THE MOLECULAR AND STRUCTURAL ANALYSIS OF THE
cag TYPE IV SECRETION SYSTEM CORE COMPLEX

By

Arwen Frick-Cheng

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

May, 2017

Nashville, Tennessee

Approved:

Mark R. Denison, M.D.

D. Borden Lacy, Ph.D

Melanie D. Ohi, Ph.D

Eric P. Skaar Ph.D, MPH

Timothy L. Cover, M.D.

To my parents and Sam- For everything, really.

ACKNOWLEDGMENTS

First things first, I would like to thank my mentor, Tim Cover for welcoming me into his lab and providing me with years of mentorship and support. Your office door is always open, and you have always been there to listen to me, discuss ideas, and plan experiments. You've also let me run wild with my project, develop my own ideas and hypothesis, which has been a thrilling experience, and frankly, the whole reason I came to graduate school. I cannot thank you enough for the guidance you have given me as a scientist.

I would also like to thank the Cover lab, all of you have been kind and welcoming to me, ready to listen to all of my questions and always ready to discuss a protocol or brainstorm a new idea. I would like to thank Brad Voss, who was always ready to get a coffee with me and talk shop, and Rhonda Caston, who has now taken Brad's mantle as official coffee buddy. I would like to thank Amber Beckett for keeping me grounded in some of my ideas, and for having a meme for *everything.* Also I would like to thank the Ohi lab for being my lab home away from home and being so kind and welcoming. I would especially like to thank Tasia Pyburn who taught me all the EM I know, and was so very patient with me throughout the entire process.

I would like to thank all my friends and family as well. The bonds made in grad school are tight, and I wouldn't have been able to do it without the constant support of all the other students struggling alongside of me. I would like to thank my parents as well, who have been supportive and understanding of my journey here, as they made a similar journey, now over thirty years ago.

And finally, I want to thank Sam, my fiancée. Thank you for putting up with a scientist's strange work schedule, thank you for listening to my science babble, thank you for being my cheerleader and constant companion.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2D	Two-dimensional
3D	Three-dimensional
ACN	Acetonitrile
BME	β-mercaptoethanol
cag PAI	Cytotoxin associated gene pathogenicity island
CCD	charge-coupled-device
CHAPS	CHAPS hydrate
cryoEM	Cryo-electron microscopy
C-terminal	Carboxy-terminal
DDM	n-Dodecyl β-D-maltoside
DM	n-Decyl β-D-maltopyranoside
DOC	sodium deoxycholate
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme linked immunosorbent assay
EM	Electron microscopy
FBS	Fetal bovine serum
GST	Glutathione S-transferase
HA	Hemagglutinin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High-performance liquid chromatography
HRP	Horseradish peroxidase
IL-8	Interleukin 8 (CXCL8)

IP	Immunopurification
kV	Kilovolt
LDAO	N,N-dimethyldodecylamine N-oxide
LPS	lipopolysaccharide
MOI	Multiplicity of infection
MS/MS	tandem mass spectrometry
MudPIT	Multidimensional protein identification technology
NMR	Nuclear magnetic resonance
NP-40	Nonidet-P40 substitute (Igepal CA-630)
N-terminal	Amino-terminal
OD	Optical density
OG	Octyl-beta-glucoside
PBS	Phosphate buffered saline
PFA	Paraformaldehyde
T4SS	Type IV Secretion System
TCA	Trichloroacetic acid
TCEP	8M urea, 100 mM Tris, pH 8.5, reduced with Tris(2-carboxyethyl)phosphine
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SXC	Strong cation exchange

RCT	Random conical tilt
RIPA	1% NP-40, 0.25% deoxycholate, 10 mM HEPES, and 100 mM NaCl pH 7.0
WT	Wild Type
Y2H	Yeast-two-hybrid

CHAPTER I

INTRODUCTION

The *Helicobacter pylori cag* Type IV Secretion System

Helicobacter pylori is a Gram-negative bacterium that colonizes the stomach in about half of the world's population (1, 2). Persistent *H. pylori* infection is a risk factor for the development of gastric cancer or peptic ulcer disease, and has been designated a Class I carcinogen by the World Health Organization (3, 4). Adverse disease outcomes associated with *H. pylori* infection occur predominantly in persons infected with strains that produce an effector protein known as CagA, which is translocated into host cells by a type IV secretion system (T4SS) (5-7). The *cagA* gene and genes required for T4SS-dependent secretion of CagA are located within a 40-kb region of the *H. pylori* chromosome known as the *cag* pathogenicity island (*cag* PAI) (5, 6). The *cag* PAI is a 40 kb region of the *H. pylori* genome that is present in almost 100% of East Asian strains and 60-70% of strains isolated from western countries (5, 8, 9). The CagA effector protein contributes to both the carcinogenesis and ulcer development associated with *H. pylori* by exerting a variety of effects within the host cell (10, 11). Upon entry into host cells, CagA causes alterations in cell signaling that are linked to gastric carcinogenesis. Src family kinases phosphorylate the C-terminus of CagA, and once phosphorylated, CagA alters the function of a variety of host proteins. One of these is the oncoprotein SHP2, which upon activation by CagA initiates a signaling cascade resulting in the activation of the Erk MAPK pathway, causing unregulated

1

cell growth (12, 13). CagA is the only protein known to be secreted by the *H. pylori cag* T4SS. The T4SS also contributes to several phenotypes independent of CagA translocation. The most well-studied of these is the *H. pylori* induced stimulation of IL-8 production in gastric epithelial cells. This is caused by peptidoglycan entry into host cells (14, 15) which is dependent on a fully functional T4SS.

<p style="text-align:center">Prototypical T4SSs</p>

T4SSs are a versatile family of secretion systems found in a wide variety of bacterial species, not just *H. pylori* (16-18). Three main functional categories of T4SSs are recognized: those that deliver effector proteins into host cells, those that export DNA, and those that import DNA (18). Effector proteins or DNA-protein complexes can be delivered into many types of recipient cells, including mammalian cells, plant cells, fungi, or other bacteria. Examples of DNA-translocating T4SSs include the VirB/D system found in *Agrobacterium tumefaciens* (which translocates T-DNA into plant cells, causing crown-gall disease) and plasmid conjugation systems (such as the *Escherichia coli* pKM101-encoded Tra system). Effector protein-translocating T4SSs are exemplified by the Dot/Icm systems in *Legionella* and *Coxiella*, and the *H. pylori cag* T4SS.

The system initially used characterize T4SSs is the VirB/VirD system present in *A. tumefaciens* (19). In the paradigm set by this system the T4SS machinery is intricate and involves three main protein subassemblies. First is the core complex (16, 20, 21), a large group of physically interacting proteins that span

the inner and outer membrane of the bacterium and is predicted to anchor the pilus. Second is the energy subcomplex, which is a group of ATPases responsible for providing energy for substrate translocation and formation of the rest of the T4SS machinery (22, 23). The final group is composed of the subunits that form the structure of the pilus, which forms when the bacteria come into contact with host cells(19). This pilus then can be used to translocate effector molecules into the host cell.

There are twelve proteins responsible for the proper formation and function of the T4SS in *A. tumefaciens* (24). In contrast, there are fifteen proteins encoded in the *H. pylori cag* PAI which are essential for activity of the *cag* T4SS, since they were indispensable for IL-8 induction (14). Interestingly, there are also three Cag proteins that are dispensable for IL-8 induction, but are required for CagA translocation, implying that there are multiple steps in CagA translocation that require more than functional T4SS machinery (14).

<u>Differences between prototypical T4SSs and the *cag* T4SS</u>

Some of *cag* proteins are predicted to be homologous to *A. tumefaciens* counterparts. However, most of these predictions are strictly based on sequence similarity, protein predictions and localization. A recent model proposed *H. pylori's* core complex is made of CagT, X, and Y, the ATPases are CagE, α, and β and the pilus is made of major pilin subunit, CagC and minor subunit, CagL (24, 25). However, these connections are tenuous at best. For example, CagY has only a 32% protein sequence similarity to its *A. tumefaciens* counterpart. CagC, the proposed major pilin subunit can be knocked out without abrogating pilus

formation (26). CagT is 208 amino acids long, as opposed to 50 in its *A. tumefaciens* counterpart. Finally, CagM and Cag3, two *H. pylori* proteins required for T4SS activity, have no homologous counterparts in *A. tumefaciens.* Furthermore, nine of the *cag* PAI genes required for CagA translocation are unique to *H. pylori* with no obvious homologs in other bacteria. This suggests that there are important differences between the *H. pylori cag* T4SS and T4SSs in other bacterial species. Multiple models for the structural organization of the *H. pylori cag* T4SS have been proposed, but there are numerous limitations and controversial features of the current models, due to a lack of detailed biochemical and structural data

One of those *cag* proteins which were required for CagA translocation, but not IL-8 induction is CagF. This stable, soluble protein when initially characterized was found to physically bind and interact with the C-terminal domain of CagA (27, 28). However, more recent studies show that CagF binds all regions of CagA (29). CagF has been proposed as a chaperone for CagA, based on its low isoelectric point and tendency to form homodimers (28), two common characteristics of chaperone proteins in Type III Secretion Systems, though CagF is not required for CagA stability. Since CagF is essential for CagA translocation, it is reasonable to presume that the CagF-CagA complex physically interacts with the *cag* T4SS machinery. However, no studies have detected any interactions between CagA, CagF and the *cag* T4SS machinery.

Structural characterization of T4SSs from multiple species

4

While *A. tumefaciens* provided the basic model of T4SSs, the most detailed structural work has been done in *E. coli.* The *E. coli* T4SS are utilized for conjugation and thus somewhat different than the *A. tumefaciens* effector molecule translocating system.

The first major subassembly of the T4SS that was described in detail was the core complex. In 2009 the Waksman group overexpressed the core complex components from the pKM101 plasmid (VirB7/TraN, VirB9/TraO, VirB10/TraF) and tagged the C-terminus of VirB10/TraF with a Strep tag (a small affinity tag), thus allowing for the selective purification of the core complex (16, 21). They were able to obtain a 15 Å cryoEM structure of the entire core complex (16), and a 2.6Å crystal structure of the upper "O-layer" of the core complex (21). These studies revealed the pKM101 core complex to be a large membrane-spanning channel with two chambers, a 14-fold symmetry and 18.5 nm in diameter. This group went on to employ a similar purification method but this time overexpressing the proteins VirB3-10 (VirB3/TrwM, VirB4/TrwK, VirB5/TrwJ,VirB6/TrwI, VirB7/TrwH, VirB8/TrwG, VirB9/TrwF and VirB10/TrwE) from the R388 conjugation plasmid (30). They were able to obtain a low-resolution, negative stain structure of this entire subassembly (30). These studies recapitulated their findings about the core complex, and went on to characterize the inner membrane complex (IMC) of the R388 T4SS. The IMC is composed of VirB3, 4, 5, 6, and 8, with a stalk-like domain connecting the core complex with the large barrel-like assemblies of the VirB4 ATPases (30).

Recently there has been work on characterization of a different T4SS core complex, namely the core complex from the Dot/Icm system in *Legionella pneumophila*. In contrast to the *E. coli* pKM101 and R388 systems the Dot/Icm T4SS used for effector protein translocation rather than bacterial conjugation. The publication reported the use of an approach that did not utilize affinity purifications to isolate T4SS subassemblies, but instead utilized differential centrifugation to isolate the core complex of the Dot/ICM T4SS (31). This study found that the core complex was composed of five proteins, DotG, H, C, F, and D. DotF corresponds to VirB10, DotH corresponds to VirB9 and DotD corresponds to VirB7, while DotC and G are specific to *Legionella*. The Dot/Icm core complex is far larger than the core complex of the *E. coli* conjugation system, about 38 nm in diameter rather than 18.5nm (31). However, is it also ring-like and contains a central channel. Its symmetry remains unknown, and thus cannot be compared to *E. coli's.*

There is clearly a large body of work describing T4SSs and their role in bacterial pathogenesis. However, T4SSs present in various bacterial species have a wide range of function and there is a high level of divergence among species in both composition of the systems and sequences of components. In particular the literature on the *cag* T4SS system is relatively sparse and therefore it is important to learn more about such an understudied and biologically relevant system.

CHAPTER II

ISOLATION OF A *cag* T4SS SUBASSEMBLY

Introduction

The *cag* T4SS is an important virulence factor present in large percentage of *H. pylori* strains. The *cag* T4SS is responsible for translocating the effector protein, CagA into the gastric epithelium. CagA itself is a *bona fide* oncoprotein (32). Transgenic mice engineered to constitutively express CagA developed tumors (33). Furthermore, epidemiological studies have shown that strains containing genes encoding the *cag* T4SS are correlated with severe disease outcomes (11). These results indicate that the *cag* T4SS is a significant virulence factor with a broad impact on public health.

Because of its relevance to public health, there has been a lot of interest in elucidating the architecture and the molecular mechanisms which regulate the *cag* T4SS. There has been a large body of work on other model systems such as *E. coli* to study and develop a basic model of how the T4SS functions, but the *cag* T4SS appears to be quite divergent from these prototypical systems. Only the two Cag proteins encoding the ATPase subunits (CagE and Cagα) and the coupling protein (Cagβ) have any significant level of sequence similarity with the model systems, and those major subunits tend to be the most highly conserved proteins of this machinery (24, 25). But even that similarity breaks down since CagE appears to be a fusion of the two prototypical ATPases, VirB3 and VirB4. Furthermore, most prototypical T4SSs require 12 proteins, whereas the

cag PAI encodes 27, clearly showing that this system is far more complex and different than other model systems.

The Cover lab has attempted to isolate subassemblies of the T4SS by using the same tagging strategies employed by the Waksman group, but to date these studies have not been successful. We have tagged other proteins that are integral members of the T4SS as determined by a requirement of these proteins for CagA translocation and IL-8 induction in an attempt to find another way to isolate components of the *cag* T4SS. To date we have tagged CagM, 3, X, V, H, I, L, U using the HA epitope tag. By tagging either CagH, I, or L, we were able to determine that these three proteins directly interacted with one another and they they were involved in pilus formation or regulation. However, the complex itself was never directly studied. The other proteins tagged did not yield protein complexes.

I decided to take an alternative and indirect approach to purify components of the *cag* T4SS. CagA is an intriguing candidate for this method: since CagA is translocated through the T4SS, a logical conclusion would be that it would physically interact with the T4SS. However, putting tags on this protein often abrogates its ability to translocate, meaning that it is likely no longer interacting with the T4SS. But there are also other proteins that interact with CagA which could then still take advantage of CagA's potential interaction with the *cag* T4SS. One such protein is CagF, which has been referred to as CagA's "chaperone" (27, 28). This protein is highly expressed and localized to both the cytosol and inner membrane of *H. pylori.* It has been called a chaperone since it is

required for the translocation of CagA and it has been confirmed as a binding

partner of CagA. My hopes were that CagF could act as a "handle" of sorts, to

purify larger complexes of the *cag* T4SS.

Materials and Methods

Bacterial strains and culture methods

Wild-type *H. pylori* strain 26695 and mutant strains were cultured on trypticase soy agar plates supplemented with 5% sheep blood, or on Brucella agar plates supplemented with 10% heat-inactivated fetal bovine serum (FBS), in room air supplemented with 5% CO_2. *H. pylori* mutant strains were selected in the presence of chloramphenicol (2.5 µg/mL), metronidazole (7.5 µg/mL), or streptomycin (50 µg/mL). Liquid cultures of *H. pylori* were grown in Brucella broth supplemented with 10% heat-inactivated FBS. *E. coli* strain DH5α was used for plasmid propagation and was grown on either Luria-Bertani (LB) agar plates or in LB broth supplemented with ampicillin (100 µg/mL).

Cell culture methods

AGS human gastric epithelial cells were grown in RPMI media supplemented with 10% FBS, 2 mM glutamine and 10 mM HEPES.

Generation of *H. pylori* strains producing HA-tagged CagF

To generate *H. pylori* strains producing HA-tagged forms of CagF, we first constructed the plasmids illustrated in **Fig. 1a**. These plasmids are modified forms of the pAD-1 plasmid (26, 34) that contain a full length copy of the *cagF* gene (along with sequences encoding an HA-epitope tag at the N-terminus or C-terminus of CagF), the *ureA* promoter and ribosomal binding site (RBS), and a chloramphenicol resistance cassette. These elements are flanked by sequences

from the *ureA* and *ureB* loci (26, 34). *H. pylori* strain 26695 or mutant strains were transformed with these plasmids (which do not replicate in *H. pylori*) and chloramphenicol-resistant transformants were selected. This approach allowed the introduction of sequences encoding HA-tagged CagF into the *H. pylori ureAB* chromosomal locus. *H. pylori* strains designated HA-CagF and CagF-HA produce CagF with N-terminal or C-terminal HA tags, respectively.

Isolation of the *H. pylori cag* T4SS core complex

H. pylori strains were grown in liquid culture for 16 h (resulting in values of optical density at 600 nm [OD_{600}] of 0.4 to 0.8). Bacterial cells were pelleted at 3,300 × *g* for 15 min at 4°C and resuspended in radioimmunoprecipitation assay (RIPA) buffer (1% NP-40, 0.25% deoxycholate, 10 mM HEPES, and 100 mM NaCl [pH 7.0] supplemented with 1 mM phenylmethylsulfonyl fluoride [PMSF] and protease inhibitors [Complete Mini Protease Inhibitor tablet; Roche]) for initial experiments or in RIPA buffer containing 300 mM NaCl for experiments involving electron microscopy and density gradients. The bacterial suspensions were sonicated on ice and incubated for 1 h at 4°C, and the bacterial lysates were then centrifuged (21,000 × *g*) to pellet insoluble material. Monoclonal anti-HA antibodies were noncovalently linked to protein G Dynabeads (Invitrogen) and incubated with the clarified bacterial lysates at room temperature for 30 min. Beads were washed three times at room temperature with the same buffers used for bacterial lysis, and proteins were selectively eluted with wash buffer containing 200 µg/ml HA peptide (YPYDVPDYA; GenScript). To isolate the core complex from *H. pylori* cocultured with gastric epithelial cells, subconfluent AGS cells were cocultured with *H. pylori*

(multiplicity of infection [MOI] of 100) for 4 h. After rinsing with phosphate-buffered saline (PBS) to remove nonadherent bacteria, adherent bacteria and epithelial cells were resuspended in RIPA buffer, and the protocol described above was used.

Analysis of IL-8 secretion

AGS human gastric epithelial cells were grown in RPMI medium (Corning) supplemented with 10% FBS, 2 mM glutamine and 10 mM HEPES and co-cultured with *H. pylori* strains at a multiplicity of infection (MOI) of 50 for four hours. IL-8 secretion was quantified using an anti-human IL-8 sandwich ELISA (R&D) (26, 34).

Western blotting

Proteins were separated on a 10% acrylamide gel, transferred to a nitrocellulose membrane, and immunoblotted using previously described rabbit polyclonal antisera to *H. pylori* Cag proteins (26), or antiserum to CagF that was generated by immunizing animals with a CagF-GST fusion protein. Horseradish peroxidase conjugated to anti-rabbit IgG (Promega) was used as a secondary antibody. Signals were generated by an enhanced chemiluminescent reaction and detected using x-ray film.

Silver stain

Proteins were separated on a 10% acrylamide gel and then stained using Pierce's silverstain kit.

Mass spectrometry analysis

Protein preparations were either run about 2 cm into a 10% Bis-Tris NuPAGE gel, stained with Colloidal Coomassie, and then subjected to in-gel trypsin digestion, or TCA-precipitated. For the latter, precipitated proteins were resolubilized in a solution of 8M urea, 100 mM Tris, pH 8.5, reduced with Tris(2-carboxyethyl)phosphine (TCEP, 10 mM), and alkylated with iodoacetamide (20 mM). Samples were then diluted back to 2M urea, trypsin added and allowed to digest overnight at 37°C. Single dimensional LC-MS/MS was performed using ThermoFisher LTQ equipped with a nano-electrospray source and attached to an Eksigent 1D+ or Nanoacuity (Waters) HPLC unit with an autosampler. Peptides were resolved via reversed phase separation on a 20 cm by 100 micron column packed emitter tip using an aqueous to organic gradient (2% to 45%). Peptide MS/MS spectra were acquired data-dependently with one full scan MS followed by 5 MS/MS scans. MudPITs were performed essentially as described previously (3, 4). Acidified peptides were loaded onto a 150 micron ID biphasic trapping column comprised of 4 cm strong cation exchange resin (LUNA 5 micron, Phenomonex) followed by 4 cm reverse phase resin (Jupiter 5 micron, 300 angstrom). The trapping column was then attached to a 20 cm (Jupiter 3 micron, 300 angstrom) 100 micron ID fused silica analytical column packed into a pulled nanospray tip. Preparations of eluted proteins and respective WT controls were analyzed by 8-step MudPIT. Salt pulses were performed by using the autosampler to inject 5 ul of ammonium acetate at 0, 100, 150, 200, 300, 500, 750, and 1000 mM concentrations. After each salt injection, peptides were separated using a 105

minute aqueous to organic gradient (2% to 35% acetonitrile (ACN) for all but the last step, which went to 98% ACN). Peptide MS/MS spectra were queried using SEQUEST (full tryptic specificity) against an *H. pylori* strain 26695 database, to which both common contaminants and reversed versions of the proteins had been appended. Resulting identifications were filtered to an estimated peptide false discovery rate (FDR) less than 5% and collated by protein using IDPicker 3.0. All reported proteins were identified based on detecting a minimum of 2 distinct peptides.

