

INVESTIGATION OF SURFACE-EXPOSED PROTEINS OF THE GASTRIC
BACTERIUM *HELICOBACTER PYLORI*

By

Bradley J. Voss

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, 2016

Nashville, Tennessee

Approved:

Eric P. Skaar, PhD, MPH

Andrew J. Link, PhD

Maria Hadjifrangiskou, PhD

Richard M. Peek, MD

Timothy L. Cover, MD

To my parents – you never stopped believing in me

ACKNOWLEDGEMENTS

First I want to thank my mentor Dr. Timothy Cover for all the years of support and guidance. He gave me a chance in the summer of 2010 and let me to do some research for a couple months. I ended up staying six years. He is probably the most patient person you will ever meet and he always treated me with the utmost respect. Despite my numerous failures, he always had something positive to say. A blot with no bands is a “very clean blot.” He is also very supportive; I am sure I told him “no” more times than he told me. Tim, thank you for allowing me to become a member of your lab and making me the scientist I am today.

I also want to thank every member of the Cover lab for tolerating me all this time. I was blessed to be able to come to work every morning to find an easy-going and supportive environment with everyone always happy to help out. I especially want to thank John and Mark who were endless sources of information and guidance that always came with a sense of humor. I also want to thank Arwen for always being willing get a cup of coffee with a joke never out of reach.

I especially want to thank all of my friends and family. My parents have always supported me in everything I have done. It did not matter if it was a school function or sporting event, they were always in the crowd cheering me on. Not least of all, they were always willing to make the long drive from Michigan to visit and catch a hockey game. I want to thank my sister Lindsay for her constant support. And finally, I want to thank my best friend and companion Whitney for being there for me every step of our journey now and into the future, wherever we end up.

This work was supported by the NIH (R01 AI068009, AI039657, and CA116087), the Department of Veterans Affairs, and the Vanderbilt University Digestive Disease Research Center (NIH grant P30DK058404).

TABLE OF CONTENTS

	Page
DEDICATION	ii
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS	xi
Chapter	
I. INTRODUCTION	1
Brief history of <i>Helicobacter pylori</i> and epidemiology	1
The surface of <i>Helicobacter pylori</i>	2
OMPs of <i>H. pylori</i>	4
Other known <i>H. pylori</i> surface-exposed virulence factors	8
Research significance and specific aims	11
Identification of surface-exposed proteins of <i>H. pylori</i>	11
Investigation of CagT, a putative VirB7 functional analog	12
II. IDENTIFICATION OF SURFACE-EXPOSED PROTEINS OF <i>HELICOBACTER</i> <i>PYLORI</i> USING COMPLEMENTARY TECHNIQUES	14
Introduction	14
Materials and Methods	18
Bacterial strains and culture conditions	18
Generation of rabbit polyclonal antisera	18
Immunoblot analysis	18
Immunogold transmission electron microscopy	19
Biotinylation of surface-exposed <i>H. pylori</i> proteins using an amine-reactive biotinylation reagent	20
Biotinylation of surface-exposed <i>H. pylori</i> proteins using a carboxyl-reactive	

biotinylation reagent	21
Purification of biotinylated proteins	22
Bacterial subcellular fractionation	22
Proteinase K treatment of intact bacteria	23
Mass spectrometric analysis of samples	24
Analysis of mass spectrometry data	25
Mapping of protein segments that are susceptible to proteinase K digestion	29
Results	30
Identification of surface-exposed proteins	30
Proteins meeting multiple criteria for surface-exposed outer membrane localization	39
Immunogold EM analysis of CagT	43
Experimental evidence of amino-terminal signal sequence cleavage	47
Multiple classes of C-terminal motifs in surface-exposed proteins	51
Susceptibility of surface-exposed proteins to proteinase K digestion and cell-surface topology of surface-exposed proteins	57
Discussion	71

III. ALTERATION OF THE *HELICOBACTER PYLORI* MEMBRANE PROTEOME IN RESPONSE TO CHANGES IN ENVIRONMENTAL SALT CONCENTRATION

Introduction	83
Methods	85
Bacterial strains and growth conditions	85
Sample preparation for proteomic analysis	86
Biotinylation of surface-exposed proteins	86
Label-free analysis	87
iTRAQ analysis	90
Statistical analysis of assigned spectra	93
<i>H. pylori</i> motility assay	96
Results	97
Salt-responsive proteins in strain 26695 identified by label-free and iTRAQ analyses	97
Salt-responsive proteins in strain 7.13 identified by label-free and iTRAQ analyses	106
Salt-responsive proteins identified in both strains 26695 and 7.13	110
Changes in abundance of surface-exposed proteins in response to salt concentration	111
Salt-dependent changes in bacterial motility	115
Discussion	118

IV. INVESTIGATION OF CAGT, A PUTATIVE VIRB7 FUNCTIONAL ANALOG	124
Introduction	124
Methods	126
Bacterial strains and growth conditions	126
Bacterial strain mutagenesis.....	126
Palmitic acid incorporation assay.....	127
Immunoblot.....	128
IL-8 induction assay.....	128
Bacterial subcellular fractionation	128
Surface-exposed FLAG-tag assay.....	129
Super resolution fluorescence microscopy	132
Results	133
Investigation of the CagT lipidation status	133
Effect of disruption of the lipbox on steady-state levels of CagT.....	136
CagT localization	144
CagT surface topology.....	147
Discussion.....	154
V. CONCLUSIONS	158
Summary.....	158
Future directions	166
Impact of surface-exposed OMPs on pathogenesis	166
Effect of identified salt-responsive proteins on pathogenesis	167
Further studies of salt-responsive changes in protein abundance.....	168
Is CagT a lipoprotein in <i>H. pylori</i> ?.....	170
<i>H. pylori</i> localization of surface-exposed lipoproteins	170
Localization of CagT in coculture.....	171
Last Remarks	173
LIST OF PUBLICATIONS	175
REFERENCES.....	176

LIST OF TABLES

Table	Page
1. Different families of predicted <i>H. pylori</i> OMPs	6
2. Summary of mass spectrometry data	33
3. Proteins meeting multiple criteria for surface-exposed outer membrane proteins	42
4. Detected amino-terminal non-tryptic cleavage events in the three biotinylated preparations.....	48
5. Surface-exposed proteins highly susceptible to proteinase K digestion, based on total assigned spectra.....	58
6. Proteins identified as susceptible to extracellular proteinase K digestion, based on analysis of individual amino acid sites	64
7. Protease susceptibility of annotated OMPs	66
8. Salt-responsive proteins in strain 26695 identified by iTRAQ analysis	99
9. Salt-responsive proteins in strain 7.13 identified by iTRAQ analysis	107
10. Salt-responsive proteins in strain 26695 identified by analysis of biotinylated intact bacteria	113
11. Salt-responsive flagellar proteins from strain 7.13.....	116

LIST OF FIGURES

Figure	Page
1. Surface structures of <i>H. pylori</i>	10
2. Criteria for identifying enriched proteins	27
3. Analysis of purified biotinylated proteins.....	31
4. Enrichment of proteins with an “outer membrane protein” annotation	37
5. Use of multiple criteria to identify surface-exposed outer membrane proteins ...	41
6. Localization of CagT by immunogold EM analysis	45
7. Experimentally detected sites of signal peptide cleavage.....	50
8. Analysis of a conserved C-terminal motif	53
9. Comparison of surface-exposed proteins identified by carboxyl-reactive biotinylation reagent compared to amine-reactive biotinylation	56
10. Susceptibility of surface-exposed proteins to proteinase K treatment	62
11. Resistance of predicted β -barrel regions to digestion by proteinase K.....	69
12. Subset of surface-exposed outer membrane proteins that are susceptible to digestion by proteinase K	77
13. Proposed model of proteinase K-susceptibility results	80
14. Distribution of R_{sc} or iTRAQ \log_2 ratios (protein abundance in bacteria from high salt conditions compared to bacteria from low salt conditions).....	95
15. Identification of salt-responsive proteinase by multiple methods	101
16. Comparison of iTRAQ results with label-free spectral counting results	104
17. Effect of salt concentration on <i>H. pylori</i> motility	117

18. Sites for CagT FLAG epitope insertion in CagT.....	131
19. Incorporation of palmitic acid into recombinant CagT	135
20. Schematic of CagT lipidation mutants	138
21. Rescue of C21S mutation.....	139
22. IL-8 induction and immunoblot of non-lipidated CagT	142
23. Triton subcellular fractionation of wt <i>H. pylori</i> and CagT _{S22D}	145
24. Localization of CagT in a Δ PAI background.....	146
25. CagT steady-state abundance and IL-8 induction phenotype	148
26. Intact cell immunolabeling of FLAG-tagged strains	150
27. Proposed model of CagT topology	151
28. Super resolution of FLAG-tagged strain 27K in pure culture	153

LIST OF ABBREVIATIONS

2D	Two-dimensional
2D-DIGE	Two-dimensional difference gel electrophoresis
ACN	Acetonitrile
AGC	Automatic gain control
Amine-PEG3-biotin	(+)-biotinyl-3,6,9,-trioxaundecanediamine
BB-FBS	Sulfite-free Brucella broth supplemented with 10% fetal bovine serum
BCA	Bicinchoninic acid
<i>cag</i> PAI	Cytotoxin associated gene pathogenicity island
cat	Chloramphenicol acetyltransferase cassette
CP,PP	Cytoplasmic and periplasmic
C-terminal	Carboxy-terminal
ECL	Enhanced chemiluminescence
EDC	1-Ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme linked immunosorbent assay
EM	Electron microscopy
FBS	Fetal bovine serum
FDR	False discovery rate
FET-BH	Fisher's exact test with Benjamini-Hochberg multiple test correction

LPS	Lipopolysaccharide
HCD	Higher-energy collisional dissociation
HPLC	High-performance liquid chromatography
IL-8	Interleukin 8 (CXCL8)
IM	Inner membrane
iTRAQ	Isobaric tags for relative absolute quantitation
LC-MS/MS	Liquid chromatography tandem mass spectrometry
LF	Label free MudPIT
LTQ	Linear trap quadrupole
m/z	Mass-to-charge ratio
MES	2-(N-morpholino)ethanesulfonic acid
MMTS	Methyl methanethiosulfonate
MOI	Multiplicity of infection
MS	Mass spectrometer
MS/MS	Tandem mass spectrometry
MSG	Monosodium glutamate
MudPIT	Multidimensional protein identification technology
NCBI BLAST	National center for biotechnology information basic local alignment search tool
N-terminal	Amino-terminal
OM	Outer membrane
OMP	Outer membrane protein
PA	Palmitic acid alkyne

PBS	Phosphate buffered saline
PFA	Paraformaldehyde
PMSF	Phenylmethanesulfonyl fluoride
RIPA	50 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA, 1% NP-40, 0.25% sodium deoxycholate
R _{sc}	Ratio of spectral counts
SCX	Strong cation exchange
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
S-NHS	N-hydroxysulfosuccinimide
SPII	Signal peptidase II
Sulfo-NHS-LC-biotin	Sulfosuccinimidyl-6-[biotin-amido]hexanoate
ΔT/T	<i>cagT</i> clean knock out complemented with <i>ureA::cagT</i>
T4SS	Type IV secretion system
TBS	Tris buffered saline
TCA	Trichloroacetic acid
TCEP	Tris(2-carboxyethyl)phosphine
TEA	Triethanolamine
TEAB	Triethyl ammonium bicarbonate
TEM	Transmission electron microscopy
TFA	Trifluoroacetic acid
TKE	50 mM Tris pH 7.4, 150 mM KCl, 10 mM EDTA

TKEZ	50 mM Tris pH 7.4, 150 mM KCl, 10 mM EDTA, 2% Zwittergent 3-14
TM	Total membrane
TNKCM	50 mM Tris, pH 7.4, 100 mM NaCl, 27 mM KCl, 1 mM CaCl ₂ , 0.5 mM MgCl ₂
TSS	Transcriptional start site
wt	Wild type
ΔPAI	<i>cag</i> PAI null <i>H. pylori</i> strain

CHAPTER I

INTRODUCTION

Brief history and epidemiology of *Helicobacter pylori*

Prior to 1984, the human stomach was considered a sterile organ. It was thought that the low gastric pH would prevent bacteria from growing or persisting in this environment (89). Any bacteria found in the stomach were considered to be transient and likely to be constituents of saliva or swallowed food. Then in 1984 a group reported isolation of a bacterium from the stomachs of patients with gastritis (89). The isolated bacterium was eventually named *Helicobacter pylori*. In 1994, the World Health Organization designated *H. pylori* as a group 1 carcinogen (52). *H. pylori* colonization results in gastric inflammation and is a risk factor for the development of gastric adenocarcinoma, gastric lymphoma, and peptic ulcer disease (4, 112).

Typically, the host is colonized by *H. pylori* early in life and colonization persists lifelong (5, 21). The actual mode of transmission is presently not understood. *H. pylori* currently colonizes approximately half of the world's population (5). The actual rate of colonization is different for different parts of the world. In Africa, for example, the rate of colonization is close to 90%, while this rate is approximately 50% in North America (29). It is worth noting that the rate of severe disease outcomes varies by region. While the rate of *H. pylori* colonization in Africa is very high, the age adjusted incidence of gastric cancer in Africa is relatively low (2%) (115). Conversely, while the rate of *H. pylori* colonization is approximately 50% in Asia, the age adjusted rate of adverse clinical

outcomes is disproportionately higher (10%) (115). There are several notable reasons for this discrepancy. One factor is geographic diversity in genetic makeup of the host (67). Individuals with alleles associated with greater production of proinflammatory cytokines or decreased production of anti-inflammatory cytokines have an increased risk of adverse disease outcomes when colonized by *H. pylori* (23, 108). Environmental conditions, such as host diet, may also contribute to adverse clinical outcomes (23). Finally, the genetic composition of *H. pylori* is a major factor (21).

H. pylori is a genetically diverse organism (11, 21). Sequence analysis of *H. pylori* strains has identified several genetic elements that are associated with worse disease outcomes (11). For example, colonization with strains of *H. pylori* that contain a genetically distinct region of the genome called the cytotoxin associated gene pathogenicity island (*cag* PAI) is highly associated with the development of gastric cancer (18, 22, 138). Another feature associated with the development of gastric cancer is the presence of genes encoding specific outer membrane proteins (OMPs) such as BabA, SabA, HopH, and type 1 HopQ (17, 54, 88, 98). The presence of the gene encoding a form of vacuolating cytotoxin called s1m1 VacA is another genetic element associated with severe disease outcomes (6).

The surface of *Helicobacter pylori*

H. pylori is a Gram negative curved bacillus. In addition to the cytoplasmic membrane present in all bacteria, Gram negative bacteria are surrounded by another membrane referred to as the outer membrane. The outer membrane acts as a semipermeable barrier that allows nutrients to pass through while keeping out possible

harmful elements such as lysozyme and other antimicrobial agents (13). In addition to being a barrier, many biologically important molecules and structures are present on the outer membrane. These include adhesins, receptors, transporters, flagella, and pili (2, 14). While the surfaces of several important pathogens have been studied in detail, the composition of the bacterial surface of *H. pylori* is still relatively undefined.

Based on sequence analysis of proteins possessing an N-terminal signal sequence and a characteristic hydrophobic C-terminal motif, *H. pylori* encodes at least 63 OMPs (3) (Table 1). The characteristic C-terminal motif consists of alternating stretches of hydrophobic amino acids likely to form transmembrane β -strands predicted to assemble a β -barrel (3, 68). The number of predicted *omp* genes is unexpectedly high considering the relative small size of the *H. pylori* genome (~1600 genes) compared to other bacteria such as *E. coli* (which also is predicted to encode approximately 60 OMPs while having a genome of ~4300 genes) (12, 132). *E. coli* expresses many copies of a few OMPs (10^5 copies of OmpA for example) while *H. pylori* expresses similar copy numbers of many different OMPs (30, 68).

Colonization with *H. pylori* can persist life-long, and expression of *H. pylori* OMPs can change over time (47, 88). Due to presence of a poly T tract preceding the transcriptional start site (TSS), the genes encoding many OMPs can undergo transcriptional variation (47, 80). Insertions and deletions within the poly T tract preceding the TSS can lead to altered transcription of *omp* genes. Another way *omp* expression may be altered in *H. pylori* is by slip-strand mispairing within polynucleotide repeats within the 5' coding region of the genes (47). A possible reason why *H. pylori* is

able to evade clearance by the host is this ability to “turn on” and “turn off” expression of possible immune targets (88).

OMPs of *Helicobacter pylori*

H. pylori attachment to epithelial cells has been shown to increase gastric inflammation and parietal cell loss *in vivo* (44). More virulent strains of *H. pylori* produce two known OMP adhesins, BabA and SabA. BabA binds host fucosylated Lewis b antigens and SabA binds host sialyl-Lewis x antigens (expressed under inflammatory conditions) (88, 93). Lewis antigens are carbohydrate structures that are produced by epithelial cells. Lewis antigens are assembled onto glycolipids and glycoproteins, and are the major determinant of host blood type (142).

H. pylori OMPs are classified into several families based on sequence analysis. *H. pylori* OMPs were initially identified by the presence of a semi-conserved carboxy-terminal domain (approximately 250 amino acids long) and were subsequently classified into two families known as Hop and Hop-related. Hop stands for “*Helicobacter* outer membrane porin” based on the observation that HopA, HopB, HopC, and HopD functioned as porins (30, 98). This group has 32 members and the semi-conserved carboxy-terminal domain is predicted to form a hydrophobic transmembrane β -barrel (3). BabA and SabA are characterized as Hop OMPs. Due to the high sequence similarity between many OMPs, it is hypothesized that these proteins fold into similar structures. HopZ, BabB, SabB, and OipA, are other Hop OMPs associated with worse disease outcomes, but unlike BabA and SabA, these OMPs lack a known host binding target (98). The distinction between the Hop and the Hop-related OMPs is that Hop

proteins encode a conserved signal sequence cleavage motif, while the Hop-related OMPs do not (3). In addition to the Hop and Hop-related OMPs, *H. pylori* is predicted to encode Hof (*H. pylori* outer membrane protein family), Hom (*H. pylori* outer membrane proteins), iron-regulated, and efflux pump OMPs (Table 1) (3). HomB is an example of an OMP associated with adverse clinical outcomes in children colonized with *H. pylori* (102, 103). The iron-regulated OMP FecA3 has been described as being essential for mouse colonization (9).

Table 1. Different families of predicted *H. pylori* OMPs (3)

Family 1: Hop			Family 2: Hof		
	Gene	Size		Gene	Size
HopZ	HP0009*	672	HofA	HP0209	450
HopD	HP0025	711	HofB	HP1083	479
HopM	HP0227	691	HofC	HP0486	528
HopA	HP0229	483	HofD	HP0487	480
HopF	HP0252	487	HofE	HP0782	455
HopG	HP0253/0254	471	HofF	HP0788	499
HopJ	HP0477	367	HofG	HP0914	514
HopH	HP0638	305	HofH	HP1167	471
HopE	HP0706	273			
SabB (HopO)	HP0722*	644	Family 3: Hom	Gene	Size
SabA (HopP)	HP0725*	653	HomA	HP0710	660
BabA (HopS)	HP1243	733	HomB	NA	668**
HopC	HP0912	515	HomC	HP0373	700
HopB	HP0913	529	HomD	HP1453	746
HopK	HP0923	369			
			Family 4: Iron-regulated	Gene	Size
HopI	HP1156	396	FecA-1	HP0686	767
HopL	HP1157	1230	FecA-2	HP0807	792
BabB (HopT)	HP0896	708	FecA-3	HP1400	841
HopQ	HP1177	641	FrpB-1	HP0876	791
HopN	HP1342	691	FrpB-2	HP0915/0916	815
HopU	HP0317	745	FrpB-3	HP1512	879
Family 1: Hop-related	Gene	Size	Family 5: Efflux pump	Gene	Size
HorA	HP0078/0079	684	HefA	HP0605	477
HorB	HP0127	286	HefD	HP0971	431
HorC	HP0324	254	HefG	HP1327	412
HorD	HP1066	200			
HorE	HP0472	186	Other	Gene	Size
HorF	HP0671	270	NlpD***	HP0506	403
HorG	HP0796	278	BamA	HP0655	916
HorH	HP1107	230		HP0694	257
HorI	HP1113	277		HP0726	305
HorJ	HP1469	248		HP0358	511
HorK	HP1501	388		HP1467	231
HorL	HP1395	242		HP0325	237
				HP0839	587
				HP1125	179
				HP1456	175
			FlgH		
			OmpP1		
			PalA		
			Lpp20		

- * Gene is out of frame due to CT repeats at 5' of gene
- ** Gene absent in strain 26695, protein size from strain J99
- *** Annotation determined via NCBI BLAST multidomain

Other known *H. pylori* surface-exposed virulence factors

In addition to a diverse repertoire of OMPs, *H. pylori* produces 2-6 unipolar sheathed flagella (Figure 1). The flagellar sheath is a continuation of the outer membrane and extends throughout the length of the flagella. HpaA is proposed to be one protein component of the flagellar sheath (83). Isogenic mutants defective in flagellar function are incapable of colonizing animal models (104). *H. pylori* uses chemotaxis to evade the low pH conditions present in the stomach (24). It is hypothesized *H. pylori* uses chemotaxis to localize itself near the gastric mucosal epithelia where the pH is closer to neutral (104). Several known adhesins, and perhaps many others, mediate bacterial binding to the surface of the host cells. Additionally, strains containing the *cag* PAI produce filamentous structures predicted to be the extracellular portion of a type 4 secretion system (T4SS) (Figure 1) (59). The only known effector protein encoded by the *cag* PAI is CagA. *H. pylori* translocates CagA into host cells in a Cag T4SS-dependent manner after the bacteria make contact with host epithelia. CagA induces many different cellular changes upon entry into host cells, resulting in various phenotypes such as actin cytoskeletal rearrangement and chromosomal instability (48, 100). A gene encoding VacA is present in virtually all strains of *H. pylori* (20). Levels of VacA expression and the form of VacA are variable among different strains. VacA is localized and secreted from the surface via a type Va secretion (autotransporter) pathway (34) (Figure 1). Proteins secreted by this pathway undergo Sec-dependent secretion across the inner membrane and are localized to the outer membrane via a carboxy-terminal transporter, or “helper” domain that interacts with the Bam complex (49, 53). The amino-terminal passenger domain is hypothesized to be threaded through the pore that forms from folding of the transporter domain (49).

Many type Va proteins are then cleaved either by a helper protease or autoproteolytic event and are either retained on the surface (as with many adhesin proteins) or are released into the extracellular space. While VacA is found in the extracellular growth supernatant, VacA is also displayed on the surface of *H. pylori*. (49). A notable feature of the surface of *H. pylori* is the observation that the outer membrane Lipid A component of LPS is characterized by low endotoxicity. The LPS O antigen is highly variable and contains Lewis antigens which mimic similar antigens found on host cells (51, 127). Finally, *H. pylori* is not known or predicted to produce a capsule or encode S-layer proteins.

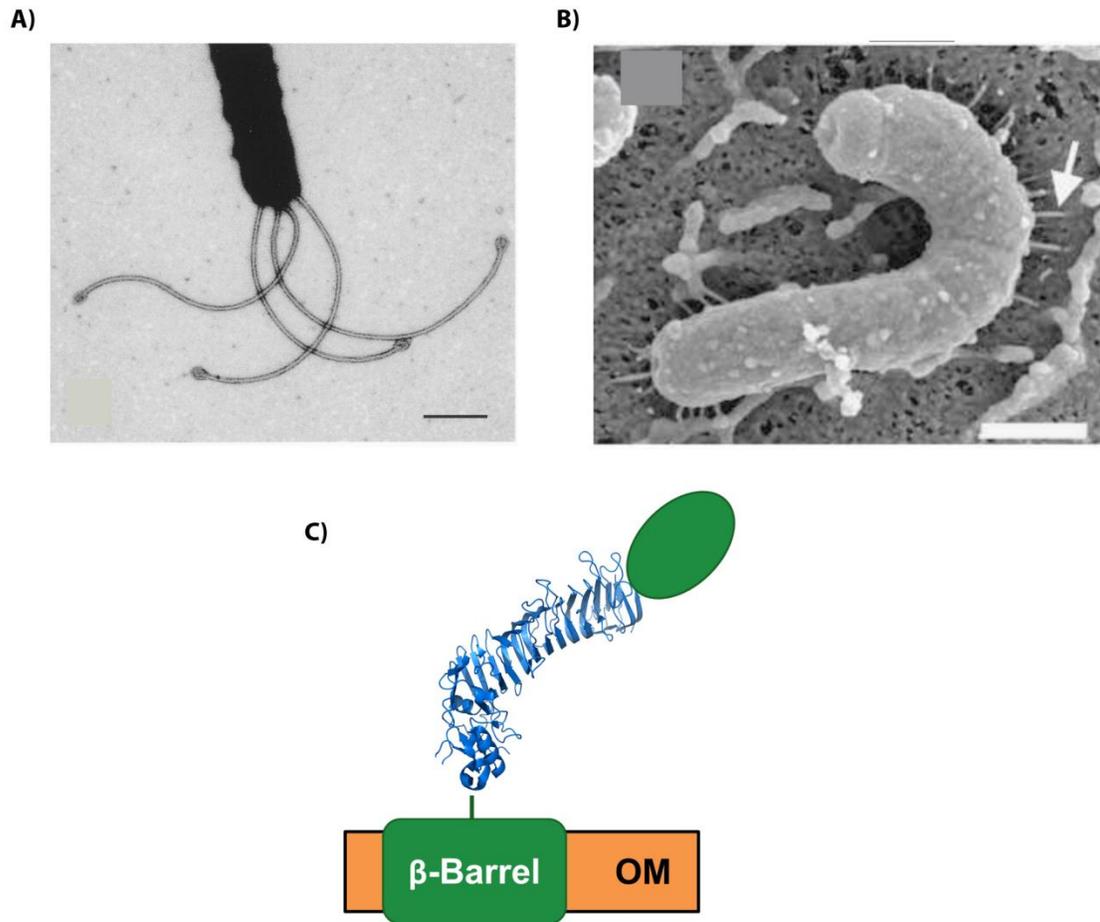


Figure 1. Surface structures of *H. pylori*. A) *H. pylori* flagella. Transmission electron microscopy of *H. pylori* strain SS1 stained with 1% phosphotungstate. Bar represents 500 nm (123). B) Cag T4SS extracellular pili. Scanning electron microscopy of *H. pylori* strain 26695 in contact with AGS epithelial cells stained with 10 nm thick gold film. Bar represents 500 nm (117). C) Cartoon of VacA on the surface of *H. pylori*. Ribbon diagram of solved VacA p55 domain crystal structure is shown in blue (42).

Research significance and specific aims

Identification of surface-exposed proteins of *H. pylori*

Bacteria must sense and adapt to changes in their environment in order to survive. The bacterial surface is the first compartment that interacts with the extracellular environment. Additionally, with bacterial pathogens such as *H. pylori*, the surface is where the bacteria make direct contact with the host, and must serve to not only adhere to surfaces and protect from the constant onslaught of the host's immune system, but is also a substantial repository of virulence determinants. *H. pylori* is a fastidious bacterium that requires rich, undefined media for growth. Additionally, it is fragile and prone to autolysis. This makes biochemical and biophysical investigation of different cellular compartments difficult to perform and interpret. Several studies have attempted to identify surface-exposed proteins of *H. pylori*. However, these typically used only a single approach and often relied on techniques not suitable for hydrophobic membrane proteins (such as OMPs and other transmembrane proteins). As a result, little is known about the composition of the surface-exposed subcellular proteome of *H. pylori*. The main goal of this thesis was to expand our knowledge of the surface-exposed protein repertoire of *H. pylori* using a combination of multiple classical approaches and novel techniques. The knowledge gained established a baseline composition of the surface-exposed proteome in order to guide future experiments into how the bacterial surface may change under various conditions simulating infection. The purpose is to gain an understanding of how *H. pylori* interacts with its environment,

how these bacteria can evade the host's immune system, and how we can develop novel therapeutic targets in order to treat or prevent *H. pylori* infection.

While *H. pylori* genetic composition (such as the presence of the *cag* PAI and genes encoding certain OMPs) is a major determinant of clinical outcomes, environmental conditions are another major determinant. Specifically, a high salt diet is associated with adverse clinical outcomes and increased inflammation *in vivo* (40, 62). Based on the observation that a high dietary salt intake results in increased gastric inflammation in *H. pylori*-infected animals, I hypothesized that many *H. pylori* proteins change in abundance when the bacteria are grown under conditions to simulate a high salt diet. These may include proteins that act to counter salt-induced osmotic stress, as well as proteins that may be contributing to adverse clinical outcomes in the host. Large-scale proteomic experiments were used to identify possible salt-responsive proteins, especially proteins present in low abundance. CagA has been previously reported as being salt-responsive (82). However, this protein is secreted into host cells through a T4SS-dependent mechanism and thus is not likely to be a source of bacterial-host surface-surface interactions. Identification of salt-responsive surface-exposed proteins would be a first step in the study of proteins contributing to adverse clinical outcomes.

Investigation of CagT, a putative VirB7 functional analog

Another important area of research focuses on the assembly and mechanism of action of the *cag* PAI-encoded T4SS (Cag T4SS). CagT is required for Cag T4SS-dependent phenotypes. The presence of a 4 amino acid motif called a lipobox within

CagT is the basis for CagT being annotated as a putative VirB7 functional analog. Other than a lipobox motif, CagT has no sequence similarity to VirB7 proteins found in other bacterial species. Within the canonical VirB/D4 T4SS found in *Agrobacterium tumefaciens*, VirB7 is a component of the VirB/D4 T4SS core complex. Based on studies of the VirB/D4 T4SS, the Cag T4SS core complex is predicted to be the scaffold to which the surface-exposed T4SS pilus is assembled upon contact with host cells. Exactly which Cag proteins are present on the surface of *H. pylori* has yet to be determined.

VirB7 in *Agrobacterium* and conjugative T4SSs functions primarily as the outer membrane anchor for the T4SS structural core (comprised of VirB7, 9 and 10) (39). In addition to the lack of sequence homology to VirB7, CagT is much larger than VirB7 (280 AA compared to 55 AA). To my knowledge, there are no previous studies that investigated potential lipidation of CagT to determine if the VirB7 annotation is appropriate. To address this, I have investigated the lipidation status and the importance of the lipobox within CagT. Any knowledge gained pertaining to CagT would potentially guide us in future investigations of how the Cag T4SS functions to deliver CagA into host cells.

CHAPTER II

IDENTIFICATION OF SURFACE-EXPOSED PROTEINS OF *HELICOBACTER PYLORI* USING COMPLEMENTARY TECHNIQUES

Introduction

Interactions between *H. pylori* and the host are dependent on specialized proteins localized on the surface of *H. pylori* (77). Many *H. pylori*-induced alterations in host cells are mediated by secreted proteins such as VacA and CagA, which are secreted by type V (autotransporter) and type IV secretion systems, respectively (20, 48). Others include *H. pylori* outer membrane protein adhesins that mediate *H. pylori* adherence to gastric epithelial cells (54, 88, 99), and surface-exposed components of the *cag* type IV secretion system (T4SS) have important roles in engaging receptors on host cells (58, 73). Some *H. pylori* outer membrane proteins (OMPs) can stimulate or inhibit inflammatory responses (121), and others can modulate the activity of the *cag* type IV secretion system (10, 56, 57).

Several approaches have been used to experimentally identify proteins localized on the surface of *H. pylori* or proteins associated with the *H. pylori* outer membrane. These have included analyses of susceptibility to protease digestion (118), immunolabeling (114), accessibility to chemical modification (119), and differential detergent solubility (8). These experimental studies have provided useful insights, but

most have been limited by the use of only a single method (thus lacking a means for validating results) or the use of methods (such as 2D gel electrophoresis) that are suboptimal for detecting membrane proteins that are present in low abundance. As an additional complication, it is difficult to separate *H. pylori* inner and outer membrane proteins using methods optimized for studies of Enterobacteriaceae (26, 30). Furthermore, *H. pylori* is prone to undergo autolysis (110), which can potentially lead to artifactual surface-exposure of proteins that have an intracellular localization in intact bacteria.

Much of our current knowledge about the outer membrane composition of *H. pylori* has been deduced from analyses of *H. pylori* genome sequences. The first analysis of an *H. pylori* genome sequence (from strain 26695) identified a family of 32 genes that were predicted to encode integral outer membrane proteins (132). These are subsequently denoted throughout the manuscript as “annotated OMPs”. A subsequent genomic analysis identified 63 *H. pylori* genes that were predicted to encode outer membrane proteins and classified these proteins into several different families (3). The largest family, designated as the “major OMP family” or Hop family, corresponded to the genes identified in the earlier genomic analysis and was divided into two subfamilies (21 Hop and 12 Hop-related or Hor proteins) (3). Among the 21 Hop proteins, at least five are reported to function as adhesins (BabA, SabA, HopZ, AlpA, and AlpB) (54, 88, 99, 107) and several are reported to have porin-like properties (25). In addition to the Hop family of OMPs, several smaller families of putative *H. pylori* OMPs have been designated, such as Hof proteins, Hom proteins, FecA-like and FrpB-like iron-regulated

proteins, and efflux pump proteins (3). The subcellular localization of many of the 63 putative OMPs has not been validated experimentally.

Most of the *in silico* efforts to identify *H. pylori* OMPs have been designed to identify OMPs with distinctive C-terminal motifs or β -barrel structures (132). Proteins exported to the surface of *H. pylori* through pathways such as the flagellar or type IV secretion systems might not possess these features, and therefore might not be successfully identified using these approaches. Thus, although the *in silico* analyses provide a valuable resource for identifying candidate proteins that are likely to be localized to the outer membrane, these analyses offer an incomplete view of the proteins that are potentially exported to the surface of *H. pylori*. Experimental studies are required in order to gain a more comprehensive understanding of the cell surface proteome, and to understand how the production of cell-surface proteins might vary under different environmental conditions.

The goals of the current study were to identify and analyze proteins that are localized on the surface of the prototype *H. pylori* reference strain 26695, to test the hypothesis that the cell-surface proteome of *H. pylori* includes additional proteins besides those that have been predicted based on *in silico* analyses, and to elucidate the cell surface topology of surface-exposed proteins. To identify surface-exposed proteins, I used multiple complementary biochemical and biophysical methods for protein separation, coupled with robust mass spectrometric methods for protein detection. I report here the identification of proteins that meet multiple criteria for surface-exposed outer membrane localization. These include numerous proteins previously known or predicted to be OMPs, as well as *cag* pathogenicity island-encoded proteins that are

required for activity of the *cag* type IV secretion system, putative lipoproteins, and additional proteins that were not previously considered to be cell-surface components. I show that the majority of Hop and Hom OMPs on the surface of intact bacteria are highly susceptible to proteolytic cleavage by an exogenous protease, whereas Hor and Hof proteins are relatively resistant. I present evidence that most of the protease-susceptible OMPs contain a large protease-susceptible extracellular domain exported beyond the outer membrane and a protease-resistant domain at the C-terminus with a predicted β -barrel structure. These features suggest that, similar to the secretion of VacA passenger domain, the N-terminal domains of protease-susceptible OMPs are exported through an autotransporter pathway. Collectively, these results provide new insights into the repertoire of surface-exposed *H. pylori* proteins that may mediate bacteria-host interactions, as well as the cell-surface topology of these proteins.

Materials and Methods

Bacterial strains and culture conditions.

Strain 26695 was isolated from a gastritis patient and fully sequenced (132). The variant of *H. pylori* strain 26695 used in this study contains an intact *cag* pathogenicity island (*cag* PAI) and a functional *cag* T4SS, but does not produce flagella (124). A Δ *cag* PAI mutant strain has been described previously (16). Bacteria were grown on either trypticase soy agar plates supplemented with 5% sheep blood or in sulfite-free Brucella broth supplemented with 10% fetal bovine serum (BB-FBS) at 37°C in room air containing 5% CO₂.

Generation of rabbit polyclonal antisera.

Anti-UreA serum was obtained from Santa Cruz. CagA, Cag3, CagM, CagT and CagX, derived from *H. pylori* 26695, were produced as recombinant proteins in *E. coli*. GST fusion proteins, MBP fusion proteins, and 6-His-tagged proteins were purified using glutathione-, maltose-, or nickel-coated beads, respectively. Rabbits were then immunized with the purified proteins. Anti-CagT serum was adsorbed with boiled Δ *cagT* mutant bacteria prior to use in immunogold EM experiments.

Immunoblot analysis.

Proteins were separated by SDS-PAGE and transferred to nitrocellulose membranes. Proteins were detected by incubating the membrane with primary antisera (diluted 1:5,000-1:10,000), followed by horseradish peroxidase-conjugated anti-rabbit

IgG as the secondary antibody (Santa Cruz). Labeled proteins were detected by enhanced chemiluminescence (ECL) methodology and visualized using X-ray film.

Immunogold transmission electron microscopy.

H. pylori was grown on trypticase soy agar plates for 24 hours. Broth cultures were inoculated at an initial density of OD₆₀₀ of approximately 0.05 and grown for 20 hours. Bacteria were harvested in late log phase, corresponding to an OD₆₀₀ of approximately 0.8, by centrifugation at 3,500 x g for 20 minutes at 4° C. Surface proteins were labeled using procedures similar to those described previously (124). Briefly, cells were washed once in PBS containing 1 mM CaCl₂ and 0.5 mM MgCl₂. Bacteria were then fixed with 2% paraformaldehyde (Electron Microscopy Sciences) and 2.5% glutaraldehyde (Electron Microscopy Sciences) in 50 mM sodium cacodylate (Electron Microscopy Sciences) for 30 min at room temperature. Bacteria were then pelleted at 5,000 x g for 10 minutes. The fixation was quenched with three washes with 0.1% glycine in 50 mM sodium cacodylate. Bacteria were then blocked with 0.1% cold water fish skin gelatin (Electron Microscopy Sciences) for 1 hour at 4° C. Bacteria were labeled with CagT antiserum for 1 hour at 4°C. Bacteria were washed twice with blocking buffer. Goat anti-rabbit antibody conjugated to 10 nm gold particles (Electron Microscopy Sciences) was applied in blocking buffer for 1 hour at room temperature. Samples were washed twice with 50 mM sodium cacodylate. Cells were then placed on formvar carbon-coated 300 mesh copper grids (Electron Microscopy Sciences) and negatively stained with 0.1% ammonium molybdate (Electron Microscopy Sciences). Samples were imaged with a FEI T-12 transmission electron microscope.

Biotinylation of surface-exposed *H. pylori* proteins using an amine-reactive biotinylation reagent.

H. pylori was grown on trypticase soy agar plates for 24 hours. Broth cultures were inoculated at an initial density of OD₆₀₀ of approximately 0.05 and grown for 20 hours. Bacteria were harvested in late log phase, corresponding to an OD₆₀₀ of approximately 0.8, by centrifugation at 3,500 x g for 20 minutes at 4° C. Surface proteins were labeled using procedures similar to those in a previous publication (119), with several modifications. These included culturing the bacteria in broth instead of plates, use of magnetic beads coated with monomeric avidin (Bioclone, BcMag beads) for purification instead of avidin-agarose, and trichloroacetic acid (TCA) precipitation instead of acetone precipitation. Briefly, the bacterial pellet was washed once, pelleted, and resuspended in labeling buffer [PBS pH 7.4 containing 1 mM CaCl₂, 0.5 mM MgCl₂, and 1.6 mM D-biotin (Sigma-Aldrich)]. Sulfo-NHS-LC-biotin (Pierce) was added (200 μM final concentration) to the bacterial suspension and incubated for 30 minutes on ice. The reaction was quenched by the addition of two volumes TNKCM (50 mM Tris, pH 7.4, 100 mM NaCl, 27 mM KCl, 1 mM CaCl₂, 0.5 mM MgCl₂) and incubated at room temperature for 10 minutes. Bacteria were pelleted, washed in TNKCM three times, and resuspended in detergent-free lysis buffer [50 mM Tris, pH 7.4, 1 mM MgCl₂, with EDTA-free protease inhibitor cocktail (Roche)]. Bacteria were lysed by sonication, and intact unbroken bacteria were pelleted at 7,000 x g for 10 minutes. The supernatant was centrifuged at 40,000 x g for 30 minutes at 4°C. The supernatant was discarded and the pellet (corresponding to a membrane fraction) was washed three times in TKE (50 mM Tris pH 7.4, 150 mM KCl, 10 mM EDTA). The pellet was solubilized in TKEZ [50 mM

Tris pH 7.4, 150 mM KCl, 10 mM EDTA, 2% Zwittergent 3-14 (Fluka), with EDTA-free protease inhibitor cocktail] for 1 hour at 4°C. This preparation was centrifuged at 100,000 x g for 1 hour at 4°C and the supernatant (predicted to be enriched in outer membrane proteins) was processed further as described below, to allow purification of biotinylated proteins. A previous study demonstrated that these methods result in minimal labeling of a periplasmic protein (119). A control sample was processed in parallel using exactly the same procedures, except that Sulfo-NHS-LC-biotin was omitted from the labeling reaction.

