INVESTIGATING THE EFFECTS OF A CYTOSINE SPECIFIC METHYLTRANSFERASE ON GENE EXPRESSION IN *HELICOBACTER PYLORI*

by

BOWEN MENG

(Under the Direction of Timothy R. Hoover)

ABSTRACT

Helicobacter pylori is a worldwide pathogen currently estimated to be actively residing in about half of the world's population. *H. pylori* strains are remarkable for the number of DNA methyltransferases encoded in their genomes, and it is becoming increasingly evident that DNA methylation has an important role in the gene expression. In this study, I characterized a C-5 cytosine DNA methyltransferase (M.Hpy99III) that is conserved in *H. pylori* strains and converts GCGC sites to $G^{m5}CGC$. A M.Hpy99III-deficient mutant in *H. pylori* G27 displayed growth and motility defects. A strong positional bias for GCGC motifs in the -13 region of *H. pylori* promoter was observed. Expression of green fluorescent protein (*gfp*) reporter genes constructed with selected GCGC-containing promoters indicated activities of some of these promoters was reduced in the M.Hpy99III-deficient mutant, which indicated many of the GCGC-containing promoters in *H. pylori* may be regulated by DNA methylation.

INDEX WORDS: *Helicobacter pylori*, Methylation, Cytosine, GFP, Gene expression, Motility, Growth

INVESTIGATING THE EFFECTS OF A CYTOSINE SPECIFIC METHYLTRANSFERASE ON GENE EXPRESSION IN *HELICOBACTER PYLORI*

by

BOWEN MENG

B.S., Virginia Polytechnic Institute and State University, 2013

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2021

INVESTIGATING THE EFFECTS OF A CYTOSINE SPECIFIC METHYLTRANSFERASE ON GENE EXPRESSION IN *HELICOBACTER PYLORI*

by

BOWEN MENG

Major Professor:	Timothy R. Hoover
Committee:	Diana M. Downs
	Robert J. Maier
	Anna C. Karls

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2021

DEDICATION

This work is dedicated to my parents and sister for their unmatched support and love throughout life and especially my graduate school years. It is with your guidance that I was able to foster my curiosity in science and teaching.

ACKNOWLEDGEMENTS

I would like to acknowledge my major professor Dr. Timothy Hoover for his tireless guidance and feedback over the many years of my graduate career. Thank you for being patient with me and giving me a chance to pursue my passions. I would also like to thank the other members of my committee, Dr. Robert Maier, Dr. Diana Downs, and Dr. Anna Karls for their critical feedback towards my research and their suggestions on how to proceed with my projects. I would like to thank all the labs in the department that allowed me to use equipment, especially the Maier lab that allowed me to use their incubators and biosafety hoods for my research. I would like to thank Dr. Stéphane Benoit and all present and past members of the Maier and Hoover labs for teaching me how to work with *Helicobacter pylori*. From a teaching perspective, I am grateful to the department as well as Dr. Jennifer Walker, Dr. Francine Scott, and Dr. Elizabeth Ottesen for allowing me to serve as instructor of record for a total of five semesters and for being role models on how to teach collegiate microbiology classes. Thank you to all my family, friends, and colleagues for helping me get through the rigors of graduate school.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Discovery of *Helicobacter pylori*

Helicobacter pylori is a Gram-negative bacterium that belongs to the phylum Campylobacterota. This novel phylum was created since the former class, *Epsilonproteobacteria,* was found to differ phylogenetically from other classes of the phylum Proteobacteria (1, 2). In 1984, Barry Marshall and Robin Warren were the first to be able to culture this unknown bacillus from gastroscopies of the antral mucosa of 100 patients (3). The pair initially observed that the bacteria were not morphologically similar to those is the genus *Campylobacter* due to the presence of sheathed flagella. As such, they named the newly discovered bacilli as "*Campylobacter pyloridis*" (4). The organism was renamed "*Campylobacter pylori*" due to *pyloridis* being grammatically incorrect, and then given its own genus in 1989 when it was renamed to "*Helicobacter pylori*" (4, 5). This was mainly due to phenotypic differences observed in *H. pylori* such as differences in cellular fatty acids, as well as 16S ribosomal RNA sequencing (5, 6).

Despite the isolation from actual patients suffering gastritis, the scientific community remained unconvinced that *H. pylori* had any definitive linkage to gastritis. There was a belief that *H. pylori* was not the causative agents of gastritis, but rather an opportunistic commensal (7). Determined, Barry Marshall fulfilled Koch's postulates using himself as the healthy individual to become infected. After drinking a culture of *H. pylori*, Marshall developed symptoms within two weeks and gastric biopsies confirmed

the presence of *H. pylori* (7). Marshall and Warren were eventually awarded the 2005 Nobel Prize in Physiology for their efforts in discovering *H. pylori* and demonstrating its role in peptic ulcer disease and other gastrointestinal ailments.

While *H. pylori* infection commonly occurs in the stomach, other *Helicobacter* species, such as *Helicobacter hepaticus,* infect the intestine and the liver of mice (8). *Helicobacter* species also have relatively specific host ranges as *H. pylori* primarily only infects humans. Other species such as *H. felis* infect cats and dogs (9, 10).

Diagnosis and Treatment of *H. pylori*

H. pylori is a medically relevant pathogen due to its presence and involvement in gastrointestinal diseases such as gastritis, peptic ulcers, and even gastric cancer (11). Detection of *H. pylori* can be achieved in a variety of techniques that encompass both invasive and non-invasive methods. The invasive techniques include endoscopies and the rapid urease test (12). Since endoscopy tends to be very expensive and often very risky to the patient, non-invasive procedures such as the urea breath test and stool antigen test are often the preferred method to detect *H. pylori* infection (12, 13). The urea breath test involves a patient drinking a solution of ^{13}C or ^{14}C labeled urea and measuring the production of CO_2 over time (12, 14). The stool antigen test involves using either polyclonal or monoclonal antibodies to detect *H. pylori* antigen in stool samples (12, 13). Combined, the two non-invasive tests offer reliable indication of *H. pylori* infection.

Once *H. pylori* is diagnosed, the standard treatment for *H. pylori* during the last few decades consists of a proton pump inhibitor, amoxicillin, and clarithromycin for 14 days (15). Variations of this treatment include using bismuth, tetracycline, and metronidazole. This triple therapy once boasted a respectable 90% eradication rate, but

the rise of antibiotic resistant strains of *H. pylori*, in particular to clarithromycin, metronidazole, and levofloxacin has reduced the eradication rate to <70% and the triple therapy is no longer recommended (15).

New strategies to treat *H. pylori* have been proposed and studied within the recent decade. Sequential therapy involves administering omeprazole and amoxicillin for 5 days then omeprazole, clarithromycin, and tinidazole for the next five days. It was shown to be very effective in treating *H. pylori* infections (16). Concomitant therapy involves simultaneously administering three antibiotics alongside a proton pump inhibitor. Studies showed that on average, concomitant therapy was comparable to sequential therapy, although one of the major downsides to concomitant therapy was the increased risk of side effects, likely due to the combined antibiotic usage (17-19). A third therapy termed hybrid therapy combines the sequential and concomitant therapies together and resulted in near complete eradication rates in one study (20). Combined, these three new therapies have been shown to be effective in replacing the failing triple therapy regimen.

Metabolism and Colonization Requirements

H. pylori is classified as a microaerophile and is auxotrophic for several amino acids (21, 22). While complex media are suitable for *H. pylori* growth, supplements such as fetal bovine serum or horse serum are often required for optimal growth (23, 24). Since it is classified as a microaerophile, *H. pylori* requires oxygen concentrations between 2% - 10% for optimal growth. As atmospheric oxygen levels are approached (roughly 21%) the viability of the organism dramatically decreases (21, 22).

In the host organism, successful colonization of *H. pylori* requires the production of flagella as well as the production of urease, which helps buffer the organism against

the extremely acidic conditions of the stomach by converting urea into ammonia to raise the pH of the surrounding area (11, 25, 26). *H. pylori* prefers a pH range between 5 and 7 for optimal growth, therefore it is no surprise that approximately 10% of the total cellular protein is urease (21, 27). The importance of urease is further corroborated by the observation that urease deficient mutants fail to colonize the stomachs of gnotobiotic piglets (25). While the main role of urease is to catalyze the conversion of urea to ammonia and carbon dioxide, the pH elevation it provides to *H. pylori* is crucial. Not only does this allow *H. pylori* to survive in a more neutral pH environment, the hydrolysis of urea has been shown to reduce the viscoelasticity of the gastric mucin so that the organism is able to swim within the mucous layer of the stomach (26, 28).

The flagella are required for motility of *H. pylori* and allows for chemotaxis of the bacterium towards the gastric epithelium where the pH is closer to neutral (29, 30). In experiments with gnotobiotic piglets, motile versions of *H. pylori* 26695 were able to persist for up to 21 days post-inoculation, whereas non-motile *H. pylori* mutants typically only persisted 6 days post-inoculation (31). In another study, motile versions of *H. pylori* 26695 were found to maintain colonization of the stomach of the piglets 87 days post-inoculation (32). In addition to surviving longer in the host, the motile *H. pylori* 26695 colonized the stomach at greater cell numbers (31).

Fully formed flagella are crucial for proper colonization, but are not completely sufficient for *H. pylori* to efficiently infect the host organism (33, 34). Other genes, such as *motB* (encodes a component of the flagellar motor stator), are essential for proper motility of *H. pylori*. In the case of the *motB* mutant, cells produced fully formed flagella but were unable to swim and had reduced success in colonizing the murine mucosa (34,

35). Chemotaxis related genes such as *cheW*, *cheY*, and *cheA* also contribute to the ability of *H. pylori* to establish an infection in the stomach. While chemotactic deficient *H. pylori* mutants colonize the host, they do so at reduced levels. Once infection is established though, the chemotactic deficient mutants maintain infection at levels similar to the corresponding wild-type parental strain provided they do not have to compete with the parental strain for host colonization (36).

Flagellar Assembly and Motility

Flagella biosynthesis in *H. pylori* is an intricately controlled process, which involves three different sigma factors and other regulatory genes (37). Synthesis of the flagella begins with the assembly of the basal body, and transcription of the genes encoding these proteins (early genes), requires the housekeeping sigma factor, σ^{80}. In contrast to *Salmonella enterica* and *Escherichia coli*, which possess a master regulator (FlhCD) that initiates the transcriptional hierarchy that governs expression of the flagellar genes, no master regulator has been identified thus far in *H. pylori* (37, 38). The basal body houses the flagellar motor and a flagellar protein export apparatus that transports axial components of the flagellum (e.g., rod, hook and filament proteins) across the cell membrane and into the lumen of the nascent flagellum. The core components of the flagellar motor include the MotA/MotB stator complex, rotor (MS- and C-rings), driveshaft (rod) and bushings (P- and L-rings) (37, 39). The *H. pylori* motor contains 18 torque-generating stator complexes, making it the largest bacterial flagellar motor described to date. The *H. pylori* motor contains several discs and rings not present in the archetypical flagella of *E. coli* and *Salmonella enterica* serovar Typhimurium (*S.*

Typhimurium), and these embellishments may have roles in generating torque or mitigating the effects of the high torque of the motor on the cell envelope (40, 41).

Once the basal body is built, the FlgS/FlgR two-component system responds to an unknown cellular cue and stimulates transcription of genes encoding the hook, hook-associated proteins, a minor flagellin, as well as other proteins required for the next stage of flagellum assembly. Transcription of these middle flagellar genes is dependent upon the alternative sigma factor σ^{54} (37, 42, 43). Upon sensing the cellular cue, FlgS undergoes autophosphorylation and then transfers the phosphate to FlgR, converting FlgR to an active form that stimulates transcription of the σ^{54} dependent genes (37, 42, 43). Although the specific cellular cue that is sensed by FlgS to initiate the signal transduction pathway has yet to be identified, FlgS binds a peptide corresponding to the N-terminal region of FlhA (a component of the flagellar protein export apparatus) with high affinity (K_d ~20 nM) (44). There are nine FlhA subunits in the flagellar protein export apparatus, and it is hypothesized that binding of FlgS monomers to the N-terminal regions of the FlhA nonamer in the export apparatus facilitates dimerization of FlgS to activate its autokinase activity (44).

Transcription of genes encoding the major flagellin (*flaA*) and other proteins required in the later stages of flagellum assembly is dependent on the alternative sigma factor σ^{28} (37). The activity of σ^{28} is negatively regulated by the anti-sigma factor FlgM (37, 39). Once the rod and hook form, FlgM inhibition is alleviated from σ^{28}, which allows the filament to be produced (37). However, unlike FlgM in *S.* Typhimurium, *H. pylori* FlgM is not exported from the cell (45). This export phenotype is thought to be due to the absence of a short amino acid region on the N-terminal domain of *H. pylori* FlgM.

Despite missing this short amino acid region, it is still able to interact with and inhibit *S.* Typhimurium σ^{28}, although export of *H. pylori* FlgM in *S.* Typhimurium is difficult (45, 46). There is evidence that the inhibitory effect of *H. pylori* FlgM on σ^{28} is alleviated by interactions with the C-terminal region of FlhA (46).

Knockouts of key genes such as *flhA*, *rpoN*, and *flgM* all have significant effects on the transcription of flagellar genes. When *flhA* and *rpoN* are absent from the cell, expression of σ^{54} dependent genes is severely downregulated in *H. pylori* strains N6 and 88-3887 (47). Curiously, in a double knockout of *flhA* and *flgM* in *H. pylori* N6, the expression of σ^{28} dependent genes is slightly upregulated as expected, but expression of σ^{54}-dependent genes is also slightly upregulated. This upregulation of σ^{54}-dependent genes expression is unexpected because the absence of *flhA* should downregulate the expression of σ^{54}-dependent genes as shown in the single *flhA* strain knockout (47). Immediately downstream of *flgM* is a cytosine-specific methyltransferase (MTase) (M. *Hpy*AVIII), which is the focus of my study. My analysis of the proteome of the M. *Hpy*AVIII mutant in *H. pylori* G27 suggested that the levels of a number of proteins involved in protection from reactive oxygen species (ROS) or metal homeostasis were diminished relative to the wild-type parental strand. Therefore, the following sections briefly review mechanisms for protection against ROS and metal homeostasis in *H. pylori*.

Reactive Oxygen Species Damage

ROS such as hydrogen peroxide (H_2O_2), the superoxide anion (O_2^-), and the hydroxyl radical (OH^-) are most commonly produced during the process of aerobic respiration (48). Accumulation of ROS in the cell can lead to permanent damage to

proteins and DNA, and so most cells have adapted specific defense mechanisms to combat potential ROS damage (48). Hydrogen peroxide (H_2O_2) is produced when superoxide dismutase catalyzes the dismutation or disproportionation of superoxide anions into molecular oxygen and hydrogen peroxide (49). Hydrogen peroxide is then converted into molecular oxygen and water through the enzymatic properties of catalase. Oscar Leow first identified catalase in 1900 and showed that catalase seemed to be ubiquitous in plants and animals (50). The ubiquity of the enzyme across multiple forms of life gives us insight on how organisms deal with the constant threat of ROS.

When *H. pylori* establishes colonization in the host organism, the innate immune system of the host is activated. It is believed that neutrophils produce the majority of the ROS that *H. pylori* has to endure (51). Superoxide anions and hydrogen peroxide are amongst the first ROS produced during a response, but over time more toxic ROS such as hypochlorous acid (HOCl) are also produced (51, 52). However, *H. pylori* has many defense mechanisms against the host immune system, and the inability of the host to completely eliminate *H. pylori* results in the overproduction of ROS, which subsequently leads to gastric tissue damage (51, 52).

In *H. pylori*, catalase (KatA) is estimated to account for 4-5% of the total cellular proteins (53). *H. pylori* SS1 does not need catalase to be able to initially colonize the murine host, however, the absence of catalase hinders the organism's ability to maintain a persistent infection (54). This is in contrast to the activity of superoxide dismutase, which has been shown to be nearly essential for the initial colonization in the murine host (55).

Interest in catalase is not only limited to its catalytic function. Catalytically defective catalase is still able to defend the cell from oxidative damage caused by HOCl

(56). Targeting catalase through the use of a DNA vaccine has also been shown to confer protective immunity against *H. pylori* in a murine model (57).

Heavy metal importance in *H. pylori*

Heavy metals such as nickel, cobalt, iron, and copper play very important roles in *H. pylori* (58). In particular, nickel is required for the proper function of urease and hydrogenease in *H. pylori*. When proteins regulating nickel storage were inactivated in *H. pylori*, colonization was decreased (58, 59).

Cobalt is crucial for nitrogen metabolism since it serves as a co-factor to RocF, which is an arginase enzyme in *H. pylori* (58, 60). However, cobalt has also been shown to be important for the ability of *H. pylori* to survive the innate immune response. *rocF* enables *H. pylori* to significantly reduce the amount of nitric oxide (NO) produced the by host immune system, which thereby allows it to survive. Absence of *rocF* results in bacterial death due to production of NO (58, 61).

Iron is able to serve as a cofactor for several enzymes and it is also involved in the electron transport chain, but proper regulation of iron concentration is the cell is crucial since an overabundance of iron can leader to cellular damage via ROS from the Fenton reaction (62, 63). Uptake of ferrous iron (Fe^{2+}) plays an important role for the successful colonization of *H. pylori*. Strains of *H. pylori* lacking *feoB* were completely unable to successfully colonize the mice (58, 64).

Copper serves as a co-factor for many different enzymes but also has roles in the colonization of *H. pylori* (58, 65). The presence of copper enhances the ability of *H. pylori* to bind to epithelial cells to initiate infection. While absence of copper did not

abolish colonization, it was diminished as compared to conditions where copper was present (65).

Restriction-modification systems in *H. pylori*

H. pylori is a naturally transformable organism that can take up exogenous DNA without chemical treatment or electroporation (66, 67). Due to the ability to uptake DNA, it would benefit *H. pylori* to be able to defend against incorporating potentially undesirable genes or other DNA elements into its genome. In bacteria, restriction modification systems (R-M systems) consist of genes encoding an endonuclease and DNA MTase that allow the bacterial cell to degrade invading foreign DNA with the endonuclease, but protect the bacterium's own DNA by methylating the sequence recognized by the endonuclease to prevent it from cutting the sequence (68). This can be thought of as a primitive bacterial immune system as the endonucleases targets DNA that is identified as non-self. However, there is a significant bias in the number of R-M systems across bacterial species (69). Despite this, there seems to be a positive correlation between genome size and total R-M systems, with larger genomes containing more R-M systems (69).

These endonucleases can serve as a barrier to successfully transforming *H. pylori* since they often degrade exogenous DNA before it integrates into the cell's genome (70). Donahue and co-workers found a solution to this conundrum by incubating the genetic material to be transformed with cell-free extracts from the *H. pylori* strain they were transforming along with S-adenosylmethionine (SAM), the methyl donor for DNA methylation reactions. This greatly increased the transformation efficiency by methylating the donor DNA *in vitro* (70). Alternatively, experiments that deleted Type II

restriction endonucleases found that competence was greatly increased in the restriction deficient strains as compared to the wild type (71).

R-M systems are divided into four different types, (Type I, II, III, IV) based on the proteins and subunits required for activity. Type I R-M systems contain three units (restriction, modification, and specificity). These subunits work together to act as a MTase and endonuclease, or if the R subunit is not present, the complex becomes just a MTase. Type II R-M systems are the simplest systems and are comprised of two units (restriction and modification) that act separately and do not complex together. Type III R-M systems have a M and a R subunit that can complex together to exhibit restriction and methylation activity. The M subunit can also independently act as just a MTase (69, 72). The final known R-M system, Type IV, has been characterized in *E. coli*. Eco57I appears to have a R subunit that is capable of exhibiting both restriction and methylation and a M subunit that is only capable of methylation (72).

DNA methylation is found through all domains of life and is characterized by the catalytically driven process of transferring a methyl group from SAM onto adenine or cytosine. In bacteria, 5-methylcytosine (5mC), N^4-methylcytosine (4mC) and N^6-methyladenine (6mA) has been observed. 4mC is specific to bacteria while CpG methylation, a form of 5mC, is specific to higher eukaryotes (73, 74) (Fig. 1.1).

H. pylori exhibits all three types of methylation in its genome using an impressive number of MTases (75, 76). Complete methylome analysis studies of *H. pylori* strains 26695 and J99 identified 17 and 22 unique methylated motifs in in these strains, respectively (75). The specific recognition sites of these methyltranferases were identified using restriction endonuclease assays and PacBio Single Molecule, Real-Time

(SMRT) sequencing (75, 77, 78). Although most of the identified MTases in *H. pylori* belong to R-M systems, in some strains, the corresponding endonuclease is either inactive or absent, but the MTase is still active (77, 78). According to the REBASE database provided by New England Biosciences, many of the MTase found in *H. pylori* are conserved across strains (76). The following sections briefly discuss the different types of DNA methylation in bacteria and how these types of DNA methylation impact gene expression.

6mA Methylation. In *Caulobacter crescentus*, a 6mA MTase CcrM, which methylates the adenine nucleotide in the 5' –GANTC – 3' motif, is critical during chromosomal replication and cell division (79). Overexpression of CcrM results in morphological abnormalities and deletion of *ccrM* is not possible (79, 80). Cell viability is highly correlated to proper expression of *ccrM* as shifting from a xylose-based medium to a glucose-based medium when *ccrM* is under the control of a xylose-inducible promoter (*xylA*) results in cell death (80). *E. coli* DNA adenine methylase (*dam*), which methylates the adenine nucleotide in the 5' –GATC – 3' motif, is critical during the mismatch repair process. Methylation allows the mismatch repair system to correctly recognize the parental strand and excise the incorrect base from the opposite strand (81, 82). Unlike *ccrM* mutants, *dam* mutants are not lethal to the cell unless they are combined with other mutations affecting the SOS response system, such as *recA* (81).

In *H. pylori*, several 6mA MTases have been characterized. M.*Hpy*AI methylates the adenine nucleotide in the 5' –CATG – 3' motif and has been shown to affect the expression of the *dnaK* operon (83-85). Abolishing activity of M.*Hpy*AI results in decreased viability during stationary phase (85). M. *Hpy*AIV is another adenine MTase in

H. pylori that recognizes and methylates the same motif as *ccrM,* 5' –GANTC – 3' (77, 79, 80, 83). Unlike *ccrM*, M. *Hpy*AIV is not essential for the cell, but loss of the MTase did result in a statistically significant decrease in transcription of *katA*. (86).

4mC Methylation. In *Caldicellulosiruptor bescii*, a m4C MTase affects the ability of the organism to undergo transformation (87). The study showed that methylation by the enzyme M.CbeI of exogenous DNA was required for successful transformation. (87). A similar observation was found in *H. pylori* in regards to the M2.*hpy*AII MTase (88). In *H. pylori*, a knockout of M2.*hpy*AII decreased the ability of *H. pylori* 26695 to successfully incorporate exogenous DNA via transformation (88). M2.*Hpy*AII was also shown to be important in the regulation of *H. pylori* virulence against human gastric cells as loss of the m4C MTase results in a dramatic reduction of *H. pylori*-induced apoptosis of human cells (88).

