eight additional cag T4SS components are depicted in their most likely localizations according to sequence prediction or experimental data described in the text. Each defined component of the cag T4SS is represented in different colors according to their codes. In addition, the effector CagA is colored in light red.
valine/leucine/isoleucine content, which is considered important for the function of VirB6 [27]. Furthermore, CagW contains an essential tryptophan residue in a conserved motif preceding the last predicted transmembrane helix similar to the VirB6 family members, which is localized in a cytoplasmic loop before transmembrane helix 4 [27].

On the other hand, although no VirB7 homolog was found in the cag PAI, the inferred cagT product contains a typical bacterial lipoprotein signal sequence (SS) of the VirB7 orthologs [33]. In addition, CagT as the other VirB7 proteins, is partially exposed on the bacterial surface [27,33], and although CagT is a much larger protein when compared with the small lipoproteins VirB7 and VirB7-like proteins, CagT can be considered as a member of a second class of ViB7 lipoproteins involved in additional functions in the cag T4SS [26].

CagV exhibits absence of obvious sequence homologies with any component of the T4S systems. Nevertheless, this protein has features similar to the VirB8 homologs. CagV is a bitopic inner membrane protein with only one predicted transmembrane helix span [27,31]. CagV also shows the three most conserved motifs among all VirB8 homologs. These conserved motifs can be located at different sequence positions within the VirB8 homologs, but the only predicted transmembrane helix always precedes the first motif (PLK) directly [31]. All these structural features suggest that CagV belongs to the VirB8 family of proteins [31].

**Unique cag T4SS components**

The cag T4SS has a more elaborated structure than other T4SSs and needs more subunits than the A. tumefaciens VirB/VirD4 system for CagA translocation and/or to induce IL-8 secretion into the target cells [34]. These additional subunits are unique and presumably represent specialized adaptations optimized for type IV secretion in the human gastric mucosal environment given that they do not show sequence similarity or topology to components of other bacterial T4SSs, and their function and localization in this secretion machinery is uncertain [39]. The cag PAI encodes a group of distinctive components, including Cag1 (Cag7/HP0520), Cag2 (Cag8/HP0521), CagQ (Cag14/HP0535), and CagP (Cag15/HP0536), which are not necessary to the CagA translocation or IL-8 induction [34]. Nevertheless, the cag PAI also encodes a group of eight distinctive components including Cag3 (Cag8/HP0522), CagM (Cag16/HP0537), CagU (Cag11/HP0531), CagI (Cag19/HP0540), and CagH (Cag20/HP0541), whose absence results in a severe impairment of CagA translocation and IL-8 induction. Meanwhile, the absence of CagF (Cag22/HP0543), CagG (Cag21/HP0542), or CagZ (Cag6/HP0526) just prevent the CagA translocation [20,34,35,38–40]. These eight unique cag T4SS components and their putative roles in assembly, localization, and interaction are described below.

**Cag3 (Cag6/HP0522)**

Cag3 is essential for cag T4SS activity given that cag3 mutant strains are defective in pilus formation, and therefore, are unable to translocate CagA and induce IL-8 in the human gastric epithelial cell line AGS [34,41–43]. cag3 is transcribed, along with cag1, cag2, and cag4 (numbered in the direction of transcription). The polycistronic mRNA is translated as Cag3, the PG hydrolase Cag4, and the proteins Cag1 and Cag2, which are not essential for cag T4SS function [22,44,45]. Alternatively, a putative promoter overlays cag3, and the contiguous upstream region also promotes the expression of cag3 and cag4 as a polycistronic mRNA molecule [44]. This protein has a molecular weight of 55 kDa and a size of 481 amino acids, possesses a pI (isoelectric point) of 8.87, and contains two predicted coiled coils and an amino-terminal SS for transport to the periplasmic space [21,22,26,27,32,42]. Cell fractionation experiments showed that the bulk of the Cag3 protein is in the membrane-associated fraction [42]. Cag3 has been detected on the external surface of H. pylori or as a component of the outer membrane [46].

