STRUCTURAL, FUNCTIONAL, AND PHYLOGENETIC ANALYSES OF THE *HELICOBACTER PYLORI* VACUOLATING TOXIN (VACA)

By

Kelly A. Gangwer

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

August, 2010

Nashville, Tennessee

Timothy L. Cover

Terence S. Dermody

Andrzej M. Krezel

H. Earl Ruley

Dana Borden Lacy

To my beloved husband, Seth, an infinite supply of encouragement, a faithful listener, but most importantly, my hero.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. D. Borden Lacy, who has been both a mentor and friend throughout this long and sometimes difficult journey. I had the unique opportunity to join Borden's lab a few months after she started her faculty position. She is a true inspiration, teaching me not only about science, but also how to juggle a family and a career. Throughout these past five years, she always had an "open door", even after we multiplied in numbers. When times were stressful, she provided gentle and enthusiastic words of encouragement. After many trips to her office in tears, I always left with a smile and a new vision of hope. Her devoted mentoring is responsible for much of my scientific and personal accomplishments. For this, I cannot thank her enough. I am forever grateful. Thank you Borden!

I had the opportunity and privilege to work together with helpful lab members that provided not only scientific advice but also some laughs along the way. Thank you Darren for showing me the initial ropes so that I could navigate through the Megalab and for providing me with the start to my scientific discoveries. Thank you Rory for making our lab a more enjoyable place to look at because of your artistic talent and creativity with orange pipet tip holders. Thank you Nicole for providing us with delectable baked goods that gave us the energy needed to get through the day. To the rest of my lab members, Mitch, Stacey, Melissa, and Desiree, thank you for patiently listening to me complain about my dissertation, my dog, and so many other important issues. Also, I would like to thank Christian Gonzalez, a student in Dr. Timothy Cover's lab, who served as a valuable collaborator. It was a pleasure sharing knowledge, ideas, reagents, and results. Furthermore, I was fortunate to have the opportunity to work together with the

Spiller lab. Specifically, Dr. Ben Spiller was instrumental in technical advice and constructive criticism.

I am thankful to Dr. Timothy “Tim” Cover, who served as my thesis committee chair, and also as a collaborator. It was such a pleasant experience interacting with Tim on a weekly basis. He was always enthusiastic and optimistic about science and I strive to emulate him in my own scientific career.

I am grateful for the remaining members of my dissertation committee, Dr. Terry Dermody, Dr. Andrzej Krezel, and Dr. Earl Ruley. Thank you for providing constructive criticism, encouragement, constant support and advice for future career goals. It was a pleasure to share my research with a group of extraordinary scientists and I have learned much about scientific debate. Furthermore, I had the privilege of obtaining outstanding training and education in the Department of Microbiology and Immunology and the Molecular Biophysics training program. I was fortunate to have access to many valuable resources and to be surrounded by excellent administrative personnel and professors.

I would like to thank my friends that have completed this journey with me and encouraged me when it seemed like I was never going to graduate and get a “real” job. To the Dove and Chalos family, thank you for adopting me into your families. I am forever grateful to them for making Nashville feel like home. I would specifically like to thank Tracy and Amanda Wilkins, whom I have known for the past two years, and have become more than just friends. I can’t say enough about what an incredible support system they have been to me in such a fragile time of my life. I will always cherish the memories we have made, and the good and bad times that we have shared. Thank you for always being there for me, going to dinner with me, keeping me company, and helping me get through some tough times. Love you guys!

Of course no acknowledgements would be complete without giving thanks to my dear family. My mom and dad have instilled many admirable qualities in me and given me a good foundation with which to meet life. Even though they are not scientists, they have taught me about hard work and self-respect, about persistence and perseverance. My mom provided compassion and never stopped believing in me. My dad has been my rock. He has always taken the time to listen to me and he has gently nudged me throughout my entire academic and personal life. He is my biggest fan. Each one of my parents, all six of them, have always expressed how proud they are of me, how they believe that I can be whomever I want and they always keep on loving me.

Lastly, I owe everything to my wonderful husband, Seth. He has been my constant companion throughout this whole journey. He has been in the seat next to me on this roller coaster, through every corner, every loop, every high and every low. We began our life together 11 years ago and I would never have been able to accomplish this feat without him. He supported me when I was in high school, when I decided on a college, when I was an undecided major and wanted to dabble in different academic areas, and when I found my true passion and transformed into a scientist. He has patiently endured many hours alone, while I worked late nights and weekends in lab. He wouldn't hesitate to bring me soda, lunch or just his company. After an unsuccessful experiment, I was fortunate to have an open ear, even though sometimes he had no idea what I was talking about. He is my honorary scientist because he would listen patiently as I practiced my seminars at home. He would never complain when I would start the first slide over and over again, looking for those perfect words for the title slide. Finally, he has provided me with inspiration, because he has always followed his dreams, worked hard to achieve his goals, and has displayed an incredible amount of respect for me. I love you Seth!

These studies were supported by the Burroughs Wellcome Fund, NIH (R01AI39657) and Vanderbilt Development funds. I was supported by the Molecular Biophysics training program (T32 GM08320) and provided with a travel grant from the graduate school to attend the 5th International Conference on Structural Analysis of Supramolecular Assemblies by Hybrid Methods.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

APS	Advanced Photon Source
CagA	Cytotoxin-associated gene
CHO	Chinese hamster ovarian
DDM	n-dodecyl beta-D-maltoside
EM	Electron microscopy
FBS	Fetal bovine serum
Git1	G protein-coupled receptor kinase interactor-1
GST	Glutathione S-transferase
His	Histidine
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kan	Kanamycin
kDa	Kilodaltons
KPi	Potassium phosphate
LFA-1	Lymphocyte function-associated antigen
MRC	Mixed raster content
OD	Optical density
OM	Outer membrane
MLST	Multi-locus sequence typing
PMA	Phorbol 12-myristate 13-acetate
RPTP	Receptor-linked protein tyrosine phosphatase
SH	Shimodaira-Hasegawa
SM	Sphingomyelin
TB	Terrific broth
VacA	Vacuolating cytotoxin

CHAPTER I

INTRODUCTION

Helicobacter pylori

As early as 1938, spiral shaped bacteria were reported to colonize the human gastric mucus layer (1, 2). Since the organisms could not be grown in culture, however, and since the stomach was considered an inhospitable environment, the results were largely ignored. In 1982, Barry J. Marshall and J. Robin Warren isolated the bacteria and made the remarkable and unexpected discovery that inflammation and ulceration in the stomach as well as ulceration of the duodenum can be the result of an infection caused by the gram-negative bacterium *Helicobacter pylori* (3, 4). When Warren and Marshall announced their findings, it was a long-standing belief in medical practice that stress and lifestyle were the major causes of peptic ulcer disease. In response to critics, Marshall underwent gastric biopsy to verify that he didn't carry the bacterium, then deliberately infected himself by ingesting a liquid culture to show that *H. pylori,* in fact, caused acute gastric illness (5). Biopsy confirmed an infection that could then be treated with antibiotics. It soon became clear that *H. pylori* causes more than 90% of duodenal ulcers and up to 80% of gastric ulcers (6-12). Warren and Marshall were awarded the Nobel Prize in Medicine in 2005 for “their discovery *of Helicobacter pylori* and its role in gastritis and peptic ulcer disease.”

More than half of the world’s human population is persistently colonized with *H. pylori* (13). In developing countries, 70 to 90% of the population carries *H. pylori*, and in developed countries the prevalence of infection is lower, ranging from 25 to 50% (14). *H. pylori* penetrates and colonizes the human gastric mucosa, an inherently inhospitable acidic environment. *H. pylori* is highly adapted to this ecologic niche and may persist for

decades, despite the host immune and inflammatory responses. In the absence of *H. pylori*, the gastric mucosa contains very few immune or inflammatory cells, however, all *H. pylori*-infected individuals develop chronic gastric inflammation (15, 16). While often asymptomatic, the chronic inflammation associated with infection makes *H. pylori* a significant risk factor for the development of several gastroduodenal diseases. These include gastric adenocarcinoma, gastric and duodenal ulceration, gastric MALT lymphoma, and gastric non-Hodgkins lymphoma (4, 11, 12, 17-25). The biological determinants of clinical outcome are poorly understood, in part because of genetic heterogeneity among *H. pylori* isolates, but also due to heterogeneity within the human hosts and variable environmental factors.

The complete genome sequences of multiple *H. pylori* strains have been determined (10, 26-30). These *H. pylori* genomes consist of circular chromosomes about 1.6 to 1.73 Mb in size (31). This relatively small size probably reflects a limited metabolic repertoire and biosynthetic capacity and is consistent with the specialization of *H. pylori* growth in its restricted niche (32). *H. pylori* strains from unrelated humans exhibit a high level of genetic diversity, and the rate of recombination in *H. pylori* is reported to be among the highest in the Eubacteria (17, 33). For example, a single host can carry multiple closely related *H. pylori* strains that are derived from one parental strain by a series of mutations, chromosomal rearrangements, and recombination events (34). The population structure of *H. pylori* is panmictic (17, 35-37). A panmictic structure is caused by recombination and horizontal genetic exchange, where foreign DNA, from intraspecies recombination, is integrated and disrupts clonal structure, resulting in newly introduced DNA that does not share the same evolutionary history as the other genes within the genome. Furthermore, *H. pylori* exhibits sequence diversity in multiple genes, including those that encode urease (38) and accessory proteins (39),

flagellin (40), vacuolating cytotoxin (VacA) (41) and cytotoxin-associated gene (CagA) (42-44).

H. pylori virulence factors

In gastric biopsy specimens, *H. pylori* are 2.5 to 5.0 μm long, 0.5 to 1.0 μm wide, and have four to six unipolar sheathed flagella (45). The lumenal pH of the human stomach is < 2, but within the gastric mucus there is a pH gradient that ranges from pH 2 to pH 7.0 (46). *H. pylori* entry into the gastric mucus layer is presumably an important feature that permits escape from the extremely low pH environment of the stomach. Flagella are essential for moving through the stomach lumen and penetrating the mucus layer, which then allows for colonization (47).

Secreted bacterial products are thought to play an important role in the pathogenesis of *H. pylori*-associated diseases. Secreted proteins are of special interest, because they come in direct contact with host tissues and may mediate important host - pathogen interactions. Urease is one important colonization factor that neutralizes the pH environment for the organism by catalyzing the hydrolysis of urea to ammonium and CO_2 (48). The enzyme is produced in abundance by all fresh clinical *H. pylori* isolates but not required for *H. pylori* growth *in vitro* (49-52).

A second important *H. pylori* virulence factor is the secreted CagA effector protein. The *cagA* gene is localized within a 40-kb chromosomal region known as the *cag* pathogenicity island (*cag* PAI) (53). *H. pylori* strains expressing CagA are associated with a significantly increased risk for development of ulcer disease or gastric cancer compared to strains that lack the *cagA* gene (54). Upon entry into cells, CagA undergoes phosphorylation by host cell kinases and induces numerous alterations in cellular signaling, leading to the designation of CagA as a “bacterial oncoprotein” (55-59).

Another important virulence factor in the pathogenesis of *H. pylori* infection is a secreted protein known as vacuolating cytotoxin (VacA). Soon after the discovery of *H. pylori* it was reported that a protein in *H. pylori* broth culture filtrates could cause the formation of large intracellular vacuoles in the cytoplasm of mammalian cells (60-62). A chromosomal gene known as *vacA* encodes the *H. pylori* protein responsible for this effect, vacuolating cytotoxin or VacA. The amino acid sequence of VacA does not have similarity to any other known bacterial or eukaryotic protein. Mature VacA toxin molecules are secreted as soluble proteins into the extracellular space, but can also remain localized on the surface of *H. pylori*.

Role of VacA in disease

All strains of *H. pylori* that have been isolated from humans (*H. pylori* is isolated from nonhuman primates as well) contain the *vacA* gene, which suggests that production of VacA is important for colonization or persistence of *H. pylori* in the human stomach. While all strains of *H. pylori* that have been isolated contain the *vacA* gene, not all VacA proteins are cytotoxic *in vitro*, which is due to allelic diversity in two regions: the signal sequence region, or s region, and the midregion, referred to as the m region (Figure 1) (63, 64). VacA molecules of the s2 type are inactive in assays for vacuolating cytotoxicity, type s1/m1 forms produce extensive vacuolation of many different cell types and type s1/m2 forms produce detectable vacuolation in a more limited range of cell types (63, 65-69). Strains of *H. pylori* that contain certain allelic forms of *vacA* are associated with an increased risk of symptomatic gastroduodenal disease compared with strains containing other allelic forms of *vacA* (63, 68-71). In particular, *H. pylori* strains that contain *vacA* alleles of the s1 type are associated with an increased risk for development of peptic ulcer disease and gastric cancer compared with strains containing *vacA* alleles of the s2 type (63, 64, 69, 70). These data can be correlated with the failure

of type s2 forms of VacA to cause detectable vacuolation *in vitro*. In addition, strains containing type m1 *vacA* alleles are associated with an increased risk for development of gastric epithelial injury and gastric cancer compared with strains containing *vacA* alleles of the m2 type (70, 72).

To experimentally investigate whether the production of VacA contributes to gastric disease, large quantities of the VacA protein have been administered directly into the stomach of mice (73). This procedure causes gastric mucosal injury and gastric inflammation. Another approach has been to infect animals with isogenic wildtype and *vacA*-null mutant strains of *H. pylori* and compare the disease pathologies. Gastric histology of infected gnotobiotic piglets suggests that VacA does not contribute to disease (74), whereas it does contribute to the pathogenesis of gastric ulceration in gerbils (75). The lack of a good animal model has made understanding the role of VacA in pathogenesis difficult to study.

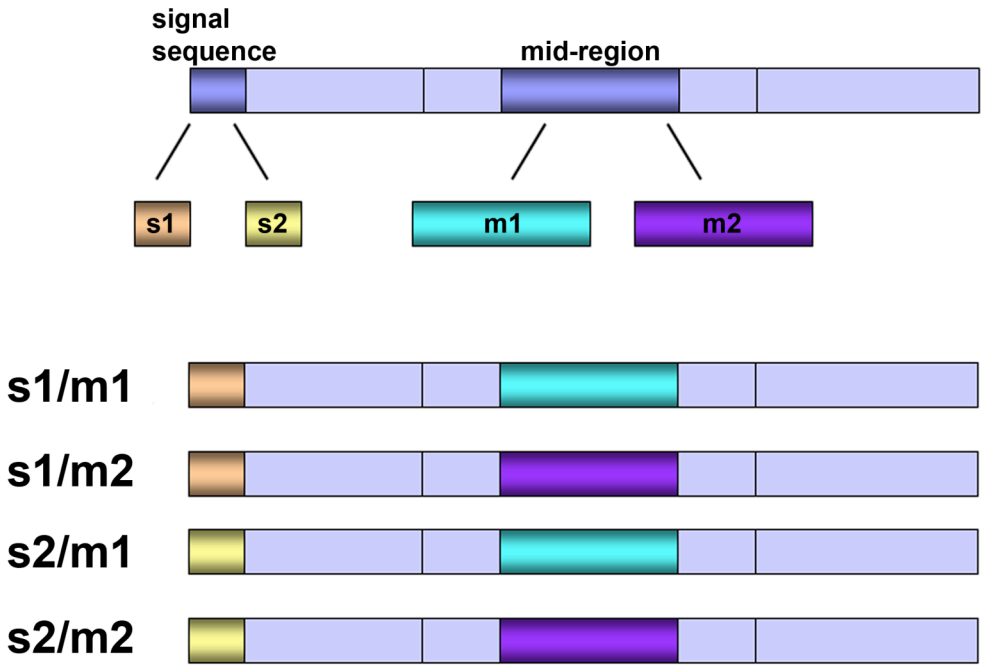


Figure 1. *vacA* gene structure. There is a high level of diversity among *vacA* alleles from different *H. pylori* strains. Allelic diversity is particularly striking near the 5' terminus of vacA (the s-region) and in the mid-region of the gene (m-region). Two main families of s-region sequences (s1 and s2) and two main families of m-region sequences (m1 and m2) have been described. Adapted from Cover *et al.* 2005 (57).

VacA effects on cells

In order to understand how VacA contributes to *H. pylori* colonization of the stomach and the development of gastroduodenal disease, the effects of VacA on human cells have been investigated *in vitro*. VacA can infect many different cell types, including gastric epithelial cells and several types of immune cells, resulting in alteration of cellular functions. Many of these activities are dependent on the formation of anion-selective channels in the membranes of host cells (61, 76-80). VacA is named for is capacity to induce extensive vacuolation in the cytoplasm of mammalian cells (55). These vacuoles arise from anion-selective channels in endosomal membranes followed by swelling of endosomal compartments (61, 78, 79, 81). To form vacuoles in response to purified VacA, a weak base such as ammonium chloride is needed (56, 82, 83). In addition to vacuolation, other channel dependent effects include depolarization of epithelial cells (84), transfer of ions and nutrients (85, 86), inhibition of T cell proliferation (87-89), and the release of cytochrome c from mitochondria resulting in the induction of epithelial cell apoptosis (90). Furthermore, the toxin is able to perturb a number of signaling pathways, independent of its channel-forming activity. For example, VacA is known to activate the G-protein coupled receptor kinase interacting protein (Git)1 and p38 (91, 92).

VacA primary structure and secretion

Gram-negative bacteria have evolved six secretion pathways (type I-VI) for the transport of proteins across the cell membrane (93, 94). They range from relatively simple structures such as type I secretion systems, composed of three subunits that only secrete one substrate protein, to complex machines such as type III and IV secretion systems, composed of more than 20 subunits that can translocate large sets of effector proteins into eukaryotic target cells.

The *vacA* gene encodes a 140-kDa protoxin, which undergoes N- and C-terminal cleavage during the secretion process to yield a mature 88-kDa toxin, p88 (41, 73, 95). Secretion is thought to occur through a Type Va or autotransporter pathway in which an N-terminal signal sequence directs the protein to the periplasm via the Sec pathway and a C-terminal β-barrel domain facilitates transport of the 'passenger' domain across the outer membrane (Figure 2) (96). Outer membrane secretion occurs in an energy independent mechanism. Following secretion, the passenger domain is cleaved from the autotransporter domain either by its own protease activity (currently none are known for VacA) or by another protease.

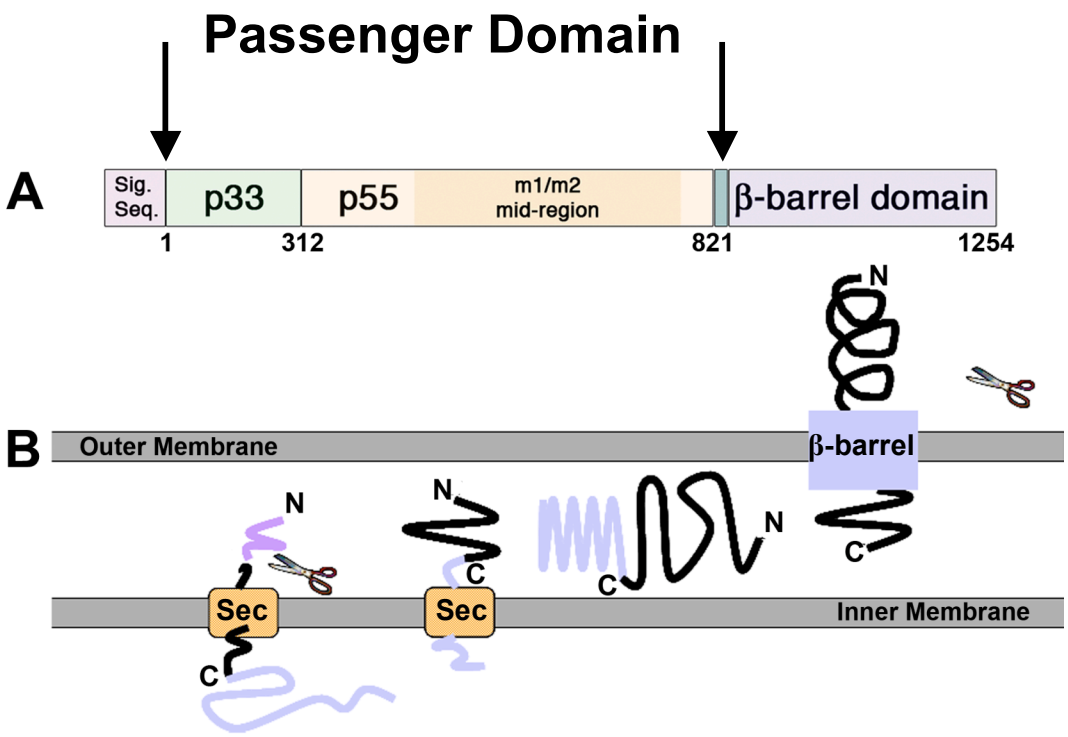


Figure 2. VacA structure and model of VacA translocation across the outer membrane. (A) The *vacA* gene encodes a 140-kDa protoxin. The passenger domain is the mature 88-kDa VacA toxin, which contains two domains, designated p33 and p55. The midregion sequence that defines type m1 and m2 forms of VacA is located within p55. (B) The signal sequence (purple) directs the protein to the periplasm through the Sec machinery. Following translocation into the periplasm and cleavage of the N-terminal signal peptide, the C-termianl β-barrel domain (blue) folds into the outer membrane. The passenger domain (black) is translocated across the outer membrane in an unfolded or partially folded state. Following export, or possibly concurrent with translocation onto the cell surface, folding of the passenger domain occurs. Cleavage of the VacA passenger domain is mediated by an unknown mechanism. Orange box in the inner membrane represents the Sec machinery. Blue box in the outer membrane represents the β-barrel domain.

VacA receptor binding

For most bacterial toxins, cellular intoxication is initiated by binding of the toxin to a receptor on the plasma membrane. Several bacterial toxins, such as diphtheria toxin (97), bind a single receptor on the surface of the host cell. In contrast, VacA has been reported to bind to multiple cell-surface components. These include two receptor-linked protein tyrosine phosphatase proteins, RPTP-α and RPTP-β (91, 98, 99). These proteins do not share sequence homology in their extracellular domains but could contain a common receptor-binding site if a common sugar moiety is involved. RPTP-β has been shown to mediate VacA induced gastric damage in a mouse intoxication model (91). Mice deficient in RPTP-β do not show mucosal damage by VacA, although gastric epithelial cell vacuolation still occurs. Instead, the RPTP-β mediated effect on Git1 signaling is proposed as the mechanism of VacA-induced gastric damage (91). The RPTP-α and RPTP-β receptors are present in both epithelial cells and T cells but it is unclear which cellular events they mediate and if they discriminate between m1 and m2 forms of VacA. Additionally, the β2 (CD18) integrin subunit has been identified as a leukocyte-specific receptor for m1-VacA on human T cells (100). VacA efficiently enters activated, primary human T lymphocytes by binding to the CD18 integrin subunit and exploiting the recycling of lymphocyte function-associated antigen (LFA)-1. Furthermore, various lipids have been shown to interact with VacA, but the importance of VacA-lipid interactions for toxin function remains to be established (61, 101-103). Most recently, sphingomyelin (SM), a key structural component of the cell membrane, has been shown to modulate the sensitivity of epithelial cells to VacA (104). It still remains unclear how SM-dependent association of VacA promotes toxin uptake.

Studies of the VacA interaction with the cell surface have been complicated by the potential for this toxin to interact with multiple cell-surface components and for the

relevant cell-surface receptors to vary among different cell types. In addition, type m1 and m2 VacA have differences in cell type specificity and might preferentially bind to different cell-surface components. Lastly, our understanding of receptor binding has been complicated by the potential for VacA to interact with cells either as a monomer or as an oligomer.

VacA oligomerization and pore-formation

It is hypothesized that oligomerization is required for VacA pore-formation (105). VacA can assemble into a variety of water-soluble oligomeric structures, including single-layered hexamers and heptamers and double-layered structures consisting of 12 or 14 subunits (60, 61, 106-109). The structures resemble 'flowers' in which a central ring is surrounded by peripheral 'petals' (Figure 3). Similar oligomeric structures have been visualized on the surface of VacA-treated cells or lipid bilayers. While the highest resolution images are 19 Å cryo-EM maps of water-soluble VacA dodecamers (107), atomic force microscopy studies suggest that membrane-associated VacA channels are probably single-layered structures (61). Moreover, water-soluble VacA oligomeric complexes lack cytotoxic activity unless they are first dissociated into monomeric components by exposure to low pH or high pH conditions (98, 101, 106, 110).

VacA can then insert into the plasma membrane to form channels (78), or can undergo endocytosis and eventually localize with late endocytic compartments or mitochondria (90, 111). VacA-induced vacuoles correspond to swollen late endosomes, which are presumed to arise as a consequence of anion flux through VacA channels in the membranes of these compartments (58, 112).

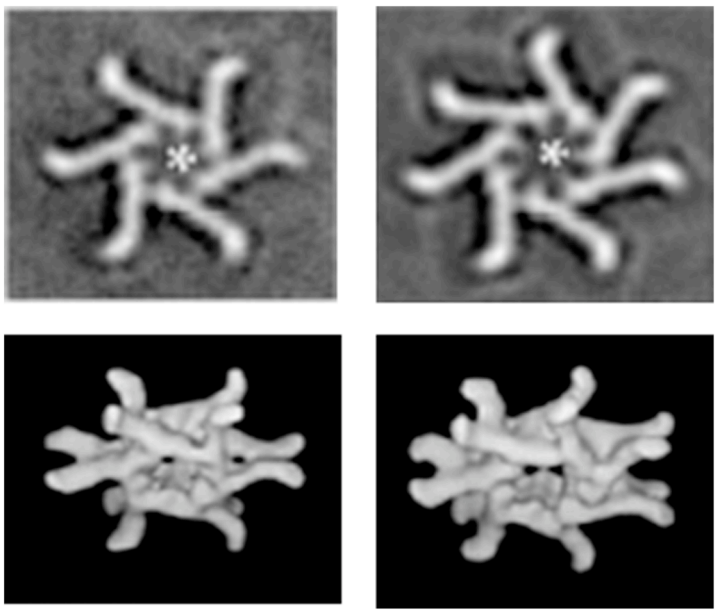


Figure 3. Structural analysis of *H. pylori* VacA toxin by cryo-negative staining EM. (*top panel*) Class averages resulting from 2D analysis of cryo-negatively stained VacA p88 particles showing well-contrasted six- or seven-sided forms. (*bottom panel*) Refined 3D reconstructions of the dodecameric and tetradecameric forms. Adapted from El-Bez *et al.* (107)

VacA domains

Two domains of VacA, p33 and p55, have been identified based on partial proteolysis of p88 into fragments of 33 kDa and 55 kDa, respectively (Figure 2A) (73). When expressed independently and then mixed, p33 and p55 can reconstitute vacuolating toxin activity (113-115). The N-terminal p33 domain (residues 1-311) contains a hydrophobic sequence (residues 6-27) involved in pore formation (105, 116), while the p55 domain (residues 312-821) contains one or more cell-binding domains (117-119). Additionally, the m1/m2 allelic forms of VacA map to the p55 region. When expressed intracellularly, the minimum portion of VacA required for cell-vacuolating activity comprises the entire p33 domain and about 110 amino acids from the N-terminus of p55 (115). Amino acid sequences within both the p33 domain (residues 49-57) (120) and p55 domain (residues 346-347) (121, 122) are required for assembly of VacA into oligomeric structures, and mutant proteins lacking these sequences fail to cause cell vacuolation. Several VacA mutant proteins have dominant-negative inhibitory effects on the capacity of wild-type VacA to cause cellular alterations (120, 123), which further supports the hypothesis that oligomeric structures are required for VacA effects on host cells.

Research objectives

Herein, I describe structural, functional, and phylogenetic analyses used to assess the mechanisms by which VacA interacts with host cells. Using X-ray crystallography, I found that the VacA m1 p55 domain adopts a β-helix fold that is characteristic of autotransporter passenger domains but unique among bacterial protein toxins (Chapter II). Analysis of VacA sequence variation in the context of the structure has allowed me to propose new models about how receptor binding and oligomerization occur (Chapter II). Efforts to obtain structural information on the VacA p33 domain

responsible for pore formation have been hindered by the insolubility of p33. Recently, Christian Gonzalez, a student in Dr. Cover's lab, has made progress in refolding the p33 domain from inclusion bodies. Christian and I have been collaborating to show that this refolded protein is well ordered and functional (Chapter III) and have developed a tractable path for crystallographic studies (Chapter V, Future Directions). Finally, the elongated β-helical structure of VacA presents a unique opportunity for studying the pressures of positive and negative selection in a structural context. I collaborated with Dr. Seth Bordenstein and Dr. Timothy Cover to study the evolution of VacA at the phylogenetic level (Chapter IV). We have demonstrated that amino acid substitutions in VacA are under strong, diversifying selection. Moreover, we describe divergence of VacA sequences into groups with distinct geographic distributions, which suggests independent bouts of VacA adaptive evolution within different human populations. Together, these studies have significantly advanced our understanding of the structural and functional correlates of VacA sequence diversity and mechanism.

CHAPTER II

CRYSTAL STRUCTURE OF THE *HELICOBACTER PYLORI* VACUOLATING TOXIN P55 DOMAIN

Introduction

Analysis of *H. pylori* strains isolated from unrelated humans indicates a high level of genetic diversity among *vacA* alleles (17, 63, 124, 125). Although frequent recombination events have eliminated phylogenetic structure from the 5' region of *vacA* alleles (17, 125), phylogenetic analysis of sequences from a *vacA* midregion indicates the existence of two large families of sequences, termed m1 and m2 (see Figure 1, Chapter I) (63, 124). Within this ~281-aa midregion, the amino acid sequences of types m1 and m2 VacA proteins are only ~55% identical. Differences in cell-type specificity have been noted for types m1 and m2 VacA proteins, a phenomenon that may result from binding of m1 and m2 VacA proteins to different cell-surface receptors (68, 126-128). With the exception of a few m1/m2 chimeras, there has been little evidence of recombination between m1 and m2 *vacA* alleles within the *vacA* midregion, and, therefore, a phylogenetic distinction between m1 and m2 VacA sequences has remained intact (124). This distinction is the basis for a widely used typing scheme for *H. pylori* isolates (63, 124) and is relevant clinically because m1 strains are associated with gastric cancer (70). To gain insights into the structural features of VacA that contribute to its cell-binding properties and to better understand the structural correlates of VacA sequence variation, we set out to determine the structure of the VacA p55 domain.

In this study, we determined a crystal structure of the VacA p55 domain refined to 2.4-Å resolution. The structure is predominantly a right-handed parallel β-helix. We

analyzed VacA protein sequences from unrelated *H. pylori* strains, including m1 and m2 forms of VacA, in the context of the structure and identified structural features of the VacA surface that may be important for interactions with host receptors. Furthermore, we docked the p55 structure into a 19-Å cryo-EM map of a VacA dodecamer and proposed a model for how VacA monomers assemble into oligomeric structures capable of membrane channel formation.

Methods

Expression and Purification of p55. A plasmid encoding the p55 domain of VacA from *H. pylori* strain 60190 (a type m1 form of VacA; GenBank accession number Q48245) with an N-terminal hexahistidine tag (113) was transformed into *E. coli* BL21(DE3). Cultures were grown at 37 °C overnight with shaking in Terrific broth (TB) supplemented with 50 ug/ml of Kanamycin (KAN). Overnight cultures were diluted 1:100 in fresh TB-Kan and grown at 37 °C for 2.5 h to an OD_{600} of 0.6-0.7. Isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.6 mM, and the proteins were allowed to express for 20 h at 23 °C. Harvested cells were resuspended in lysis buffer (50 mM KPhosphate, 300 mM NaCl, 20 mM imidazole at pH 8.0) using a pre-chilled Dounce homogenizer and lysed by three passages through a pressure dispersion homogenizer at 20K psi. Cell lysates were centrifuged at 20K rpm (48,000 x g) for 20 min at 4 °C. Wildtype p55 was purified by Ni-affinity, ion exchange and gel filtration chromatography and concentrated to 5.5 mg/ml in a 100 mM NaCl, 50 mM KPi buffer. The sizing column retention time suggested a molecular weight of ~130 kDa. We hypothesize that this result reflects the presence of an elongated p55 dimer. Seleno-methionine p55 was produced from *E.coli* BL834(DE3) in a minimal medium (M9 salts plus glucose, amino

acid supplements, vitamins and minerals) in which 40 mg/liter L-cysteine and L-seleno-methionine were added. Purification methods for seleno-methionine p55 were similar to those used for the native protein, except 1 mM DTT and 5 mM methionine were incorporated to all buffers before affinity chromatography.

Crystallization and Preparation of Heavy Atom Derivatives. The native and seleno-methionine p55 were concentrated in a 100 mM NaCl, 50 mM KPhosphate buffer and crystallized at 21 °C by the hanging-drop vapor diffusion method in which protein and precipitant were mixed in a 1:1 ratio. Crystals grew from starting protein concentrations of 5.5 mg/ml and reservoirs containing 18-23% PEG1500 (Figure 4A). Heavy atom derivatives were prepared by soaking crystals in 50 mM potassium tetrabromoplatinate (IV) (K_2PtBr_6) for 3 d, 5 mM 1,4-diacetoxymercuri-2,3-dimethoxybutane for 24 h, or 5 mM gold chloride ($HAuCl_4$) for 24 h. Crystals were mounted on cryo loops, sequentially soaked in cryoprotectant solutions containing 25% PEG1500, and 5%, 10%, and 15% glycerol, and flash-cooled in liquid nitrogen.

Structure Determination and Refinement. X-ray data were collected from single crystals at 100 K on beamline ID-22 at the Advanced Photon Source (Argonne, IL) by using a MAR300 Image Plate Detector. All the diffraction data were indexed, integrated, scaled, and merged with HKL2000 (Figure 4B, Table 1) (129). Native and derivative crystals were spacegroup $P3_221$ and had unit cell dimensions of a=b=57 Å and c=260 Å. Phases were determined by MIRAS using a native protein crystal and seleno-methionine, platinum, mercury, and gold derivatives (Table 1). Heavy atom positions were located and refined with the autoSHARP program suite (130). The structure was subjected to iterative rounds of model building in O (131) and refinement in CNS (132) before applying TLS refinement (133) in REFMAC (134). The refined model (R_{cryst}=18.1%, R_{free}=22.2%) consists of amino acids 355-366, 377-811 and 206 water molecules. SDS-PAGE and N-terminal sequence analysis of dissolved crystals

indicated that proteolysis had occurred at the N-terminus of p55 during crystallization, thus explaining the absence of residues 312-354 and 367-376 from the crystal structure. Proteolysis could be prevented by the addition of protease inhibitors, but crystals did not form under these conditions.

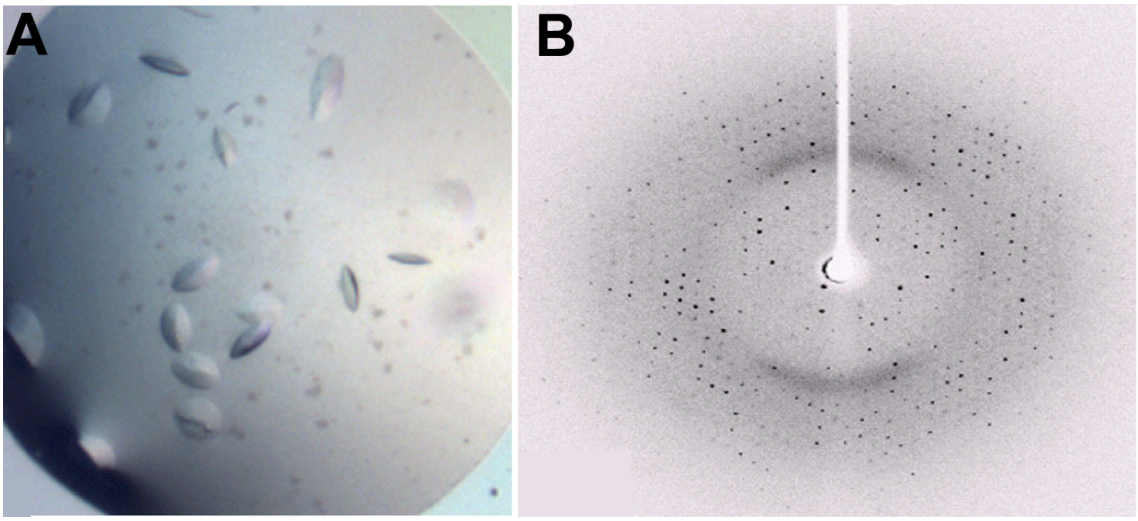


Figure 4. Crystals of VacA p55 diffract to 2.4 Å. (A) Crystals of the VacA p55 domain can be grown reproducibly in hanging drops to dimensions of ~ 200 x 150 x 50 µm. Crystals grow from precipitate after 10 days in 18-23 % PEG1500. (B) The crystals diffract to 2.4 Å resolution at the APS synchrotron X-ray source.

