STRUCTURAL AND FUNCTIONAL STUDIES OF TRANSCRIPTIONAL REGULATION IN *HELICOBACTER PYLORI*

By

Brendan Nathaniel Borin

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Biological Sciences

December, 2008

Nashville, Tennessee

Approved:

Andrzej M. Krezel, Ph.D.

Gerald Stubbs, Ph.D.

Charles R. Sanders, Ph.D.

Timothy L. Cover, M.D.

Brandt F. Eichman, Ph.D.

ACKNOWLEDGEMENTS

I am very grateful for the help I have received from many people over the last several years. First and foremost, I thank my adviser, Dr. Andrzej Krezel, who was personally involved in training me in all aspects of our research from protein expression to structure determination.

I also thank the other members of my committee, Drs. Gerald Stubbs, Chuck Sanders, Tim Cover, and Brandt Eichman for their valuable advice.

I must thank Dr. Andrei Popescu, a former graduate student in our laboratory who initiated the work on *Helicobacter pylori*. He produced several of the protein samples with which I worked. In addition, his advice during the early years of my training was invaluable.

Much of the functional work on *Helicobacter pylori* has been done in close collaboration with Drs. John Loh and Tim Cover. Their expertise in working with *H. pylori* has been essential to our progress on those projects. John has provided mutant strains for us, as well as the microarrays.

I am thankful for financial support from the Molecular Biophysics Training Grant, which provided funding for two years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2D	two-dimensional
3D	three-dimensional
αCTD	C-terminal domain of the α subunit of RNA polymerase
Å	Angstrom (10^{-10} m)
BB	Brucella broth
BMRB	Biological Magnetic Resonance Data Bank (BioMagResBank)
bp	base pairs
BS^3	Bis(Sulfosuccinimidyl) suberate
cat	chloramphenicol acetyl transferase
cDNA	complementary DNA
CE	capillary electrophoresis
CFU	colony-forming unit
COSY	correlation spectroscopy
DNA	deoxyribonucleic acid
dsDNA	double-stranded DNA
DSS	2,2-dimethyl-2-silapentane-5-sulfonic acid
EDTA	ethylenediaminetetraacetic acid
FBS	fetal bovine serum
FPLC	fast performance liquid chromatography
h	hour
HK	histidine kinase

HSQC	heteronuclear single quantum coherence
Hz	Hertz (s^{-1})
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	kiloDalton
LB	Luria-Bertani broth
MALT	mucosa-associated lymphoid tissue
MES	2-(N-Morpholino)-Ethanesulfonic Acid
µL	microliter
mL	milliliter
µM	micromolar
mM	millimolar
mRNA	messenger ribonucleic acid
NDSB	non-detergent sulfobetaine
NECEEM	nonequilibrium capillary electrophoresis of equilibrium mixtures
Ni-NTA	nickel-nitrilotriacetic acid
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
OD	optical density at 600 nm
ORF	open reading frame
PCR	polymerase chain reaction
PDB	Protein Data Bank
Pol	eukaryotic RNA polymerase

ppm	parts per million
qPCR	quantitative PCR
RDC	residual dipolar coupling
RHH	ribbon-helix-helix
RM	restriction-modification
RMSD	root-mean-square deviation
RNA	ribonucleic acid
RNAP	RNA polymerase
rRNA	ribosomal ribonucleic acid
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
TBE	Tris, Borate, EDTA
TCEP	Tris(2-Carboxyethyl) phosphine
TOCSY	total correlation spectroscopy
Tris	tris(hydroxymethyl)aminomethane
UV	ultraviolet
WT	wild type

CHAPTER I

INTRODUCTION

Discovery of *Helicobacter pylori*

Helicobacter pylori (H. pylori) is a Gram-negative, pathogenic bacterium that infects half the world's population and is responsible for the majority of cases of gastric and duodenal ulcers. It is also the most important risk factor for gastric cancer. Uniquely adapted to survive the low pH conditions of the human stomach, if left untreated it will establish an infection that will persist for the remainder of the infected individual's life. In the most severe cases, long-term infection can lead to gastric cancer. Complications of *H. pylori* infections have incredible human health costs and economic costs both in the United States and worldwide. According to the U.S. Centers for Disease Control website, between 500,000 and 850,000 new cases of peptic ulcers are reported each year. New and recurring ulcers lead to over one million hospitalizations and 6500 deaths annually. In addition, the economic costs of peptic ulcer disease are estimated to be around $6 billion in the United States alone.

Given the prevalence of gastric diseases that are now associated with *H. pylori*, which has accompanied humanity since the beginning of the species, it is surprising that the discovery of *H. pylori* and the understanding of its role in these diseases has been so recent. Before the discovery of *H. pylori*, and even several years after, the medical community considered stress and lifestyle factors such as eating spicy foods to be the cause of ulcers. In 1982, Barry J. Marshall and J. Robin Warren discovered and cultivated a new bacterium from biopsies of patients with gastric inflammation and ulcers (Warren and Marshall 1983). In a brave attempt to

prove the gastric pathogenicity of *H. pylori*, Marshall drank a culture of the bacteria and soon after, suffered from typical symptoms of acute gastritis. Unfortunately, given the firm belief by the medical community that no organism could thrive in the stomach, most did not regard this as sufficient proof that the bacterium could be the cause of ulcers. Subsequent work over the next several years firmly established the causative role of *H. pylori* in ulcers and more severe gastric diseases, including cancer. In 1994, the World Health Organization declared *H. pylori* to be a class I carcinogen (IARC 1994). Marshall and Warren were awarded the Nobel Prize in Medicine in 2005.

Although our awareness of *H. pylori* is rather recent, it appears that we have a very long history with the bacterium. A recent genetic analysis of hundreds of strains from around the world found that genetic differentiation among *H. pylori* strains increases, while genetic diversity decreases, with distance from East Africa. These genetic trends, along with the timescales associated with them, closely parallel those of humans, confirming that *H. pylori* migrated from East Africa along with its human hosts (Linz, Balloux *et al.* 2007). It is unlikely that *H. pylori* is zoonotic because there is no known reservoir outside of humans. In fact, the most closely related species, *Helicobacter acinonychis*, which infects large cats, is thought to have jumped from humans to cats approximately 200,000 years ago (Eppinger, Baar *et al.* 2006). Because of its high degree of genetic variability and the fact that it infects such a large percentage of the population, *H. pylori* genetics can in some cases be an incredibly useful tool for other areas of science. One extremely interesting study showed that *H. pylori* genetic markers could determine human population movements even when human genetic markers were useless (Wirth, Wang *et al.* 2004).

H. pylori is a member of the ε-proteobacteria, which includes *Helicobacter* species as well as closely related *Campylobacter* and *Wolinella* species. Many species establish commensal infections within their natural hosts, but can cause disease in humans. *Campylobacter jejuni* is a commensal organism in chickens and other birds, but is a pathogen in humans, capable of causing severe, bloody diarrhea (Young, Davis *et al.* 2007). *Wolinella succinogenes*, which is the most closely related species to *H. pylori* outside of the *Helicobacter* genus, infects bovines and is considered to be nonpathogenic (Baar, Eppinger *et al.* 2003). Arcobacter species are found in livestock and various water sources. They cause a less serious diarrhea than *Campylobacter* and are considered emerging pathogens (Snelling, Matsuda *et al.* 2006). Besides these terrestrial species, more than 90% of bacteria in deep-sea hydrothermal vents are members of the ε-proteobacteria (Nakagawa, Takaki *et al.* 2007).

In their original Lancet paper (Warren and Marshall 1983) and a follow-up the next year (Marshall and Warren 1984), Marshall and Warren described several features of *H. pylori* that led to its identification as a new species. The bacteria have a curved rod shape with 1-2 spirals (Figure 1) that is quite unique among ε-proteobacteria and may bestow upon it the ability to penetrate the gastric mucosa (Andersen 2007). They have 4-6 flagella, all located at one end of the cell. The flagella are an essential colonization factor, giving the bacterium the ability to penetrate the mucus layer of the stomach and to escape regions of high acidity (Croxen, Sisson *et al.* 2006). They are quite different from other bacterial flagella in many ways. Sheaths protect them from depolymerization in the acidic environment, and they contain only seven protofilaments instead of the usual eleven (Galkin, Yu *et al.* 2008). The differences render them unrecognizable by toll-like receptor 5 (Andersen-Nissen, Smith *et al.* 2005). Marshall and Warren also noticed that *H. pylori* are microaerophilic and slow-growing. In fact, they did not

recover bacteria from their first samples because they threw out the incubation plates after only two days. *H. pylori* have a long doubling time, on the order of four hours, in part because of their constant mobility, which consumes large amounts of energy.

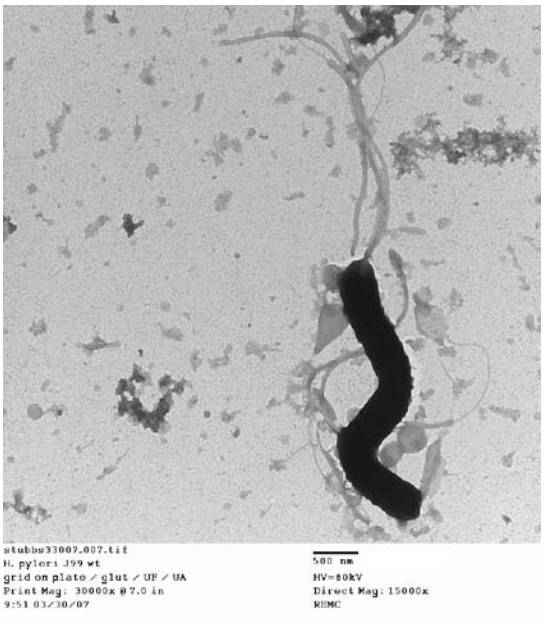

Figure 1. Electron micrograph of a *Helicobacter pylori* J99 cell at 15,000x magnification (taken by Amy Kendall)

The genomes of six different strains of *H. pylori* have been sequenced. The 26695 strain (GenBank AE000511) was the first to be sequenced (Tomb, White *et al.* 1997), and in fact, it was the seventh complete sequenced bacterial genome (Josenhans, Beier *et al.* 2007). Following that, the J99 strain (GenBank AE001439) was sequenced (Alm, Ling *et al.* 1999), allowing a comparison of the two genomes. The entire genome of the HPAG1 strain (GenBank CP000241), associated with higher incidences of chronic atrophic gastritis and gastric adenocarcinoma, has also been sequenced (Oh, Kling-Backhed *et al.* 2006). More recently, sequences for the G27

(GenBank CP001173), Shi470 (GenBank CP001072), and P12 (GenBank CP001217) strains have become available.

Disease

Possible Outcomes

H. pylori is the underlying cause of a wide range of gastric diseases. It is important to note, however, that the majority of infections do not lead to severe disease. Once an infection has been established, *H. pylori* adopts a strategy of persistence rather than acute pathogenicity. Although infection leads to chronic gastritis in all cases (Kandulski, Selgrad *et al.* 2008), this itself does not usually cause any discomfort, and the person will remain unaware of the infection. *H. pylori* has several effects on the epithelial tissue. The bacteria stimulate cytokine production that leads to an inflammatory reaction. The chronic inflammatory state can eventually lead to changes such as atrophy and fibrosis (Ernst and Gold 2000). A small percentage of those infected will at some point in their lives suffer from a gastric or duodenal ulcer, which is an erosion of the mucosal lining of the stomach or the duodenum. The underlying tissue becomes more exposed to acid, leading to a painful experience. The lifetime risk of developing an ulcer has been estimated to be around 10% (Ernst and Gold 2000), but varies between 3 and 25% depending on a number of factors (Kandulski, Selgrad *et al.* 2008).

Gastric cancer is a less common, but more serious, eventual outcome than ulcers. Two types of gastric cancer are associated with *H. pylori* infection – mucosa associated lymphoid tissue (MALT) lymphoma and gastric adenocarcinoma. MALT lymphoma develops as a result of chronic stimulation of T-cells, whose cytokines cause an expansion of B-cells that can invade the epithelial cell compartment (Ernst and Gold 2000; Kandulski, Selgrad *et al.* 2008). During the unchecked proliferation of cells, genetic alterations can occur, including a specific

translocation that leads to oncogenesis (Swisher and Barbati 2007). Gastric adenocarcinoma is a cancer of the epithelial cells of the stomach. Inflammatory responses, including oxygen radicals, from white blood cells recruited to the epithelium in response to *H. pylori* infection can induce greater cell proliferation and eventually mutations that lead to carcinogenesis (Correa 2004).

Risk Factors

Several risk factors have been identified for both the initial infection with *H. pylori* and also the subsequent development of more severe disease. Rates of infection vary widely in different regions of the world. One glance at a map (Figure 2) showing regional rates of infection and it is clear that less developed regions have the highest rates. Poor socioeconomic conditions, which are sometimes accompanied by crowded and unsanitary conditions and poor nutrition are associated with higher rates of infection. Many studies in children have shown that not only are the rates of infection in poor communities higher, but rates of spontaneous clearance are also lower than in wealthier communities (Kivi and Tindberg 2006).

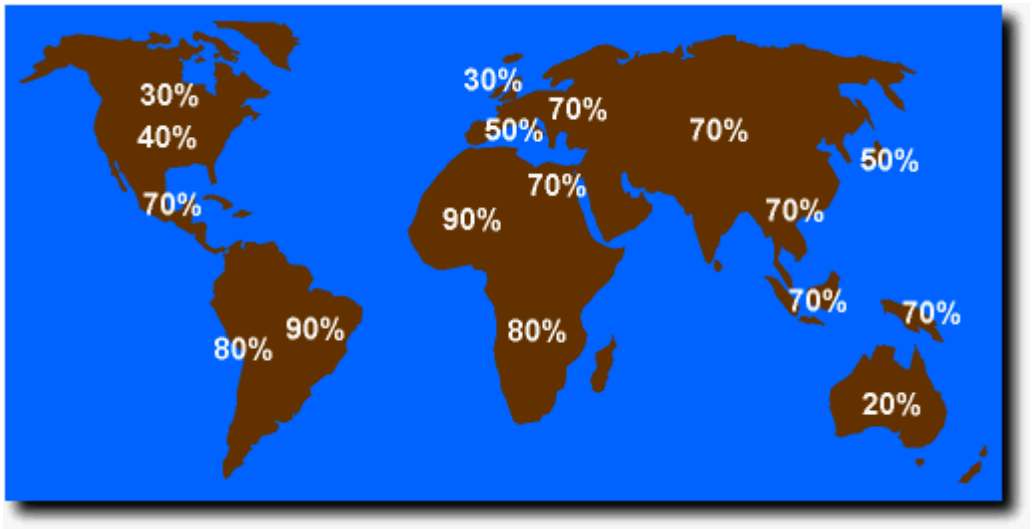

Figure 2. World map showing rates of *Helicobacter pylori* infection in various regions.

Much of a person's risk for becoming infected with *H. pylori* depends on the infection status of family members. The route by which *H. pylori* spreads from person to person is not known, but genetic evidence suggests that it is primarily acquired during childhood from close family members. The most prevalent route of infection is from mother to child. No free-living sources of the bacteria have been found. Fecal-oral and oral-oral routes of infection have been suggested but not firmly established. In healthy people with normal bowel movements, there is usually no culturable *H. pylori* in the feces. Large amounts of culturable *H. pylori* have been found in vomitus, however (Amieva and El-Omar 2008).

For both ulcers and gastric cancer, *H. pylori* infection is the single greatest risk factor. 70-75% of gastric ulcers and 90-95% of duodenal ulcers are caused by *H. pylori* (Ernst and Gold 2000), and it is the most important risk factor for gastric cancer (Kandulski, Selgrad *et al.* 2008). Like many other ε-proteobacteria in their natural hosts, *H. pylori* "prefers" to establish a commensal infection rather than causing severe disease which could risk the life of its host. The eventual outcome of infection, however, depends on environmental, host, and bacterial virulence factors.

Researchers have investigated bacterial genetic factors that can help predict long-term outcomes. *H. pylori* strains can be classified as either type I or type II, depending on expression of the two genes cagA and vacA (Xiang, Censini *et al.* 1995). CagA is one protein expressed from the CAG (cytotoxin associated gene) pathogenicity island, which also encodes for the components of a type 4 secretion system. CagA is secreted into host cells, where it is phosphorylated by host kinases (Higashi, Tsutsumi *et al.* 2002; Saadat, Higashi *et al.* 2007). It can then activate several different signal transduction pathways. Individuals harboring CagA+ strains have greater inflammatory responses and higher levels of IL-8 secretion than individuals

with CagA- strains (Peek, Miller *et al.* 1995). Although cagA+ strains increase the risk of cancer two-fold (Kandulski, Selgrad *et al.* 2008), 70-95% of all strains have cagA, and only a very small percentage of people actually develop cancer (Ernst and Gold 2000). Approximately 1% will eventually develop gastric adenocarcinoma, while development of MALT lymphoma is even rarer. VacA is a secreted toxin that forms hexameric chloride channels in membranes (Iwamoto, Czajkowsky *et al.* 1999). Some of its reported effects include causing osmotic swelling in endosomes (Amieva and El-Omar 2008), altering tight junctions to allow passage of nutrients such as Fe^{3+} and Ni^{2+} (Papini, Satin *et al.* 1998), and inhibition of T cell activation (Gebert, Fischer *et al.* 2003).

Probably more important than bacterial factors for determining the outcome of *H. pylori* infection are host factors. It has been suggested that more severe diseases are the result of poorly regulated immune responses (Ernst and Gold 2000), although the immune response is, in turn, affected by certain bacterial factors. Some individuals have cytokine profiles that indicate a greater inflammatory response to infection. Polymorphisms in the interleukin-1 gene cluster have been shown to be particularly important (El-Omar, Carrington *et al.* 2000).

Treatment

Fortunately, many of the diseases associated with *H. pylori* are preventable or even curable now that we know to treat the underlying cause with antibiotics. Some of the most convincing evidence that advanced Marshall and Warren's hypothesis was the drastic reduction in recurrence rates of ulcers following antibiotic treatment (Ernst and Gold 2000). Treatments can be quite complex, with upwards of a dozen pills taken at specific times each day for several weeks. It is rare for a single agent to be administered. Drug regimens are usually known as dual,

triple, or even quadruple therapies, depending on the number of drugs involved. Early dual therapies consisted of a proton pump inhibitor (PPI) and either amoxicillin or clarithromycin.

Bismuth is an agent that is commonly used against *H. pylori*. It has very interesting antibacterial properties. Bismuth interferes with the cell wall (Stratton, Warner *et al.* 1999) and adhesion to gastric epithelial cells and may inhibit urease and phospholipase (Ottlecz, Romero *et al.* 1999). Its overall effects on the cells appear similar to those of iron deprivation (Bland, Ismail *et al.* 2004). Dual therapies consisting of bismuth and one antibiotic do not have acceptable eradication rates; however, triple therapies consisting of bismuth, tetracycline, and either metronidazole or clarithromycin have eradication rates greater than 95% and are the most effective and least costly treatments (Salcedo and Al-Kawas 1998). The most intense drug regimens are called quadruple therapies, and consist of bismuth, a proton-pump inhibitor, and two antibiotics.

Although the best therapies can approach eradication rates of around 98%, the effective rates of all therapies are reduced by two principal factors – noncompliance and antibiotic resistance. Noncompliance is a result of the length and complexity of the drug regimens as well as adverse effects of some drugs. Treatments often require taking several pills each day at precise times. More complex treatments involving more drugs that should have higher eradication rates may suffer more from noncompliance. Adverse effects include those of clarithromycin, which can cause a bitter, metallic taste sensation. Metronidazole can do the same, as well as causing nausea or a disulfiramlike reaction (Salcedo and Al-Kawas 1998). One common adverse effect due to antibiotic use in general is diarrhea, due to full or partial eradication of naturally occurring, beneficial bacteria in the intestines (McFarland 2008).

Antibiotic Resistance

As with many bacterial pathogens, we are now facing a situation where antibiotic resistance is on the rise. At this point, virtually all infections have some antibiotic resistance (Graham and Shiotani 2008). Although *H. pylori* does not easily develop resistance to some drugs such as amoxicillin and tetracycline, there are now known strains with resistance to each. The mechanisms of resistance have been identified and are really quite simple – a single amino acid substitution in a penicillin-binding protein in the case of amoxicillin (Gerrits, Schuijffel *et al.* 2002) and a triple base-pair substitution in the 16S rRNA gene in the case of tetracycline (Gerrits, de Zoete *et al.* 2002). Clarithromycin, which has the greatest antibacterial activity against *H. pylori*, and metronidazole are often rendered less effective because of resistant strains (Salcedo and Al-Kawas 1998). In order to combat resistance, dual therapies were soon replaced by triple therapies, which consisted of a proton-pump inhibitor or bismuth and 2 antibiotics among amoxicillin, clarithromycin, and metronidazole. Unfortunately, triple therapies containing clarithromycin now only boast a 50-79% eradication rate, far below the 95% rate which is often considered to be the threshold for an effective antibacterial treatment (Graham and Shiotani 2008).

Genetic Variability

H. pylori has an extraordinary ability to adapt to changing circumstances, and it is this ability that enables it to establish permanent infections. The key to its adaptability is its genetic variability. *H. pylori* is one of the most genetically variable species, and each carrier harbors his own strain or strains (Suerbaum and Josenhans 2007). Different isolates can vary quite significantly in their genomic content. One microarray study of several strains showed that 25%

of genes from the sequenced J99 and 26695 strains are missing in at least one of the tested strains (Gressmann, Linz *et al.* 2005). Many of the strain-specific genes encode for restriction enzymes, transposases, and outer membrane proteins (Salama, Guillemin *et al.* 2000). Most of these genes have high GC content and were likely picked up by horizontal gene transfer. Genetic variability is a trait common to many ε-proteobacteria, and many mechanisms contribute to it. *H. pylori* has one of the highest mutation rates among bacteria, and also a high rate of exogenous DNA uptake and homologous recombination. Analysis of the sequenced genomes failed to identify homologs of several DNA repair genes that can be found in *E. coli* (Kang and Blaser 2006). In one study of genetic variability over time within a single host, it was found that both the overall content of genes as well as individual gene sequences could change significantly over the course of several years (Israel, Salama *et al.* 2001). A later study, however, showed that although most genetic changes were due to homologous recombination, very few recombination events resulted in loss or gain of genes (Kraft, Stack *et al.* 2006). In one study of homologous recombination in pathogenic bacteria, *H. pylori* was found to have the highest recombination rate (Perez-Losada, Browne *et al.* 2006). Several instances of recombination between two strains infecting a single patient over the course of long-term infection have been detected (Kersulyte, Chalkauskas *et al.* 1999). In addition, slipped-strand mispairing appears to be a common mechanism of switching on and off several genes (Suerbaum and Josenhans 2007). Many of the changes that occur allow the bacteria to evade the immune system, a task it performs so well that it appears that those with an immune deficiency do not have a higher rate of infection than the general population (Suerbaum and Josenhans 2007). All of these mechanisms allow *H. pylori* to adapt to specific hosts, specific regions of the stomach, and to changing conditions over the lifetime of their host.

Transcriptional Regulation in *H. pylori*

Transcription is the process by which cells read genomic information encoded by specific sequences of DNA and produce RNA with the same sequences, except for replacing thymine with uracil. The process is essential, as the RNAs that are produced play several vital roles within the cells. Messenger RNA (mRNA) contains the necessary information for ribosomes to produce new proteins. Ribosomal RNAs are necessary components of the ribosomes. Transfer RNAs are coupled to specific amino acids that they deliver to the ribosome for incorporation into the growing polypeptide chain. Transcription is carried out by DNA-dependent RNA polymerase, a multisubunit enzyme with core composition $\alpha_2\beta\beta'\omega$ in bacteria (Burgess 1969). The situation is more complicated in eukaryotes, where there are three different RNA polymerases, each of which has many more subunits than the bacterial RNA polymerase (Cramer, Armache *et al.* 2008).

The protein content of a cell is not constant through time. Cells must continually be able to change the levels of certain proteins in response to different situations, such as different growth phases or stages of the cell cycle or in response to environmental signals. Therefore, regulation of the transcriptional process is essential. There are several mechanisms of regulating transcription, including the use of sequence-specific transcription factors that bind DNA and recruit RNA polymerase, repressors that bind DNA and prevent RNA polymerase from binding the promoter sequence, and sequence-specific sigma factors that are considered subunits of RNA polymerase holoenzyme. Slipped-strand mispairing can completely turn a gene on or off.

With approximately 1,500 genes, *H. pylori* has a relatively small genome and very few transcriptional regulators described to date (Tomb, White *et al.* 1997; Alm, Ling *et al.* 1999). Initial identification of several transcriptional regulators by homology resulted from sequencing

efforts. The small number of regulators has been explained by the fact that *H. pylori* has only one environmental niche and must respond to only a small number of stimuli compared to free living bacteria. It has been reported that the number of transcriptional regulators in bacteria increases proportionally to the square of the number of genes in the genome (van Nimwegen 2003). Having a genome that is only approximately one-third the size of that of *E. coli*, one would expect to find many fewer transcriptional regulators. Nevertheless, our understanding of the regulation of metabolic processes and environmental responses in *H. pylori* is far from complete, and there are surely more transcriptional regulators to be discovered.

The first level of transcriptional regulation is found in the RNA polymerase itself. Although the core subunits, $\alpha_2\beta\beta'\omega$, form a catalytically-competent enzyme, called the apoenzyme, promoter-specific transcription requires a σ subunit, which binds tightly to the core, forming the holoenzyme (Murakami and Darst 2003). *H. pylori* has only three sigma factors, σ^{80}, σ^{54}, and σ^{28}. σ^{80} is similar to σ^{70} from *E. coli* in that it is responsible for transcription of most housekeeping genes. There are many other ways in which activity of RNA polymerase can be modified at individual promoters. The major transcriptional responses of *H. pylori* and many of the regulatory proteins involved in them will be summarized below.

Two-Component Systems

Two-component signaling systems are widely found in bacteria as well as eukaryotes outside of the animal kingdom. Composed of a histidine kinase (HK) and a response regulator, they are important for a wide range of responses, including chemotaxis, quorum sensing, and in eukaryotes, hormone-dependent development. HK's are transmembrane proteins with an extracellular sensor domain and signal transducing intracellular domains. They usually function

as dimers and in response to the proper signal, autophosphorylate a conserved histidine found in the dimerization domain. The catalytic domain is responsible for phosphorylating a conserved aspartate in the response regulator. Structures have been determined of the intracellular dimerization and catalytic domains for several HK's, and they appear to be structurally well-conserved. The extracellular sensing domains, however, respond to a wide variety of signals, and there is no common structural motif among them (Wolanin, Thomason *et al.* 2002).

H. pylori has relatively few HK's and response regulators (4 and 7, respectively) (Marais, Mendz *et al.* 1999) compared to other bacteria such as *E. coli*, which has more than 30 two-component systems. The number of HK's in an organism is thought to reflect the complexity of its lifecycle and the number of environmental signals it must respond to. The low number of HK's in *H. pylori* is indicative of the fact that the bacterium only inhabits the human stomach and that it has very little competition from other microorganisms (Beier and Frank 2000).

Acid Response

Regular exposure to low pH in the human stomach is the most significant environmental challenge faced by the gastric pathogen *Helicobacter pylori*. The acid response is the most important and complex response of *H. pylori*. Sachs has described its unique acid response as "acid acclimation" because although several other bacteria such as *Salmonella typhimurium, Vibrio cholera, and Escherichia coli* can mount a temporary response allowing them to survive the acid of the stomach while they pass through, they are incapable of colonizing their host and continuing to grow in such low pH conditions (Sachs, Weeks *et al.* 2005). Much research has been done on the acid response, but because of its complexity, our understanding of it is still limited.

Several array-based gene transcription studies have shown that between 100-200 genes are differentially regulated in response to acid (Ang, Lee *et al.* 2001; Merrell, Goodrich *et al.* 2003; Wen, Marcus *et al.* 2003). Within the general acid response, there are several specific responses including raising the local pH, dealing with the increased divalent metal ion concentration, maintenance of cell membranes and the cell wall, and increased motility. Although there is not complete agreement between the different studies, several key components of the acid response have been identified. Of critical importance is the enzyme urease, a nickel-dependent enzyme that catalyzes a reaction that converts urea to ammonia and carbon dioxide. Ammonia is quickly protonated, removing an excess proton from the cytoplasm. Carbon dioxide is an important product as well. *H. pylori* has both cytoplasmic and periplasmic versions of carbonic anhydrase, which uses carbon dioxide to produce carbonate ions that buffer the cytoplasm and periplasm against sudden changes of pH. The periplasmic carbonic anhydrase may be one of the key factors that allows *H. pylori* to survive low pH longer than most other bacteria (Marcus, Moshfegh *et al.* 2005).

One of the important regulatory systems involved in the acid response is the two-component system ArsRS (HP0165/HP0166). ArsS (HP0165) is a histidine kinase known to act as an acid sensor (Pflock, Dietz *et al.* 2004), and several studies have identified genes that are targeted by its cognate response regulator ArsR (HP0166) (Dietz, Gerlach *et al.* 2002; Forsyth, Cao *et al.* 2002; Pflock, Kennard *et al.* 2005; Pflock, Finsterer *et al.* 2006; Wen, Feng *et al.* 2006; Wen, Feng *et al.* 2007). Metal ion-dependent regulators are also involved in the acid response. Some of the important regulatory proteins involved in the metal ion response include the two-component system HP1364/HP1365, the nickel responsive regulator NikR (Contreras, Thiberge *et al.* 2003), and the ferric uptake regulator Fur. Fur was the first regulatory protein

shown to be involved in the acid response (Bijlsma, Waidner *et al.* 2002), and array studies have implicated both Fur and NikR in the acid response (Bury-Mone, Thiberge *et al.* 2004). Recently, it was also shown that the CrdRS (HP1364/HP1365) two-component system is required for an effective acid response (Loh and Cover 2006), although there appear to be strain-specific differences (Pflock, Muller *et al.* 2007).

Metal-Ion Responses

Closely associated with the acid response are the metal ion responses. Low pH conditions increase the solubility of divalent metal ions and raise their effective concentrations available to cells (Krishnaswamy and Wilson 2000). Many divalent metal ions are essential in trace amounts, but at higher concentrations, they can have toxic effects. Intracellular concentrations of many metal ions are precisely controlled. Two repressors, NikR and Fur, are involved in the homeostasis of nickel and iron, respectively. Both have been shown to be important regulators of the acid response as well (Bury-Mone, Thiberge *et al.* 2004). NikR was discovered in *E. coli* (De Pina, Desjardin *et al.* 1999), where it binds the promoter of the *nikABCDE* operon (Chivers and Sauer 1999), repressing its transcription. The *H. pylori* homolog of NikR was found to have a more significant role in transcriptional regulation, affecting not only the nickel uptake mechanism, but also urease, iron uptake genes, and motility genes (Contreras, Thiberge *et al.* 2003). It was later shown to be both an activator and a repressor, depending on the position of its binding site within different promoters (Ernst, Kuipers *et al.* 2005). The ferric uptake regulator Fur can also induce or repress different sets of genes in an iron-dependent way, inducing iron storage and energy genes while repressing metal metabolism, motility, and cell wall synthesis genes (Ernst, Bereswill *et al.* 2005).

Copper is another metal that is an essential cofactor in many proteins, but toxic at high concentrations. *H. pylori* has a system that specifically regulates intracellular copper concentration. CrdRS is a two-component system that responds to high copper concentrations and upregulates transcription of the copper resistance determinant protein CrdA (Waidner, Melchers *et al.* 2005). CrdA, along with CrdB, CzcA, and CzcB form a copper efflux system (Waidner, Melchers *et al.* 2002).

Motility

Motility is an essential colonization factor and a hierarchically regulated process that requires at least 40 genes involved in flagellar biosynthesis and chemotaxis. Building the flagella correctly and in the right number, as well as responding correctly to environmental stimuli requires a complex set of regulatory mechanisms. Aside from the housekeeping σ^{80} transcription factor, which promotes transcription of one set of motility genes, the other two σ factors, σ^{54} and σ^{28}, are responsible for transcription of different sets of motility genes. σ^{54} requires the two-component system FlgRS for its function. A number of other proteins are involved in motility regulation (Niehus, Gressmann *et al.* 2004).

Adhesion Response

H. pylori must always be ready to move to escape acid or avoid being washed away with the gastric mucus that is constantly being overturned. At any given time, however, about 20% of cells are adhered to the gastric epithelial cell surface layer (Amieva and El-Omar 2008). From this position, they are protected from the acidic conditions in the lumen of the stomach by the

gastric mucus. Many proteins have been identified as adhesins that contribute to these interactions.

Adhesion is accompanied by transcriptional changes in several genes. One transcriptional array-based study found many upregulated and downregulated genes (Kim, Marcus *et al.* 2004). Upregulated genes include the paralyzed flagella protein (HP1274), outer membrane proteins, and HP0222. The most noticeable class of downregulated genes are those involved in motility, including the major flagellin flaA. It makes sense for *H. pylori* to conserve energy by downregulating motility upon adhesion to the gastric epithelial cell layer because they are protected from acid by the mucus layer. We do not yet know what transduces the signal upon adhesion to cause expression changes, nor do we know what transcriptional regulators are involved. Among the upregulated and downregulated genes, the only known transcriptional regulator is HP0222, which was discovered in our laboratory.

Stringent Response

The stringent response in bacteria is a survival strategy used in times of low nutrient conditions that can be triggered by uncharged tRNAs (Jain, Kumar *et al.* 2006) or low levels of phosphorous, iron, or a carbon source (Srivatsan and Wang 2008). One study reported a lack of a stringent response in *H. pylori* (Scoarughi, Cimmino *et al.* 1999) even though it had been found in the closely related *Campylobacter jejuni* (Gaynor, Wells *et al.* 2005). Subsequent studies, however, found an active stringent response. Like other bacteria, *H. pylori* produces ppGpp in response to low nutrient conditions and exhibits a typical stringent response, including much lower transcription of ribosomal RNA (Wells and Gaynor 2006). *H. pylori* mutants lacking

SpoT, a (p)ppGpp synthetase and hydrolase, are unable to survive aerobic shock or acid and are less capable of surviving during stationary phase growth (Mouery, Rader *et al.* 2006).

Growth Phase

As *H. pylori* cultures begin their transition from log phase growth into stationary phase growth, a large number of genes, especially virulence genes, are either induced or repressed (Thompson, Merrell *et al.* 2003). Iron uptake proteins are repressed, while iron storage proteins are induced. Many of the acid response genes such as urease and carbonic anhydrase are repressed. Ribosomal genes are also repressed. Several papers have addressed the role of growth phase in the regulation of motility (Niehus, Ye *et al.* 2002; Loh, Forsyth *et al.* 2004; Rader, Campagna *et al.* 2007).

Nuclear Magnetic Resonance

NMR as a Tool for Structural Biology

The primary experimental method used in our laboratory is nuclear magnetic resonance (NMR). NMR is an incredibly useful technique in many fields of science. In medicine, MRI is used as a diagnostic imaging tool. NMR is used by chemists to determine structures of small molecules and monitor the progress of reactions. It is also used in materials science and oil exploration (Kleinberg 2001). We use NMR for its ability to provide information about the atomic-level structures of biomacromolecules. Along with X-ray crystallography, it is one of the two methods capable of determining high-resolution structures of biomacromolecules. Our laboratory focuses on determining structures of proteins. High-resolution structures of proteins are useful for figuring out how they perform their function and in some cases, for helping determine their functions in the first place. Structures can help elicit the details of how proteins

interact with other proteins, DNA, and small molecule ligands and can even aid in the design of new drugs that will bind with high affinity to a specific protein target.

As a method for doing structural biology, NMR is much newer than X-ray crystallography. This is reflected in the relative numbers of NMR and X-ray structures in the Protein Data Bank (PDB) (Bernstein, Koetzle *et al.* 1977; Berman, Westbrook *et al.* 2000). When a structure is determined, its atomic coordinates are usually deposited in the PDB, which serves as a central repository for protein structures. As of October, 2008, there are 53,917 structures in the PDB. NMR methods were used to determine 7,546 of them, or only about 14%. While the first X-ray protein structures were determined in the 1960s, the first NMR protein structure was only determined in 1983 in the lab of Kurt Wüthrich (Williamson, Havel *et al.* 1985).