Additionally, affinity purification of proteins expressed by H. pylori during growth, in the absence of AGS cell contact showed that Cag3, within the periplasmic space, forms oligomers and interacts with the cag T4SS core components in the membrane-associated fraction [32,42]. Coimmunoprecipitation and cross-linking approaches showed that Cag3 specifically interacts with CagM, CagT, CagX, and CagY, which together form part of the cag T4S membrane-spanning core complex (Fig. 1) [42]. The interaction of Cag3 with CagT, another component required for pilus formation, represents the most specific Cag3 protein partner based on the abundance of CagT peptides identified by affinity purification followed by mass spectrometry approaches [42]. Another characteristic of the Cag3/CagT interaction is that it is required to stabilize both proteins and maintain their optimal levels in the outer membrane subcomplex of H. pylori T4SS [42]. Consistent with these studies, an in-depth analysis indicated that Cag3 specifically cosediments...
with CagM, CagT, CagX, and CagY in high-molecular-mass fractions in glycerol density gradients [47]. This analysis, using negative-stain single-particle electron microscopy, revealed that Cag3 is a component of the membrane-spanning core complex that consists of an outer ring of 41 nm in diameter and a 19 nm diameter central ring that are connected to each other by spoke-like links [47]. A mutant missing the Cag3 component has complexes with a well-defined central ring but a poorly defined outer ring. Furthermore, the localization of Cag3 by immunogold labeling analysis revealed gold particles in the periphery of the outer ring of the complexes from the wild-type (WT) strain, whereas a mutant lacking Cag3 did not show these particles, confirming that Cag3 is a peripheral component of the membrane-spanning core complex [47]. Collectively, these results indicate that Cag3 helps to form and stabilize VirB7 steady-state levels of the membrane-spanning core complex, which is required for pilus formation and the full activity of the cag T4S machinery.

**CagM (Cag16/HP0537)**

CagM is an additional unique component with functional relevance for the cag T4SS. Mutagenesis analyses demonstrated that ΔcagM strains are defective in pilus formation, and consequently, do not translocate CagA, release peptidoglycan, or induce IL-8 secretion in host gastric cells [7, 20, 32, 34, 43, 48]. cagM forms an operon with the downstream gene cagN, which is expressed as a polycistronic mRNA [44, 45]. cagN (cag17/ hp0538) deletion has an intermediate effect on CagA delivery and IL-8 induction phenotype [49, 50], and there is no evidence that CagM and CagN interact. Curiously, cagN can also be expressed from its own promoter as monocistronic mRNA [45]. CagM is a 44 kDa protein made of 376 amino acids with a calculated pI of 9.25–9.29; it contains at least three coiled coils and a predicted SS driving its export to the periplasm space [27, 32, 48].

Cell fractionation experiments, 2-DE, and immunoblot analyses showed that CagM is primarily found in the membrane-bound fraction [27, 47, 48]. CagM is more abundant in the outer membrane fraction than in the inner membrane fraction, where it is found as a surface-exposed outer membrane protein, together with CagT and Cag3 [27, 46]. In the membrane-bound fraction, it oligomerizes with itself and interacts with other cag PAI-encoded proteins, such as CagT, CagX, CagY, and Cag3 (Fig. 1) [27, 32, 42, 47]. CagM, like Cag3, is one of the five Cag components of the ring-shaped complex, along with CagT, CagX, and CagY, which show some similarities in sequence or function to components of the membrane-spanning core complex of VirB7, VirB9, and VirB10 of *A. tumefaciens* T4SS, respectively [19, 27, 33, 47]. A negative-stain single-particle electron microscopy analysis revealed that ring-shaped complexes could not be recovered from a ΔcagM mutant [47]. Such an absence in the membrane fraction could be explained by the reduced stability of the heterodimer Cag3/CagT; the ΔcagM mutants produced significantly lower amounts of CagT (60%) [27, 34, 42].