Table 1. X-ray data collection and refinement statistics

Data Collection	P55 native	Selenium	Mercury	Platinum	Gold
Wavelength, Å	0.979	0.979	1.001	1.071	1.040
Resolution (outer shell), Å	50.0-2.4 (2.5-2.4)	50.0-2.9 (3.0-2.9)	50.0-2.9 (3.0-2.9)	50.0-3.6 (3.8-3.6)	50.0-2.9 (3.0-2.9)
R_{merge}*, %	7.6 (41.1)	11.0 (41.0)	14.4 (28.5)	11.3 (25.0)	10.9 (36.8)
Mean I/σI	11.8 (2.4)	15.5 (3.0)	10.0 (3.5)	10.7 (2.6)	12.0 (2.3)
Completeness, %	95.2 (95.0)	99.5 (95.8)	93.4 (91.3)	97.6 (82.1)	98.3 (91.9)
Redundancy	3.9 (3.5)	8.3 (5.4)	3.8 (2.7)	4.4 (2.0)	5.2 (3.6)
Unique observations	19,916 (1,705)	20,530 (2,000)	25,748 (2,144)	10,578 (901)	19,584 (1921)
Phasing Power					
ISO [acentric/centric]		0.803/0.683	0.591/0.542	0.515/0.441	0.495/0.448
ANO		0.860	0.716	0.333	0.674

Refinement	
R_{cryst}/ R_{free}, % †	19.5/23.8
No. protein atoms	3353
No. solvent waters	123
Bond length rmsd, Å	0.029
Bond angle rmsd, °	2.511
Avg. protein B, Å²	59.6
Ramachandran plot, %	
Most favored regions	82.9
Allowed regions	14.3
Generously allowed regions	2.6
Disallowed regions	0.3

Outer resolution bin statistics are given in parentheses.

*R_{merge} = $S_{hkl}(S_i|I_{hkl,i} - \langle I_{hkl}\rangle))/S_{hkl,i}\langle I_{hkli}\rangle$, where $I_{hkl,i}$ is the intensity of an individual measurement of the reflection with Miller indices h, k, and l, and $\langle I_{hkl}\rangle$ is the mean intensity of that reflection.

†R_{cryst} = $S||F_{obs,hkl}| - |F_{calc,hkl}||/|F_{obs,hkl}|$, where $|F_{obs,hkl}|$ and $|F_{calc,hkl}|$ are the observed and calculated structure factor amplitudes. R_{free} is equivalent to R_{cryst} but calculated with reflections (5%) omitted from the refinement process.

‡Calculated with the program PROCHECK (135).

Results and Discussion

Crystal structure of the VacA p55 domain

The crystal structure of the VacA p55 domain was determined to a minimum Bragg spacing of d_{min} = 2.4 Å using experimental phases from multiple isomorphous replacement and anomalous scattering (Table 1). The p55 structure is predominantly a right-handed parallel β-helix (residues 355-735) but has a small globular domain at the C-terminus (residues 736-811) with mixed α/β secondary structure elements (Figure 5). The structure resembles a 'sock' in which the C-terminal domain curves from the 'heel' and extends to the tip of the 'foot'. The β-helical 'calf' is 65 Å long with widths of 25-31Å, while the C-terminal 'foot' is approximately 13 x 17 x 42 Å. In the crystal, pairs of p55 molecules meet at their N-terminal strands about a crystallographic two-fold axis (Figure 6). This packing is consistent with the elution of p55 as a dimer from gel filtration columns (see Methods) and a previous EM study indicating that a p55 domain secreted by *H. pylori* formed dimers (118). VacA lacks sequence homology with other known proteins but is structurally similar to other β-helix protein structures. The closest structural homologs, as assessed by the three-dimensional search algorithm DALI, are the inulin fructotransferase (2inu-A, Z = 13.9), P22 tailspike protein (1tyv, Z = 13.1), and the autotransporter protein P.69 pertactin (1dab-A, Z = 12.5).

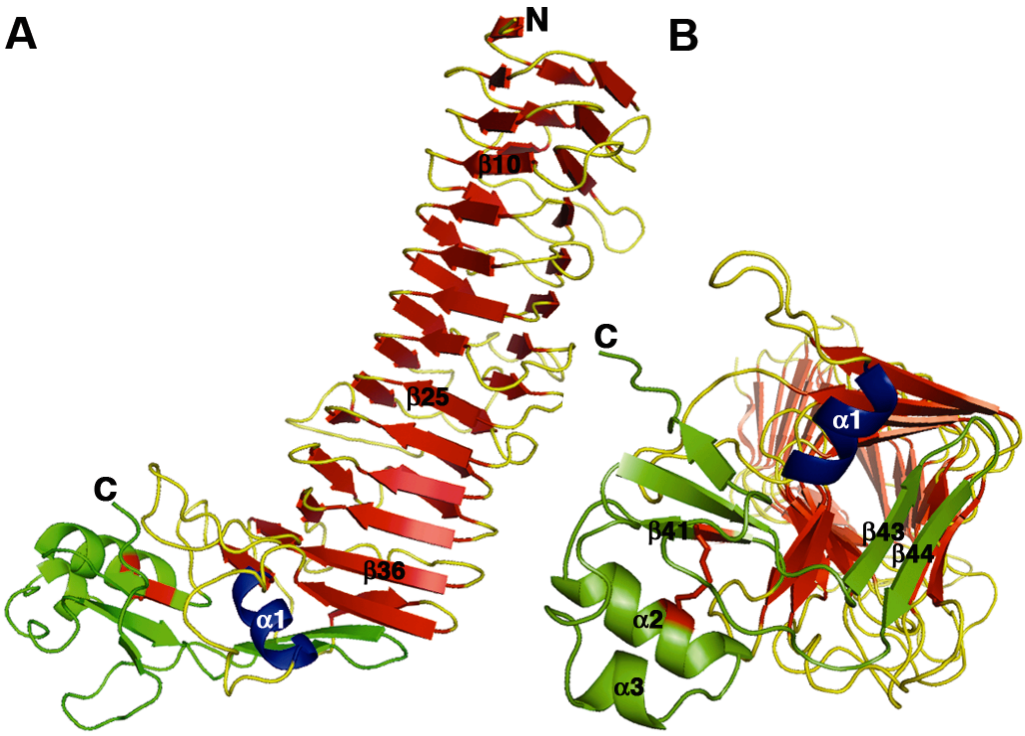


Figure 5. VacA p55 structure. (A) The VacA p55 fragment adopts a β-helix structure that is composed of three parallel β-sheets (red) connected by loops of varying length and structure (yellow). The α-helix in blue (α1) is contained within one of these loops but is highlighted in blue to show how it caps the end of the β-helix. The C-terminal domain (green) has a mixture of α/β secondary structure elements and contains a disulfide bond (red) not previously observed in an autotransporter passenger domain structure. A few of the secondary structural elements are labeled to serve as landmarks and correlate to sequences depicted in Figure 12. (B) The C-terminus of the β-helix is capped by a β-hairpin from the C-terminal domain (green) and the α1 α-helix (blue) located in one of the long β-helix loops. This view represents a rotation of the molecule in a by ~90 ° into the plane of the page.

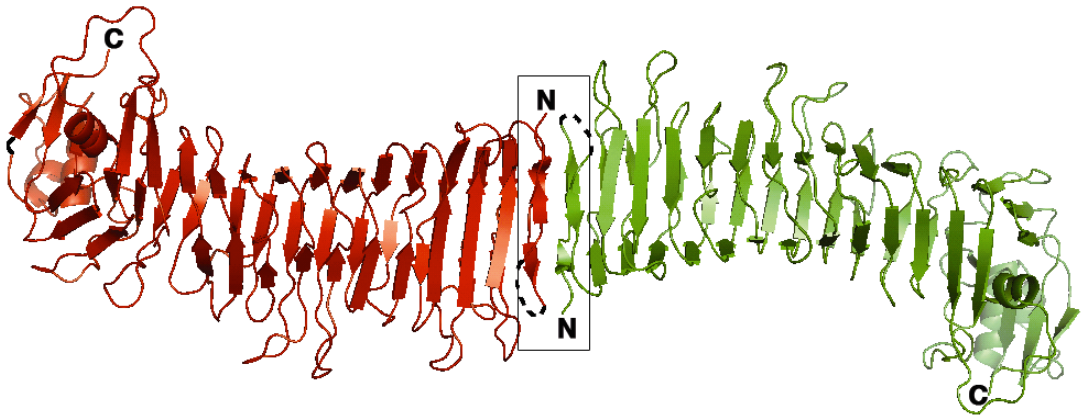


Figure 6. The VacA p55 dimer forms an elongated β-helical structure. Gel filtration and electron microscopy (118) indicate that VacA p55 exists as a dimer in solution. While the p55 protein crystallized with one molecule per asymmetric unit, a two-fold symmetry axis was present that revealed an extended β-helix structure formed from a head-to-head assembly of two molecules (red and green). The fact that p88 does not form dimers suggests that this interface (boxed) is occluded by p33. We propose that p33 adopts a β-helix structure that extends from the N-terminal end of p55.

Comparison of the VacA p55 structure with structures of other autotransporter passenger domains

Crystal structures have been determined for two other type Va autotransporter passenger domains: pertactin (an adhesin from *B. pertussis*) and Hbp (a hemoglobin protease from *E. coli*). Pertactin (136), Hbp (137), and VacA p55 do not share sequence similarities but all contain a β-helix fold (Figure 7). The β-helix fold is composed of multiple ~25 amino acid quasi-repeats, each of which forms a single rung of the helix (Figure 8). Each rung contains three β-strands and three turn regions named β1, T1, β2, T2, β3 and T3 in their order of occurrence. β1 and β2 are antiparallel β-strands and β3 is approximately perpendicular to β2. Conservation of the β-helix fold among autotransporter passenger domains suggests that this structural feature is required for efficient secretion across the outer membrane and folding (138). Specifically, folding studies with pertactin indicate that the β-helix can form slowly but reversibly in solution, without aggregation, and with the formation of a stable core in the C-terminal half of the β-helix (138). The C-terminal region of BrkA, an autotransporter whose structure is predicted to resemble that of pertactin, was shown to mediate BrkA folding, even when added *in trans* (139). A C-terminal β-helix cap may be important as a nucleus and/or chaperone for secretion and folding (139). These observations are consistent with current models of AT transport across the outer membrane (OM) in that they suggest the passenger domain can remain unfolded without aggregation in the periplasm, that the protein can fold in a C- to N-terminal fashion once threaded through the OM, and that the C-terminal β-helix cap is an important nucleus for proper folding. The end of the VacA p55 β-helix is capped by a β-hairpin, similar to that observed in the structure of pertactin (136). There is also a short α -helix in this region of VacA that appears to be unique among β -helix structures (Figure 5B).

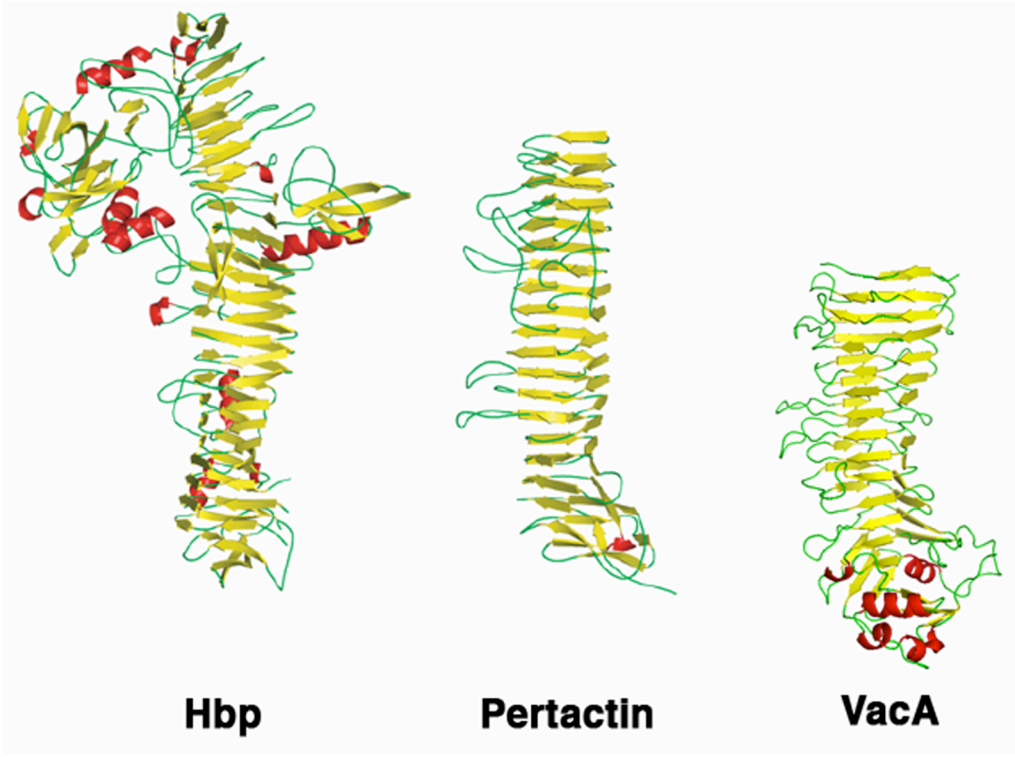


Figure 7. Passenger domain structures. Ribbon diagrams of the passenger domains from three autotransporter structures: Hbp (a hemoglobin protease from *E. coli*), pertactin (an adhesin from *Bordetella pertussis*) and VacA. β-strands are yellow, loops are green, and helices are red. Adapted from Barnard *et al.* (140).

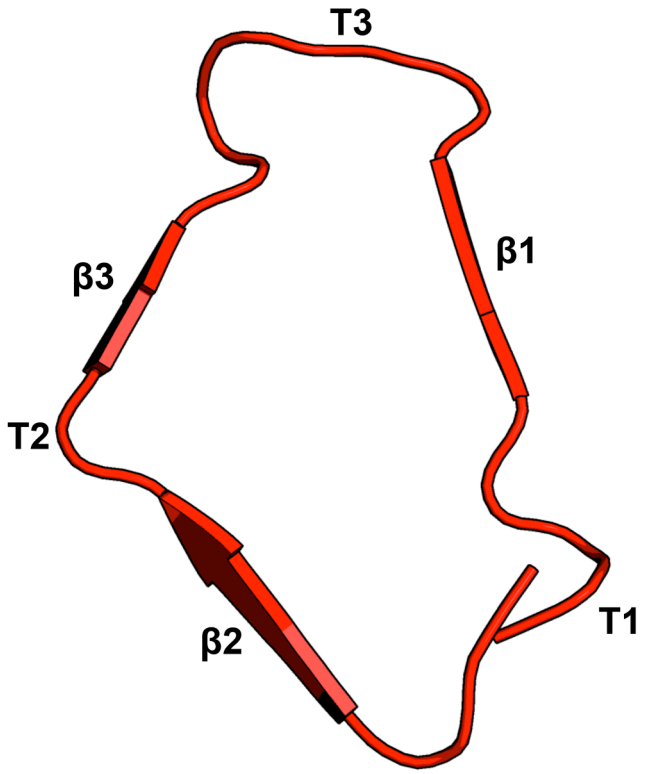


Figure 8. Nomenclature of the rungs from parallel β-helix proteins. A rung of VacA p55 domain is shown and viewed down from the parallel β-helix axis. The chain enters the coil from the T1 loop region. The three parallel β-strands and the turns between them from a single rung are shown and named β1, T1, β2, T2, β3 and T3.

A notable difference between VacA p55 and the two other autotransporter passenger domain structures is the presence of a disulfide in the p55 C-terminal sub-domain (Figure 5B). The low cysteine content observed in autotransporter sequences is consistent with a model in which passenger domains translocate across the outer membrane in an unfolded conformation. Nevertheless, a number of autotransporter sequences have a single, closely-spaced pair of cysteines near the C-terminus of their passenger domains (66). In VacA, these cysteines are either 11 or 13 amino acids apart. Mutation of either of these cysteines to serine results in a decrease in toxin secretion but has no effect on the vacuolating activity of the toxin (66) and (M. McClain, unpublished results). Here, we show experimentally that these cysteines form a disulfide and that they are positioned to buttress both the β-helix cap and a conserved C-terminal pocket (discussed below). Finally, another feature of p55 that differs from other autotransporter passenger domains is the presence of multiple kinks that disrupt what would otherwise be continuous β-sheets. We have divided the VacA p55 β-helix into five sub-domains to reflect these disruptions (Figure 9) and note that the divisions correlate with predicted sites of homologous recombination (as discussed further below).

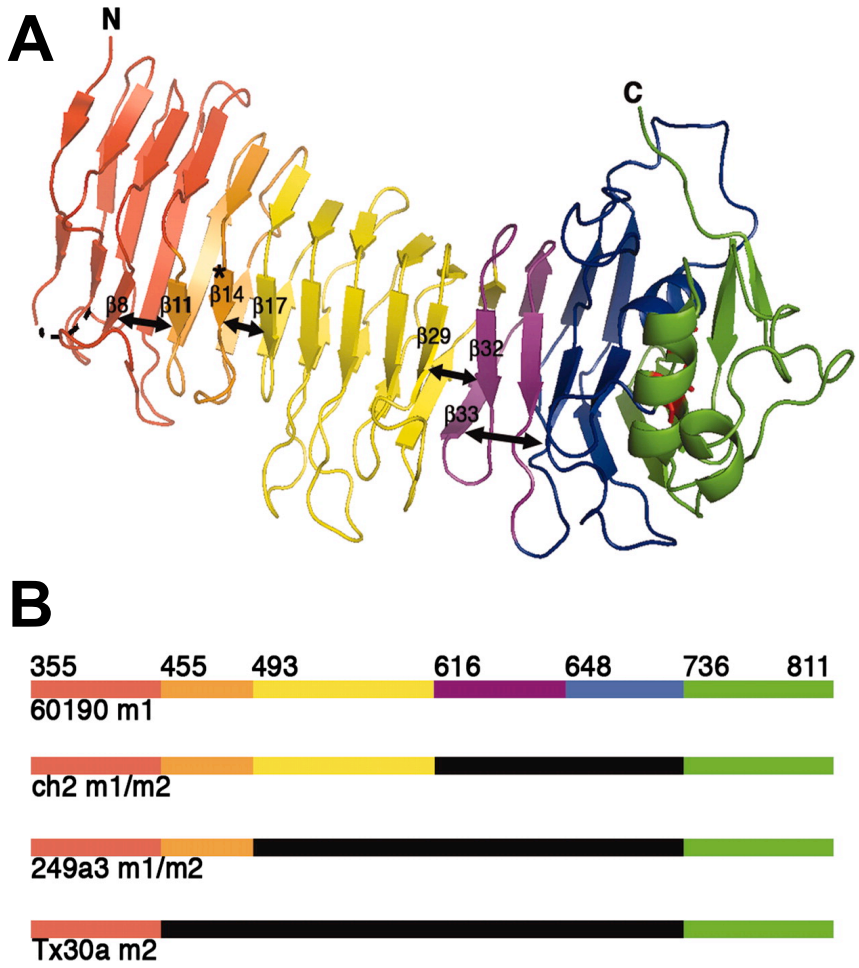


Figure 9. The VacA p55 structure reveals marked disruptions in β-sheet contacts. (A) Breaks in the β-sheet contacts (depicted with double-headed arrows) result in sub-domains (SD) within the β-helix: SD-1 (red, 355-454), SD-2 (orange, 455-493), SD-3 (yellow, 494-615), SD-4 (purple, 616-647), SD-5 (blue, 648-735). The C-terminal domain and disulfide bond are colored green and red, respectively, and the site of the m2 23 amino acid insertion is indicated by an asterisk. (B) This schematic shows prototype m1 and m2 VacA proteins (strains 60190 and Tx30a, respectively) and two naturally occurring m1m2 chimeras. The breaks in β-sheet contacts (A) correlate with sites of homologous recombination (see also Figure 13). In this schematic the mid-region that distinguishes type m1 and type m2 sequences is colored orange/yellow/purple/blue (m1) or black (m2).

Sequence variation among VacA proteins

In an effort to understand how genetic differences between m1 and m2 *vacA* alleles relate to structural and functional differences in the proteins they encode, we have examined VacA sequence polymorphisms in the context of the p55 structure. We identified sequences for 62 m1 and 27 m2 VacA proteins in GenBank that were complete in the p55 region. In addition, we identified 3 m1/m2 chimeric VacA proteins for which sites of recombination between m1 and m2 sequences were easily recognizable. We aligned these sequences in various combinations using the program ClustalW (141) and mapped the sequence similarity scores to the three-dimensional m1 p55 structure using ESPript (142). The majority of residues pointing into the interior of the β-helix are either strictly or highly conserved, consistent with the idea that mutation of buried residues would result in a detrimental loss of structure and, therefore, function (Figure 10). The surface-exposed residues are also fairly conserved when exclusively m1 or exclusively m2 sequences are analyzed (Figure 11). However, when m1 and m2 sequences are analyzed together, the surface-exposed residues are highly variable (Figure 12). There are only two regions of the surface with notable sequence conservation. One is located at the N-terminus (Figure 12A) and will be discussed in the next section with respect to protein oligomerization. The second conserved surface is located at the C-terminus of the β-helix in a cavity formed by two loops (residues 668-678 and 730-734) and the disulfide-linked C-terminal domain (Figure 12B). The strict sequence conservation, the fact that many binding sites are located in clefts or cavities, and the fact that this is the only cavity observed in the VacA p55 structure, suggest that this area may represent a receptor binding site that is shared by m1 and m2 forms of the toxin.

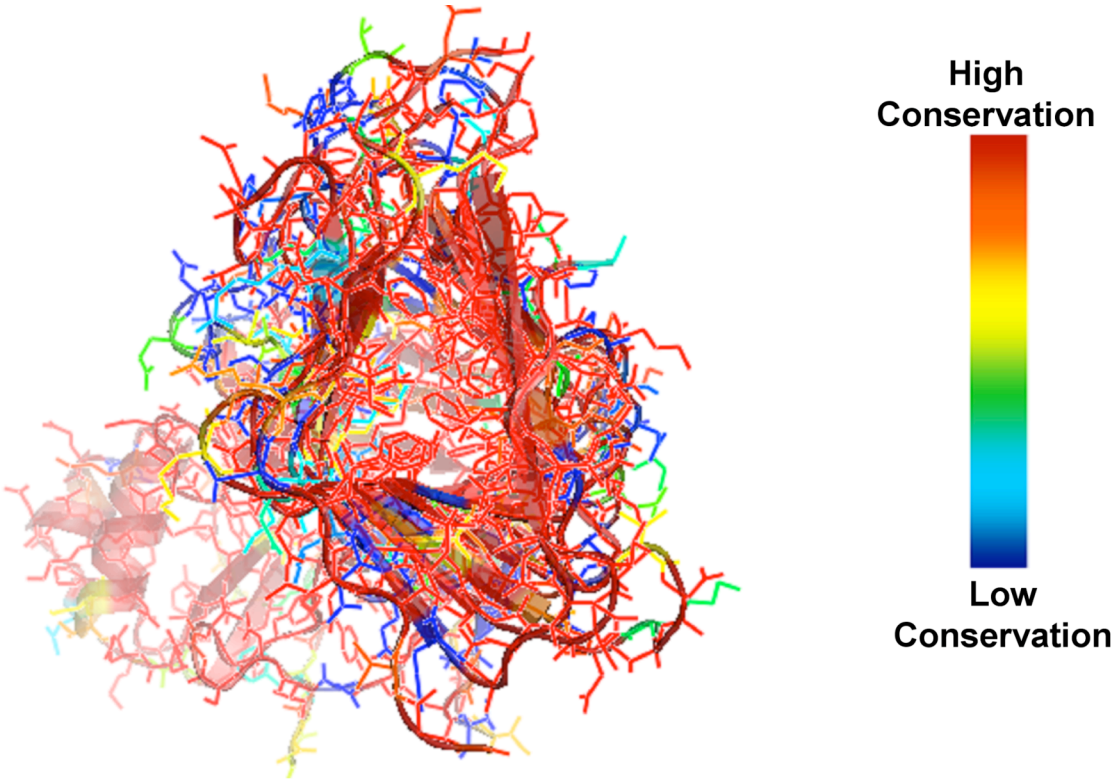


Figure 10. Alignment of m1 and m2 sequences mapped to the structure of VacA p55 from strain 60190 (an m1 sequence). 62 m1and 27 m2 sequences were aligned to that of VacA from m1 strain 60190 and scored with a Risler matrix according to the extent of sequence variation. Scores were displayed on the p55 structure with a color ramp (red, orange, yellow, green, light blue, dark blue) in which strictly conserved residues are colored red, and the most variable residues are colored dark blue. Residue side chains are shown as sticks. The color-ramp of the alignment scores indicates a high level of sequence conservation of residues pointed in the interior of the β-helix with low sequence conservation for surface-exposed residues. This view represents a rotation of the molecule into the plane of the page, from N- to C-terminus, looking down the β-helix.

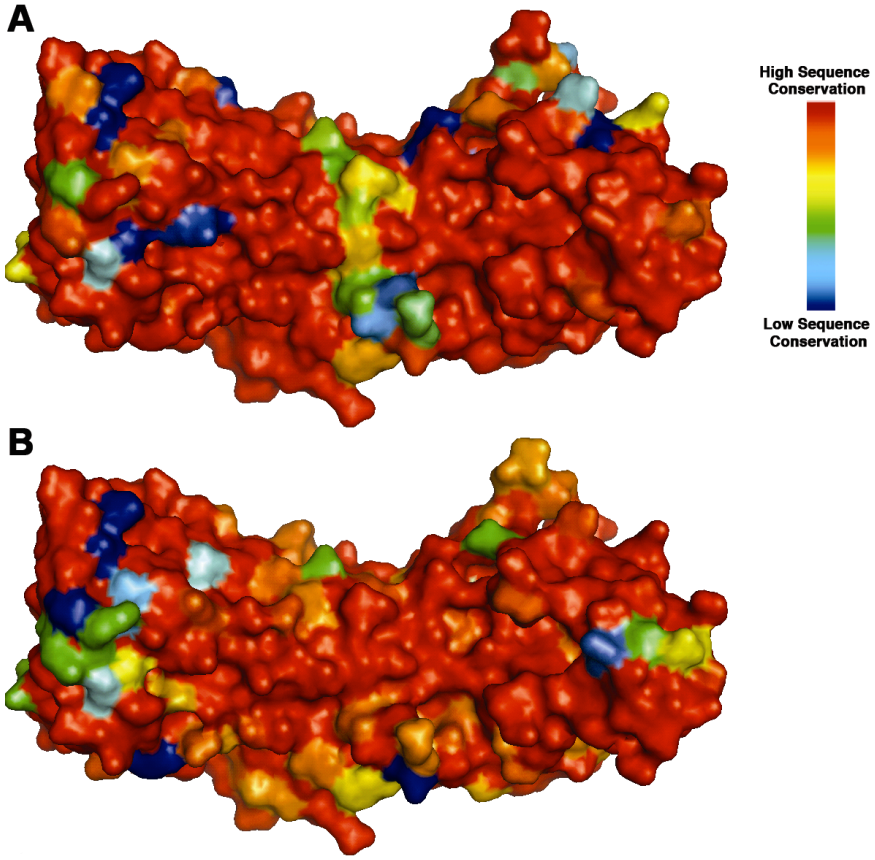


Figure 11. Alignment of exclusively m1 or exclusively m2 sequences mapped to the structure of VacA p55 from strain 60190 (an m1 sequence). (A) 49 m1 sequences were aligned to that of VacA from m1 strain 60190 and scored with a Risler matrix as described in Figure 10. The color-ramp of the alignment scores indicates a high level of sequence conservation among m1 strains. The central stripe of sequence diversity corresponds to β26, β29, and loop residues 597-600. (B) 27 m2 sequences were aligned to each other for calculating an alignment score and to the sequence of m1 VacA from strain 60190 for the purpose of displaying those scores on the three-dimensional structure of m1 VacA p55. While the m1 structure is not an entirely accurate representation of the m2 surface structure due to the presence of the 23 amino acid insertion and other sequence variations in the m2 mid-region, the analysis suggests that the surfaces of VacA proteins from m2 strains will be very similar.

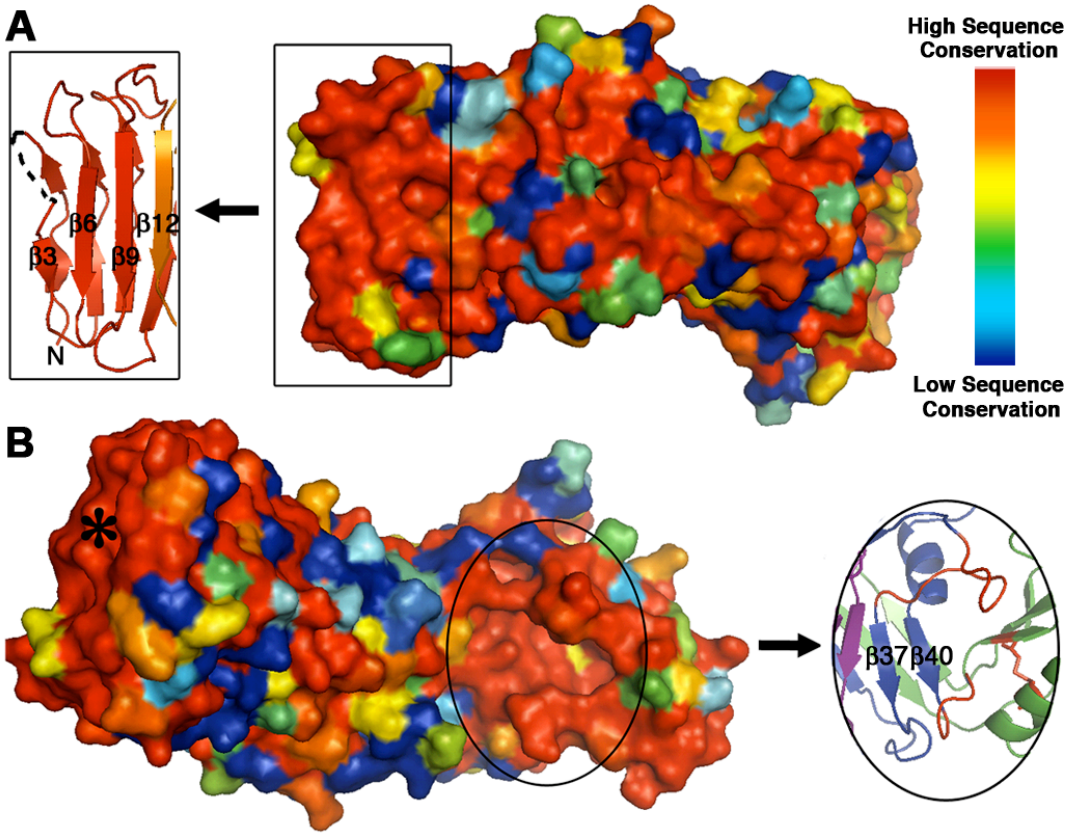


Figure 12. The VacA p55 structure has two patches of conserved residues. (A) 62 m1, 27 m2, and 3 m1/m2 chimeric VacA sequences were aligned and scored as described in Figure 10. One conserved region is at the N-terminus of the protein (boxed) and correlates to the β3, β6, and β9 strands of the β-helix. We propose that this surface is important for VacA oligomerization. (B) Rotation of the molecule in view (A) by ~180˚ around the long axis of the protein reveals the second conserved region of the VacA p55 structure. This surface is located in a pocket at the C-terminus of the β -helix (circled) and may represent a common receptor-binding site for all m1 and m2 VacA proteins. The pocket is formed by two long β -helix loops (red) and the disulfide bond (also in red) of the C-terminal domain. While the asterisked region is also highly conserved, this correlates to the N-terminal end of the β-helix that we predict to be buried when p33 is present.

Because multiple surface-exposed regions are highly divergent, it is difficult to identify a single region in the p55 structure that can explain the different binding properties of m1 and m2 VacA proteins (Figure 12). The biggest difference between m1 and m2 sequences is the presence of a 23 amino acid insertion in m2 sequences (Figure 9A, asterisk; and Figure 13). The insertion is notable in that its sequence is an imperfect repeat of the sequence located immediately upstream, and the length approaches the average length of a single β-helix rung (~25 amino acids) (Figure 13). The 23 amino acid m2 insertion could therefore result in an additional β-helix rung, similar in structure to the rung that precedes it. Based on an analysis of engineered m1/m2 chimeras, a region important for m1-specific binding to HeLa cells has been mapped to residues 460-496 in the m1 sequence (126). The m2 23 amino acid insertion is located within this region, but an engineered chimera in which these 23 m2 amino acids were inserted into an m1 sequence did not result in a loss of HeLa cell toxicity. Excluding the m2 insertion, there are only ten amino acids in the 460-496 m1 sequence that consistently differ between the m1 and m2 forms (Figure 13). The position of these residues in the context of the 23 amino acid m2 insertion is impossible to predict, and therefore, an understanding of structural features that give rise to VacA cell-type specificity may need to await analysis of an m2 p55 structure.

Analysis of the VacA sequence alignments in the context of the m1 p55 structure reveals that disruptions of β-sheet contacts within the β-helix correlate with sites of homologous recombination events. For example, VacA residue D455 (strain 60190) represents the beginning of the mid-region that defines m1 and m2 strains (Figure 9B and Figure 13) and follows a kink in the structure that disrupts β-sheet contacts between β8 and β11 (Figure 9A and Figure 13). This structural disruption and the associated sequence transition are illustrated with a change from red to orange (Figure 9 and Figure 13). Other examples of a correlation between VacA sub-domains and recombination

events are evident based on analysis of rare naturally occurring m1/m2 chimeras, which have m1 character in the N-terminal portion of the VacA mid-region but m2 character in the C-terminal portion. The transition point from m1 to m2 in the chimeric VacA from strain 249a3 is at D493, while the transition point for the ch2 and v225 strains of VacA is at N616 (Figure 9B and Figure 13). As illustrated in Figure 9A, these sequence transitions correlate with disruptions in β-sheet contacts between β14 and β17 and between β29 and β32, respectively. The final sub-domain begins at residue G648. While we did not identify a sequence with evidence of recombination at this site, the sub-domain (blue) has longer loops than other sub-domains and does not make β-sheet contacts with β33 from the previous sub-domain (Figure 9A).

In summary, the polymorphisms that differentiate m1 and m2 forms correspond almost entirely to surface-exposed residues. Diversification in these surface-exposed regions of VacA may have been driven by immune selective pressures. Recombination occurs commonly in *H. pylori* (17, 125), but m1/m2 chimeras are only rarely identified (124). This suggests that *H. pylori* strains with intact m1 or intact m2 VacA sequences have favorable functional properties, and thus, have a selective advantage compared to strains containing chimeric m1/m2 sequences (68, 126, 127). The rare naturally occurring m1/m2 chimeras shown in Figure 9B and Figure 13 probably arose as the result of recombination events that maintained favorable structural and functional properties, due to recombination at sites corresponding to breaks in the β-helix.