5mC Methylation. 5mC has been shown to be very crucial for proper mammalian development. Mutation of the methylase responsible for CpG methylation results in embryonic lethality as reported by Li and co-workers (89). However, cytosine methylation also has its drawbacks as one of the most frequent mutations in human DNA (cytosine to thymine) is thought to be mediated by the methylation process (90, 91). Methylated cytosines are able to spontaneously deaminate directly into thymine, whereas thymine goes through an uridine intermediate and is changed to a thymine after DNA replication if the uridine is not corrected (91). The rate of deamination of methylated cytosines has also been shown to be approximately 2-3 faster than unmethylated cytosines (92).

In humans, CpG methylation has been studied extensively and has been shown to be an important factor in gene expression and regulation. In fact, over 70% of promoters in the human genome contain high CpG content (93). Since mammalian genetic material is associated with chromatin, methylation can alter chromatin structure, which then leads to alterations in gene expression. (94)

Although the role of 5mC in bacterial systems is not as clear cut as it is in eukaryotic systems, several groups have shown potential regulatory functions for 5mC. Kahramanoglou *et al*. discovered in *E. coli* K12 MG1655 the internal cytosine of the 5'– CCWGG – 3' motif was fully converted to 5mC during stationary phase, but the extended 5'– CCCWGG – 3' motif was only partially methylated during exponential growth (95). The group also discovered that in a 5mC-deficient mutant expression of several genes responsible for the stress response are upregulated, including *rpoS* (95).

Chao *et al*. characterized the VchM MTase in *Vibrio cholerae*. VchM modifies the first cytosine in the 5'– RCCGGY – 3' motif to 5mC (96, 97). Absence of this MTase significantly hinders the ability of *V. cholera*e in its ability to compete with wild-type strains in competition assays. VchM deficient strains also appear to have a slight growth defect, which is corroborated by its failure to colonize mice at wild-type levels (97).

In *H. pylori*, m5C has been shown to be important for proper gene expression. M. *Hpy*AVIBM converts the first cytosine in the 5'– CCTC – 3' motif to 5mC and inactivation of this MTase results in different effects depending on the strain (77, 78, 83, 98). In *H. pylori* SS1, absence of M. *Hpy*AVIBM methylation results in increased motility as compared to the wild type. Conversely, in *H. pylori* AM5, loss of the MTase

results in the increase in expression of genes associated with virulence, such as *cagA* and *vacA* (98).

M. *Hpy*AVIII is a 5mC MTase that methylates the first cytosine in the 5'– GCGC – 3' motif (77, 78, 83). While virtually all *H. pylori* strains sequenced to date contain M. *Hpy*AVIII, only around 10% of the ~500 *H. pylori* sequenced strains appear to encode the cognate restriction endonuclease (99). The majority of strains containing the endonuclease are of African descent, while the few strains that are not directly related to the African strains are thought to have acquired the endonuclease though recombination events (99, 100). Absence of the endonuclease does not result in a nonfunctional MTase, suggesting an important role for this 5mC MTase in *H. pylori* gene regulation (77, 78, 100).

In *H. pylori* J99, loss of M. *Hpy*AVIII results in several distinct phenotypes. *H. pylori* J99 has a substantial growth defect that is likely due to stressful conditions as evidenced by the abundance of cells in the coccoid state very early in liquid cultures (100). *H. pylori* is thought to transition to coccoid form only when it encounters unfavorable environmental conditions for growth. Once transformed to coccoid form, the cells then become unculturable and cannot be revived regardless of growth conditions (101). This transition to coccoid form could explain why *H. pylori* J99 M. *Hpy*AVIII-deficient mutant has such a significant growth defect since the same phenomenon was not observed in the wild type parental strain (100).

The J99 M. *Hpy*AVIII mutant also exhibited increased susceptibility to copper. This is thought to be due to decreased expression of *crdR*, which is part of the CrdRS two-component system responsible for protecting the bacterium against copper stress

(100, 102). The presence of a 5'– GCGC – 3' motif within 50 base pairs upstream of the transcriptional start site provides evidence for M. *Hpy*AVIII methylation to be a key factor of regulation for *crdR* (100). In addition to *crdR*, one of the genes associated with iron transport, *feoB*, was also found to be downregulated in the M. *Hpy*AVIII mutant (100). Natural competence was also affected in the J99 M. *Hpy*AVIII mutant resulting in a reduction in the transformation frequency with a chloramphenicol resistance gene (100).

Site-directed mutations of GCGC regions near *jhp0832*, which is part of a toxin-antitoxin system in *H. pylori*, showed that GCGC motifs have an effect on the transcription of this operon. *jhp0832* encodes an antitoxin and is homologous to HP0893, which has been shown to be an antitoxin in *H. pylori* strain 26695 (100, 103). Of the three GCGC motifs associated with *jhp0832* that were changed to GAGC, only the motif that overlapped the -10 promoter region had a measurable effect on the transcription of the operon. When the other two GCGC motifs were altered, there was no decrease in gene expression unless the GCGC motif at the -10 position was also altered (100).

Research Summary

The goal of my study was to understand how methylation of GCGC motifs by the MTase M. *Hpy*AVIII and its homologs affects gene expression in *H. pylori*. In Chapter 2, I report on bioinformatic studies that examined the occurrence GCGC motifs within predicted promoter regions of the *H. pylori* 26695 genome. Of all the methylation motifs observed in strain 26695, only the GCGC motifs had a strong positional bias. The predicted promoter regions were identified in a transcriptome analysis that employed a differential RNA-seq approach by Sharma and co-workers (104). In the differential RNA-

seq assay, isolated RNA was treated with terminator exonuclease, which degrades 5'-phosphate RNA but not 5'-triphosphate RNA, to discriminate primary transcripts from processed transcripts that have a 5'-phosphate. A total of 211 promoter regions contained a GCGC motif. Interestingly, our mining of the transcriptome dataset revealed the GCGC motifs in the promoter regions displayed a strong bias for -13 position relative to the transcriptional start site. Twenty-two of the promoters that had a GCGC motif in the -13 region were associated with genes that were differentially regulated in the *H. pylori* J99 JHP1050 (M.*Hpy*AVIII homolog) mutant. These findings indicated the activities of these 22 GCGC-containing promoters from *H. pylori* J99 may be influenced directly by methylation of the GCGC motif. Expression of green fluorescent protein reporter genes constructed with various GCGC-containing promoters in *H. pylori* G27 wild type and a mutant in which the M.*Hpy*AVIII homolog had been inactivated suggested the activities of some of the promoters are either stimulated or inhibited by methylation of the GCGC motif.

While the study by Estibariz and co-workers investigated the effects of M. *Hpy*AVIII across several different strains of *H. pylori*, my studies expanded on the previous study to investigate how inactivation of M.*Hpy*AVIII homologs in other *H. pylori* strains (G27, ATCC 43504, B128, and X47-2AL) affected growth in liquid cultures, which is reported in Chapter 2. (100). In addition, I inactivated M.*Hpy*AVIII in *H. pylori* 26695, which had been done in the previous study by Estibariz and co-workers (100). As observed for *H. pylori* J99, loss of the M.*Hpy*AVIII homolog in *H. pylori* strains G27, B128 and 26695 resulted in significant decreases in growth rate, but not in *H. pylori* strains ATCC 43504 or X47-2AL. The reduced growth rate of the M.*Hpy*AVIII

mutant in *H. pylori* 26695 was not consistent with a previous report (100), which may be due to genotypic variations between the *H. pylori* 26695 strain in our lab and that of the lab of Estibariz and co-workers.

In addition to examining how loss of M.*Hpy*AVIII homologs in various *H. pylori* strains impacted growth, we examined how loss of the M.*Hpy*AVIII homolog in *H. pylori* G27 affected the proteome of *H. pylori* G27, as well as sensitivity to hydrogen peroxide and motility. The rationale for examining sensitivity to hydrogen peroxide and motility in the *H. pylori* G27 M.*Hpy*AVIII mutant were: (i) the initial proteomic analysis of the M.*Hpy*AVIII mutant suggested levels of catalase (KatA) were reduced in the mutant; and (ii) the gene encoding M.*Hpy*AVIII homologs in various *H. pylori* strains is associated with genes involved in flagellum biosynthesis. Consistent with the results from the initial proteomic analysis, the M.*Hpy*AVIII mutant displayed increased sensitivity to hydrogen peroxide. The *H. pylori* G27 M.*Hpy*AVIII mutant was less motile in the soft agar medium that we normally use for assessing motility, with the diameters of the swim halos of the mutant being ~40-50% of those of the parental strain. The reduced motility phenotype was dependent on the presence of $FeSO_4$ in the medium, and curiously, was only observed for the *H. pylori* G27 M.*Hpy*AVIII mutant, as the motilities of the M.*Hpy*AVIII mutants in *H. pylori* strains B128 or ATCC 43504 were indistinguishable from their parental strains and did not display significant differences in the presence or absence of added $FeSO_4$. It is not clear if the reduced motility of the *H. pylori* G27 M.*Hpy*AVIII mutant in the presence of $FeSO_4$ is due to an altered chemotaxis response or a deficiency in flagellar motor function. Growth curve data suggests the altered motility

is not due to a growth defect as the doubling times of the mutant with and without $FeSO_4$ were essentially the same.

The results of my research on M.*Hpy*AVIII and its homologs in *H. pylori* strains provides new information on the role this 5mC MTase plays in gene regulation. In addition, the results from my studies provide insights into why these MTases have been retained in *H. pylori* strains in the absence of a cognate RE. Finally, the work presented here open new avenues for exploring the potential roles of other DNA MTases in *H. pylori* gene regulation and physiology.

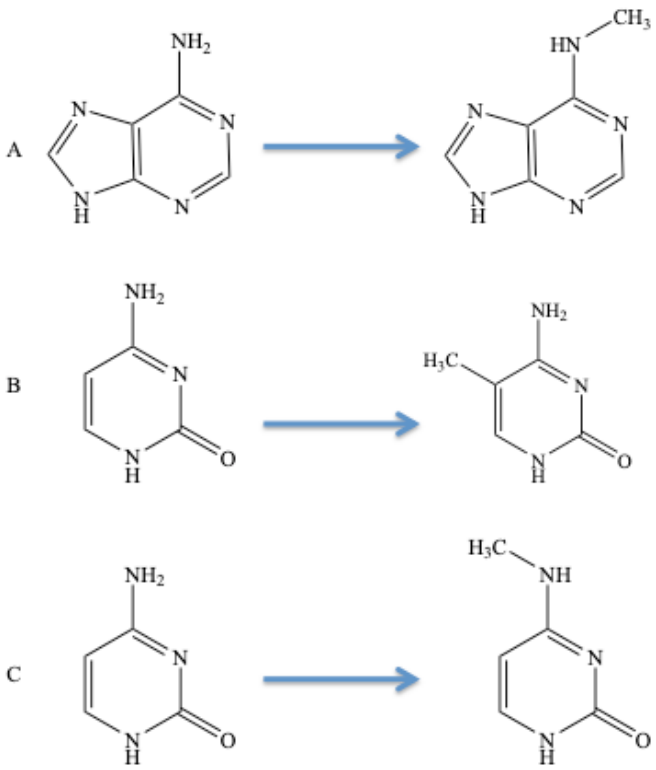

Figure 1.1: Diagram of un-modified nucleotides and their methylated counterparts. Un-modified nucleotides are to the left of the arrow. (A) 6mA methylation, (B) 5mC methylation, (C), 4mC methylation

References

1. **Waite DW, Vanwonterghem I, Rinke C, Parks DH, Zhang Y, Takai K, Sievert SM, Simon J, Campbell BJ, Hanson TE, Woyke T, Klotz MG, Hugenholtz P.** 2017. Comparative Genomic Analysis of the Class *Epsilonproteobacteria* and Proposed Reclassification to Epsilonbacteraeota (phyl. nov.). Front Microbiol **8:**682.
2. **Waite DW, Vanwonterghem I, Rinke C, Parks DH, Zhang Y, Takai K, Sievert SM, Simon J, Campbell BJ, Hanson TE, Woyke T, Klotz MG, Hugenholtz P.** 2018. Erratum: Addendum: Comparative Genomic Analysis of the Class *Epsilonproteobacteria* and Proposed Reclassification to Epsilonbacteraeota (phyl. nov.). Front Microbiol **9:**772.
3. **Marshall BJ, Warren JR.** 1984. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. Lancet **1:**1311-1315.
4. **Marshall BJ, Goodwin CS.** 1987. Notes: Revised Nomenclature of *Campylobacter pyloridis*. International Journal of Systematic Bacteriology **37:**68-68.
5. **Goodwin CS, Armstrong JA, Chilvers T, Peters M, Collins MD, Sly L, McConnell W, Harper WES.** 1989. Transfer of *Campylobacter pylori* and *Campylobacter mustelae* to *Helicobacter* gen. nov. as *Helicobacter pylori* comb. nov. and *Helicobacter mustelae* comb. nov., Respectively. International Journal of Systematic Bacteriology **39:**397-405.

6. **Goodwin CS, McCulloch RK, Armstrong JA, Wee SH.** 1985. Unusual cellular fatty acids and distinctive ultrastructure in a new spiral bacterium (*Campylobacter pyloridis*) from the human gastric mucosa. J Med Microbiol **19:**257-267.

7. **Marshall BJ, Armstrong JA, McGechie DB, Glancy RJ.** 1985. Attempt to fulfil Koch's postulates for pyloric *Campylobacter*. The Medical journal of Australia **142:**436-439.

8. **Fox JG, Ge Z, Whary MT, Erdman SE, Horwitz BH.** 2011. *Helicobacter hepaticus* infection in mice: models for understanding lower bowel inflammation and cancer. Mucosal Immunol **4:**22-30.

9. **Solnick JV, Schauer DB.** 2001. Emergence of diverse *Helicobacter* species in the pathogenesis of gastric and enterohepatic diseases. Clinical microbiology reviews **14:**59-97.

10. **Paster BJ, Lee A, Fox JG, Dewhirst FE, Tordoff LA, Fraser GJ, O'Rourke JL, Taylor NS, Ferrero R.** 1991. Phylogeny of *Helicobacter felis* sp. nov., *Helicobacter mustelae*, and related bacteria. Int J Syst Bacteriol **41:**31-38.

11. **Ernst PB, Gold BD.** 2000. The disease spectrum of *Helicobacter pylori*: the immunopathogenesis of gastroduodenal ulcer and gastric cancer. Annu Rev Microbiol **54:**615-640.

12. **Ricci C, Holton J, Vaira D.** 2007. Diagnosis of *Helicobacter pylori*: invasive and non-invasive tests. Best Pract Res Clin Gastroenterol **21:**299-313.

13. **Vaira D, Vakil N.** 2001. Blood, urine, stool, breath, money, and *Helicobacter pylori*. Gut **48:**287-289.

14. **Graham DY, Klein PD, Evans DJ, Jr., Evans DG, Alpert LC, Opekun AR, Boutton TW.** 1987. *Campylobacter pylori* detected noninvasively by the 13C-urea breath test. Lancet **1:**1174-1177.

15. **Papastergiou V, Georgopoulos SD, Karatapanis S.** 2014. Treatment of *Helicobacter pylori* infection: meeting the challenge of antimicrobial resistance. World J Gastroenterol **20:**9898-9911.

16. **Zullo A, Rinaldi V, Winn S, Meddi P, Lionetti R, Hassan C, Ripani C, Tomaselli G, Attili AF.** 2000. A new highly effective short-term therapy schedule for *Helicobacter pylori* eradication. Aliment Pharmacol Ther **14:**715-718.

17. **Kim SY, Choi DJ, Chung JW.** 2015. Antibiotic treatment for *Helicobacter pylori*: Is the end coming? World J Gastrointest Pharmacol Ther **6:**183-198.

18. **Zullo A, Scaccianoce G, De Francesco V, Ruggiero V, D'Ambrosio P, Castorani L, Bonfrate L, Vannella L, Hassan C, Portincasa P.** 2013. Concomitant, sequential, and hybrid therapy for *H. pylori* eradication: A pilot study. Clinics and Research in Hepatology and Gastroenterology **37:**647-650.

19. **Wu DC, Hsu PI, Wu JY, Opekun AR, Kuo CH, Wu IC, Wang SS, Chen A, Hung WC, Graham DY.** 2010. Sequential and concomitant therapy with four drugs is equally effective for eradication of *H pylori* infection. Clin Gastroenterol Hepatol **8:**36-41 e31.

20. **Hsu PI, Wu DC, Wu JY, Graham DY.** 2011. Modified sequential *Helicobacter pylori* therapy: proton pump inhibitor and amoxicillin for 14 days with

clarithromycin and metronidazole added as a quadruple (hybrid) therapy for the final 7 days. Helicobacter **16:**139-145.

21. **Kangatharalingam N, Amy PS.** 1994. *Helicobacter pylori* comb. nov. Exhibits Facultative Acidophilism and Obligate Microaerophilism. Appl Environ Microbiol **60:**2176-2179.

22. **Nedenskov P.** 1994. Nutritional requirements for growth of *Helicobacter pylori*. Appl Environ Microbiol **60:**3450-3453.

23. **Albertson N, Wenngren I, Sjostrom JE.** 1998. Growth and survival of *Helicobacter pylori* in defined medium and susceptibility to Brij 78. J Clin Microbiol **36:**1232-1235.

24. **Xia HX, English L, Keane CT, O'Morain CA.** 1993. Enhanced cultivation of *Helicobacter pylori* in liquid media. J Clin Pathol **46:**750-753.

25. **Eaton KA, Brooks CL, Morgan DR, Krakowka S.** 1991. Essential role of urease in pathogenesis of gastritis induced by *Helicobacter pylori* in gnotobiotic piglets. Infect Immun **59:**2470-2475.

26. **Scott DR, Weeks D, Hong C, Postius S, Melchers K, Sachs G.** 1998. The role of internal urease in acid resistance of *Helicobacter pylori*. Gastroenterology **114:**58-70.

27. **Bauerfeind P, Garner R, Dunn BE, Mobley HLT.** 1997. Synthesis and activity of *Helicobacter pylori* urease and catalase at low pH. Gut **40:**25-30.

28. **Celli JP, Turner BS, Afdhal NH, Keates S, Ghiran I, Kelly CP, Ewoldt RH, McKinley GH, So P, Erramilli S, Bansil R.** 2009. *Helicobacter pylori* moves

through mucus by reducing mucin viscoelasticity. P Natl Acad Sci USA **106:**14321-14326.

29. **Croxen MA, Sisson G, Melano R, Hoffman PS.** 2006. The *Helicobacter pylori* chemotaxis receptor TlpB (HP0103) is required for pH taxis and for colonization of the gastric mucosa. J Bacteriol **188:**2656-2665.

30. **Schreiber S, Konradt M, Groll C, Scheid P, Hanauer G, Werling HO, Josenhans C, Suerbaum S.** 2004. The spatial orientation of *Helicobacter pylori* in the gastric mucus. P Natl Acad Sci USA **101:**5024-5029.

31. **Eaton KA, Morgan DR, Krakowka S.** 1992. Motility as a factor in the colonisation of gnotobiotic piglets by *Helicobacter pylori*. J Med Microbiol **37:**123-127.

32. **Eaton KA, Morgan DR, Krakowka S.** 1990. Persistence of *Helicobacter pylori* in conventionalized piglets. J Infect Dis **161:**1299-1301.

33. **Josenhans C, Labigne A, Suerbaum S.** 1995. Comparative ultrastructural and functional studies of *Helicobacter pylori* and *Helicobacter mustelae* flagellin mutants: both flagellin subunits, FlaA and FlaB, are necessary for full motility in *Helicobacter* species. J Bacteriol **177:**3010-3020.

34. **Ottemann KM, Lowenthal AC.** 2002. *Helicobacter pylori* uses motility for initial colonization and to attain robust infection. Infect Immun **70:**1984-1990.

35. **Chun SY, Parkinson JS.** 1988. Bacterial motility: membrane topology of the *Escherichia coli* MotB protein. Science **239:**276.

36. **Terry K, Williams SM, Connolly L, Ottemann KM.** 2005. Chemotaxis plays multiple roles during *Helicobacter pylori* animal infection. Infect Immun **73:**803-811.

37. **Tsang J, Hoover TR.** 2014. Themes and Variations: Regulation of RpoN-Dependent Flagellar Genes across Diverse Bacterial Species. Scientifica (Cairo) **2014:**681754.

38. **Liu X, Matsumura P.** 1994. The FlhD/FlhC complex, a transcriptional activator of the *Escherichia coli* flagellar class II operons. J Bacteriol **176:**7345-7351.

39. **Macnab RM.** 2003. How bacteria assemble flagella. Annu Rev Microbiol **57:**77-100.

40. **Beeby M, Ribardo DA, Brennan CA, Ruby EG, Jensen GJ, Hendrixson DR.** 2016. Diverse high-torque bacterial flagellar motors assemble wider stator rings using a conserved protein scaffold. P Natl Acad Sci USA **113:**E1917-1926.

41. **Qin Z, Lin WT, Zhu S, Franco AT, Liu J.** 2017. Imaging the motility and chemotaxis machineries in *Helicobacter pylori* by cryo-electron tomography. J Bacteriol **199:**e00695-00616.

42. **Spohn G, Scarlato V.** 1999. Motility of *Helicobacter pylori* is coordinately regulated by the transcriptional activator FlgR, an NtrC homolog. J Bacteriol **181:**593-599.

43. **Beier D, Frank R.** 2000. Molecular characterization of two-component systems of *Helicobacter pylori*. J Bacteriol **182:**2068-2076.

44. **Tsang J, Hirano T, Hoover TR, McMurry JL.** 2015. *Helicobacter pylori* FlhA Binds the Sensor Kinase and Flagellar Gene Regulatory Protein FlgS with High Affinity. J Bacteriol **197:**1886-1892.

45. **Josenhans C, Niehus E, Amersbach S, Horster A, Betz C, Drescher B, Hughes KT, Suerbaum S.** 2002. Functional characterization of the antagonistic flagellar late regulators FliA and FlgM of *Helicobacter pylori* and their effects on the *H.pylori* transcriptome. Molecular Microbiology **43:**307-322.

46. **Rust M, Borchert S, Niehus E, Kuehne SA, Gripp E, Bajceta A, McMurry JL, Suerbaum S, Hughes KT, Josenhans C.** 2009. The *Helicobacter* pylori anti-sigma factor FlgM is predominantly cytoplasmic and cooperates with the flagellar basal body protein FlhA. J Bacteriol **191:**4824-4834.

47. **Niehus E, Gressmann H, Ye F, Schlapbach R, Dehio M, Dehio C, Stack A, Meyer TF, Suerbaum S, Josenhans C.** 2004. Genome-wide analysis of transcriptional hierarchy and feedback regulation in the flagellar system of *Helicobacter pylori*. Mol Microbiol **52:**947-961.

48. **Giorgio M, Trinei M, Migliaccio E, Pelicci PG.** 2007. Hydrogen peroxide: a metabolic by-product or a common mediator of ageing signals? Nat Rev Mol Cell Biol **8:**722-728.