Biotinylation of surface-exposed *H. pylori* proteins using a carboxyl-reactive biotinylation reagent.

H. pylori was cultured as described for the biotinylation experiments above. The bacterial pellet was washed in MES buffer (100 mM MES pH 6.0, 500 mM NaCl, 5 mM biotin), pelleted at 3,500 x g for 10 minutes and resuspended in MES buffer containing 2 mM 1-Ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride (EDC, Pierce) and 5 mM *N*-hydroxysulfosuccinimide (S-NHS, Pierce). The reaction was incubated for 15 minutes at 4°C. The pH was raised from 6.0 to 7.3 with 1 M sodium carbonate. Amine-PEG3-biotin was added to the reaction (24 mM final concentration) and incubated for 2 hours at 4°C. Two volumes of TNKCM (50 mM Tris, pH 7.4, 100 mM NaCl, 27 mM KCl, 1 mM CaCl₂, 0.5 mM MgCl₂) were added to the reaction and incubated at room temperature for 10 minutes. Bacteria were pelleted, washed in TNKCM three times, and resuspended in detergent-free lysis buffer [50 mM Tris, pH 7.4, 1 mM MgCl₂, with EDTA-free protease inhibitor cocktail (Roche)]. Bacteria were lysed by sonication, and intact unbroken bacteria were pelleted at 7,000 x g for 10 minutes. The supernatant was

centrifuged at 40,000 x g for 30 minutes at 4°C. The supernatant was removed and the total membrane pellet was washed three times in TKE (50 mM Tris pH 7.4, 150 mM KCl, 10 mM EDTA). Outer membrane proteins were solubilized in TKEZ (50 mM Tris pH 7.4, 150 mM KCl, 10 mM EDTA, 2% Zwittergent 3-14 (Fluka), with EDTA-free protease inhibitor cocktail) for 1 hour at 4°C (72). Insoluble proteins were pelleted by centrifugation at 100,000 x g for 1 hour at 4°C and the supernatant (containing outer membrane proteins) was processed further as described below. A control sample was processed in parallel using exactly the same procedures, except that EDC, S-NHS, and amine-PEG3-Biotin were omitted from the labeling reaction.

Purification of biotinylated proteins.

To purify biotinylated proteins from the preparations described above, the samples were diluted 10 times in phosphate buffer (100 mM NaPO₄ pH 7.4, 150 mM NaCl). Monomeric avidin magnetic beads (Bioclone) were prepared according to the manufacturer's instructions and equilibrated in washing buffer (100 mM NaPO₄ pH 7.4, 150 mM NaCl, 0.2% Zwittergent 3-14). The sample was incubated with equilibrated beads at room temperature, and bound proteins were eluted from the beads with 100 mM NaPO₄ pH 7.4, 150 mM NaCl, 0.2% Zwittergent 3-14, 5 mM D-Biotin (Sigma) for 30 minutes at room temperature. Eluted proteins were precipitated with a final concentration of 25% TCA.

Bacterial subcellular fractionation.

H. pylori was grown on trypticase soy agar plates for 24 hours and then grown for 20 hours in Brucella broth to late log phase, corresponding to an OD₆₀₀ of approximately

0.8. Bacteria were pelleted by centrifugation at 3,500 x g for 20 min at 4°C. The bacterial cells were then fractionated as described previously (111). In brief, the bacterial pellet was washed in 10 mM Tris pH 8.0 and resuspended in 10 mM Tris pH 8.0 with EDTA-free protease inhibitor cocktail. Bacteria were lysed by sonication. Intact bacteria were pelleted at 7,000 x g for 10 minutes. A total membrane fraction was collected by centrifuging the supernatant at 230,000 x g for 1 hour at 4°C. Membrane proteins were solubilized with Triton X-100 buffer (10 mM Tris pH 8.0, 1% Triton X-100, EDTA-free protease inhibitor cocktail) for 30 minutes at 4°C. This preparation was centrifuged at 230,000 x g for 45 minutes at 4°C, yielding a soluble fraction (which is predicted to be enriched in inner membrane proteins) and an insoluble fraction (which is predicted to be enriched in outer membrane proteins).

Proteinase K treatment of intact bacteria.

H. pylori was grown on trypticase soy agar plates for 24 hours and then grown for 20 hours in BB-FBS broth to late log phase, corresponding to an OD₆₀₀ of approximately 0.8. Bacteria were pelleted by centrifugation at 3,500 x g for 20 min at 4°C. Bacteria were then washed with PBS and resuspended in Tris buffer (50 mM Tris pH 7.4, 150 mM NaCl, 1 mM CaCl₂, 1 mM MgCl₂, 1 mM KCl). Proteinase K was added to bacterial suspensions (final concentrations of either 100 µg mL⁻¹ at 22°C or 1 mg mL⁻¹ at 4°C) and incubated for 30 min (111, 114, 124). PMSF was added to a final concentration of 2 mM. Bacteria were pelleted and resuspended in Tris buffer with protease inhibitor cocktail (Roche). Bacteria were then lysed by sonication and intact cells were removed by centrifugation at 5,000 x g for 10 min. Total membranes were collected by centrifugation at 200,000 x g for 30 min at 4°C. Membrane pellets were washed in Tris

buffer and solubilized in RIPA buffer (50 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA, 1% NP-40, 0.25% sodium deoxycholate) containing protease inhibitor cocktail (111, 114, 124).

Mass spectrometric analysis of samples.

To provide a comprehensive analysis of the protein content in samples, the samples were analyzed by multidimensional protein identification technology (MudPIT) (19, 78). 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 proteins were digested with trypsin in solution. MudPITs were performed essentially as described previously (87), using a ThermoFisher LTQ equipped with a nano-electrospray source and attached to an Eksigent 1D+ or Nanoacuity (Waters) HPLC unit with an autosampler. Peptide MS/MS spectra were acquired data-dependently with one full scan MS followed by 5 MS/MS scans. Acidified peptides were loaded onto a 150 micron ID biphasic trapping column comprised of 4 cm strong cation exchange resin (LUNA 5 micron, Phenomenex) 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 biotinylated proteins and respective unlabeled controls were analyzed by 8-step MudPIT. Preparations from the proteinase K experiment and cell fractions were analyzed by 11-step MudPIT. Salt pulses were performed by using the autosampler to inject 5 μ l of ammonium acetate at 0, 100, 150, 200, 300, 500, 750, and 1000 mM concentrations for the 8-step analyses used in analyzing the biotinylated preparations, and 0, 50, 75, 100, 150, 200, 250, 300,

500, 750, and 1000 mM for the 11-step analyses. 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) and Myrimatch (semi-tryptic specificity) against an *H. pylori* strain 26695 database, to which both common contaminants and reversed versions of the proteins had been appended. For analyses of mass spectrometry data from protease susceptibility experiments, we restricted the analysis to peptides that could be uniquely matched to a single protein. For analyses of mass spectrometry results from all other experiments, we analyzed all peptides detected, including those that could be matched to more than one *H. pylori* protein. Resulting identifications were filtered to an estimated peptide and protein false discovery rate (FDR) less than 5% and collated by protein using IDPicker 3.0. All subsequent analyses were performed in Excel and R. All reported proteins were identified based on detecting a minimum of 2 distinct peptides.

Analysis of mass spectrometry data.

All experiments were done in triplicate and the resulting spectral data were merged prior to analysis (135). Proteins were identified as enriched in one set of preparations compared to another, based on both analysis of fold enrichment [comparing normalized ratios of numbers of assigned spectra, calculated using either QuasiTel model-generated rates or R_{sc} values, which both account for abundance of spectral counts when analyzing normalized data] and the statistical significance of differences in abundance of spectral counts assigned to each protein (using Fisher's exact test with a Bonferonni correction) (76, 101). Criteria for identifying proteins that

were enriched in one set of samples compared to another were chosen based on comparative analysis of data for spectral proteins previously annotated as outer membrane proteins (likely to be surface-exposed) and data for proteins annotated either as ribosomal proteins, annotated inner membrane proteins, or other non-outer membrane proteins based on annotation (not likely to be surface-exposed) (Figure 2), (135). The application of the selected cutoffs resulted in a protein false discovery rate of <10% (Figure 2). I then applied these same cutoff criteria in analysis of spectral data corresponding to all proteins identified in the respective preparations (135).

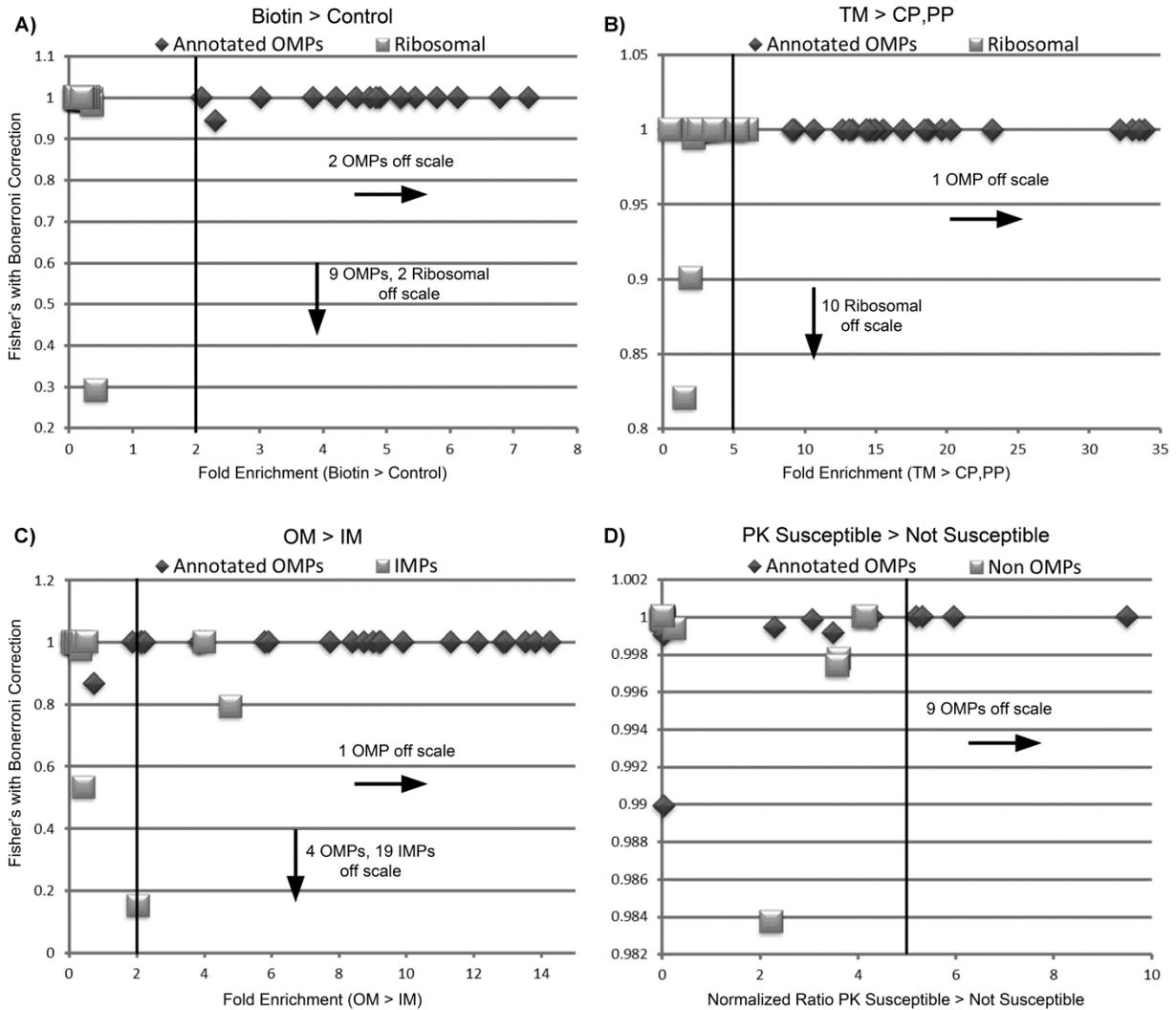


Figure 2. Criteria for identifying enriched proteins. To identify proteins enriched in each preparation of interest compared to a control or comparison preparation, I calculated the fold enrichment for each protein (based on analysis of spectral counts, using QuasiTel) and also calculated the statistical significance of differences in numbers of spectral counts assigned to each protein in the two preparations (using Fisher's exact test with a Bonferonni correction). The x axes indicate fold enrichment and the y axes indicate values corresponding to $(1 - p \text{ value})$. I then analyzed data for proteins annotated as outer membrane proteins compared to data for proteins predicted to have a non-surface-exposed localization. These analyses allowed us to choose appropriate criteria for identifying proteins that were enriched in one sample compared to another sample. False discovery rates were calculated by comparing a positive control group of proteins (annotated outer membrane proteins for panels A and B, or an expanded list of

putative outer membrane proteins selected based on previous studies and protein annotation for panel C) to a negative control group of proteins (27 ribosomal proteins for Biotin>Control and TM > CP, PP; or a set of predicted inner membrane proteins (IMP) that were selected based on previous studies and protein annotation)(135). For each analysis, I defined a p value of < 0.05 with Bonferonni correction as significant, and arbitrarily selected cutoffs for fold-enrichment to allow the identification of a maximum number of positive control proteins while limiting the false discovery rate to < 10%. Proteins were considered enriched if the numbers of assigned spectra were significantly higher in the preparation of interest compared to control preparations, and if there was a minimum fold enrichment compared to the respective counterpart. (A) Biotin compared to unlabeled control (Biotin > Control), (B) membrane proteins compared to soluble proteins (TM > CP, PP), (C) Triton X-100 insoluble proteins compared to Triton X-100 soluble proteins (OM > IM). (D) Proteinase K-treated bacteria compared to non-treated bacteria (PK susceptible > PK non-susceptible). Vertical line indicates the minimum levels of enrichment selected as cutoffs: Biotin > Control = 2, TM > CP, PP = 5.5, OM > IM = 2, PK susceptible > PK non-susceptible = 5.

Mapping of protein segments that are susceptible to proteinase K digestion.

To systematically identify sites within individual proteins that were susceptible to proteinase K digestion, I analyzed the spectral data from the proteinase K experiment described above (conducted at 4°C). Spectral coverage data from IDPicker 3.0 files were uploaded into R. For each protein to which peptides had been matched, I then quantified the spectral coverage at individual amino acid positions. I analyzed all amino acid sites for which ≥ 5 assigned spectra were detected in the untreated control sample (designated as 'residues available for analysis'), and restricted the analysis to proteins with >100 residues available for analysis. Residues with a ≥ 5 -fold greater number of assigned spectra in the untreated control samples compared to the protease-treated samples were designated as protease-susceptible sites. Residues with a < 2 -fold greater number of assigned spectra in the untreated control samples compared to the protease-treated samples were designated as protease-non-susceptible sites. Criteria for the designation of protease-susceptible proteins were chosen based on a comparative analysis of annotated OMPs and proteins predicted to have an inner membrane, cytoplasmic, or periplasmic localization (indicated as "non-OMPs) (Figure 2) (135). Proteins with a normalized ratio of susceptible residues to non-susceptible residues of ≥ 5 and exhibiting a statistically significant increased ratio of susceptible residues to non-susceptible residues (determined by Fisher's exact test with Bonferroni correction) were considered susceptible to protease treatment (Figure 2) (135).

Results

Identification of surface-exposed proteins.

To identify proteins localized on the surface of *H. pylori*, I used a cell-impermeable amine-reactive biotin conjugation reagent to label surface-exposed proteins on intact bacterial cells. SDS-PAGE and silver staining of proteins purified from the biotinylated bacteria revealed the presence of numerous proteins (Figure 3A). In parallel, I generated a mock preparation of proteins purified from non-biotinylated bacteria. This unlabeled control preparation contained a relatively low abundance of proteins compared to the sample of purified biotinylated proteins (Figure 3A). As a first step in analyzing and comparing these preparations, I immunoblotted the two samples with several available antisera. CagA and UreA were equally abundant in both preparations, whereas Cag3, CagM, CagT and CagX were more abundant in the biotinylated preparations (Figure 3B). These results are consistent with an enrichment of specific proteins in the preparation of biotinylated proteins compared to the unlabeled control sample.

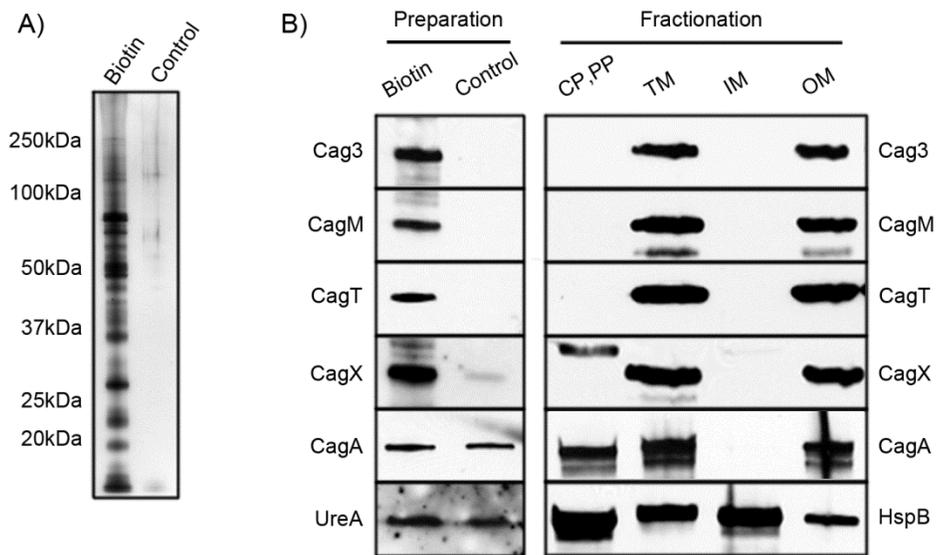


Figure 3. Analysis of purified biotinylated proteins. (A) Biotinylated proteins were purified from biotinylated bacteria, and a control preparation was generated from non-biotinylated bacteria. Equal volumes of each preparation were analyzed by silver staining. Specific bands were visible in the preparation of biotinylated proteins but not in the control preparation. **(B)** The preparations from panel A, as well as subcellular fractions, were immunoblotted with antisera to five proteins encoded by the *cag* PAI, UreA, and HspB. Equal volumes of biotinylated and unlabeled control preparations were loaded into each lane, and standardized amounts of subcellular fractions (25 ug of total protein) were loaded into each lane. CP,PP – soluble proteins predicted to have a cytoplasmic or periplasmic localization; TM, insoluble proteins corresponding to a total membrane preparation; IM, Triton-soluble membrane proteins, predicted to have an inner membrane localization; OM, Triton-insoluble membrane proteins, predicted to have an outer membrane localization.

I generated three independent preparations of the biotinylated and unlabeled control samples, and analyzed the protein contents of these samples by multi-dimensional protein identification technology (MudPIT). Consistent with the results of SDS-PAGE and silver staining, the total number of spectra confidently matched to *H. pylori* proteins (spectral counts) was about 5 times higher in analyses of the biotinylated preparations than in analyses of the control preparations (Table 2) (135). In total, I identified >500 *H. pylori* proteins in these samples (135). The detection of numerous proteins in the control preparations despite very few bands being visible on the silver-stained SDS-PAGE gel is attributed to the high sensitivity of the mass spectrometric detection methods. I employed a statistical approach (based on analysis of spectral counts) to compare the biotinylated preparations with the unlabeled control preparations, and identified 85 proteins that were significantly enriched in the biotinylated preparations compared to the control preparations (Table 2, Figure 2).

Table 2. Summary of mass spectrometry data

Sample	Number of assigned spectra ^l	Number of proteins detected ^m
Biotin^a	82,844	1,055
Control^b	16,009	941
Biotin>Control^c	NA	85
TM^d	90,604	1,120
CP,PP^e	78,773	1,066
TM>CP,PP^f	NA	192
OM^g	64,478	1,095
IM^h	74,320	1,114
OM>IMⁱ	NA	156
Untreated 4°C^j	92,644	1,063
Untreated RT^j	75,482	1,056
PK 4°C^j	100,521	1,064
PK RT^j	103,021	1,063
Untreated>PK 4°C^k	NA	7
Untreated>PK RT^k	NA	6

^abiotinylated preparations

^bunlabeled control preparations

^csubset of proteins that were significantly enriched in biotinylated preparations compared to unlabeled control preparations

^dinsoluble proteins (in the absence of detergent), predicted to localize to the total membrane (TM) compartment

^esoluble proteins (in the absence of detergent), predicted to localize to the cytoplasmic and periplasmic compartments (CP, PP)

^fsubset of proteins that were significantly enriched in the insoluble preparations (TM) compared to soluble preparations (CP, PP)

^gTriton X-100-insoluble proteins, predicted to localize to the outer membrane (OM) compartment

^hTriton X-100-soluble proteins, predicted to localize to the inner membrane (IM) compartment

ⁱ subset of proteins that were significantly enriched in the Triton X-100-insoluble preparations compared to the Triton X-100-soluble preparations

^j Intact bacteria were incubated in buffer or treated with proteinase K (PK) at 4°C or room temperature (RT).

^k subset of proteins for which assigned peptides were detected at significantly higher levels in untreated preparations than in proteinase K-treated preparations.

^l Results represent the sum of assigned ms/ms spectra, based on analysis of three combined biological replicates.

^m Results represent the total number of identified proteins, based on analysis of three combined biological replicates.

NA, not applicable.

I also used biophysical methods independent of biotin labeling to identify proteins associated with the *H. pylori* outer membrane. This approach was based on the reasoning that most surface-exposed proteins are likely to be associated with the bacterial outer membrane. As a first step, I used detergent-free conditions to generate preparations of soluble *H. pylori* proteins (predicted to contain cytoplasmic and periplasmic compartments) and insoluble *H. pylori* proteins (predicted to be derived from both inner and outer membranes, subsequently referred to as the total membrane fraction). Then I fractionated the total membrane preparation based on differential protein solubility in Triton X-100, which allows separation of inner membrane proteins and outer membrane proteins (111). Immunoblotting of these fractions with the same antisera used the analysis of the biotinylated samples in Figure 3B revealed that four Cag proteins (Cag3, CagM, CagT and CagX) were more abundant in the total membrane fraction than in the cytoplasmic/periplasmic fraction, and also more abundant in the outer membrane fraction than in the inner membrane fraction (Figure 3B). I generated three independent preparations of each fraction, and the protein composition of these samples was analyzed by MudPIT. I identified 191 proteins that were significantly enriched in the total membrane preparations compared to the soluble protein preparations (Table 2 and Figure 2) (135), and I identified 155 proteins that were significantly enriched in the Triton-insoluble outer membrane fractions compared to the Triton-soluble inner membrane fractions (Table 2, and Figure 2) (135).

To evaluate the efficacy of the procedures described above for enrichment of putative outer membrane proteins, I analyzed the presence in each preparation of

proteins previously predicted by *Tomb et al.* to be outer membrane proteins (“annotated OMPs”) (132) and subsequently classified as Hop or Hop-related proteins (3). There was a significant enrichment of proteins with an outer membrane protein annotation in preparations of biotinylated proteins compared to preparations of control (non-biotinylated) proteins ($p < 0.0001$) (Figure 4). Similarly, there was an enrichment in annotated OMPs when comparing the total membrane preparations with the soluble protein preparations (cytoplasmic/periplasmic proteins) ($p < 0.0001$), and when comparing the outer membrane preparations with the inner membrane preparations ($p < 0.0001$) (Figure 4). These analyses indicated that, as expected, each of the experimental approaches resulted in an enrichment of annotated outer membrane proteins.

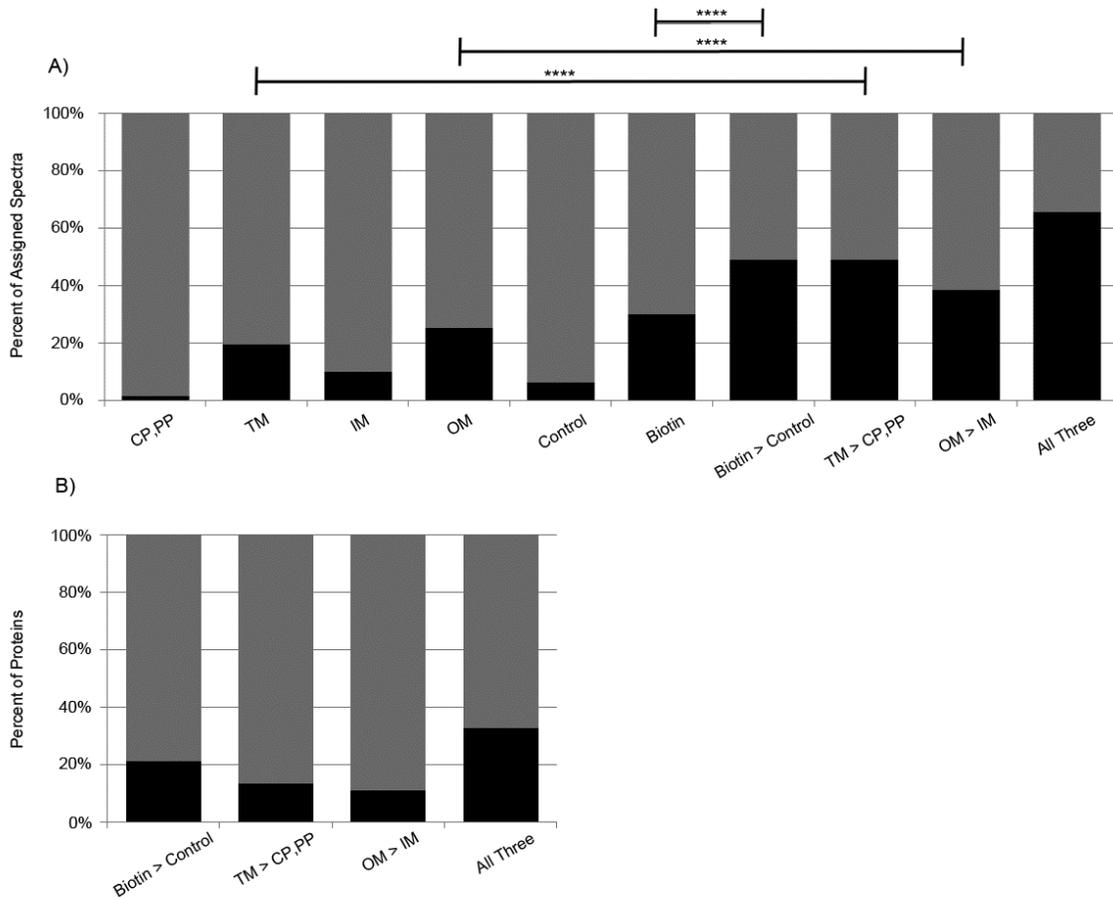


Figure 4. Enrichment of proteins with an “outer membrane protein” annotation. The proportion of spectra and proportion of proteins annotated as outer membrane proteins were analyzed for each set of preparations. **(A)** Black bars indicate the proportion of spectral counts in each preparation that were attributed to proteins previously annotated as Hop or Hor outer membrane proteins. **(B)** Black bars indicate the proportion of proteins identified in each preparation that were previously annotated as Hop or Hor outer membrane proteins. CP, PP: Soluble proteins, predicted to be localized to cytoplasmic and periplasmic compartments; TM: insoluble proteins predicted to be localized to membranes; IM: Triton X-100-soluble proteins, predicted to localize to the inner membrane; OM: Triton X-100-insoluble proteins, predicted to localize to the outer membrane; Control: unlabeled control used in biotinylation experiment; Biotin: preparation of purified biotinylated proteins. Biotin > Control: proteins identified as enriched in the biotinylation preparation compared to control. TM > CP, PP: proteins enriched in the TM preparation compared to the CP, PP preparation. OM > IM: proteins enriched in the OM fraction compared to the IM fraction. All Three: proteins

meeting three criteria consistent with outer membrane localization (enriched in the biotinylated, TM and OM preparations, compared to respective controls). All data represent the combined results from three independent experiments. ****, $p < 0.001$ for the indicated comparisons (Fisher's Exact test). Values for the set of proteins designated "All Three" in section (A) were significantly higher than values for all other protein sets ($p < 0.001$).

Proteins meeting multiple criteria for surface-exposed outer membrane localization.

I next sought to identify proteins that met multiple criteria for surface-exposed outer membrane localization, thereby providing stronger evidence compared to the use of only a single method. Specifically, I sought to identify proteins that met three criteria: (i) significantly enriched in the biotinylated preparations compared to the control (non-biotinylated) preparations, (ii) significantly enriched in the total membrane preparation compared to the soluble fraction, and (iii) significantly enriched in the Triton-insoluble outer membrane fraction compared to the Triton-soluble inner membrane fraction. Thirty nine proteins (localized in the center of the Venn diagram shown in Figure 5A) met all three of these criteria (Table 3). In an analysis of the detected peptides corresponding to this set of 39 proteins, 66% of the spectra were assigned to proteins previously annotated by *Tomb et al.* as outer membrane proteins (132) (Figure 5B), which is significantly higher than the corresponding values for any of the individual preparations (Figure 4). Thirteen of these 39 proteins were annotated by *Tomb et al.* as predicted OMPs (corresponding to 12 Hop and 1 Hop-related or Hor protein) (132), and 20 were annotated by *Alm et al.* as predicted OMPs [including three annotated as iron(III) dicitrate transport (FecA) proteins, an iron-regulated FrpB protein and the VacA-like protein FaaA (HP0609/0610)] (3, 114). The set of 39 proteins also includes numerous proteins not previously annotated as OMPs. These include three Cag proteins (Cag3, CagM, and CagT), each of which is required for the function of the *cag* T4SS that translocates the effector protein CagA into gastric epithelial cells (*cag* T4SS) (35), several annotated as lipoproteins (including Lpp20 and HP1571) (66), two homologs of the flagellar sheath adhesin HpaA (HP0410 and HP0492) (27, 119), a

protein annotated as “protective surface antigen D15” which corresponds to an outer membrane assembly protein BamA ortholog (HP0655), a secreted serine protease (HtrA, HP1019)(15, 50), a protein annotated as a conserved secreted protein or peptidyl prolyl cis-trans isomerase (HP0977), a plasminogen-binding protein (PgbB, HP0863) (61), and several proteins identified annotated as “hypothetical proteins” (Table 3, Figure 5C).

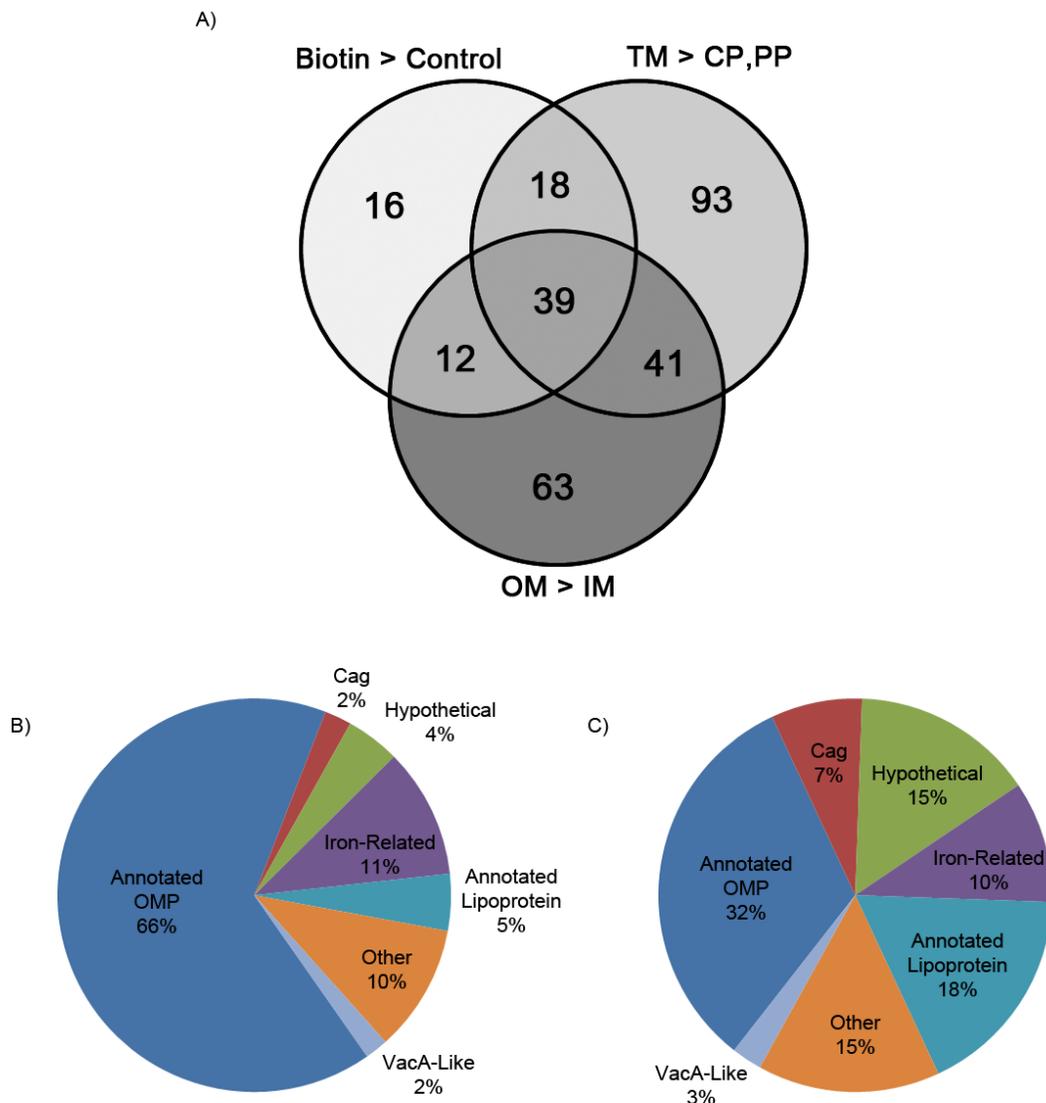


Figure 5. Use of multiple criteria to identify surface-exposed outer membrane proteins. **A)** Venn diagram of proteins identified as enriched in the biotinylated preparation compared to the control preparation (Biotin > Control), enriched in the total membrane preparation compared to the soluble fraction (TM > CP, PP), and enriched in a Triton-insoluble preparation compared to a Triton-soluble preparation (OM > IM). Thirty nine proteins met all three criteria. **B-C)** Characteristics of the 39 proteins in the center segment of (A), based on proportion of spectral counts (B) or proportion of annotated proteins in each class (C). Annotated OMPs represent Hop and Hor OMPs.

Table 3: Proteins meeting multiple criteria for surface-exposed outer membrane proteins^a

Original annotation	Alternate names	Gene number
outer membrane protein (Omp2)	HopD	HP0025
hypothetical protein		HP0036
hypothetical protein		HP0120
outer membrane protein (Omp4)	HorB	HP0127
outer membrane protein (Omp7)	HopF/HopX	HP0252
hypothetical protein		HP0358
putative neuraminylactose-binding hemagglutinin homolog (HpaA)		HP0410
outer membrane protein (Omp12)	HopJ	HP0477
lipoprotein, putative	HpaA	HP0492
cag pathogenicity island protein (Cag3)	Cag3	HP0522
cag pathogenicity island protein (Cag12)	CagT	HP0532
cag pathogenicity island protein (Cag16)	CagM	HP0537
penicillin-binding protein 1A (PBP-1A)		HP0597
membrane fusion protein (MtrC)	HefB	HP0606
hypothetical protein ^b	FaaA	HP0609 ^b
toxin-like outer membrane protein ^b	FaaA	HP0610 ^b
outer membrane protein (Omp13)	HopH/OipA	HP0638
protective surface antigen D15	BamA	HP0655
iron(III) dicitrate transport protein (FecA)	FecA-1	HP0686
lipoprotein, putative		HP0746
iron(III) dicitrate transport protein (FecA)	FecA-2	HP0807
lipoprotein, putative		HP0838
lipoprotein, putative	PgbB	HP0863
outer membrane protein (Omp19)	BabB/HopT	HP0896
outer membrane protein (Omp20)	HopC/AlpA	HP0912
outer membrane protein (Omp21)	HopB/AlpB	HP0913
outer membrane protein (Omp22)	HopK	HP0923
conserved hypothetical secreted protein		HP0977
lipoprotein, putative		HP1002
serine protease (HtrA)		HP1019
outer membrane protein (Omp25)	HopI	HP1156
outer membrane protein (Omp26)	HopL	HP1157
hypothetical protein		HP1173
outer membrane protein (Omp27)	HopQ	HP1177
outer membrane protein (Omp28)	BabA/HopS	HP1243
paralysed flagella protein (PflA)		HP1274
iron(III) dicitrate transport protein (FecA)	FecA-3	HP1400
membrane-associated lipoprotein (Lpp20)		HP1456
iron-regulated outer membrane protein (FrpB)	FrpB-3	HP1512
rare lipoprotein A (RlpA)		HP1571

^a These proteins correspond to 39 that are depicted in the center of the Venn diagram shown in Figure 5A.

^b HP0609 and HP0610 were originally annotated as two separate proteins, but correspond to a single protein (VacA-like protein FaaA).

I also analyzed the proteins that met only two criteria for outer membrane localization, corresponding to relevant segments of the Venn diagram in Figure 5A. One subset of 18 proteins (enriched in the biotinylation preparations and the total membrane preparations, but not in the Triton-insoluble membrane preparations) included six annotated OMPs (BabC or HopU, HP0317; HomA, HP0710; HopM, HP0227; HopA, HP0229; HorF, HP0671; HopN, HP1342) and HpaA (a flagellar sheath adhesin, HP0797) (60). A high proportion of the proteins in the other segments of the Venn diagram (Figure 5A) corresponded to proteins predicted to have an intracellular localization, and these probably represent false positive results.

Collectively, these experiments detected numerous proteins on the surface of *H. pylori* that had previously been predicted to have outer membrane localization (“annotated OMPs”), as well as additional proteins that were not previously predicted to be localized on the *H. pylori* surface. The evidence for surface localization is strongest for the set of proteins meeting multiple criteria for surface-exposed outer membrane localization, depicted in the center of the Venn diagram in Figure 5A and Table 3.

Immunogold EM analysis of CagT.

To provide further evidence for the surface-exposed localization of proteins not previously annotated as OMPs, I selected one of the proteins that met multiple criteria for outer membrane localization (CagT) and undertook further analysis of this protein by immunogold electron microscopy. In immunogold labeling experiments with an anti-CagT antiserum, I detected a significantly greater number of gold particles on the

surface of wild-type bacteria than on the surface of Δcag PAI mutant bacteria, which fail to produce CagT ($p < 0.0001$) (Figure 6). In images of the wild-type bacteria, multiple gold particles were often clustered together, which suggests that CagT is non-randomly distributed on the bacterial surface. These immunogold labeling studies corroborate the surface localization of CagT, and provide further validation of the experimental approaches used in this study for the identification of surface-exposed proteins.