Results

<u>Purification of CagF and co-purification of additional Cag proteins</u>

Previous studies have shown that the *H. pylori* CagA effector protein interacts with another *cag* PAI-encoded protein, CagF (28, 29, 35). CagF is required for CagA translocation into eukaryotic host cells, and it has been proposed that CagF is a chaperone for CagA (28, 29, 35). We hypothesized that the CagF-CagA complex physically interacts with components of the *cag* T4SS machinery and that the CagF-CagA complex could be used as bait for purifying subassemblies of the *cag* T4SS. To test this hypothesis, we generated three *H. pylori* strains that produced hemagglutinin (HA) epitope-tagged forms of CagF; two (HA-CagF and Δ*cagF*/HA-CagF) produced CagF with an N-terminal tag, and one (CagF-HA) produced CagF with a C-terminal tag (**Fig. 1a and b**) The HA-CagF and CagF-HA strains retained the endogenous *cagF* gene (encoding an untagged form of CagF), whereas the Δ*cagF*/HA-CagF strain encoded only the tagged form of CagF. All three strains retained the ability to induce IL-8 secretion in human gastric epithelial (AGS) cells, indicating that the *cag* T4SS is functional in each of these strains (**Fig. 1c**).
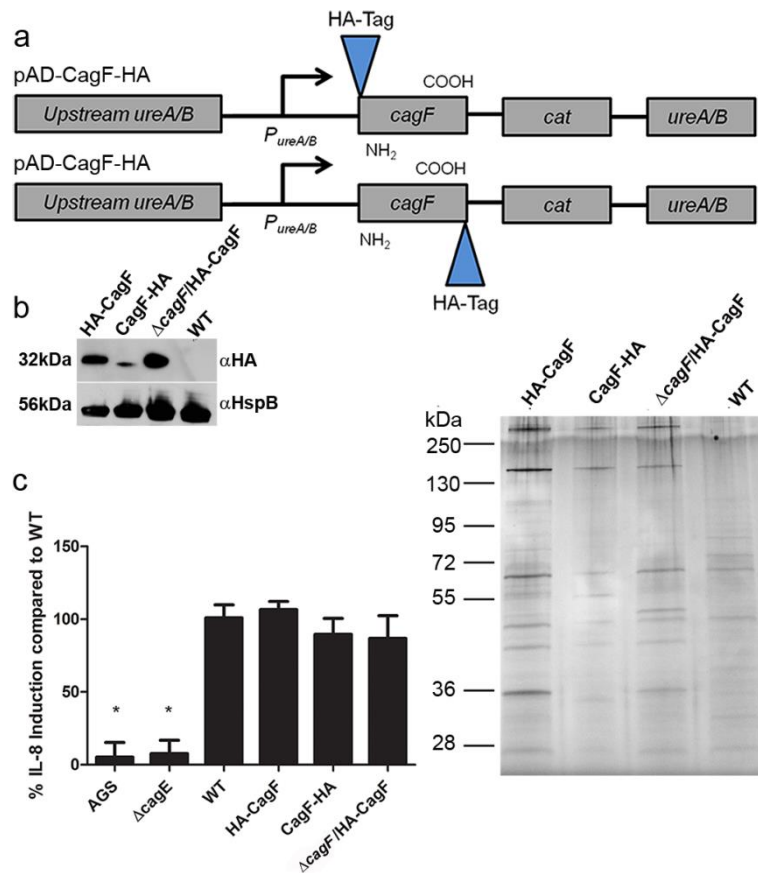
Figure 1: Generation of H. pylori strains that produce HA-tagged CagF. (a) Schematic of plasmids encoding epitope-tagged forms of CagF (with either an N-terminal or C-terminal epitope HA tag). The sequences were introduced into the H. pylori ureAB chromosomal locus, yielding strains designated HA-CagF or CagF-HA, respectively. The HA-CagF and Cag-HA strains retained the endogenous cagF gene (encoding an untagged form of CagF). We also generated a ΔcagF/HA-CagF strain, which encodes only the tagged form of CagF. In this strain, the endogenous copy of cagF was deleted, and a gene encoding HA-CagF was introduced into the urease locus. (b) Immunoblotting of the indicated strains with an anti-HA antibody showed the production of HA-tagged forms of CagF. (c) AGS cells were infected with the wild-type (WT) strain or the indicated strains producing HA-tagged CagF, and IL-8 production was quantified by ELISA. AGS indicates results for cells without added H. pylori. Values represent means ± SD, based on analysis of at least six replicate samples. Levels of IL-8 production induced by mutant strains were compared to levels induced by the WT strain (ANOVA followed by Dunn's multiple-comparison test). *, $P < 0.001$. (d) SDS-PAGE and silver stain analysis of preparations immunopurified from strains producing HA-tagged CagF (using a monoclonal HA antibody) and analysis of a control preparation from an untagged WT strain that was processed in parallel.

We immunopurified HA-tagged CagF from lysates of the three *H. pylori* strains using a monoclonal HA antibody, and an untagged wild-type (WT) strain was processed in parallel as a negative control. SDS-PAGE and silver stain analyses showed distinct proteins in preparations derived from the strains producing HA-tagged CagF that were absent in preparations from the untagged WT control strain (**Fig. 1d**). Immunoblotting showed that CagF and CagA were immunopurified from the strains producing HA-tagged CagF, whereas neither of these proteins was immunopurified from the untagged WT control strain (**Fig. 2**). We then immunoblotted the samples with a panel of antibodies against several components of the *cag* T4SS and found that Cag3, CagT, CagM, CagX, and CagY copurified with CagF and CagA (**Fig. 2**). Since similar results were obtained with all three strains producing HA-tagged CagF (**Fig. 2**), we conducted subsequent experiments using strains producing CagF with an N-terminal HA tag (either HA-CagF or Δ*cagF*/HA-CagF).
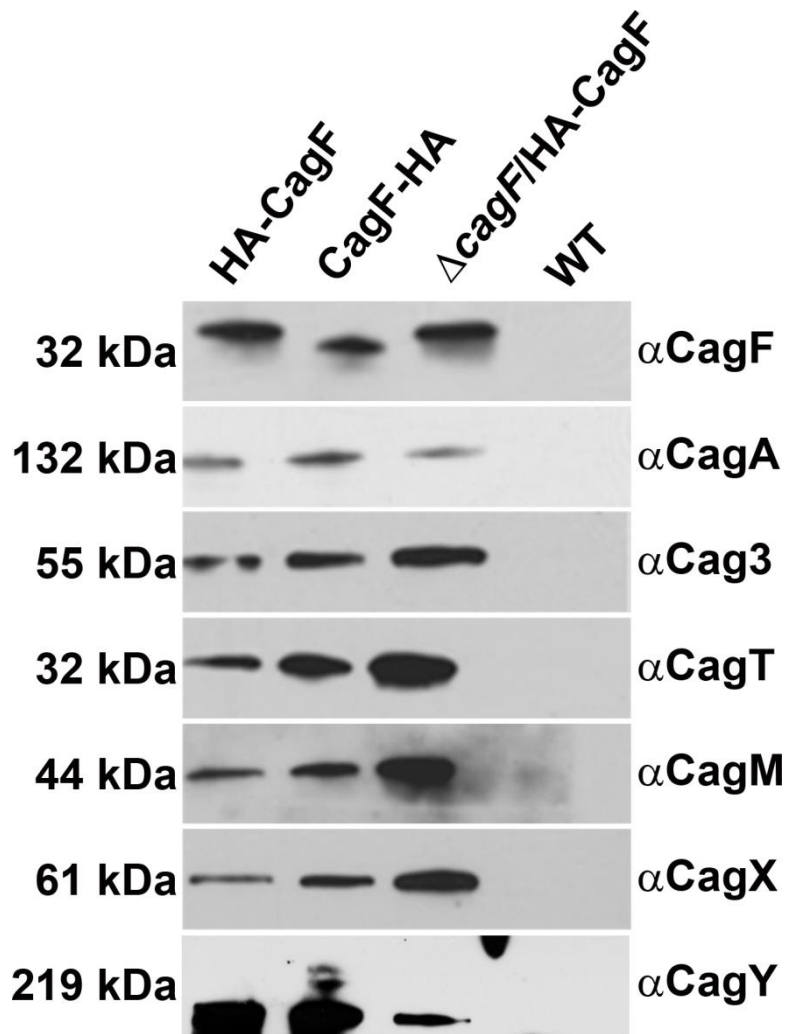
Figure 2: Analysis of proteins that copurify with CagF. Immunoblot analysis of preparations immunopurified from three strains producing HA-tagged CagF using a monoclonal HA antibody and of a control preparation from an untagged WT strain that was processed in parallel. Samples were immunoblotted with the indicated antisera. Results are representative of at least two independent experiments.

Analysis by mass spectrometry indicated that the seven Cag proteins described above (CagF, CagA, Cag3, CagT, CagM, CagX, and CagY) were the most abundant proteins in preparations derived from the HA-CagF-producing strain and that each of these proteins was enriched in preparations from the HA-CagF-producing strain compared to a control preparation from the untagged WT strain (**Table 1**). The copurification of several *H. pylori* proteins encoded by genes outside the *cag* PAI and produced at high levels by the bacteria (**Table 1**) most likely reflects nonspecific interactions.

Table 1: Proteins that co-purify with HA-CagF

Identified Proteins	Accession Number	Spectral Counts (HA-CagF[a])	Spectral Counts (WT[a])
CagA	HP0547	178	0
CagF	HP0543	56	1
CagX	HP0528	40	0
CagM	HP0537	38	0
Cag3	HP0522	44	0
CagT	HP0532	26	0
CagY	HP0527	18	0
Protease	HP1350	28	0
VacA	HP0887	16	8
GroEL	HP0010	12	0
Omp18	HP1125	6	0
EF-Tu	HP1205	5	1
hypothetical protein	HP0052	1	7
FrpB	HP0876	1	7
Total Cag Spectra		400	1
Total Non-Cag Spectra		69	23
Total Spectra		469	24

[a]HA-CagF was immunoaffinity purified from an *H. pylori* strain expressing this protein using a monoclonal anti-HA antibody, and a wild-type strain (WT) was processed in parallel as a control. The table shows numbers of spectral counts detected in each preparation by Multidimensional Protein Identification Technology (MudPIT).

<u>Purification of CagF in co-culture with gastric epithelial cells</u>

The immunopurifications described above were performed using lysates of *H. pylori* grown in liquid culture. CagA is not actively secreted by the bacteria under such conditions, but *H. pylori* contact with gastric epithelial cells triggers translocation of CagA into the epithelial cells (12, 36). We hypothesized that *H. pylori* contact with gastric epithelial cells leads to activation of the *cag* T4SS, possibly resulting in altered interactions of the CagF-CagA complex with T4SS components. Therefore, we cocultured AGS gastric epithelial cells with HA-CagF-producing *H. pylori* (Δ*cagF*/HA-CagF) and purified HA-CagF from lysates of the coculture. AGS cells cocultured with untagged WT bacteria were processed in parallel as a negative control. Mass spectrometry analysis revealed that preparations isolated from the coculture contained the same seven Cag proteins detected in the earlier experiments (**Table 2**). Low levels of several other Cag proteins were also detected (**Table 2**). Importantly, Cag3, CagT, CagM, CagX, and CagY reproducibly co-purified with CagF and CagA in experiments conducted with either bacterial cell-epithelial cell co-cultures or bacterial liquid cultures.

Table 2: Proteins that co-purify with HA-CagF when H. pylori ΔcagF/HA-CagF is co-cultured with AGS cells

Identified Proteins	Accession Number	Assigned Spectral Counts (ΔcagF/HA-CagF[a])	Assigned Spectral Counts (WT[a])
cag pathogenicity island protein Cag3	HP0522	58	0
cag pathogenicity island protein CagY	HP0527	165	1
cag pathogenicity island protein CagX	HP0528	86	0
cag pathogenicity island protein CagW	HP0529	2	1
cag pathogenicity island protein CagV	HP0530	7	0
cag pathogenicity island protein CagT	HP0532	36	0
cag pathogenicity island protein CagM	HP0537	50	0
cag pathogenicity island protein CagL	HP0539	2	0
cag pathogenicity island protein CagI	HP0540	3	0
cag pathogenicity island protein CagG	HP0542	3	1
cag pathogenicity island protein CagF	HP0543	51	2
cag pathogenicity island protein CagE	HP0544	12	0
cag pathogenicity island protein CagA	HP0547	882	11
chaperone and heat shock protein (groEL)	HP0010	23	16
co-chaperone (groES)	HP0011	2	0
outer membrane protein (omp2)	HP0025	5	3
citrate synthase (gltA)	HP0026	13	3
urease accessory protein (ureG)	HP0068	3	3
urease beta subunit (urea amidohydrolase) (ureB)	HP0072	2	2
urease, alpha subunit (ureA) [3.5.1.5]	HP0073	1	3
methyl-accepting chemotaxis transducer (tlpC)	HP0082	1	21
ribosomal protein S9 (rpsI)	HP0083	2	2
conserved hypothetical protein	HP0086	2	0
chaperone and heat shock protein 70 (dnaK)	HP0109	4	4
DNA topoisomerase I (topA)	HP0116	1	1
ribosomal protein L20 (rplT)	HP0126	1	3
hypothetical protein	HP0130	3	1
cell binding factor 2	HP0175	6	3
translation elongation factor EF-P (efp)	HP0177	3	0
chaperone and heat shock protein C62.5 (htpG)	HP0210	1	1
peptide methionine sulfoxide reductase (msrA) [1.8.4.6]	HP0224	1	1
hypothetical protein	HP0231	0	2
ATP-dependent RNA helicase, DEAD-box	HP0247	4	4

family (deaD)			
conserved hypothetical protein	HP0248	2	1
ATP-dependent nuclease (addB)	HP0275	2	0
ribosomal protein L21 (rplU)	HP0296	4	3
hypothetical protein	HP0305	4	4
outer membrane protein (omp9)	HP0317	6	5
conserved hypothetical protein	HP0318	2	1
cell division inhibitor (minD)	HP0331	1	7
adhesin-thiol peroxidase (tagD)	HP0390	5	4
ribosomal protein S1 (rpsA)	HP0399	5	2
oligoendopeptidase F (pepF)	HP0470	3	1
conserved hypothetical secreted protein	HP0506	2	0
ribosomal protein L9 (rplI)	HP0514	2	0
transcription termination factor Rho (rho)	HP0550	3	1
ribosomal protein L31 (rpmE)	HP0551	1	1
acyl carrier protein (acpP)	HP0559	2	0
3-ketoacyl-acyl carrier protein reductase (fabG)	HP0561	2	0
aminopeptidase a-i (pepA)	HP0570	3	0
ferredoxin oxidoreductase, alpha subunit	HP0589	2	2
lipoprotein, putative	HP0596	1	1
hemolysin secretion protein precursor (hylB)	HP0599	3	0
tetrahydrodipicolinate N-succinyltransferase (dapD)	HP0626	4	1
aspartate ammonia-lyase (aspA)	HP0649	0	2
processing protease (ymxG)	HP0657	4	5
N-methylhydantoinase	HP0696	1	3
hypothetical protein	HP0720	2	0
hypothetical protein	HP0721	5	4
phosphoribosylpyrophosphate synthetase (prsA)	HP0742	1	3
conserved hypothetical protein	HP0747	1	2
aconitase B (acnB) [4.2.1.3]	HP0779	2	0
preprotein translocase subunit (secA)	HP0786	3	1
iron(III) dicitrate transport protein (fecA)	HP0807	4	0
thioredoxin (trxA)	HP0824	5	2
thioredoxin reductase (trxB)	HP0825	2	0
inosine-5'-monophosphate dehydrogenase (guaB)	HP0829	0	2
catalase [1.11.1.6]	HP0875	19	5
vacuolating cytotoxin	HP0887	36	17
hydrogenase expression-formation protein (hypB)	HP0900	5	0

outer membrane protein (omp20)	HP0912	4	1
hypothetical protein	HP0958	2	1
polyphosphate kinase (ppk)	HP1010	1	5
protease (pqqE)	HP1012	4	2
serine protease (htrA)	HP1019	6	4
ferric uptake regulation protein (fur)	HP1027	1	2
acetyl-CoA synthetase (acoE)	HP1045	1	1
translation initiation factor IF-2 (infB)	HP1048	1	1
gamma-glutamyltranspeptidase (ggt) [2.3.2.2]	HP1118	1	4
peptidoglycan associated lipoprotein precursor (omp18)	HP1125	2	1
ATP synthase F1, subunit beta (atpD)	HP1132	7	7
ATP synthase F1, subunit alpha (atpA)	HP1134	2	3
ribosomal protein L19 (rpl1S)	HP1147	7	5
flavodoxin (fldA)	HP1161	10	5
outer membrane protein (omp27)	HP1177	2	0
carbonic anhydrase	HP1186	2	1
translation elongation factor EF-G (fusA)	HP1195	7	0
ribosomal protein S7 (rpsG)	HP1196	3	3
ribosomal protein S12 (rpsL)	HP1197	0	2
DNA-directed RNA polymerase, beta and beta' subunit (rpoBC)	HP1198	16	16
ribosomal protein L7-L12 (rplL)	HP1199	2	0
ribosomal protein L1 (rplA)	HP1201	0	2
translation elongation factor EF-Tu (tufB)	HP1205	21	10
DNA-directed RNA polymerase, alpha subunit (rpoA)	HP1293	3	1
ribosomal protein S4 (rpsD)	HP1294	2	1
ribosomal protein S11 (rpsK)	HP1295	0	2
ribosomal protein S13 (rpsM)	HP1296	5	4
ribosomal protein L15 (rplO)	HP1301	3	2
ribosomal protein L6 (rplF)	HP1304	4	2
ribosomal protein S8 (rpsH)	HP1305	2	2
ribosomal protein L24 (rplX)	HP1308	0	2
ribosomal protein L14 (rplN)	HP1309	1	1
ribosomal protein S3 (rpsC)	HP1313	1	1
ribosomal protein L2 (rplB)	HP1316	8	5
ribosomal protein L3 (rplC)	HP1319	4	7
ribosomal protein S10 (rpsJ)	HP1320	3	1
hypothetical protein	HP1334	3	0
protease	HP1350	11	6
conserved hypothetical protein	HP1414	2	1
conserved hypothetical ATP-binding protein	HP1430	5	14

peptidyl-prolyl cis-trans isomerase B, cyclosporin-type rotamase (ppi)	HP1441	1	1
gerC2 protein (gerC2)	HP1483	1	1
iron-regulated outer membrane protein (frpB)	HP1512	2	2
iron(III) ABC transporter, periplasmic iron-binding protein (ceuE)	HP1561	3	1
iron(III) ABC transporter, periplasmic iron-binding protein (ceuE)	HP1562	4	1
alkyl hydroperoxide reductase (tsaA)	HP1563	12	10
outer membrane protein	HP1564	0	2
Total Cag Spectra		1357	16
Total Non-Cag Spectra		409	299
Total Spectra		1766	315

[a] ΔcagF/HA-CagF was immunoaffinity purified from an *H. pylori* strain expressing this protein using a monoclonal anti-HA antibody, and a wild-type strain (WT) was processed in parallel as a control. The protein content of preparations was analyzed by Multidimensional Protein Identification Technology (MudPIT). The list of proteins identified was filtered via IDPicker 3.0 using a 5% peptide false discovery rate (FDR) and a minimum of 2 distinct peptides per protein. The table shows numbers of assigned spectral counts.

Discussion

In this study, I isolated the putative core complex of the *cag* T4SS. Our lab has tried to isolate large subassemblies of the *cag* T4SS by tagging Cag proteins associated with various regions of the *cag* T4SS. We tagged inner membrane core members CagV and CagU, but were only able to purify the protein that was directly tagged (data not shown). We then tagged three proteins that have no homology to other proteins in T4SSs, Cag H, I, and L. We were able show that these three proteins were interacting with one another, and involved in regulation or formation of the pilus, but no further efforts were made to analyze the complex itself. And then finally, we tagged proteins predicted to be part of the core complex, Cag3, M, X, and Y. These attempts yielded similar results to the inner membrane proteins: we could only isolate the tagged proteins.

It was clear that we needed to take a different approach in purifying modules of the *cag* T4SS. I hypothesized that we needed to use a more indirect approach; perhaps the affinity association with the tag was disrupting the protein-protein interactions of this complex macromolecular machine. One potential protein that could accomplish this could be CagA, the only known effector protein of the *cag* T4SS. Since it requires the T4SS for its translocation, it has the potential to physically interact with the machinery, though this interaction has never been shown. However, whenever our lab has tried to tag CagA, the protein lost its ability to be translocated. This suggests that the tag is interfering with the ability of CagA to interface with the T4SS, thus making it unsuitable for my purposes.

However, there was another protein that could utilize the indirect interaction between CagA and the T4SS. This protein is CagF, which is the chaperone protein of CagA (28, 29). Several studies have shown that this protein physically interacts with CagA, and is required for CagA to be successfully translocated into eukaryotic cells. Furthermore, when CagF is tagged, CagA is still translocated, meaning that there would be no disruption of the interaction that I sought to take advantage of.

I tagged CagF with an HA (HA-CagF) epitope to enable a specific purification and elution from *H. pylori.* I immunopurified CagF from bacterial lysates and was also able to isolate its cognate binding partner, the effector protein CagA, confirming the efficacy of my IP. However, I was also able to isolate five other Cag proteins: Cag3, T, M, X and Y. This was a specific purification, since I was unable to isolate any of these proteins from an untagged strain of *H. pylori.* Those five proteins, Cag3, T, M, X, and Y correspond to the putative core complex of the *cag* T4SS. This is the first time that the core complex has been purified from *H. pylori*, and also the first time that an interaction between the effector protein CagA and the *cag* T4SS has been observed.

There are several assays to assess the activity of the *cag* T4SS. These assays are grouped into two major categories: one, where the translocation of CagA is determined, and the other where the actual formation of the T4SS is assessed.

There are two ways to determine if CagA has been translocated. First, when CagA is translocated into eukaryotic cells, it is phosphorylated by the host cell kinase Src (12). Thus, an immunoblot using specific phospho-antibodies can be used to determine is CagA has been phosphorylated and therefore, translocated. The second method is far less quantitative and relies on cell morphology. When CagA is translocated into host cells, it induces actin cytoskeletal rearrangements, causing the cell to form long, thin protrusions. These are referred to as the "hummingbird" phenotype (37).

There are several different methods to assess formation of the T4SS. These assays will still produce a positive result in the absence of CagA, since the *cag* T4SS can still form without its effector protein. When the T4SS is elaborated, it is hypothesized that it requires the use of a peptidoglycan hydrolase (Cag4), to digest away the thin layer of peptidoglycan that exists between the inner and outer membrane so the pilus can be formed in the extracellular space (15). It is hypothesized that this process will introduce peptidoglycan into the host cells when the pilus interfaces with the eukaryotic cell. The introduction of this bacterial polymer will strongly induce innate immune signaling through NOD1 leading to the robust de-repression of the master-regulator NF-κB (15). This gene's upregulation can be detected via a reporter cell line wherein NF-κB has been fused with luciferase. One of the major genes that is upregulated by this increase in NF-κB expression is the host cytokine, IL-8. This provides another simple readout for *cag* T4SS function, where IL-8 production can be measured and quantified through ELISA. The final approach to measure *cag* T4SS activity is through

28

microscopy. Utilizing SEM, the actual elaboration of the pilus structure in contact with eukaryotic cells can be visualized (34).