49. **McCord JM, Fridovich I.** 1969. Superoxide dismutase. An enzymic function for erythrocuprein (hemocuprein). The Journal of biological chemistry **244:**6049-6055.

50. **Loew O.** 1900. A New Enzyme of General Occurrence in Organismis. Science **11:**701-702.

51. **Davies GR, Simmonds NJ, Stevens TR, Sheaff MT, Banatvala N, Laurenson IF, Blake DR, Rampton DS.** 1994. *Helicobacter pylori* stimulates antral mucosal reactive oxygen metabolite production in vivo. Gut **35:**179-185.

52. **Handa O, Naito Y, Yoshikawa T.** 2010. *Helicobacter pylori:* a ROS-inducing bacterial species in the stomach. Inflamm Res **59:**997-1003.

53. **Mahawar M, Tran V, Sharp JS, Maier RJ.** 2011. Synergistic roles of *Helicobacter pylori* methionine sulfoxide reductase and GroEL in repairing oxidant-damaged catalase. The Journal of biological chemistry **286:**19159-19169.

54. **Harris AG, Wilson JE, Danon SJ, Dixon MF, Donegan K, Hazell SL.** 2003. Catalase (KatA) and KatA-associated protein (KapA) are essential to persistent colonization in the *Helicobacter pylori* SS1 mouse model. Microbiology **149:**665-672.

55. **Seyler RW, Jr., Olson JW, Maier RJ.** 2001. Superoxide dismutase-deficient mutants of *Helicobacter pylori* are hypersensitive to oxidative stress and defective in host colonization. Infect Immun **69:**4034-4040.

56. **Benoit SL, Maier RJ.** 2016. *Helicobacter* Catalase Devoid of Catalytic Activity Protects the Bacterium against Oxidative Stress. The Journal of biological chemistry **291:**23366-23373.

57. **Miyashita M, Joh T, Watanabe K, Todoroki I, Seno K, Ohara H, Nomura T, Miyata M, Kasugai K, Tochikubo K, Itoh M, Nitta M.** 2002. Immune responses in mice to intranasal and intracutaneous administration of a DNA vaccine encoding *Helicobacter pylori*-catalase. Vaccine **20:**2336-2342.

58. **Kusters JG, van Vliet AH, Kuipers EJ.** 2006. Pathogenesis of *Helicobacter pylori* infection. Clin Microbiol Rev **19:**449-490.

59. **Benoit SL, Miller EF, Maier RJ.** 2013. *Helicobacter pylori* stores nickel to aid its host colonization. Infect Immun **81:**580-584.

60. **McGee DJ, Radcliff FJ, Mendz GL, Ferrero RL, Mobley HL.** 1999. *Helicobacter pylori rocF* is required for arginase activity and acid protection in vitro but is not essential for colonization of mice or for urease activity. J Bacteriol **181:**7314-7322.

61. **Gobert AP, McGee DJ, Akhtar M, Mendz GL, Newton JC, Cheng Y, Mobley HL, Wilson KT.** 2001. *Helicobacter pylori* arginase inhibits nitric oxide production by eukaryotic cells: a strategy for bacterial survival. P Natl Acad Sci USA **98:**13844-13849.

62. **Pich OQ, Merrell DS.** 2013. The ferric uptake regulator of *Helicobacter pylori*: a critical player in the battle for iron and colonization of the stomach. Future Microbiol **8:**725-738.

63. **Touati D.** 2000. Iron and oxidative stress in bacteria. Archives of biochemistry and biophysics **373:**1-6.

64. **Velayudhan J, Hughes NJ, McColm AA, Bagshaw J, Clayton CL, Andrews SC, Kelly DJ.** 2000. Iron acquisition and virulence in *Helicobacter pylori*: a major role for FeoB, a high-affinity ferrous iron transporter. Mol Microbiol **37:**274-286.

65. **Montefusco S, Esposito R, D'Andrea L, Monti MC, Dunne C, Dolan B, Tosco A, Marzullo L, Clyne M.** 2013. Copper promotes TFF1-mediated *Helicobacter pylori* colonization. PLoS One **8:**e79455.

66. **Nedenskov-Sorensen P, Bukholm G, Bovre K.** 1990. Natural competence for genetic transformation in *Campylobacter pylori*. J Infect Dis **161:**365-366.

67**. Tsuda M, Karita M, Nakazawa T.** 1993. Genetic transformation in *Helicobacter pylori*. Microbiol Immunol **37:**85-89.

68. **Wilson GG, Murray NE.** 1991. Restriction and modification systems. Annu Rev Genet **25:**585-627.

69. **Vasu K, Nagaraja V.** 2013. Diverse functions of restriction-modification systems in addition to cellular defense. Microbiol Mol Biol Rev **77:**53-72.

70. **Donahue JP, Israel DA, Peek RM, Blaser MJ, Miller GG.** 2000. Overcoming the restriction barrier to plasmid transformation of *Helicobacter pylori*. Mol Microbiol **37:**1066-1074.

71. **Zhang XS, Blaser MJ.** 2012. Natural transformation of an engineered *Helicobacter pylori* strain deficient in type II restriction endonucleases. J Bacteriol **194:**3407-3416.

72. **Janulaitis A, Petrusyte M, Maneliene Z, Klimasauskas S, Butkus V.** 1992. Purification and properties of the Eco57I restriction endonuclease and methylase--prototypes of a new class (type IV). Nucleic Acids Res **20:**6043-6049.

73. **Cheng X.** 1995. DNA modification by methyltransferases. Curr Opin Struct Biol **5:**4-10.

74. **Sanchez-Romero MA, Cota I, Casadesus J.** 2015. DNA methylation in bacteria: from the methyl group to the methylome. Curr Opin Microbiol **25:**9-16.

75. **Krebes J, Morgan RD, Bunk B, Sproer C, Luong K, Parusel R, Anton BP, Konig C, Josenhans C, Overmann J, Roberts RJ, Korlach J, Suerbaum S.**

2014. The complex methylome of the human gastric pathogen *Helicobacter pylori*. Nucleic Acids Res **42:**2415-2432.

76. **Roberts RJ, Vincze T, Posfai J, Macelis D.** 2015. REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res **43:**D298-299.

77. **Lin LF, Posfai J, Roberts RJ, Kong H.** 2001. Comparative genomics of the restriction-modification systems in *Helicobacter pylori*. P Natl Acad Sci USA **98:**2740-2745.

78. **Kong H, Lin LF, Porter N, Stickel S, Byrd D, Posfai J, Roberts RJ.** 2000. Functional analysis of putative restriction-modification system genes in the *Helicobacter pylori* J99 genome. Nucleic Acids Res **28:**3216-3223.

79. **Zweiger G, Marczynski G, Shapiro L.** 1994. A Caulobacter DNA methyltransferase that functions only in the predivisional cell. J Mol Biol **235:**472-485.

80. **Stephens C, Reisenauer A, Wright R, Shapiro L.** 1996. A cell cycle-regulated bacterial DNA methyltransferase is essential for viability. P Natl Acad Sci USA **93:**1210-1214.

81. **Peterson KR, Wertman KF, Mount DW, Marinus MG.** 1985. Viability of *Escherichia coli* K-12 DNA adenine methylase (dam) mutants requires increased expression of specific genes in the SOS regulon. Mol Gen Genet **201:**14-19.

82. **Lu AL, Clark S, Modrich P.** 1983. Methyl-directed repair of DNA base-pair mismatches in vitro. P Natl Acad Sci USA **80:**4639-4643.

83. **Vitkute J, Stankevicius K, Tamulaitiene G, Maneliene Z, Timinskas A, Berg DE, Janulaitis A.** 2001. Specificities of eleven different DNA methyltransferases of *Helicobacter pylori* strain 26695. J Bacteriol **183:**443-450.

84. **Xu Q, Peek RM, Jr., Miller GG, Blaser MJ.** 1997. The *Helicobacter pylori* genome is modified at CATG by the product of *hpyIM*. J Bacteriol **179:**6807-6815.

85. **Donahue JP, Israel DA, Torres VJ, Necheva AS, Miller GG.** 2002. Inactivation of a *Helicobacter pylori* DNA methyltransferase alters *dnaK* operon expression following host-cell adherence. FEMS Microbiol Lett **208:**295-301.

86. **Skoglund A, Bjorkholm B, Nilsson C, Andersson AF, Jernberg C, Schirwitz K, Enroth C, Krabbe M, Engstrand L.** 2007. Functional analysis of the M.HpyAIV DNA methyltransferase of *Helicobacter pylori*. J Bacteriol **189:**8914-8921.

87. **Chung D, Farkas J, Huddleston JR, Olivar E, Westpheling J.** 2012. Methylation by a unique alpha-class N4-cytosine methyltransferase is required for DNA transformation of *Caldicellulosiruptor bescii* DSM6725. PLoS One **7:**e43844.

88. **Kumar S, Karmakar BC, Nagarajan D, Mukhopadhyay AK, Morgan RD, Rao DN.** 2018. N4-cytosine DNA methylation regulates transcription and pathogenesis in *Helicobacter pylori*. Nucleic Acids Res **46:**3429-3445.

89. **Li E, Bestor TH, Jaenisch R.** 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell **69:**915-926.

90. **Bestor TH, Verdine GL.** 1994. DNA methyltransferases. Current opinion in cell biology **6:**380-389.

91. **Bestor TH, Coxon A.** 1993. Cytosine methylation: the pros and cons of DNA methylation. Curr Biol **3:**384-386.

92. **Shen JC, Rideout WM, 3rd, Jones PA.** 1994. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. Nucleic Acids Res **22:**972-976.

93. **Saxonov S, Berg P, Brutlag DL.** 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. P Natl Acad Sci USA **103:**1412-1417.

94. **Geiman TM, Robertson KD.** 2002. Chromatin remodeling, histone modifications, and DNA methylation-how does it all fit together? J Cell Biochem **87:**117-125.

95. **Kahramanoglou C, Prieto AI, Khedkar S, Haase B, Gupta A, Benes V, Fraser GM, Luscombe NM, Seshasayee AS.** 2012. Genomics of DNA cytosine methylation in *Escherichia coli* reveals its role in stationary phase transcription. Nat Commun **3:**886.

96. **Banerjee S, Chowdhury R.** 2006. An orphan DNA (cytosine-5-)-methyltransferase in *Vibrio cholerae*. Microbiology **152:**1055-1062.

97. **Chao MC, Zhu S, Kimura S, Davis BM, Schadt EE, Fang G, Waldor MK.** 2015. A Cytosine Methyltransferase Modulates the Cell Envelope Stress Response in the Cholera Pathogen [corrected]. PLoS genetics **11:**e1005666.

98. **Kumar R, Mukhopadhyay AK, Ghosh P, Rao DN.** 2012. Comparative transcriptomics of *H. pylori* strains AM5, SS1 and their *hpyAVIBM* deletion mutants: possible roles of cytosine methylation. PLoS One **7:**e42303.

99. **Chen IA, Chu K, Palaniappan K, Pillay M, Ratner A, Huang J, Huntemann M, Varghese N, White JR, Seshadri R, Smirnova T, Kirton E, Jungbluth SP, Woyke T, Eloe-Fadrosh EA, Ivanova NN, Kyrpides NC.** 2019. IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. Nucleic Acids Res **47:**D666-D677.

100. **Estibariz I, Overmann A, Ailloud F, Krebes J, Josenhans C, Suerbaum S.** 2019. The core genome m5C methyltransferase JHP1050 (M.Hpy99III) plays an important role in orchestrating gene expression in *Helicobacter pylori*. Nucleic Acids Res **47:**2336-2348.

101. **Azevedo NF, Almeida C, Cerqueira L, Dias S, Keevil CW, Vieira MJ.** 2007. Coccoid form of *Helicobacter pylori* as a morphological manifestation of cell adaptation to the environment. Appl Environ Microbiol **73:**3423-3427

102. **Waidner B, Melchers K, Stahler FN, Kist M, Bereswill S.** 2005. The *Helicobacter pylori* CrdRS two-component regulation system (HP1364/HP1365) is required for copper-mediated induction of the copper resistance determinant CrdA. J Bacteriol **187:**4683-4688.

103. **Han KD, Ahn DH, Lee SA, Min YH, Kwon AR, Ahn HC, Lee BJ.** 2013. Identification of chromosomal HP0892-HP0893 toxin-antitoxin proteins in *Helicobacter pylori* and structural elucidation of their protein-protein interaction. The Journal of biological chemistry **288:**6004-6013.

104. **Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermuller J, Reinhardt R, Stadler PF, Vogel J.** 2010. The primary transcriptome of the major human pathogen *Helicobacter pylori*. Nature **464:**250-255.

CHAPTER 2

BIOINFORMATIC ANALYSIS AND FUNCTIONAL CHARACTERIZATION OF THE MOTIF METHYLATED BY A CONSERVED ^{M5}C DNA METHYLTRANSFERASE IN *HELICOBACTER PYLORI*[1]

[1] Meng, B., Epp, N., Wijaya, W., Mrázek, J. and Hoover, T.R. To be submitted to *Frontiers in Microbiology*

Abstract

Helicobacter pylori colonizes the human stomach where it can cause peptic ulcers and other gastric diseases. *H. pylori* genomes contain several genes encoding Type II restriction-modification systems, which consist of a restriction endonuclease and DNA methyltransferase (MTase). *H. pylori* J99 M. Hpy99III (JHP1050) is a 5-methylcytosine (^{m5}C) MTase that converts GCGC motifs to G^{m5}CGC. Disrupting JHP1050 in *H. pylori* J99 was reported to alter expression of genes involved in various activities, suggesting methylation of GCGC motifs by JHP1050 influences transcription of certain genes. Examination of a transcriptional start site (TSS) dataset for *H. pylori* 26695 identified 211 putative promoter regions that contained a GCGC motif. GCGC motifs in these promoters displayed a bias for -13 position relative to the TSS, and several of these promoters were associated with genes that were differentially regulated in the *H. pylori* J99 JHP1050 mutant. A position-specific score matrix generated from an alignment of GCGC-containing promoters was used to identify similar motifs in the genomes of *H. pylori* 26695. Compared to randomized genome sequences, motifs resembling GCGC-containing promoters were overrepresented in intergenic regions and most of them appeared to be functional promoters. In contrast, the frequency of potential GCGC-containing promoters in protein-coding regions did not differ significantly from the random model and most of these motifs did not appear to be active promoters, suggesting they had arisen in the genome randomly and were not maintained by selection. Studies involving green fluorescent protein (*gfp*) reporter genes constructed with six potential GCGC-containing promoters in wild-type *H. pylori* G27 and 26695and a M.Hpy99III-deficient mutant identified three promoters whose activities appeared to be dependent on

the methylation of the GCGC motif. As reported for *H. pylori* J99, loss of M.Hpy99III homologs in *H. pylori* strains G27 and 26695 significantly decreased growth rates, suggesting that methylation of GCGC motifs in these strains is important for the appropriate expression of one or more genes that are critical for normal cell growth.

Introduction

Helicobacter pylori is a medically relevant organism that colonizes the stomach of about half the human population worldwide (1, 2). Infection of the gastric mucosa by *H. pylori* is the major factor for peptic ulcer disease and chronic gastritis, and a major risk factor for gastric cancer and mucosa-associated lymphoid tissue lymphoma (3-5). Several important factors facilitate host colonization by *H. pylori*, including urease, catalase and motility (6-8).

A remarkable feature of *H. pylori* genomes is the high number of genes encoding DNA restriction-modification (R-M) systems compared to the genomes of other bacteria (9-11). There are four types of R-M systems, with Type II R-M systems being the simplest and most prevalent in bacteria (11). Type II R-M systems consist of a restriction endonuclease and DNA methyltransferase (MTase) that act independently of each other. The restriction endonuclease recognizes and cuts a defined DNA motif, while the MTase methylates a specific nucleotide within the motif to prevent restriction of the bacterium's DNA. The primary major role of bacterial R–M systems is to protect the genome of the cell by restriction of incoming foreign DNA (12), although bacterial MTases have additional roles in gene expression, DNA replication, cell cycle control and chromosome maintenance (13, 14). *H. pylori* is naturally competent, which may account for the high number of R-M systems in *H. pylori* as this could serve to protect the bacterium from

invading foreign DNA (15). Although most MTases are associated with a restriction enzyme, some MTases in *H. pylori* and other bacteria are orphan MTases that lack cognate restriction enzymes.

MTases catalyze the addition of a methyl group from the donor S-adenosyl methionine to adenine or cytosine. Three types of DNA methylation have been identified in bacteria, N^6-methyladenine (^{m6}A), N^4-methylcytosine (^{m4}C), and ^{m5}C (16). Methylomes of *H. pylori* strains are highly diverse due to the different sets of MTases that the strains possess (9, 17, 18). Examination of base modifications of the *H. pylori* UM032 genome revealed 17 methylated sequence motifs (18), while methylome analyses of *H. pylori* strains 26695 and J99-R3 identified 17 and 22 methylated sequence motifs, respectively (9). Some *H. pylori* MTase genes are phase variable due to the presence of homopolymeric nucleotide repeats, which are susceptible to length changes resulting from slipped strand mispairing (9, 19). Despite the variability of *H. pylori* methylomes, a few target motifs appear to be methylated in all or nearly all *H. pylori* strains based on a study by Vale and co-workers in which they examined genomic DNA from 221 *H. pylori* strains for susceptibility to cleavage by 29 methylation-sensitive restriction enzymes (20).

MTase JHP1050 (M.Hpy99III) from *H. pylori* strain J99 methylates GCGC sequences to generate G^{m5}CGC motifs, and based on gene sequence analysis, M.Hpy99III homologs were predicted to be present and active in all 459 *H. pylori* strains examined by Estibariz and co-workers (21). The gene encoding the cognate restriction endonuclease for M.Hpy99III (*jhp1049*) was found in only 61 of the 459 *H. pylori* strains analyzed, but only 15 of these genes were predicted to be functional and the rest were pseudogenes with premature stop codons and/or frameshift mutations (21). Interestingly,

the presence of *jhp1049* homologs among *H. pylori* strains displays a biased phylogeographic distribution as a majority of strains that carry the gene belong to a population with substantial African ancestry (21).

Transcriptome analysis of *H. pylori* J99, *H. pylori* BCM-300, and mutants of the two strains where the gene encoding the M.Hpy99III homolog was inactivated showed that loss of the MTase resulted in altered expression of 225 genes in J99 and 29 genes in BCM-300, 10 of which were differentially regulated in both mutant strains (21). Growth rate, as well as expression of genes involved in important phenotypic traits, including adherence to host cells, natural competence, cell shape and copper sensitivity, were affected in the *H. pylori* J99 M.Hpy99III-deficient mutant (21). Taken together, these observations suggest methylation of GCGC sequences by M.Hpy99III plays an important role in gene expression in *H. pylori*.

The role of ^{m5}C in gene expression in mammalian systems is well characterized. When present near gene regulatory regions in mammalian DNA, ^{m5}C generally interferes with gene expression by inhibiting the binding of transactivators to their cognate sequence motifs or by interacting with methyl-CpG-binding proteins that recruit other proteins to the locus, which leads to gene repression (22-24). Some transactivators though, such as the Epstein-Barr virus transcription factor Zta, recognize ^{m5}C within DNA sequence motifs to activate epigenetically repressed viral promoters (25, 26).

In contrast to mammalian systems, little is known about how ^{m5}C influences gene expression in bacteria. *H. pylori* M.HpyAVIBM is an orphan ^{m5}C MTase that methylates CCTC sequences to generate ^{m5}CCTC motifs (27, 28). Deletion of *hpyAVIBM* in *H. pylori* strains AM5 and SS1 affected expression of a number of genes with roles in

motility, adhesion and virulence, although for some genes deleting *hpyAVIBM* in the two strains had opposing effects on expression (29). *Escherichia coli* K-12 strains possess the orphan ^{m5}C MTase Dcm that recognizes the sequence CCWGG. Loss of Dcm results in increased expression of several categories of genes (30-32). Dcm has been proposed to repress expression of *rpoS*, and loss of Dcm results in increased levels of RpoS, which leads to elevated expression of RpoS-dependent genes (30). Loss of the orphan ^{m5}C MTase VchM in *Vibrio cholerae* results in attenuated growth both *in vivo* in infant mice and *in vitro* in competition with wild-type cells (33). Transcriptome analysis identified 134 genes with elevated transcript levels in a Δ*vchM* mutant compared to the wild-type strain, but only about half of these genes contained the methylation target sequence (33). Moreover, mutating sites methylated by VchM in three of four genes up-regulated in the Δ*vchM* mutant had no significant effect on transcript levels, although mutating all three motifs in one gene did result in ~5-fold increase in transcript abundance (33). These observations indicate that while ^{m5}C may directly influence expression of some genes in *V. cholerae*, many of the genes whose expression was altered in the Δ*vchM* mutant are indirectly affected by ^{m5}C.

In the case of *H. pylori* M.Hpy99III, GCGC motifs that overlap promoters may influence gene expression directly, while methylation of other GCGC sequences may have an indirect effect on gene expression. As an example of the former, Estibariz and co-workers identified a GCGC motif in the extended -10 region of the promoter for the *jhp0832*/*jhp0831* operon (encodes a toxin-antitoxin system) that appeared to be required for gene expression (21). In the present study, we analyzed the frequency of GCGC motifs in the *H. pylori* 26695 genome and the positional distribution of GCGC motifs

within predicted promoter regions. A strong bias for location of the GCGC motif near the -13 position of promoters was observed, suggesting that methylation of the GCGC motif has a role in the function of a significant number of promoters and has been maintained by selective pressure. Expression of green fluorescent protein (GFP) reporter genes under control of *H. pylori* GCGC-containing promoters suggested further that methylation of the GCGC motif was important for promoter activity.

Results

Comparison of *H. pylori* promoter sequences containing GCGC motifs. A search of a transcriptional start site (TSS) database for the *H. pylori* 26695 transcriptome (34) revealed 211 putative promoters that had a GCGC motif within 50 nucleotides of the TSS (Table S2.1). About 39% of the sequences were predicted primary or secondary promoters of genes, while ~14% were within coding regions of genes and in the same orientation (i.e., internal promoters) and ~47% were within coding regions of genes but in the opposition orientation (i.e., antisense promoters). The positions of the GCGC motifs within the promoter regions are not distributed uniformly as there is a marked preference around position -13 (measured from the 3'-end of the GCGC motif) (Fig. 2.1). GCGC motifs appeared to be distributed randomly across the rest of the promoter region with the exception of the area spanning positions -5 to -10 for which there was a single GCGC motif in the 211 promoter sequences. The excess of GCGC counts at position -13 (50 GCGC sites) relative to any other position in the promoter region is highly significant with p-value <0.00001 using a two-tail binomial test. The bias against GCGC motifs in the -5 to -10 region suggests a high G + C content in the region is deleterious to promoter function. Consistent with this hypothesis, a sequence logo generated for all of the

predicted promoter sequences identified from the *H. pylori* 26695 transcriptome (34) revealed a preference for A or T residues in the region from -7 to -12 (Fig. S2.1). Our further investigation focused on the subset of *H. pylori* promoters with a GCGC motif in the -13 region, which we hereafter refer to as GCGC-containing promoters.