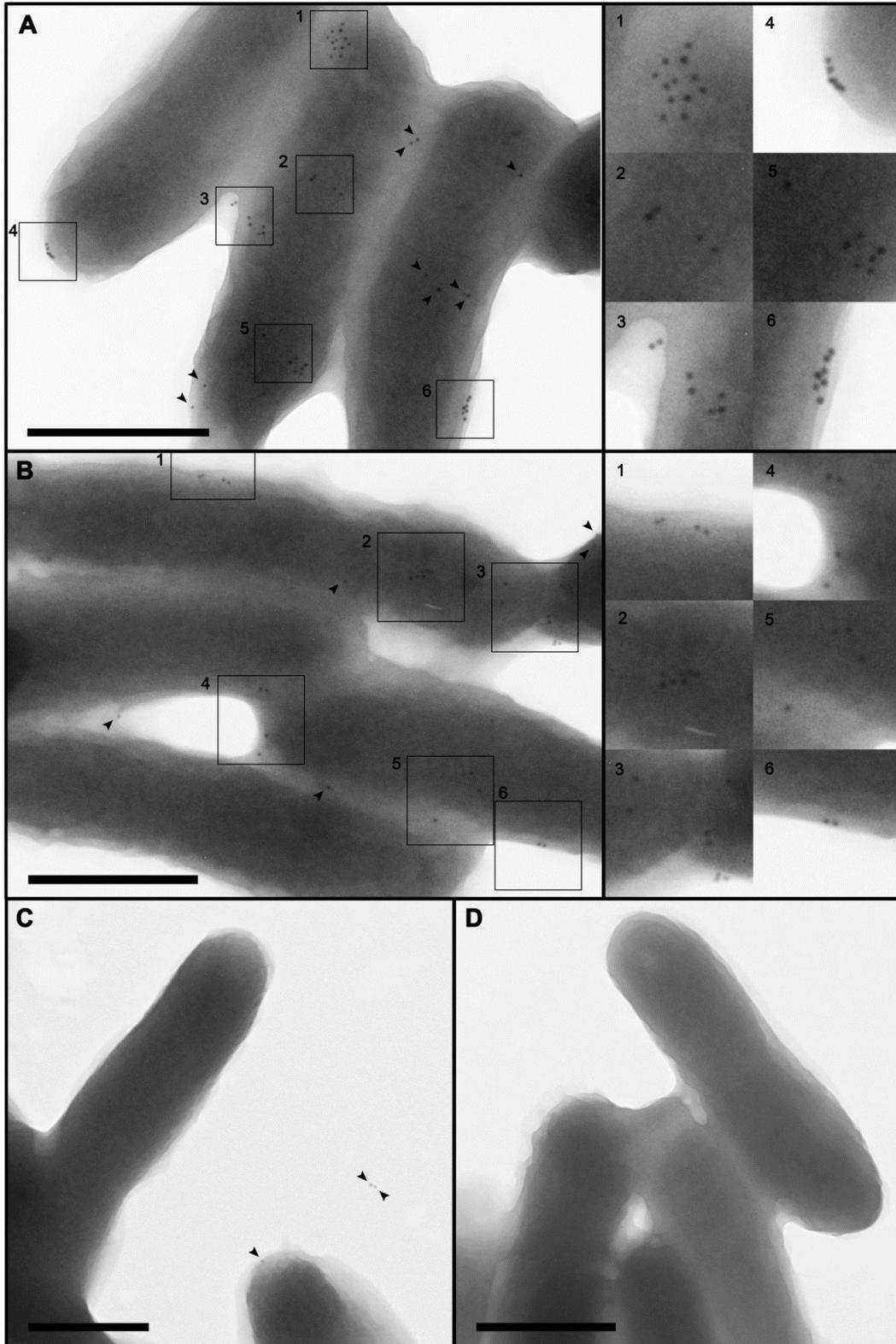


Figure 6. Localization of CagT by immunogold EM analysis. (A-C) *H. pylori* strains were immunolabeled with rabbit antiserum to CagT, followed by secondary antibodies

conjugated to 10 nm gold particles. Panels A and B, wild-type strain. Panel C, Δcag PAI mutant strain (Δ PAI). Higher magnification images of the regions in boxes, containing a high density of gold particles, are shown at right. Arrowheads designate additional gold particles. **D)** Immunogold labeling of the WT strain using secondary antibodies conjugated to gold particles, without primary antiserum. The number of gold particles detected on WT bacteria was significantly higher than the number detected on the *cag* Δ PAI mutant strain (mean 3.6 gold particles per WT cell and 1.8 gold particles per *cag* Δ PAI mutant cell, based on analysis of >200 bacteria of each type, p-value < 0.0001, Welch two sample *t*-test). Bars, 500 nm.

Experimental evidence of amino-terminal signal sequence cleavage.

Many surface-exposed proteins are predicted to undergo cleavage of an N-terminal signal sequence during Sec-dependent export across the inner membrane. To experimentally identify surface-exposed proteins that underwent cleavage of amino-terminal signal sequences, I analyzed the spectral data from the biotinylation experiments and sought to identify peptides that had been generated by non-tryptic cleavage events occurring near the amino-termini of proteins. Fourteen of the 39 surface-exposed proteins shown in Table 3 showed evidence of amino-terminal non-tryptic cleavage events consistent with Sec-dependent signal sequence processing (Table 4) (135). An alignment of these sequences, relative to the sites of non-tryptic cleavage, is shown in Figure 7. The consensus sequence includes a stretch of approximately 10 uncharged or hydrophobic residues upstream of the cleavage site, a semi-conserved leucine at position -3, and an alanine at position -1. An acidic patch of 1-3 residues was detected immediately downstream from the cleavage site (135). Interestingly, the signal peptide prediction software SignalP 4.1 (109) predicted the existence of only seven of these 14 cleavage sites, and for one of the proteins (HP1156, HopI), the cleavage site predicted by SignalP differed by one amino acid position compared to what was experimentally detected. This discrepancy suggests that the currently available signal peptide prediction software is not optimally designed for analysis of signal peptides in ϵ -Proteobacteria such as *H. pylori*.

Table 4. Detected amino-terminal non-tryptic cleavage events in the three biotinylated preparations (a)

42 Surface-exposed proteins with detected N-term non-tryptic cleavage in the three biotinylated preparations			
Gene number	Detected cleavage position (b)	Sequence (c)	Signalp 4.1 (d) Description
HP0025	20/21	GSILVSAISA-EDNGFFVVSAG	20/21 outer membrane protein (omp2)
HP0252	23/24	LIASMSLICRA-EEDGAFVYTD	23/24 outer membrane protein (omp7)
HP0537	18/19	SLVAFGVLSA-NVEQFGSFFEN	NO cag pathogenicity island protein (cag16)
HP0609	70/71	GLGFVSATSA-EDYNSSVYMWL	NO hypothetical protein
HP0655	20/21	FACINTSVRA-LENDGSKPMD	NO protective surface antigen D15
HP0896	19/20	SLSLSFILHA-EDDGFYTSVG	19/20 outer membrane protein (omp19)
HP0912	21/22	LALCASISYA-EDDGGFFTYG	21/22 outer membrane protein (omp20)
HP0913	41/42	LSLSFNVPVGA-EEDGGFMFRG	NO outer membrane protein (omp21)
HP1156	27/28	PLFFLNPLAA-EDDGFVMSVS	26/27 outer membrane protein (omp25)
HP1173	26/27	IGILSVQLDA-RSFVDGDDIDI	26/27 hypothetical protein
HP1177	21/22	LTTAASILHA-EDNGVFLSYG	21/22 outer membrane protein (omp27)
HP1243	20/21	GSILVSTLSA-EDDGFYTSVG	20/21 outer membrane protein (omp28)
HP1400	22/23	SLICQESLFA-KEKDYTLGKV	NO iron(III) dictrate transport protein (fecA)
HP1512	21/22	SCIIVASNLQA-QETHTTLGKV	NO iron-regulated outer membrane protein (frpB)

(a) Three biotinylated preparations were analyzed for spectral coverage. Fourteen of the 39 surface-exposed proteins (Table 3) were found to undergo non-tryptic cleavage at the amino-terminus, consistent with sec-dependent cleavage.

(b) Amino acid number, "/" indicates cleavage

(c) Amino acid sequence around cleavage, "-" indicates cleavage

(d) Amino acid number of predicted cleavage site determined by SignalP, "NO" indicates no sec-dependent cleavage predicted

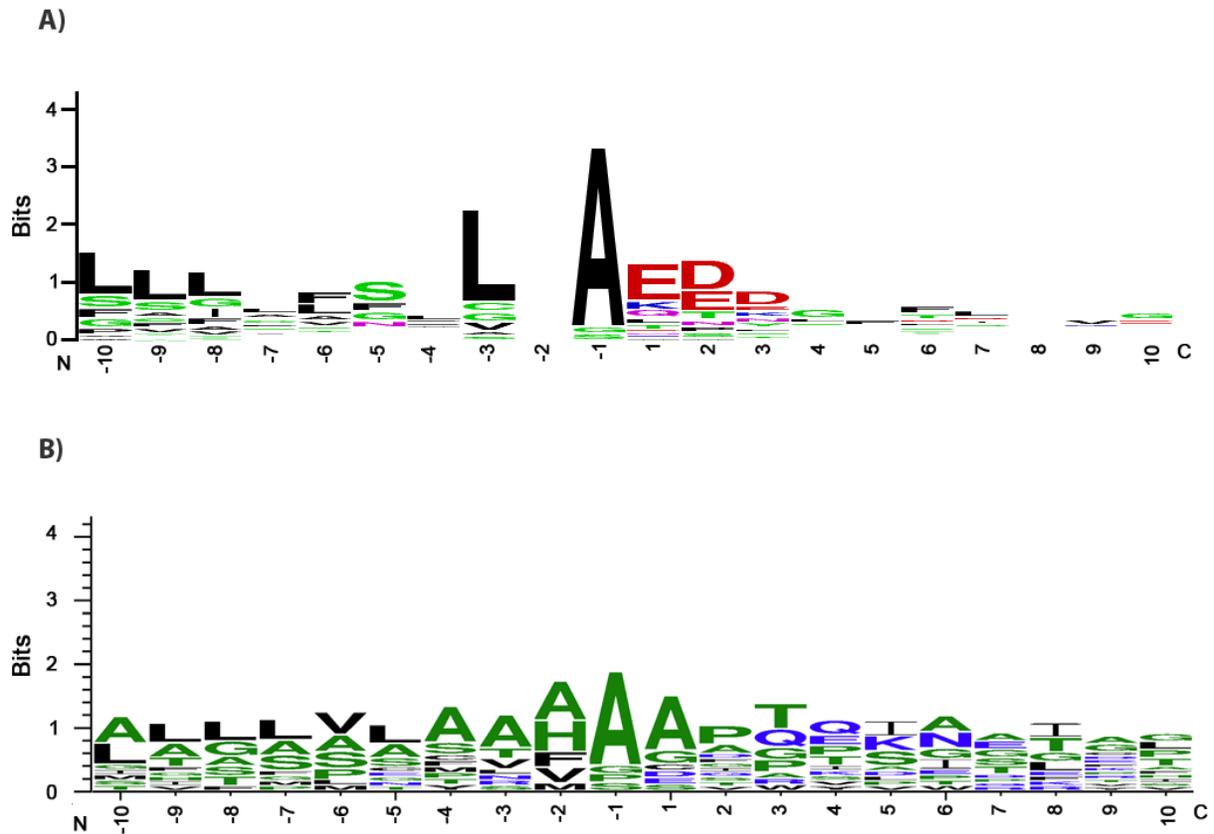


Figure 7. Experimentally detected sites of signal peptide cleavage. A) Identified peptides in the preparations of biotinylated proteins that corresponded to non-tryptic cleavage events and that could be matched to sites near the N-termini of proteins, suggesting the occurrence of signal peptide cleavage. This analysis was restricted to peptides that could be matched to a single protein. Amino-terminal non-tryptic cleavage was detected for 14 of the 39 surface-exposed proteins shown in Tables 3 and 4. In the alignment, the numbering of positions is relative to the detected non-tryptic cleavage site. B) Aligned 14 confirmed signal sequences and amino-terminal peptides from multiple other Gram-negative bacterial species from Signal Peptide Website (<http://signalpeptide.de>). Sequences were then aligned using WebLogo (weblogo.berkeley.edu).

Multiple classes of C-terminal motifs in surface-exposed proteins.

Targeting of proteins to the outer membrane is often dependent on the presence of a C-terminal motif required for interaction with the β -barrel assembly machinery (Bam) complex (13, 45). Therefore, I analyzed the C-terminal sequences of the 39 surface-exposed proteins shown in Table 3. The most commonly identified amino acids at the C-terminal position were phenylalanine (n = 16), lysine (n = 9) or tyrosine (n = 5) (135).

There were several similarities in the C-terminal sequences of the proteins terminating in phenylalanine (F) or tyrosine (Y). Specifically, 13 of the 39 surface-exposed proteins shown in Table 3 contain closely related C-terminal sequences characterized by a conserved tyrosine at the -5 position, a conserved arginine at the -13 position, either a valine (7 proteins) or phenylalanine (6 proteins) at position -9, and semi-conserved tyrosines at position -8 (11 proteins) and -15 (10 proteins) (Figure 8). Thus, the consensus C-terminal motif for these proteins is [Y/F X R XXX V/F Y XX Y XXX F/Y], and three of these proteins (HopQ, BabA, and BabB) contain an identical C-terminal motif (VYLN YVFAY). All 13 of these C-terminal sequences contain hydrophobic residues at the -1, -3, -5, -7, -8, -9 and -15 positions, which is consistent with a motif that allows insertion of proteins into the outer membrane through a BamA (or Omp85)-dependent process (13, 128). Twelve of the 13 proteins containing this motif are classified as Hop proteins and one as a Hop-related protein (HorB, HP0127). Eight of the 21 proteins with C-terminal phenylalanines or tyrosines, including the VacA-like protein FaaA, contain C-terminal sequences that are somewhat different from the

consensus described above, but still contain hydrophobic residues at the -3, -5, -7, and -9 positions.

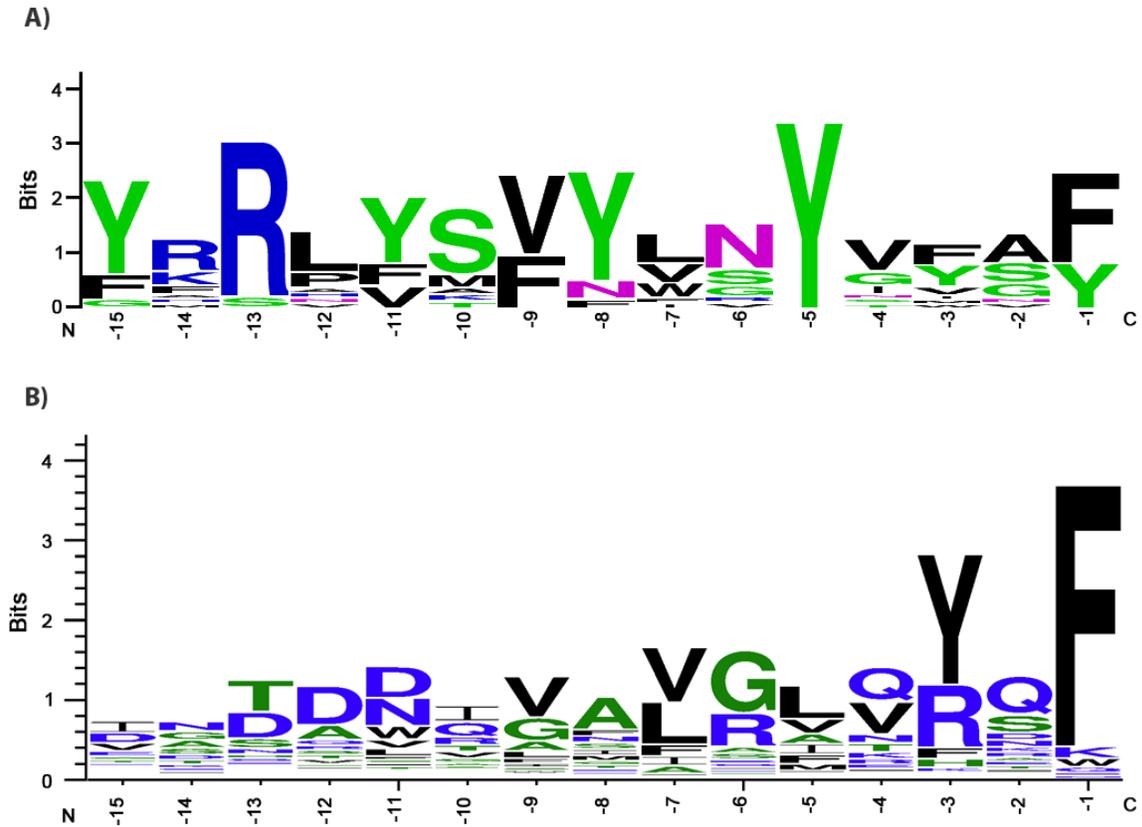


Figure 8. Analysis of a conserved C-terminal motif. A) Among the 39 surface-exposed proteins shown in Table 3, 15 contain a distinctive C-terminal motif, characterized by conserved residues at positions -15, -13, -9, -5 and -1 (135). Numbers correspond to the distance from the carboxyl-terminus. B) 100 randomly selected OMPs [collected from OMPdb (ompdb.org)] from other Gram-negative bacterial species were aligned. Sequences were aligned using WebLogo (weblogo.berkeley.edu).

Eighteen of the 39 proteins did not contain a C-terminal phenylalanine or tyrosine; in nine cases the C-terminal amino acid was lysine and in two cases the C-terminal amino acid was arginine. There was no apparent conservation of C-terminal sequences among the 9 proteins with C-terminal lysines or among the larger set of 18 proteins terminating in amino acids other than phenylalanine or tyrosine. However, many of these had C-terminal segments that were characterized by an abundance of basic residues (lysine and arginine) (135). Interestingly, several of these were previously annotated as lipoproteins. I hypothesized that the amine-reactive biotinylation reagent might preferentially label these C-terminal basic residues, leading to an increased representation of these proteins in the preparation of purified biotinylated proteins. To address this possibility, I biotinylated surface-exposed proteins with a cell-impermeable carboxyl-reactive biotinylation reagent, and then purified and analyzed the biotinylated proteins. By using this approach, I identified 32 proteins that were significantly enriched in the biotinylated preparation compared to the unlabeled control (135). Twenty-seven of these 32 proteins were also identified as surface-exposed in experiments using the amine-linked biotinylation method (Figure 9). Thus, the surface-exposed proteins identified when using a carboxyl-reactive biotinylation reagent were similar to those identified when using the amine-reactive biotinylation reagent. Eight of the 27 surface-exposed proteins identified using both methods terminated in either a lysine or arginine, and the proportion of surface-exposed proteins terminating in either lysine or arginine was similar in experiments using the two methods (3 of 32 with the carboxyl-reactive method and 9 of 85 with the amine-reactive method). This provides evidence that the

proteins identified using the amine-linked biotinylation approach were not identified simply as a consequence of preferentially labeling C-terminal basic residues.

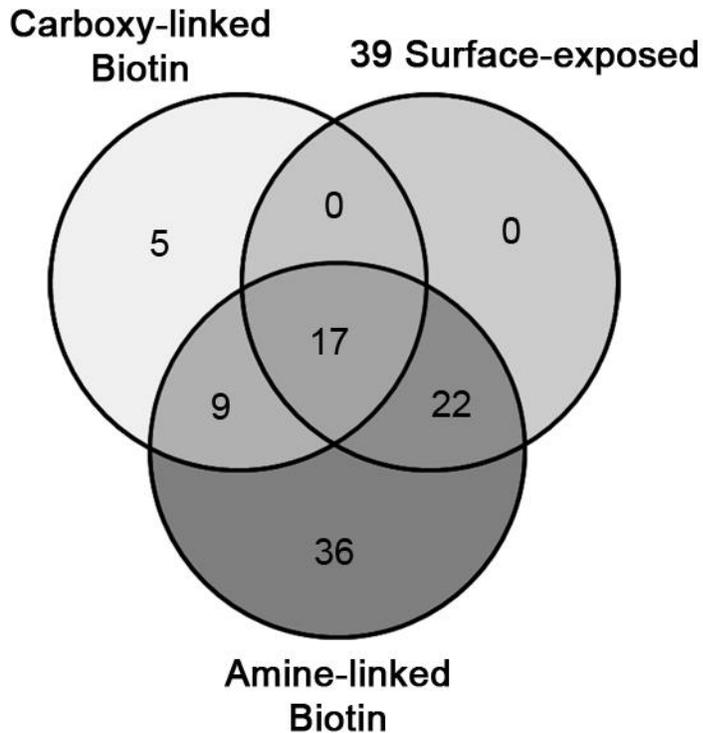


Figure 9. Comparison of surface-exposed proteins identified by carboxyl-reactive biotinylation compared to amine-reactive biotinylation. In experiments with an amine-reactive biotinylation reagent, I identified 85 proteins that were significantly enriched in the biotinylated preparation compared to a control preparation (135). In experiments with a carboxyl-reactive biotinylation reagent, I identified 32 proteins that were significantly enriched in the biotinylated preparation compared to a control preparation (135). The Venn diagram compares proteins identified using these two biotinylation methods, as well as the set of 39 proteins shown in Table 3. The proteins identified in experiments with the carboxyl-reactive biotinylation reagent were similar to the proteins identified in experiments using an amine-reactive biotinylation reagent.

Susceptibility of surface-exposed proteins to proteinase K digestion and cell surface topology of surface-exposed proteins.

I next undertook experiments designed to assess the susceptibility of surface-exposed proteins on intact bacteria to digestion by an exogenous protease. These experiments provided an additional approach for localizing specific proteins to the surface of *H. pylori*, and also allowed us to test the hypothesis that specific groups of surface-exposed OMPs might differ in susceptibility to proteolytic cleavage. I treated intact bacteria with proteinase K under conditions previously shown to preferentially digest surface-exposed proteins but not intracellular proteins (111, 124). I then used MudPIT to identify proteins for which there were significantly reduced numbers of assigned spectra detected in protease-treated bacteria compared to control bacteria. I identified 7 such proteins that were highly susceptible to proteolysis: VacA (HP0887), BabA (HP1243), BabB (HP0896), BabC (HopU, HP0317), HopQ (HP1177), HomA (HP0710), and HomD (Table 5) (135). Three of which (BabA, BabB, and HopQ) were identified using the biotinylation and fractionation methods. Proteinase K treatment did not cause any significant reduction in the abundance of peptides originating from inner membrane proteins, such as FixO (HP0145) and FtsH (HP1069) (135). The results were similar in experiments conducted at 4°C and at 22°C (Table 5).

Table 5. Surface-exposed proteins highly susceptible to proteinase K digestion, based on total assigned spectra^a

Original Annotation	Alternative names	Gene Number	Spectral Counts ^b				p Value ^c		Fold Change ^d	
			Untreated	Prot K	Untreated	Prot K	4°C	22°C	4°C	22°C
vacuolating cytotoxin	VacA	HP0887	986	185	793	183	< 0.0001	< 0.0001	5.85	5.98
outer membrane protein (Omp9)	HopU	HP0317	653	92	567	24	< 0.0001	< 0.0001	7.75	32
outer membrane protein (Omp27)	HopQ	HP1177	1446	773	1229	760	< 0.0001	< 0.0001	2.06	2.24
outer membrane protein (Omp19)	Babb/HopT	HP0896	577	95	510	33	< 0.0001	< 0.0001	6.63	21
outer membrane protein (Omp28)	BabA/Hops	HP1243	341	20	273	10	< 0.0001	< 0.0001	18.23	35.8
conserved hypothetical protein	HomA	HP0710	264	114	184	107	< 0.0001	< 0.0001	2.52	2.36
conserved hypothetical protein	HomD	HP1453	45	13	37	17	< 0.005	0.1463	3.68	2.94
Total assigned spectra			89,222	97,275	89,222	99,781				

^a Intact *H. pylori* were incubated with proteinase K at 4°C or 22°C, and the treated bacteria (as well as untreated controls) were then analyzed by MudPIT.

^b The combined results of three independent experiments for each condition (untreated or proteinase K-treated) are shown.

^c Results for proteinase K-treated bacteria were compared with results for untreated bacteria, using Fisher's exact test with Bonferonni multiple test correction.

^d The magnitude of changes in protein abundance in response to proteinase K digestion was calculated based on analysis of $2^{R_{sc}}$.

The methodology described above, based on analysis of all spectra assigned to individual proteins, allowed us to identify proteins that were highly susceptible to proteolytic digestion by an exogenous protease, but this is probably not a sensitive approach for identifying proteins that undergo proteolysis at relatively few sites. Therefore, I systematically analyzed the protein coverage of peptides with assigned spectra for each detected protein at the level of individual amino acids, which permitted a more in-depth assessment of sites where proteins had undergone proteolysis. The identification of protease-susceptible sites using this approach is illustrated in the top panel of Figure 10. Using this approach, six of the seven proteins described above were identified as susceptible to proteolytic digestion, and 9 additional protease-susceptible proteins were identified (Figure 10 and Table 6). 8 proteins were also identified using biotinylation and fractionation (AlpA, AlpB, BabA, BabB, HopF, HopQ, OipA, and FaaA). HomD, identified as protease-susceptible in the earlier analysis (Table 5), was excluded from the current analysis because the mass spectral data did not meet the inclusion criteria. Thirteen of the 15 protease-susceptible proteins met multiple criteria for surface-exposed localization in the earlier biotinylation and cell fractionation experiments (135). Additionally, VacA was identified as susceptible to proteinase K treatment. The other protein exhibiting evidence of protease susceptibility (cell division inhibitor MinD) is presumed to represent a false positive result. A systematic comparison of the protease-susceptibility of annotated OMPs detected in the proteomic analysis revealed that the majority of Hop proteins were susceptible to proteolytic digestion, but several Hop proteins (HopJ, HopK, HopE, HopG) were

resistant (Table 7). In comparison to Hop proteins, the Hof and Hor families of OMPs were relatively resistant to extracellular protease digestion (Table 7).

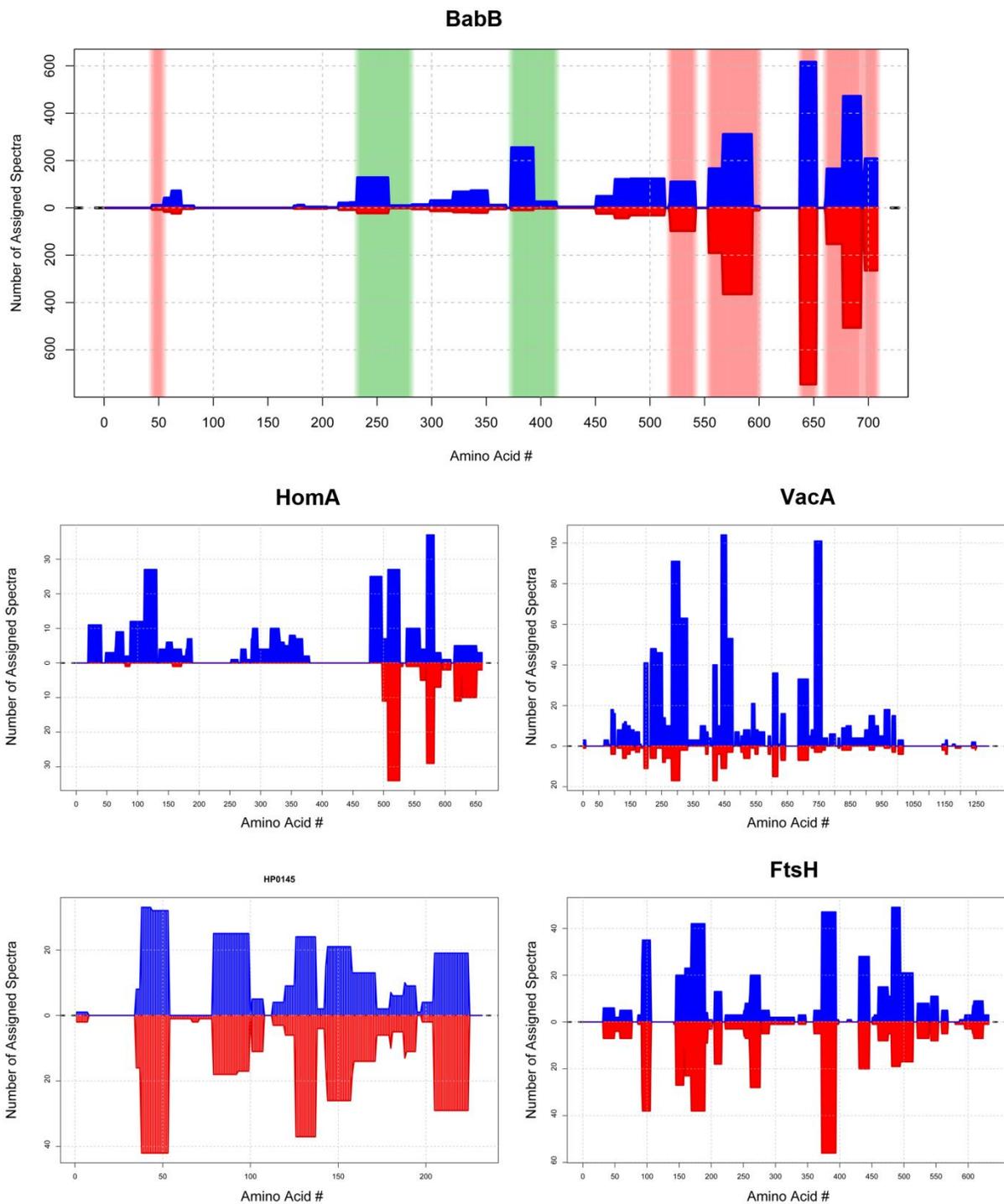


Figure 10. Susceptibility of surface-exposed proteins to proteinase K treatment. Intact *H. pylori* cells were treated with proteinase K (1 mg mL^{-1}) or buffer control at 4°C , and were then analyzed by MudPIT. Three experiments were performed and the results were merged. I then analyzed the assigned spectra for each detected protein at

the level of individual amino acids, and sites susceptible to proteolytic cleavage or resistant to proteolytic cleavage were identified, based on criteria defined in Methods. Blue bars (top half of graphs) depict numbers of assigned spectra from experiments with control (untreated) bacteria, red bars (bottom half of graphs) depict numbers of assigned spectra from experiments with proteinase K-treated bacteria, and orange lines indicate the ratio of assigned spectra untreated compared to protease-treated bacteria. Data for 5 representative proteins are shown. In the top panel, the green shading illustrates protease-susceptible regions in BabA (exhibiting a ≥ 5 fold difference in abundance of spectral counts when comparing untreated bacteria and protease-treated bacteria) and the pink shading illustrates protease-resistant regions (exhibiting a < 2 -fold difference in abundance of spectral counts). The designation of protease-susceptible or protease-resistant regions was restricted to residues for which ≥ 5 assigned spectra were detected in the untreated samples. The N-terminal portions of BabB, HomA and VacA were highly susceptible to protease digestion, whereas intracellular proteins such as the cytochrome c oxidase subunit FixO (HP0145) and the cell division protein FtsH were resistant to protease digestion.

Table 6. Proteins identified as susceptible to extracellular proteinase K digestion, based on analysis of individual amino acid sites

Original annotation	Alternative names	Gene number	Number Amino Acids				Normalized Ratio PK susceptible > not Susceptible ^d
			Protein length ^a	# residues available for analysis ^b	# residues susceptible to protease ^c		
outer membrane protein (Omp5)	HopM	HP0227	691	356	143	26.9	
outer membrane protein (Omp6)	HopA	HP0229	483	283	197	33.7	
outer membrane protein (Omp7)	HopF	HP0252	487	296	85	6.0	
outer membrane protein (Omp9)	HopU/BabC	HP0317	745	480	157	16.5	
cell division inhibitor	Mind	HP0331	268	121	40	8.5	
toxin-like outer membrane protein	FaaA	HP0610 ^e	1943	199	39	5.3	
outer membrane protein (Omp13)	HopH/OipA	HP0638	305	120	31	5.2	
conserved hypothetical protein	HomA	HP0710	660	283	205	38.6	
vacuolating cytotoxin	VacA	HP0887	1290	642	377	57.7	
outer membrane protein (Omp19)	BabB	HP0896	708	442	92	9.5	
outer membrane protein (Omp20)	HopC/AlpA	HP0912	515	360	208	23.4	
outer membrane protein (Omp21)	HopB/AlpB	HP0913	529	273	135	14.4	
outer membrane protein (Omp27)	HopQ	HP1177	641	495	367	45.7	
outer membrane protein (Omp28)	BabA	HP1243	733	444	190	21.2	
outer membrane protein (Omp29)	HopN	HP1342	691	356	143	26.9	

^a Total number of amino acids in the annotated gene.

^b Total number of amino acids with ≥ 5 assigned spectra in the untreated samples.

^c Total number of amino acids with ≥ 5 -fold more assigned spectra in the untreated samples compared to proteinase K treated samples.

^d Ratio of PK susceptible residues compared to non-susceptible residues normalized by total residues for each category identified.

^e HP0609 and HP0610 were originally annotated as two separate proteins, but correspond to a single protein (VacA-like protein FaaA).

Table 7. Protease susceptibility of annotated OMPs

OMP	Gene Number	Number Amino Acids			
		Length	# residues available for analysis ^a	# residues susceptible to protease ^b	% residues susceptible to protease ^c
Hop Family					
HopH	HP0638	305	120	31	25.83
HopJ	HP0477	367	129	0	0
HopK	HP0923	369	129	0	0
HopE	HP0706	273	162	0	0
HopL	HP1157	1230	221	34	15.38
HopG	HP0254	431	233	0	0
HopI	HP1156	696	259	46	17.76
HopB/AlpB	HP0913	529	273	135	49.45
HopA	HP0229	483	283	197	69.61
HopF	HP0252	487	296	85	28.72
HopM	HP0227	691	356	143	40.17
HopN	HP1342	691	356	143	40.17
HopC/AlpA	HP0912	515	360	208	57.78
HopD	HP0025	711	436	52	11.93
HopT/BabB	HP0896	708	442	92	20.81
HopS/BabA	HP1243	733	444	190	42.79
HopU/BabC	HP0317	745	480	157	32.71
HopQ	HP1177	641	495	367	74.14
Hop-Related Family					
HorI	HP1113	277	24	0	0
HorC	HP0324	254	77	20	25.97
HorE	HP0472	186	85	0	0
HorL	HP1395	242	85	20	23.53
HorF	HP0671	270	123	0	0
HorB	HP0127	286	125	0	0
HorJ	HP1469	248	131	0	0
HorH	HP1107	230	132	25	18.94
HorK	HP1501	388	221	0	0
Hom Family					
HomC	HP0373	700	25	25	100
HomD	HP1453	746	26	26	100
HomA	HP0710	660	283	205	72.44
Hof Family					
HofD	HP0487	480	87	0	0
HofF	HP0788	499	94	0	0
HofG	HP0914	514	177	0	0
HofB	HP1083	479	266	10	3.76

HofC	HP0486	528	326	0	0
HofE	HP0782	455	0	0	0

^a Total number of amino acids with ≥ 5 assigned spectra in the untreated samples.

^b Total number of amino acids with ≥ 5 -fold more assigned spectra in the untreated samples compared to proteinase K-treated samples.

^c Percentage of total number of amino acids with ≥ 5 assigned spectra in the untreated samples with ≥ 5 -fold more assigned spectra in the untreated samples compared to the proteinase K-treated samples.

In an analysis of the regions within individual proteins that were susceptible or resistant to proteolysis, I noted that many annotated OMPs (especially those in the Hop family) were highly susceptible to protease digestion within the amino-terminal portion of the protein (comprising roughly two-thirds of the entire length of the protein), whereas the carboxy-terminal regions of these proteins were protease-resistant. This result was observed for 10 proteins (BabA, BabB, BabC, HopQ, HomA, HopA, HopM, HopN, AlpA, and AlpB) (Figures 10 and 11 and data not shown). The protease-resistant C-terminal regions of these proteins ranged from about 150 to 250 amino acids in length. Analysis of these proteins with the β -barrel prediction program BOCTOPUS predicts that the C-terminal 160-300 amino acids of these proteins correspond to multiple trans-membrane β strand segments consistent with a β -barrel structure and that predicted trans-membrane segments are absent or infrequently present elsewhere within these proteins (Figure 11). These data suggest that most of the protease-susceptible OMPs contain a large protease-susceptible extracellular domain exported beyond the outer membrane, and a protease-resistant β -barrel domain at the C-terminus.

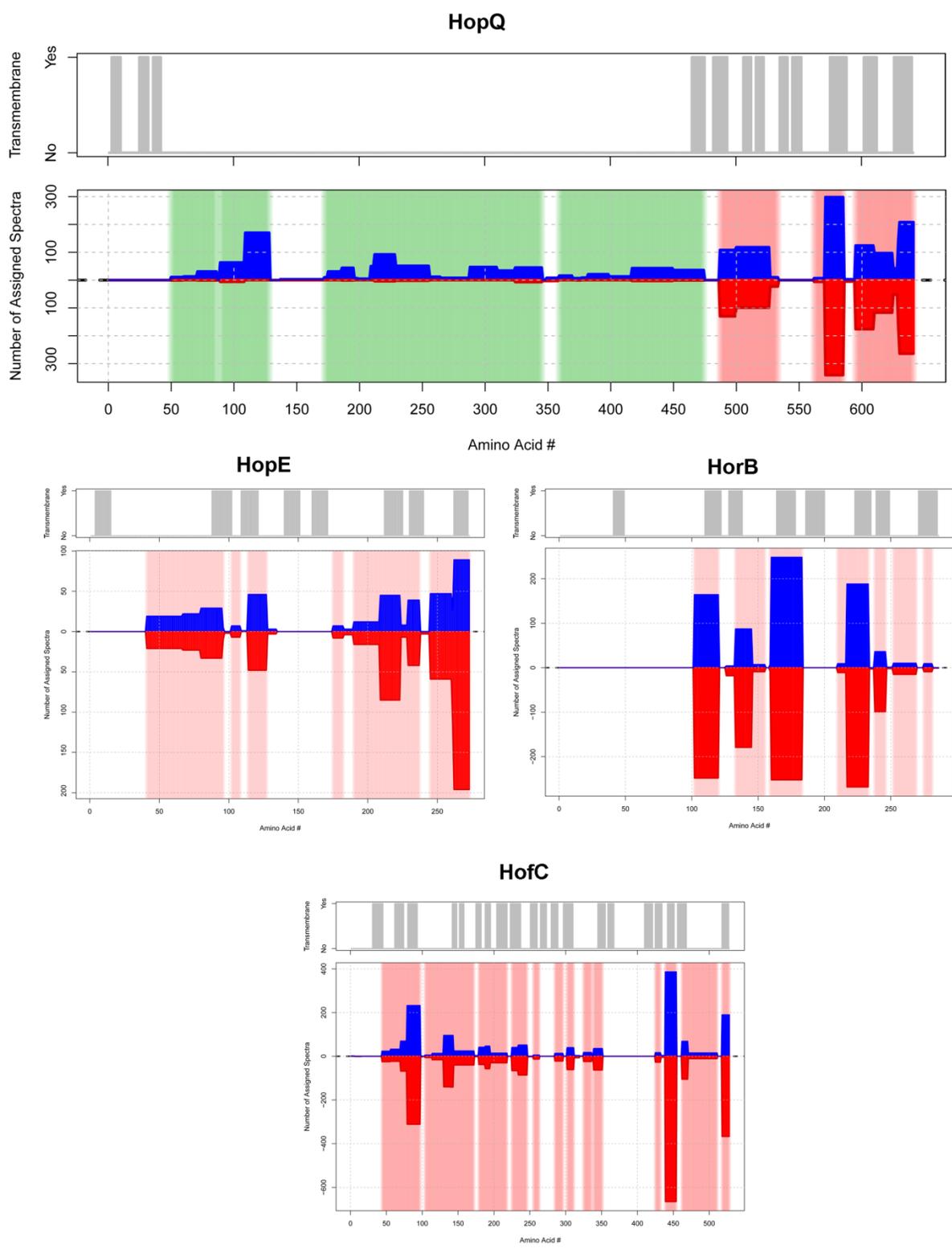


Figure 11. Resistance of predicted β -barrel regions to digestion by proteinase K. This Figure shows an analysis of the protease-susceptibility of 4 *H. pylori* OMPs, using

the approach shown in Figure 10. Blue bars (top half of graphs) depict numbers of assigned spectra from experiments with control (untreated) bacteria and red bars (bottom half of graphs) depict numbers of assigned spectra from experiments with proteinase K-treated bacteria. The green shading illustrates protease-susceptible regions (exhibiting a ≥ 5 fold difference in abundance of spectral counts when comparing untreated bacteria and protease-treated bacteria) and the pink shading illustrates protease-resistant regions (exhibiting a < 2 -fold difference in abundance of spectral counts). The designation of protease-susceptible or protease-resistant regions was restricted to residues for ≥ 5 assigned spectra were detected in the untreated samples. The locations of trans-membrane β strands were predicted using the program BOCTOPUS, and are shown at the top of each panel (grey indicates predicted trans-membrane β -strands). A large N-terminal portion of HopQ is highly susceptible to proteolytic digestion, whereas the C-terminal portion of HopQ is resistant to proteolytic digestion. The other 3 OMPs (HopE, HorB, and HofC) are relatively resistant to proteolysis. The resistant C-terminal region of HopQ corresponds to a region predicted by BOCTOPUS to have multiple trans-membrane β strands, consistent with a β -barrel structure. Three protease-resistant OMPs (HopE, HorB and HofC) are each predicted to have predominantly β -barrel structure.