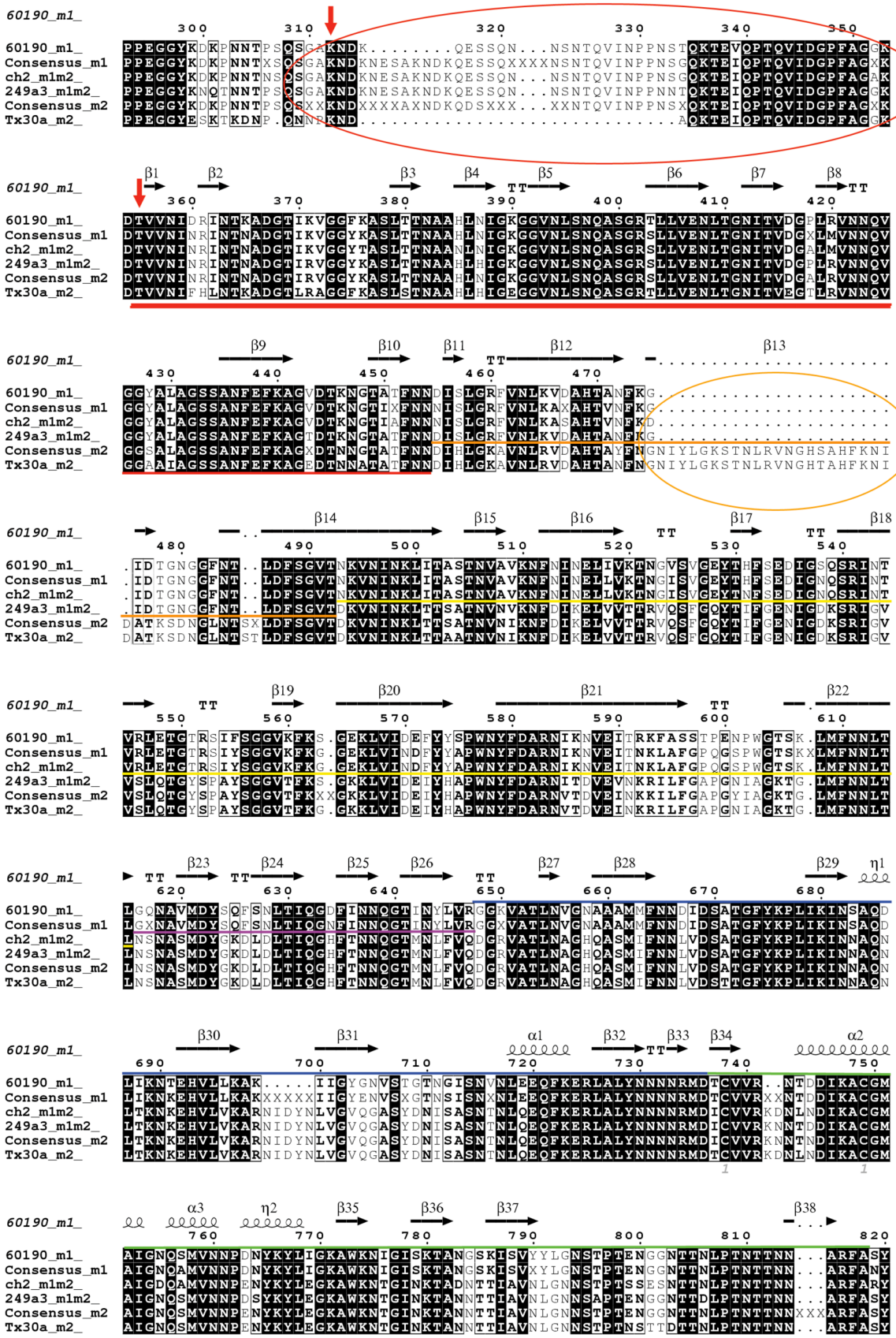

60190_m1_
60190_m1_
Consensus_m1
ch2_m1m2_
249a3_m1m2_
Consensus_m2
Tx30a_m2_

Figure 13. Alignment of m1, m2 and m1/m2 chimeric VacA sequences reveal sites of homologous recombination. VacA sequences correspond to GenBank accession numbers Q48245 (60190), Q9KJA6 (ch2), Q6DLS8 (249a3) and Q48253 (Tx30a). The numbering corresponds to the VacA sequences from strain 60190. Consensus sequences, based on the analysis of 62 m1 and 27 m2 sequences, are also shown. The secondary structural elements are identified as β (β-strand), TT (tight turn or kink), η (3_{10}-helix), or α (α-helix). Sequences were aligned in CLUSTALW (141) and displayed in ESPript (142). Strictly conserved residues are shown in white letters with a black background. Residues that are similar are boxed in bold letters with white backgrounds. The linkage between the p33 and p55 domains of VacA is sensitive to protease and thought to adopt a flexible loop structure since p33 and p55 remain associated after cleavage. The first residue of p55, defined as K312 (red arrow), may also reside in a flexible loop as the N-terminal portion of p55 (circled in red) was removed by proteolysis during crystallization. The first residue observed in the crystal structure was T355 (second red arrow). Residue D455 (where orange color begins) represents the beginning of the mid-region that defines m1 and m2 strains. The most notable feature of m2 strains is a 23 amino acid insertion between residues G475 and I476 of strain 60190 (orange circle). This insertion is an imperfect repeat of the preceding sequence (starting at residue D455) and is absent in both of the m1/m2 chimeric strains. The first residue where the chimeric VacA from strain 249a3 begins to adopt m2 character aligns to N493 of strain 60190. The sequence that follows is colored yellow to highlight this transition. The first residue where the chimeric VacA from strains ch2 and v225 begins to adopt m2 character aligns to G616 of VacA strain 60190. (Only the ch2 sequence is shown for clarity.) The sequence that follows is colored purple to highlight this transition. The structure is colored in blue from 648-734 to indicate the fifth sub-domain of the β-helix. The C-terminal domain is colored green.

Assembly of VacA into oligomeric structures

In order to understand the p55 structure in the context of VacA oligomers, we have docked p55 into the 19 Å cryo-EM map of the wild-type dodecamer (107) using the program COLORES (Figure 14A) (143). The elongated shape of the β-helix and the curve of the C-terminal 'foot' allow for unambiguous placement of 12 p55 subunits into the 'petal-like' features of the map. The docking suggests that contacts between the p55 components of the two hexameric layers are mediated by three loops and that the N-terminal end of the p55 β-helix extends into the central density of the map (Figure 14B). The conserved pocket that we propose as a potential common receptor binding site for m1 and m2 forms of VacA is located on the sides of the 'petals' and would be fully accessible to cell-surface receptors when VacA is assembled in either a single-layered or double-layered oligomeric structure (Figure 14B).

We hypothesize that a large portion of p33 will adopt and extend the β-helix fold observed in p55. This idea is supported by two observations. First, an extended β-helix structure is consistent with the shape of a p88 monomer obtained by EM (Figure 14C, inset) (107). Second, the program BetaWrapPro identifies the stretch of amino acids between residues 120 and 249 as a five-coil β-helix and predicts additional β-helix strands in the region between 252 and 288 (144). BetaWrapPro uses profile wrapping for prediction and comparative modeling of β -helices and has been shown to identify the β-helix motif with high sensitivity and selectivity (144).

Based on our prediction that a large portion of p33 will extend the β-helix fold observed in p55, we suggest a model in which VacA oligomerization is mediated by contacts between p33 and the N-terminus of p55 from a neighboring subunit (Figure 14C). Specifically, one would predict the p55 oligomerization surface to contain the β-strands β3 and β6. The surface-exposed residues of β3 and β6 are accessible within the central density of the EM map and are strictly conserved among the 92 m1 and m2

sequences we surveyed, suggesting selective pressure to preserve this surface (Figure 14B). An important functional role of this region is also supported by biochemical data indicating that VacA residues 1-422 can induce vacuolation in cultured cells when expressed intracellularly but p33, p55, and VacA residues 1-394 cannot (115, 122). Finally, this model is supported by a yeast two-hybrid experiment in which p33 and residues 313-478 of p55 were shown to interact (114). The oligomerization surface may also contain parts of the flexible loop (312-354) that was not visible in this structure. Much of this loop is strictly conserved among the m1 and m2 VacA sequences (Figure 13).

p33 (residues 1-311) contains a putative α-helical pore-forming domain at its N-terminus (116, 145). Comparison of the VacA wild-type and Δ6-27 dodecamer structures by cryo EM suggests that the N-terminal domain of p33 is located in the central density of the structure (107). Secondary structure prediction analyses suggest that p33 has α-helical structural elements between residues 1 and 71 (144). Short β -strands are then predicted to begin at residue 87 and extend through a region (residues 120-288) that is predicted to have a β-helical structure, based on BetaWrapPro. We therefore anticipate that a single subunit of VacA will adopt a roughly symmetric shape within the context of the oligomer and that residues in the N-terminal region of p33 (likely C-terminal to the pore-forming region) would be positioned to mediate oligomerization with the N-terminal region of a neighboring p55 subunit (Figure 14C).

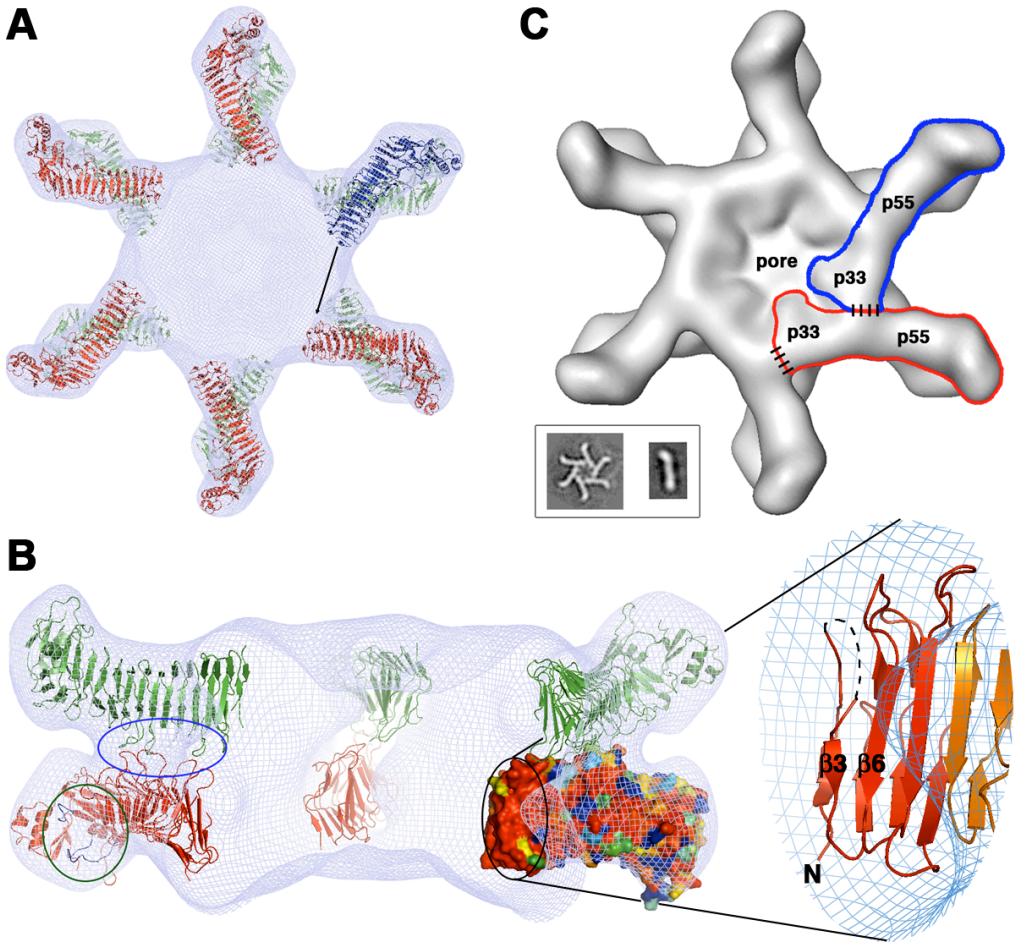


Figure 14. Docking the p55 crystal structure into a 19Å cryo-EM map of the VacA dodecamer results in a model for oligomerization. (A) 12 p55 subunits are shown docked into a 19Å cryo-EM map of a VacA dodecamer. An arrow is shown to indicate the space that the blue molecule will occupy if p33 extends the β-helix structure of p55. (B) The view in (A) is rotated by 90˚ such that the blue molecule moves toward the reader and is now located on the bottom of the dodecamer. The blue molecule is not visible in this side view, however, because the view has been sliced so that only the 'back' of the structure is visible. On the left, a blue circle highlights three loops that mediate p55-p55 contacts between the two layers. Circled in green are the two loops that line the conserved pocket that we propose as a common receptor-binding site. This pocket is located on the side of the molecule and would therefore be accessible in both single layered and bilayered forms of the toxin. The black circle contains the other conserved surface in p55 (also shown in Figure 12A). This surface protrudes into the central ring of the VacA oligomer and would be accessible to the blue molecule in (A) for contacts that could mediate oligomerization within a hexameric or heptameric plane. We have zoomed in on the secondary structures that contribute to this surface. (C) We propose an oligomerization model in which p33 interacts with the N-terminal portion of p55 from the neighboring subunit. Regions of contact between p33 and p55 are depicted with dashed lines. The inset shows EM images of a VacA hexamer and a VacA monomer. The shape of a VacA hexamer (inset) is similar to the shape of a single layer within the dodecamer. The rod-like shape of the p88 VacA monomer (inset) supports a model in which the β-helix observed in p55 will extend into p33.

A protective *H. pylori* vaccine would potentially reduce the incidence of *H. pylori* infection and the serious complications of peptic ulcer disease and gastric adenocarcinoma (146), but the high level of diversity among *H. pylori* strains and the high frequency of genetic recombination among strains are likely to present significant challenges to vaccine development. VacA is a candidate vaccine antigen since immunization with VacA confers protective immunity in a mouse model of *H. pylori* infection (146, 147). We show here that two VacA surfaces are strictly conserved among all surveyed m1 and m2 sequences and propose that these surfaces are under selective pressure to be preserved in order to mediate receptor binding and oligomerization functions. Efforts to selectively target these regions could result in progress toward the development of a vaccine that confers protection against multiple strains of *H. pylori*.

CHAPTER III

RECONSTITUTION OF *HELICOBACTER PYLORI* VACA TOXIN FROM PURIFIED COMPONENTS

Introduction

The 88 kDa VacA monomers secreted by *H. pylori* can assemble into large water-soluble oligomeric complexes. These flower-shaped structures can be either single-layered (containing 6-9 subunits) or bilayered (containing 12-14 subunits) (60, 106, 107, 109). Similar oligomeric structures have been visualized on the surface of VacA-treated cells or lipid bilayers (60-62). Amino acid sequences within both the p33 domain (residues 49-57) and p55 domain (residues 346-347) are required for assembly of VacA into these oligomeric structures, and mutant proteins lacking these sequences fail to cause cell vacuolation (120, 121). Several VacA mutant proteins have dominant-negative inhibitory effects on the ability of wild-type VacA to cause cellular alterations (105, 120, 123), which further supports the hypothesis that oligomeric structures are required for VacA effects on host cells. Water-soluble VacA oligomeric complexes lack cytotoxic activity unless they are first dissociated into monomeric components by exposure to low pH or high pH conditions, which suggests that VacA monomeric components reassemble into membrane channels when in contact with host cells (61, 72, 78, 79). Although the structure of water-soluble VacA oligomeric complexes has been investigated in detail, the conditions that promote oligomerization of VacA are not well understood.

While an X-ray crystal structure for p55 exists (Chapter II) and p88 oligomers have been visualized by cryo-electron microscopy (60, 107), a detailed analysis of p33 has been hindered by an inability to purify this domain in an active form. In this study, we expressed and purified a recombinant form of p33 under denaturing conditions and optimized conditions for the refolding of soluble protein. We show that refolded p33 can be added to purified p55 *in trans* to cause vacuolation of HeLa cells and inhibition of IL-2 production by Jurkat cells, effects identical to those produced by the p88 toxin secreted from *H. pylori*. The p33 protein markedly enhances the cell-binding properties of p55. Size exclusion chromatography experiments suggest that p33 and p55 assemble into a complex consistent with the size of a p88 monomer. Electron microscopy reveals small rod-shaped structures that can convert to oligomeric flower-shaped structures in the presence of detergent. We propose that the ability to reconstitute a functional VacA toxin from two purified fragments *in trans* is facilitated by the elongated β-helical structure of VacA.

Methods

Purification of p88 VacAfrom H. pylori broth culture supernatant. H. pylori strain 60190 was grown in broth culture and VacA was purified from the culture supernatant as described previously (60, 107).

Plasmids for expression of p33 and p55 VacA fragments. Plasmids encoding the p33 and p55 domains of VacA from *H. pylori* strain 60190 (a type m1 form of VacA; GenBank accession number Q48245), as well as a p33Δ6-27 mutant protein, have been described previously (114, 123, 148). The p33 proteins contain a C-terminal hexahistidine tag and the p55 protein contains an N-terminal hexahistidine tag.

*Expression and purification of recombinant VacA proteins***.** VacA p55 was purified as described previously (148). VacA p33 was expressed in *E. coli* BL21 (DE3) by culturing in TB (Fisher) supplemented with 25 µg of Kanamycin/ml (TB-KAN) at 37 °C overnight with shaking. Cultures were diluted 1:100 in TB-KAN and grown at 37 °C until they reached an OD_{600} of 0.6. Cultures were induced with a final IPTG concentration of 0.5 mM and incubated at 37 °C for 2 hours (114).

VacA p33 proteins were purified from inclusion bodies. Briefly, IPTG-induced cultures were pelleted, washed in 0.9% NaCl, and resuspended (10 ml per liter of culture) in sonication buffer [10 mM Tris pH 7.5, 100 mM NaCl, 1 mM EDTA, a protease inhibitor tablet, and lysozyme 20,000 U/ml (Ready-lyse, Epicentre)]. The cells were incubated at room temperature for 15 minutes with shaking, and sonicated with six 20 watt bursts (45 seconds per burst with 15 s cooling periods). Lysed bacterial cells were centrifuged to pellet the inclusion bodies. The insoluble inclusion body pellet was resuspended in buffer containing 100 mM NaH_2PO_4, 10 mM Tris, and 8 M urea (pH 8.0) at 5 ml per gram wet weight, and incubated for 1 hour at room temperature. The samples were centrifuged, and the resulting supernatant was added to Ni-NTA beads (Novagen) at a ratio of 4 ml of supernatant per 1 ml of beads. The protein-bead mixture was incubated for 1 hour at room temperature before loading it into a column and allowing it to flow by gravity. The column was washed with 10 column volumes of 100 mM NaH_2PO_4, 10 mM Tris, 10 mM imidazole, and 8 M urea (pH 6.3), followed by 100 mM NaH_2PO_4, 10 mM Tris, and 8 M urea (pH 5.9). The p33 protein was eluted from the column with 100 mM NaH_2PO_4, 10 mM Tris, and 8 M urea (pH 4.5). Successful expression and purification of p33 was confirmed by mass spectrometry.

*Refolding of VacA p33***.** The denatured VacA p33 protein was refolded by dialyzing the protein against a buffer containing 55 mM Tris, 21 mM NaCl, 0.88 mM KCl, 1.1 M guanidine, and 880 mM arginine (pH 8.2) for 24 hours. The protein then was dialyzed in

two other buffers, each for 24 h. The first reduced the guanidine to 800 mM and the arginine to 500 mM, and the second reduced the arginine to 250 mM and maintained an 800 mM guanidine concentration (149). Further reductions in the arginine or guanidine concentrations resulted in precipitation of p33 VacA. The dialyzed protein was then centrifuged and filtered (0.2 micron) to remove any insoluble protein.

*Cell culture assays***.** HeLa cells were grown in minimal essential medium (modified Eagle's medium containing Earle's salts) supplemented with 10% fetal bovine serum (FBS) in a 5% CO_2 atmosphere at 37ºC. Jurkat lymphocytes (clone E6-1) (ATCC TIB-152) were grown in RPMI 1640 medium containing 2 mM L-glutamine, 1.5 g/liter sodium bicarbonate, 4.5 g/l glucose, 10 mM HEPES and 1.0 mM sodium pyruvate supplemented with 10% FBS in a 5% CO_2 atmosphere at 37 °C.

For vacuolating assays, HeLa cells were seeded at 1.2 x 10^4 cells/well into 96-well plates 24 hours prior to the addition of VacA proteins. The recombinant p33 and p55 proteins (each at 1 mg/ml) were pre-mixed at a 1:1 mass ratio. Preparations of p33, p55 or the p33/p55 mixture were then added to the tissue culture media overlying HeLa cells (supplemented with 10 mM ammonium chloride) at concentrations of 10 µg/ml (or 5 µg/ml of each protein in the case of the p33/p55 mixture) for 9 hours at 37 °C. After incubation, cell vacuolation was examined by inverted light microscopy and quantified by neutral red uptake assay (150). For dominant negative assays, refolded p33 Δ6-27 protein was mixed with the refolded p33/purified p55 mixture (1 mg/ml each) in a 1:1:1 mass ratio and added to the tissue culture media overlying the cells at a final protein concentration of 10µg/ml for 9 hours at 37 °C, and vacuolation was quantified (123).

To analyze the capacity of VacA to inhibit IL-2 secretion by Jurkat T cells, Jurkat cells were plated at 1 x 10^5 cells/well, and the recombinant p33 and p55 were added to cells either individually or as a p33/p55 mixture (1:1 mass ratio) at concentrations of 6 µg/ml (or 3µg/ml in the case of the p33/p55 mixture), which corresponds to about a 1.7:1

molar ratio, for 30 minutes at 37 °C. The use of excess p33 on a molar basis compensated for the possibility that refolding of denatured p33 might be less than 100% efficient. After incubation, 0.05 μg/ml of phorbol 12-myristate 13-acetate (PMA) and 0.5 μg/ml of ionomycin were added for 24 hours at 37°C. The cells were then centrifuged at 2,000 rpm for 7 minutes and the supernatants were tested for IL-2 by ELISA, according to the manufacturer's protocol (R&D Systems Human IL-2 Immunoassay) (151).

Interactions of p33 and p55 with HeLa cells. Purified p55 was labeled with Alexa 488 (Molecular Probes) according to the manufacturer's instructions. HeLa cells were incubated with Alexa 488-labeled p55 alone (10 ug/ml) or a mixture of labeled p55 plus purified refolded p33 (5 ug/ml of each) at 37 °C degrees. Alternatively, cells were incubated with purified Alexa 488-labeled p55 plus purified refolded c-Myc-tagged p33 protein (113). Cells were fixed with 4% formaldehyde. To detect the c-Myc-tagged p33 protein, cells were permeabilized with methanol, and the c-Myc tag was detected by indirect immunofluorescence using anti-c-Myc antibody and a Cy3-conjugated secondary antibody. Cells were viewed with an LSM 510 confocal laser scanning inverted microscope.

*Size exclusion chromatography***.** Gel filtration was performed using eitherSuperdex 200 10/300 GL high-resolution resin or Superdex 200 10/300 prep grade resin, equilibrated in 55 mM Tris (pH 8.0), 21 mM NaCl, 0.88 mM KCl, 800 mM guanidine, and arginine (either 800 mM or 250 mM). Protein samples (500 μl) were first injected onto the gel filtration column individually at a final concentration of 0.75 mg/ml for the p33 protein and 0.4 mg/ml for the p55 protein. To analyze p33/p55 mixtures, the appropriate sizing column fractions corresponding to either p33 or p55 were each concentrated to 1 mg/ml. VacA p33 was added to p55 in a 2:1 v/v ratio, the mixture incubated for 45 minutes at 4 °C, and the p33/p55 mixture was then applied to a gel filtration column. Retention

volumes of bovine thyroglobulin, alcohol dehydrogenase, bovine serum albumin, and carbonic anhydrase were used as standards to calculate the molecular masses of the purified VacA proteins.

Electron microscopy. To visualize the morphology of p33/p55 mixtures, appropriate gel filtration fractions containing these proteins were analyzed by electron microscopy (EM) using conventional negative staining as described (152). Protein solutions were diluted to appropriate final concentrations (25 to 100 µg/ml) and 2.5 µl aliquots were spotted onto glow-discharged copper-mesh grids (EMS) for approximately 1 min. In some experiments, p33/p55 mixtures were mixed 9:1 (v/v) with Brucella broth (153), bovine total heart extract (Avanti) solubilized in chloroform (25 mg/ml), chloroform, or n-dodecyl beta-D-maltoside (DDM, Anatrace) prior to EM analysis. The final concentration of DDM was 0.34 mM, which corresponds to twice the critical micelle concentration. The grids were washed in 5 drops of water followed by one drop of 0.7% uranyl formate. Grids were then incubated on one drop of 0.7% uranyl formate for 1 min, blotted against filter paper and allowed to air dry. Images of wild-type p88 or p33/p55 mixed with Brucella broth, DDM, chloroform, or bovine heart lipids in chloroform were recorded using a FEI morgagni run at 100 kV at a magnification of 36,000X. Images were collected on ATM 1Kx1K CCD camera. Images of p88 used for multi-reference alignment were collected on a FEI 120 KV electron microscope at a magnification of 67,000X. Images were recorded on DITABIS digital imaging plates (Pforzheim, Germany). The plates were scanned on a DITABIS micron scanner (Pforzheim, Germany), converted to mixed raster content (mrc) format, and binned by a factor of 2 yielding final images with 4.48 Å/pixel. Images of p33/p55 in DDM purified by gel filtration were taken on a 200 kV FEI electron microscope equipped with a field emission electron source and operated at an acceleration voltage of 120 kV at a magnification of 100,000X. Images were collected using a Gatan 4Kx4K CCD camera. CCD images were converted to mrc format and

binned by a factor of 4 resulting in final images with 4.26 Á/pixel. Images of both p88 and p33/p55 were taken under low-dose conditions using a defocus value of –1.5 um.

For alignment and averaging of p88 VacA and p33/p55 VacA in DDM, 9,871 and 1,273 images of p88 and p33/p55 VacA particles, respectively, were selected with Boxer and windowed with a 120 pixel side length (154). Image analysis was carried out with SPIDER and the associated display program WEB (155). The images were rotationally and translationally aligned and subjected to 10 cycles of multi-reference alignment and K-means classification. For analysis of p88 VacA, alignment particles were first classified into 20 class averages and 7 representative classes then were chosen as references for another cycle of multi-reference alignment. For analysis of p33/p55 VacA, alignment particles were first classified into 10 class averages and then four representative projections were chosen as references for another cycle of multi-reference alignment.

Results

Expression, purification and refolding of recombinant p33 VacA

In previous studies, it has not been possible to purify a functionally active form of the p33 domain (113). We attempted to purify the p33 VacA from *E. coli* extracts under native conditions but were unsuccessful. Therefore, we expressed and purified the recombinant p33 under denaturing conditions and then used dialysis to reduce the concentration of denaturants and allow the protein to refold. After the p33 protein was refolded, it eluted as a well-defined peak by size exclusion chromatography (Figure 15).

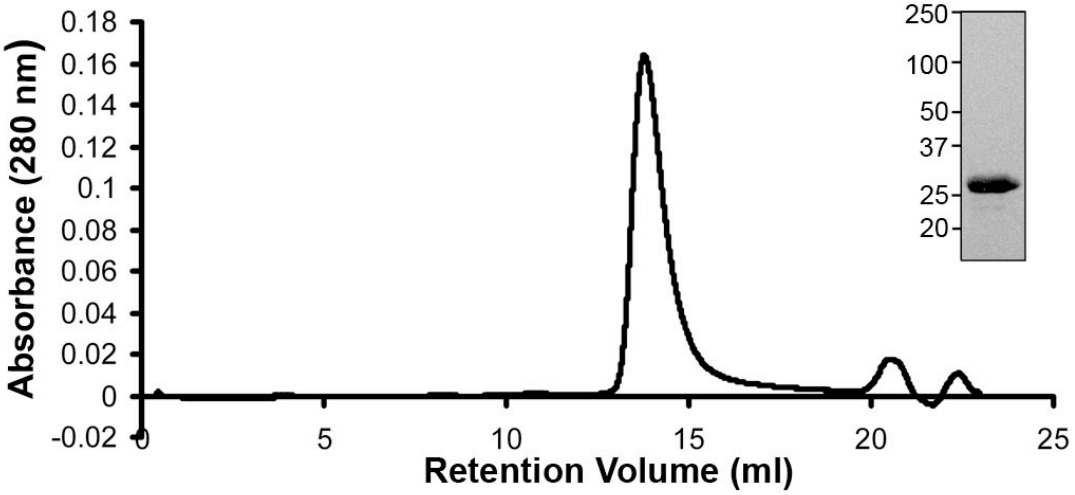


Figure 15. Purification of recombinant p33 VacA. SDS-PAGE and Coomassie blue stain of p33 VacA purified under denaturing conditions (inset). Gel filtration chromatography (Superdex 200 10/300 GL high-resolution resin) of p33 VacA after protein refolding, using buffer containing 800 mM guanidine and 800 mM arginine, as described in Methods.

Refolded p33 mixed with purified p55 causes cellular alterations

To test the activity of the purified p33 and p55 proteins, we added these proteins individually and in combination to HeLa cells, and analyzed the capacity of the proteins to cause cell vacuolation, a hallmark of VacA activity. No detectable vacuolating activity was observed when the p33 or p55 protein was added to cells individually, as demonstrated by neutral red uptake and light microscopic examination of the cells (Figure 16A). Similarly, none of the buffers alone or in combination exhibited any detectable activity. In contrast, a mixture of the purified p33 and p55 proteins caused extensive vacuolation of HeLa cells (Figure 16A and 16B). The potency of the p33/p55 mixture was slightly lower than that of the p88 VacA protein purified from *H. pylori* broth culture supernatant (Figure 16B). A mixture of p55 plus heat-denatured p33 failed to cause any detectable effects on cells (Figure 17).

Previous studies have shown that VacA from *H. pylori* inhibits production of IL-2 by Jurkat cells (88). To test whether p33 and p55 proteins exhibit a similar activity, we incubated Jurkat cells with the purified p33 and p55 proteins individually and in combination. When added individually, neither p33 nor p55 had any effect on IL-2 secretion (Figure 16C). In contrast, the p33/p55 mixture inhibited IL-2 secretion from Jurkat cells (Figure 16C and Figure 16D). The potency of the p33/p55 mixture was slightly lower than that of the p88 VacA protein purified from *H. pylori* (Figure 16D). Collectively, these results indicate that the refolded p33 is biologically active and is capable of causing alterations in eukaryotic cells.

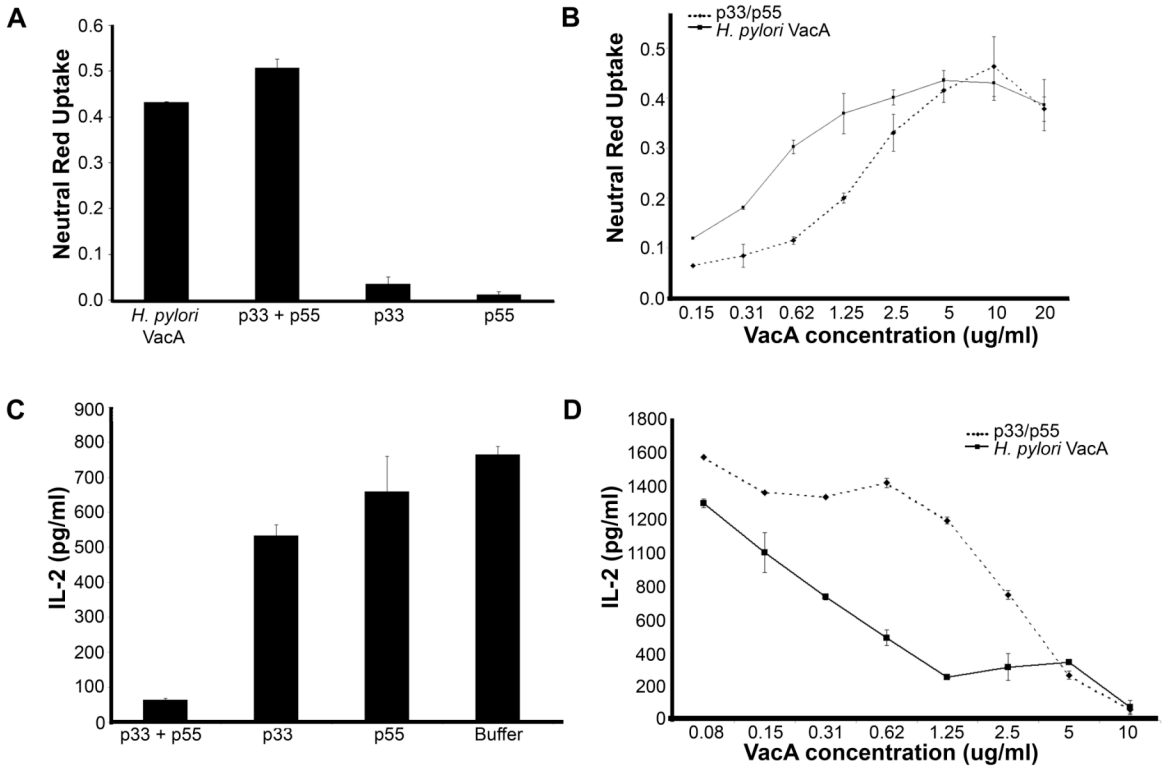


Figure 16. Effects of p33 and p55 VacA proteins on HeLa cells and Jurkat cells. Purified refolded p33 and purified p55 (each 1 mg/ml) were mixed together in a 1:1 mass ratio, which ensured an excess of p33 on a molar basis. The p88 VacA protein purified from *H. pylori* culture supernatant was acid-activated prior to contact with cells (106, 110), whereas the p33 and p55 preparations were not acid-activated. (A) HeLa cells were incubated with the purified VacA proteins at a final concentration of 10 µg/ml (or 5 µg/ml of each protein in the case of p33/p55 mixture). Cell vacuolation was quantified by neutral red uptake assay (OD 540 nm). (B) HeLa cells were incubated with the indicated final concentrations of a p33/p55 mixture (20 µg/ml corresponds to 10 µg/ml p33 and 10 µg/ml p55), or the p88 form of VacA purified from *H. pylori* broth culture supernatant. Cell vacuolation was quantified by neutral red uptake assay. (C) Jurkat cells were incubated with the indicated purified VacA proteins at a concentration of 6 µg/ml (or 3 µg/ml of each protein in the case of the p33/p55 mixture) for 30 min at 37°C. The cells were then stimulated and IL-2 secretion was measured as described in Experimental Procedures. (D) Jurkat cells with the indicated were incubated with the indicated final concentrations of a p33/p55 mixture or the p88 form of VacA purified from *H. pylori* culture supernatant. The cells were then stimulated and IL-2 secretion was measured as described in Methods. Results represent mean ± standard deviation, based on analysis of triplicate samples.

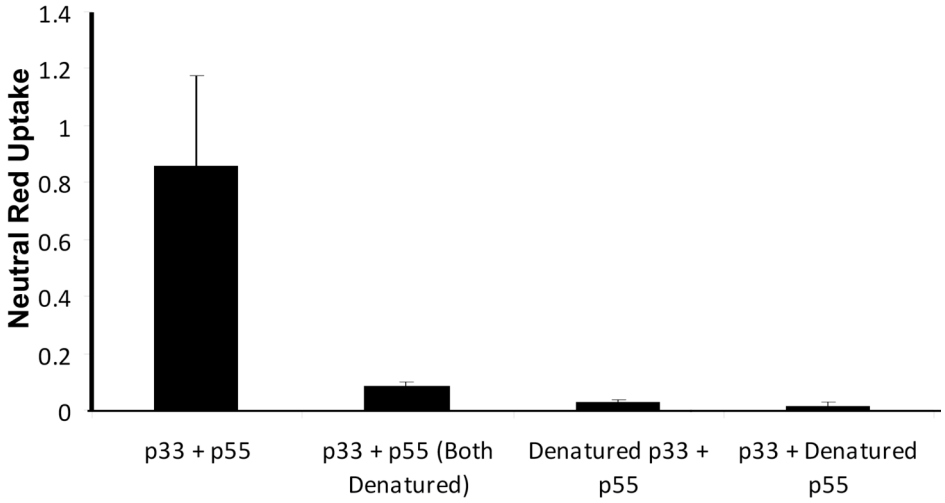


Figure 17. Effects of denatured p33 and p55 VacA proteins on HeLa cells. p33 was refolded as described in Materials and Methods. Purified p55 and p33 (each 1 mg/ml) were mixed together in a 1:1 mass ratio, which ensured an excess of p33 on a molar basis. HeLa cells were incubated with the indicated heat denatured purified VacA p33/p55, or untreated p33/p55 as a control, at a final concentration of 10 µg/ml. Cell vacuolation was quantified by neutral red uptake (OD 540 nm). Results represent mean ± standard deviation, based on analysis of triplicate samples.

Refolded p33 Δ6-27 exhibits a dominant negative effect

When certain mutant VacA proteins (e.g. VacA Δ6-27) are mixed with wildtype VacA, the mutant proteins can act as dominant-negative inhibitors of wildtype VacA activity (67, 105, 120, 121, 123). To further validate the new methods for expression and refolding of p33 proteins, we expressed, purified and refolded the p33 Δ6-27 protein under the same conditions used for purification and refolding of the p33 wild-type protein. When added to cells individually or in combination with purified p55, the p33 Δ6-27 protein did not cause detectable cell vacuolation (Figure 18A). To test for dominant-negative properties of the mutant protein, we pre-mixed the p33 Δ6-27 protein with p33/p55 mixtures that were known to be active (Figure 18A). When this p33/p55/p33 Δ6-27 mixture was added to cells, no detectable vacuolation was observed, indicating that the mutant protein exhibited a dominant-negative effect (Figure 18B). Thus, the purified refolded p33 Δ6-27 protein, when mixed with purified p55, exhibited dominant-negative inhibitory properties similar to those of the p88 Δ6-27 protein.