**CagF (Cag22/HP0543)**

CagF is a cag PAI-encoded protein located within the operon *cagC-cagL* or alternatively within the operon *cagF-cagL* [45]. CagF is a 31–35 kDa protein of 268 amino acids, with predicted pI of 4.5–4.64 and shows high immunoreactivity in humans [39, 51–53]. Bioinformatics approaches indicate that this protein does not contain an SS, but it does have a coiled-coil motif [52]. The CagF immunolocalization in bacterial cell fractions show that CagF antibodies are bound to the inner membrane and cytoplasmic fractions (Fig. 1) [52, 53]. CagF is required for the translocation of CagA, and although it has been reported that it is not required for secretion of IL-8 [50], CagF has an important effect on the induction of secretion in the target cells [20, 53]. The interaction between CagF and CagA is very strong and direct, as it is the only prominent protein that CagA coprecipitates with [27, 52, 53]. However, CagF is not delivered with CagA into host cells, so it has been proposed that CagF is a chaperone-like protein that recruits CagA before it interacts with the cag T4SS apparatus [52, 53]. CagA, like other effectors in T4SS, in addition to requiring a secretion-targeting signal within the N- and C-terminal domains, it needs a chaperone in order to be recognized as T4SS substrate before being delivered to the membrane-spanning core complex of the T4S system [52–54]. Although chaperones are proteins that are typically characterized by a low molecular weight and low pI, the molecular weight of CagF is larger [52]. Nevertheless, the CagF cytoplasmic and inner membrane localization and the fact that is not delivered together with CagA within target cells are consistent with its chaperone-like function [52]. Indeed, isothermal titration calorimetry showed that CagF engages CagA with nm affinity to form a 1:1 complex. Peptide arrays and isothermal titration calorimetry also showed that the coiled-coil motif and the C-terminal helix within CagF bind to
domains I-III and domain IV of CagA, respectively [55]. This strong interaction stabilizes CagA and ensures its intact delivery into target cells by the cag T4SS. In contrast, the interaction of CagF and CagA with the core complex components is relatively weak, given that CagA and CagF are detected in the top fraction of a glycerol gradient centrifugation, separately from the core complex, which is in the bottom fraction. This is consistent with an interaction between an effector protein and secretion machinery [47]. Thus, when CagF was used as bait in an immunopurification strategy in a cagA mutant strain, it did not coprecipitate the core complex components, indicating that the CagA effector directly attaches to the core complex [47].

**CagH (Cag20/HP0541)**

CagH is a 39 kDa protein made of 370 amino acids. It lacks a SS, but it is detectable in the membrane fraction [26,38]. It has a conserved flagellar hook-associated protein K motif and it is a bitopic inner membrane protein with only one predicted transmembrane helix span (Fig. 1) [27,38]. CagH shares a weak sequence similarity with CagL and is essential for CagA translocation and the full induction of IL-8 in target cells [20,38]. A mass spectrometry-based analysis revealed that CagH coprecipitates with CagI and the integrin-binding protein CagL [37,38]. Within this complex, CagH and CagI are proposed to strongly interact with the membrane given that they cannot be solubilized by sonication [38]. These two proteins, in addition to having a close interaction, are transcribed as a polycistronic mRNA that includes the genes cagC-cagL and/or cagF-cagL [45]. CagH regulates the number and proportion of pili that form during the interaction between the bacteria and the host. Scanning electron microscopy revealed that a mutant strain lacking this gene exhibits a larger number of elongated and thickened pili than the WT strain [38]. Mutation of cagH reduces the stability of CagL and interferes with the copurification of CagL and CagI [38]. Moreover, the absence of CagH in a cagI mutant strain suggests that CagH stability is CagI dependent [38]. These three proteins share a unique, conserved C-terminal hexapeptide motif that is required for the functional structure of the cag T4SS [38]. The C-terminal motifs of CagI and CagL are required for pilus formation, whereas bacteria with a truncated motif in CagH express pili that are similar in shape and number to the WT strain but are nonfunctional, thus, demonstrating the important role of the C-terminal motif of CagH in pilus biogenesis and function [38].

**CagI (Cag19/HP0540)**

CagI is a protein of 41.5 kDa with a clearly predicted SS for export into the periplasm space [38]. CagI, like CagL, has a conserved C-terminal hexapeptide motif that is required for pilus formation, and consequently, for CagA translocation and IL-8 induction in the target cells [38]. CagI is encoded within the operon cagC-cagL and/or within the operon cagF-cagL and is translated along, among others, with CagH and the putative β1 integrin ligand, CagL [45]. Interestingly, the stability of the CagH, CagL, and CagI components decreases if one or more of these interacting proteins is missing [38,56]. Indeed, CagI stability depends on several components of the cag PAI. For example, in isogenic mutants in cagX, cagY, cagH, or cagG from *H. pylori* strain P12 grown in the absence of host cells, there is essentially no detectable CagI [56]. Similarly, the deletion of cag3, cagW, cagV, cagU, cagM, cagL, or cagE significantly reduces the levels of CagI [56]. Similar results were obtained for *H. pylori* strain 26695 grown in the absence of target cells, where CagI was not detected in the isogenic mutants of cagY, cagX, cagV, cagT, cagM, cag3, and cagG [41]. In contrast, the deletion of cagI does not affect the stability of CagX, CagT, CagM, CagF, or CagZ, but it reduces CagL levels, and CagH becomes undetectable. However, cagI expression is not affected by the deletion of the cagF, cagZ, or cagA genes [41,56].