Discussion

In this study, I used multiple complementary methods to identify and analyze proteins localized on the surface of *H. pylori*. I identified 39 proteins that met 3 criteria for surface localization or outer membrane localization (i.e. identification by biotinylation, enriched in a membrane fraction compared to soluble fraction, and enriched in a Triton-insoluble outer membrane fraction compared to a Triton-soluble inner membrane fraction), thereby providing stronger evidence for surface localization compared to the use of only a single method. Among these 39 proteins, many were previously known or predicted to be localized on the bacterial surface, whereas others were not previously known to be surface-exposed or associated with the outer membrane.

My primary method involved biotinylation of surface-exposed proteins with a membrane-impermeable biotinylation reagent. A potential concern with biotinylation-based methods is the possibility that periplasmic or inner membrane proteins might be labeled, but a previous study showed that the approach used in this study results in minimal labeling of the abundant *E. coli* periplasmic protein MalE (119). The biotinylation experiments allowed us to identify 85 putative surface-exposed *H. pylori* proteins (135). Eighteen putative surface-exposed proteins were identified in a previous study that used similar biotinylation-based methodology (119). Three surface-exposed proteins were identified in both the current study and the previous study: an HpaA-like protein (HP0410), CagM, and the secreted serine protease HtrA (15, 50, 119). There were several differences when comparing the methodology of the current study with that

used in a previous study (119). First, the current study used a MudPIT-based methodology for identifying proteins, whereas the previous study used 2D gel-based methods; limitations in the use of 2D gel methods for detection of membrane proteins may account for the detection of relatively few annotated OMPs in the previous study. Second, the current study compared preparations of biotinylated proteins with control non-biotinylated preparations, which allowed the exclusion of proteins that interacted non-specifically with the beads used in the purification process. Third, the current study used monomeric avidin-coated magnetic beads instead of avidin agarose beads. Finally, the current study analyzed bacteria grown in broth culture, whereas the previous study analyzed bacteria grown on plates. Collectively, these modified approaches were helpful in enhancing the detection of outer membrane proteins and reducing contamination from cytoplasmic proteins. To determine whether the bacterial growth conditions had a major influence on the results, I conducted experiments in which I compared preparations of biotinylated proteins derived from bacteria grown in broth culture with preparations of biotinylated proteins derived from bacteria grown on blood agar plates. In comparison to preparations of biotinylated proteins from bacterial grown in broth, preparations of biotinylated proteins from bacteria grown on solid medium had a lower proportion of assigned spectra for annotated OMPs (relative proportion of total assigned spectra, plate vs. broth = 0.5) and a higher proportion of assigned spectra for annotated intracellular proteins (relative proportion of total assigned spectra, plate vs. broth = 9.9) (135).

The set of proteins meeting three criteria for surface localization (Table 3) included three *cag* pathogenicity island (PAI)-encoded proteins that are components of

the *cag* T4SS (Cag3, CagM and CagT) (33). I also detected surface-exposed proteins that were previously annotated as lipoproteins. Among the proteins shown in Table 3, five proteins contain a sequence motif known as a lipobox (required for the recognition of posttranslational lipidation machinery), based on analysis by DOLOP (7) [HpaA paralog HP0492, CagT HP0532, predicted lipoprotein HP1002, Lpp20 (HP1456), and predicted lipoprotein HP1571]. In *E. coli* and many other well-studied Gram negative species, most lipoproteins are anchored to either the outer membrane or the inner membrane and face the periplasm (91), but some lipoproteins may also be surface-exposed and exported to the bacterial surface by an unknown mechanism (for example, JlpA in *C. jejuni* and TbpB in *N. meningitidis*) (13, 75, 95). The findings in the current study suggest that the surface of *H. pylori* is decorated with multiple lipoproteins. The Lol system, which is required for lipoprotein export, is considerably different in *H. pylori* compared to many other bacterial species (77), and this could potentially be a factor that influences the localization of *H. pylori* lipoproteins. Finally, I detected multiple surface-exposed that are annotated as “hypothetical proteins”. Further study will be required to investigate the localization and functions of these proteins in greater detail.

Several proteins previously reported to be localized to the surface of *H. pylori* are not included on the list of proteins in Table 3, which indicates that this is not a comprehensive list of the proteins exported to the *H. pylori* surface. Therefore, I also analyzed the proteins that met two criteria consistent with a surface-exposed outer membrane localization (Figure 5) (135). This analysis revealed that the set of proteins enriched in the biotinylated preparations and enriched in the total membrane preparations (but not enriched in the Triton-insoluble outer membrane preparation)

contained numerous proteins that are known or predicted to be localized to the outer membrane, such as HopA, HopM, HopN, HopU/BabC, and HorF.

Although the experiments reported in this manuscript provide a much more comprehensive understanding of the *H. pylori* cell-surface proteome than has been previously reported, there are several limitations of the current study. First, I analyzed mainly bacteria that were cultured in a single growth condition (late log phase from broth culture). These conditions were selected based on the results of a previous microarray gene expression study, which reported that several virulence-associated genes were maximally expressed in the late-log phase of growth (130). Additional surface-exposed proteins might be detectable if alternate conditions were studied. Identification of surface-exposed proteins under well-defined laboratory conditions, as described in this study, provides an important foundation for future studies investigating how the surface proteins of *H. pylori* change under alternate growth conditions. Second, the mass spectrometry-based approach used in these studies might fail to detect surface proteins that contain a small number of tryptic cleavage sites, surface proteins with extensive post-translational modifications, or surface proteins expressed in low abundance. Finally, I studied only a single strain, which is known to lack production of at least four outer membrane proteins [HopP/SabA, HopO/SabB, HopZ, and HomB] (3, 132) and lacks flagella (124).

Many of the surface-exposed proteins identified in this study are predicted to be exported through a process that involves Sec-dependent cleavage of an amino-terminal signal sequence. In support of this view, I detected non-tryptic cleavage events near the amino-terminus of numerous proteins. The presence in many of these proteins of a

conserved C-terminal motif ending in phenylalanine or tyrosine, and containing hydrophobic residues at -1, -3, -5, -7, -8, -9 and -15 positions, is consistent with insertion of these proteins into the outer membrane through a Bam-dependent process (13, 45, 128). Most of the Hop or Hor proteins contained the consensus C-terminal motif [Y/F X R XXX V/F Y XX Y XXX F/Y]. Surprisingly, many of the surface-exposed proteins detected in this study, including several putative lipoproteins, did not contain a C-terminal motif terminating in phenylalanine or tyrosine, but instead terminated in lysine. The mechanism by which these and other proteins lacking a typical C-terminal motif are exported to the bacterial surface is not known.

As an additional approach for analyzing the cell-surface proteome of *H. pylori*, I subjected intact bacteria to treatment with proteinase K and sought to identify a subset of surface-exposed proteins that were susceptible to proteolysis. Initial studies, based on analyzing the total numbers of spectral counts assigned to individual proteins, allowed us to identify 7 surface-exposed proteins [VacA, BabA, HopQ, BabB, BabC (HopU), HomA, and HomD] that were highly susceptible to protease digestion. VacA is a toxin that is secreted through an autotransporter pathway (20, 34), and subsequently can be released as a soluble protein into the extracellular space or remain attached to the bacterial surface (20, 55). BabA is an *H. pylori* adhesin that binds to the Lewis b antigen on host cells (54), and HopQ is an outer membrane protein that influences the activity of the *cag* type IV secretion system (10). In contrast to VacA, BabA, and HopQ, virtually nothing is known about the functions of BabB, BabC, HomA and HomD. BabA, BabB, BabC and HopQ are all classified as Hop proteins, whereas HomA and HomD are classified within a small family of OMPs known as Hom proteins (3). A previous

study used proteinase K treatment coupled with 2D gel electrophoretic methods to identify five protease-susceptible surface-exposed *H. pylori* proteins (118). Two of the proteinase K-susceptible proteins identified in the previous study (VacA and HopQ) (118) were also shown to be protease-susceptible in the current study.

I also analyzed the protein coverage of peptides with assigned spectra from the proteinase K experiments at the level of individual amino acids, which provided a more sensitive approach for detecting proteins that were susceptible to proteolytic cleavage. By using this approach, I identified 15 proteins that were protease-susceptible. All but one of these were identified as enriched in at least two out of the three biotinylation and cell fractionation experiments (VacA was enriched in TM and OM, HomA and BabC were enriched in the biotinylated preparations and the TM fractionations) (Figure 12). Conversely, the majority of surface-exposed proteins identified in the biotinylation and cell fractionation experiments were not susceptible to proteolytic cleavage. Thus, the proteinase K experiments identified a specialized subset of protease-susceptible surface-exposed proteins with properties that differ from those of protease-resistant surface-exposed proteins.

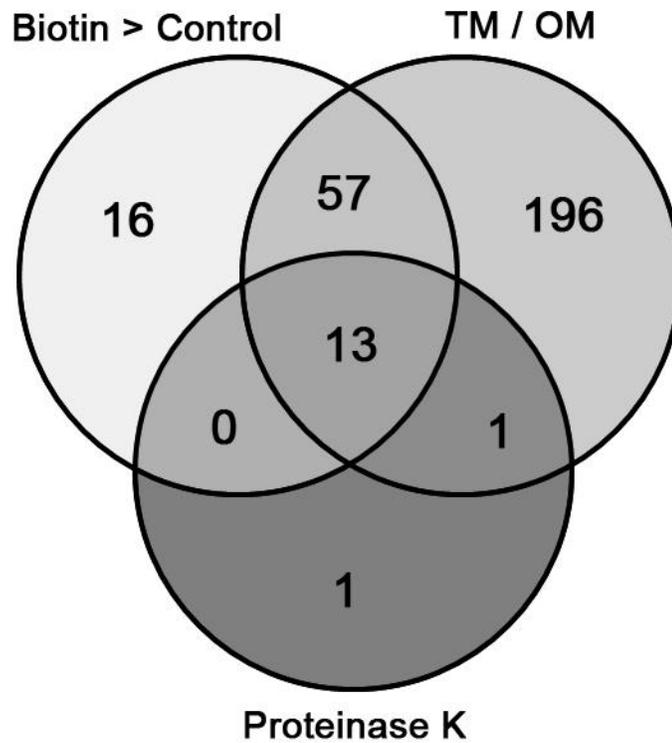


Figure 12. Subset of surface-exposed outer membrane proteins that are susceptible to digestion by proteinase K. Among the identified using methods based on biotinylation or differential detergent solubility (Figure 5 and Table 3), 13 were susceptible to digestion by extracellular proteinase K. TM / OM corresponds to proteins enriched in either the TM > CP,PP or OM > IM.

The set of protease-susceptible proteins included numerous OMPs classified within the Hop or Hom family, as well as VacA and a VacA-like protein (FaaA) (Table 6). Among the protease-susceptible Hop and Hom proteins, the C-terminal regions were resistant to protease digestion, whereas the N-terminal portions were susceptible. These data suggest that the N-terminal portions of these OMPs are exported beyond the outer membrane. The protease-resistant C-terminal domains of these OMPs are predicted to have multiple trans-membrane β -strands consistent with a β -barrel structure (based on the use of the program BOCTOPUS), and all contain the conserved C-terminal motif shown in Figure 8. Based on similarities between the protease susceptibility patterns of VacA and OMPs, I suggest that the N-terminal domains of these OMPs are exported beyond the outer membrane through an autotransporter (type V secretion) pathway, similar to the secretion of the VacA passenger domain (20, 34). In contrast to VacA, which undergoes proteolysis to cleave the passenger domain from the β -barrel domain, there is not yet any evidence that the protease-susceptible OMPs undergo similar proteolytic cleavage.

Although a subset of annotated OMPs were susceptible to proteolytic digestion, many others were relatively resistant (Table 7). All three Hom proteins detected in this study (HomA, HomC and HomD) and many of the Hop proteins exhibited evidence of susceptibility to proteolytic cleavage, whereas Hof and Hor proteins were relatively resistant (Table 7). There are several possible explanations that may account for the observed differences in susceptibility to proteolytic cleavage, including variation among the proteins in the proportion of β -barrel structure, differences among the proteins in the

orientation of large loops (projecting toward the periplasm or the bacterial surface), differences in physical accessibility of extracellular domains to the protease active site, and differences in amino acid composition. In contrast to the protease-susceptible Hop proteins, at least one of the Hop proteins resistant to proteolytic cleavage (HopE) is predicted to have a structure comprised of predominantly transmembrane β -strands, based on analysis by BOCTOPUS (Figure 11), which likely accounts for its resistance to proteolysis. Similarly, HorB and HofC are predicted to have structures comprised of predominantly transmembrane β -strands, and these proteins were resistant to proteolysis (Figure 11). Thus, the current data suggest that β -barrel-containing segments of *H. pylori* OMPs are relatively resistant to proteolysis, whereas large extracellular domains of *H. pylori* OMPs with non- β -barrel structures are often susceptible to proteolysis (Figure 13).

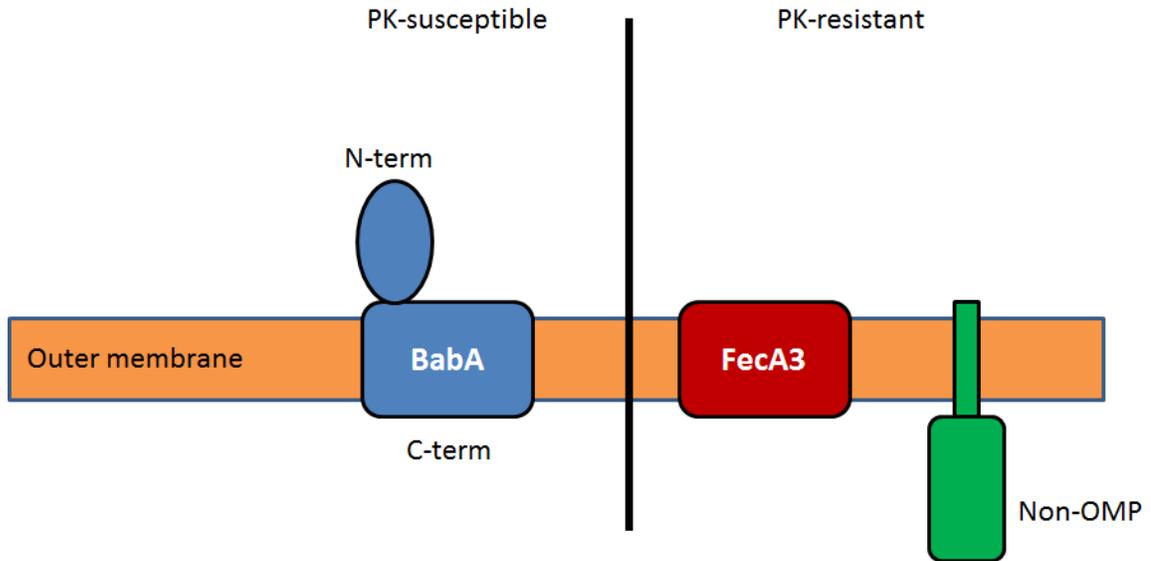


Figure 13. Proposed model of proteinase K-susceptibility results. The amino-terminal portions of 13 OMPs were highly susceptible to extracellular proteinase K digestion. The carboxy-terminal portions of these OMPs were resistant to proteinase K. OMP BabA is illustrated as a representative example. Regions of protease-resistance corresponded to regions of proteins with many predicted trans-membrane β strands. Proteins with predicted trans-membrane β strands occurring throughout the length of the amino acid sequence also were protease-resistant. OMP FecA3 is a representative example. Also, non-OMPs identified on the surface using biotinylation and fractionation were not susceptible to extracellular proteinase-K (exception was VacA-paralog FaaA). I speculate these proteins either do not contain large extracellular domains or are innately resistant to proteinase K.

The passenger domains of most proteins secreted through an autotransporter (type V secretion) pathway, including *H. pylori* VacA, have β -helical structures (42, 63). Interestingly, analysis of the protease-sensitive putative autotransporter proteins identified in this study using the program BetaWrapPro predicted the existence of β -helical structure within VacA (consistent with the crystal structure of this protein) (42, 63), but not within the other protease-sensitive OMPs. Thus, it seems likely that the N-terminal portions of many protease-susceptible *H. pylori* OMPs differ in structure compared to most known autotransporter passenger domains. In support of this view, a recent X-ray crystallographic study of the N-terminal region of a *H. pylori* SabA (an OMP that is related to BabA, but not produced by the *H. pylori* strain analyzed in this study) revealed primarily α -helical motifs instead of the β -helical motifs characteristic of most autotransporter passenger domains (63, 105).

In summary, these experiments provide valuable new insights into the repertoire of proteins present on the surface of *H. pylori*. In addition, my detailed analysis of the susceptibility of surface proteins to proteinase K digestion provides the first systematic analysis of the topology of *H. pylori* cell-surface proteins, and provides experimental evidence in support of an autotransporter mode of export for many of these proteins. Several of the surface-exposed proteins identified in this study are known to have important roles in mediating *H. pylori*-host interactions, and the functions of others are not yet known. The surface-exposed proteins with large extracellular domains identified in this study are especially likely to be localized in close proximity to host cells, and

therefore, I suggest that these proteins in particular may have important roles in mediating interactions with the host.

CHAPTER III

ALTERATION OF THE *HELICOBACTER PYLORI* MEMBRANE PROTEOME IN RESPONSE TO CHANGES IN ENVIRONMENTAL SALT CONCENTRATION

Introduction

Multiple risk factors for the development of symptomatic gastric disease in *H. pylori*-infected humans have been identified. High dietary salt intake is one such factor, and has been identified as a risk factor for gastric cancer (23, 36, 37, 133). Currently, the reason for this is unknown. My hypothesis is high levels of environmental salt alter the abundance of several surface-exposed *H. pylori* proteins and may be contributing to adverse clinical outcomes. This hypothesis is based on the observation that CagA is more abundant in bacteria grown under high salt conditions compared to low salt conditions (81, 82). Therefore, it is hypothesized that salt-induced alterations in CagA production contribute to the increased gastric cancer risk associated with a high salt diet (23, 40). There are likely other proteins that also change in abundance when *H. pylori* is grown under high salt conditions

Thus far, there has been relatively little progress in analyzing the regulation of *H. pylori* membrane protein production in response to changes in environmental salt concentration. A previous study used 2D-DIGE proteomic methods to analyze *H. pylori* lysates and identified 31 *H. pylori* proteins that are regulated in response to changes in

the salt concentration of the bacteriologic culture medium (81), but 2D-DIGE is considered suboptimal for detecting and monitoring the abundance of membrane proteins and low-abundance proteins (137). Therefore, the previous 2D-DIGE study may have failed to detect various salt-responsive changes in the abundance of *H. pylori* membrane proteins.

In the current study, I sought to identify *H. pylori* membrane proteins that change in abundance in response to alterations in environmental salt concentrations. I used both a label-free spectral counting approach, as well as Isobaric Tags for Relative Absolute Quantitation (iTRAQ). I selected the laboratory strain 26695 for analysis since it is a prototype strain with a detailed genome annotation, and I also analyzed strain 7.13, which can colonize Mongolian gerbils and promote the development of gastric cancer in this model (38). I report on salt-responsive changes in protein abundance that were detected in both *H. pylori* strains, as well as strain-specific responses. I discuss how the observed changes in the bacteria in response to high salt conditions may influence interactions between *H. pylori* and gastric epithelial cells, as well as the host immune response to the bacteria, and propose that these changes may account, at least in part, for the increased risk of gastric cancer associated with a high salt diet.

Methods

Bacterial strains and growth conditions

H. pylori strain 26695 is a reference strain with a fully annotated genome sequence that was originally isolated from a patient that had gastritis (28, 132). *H. pylori* strain 7.13 can colonize the stomach in a Mongolian gerbil model, sometimes resulting in gastric cancer (40, 96). The complete genome sequence of a closely related strain (strain B8) was used in the current study when analyzing proteomic data (32). Both strains have a functional *cag* T4SS. Strain 26695 lacks flagella and does not produce several OMPs (including HopZ, SabA, SabB, and HomB) (132). Strain 7.13 does not produce VacA (32, 92). For the proteomic studies, *H. pylori* strains were grown at 37°C in ambient air containing 5% CO₂ on either trypticase soy agar plates supplemented with 5% sheep blood, or in sulfite-free Brucella broth supplemented with 10% fetal bovine serum (BB-FBS) and containing varying concentrations of added sodium chloride: 0.25% (low salt), 0.5% (regular salt), or 1.1% (high salt) (81). All proteomic analyses of the effect of salt concentration on protein abundance were performed by comparing bacteria grown in low salt conditions (0.25% added sodium chloride) with bacteria grown in high salt conditions (1.1% added sodium chloride). Bacteria were grown in the presence of 0.5% added sodium chloride for purposes of generating an unlabeled control preparation, as part of biotinylation experiments. The salt concentrations used in these experiments do not cause substantial alterations in bacterial growth (82).

Sample preparation for proteomic analysis

Bacteria were grown until late-log phase (absorbance 600 nm of approximately 0.7) as described previously (135). Bacteria were pelleted at 3,500 x g and resuspended in TNKCM (50 mM Tris, pH 7.4, 100 mM NaCl, 27 mM KCl, 1 mM CaCl₂, 0.5 mM MgCl₂). Bacteria were pelleted again and resuspended in lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA, protease inhibitor cocktail). Lysates were generated via sonication (4 pulses on ice at 15 sec each, with 45 sec rest between pulses). Unbroken cells were pelleted at 4,500 x g for 10 min. Total membrane fractions were generated by ultracentrifugation of the lysates at 40,000 x g for 30 min. Previous studies have shown that these total membrane fractions are enriched in inner and outer membrane proteins, but also contain detectable cytoplasmic and periplasmic proteins (119, 135). Total membrane pellets were rinsed with lysis buffer and solubilized in RIPA buffer (50 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA, 1% NP-40, 0.25% sodium deoxycholate, protease inhibitor cocktail). Non-solubilized debris was removed by centrifugation at 21,000 x g for 10 min. Total protein concentration was determined with BCA reagents (Pierce).

Biotinylation of surface-exposed proteins

Bacteria were grown to late-log phase in medium containing either 0.25% NaCl (low salt) or 1.1% NaCl (high salt), and surface-exposed proteins on intact bacteria were biotinylated and purified as described in Chapter II. To generate an unlabeled control preparation, bacteria were grown in medium containing 0.5% NaCl (normal salt), and the non-biotinylated bacteria were processed in parallel with the biotinylated bacteria.

This biotinylation protocol allowed labeling of proteins on the surface of strain 26695, but similar experiments with strain 7.13 were less successful, possibly due to fragility of the latter strain during the multiple steps required for biotinylation.

Label-free analysis

For each strain analyzed, one culture was grown in high salt medium (1.1% NaCl) and one culture was grown in low salt medium (0.25% NaCl). Proteins in the membrane preparations were precipitated with ice-cold acetone overnight at -20°C. Following precipitation, samples were centrifuged at 18,000 x g at 4°C, and precipitates were washed with cold acetone, dried, and reconstituted in 8M urea in 250 mM Triethyl ammonium bicarbonate (TEAB) (pH 8.0). Samples were reduced with Tris(2-carboxyethyl)phosphine (TCEP), alkylated with methyl methanethiosulfonate (MMTS), diluted with TEAB to obtain a solution containing 2M urea, and digested with sequencing-grade trypsin overnight at 37°C. For label-free 2D-LC-MS experiments, peptides were loaded onto a self-packed biphasic C18/SCX MudPIT column using a helium-pressurized cell (pressure bomb). The MudPIT column consisted of 360 x 150 µm i.d. fused silica, which was fritted with a filter-end fitting (IDEX Health & Science) and packed with 6 cm of Luna SCX material (5 µm, 100Å) followed by 4 cm of Jupiter C18 material (5 µm, 300Å, Phenomenex). Once the sample was loaded, the MudPIT column was connected using an M-520 microfilter union (IDEX Health & Science) to an analytical column (360 µm x 100 µm i.d.), equipped with a laser-pulled emitter tip and packed with 20cm of C18 reverse phase material (Jupiter, 3µm beads, 300Å, Phenomenex). Using an Eksigent NanoLC HPLC and Autosampler, MudPIT analysis

was performed with an 11-step salt pulse gradient (25, 50, 75, 100, 150, 200, 250, 300, 500, 750 mM, 1M ammonium acetate). Following each salt pulse, peptides were gradient-eluted from the reverse phase analytical column at a flow rate of 500 nL/min, and the mobile phase solvents consisted of 0.1% formic acid, 99.9% water (solvent A) and 0.1% formic acid, 99.9% acetonitrile (solvent B). For the peptides from the first 10 SCX fractions, the reverse phase gradient consisted of 2–40 %B in 90 min, followed by a 15 minute equilibration at 2 %B. For the last SCX-eluted peptide fractions, the peptides were eluted from the reverse phase analytical column using a gradient of 2-98 %B in 100 min, followed by a 10 min equilibration at 2 %B. Peptides were introduced via nano-electrospray into an LTQ-Orbitrap Velos mass spectrometer (Thermo Scientific), and the data were collected using a 17-scan event data-dependent method. Full scan (m/z 350-2000) spectra were acquired with the Orbitrap as the mass analyzer (resolution 60,000), and the 16 most abundant ions in each MS scan were selected for collision-induced dissociation in the LTQ. An isolation width of 2 m/z, activation time of 10 ms, and 35% normalized collision energy were used to generate MS/MS spectra. The MSⁿ AGC target value was set to 1×10^4 , and the maximum injection time was set to 100 ms. Dynamic exclusion was enabled, using a repeat count of 1 within 10 sec and exclusion duration of 30 sec. For peptide identifications, tandem mass spectra were converted into DTA files using Scansifter and were searched using a custom version of Sequest (Thermo Fisher Scientific) operating on the Vanderbilt ACCRE computing cluster (85). MS/MS spectra were searched against a concatenated forward and reverse (decoy) database containing the *H. pylori* 26695 or B8 strain subsets of UniprotKB protein database (www.uniprot.org). Additional search parameters included:

trypsin enzyme specificity, monoisotopic masses used for searching product ions, and allowance of oxidation of methionine and MMTS modification of cysteine as variable modifications. Scaffold 4.1.1 (Proteome Software) was used to summarize and compare search results, where a minimum probability threshold of 95% was required for peptide identifications and data were filtered to a false-discovery rate (FDR) of <3% at the protein level.

To analyze preparations from biotinylated bacteria, shotgun proteomic analysis was performed by first resolving 50 µg of protein preparations approximately 1 cm using a 10% Novex precast gel, and then performing in-gel tryptic digestion at 37°C to recover peptides (125). These peptides were analyzed via MudPIT essentially as described in (86, 90) and above, with the exception that data were acquired using an LTQ mass spectrometer (Thermo Scientific) and MudPIT separations performed using a NanoAquity HPLC system (Waters). Unlike the analysis of total membrane preparations, where data were collected over 11 salt steps, only 8 steps (~ 16 hr) were utilized for analyzing this less complicated cell surface proteome. Both the intact masses (MS) and 5 data-dependent fragmentation (MS/MS) spectra of the peptides were collected. The peptide MS/MS spectral data were searched against the *H. pylori* 26695 protein databases using Sequest (141). Resulting identifications were collated and filtered using IDPicker 3 (84), where a minimum probability threshold of 95% was required for peptide identifications and data were filtered to a false-discovery rate (FDR) of <3% at the protein level. When analyzing data from biotinylated preparations, only spectra assigned to unique proteins were considered for spectral counting.

iTRAQ analysis

One 2-plex iTRAQ experiment was conducted for each *H. pylori* strain assayed. Total membrane preparations were generated as described above, and proteins were precipitated with ice-cold acetone overnight at -20°C. Following precipitation, samples were centrifuged at 18,000 x g at 4°C, and precipitates were washed with cold acetone, dried, and reconstituted in 8M urea in 250 mM TEAB (pH 8.0). Samples were reduced with 5 μ L of 50 mM TCEP, alkylated with 2.5 μ L of 200 mM MMTS, diluted 4-fold with TEAB to obtain a final solution containing 2M urea, and digested with sequencing-grade trypsin overnight. Peptides were then labeled with iTRAQ reagents according to the manufacturer's instructions (reporter tags: strain 26695, low salt = 114, high salt = 117; strain 7.13, low salt = 115, high salt = 117) (AB Sciex). For 50 μ g of protein, 1 unit of labeling reagent was used. Labeling reagent was reconstituted in ethanol such that each protein sample was labeled at a final concentration of 90% ethanol, and labeling was performed for 2 hr. The resulting labeled peptides were then desalted by a modified Stage-tip method (116). iTRAQ-labeled samples were mixed and acidified with TFA. A disc of C18 extraction membrane (Chrom Tech. Inc. C18 SPE Empore disk) was cored with a 16-gauge needle, and the cored piece of membrane was fitted tightly into a 200 μ L pipet tip. Three milligrams of C18 resin (Phenomenex Jupiter C18, 5 μ m particle size) was suspended in 200 μ L of methanol and loaded into the pipet tip containing the cored C18 membrane. The C18 material was packed into the tip using centrifugation to form a resin-packed C18 clean-up tip (resin tip). Multiple resin tips were used for desalting the mixed 2-plex iTRAQ samples. Resin tips were equilibrated with 0.1% TFA in HPLC-grade water, labeled peptides were loaded into the tip by centrifugation, washed with

0.1% TFA, and eluted with 100 μ l of 80% ACN containing 0.1%TFA. All eluted peptides were pooled together and dried by speed-vac centrifugation. Peptides were reconstituted in 0.1% formic acid, and peptides were loaded onto a self-packed biphasic C18/SCX MudPIT column using a helium-pressurized cell (pressure bomb). The MudPIT column consisted of 360 x 150 μ m i.d. fused silica, which was fritted with a filter-end fitting (IDEX Health & Science) and packed with 6 cm of Luna SCX material (5 μ m, 100 \AA) followed by 4 cm of Jupiter C18 material (5 μ m, 300 \AA , Phenomenex). Once the sample was loaded, the MudPIT column was connected using an M-520 microfilter union (IDEX Health & Science) to an analytical column (360 μ m x 100 μ m i.d.), equipped with a laser-pulled emitter tip and packed with 20 cm of C18 reverse phase material (Jupiter, 3 μ m beads, 300 \AA , Phenomenex). Using an Eksigent NanoLC HPLC and Autosampler, MudPIT analysis was performed with a 13-step salt pulse gradient (0, 25, 50, 75, 100, 150, 200, 250, 300, 500, 750 mM, 1M, and 2M ammonium acetate). Following each salt pulse, peptides were gradient-eluted from the reverse analytical column at a flow rate of 500 nL/min, and the mobile phase solvents consisted of 0.1% formic acid, 99.9% water (solvent A) and 0.1% formic acid, 99.9% acetonitrile (solvent B). For the peptides from the first 11 SCX fractions, the reverse phase gradient consisted of 2–50 %B in 90 min, followed by a 15 min equilibration at 2 %B. For the last 2 SCX-eluted peptide fractions, the peptides were eluted from the reverse phase analytical column using a gradient of 2-98 %B in 100 min, followed by a 10 min equilibration at 2 %B. Peptides were introduced via nano-electrospray into a Q Exactive mass spectrometer (Thermo Scientific). The Q Exactive was operated in the data-dependent mode acquiring HCD MS/MS scans ($R = 17,500$) after each MS1 scan ($R =$

70,000) on the 18 most abundant ions using an MS1 ion target of 1×10^6 ions and an MS2 target of 5×10^4 ions. The maximum ion time for MS/MS scans was set to 100 ms, the HCD-normalized collision energy was set to 26, dynamic exclusion was set to 25 sec, and peptide match and isotope exclusion were enabled. Mass spectra were processed using the Spectrum Mill software package (version B.04.00) (Agilent Technologies). MS/MS spectra acquired on the same precursor m/z (± 0.01 m/z) within ± 1 sec in retention were merged. MS/MS spectra of poor quality which failed the quality filter by not having a sequence tag length >1 were excluded from searching. A minimum matched peak intensity requirement was set to 50%. For peptide identification, MS/MS spectra were searched against either the *H. pylori* strain 26695 or strain B8 subsets of the UniprotKB protein database (www.uniprot.org). Additional search parameters included: trypsin enzyme specificity with a maximum of three missed cleavages, ± 20 ppm precursor mass tolerance, ± 20 ppm (HCD) product mass tolerance, and fixed modifications including MMTS alkylation of cysteines and iTRAQ labeling of lysines and peptide N-termini. Oxidation of methionine was allowed as a variable modification. Autovalidation was performed such that peptide assignments to mass spectra were designated as valid following an automated procedure during which score thresholds were optimized separately for each precursor charge state and the maximum target-decoy-based false-discovery rate (FDR) was set to 1.0%. To obtain iTRAQ protein ratios, the median was calculated for all peptides assigned to each protein. Because of a high interrelatedness among *H. pylori* outer membrane proteins (OMPs), only peptides that could be assigned to a single OMP were considered when calculating iTRAQ protein ratios.

Statistical analysis of assigned spectra

For label-free analysis of total membrane preparations, the identification of proteins enriched in one preparation compared to another was based on two criteria. First, the relative abundance of assigned spectra was calculated using a normalized ratio of spectral counts (R_{sc}): $\log_2 \left(\frac{a+0.5}{b+0.5} \right) + \log_2 \left(\frac{b_{total}-b+0.5}{a_{total}-a+0.5} \right)$ (101). For the identification of salt-responsive proteins, “a” indicates high salt and “b” indicates low salt. For the biotinylation experiment to identify biotinylated proteins, “a” indicates the biotinylated preparation and “b” indicates the unlabeled control. As a second criteria, the significance of differences in numbers of assigned spectra was calculated using a Fisher’s exact test (FET) with Benjamini-Hochberg (BH) multiple test correction. R_{sc} values were normalized to mean of zero prior to analysis. Proteins designated as salt-responsive had a normalized R_{sc} of < -0.585 or > 0.585 (corresponding to a ≥ 1.5 fold change), and a FET BH p value < 0.05 .

Triplicate samples of biotinylated preparations were analyzed and the resulting assigned spectra were merged prior to analysis (135). Proteins in biotinylated samples were identified as enriched over the mock control preparation (no biotinylation) with a FET BH p value < 0.05 and $R_{sc} > 1$ (135). Salt-responsive proteins were identified as described above for the label-free analysis.

To analyze iTRAQ data, frequency distribution histograms for \log_2 protein ratios were generated using Graphpad Prism 5.0. Ratios were binned to a value of 0.1. \log_2 ratios were then fitted to a normal distribution using least squares (non-linear) regression. The mean derived from the Gaussian fit was used to normalize \log_2 ratios

for a bin center of zero. The mean and standard deviation derived from the Gaussian fit were used to generate a Z-score for each Log_2 ratio and calculate p values that were subsequently corrected for multiple comparisons by the BH method (129). Proteins with a $p < 0.05$ were defined as significantly different in abundance. For strains 26695 and 7.13, p values of 0.05 correspond approximately to absolute Log_2 ratios of 1 and 0.8, respectively.

Non-linear regression analyses were performed in Graphpad Prism (Figure 14). Z-score, Fisher's exact test, Student's t -test, and multiple test corrections were calculated in R (<http://www.R-project.org>).

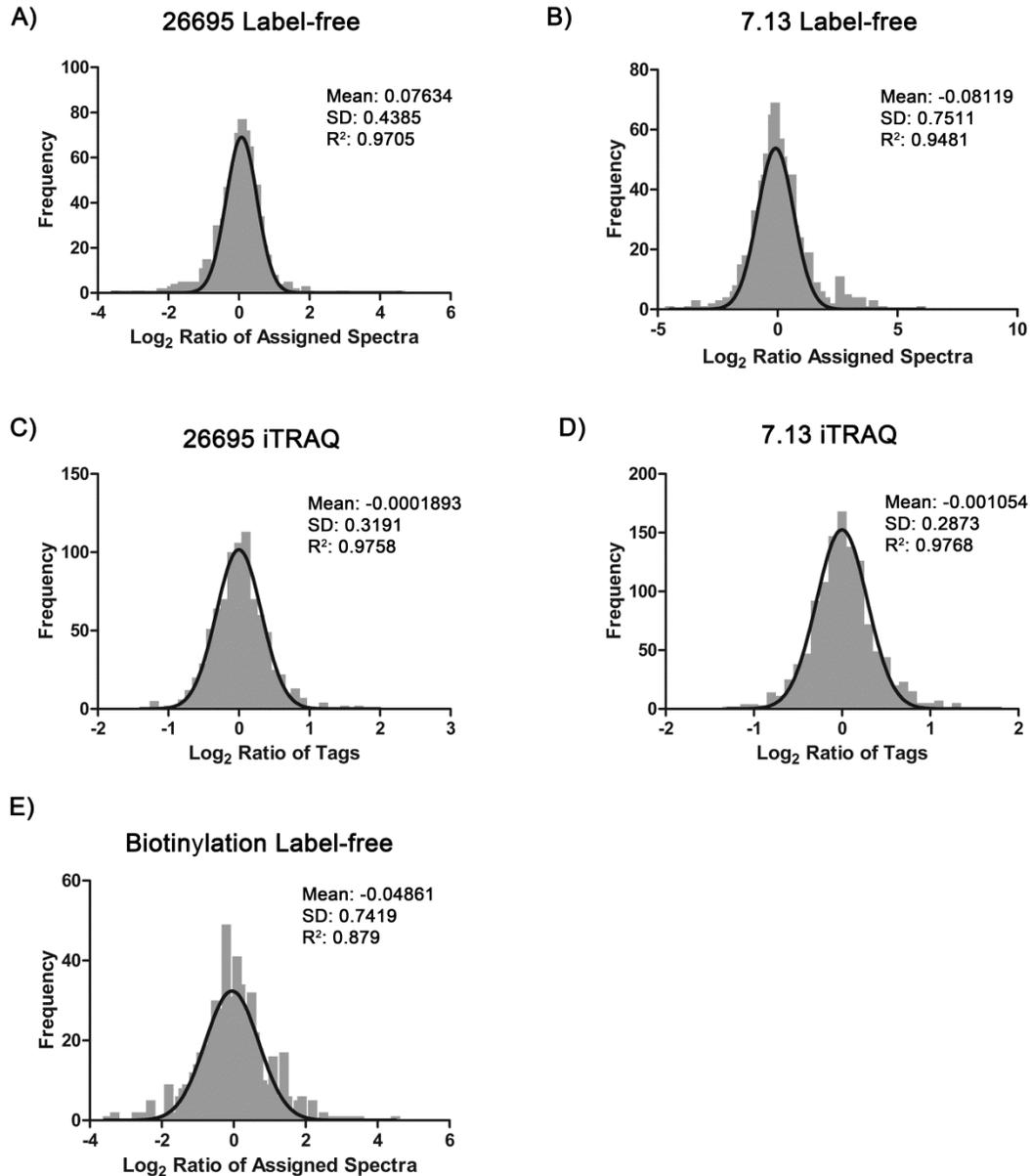


Figure 14. Distribution of R_{sc} or iTRAQ \log_2 ratios (protein abundance in bacteria from high salt conditions compared to bacteria from low salt conditions). A) 26695 label-free analysis ($n = 826$), B) 7.13 label-free analysis ($n = 1,091$), C) 26695 iTRAQ ($n = 849$), D) 7.13 iTRAQ ($n = 1,178$). and E) biotinylated preparations ($n = 643$). Curves were generated by non-linear regression analysis of R_{sc} spectral counting ratios or iTRAQ \log_2 ratios, and the data were grouped with bin values of 0.1.

H. pylori motility assay

H. pylori motility was analyzed as described previously (114). Briefly, overnight broth cultures of *H. pylori* 7.13 were back-diluted to an optical density at 600 nm of 0.1. 1- μ l aliquots of the bacterial suspensions were then stabbed into Brucella soft agar plates containing 5% FBS, 0.35% agar, and varying concentrations of sodium chloride. Diameters of bacterial halos were measured following a 5 day incubation of the plates at 37°C in a microaerobic atmosphere generated by GasPak EZ Campy packs (BD). At least 12 replicates were assayed for each condition.