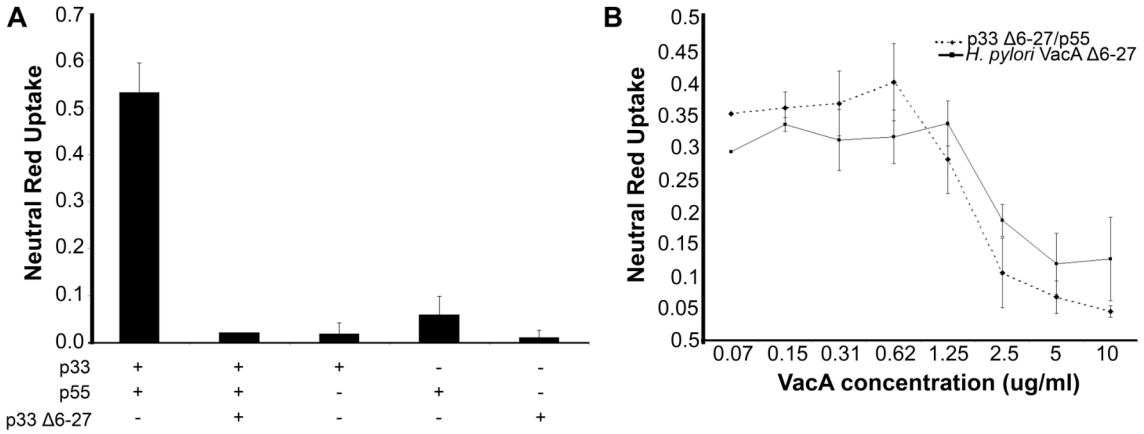


Figure 18. Refolded p33 ∆6-27 exhibits dominant negative properties. (A) VacA p33 ∆6-27 was purified and refolded as described in Experimental Procedures. Purified p33 ∆6-27 was mixed with p55 and p33 (each 1 mg/ml) at a 1:1:1 mass ratio. HeLa cells were then incubated with the indicated recombinant VacA proteins (either individually or in a mixture) at a final concentration of 10 μg/ml for 9 hours at 37°C. Cell vacuolation was quantified by neutral red uptake assay. (B) Wildtype p88 VacA (5 μg/ml) was incubated with the indicated concentrations of the VacA p33Δ6-27/p55 mixture or the p88 Δ6-27 VacA protein purified from *H. pylori* culture supernatant. Cell vacuolation was quantified by neutral red uptake assay. Results represent the mean ± standard deviation, based on analysis of triplicate samples.

Interactions of p33 and p55 with HeLa cells

Several previous studies have shown that sequences within the p55 domain contribute to the binding of p88 VacA to cells, and it has been suggested that p55 functions as a cell-binding domain (68, 118). To investigate the cell-binding properties of p55 in further detail, we incubated HeLa cells with purified fluorescently-labeled p55. Very little if any interaction of purified p55 with HeLa cells was observed (data not shown). In contrast, when p55 was incubated with HeLa cells in the presence of purified refolded p33, a marked increase in the binding and uptake of p55 by cells was observed (data not shown). Thus, p33 markedly enhanced the cell-binding properties of p55. Further studies indicated that when a mixture of p33 and p55 was incubated with cells, both p33 and p55 bound to the cell surface (data not shown). These properties of purified p33 and p55 proteins are consistent with previously observed properties of p33 and p55 proteins contained in crude *E. coli* extracts (113).

Interaction of refolded p33 with purified p55

To investigate potential interactions among the purified p33 and p55 proteins, we performed size exclusion chromatography experiments. When the refolded wildtype p33 protein was analyzed, a peak with a predicted mass of about 96 kDa was observed (Figure 19, red peak with star). When the purified p55 protein was analyzed, a peak with an approximate molecular mass of 178 kDa observed (Figure 19, green peak with star). When the p33/p55 mixture was analyzed, a peak with a predicted mass of 86 kDa was observed (Figure 19, blue peak with star), the 96 kDa peak (corresponding to p33 alone) was lost, and the 178 kDa peak (corresponding to p55 alone) was minimized. Representative fractions were tested by SDS-PAGE and Coomassie blue staining; this revealed an approximate 33 kDa band for the VacA 96 kDa peak, a 55 kDa band for the 178 kDa peak, and two protein bands of 33 kDa and 55 kDa for the 86 kDa peak (Figure

19B). When tested in cell culture assays, the p33/p55 mixture corresponding to the blue peak in Figure 19 caused cell vacuolation with a potency similar to that shown in Figure 16. Taken together, these results suggest that the refolded p33 protein interacts with the purified p55 protein to yield a p33/p55 complex. Moreover, these data suggest that p33 oligomers and p55 oligomers must undergo disassembly in order to interact with each other and form 88 kDa p33/p55 complexes.

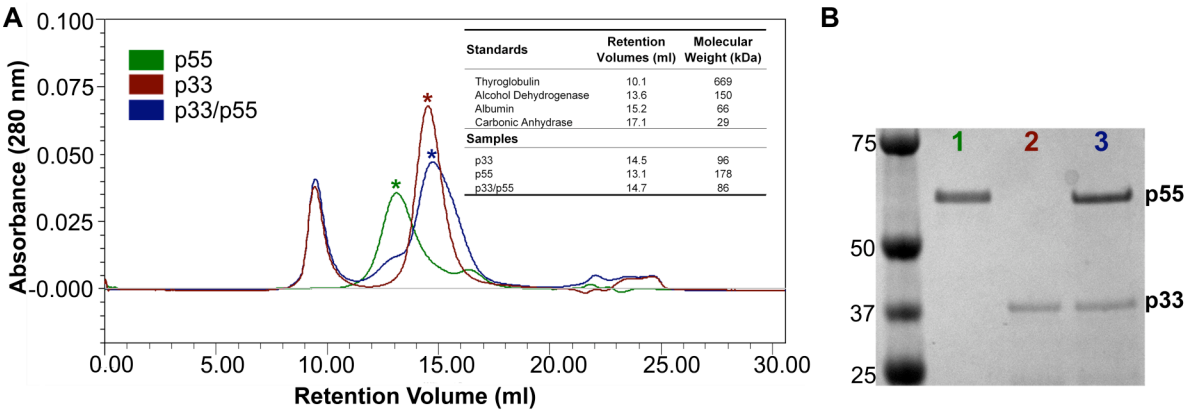


Figure 19. Analysis of p33 and p55 proteins by gel filtration. (A) Size exclusion chromatography (Superdex 200 10/300 prep grade resin) of refolded p33 (red peak), purified p55 (green peak) or a mixture of the two proteins (blue peak). Refolded p33 and purified p55 (each 1 mg/ml) were mixed at a 2:1 mass ratio and injected into the sizing column, as described in Materials and Methods. The buffer contained 800 mM guanidine and 250 mM arginine. The inset shows retention volumes of p33, p55, and the p33/p55 mixture in comparison to standard proteins. (B) The lower-molecular-mass peaks (stars) from each of the size exclusion chromatography experiments shown in panel A were analyzed by SDS-PAGE and Coomassie blue staining.

Assembly of p33/p55 complexes into oligomeric structures

The p88 VacA protein secreted by *H. pylori* can assemble into water-soluble oligomers (60, 106, 107, 109). To investigate the possibility that p33 and p55 domains might assemble into similar structures, we visualized the p33/p55 mixture (purified by gel filtration as a monomeric complex) by EM. VacA p88 oligomers purified from *H. pylori* culture supernatant (and exchanged into guanidine- and arginine-containing buffer by gel filtration) were analyzed as a control. As expected, large flower-like structures were visualized in preparations of *H. pylori* p88 VacA (Figure 20A). In contrast, the p33/p55 mixture consisted mainly of small rod-like particles (Figure 20B), similar to the appearance of p88 monomers produced by *H. pylori* (106, 107).

To explain why p88 proteins in *H. pylori* broth culture supernatant readily assemble into flower-like oligomeric structures whereas purified p33 and p55 proteins do not, we hypothesized that the broth culture medium used for growth of *H. pylori* (a nutrient-rich medium prepared from yeast extract and animal tissue, known as Brucella broth) might contain factors that promote VacA oligomerization. To test this hypothesis, we examined the appearance of the p33/p55 mixture by EM, either in the presence or absence of added Brucella broth. In the presence of added Brucella broth, increased formation of flower-shaped complexes was detected (Figure 20C). These experiments indicated that Brucella broth stimulates the oligomerization of p33/p55 mixtures into oligomeric structures similar to those formed by p88 VacA from *H. pylori*.

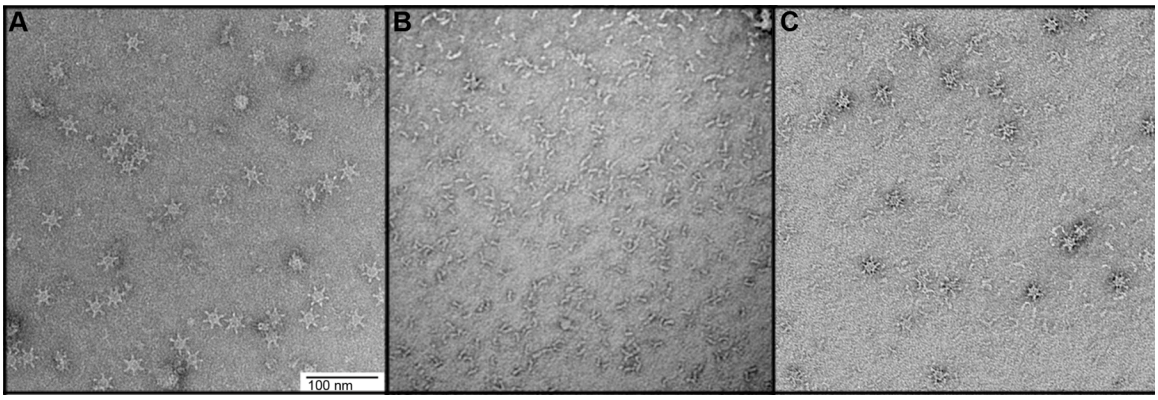


Figure 20. Assembly of p33 and p55 proteins into oligomeric structures. EM analysis of (A) p88 purified from *H. pylori* culture supernatant, and then exchanged into a guanidine-containing buffer by gel filtration, or (B) a mixture of refolded p33 and p55 eluted from the sizing column (corresponding to Figure 19, blue peak with star). The p33/p55 preparation shown in panel B was then mixed with different components as described in Methods and analyzed by EM. (C) p33/p55 plus Brucella broth.

High-resolution imaging of p33/p55 oligomeric complexes

We reasoned that the complex mixture of components in Brucella broth, including numerous membrane-derived factors, promoted the formation of flower-like oligomers. In an effort to stimulate VacA oligomerization using more refined conditions, we incubated p33/p55 mixtures with various additives designed to create an amphipathic environment, including bovine heart total extract solubilized in chloroform, chloroform alone, and detergent (DDM). Each of these additives promoted oligomerization of the p33/p55 monomeric complexes into flower-like oligomeric structures. The VacA oligomers formed in the presence of bovine heart extract or chloroform had a more heterogeneous appearance than the VacA oligomers formed in the presence of DDM, and therefore, we studied the latter oligomers in further detail. To permit higher resolution imaging, p33/p55 monomeric complexes (corresponding to the 86 kDa blue peak in Figure 19) were mixed with detergent, dialyzed, and passed over a gel filtration column in the presence of detergent and arginine and absence of guanidine. Under these conditions, the 86 kDa peak was minimized and a high-molecular mass (>300 kDa) peak was observed. High-molecular mass complexes containing wild-type p33 and p55 were isolated and analyzed further by EM. The appearance of these oligomers (Figure 21A) was similar to that of the p33/p55 complexes shown in Figure 20 and p88 oligomers isolated from *H. pylori* broth culture supernatant (Figure 21B). To further characterize the structural features of p33/p55 oligomers, approximately 1,300 particles were classified into 10 groups and four classes were chosen as references for an additional round of reference based-alignment (Figure 21C). To directly compare the structural organization of p33/p55 oligomers with that of p88 oligomers purified from *H. pylori* broth culture supernatant, class averages of p88 oligomers were also generated.

Because p88 oligomers seemed to adopt a larger number of conformations than p33/p55 oligomers, a larger number of p88 images were classified. Approximately 10,000 particles of p88 VacA were classified into 20 class averages and seven classes were chosen for an additional round of reference based-alignment (Figure 21D).

The result of the p33/p55 alignment (Figure 21C) shows that the majority of the p33/p55 oligomers are composed of six or seven subunits (67%, Figure 21C, panels 1 and 2), with one smaller class composed of an oligomer with 12 visible subunits (22%, Figure 21C, panel 3), and one class representing poorly formed oligomers (Figure 21C, panel 4). The 12-subunit complex may represent a double-layered oligomer with the two layers splayed (60, 107). The overall appearances of hexameric and heptameric p33/p55 oligomers are reminiscent of single-layer hexameric and heptameric oligomers formed by p88 VacA (Figure 21D, panels 1 and 2) (60, 106, 107). These single-layered oligomers exhibit a striking chirality, which suggests that one surface adsorbs preferentially to the support film. In contrast to the p33/p55 oligomers, a majority of the p88 oligomers exist as double-layered complexes containing 12-14 subunits (77%, Figure 21D, panels 3 though 7) (60, 106, 107). Importantly, difference maps created between averages of p33/p55 and p88 single-layer heptameric and hexameric oligomers did not show any statistically relevant difference peaks, which indicates that these oligomeric forms are structurally equivalent.

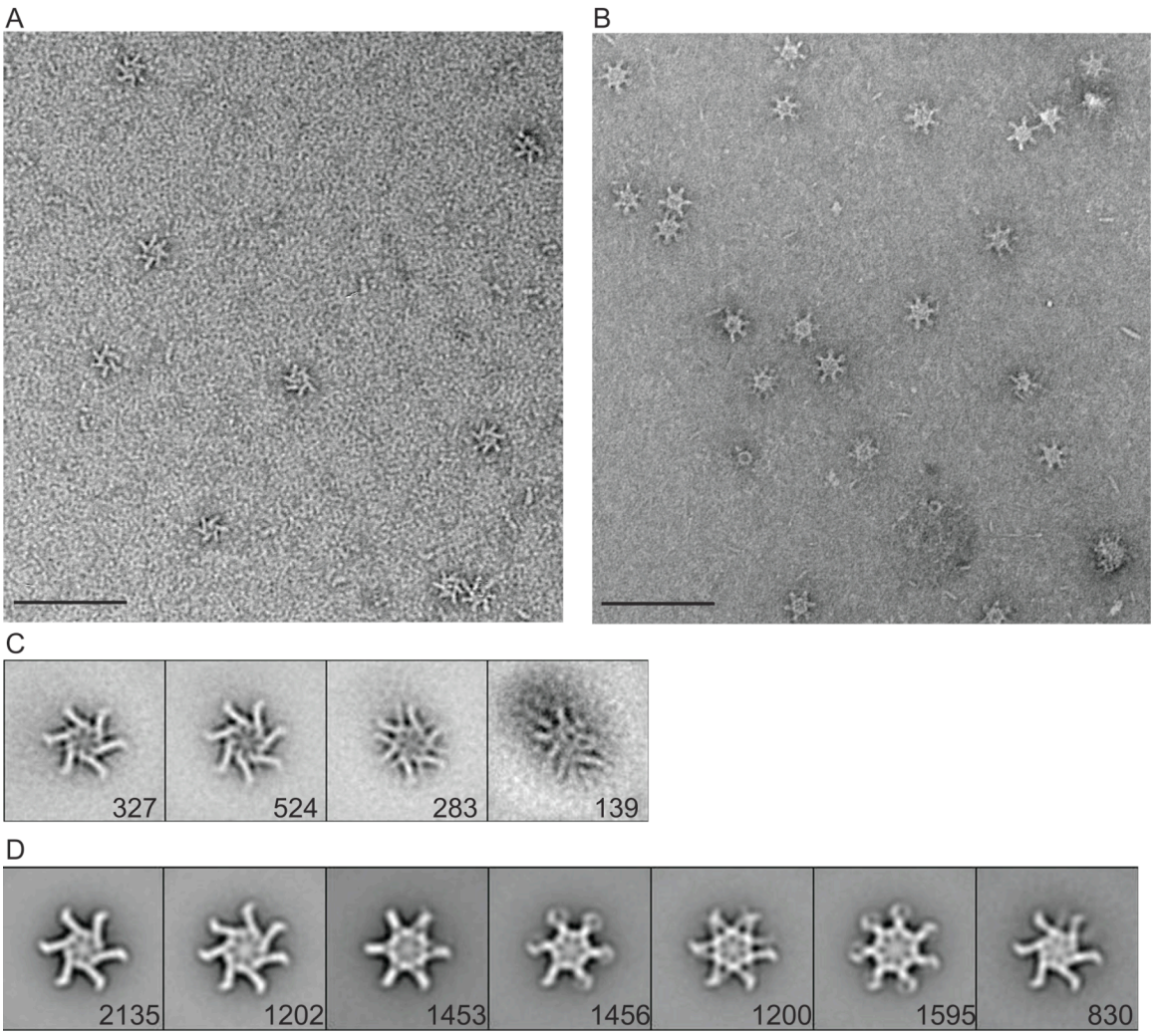


Figure 21. Analysis of p33/p55 VacA oligomers in negative stain. Mixtures of refolded p33 and p55 eluted from the sizing column (corresponding to Figure 19, blue peak with star) were mixed with DDM, and then dialyzed overnight in buffer containing 55 mM Tris pH 8.0, 21 mM NaCl, 0.88 mM KCl, 250 mM arginine, and DDM. The protein was passed over a gel filtration column that was equilibrated with dialysis buffer containing DDM, and VacA oligomers eluting in a high-molecular-mass fraction were then analyzed by EM. (A) Representative image of negative stained p33/p55 VacA oligomers eluting in a high-molecular-mass fraction. Scale bar, 100 nm. (B) Representative image of negative stained p88 VacA oligomers isolated from *H. pylori* broth culture supernatant. Scale bar, 100 nm. (C) Four class averages of p33/p55 VacA particles in negative stain generated from reference-based alignment. The number of particles in each projection average is shown in the lower right corner of each average. Side length of individual panels is 511 Å. (D) Seven representative class averages of p88 VacA particles in negative stain generated from reference-based alignment. The number of particles in each projection average is shown in the lower right corner of each average. Side length of individual panels is 538 Å.

Discussion

In this study, we demonstrate that a functionally active form of *H. pylori* VacA can be reconstituted from two purified VacA fragments (p33 and p55). Previously, the p55 fragment was purified and its crystal structure determined (148), but it was not possible to purify a soluble, functionally-active form of p33. In the current study, we purified the p33 domain under denaturing conditions and then employed a series of steps designed to allow the protein to refold and remain soluble. We found that the refolded p33 protein was soluble in a buffer containing 800 mM guanidine and 250 mM arginine, but upon removal of these additives, the p33 protein became insoluble. Nevertheless, in comparison to denatured p33, the refolded p33 protein exhibited functional activity when mixed with the p55 fragment, which suggests that the p33 protein was successfully refolded.

Previous studies reported that a mixture of *E. coli* lysates containing VacA p33 and p55 can cause vacuolation of HeLa cells (113), and intracellular co-expression of p33 and p55 in HeLa cells results in cell vacuolation (115, 122). However, there are numerous limitations associated with the use of crude *E. coli* lysates or intracellular expression systems. By using purified p33 and p55 proteins in the current study, we were able to monitor the process by which p33 and p55 proteins interact to yield a functionally active VacA protein. Specifically, we demonstrate that the p33 and p55 proteins were purified with molecular masses of about 96 kDa and 178 kDa, respectively. The mass of the p33 protein is consistent with a trimeric form, but efforts to validate this by EM were unsuccessful. The mass of the p55 protein is consistent with a trimer as well, but the crystal structure of p55 revealed a head-to-head packed dimer that adopts an elongated dumbbell shape (148). The elongated shape and the unusual

buffer conditions likely account for the high apparent molecular mass of p55 on the sizing column. Upon mixing of the p55 and p33 preparations, the p55 and p33 homo-oligomers each dissociated to yield a p55/p33 complex with a mass of about 86 kDa, corresponding to a complex containing one p55 subunit and one p33 subunit. The monomeric complexes were visible by EM as elongated rods, consistent with the hypothesis that p33 extends the β-helix structure observed in p55 (148).

The ability to reconstitute a functional protein from two individually-expressed component domains is somewhat unusual among bacterial protein toxins, and unusual among proteins in general. This phenomenon is probably facilitated by the elongated β-helical structure of VacA, a structural feature that is predicted to be shared by many autotransporter passenger domains (139, 148, 156). The VacA p55 domain consists predominantly of a β-helix, composed of multiple ~25 amino acid repeats, each of which forms a 3 β-strand triangle-shaped rung (148). Adjacent coils are held together by backbone hydrogen bonds. The β-helix is therefore very different from globular proteins where adjacent structural elements are held together with an intricate arrangement of sidechain interactions. Based on computer modeling, the VacA p33 domain is also predicted to comprise a β-helical structure, and it is predicted that the p88 protein comprises an elongated continuous β-helical structure (148). In the current experiments, we speculate that the C-terminal coil of p33 interacts with the N-terminal coil of p55, recapitulating the structural relationship that exists between these two domains in the intact p88 VacA protein (148). If generally true, the ability to assemble proteins *in trans* using a β-helical packing motif could have exciting implications for the generation of novel protein functionalities.

A distinctive property of the p88 VacA protein secreted by *H. pylori* is its ability to assemble into water-soluble, flower-shaped oligomeric structures (60, 106, 107, 109). In the current study, we observed that purified p33 and p55 proteins interact to form ~86

kDa complexes, but do not readily assemble into oligomeric structures when maintained in buffer alone. One possible explanation is that the guanidine and arginine constituents of the buffer (required for maintenance of p33 solubility) prevent VacA oligomerization; however, we observed that these agents did not cause disassembly of p88 oligomers purified from *H. pylori* culture supernatant. We hypothesized that *H. pylori* broth culture supernatant might contain factors (either components of the rich Brucella broth medium used for culture of *H. pylori,* or additional *H. pylori* products) that allow VacA oligomers to form. We observed that, indeed, the addition of freshly prepared Brucella broth (not previously cultured with *H. pylori*) to purified p33/p55 mixtures promoted assembly of VacA into oligomeric structures. Similarly, the addition of detergent also stimulated oligomerization. We speculate that oligomerization is stimulated by exposure to an amphipathic environment, and that the oligomerization observed in these experiments mimics the process by which VacA oligomerizes when in contact with membranes of host cells.

The p88 VacA protein is typically purified in an oligomeric form from *H. pylori* broth culture supernatant (56, 60, 106, 107, 109), and monomeric forms of p88 VacA have been relatively difficult to purify. When added to cultured eukaryotic cells, purified p88 VacA oligomers lack detectable activity in most assays unless the oligomers are first exposed to low pH or high pH conditions, which results in oligomer disassembly; oligomers have been observed to reassemble if the pH is returned to neutral (61, 78, 80, 98, 101, 106, 110). A current model presumes that VacA monomers interact with the cell surface, and then reassemble into oligomeric complexes that function as membrane channels. In the current study, we demonstrate that a mixture of purified p33 and p55 proteins is fully active in cell culture assays in the absence of low pH or high pH activation. Since the p33/p55 mixture predominantly consists of a p88 complex (Figures

19 and 20), this provides additional support for a model in which VacA monomers interact with the plasma membrane.

Several lines of evidence indicate that oligomerization of p88 VacA is required for VacA-induced cellular alterations. Specifically, mutant forms of VacA that fail to oligomerize properly lack detectable activity in most cell-based assays (120, 121), and dominant-negative mutant forms of VacA interact with wild-type VacA and thereby abrogate wild-type VacA activity (67, 105, 120, 121). VacA oligomerization is presumably required for membrane channel formation (61, 78, 79), and we speculate that oligomerization also may be required for VacA internalization. Potentially oligomerization of VacA occurs preferentially within lipid raft components of the plasma membrane (62, 103, 157). VacA oligomeric structures have been visualized on the surface of VacA-treated cells or lipid bilayers (60-62), and in contrast to double-layered oligomeric forms of VacA found in *H. pylori* culture supernatant, there is evidence that the VacA oligomeric complexes formed on the surface of cells are single-layered (61). In the current study, we observed that detergent promoted assembly of p33/p55 mixtures into predominantly single-layered oligomeric structures. Therefore, the complexes visualized in the current study are predicted to be useful models for VacA channels that form in the context of human cells.

The reconstitution of VacA activity from purified p33 and p55 components probably involves a complex series of molecular events. An initial step involves disassembly of p33 and p55 homo-oligomers and formation of a p33-p55 complex. Potentially the presence of p55 disrupts p33-p33 interactions, or the presence of p33 may disrupt p55-p55 interactions. An important observation is that neither p33 nor p55 bound to cells when added individually, whereas the p33/p55 mixture exhibited binding to cells (data not shown). One possible explanation is that the homo-oligomeric forms of p33 and p55 lack cell-binding activity, and cell-binding surfaces become exposed upon

disassembly of the homo-oligomeric complexes. Alternatively, the receptor-binding site(s) may span both domains. Finally, the assembly of p33/p55 complexes into higher-order flower-shaped oligomers may stabilize the interaction of VacA with the surface of eukaryotic cells, and also may be required for insertion of VacA into membranes and channel formation.

CHAPTER IV

MOLECULAR EVOLUTION OF *HELICOBACTER PYLORI* VACUOLATING TOXIN (VACA)

Introduction

H. pylori strains from unrelated humans exhibit a high level of genetic diversity that is adapted for colonization of the human stomach (17, 18). The population structure of *H. pylori* is panmictic, and the rate of recombination in *H. pylori* is reported to be among the highest in the Eubacteria (17, 33). Multilocus sequence analysis of housekeeping genes has revealed the presence of at least nine different *H. pylori* populations or subpopulations that are localized to distinct geographic regions (158, 159). Analysis of these sequences suggests that *H. pylori* has spread throughout the world, concurrently with the major events of human dispersal, and thus *H. pylori* is potentially a useful marker for the geographic migrations of human populations (158). While the genetic diversity patterns are consistent with this theory, there has not yet been any formal phylogenetic assessment of the ancestry of *H. pylori* in human populations.

All strains of *H. pylori* contain a chromosomal *vacA* gene, but individual strains differ considerably in levels of VacA activity (57, 63). Genetic variation at this locus could be under strong selection as *H. pylori* adapts to the host immune response, colonizes new human hosts, or inhabits different host environments. Previous phylogenetic studies analyzed relatively small numbers of *vacA* sequence fragments and focused mainly on analysis of nucleotide sequences rather than amino acid sequences (17, 63, 69, 124, 125, 160). Two studies analyzed *vacA* sequence encoding a fragment

of the p33 domain and did not detect any recognizable phylogenetic structure (star or bush type pattern), presumably due to the presence of extensive recombination (17, 125). Other studies analyzed different regions of VacA and detected polymorphisms that allow classification of *vacA* alleles into distinct families (designated s1/s2, i1/i2, or m1/m2) depending on the presence of signature sequences in different regions of VacA (63, 124, 161). In general, strains containing *vacA* alleles classified as s1, i1, or m1 have been associated with an increased risk of ulcer disease or gastric cancer compared to strains containing *vacA* alleles classified as s2, i2, or m2 (63, 70, 161). Geographic differences have been detected within several of these *vacA* regions (64, 69, 162-164).

H. pylori strains that produce an active VacA protein (type s1 VacA) typically express CagA, and strains that produce inactive VacA proteins (type s2 VacA) typically lack the *cagA* gene (63). *vacA* and the *cag* PAI localize to distant sites on the *H. pylori* chromosome, and therefore, the basis for this association has been unclear. There are complex relationships between the cellular effects of VacA and CagA, whereby VacA can downregulate CagA's effects on epithelial cells, or vice versa (165-168). This functional interaction between VacA and CagA may represent a mechanism that allows *H. pylori* to minimize damage to gastric epithelial cells or minimize mucosal inflammation, thereby allowing it to persistently colonize the stomach.

In this chapter, we show that phylogenetic reconstructions of amino acid sequences specify three VacA groups with distinct geographic distributions. Divergence of the groups is principally due to sequence changes in the p55 domain, a central region responsible for binding of the toxin to host cells. Population genetic analyses specify that divergence of the p55 domain into three groups occurred by natural selection. The positively selected sites in the p55 domain map to surface-exposed residues in the VacA crystal structure. Moreover, a phylogenetic tree for the virulence determinant CagA is

comprised of three groups that are phylogeographically-similar to those of the VacA tree, suggesting coevolution of VacA and CagA due to common selective pressure. Finally, phylogenetic analysis of the origins of VacA using the close relative *H. acinonychis* as an outgroup reveals that the ancestry of VacA is strikingly different from the African origins that typify the core genome. Taken together, these results indicate previously unrecognized positive selection and coevolution in the VacA and CagA virulence determinants of *H. pylori*. The correspondence of positively selected VacA sites to surface-exposed residues of p55 is consistent with immune-mediated positive selection, since antibody responses are directed against p55.

Methods

VacA sequences. One hundred deduced VacA amino acid sequences (86 full-length gene sequences and 14 that were complete within the region encoding the p55 domain) were retrieved from Genbank. These sequences originated from *H. pylori* strains that were isolated from humans in many different regions of the world. To obtain additional VacA sequences of African origin, we analyzed *vacA* in five *H. pylori* strains that were isolated from patients in Africa and previously classified by MLST analysis as HpAfrica2 (strains 191.9 and 501.9), HspSAfrica (cc2c) or HspW Africa (D1a and D1b) (158, 159). The *vacA* locus was amplified using Expand Long Template PCR System (Roche Applied Science) with the primers described in Table 2 and the *vacA* sequences were determined. Sequences are deposited in GenBank (accession numbers pending). The *vacA* sequences from strains D1a and D1b each contained a frameshift mutation and were excluded from subsequent phylogenetic analyses. Sequences were aligned with MUSCLE and edited manually in MacClade 4.08 (169). The total length of aligned sequences was 1354 amino acids. All indels and hypervariable regions were removed,

resulting in a final alignment length of 1135 amino acids for the unrooted analysis and 971 amino acids for the rooted analysis.

Table 2 : Primers used for PCR-amplication of *vacA* from additional African strains.

Forward[1]	AAAATCGCTTTGATGGACACCCCA
Reverse[1]	CAAGCTTCCACGCCAATCGCTATGACT
Forward[2]	CTGCTGTAGGAACGGTCTC
Reverse[2]	ATCATCGCATTACTCAAGCTCAAAGTTTG

[1]These primers were used to amplify full-length *vacA* from African strains 191.9, 501.9, D1a, and D1b.

[2]These primers were used to amplify a *vacA* fragment (encodes p33, p55, SAP and N-terminal region of C-terminal β-Barrel domain) from African strain cc2c.

VacA reference sequences. VacA from strain 60190 (Genbank accession no. Q484245) was used as the reference sequence for amino acid numbering, in which residue 1 refers to alanine-1 of the secreted 88-kDa VacA protein. It is the prototype s1/m1 form of VacA and the crystal structure of the p55 domain of VacA from this strain has been determined (148). VacA from strain 95-54 (Genbank accession no. U95971) and strain Tx30a (Genbank accession no. Q48253) were used as reference sequences for s1/m2 and s2/m2 proteins, respectively (63, 68).

Criteria for classification of VacA sequences. VacA sequences were classified as m1 or m2 based on the absence or presence, respectively, of a 21-amino-acid insert within the p55 domain (between amino acid 475 and 476) (Figure 13) (63). We identified and excluded two m1/m2 chimeric VacA proteins (from strains ch2 and v225) in which tracts of recombination between m1 and m2 sequences were identifiable by eye. VacA sequences were classified as s1 or s2, based on the absence or presence, respectively, of a 9 amino acid insertion in the signal sequence region (63). VacA sequences were classified as i1 or i2, based on amino acid substitutions that fall into two clusters that have been previously denoted as cluster B and C (161).

Phylogenetic analyses. The unrooted phylogenetic distance trees for VacA and CagA were created using the Neighbor Joining method with the Jukes Cantor genetic distance model in Geneious 4.6.5 (170). Support for nodes on the neighbor-joining tree was assessed by 2000-replicates of bootstrap, and the majority rule consensus tree is shown.

Population genetic tests of selection. A sliding window analysis of dN/dS ratios was performed using VacA sequences from strains 60190 (m1 type) and 95-54 (m2 type) with the program DnaSP (171). Sliding window parameters included a window size of 50 bases and a step size of 10 bases. For further analyses, a total of 45 VacA sequences, corresponding to 15 VacA amino acid sequences from each VacA group,

were retrieved from Genbank. These strains are shown (bold) in Figure 23. Additionally, a total of 32 CagA sequences, corresponding to CagA amino acid sequences from each CagA group (6 from Group 1, 15 from Group 2, and 11 from Group 3), were retrieved. Sequences were assembled and aligned with Geneious and edited manually in MacClade 4.08 (169). All indels and hypervariable regions were removed. The McDonald-Kreitman test (172) was carried out on full-length *vacA*, individual regions of *vacA,* and full-length *cagA* sequences with the exclusion of low frequency variants less than or equal to 15% to reduce artifacts associated with detecting adaptive evolution. Next, we investigated the codon sites under selection using the programs omegaMap (173) and PAMLv4.3 (174, 175). The Bayesian method implemented in omegaMap incorporates intragenic recombination and does not assume a known fixed genealogy, so that recombination does not inflate the false detection rate of positive sites. The same *vacA* sequences from the MKT were used in omegaMap (15 sequences from each Group for a total of 45 sequences). We used 500,000 Markov-chain Monte Carlo samplings, and the first 50,000 iterations were discarded as a burn-in. The variable model was chosen with the average block length of 30 for both ω and ρ. One objective and one subjective approach to prior specification were used. First, inverse distributions were used for ω and ρ, and improper inverse distributions were used for the other parameters (μ, κ, and φ). Starting values for μ, κ, and φ were, respectively, 0.1, 1, and 1. In the second approach to prior specification, exponential distributions were used for all parameters (starting values were μ =0.1, κ = 1, φ = 1, ω = 1 and ρ = 0.001). Two independent runs were performed and both runs converged. For PAML, twelve full-length *vacA* sequences that encompass the range of genetic variation in VacA were selected for analysis. Codons that were subject to positive selection were identified using Nsites models (M1a, M7) that do not permit positive selection compared to models (M2a, M8) that permit sites to evolve under positive selection.

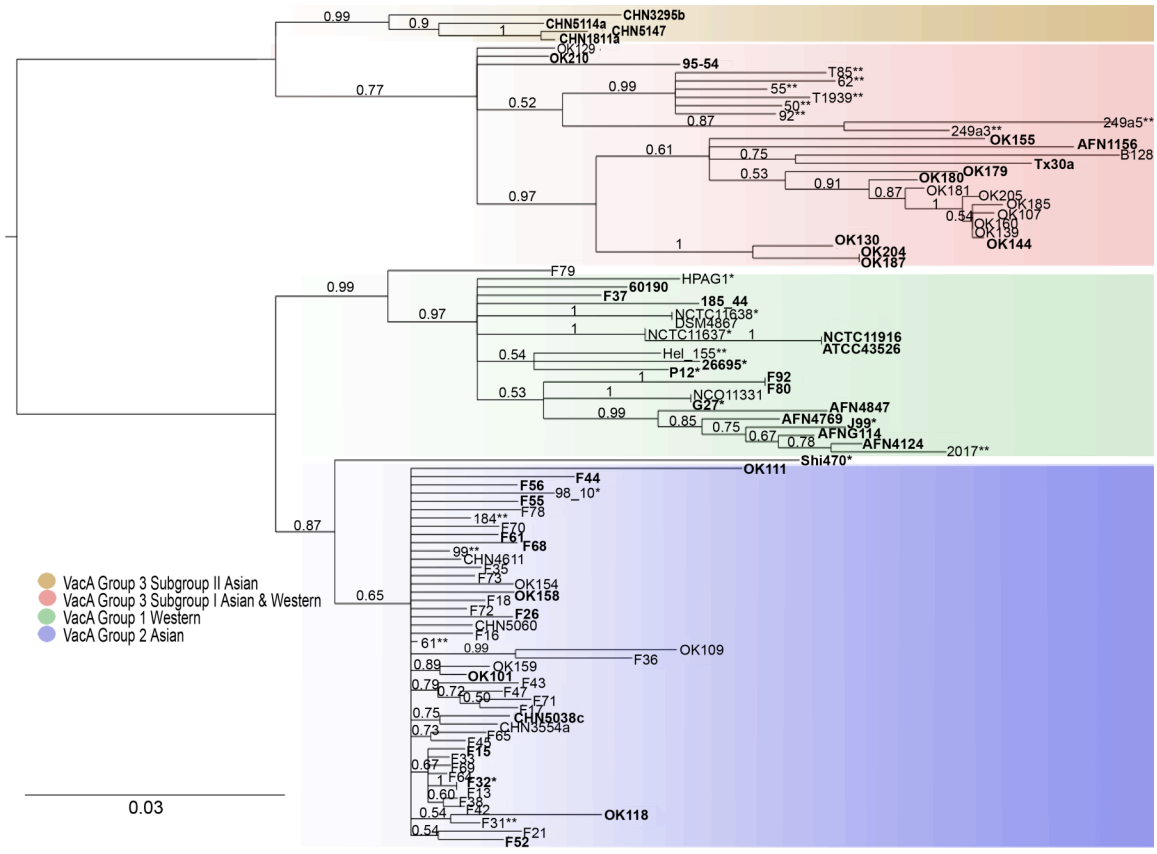


Figure 22. Ladderized version of the neighbor-joining VacA phylogenetic tree shown in Figure 23. This diagram shows the taxa used in the analyses and their evolutionary relationships. Single asterisks indicate taxa for which whole genome sequences are available; sequences from these strains were included in the MLST analysis and SH analysis (Figure 33 and Figure 35). Double asterisks indicate taxa that were incomplete in the p33 or β-barrel region. Bold labeling indicates taxa used in the omegaMap and MKT analysis (Table 3). The numbers represent posterior probability values for each node.