The various assays described above show an array of diverse techniques that take advantage of the multitude of perturbations that the *cag* T4SS causes. However, all of these assays require the presence of eukaryotic cells and we have been unable to assess the functionality of the *cag* T4SS from pure bacterial culture. The results reported here indicate that the core complex is assembled prior to *H. pylori* contact with host cells and CagA translocation. We speculate that CagA is bound to assembled core complexes, ready to be translocated as soon as the bacteria receive the appropriate signal, making the act of translocation more efficient. Thus far, the molecular mechanisms that regulate CagA secretion (i.e., inhibition of secretion during bacterial growth in broth culture and activation of secretion when *H. pylori* contacts gastric epithelial cells) remain poorly understood. But these results represent an important first step in understanding them, as well as the first time that there has been a reported interaction between CagA and a major structural component of the *cag* T4SS.

I also wanted to investigate the possibility that the *cag* T4SS might be reorganized when the bacteria are in contact with host cells. To do this I performed my immunoprecipitation from bacteria co-cultured with gastric epithelial cells. There are many other components besides the core complex that form the *cag* T4SS. Two other major subassemblies are the ATPases which are predicted to form large, multimeric complexes, an inner membrane core which connects the core complex and the ATPases, and finally the pilus which protrudes into the

extracellular space. Since the pilus is a component that only appears to form when the bacteria are in contact with host cells, and the ATPases are required for pilus formation, I hypothesized that I could purify these components from a co-culture model. Based on mass spec results, I purified the same seven proteins (CagF, A, 3, T, M, X, and Y) that I did from the experiments done in pure culture. There were other Cag proteins identified in this preparation, however, their spectral counts were so low that is it unlikely these hits are actually significant.

Overall, this is the first time that the predicted core complex of the *cag* T4SS has ever been isolated. This study is a key initial step in characterizing an important virulence factor of *H. pylori,* which has presented a considerable challenge to both our lab and others*.*

CHAPTER III

COMPOSTION OF THE CORE COMPLEX

Introduction

In the previous chapter I showed that I could isolate the predicted members of the *cag* T4SS core complex. The purification was a specific one and used the chaperone protein CagF. The same seven proteins were isolated (CagF, A, 3, T, M, X, and Y) if the immunoprecipitation was performed from pure bacterial lysates or from *H. pylori* co-cultured with gastric epithelial cells.

This result represents an important first step in defining the structure of the *cag* T4SS. Previous studies of the core complex and subassmblies of the *cag* T4SS focused on the protein-protein interactions of the various components. Work from the Cover lab in 2006 used a yeast-two-hybrid (Y2H) system to probe for different interactions among Cag proteins (38). Of the predicted core complex proteins, they found that CagX and the C-terminus of CagY strongly interacted in a reciprocal manner, and to a lesser extent CagM and CagX interacted, and Cag3 and CagT interacted. However, a Y2H for bacteria is not a very biologically relevant assay since bacterial proteins are expressed in the highly divergent eukaryotic system and are not localized to their normal compartments. Y2H studies mainly show that proteins *can* interact but do not indicate if they *do* interact.

Several years later, this study was followed up by two more studies in 2013 in which the interactions were assessed in *H. pylori*. One study by Pinto-Santini *et al*, focusing on the protein Cag3, performed immunoprecipitations with a polyclonal Cag3 antibody (39). Their work indicated that Cag3 seemed to be interacting with a large majority of Cag proteins, namely CagM, T, X, Y, D, 1, α, β, A, F, C, and E. However, only the interactions with CagT and CagM were confirmed with alternative approaches; the interaction with CagT was detected with a reciprocal IP, and the interaction with CagM was detected with cross-linking. It is highly likely that many of the other interactions could be non-specific ones due to the use of a polyclonal antibody. The other study was performed by Kutter et al, who reported IP experiments indicating that CagY interacted with CagX, CagX interacted with CagM and that CagM interacted with CagT. All of those interacted were confirmed by reciprocal IP (35).

These studies give a snapshot of the various protein interactions, but hopefully my studies of the entire protein complex can provide a more complete picture, instead of analyzing only one or two binding partners. Furthermore, I would like to be able to purify it in an alternative manner.

In this chapter, I describe experiments deigned to study the core complex composition in greater detail, to determine the role that any other Cag proteins might have in the formation of the core complex, and to further define the protein-protein interactions that make up the core complex itself.

Materials and Methods

Isolation of the *H. pylori cag* T4SS core complex using a FLAG

immunoprecipitation

H. pylori strains were grown in liquid culture for 16 h (resulting in values of

optical density at 600 nm [OD_{600}] of 0.4 to 0.8). Bacterial cells were pelleted at

3,300 × *g* for 15 min at 4°C and resuspended in radioimmunoprecipitation assay

(RIPA) buffer (1% NP-40, 0.25% deoxycholate, 10 mM HEPES, and 100 mM NaCl

[pH 7.0] supplemented with 1 mM phenylmethylsulfonyl fluoride [PMSF] and

protease inhibitors [Complete Mini Protease Inhibitor tablet; Roche]). The bacterial

suspensions were sonicated on ice and incubated for 1 h at 4°C, and the bacterial

lysates were then centrifuged (21,000 × *g*) to pellet insoluble material. Monoclonal

anti-FLAG antibodies were noncovalently linked to protein G Dynabeads

(Invitrogen) and incubated with the clarified bacterial lysates at 4°C for 2 hr. Beads

were washed three times at room temperature with the half-strength RIPA buffer,

and proteins were selectively eluted with wash buffer containing 200 µg/ml FLAG

peptide (DYKDDDDK; GenScript).

Separation of proteins by gradient centrifugation

15% and 45% solutions of glycerol were prepared in RIPA buffer containing

300 mM NaCl. A 15-45% 5 mL gradient was poured using a gradient mixer (Jules

INC.). Core complex preparations isolated by immunoaffinity methods were

applied to the top of the gradient, and the gradient was then centrifuged in a SW

55-Ti rotor for 16 hours at 15,000 rpm at 4°C. Fractions (each 625 µl) were

collected from the top of the gradient, and were subsequently analyzed by immunoblotting and mass spec.

Generation of *H. pylori* strains producing FLAG-tagged CagT and other mutant strains

Unmarked Δ*cagX,* Δ*cag3,* Δ*cagM,* mutant strains have been described previously (26). Unmarked Δ*cagT,* Δ*cagF,* and *ΔcagY* mutant strains were generated using previously described contra-selection methods involving a *cat-rdxA* cassette (34) or a *cat-rpsL* cassette (26), and a marked *cagA* mutant strain (*cagA::kan*) was generated by inserting a kanamycin resistance cassette. These mutant strains were further modified by the introduction of sequences encoding HA-CagF into the *ureAB* chromosomal locus, thereby allowing expression of CagF with an N-terminal HA tag.

To make the FLAG tagged CagT construct, a point mutant were generated where a FLAG tag was inserted after the 27th amino acid of the protein using the QuikChange Multi site-directed mutagenesis kit (Agilent) and then transformed into a cagT::cat/rdxA strain.

Results

Isolation of two distinct Cag protein complexes

To investigate whether the proteins co-purifying with CagF were constituents of one or more protein complexes, we analyzed immunopurified samples prepared as described above by velocity sedimentation in glycerol density gradients. Immunoblot analysis indicated that Cag3, CagT, CagM, CagX, and CagY co-sedimented in the bottom fractions of the gradient (i.e., in the high-molecular-mass fractions) (**Fig. 3a**). This result provides evidence that Cag3, CagT, CagM, CagX, and CagY are components of a large protein complex. CagA and CagF were detected in the first fraction (taken from the top of the gradient), separately from the other Cag proteins, suggesting that CagF and CagA dissociate from the larger complex during gradient centrifugation. Mass spectrometry analysis of the gradient fractions (**Fig. 3b**; see also **Table 3**) confirmed that Cag3, CagT, CagM, CagX, and CagY were all distributed in the same gradient fractions (supporting the conclusion that these are components of a protein complex), whereas CagF and CagA were found predominantly in fractions from the top of the gradient (**Fig. 3b**). Each of the non-Cag proteins present in the sample was detected predominantly in the first fraction (taken from the top of the gradient) instead of the fractions containing Cag3, CagT, CagM, CagX, and CagY (**Fig. 3b**). Thus, the stringent conditions of the glycerol density gradient allowed purification to homogeneity of a large stable complex of Cag proteins distinct from CagA, CagF, and non-Cag proteins (**Fig. 3b**; see also **Table 3**)

35
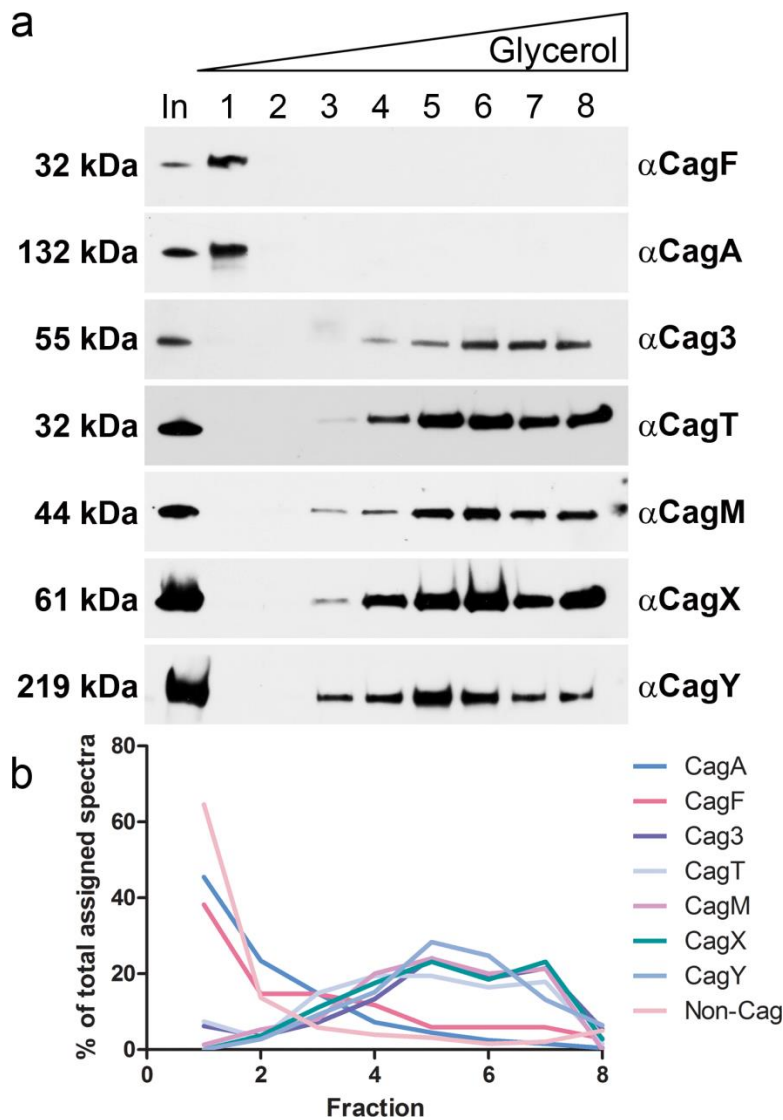
Figure 3: Sedimentation of *H. pylori cag* T4SS core complex in density gradients. (a) Preparations resulting from immunoaffinity purification of HA-CagF were analyzed by velocity sedimentation in glycerol density gradients, and gradient fractions (collected from the top of the gradient) were immunoblotted with the indicated antisera. In, input sample before it was applied to the gradient. Results are representative of two independent experiments. (b) The protein content of gradient fractions was analyzed by 1D mass spectrometry (see **Table 3** for complete results). The distribution of CagF, CagA, Cag3, CagT, CagM, CagX, CagY, and total non-Cag proteins in individual gradient fractions is shown.

Table 3: Protein composition of fractions resulting from the sedimentation of an HA-CagF-immunopurified sample in a density gradient

Identified proteins	Fractions								Total Assigned Spectra
	1	2	3	4	5	6	7	8	
CagF	13	5	5	4	2	2	2	1	34
CagA	202	104	66	32	20	11	7	2	444
Cag3	7	4	8	15	27	21	24	6	112
CagT	5	2	10	13	13	11	12	1	67
CagM	1	4	6	15	18	15	16	0	75
CagX	0	4	12	19	25	20	25	3	108
CagY	0	7	23	38	71	62	33	16	250
chaperone and heat shock protein (groEL)	24	10	8	1	1	1	0	2	47
protease	13	6	2	0	0	0	0	0	21
alkyl hydroperoxide reductase (tsaA)	11	4	2	0	0	0	0	0	17
translation elongation factor EF-Tu (tufB)	21	7	1	0	0	0	0	0	29
outer membrane protein (omp2)	12	0	1	1	0	0	1	1	16
hydantoin utilization protein A (hyuA)	7	1	1	0	0	1	0	1	11
DNA-directed RNA polymerase, beta and beta' subunit (rpoBC)	6	2	1	1	5	0	4	2	21
superoxide dismutase (sodB)	6	3	1	2	1	0	1	0	14
outer membrane protein (omp32)	4	1	1	0	0	0	0	0	6
excinuclease ABC subunit B (uvrB)	3	0	1	0	0	0	0	1	5
vacuolating cytotoxin	3	0	1	1	0	0	0	0	5
S-adenosylmethionine synthetase 2 (metX)	3	0	1	0	0	0	0	0	4
hypothetical protein	1	1	1	0	0	0	0	3	6
chaperone and heat shock protein 70 (dnaK)	14	3	0	0	0	0	0	0	17
iron-regulated outer membrane protein (frpB)	13	3	0	0	0	0	0	0	16
catalase [1.11.1.6]	11	3	0	0	0	0	0	0	14
hypothetical protein	11	2	0	0	1	1	0	0	15
translation elongation factor EF-G (fusA)	9	1	0	1	0	0	0	1	12
outer membrane protein (omp4)	8	0	0	1	0	0	0	0	9
adhesin-thiol peroxidase (tagD)	6	2	0	0	0	0	0	0	8

	1	2	3	4	5	6	7	8	Total
cell binding factor 2	6	0	0	0	0	0	0	0	6
outer membrane protein (omp27)	5	0	0	0	0	0	0	0	5
thioredoxin reductase (trxB)	5	0	0	0	0	0	0	0	5
conserved hypothetical protein	4	1	0	0	0	0	0	1	6
lipoprotein, putative	4	0	0	0	0	0	0	0	4
peptidoglycan associated lipoprotein precursor (omp18)	3	0	0	2	3	1	1	0	10
protective surface antigen D15	3	0	0	0	0	1	0	1	5
aminopeptidase a-i (pepA)	3	0	0	1	0	0	0	1	5
flavodoxin (fldA)	3	1	0	1	0	0	0	0	5
ribosomal protein L1 (rplA)	3	0	0	0	0	0	0	0	3
ribosomal protein S1 (rpsA)	3	0	0	0	0	0	0	0	3
outer membrane protein (omp20)	3	0	0	0	0	0	0	0	3
outer membrane protein (omp9)	3	0	0	0	0	0	0	1	4
outer membrane protein (omp31)	3	0	0	0	0	0	0	0	3
putative neuraminyllactose-binding hemagglutinin homolog (hpaA)	3	0	0	0	0	0	0	0	3
conserved hypothetical secreted protein	3	0	0	0	0	0	0	0	3
threonyl-tRNA synthetase (thrS) [6.1.1.3]	2	1	0	3	0	0	0	1	7
ribosomal protein S3 (rpsC)	0	0	0	0	1	1	1	3	6

[a] HA-CagF was immunoaffinity purified from an *H. pylori* strain expressing this protein using a monoclonal anti-HA antibody, and the preparation was then subjected to velocity sedimentation in a glycerol density gradient. The protein contents of gradient fractions (collected from the top of the gradient) were analyzed by 1-D mass spectrometry. The numbering along the top of the table refers to fraction number. The table shows numbers of spectral counts detected in each of the eight fractions.

Purification of CagT and co-purification of additional Cag proteins

The previously described experiments indicated that through the IP of HA-CagF I can isolate the core complex of the *cag* T4SS system, as well at the CagF-CagA complex. However, I wanted further confirmation of these results, so I sought out another way to purify the core complex. Our lab has attempted to purify subassemblies of the *cag* T4SS by directly tagging proteins which are a part of it with llimited success. A former labmate, Bradley Voss was attempting to identify surface exposed residues of CagT and thus looked for regions of low hydrophobicity within the protein. He identified eight regions of low hydrophobicity and made eight different constructs where a FLAG epitope was inserted into each of these regions, and then expressed the resulting FLAG-tagged forms of CagT from the urease locus.

Of these eight different strains, only three retained their ability to induce IL-8 production in gastric epithelial cells, indicating that the introduction of the FLAG epitope had not disrupted T4SS function. Of these three constructs, I chose ΔCagT/ureA::27K (T/27KFLAG) since it had the highest basal level of CagT expression compared to the other two. I then immunopurified T/27KFLAG from bacterial lysates. Immunoblotting showed that CagT was immunopurified from the strains producing FLAG-tagged CagT, and not from the untagged WT control (**Fig. 4**). I then immunoblotted the samples with a panel of antibodies against several components of the *cag* T4SS and found that Cag3, CagM, CagX, and CagY copurified with CagT (**Fig. 4**).
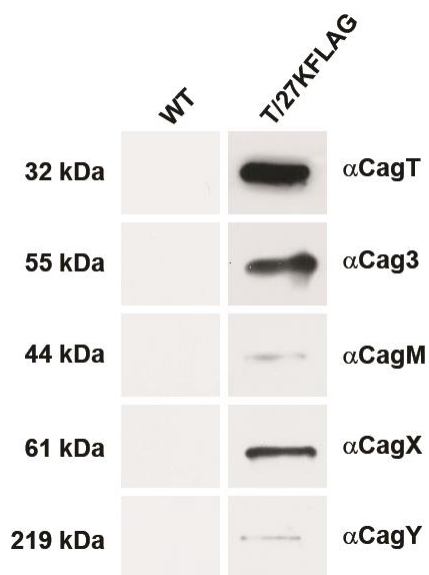
Figure 4: The FLAG-tagging of CagT allows for purification of the *cag* core complex. Immunoblot analysis of preparations immunopurified from a strain producing FLAG-tagged CagT using a monoclonal FLAG antibody and of a control preparation from an untagged WT strain that was processed in parallel. Samples were immunoblotted with the indicated antisera.

Analysis of complexes formed by strains containing mutations in *cag* genes
encoding integral components of the core complex

We hypothesized that several of the Cag proteins identified in the
immunopurification experiments would be required for assembly or stability of the
core complex, whereas others would be nonessential for core complex assembly
(despite being required for T4SS function). For example, the Cag proteins that
have orthologs in T4SSs of multiple bacterial species are likely to be required for
core complex formation, whereas proteins unique to *H. pylori* (Cag3, CagT, and
CagM) might be dispensable. To test this hypothesis, we analyzed mutant strains
in which *cagY*, *cagX*, *cagT*, *cag3*, and *cagM* were individually deleted, and each
mutant was engineered to produce HA-CagF, thereby permitting isolation of the
core complex. Consistent with the results of previous studies (14), these five
mutant strains were unable to induce IL-8 secretion in AGS cells (indicating the
absence of a functional *cag* T4SS) (**Fig. 5a**). We also generated a *cagA* mutant
strain that expressed HA-CagF, and, consistent with previous results (14), this
strain retained the capacity to induce IL-8 secretion (**Fig. 5a**). We then sought to
isolate Cag protein complexes from the mutant strains by immunopurifying HA-
tagged CagF (using the method shown in **Fig. 1** and **2**) and analyzed the resulting
preparations by immunoblotting. In experiments performed with Δ*cagX* or Δ*cagY*
mutants, we successfully purified CagF and CagA but were unable to copurify core
complex components (**Fig. 5b** and **c**), despite evidence that these proteins were
present in lysates from these mutants (see **Fig. 5c**). Similarly, in an experiment
performed with the *cagA* mutant strain, we successfully purified CagF but were

unable to copurify any components of the core complex (**Fig. 5b** and **c**). In

contrast, we were able to immunopurify multiple Cag proteins as well as CagF and

CagA from the Δ*cag3*, Δ*cagT*, *and* Δ*cagM* mutant strains. In experiments

performed with the Δ*cag3* strain, all components of the core complex (except

Cag3) copurified with CagF and CagA. Preparations immunopurified from the

Δ*cagT* strain lacked Cag3 and CagT, and preparations immunopurified from the

Δ*cagM* strain lacked Cag3, CagT, and CagM (**Fig. 5b** and **c**). These experiments

provided evidence that incomplete protein complexes can assemble in the

absence of Cag3, CagT, or CagM (**Fig.5b** and **c**).
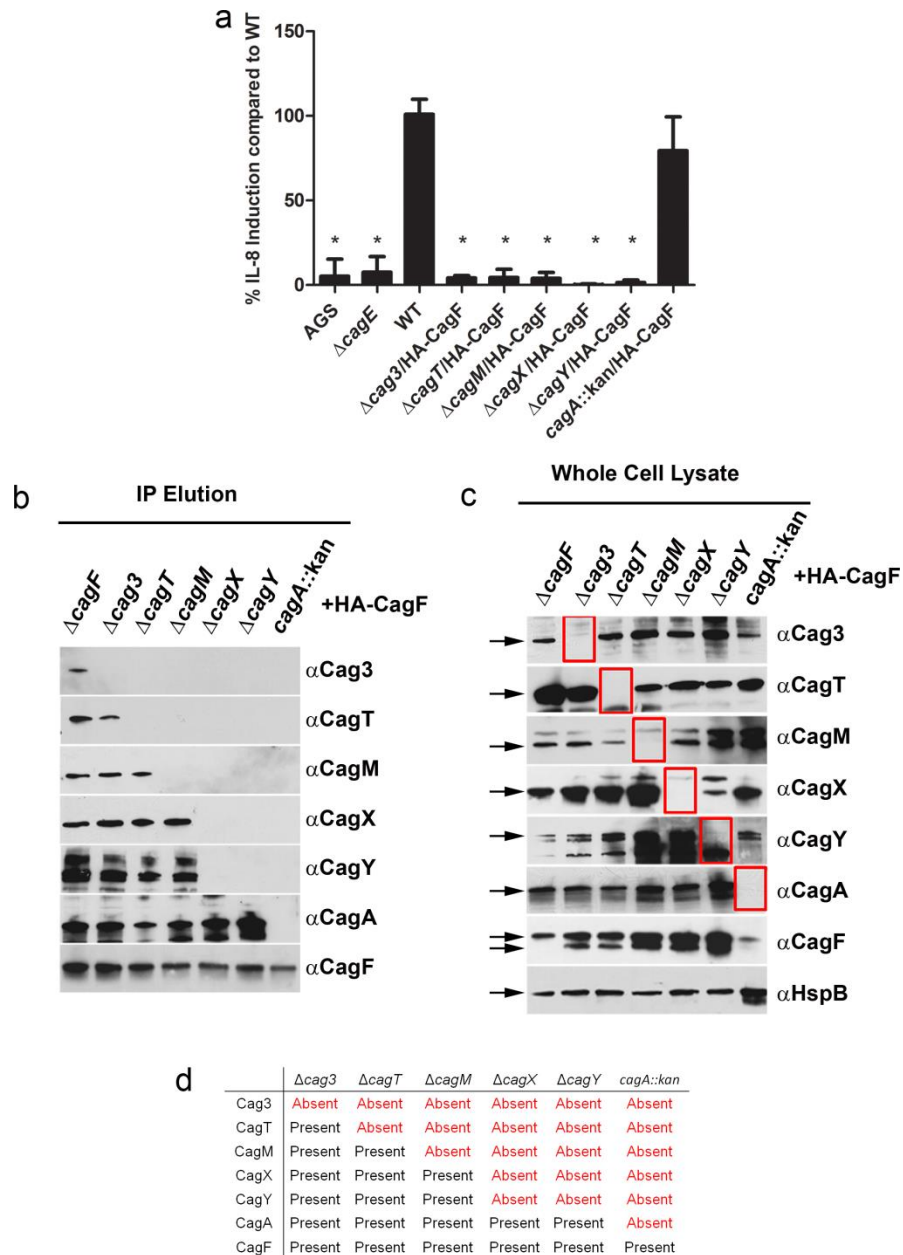
Figure 5: Immunopurification of HA-CagF from a panel of *cag* mutant strains. (a) AGS cells were cocultured with the WT strain or the indicated mutants, and IL-8 secretion was quantified by enzyme-linked immunosorbent assay (ELISA). "AGS" indicates cells without added *H. pylori*. Values represent means ± standard deviations (SD), based on analysis of at least six replicate samples. Levels of IL-8 production induced by mutant strains were compared to levels induced by the WT strain (analysis of variance [ANOVA] followed by Dunn's multiple-comparison test). *, $P < 0.001$. (b) HA-CagF was immunoaffinity purified from the strains analyzed in panel A. The affinity-purified samples (immunopurification [IP] elution) were then immunoblotted with the indicated antisera. Strains are designated at the top of the panel, and antisera are designated at the right side of the panel. (c) Analysis of Cag protein production in a panel of *cag* mutant strains. Immunoblotting of whole-