We compared the sequences of the GCGC-containing promoters to determine if they shared sequence similarity. Since the distance between core promoter elements and the TSS may vary due to various factors, including promoter sequence, DNA topology and concentrations of initiation nucleotide triphosphates (35-39), we aligned the sequences of 79 promoters where the GCGC motif was positioned from -11 to -15 relative to the predicted TSS (Table S2.2). Thirty-seven of the promoters are listed as primary or secondary promoters in the TSS database, while 11 were listed as internal and 31 were listed as antisense promoters (34). Sequence alignment of the 79 GCGC-containing promoters revealed nucleotides at positions -7 (T), -11 (A) and -12 (T) were highly conserved, and also indicated a weaker preference for A at positions -8 and -9 (Fig. 2.2).

Sequences resembling GCGC-containing promoters within protein-coding regions may have arisen by chance in the absence of selection. Several recent studies raised questions about the roles of intragenic promoters (40-45). While many intragenic promoters are verifiably active (40-45), their physiological roles are often obscure. A combination of experiments and computational simulations suggested that promoters, as well as transcription factor binding sites in general, are likely to arise by chance even in the absence of selection, which could be beneficial in promoting evolutionary adaptations by increasing the dynamics of evolution of gene regulatory networks. To investigate

whether the internal and antisense GCGC-containing promoters in *H. pylori* can arise by chance in the absence of selection, we used the model described by Mrázek and Karls (46).

Comparisons of position-specific score matrix (PSSM) score distributions using the PSSM derived from the alignment of the GCGC-containing promoters in the *H. pylori* 26695 genome and 1000 randomized genomes are shown in Figure 2.3. Only the right tail of the distribution with PSSM scores > 0 is shown. For PSSM scores close to zero, the simulations were expected to match the observed values because such low scores typically do not indicate active promoters. This was true for most transcription factors investigated in Mrázek and Karls (46). However, the values for the predicted GCGC-containing promoters were systematically lower for the random sequences than for the actual *H. pylori* genome (Fig. 2.3). This indicates that even the more complex "m1c1" model to generate the random sequences, which reproduces the codon usage, dinucleotide content, and sequence heterogeneity of the actual genome at the scale of individual genes, is not an accurate representation of the null hypothesis that would ideally consider all biases affecting the genome sequence except those resulting directly from the selection on active GCGC-containing promoters. Nevertheless, the model matches the overall shape of the distribution of PSSM scores in the genome up to the PSSM value of ~11, after which the deviation from the random sequences begins to increase. This is clearly visible in the intergenic sequences where physiologically important promoters are most likely to occur and the increasing deviation from the random model at the higher PSSM score values is indicative of a number of active GCGC-containing promoters maintained by selection. In the absence of selection, the curve for the actual genome would be

expected to match approximately the shape of that for the random sequences. This was the case for sequences from the protein-coding regions as the values for the model and actual genome tended to converge for higher PSSM score values, with the exception of four potential promoters in the antisense strand with unexpectedly high PSSM scores >20. The four potential antisense promoters with unexpectedly high PSSM scores were located in *bcp* (HP0136, encodes bacterioferritin co-migratory protein), *lpxL* (HP0280, encodes lipid A acyl transferase; (47)), *valS* (HP1153, encodes valyl-tRNA synthetase) and *alaS* (HP1241, encodes alanyl-tRNA synthetase). The TSS data for *H. pylori* 26695 (34) confirmed the sequences in *lpxL*, *valS* and *alaS* as functional antisense promoters, although the promoter in the coding sequence of *valS* was reported to be the primary promoter for the sRNA encoding HPnc6160.

To evaluate the efficacy of PSSM scores in identifying GCGC-containing sequences with potential promoter activity in *H. pylori*, we generated a PSSM from the alignment of the GCGC-containing promoters from the *H. pylori* 26695 transcriptome (34) using the program Motif Locator (https://www.cmbl.uga.edu//software.html). The PSSM was used by the program to scan the *H. pylori* 26695 genome sequence for all words of 51 nucleotides in length with a PSSM score higher than the default score cutoff, which was set as 13.148, the tenth percentile of all scores for the motifs in the training set. The Motif Locator analysis identified 287 sequences that met or exceeded this cutoff score in the *H. pylori* 26695 genome, which was well above the number of sequences in the training set and indicated that only ~24% of the identified sequences appeared to be active promoters under the conditions used for the *H. pylori* 26695 transcriptome analysis (34). Approximately 62% of the motifs with high PSSM scores that were located in

intergenic regions and in the correct orientation to function as a primary or secondary promoter appeared to be active promoters (Table 2.1). In contrast, only about 19% of the motifs with high PSSM scores that were located in protein-coding regions appeared to be active promoters (Table 2.1). These data suggest that most sequences with high PSSM scores that are located in protein-coding regions are not active promoters maintained by selection, whereas sequences with high PSSM scores that are positioned in intergenic regions have a much higher probability of being active promoters.

Since many of the sequences with high PSSM scores did not appear to be functional promoters, we searched for DNA elements that might account for promoter activity. To address this issue, we used the MEME program in the MEME suite (https://pubmed.ncbi.nlm.nih.gov/25953851/) in the standard mode as well as in the differential mode, which searches specifically for motifs in a target set of sequences relative to a control set of sequences. For the analysis, we searched for motifs within 50, 100, 200, 300 or 500 nucleotides of the GCGC motif that were enriched in the set of 69 GCGC-containing motifs with promoter activity (i.e., were from the training set used to generate the PSSM), but were underrepresented in a set of 110 GCGC-containing motifs that appeared to lack promoter activity (i.e., were not associated with a TSS). Only a minor difference between the two sets of sequences was noted in the region immediately downstream of the GCGC motif. Within the set of sequences that had promoter activity, the motif GCGCTANAAT was enriched, while in the set of sequences that appeared to lack promoter activity the motif GCGCTAAAAA was enriched.

Identification of GCGC-containing promoters whose activity may be influenced by methylation. Estibariz and co-workers identified 225 genes that were differentially

expressed (differentially expressed genes; DEGs) in the *H. pylori* J99 *jhp1050* mutant, and provided evidence that the GCGC-containing promoter for one of the DEGs, *jhp0832*, was responsive to methylation (21). To examine the possibility that methylation of GCGC-containing promoters may have accounted for the differential expression of other genes in the *H. pylori* J99 *jhp1050* mutant (21), we used the Motif Locator program to identify potential GCGC-containing promoters in the vicinity of these genes. For the analysis, the same alignment of sequences used for analysis of the *H. pylori* 26695 genome was used as the training set. The Motif Locator analysis identified 272 sequences with high PSSM scores in the *H. pylori* J99 genome, which was comparable to the number of motifs (n = 287) identified in the *H. pylori* 26695 genome. Motifs identified in the search were checked to identify those associated with DEGs in the *H. pylori* J99 *jhp1050* mutant (21), and for those that were, the sequences were cross-checked against the *H. pylori* 26695 TSS database (34) to identify sequences that likely corresponded to active promoters. Twenty-two potential GCGC-containing promoters were identified in the *H. pylori* J99 genome, 14 of which were predicted primary or secondary promoters (Table 2.2). Most of these predicted GCGC-promoters were located upstream of genes that were down-regulated in the *H. pylori jhp1050* mutant, although three of the putative promoters were upstream of genes that were up-regulated in the *jhp1050* mutant (Table 2.2). Four of the putative GCGC-containing promoters were internal promoters and four were antisense promoters, and half of the genes bearing these promoters or the adjacent genes were up-regulated in the *jhp1050* mutant, while the other half were down-regulated in the *jhp1050* mutant (Table 2.2). These findings suggest the activities of these 22

potential GCGC-containing promoters from *H. pylori* J99 are influenced directly by methylation of the GCGC motif.

Expression of some GFP reporter gene constructs bearing GCGC-containing promoters is responsive to DNA methylation. The presence of predicted GCGC-containing promoters within intergenic regions located directly upstream of many of the DEGs in *H. pylori* J99 suggested methylation of the GCGC motifs in these promoters may affect their activities. To examine this hypothesis, as well as extend the survey of potential GCGC-containing promoters from other *H. pylori* strains, we constructed green fluorescent protein (*gfp*) reporter genes with four predicted GCGC-containing promoters from *H. pylori* G27. Three of the putative *H. pylori* G27 promoters corresponded to primary GCGC-containing promoters in *H. pylori* 26695 (*icd*, *cah* and HP0893 (HPG27_846); encode isocitrate dehydrogenase, α-carbonic anhydrase and a predicted antitoxin of a toxin-antitoxin system), while the fourth promoter corresponded to an antisense GCGC-containing promoter within HP0914 (HPG27_865; encodes a putative outer membrane protein) (34).In addition, *gfp* reporter genes were constructed with two predicted GCGC-containing promoters of DEGs in the *H. pylori* J99 *jhp1050* mutant (21). One of the promoters was the potential primary promoter for *jhp0160* (encodes a hypothetical protein), which was down-regulated in the *jhp1050* mutant (21). The other promoter was a potential anti-sense promoter in the coding sequence of *jhp0334* (*kgtP*, encodes α-ketoglutarate permease), which was up-regulated in the *jhp1050* mutant (21). The *gfp* reporter genes were introduced into wild-type *H. pylori* G27 and a HPG27_1066 (encodes the M.Hpy99III ortholog) mutant on the shuttle vector pHel3. We confirmed the M.Hpy99III ortholog in *H. pylori* G27 was functional by digesting gDNA from wild type

and the HPG27_1066 mutant with the restriction enzyme HinP1I, which cuts DNA at unmethylated GCGC sites but not methylated sites (Fig. S2.2). The *icd-gfp* and *hpg27*_846-*gfp* reporter genes were expressed at high levels in the wild-type strain, but expression levels of the reporter genes were reduced ~2-fold and ~4-fold, respectively, in the HPG27_1066 mutant relative to wild type (Fig. 2.4). The *jhp0160-gfp* reporter gene was expressed at a lower level compared to the *icd-gfp* and *hpg27*_846-*gfp* reporter genes, and expression levels of this *gfp* reporter gene were reduced ~7-fold in the HPG27_1066 mutant (Fig. 2.5B). In contrast, the *cah-gfp*, *hpg27*_865-*gfp* and *jhp0334-gfp* reporter genes were expressed at very low levels in the wild-type strain and there were no significant differences in expression in the HPG27_1066 mutant (Fig. 2.4). The sequence of the anti-sense promoter in HP0914 matches perfectly with that in HPG27_865 (Table S2.2), suggesting that the low level of expression of the HPG27_865-*gfp* reporter gene is not due to sequence differences between the two *H. pylori* strains.

To determine if the GCGC motif is critical for activity of the HPG27_846 promoter, we altered the motif to GTGC. GFP expression from the reporter gene with the altered HPG27_846 promoter was reduced ~5-fold in wild-type *H. pylori* G27 (Fig. 2.4D). Interestingly, GFP expression from the reporter gene with the altered HPG27_846 promoter was reduced further in the HPG27_1066 mutant (Fig. 2.4D). The reason for the reduce GFP expression in the HPG27_1066 mutant is not known. There are two other GCGC motifs within the *gfp* coding sequence, but it is not clear how these motifs might influence expression of the *gfp* reporter gene.

Deletion of HPG27_1066 results in slower growth and altered expression of certain proteins in *H. pylori* G27. Estibariz and co-workers reported that loss of M.Hpy99III in

H. pylori J99 resulted in significantly slower growth (21). We observed a similar growth defect in the HPG27_1066 mutant, which had a doubling time that was more than twice that of the wild type strain in BHI medium supplemented with horse serum (Table 2.3). To determine if loss of the M.Hpy99III homolog in other *H. pylori* strains resulted in similar growth defects, we disrupted the gene encoding the M.Hpy99III homolog in *H. pylori* strains B128, ATCC 43504, X47-AL2 and 26695 and measured the growth rates of the resulting mutants. Only the *H. pylori* 26695 M.Hpy99III-deficient mutant exhibited growth that was significantly slower compared to the parental strain (Table 2.3). The result with the *H. pylori* 26695 mutant conflicted with the previous report by Estibariz and co-workers (21), which may reflect differences in lab strains or growth conditions used in our growth assays.

Examination of the protein profiles of wild-type *H. pylori* G27 and the HPG27_1066 mutant revealed several apparent differences in the relative amounts of proteins in the two strains (Fig. 2.6). These differences were most evident in the insoluble protein fractions of the two strains. To identify proteins that were differentially expressed in the wild-type and mutant strains, the protein samples were subjected to SDS-PAGE and protein bands were excised from the resulting gel and analyzed by in-gel trypsin digestion of the proteins followed by liquid chromatography-mass spectrometry (LC-MS/MS). Nine proteins were identified as being present at lower levels in the HPG27_1066 mutant and nine proteins were identified as being present at elevated levels in the mutant (Table 2.4). None of the genes encoding the potentially differentially expressed proteins had an associated GCGC-containing promoter. Consistent with the dogma that the number of ribosomes per cell increases with growth rate, three of the

proteins that were present at lower levels in the HPG27_1066 mutant were ribosomal proteins (L2 and S1) or involved in translation (elongation factor G). Catalase (KatA) levels also appeared to be reduced in the HPG27_1066 mutant (Table 2.4). To confirm that the HPG27_1066 mutant did indeed have lower amounts of catalase, we compared the sensitivities of the mutant and wild-type *H. pylori* G27 strains to H_2O_2 (Fig 2.7). As shown in Figure 2.7, the HPG27_1066 mutant was more sensitive to H_2O_2 than the wild-type parental strain.

Although levels of flagellar proteins appeared to be unaffected in the HPG27_1066 mutant, we examined the motility of the mutant in soft agar medium. For *H. pylori* G27 strain, loss of HPG27_1066 resulted in decreased in motility when $FeSO_4$ was included in the motility medium (Fig. 2.8). Omission of $FeSO_4$ resulted in a statistically significant increase in motility, however, the motility was not restored to wild-type levels. Motility of the *H. pylori* B128 M.Hpy99III-deficient mutant was also reduced compared to the wild-type parental strain, but in contrast to the *H. pylori* G27 mutant, the presence of $FeSO_4$, in the motility medium had no influence on the size of the swim halo (Fig. 2.8).

Discussion

As demonstrated by Estibariz and co-workers (21), orthologs of the ^{m5}C MTase M. Hpy99III play a critical role in *H. pylori* biology as loss of the MTase in *H. pylori* strains J99 and BCM-300 resulted in altered expression of numerous genes in the two strains, many of which are involved in cellular functions required for host colonization. While the altered expression of some of these genes is due likely to indirect effects, Estibariz and co-workers showed that a GCGC motif in the extended -10 region of the

jhp0832 promoter apparently needs to be methylated for optimal promoter activity (21). To ascertain the extent to which GCGC motifs occur within *H. pylori* promoter regions, we searched a dataset of predicted promoter sequences for *H. pylori* 26695 generated by Sharma and co-workers in a transcriptome analysis (34). Approximately 11% of the predicted promoters in *H. pylori* 26695 (211 of 1910) possessed a GCGC motif within 50 nucleotides of the TSS. Interestingly, the GCGC motif showed a strong bias for the -13 region (Fig. 2.1), suggesting the positioning of the GCGC motif in this region has functional significance in many of these promoters and has been maintained in the *H. pylori* 26695 genome by selective pressure.

Using a PSSM generated from a training set of GCGC-containing promoter sequences to search the *H. pylori* 26695 genome for potential GCGC-containing promoters worked well for intergenic sites as it produced a reasonably acceptable false positive rate (~38%), whereas the false positive rate for sequences within protein-coding regions was quite high (~81%) (Table 2.1). Comparing the sequences of GCGC-containing promoters with sequences that were not associated with identified promoters but had high PSSM scores failed to yield insight into DNA elements that may be important for promoter function. The large number of sequences with high PSSM scores that were not associated with putative promoters, particularly motifs that were located within protein-coding regions, suggest that most of these sites arise by chance in the absence of natural selection. Consistent with this hypothesis, PSSM score distributions using the PSSM derived from the alignment of the GCGC-containing promoters did not differ between the *H. pylori* 26695 genome and randomized genomes for intragenic sequences (Fig. 2.3). In contrast, the PSSM score distribution for intergenic sequences in

the *H. pylori* 26695 genome deviated from that of the randomized genomes when the PSSM value exceeded 11 (Fig. 2.3), which is indicative of maintenance of active GCGC-containing promoters within intergenic regions by selection.

Using a TSS dataset from *H. pylori* 26695 to evaluate a transcriptome dataset that identified DEGs in the *H. pylori jhp1050* mutant, we identified several GCGC-containing promoters whose activities are likely to be affected directly by methylation of the GCGC motif (Table 2). Assays with *gfp* reporter genes bearing predicted GCGC-containing promoters from *H. pylori* strains G27 or J99 provided supporting evidence that the activities of some of these promoters are dependent on methylation (Fig. 2.4, Fig. 2.5). Only three of the six predicted GCGC-containing promoters examined (*icd*, HPG27_846 and *jhp0160* promoters) had moderate to high activity in the *gfp* reporter gene assays. These promoters have two T residues immediately upstream of the GCGC motif, while the potential GCGC-containing promoters that had low activity in the *gfp* reporter gene assays (*cah*, HPG27_865 and *jhp0334* promoters) have a GA or TA dinucleotide immediately upstream of the GCGC motif (Table S2.2). Although it is a small sample size, these data suggest the presence of a T residue upstream of the GCGC motif is a critical determinant for promoter strength. Two of the GCGC-containing promoters examined the *gfp* reporter gene assays were potential anti-sense promoters that would interfere with expression of the genes in which they were located (HPG27_865 and *jhp0334*). These GCGC-containing anti-sense promoters may have arisen by chance and are tolerated due to their low intrinsic activities. Alternatively, the anti-sense promoters may have regulatory roles under specific environmental conditions or require *cis*-acting

elements that were not included in the *gfp* reporter gene constructs, which might account for the low activities of the promoters in the *gfp* reporter gene assays.

An interesting question is how does methylation of the GCGC motif influence promoter activity? Methylation of GCGC motifs may influence the activity of GCGC-containing promoters by affecting base readout in the DNA major groove or shape readout in the DNA minor groove (48), and thereby impact interaction between RNA polymerase and the promoter. Each nucleotide base pair has a unique array of group signatures in the major groove that may be recognized by specific amino acid residues in protein-DNA interactions. Methylation of C-5 in cytosine alters the functional group signature in the major groove by replacing a nonpolar hydrogen with a bulkier and more hydrophobic methyl group. The 5-methyl group in ^{m5}C and thymine is situated at the major groove edge and the ^{m5}C-G base pair can be contacted through hydrophobic interactions, similar to how thymine is contacted in unmethylated DNA.

If the methyl group of ^{m5}C and thymine is sufficient to confer binding specificity, then in principle, the two nucleotides should be able to substitute for each other in DNA-binding sites. In studies with synthetic *E. coli lac* operators, substitution of a specific A-T base pair with a G-C base pair, but not a G-^{m5}C base pair, decreased the stability of the LacI-operator complex (49). Replacing the A-T base pair with an A-U base pair also reduced the stability of the LacI-operator complex, which implicates the thymine 5-methyl group as the only important functional group recognized by LacI at this base pair since the A-U base pair has the same group signature as the A-T base pair with the exception of the 5-methyl group (49).

Future studies can address these issues as well as others. For example, it will be interesting to test the hypothesis that changing the GCGC motif to GTAC mimics the $G^{m5}CGC$ motif and allows promoter activity. In addition, characterization of additional GCGC-containing promoters will identify contextual elements that are important for robust promoter activity.

Materials and Methods

Bioinformatic analysis. Computer simulations were used to assess the occurrence of the promoter motif under the conditions of a null hypothesis, which assumed the motif is not subject to direct selective constraints but is potentially influenced by other biases, including biased codon and amino acid usage, dinucleotide usage biases, or local variance in GC content. For the simulation, 1,000 randomized genomes were generated from the *H. pylori* 26695 genome using the 'm1c1' model in Genome Randomizer (50). In the modeling, the genome is separated into segments consisting of protein-coding genes and intergenic regions. Each intergenic region is modeled as a first order Markov chain using the nucleotide alphabet and each gene as a first order Markov chain using the codon alphabet, where the next codon probability depends on the last base of the previous codon. Consequently, the model reproduces the compositional heterogeneity of the sequence at the scale of individual genes (e.g., GC-rich genes or intergenic segments in the AT-rich *H. pylori* genome retain their lower AT content) as well as asymmetry between the sense (coding) and antisense (template) strand, and between the leading and lagging strand with respect to the direction of replication (i.e., GC-skew). Moreover, the model reproduces the dinucleotide frequencies in each intergenic region and codon

frequencies as well as frequencies of dinucleotides spanning adjacent codons in each gene.

Motif Locator (https://www.cmbl.uga.edu//software/motloc.html) was used to identify potential GCGC-containing promoters in the *H. pylori* J99 genome. Sequences of 76 potential GCGC-containing promoters identified from a transcriptome analysis of *H. pylori* 26695 (34) were aligned with the GCGC motif as the reference point. Sequences were 51 nucleotides in length and corresponded to the TSS identified in the transcriptome analysis along with 50 nucleotides of sequence upstream of the TSS (34). Motif Locator converted the alignment into a position-specific score matrix (PSSM), then scanned the *H. pylori* J99 genome sequence for all words of 51 nucleotides in length with a PSSM score higher than a given cutoff S_0. The default score cutoff, which is the tenth percentile among all scores for the 78 motif sequences in the alignment, was used for the analysis. The primary output of the program is a set of coordinates in the analyzed DNA sequence of motifs similar to those in the alignment. These coordinates can be subsequently passed to other programs (*r*-scan statistics, pattern vicinity analysis) to provide additional information about the distribution of the matching motifs in the analyzed sequence and with respect to genes.

Bacterial strains and growth conditions. *E. coli* DH5α was used for cloning and plasmid construction. *E. coli* strains were grown in Luria-Bertani broth or agar medium supplemented with kanamycin (30 μg/mL) or ampicillin (100 μg/mL) when appropriate. For routine growth of *H. pylori* strains, the cultures were grown microaerobically under an atmosphere consisting of 10% CO_2, 4% O_2 and 86% N_2 at 37°C on tryptic soy agar (TSA) supplemented with 5% heat-inactivated horse serum. When required,

erythromycin (100 μg/mL) or Kanamycin (30 μg/mL) was added into the *H. pylori* growth medium. Liquid cultures of *H. pylori* strains were grown microaerobically at 37°C in bottles containing 75% N_2, 10% H_2, 10% O_2, and 5%CO_2 supplemented with 5% heat-inactivated horse serum.