Reconstruction of a vacA pseudogene from H. acinonychis. The entire *vacA* pseudogene of *H. acinonychis*, corresponding to approximately nucleotides 443900 to 439500 in the genome sequence of strain Sheeba (176), was translated in all 3 reading frames, and the translated fragments with homology to *H. pylori* VacA were then concatenated. The VacA protein encoded by the reconstructed *H. acinonychisvacA* pseudogene consists of 1310 amino acids. A BLAST search indicates that the reconstructed *H. acinonychis* VacA sequence exhibits 64% amino acid identity to its closest match in *H. pylori* and retains a high level of relatedness to *H. pylori* VacA throughout the entire sequence.

MLST Analysis. Nucleotide sequences were retrieved from the *H. pylori* multilocus sequence typing database (http://pubmlst.org/helicobacter). This database contains sequence data (398 to 627 nucleotides per gene) for seven housekeeping genes (*atpA, efp, mutY, ppa, trpC, ureI, yphC*) (158). Concatenated nucleotide sequences were aligned using MUSCLE and edited manually in MacClade 4.08 (169). To permit rooting of an MLST tree, we retrieved orthologous sequences from the *H. acinonychis* genome (176). PhyloBayes and MrBayes inference methods were used to generate the rooted MLST trees and posterior probability values.

Rooted Phylogenetic Analyses of VacA sequences. PhyloBayes 2.3 was used to infer phylogenetic relationships of the rooted tree based on PhyloBayes and MrBayes inference methods. These analyses were performed with the site homogeneous models of Jones-Taylor-Thorton (JTT), and Whelan and Goldman (WAG), and the category amino acid site-heterogeneous mixture model (CAT) to suppress tree artifacts associated with long branch attraction (177). For all PhyloBayes and MrBayes analyses, at least two independent runs were performed with free equilibrium frequencies inferred from the data and gamma distributed rate variation with four discrete categories. Burn-

ins up to 20% of the sampled trees were used until a maxdiff value < 0.15 was achieved to ensure chain equilibration.

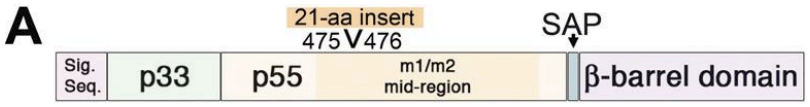
Shimodair-Hasegawa Test. We tested the significance of topological differences in VacA and MLST phylogenetic trees using the Shimodaira-Hasegawa (SH) test (178). The SH test compares the likelihood score (-lnL) of a given data set across its ML tree versus the -lnL of that data set across alternative topologies, which in this case are the ML phylogenies for other data sets. The differences in the -lnL values are evaluated for statistical significance using bootstrap (1000 replicates) based on RELL sampling.

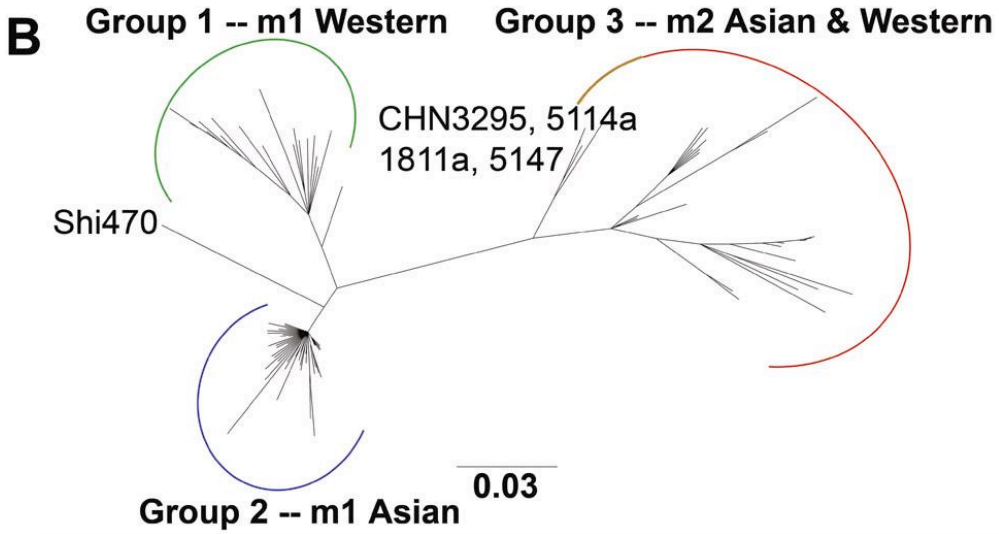
Results & Discussion

VacA phylogeography

An unrooted phylogenetic analysis of VacA protein sequences from 100 different *H. pylori* strains demonstrated three distinct groups (Figure 23 and Figure 24), corresponding to strains predominantly of Western origin (Group 1), Asian origin (Group 2), or Western and Asian origin (Group 3). Within Group 3 there is a subgroup of 4 sequences (from strains CHN5147, CHN1811a, CHN5114a and CHN3295b; designated subgroup II), all of which were from *H. pylori* strains isolated in Shanghai, China (164). *H. pylori* strain Shi470, located in the tree between Group 1 Western and Group 2 Asian populations, was isolated from an Amerindian patient in the Amazon (179). We examined the sequences within each group to detect previously described VacA signature sequences (63, 124, 161). Based on an analysis of indels (insertions and deletions, which were excluded when generating the tree) all of the sequences in Group 1 and Group 2 were classified as type m1, and all of the sequences in Group 3 were classified as type m2 (Figure 23). All of the VacA sequences in Groups 1 and 2 contain

a type s1 signal sequence region; three sequences within Group 3 contain a type s2 signal sequence, and the remainder contain type s1 signal sequences (Figure 23).





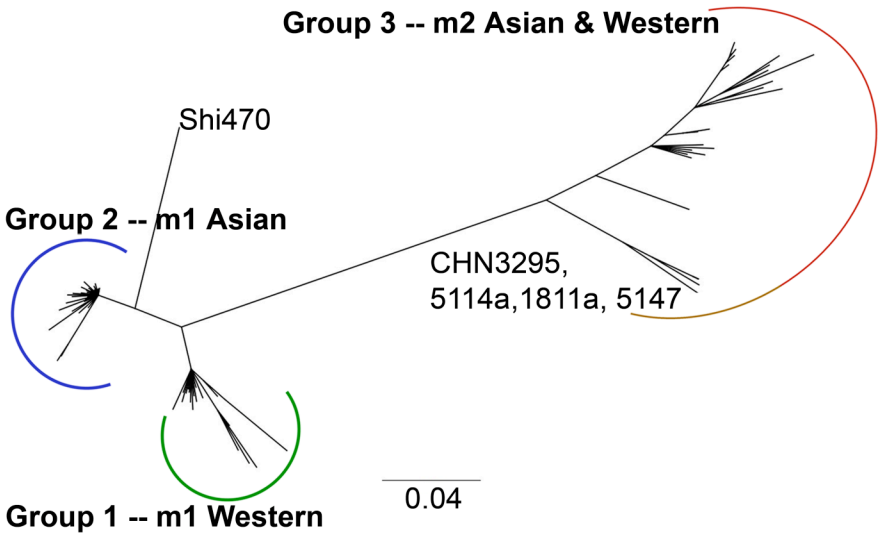
Classification	# of Strains	Origin		VacA Type			
		Western	Asian	s1	s2	m1	m2
Group 1	22	19	3	22	0	22	0
Group 2	46	0	46	46	0	46	0
Group 3							
Group 3 Subgroup I	26	5	21	23	3	0	26
Group 3 Subgroup II	4	0	4	4	0	0	4
Total	**98**	**24**	**74**	**95**	**3**	**68**	**30**

Figure 23. Analysis of VacA phylogeography. (A) The *vacA* gene encodes a 140-kDa protoxin, which undergoes cleavage to yield a signal sequence, a secreted 88-kDa toxin, a secreted alpha peptide, and a C-terminal β-barrel domain. The mature 88-kDa VacA toxin contains two domains, designated p33 and p55. The midregion sequence that defines type m1 and m2 forms of VacA is located within p55. A 21-amino acid insertion is present in m2 forms, but not m1 forms of VacA. (B) Neighbor-joining phylogenetic tree of 86 complete amino acid sequences and 14 partial sequences of VacA. Three major groups are evident. The chart shows the number of strains analyzed and characteristics of VacA sequences in each group of the tree. Group 1 comprises type m1 sequences mainly from strains of Western origin, Group 2 comprises m1 sequences of Asian origin and Group 3 comprises m2 sequences of both Asian and Western origin.

Analysis of VacA structural domains

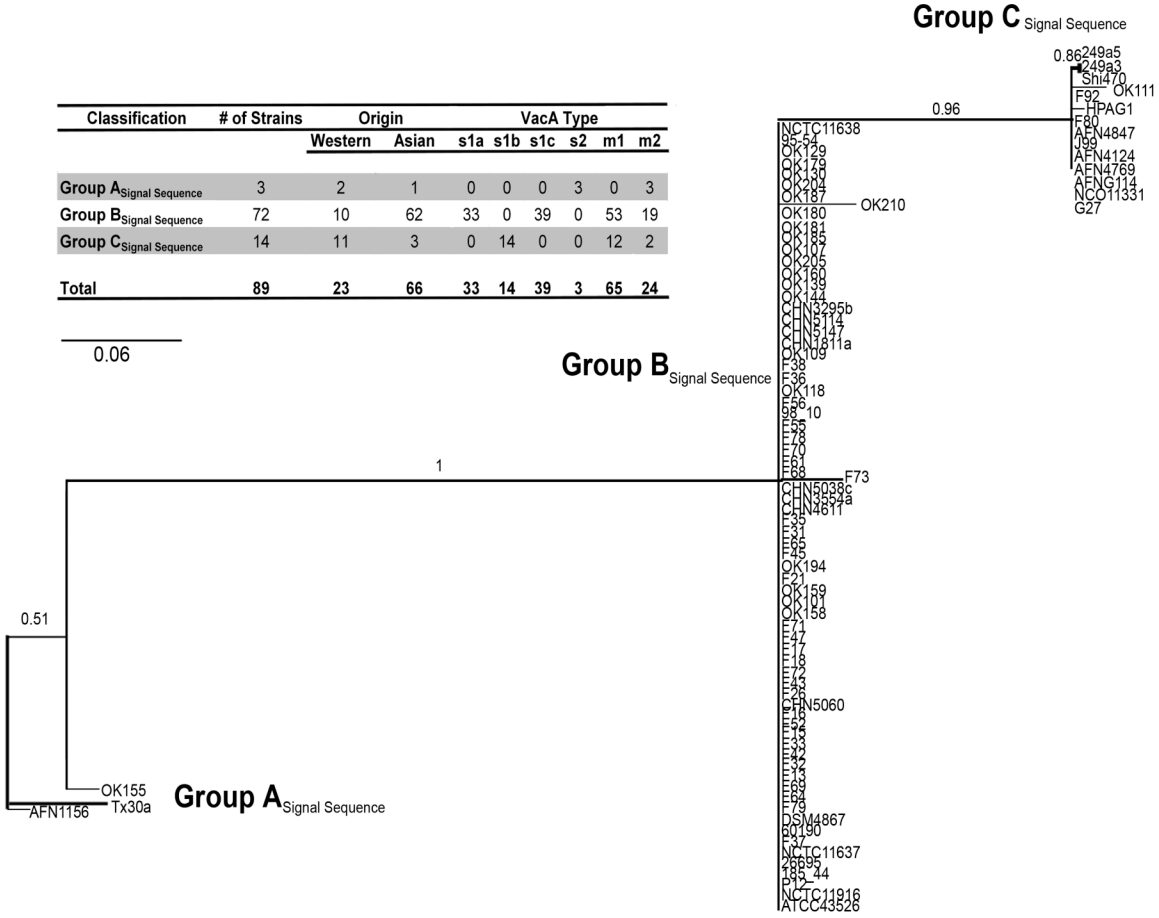
Next we performed phylogenetic analyses restricted to five putative structural domains of VacA (Figure 23A): p55, signal sequence region, p33, secreted alpha peptide (SAP), and the C-terminal β-barrel region (Figures 24, 25, 26, 27, and 28). The tree for the p55 domain (427 aligned amino acids) yielded a three-group pattern (Figure 24) that overlaps with the phylogeography of full-length VacA (Figure 23B). The other regions exhibited tree structures (see Figures 25-28) substantially different from those of full-length VacA or p55 trees. The phylogenetic structure of the signal sequence domain (Figure 25) is consistent with data reported in previous studies, reflecting a divergence between s1 and s2 forms of VacA and the presence of several s1 subtypes (s1a, s1b, s1c). In our analysis of the p33 domain, most of the sequences localized within a single group (Group 1), but 14 sequences were highly divergent (Group 2) (Figure 26). This branching pattern reflects differences within the intermediate region (type i1 and i2). The sequences in p33 Group 1 are characterized as type i1, with the exception of two sequences that appear to be i1-i2 hybrids, and sequences in p33 Group 2 are exclusively characterized as type i2. The branching pattern observed in an analysis of the p33 domain (Figure 26) is strikingly different from the results published in previous studies that analyzed ~500 bp of nucleotide sequences encoding a fragment of the p33 region; these previous studies described a single “star” pattern, and detected extensive recombination that obliterated any phylogenetic structure (17, 125). The phylogenetic structures of the alpha peptide and C-terminal β-barrel region (Figure 27 and Figure 28, respectively) have not been previously investigated, but distinct groups were also detected within these trees. Phylogenetic trees of the p55 domain, SAP, and beta-barrel domain each contained groups that comprised exclusively sequencess of East Asian origin. Other features of the phylogenetic structures, such as the divergence of m1 and m2 sequences within the p55 region, were independent of geography. Based on the

phylogeography patterns of the other domains, we can conclude that the localization of full-length VacA sequences to three main groups (Figure 23) is determined primarily by sequence divergence in the p55 region (Figure 24).



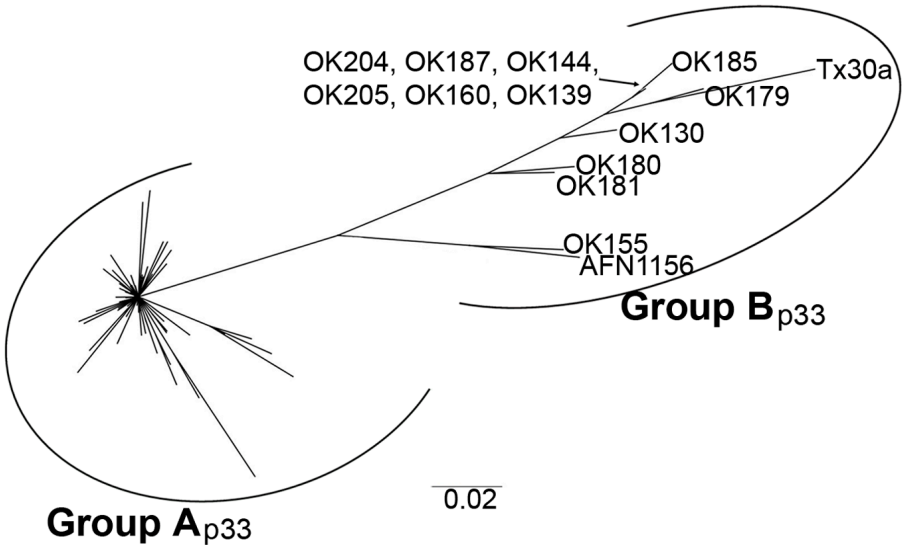
Classification	# of Strains	Origin		VacA Type			
		Western	Asian	s1	s2	m1	m2
Group 1	23	19	4	23	0	23	0
Group 2	46	0	46	46	0	46	0
Group 3							
Group 3 Subgroup I	26	5	21	23	3	0	26
Group 3 Subgroup II	4	0	4	4	0	0	4
Total	**99**	**24**	**75**	**96**	**3**	**69**	**30**

Figure 24. Neighbor-joining phylogeny of the VacA p55 domain. Three main groups are detected within this tree, designated Groups 1-3. The chart shows the number of strains analyzed and characteristics of VacA sequences in each group of the tree. This tree maintains the same pattern as the VacA full-length tree shown in Figure 1. The nomenclature for the primary VacA p55 groups (Groups 1, 2, and 3) is consistent with the nomenclature of groups in the full-length VacA tree (Figure 23).



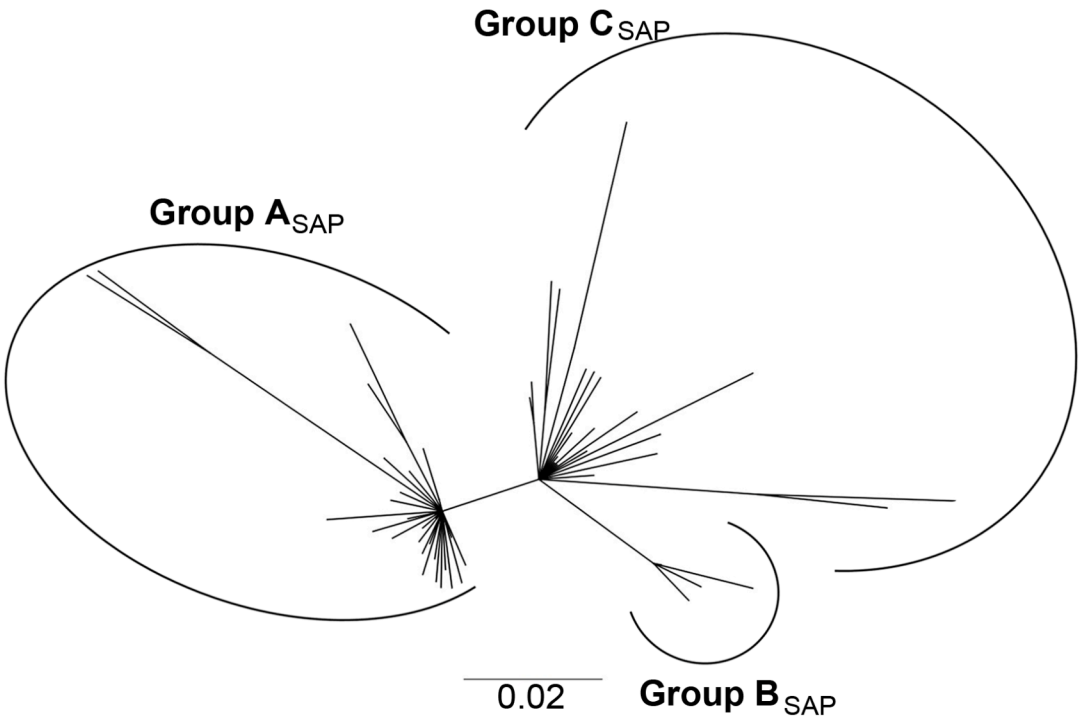
Classification	# of Strains	Origin		VacA Type					
		Western	Asian	s1a	s1b	s1c	s2	m1	m2
Group A Signal Sequence	3	2	1	0	0	0	3	0	3
Group B Signal Sequence	72	10	62	33	0	39	0	53	19
Group C Signal Sequence	14	11	3	0	14	0	0	12	2
Total	89	23	66	33	14	39	3	65	24

Figure 25. Neighbor-joining phylogeny of VacA signal sequences. This tree is a ladderized tree revealing three main groups (Groups A-$C_{Signal\ Sequence}$). The phylogenetic structure of the signal sequence domain is consistent with data reported in previous studies (63), reflecting a divergence between s1 and s2 forms of VacA and the presence of several s1 subtypes (s1a, s1b, s1c). Group $A_{Signal\ Sequence}$ contains three sequences that are highly divergent from all of the other sequences; these three sequences contain an insertion characteristic of type s2 VacA signal sequences, whereas all of the others lack this insertion (type s1). Group $B_{Signal\ Sequence}$ contains s1a and s1c sequences, whereas Group $C_{Signal\ Sequence}$ contains exclusively s1b sequences. In agreement with previous reports (63, 64), we detected five amino acid differences between subtype s1a and s1b sequences. Subtype s1c differs from subtype s1a at two amino acid positions and differs from s1b at seven positions. Type s1c sequences were detected exclusively in Asian strains and type s1b sequences were detected predominantly in Western strains.



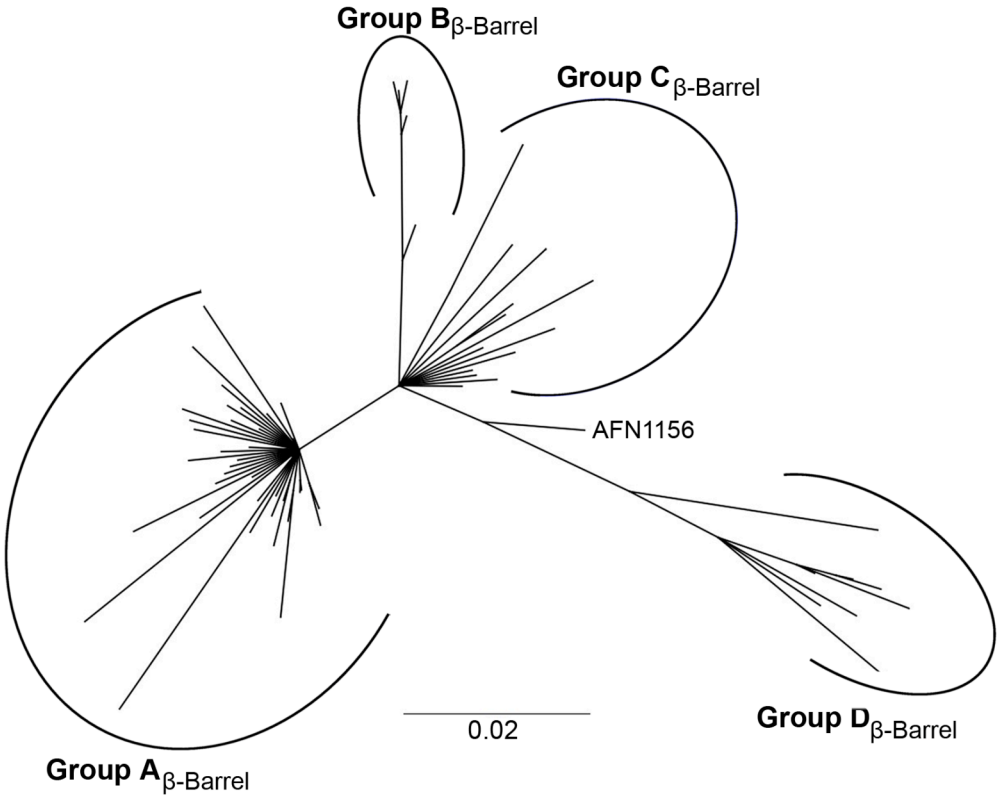
Classification	# of Strains	Origin		VacA Type					
		Western	Asian	s1	s2	m1	m2	i1	i2
Group A_{p33}	86	23	63	86	0	70	16	84	0
Group B_{p33}	14	2	12	11	3	0	14	0	14
Total	**100**	**25**	**75**	**97**	**3**	**70**	**30**	**84**	**14**

Figure 26. Neighbor-joining phylogeny of the VacA p33 domain. Two main groups are evident, designated Group A_{p33} and Group B_{p33}. The chart shows the number of strains analyzed and characteristics of VacA sequences in each group of the tree. The sequences in Group A_{p33} were localized in Groups 1, 2, 3 of the full-length VacA tree (Figure 23), and sequences in Group B_{p33} were all localized in Group 3 of the full-length VacA tree. Divergence between Group A_{p33} and Group B_{p33} reflects differences within the VacA intermediate region (161). The sequences in Group $A_{p33}$are characterized as type i1, with the exception of two sequences that appear to be i1-i2 hybrids, and sequences in Group $B_{p33}$are exclusively characterized as type i2.



Classification	# of Strains	Origin		VacA Type			
		Western	Asian	s1	s2	m1	m2
Group A_{SAP}	48	1	47	48	0	43	5
Group B_{SAP}	14	0	14	13	1	0	14
Group C_{SAP}	26	21	5	24	2	23	3
Total	**88**	**22**	**66**	**85**	**3**	**66**	**22**

Figure 27. Neighbor-joining phylogeny of the VacA secreted alpha peptide (SAP) domain. The phylogeny of the alpha peptide region has not been previously investigated, but distinct groups are detected within this tree, designated Groups A-C_{SAP}. The chart shows the number of strains analyzed and characteristics of VacA sequences in each group of the tree. The SAP phylogenetic tree contains 3 groups that reflect the geographic origin of strains (two groups are predominantly from strains of Asian origin and one group is predominantly of Western origin) (Groups A-C_{SAP}). Group A_{SAP} corresponds to a mixture of the strains in Groups 1 and 3 of full-length VacA tree (Figure 23), Group B_{SAP} corresponds to Group 3 of the full-length tree, and Group C_{SAP} corresponds to Groups 2 and 3 of the full-length tree.



Classification	# of Strains	Origin		VacA Type			
		Western	Asian	s1	s2	m1	m2
Group $A_{\beta\text{-Barrel}}$	50	1	49	50	0	12	8
Group $B_{\beta\text{-Barrel}}$	10	0	10	9	1	0	10
Group $C_{\beta\text{-Barrel}}$	17	14	3	16	1	14	3
Group $D_{\beta\text{-Barrel}}$	9	6	3	9	0	9	0
Total	**86**	**21**	**65**	**84**	**2**	**65**	**21**

Figure 28. Neighbor-joining phylogeny of the VacA C-terminal β-barrel domain. The phylogeny of the C-terminal β-barrel region has not been previously investigated, but four distinct groups are detected, designated Groups A-$D_{\beta\text{-barrel}}$. The chart shows the number of strains analyzed and characteristics of VacA sequences in each group of the tree. The β-barrel phylogenetic tree contains 4 groups that reflect the geographic origin of strains (two groups are predominantly from strains of Western origin and two groups are predominantly of Asian origin) (Groups A-$D_{\beta\text{-barrel}}$). Group $A_{\beta\text{-barrel}}$ corresponds to a mixture of the strains in Groups 2 and 3 of full-length VacA (Figure 23), Group B $_{\beta\text{-barrel}}$ corresponds to Group 3 of the full-length tree, Group C $_{\beta\text{-barrel}}$ consists of a mixture of strains from Groups 1 and 3 of the full-length tree, and Group D $_{\beta\text{-barrel}}$ consists of mainly strains from Group 1 of the full-length VacA tree.

Sliding window and McDonald-Kreitman tests

A relatively high level of divergence within the p55 VacA cell-binding domain compared to other domains may reflect relaxed constraint on that portion of the sequence or positive selection if amino acid replacements confer a selective advantage. In the latter case, we expect to observe an accumulation of nonsynonymous changes at a rate higher than that of synonymous changes (dS). To investigate the evolutionary pressures acting on VacA, we first analyzed *vacA* sequences for positive selection (dN/dS > 1) using a sliding window analysis with full-length *vacA* sequences from strains 60190 (type m1 Western, Group 1) and 95-54 (type m2, Group 3). The crystal structure of the p55 domain is available for VacA from strain 60190 (148), and VacA from strain 95-54 is known to exhibit a different cell-type specificity compared to VacA from strain 60190 (68). dN /dS ratios greater than one were observed in mainly one portion of the *vacA* sequence - the p55 cell-binding domain (Figure 29).

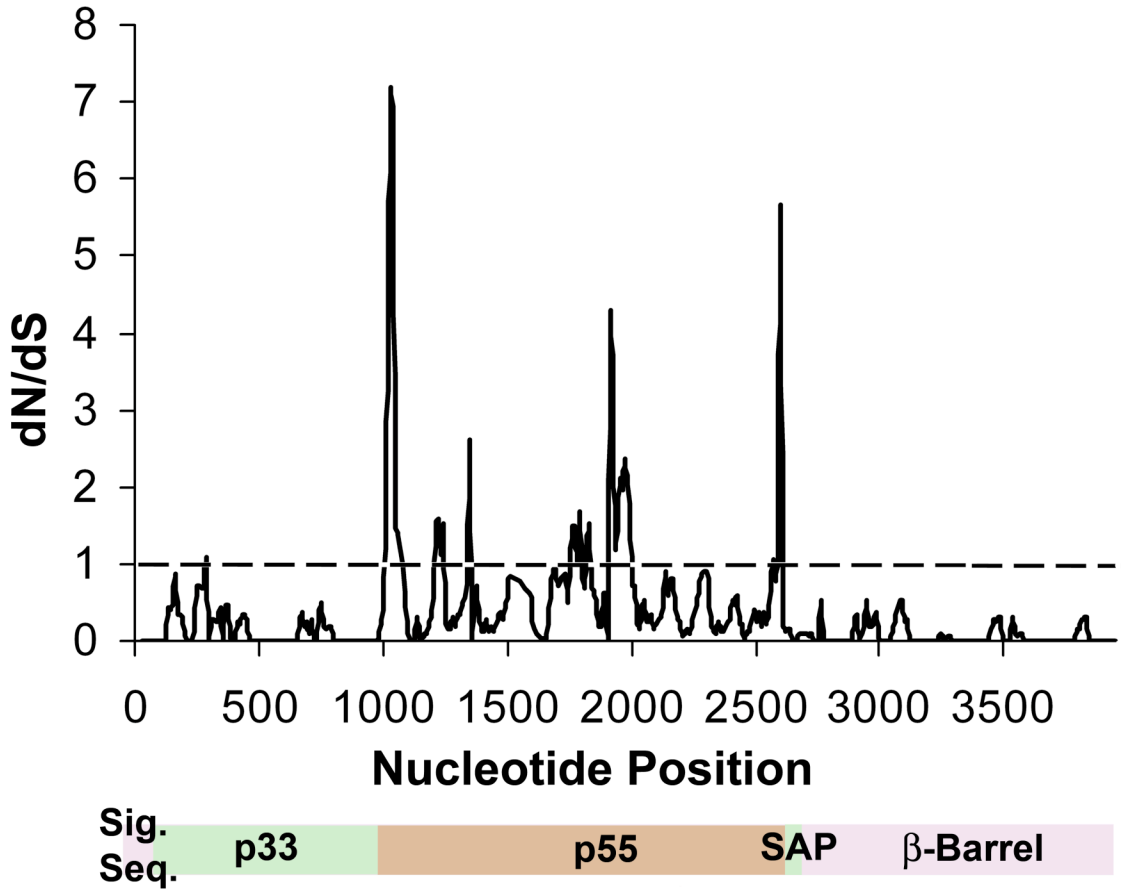


Figure 29. Sliding window analysis of VacA from *H. pylori* strains 60190 and 95-54. VacA sequences from strains 60190 (Western m1 type) and 95-54 (Western m2 type) were aligned and dN/dS ratios were calculated using DnaSP with a sliding window of 50 bases and a 10-bp step size. dN/dS > 1 indicates positive selection.

To follow up the initial observation of elevated dN/dS ratios in the portion of *vacA* encoding the p55 cell-binding domain, we collected full-length DNA sequences of 15 *vacA* alleles from each of the three main groups (Groups 1-3, Figure 22). We used the McDonald-Kreitman test (MKT) (172) to investigate if adaptive evolution in the p55 domain is driving the divergence of the three groups. The MKT analyzes the neutral theory prediction that the ratio of synonymous-to-nonsynonymous polymorphism (Ps/Pn) within groups should be the same as the ratio of synonymous-to-nonsynonymous divergence (Ds/Dn) between groups, and was used previously to detect positive selection in an *H. pylori sel1* homolog (180). The results indicate a significant deviation from neutrality when analyzing full-length *vacA* sequences and the p55 domain (Table 3) ($p < 0.001$), but not when analyzing the p33 domain or other regions. Excess nonsynonymous fixation, one signature of adaptive protein evolution, causes the Neutrality Index (NI) in the MKT to be less than 1. For all statistically-significant MKT comparisons, the NI was < 0.53. These results confirm the sliding window analysis and indicate that the divergence in the p55 cell-binding domain is in part due to strong positive selection.

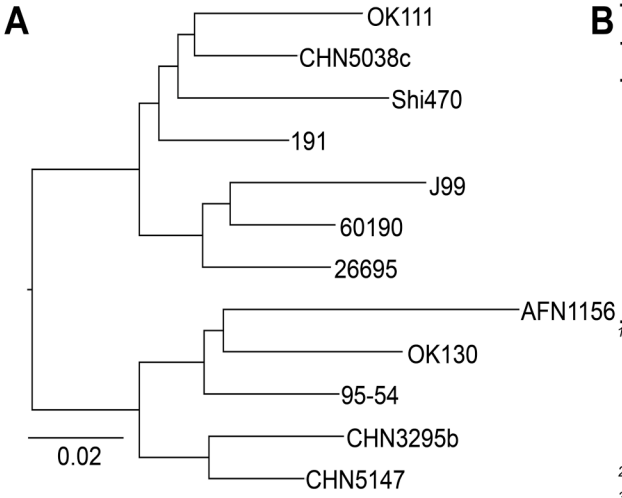
Table 3. Analysis of positive selection in *vacA* using the McDonald-Kreitman Test.

Domain of VacA	Dn	Ds	Pn	Ps	p-value	NI[1]	α-value[2]
Group 2 (m1 Asian) vs. Group 3 (m2)							
Full-length	146.06	68.03	248	286	0	0.408	0.591
Signal Sequence	0	0	17	19	N/A	N/A	N/A
p33	0	0	57	75	N/A	N/A	N/A
p55	153.49	75.68	99	92	0.001	0.53	0.469
SAP	1	0	22	22	0.322	0	1
β-Barrel	0	0	72	100	N/A	N/A	N/A
Group 1 (m1 Western) vs. Group 3 (m2)							
Full-length	132.02	63.09	353	393	0	0.429	0.57
Signal Sequence	0	0	21	19	N/A	N/A	N/A
p33	3	2.01	78	104	0.446	0.501	0.498
p55	130.93	61.1	143	140	0	0.476	0.523
SAP	3.02	3.07	33	29	0.862	1.157	-0.157
β-Barrel	1	1	78	101	0.856	0.774	0.225
Group 2 (m1 Asian) vs. Group 1 (m1 Western)							
Full-length	46.51	31.81	194	314	0	0.422	0.577
Signal Sequence	0	0	12	16	N/A	N/A	N/A
p33	4.01	5.08	31	75	0.348	0.523	0.476
p55	32.64	17.62	70	110	0.001	0.343	0.656
SAP	6.08	2.03	27	29	0.154	0.311	0.688
β-Barrel	4.01	7.18	54	84	0.828	1.15	-0.15

[1] The Neutrality Index (NI) was calculated from the ratio NI = (Pn/Ps)/(Dn/Ds), where P = polymorphic within the population, D = divergence or fixed difference between populations, n = nonsynonmous, and s = synonymous.
[2] α-value is the proportion of adaptive substitutions that ranges from -∞ to 1 and is estimated as 1-NI.

Codons under selection in VacA

We next used the program PAML to detect specific amino acids experiencing positive selection. We aligned and analyzed the p55 region-encoding sequences of 12 *vacA* genes that encompass the genetic diversity in the unrooted phylogenetic analysis (Figure 30A). Models that allow codons to evolve under positive selection (M8 and M2a) fit the dataset significantly better than do models that do not permit positive selection (M7 and M1a) ($p < 0.0001$, Figure 30B). The M1a vs. M2a analysis identified 14 amino acids that were undergoing positive selection and seven were significantly selected (posterior probability > 0.85). All seven were surface-exposed in the p55 domain crystal structure (Figure 30C) and localized to regions identified by the sliding window analysis (Figure 29), signifying that adaptive evolution is potentially due to selection for receptor-binding determinants that mediate VacA interactions with host cells or immune escape.