43

cell lysates from mutant strains, each expressing HA-CagF from a gene in the urease locus, with the indicated antisera. HspB (a GroEL heat shock protein) was analyzed as a loading control. Strains are designated at the top of the figure, and antisera are designated to the right of the figure. Arrows indicate immunoreactive bands corresponding to Cag proteins, and other bands correspond to cross-reactive proteins. Two CagF bands correspond to the tagged and the untagged forms of the protein. (d) Summary of results from the experiments presented in panel B, indicating the presence or absence of Cag proteins in the preparations immunopurified from mutant strains. Results are representative of three independent experiments.

Discussion

The methodology used in this study for isolating the *H. pylori cag* T4SS core complex relied on immunopurification of CagF, a *cag* PAI-encoded protein that binds the CagA effector protein (27-29). Importantly, velocity sedimentation experiments performed with glycerol gradients indicated that CagF and CagA are not integral components of the core complex. The CagF-based purification strategy conducted with a *cagA* mutant strain failed to yield core complex components, which suggests that CagA directly interacts with the core complex since CagA isn't required for *cag* T4SS activity. The interaction of CagA (and possibly CagF) with the core complex is presumably relatively weak, as expected for an interaction of an effector protein with secretion machinery.

The velocity density gradients also showed that Cag3, T, M, X, and Y also co-sedimented. These five proteins are the predicted members of the core complex, and since they co-sedimented it shows that they form a true complex. Mass spec results also showed that none of the contaminating proteins from the immunopurification co-sedimented with the core complex members.

Each of the five protein components of the core complex (CagX, CagY, CagM, CagT, and Cag3) is required for *cag* T4SS-dependent phenotypes, including CagA translocation into host cells and induction of IL-8 secretion by gastric epithelial cells (14). Most of these proteins are also required for production of T4SS-associated pilus structures at the bacterium-host cell interface (26). Previous studies have localized all 5 of these proteins at the outer membrane of *H. pylori* (40, 41). CagX and CagY exhibit low-level sequence relatedness to VirB9

and VirB10 components of T4SSs in other bacterial species, but CagM, CagT, and Cag3 are not closely related to any other bacterial proteins. Sequence analysis of CagT suggests that it may be a lipoprotein, leading to the suggestion that it may be a VirB7 homolog (7). CagA exhibits a high level of sequence variation in comparisons of *H. pylori* strains from different geographic areas, but, in contrast, there is relatively little phylogeographic variation in sequences of the five core complex components (41).

The mechanism of core complex formation in *H. pylori* is not at all understood. In model organisms it is hypothesized that the core complex components form in an energy independent manner in the periplasm of the bacteria (19, 42). Four of the five Cag proteins that form the core complex, (Cag3, T, M, and X) all have single sequences, meaning that they are transported past the inner membrane and into the periplasm in a Sec-dependent manner, and all of them are detected in the outer membrane (40).

However, CagY does not possess a signal sequence. That being said, this large protein is also present on the outer membrane and has an alpha-helical domain near its N-terminus that is predicted to be a transmembrane domain (35, 43). Therefore, CagY could be transported past the inner membrane in a Sec-independent manner, with the predicted alpha-helical region inserting into the membrane.   I hypothesize that this region inserts into the inner membrane rather than the outer membrane, so that CagY spans from the outer membrane, through the periplasm and inner membrane to then reach into the cytoplasm. Interestingly, the C-terminus is the region of CagY that possesses the sequence similarity with

VirB10, and I believe this is the part that interacts with the rest of the core complex components in the outer membrane. However, the CagY N-terminus is composed of repeat regions and is entirely unique to *H pylori* (43, 44).   I hypothesize that this region interacts with the CagA-CagF complex, since the latter two proteins are localized to the cytoplasm and inner membrane, and CagY is the only component of the core complex that is not wholly localized to the outer membrane. Furthermore, in a *ΔcagY* strain I can no longer isolate the core complex, even though components of the core complex are still present in bacterial lysates. It is possible that CagY is required for the stability and formation of the core complex, but, it is also possible that CagY is required for the direct interaction of the CagA-CagF complex with the core complex. Another line of evidence that lends credence to the latter hypothesis is the fact that a *ΔcagY* strain can still form pili (44). It seems highly unlikely that the pilus could still form without an integral structural component of the *cag* T4SS.

I was also unable to isolate core complexes from a *ΔcagX* strain. I believe that CagX is required either for the stability of the core complex or potentially as a nucleating factor since CagX is required for T4SS dependent phenotypes.   It seems unlikely that it directly interacts with CagA-CagF since CagX is solely localized to the outer membrane.

Strains with either *cag3, T,* or *M* deleted still yielded partially formed complexes. In a *Δcag3* strain, the only protein missing from the core complex is Cag3, and then in the *ΔcagT* strain, the complex is missing Cag3 and CagT, and then finally in a *ΔcagM* strain, the core complex is missing Cag3, T and M. This

suggests that these three proteins directly interact with one another in a stepwise manner. These proteins could potentially be on the periphery of the core complex, though there would need to be localization studies to confirm this theory.

I also found another way to isolate the core complex by utilizing a FLAG tag on CagT. This experiment gives the results from the CagF IP further credence, since CagT is an integral member of the core complex. In future studies, this can also be an invaluable tool to determine the role of CagX, Y and A in core complex stability and formation. I would follow the same strategy as I did in making the *cag* knock-out strains that could express HA-CagF, though instead of using HA-CagF I would use T/27KFLAG, and then assess core complex formation through western blotting. This way I could determine if CagX, Y and A are truly required for core complex formation, or simply for direct interaction with the CagA-CagF complex since, I am purifying the complex in a directly tagged manner.

Overall, this study provides biochemical and biophysical characterization of the core complex of the *cag* T4SS. I showed that all five predicted proteins of the core complex co-sediment through the glycerol gradient, and this complex is separate from the CagA-CagF complex. I also showed that I could purify the same core complex using a FLAG tagged copy of CagT, confirming my earlier results. Finally, I showed that CagX, Y and A are required for purification of the core complex using HA-CagF, while Cag3, T, and M are dispensable. All together these results provide a solid groundwork for understanding the composition and structural organization of the *cag* T4SS core complex.

CHAPTER IV

NEGATIVE STAIN TRANSMISSION ELECTRON

MICROSCOPY ANALYSIS OF THE *cag* T4SS

CORE COMPLEX

Introduction

In the previous chapter I showed that I could isolate five predicted members of the core complex and those five proteins formed a complex that would co-migrated through a density gradient. This represents the first time that the *cag* core complex has been isolated.

The next logical step is to analyze the structure of this large complex. Negative stain EM provides an attractive option; since the complex is as large as it is, I should be able to visualize it with relative ease. Furthermore, I have a panel of mutants which should allow me to isolate complexes that are missing certain proteins. By imaging these mutants and comparing them to WT core complex, I can localize these various proteins based on the change in structure. I can also use techniques such as immunogold labeling to localize various proteins.

In the first chapter I showed that I could isolate core complexes from bacteria that had been co-cultured with gastric epithelial cells. There was no difference in the protein composition of these core complexes compared to ones that were isolated from pure culture. However, while the protein content remains the same, there could be some kind of conformational shift that occurs resulting in

CagA translocation into host cells. EM would be an excellent venue to assess any

kind of major conformational shift in response to these changes in conditions.

Materials and Methods

Negative-stain electron microscopy

Samples were prepared for EM as previously described (45). In brief, 5-µl aliquots of samples were absorbed to a glow-discharged 400-mesh copper grid covered with carbon-coated collodion film. Grids were washed in four drops of water and then stained with two drops of uranyl formate (0.75%). Samples were imaged on a Morgagni electron microscope (FEI, Hillsboro, OR) operated at an acceleration voltage of 100 kV and equipped with a 1kx1k charge-coupled-device (CCD) camera (ATM). Images were recorded at a magnification of ×28,000. Images used to generate class averages were collected using a TF20 electron microscope (FEI, Hillsboro, OR) equipped with a field emission gun with an acceleration voltage of 200 kV under low-dose conditions at ×62,000 magnification and a defocus value of ~1.5 µm on a 4kx4k Gatan Ultrascan CCD camera.

Class average generation

Images were converted to mrc (mixed raster content) and binned using a value of 2, resulting in images of 3.5 Å/pixel. For alignment and averaging, 445, 416, and 149 images of wild-type complexes, Δ*cag3*complexes, and Δ*cagT* complexes, respectively, were selected with Boxer and windowed with a 180-pixel side length (63 nm). Image analysis was conducted with Spider (46). The images were rotationally and translationally aligned and subjected to 10 cycles of multireference alignment and K-means classification. For analysis of wild-type complexes, particles were classified into 10 class averages and then 5

51

representative classes were chosen as references for multireference alignment. For analysis of Δ*cag3* complexes, particles were classified into 20 class averages and then 4 representative projections were chosen as references for another cycle of multireference alignment. For analysis of Δ*cagT* complexes, particles were first classified into 10 class averages and then 3 representative projections were chosen as references for another cycle of multireference alignment.

<div align="center">

Immunogold labeling of protein complexes
</div>

Immunogold labeling of protein complexes was performed on glow-discharged carbon-coated grids. Purified core complex preparations were applied to grids and blocked with 0.1% cold fish skin gelatin in 50 mM sodium cacodylate for 30 minutes at 4°C in a sealed container. Excess blocking buffer was removed with filter paper, and then the grid was incubated with primary antibody (rabbit anti-Cag3 serum (34)) in 50 mM sodium cacodylate, at a dilution of 1:50 for 45 minutes at room temperature. Excess primary antibody was removed using filter paper, and the grid was rinsed three times for five minutes each in 10 mM HEPES with 300 mM NaCl at pH 7.0 at room temperature. Excess PBS was removed, and then the grid was incubated in secondary anti-rabbit antibody conjugated to 5 nm gold particles (EMS) at a dilution of 1:50 in 0.1% cold fish skin gelatin in 50 mM sodium cacodylate for 45 minutes at 4°C. The grids were rinsed three times in 10 mM HEPES with 300 mM NaCl at pH 7.0 at room temperature, and stained with 0.75% uranyl formate prior to transmission EM.

Results

Ultrastructure analysis of the core complex

To analyze the ultrastructure of the *cag* T4SS core complex, *H. pylori* strain HA-CagF (containing an intact *cag* PAI) was grown in liquid culture, complexes were isolated using the CagF-based immunopurification procedure, and the resulting complexes were visualized using negative-stain single-particle EM. EM analysis revealed homogeneous ring-shaped complexes, characterized by well-defined central and outer rings (**Fig. 6a**). We also used negative-stain single-particle EM to analyze fractions of a glycerol gradient to which the immunopurified sample had been applied, and detected the ring-shaped complexes in the gradient fractions that contained Cag3, CagT, CagM, CagX, and CagY but not in the fractions collected from the top of the gradient (see **Fig. 7**). Since the gradient fractions containing these five Cag proteins were free of other non-Cag proteins (**Fig. 7b**; see also **Table 3**), this result provides evidence that the complexes visible by EM were composed of Cag proteins.

CagA is not actively secreted during *H. pylori* growth in liquid culture, but coculture of *H. pylori* with gastric epithelial cells is a trigger for translocation of CagA into host cells (12, 36). Therefore, although we did not detect substantial changes in the protein composition of the immunopurified samples in response to epithelial cell contact (see **Table 2**), we hypothesized that the structural organization of the complexes might change when *H. pylori* contacts gastric epithelial cells. To investigate this possibility, we cocultured AGS gastric epithelial

53

cells with HA-CagF-producing *H. pylori,* purified HA-CagF from lysates of the

coculture, and then analyzed the immunopurified core complexes by negative-

stain EM. The protein complexes isolated from *H. pylori-*AGS cocultures were

similar in appearance to the complexes isolated from *H. pylori* liquid cultures, but

the complexes isolated from cocultures were present in relatively low abundance

and had a less homogeneous appearance (suggestive of degradation) than

complexes isolated from liquid cultures (**Fig. 6b**). Similarly, the complexes in

fractions collected from glycerol gradients had suboptimal morphology and

exhibited increased clumping compared to complexes in preparations that were

not passed through a gradient (see **Fig. 7**). Consequently, all of the subsequent

experiments focused on complexes isolated from *H. pylori* grown in liquid culture

without use of gradient procedures.

To further characterize the structural features of the complexes, we

collected a data set containing a larger number of particles, classified them into 10

class averages using reference-free alignment (**Fig. 8a**), and then selected five of

these as references for an additional round of reference-based alignment (see **Fig.**

8a and **b**). This analysis revealed well-defined classes, corresponding to *en face*

and side views of the T4SS complex (**Fig. 6c**). The *en face* class shows a well-

defined complex, 41 nm in diameter, composed of a central and outer ring

connected by short linkers. The central ring is 19 nm in diameter. Fourteen linkers

or "spokes" connect the central and outer ring. The side view reveals a stalk-like

structure about 40 nm in length, most likely corresponding to the central ring seen

in the *en face* views that extends away from the ring-like portions of the complex.
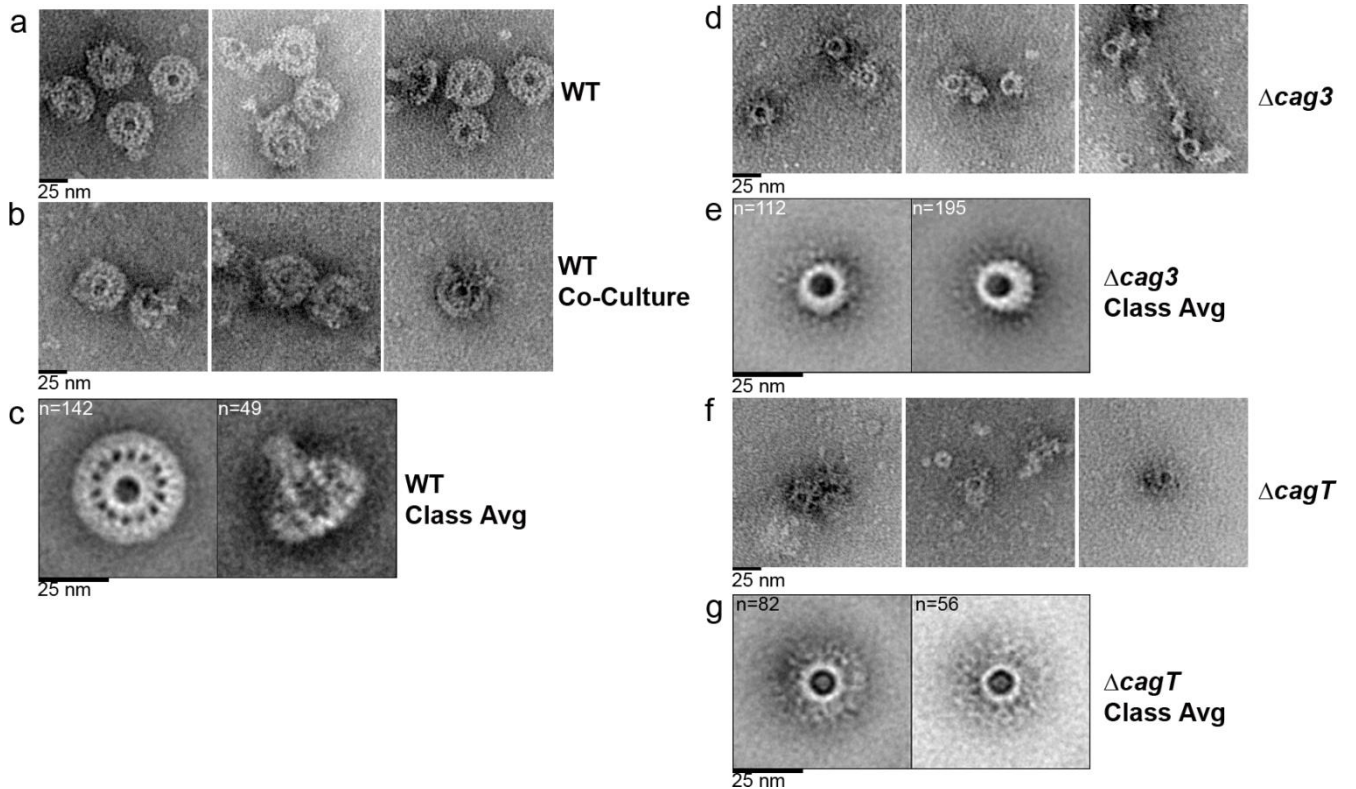
Figure 6: Negative-stain EM analysis of the *H. pylori cag* T4SS core complex.
(a) Representative negative-stain images of WT *cag* T4SS complexes purified
from a strain with an intact *cag*PAI that was engineered to produce HA-CagF. (b)
Representative negative-stain images of WT *cag* T4SS complexes purified from an
HA-CagF-producing strain (Δ*cagF*/HA-CagF) cocultured with AGS cells. (c) Class
averages (Class Avg) showing an *en face* view and a side view of WT complexes.
(d) Representative negative-stain images of *cag* T4SS complexes purified from a
Δ*cag3*strain engineered to produce HA-CagF. (e) Class averages of the *cag*T4SS
complex purified from a Δ*cag3* strain. (f) Representative negative-stain images of
cag T4SS complexes purified from a Δ*cagT*strain engineered to produce HA-
CagF. (g) Class averages of the *cag*T4SS complex purified from a Δ*cagT* strain.
Numbers of particles used for generating each average are shown in upper left
corner of panels. All scale bars represent 25 nm.