Although NMR and X-ray crystallography methods are both capable of producing high-resolution structures, the two methods each have their own advantages and disadvantages. For X-ray crystallography, the main difficulty lies in producing crystals of the protein. Crystal formation may only occur in a narrow range of conditions, and sometimes thousands of conditions must be tested. One advantage of X-ray crystallography, however, is that once good, diffracting crystals can be produced and the phase problem can be solved, the time and effort demanded of the researcher for data collection and structure computation is less than that for determining an NMR structure. The other advantage of X-ray crystallography is that it is not limited by molecular size. Although new NMR methods are continually being developed that push the size limit higher and higher, structure determination in general becomes exponentially more difficult with increasing molecular weight. NMR methods have two main advantages over X-ray crystallography. One is that crystal formation is not required. Once a soluble protein

sample has been obtained, one can begin collecting useful data. The other advantage is that the protein is in solution, which is closer to its functional, physiological conditions than in a packed crystal. It has become clear that dynamic motions of proteins are often critical for their functions. NMR methods have been developed to gain insight into these motions, which cannot be easily discerned by X-ray crystallography.

Discovery of Nuclear Magnetic Resonance

As an experimental tool, NMR is based on the fact that certain atomic nuclei have spin, an intrinsic angular momentum that gives rise to a magnetic moment. One of the experiments that directly led to the development of the concept and theory of spin was the Stern-Gerlach experiment (Gerlach and Stern 1922). In this experiment, silver ions were sent through an inhomogeneous magnetic field, and their deflections were recorded. Contrary to classical theory, which predicted that the deflections would vary continuously with the direction of their spin angular momentum vectors, all of the ions were found to have deflected along one of two paths. The Stern-Gerlach experiment provided some of the most direct evidence in favor of quantum theory. The original experiment done with silver ions measured the deflection due to electron spin angular momentum, which is much greater than that of a proton, however. Stern went on to measure the magnetic moment of the proton, for which he won the Nobel prize in 1943. Isidor Rabi's molecular beam resonance experiments, which used oscillating magnetic fields to induce transitions between nuclear magnetic spin states, allowing measurement of magnetic moments (Rabi 1939), earned him a Nobel prize in 1944. Nuclear magnetic resonance was finally detected in bulk matter in 1946 by two independent groups led by Felix Bloch and Edward Purcell. Bloch's group detected an NMR signal from water (Bloch, Hansen *et al.* 1946) while

Purcell's group detected a signal from paraffin (Purcell, Torrey *et al.* 1946). They shared the Nobel prize for their work in 1952.

Nuclear Magnetic Resonance Spectroscopy

The NMR phenomenon arises from the spin of nuclear particles, a property that is purely quantum mechanical. A deep understanding of NMR requires a good knowledge of quantum mechanics; however, the basic magnetic resonance phenomenon can be explained in classical terms because the two regimes predict the same results (Hanson 2008). Because nuclear spins have intrinsic angular momentum, as well as a magnetic moment, their magnetic moments begin to precess when placed in a magnetic field. The precessional frequency is called the Larmor frequency and is governed by the following equation:

$$\omega = \gamma B$$

The Larmor frequency ω depends on the gyromagnetic ratio γ, which is a constant for a given nucleus, and the strength of the magnetic field, B, in which it is placed. For field strengths commonly used in biomolecular NMR, Larmor frequencies of protons are typically on the order of hundreds of MHz, in the radiofrequency range. All nuclei of a given species should precess with the exact same frequency when placed in a homogeneous magnetic field, but in practice, a number of other phenomena slightly perturb the effective field felt by each nucleus. Nuclei in different parts of a molecule will have slightly different Larmor frequencies, and this makes NMR a powerful tool for structure determination.

Chemical shift is one of the most fundamental concepts in NMR. Although all nuclei of a given isotope have the same gyromagnetic ratio, they will not all have the exact same Larmor frequency due to slight variability of the magnetic field strength at different positions in a

molecule. These variations are primarily due to differences in the local electronic structure around each nucleus. Differences in electron density due to electronegativity as well as currents of valence electrons induced by the static field give rise to small secondary magnetic fields that modify the net field felt by the nuclear spins. The secondary field at a given position is anisotropic (orientation dependent) with respect to the orientation of the molecule in the static field; however, in solution NMR, rapid tumbling of the molecule gives rise to an averaging process that causes each nucleus at a given position to have only a single frequency. The secondary fields change the Larmor frequencies of the nuclei by just a few parts per million (ppm) compared to a standard reference frequency, and this change expressed in ppm is known as the chemical shift of the nucleus. Remarkably, these small effects can easily be detected, and they inherently contain a great deal of useful structural information. Patterns of backbone atom nuclei chemical shifts can be used to obtain accurate secondary structure predictions (Wishart and Sykes 1994).

J-coupling (scalar coupling) is an interaction between two spins that alters the Larmor frequency of one spin depending on the spin state of its coupling partner. The J-coupling interaction is mediated through chemical bonds, and in particular, through s-orbital electrons due to the dominant Fermi contact term. In biological applications, most relevant nuclei have spin ½, leading to two stationary spin states which, at room temperature, are roughly evenly populated. If these nuclei are J-coupled to another nucleus, they will split the signal of the coupled nucleus into two signals of equal intensity. The frequency difference of the two signals in Hz indicates the strength of the coupling interaction. Like chemical shifts, J-couplings contain a lot of structural information; however, depending on the situation, the splittings of the resonance signals can be either very useful or very inconvenient. For small molecule applications, the

splitting, which distribute the observed signal of one nucleus into multiple proportionally weaker signals, are useful because of the structural information they provide. The reduction in intensity of the signals is not usually a problem because small molecules have sharp, intense lines to begin with. For biomacromolecules, decoupling methods are often used to suppress the splittings. Large molecules already have wider, less intense lines than small molecules. J-coupling further reduces the intensities of the lines. The greatest problem, though, is that spectra of large molecules already suffer from the overlap problem because of the large number of signals. Splittings due to J-coupling only make this problem worse.

Dipolar coupling is another interaction between spins, but it is mediated directly through space instead of through chemical bonds. Like J-coupling, the dipolar interaction affects the Larmor frequency of one spin according to the spin state of its coupling partner; however, the dipolar interaction is much stronger than scalar coupling, with splittings on the order of several kHz. In isotropic solution, splitting due to the dipolar coupling interaction is not observed because the rapid tumbling of the molecules averages the interaction to zero. It is possible, however, to alter the sample conditions to induce a slight orientation of the molecules. The slight orientation means that the motional averaging will not reduce the dipolar coupling to zero. There will be a small, observable splitting called the residual dipolar coupling (RDC), and these can be used to refine structures.

Although dipolar couplings are not observed directly in solution NMR, the dipolar interactions are responsible for a large part of the relaxation of signals. Tumbling of the molecules in solution creates random fluctuations in the field that one spin feels due to other, nearby spins. Some of the frequencies of these fluctuations are capable of inducing spin-flips, which cause relaxation.

One particular type of relaxation, called cross-relaxation, occurs when mutual spin-flips cause magnetization to be transferred from one spin to another. This phenomenon gives rise to the nuclear Overhauser effect (NOE). The original Overhauser effect described a similar phenomenon, in which polarization is transferred from electrons to nuclear spins (Overhauser 1953). The efficiency of polarization transfer from one nucleus to another by NOE is inversely proportional to the 6^{th} power of the distance between them. Although internal protein motions and "spin diffusion" (the multistep propagation of polarization by cross-relaxation) preclude calculation of exact distances between nuclei, the measured NOEs can be converted to approximate distance restraints. NOEs are extremely important for NMR structure determination because they give rise to the most important and commonly used restraints for structure calculations.

Modern NMR experiments rely on the fact that the bulk magnetization that arises from many spins in a sample can be manipulated using precise sequences of radio frequency pulses and time delays. During the course of structure determination, a number of different pulse sequences are employed to obtain different types of information about the molecule of interest. When a sample is first placed in the magnet, relaxation processes cause the buildup of a small net magnetization that is parallel to the magnetic field of the instrument. Short radiofrequency pulses applied through a coil that surrounds the sample "rotate" the net magnetization vector about an axis perpendicular to the static field. Combinations of these pulse-induced rotations of the magnetization, along with relaxation and magnetization transfer due to J-coupling and dipolar coupling that occur during time delays define different NMR experiments that contain a wealth of information about relationships of observed nuclei within molecules.

After the pulse sequence is finished, the same coil that is used to transmit radiofrequency pulses to the sample is then used to detect the signal coming from the sample. This signal is composed of individual signals from all nuclei in the sample, each with a slightly different frequency of oscillation. The time-domain data points that are detected and stored are subjected to a Fourier transform to convert them into the corresponding frequency spectrum.

Structure Determination

The general strategy for structure determination by NMR was laid out in Wüthrich's seminal work, NMR of Proteins and Nucleic Acids (Wüthrich 1986). The process can be divided into three phases – spin system identification and sequential assignments, collection of restraints, and structure calculations. At the time, the sequential assignment process utilized 2D ^{1}H experiments, such as COSY, NOESY, and TOCSY. COSY and TOCSY experiments were used to identify spin systems (all protons that can be linked by J-coupling). NOESY experiments were used to determine which spin systems occurred next to each other in the sequence. Known patterns of spins for each amino acid allowed the identification of unique two, three, and four residue sequence elements that could be assigned unequivocally and serve and anchor points for additional assignments. Once spin systems are assigned to specific residues within the protein and all of the sidechain protons are assigned to specific atoms, NOESY crosspeaks can be assigned and converted into distance restraints between two specified atoms. Using the experimentally determined distance restraints as well as the covalent structure of the polypeptide, distance geometry methods were employed to determine 3D structures (Havel, Kuntz *et al.* 1983).

Although the main ideas outlined by Wüthrich are still valid today, there has been incredible progress concerning each step in the process of structure determination, and there are now many more tools available to the researcher. The wide variety of expression vectors allow overexpression of a protein of interest in bacterial and eukaryotic hosts in isotopically enriched forms. Field strengths of spectrometers in use today are much higher, increasing both the sensitivity and resolution of the experiments. Pulsed field gradients allow for better water suppression and are used in place of time-consuming phase cycles for coherence selection. Isotope labeling strategies allow more complicated pulse sequences and pushed higher the maximum size of biomacromolecules that can be studied by NMR. The sequential assignment process is now easier due to ^{15}N and ^{13}C labeling and the development of 3D triple-resonance experiments (Sattler, Schleucher *et al.* 1999). ^{2}H labeling drastically reduces relaxtion due to dipolar interactions, and in combination with TROSY experiments (Pervushin, Riek *et al.* 1997), has allowed NMR studies of much larger molecules. New types of measurements, including residual dipolar couplings (RDCs) (Lipsitz and Tjandra 2004) and paramagnetic relaxation enhancements (PREs), give rise to additional restraints that are complementary to NOE-derived distance restraints and can be used to refine structures. Simulated annealing algorithms using torsion angle dynamics provide efficient methods to calculate structures (Guntert, Mumenthaler *et al.* 1997). Because structure determination is a very iterative process involving dozens to hundreds of rounds of NOE assignments and structure calculations, the incredible increase in computational power has significantly reduced the time needed at each step and iteration of this process.

Outline

Our work on transcriptional regulation in *Helicobacter pylori* started several years ago as a small scale structural genomics project. Several targets of unknown structure and function were chosen based on possible biological interest and certain qualities that were predicted to make them suitable for NMR structural work (Popescu 2004). This work resulted in the structure of one particular target, HP0222, that turned out to be a very interesting DNA-binding protein (Popescu, Karpay *et al.* 2005).

Since then, work in our laboratory has expanded to several other aspects of transcriptional regulation in *H. pylori*. Figure 3 shows a few of the proteins involved in transcriptional regulation that are targets of study in our laboratory. The major part of my effort over the last several years has been determining the function of HP0222, but we have also made progress on other projects. This thesis describes both structural and functional results we have obtained on several of the projects. Chapters II and III cover the structure determination of the RNA polymerase α-subunit C-terminal domain and HP0564, a protein homologous to HP0222. Chapters IV and V cover our progress toward determining the functions of HP0222 and HP0564. Appendix A covers work done on other projects dealing with *H. pylori* proteins.

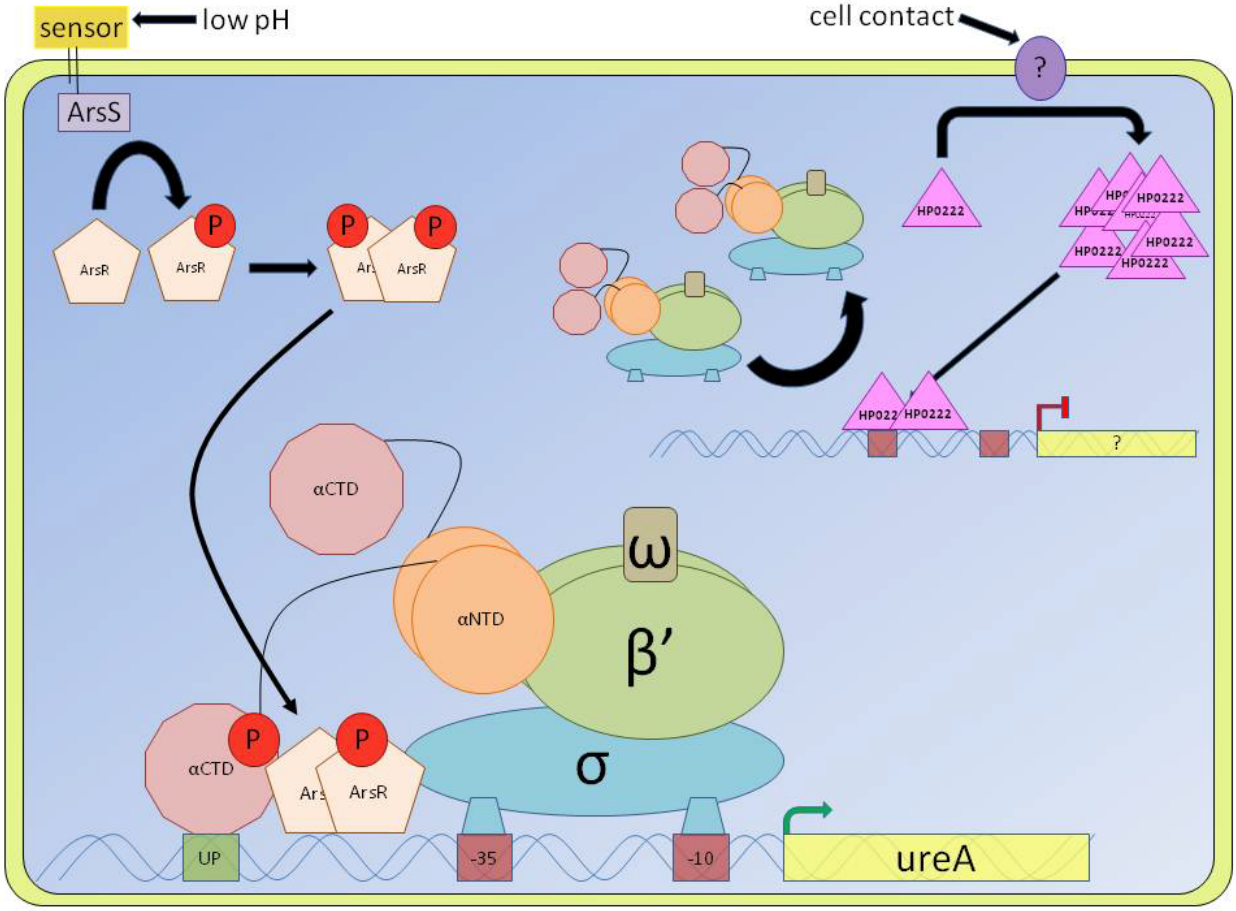

Figure 3. Diagram showing some proteins involved in transcriptional regulation and important environmental responses of *Helicobacter pylori*. Targets of structural and functional study in our laboratory include the RNA polymerase α subunit C-terminal domain, the response regulator ArsR, the periplasmic sensor domain of the histidine kinase ArsS, and the ribbon-helix-helix transcriptional repressor HP0222.

CHAPTER II

STRUCTURE DETERMINATION OF THE *H. PYLORI* RNA POLYMERASE α SUBUNIT C-TERMINAL DOMAIN

A large part of the material presented in this chapter has been submitted for publication.

Abstract

Bacterial RNA polymerase is a large, multi-subunit enzyme responsible for transcription of genomic information. The C-terminal domain of the α subunit of RNA polymerase (αCTD) functions as a DNA and protein recognition element localizing the polymerase on promoter sequences. Despite the high degree of conservation of the subunits among bacteria, *Helicobacter pylori* RNA polymerase has several distinctive features. We have determined the tertiary structure of *H. pylori* αCTD. It is larger and includes a highly amphipathic helix near the C-terminal end that is not present in other structurally determined αCTDs. Residues within this helix are highly conserved among ε-proteobacteria. NMR experiments show that the *H. pylori* αCTD can bind DNA similarly to other αCTDs. By contrast, the sequence and structural differences modeled into the context of transcriptional complexes suggest novel interactions with transcription factors.

Introduction

RNA polymerase (RNAP) is the essential protein complex that transcribes genomic information contained in the template strand of DNA, producing RNA that can be translated into a protein sequence or function on its own. Catalytically important regions of the enzyme are

conserved on the sequence, structural, and functional levels across all kingdoms of life. In bacteria, the catalytically competent core of the enzyme has subunit composition $\alpha_2\beta\beta'\omega$ and a total molecular weight of about 400 kDa (Burgess 1969). The β and β' subunits are the largest and contain the catalytic site. Each is bound to one of the α subunits, which stabilize the complex by dimerizing through their N-terminal domains. The ω subunit has both structural and functional roles in the complex (Mathew and Chatterji 2006). Promoter-specific transcription is driven by σ subunits (factors), which bind the core complex and recognize specific promoter elements.

Although the catalytically competent core subunits of RNAP from bacteria have homologs in archaea and eukaryotes, there are major differences in the RNAPs from the different kingdoms. Whereas in bacteria, there is a single RNAP that is responsible for transcribing all RNAs, in eukaryotes, there are three different RNAPs - Pol I, Pol II, and Pol III. Each has a distinct function. Pol I transcribes ribosomal RNA (rRNA). Pol II transcribes messenger RNA (mRNA). Pol III transcribes transfer RNAs (tRNA) and other small RNAs (Cramer 2002). In addition, the eukaryotic RNA polymerases, as well as the archaeal RNA polymerase, have much more complex subunit compositions, with 14, 12, and 17 subunits for eukaryotic Pol I, Pol II, and Pol III, respectively (Cramer, Armache *et al.* 2008). Among them, there is a 10 subunit, structurally conserved core. The homodimer of α subunits (α_2) that recruits the β and β' subunits and stabilizes the complex in bacteria, is replaced by a heterodimer in eukaryotes. The subunits comprising this heterodimer do not have C-terminal domains like the α subunits from bacteria.

Structure and Assembly of Bacterial RNA Polymerase

The order in which subunits are added to the RNAP complex has been determined (Ishihama 1981). First, an α_2 homodimer is formed through the N-terminal domains of α. Then,

the β subunit is recruited. By convention, it is bound to the α_I subunit. Finally, the β' subunit, in association with the ω subunit, is incorporated, binding the α_{II} subunit, forming the catalytically competent apoenzyme core. The ω subunit binds both N-terminal and C-terminal regions of β' simultaneously, assisting its incorporation into the $\alpha_2\beta$ subcomplex (Ghosh, Ishihama *et al.* 2001). The core can then associate with one of a number of σ factors to form the holoenzyme (Figure 4A).

Over the last decade, a lot of structural work has been done on both individual subunits and the full RNAP complex from bacteria, archaea, and eukaryotes. Crystal structures of the bacterial apoenzyme from *Thermus aquaticus* (Zhang, Campbell *et al.* 1999) and the holoenzyme from *Thermus thermophilus* (Vassylyev, Sekine *et al.* 2002) have been determined. Structures of RNA polymerase II from yeast (Cramer, Bushnell *et al.* 2001; Armache, Mitterweger *et al.* 2005) (Figure 4B) and the archaeal RNA polymerase from *Sulfolobus solfataricus* (Hirata, Klein *et al.* 2008) (Figure 4C) have allowed detailed structural comparisons (Cramer 2002).

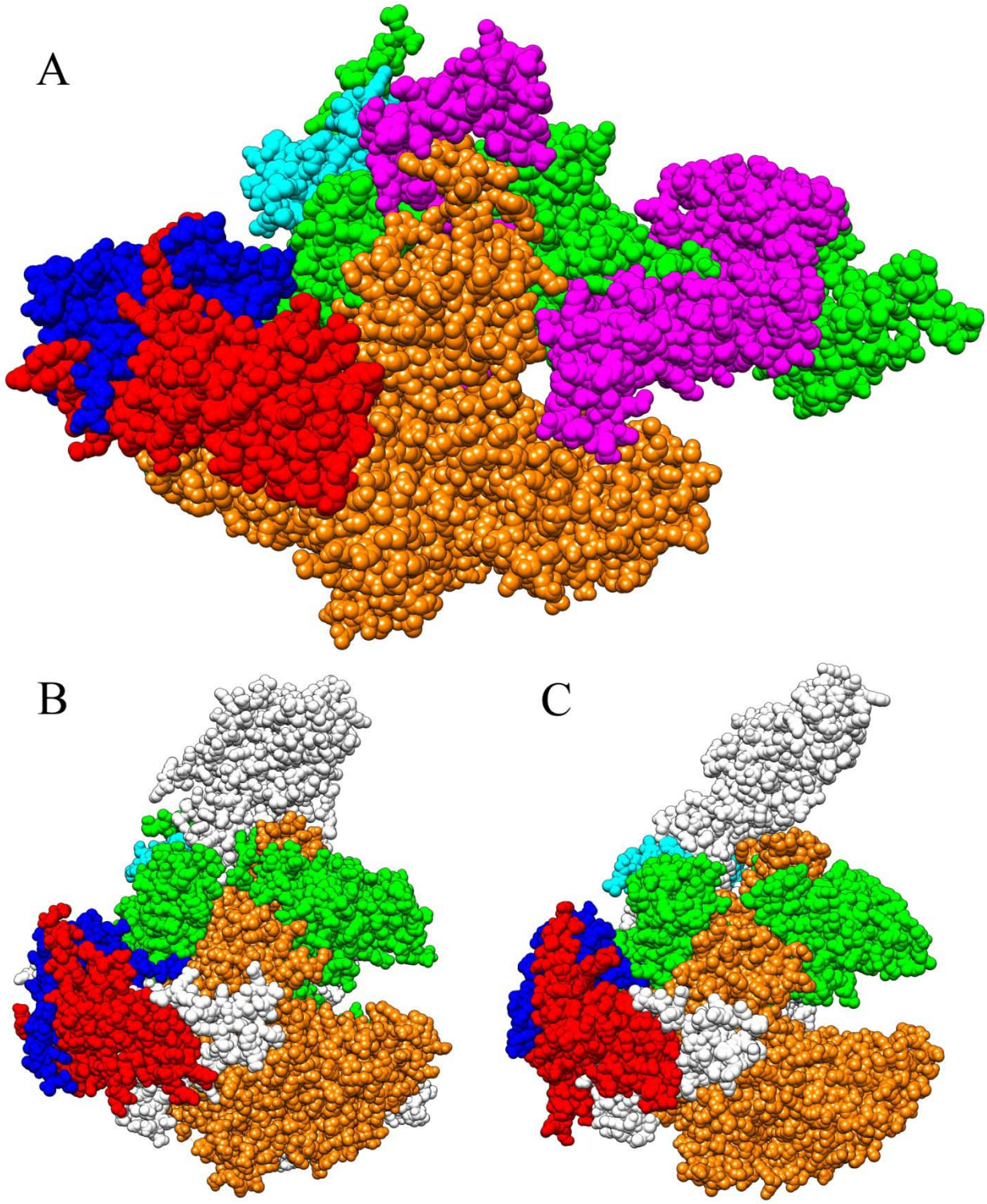

Figure 4. RNA polymerase from A) bacterium (*Thermus thermophilus*) B) eukaryote (*Saccharomyces cerevisiae*) C) archaeon (*Sulfolobus solfataricus*). In A), α_I and α_{II} are colored red and blue; β and β' are colored orange and green; ω is colored cyan; σ is colored magenta. The corresponding subunits in the eukaryotic and archaeal RNAPs have the same color. Additional subunits are colored white.

Transcription

Upon binding of the apoenzyme to the σ factor, the resulting holoenzyme is capable of recognizing specific promoter sequences at -10 and -35 bp upstream of the transcriptional start site. Initial binding to DNA forms the closed RNAP-promoter complex. The DNA sequence around the -10 site usually consists of several A-T basepairs that transiently unwind. Conserved aromatic residues in the sigma factor stack against the unwound bases, stabilizing the transiently melted form. Unwinding of several basepairs around the -10 site extending past the start site forms the transcription bubble and the open RNAP-promoter complex. In a process called abortive initiation, RNAP will repeatedly transcribe short segments of RNA, which will be prematurely released. Eventually, it will transcribe an RNA that reaches about 12 nucleotides in length. At this point, a sequence of conformational changes destabilizes interactions between the core RNAP and the sigma factor, allowing the RNAP to escape from the promoter, forming the ternary elongation complex (TEC). The TEC consists of the RNAP, DNA, and RNA, but it is unknown whether the sigma factor is completely dissociated from RNAP during elongation (Murakami and Darst 2003). The TEC is highly stable and processive, dissociating when it reaches specific termination signals that consist of a stable RNA hairpin followed by a U-rich tract (Herbert, Greenleaf *et al.* 2008).

α Subunit C-Terminal Domain

Despite the high degree of evolutionary conservation of RNA polymerase, *Helicobacter pylori* RNA polymerase demonstrates several structural differences from other known bacterial RNA polymerases. The β and β' subunits are about 45% identical to their *E. coli* counterparts. Although they are usually expressed as separate gene products in bacteria, they are found as one

fused gene product in *Helicobacter* and *Wolinella* species (Zakharova, Paster *et al.* 1999), and this fusion may confer a selective advantage (Dailidiene, Tan *et al.* 2007). Also, the primary specificity subunit of the RNA polymerase holoenzyme, σ^{80}, has diverged significantly compared to σ^{70} from *E. coli* and other bacteria (Solnick, Hansen *et al.* 1997). Sequence alignments between *H*. pylori and *E. coli* for all the subunits show that the α subunits (28% identical, 50% similar) have diverged even more than the σ subunits (32% identical, 51% similar). Table 1 shows sequence identities and similarities of all subunits from *H. pylori* and *E. coli*. Values were determined using the EMBOSS global alignment tool (Rice, Longden *et al.* 2000). Identity and similarity decrease away from the catalytic core. The ω subunits have unusually low similarity because they are small proteins and each sequence has a stretch of residues not present in the other.

Table 1. Sequence identity and similarity of *H. pylori* and *E. coli* RNAP subunits

Subunit	**Identity**	**Similarity**
α	28.4	49.7
αCTD	25.2	39.8
β	46.0	65.1
β'	45.3	61.5
ω	23.8	33.7
σ^{80}/ σ^{70}	32.1	50.9

The C-terminal domains of the α subunits (αCTDs) are separated from the dimerizing N-terminal domains by long, flexible linkers (Jeon, Yamazaki *et al.* 1997), and are essential for growth in *E. coli* and most likely all other bacteria (Hayward, Igarashi *et al.* 1991). They are

known to play an important role in transcription of certain genes. A highly conserved region of the domain interacts with upstream elements, binding the minor groove of A+T rich sequences that are often found near the consensus -10 and -35 elements of some promoters. This interaction drastically raises transcription levels at the *E. coli* rrnB P1 promoter (Ross, Gosink *et al.* 1993). The ways in which the αCTDs stimulate transcription are complex and varied. Several interaction surfaces of the domain are used to contact DNA and different transcription factors (Benoff, Yang *et al.* 2002; McLeod, Aiyar *et al.* 2002; Dangi, Gronenborn *et al.* 2004).

A sequence alignment of αCTDs (Figure 5) shows a highly conserved domain that consists of four α-helices and a long, ordered N-terminal loop. The conservation does not extend to several C-terminal residues. The first determined structure of an αCTD, from *E. coli*, showed that these C-terminal residues are, in fact, well-ordered and contribute key hydrophobic core interactions (Jeon, Negishi *et al.* 1995). In *H. pylori*, the αCTD C-terminal segment is several residues longer than in *E. coli*, and the two sequences are not at all similar. We determined the solution structure of αCTD from *H. pylori*. Our structure shows that not only are most of the C-terminal residues ordered, but that they form a fifth helix that is highly amphipathic and contributes more to the hydrophobic core than the corresponding sequence present in *E. coli* αCTD.

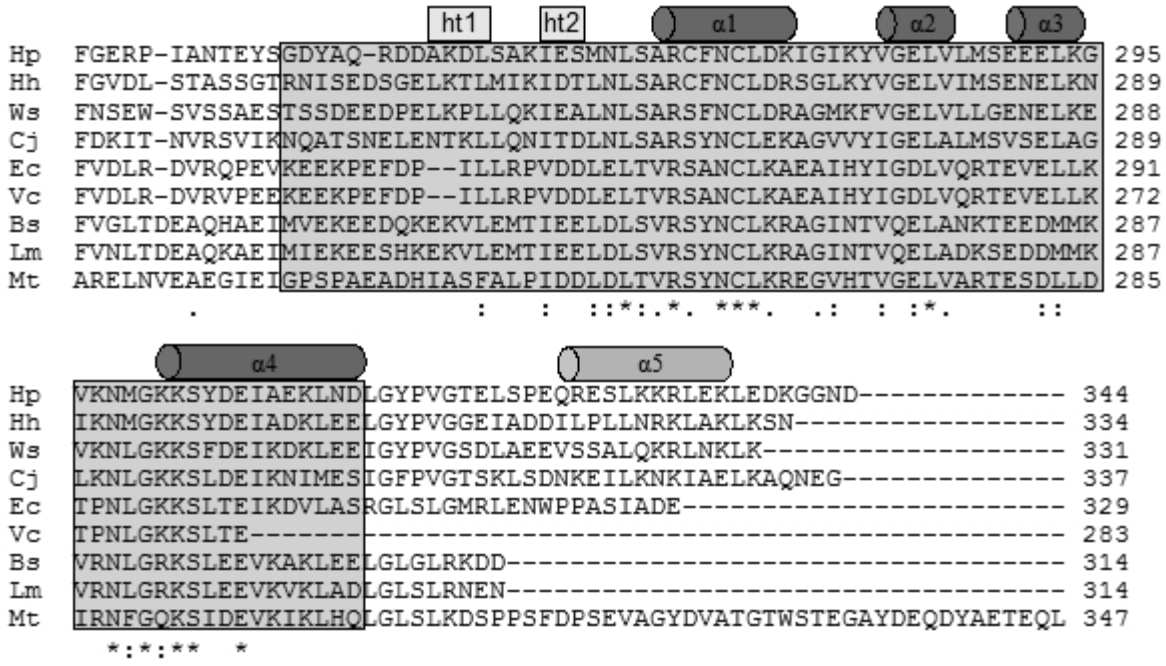

Figure 5. Amino acid sequence alignment of several bacterial αCTDs including part of the interdomain linker. PFAM-defined conserved domain is shaded gray. The two helical turns are indicated by boxes above the sequences and are labeled ht1 and ht2. α-helices are indicated by cylinders above the sequences and are labeled α1-α5. Helix α5 is an additional α-helix found in *H. pylori*. The conservation in all sequences presented is indicated by symbols in the bottom line; * invariant, : very highly conserved, . conserved. Numbers correspond to positions in the native sequence of each protein. Hp *Helicobacter pylori*; Hh *Helicobacter hepaticus*; Ws *Wolinella succinogenes*; Cj *Campylobacter jejuni*; Ec *Escherichia coli*; Vc *Vibrio cholerae*; Bs *Bacillus subtilis*; Lm *Listeria monocytogenes*; Mt *Mycobacterium tuberculosis.*

We have modeled *H. pylori* αCTD into known three-dimensional structures of the αCTD-containing complexes. Although much work has been done to describe αCTD interactions both structurally and functionally in *E. coli* and *Bacillus subtilis,* this type of analysis has not yet been done for *H. pylori*. Our models show that in many cases, the *H. pylori* αCTD would not be able to form the homologous protein-protein interactions. The inability of *H. pylori* and *E. coli* RNA polymerases to transcribe genes downstream of some promoters from the other bacterium due to differences in the σ subunits has been documented (Beier, Spohn *et al.* 1998; Shirai, Fujinaga *et al.* 1999). Our modeling suggests that the *H. pylori* αCTD interacts with cognate transcription

factors in novel ways and would also cause additional incompatibilities with some *E. coli* promoters.

Experimental Procedures

Protein Expression and Purification

Residues 231-344 encoded by the JHP1213 (HP1293, in sequenced strain 26695) gene from *H. pylori* strain J99, corresponding to the C-terminal domain of the α-subunit of RNA-polymerase and the flexible, interdomain linker were PCR amplified from genomic DNA using the following oligonucleotide primers: forward- 5'-gacggatccctgggcgtttttggcgaaag, reverse- 5'-gacggtaccgtttgtgtctcatcagtcgttacctcc. The PCR product was cloned into a modified pET vector that introduced an N-terminal, 12 residue His_6 tag (MRGSHHHHHHGS). Transformed *E. coli* BL21 (DE3) cells were grown in LB media to OD_{600}=1 and induced with 0.4 mM IPTG for 3 h. Cells were collected and lysed by sonication in binding buffer (20 mM Tris-HCl, 5 mM imidazole, 0.5 M NaCl, 8 M urea, pH 7.9). Cell extract was centrifuged to remove insoluble debris and filtered through a 2.7 μm filter. Soluble proteins were loaded onto Ni-NTA resin, washed (20 mM Tris-HCl, 30 mM imidazole, 0.5 M NaCl, 8 M urea, pH 7.9), and eluted (20 mM Tris-HCl, 0.5 M NaCl, 100 mM EDTA, 8 M urea, pH 7.9). Refolding was achieved by extensive dialysis against NMR sample buffer (25 mM KH_2PO_4, 225 mM KCl, 1mM TCEP, pH 7.3). Isotope-labeled samples were prepared by growing cells in M9 minimal media with $^{15}NH_4Cl$ and/or ^{13}C-u-glucose (CIL, Andover, MA) as sole sources of nitrogen and/or carbon.

NMR Experiments

NMR experiments were performed on Bruker Avance 600 and 800 MHz spectrometers equipped with cryoprobes. Samples were prepared at approximately 1 mM protein concentration in 25 mM KH_2PO_4, 225 mM KCl, 1mM TCEP, pH 7.3 and placed in 3 mm tubes to reduce the total salt in the detection volume. All experiments were performed at 25 °C. Singly-labeled ^{15}N samples were used to acquire 2D HSQC experiments. Doubly-labeled ^{15}N,^{13}C samples were used to acquire 3D CBCANH, CCCONH, HCCCONH, and HCCH-TOCSY experiments used for backbone and sidechain assignments. 3D ^{15}N and ^{13}C NOESY-HSQC experiments were used for assigning NOESY crosspeaks used in structure calculations.

A heteronuclear $\{^1H\}$-^{15}N NOE experiment was recorded to measure the backbone dynamics. Peak assignments from a ^1H-^{15}N HSQC were transferred to the spectra with and without saturation, and for each residue, the ratio of the intensities of the peak in the two spectra was taken as a measure of the steady-state heteronuclear NOE. Residues with peaks that overlapped in the spectrum (I239, Y244, D253, K255, D256, L293, K301, Y304, E309, D344) were excluded.

Chemical shift perturbation experiments were performed to observe DNA-binding and protein-protein interactions by αCTD. Protein concentration was decreased to 100 μM to minimize nonspecific binding to DNA. After addition of DNA, the salt concentration was gradually decreased to 50 mM. At this salt concentration, changes caused by protein-DNA interactions could be observed in the spectrum. Because αCTD tends to aggregate at such low salt concentrations, 50 mM arginine and 50 mM glutamate were added to the sample as a salt substitute (Hautbergue and Golovanov 2008). A fragment of dsDNA from the promoter region of gene HP1408 from strain 26695, with sequence 5'-gataaaataataaaaacgcatcattaaccattgattga, was

used. It contains a postulated αCTD binding site (Ross, Gosink *et al.* 1993) as well as a possible binding site for the acid-response regulator ArsR (Dietz, Gerlach *et al.* 2002). After mixing αCTD and DNA at 100 μM each, ArsR DNA-binding domain was added to the sample at 200 μM concentration to account for the two potential ArsR binding sites on the DNA fragment.