CagI directly interacts with the host factor β1 integrin, which acts as a host cell-surface receptor on gastric epithelial cells [37,41]. A recent study provided evidence that CagI may be a periplasmic protein that is only loosely associated with the outer membrane, with only a few molecules partially exposed on the bacterial surface, especially on the pilus-like structure (Fig. 1) [41]. Nevertheless, assays performed in the absence of the target cells showed that CagI is not involved in CagA translocation to the surface of the bacterium, given that CagA is detected on the bacterial cell surface of cagI mutant strains [41]. However, CagI is essential for transporting CagA in the presence of gastric epithelial cells, given that cagI mutant strains fail to form pili [38]. Likewise, CagL was detected on pili [57], and given that it interacts with CagI and CagH [38,41,56], it is probable that these three proteins form part of the pili together with CagY and CagC [26,35,43,58]. Nevertheless, it has not been experimentally demonstrated that CagI and CagH are pili components [38,56,59]. Interestingly, in contrast to cagI mutant strains that fail to form pili, the cagC and cagY mutants were defective in T4SS function but retained the capacity for pili formation [43].
results are probably due to the substantially larger size of CagY (VirB10), which could give a different function from those of VirB10 in other bacterial species. In the case of CagC, it exhibits a weak sequence relationship with other VirB2 orthologs and it is not the major component of the pili, as it occurs in *A. tumefaciens* (Fig. 1) [29,43].

**CagZ (Cag6/HP0526)**

Based on its primary sequence, CagZ is not homologous to any proteins of the VirB/VirD4 secretion machinery of *A. tumefaciens* or any of the components of T4SSs. Nevertheless, *cagZ* mutant strains are severely impaired in IL-8 induction and are unable to transport CagA into AGS cells [34,40]. *cagZ* is transcribed into a polycistronic mRNA that includes the genes *virD4* and *virB11* [44,45]. CagZ is a 23–24 kDa protein, with a predicted acid pI value of 5–5.11, and it is 199 amino acids long [32,60]. Interestingly, CagZ does not have a predicted amino-terminal SS, but is found in both the soluble and membrane fractions [32,40]. The determination of its three-dimensional crystal structure showed that CagZ is comprised of a single compact L-shaped domain containing seven alpha helices running antiparallel to each other [60]. Seventy percent of its residues are in an alpha-helix conformation; there are no beta-sheet domains present, and it has a disordered C-terminal end. These three-dimensional structures of CagZ do not present structural homologs; thus, CagZ is considered to represent a new type of protein folding [60].

In coimmunoprecipitation assays of Cag fusion proteins expressed in *Escherichia coli*, CagZ was found to interact with CagV, CagM, CagX, CagS, CagI, and Cag5 [32,40]. Curiously, of all these interactions, the most important is the interaction between CagZ and Cag5, which has a stabilizing effect on Cag5 [40]. Additionally, pull-down experiments showed that the interaction of the coupling protein homolog Cag5 with CagA is independent of the presence of CagZ [40]. Thus, the presence of CagA is not required for the interaction between Cag5 and CagZ [40]. The binding of CagZ and Cag5 to the membrane-spanning core complex could be a means to recruit CagA to the translocation channel of the *cag* T4S system (Fig. 1) [40]. In this context, the potential role of CagZ as a chaperone can be inferred by an analogy with the type III secretion system chaperone proteins [40,60]. CagZ, like these chaperons, is typically characterized by a low pI and a low molecular weight, with a primarily negative molecular surface and a cluster of negative residues in two of its helices [32,60].

**CagU (Cag11/HP0531)**

CagU is an inner membrane protein of ~ 24 kDa (218 amino acids) [26,27]. It does not have a predicted SS but has three predicted transmembrane helices between residues 101–119, 140–158, and 178–196 [26,27]. As with CagZ, CagU also has an unstructured region located, in particular, between residues 1–21 [61]. CagU is transcribed as a polycistronic mRNA that includes the gene *cagT* (VirB7) [44,45]. CagU seems to be unique as no interaction between CagU and any of the other structural components of the *cag* T4SS has been described so far, even with its own operon partner, CagT [44,59]. *cagU* mutants are unable to transport CagA or induce IL-8 secretion in target cells, but the reasons for these defects are not completely understood. It has been reported that the operon *cagU-cagT* is required for pili production [62], but only CagT has been identified as necessary for the formation of pili by a mutation and complementation experiment [43]. Nevertheless, it is known that the absence of CagU may influence CagI stability; the deletion of *cagU* in isogenic mutants of the *H. pylori* strain P12 significantly reduced the levels of CagI, which plays a main role in pili formation [56]. Additionally, CagU may contribute to formation of a cytoplasmic membrane pore with CagH and CagW (VirB6), which has been proposed as an inner membrane-associated structural component (Fig. 1) [26,27].