55

a Fraction 1

25 nm

b Fraction 6

25 nm
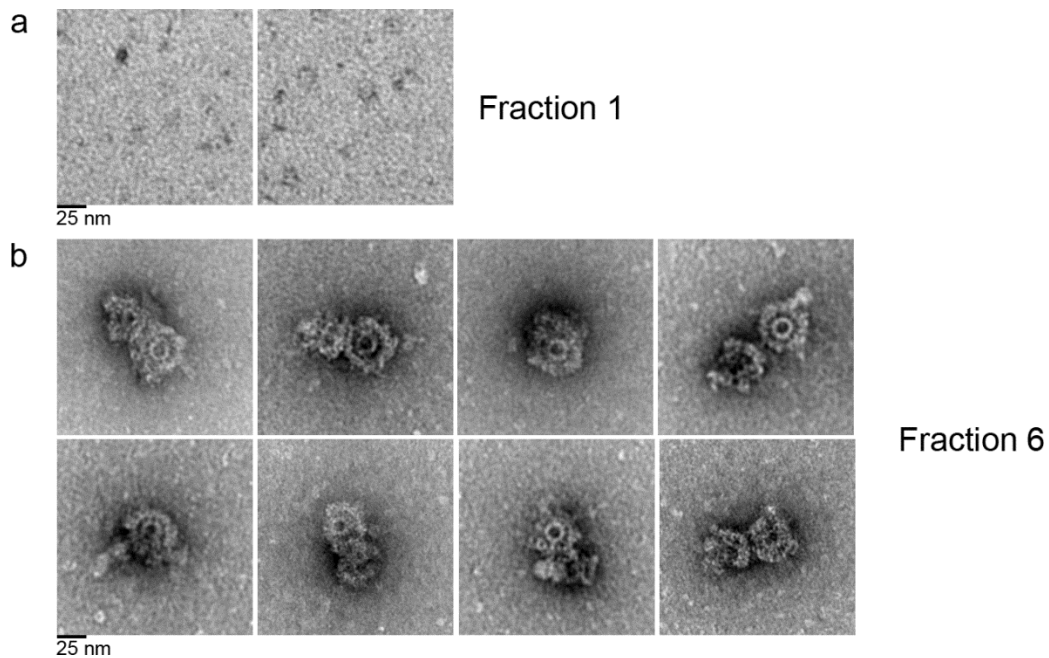
Figure 7: Negative-stain EM analysis of glycerol gradient fractions containing the *cag* T4SS core complex. Preparations resulting from immunoaffinity purification of HA-CagF were analyzed by velocity sedimentation in glycerol density gradients as shown in **Fig. 4**, and gradient fractions were analyzed by negative-stain EM. (a) Representative negative-stain EM images of fraction 1 (top of the gradient), which contains CagA and CagF but no core complex components. (b) Representative negative-stain EM images of fraction 6 (near bottom of the gradient), which contains *cag* core complex components. Fraction 1 contained no complexes as visualized by EM (a), while *cag* T4SS complexes were observed in images taken from fraction 6 (b). Scale bar, 25 nm.
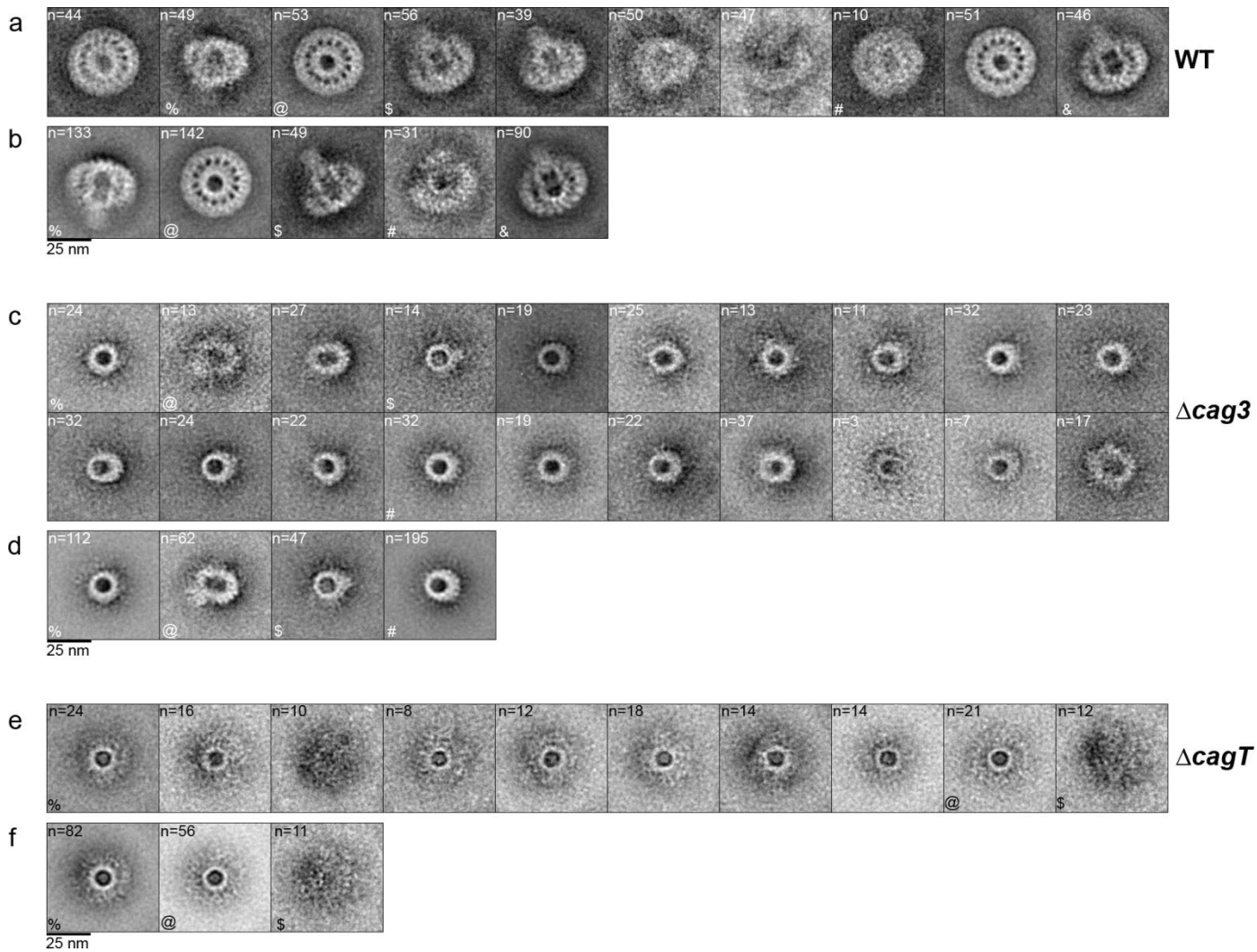
Figure 8: Negative-stain 2D class averages of WT *cag* T4SS complexes and complexes purified from Δ*cag3* or Δ*cagT* mutant strains. (a) Reference-free alignment of 445 particles isolated from an HA-CagF-producing strain with an intact *cag* PAI into 10 classes. Five classes (marked with %, @, $, #, and & symbols) were chosen as references for an additional round of multireference alignment. (b) Result of reference-based alignment. (c) Reference-free alignment of 416 particles isolated from a Δ*cag3*/HA-CagF-producing strain into 20 classes. Four classes (marked with %, @, $, and # symbols) were chosen as references for an additional round of multireference alignment. (d) Result of reference-based alignment. (e) Reference-free alignment of 149 particles isolated from a Δ*cagT*/HA-CagF-producing strain into 10 classes. Three classes (marked with %, @, and $ symbols) were chosen as references for an additional round of multireference alignment. (f) Result of reference-based alignment. Numbers of particles used for generating each average are shown in upper left-hand corner of the panels. Side length of panels, 63 nm.

57

<u>Ultrastructure analysis of complexes formed by mutant strains</u>

The immunoprecipitation analyses of mutant strains shown in **Fig. 5** suggested that incomplete complexes could be formed in the absence of Cag3, CagT, or CagM (**Fig. 5b** and **d**). To investigate these incomplete complexes further, we applied the CagF-based immunopurification procedure to these mutant strains and then analyzed the resulting complexes by negative-stain EM. We first imaged preparations derived from the Δ*cagX* mutant (containing CagA and CagF but not any components of the core complex, based on immunoblotting results) (**Fig. 5b** and **d**). Consistent with the absence of core complex components detectable by immunoblotting, we did not visualize any ring-like structures in these preparations.

We then imaged complexes isolated from the Δ*cag3* mutant (lacking Cag3 but containing CagX, CagY, CagT, and CagM, based on immunoblot analysis results). The complexes detected in these preparations contained a well-defined ring, but there was a marked reduction in peripheral content and no evidence of a well-organized outer ring (**Fig. 6d**). To further characterize the structural features of complexes purified from the Δ*cag3* mutant, we collected a data set of particles and generated class averages (**Fig. 8c** and **d**). These averages revealed *en face* views of a 19-nm-diameter ring (**Fig. 6e**). Unlike images and classes of wild-type T4SS core complexes, the 2D class averages of complexes isolated from the Δ*cag3* mutant contained no outer ring or connecting spokes (**Fig. 6e**). The absence of peripheral components in complexes from the Δ*cag3* mutant without any obvious perturbation of the 19-nm-diameter ring provides evidence that Cag3

is localized to the periphery of the complex. To further test this hypothesis, we immunolabeled purified WT core complexes as well as complexes from the Δ*cag3* mutant with primary polyclonal antibodies reactive to Cag3, followed by a secondary antibody conjugated to 5-nm-diameter gold particles. Imaging of the labeled wild-type complexes by negative-stain EM revealed gold particles in the periphery of the complexes, but no gold particles were associated with the central ring (**Fig. 9a**). Mutant complexes lacking Cag3 were not labeled with gold particles (**Fig. 9b**), demonstrating the specificity of the immunolabeling. These immunogold EM results corroborated the conclusion that Cag3 is a peripheral component of the *cag* T4SS core complex.

We also imaged complexes isolated from the Δ*cagT* mutant (lacking both Cag3 and CagT). Similarly to complexes isolated from the Δ*cag3* mutant, the complexes isolated from the Δ*cagT* mutant contained a single well-defined ring but no well-organized outer ring (**Fig. 6f**). The Δ*cagT* mutant complexes were less abundant than the Δ*cag3* mutant complexes. In order to characterize the structural features of complexes purified from the Δ*cagT* mutant, we collected a data set of particles and generated class averages (**Fig. 8e** and **f**). The *en face* views indicate that the diameter of the ring in complexes from the Δ*cagT* mutant is similar to the diameters of the rings in complexes from the Δ*cag3* mutant and WT strains (**Fig. 6g**), but the ring in complexes from the Δ*cagT* mutant appears to have a reduced thickness. Unlike images and classes of wild-type T4SS core complexes, the 2D class averages of complexes isolated from the Δ*cagT* mutant contained no outer ring or connecting spokes (**Fig. 6g**). Immunogold labeling of CagT was attempted

using complexes purified from WT and Δ*cagT* strains; however, the results were inconclusive due to the nonspecific binding of the anti-CagT polyclonal antibody under these conditions. We also analyzed preparations derived from the Δ*cagM* mutants (which lack Cag3, CagT, and CagM) but were unable to visualize any ring-shaped complexes in these preparations. The absence of detectable ring-like complexes in preparations from the Δ*cagM* mutant could be due to a reduced stability of these complexes, or, alternatively, ring-shaped complexes may fail to assemble in the absence of CagM. Collectively, the structural differences between mutant and WT complexes, combined with specific antibody labeling, confirm that the ring-like structures seen by negative-stain EM correspond to the T4SS core complex.
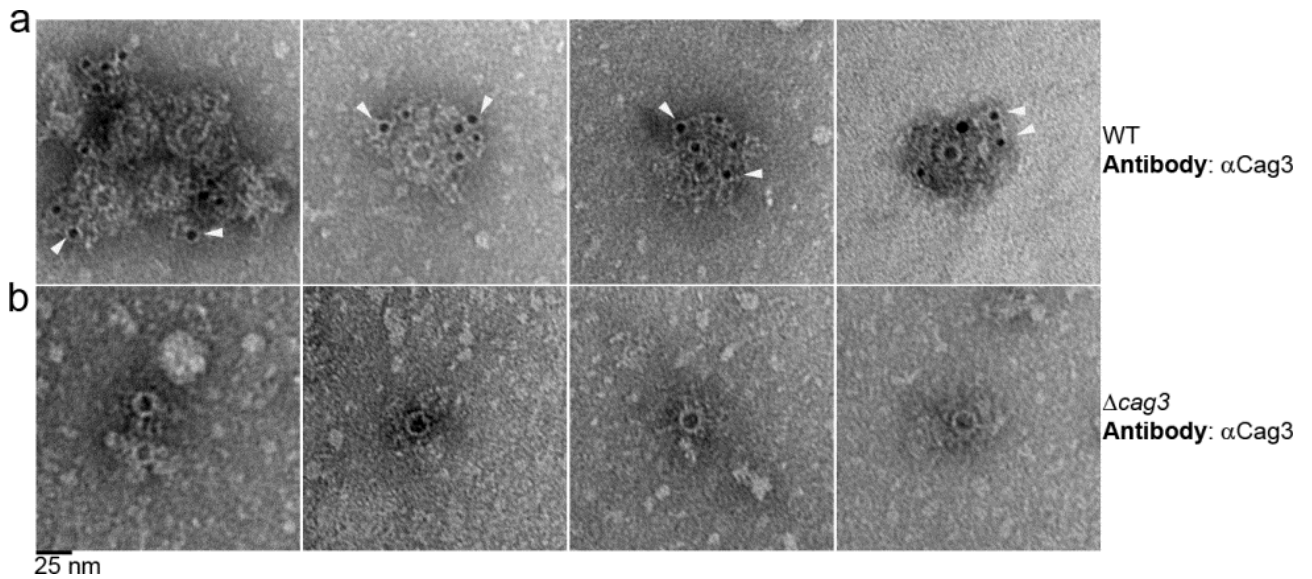
Figure 9: Localization of Cag3 by immunogold labeling and negative-stain EM. Negative-stain EM images of WT core complexes, isolated from a strain with an intact *cag* PAI (a) or a Δ*cag3* mutant (b) and subjected to immunogold labeling with primary anti-Cag3 antiserum followed by 5-nm-diameter-gold-particle-conjugated secondary antibodies, are shown. Arrowheads indicate representative gold particles. Images are representative of results from three independent experiments. Scale bar, 25 nm.

Analysis of complexes formed by mutant strains of certain *cag* genes required for T4SS activity

Every integral member of the core complex is required for *H. pylori* induced IL-8 production in gastric epithelial cells and for the translocation of CagA. However, there are ten additional Cag proteins that are also required for IL-8 induction, namely Cag4, Cagα, CagW, CagV, CagU, CagH, CagI, CagL, CagE and CagC.   These Cag proteins could potentially be playing a role in core complex formation. To address this hypothesis, I analyzed mutant strains where *cagV, cagH, cagI, cagL, cagE* and *cagC* had been individually deleted and then engineered to express HA-CagF, allowing purification of the core complex. I assessed core complex formation through either western blotting or electron microscopy. Through these analyses, I found that six of the proteins I tested were required for core complex formation. (**Fig. 10**).
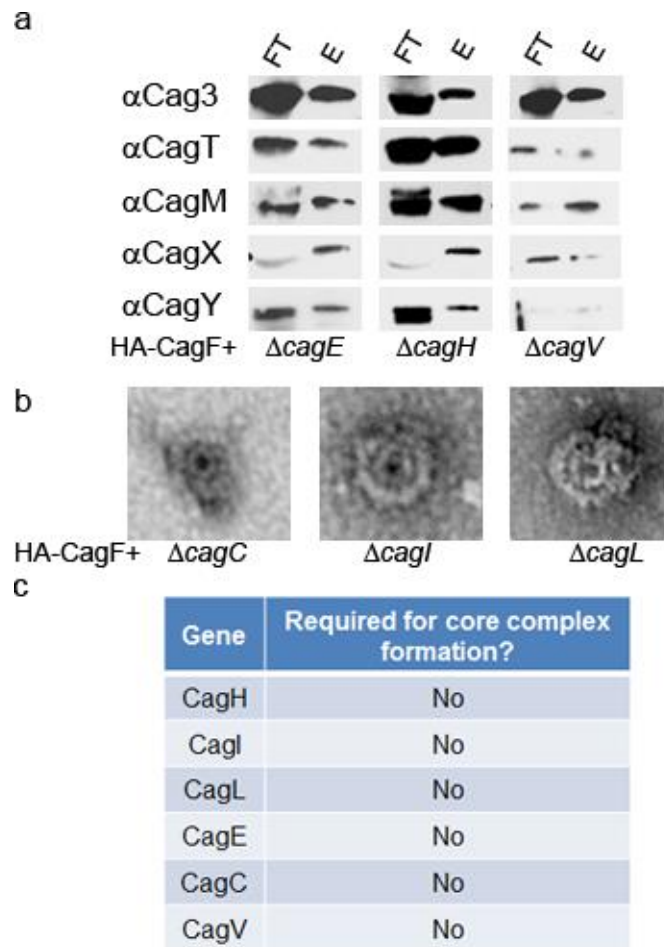
Figure 10: Immunopurification of HA-CagF from a panel of *cag* mutant strains. (a) HA-CagF was immunoaffinity purified from indicated mutant strains which each produced HA-CagF and analyzed by western blotting with indicated antibodies. Strains are designated at the bottom of the panel, and antisera are designated at the left side of the panel. E refers to elution, and FT is flow-through from the elution and acts as a positive control. (b) Negative stain EM of HA-CagF immunopurifications from indicated mutant strains. Strains are designated at the bottom of the panel. (c) Summary of results from the experiments presented in panel a and b, indicating the requirement for core complex formation from mutant strains as determined by either western blotting or EM analysis.
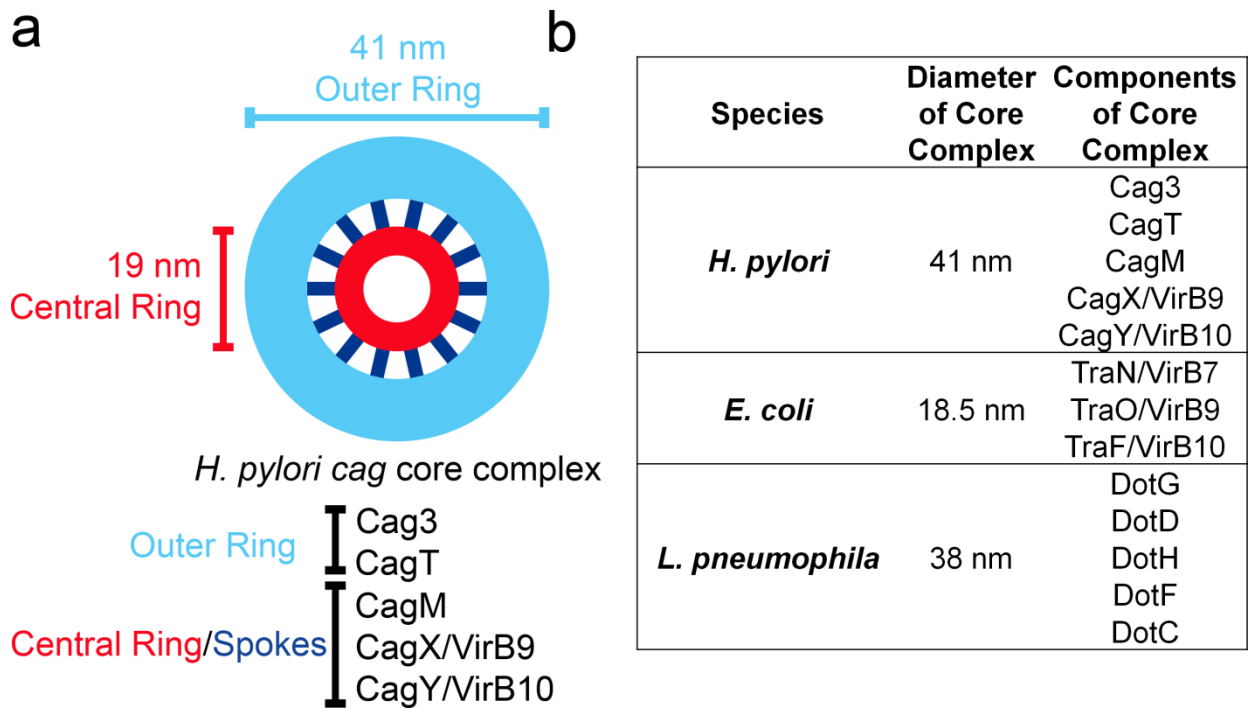
Figure 11: Schematic model of the *H. pylori cag* T4SS core complex. (a) Prominent features of the complex include an outer ring (light blue) and central ring (red) connected by 14 spokes (dark blue). Analysis of complexes from mutant strains suggests that Cag3 and CagT are localized to the periphery of the intact wild-type complex. (b) Comparison of the diameters and components of T4SS core complexes from *H. pylori*, *E. coli*, and *L. pneumophila*.

Discussion

EM analysis of the isolated core complexes revealed homogeneous ring-shaped structures, 41 nm in diameter, characterized by a central ring connected to an outer ring by thin spoke-like linkers (**Fig. 11a**). The presence of 14 visible spoke-like linkers in the class averages suggests that the complex may have 14-fold rotational symmetry, but three-dimensional (3D) structural analysis will be required for an accurate assessment of the symmetry. A side view of the complex reveals a long stalk-like domain 40 nm in length that likely corresponds to the central ring seen in the *en face* view. Since all of the core complex components identified in the current study (CagX, CagY, CagT, CagM, and Cag3) have been detected previously on the external surface of *H. pylori* or as components of the outer membrane (35, 40), we propose that the complex is associated primarily with the outer membrane.

Analyses of complexes isolated from a Δ*cag3* mutant strain, along with immunogold labeling experiments, indicate that Cag3 is localized to the periphery of the core complex. Specifically, the complexes formed by the Δ*cag3* mutant strain (containing CagX, CagY, CagM, and CagT, based on immunoblot analysis results) have an intact 19-nm-diameter ring but exhibit a significantly reduced overall diameter compared to wild-type complexes, attributable to loss of the outer ring and spoke-like linkers. We speculate that the substantial loss of peripheral components in complexes from the Δ*cag3* mutant reflects not only the absence of Cag3 but also an alteration in the structure of residual peripheral components so that they are structurally disordered and no longer visible by EM. Complexes

isolated from a Δ*cagT* mutant strain (containing CagX, CagY, and CagM, based on immunoblot analysis results) also contained an intact 19-nm-diameter ring and exhibited a loss of visible peripheral components. The complexes from the Δ*cagT* mutant appeared similar to the Δ*cag3* complexes except that the ring appeared to have a reduced thickness. The incomplete complexes derived from Δ*cag3* or Δ*cagT* mutant strains might correspond to intermediate or premature stages in the assembly process that leads to formation of the complete core complex.

Previous studies have analyzed the 3D structure of the T4SS core complex encoded by the *E. coli* pKM101 conjugative plasmid, which is composed of three proteins, VirB7/TraN, VirB9/TraO, and VirB10/TraF (16). The pKM101 T4SS core complex is organized into a ring-like structure, 18.5 nm in diameter, with 14-fold symmetry (16) and is localized primarily to the outer membrane. Core complexes isolated from the *A. tumefaciens* VirB/D4 T4SS are similar in size (about 20 nm in diameter) (47). Therefore, the *H. pylori cag* T4SS core complex (41 nm in diameter) is substantially larger in size than either the pKM101-encoded core complex (21) or *A. tumefaciens* core complex (47) (**Fig. 11b**). Moreover, the *H. pylori* core complex is composed of five Cag proteins, whereas core complexes from the pKM101 and *A. tumefaciens* T4SSs contain only three (VirB7/TraN, VirB9/TraO, and VirB10/TraF) (16, 21) (**Fig. 11b**). Despite these differences in size and overall architecture, it is notable that 14 linkers or spokes connecting the central ring to the outer ring could be visualized in the class average of *H. pylori cag* T4SS complexes, which suggests that the *cag* T4SS core complex may have 14-fold rotational symmetry similar to that of the pKM101 core complex. It is also

notable that the central ring of the *H. pylori cag* T4SS core complex has a diameter (19 nm) similar to those seen with the T4SS core complexes of *E. coli* and *A. tumefaciens* (16, 21, 47), suggesting that the overall structural organization may be conserved.

