ROLE OF UNIQUE *HELICOBACTER PYLORI* PROTEINS IN THE *CAG*

PATHOGENICITY ISLAND-ENCODED TYPE IV SECRETION SYSTEM

By

Carrie Leigh Shaffer

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Microbiology and Immunology

December, 2011

Nashville, Tennessee

Approved:

D. Borden Lacy, PhD

Seth R. Bordenstein, PhD

Timothy L. Cover, MD

Andrew J. Link, PhD

Eric P. Skaar, PhD, MPH

Roy Zent, MD, PhD

For my dad – you've always been my hero.

Acknowledgements

First and foremost, I would like to thank my incredible mentor Dr. Timothy L. Cover for inviting me to join his lab and pursue all of my varied curiosities. I appreciate Dr. Cover's continual patience, his kindness, and most important, his sense of humor. He always sees the silver lining in every situation, and provides enormous encouragement during difficult times, including millennia flooding in a metropolitan area. His enthusiasm for our work, and an ever curious mind provided a source of inspiration for me and all of the members of his lab. Tim, I am honored to have been a member of your lab, and thankful for the guidance you have provided that has made me the scientist that I am today.

I would also like to thank my esteemed committee for their guidance and very helpful suggestions throughout my training. Thank you to my Chair, Dr. Borden Lacy, for your support; Dr. Eric Skaar for your sense of humor, enthusiasm for teaching, and for a fellow bacteriologist's insight; Dr. Andy Link for your sarcastic (and very funny) comments, and expertise in mass spectrometry; and Dr. Seth Bordenstein and Dr. Roy Zent for providing support to my project from an evolutionary biology and human host perspective, respectively. Thank you to the faculty of the Department of Microbiology and Immunology, as well as the amazing Jean Tidwell. I'd like to thank my classmates for all of the laughter, hilarious comments, and camaraderie during our training. I especially thank Major John Schmitt for keeping me sane.

Thank you to members of the Cover Lab for helpful ideas and suggestions in lab meetings, and for valuable contributions in collaborative efforts. Thank you to my pseudo-mentor Dr. Hayes McDonald for hours of (often hilarious) conversation, and for

teaching me the intricacies of mass spectrometry. Thank you to the most wonderful, knowledgeable Ms. Salisha Hill for all of the beautiful mass spectrometry results – and a special thanks to Salisha and Dr. Kristie Rose for all of the laughter during my frequent visits to the mass spectrometry facility. I would like to thank Kristie for all of her expertise and perspective during collaborative efforts, especially in projects involving very strange, post-translationally modified proteins.

And finally, thank you to my friends and family for your continual support and encouragement. I would like to especially thank my mom, dad, and sister for teaching me how to laugh, and for never doubting me. Your support and love has allowed me to succeed. Thank you to my best friend, Jamie Floyd, for being my partner-in-crime, my fellow explorer in all of our travel expeditions, and for the endless laughs, the obnoxious and ridiculous number of inside jokes that we share, consuming massive amounts of food (most meals enjoyed in silence) together, and for endless support and friendship. Thank you to one of my dearest friends Dr. Stacy Duncan, for bringing joy to the lab environment – I will miss the sounds of our laughter ringing through the lab. Thank you to Brad Voss for helping me deal with a demolished condo after the Nashville flood of 2010, and for sharing countless numbers of meals, and many, many laughs.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

PMN	polymorphonuclear neutrophil
IARC	International Agency for Research on Cancer
MALT	mucosa-associated lymphoid tissue
MLST	multilocus sequence typing
PAI	pathogenicity island
T4SS	type IV secretion system
Y2H	yeast 2-hybrid
WT	wild-type
MudPIT	Multidimensional Protein Identification Technology
SCX	strong cation exchange
MS	mass spectrometry
GST	glutathione S-transferase
SDS	sodium dodecyl sulfate
PAGE	polyacrylamide gel electrophoresis
HA	hemagglutinin
FBS	fetal bovine serum
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
ELISA	enzyme-linked immunosorbent assay
MOI	multiplicity of infection
EDTA	ethylenedinitrilo-tetraacetic acid
NP-40	nonyl phenoxypolyethoxylethanol
PCR	polymerase chain reaction
PBS	phosphate-buffered saline

RIPA	radio-immunoprecipitation assay
DTT	dithiothreitol
TBS	tris-buffered saline
HPLC	high pressure liquid chromatography
ACN	acetonitrile
FDR	false discovery rate
PMSF	phenylmethylsulphonyl fluoride
TEM	transmission electron microscopy
FESEM	field emission scanning electron microscopy
ANOVA	analysis of variance
CA	carbonic anyhdrase
PK	proteinase K
OMP	outer membrane protein
T3SS	type III secretion system
PSSM	position-specific scoring matrix
CDD	conserved domain database
DSP	dithiobis[succinimidyl propionate]
DMSO	dimethyl sulfoxide
TCA	trichloroacetic acid
RT-PCR	real-time polymerase chain reaction
Fur	ferric uptake regulator
TAP	tobacco acid pyrophosphatase
MRM	multiple reaction monitoring

CHAPTER I

INTRODUCTION

Bacteriology is a risky business

Oftentimes, the difference between enormous success and miserable failure is having the courage to take calculated risks. Risky behavior frequently results in adverse consequences; however, a high risk-high reward mentality combined with the intestinal fortitude required to participate in such behavior will, on occasion, result in spectacular success. Calculated risk was the driving force behind the team of Dr. Barry Marshall and Dr. J. Robin Warren, two Australian physicians who in 1984 were first able to culture a bacterium related to *Campylobacter* from a gastric biopsy acquired from a patient presenting with gastritis [1]. This novel bacterium was described as small, spiral-shaped, and having up to five unipolar flagella, differentiating it from *Campylobacter*, which have single polar flagella. Although an association between this novel 'Pyloric *Campylobacter*' and gastritis had been described in multiple countries by 1985 [2,3], it remained unclear whether this bacterium was the primary cause of gastritis, or if the bacterium was an opportunistic commensal of the abnormal gastric mucosa. Moreover, the contribution, if any, that the bacteria were making to the development of gastric diseases other than gastritis could not be ascertained. Therefore, Marshall considered it important to fulfill Koch's Postulates, utilizing himself as the voluntary human subject [4].

Marshall consumed a liquid culture of 'Pyloric *Campylobacter*' containing approximately 10^9 colony forming units, and developed acute gastritis within 10 days post-ingestion. Gastric biopsy at 10 days post-infection demonstrated the presence of polymorphonuclear neutrophil (PMN) infiltrates, abnormal gastric epithelial cells, and 'Pyloric *Campylobacter*' attached to the luminal surface of the antral epithelial cells [4].

1

At 14 days post-ingestion, Marshall initiated antibiotic therapy, which was able to completely resolve his symptoms within 24 hours [4].

This risky experiment was the first to report a direct association of infection by 'Pyloric *Campylobacter*' (later renamed *Helicobacter pylori*) with development of acute gastritis, providing resolution to a medical curiosity that had persisted for decades. Marshall and Warren's research reintroduced the concept, first proposed at the turn of the century, that a microorganism was the causative agent of gastric disease. About a decade after Marshall's famous experiment (or period of questionable sanity, depending on how one views his methodology), the World Health Organization International Agency for Research on Cancer (IARC) designated *H. pylori* infection as a class I carcinogen [5], recognizing the important role of the bacterium as a contributing factor in the development of gastric cancer. In 2005, Marshall's calculated risk, in addition to work performed with Dr. Warren, culminated in international recognition when the Australian duo received the Nobel Prize in Physiology or Medicine 'for their discovery of the bacterium *Helicobacter pylori* and its role in gastritis and peptic ulcer disease.'

Clinical features of infection by *Helicobacter pylori* and its role in causation of severe gastric disease

Arguably one of the most successful human pathogens, the Gram-negative *H. pylori* has been in intimate association with the human host for an estimated 60,000-100,000 years [6,7,8,9], and currently colonizes the gastric mucosa of approximately 50% of the global population [10]. Infection by *H. pylori* is almost always accompanied by superficial chronic gastritis [11]. In the absence of effective antibiotic therapy, colonization of the stomach by *H. pylori* persists for decades or the lifetime of the individual. The majority of *H. pylori*-infected individuals do not develop symptoms related to their infection; however, the presence of *H. pylori* is associated with an increased risk for the development of several severe gastric diseases including peptic

ulcers, gastric adenocarcinoma, and mucosa-associated lymphoid tissue (MALT) lymphoma [5,12,13,14,15,16,17,18,19,20,21]. Gastric adenocarcinoma is the second leading cause of cancer-related death world-wide. The incidence of gastric cancer varies geographically, with the highest rates occuring in Japan, China, South America, and Eastern European nations [22].

Epidemiological case-control studies have demonstrated a significantly higher incidence of gastric cancer in *H. pylori*-infected individuals than in uninfected persons. Infection by *H. pylori* is reported to increase the risk of gastric cancer 2-16 fold [21]. Additionally, *H. pylori* infection is significantly more prevalent in patients with peptic ulcer disease and gastric lymphoma, as compared to age-matched controls [18,20]. Recent evidence suggests that infection with *H. pylori* contributes to iron deficiency anemia [23]. However, in some cases, infection by *H. pylori* may be beneficial to the host by serving to protect against development of certain diseases, including esophageal reflux disease, and cancer of the cardia and esophagus. In addition, infection by *H. pylori* may lead to secondary effects that may help to prevent *Mycobacterium tuberculosis* infection from progressing to active tuberculosis [24], and may play a role in providing protection against development of asthma [25]. The protective properties of infection by *H. pylori* are intriguing, but nonetheless, many years of investigation will be required to determine the mechanisms by which *H. pylori* may provide benefits to its host.

Genetic diversity of *Helicobacter pylori* and the *cag* pathogenicity island

H. pylori infection can result in a broad range of clinical outcomes. The clinical outcomes of infection are likely determined by a combination of factors, including host genetic and physiologic factors, environmental factors, and genetic diversity of the bacterial strain. Nearly every strain of *H. pylori* is unique at the DNA level [26]. Comparative analysis of multiple *H. pylori* strains using methods such as multilocus sequence typing (MLST) [6,8,9,27,28] and DNA microarray [29,30] have revealed

considerable differences among *H. pylori* strains [31,32]. One striking difference in the gene content among strains is the presence or absence of a 40-kilobase chromosomal region referred to as the *cag* pathogenicity island (*cag* PAI) [33,34]. The *cag* PAI is harbored by approximately 50% of *H. pylori* isolates in the United States, and can be present in as many as 90% of *H. pylori* isolates from East Asian nations [35]. Additionally, analysis of *H. pylori* isolates from patients with gastritis compared with *H. pylori* isolates from patients with gastric cancer or peptic ulcer disease revealed several genetic markers associated with higher virulence. The *cag* PAI, type s1 *vac*A alleles, type I *hop*Q alleles, and functional *bab*A alleles are more common in strains associated with severe gastric disease, with several of these virulence determinants frequently occurring together in a single isolate [21,36,37,38,39,40,41].

The G+C content of the *cag* PAI is substantially different from the G+C content of the *H. pylori* chromosome, suggesting that the island was acquired via horizontal gene transfer from an unidentified source [31]. Annotation of the *cag* PAI indicates that approximately 27 proteins are encoded within the island, some of which are homologous to components of the prototypical *Agrobacterium* type IV secretion system (T4SS) (Figure 1-1) [42]. However, in some cases, the level of sequence homology of *cag* genes to *vir* genes is quite weak, demonstrating a phylogenetic divergence from a common ancestor. Table 1 lists multiple commonly used *cag* gene designations.

Table 1-1. Genetic analysis of proteins encoded by the *cag* pathogenicity island

Gene Number [a]	Protein	Gene Name	Homology [b]	Function	Phenotype of strains with mutation in indicated genes	
					CagA Translocation [c]	IL-8 Induction [d]
HP0547	CagA	cag26		Effector	-	-
HP0546	CagC	cag25	VirB2		-	+
HP0545	CagD	cag24			++	+
HP0544	CagE	cag23	VirB4	ATPase	-	-
HP0543	CagF	cag22			-	+
HP0542	CagG	cag21			-	+
HP0541	CagH	cag20			-	-
HP0540	CagI	cag19			-	+
HP0539	CagL	cag18			-	-
HP0538	CagN	cag17			+	+
HP0537	CagM	cag16			-	-
HP0536	CagP	cag15			++	+
HP0535	CagQ	cag14			++	+
HP0534	CagS	cag13			++	+
HP0532	CagT	cag12	VirB7		-	-
HP0531	CagU	cag11			-	-
HP0530	CagV	cag10	VirB8		-	-
HP0529	CagW	cag9			-	-
HP0528	CagX	cag8	VirB9		-	-
HP0527	CagY	cag7	VirB10		-	-
HP0526	CagZ	cag6			+	+
HP0525			VirB11	ATPase	-	-
HP0524	Cag5	cag5	VirD4	Coupling protein	-	+
HP0523	Cag4	cag4	VirB1	Lytic transglycosylase	-	-
HP0522	Cag3	cag3			-	-
HP0521	Cag2	cag2			++	+
HP0520	Cag1	cag1			++	+

[a] Based on the *H. pylori* 26695 genome annotation.
[b] Based on comparison to the *Agrobacterium tumefaciens* T4SS
[c] CagA translocation into AGS cells; +, yes; -, no.
[d] *H. pylori*-induced IL-8 secretion by AGS cells; +, yes; - no.
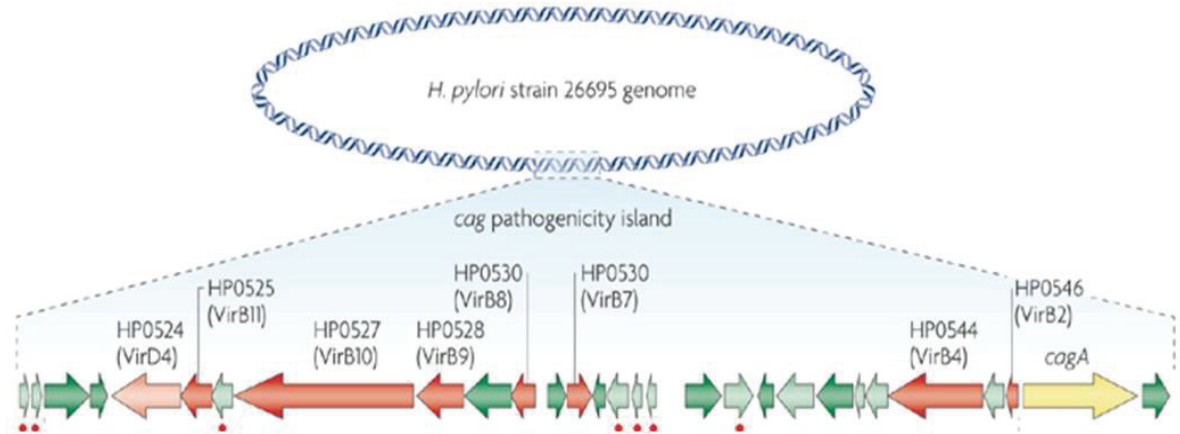
Figure 1-1. Gene arrangement of the *cag* pathogenicity island (*cag* PAI) in *H. pylori* 26695. The *cag* PAI is a 40-kilobase chromosomal region which encodes 27 proteins that are involved in the formation of a type IV secretion system (T4SS). Some of the genes within the *cag* PAI have similarity to components of the T4SS of the plant pathogen *Agrobacterium tumefaciens* (red arrows). The T4SS is utilized to translocate the oncoprotein CagA (yellow arrow) into the host gastric epithelial cell. In addition to CagA translocation, the T4SS also induces the production of IL-8 by gastric epithelial cells. Genes indicated by dark red and dark green arrows are required for both CagA translocation and induction of IL-8. Some genes are required for CagA translocation, but are not involved in eliciting IL-8 synthesis and secretion by gastric epithelial cells (light red and light green arrows). Seven genes encoded within the *cag* PAI are not required for either CagA translocation or induction of IL-8 secretion (indicated by red dots). Image reprinted from [7].

Helicobacter pylori *cag* PAI-dependent effects on the host gastric epithelium

CagA, a 140 kilodalton immunodominant antigen, enters gastric epithelial cells and causes numerous cellular alterations. Additional proteins encoded by the *cag* PAI assemble into a type IV secretion system (T4SS) that ushers CagA into gastric epithelial cells [43,44]. Following translocation, CagA is phosphorylated by host cell kinases at tyrosine residues contained within EPIYA motifs in the C-terminal region of the protein [43,44,45]. CagA, in both phosphorylated and non-phosphorylated forms, is able to interact with host cell signaling proteins, resulting in an assortment of consequences, including cytoskeletal alterations, disruption of cellular junctions, and altered cellular adhesion and polarity [42,45,46,47].

Within gastric epithelial cells, phosphotyrosine-CagA activates SHP-2 [48], leading to dephosphorylation of multiple cell proteins, and inactivation of the catalytic activity of c-Src tyrosine kinase [49]. This cascade inactivates multiple cellular proteins, including the actin-binding protein cortactin [49], prompting the gastric epithelium to undergo cytoskeletal rearrangements that ultimately result in a vast morphological change referred to as the hummingbird phenotype. CagA has been shown to interact with Grb-2, resulting in activation of the Ras/MEK/ERK signaling pathway, cell scattering, membrane ruffling, and cell proliferation [50]. Interaction of CagA with the hepatocyte growth factor receptor c-Met results in an increase in cell motility [51]. CagA has been shown to co-localize with the intercellular junction proteins ZO-1 and JAM, resulting in the disruption of tight junctions to allow for bacterial invasion to the basolateral surface of polarized epithelial cells [47]. Furthermore, CagA has the ability to stimulate a change in polarized cell monolayers, leading to a shift from a polarized to an invasive phenotype [52]. Notably, activation of β-catenin by CagA leads to a transcriptional upregulation of oncogenes [53].

7

Numerous proteins encoded by the *cag* PAI are required for the activation of pro-inflammatory signal transduction pathways in gastric epithelial cells. The most well characterized *H. pylori*-induced pro-inflammatory response is the expression and secretion of the cytokine IL-8 by gastric epithelial cells, in response to co-culture with *cag* PAI-positive strains [54,55]. One mechanism leading to IL-8 induction involves entry of *H. pylori* peptidoglycan into the epithelial cell cytoplasm, where it is recognized by the pathogen recognition molecule NOD1; entry of peptidoglycan into the cytoplasm occurs through a *cag* PAI-dependent process [56]. A second mechanism leading to IL-8 induction involves translocation of CagA. The CagA protein can induce upregulation of IL-8 transcription via activation of the Ras/Raf signaling pathway [57]; however, most reports suggest that IL-8 activation is dependent upon *cag* PAI proteins other than CagA, and IL-8 stimulation occurs in a CagA-independent manner that is strain-specific [57,58,59,60]. Additional CagA-independent effects of the *cag* PAI on the host cell include alteration of epithelial gene transcription (such as MMP-7 [61]), and transactivation of the EGF receptor [62,63].

CagL is a 26 kDa protein component of the *cag* T4SS that lacks homologs in other bacterial species. *H. pylori* mutants lacking *cagL* are defective in the ability to translocate CagA into host cells and do not stimulate production of IL-8 by gastric epithelial cells [55,64]. An important feature of CagL is its ability to bind α5β1 integrin [64]. CagL contains a canonical integrin-binding RGD motif, but there is not uniform agreement about the functional significance of this RGD motif in mediating binding of *H. pylori* to α5β1 integrin [64,65]. A recent study reported that recombinantly expressed CagL functionally mimics fibronectin in its ability to induce focal adhesion formation in mouse fibroblasts and stimulate spreading of AGS cells [66]. Another study reported that CagL mediates dissociation of ADAM17 from α5β1 integrin [67]. This leads to NFκB-mediated repression of the gastric H,K-adenosine triphosphate α-subunit (HKα), which

may ultimately contribute to transient hypochlorhydria in *H. pylori*-infected individuals [67]. Recently it was reported that, in addition to CagL, three other *H. pylori* proteins encoded by the *cag* PAI (CagA, CagI, and CagY) can interact with β1 integrin [65]. Interactions of these three Cag proteins with β1 integrin have been detected using several approaches, including yeast two-hybrid screening, analysis of interactions between these proteins and integrin on the surface of gastric epithelial cells, and surface plasmon resonance analysis [65].

Versatility of bacterial type IV secretion systems

Type IV secretion systems (T4SS) are utilized by a wide variety of bacterial species for delivery of effector proteins or DNA-protein complexes into an assortment of recipient cells, including mammalian cells, plant cells, fungi, or other bacteria [68,69,70,71,72]. T4SSs are distinguished from other bacterial secretion system types as a result of their ability to translocate both proteins and nucleoprotein complexes into the recipient cell. The T4SS family is subdivided into three subfamilies of translocation families, each of which can contribute to bacterial pathogenesis. The largest subfamily consists of DNA conjugation systems, and these T4SSs are present in most species of Gram-negative and Gram-positive bacteria [69]. The function of these conjugation systems is to mediate DNA transfer between bacteria of the same or different species, as well as introduction of bacterial DNA into phylogentically diverse recipients such as fungi, plant, or mammalian cells. These systems contribute to pathogenicity by allowing bacteria to constantly expand or reduce their genomic content, to allow for enhancement of bacterial fitness in a variety of environmental conditions which may include infection of a human host. This family of T4SSs can mediate the widespread dissemination of mobile genetic elements between bacteria, which can lead to rapid emergence of antibiotic resistance among populations of clinically relevant pathogens.

The second subfamily consists of DNA uptake and release systems that function in the absence of host cell contact [69]. This is the most recently discovered subfamily of T4SSs, and is currently comprised of two DNA competency (uptake) systems – the *Campylobacter jejuni* Cjb/VirB system, and the *Helicobacter pylori* ComB system [73,74,75] – as well as the F-plasmid Tra-like DNA-release system encoded by the gonococcal genetic island of *Neisseria gonorrhoeae* [76]. As with DNA conjugation machinery, the DNA uptake and release subfamily of T4SSs likely contributes to genome plasticity via exchange of mobile elements that contain genes which are advantageous for survival.

The third subfamily, perhaps most relevant to bacterial pathogenesis, is the type IV secretion effector/translocator systems. This subfamily has an important role in the infection process of several prominent bacterial pathogens of both plants and mammals [69]. These T4SSs are similar to type III secretion systems in that they deliver their effector molecules through a needle-like appendage, and both systems require direct contact of bacteria with the host cell for delivery of cargo to the eukaryotic cytoplasm. The prototypical effector/translocator T4SS is present in the plant pathogen *Agrobacterium tumefaciens* [69,71]. In addition to *H. pylori*, T4SSs are present in a number of medically-relevant Gram-negative bacteria including *Legionella pneumophlia, Bordetella pertussis, Rickettsia prowazekii, Coxiella burnetii*, and *Brucella suis* [69]. Of note, *Bordetella pertussis* utilizes an effector/translocator T4SS to export pertussis toxin to the extracellular environment, where it is able to bind and enter the host cell through the endosomal pathway [69].

Our current understanding of bacterial T4SSs is based in large part on elegant studies of the *Agrobacterium tumefaciens* T4SS, which translocates bacterial DNA into plant cells [69,71]. The *Agrobacterium* VirB-VirD system is the most extensively characterized T4SS, and can be utilized as a model system for understanding the

10

function of T4SS in other bacterial species. The *Agrobacterium* T4SS is comprised of 11 proteins, termed VirB1-VirB11, encoded by the *virB* operon, and one additional protein encoded by the *virD* operon (VirD4) [69,71]. These Vir proteins have been classified into three functional categories according to cellular localization. VirB4, VirB11, and VirD4 are present in the inner membrane; VirB6, VirB7, VirB8, VirB9, VirB10 are involved in formation of a core complex; and the T4SS pilus is comprised of the major pilin component VirB2, and minor pilin component VirB5 [71].

Organization of the *cag* PAI-encoded type IV secretion system

Thus far there has been relatively little study of the *H. pylori cag* T4SS. Initial identification of bacterial genes required for function of the *H. pylori cag* T4SS was accomplished by systematic transposon mutagenesis of genes within the *cag* PAI [55], or by inserting antibiotic cassettes into selected genes within the *cag* PAI [33,34,77]. One study reported that 17 *cag* genes are required for translocation of CagA into gastric epithelial cells, and 14 *cag* genes influence the secretion of IL-8 (Table 1) [55]. Several of these genes encode products that exhibit low-level sequence similarity to components found in the T4SSs of *A. tumefaciens* and other bacterial species [42,43,46]. At least nine of the *cag* genes that are reported to be required for function of the *cag* T4SS lack homologs in other bacterial species (Table 1) [42,55,78]. The functions of the encoded proteins remain largely uncharacterized. Structural analyses have been undertaken for several components of the *cag* T4SS [79], and several studies have shown that Cag proteins can interact to form subassemblies [78,80,81]. Overall, there is only a very limited understanding of the structural organization and architecture of the *cag* T4SS apparatus.

Few structural studies of components of the *cag* T4SS have been performed. Currently, crystal structures are available for a small number of Cag proteins. Available structures include the *H. pylori* VirB11 homolog [82], an ATPase that organizes into

11

hexameric structures and is required for CagA translocation; CagZ, a protein that interacts with and stabilizes Cag5 (a VirD4 homolog) [83]; CagS, a protein that does not have homologous components in other T4SSs, but is not essential for CagA translocation or induction of IL-8 synthesis and secretion [84]; CagD, a unique component that forms covalent disulfide-linked homodimers [85], and is reported by one study, but not another, to have a role in CagA translocation; and 12 amino acid residues of CagA in co-crystal structure with MARK2 [86], one of approximately 20 host proteins with which CagA interacts.

Multiple studies have employed immunoprecipation or yeast 2-hybrid (Y2H) methods to evaluate Cag protein-protein interactions. Y2H screening of Cag protein interactions has produced variable results. One study utilized Y2H methodology in an attempt to generate a comprehensive protein interaction map of *H. pylori*, but detected very few Cag protein interactions [87,88]. Other studies utilizing Y2H methods discovered multiple interactions among Cag proteins [81]. Perhaps the most well characterized Cag protein-protein interaction is that of the effector molecule CagA with its chaperone-like protein CagF. This CagA-CagF interaction has been described in a number of publications, and has been detected using immunoprecipitation approaches [89]. Recently, Y2H methods were used to demonstrate an interaction between the coupling protein Cag5 (VirD4) and CagA, as well as a stabilizing interaction between CagZ and Cag5 [90].

Immunoprecipitation methods have been used to identify a number of potential Cag-Cag interactions. Cag5, CagT, CagN, CagD, and Cag1 were suggested to co-immunoprecipitate with Cag3 [78]. In addition, low levels of CagM, CagX, CagC, CagE, and the VirB11 homolog were also present in Cag3 immunoaffinity purifications [78]. Y2H analysis also indicates that Cag3 can interact with CagV, CagT, CagM, and CagG [81]. The interaction between Cag3 and CagT has been detected by multiple

independent methods [78,81]. Y2H screening and immunoprecipitation methods were also exploited to identify a Cag subassembly at the outer bacterial membrane consisting of CagM and the Vir homologs CagY, CagT, and CagX [80].

Several studies have reported localization of Cag proteins to specific subcellullar locations. Cag3, a protein unique to *H. pylori*, was found to be surface-exposed on the outer membrane, with a small proportion of the protein distributed to the soluble fraction [78]. CagN, a protein not required for CagA translocation and IL-8 induction, is a membrane-associated protein that is processed at the C-terminus [91]. CagY, a large protein that has low-level relatedness to VirB10, may also be localized to multiple subcellular locations, including the surface of the bacteria [92].

Surface-exposed pili are an important feature of T4SSs [69,93]. The T4SS pili of *A. tumefaciens* are comprised of VirB2 (the major pilin subunit) and VirB5 (the minor pilin subunit) [69,93]. When *H. pylori* is in contact with gastric epithelial cells, the bacteria express pili that are associated with the *cag* T4SS (Figure 1-2 A, B) [64,92,94,95]. Relatively little is known about the composition and biogenesis of these *H. pylori* pili. In contrast to T4SS pili from other bacterial species, the *H. pylori* T4SS pili are reported to be sheathed organelles [92]. One study reported that CagY may be a component of the pilus sheath [92]. It is thought that the presence of numerous repeated sequences in *cagY* allows for frequent intragenic recombination events, producing in-frame rearrangements [96]. These intragenic rearrangements may play a key role in the ability of the bacteria to evade the host immune response by increasing antigenic variation [96].
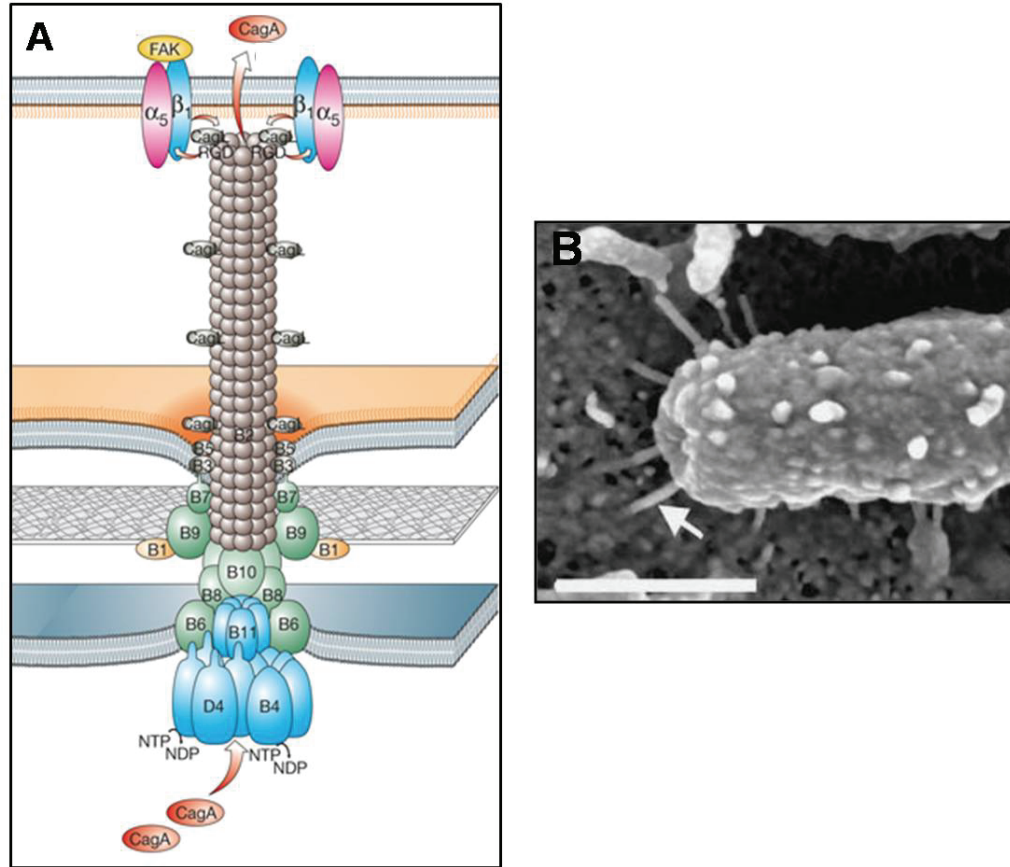
Figure 1-2. Depiction of the *cag* T4SS of *H. pylori*. A. Our current understanding of the *cag* T4SS structure and organization is based largely upon studies of the T4SS of *Agrobacterium*. The *cag* T4SS is a bi-membrane-spanning macromolecular structure which includes a pilus that extends from the bacterium to the surface of the host gastric epithelial cell. The proposed model of the *cag* T4SS is shown, with location and assembly of the Vir homologs (VirB1-VirB11) depicted according to the corresponding localization within the *Agrobacterium* T4SS. The unique *H. pylori* protein CagL contains an RGD motif and is thought to interact with and activate host α5β1 integrin to facilitate translocation of CagA (image reprinted from [64]). **B.** When in intimate contact with gastric epithelial cells, *cag* PAI-positive strains of *H. pylori* generate T4SS pili. These T4SS pili can be imaged by scanning electron microscopy (image reprinted from [92]).

H. pylori CagC is reported to be a VirB2 homolog [94], but there is not yet convincing evidence that CagC is a major component of *H. pylori* pili. Several of the *H. pylori* proteins (CagY, CagT, and CagX) that have been localized to the pili by immunogold staining [92,95] are distantly related to core complex components in other T4SS (VirB10, VirB7, and VirB9, respectively). *H. pylori* CagL has also been localized to pili by immunogold staining and it is suggested that it is a minor pilus component (Figure 1-3) [64]; however, CagL and VirB5 do not exhibit any detectable sequence similarity. In various studies, CagL has been detected as a constituent of the T4SS pilus (Figure 1-3 B) [64], on the surface of *H. pylori* (Figure 1-3 A, C) [66], or as a protein localized to the soluble fraction of *H. pylori* lysates [80,92]. Taken together, these findings suggest that the pili associated with the *H. pylori cag* T4SS are considerably different from the T4SS pili of *A. tumefaciens* and other known bacterial T4SSs.
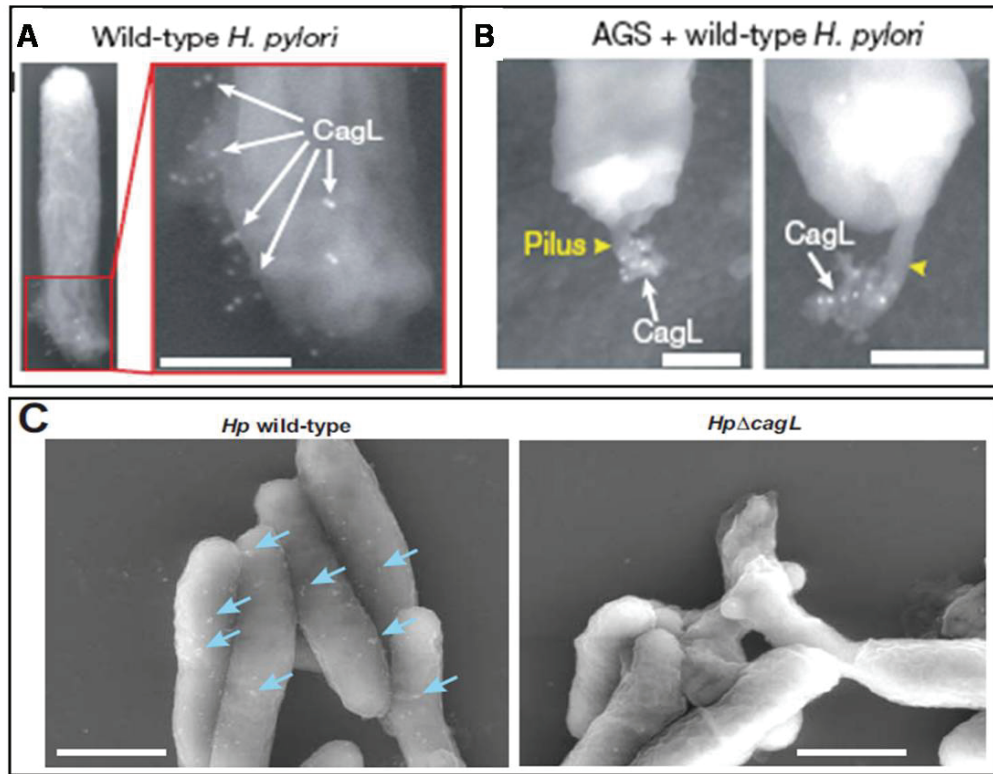
Figure 1-3. Localization of CagL to the *cag* T4SS pilus and surface of *H. pylori*. Immunogold labeling was utilized to localize CagL to the surface of WT *H. pylori* in the absence of host cell contact **(A, C)**. When in contact with gastric epithelial cells, CagL can be localized along the length and tip of T4SS pilus structures **(B)**. CagL cannot be detected by immunogold labeling in a Δ*cagL* isogenic mutant **(C)**. Panels A and B reprinted from [64]; Panel C reprinted from [66].