Inactivation of gene encoding M.*Hpy*99III homolog in *H. pylori* G27. HPG27_1066 in *H. pylori* G27, which is the ortholog of JHP1050 (encodes M.*Hpy*99III) and HP1121 in *H. pylori* strains J99 and 26695, respectively, was disrupted with an erythromycin-resistance cassette (*ermB*) as follows. iProof DNA polymerase (Bio-Rad) was used for PCR with genomic DNA (gDNA) from *H. pylori* B128 purified using the Wizard® Genomic DNA Purification Kit (Promega) as the template. Primers used for PCR are listed in Table S2.3. Primer set hp1121up_F2 and hp1121up_R2 was used to amplify a 505-bp DNA fragment that included 413 bp of DNA upstream of the predicted start codon of HP1121 plus 92 bp of HP1121 coding sequence. Primer set hp1121d_F2 and hp1121d_R2 was used to amplify a 576-bp DNA fragment that included 489 bp of DNA downstream of the stop codon of HP1121 plus 87 bp of HP1121 coding sequence. Primer set ermB_OLF and ermB_OLR was used to amplify an erythromycin-resistance cassette from pSB167 (51). Primers ermB_OLF and hp1121up_R2 shared complementary sequence, as did primers ermB_OLR and hp1121d_F2, and the three amplicons were joined by overlapping PCR. The resulting amplicon was transformed into *H. pylori* B128 and erythromycin-resistant transformants were selected on TSA supplemented with erythromycin. After confirmation of proper insertion into strain B128 via PCR and DNA sequencing, primer set hp1121up_F2 and hp1121d_R2 were then used to amplify the B128 genomic DNA with the erythromycin insertion. The resulting PCR product was

transferred into *H. pylori* G27 via natural transformation. PCR and DNA sequencing confirmed replacement of HPG27_1066 with the *ermB* cassette in the *H. pylori* G27 chromosome. Inactivation of HPG27_1066 was confirmed further by showing that gDNA from the mutant was digested with HinP1I (New England Biolabs), which recognizes the sequence GCGC but is sensitive to DNA methylation. The same method to inactivate HPG27_1066 in *H. pylori* G27 was used to create methylase deficient strains of 26695, X47-2AL, and ATCC 43504.

Construction of *gfp* reporter genes. A synthetic gene containing the coding region of *gfp* in which the codon usage was optimized for *H. pylori* was synthesized by Genewiz (Fig. S2.3). In addition to the *gfp* coding region, the synthetic gene contained the *H. pylori ureA* ribosome binding site and the *H. pylori* G27 HPG27_846 promoter region, which was flanked by XhoI and NheI restriction sites. A BamHI restriction site was introduced immediately downstream of the *gfp* stop codon in the synthetic gene. The synthetic gene, which was provided by the supplier in plasmid pUC57-Kan, was cloned into the XhoI and BamHI restriction sites in the shuttle vector pHel3 (52) to generate plasmids pECBM27 – pECBM30. To construct reporter genes in which *gfp* was under control of other promoters, DNA corresponding to the promoter regions were amplified by PCR using iProof DNA polymerase (Bio-Rad) and gDNA from *H. pylori* J99 as the template. Primers used for PCR are listed in Table S2.3, and each primer contained a XhoI or NheI restriction site to facilitate exchange of the HPG27_846 promoter in plasmid pECBM30 with the new promoter region. A 320 base pair fragment containing roughly 300 base pairs upstream of the GCGC motif (some fragments were shortened due to internal restriction sites) and 3 base pairs downstream of the predicted transcriptional

start site was amplified. For each reporter plasmid construct, the promoter region was sequenced (Eton Biosciences) to confirm that the reporter plasmid construct was correct. The reporter gene for HPG27_865 used gDNA from G27 as template and amplified a region approximately 50 base pairs upstream of the GCGC motif and 5 base pairs downstream of the predicted transcriptional start site.

Growth Curves. *H. pylori* strains were grown for 3-4 days and then inoculated into bottles containing Brain Heart Infusion (BHI) pH 6.5 + 5% heat-inactivated horse serum for growth. After overnight growth in the bottles, each culture was sub-cultured into new bottles to a starting OD_{650} of 0.01 (Strain 26695 was sub-cultured to a starting OD_{650} of 0.02 due to slower growth). Each strain was inoculated into three bottles per experiment. Hourly measurements were taken after inoculation. Growth curve data was graphed and doubling time was calculated from using the slope from the line of best fit of the predicted exponential phase. At least two to three biological replicates were used to calculate growth rate and the two-tailed Student's t-test was used to calculate statistical significance.

Motility Plate assay. *H. pylori* strains were grown routinely as previously described on tryptic soy agar (TSA) supplemented with 5% heat-inactivated horse for 3-4 days before being stab inoculated into the middle of Mueller Hinton Broth plates supplemented with 0.4% Noble Agar, 5% heat-inactivated horse serum, 7 μM $FeSO_4$, and 20 mM 2-(*N*-morpholino) ethanesulfonic acid pH 6.0 (MES). After inoculation, the plates were then incubated under microaerophillic conditions for one week prior to recording results. Following incubation, the diameters of swim halos were recorded. At least three replicates were reported for studies including $FeSO_4$ and one to two replicates were

reported for studies without $FeSO_4$. The two-tailed Student's t-test was used to calculate statistical significance.

Restriction Digestion analyses. Genomic DNA from *H. pylori* strains were harvested from plates using the Wizard® Genomic DNA Purification Kit (Promega) after being grown for 4 - 5 days. Following DNA extraction, 1μg of gDNA was digested with HinP1I (New England Biolabs) for 1 hour and the resulting digestion reaction was run out on a 0.8% agarose gel.

Catalase Assay. After 3-4 days of growth, *H. pylori* strains were harvested and resuspended to a final OD_{600} of 0.500 in BHI. Approximately 100 μL of each strain was spread plated onto Tryptic Soy Agar and a sterile filter disk (6 mm in diameter) was placed in the middle of each plate. Approximately 10 μL of 30% H_2O_2 was then placed onto each filter and the plates were incubated for 2 days under microaerophilic conditions. Zones of inhibition were measured following incubation from three technical replicates each consisting of three biological replicates. The two-tailed Student's t-test was used to determine statistical significance.

***gfp* reporter gene assays.** *H. pylori* cells were grown on TSA medium supplemented with kanamycin and after 4-5 days, cells were harvested and resuspended in phosphate buffer (pH 8.0, 15 mM KH_2PO_4 + 17 mM K_2HPO_4) to a final OD_{600} of 0.5. Samples were placed into in a 96-well plate with a black bottom (Greiner). A Biotek Synergy™ Mx Microplate Reader was used to detect fluorescence at 462 nm excitation and 510 nm absorbance with a gain of 75. Three biological replicates as well as three technical replicates (G27 promoters only) were used to collect data. Statistical significance was calculated using the two-tailed Student's t-test

Preparation of samples for proteomic analysis. Cells of *H. pylori* strains were grown for 5 days on TSA medium supplemented with horse serum, harvested, and resuspended in 20 mL phosphate buffer (15 mM KH_2PO_4 + 17 mM K_2HPO_4, pH 8.0). Cells were pelleted by centrifugation at 3,000 x g for 10 minutes and resuspended in phosphate buffer prior to being lysed by French press (14,000 psi cell pressure) or in a One Shot Cell Disrupter set to 20,000psi (Pressure BioSciences Inc.). Afterwards, cells were spun at 5,600 x g for 10 minutes to pellet cell debris and unbroken cells, and the resulting supernatant was centrifuged at 100,000 x g for one hour to separate the soluble and insoluble fractions. Following centrifugation, the resulting supernatant was transferred to a separate tube and the pellet was resuspended in 1X Phosphate Buffer pH 8.0. A BCA assay (Thermo-Fischer) was used to quantify protein concentration and the fractions were then visualized on a 12% SDS-PAGE gel. Areas of the resulting gel where protein band intensity visually differed between the wild-type and mutant strains were excised and were sent to be analyzed via mass spectrometry.

Author Contributions

Planned and designed the study: BM, JM, TRH, Performed the experiments: BM, NE, Analyzed the data: BM, WW, JM, TRH, Wrote the manuscript: BM, JM, TRH.

Acknowledgements

We thank Rob Maier for providing equipment for culturing *H. pylori* strains, Ron Orlando for proteomic analysis, and Andrew Bohannon for technical assistance. This work was supported by NIH grant AI140444 to T.R.H.

Table 2.1: Distribution of potential GCGC-containing promoters in the *H. pylori* 26695 genome predicted from the Motif Locator program.

location and orientation of motifs	coding regions		intergenic/partial overlap	
	sense	antisense	sense	antisense
number of motifs	106	125	42	14
number of motifs associated with TSS	18	25	26	1

Table 2.2: Putative GCGC-containing promoters in *H. pylori* 26695 that are associated with differentially expressed genes in *H. pylori* J99 *jhp1050* mutant.

	H. pylori 26695		*H. pylori* J99			
[1]gene	[1]locus tag (26695)	PSSM score	locus tag (J99)	PSSM score	[2]mutant vs WT	comments
primary and secondary promoters						
groES	HP0011	13.660	jhp0009	17.028	up	
hrcA	HP0111	<8.000	jhp0103	16.614	up	ATGC motif instead of GCGC motif in *H. pylori* 26695
	HP0174	15.387	jhp0160	14.319	down	
addB	HP0275	17.261	jhp0260	15.987	up	
mraY	HP0493	18.213	jhp0445	18.786	no	divergent gene (jhp0444) is down-regulated in mutant
feoB	HP0687	17.447	jhp0627	17.509	down	
	HP0813	18.451	jhp0749	15.246	down	
	HP0895	18.377	jhp0832	18.092	down	
	HP0920	19.200	jhp0854	18.583	no/down	downstream gene (*carB*) is down-regulated in mutant
	HP1175	20.931	jhp1102	20.847	down	
cah	HP1186	17.271	jhp1112	18.241	down	
trmU	HP1335	15.240	jhp1254	15.185	down	
	HP1365	17.935	jhp1283/ jhp1443	15.899	both down	
	HPt26 (tRNA-Pro)	20.700	jhp_t12	20.015	NR/down	downstream gene (*pyrD*) is down-regulated in mutant
internal promoters						
kdsB	HP0230	18.299	jhp0215	10.708	no/up	357 nts from end of gene; downstream gene (*hopA*) is up-regulated in mutant
mraW	HP0707	18.201	jhp0646	18.012	no/down	18 nts upstream of next gene in operon (jhp0647), which is down-regulated in mutant
coaX	HP0862	17.776	jhp0796	13.761	down	249 nts from end of gene
kgtP	HP1091	13.971	jhp0334	13.868	up	1012 nts from end of gene
antisense promoters						
	HP0728	17.674	jhp0665	17.644	down	320 nts from start of gene
hopG	HP0914	19.674	jhp0850	19.707	up	1476 nts from start of gene
	HP0951	16.555	jhp0885	18.068	down	470 nts from start of gene
kgtP	HP1091	16.813	jhp0334	14.903	up	137 nts from start of gene

[1]Only first gene in operon is indicated.

[2]Indicates if expression of the gene was up-regulated (up), down-regulated (down), or unaffected (no) in the *H. pylori* J99 M.Hpy99III (*jhp1050*) mutant in the RNA-seq assay (21). Data for tRNAs were not reported (NR) for the RNA-seq assay.

Table 2.3: Growth rates of M.Hpy99III-deficient mutants in various *H. pylori* strains.

	Doubling Times (h)	
***H. pylori* strain**	**Wild type**	**M.Hpy99III-deficient mutant**
G27	4.1 ± 0.94	8.6 ± 1.8*
B128	4.1 ± 0.57	6.6 ± 2.5
ATCC 43504	4.9 ± 0.53	4.1 ± 0.30
X47-AL2	5.1 ± 1.2	6.0 ± 0.65
26695	5.7 ± 1.8	8.1 ± 2.3*

*p-value < 0.05

Table 2.4: Proteins that appear to be differentially expressed in wild-type *H. pylori* G27 and the HPG27_1066 mutant.

Proteins that appear to be at lower levels in the HPG27_1066 mutant					
Gene	**Protein function**	**[a]WT score**	**WT total peptides**	**[a]Mutant score**	**Mutant total peptides**
ureB	urease beta subunit	1753.85	28	175.83	4
katA	catalase	960.21	18	82.68	3
rplB	ribosomal protein L2	871.71	17	497.14	10
HPG27_1499	iron (III) ABC transport protein	583.97	11	0	0
rpsA	ribosomal protein S1	509.86	11	191.88	5
cagA	exotoxin CagA	401.1	10	115.56	3
rnj	ribonuclease J	364.05	10	187.28	6
frdA	fumarate reductase subunit	111.04	3	0	0
fusA	elongation factor G	90.87	3	0	0
Proteins that appear to be at elevated levels in the HPG27_1066 mutant					
hopU	outer membrane protein	770.16	10	1199.21	18
HPG27_1436	TonB-dependent receptor	112.44	2	1086.13	21
sabB	outer membrane protein	0	0	429.06	9
rpoB	RNA polymerase subunit	144.56	4	313.88	9
HPG27_1459	iron (III) dicitrate transport protein	123.96	4	235.85	8
babA	outer membrane protein	30.88	1	215.11	6
mcpB	methyl-accepting chemotaxis protein	68.73	2	199.3	5
cagZ	sodium:calcium antiporter	0	0	117.82	5
guaB	IMP dehydrogenase	0	0	117.68	3

[a]Scores are derived from total number of peptides identified, the quality of the agreement between the experimental/theoretical spectra, and the coverage of protein length.

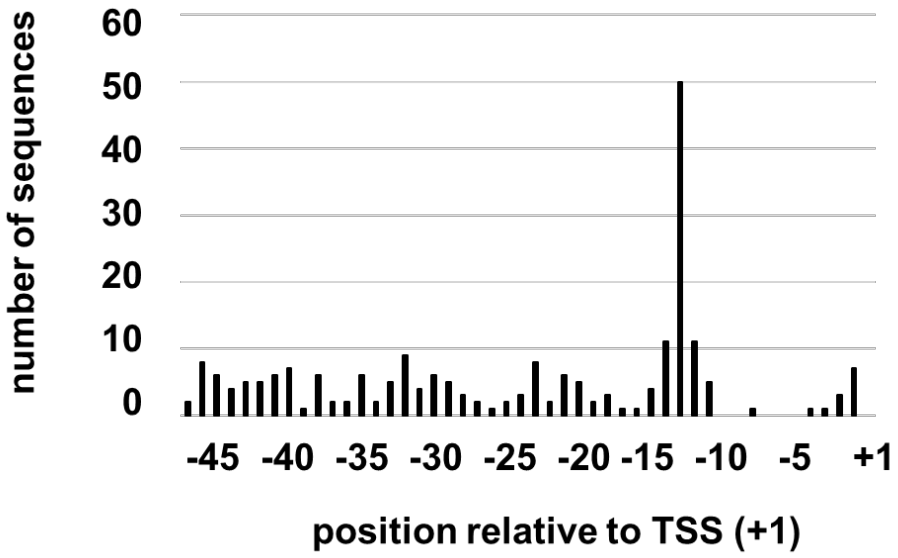

Figure 2.1: Positional distribution of GCGC motifs in *H. pylori* 26695 promoter regions. The position of the 3'-end of the GCGC motif relative to the transcriptional start site (TSS) was determined for 211 putative promoter sequences that contain a GCGC motif within 50 nucleotides of the TSS. There were 217 non-overlapping GCGC motifs and 7 overlapping GCGC motifs (i.e., GCGCGC) in the promoter sequences.

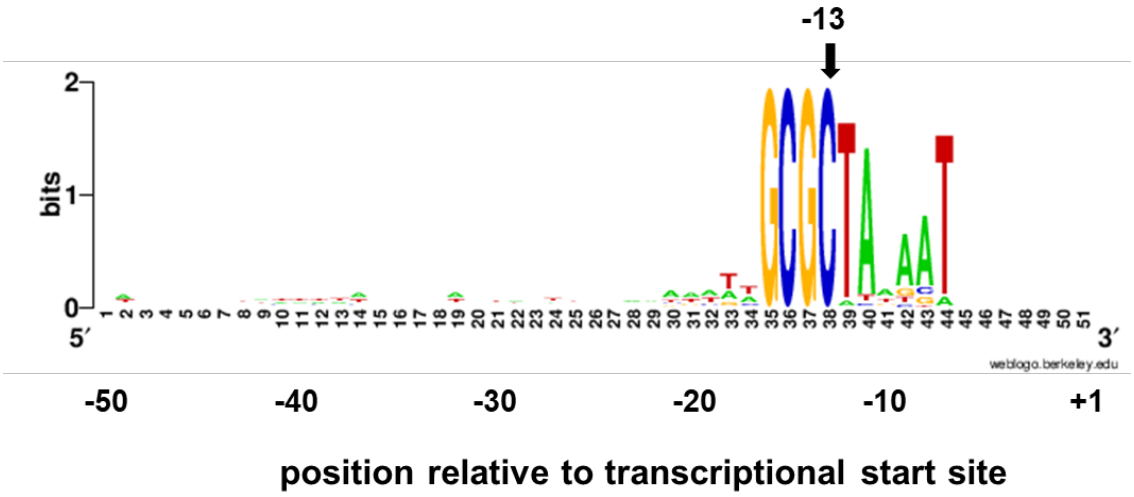

Figure 2.2: **Conservation of *H. pylori* 26695 promoters with a GCGC motif near position -13.** Promoter sequences were inferred from the TSS database reported by Sharma and co-workers (34). The 3'-ends of the GCGC motifs in the promoters were positioned from -11 to -15 relative to the TSS, with -13 being the most common position. The GCGC motifs were aligned for 80 promoter sequences and from the alignment a sequence logo was generated using WebLogo (https://weblogo.berkeley.edu/logo.cgi) (53).

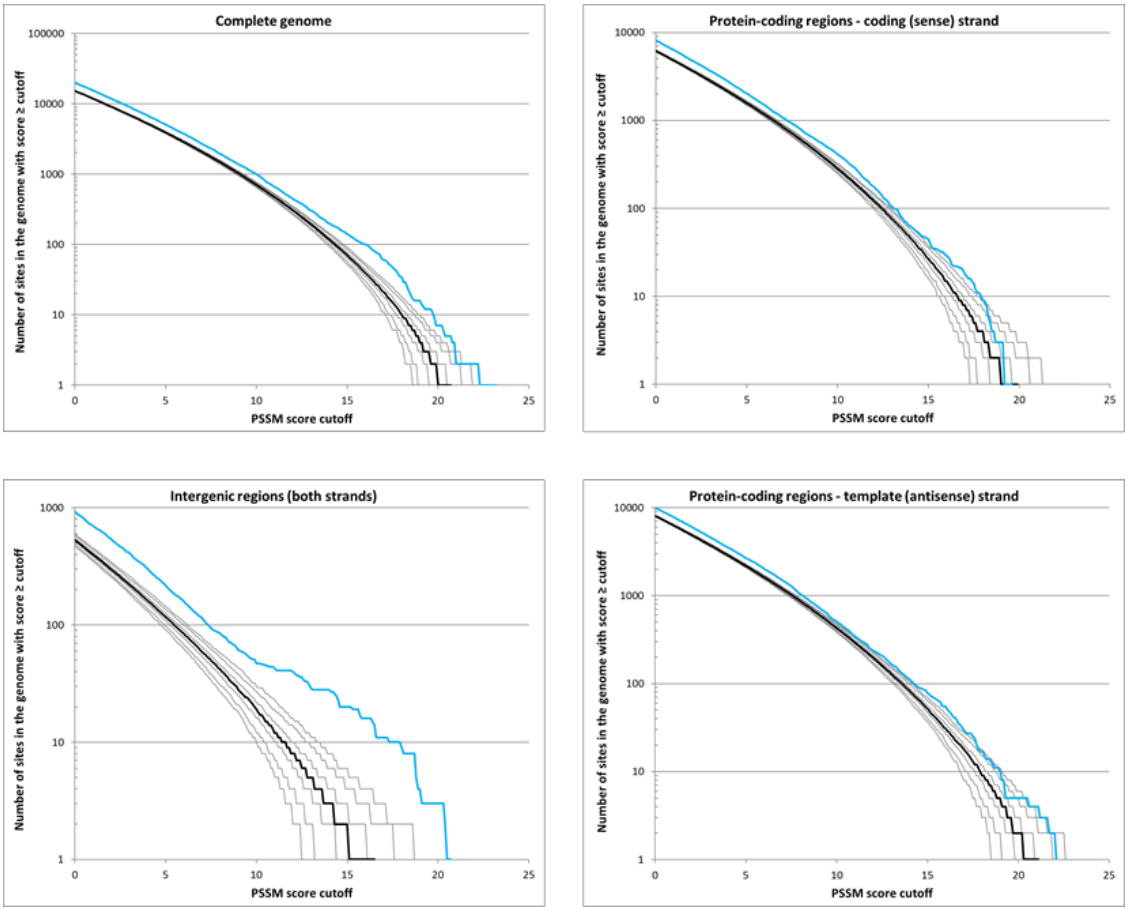

Figure 2.3: Reverse cumulative distributions of PSSM scores for sequences that resemble GCGC-containing promoters in *H. pylori* 26695 genome. The blue line represents the observed values for the *H. pylori* 26695 genome and the black lines represent the random sequences. The y-axis indicates the number of sites in the genome with scores greater or equal to the PSSM score cutoff indicated on the x-axis. The thick black line represents the median for 1000 random sequences and the thin black lines correspond to the 90th, 95th and 99th percentiles.

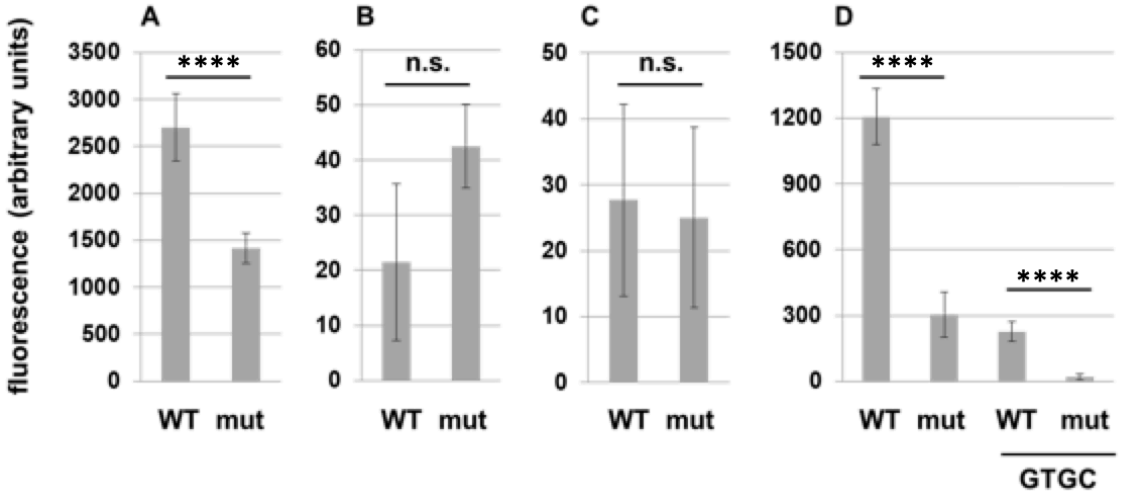

Figure 2.4: Expression of GFP reporter genes bearing putative GCGC-promoters from *H. pylori* G27 in wild type and HPG27_1066 mutant. (**A**) GFP reporter gene containing *icd* (HPG27_24) promoter; (**B**) GFP reporter gene containing *cah* (HPG27_1129) promoter; (**C**) GFP reporter gene containing HPG27_865 promoter; (**D**) GFP reporter genes containing the wild-type HPG27_846 promoter or the promoter where GCGC motif was altered to GTGC. Background fluorescence is subtracted out from data. Asterisks (****) denote differences in GFP fluorescence that were statistically significant (*p*-value <0.0001), while n.s. indicates differences that were not significant. All strains were evaluated using three technical replicates consisting of three biological replicates each. Statistical significance was determined using a two-tailed Student's t-test.