Structure Determination

NMR data were processed using TOPSPIN 2.0.b.6 (Bruker, Billerica, MA) and analyzed using SPARKY (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). Structure calculations were performed using CYANA (Guntert, Mumenthaler *et al.* 1997) version 2.1 with 25,000 steps for each structure. Distance restraints were calibrated automatically using CYANA routines. Hydrogen bond restraints were included only in later stages of calculations when they could be identified in a majority of structures. A total of 119 backbone φ and ψ dihedral angle restraints calculated using TALOS (Cornilescu, Delaglio *et al.* 1999) were used in the calculations. For the final round of calculations, 500 structures were calculated in CYANA, and the 50 with the lowest target function were energy-minimized using AMBER 9 (Case, Cheatham *et al.* 2005) with 3,000 steps of steepest descent energy minimization. Energy-minimized structures were analyzed with AQUA and PROCHECK-NMR (Laskowski, Rullmannn *et al.* 1996). The final ensemble consists of the 15 structures with the lowest energies and the best non-bonded backbone geometry. The PDB accession code for the ensemble is 2k8n. Chimera (Pettersen, Goddard *et al.* 2004) was used for visualization of structures and for producing figures. Delphi (Gilson 1987) was used for calculating the electrostatic surface potentials.

Results

Sequence Alignment

The bacterial RNA polymerase α subunit C-terminal domain family (Pfam (Finn, Tate *et al.* 2008) PF03118) is highly conserved with a length of about 68 residues (Figure 5). The average identity between members of this domain family is 45%. The conservation does not extend to a variable number of C-terminal residues of the polypeptide. Figure 5 shows a sequence alignment of nine bacterial αCTDs including the flexible linker between the N-terminal domain (NTD) and the CTD. The C-terminal sequences of the proteins are not conserved in terms of length or identity; however, among ε-proteobacteria, there is very good conservation in positions occupied by hydrophobic and charged residues.

Structure Determination

To determine the structure of the αCTD from *H. pylori*, we isolated a fragment consisting of the core domain, the interdomain linker, and the C-terminal residues to the end of the native protein sequence. We used NMR methods to determine a solution structure. Our final structural ensemble consists of 15 models with an average backbone RMSD of 0.41 Å and heavy atom RMSD of 0.93 Å (Table 2). Similar to other bacterial αCTDs, the well-conserved domain consists of four helices with short connecting loops, and a longer N-terminal, ordered loop (A254-S267) containing two helical turns that contributes residues critical to the hydrophobic core. Analyzing the final ensemble, we noticed that the backbone carbonyl oxygen atoms of N-terminal loop residues A259 and L257 act as hydrogen bonding partners for the backbone amide hydrogen atoms of helix 2 terminal residues V282 and G283, respectively. During the course of resonance assignments, the backbone amide proton of S267, also in the N-terminal loop, was

found far downfield shifted at 10.43 ppm, suggesting a strong hydrogen bond. Initial structure calculations showed the amide proton to be in the vicinity of the sidechain oxygens of E306. This glutamate is highly conserved, and known crystal structures confirmed these hydrogen bonds, so they were added as structural restraints in the remaining calculations. Helices 2 and 3 are rather short, while helices 1 and 4 are longer and contribute residues that are required for DNA-binding.

The NMR structure of the *E. coli* αCTD showed that the C-terminal loop, comprised of residues not included in the core domain alignment, is well-ordered and makes important hydrophobic contacts (Jeon, Negishi *et al.* 1995). This is also the case for the *H. pylori* αCTD, except that after a short loop, there is a fifth helix (Figure 6). Helix 5 is amphipathic, contributing three leucine sidechains that make many contacts with other hydrophobic residues in the hydrophobic core. Nearly all other residues in this helix have charged sidechains.

Table 2. Structural statistics for ensemble of 15 structures of *H. pylori* αCTD

Nonredundant NOE restraints	957
Intraresidue	316
Short	252
Medium	169
Long	220
Hydrogen bond restraints	36
TALOS dihedral angle restraints	119
Average CYANA target function	1.52
Number of violations > 0.2 Å	0
Average AMBER energies (± standard deviation)	
Input structures	3953
Energy minimized structures	-4354 (± 24)
Average Ramachandran statistics from PROCHECK (residues	
Most favored (%)	90.0
Additionally allowed (%)	9.4
Generously allowed (%)	0.6
Disallowed (%)	0
Average RMSD from mean structure (Å, residues 254-337)	
Backbone (N,C^{α},C',O)	0.41
Heavy atoms	0.93

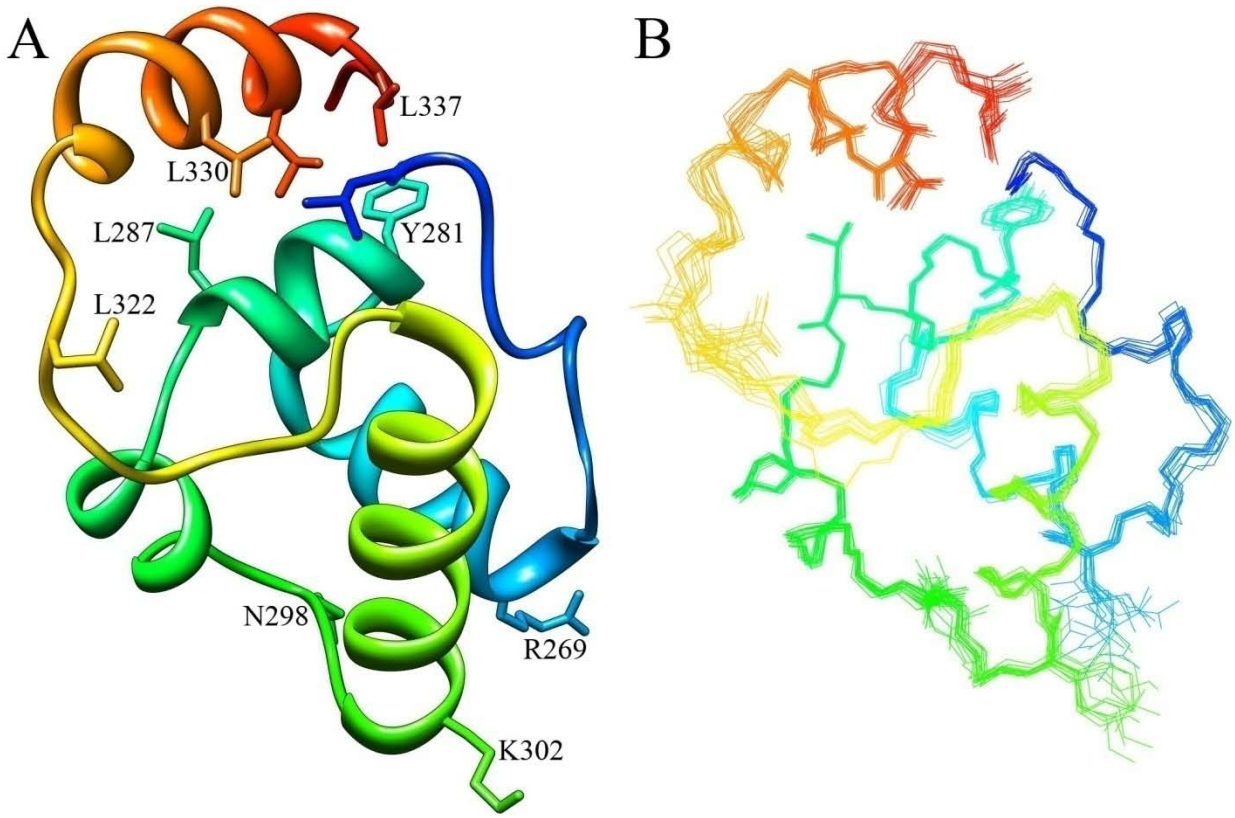

Figure 6. A) Ribbon diagram of the *H. pylori* RNA polymerase αCTD. Rainbow coloring, N-terminus blue and C-terminus red, was used. The fifth, C-terminal helix is visible at the top in orange and red. Sidechains of the three leucines from the helix as well as of other hydrophobic residues that they contact are shown. Well-conserved DNA-binding residues (R269, N298, K302) are also indicated. B) Wire diagram of ensemble of 15 conformers in the same orientation as in A).

Table 3. PROCHECK-NMR and AMBER statistics for 50 conformers

conformer #	most favorable	additionally allowed	generously allowed	disallowed	initial energy	final energy
17	90.8	9.2	0	0	2016	-4305
35	90.8	7.9	1.3	0	2346	-4373
8	90.8	7.9	1.3	0	2244	-4370
30	90.8	7.9	1.3	0	5917	-4359
48	90.8	7.9	1.3	0	3563	-4353
27	90.8	7.9	1.3	0	4408	-4320
3	89.5	10.5	0	0	2042	-4392
12	89.5	10.5	0	0	1484	-4359
40	89.5	10.5	0	0	7918	-4357
43	89.5	10.5	0	0	5471	-4355
7	89.5	10.5	0	0	3428	-4344
34	89.5	10.5	0	0	8481	-4340
38	89.5	10.5	0	0	3549	-4335
22	89.5	9.2	1.3	0	3744	-4390
23	89.5	9.2	1.3	0	2689	-4365
39	89.5	9.2	1.3	0	2379	-4361
10	89.5	9.2	1.3	0	4109	-4356
1	89.5	9.2	1.3	0	2565	-4338
14	89.5	9.2	1.3	0	1485	-4317
46	88.5	10.5	1.3	0	1705	-4354
31	88.2	11.8	0	0	2989	-4347
21	88.2	11.8	0	0	3368	-4343
42	88.2	11.8	0	0	7795	-4330
36	88.2	11.8	0	0	1263	-4324
24	88.2	11.8	0	0	1356	-4323
47	88.2	11.8	0	0	7682	-4321
26	88.2	10.5	1.3	0	8964	-4366
28	86.8	13.2	0	0	1212	-4363
37	86.8	13.2	0	0	6117	-4358
5	86.8	13.2	0	0	1512	-4356
6	86.8	13.2	0	0	3438	-4353
4	86.8	13.2	0	0	6615	-4347
19	86.8	13.2	0	0	1836	-4341
9	86.8	13.2	0	0	5822	-4335
44	86.8	13.2	0	0	9453	-4335
45	86.8	13.2	0	0	7908	-4333
2	86.8	13.2	0	0	1797	-4330
20	86.8	11.8	1.3	0	5848	-4357
25	86.8	11.8	1.3	0	2280	-4335
11	86.8	11.8	1.3	0	3720	-4319
18	85.5	14.5	0	0	3267	-4357
16	85.5	14.5	0	0	3146	-4348
13	85.5	14.5	0	0	1350	-4325

"Table 3, continued"

49	85.5	14.5	0	0	1262	-4312
50	85.5	14.5	0	0	3036	-4301
15	85.5	14.5	0	0	1710	-4285
29	84.2	15.8	0	0	1097	-4357
32	84.2	15.8	0	0	3223	-4337
41	88.2	10.5	0	1.3	1625	-4338
33	88.2	10.5	0	1.3	4481	-4334

A heteronuclear $\{^{1}H\}$-^{15}N NOE experiment (Figure 7) was recorded to determine the point at which the flexible, interdomain linker ends and the stable, structured domain begins, and also to determine the mobility of the C-terminal residues. The first residue with a steady-state NOE greater than 0.5 is L257, which is also involved in many hydrophobic core contacts. Residue A254 has an NOE of 0.5, on the borderline between the flexible and structured regions. This is consistent with the fact that we found just a few proton-proton NOE interactions between A254 and the structured domain. The heteronuclear steady-state NOEs also showed that the entire region corresponding to the well-conserved domain up to the end of helix 4 has a stable conformation. There is a steady decline in the heteronuclear NOEs from Y316 to L322. Our structural ensemble shows that this region has a greater range of conformations than the rest of the structured domain. On the other hand, the backbone of residues S323 to L337 appears well-ordered based on the observed heteronuclear NOEs. This is consistent with the formation of a helix that contributes residues to the hydrophobic core.

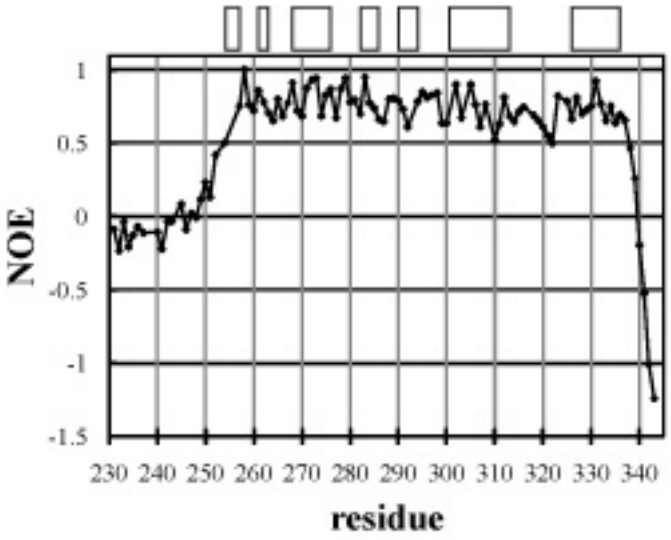

Figure 7. Heteronuclear $\{^{1}H\}$-^{15}N NOE values for backbone amides of αCTD. NOE values were calculated as a ratio of intensities of peaks in the spectra with and without saturation, and are plotted against the residue numbers. The two helical turns and the five α-helices are indicated by boxes above the graph.

DNA and Protein Interactions

Chemical shift perturbation experiments were performed to confirm DNA-binding by *H. pylori* αCTD. Optimal DNA recognition sequences have only been found for the *E. coli* αCTD (Estrem, Ross *et al.* 1999). We used a fragment of the promoter region of the *H. pylori* gene HP1408, which contains an element that is thought to be bound by the acid-response regulator ArsR (Dietz, Gerlach *et al.* 2002) as well as a sequence that appears similar to known αCTD binding sites in other bacteria. All amide peaks of αCTD broadened upon addition of the DNA fragment, indicating a binding event that formed a larger complex and affected the relaxation properties of the resonance signals. Several peaks nearly disappeared, and all of those peaks mapped to residues found on the DNA-binding surface of the protein (Figure 8). The additional broadening of peaks at the interface of the complex is a result of intermediate chemical exchange, wherein the resonant frequencies of nuclei near the interface differ depending on whether they are in a bound or unbound state and the binding and release events are occurring at

a rate on the order of the difference in the resonant frequencies of the nuclei in the two states. No additional changes in the HSQC spectrum of αCTD were observed when the ArsR DNA-binding domain was added to the sample. However, we do not know whether the ArsR DNA-binding domains bind strongly enough to this DNA fragment to allow such observations.

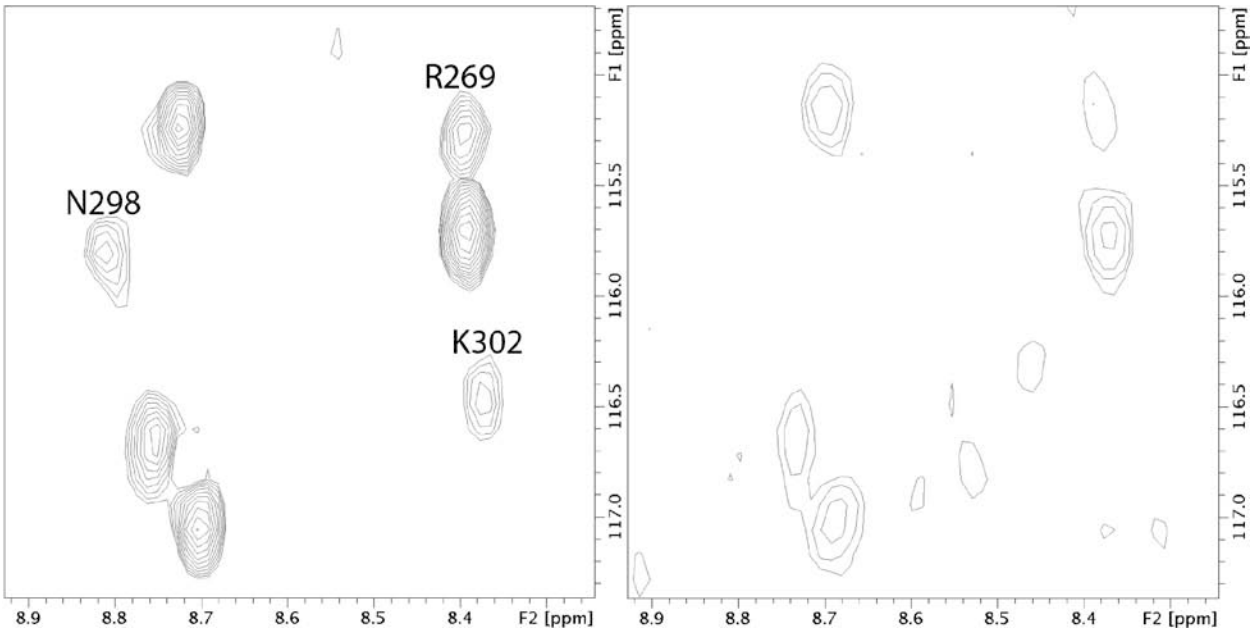

Figure 8. Sections of ^{1}H-^{15}N HSQC spectra of *H. pylori* αCTD. A) αCTD alone. B) αCTD with bound DNA. Binding of DNA causes all peaks in the spectrum to broaden, but peaks corresponding to some of the residues on the DNA-binding surface have disappeared.

Model Building

Although the core of the domain is highly similar to that of other αCTDs, especially around the DNA-binding surface, we wanted to gain more insight into its possible protein-protein interactions. We produced electrostatic surface potentials for both the *H. pylori* and *E. coli* αCTDs (Figure 9). Overall, the two proteins have very similar isoelectric points (5.31 for *H. pylori* vs. 5.41 for *E. coli*); however, there are significant differences in surface charge

distribution. The surface area of the C-terminal segments in the *H. pylori* αCTD is much more charged than that of *E. coli*, which is fairly hydrophobic.

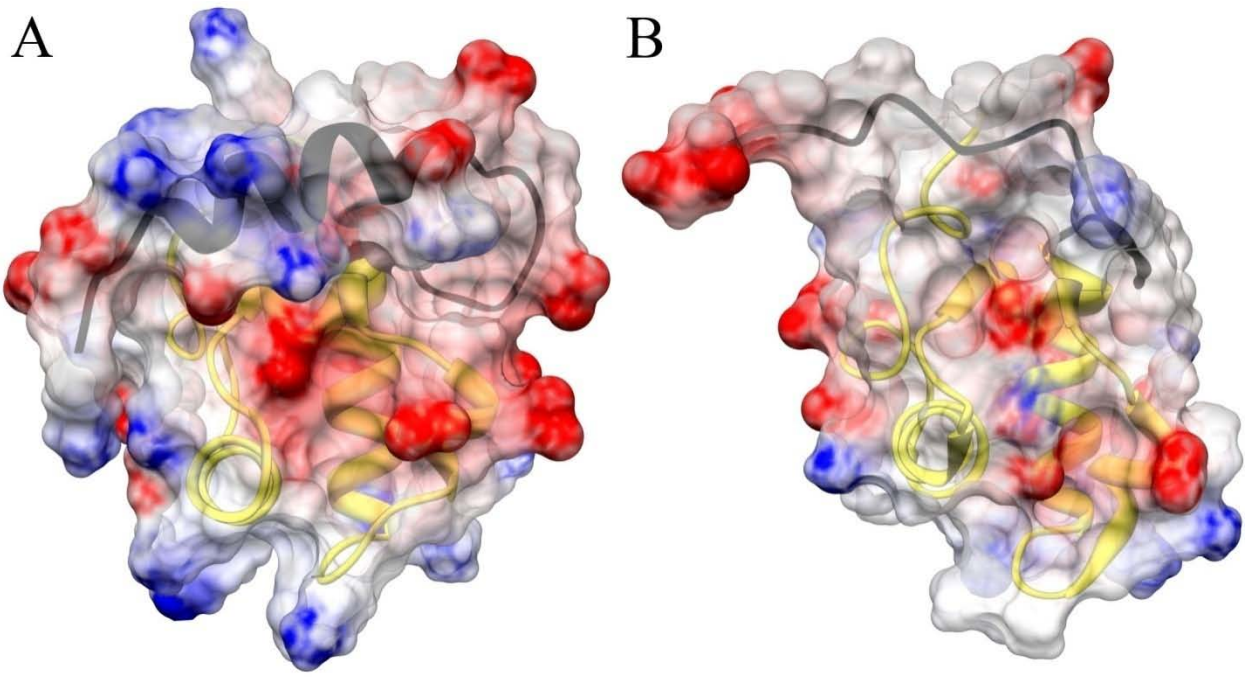

Figure 9. Electrostatic surface potentials. A) *H. pylori* αCTD. B) *E. coli* αCTD. Blue indicates positive charge; red indicates negative charge. The structurally divergent C-terminal fragments are indicated by black coloring of the ribbons. The C-terminal helix from *H. pylori* αCTD presents a more charged surface to the solvent than the hydrophobic ordered loop from the *E. coli* αCTD.

There are a few high-resolution structures that include αCTD in complex with other proteins or DNA. We used the structure of the *E. coli* αCTD in complex with the catabolite activator protein (CAP) and DNA (Benoff, Yang *et al.* 2002) to model possible interactions of *H. pylori* αCTD. CAP, also known as the cAMP receptor protein, activates transcription at promoters such as P_{lac} and P_{gal} in *E. coli*. Given the conservation of both sequence and structure around the DNA-binding site, the *H. pylori* αCTD interaction with DNA is likely very similar to

that of other αCTDs. Our structure was superimposed on the *E. coli* αCTD in complex with CAP (Figure 10A). Only a few minor conformational changes of the sidechains of the DNA-binding face would be required for the *H. pylori* αCTD to fit at this interface.

Although the structures of the *H. pylori* and *E. coli* αCTDs superimpose quite closely in the model, the backbone of helix 3 from the *H. pylori* αCTD is found much closer to CAP. At the αCTD/CAP interface (Figure 10B), there is a steric clash due to a bulkier residue (E291) that would probably preclude the interaction. This model orients αCTD with respect to the DNA and other possible DNA-bound proteins and shows that the flexible portion of the C-terminal segment that precedes the amphipathic helix extends further away from the core of the protein than in *E. coli*. It is also well-placed for interacting with other proteins bound to promoter sequences. This region of the protein is part of the "287 determinant" in the *E. coli* αCTD and is known to be important for protein-protein interactions with CAP and the *E. coli* FNR protein, which activates transcription in response to oxygen starvation (Lee, Wing *et al.* 2000).

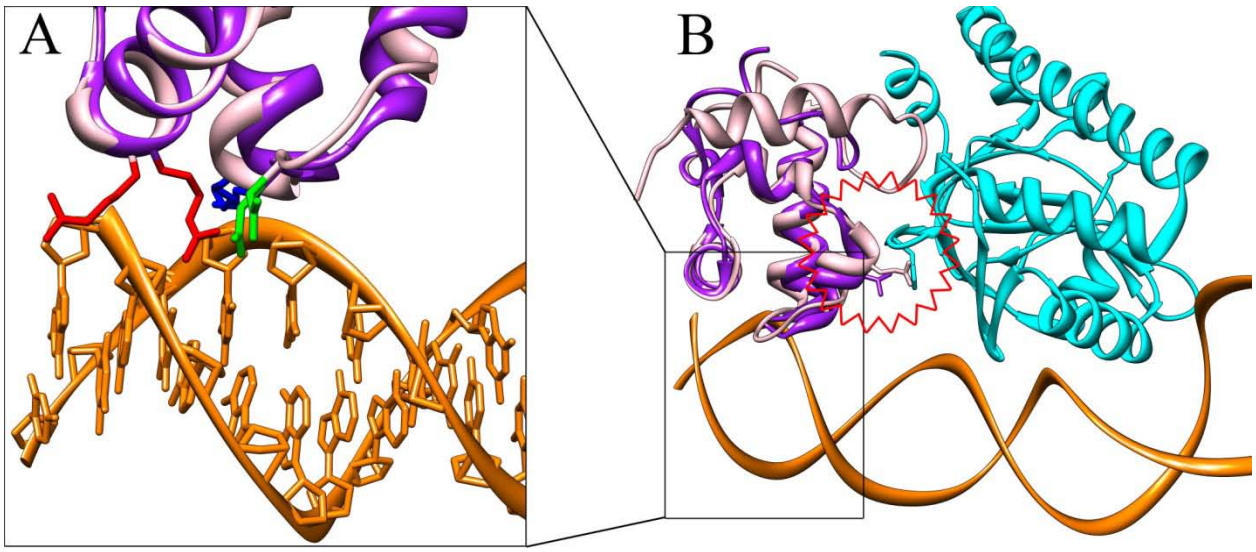

Figure 10. Structure of *H. pylori* αCTD superimposed on the structure of *E. coli* αCTD in complex with DNA and CAP. *H. pylori* αCTD shown in pink; *E. coli* αCTD shown in purple; CAP shown in cyan; DNA shown in orange. A) Interaction between αCTD and DNA. Well-conserved DNA binding residues, R269(R265), N298(N294), and K302(K298) are shown in red, green, blue, respectively. Residue numbers are given for *H. pylori* (*E. coli*). B) Interaction between αCTD and CAP. The position of helix 3 as well as its particular residue sidechains might interfere with CAP binding in the red outlined area.

Discussion

The structure of the conserved core domain of *H. pylori* αCTD (residues 257-318 in *H. pylori* numbering) is similar to that of *E. coli* αCTD (PDB accession code 1coo), with a backbone RMSD of 2 Å. The most noticeable difference between the *H. pylori* αCTD and the *E. coli* αCTD is in the C-terminal, sequentially divergent segment. Although both segments provide critical hydrophobic core residues, their backbone conformations are very different. While the *E. coli* αCTD C-terminal segment is well-ordered and extended without a regular secondary structure, the *H. pylori* αCTD segment forms a highly amphipathic α-helix. *H. pylori* αCTD also contains additional hydrophobic residues in helix 3. L287 from *H. pylori* replaces Q283 from *E. coli* and makes contact with L330 from helix 5.

H. pylori αCTD superimposes more closely on the αCTD from *Bacillus subtilis* (Newberry, Nakano *et al.* 2005) over the core domain (residues 257-318), with a backbone RMSD of only 1 Å. The C-terminal ends of the *H. pylori* and *B. subtilis* αCTDs, however, are even more divergent than *H. pylori* and *E. coli*. The *B. subtilis* αCTD does not loop back around to make hydrophobic contacts with the N-terminal loop or the beginning of helix 3. This can be explained by the fact that the *B. subtilis* αCTD helix 3 is much more hydrophilic than that of *H. pylori* αCTD. Y281 from *H. pylori*, which is at the N-terminal end of helix 3, makes contact with L334 and L337 from helix 5. In *B. subtilis*, the corresponding residue is a threonine (T273). Hydrophobic residues L287 and M288 at the end of helix 3 from *H. pylori* correspond to residues N279 and K280 from *B. subtilis*.

Many different interaction surfaces are used by αCTD to interact with both DNA and other proteins. In *E. coli*, three determinants have been identified that are required for transcription activation by CAP (Benoff, Yang *et al.* 2002). The "265 determinant", which includes residues from helix 1, the turn between helices 3 and 4, and helix 4 that are used in binding the minor groove of DNA, is highly conserved in all αCTDs. The "261 determinant", composed of residues in the N-terminal loop that may come into contact with σ^{70}, is similar in terms of structure and sequence identity. The "287 determinant", which includes residues in helix 3 and the first part of the C-terminal loop, is similar to *H. pylori* αCTD in terms of sequence identity in helix 3 residues, but is different in terms of both sequence and structure in the C-terminal loop. The *E. coli* Fis transcriptional activator interacts with residues 271-273 (McLeod, Aiyar *et al.* 2002) (*E. coli* numbering), which are not similar in *H. pylori*. MarA, another *E. coli* transcription activator, actually interacts with αCTD at the same site where αCTD would normally interact with DNA (Dangi, Gronenborn *et al.* 2004).

At this point, there are no reports of specific interactions between *H. pylori* transcription factors and αCTD or other subunits of RNA polymerase, nor do we know consensus DNA-binding sequences for many of the transcription factors. Therefore, we used modeling to explore possible protein-protein interactions of αCTD, and especially the role that the unique C-terminal segment might play in them. The models suggest that the *H. pylori* αCTD would not be able to interact with transcription factors from *E. coli*, the system in which these interactions have been best studied. Our model based on the αCTD/CAP/DNA complex shows that the part of the C-terminal segment that is less well-ordered is perfectly situated for making contacts with transcription factors on DNA. Although it does not place the amphipathic helix in a position where it can directly interact with the adjacent CAP protein, it is difficult to say whether or not it is involved in protein-protein interactions due to a lack of other experimentally determined complex structures showing alternative interaction modes.

The αCTD from *H. pylori* and other ε-proteobacteria features a fifth helix with a well-conserved pattern of hydrophobic and charged residues, even though these bacteria have some of the most rapidly evolving genomes. Studies have shown that *H. pylori* has one of the highest rates of mutation due to its apparent lack of several DNA repair enzymes (Wang, Humayun *et al.* 1999) and also one of the highest rates of exogenous DNA uptake and homologous recombination of all bacteria (Suerbaum and Josenhans 2007). The fact that the features of αCTDs are highly conserved on both the tertiary and primary levels indicates that they most likely serve a specific function. If they did not, one would expect to see greater sequence diversity. Given that the αCTDs from other bacteria use so many different interaction surfaces for protein-protein interactions, it is possible that the fifth helix forms a required interaction site for other transcriptional regulators. Although interactions between αCTD and several

transcriptional regulators have been studied in *E. coli*, this work has not yet been done for *H. pylori*. Compared to *E. coli*, *H. pylori* has a relatively small genome (Tomb, White *et al.* 1997; Alm, Ling *et al.* 1999), with fewer identified transcriptional regulators. We look forward to determining which transcriptional regulators interact with αCTD and mapping their interaction surfaces.

Conclusions and Future Directions

Finding an additional helix in a protein that has such great overall sequence conservation is exciting. In the future, we would like to determine the importance of this helix as well as all of the other structural differences. Although the parts of RNAP that are directly involved in the catalytic process of transcription are highly conserved among all organisms, parts that are involved in the regulation of transcription are less conserved. Overall, the regulation of transcription is species-specific. Therefore, no single organism can serve as an adequate model organism for transcriptional regulation. It is worthwhile studying the interactions of RNAP with transcription factors. RNA polymerase is a promising target for new antibiotics because of significant structural differences compared to eukaryotic RNA polymerases (Artsimovitch and Vassylyev 2006). If unique features of the *H. pylori* RNAP could be targeted, patients would benefit from eradication of their *H. pylori* infections without suffering from side effects due to the killing of all other intestinal bacteria.

At this point, we speculate that the C-terminal helix of the αCTD from *H. pylori* and other ε-proteobacteria plays a role in protein-protein interactions. The amphipathic helix presents several charged residues to the solvent, in contrast to the hydrophobic, ordered C-

terminal residues of *E. coli*. To determine the importance of these charged residues, we would mutate them one by one or in combination.

We would also like to determine interacting partners of αCTD and map their interaction surfaces. This would be a long-term project because we do not yet know which transcription factors might interact with it, nor do we have consensus DNA-binding sequences for all of them.

CHAPTER III

STRUCTURE DETERMINATION OF HP0564

A large part of the material presented in this chapter has been taken from published work: Borin BN, Krezel AM. Structure of HP0564 from Helicobacter pylori identifies it as a new transcriptional regulator. Proteins 2008 Oct;73(1):265-8.

Introduction

As part of a small-scale *H. pylori* structural genomics project in our laboratory, several target genes were identified that satisfied several criteria: i) no sequence homologs in bacteria, ii) sequence properties suggesting feasibility of structural determination by NMR, iii) some hint of biological interest (Popescu 2004). From among the final list of target genes, HP0222 was chosen for its potential relevance to two major environmental responses of *H. pylori.* Microarray studies had shown HP0222 to be highly upregulated upon exposure to low pH (Ang, Lee *et al.* 2001) as well as upon attachment to gastric epithelial cells (Kim, Marcus *et al.* 2004).

Through structure determination, HP0222 was found to be a member of the ribbon-helix-helix (RHH) superfamily of transcriptional regulators (Popescu, Karpay *et al.* 2005). Among all known protein sequences, HP0222 has a single sequence homolog that is encoded by another *Helicobacter pylori* gene, HP0564. We determined the structure of HP0564 to confirm that it is also a member of the RHH superfamily of transcriptional regulators and also to compare the structure of HP0222 and HP0564 (Borin and Krezel 2008). Although it has no assigned function, our structural analysis indicates that it is a member of the ribbon-helix-helix

superfamily (Pfam protein domain family PF01402) of transcriptional regulators. These proteins bind to specific DNA sequences with high affinity and usually act as repressors.

Ribbon-Helix-Helix Proteins

Fold

The RHH fold was first described in 1989 with the crystal structure of the *E. coli* MetJ repressor (Rafferty, Somers *et al.* 1989). Since then, over 2000 sequences have been putatively identified as belonging to the RHH family, but only a few have been characterized biochemically or structurally and are confirmed RHH proteins. So far, RHH proteins have been found in prokaryotes, archaea, and bacteriophages. Putative RHH sequences from eukaryotes do not demonstrate certain sequence motifs present in known RHH proteins (Schreiter and Drennan 2007).

RHH proteins are named for their characteristic secondary structural elements. All have a short β-ribbon followed by an α-helix (helix α1), a short loop, and a second α-helix (helix α2) (Figure 11b). The β-ribbon consists of approximately seven residues and is followed immediately by helix α1. Both helices consist of about fourteen residues, and they are separated by a three-residue linker. In solution, RHH proteins are found as intertwined homodimers, where dimerization produces a double-stranded, antiparallel β-sheet with solvent-exposed sidechains that are used in making sequence-specific contacts with DNA. The hydrophobic core of the protein is rather small, consisting of the sidechains of three residues from the β-sheet, one residue from helix α1, and three residues from helix α2, for each of the subunits. Branched chain hydrophobic amino acids are very well-conserved at these positions (Schreiter and Drennan 2007).

There are only minor differences in the structures of some of the RHH proteins, mainly in the lengths of the helices and the linker between them. Overall, the pairwise backbone RMSD between known RHH structures is quite low (Schreiter and Drennan 2007). Despite having very similar structures, sequence identity among RHH proteins is very low, making it difficult to recognize new ones from their sequences alone. HP0222 and HP0564 from *H. pylori* have significant sequence homology, but neither produces any hits from other species in a BLAST search.

Function

In general, RHH proteins bind DNA through their N-terminal β-sheets. More specifically, they have very diverse functions within cells. Arc (Susskind 1983) and Mnt (Levine, Truesdell *et al.* 1975) are both involved in regulation of the bacteriophage lytic cycle. CopG (del Solar, Acebo *et al.* 1995) and omega (de la Hoz, Ayora *et al.* 2000) are involved in plasmid copy number control. MetJ represses transcription of genes that encode for proteins in the methionine biosynthesis pathway (Saint-Girons, Parsot *et al.* 1988). Most RHH proteins act as repressors; however, there have been some reports of positive regulation, as well. In *E. coli*, NikR responds to high concentration of Ni^{2+} and represses transcription of the nikABCDE operon, which encodes for a nickel transport system (De Pina, Desjardin *et al.* 1999). In *H. pylori*, however, NikR not only represses the NixA nickel transporter, it also upregulates transcription of urease (Ernst, Kuipers *et al.* 2005). AlgZ has been reported to be a positive regulator of the alginate synthesis operon in *Pseudomonas aeruginosa* (Baynham, Brown *et al.* 1999).

DNA-Binding

RHH proteins bind specific DNA sequences by placing their β-sheets in the major groove of DNA. The original model of DNA-binding by RHH proteins, however, suggested that the second α-helices were placed into the major groove, consistent with the fact that all known transcription factors at the time, including the helix-turn-helix family, used α-helices to bind DNA. In fact, the RHH family was the first that was found to bind DNA using a β-sheet instead of an α-helix (Breg, van Opheusden *et al.* 1990; Somers, Rafferty *et al.* 1994).

Functional RHH protein/DNA complexes always involve higher orders of protein oligomerization. The resulting cooperative binding effectively increases the binding affinity. Each dimer recognizes a specific DNA sequence, termed a half-site because many RHH proteins tetramerize, with each dimer binding one half-site. Usually, binding sites contain inverted-repeat half sites. MetJ, however, binds a site consisting of five tandemly-repeated half-sites (Davidson and Saint Girons 1989). Spacing between half-sites is usually small, but there are cases where the spacing is quite large. NikR tetramerizes in a nickel-dependent way through an additional domain, placing the two RHH domains on half-sites that are 16 bp apart (Contreras, Thiberge *et al.* 2003).