Results

Salt-responsive proteins in strain 26695 identified by label-free and iTRAQ analyses

As an initial approach for identifying *H. pylori* proteins that differ in abundance depending on the sodium chloride concentration of the culture medium, I conducted experiments with the well-characterized *H. pylori* reference strain 26695. I cultured strain 26695 in either low-salt medium or high-salt medium, and generated subcellular fractions enriched in bacterial membrane proteins. I then analyzed the preparations by label-free MudPIT. Among 806 *H. pylori* proteins identified in the preparations, there was a significant difference in the number of assigned spectra for 66 proteins when comparing preparations from bacteria grown under low-salt and high-salt conditions (136). Thirty-eight proteins were more abundant in preparations from bacteria grown in low salt conditions, and twenty-eight proteins were more abundant in preparations from bacteria grown in high salt conditions (136).

As another approach to identify salt-responsive proteins, I used iTRAQ methods to analyze the membrane preparations from strain 26695, grown under high and low salt conditions, and identified 849 *H. pylori* proteins in the preparations. By using Z-score statistical analysis, I identified 22 of these proteins that were significantly different in abundance depending on the salt concentration (Table 8). Nine proteins were more abundant in bacteria grown under low salt conditions. These included the vacuolating toxin VacA, a VacA-like protein (FaaA), an outer membrane protein (FecA3) and multiple hypothetical proteins. Two proteins identified in this analysis (LysA and a hypothetical protein) are encoded by adjacent genes (HP0290 and HP0291) that are

predicted to be members of an operon. Thirteen proteins were more abundant in bacteria grown under high salt conditions (Table 8). These included CagA, flagellar hook protein FlgE2 (HP0908), a fibronectin domain containing protein (HP0746), an amidase (AmiE), a sodium and chloride-dependent transporter (HP0497), a conserved hypothetical secreted protein (HP1464), and multiple other hypothetical proteins. A protein product of HP0964, previously annotated as a pseudogene, was also identified as more abundant in bacteria grown under high salt conditions (Table 8).

Table 8: Salt-responsive proteins in strain 26695 identified by iTRAQ analysis

Proteins more abundant in low salt compared to high salt

Accession	Uniprot	KEGG	Fold Difference	Description
HP1400* [#]	O25950	Transport	2.53	iron(III) dicitrate transport protein (FecA3)
HP0664* ^{&}	K4NED0	NA	2.36	hypothetical protein
HP0887*	P55981	Pathogenesis	2.36	vacuolating cytotoxin
HP0802	O08315	riboflavin biosynthetic process	2.33	GTP cyclohydrolase II (RibA)
HP0290*	P56129	lysine biosynthetic process	2.30	diaminopimelate decarboxylase (dap decarboxylase) (LysA)
HP1532	O26060	glutamine metabolic process	2.27	glucosamine fructose-6-phosphate aminotransferase (isomerizing) (GlmS)
HP0721	O25423	NA	2.08	hypothetical protein
HP0291	O25064	chorismate metabolic process	2.02	hypothetical protein
HP0610* [#]	O25331	NA	2.01	toxin-like outer membrane protein (FaaA)

Proteins more abundant in high salt compared to low salt

Accession	Uniprot	KEGG	Fold Difference	Description
HP0964*	O25617	NA	3.86	hypothetical protein
HP0574	O25298	pentose-phosphate shunt	3.38	ribose 5-phosphate isomerase B (RpiB)
HP0547*	P55980	toxin transport	3.16	cag pathogenicity island protein (CagA)
HP1584	P55996	proteolysis	3.15	sialoglycoprotease (Gcp)
HP1248	P56123	nucleic acid phosphodiester bond hydrolysis	2.92	virulence associated protein homolog (VacB)
HP0294*	O25067	nitrogen compound metabolic process	2.76	aliphatic amidase (AmiE)
HP1464*	O26000	NA	2.58	conserved hypothetical secreted protein
HP0908 [!]	O25566	NA	2.37	flagellar hook (FlgE2)
HP0497	O25239	transmembrane transport	2.37	sodium- and chloride-dependent transporter
HP1588* ^{&}	O26107	NA	2.32	conserved hypothetical protein
HP0337	O25104	NA	2.23	hypothetical protein
HP1124	O25749	NA	2.13	hypothetical protein
HP0746* [#]	O25442	NA	1.97	fibronectin domain-containing protein

* Also significant in label-free analysis

Also significant in biotinylation experiments

& Significant in iTRAQ and label-free analyses of strain 26695, but not significant in either analysis of strain 7.13

[!] Strain 26695 produces two distinct FlgE proteins, FlgE1 (HP0870) and FlgE2 (HP0908).

NA: not assigned

Eleven (50%) of the 22 proteins identified as salt responsive in the iTRAQ experiment were also identified as salt responsive in the label-free MudPIT experiment (Table 8 and Figure 15A). Six of the eleven proteins were more abundant in bacteria grown under high salt conditions -- CagA, AmiE, fibronectin-domain containing protein (HP0476), a product of the annotated pseudogene HP0964, and two hypothetical proteins (HP1464 and HP1588). Five of the eleven proteins were more abundant in bacteria grown under low salt conditions -- LysA (HP0290), FaaA, HP0664 (hypothetical protein), VacA, and FecA3.

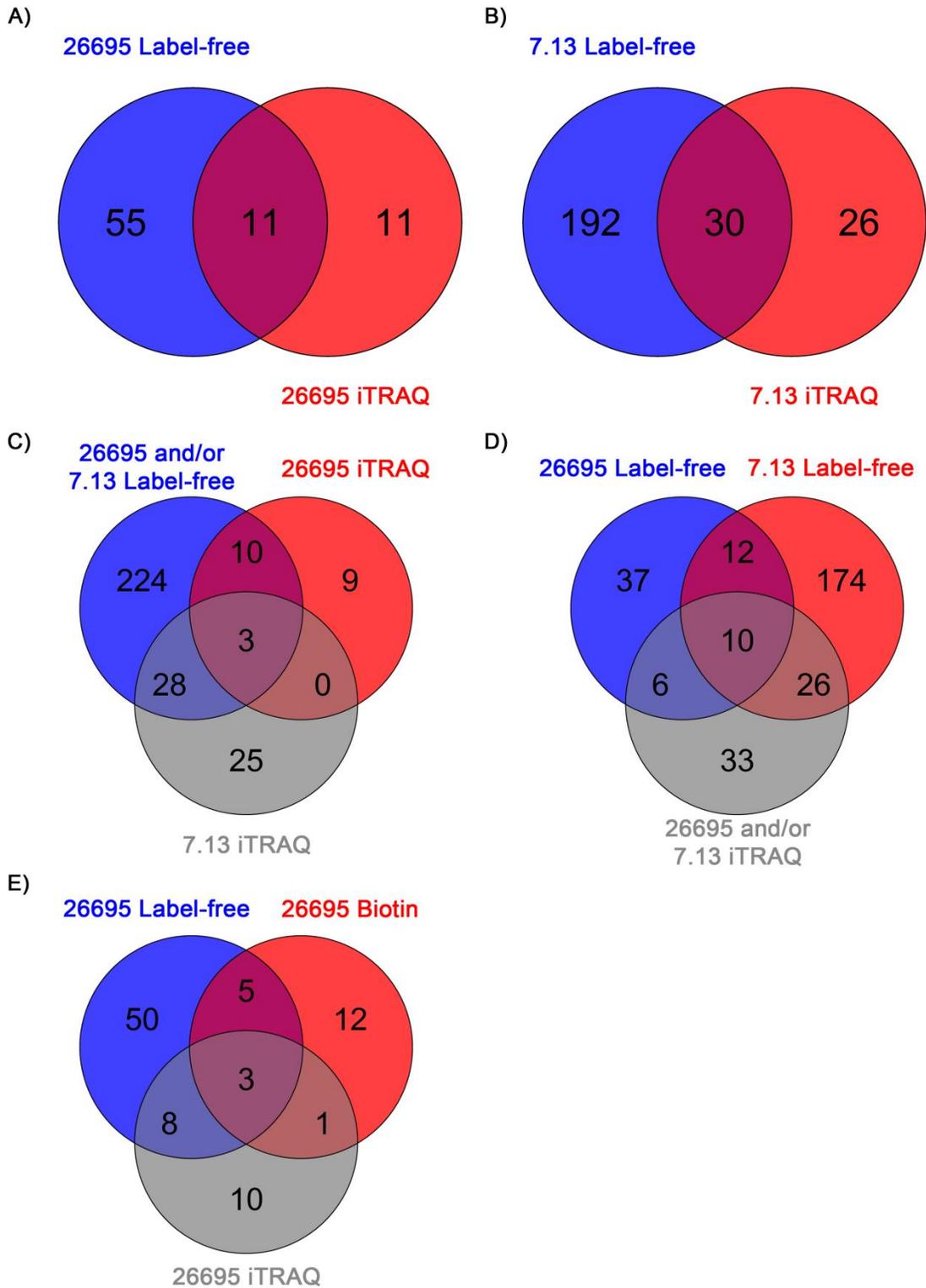


Figure 15. Identification of salt-responsive proteins by multiple methods. Venn diagrams showing: A) salt-responsive proteins in strain 26695 detected by label free

methods compared to iTRAQ methods, B) salt-responsive proteins in strain 7.13 detected by label-free methods compared to iTRAQ methods, C) the numbers of salt-responsive proteins identified by iTRAQ analysis of the indicated strains. “26695 and/or 7.13 label-free” indicates salt-responsive proteins identified by label-free analysis of at least one strain. D) Salt-responsive proteins identified by label-free analysis of the indicated strains. “26695 and/or 7.13 iTRAQ” indicates salt-responsive proteins identified by iTRAQ analysis of at least one strain. E) Salt-responsive proteins in strain 26695 identified by the three indicated methods.

As another approach to compare the results obtained by the label-free and iTRAQ methods, I analyzed the 66 salt-responsive proteins identified in the label-free experiments, and evaluated if the directionality of changes in protein abundance detected in the label-free experiments matched the directionality of changes in protein abundance detected in the iTRAQ experiments (Figure 16A, 16B). In the label-free experiments, 28 proteins were upregulated under high salt conditions and 38 were upregulated under low salt conditions. Similar salt-induced changes were observed in the iTRAQ experiments ($p < 0.001$, Student's t-test, Figure 16B). The same patterns were observed even when excluding the proteins that were identified by Z-score analysis of the iTRAQ data (Table 8, Figure 16B). These findings illustrate that the two different methods yielded similar patterns of results, and suggest that the changes in protein abundance were not limited to the relatively small set of salt-responsive proteins identified in the statistical analysis of the iTRAQ data.

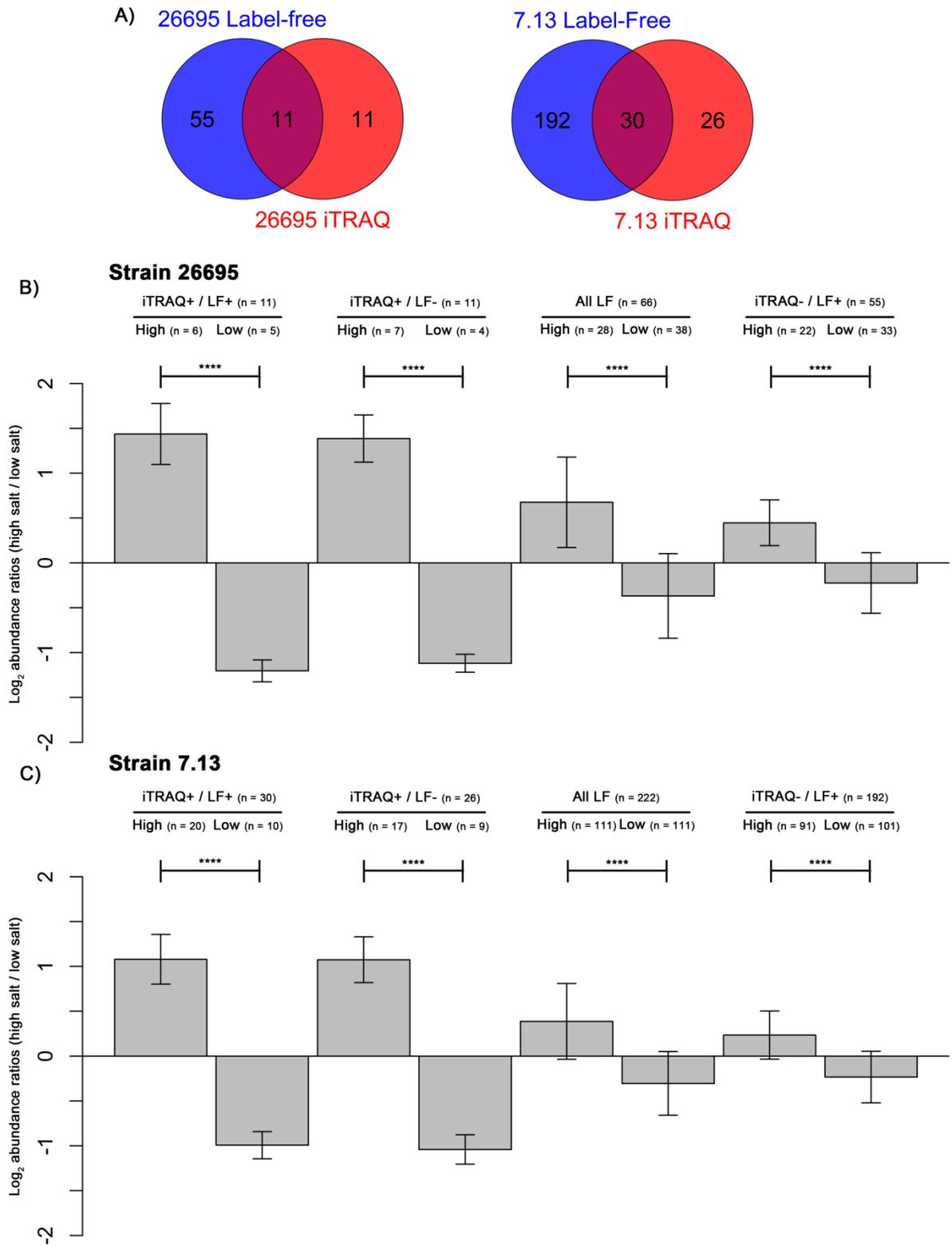


Figure 16. Comparison of iTRAQ results with label-free spectral counting results.
 A) Venn diagrams illustrating the number of salt-responsive proteins identified in strain

26695 or strain 7.13 by label-free and iTRAQ methodologies. B) iTRAQ log₂ ratios (mean + SD) for multiple sets of proteins from strain 26695, corresponding to different segments of the Venn diagram in panel A. C) iTRAQ log₂ ratios for multiple sets of proteins from strain 7.13, corresponding to different segments of the Venn diagram in panel A. The y-axis shows mean values for the indicated proteins (abundance in high salt conditions compared to low salt conditions), based on results of the iTRAQ experiments. iTRAQ+/LF+ indicates salt-responsive proteins identified by both iTRAQ and label-free approaches. "All LF" indicates all salt-responsive proteins identified by label-free methods, without considering whether or not they were identified by iTRAQ. iTRAQ-/LF+ indicates salt-responsive proteins identified by label-free methods but not iTRAQ. "High" indicates the group of proteins that exhibited increased abundance under high salt conditions compared to low salt conditions in the label-free experiments, and "low" indicates the group of proteins that exhibited decreased abundance under high salt conditions compared to low salt conditions in the label-free experiments. iTRAQ+/LF- indicates a control group (salt-responsive proteins identified by iTRAQ but not label-free methods); in this case, "high" and "low" correspond to categories based on the iTRAQ data. ****, Student's *t*-test $p < 0.001$. LF, Label-free.

Salt-responsive proteins in strain 7.13 identified by label-free and iTRAQ analyses

I next investigated the effects of changes in environmental salt concentration on protein abundance in *H. pylori* strain 7.13. In contrast to strain 26695, which lacks flagella and fails to colonize animal models, strain 7.13 produces flagella and can promote the development of gastric cancer in a Mongolian gerbil model (40, 96). I cultured strain 7.13 in either low-salt medium or high-salt medium, prepared membrane fractions from the bacteria and analyzed the preparations by label-free MudPIT. A total of 1054 *H. pylori* proteins were identified in these preparations, and 222 of these proteins differed in abundance, depending on the salt concentration of the medium (136). Among the 222 proteins, 111 were more abundant in bacteria grown under low salt conditions, and 111 other proteins were more abundant in bacteria grown under high salt conditions (136). Next I analyzed membrane preparations of strain 7.13 grown under high and low salt conditions using iTRAQ. Among 1185 *H. pylori* proteins detected, I identified 56 proteins that were significantly different in abundance (Table 9). Nineteen proteins were more abundant in bacteria grown under low salt conditions (Table 9). These included two flagellar proteins (FlgE and FlaG), an iron (III) dicitrate transport protein (FecA3), and two subunits of quinone-reactive Ni-Fe hydrogenase. Thirty nine proteins were more abundant in bacteria grown under high salt conditions (Table 9). These included CagA, CagT, a metalloprotease (HP0506), and several outer membrane proteins (HcrK, HopA, and HopD).

Table 9: Salt-responsive proteins in strain 7.13 identified by iTRAQ analysis

Proteins more abundant in low salt compared to high salt

Accession	Uniprot	KEGG	Fold difference	Description	26695 Homolog Locus Tag
HPB8_1702*	D7FGF2	Transport	2.45	Putative iron(III) dicitrate transport protein (FecA3)	HP1400
HPB8_1645	D7FG95	metabolic process	2.43	Membrane-bound lytic murein transglycosylase D	HP1572
HPB8_831	D7FDY1	oxidation-reduction process	2.32	Quinone-reactive Ni/Fe hydrogenase, small subunit (HydA)	HP0631
HPB8_1116	D7FER6	NA	2.26	Putative uncharacterized protein	HP0467
HPB8_258* ^{&}	D7FCA8	heme binding	2.18	Cytochrome c553	HP1227
HPB8_625	D7FDC5	cellular aromatic compound metabolic process	2.15	4-oxalocrotonate tautomerase (Dmpl)	HP0924
HPB8_829*	D7FDX9	NA	2.1	Modulator of drug activity B (MDA66)	HP0630
HPB8_958* ^{&}	D7FEA8	NA	2.07	Flagellar protein FlaG	HP0751
HPB8_844*	D7FDZ4	oxidation-reduction process	2.05	NAD(P)H-flavin oxidoreductase	HP0642
HPB8_822	D7FDX2	cell wall organization	2.02	UDP-N-acetylmuramate--L-alanine ligase	HP0623
HPB8_188	D7FC38	NA	2.00	Putative uncharacterized protein	NA
HPB8_832* ^{&}	D7FDY2	oxidation-reduction process	1.98	Quinone-reactive Ni/Fe hydrogenase, large subunit (HydB)	HP0632
HPB8_176* ^{&}	D7FC26	translation	1.87	50S ribosomal protein L18	HP1303
HPB8_153	D7FC03	NA	1.83	Putative uncharacterized protein	HP1326
HPB8_610	D7FDB0	NA	1.83	Putative uncharacterized protein	HP0938
HPB8_1264* ^{&}	D7FF65	transmembrane transport	1.8	Dipeptide ABC transporter, periplasmic dipeptide-binding protein (DppA)	HP0298
HPB8_195	D7FC45	NA	1.78	Putative uncharacterized protein	HP1286
HPB8_1079* [!]	D7FEM9	bacterial-type flagellum-dependent cell motility	1.77	Flagellar hook protein FlgE3	HP0870
HPB8_2*	D7FBK2	NA	1.76	Putative uncharacterized protein	HP1527

Proteins more abundant in high salt compared to low salt

Accession	Uniprot	KEGG	Fold difference	Description	26695 Homolog Locus Tag
HPB8_475* ^{&}	D7FCX5#	NA	3.47	Putative uncharacterized protein	HP0423 HP1412
HPB8_1569	D7FG19#	NA	3.47	Putative uncharacterized protein	NA
HPB8_1195*	D7FEZ5	NA	3.07	Putative uncharacterized protein	HP1076
HPB8_524	D7FD24%	NA	2.56	Putative uncharacterized protein	NA
HPB8_555	D7FD55%	NA	2.56	Putative uncharacterized protein	HP1405
HPB8_1336* ^{&}	D7FFD6	NA	2.53	Outer membrane protein HopA	HP0229
HPB8_1699	D7FGE9	NA	2.53	Putative uncharacterized protein	HP1397
HPB8_957* ^{&}	D7FEA7	NA	2.48	Putative uncharacterized protein	HP0750
HPB8_1043	D7FEJ3	NA	2.43	Putative uncharacterized protein	HP0833
HPB8_385*	D7FCN5	NA	2.4	cysteine-rich protein X HcpX	HP1117
HPB8_513*	D7FD13	NA	2.23	Putative uncharacterized protein	NA
HPB8_877*	D7FE27	NA	2.21	Putative uncharacterized protein	HP0674

HPB8_p0002	D7GAX3	NA	2.2	Putative uncharacterized protein	NA
HPB8_641*	D7FDE1	DNA duplex unwinding	2.17	Rep helicase, single-stranded DNA-dependent ATPase (Rep)	HP0911
HPB8_1582	D7FG32	polysaccharide metabolic process	2.17	Mannose-1-phosphate guanylyltransferase	HP0043
HPB8_622	D7FDC2	proteolysis	2.16	Protease HtpX homolog	HP0927
HPB8_1057*	D7FEK7	DNA modification	2.15	Type I restriction enzyme, S subunit	HP0790
HPB8_1600* ^{&}	D7FG51	NA	2.12	Outer membrane protein HopD	HP0025
HPB8_872	D7FE22	metabolic process	2.04	NAD-dependent deacetylase	NA
HPB8_28	D7FBM8	NA	1.99	Outer membrane protein HorK	HP1501
HPB8_711*	D7FDL1	NA	1.98	Cag pathogenicity island protein T	HP0532
HPB8_690	D7FDJ0	metabolic process	1.97	Uncharacterized metalloprotease yebA	HP0506
HPB8_89	D7FBT9	translation	1.95	50S ribosomal protein L34	HP1447
HPB8_741*	D7FDP1	toxin transport	1.92	Cytotoxin-associated protein A	HP0547
HPB8_998	D7FEE8	DNA modification	1.89	Type I restriction enzyme, S subunit	HP0790
HPB8_1204	D7FF04	NA	1.87	Putative uncharacterized protein	HP1085
HPB8_794*	D7FDU4	NA	1.84	Putative uncharacterized protein	HP0596
HPB8_1567* ^{&}	D7FG17	NA	1.82	Putative uncharacterized protein	HP1411 HP0424
HPB8_865	D7FE15	mRNA processing	1.8	Ribonuclease 3	HP0662
HPB8_537	D7FD37	DNA topological change	1.79	DNA topoisomerase I	HP0440
HPB8_897	D7FE47	metabolic process	1.79	Acetone carboxylase, alpha subunit	HP0696
HPB8_953*	D7FEA3	NA	1.79	Putative uncharacterized protein	HP0746
HPB8_1547* ^{&}	D7FFZ7	NA	1.79	Putative uncharacterized protein comL	HP1378
HPB8_319*	D7FCG9	NA	1.78	Outer membrane protein HopQ	HP1177
HPB8_729	D7FDN0	NA	1.78	Putative uncharacterized protein	HP0513
HPB8_1047*	D7FEJ7	NA	1.78	Putative uncharacterized protein	HP0836 HP0837
HPB8_972*	D7FEC2	NA	1.77	Putative uncharacterized protein	HP0764
HPB8_24	D7FBM4	oxidation-reduction process	1.75	Riboflavin biosynthesis protein (RibG)	HP1505
HPB8_375*	D7FCM5	cell outer membrane	1.75	Peptidoglycan-associated lipoprotein	HP1125

* Also significant in label-free analysis

[&] Significant in iTRAQ and label-free analyses of strain 7.13, but not significant in either analysis of strain 26695

[!] Strain 7.13 produces two distinct FlgE proteins, FlgE1 and FlgE3.

NA: not assigned

Among the 56 proteins produced by strain 7.13 that were identified as differentially abundant in iTRAQ experiments, 30 (54%) were identified as differentially abundant by both iTRAQ and label-free MudPIT (Table 9 and Figure 15B). Among these 30 proteins, 10 were enriched in bacteria grown under low salt conditions; these included the outer membrane protein FecA3 and flagellar hook protein FlgE3. The set of 20 proteins enriched in bacteria grown under high salt conditions included CagA, HopQ, CagT, HopD, HopA, and ComL.

I analyzed the 222 salt-responsive proteins from strain 7.13 that were identified in the label-free experiments, and evaluated if the directionality of changes in protein abundance detected in the label-free experiments matched the directionality of changes in protein abundance detected in the iTRAQ experiments (Figure 16A, 16C). In the label-free experiments, 111 proteins were upregulated under high salt conditions and 111 were upregulated under low salt conditions. Similar salt-induced changes were observed in the iTRAQ experiments ($p < 0.001$, Student's t-test, Figure 16C). The same patterns were observed even when excluding the proteins that were identified by Z-score analysis of the iTRAQ data (Table 9, Figure 16C). Thus, similar to the conclusions reached in studies of strain 26695, this analysis of the 7.13 data indicates that the two methodologic approaches yielded similar patterns of results, and suggest that the changes in protein abundance were not limited to the relatively small set of salt-responsive proteins identified in the statistical analysis of the iTRAQ data.

Salt-responsive proteins identified in both strains 26695 and 7.13

The analysis of two different *H. pylori* strains in the experiments described above provided us with an opportunity to detect salt-responsive changes in protein abundance conserved in multiple strains. I identified 22 proteins that were salt-responsive in both strain 26695 and 7.13, based on a label-free spectral-counting approach (136). By using iTRAQ, I identified 3 proteins that were salt-responsive in both strains (CagA, FecA3, and fibronectin domain protein) (Figure 15C) (136). Among the 22 proteins identified as salt-responsive in both 26695 and 7.13 based on label-free spectral counting experiments, 10 were identified as salt-responsive in at least one of the two iTRAQ analyses (Figure 15D). These included FaaA, FecA3, CagA, fibronectin-domain containing protein HP0746, HopQ, and HP0596 (corresponding to tumor necrosis factor alpha-inducing protein). Forty one proteins were identified as salt-responsive by at least one of the label-free analyses and also by at least one of the iTRAQ analyses (Figure 15C).

In addition to the salt-responsive alterations detected in both *H. pylori* strains, I also detected strain-specific responses to changes in salt concentrations (i.e., proteins that were identified as significantly different in abundance based on both label-free and iTRAQ methods for one strain, but not identified as salt-responsive by either method in the other strain) (Tables 8 and 9). For example, the outer membrane proteins HopA and HopD, flagellar protein FlaG, and the competence protein ComL were salt-responsive in strain 7.13 but not in strain 26695. In total I identified 13 proteins as salt-responsive in one strain but not the other (11 proteins from 7.13 and 2 proteins from 26695) (Tables 8 and 9). Additionally, I detected strain-specific differences in the directionality of salt-

induced changes. AmiE was significantly more abundant in strain 26695 grown under high salt conditions (based on label-free spectral counting and iTRAQ), whereas AmiE was significantly more abundant in strain 7.13 grown under low salt conditions (based on label-free spectral counting). I also identified strain-specific differences in effects of salt on CheA and Omp8 (HopG), based on a label-free spectral counting approach (136).

Changes in abundance of surface-exposed proteins in response to salt concentration

The label-free and iTRAQ experiments were performed by analyzing total membrane fractions, which contain both inner and outer membrane proteins. I hypothesized that, in comparison to analyses of total membrane fractions, low-abundance salt-responsive proteins present on the bacterial surface might be identified more efficiently by analyzing preparations enriched for surface-exposed proteins. Therefore, I grew *H. pylori* in either high salt or low salt concentrations as before, labeled surface-exposed proteins of intact bacteria with amine-reactive, cell-impermeable biotin, and purified the biotinylated proteins. I first analyzed the preparations of purified biotinylated proteins compared to mock preparations from control bacteria (not labeled with biotin and grown under normal salt conditions), using label-free MudPIT. Among 639 *H. pylori* proteins detected in these preparations, I identified 146 as putative surface-exposed proteins (136), many of which were also identified by the same approach in a previous study (135). I next compared the biotinylated preparations from bacteria grown in high salt conditions with biotinylated

preparations from bacteria grown in low salt conditions, and identified 21 of the 146 putative surface-exposed proteins as differentially abundant (Table 10). Most of these (15 of 21) were more abundant in the preparations from bacteria grown under low salt conditions than in the preparations from bacteria grown under high salt conditions. These included VacA-like proteins ImaA (HP0289) and FaaA (HP0609/10), BamA, ComH, FecA3, several annotated OMPs (AlpB, HorB, and HP0358), two conserved hypothetical secreted proteins (HP0235 and HP1117), and several hypothetical proteins (Table 10). Proteins identified as enriched in the high salt preparations compared to low salt preparations included HP1124 (a hypothetical protein that may be an outer membrane channel, based on BLAST conserved domain search), and fibronectin type 3 domain protein (HP0746) (Table 10).

Table 10. Salt-responsive proteins in strain 26695 identified by analysis of biotinylated intact bacteria

Proteins more abundant in low salt compared to high salt

Accession	Assigned Spectra			Fold difference			Uniprot	KEGG	Description
	Biotinylated High Salt	Biotinylated Low Salt	Unlabeled Normal Salt Control	Low vs. High	High vs. Control	Low vs. Control			
HP0289	3	20	0	6.58	2.16	13.33	O25063	Protein Binding	toxin-like outer membrane protein (ImaA)
HP0235	8	25	0	3.37	5.25	16.57	O25021	metabolic process	conserved hypothetical secreted protein
HP1117	55	131	5	2.67	3.09	7.7	O25742	NA	conserved hypothetical secreted protein
HP0609*	62	141	18	2.55	1.03	2.46	O25330	protein binding	toxin-like outer membrane protein (FaaA)
HP0913	142	291	25	2.31	1.69	3.63	O25571	NA	outer membrane protein (omp21)
HP0610*#	153	307	30	2.26	1.52	3.2	O25331	NA	toxin-like outer membrane protein (FaaA)
HP0645*	55	97	2	1.98	6.82	12.6	O25362	NA	soluble lytic murein transglycosylase (Slt)
HP1400*#	140	236	9	1.9	4.49	7.95	O25950	transport	iron(III) dicitrate transport protein (FecA3)
HP0358*	155	260	4	1.89	10.47	18.45	O25125	NA	hypothetical protein
HP1527*	192	314	7	1.84	7.74	13.31	O26055	NA	ComH
HP1333	51	83	1	1.82	10.55	18.01	O25891	NA	hypothetical protein
HP0655	163	256	12	1.77	3.96	6.55	O25369	protein insertion into membrane	BamA
HP0252	82	123	1	1.68	16.85	26.55	O25034	NA	outer membrane protein (omp7)
HP1173	163	233	4	1.61	11.04	16.59	O25787	NA	hypothetical protein
HP0127	161	218	15	1.52	3.17	4.51	O24941	NA	outer membrane protein (omp4)

Proteins more abundant in high salt compared to low salt

Accession	Assigned Spectra			Fold difference			Uniprot	KEGG	Description
	Biotinylated High Salt	Biotinylated Low Salt	Unlabeled Normal Salt Control	Low vs. High	High vs. Control	Low vs. Control			
HP1376	16	2	0	5.88	10.22	1.63	O25928	lipid A biosynthetic process	(3R)-hydroxymyristoyl-(acyl carrier protein) dehydratase (FabZ)

HP1187	36	13	0	2.41	22.59	8.79	O25799	NA	hypothetical protein
HP1561	68	32	3	1.88	6.05	3.02	O26082	NA	iron(III) ABC transporter, periplasmic iron-binding protein (CeuE)
HP1124#	608	326	10	1.67	17.68	9.99	O25749	NA	hypothetical protein
HP0746*#	238	140	1	1.52	48.88	30.31	O25442	NA	fibronectin type 3 domain protein
HP0103*	91	53	1	1.52	18.83	11.59	O24929	signal transduction	methyl-accepting chemotaxis protein (tlpB)

* Also significant in label-free analysis of 26695 total membrane preparations

Also significant in 26695 iTRAQ experiment

Three proteins produced by strain 26695 (FecA3, fibronectin-domain protein HP0746, and FaaA) were identified as salt-responsive based on all three of the experimental approaches (label-free MudPIT, iTRAQ, and biotinylation) (Figure 15E), and 14 proteins were identified as salt-responsive by two of the three methods (Figure 15E). In addition to the proteins listed previously (detected by both label-free methods and iTRAQ), these included FaaA (HP0609) and three hypothetical proteins (HP0358, HP1527, and HP1124).

Salt-dependent changes in bacterial motility

Label-free and iTRAQ analysis of strain 7.13 revealed that the abundance of multiple proteins associated with flagellar functions was reduced when the bacteria were cultured in high salt medium (Table 11). I hypothesized that these shifts in protein abundance would have an effect on bacterial motility. To test this hypothesis, we analyzed the motility of *H. pylori* 7.13 in soft agar containing a range of sodium chloride concentrations. We observed a significant decrease in bacterial motility in response to high salt conditions, and the effect of salt concentration on motility was dose-dependent (Figure 17).

Table 11. Salt-responsive flagellar proteins from strain 7.13

Log₂ R_{sc} *	iTRAQ ratio**	Protein
-1.75	-0.67	Flagellar hook capping protein FlgD
-1.57	-0.46	Flagellar hook protein FlgE1
-1.31	-0.82	Flagellar hook protein FlgE3
-1.75	-0.76	Flagellar hook-associated protein 3 FlgL
-1.52	-1.05	Flagellar protein FlaG
-1.02	-0.77	Flagellin A
-2.16	-0.55	Flagellin FliC

* Calculated using the formula: $\log_2 \frac{H+0.5}{L+0.5} + \log_2 \frac{T_L - L + 0.5}{T_H - H + 0.5}$. Where H = high salt assigned spectra, L = low salt assigned spectra, T_H = total high salt spectra, T_L = total low salt spectra

** log₂ ratio of high salt tag intensity over low salt tag intensity

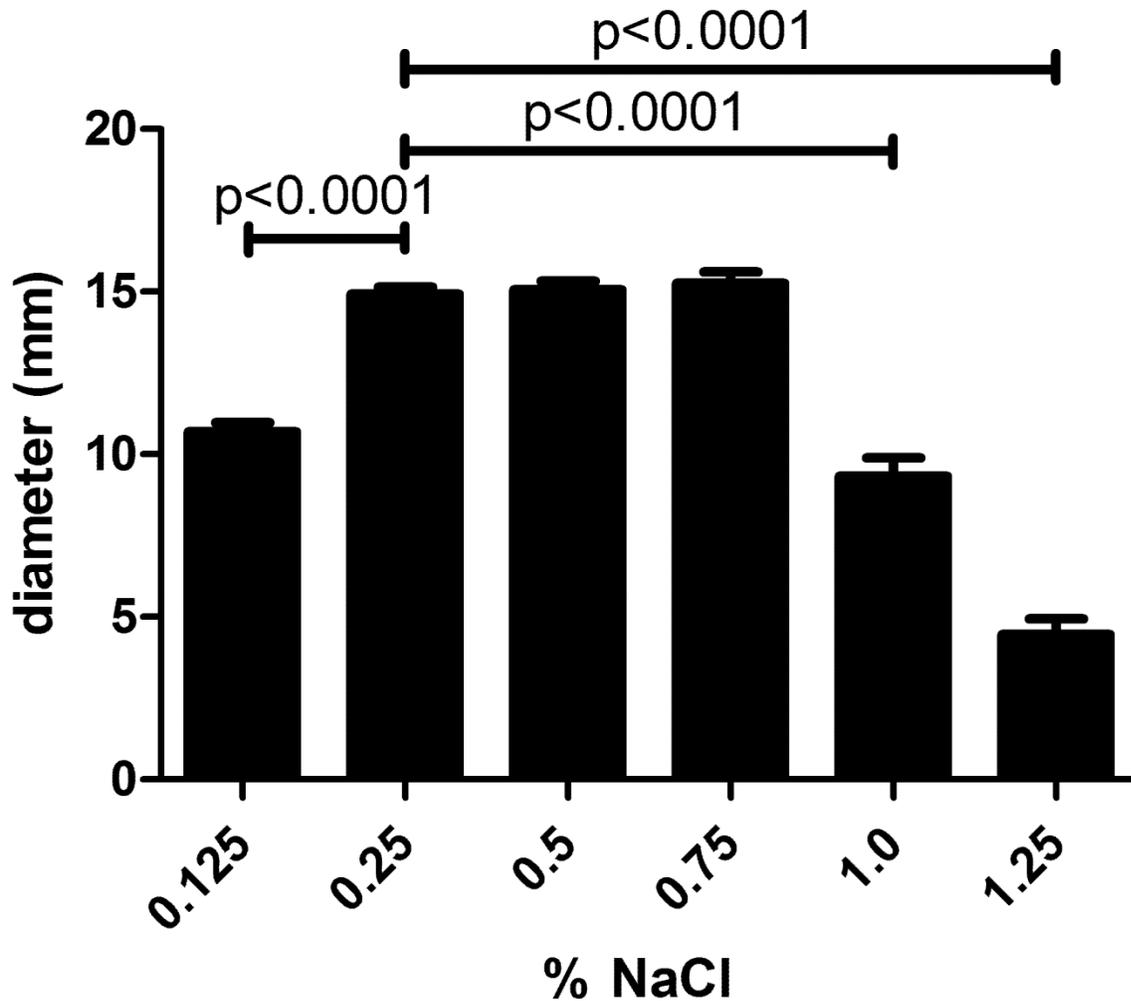


Figure 17. Effect of salt concentration on *H. pylori* motility. *H. pylori* strain 7.13 was stabbed into Brucella soft agar plates containing 5% FBS, 0.35% agar, and the indicated concentrations of added sodium chloride. Diameters of bacterial halos were measured after incubation for 5 days. The indicated p values were calculated by Student's T test.

Discussion

In this study, I used both label-free and iTRAQ methods to identify *H. pylori* membrane proteins that change in abundance in response to alterations in environmental salt concentrations. In addition, I analyzed the responses of two different *H. pylori* strains to alterations in salt concentrations. These methods allowed us to identify salt-responsive proteins in a given strain by more than one proteomic approach, as well as proteins that were salt-responsive in more than one *H. pylori* strain.

When comparing the iTRAQ results to the label-free results for a particular strain, I was able to confirm approximately 50% of the iTRAQ results by label-free methods (Figure 15). Conversely, a lower proportion of the label-free results was confirmed by iTRAQ results. This pattern is attributable to the use of different criteria for identifying salt-responsive proteins in the iTRAQ experiments compared to the label-free experiments. The differences in iTRAQ and label-free results also reflect distinct strengths and limitations of the two approaches. For example, label-free spectral counting is optimal for detecting differences in abundance of proteins with many assigned spectra, whereas iTRAQ is able to quantify differences in the abundance of proteins with fewer assigned spectra.

In a previous study of strains 26695 and 7.13, 31 salt-responsive *H. pylori* proteins were identified by using 2D-DIGE (81). Several salt-responsive proteins identified in the previous 2D-DIGE analysis (for example, CagA, HopQ, FaaA, and HP1588) were also identified by multiple methods in the current study. Several other proteins designated as salt-responsive in the previous 2D-DIGE analysis (HP0014,

HP0221, HP0289, HP0377, HP0630, and HP0680) underwent similar salt-dependent changes in abundance in the current experiments, but the changes were not deemed to be statistically significant. Multiple factors account for the relatively small numbers of salt-responsive proteins identified in both the previous 2D-DIGE study and the current study. For example, the current study focused on membrane-associated proteins, whereas the 2D-DIGE was performed with whole cell lysate. Another important consideration is that 2D-DIGE is suboptimal for analysis of membrane proteins and low-abundance proteins. The methods utilized in this study permitted an analysis of such proteins with a higher level of sensitivity than was possible using 2D-DIGE. Thus, I were able to identify several salt-responsive proteins that were not previously demonstrated. For example, fibronectin domain-containing protein HP0746 and the outer membrane protein FecA3 were shown to be salt-responsive in the current study by multiple methods and in two different *H. pylori* strains. Currently very little is known about the function of HP0746. BLAST analysis indicates that this protein contains multiple fibronectin type 3 domains, suggesting that HP0746 may function as an adhesin or an extracellular glycohydrolase (43, 79). FecA3 is one of several annotated outer membrane iron dicitrate transporter proteins. The proteomic data indicate that both *H. pylori* strains analyzed in this study produce three distinct FecA proteins (FecA1, FecA2, and FecA3). FecA1 and FecA2 have been previously reported to be regulated by Fur (9, 134) and the current results indicate that FecA3 is regulated by environmental salt concentrations.