Only two components of the *H. pylori cag* T4SS core complex exhibit sequence similarity to components of previously studied T4SS core complexes. Specifically, CagX and CagY exhibit sequence relatedness to VirB9 and VirB10 (**Fig. 11b**). We propose that CagX and CagY are localized to the central ring of the *cag* T4SS core complex. Consistent with this view, the complexes isolated from Δ*cag3* and Δ*cagT* mutants contained CagX and CagY and had intact rings (with diameters similar to that of the central ring of WT complexes) but lacked the structurally organized outer ring and connecting spokes visualized in WT T4SS complexes.

The exact composition of core complexes in effector protein-translocating T4SSs has not previously been determined for most bacterial species, though the *Legionella* Dot/Icm core complex is reported to contain at least five proteins (DotC, DotD, DotF, DotG, and DotH) (31). Recently published EM images of the *Legionella* Dot/Icm T4SS core complex indicate that it is organized as a circular structure 38 nm in diameter (31). Thus, the size of the *H. pylori* core complex analyzed in the current study is more similar to that of the *Legionella* Dot/Icm core complex than to those of *Agrobacterium* or pKM101 core complexes (**Fig. 11b**). Another notable similarity between the *H. pylori cag* and *Legionella* Dot/Icm T4SSs is that the core complexes of the two systems have constituents not widely

conserved in T4SSs of other bacterial species. Three *H. pylori* core complex constituents (Cag3, CagM, and CagT) seem to be unique species-specific adaptations, and three *Legionella* core complex constituents (DotC, DotF, and DotD) are absent fromT4SSs in most other bacterial species. The *H. pylori cag* T4SS translocates CagA across the plasma membrane of gastric epithelial cells, whereas the *Legionella* Dot T4SS translocates several hundred effector proteins across endosomal membranes in eukaryotic cells. The existence of specialized adaptations in *H. pylori* and *Legionella* T4SS core complexes is consistent with the considerably different functions of the T4SSs in these two species.

In future studies, it will be important to investigate the molecular mechanisms by which the *H. pylori cag* T4SS mediates CagA translocation into host cells. One possibility is that CagA may pass directly through the central pore visible within the core complex. Analysis of the crystal structure of CagA (amino-terminal residues 1 to 876) indicates that it has dimensions of 8 by 11 by 5.5 nm (48), so either a folded CagA protein or an unfolded CagA protein could conceivably fit through the ~10-nm pore of the core complex. The mechanisms by which CagA enters host cells after translocation across the bacterial envelope remain a topic of inquiry. T4SS-associated pili assembled at the bacterium-host cell interface presumably contribute to the entry of CagA into host cells (34, 49, 50), and it has been suggested that binding of CagA to phosphatidylserine or integrins on the surface of host cells also may be important for entry into host cells (51, 52).

There are 18 proteins encoded the *cag* PAI that are required for T4SS

activity, including the five proteins that form the core complex. It is possible that

these other Cag proteins that are required for T4SS activity because they act as

accessory proteins for the core complex, and might aid its formation or

function.   However, some are more likely candidates than others. For example,

both CagE and Cagα are predicted to be ATPases that provide the energy for

CagA translocation or pilus formation. The core complex most likely forms in the

periplasm or outer membrane which lack ATP and therefore core complex

assembly probably does not require the use of ATPases.   And, core complex

does not require CagE for its formation, but the effect that Cagα might have on the

core complex is still unknown.

CagC is another Cag protein required for T4SS function, and its own

function is unknown; it was predicted to be the major pilin subunit, but when it is

deleted, *H. pylori* still can form pili (26). However, when CagC is deleted, I can still

isolate the core complex, so it is not involved in that function.

CagH, I, and L are all required for T4SS phenotypes, and interestingly,

these proteins form a complex. The deletion of *cagH* results in a hyper-pilated

bacteria, whilst the deletion of either *cagI* or *cagL* results in a lack of pilus

production (34). However, the loss of any of these proteins still results in the

formation of the core complex, so these proteins are not accessory proteins to the

core complex.

One gene that I predict might be involved in core complex assembly would be Cag4, which is predicted to be a peptidoglycan hydrolase. It is predicted to digest the peptidoglycan layer in the periplasm, thus allowing the T4SS machinery to span from the outer membrane to the cytoplasm. I do not think that Cag4 would directly regulate the core complex formation per se, but, without the opening in the peptidoglycan, I think that the core complex would form in the periplasm, but be unable to integrate into the outer membrane. Under these circumstances the core complex would then be swiftly degraded by the proteases in the periplasm.

CHAPTER V

OPTIMIZATION OF *cag* T4SS CORE COMPLEX

PURIFICATION FOR 3D RECONSTRUCTION

Introduction

In the previous chapter I utilized negative stain electron microscopy to better visualize the structural details of the *cag* T4SS core complex. Using this technique, I discovered that the core complex is a large ring-like structure about 40 nm in diameter, containing a central ring that is connected to an outer ring by 14 spokes. Furthermore, I was to localize Cag3 and CagT to the outer ring, through immunogold labeling and mutagenesis studies.   However there are several shortcomings of the negative stain data. First, there is a resolution limit of 20Å. And secondly, samples are that are imaged with electron microscopy are done so under high vacuum which causes severe flattening and dehydrating effects (45).

I wanted to improve and refine my initial findings and learn more details about the structure of the core complex. NMR (nuclear magnetic resonance) and crystallography are two traditional methods that are used for this purpose.   However, NMR is only effective for small proteins less than 50 kDa in size, and thus is not at all appropriate for this large complex. Crystallography can resolve the structure of large proteins and even large protein complexes. However, it is very difficult for large protein complexes to crystallize, and crystallization requires milligrams of material, which cannot be obtained with my current IP methodology.

In the past, low resolution structures have been obtained using negative stain data with a technique called random conical tilt (RCT). In this technique, images are collected at 0° and 60°, and then particle pairs at picked, so there is data about one specific particle at both 0° and 60°. The untilted data allows us to sort the particles into their respective classes, as well as sorting out bad particles or mispicked ones, and then we can model the 3D structure based on the tilted data . This technique at its best will produce a structure at about 20-18Å, but will produce a 3D model, which is more information than I currently have. However, the most optimal approach is cryo-electron microscopy (cryoEM). In this technique, the samples are plunged in vitreous ice, which doesn't form damaging ice crystals, and thus preserves the sample in its biologically-relevant state. When encased in this vitreous ice, the sample will also be protected from the flattening effects of the vacuum present in the microscope. With advances in both camera and microscope technology we can now achieve atomic resolution of protein complexes using cryoEM in their near biological state. Furthermore, this technique is far more successful than x-ray crystallography in obtaining high-resolution structural data of large multiprotein complexes.

The technique to obtain high-resolution data from cryoEM data is called "single particle analysis" (SPA). In brief, a large dataset of tens of thousands of particles (or hundreds of thousands) are collected and all of these particles are averaged together to then provide the data for a 3D model. The assumption is that these particles are randomly orientated in the vitreous ice, giving us several viewpoints of the same complex and allowing us to accurately model in 3D

72

space. It is a powerful technique, but it comes with several limitations. Samples

cannot be stained to keep resolution under 20Å, so all the contrast of the images

must come from the size of the particle. Luckily, the core complex is quite large,

so it should be relatively easy to visualize. Furthermore there can be very little

detergent present in cryo samples. Unfortunately, since the core complex is a

membrane associated complex, there is a high level of detergent present in the

sample, so the concentration of detergent needs to be reduced to allow imaging.

Also, since only very large datasets can provide high resolution data, the sample

has to be concentrated; otherwise the man hours needed to gather enough images

would simply be impractical. I need to greatly increase the concentration of my

sample; it was even difficult to get enough images for negative stain 2D

classification. And finally, since the assumption is that the particles are randomly

oriented in the vitreous ice, they cannot be touching, or clustering about one

another, since this forces the particles into preferential orientations and biases the

dataset. In the negative stain EM that I have, I can tell that the core complex tends

to cluster with itself so this is another hurdle that needs to be addressed.

In this chapter I will show all my efforts to modify my IP protocol to produce

a sample amenable to cryoEM as well as my attempts to make a 3D structure of

the core complex using RCT.

Materials and Methods

Random conical tilt imaging

Samples for EM were prepared as described above. Tilt pairs were collected at -60° and 0° using a TF20 electron microscope (FEI, Hillsboro, OR) equipped with a field emission gun with an acceleration voltage of 200 kV under low-dose conditions at ×25,000 magnification and a defocus value of ~1.5 μm on a 4kx4k Gatan Ultrascan CCD camera.

Class average generation for RCT

Images were converted to mrc (mixed raster content) and binned using a value of 2, resulting in images of 4.4 Å/pixel. For alignment and averaging, 2,526 paired particles of wild-type complexes were selected with WEB and windowed with a 100-pixel side length (44 nm). Image analysis was conducted with Spider (**47**). The images were rotationally and translationally aligned and subjected to 10 cycles of multireference alignment and K-means classification. Particles were classified into 40 class averages and then 6 representative classes were chosen as references for multireference alignment.

Results

Optimization of core complex isolation protocol for cryoEM

Negative stain EM and class averaging provided me with a solid overall morphology of the core complex. Next, I wanted to be able to obtain high-resolution data of this complex. To do this, I turned to cryo-electron microscopy (cryoEM). This technique plunges the sample in liquid ethane, creating vitreous ice, protecting the sample from the flattening and dehydration effects present in EM.

However, the conditions amenable for cryoEM are strict. The sample ideally will have no detergents, though low levels are tolerated. Furthermore, the sample has to be concentrated, ideally with several particles per viewfield. And finally, the particles need to be spaced out and not clustered together. Unfortunately my current protocol to isolate the core complex provides a sample that has both a high detergent content and dilute concentration, not to mention that the core complex has a tendency to cluster.

I first attempted to find another way to purify the core complex. IPs usually do not provide significant yield, and also my method is an indirect one, since CagF isn't an integral part of the core complex. In the previous chapter, I mentioned that I had been able to purify the core complex using a FLAG tag to CagT. However, that purification was even more dilute that using the HA-CagF (**Table 4**). I also tagged CagX, Cag3, and CagM with HA tags (see table 4) but none of these purifications yielded the core complex.   Then I turned to using a variety of small

affinity tags. The strep-tag is a small eight amino acid tag that can bind to strep-

tactin resin (an engineered form of strepavidin) and then be specifically eluted

using desthiobiotin. The core complex from *E. coli* was purified using a strep tag

on the C-terminus of VirB10. However I was unable to purify the core complex

with strep-tagged copies of CagA, CagF, or CagY (**Table 4**). After the failure of

the strep tags, I turned to another small affinity tag, namely a 13xHis tag. I tagged

both CagF and Cag3 with a 13xHis tag, both at their N-and C-terminus. Cag3 has

a signal sequence, so the His tag was placed after the signal sequence, therefore

after modification of the protein by the Sec machinery the His tag is at the N-

terminus. I was able to recover the core complex from the strain producing the N-

terminally tagged Cag3 construct based on western blot analysis. However, I was

not able to visualize any core complex from this strain using EM. This was likely

due to the high level of contaminating proteins in the sample. (**Table 4**) Thus far, it

seems that the most consistent way to isolate the core complex has been by using

the HA-CagF construct.

Table 4: Alternative methods of tagging to isolate the core complex

	Proteins purified (assessed by immunoblot)	EM results
ΔCag3/Cag3-HA$_{61H}$	Cag3	N/A
ΔCag3/Cag3-13xHis$_{22A}$	Cag3, T, M, X, and Y	Unable to visualize core complex
ΔCag3/Cag3-13xHis	None	N/A
ΔCagM/CagM-HA$_{18A}$	CagM	N/A
ΔCagM/ CagM-HA$_{232D}$	none	N/A
ΔCagX/CagX-HA$_{26A}$	none	N/A
ΔCagT/CagT-FLAG$_{55K}$	CagT, X	N/A
ΔCagT/CagT-FLAG$_{106K}$	none	N/A
ΔCagT/CagT-FLAG$_{210Y}$	Cag3, T, M, and X	Unable to visualize any complex
ΔCagT/CagM-HA$_{18A}$	none	N/A
ΔCagT/CagX-HA$_{26A}$	none	N/A
CagY-strep	none	N/A
ΔCagF/Strep-CagF	none	N/A
ΔCagF/13xHis-CagF	CagY (non-specific)	N/A
ΔCagF/CagF-13xHis	CagT (non-specific)	N/A
Strep-CagA	none	N/A

Then I turned to changing the detergent conditions to find a better mix, and potentially lower the detergent concentration compared to the RIPA mix I typically used, which contains 1% NP-40 and 0.25% DOC. However, after extensive testing (**Table 5**), the only other detergent that seem to work are 1% or 0.5% NP-50 paired with 0.5% CHAPS. But, EM revealed that the core complexes obtained using these conditions are even more dilute, and there seemed to be residual membrane still bound to the complexes, obscuring much of their morphological detail. Increasing the NaCl concentration from 100mM to 300 mM in RIPA marginally improved the EM, so I moved to that concentration.

Table 5: Alternative detergent conditions to isolate the core complex

Lysis[a]	Wash/elution	Temp/ timing	Proteins purified (assessed by immunoblot)[b]	EM results[c]
1% NP-40, 0.25%DOC, 100 mM NaCl, 20 mM HEPES. **Volume***: 10 mL*	0.3%DDM, 100 mM NaCl, 20 mM HEPES	All at 4°C	CagA, CagF	N/A
20 mM HEPES, 100 mM NaCl, 1 mM EDTA (0.5% DDM added after sonication) **Volume***: 10 mL*	0.3%DDM, 100 mM NaCl, 20 mM HEPES	All at 4°C	CagA, CagF	N/A
20 mM HEPES, 100 mM NaCl, 1 mM EDTA (0.5% DDM added after sonication)	0.3%DDM, 100 mM NaCl, 20 mM HEPES	All at 4°C, lysate over beads for an hour	CagA, CagF	N/A
0.05%DDM, 20 mM HEPES, 100 mM NaCl, 1 mM EDTA	0.3%DDM, 100 mM NaCl, 20 mM HEPES	All at 4°C, lysate over beads for an hour	CagA, CagF	N/A
0.5%DDM, 20 mM HEPES, 100 mM NaCl, 1 mM EDTA	0.01%DDM, 100 mM NaCl, 20 mM HEPES	All at 4°C, lysate over beads for an hour	CagA, CagF	N/A
2% OG, 20 mM HEPES, 100 mM NaCl, 1 mM EDTA	0.7%OG, 100 mM NaCl, 20 mM HEPES	All at 4°C, lysate over beads for an hour	CagA, CagF	N/A
2% OG, 20 mM HEPES, 100 mM NaCl, 1 mM EDTA	0.7%OG, 100 mM NaCl, 20 mM HEPES	All at 4°C	CagA, CagF	N/A
2% OG, 20 mM HEPES, 100 mM NaCl, 1 mM EDTA	0.7%OG, 100 mM NaCl, 20 mM HEPES	RT	CagA, CagF	N/A
1%DM, 20 mM HEPES, 100 mM NaCl, 1 mM EDTA	0.1%DM, 100 mM NaCl, 20 mM HEPES	RT	CagA, CagF	N/A
2% OG, 0.25% DOC, 20 mM HEPES, 100 mM NaCl, 1 mM EDTA	0.7%OG, 100 mM NaCl, 20 mM HEPES	RT	CagA, CagF	N/A
1%DM, 0.25% DOC, 20 mM HEPES, 100 mM	0.1%DM, 100 mM NaCl, 20	RT	CagA, CagF	N/A

NaCl, 1 mM EDTA	mM HEPES			
1% NP-40, 0.25%DOC, 100 mM NaCl, 1 mM EDTA, 20 mM HEPES	0.5%LDAO, 100 mM NaCl, 20 mM HEPES	RT	CagA, CagF	N/A
2% LDAO, 100 mM NaCl, 1 mM EDTA, 20 mM HEPES	0.5%LDAO, 100 mM NaCl, 20 mM HEPES	RT	CagA, CagF	N/A
1% NP-40, 0.25%DOC, 100 mM NaCl, 20 mM HEPES	0.3% DDM, 100 mM NaCl, 20 mM HEPES.	RT	CagA, CagF	N/A
1% NP-40, 0.5%CHAPS, 100 mM NaCl, 20 mM HEPES	Unchanged from lysis conditions	RT	Cag3, T, M, X, Y, A, F	unknown
0.5% NP-40, 0.5%CHAPS, 100 mM NaCl, 20 mM HEPES.	Unchanged from lysis conditions	RT	Cag3, T, M, X, Y, A, F	Very poor, but can see core complex
1% NP-40, 0.25%DOC, 300 mM NaCl, 20 mM HEPES.	Unchanged from lysis conditions	RT	N/A	Clear core complex but dilute
1% NP-40, 0.25%DOC, 300 mM NaCl, 20 mM HEPES, pH 9.0	Unchanged from lysis conditions	RT	N/A	Clear core complex but dilute
1% NP-40, 0.25%DOC, 300 mM NaCl, 20 mM HEPES, pH 11.0	Unchanged from lysis conditions	RT	N/A	Clear core complex but dilute
1% NP-40, 0.25%DOC, 500 mM NaCl, 20 mM HEPES	Unchanged from lysis conditions	RT	N/A	Clear core complex but dilute
1% NP-40, 0.25%DOC, 100 mM KCl NaCl, 20 mM HEPES	Unchanged from lysis conditions	RT	N/A	Clear core complex but dilute
1% NP-40, 0.25%DOC, 300 mM KCl NaCl, 20 mM HEPES	Unchanged from lysis conditions	RT	N/A	Clear core complex but dilute

[a]All lysis was performed by sonication on ice
[b]N/A indicates that immunoblotting was not performed on that condition
[c]N/A indicates that EM was not performed on that condition

Then, I explored several options to concentrate or increase the yield of the core complex. Unfortunately, it seemed that most of the traditional methods of concentration were not amenable to this sample (**Table 6**). Every attempt to use a concentrator column resulted in aggregation of the core complex or the complex binding to the concentrator itself. Fixation did not solve this problem, nor pre-clearing the concentrator with BSA.

I also attempted to use an ammonium sulfate precipitation and speed-vac to concentrate the sample, but neither of these techniques were successful (**Table 6**). I then tried longer incubations of the sample on EM grids. I saw an increase in number of complexes after an incubation of two hours, but they were highly clustered. An overnight incubation seemed to cause the core complex to fall apart (**Table 6**).

There is a relatively new technique that has been used in cryoEM which utilizes "amphipols" to reduce detergent concentrations (53). These are amphipathic polymers, which exchange with the detergents surrounding the protein, keeping them soluble (54). Then the free detergent can be bound and removed using polystyrene bio-beads. I was hoping that removing the detergents, but still having the amphiphols bound around the core complex would stabilize the core complex and reduce clustering. However, the exchange with biobeads must be done overnight, and I could not visualize the core complex after such a long incubation (**Table 6**), an observation which falls in line with the one I made when I incubated the sample overnight on EM grids.

Then I used increasing amounts of HA-antibody, hoping to saturate the binding of the core complex. I had been using 5 ug with 50 ul of dynabeads, and then I tested up to 30 ug of antibody per 50 ul of dynabeads. To test the yield of the IP, I immunblotted for amounts of CagF which is the being pulled down. I saw a slight increase of CagF at 10 ug of antibody, but I did not see any more increase in CagF for any of the higher concentrations of antibody. (**Table 6**). Finally, I then tried to scale up the IP. Normally, I use 25 mL of liquid bacterial culture with 50 ul of dynabeads, 10 ug of HA-antibody, and eluted in 200 ul. To scale up, I used 125 mL of liquid bacterial culture with 250 ul of dynabeads, 50 ug of HA-antibody, and eluted in 200 ul, which would be expected to increase the yield and concentration five-fold. This resulted in increased concentration of core complex (assessed by EM), though the particles were highly clustered (**Table 6**).

Table 6: Techniques to attempt to concentrate HA-CagF IP elution

Process[a]	Conditions	EM results
Concentrator column	Pre-clear with 5% BSA, 1 mDa nitrocellulose column, 4X concentrated	Aggregated
	Pre-clear with 5% BSA, 1 mDa PES column 2X concentrated	Unchanged from unconcentrated
	Pre-clear with 5% BSA, 1 mDa PES column 4X concentrated	Aggregated
	Pre-clear with 5% BSA, 1 mDa PES column 8X concentrated	Aggregated
Fixation	0.1% Glutaraldehyde, 5 minutes	Looks good
	0.1% Glutaraldehyde, 10 minutes	Aggregated
	0.1% Glutaraldehyde, 20 minutes	Aggregated
	0.1% Glutaraldehyde, 30 minutes	Aggregated
	0.1% Glutaraldehyde, 40 minutes	Aggregated
	0.1% Glutaraldehyde, 5 minutes, concentrate with 1 mDa PES column	Aggregated
	1% PFA, 5 minutes	Aggregated
	1% PFA, 10 minutes	Aggregated
	1% PFA, 15 minutes	Aggregated
	1% PFA, 20 minutes	Aggregated
	1% PFA, 15 minutes, concentrate with 1 mDa PES column	Aggregated
Ammonium sulfate precipitation	50% solution, spin out protein, re-suspend in 20 ul buffer	Unable to see complex
Speed-vac	At 4C down to 50 ul	Complex made more dilute
Reduction of detergents using amphipols	Add amphipols to IP elution, remove detergent with biobeads (O/N)	Unable to see complex
EM	Incubate EM grids in IP elution for 2 hours	Concentrated, but clustering
	Incubate EM grids in IP	Unable to see

	elution for O/N	complex
Modifying the IP protocol	10 ug of HA-antibody/50 uL dynabeads	Slightly more core complex than normal
	15 ug of HA-antibody/50 uL dynabeads	Same as 10 ug HA antibody
	20 ug of HA-antibody/50 uL dynabeads	Same as 10 ug HA antibody
	25 ug of HA-antibody/50 uL dynabeads	Same as 10 ug HA antibody
	30 ug of HA-antibody/50 uL dynabeads	Same as 10 ug HA antibody
	Use 125 mL starting culture, elute in 200 uL for an hour	Concentrated, but clustering

[a]All samples were prepared using the RIPA buffer referenced in the first chapter unless otherwise stated.