Materials and Methods

Protein Expression and Purification

For structural work, a shortened construct of HP0564 was created that lacks the flexible N-terminal 20 residues as well as the C-terminal 7 residues, leaving only the stably folded region. The corresponding sequence was PCR amplified from genomic DNA of *H. pylori* strain J99 and cloned into a modified pET vector with an N-terminal, 12 residue His_6 tag

(MRGSHHHHHHGS). Transformed *Escherichia coli* BL21 (DE3) cells were grown in LB media to the OD_{600}=1 and induced with 0.4 mM IPTG for 3 h. Cells were spun down, resuspended in binding buffer (20 mM Tris, 0.5 M NaCl, 5 mM imidazole, 8 M urea, pH 7.9), and disrupted by sonication (6x30 seconds). Filtered (1 µm) cell extract was loaded on a Ni-NTA column, followed by a 100 mL wash (20 mM Tris, 0.5 M NaCl, 30 mM imidazole, 8 M urea, pH 7.9) and elution (20 mM Tris, 0.5 M NaCl, 0.1 M EDTA, 8 M urea, pH 7.9). Refolding was achieved by dialysis against distilled water. No additional protein bands could be detected by tricine, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

Isotope-labeled samples were prepared by growing cells in M9 minimal media supplemented with $^{15}NH_4Cl$ and/or ^{13}C-u-glucose (CIL, Andover, MA). All other aspects of the expression and purification of labeled samples were identical to those used for natural abundance protein.

Crosslinking Experiments

Crosslinking experiments were performed with BS^3 (Pierce, Rockford, IL). Reaction buffer was 20 mM NaH_2PO_4, pH 7.0. Crosslinker was dissolved in reaction buffer to 10 mM stock concentration immediately prior to setting up reactions. All reactions were in 20 µL, consisting of 17 µL reaction buffer, 2 µL HP0564 (10 µg/µL), and 1 µL of an appropriate dilution of BS^3. Final concentrations of 0, 0.005, 0.05, and 0.5 mM BS^3 were used. Reactions were allowed to proceed for 3 minutes before being quenched with 5 µL of 1 M Tris, pH 7.5. All reactions were run on a 10% SDS-PAGE gel and stained with Coomassie Blue.

Gel Filtration Experiments

Size exclusion separations were performed on a Superdex 75 10/30 FPLC column (Pharmacia, Piscataway, NJ) at 4 °C in 50 mM KH_2PO_4, pH 4.0. Elution was followed by UV absorption at 214 nm. The calibration curve used to calculate the molecular weight was prepared with ubiquitin, thioredoxin, and ovalbumin run under identical conditions.

NMR Experiments

NMR experiments were performed on Bruker Avance 600 and 800 MHz spectrometers at 25 °C. Samples were prepared at 1 mM monomer concentration in 50 mM KH_2PO_4, pH 4.0. Natural abundance protein was used to acquire ^{1}H 2D NOESY spectra using mixing times of 25, 50, and 100 ms. Singly-labeled ^{15}N and ^{13}C samples were used to acquire 2D HSQCs. Doubly-labeled ^{15}N,^{13}C samples were used to acquire 3D HNCO, CBCANH, CCCONH, and HCCCONH experiments used for backbone and sidechain assignments (Sattler, Schleucher *et al.* 1999). Over 98% of backbone resonances (H^N, N, C^α, H^α, C') and 85% of commonly assignable carbon and proton sidechain resonances were assigned.

Structure Calculations

NMR data were processed using XWINNMR (Bruker, Billerica, MA) and analyzed using SPARKY (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). Structure calculations were performed using CYANA (Guntert, Mumenthaler *et al.* 1997) version 2.1 with 25,000 steps for each structure. NOE crosspeaks corresponding to both intramolecular and intermolecular interactions were assigned manually, and intensities were automatically converted to distance restraints using built-in CYANA routines. Given the small

size of the protein (7.8 kDa monomer), the 2D NOESY was sufficiently resolved to assign all crosspeaks. 3D heteronuclear-resolved NOESY spectra were recorded, but offered no additional distance information and were not used in structure calculations. In the initial stages of calculations, only NOE-derived restraints were used. Hydrogen bond restraints were added in later stages when they could be identified in a majority of calculated structures. In the last stage, out of 1,000 initial structures, the 50 with the lowest target function values were minimized in AMBER (Case, Cheatham *et al.* 2005) version 9 using 10,000 steps of conjugate gradient energy minimization (Table 4). Of these 50 energy-minimized structures, the 20 with the lowest nonbonded backbone energies were used in the final ensemble, which was analyzed using AQUA and PROCHECK-NMR (Laskowski, Rullmannn *et al.* 1996) (Table 5). The PDB entry, including the structural ensemble as well as the restraints used in structure calculations, has the PDB accession code 2k1o. BMRB entry 15761 contains ^{1}H, ^{13}C, and ^{15}N chemical shift assignments. Chimera (Pettersen, Goddard *et al.* 2004) was used for interactive analysis and figure production.

Table 4. PROCHECK-NMR and AMBER statistics for 50 conformers

conformer #	most favorable	additionally allowed	generously allowed	disallowed	initial energy	final energy
6	91.2	8.8	0	0	3987.90	-5111.10
19	90	10	0	0	3683.10	-5133.20
10	90	10	0	0	3967.00	-5119.90
25	90	10	0	0	3897.60	-5112.90
47	90	8.8	1.2	0	3931.80	-5126.20
1	90	8.8	1.2	0	3953.10	-5120.00
17	90	8.8	1.2	0	3910.20	-5102.70
20	88.8	11.2	0	0	3999.70	-5154.80
31	88.8	11.2	0	0	3815.30	-5126.00
28	88.8	11.2	0	0	3830.00	-5122.60
18	88.8	11.2	0	0	3404.00	-5122.20
37	88.8	11.2	0	0	3924.00	-5120.10
39	88.8	10	1.2	0	3945.20	-5126.00
36	87.5	12.5	0	0	3928.90	-5148.60
38	87.5	12.5	0	0	4081.30	-5134.40
16	87.5	12.5	0	0	4031.20	-5123.00
15	87.5	12.5	0	0	3943.40	-5120.60
29	87.5	12.5	0	0	3867.50	-5116.80
13	87.5	12.5	0	0	3794.60	-5111.60
30	87.5	12.5	0	0	3981.30	-5102.70
34	87.5	12.5	0	0	3979.80	-5092.00
49	87.5	12.5	0	0	3823.60	-5081.20
7	87.5	12.2	0	0	3693.50	-5087.60
44	87.5	11.2	1.2	0	3916.50	-5165.20
35	87.5	11.2	1.2	0	4007.70	-5134.00
5	87.5	11.2	1.2	0	3824.40	-5116.70
22	86.2	13.8	0	0	3745.40	-5153.70
4	86.2	13.8	0	0	3978.90	-5147.50
40	86.2	13.8	0	0	3986.60	-5139.20
12	86.2	13.8	0	0	3605.10	-5139.00
2	86.2	13.8	0	0	4067.20	-5137.10
24	86.2	13.8	0	0	3980.30	-5133.50
8	86.2	13.8	0	0	3608.70	-5124.80
45	86.2	13.8	0	0	3882.90	-5123.80
33	86.2	13.8	0	0	3909.90	-5112.50
42	86.2	12.5	1.2	0	3842.90	-5123.10
11	86.2	12.5	1.2	0	3937.30	-5114.30
50	86.2	12.5	1.2	0	3790.90	-5112.20
27	86.2	12.5	1.2	0	3815.10	-5103.70
41	85	15	0	0	3967.90	-5145.30
21	85	15	0	0	3794.70	-5136.60
14	85	15	0	0	3876.60	-5135.20
32	85	15	0	0	3873.40	-5125.10

"Table 4, continued"

23	85	15	0	0	3961.60	-5116.70
46	85	12.5	2.5	0	4005.00	-5140.60
48	83.8	16.2	0	0	3772.90	-5112.20
3	87.5	11.2	0	1.2	3969.60	-5126.10
26	87.5	11.2	0	1.2	3963.00	-5116.80
9	86.2	12.5	0	1.2	3895.60	-5136.10
43	86.2	12.5	0	1.2	3686.90	-5106.40

Table 5. Structural statistics for ensemble of 20 structures of HP0564 (JHP0511)

NOE restraints	797
Intraresidue	222
Short	220
Medium	134
Long	221
Intramolecular	626
Intermolecular	171
Hydrogen bonds per dimer	44
Average CYANA target function	0.11
Number of violations > 0.2 Å	0
Average AMBER energies (± standard deviation)	
Input structures	-3894 (± 146)
Energy minimized structures	-5123 (± 13)
Average Ramachandran statistics from PROCHECK (residues 23-62)	
Most favored (%)	88.8
Additionally allowed (%)	10.9
Generously allowed (%)	0.2
Disallowed (%)	0
Average RMSD from mean structure (Å, residues 23-62)	
Backbone (N,C^{α},C',O)	0.59
Heavy atoms	1.08

Results and Discussion

A Genbank search with the DNA or protein sequence of *Helicobacter pylori* HP0564 (Uniprot Q9ZLR7_HELPJ) yields no orthologs and only one paralog, HP0222 (Popescu, Karpay *et al.* 2005) (Uniprot Q9ZML0_HELPJ), which has 17 identical residues out of 40 in the stably folded region consisting of residues 23-62 (Figure 11b). The structure of HP0564 shows it to be a member of the ribbon-helix-helix (RHH) superfamily of transcriptional regulators (Figure 11a). The ensemble of 20 conformers of HP0564 is shown in (Figure 12). A DALI (Holm and Sander 1996) search yielded the Arc repressor (PDB accession code 1baz), CopG (PDB accession code 1ea4), and HP0222 (PDB accession code 1x93) as its closest structural relatives, all with Z-scores greater than 5.0. These proteins are always found in solution as dimers (Breg, van Opheusden *et al.* 1990). Dimerization creates an antiparallel double-stranded β-sheet with several sidechains exposed to solvent that are used in making sequence-specific contacts with DNA (Raumann, Rould *et al.* 1994). Upon binding DNA, proteins in this superfamily form tetramers or higher order oligomers, where each dimer binds several base pairs of DNA.

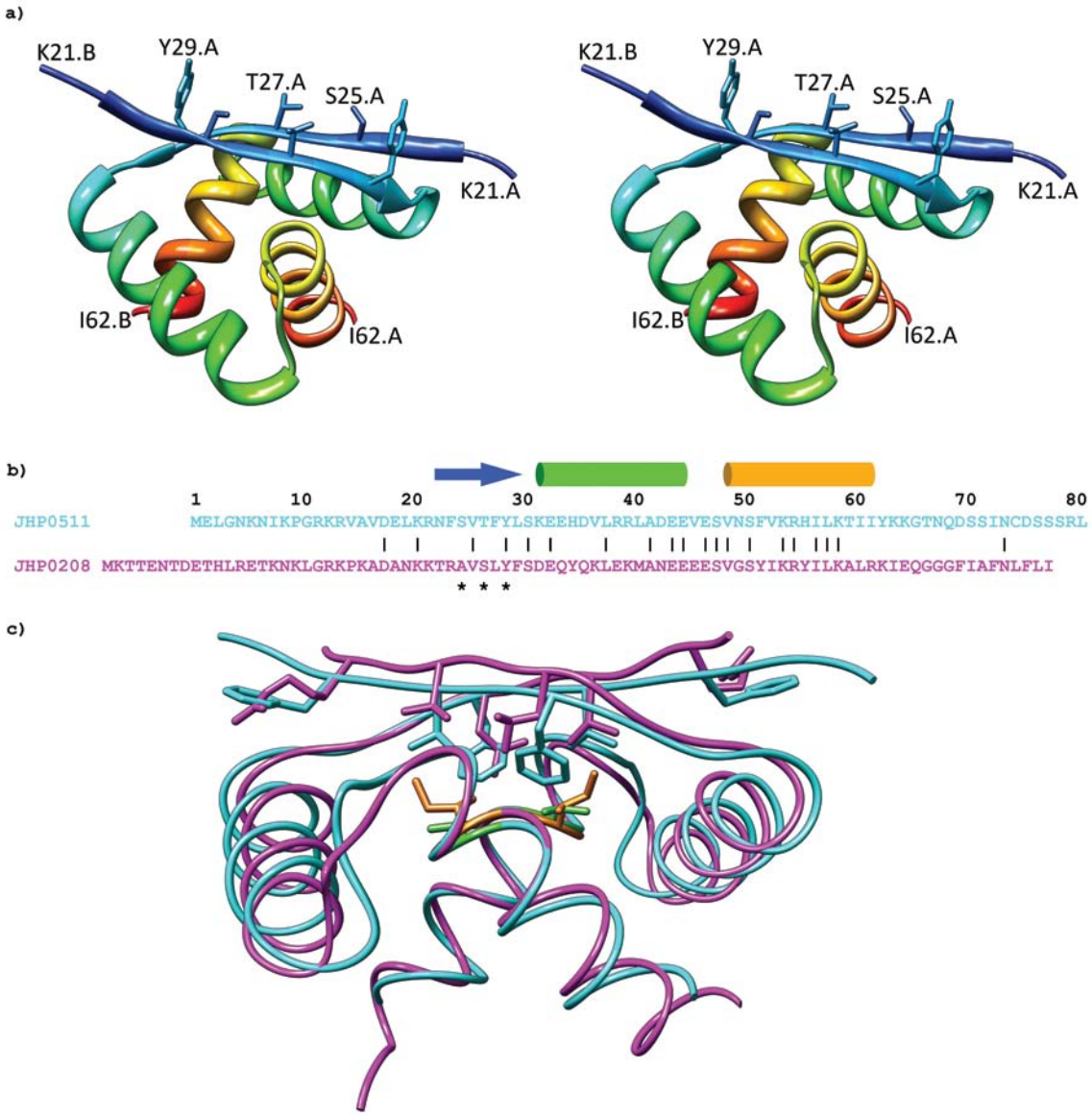

Figure 11. a) Stereo view ribbon diagram of HP0564 showing residues 21-62 of each subunit. Side chain heavy atoms of β-sheet residues S25, T27, and Y29 that make up the DNA binding interface of ribbon-helix-helix proteins are labeled for one subunit. b) Sequence alignment of JHP0511 (HP0564) and JHP0208 (HP0222) from the J99 strain of *H. pylori.* Numbering is according to the JHP0511 sequence. Secondary structural elements are indicated for HP0564, with an arrow representing the β-strand and cylinders representing the α-helices. DNA-binding residues from the β-sheet are indicated by asterisks. c) Superposition of HP0564 (cyan) on HP0222 (magenta). The β-sheet of HP0564 packs more closely to the α-helices than in HP0222, possibly due to its less bulky valine (green) at position 53 compared to the isoleucine (orange) of HP0222.

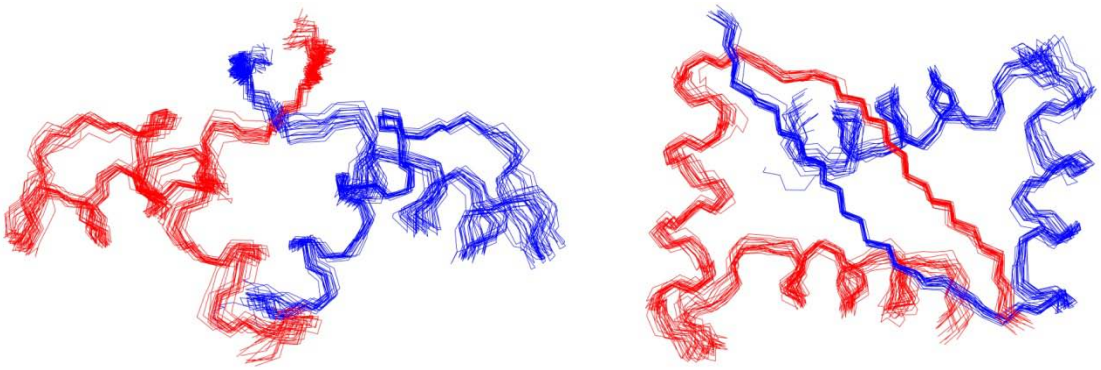

Figure 12. Two views of the ensemble of 20 conformers of HP0564. The two subunits are colored red and blue.

Chemical crosslinking was performed to confirm that HP0564 could form dimers (Figure 13). The amount of dimer and species corresponding to higher-order oligomers increased with increasing BS^3 concentration. Without crosslinking, traces of noncovalent dimers were present on SDS-PAGE gels. In experiments with HP0222 (Popescu, Karpay *et al.* 2005), we did not observe the higher-order, cross-linked forms. Gel filtration experiments showed only stable dimers in solution (Figure 14).

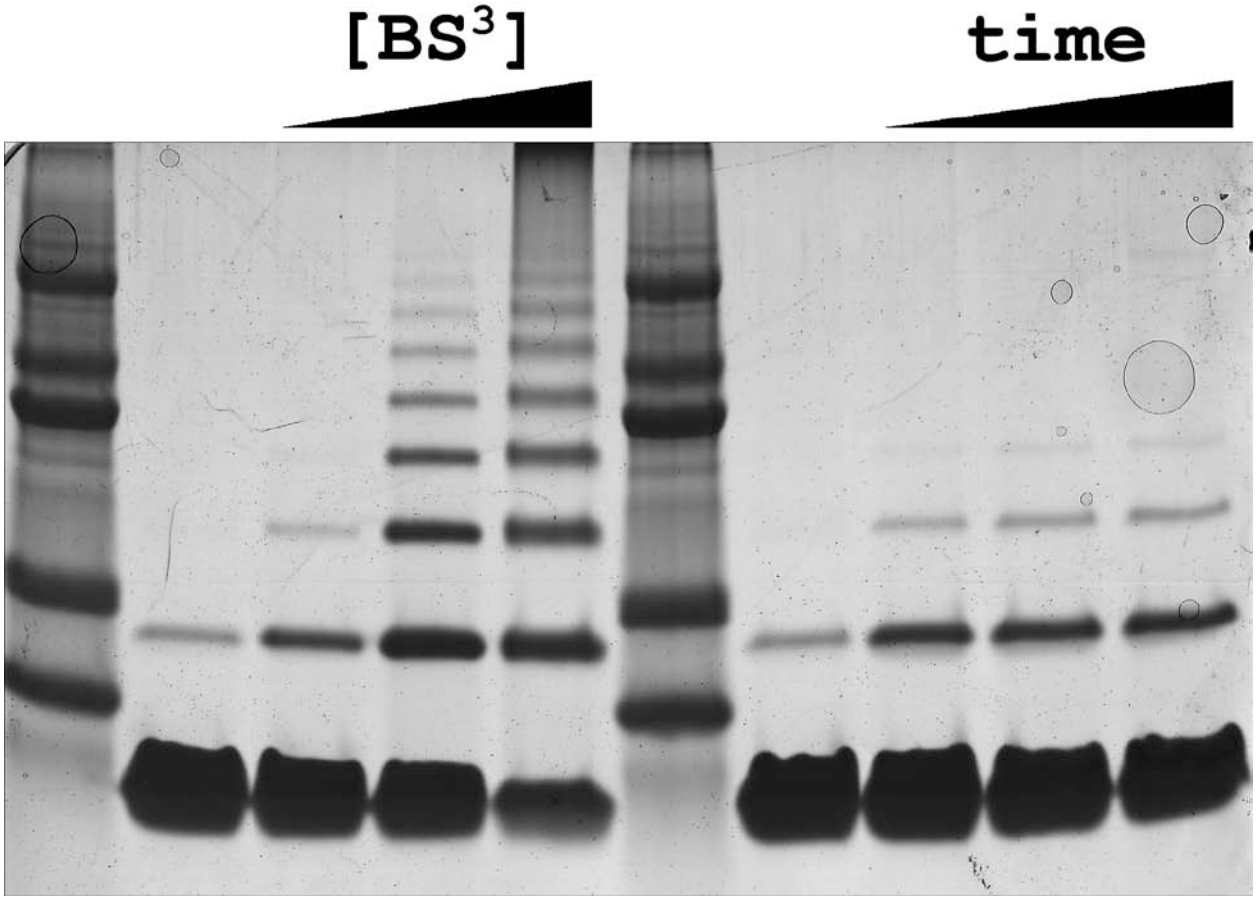

Figure 13. BS^3 crosslinking of HP0564. On the left side of the figure, [BS^3] values are 0, 100 µM, 1 mM, 10 mM, and the reaction was allowed to proceed for 3 minutes. On the right side of the figure, 100 µM BS^3 was used, and the reactions were allowed to proceed for 0, 5, 30, and 60 minutes.

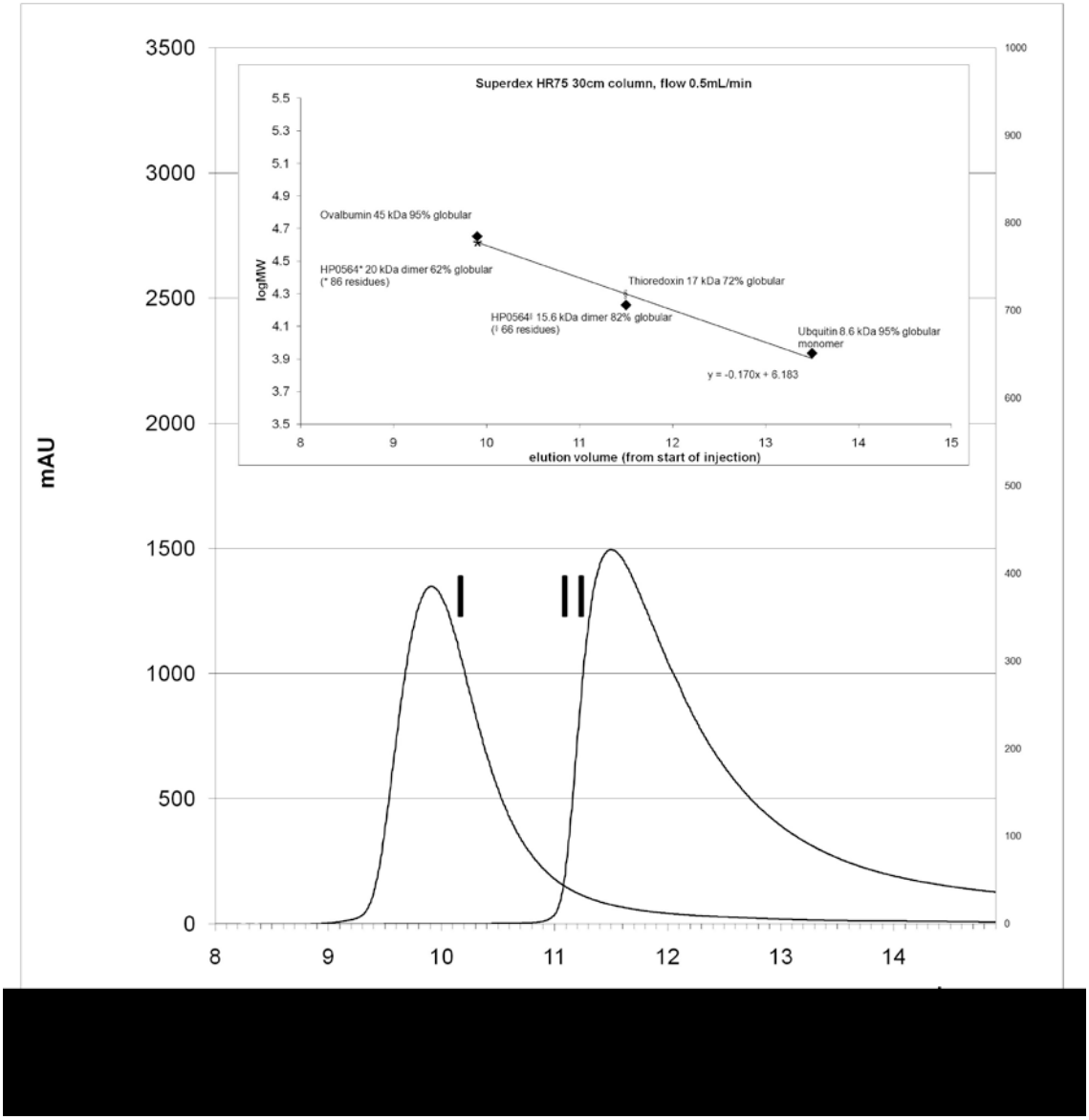

Figure 14. Gel filtration of HP0564. Experiments were performed on a Superdex 75 10/30 column equilibrated with 50 mM potassium phosphate, pH 4.0. Peak I corresponds to full-length HP0564, while peak II corresponds to the shortened construct. The inset at the top of the graph is a calibration curve using ovalbumin, thioredoxin, and ubiquitin.

The DNA binding residues are not conserved between the β-sheets of HP0222 and HP0564. The presence of intact HP0564 does not complement a deletion of HP0222. HP0222 null mutants are viable, but show significantly slower growth than parent wild-type strains. This suggests that they are not functionally redundant and will bind different DNA sequences and regulate different genes. Structurally, the two proteins are very similar, with a backbone RMSD of 1.24 Å (Figure 11c). Superimposing HP0222 and HP0564, one can see that the β-sheet in HP0564 packs more closely to the helices than in HP0222, possibly due to the less bulky valine at position 53 compared to isoleucine in HP0222. Although there are no absolutely conserved amino acids in the RHH family (Schreiter and Drennan 2007), the HP0564 sequence agrees with the sequence motifs featured in all RHH proteins, including the alternating hydrophilic and hydrophobic residues within the β-sheet and the hydrophobic core residues – F24, V26, F28 from the β-sheet, L38 from α-helix 1, and V53, I57, I61 from α-helix 2. All of these residues are involved in making contacts with residues from the other subunit in the dimer.

It is exciting to discover a new transcriptional regulator in *Helicobacter*. Because so few transcriptional regulators have been identified, and because of the conservation of HP0564 in multiple strains of *H. pylori*, we expect HP0564 to play an important role in transcriptional regulation. We are working on determining its cognate DNA-binding sequence and its function in the cell.

CHAPTER IV

FUNCTIONAL CHARACTERIZATION OF HP0222

Introduction

Through structure determination, we identified HP0222 and HP0564 as two new DNA-binding transcriptional regulators in *H. pylori*. These results were exciting and somewhat fortunate because there is no guarantee that determining the structure of an unknown protein will give a clue to its specific function. In our case, the dearth of known transcriptional regulators in *H. pylori* made the discovery all the more interesting.

In order to pursue these projects further, we are attempting to fully characterize the functions of our novel RHH proteins. Protein function can be divided into three general categories – biochemical, biological, and phenotypic. We know that the basic biochemical function of RHH proteins is to bind DNA, and we expect them to act as repressors of transcription, but the next step is to determine the consensus DNA-binding sequences of our proteins. We would also like to determine their biological roles in the cell. Our goal is to determine all genes that are regulated by HP0222 and HP0564 and identify a specific pathway or response in which they play a critical part. In addition, we hope that our results can explain the phenotypic effects that we have observed in our HP0222^{-} and HP0564^{-} null mutant strains. We decided to attack the problem from many angles using several experimental methods.

Functional work on these proteins is an ongoing project that we are working on in collaboration with the laboratory of Tim Cover.

Phenotypic Analysis

Mutants

To determine the functions of our protein, we created mutant strains lacking functional copies of the HP0222 gene. HP0222$^-$ mutant strains were created previously in the J99 and HPK5 strains (Popescu, Karpay *et al.* 2005).

One phenotype that the HP0222$^-$ strains exhibit is a greater adhesiveness than the WT strains. After pelleting cells, it is more difficult to resuspend the mutants. This could indicate a difference in the complement of outer membrane components of the strains.

Another phenotype that we observed was slower growth. When *H. pylori* are spread across a blood-agar plate, they form a shiny, translucent layer within about 24 hours. The HP0222$^-$ strain appeared to take longer to form a fully grown layer.

Growth Kinetics

To further quantify the apparent slower growth of HP0222$^-$, growth curves were obtained from liquid cultures to determine whether the mutation would affect growth rate (Figure 15). Because one of the original acid response transcription profiling studies showed HP0222 to be highly upregulated upon exposure to acid, our initial growth experiments were performed using pH buffered growth media at pH 5 and 7. Cultures of J99 and HPK5 wild-type and HP0222$^-$ strains were inoculated in Brucella broth (Bacto proteose peptone No. 3 10g/L, Bacto tryptone 10g/L, Bacto yeast extract 2g/L (all Bacto products from BD), NaCl 5g/L, glucose 1g/L) medium supplemented with 10% fetal bovine serum (Atlanta Biologicals) at OD_{600} 0.1 for overnight growth at 37°C and 5% CO_2. The following morning, the overnight cultures were used to inoculate 1.5 mL of pH buffered media at approximately OD 0.1 in 24 well plates. pH buffered

media consisted of Brucella broth, 10% FBS, buffered with 50 mM Tris and 50 mM MES and adjusted to either pH 5 or pH 7. Both WT and HP0222^{-} strains were tested at pH 5 and pH 7, and all conditions were tested in triplicate. OD readings were taken at the beginning and at 4 hour intervals for 32 hours.

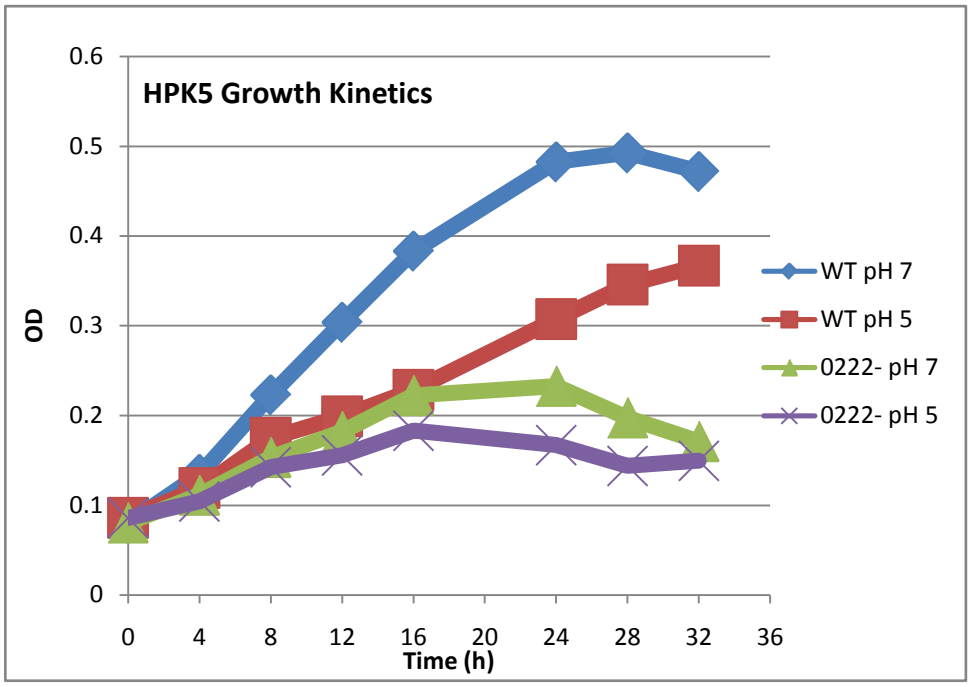

Figure 15. pH-dependent growth curves for *H. pylori* HPK5 wild type (WT) and HP0222^{-} strains. Bacteria were grown in pH-adjusted Brucella broth with 10% FBS.

The presence of buffers in the media slowed the growth of *H. pylori* compared to unbuffered media even at pH 7; however, we found it to be necessary due to the ability of *H. pylori* to rapidly raise the pH of the unbuffered acidic media to neutral. Although we expected to see an effect on growth at acidic pH, instead we found an effect primarily at neutral pH, where HP0222^{-} growth was significantly retarded compared to wild type. At acidic pH, neither the wild-type nor the mutant strain grew very well, and the difference in growth was insignificant. In addition to

the slower growth, HP0222⁻ strains at either pH reached their saturation point and began to decline at a much lower OD than the wildtype strains. Results from J99 and HPK5 were very similar.

Growth kinetics experiments were repeated many times, often to check that our strains were fully motile (see discussion). After finding that there was not a significant difference in growth rate at low pH, we simplified the composition of our growth media by removing the buffering compounds, allowing the bacteria to grow significantly faster. We also found that filter sterilizing the media instead of heat sterilizing it improved the growth rates of the bacteria. Figure 16 shows growth curves for J99 WT and HP0222⁻ strains.

At later times in the growth curves, the measured OD is often much lower than the previous measurement. This appears to be a feature of *H. pylori* growth. When they have reached saturation, they tend to die and rupture more easily than other bacteria such as *E. coli*. We observe this decline in all growth experiments.

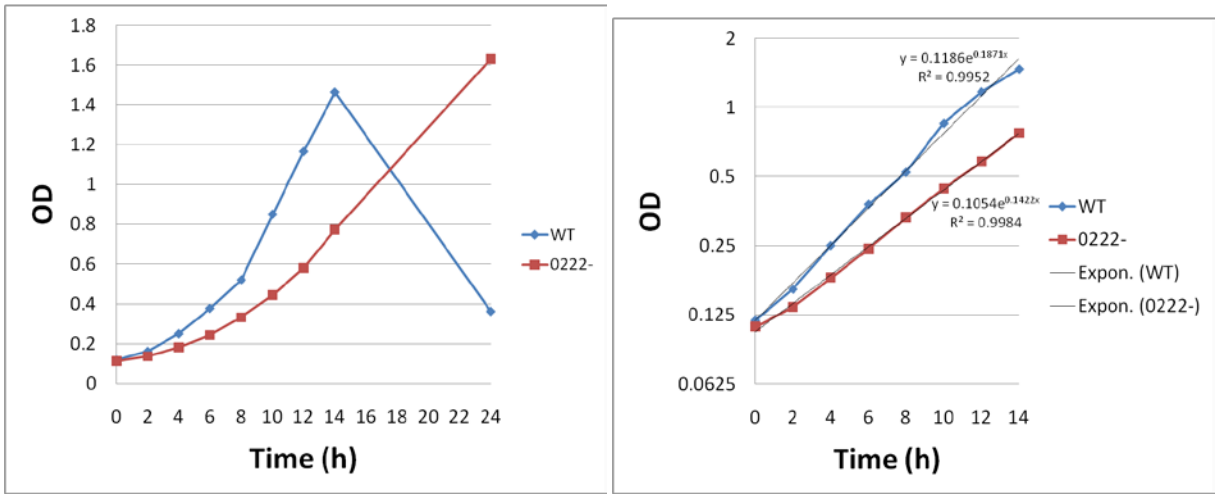

Figure 16. Growth curves of J99 WT and HP0222⁻ strains. The plot on the left shows the measured values. The plot on the right is on a logarithmic scale between 0 and 14 hours. Trendlines are shown, and their equations were used to extract doubling times for each strain.

Trendline equations were converted to base 2 for extraction of doubling times, which were taken as the inverse of the coefficient in front of x in the equations.

$$\text{WT}: y = 0.1186e^{0.1871x} = 0.1186 * 2^{0.2699x} \rightarrow \text{doubling time} = 3.7\ \text{h}$$

$$\text{HP0222}^{-}: y = 0.1054e^{0.1422x} = 0.1054 * 2^{0.2052x} \rightarrow \text{doubling time} = 4.9\ \text{h}$$

HP0222^{-} grows significantly slower than WT, with a doubling time of only 4.9 hours, versus 3.7 hours for WT.

pH-Dependent Survival Rate Experiments

Growth kinetics experiments provide information about the aggregate growth rates of strains, but they do not tell us about the fitness of individual cells. To determine whether HP0222^{-} cells simply grow slower or die more often, we performed survival rate experiments at various pH values. J99 wildtype and HP0222^{-} strains were grown for 24 hours in buffered Brucella broth at pH 5,6 and 7. A small aliquot was taken from each culture of J99 wildtype and HP0222^{-} at pH 5,6 and 7 and diluted 10,000x. 50 µL of the dilution was plated onto blood-agar plates and incubated at 37°C for 2 days (Figure 17). Colonies were counted manually, and the number of colony-forming units (CFU) per mL was determined by multiplying the number of colonies counted by 200,000 (Figure 17). Survival rates were determined by dividing the CFU by the total number of bacteria in the culture, as estimated by the final OD measurement (Table 6).