Despite the efforts of multiple talented groups and more than two decades of research, there is still much to discover about *H. pylori*, and in particular, significant knowledge gaps exist in the area of the architecture and topology of the *cag* T4SS. It is unclear how the *cag* T4SS assembles prior to translocation of CagA into gastric epithelial cells. Furthermore, the molecular mechanism by which CagA is shuttled into the host cell has not been elucidated. The function of many Cag proteins, especially those without homologs in other bacterial T4SSs, remains unexplored.

This thesis primarily focuses on the requirement of non-homologous Cag proteins in mediating translocation of CagA through biogenesis of the T4SS pilus structure at the interface of the bacteria and host cell. In Chapter II, we utilize a robust mass spectrometry method to identify components of a T4SS subassembly that contains the pilus component CagL, establishing a powerful new tool for analyzing Cag protein-protein interactions. We provide evidence that components of this CagL-containing subassembly are required for CagA translocation and induction of IL-8 synthesis and secretion, and that these components share multiple properties including co-transcription and protein sequence relatedness.

In Chapter III, we demonstrate that each of the proteins identified in the CagL T4SS subassembly plays a key role in formation of the T4SS pilus between *H. pylori* and the surface of gastric epithelial cells. We show that CagL and the novel protein CagI are required for formation of the T4SS pili. CagH, a protein that interacts with CagL and CagI, is not required for T4SS formation. However, *H. pylori* that are deficient in *cagH* exhibit a hyperpiliated phenotype, producing T4SS pili that have a dramatically altered morphology compared to those of wild-type (WT) *H. pylori*. We identify a conserved flagellar hook protein K (FlgK) domain near the N-terminus of CagH that is evolutionarily-related to FlgK of *Burkholderia* species. This domain may be contributing to the role of CagH as a modified capping protein or molecular ruler involved in controlling dimensions

17

of the Cag T4SS pilus. Furthermore, we identify nearly identical hexapeptide C-terminal motifs present in CagH, CagI, and CagL that also contribute to T4SS pilus biogenesis. We show that this motif in CagL and CagI is required for pilus formation. In contrast, bacteria expressing CagH that is deficient in this C-terminal motif (CagHΔCT) produce pili with morphology and abundance similar to those of the WT strain. These T4SSs are nonetheless non-functional, as deletion of the motif results in ablation of CagA translocation and induction of IL-8 synthesis and secretion by gastric epithelial cells.

Finally, in Chapter IV, we utilize chemical cross-linking, immunoaffinity purification, and mass spectrometry strategies to identify CagA protein interactions occurring in *H. pylori* in the context of host-cell contact. We identify several CagA interactions occurring in bacteria in contact with the host cell that differ from interactions of CagA in bacteria that have not been co-cultured with gastric epithelial cells. Chemical cross-linking strategies are also employed to analyze interactions of the unique T4SS component CagU. Experiments with epitope-tagged CagU suggest that CagU may homodimerize during T4SS apparatus assembly. Furthermore, we utilize mass spectrometry to analyze sheared preparations derived from WT, hyperpiliated Δ*cagH* bacteria, and non-piliated bacteria that had been in contact with gastric epithelial cells. These sheared preparations are predicted to contain surface components, such as pili, and analysis of these preparations may allow the composition of pili to be determined. Findings described in this thesis provide important insights into the architecture and function of the *cag* T4SS, and establish the foundation for additional studies aimed to elucidate the assembly of the *cag* PAI-encoded T4SS, as well as the molecular mechanism by which the oncoprotein CagA is translocated into the host gastric cell.

CHAPTER II

IDENTIFICATION OF *HELICOBACTER PYLORI* CAG PROTEINS THAT INTERACT WITH THE TYPE IV SECRETION SYSTEM PILUS COMPONENT CAGL

Introduction

The *H. pylori cag* pathogenicity island (*cag* PAI) encodes components of a type IV secretion system (T4SS) that translocates the bacterial oncoprotein CagA and peptidoglycan into gastric epithelial cells [42,43,44,46,97]. Presumably, the assembly and function of the T4SS apparatus is dependent upon multiple Cag protein-protein interactions. Analysis of protein-protein interactions has been a well-established method utilized to understand the assembly of many bacterial T4SSs [81,98,99,100]. Yeast 2-hybrid (Y2H) analyses from multiple groups have identified a number of homotypic and heterotypic interactions among Cag proteins, providing evidence that Cag proteins physically interact to form T4SS subassemblies and complexes [81,88,90,101]. However, Y2H analyses are often associated with a high rate of both false-positive and false-negative results, and it is possible that Cag protein interactions identified in yeast do not occur in *H. pylori* [81]. Classic immunoprecipitation methods have also provided evidence of Cag protein interactions among endogenously-expressed Cag proteins [78,80]; however, this methodology is often limited by the lack of antibodies directed against potential interacting partners. Furthermore, although some components of the Cag T4SS have homology to constituents of the *Agrobacterium* T4SS, substantial differences are likely to exist between the two secretion systems.

In order to analyze Cag protein-protein interactions, we sought to overcome limitations associated with traditional co-immunoprecipitation methods. Moreover, we endeavored to identify Cag protein interactions that occur endogenously within *H. pylori* in the absence of host cell contact, and conditions under which the T4SS has formed in

the presence of host cell contact. We therefore employed an immunoaffinity purification method to isolate Cag protein complexes or subassemblies, and identified all potential interacting proteins by a sophisticated and robust mass spectrometry method known as Multidimensional Protein Identification Technology (MudPIT).

CagL is a specialized component of the *cag* T4SS that binds the host receptor α5β1 integrin. *H. pylori* CagL has also been localized to pili by immunogold staining and it is suggested that it is a minor pilus component [64]; however, CagL and VirB5 do not exhibit any detectable sequence similarity. CagL contains a canonical integrin-binding RGD motif, but there is not uniform agreement about the functional significance of this RGD motif in mediating binding of *H. pylori* to α5β1 integrin. Recently it was reported that, in addition to CagL, three other *H. pylori* proteins encoded by the *cag* PAI (CagA, CagI, and CagY) can interact with β1 integrin [65]. Since CagL is an important component of the *cag* T4SS that interacts directly with epithelial cells, we hypothesized that CagL might physically interact with other T4SS components, and that such interactions might be required for CagL export, localization, stability, or activity. Here, we show that two other proteins encoded by the *cag* PAI (CagH and CagI) co-purify with CagL, and we provide evidence that these three proteins are components of one or more subassemblies associated with the *cag* T4SS. CagL, CagI, and CagH are not homologous to components of T4SSs in other bacterial species, but we demonstrate that all three proteins are essential for activity of the *cag* T4SS.

Methods

Bacterial strains and growth conditions. *H. pylori* 26695 and its *cag* isogenic mutant derivative strains were grown on trypticase soy agar plates supplemented with 5% sheep blood or Brucella agar plates supplemented with 5% fetal bovine serum at 37°C in room air containing 5% CO_2. *H. pylori* mutant strains were selected based on resistance to chloramphenicol (5 µg/ml), kanamycin (10 µg/ml), or metronidazole (7.5 to

15 µg/ml).   *E. coli* strain DH5α, used for plasmid propagation, was grown on Luria-Bertani agar plates or in Luria-Bertani liquid medium supplemented with ampicillin (50 µg/ml), chloramphenicol (25 µg/ml), or kanamycin (10 µg/ml), as appropriate.

Generation of rabbit polyclonal anti-CagL serum.   CagL derived from *H. pylori* 26695 (and lacking a putative signal sequence) was expressed as a GST fusion protein from pGEX-6P-1 vector (GE Healthcare, formerly Amersham). CagL-GST was purified using gluthathione beads [81].   Rabbits were then immunized with purified CagL-GST.

Immunoblot analysis. To detect expression of Cag proteins, individual samples were separated by SDS-PAGE (4-20% gradient), transferred to a nitrocellulose membrane, and subsequently immunoblotted using rabbit polyclonal antiserum raised against recombinant CagL, or mouse monoclonal antibodies reactive with FLAG or HA epitopes (Sigma).   To confirm similar loading of samples, immunoblotting using either a monoclonal mouse antibody (Santa Cruz) or a rabbit polyclonal antiserum to *H. pylori* HspB, a GroEL homolog, was utilized [102].   Horseradish peroxidase-conjugated anti-rabbit IgG or anti-mouse IgG was used as the second antibody. Signals were generated by enhanced chemiluminescence reaction and detection by exposure to X-ray film.

Cell culture methods. AGS human gastric epithelial cells were grown in the presence of 5% CO_2 in RPMI medium containing 10% FBS, 2 mM L-glutamine, and 10 mM HEPES buffer.

IL-8 secretion by gastric cells in contact with H. pylori.   *H. pylori* strains were co-cultured with AGS cells at a multiplicity of infection of 100:1, and IL-8 secretion was analyzed using an anti-human IL-8 sandwich ELISA (R&D). Levels of IL-8 secreted by AGS cells in contact with isogenic *cag* mutants were compared to levels secreted by AGS cells in response to infection with WT *H. pylori* 26695.

CagA translocation assay. Translocation of CagA into AGS cells was analyzed by co-culturing *H. pylori* strains with AGS cells and detecting tyrosine phosphorylation of CagA,

21

as previously described [81,103]. Briefly, *H. pylori* and AGS human gastric cells were co-cultured at a MOI of 100:1 for 24 h at 37°C. Cells were lysed in NP-40 lysis buffer containing Complete Mini EDTA-free Protease Inhibitor (Roche) and 2 mM sodium orthovanadate, and CagA translocation was assessed by separating the soluble fraction using 7.5% SDS-PAGE and immunoblotting with an anti-phosphotyrosine antibody (α-PY99, Santa Cruz).

Synthesis of a cat-rdxA cassette. To facilitate introduction of unmarked mutations into *H. pylori* chromosomal genes of interest, a *cat-rdxA* cassette was synthesized and cloned into pUC57 vector (Genscript). This cassette confers resistance to chloramphenicol mediated by the chloramphenicol acetyl-transferase (*cat*) gene from *Campylobacter coli*, and susceptibility to metronidazole is mediated by an intact *rdxA* gene (HP0954) from *H. pylori* 26695. Both genes are in the same orientation, with *cat* upstream from *rdxA*. Expression of *cat* is driven by the *C. coli cat* promoter, and expression of *rdxA* is driven by the *H. pylori vacA* promoter. The two genes are separated by a presumed *C. coli cat* terminator, and a ΦT7 terminator is located at the 3'-distal end of the *rdxA* gene.

Mutagenesis of H. pylori cag genes. To construct a Δ*cagL* mutant strain, we PCR-amplified *cagL* along with approximately 0.5 kb of flanking DNA from *H. pylori* 26695 genomic DNA using Amplitaq Gold (ABI), and cloned the PCR product into pGEM-T Easy (Promega). By using this plasmid as a template for inverse PCR and then ligating the PCR product, we generated a modified plasmid lacking *cagL* and containing a BamHI site at the site of the deletion. A kanamycin resistance cassette was cloned into the BamHI site to yield pCSLK2-4. *H. pylori* 26695 was transformed with pCSLK2-4 (which is unable to replicate in *H. pylori*), and single colonies resistant to kanamycin were selected. PCR and sequencing were used to confirm that the kanamycin cassette had inserted into the *cagL* locus in the same orientation as operon transcription.

To generate unmarked mutant strains, we used a new method that is a variant of counterselection methods used previously in *H. pylori* (Figure 2-3) [104,105]. As a first step, the *rdxA* gene, which confers resistance to metronidazole, as well as approximately 0.5 kb of flanking DNA on each side, was PCR-amplified from *H. pylori* 26695 and cloned into pGEM-T Easy to yield pMM670. By using pMM670 as the template for inverse PCR and then ligating the PCR product, we generated a modified plasmid (pMM672) in which the coding region of *rdx*A was deleted. Transformation of *H. pylori* 26695 with pMM672 (which is unable to replicate in *H. pylori*) and selection for metronidazole-resistant colonies resulted in the recovery of a mutant in which the *rdx*A locus was deleted (*H. pylori* Δ*rdxA*) (Figure 2-3A).

For each *cag* gene of interest, we PCR-amplified the relevant gene as well as approximately 0.5 kb of flanking DNA on each side from *H. pylori* 26695 genomic DNA, and cloned these sequences into pGEM-T Easy. By using these plasmids as templates for inverse PCR and then ligating the PCR products, we generated modified plasmids lacking the *cag* gene of interest and containing a BamHI site at the site of the deletion. A *cat-rdxA* cassette (described above) was then cloned into the BamHI site. Plasmids containing the *cat-rdxA* cassette (which are unable to replicate in *H. pylori*) were transformed into *H. pylori* Δ*rdxA*, and single colonies resistant to chloramphenicol and susceptible to metronidazole were selected (Figure 2-3A, B). PCR analysis confirmed that the *cat-rdxA* cassette had inserted into the desired chromosomal site and that the relevant *cag* genes were deleted. To generate unmarked mutants, strains containing the *cat-rdxA* cassette were transformed with plasmids harboring the desired mutations, and transformants resistant to metronidazole were selected (Figure 2-3A, B). A similar approach was used to introduce a gene encoding a C-terminally truncated form of CagL into the endogenous *cagL* locus.

To complement mutants in *cis* at a heterologous chromosomal locus, we used a plasmid derived from pAD1 [106], which allows genes of interest to be introduced into the *H. pylori* chromosomal *ureA* locus. The modified pAD1 plasmid contains a chloramphenicol resistance cassette, restriction sites to allow cloning of a gene of interest into a site downstream from the *ureA* promoter and a ribosomal binding site, and flanking sequences derived from the *ureA* locus. Plasmids were constructed to allow expression of epitope-tagged forms of CagH, CagI, and CagL, and in addition, plasmids were constructed to allow expression of untagged forms of CagI and CagL. CagL was expressed as a protein containing a hemagglutinin (HA) tag introduced at residue 22 (immediately after a putative signal sequence) (Figure 2-1D). CagH was expressed as a protein containing an N-terminal HA tag, and CagI was expressed with an internal FLAG epitope introduced at residue 51 in a region downstream of the predicted signal sequence (Figure 2-3C), which is predicted to be surface exposed based on hydrophilicity plots (ProtScale in Expasy) [107]. *H. pylori* strains were transformed with these plasmids, and chloramphenicol-resistant colonies were selected. Expression of the epitope-tagged proteins was verified by immunoblotting using monoclonal anti-HA or anti-FLAG antibodies, respectively.

Immunoaffinity purification of Cag proteins. WT or Δ*cagL* isogenic mutant bacteria cultured on solid media were harvested after 48 h of growth, washed once in phosphate-buffered saline (PBS), and lysed overnight at 4°C in RIPA buffer (10 mM Tris, 100 mM NaCl, 1% NP-40, 0.25% deoxycholic acid, Complete Mini EDTA-free Protease Inhibitor (Roche), pH 7.2). Cellular debris was removed by high speed centrifugation. CagL was immunoaffinity purified from bacterial lysate using polyclonal anti-CagL antibodies that had been covalently cross-linked to a Protein A support (Dynal, Invitrogen). Alternatively, CagL-HA and CagH-HA were purified from strains expressing these proteins using monoclonal anti-HA antibodies that had been either covalently cross-

linked or non-covalently bound to a Protein G support (Dynal, Invitrogen). Immobilized anti-FLAG M2 antibody (Sigma) was utilized for immunoaffinity purification of CagI-FLAG. Target proteins were immunoaffinity purified at room temperature. Immunoaffinity purified proteins were washed in 100 bed volumes of PBS containing 0.1% Tween-20 (PBST 0.1%) prior to elution by boiling in SDS buffer (0.3 M Tris-HCl, 1% SDS, 10% glycerol, 100 mM DTT, Pierce), or elution by HA or FLAG peptide competition (20 µg/ml in PBS for HA peptide [Sigma], final concentration; 20 µg/ml in TBS for FLAG peptide [Sigma], final concentration). In each case, the presence of the targeted protein in the immunoaffinity purified sample was verified by immunoblotting.

To purify Cag proteins from bacteria attached to gastric epithelial cells, the WT strain and a strain expressing CagH-HA were each co-cultured with AGS cells (80% confluent) at an MOI of 100 for 5 hours prior to the addition of RIPA buffer and mechanical disruption of the cell monolayers. Cells were lysed by 3 pulses of sonication (10 s for each pulse), followed by overnight incubation at 4°C in RIPA buffer. Cleared lysates were incubated with immobilized anti-HA monoclonal antibody (Sigma) at room temperature. Immunoaffinity-purified preparations were washed with 100 bed volumes of PBST 0.1%. Bound proteins were eluted by HA peptide competition.

Mass spectrometric analysis of samples. To provide a comprehensive analysis of the protein content in immunoaffinity purified samples, the samples were analyzed by multidimensional protein identification technology (MudPIT). Purified proteins were eluted from the beads, run about 2 cm into a 10% NuPAGE gel, and then subjected to in-gel digestion with trypsin and peptide extraction. The resulting peptide mixtures were analyzed via MudPIT essentially as described [108]. Briefly, peptides were loaded via pressure cell (New Objective) onto a biphasic pre-column fritted using an Upchurch M-520 filter union (IDEX). This 100-µm fused silica microcapillary column was packed with 3 cm of 4-µm C18 reverse-phase resin (Jupiter, Phenomenex) followed by 5 cm of

strong cation-exchange resin (Luna SCX, Phenomenex). Once loaded, it was then placed in-line with a 100 μm × 20 cm, C18 packed emitter tip column (Jupiter C18, 3 μm, 300 Å, Phenomonex) coupled to an LTQ ion trap mass spectrometer equipped with an Eksigent NanoLC-AS1 Autosampler 2.08, an Eksigent NanoLC-1D plus HPLC pump, and nanospray source. Multidimensional separations were accomplished using 5 μl autosampled pulses of ammonium acetate in 0.1% formic acid (25, 50*, 75, 100, 150*, 200, 250*, 300, 500*, 750*, 1000* mM pulses - * shorter MudPITs containing these salt pulses were performed on isolated proteins) followed by a 105 min reversed phase gradient from water 0.1% formic acid to 45% ACN 0.1% formic acid. Tandem mass spectra were collected throughout the runs in a data dependent manner using dynamic exclusion to improve data acquisition of lower intensity peptides. These spectra were extracted from the instrument files using ScanSifter and searched using SEQUEST [109] against an *H. pylori* strain 26695 database that also contained common contaminants and reversed versions of the *H. pylori* proteins. Identifications were filtered to an estimated desired false discovery rate (FDR) using reverse database hits and collated to proteins using IDPicker [110]. All reported proteins were identified with a minimum of 2 distinct peptides.

Real-time PCR. Total RNA was isolated from *H. pylori* using Trizol Reagent (Gibco), according to the manufacturer's protocol. RNA samples were refined using the RNeasy Mini Kit (Qiagen), and on-column RNase-free DNase digestion. cDNA synthesis was performed on 100 ng of purified RNA using the iScript cDNA synthesis kit (BioRad). As a control, first strand cDNA reactions were carried out in parallel without reverse transcriptase. Real time PCR was executed in triplicate on an ABI StepOne Real Time PCR machine, using TaqMan MGB chemistry. Abundance of *cagH*, *cagI*, and *cagL* transcripts in *cag* isogenic mutant strains was calculated using the ΔΔCT method, with

each transcript signal normalized to the abundance of the *recA* internal control and comparison to the normalized transcript levels of WT *H. pylori*.

Bacterial subcellular fractionation. Wild-type *H. pylori* strain 26695, as well as strains expressing CagL-HA, CagH-HA or CagI-FLAG, were grown on solid media for 24 h prior to harvest, washing, and resuspension in sonication buffer (10 mM Tris-HCL, pH 8.0, Complete Mini EDTA-Free Protease Inhibitor (Roche)). Bacteria were lysed by 5 pulses of sonication. Unbroken bacteria and cellular debris were removed from the lysate by centrifugation at 4500 x g for 10 min. The supernatant was separated by ultracentrifugation (1h at 250,000 x g, 4°C) into a soluble fraction and a total membrane fraction. Proteins in the soluble fraction were concentrated by methanol-chloroform precipitation [111] prior to solubilization in reducing SDS-PAGE buffer (Pierce). The total membrane fraction was washed three times in sonication buffer, and resuspended in reducing SDS-PAGE buffer (Pierce). Subcellular fractions were immunoblotted with anti-CagL, anti-HA and anti-FLAG antibodies, followed by appropriate secondary antibodies, to detect untagged CagL, CagL-HA, CagH-HA, and CagI-FLAG, respectively.

Cleavage of surface proteins by proteinase K. Susceptibility of surface-exposed *H. pylori* proteins to digestion with proteinase K was assessed using a modification of previous protocols [102]. *H. pylori* expressing either CagH-HA, CagI-FLAG, or CagL-HA were grown for 24 h on blood agar plates prior to harvesting and washing in PBS. Bacteria were resuspended in RMPI medium, or RPMI medium containing proteinase K (0-20 µg/ml). Bacteria were incubated for 30 min on ice, and proteinase K activity then was abrogated by addition of PMSF (final concentration 2 mM). After washing in RPMI containing 2 mM PMSF, the bacteria were resuspended in SDS sample buffer and analyzed by SDS-PAGE and immunoblotting. CagH-HA, CagI-FLAG, and CagL-HA were detected with monoclonal anti-HA or monoclonal anti-FLAG antibodies. Cleavage of VacA, which is known to be present on the *H. pylori* surface [112,113], was assessed

27

using an anti-VacA antibody [114]. Cleavage of carbonic anhydrase was assessed using an antiserum raised against a periplasmic domain of this protein [115]. Densitometry was accomplished using Image J software [116]; band densities were normalized to HspB and compared to control samples that were not treated with proteinase K.

Flow cytometry. Bacteria were grown overnight in Brucella broth supplemented with 5% fetal bovine serum. Bacterial cells were fixed in 4% paraformaldehyde before staining with primary antibody (monoclonal anti-HA [Sigma] or monoclonal anti-FLAG [Sigma]) in PBS supplemented with 50 mM EDTA and 0.1% BSA, followed by staining with secondary antibody (goat F(ab')$_2$ anti-mouse IgG labeled with Alexa Fluor 488) in PBS supplemented with 0.1% BSA. Flow cytometry data were collected on a Becton Dickenson LSR2 instrument using Becton Dickenson FACS Diva 6.1.3 software. 20,000 single cell events were collected for each sample and analyzed using Tree Star Flow Jo 8.8.2 software.

Immunoelectron microscopy. *H. pylori* cells were grown on blood agar plates for 48 h as described above. Immunogold labeling was performed on whole bacterial cells as previously described with some modifications. Briefly, bacterial cells were lifted onto formvar-coated 200 mesh copper TEM grids (Electron Microscopy Sciences). Cells were washed three times with pre-warmed PBS and incubated in 1% gelatin in PBS for 30 min at 37°C. After blocking, mouse monoclonal primary antibodies (anti-HA or anti-FLAG) were applied to the sample grids and incubated for 30 min at 37°C. Subsequently, grids were washed three times with pre-warmed PBS and a second blocking step was performed using 1% gelatin in PBS for 30 min at 37°C. Then goat anti-mouse secondary antibodies conjugated to 10 nm colloidal gold particles (Ted Pella, Inc.) were applied and incubated for 30 min at 37°C. Afterward, grids were washed three times with prewarmed PBS and cells were negatively stained with 1% ammonium

molybdate before being visualized with a Philips CM-12 transmission electron microscope.

Statistical Analysis.   The statistical significance of differences in numbers of spectral counts of proteins detected in different immunoaffinity purified preparations was determined by G-test, as previously described [117]. The statistical significance of differences in IL-8 production or *cag* gene expression when comparing WT *H. pylori* with mutant strains was determined by one-way ANOVA followed by Dunnett's post-hoc test for multiple comparisons against a single control [118].   Statistical analysis of pilus dimensions was performed using a two-tailed Student's T-Test.

Ethics Statement.   This study was carried out in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocol was approved by the Institutional Animal Care and Use Committee of Vanderbilt University School of Medicine (M/07/292).

Results

Identification of Cag proteins that co-purify with CagL. Previous studies have shown that CagL is an important component of the *H. pylori cag* T4SS that can bind α5β1 integrin [55,64,65]. We hypothesized that CagL interacts with other components of the *cag* T4SS to form one or more subassemblies. Therefore, we sought to identify *H. pylori* proteins that co-purify with CagL. To facilitate these studies, a *cagL*-deficient *H. pylori* mutant strain (Δ*cagL*) was generated as described in Methods. As expected, CagL expression was detected by immunoblot analysis in the wild-type (WT) strain but not in the Δ*cagL* mutant strain (Figure 2-1A). Consistent with previous reports [55,64,65], the Δ*cagL* mutant strain was defective in its ability to stimulate IL-8 secretion by gastric epithelial cells (Figure 2-1B) and defective in its ability to translocate CagA into gastric epithelial cells (Figure 2-1C). Complementation with an epitope-tagged form of CagL restored the ability of the bacteria to translocate CagA into host cells and induce IL-8 secretion (Figure 2-1A-D).
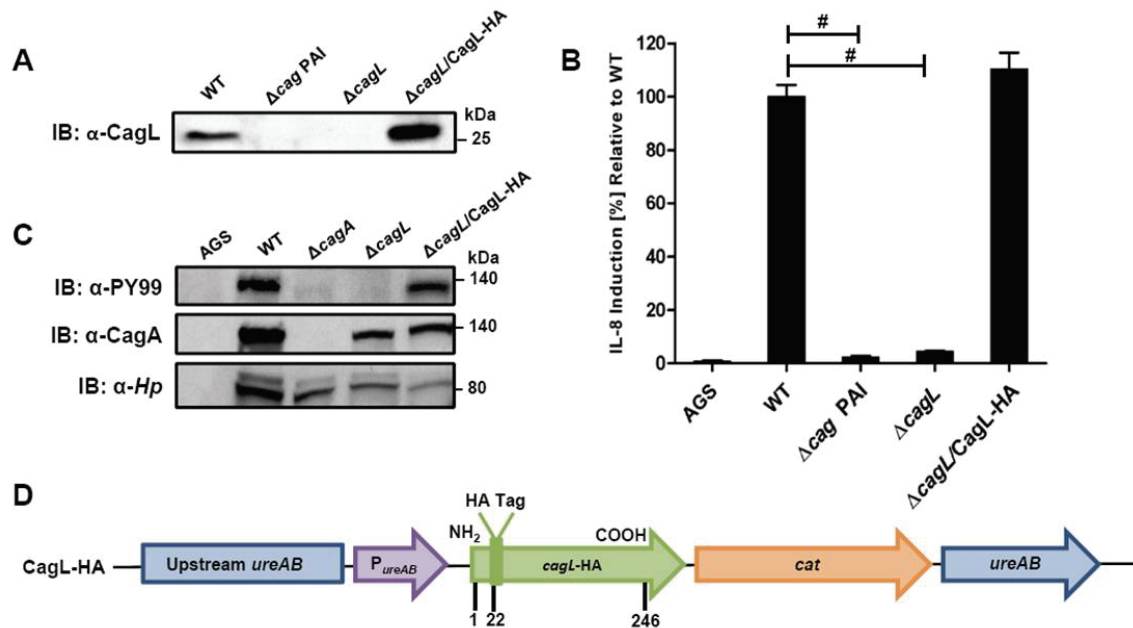
Figure 2-1. Functional analysis of CagL. (A) Immunoblot (IB) analysis indicates that CagL is expressed in wild-type *H. pylori* 26695 (WT), but not in Δ*cag* PAI or Δ*cagL* mutants. Complementation of Δ*cagL* with a gene encoding an epitope-tagged form of CagL (CagL-HA) results in restoration of CagL expression. **(B)** AGS cells were infected with WT *H. pylori* or the indicated mutants, and levels of IL-8 secreted by AGS cells were quantified by ELISA. A wild-type strain was able to induce IL-8 secretion, whereas mutants deficient in *cagL* were unable to induce IL-8 secretion. Complementation of CagL rescued the ability of Δ*cagL* to stimulate IL-8 secretion. Values from six replicate samples were compared to the wild-type control by ANOVA followed by Dunnett's *post hoc* test; # indicates $p < 0.05$. **(C)** AGS cells were co-cultured with WT *H. pylori* or the indicated mutants, and tyrosine-phosphorylated CagA was detected using an anti-phosphotyrosine antibody (α-PY99). Samples also were immunoblotted with an anti-CagA antibody and antiserum reactive with multiple *H. pylori* proteins (α-*Hp*). The Δ*cagL* mutant was defective in ability to translocate CagA into AGS cells. **(D)** Schematic of a construct encoding an epitope-tagged form of CagL (CagL-HA), used to complement Δ*cagL* in *cis* from the *ureA* locus. P*ureAB*, *ure*AB promoter; *cat*, chloramphenicol resistance mediated by chloramphenicol acetyltransferase gene of *C. coli*.

We used a polyclonal anti-CagL serum to immunoaffinity-purify CagL from a wild-type *H. pylori* strain, and a Δ*cagL* mutant strain was processed in parallel. As expected, immunoblotting studies showed that CagL was immunoaffinity-purified from WT lysate but not Δ*cagL* lysate (data not shown). The protein content of each sample was analyzed using multidimensional protein identification technology (MudPIT). Numerous proteins were detected in each sample, but we focused in particular on proteins encoded by the *cag* PAI, since many of these are known or predicted to comprise components of the *cag* T4SS. CagL was identified by MudPIT in the preparation derived from WT lysate, but was not detected in the preparation derived from Δ*cagL* lysate (Table 2-1). In addition, the relative abundance of CagH and CagI peptides detected in affinity-purified preparations derived from the WT strain was significantly higher than the number of CagH and CagI peptides detected in preparations derived from the Δ*cagL* mutant strain (Table 2-1).

Table 2-1. Cag proteins that co-purify with CagL

		anti-CagL [b]			
Gene Number [a]	**Protein**	**WT**	**Δ*cagL***	**WT**	**Δ*cagL***
HP0539	CagL	19 [**]	0	21 [***]	0
HP0540	CagI	33 [***]	0	29 [***]	0
HP0541	CagH	17 [**]	0	19 [***]	0
HP0547	CagA	33	13	32	27
HP0522	Cag3	2	0	0	0
HP0527	CagY	4	1	3	3
Total Spectral Counts		3009	1054	2829	2609

[a] Based on the *H. pylori* 26695 genome annotation
[b] CagL was affinity purified from the WT strain (cultured for 48 h) using anti-CagL polyclonal antiserum, and a Δ*cagL* mutant was processed in parallel as a control. The Table shows numbers of spectral counts observed by MudPIT analysis for each identified Cag protein. The Table shows results from two independent experiments.
** p<0.01; *** p<0.001 when comparing WT with the Δ*cagL* mutant, according to the G-test likelihood ratio, post-spectral count normalization.

Peptides identified for CagH, CagI, and CagL covered 33% to 51% of the respective proteins (Figure 2-2A). CagH and CagI co-purified with CagL from the WT strain in six independent experiments (data not shown). CagA and several other Cag proteins were also detected in the affinity purified preparations, but the numbers of spectral counts corresponding to these proteins were not significantly different when comparing preparations derived from WT and ΔcagL mutant strains (Table 2-1). Many other Cag proteins were detected by mass spectrometric analysis of crude bacterial lysate (Table 2-2) but were not detected in the affinity purified preparations. These initial experimental results, demonstrating co-purification of CagL, CagH, and CagI, suggested that these three proteins are components of one or more subassemblies associated with the T4SS.

A

CagH, 39.4 kDa; 33.0% Coverage

MAGTQAIYESSSAGFLSEISSIISSTSGVAGPFAGIVAGAMSAAIIPIVVGFTNPQMTAIMTQYN
QSIAEAVSMPMKAANQQYNQLYQGFNDQSMAVGNNILNISKLTGEFNVQGNTQGAQISAVNS
QIASILASNTTPKNPSAIEAYATNQIAVPSVPTTVEMMSGILGNITSAAPKYALALQEQLRSQAS
NSSMNDTADSLDSCTALGALVGSSKVFFSCMQISMTPMSVSMPTVYAKYQALXTNALTSGTN
PMTTPACPIGDKVLAVYCYAEKVAEILREYYIEFVKNNTNLLQNASQMILNQSGLATSTYDTQA
ISNISSLYNYNIVANKSFLKSHLTYLDYIKNKLKGQKDSYLTERVQTKIIVK

CagI, 41.6 kDa; 43.8% Coverage

MKCFLSIFSFLTFCGLSLNGTEVVITLEPALKAIQADAQAKQKTAQAELKAIEAQSSAKEKAIQ
AQIEGELRTQLATMSAMLKGANGVINGVNGMTGGFFAGSDILLGVMEGYSSALSALGGNVK
MIVEKQKINTQTEIQNMQIALQKNNEIIKLKMNQQNALLEALKNSFEPSVTLKTQMEMLSQAL
GSSSDNAQYIAYNTIGIKAFEETLKGFETWLKVAMQKATLIDYNSLTGQALFQSAIYAPALSFF
SSMGAPFGIIETFTLAPTKCPYLDGLKISACLMEQVIQNYRMIVALIQNKLSDADFQNIAYLNGI
NGEIKTLKGSVDLNALIEVAILNAENHLNYIENLEKKADLWEEQLKLERETTARNIASSKVIVK

CagL, 26.8 kDa; 51.1% Coverage

MKTLVKNTISSFLLLSVLMAEDITSGLKQLDSTYQETNQQVLKNLDEIFSTTSPSANNEMGEED
ALNIKKAAIALRGDLALLKANFEANELFFISEDVIFKTYMSSPELLLTYMKINPLDQNTAEQQC
GISDKVLVLYCEGKLKIEQEKQNIRERLETSLKAYQSNIGGTASLITASQTLVESLKNKNFIKGI
RKLMLAHNKVFLNYLEELDALERSLEQSKRQYLQERQSSKIIVK

B

53	SPSANNEMGEE-DALNIKKAAIALRGDLALLKANFEANELFFISEDVIFKTYMSSPELLL	111
	S ++N+ M + D+L+ A ALG K F ++ V T + + L	
189	SQASNSSMNDTADSLDSCTALGALVGSS---KVFFSCMQISMTPMSVSMPTVYAKYQALX	245

112	TYMKINPLDQNTAEQQCGISDKVLVLYCEGKLKIEQEKQNIRERLETSLKAYQSNIGGTA	171
	T + + T C I DKVL +YC E+ + +RE +K + + +	
246	TNALTSGTNPMTT?A-CPIGDKVLAVYCYA----EKVAEILREYYIEFVKNNTNLLQNAS	300

172	SLITASQTLVESLKNKNFIKGIRKLM---LAHNKVFL-NYLEELDALERSLEQSKRQYLQ	227
	+I L S + I I L + NK FL ++L LD ++ L+ K YL	
301	QMILNQSGLATSTYDTQAISNISSLYNYNIVANKSFLKSHLTYLDYIKNKLKGQKDSYLT	360

228	ERQSSKIIVK	237	**CagL**
	ER +KIIVK		
361	ERVQTKIIVK	370	**CagH**

C

Figure 2-2. Related properties of CagH, CagI, and CagL. (A) CagH and CagI were each co-immunoaffinity purified with the target protein CagL from WT lysate. CagH, CagI, and CagL peptides identified by MudPIT (pooled data from all CagL purifications shown in Table 2-1 and Table 2-3) are indicated. N-terminal signal peptides (underlined) are predicted to be present in CagI and CagL, based on SignalP analysis [119,120]. **(B)** An alignment of CagH and CagL protein sequences reveals sequence similarity between the C-terminal portion of CagH and the corresponding portion of CagL (32% amino acid identity, 47% similarity in the region shown). **(C)** Operon structure of the *cag* PAI region spanning from *cagA* to *cagM*. Black arrows indicate transcriptional start points. *cagH, cagI,* and *cagL* are contiguous genes within an operon spanning from *cagC* to *cagL* and a suboperon spanning from *cagF* to *cagL*.