**CagG (Cag21/HP0542)**

The CagG consists of 142 amino acids with a predicted pI between 4 and 6 and a SS that suggests its localization is in the periplasm (Fig. 1) [22,26,56]. In *H. pylori*, *cagG* mutant strains are incapable of delivering CagA into gastric epithelial cells, even though they retain the capacity to induce IL-8 production [34].

*cagG* is transcribed as a polycistronic mRNA that includes the genes *cagC-cagL* and/or *cagF-cagL* [45]. Curiously, although *cagG* and *cagH* are expressed simultaneously, there is no evidence that their products, CagG and CagH, interact. In isogenic mutants in *cagG* from *H. pylori* strain P12 grown in the absence of host cells, there is essentially no detectable CagI and CagL [56]. Similar results were obtained for *H. pylori* strain 26695 grown in the absence of target cells, where CagH and CagI were not detected and no pili-like structure was observed in the isogenic mutants of *cagG* [41]. However, as none of these studies included complementation of *cagG* mutants, the actual contribution of this protein to all these phenotypes is not clear. However, the fact that these genes form an...
components leads to erroneous protein oncoprotein CagA and peptidoglycan into host target cells. Despite recent studies on the VirB/VirD4 homolog components of the cag T4SS, the roles that they play in the secretion of CagA and the induction of IL-8 in gastric epithelial cells remain unclear. A lack of knowledge of the functional interactions and structural characteristics are particularly evident for these unique constituents of cag T4SS. Multiple lines of evidence indicate that this lack of knowledge is partially due to the fact that the encoded proteins do not act individually but rather in complexes that form multimeric structures.

Conclusion

Highly virulent H. pylori strains harbor a cag PAI encoding a T4SS. This T4SS forms a structure analogous to a macromolecular syringe to inject the oncoprotein CagA and peptidoglycan into host target cells. The presence of one or several of the functionally interacting complexes is necessary to clarify the assembly and structural organization of these unique components of the cag T4SS. Therefore, this new knowledge will help to understand the complex architecture of this apparatus of secretion and the molecular mechanisms that regulate the activation of CagA secretion when the cag T4SS contacts gastric epithelial cells.

Acknowledgements

This work was supported by CONACYT grant no. V51136-M.

Author contributions

EM generated the study design and wrote the manuscript and the final revision; MF-E contributed to the literature search and wrote the manuscript and the final revision; GRA-G generated the study design, contributed to the literature review, and coordinated the manuscript writing and the final revision.