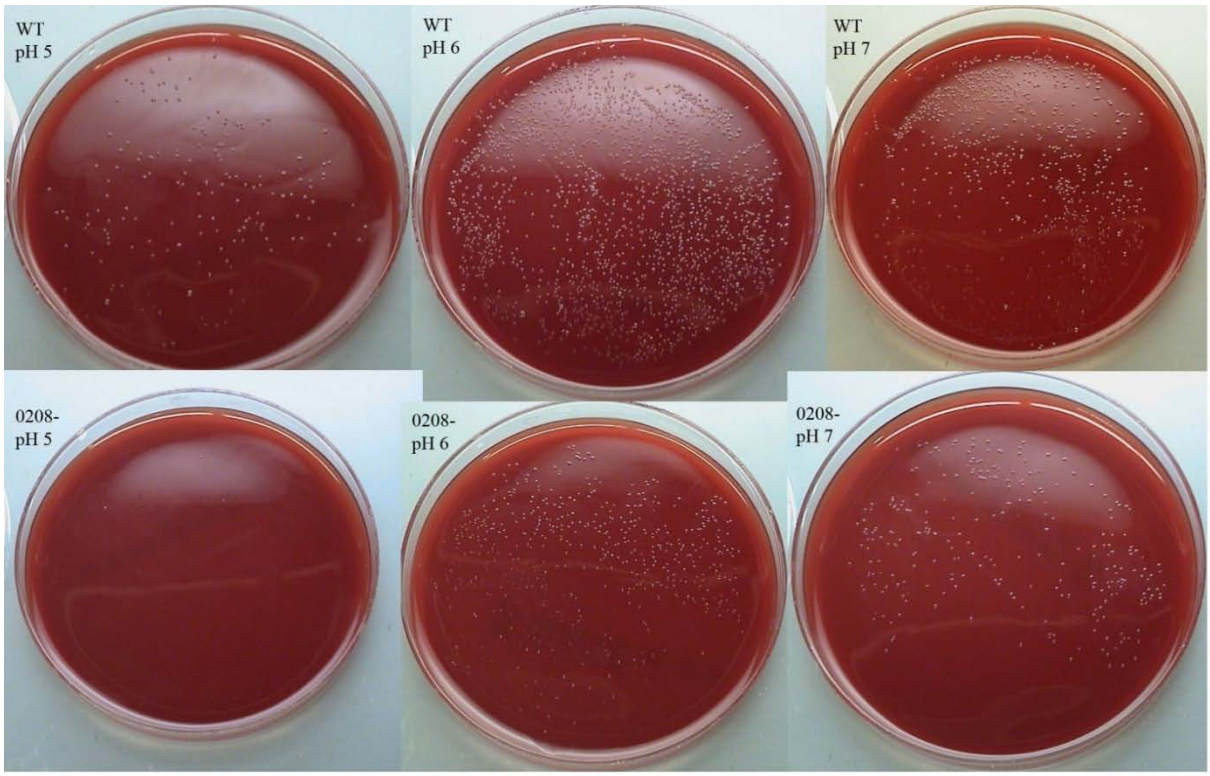

Figure 17. pH survival rate experiments. WT and HP0222⁻ strains were grown in Brucella broth adjusted to pH 5, 6, or 7. Cultures were diluted and plated onto blood-agar plates, and colony-forming units were counted.

Table 6. Survival rates for J99 WT and HP0222⁻ strains at pH 5,6,7.

WT	**CFU**	**Final OD**	**Survival Rate**	**Average (± spread)**
pH 5 #1	222	0.29	15.3%	12.8 ± 2.6%
#2	162	0.32	10.2%	
pH 6 #1	1652	1.17	28.2%	29.8 ± 1.6%
#2	1841	1.18	31.3%	
pH 7 #1	993	1.16	17.1%	16 ± 1.1%
#2	872	1.17	14.9%	
HP0222 mutant				
pH 5 #1	4	0.27	2.94%	4 ± 1.1%
#2	7	0.28	5.05%	
pH 6 #1	660	0.69	19.1%	16.5 ± 2.6%
#2	485	0.70	13.9%	
pH 7 #1	340	0.54	12.7%	11.5 ± 1.3%
#2	271	0.53	10.2%	

After finding that the HP0222^{-} mutation slowed growth at neutral and acidic pH, we wanted to determine whether the mutant strain was simply growing slower, or also dying at a faster rate. We performed experiments to measure the number of viable, colony-forming units (CFU) in order to calculate survival rates. Table I shows the number of CFU from each plate, the final OD measurement of the culture, and the survival rates for each strain and pH. For both strains, growth rates and survival rates were optimal at pH 6, with a sharp decline in growth-rate, final OD measurement, and survival rate at pH 5. Between the two strains, wild type always showed a higher survival rate. Viable CFU of HP0222^{-} were drastically decreased at pH 5.

Motility Assays

H. pylori strains were grown on blood-agar plates for 24 hours and transferred to Brucella agar (0.5%), 10% FBS plates by covering the tip of a disposable plastic needle in a lawn of cells from the blood-agar plates and quickly stabbing it vertically down to the bottom of the Brucella agar plate. Plates were incubated at 37 °C and 5% CO_2 for 7 days. Each plate was inoculated with wildtype and HP0222^{-}. Photographs of plates were taken after 7 days (Figure 18).

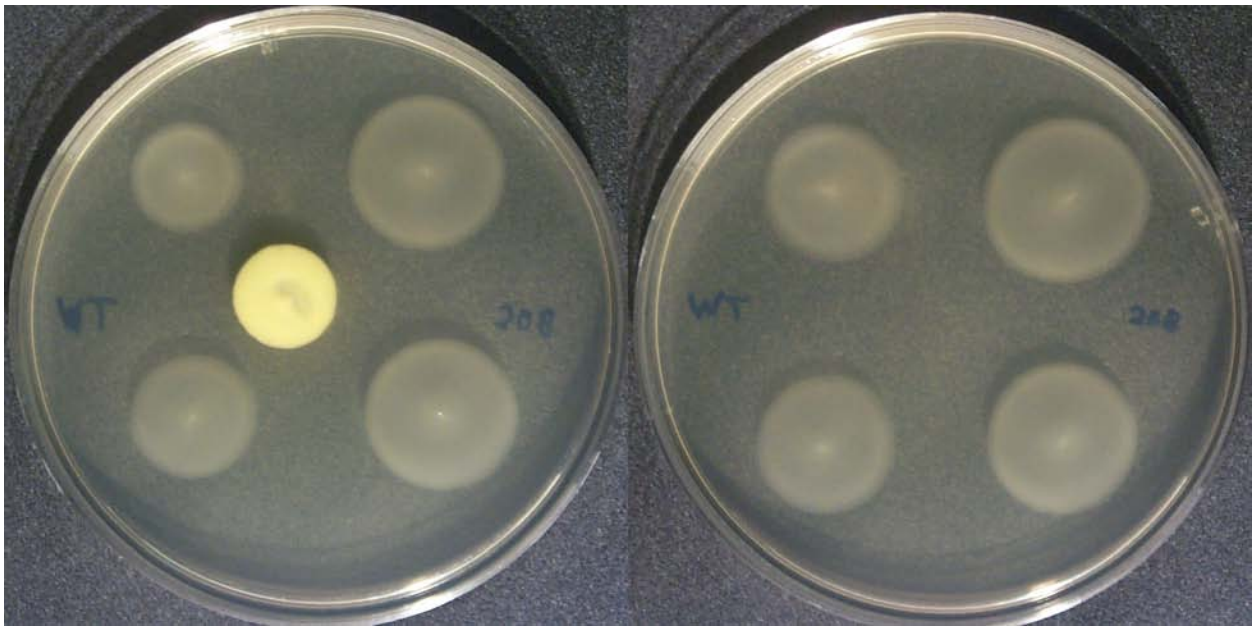

Figure 18. Motility assays of *H. pylori* J99 WT and HP0222^{-} strains on semi-solid agar (0.5%) plates. Each plate shows two WT spots on the left and two HP0222^{-} spots on the right. Photographs were taken after 7 days.

All four HP0222^{-} spots have a larger diameter than the WT spots, while the densities are roughly equal. HP0222^{-} appears to be hypermotile compared to WT. This result makes sense in light of the growth curve results. Motility is known to use large amounts of energy, so hypermotile strains might be expected to grow slower than WT strains.

Biological Analysis

Microarrays

Because of the difficulty of the SELEX experiments, we had no hints which genes could be directly regulated by HP0222. We decided to perform microarray experiments to compare transcript levels of all *H. pylori* genes between WT and HP0222^{-} strains.

H. pylori J99 WT and HP0222^{-} cultures were inoculated in the morning in 10 mL Brucella broth supplemented with 10% FBS to OD 0.1 from overnight cultures. After 8 hours, WT cultures had grown to OD approximately 0.4, while mutant strains grew to OD approximately 0.35. 2 mL from each culture were spun down gently, and 1.5 mL of culture medium were removed. Cell pellets were resuspended in the remaining 500 µL of culture medium. 1 mL Bacterial Protect Reagent (Qiagen) was added, and the mixture was vortexed. After centrifugation, cell pellets were lysed by resuspending in 1 mL Trizol Reagent (Invitrogen) heated to 65 °C. RNA was precipitated with isopropanol and washed with 70% ethanol. DNA contamination was removed by redissolving precipitated nucleic acid in 90 µL water and adding 10 µL RQ1 RNase-free DNase (Promega) and incubating at 37 °C for 1 hour. Purified RNA was cleaned up using the RNeasy Mini Kit (Qiagen).

H. pylori microarrays containing sequences from all genes in the 26695 strains as well as unique sequences from the J99 strain were obtained from the Pathogen Functional Genomics Resource Center of the J. Craig Venter Institute by Dr. John Loh. Quality control analysis of RNA samples and labeled cDNA synthesis, as well as hybridization and reading of the arrays were performed by the staff in the Vanderbilt Microarray Shared Resource facility.

Large datasets are required to obtain statistically significant results from microarray experiments because of the various sources of error. Each gene is spotted on the microarray three times in different regions to control for variations in spotting and hybridization to different regions of the array. To control for random variations in transcript levels, three cultures each of WT and HP0222^{-} were grown. Data from three microarrays, each hybridized with labeled cDNA from one WT and one HP0222^{-}, were obtained, and all results were averaged.

Because most RHH proteins act as repressors, we graphed the data as HP0222⁻/WT ratios and looked primarily for increases in the levels of transcription (Figure 19). The left side of the graph, up to about gene 1600 represents sequences derived from the 26695 strain, while the right side of the graph represents sequences from the J99 strain. There is clearly a difference in the mean and standard deviation of the ratios from sequences derived from the two strains, which may be due to the fact that the 26695 sequences likely contain several mismatches when binding cDNA from our J99 samples.

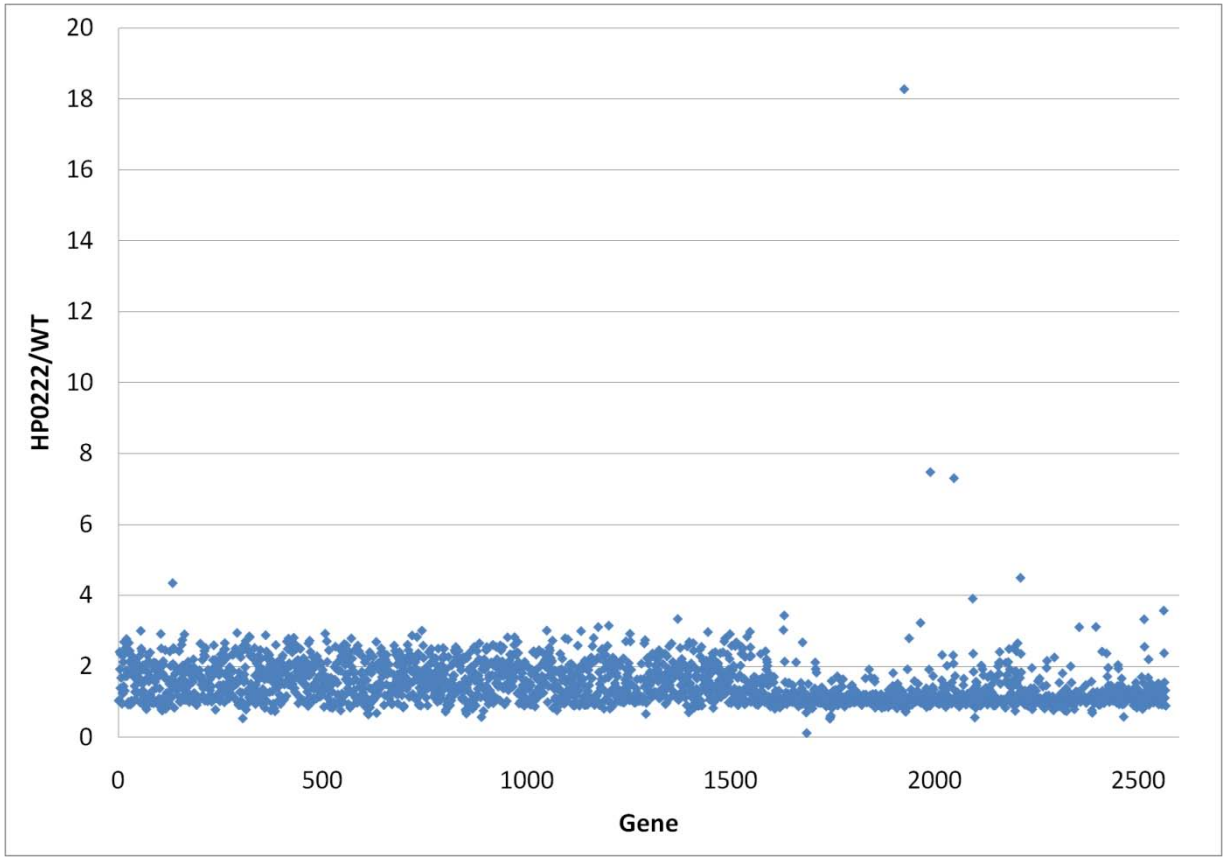

Figure 19. Graph of HP0222⁻ to WT transcript ratios for all 26695 and some J99 genes, averaged over three microarrays. For 26695 genes on the array, the average transcript ratio was 1.63, with a standard deviation of 0.53. For the J99 genes, the average ratio was 1.22, with a standard deviation of 0.71.

An additional set of three arrays was obtained using new isolates of both WT and HP0222^{-} strains. Data from all six arrays were averaged. Genes were accepted as upregulated in the mutant strain if their HP0222^{-} to WT transcript ratios were at least two standard deviations above the mean, and if they were consistently upregulated in at least five of the six samples (Table 7).

Table 7. Microarray results for HP0222^{-}

Gene (J99 #)	**Gene (26695 #)**	**Average (HP0222^{-}/WT)**	**Gene Product/Function**
JHP0572	HP0629	20.5	unknown
JHP0662	HP0725	4.85	HopP – outer membrane protein
JHP0954	NA	3.35	unknown
JHP1297	NA	2.14	Type III restriction enzyme
JHP0742	HP0806	2.10	Unknown, conserved

Real-Time PCR

Real-time PCR is often used to quantify levels of transcripts and validate microarray data. RNA was obtained as described for the microarray experiments. cDNA synthesis was performed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Forward and reverse primers for the coding region of each gene were designed using Primer3 (Rozen and Skaletsky 2000) with the following input parameters: 20bp length, 100-250bp product length, primer T_m 60 °C. For each gene, a master mix was produced containing Maxima SYBR-Green qPCR Master Mix (Fermentas) and primers (final concentration of 2 uM each). 23 µL of this master mix was pipetted into each well, followed by 2 µL of a 200x dilution of cDNA template, for a total reaction volume of 25 µL. Reactions were run on a BioRad iCycler running MyIQ

software. The hot-start enzyme was activated by a 10 minute interval at 95 °C, followed by 45 cycles of 95 °C 20 s, 55 °C 30 s, 72 °C 30 s. Upon completion of the PCR reaction, a melt curve was acquired by raising the temperature 0.5 °C every 10 s from 55 °C to 95 °C.

Several genes were tested by real-time PCR, including JHP0572 and JHP0662, as well as many genes involved in motility. Results confirmed that JHP0662, the outer membrane protein *hopP*, is upregulated in our mutant. JHP0572 was upregulated only slightly. Many genes involved in motility, including the major flagellin *flaA*, JHP0424, JHP1051, JHP1047, and JHP1048 were also slightly upregulated.

Biochemical Analysis

SELEX

Our initial strategy for determining the function of HP0222 in *H. pylori* was to use the SELEX (Tuerk and Gold 1990) (Systematic Evolution of Ligands by Exponential Enrichment) technique to determine the consensus DNA-binding sequence for the protein. As the name suggests, SELEX is a technique whereby one starts with a pool of a large number of possible ligands and progressively selects for only the strongest binding ligands through a series of selection and amplification steps. Traditionally, the ligands are nucleic acids (RNA or DNA), and the targets are proteins; however, the protein targets do not always have nucleic acid binding functions in the cell. In fact, many very tightly binding aptamers (usually single-stranded nucleic acids) have been found for non-nucleic acid binding proteins (Djordjevic 2007). Figure 20 shows the general flow of a SELEX experiment. At the beginning of the experiment, one starts with a large pool of nucleic acid sequences. This pool is bound to the target protein, and the mixture is put through a separation procedure that theoretically allows one to retain only nucleic

acids that bind tightly to the protein. Many separation techniques have been developed for SELEX, including nitrocellulose membrane filtration, affinity tags, column matrices, cross-linking, gel electrophoresis, capillary electrophoresis, and centrifugation (Gopinath 2007). In practice, because the tightly binding nucleic acids are far outnumbered by weakly or moderately tight binding nucleic acids, it is nearly impossible to obtain only the most strongly binding sequences in one step. After the first separation procedure, the bound nucleic acids are removed and used as templates in a PCR reaction, which amplifies both the tight-binding nucleic acids and any others that passed through the separation procedure. The resulting products are the pool that will be used in the second round of SELEX. If the round was successful, then the tight-binding sequences will be enriched relative to weakly-binding sequences. The SELEX procedure ends when one has a number of very similar sequences that can be aligned to determine the consensus binding sequence. Typically, a SELEX experiment will require 12-15 rounds of binding, separation, and amplification.

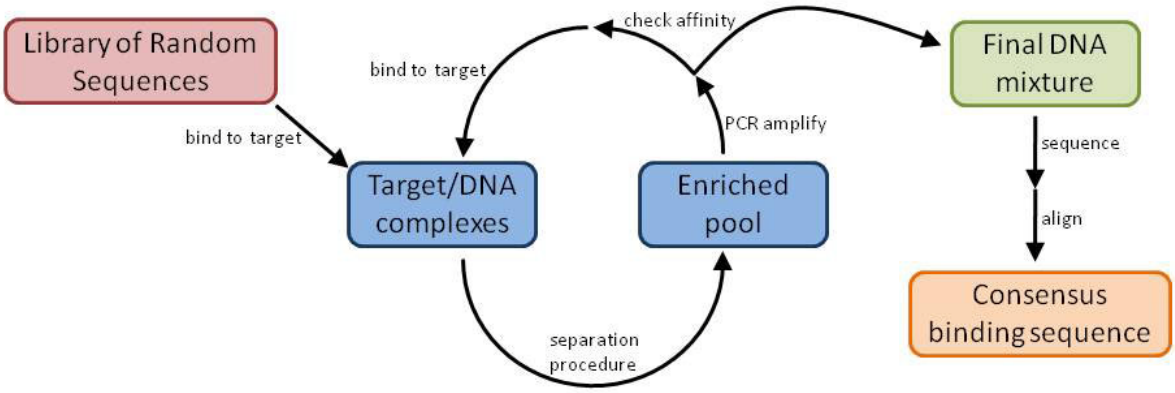

Figure 20. General SELEX procedure. Starting with a large pool of sequences, an iterative procedure of binding, separation, and enrichment is carried out until the bulk affinity of the current pool of sequences for the target is sufficiently high. Sequencing and alignment hopefully yields a consensus sequence.

Because we were attempting to determine the consensus DNA-binding sequence for a known transcriptional regulator with natural DNA-binding functions, we began by performing genomic SELEX experiments. The initial pool of DNA in genomic SELEX is obtained by purifying genomic DNA from an organism and cutting it up into smaller fragments, usually using restriction enzymes. The advantage of this type of SELEX is that the number of different sequences in the initial pool is many orders of magnitude smaller than in the typical random sequence pool. It is also almost guaranteed to contain at least one good binding sequence for a target protein whose function is to bind DNA, assuming that the fragmentation procedure does not disrupt that sequence. One disadvantage is that it is more difficult to amplify the selected sequences after the first round.

For genomic SELEX, we followed a published procedure (Dietz, Gerlach *et al.* 2002). His_6-tagged HP0222 was bound to MagneHis Ni-Particles (Promega) and incubated with J99 genomic DNA that had been completely digested with the ApoI restriction enzyme (Figure 21A). Because HP0222 binds nonspecifically to too many fragments in interaction buffer (IB) (20 mM Tris-HCl, pH 7.5, 10 mM MgCl2, 100 mM KCl, 20 mM imidazole), we selected strongly binding fragments using a series of washes with increasing salt concentration. The wash steps consisted of IB supplemented with NaCl (375mM, 563mM, 857mM, 1.5M) (Figure 21B). Eluted fragments were ligated on both ends to a double-stranded sticky end adapter. The adapted fragments were PCR amplified using one primer corresponding to the adapter sequence as both the forward and reverse primer, and the resulting products were used in a second capture experiment (Figure 21C).

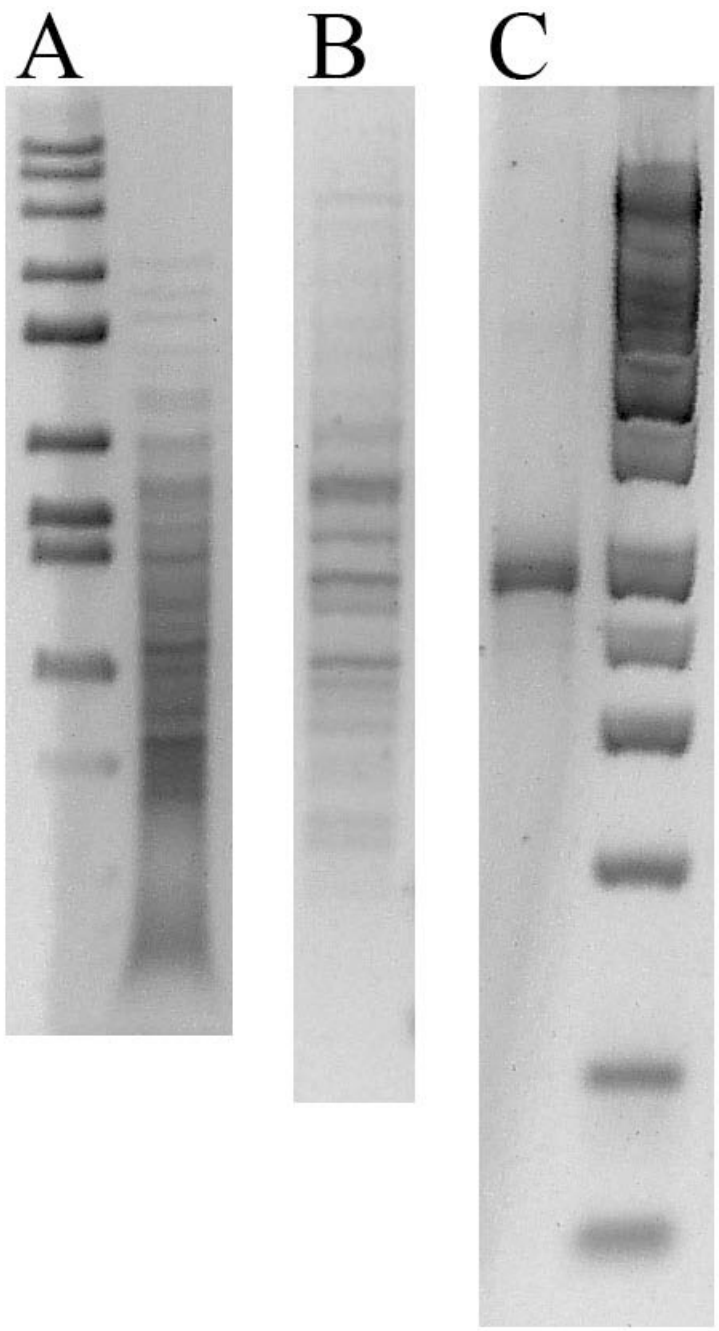

Figure 21. Genomic SELEX results, A) left lane – molecular weight standards; right lane – successful ApoI digest of J99 genomic DNA, with individual bands visible, B) DNA washed off from 1.5 M NaCl wash, C) left lane – single strong band after amplification of DNA in B) and a second round of selection; right lane – molecular weight standards. DNA fragments were separated and stained in 2% agarose gels, 0.5x TBE, pH 8.5, with GelRed stain (Biotium).

Most of our attempts at SELEX utilized a random sequence pool of DNA. This method is completely unbiased because every possible sequence of a specified length is present in the initial pool of DNA; however, the sheer number of different sequences and the extremely small number of copies of the most tightly binding sequences in the initial pool make it more difficult to detect and recover those sequences. Unlike genomic SELEX, it is easy to design the oligos such that they can be amplified after the first separation procedure without an additional adapter ligation step. Table 8 shows all of the template oligos used in our SELEX experiments.

Table 8. Template oligos used for SELEX experiments

Oligo	**Sequence**
SELEX1	CGCTGACTGACTGAGCCGCCG**NNNNNNNNNNNNNNN**GGCTCAGCTCACCTCAGCCCG
SELEX2	(reverse complement of SELEX1)
SELEX3	CGATGACTGACTGACCTGTGC**NNNNNNNNNNNNNNN**GGTTCAGGTCAAGTCAGCACG
SELEX4	CGATGACTCACTGACCTGCTCTACACG**NNNNNNNNNNNNNNNNNNN**GGAACGAATGCCTTGTCTACTGAGTGC
SELEX5	CCCAAGCTTAATACGACGCACTATAGGGAGGAT**NNNNNNNNNNNNNNNNNNN**TTGCAGCATCGTGAACTAGGATCCGGG
SELEX6	CCCAAGCTTAATACGACGCACTATAGGGAGCTA**NNNNNNNNNNNNNNNNNNN**CGTGTAGAGCAGGTCAGTGAGTCAGGATCCACG
SELEX7	CGCAAGCTTCATACGACGCACTCATGGGAGCTAAGCACTAACTGCC**NNNNNNNN**CGACCATGCTGAACGTGTAGAGCAGGTCAGTGAGTCCGGATCCCGC
SELEX8	CGCAAGCTTCATACGACGCACTCATGGGAGCTAAGCACTTTTAAAAN**NNNNNNNNNNNN**CATGCTGAACGTGTAGAGCAGGTCAGTGAGTCCGGATCCCG
SELEX9	CGCAAGCTTCATACGACGCACTCATGGGAGCTAAGCACTA**NNNNNNNNNNNNNNNNNNN**TGCTGAACGTGTAGAGCAGGTCAGTGAGTCCGGATCCCGC

Although conceptually SELEX appears to be rather straightforward, in practice it can be very challenging. The technique is simply more suitable for some proteins than for others. The SELEX procedure has been thoroughly analyzed mathematically and a number of potential experimental pitfalls have been identified (Irvine, Tuerk *et al.* 1991; Levine and Nilsen-Hamilton

2007). One practical obstacle is that the protein concentration used must be optimized to differentiate its specific and nonspecific interactions with DNA. Without knowing these affinities beforehand, the initial protein concentration must be optimized; however, even this is difficult because of the necessity to detect and isolate complexes at very low concentration. In successive rounds, the protein concentration must be lowered to obtain better selection of only the strongest binding sequences. If specific interactions are not very strong ($K_d \sim 10^{-9}$), it is possible, however, to ruin an experiment by losing the enriched tight-binding sequences if the protein concentration is not high enough.

Detection of bound nucleic acids is a significant problem, especially during the early rounds of SELEX. When using a large pool of sequences, there are only a few molecules that will bind the target protein with high affinity. If one is using a gel-based selection procedure, it is often difficult to detect the position of the shifted band, even when using radioactively labeled DNA. Many successful SELEX experiments rely on the existence of a known reference binding sequence, which is run side-by-side on the gel next to the protein-DNA mixture. The position of the shifted reference sequence indicates where to cut a fragment from the lane containing the SELEX DNA. In our case, because we do not have a known binding sequence, we tried a brute force method of finding the shifted band. The entire lane above the band of unshifted DNA was cut out and sliced into eight pieces. DNA was extracted separately from all slices and used as template in PCR reactions to determine which slice contained the most DNA. Unfortunately, no difference in the amount of DNA from each slice could be determined (Figure 22C). It appears that DNA is randomly captured in the gel.

As SELEX proceeds and the pool of DNA is narrowed down to a number of similar sequences, problems with multitemplate PCR begin to appear. Typical problems that occur

include bias, in which the ratios of concentrations of final products do not accurately reflect the ratios of the starting templates, heteroduplex formation, which occurs when one template strand anneals to the complementary strand of a different template strand, and chimeric products, which are sequences that are generated from more than one template (Kanagawa 2003). Although we cannot determine how significantly the problem of PCR bias has affected our results, heteroduplex formation has certainly been a problem. In many applications, heteroduplexes are a problem because when cloned, cellular enzymes repair the mismatch by randomly selecting a template strand, which can artificially increase sequence diversity (Speksnijder, Kowalchuk *et al.* 2001). In a SELEX experiment, the heteroduplexes are also a problem because the mismatches are not resolved through cloning and intracellular repair between selection rounds, and the bulges created by the mismatches can render the heteroduplexes unsuitable for binding to the target protein. Mismatches appear on gels as a slightly shifted band (Figure 22A). Using capillary electrophoresis to analyze PCR products, it was shown that in PCR reactions containing heterogeneous templates, formation of the desired products ceases before exhaustion of primers and that these products are converted into by-products within just a few cycles (Musheev and Krylov 2006). A method of removing heteroduplexes called reconditioning PCR has been developed. In this procedure, one dilutes the heteroduplex products ten-fold, adding fresh primers and subjects the mixture to three cycles of PCR (Thompson, Marcelino *et al.* 2002). We have used this method with limited success. It is best to simply reduce the number of PCR cycles in each reaction to limit the formation of undesired by-products in the first place (Kanagawa 2003); however, determining the optimum number of PCR cycles can be tricky because one must balance the need to create enough of the desired product for the next round

with limiting by-product formation, and the optimum number of cycles may change from round to round.

A SELEX experiment can be performed more efficiently with fewer rounds of separation and amplification if there is a greater disparity between the target protein's affinities for specifically and nonspecifically bound sequences. At least two orders of magnitude is desirable. Proteins with unusually high affinities for nonspecific sequences will not be able to significantly enrich the specific aptamer from round to round, requiring a greater number of rounds before completion. Performing more rounds of separation and amplification increases the risk of failure due to other problems mentioned above.

Many DNA-binding proteins feature a floppy, positively-charged N-terminal or C-terminal segment that interacts with the phosphate backbone of DNA and increases the affinity of the protein for DNA without increasing its specificity (Crane-Robinson, Dragan *et al.* 2006). Among the well-characterized RHH proteins, HP0222 and its *H. pylori* homolog HP0564 are two of the four that have N-terminal tails of greater than 20 amino acids. While determining the appropriate concentration of HP0222 for the binding reactions, we found that HP0222 has a high affinity for nonspecific DNA, being able to shift the entire band of random sequences at a concentration of 1 μM (Figure 22B). We also tried SELEX experiments with a mutant version of HP0222 lacking the N-terminal tail.

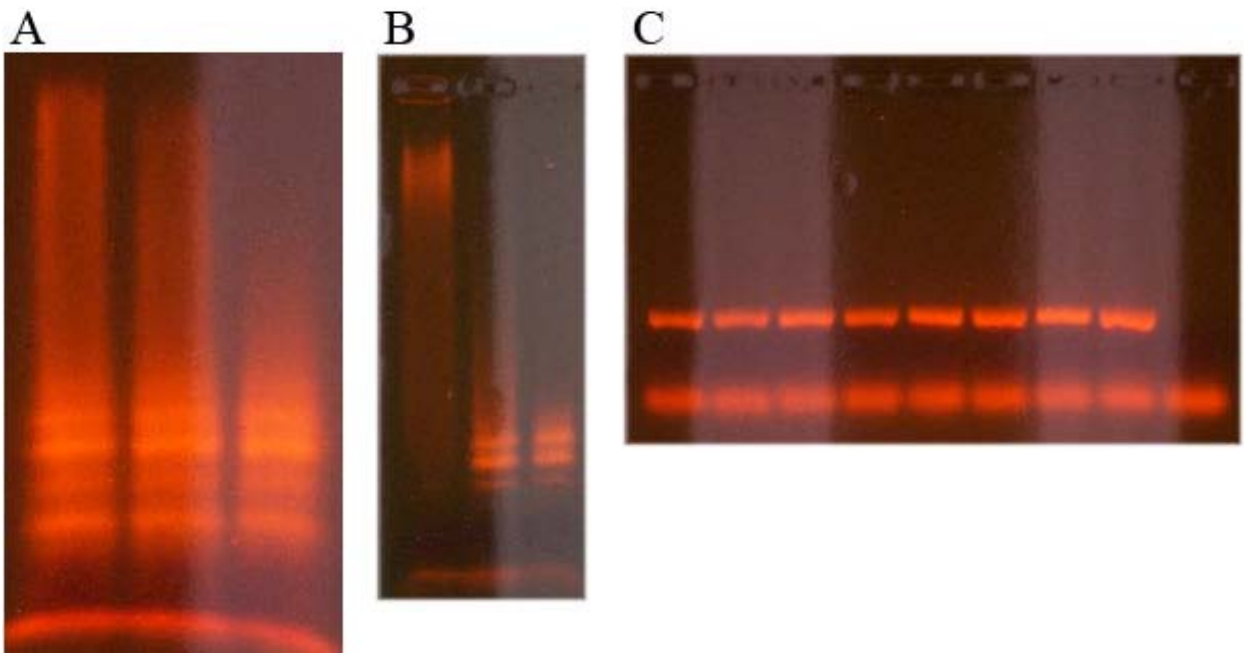

Figure 22. Agarose gels showing several pitfalls of the SELEX procedure. A) Ladder of bands represent a series of double-stranded DNA species with increasing numbers of mismatched bases resulting from similar but nonidentical template sequences and depletion of PCR primers. B) Left lane shows the entire pool of sequences shifted by 1 μM HP0222 protein during an early round of SELEX, resulting in no selection or enrichment of the strongest binding sequences. C) Each lane shows PCR products from amplification of DNA contained in one of eight equal-sized segments cut out of a lane from the previous separation. This experiment shows that DNA is randomly captured in the agarose gel matrix, reducing the efficiency of selection.

Our final attempts at SELEX utilized a different separation technique – capillary electrophoresis (CE). Using the nonequilibrium capillary electrophoresis of equilibrium mixture (NECEEM) method (Berezovski, Drabovich *et al.* 2005), CE-based SELEX methods offer several advantages over other methods. Separation is highly efficient, and protein-bound DNA passes through the capillary before unbound DNA, minimizing the amount of unbound DNA that passes through to the next round. Detection sensitivity is much greater than staining a gel, even when using UV detection. Laser-induced fluorescence detection can be used to detect much smaller quantities of bound DNA, and in one case, the entire SELEX procedure could be

completed in only one round of selection (Berezovski, Drabovich *et al.* 2005). In fact, SELEX experiments have been performed using a procedure called non-SELEX, in which the amplification steps between rounds are eliminated (Berezovski, Musheev *et al.* 2006). Despite the apparent advantages of CE-based SELEX methods, we were unable to obtain any results on our protein.

Discussion

Motility Variants

One interesting observation we made while working with cultures on plates and in liquid media is that when left on a plate for more than 24 hours or in liquid media for more than about 12 hours, both wild type and HP0222^{-} strains became nonmotile. We first noticed this in liquid culture when the doubling times of both strains steadily decreased from about 4 hours for wild type and 5 hours for HP0222^{-} to about 1.9 hours for both strains. We ran motility assays on these long term cultured strains and confirmed that they are nonmotile. It appears that when nutrients are depleted, there is a strong selection for nonmotile variants that do not use as much energy. The selection for nonmotile variants appears to be especially strong for the hypermotile, slower growing HP0222^{-} cells. It is interesting to note that the nonmotile HP0222^{-} strain grows just as fast as nonmotile WT, suggesting that the slow-growth phenotype of HP0222^{-} is primarily due to the hypermotility.