Table 2-2. Cag proteins identified in *H. pylori* whole cell lysate

Gene Number [a]	Protein	Homology [b]	SDS [c]	RIPA [c]
HP0520	Cag1		12	10
HP0522	Cag3		56	3
HP0524	Cag5	VirD4	38	14
HP0525	Cagα	VirB11	16	4
HP0526	CagZ		4	7
HP0527	CagY	VirB10	43	5
HP0528	CagX	VirB9	21	1
HP0529	CagW		8	10
HP0530	CagV	VirB8	29	4
HP0531	CagU		2	5
HP0532	CagT	VirB7	26	1
HP0534	CagS		0	2
HP0536	CagP		2	0
HP0537	CagM		20	6
HP0538	CagN		1	1
HP0540	CagI		0	4
HP0541	CagH		0	2
HP0543	CagF		7	56
HP0544	CagE	VirB4	67	15
HP0545	CagD		6	12
HP0546	CagC	VirB2	7	0
HP0547	CagA		268	114
Total Spectral Counts			27688	33569

[a] Based on the *H. pylori* 26695 genome annotation
[b] Based on comparison to the *Agrobacterium tumefaciens* T4SS
[c] *H. pylori* 26695 was cultured for 48 h and then solubilized in either 1% SDS or RIPA buffer. The Table shows numbers of raw spectral counts observed by MudPIT analysis for each identified Cag protein.

As another approach for identifying proteins that co-purify with CagL, we undertook experiments with a complemented Δ*cagL* mutant strain expressing CagL with an N-terminal hemagglutinin (HA) epitope tag (Figure 2-1D). We immunoaffinity-purified CagL-HA from lysate of the strain expressing CagL-HA, using an immobilized monoclonal anti-HA antibody; the WT strain (expressing CagL without an HA tag) was processed in parallel in the same manner. CagL-HA and co-purifying proteins were eluted from the anti-HA affinity column by HA peptide competition, and MudPIT was utilized to analyze the total protein content of the samples. Mass spectrometry identified only three Cag proteins in the sample purified from the strain expressing CagL-HA: the target protein CagL-HA, CagI, and CagH (Table 2-3). The relative abundance of CagL and CagI peptides detected in the preparation from the CagL-HA-expressing strain was significantly higher than the number of CagL and CagI peptides detected in the preparation from the WT strain. Of note, compared to affinity-purified preparations of CagL obtained using polyclonal antiserum, the preparation of CagL-HA obtained using a monoclonal antibody and peptide elution had a much higher level of purity (Table 2-3 and Table 2-2).

Table 2-3. Cag proteins that co-purify with epitope-tagged CagH, CagI, and CagL

Gene Number [a]	Protein	anti-HA [b]			anti-FLAG [c]	
		WT	CagH-HA	CagL-HA	WT	CagI-FLAG
HP0540	CagI	0	203 ***	10 ***	0	170 ***
HP0541	CagH	1	91 ***	2	0	19 ***
HP0539	CagL	0	98 ***	21 ***	0	19 ***
HP0547	CagA	0	2	0	47	87
HP0544	CagE	0	0	0	6 *	1
HP0535	CagQ	1	4	0	0	0
HP0530	CagV	0	0	0	4	4
HP0543	CagF	0	0	0	2	2
HP0528	CagX	0	0	0	2	6
HP0527	CagY	0	0	0	5	4
HP0520	Cag1	0	0	0	3	2
Total Spectral Counts		208	1428	237	6310	10942

[a] Based on the *H. pylori* 26695 genome annotation

[b] CagH-HA or CagL-HA were affinity purified from strains expressing these proteins using an anti-HA antibody, and a WT strain was processed in parallel as a control. The Table shows numbers of spectral counts observed by MudPIT analysis for each identified Cag protein.

[c] CagI-FLAG was affinity purified from a strain expressing this protein using an anti-FLAG antibody, and a WT strain was processed in parallel as a control. The Table shows numbers of spectral counts observed by MudPIT analysis for each identified Cag protein.

* p<0.05; *** p<0.001 when compared to WT control, according to the G-test likelihood ratio, post-spectral count normalization.

To further investigate potential interactions among CagH, CagI, and CagL, we attempted to generate polyclonal antisera against CagH and CagI, but this failed to yield antisera that recognized the corresponding *H. pylori* proteins. Therefore, we generated unmarked non-polar Δ*cagH* and Δ*cagI* mutants (Figure 2-3A, B), as well as complemented mutant strains expressing epitope-tagged forms of CagH or CagI in a heterologous chromosomal locus (Figure 2-3C). CagH-HA and CagI-FLAG proteins of the expected masses were detected in the complemented bacteria (but not WT bacteria) by immunoblotting using monoclonal antibodies against the epitope tags (Figure 2-4A). We immunoaffinity-purified CagH-HA from lysate of the strain expressing this protein, and lysate from the WT strain (expressing CagH without an HA tag) was processed in parallel. The relative abundance of CagH, CagI and CagL detected in the preparation from the CagH-HA-expressing strain was significantly higher than the number of CagH, CagI and CagL peptides detected in the preparation from the WT strain (Table 2-3 and Table 2-2).
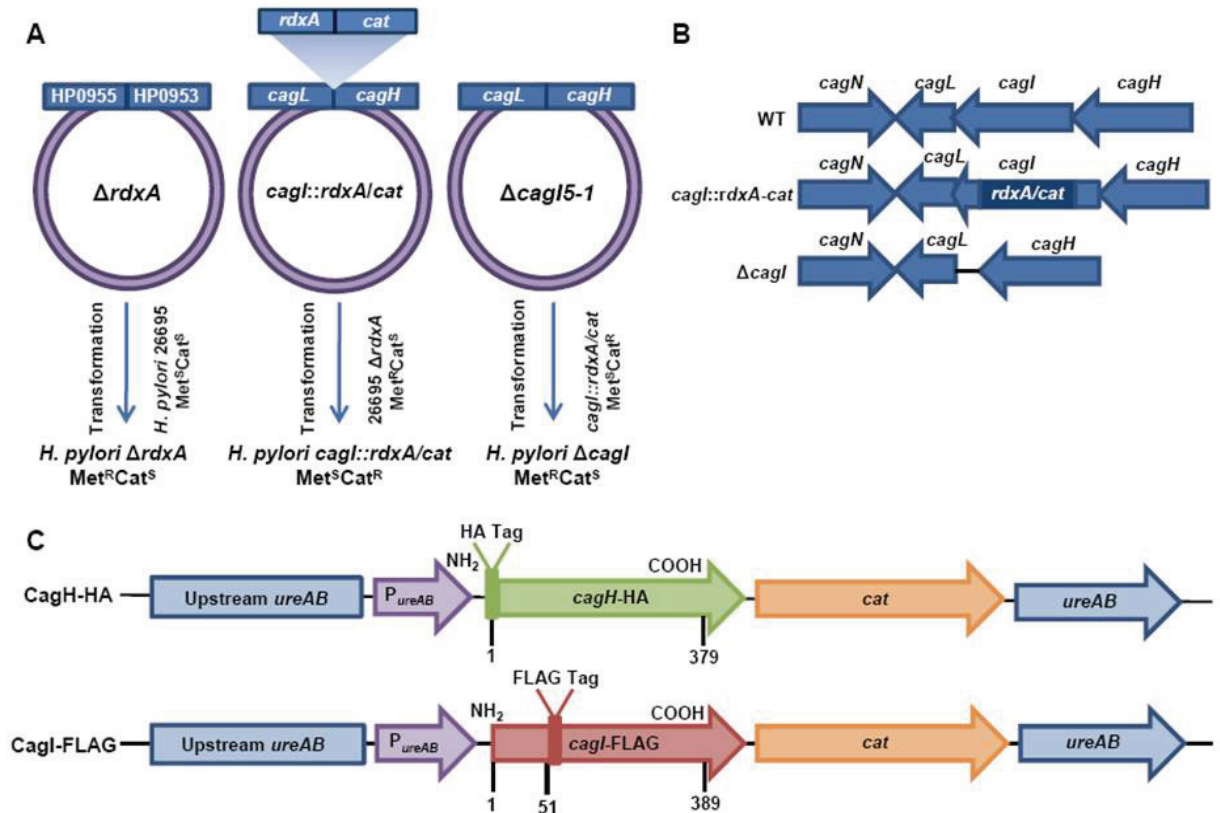
Figure 2-3. Generation of unmarked, in-frame isogenic mutants. **(A)** A metronidazole-resistant strain of *H. pylori* was generated by transforming the WT strain with pMM672 (p-Δ*rdxA*); metronidazole-resistant transformants contained a deletion of the *rdxA* locus. The *cagI* chromosomal locus was deleted by transforming the metronidazole-resistant *H. pylori* with p-*cagI*::*rdxA*/*cat;* this resulted in chloramphenicol-resistant, metronidazole-sensitive colonies, designated *cagI*::*rdxA*/*cat*. To generate a mutant containing an unmarked *cagI* deletion, *H. pylori cag*I::*rdx*A/*cat* was transformed with pΔ*cagI*4-1, and metronidazole-resistant colonies were selected. **(B)** Schematic illustration of the chromosomal *cagH-cagI-cagL* region in the WT strain, *cagI::rdxA-cat* mutant, and Δ*cagI* unmarked mutant. A similar approach was used to generate an unmarked Δ*cagH* mutant strain. **(C)** Schematic of plasmids encoding epitope-tagged forms of CagH and CagI. P$_{ureAB}$, *ure*AB promoter; *cat*, chloramphenicol resistance mediated by chloramphenicol acetyltransferase gene of *C. coli*. The hemagglutinin (HA) epitope sequence was inserted into CagH at the N-terminus (CagH-HA), and the FLAG epitope was inserted at residue 51 of CagI (CagI-FLAG). Numbers indicate position of residues in the epitope tagged proteins. These plasmids were introduced into Δ*cagH* and Δ*cagI* mutants, respectively, and the resulting complemented strains contained sequences encoding CagH-HA or CagI-FLAG, inserted into the *ureA* locus.

40
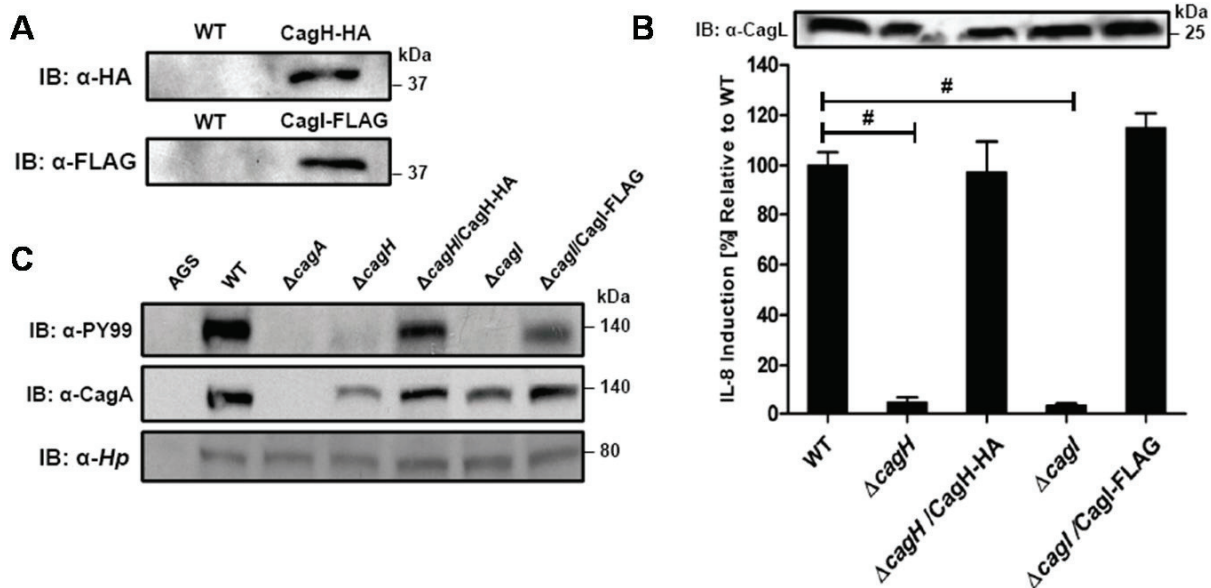
Figure 2-4. Functional analysis of CagH and CagI. **(A)** Immunoblot analysis of CagH-HA and CagI-FLAG in the complemented Δ*cagH* and Δ*cagI* mutant strains, using monoclonal antibodies directed against the epitope tags. **(B)** WT *H. pylori* or the indicated mutants were co-cultured with AGS cells for 4.5 h, and IL-8 expression was analyzed by ELISA. Δ*cagH* and Δ*cagI* mutants were defective in ability to induce IL-8 secretion, and complementation corrected the defect. Values from six replicate samples were compared to the wild-type control by ANOVA followed by Dunnett's *post hoc* correction; # indicates $p < 0.05$. CagL remained detectable in all strains throughout the assay, as demonstrated by anti-CagL immunoblot. **(C)** AGS cells were co-cultured with WT *H. pylori* or the indicated mutants, and tyrosine-phosphorylated CagA was detected with an anti-phosphotyrosine antibody (α-PY99). Samples also were immunoblotted with an anti-CagA antibody and antiserum reactive with multiple *H. pylori* proteins (α-*Hp*). CagA translocation was impaired in the Δ*cagH* and Δ*cagI* mutants.

In a similar manner, we purified CagI-FLAG from the strain expressing this epitope-tagged protein. The relative abundance of CagI, CagH, and CagL detected in the preparation from the CagI-FLAG-expressing strain was significantly higher than the number of corresponding peptides detected in the preparation from the WT strain (Table 2-3). These experiments recapitulated the co-purification of CagH, CagI, and CagL that was observed in initial purifications utilizing polyclonal anti-CagL antisera (Table 2-1). In individual experiments, we sometimes detected levels of other proteins besides CagH, CagI, and CagL that were significantly increased in affinity-purified preparations compared to control preparations (Table 2-4), but these results were not consistently reproduced and were not recapitulated when alternate members of a putative CagH-CagI-CagL complex were targeted for purification.

Table 2-4. Proteins detected in immunoaffinity-purified preparations of CagH-HA and CagL-HA

Gene Number [a]	Protein	WT [b]	CagH-HA [b]	CagL-HA [b]
HP0539	CagL	0	98 ***	21 ***
HP0540	CagI	0	203 ***	10 ***
HP0541	CagH	1	91 ***	2
HP0547	CagA	0	2	0
HP0535	CagQ	1	4	0
HP0010	chaperone and heat shock protein	16	72	19
HP0011	co-chaperone	0	4	0
HP0025	outer membrane protein 2	1	9	3
HP0027	isocitrate dehydrogenase	1	4	1
HP0061	hypothetical protein	1	0	1
HP0072	urease beta subunit	7	0	0
HP0073	urease, alpha subunit	6	0	0
HP0082	methyl-accepting chemotaxis transducer	0	12	0
HP0083	ribosomal protein S9	1	1	0
HP0086	conserved hypothetical protein	0	3	0
HP0099	methyl-accepting chemotaxis protein	1	22	0
HP0103	methyl-accepting chemotaxis protein	1	1	0
HP0109	chaperone and heat shock protein 70	2	11	1
HP0111	hypothetical protein	1	1	0
HP0116	DNA topoisomerase I	0	7	0
HP0153	recombinase	0	8	0
HP0175	cell binding factor 2	2	2	1
HP0176	fructose-bisphosphate aldolase	0	2	0
HP0197	S-adenosylmethionine synthetase 2	0	5	0
HP0210	chaperone and heat shock protein	0	4	0
HP0224	peptide methionine sulfoxide reductase	2	5	0
HP0231	hypothetical protein	2	0	1
HP0243	neutrophil activating protein	12	0	3
HP0264	ATP-dependent protease binding subunit	0	8	0
HP0296	ribosomal protein L21	1	7	0
HP0297	ribosomal protein L27	1	1	0
HP0305	hypothetical protein	2	2	0
HP0310	conserved hypothetical protein	0	2	0
HP0317	outer membrane protein 9	1	25	1
HP0319	arginyl-tRNA synthetase	1	1	0
HP0370	biotin carboxylase	0	3	0
HP0371	biotin carboxyl carrier protein	0	38 **	0
HP0390	adhesin-thiol peroxidase	2	5	0
HP0392	histidine kinase	0	16	0
HP0397	phosphoglycerate dehydrogenase	0	2	0
HP0410	neuraminyllactose-binding hemagglutinin homolog	0	14	0

HP0415	conserved hypothetical integral membrane protein	0	3	0
HP0417	methionyl-tRNA synthetase	0	8	0
HP0434	hypothetical protein	0	2	1
HP0448	hypothetical protein	0	0	2
HP0485	catalase-like protein	3	0	1
HP0486	hypothetical protein	0	3	0
HP0516	heat shock protein	0	2	0
HP0550	transcription termination factor Rho	0	2	0
HP0558	beta ketoacyl-acyl carrier protein synthase II	0	4	0
HP0561	3-ketoacyl-acyl carrier protein reductase	2	6	1
HP0570	aminopeptidase	3	2	1
HP0588	ferrodoxin-like protein	0	2	1
HP0589	ferredoxin oxidoreductase, alpha subunit	1	8	6
HP0590	ferredoxin oxidoreductase, beta subunit	0	3	0
HP0591	ferredoxin oxidoreductase, gamma subunit	1	3	0
HP0596	lipoprotein, putative	0	3	0
HP0599	hemolysin secretion protein precursor	0	18	0
HP0605	hypothetical protein	0	16	0
HP0606	membrane fusion protein	3	4	4
HP0607	acriflavine resistance protein	2	4	0
HP0629	hypothetical protein	1	1	0
HP0632	quinone-reactive Ni-Fe hydrogenase, large subunit	0	3	0
HP0649	aspartate ammonia-lyase	0	3	0
HP0657	processing protease	3	0	0
HP0690	acetyl coenzyme A acetyltransferase	0	2	0
HP0695	hydantoin utilization protein A	7	17	4
HP0696	N-methylhydantoinase	4	13	3
HP0697	hypothetical protein	0	4	0
HP0786	preprotein translocase subunit	0	3	0
HP0797	flagellar sheath adhesin	0	3	1
HP0825	thioredoxin reductase	0	2	0
HP0829	inosine-5'-monophosphate dehydrogenase	3	0	1
HP0859	ADP-L-glycero-D-mannoheptose-6-epimerase	0	2	0
HP0875	catalase	17	0	15
HP0883	Holliday junction DNA helicase	0	31	0
HP0887	vacuolating cytotoxin	7	30	10
HP0913	outer membrane protein 21	0	1	1
HP0978	cell division protein	0	2	0
HP1019	serine protease	1	8	5
HP1069	cell division protein	0	5	0
HP1118	gamma-glutamyltranspeptidase	4	0	3
HP1125	peptidoglycan associated lipoprotein precursor	0	3	0
HP1126	colicin tolerance-like protein	1	1	1
HP1131	ATP synthase F1, subunit epsilon	0	2	0
HP1132	ATP synthase F1, subunit beta	1	24	5

44

HP1134	ATP synthase F1, subunit alpha	1	36	4
HP1135	ATP synthase F1, subunit delta	0	7	0
HP1147	ribosomal protein L19	0	4	0
HP1152	signal recognition particle protein	0	2	0
HP1172	glutamine ABC transporter	0	20	0
HP1195	translation elongation factor EF-G	0	7	0
HP1198	DNA-directed RNA polymerase, beta and beta' subunit	1	2	0
HP1200	ribosomal protein L10	0	1	1
HP1205	translation elongation factor EF-Tu	6	73	3
HP1241	alanyl-tRNA synthetase	0	8	0
HP1243	outer membrane protein 28	1	31	1
HP1245	single-strand DNA-binding protein	0	125 [***]	0
HP1279	anthranilate isomerase	0	30	0
HP1286	conserved hypothetical secreted protein	3	1	2
HP1293	DNA-directed RNA polymerase, alpha subunit	0	5	0
HP1301	ribosomal protein L15	4	0	2
HP1302	ribosomal protein S5	0	3	0
HP1304	ribosomal protein L6	2	5	1
HP1310	ribosomal protein S17	0	2	0
HP1314	ribosomal protein L22	1	1	0
HP1318	ribosomal protein L4	9	1	4
HP1319	ribosomal protein L3	0	6	1
HP1320	ribosomal protein S10	0	4	0
HP1325	fumarase	0	2	0
HP1345	phosphoglycerate kinase	0	9	0
HP1350	protease	23	29	71 [***]
HP1375	UDP-N-acetylglucosamine acyltransferase	0	5	0
HP1385	fructose-1,6-bisphosphatase	0	2	0
HP1399	arginase	1	1	0
HP1428	conserved hypothetical protein	1	1	0
HP1454	hypothetical protein	2	0	0
HP1457	hypothetical protein	0	2	0
HP1460	DNA polymerase III alpha-subunit	2	1	0
HP1488	conserved hypothetical secreted protein	1	6	0
HP1496	general stress protein	0	3	0
HP1542	hypothetical protein	0	4	0
HP1554	ribosomal protein S2	0	2	0
HP1563	alkyl hydroperoxide reductase	19	18	15
HP1588	conserved hypothetical protein	1	21	2
	Total Spectral Counts	208	1428	237

[a] Based on the *H. pylori* 26695 genome annotation
[b] CagH-HA and CagL-HA were affinity purified from strains expressing these proteins using an anti-HA antibody, and a WT strain was processed in parallel as a control. The Table shows numbers of spectral counts observed by MudPIT analysis for each identified protein.
** p<0.01; *** p<0.001 when compared to WT control, according to the G-test likelihood ratio, post-spectral count normalization.

Further analysis revealed that there is weak sequence similarity between the amino acid sequences of CagL and CagH (32% amino acid identity and 47% similarity in the C-terminal region, Figure 2-2B). CagL, CagI, and CagH are encoded by contiguous genes in an operon (Figure 2-2C) [121]. Other genes included in this operon, upstream from *cagH, cagI*, and *cagL*, include *cagC* (encoding the putative T4SS major pilin component [94]), *cagE* (homologous to *virB4*, encoding an ATPase required for T4SS function), and *cagF* (encoding a chaperone protein that binds to CagA [122]). Since we observed highly reproducible co-purification of CagH, CagI, and CagL, and because genes within an operon often have related functions, we proceeded to undertake further studies of all three proteins.

CagL stability is impaired in the absence of CagH and CagI. Previous studies have reported decreased stability of T4SS components if one or more interacting proteins are absent [80,90,93,123,124]. Based on the observed co-purification of CagI and CagH with CagL, we hypothesized that CagL stability might be impaired in the absence of CagI or CagH. To assess CagL protein stability, we analyzed bacterial lysates prepared from the WT strain, Δ*cagI* mutant, and Δ*cagH* mutant, each cultured for either 24 h or 48 h. CagL could be detected by immunoblotting of lysates from all three strains at 24 h of growth (Figure 2-5A). However, after 48 h of growth, CagL levels were markedly reduced in lysates from the Δ*cagI* and Δ*cagH* mutants compared to levels in WT lysate. CagL stability at the 48 h timepoint was increased by complementation of the mutant strains (Figure 2-5A). We also evaluated the stability of CagL in the absence of two Cag proteins selected as controls (CagA and CagF). We did not detect significant co-purification of CagA or CagF with CagL in the previous experiments (Table 2-1, Table 2-3), and neither CagA nor CagF is required for functionality of the T4SS, based on analysis of IL-8 induction [55]. Absence of CagA or CagF did not have any detectable effect on CagL stability during early or late phases of growth (Figure 2-5A).

46
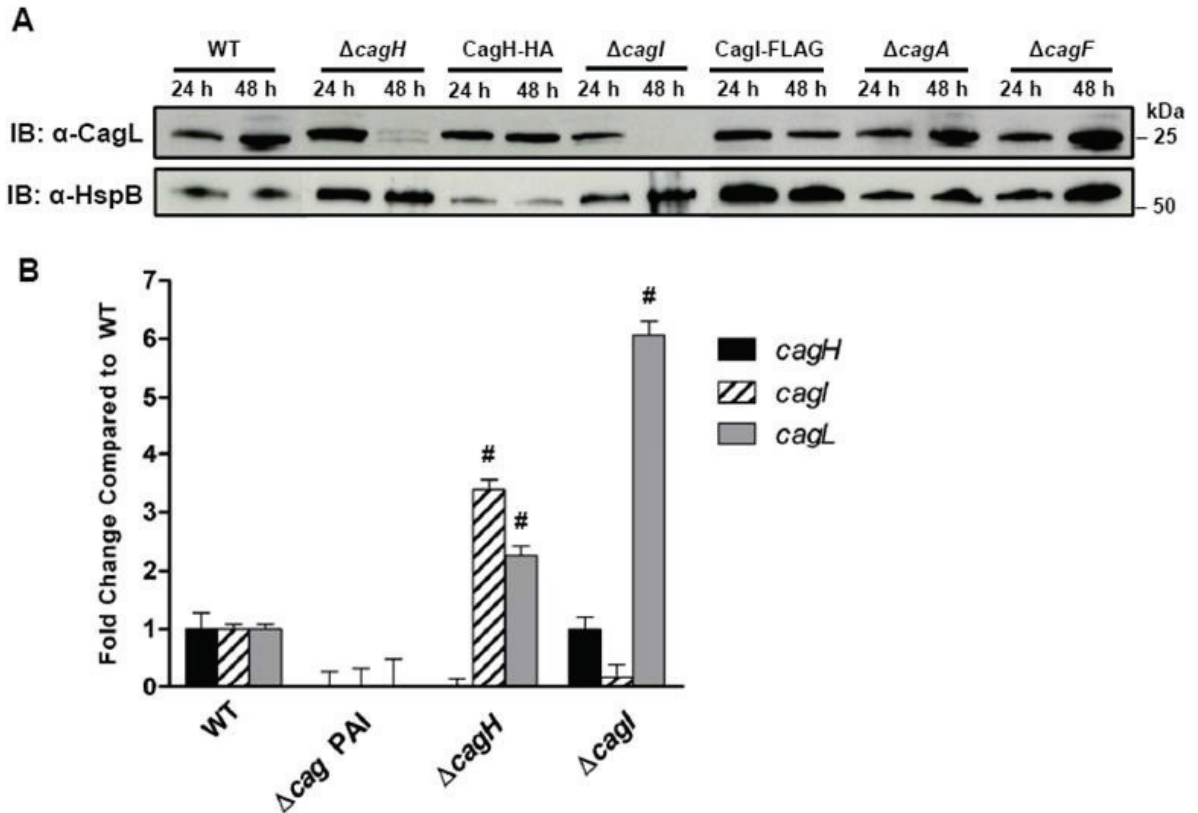
Figure 2-5. Analysis of CagL stability in Δ*cagH* and Δ*cagI* mutants. (A) *H. pylori* strains were cultured for 24 or 48 hours on solid media prior to lysis and immunoblotting to detect expression of CagL.   Expression of HspB, a heat shock protein homolog, was analyzed as a control. In the Δ*cagH* and Δ*cagI* mutants, levels of CagL were markedly lower at the 48 h timepoint than at the 24 h timepoint.   CagL levels remained stable at both timepoints in the WT strain and the complemented mutants expressing CagH-HA or CagI-FLAG. CagL stability is unaffected by absence of Cag proteins that do not interact with CagL (Δ*cagA* and Δ*cagF*).   **(B)** RT-PCR analysis of *cagH*, *cagI*, and *cagL* transcription in the indicated WT and mutant bacteria, cultured for 48 h. *cagL* transcript levels are upregulated in the Δ*cagH* and Δ*cagI* mutants, compared to levels detected in the WT strain. # indicates p<0.05 according to one-way ANOVA with Dunnett's *post hoc* correction.

We also investigated the possibility that the observed reductions in CagL protein expression in Δ*cagH* and Δ*cagI* mutants might be attributable to a reduction in *cagL* transcription. Analysis of *cagL* transcription in each of these strains by real-time PCR indicated that *cagL* transcription was not diminished, but was in fact increased in the Δ*cagH* and Δ*cagI* mutants compared to the WT strain at 48 h of growth (Figure 2-5B). Thus, the observed reduction in CagL levels in the Δ*cagH* and Δ*cagI* mutant strains is attributed to reduced stability of CagL, rather than decreased transcription of *cagL*.

To further investigate potential relationships among CagL, CagI, and CagH, we analyzed whether co-purification of CagI with CagL was dependent on the presence of CagH. We used polyclonal anti-CagL antiserum to immunoaffinity purify CagL from both the WT strain and a Δ*cag*H mutant strain, each cultured for 24 h (which permitted stable expression of CagL in both strains). The samples were then analyzed by mass spectrometry. As expected, CagL was immunoaffinity purified from both the WT strain and the Δ*cagH* mutant (Table 2-5). CagI was co-purified along with CagL from the WT strain, but not from the Δ*cagH* mutant (Table 2-5). In a similar manner, we investigated whether co-purification of CagH with CagL was dependent on the presence of CagI. CagH was co-purified with CagL from both the WT strain and Δ*cagI* mutant; however, significantly lower levels of CagH were co-purified from the Δ*cagI* mutant strain (Table 2-5).

In summary, these experiments indicate that CagL stability is reduced in the absence of CagH or CagI, and co-purification of CagH or CagI with CagL is dependent on the presence of all three proteins. Taken together with the previous results (Tables 2-1, 2-3, 2-5), these data provide evidence that CagL, CagI, and CagH are components of one or more subassemblies associated with the *cag* T4SS.

Table 2-5. Cag proteins that co-purify with CagL in the absence of CagH or CagI

Gene Number [a]	Protein	WT [b]	ΔcagH [b]	ΔcagI [b]
HP0539	CagL	43	25 *	22
HP0540	CagI	60	0 ***	0 ***
HP0541	CagH	22	0 ***	3 **
HP0530	CagV	16	13	11
HP0544	CagE	4	0	2
HP0547	CagA	82	74	42 *
HP0524	Cag5	6	9	3
HP0529	CagW	2	1	6
HP0526	CagZ	0	1	2
HP0543	CagF	1	2	1
HP0528	CagX	2	0	1
HP0527	CagY	1	3	0
Total Spectral Counts		8665	8768	6667

[a] Based on the *H. pylori* 26695 genome annotation

[b] CagL was affinity purified from the WT strain, Δ*cagH* and Δ*cagI* mutants (each cultured for 24 h) using anti-CagL polyclonal antiserum. The Table shows numbers of spectral counts observed by MudPIT analysis for each identified Cag protein.

*p<0.05; ** p<0.01; *** p<0.001 when comparing each of the mutant strains with WT, according to the G-test likelihood ratio, post-spectral count normalization.

Functional properties of CagH and CagI. A previous study, which analyzed mutant strains containing transposon insertions in individual *cag* genes, provided an important foundation for identifying genes in the *cag* PAI that are required for functionality of the *H. pylori cag* T4SS. This study reported that both CagH and CagI were required for CagA translocation, and that CagH (but not Cag*I)* was required for IL-8 induction [55]. A limitation of the methodologic approach was that unrecognized spontaneous secondary mutations or polar effects associated with transposon insertion into the *cag* PAI could not be confidently excluded. To determine definitively whether CagH and CagI were required for the function of the *cag* T4SS, we analyzed the unmarked Δ*cagH* and Δ*cagI* deletion mutant strains and complemented mutants described in the previous section.

In comparison to the WT strain, both the Δ*cagH* and Δ*cagI* mutants were defective in their ability to induce IL-8 secretion in gastric epithelial cells (Figure 2-4B). Complementation of these mutants with genes encoding CagH-HA and CagI-FLAG rescued the ability of the respective Δ*cagH* and Δ*cagI* mutants to induce IL-8 secretion at WT levels (Figure 2-4B). Immunoblot analysis indicated that CagL remained intact in the mutant strains throughout the assay (Figure 2-4B inset), which indicates that the inability of these mutant strains to induce IL-8 expression is attributable to absence of either CagH or CagI, respectively, rather than instability of CagL. The Δ*cagH* and Δ*cagI* mutants were also defective in their ability to translocate CagA into gastric epithelial cells (Figure 2-4C). The ability of these mutants to translocate CagA at levels comparable to WT was rescued by complementation with epitope-tagged versions of the corresponding proteins (Figure 2-4C). These results indicate that CagH and CagI, similar to CagL, are required for proper function of the *cag* T4SS.