B PAML analyses of *vacA* p55 domain

Analysis[1]	-2(ln λ)	DF	p-Value	Positively Selected Sites
M1a vs. M2a	147.8	2	p < 0.0001	334T, 360D, 362I, 430L, 443V[2], 524V[3], 650K, 663M[2], 667D[3], 708T[3], 719E[3], 743T[2], 817F, 819S
M7 vs. M8	162.62	2	p < 0.0001	334T, 360D, 362I, 430L, 443V[2], 524V[3], 650K[2], 663M[3], 7O4G, 708T[3], 712G, 719E[3], 743T[3], 799E, 817F, 819S

[1]Codons that were subject to positive selection were identified using Nsites models (M1a, M7) that do not permit positive selection compared to models (M2a, M8) that do permit positive selection. The strength of positive selection was calculated by comparing twice the log likelihood (λ) difference in a chi-square test with two degrees of freedom (DF).
[2]Under positive selection with a posterior probability of > 0.95
[3]Under positive selection with a posterior probability of > 0.85

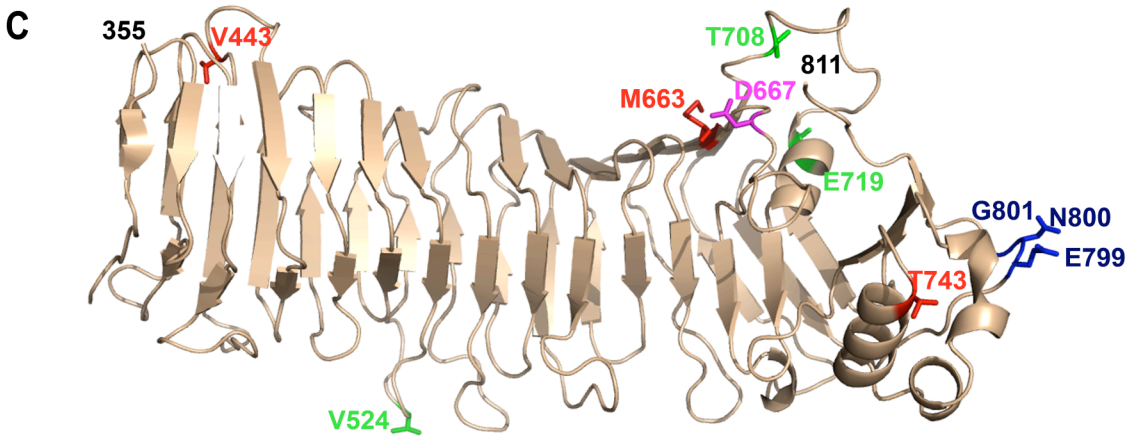


Figure 30. Positive selection within the *vacA* p55 domain identified by PAML and omegaMap. (A) Neighbor-joining tree of VacA sequences selected for the PAML analysis. (B) Summary of amino acid sites in the p55 domain that are under positive selection, based on PAML analysis. Results obtained using two models (M1a vs. M2a and M7 vs. M8) are shown. (C) Three-dimensional structure of the p55 domain from *H. pylori* strain 60190, with positively-selected amino acids identified in the former PAML analysis, mapped in color (posterior probability > 0.85, green or > 0.95, red). The p55 domain crystal structure comprises residues 355 to 811. Residue 667 is labeled in magenta and is the amino acid that both PAML and omegaMap determined to be under significant positive selection (posterior probability > 0.95). Residues E799, N800, G801 are labeled in blue and are the additional residues in the type m2 VacA sequences (Group 3) that omegaMap determined to be under significant positive selection.

Selection in the presence of recombination

The processes of recombination and natural selection can impact genetic variation in similar ways and create artifacts for interpreting results. In particular, hotspots of recombination can inflate dN/dS ratios to produce a false positive result of adaptive evolution (181, 182). To control for this potential artifact, we implemented the program omegaMap on the 45 *vacA* alleles assembled for the MKT to evaluate rates of recombination and dN/dS ratios across the p55-encoding region of the *vacA* gene. The distribution of recombination and selection confirms that different regions of the protein experience different evolutionary pressures. In particular, the average population rate of recombination (rho) per codon within the p55 domain is 0.48 ± 0.01 (mean ± standard error), which is significantly lower than the average rho for all codons outside of the p55 domain, 0.84 ± 0.01 (Mann-Whitney U test, $p < 0.0001$). Lower recombination rates in the p55 domain are consistent with our finding that p55 harbors the amino acid differences that structure the divergence of the full-length VacA sequences into three distinct groups. Consistent with the results of the sliding window, McDonald-Kreitman, and PAML analyses, the omegaMap analysis also identifies positively-selected sites with a dN/dS ratio > 1 and a posterior probability > 0.95. Residue D667 was found to be under strong, positive selection (dN/dS = 5.3, posterior probability = 1.0) and residue M663 was marginally nonsignificant (dN/dS = 1.6, posterior probability = 0.78). Notably, these positively-selected sites were recapitulated among the sites identified in the PAML analysis (Figure 30). These surface-exposed residues are adjacent to each other in the VacA p55 domain crystal structure (Figure 30C, magenta and red). Finally, we analyzed sites undergoing positive selection within each group (N = 15 alleles each) and found three additional residues (E799, N800, G801) in sequences from Group 3 (type m2 VacA) that are positively selected (dN/dS = 2.4, posterior probability = 0.95, 0.96, 0.95,

respectively) (Figure 30C, blue) and one adjacent residue, G802, that was marginally nonsignificant (dN/dS = 1.5, posterior probability = 0.74).

The β-helical structure of VacA presents a unique opportunity for studying the pressures of positive and negative selection in a structural context, since the residues are in an alternating pattern of facing inside and outside the β-barrel. All residues identified to be under positive selection are surface exposed. Initially, since the selection was in the p55 region, we speculated that the surface-exposed amino acids were important for receptor binding. Since we are able to obtain purified p33 (Chapter III), and it is functional when mixed with p55, we had a recombinant system to test the effect of mutations. Two adjacent residues (M663 and D667) identified to be under positive selection, were chosen for site-directed mutagenesis, resulting in two double mutants; M663A/D667A or M663A/D667K. The mutants were tested in three cell culture assays (methods described in Chapter III, cell culture assays): vacuolation, IL-2 secretion and cell apoptosis (Figure 31A-C). In each assay, the mutant p55 mixed with p33 behaved similarly to wildtype.

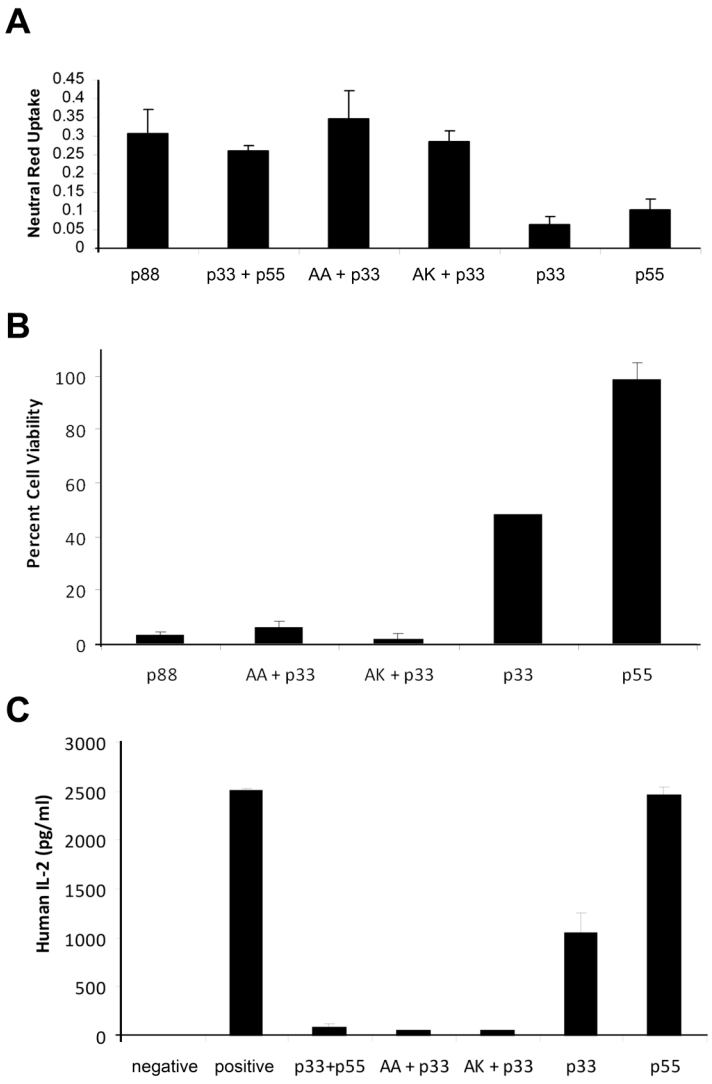


Figure 31. Effects of wildtype p33 and wildtype and mutant p55 VacA proteins on HeLa, AZ-521 and Jurkat cells. p33 was refolded as described in Methods. Wildtype and mutant purified p55 and wildtype p33 (each 1 mg/ml) were mixed together in a 1:1 mass ratio, which ensured an excess of p33 on a molar basis. The p88 VacA protein purified from *H. pylori* culture supernatant was acid-activated prior to contact with cells (106, 110), whereas the p33 or p55 preparations were not acid-activated. M663A /D667A p55 (AA) mutant, and M663A/D667K p55 (AK) mutant were tested in the activity assays. (A) HeLa cells were incubated with the indicated purified VacA proteins at a final concentration of 10 μg/ml (or 5 μg/ml of each protein in the case of p33/p55 mixture). Cell vacuolation was quantified by neutral red uptake assay (OD 540 nm). (B) AZ-521 cells were incubated with a p33/p55 mixture at the indicated final concentrations (20 μg/ml corresponds to 10 μg/ml p33 and 10 μg/ml p55). Cell viability was quantified by CellTiter-Glo. (C) Jurkat cells were incubated with the indicated purified VacA proteins at a concentration of 6 μg/ml (or 3 μg/ml of each protein in the case of the p33/p55 mixture), and IL-2 secretion was measured as described in Materials and Methods. Results represent mean ± standard deviation, based on analysis of triplicate samples.

We hypothesized that these residues may be important for secretion, since they were not important in the *in vitro* cell culture assays. The M663A/D667A mutant was introduced, via homologous recombination, into VacA from *H. pylori*. There was no effect on the secretion of the mutant VacA compared to wildtype (Figure 32).

Since the functional tests were negative, we hypothesize that this selection in the p55 region may be important for antibody evasion or maybe a consequence of immune selective pressure. Previous studies have shown that anti-VacA antibody responses are primarily against the p55 domain rather than the p33 domain, and these antibodies can neutralize VacA activity. Therefore, positive selection within the p55 domain may have been driven by evasion of host immune responses. *H. pylori* is known to live and colonize the gastric mucosa layer and can persist for a long period of time. Immune evasion may be an important feature for the successful colonization and persistence of *H. pylori*.

Figure 32. Immunoblot showing secretion of VacA p88 and M663A/D667A VacA p88 from *H. pylori*. Lanes 1 and 2 are bacterial pellet and supernatant, respectively, from wildtype VacA. Lanes 3 and 4 are bacterial pellet and supernatant from M663A /D667A mutant VacA p88.

Rooted phylogenetic analyses of VacA and housekeeping genes

The three-group phylogenetic structure of VacA (Figure 23) has been shaped by positive selection associated with distinct geographic distributions. *H. pylori* housekeeping genes also exhibit geographic diversity (158, 159), but presumably there has been very little positive selection upon housekeeping genes because the genes used for these analyses are conserved and typically under purifying selection. Therefore, we hypothesized that there would be marked differences in the evolutionary history and phylogeography of *vacA* compared to housekeeping genes.

To generate a rooted tree of housekeeping gene sequences, we selected a total of 61 sequences from representative *H. pylori* strains that previously had been classified into nine geographically distinct populations and subpopulations based on MLST analysis (158), as well as strains that were used in the analysis of VacA. We used the corresponding housekeeping genes from the close relative *H. acinonychis* to root the tree (176). The Bayesian root of this tree is confidently positioned in taxa classified by MLST analysis as HpAfrica2, a population currently found almost exclusively in South Africa (158, 159) (Figure 33), and the next most closely related taxa are also of African origin (classified as HpSouth Africa or HpWest Africa subpopulations). To confirm the rooting position in African populations, we excluded the HpAfrica2 taxa, repeated the analysis, and again observed the rooting position in the set of South African taxa. Rooting of the MLST tree within African taxa is consistent with previous reports of an ancient African origin for *H. pylori* in humans (158, 159). The phylogenetic analysis suggests a spread of *H. pylori* from Africa to Europe, and then from Europe to East Asia and the American Hemisphere**.** Previous analyses of genetic diversity in *H. pylori* indicated that the global spread of *H. pylori* mirrored that of its human hosts (159). The current analysis confirms the hypothesis that *H. pylori,* indeed, has African origins. If we crudely extrapolate human migration patterns from the *H. pylori* phylogeny as has been

done in other population genetic studies, our analysis suggests that migration from the Eastern coast of Africa led to human colonization of Europe before Asia. The basic reason is that the earliest branching taxa from the African *H. pylori* lineages are classified as HpEuropean, while the HpAsia2 populations (currently found in Northern India and Southeast Asia) branch off from the ancestors of the European taxa. Deducing the routes of human migration is complex, and will require synthesizing other evidence beyond the *H. pylori* data.

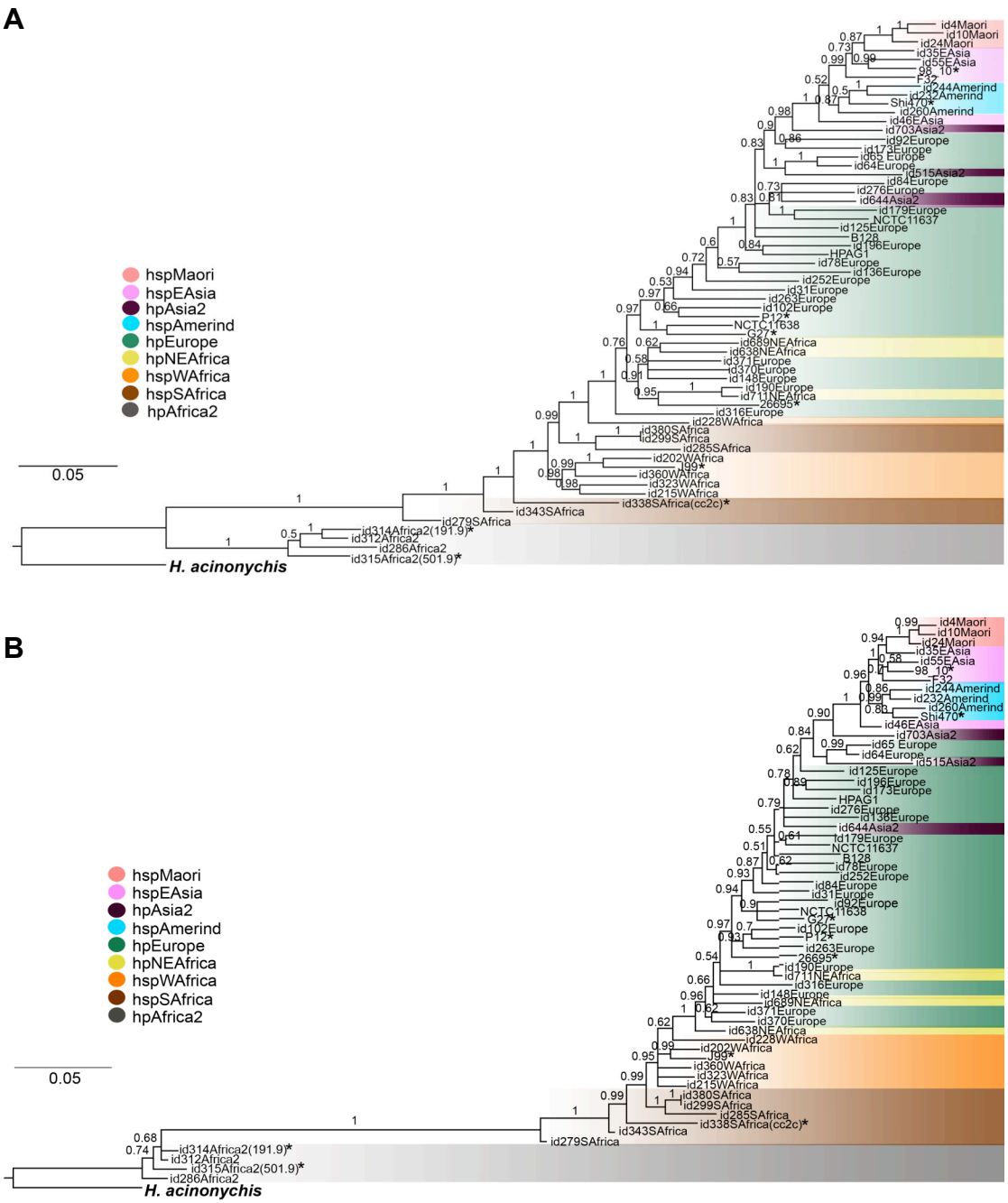


Figure 33. Rooted (A) MrBayes and (B) PhyloBayes phylogenetic trees of concatenated housekeeping gene. Nucleotide sequences of 7 housekeeping genes (*atpA, efp, mutY, ppa, trpC, urel, yphC*) from 61 strains of *H. pylori* and 1 outgroup, *H. acinonychis,* were analyzed. Sequences representing nine previously described populations or subpopulations of *H. pylori* (158, 159) are color-coded on the tree: hspMaori, hspEAsia, hpAsia2, hspAmerind, hpEurope, hspWAfrica, hpNEAfrica, hspSAfrica, and hpAfrica2. The numbers represent Bayesian posterior probability values for each node. Asterisks indicate strains that were also analyzed in Figure 34. Strains are labeled with "id numbers" that are used in an *H. pylori* multilocus sequence typing database (http://pubmlst.org/helicobacter), and the population assignments of strains based on MLST analysis (158, 159) are indicated.

We next generated a rooted tree of full-length VacA sequences (Figure 34) using the *vacA* pseudogene in *H. acinonychis* as an outgroup (176). We aligned the VacA amino acid sequence of *H. acinonychis* with 24 ingroup taxa, corresponding to *H. pylori* VacA sequences that were representative of VacA Groups 1-3 (Figure 23). To increase representation of African VacA sequences in this tree, we determined the *vacA* sequences of three additional *H. pylori* strains of African origin (including two strains classified as HpAfrica2 based on MLST analysis). The resulting 971-amino-acid sequence alignment yields Bayesian and PhyloBayesian phylogenies (Figure 34) showing that the VacA root is confidently positioned in the CHN3295 and CHN5147 taxa. These strains, isolated in Shanghai, China, (164) have m2 sequence characteristics and belong to the Group 3 subgroup II of the full-length VacA tree (Figure 23). The next earliest branching lineage includes the Group 3 subgroup I strains, which also have m2 sequence characteristics and were isolated in many different geographic locations. Three observations suggest the VacA rooting position is accurate: (i) First, the rooting position in m2 Asian strains issupported with two different inference methods, MrBayes and PhyloBayes; (ii) three different models of evolution (CAT, WAG, JTT) and removal or addition of other m2 Asian strains and hypervariable regions does not alter the rooting position; in particular, the probabilistic inference model, CAT, accounts for across-site heterogeneities and can handle model misspecifications associated with long branch attraction (171) artifacts (177); and (iii) the maximum amino acid identity of the outgroup with Asian m2 (65.4%) or non-Asian m2 sequences (64.0%) is greater than that of the outgroup with Asian m1 (57.4%) or non-Asian m1 sequences (51.1%). Thus, a comparison of the rooted trees confirms that there are substantial differences in the phylogenetic structure of VacA compared to housekeeping gene sequences. Furthermore, to statistically test the topological incongruence between the VacA and

MLST phylogenies, we compared the ML phylogenies of 14 taxa common to both datasets using the Shimodaira-Hasegawa (SH) test (Figure 35). The MLST and VacA datasets show unambiguous significant differences ($p < 0.0001$ for both, Figure 35A and B, respectively), indicating the VacA toxin gene has not coevolved with the core genome of *H. pylori*.

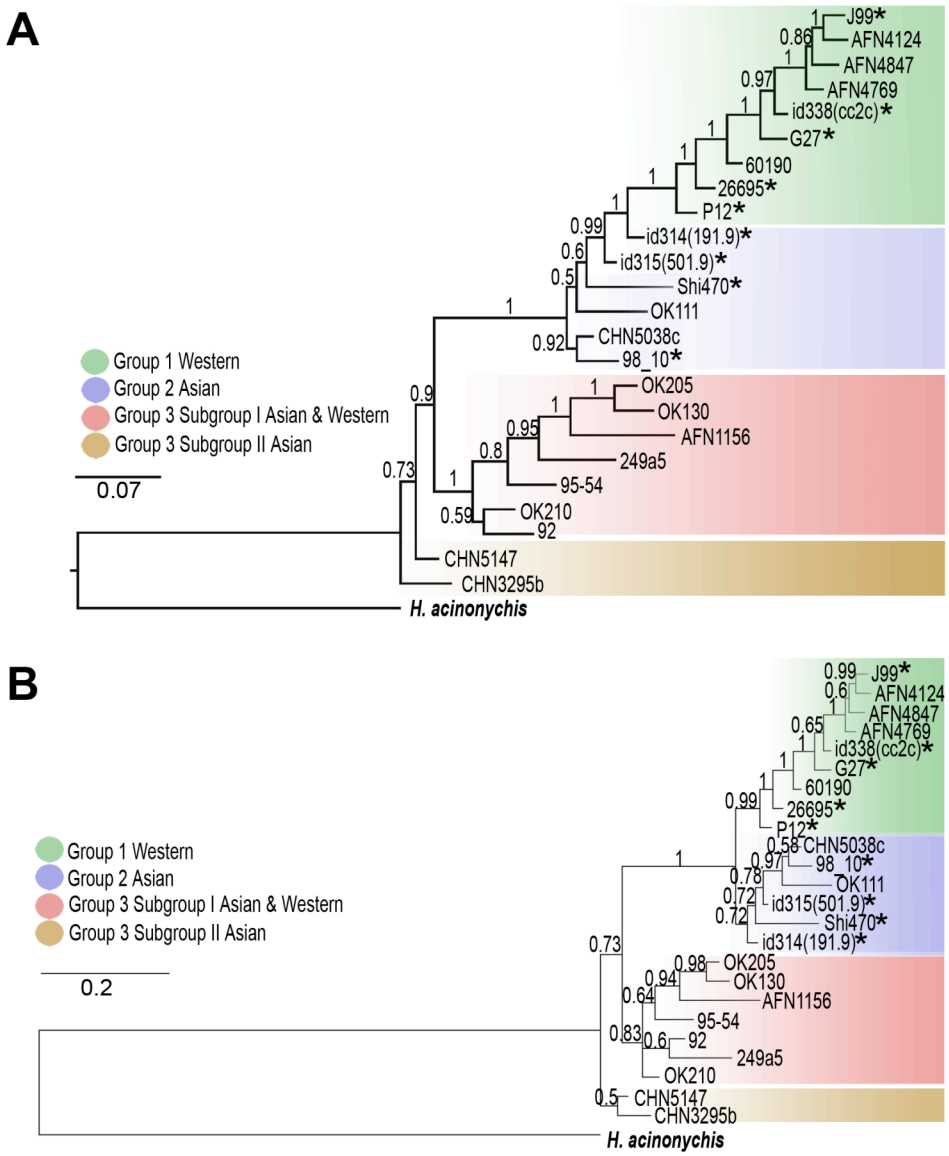


Figure 34. Rooted (A) MrBayes and (B) PhyloBayes phylogenetic trees of VacA sequences. Twenty-four representative VacA sequences from Groups 1-3 (Figure 23) and 1 outgroup, deduced from the reconstructed *H. acinonychis vacA* pseudogene were analyzed. Four major groups are indicated: m2 Asian group (corresponding to the Group 3 subgroup, brown, in Figure 23), m2 Western group (Figure 23 Group 3, red), m1 Asian group (Figures 23 Group 2, blue), and m1 Western group (Figure 23 Group 1, green). The numbers represent posterior probability values for each node. Asterisks indicate strains that were also analyzed in Figure 33.

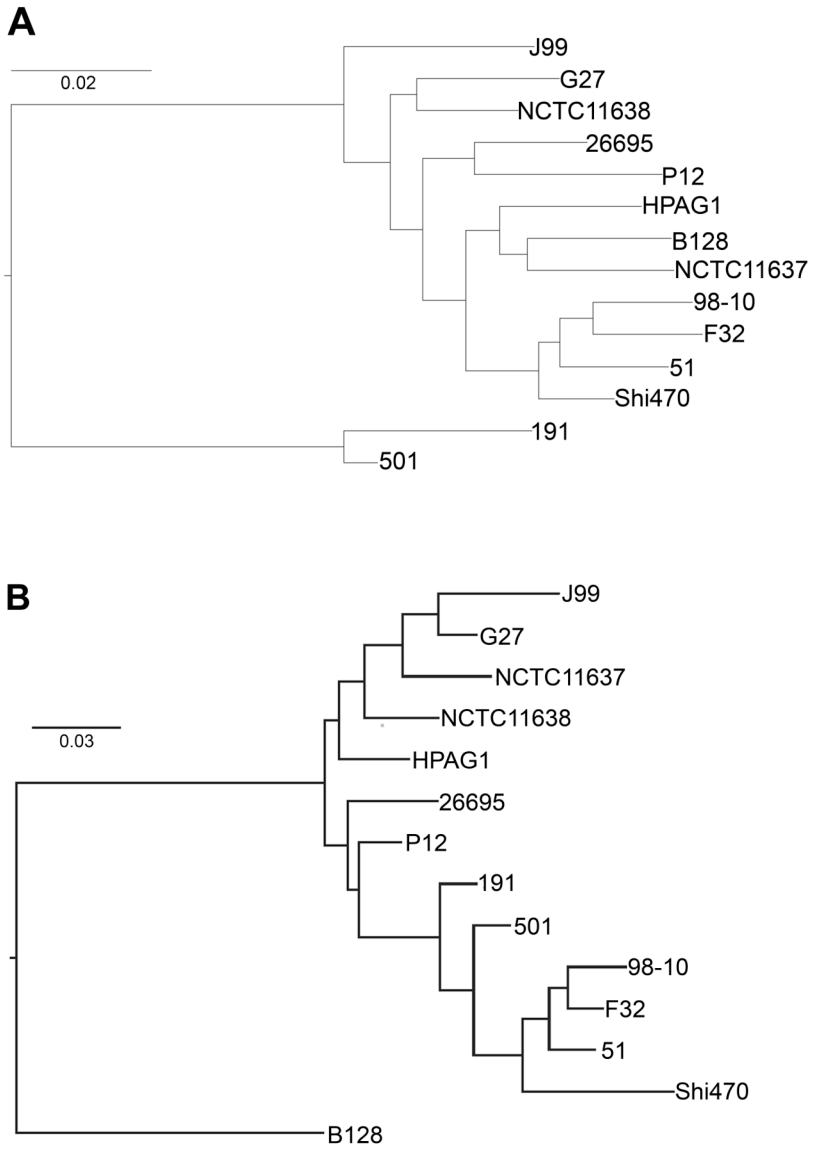
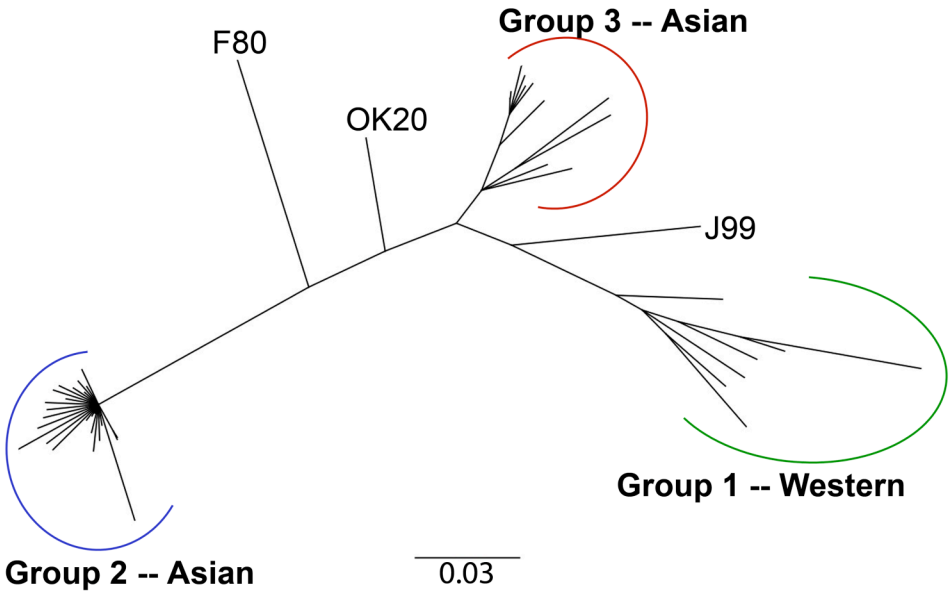


Figure 35. Topological differences in VacA and MLST phylogenetic trees using the Shimodaira-Hasegawa (SH) test. The Neighbor-joining phylogenetic trees from the Shimodaira-Hasegawa test shows the 14 taxa used in the analyses and evolutionary relationships for (A) MLST and (B) VacA. The taxa used were identical for both the MLST and VacA trees.

CagA phylogeography

In *H. pylori* strains 26695 and J99, *vacA* and *cagA* (encoding the secreted effector protein CagA) are located ~350 kb apart in the genomes. VacA can downregulate CagA's effects on epithelial cells and CagA can protect cells against the apoptotic effects of VacA (166, 168). Furthermore, VacA can counteract the ability of CagA to activate NFAT in gastric epithelial cells (167). We thus hypothesize that these two genes share an evolutionary history characterized by co- or counter-adaptations in response to a common selective pressure. We identified 46 *H. pylori* strains for which both VacA and CagA sequences were available. Phylogenetic analysis of the full-length CagA sequences revealed three groups (Figure 36), similar to the phylogeny of VacA (Figure 23). In the CagA analysis, Group 1 consists of seven sequences that are predominantly from strains of Western origin (Figure 36). The corresponding VacA sequences from these strains are characterized as m1 Western and are found in Group 1 in the full-length VacA tree (Figure 23). CagA Group 2 consists of 25 exclusively Asian sequences; the corresponding VacA sequences are characterized as m1 Asian and are found in Group 2 in the full-length VacA tree. Finally, CagA Group 3 consists of 11 exclusively Asian sequences; the corresponding VacA sequences are characterized as m2 and are found in Group 3 in the full-length VacA tree. Thus, the CagA phylogenetic tree strikingly resembles the full-length VacA tree and suggests that CagA and VacA are coevolving due to a similar selective pressure. A McDonald-Kreitman test indicates that positive selection has significantly shaped CagA divergence of Group 2 from both Group 3 ($p < 0.005$, NI = 0.58) and Group 1 ($p < 0.018$, NI = 0.66). These findings indicate that both VacA and CagA are under positive selection and suggest for the first time that these virulence determinants coevolved in concert. The resulting diversity in VacA and CagA sequences potentially modulates the activities of these

proteins and could be a factor that influences the risk for development of peptic ulceration or gastric cancer in *H. pylori*-infected persons.



Classification	# of Strains	Origin		VacA Type			
		Western	Asian	s1	s2	m1	m2
Group 1	7	5	2	7	0	7	0
Group 2	25	0	25	25	0	24	1
Group 3	11	0	11	10	1	0	11
Total	**43**	**5**	**38**	**42**	**1**	**31**	**12**

Figure 36. Analysis of CagA phylogeography. Neighbor-joining phylogenetic tree of 46 CagA amino acid sequences. Three major groups are evident: Group 1 consists predominantly of sequences from strains of Western origin, Group 2 consists of Asian sequences, and Group 3 consists of Asian sequences. The chart shows the number of strains analyzed and characteristics of VacA sequences in each group of the tree. CagA sequences shown in Groups 1 and 2 correspond to *H. pylori* strains containing type m1 VacA (Groups 1 and 2 of Figure 23), whereas CagA sequences shown in Group 3 correspond to strains containing type m2 VacA (Group 3 of Figure 23). The nomenclature for the primary CagA groups (Groups 1, 2, and 3) is consistent with the nomenclature of groups in the full-length VacA tree (Figure 23).

In summary, our key findings indicate for the first time that (i) VacA sequences are divided into three distinct groups on the basis of amino acid sequences, and different VacA domains exhibit different evolutionary histories reflecting diverse molecular evolutionary pressures. (ii) VacA has undergone strong divergence and positive selection in the p55 domain (iii) The phylogeographic features of VacA and CagA are surprisingly similar, yet markedly different from the phylogeographic features of housekeeping genes, which reflect a global spread of *H. pylori* out of Africa. This result suggests that there is a related selective pressure on both VacA and CagA. Since there is substantial physical distance between *vacA* loci and *cag* genes within the bacterial genome, the selection of a *vacA/cagA* coadapted genotype may be due to a form of pseudolinkage of functionally-interacting genes that perhaps balances proinflammatory and anti-inflammatory characteristics of strains to facilitate long-term colonization of the human gastric mucosa. The present work therefore provides the first evidence for selection-driven changes in *H. pylori* virulence determinants associated with peptic ulceration and gastric cancer.

CHAPTER V

CONCLUSIONS

Summary

H. pylori is the causative agent of most cases of peptic ulcer disease and a major risk factor for the development of gastric cancer. Experimental and epidemiological studies suggest that VacA is an important virulence factor in *H. pylori* pathogenesis and linked to severe gastric tissue damage (183). I*n vitro*, VacA induces various cytotoxic effects, while *in vivo*, the relevance of these phenomena for *H. pylori* gastric infection and alteration is mostly undefined. Structural and functional studies of VacA should provide information needed to address this knowledge gap.

Prior to undertaking this thesis project, two putative VacA domains had been described in the literature. The mature secreted 88 kDa VacA can undergo proteolysis into 33- and 55- kDa fragments termed p33 and p55, respectively. This thesis has provided insight into the importance of these domains for VacA assembly into oligomers, the structural and functional correlates of VacA sequence diversity as well as how molecular evolution has shaped VacA.

Chapter II of this thesis describes the use of x-ray crystallography to determine an atomic structure of the p55 receptor-binding domain. The structure is an elongated β-helix molecule that is typical of autotransporter proteins. VacA sequence polymorphisms were examined in the context of the structure in an effort to understand how genetic differences between m1 and m2 *vacA* alleles relate to structural and functional differences in the proteins they encode. In addition, docking of the high-resolution p55 domain in the dodecamer VacA p88 EM map and the identification of a highly conserved

N-terminus of p55, as shown by analysis of VacA sequence polymorphisms associated with the m1 and m2 forms of VacA, provide strong evidence that the p33 domain extends the β-helix and interacts with the neighboring p55 domain. This interaction plays an important role in VacA oligomerization.

Chapter III describes the expression and refolding of p33. This chapter also describes how purified p33, when added *in trans* to purified p55, reconstitutes a functional protein. Furthermore, it was demonstrated that both domains were required for VacA binding and internalization into host cells. Also, purified p33 and p55 assemble into oligomeric structures in the presence of amphipathic molecules. In the presence of detergent, single-layered forms are the predominant species. This is interesting because we hypothesize that single-layered forms represent the relevant physiological structures required for pore formation. VacA p88 from *H. pylori* must first be acid-activated in order to cause an effect on cells, whereas purified p33 and p55 do not require this pre-treatment for activity. Finally, the recombinant expression of active p33 is a major step forward in our efforts to further understand VacA structure-function relationships, since this fragment can be produced in the large amounts required for future crystallography experiments.

In Chapter IV, we analyzed the molecular evolutionary forces acting on VacA. *vacA* genes from a large set of *H. pylori* strains were analyzed in a population genetic context. VacA sequences cluster into three groups with distinct geographic distributions and adaptive evolution in VacA is restricted to the p55 domain. Since the crystal structure was determined in Chapter II, we could map positively selected sites to the structure. The specific amino acids that were identified to be under positive selection map to surface exposed residues. It is possible that the correspondence of positively selected VacA sites to surface-exposed residues of p55 is important for immune evasion or may have arisen in response to immune selective pressures. In this chapter, we

confirmed, by a rooted phylogenetic analysis of *H. pylori* housekeeping genes, the theory that *H. pylori* originated in Africa and spread throughout the world along with the major human migration events. Rooted phylogenetic analysis of the origins of VacA reveals that the ancestry of VacA is strikingly different from the African origins that typify the core genome. Finally, the virulence determinant CagA is also under positive selection and has a three-group phylogeny that is strikingly similar to that of VacA, suggesting that the natural selection of VacA and CagA have been coupled.

Future Directions

We are interested in further dissecting the role of VacA in *H. pylori* colonization, persistence, and pathogenesis. Specifically, we are interested in understanding the relevant receptors for inflammatory and immunomodulatory responses, the molecular details of how VacA binds those receptors, the cellular events required for oligomerization and pore-formation, and the VacA pore structure. Moreover, we would like to understand if vacuolation is relevant to the function of VacA or a 'by-product'. Additionally, there are numerous papers that suggest VacA inserts into the mitochondria (90, 184-188), but the trafficking of VacA is not clear. Finally, an animal model that more accurately reflects the complex balance between inflammation and persistence is needed to further address the effect of VacA *in vivo*. We plan to address VacA receptor binding, oligomerization, and pore formation in the following future experiments. These future studies are expected to lead to important new insights into the role that VacA plays in bacterial colonization of the host and evasion of host immune defenses.