We do not know exactly how the strains become nonmotile. The fact that the doubling times in liquid culture gradually decrease, however, suggests that there is a selection for a preexisting population of nonmotile cells. There are reports describing such subpopulations. *fliP*, which encodes for a protein involved in flagellar export is subject to slipped-strand

mispairing transcriptional regulation. WT strains have a stretch of 8 cytidines in the promoter region. Variants with an additional cytidine are completely nonmotile (Josenhans, Eaton *et al.* 2000). At higher bacterial densities, when energy resources have been depleted, nonmotile variants may have higher survival rates, allowing them to dominate the population. From this experience, we learned that it is important to regularly check growth rates and motility. We were able to recover motile variants again by inoculating the strains into soft-agar motility plates and extracting motile bacteria from the edge of the colony after several days. All experiments since then have been done using bacteria that have been taken from our permanent motile stocks and passed on no more than two or three times.

Adhesion to gastric epithelial cells has been found to upregulate HP0222, with a concomitant downregulation of several motility genes. Conditions have not yet been found where HP0222 is expressed at lower levels than WT, except for our mutants. There are several reports of hypermotile mutants in other bacteria that have a greater ability to colonize their hosts (Jones, Marston *et al.* 2004; Martinez-Granero, Rivilla *et al.* 2006; Haugen, Pellett *et al.* 2007). In particular, one study found that mutations that restored motility in a poorly motile strain of *Campylobacter jejuni*, a bacterium related to *H. pylori*, also restored its colonization ability (Jones, Marston *et al.* 2004). It is not always the case, however, that hypermotility is linked to greater colonization ability. One study found that nonmotile mutants of a strain of *Escherichia coli* grew faster and were better able to colonize the mouse intestine (Gauger, Leatham *et al.* 2007). It appears that in some cases hypermotility may be able to offset the slower growth to allow better colonization ability, but not always. This makes it difficult to predict the colonization ability of our HP0222^{-} strain, and it is something we would like to test in the future.

Microarrays

Whenever deleting a single gene from the genome, it is possible that polar effects on nearby genes may play a role in any observable phenotypes. Although noncoding sequences in the genome between HP0222 and its flanking genes, HP0221 and HP0223, suggest that they are not part of the same operon, this remains a possibility. Transcript levels of these two genes were checked in our array data and also by real-time PCR. Neither gene is significantly affected by the deletion of the functional copy of HP0222.

Results from the growth kinetics experiments were crucial for designing the microarray experiments. Because of the large number of changes in transcriptional levels of many proteins upon the switch to stationary phase growth, it was important to ensure that RNA was extracted during log phase growth. Overnight cultures would not be acceptable. Our procedure involved inoculating new cultures to OD 0.1 from overnight cultures and allowing them to grow for eight hours to approximately OD 0.4, well within log-phase growth according to previous experiments.

CHAPTER V

FUNCTIONAL CHARACTERIZATION OF HP0564

Introduction

Based on the structural similarity to RHH proteins, HP0564 is another *H. pylori* protein that is likely to play a role in transcriptional regulation. Its β-sheet residues are not the same as those of HP0222; therefore, we expect it to bind different sequences and regulate different genes. In the case of HP0222, two previous microarray-based transcriptional studies found upregulated transcription in different conditions. There is no literature mentioning HP0564. We used many of the same experimental methods to determine the function of HP0564 as we did for HP0222.

Phenotypic Analysis

Mutants

Creating mutant strains in *H. pylori* is not as straightforward as it is in *E. coli*. Many of the tools for doing molecular biology in *H. pylori* are not nearly as developed. Although *H. pylori* are naturally competent, they also have a large number of restriction-modification (RM) systems, and each strain has its own set of functional RM systems. Although some endogenous *H. pylori* plasmids have been found, our experience with them suggests that they are not stable in every strain.

The HP0564^{-} strain that we used for the following experiments was produced by Dr. John Loh. The JHP0511 coding sequence from J99 plus regions flanking upstream and downstream were PCR amplified and cloned into pGEM-T Easy vector. Inverse PCR was then

used to generate deletions within the JHP0511 ORF while introducing BglII restriction sites. The inverse PCR product was digested with BglII and recircularized after a ligation step using T4 DNA ligase, and stored. To insert an antibiotic cassette within the coding sequence of JHP0511, the re-circularized plasmid was digested with BglII and the overhangs were filled by DNA polymerase I large (Klenow) fragment. A blunt ended HindIII fragment containing the cat (chloramphenicol acetyl transferase) cassette was excised from plasmid pCM7 and ligated into the blunt ended BglII site within the JHP0511 ORF. Resultant plasmids were transformed into E. coli DH5α, and clones harboring the desired insertion screened by colony PCR. Plasmids were subsequently sequenced to confirm that the cat cassette had been inserted in the appropriate orientation. Acceptable plasmids were subsequently used for natural transformation of H. pylori strain J99, and the cat insertion into the H. pylori genome confirmed by PCR analysis.

Growth Kinetics

Growth curves were obtained for HP0564^{-} in the same way as described for HP0222^{-} (Figure 23).

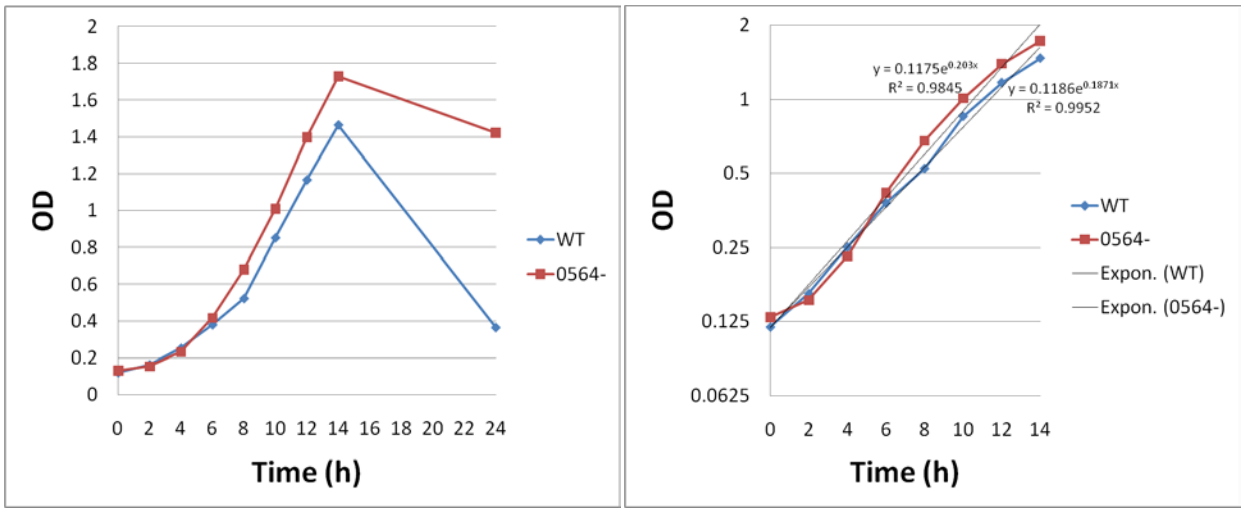

Figure 23. Growth curves of J99 WT and HP0564⁻ strains. The plot on the left shows the measured values. The plot on the right is on a logarithmic scale between 0 and 14 hours. Trendlines are shown, and their equations were used to extract doubling times for each strain.

Trendline equations were converted to base 2 for extraction of doubling times, which were taken as the inverse of the coefficient in front of x in the equations.

$$\text{WT}: y = 0.1186e^{0.1871x} = 0.1186 * 2^{0.2699x} \rightarrow \text{doubling time} = 3.7\ \text{h}$$

$$\text{HP0564}^{-}: y = 0.1175e^{0.2030x} = 0.1175 * 2^{0.2929x} \rightarrow \text{doubling time} = 3.4\ \text{h}$$

Over several experiments, HP0564⁻ consistently grew faster than WT.

Motility Assays

Motility assays were performed for HP0564⁻ at the same time as for HP0222⁻ (Figure 24). The results were much different.

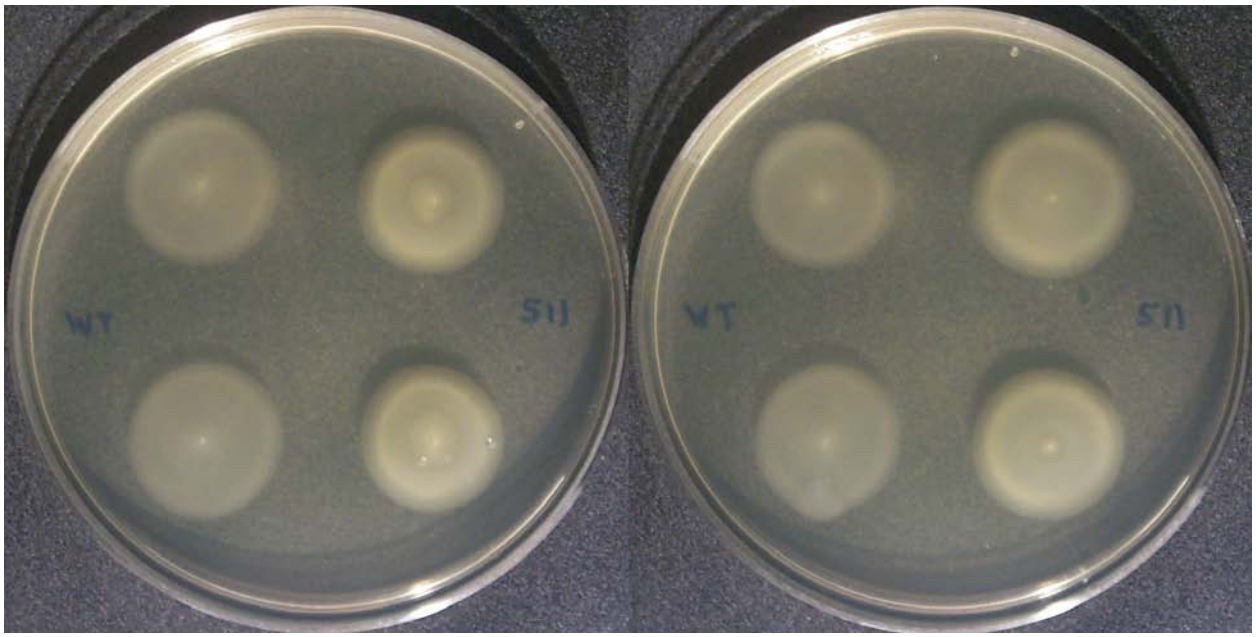

Figure 24. Motility assays of J99 WT and HP0564^{-} strains in semi-solid agar (0.5%) plates. Each plate shows two WT spots on the left and two HP0564^{-} spots on the right. Photographs were taken after 7 days.

All four HP0564^{-} spots have diameters that are on par with those of WT; however, the HP0564^{-} spots clearly have higher densities than WT. On these transparent plates, we also noticed that HP0564^{-} has an intense yellow color at higher densities.

Biological Analysis

Microarrays

Microarrays were used to determine the transcription profiles of the HP0564^{-} strain. Means and standard deviations of mutant to WT transcript ratios were similar to those for HP0222^{-}. Figure 25 plots the average ratio over 3 arrays for all genes on the array.

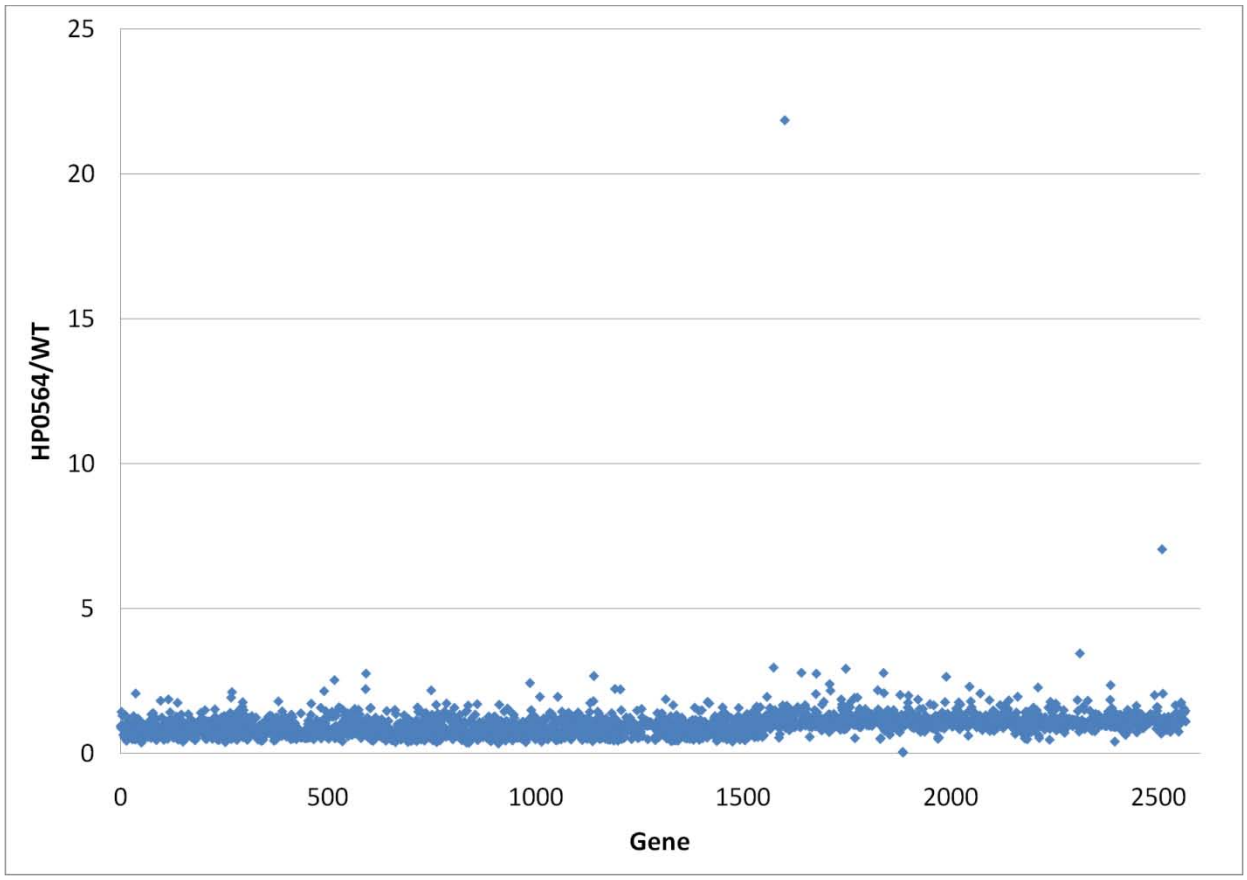

Figure 25. Graph of HP0564^{-} to WT transcript ratios for all 26695 and some J99 genes, averaged over three microarrays. For 26695 genes on the array, the average transcript ratio was 0.90, with a standard deviation of 0.31. For the J99 genes, the average ratio was 1.18, with a standard deviation of 0.74.

Table 9 shows a list of genes that were upregulated consistently across all the arrays with a transcript ratio at least two standard deviations above the mean. More genes were found that satisfied these requirements than for HP0222^{-}, which may be explained by the fact that only three arrays were used in the averaging instead of six. Two of the genes, JHP1101, a glucose/galactose transporter, and JHP0299, an ABC transporter, were found to be upregulated for both the 26695 and J99 sequences present on the array. Other genes of note that may help explain the observed phenotypes include the outer membrane proteins HorA and HofA, as well as the molybdate ABC transporter JHP0425.

Table 9. Microarray results for HP0564⁻

Gene (J99 #)	**Gene (26695 #)**	**Average (HP0564/WT)**	**Gene Product/Function**
JHP0073	HP0079	21.9	HorA – outer membrane protein
JHP1410	HP1521	7.06	restriction enzyme
JHP1101	HP1174	3.47	glucose/galactose transporter
JHP0036	HP0041/0042	2.98	ComB10 competence protein
JHP0299	HP0611/0612	2.94	ABC transporter
JHP0146	HP0158	2.80	unknown
JHP0425	HP0473	2.79	molybdate ABC transporter
JHP0195	HP0209	2.77	HofA – outer membrane protein
Gene (26695 #)	**Gene (J99 #)**	**Average (HP0564/WT)**	**Gene Product/Function**
HP1238	JHP1159	2.22	formamidase
HP1017	JHP0406	2.44	arginine permease
HP0611/0612	JHP0299	2.23/2.77	ABC transporter
HP1174	JHP1101	2.69	glucose/galactose transporter

Real-time PCR

At this point, we have not yet been able to perform a thorough real-time PCR validation of the microarray results.

Discussion

Observed Phenotypes

The HP0564⁻ strain behaved very differently from HP0222⁻ compared to WT in our experiments. HP0564⁻ consistently grew slightly faster than WT, and on motility agar plates, formed discs that were about the same size, if not slightly smaller, than WT. It is possible that

the upregulation of the glucose/galactose transporter JHP1101 is responsible for its faster growth; however, if it is truly less motile than WT, that could also explain its faster growth.

The yellowish color we observed for HP0564^{-} at higher densities both on plates and in liquid culture was very interesting. We do not have an explanation for this, but given the upregulation of transporters and outer membrane proteins, we can speculate that a molybdate compound or some other compound that is accumulating in the cell is responsible for giving the mutant its color.

Microarray Experiments

The microarray results for the HP0564^{-} strain looked promising. Means and standard deviations of transcript ratios were reasonable, and only a few genes were found to be highly upregulated. It is possible that the upregulated genes could explain the observed phenotypes; however, more work is needed to confirm the results and to determine whether the effects are a direct result of a lack of repression by HP0564 in our mutant.

We checked the transcript levels of the genes flanking HP0564. In this mutant, HP0563 was also highly downregulated. Because the antibiotic resistance cassette was inserted into the coding region of HP0564, it could not interfere with HP0563. Most likely, the homologous recombination event disrupted HP0563. There is no known function for HP0563, and at this point, we do not know what affects it might have on transcriptional regulation, if any.

Conclusions and Future Directions

There are many questions that remain to be answered concerning both the HP0222^{-} and HP0564^{-} strains. Microarray studies are very useful for providing a starting point for further

experiments because they cover all of the genes in the genome. One must be careful when interpreting microarray experiments, however. Even with good statistics, it is possible to get false positives due to nonspecific binding. In addition, genes that are truly upregulated may not be directly affected by the condition tested. In our case, we do not yet know whether our proteins bind the upstream promoter sequences of the genes that were found to be upregulated. Further experiments must be performed to determine whether the observed upregulation is due to direct binding or to other, indirect effects.

We would also like to know the consensus DNA binding sequences of our proteins. Our attempts to use SELEX to determine them have not been successful. We hope that we can find genes whose promoter sequences contain binding sites for our proteins and that we can narrow them down enough to recognize the consensus sequences by alignments of all the promoter sequences bound.

HP0222 was mentioned before in the literature as one of the most highly upregulated genes upon adhesion to AGS cells (Kim, Marcus *et al.* 2004). We do not know whether HP0564 is upregulated or downregulated in some condition or what that condition might be. We would like to know what conditions or proteins are involved in regulating the transcript levels of HP0222 and HP0564, but these questions will be much more difficult to answer.

We would also like to determine structures of our RHH proteins in complex with their consensus DNA binding sequences. There are examples of two different RHH proteins having the same set of DNA-binding residues in their β-sheets while recognizing different DNA sequences. Perhaps with more determined complex structures, we will be able to explain how structure determines DNA-binding specificity.

APPENDIX A

OTHER *H. PYLORI* PROJECTS

ArsS Sensor Domain

ArsS (HP0165) plays a crucial role in the acid response of *Helicobacter pylori*. It is a histidine kinase that senses extracellular, low pH conditions and phosphorylates the essential response regulator HP0166, resulting in altered transcript levels of many of the genes known to be involved in the acid response.

The amino acid sequence of the extracellular sensor domain has no homology to any protein except for similar pH sensors in *Helicobacter* species and *Wolinella succinogenes*. We were very interested in determining its structure to gain insight into how it senses low pH and transduces a signal across the membrane.

We tried expressing two constructs of this protein – one containing just the 100 extracellular domain residues and another containing the domain with the flanking transmembrane helices on either side (Figure 26). Both constructs were expressed from our modified pET vector that incorporates an N-terminal His_6 tag.

There was no expression of the construct containing the two flanking transmembrane helices, but the sensor domain alone was expressed very well. None of the domain, however, was found in the soluble fraction. To purify the protein in larger scale expressions, we performed inclusion body purifications using Inclusion Body Solubilization Reagent (Pierce) following the manufacturer's standard protocol.

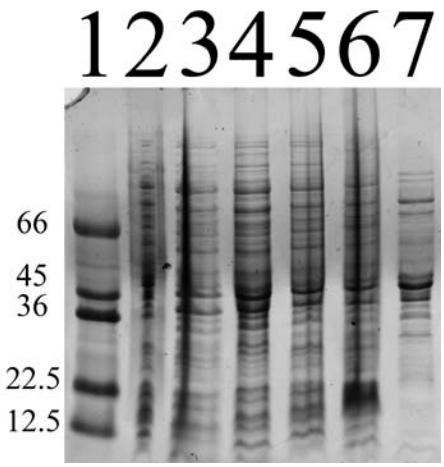

Figure 26. Expression tests of both ArsS sensor domain constructs. 1) molecular weight standards (sizes indicated at left), 2) transmembrane construct preinduced, 3) transmembrane construct postinduced, 4) soluble transmembrane construct postinduced, 5) sensor domain preinduced, 6) sensor domain postinduced, 7) soluble sensor domain postinduced

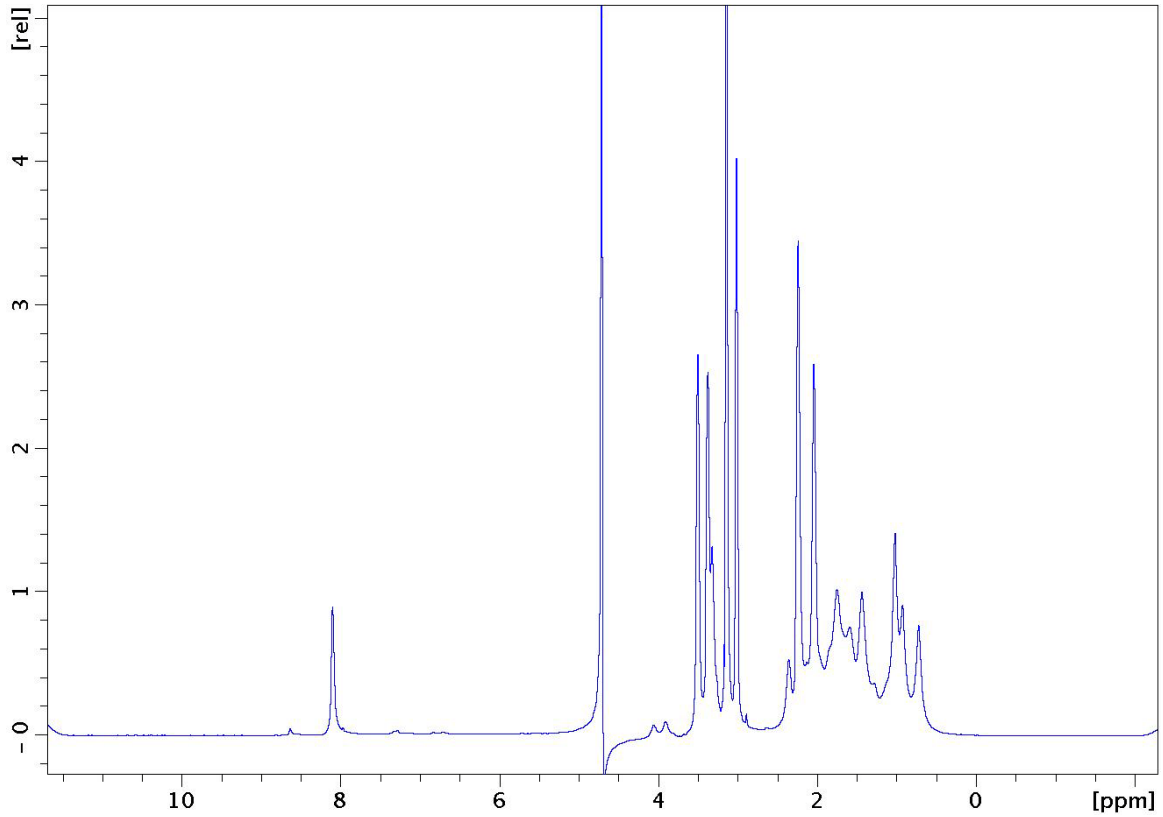

Figure 27. 1D ^{1}H NMR spectrum of the ArsS sensor domain. The sample is most likely aggregated, with visible peaks corresponding to unfolded protein.

In solution, the ArsS sensor behaved very poorly. Many sample conditions were tested including a range of pH values from 3 to 10 and salt concentrations from 0 to 500 mM. Several additives were tested, including the divalent metal ions Ni^{2+}, Mg^{2+}, Ca^{2+}, several non-detergent sulfobetaines (NDSBs), the amino acids glycine and arginine, and detergents such as CHAPS, Tween 20, and Triton X-100. All sample conditions resulted in a solution with gel-like consistency. A typical 1D ^{1}H NMR spectrum is shown in Figure 27. Based on the behavior of the sample, we expect that the protein may be largely aggregated. The part that can be seen in the NMR spectrum is unfolded. All amide peaks have collapsed to the random coil shift of 8.2 ppm. Aliphatic peaks are bunched into just a few broad peaks.

We have also tried crystallization of the sensor protein. A sample of protein was sent to the high-throughput screening lab at the Hauptman-Woodward Medical Research Institute, which tests 1536 conditions. Several hits were obtained, all in high concentrations of ammonium sulfate. We were unable to grow crystals in any of those conditions here.

JHP1348

JHP1348 (HP1455 in the 26695 strain) was another protein identified in the initial screen for structural targets from *H. pylori* (Popescu 2004). It is a 12.7 kDa protein of 113 amino acids and no assigned function. A BLAST search of the sequence yields only very similar proteins from other *Helicobacter* species, all of which are listed as hypothetical.

The N-terminal 14 residues of JHP1348 form a highly hydrophobic region that is predicted to be a portion of a cleaved signal sequence. We expressed JHP1348 in E. coli BL21 with an N-terminal His_6 tag (MRGSHHHHHHGS) and lacking 14 residues of the signal sequence. Uniformly $^{15}N/^{13}C$-labeled and ^{15}N-labeled proteins were produced by growing cells

in M9 media supplemented with ^{15}N-NH_4Cl and ^{13}C-glucose. Proteins were initially purified on a Ni-NTA column. Purification was enhanced by exploiting the pH-dependent solubility of JHP1348. The pH was lowered to 2.5, producing a pellet containing unwanted contaminants. Supernatant containing JHP1348 protein was collected, and the protein was precipitated by raising the pH to 7.5. The pellet was washed with 100 mM Tris buffer, pH 7.5, resuspended in H_2O, pH 4.0, and lyophilized.

NMR samples were produced by dissolving protein to 1 mM concentration in 90% H_2O/10% D_2O, pH 4.0. Spectra were recorded at 27° C on a Bruker Avance600 spectrometer equipped with a 5 mm cryoprobe. ^{1}H chemical shifts were referenced directly to 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) at 0.00 ppm, and ^{13}C and ^{15}N chemical shifts were referenced indirectly using absolute frequency ratios (Wishart, Bigam *et al.* 1995). Backbone assignments as well as C^{β} and H^{β} assignments were made based on CBCANH, CBCA(CO)NH, HBHANH, HBHA(CO)NH, and HNCO experiments (Sattler, Schleucher *et al.* 1999). Sidechain resonances were assigned using CC(CO)NH, H(CC)(CO)NH, HCCH-TOCSY, and ^{1}H-^{13}C HSQC experiments. Three-dimensional ^{15}N- and ^{13}C-edited NOESY-HSQC experiments were used to assign asparagine and glutamine sidechain NH_2 resonances and aromatic sidechain resonances to the correct residue. The 2D ^{1}H-^{15}N HSQC (Figure 28) was run on ^{15}N-labeled sample. NMR data were processed using XWINNMR ver. 2.6 (Bruker) and analyzed with Sparky ver. 3.111 (Goddard, T.D., Kneller, D.G. SPARKY 3, University of California, San Francisco). Secondary structure prediction was performed using the PECAN server at NMRFAM. Results are shown in Figure 30. Structurally, JHP1348 belongs to the mixed alpha/beta class.

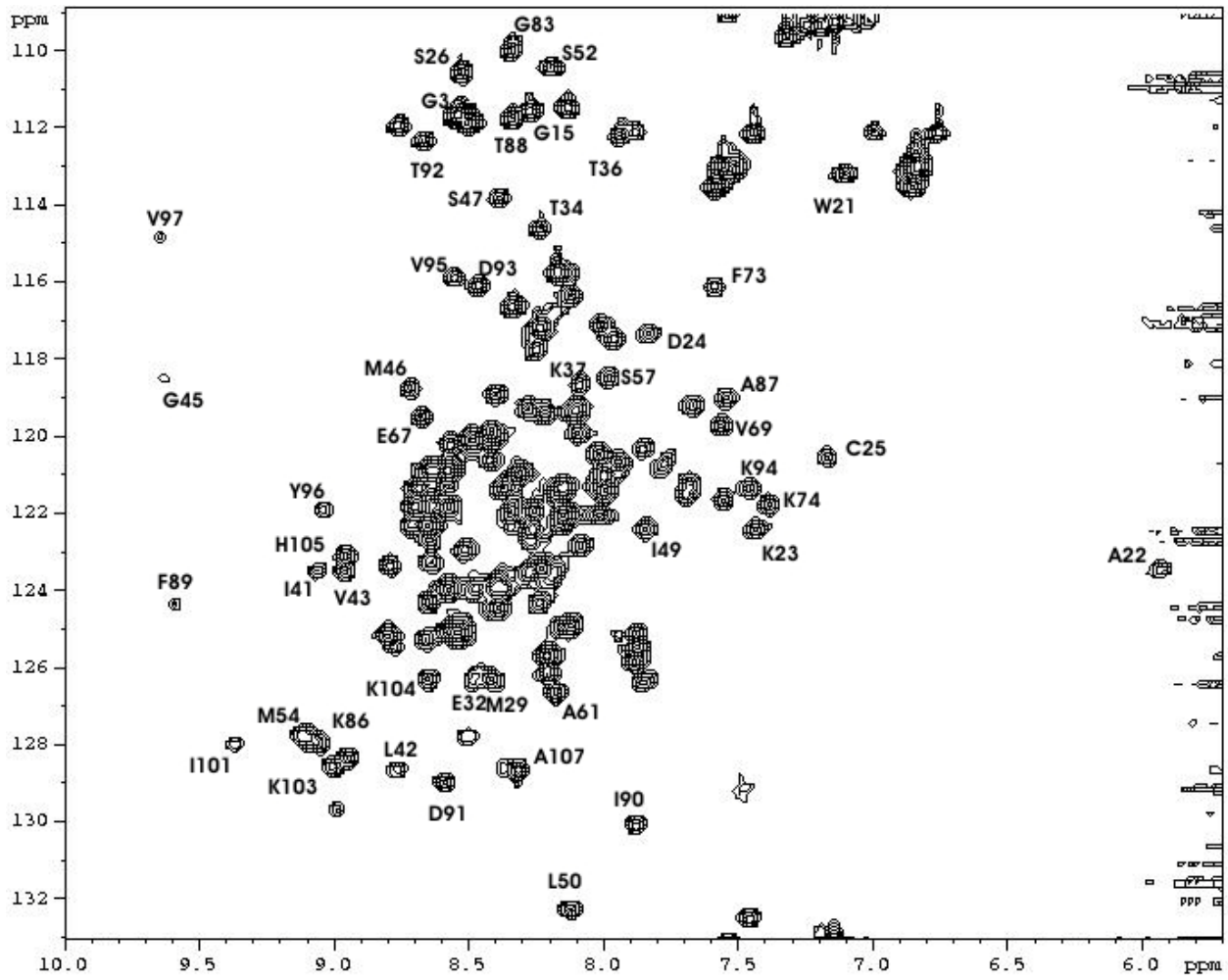

Figure 28. 600 MHz 2D ^{1}H-^{15}N HSQC spectrum of ^{15}N-labeled JHP1348 at 27 °C, pH 4.0. Sequential assignments are indicated with the one-letter amino acid code and residue number.

Experimental Procedures

Crosslinking experiments were performed to determine the oligomeric state of JHP1348 (Figure 29). The crosslinker BS^3, with an 11.4 Å spacer, was used. Reactions were set up in 10 mM NaH_2PO_4, pH 7.0, with 8 µg protein in each reaction. A control reaction without BS^3 is shown in lane 2 of Figure 29, followed by reactions with 5, 50, and 500 µM BS^3. In all lanes, including the control reaction, a dimer of JHP1348 can be seen. The amount of dimer, as well as very large molecular weight species, increases with increasing crosslinker concentration.

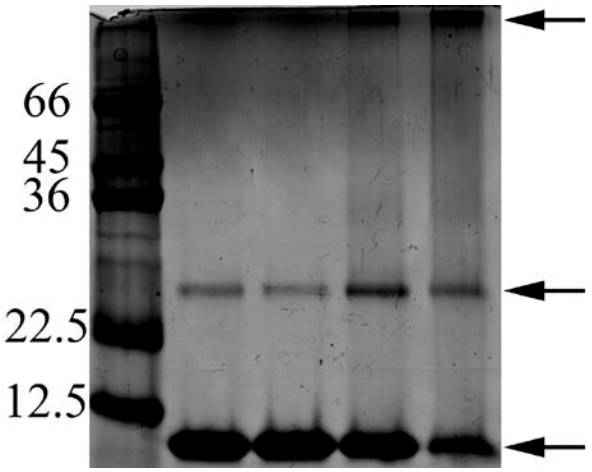

Figure 29. BS^3 crosslinking reactions with JHP1348 protein. Sizes of molecular weight standards are indicated to the left. Arrows on the right indicate monomer, dimer, and high molecular weight oligomers. The monomer is 11.4 kDa, so we expect the middle band to represent a dimer.

The sequential assignment process for JHP1348 was not quite straightforward. Although the ^{1}H-^{15}N HSQC spectrum was of reasonable quality, the number of backbone amide peaks was more than expected. These additional peaks were less intense than the others, and we suspected that these peaks corresponded to a partially folded form of the protein. The population of the partially folded form was significant enough to give rise to strong signals in 3D triple resonance experiments such as CBCANH. We chose the set of peaks with greater intensity for two reasons: i) the chemical shift dispersion of the amide protons for this set of peaks was greater than for the weaker peaks, so it was assumed that these peaks represented the fully folded form, ii) NOE crosspeaks in the ^{15}N-edited NOESY experiment were very weak or nonexistent for the set of lower intensity peaks.

Excluding the His_6 tag, we were able to assign nearly all backbone ^{1}H, ^{15}N, and ^{13}C resonances (Borin, Popescu *et al.* 2005). Only C′ resonances assignments for L18, which precedes a proline, and E111, which is the final residue, are missing. Over 87% of sidechain ^{1}H,

^{13}C, and ^{15}N resonances have been assigned. The ^{1}H, ^{13}C, and ^{15}N chemical shifts have been deposited in the BioMagResBank under accession number 6640.

Using the assigned chemical shifts of JHP1348, a secondary structure prediction was obtained using the PECAN server (Figure 30). The first 20 residues are not predicted to have regular secondary structure. The next 7 residues are weakly predicted to be helical. Following that is a β strand, a long helix and two more β strands.