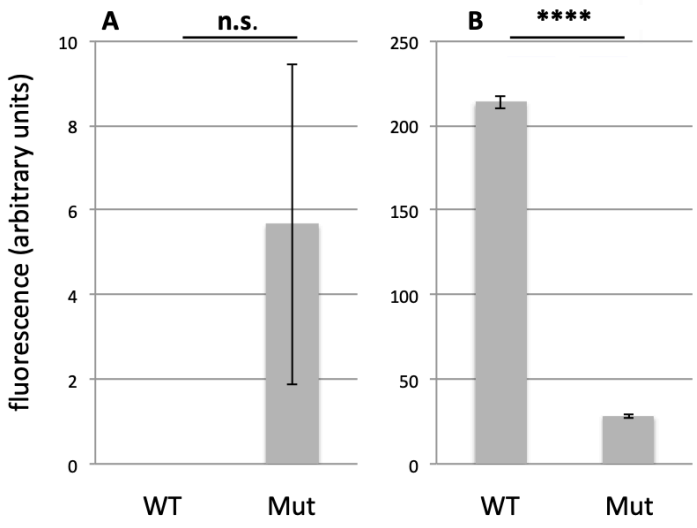

Figure 2.5: Expression of GFP reporter genes in *H. pylori* G27 wild type and HPG27_1066 mutant bearing predicted GCGC-promoters from differentially expressed genes identified in *H. pylori* J99 *jhp1050* mutant. (**A**) GFP reporter gene containing *kgtP* (JHP_0334) promoter; (**B**) GFP reporter gene containing JHP_0160 promoter; Background fluorescence is subtracted out from data and the GFP activity from the WT strain expressing the *kgtP* was indistinguishable from the normal background fluorescence. Asterisks (****) denote differences in GFP fluorescence that were statistically significant (p-value <0.0001), while n.s. indicates differences that were not significant. All strains were evaluated with three replicates. Statistical significance was determined using a two-tailed Student's t-test.

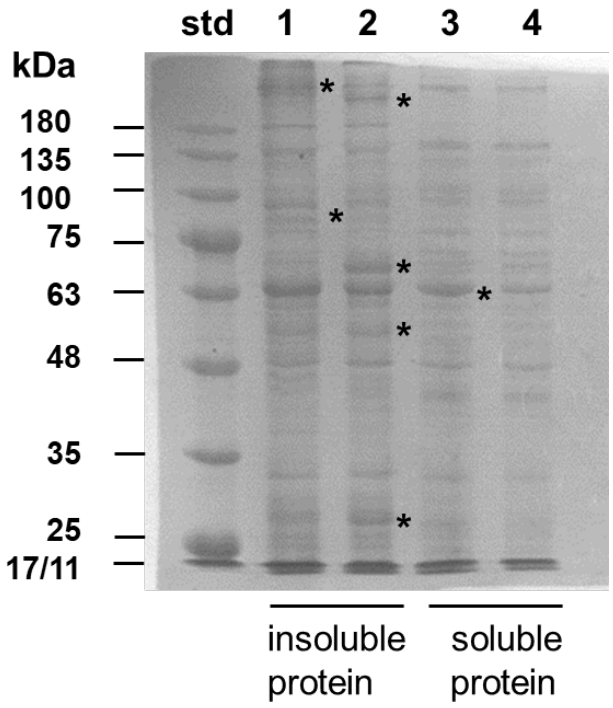

Figure 2.6: SDS-PAGE of soluble and insoluble protein fractions of wild-type *H. pylori* G27 and the HPG27_1066 mutant. Crude cell extracts of the wild-type and mutant strains were subjected to high-speed centrifugation to separate soluble proteins (cytosolic and periplasmic proteins) from insoluble proteins (i.e., ribosome, membrane-bound and membrane-associated proteins). 10 μg of each protein sample was loaded onto a 12% SDS-PAGE gel. Lane 1, insoluble, mutant protein fraction; lane 2, insoluble, wild-type protein fraction; lane 3 soluble, mutant protein fraction; lane 4, soluble, wild-type protein fraction. Asterisks indicate protein bands that appeared to differ in their intensities between the wild-type and mutant samples. Protein ladder (std) was BLUEstain™ 3 (GoldBio) and masses of the protein standards are indicated.

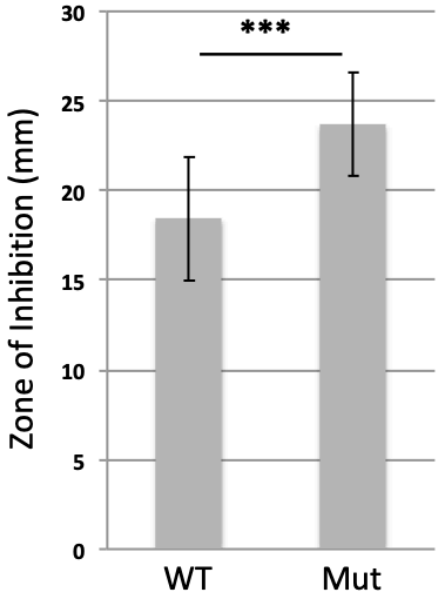

Figure 2.7: The HPG27_1066 mutant is more sensitive to hydrogen peroxide than the wild-type parental strain. *H. pylori* strains were tested for sensitivity to H_2O_2 by spreading cells on an agar medium, placing sterile filter disks on the agar medium, and then applying an H_2O_2 solution to the filters. After 2 days, zones of inhibition around the disks were measured. Asterisks (***) indicate results that are statistically significant at *p*-value <0.001. Three technical replicates were conducted with three biological replicates in each experiment. Statistical significance was determined using the two-tailed Student's t-test.

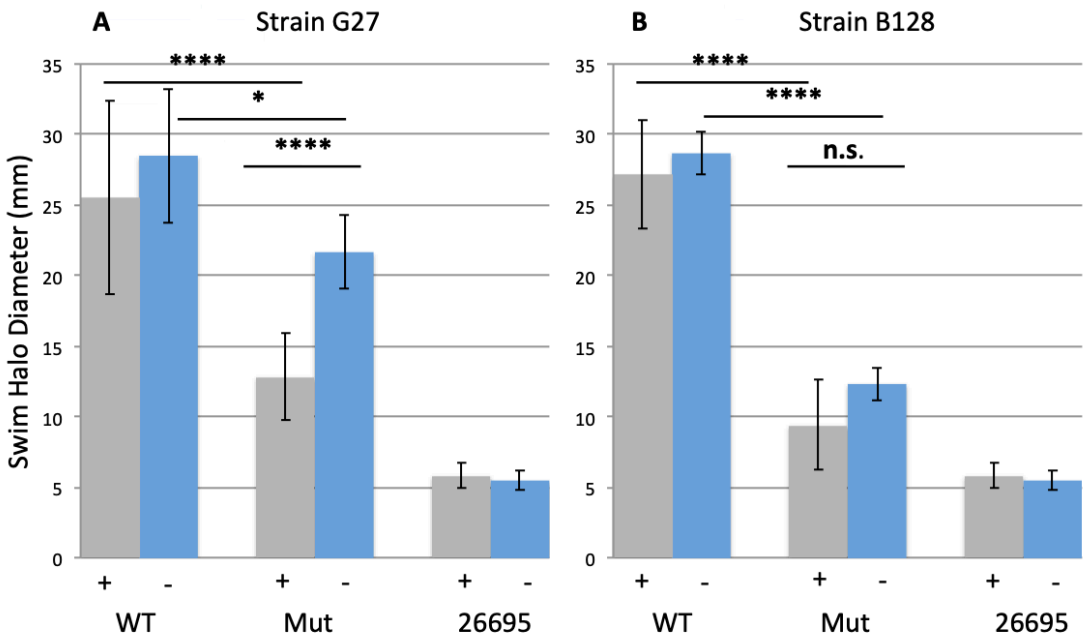

Figure 2.8: Motility of *H. pylori* wildtype and methylase deficient strains on soft agar with and without the addition of $FeSO_4$. Motility assay of (A) strain G27 and (B) strain B128 with accompanying swim halo diameters. (+) Indicates the addition of $FeSO_4$ and (-) represents omission of $FeSO_4$ from the agar medium. Where indicated, (WT) refers to the wildtype strain and (Mut) refers to the methylase deficient strain. At least three replicates were reported for studies including $FeSO_4$ and one to two replicates were reported for studies without $FeSO_4$. A two-tailed Student's t-test was used to calculate significance. Asterisks denote statistically significant differences in motility (**** p-value <0.0001, *p-value <0.02) while n.s. indicates differences that are not significant. *H. pylori* 26695 served as a negative control for the motility assay due to it being non-motile.

Table S2.1: Alignment of sequences of GCGC-containing promoters in *H. pylori* 26695 and corresponding sequences from *H. pylori* J99.

[1]gene	locus tag	[2]sequence (-50 to +1)
primary and secondary promoters		
groES	HP0011 jhp0009	CTTTGTTTTTATGGCTTGACTTATCCCTAAAAAT**GCGC**TATAGTTATGTCG GTTTGGTTTTATGGCTTGACTTATCCCTAAAAAT**GCGC**TATAGTTATGTCG
icd	HP0027 jhp0023	AAAAAAAGATTTTTAAGGGTTATATAGTATTTTT**GCGC**TAGTATAGTTACT AAAAAAAGATTTTTAAGGGTTATATAGTATTTTT**GCGC**TAGTATAGTTACT
hrcA	HP0111 jhp0103	TTTTTTAAAACAAGTGATAGATTATTAACTTTTT*ATGC*TATAATGCGAGGG CTTGTTTAAAAAAGTGATAGATTATTAACTTTTT**GCGC**TATAATGCGAGGG
	HP0170 jhp0156	TTTAGAGGGCGTTTTGATTGCTAAAGCTTAAGT**GCGC**TAAAATGACAGCCG TTTAGAGGGCGTTTTGATTGCTAAAGCTTAAGT**GCGC**TAAAATGATAGCCG
	HP0174 jhp0160	TTGATAAGCTTACTATGTAAGATTAAGGGACTTT**GCGC**TTAAATTACCCTA TTGATAAGCTTACTATATAAGATTAAGGGACTTT**GCGC**TTAAATACCCCTT
addB	HP0275 jhp0260	CGGGCATTACACAAAAAACCAAAGAAACTAAAGT**GCGC**TAAAATTCACTTC CGGGCATCACACAAAAAACCAAAGAAACTAAAGT**GCGC**TAAAATTCACTAT
	HP0316	TTTTTCTAACCTTTTTAATCCCACTCTATCGCTT**GCGC**TATAATGAATTGG
mraY	HP0493 jhp0445	CTACCCTATTCATGCTATTTTTAAGTAAATCAATTT**GCGC**TACAATTTTCT AAACCGTAATTATACTATTTTTAAGCAAATCAATTT**GCGC**TACAATTTTCC
	HP0565 jhp0512	TAATAGTAATGATTTTGTTGCAATTAAACATAAT**GCGC**TATCATTTCTAGG TAATTAAAATGATTTTGTCGTAATTAAACATAAT**GCGC**TATCATTCCTAGG
	HP0624 jhp0568	TAGCATTAAACCTCTTCATCAGTTAAATTCTTAT**GCGC**TATTATACTCTTT TAGCATCAAACCTCTTCATCAGTTAAATTCTTAT**GCGC**TATTATACTCTTT
feoB	HP0687 jhp0627	GTCTCGCTTTTATTATATAACAATTAGTCTTATA**GCGC**TATAGTGCTGACT GTCTCGCTTTTATTATATAACAATTAGTCTTATA**GCGC**TATAGTGCCAACT
prsA	HP0742 jhp0679	GCTTGGTTGAAAACAAACGCTTTTAGTTTGGGT**GCGC**TACAATTGATTTTA GCTTGGTTGAAAACAAACGCTTTTAGTTCGGGT*GTGT*TACAATTGATTTTA
lex2B	HP0805 jhp0741	ATCATCTAACATGATGCGTATTTTACTAAAATTT**GCGC**TAGAATGTAAGTC ATTATCTAGCATGATGCGTATTTTACTAAAATTT**GCGC**TAGAATGTAAGGC
	HP0813 jhp0749	TAGAAGCGCTATTCTAGTCAAAATCGCCTTATTTTT**GCGC**TATGATTTTCA AGAGACCGCTATTCTAGTCAAAATCGCCTTATTTTT**GCGC**TATGATATTAT
	HP0893 jhp0830	CCACACCCCTTATAACGCTTAAACCAAATCGCTT**GCGC**TATAATGAACTGA ATTTTCTAGCCTTTTTAATACCGCTCTATCGCTT**GCGC**TATAATGAATCGG
	HP0895 jhp0832	TTTTTCTAACCTTTTTAATCCCACTCTATCGCTT**GCGC**TATAATGAATTGG ATTTTCTAGCCTTTTTAATACCGCTCTATCGCTT**GCGC**TATAATGAATCGG
hypB	HP0900 jhp0837	AAAAAGTCGCTAAATCATAGTCGTTTAATTAAGT**GCGC**TACGATACGAAAA AAGAAGTCGTTTAATGATAGTCGTTTAATTAAG*ACGC*TACAATACGAAAA
	HP0920 jhp0854	TTAGGGGGTTTTGTTTGGTATTTTACTACATTTT**GCGC**TATTATATTGGAC TTAGGGGGTTTTGTTTGGTATTTTACTACATTTT**GCGC**TATTATATTGAAC
	HP1079 jhp0346	CTCAAGTTCTTTTAAAATTTTGTCGTAATTGCAT**GCGC**TATAATGCCAAAA ATTTATTTCCTATTAAATTTTTTCGTAATCGCAT*GCGT*TATAATATGAAAA
fldA	HP1161 jhp1088	AAAATGGGTTTTTAAGCAAGATGTTTAATTAAAT**GCGC**TAGACTACGCCCA AAAATGGGTTTTTAAGCAAGATGTTTAATTAAAT**GCGC**TAGACTACGCCCA
	HP1175 jhp1102	AATCAAGCATCTTAAGGCAATTTAAAATAAAATA**GCGC**TAGCATGACCCTA AATCAAGCGTCTTAAGGCAATTTAAAATAAAATA**GCGC**TAACATGACCCCA
cah	HP1186	ATCTATTTTTTGTAACTGCGGTCATTGTGGATTAG**GCGC**TAGAATTAATTG

	jhp1112	ATCTATTTTTTGTAACTGCGGTCATTATTGATTAA**GCGC**TAGAATTAATTG
	HP1321 jhp1241	TTTTACTTCCTTTTTATAAGGAAATATAAAAAAT**GCGC**TACGATTATCTTA TTTTACTTCCTTTTTATAAGGAAATAGAAAAAAAT**GCGC**TACGATTGTCTT
trmU	HP1335 jhp1254	AATGGTTAAAAAAAGGTGGTTATGAATGGTTTTT**GCGC**TAGACTACGAGCC AATGGTTAAAAAAAGGTGGTTATGAATGGTTTTT**GCGC**TAGACTACGAGCC
	HP1365 jhp1283 / jhp1443	TTTTTTATAACCGCTTTTAGTATAAAGGCAATTA**GCGC**TACAATACCCCCA CATGATCGCTTTTGGTATAAAGGCTATAGAATTA**GCGC**TACAATACCCCCA CATGATCGCTTTTGGTATAAAGGCTATAGAATTA**GCGC**TACAATACCCCCA
mod	HP1368	CATAGGAAGTATTATAACAAAGCCCGCTCATTAA**GCGC**TACAATACCCCCA
	HPt26 (tRNA-Pro) jhp_t12	AAAAAAGAGTTTTTAAACAACAACTAAGCAAACTC**GCGC**TACAATCTCCCT AAAAAAGAGTTTTTAAACAACAACTAAGCAAACTC**GCGC**TACAATACCCCT
internal promoters		
glyA	HP0183 jhp0171	TAACGAATAAATACGCTGAAGGCTACCCTAACAA**GCGC**TATTATGGAGGCT TAACGAATAAATACGCTGAAGGCTATCCTAACAA*ACGC*TATTATGGGGGCT
kdsB	HP0230 jhp0215	AAGCCACCAAAAACGCCCCTTTCATGGCGACTT**GCGC**TAAAGTTATTGATG AAGCCACCCAAAACGCCCCTTTCATGGCGACTT**GCGC**CAAAGTCATTGATG
ycf5	HP0378 jhp1003	TAATGAGGCTCCAATTGACATGCATGGGGGAAAAA**GCGC**TAAAATTGAGCG CAATGAAGGCCAAATTGACATGCATGGGGGCAAAA**GCGC**TAAAATAGAACG
hsdR	HP0464 jhp0416	TAAAAGTCTCACGCGCCATTAACACGCTCAAAGA**GCGC**TACAATTTAATCC TAAAAGTCTCACGCGCCATTAACACGCTTAAAGA**GCGC**TACAATTTAATCC
	HP0571 jhp0518	CATGCGTTAGGGTGGCTTAAAAAACACCCTTAT**GCGC**TAATATTACTATTA CATGCGTTAGAGTGGCTTAAAAAACACCCTTAT**GCGC**TAATATTACTATTA
	HP0580 jhp0527	AGCTCCAACCAAGCCTAACCCCCTTTAAAGATT**GCGC**TATTATGGCGTTTA AGCTCCAACCAAGCTTAACCCCCTTTAAAGACT**GCGC**TGTCATGGCGTTTA
	HP0668	CAGCGTTAGATAGCATCGTCTTTTTTGATGGCAAAA**GCGC**TATGGTGGATA
mraW	HP0707 jhp0646	CCAGAAGAAATTAAAAACAACAGGCGTTCACGAA**GCGC**TAAAATGAGGGTG CCAGAAGAAATTAAAAACAACAGGCGTTCACGAA**GCGC**TAAAATGAGGGTG
	HP0760 jhp0697	AATACGCTAAACGCATGCAAGCTTTAGAAGAGATC**GCGC**TAGAATTTGATG AATACGCTAAACGCATGCAAGCTTTAGAAGAGATC*GCGT*TAGAATTTGATG
thiB	HP0843 jhp0781	CGCGCCTTTTATTGTCAATGATGAGGTGCAACTC**GCGC**TAGAATTAAAGGC CACGCCTTTTATTATTAATGATGAGGTGCGACTC**GCGC**TAGAATTAAAGGC
coaX	HP0862 jhp0796	GGTTTAGCCCAATATATTCATGCGTATAAAAAAA**GCGC**TAAAATTTTAGAG GGTTTAGCCCAATATGTCCATGCGTATAAAAAAA**GCGC**GAAAATCTTAGAG
rarR	HP1026 jhp0398	CGACGCTAGAGCGTTATTAAACCTTTTAGATTTGA**GCGC**TAAAATAGAAGA CGATGCCAGAGCGTTATTAAACCTTTTAGATTTGA**GCGC**TAAAATAGAAAA
kgtP	HP1091 jhp0334	AAAACTTCTATGGTGTATTCCATTATCCTTATG**GCGC**TAGGCTCTTTTTTG AAAACTTCCATGGTGTATTCCATTATCCTTATG**GCGC**TAGGCTCTTTCATG
pyrE	HP1257 jhp1178	GACATTATCACTACGGGAAAATCCGCTATGGAAT**GCGC**TAAAGTTTTAGAA GACATCATCACTACGGGAAAATCCGCTATGGAAT**GCGC**TAAAGTTTTAGAA
fecA	HP1400 jhp1426	ACTAGGGATTTTCAAGTGAGCTCTAGCTACAATA**GCGC**TAACATGGTTACT ACTAGGGACTTTCAGGTGAGCTCTAGCTACAATA**GCGC**TAACATGGTTACT
trbB	HP1421 jhp1316	GCGTGGGGCTTTTAGTGGGCAAGCAAGAAAACAC**GCGC**TTTAATTATGAAG GCGTGGGGCTTTTAGTGGGCAAGCAAGAAAACAC**GCGC**TTTAATTATGAAG
recG	HP1523 jhp1412	ATGAAAAATTTAGAGCGCAAAAAATTGCAATTTG**GCGC**TAAAATCGCATGC ATGAAAAATTTAGAGCGCAAAAAATTGCAATTCA*ACGC*CAAAATCGCATGC