References

1 Kim SS, Ruiz VE, Carroll JD & Moss SF (2011) Helicobacter pylori in the pathogenesis of gastric cancer and gastric lymphoma. Cancer Lett 305, 228–238.
2 The Eurogast Study Group (1993) An international association between Helicobacter pylori infection and gastric cancer. Lancet 341, 1359–1362.
3 Backert S & Blaser MJ (2016) The role of CagA in the gastric biology of Helicobacter pylori. Cancer Res 76, 203–222.
4 Brandt S, Kwok T, Hartig R, König W & Backert S (2005) NF-kappaB activation and potentiation of proinflammatory responses by the Helicobacter pylori CagA protein. Proc Natl Acad Sci USA 102, 9300–9305.
5 Hayashi T, Senda M, Morohashi H, Higashi H, Horio M, Kashiba Y, Nagase L, Sasaya D, Shimizu T, Venugopal N et al. (2012) Tertiary structure-function analysis reveals the pathogenic signaling potentiation mechanism of Helicobacter pylori oncogenic effector CagA. Cell Host Microbe 12, 20–33.
6 Nesic’ D, Bui L, Lu X & Stebbins CE (2014) Structure of the Helicobacter pylori CagA oncoprotein bound to the human tumor suppressor ASPP2. Proc Natl Acad Sci USA 111, 1562–1567.
7 Viala J, Chaput C, Boneca IG, Cardona A, Girardin SE, Moran AP, Athman R, Mémé S, Huere MR, Coyle AJ et al. (2004) Nod1 responds to peptidoglycan delivered by the Helicobacter pylori cag pathogenicity island. Nat Immunol 5, 1166–1174.
8 Hatakeyama M (2014) Helicobacter pylori CagA and gastric cancer: a paradigm for hit-and-run carcinogenesis. Cell Host Microbe 15, 306–316.
9 Kusters JG, van Vliet AH & Kuipers EJ (2006) Pathogenesis of Helicobacter pylori infection. Clin Microbiol Rev 19, 449–490.
10 Hamilton HL, Dominguez NM, Schwartz KJ, Hackett KT & Dillard JP (2005) Neisseria gonorrhoeae secretes chromosomal DNA via a novel type IV secretion system. Mol Microbiol 55, 1704–1721.
11 Hofreuter D, Odenbreit S & Haas R (2001) Natural transformation competence in Helicobacter pylori is mediated by the basic components of a type IV secretion system. Mol Microbiol 41, 379–391.
12 Tegtmeyer N, Wessler S & Backert S (2011) Role of the cag-pathogenicity island encoded type IV secretion system in Helicobacter pylori pathogenesis. FEBS J 278, 1190–1202.
13 Trokter M, Felisberto-Rodrigues C, Christie PJ & Waksman G (2014) Recent advances in the structural and molecular biology of type IV secretion systems. Curr Opin Struct Biol 27, 16–23.
14 Pituschke A & Hirt H (2010) New insights into an old story: Agrobacterium-induced tumour formation in plants by plant transformation. EMBO J 29, 1021–1032.
15 Lai EM & Kado CI (1998) Processed VirB2 is the major subunit of the promiscuous pilus of Agrobacterium tumefaciens. J Bacteriol 180, 2711–2717.
16 Eisenbrandt R, Kalkum M, Lai EM, Lurz R, Kado CI & Lanka E (1999) Conjugative pili of IncP plasmids, and the Ti plasmid T pilus are composed of cyclic subunits. J Biol Chem 274, 22548–22555.

17 Lai EM, Eisenbrandt R, Kalkum M, Lanka E & Kado CI (2002) Biogenesis of T pili in Agrobacterium tumefaciens requires precise VirB2 propilin cleavage and cyclization. J Bacteriol 184, 327–330.

18 Schmidt-Eisenlohr H, Domke N, Angerer C, Wanner G, Zambryski PC & Baron C (1999) Vir proteins stabilize VirB5 and mediate its association with the T pilus of Agrobacterium tumefaciens. J Bacteriol 181, 7485–7492.

19 Christie PJ, Atmakuri K, Krishnamoorthy V, Jakubowski S & Cascales E (2005) Biogenesis, architecture, and function of bacterial type IV secretion systems. Annu Rev Microbiol 59, 451–485.

20 Selbach M, Moese S, Meyer TF & Backert S (2002) Functional analysis of the Helicobacter pylori cag pathogenicity island reveals both VirD4-CagA-dependent and VirD4-CagA-independent mechanisms. Infect Immun 70, 665–671.

21 Alm RA, Ling LS, Moir DT, King BL, Brown ED, Doig PC, Smith DR, Noonan B, Guild BC, deJonge BL et al. (1999) Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori. Nature 397, 176–180.

22 Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, Ketchum KA, Klenk HP, Gill S, Dougherty BA et al. (1997) The complete genome sequence of the gastric pathogen Helicobacter pylori. Nature 388, 539–547.

23 Censini S, Lange C, Xiang Z, Crabtree JE, Ghia P, Borodovsky M, Rappuoli R & Covacci A (1996) cag, a pathogenicity island of Helicobacter pylori, encodes type I-specific and disease-associated virulence factors. Proc Natl Acad Sci USA 93, 14648–14653.

24 Bourzac KM & Guillermín K (2005) Helicobacter pylori-host cell interactions mediated by type IV secretion. Cell Microbiol 7, 911–919.

25 Covacci A, Telford JL, Del Giudice G, Parsonnet J & Rappuoli R (1999) Helicobacter pylori virulence and genetic geographic. Science 284, 1328–1333.

26 Fischer W (2011) Assembly and molecular mode of action of the Helicobacter pylori Cag type IV secretion apparatus. FEBS J 278, 1203–1212.