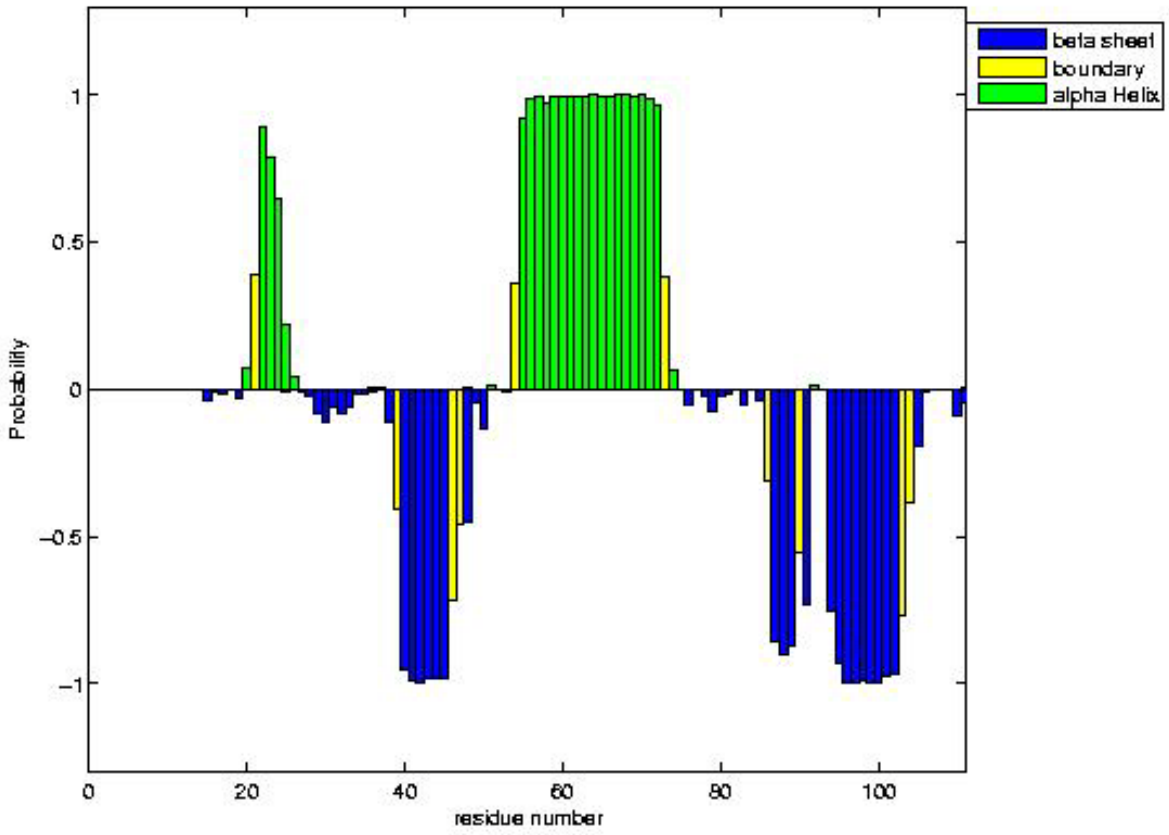

Figure 30. Secondary structure prediction from the PECAN server at NMRFAM based on chemical shifts.

A preliminary structure of JHP1348 was calculated using the more intense set of peaks for the N-terminal region. It has not yet been energy-minimized. JHP1348 features a three-strand, antiparallel β-sheet at its core (Figure 31). On one side of the sheet, a long α-helix crosses diagonally. The other side of the β-sheet, which is extremely hydrophobic, is partially covered by the N-terminal region, which forms a helical turn, consistent with the weak helical secondary structure prediction. If the structure had been calculated with the alternative set of N-terminal peaks, this region would not have been helical, and it would not have come into contact with the β-sheet.

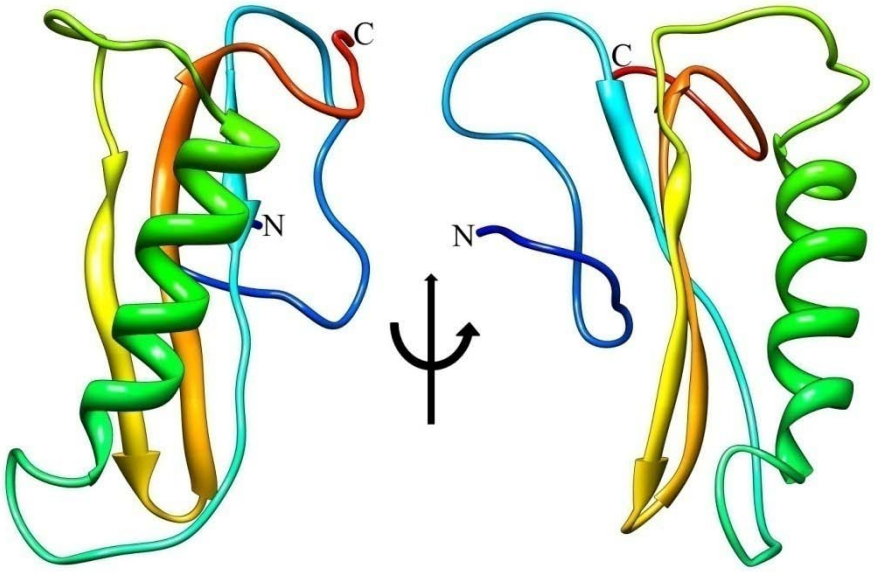

Figure 31. Two views of the JHP1348 structure calculated in CYANA

In order to get hints as to the possible function of JHP1348, the structure was submitted to the DALI server. DALI (distance matrix alignment) compares a given structure with all known structures in the PDB. Although many hits were returned by DALI, visual inspection of the structures was necessary to find ones that looked similar to JHP1348. Based on structural

similarity, JHP1348 is most likely a pantetheinyl transferase. These proteins are found in fatty acid synthesis and non-ribosomal peptide synthesis pathways where they cleave coenzyme A and transfer the pantetheinyl portion to a target protein. The active site is formed at the interface of two subunits in either a dimer or trimer. If JHP1348 is a pantetheinyl transferase, then our protein sample is not in the correct oligomeric state. Our NOESY data indicate that an N-terminal region of the protein is folded back across the hydrophobic side of the beta-sheet, blocking the dimerization interface. We tried expressing full-length and further truncated proteins, but they only increased hydrophobic aggregation.

APPENDIX B

CALCULATION SCRIPTS AND ASSIGNMENT TABLES FOR αCTD

CYANA Input Files

init.cya – contains initialization parameters for CYANA

```
nproc=4
name:=alpha
amberlib
read seq $name.seq
```

calc.cya – reads peak information, calculates distance restraints, and calculates structures

```
peaks       := n15.peaks,c13.peaks  # names of NOESY peak lists
prot        := alpha.prot           # names of chemical shift lists
constraints :=                      # additional (non-NOE) constraints
tolerance   := 0.035,0.035,0.6      # chemical shift tolerances
calibration :=                      # NOE calibration parameters
rmsdrange   := 24..107              # residue range for RMSD calculation
peakcheck prot=$prot info=full
calibration prot=$prot peaks=$peaks
peaks calibrate "**" simple
stereoassigns
distance modify info=full
caltab
write upl alpha.upl
ramaaco minimal
rotameraco
write aco alpha.aco
./init
read upl alpha.upl
read aco alpha.aco
read aco talos.aco append
hbonds
stereoassigns
distance stat info=full
seed=44465
calc_all 500 steps=25000
overview alpha.ovw structures=50 hbond=35 range=$rmsdrange pdb
```

alpha.prot – contains chemical shifts for all assigned atoms

atom #	**cs**	**dev.**	**name**	**res #**
1	55.53	0	CA	1
2	42.65	0	CB	1
3	24.98	0	CD1	1
4	23.58	0	CD2	1
5	27.1	0	CG	1
6	4.373	0.011	HA	1
7	0.862	0	QQD	1
8	45.41	0	CA	2
9	3.899	0.002	HA	2
10	8.313	0.002	HN	2
11	110.1	0.037	N	2
12	62.5	0	CA	3
13	32.69	0	CB	3
14	21	0	CG1	3
15	20.31	0	CG2	3
16	4.029	0.005	HA	3
17	1.927	0.003	HB	3
18	7.865	0.007	HN	3
19	120	0.045	N	3
20	0.754	0.003	QQG	3
21	58	0	CA	4
22	39.54	0	CB	4
23	4.617	0	HA	4
24	3.173	0.004	HB2	4
25	2.982	0.003	HB3	4
26	8.342	0.004	HN	4
27	124.2	0.054	N	4
30	45.34	0	CA	5
31	3.918	0	HA1	5
32	3.832	0	HA2	5
33	8.189	0.001	HN	5
34	111.3	0.016	N	5
35	56.45	0	CA	6
36	30.57	0	CB	6
37	36.31	0	CG	6
38	4.284	0	HA	6
39	2.029	0	HB2	6
40	1.917	0	HB3	6
41	8.222	0.004	HN	6
42	121.5	0.036	N	6
43	2.231	0	QG	6
44	8.323	0.001	HN	7
45	124	0.022	N	7
46	63.22	0	CA	8
47	32.09	0	CB	8
48	50.74	0	CD	8
49	27.81	0	CG	8
50	61.23	0	CA	9
51	38.86	0	CB	9
52	13.2	0.092	CD1	9
53	27.39	0.042	CG1	9
54	17.6	0	CG2	9
55	4.1	0.005	HA	9
56	1.82	0.006	HB	9
57	1.484	0.005	HG12	9
58	1.176	0.003	HG13	9
59	8.177	0.004	HN	9
60	122.2	0.046	N	9
61	0.851	0.016	QD1	9
62	0.897	0.001	QG2	9
63	52.46	0	CA	10
64	19.53	0	CB	10
65	8.328	0.003	HN	10
66	128.8	0.057	N	10
67	1.376	0.006	QB	10
68	53.47	0	CA	11
69	39.35	0	CB	11
70	8.422	0.008	HN	11
71	119.4	0.052	N	11
72	62.17	0	CA	12
73	69.83	0.104	CB	12
74	21.61	0	CG2	12
75	4.307	0.009	HA	12
76	4.213	0.015	HB	12
77	8.082	0.001	HN	12
78	114.9	0.004	N	12
79	1.156	0.005	QG2	12
80	56.78	0	CA	13
81	30.25	0	CB	13
82	36.31	0	CG	13
83	4.258	0	HA	13
84	8.371	0.002	HN	13
85	123.7	0.04	N	13
86	1.906	0	QB	13
87	2.184	0	QG	13
88	57.97	0	CA	14
89	38.91	0	CB	14
90	4.244	0	HA	14

alpha.prot, continued

91	3.036	0	HB2	14
92	2.943	0	HB3	14
93	8.167	0.003	HN	14
94	122.1	0.043	N	14
95	58.34	0	CA	15
96	64.15	0	CB	15
97	4.421	0	HA	15
98	8.212	0.006	HN	15
99	119.2	0.079	N	15
100	3.823	0	QB	15
101	45.46	0	CA	16
102	3.888	0.001	HA2	16
103	7.773	0.007	HN	16
104	111.1	0.008	N	16
105	54.49	0	CA	17
106	41.26	0	CB	17
107	4.572	0.004	HA	17
108	2.609	0.008	HB2	17
109	2.516	0.014	HB3	17
110	8.164	0.003	HN	17
111	121.2	0.029	N	17
112	58.27	0.031	CA	18
113	38.67	0.012	CB	18
114	4.457	0.004	HA	18
115	3.052	0.012	HB2	18
116	2.962	0.003	HB3	18
117	8.089	0.004	HN	18
118	121.8	0.032	N	18
119	7.09	0	QD	18
120	52.74	0	CA	19
121	19.16	0	CB	19
122	4.533	0	HA	19
123	8.07	0.003	HN	19
124	125.4	0.027	N	19
125	1.33	0.002	QB	19
126	56.01	0	CA	20
127	29.33	0	CB	20
128	33.98	0	CG	20
129	4.212	0	HA	20
130	8.104	0.001	HN	20
131	119.8	0.024	N	20
132	1.966	0	QB	20
968	55.74	0	CA	21
133	4.29	0.004	HA	21
134	1.859	0	HB2	21
135	1.757	0	HB3	21
136	8.203	0.006	HN	21
137	123.2	0.031	N	21
138	1.597	0	QG	21
139	54.43	0	CA	22
140	41.97	0	CB	22
141	4.639	0.007	HA	22
142	2.632	0	QB	22
143	54.99	0	CA	23
144	41.07	0	CB	23
145	4.55	0.001	HA	23
146	8.357	0.004	HN	23
147	122.5	0.058	N	23
148	2.707	0.003	QB	23
149	53.07	0.147	CA	24
150	18.71	0.085	CB	24
151	4.26	0.007	HA	24
152	8.162	0.015	HN	24
153	123.8	0.032	N	24
154	1.387	0.003	QB	24
155	57.76	0	CA	25
156	32.69	0	CB	25
157	28.84	0	CD	25
158	42.32	0	CE	25
159	24.94	0	CG	25
160	4.161	0.004	HA	25
161	8.058	0.007	HN	25
162	120.6	0.019	N	25
163	1.867	0.012	QB	25
164	1.733	0	QD	25
165	3.019	0	QE	25
166	1.419	0.016	QG	25
167	55.61	0.181	CA	26
168	4.538	0.011	HA	26
169	8.254	0.007	HN	26
170	120.5	0.084	N	26
171	2.746	0.007	QB	26
172	56.17	0.149	CA	27
173	41.46	0.074	CB	27
174	25.64	0.079	CD1	27
175	22.6	0.047	CD2	27
176	26.82	0	CG	27
177	4.009	0.01	HA	27
178	1.233	0.008	HB2	27
179	1.756	0.008	HB3	27
180	1.612	0.005	HG	27

alpha.prot, continued

181	8.09	0.006	HN	27
182	119.5	0.048	N	27
183	0.65	0.005	QD1	27
184	0.276	0.003	QD2	27
185	57.68	0.235	CA	28
186	63.6	0.077	CB	28
187	4.237	0.004	HA	28
188	7.918	0.01	HN	28
189	113.5	0.058	N	28
190	3.97	0.011	QB	28
191	52.09	0.064	CA	29
192	19.39	0.049	CB	29
193	4.347	0.008	HA	29
194	7.026	0.008	HN	29
195	123.4	0.028	N	29
196	1.498	0.007	QB	29
197	56.09	0.142	CA	30
198	33.09	0	CB	30
199	28.69	0	CD	30
200	42	0	CE	30
201	25.84	0	CG	30
202	4.91	0.007	HA	30
203	1.863	0.004	HB2	30
204	1.809	0.003	HB3	30
205	8.842	0.005	HN	30
206	121.1	0.043	N	30
207	3.055	0	QE	30
209	65.19	0.075	CA	31
210	38.08	0.069	CB	31
211	14.82	0.057	CD1	31
212	26.92	0.214	CG1	31
213	18.79	0.053	CG2	31
214	3.836	0.007	HA	31
215	1.859	0.005	HB	31
216	1.472	0.008	HG12	31
217	1.031	0.012	HG13	31
218	8.529	0.012	HN	31
219	120.6	0.087	N	31
220	0.584	0.007	QD1	31
221	0.781	0.005	QG2	31
222	59.65	0.081	CA	32
223	29.04	0	CB	32
224	36.26	0	CG	32
225	3.987	0.004	HA	32
226	9.534	0.006	HN	32
227	124.1	0.056	N	32
228	2.126	0.006	QB	32
229	2.248	0.008	QG	32
230	60.18	0.143	CA	33
231	64.04	0.062	CB	33
232	4.349	0.002	HA	33
233	7.929	0.005	HN	33
234	115.8	0.025	N	33
235	4.076	0.005	QB	33
972	30.6	0	CB	34
236	17.4	0.047	CE	34
237	4.238	0.005	HA	34
973	2.209	0.005	HB2	34
974	2.063	0.018	HB3	34
975	2.486	0.004	HG2	34
976	2.343	0.017	HG3	34
238	7.652	0.009	HN	34
239	120.1	0.027	N	34
241	1.941	0.002	QE	34
243	54.24	0.073	CA	35
244	36.94	0.308	CB	35
245	4.369	0.006	HA	35
246	2.969	0.008	HB2	35
247	2.727	0.003	HB3	35
248	7.489	0.005	HD21	35
249	6.735	0.012	HD22	35
250	7.876	0.007	HN	35
251	116.1	0.111	N	35
252	113.6	0.04	ND2	35
253	54.25	0.082	CA	36
254	43.62	0.179	CB	36
255	22.56	0.044	CD1	36
256	22.61	0.002	CD2	36
257	25.87	0	CG	36
258	4.399	0.007	HA	36
977	1.856	0.007	HB2	36
978	1.496	0.004	HB3	36
259	1.664	0.002	HG	36
260	8.624	0.008	HN	36
261	119.3	0.044	N	36
982	0.882	0.005	QD1	36
983	0.88	0.008	QD2	36
263	0.891	0.002	QQD	36
264	58.85	0.166	CA	37
265	63.25	0.142	CB	37
266	4.206	0.006	HA	37

alpha.prot, continued

267	10.44	0.009	HN	37
268	120.9	0.052	N	37
269	4.187	0.003	QB	37
270	56	0.061	CA	38
271	17.93	0.061	CB	38
272	4.175	0.009	HA	38
273	8.71	0.008	HN	38
274	125.7	0.043	N	38
275	1.487	0.004	QB	38
276	59.54	0	CA	39
277	29.88	0	CB	39
278	43.47	0	CD	39
279	27.42	0	CG	39
280	4.103	0.005	HA	39
281	8.452	0.006	HN	39
282	116.7	0.071	N	39
283	1.742	0	QB	39
284	3.278	0	QD	39
285	1.895	0	QG	39
286	62.5	0	CA	40
287	26.93	0	CB	40
288	4.034	0.016	HA	40
289	7.6	0.004	HN	40
290	121.9	0.034	N	40
291	2.967	0.008	QB	40
292	62.53	0	CA	41
293	39.35	0	CB	41
294	3.783	0.005	HA	41
295	3.084	0.005	HB2	41
296	3.003	0.01	HB3	41
990	7.111	0.002	HD1	41
991	7.123	0.005	HD2	41
992	7.256	0.001	HE1	41
993	7.236	0.006	HE2	41
297	8.982	0.003	HN	41
298	118.9	0.049	N	41
301	56.65	0	CA	42
302	38.42	0.041	CB	42
303	4.212	0.004	HA	42
304	2.995	0.004	HB2	42
305	2.759	0.005	HB3	42
306	7.7	0.008	HD21	42
307	7.031	0.018	HD22	42
308	8.793	0.01	HN	42
309	116.6	0.039	N	42
310	113.4	0.045	ND2	42
311	62.38	0.059	CA	43
312	26.47	0.063	CB	43
313	4.189	0.007	HA	43
314	3.183	0.004	HB2	43
315	2.925	0.007	HB3	43
316	7.486	0.014	HN	43
317	117.6	0.041	N	43
318	56.6	0	CA	44
319	41.84	0	CB	44
320	26.51	0.042	CD1	44
321	21.94	0.058	CD2	44
322	26.31	0	CG	44
323	3.994	0.018	HA	44
324	1.681	0.008	HG	44
325	7.121	0.007	HN	44
326	118.8	0.053	N	44
327	1.932	0.003	QB	44
328	0.765	0.002	QD1	44
329	0.714	0.008	QD2	44
330	57.35	0	CA	45
331	42.72	0.181	CB	45
332	4.086	0.001	HA	45
333	2.165	0.004	HB2	45
334	1.816	0.003	HB3	45
335	8.738	0.008	HN	45
336	121.2	0.101	N	45
337	58.52	0	CA	46
338	32.09	0	CB	46
339	28.15	0	CD	46
340	42.29	0	CE	46
341	25.41	0	CG	46
342	4.042	0.008	HA	46
343	1.953	0.002	HB2	46
344	1.859	0	HB3	46
345	7.875	0.009	HN	46
346	117.2	0.069	N	46
347	1.69	0.002	QD	46
348	2.995	0	QE	46
349	1.587	0.007	QG	46
350	61.13	0.06	CA	47
351	37.87	0.066	CB	47
352	15.35	0.05	CD1	47
353	27.08	0.127	CG1	47
354	18.3	0.06	CG2	47
355	4.553	0.004	HA	47

alpha.prot, continued

356	2.229	0.006	HB	47
357	6.952	0.006	HN	47
358	111.4	0.046	N	47
359	0.933	0.005	QD1	47
360	1.518	0.008	QG1	47
361	1.05	0.003	QG2	47
362	45.86	0.08	CA	48
363	3.761	0.008	HA1	48
364	4.128	0.01	HA2	48
365	7.704	0.007	HN	48
366	110.2	0.062	N	48
367	61.19	0	CA	49
368	36.22	0.073	CB	49
369	11.45	0.08	CD1	49
370	17.82	0.136	CG2	49
371	3.902	0.005	HA	49
372	1.644	0.01	HB	49
373	1.319	0.008	HG12	49
374	1.299	0.007	HG13	49
375	7.431	0.012	HN	49
376	122.4	0.028	N	49
377	0.706	0.006	QD1	49
998	1.299	0.017	QG1	49
378	0.766	0.006	QG2	49
379	57.05	0.118	CA	50
380	28.97	0.113	CB	50
381	24.62	0.11	CD	50
382	42.12	0	CE	50
383	23.05	0	CG	50
384	4.335	0.005	HA	50
970	1.603	0.002	HB2	50
971	1.563	0	HB3	50
385	9.493	0.008	HN	50
386	127.1	0.068	N	50
388	1.245	0.012	QD	50
389	2.936	0.009	QE	50
390	0.995	0.008	QG	50
391	56.85	0.075	CA	51
392	42.04	0.186	CB	51
393	5.385	0.006	HA	51
394	3.231	0.004	HB2	51
395	2.58	0.014	HB3	51
396	9.047	0.008	HN	51
397	121.5	0.072	N	51
398	7.022	0.007	QD	51
399	6.768	0.014	QE	51
400	66.98	0.115	CA	52
401	32.2	0	CB	52
402	21.3	0.165	CG1	52
403	22.51	0.071	CG2	52
404	3.398	0.006	HA	52
405	2.029	0.008	HB	52
406	9.17	0.013	HN	52
407	124.3	0.028	N	52
408	0.81	0.009	QG1	52
409	0.976	0.043	QG2	52
410	48.24	0.06	CA	53
411	3.652	0.004	HA1	53
412	3.653	0.008	HA2	53
413	8.949	0.01	HN	53
414	104.1	0.041	N	53
415	59.98	0	CA	54
416	30.94	0	CB	54
417	40.13	0.061	CG	54
418	3.901	0.004	HA	54
419	7.669	0.004	HN	54
420	117.6	0.034	N	54
421	2.288	0.003	QB	54
422	2.461	0.005	QG	54
423	57.05	0.052	CA	55
424	42.51	0.067	CB	55
425	24.01	0.094	CD1	55
984	24.63	0	CD2	55
426	26.7	0	CG	55
427	4.202	0.006	HA	55
428	2.208	0.002	HB2	55
429	1.42	0.006	HB3	55
987	1.518	0	HG	55
430	7.434	0.005	HN	55
431	116.9	0.049	N	55
988	0.891	0.004	QD1	55
989	0.899	0	QD2	55
433	64.2	0.109	CA	56
434	31.86	0.042	CB	56
435	22.62	0.182	CG1	56
436	21.06	0.065	CG2	56
437	3.893	0.007	HA	56
438	1.961	0.005	HB	56
439	6.924	0.017	HN	56
440	118.5	0.07	N	56
441	0.774	0.006	QG1	56

alpha.prot, continued

442	0.719	0.002	QG2	56
443	56	0.119	CA	57
444	43.45	0.311	CB	57
445	22.7	0.128	CD1	57
446	23.51	0.05	CD2	57
447	26.42	0	CG	57
448	4.061	0.008	HA	57
449	1.97	0.009	HB2	57
450	1.774	0.005	HB3	57
451	1.604	0.006	HG	57
452	6.994	0.012	HN	57
453	116.7	0.038	N	57
454	0.77	0.006	QD1	57
455	0.855	0.004	QD2	57
456	56.79	0	CA	58
457	34.83	0	CB	58
458	16.86	0.129	CE	58
459	32.51	0.108	CG	58
460	4.401	0.005	HA	58
461	2.448	0.006	HB2	58
462	2.167	0.009	HB3	58
463	2.457	0.008	HG2	58
464	2.985	0.008	HG3	58
465	7.253	0.006	HN	58
466	120.8	0.085	N	58
467	1.87	0.003	QE	58
468	56.95	0	CA	59
469	65.63	0.053	CB	59
470	4.616	0.005	HA	59
471	8.784	0.006	HN	59
472	118.5	0.046	N	59
473	4.422	0.009	QB	59
474	60.18	0.222	CA	60
475	30.01	0	CB	60
476	35.87	0	CG	60
477	3.966	0.009	HA	60
478	2.06	0	HB2	60
479	1.981	0	HB3	60
480	9.033	0.004	HN	60
481	123	0.07	N	60
482	2.407	0.007	QG	60
483	59.71	0	CA	61
484	29.38	0	CB	61
485	36.58	0.091	CG	61
486	3.866	0.005	HA	61
487	8.461	0.007	HN	61
488	117.1	0.019	N	61
489	1.984	0	QB	61
490	2.269	0.006	QG	61
491	59.17	0	CA	62
492	30.13	0	CB	62
493	37.6	0	CG	62
494	3.932	0.011	HA	62
495	7.828	0.004	HN	62
496	121.6	0.028	N	62
497	1.953	0.007	QB	62
498	2.32	0.005	QG	62
499	58.32	0.19	CA	63
500	42.53	0.078	CB	63
501	25.06	0.173	CD1	63
502	24.63	0.071	CD2	63
503	27.53	0.001	CG	63
504	4.164	0.009	HA	63
505	1.767	0.011	HG	63
506	8.384	0.007	HN	63
507	122.4	0.021	N	63
508	1.862	0.003	QB	63
509	0.923	0.009	QD1	63
510	0.97	0.003	QD2	63
511	59.11	0.055	CA	64
512	32.88	0.063	CB	64
513	29.5	0.067	CD	64
514	41.62	0.055	CE	64
515	26.61	0.215	CG	64
516	3.819	0.006	HA	64
517	8.196	0.006	HN	64
518	115.7	0.044	N	64
519	1.598	0.011	QB	64
520	1.312	0.008	QD	64
521	2.439	0.004	QE	64
522	0.695	0.005	QG	64
523	44.76	0.084	CA	65
524	4.138	0.005	HA1	65
525	3.614	0.006	HA2	65
526	7.477	0.005	HN	65
527	105.5	0.083	N	65
528	63.89	0.079	CA	66
529	32.19	0.049	CB	66
530	22.69	0.117	CG1	66
531	21.96	0.061	CG2	66
532	3.69	0.006	HA	66

alpha.prot, continued

533	2.115	0.006	HB	66
534	7.088	0.009	HN	66
535	124.5	0.028	N	66
536	1.162	0.004	QG1	66
537	0.936	0.007	QG2	66
538	58.58	0.029	CA	67
539	32.17	0	CB	67
540	28.88	0	CD	67
541	42.17	0	CE	67
542	24.49	0	CG	67
543	3.927	0.005	HA	67
544	8.374	0.005	HN	67
545	130.9	0.071	N	67
546	1.752	0.007	QB	67
547	2.985	0	QE	67
548	1.416	0.01	QG	67
549	55.12	0.011	CA	68
550	37.34	0.238	CB	68
551	4.312	0.004	HA	68
552	3.147	0.004	HB2	68
553	3.045	0.005	HB3	68
554	7.757	0.003	HD21	68
555	7.003	0.008	HD22	68
556	8.863	0.009	HN	68
557	117.2	0.06	N	68
558	114.9	0.032	ND2	68
559	56.06	0	CA	69
560	32.37	0	CB	69
561	16.77	0.079	CE	69
562	31.93	0.039	CG	69
563	4.639	0.004	HA	69
564	1.808	0.008	HB2	69
565	2.319	0	HB3	69
566	2.98	0.004	HG2	69
567	2.503	0.007	HG3	69
568	7.891	0.006	HN	69
569	122	0.035	N	69
570	1.989	0.006	QE	69
571	8.06	0.005	HN	70
572	115.5	0.23	N	70
573	4.099	0	QA	70
574	60.05	0	CA	71
575	32.87	0	CB	71
576	28.43	0	CD	71
577	42.41	0	CE	71
578	24.93	0	CG	71
579	4.08	0.001	HA	71
580	1.895	0	QB	71
581	1.76	0	QD	71
582	3.072	0	QE	71
583	1.513	0	QG	71
584	59.7	0.186	CA	72
585	32.16	0	CB	72
586	29.03	0	CD	72
587	42.19	0	CE	72
588	25.33	0	CG	72
589	4.23	0.004	HA	72
590	8.411	0.007	HN	72
591	117.9	0.067	N	72
592	1.911	0.006	QB	72
593	1.696	0	QD	72
594	3.052	0	QE	72
595	1.57	0	QG	72
596	58.68	0	CA	73
597	64.02	0	CB	73
598	4.47	0	HA	73
599	7.862	0.005	HN	73
600	118.3	0.097	N	73
601	3.876	0	QB	73
602	62.77	0.069	CA	74
603	38.11	0	CB	74
604	3.995	0.009	HA	74
605	2.954	0.009	HB2	74
606	3.067	0.004	HB3	74
607	8.317	0.005	HN	74
608	124.6	0.065	N	74
609	7.026	0.009	QD	74
610	6.698	0.005	QE	74
611	57.43	0.043	CA	75
612	40.06	0	CB	75
613	4.21	0.003	HA	75
614	2.812	0.002	HB2	75
615	2.657	0.004	HB3	75
616	8.735	0.003	HN	75
617	119.9	0.093	N	75
618	59.65	0	CA	76
619	29.79	0.114	CB	76
620	35.91	0.099	CG	76
621	4.091	0.002	HA	76
622	2.399	0.003	HB2	76
623	2.23	0.003	HB3	76

alpha.prot, continued

624	2.491	0.013	HG2	76
625	2.131	0.005	HG3	76
626	7.686	0.009	HN	76
627	121.8	0.016	N	76
628	64.06	0.119	CA	77
629	38.14	0.084	CB	77
630	15.29	0.134	CD1	77
631	29.97	0	CG1	77
632	18.73	0.141	CG2	77
633	3.576	0.005	HA	77
634	1.763	0.003	HB	77
635	7.916	0.013	HN	77
636	121.1	0.024	N	77
637	0.849	0.004	QD1	77
638	1.667	0.005	QG1	77
639	0.844	0.012	QG2	77
640	55.5	0	CA	78
641	17.31	0.067	CB	78
642	3.8	0.003	HA	78
643	8.841	0.008	HN	78
644	123.8	0.029	N	78
645	1.187	0.004	QB	78
646	59.45	0	CA	79
647	29.56	0	CB	79
648	36.33	0	CG	79
649	4.054	0.004	HA	79
650	8.062	0.006	HN	79
651	120.5	0.011	N	79
652	2.248	0.001	QB	79
653	2.411	0.001	QG	79
654	58.77	0	CA	80
655	32.03	0	CB	80
656	26.07	0	CD	80
657	42.74	0	CE	80
658	25.44	0	CG	80
659	4.17	0.006	HA	80
660	1.831	0	HD2	80
661	1.706	0	HG2	80
662	1.589	0	HG3	80
663	7.838	0.007	HN	80
664	120.5	0.042	N	80
665	1.975	0.007	QB	80
666	3.086	0.002	QE	80
667	58.41	0.051	CA	81
668	39.72	0.059	CB	81
669	24.72	1.812	CD1	81
670	26.08	0.004	CD2	81
671	27.21	0	CG	81
672	3.955	0.009	HA	81
673	2.076	0.006	HB2	81
674	1.675	0.006	HB3	81
675	8.712	0.009	HN	81
676	120.8	0.091	N	81
677	0.743	0.006	QD1	81
989	0.74	0.002	QD2	81
988	0.747	0.007	QQD	81
678	57.79	0	CA	82
679	38.89	0.124	CB	82
680	4.062	0.005	HA	82
681	2.943	0.003	HB2	82
682	3.049	0.007	HB3	82
683	7.854	0.001	HD21	82
684	6.852	0.001	HD22	82
685	7.885	0.003	HN	82
686	118.9	0.053	N	82
687	115.5	0.023	ND2	82
688	57.48	0	CA	83
689	40.75	0	CB	83
690	4.357	0.002	HA	83
691	2.909	0.01	HB2	83
692	2.724	0.008	HB3	83
693	8.239	0.005	HN	83
694	122.3	0.02	N	83
695	55.8	0	CA	84
696	43.53	0.139	CB	84
697	25.86	0.203	CD1	84
698	23.84	0.127	CD2	84
699	26.73	0	CG	84
700	4.262	0.007	HA	84
701	1.981	0.005	HB2	84
702	1.778	0.004	HB3	84
703	7.616	0.009	HN	84
704	119	0.031	N	84
705	0.886	0.007	QD1	84
706	0.893	0.005	QD2	84
707	44.07	0	CA	85
708	4.054	0.004	HA1	85
709	3.574	0.002	HA2	85
710	7.692	0.003	HN	85
711	105	0.057	N	85
712	55.14	0	CA	86

alpha.prot, continued

713	40.19	0	CB	86
714	4.719	0	HA	86
716	2.945	0.004	HB3	86
994	6.769	0.009	HD1	86
995	6.76	0	HD2	86
996	6.692	0.006	HE1	86
997	6.69	0	HE2	86
717	7.929	0.008	HN	86
718	119.5	0.061	N	86
721	62.82	0	CA	87
722	31.67	0	CB	87
723	49.81	0.044	CD	87
724	4.32	0.003	HA	87
725	3.476	0.004	HD2	87
726	3.298	0.008	HD3	87
985	2.011	0.004	HG2	87
986	1.867	0.005	HG3	87
727	2.441	0.004	QB	87
728	64.95	0.081	CA	88
729	31.79	0.019	CB	88
730	20.93	0.05	CG1	88
731	21.13	0.067	CG2	88
732	3.837	0.01	HA	88
733	1.966	0.008	HB	88
734	9.157	0.017	HN	88
735	123.9	0.04	N	88
736	1.073	0.015	QG1	88
737	1.071	0.013	QG2	88
738	45.03	0	CA	89
739	4.332	0.015	HA1	89
740	3.832	0.006	HA2	89
741	9.247	0.035	HN	89
742	115.9	0.256	N	89
743	64.35	0.116	CA	90
744	69.76	0.03	CB	90
745	21.29	0.059	CG2	90
746	4.088	0.016	HA	90
747	4.165	0.005	HB	90
748	8.026	0.007	HN	90
749	119.8	0.113	N	90
750	1.22	0.003	QG2	90
751	55.66	0.118	CA	91
752	30.78	0	CB	91
753	36.28	0	CG	91
754	4.467	0.008	HA	91
755	1.991	0.006	HB2	91
756	1.942	0	HB3	91
757	2.3	0	HG2	91
758	2.183	0	HG3	91
759	8.773	0.007	HN	91
760	129.2	0.157	N	91
761	54.16	0.125	CA	92
762	43.38	0.327	CB	92
763	25.89	0.15	CD1	92
764	27.16	0	CG	92
765	4.585	0.01	HA	92
766	1.597	0.007	HB2	92
767	1.471	0.004	HB3	92
768	1.684	0.009	HG	92
769	8.498	0.004	HN	92
770	127.2	0.074	N	92
771	0.738	0.004	QD1	92
772	0.771	0.006	QD2	92
773	56.96	0	CA	93
774	62.65	0	CB	93
775	4.738	0	HA	93
776	9.063	0.005	HN	93
777	121.6	0.058	N	93
778	4.046	0	QB	93
779	66.4	0.137	CA	94
780	31.73	0.111	CB	94
781	49.96	0.085	CD	94
969	28.1	0	CG	94
782	4.207	0.004	HA	94
783	2.426	0.013	HB2	94
784	2.406	0.001	HB3	94
785	2.091	0.001	HG2	94
786	2.067	0	HG3	94
787	3.962	0.007	QD	94
788	60.26	0	CA	95
789	29	0	CB	95
790	36.88	0	CG	95
791	4.108	0.005	HA	95
979	2.06	0	HB2	95
980	1.967	0	HB3	95
792	8.816	0.002	HN	95
793	118	0.075	N	95
795	2.295	0.002	QG	95
796	58.87	0	CA	96
797	29.48	0	CB	96
798	34.74	0.125	CG	96