Localization of CagH, CagI, and CagL.   Since expression of CagH and CagI proteins in *H. pylori* has not been detected previously, we sought to analyze the subcellular localization of these proteins.   As a first approach, we investigated potential surface exposure of CagH, CagI, and CagL by analyzing the susceptibility of these proteins, as well as several controls, to digestion by proteinase K.   As expected, incubation of the intact bacteria with proteinase K was sufficient to cleave VacA, a protein known to be localized on the bacterial surface (Figure 2-6A) [112,113].   In contrast, carbonic anhydrase (CA), a protein known to have a periplasmic domain[115], was not susceptible to cleavage by proteinase K.   Similarly, CagV (a VirB8 homolog), which is predicted to have domains in the periplasmic space [79,80,125], was also resistant to cleavage by proteinase K (data now shown).   CagH-HA, CagI-FLAG, and CagL-HA were each susceptible to cleavage by proteinase K (Figure 2-6A), but in contrast to VacA, these three Cag proteins were incompletely digested.   This suggests that CagH, CagI, and CagL are present on the bacterial surface, and in addition, another pool of these proteins exists in a site that is not susceptible to the protease.   Consistent with the susceptibility of CagH-HA, CagI-FLAG, and CagL-HA to cleavage by proteinase K, surface exposure of all three proteins was detected by flow cytometry analysis (Figure 2-6B).
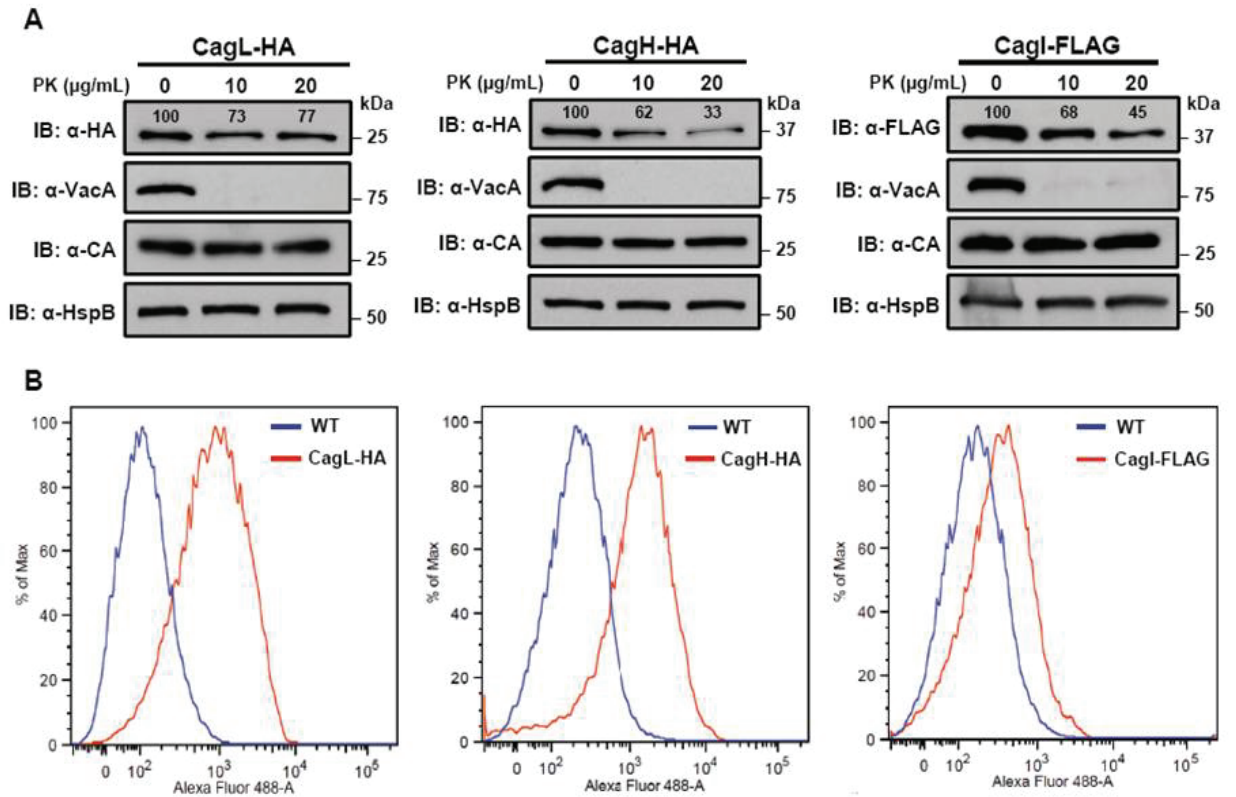
Figure 2-6. Localization of CagH and CagI. (A) Whole bacteria expressing CagL-HA, CagH-HA, or CagI-FLAG were harvested after 24 h of growth and subjected to proteolytic digestion by Proteinase K (PK). Susceptibility to cleavage by Proteinase K was monitored by immunoblotting using antibodies against the appropriate epitope tags. As a control, cleavage of VacA (a surface-exposed protein) and carbonic anhydrase (CA, an inner membrane protein with a periplasmic domain) was monitored. Heat shock protein (HspB) served as a loading control, and densitometry was performed as described in Methods. Numbers above bands representing CagL-HA, CagH-HA, and CagI-FLAG indicate the percent of protein remaining after cleavage by PK. **(B)** Flow cytometry analysis indicates that CagL-HA, CagH-HA, and CagI-FLAG proteins are detectable on the bacterial surface.

Immunogold labeling studies also revealed localization of CagH-HA, CagI-FLAG, and CagL-HA on the surface of *H. pylori* (Figure 2-7).   Typically fewer than 10 gold particles per bacterial cell were visualized on the peripheral margins of bacteria, which suggests either that the number of CagH, CagI, and CagL proteins on the bacterial surface under these conditions is relatively small, or that there may be limitations with the use of these monoclonal antibodies for immunogold labeling.
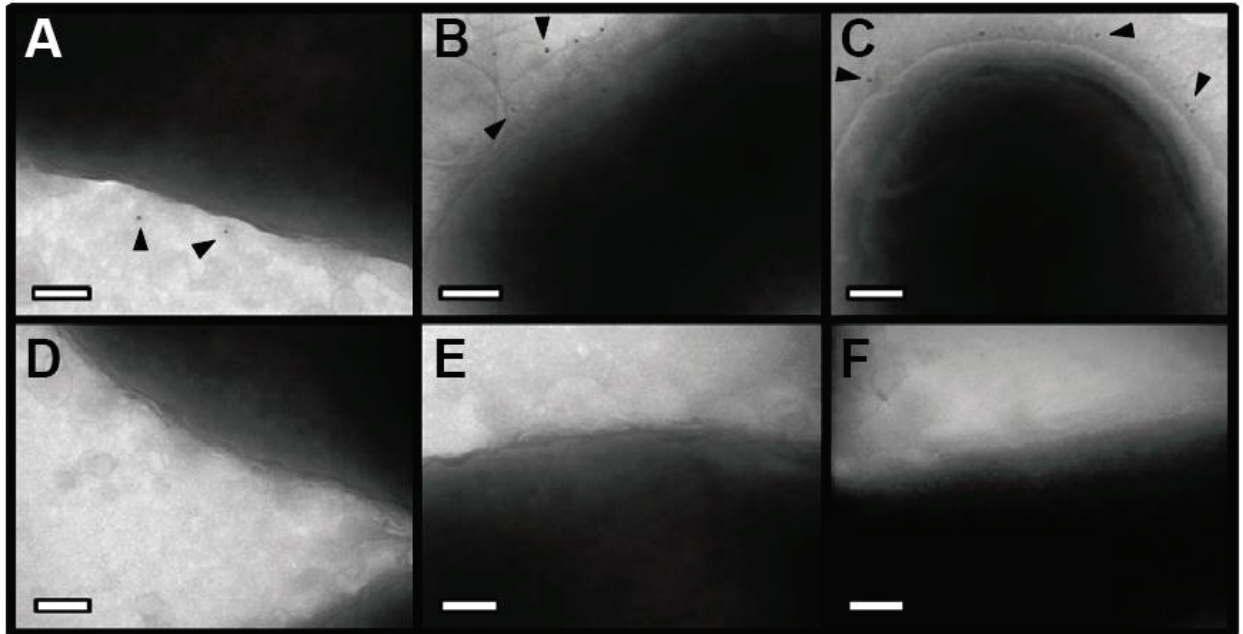
Figure 2-7. **Immunoelectron microscopy analysis of whole bacterial cells.** *H. pylori* strains were grown on blood agar plates and immunogold labeling was performed as described in Methods. **(A)** Δ*cagH*/CagH-HA, stained with anti-HA and gold-labeled secondary antibody. **(B)** Δ*cagI*/CagI-FLAG, stained with anti-FLAG and gold-labeled secondary antibody. **(C)** Δ*cagL*/CagL-HA, stained with anti-HA and gold-labeled secondary antibody. **(D)** WT strain, stained with anti-HA and gold-labeled secondary antibody. **(E)** WT strain, stained with anti-FLAG and gold-labeled secondary antibody. **(F)** WT strain, stained with secondary antibody only (representative image for all strains when treated with secondary antibody only). Magnification bars indicate 100 nm. Arrowheads indicate location of gold-labeled secondary antibody.

As another approach for analyzing the subcellular localization of CagH, CagI, and CagL, *H. pylori* strains expressing CagH-HA, CagI-FLAG, or CagL-HA were sonicated and fractionated in the absence of detergent. Soluble fractions (which are expected to contain cytoplasmic and periplasmic proteins) and total membrane fractions were analyzed by immunoblotting using monoclonal antibodies directed against the appropriate epitope tags. In agreement with a previous report [80,92], the native form of CagL was localized mainly in the soluble fraction, and lesser amounts were present in the membrane fraction; similar results were observed for CagL-HA (Figure 2-8A, B, C). Both CagH-HA and CagI-FLAG were localized exclusively in the membrane fraction (Figure 2-8B, C). The insolubility of CagH and CagI in this experiment compared to previous experiments is attributed to the absence of detergent in these fractionated samples and the presence of detergent in earlier samples. We propose that, when complexed together, CagL is loosely tethered to membrane-associated CagH and CagI, and that in the absence of detergent, the shear forces or heat generated by sonication are sufficient to disrupt this interaction and solubilize CagL, but not CagH and CagI. An additional possibility is that the CagH-CagI-CagL complexes include the relatively small portion of membrane-associated CagL.
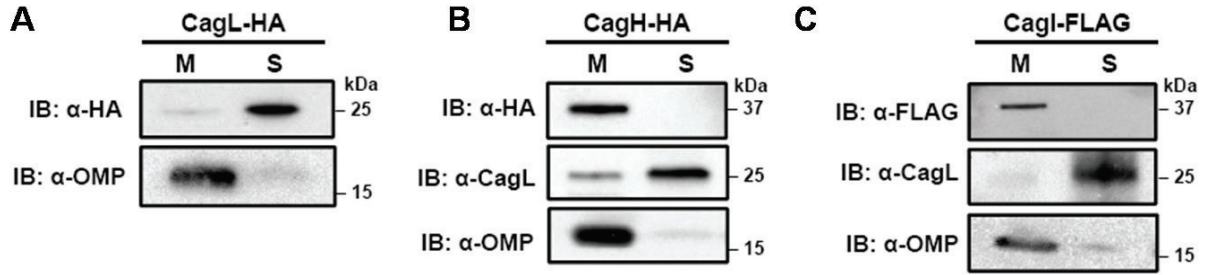
Figure 2-8. Localization of CagL, CagH, and CagI in detergent-free lysates. **(A)** Bacteria expressing epitope-tagged CagL (CagL-HA), **(B)** CagH (CagH-HA), or **(C)** CagI (CagI-FLAG) were fractionated into total membrane (M) and soluble (S) fractions as described in Methods. CagL-HA, CagH-HA, CagI-FLAG, and untagged CagL were detected by immunoblotting. As a control, fractions were immunoblotted with a monoclonal antibody that recognizes an *H. pylori* outer membrane protein (α-OMP) (Santa Cruz).

In summary, these experiments provide evidence that CagL, CagH, and CagI exist in multiple subcellular locations, including the surface of *H. pylori.* The existence of these proteins in multiple subcellular locations potentially reflects multiple stages of T4SS assembly. Localization of CagL mainly to the soluble fraction as well as on the bacterial surface helps to reconcile apparent contradictions in previous publications, which reported CagL localization to either the soluble fraction or surface-exposed sites [64,66,80,92]. In agreement with a previous study [80], we propose that CagL exists mainly in a periplasmic pool, but it can also be associated with the outer membrane, perhaps loosely tethered to the cell surface. Distribution of at least two other Cag proteins (Cag3 and CagY) into multiple pools with different localizations has been suggested previously [78,92]. Complexes comprised of CagH, CagI, and CagL could potentially exist in multiple sites, including the bacterial surface and the periplasm (where periplasmic CagL may interact with membrane-associated CagH and CagI proteins).

Discussion

One of the important mechanisms by which *H. pylori* infection leads to severe gastric disease is through the actions of the bacterial oncoprotein CagA [45,46,126]. Translocation of CagA into gastric epithelial cells occurs through a T4SS-mediated process and requires multiple proteins encoded by the *cag* PAI [42,43,79,97,125]. Several *H. pylori* proteins required for CagA translocation are distantly related to components of T4SSs in other bacterial species and presumably have conserved functions [42,55,69,80]. Herein, we provide new insights into three components of the *cag* T4SS that lack homologs in other T4SSs – CagH, CagI, and CagL.

Prior to the current study, it was known that CagL can bind α5β1 integrin and can cause several alterations in host cells [55,64,65,66,67]. CagL was localized in various studies to several bacterial subcellular sites, including a soluble bacterial fraction [80,92],

the bacterial surface [66], and pili on the surface of *H. pylori* [64].   We reasoned that CagL might physically interact with other T4SS components, and that such interactions might be required for CagL export, localization, stability, or activity.   Therefore, we conducted studies designed to identify *H. pylori* proteins that interact with CagL. In initial experiments involving affinity purification of CagL, we observed a highly reproducible co-purification of CagH and CagI with CagL, and we subsequently detected the same pattern of co-purification when either CagH or CagI was targeted for purification. We detected evidence of CagH-CagI-CagL interactions not only in experiments involving bacteria grown in pure culture, but also in experiments involving *H. pylori* that were attached to gastric epithelial cells. We were unable to undertake analysis of direct interactions among CagL, CagI and CagH because recombinant CagI and CagH could not be expressed in a soluble form under non-denaturing conditions.   However, several other results provided evidence of a physical relationship among these proteins. Specifically, CagL stability was decreased in the absence of either CagI or CagH, and co-purification of CagI or CagH with CagL was dependent on the presence of all three proteins.   All three proteins are required for functional activity of the *cag* T4SS, and all three proteins have a role in formation of pili at the interface between *H. pylori* and gastric epithelial cells. Other intriguing relationships among these proteins include their linkage within a single operon in the *cag* PAI [121], and ability of both CagI and CagL to bind β1 integrin [64,65]. Collectively, the data suggest that CagL, CagI, and CagH are components of one or more T4SS subassemblies involved in pilus biogenesis.

Recognition of the relationships among CagH, CagI, and CagL in the current study was facilitated by using immunoaffinity purifications and a robust multidimensional mass spectrometry protein identification technique. In comparison to traditional co-immunoprecipitation experiments, which rely upon the availability of antibodies to detect potential interacting partners of a target protein by immunoblotting, the mass

spectrometry-based approach used in the current study allowed a comprehensive analysis that was not limited by the availability of antibodies. In addition, this approach affords advantages compared to two-hybrid screens, because it allows detection of interactions among native proteins expressed in *H. pylori,* as well as the potential to detect indirect secondary and tertiary interactions. Analysis of the appropriate control strains allowed for assessment of non-specific protein interactions and statistical evaluation of results.

Prior to the current study, CagH and CagI had not been investigated in detail. Neither CagH nor CagI was detected in analyses of *H. pylori* using 2D gel proteomic methodology [81,127], and immunologic detection of CagI expression has been hindered by difficulty in raising antibodies against the protein [65]. Similarly, our efforts to raise polyclonal antisera against CagH or CagI proteins failed to yield antisera that recognized the appropriate antigens. However, we were able to detect CagH and CagI expression by mass spectrometry, and we were also able to detect epitope-tagged forms of these *H. pylori* proteins. One study reported that CagI was capable of binding β1 integrin in a yeast two hybrid assay, and CagI interactions with integrin were confirmed by analyzing binding of a CagI-GST fusion protein to cell lines deficient in β1 integrin or cell lines that had been genetically complemented with human β1 integrin [65]. Another study reported that *H. pylori* mutant strains containing transposon insertions in *cagH* or *cagI* were defective in their ability to translocate CagA into host cells, and that CagH (but not CagI) was required for induction of IL-8 synthesis and secretion by cultured gastric epithelial cells [55]. Herein, we report that *cagH* and *cagI* mutant strains were defective in both CagA translocation and IL-8 induction. The discrepancy in results pertaining to CagI and IL-8 induction might be attributable to differences in the mutagenesis methods utilized to generate Δ*cagI* isogenic mutants. We analyzed unmarked strains containing deletions of the relevant *cag* genes, and complementation of the unmarked mutants

confirmed that CagH and CagI are required for CagA translocation and stimulation of IL-8 secretion by cultured gastric epithelial cells.

Localization analyses performed in the current study indicated that CagH, CagI, and CagL can all be exported to the surface of *H. pylori,* and in addition, provided evidence for the existence of non-surface-exposed pools of these proteins (Figures 2-6, 2-8). The existence of these proteins in multiple subcellular locations potentially reflects multiple stages of T4SS assembly. Detection of CagL on the surface of *H. pylori* in the current study agrees with previous reports, which reported localization of CagL to either the bacterial surface or to pilus structures, based on immunogold EM staining [64,66]. Other lines of evidence supporting a surface-exposed localization of CagI and CagL include the ability of both proteins to bind β1 integrin [64,65], signatures of positive selection in the genes encoding these two proteins, and the presence of a predicted signal sequence in both proteins (Figure 2-2A) [128]. The detection of a non-surface-exposed pool of CagL in the current study is consistent with results of previous studies, which detected CagL in a soluble bacterial fraction consistent with the periplasm [80,92].

Determining the subcellular site at which CagH, CagI, and CagL assemble into complexes is complicated by the presence of these proteins in multiple subcellular sites. All three proteins can localize to the bacterial surface, and this is one site where the formation of CagH-CagI-CagL complexes may be relevant. In addition, since a large proportion of CagL is localized to a soluble fraction, it seems likely that periplasmic CagL may interact with membrane-associated CagH or CagI proteins that have domains in the periplasm. We detected evidence of CagH-CagI-CagL complexes not only in pure bacterial cultures, but also in co-cultures of *H. pylori* with gastric epithelial cells (Tables 2-1, 2-3, 3-2). Since the CagH-CagI-CagL complexes were detected in pure bacterial cultures (which lack detectable surface pili), the complexes must clearly be localized to sites distinct from pili under these conditions.

Because CagH, CagI, and CagL are present on the surface of *H. pylori* under conditions in which the T4SS is not fully assembled, and CagA is not translocated, we speculate that these proteins are involved in biogenesis of the T4SS pilus structure when *H. pylori* is in contact with gastric epithelial cells. The requirement of these unique surface-exposed Cag proteins in formation of the T4SS pilus structure sheds light on the dramatic differences that exist between the *H. pylori* T4SS and T4SSs of other Gram-negative bacteria.

CHAPTER III

ROLE OF UNIQUE *HELICOBACTER PYLORI* CAG PROTEINS IN BIOGENESIS OF THE TYPE IV SECRETION SYSTEM PILUS AT THE INTERFACE OF BACTERIA AND GASTRIC EPITHELIAL CELLS

Introduction

Surface-exposed pili are an important feature of T4SSs [69,93]. The T4SS pili of *A. tumefaciens* are comprised of VirB2 (the major pilin subunit), VirB5 (the minor pilin subunit), and the pilus-associated protein VirB7 [69,93,129]. In the *Agrobacterium* T4SS, the presence of the membrane-associated and surface-exposed T pilus structure is indispensable for delivery of the nucleoprotein complex cargo to its host cell. VirB2 is processed and cyclized via peptide bond linkage of the N- and C-terminal residues, generating the T pilin [129]. The T pilin subunits are exported across both bacterial membranes to the surface of the cell, where the pilin subunits are able to be assembled into a filamentous pilus structure having a diameter of approximately 10 nm [129]. All 11 Vir proteins encoded by the *vir* operon are required for transport of the VirB2 pilin subunits to the surface of the bacteria. However, processing and cyclization of VirB2 propilin does not require the presence of any other Vir protein [129].

Thus far, the mechanisms by which the *cag* T4SS delivers CagA and peptidoglycan to the host cell in order to mediate myriad diverse effects are unknown. The possibility exists that in addition to CagA and peptidoglycan [43,46,56], multiple other *H. pylori* proteins can utilize the *cag* PAI-encoded T4SS as a conduit for delivery into the gastric epithelium. In many Gram-negative pathogens, including *Salmonella*, *Shigella*, and *Yersinia*, a type III secretion system (T3SS) is exploited for delivery of multiple effector proteins to the host cell [130,131]. T3SSs purified from *Salmonella* and *Shigella* clearly contain a surface-exposed needle-like appendage which presumably serves to inject effector proteins directly into the cytoplasm of the recipient cell

[130,131,132]. Despite extensive characterization of T3SS apparatuses, including analysis by electron microscopy [131,132,133,134] and generation of atomic-scale resolution models of T3SS complexes by crystallization [131,135,136], only low resolution analyses of T4SS pilus organelles exist for the T pilus of *Agrobacterium* and pilus structures involved in bacterial conjugation [137,138].

When *H. pylori* is in contact with gastric epithelial cells, the bacteria express pili that are associated with the *cag* T4SS [64,92,94,95]. Relatively little is known about the composition and biogenesis of these *H. pylori* pili. In contrast to T4SS pili from other bacterial species, the *H. pylori* T4SS pili are reported to be sheathed organelles [92]. *H. pylori* CagC is reported to be a VirB2 homolog [94], but there is not yet convincing evidence that CagC is a major component of *H. pylori* pili. Several of the *H. pylori* proteins (CagY, CagT, and CagX) that have been localized to the pili by immunogold staining [92,95] are distantly related to core complex components in other T4SS (VirB10, VirB7, and VirB9, respectively). *H. pylori* CagL has also been localized to pili by immunogold staining and it is suggested that it is a minor pilus component [64]; however, CagL and VirB5 do not exhibit any detectable sequence similarity. Taken together, these findings suggest that the pili associated with the *H. pylori cag* T4SS are considerably different from the T4SS pili of *A. tumefaciens* and other known bacterial T4SSs.

Based on surface-exposure of CagH, CagI, and CagL in the absence of host cell contact, localization of CagL to a pilus structure formed when *H. pylori* is in contact with gastric epithelial cells, and a potential interaction of CagL and CagI with β1 integrin on the surface of gastric cells, we hypothesized that CagH, CagI, and CagL each have a role in biogenesis of the T4SS pilus. Herein, we utilize scanning electron microscopy studies to reveal that CagI and CagL are required for formation of pili at the interface between *H. pylori* and gastric epithelial cells. We demonstrate that a Δ*cagH* mutant

63

strain produces pili with a distorted morphology compared to the pili of wild-type bacteria. This suggests that CagH functions as a regulator of pilus dimensions. We identify a conserved flagellar hook protein K (FlgK) domain within the N-terminal region of CagH that may contribute to the ability of CagH to control T4SS pilus dimensions. Furthermore, we show that CagH, CagI, and CagL proteins contain a conserved C-terminal hexapeptide motif that is critical for T4SS functionality, and in the case of CagI and CagL, this motif is required for T4SS pilus formation. These studies highlight the important functions of these unique *H. pylori cag* T4SS components and illustrate the marked variation that exists among bacterial T4SSs.

Methods

Bacterial strains and growth conditions, AGS cell culture, mutagenesis of *H. pylori*, IL-8 ELISA, CagA translocation assay, bacterial subcellular fractionation, and statistical analysis were performed as described in Chapter II.

FESEM of H. pylori in contact with gastric epithelial cells. H. pylori and AGS human gastric cells were co-cultured at a MOI of 100:1 on tissue culture-treated coverslips (BD Biosciences) for 4 h at 37°C in the presence of 5% CO_2. Cells were fixed with 2.0% paraformaldehyde, 2.5% glutaraldehyde in 0.05M sodium cacodylate buffer for 1 h at 37°C. Coverslips were washed with sodium cacodylate buffer and secondary fixation was performed with 1% osmium tetroxide at room temperature for 2 h. Coverslips were washed with sodium cacodylate buffer and dehydrated with sequential washes of increasing concentrations of ethanol. Samples were then dried at the critical point, mounted onto sample stubs, grounded with a thin strip of silver paint at the sample edge, and sputter-coated with gold before viewing with a Zeiss Supra 35V FEG SEM. Analysis of pilus dimensions and image analysis was performed using Image J software.

Results

CagH, CagI, and CagL have key roles in T4SS pilus formation. To investigate a possible mechanism by which CagH, CagI, and CagL might contribute to activity of the *cag* T4SS, we used scanning electron microscopy (FESEM) to analyze interactions of *H. pylori* with AGS gastric epithelial cells. Specifically, we tested the hypothesis that CagH, CagI, or CagL might be required for formation of pili at the bacterial-host cell interface. As expected, we detected pili at the interface between WT *H. pylori* and AGS cells (Figure 3-1A), and consistent with previous reports [64,92,95], these pili were not visualized when a Δ*cag* PAI mutant (Figure 3-1B), *cagT* (*virB7* homolog) mutant, or *cagE* mutant (data not shown) were co-cultured with AGS cells. Adherent WT bacteria exhibited pili on surfaces that were contiguous to the epithelial cells, and we did not observe pili on non-adherent bacteria. The dimensions of the pili were approximately 14 nm in width and 65 nm in length (Table 3-1), and the pili were distributed along the lateral and polar surfaces of the bacteria. When co-cultured with AGS cells, neither the Δ*cagL* nor the Δ*cagI* mutant strains produced detectable pili (Figure 3-1C, F). This defect was rescued by complementation of the Δ*cagI* and Δ*cagL* mutants (Figure 3-1D, E, G).
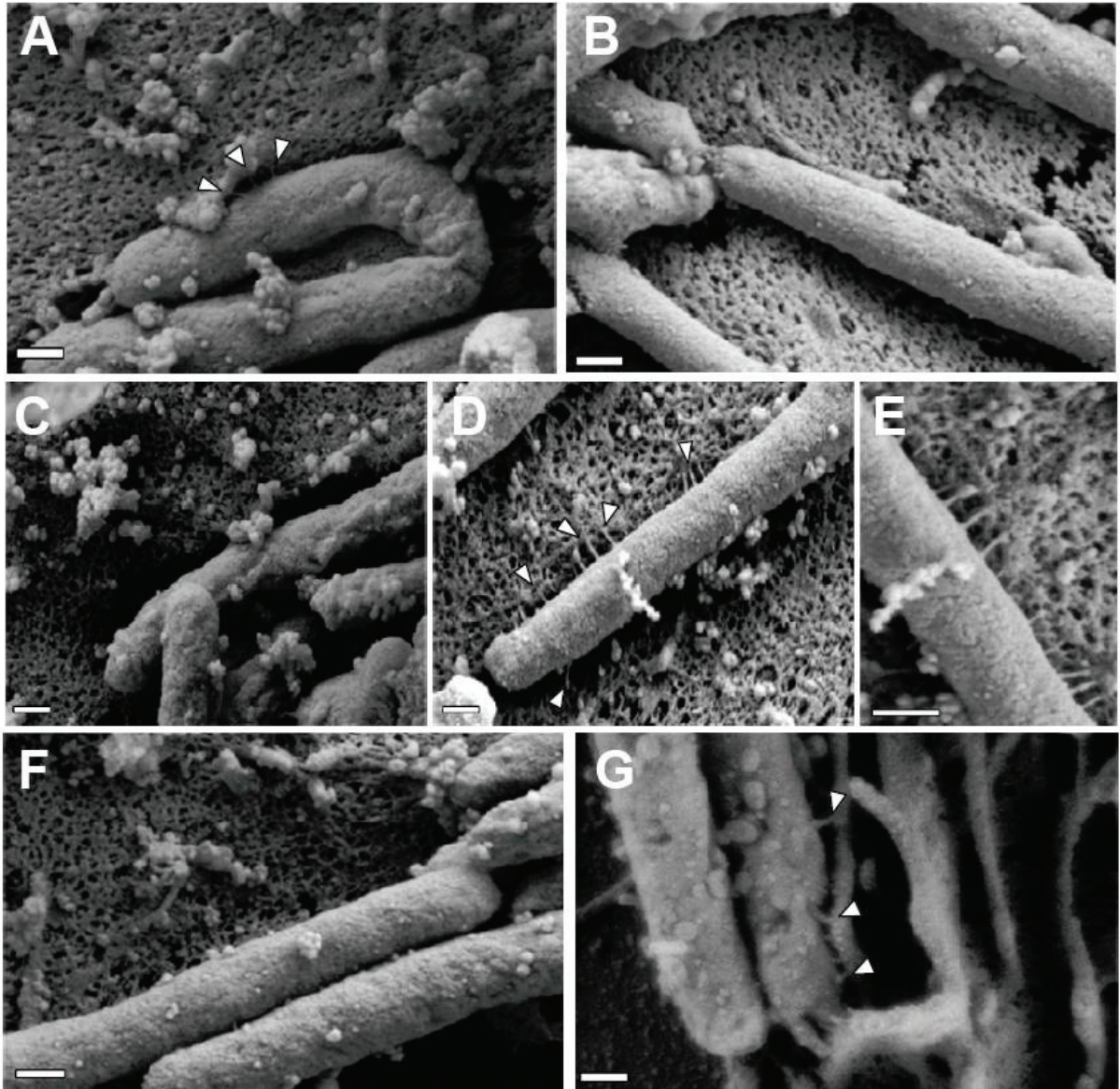
Figure 3-1. Role of CagI and CagL in pilus formation. WT *H. pylori*, mutant strains, and complemented mutant strains were co-cultured with AGS cells and then analyzed by FESEM. **(A)** WT. **(B)** Δ*cag* PAI. **(C)** Δ*cagL*. **(D)** Δ*cagL*/CagL-HA. **(E)** High magnification of pilus structures observed in Δ*cagL*/CagL-HA (panel D). **(F)** Δ*cagI*. **(G)** Δ*cagI/*CagI. Arrowheads indicate T4SS pilus structures. Magnification bars indicate 200 nm.

Table 3-1. Analysis of type IV secretion system pili [a]

	Width (nm)	Length (nm)	Pili/cell
WT	13.7 ± 2.4	65.1 ± 30.7	3.8 ± 2.6
Δ*cagH*	21.5 ± 4.3 ***	115.1 ± 65.1 ***	8.0 ± 4.3 ***
Δ*cagH*/CagH-HA	14.4 ± 2.8	65.6 ± 29.9	3.3 ± 2.9
Δ*cagH*/CagH-HAΔCT	15.8 ± 5.1	53.4 ± 18.5	3.5 ± 2.7
Δ*cagI*/CagI	15.7 ± 3.8	91.1 ± 46.1	3.1 ± 1.7
Δ*cagL*/CagL-HA	14.6 ± 2.8	81.7 ± 27.2	3.8 ± 2.8

[a] Quantification of pilus width, length, and number is based on a minimum of 2 biological replicates and a minimum of 5 representative micrographs containing a total of at least 20 adherent *H. pylori* cells.

***, $p < 0.001$ compared to WT, based on two-tailed T-test analysis.

In contrast to the Δ*cagI* and Δ*cagL* mutant strains, the Δ*cagH* mutant strain was capable of producing pili. However, there were several striking differences when comparing the Δ*cagH* mutant strain (Figure 3-2B, C, D, F) with the WT strain (Figure 3-2A, E). First, the number of pili produced by the Δ*cagH* mutant was significantly higher than the number of pili produced by the WT strain (8.0 ± 4.3 vs. 3.8 ± 2.6 visible pili per adherent bacteria, $p < 0.0001$) (Table 3-1). Second, the length of pili produced by the Δ*cagH* mutant was significantly greater than that of pili produced by the WT strain (115.1 ± 65.1 nm vs. 65.1 ± 30.7 nm, $p < 0.0001$). Third, the width of pili produced by the Δ*cagH* mutant was significantly greater than that of pili produced by the WT strain (21.5 ± 4.3 nm vs. 13.7 ± 2.4 nm, $p < 0.0001$). A complemented Δ*cagH* mutant strain produced pili that were indistinguishable from the pili produced by the WT strain (Figure 3-2G and Table 3-1). These results suggest that CagH may be a regulator of pilus dimensions. As considered further in the Discussion, an analysis of highly conserved domains within CagH (Figure 3-2H) provides clues into possible mechanisms by which CagH controls dimensions of the *H. pylori* T4SS pilus.

H

| FlgK PSSM 1 | FlgK PSSM 2 | FlgK PSSM 3 |

Y. pestis YscP NNHDFNLSPLAEGATNRAHLYQQFDDRVESIINALMPLAPFLEGVTCETGTSSESPCEPS
S. enterica Typhi FliK KDARITLADLQAAGGKLSKLKTSALAALSKTAQDLASLSALFAMLPGCPVATPVAGETPA
H. pylori CagH SIAEAVSMPMKAANQQYNQLYQGFNDQSMAVGNNINTQGAQLSAVNSQIASIASNTTPKN
B. mallei FlgK SNAQTLASQLVAAGQQYSQLRQSVNSQLTDTVTQINSYTSQLAQLNEQIASASSQGQPPN
B. pseudomallei FlgK SNAQTLASQLVAAGQQYSQLRQSVNSQLTDTVTQINSYTSQLAQLNEQIASASSQGQPPN

Figure 3-2. Analysis of pili formed by a Δ*cagH* mutant strain. WT *H. pylori*, a Δ*cagH* mutant, and a complemented mutant (Δ*cagH*/CagH-HA) were co-cultured with AGS cells and then analyzed by FESEM. **(A)** WT strain. **(B, C, D)** Multiple images of the Δ*cagH* mutant. **(E)** High magnification of panel A (WT strain). **(F)** High magnification of panel B (Δ*cagH* mutant). **(G)** Δ*cagH*/CagH-HA. Magnification bars indicate 200 nm. **(H)** Use of a position-specific scoring matrix (PSSM) within the NCBI Conserved Domain Database [139] identified a conserved flagellar hook protein K (FlgK) domain in CagH. The alignment compares the region of CagH corresponding to the FlgK domain, FlgK domain sequences of *Burkholderia* spp., and sequences of two members of the flagellar hook superfamily that function as molecular rulers (FliK from *Salmonella* and YscP from *Yersinia*) [140,141,142,143]. Arrows above the alignment indicate three portions of the FlgK domain identified in CagH.