27 Kutter S, Buhrdorf R, Haas J, Schneider-Brachert W, Haas R & Fischer W (2008) Protein subassemblies of the Helicobacter pylori Cag type IV secretion system revealed by localization and interaction studies. J Bacteriol 190, 2161–2171.

28 Zhong Q, Shao S, Mu R, Wang H, Huang S, Han J, Huang H & Tian S (2011) Characterization of peptidoglycan hydrolase in Cag pathogenicity island of Helicobacter pylori. Mol Biol Rep 38, 503–509.

29 Terradot L & Waksman G (2011) Architecture of the Helicobacter pylori Cag-type IV secretion system. FEBS J 278, 1213–1222.

30 Backert S, Fronzes R & Waksman G (2008) VirB2 and VirB5 proteins: specialized adhesins in bacterial type-IV secretion systems? Trends Microbiol 16, 409–413.

31 Buhrdorf R, Förster C, Haas R & Fischer W (2003) Topological analysis of a putative virB8 homologue essential for the cag type IV secretion system in Helicobacter pylori. Int J Med Microbiol 293, 213–217.

32 Busler VJ, Torres VJ, McClain MS, Tirado O, Friedman DB & Cover TL (2006) Protein-protein interactions among Helicobacter pylori cag proteins. J Bacteriol 188, 4787–4800.

33 Akopyants NS, Clifton SW, Kersulyte D, Crabtree JE, Youree BE, Reece CA, Bukanov NO, Drazek ES, Roe BA & Berg DE (1998) Analyses of the cag pathogenicity island of Helicobacter pylori. Mol Microbiol 28, 37–53.

34 Fischer W, Pilz J, Buhrdorf R, Gebert B, Odenbreit S & Haas R (2001) Systematic mutagenesis of the Helicobacter pylori cag pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8. Mol Microbiol 42, 1337–1348.

35 Tegtmeier N, Lind J, Schmid B & Backert S (2014) Helicobacter pylori CagL Y58/E59 mutation turns-off type IV secretion-dependent delivery of CagA into host cells. PLoS ONE 9, e97782.

36 Conradi J, Tegtmeier N, Wózna M, Wissbrock M, Michalec C, Gagell C, Cover TL, Frank R, Sewald N & Backert S (2012) An RGD helper sequence in CagL of Helicobacter pylori assists in interactions with integrins and injection of CagA. Front Cell Infect Microbiol 2, 1–15.

37 Jiménez-Soto LF, Kutter S, Sewald X, Ertl C, Weiss E, Kapp U, Rohde M, Pirch T, Jung K, Retta SF et al. (2009) Helicobacter pylori type IV secretion apparatus exploits 1 integrin in a novel RGD-independent manner. PLoS Pathog 5, e1000684.

38 Shaffer CL, Gaddy JA, Loh JT, Johnson EM, Hill S, Hennig EE, McClain MS, McDonald WH & Cover TL (2011) Helicobacter pylori exploits a unique repertoire of type IV secretion system components for pilus assembly at the bacteria-host cell interface. PLoS Pathog 7, e1002237.

39 Cendron L & Zanotti G (2011) Structural and functional aspects of unique type IV secretory components in the Helicobacter pylori cag-pathogenicity island. FEBS J 278, 1223–1231.

40 Jurik A, Haussner E, Kutter S, Pattis I, Prassl S, Weiss E & Fischer W (2010) The coupling protein CagB and its interaction partner CagZ are required for type IV secretion of the Helicobacter pylori CagA protein. Infect Immun 78, 5244–5251.
41 Kumar N, Shariq M, Kumari R, Tyagi RK & Mukhopadhyay G (2013) Cag type IV secretion system: CagI independent bacterial surface localization of CagA. PLoS ONE 8, e74620.

42 Pinto-Santini DM & Salama NR (2009) Cag3 is a novel essential component of the Helicobacter pylori Cag type IV secretion system outer membrane subcomplex. J Bacteriol 191, 7343–7352.

43 Johnson EM, Gaddy JA, Voss BJ, Hennig EE & Cover TL (2014) Genes required for assembly of pili associated with the Helicobacter pylori cag type IV secretion system. Infect Immun 82, 3457–3470.

44 Ta LH, Hansen LM, Sause WE, Shiva O, Millstein A, Ottemann KM, Castillo AR & Solnick JV (2012) Conserved transcriptional unit organization of the cag pathogenicity island among Helicobacter pylori strains. Front Cell Infect Microbiol 2, 1–14.