Table 7: **Techniques to attempt to reduce clustering of core complex**

[a]Addition to buffer	EM results
50 mM Mg₂SO₄ +1% NP-40, 0.25%DOC, 300 mM NaCl, 20 mM HEPES	Clustered
100 mM Mg₂SO₄ +1% NP-40, 0.25%DOC, 300 mM NaCl, 20 mM HEPES	Clustered
200 mM Mg₂SO₄ +1% NP-40, 0.25%DOC, 300 mM NaCl, 20 mM HEPES	Less complex
50 mM NH₄SO₄ +1% NP-40, 0.25%DOC, 300 mM NaCl, 20 mM HEPES	Less complex
100 mM NH₄SO₄ +1% NP-40, 0.25%DOC, 300 mM NaCl, 20 mM HEPES	Less complex
200 mM NH₄SO₄ +1% NP-40, 0.25%DOC, 300 mM NaCl, 20 mM HEPES	Less complex
10 mM MgCl₂ +1% NP-40, 0.25%DOC, 300 mM NaCl, 20 mM HEPES	Clustered
20 mM MgCl₂ +1% NP-40, 0.25%DOC, 300 mM NaCl, 20 mM HEPES	Clustered
50 mM MgCl₂ +1% NP-40, 0.25%DOC, 300 mM NaCl, 20 mM HEPES	Clustered
2 mM BME +1% NP-40, 0.25%DOC, 300 mM NaCl, 20 mM HEPES	More aggregation
5 mM BME +1% NP-40, 0.25%DOC, 300 mM NaCl, 20 mM HEPES	More aggregation
10 mM BME +1% NP-40, 0.25%DOC, 300 mM NaCl, 20 mM HEPES	More aggregation
125 mM Arginine +1% NP-40, **0.025%DOC**, 300 mM NaCl, 200 mM HEPES	Clustered
250 mM Arginine +1% NP-40, **0.025%DOC**, 300 mM NaCl, 200 mM HEPES	Clustered
500 mM Arginine +1% NP-40, **0.025%DOC**, 300 mM NaCl, 200 mM HEPES	Less complex
125 mM Glycine +1% NP-40, **0.025%DOC**, 300 mM NaCl, 200 mM HEPES	Decreased clustering
250 mM Glycine +1% NP-40,	Clustered

0.025%DOC, 300 mM NaCl, 200 mM HEPES	
500 mM Glycine +1% NP-40, **0.025%DOC**, 300 mM NaCl, 200 mM HEPES	Less complex
125 mM Arginine+125 mM Glycine +1% NP-40, **0.025%DOC**, 300 mM NaCl, 200 mM HEPES	Less complex
125 mM Glycine +0.5% NP-40, **0.025%DOC**, 300 mM NaCl, 200 mM HEPES	Increased clustering
125 mM Glycine +0.5% NP-40, 300 mM NaCl, 200 mM HEPES	Increased clustering
125 mM Glycine +0.1% NP-40, **0.025%DOC**, 300 mM NaCl, 200 mM HEPES	Increased clustering
125 mM Glycine +0.1% NP-40, 300 mM NaCl, 200 mM HEPES	Increased clustering

[a]The base buffer is 1%NP-40, 0.25% DOC, 300 mM NaCl, and 20 mM HEPES at pH 7.0 unless otherwise stated.

I now had found a way to obtain a higher concentration of the particles, but now I had to find a way to reduce the clustering. There are several classes of additives that can be used to reduce protein aggregation, namely kosmotropes, chaotropes, reducing agents and amino acids (55). I tested a wide variety of these additives (**Table 7**), and found that the addition of 125 mM glycine, while not completely eliminating the clustering substantially improved it. I then plunged and attempted to image these samples using cryoEM (with the help of my collaborator, Tasia Pyburn), but the detergent levels were too high and we could not image the complex. I then tried to reduce the amounts of detergent in this sample (**Table 7**), but all reductions in detergent content simply increased the clustering.

<u>3D reconstruction using random conical tilt</u>

Since there were so many hurdles with the cryoEM I decided to attempt Random Conical Tilt (RCT). This technique would allow me to obtain a 3D model of the core complex, though it would not be high resolution since it will still be utilizing negative stain EM. I would image particles at both 0 and 60 and then pick particle "pairs". The untitled data allows us to sort the particle sin to their respective classes, as well as sorting bad particles or mispicked ones, and then we can model the 3D structure based on the tilted data.

I used the concentrated sample which contains 125 mM glycine to image for the RCT. I took 218 paired images (436 total images).   I imaged at 25,000x, a lower magnification than the one used for the initial generation of class averages, in order to have more particles per field of view. Using the program EMAN2, I

88

picked about 1,000 paired particles, and then attempted a reference free-alignment to classify them into averages. However, the averages formed were of poor quality, and only consisted of *en face* views of the core complex, even though during the process of picking I saw many side views (**data not shown**). I repicked this data set using the program WEB, and this program produced far better averages (**data not shown**) including side views. This meant that the issue was not with the quality of the data, but actually with the program.

I then went on and gathered 282 more paired images for a total of 500 paired images including the initial set. I used the program WEB to pick 2,526 particles, which were then classified into six reference-based classes, and these looked very similar to the averages I previously generated (**Fig. 12**). Four of the six classes were of the core complex, so these classes were merged together, giving us 1,650 particles that I then attempted to reconstruct into a 3D projection. However, this did not produce a proper structure, only a strange disc, with unconnected densities (data not shown).
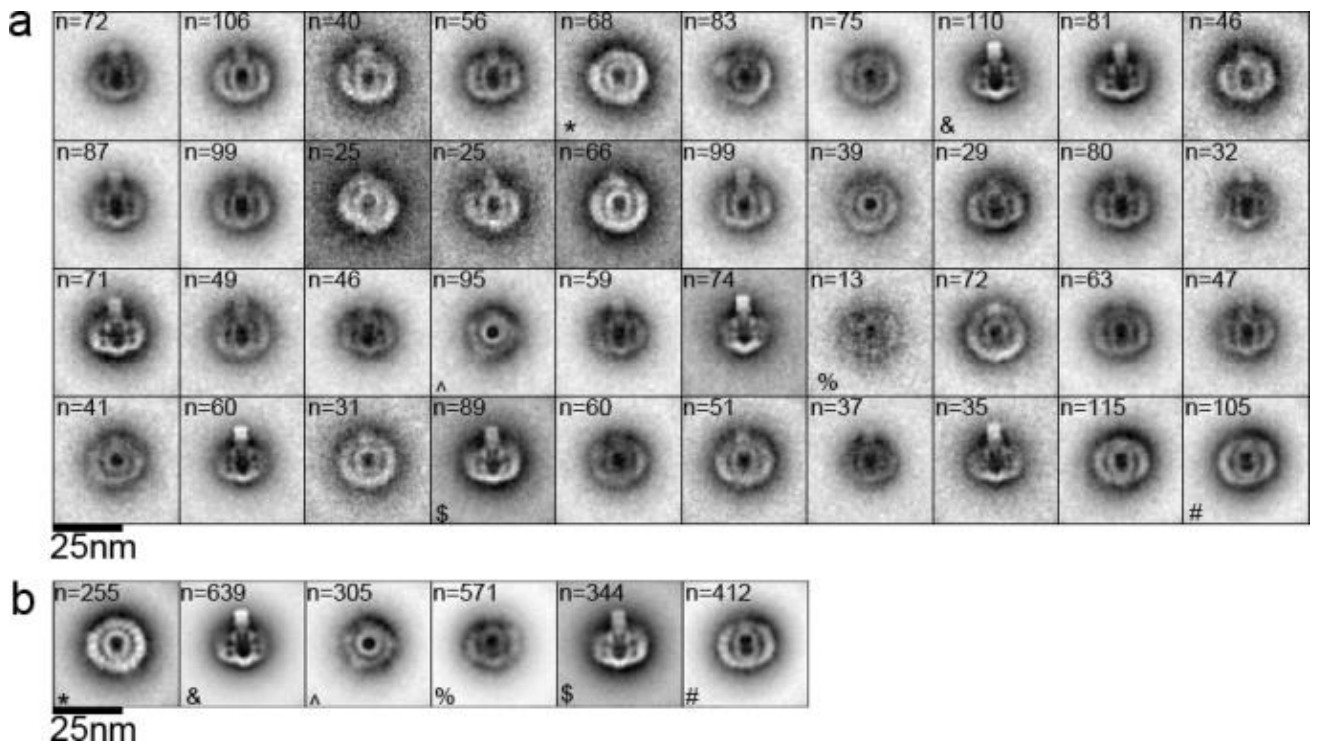
Figure 12: **Negative-stain 2D class averages of WT *cag* T4SS from RCT** (a) Reference-free alignment of 2,526 particles isolated from an HA-CagF-producing strain with an intact *cag* PAI into 40 classes. Six classes (marked with *, &, ^, %, and $ symbols) were chosen as references for an additional round of multireference alignment. (b) Result of reference-based alignment. Numbers of particles used for generating each average are shown in upper left-hand corner of the panels. Side length of panels, 44 nm.

90

Discussion

In this chapter, I attempted to optimize my preparation of the core complex to facilitate cryoEM analysis.   There are several issues with my current preparation that makes it inappropriate for cryoEM. First, the sample is far too dilute to be efficiently imaged with cryoEM. Second, the detergent concentrations are incompatible with cryoEM. And finally, the core complex tends to cluster, meaning that it takes on preferential orientations, which is not ideal for a cryoEM dataset. However, despite all these issues, there are a lot of advantages of using cryoEM to resolve the structure of the core complex. First, the core complex is very large, this means that once it is in the vitreous ice, it should be very easy to see, which is often an issue in cryoEM samples. And, second, it appears that the core complex has a high level of symmetry, potentially 14-fold. This is a huge advantage when it comes to single particle analysis.   For example, if I had picked 1,000 particles of the core complex, assuming that it has a 14-fold symmetry, I have the equivalent of 14,000 particles worth of data. So, despite the hurdles presented with this sample, cryoEM seems to be one of the best options open to me to study the structure of the core complex.

At first, I looked for ways to increase the yield of the purification. The way that I have been able to purify the core complex is by immunopurifying CagF, which purifies the CagA-CagF complex, and it appears that this complex in turn interacts with the core complex. However the interaction with CagF-CagA complex and the core complex is not a very strong one; the two complexes dissociate from one another when applied to a glycerol gradient (see chapter

3). So, I thought that if I could find a way to purify the core complex by tagging one of the proteins that compose the core complex, perhaps this could increase the yield, since I would be relying on a more direct interaction. While I was able to purify the core complex with a specific FLAG tag on CagT, the yield was very poor. I also used HA tags on Cag3, CagX, and CagM, but I was unable to purify the core complex with these constructs. I believe that the presence of the antibody is sterically hindering the formation of the core complex; I am able to purify the protein that is tagged, just not any of its binding partners. Since I was able to use the CagT-FLAG$_{27K}$ construct to purify the core complex, this implies that perhaps there are flexible loops or regions that are surface exposed that do not interfere with core complex assembly. Indeed, it makes sense that the CagT-FLAG$_{27K}$ construct could purify the core complex, since it was originally designed so that the FLAG epitope was inserted into predicted surface exposed, hydrophilic regions of CagT. However, I tested the four other constructs of FLAG-tagged CagT that were also predicted to be surface exposed, and none of them worked, so this method to place tags is not 100% reliable.

Since I was having so many issues with directly tagging the core complex members, I returned to the interaction that I knew worked, the purification of CagF. Using an affinity tag should allow me to purify a larger amount of sample. However introduction of either His-tags or Strep-tags on CagF did not yield any core complex, even if the tags were placed in the exact same position as the HA-tag. I also strep-tagged CagA and CagY, as well as placing His-tags on Cag3, but none of these strategies worked.

The IP of HA-CagF to date is the most reliable way to purify the core complex. Fortunately, cryoEM does not need a large volume of sample, just a concentrated one. Since the IP elution yields 200uL of purified sample, I could concentrate the sample with a concentrator column with a molecular weight cut-off, greatly reducing the volume, but still having plenty for EM. However, the core complex seemed to bind to the membranes of concentrator columns, and I could not fix the issue with either pre-clearing it with BSA or fixing the core complex beforehand. Other attempts of concentration by reducing the volume, namely by ammonium sulfate precipitation and speed-vac, were also unsuccessful.

So then I turned to modifying the IP protocol itself and found that scaling up the IP actually increased the yield, though concentrating the core complex increased its tendency to cluster. It seems as though this is an inherent trait of the complex-perhaps hydrophobic regions of the proteins are associating with one another. Most of the proteins do have large hydrophobic stretches since they are all membrane associated. The addition of glycine somewhat alleviated the clustering problem, but it did not entirely eliminate the problem.

And, finally, it appears that only a very specific detergent mix will allow an HA IP of HA-CagF purify the core complex. I have tried various different combinations of more EM friendly detergents such as DDM, DM, OG and LDAO. In these purifications I could still isolate CagF and CagA, meaning the IP was indeed working, but I could no longer isolate the core complex. This implies that there is a very specific and potentially weak interaction of the CagA-CagF complex with the core complex that can only be preserved with a certain detergent

mix. The only detergents that could preserve this interaction were either NP-40 and DOC, or NP-40 and CHAPS. CHAPS and DOC have similar properties, though DOC is an ionic detergent and CHAPS is zwitterionic and considered a slightly milder than DOC. However, these detergents are not compatible with cryoEM. And while the amount of NP-40 and DOC can be lowered and still recover core complex, the complex clusters even more, and the reduction in detergent is still not enough to be used in cryoEM.

There are other techniques that utilize EM to provide structural data. One of these is RCT, which uses negative stain data. While it cannot reach the atomic resolution of cryoEM, it can provide a 3D model of around 20-18Å in resolution, which would provide me with more information that I currently have. Like cryoEM, it relies on single particle analysis and requires a relatively large dataset of particles, though not nearly as many for cryoEM. I found a way to increase my yield and decrease the clustering of core complex, so this provided an excellent sample for RCT. While I got class averages that matched previous class averages, I was unable to obtain a 3D structure. All the back projections I attempted yielded mostly noise and no meaningful data. On possible problem is that I analyzed about 1,200 particles when attempting to build a 3D model which is on the low end of the spectrum even for RCT.

CHAPTER VI

CONCLUSIONS

Summary

Helicobacter pylori is a ubiquitous gastric pathogen that infects over half of the world's population (1, 56). This bacterium is a common causative agent of gastric cancer, which is the second leading cause of cancer deaths in the world. Due to this connection, *H. pylori* has been designated a Class I carcinogen by the WHO.

There is a high level genetic diversity among *H. pylori* strains, and strains producing certain virulence factors are associated with adverse clinical outcomes. One of these virulence factors is the *cag* PAI, a 40 kb region on the bacterial chromosome (5). The *cag* PAI encodes the bacterial effector protein, CagA as well as a T4SS. CagA is a bona fide oncoprotein, and contributes to the gastric cancer associated with long-term *H. pylori* infection (33, 57).

Unsurprisingly, CagA and the *cag* T4SS have been the focus of many studies over the years. But there have been many hurdles to isolating subassemblies of the *cag* T4SS. This study is the first to isolate a large subassembly of the *cag* T4SS and study it in isolation.

Directly tagging proteins that are integral members of the *cag* T4SS has not allowed isolation of protein complexes in the past, so I turned to a more indirect approach. I put a tag on CagF, which is a chaperone of the effector protein CagA. Previous studies have shown that CagA and CagF are binding partners,

but have not shown any other interactions involving these proteins (27, 28). However, in my hands, I was able to not only purify CagA and CagF using an HA-tagged version of CagF, but also five other Cag proteins, namely Cag3, T, M, X and Y. These five proteins correspond to the core complex of the *cag* T4SS. These results are first time that CagA has been shown to interact with a major subassembly of the *cag* T4SS. Furthermore, the set of proteins interacting with CagA and CagF does not change if the IP is performed from bacteria co-cultured with gastric epithelial cells.

To determine if these purified proteins formed a complex, I performed differential density centrifugation and I found that I was purifying two separate complexes: CagA and CagF co-purified with one another, and that Cag3, T, M, X and Y co-purified with one another. Those five proteins, Cag3, T, M, X and Y are all predicted to be the core complex of the T4SS, and thus by showing that they co-migrated through a gradient, I have an important piece of evidence that I am indeed purifying the core complex of the *cag* T4SS. It is worth noting that while I can purify the core complex using the CagA-CagF complex as bait; it appears that the interaction with the core complex is a transient one.

I then went on to study the protein-protein interactions that make up the core complex. I deleted the gene of each protein that was in the core complex (as well as CagA), and then re-introduced an HA-tagged copy of CagF into these mutant strains so I could isolate these mutant core complexes and determine their protein composition. Through these studies I found that CagA, CagX and CagY are all required for the purification of the core complex. Every member of the core

complex is required for T4SS activity, so it is doubtful that the core complex is not being formed in the CagA mutant ; rather, I hypothesize that it is directly interacting with the core complex. Therefore, in the absence of CagA, CagF does not interact with the core complex, and I can no longer purify it. CagX or CagY could be required for either the formation or the stability of the core complex, though at this juncture we cannot rule out the possibility that they are also required for the direct interaction with the CagF-CagA complex.

The three other proteins that form the core complex, Cag3, CagT, and CagM are not required for the purification of the core complex, and I am able to purify mutant complexes lacking these compoents. The Cag3 mutant is only missing Cag3, while the CagT mutant is missing Cag3 and CagT, and finally the CagM mutant is missing Cag3, T and M. It appears then that these three proteins are directly interacting with one another.

Every member of the core complex is required for the activity of the T4SS as well as active CagA translocation. However, there are thirteen other Cag proteins beyond these five that are also required for T4SS activity. It is possible then, that some of these Cag proteins could also be aiding in the formation of the core complex, and thus their loss would cause the loss of T4SS activity and CagA translocation. To test this, I used the same strategy as I did with the integral members of the core complex; I deleted the genes encoding Cag proteins required for T4SS function, and then engineered these mutant strains to express an HA-tagged form of CagF to purify the core complex. Thus far, of the thirteen proteins

97

required, I have been able to test six mutants, namely Cag H, I, L, E, V and C. None of these proteins are required for core complex formation.

After these biochemical and biophysical characterizations, I turned to negative stain EM to begin to understand the structure of this complex. *En face* views revealed that the WT core complex had a large, ring-like structure, containing an outer ring and a central ring connected by 14 spokes. The entire diameter is 41 nm, while the central ring is 19 nm. Side views also revealed a large stalk protruding from the central ring that is 41 nm in diameter. I was also able to visualize these complexes from the fractions in the glycerol gradients where the core complex was present, providing further evidence that I am indeed visualizing the core complex and not some contaminating complex. WT core complexed isolated from co-culture of *H. pylori* and gastric epithelial cells have the exact same morphology as ones isolated from pure culture, so there is no evidence of a conformational shift caused by active CagA translocation.

I was able to image two mutant core complexes as well, the Cag3 mutant (missing Cag3) and the CagT mutant (missing Cag3 and CagT). These two complexes had a remarkably similar morphology: the outer ring was lost, with only the central ring remaining. These data imply that Cag3 and CagT are on the outer ring of the core complex, a claim further supported by immunogold labeling which localizes Cag3 to periphery. Cag3 and CagT probably do not make up the *entire* outer ring. Both the Cag3 and the CagT mutants have the exact same morphology where the entire outer ring is missing, and the central ring remains completely intact, even though the CagT mutant is missing both CagT and Cag3. What could

98

be happening is that the loss of the Cag3 causes a loss in the structural integrity of the outer ring, so I can no longer visualize it with EM. That same loss persists in a CagT (which lacks both Cag3 and CagT) mutant, but since I was already visualizing a loss of structural integrity, there doesn't seem to be an additional with the loss of CagT, which tells me that CagT is likewise on the periphery and not in the central ring.

Finally, I wanted to be able to obtain a high-resolution structure of the core complex using single particle cryoEM. Over the past few years, I have attempted many different strategies to make this purification amenable for cryo conditions: changing which protein is tagged, adjusting the mixture of detergents, and attempting various concentrating strategies. None of these strategies to date have been successful. I also tried to obtain a low-resolution 3D structure of the core complex using RCT, but the data set I picked did not yield a 3D structure.

Despite these setbacks, I was still able to further our knowledge about the *cag* T4SS, and more broadly as T4SS as a whole.   Prior to my studies of the *cag* T4SS all the detailed structural work of T4SS core complexes had been utilizing *E coli* conjugation systems. This literature showed the core complex to be a small ring-like structure about 19 nm in diameter, composed of three proteins (16, 21). In contrast, the *cag* core complex is much larger in size, at 41 nm, and composed of five proteins instead of three. Interestingly though, the central ring of the *cag* core complex is the same size as the *E. coli* core complex and composed of the only proteins with sequence similarity to other systems beyond *H. pylori.*

The body of literature on effector protein translocating T4SSs is quite

small. However, a few years ago, EM was used to define the structure of

Legionella's Dot/Icm core complex (31). This core complex had far more

similarities with the *cag* core complex than the *E. coli* core complex. The Dot/Icm

core complex is also composed of five proteins, and is likewise quite large, about

38 nm in diameter. It has a similar overall structure though, namely a large ring

with a central pore, though it does not appear to have spokes connecting the pore

to the outer ring (31). Another study used cryoEM tomography to visualize the

Dot/Icm core complex *in situ*, which further emphasized the similarities between

the Dot/Icm core complex and the *cag* core complex. The cryoEM tomography

revealed that the Dot/Icm core complex was a large complex with a stalk-like

domain which points into the periplasm and inner membrane (58). Strikingly, an

overlay of the *cag* core complex perfectly matches the Dot/Icm core complex in

both size and overall morphology.

I believe that this morphology, now seen in both *H. pylori* and *L.*

pneumophila could be a paradigm of effector protein translocating T4SSs. This

illustrates the vast difference between conjugative T4SSs and effector

translocating T4SSs, unsurprisingly since their functions are quite diverse. The

diversity of T4SS requires continued research, since a single model is not

adequate to describe the intricacies of this system.

Future Directions

My work detailed here provides important new insights into the structural

organization of the *cag* T4SS core complex, but there is still much that needs to be

elucidated. Foremost, obtaining a high-resolution structure of the core complex is an important goal. I have attempted to modify my purification methods and conditions in order to make the preparation compatible with cryoEM, but thus far these efforts have not been successful. There are still more avenues to attempt, including modifying the protocol with amphipols, or perhaps even trying to express parts or the entirety of the core complex recombinantly. Furthermore, there are other small tags that I have not yet taken advantage of yet. For example, the avi-tag is a 15-amino acid tag that is specifically biotinylated in the presence of biotin, biotin ligase and ATP. I could then use streptavidin to pull down the biotinylated protein, and perhaps purify the core complex, if I tag the appropriate protein. I would want to tag CagF with this since this was the protein I tagged to initially purify the core complex, as well as Cag3, since I was able to nonspecifically purify the core complex using a His-tag on Cag3. Presumably, the high specificity of the biotin-streptavidin interaction will yield a cleaner more highly purified preparation than that which was obtained using a 13xHis tag.