Identify the region of VacA important for binding carbohydrate

A number of VacA glycoprotein receptors have been reported in the literature. We were unable to detect specific, saturable binding to gastric epithelial cells using flow cytometry and radiolabeling of the ligand (Appendix). We were particularly interested in pursuing structural studies on the VacA αLβ2 integrin complex but were unable to detect an interaction using biochemical methods (data not shown). Since we used a CHO3.2.8.1 cell line to express the integrin ectodomain, we hypothesized that the failure to see binding was a result of improper integrin glycosylation. To better understand the glycan binding profile of VacA, we sent VacA p55 and the full-length VacA p88 to a glycan array-screening core, the Consortium for Functional Glycomics. The strongest interactions, binding to sialylated 2-3 structures, were observed for p88, but not p55 (Figure 37). Additionally, when p88 was acid activated, binding to 2-3 sialyated structures was still observed (Figure 37B). Only one sugar (GlcAb1-3GlcNAcb-Sp8) was found to bind the p55 region of VacA (Figure 37C). This suggests that the binding to carbohydrates stems from the p33 region of VacA and oligomerization is not required (Figure 37).

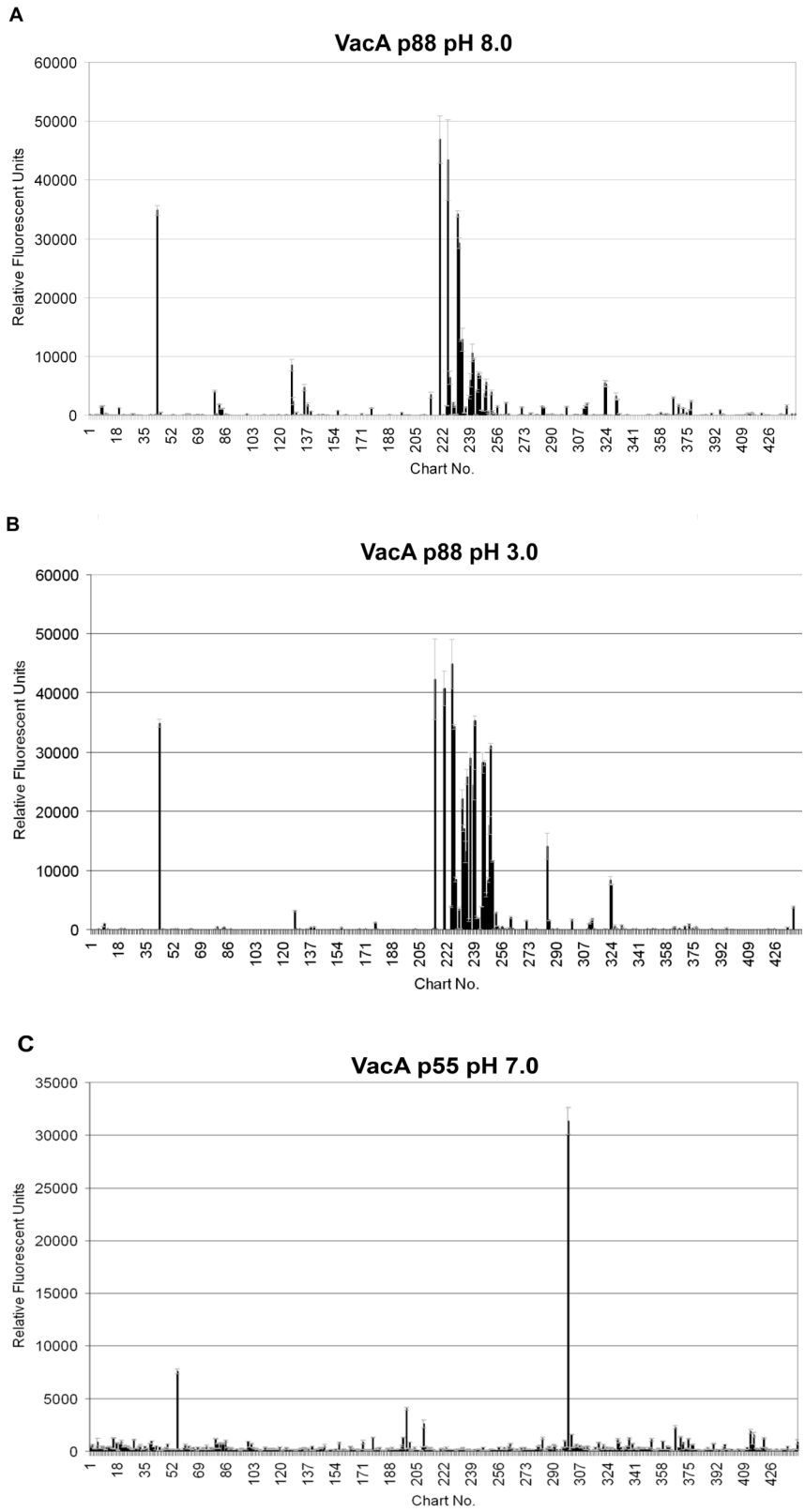


Figure 37. VacA p88 interaction with the glycan array. (A) VacA p88 binding at pH 8.0 (B) VacA p88 binding at pH 3.0 (C) VacA p55 binding at pH 7.0

One goal for future studies is to verify and visualize this interaction to 2-3 sialic acid, identify the region of VacA important for this interaction, and generate and test VacA mutants in cell-based assays. We will purify oligomeric VacA p88, add 2-3 sialylated sugar bound to streptavidin and detect binding by adding a nano-gold molecule linked to biotin. We can visualize the gold molecule by EM and identify the region of VacA p88 important for this interaction. We will generate mutations within this region and test these mutants for binding, pore formation, and signaling in gastric epithelial cells.

Determine a high-resolution crystal structure of VacA p88 or VacA p33/p55

The structure of VacA p55 (Chapter II) allowed us to define the structural features of VacA that may contribute to its cell-binding and oligomerization properties as well as the structural correlates of VacA sequence variation. However, the structural picture of VacA is not complete, because an atomic view of the p33 domain does not exist. We hypothesize that a high-resolution crystal structure of p33 will provide further insight into how VacA assembles into higher-ordered oligomers and allow us to begin to dissect how VacA forms pores. Additionally, we will be able to test our hypothesis that p33 adopts and extends the β-helix fold observed in p55.

VacA p88 forms a heterogeneous mixture of water-soluble oligomeric structures and is therefore not ideal for crystallization studies. With the advent of the p33 expression system described in Chapter III, we are able to obtain large quantities of refolded p33. We have screened over 480 buffer conditions in attempts to crystallize p33 but the protein precipitates in the majority of the conditions and high levels of guanidine and arginine are required in the buffer. We mixed p33 and p55 together in an attempt to reduce the amounts of guanidine and arginine. However, even though it formed monomeric p88 (Chapter III), the guanidine and arginine could not be reduced.

We further demonstrated that p33/p55, in the presence of detergent and absence of guanidine, formed oligomers. We hypothesized that these oligomers were all single-layered and may form a more homogeneous sample of oligomers. However, even though the majority were single-layered (Chapter III, Figure 21), they were still composed of hexamers and heptamers and double-layered forms were still present.

To circumvent these obstacles, we have constructed an oligomerization deficient mutant. It has been shown that residues 346-347, which are located in the p55 domain, are required for oligomerization (121). A deletion mutant was made in VacA p88 from *H. pylori* using a homologous recombination system. The expression level of delta 346-347 VacA p88 was decreased significantly and there is currently no purification system available for monomeric p88. However, we can take advantage of our recombinant expression system for p33 and p55 proteins. Amino acids 346-347 were deleted in the recombinant p55 and then mixed with wildtype, recombinant p33. This forms monomeric p88 and does not oligomerize. I setup crystal trays with this protein at 6 mg/ml, and 35% of the wells remained soluble. This is a promising result, and future studies will include screening more buffer conditions, using a higher protein concentration, and setting up crystal trays at different temperatures.

Amino acids 49-57 in the p33 domain are also required for oligomerization. These amino acids were deleted in the recombinant p33 and mixed with wildtype, recombinant p55. Again, this forms monomeric p88 as judged by gel filtration chromatography. The oligomerization deficient mutants provide an attractive system for future crystallography studies. Additionally, we can use the oligomerization deficient mutants to further dissect channel-dependent and independent activities and signaling functionality on cells. We know these mutants are not functional in the vacuolation assay since oligomerization is important for vacuolation. We do not know, however, if an oligomerization deficient mutant will be functional in the Git1 or p38-signaling assay.

VacA p55 can activate Git1; however, p55 forms dimers, which may be sufficient for activation (87). VacA oligomers may crosslink the receptors, resulting in the induction of signaling or internalization pathways.

An alternate path to obtain the crystal structure of p88 is to recombinantly express p88. We have tried to express p88 recombinantly using *E.coli*; however, there is low expression and there is breakdown of p88 into p33 and p55. *Bacillus megaterium* has been gaining more popularity as an efficient alternative to *E.coli*. *B. megaterium* as an expression host relies on the xylose operon as a regulatory element and has the advantage that none of the alkaline proteases are present. This enables expression of foreign proteins without degradation. Moreover, no endotoxins are found in the cell wall and protein yields are exceptionally good. I have recently cloned p88 into the *B. megaterium* vector. If the protein expresses, the first step will be to see if the protein oligomerizes by gel filtration and EM. If so, we can easily construct an oligomerization deficient mutant in this recombinant system, and use this protein for crystallography studies. Finally, another alternate approach would be to express oligomerization-deficient p88 to inclusion bodies and refold the protein. This approach was used for the expression and purification of p33 in Chapter III.

Determine the structure of VacA in a pore state

Following receptor binding and oligomerization, VacA is thought to insert into the membrane and is then internalized which leads to the formation of vacuoles. Expression of VacA in transfected cells results in cell vacuolation, which suggests that VacA has an intracellular site of action. This vacuolating activity is hypothesized to be the consequence of anion-selective channel formation in late endosomal compartments (61, 78, 79, 81, 189)

A 2D crystal structure of VacA on lipid bilayers would reveal the structural features important for pore formation. Atomic force microscopy has shown that below pH 5, VacA associates with anionic lipid bilayers to form hexameric membrane-associated complexes (61) and forms pores across planar lipid bilayers.

Czajkowsky *et al* have produced VacA 2D crystals on anionic phospholipid bilayers. The VacA crystals reveal a hexagonal pattern of rings and possess the same general characteristics as observed both with the water-soluble dodecamer and with the isolated membrane-associated oligomers observed at pH < 5. Height measurements suggest that the oligomer is single-layered. Furthermore, pore formation presumably requires that VacA must insert and span the entire bilayer. The VacA hexamer is only about 4.8 nm and protrudes only 2.9 nm from the bilayer, suggesting that a large conformational change is required during insertion. We plan to recreate these 2D crystals by using the methods described previously (61). Electron crystallography will allow for the reconstruction of a three-dimensional model of the VacA pore state.

Currently, the highest resolution image of a VacA oligomer is a 19 A EM map of the VacA dodecamer (107). In collaboration with the Ohi lab, we will use cryo-electron microscopy to determine a high-resolution structure of a VacA oligomer. Progress has been hampered because VacA from *H. pylori* forms a heterogeneous mixture of oligomers. To help improve homogeneity, we can reconstitute the VacA oligomers by mixing the purified domains (p33 and p55) with detergent (Chapter III). By using this process, we obtain mostly single-layered species (Chapter III, Figure 21), which will improve the number of EM particles and improve particle collection. By comparing this structure to the 3D structure obtained by 2D Electron crystallography, we can begin to elucidate how this toxin inserts into the membrane and forms pores.

Concluding Remarks

The future directions I have outlined rely heavily on the tools of structural biology, methods I have come to appreciate over the course of my graduate work. The pictures these methods provide should offer the molecular detail needed to dissect the physiological roles of VacA in *H. pylori* function.

APPENDIX

FORAYS INTO RECEPTOR BINDING

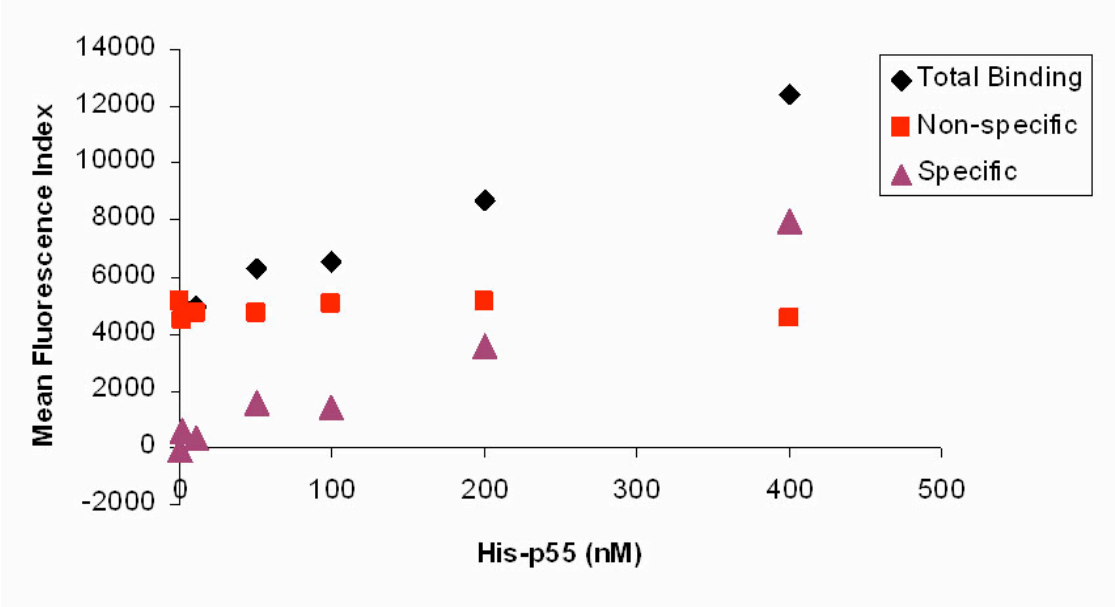
Introduction

The molecular basis for the role of VacA in pathogenesis has been complicated by the fact that it binds multiple receptors and can trigger a variety of cellular effects. Several lines of evidence indicate that the initial interaction of VacA with its target cell is mediated by interaction with receptor(s), which trigger receptor-mediated signaling events or the formation of membrane channels (76, 91, 190-192). VacA may contain multiple receptor-binding sites, each of which is optimized for binding to a different receptor (128). Moreover, there appear to be significant functional differences depending on the type of *vacA* allele present, m1 or m2. Type m1 and m2 forms of VacA can differ in their cell-type specificity, implying the roles of discrete receptors and/or sequence polymorphisms in the receptors from different human populations. Furthermore, m1 strains of *H. pylori* are associated at higher frequency with gastric cancer (87-89). The underlying process by which VacA interacts with multiple components of the plasma membrane is poorly understood.

Identification of the VacA receptor binding sites can be accomplished with site-directed mutagenesis and an assay in which the receptor-binding affinities of different VacA mutants can be compared. Past experiments, using *H. pylori* purified VacA, have suggested that VacA binds either non-specifically or to abundant, low-affinity receptors on HeLa cells (191). With the advent of a GST-fusion and flow cytometry assay, Wang and colleagues have shown specificity and saturability in binding and that the GST-VacA fusion has a binding activity similar to that of the native toxin (119, 128). The use of the

N-terminal GST-fusion protein inhibits VacA oligomerization. Kinetic and Scatchard analysis revealed a specific high-affinity receptor interaction with a measured affinity of K_D ~ 5 nm (119, 128). This value, however, may need to be reconsidered because of the capacity of GST to dimerize.

A specific, saturable interaction has never been demonstrated with the p55 domain alone, even though several studies have implicated the importance of this domain in binding to host cells (68, 117-119, 128). We wanted to develop a sensitive assay to detect subtle differences in binding between m1 and m2 forms of purified VacA p55 domain in an effort to identify specific amino acids important for their interactions with host cells. We developed a flow cytometry binding assay where we incubated trypsinized cells (10^5) with His-p55 VacA. Cells were incubated with anti-His antibody, and incubated with goat anti-rabbit phycoerythrin. Flow cytometric analyses were carried out at the Vanderbilt Flow Cytometry Core Laboratory. For competition experiments, cells were incubated with a 100-fold molar excess of the cleaved (His removed) wildtype VacA p55. With this assay, we were able to show specific binding, however, the binding was not saturable and we could not calculate a binding affinity (Figure 38). In an effort to gain sensitivity, we developed a radioligand binding assay, where we added ^{125}I -p55 (hot) to AGS cells in the absence or presence of 100 fold excess of unlabeled p55 (cold) (Figure 39). Again, binding was not saturated. A possible explanation is that VacA binds to multiple cell surface components on gastric epithelial cells due to its “sticky” nature. Finally, evidence for a specific receptor is lacking and needs to be reevaluated.



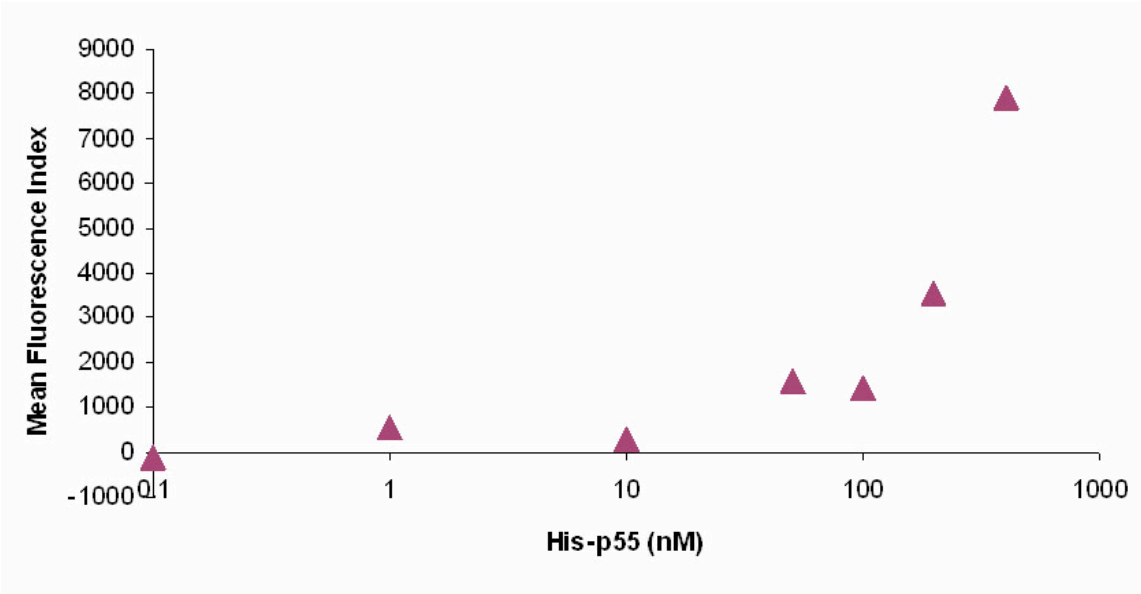


Figure 38. Analysis of VacA p55 binding to Human Adenocarcinoma Gastric Epithelial Cells (AGS) by flow cytometry. AGS cells were incubated in medium, 2% fetal bovine serum with His-p55 alone or plus a 100-fold molar excess of p55 (His-tag removed) at various concentrations. Specific binding is the difference between total binding and nonspecific binding. The lower panel represents a log scale of specific binding.

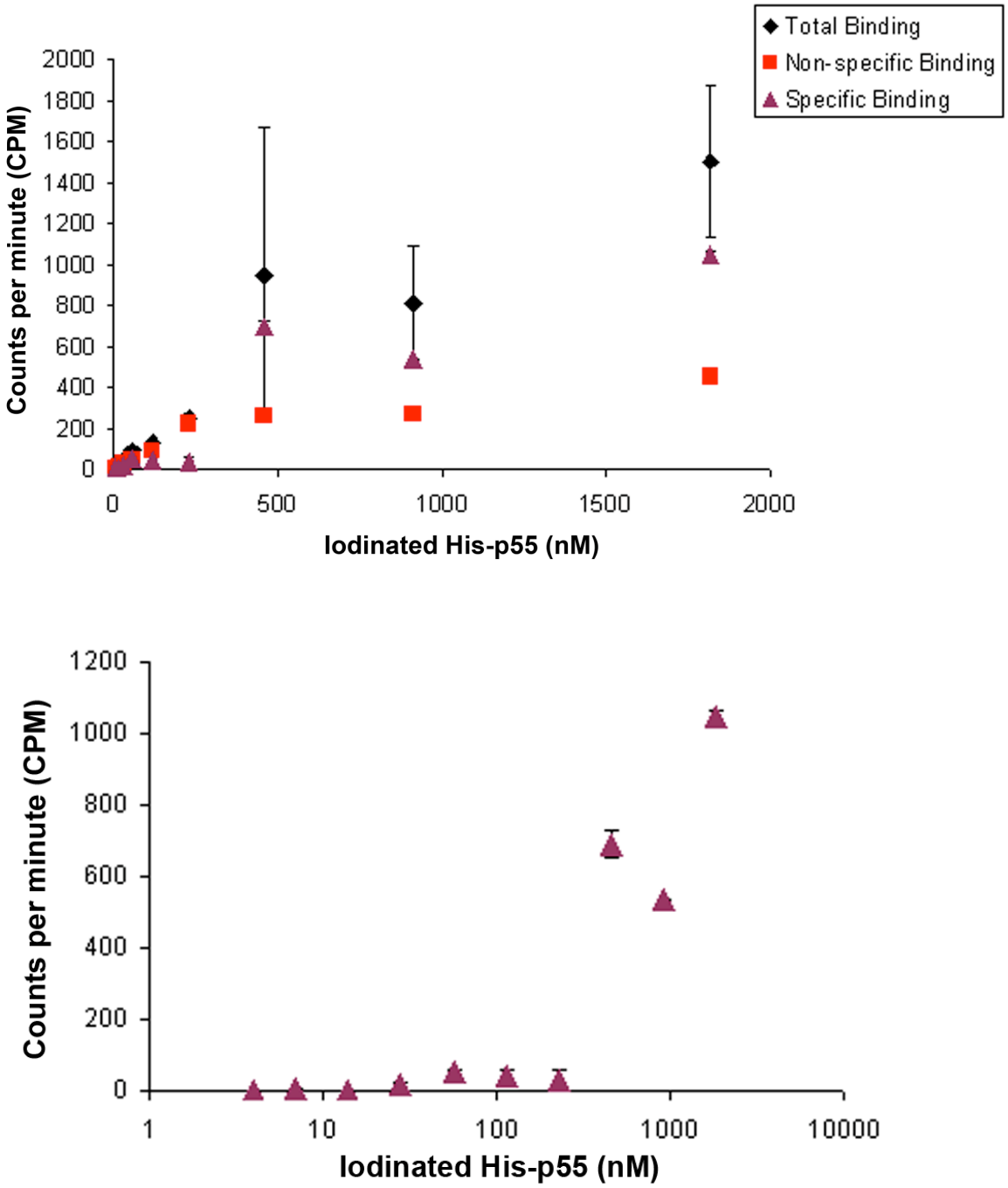


Figure 39. Analysis of VacA p55 binding to Human Adenocarcinoma Gastric Epithelial Cells (AGS) by radioiodination. AGS cells were incubated in medium, 2% fetal bovine serum with ^{125}I 'hot' His-p55 alone or plus a 100-fold molar excess of cold His-p55 at various concentrations. Specific binding is the difference between total binding and nonspecific binding. The lower panel represents a log scale of specific binding.

LIST OF PUBLICATIONS

Gangwer, K. A.,Mushrush, D. J., Stauff, D. L., Spiller, B., McClain, M. S., Cover, T. L., and Lacy, D. B. 2007. Crystal structure of the *Helicobacter pylori* vacuolating toxin p55 domain.*Proc Natl Acad Sci U S A* **104** (41), 16293-8**.**

*Gonzalez-Rivera, C., *Gangwer, K.A**.,** McClain, M.S., Ilyas, E.M., Chambers, M.G., Ohi, M.D., Lacy, D.B., and Cover, T.L. 2010. Reconstitution of *Helicobacter pylori* VacA toxin from purified components. *Biochemistry* **49** (27), 5743-52.

*These authors contributed equally to this work

Gangwer, K.A., Shaffer, C.L., Lacy, D.B., Cover, T.L., Bordenstein, S.R. 2010 Molecular Evolution of *Helicobacter pylori* Vacuolating Toxin (VacA). **In preparation.**

BIBLIOGRAPHY

1. **Doenges, J.** 1938. Spirochaetes in gastric glands of macacus rhesus and humans without definite history of related disease. Proc Soc Exp Biol Med **38,** 536-538.

2. **Fung, W. P., Papadimitriou, J. M., & Matz, L. R.** 1979. Endoscopic, histological and ultrastructural correlations in chronic gastritis. Am J Gastroenterol **71,** 269-279.

3. **Warren, J. R., and Marshall, B.J.** 1983. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. Lancet **1,** 1273-1275.

4. **Marshall, B. J. & Warren, J. R.** 1984. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. Lancet **1,** 1311-1315.

5. **Marshall, B. J., Armstrong, J. A., McGechie, D. B., & Glancy, R. J.** 1985. Attempt to fulfil Koch's postulates for pyloric *Campylobacter*. Med J Aust **142,** 436-439.

6. **Wyatt, J. I., Rathbone, B. J., Dixon, M. F., & Heatley, R. V.** 1987. *Campylobacter pyloridis* and acid induced gastric metaplasia in the pathogenesis of duodenitis. J Clin Pathol **40,** 841-848.

7. **Marshall, B. J., McGechie, D. B., Rogers, P. A., & Glancy, R. J.** 1985. Pyloric *Campylobacte*r infection and gastroduodenal disease. Med J Aust **142,** 439-444.

8. **Carrick, J., Lee, A., Hazell, S., Ralston, M., & Daskalopoulos, G.** 1989. *Campylobacter pylori*, duodenal ulcer, and gastric metaplasia: possible role of functional heterotopic tissue in ulcerogenesis. Gut **30,** 790-797.

9. **Johnston, B. J., Reed, P. I., & Ali, M. H.** 1986. Campylobacter like organisms in duodenal and antral endoscopic biopsies: relationship to inflammation. Gut **27,** 1132-1137.

10. 1994. NIH Consensus Conference. *Helicobacter pylori* in peptic ulcer disease. NIH Consensus Development Panel on Helicobacter pylori in Peptic Ulcer Disease. Jama **272,** 65-69.

11. **Blaser, M. J.** 1990. *Helicobacter pylori* and the pathogenesis of gastroduodenal inflammation. J Infect Dis **161,** 626-633.

12. **Blaser, M. J., Perez-Perez, G. I., Lindenbaum, J., Schneidman, D., Van Deventer, G., Marin-Sorensen, M., & Weinstein, W. M.** 1991. Association of infection due to *Helicobacter pylori* with specific upper gastrointestinal pathology. Rev Infect Dis **13 Suppl 8,** S704-708.

13. **Taylor, D. N. & Blaser, M. J.** 1991. The epidemiology of *Helicobacter pylori* infection. Epidemiol Rev **13,** 42-59.

14. **Pounder, R. E. & Ng, D.** 1995. The prevalence of *Helicobacter pylori* infection in different countries. Aliment Pharmacol Ther **9 Suppl 2,** 33-39.

15. **Dixon, M. F., Genta, R. M., Yardley, J. H., & Correa, P.** 1996. Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994. Am J Surg Pathol **20,** 1161-1181.

16. **Langenberg, W., Rauws, E. A., Houthoff, H. J., Oudbier, J. H., van Bohemen, C. G., Tytgat, G. N., & Rietra, P. J.** 1988. Follow-up study of individuals with untreated *Campylobacter* pylori-associated gastritis and of noninfected persons with non-ulcer dyspepsia. J Infect Dis **157,** 1245-1249.

17. **Suerbaum, S., Smith, J. M., Bapumia, K., Morelli, G., Smith, N. H., Kunstmann, E., Dyrek, I., & Achtman, M.** 1998. Free recombination within *Helicobacter pylori*. Proc Natl Acad Sci U S A **95,** 12619-12624.

18. **Blaser, M. J. & Berg, D. E.** 2001. *Helicobacter pylori* genetic diversity and risk of human disease. J Clin Invest **107,** 767-773.

19. **Nomura, A., Stemmermann, G. N., Chyou, P. H., Kato, I., Perez-Perez, G. I., & Blaser, M. J.** 1991. *Helicobacter pylori* infection and gastric carcinoma among Japanese Americans in Hawaii. N Engl J Med **325,** 1132-1136.

20. **Parsonnet, J., Friedman, G. D., Vandersteen, D. P., Chang, Y., Vogelman, J. H., Orentreich, N., & Sibley, R. K.** 1991. *Helicobacter pylori* infection and the risk of gastric carcinoma. N Engl J Med **325,** 1127-1131.

21. **Bayerdorffer, E., Neubauer, A., Rudolph, B., Thiede, C., Lehn, N., Eidt, S., & Stolte, M.** 1995. Regression of primary gastric lymphoma of mucosa-associated lymphoid tissue type after cure of *Helicobacter pylori* infection. MALT Lymphoma Study Group. Lancet **345,** 1591-1594.

22. **Parsonnet, J., Hansen, S., Rodriguez, L., Gelb, A. B., Warnke, R. A., Jellum, E., Orentreich, N., Vogelman, J. H., & Friedman, G. D.** 1994. *Helicobacter pylori infection* and gastric lymphoma. N Engl J Med **330,** 1267-1271.

23. **Wotherspoon, A. C., Doglioni, C., Diss, T. C., Pan, L., Moschini, A., de Boni, M., & Isaacson, P. G.** 1993. Regression of primary low-grade B-cell gastric lymphoma of mucosa-associated lymphoid tissue type after eradication of *Helicobacter pylori*. Lancet **342,** 575-577.

24. **Wotherspoon, A. C., Ortiz-Hidalgo, C., Falzon, M. R., & Isaacson, P. G.** 1991. *Helicobacter pylori*-associated gastritis and primary B-cell gastric lymphoma. Lancet **338,** 1175-1176.

25. **Zucca, E., Bertoni, F., Roggero, E., Bosshard, G., Cazzaniga, G., Pedrinis, E., Biondi, A., & Cavalli, F.** 1998. Molecular analysis of the progression from Helicobacter pylori-associated chronic gastritis to mucosa-associated lymphoid-tissue lymphoma of the stomach. N Engl J Med **338,** 804-810.

26. **Alm, R. A., Ling, L. S., Moir, D. T., King, B. L., Brown, E. D., Doig, P. C., Smith, D. R., Noonan, B., Guild, B. C., deJonge, B. L.*, et al.*** 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. Nature **397,** 176-180.

27. **Oh, J. D., Kling-Backhed, H., Giannakis, M., Xu, J., Fulton, R. S., Fulton, L. A., Cordum, H. S., Wang, C., Elliott, G., Edwards, J.*, et al.*** 2006. The complete genome sequence of a chronic atrophic gastritis *Helicobacter pylori* strain: evolution during disease progression. Proc Natl Acad Sci U S A **103,** 9999-10004.

28. **Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A.*, et al.*** 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature **388,** 539-547.

29. **Baltrus, D. A., Amieva, M. R., Covacci, A., Lowe, T. M., Merrell, D. S., Ottemann, K. M., Stein, M., Salama, N. R., & Guillemin, K.** 2009. The complete genome sequence of *Helicobacter pylori* strain G27. J Bacteriol **191,** 447-448.

30. **Fischer, W., Windhager, L., Rohrer, S., Zeiller, M., Karnholz, A., Hoffmann, R., Zimmer, R., & Haas, R.** Strain-specific genes of *Helicobacter pylori*: genome evolution driven by a novel type IV secretion system and genomic island transfer. Nucleic Acids Res.

31. **Beji, A., Megraud, F., Vincent, P., Gavini, F., Izard, D., & Leclerc, H.** 1988. GC content of DNA of *Campylobacter pylori* and other species belonging or related to the genus Campylobacter. Ann Inst Pasteur Microbiol **139,** 527-534.

32. **Cover, T. L., Berg, D.E., Blaser, M.J, Mobley, L.T** 2001. in *Principles of Bacterial Pathogenesis*, ed. Press, A.

33. **Go, M. F., Kapur, V., Graham, D. Y., & Musser, J. M.** 1996. Population genetic analysis of *Helicobacter pylori* by multilocus enzyme electrophoresis: extensive allelic diversity and recombinational population structure. J Bacteriol **178,** 3934-3938.

34. **Israel, D. A., Salama, N., Krishna, U., Rieger, U. M., Atherton, J. C., Falkow, S., & Peek, R. M., Jr.** 2001. *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. Proc Natl Acad Sci U S A **98,** 14625-14630.

35. **Achtman, M., Azuma, T., Berg, D. E., Ito, Y., Morelli, G., Pan, Z. J., Suerbaum, S., Thompson, S. A., van der Ende, A., & van Doorn, L. J.** 1999. Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. Mol Microbiol **32,** 459-470.

36. **Hazell, S. L., Andrews, R. H., Mitchell, H. M., & Daskalopoulous, G.** 1997. Genetic relationship among isolates of *Helicobacter pylori*: evidence for the existence of a Helicobacter pylori species-complex. FEMS Microbiol Lett **150,** 27-32.

37. **Salaun, L., Audibert, C., Le Lay, G., Burucoa, C., Fauchere, J. L., & Picard, B.** 1998. Panmictic structure of *Helicobacter pylori* demonstrated by the comparative study of six genetic markers. FEMS Microbiol Lett **161,** 231-239.

38. **Foxall, P. A., Hu, L. T., & Mobley, H. L.** 1992. Use of polymerase chain reaction-amplified *Helicobacter pylori* urease structural genes for differentiation of isolates. J Clin Microbiol **30,** 739-741.

39. **Desai, M., Linton, D., Owen, R. J., & Stanley, J.** 1994. Molecular typing of *Helicobacter pylori* isolates from asymptomatic, ulcer and gastritis patients by urease gene polymorphism. Epidemiol Infect **112,** 151-160.

40. **Forbes, K. J., Fang, Z., & Pennington, T. H.** 1995. Allelic variation in the *Helicobacter pylori* flagellin genes flaA and flaB: its consequences for strain typing schemes and population structure. Epidemiol Infect **114,** 257-266.

41. **Cover, T. L., Tummuru, M. K., Cao, P., Thompson, S. A., & Blaser, M. J.** 1994. Divergence of genetic sequences for the vacuolating cytotoxin among *Helicobacter pylori* strains. J Biol Chem **269,** 10566-10573.

42. **Garner, J. A. & Cover, T. L.** 1995. Analysis of genetic diversity in cytotoxin-producing and non-cytotoxin-producing *Helicobacter pylori* strains. J Infect Dis **172,** 290-293.

43. **Kato, S., Onda, M., Matsukura, N., Tokunaga, A., Matsuda, N., Yamashita, K., & Shields, P. G.** 1996. Genetic polymorphisms of the cancer related gene and *Helicobacter pylori* infection in Japanese gastric cancer patients. An age and gender matched case-control study. Cancer **77,** 1654-1661.

44. **Miehlke, S., Kibler, K., Kim, J. G., Figura, N., Small, S. M., Graham, D. Y., & Go, M. F.** 1996. Allelic variation in the cagA gene of *Helicobacter pylori* obtained from Korea compared to the United States. Am J Gastroenterol **91,** 1322-1325.

45. **Goodwin, C. S. & Armstrong, J. A.** 1990. Microbiological aspects of *Helicobacter pylori* (*Campylobacter pylori*). Eur J Clin Microbiol Infect Dis **9,** 1-13.

46. **Quigley, E. M. & Turnberg, L. A.** 1987. pH of the microclimate lining human gastric and duodenal mucosa *in vivo*. Studies in control subjects and in duodenal ulcer patients. Gastroenterology **92,** 1876-1884.

47. **Dunn, B. E., Cohen, H., & Blaser, M. J.** 1997. *Helicobacter pylori*. Clin Microbiol Rev **10,** 720-741.

48. **Mobley, H. L., Island, M. D., & Hausinger, R. P.** 1995. Molecular biology of microbial ureases. Microbiol Rev **59,** 451-480.

49. **Cussac, V., Ferrero, R. L., & Labigne, A.** 1992. Expression of *Helicobacter pylori* urease genes in *Escherichia coli* grown under nitrogen-limiting conditions. J Bacteriol **174,** 2466-2473.

50. **Dunn, B. E., Campbell, G. P., Perez-Perez, G. I., & Blaser, M. J.** 1990. Purification and characterization of urease from *Helicobacter pylori*. J Biol Chem **265,** 9464-9469.

51. **Hu, L. T. & Mobley, H. L.** 1990. Purification and N-terminal analysis of urease from *Helicobacter pylori*. Infect Immun **58,** 992-998.

52. **Labigne, A., Cussac, V., & Courcoux, P.** 1991. Shuttle cloning and nucleotide sequences of *Helicobacter pylori* genes responsible for urease activity. J Bacteriol **173,** 1920-1931.

53. **Blaser, M. J., Perez-Perez, G. I., Kleanthous, H., Cover, T. L., Peek, R. M., Chyou, P. H., Stemmermann, G. N., & Nomura, A.** 1995. Infection with *Helicobacter pylori* strains possessing cagA is associated with an increased risk of developing adenocarcinoma of the stomach. Cancer Res **55,** 2111-2115.