leuS	HP1547 jhp1452	CAGCTTATCACTCAAGGCATGGTTTTAAAAAATG**GCGC**TAAGATGAGCAAG CAGCTTATCACTCAAGGCATGGTCTTAAAAGATG*GTGC*TAAGATGAGCAAA
antisense promoters		
putP	HP0055 jhp0047	AGGCACATCTCTAATGGAGCGGATAGACATGAA**GCGC**ACTAAAATATGGGG AGGCACATCTCTAATGGAGCGGATAGACATGAA**GCGC**ACTAAAATATGGGG
frdC	HP0193 jhp0179	TTGAGACTAAAAACATGTGTGCTATCATAAAGA**GCGC**TAAAATCAAGCCCG TAGAGACCAAAAACATGTGCGCTATCATAAAGA**GCGC**TAAAATCAAGCCCG
	HP0232 jhp0217	TTGCTTAACACGCGTTCATTGCCTTGCAAATC**GCGC**ACTATAATGTGGGTT TTGCTCAACACGCGTTCATTGCCTTGCAAATC**GCGC**ACTACAATGTGGGTT
	HP0271 jhp0256	CTTTTCGCATAAAACCTCTTTAGCTTTACTCATC**GCGC**TAAAATAAGTGTA CTTTTCGCATAAAACCTCTTTAGCTTTACTCATC**GCGC**TAAAATAGGTGTA
	HP0271 jhp0256	TGTAACTAAAACGCATGCTTTTCATTCTCCATGC**GCGC**TAAAATGTTCAAG TGTAACTAAAACGCATGCTTTTCATTCTCCATGC**GCGC**TAAAACATTCAAC
ibpB	HP0280 jhp0265	CCCAAACACCCCCTACCATAATTTTCATAATATT**GCGC**TAAAGTCGTGCCT CCCAAACACCCCCTACCATAATTTTCATAATATT**GCGC**CAAAGTCGTGCCT
dppA	HP0298 jhp0283	TCCATCTGGCGTTCAAACGAAAACAGCACATCTTTA**GCGC**TAAACTCTACT TCCATTTGGCGTTCAAAAGAAAACAGCACGTCTTTA**GCGC**TAAACTCTACT
nadE	HP0329 jhp0312	AGAATAATCGTATAAAAAAGCCATGCGCAATCTT**GCGC**AAAAATTCCCCTT AGAATAGTCGTATAAAAAAGCCATGCGTAACCTT**GCGC**AAAAATTCCCCTT
dsbC	HP0377 jhp1004	ATGCTCTTTGGAGTAGCTGATATTGACATAGTAA**GCGC**TAAAATGCTCTTT ATGCTCTTTAGAATAGCTGATATTGACATAGTAA*GCAC*TAAAATGCTCTTT
fliN	HP0584 jhp0531	TGCAACAAAGGGATTTGCGTAGAGCCAAGTTCT**GCGC**TAAAAACAATCTCC TGCAATAAAGGGATTTGCGTAGAGCCAAGTTCC**GCGC**TAAAAACAATCTCC
mod	HP0593	CCAAAATACCCCGTTTTCATATTCCCACTTTTT**GCGC**TAGGATTATCATTA
	HP0595 jhp0542	GTTGAAATCCGCCCCCACGAACGACCCCACCGTC**GCGC**TATAATAGCTATC GTTAAAATCCGCCCCCACGAAAGCCCCCACCGTC**GCGC**TGTAATAGCTATC
cheV	HP0616 jhp0559	CTGCACCACTCGTTCTTCATCAAAACGAGTGATA**GCGC**TAAGCTTACCCTC CTGCACCACTCGTTGTTCATCAAAACGGGTGATA**GCGC**TAAGCTTACCCTC
	HP0637 jhp0580	CCACTGAATTGTAAAAGCCGTTTAAAGAGCCATAA**GCGC**TATAATTAGTGA CCACTGAATTGTAAAAGCCGTTTAGAGAACCATAA*GCGT*TATAATTAGTGA
	HP0728 jhp0665	CATCAAAAACCATTCCAGCCTGTCATTCAAATGGT**GCGC**TAAAATCAAATG CATCAAAAACCATTCCAGCCTGTCATTCAAATGGT**GCGC**TAAAATCAAATG
	HP0745 jhp0682	GAATAACGGCTTTTTTCTTTTTTGACCGCTTTGA**GCGC**TATCATTTTTAAG GAATAGCGGCTTTTTTCTTTTTTAGCCGCTTTGA**GCGC**TATCATTTTTAAG
	HP0852 jhp0788	GCTCTTTAAAAAACCGCTGAATTTTTCATTTTTC**GCGC**TAAGCTCTGAAGC GCTCTTTAAAAAACCGCTGAATTTTTCATTTTTC**GCGC**TAAGCTCCAAAGC
murJ	HP0885 jhp0817	AACCCCCACCAAAAAAAACCACGCCTTTTGCAA**GCGC**TGTAAGATTAGATC AACCCCCACCAAAAAAAACCACGCCTTTTGCAA*GCGT*TGTAAGATTAAATC
vacA	HP0887 jhp0819	TTTAGCGTTGAAATCCACTCTCGTGGTGCGATCA**GCGC**TATCCTTATAGCT TTTAGCGTTGAAATTCACTCTCGTGGTGCGATCA**GCGC**TATCCTTATAGCT
	HP0910 jhp0846	TTTTGTAAGTCCAGGTTTTGGTTATAAGGGAATA**GCGC**TAAAATAGAGCCG TTTTGTAAATCTAAGTTTTGGTTATAAGGGAATA**GCGC**TAAAATAGAGCCG
rep	HP0911 jhp0847	TGGCGTGTGTTCAGGGTTTTTATTTGAAAGCCAT**GCGC**TAAAATCTTCTTG TGGCGCGTGTTCAGGGTTTTTACTAGAAAGCCAA**GCGC**TAAAATCTTCTTG
hopG	HP0914 jhp0850	GTTGGTCATCATGTAACTTCTGTCTTGGTAATTA**GCGC**TAAAATCGCCATC GTTGGTCATCATGTAACTTCTGTCTTGGTAATTA**GCGC**TAAAATCGCCATC
	HP0922 jhp0856	AGTGCTATTAAAGTTATAAGTCCCTTGATTGAAA**GCGC**TATGATTGAAGCT GGCGCTGTTAAAGTGATAAGTCCCTTGATTGAAA*TTGC*TGTTATTGAAGCT
	HP0936	AAAGTATAAAACTATGGTATTGACTAAACTTG**GCGC**GTTAGAATAAAATTC

alr	HP0941 jhp0876	GACACAAGCGTCTTTAGGGACAATGCTTTTGACT**GCGC**TAAAATTATGCCT GACACAAGCGTCTTTAGGGACAATATTTTTGACT*GCAT*TAAAATTATGCCT
	HP0951 jhp0885	AAATTTAGGCTTTTAGCTTTCAAACAAGAGGGGT**GCGC**TAGAATAAACCCT AAATCCAAGCTTTTAGCTTTCAAGCAAGAGGGGT**GCGC**TAAAATAAACCCT
	HP0971 jhp0905	GGTCAAACTCTCCCCTAAAGTCTCTTCTAAAA**GCGCGC**TAATATCAATGTA GGTCAAGCTCTCCCCTAAAGTTTCTTCTAAAA**GCGCGC**TAATATCAATGTA
bioA	HP0976 jhp0910	GCTTGCACAATAAAACGGCTTGTTTTAAATATCTT**GCGC**TATAAATATGCA GCTTGCACCATAAAACGGCTTGTTTTAAATATTTC**GCGC**TATAAATATGCA
copA	HP1072 jhp0353	AGCGATGACATGTTCGCTACTCTTTTCAATACTGA**GCGC**TAAACTCAATAA AGCAATGACATGTTCGCTGCTCTTTTCAATGCTGC*TCGC*TAAACTCAATAG
tktA	HP1088 jhp0337	AGATTCGCCTTTTTTAAAGGTGGGGTAATGGATG**GCGC**TAAAATCAAAATC AGATTCGCCTTTTTTAAAGGCGGGGTAATTAATG*GCGT*TAAAATCAAAATT
kgtP	HP1091 jhp0334	AACCCTAGCATGAAAACTAAAAAAGCTGAGATTA**GCGC**TAAAGTGGGGTCG AACCCTAGCATAAAAACTAAAAAAGCTGAGATGA**GCGC**TAGAGTAGGGTCA
valS	HP1153 jhp1080	ATACAAAGTCTTATACCCATCCATGCGTTTGTAAC**GCGC**TAAAATATCTTG ATACAAGGTCTTATACCCGTCCATGCGTTTATAGC*GCAC*TAAAATATCTTG
alaS	HP1241 jhp1162	CCCACAAAAACATTAGGTGCATAAGCGTTTAAAATA**GCGC**TAAAATCGGCG CCTACAAATTCATTAGGTGCATAAGCGTTTAAAATA**GCGC**TAAAATCGGCG
pcrA	HP1553 jhp1446	AATGATTTTAGACTCTACAAATTGAGACAATTTT**GCGC**TAGATTCCGTGCT AATGATTTTAGACTCTACCAATTGAGACAATTTC**GCGC**TAGATTCCGTGCT

[1]Only first gene in operon is indicated.

[2]Sequence for *H. pylori* 26695 in top line and corresponding sequence from *H. pylori* J99 in bottom line. GCGC motifs in the -13 region of the promoter are indicated in bold, and when the GCGC motif is absent the corresponding sequence is in italics.

Table S2.2: Sequenced comparison of GCGC-containing promoters from *H. pylori* strains 26695 and G27 included in *gfp* reporter genes.

gene	sequence (TSS and 50 nucleotides of upstream sequence)[1]
icd (26695) (G27)	AAAAAAAGATTTTTAAGGGTTATATAGTATTTTT**GCGC**TAGTATAGTTACT AAAAAAAGATTTTTAGGGGTTATATAGTATTTTT**GCGC**TAGTATAGTTACT
cah (26695) (G27)	ATCTATTTTTTGTAACTGCGGTCATTGTGGATTAG**GCGC**TAGAATTAATTG ATCTATTTTTTGTAACTGCGGTCATTGTTTATTGA**GCGC**TAGAATTAATTG
HP0893 HPG27_846 846 mutant	CCACACCCCTTATAACGCTTAAACCAAATCGCTT**GCGC**TATAATGAACTGA TCAAACCCCTTATAACGCTTAAACCAAATCGCTT**GCGC**TATAATGAGCTGA TCAAACCCCTTATAACGCTTAAACCAAATCGCTT**GTGC**TATAATGAGCTGA
HP0914 HPG27_865	GTTGGTCATCATGTAACTTCTGTCTTGGTAATTA**GCGC**TAAAATCGCCATC GTTGGTCATCATGTAACTTCTGTCTTGGTAATTA**GCGC**TAAAATCGCCATC
HP0174 jhp0160	TTGATAAGCTTACTATGTAAGATTAAGGGACTTT**GCGC**TTAAATTACCCTA TTGATAAGCTTACTATATAAGATTAAGGGACTTT**GCGC**TTAAATACCCCTT
kgtP (26695) (J99)	AACCCTAGCATGAAAACTAAAAAAGCTGAGATTA**GCGC**TAAAGTGGGGTCG AACCCTAGCATAAAAACTAAAAAAGCTGAGATGA**GCGC**TAGAGTAGGGTCA

[1]GCGC motifs are in bold-face. Nucleotide sequences in the *H. pylori* G27 sequences that differ from those in the *H. pylori* 26995 sequences are indicated in red. The mutation in the GCGC motif of the HPG27_846 promoter (846 mutant) is underlined.

Table S2.3: Primer sequences used in this study

Hp1121_up_F2	5'-GCCGATACTTGCTGTAATGGA-3'	Constructing PCR product for inactivation of M.*Hpy*99III
Hp1121_up_R2	5'-TTGCGATACCGTCGAGATGGCTTCATG ATTGATTTCTGC-3'	Constructing PCR product for inactivation of M.*Hpy*99III
Hp 1121_d_F2	5'-TACTAAAGGGAATGTAGATAAGCAAG C CATTTGCTATCCCAA-3'	Constructing PCR product for inactivation of M.*Hpy*99III
Hp 1121_d_R2	5'-GGGGAATGATTTGTTGGATGAG-3'	Constructing PCR product for inactivation of M.*Hpy*99III
ermB_OLF	5'-ATCTCGACGGTATCGCAA-3'	Constructing PCR product for inactivation of M.*Hpy*99III
ermB_OLR	5'-TTATCTACATTCCCTTTAGTA-3'	Constructing PCR product for inactivation of M.*Hpy*99III
HPG27_865Fwd	5'-TCGATCCTCGAGAGCGTGAGGTTGGTC ATCATGTAA -3'	Constructing GFP reporter plasmid for HPG27_865
HPG27_865Rev	5'-TACATTGCTAGCACCCAGATGGCGATT TTAGCGCTA-3'	Constructing GFP reporter plasmid for HPG27_865
JHP_0334_GFP_Fwd	5'-TCCAGCCTCGAGCATAAAAACCTTTTT TACCATTCT-3'	Constructing GFP reporter plasmid for JHP_0334
JHP_0334_GFP_Rev	5'-TACATTGCTAGCCAATGACCCTACTCT AGCGCTCAT-3'	Constructing GFP reporter plasmid for JHP_0334
JHP_0160_GFP_Fwd	5'-TCCAGCCTCGAGATTTAAAGTGGGTGA AAAATGTTC-3'	Constructing GFP reporter plasmid for JHP_0160
JHP_0160_GFP_Rev	5'-TACATTGCTAGCTTTAAGGGGTATTTA AGCGCAAAG-3'	Constructing GFP reporter plasmid for JHP_0160

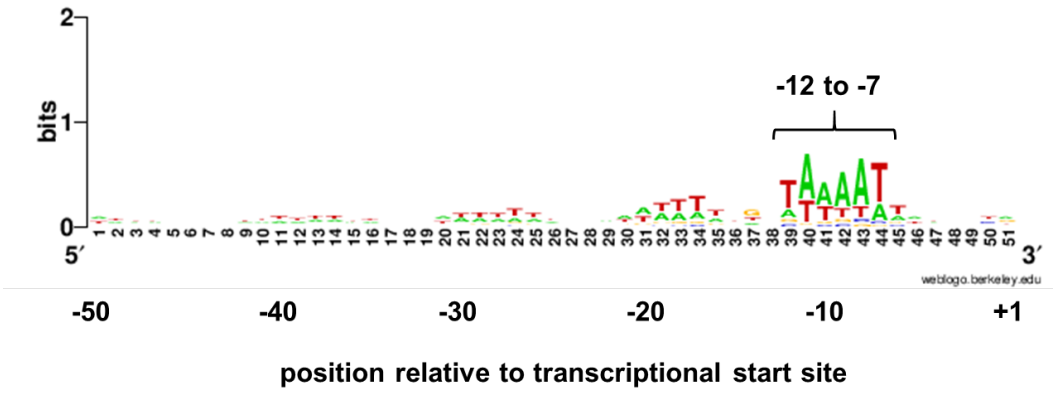

Figure S2.1: Sequence logo for predicted *H. pylori* 26695 promoters. WebLogo was used to generate a sequence logo from an alignment of 1914 promoter sequences identified from a *H. pylori* 26695 transcriptome analysis (34). The sequence logo shows a strong preference for A or T residues in positions -12 to -7.

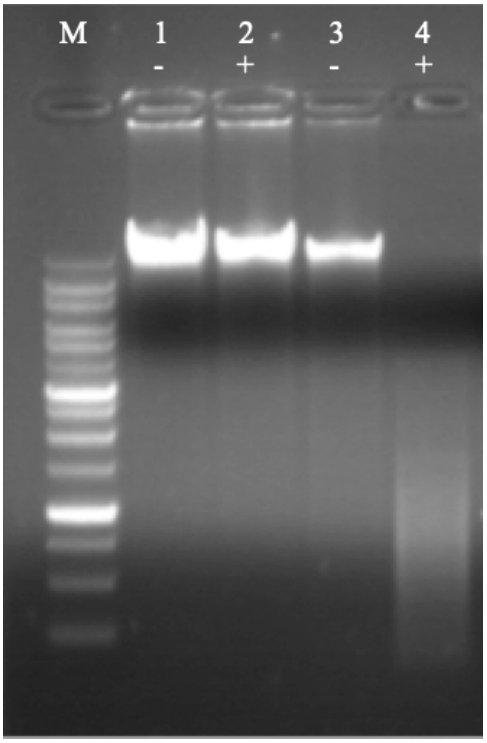

Figure S2.2: Digestion of gDNA from wild-type *H. pylori* G27 and the HPG27_1066 mutant with HinP1I. DNA from wild-type *H. pylori* G27 (lanes 1 and 2) or the HPG27_1066 mutant (lanes 3 and 4) was treated with (+) or without (-) HinP1I then subjected to agarose gel electrophoresis. Marker is 1kb PLUS™ DNA Ladder (GoldBio)

XhoI HPG27_846 promoter region

CTCGAGAGCTGTTTTAATCCACTCAAACCCCTTATAACGCTTAAACCAAATCGCTT**GCGC**TATAATGAGCTGATATTT

NheI *ureA* RBS start

TAAAGCTAGCAAGAATAGGAGAATGAG*ATG*GTGAGCAAAGGGGAAGAATTATTTACCGGGGTGGTGCCTATTTTAGTG

GAATTAGATGGGGATGTGAATGGGCATAAATTTAGCGTGAGCGGGGAAGGGGAAGGGGATGCTACCTATGGGAAATTA

ACCTTAAAATTTATTTGCACCACCGGGAAATTACCTGTGCCTTGGCCTACCTTAGTGACCACCTTAACCTATGGGGTG

CAATGCTTTAGCAGGTATCCTGATCATATGAAACAACATGATTTTTTTAAAAGCGCTATGCCTGAAGGGTATGTGCAA

GAAAGGACCATTTTTTTTAAAGATGATGGGAATTATAAAACCAGGGCTGAAGTGAAATTTGAAGGGGATACCTTAGTG

AATAGGATTGAATTAAAAGGGATTGATTTTAAAGAAGATGGGAATATTTTAGGGCATAAATTAGAATATAATTATAAT

AGCCATAATGTGTATATTATGGCTGATAAACAAAAAAATGGGATTAAAGTGAATTTTAAAATTAGGCATAATATTGAA

GATGGGAGCGTGCAATTAGCTGATCATTATCAACAAAATACCCCTATTGGGGATGGGCCTGTGTTATTACCTGATAAT

CATTATTTAAGCACCCAAAGCGCTTTAAGCAAAGATCCTAATGAAAAAAGGGATCATATGGTGTTATTAGAATTTGTG

stop BamHI

ACCGCTGCTGGGATTACCTTAGGGATGGATGAATTATATAAA*TAA*GGATCC

Figure S2.3: Template used for constructing *gfp* reporter genes. The DNA sequence indicated above was synthesized by Genewiz and cloned into plasmid pUC57-Kan. Unique XhoI, NheI and BamHI restriction sites (underlined) were included to facilitate introduction of the reporter gene into other plasmids and swapping promoter regions. A 76-bp DNA sequence containing the HPG27_846 promoter region and flanked by XhoI and NheI sites was included in the synthesized gene. Start and stop codons for *gfp* are indicated in italics. Codon usage of *gfp* was optimized for *H. pylori*. The ribosome binding site of *H. pylori ureA* (in red) was introduced between the NheI site and start codon of *gfp*.

References

1. **Amieva MR, El-Omar EM.** 2008. Host-bacterial interactions in *Helicobacter pylori* infection. Gastroenterology **134:**306-323.
2. **Atherton JC, Blaser MJ.** 2009. Coadaptation of *Helicobacter pylori* and humans: ancient history, modern implications. J Clin Invest **119:**2475-2487.
3. **Blaser MJ.** 1993. *Helicobacter pylori*: microbiology of a 'slow'bacterial infection. Trends Microbiol. **1:**255-260.
4. **Cover TL, Blaser MJ.** 1992. *Helicobacter pylori* and gastroduodenal disease. Annu. Rev. Med. **43:**135-145.
5. **Kuipers EJ.** 1997. *Helicobacter pylori* and the risk and management of associated diseases: gastritis, ulcer disease, atrophic gastritis and gastric cancer. Aliment Pharmacol Ther 11 Suppl **1:**71-88.
6. **Eaton KA, Brooks CL, Morgan DR, Krakowka S.** 1991. Essential role of urease in pathogenesis of gastritis induced by *Helicobacter pylori* in gnotobiotic piglets. Infect. Immun. **59:**2470-2475.
7. **Eaton KA, Morgan DR, Krakowka S.** 1992. Motility as a factor in the colonisation of gnotobiotic piglets by *Helicobacter pylori*. J. Med. Microbiol. **37:**123-127.
8. **Harris AG, Wilson JE, Danon SJ, Dixon MF, Donegan K, Hazell SL.** 2003. Catalase (KatA) and KatA-associated protein (KapA) are essential to persistent colonization in the *Helicobacter pylori* SS1 mouse model. Microbiology **149:**665-672.

9. **Krebes J, Morgan RD, Bunk B, Sproer C, Luong K, Parusel R, Anton BP, Konig C, Josenhans C, Overmann J, Roberts RJ, Korlach J, Suerbaum S.** 2014. The complex methylome of the human gastric pathogen *Helicobacter pylori*. Nucleic Acids Res **42:**2415-2432.

10. **Roberts RJ, Vincze T, Posfai J, Macelis D.** 2015. REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res **43:**D298-299.

11. **Vasu K, Nagaraja V.** 2013. Diverse functions of restriction-modification systems in addition to cellular defense. Microbiol Mol Biol Rev **77:**53-72.

12. **Loenen WA, Dryden DT, Raleigh EA, Wilson GG, Murray NE.** 2014. Highlights of the DNA cutters: a short history of the restriction enzymes. Nucleic Acids Res **42:**3-19.

13. **Low DA, Casadesus J.** 2008. Clocks and switches: bacterial gene regulation by DNA adenine methylation. Curr Opin Microbiol **11:**106-112.

14. **Wion D, Casadesus J.** 2006. N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. Nat Rev Microbiol **4:**183-192.

15. **Tsuda M, Karita M, Nakazawa T.** 1993. Genetic transformation in *Helicobacter pylori*. Microbiol Immunol **37:**85-89.

16. **Cheng X.** 1995. DNA modification by methyltransferases. Curr Opin Struct Biol **5:**4-10.

17. **Furuta Y, Namba-Fukuyo H, Shibata TF, Nishiyama T, Shigenobu S, Suzuki Y, Sugano S, Hasebe M, Kobayashi I.** 2014. Methylome diversification through

changes in DNA methyltransferase sequence specificity. PLoS Genet **10:**e1004272.

18. **Lee WC, Anton BP, Wang S, Baybayan P, Singh S, Ashby M, Chua EG, Tay CY, Thirriot F, Loke MF, Goh KL, Marshall BJ, Roberts RJ, Vadivelu J.** 2015. The complete methylome of *Helicobacter pylori* UM032. BMC Genomics **16:**424.

19. **Srikhanta YN, Gorrell RJ, Steen JA, Gawthorne JA, Kwok T, Grimmond SM, Robins-Browne RM, Jennings MP.** 2011. Phasevarion mediated epigenetic gene regulation in *Helicobacter pylori*. PLoS One **6:**e27569.

20. **Vale FF, Megraud F, Vitor JM.** 2009. Geographic distribution of methyltransferases of *Helicobacter pylori*: evidence of human host population isolation and migration. BMC Microbiol **9:**193.

21. **Estibariz I, Overmann A, Ailloud F, Krebes J, Josenhans C, Suerbaum S.** 2019. The core genome m5C methyltransferase JHP1050 (M.Hpy99III) plays an important role in orchestrating gene expression in *Helicobacter pylori*. Nucleic Acids Res **47:**2336-2348.

22. **Baubec T, Schubeler D.** 2014. Genomic patterns and context specific interpretation of DNA methylation. Curr Opin Genet Dev **25:**85-92.

23. **Klose RJ, Bird AP.** 2006. Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci **31:**89-97.

24. S**mith ZD, Meissner A.** 2013. DNA methylation: roles in mammalian development. Nat Rev Genet **14:**204-220.

25. **Bergbauer M, Kalla M, Schmeinck A, Gobel C, Rothbauer U, Eck S, Benet-Pages A, Strom TM, Hammerschmidt W.** 2010. CpG-methylation regulates a class of Epstein-Barr virus promoters. PLoS Pathog **6:**e1001114.

26. **Bhende PM, Seaman WT, Delecluse HJ, Kenney SC.** 2004. The EBV lytic switch protein, Z, preferentially binds to and activates the methylated viral genome. Nat Genet **36:**1099-1104.

27. **Lin LF, Posfai J, Roberts RJ, Kong H. 2001**. Comparative genomics of the restriction-modification systems in *Helicobacter pylori*. Proc Natl Acad Sci U S A **98:**2740-2745.

28. **Vitkute J, Stankevicius K, Tamulaitiene G, Maneliene Z, Timinskas A, Berg DE, Janulaitis A.** 2001. Specificities of eleven different DNA methyltransferases of *Helicobacter pylori* strain 26695. J Bacteriol **183:**443-450.