A comparative analysis of strains 26695 and 7.13 identified numerous proteins that were salt-responsive in both strains. Recapitulation of the findings in multiple strain

backgrounds provides added confidence in the designation of these proteins as salt-responsive. I also observed 14 examples of strain-specific responses (proteins that were identified as salt responsive in one strain based on both label-free and iTRAQ methods, but not identified as salt responsive in the other strain by either proteomic method). For example, HopA, HopD, FlaG, and ComL were salt-responsive in strain 7.13 but not in strain 26695. Strikingly, salt had opposite effects on AmiE in the two strains (more abundant in high salt conditions for strain 26695 and more abundant in low salt conditions for strain 7.13). These strain-specific differences might be attributable to the high level of genetic diversity that exists among *H. pylori* strains. For example, there might be variations in nucleotide sequence motifs recognized by salt-responsive regulators, or there might be variations in the regulatory systems that mediate salt-responsive changes. The observed strain-specific variation in results is consistent with previous reports of strain-specific responses to environmental stimuli in *H. pylori* (41).

Previous studies have shown that transcriptional alterations account for the changes in abundance of several *H. pylori* proteins in response to altered salt concentration (41, 81, 82). However, it seems possible that some of the observed changes may also be mediated by post-transcriptional changes (including changes in transcript stability or changes in protein stability). Analysis of promoter regions and untranslated regions of genes encoding the salt-responsive proteins identified in this study did not readily identify any conserved sequence motifs that might be mechanistically involved in the observed salt-responsive changes. Further studies will

be required to elucidate the molecular mechanisms underlying the observed salt-induced changes in abundance of individual proteins.

My analysis of strain 7.13 revealed that there was a lower abundance of multiple proteins associated with flagellar activity when the bacteria were grown in high salt conditions. Most of these changes were not detected with strain 26695, which does not produce flagella; flagellar hook protein FlgE2 (HP0908) was the only salt responsive protein associated with flagellar activity detected in the iTRAQ analysis of strain 26695. Concordant with the proteomic data for strain 7.13, we detected decreased bacterial motility of this strain under high salt conditions. Previous studies showed that the salt concentrations employed in these experiments have minimal effects on the growth rate of *H. pylori* (82), but it is possible that slightly impaired growth at high salt concentrations may contribute to the reductions in halo size observed in the soft agar assay. Changes in *H. pylori* motility in response to altered salt concentration could have important consequences for the bacteria. For example, if the bacteria are less motile, this may impair their ability to move within the mucus layer in response to pH or nutrient availability, which could potentially lead to alterations in bacterial load. Studies in mouse models have shown that there is typically a reciprocal relationship between *H. pylori* density and severity of the gastric inflammatory response (1), and therefore, alterations in the bacterial load may lead to changes in the gastric mucosal inflammatory response. In addition, decreased motility may facilitate increased bacterial adhesion to the host tissue. Similar to the results of the current study, decreased bacterial motility in response to high salt concentrations has also been observed in studies of other bacterial species (126).

Among the salt-responsive proteins identified in this study, several have previously been shown to have important roles in *H. pylori* colonization of the stomach. For example, FecA3 is required for *H. pylori* colonization of the mouse stomach (9), and a *faaA* mutant was outcompeted by a wild type strain (114). The impaired colonization of the *faaA* mutant may be attributable to flagellar abnormalities and a motility defect (114). The functions of most of the other salt-responsive proteins *in vivo* have not yet been studied in any detail.

In summary, these results provide new insight into the effect of environmental salt concentration on the *H. pylori* membrane proteome. In future studies, it will be important to analyze further the functional consequences of the observed changes in protein abundance. In addition to the observed changes in motility that occur in response to altered salt concentrations, I speculate that altered abundance of certain salt-responsive proteins may enhance the ability of *H. pylori* to survive salt stress. I further speculate that the changes in abundance of surface-exposed proteins that occur following exposure of *H. pylori* to high salt concentrations may lead to altered interactions of the bacteria with gastric epithelial cells and immune cells (for example, changes in bacterial adherence and changes in production of cytokines), which could influence the severity of the gastric inflammatory response. In future studies, it will also be important to analyze *H. pylori* strains isolated from patients who chronically consume a high salt diet and determine if these strains exhibit differences in proteomic profiles compared to isolates from patients who consume low salt diets. Ultimately, studies of *H. pylori* responses to altered salt concentrations may lead to a better understanding of

the mechanisms by which a high salt diet contributes to an increased risk of gastric adenocarcinoma.

CHAPTER IV

INVESTIGATION OF CAGT, A PUTATIVE VIRB7 FUNCTIONAL ANALOG

Introduction

As described in Chapter II, three proteins encoded by the *cag* PAI (Cag3, CagM, and CagT) were identified as surface-exposed. Chapter III identified CagT as being salt-responsive in strain 7.13. Colonization with *H. pylori* strains encoding the *cag* PAI is strongly associated with the development of gastric cancer (21). The *cag* PAI encodes multiple proteins that are components of a T4SS (Cag T4SS). IL-8 induction and extracellular filamentous structure (T4SS pilus) formation when *H. pylori* is cocultured with AGS gastric epithelial cells are two Cag T4SS-dependent phenotypes. Deletion of *cagT* abolishes these phenotypes (59). These data suggest CagT is an essential component of the Cag T4SS. The putative function of CagT is inferred by its annotation as a VirB7 homolog. However, this annotation is misleading due to the fact that CagT has almost no sequence similarity to VirB7. CagT is also much larger in terms of amino acid length than VirB7 (280 AA vs. 55 AA). The only sequence similarity CagT has with VirB7 is the presence of a 4 AA motif called a lipobox. A lipobox is a predictor of post-translational lipidation of a bacterial protein.

A bacterial lipoprotein is synthesized in several steps. First the newly synthesized protein is recognized and secreted via the Sec pathway. Then, before the signal sequence is cleaved, the lipidation synthesis machinery recognizes the lipobox. The first modification is the addition of a diacylglycerol on the side chain of the conserved

cysteine within the lipobox. Then the amino acids preceding the cysteine are cleaved via the signal peptidase II (SPII). Finally, a fatty acid is ligated to the amino terminus of the cysteine. In order to be localized to the outer membrane, the newly formed lipoprotein is ushered across the periplasm to the outer membrane. The canonical model of lipoprotein outer membrane localization (based on studies in *E. coli*) is via the localization of lipoproteins (Lol) pathway (70). If the amino acid immediately following the modified cysteine is an aspartic acid, this is called a Lol avoidance or inner membrane retention signal. The newly synthesized lipoprotein is recognized by the Lol machinery and localized to the outer membrane in the absence of the inner membrane retention signal. Despite the presence of conserved lipoprotein synthesis machinery in *H. pylori* (132), the model of lipoprotein localization must be viewed with caution because only two of the five Lol proteins are present in *H. pylori* based on sequence homology (77).

Surface localization of lipoproteins is a process not well understood for most bacteria, including *H. pylori* (70). In this thesis chapter, I describe investigation of CagT surface localization and CagT lipidation. I used a genetic approach to investigate the effect of lipidation of CagT on known Cag T4SS phenotypes and CagT localization. Finally, I expand on the finding from Chapter II that CagT is surface-exposed. To do this I used an approach involving intact-cell surface labeling of FLAG-tag insertions into CagT to investigate the surface-exposed topology of CagT.

Methods

Bacterial strains and growth conditions

H. pylori strain 26695 was grown at 37°C in ambient air containing 5% CO₂ on either trypticase soy agar plates supplemented with 5% sheep blood, or in sulfite-free Brucella broth supplemented with 10% fetal bovine serum (BB-FBS). When selecting for mutants, Brucella agar-10% FBS or 5% sheep blood trypticase soy agar plates were supplemented with 5 µg/mL chloramphenicol, 15 µg/mL metronidazole, or 10 µg/mL kanamycin. *Escherichia coli* strain MG1655 was grown at 37°C in ambient air on either LB plates or LB broth supplemented with 100 µg/mL ampicillin.

Bacterial strain mutagenesis

Marked *cagT* mutagenesis was performed by inserting a *cat/rdxA* cassette in *cagT* as described previously (59). For unmarked mutagenesis of *cagT* in *H. pylori*, the genomic region of *cagT* (HP0532) with 500b flanking sequence (HP0532+500) from strain 26695 was amplified using Amplitaq Gold (ThermoFisher) and ligated into pGEMT-Easy (Promega) using TA cloning (HP0532+500pGEMT). Using inverse PCR, *BamHI* restriction sites were inserted in the 5' and 3' locations of HP0532 (Vent polymerase, NEB). HP0532 was then excised using *BamHI* (ThermoFisher) and the vector fragment (containing the HP0532 5' and 3' 500b flanking sequences) was gel purified. This fragment was circularized using T4 ligase (ThermoFisher) and transformed into DH5α. This plasmid was then used to transform 26695Δ*rdxA*, *cagT::cat/rdxA*, and transformants resistant to metronidazole were selected to generate 26695Δ*rdxA*,Δ*cagT*. Point mutations and FLAG epitope insertions were generated using

the HP0532+500pGEMT plasmid with QuikChange Multi site-directed mutagenesis kit (Agilent) and transformed into 26695 Δ *rdxA*, *cagT::cat/rdxA*.

Wild type *cagT* and *cagT*_{C21S} were introduced in *E. coli* strain MG1655 by insertion in the pBAD vector and selection using ampicillin resistance. Recombinant CagT was expressed *E. coli* strains using arabinose induction.

Palmitic acid incorporation assay

E. coli was grown overnight in 5 mL LB cultures and another 5 mL culture was subcultured at a starting Abs600 of 0.1 until the density reached an Abs600 of 0.5. Then recombinant CagT was expressed using a final concentration of 0.2% arabinose (Sigma), and lipoproteins were labeled using 25 μ M palmitic acid alkyne (PA) (Cayman) and induced for 4 hours at 37°C. 1 mL of culture was collected and washed twice using 1 mL PBS. Bacteria were pelleted and frozen at -70°C until needed.

Bacterial pellets were resuspended in 200 μ L triethanolamine (TEA) with lysozyme and Omnicleave (1 μ L/mL). Bacteria were incubated on ice for 5 minutes and lysed using freeze-thaw (3x). 2 mM final concentration PMSF was then added to the lysate. Membranes were solubilized using a 0.5% final concentration of SDS. To detect labeled lipoproteins, Cyazine 7.5 azide (Cy 7.5) (Lumiprobe) was conjugated to the palmitic acid alkyne using a Click-iT labeling kit according to manufactures instructions (ThermoFisher). Labeled proteins were visualized using an Odyssey Imaging System (Licor).

Immunoblotting

Protein lysates were resolved using 10% TGX SDS-PAGE gels (BioRad) and proteins were transferred to nitrocellulose membranes. Membranes were blocked using either 2% milk PBS-Tween 20 (0.1%) or Odyssey blocking buffer (LI-COR) with 0.1% Tween 20. Proteins were detected by incubating the membrane with primary antisera (diluted 1:5,000-1:10,000), followed by either horseradish peroxidase-conjugated anti-rabbit IgG (Santa Cruz), horseradish peroxidase-conjugated anti-mouse IgG (Perkin Elmer), or IRDye 680RD anti-Rabbit (LI-COR) as the secondary antibody. CagT antiserum was preabsorbed to *E. coli* strain BL21(DE3) prior to immunoblotting. Labeled proteins were detected either by ECL methodology and visualized using X-ray film, or with immunofluorescence visualized with an Odyssey imaging system.

IL-8 induction assay

AGS cells were grown using RPMI with 25 mM HEPES and 10% FBS and cocultured with *H. pylori* grown overnight in BB-FBS with a MOI of 100. Cocultures were incubated for 4 hours at 37°C with 5% CO₂. Supernatant was collected and IL-8 induction was determined using IL-8 ELISA according to manufacturer specifications (Biolegend).

Bacterial subcellular fractionation

H. pylori was grown on trypticase soy agar plates for 24 hours and then grown for 20 hours in Brucella broth to late log phase, corresponding to an OD₆₀₀ of approximately 0.8. Bacteria were pelleted by centrifugation at 3,500 x g for 20 min at 4°C. The bacterial cells were then fractionated as described previously in Chapter II. After

fractionation, outer membrane pellets were solublized in RIPA lysis buffer with protease inhibitor cocktail (Roche) and boiled for 10 minutes. Insoluble debris was removed via spinning at 21,000xg for 10 minutes. Protein concentration was determined via BCA assay (Thermo).

Surface-exposed FLAG tag assay

The FLAG epitope (DYKDDDDK) was inserted in various locations throughout the length of CagT that were predicted to have low hydrophobicity and thus would be more likely to be solvent exposed (27K, 55K, 106K, 110Q, 115R, 145K, 210Y, and 260L) (Figure 18). *H. pylori* was grown in 25 mL BB-FBS overnight (Abs600 approximately 0.5-0.6). Approximately the same number of bacteria were collected and processed in parallel. Bacteria were pelleted and washed three times in PBS. 1.5×10^8 bacteria were added to a lysine-coated 96-well black-walled plate. For studies with permeabilized cells, EDTA was added to a final concentration of 10 mM and the bacteria were sonicated once at 25% maximum output for 10 seconds. Plates were spun to promote bacterial adherence. Bacteria were fixed using paraformaldehyde in PBS at a final concentration of 2% and incubated for 10 minutes at room temperature. Non-adherent bacteria were removed and fixation was quenched three times using TBS. Bacteria were blocked using Odyssey blocking reagent (PBS based, LI-COR) for 45 minutes at room temperature. Bacteria were labeled using anti-FLAG antibody (1:2,500. M2 clone, Sigma) for 2 hours. Excess antibody was removed by three washes with PBS with 5 minute incubations. The anti-FLAG was labeled using anti-mouse antibody (1:1,000, IRDye 800 CW, LI-COR) for one hour. Excess antibody was removed using

three washes with PBS with 5 minute incubations. Plates were visualized with an Odyssey imaging system. Signal intensity was measured using Image Studio Lite.

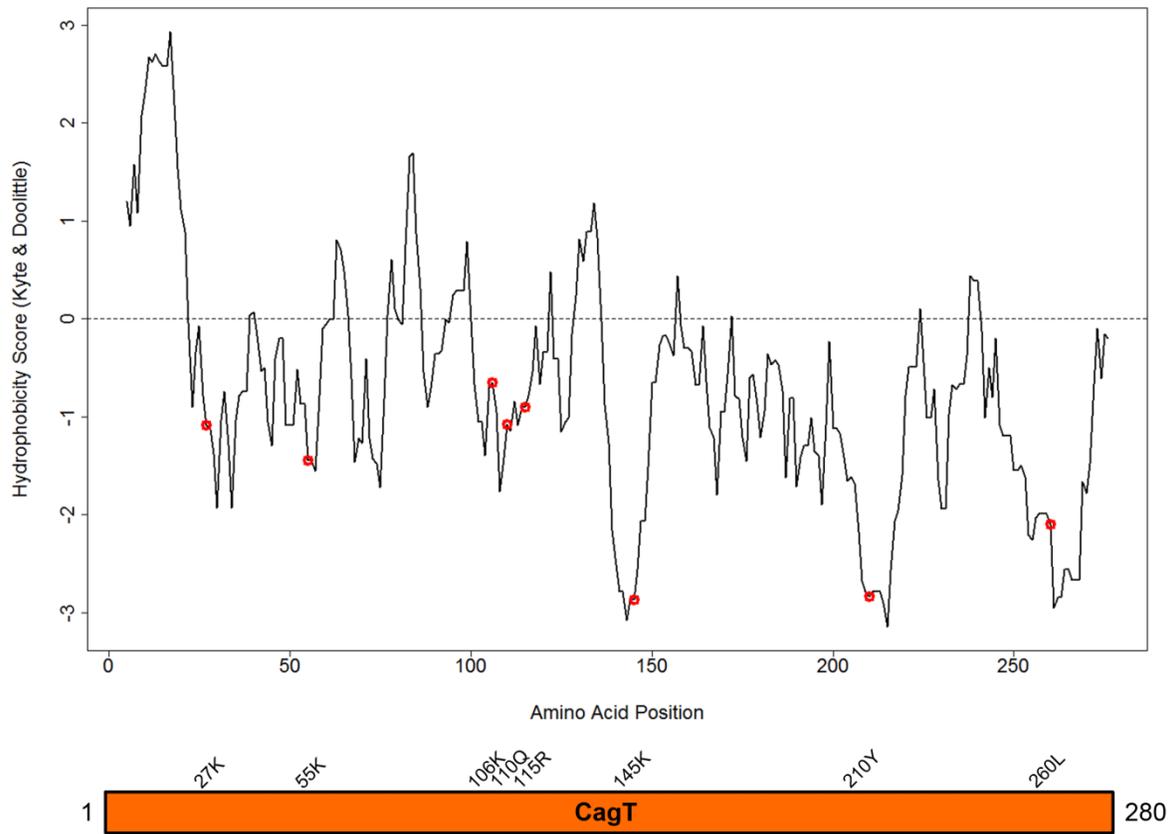


Figure 18. Sites for CagT FLAG epitope insertion in CagT. Hydrophobicity plot for the amino acid sequence of CagT. Red dots indicate positions chosen for FLAG epitope (DYKDDDDK) insertion.

Super resolution fluorescence microscopy

Bacteria were grown in liquid BB-FBS media overnight until the Abs600 was approximately 0.7. 1 mL aliquots of culture were collected and pelleted at 3500 x g for 5 min at 4°C. Bacteria were washed and pelleted 3 times in PBS. The bacterial suspension was then concentrated 10 times and spotted on a poly-L-lysine coverslip (Corning). Bacteria were allowed to adhere for 2 minutes at room temperature. Excess bacteria were aspirated and adherent bacteria were fixed with 2% PFA in PBS for 10 minutes. Fixation was stopped by the addition of TBS three times for 5 minutes. Coverslips were blocked using Odyssey blocking reagent for 1 hour (PBS based, LICOR). The FLAG tag was labeled using 1:1000 M2 anti-FLAG (Sigma) for 2 hours at room temperature in blocking buffer. Bacteria were washed with PBS and anti-FLAG antibody was detected using anti-mouse Alexa Fluor 546 (Thermo) for 1 hour at room temperature in blocking buffer. Bacteria were washed three times in PBS. Bacteria were stained using Sybr Green I in PBS for 30 minutes (Thermo). Slides were mounted using Prolong Gold (Thermo). Samples were imaged on OMX super resolution microscope using 488 nm and 568 nm lasers and a 60x Plan-Aprochromat objective (GE Lifesciences). Bacterial dimensions were measured using ImageJ.

Results

Investigation of CagT lipidation status

Classically, lipoproteins are identified by either radiolabeled palmitic acid incorporation or by a globomycin-dependent SDS-PAGE gel shift. *H. pylori* does not take up palmitic acid (likely due to the lack of a long-chain fatty-acid transporter FadL analog) (data not shown) (99), and is not sensitive to globomycin (97). To overcome these limitations, I analyzed possible lipid incorporation of CagT expressed as a recombinant protein in *E. coli* (rCagT). *E. coli* strains expressing CagT were labeled with palmitic acid alkyne. Proteins modified with the palmitic acid alkyne were detected via conjugation with Cyazine 7.5 azide. Proteins were resolved using SDS-PAGE transferred to nitrocellulose membranes. CagT production was determined using CagT antisera and fluorescent conjugated anti rabbit immunoglobulin. Cyazine 7.5-coupled proteins and immunolabeled CagT was visualized for fluorescence using an Odyssey imaging system. Analysis of empty vector *pBAD*, uninduced *E. coli*, unlabeled *cagTpBAD*, and non-lipidated *cagT_{C21S}pBAD* did not reveal a band at the approximate size of CagT (30 kDa) for either the Cyazine 7.5 channel or the anti-rabbit fluorescent channel. However, analysis of induced *cagTpBAD* labeled with palmitic acid did reveal a band at the appropriate size for both the Cyazine 7.5 and anti-rabbit channels. In addition to the expected size of CagT, CagT antisera also labeled a band of larger size in all induced samples (Figure19). The larger CagT band was likely the unprocessed, full-length CagT prior to signal sequence cleavage. With the observation that

*cagT*_{C21S}*pBAD* strain had the larger size band and not the expected size band was the phenotype we would expect upon globomycin treatment (97). This suggests the *cagT*_{C21S}*pBAD* strain only is capable of producing the immature, full-length form of CagT and is unable to cleave the amino-terminal signal sequence. Overall, these data suggest CagT recombinantly expressed in *E. coli* is lipidated.

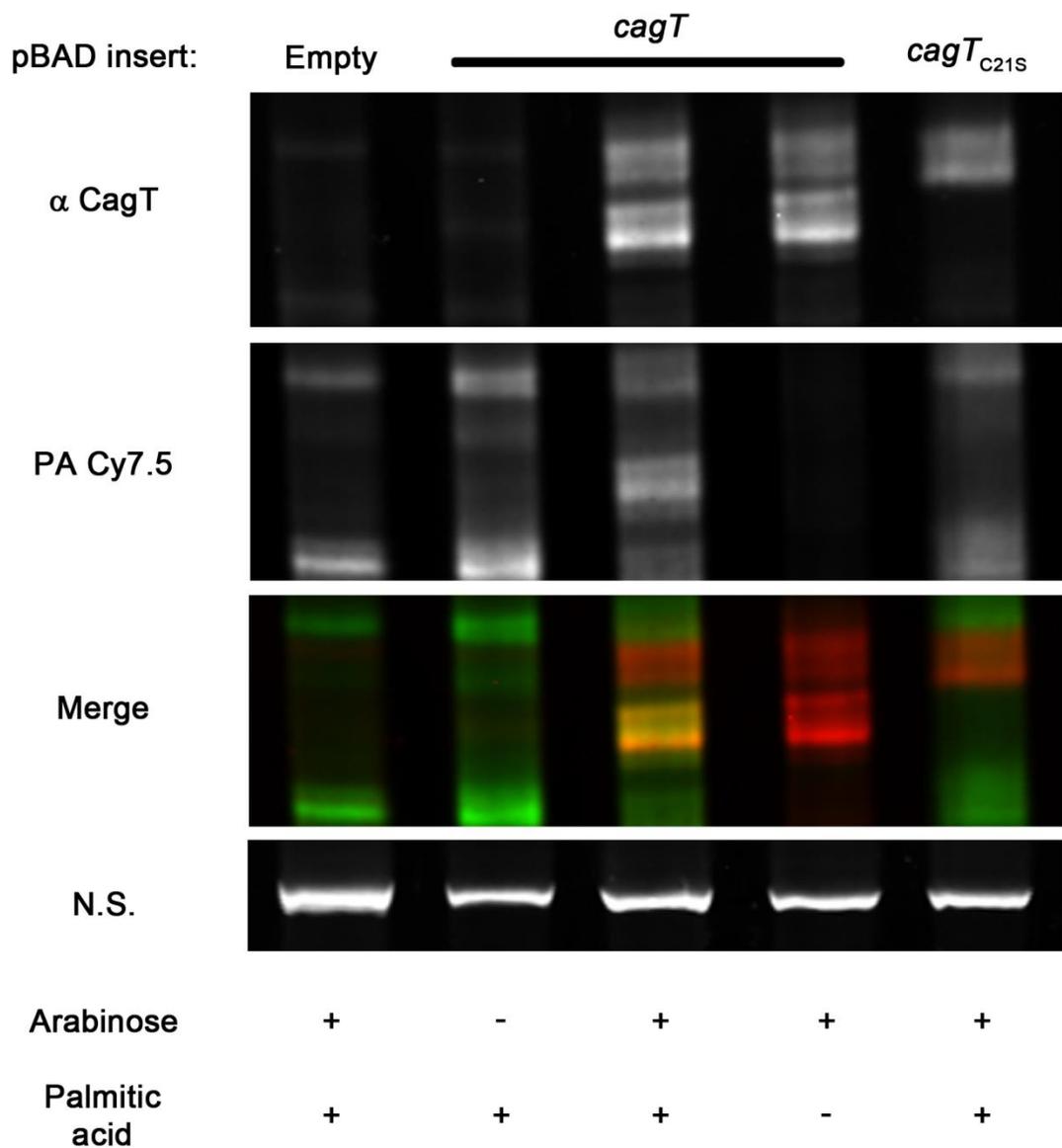


Figure 19. Incorporation of palmitic acid into recombinant CagT. *E. coli* strain MG1655 was transformed with either *cagT* or *cagT*_{C21S}. Strains were grown until mid-log, induced with 0.2% arabinose, and lipoproteins were labeled with palmitic acid alkyne. Palmitic acid incorporation was detected by labeling cell lysates with Cyazine 7.5 azide via Click chemistry (PA Cy7.5). N.S.: non-specific band representing loading control.

Effect of disruption of the lipobox on steady-state levels of CagT

The only conserved amino acid within the lipobox motif among different bacterial species is a cysteine (7), which is post translationally lipidated. With studies of *A. tumefaciens*, removal of the VirB7 lipobox cysteine resulted in bacteria unable to produce a T-pilus (120). With studies of the *E. coli* VirB/D4 T4SS (Tra T4SS), it has been shown that removal of the lipobox cysteine in VirB7 (TraN) results in misslocalization of VirB7, VirB9, and VirB10 to the inner membrane (39). Classical experiments to determine if a bacterial protein is lipidated (globomycin-dependent SDS-PAGE size shift and palmitic acid incorporation) have yet to be successful in *H. pylori*. Therefore, direct detection of CagT lipidation in *H. pylori* is currently impossible. To investigate if CagT is lipidated when expressed in *H. pylori*, I generated a strain of *H. pylori* encoding a gene for CagT_{C21S} (*cagT::cagT_{C21S}*), similar to what was done previously in *E. coli* with TraN. CagT_{C21S} in this strain was undetectable via immunoblot. Sequencing of the gene region confirmed that no secondary mutations in the *cagT* gene were present. In addition to the reduction in steady-state abundance of CagT, the CagT_{C21S} strain no longer induced IL-8 upon coculture with AGS gastric epithelial cells. These data suggest the steady-state abundance of CagT in this strain was too low to allow assembly of a functional Cag T4SS.

To rescue this decreased steady-state abundance phenotype, I reintroduced the four amino acids corresponding to the CagT lipobox at sites either N-terminal (Rescue A) or C-terminal (Rescue B) to the C21S mutation (Figure 20). Both of these strains had greater levels of steady-state abundance of CagT compared to the CagT_{C21S} strain. Further, both strains were capable of inducing IL-8 upon coculture with AGS cells

(Figure 21). These data indicate that reintroducing a wild type (wt) lipobox within CagT_{C21S} was sufficient to restore CagT function.

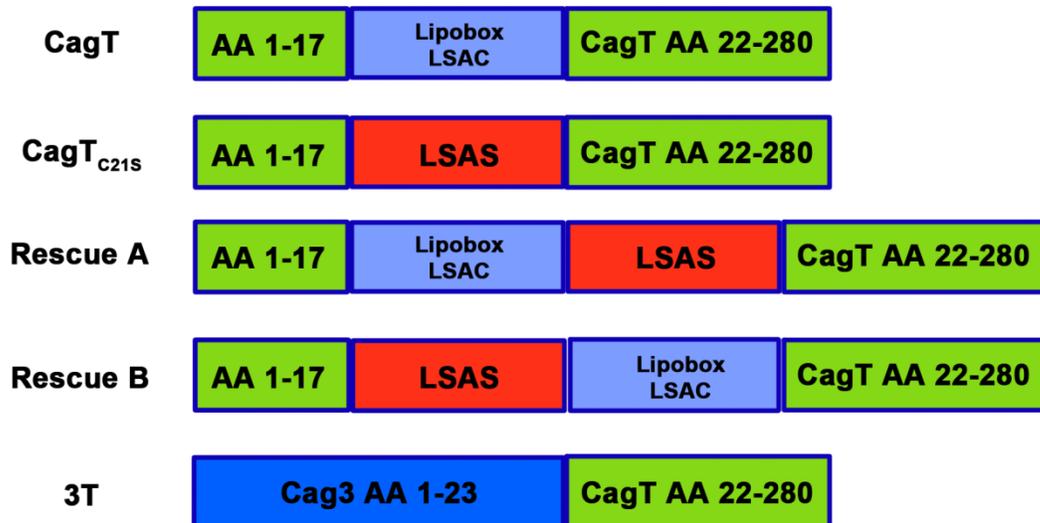


Figure 20. Schematic of CagT lipidation mutants. CagT: wt CagT sequence with intact lipobox motif LSAC. CagT_{C21S}: CagT with cysteine to serine lipobox mutation resulting in a predicted non-lipidated form of CagT. Rescue A and Rescue B: Lipobox disruption complementation via insertion of wt lipobox sequence either N-terminal or C-terminal to C21S mutation. 3T: Cag3-CagT fusion, signal sequence of Cag3 replaced signal sequence and lipobox of CagT. Construct is predicted to generate a non-lipidated form of CagT using the Cag3 secretion signal. Green corresponds to wt CagT sequence. Light blue corresponds to wt CagT lipobox sequence. Red corresponds to C21S mutation (non-lipidated CagT). Dark blue corresponds to wt Cag3 sequence until predicted signal peptidase cleavage site.

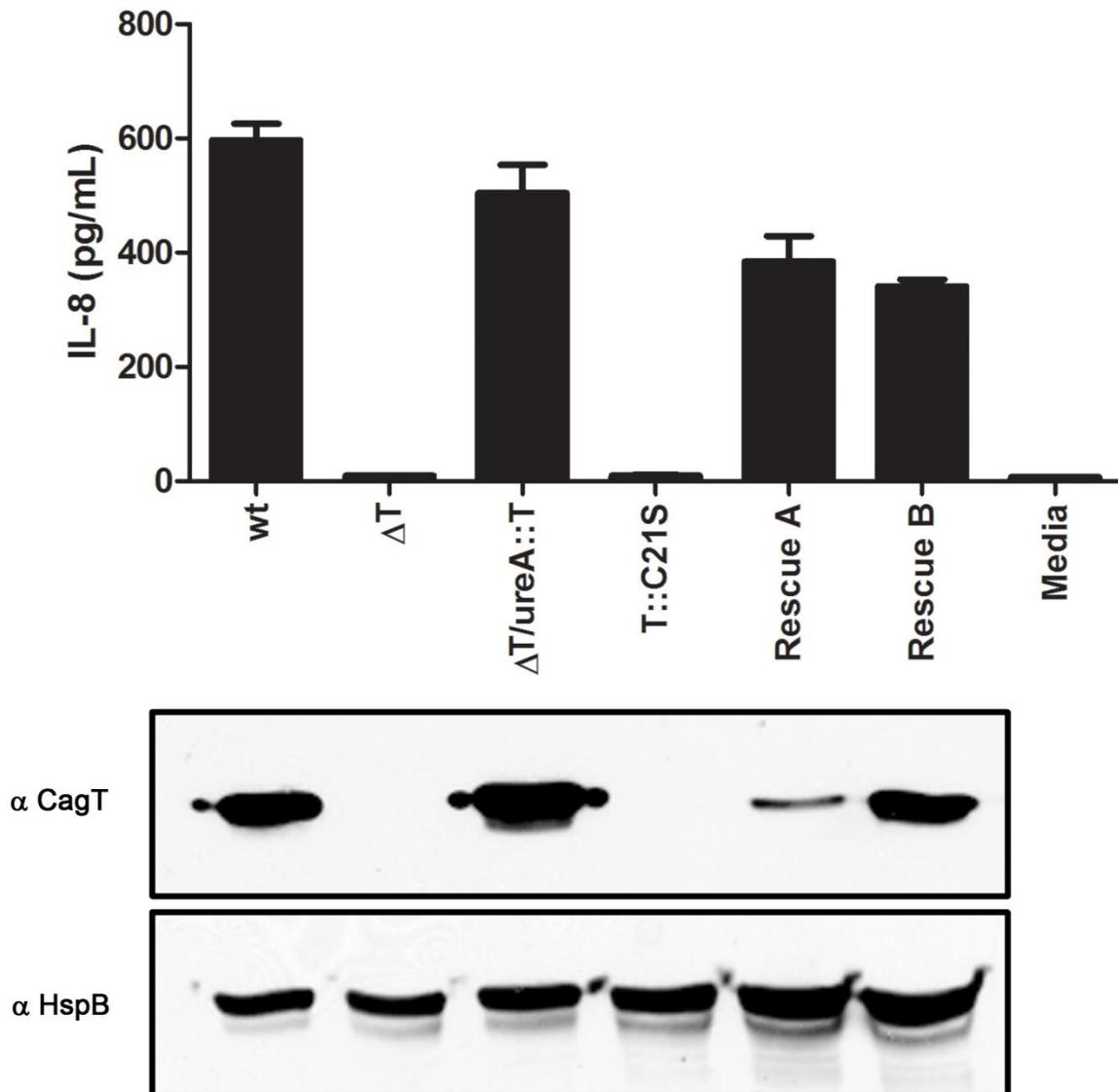


Figure 21. Rescue of C21S mutation. Top panel: IL-8 levels in supernatant after 4 hour coculture of AGS cells with indicated *H. pylori* strains. Bottom panel: CagT and HspB immunoblot of whole cell lysates of indicated *H. pylori* strains. T::C21S: *cagT* replaced by gene for non-lipidated CagT (CagT_{C21S}). Genes encoding Rescue A and Rescue B inserted in the *cagT* endogenous locus.

CagT expressed in *E. coli* was observed to be lipidated (Figure 19). Disruption of the lipobox in CagT resulted in lower CagT steady-state abundance and this strain was deficient in IL-8 induction. To investigate whether the reduction in IL-8 induction and CagT steady-state abundance were due to loss of CagT lipidation or due to the C21S mutation (unrelated to loss of lipidation), I constructed another non-lipidated form of CagT. This form of CagT completely lacked a lipobox, and contained a signal peptidase I cleavage site instead of a predicted SP11 cleavage site. I achieved this by generating a gene fusion of nucleotides encoding the body of CagT (AA 22-280) with the secretion signal of the Cag3 (AA 1-22) (Figure 20). Cag3 is a protein not predicted to be lipidated and is also a surface-exposed Cag protein (Table 3). The resulting fusion was designated 3T. Lysate was resolved using SDS-PAGE and abundance of CagT was estimated by immunoblotting with CagT antisera. While CagT abundance was below the level of detection when immunoblotting 20 μ g of lysate, the 3T strain was capable of inducing IL-8 when cocultured with AGS cells (Figure 22). While unexpected, this was a reproducible result. It is possible that 3T is not detectable via immunoblot because CagT antiserum was not able to bind it. However, this is not likely because MudPIT LC-MS/MS of 3T membrane fractions were also unable to assign spectra to CagT (data not shown). Sequence analysis of the gene encoding 3T did not identify any secondary mutations. An alternative explanation is that the Cag T4SS requires very few copies of CagT to function, and the amount of CagT present, while sufficient for IL-8 induction, was below the limit of detection via immunoblot. By developing a much longer exposure an overloaded 3T lysate immunoblot, I did detect a very faint band of appropriate size of

CagT (Figure 22). These data suggest lipidation is not required for CagT function but is required for wt steady-state abundance of CagT.

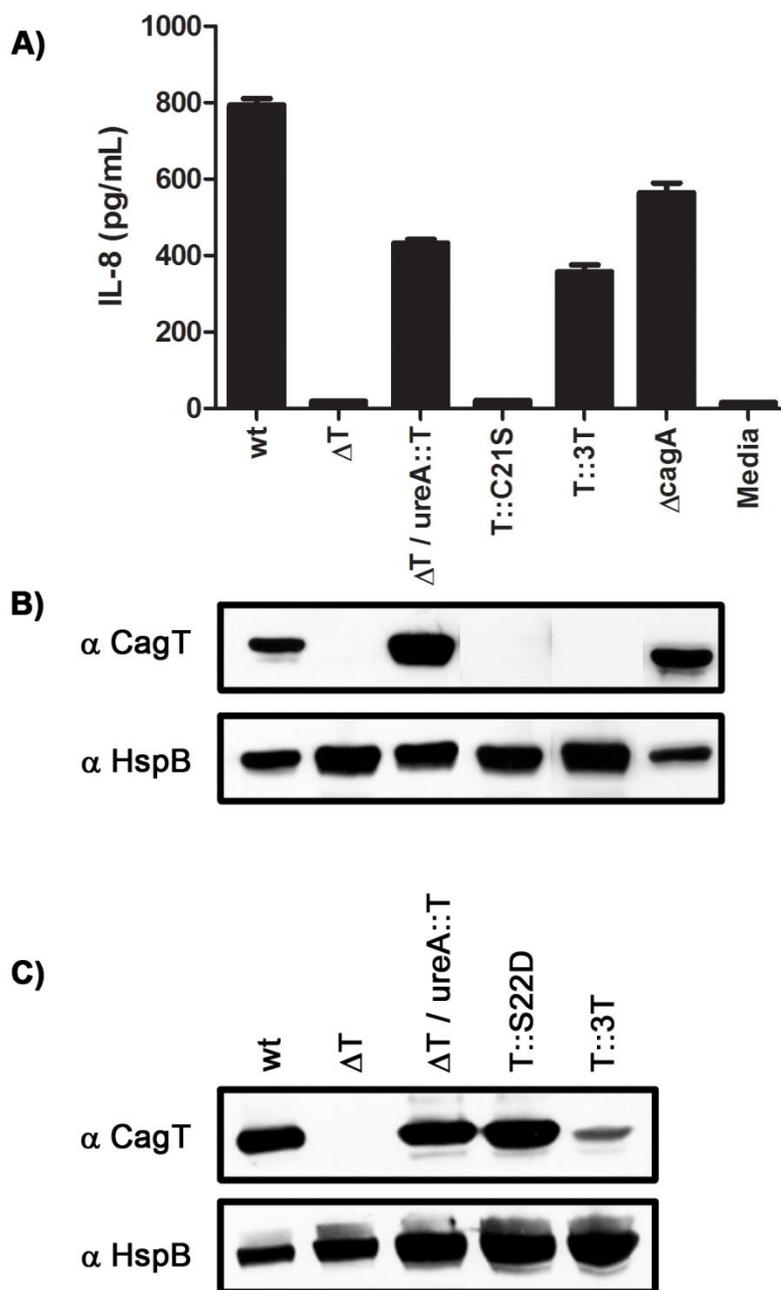


Figure 22. IL-8 induction and immunoblot of non-lipidated CagT. A) Amount of cytokine IL-8 in AGS cell supernatant after 4 hour coculture with indicated *H. pylori* strain. B) Immunoblot of 20 μ g whole cell lysate using either CagT antisera or HspB antisera. Lanes are indicated as in A). C) Immunoblot of whole bacteria using either

CagT antisera or HspB antisera. wt = 26695, $\Delta T = \Delta cagT$, $\Delta T / ureA::T = \Delta cagT$ with wt *cagT* expressed in the *ureA* locus, T::S22D= *cagT*_{S22D}, T::3T = Cag3-CagT (Figure 20).

CagT localization

CagT was identified as being surface-exposed using immunogold TEM (Figure 6). This finding is significant because lipoprotein localization is poorly understood in *H. pylori*, and *H. pylori* lacks several Lol complex protein homologs. To investigate whether CagT is shuttled to the outer membrane in a Lol-dependent pathway, I inserted a canonical +2 Lol-avoidance point mutation (CagT_{S22D}) in the *cagT* endogenous locus. CagT from this strain was detected by immunoblotting in similar abundance compared to the wt strain (Figure 22). Further, CagT_{S22D} was found in the Triton-insoluble outer membrane fraction similar to wt CagT (Figure 23). This may suggest that either CagT uses a Lol-independent localization pathway, the *H. pylori* Lol-avoidance mechanism is non-canonical, or that *H. pylori* lacks a functional Lol complex.

Independent of the Lol complex, proper localization of CagT may be dependent on other Cag proteins. To investigate this, I inserted *cagT* in the *ureA* locus of a 26695 Δ PAI strain (Figure 24). CagT in this strain localized to the outer membrane, similar to its localization in wild type bacteria. This suggests CagT localization is independent of other Cag proteins. However, a comparison of the α CagT / α HspB signal ratios of Δ T/T (1.8) and Δ PAI/T (0.4) suggested that some element of the *cag* PAI is required for full wt levels of steady-state CagT (Figure 24).