Besides the high resolution structure, there are several other avenues of research that are high priority for further study. In all my studies I have only ever been able to purify the core complex and none of the other complexes of the *cag* T4SS. The core complex should have proximity to the inner membrane core, ATPases, and pilus, according to models in *E. coli.* This could represent yet another divergence between these two systems, but this cannot be proven with a lack of data. In order to delve into this question I would like to take advantage of a relatively new technology: BioID. This technology takes advantage of a point

mutant of biotin ligase (BirA*); instead of biotinylating at specific residues, this enzyme will promiscuously biotinylate all proteins within a ~10nm proximity (59). If I engineered this protein into members of the T4SS I could then identify the proteins biotinylated and I could start to build a map of the proteins within the T4SS. One drawbacks of this technique, is the large size of BirA* (372 amino acids) which might disrupt the structure of the core complex. CagF and CagA are potentially good candidates for fusion, since they lightly interact with the core complex, but are not integral members. By identifying the proteins within close proximity to CagA or CagF, I could potentially learn if CagA interacts with CagY or CagX. Furthermore, I could perform the same experiment when the bacteria is actively translocating CagA and in contact with gastric epithelial cells. This could potentially reveal new interactions between CagA and CagF and these proteins would presumably be important for CagA translocation, and also could indicate a change or shift in the T4SS machinery to allow for translocation.

One member of the core complex that I could potentially label with this form of biotin ligase would be CagY. It is a very large protein, and I hypothesize that the CagY N-terminus does not interact with the rest of the core complex members, merely interfaces with CagA. I hope then, that by introducing in BirA* to this region I could assess the proteins closest to CagY, and it would be interesting to see if that would be the ATPases, or the inner membrane core. This could also be used on a variety of under studied Cag protein to identify their interactions with other members of the *cag* PAI and elucidate their role in the T4SS machinery.

Besides investigating the intimate interactions between the various components of the Cag machinery, there is a lot of value in figuring out the proteins required for core complex formation. Two proteins that might be required for the formation of the core complex are CagX and CagY. If either one of those genes are deleted, then the core complex can no longer be isolated using HA-CagF. However this could mean that either CagX and CagY are required for the interaction of CagF with the core complex, or that they are required for the formation of the core complex. To answer this question, I need an alternative way to purify the core complex, and luckily I can do this with the FLAG-tagged version of CagT. I could engineer both of these mutant strains to express the FLAG-tagged CagT and then assess isolation of the core complex by western blotting. I hypothesize that in the absence of CagY I will still be able to isolate the core complex. I think in this *cag* system it acts more as an adaptor protein to link CagA and the *cag* T4SS machinery rather than an integral member of the machinery. While CagY is required for both CagA translocation and IL-8 induction, in its absence pili are still formed. Therefore I think it is far more likely that CagX is required for formation of the core complex. This same strategy could be applied to assess if CagA is required for core complex formation, though I consider this to be unlikely since IL-8 induction is still preserved in the absence of CagA.

Another interesting question is to determine where and how CagA is interacting with the core complex. Hopefully the bioID technique could shed some light on the matter. I have not yet assessed if CagA and CagF are purified when the core complex is purified using the CagT-FLAG construct, but if they can be,

this could open up some avenues of research. We know that CagF is required for

CagA translocation, but we do not know through which mechanism. As a matter of

fact, we know very little of the mechanism of CagA translocation at all. Work by

Bonsor et al showed that CagF bound to CagA in multiple regions, and they

hypothesized that this binding would "open up" CagA, and thus reveal the residues

required to interface with the T4SS. I could test this hypothesis by deleting CagF

and seeing if CagA will still interact with the T4SS, through either the BioID

method, or with the CagT-FLAG construct.

It is also important to determine how CagA can translocate through the

T4SS, now that we have an idea of what part of that structure looks like. One

possibility is that CagA may pass directly through the central pore visible within the

core complex. Recently, 876 residues of the amino-terminus of CagA have been

crystalized, revealing dimensions of 8 by 11 by 5.5 nm (48). A structure of this

size could conceivably fit through the ~10-nm pore of the core complex while still

folded. However, the mechanisms by which CagA enters host cells after

translocation across the bacterial envelope remain a topic of inquiry. It is

presumed that the T4SS-associated pili assembled at the bacterium-host cell

interface contribute to the entry of CagA into host cells (34, 49, 50). And it has

been suggested that binding of CagA to phosphatidylserine or integrins on the

surface of host cells also may be important for entry into host cells (51, 52).

Another question is to determine the orientation of the core complex within

the bacterial membrane. It is tempting to speculate that the stalk portion is

oriented to the periplasm and inner membrane that CagA interacts with the

stalk. Interestingly, there is cryotomography data of the *L. pneumophila* core complex *in situ* which shows that is has very similar dimensions as the *cag* T4SS core complex, as well as a long, stalk-like domain, and stalk-like domain points into the periplasm (58). So, it is likely that the *cag* T4SS has a similar orientation. However, it would be ideal to have experimental data to support this hypothesis. Fortunately there are some tools at our disposal to answer this question. Previous work in our lab optimized a protocol to biotinylate the surface exposed residues of outer membrane proteins in *H. pylori*. I could use this same protocol, and then isolate the core complex and determine by immunoblotting which proteins are biotinylated. I would then complement this approach by using streptavidin conjugated to nano gold particles to identify by EM which areas of the core complex are biotinylated. If the stalk is labeled, it is likely that the stalk is surface exposed; if the stalk is not labeled and the ring like regions are labeled, this would suggest that the stalk is pointing into the periplasm.

Finally there were some interesting results from IPs I performed from bacteria in contact with gastric epithelial cells that I was unable to follow up. Namely, when we performed mass spec on the samples, we identified not only bacterial proteins also identified eukaryotic proteins that were enriched by the IP. We recovered Par1b/MARK2, which has been previously identified as a binding partner of CagA (60). We also recovered significant amounts of a "general transcription factor II (TFII-IB)" as well as both the S1008 and S100A9 subunits of the antimicrobial protein, calprotectin. It is important to note that these interactions could either be between the core complex components or CagA; the eukaryotic

105

cells are lysed with the bacterial cells, and incubated together. Since CagF is the protein I am directly pulling down, this interaction can either purify CagA alone, or CagA and the core complex. To determine which proteins the eukaryotic proteins are interacting with, it would be vital to perform a glycerol gradient on the samples, and assess where the eukaryotic proteins co-migrate.

However, both of these are intriguing results for different reasons. The nuclear isoform of TFII-IB has been shown to suppress the expression of c-fos by binding to its promoter (61), while *cagA* positive strains will actually induce the expression of c-fos (62). It is possible then, that CagA could be binding to TFII-IB, thus making it unable to repress expression of c-fos. EMSA of TFII-IB with this promoter sequence either incubated with or without CagA could start to answer this question.   What is intriguing about the calprotectin result is that previous studies have shown that calprotectin actually lowers the activity of the *cag* T4SS (63). Calprotectin could be directly binding the core complex, and somehow, suppressing its efficiency of formation, since the previous study assesses T4SS function through analysis of IL-8 induction and CagA translocation.

There are a variety of directions for the study of the T4SS. The focus can be on the actual mechanism of CagA translocation, others can elucidate the various protein interactions that make up the machinery of the T4SS and finally, there can also be focus on the interaction that the effector protein or the T4SS itself makes with host eukaryotic proteins. Regardless of which avenue is chosen, all of these studies will provide important insights to a highly relevant virulence factor of the gastric pathogen *H. pylori.*

REFERENCES

1.   Robin Warren J, Marshall B. 1983. UNIDENTIFIED CURVED BACILLI ON GASTRIC EPITHELIUM IN ACTIVE CHRONIC GASTRITIS. The Lancet 321:1273-1275.

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

3.   Forman D, Newell DG, Fullerton F, Yarnell JW, Stacey AR, Wald N, Sitas F. 1991. Association between infection with Helicobacter pylori and risk of gastric cancer: evidence from a prospective investigation. BMJ 302:1302-1305.

4.   Fuchs CS, Mayer RJ. 1995. Gastric Carcinoma. New England Journal of Medicine 333:32-41.

5.   Censini S, Lange C, Xiang Z, Crabtree JE, Ghiara P, Borodovsky M, Rappuoli R, Covacci A. 1996. cag, a pathogenicity island of Helicobacter pylori, encodes type I-specific and disease-associated virulence factors. Proceedings of the National Academy of Sciences 93:14648-14653.

6.   Covacci A, Censini S, Bugnoli M, Petracca R, Burroni D, Macchia G, Massone A, Papini E, Xiang Z, Figura N. 1993. Molecular characterization of the 128-kDa immunodominant antigen of Helicobacter pylori associated with cytotoxicity and duodenal ulcer. Proceedings of the National Academy of Sciences 90:5791-5795.

7.   Akopyants NS, Clifton SW, Kersulyte D, Crabtree JE, Youree BE, Reece CA, Bukanov NO, Drazek ES, Roe BA, Berg DE. 1998. Analyses of the cag pathogenicity island of Helicobacter pylori. Molecular Microbiology 28:37-53.

8.   Alm RA, Ling LS, Moir DT, King BL, Brown ED, Doig PC, Smith DR, Noonan B, Guild BC, deJonge BL, Carmel G, Tummino PJ, Caruso A, Uria-Nickelsen M, Mills DM, Ives C, Gibson R, Merberg D, Mills SD, Jiang Q, Taylor DE, Vovis GF, Trust TJ. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori. Nature 397:176-180.

9.   Tomb J-F, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, Ketchum KA, Klenk HP, Gill S, Dougherty BA, Nelson K, Quackenbush J, Zhou L, Kirkness EF, Peterson S, Loftus B, Richardson D, Dodson R, Khalak HG, Glodek A, McKenney K, Fitzegerald LM, Lee N, Adams MD, Hickey EK, Berg DE, Gocayne JD, Utterback TR, Peterson JD, Kelley JM, Cotton MD, Weidman JM, Fujii C, Bowman C, Watthey L, Wallin E, Hayes WS, Borodovsky M, Karp PD, Smith HO, Fraser CM, Venter JC. 1997. The

complete genome sequence of the gastric pathogen Helicobacter pylori. Nature 388:539-547.

10. Blaser MJ, Perez-Perez GI, Kleanthous H, Cover TL, Peek RM, Chyou PH, Stemmermann GN, Nomura A. 1995. Infection with Helicobacter pylori Strains Possessing cagA Is Associated with an Increased Risk of Developing Adenocarcinoma of the Stomach. Cancer Research 55:2111-2115.

11. 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.

12. Stein M, Bagnoli F, Halenbeck R, Rappuoli R, Fantl WJ, Covacci A. 2002. c-Src/Lyn kinases activate Helicobacter pylori CagA through tyrosine phosphorylation of the EPIYA motifs. Molecular Microbiology 43:971-980.

13. Higashi H, Tsutsumi R, Muto S, Sugiyama T, Azuma T, Asaka M, Hatakeyama M. 2002. SHP-2 Tyrosine Phosphatase as an Intracellular Target of Helicobacter pylori CagA Protein. Science 295:683-686.

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

15. Viala J, Chaput C, Boneca IG, Cardona A, Girardin SE, Moran AP, Athman R, Memet S, Huerre MR, Coyle AJ, DiStefano PS, Sansonetti PJ, Labigne A, Bertin J, Philpott DJ, Ferrero RL. 2004. Nod1 responds to peptidoglycan delivered by the Helicobacter pylori cag pathogenicity island. Nat Immunol 5:1166-1174.

16. Fronzes R, Schäfer E, Wang L, Saibil HR, Orlova EV, Waksman G. 2009. Structure of a Type IV Secretion System Core Complex. Science 323:266-268.

17. Wallden K, Rivera-Calzada A, Waksman G. 2010. Microreview: Type IV secretion systems: versatility and diversity in function. Cellular Microbiology 12:1203-1212.

18. Alvarez-Martinez CE, Christie PJ. 2009. Biological Diversity of Prokaryotic Type IV Secretion Systems. Microbiology and Molecular Biology Reviews 73:775-808.

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

20. Krall L, Wiedemann U, Unsin G, Weiss S, Domke N, Baron C. 2002. Detergent extraction identifies different VirB protein subassemblies of the type IV secretion machinery in the membranes of Agrobacterium tumefaciens. Proceedings of the National Academy of Sciences 99:11405-11410.

21. Chandran V, Fronzes R, Duquerroy S, Cronin N, Navaza J, Waksman G. 2009. Structure of the outer membrane complex of a type IV secretion system. Nature 462:1011-1015.

22. Cascales E, Christie PJ. 2004. Definition of a bacterial type IV secretion pathway for a DNA substrate. Science 304:1170-1173.

23. Cascales E, Christie PJ. 2004. Agrobacterium VirB10, an ATP energy sensor required for type IV secretion. Proc. Natl. Acad. Sci. U. S. A. 101:17228-17233.

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

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

26. Johnson EM, Gaddy JA, Voss BJ, Hennig EE, Cover TL. 2014. Genes Required for Assembly of Pili Associated with the Helicobacter pylori cag Type IV Secretion System. Infection and Immunity 82:3457-3470.

27. 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. Infection and Immunity 74:273-281.

28. 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.

29. Bonsor DA, Weiss E, Iosub-Amir A, Reingewertz TH, Chen TW, Haas R, Friedler A, Fischer W, Sundberg EJ. 2013. Characterization of the Translocation-competent Complex between the Helicobacter pylori Oncogenic Protein CagA and the Accessory Protein CagF. Journal of Biological Chemistry 288:32897-32909.

30. Low HH, Gubellini F, Rivera-Calzada A, Braun N, Connery S, Dujeancourt A, Lu F, Redzej A, Fronzes R, Orlova EV, Waksman G. 2014. Structure of a type IV secretion system. Nature 508:550-553.

31. Kubori T, Koike M, Bui XT, Higaki S, Aizawa S-I, Nagai H. 2014. Native structure of a type IV secretion system core complex essential for Legionella pathogenesis. Proceedings of the National Academy of Sciences 111:11804-11809.

32. Hatakeyama M. 2008. Linking epithelial polarity and carcinogenesis by multitasking Helicobacter pylori virulence factor CagA. Oncogene 27:7047-7054.

33. Ohnishi N, Yuasa H, Tanaka S, Sawa H, Miura M, Matsui A, Higashi H, Musashi M, Iwabuchi K, Suzuki M, Yamada G, Azuma T, Hatakeyama M. 2008. Transgenic expression of Helicobacter pylori CagA induces gastrointestinal and hematopoietic neoplasms in mouse. Proceedings of the National Academy of Sciences 105:1003-1008.

34. Shaffer CL, Gaddy JA, Loh JT, Johnson EM, Hill S, Hennig EE, McClain MS, McDonald WH, Cover TL. 2011. <italic>Helicobacter pylori</italic> Exploits a Unique Repertoire of Type IV Secretion System Components for Pilus Assembly at the Bacteria-Host Cell Interface. PLoS Pathog 7:e1002237.

35. Kutter S, Buhrdorf R, Haas J, Schneider-Brachert W, Haas R, Fischer W. 2008. Protein Subassemblies of the Helicobacter pylori Cag Type IV Secretion System Revealed by Localization and Interaction Studies. Journal of Bacteriology 190:2161-2171.

36. Odenbreit S, Püls J, Sedlmaier B, Gerland E, Fischer W, Haas R. 2000. Translocation of Helicobacter pylori CagA into Gastric Epithelial Cells by Type IV Secretion. Science 287:1497-1500.

37. Bourzac KM, Botham CM, Guillemin K. 2007. Helicobacter pylori CagA Induces AGS Cell Elongation through a Cell Retraction Defect That Is Independent of Cdc42, Rac1, and Arp2/3. Infection and Immunity 75:1203-1213.

38. Busler VJ, Torres VJ, McClain MS, Tirado O, Friedman DB, Cover TL. 2006. Protein-Protein Interactions among Helicobacter pylori Cag Proteins. Journal of Bacteriology 188:4787-4800.

39. Pinto-Santini DM, Salama NR. 2009. Cag3 Is a Novel Essential Component of the Helicobacter pylori Cag Type IV Secretion System Outer Membrane Subcomplex. Journal of Bacteriology 191:7343-7352.

40. Voss BJ, Gaddy JA, McDonald WH, Cover TL. 2014. Analysis of Surface-Exposed Outer Membrane Proteins in Helicobacter pylori. Journal of Bacteriology 196:2455-2471.

41. Olbermann P, Josenhans C, Moodley Y, Uhr M, Stamer C, Vauterin M,

Suerbaum S, Achtman M, Linz B. 2010. A Global Overview of the Genetic and Functional Diversity in the <italic>Helicobacter pylori cag</italic> Pathogenicity Island. PLoS Genet 6:e1001069.

42. Christie PJ. 2004. Type IV secretion: the *Agrobacterium* VirB/D4 and related conjugation systems. Biochimica et Biophysica Acta (BBA)- Molecular Cell Research 1694:219-234.

43. Delahay RM, Balkwill GD, Bunting KA, Edwards W, Atherton JC, Searle MS. 2008. The Highly Repetitive Region of the Helicobacter pylori CagY Protein Comprises Tandem Arrays of an α-Helical Repeat Module. Journal of Molecular Biology 377:956-971.

44. Barrozo RM, Cooke CL, Hansen LM, Lam AM, Gaddy JA, Johnson EM, Cariaga TA, Suarez G, Peek RM, Jr., Cover TL, Solnick JV. 2013. Functional Plasticity in the Type IV Secretion System of Helicobacter pylori. PLOS Pathogens 9:e1003189.

45. Ohi M, Li Y, Cheng Y, Walz T. 2004. Negative staining and image classification—powerful tools in modern electron microscopy. Biological procedures online 6:23-34.

46. Shaikh TR, Gao H, Baxter WT, Asturias FJ, Boisset N, Leith A, Frank J. 2008. SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. Nat. Protocols 3:1941-1974.

47. Sarkar MK, Husnain SI, Jakubowski SJ, Christie PJ. 2013. Isolation of Bacterial Type IV Machine Subassemblies, p. 187-204. *In* Delcour AH (ed.), Bacterial Cell Surfaces: Methods and Protocols. Humana Press, Totowa, NJ.

48. Hayashi T, Senda M, Morohashi H, Higashi H, Horio M, Kashiba Y, Nagase L, Sasaya D, Shimizu T, Venugopalan N, Kumeta H, Noda Nobuo N, Inagaki F, Senda T, Hatakeyama M. 2012. Tertiary Structure-Function Analysis Reveals the Pathogenic Signaling Potentiation Mechanism of Helicobacter pylori Oncogenic Effector CagA. Cell Host & Microbe 12:20-33.

49. Rohde M, Püls J, Buhrdorf R, Fischer W, Haas R. 2003. A novel sheathed surface organelle of the Helicobacter pylori cag type IV secretion system. Molecular microbiology 49:219-234.

50. Kwok T, Zabler D, Urman S, Rohde M, Hartig R, Wessler S, Misselwitz R, Berger J, Sewald N, Konig W, Backert S. 2007. Helicobacter exploits integrin for type IV secretion and kinase activation. Nature 449:862-866.

51. Murata-Kamiya N, Kikuchi K, Hayashi T, Higashi H, Hatakeyama M. 2010.

Helicobacter pylori Exploits Host Membrane Phosphatidylserine for Delivery, Localization, and Pathophysiological Action of the CagA Oncoprotein. Cell Host & Microbe 7:399-411.

52.   Kaplan-Türköz B, Jiménez-Soto LF, Dian C, Ertl C, Remaut H, Louche A, Tosi T, Haas R, Terradot L. 2012. Structural insights into Helicobacter pylori oncoprotein CagA interaction with β1 integrin. Proceedings of the National Academy of Sciences 109:14640-14645.

53.   Liao M, Cao E, Julius D, Cheng Y. 2013. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. Nature 504:107-112.

54.   Tribet C, Audebert R, Popot J-L. 1996. Amphipols: Polymers that keep membrane proteins soluble in aqueous solutions. Proceedings of the National Academy of Sciences 93:15047-15050.

55.   Bondos SE, Bicknell A. 2003. Detection and prevention of protein aggregation before, during, and after purification. Analytical Biochemistry 316:223-231.

56.   Cover TL, Blaser MJ. 2009. Helicobacter pylori in Health and Disease. Gastroenterology 136:1863-1873.

57.   Hatakeyama M. 2014. Helicobacter pylori CagA and Gastric Cancer: A Paradigm for Hit-and-Run Carcinogenesis. Cell Host & Microbe 15:306-316.

58.   Ghosal D, Chang Y-W, Jeong KC, Vogel JP, Jensen GJ. 2016. Structure of the Legionella Dot/Icm type IV secretion system in situ by electron cryotomography. bioRxiv.

59.   Kim DI, KC B, Zhu W, Motamedchaboki K, Doye V, Roux KJ. 2014. Probing nuclear pore complex architecture with proximity-dependent biotinylation. Proceedings of the National Academy of Sciences 111:E2453-E2461.

60.   Saadat I, Higashi H, Obuse C, Umeda M, Murata-Kamiya N, Saito Y, Lu H, Ohnishi N, Azuma T, Suzuki A, Ohno S, Hatakeyama M. 2007. Helicobacter pylori CagA targets PAR1/MARK kinase to disrupt epithelial cell polarity. Nature 447:330-333.

61.   Roy AL. 2007. Signal-Induced Functions of the Transcription Factor TFII-I. Biochimica et biophysica acta 1769:613-621.

62.   Meyer-ter-Vehn T, Covacci A, Kist M, Pahl HL. 2000. Helicobacter pylori Activates Mitogen-activated Protein Kinase Cascades and Induces Expression of the Proto-oncogenes c-fos and c-jun. Journal of Biological Chemistry 275:16064-16072.

63.   Gaddy JA, Radin JN, Loh JT, Piazuelo MB, Kehl-Fie TE, Delgado AG, Ilca

FT, Peek RM, Cover TL, Chazin WJ, Skaar EP, Scott Algood HM. 2014. The Host Protein Calprotectin Modulates the Helicobacter pylori cag Type IV Secretion System via Zinc Sequestration. PLOS Pathogens 10:e100445