45 Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeis S, Sittka A, Chabas S, Reiche K, Hackermüller J, Reinhardt R et al. (2010) The primary transcriptome of the major human pathogen Helicobacter pylori. Nature 464, 250–255.

46 Voss BJ, Gaddy JA, McDonald WH & Cover TL (2014) Analysis of surface-exposed outer membrane proteins in Helicobacter pylori. J Bacteriol 196, 2455–2471.

47 Frick-Cheng AE, Pyburn TM, Voss BJ, McDonald WH, Ohi MD & Cover TL (2016) Molecular and structural analysis of the Helicobacter pylori cag type IV secretion system core complex. MBio 7, 167–176.

48 Ling F, Wang X, Dai D, Yu M, Chen C, Qian J, Liu C, Zhang Y, Ding J, Guan XW et al. (2013) The Helicobacter pylori Protein CagM is located in the transmembrane channel that is required for CagA translocation. Curr Microbiol 67, 531–536.

49 Bourzac KM, Satkamp LA & Guillemin K (2006) The Helicobacter pylori cag pathogenicity island protein CagN is a bacterial membrane-associated protein that is processed at its C terminus. Infect Immun 74, 2537–2543.

50 Fischer W, Püls J, Buhrdorf R, Gebert B, Odenbreit S & Haas R (2003) Systematic mutagenesis of the Helicobacter pylori cag pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8. Mol Microbiol 47, 1759.

51 Seydel A, Tasca E, Berti D, Rappuoli R, Del Giudice G & Montecucco C (2002) Characterization and immunogenicity of the CagF protein of the cag pathogenicity island of Helicobacter pylori. Infect Immun 70, 6468–6470.

52 Couturier MR, Tasca E, Montecucco C & Stein M (2006) Interaction with CagF is required for translocation of CagA into the host via the Helicobacter pylori type IV secretion system. Infect Immun 74, 273–281.

53 Pattis I, Weiss E, Lauks R, Haas R & Fischer W (2007) The Helicobacter pylori CagF protein is a type IV secretion chaperone-like molecule that binds close to the C-terminal secretion signal of the CagA effector protein. Microbiology 153, 2896–2909.

54 Holhfeld S, Pattis I, Püls J, Plano GV, Haas R & Fischer W (2006) A C-terminal translocation signal is necessary, but not sufficient for type IV secretion of the Helicobacter pylori CagA protein. Mol Microbiol 59, 1624–1637.

55 Bonso DA, Weiss E, Iosub-Amir A, Reingewertz TH, Chen TW, Haas R, Friedler A, Fischer W & Sundberg EJ (2013) Characterization of the translocation-competent complex between the Helicobacter pylori oncogenic protein CagA and the accessory protein CagF. J Biol Chem 288, 32897–32909.

56 Pham KT, Weiss E, Jiménez Soto LF, Breithaupt U, Haas R & Fischer W (2012) CagI is an essential component of the Helicobacter pylori cag type IV secretion system and forms a complex with CagL. PLoS ONE 7, e53541.

57 Kwok T, Zabler D, Urman S, Rohde M, Hartig R, Wessler S, Misselwitz R, Berger J, Sewald N, König W et al. (2007) Helicobacter exploits integrin for type IV secretion and kinase activation. Nature 449, 862–866.

58 Barrozo RM, Cooke CL, Hansen LM, Lam AM, Gaddy JA, Johnson EM, Cariaga TA, Suarez G, Peek RM Jr, Cover TL et al. (2013) Functional plasticity in the type IV secretion system of Helicobacter pylori. PLoS Pathog 9, e1003189.

59 Backert S, Tegtmeyer N & Fischer W (2015) Composition, structure and function of the Helicobacter pylori cag pathogenicity island encoded type IV secretion system. Future Microbiol 10, 955–965.

60 Cendron L, Seydel A, Angelini A, Battistutta R & Zanotti G (2004) Crystal structure of CagZ, a protein from the Helicobacter pylori pathogenicity island that encodes for a type IV secretion system. J Mol Biol 340, 881–889.

61 Dostáňová Z, Cszimkó V, Tompa P & Simon I (2005) IUPred web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 21, 3433–3434.

62 McClain MS, Duncan SS, Gaddy JA & Cover TL (2013) Control of gene expression in Helicobacter pylori using the Tet repressor. J Microbiol Methods 95, 336–341.