29. **Kumar R, Mukhopadhyay AK, Ghosh P, Rao DN.** 2012. Comparative transcriptomics of *H. pylori* strains AM5, SS1 and their hpyAVIBM deletion mutants: possible roles of cytosine methylation. PLoS One **7:**e42303.

30. **Kahramanoglou C, Prieto AI, Khedkar S, Haase B, Gupta A, Benes V, Fraser GM, Luscombe NM, Seshasayee AS.** 2012. Genomics of DNA cytosine methylation in *Escherichia coli* reveals its role in stationary phase transcription. Nat Commun **3:**886.

31. **Militello KT, Simon RD, Qureshi M, Maines R, VanHorne ML, Hennick SM, Jayakar SK, Pounder S.** 2012. Conservation of Dcm-mediated cytosine DNA methylation in *Escherichia coli*. FEMS Microbiol Lett **328:**78-85.

32. **Militello KT, Mandarano AH, Varechtchouk O, Simon RD.** 2014. Cytosine DNA methylation influences drug resistance in *Escherichia coli* through increased sugE expression. FEMS Microbiol Lett **350:**100-106.

33. **Chao MC, Zhu S, Kimura S, Davis BM, Schadt EE, Fang G, Waldor MK.** 2015. A cytosine methyltransferase modulates the cell envelope stress response in the cholera pathogen [corrected]. PLoS Genet **11:**e1005666.

34. **Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermuller J, Reinhardt R, Stadler PF, Vogel J.** 2010. The primary transcriptome of the major human pathogen *Helicobacter pylori*. Nature **464:**250-255.

35. **Aoyama T, Takanami M.** 1985. Essential structure of *E. coli* promoter II. Effect of the sequences around the RNA start point on promoter function. Nucleic Acids Res **13:**4085-4096.

36. **Jeong W, Kang C.** 1994. Start site selection at lacUV5 promoter affected by the sequence context around the initiation sites. Nucleic Acids Res **22:**4667-4672.

37. **Vvedenskaya IO, Vahedian-Movahed H, Zhang Y, Taylor DM, Ebright RH, Nickels BE.** 2016. Interactions between RNA polymerase and the core recognition element are a determinant of transcription start site selection. Proc Natl Acad Sci U S A **113:**E2899-2905.

38. **Walker KA, Osuna R.** 2002. Factors affecting start site selection at the *Escherichia coli* fis promoter. J Bacteriol **184:**4783-4791.

39. **Winkelman JT, Vvedenskaya IO, Zhang Y, Zhang Y, Bird JG, Taylor DM, Gourse RL, Ebright RH, Nickels BE.** 2016. Multiplexed protein-DNA cross-linking: Scrunching in transcription start site selection. Science **351:**1090-1093.

40. **Bono AC, Hartman CE, Solaimanpour S, Tong H, Porwollik S, McClelland M, Frye JG, Mrazek J, Karls AC.** 2017. Novel DNA Binding and Regulatory Activities for sigma(54) (RpoN) in *Salmonella enterica* Serovar Typhimurium 14028s. J Bacteriol **199:**e00816-00816.

41. **Bonocora RP, Smith C, Lapierre P, Wade JT.** 2015. Genome-Scale Mapping of *Escherichia coli* sigma54 Reveals Widespread, Conserved Intragenic Binding. PLoS Genet **11:**e1005552.

42. **Fitzgerald DM, Smith C, Lapierre P, Wade JT.** 2018. The evolutionary impact of intragenic FliA promoters in proteobacteria. Mol Microbiol **108:**361-378.

43. **Ishihama A, Shimada T, Yamazaki Y.** 2016. Transcription profile of *Escherichia coli*: genomic SELEX search for regulatory targets of transcription factors. Nucleic Acids Res **44:**2058-2074.

44. **Minch KJ, Rustad TR, Peterson EJ, Winkler J, Reiss DJ, Ma S, Hickey M, Brabant W, Morrison B, Turkarslan S, Mawhinney C, Galagan JE, Price ND, Baliga NS, Sherman DR.** 2015. The DNA-binding network of *Mycobacterium tuberculosis*. Nat Commun **6:**5829.

45. **Samuels DJ, Frye JG, Porwollik S, McClelland M, Mrazek J, Hoover TR, Karls AC.** 2013. Use of a promiscuous, constitutively-active bacterial enhancer-binding protein to define the sigma(5)(4) (RpoN) regulon of *Salmonella* Typhimurium LT2. BMC Genomics **14:**602.

46. **Mrazek J, Karls AC.** 2019. In silico simulations of occurrence of transcription factor binding sites in bacterial genomes. BMC Evol Biol **19:**67.

47. **Stead CM, Beasley A, Cotter RJ, Trent MS.** 2008. Deciphering the unusual acylation pattern of *Helicobacter pylori* lipid A. J Bacteriol **190:**7012-7021.

48. **Dantas Machado AC, Zhou T, Rao S, Goel P, Rastogi C, Lazarovici A, Bussemaker HJ, Rohs R.** 2015. Evolving insights on how cytosine methylation affects protein-DNA binding. Brief Funct Genomics **14:**61-73.

49. **Fisher EF, Caruthers MH.** 1979. Studies on gene control regions XII. The functional significance of a lac operator constitutive mutation. Nucleic Acids Res **7:**401-416.

50. **Mrazek J.** 2006. Analysis of distribution indicates diverse functions of simple sequence repeats in *Mycoplasma* genomes. Mol Biol Evol **23:**1370-1385.

51. **Mehta NS, Benoit SL, Mysore J, Maier RJ.** 2007. In vitro and in vivo characterization of alkyl hydroperoxide reductase mutant strains of *Helicobacter hepaticus*. Biochim Biophys Acta **1770:**257-265.

52. **Heuermann D, Haas R.** 1998. A stable shuttle vector system for efficient genetic complementation of *Helicobacter pylori* strains by transformation and conjugation. Mol Gen Genet **257:**519-528.

53. **Crooks GE, Hon G, Chandonia JM, Brenner SE.** 2004. WebLogo: a sequence logo generator. Genome Res **14:**1188-1190.

CHAPTER 3

CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions

Helicobacter pylori is an extraordinary pathogen that is able to colonize one of the harshest environments in the human body. From motility to urease activity, it is clear that the organism utilizes a variety of mechanisms to ensure its survival in the host. With the rise of antibiotic resistant bacteria and failing treatments against once treatable diseases, it has become a race against time to develop effective treatments against diseases that were once thought to be a minor problem. C-5 cytosine (^{m5}C) methylation in humans has been well characterized in regards to its effect on gene regulation, but our knowledge about methylation in bacteria is rudimentary. While the effects of ^{m5}C DNA methyltransferases (MTase), such as *ccrM* in *Caulobacter crescentus* and *dam* methlyase in *E. coli*, have been characterized, the effect of methyltransferases in *H. pylori* remain understudied (1-4)

My work expanded on a previous study on JHP1050 (M.Hpy99III) that showed effects on global gene expression and various phenotypic traits in the *H. pylori* J99 methylase-deficient mutant (5). My work centered on the homolog of this methylase (HPG27_1066) in *H. pylori* G27, but also examined other strains of *H. pylori*. My main motivation of studying this MTase is due to a growth defect in *H. pylori* G27, which is similar to the growth defect observed in J99 (5). Additionally, I discovered that in a soft agar motility medium, the MTase-deficient strain of *H. pylori* G27 is less motile than the

wild-type parental strain and the motility of the MTase mutant is inhibited further by the inclusion of $FeSO_4$ in the medium (Fig. 2.8). Of the strains that I tested, *H. pylori* G27 and *H. pylori* 26695 were the only ones that had statistically significant differences in growth in the MTase mutants (Table 2.3). This is supported by the Estibariz study in that not all *H. pylori* strains lacking this methyltransferase have a growth defect (5). However, the growth defect I observed for *H. pylori* 26695 MTase mutant did not coincide with what was reported by Estibariz and co-workers (5), which may have been due to genotypic differences between the strains used in my study and the previous study.

Bioinformatics studies showed the GCGC motif that is methylated by homologs of M.Hpy99III is frequently located near the -13 position of the promoter (Fig 2.1). Examination of other DNA methylation motifs in *H. pylori* 26695 indicated the GCGC motif is unique in its strong positional bias within promoter regions (C. Bazal, unpublished results). These findings suggest that methylation of GCGC motifs situated in the -13 region of promoters is important for promoter activity and the motif has been maintained at this position through selective pressure. Several observations support this hypothesis. First, in the vast majority of *H. pylori* strains whose genomes have been sequenced the M.Hpy99III homolog is an orphan DNA MTase and the cognate restriction enzyme has been lost (Cite IMG). This observation suggests strongly that during the course of *H. pylori* evolution, the M.Hpy99III homolog acquired an important physiological function that was distinct from its original role in protecting *H. pylori* genomic DNA from restriction. Second, I identified three GCGC-containing promoters whose activities were diminished in the *H. pylori* G27 M.Hpy99III-deficient mutant. Third, I found that changing the GCGC motif to GTGC in the HPG27_846 promoter

resulted in a dramatic decrease in promoter activity (Fig 2.4). A similar observation was reported for the corresponding promoter in *H. pylori* J99 (5). As indicated above, loss of the M.Hpy99III homolog does not result in an *in vitro* growth defect in all *H. pylori* strains. We postulate though that M.Hpy99III homologs are important for transcription initiation of genes needed for successful colonization and persistence of *H. pylori* in the human gastric mucosa.

Future Directions

There are several interesting areas of research to pursue regarding the role of the M.Hpy99III homologs in *H. pylori* biology. One of the first questions to address is whether the ^{m5}C MTase is required for colonization or persistence of *H. pylori* in the gastric mucosa. Mouse or Mongolian gerbil animal models could be used for such studies.

Another interesting avenue of research would be to characterize additional GCGC-containing promoters to identify contextual features that are important for robust promoter activity. Although the sample size was small in my study, results from the *gfp* reporter gene studies suggested the presence a thymine immediately preceding the GCGC motif was important for promoter activity (Fig 2.4, Fig. 2.5). Methylation of GCGC motifs located at positions within the promoter besides the -13 region might affect promoter activity. Therefore, characterization of GCGC-containing promoters where the GCGC motif is outside the -13 region would be an interesting line of investigation as it might lead to the identification of other promoter elements that are affected directly by GCGC methylation.

An important line of inquiry is how does methylation of the GCGC motif affects promoter activity? As illustrated in Figure 3.1, methylation of cytosine at C5 affects base-readout and potential protein-DNA interactions due to changes in hydrophobic contacts in the major groove (6). When the C5 position of the cytosine is methylated, the resulting ^{5m}C-G pair mimics the T-A base pair at the edge of the major groove by replacing a nonpolar hydrogen with a bulkier and more hydrophobic methyl group (6). Note that the base-readout for the interior of major groove in the ^{5m}C-G pair is not altered and differs from that of the T-A pair. If the methyl group of ^{m5}C and thymine is sufficient to confer binding specificity by RNA polymerase holoenzyme (RNAP-holoenzyme), then the two nucleotides would be expected to substitute for each other in DNA-binding sites. It is not unreasonable to postulate that such a substitution would allow the promoter to retain its function as substitution of a specific A-T base pair with a G-^{m5}C base pair, but not a G-C base pair, allowed LacI to bind the synthetic operator (6). Therefore, it would be interesting to determine if substituting the GCGC motif in GCGC-containing promoters with GTAC retains promoter function. Failure of the synthetic promoter with the GTAC substitution would suggest base-readout for the interior of the major groove or electrostatic interactions within the minor groove are important determinants for promoter function.

Determining which step in transcription initiation is affected by methylation of the GCGC motif in GCGC-containing promoters is another interesting issue. Transcription initiation is a multi-step process, which includes initial binding of RNAP-holoenzyme to the promoter (i.e., closed complex formation), isomerization of the closed promoter complex to an open complex, formation of abortive transcripts, and promoter

clearance (8). Methylation of the GCGC motif could affect any of these steps in transcription initiation. It seems most likely that methylation would affect the affinity of RNAP-holoenzyme for the promoter, a hypothesis that could be examined with ChIP-seq assays with the wild-type parental strain and the HPG27_1066 mutant.

Finally, it would be informative to determine the basis for growth defect in *H. pylori* G27 HPG27_1066 mutant. The reduced growth rates of the M.Hpy99III-deficient mutants most likely result from the aberrant transcription of one or more genes critical for normal growth. Altered expression of these genes could have a cumulative or synergistic effect on growth of the M.Hpy99III-deficient mutants. Candidates for genes whose altered expression might impact growth include ones involved in metal homeostasis. Estibariz and co-workers noted that the *H. pylori* J99 M.Hpy99III-deficient mutant had increased sensitivity to heavy metals (5). My data on the motility of the HPG27_1066 mutant suggested iron regulation may be affected as the size of the swim halo was diminished by the addition of $FeSO_4$ to the motility agar medium (Fig. 2.8). Estibariz *et al*. reported that *feoB*, which encodes a ferrous sulfate transporter, was down-regulated in the *H. pylori* J99 M.Hpy99III-deficient mutant (5, 9). While it is unknown if *feoB* was down-regulated in the HPG27_1066 mutant, and if so, was responsible for the observed growth phenotype, the predicted *feoB* promoter in *H. pylori* G27 does have a GCGC motif in the -13 region.

Another candidate gene whose altered expression in the M.Hpy99III-deficient mutants might impact growth is *icd*, which encodes the tricarboxlic acid cycle enzyme isocitrate dehydrogenase. My studies showed that the *icd-gfp* reporter gene was down-regulated in the HPG27_1066 mutant (Fig. 2.4). Moreover, preliminary proteomic data of

the soluble protein fractions of wild-type *H. pylori* G27 and the HPG27_1066 mutant suggested isocitrate dehydrogenase levels were reduced in the mutant (Table 3.1).

Yet another candidate gene whose altered expression might impact growth is *hrcA*, which encodes a heat shock repressor that negatively regulates expression of operons encoding major chaperones in *H. pylori* (*groESL*, and *hrcA-grpE-dnaK*) as well as genes encoding an outer membrane protein (*omp16*) and a C_4-dicarboxylic acid transport protein (*dcuA*) (10). The predicted promoter for *hrcA* in *H. pylori* G27 has a GCGC motif at the -13 position, but sequence of the corresponding site in the *H. pylori* 26695 *hrcA* promoter is ATGC (Table 3.2). Such heterogeneity in occurrence of GCGC motifs in promoters likely accounts for phenotypic differences in M.Hpy99III-deficient mutants in various *H. pylori* strains. Levels of DnaK appeared to be reduced in the HPG27_1066 mutant relative to wild type (Table 3.1), which is consistent with the hypothesis that methylation of the GCGC motif in the *hrcA-grpE-dnaK* promoter in *H. pylori* G27 is required for optimal promoter activity. Interestingly, disrupting *hrcA* in *H. pylori* G27 resulted in reduced motility and reduced transcript levels of several flagellar genes (10). If expression of *hrcA* was inhibited in the HPG27_1066 mutant, this might account for the motility defect of the mutant.

RNA-seq analysis of the M.Hpy99III-deficient mutants in *H. pylori* strains G27 and B128 would provide valuable data for identifying other candidate genes whose altered expression account for the growth defects of these mutants. Promoters of candidate genes could be validated for their roles in the growth defects of the M.Hpy99III-deficeint mutants by replacing them with promoter alleles from *H. pylori* strains whose growth rates are not impacted by loss of the M.Hpy99III homolog.

My work on HPG27_1066 has revealed several novel findings not previously reported, in particular a motility and catalase phenotype. Continued research on the GCGC promoters in *H. pylori* G27 will answer several questions that have arisen due to the study. Understanding how methylation regulates gene expression in *H. pylori* will allow us to better understand the organism as a whole and how methylation regulates gene expression in a bacterial system.

Table 3.1: Proteins that appear to be present at differing levels in wild type and HPG27_1066 mutant.

soluble proteins				
Accession	**Gene**	**Description**	**#WT peptides found**	**#Mutant peptides found**
Higher number of peptides detected in mutant sample				
ACI27174.1	*htrA*	protease DO	0	7
ACI27994.1	*rpsM*	ribosomal protein S13	0	3
ACI27347.1	*dapD*	tetra hydro dipicolinate N succinyl transferase	0	3
ACI28098.1	*rnj*	ATP/GTP binding protein	3	9
ACI26908.1	*recA*	recombinase	2	6
ACI27549.1	*hup*	DNA-binding protein HU	2	5
ACI28001.1	*rplF*	ribosomal protein L6	2	5
ACI26943.1	*frdA*	fumarate reductase flavo protein subunit	2	5
ACI27991.1	*rpoA*	RNA polymerase alpha subunit	3	7
ACI27829.1	*atpD*	ATP synthase F1 subunit beta	5	11
Lower number of peptides detected in mutant sample				
ACI27765.1	HPG27_1012	hypothetical protein	3	0
ACI27935.1	*alaS*	alanyl tRNA synthetase	3	0
ACI27056.1	*pgdA*	hypothetical protein	4	0
ACI26800.1	*icd*	isocitrate dehydrogenase	6	0
ACI26959.1	*htpG*	chaperone and heat-shock protein 90	10	4
ACI27010.1	*clpB*	ATP-dependent protease binding subunit heat-shock protein	20	8
ACI26870.1	*dnaK*	chaperone and heat-shock protein 70	27	13
Insoluble proteins				
Higher number of peptides detected in mutant sample				
ACI27064.1	*babA*	outer membrane protein	0	5
ACI27328.1	*hefC*	cytoplasmic pump protein of the *hefABC* efflux system	0	3
ACI27605.1	*hopU*	outer membrane protein	3	12
ACI27174.1	*htrA*	protease DO	7	19
ACI27831.1	*atpA*	ATP synthase F1 subunit alpha	3	8

ACI27870.1	*hopQ*	outer membrane protein	5	13
ACI26799.1	*gltA*	citrate synthase	2	5
ACI26978.1	*dsbG*	disulphide isomerase	3	7
ACI27902.1	*tuf*	translation elongation factor EF Tu	5	10
ACI27209.1	*hofC*	outer membrane protein	2	4
Lower number of peptides detected in mutant sample				
ACI26870.1	*dnaK*	chaperone and heatshock protein 70	8	2
ACI27595.1	*vacA*	vacuolating cytotoxin A	6	2

Table 3.2: Comparison of *hrcA* promoter sequence between *H. pylori* strains 26695 and G27

gene	sequence (TSS and 50 nucleotides of upstream sequence)[1]
hrcA (26695)	TTTTTTAAAACAAGTGATAGATTATTAACTTTTTATGCTATAATGCGAGGG
(G27)	TTTTTTAAAACAATTGATAGATTATTAACTTTTT**GCGC**TATAATTCGAGGG

[1]GCGC motifs are in bold-face. Nucleotide sequences in the *H. pylori* 26695 sequences that differ from those in the *H. pylori* G27 sequences are indicated in red.

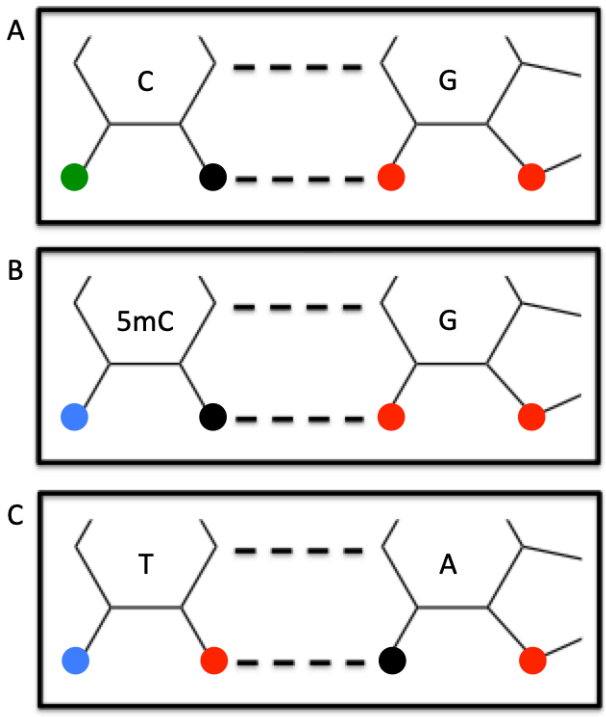

Figure 3.1: Base readout for the major groove of DNA base pairs. (A) denotes a C-G base pair, (B) denotes a 5mC – G base pair and (C) denotes a T-A base pair. A black circle represents a hydrogen bond donor, a red circle represents a hydrogen bond acceptor, a blue circle represents a methyl group, and a green circle represents a nonpolar hydrogen. Only functional groups related to the major groove are indicated (6).

References:

1. **Zweiger G, Marczynski G, Shapiro L.** 1994. A *Caulobacter* DNA methyltransferase that functions only in the predivisional cell. J Mol Biol **235:**472-485.
2. **Stephens C, Reisenauer A, Wright R, Shapiro L.** 1996. A cell cycle-regulated bacterial DNA methyltransferase is essential for viability. P Natl Acad Sci USA **93:**1210-1214.
3. **Peterson KR, Wertman KF, Mount DW, Marinus MG.** 1985. Viability of *Escherichia coli* K-12 DNA adenine methylase (dam) mutants requires increased expression of specific genes in the SOS regulon. Mol Gen Genet **201:**14-19.
4. **Lu AL, Clark S, Modrich P.** 1983. Methyl-directed repair of DNA base-pair mismatches in vitro. P Natl Acad Sci USA **80:**4639-4643.
5. **Estibariz I, Overmann A, Ailloud F, Krebes J, Josenhans C, Suerbaum S.** 2019. The core genome m5C methyltransferase JHP1050 (M.Hpy99III) plays an important role in orchestrating gene expression in *Helicobacter pylori*. Nucleic Acids Res **47:**2336-2348.
6. **Dantas Machado AC, Zhou T, Rao S, Goel P, Rastogi C, Lazarovici A, Bussemaker HJ, Rohs R.** 2015. Evolving insights on how cytosine methylation affects protein-DNA binding. Brief Funct Genomics **14:**61-73.

7. **Fisher EF, Caruthers MH.** 1979. Studies on gene control regions XII. The functional significance of a lac operator constitutive mutation. Nucleic Acids Res **7:**401-416.

8. **Browning DF, Busby SJ.** 2016. Local and global regulation of transcription initiation in bacteria. Nat Rev Microbiol **14:**638-650.

9. **Velayudhan J, Hughes NJ, McColm AA, Bagshaw J, Clayton CL, Andrews SC, Kelly DJ.** 2000. Iron acquisition and virulence in *Helicobacter pylori*: a major role for FeoB, a high-affinity ferrous iron transporter. Mol Microbiol **37:**274-286.

10. **Roncarati D, Danielli A, Spohn G, Delany I, Scarlato V.** 2007. Transcriptional regulation of stress response and motility functions in *Helicobacter pylori* is mediated by HspR and HrcA. J Bacteriol **189:**7234-7243.