In all of the initial experiments (Tables 2-1, 2-3, 2-6), interactions among CagH, CagI, and CagL were detected in *H. pylori* that were cultured in the absence of gastric epithelial cells. Our electron microscopy experiments indicate that pili associated with the *cag* T4SS are not detected when the bacteria are cultured in the absence of gastric epithelial cells (data not shown), and in agreement with previous reports [64,92], we found that contact of *H. pylori* with gastric epithelial cells stimulates the formation of pili that are associated with the *cag* T4SS.   This suggests that the *cag* T4SS is not yet fully assembled when bacteria are cultured in the absence of gastric epithelial cells.   To determine whether interactions among CagH, CagI, and CagL are detectable under conditions in which the *cag* T4SS is fully assembled, we sought to detect the presence of this subassembly in *H. pylori* that are co-cultured with gastric epithelial cells.   We co-cultured bacteria expressing CagH-HA with AGS cells, affinity purified CagH-HA from adherent bacteria, and then analyzed the protein content of the affinity purified preparation. Wild-type bacteria were co-cultured with AGS cells and were processed in parallel as a control. As shown in Table 3-2, we again observed co-purification of CagH, CagI, and CagL.   Therefore, a CagH-CagI-CagL subassembly is detectable not only when bacteria are cultured in the absence of gastric epithelial cells, but also under conditions in which the *cag* T4SS is fully assembled.

Table 3-2. Co-purification of CagH, CagI, and CagL from *H. pylori* attached to gastric epithelial cells

Gene Number [a]	Protein	anti-HA [b]	
		WT	CagH-HA
HP0539	CagL	0	24 [***]
HP0540	CagI	0	25 [***]
HP0541	CagH	0	20 [**]
HP0527	CagY	0	2
Total Spectral Counts		35	220

[a] Based on the *H. pylori* 26695 genome annotation

[b] An *H. pylori* strain expressing CagH-HA and a WT strain were each co-cultured with AGS cells for 5 h. CagH-HA was affinity purified using an anti-HA antibody, and the WT co-culture sample was processed in parallel as a control. The Table shows numbers of spectral counts observed by MudPIT analysis for each identified Cag protein.

** $p < 0.01$; *** $p < 0.001$ when compared to WT control, according to the G-test likelihood ratio, post-spectral count normalization.

Since CagL was previously detected as a pilus component [64], we hypothesized that CagI and CagH might also localize to pili. To test this hypothesis, we used scanning EM and immunogold labeling studies to analyze *H. pylori* that were co-cultured with gastric epithelial cells.   Using this approach, we were unable to detect localization of CagH or CagI to the pili, and we were also unable to detect CagH or CagI in any other sites.   Since we were able to detect surface localization of CagH, CagI, and CagL in transmission EM experiments and flow cytometry experiments but not FESEM experiments, there are probably limitations associated with the use of these monoclonal antibodies for immunogold labeling in the context of FESEM. Specifically, the FESEM methodology requires multiple extra washing steps (a series of seven sequential ethanol dehydration steps and three liquid carbon dioxide washing steps) that are not required for transmission EM, and monoclonal antibodies often are considered suboptimal compared to polyclonal antisera for immunogold EM studies [144].

Analysis of a conserved C-terminal motif.   Careful inspection of the sequences of CagH, CagI, and CagL revealed that all three proteins contain a conserved C-terminal motif (Figure 3-3) consisting of the distal six amino acids of each protein.   The C-terminal motifs of CagH, CagI, and CagL are encoded by divergent DNA sequences (Figure 3-3).

Figure 3-3. Identification of a conserved C-terminal motif in CagH, CagI, and CagL. A conserved 6-amino acid C-terminal motif is present in CagH, CagI, and CagL (WebLogo [145]). Numbers indicate position of the amino acid residue from the C-terminus of CagH, CagI, or CagL. This C-terminal motif is encoded by divergent DNA sequences in *cagH, cagI*, and *cagL*.

To investigate whether this C-terminal motif is functionally important, we generated *H. pylori* mutant strains expressing forms of CagH, CagI, or CagL in which this hexapeptide motif was deleted, as described in Methods. Deletion of the C-terminal motif individually in CagH, CagI, and CagL resulted in reduced levels of the mutated proteins compared to levels of the corresponding WT proteins (Figure 3-4A), along with a marked reduction in ability of the mutant bacteria to induce IL-8 secretion by AGS cells (Figure 3-4B) and abolishment of CagA translocation (Figure 3-4C).
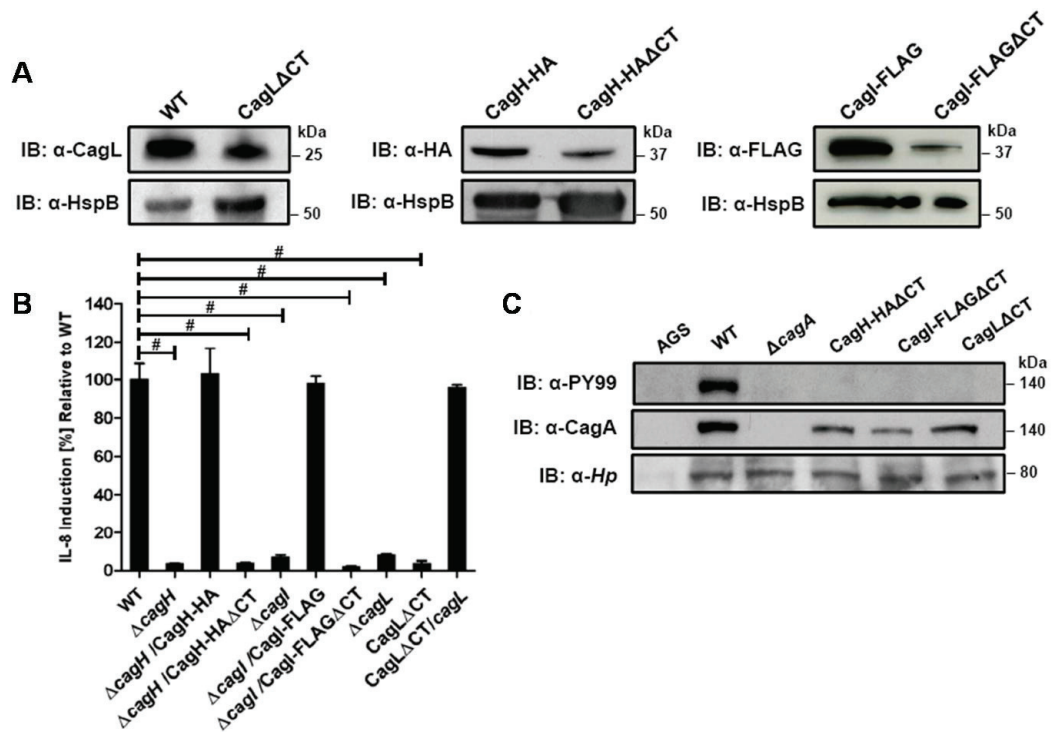
Figure 3-4. Functional analysis of a conserved C-terminal motif in CagH, CagI, and CagL. **(A)** Immunoblot analysis of mutants expressing C-terminally truncated forms of CagL, CagH-HA, or CagI-FLAG, each lacking a six-residue C-terminal motif. CagLΔCT was expressed from the endogenous *cagL* locus in the *cag* PAI, and CagH-HAΔCT and CagI-FLAGΔCT were expressed from the *ureA* locus. CagL expression was detected with anti-CagL serum, CagH-HA expression was detected with an anti-HA monoclonal antibody, and CagI-FLAG expression was detected with an anti-FLAG antibody. Truncated forms of CagL, CagH, and CagI were detected in each of the mutant strains, but the levels of the truncated proteins were reduced compared to levels of the full-length proteins. **(B)** AGS cells were co-cultured with WT *H. pylori* or the indicated mutants, and IL-8 expression was analyzed by ELISA as described in Methods. Values from six replicate samples were compared to the wild-type control by ANOVA followed by Dunnett's *post hoc* correction; # indicates p<0.05. **(C)** AGS cells were co-cultured with WT *H. pylori* or the indicated mutants, and tyrosine-phosphorylated CagA was detected with an anti-phosphotyrosine antibody (α-PY99).

76

In contrast to wild-type CagH-HA and CagI-FLAG, which each localize to the membrane fraction (Figure 2-8 and 3-5A), CagH-HAΔCT and CagI-FLAGΔCT are partially mislocalized to the soluble fraction (Figure 3-5A). Furthermore, in the absence of the C-terminal motif of CagH-HA or CagI-FLAG, CagL stability is markedly reduced at the 48 h growth timepoint compared to CagL stability at the 24 h growth timepoint (Figure 3-5B). This result further supports the conclusion that CagH, CagI, and CagL are all members of a protein subassembly.
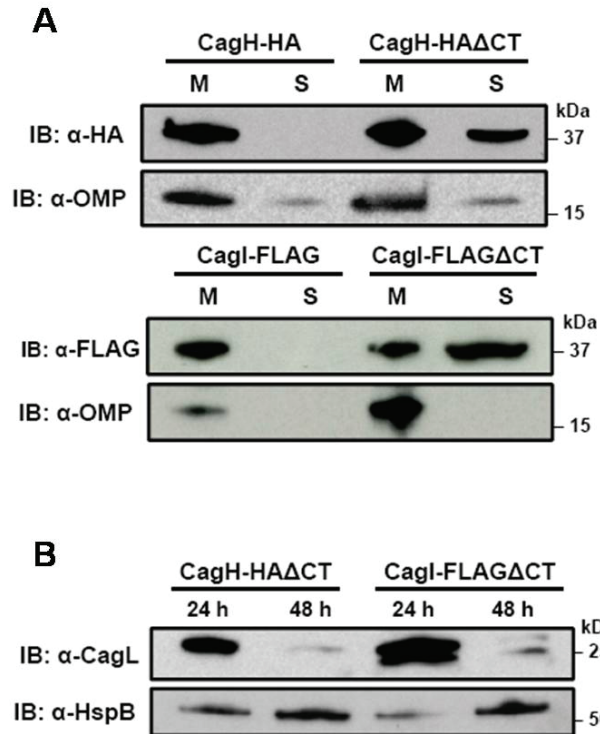
Figure 3-5. Localization of C-terminally truncated CagH and CagI. (A) Bacteria expressing CagH-HA, a C-terminally truncated form of CagH-HA (CagH-HAΔCT), CagI-FLAG, or a C-terminally truncated form of CagI (CagI-FLAGΔCT) were fractionated into total membrane (M) and soluble (S) fractions as described in Methods, and the tagged proteins were detected by immunoblotting. Intact CagH-HA and CagI-FLAG proteins were detected mainly in the membrane fraction. Deletion of the C-terminal motif from CagH-HA (CagH-HAΔCT) or CagI (CagI-FLAGΔCT) results in partial mislocalization of these proteins to the soluble fractions. **(B)** Bacteria expressing C-terminally truncated CagH (CagH-HAΔCT) or CagI (CagI-FLAGΔCT) were cultured for 24 or 48 h prior to lysis and immunoblotting to detect CagL. HspB was monitored as a loading control. CagL levels are markedly reduced at 48 h of growth in strains expressing truncated CagH or CagI, compared to strains expressing full-length proteins.

Finally, we conducted experiments to determine whether the C-terminal motifs found in CagH, CagI, and CagL were required for formation of pili at the interface between *H. pylori* and gastric epithelial cells.   Mutant strains lacking the C-terminal motif in CagH, CagI, or CagL were co-cultured with AGS cells, and the samples were then analyzed by FESEM.   Bacteria expressing CagI-FLAGΔCT or CagLΔCT each failed to produce detectable pili (Figure 3-6C, D), and thus had a phenotype indistinguishable from Δ*cagI* and Δ*cagL* mutants (Figure 3-6).   Bacteria expressing CagH-HAΔCT expressed pili that were indistinguishable from those produced by the WT strain (Figure 3-6A, B and Table 3-1).   Despite producing normal-appearing pili, the CagH-HAΔCT mutation rendered the *cag* T4SS non-functional (Figure 3-4B, C).   These experiments reveal that physical contact of the T4SS secretion pilus with gastric epithelial cells is not sufficient to induce IL-8 production, and indicate that the C-terminal motif present in CagH, CagI, and CagL is critical for functionality of the T4SS.
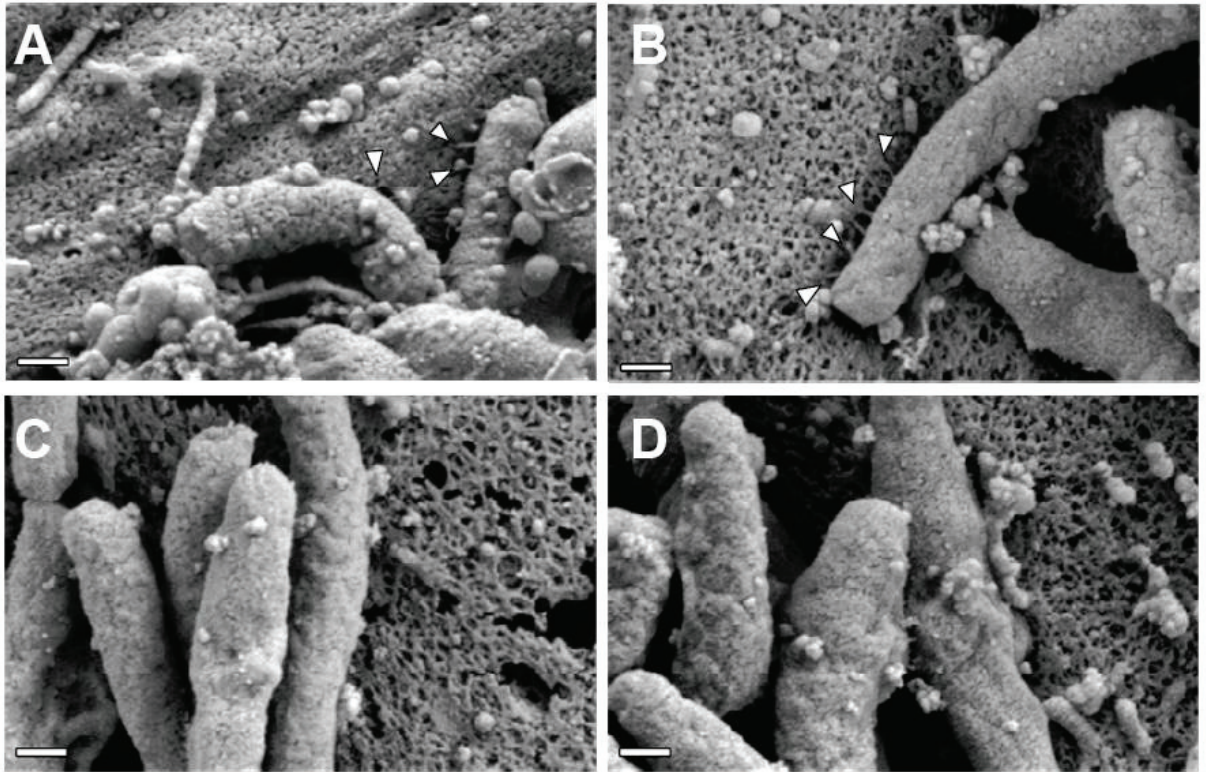
Figure 3-6. The C-terminal motif of CagI and CagL is required for pilus formation.
WT *H. pylori* or mutant strains expressing C-terminally (CT) truncated forms of CagH,
CagI, or CagL were co-cultured with AGS cells and imaged by FESEM. **(A)** WT. **(B)**
CagH-HAΔCT. **(C)** CagI-FLAGΔCT. **(D)** CagLΔCT. Magnification bars indicate 200 nm.

Discussion

Herein, we describe the requirement for both CagL and CagI are required for formation of T4SS pili. Complementation of Δ*cagL* and Δ*cagI* mutant strains resulted in restoration of pilus formation and T4SS function. In previous studies, CagY, CagT, and VirB11 ATPase were reported to be required for formation of pili [64,92,95], but the role of these proteins in *H. pylori* pilus formation has not yet been verified by testing of complemented mutant strains. CagT, CagY, and the VirB11 ATPase are each homologous to proteins found in T4SSs of other bacteria, but in contrast, CagI and CagL do not have homologues in other T4SSs.

A previous study used immunogold labeling methods and polyclonal antiserum to detect localization of CagL to pilus structures [64]. We were unable to detect localization of CagH or CagI to pilus structures, probably due to limitations associated with the use of monoclonal antibodies for immunogold labeling in the context of scanning electron microscopy. Specifically, the FESEM methodology requires multiple extra washing steps (a series of seven sequential ethanol dehydration steps and three liquid carbon dioxide washing steps) that are not required for transmission EM, and monoclonal antibodies often are considered suboptimal compared to polyclonal antisera for immunogold EM studies [144].   Therefore, at present it is not known whether CagL localization to pili requires dissociation of CagL from CagH and CagI, or whether all three proteins eventually localize to pili.

Interestingly, a Δ*cagH* mutant strain was hyperpiliated and expressed pili that were elongated and thickened in comparison to pili of the WT strain.   The increased number of pili visualized in images of the Δ*cagH* mutant strain might be attributable to their increased thickness and decreased fragility, resulting in a greater likelihood that the structures are preserved and visualized. Complementation of the Δ*cagH* mutant strain resulted in production of pili with a WT morphology. These results suggest that CagH is

81

a regulator of pilus dimensions. Analysis of the CagH sequence using a conserved domain database search (NCBI) [139] indicates that it contains a flagellar hook-associated protein (FlgK) domain (PRK06945) (Figure 3-2H). The FlgK domain within CagH is most closely related to the corresponding domains found in FlgK proteins of *Burkholderia* spp*.,* rather than the corresponding domain found in *H. pylori* FlgK. The FlgK domain in CagH also exhibits similarity to portions of the FliK flagellar protein of *Salmonella* and the YscP type III secretion system protein of *Yersinia pestis,* two proteins that are known to be determinants of either flagellar hook length [140,141,142,146,147,148] or type III secretion system needle length [143]. We speculate that, analogous to the role of FlgK (a flagellar hook-junction protein) in terminating flagellar hook assembly, CagH may have a role in terminating pilus assembly. Alternatively, analogous to FliK and YscP, CagH may serve as a molecular ruler to control the dimensions of *cag* T4SS pili in *H. pylori*. It will be important in future studies to dissect the molecular mechanisms by which CagH regulates pilus dimensions.

Similar to previous studies [64,92,94,95], we observed that *H. pylori* contact with gastric epithelial cells stimulated the production of pili, and in agreement with previous studies [64,95], we observed that Δ*cag* PAI mutant, Δ*cagT* mutant, and Δ*cagE* mutant strains failed to produce pili. However, the dimensions of the pili visualized in the current study differed from what was reported in a previous study. Specifically, a previous study [92] reported that the pili were between 45 and 75 nm in width, depending on presence or absence of a sheath structure, whereas in the current study, we found that the pili were much thinner. The same strain of *H. pylori* was used in both studies, so this difference is probably not attributable to strain-dependent variation. We speculate that the difference in reported dimensions might be attributable to differences in methods for scanning EM, such as the use of a relatively thin layer of gold in the current study and a thicker layer of coating in the previous study.

A striking feature shared by CagH, CagI, and CagL is a conserved C-terminal hexapeptide motif (Figure 3-3). This C-terminal motif is not found in any other protein encoded by *H. pylori* strain 26695. We show that this motif is required for functional activity of the T4SS, and that the C-terminal motifs of CagI and CagL are required for pilus formation. Deletion of the C-terminal motif leads to partial mislocalization of CagH and CagI to a soluble fraction instead of a membrane fraction (Figure 3-5A). Based on the observed mislocalization of C-terminally truncated CagH and CagI proteins, we hypothesize that this C-terminal motif plays a role in protein sorting or localization. In several other bacterial species, C-terminal motifs are required for recognition of proteins by the T4SS machinery. For example, RalF, an effector molecule of the *Legionella* Dot/Icm system, has a C-terminal motif consisting of hydrophobic residues flanked by lysine/arginine moieties [149]. Mutagenesis studies indicated that deletion of the distal 3 terminal amino acid residues of RalF leads to loss of recognition by the T4SS machinery, and thus loss of translocation [149]. C-terminal translocation signals in T4SS substrates also have been detected in *A. tumefaciens* and *Bartonella henselae* [150,151]. In a similar manner, we propose that the conserved C-terminal motifs in CagH, CagI, and CagL target these proteins to appropriate sites within the *H. pylori* T4SS. Interestingly, a mutant strain expressing a C-terminally truncated form of CagH produced normal-appearing pili, but this strain was nevertheless defective in both CagA translocation and ability to stimulate IL-8 induction. This result reveals that contact of *H. pylori* pili with the AGS cell surface is not sufficient to induce IL-8 production.

T4SSs are a diverse collection of macromolecular machines, found in a broad phylogenetic range of bacteria. It has been proposed that all T4SSs employ a common mechanism for secretion across the cytoplasmic membrane, and that diversity has arisen primarily as a consequence of variation in the types of cell envelopes that must be spanned, along with variation in donor-host cell interactions [70]. Our current

83

understanding of the architecture of T4SSs is based on elegant work with several model systems, including the VirB/VirD4 system of *A. tumefaciens* and several plasmid conjugation systems. There are limitations, however, when comparing these model systems with the T4SSs found in distantly related bacteria. Several components of the *H. pylori cag* T4SS are distantly related to components of the VirB/VirD4 system, and as highlighted in the current study, other key constituents of the *cag* T4SS lack homology to components of T4SSs in other bacterial species. In addition to the three *H. pylori* proteins analyzed in the current study, *cag* T4SS functionality probably requires at least six other proteins that lack homology to VirB/VirD4 proteins [55]. In future studies, it will be important to investigate the functional roles of these proteins in further detail, and from a broader perspective, it will be important to investigate the structural correlates of biological diversity among T4SSs.

CHAPTER IV

FURTHER STUDIES OF THE *HELICOBACTER PYLORI CAG* TYPE IV SECRETION SYSTEM

Introduction

The *cag* pathogenicity island-encoded T4SS of *H. pylori* differs considerably from the prototypical T4SS found in the plant pathogen *Agrobacterium*. The *Agrobacterium* T-DNA transfer system is comprised of 11 essential components (VirB1-VirB11) and one essential accessory protein (VirD4) that is utilized during recognition and delivery of translocation substrate to the T4SS apparatus [69,129]. The *cag* T4SS is distantly related to T4SSs found in many other bacteria, and only a few of the predicted 27 *cag* genes encode components have recognizable sequence similarity to Vir proteins [7,55]. Only the ATPases CagE (VirB4) and VirB11 homolog, the presumed structural components CagX (VirB9) and CagY (VirB10), and the coupling protein Cag5 (VirD4) have obvious homology to Vir components [79,125]. However, the VirB10 homolog CagY is much larger than its counterpart in the *Agrobacterium* T4SS, and contains numerous repeated sequences that are not present in the sequence of VirB10 [96,125,152].

In addition to the use of DNA and protein sequence similarity to identify T4SS protein homologs in the Cag T4SS, protein topology predictions, functional studies, and protein localization studies have also been employed to identify and analyze VirB homologs. Based on these analyses, it is suggested that the Cag T4SS contains homologs to the lytic transglycosylase VirB1 (Cag4), the major pilin VirB2 (CagC), the minor pilus component VirB5 (CagL), and the structural components VirB8 (CagV), VirB7 (CagT), and VirB6 (CagW) [80,94,153,154,155].

85

The current model of assembly of the Cag T4SS is based largely on studies of T4SS assembly in other bacterial species. Recently, structural details of the core complex of a conjugative T4SS have emerged with determination of the crystal structure encoded by the *E. coli* plasmid pKM101 [68,156]. The core complex of this T4SS was determined to contain 14 monomers of each component VirB7 (CagT), VirB9 (CagX), and VirB10 (CagY) [68]. Assembly of this core complex was not dependent upon the presence of additional Vir homologs [68]. Based on the presence of homologs to these Vir components in the Cag T4SS, it is reasonable to presume that assembly of the *H. pylori* T4SS core complex assembles in a manner similar to the pKM101 conjugative T4SS [125].

Assembly of the T4SS core complex in *H. pylori* is likely to occur under basal conditions (i.e., bacteria cultured in the absence of gastric epithelial cells), but it is unlikely that the T4SS would assemble constitutively in the absence of host cell contact. Based on what is observed in *Agrobacterium*, it seems possible that assembly and localization of the *cag* T4SS may occur preferentially at either the bacterial pole or along the lateral surface. In *Agrobacterium*, assembly of the T4SS occurs in a spatial manner, with the polar positioning of the secretion system directed by the action of the nucleating factor VirB8 [157], potentially through the action of the transglycosylase VirB1 [70]. CagV, the VirB8 homolog in *H. pylori*, may contribute to formation or preferential localization of the Cag T4SS in a similar manner [125]. CagV may act as a nucleator by interacting with core components that are required for their proper localization and activity.

Formation of the inner channel of the T4SS is thought to require additional essential components. One component predicted to be required for formation of the secretion system channel which opens to the cytoplasmic space is the VirB6 homolog

CagW [80]. CagW is thought to contain several transmembrane helices, with a presumed localization to the inner membrane [80]. Another essential component of the Cag T4SS is CagU [55]. CagU does not have a homolog in other bacterial T4SSs. It is also predicted to be localized to the inner membrane [125], where it may be involved in formation of the secretion channel. CagH has been previously predicted to be localized to the inner membrane and is speculated to be involved in channel formation [80]. However, we have previously demonstrated surface-exposure of CagH (Chapter II) [158]. This observation does not preclude the possibility that CagH is involved in forming the secretion system channel, and may be localized to several sites within the T4SS apparatus. Functional studies of proteins potentially involved in secretion system channel formation, such as CagU and CagW, have not been performed.

One of the most striking differences that exists between the *H. pylori* Cag T4SS and virulence-associated T4SSs of other bacterial species is the observation that CagA is the only protein effector that is translocated by the *cag* T4SS into the host cell [69]. In contrast, virulence-associated T4SSs of other bacteria are known to translocate multiple protein effectors, or nucleoprotein complexes [69]. Although multiple host cellular consequences can be attributed to interactions of CagA with host proteins [42,43,45,46,47], the mechanism by which *H. pylori* injects the CagA oncoprotein into the cytoplasm of the host cell is largely unknown. Like other T4SS effector proteins [70], CagA contains a C-terminal translocation signal comprised of approximately 20 largely basic residues [122]. In contrast to other T4SS effector proteins, the C-terminal translocation signal of CagA is necessary but not sufficient for translocation [125].

CagA must be directed to the secretion system channel in order to be translocated. There are multiple essential Cag proteins known to be localized to the bacterial cytoplasm, any of which may serve as a CagA translocation signal recognition

protein. One of these proteins, CagF, is believed to be the chaperone-like molecule, based on numerous protein-protein interaction studies [89,122]. Although CagF interacts with approximately 100 residues of in the C-terminal region of CagA, the predicted C-terminal translocation signal of CagA is not encompassed within the CagF-interacting region [122]. This suggests that other Cag proteins required for CagA translocation may be involved in recognizing the C-terminal secretion signal of CagA. Two possible proteins that may fulfill such a role are CagZ and the coupling protein homolog Cag5 [90]. The most likely scenario for recognition of the CagA C-terminal translocation signal involves interaction of the coupling protein Cag5 with CagA, as Cag5 is required for CagA translocation, but is not necessary for stimulation of IL-8 synthesis and secretion by cultured gastric epithelial cells [55,60]. Furthermore, Cag5 has been reported to interact with core components of the T4SS apparatus [81]. Similar to coupling proteins in other T4SSs, Cag5 has ATPase activity and contains several transmembrane helices, with a C-terminal segment that protrudes into the cytoplasm [80]. The positioning of the coupling protein at the base of the T4SS apparatus supports the view that it has a role in controlling T4SS substrate specificity.

In contrast to the process of nucleoprotein complex translocation in the *Agrobacterium* T4SS, the molecular mechanism of CagA translocation through the *H. pylori* Cag T4SS remains largely uncharacterized. Interactions of nucleoprotein complex with structural components of the Vir T4SS have been well defined [159]. Although homologous structural components exist in the *cag* T4SS, the vast differences between Cag proteins and their counterparts in *Agrobacterium*, as well as the presence of many additional proteins in the Cag T4SS that have no homologous components in other T4SSs [7,55], suggests that the CagA translocation process varies considerably from that of nucleoprotein delivery. Many Cag protein interactions with other Cag proteins

that have been identified in yeast 2-hybrid assays [64,81,90] have not been validated in *H. pylori*. Furthermore, it is unclear whether the T4SS pilus structure is utilized as a conduit for CagA and peptidoglycan translocation, or if the pilus is simply required during infection to draw the bacteria closer to the surface of the host epithelial cell prior to release and delivery of CagA. Surprisingly, the composition of the pilus structure itself has not been conclusively determined. Analysis of CagA protein-protein interactions during transport into the host epithelial cell may help to elucidate the architecture and topology of the T4SS apparatus.

This chapter describes the use of chemical cross-linking, immunoaffinity purification, and mass spectrometry strategies to provide further insights into the structural organization of the *cag* T4SS. As one approach, we conducted experiments to identify CagA protein interactions occurring in *H. pylori* in the context of bacteria-host cell contact. We identify several CagA interactions occurring in bacteria in contact with the host cell that differ from interactions of CagA in bacteria that have not been co-cultured with gastric epithelial cells. In another series of experiments, we analyze the suspected inner channel component CagU. The experiments presented in this chapter provide evidence suggesting that CagU may homodimerize during T4SS apparatus assembly, and we identify candidate proteins with which CagU may interact. Finally, we describe experiments designed to elucidate the composition of the *cag* T4SS pilus organelle. The findings presented in this chapter provide the foundation for additional studies directed toward understanding the assembly of the *cag* PAI-encoded T4SS, as well as the molecular mechanism by which the oncoprotein CagA is translocated into the host gastric cell.

Methods

Bacterial strains and growth conditions, AGS cell culture, mutagenesis of *H. pylori,* and bacterial subcellular fractionation were performed as described in the Methods section of Chapter II.

Reversible cross-linking of whole bacteria cells. *H. pylori* were grown for 48 h on blood agar prior to harvesting. Bacteria were washed once in 1X PBS, and whole cells were pelleted by centrifugation at 5,000 rpm for 10 min. Cells were re-suspended in 1 ml PBS containing 0.5 mM Dithiobis[succinimidyl propionate] (DSP). Cells that remained un-cross-linked were mock treated by addition of 50 µl dimethyl sulfoxide (DMSO) [equivalent to the volume of DSP] to the PBS. Cross-linking occurred by incubation at room temperature with constant rotation. At the end of the incubation, cross-linking was quenched by addition of 20 mM Tris, final concentration, for 15 min with constant rotation. Whole cells were pelleted at the end of the quenching reaction by centrifugation at 7,000 x g for 5 min. Cell pellets were lysed by re-suspension in 1 ml RIPA buffer (10 mM Tris, 100 mM NaCl, 1% NP-40, 0.25% deoxycholic acid, Complete Mini EDTA-free Protease Inhibitor (Roche), pH 7.2), and 5 pulses of sonication on ice (10 s each pulse). Complete cell lysis occurred overnight at 4°C in RIPA buffer. Cellular debris was removed by high speed centrifugation. Immunoaffinity purification of Cag protein from bacterial lysate was performed using polyclonal anti-CagA antibodies that had been covalently cross-linked to a Protein G support (Dynal, Invitrogen), or anti-HA monoclonal antibodies (Sigma) immobilized to a Protein G support (Dynal, Invitrogen). Target proteins were immunoaffinity purified at room temperature. Immunoaffinity purified proteins were washed in 100 bed volumes of PBS containing 0.1% Tween-20 (PBST 0.1%). Binding partners of CagA were eluted by incubation of the beads in 50 mM DTT for 1 h at 37°C with constant rotation to reverse the DSP cross-links. Immunoaffinity-purified CagA was eluted from the bead by boiling in 1X SDS buffer (0.3

90

M Tris-HCl, 1% SDS, 10% glycerol, 100 mM DTT, Pierce) for 5 min. Purified CagU-HA complexes were eluted from the anti-HA beads by HA peptide competition (20 μg/ml HA peptide (Sigma), final concentration). In each case, the presence of the targeted protein in the immunoaffinity purified sample was verified by immunoblotting.

Reversible cross-linking of H. pylori attached to gastric epithelial cells. H. pylori was co-cultured with AGS cells (80% confluent) at an MOI of 100 for 5 h prior to removal of unattached bacteria by several washes of the monolayer with 10 ml PBS. Cross-linking of *H. pylori* and AGS cells occurred by addition of 2 ml PBS containing 0.5 mM DSP per flask (2 x 10^7 AGS cells per flask) for 30 min at room temperature. The cross-linking reaction was quenched by addition of Tris (20 mM final concentration) to each flask for 15 min with occasional agitation. The cross-linking solution was removed, and the cross-linked monolayers were washed once in PBS to remove any detached AGS cells.

Differential lysis of eukaryotic cells cross-linked to H. pylori. H. pylori were added to AGS monolayers at an MOI of 100, and were subsequently cross-linked surface of AGS cells, as previously described. The cross-linked monolayers were treated with 0.05% saponin. Monolayers were harvested by mechanical disruption, and eukaryotic cells were lysed occurred at room temperature for 10 min with constant rotation. Cellular debris and *H. pylori* were separated from the soluble eukaryotic fraction by centrifugation at 5,000 x g for 5 min.

Shearing of H. pylori surface structures. H. pylori that had been co-cultured and chemically cross-linked to the surface of AGS cells were pelleted by centrifugation at 5,000 x g for 5 min following differential lysis of the eukaryotic cells. The bacterial pellets were gently re-suspended in 1 ml of 1X PBS. Shearing of *H. pylori* surface structures was performed by passage of the re-suspended bacterial pellets three times through a 25 gauge needle. The bacterial suspension was the centrifuged at 13,000 x g for 5 min to separate the bacterial fraction (pellet) from the surface structure-enriched fraction

(supernatant). Proteins in the sheared fraction were concentrated by precipitation using 20% trichloroacetic acid (TCA). The presence of suspected pilus components was analyzed by immunoblotting of the precipitated protein samples. Proteins present in the sheared fraction were analyzed by MudPIT mass spectrometry.

Purification of CagA from cross-linked H. pylori. Bacterial pellets remaining after shearing of surface structures were lysed by sonication in RIPA buffer (5 pulses each pellet, 10 s each pulse), followed by overnight incubation with constant rotation at 4°C. CagA was immunoaffinity purified from the cross-linked bacteria lysates as previously described in the Methods. Proteins interacting with CagA were eluted by incubation of the beads in 50 mM DTT [160], and CagA was eluted from the beads by boiling in 1X SDS buffer. The presence of CagA and its interacting partner CagF was verified by immunoblot. Additional proteins that interact with CagA were identified by mass spectrometry as described in Chapter II.