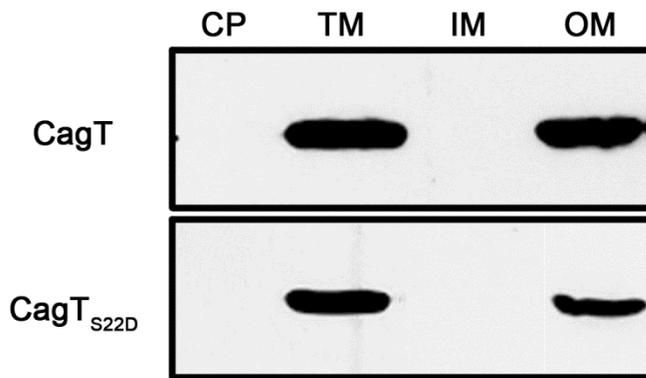


Figure 23. Triton subcellular fractionation of wt *H. pylori* and CagT_{S22D}. Both panels are immunoblots for CagT. 25 μ g total protein was loaded into each lane. CP: cytoplasmic fraction (detergent-free soluble lysate), TM: total membrane fraction (detergent-free insoluble lysate), IM: inner membrane fraction (Triton-X100 soluble TM lysate), OM: outer membrane fraction (Triton-X100 insoluble TM lysate)

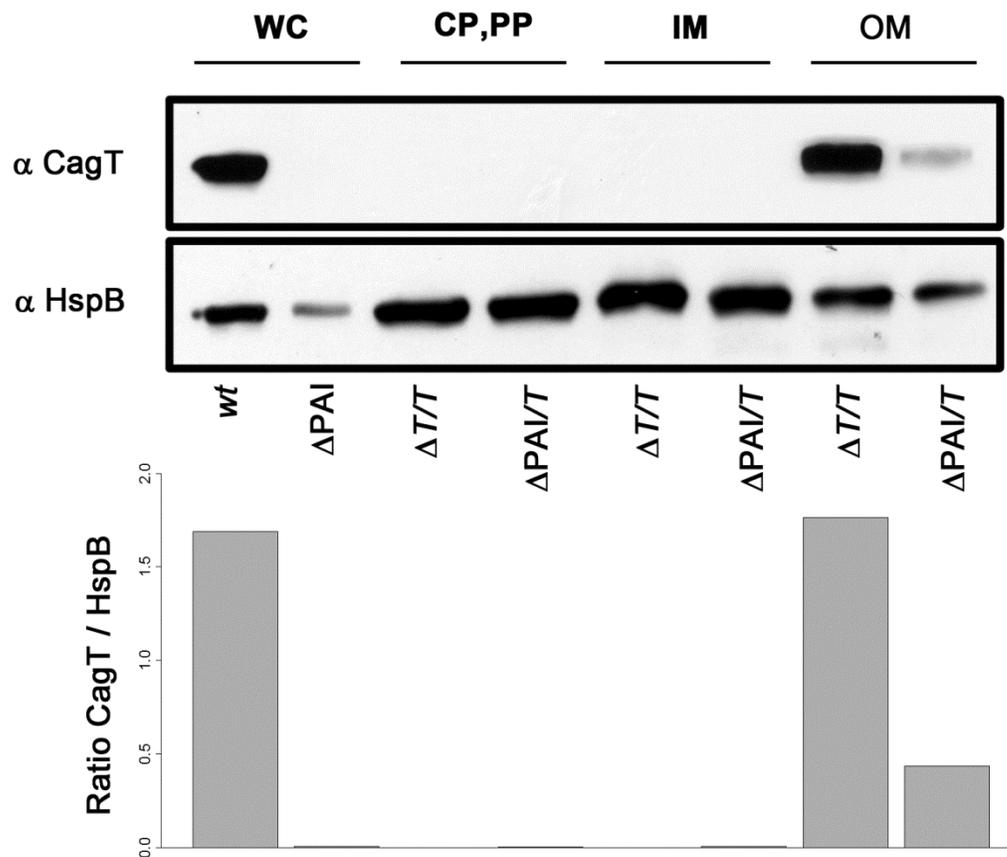


Figure 24. Localization of CagT in a Δ PAI background. Top: immunoblot of cellular fractionations for indicated *H. pylori* strains. WC fraction is lysate from wt and Δ PAI strains, CP,PP, IM, and OM fractions were generated with Δ T/T and Δ PAI/T strains. Bottom: densitometry of the α CagT signal / α HspB signal for each lane. CP: soluble protein lysate (cytoplasmic fraction), IM: Triton-soluble total-membrane fraction (inner membrane fraction), OM: Triton-insoluble total-membrane fraction (outer membrane fraction), WC: whole cell lysate. wt: unmodified strain 26695, Δ T: clean deletion of *cagT* (HP0532), Δ T/T: clean deletion of *cagT* with *cagT* complemented in *cis* (*ureA::cagT*), Δ PAI/T, *cag* PAI::chloramphenicol resistance cassette (*cat*) with *cagT* complemented in *cis* (*ureA::cagT*).

CagT surface topology

To characterize the surface-exposed topology of CagT, I inserted a FLAG epitope (DYKDDDDK) in various locations within CagT that were predicted to have low hydrophobicity (Figure 18). The resulting strains were named for the amino acid number and residue name immediately preceding the FLAG-tag insertion. Each *H. pylori* strain was cocultured with AGS gastric epithelial cells for 4 hours at an MOI of 100 to determine if insertion of the FLAG epitope disturbed Cag T4SS function (Figure 25). Strains 27K and 260L had similar steady-state levels of CagT compared to wt, and had a similar IL-8 induction phenotype. Strain 210Y had wt levels of CagT, but was unable to induce IL-8. Strain 145K had slightly less steady-state levels of CagT, but was still capable of wt levels of IL-8 induction. Strain 55K had similar steady-state levels of CagT compared to strain 145K, but unlike 145K, 55K was unable to induce IL-8. Strains 106K, 110Q, and 115R had very low steady-state amounts of CagT and were unable to induce IL-8 (Figure 25).

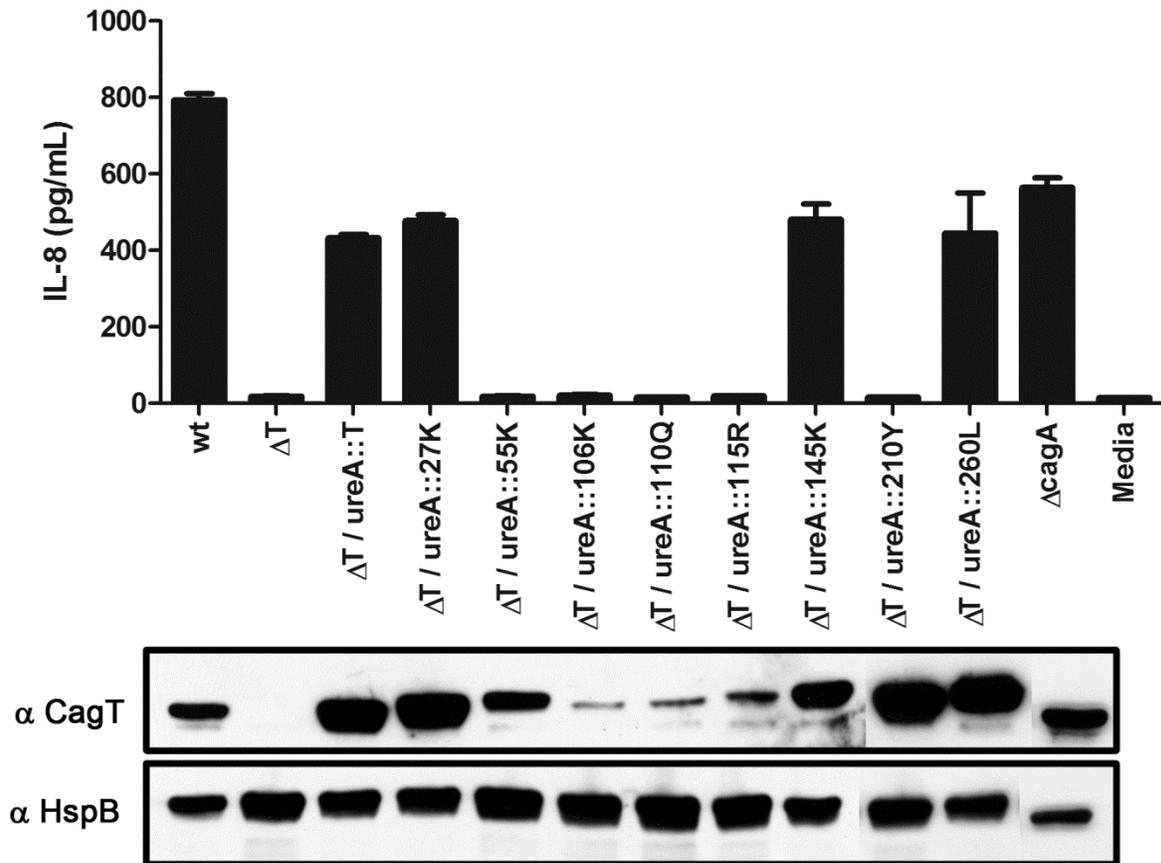


Figure 25. CagT steady-state abundance and IL-8 induction phenotype. Top: IL-8 induction of AGS cells after 4 hour coculture with an MOI of 100. Bottom: immunoblot for CagT and HspB. Numbers and letters correspond to preceding amino acid number for FLAG insertion.

Immunolabeling of intact bacteria from the FLAG-tag CagT panel identified the 27K epitope as being surface-exposed. After permeabilizing cells with 10 mM EDTA and sonicating once, the remaining FLAG-tagged strains reacted with the anti-FLAG antibody, whereas the *cagT* knock out complemented with *wt cagT* did not (Figure 26). This suggests the labeling was specific and the FLAG tag from the other strains was still solvent-exposed, just inaccessible under non-lytic conditions (Figure 26). 27K is within close proximity to the predicted post-translational lipidation site (C21). It is likely that the lipid moiety is inserted within the outer-leaflet of the outer membrane. Overall these results suggest CagT is a transmembrane lipoprotein, with its amino-terminus localized on the extracellular side of *H. pylori* and its carboxy-terminus localized within the periplasmic space (Figure 27).

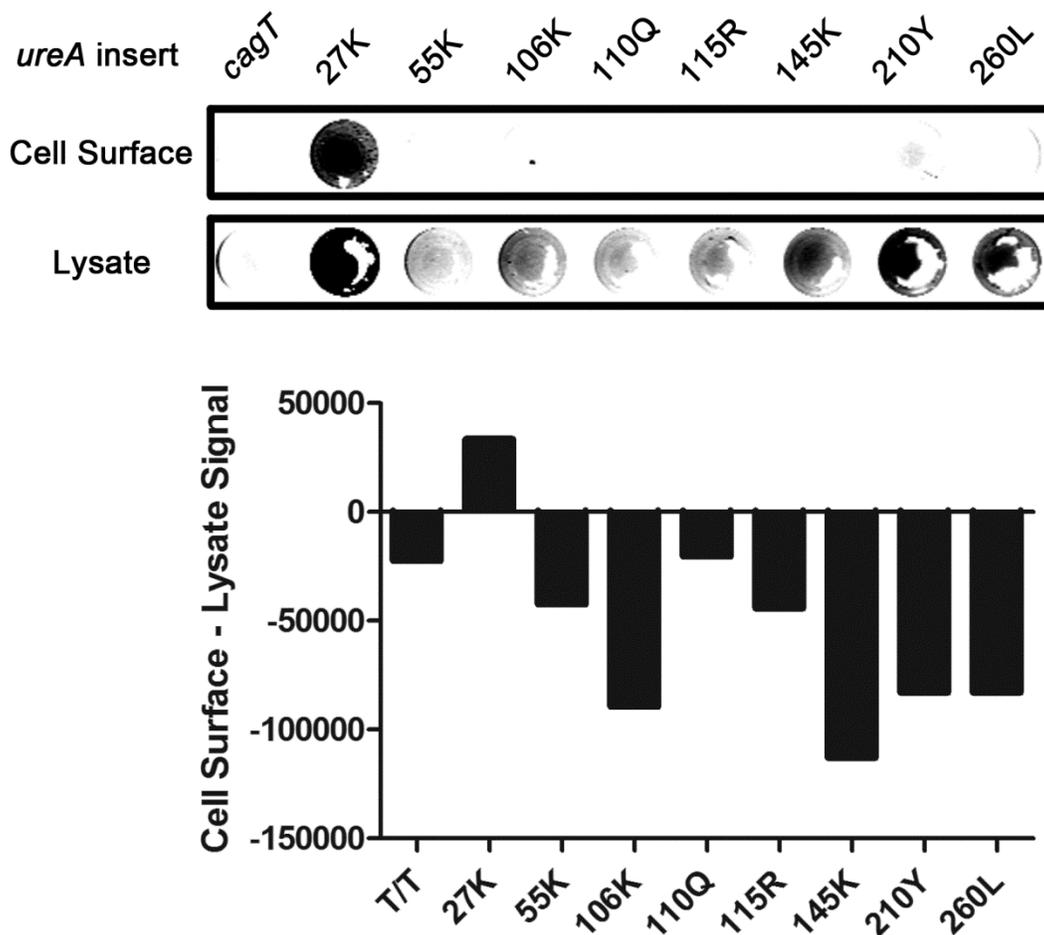


Figure 26. Intact cell immunolabeling of FLAG-tagged strains. Top: either intact cells (cell surface) or lysed (lysate) cells were applied to a lysine coated 96-well plate and probed with α FLAG antibody. Panel is representative image of duplicated samples. Bottom: average signal intensity from the lysed cells subtracted from the cell surface wells. All strains are from a $\Delta cagT$ background. Complementation of the $\Delta cagT$ was achieved by insertion nucleic acid sequence encoding either wt CagT (T/T), or FLAG-tagged CagT, in the *ureA* locus. Intact bacteria were bound to plates in PBS, permeabilized bacteria were sonicated in PBS with 10 mM EDTA. FLAG tag was detected using anti-FLAG immunoglobulin (Sigma) and anti-mouse IRDye 800 CW, LI-COR. Signal intensity was measured using Image Studio Lite (LI-COR).

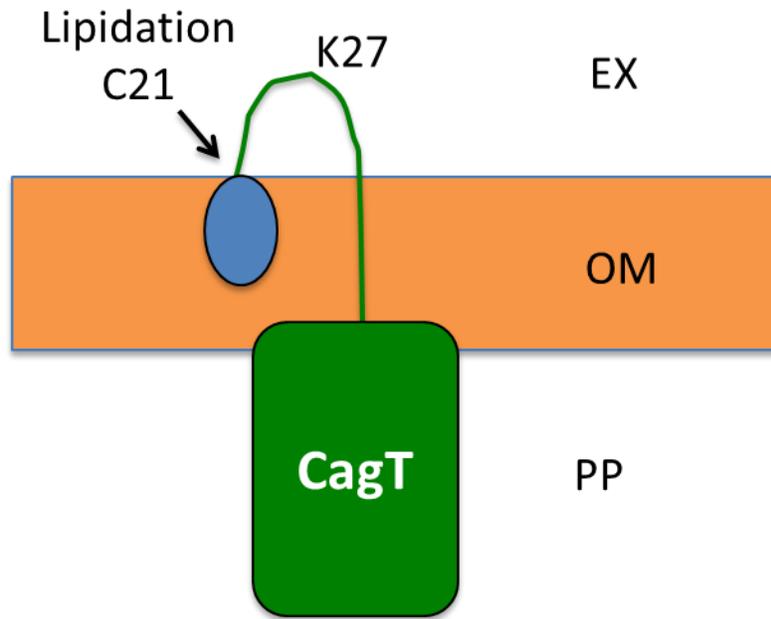


Figure 27. Proposed model of CagT topology. Blue oval: predicted lipoprotein post-translational modification. C21 and K27: amino acid positions, EX: extracellular, OM: outer membrane, PP: periplasm.

In Chapter II I confirmed CagT was surface-exposed using immunogold TEM (Figure 6). I observed CagT labeling as occurring in clusters. This was subsequently confirmed by another group (122). To further study this clustering phenotype I conducted studies with the FLAG-tagged CagT strain K27. By using super-resolution fluorescence microscopy I observed a FLAG tag distribution pattern with the strain K27 similar to CagT in wt bacteria. These results suggest CagT in the strain K27 distributes throughout *H. pylori* similar to the CagT distribution pattern I previously observed in wt bacteria (Figures 6 and 28).

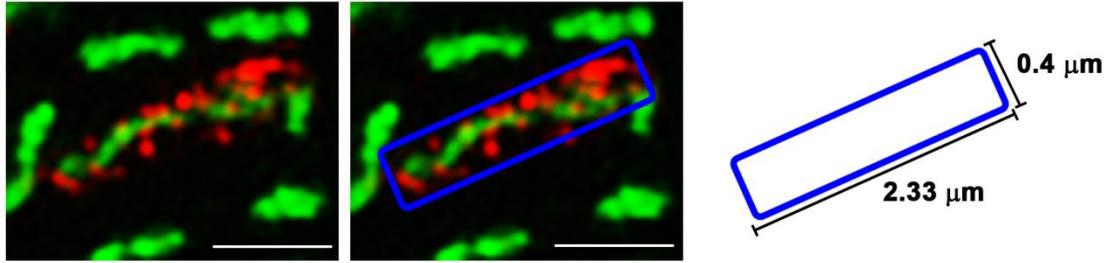


Figure 28. Super resolution of FLAG-tagged strain 27K in pure culture. Bacteria were fixed and labeled using anti-FLAG M2 (Sigma) and imaged on OMX super resolution microscope with a 60x Plan-Aprochromat objective (GE Lifescience). Green: Sybr Green I, red: anti-mouse Alexa Fluor 546, blue box: outline of single *H. pylori* bacteria, white bar = 1 μm . Typical dimensions of *H. pylori*: 2-4 μm x 0.5-1 μm .

Discussion

In this study I investigated the effect of post-translational modification on the localization and activity of the lipoprotein CagT, and also investigated CagT surface topology. I conclude that CagT, like VirB7, is most likely a lipoprotein based on studies of recombinant CagT in *E. coli* strain MG1655. Upon induction of CagT production, there was also a palmitic acid – cyazine 7.5 (PA Cy7.5) band of approximate size to expected size of CagT. CagT antiserum also labeled a similar sized band. This suggests recombinant CagT incorporated the palmitic acid. A strain producing lipidation-deficient CagT_{C21S} was unable to produce this palmitated band. This result suggests the C21 of CagT is the site of lipidation. Direct detection of such a modification in *H. pylori* is impossible due to its inability to take up palmitate (several attempts were unsuccessful), as well as being resistant to the signal peptidase II inhibitor globomycin (97).

Mutation of the cysteine predicted to be the site of post-translational lipidation results in a significant reduction in steady-state amounts of CagT. This C21S mutation also resulted in the inability of *H. pylori* to induce IL-8, similar to a clean *cagT* deletion (Figures 21 and 22). The abundance of CagT_{C21S} was too low to be detectable via immunoblot. Wild type phenotypes can be restored upon addition of the four amino acid wild-type lipobox either before or after the C21S substitution. Another strain that encoded a predicted non-lipidated CagT (designated 3T) had a steady-state level of 3T almost below the level of detection with immunoblot. Despite the inability to detect 3T at CagT wt levels on an immunoblot, this strain was still able to induce IL-8 production upon coculture with AGS cells. This suggests while the steady-state abundance of 3T is

very low, the few copies of 3T are functional. I can conclude that CagT does not require lipidation for function. However, lipidation of CagT is required for wild type steady state abundance of CagT.

Outer membrane localization of the *E. coli* lipoprotein Lpp requires the Lol complex (131). Lipoproteins with an aspartic acid immediately after the lipidated cysteine (often referred to as the Lol avoidance signal) will be retained in the inner membrane (140). When I substituted the +2 amino acid of CagT with an aspartic acid (CagT_{S22D}), I still detected CagT in the outer membrane. To investigate the possibility of CagT localization being a Cag-dependent function, I inserted *cagT* within the *ureA* locus in a strain of *H. pylori* missing the entire *cag* PAI. Like the CagT_{S22D} strain, this Δ PAI/*cagT* still localized CagT properly. This suggests proper localization of CagT requires an endogenous cellular process of *H. pylori*, likely to be required for localization of other non-Cag lipoproteins. Although insertion of a canonical Lol-avoidance signal had no effect on CagT OM localization, it is possible that CagT may be utilizing the Lol machinery. Other residues may be required for inner membrane retention (74). Previously, I have identified the hypothetical protein HP1457 as an inner membrane protein (135). This protein is predicted to be a lipoprotein based on the presence of a canonical lipobox and predicted signal peptidase II cleavage site. The +2 residue for HP1457 is not an aspartic acid; like CagT, the +2 residue is a serine. This would suggest *H. pylori* may use an alternative Lol-avoidance signal. For example, one group reported that the +3 and +4 residues are determinants of lipoprotein inner membrane retention in *Pseudomonas aeruginosa* (94). Future studies would be needed to investigate the possibility of novel inner membrane retention signals in *H. pylori*.

CagT is capable of outer membrane localization in a Δ PAI background. However, comparing the α CagT / α HspB signal ratios of Δ T/T and Δ PAI/T suggested that full wt levels of steady-state CagT requires an unknown element of the *cag* PAI. These data conflict with a previous study that used isogenic knockouts of different *cag* genes (72). Results of the previous study suggested steady-state amounts of CagT did not depend on the presence of Cag3, Cag4, VirB11, CagY, CagX, CagW, CagV, CagU, CagM, CagL, CagH, CagE, or CagC. However, this represents approximately half of the predicted Cag proteins encoded in the *cag* PAI. Other proteins may be required for the maintenance of steady-state abundance of CagT. Further, it is possible that an untranslated element influences transcript abundance, transcript stability, or translation efficiency. Future studies are needed to investigate these possibilities further.

My data suggest CagT is a surface-exposed lipoprotein. Outer membrane sorting of lipoproteins in ϵ proteobacteria is poorly understood. Surface localization is especially confounding as no shared system for surface localization of lipoproteins present across bacterial species has been described. While CagT immunogold TEM suggests CagT is surface-exposed, it was unclear if CagT was entirely surface-exposed, or only partially. Using FLAG-tagged CagT, I have shown that at least the amino-terminal domain of CagT, near the predicted site for lipidation, is surface-exposed, while the carboxy-terminus of CagT is not surface-exposed.

Overall, these studies strengthen our knowledge of CagT, an essential element of the Cag T4SS. More broadly, these studies may help guide future studies into how a surface-exposed lipoprotein is localized properly in *H. pylori*. Previous studies have identified several lipoproteins as being required for colonization of host tissue.

Investigation into the mechanism of lipoprotein synthesis and localization may help identify a potential therapeutic drug target for the treatment of *H. pylori*.

CHAPTER V

CONCLUSIONS

Summary

Helicobacter pylori is a human gastric bacterium that currently colonizes approximately half of the world's population. Colonization usually occurs in childhood, can persist lifelong, and is a strong risk factor for development of peptic ulcers or gastric cancer. There are many factors that influence disease outcome. These include host genetics, environmental conditions (host diet), and bacterial genetic makeup.

H. pylori is a very genetically diverse organism. Human migration has been mapped by analyzing the genetic profile of *H. pylori* strains from different geographically distinct locations (31). Strains of *H. pylori* that contain certain genetic elements, such as the *cag* PAI and the gene encoding s1m1 VacA, are associated with adverse clinical outcomes, compared to strains that lack these features (21).

Other major *H. pylori* virulence-associated factors are specific OMPs such as BabA and SabA (88, 93). These OMPs function as adhesins. BabA binds fucosylated Lewis b antigens and SabA binds Lewis x antigens. The amino-terminal domains of both BabA and SabA have recently been crystallized. Both have similar structures, characterized by a “golf club” like appearance. The “handle” domain connects the extracellular “head” domain to the carboxy-terminal β -barrel (106). The respective Lewis antigens bind at the top of this “head” domain (46). Considering the gastric environment where *H. pylori* colonizes, adhesion to host tissue is an important biological function. The bacteria must avoid being shed with the continual turnover of the gastric mucus

layer to prevent being exposed to the low pH of the gastric lumen. By having strong adhesive properties, the bacteria minimize this risk. Other OMPs with amino acid sequences similar to BabA and SabA are HopQ, BabB, SabB, HopZ, and OipA (HopH). These OMPs are also associated with pathogenesis (98). While these proteins all have similar amino acid sequences (primarily within the carboxy-terminal domain) (3), it is currently not known if any of these proteins function as adhesins or if these OMPs have other biological functions.

While we can use sequence analysis to find genetic elements associated with disease, it is unknown if these gene products are localized to the bacterial surface. The bacterial surface is the interface between host and pathogen and likely to contain many virulence determinants. Previous studies have attempted to identify surface-exposed proteins in *H. pylori* (118, 119). However, these studies often only employed a single approach. In order to more confidently identify surface-exposed proteins, I used multidimensional approach described in Chapter II (on-cell biotinylation, multiple cellular fractionations, and protease susceptibility). By combining several approaches in the same study, I was able to identify 45 putative surface-exposed proteins, more confidently than if I only used a single approach. Thirty-nine surface-exposed proteins were identified using on-cell biotinylation coupled with multidimensional cellular fractionation. Five additional proteins were identified by combining protease-susceptibility with on-cell biotinylation and 1 of 2 methods of cellular fractionation. VacA was identified as surface-exposed by coupling proteinase-K susceptibility with both membrane fractionation methods. Of the thirty-nine identified proteins, thirteen (32%) were annotated as either Hop (12 of 13) or Hop-related (1 of 13) OMPs. Seven OMPs

from other families were also identified. Four were from family 4 (iron-regulated) and three were from the “other” category.

I was also able to characterize the surface topology of several outer membrane proteins. Seven of the thirteen family 1 OMPs identified using the biotinylation and fractionation approaches were susceptible to digestion by an extracellular protease. An additional five protease-susceptible OMPs (4 Hop, 1 Hom) were not among the proteins identified using the multidimensional approach (BabC, HopA, HomA, HopM, and HopN). While 12 were considered significant by topology mapping criteria described in Chapter II, 21 of the 36 annotated Hop, Hop-related, Hom, and Hof families of OMPs displayed some level of protease susceptibility. All proteins from the Hop OMP family identified using the biotinylation and fractionation methods were identified as having some amount to protease susceptibility. Based on these results, it is clear many proteins in the Hop OMP family possess an extracellular domain that may be interacting with the extracellular environment and possibly making direct contact with the host. Future studies will be needed to expand on these findings including efforts to determine whether these proteins function as adhesins and what their binding targets are.

In Chapter III of this thesis. I identified salt-responsive membrane proteins of *H. pylori*. One of the environmental conditions associated with adverse outcomes of *H. pylori* infection is high dietary salt. While the host may undergo physiological changes in response to high salt conditions, I was interested in identifying salt-induced changes in *H. pylori* membrane protein abundance. I hypothesized proteins that changed in abundance in response to environmental salt (a known condition associated with an increased risk in the development of gastric cancer) may contribute to pathogenesis.

Sequence analysis of pathogenic vs non-pathogenic strains may identify potential virulence factors that are present in one strain and absent in another. However, what if a specific protein only contributes to virulence when steady-state abundance is altered? Sequence analysis would identify the gene encoding this protein in both pathogenic and non-pathogenic strains and thus would not be readily identified.

I focused this study on membrane proteins because this cellular fraction would contain the surface-exposed proteins identified in Chapter II. I used a multidimensional proteomic approach to identify salt-responsive membrane proteins. Specifically, I combined label-free Multidimensional protein identification technology (label-free MudPIT) with isobaric tag for absolute protein quantification (iTRAQ). I analyzed both the laboratory strain 26695 and the infection model strain 7.13 and assayed for proteins significantly changing in abundance in response to environmental salt.

I was able to identify a variety of proteins that were changing in abundance in response to environmental salt using this approach. Approximately half of the proteins identified as salt responsive using iTRAQ were also identified as significantly different in abundance between the two salt conditions using label-free MudPIT (Figure 16). Upon further analysis of the data, I noticed that vast majority of proteins identified using both methods were trending in the same direction (Figure 16). This emphasizes the power of our statistical analysis methods, giving us greater confidence in our list of salt-responsive proteins.

By comparing these results to the identified surface-exposed proteins in Chapter II, I identified three surface-exposed salt-responsive proteins in strain 26695 and five in

strain 7.13. Two salt-responsive surface proteins identified in both strains were the iron(III)-dicitrate transport protein FecA3 (also an OMP from the iron-regulated OMP family) and putative lipoprotein HP0746. FecA3 has been previously reported as being required for animal colonization. FecA3 had fewer assigned spectra from bacteria grown in high salt conditions compared to bacteria grown in low salt conditions. Animals with a high salt diet and infected with *H. pylori* developed a stronger inflammatory response than animals with a normal salt diet or uninfected animals (40). It is possible that changes in bacterial protein abundance influence this disease outcome. Based on FecA3 having fewer assigned spectra in high salt conditions, I speculate that FecA3 can be anti-inflammatory, or make the bacteria less adhesive to host tissue in higher abundance. The hypothetical protein HP0746 has a canonical lipobox so it is predicted to be lipidated. This protein was identified as having a greater number of assigned spectra in the high salt conditions compared to the low salt conditions. HP0746 also contains a multidomain identified by NCBI BLAST as a fibronectin type III domain (not to be confused with a fibronectin binding domain). Bacterial proteins with this fibronectin type III domain have been studied in gram positive organisms and these proteins may have enzymatic properties (64). In contrast to the speculation that FecA3 may be less inflammatory in higher concentrations, it is possible that HP0746 is proinflammatory. The increased abundance of HP0746 in high salt may be contributing to the increased host pathology with this condition. The VacA-paralog FaaA was identified as salt responsive in strain 26695 and HopD, HopQ, and CagT were identified as salt-responsive in strain 7.13.

Strain 26695 does not produce flagella, but strain 7.13 does. Several proteins in strain 7.13 with flagellar function annotations were identified as more abundant in low salt conditions than high salt conditions (Table 11). I hypothesized that due to the decreased abundance of flagellar proteins in high salt conditions, *H. pylori* would be less motile in a high salt environment. We used a soft agar colony diffusion assay to test this possible motility phenotype. In agreement with my hypothesis, I observed a reduced colony diameter under high salt conditions compared to low salt conditions (Figure 17). It is possibly that by reducing motility, *H. pylori* would be more adherent to host tissue under high salt conditions. Overall, these studies greatly expanded our knowledge on how the membrane proteome of *H. pylori* changes in response to environmental salt concentrations.

Chapter IV investigated CagT, a surface-exposed component of the Cag T4SS. In Chapter II, CagT was identified as surface-exposed using a multidimensional proteomic approach, and was confirmed using immunogold TEM. In Chapter III, CagT was identified in strain 7.13 as being salt-responsive (more abundant in high salt using label-free MudPIT and iTRAQ). Prior to my experiments in Chapter IV, there was very little data about any aspect of CagT, and much of its function was inferred because of a VirB7 annotation. This annotation was based solely on the presence of a shared four amino acid motif (named a lipobox) found in both CagT and VirB7. A lipobox is a predictor of post translational lipidation. However, the lipobox is the only distinguishing feature shared between CagT and VirB7. The most obvious difference between CagT and VirB7 is the length of the amino acid sequences. VirB7 is 55 amino acids in length while CagT is 280 amino acids in length.

CagT is required for all currently known Cag T4SS-dependent phenotypes. This highlights the importance of CagT for the activity of the Cag T4SS. My studies of CagT centered around two themes: investigation of the putative lipobox, and the surface-exposed topology of CagT. The significance of these studies is highlighted by the fact that surface-exposed lipoproteins are an emerging field of bacteriology and recently have been gaining appreciation for their importance in bacterial physiology and pathogenicity (70, 71, 113, 139).

Lipoproteins are classically identified by two main methods, globomycin sensitivity and palmitate incorporation. Unfortunately, *H. pylori* is resistant to globomycin and does not take up palmitate. I employed studies with recombinant CagT in order to investigate whether CagT is indeed a lipoprotein. I recombinantly expressed CagT in the *E. coli* strain MG1655 and labeled lipoproteins using labeled palmitate. I was able to identify rCagT (recombinant CagT) as being lipidated. I used a genetic approach to further study the lipidation of CagT in *H. pylori*. Previous studies of the lipoprotein VirB7 described how replacing the lipid-modified cysteine with a serine resulted in a misslocalization phenotype (39). In my studies, this same mutation in CagT resulted in CagT no longer being detectable via immunoblot (Figures 21 and 22). This reduced steady-state CagT was rescued by insertion of the four amino acid lipobox either before or after the disrupted lipobox. Further, a predicted non-lipidated fusion of the Cag3 signal sequence (predicted signal peptidase I cleavage site) to the CagT sequence beyond the lipobox was poorly detectable via immunoblot. These data suggest the lipobox is required for steady-state abundance of CagT. While CagT was undetectable via immunoblot in the strain expressing a Cag3-CagT fusion, this strain was still capable

of inducing IL-8 upon coculture with AGS cells, suggesting that it possessed a functional Cag T4SS, a surprising but reproducible phenotype. This result suggests lipidation of CagT is not required for its function and indicates that lipidation is required for steady-state abundance of CagT.

I also investigated the surface topology of CagT. I inserted an eight amino acid FLAG epitope (DYKDDDDK) at multiple sites throughout the length of CagT in regions of predicted low hydrophobicity (Figure 18). Immunoblotting of each of these CagT-FLAG strains for CagT identified the middle region of CagT as a region required for intact steady-state abundance of CagT (Figure 25). Strains 55K, 106K, 110Q, 115R, and 145K all had lower steady-state abundance of CagT compared to unlabeled CagT. Strains 27K, 210Y, and 260L had steady-state abundance of CagT similar to unlabeled CagT. Using these strains, I identified only the amino terminus of CagT (27K) as being surface-exposed (Figure 26). Because 27K is very close to the predicted SPII cleavage site (predicted SPII cleavage occurs between residues 20 - 21), it is likely the lipid modification is localized to the outer leaflet of the outer membrane (69).

Future Directions

Impact of surface-exposed OMPs on pathogenesis

The importance of the studies described in this thesis is based on the premise that surface-exposed *H. pylori* proteins influence disease outcome. This premise is based on the observation that OMPs BabA and SabA are known adhesins and patients colonized with *H. pylori* strains expressing these and other closely related OMPs have an increased risk of gastric cancer or peptic ulcer disease. In Chapter II, I identified 44 putative surface-exposed outer membrane proteins in *H. pylori*. Annotated OMPs were the most abundant family of proteins identified in Chapter II. Many of the identified OMPs (BabA, BabB, BabC, HopA, HopB, HopC, HopF, HopH, HopM, HopN, HopQ and HomA) displayed significant sensitivity to an extracellular protease, suggesting the presence of extensive extracellular domains. HomD was also identified as being protease-susceptible due to an observed decreased in total assigned spectra after protease treatment. However, there was insufficient spectral coverage for the topology mapping analysis described in Chapter II. With studies of BabA and SabA, the adhesive properties of these proteins have been found to be localized within the amino-terminal domain. Many of the OMPs identified in Chapter II have sequence similarity at the carboxy-terminus and sequence variability at the amino-terminus. I observed the protease-sensitivity specifically at the amino-terminus of these OMPs, suggesting the amino-terminal domain is exposed at the extracellular side of the outer membrane. Based on this, I hypothesize that many of the OMPs identified in Chapter II that display

significant protease-sensitivity also function as adhesins, similar to BabA and SabA. As a future approach to test this hypothesis I would generate a mutant panel for each of the 12 identified protease-susceptible OMPs. Each of these mutants could be tested for adhesive properties in either animal model or human stomach explant tissue samples (88). Mutant *H. pylori* that are less adherent to these tissue samples would be candidates for competition assays with wildtype bacteria in an animal model. Mutant bacteria that are significantly less abundant in the stomach after infection would suggest the corresponding OMP is likely acting as an adhesin.

A separate but related approach could be used to identify host cell receptors for *H. pylori* OMPs. BabA and SabA bind Lewis b and Lewis X antigens. Both antigens are carbohydrates. It is likely that other adhesive OMPs share this carbohydrate binding property with similar antigens. To identify putative host cell receptors for additional protease-susceptible OMPs, the OMP mutant panel deficient in production of the twelve protease-susceptible OMPs could be used in the screening of a mammalian glycan array. Candidates would be confirmed using adherence assays for cell lines (capable of expressing or unable to produce these glycans).

Effect of salt-responsive proteins on pathogenesis

I identified several surface-exposed proteins as being salt-responsive. However, it is unknown if the identified salt-responsive proteins influence host diet-dependent pathogenesis *in vivo*. A possible experiment to address this would be to infect animals with mutant strains of *H. pylori* deficient in the production of the identified salt-

responsive proteins. Priority would be given to salt-responsive proteins identified in both 26695 and 7.13 (FecA3 and HP0746). *H. pylori* mutant strains harboring mutations in these genes could be used for a competition assay with wild type bacteria in animals fed either high or low salt diets.

Further studies of salt-responsive changes in protein abundance

We identified several proteins changing in steady-state abundance in response to environmental NaCl concentration. Is the change due to the Na⁺, Cl⁻, or a more generally osmotic stress response? To investigate this, I propose comparing *H. pylori* grown with NaCl to bacteria grown in the presence of other salts, such as KCl (as a source of Cl⁻) or sodium bicarbonate (as a source of Na⁺). Generalized osmotic stress can be induced using a salt lacking both the sodium and the chloride, such as lithium carbonate. Each of these culture conditions could be applied in a study similar to Chapter III. The salt-responsive proteins identified in Chapter III could be compared to proteins identified as changing in relative abundance using these alternative additives. This would allow us to classify proteins as being either sodium-responsive, chloride-responsive, or as changing in abundance in response to osmotic stress. These studies may help us identify the specific stimuli responsible for the observed changes in protein abundance and guide future investigations of the mechanism of how *H. pylori* responds to environmental conditions.

While we were able to identify several proteins changing in abundance in response to environmental salt concentration, it is unknown whether the observed

changes are a result of changes in transcription or other alterations. I propose to perform RNA-Seq on *H. pylori* grown under the conditions from Chapter III. Changes in transcription identified with RNA-Seq may then be further validated by multiple reaction monitoring (MRM) LC-MS/MS. Ultimately an important goal is to define mechanism by which *H. pylori* alters protein abundance in response to changes in environmental salt concentration. Differentiating between transcriptional-dependent and transcriptional-independent changes in salt-responsive proteins is an important step toward this goal. An additional approach would be to utilize protein network databases. Proteins predicted to be involved in similar pathways or protein-protein interactions also identified as being salt-responsive would be a strong rationale for further investigation.

Another environmental additive that deserves further study is monosodium glutamate (MSG). I suggest this because areas of the world with high salt diets that correlate with higher levels of gastric cancer are generally also areas of the world that consume higher levels of MSG. Additionally, glutamate is an intermediate of the citric acid cycle (glutamate is converted to α -ketoglutarate by glutamate dehydrogenase). The extra glutamate in MSG may alter the metabolic activity of *H. pylori*, thus stimulating changes in protein abundance. Additionally the conversion of glutamate to α -ketoglutarate by glutamate dehydrogenase produces ammonia, which may aid *H. pylori* survival in the acidic environment of the stomach.

Is CagT a lipoprotein in *H. pylori*?

CagT expressed in *E. coli* strain MG1655 appears to be lipidated based on a palmitic acid incorporation assay. Other groups who previously investigated lipidation of *H. pylori* proteins performed experiments based on analysis of recombinant proteins (97, 99). To my knowledge, no group has been able to directly analyze lipidation in *H. pylori*. This is likely due to *H. pylori* being insensitive to globomycin (SPII inhibitor) and unable to take up palmitate (97, 99)(data not shown). One preliminary approach to identify post-translational lipidation in *H. pylori* would be to treat *H. pylori* lysate with a lipase to cleave the diacylglycerol on the modified cysteine side chain. Then it would be possible to assay for a decrease in size (expected decrease of about 0.5 kDa) of CagT via immunoblot. This would provide preliminary evidence of CagT lipidation in *H. pylori*. Spectra of lipase-treated and untreated CagT could be compared for further validation.