54. **Odenbreit, S., Puls, J., Sedlmaier, B., Gerland, E., Fischer, W., & Haas, R.** 2000. Translocation of *Helicobacter pylori* CagA into gastric epithelial cells by type IV secretion. Science **287,** 1497-1500.

55. **Leunk, R. D., Johnson, P. T., David, B. C., Kraft, W. G., & Morgan, D. R.** 1988. Cytotoxic activity in broth-culture filtrates of *Campylobacter pylori*. J Med Microbiol **26,** 93-99.

56. **Cover, T. L. & Blaser, M. J.** 1992. Purification and characterization of the vacuolating toxin from *Helicobacter pylori*. J Biol Chem **267,** 10570-10575.

57. **Cover, T. L. & Blanke, S. R.** 2005. *Helicobacter pylori* VacA, a paradigm for toxin multifunctionality. Nat Rev Microbiol **3,** 320-332.

58. **de Bernard, M., Cappon, A., Del Giudice, G., Rappuoli, R., & Montecucco, C.** 2004. The multiple cellular activities of the VacA cytotoxin of *Helicobacter pylori.* Int J Med Microbiol **293,** 589-597.

59. **Fischer, W., Prassl, S., & Haas, R.** 2009. Virulence mechanisms and persistence strategies of the human gastric pathogen *Helicobacter pylori*. Curr Top Microbiol Immunol **337,** 129-171.

60. **Adrian, M., Cover, T. L., Dubochet, J., & Heuser, J. E.** 2002. Multiple oligomeric states of the *Helicobacter pylori* vacuolating toxin demonstrated by cryo-electron microscopy. J Mol Biol **318,** 121-133.

61. **Czajkowsky, D. M., Iwamoto, H., Cover, T. L., & Shao, Z.** 1999. The vacuolating toxin from *Helicobacter pylori* forms hexameric pores in lipid bilayers at low pH. Proc Natl Acad Sci U S A **96,** 2001-2006.

62. **Geisse, N. A., Cover, T. L., Henderson, R. M., & Edwardson, J. M.** 2004. Targeting of *Helicobacter pylori* vacuolating toxin to lipid raft membrane domains analysed by atomic force microscopy. Biochem J **381,** 911-917.

63. **Atherton, J. C., Cao, P., Peek, R. M., Jr., Tummuru, M. K., Blaser, M. J., & Cover, T. L.** 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.

64. **van Doorn, L. J., Figueiredo, C., Sanna, R., Pena, S., Midolo, P., Ng, E. K., Atherton, J. C., Blaser, M. J., & Quint, W. G.** 1998. Expanding allelic diversity of *Helicobacter pylori vacA*. J Clin Microbiol **36,** 2597-2603.

65. **Letley, D. P. & Atherton, J. C.** 2000. Natural diversity in the N terminus of the mature vacuolating cytotoxin of Helicobacter pylori determines cytotoxin activity. J Bacteriol **182,** 3278-3280.

66. **Letley, D. P., Rhead, J. L., Bishop, K., & Atherton, J. C.** 2006. Paired cysteine residues are required for high levels of the *Helicobacter pylori* autotransporter VacA. Microbiology **152,** 1319-1325.

67. **McClain, M. S., Cao, P., Iwamoto, H., Vinion-Dubiel, A. D., Szabo, G., Shao, Z., & Cover, T. L.** 2001. A 12-amino-acid segment, present in type s2 but not type s1 *Helicobacter pylori* VacA proteins, abolishes cytotoxin activity and alters membrane channel formation. J Bacteriol **183,** 6499-6508.

68. **Pagliaccia, C., de Bernard, M., Lupetti, P., Ji, X., Burroni, D., Cover, T. L., Papini, E., Rappuoli, R., Telford, J. L., & Reyrat, J. M.** 1998. The m2 form of the *Helicobacter pylori* cytotoxin has cell type-specific vacuolating activity. Proc Natl Acad Sci U S A **95,** 10212-10217.

69. **Van Doorn, L. J., Figueiredo, C., Megraud, F., Pena, S., Midolo, P., Queiroz, D. M., Carneiro, F., Vanderborght, B., Pegado, M. D., Sanna, R.*, et al.*** 1999. Geographic distribution of vacA allelic types of *Helicobacter pylori*. Gastroenterology **116,** 823-830.

70. **Figueiredo, C., Machado, J. C., Pharoah, P., Seruca, R., Sousa, S., Carvalho, R., Capelinha, A. F., Quint, W., Caldas, C., van Doorn, L. J.*, et al.*** 2002. *Helicobacter pylori* and interleukin 1 genotyping: an opportunity to identify high-risk individuals for gastric carcinoma. J Natl Cancer Inst **94,** 1680-1687.

71. **van Doorn, L. J., Figueiredo, C., Sanna, R., Plaisier, A., Schneeberger, P., de Boer, W., & Quint, W.** 1998. Clinical relevance of the cagA, vacA, and iceA status of *Helicobacter pylori*. Gastroenterology **115,** 58-66.

72. **Atherton, J. C., Peek, R. M., Jr., Tham, K. T., Cover, T. L., & Blaser, M. J.** 1997. Clinical and pathological importance of heterogeneity in *vacA*, the vacuolating cytotoxin gene of *Helicobacter pylori*. Gastroenterology **112,** 92-99.

73. **Telford, J. L., Ghiara, P., Dell'Orco, M., Comanducci, M., Burroni, D., Bugnoli, M., Tecce, M. F., Censini, S., Covacci, A., Xiang, Z.*, et al.*** 1994. Gene structure of the *Helicobacter pylori* cytotoxin and evidence of its key role in gastric disease. J Exp Med **179,** 1653-1658.

74. **Eaton, K. A., Cover, T. L., Tummuru, M. K., Blaser, M. J., & Krakowka, S.** 1997. Role of vacuolating cytotoxin in gastritis due to *Helicobacter pylori* in gnotobiotic piglets. Infect Immun **65,** 3462-3464.

75. **Ogura, K., Maeda, S., Nakao, M., Watanabe, T., Tada, M., Kyutoku, T., Yoshida, H., Shiratori, Y., & Omata, M.** 2000. Virulence factors of *Helicobacter pylori* responsible for gastric diseases in Mongolian gerbil. J Exp Med **192,** 1601-1610.

76. **Gauthier, N. C., Ricci, V., Gounon, P., Doye, A., Tauc, M., Poujeol, P., & Boquet, P.** 2004. Glycosylphosphatidylinositol-anchored proteins and actin cytoskeleton modulate chloride transport by channels formed by the *Helicobacter pylori* vacuolating cytotoxin VacA in HeLa cells. J Biol Chem **279,** 9481-9489.

77. **Iwamoto, H., Czajkowsky, D. M., Cover, T. L., Szabo, G., & Shao, Z.** 1999. VacA from *Helicobacter pylori*: a hexameric chloride channel. FEBS Lett **450,** 101-104.

78. **Szabo, I., Brutsche, S., Tombola, F., Moschioni, M., Satin, B., Telford, J. L., Rappuoli, R., Montecucco, C., Papini, E., & Zoratti, M.** 1999. Formation of anion-selective channels in the cell plasma membrane by the toxin VacA of *Helicobacter pylori* is required for its biological activity. EMBO J **18,** 5517-5527.

79. **Tombola, F., Carlesso, C., Szabo, I., de Bernard, M., Reyrat, J. M., Telford, J. L., Rappuoli, R., Montecucco, C., Papini, E., & Zoratti, M.** 1999. *Helicobacter pylori* vacuolating toxin forms anion-selective channels in planar lipid bilayers: possible implications for the mechanism of cellular vacuolation. Biophys J **76,** 1401-1409.

80. **Tombola, F., Oregna, F., Brutsche, S., Szabo, I., Del Giudice, G., Rappuoli, R., Montecucco, C., Papini, E., & Zoratti, M.** 1999. Inhibition of the vacuolating and anion channel activities of the VacA toxin of *Helicobacter pylori*. FEBS Lett **460,** 221-225.

81. **Morbiato, L., Tombola, F., Campello, S., Del Giudice, G., Rappuoli, R., Zoratti, M., & Papini, E.** 2001. Vacuolation induced by VacA toxin of *Helicobacter pylori* requires the intracellular accumulation of membrane permeant bases, Cl(-) and water. FEBS Lett **508,** 479-483.

82. **Cover, T. L., Vaughn, S. G., Cao, P., & Blaser, M. J.** 1992. Potentiation of *Helicobacter pylori* vacuolating toxin activity by nicotine and other weak bases. J Infect Dis **166,** 1073-1078.

83. **Li, Y., Wandinger-Ness, A., Goldenring, J. R., & Cover, T. L.** 2004. Clustering and redistribution of late endocytic compartments in response to *Helicobacter pylori* vacuolating toxin. Mol Biol Cell **15,** 1946-1959.

84. **Papini, E., Satin, B., Norais, N., de Bernard, M., Telford, J. L., Rappuoli, R., & Montecucco, C.** 1998. Selective increase of the permeability of polarized epithelial cell monolayers by *Helicobacter pylori* vacuolating toxin. J Clin Invest **102,** 813-820.

85. **Tombola, F., Morbiato, L., Del Giudice, G., Rappuoli, R., Zoratti, M., & Papini, E.** 2001. The *Helicobacter pylori* VacA toxin is a urea permease that promotes urea diffusion across epithelia. J Clin Invest **108,** 929-937.

86. **Debellis, L., Papini, E., Caroppo, R., Montecucco, C., & Curci, S.** 2001. *Helicobacter pylori* cytotoxin VacA increases alkaline secretion in gastric epithelial cells. Am J Physiol Gastrointest Liver Physiol **281,** G1440-1448.

87. **Boncristiano, M., Paccani, S. R., Barone, S., Ulivieri, C., Patrussi, L., Ilver, D., Amedei, A., D'Elios, M. M., Telford, J. L., & Baldari, C. T.** 2003. The *Helicobacter pylori* vacuolating toxin inhibits T cell activation by two independent mechanisms. J Exp Med **198,** 1887-1897.

88. **Gebert, B., Fischer, W., Weiss, E., Hoffmann, R., & Haas, R.** 2003. *Helicobacter pylori* vacuolating cytotoxin inhibits T lymphocyte activation. Science **301,** 1099-1102.

89. **Sundrud, M. S., Torres, V. J., Unutmaz, D., & Cover, T. L.** 2004. Inhibition of primary human T cell proliferation by *Helicobacter pylori* vacuolating toxin (VacA) is independent of VacA effects on IL-2 secretion. Proc Natl Acad Sci U S A **101,** 7727-7732.

90. **Galmiche, A., Rassow, J., Doye, A., Cagnol, S., Chambard, J. C., Contamin, S., de Thillot, V., Just, I., Ricci, V., Solcia, E.*, et al.*** 2000. The N-terminal 34 kDa fragment of *Helicobacter pylori* vacuolating cytotoxin targets mitochondria and induces cytochrome c release. Embo J **19,** 6361-6370.

91. **Fujikawa, A., Shirasaka, D., Yamamoto, S., Ota, H., Yahiro, K., Fukada, M., Shintani, T., Wada, A., Aoyama, N., Hirayama, T.*, et al.*** 2003. Mice deficient in protein tyrosine phosphatase receptor type Z are resistant to gastric ulcer induction by VacA of *Helicobacter pylori*. Nat Genet **33,** 375-381.

92. **Nakayama, M., Kimura, M., Wada, A., Yahiro, K., Ogushi, K., Niidome, T., Fujikawa, A., Shirasaka, D., Aoyama, N., Kurazono, H.*, et al.*** 2004. *Helicobacter pylori* VacA activates the p38/activating transcription factor 2-mediated signal pathway in AZ-521 cells. J Biol Chem **279,** 7024-7028.

93. **Henderson, I. R., Nataro, J. P., Kaper, J. B., Meyer, T. F., Farrand, S. K., Burns, D. L., Finlay, B. B., & St Geme, J. W., 3rd** 2000. Renaming protein secretion in the gram-negative bacteria. Trends Microbiol **8,** 352.

94. **Pukatzki, S., Ma, A. T., Sturtevant, D., Krastins, B., Sarracino, D., Nelson, W. C., Heidelberg, J. F., & Mekalanos, J. J.** 2006. Identification of a conserved bacterial protein secretion system in Vibrio cholerae using the Dictyostelium host model system. Proc Natl Acad Sci U S A **103,** 1528-1533.

95. **Schmitt, W. & Haas, R.** 1994. Genetic analysis of the *Helicobacter pylori* vacuolating cytotoxin: structural similarities with the IgA protease type of exported protein. Mol Microbiol **12,** 307-319.

96. **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.

97. **Mekada, E., Okada, Y., & Uchida, T.** 1988. Identification of diphtheria toxin receptor and a nonproteinous diphtheria toxin-binding molecule in Vero cell membrane. J Cell Biol **107,** 511-519.

98. **Yahiro, K., Niidome, T., Kimura, M., Hatakeyama, T., Aoyagi, H., Kurazono, H., Imagawa, K., Wada, A., Moss, J., & Hirayama, T.** 1999. Activation of *Helicobacter pylori* VacA toxin by alkaline or acid conditions increases its binding to a 250-kDa receptor protein-tyrosine phosphatase beta. J Biol Chem **274,** 36693-36699.

99. **Yahiro, K., Wada, A., Nakayama, M., Kimura, T., Ogushi, K., Niidome, T., Aoyagi, H., Yoshino, K., Yonezawa, K., Moss, J.*, et al.*** 2003. Protein-tyrosine phosphatase alpha, RPTP alpha, is a *Helicobacter pylori* VacA receptor. J Biol Chem **278,** 19183-19189.

100. **Sewald, X., Gebert-Vogl, B., Prassl, S., Barwig, I., Weiss, E., Fabbri, M., Osicka, R., Schiemann, M., Busch, D. H., Semmrich, M.*, et al.*** 2008. Integrin subunit CD18 Is the T-lymphocyte receptor for the *Helicobacter pylori* vacuolating cytotoxin. Cell Host Microbe **3,** 20-29.

101. **Molinari, M., Galli, C., de Bernard, M., Norais, N., Ruysschaert, J. M., Rappuoli, R., & Montecucco, C.** 1998. The acid activation of *Helicobacter pylori* toxin VacA: structural and membrane binding studies. Biochem Biophys Res Commun **248,** 334-340.

102. **Moll, G., Papini, E., Colonna, R., Burroni, D., Telford, J., Rappuoli, R., & Montecucco, C.** 1995. Lipid interaction of the 37-kDa and 58-kDa fragments of the *Helicobacter pylori* cytotoxin. Eur J Biochem **234,** 947-952.

103. **Schraw, W., Li, Y., McClain, M. S., van der Goot, F. G., & Cover, T. L.** 2002. Association of *Helicobacter pylori* vacuolating toxin (VacA) with lipid rafts. J Biol Chem **277,** 34642-34650.

104. **Gupta, V. R., Patel, H. K., Kostolansky, S. S., Ballivian, R. A., Eichberg, J., & Blanke, S. R.** 2008. Sphingomyelin functions as a novel receptor for *Helicobacter pylori* VacA. PLoS Pathog **4,** e1000073.

105. **Vinion-Dubiel, A. D., McClain, M. S., Czajkowsky, D. M., Iwamoto, H., Ye, D., Cao, P., Schraw, W., Szabo, G., Blanke, S. R., Shao, Z.*, et al.*** 1999. A dominant negative mutant of *Helicobacter pylori* vacuolating toxin (VacA) inhibits VacA-induced cell vacuolation. J Biol Chem **274,** 37736-37742.

106. **Cover, T. L., Hanson, P. I., & Heuser, J. E.** 1997. Acid-induced dissociation of VacA, the *Helicobacter pylori* vacuolating cytotoxin, reveals its pattern of assembly. J Cell Biol **138,** 759-769.

107. **El-Bez, C., Adrian, M., Dubochet, J., & Cover, T. L.** 2005. High resolution structural analysis of *Helicobacter pylori* VacA toxin oligomers by cryo-negative staining electron microscopy. J Struct Biol **151,** 215-228.

108. **Lanzavecchia, S., Bellon, P. L., Lupetti, P., Dallai, R., Rappuoli, R., & Telford, J. L.** 1998. Three-dimensional reconstruction of metal replicas of the *Helicobacter pylori* vacuolating cytotoxin. J Struct Biol **121,** 9-18.

109. **Lupetti, P., Heuser, J. E., Manetti, R., Massari, P., Lanzavecchia, S., Bellon, P. L., Dallai, R., Rappuoli, R., & Telford, J. L.** 1996. Oligomeric and subunit structure of the *Helicobacter pylori* vacuolating cytotoxin. J Cell Biol **133,** 801-807.

110. **de Bernard, M., Papini, E., de Filippis, V., Gottardi, E., Telford, J., Manetti, R., Fontana, A., Rappuoli, R., & Montecucco, C.** 1995. Low pH activates the vacuolating toxin of *Helicobacter pylori*, which becomes acid and pepsin resistant. J Biol Chem **270,** 23937-23940.

111. **Gauthier, N. C., Monzo, P., Gonzalez, T., Doye, A., Oldani, A., Gounon, P., Ricci, V., Cormont, M., & Boquet, P.** 2007. Early endosomes associated with dynamic F-actin structures are required for late trafficking of *H. pylori* VacA toxin. J Cell Biol **177,** 343-354.

112. **Genisset, C., Puhar, A., Calore, F., de Bernard, M., Dell'Antone, P., & Montecucco, C.** 2007. The concerted action of the *Helicobacter pylori* cytotoxin VacA and of the v-ATPase proton pump induces swelling of isolated endosomes. Cell Microbiol **9,** 1481-1490.

113. **Torres, V. J., Ivie, S. E., McClain, M. S., & Cover, T. L.** 2005. Functional properties of the p33 and p55 domains of the *Helicobacter pylori* vacuolating cytotoxin. J Biol Chem **280,** 21107-21114.

114. **Torres, V. J., McClain, M. S., & Cover, T. L.** 2004. Interactions between p-33 and p-55 domains of the *Helicobacter pylori* vacuolating cytotoxin (VacA). J Biol Chem **279,** 2324-2331.

115. **Ye, D., Willhite, D. C., & Blanke, S. R.** 1999. Identification of the minimal intracellular vacuolating domain of the *Helicobacter pylori* vacuolating toxin. J Biol Chem **274,** 9277-9282.

116. **McClain, M. S., Iwamoto, H., Cao, P., Vinion-Dubiel, A. D., Li, Y., Szabo, G., Shao, Z., & Cover, T. L.** 2003. Essential role of a GXXXG motif for membrane channel formation by *Helicobacter pylori* vacuolating toxin. J Biol Chem **278,** 12101-12108.

117. **Garner, J. A. & Cover, T. L.** 1996. Binding and internalization of the *Helicobacter pylori* vacuolating cytotoxin by epithelial cells. Infect Immun **64,** 4197-4203.

118. **Reyrat, J. M., Lanzavecchia, S., Lupetti, P., de Bernard, M., Pagliaccia, C., Pelicic, V., Charrel, M., Ulivieri, C., Norais, N., Ji, X.*, et al.*** 1999. 3D imaging

of the 58 kDa cell binding subunit of the *Helicobacter pylori* cytotoxin. J Mol Biol **290,** 459-470.

119. **Wang, H. J. & Wang, W. C.** 2000. Expression and binding analysis of GST-VacA fusions reveals that the C-terminal approximately 100-residue segment of exotoxin is crucial for binding in HeLa cells. Biochem Biophys Res Commun **278,** 449-454.

120. **Genisset, C., Galeotti, C. L., Lupetti, P., Mercati, D., Skibinski, D. A., Barone, S., Battistutta, R., de Bernard, M., & Telford, J. L.** 2006. A *Helicobacter pylori* vacuolating toxin mutant that fails to oligomerize has a dominant negative phenotype. Infect Immun **74,** 1786-1794.

121. **Ivie, S. E., McClain, M. S., Torres, V. J., Algood, H. M., Lacy, D. B., Yang, R., Blanke, S. R., & Cover, T. L.** 2008. *Helicobacter pylori* VacA subdomain required for intracellular toxin activity and assembly of functional oligomeric complexes. Infect Immun **76,** 2843-2851.

122. **Ye, D. & Blanke, S. R.** 2002. Functional complementation reveals the importance of intermolecular monomer interactions for *Helicobacter pylori* VacA vacuolating activity. Mol Microbiol **43,** 1243-1253.

123. **Torres, V. J., McClain, M. S., & Cover, T. L.** 2006. Mapping of a domain required for protein-protein interactions and inhibitory activity of a *Helicobacter pylori* dominant-negative VacA mutant protein. Infect Immun **74,** 2093-2101.

124. **Atherton, J. C., Sharp, P. M., Cover, T. L., Gonzalez-Valencia, G., Peek, R. M., Jr., Thompson, S. A., Hawkey, C. J., & Blaser, M. J.** 1999. Vacuolating cytotoxin (vacA) alleles of *Helicobacter pylori* comprise two geographically widespread types, m1 and m2, and have evolved through limited recombination. Curr Microbiol **39,** 211-218.

125. **Gottke, M. U., Fallone, C. A., Barkun, A. N., Vogt, K., Loo, V., Trautmann, M., Tong, J. Z., Nguyen, T. N., Fainsilber, T., Hahn, H. H.*, et al.*** 2000. Genetic variability determinants of *Helicobacter pylori*: influence of clinical background and geographic origin of isolates. J Infect Dis **181,** 1674-1681.

126. **Ji, X., Fernandez, T., Burroni, D., Pagliaccia, C., Atherton, J. C., Reyrat, J. M., Rappuoli, R., & Telford, J. L.** 2000. Cell specificity of *Helicobacter pylori* cytotoxin is determined by a short region in the polymorphic midregion. Infect Immun **68,** 3754-3757.

127. **Skibinski, D. A., Genisset, C., Barone, S., & Telford, J. L.** 2006. The cell-specific phenotype of the polymorphic vacA midregion is independent of the appearance of the cell surface receptor protein tyrosine phosphatase beta. Infect Immun **74,** 49-55.

128. **Wang, W. C., Wang, H. J., & Kuo, C. H.** 2001. Two distinctive cell binding patterns by vacuolating toxin fused with glutathione S-transferase: one high-affinity m1-specific binding and the other lower-affinity binding for variant m forms. Biochemistry **40,** 11887-11896.

129. **Otwinowski, Z. M., W.** 1997. Methods in Enzymology **276,** 307-326.

130. **Fortelle, E. d. l. & Bricogne, G.** 1997. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. Methods in Enzymology **276,** 472-494.

131. **Jones, T. A., Zou, J. Y., Cowan, S. W., & Kjeldgaard, M.** 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A **47 (Pt 2),** 110-119.

132. **Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S.*, et al.*** 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr **54,** 905-921.

133. **Winn, M. D., Isupov, M. N., & Murshudov, G. N.** 2001. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta Crystallogr D Biol Crystallogr **57,** 122-133.

134. **Murshudov, G. N., Vagin, A. A., & Dodson, E. J.** 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr **53,** 240-255.

135. **Morris, A. L., MacArthur, M. W., Hutchinson, E. G., & Thornton, J. M.** 1992. Stereochemical quality of protein structure coordinates. Proteins **12,** 345-364.

136. **Emsley, P., Charles, I. G., Fairweather, N. F., & Isaacs, N. W.** 1996. Structure of *Bordetella pertussis* virulence factor P.69 pertactin. Nature **381,** 90-92.

137. **Otto, B. R., Sijbrandi, R., Luirink, J., Oudega, B., Heddle, J. G., Mizutani, K., Park, S. Y., & Tame, J. R.** 2005. Crystal structure of hemoglobin protease, a heme binding autotransporter protein from pathogenic *Escherichia coli*. J Biol Chem **280,** 17339-17345.

138. **Junker, M., Schuster, C. C., McDonnell, A. V., Sorg, K. A., Finn, M. C., Berger, B., & Clark, P. L.** 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.

139. **Oliver, D. C., Huang, G., Nodel, E., Pleasance, S., & Fernandez, R. C.** 2003. A conserved region within the *Bordetella pertussis* autotransporter BrkA is necessary for folding of its passenger domain. Mol Microbiol **47,** 1367-1383.

140. **Barnard, T. J., Dautin, N., Lukacik, P., Bernstein, H. D., & Buchanan, S. K.** 2007. Autotransporter structure reveals intra-barrel cleavage followed by conformational changes. Nat Struct Mol Biol **14,** 1214-1220.

141. **Thompson, J. D., Higgins, D. G., & Gibson, T. J.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through

sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22,** 4673-4680.

142. **Gouet, P., Courcelle, E., Stuart, D. I., & Metoz, F.** 1999. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics **15,** 305-308.

143. **Wriggers, W., Milligan, R. A., & McCammon, J. A.** 1999. Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. J Struct Biol **125,** 185-195.

144. **McDonnell, A. V., Menke, M., Palmer, N., King, J., Cowen, L., & Berger, B.** 2006. Fold recognition and accurate sequence-structure alignment of sequences directing beta-sheet proteins. Proteins **63,** 976-985.

145. **Kim, S., Chamberlain, A. K., & Bowie, J. U.** 2004. Membrane channel structure of *Helicobacter pylori* vacuolating toxin: role of multiple GXXXG motifs in cylindrical channels. Proc Natl Acad Sci U S A **101,** 5988-5991.

146. **Ruggiero, P., Peppoloni, S., Rappuoli, R., & Del Giudice, G.** 2003. The quest for a vaccine against *Helicobacter pylori*: how to move from mouse to man? Microbes Infect **5,** 749-756.

147. **Marchetti, M., Arico, B., Burroni, D., Figura, N., Rappuoli, R., & Ghiara, P.** 1995. Development of a mouse model of *Helicobacter pylori* infection that mimics human disease. Science **267,** 1655-1658.

148. **Gangwer, K. A., Mushrush, D. J., Stauff, D. L., Spiller, B., McClain, M. S., Cover, T. L., & Lacy, D. B.** 2007. Crystal structure of the *Helicobacter pylori* vacuolating toxin p55 domain. Proc Natl Acad Sci U S A **104,** 16293-16298.

149. **Middelberg, A. P.** 2002. Preparative protein refolding. Trends Biotechnol **20,** 437-443.

150. **Cover, T. L., Puryear, W., Perez-Perez, G. I., & Blaser, M. J.** 1991. Effect of urease on HeLa cell vacuolation induced by *Helicobacter pylori* cytotoxin. Infect Immun **59,** 1264-1270.

151. **Algood, H. M., Torres, V. J., Unutmaz, D., & Cover, T. L.** 2007. Resistance of primary murine CD4+ T cells to *Helicobacter pylori* vacuolating cytotoxin. Infect Immun **75,** 334-341.

152. **Ohi, M., Li, Y., Cheng, Y., & Walz, T.** 2004. Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. Biol Proced Online **6,** 23-34.

153. **Hawrylik, S. J., Wasilko, D. J., Haskell, S. L., Gootz, T. D., & Lee, S. E.** 1994. Bisulfite or sulfite inhibits growth of *Helicobacter pylori*. J Clin Microbiol **32,** 790-792.

154. **Ludtke, S. J., Baldwin, P. R., & Chiu, W.** 1999. EMAN: semiautomated software for high-resolution single-particle reconstructions. J Struct Biol **128,** 82-97.

155. **Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., & Leith, A.** 1996. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J Struct Biol **116,** 190-199.

156. **Dautin, N. & Bernstein, H. D.** 2007. Protein secretion in gram-negative bacteria via the autotransporter pathway. Annu Rev Microbiol **61,** 89-112.

157. **Patel, H. K., Willhite, D. C., Patel, R. M., Ye, D., Williams, C. L., Torres, E. M., Marty, K. B., MacDonald, R. A., & Blanke, S. R.** 2002. Plasma membrane cholesterol modulates cellular vacuolation induced by the *Helicobacter pylori* vacuolating cytotoxin. Infect Immun **70,** 4112-4123.

158. **Falush, D., Wirth, T., Linz, B., Pritchard, J. K., Stephens, M., Kidd, M., Blaser, M. J., Graham, D. Y., Vacher, S., Perez-Perez, G. I.*, et al.*** 2003. Traces of human migrations in *Helicobacter pylori* populations. Science **299,** 1582-1585.

159. **Linz, B., Balloux, F., Moodley, Y., Manica, A., Liu, H., Roumagnac, P., Falush, D., Stamer, C., Prugnolle, F., van der Merwe, S. W.*, et al.*** 2007. An African origin for the intimate association between humans and *Helicobacter pylori*. Nature **445,** 915-918.

160. **Ito, Y., Azuma, T., Ito, S., Suto, H., Miyaji, H., Yamazaki, Y., Kohli, Y., & Kuriyama, M.** 1998. Full-length sequence analysis of the vacA gene from cytotoxic and noncytotoxic *Helicobacter pylori*. J Infect Dis **178,** 1391-1398.

161. **Rhead, J. L., Letley, D. P., Mohammadi, M., Hussein, N., Mohagheghi, M. A., Eshagh Hosseini, M., & Atherton, J. C.** 2007. A new *Helicobacter pylori* vacuolating cytotoxin determinant, the intermediate region, is associated with gastric cancer. Gastroenterology **133,** 926-936.

162. **Yamazaki, S., Yamakawa, A., Okuda, T., Ohtani, M., Suto, H., Ito, Y., Yamazaki, Y., Keida, Y., Higashi, H., Hatakeyama, M.*, et al.*** 2005. Distinct diversity of vacA, cagA, and cagE genes of *Helicobacter pylori* associated with peptic ulcer in Japan. J Clin Microbiol **43,** 3906-3916.

163. **Pan, Z. J., Berg, D. E., van der Hulst, R. W., Su, W. W., Raudonikiene, A., Xiao, S. D., Dankert, J., Tytgat, G. N., & van der Ende, A.** 1998. Prevalence of vacuolating cytotoxin production and distribution of distinct vacA alleles in *Helicobacter pylori* from China. J Infect Dis **178,** 220-226.

164. **Ji, X., Frati, F., Barone, S., Pagliaccia, C., Burroni, D., Xu, G., Rappuoli, R., Reyrat, J. M., & Telford, J. L.** 2002. Evolution of functional polymorphism in the gene coding for the *Helicobacter pylori* cytotoxin. FEMS Microbiol Lett **206,** 253-258.

165. **Argent, R. H., Thomas, R. J., Letley, D. P., Rittig, M. G., Hardie, K. R., & Atherton, J. C.** 2008. Functional association between the *Helicobacter pylori* virulence factors VacA and CagA. J Med Microbiol **57,** 145-150.

166. **Tegtmeyer, N., Zabler, D., Schmidt, D., Hartig, R., Brandt, S., & Backert, S.** 2009. Importance of EGF receptor, HER2/Neu and Erk1/2 kinase signalling for host cell elongation and scattering induced by the *Helicobacter pylori* CagA protein: antagonistic effects of the vacuolating cytotoxin VacA. Cell Microbiol **11,** 488-505.

167. **Yokoyama, K., Higashi, H., Ishikawa, S., Fujii, Y., Kondo, S., Kato, H., Azuma, T., Wada, A., Hirayama, T., Aburatani, H.*, et al.*** 2005. Functional antagonism between *Helicobacter pylori* CagA and vacuolating toxin VacA in control of the NFAT signaling pathway in gastric epithelial cells. Proc Natl Acad Sci U S A **102,** 9661-9666.

168. **Oldani, A., Cormont, M., Hofman, V., Chiozzi, V., Oregioni, O., Canonici, A., Sciullo, A., Sommi, P., Fabbri, A., Ricci, V.*, et al.*** 2009. *Helicobacter pylori* counteracts the apoptotic action of its VacA toxin by injecting the CagA protein into gastric epithelial cells. PLoS Pathog **5,** e1000603.

169. **Maddison, D. R. M., W.P.eds** 2005. Macclade 4: Analysis of Phylogeny and Character Evolution. (Sinauer Associates, Inc., Sunderland, MA).

170. **Drummond, A. J., Ashton, B., Cheung, M., Heled, J., Kearse, M., Moir, R., Stones-Havas, T., Thierer, T., & Wilson, A.** 2009. Geneious 4.6.5. Geneious 4.6.5.

171. **Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X., & Rozas, R.** 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19,** 2496-2497.

172. **McDonald, J. H. & Kreitman, M.** 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature **351,** 652-654.

173. **Wilson, D. J. & McVean, G.** 2006. Estimating diversifying selection and functional constraint in the presence of recombination. Genetics **172,** 1411-1425.

174. **Yang, Z.** 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci **13,** 555-556.

175. **Yang, Z.** 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol **24,** 1586-1591.

176. **Eppinger, M., Baar, C., Linz, B., Raddatz, G., Lanz, C., Keller, H., Morelli, G., Gressmann, H., Achtman, M., & Schuster, S. C.** 2006. Who ate whom? Adaptive *Helicobacter* genomic changes that accompanied a host jump from early humans to large felines. PLoS Genet **2,** e120.

177. **Lartillot, N., Brinkmann, H., & Philippe, H.** 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol Biol **7 Suppl 1,** S4.

178. **Lewis, R. L., Beckenbach, A. T., & Mooers, A. O.** 2005. The phylogeny of the subgroups within the melanogaster species group: likelihood tests on COI and COII sequences and a Bayesian estimate of phylogeny. Mol Phylogenet Evol **37,** 15-24.

179. **Kersulyte, D., Mukhopadhyay, A. K., Velapatino, B., Su, W., Pan, Z., Garcia, C., Hernandez, V., Valdez, Y., Mistry, R. S., Gilman, R. H.*, et al.*** 2000. Differences in genotypes of *Helicobacter pylori* from different human populations. J Bacteriol **182,** 3210-3218.

180. **Ogura, M., Perez, J. C., Mittl, P. R., Lee, H. K., Dailide, G., Tan, S., Ito, Y., Secka, O., Dailidiene, D., Putty, K.*, et al.*** 2007. *Helicobacter pylori* evolution: lineage- specific adaptations in homologs of eukaryotic Sel1-like genes. PLoS Comput Biol **3,** e151.

181. **Shriner, D., Nickle, D. C., Jensen, M. A., & Mullins, J. I.** 2003. Potential impact of recombination on sitewise approaches for detecting positive natural selection. Genet Res **81,** 115-121.

182. **Reed, F. A. & Tishkoff, S. A.** 2006. Positive selection can create false hotspots of recombination. Genetics **172,** 2011-2014.

183. **Montecucco, C., E. Papini, M. de Bernard, J.L Telford, and R. Rappuoli** 1999. in *The comprehensive sourcebook of bacterial protein toxins* (In J. E. Alouf, and J. H. Freer (ed.), , San Diego, California), pp. 264-286.

184. **Domanska, G., Motz, C., Meinecke, M., Harsman, A., Papatheodorou, P., Reljic, B., Dian-Lothrop, E. A., Galmiche, A., Kepp, O., Becker, L.*, et al.*** *Helicobacter pylori* VacA toxin/subunit p34: targeting of an anion channel to the inner mitochondrial membrane. PLoS Pathog **6,** e1000878.

185. **Kimura, M., Goto, S., Wada, A., Yahiro, K., Niidome, T., Hatakeyama, T., Aoyagi, H., Hirayama, T., & Kondo, T.** 1999. Vacuolating cytotoxin purified from *Helicobacter pylori* causes mitochondrial damage in human gastric cells. Microb Pathog **26,** 45-52.

186. **Menaker, R. J., Ceponis, P. J., & Jones, N. L.** 2004. *Helicobacter pylori* induces apoptosis of macrophages in association with alterations in the mitochondrial pathway. Infect Immun **72,** 2889-2898.

187. **Willhite, D. C. & Blanke, S. R.** 2004. *Helicobacter pylori* vacuolating cytotoxin enters cells, localizes to the mitochondria, and induces mitochondrial membrane permeability changes correlated to toxin channel activity. Cell Microbiol **6,** 143-154.

188. **Willhite, D. C., Cover, T. L., & Blanke, S. R.** 2003. Cellular vacuolation and mitochondrial cytochrome c release are independent outcomes of *Helicobacter pylori* vacuolating cytotoxin activity that are each dependent on membrane channel formation. J Biol Chem **278,** 48204-48209.

189. **Papini, E., de Bernard, M., Milia, E., Bugnoli, M., Zerial, M., Rappuoli, R., & Montecucco, C.** 1994. Cellular vacuoles induced by *Helicobacter pylori* originate from late endosomal compartments. Proc Natl Acad Sci U S A **91,** 9720-9724.

190. **Massari, P., Manetti, R., Burroni, D., Nuti, S., Norais, N., Rappuoli, R., & Telford, J. L.** 1998. Binding of the *Helicobacter pylori* vacuolating cytotoxin to target cells. Infect Immun **66,** 3981-3984.

191. **McClain, M. S., Schraw, W., Ricci, V., Boquet, P., & Cover, T. L.** 2000. Acid activation of *Helicobacter pylori* vacuolating cytotoxin (VacA) results in toxin internalization by eukaryotic cells. Mol Microbiol **37,** 433-442.

192. **Ricci, V., Sommi, P., & Romano, M.** 2000. The vacuolating toxin of *Helicobacter pylori*: a few answers, many questions. Dig Liver Dis **32 Suppl 3,** S178-181.