Results

Chemical cross-linking and immunoaffinity purification of T4SS complexes containing CagU. CagU, a 25 kDa protein that is required for the translocation of CagA and induction of IL-8 synthesis and secretion [55], has no homologous component in T4SSs of other bacterial species. It is predicted to contain several transmembrane helices, and is thought to be localized to the inner bacterial membrane. We hypothesized that CagU interacts with other components of the *cag* T4SS to form one or more subassemblies that are required for formation of the T4SS apparatus. Therefore, we sought to identify *H. pylori* proteins that co-purify with CagU. To facilitate these studies, CagU was expressed in *H. pylori* with an N-terminal HA epitope tag, as described in the Methods section (Chapter II), using a system in which CagU-HA was expressed from the *ureA* locus in WT 26695. Expression of CagU-HA of the expected size was verified by immunoblot of whole cell lysate using monoclonal anti-HA. A single

band corresponding to the predicted molecular mass of CagU (approximately 25 kDa) was detected in SDS lysate of CagU-HA-expressing bacteria, but not an SDS lysate of WT bacteria (Figure 4-1). Integration into the chromosomal *ureA* locus was verified by PCR amplification of the relevant region from 26695/CagU-HA genomic DNA, followed by sequencing of the PCR product.
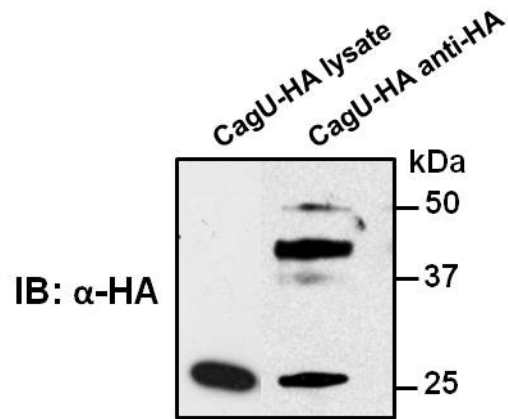
Figure 4-1. Immunoaffinity purification of CagU-HA from whole cell lysate. N-terminally epitope-tagged CagU-HA was expressed from the *ureA* locus in WT *H. pylori*. A single band corresponding to the predicted molecular mass of CagU (25 kDa) can be detected in whole cell lysate derived from bacteria expressing CagU-HA (CagU-HA lysate). Immunoblot analysis of CagU-HA that had been immunoaffinity purified from bacterial lysate by monoclonal anti-HA (CagU-HA anti-HA) contains two distinct bands corresponding to molecular masses of approximately 25 and 45 kDa.

As a next step, CagU-HA was immunoffinity purified using monoclonal anti-HA immobilized to a Protein G support, and eluted from the antibody by HA peptide competition. Immunoaffinity purification of CagU-HA was confirmed by immunoblot (Figure 4-1). In contrast to whole cell lysate in which a single band corresponding to the predicted molecular mass of CagU-HA was detected, two distinct bands were detected by monoclonal anti-HA immunoblot of preparations of purified CagU-HA (Figure 4-1). In addition to the band corresponding to the CagU-HA monomer in whole cell lysate at approximately 25 kDa, a strong band at approximately 45 kDa is present in immunoaffinity purified preparations of CagU-HA. The appearance of this 45 kDa band had not previously been observed in whole cell lysate derived from bacteria expressing CagU-HA. Furthermore, the presence of the 45 kDa band in preparations of affinity-purified CagU-HA purifications was not dependent upon the location of the HA epitope tag. Specifically, a 25 kDa and 45 kDa band were also detected by monoclonal anti-HA immunoblot analysis of purified CagU-HA in experiments using CagU with the HA epitope relocated to the C-terminus (data not shown). It is possible that the 45 kDa band detected in the CagU-HA immunoaffinity purification preparations represents a homo- or heterodimer of CagU-HA that is covalently bonded, or alternatively, it may be a homo- or heterodimer of CagU-HA that is non-covalently associated and resistant to dissociation by SDS. Because immunoaffinity purified CagU-HA was specifically eluted from the anti-HA support by peptide competition, we believe that the 25 kDa band and 45 kDa represent CagU-HA species, rather than mouse IgG heavy chain (50 kDa) or light chain (25 kDa).

In order to further characterize CagU-HA interactions, we employed a chemical cross-linking strategy prior to immunoaffinity purification of CagU-HA. We hypothesized that chemical cross-linking by a cell membrane-permeable cross-linker (DSP) would capture transient or stable interactions of CagU-HA with other proteins that might be

required for formation of the *cag* T4SS apparatus. When lysate derived from bacteria that had been treated with DSP was analyzed, we observed a 25 kDa band with strong signal intensity by monoclonal anti-HA immunoblot, and a band of approximately 45 kDa with much weaker signal intensity (Figure 4-2, lane 2). Immunoaffinity purification of CagU-HA from lysate derived from *H. pylori* cross-linked by DSP treatment recapitulated results of initial CagU-HA purifications, with the presence of two bands corresponding to approximately 25 kDa and 45 kDa (Figure 4-2).
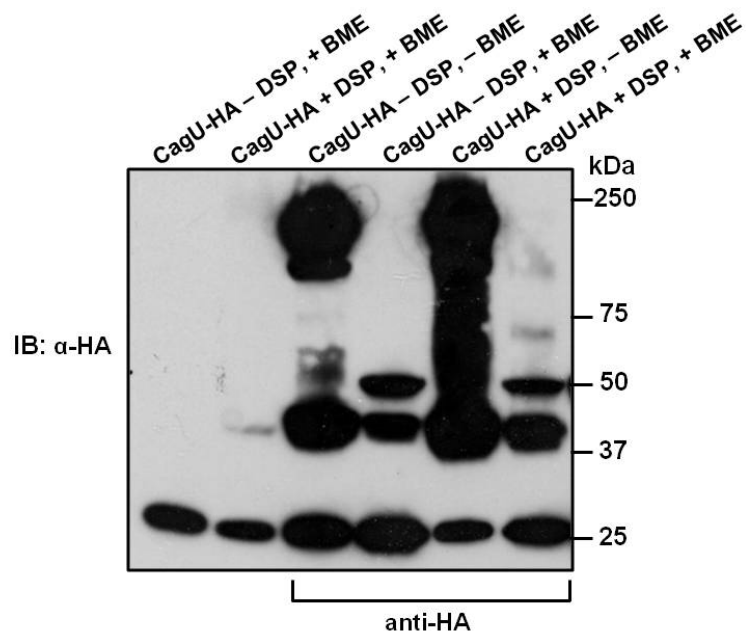
Figure 4-2. Chemical cross-linking strategy to capture CagU-HA interactions. *H. pylori* expressing CagU-HA was cultured for 48 h on solid media. Lanes 1 and 2 of show immunoblot of CagU-HA lysate prior to immunoaffinity purification; Lanes 3-6 show immunoblot analysis of samples in which CagU-HA was immunoaffinity purified from bacteria that had either been mock treated (– DSP) or treated (+ DSP) with the cell-permeable cross-linker DSP. CagU-HA could be immunoaffinity purified from both cross-linked and mock cross-linked bacteria. Addition of the reducing agent BME was able to cleave DSP cross-links. The prominent band at 50 kDa represents the presence of mouse IgG heavy chain.

Anti-HA monoclonal immunoblot analysis of CagU-HA purifications from bacteria treated with DSP revealed a smear of CagU-HA protein complexes under conditions in which the cross-links were not reversed (– BME), indicating efficient cross-linking of CagU-HA to other *H. pylori* proteins (Figure 4-2). Addition of reducing agent to these cross-linked immunoaffinity purified preparations efficiently cleaved the DSP cross-links, as demonstrated by appearance of two prominent bands corresponding to 25 kDa and 45 kDa, and to a lesser extent, a weak band at approximately 70 kDa (Figure 4-2, lane 6). These results suggest that CagU is able to interact with a number of *H. pylori* proteins, and may potentially form homo- or heterodimers. Alternatively, the 45 kDa band present in immunoaffinity purifications of CagU-HA may be represent enrichment of a population of CagU that is post-translationally modified. The 45 kDa band is clearly distinct from that of IgG heavy chain at 50 kDa (Figure 4-2). In future analyses, we will use an HRP-labeled anti-HA monoclonal antibody for immunoblot of CagU-HA purifications in order to definitively distinguish between CagU-HA species and the heavy chain of mouse IgG.

We next sought to identify proteins that interact with CagU-HA, by using mass spectrometry. Immunoaffinity purifications of CagU-HA from bacteria that had not been treated with chemical cross-linker were compared to CagU-HA purifications obtained from bacteria that had been treated with DSP prior to purification. CagU-HA complexes were eluted with HA peptide competition, and each preparation was analyzed by MudPIT (Table 4-1). As expected, we identified the target protein CagU in each of the purification samples. There were a number of striking differences when comparing preparations derived from cross-linked bacteria to preparations from bacteria that had not been cross-linked prior to CagU-HA purification. The proteins present in the preparations from cross-linked bacteria, but absent in preparations from the mock treated bacteria, represent many potential CagU-HA interaction partners (Table 4-1).

The only Cag protein that statistically differed when comparing the cross-linked and non-cross-linked CagU-HA preparations was the presence of the VirB6 homolog CagW in the sample that had been treated with DSP. Of note, the protease encoded by *prc* (HP1350), a non-Cag protein, appears to be associated with CagU-HA in the non-cross-linked form, and to a lesser extent, the DSP treated CagU-HA. Prc is known to be surface exposed, and contains a highly conserved PDZ domain characteristic of C-terminal processing proteases [161,162]. The carboxy-terminal proteases are often involved in post-translational protein processing, maturation of proteins, or disassembly or degradation of proteins in bacteria. The PDZ domain is required for binding of the protease to C-terminal peptides of target proteins, and may also be implicated in binding of the protease to internal recognition sites as well as lipids [161,162,163]. As considered further in the Discussion, we hypothesize that Prc carboxy-terminal protease plays a role in processing of T4SS components.

Table 4-1. Cag proteins that interact with CagU-HA in the presence or absence of chemical cross-links

Gene Number [a]	Protein	anti-HA	
		CagU-HA – DSP [b]	CagU-HA + DSP [c]
HP0531	CagU	15	15
HP0547	CagA	4	9
HP0527	CagY	5	1
HP0529	CagW	0	9 ***
HP1350	Prc Protease	611 ***	66
Total Spectral Counts		1256	1079

[a] Based on the *H. pylori* 26695 genome annotation
[b] Bacteria expressing CagU-HA were cultured in the absence of host cell contact. These bacteria were mock cross-linked, and CagU-HA was immunoaffinity purified using an anti-HA antibody. The Table shows numbers of spectral counts observed by MudPIT analysis for each identified Cag protein and the Prc protease.
[c] Bacteria expressing CagU-HA were cultured in the absence of host cell contact. These bacteria were cross-linked by treatment with DSP, and CagU-HA was immunoaffinity purified using an anti-HA antibody. The Table shows numbers of spectral counts observed by MudPIT analysis for each identified Cag protein and the Prc protease.
*** p<0.001, according to the G-test likelihood ratio, post-spectral count normalization.

Localization of CagU. The mass spectrometry results suggested that CagU-HA specifically interacts with the VirB6 homolog (CagW), based on experiments in which chemical cross-linkers are utilized to capture protein-protein interactions occurring endogenously in *H. pylori.* VirB6 is a known secretion system channel protein in the *Agrobacterium* T4SS apparatus that is localized to the bacterial pole [157,164]. It is therefore reasonable to suspect that CagW has a similar localization within the *H. pylori cag* T4SS. CagW is predicted to encode several transmembrane helices, and computer modeling predicts localization of CagW to the inner bacterial membrane [80]. Based on the apparent interaction of CagU-HA with CagW detected by mass spectrometry, as well as the prediction of transmembrane helices within the CagU sequence [125], we hypothesized that CagU-HA is a membrane-associated protein. In order to test this hypothesis, we fractionated *H. pylori* grown on solid media into a total membrane fraction and a soluble fraction (representing total cytoplasmic and periplasmic proteins). We analyzed each protein fraction by immunoblot. Immunoblot analysis by anti-HA indicates that CagU-HA is localized to the total membrane fraction (Figure 4-3). We were unable to localize CagW to a particular sub-cellular fraction due to lack of antibodies against this protein.
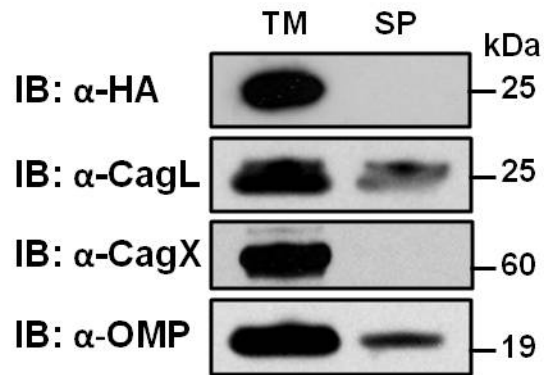
Figure 4-3. Localization of CagU. Bacteria expressing epitope-tagged CagU (CagU-HA) were fractionated into total membrane (TM, total membrane) and soluble fractions containing the cytoplasmic and periplasmic proteins (SP, soluble proteins) as described in Methods.   CagU-HA was detected by immunoblotting (α-HA).   As a fractionation control, fractions were immunoblotted with a monoclonal antibody that recognizes an *H. pylori* outer membrane protein (α-OMP) (Santa Cruz). CagL, a protein known to be localized to multiple subcellular fractions, and CagX, a protein known to be localized to the bacterial membrane, were utilized as additional fractionation controls.

CagU has no known homolog in other bacterial secretion systems. However, we speculate that CagU may be involved in formation of the *H. pylori* T4SS channel via interactions with the VirB6 homolog CagW. As considered further in the Discussion, CagU may serve as a functional homolog to IcmQ, a channel forming component of the *Legionella pneumophila* T4SS.

Analysis of CagA interactions during type IV secretion. The molecular mechanism by which CagA is shuttled into the host gastric epithelium via the *cag* T4SS is virtually unknown. We hypothesized that CagA must interact with multiple proteins in order to be translocated into the host cell, and that such protein interactions would occur sequentially as CagA is first recognized, directed to the T4SS channel, and finally translocated through the T4SS pilus into the host cell.

In order to analyze CagA interactions during type IV secretion, we employed a chemical cross-linking strategy to capture both stable and transient CagA interactions in bacteria that were harvested from solid media, or *H. pylori* that had been co-cultured with gastric epithelial cells [160]. *H. pylori* that had been in contact with gastric epithelial cells for a period of 5 h (a period of time sufficient for the T4SS to form and translocate CagA) were chemically cross-linked by DSP as described in the Methods. Subsequent to cross-linking, the gastric epithelial monolayer and attached bacteria were harvested by mechanical disruption, and the eukaryotic cells were differentially lysed by addition of saponin (a detergent that is insufficient to lyse bacterial cells). The bacterial pellet remaining after lysis and removal of the soluble eukaryotic fraction was sonicated and lysed overnight in RIPA buffer. Anti-CagA polyclonal sera was covalently immobilized to a Protein G support, and utilized to immunoaffinity purify CagA from the cross-linked bacterial cell lysate. Subsequent to CagA purification, proteins that interact with CagA were specifically eluted from the CagA protein by addition of reducing agent (DTT) to reverse the DSP cross-links. In this strategy, the target protein (CagA) remains

associated with the antibody-bead support, while interacting proteins are eluted and can be identified by mass spectrometry. Elution of CagA from the antibody-bead subsequent to DTT elution of interacting partners was accomplished by boiling in SDS buffer.

We were able to successfully immunoaffinity purify CagA from *H. pylori* grown on solid medium, as well as from WT *H. pylori* and several mutant strains that had been co-cultured with gastric epithelial cells prior to chemical cross-linking. We purified similar quantities of CagA from a non-piliated strain (Δ*cagL*), a hyperpiliated strain (Δ*cagH*), and a complemented strain with WT phenotype (CagH-HA) that had each been co-cultured with AGS cells for 5 h. Therefore, our ability to purify CagA from *H. pylori* attached to gastric epithelial cells was not dependent upon formation of T4SS pili. Our ability to purify CagA was also not impaired by introduction of covalent chemical cross-links in *H. pylori* via treatment with DSP (Figure 4-4).
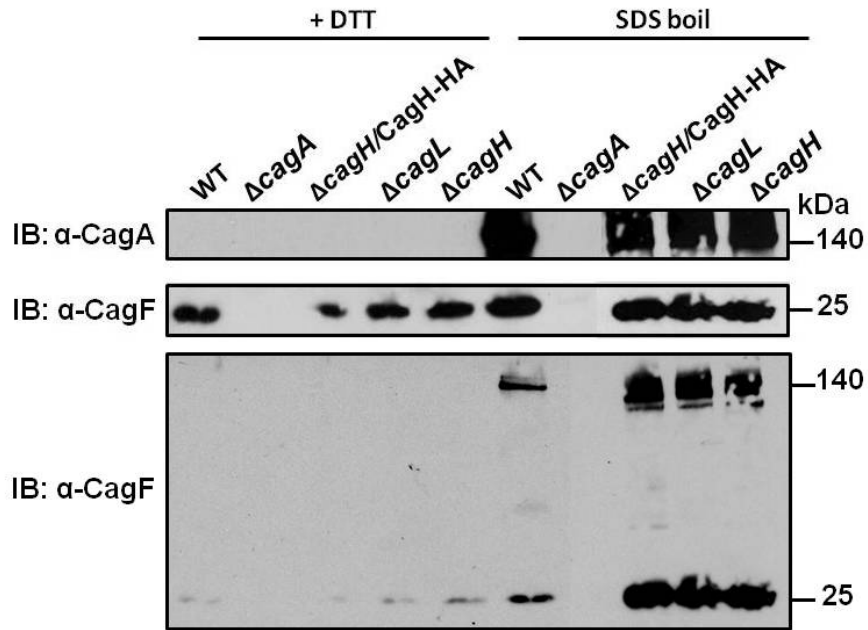
Figure 4-4. Purification of CagA and interacting proteins from *H. pylori*. *H. pylori* was grown on solid media (WT and Δ*cagA*) or cultured with gastric epithelial cells for 5 h (Δ*cagH*/CagH-HA, Δ*cagL*, Δ*cagH*) prior to cross-linking with the membrane permeable cross-linker DSP. CagA was immunoaffinity purified from cross-linked *H. pylori* lysates, and interacting proteins were specifically eluted by cleavage of DSP cross-links by incubation in 50 mM DTT (left portion of immunoblot) [160]. The known CagA interacting partner CagF could be detected in DTT elutions from plate-grown *H. pylori,* as well as various *H. pylori* strains in contact with gastric epithelial cells. CagF was not eluted from mock purifications of DSP cross-linked lysate derived from a Δ*cagA* isogenic mutant control strain. CagA was immunoffinity purified in similar levels from each strain, and could be recovered from the anti-CagA matrix by boiling in SDS (right portion of immunoblot). Cleavage of DSP cross-links by DTT was not completely efficient, as demonstrated by anti-CagF immunoblot of SDS elutions from the anti-CagA matrix (lower anti-CagF panel). CagA-CagF complexes can be detected in the immunoblot region corresponding to approximately 140-250 kDa.

As a control, isogenic mutant bacteria deficient in CagA (Δ*cagA*) that had been grown on solid medium were treated with DSP, and anti-CagA was used to mock-purify bacterial lysate derived from this strain. CagF, a well-characterized protein that is known to interact with the C-terminal region of CagA, was found to co-purify with CagA from WT, Δ*cagH*, and Δ*cagL* lysates (Figure 4-4), but not a mock purification of Δ*cagA* lysate. Cleavage of DSP cross-links by DTT was not completely efficient, as demonstrated by anti-CagF immunoblot of CagA elutions, which reveal residual CagF that remained bound to CagA, or non-specifically to the Protein G bead support (band corresponding to approximately 140-250 kDa, lower panel of Figure 4-4).

Proteins eluted from immunoaffinity-purified CagA by DTT cleavage of DSP cross-links were analyzed by mass spectrometry. Preparations from *H. pylori* that had not been in contact with gastric epithelial cells were compared to preparations from bacteria that had been cultured with gastric epithelial cells for 5 h prior to cross-linking and CagA purification. In both cases, only a few Cag proteins were identified in addition to CagA (Table 4-2). As expected, the CagA chaperone-like molecule CagF was identified by mass spectrometry. CagZ, a cytoplasmic protein thought to stabilize the coupling protein Cag5 [90], and CagY, a presumed structural component of the T4SS apparatus [92], were also identified in anti-CagA purifications by mass spectrometry. There were no significant differences in the Cag proteins immunoaffinity purified by anti-CagA from *H. pylori* cultured either in the presence or absence of host cell contact (Table 4-2).

Table 4-2. Cag proteins purified by anti-CagA from *H. pylori* in the presence or absence of host cell contact

		anti-CagA	
Gene Number [a]	Protein	WT + DSP [b]	CagH-HA + DSP [c]
HP0547	CagA	214	175
HP0522	CagF	13	9
HP0526	CagZ	4	3
HP0527	CagY	4	3
Total Bacterial Protein Spectral Counts		4515	3830

[a] Based on the *H. pylori* 26695 genome annotation

[b] CagA was immunoaffinity purified from bacteria that had been grown on solid media prior to DSP cross-linking. DTT was utilized to reverse DSP chemical cross-links in order to elute proteins from the anti-CagA support. Polyclonal anti-CagA was utilized for CagA purification. The Table shows numbers of spectral counts observed by MudPIT analysis for each identified Cag protein.

[c] CagA was immunoaffinity purified from CagH-HA bacteria (WT phenotype) that had been co-cultured with AGS cells for 5 h prior to chemically cross-linking by DSP. DTT was utilized to reverse DSP chemical cross-links in order to elute proteins from the anti-CagA support. Polyclonal anti-CagA utilized for CagA purification. The Table shows numbers of spectral counts observed by MudPIT analysis for each identified Cag protein.

In addition to a few Cag proteins, many other proteins were co-immunoaffinity purified with CagA from *H. pylori*. Numerous outer membrane proteins (OMP) were detected in significantly greater abundance in preparations from *H. pylori* co-cultured with gastric epithelial cells than in *H. pylori* that had not been in contact with host cells (Table 4-3). For example, significantly more HopQ was co-purified with CagA purified from chemically cross-linked *H. pylori* in contact with gastric epithelial cells than from cross-linked *H. pylori* alone. In addition, more iron-regulated outer membrane protein FrpB and putative lipoprotein HP0596 were co-purified with CagA from *H. pylori* that had been in contact with gastric cells compared to the plate-cultured bacteria control (Table 4-3). Taken together, these results suggest that CagA is able to differentially interact with a number of proteins depending upon the presence or absence of host cell contact. Furthermore, these results indicate that upon contact with the host gastric cell, CagA may become associated with one or several outer membrane protein complexes, which may also contain important bacterial outer membrane proteins and adhesins.

Table 4-3. Selected outer membrane proteins purified by anti-CagA from *H. pylori* in the presence or absence of host-cell contact

		anti-CagA	
Gene Number [a]	Protein (gene name)	WT + DSP [b]	CagH-HA + AGS + DSP [c]
HP1512	FrpB	0	13[***]
HP0913	HopB (*omp21*)	0	5[**]
HP0912	HopC (*omp20*)	0	17[***]
HP0025	HopD (*omp2*)	0	25[***]
HP0706	HopE (*omp15*)	0	6[**]
HP0227	HopM (*omp5*)	0	8[***]
HP1177	HopQ (*omp27*)	0	15[***]
HP0317	HopU (*omp9*)	0	22[***]
HP0596	Lipoprotein	1	22[***]
HP1456	Lipoprotein	0	12[***]
Total Bacterial Protein Spectral Counts		4254	3421

[a] Based on the *H. pylori* 26695 genome annotation

[b] CagA was immunoaffinity purified from bacteria that had been grown on solid media prior to DSP cross-linking. Polyclonal anti-CagA utilized for CagA purification. DTT was utilized to reverse DSP chemical cross-links in order to elute proteins from the anti-CagA support.   The Table shows numbers of spectral counts observed by MudPIT analysis for selected outer membrane proteins.

[c] CagA was immunoaffinity purified from CagH-HA bacteria (WT phenotype) that had been co-cultured with AGS cells for 5 h prior to chemically cross-linking by DSP. Polyclonal anti-CagA utilized for CagA purification. DTT was utilized to reverse DSP chemical cross-links in order to elute proteins from the anti-CagA support.   The Table shows numbers of spectral counts observed by MudPIT analysis for selected outer membrane proteins.

, $p < 0.01$,* $p < 0.001$, according to the G-test likelihood ratio, post-spectral count normalization.

Shearing of *H. pylori* surface structures. Upon contact with the host cell, *H. pylori* may alter its surface properties to facilitate adhesion and optimal interactions with host cells. One important surface structure that forms in *cag* PAI-positive *H. pylori* upon contact with host gastric epithelial cells is the T4SS pilus [64,94,95]. The composition of these pilus structures is largely unknown. It is thought that *H. pylori* T4SS pilus structures may contain several Cag proteins, including the VirB2 homolog CagC [94], the VirB10 homolog CagY [92], and the VirB5 functional homolog CagL [64]. In addition, the effector protein CagA is thought to be associated with the pilus structure as it is translocated into the host gastric epithelial cell [64]. We hypothesized that chemical cross-linking of *H. pylori* to the surface of gastric epithelial cells would preserve fragile surface features such as the T4SS pilus, and that such cross-linked surface structures could be mechanically sheared from the surface of the bacteria after the gastric epithelial cells had been removed through a differential lysis procedure.

In an attempt to enrich for T4SS pili and other surface features of *H. pylori* formed in response to attachment to gastric epithelial cells, we stimulated formation of surface structures by co-culture of several bacterial strains with gastric epithelial cells. Included in the analysis were a non-piliated Δ*cagL* strain, a strain producing many non-functional T4SS pili (Δ*cagH*), and Δ*cagH*/CagH-HA, which produces functional T4SS pili at WT levels. We utilized DSP to cross-link the *H. pylori* strains to the host cell after a period of 5 h as described in the Methods. This period of time is sufficient to induce formation of T4SS pili, and in the case of Δ*cagH*/CagH-HA, sufficient for CagA translocation and IL-8 induction. Surface structures were mechanically sheared from the surface of the bacteria after removal of the host cell monolayer by selective lysis in saponin.

We first analyzed each sheared fraction for the presence of suspected pilus components by immunoblot. We were able to detect appreciable amounts of CagA in

the sheared fractions from CagH-HA and the hyperpiliated Δ*cagH*. Markedly reduced levels of CagA were detected in sheared fractions from the non-piliated Δ*cagL* strain (Figure 4-5), even though Δ*cagL* bacteria express WT levels of CagA. CagL, a suspected pilus component [64], was also detectable by immunoblot in sheared preparations derived from the hyperpiliated Δ*cagH* strain (Figure 4-5). However, CagL could not be detected by immunoblot analysis of sheared preparations derived from CagH-HA bacteria that produce functional T4SS pili at WT levels. We were unable to detect the presence of CagY in sheared fractions from *H. pylori* co-cultured with gastric epithelial cells (data not shown). These findings could be reproduced in multiple biological replicates. Because we could detect differences in the levels of specific Cag proteins in sheared preparations derived from different bacterial mutant strains, we sought to investigate the composition of these sheared fractions by mass spectrometry. Because sheared fractions derived from the hyperpiliated Δ*cagH* strain appeared to contain higher levels of CagA and CagL proteins as determined by immunoblot analysis, we selected this sheared fraction for further analysis by mass spectrometry (Table 4-4).
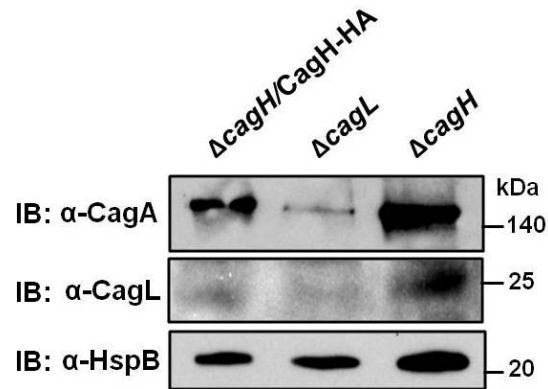
Figure 4-5. Analysis of sheared fractions from several *H. pylori* strains. Surface features sheared from *H. pylori* after 5 h of host cell contact and chemical cross-linking to preserve structural integrity were analyzed by immunoblot. CagA was readily detected in sheared fractions derived from Δ*cagH*/CagH-HA bacteria (which have a functional T4SS with pili that are the same size and abundance as those of WT *H. pylori*) as well as bacteria that produce abundant T4SS pili that are nonetheless non-functional (Δ*cagH*). CagL could also be detected in sheared fractions derived from the hyperpiliated Δ*cagH* strain. Sheared preparations derived from Δ*cagL*, which does not produce T4SS pili, contained markedly reduced levels of CagA.

Table 4-4. Cag proteins identified in sheared fractions of a hyperpiliated *H. pylori* strain

Gene Number [a]	Protein	Δ*cagH* Sheared Fraction [b]
HP0520	Cag1	3
HP0543	CagF	5
HP0545	CagD	16
HP0547	CagA	21
Total Bacterial Protein Spectral Counts		6437

[a] Based on the *H. pylori* 26695 genome annotation

[b] Δ*cagH* was co-cultured with gastric epithelial cells for 5 h prior to cross-linking by DSP to preserve morphology of structural surface features. Bacteria were pelleted after lysis of eukaryotic cells by saponin, and surface features were removed from *H. pylori* by mechanical shearing. The Table shows numbers of spectral counts observed by MudPIT analysis of identified Cag proteins.

As expected, CagA was identified by mass spectrometry as a component of the sheared fraction derived from Δ*cagH*. In addition to CagA, CagD was also identified as a component of the sheared preparation (Table 4-4). CagD has been characterized as being membrane-associated, and is thought to be present on the surface of *H. pylori* [85]. Two additional Cag proteins, CagF and Cag1, were identified in lesser quantities. These two Cag proteins are hypothesized to be localized to the inner bacterial membrane or cytoplasm [78,89,122], and may be present in low abundance in the sheared fractions as a result of bacterial lysis during the mechanical shearing process.

Our ability to detect additional Cag proteins in the sheared fraction of *H. pylori* by mass spectrometry is possibly limited by the overwhelming signal from gastric proteins present in the preparations. It is likely that gastric membrane components are covalently cross-linked to the surface of *H. pylori* during DSP treatment, and that such membrane fragments are also sheared from the surface of *H. pylori*. Although a differential lysis procedure is utilized in an attempt to enrich for bacterial cells from co-culture [43], residual gastric proteins that remain cross-linked to the surface of *H. pylori* are much more abundant than *H. pylori* proteins in bacterial sheared fractions, as indicated by mass spectrometry analysis (11,103 spectral counts representing human gastric proteins versus 6437 total spectral counts representing *H. pylori* proteins).

In addition to a few Cag proteins, a number of non-Cag proteins were identified in the Δ*cagH* sheared fraction by mass spectrometry. Several outer membrane proteins, as well as multiple membrane-associated lipoproteins were present in the Δ*cagH* sheared fraction. We also identified several proteins known to be localized to the surface of *H. pylori*, including the major virulence factor VacA, and surface-associated carboxy-terminal protease Prc. Taken together with immunoblot analysis, the mass spectrometry results indicate that we are able to shear a number of proteins from the surface of *H. pylori* after the bacteria have been in contact with gastric epithelial cells.

However, it is difficult to interpret the significance of proteins identified in *H. pylori* sheared preparations by mass spectrometry due to lack of an appropriate control. In future investigations, it will be important to optimize shearing protocols in order to increase our ability to detect low abundant bacterial proteins in sheared preparations, and distinguish between proteins that are surface-exposed and those that are present in sheared preparations as a result of non-specific lysis of *H. pylori* during the shearing process.

Discussion

The *H. pylori* T4SS encoded by the *cag* pathogenicity island is markedly different from T4SSs found in other bacterial species. Although nine components of the *cag* T4SS have a low degree of sequence similarity to constituents of the T4SS found in *Agrobacterium* [7,55,125], the function and localization of these homologs within the *cag* T4SS remain largely uncharacterized. Furthermore, the functions of many components unique to the *H. pylori* T4SS are virtually unknown. Major deviations of the *cag* T4SS from the prototypical *Agrobacterium* T4SS are likely a result of strong evolutionary pressures that have guided divergence and adaptation of the *H. pylori* T4SS system to the human host [128]. Furthermore, *H. pylori* has refined its T4SS over the course of many thousands of years of intimate association with humans to facilitate efficient translocation of CagA, and possibly other, effector molecules into the host cell [6,7,8]. It is reasonable to assume that incorporation of multiple unique, essential components into the *cag* T4SS, along with alteration of homologous components, has occurred in order to effectively subvert the host defense system during chronic infection. The structure of the *cag* T4SS has ultimately evolved to translocate effector cargo into the host cell in a specific, well-defined manner.

CagU, a component of the Cag T4SS that is essential for translocation of CagA and induction of IL-8 synthesis and secretion by cultured gastric epithelial cells [55], is unique to *H. pylori*. Thus far, the function of CagU is not understood. It is predicted to be an inner membrane protein that may be a structural component of the T4SS apparatus [125]. We hypothesized that CagU interacts with other components of the *cag* T4SS to form one or more subassemblies. Using immunoaffinity purification methods to purify epitope-tagged CagU from *H. pylori* lysate and subsequent analysis of purified proteins by mass spectrometry, we were able to detect multiple candidate interactions of CagU with other Cag proteins (Figure 4-1, Table 4-1). Chemical cross-linking strategies were also utilized to detect probable interactions of CagU with the VirB6 homolog CagW (Figure 4-3, Table 4-1). CagW is thought to serve a role similar to VirB6 as a constituent of the inner channel of the T4SS apparatus [80]. However, the role of CagW in the *cag* T4SS has not been elucidated. Both CagU and CagW are predicted to contain several transmembrane helices, suggesting localization to the bacterial membrane [80]. We demonstrate membrane localization of CagU to the total bacterial membrane fraction in a detergent-free lysate of *H. pylori*, providing the first direct evidence that CagU is localized to the membrane (Figure 4-3).

Interestingly, when CagU-HA was purified from *H. pylori* in the absence of chemical cross-linkers, we recovered an abundance of the surface-exposed carboxy-terminal protease encoded by *prc* (HP1350) (Table 4-1). In the presence of chemical cross-linker, we recover 10-fold less Prc compared to the CagU-HA purifications in which chemical cross-linker has been omitted. Co-purification of Prc with CagU-HA is reproducible (data not shown), and the relative abundance of Prc by spectral count is astoundingly high in CagU-HA purifications. Prc accounts for approximately 49% of the total protein content of CagU-HA purifications from un-cross-linked bacteria, versus approximately 6% of the total protein content of CagU-HA purifications from bacteria

treated with chemical cross-linker (Table 4-1). The significance of the observed association of this carboxy-terminal protease with CagU-HA is unclear. The high abundance of the protease in association with un-cross-linked CagU may indicate the presence of oligomers of Prc in complex with CagU-HA. The protein coverage (i.e., the proportion of protein represented by detected peptides) of Prc that co-purifies with CagU-HA is similar in the presence (44%) or absence (40%) of cross-linker. Prc may be required for post-translational processing of CagU prior to incorporation of CagU into the T4SS apparatus. It is also possible that association of CagU-HA with Prc occurs on the surface of the cell, as Prc is known to be surface-exposed. Furthermore, co-purification of Prc with other Cag proteins, including the surface-exposed pilus component CagL, has been detected (Table 2-4) [158]. It will therefore be important to elucidate the potential role that Prc may have in processing CagU or other bacterial or host proteins, and whether this carboxy-terminal protease plays a role in formation and function of the *H. pylori* T4SS.