H. pylori localization of surface-exposed lipoproteins

Experiments showing immunogold and FLAG-tag labeling of CagT in intact *H. pylori* provide evidence that CagT is surface-exposed. Localization of lipoproteins to the bacterial surface is a little understood process. Expression of CagT in a strain of *H. pylori* deficient for the *cag* PAI still shows CagT localized to the Triton X-100 insoluble outer membrane fraction. This suggests CagT does not require other Cag proteins for proper localization. It is likely CagT is localized to the outer membrane by a cellular process also utilized by other *H. pylori* lipoproteins. Flagellar sheath protein HpaA (HP0797) and “a secreted protein involved in flagellar motility” (HP0232) are other

putative *H. pylori* lipoproteins (7, 97). Genes encoding both HpaA and HP0232 have been described as being required for colonization (65). This suggests that inhibition of lipoprotein processing in *H. pylori* may be a novel therapeutic approach. An approach to identify a possible localization chaperone would be to immunopurify CagT and identify proteins coenriched. Assuming genes that are required for colonization are not essential for bacterial survival, another approach would be to screen a mutant *H. pylori* library for strains that no longer properly localize CagT on its surface, by magnetic-associated cell sorting (MACS). Strains without antibody labeled CagT would be analyzed via deep sequencing for identification of disrupted genes that may be involved in lipoprotein localization. Candidate genes would be disrupted via site-directed mutagenesis and CagT localization would be analyzed by cellular fractionation and on-cell immunolabeling.

Localization of CagT in coculture

Images generated from scanning electron microscopy of *H. pylori* in coculture with gastric epithelial cells have revealed filamentous structures (Cag T4SS pili) at the host-bacterial interface (96, 124). Immunogold TEM of CagT from *H. pylori* grown in pure culture showed labeling of CagT on the bacterial surface. It is unclear whether Cag proteins change localization in response to contact with host cells, or if Cag T4SS pili only form at Cag T4SS cores in close proximity to the host tissue. Use of the FLAG-tagged CagT strain 27K would be a way to begin to investigate whether CagT changes its distribution on the surface of *H. pylori* in response to host-cell contact. By using immunofluorescence (Figure 28) or immunogold microscopy, I would compare CagT-

FLAG surface labeling in *H. pylori* in pure culture to bacteria in coculture with AGS gastric epithelial cells.

Last Remarks

An in-depth knowledge of the composition of the bacterial surface is important for understanding how a bacterium interacts with its environment and how pathogens cause disease in the host. Many virulence determinants are present on the bacterial surface and are within close proximity to the host. *H. pylori* encodes many OMPs that are associated with pathogenesis and two of these OMPs are known adhesins, BabA and SabA. There are approximately 30 OMPs with a large amount of sequence similarity to BabA and SabA (OMP family 1). Protease-susceptibility assays and structural studies (crystallography) suggest these and other related OMPs possess a large extracellular domain and are likely interacting with host tissue. Further study of the OMPs identified in this thesis is important for advancing our understanding of how *H. pylori* interacts with the host and ultimately how *H. pylori* causes disease.

In this thesis I describe how *H. pylori* can adapt to changes in its environment. While I investigate extracellular salt concentration as a model for a high salt diet, there are many other potential environmental factors (such as nutrient availability) that may affect *H. pylori* and its ability to cause disease. I mentioned MSG as a source of glutamate that may potentially alter the metabolic activity in *H. pylori* as an example. Other environmental factors including dietary metal intake or alternative carbon sources may also be factors that contribute to pathogenesis.

The gene encoding CagT is located on a pathogenicity island that is strongly correlated with severe disease outcomes. Gene products of this pathogenicity island are predicted to form a Cag T4SS that translocates CagA into host cells. However, currently

there is little functional data for most of the Cag T4SS components and very little is known about how the Cag T4SS translocates CagA into host cells. Understanding how this process occurs could potentially lead to development of new therapeutic approaches to combat *H. pylori* infection.

In summary, study of the bacterial cell-surface proteome is an important area of research that will enable us to better understand how *H. pylori* causes disease in its human host and how we may better treat infection in the future.

LIST OF PUBLICATIONS

Johnson EM, Gaddy JA, **Voss BJ**, Hennig EE, Cover TL. 2014. Genes required for assembly of pili associated with the *Helicobacter pylori* Cag type IV secretion system. *Infect Immun.* 82:3457-3470.

Voss BJ, Gaddy JA, McDonald WH, Cover TL. 2014. Analysis of surface-exposed outer membrane proteins in *Helicobacter pylori*. *J Bacteriol.* 196:2455-2471.

Voss BJ and Cover TL. 2015. Biotinylation and purification of surface-exposed *Helicobacter pylori* proteins. *Bio-protocol* 5(8): e1455. <http://www.bio-protocol.org/e1455>

Voss BJ, Loh JT, Hill S, Rose KL, McDonald WH, Cover TL. 2015. Alteration of the *Helicobacter pylori* membrane proteome in response to changes in environmental salt concentration. *Proteomics Clin. Appl.* 9:1021-1034.

Snider CA, **Voss BJ**, McDonald WH, Cover TL. 2016. Growth phase-dependent composition of the *Helicobacter pylori* exoproteome. *J. Proteomics* 130:94-107.

Snider CA, **Voss BJ**, McDonald WH, Cover TL. 2015. Supporting data for analysis of the *Helicobacter pylori* exoproteome. *Data in Brief* 5:560-563.

Frick-Cheng AE, Pyburn TM, **Voss BJ**, McDonald WH, Ohi MD, Cover TL. 2016. Molecular and Structural Analysis of the *Helicobacter pylori* cag Type IV Secretion System Core Complex. *mBio* 7:e02001-15.

Voss JB, et al. 2016. Investigation of CagT, a putative VirB7 functional analog. [Manuscript in preparation]

REFERENCES

1. **Algood HM, Cover TL.** 2006. *Helicobacter pylori* persistence: an overview of interactions between *H. pylori* and host immune defenses. Clin Microbiol Rev **19**:597-613.
2. **Allen WJ, Phan G, Waksman G.** 2012. Pilus biogenesis at the outer membrane of Gram-negative bacterial pathogens. Curr Opin Struct Biol **22**:500-506.
3. **Alm RA, Bina J, Andrews BM, Doig P, Hancock RE, Trust TJ.** 2000. Comparative genomics of *Helicobacter pylori*: analysis of the outer membrane protein families. Infect Immun **68**:4155-4168.
4. **Amieva MR, El-Omar EM.** 2008. Host-bacterial interactions in *Helicobacter pylori* infection. Gastroenterology **134**:306-323.
5. **Atherton JC, Blaser MJ.** 2009. Coadaptation of *Helicobacter pylori* and humans: ancient history, modern implications. J Clin Invest **119**:2475-2487.
6. **Atherton JC, Cao P, Peek RM, Jr., Tummuru MK, Blaser MJ, Cover TL.** 1995. Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. Association of specific *vacA* types with cytotoxin production and peptic ulceration. J Biol Chem **270**:17771-17777.
7. **Babu MM, Priya ML, Selvan AT, Madera M, Gough J, Aravind L, Sankaran K.** 2006. A database of bacterial lipoproteins (DOLOP) with functional assignments to predicted lipoproteins. J Bacteriol **188**:2761-2773.
8. **Baik SC, Kim KM, Song SM, Kim DS, Jun JS, Lee SG, Song JY, Park JU, Kang HL, Lee WK, Cho MJ, Youn HS, Ko GH, Rhee KH.** 2004. Proteomic analysis of the sarcosine-insoluble outer membrane fraction of *Helicobacter pylori* strain 26695. J Bacteriol **186**:949-955.
9. **Baldwin DN, Shepherd B, Kraemer P, Hall MK, Sycuro LK, Pinto-Santini DM, Salama NR.** 2007. Identification of *Helicobacter pylori* Genes That Contribute to Stomach Colonization. Infect Immun **75**:1005-1016.

10. **Belogolova E, Bauer B, Pompaiah M, Asakura H, Brinkman V, Ertl C, Bartfeld S, Nechitaylo TY, Haas R, Machuy N, Salama N, Churin Y, Meyer TF.** 2013. *Helicobacter pylori* outer membrane protein HopQ identified as a novel T4SS-associated virulence factor. *Cell Microbiol* **15**:1896-1912.
11. **Blaser MJ, Berg DE.** 2001. *Helicobacter pylori* genetic diversity and risk of human disease. *J Clin Invest* **107**:767-773.
12. **Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y.** 1997. The Complete Genome Sequence of *Escherichia coli* K-12. *Science* **277**:1453-1462.
13. **Bos MP, Robert V, Tommassen J.** 2007. Biogenesis of the gram-negative bacterial outer membrane. *Annu Rev Microbiol* **61**:191-214.
14. **Braun V.** 1995. Energy-coupled transport and signal transduction through the Gram-negative outer membrane via TonB-ExbB-ExbD-dependent receptor proteins. *FEMS Microbiol Rev* **16**:295-307.
15. **Bumann D, Aksu S, Wendland M, Janek K, Zimny-Arndt U, Sabarth N, Meyer TF, Jungblut PR.** 2002. Proteome analysis of secreted proteins of the gastric pathogen *Helicobacter pylori*. *Infect Immun* **70**:3396-3403.
16. **Busler VJ, Torres VJ, McClain MS, Tirado O, Friedman DB, Cover TL.** 2006. Protein-protein interactions among *Helicobacter pylori* *cag* proteins. *J Bacteriol* **188**:4787-4800.
17. **Cao P, Cover TL.** 2002. Two different families of *hopQ* alleles in *Helicobacter pylori*. *J Clin Microbiol* **40**:4504-4511.
18. **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. *Proc Natl Acad Sci U S A* **93**:14648-14653.
19. **Chen EI, Hewel J, Felding-Habermann B, Yates JR.** 2006. Large Scale Protein Profiling by Combination of Protein Fractionation and Multidimensional Protein Identification Technology (MudPIT). *Mol Cell Proteomics* **5**:53-56.

20. **Cover TL, Blanke SR.** 2005. *Helicobacter pylori* VacA, a paradigm for toxin multifunctionality. *Nat Rev Microbiol* **3**:320-332.
21. **Cover TL, Blaser MJ.** 2009. *Helicobacter pylori* in health and disease. *Gastroenterology* **136**:1863-1873.
22. **Cover TL, Dooley CP, Blaser MJ.** 1990. Characterization of and human serologic response to proteins in *Helicobacter pylori* broth culture supernatants with vacuolizing cytotoxin activity. *Infect Immun* **58**:603-610.
23. **Cover TL, Peek RM, Jr.** 2013. Diet, microbial virulence, and *Helicobacter pylori*-induced gastric cancer. *Gut microbes* **4**:482-493.
24. **Croxen MA, Sisson G, Melano R, Hoffman PS.** 2006. The *Helicobacter pylori* Chemotaxis Receptor TlpB (HP0103) Is Required for pH Taxis and for Colonization of the Gastric Mucosa. *J Bacteriol* **188**:2656-2665.
25. **Doig P, Exner MM, Hancock RE, Trust TJ.** 1995. Isolation and characterization of a conserved porin protein from *Helicobacter pylori*. *J Bacteriol* **177**:5447-5452.
26. **Doig P, Trust TJ.** 1994. Identification of surface-exposed outer membrane antigens of *Helicobacter pylori*. *Infect Immun* **62**:4526-4533.
27. **Duncan SS, Valk PL, McClain MS, Shaffer CL, Metcalf JA, Bordenstein SR, Cover TL.** 2013. Comparative genomic analysis of East Asian and non-Asian *Helicobacter pylori* strains identifies rapidly evolving genes. *PLoS one* **8**:e55120.
28. **Eaton KA, Morgan DR, Krakowka S.** 1992. Motility as a factor in the colonisation of gnotobiotic piglets by *Helicobacter pylori*. *J Med Microbiol* **37**:123-127.
29. **Eusebi LH, Zagari RM, Bazzoli F.** 2014. Epidemiology of *Helicobacter pylori* Infection. *Helicobacter* **19**:1-5.
30. **Exner MM, Doig P, Trust TJ, Hancock RE.** 1995. Isolation and characterization of a family of porin proteins from *Helicobacter pylori*. *Infect Immun* **63**:1567-1572.

31. **Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, Blaser MJ, Graham DY, Vacher S, Perez-Perez GI, Yamaoka Y, Mégraud F, Otto K, Reichard U, Katzowitsch E, Wang X, Achtman M, Suerbaum S.** 2003. Traces of Human Migrations in *Helicobacter pylori* Populations. *Science* **299**:1582-1585.
32. **Farnbacher M, Jahns T, Willrodt D, Daniel R, Haas R, Goesmann A, Kurtz S, Rieder G.** 2010. Sequencing, annotation, and comparative genome analysis of the gerbil-adapted *Helicobacter pylori* strain B8. *BMC genomics* **11**:335.
33. **Fischer W.** 2011. Assembly and molecular mode of action of the *Helicobacter pylori* Cag type IV secretion apparatus. *FEBS J* **278**:1203-1212.
34. **Fischer W, Buhrdorf R, Gerland E, Haas R.** 2001. Outer membrane targeting of passenger proteins by the vacuolating cytotoxin autotransporter of *Helicobacter pylori*. *Infect Immun* **69**:6769-6775.
35. **Fischer W, Puls J, Buhrdorf R, Gebert B, Odenbreit S, Haas R.** 2001. Systematic mutagenesis of the *Helicobacter pylori* cag pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8. *Mol Microbiol* **42**:1337-1348.
36. **Fox JG, Dangler CA, Taylor NS, King A, Koh TJ, Wang TC.** 1999. High-salt diet induces gastric epithelial hyperplasia and parietal cell loss, and enhances *Helicobacter pylori* colonization in C57BL/6 mice. *Cancer Res* **59**:4823-4828.
37. **Fox JG, Rogers AB, Ihrig M, Taylor NS, Whary MT, Dockray G, Varro A, Wang TC.** 2003. *Helicobacter pylori*-associated gastric cancer in INS-GAS mice is gender specific. *Cancer Res* **63**:942-950.
38. **Franco AT, Israel DA, Washington MK, Krishna U, Fox JG, Rogers AB, Neish AS, Collier-Hyams L, Perez-Perez GI, Hatakeyama M, Whitehead R, Gaus K, O'Brien DP, Romero-Gallo J, Peek RM.** 2005. Activation of β -catenin by carcinogenic *Helicobacter pylori*. *Proc Natl Acad Sci U S A* **102**:10646-10651.
39. **Fronzes R, Schafer E, Wang L, Saibil HR, Orlova EV, Waksman G.** 2009. Structure of a type IV secretion system core complex. *Science* **323**:266-268.
40. **Gaddy JA, Radin JN, Loh JT, Zhang F, Washington MK, Peek RM, Jr., Algood HM, Cover TL.** 2013. High dietary salt intake exacerbates *Helicobacter pylori*-induced gastric carcinogenesis. *Infect Immun* **81**:2258-2267.

41. **Gancz H, Jones KR, Merrell DS.** 2008. Sodium chloride affects *Helicobacter pylori* growth and gene expression. *J Bacteriol* **190**:4100-4105.
42. **Gangwer KA, Mushrush DJ, Stauff DL, Spiller B, McClain MS, Cover TL, Lacy DB.** 2007. Crystal structure of the *Helicobacter pylori* vacuolating toxin p55 domain. *Proc Natl Acad Sci U S A* **104**:16293-16298.
43. **Giglio KM, Fong JC, Yildiz FH, Sondermann H.** 2013. Structural Basis for Biofilm Formation via the *Vibrio cholerae* Matrix Protein RbmA. *J Bacteriol* **195**:3277-3286.
44. **Guruge JL, Falk PG, Lorenz RG, Dans M, Wirth H-P, Blaser MJ, Berg DE, Gordon JI.** 1998. Epithelial attachment alters the outcome of *Helicobacter pylori* infection. *Proc Natl Acad Sci U S A* **95**:3925-3930.
45. **Hagan CL, Silhavy TJ, Kahne D.** 2011. beta-Barrel membrane protein assembly by the Bam complex. *Annu Rev Biochem* **80**:189-210.
46. **Hage N, Howard T, Phillips C, Brassington C, Overman R, Debreczeni J, Gellert P, Stolnik S, Winkler GS, Falcone FH.** 2015. Structural basis of Lewisb antigen binding by the *Helicobacter pylori* adhesin BabA. *Science Advances* **1**.
47. **Harvey VC, Acio CR, Bredehoft AK, Zhu L, Hallinger DR, Quinlivan-Repasi V, Harvey SE, Forsyth MH.** 2014. Repetitive Sequence Variations in the Promoter Region of the Adhesin-Encoding Gene *sabA* of *Helicobacter pylori* Affect Transcription. *J Bacteriol* **196**:3421-3429.
48. **Hatakeyama M.** 2011. Anthropological and clinical implications for the structural diversity of the *Helicobacter pylori* CagA oncoprotein. *Cancer science* **102**:36-43.
49. **Henderson IR, Navarro-Garcia F, Desvaux M, Fernandez RC, Ala'Aldeen D.** 2004. Type V Protein Secretion Pathway: the Autotransporter Story. *Microbiology and Molecular Biology Reviews* **68**:692-744.
50. **Hoy B, Lower M, Weydig C, Carra G, Tegtmeyer N, Geppert T, Schroder P, Sewald N, Backert S, Schneider G, Wessler S.** 2010. *Helicobacter pylori* HtrA is a new secreted virulence factor that cleaves E-cadherin to disrupt intercellular adhesion. *Embo Reports* **11**:798-804.

51. **Hug I, Couturier MR, Rooker MM, Taylor DE, Stein M, Feldman MF.** 2010. *Helicobacter pylori* lipopolysaccharide is synthesized via a novel pathway with an evolutionary connection to protein N-glycosylation. *PLoS Pathog* **6**:e1000819.
52. **IARC.** 1994. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Schistosomes, Liver Flukes, and *Helicobacter pylori*. **61**.
53. **Ieva R, Bernstein HD.** 2009. Interaction of an autotransporter passenger domain with BamA during its translocation across the bacterial outer membrane. *Proc Natl Acad Sci* **106**:19120-19125.
54. **Ilver D, Arnqvist A, Ogren J, Frick IM, Kersulyte D, Incecik ET, Berg DE, Covacci A, Engstrand L, Boren T.** 1998. *Helicobacter pylori* adhesin binding fucosylated histo-blood group antigens revealed by retagging. *Science* **279**:373-377.
55. **Ilver D, Barone S, Mercati D, Lupetti P, Telford JL.** 2004. *Helicobacter pylori* toxin VacA is transferred to host cells via a novel contact-dependent mechanism. *Cell Microbiol* **6**:167-174.
56. **Ishijima N, Suzuki M, Ashida H, Ichikawa Y, Kanegae Y, Saito I, Boren T, Haas R, Sasakawa C, Mimuro H.** 2011. BabA-mediated adherence is a potentiator of the *Helicobacter pylori* type IV secretion system activity. *J Biol Chem* **286**:25256-25264.
57. **Jimenez-Soto LF, Clausen S, Sprenger A, Ertl C, Haas R.** 2013. Dynamics of the Cag-type IV secretion system of *Helicobacter pylori* as studied by bacterial co-infections. *Cell Microbiol* **15**:1924-1937.
58. **Jimenez-Soto LF, Kutter S, Sewald X, Ertl C, Weiss E, Kapp U, Rohde M, Pirch T, Jung K, Retta SF, Terradot L, Fischer W, Haas R.** 2009. *Helicobacter pylori* type IV secretion apparatus exploits beta1 integrin in a novel RGD-independent manner. *PLoS Pathog* **5**:e1000684.
59. **Johnson EM, Gaddy JA, Voss BJ, Hennig EE, Cover TL.** 2014. Genes Required for Assembly of Pili Associated with the *Helicobacter pylori* cag Type IV Secretion System. *Infect Immun* **82**:3457-3470.

60. **Jones AC, Logan RP, Foyne S, Cockayne A, Wren BW, Penn CW.** 1997. A flagellar sheath protein of *Helicobacter pylori* is identical to HpaA, a putative N-acetylneuraminylactose-binding hemagglutinin, but is not an adhesin for AGS cells. *J Bacteriol* **179**:5643-5647.
61. **Jonsson K, Guo BP, Monstein HJ, Mekalanos JJ, Kronvall G.** 2004. Molecular cloning and characterization of two *Helicobacter pylori* genes coding for plasminogen-binding proteins. *Proc Natl Acad Sci U S A* **101**:1852-1857.
62. **Joossens JV, Hill MJ, Elliott P, Stamler R, Stamler J, Lesaffre E, Dyer A, Nichols R, Kesteloot H.** 1996. Dietary Salt, Nitrate and Stomach Cancer Mortality in 24 Countries. European Cancer Prevention (ECP) and the INTERSALT Cooperative Research Group
Int J Epidemiol **25**:494-504.
63. **Junker M, Schuster CC, McDonnell AV, Sorg KA, Finn MC, Berger B, Clark PL.** 2006. Pertactin beta-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins. *Proc Natl Acad Sci U S A* **103**:4918-4923.
64. **Kataeva IA, Seidel RD, Shah A, West LT, Li X-L, Ljungdahl LG.** 2002. The Fibronectin Type 3-Like Repeat from the *Clostridium thermocellum* Cellobiohydrolase CbhA Promotes Hydrolysis of Cellulose by Modifying Its Surface. *Appl Environ Microbiol* **68**:4292-4300.
65. **Kavermann H, Burns BP, Angermüller K, Odenbreit S, Fischer W, Melchers K, Haas R.** 2003. Identification and Characterization of *Helicobacter pylori* Genes Essential for Gastric Colonization. *J Exp Med* **197**:813-822.
66. **Keenan J, Oliaro J, Domigan N, Potter H, Aitken G, Allardyce R, Roake J.** 2000. Immune Response to an 18-Kilodalton Outer Membrane Antigen Identifies Lipoprotein 20 as a *Helicobacter pylori* Vaccine Candidate. *Infect Immun* **68**:3337-3343.
67. **Kodaman N, Pazos A, Schneider BG, Piazuolo MB, Mera R, Sobota RS, Sicinschi LA, Shaffer CL, Romero-Gallo J, de Sablet T, Harder RH, Bravo LE, Peek RM, Wilson KT, Cover TL, Williams SM, Correa P.** 2014. Human and *Helicobacter pylori* coevolution shapes the risk of gastric disease. *Proc Natl Acad Sci* **111**:1455-1460.

68. **Koebnik R, Locher KP, Van Gelder P.** 2000. Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Mol Microbiol* **37**:239-253.
69. **Konovalova A, Perlman DH, Cowles CE, Silhavy TJ.** 2014. Transmembrane domain of surface-exposed outer membrane lipoprotein RcsF is threaded through the lumen of β -barrel proteins. *Proc Natl Acad Sci* **111**:E4350-E4358.
70. **Konovalova A, Silhavy TJ.** 2015. Outer membrane lipoprotein biogenesis: Lol is not the end. *Philos Trans R Soc Lond B Biol Sci* **370**.
71. **Kovacs-Simon A, Titball RW, Michell SL.** 2011. Lipoproteins of Bacterial Pathogens. *Infect Immun* **79**:548-561.
72. **Kutter S, Buhrdorf R, Haas J, Schneider-Brachert W, Haas R, Fischer W.** 2008. Protein subassemblies of the *Helicobacter pylori* Cag type IV secretion system revealed by localization and interaction studies. *J Bacteriol* **190**:2161-2171.
73. **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.
74. **Lewenza S, Mhlanga MM, Pugsley AP.** 2008. Novel Inner Membrane Retention Signals in *Pseudomonas aeruginosa* Lipoproteins. *J Bacteriol* **190**:6119-6125.
75. **Lewis LA, Rohde K, Gipson M, Behrens B, Gray E, Toth SI, Roe BA, Dyer DW.** 1998. Identification and Molecular Analysis of *lbpBA*, Which Encodes the Two-Component Meningococcal Lactoferrin Receptor. *Infect Immun* **66**:3017-3023.
76. **Li M, Gray W, Zhang H, Chung CH, Billheimer D, Yarbrough WG, Liebler DC, Shyr Y, Slebos RJ.** 2010. Comparative shotgun proteomics using spectral count data and quasi-likelihood modeling. *J Proteome Res* **9**:4295-4305.
77. **Liechti G, Goldberg JB.** 2012. Outer membrane biogenesis in *Escherichia coli*, *Neisseria meningitidis*, and *Helicobacter pylori*: paradigm deviations in *H. pylori*. *Front Cell Infect Microbiol* **2**:29.

78. **Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR.** 1999. Direct analysis of protein complexes using mass spectrometry. *Nat Biotech* **17**:676-682.
79. **Little E, Bork P, Doolittle R.** 1994. Tracing the spread of fibronectin type III domains in bacterial glycohydrolases. *J Mol Evol* **39**:631-643.
80. **Liu H, Fero JB, Mendez M, Carpenter BM, Servetas SL, Rahman A, Goldman MD, Boren T, Salama NR, Merrell DS, Dubois A.** 2015. Analysis of a single *Helicobacter pylori* strain over a 10-year period in a primate model. *Int J Med Microbiol* **305**:392-403.
81. **Loh JT, Friedman DB, Piazuolo MB, Bravo LE, Wilson KT, Peek RM, Jr., Correa P, Cover TL.** 2012. Analysis of *Helicobacter pylori* cagA promoter elements required for salt-induced upregulation of CagA expression. *Infect Immun* **80**:3094-3106.
82. **Loh JT, Torres VJ, Cover TL.** 2007. Regulation of *Helicobacter pylori* cagA expression in response to salt. *Cancer Res* **67**:4709-4715.
83. **Lundström AM, Blom K, Sundaeus V, Bölin I.** 2001. HpaA shows variable surface localization but the gene expression is similar in different *Helicobacter pylori* strains. *Microb Pathog* **31**:243-253.
84. **Ma Z-Q, Dasari S, Chambers MC, Litton MD, Sobecki SM, Zimmerman LJ, Halvey PJ, Schilling B, Drake PM, Gibson BW, Tabb DL.** 2009. IDPicker 2.0: Improved Protein Assembly with High Discrimination Peptide Identification Filtering. *J Proteome Res* **8**:3872-3881.
85. **Ma Z-Q, Tabb DL, Burden J, Chambers MC, Cox MB, Cantrell MJ, Ham A-JL, Litton MD, Oreto MR, Schultz WC, Sobecki SM, Tsui TY, Wernke GR, Liebler DC.** 2011. Supporting tool suite for production proteomics. *Bioinformatics* **27**:3214-3215.
86. **MacCoss MJ, McDonald WH, Saraf A, Sadygov R, Clark JM, Tasto JJ, Gould KL, Wolters D, Washburn M, Weiss A, Clark JI, Yates JR.** 2002. Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc Natl Acad Sci U S A* **99**:7900-7905.

87. **MacCoss MJ, McDonald WH, Saraf A, Sadygov R, Clark JM, Tasto JJ, Gould KL, Wolters D, Washburn M, Weiss A, Clark JI, Yates JR, 3rd.** 2002. Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc Natl Acad Sci U S A* **99**:7900-7905.
88. **Mahdavi J, Sonden B, Hurtig M, Olfat FO, Forsberg L, Roche N, Angstrom J, Larsson T, Teneberg S, Karlsson KA, Altraja S, Wadstrom T, Kersulyte D, Berg DE, Dubois A, Petersson C, Magnusson KE, Norberg T, Lindh F, Lundskog BB, Arnqvist A, Hammarstrom L, Boren T.** 2002. *Helicobacter pylori* SabA adhesin in persistent infection and chronic inflammation. *Science* **297**:573-578.
89. **Marshall B, Warren JR.** 1984. UNIDENTIFIED CURVED BACILLI IN THE STOMACH OF PATIENTS WITH GASTRITIS AND PEPTIC ULCERATION. *The Lancet* **323**:1311-1315.
90. **Martinez MN, Emfinger CH, Overton M, Hill S, Ramaswamy TS, Cappel DA, Wu K, Fazio S, McDonald WH, Hachey DL, Tabb DL, Stafford JM.** 2012. Obesity and altered glucose metabolism impact HDL composition in CETP transgenic mice: a role for ovarian hormones. *J Lipid Res* **53**:379-389.
91. **Masuda K, Matsuyama S, Tokuda H.** 2002. Elucidation of the function of lipoprotein-sorting signals that determine membrane localization. *Proc Natl Acad Sci U S A* **99**:7390-7395.
92. **McClain MS, Shaffer CL, Israel DA, Peek RM, Jr., Cover TL.** 2009. Genome sequence analysis of *Helicobacter pylori* strains associated with gastric ulceration and gastric cancer. *BMC genomics* **10**:3.
93. **Moonens K, Gideonsson P, Subedi S, Bugaytsova J, Romaõ E, Mendez M, Nordén J, Fallah M, Rakhimova L, Shevtsova A, Lahmann M, Castaldo G, Brännström K, Coppens F, Lo Alvin W, Ny T, Solnick Jay V, Vandebussche G, Oscarson S, Hammarström L, Arnqvist A, Berg Douglas E, Muyldermans S, Borén T, Remaut H.** Structural Insights into Polymorphic ABO Glycan Binding by *Helicobacter pylori*. *Cell Host Microbe* **19**:55-66.
94. **Narita S-i, Tokuda H.** 2007. Amino Acids at Positions 3 and 4 Determine the Membrane Specificity of *Pseudomonas aeruginosa* Lipoproteins. *J Biol Chem* **282**:13372-13378.

95. **Nothhaft H, Szymanski CM.** 2010. Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol* **8**:765-778.
96. **Noto JM, Gaddy JA, Lee JY, Piazuolo MB, Friedman DB, Colvin DC, Romero-Gallo J, Suarez G, Loh J, Slaughter JC, Tan S, Morgan DR, Wilson KT, Bravo LE, Correa P, Cover TL, Amieva MR, Peek RM, Jr.** 2013. Iron deficiency accelerates *Helicobacter pylori*-induced carcinogenesis in rodents and humans. *J Clin Invest* **123**:479-492.
97. **O'Toole PW, Janzon L, Doig P, Huang J, Kostrzynska M, Trust TJ.** 1995. The putative neuraminylactose-binding hemagglutinin HpaA of *Helicobacter pylori* CCUG 17874 is a lipoprotein. *J Bacteriol* **177**:6049-6057.
98. **Odenbreit S, Swoboda K, Barwig I, Ruhl S, Boren T, Koletzko S, Haas R.** 2009. Outer membrane protein expression profile in *Helicobacter pylori* clinical isolates. *Infect Immun* **77**:3782-3790.
99. **Odenbreit S, Till M, Hofreuter D, Faller G, Haas R.** 1999. Genetic and functional characterization of the *alpAB* gene locus essential for the adhesion of *Helicobacter pylori* to human gastric tissue. *Mol Microbiol* **31**:1537-1548.
100. **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. *Proc Natl Acad Sci U S A* **105**:1003-1008.
101. **Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, Resing KA, Ahn NG.** 2005. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics* **4**:1487-1502.
102. **Oleastro M, Cordeiro R, Ferrand J, Nunes B, Lehours P, Carvalho-Oliveira I, Mendes AI, Penque D, Monteiro L, Megraud F, Menard A.** 2008. Evaluation of the clinical significance of *homB*, a novel candidate marker of *Helicobacter pylori* strains associated with peptic ulcer disease. *J Infect Dis* **198**:1379-1387.
103. **Oleastro M, Monteiro L, Lehours P, Megraud F, Menard A.** 2006. Identification of markers for *Helicobacter pylori* strains isolated from children with peptic ulcer disease by suppressive subtractive hybridization. *Infect Immun* **74**:4064-4074.

104. **Ottemann KM, Lowenthal AC.** 2002. *Helicobacter pylori* Uses Motility for Initial Colonization and To Attain Robust Infection. *Infect Immun* **70**:1984-1990.
105. **Pang SS, Nguyen ST, Perry AJ, Day CJ, Panjekar S, Tiralongo J, Whisstock JC, Kwok T.** 2013. The three-dimensional structure of the extracellular adhesion domain of the sialic acid-binding adhesin SabA from *Helicobacter pylori*. *J Biol Chem*.
106. **Pang SS, Nguyen STS, Perry AJ, Day CJ, Panjekar S, Tiralongo J, Whisstock JC, Kwok T.** 2014. The Three-dimensional Structure of the Extracellular Adhesion Domain of the Sialic Acid-binding Adhesin SabA from *Helicobacter pylori*. *J Biol Chem* **289**:6332-6340.
107. **Peck B, Ortkamp M, Diehl KD, Hundt E, Knapp B.** 1999. Conservation, localization and expression of HopZ, a protein involved in adhesion of *Helicobacter pylori*. *Nucleic Acids Res* **27**:3325-3333.
108. **Peek RM, Jr., Blaser MJ.** 2002. *Helicobacter pylori* and gastrointestinal tract adenocarcinomas. *Nat Rev Cancer* **2**:28-37.
109. **Petersen TN, Brunak S, von Heijne G, Nielsen H.** 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**:785-786.
110. **Phadnis SH, Parlow MH, Levy M, Ilver D, Caulkins CM, Connors JB, Dunn BE.** 1996. Surface localization of *Helicobacter pylori* urease and a heat shock protein homolog requires bacterial autolysis. *Infect Immun* **64**:905-912.
111. **Pham KT, Weiss E, Jimenez Soto LF, Breithaupt U, Haas R, Fischer W.** 2012. CagI is an essential component of the *Helicobacter pylori* Cag type IV secretion system and forms a complex with CagL. *PLoS one* **7**:e35341.
112. **Polk DB, Peek RM, Jr.** 2010. *Helicobacter pylori*: gastric cancer and beyond. *Nat Rev Cancer* **10**:403-414.
113. **Pride AC, Herrera CM, Guan Z, Giles DK, Trent MS.** 2013. The Outer Surface Lipoprotein VolA Mediates Utilization of Exogenous Lipids by *Vibrio cholerae*. *mBio* **4**.

114. **Radin JN, Gaddy JA, Gonzalez-Rivera C, Loh JT, Algood HM, Cover TL.** 2013. Flagellar localization of a *Helicobacter pylori* autotransporter protein. *mBio* **4**:e00613-00612.
115. **Rahman R, Asombang AW, Ibdah JA.** 2014. Characteristics of gastric cancer in Asia. *World J Gastroenterol* **20**:4483-4490.
116. **Rappsilber J, Ishihama Y, Mann M.** 2003. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **75**:663-670.
117. **Rohde M, Puls J, Buhrdorf R, Fischer W, Haas R.** 2003. A novel sheathed surface organelle of the *Helicobacter pylori* *cag* type IV secretion system. *Mol Microbiol* **49**:219-234.
118. **Sabarth N, Hurvitz R, Schmidt M, Zimny-Arndt U, Jungblut PR, Meyer TF, Bumann D.** 2005. Identification of *Helicobacter pylori* surface proteins by selective proteinase K digestion and antibody phage display. *J Microbiol Methods* **62**:345-349.
119. **Sabarth N, Lamer S, Zimny-Arndt U, Jungblut PR, Meyer TF, Bumann D.** 2002. Identification of surface proteins of *Helicobacter pylori* by selective biotinylation, affinity purification, and two-dimensional gel electrophoresis. *J Biol Chem* **277**:27896-27902.
120. **Sagulenko V, Sagulenko E, Jakubowski S, Spudich E, Christie PJ.** 2001. VirB7 Lipoprotein Is Exocellular and Associates with the *Agrobacterium tumefaciens* T Pilus. *J Bacteriol* **183**:3642-3651.
121. **Sause WE, Castillo AR, Ottemann KM.** 2012. The *Helicobacter pylori* autotransporter ImaA (HP0289) modulates the immune response and contributes to host colonization. *Infect Immun* **80**:2286-2296.
122. **Schätzle S, Specht M, Waidner B.** 2015. Coiled Coil Rich Proteins (Ccrp) Influence Molecular Pathogenicity of *Helicobacter pylori*. *PloS one* **10**:e0121463.
123. **Seong Kim J, Hoon Chang J, Il Chung S, Sun Yum J.** 1999. Molecular Cloning and Characterization of the *Helicobacter pylori* *fliD* Gene, an Essential Factor in Flagellar Structure and Motility. *J Bacteriol* **181**:6969-6976.

124. **Shaffer CL, Gaddy JA, Loh JT, Johnson EM, Hill S, Hennig EE, McClain MS, McDonald WH, Cover TL.** 2011. *Helicobacter pylori* exploits a unique repertoire of type IV secretion system components for pilus assembly at the bacteria-host cell interface. *PLoS Pathog* **7**:e1002237.
125. **Shevchenko A, Wilm M, Vorm O, Mann M.** 1996. Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels. *Anal Chem* **68**:850-858.
126. **Soutourina OA, Semenova EA, Parfenova VV, Danchin A, Bertin P.** 2001. Control of Bacterial Motility by Environmental Factors in Polarly Flagellated and Peritrichous Bacteria Isolated from Lake Baikal. *Appl Environ Microbiol* **67**:3852-3859.
127. **Stead CM, Zhao J, Raetz CR, Trent MS.** 2010. Removal of the outer Kdo from *Helicobacter pylori* lipopolysaccharide and its impact on the bacterial surface. *Mol Microbiol* **78**:837-852.
128. **Struyvé M, Moons M, Tommassen J.** 1991. Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein. *J Mol Biol* **218**:141-148.
129. **Thissen D, Steinberg L, Kuang D.** 2002. Quick and Easy Implementation of the Benjamini-Hochberg Procedure for Controlling the False Positive Rate in Multiple Comparisons. *J Educ Behav Stat* **27**:77-83.
130. **Thompson LJ, Merrell DS, Neilan BA, Mitchell H, Lee A, Falkow S.** 2003. Gene expression profiling of *Helicobacter pylori* reveals a growth-phase-dependent switch in virulence gene expression. *Infect Immun* **71**:2643-2655.
131. **Tokuda H, Matsuyama S-i.** 2004. Sorting of lipoproteins to the outer membrane in *E. coli*. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1693**:5-13.
132. **Tomb JF, 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, Fitzgerald 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.
133. **Tsugane S.** 2005. Salt, salted food intake, and risk of gastric cancer: epidemiologic evidence. *Cancer science* **96**:1-6.
134. **Van Vliet AHM, Stoof J, Vlasblom R, Wainwright SA, Hughes NJ, Kelly DJ, Bereswill S, Bijlsma JJE, Hoogenboezem T, Vandenbroucke-Grauls CMJE, Kist M, Kuipers EJ, Kusters JG.** 2002. The Role of the Ferric Uptake Regulator (Fur) in Regulation of *Helicobacter pylori* Iron Uptake. *Helicobacter* **7**:237-244.
135. **Voss BJ, Gaddy JA, McDonald WH, Cover TL.** 2014. Analysis of Surface-Exposed Outer Membrane Proteins in *Helicobacter pylori*. *J Bacteriol* **196**:2455-2471.
136. **Voss BJ, Loh JT, Hill S, Rose KL, McDonald WH, Cover TL.** 2015. Alteration of the *Helicobacter pylori* membrane proteome in response to changes in environmental salt concentration. *Proteomics Clin Appl* **9**:1021-1034.
137. **Wu CC, Yates JR, 3rd.** 2003. The application of mass spectrometry to membrane proteomics. *Nat Biotechnol* **21**:262-267.
138. **Xiang Z, Censini S, Bayeli PF, Telford JL, Figura N, Rappuoli R, Covacci A.** 1995. Analysis of expression of CagA and VacA virulence factors in 43 strains of *Helicobacter pylori* reveals that clinical isolates can be divided into two major types and that CagA is not necessary for expression of the vacuolating cytotoxin. *Infect Immun* **63**:94-98.
139. **Xie F, Li G, Zhang W, Zhang Y, Zhou L, Liu S, Liu S, Wang C.** 2016. Outer membrane lipoprotein VacJ is required for the membrane integrity, serum resistance and biofilm formation of *Actinobacillus pleuropneumoniae*. *Vet Microbiol* **183**:1-8.
140. **Yamaguchi K, Yu F, Inouye M.** 1988. A single amino acid determinant of the membrane localization of lipoproteins in *E. coli*. *Cell* **53**:423-432.
141. **Yates JR, Eng JK, McCormack AL, Schieltz D.** 1995. Method to Correlate Tandem Mass Spectra of Modified Peptides to Amino Acid Sequences in the Protein Database. *Anal Chem* **67**:1426-1436.

142. **Yuriev E, Farrugia W, Scott AM, Ramsland PA.** 2005. Three-dimensional structures of carbohydrate determinants of Lewis system antigens: Implications for effective antibody targeting of cancer. *Immunol Cell Biol* **83**:709-717.