CagU shares some similarities to the IcmQ component of the T4SS in *Legionella*. A BLAST search query of the CagU sequence against the *Legionella* proteome identifies weak sequence similarity to IcmQ, the only *Legionella* T4SS component with detectable similarity to CagU. CagU may serve a similar role to IcmQ in formation of the *H. pylori* T4SS. In the *Legionella* T4SS, IcmQ, a 22 kDa protein, is involved in formation of the T4SS channel [165]. IcmQ is essential for T4SS function, intracellular growth, high multiplicity of infection, and proper intracellular trafficking during *Legionella* infection. IcmQ is prevented from forming the secretion system channel pore by association with its chaperone IcmR [166]. In the absence of IcmR, IcmQ forms high molecular weight complexes [167], a phenotype that is reminiscent of the behavior of T3SS substrates in the absence of their corresponding chaperone. In addition, the IcmQ chaperone IcmR forms SDS-resistant dimers [165]. Unlike CagU, which is associated with the bacterial

membrane, IcmQ and IcmR are soluble proteins [166]. However, in spite of its soluble subcellular localization, IcmQ efficiently inserts into lipid membranes during formation of the *Legionella* T4SS apparatus [166] in a manner similar to soluble membrane-active toxins such as *Staphylococcus aureus* α-toxin [168]. IcmQ also has similar pore-forming activity to *Shigella flexneri* IpaB [169] and the *Agrobacterium* VirE2 protein [170], which are both translocated into the host cell. In addition, IcmQ can be detected on the surface of *Legionella* during bacterial contact with the host cell [166]. It is proposed that during infection, a portion of the available IcmQ is translocated to the outer part of the translocation machinery, and that regulation of the pore-forming capacity of IcmQ by association with IcmR could be involved in controlling the opening or closing of the T4SS channel when in contact with the host cell [166].

Based on these observations, we speculate that CagU may be involved in formation of the *H. pylori* T4SS channel. The appearance of high molecular weight complexes of CagU-HA during the immunoprecipitation procedure, but not in bacterial lysate, may be the result of disruption of CagU/chaperone interactions. In a manner analogous to the IcmQ/IcmR interaction, a chaperone may bind CagU and prevent formation of high molecular weight complexes. In our studies, it is possible that CagU is interacting with a chaperone molecule similar to that of IcmR, and forming SDS-resistant heterodimers that we are able to detect by immunoblot. Another possibility is that the immunoprecipitation procedure brings CagU-HA monomers within close proximity, permitting formation of CagU-HA complexes in a manner similar to what may occur during formation of the T4SS channel. These oligomeric CagU complexes may fold or associate in such a way that the complexes are resistant to denaturation by SDS. It is also possible that CagU is post-translationally modified (for example, by addition of N-linked glycans), which could also result in drastic shifts in migration patterns during SDS-PAGE resolution of CagU-HA. It will be important in future investigations to discern

whether CagU is post-translationally modified or forms multimeric complexes during T4SS apparatus assembly.

Chemical cross-linking strategies to capture stable or transient interactions of Cag proteins have been used in only a limited numbers of studies. The use of the reversible, membrane-permeable chemical cross-linker DSP was successfully utilized to detect candidate interactions of both CagU (Figure 4-3, Table 4-1) and CagA (Figure 4-4, Table 4-2) with other Cag proteins. The use of DSP cross-linking strategies and purification of CagA complexes was of particular interest. We were able to cross-link CagA complexes in *H. pylori* harvested from solid media, as well as *H. pylori* that had been co-cultured with gastric epithelial cells (Figure 4-4). Surprisingly, CagA interactions captured by chemical cross-linking did not differ when comparing bacteria grown in the presence or absence of host cells (Table 4-2). In contrast to *H. pylori* grown on solid media, CagA was associated with a number of outer membrane proteins in response to contact with host gastric cells (Table 4-3). This result suggests that the localization of CagA changes in response to host cell contact, during the process of translocation into the gastric epithelial cell. The association of CagA with multiple outer membrane proteins in response to contact with the host cell may also be the result of association of CagA with T4SS pilus structures that are anchored to the outer membrane. These results suggest that CagA is associated with one or several complexes on the surface of *H. pylori* during interaction with the host cell.

In experiments performed under conditions in which *H. pylori* were attached to gastric epithelial cells, the preparations of purified CagA contained a substantial amount of human components. These human components were probably present in CagA purifications as a result of introduction of chemical cross-links between the surface of the bacteria and the surface of the gastric cells. It is likely that the high signal intensity from the abundant human proteins in the CagA purification samples greatly diminish the

119

signal intensity generated by the low-abundant *H. pylori* proteins. Eukaryotic cell lysis and CagA purification methods will require optimization in future studies in order to increase the signal-to-noise ratio of low-abundant Cag proteins for detection by mass spectrometry.

Another potentially valuable approach for identifying surface features of *H. pylori* that are generated in response to bacterial contact with gastric epithelial cells is the use of chemical cross-linking to preserve surface structure morphology, prior to mechanical shearing to remove the structures from the surface of the bacteria. One important structure that is formed only in response to host cell contact in *cag* PAI-positive *H. pylori* strains is the T4SS pilus [64,92,95]. The pilus structure is required for CagA translocation [64], and we have shown that WT pili are necessary, but not sufficient, for induction of IL-8 synthesis and secretion [158].

We were able to identify a few Cag proteins by mass spectrometry analysis of sheared fractions derived from a hyperpiliated Δ*cagH* strain of *H. pylori* that had been in contact with gastric cells (Table 4-4). Furthermore, we were able to detect the presence of CagA and CagL in sheared fractions of Δ*cagH* bacteria or the complemented CagH-HA, but not a non-piliated Δ*cagL* strain (Figure 4-5). In addition to CagA, mass spectrometry analysis of Δ*cagH* sheared fractions indicated the presence of CagD, and to a lesser extent, Cag1 and the CagA chaperone-like molecule CagF. It is likely that the presence of small quantities of CagF in the sheared fraction is the result of non-specific bacterial lysis during the shearing protocol. Interestingly, Cag3, a protein known to be localized to multiple subcellular sites including the surface of the bacteria, was found to interact with the accessory proteins CagD and Cag1 when immunoaffinity purified from *H. pylori* lysate [78]. The requirement for CagD in the process of CagA translocation and induction of IL-8 is unclear [55,85]. However, CagD is thought to be a component of both periplasmic and outer membrane T4SS subassemblies based on

protein localization studies and analysis of CagD surface-exposure by immunofluorescence microscopy [85].

In addition to the presence of a few Cag proteins in sheared fractions from Δ*cagH* bacteria, several known outer membrane proteins were identified in the fractions by mass spectrometry. The sheared fraction also contained the important surface-exposed toxin VacA, multiple outer membrane proteins, and the surface-exposed carboxy-terminal protease encoded by *prc*. Taken together with immunoblot analysis of sheared fractions obtained from several strains, these results suggest that the composition of the surface of *H. pylori* is complex, and the surface of the bacteria varies depending on the presence or absence of a functional T4SS.

As with the case of CagA purifications from *H. pylori* that had been cross-linked to the surface of gastric cells, sheared fractions derived from Δ*cagH* bacteria contained a high proportion of human gastric proteins. The presence of the gastric proteins in the sheared fractions may have diminished our capability to detect the low abundant *H. pylori* proteins in the sheared preparation. Optimization of the shearing protocol and enrichment of surface proteins from the sheared fraction will be imperative for future studies of *H. pylori* proteins exposed to the extracellular milieu during infection and in response to host cell contact.

The studies presented in this chapter provide the foundation for additional investigations of Cag protein-protein interactions. Despite nearly two decades of work, significant knowledge gaps exist in multiple facets of type IV secretion processes that occur in *H. pylori* and other pathogenic bacteria. The methods employed in this Chapter may be applied to future investigations of Cag interactions during assembly and function of the *H. pylori* T4SS, and such interaction studies may provide the basis for functional examination that will shed light on the molecular mechanisms of CagA translocation into the host cell. Because infection by *H. pylori* that harbors a functional T4SS can lead to

development of severe gastric disease, it is imperative that we understand the process by which CagA is introduced into the host cell so that we may then formulate strategies to preclude its adverse cellular effects.

CHAPTER V

CONCLUSIONS

Summary

The type IV secretion system (T4SS) encoded by the *cag* pathogenicity island (*cag* PAI) is a major virulence determinant of *H. pylori* [42,45,46,97], and it is present in approximately 50% of all *H. pylori* strains [35]. While the T4SSs of other bacterial species have been characterized in detail, significant gaps exist in our understanding of the architecture and topology of the *cag* T4SS. It is unclear how the T4SS apparatus assembles prior to translocation of CagA into gastric epithelial cells. Moreover, the molecular mechanism by which CagA is translocated into the host cell has not been elucidated. The functions of many Cag proteins, especially those without homologs in other bacterial T4SSs, remain largely unexplored.

In my thesis, I have established methods for enriching Cag protein complexes and identifying candidate protein-protein interactions by mass spectrometry. Furthermore, I have identified a number of Cag proteins that lack homologs in other T4SSs and are involved in biogenesis of the T4SS pilus structure at the bacteria-host cell interface. Additionally, I have provided preliminary evidence that chemical cross-linking and immunoaffinity purification strategies can be utilized to identify protein-protein interactions occurring between unique *H. pylori* Cag proteins and Cag proteins that are homologous to components of the prototypical T4SS found in *Agrobacterium*.

In Chapter II, we utilized immunoaffintiy purification and subsequent mass spectrometry analysis of purified proteins to identify a T4SS subassembly that contains the T4SS pilus component CagL and two other previously uncharacterized proteins, CagH and CagI. The methods presented in Chapter II established a powerful new tool

for screening Cag protein-protein interactions. Understanding Cag protein interactions in the presence and absence of host cell contact may provide important clues as to how the Cag T4SS assembles. Furthermore, developing a robust methodology for high-throughput screening of potential Cag protein-protein interactions is instrumental in dissecting the architecture of the T4SS apparatus.

We provided evidence that CagH, CagI, and CagL are each required for CagA translocation and induction of IL-8 synthesis and secretion by cultured gastric epithelial cells. In addition, these components share multiple properties, including co-transcription and protein sequence relatedness. We demonstrated that formation of the CagHIL complex occurred in both the presence and absence of bacteria-host cell contact. Maintenance of the complex was dependent upon the presence of all three proteins. Furthermore, deletion of either *cagH* or *cagI* resulted in markedly decreased stability of CagL, supporting our hypothesis that these three proteins are constituents of a T4SS subassembly or complex. CagH, CagI, and CagL are all detectable on the surface of *H. pylori* in the absence of host cell contact. We demonstrated that these three proteins may exist in multiple cellular sites, including the periplasm and the bacterial membrane.

In Chapter III, we provided evidence that CagH, CagI, and CagL each play a role in biogenesis of the T4SS pilus structure at the bacteria-host cell interface. We demonstrated that CagL and CagI are required for formation of the T4SS pili, and that formation of T4SS pili was not dependent upon the presence of CagH. However, *H. pylori* that are deficient in *cagH* exhibit a hyperpiliated phenotype, producing T4SS pili that have a dramatically altered morphology compared to those of wild-type (WT) *H. pylori*. To our knowledge, description of a hyperpiliated phenotype has not been described in the literature for T4SSs of any bacterial species.

We identified a conserved flagellar hook protein K (FlgK) domain near the N-terminus of CagH that is evolutionarily-related to the corresponding domain in FlgK of

124

Burkholderia species. This domain may contribute to the role of CagH as a modified capping protein or molecular ruler involved in controlling dimensions of the Cag T4SS pilus. Furthermore, we identified nearly identical hexapeptide C-terminal motifs present in CagH, CagI, and CagL that also contribute to T4SS pilus biogenesis. We showed that this motif in CagI and CagL is required for pilus formation. In contrast, bacteria expressing CagH that lacks this C-terminal motif produce pili with morphology and abundance similar to those of the WT strain. These T4SSs are nonetheless non-functional, as deletion of the CagH C-terminal motif results in ablation of both CagA translocation and induction of IL-8 synthesis and secretion by gastric epithelial cells. We demonstrate that deletion of this motif from CagH or CagI results in partial mislocalization of the truncated proteins to the soluble fraction, rather than the total membrane fraction, of a detergent-free *H. pylori* lysate. Based on the observed mislocalization of C-terminally truncated CagH and CagI proteins, we hypothesize that this C-terminal motif plays a role in protein sorting or localization.

The studies presented in Chapters II and III highlight the marked variation that exists among bacterial T4SSs, and illustrate the important role that is played by proteins that are unique to *H. pylori* in the assembly and function of the *cag* PAI-encoded T4SS. Although the basis for most of the studies of the *H. pylori* T4SS have been elegant studies of the VirB/VirD4 *Agrobacterium* T4SS, there are limitations when comparing this model system with the T4SSs found in distantly related bacteria. Several components of the *H. pylori cag* T4SS are distantly related to components of the VirB/VirD4 system, and as highlighted in Chapters II and III, other key constituents of the *cag* T4SS lack homology to components of T4SSs in other bacterial species. In addition to CagH, CagI, and CagL, *H. pylori* T4SS functionality likely requires at least six other proteins that lack homology to VirB/VirD4 proteins [55]. In future studies, it will be important to investigate the functional roles of these proteins in further detail, and from a broader

perspective, it will be important to investigate the structural correlates of biological diversity among T4SSs.

In Chapter IV, we further developed our methodology for identification of subassemblies or complexes involved in assembly and function of the *cag* PAI-encoded T4SS. We employed a chemical cross-linking strategy prior to immunoffinity purification and identification of proteins by mass spectrometry to analyze interactions of the unique T4SS component CagU. Experiments with epitope-tagged CagU suggest that CagU may homo- or heterodimerize during T4SS apparatus assembly. Analysis of CagU that had been chemically cross-linked to interacting proteins by mass spectrometry reveals a probable interaction of CagU with the VirB6 homolog CagW, which was not identified in the absence of cross-linking. We utilized a similar chemical cross-linking strategy to identify candidate interactions of the effector molecule CagA in both the presence and absence of host cell contact. We identified several CagA interactions occurring in bacteria in contact with the host cell that differed from interactions of CagA in bacteria that had not been co-cultured with gastric epithelial cells.

Finally, we utilized mass spectrometry to analyze sheared preparations, which are predicted to contain surface-exposed structural components of *H. pylori*. CagA was a component of sheared preparations derived from WT and hyperpiliated Δ*cagH* bacteria that had been in contact with gastric epithelial cells, but was not a major constituent of sheared preparations derived from a non-piliated Δ*cagL* mutant that had been co-cultured with gastric epithelial cells. Findings described in Chapter IV provide the foundation for additional studies to elucidate the assembly of the *cag* PAI-encoded T4SS, as well as the molecular mechanism by which the oncoprotein CagA is translocated into the host gastric cell.

Future Directions

Elucidate the mechanism by which dimensions and abundance of T4SS pili are controlled. We have shown that *H. pylori* that are deficient in *cagH* exhibit a hyperpiliated phenotype, and produce T4SS pili that have a drastically altered morphology compared to pili produced by WT bacteria (Figure 3-2). A striking feature of CagH is the presence of a large domain spanning approximately 60 amino acid residues near the N-terminus that shares remarkable conservation to flagellar hook protein K (FlgK) sequences of *Burkholderia* spp. (Figure 3-2H) [158]. FlgK is flagellar hook-associated capping protein that is utilized in flagellar systems to cap polymerization of the flagellar hook in order to signal the switch to filament polymerization during flagella biosynthesis [146,148].

In an attempt to further understand the role, if any, that the conserved FlgK domain of CagH plays in controlling dimensions of the T4SS pilus, a CagH mutant has been generated in which 36 amino acid residues within the FlgK domain were deleted (CagH-HAΔFlgK). The FlgK domain of CagH could not be recognized by the Conserved Domain Database (CDD) within NBCI when these residues were deleted. CagH-HAΔFlgK was expressed in Δ*cagH* bacteria at significantly decreased levels compared to full-length CagH-HA (data not shown), indicating that the stability of CagH is compromised in the absence of residues within the FlgK domain. Potentially, deletion of these residues may cause CagH to be misfolded, resulting in decreased detection of CagHΔCT expression. We will utilize our established scanning electron microscopy analysis to assess whether bacteria expressing CagH-HAΔFlgK cells produce pili with abundance and dimensions similar to those of either WT or Δ*cagH* bacteria. Because of the decreased stability of CagH in the absence of the FlgK domain residues, we hypothesize that bacteria expressing CagH-HAΔFlgK will have a T4SS pilus phenotype that mimics a Δ*cagH* strain.

We will also investigate if CagH is serving as a regulator to signal increased T4SS pilus biogenesis in the presence of host cell contact. During infection, it is probable that *H. pylori* senses changes in the environment such as available iron or pH of the extracellular milieu [103,171,172]. We will investigate changes in *cagH* transcription by RT-PCR in response to a number of environmental conditions, including changes in pH, changes in salt concentration, and changes in iron availability. The level of *cagH* transcription, as well as transcription of *cagI* and *cagL*, in the presence of host cell contact will be established. We will compare *cag* gene transcription levels in the absence of host cells to the level of transcription of the corresponding genes in a co-culture system in which we can control availability of exogenous iron, as well as shifts in pH. It is also possible that CagH is acting as a DNA binding protein that can negatively regulate transcription of genes within the *cagHIL* operon. If this is the case, complementation of the Δ*cagH* mutant at a secondary chromosomal site would result in restoration of WT levels of *cagI* and *cagL* transcription as compared to the nearly 5 fold up-regulation of *cagI* and *cagL* transcription occurring in the Δ*cagH* mutant in the absence of host cell contact (Figure 2-5).

Transcription of *cagH* is known to be induced during iron starvation [172]. In iron-deplete conditions, *H. pylori* is known to have increased virulence in the animal model (J Noto *et al.*, abstract presented at CHRO, Vancouver, Canada 2011). It is reasonable to suspect that *cagI* and *cagL*, which are required for T4SS pilus formation and are encoded by overlapping genes downstream of *cagH*, are negatively regulated by the presence of available iron, potentially through the action of the *H. pylori* ferric uptake regulator (Fur) protein [172]. In other words, when iron is available, the Fur-Fe^{2+} complex may bind a Fur consensus region near the *cagHIL* operon too down-regulate expression of *cagI* and *cagL*, resulting in loss of T4SS pilus biogenesis. However, in the absence of available iron, a condition that is expected to be encountered during infection

of the host, Fur is unable to bind a nearby Fur consensus sequence, resulting in up-regulation of *cagI* and *cagL* transcription, and a subsequent increase in T4SS pilus biogenesis at the bacteria-host cell interface. Recently, it has been shown that CagA translocation is an important event that alters the host cell physiology to allow *H. pylori* to obtain iron during infection [173]. It is therefore plausible that formation of the T4SS apparatus occurs in response to iron deficiency, as formation and functionality of the T4SS are important during the process of iron nutrient acquisition.

Although a comprehensive map of transcription start sites within the *H. pylori* 26695 genome has been generated [121], we will analyze the structure of the *cagHIL* operon in more detail. To do so, we will utilize multiple approaches, including primer extension analysis to map the 5' end of RNA transcripts, and Tobacco Acid Pyrophosphatase (TAP) assay to map the transcription start sites within the *cagHIL* operon [174]. In addition to the typical *cagH*-containing transcript, an anti-sense transcript is also encoded within the same region of chromosomal DNA [121]. This *cagH* antisense transcript may negatively regulate the *cagHIL* operon. Furthermore, *cagI* is predicted to encode an internal transcription start site for *cagL* [121]. These alternate transcription start sites may be involved in regulating levels of gene transcription within the *cagHIL* operon. For example, transcription of the *H. pylori* urease operon produces three main transcripts, two of which are degraded post-transcriptionally in order to generate various transcription products; cleavage of these urease polycistrons occurs in response to shifts in environmental pH [174]. Post-transcriptional control of operons within the *cag* PAI may be occurring in a similar manner. It is also possible that deletion of a Fur consensus region predicted in *cagI* results in loss of regulation of *cagL* transcription by Fur, and may account for the nearly 10 fold increase of *cagL* transcript levels observed in a Δ*cagI* mutant (Figure 2-5) in the absence of host cell contact [158].

Determine the role that non-homologous components play in formation of the T4SS apparatus. CagU is one of approximately nine non-homologous Cag proteins required for CagA translocation and function of the T4SS [55]. We have successfully generated a strain of *H. pylori* that expresses an N-terminal HA epitope-tagged form of CagU from the *ureA* chromosomal locus. In Chapter IV, we demonstrated interaction of this epitope-tagged CagU with other components of the T4SS, including the VirB6 homolog CagW. This result, along with a previous report analyzing a *cagU* mutant [55], suggests that CagU is likely involved in formation of the T4SS apparatus. Construction of Δ*cagU* mutant will be important for future studies.

A Δ*cagU* mutant will be generated using the methods presented in Chapter II, whereby the CagU coding region will be replaced with an in-frame kanamycin resistance cassette that is transcribed in the same direction as *cagU* and the downstream gene *cagV*. Maintenance of *cagV* transcription will be monitored by anti-CagV immunoblot to ensure that insertion of the kanamycin resistance cassette into the *cagU* locus does not have polar effects that disrupt transcription of the essential T4SS gene *cagV*. The Δ*cagU* mutant will then be assayed for loss of T4SS function by analysis of IL-8 induction by gastric epithelial cells co-cultured with Δ*cagU* bacteria, as well as CagA translocation in the same co-culture system. We expect that deletion of *cagU* will ablate CagA translocation as well as the ability of the bacteria to induce secretion of IL-8 [55]. The Δ*cagU* strain will be complemented in a heterologous chromosomal locus by a construct designed to express either N- or C-terminally epitope-tagged CagU-HA [158]. We anticipate that complementation of Δ*cagU* with either of these constructs will restore the T4SS phenotype to that of WT *H. pylori*.

We will repeat CagU-HA immunoaffinity purifications using our previously established protocol and peptide competition elution of CagU-HA complexes from *H. pylori* lysate. CagU-HA purifications will be performed under conditions in which the

T4SS is formed and functional (bacteria-gastric epithelial cell co-culture), or conditions under which the T4SS is assembled at basal levels, or exists in an intermediate state of assembly (bacteria in the absence of host cell contact). We will evaluate purified CagU-HA by immunoblot, and identify co-purified proteins by mass spectrometry. We expect to recapitulate the results presented in Chapter IV with respect to detection of CagU-HA species that can be detected at multiple molecular masses by immunoblot. In order to discern whether CagU is dimerizing during the purification process, we will utilize targeted peptide mass spectrometry (MRM) to identify specific CagU peptides in various molecular mass ranges of an SDS-PAGE gel. This approach may allow us to quantify abundance of CagU existing in the monomer form (at the predicted molecular mass of approximately 25 kDa), for comparison to CagU peptides at a higher molecular mass (of approximately 45 kDa), which could represent a dimerized or complexed form of CagU.

We will also evaluate the possibility that CagU is acquiring post-translational modifications, such as N-linked glycosylations, that could be leading to the apparent size shift to approximately 45 kDa after enrichment by immunoaffinity purification. CagU is predicted to encode several N-linked glycosylation sites (5 glycosylation sites predicted by NetNGlyc server) distributed throughout amino acid residues 1-28. In order to assess whether the shift in molecular mass is due to the presence of an N-linked glycan, we will purify CagU-HA from bacteria alone, or bacteria in co-culture with gastric epithelial cells, and treat the purified CagU-HA with PNGase F to cleave any N-linked glycans from the purified protein. We will monitor any shift in molecular mass of CagU-HA by immunoblot. If treatment of CagU-HA with PNGase F results in loss of the 45 kDa band and appearance of one or several new lower molecular mass bands, we will repeat the CagU-HA purifications and titrate the level of PNGaseF enzyme in order to estimate the number of glycosylation events occurring on CagU. CagU-HA purifications that have been treated with PNGase F will be analyzed in collaboration with investigators in the

Vanderbilt Mass Spectrometry and Proteomics Research Facility. If CagU is glycosylated and present on the surface of the bacteria, it may play a role in adherence to the gastric epithelial cell surface during type IV secretion.

We will also evaluate the localization of CagU-HA in *H. pylori* in more detail in order to ascertain if CagU is localized to the inner or outer bacterial membrane. To do so, we will evaluate CagU-HA susceptibility to cleavage by trypsin and proteinase K, utilize immunofluorescence to detect CagU on the surface of bacteria in co-culture with gastric epithelial cells, and utilize flow cytometry or immunogold labeling to detect surface-exposure of CagU in the absence of host cell contact [158]. CagU is predicted to be an inner membrane protein, but as is the case with CagH, which was also predicted to be localized to the inner bacterial membrane, CagU may exist in multiple subcellular sites, which may include the surface of the bacteria. Additionally, we will utilize our scanning electron microscopy assay to evaluate the ability of Δ*cagU* bacteria to form T4SS pili in response to contact with gastric epithelial cells. If CagU is a channel protein required for assembly and formation of the T4SS core complex, we would expect a Δ*cagU* strain to be deficient in T4SS pilus formation. Furthermore, if CagU is involved in formation of the central channel of the secretion system, other components of the T4SS pilus structure, such as CagL, may not reach the surface of the bacteria. This phenomenon could be assayed by analyzing Cag protein localization to the bacterial surface as determined by flow cytometry, immunogold labeling, or protection from cleavage by proteinase K in Δ*cagU* bacteria.

Mass spectrometry results suggest that CagU potentially interacts with the surface-exposed carboxy-terminal protease encoded by *prc* (HP1350). This protease has also been identified in immunoaffinity purifications of CagL, a T4SS pilus component that is known to be localized to the outer membrane in the absence of host cell contact (Table 2-4). The protease encoded by *prc* is also known to be surface-exposed

[175,176]. We will investigate the significance of the CagU-protease interaction, and the interaction of this protease with other Cag proteins, by evaluating effect of deletion of *prc* on type IV secretion. We hypothesize that this protease is involved in processing of multiple *cag* T4SS components, and that deletion of the protease will abolish T4SS function. It is possible that deletion of *prc* will have numerous cellular effects. If the Δ*prc* strain is viable and has growth characteristics similar to those of WT *H. pylori*, we will evaluate T4SS function by assay of CagA translocation, and induction of IL-8 synthesis and secretion by gastric epithelial cells in co-culture with the Δ*prc* isogenic mutant. If deletion of *prc* abrogates T4SS function, it will be the first report of a gene encoded outside of the *cag* PAI involved in assembly and function of the *H. pylori cag* T4SS.

Identify T4SS pilus components in sheared fractions from *H. pylori*-gastric epithelial cell co-culture. We have previously utilized chemical cross-linking strategies to detect protein-protein interactions that occur when *H. pylori* is co-cultured with gastric epithelial cells. However, identification of proteins or structures sheared from the surface of *H. pylori* that had been cross-linked to gastric epithelial cells was largely unsuccessful. The greatest obstacle to identification of *H. pylori* proteins present in sheared preparations from a bacteria-gastric cell mixture by mass spectrometry is the overwhelming proportion of human gastric proteins that are also present in the sheared preparations. The protocol for selectively lysing eukaryotic cells in a gastric epithelial cell–*H. pylori* co-culture involves incubation of harvested cell monolayers in low concentrations of saponin. However, when chemical cross-linker is added to the co-culture system, it becomes more difficult to lyse the eukaryotic cells, and large sections of eukaryotic membranes presumably remain chemically cross-linked to the surface of *H. pylori*. Therefore, we will optimize selective eukaryotic cell lysis conditions and recovery of attached *H. pylori* from gastric cell co-culture for subsequent bacterial manipulation and downstream mass spectrometry applications.

133

Mechanical shearing to enrich for surface structures of *H. pylori* will be employed after optimization of eukaryotic lysis conditions. As an additional step for enrichment of Cag proteins from sheared fractions, we will utilize immunoaffinity purification strategies to isolate T4SS components of interest. This enrichment step may be performed under conditions in which the bacteria have been treated with chemical cross-linker, or remained un-cross-linked in co-culture with gastric epithelial cells. Alternatively, chemical cross-linker may be utilized to preserve surface structure morphology in the context of bacteria-host cell co-culture, and the chemical cross-links can be cleaved by addition of DTT to the sheared fraction prior to isolation of Cag proteins. We hypothesize that enrichment of Cag proteins from the sheared preparations will allow us to identify a number of bacterial proteins by mass spectrometry that would normally be difficult to detect in a very complex sheared fraction that contains both host and bacterial proteins. Initially, we will focus on isolating Cag proteins thought to be involved in formation of the T4SS pilus structure, including CagL, CagI, and CagH. The use of monoclonal antibodies for epitope-tagged protein purifications will be preferential to polyclonal sera to allow for specific elution of purified protein complexes by peptide competition. Epitope-tagged versions of several suspected T4SS pilus components have been generated, including CagI and CagL.

Evaluate protein-protein interactions and functions of additional T4SS components. In addition to studies of non-homologous Cag proteins within the *H. pylori* T4SS, it will also be important to evaluate interactions and functions of Cag proteins that have homology to T4SS components found in *Agrobacterium*. Although low-level sequence similarity to Vir proteins can be detected in nine components of the *cag* T4SS [55,125], the *H. pylori* homologs have likely diverged from ancestral Vir proteins. Therefore, we hypothesize that although some Cag proteins have maintained sequence homology, the *H. pylori* Vir homologs have adapted to function within the human host,

and by doing so, perform functions in the *cag* T4SS that are different than the functions of the corresponding proteins within the *Agrobacterium* T4SS.

We have previously generated polyclonal antisera against a number of Vir homologs in *H. pylori*, including CagT (VirB7), CagV (VirB8), CagY (VirB10), CagX (VirB9), and CagC (VirB2). We hypothesize that many of these Vir homologs will interact with other Cag components that are involved in formation of the T4SS apparatus. We will utilize chemical cross-linking in bacteria alone, or bacteria in co-culture with gastric epithelial cells in order to compare Cag protein interactions that potentially differ depending on whether the bacteria are cultured in the presence or absence of host cell contact. Mass spectrometry and immunoblot will be utilized to evaluate Cag protein purification, and to assess co-purification of additional Cag proteins. We will generate the appropriate isogenic mutants to serve as negative controls in immunoaffinity purifications. Epitope-tagged versions of Cag proteins of particular interest will be generated for use in further analyses. These epitope-tagged proteins are often better suited for isolation and subsequent identification of co-purifying proteins by mass spectrometry because of our ability to specifically compete purified complexes from the monoclonal antibody.

We predict that we will identify a number of Cag protein interactions among Vir homologs and non-homologous components of the *cag* T4SS. As demonstrated with CagU, a component of the T4SS that is unique to *H. pylori*, Vir homologs (such as CagW) may interact with one or numerous novel components during T4SS apparatus assembly. In addition, we predict that in many cases, Cag proteins encoded within an operon will interact and serve related functions within the *cag* T4SS. In studies presented in Chapters II and III, we established that *cag* genes positioned within an operon (*cagHIL*) encode proteins that have related functions and localization to the same subcellular site, and our analyses indicate that each encoded protein has an

135

essential role in biogenesis of the T4SS pilus organelle at the bacteria-host cell interface [158]. Further investigation of protein-protein interactions occurring among components of the *H. pylori* T4SS may therefore shed light on the function or localization of unique proteins encoded by the *cag* PAI.

LIST OF PUBLICATIONS

Shaffer, CL, JA Gaddy, JT Loh, EM Johnson, S Hill, EE Hennig, MS McClain, WH McDonald, TL Cover (2011). *Helicobacter pylori* exploits a unique repertoire of type IV secretion system components for pilus assembly at the bacteria-host cell interface. *PLoS Pathogens* 7(9): e1002237.

Loh, JT, CL Shaffer, MB Piazuelo, LE Bravo, MS McClain, P Correa, TL Cover (2011). Analysis of *cagA* in *Helicobacter pylori* Strains from Colombian populations with contrasting gastric cancer risk reveals a biomarker for disease severity. *Cancer Epidemiology, Biomarkers & Prevention* 20(10):2237-49.

Saltet de Sablet, T*, MB Piazuelo*, CL Shaffer*, BG Schneider, M Asim, R Chaturvedi, LE Bravo, L Sicinschi, AG Delgado, RM Mera, DA Israel, J Romero-Gallo, RM Peek, Jr., TL Cover, P Correa, KT Wilson (2011) Phylogeographic origin of *Helicobacter pylori* is a determinant of gastric cancer risk. *Gut* 60(9):1189-95. *Co-First Author

Gangwer, KA, CL Shaffer, S Suerbaum, DB Lacy, TL Cover, SR Bordenstein (2010). Molecular evolution of the *Helicobacter pylori* vacuolating toxin gene *vacA*. *J Bacteriology* 192: 6126-6135.

McClain MS, CL Shaffer, DA Israel, RM Peek, Jr., TL Cover (2009). Genome sequence analysis of *Helicobacter pylori* strains associated with gastric ulceration and gastric cancer. *BMC Genomics* 10: 3.

BIBLIOGRAPHY

1. Marshall BJ, Warren JR (1984) Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. Lancet 1: 1311-1315.
2. Langenberg ML, Tytgat GNJ, Schipper MEI, Rietra PJGM, Zanen HC, *et al.* (1984) *Campylobacter*-like organisms in the stomach of patients and healthy individuals. The Lancet 323: 1348-1349.
3. Jones D, Lessells, AM, Eldridge, J (1984) *Campylobacter* like organisms on the gastric mucosa: culture, histological, and serological studies. Journal of Clinical Pathology 37: 1002-1006.
4. Marshall BJ, Armstrong JA, McGechie DB, Glancy RJ (1985) Attempt to fulfil Koch's postulates for pyloric *Campylobacter*. Med J Aust 142: 436-439.
5. Anonymous. Schistosomes, liver flukes and *Helicobacter pylori* IARC monographs on the evaluation of carcinogenic risks to humans. 1994 7-14 June 1994; Lyon, France. pp. 1-241.
6. Linz B, Balloux F, Moodley Y, Manica A, Liu H, *et al.* (2007) An African origin for the intimate association between humans and *Helicobacter pylori*. Nature 445: 915-918.
7. Suerbaum S, Josenhans C (2007) *Helicobacter pylori* evolution and phenotypic diversification in a changing host. Nat Rev Microbiol 5: 441-452.
8. Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, *et al.* (2003) Traces of human migrations in *Helicobacter pylori* populations. Science 299: 1582-1585.
9. Ghose C, Perez-Perez GI, Dominguez-Bello MG, Pride DT, Bravi CM, *et al.* (2002) East Asian genotypes of *Helicobacter pylori* strains in Amerindians provide evidence for its ancient human carriage. Proc Natl Acad Sci U S A 99: 15107-15111.
10. Suerbaum S, Michetti P (2002) *Helicobacter pylori* infection. N Engl J Med 347: 1175-1186.
11. Dooley CP, Cohen H, Fitzgibbons PL, Bauer M, Appleman MD, *et al.* (1989) Prevalence of *Helicobacter pylori* infection and histologic gastritis in asymptomatic persons. N Engl J Med 321: 1562-1566.
12. Figueiredo C, Machado J, Pharoah P, Seruca R, Sousa S, *et al.* (2002) *Helicobacter pylori* and interleukin 1 genotyping: an opportunity to identify high-risk individuals for gastric carcinoma. J Natl Cancer Inst 94: 1680 - 1687.
13. Nomura A, Stemmermann GN, Chyou PH, Kato I, Pérez-Pérez GI, *et al.* (1991) *Helicobacter pylori* Infection and gastric carcinoma among Japanese Americans in Hawaii. N Engl J Med 325: 1132-1136.
14. Wotherspoon AC, Ortiz-Hidalgo C, Falzon MR, Isaacson PG (1991) *Helicobacter pylori*-associated gastritis and primary B-cell gastric lymphoma. Lancet 338: 1175-1176.
15. Wotherspoon AC, Doglioni C, Diss TC, Pan L, Moschini A, *et al.* (1993) Regression of primary low-grade B-cell gastric lymphoma of mucosa- associated lymphoid tissue type after eradication of *Helicobacter pylori*. Lancet 342: 575-577.
16. Parsonnet J, Friedman GD, Vandersteen DP, Chang Y, Vogelman JH, *et al.* (1991) *Helicobacter pylori* infection and the risk of gastric carcinoma. N Engl J Med 325: 1127-1131.
17. Parsonnet J, Hansen S, Rodriguez L, Gelb AB, Warnke RA, *et al.* (1994) *Helicobacter pylori* infection and gastric lymphoma. N Engl J Med 330: 1267-1271.

18. Graham DY, Lew GM, Klein PD, Evans DG, Evans DJ, Jr., *et al.* (1992) Effect of treatment of *Helicobacter pylori* infection on the long-term recurrence of gastric or duodenal ulcer. A randomized, controlled study. Ann Intern Med 116: 705-708.

19. Nomura A, Stemmermann GN, Chyou PH, Perez-Perez GI, Blaser MJ (1994) *Helicobacter pylori* infection and the risk for duodenal and gastric ulceration. Ann Intern Med 120: 977-981.

20. Anonymous (1994) NIH Consensus Conference. *Helicobacter pylori* in peptic ulcer disease. NIH Consensus Development Panel on *Helicobacter pylori* in Peptic Ulcer Disease. Jama 272: 65-69.

21. Peek RM, Jr., Blaser MJ (2002) *Helicobacter pylori* and gastrointestinal tract adenocarcinomas. Nat Rev Cancer 2: 28-37.

22. Fuchs CS, Mayer RJ (1995) Gastric carcinoma. N Engl J Med 333: 32-41.

23. Annibale B, Capurso G, Chistolini A, D'Ambra G, DiGiulio E, *et al.* (2001) Gastrointestinal causes of refractory iron deficiency anemia in patients without gastrointestinal symptoms. The American Journal of Medicine 111: 439-445.

24. Perry S, de Jong BC, Solnick JV, Sanchez MdIL, Yang S, *et al.* (2010) Infection with *Helicobacter pylori* Is Associated with Protection against Tuberculosis. PLoS ONE 5: e8804.

25. Chen Y, Blaser MJ (2008) *Helicobacter pylori* colonization is inversely associated with childhood asthma. J Infect Dis 198: 553-560.

26. Blaser MJ, Berg DE (2001) *Helicobacter pylori* genetic diversity and risk of human disease. J Clin Invest 107: 767-773.

27. de Sablet T, Piazuelo MB, Shaffer CL, Schneider BG, Asim M, *et al.* Phylogeographic origin of *Helicobacter pylori* is a determinant of gastric cancer risk. Gut. doi:10.1136/gut.2010.234468.

28. McClain MS, Shaffer CL, Israel DA, Peek RM, Jr., Cover TL (2009) Genome sequence analysis of *Helicobacter pylori* strains associated with gastric ulceration and gastric cancer. BMC Genomics 10: 3.

29. Israel DA, Salama N, Krishna U, Rieger UM, Atherton JC, *et al.* (2001) *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. Proc Natl Acad Sci U S A 98: 14625-14630.

30. Salama N, Guillemin K, McDaniel TK, Sherlock G, Tompkins L, *et al.* (2000) A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. Proc Natl Acad Sci U S A 97: 14668-14673.

31. Tomb J-F, White O, Kerlavage AR, Clayton RA, Sutton GG, *et al.* (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature 388: 539-547.

32. Alm RA, Ling LS, Moir DT, King BL, Brown ED, *et al.* (1999) Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. Nature 397: 176-180.

33. Censini S, Lange C, Xiang Z, Crabtree JE, Ghiara P, *et al.* (1996) *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. Proc Natl Acad Sci 93: 14648-14653.

34. Akopyants NS, Clifton SW, Kersulyte D, Crabtree JE, Youree BE, *et al.* (1998) Analyses of the *cag* pathogenicity island of *Helicobacter pylori*. Molecular Microbiology 28: 37-53.

35. Gressmann H, Linz B, Ghai R, Pleissner KP, Schlapbach R, *et al.* (2005) Gain and loss of multiple genes during the evolution of *Helicobacter pylori*. PLoS Genet 1: e43.

36. Atherton J, Cao P, Peek R, Tummuru M, Blaser M, *et al.* (1995) Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. Association of specific *vacA* types with cytotoxin production and peptic ulceration. J Biol Chem 270: 17771 - 17777.

37. Parsonnet J, Friedman GD, Orentreich N, Vogelman H (1997) Risk for gastric cancer in people with CagA positive or CagA negative *Helicobacter pylori* infection. Gut 40: 297-301.

38. Blaser MJ, Perez-Perez GI, Kleanthous H, Cover TL, Peek RM, *et al.* (1995) Infection with *Helicobacter pylori* strains possessing *cagA* is associated with an increased risk of developing adenocarcinoma of the stomach. Cancer Res 55: 2111-2115.

39. van Doorn LJ, Figueiredo C, Sanna R, Plaisier A, Schneeberger P, *et al.* (1998) Clinical relevance of the *cagA, vacA*, and *iceA* status of *Helicobacter pylori*. Gastroenterology 115: 58-66.

40. Gerhard M, Lehn N, Neumayer N, Boren T, Rad R, *et al.* (1999) Clinical relevance of the *Helicobacter pylori* gene for blood-group antigen-binding adhesin. Proc Natl Acad Sci U S A 96: 12778-12783.

41. Cao P, Cover TL (2002) Two different families of *hopQ* alleles in *Helicobacter pylori*. J Clin Microbiol 40: 4504-4511.

42. Bourzac KM, Guillemin K (2005) *Helicobacter pylori*-host cell interactions mediated by type IV secretion. Cell Microbiol 7: 911-919.

43. Odenbreit S, Puls J, Sedlmaier B, Gerland E, Fischer W, *et al.* (2000) Translocation of *Helicobacter pylori* CagA into gastric epithelial cells by type IV secretion. Science 287: 1497-1500.

44. Segal ED, Cha J, Lo J, Falkow S, Tompkins LS (1999) Altered states: involvement of phosphorylated CagA in the induction of host cellular growth changes by *Helicobacter pylori*. Proc Natl Acad Sci USA 96: 14559-14564.

45. Hatakeyama M (2004) Oncogenic mechanisms of the *Helicobacter pylori* CagA protein. Nat Rev Cancer 4: 688-694.

46. Backert S, Selbach M (2008) Role of type IV secretion in *Helicobacter pylori* pathogenesis. Cellular Microbiology 10: 1573-1581.

47. Amieva MR, Vogelmann R, Covacci A, Tompkins LS, Nelson WJ, *et al.* (2003) Disruption of the epithelial apical-junctional complex by *Helicobacter pylori* CagA. Science 300: 1430-1434.

48. Higashi H, Tsutsumi R, Muto S, Sugiyama T, Azuma T, *et al.* (2002) SHP-2 tyrosine phosphatase as an intracellular target of *Helicobacter pylori* CagA protein. Science 295: 683-686.

49. Selbach M, Moese S, Hurwitz R, Hauck CR, Meyer TF, *et al.* (2003) The *Helicobacter pylori* CagA protein induces cortactin dephosphorylation and actin rearrangement by c-Src inactivation. Embo J 22: 515-528.

50. Mimuro H, Suzuki T, Tanaka J, Asahi M, Haas R, *et al.* (2002) Grb2 is a key mediator of *Helicobacter pylori* CagA protein activities. Mol Cell 10: 745-755.

51. Churin Y, Al-Ghoul L, Kepp O, Meyer TF, Birchmeier W, *et al.* (2003) *Helicobacter pylori* CagA protein targets the c-Met receptor and enhances the motogenic response. J Cell Biol 161: 249-255.

52. Bagnoli F, Buti L, Tompkins L, Covacci A, Amieva MR (2005) *Helicobacter pylori* CagA induces a transition from polarized to invasive phenotypes in MDCK cells. Proc Natl Acad Sci U S A 102: 16339-16344.

53. Franco AT, Israel DA, Washington MK, Krishna U, Fox JG, *et al.* (2005) Activation of β-catenin by carcinogenic *Helicobacter pylori*. Proc Natl Acad Sci U S A 102: 10646-10651.

54. Guillemin K, Salama NR, Tompkins LS, Falkow S (2002) Cag pathogenicity island-specific responses of gastric epithelial cells to *Helicobacter pylori* infection. Proc Natl Acad Sci U S A 99: 15136-15141.

55. Fischer W, Puls J, Buhrdorf R, Gebert B, Odenbreit S, *et al.* (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.

56. Viala J, Chaput C, Boneca IG, Cardona A, Girardin SE, *et al.* (2004) Nod1 responds to peptidoglycan delivered by the *Helicobacter pylori* cag pathogenicity island. Nat Immunol 5: 1166-1174.

57. Brandt S, Kwok T, Hartig R, Konig W, Backert S (2005) NF-kappaB activation and potentiation of proinflammatory responses by the *Helicobacter pylori* CagA protein. Proc Natl Acad Sci U S A 102: 9300-9305.

58. Kim SY, Lee YC, Kim HK, Blaser MJ (2006) *Helicobacter pylori* CagA transfection of gastric epithelial cells induces interleukin-8. Cell Microbiol 8: 97-106.

59. Argent RH, Hale JL, El-Omar EM, Atherton JC (2008) Differences in *Helicobacter pylori* CagA tyrosine phosphorylation motif patterns between western and East Asian strains, and influences on interleukin-8 secretion. Journal of Medical Microbiology 57: 1062-1067.

60. 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.

61. Crawford HC, Krishna US, Israel DA, Matrisian LM, Washington MK, *et al.* (2003) *Helicobacter pylori* strain-selective induction of matrix metalloproteinase-7 in vitro and within gastric mucosa. Gastroenterology 125: 1125-1136.

62. Keates S, Keates AC, Nath S, Peek RM, Jr., Kelly CP (2005) Transactivation of the epidermal growth factor receptor by *cag+ Helicobacter pylori* induces upregulation of the early growth response gene Egr-1 in gastric epithelial cells. Gut 54: 1363-1369.

63. Keates S, Sougioultzis S, Keates AC, Zhao D, Peek RM, Jr., *et al.* (2001) *cag+ Helicobacter pylori* induce transactivation of the epidermal growth factor receptor in AGS gastric epithelial cells. J Biol Chem 276: 48127-48134.

64. Kwok T, Zabler D, Urman S, Rohde M, Hartig R, *et al.* (2007) *Helicobacter* exploits integrin for type IV secretion and kinase activation. Nature 449: 862-866.

65. Jiménez-Soto LF, Kutter S, Sewald X, Ertl C, Weiss E, *et al.* (2009) *Helicobacter pylori* Type IV Secretion Apparatus Exploits β1 Integrin in a Novel RGD-Independent Manner. PLoS Pathog 5: e1000684.

66. Tegtmeyer N, Hartig R, Delahay RM, Rohde M, Brandt S, *et al.* (2010) A Small Fibronectin-mimicking Protein from Bacteria Induces Cell Spreading and Focal Adhesion Formation. Journal of Biological Chemistry 285: 23515-23526.

67. Saha A, Backert S, Hammond CE, Gooz M, Smolka AJ (2010) *Helicobacter pylori* CagL Activates ADAM17 to Induce Repression of the Gastric H, K-ATPase α Subunit. Gastroenterology 139: 239-248.

68. Fronzes R, Schafer E, Wang L, Saibil HR, Orlova EV, *et al.* (2009) Structure of a type IV secretion system core complex. Science 323: 266-268.

69. Cascales E, Christie PJ (2003) The versatile bacterial type IV secretion systems. Nat Rev Microbiol 1: 137-149.

70. Alvarez-Martinez CE, Christie PJ (2009) Biological diversity of prokaryotic type IV secretion systems. Microbiol Mol Biol Rev 73: 775-808.

71. Yeo HJ, Waksman G (2004) Unveiling molecular scaffolds of the type IV secretion system. J Bacteriol 186: 1919-1926.

72. Fronzes R, Christie PJ, Waksman G (2009) The structural biology of type IV secretion systems. Nat Rev Micro 7: 703-714.

73. 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.

74. Bacon DJ, Alm RA, Hu L, Hickey TE, Ewing CP, *et al.* (2002) DNA Sequence and Mutational Analyses of the pVir Plasmid of *Campylobacter jejuni* 81-176. Infect Immun 70: 6242-6250.

75. Hofreuter D, Odenbreit S, Püls Jr, Schwan D, Haas R (2000) Genetic competence in *Helicobacter pylori*: mechanisms and biological implications. Research in Microbiology 151: 487-491.

76. Dillard JP, Seifert HS (2001) A variable genetic island specific for *Neisseria gonorrhoeae* is involved in providing DNA for natural transformation and is found more often in disseminated infection isolates. Molecular Microbiology 41: 263-277.

77. Li SD, Kersulyte D, Lindley IJ, Neelam B, Berg DE, *et al.* (1999) Multiple genes in the left half of the cag pathogenicity island of *Helicobacter pylori* are required for tyrosine kinase-dependent transcription of interleukin-8 in gastric epithelial cells. Infect Immun 67: 3893-3899.

78. 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.

79. Terradot L, Waksman G (2011) Architecture of the *Helicobacter pylori* Cag-type IV secretion system. FEBS Journal 278: 1213-1222.

80. Kutter S, Buhrdorf R, Haas J, Schneider-Brachert W, Haas R, *et al.* (2008) Protein subassemblies of the *Helicobacter pylori* Cag type IV secretion system revealed by localization and interaction studies. J Bacteriol.

81. Busler VJ, Torres VJ, McClain MS, Tirado O, Friedman DB, *et al.* (2006) Protein-Protein Interactions among *Helicobacter pylori* Cag Proteins. J Bacteriol 188: 4787-4800.

82. Yeo HJ, Savvides SN, Herr AB, Lanka E, Waksman G (2000) Crystal structure of the hexameric traffic ATPase of the *Helicobacter pylori* type IV secretion system. Mol Cell 6: 1461-1472.

83. 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.

84. Cendron L, Tasca E, Seraglio T, Seydel A, Angelini A, *et al.* (2007) The crystal structure of CagS from the *Helicobacter pylori* pathogenicity island. Proteins: Structure, Function, and Bioinformatics 69: 440-443.

85. Cendron L, Couturier M, Angelini A, Barison N, Stein M, *et al.* (2009) The *Helicobacter pylori* CagD (HP0545, Cag24) Protein Is Essential for CagA Translocation and Maximal Induction of Interleukin-8 Secretion. Journal of Molecular Biology 386: 204-217.

86. Nesic D, Miller MC, Quinkert ZT, Stein M, Chait BT, *et al.* (2010) *Helicobacter pylori* CagA inhibits PAR1-MARK family kinases by mimicking host substrates. Nat Struct Mol Biol 17: 130-132.

87. Rain JC, Selig L, De Reuse H, Battaglia V, Reverdy C, *et al.* (2001) The protein-protein interaction map of *Helicobacter pylori*. Nature 409: 211-215.

88. Terradot L, Durnell N, Li M, Li D, Ory J, *et al.* (2004) Biochemical characterization of protein complexes from the *Helicobacter pylori* protein interaction map: strategies

for complex formation and evidence for novel interactions within type IV secretion systems. Mol Cell Proteomics.

89. 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.

90. Jurik A, Hausser E, Kutter S, Pattis I, Prassl S, *et al.* (2010) The Coupling Protein Cagβ and Its Interaction Partner CagZ Are Required for Type IV Secretion of the *Helicobacter pylori* CagA Protein. Infect Immun 78: 5244-5251.

91. 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.

92. Rohde M, Puls J, Buhrdorf R, Fischer W, Haas R (2003) A novel sheathed surface organelle of the *Helicobacter pylori cag* type IV secretion system. Mol Microbiol 49: 219-234.

93. Yuan Q, Carle A, Gao C, Sivanesan D, Aly KA, *et al.* (2005) Identification of the VirB4-VirB8-VirB5-VirB2 pilus assembly sequence of type IV secretion systems. J Biol Chem 280: 26349-26359.

94. Andrzejewska J, Lee SK, Olbermann P, Lotzing N, Katzowitsch E, *et al.* (2006) Characterization of the pilin ortholog of the *Helicobacter pylori* type IV *cag* pathogenicity apparatus, a surface-associated protein expressed during infection. J Bacteriol 188: 5865-5877.

95. Tanaka J, Suzuki T, Mimuro H, Sasakawa C (2003) Structural definition on the surface of *Helicobacter pylori* type IV secretion apparatus. Cell Microbiol 5: 395-404.

96. Aras RA, Fischer W, Perez-Perez GI, Crosatti M, Ando T, *et al.* (2003) Plasticity of repetitive DNA sequences within a bacterial (Type IV) secretion system component. J Exp Med 198: 1349-1360.

97. Tegtmeyer N, Wessler S, Backert S (2011) Role of the cag-pathogenicity island encoded type IV secretion system in *Helicobacter pylori* pathogenesis. FEBS Journal 278:1190-1202.

98. Anderson LB, Hertzel AV, Das A (1996) *Agrobacterium tumefaciens* VirB7 and VirB9 form a disulfide-linked protein complex. Proc Natl Acad Sci U S A 93: 8889-8894.

99. Baron C, Thorstenson YR, Zambryski PC (1997) The lipoprotein VirB7 interacts with VirB9 in the membranes of *Agrobacterium tumefaciens*. J Bacteriol 179: 1211-1218.

100. Das A, Xie YH (2000) The *Agrobacterium* T-DNA transport pore proteins VirB8, VirB9, and VirB10 interact with one another. J Bacteriol 182: 758-763.

101. Kutter S, Buhrdorf R, Haas J, Schneider-Brachert W, Haas R, *et al.* (2008) Protein Subassemblies of the *Helicobacter pylori* Cag Type IV Secretion System Revealed by Localization and Interaction Studies. J Bacteriol 190: 2161-2171.

102. Ivie S, McClain M, Algood H, Lacy DB, Cover T (2010) Analysis of a beta-helical region in the p55 domain of *Helicobacter pylori* vacuolating toxin. BMC Microbiology 10: 60.

103. Loh JT, Torres VJ, Cover TL (2007) Regulation of *Helicobacter pylori cagA* expression in response to salt. Cancer Res 67: 4709-4715.

104. Dailidiene D, Dailide G, Kersulyte D, Berg DE (2006) Contraselectable Streptomycin Susceptibility Determinant for Genetic Manipulation and Analysis of *Helicobacter pylori*. Appl Environ Microbiol 72: 5908-5914.

105. Copass M, Grandi G, Rappuoli R (1997) Introduction of unmarked mutations in the *Helicobacter pylori vacA* gene with a sucrose sensitivity marker. Infect Immun 65: 1949-1952.

106. Loh JT, Forsyth MH, Cover TL (2004) Growth phase regulation of *flaA* expression in *Helicobacter pylori* is *luxS* dependent. Infect Immun 72: 5506-5510.

107. Hopp TP, Woods KR (1981) Prediction of protein antigenic determinants from amino acid sequences. Proc Natl Acad Sci U S A 78: 3824-3828.

108. MacCoss MJ, McDonald WH, Saraf A, Sadygov R, Clark JM, *et al.* (2002) Shotgun identification of protein modifications from protein complexes and lens tissue. Proc Natl Acad Sci U S A 99: 7900-7905.

109. Yates JR, Eng JK, McCormack AL, Schieltz D (1995) Method to Correlate Tandem Mass Spectra of Modified Peptides to Amino Acid Sequences in the Protein Database. Analytical Chemistry 67: 1426-1436.

110. Ma ZQ, Dasari S, Chambers MC, Litton MD, Sobecki SM, *et al.* (2009) IDPicker 2.0: Improved protein assembly with high discrimination peptide identification filtering. J Proteome Res 8: 3872-3881.

111. Wessel D, Flügge UI (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. Analytical Biochemistry 138: 141-143.

112. Ilver D, Barone S, Mercati D, Lupetti P, Telford J (2004) *Helicobacter pylori* toxin VacA is transferred to host cells via a novel contact-dependent mechanism. Cell Microbiol 6: 167 - 174.

113. Fischer W, Buhrdorf R, Gerland E, Haas R (2001) Outer membrane targeting of passenger proteins by the vacuolating cytotoxin autotransporter of *Helicobacter pylori*. Infection and Immunity 69: 6769-6775.

114. Schraw W, Li Y, McClain M, Goot F, Cover T (2002) Association of *Helicobacter pylori* vacuolating toxin (VacA) with lipid rafts. J Biol Chem 277: 34642 - 34650.

115. Marcus EA, Moshfegh AP, Sachs G, Scott DR (2005) The Periplasmic α-Carbonic Anhydrase Activity of *Helicobacter pylori* Is Essential for Acid Acclimation. J Bacteriol 187: 729-738.

116. Abramoff MD, Magelhaes, P.J., Ram, S.J (2004) Image Processing with ImageJ. Biophotonics International 11: 36-42.

117. Zhang B, VerBerkmoes NC, Langston MA, Uberbacher E, Hettich RL, *et al.* (2006) Detecting Differential and Correlated Protein Expression in Label-Free Shotgun Proteomics. Journal of Proteome Research 5: 2909-2918.

118. Zar JH (1999) Biostatistical Analysis. New Jersey: Prentice Hall.

119. Nakai K, Kanehisa M (1991) Expert system for predicting protein localization sites in gram-negative bacteria. Proteins: Structure, Function, and Bioinformatics 11: 95-110.

120. Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997) A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Int J Neural Syst 8: 581-599.

121. Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeisz S, *et al.* (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. Nature 464: 250-255.

122. Pattis I, Weiss E, Laugks 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.

123. Fernandez D, Spudich G, Zhou X, Christie P (1996) The *Agrobacterium tumefaciens* VirB7 lipoprotein is required for stabilization of VirB proteins during assembly of the T-complex transport apparatus. J Bacteriol 178: 3168-3176.

124. Hapfelmeier S, Domke N, Zambryski PC, Baron C (2000) VirB6 Is Required for Stabilization of VirB5 and VirB3 and Formation of VirB7 Homodimers in *Agrobacterium tumefaciens*. J Bacteriol 182: 4505-4511.

125. Fischer W (2011) Assembly and molecular mode of action of the *Helicobacter pylori* Cag type IV secretion apparatus. FEBS Journal 278: 1203-1212.

126. Cover TL, Blaser MJ (2009) *Helicobacter pylori* in health and disease. Gastroenterology 136: 1863-1873.

127. Backert S, Kwok T, Schmid M, Selbach M, Moese S, *et al.* (2005) Subproteomes of soluble and structure-bound *Helicobacter pylori* proteins analyzed by two-dimensional gel electrophoresis and mass spectrometry. Proteomics 5: 1331-1345.

128. Olbermann P, Josenhans C, Moodley Y, Uhr M, Stamer C, *et al.* (2010) A Global Overview of the Genetic and Functional Diversity in the *Helicobacter pylori* cag Pathogenicity Island. PLoS Genet 6: e1001069.

129. Lai E-M, 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.

130. Blocker A, Komoriya K, Aizawa S-I (2003) Type III secretion systems and bacterial flagella: Insights into their function from structural similarities. Proc Natl Acad Sci U S A 100: 3027-3030.

131. Galan JE, Wolf-Watz H (2006) Protein delivery into eukaryotic cells by type III secretion machines. Nature 444: 567-573.

132. Kubori T, Matsushima Y, Nakamura D, Uralil J, Lara-Tejero M, *et al.* (1998) Supramolecular Structure of the *Salmonella typhimurium* Type III Protein Secretion System. Science 280: 602-605.

133. Tamano K, Aizawa S-I, Katayama E, Nonaka T, Imajoh-Ohmi S, *et al.* (2000) Supramolecular structure of the *Shigella* type III secretion machinery: the needle part is changeable in length and essential for delivery of effectors. Embo J 19: 3876-3887.

134. Sekiya K, Ohishi M, Ogino T, Tamano K, Sasakawa C, *et al.* (2001) Supermolecular structure of the enteropathogenic *Escherichia coli* type III secretion system and its direct interaction with the EspA-sheath-like structure. Proc Natl Acad Sci U S A 98: 11638-11643.

135. Deane JE, Roversi P, Cordes FS, Johnson S, Kenjale R, *et al.* (2006) Molecular model of a type III secretion system needle: Implications for host-cell sensing. Proc Natl Acad Sci U S A 103: 12529-12533.

136. Yip CK, Kimbrough TG, Felise HB, Vuckovic M, Thomas NA, *et al.* (2005) Structural characterization of the molecular platform for type III secretion system assembly. Nature 435: 702-707.

137. Aly KA, Baron C (2007) The VirB5 protein localizes to the T-pilus tips in *Agrobacterium tumefaciens*. Microbiology 153: 3766-3775.

138. Aguilar J, Zupan J, Cameron TA, Zambryski PC (2010) *Agrobacterium* type IV secretion system and its substrates form helical arrays around the circumference of virulence-induced cells. Proc Natl Acad Sci U S A 107: 3758-3763.

139. Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, *et al.* (2009) CDD: specific functional annotation with the Conserved Domain Database. Nucleic Acids Research 37: D205-D210.

140. Erhardt M, Hirano T, Su Y, Paul K, Wee DH, *et al.* (2010) The role of the FliK molecular ruler in hook-length control in *Salmonella enterica*. Molecular Microbiology 75: 1272-1284.

141. Kawagishi I, Homma M, Williams A, Macnab R (1996) Characterization of the flagellar hook length control protein FliK of *Salmonella typhimurium* and *Escherichia coli*. J Bacteriol 178: 2954-2959.

142. Williams A, Yamaguchi S, Togashi F, Aizawa S, Kawagishi I, *et al.* (1996) Mutations in *fliK* and *flhB* affecting flagellar hook and filament assembly in *Salmonella typhimurium*. J Bacteriol 178: 2960-2970.

143. Journet L, Agrain Cl, Broz P, Cornelis GR (2003) The Needle Length of Bacterial Injectisomes Is Determined by a Molecular Ruler. Science 302: 1757-1760.

144. Zerovnik E, Zavasnik-Bergant V, Kopitar-Jerala N, Pompe-Novak M, Skarabot M, *et al.* (2002) Amyloid Fibril Formation by Human Stefin B in vitro: Immunogold Labelling and Comparison to Stefin A. Biological Chemistry 383: 859-863.

145. Crooks GE, Hon G, Chandonia J-M, Brenner SE (2004) WebLogo: A Sequence Logo Generator. Genome Research 14: 1188-1190.

146. Muramoto K, Makishima S, Aizawa S-I, Macnab RM (1999) Effect of Hook Subunit Concentration on Assembly and Control of Length of the Flagellar Hook of *Salmonella*. J Bacteriol 181: 5808-5813.

147. Kutsukake K, Minamino T, Yokoseki T (1994) Isolation and characterization of FliK-independent flagellation mutants from *Salmonella typhimurium*. J Bacteriol 176: 7625-7629.

148. Makishima S, Komoriya K, Yamaguchi S, Aizawa S-I (2001) Length of the Flagellar Hook and the Capacity of the Type III Export Apparatus. Science 291: 2411-2413.

149. Nagai H, Cambronne ED, Kagan JC, Amor JC, Kahn RA, *et al.* (2005) A C-terminal translocation signal required for Dot/Icm-dependent delivery of the *Legionella* RalF protein to host cells. Proc Natl Acad Sci U S A 102: 826-831.

150. Vergunst AC, van Lier MCM, den Dulk-Ras A, Grosse Stüve TA, Ouwehand A, *et al.* (2005) Positive charge is an important feature of the C-terminal transport signal of the VirB/D4-translocated proteins of *Agrobacterium*. Proc Natl Acad Sci U S A 102: 832-837.

151. Schulein R, Guye P, Rhomberg TA, Schmid MC, Schöder G, *et al.* (2005) A bipartite signal mediates the transfer of type IV secretion substrates of *Bartonella henselae* into human cells. Proc Natl Acad Sci U S A 102: 856-861.

152. Liu G, McDaniel TK, Falkow S, Karlin S (1999) Sequence anomalies in the Cag7 gene of the *Helicobacter pylori* pathogenicity island. Proc Natl Acad Sci U S A 96: 7011-7016.

153. Zahrl D, Wagner M, Bischof K, Bayer M, Zavecz B, *et al.* (2005) Peptidoglycan degradation by specialized lytic transglycosylases associated with type III and type IV secretion systems. Microbiology 151: 3455-3467.

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

155. 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*. International Journal of Medical Microbiology 293: 213-217.

156. Chandran V, Fronzes R, Duquerroy S, Cronin N, Navaza J, *et al.* (2009) Structure of the outer membrane complex of a type IV secretion system. Nature 462: 1011-1015.

157. Judd PK, Kumar RB, Das A (2005) Spatial location and requirements for the assembly of the *Agrobacterium tumefaciens* type IV secretion apparatus. Proc Natl Acad Sci U S A 102: 11498-11503.

158. Shaffer CL, Gaddy JA, Loh JT, Johnson EM, Hill S, *et al.* (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.

159. Cascales E, Christie PJ (2004) Definition of a Bacterial Type IV Secretion Pathway for a DNA Substrate. Science 304: 1170-1173.

160. Smith AL, Friedman DB, Yu H, Carnahan RH, Reynolds AB (2011) ReCLIP (Reversible Cross-Link Immuno-Precipitation): An Efficient Method for Interrogation of Labile Protein Complexes. PLoS ONE 6: e16206.

161. Fanning AS, Anderson JM (1996) Protein-protein interactions: PDZ domain networks. Current Biology 6: 1385-1388.

162. Beebe KD, Shin J, Peng J, Chaudhury C, Khera J, *et al.* (2000) Substrate Recognition through a PDZ Domain in Tail-Specific Protease. Biochemistry 39: 3149-3155.

163. Tadokoro A, Hayashi H, Kishimoto T, Makino Y, Fujisaki S, *et al.* (2004) Interaction of the *Escherichia coli* Lipoprotein NlpI with Periplasmic Prc (Tsp) Protease. Journal of Biochemistry 135: 185-191.

164. Judd PK, Kumar RB, Das A (2005) The type IV secretion apparatus protein VirB6 of *Agrobacterium tumefaciens* localizes to a cell pole. Molecular Microbiology 55: 115-124.

165. Raychaudhury S, Farelli JD, Montminy TP, Matthews M, Ménétret J-Fo, *et al.* (2009) Structure and Function of Interacting IcmR-IcmQ Domains from a Type IVb Secretion System in *Legionella pneumophila*. Structure 17: 590-601.

166. Duménil G, Montminy TP, Tang M, Isberg RR (2004) IcmR-regulated Membrane Insertion and Efflux by the *Legionella pneumophila* IcmQ Protein. Journal of Biological Chemistry 279: 4686-4695.

167. Duménil G, Isberg RR (2001) The *Legionella pneumophila* IcmR protein exhibits chaperone activity for IcmQ by preventing its participation in high-molecular-weight complexes. Molecular Microbiology 40: 1113-1127.

168. Bhakdi S, Tranum-Jensen J (1991) Alpha-toxin of *Staphylococcus aureus*. Microbiol Mol Biol Rev 55: 733-751.

169. De Geyter C, Wattiez R, Sansonetti P, Falmagne P, Ruysschaert J-M, *et al.* (2000) Characterization of the interaction of IpaB and IpaD, proteins required for entry of *Shigella flexneri* into epithelial cells, with a lipid membrane. European Journal of Biochemistry 267: 5769-5776.

170. Dumas F, Duckely M, Pelczar P, Van Gelder P, Hohn B (2001) An *Agrobacterium* VirE2 channel for transferred-DNA transport into plant cells. Proc Natl Acad Sci U S A 98: 485-490.

171. Merrell DS, Goodrich ML, Otto G, Tompkins LS, Falkow S (2003) pH-regulated gene expression of the gastric pathogen *Helicobacter pylori*. Infect Immun 71: 3529-3539.

172. Merrell DS, Thompson LJ, Kim CC, Mitchell H, Tompkins LS, *et al.* (2003) Growth phase-dependent response of *Helicobacter pylori* to iron starvation. Infect Immun 71: 6510-6525.

173. Tan S, Noto JM, Romero-Gallo J, Peek RM, Jr., Amieva MR (2011) *Helicobacter pylori* Perturbs Iron Trafficking in the Epithelium to Grow on the Cell Surface. PLoS Pathog 7: e1002050.

174. Akada JK, Shirai M, Takeuchi H, Tsuda M, Nakazawa T (2000) Identification of the urease operon in *Helicobacter pylori* and its control by mRNA decay in response to pH. Molecular Microbiology 36: 1071-1084.

175. Sabarth N, Lamer S, Zimny-Arndt U, Jungblut PR, Meyer TF, *et al.* (2002) Identification of surface proteins of *Helicobacter pylori* by selective biotinylation, affinity purification, and two-dimensional gel electrophoresis. J Biol Chem 277: 27896-27902.

176. Smith TG, Lim J-M, Weinberg MV, Wells L, Hoover TR (2007) Direct analysis of the extracellular proteome from two strains of *Helicobacter pylori*. Proteomics 7: 2240-2245.