alpha.prot, continued

799	4.085	0.005	HA	96
800	2.069	0	HB2	96
801	1.932	0.005	HB3	96
802	7.616	0.012	HE21	96
803	6.985	0.004	HE22	96
804	2.52	0.008	HG2	96
805	2.408	0.006	HG3	96
806	7.906	0.008	HN	96
807	122.2	0.037	N	96
808	112.8	0.02	NE2	96
809	60.33	0.127	CA	97
810	30.36	0	CB	97
811	43.11	0.03	CD	97
812	28.52	0.074	CG	97
813	3.753	0.008	HA	97
814	8.756	0.01	HN	97
815	121.2	0.045	N	97
816	1.854	0.002	QB	97
817	3.255	0.008	QD	97
981	1.408	0.008	QG	97
818	59.38	0	CA	98
819	29.38	0	CB	98
820	36.03	0	CG	98
821	4.043	0.006	HA	98
822	8.134	0.003	HN	98
823	119.3	0.039	N	98
824	2.088	0.02	QB	98
825	2.327	0.001	QG	98
826	61.51	0.061	CA	99
827	63	0	CB	99
828	4.156	0.006	HA	99
829	4.022	0	HB2	99
830	3.946	0.003	HB3	99
831	8.019	0.004	HN	99
832	114.8	0.065	N	99
833	58.06	0	CA	100
834	41.53	0.078	CB	100
835	25.82	0.094	CD1	100
836	23.32	0	CD2	100
837	27.21	0	CG	100
838	4.087	0.017	HA	100
839	1.861	0.001	HB2	100
840	1.568	0.003	HB3	100
841	1.411	0.016	HG	100
842	8.065	0.013	HN	100
843	124.3	0.047	N	100
844	0.739	0.001	QD1	100
845	60.33	0.118	CA	101
846	32.46	0	CB	101
847	29.37	0	CD	101
848	41.98	0	CE	101
849	25.84	0	CG	101
850	3.772	0.006	HA	101
851	8.302	0.009	HN	101
852	119.8	0.05	N	101
853	1.873	0.005	QB	101
854	1.629	0.01	QD	101
855	2.951	0	QE	101
856	1.437	0	QG	101
857	59.15	0.132	CA	102
858	32.33	0	CB	102
859	28.61	0	CD	102
860	42.19	0	CE	102
861	25.41	0	CG	102
862	4.05	0.008	HA	102
863	1.934	0	HB2	102
864	1.866	0	HB3	102
865	7.951	0.013	HN	102
866	118.4	0.066	N	102
867	1.632	0.008	QD	102
868	2.971	0	QE	102
869	1.506	0.007	QG	102
870	59.13	0.179	CA	103
871	30.27	0	CB	103
872	43.62	0	CD	103
873	27.82	0	CG	103
874	3.998	0.012	HA	103
875	1.964	0	HB2	103
876	1.904	0	HB3	103
877	7.863	0.012	HN	103
878	121.2	0.072	N	103
879	3.094	0	QD	103
880	1.694	0	QG	103
881	57.2	0.12	CA	104
882	40.93	0	CB	104
883	21.84	0.063	CD1	104
884	25.88	0.162	CD2	104
885	26.26	0	CG	104
886	3.809	0.007	HA	104
887	1.348	0.024	HG	104
888	8.148	0.015	HN	104

alpha.prot, continued

889	119	0.073	N	104
890	1.801	0.006	QB	104
891	0.463	0.004	QD1	104
892	0.653	0.007	QD2	104
893	59.21	0	CA	105
894	29.66	0	CB	105
895	36.76	0	CG	105
896	4.014	0.009	HA	105
897	7.987	0.004	HN	105
898	120.2	0.042	N	105
899	2.255	0.01	QB	105
900	2.475	0.001	QG	105
901	57.65	0.135	CA	106
902	32.23	0.07	CB	106
903	28.71	0	CD	106
904	42.29	0	CE	106
905	25.05	0.146	CG	106
906	4.22	0.004	HA	106
907	1.965	0.003	HB2	106
908	1.876	0.008	HB3	106
909	1.589	0	HG2	106
910	1.535	0.005	HG3	106
911	7.517	0.004	HN	106
912	118.9	0.024	N	106
913	1.707	0.006	QD	106
914	2.985	0	QE	106
915	56.78	0.115	CA	107
916	41.7	0.088	CB	107
917	25.16	0.048	CD1	107
918	22.7	0.126	CD2	107
919	26	0.187	CG	107
920	4.113	0.007	HA	107
921	1.718	0.006	HB2	107
922	1.456	0.009	HB3	107
923	1.713	0.002	HG	107
924	7.503	0.007	HN	107
925	120.7	0.066	N	107
926	0.449	0.004	QD1	107
927	0.769	0.006	QD2	107
928	56.67	0	CA	108
929	30.78	0	CB	108
930	36.58	0	CG	108
931	4.324	0.001	HA	108
932	2.42	0	HB2	108
933	2.333	0	HB3	108
934	7.671	0.007	HN	108
935	118.9	0.038	N	108
936	3.005	0	QG	108
937	54.8	0	CA	109
938	41.25	0	CB	109
939	4.626	0.003	HA	109
940	2.805	0	HB2	109
941	2.706	0	HB3	109
942	8.042	0.004	HN	109
943	122.2	0.072	N	109
944	56.56	0	CA	110
945	32.81	0	CB	110
946	28.63	0	CD	110
947	42.29	0	CE	110
948	24.67	0	CG	110
949	4.371	0.003	HA	110
950	8.309	0.013	HN	110
951	122.9	0.057	N	110
952	1.916	0	QB	110
953	1.721	0	QD	110
954	3.034	0	QE	110
955	1.492	0.005	QG	110
956	45.59	0	CA	111
957	4.005	0.002	HA1	111
958	8.452	0.004	HN	111
959	110.4	0.05	N	111
960	45.46	0	CA	112
961	3.999	0	HA1	112
962	8.305	0.001	HN	112
963	109.7	0.029	N	112
964	53.21	0	CA	113
965	39.58	0	CB	113
966	8.364	0.002	HN	113
967	119.7	0.046	N	113

hbonds.cya – contains hydrogen bond restraints

```
hbond HN 33 O 30
hbond HN 34 O 31

hbond HN 42 O 38
hbond HN 43 O 39
hbond HN 44 O 40
hbond HN 45 O 41
hbond HN 46 O 42
hbond HN 47 O 43
hbond HN 48 O 44

hbond HN 52 O 29

hbond HN 55 O 51
hbond HN 56 O 52
hbond HN 57 O 53

hbond HN 63 O 59
hbond HN 64 O 60

hbond HN 74 O 70
hbond HN 75 O 71
hbond HN 76 O 72
hbond HN 77 O 73
hbond HN 78 O 74
hbond HN 79 O 75
hbond HN 80 O 76
hbond HN 81 O 77
hbond HN 82 O 78
hbond HN 83 O 79
hbond HN 84 O 80

hbond HN 100 O 96
hbond HN 101 O 97
hbond HN 102 O 98
hbond HN 103 O 99
hbond HN 104 O 100
hbond HN 105 O 101
hbond HN 106 O 102
hbond HN 107 O 103

hbond HN 37 OE1 76
hbond HN 37 OE2 76
```

talos.aco – contains dihedral angle restraints from TALOS output

7 ARG+	PHI	-137.9	-43.5	45 ASP-	PSI	-51.3	-29.1
7 ARG+	PSI	118.6	179.8	46 LYS+	PHI	-81.2	-52.3
18 TYR	PHI	-79.4	-48.9	46 LYS+	PSI	-55	-8.2
18 TYR	PSI	-52.9	-29.4	47 ILE	PHI	-130.7	-64.2
23 ASP-	PHI	-87.6	-45.1	47 ILE	PSI	-34.8	30.6
23 ASP-	PSI	-58.6	-16.4	49 ILE	PHI	-115.7	-56.6
24 ALA	PHI	-88.2	-41.4	#49 ILE	PSI	107.9	153.2
24 ALA	PSI	-56.2	-16.1	#50 LYS+	PHI	-107.4	-54.1
25 LYS+	PHI	-99.6	-38.3	#50 LYS+	PSI	62.1	165.9
25 LYS+	PSI	-84.3	32.6	52 VAL	PHI	-76.2	-46.4
26 ASP-	PHI	-101.5	-40.4	52 VAL	PSI	-52.3	-25.1
26 ASP-	PSI	-85.2	32.8	53 GLY	PHI	-72.9	-52.9
27 LEU	PHI	-100.3	-41.6	53 GLY	PSI	-49	-28.9
27 LEU	PSI	-70	16.8	54 GLU-	PHI	-77.8	-56.7
#31 ILE	PHI	-73.8	-47.1	54 GLU-	PSI	-50.8	-29.4
#31 ILE	PSI	-47.6	-26.2	56 VAL	PHI	-97.7	-46.3
#32 GLU-	PHI	-80.2	-57	56 VAL	PSI	-50.4	-25.1
#32 GLU-	PSI	-51.1	-31.1	57 LEU	PHI	-102.3	-55.1
#33 SER	PHI	-87	-51.3	57 LEU	PSI	-47.7	-9.6
#33 SER	PSI	-53.5	-17.9	59 SER	PHI	-137.3	-46.6
#34 MET	PHI	-71.7	-51.6	59 SER	PSI	121.7	196.1
#34 MET	PSI	-53.3	-33.3	60 GLU-	PHI	-70.8	-47.6
#35 ASN	PHI	-91.4	-49.6	60 GLU-	PSI	-52.5	-26.9
#35 ASN	PSI	-63.3	15	61 GLU-	PHI	-73.5	-53.5
#36 LEU	PHI	-147.2	-45.3	61 GLU-	PSI	-53.1	-27.8
#36 LEU	PSI	83.9	165.1	62 GLU-	PHI	-77	-57
38 ALA	PHI	-78	-44.5	62 GLU-	PSI	-51.7	-29.6
38 ALA	PSI	-59.6	-30	63 LEU	PHI	-72.6	-52.6
39 ARG+	PHI	-74.2	-54.2	63 LEU	PSI	-54.1	-28
39 ARG+	PSI	-50.8	-26	66 VAL	PHI	-114.9	-56.2
40 CYS	PHI	-79.6	-58.1	66 VAL	PSI	107.3	139
40 CYS	PSI	-50.9	-30.7	67 LYS+	PHI	-80.1	-41
41 PHE	PHI	-74.1	-50.5	67 LYS+	PSI	-62.2	-22
41 PHE	PSI	-55	-34.3	71 LYS+	PHI	-77.9	-50.6
42 ASN	PHI	-69.4	-49.4	71 LYS+	PSI	-56.6	-15
42 ASN	PSI	-57.1	-20.9	72 LYS+	PHI	-77.3	-48.5
43 CYS	PHI	-72.4	-52.4	72 LYS+	PSI	-60.3	-18.1
43 CYS	PSI	-63.2	-17.1	73 SER	PHI	-88.2	-54.6
44 LEU	PHI	-74.7	-54.7	73 SER	PSI	-59.8	-8.2
44 LEU	PSI	-48.4	-28.4	74 TYR	PHI	-73.6	-42.9
45 ASP-	PHI	-76.3	-52.8	74 TYR	PSI	-68.8	-25.5

talos.aco, continued

75 ASP-	PHI	-77.8	-50.4
75 ASP-	PSI	-51.1	-24.4
76 GLU-	PHI	-78.2	-57.9
76 GLU-	PSI	-52.2	-27.5
77 ILE	PHI	-71.3	-51.3
77 ILE	PSI	-55.5	-28.9
78 ALA	PHI	-73.6	-52.3
78 ALA	PSI	-54.1	-26.3
79 GLU-	PHI	-77.3	-53.7
79 GLU-	PSI	-53.3	-33.3
80 LYS+	PHI	-81.2	-55
80 LYS+	PSI	-51.9	-30
81 LEU	PHI	-77.6	-54.7
81 LEU	PSI	-52.1	-32.1
82 ASN	PHI	-72.4	-52.4
82 ASN	PSI	-53.3	-33.3
83 ASP-	PHI	-76.4	-53.2
83 ASP-	PSI	-54.2	-12.8
86 TYR	PHI	-135.4	-56
86 TYR	PSI	67	173.5
88 VAL	PHI	-113.8	-26.1
88 VAL	PSI	113.7	148.4
90 THR	PHI	-100.8	-67.3
90 THR	PSI	113.5	146.8
91 GLU-	PHI	-119.2	-64
91 GLU-	PSI	112.9	156.5
#93 SER	PHI	-127.6	-41
#93 SER	PSI	84.2	202.3
95 GLU-	PHI	-79.5	-53.7
95 GLU-	PSI	-54.9	-21.7
96 GLN	PHI	-74.5	-54.5
96 GLN	PSI	-52.9	-32.9
97 ARG+	PHI	-77.5	-51.5
97 ARG+	PSI	-55.7	-27.5
98 GLU-	PHI	-80	-54.3
98 GLU-	PSI	-58	-17.4
99 SER	PHI	-74.5	-54.5
99 SER	PSI	-49.9	-29.9
100 LEU	PHI	-77.3	-50.9
100 LEU	PSI	-50.8	-30.8
101 LYS+	PHI	-76.9	-50.3
101 LYS+	PSI	-53.9	-28.4
102 LYS+	PHI	-73.4	-53.4
102 LYS+	PSI	-52.4	-32.4
103 ARG+	PHI	-74.5	-53.5
103 ARG+	PSI	-53	-29.1
104 LEU	PHI	-77.7	-54.2
104 LEU	PSI	-44.8	-24.8
105 GLU-	PHI	-82	-49
105 GLU-	PSI	-54.6	-27.2
106 LYS+	PHI	-73.1	-52.8
106 LYS+	PSI	-57.3	-17.9
107 LEU	PHI	-95.2	-45.2
107 LEU	PSI	-74.1	16

stereoassigns.cya – contains stereospecific assignments

```
atom stereo "27 HB2 HB3 QD1 QD2"
atom stereo "36 QD1 QD2"
atom stereo "41 HD1 HD2"
atom stereo "41 HE1 HE2"
atom stereo "44 QD1 QD2"
atom stereo "52 QG1 QG2"
atom stereo "55 QD1 QD2"
atom stereo "56 QG1 QG2"
atom stereo "57 QD1 QD2"
atom stereo "63 QD1 QD2"
atom stereo "66 QG1 QG2"
atom stereo "81 QD1 QD2"
atom stereo "84 QD1 QD2"
atom stereo "86 HD1 HD2"
atom stereo "86 HE1 HE2"
atom stereo "88 QG1 QG2"
atom stereo "92 QD1 QD2"
atom stereo "100 QD1 QD2"
atom stereo "104 QD1 QD2"
atom stereo "107 QD1 QD2"
```

alpha.upl – contains distance restraints based on NOEs

```
55   LEU   HN   56   VAL   HN   3.65   #peak   3
57   LEU   HN   58   MET   HN   3.64   #peak   5
58   MET   HN   59   SER   HN   4.81   #peak   444
54   GLU-  HN   55   LEU   HN   3.69   #peak   1
53   GLY   HN   54   GLU-  HN   4.34   #peak   8
52   VAL   HN   53   GLY   HN   4.13   #peak   11
50   LYS+  HN   51   TYR   HN   3.39   #peak   12
48   GLY   HN   49   ILE   HN   3.56   #peak   15
47   ILE   HN   48   GLY   HN   3.37   #peak   16
46   LYS+  HN   48   GLY   HN   4.45   #peak   17
47   ILE   HN   49   ILE   HN   4.35   #peak   337
45   ASP-  HN   46   LYS+  HN   3.46   #peak   21
44   LEU   HN   45   ASP-  HN   3.52   #peak   22
43   CYS   HN   44   LEU   HN   3.48   #peak   25
42   ASN   HN   43   CYS   HN   3.64   #peak   26
41   PHE   HN   42   ASN   HN   3.78   #peak   28
40   CYS   HN   41   PHE   HN   3.65   #peak   31
35   ASN   HN   36   LEU   HN   4.56   #peak   33
32   GLU-  HN   33   SER   HN   4.13   #peak   35
33   SER   HN   34   MET   HN   3.71   #peak   36
60   GLU-  HN   61   GLU-  HN   3.84   #peak   40
```

“alpha.upl, continued”

61	GLU-	HN	62	GLU-	HN	3.67	#peak	43
62	GLU-	HN	63	LEU	HN	3.6	#peak	45
63	LEU	HN	64	LYS+	HN	3.64	#peak	46
64	LYS+	HN	65	GLY	HN	3.59	#peak	48
65	GLY	HN	66	VAL	HN	3.56	#peak	51
68	ASN	HN	69	MET	HN	4.15	#peak	53
75	ASP-	HN	76	GLU-	HN	3.63	#peak	54
76	GLU-	HN	77	ILE	HN	3.56	#peak	56
77	ILE	HN	78	ALA	HN	3.58	#peak	58
78	ALA	HN	79	GLU-	HN	3.52	#peak	61
79	GLU-	HN	80	LYS+	HN	3.44	#peak	62
80	LYS+	HN	81	LEU	HN	3.56	#peak	93
82	ASN	HN	83	ASP-	HN	3.47	#peak	65
83	ASP-	HN	84	LEU	HN	3.53	#peak	66
85	GLY	HN	86	TYR	HN	3.38	#peak	69
86	TYR	HN	86	TYR	HD1	3.83	#peak	70
89	GLY	HN	90	THR	HN	3.45	#peak	72
98	GLU-	HN	99	SER	HN	3.46	#peak	74
100	LEU	HN	101	LYS+	HN	3.46	#peak	76
101	LYS+	HN	102	LYS+	HN	3.52	#peak	77
102	LYS+	HN	103	ARG+	HN	3.39	#peak	80
103	ARG+	HN	104	LEU	HN	3.53	#peak	81
104	LEU	HN	105	GLU-	HN	3.49	#peak	83
105	GLU-	HN	106	LYS+	HN	3.42	#peak	85
107	LEU	HN	108	GLU-	HN	3.7	#peak	88
96	GLN	HN	97	ARG+	HN	3.54	#peak	91
28	SER	HN	29	ALA	HN	3.39	#peak	97
29	ALA	HN	30	LYS+	HN	4.33	#peak	99
30	LYS+	HN	31	ILE	HN	4.75	#peak	100
73	SER	HN	74	TYR	HN	3.95	#peak	102
74	TYR	HN	75	ASP-	HN	3.69	#peak	104
95	GLU-	HN	96	GLN	HN	4.24	#peak	105
27	LEU	HN	28	SER	HN	4.04	#peak	108
26	ASP-	HN	27	LEU	HN	4.5	#peak	109
39	ARG+	HN	40	CYS	HN	4.36	#peak	32
109	ASP-	HN	109	ASP-	HB3	4.01	#peak	111
109	ASP-	HN	109	ASP-	HB2	4.01	#peak	112
93	SER	HN	93	SER	QB	3.9	#peak	114
76	GLU-	HN	76	GLU-	HB2	3.52	#peak	118
47	ILE	HN	47	ILE	QD1	3.89	#peak	113
49	ILE	HN	49	ILE	QD1	3.72	#peak	148
77	ILE	HN	77	ILE	QG1	3.62	#peak	125
67	LYS+	HN	67	LYS+	QB	3.35	#peak	129
3	VAL	HN	3	VAL	QQG	4.05	#peak	135
9	ILE	HN	9	ILE	HB	3.7	#peak	142

“alpha.upl, continued”

9	ILE	HN	9	ILE	HG12	4.44	#peak	143
9	ILE	HN	9	ILE	HG13	4.44	#peak	144
9	ILE	HN	9	ILE	QG2	4.45	#peak	145
10	ALA	HN	10	ALA	QB	4.17	#peak	146
19	ALA	HN	20	GLN	HA	4.44	#peak	158
19	ALA	HN	19	ALA	QB	3.69	#peak	160
19	ALA	QB	20	GLN	HN	4.52	#peak	161
27	LEU	HN	27	LEU	HG	3.77	#peak	163
21	ARG+	HN	21	ARG+	QG	4.73	#peak	168
23	ASP-	HN	23	ASP-	QB	3.71	#peak	169
22	ASP-	HA	23	ASP-	HN	3.52	#peak	171
23	ASP-	HN	24	ALA	HN	4.84	#peak	172
23	ASP-	QB	24	ALA	HN	4.32	#peak	175
79	GLU-	HN	79	GLU-	QG	4.12	#peak	177
25	LYS+	HN	25	LYS+	QG	4.16	#peak	178
78	ALA	QB	79	GLU-	HN	3.39	#peak	179
25	LYS+	HN	25	LYS+	QB	3.7	#peak	180
76	GLU-	HA	79	GLU-	HN	3.27	#peak	182
26	ASP-	HN	26	ASP-	QB	3.8	#peak	186
25	LYS+	QB	26	ASP-	HN	4.1	#peak	187
27	LEU	HN	27	LEU	HB3	4.14	#peak	191
26	ASP-	QB	27	LEU	HN	4.48	#peak	192
26	ASP-	HA	28	SER	HN	4.85	#peak	195
28	SER	HN	28	SER	QB	3.97	#peak	197
27	LEU	HB3	28	SER	HN	4.8	#peak	198
27	LEU	HG	28	SER	HN	4.81	#peak	199
29	ALA	HN	52	VAL	QG1	4.52	#peak	201
29	ALA	HN	52	VAL	QG2	4.37	#peak	202
29	ALA	HN	52	VAL	HB	4.27	#peak	204
28	SER	QB	29	ALA	HN	4.16	#peak	205
26	ASP-	HA	29	ALA	HN	4.46	#peak	208
27	LEU	HN	29	ALA	HN	4.75	#peak	209
30	LYS+	HN	30	LYS+	HB2	3.57	#peak	212
30	LYS+	HN	30	LYS+	HB3	3.57	#peak	213
29	ALA	QB	30	LYS+	HN	3.37	#peak	214
31	ILE	HN	31	ILE	QG2	3.53	#peak	216
31	ILE	HN	31	ILE	HG13	3.84	#peak	217
31	ILE	HN	31	ILE	HG12	3.84	#peak	218
31	ILE	HN	31	ILE	HB	3.84	#peak	219
31	ILE	HN	50	LYS+	HA	4.65	#peak	221
31	ILE	HN	51	TYR	QD	4.62	#peak	224
31	ILE	HN	32	GLU-	HN	4.17	#peak	227
32	GLU-	HN	41	PHE	HE2	4.26	#peak	229
30	LYS+	HA	32	GLU-	HN	4.65	#peak	230
32	GLU-	HN	32	GLU-	QG	4.15	#peak	233

“alpha.upl, continued”

30	LYS+	HB2	32	GLU-	HN	4.2	#peak	235
31	ILE	HG12	32	GLU-	HN	4.58	#peak	236
31	ILE	HG13	32	GLU-	HN	4.58	#peak	237
31	ILE	QG2	32	GLU-	HN	4.39	#peak	238
32	GLU-	QB	33	SER	HN	4.6	#peak	240
32	GLU-	QG	33	SER	HN	4.5	#peak	241
31	ILE	HA	33	SER	HN	4.54	#peak	243
34	MET	HN	35	ASN	HN	4.44	#peak	247
33	SER	QB	34	MET	HN	4.64	#peak	250
32	GLU-	HA	34	MET	HN	4.69	#peak	251
31	ILE	HA	34	MET	HN	4.52	#peak	252
34	MET	HN	34	MET	HG2	3.98	#peak	253
34	MET	HN	34	MET	HG3	3.98	#peak	254
34	MET	HN	34	MET	HB2	3.94	#peak	255
34	MET	HN	34	MET	HB3	3.94	#peak	736
34	MET	HN	36	LEU	HG	4.57	#peak	257
34	MET	HN	36	LEU	QD1	4.29	#peak	258
35	ASN	HB2	36	LEU	HN	5.08	#peak	262
34	MET	HB3	36	LEU	HN	4.75	#peak	263
36	LEU	HN	36	LEU	HB2	4.1	#peak	264
36	LEU	HN	36	LEU	HG	3.5	#peak	265
36	LEU	HN	36	LEU	HB3	4.1	#peak	266
36	LEU	HN	36	LEU	QQD	3.62	#peak	267
36	LEU	QD2	37	SER	HN	3.9	#peak	268
36	LEU	HB3	37	SER	HN	4.32	#peak	269
36	LEU	HG	37	SER	HN	4.83	#peak	270
36	LEU	HB2	37	SER	HN	4.32	#peak	271
37	SER	HN	40	CYS	QB	4.38	#peak	272
37	SER	HN	37	SER	QB	3.7	#peak	273
36	LEU	HA	37	SER	HN	3.2	#peak	274
38	ALA	QB	39	ARG+	HN	3.71	#peak	87
61	GLU-	HN	61	GLU-	QB	3.09	#peak	278
37	SER	QB	40	CYS	HN	4.49	#peak	282
40	CYS	HN	40	CYS	QB	3.45	#peak	284
38	ALA	QB	40	CYS	HN	4.74	#peak	285
41	PHE	HN	41	PHE	HB3	3.58	#peak	288
41	PHE	HN	41	PHE	HB2	3.58	#peak	289
38	ALA	HA	41	PHE	HN	4.17	#peak	292
41	PHE	HD1	42	ASN	HN	3.93	#peak	293
39	ARG+	HA	42	ASN	HN	4.3	#peak	295
42	ASN	HB3	43	CYS	HN	3.94	#peak	300
42	ASN	HB2	43	CYS	HN	3.94	#peak	301
43	CYS	HN	43	CYS	HB2	3.63	#peak	302
40	CYS	HA	43	CYS	HN	4.27	#peak	303
41	PHE	HA	44	LEU	HN	4.71	#peak	307

“alpha.upl, continued”

43	CYS	HB2	44	LEU	HN	4.24	#peak	100
44	LEU	HN	44	LEU	QB	3.53	#peak	310
44	LEU	HN	44	LEU	HG	3.71	#peak	311
31	ILE	QD1	45	ASP-	HN	4.56	#peak	313
81	LEU	HN	81	LEU	HB3	3.82	#peak	558
97	ARG+	HN	97	ARG+	QB	3.26	#peak	315
45	ASP-	HN	45	ASP-	HB2	3.47	#peak	316
97	ARG+	HN	97	ARG+	QD	4.54	#peak	318
42	ASN	HA	45	ASP-	HN	3.8	#peak	322
44	LEU	HN	46	LYS+	HN	4.77	#peak	323
43	CYS	HA	46	LYS+	HN	4.07	#peak	324
46	LYS+	HN	46	LYS+	HB2	3.44	#peak	327
46	LYS+	HN	46	LYS+	HB3	3.44	#peak	328
46	LYS+	HN	46	LYS+	QD	4.2	#peak	329
46	LYS+	HN	46	LYS+	QG	3.56	#peak	330
47	ILE	HN	47	ILE	QG1	3.41	#peak	331
46	LYS+	HB3	47	ILE	HN	4.53	#peak	332
47	ILE	HN	47	ILE	HB	4.18	#peak	333
47	ILE	HN	48	GLY	HA1	5.03	#peak	334
46	LYS+	HN	47	ILE	HN	3.71	#peak	19
48	GLY	HN	49	ILE	HB	4.76	#peak	342
47	ILE	QG1	48	GLY	HN	4.25	#peak	343
47	ILE	QG2	48	GLY	HN	4	#peak	122
47	ILE	QG1	49	ILE	HN	3.95	#peak	348
49	ILE	HN	49	ILE	HB	3.33	#peak	349
49	ILE	HN	50	LYS+	HN	4.82	#peak	353
49	ILE	HA	50	LYS+	HN	3.28	#peak	355
49	ILE	HB	50	LYS+	HN	4.78	#peak	356
50	LYS+	HN	50	LYS+	QD	3.98	#peak	357
50	LYS+	HN	50	LYS+	QG	4.47	#peak	358
49	ILE	QG2	50	LYS+	HN	3.92	#peak	359
92	LEU	QD2	93	SER	HN	3.88	#peak	360
50	LYS+	QG	51	TYR	HN	4.36	#peak	361
49	ILE	QG1	51	TYR	HN	4.14	#peak	362
92	LEU	HB3	93	SER	HN	4.29	#peak	363
92	LEU	HB2	93	SER	HN	4.29	#peak	364
49	ILE	HA	51	TYR	HN	4.02	#peak	367
91	GLU-	HA	93	SER	HN	4.17	#peak	369
92	LEU	HA	93	SER	HN	3.16	#peak	370
51	TYR	HN	51	TYR	QD	3.98	#peak	372
31	ILE	HN	52	VAL	HN	4.5	#peak	373
29	ALA	HN	52	VAL	HN	3.93	#peak	374
51	TYR	HA	52	VAL	HN	3.28	#peak	375
30	LYS+	HA	52	VAL	HN	4.14	#peak	376
87	PRO	HA	88	VAL	HN	3.08	#peak	377

“alpha.upl, continued”

51	TYR	HB2	52	VAL	HN	4.59	#peak	380
51	TYR	HB3	52	VAL	HN	4.59	#peak	381
52	VAL	HN	52	VAL	HB	3.54	#peak	382
88	VAL	HN	88	VAL	HB	3.36	#peak	383
29	ALA	QB	52	VAL	HN	4.56	#peak	384
52	VAL	HN	52	VAL	QG2	3.26	#peak	385
52	VAL	HN	52	VAL	QG1	3.74	#peak	386
52	VAL	QG1	53	GLY	HN	4.02	#peak	389
52	VAL	QG2	53	GLY	HN	4.69	#peak	390
27	LEU	HB2	53	GLY	HN	4.53	#peak	391
52	VAL	HB	53	GLY	HN	3.9	#peak	392
51	TYR	HB3	53	GLY	HN	4.6	#peak	393
51	TYR	HB2	53	GLY	HN	4.6	#peak	394
27	LEU	HA	53	GLY	HN	4.22	#peak	396
51	TYR	HN	54	GLU-	HN	4.78	#peak	397
51	TYR	QD	54	GLU-	HN	4.66	#peak	398
54	GLU-	HN	54	GLU-	QG	3.89	#peak	402
54	GLU-	HN	54	GLU-	QB	4.13	#peak	403
27	LEU	HB3	54	GLU-	HN	4.76	#peak	404
54	GLU-	HN	57	LEU	QD1	4.05	#peak	405
54	GLU-	HN	104	LEU	QD2	4.3	#peak	406
49	ILE	QG2	55	LEU	HN	3.8	#peak	407
52	VAL	HA	55	LEU	HN	4.62	#peak	410
55	LEU	HB2	56	VAL	HN	4.33	#peak	416
55	LEU	HB3	56	VAL	HN	4.33	#peak	418
55	LEU	QD1	56	VAL	HN	4.36	#peak	419
56	VAL	HN	56	VAL	QG1	3.23	#peak	420
57	LEU	HN	57	LEU	HG	3.89	#peak	423
57	LEU	HN	57	LEU	HB3	3.91	#peak	424
57	LEU	HN	57	LEU	HB2	3.91	#peak	425
55	LEU	HA	58	MET	HN	4.43	#peak	429
56	VAL	HA	58	MET	HN	4.32	#peak	431
58	MET	HN	58	MET	HB2	3.95	#peak	433
58	MET	HN	58	MET	HB3	3.95	#peak	434
57	LEU	HB3	58	MET	HN	4.62	#peak	435
59	SER	HN	62	GLU-	QB	4.05	#peak	436
59	SER	HN	62	GLU-	QG	3.67	#peak	438
58	MET	HG3	59	SER	HN	4.74	#peak	439
57	LEU	HA	59	SER	HN	3.93	#peak	440
59	SER	HN	59	SER	QB	3.1	#peak	441
59	SER	HN	60	GLU-	HN	4.7	#peak	445
59	SER	HA	60	GLU-	HN	3.25	#peak	446
59	SER	QB	60	GLU-	HN	3.69	#peak	447
60	GLU-	HN	60	GLU-	QG	4.57	#peak	449
60	GLU-	HN	60	GLU-	HB2	3.5	#peak	450

“alpha.upl, continued”

60	GLU-	HN	60	GLU-	HB3	3.5	#peak	451
60	GLU-	HN	88	VAL	QG2	4.36	#peak	452
61	GLU-	HN	61	GLU-	QG	3.75	#peak	453
59	SER	QB	61	GLU-	HN	4.24	#peak	455
59	SER	HA	61	GLU-	HN	4.6	#peak	456
59	SER	QB	62	GLU-	HN	4.66	#peak	457
62	GLU-	HN	62	GLU-	QG	3.3	#peak	459
62	GLU-	HN	62	GLU-	QB	3.54	#peak	460
63	LEU	HN	63	LEU	HG	3.48	#peak	461
63	LEU	HN	63	LEU	QB	3.34	#peak	462
62	GLU-	QG	63	LEU	HN	3.9	#peak	464
64	LYS+	HN	74	TYR	QE	4.65	#peak	467
63	LEU	QB	64	LYS+	HN	3.87	#peak	470
64	LYS+	HN	64	LYS+	QB	3.27	#peak	471
64	LYS+	HN	64	LYS+	QD	3.62	#peak	472
64	LYS+	HN	64	LYS+	QG	4.2	#peak	473
64	LYS+	QB	65	GLY	HN	3.97	#peak	475
62	GLU-	HA	65	GLY	HN	4.37	#peak	478
63	LEU	HN	65	GLY	HN	4.94	#peak	480
63	LEU	HA	66	VAL	HN	3.84	#peak	481
64	LYS+	HA	66	VAL	HN	4.47	#peak	482
66	VAL	HN	66	VAL	QG1	3.12	#peak	486
66	VAL	HN	66	VAL	QG2	3.72	#peak	487
47	ILE	QG2	67	LYS+	HN	4.48	#peak	489
67	LYS+	HN	67	LYS+	QG	4.36	#peak	491
66	VAL	HA	67	LYS+	HN	3.02	#peak	492
66	VAL	HN	67	LYS+	HN	4.61	#peak	493
67	LYS+	HA	68	ASN	HN	3.57	#peak	495
67	LYS+	QG	68	ASN	HN	4.88	#peak	496
66	VAL	QG2	69	MET	HN	4.02	#peak	497
69	MET	HN	69	MET	HB2	3.44	#peak	498
96	GLN	HN	96	GLN	HB3	3.97	#peak	499
96	GLN	HN	96	GLN	HB2	3.97	#peak	500
69	MET	HN	69	MET	HB3	3.44	#peak	501
96	GLN	HN	96	GLN	HG3	3.74	#peak	502
96	GLN	HN	96	GLN	HG2	3.74	#peak	503
69	MET	HN	69	MET	HG2	4.43	#peak	504
68	ASN	HB2	69	MET	HN	4.67	#peak	505
67	LYS+	HA	69	MET	HN	4.11	#peak	506
72	LYS+	HN	72	LYS+	QB	3.82	#peak	510
72	LYS+	QB	73	SER	HN	4.03	#peak	513
74	TYR	HN	74	TYR	HB2	3.58	#peak	515
74	TYR	HN	74	TYR	HB3	3.58	#peak	516
71	LYS+	HA	74	TYR	HN	3.93	#peak	518
74	TYR	QD	75	ASP-	HN	3.8	#peak	519

“alpha.upl, continued”

74	TYR	HB3	75	ASP-	HN	3.96	#peak	522
74	TYR	HB2	75	ASP-	HN	3.96	#peak	523
75	ASP-	HN	75	ASP-	HB2	3.67	#peak	524
75	ASP-	HN	75	ASP-	HB3	3.67	#peak	525
76	GLU-	HN	76	GLU-	HG3	4.16	#peak	526
75	ASP-	HB3	76	GLU-	HN	4.11	#peak	527
75	ASP-	HB2	76	GLU-	HN	4.11	#peak	528
74	TYR	HA	77	ILE	HN	4.42	#peak	531
76	GLU-	HB3	77	ILE	HN	4.2	#peak	534
77	ILE	HN	77	ILE	QG2	3.5	#peak	535
77	ILE	QG2	78	ALA	HN	3.63	#peak	536
78	ALA	HN	78	ALA	QB	3.02	#peak	537
77	ILE	QG1	78	ALA	HN	4.39	#peak	538
76	GLU-	HA	78	ALA	HN	4.53	#peak	542
75	ASP-	HA	78	ALA	HN	4.14	#peak	543
80	LYS+	HN	80	LYS+	HG3	4.13	#peak	544
80	LYS+	HN	80	LYS+	HG2	4.13	#peak	545
80	LYS+	HN	80	LYS+	QB	3.54	#peak	547
77	ILE	HA	80	LYS+	HN	4.2	#peak	548
81	LEU	HN	82	ASN	HN	3.61	#peak	550
78	ALA	HA	81	LEU	HN	4.22	#peak	553
77	ILE	HA	81	LEU	HN	4.74	#peak	554
81	LEU	HN	81	LEU	HB2	3.82	#peak	555
80	LYS+	QB	81	LEU	HN	3.9	#peak	556
45	ASP-	HN	45	ASP-	HB3	3.47	#peak	557
55	LEU	QD1	81	LEU	HN	3.56	#peak	559
81	LEU	QD1	82	ASN	HN	4.49	#peak	560
81	LEU	HB3	82	ASN	HN	4.23	#peak	561
81	LEU	HB2	82	ASN	HN	4.23	#peak	562
80	LYS+	HA	83	ASP-	HN	3.82	#peak	568
83	ASP-	HN	83	ASP-	HB3	3.59	#peak	571
84	LEU	HN	84	LEU	QD1	3.76	#peak	572
83	ASP-	HB3	84	LEU	HN	4.28	#peak	575
83	ASP-	HB2	84	LEU	HN	4.28	#peak	576
81	LEU	HA	84	LEU	HN	4.19	#peak	577
82	ASN	HA	84	LEU	HN	4.54	#peak	578
80	LYS+	HA	84	LEU	HN	4.67	#peak	579
84	LEU	HN	86	TYR	HN	4.26	#peak	582
84	LEU	HN	85	GLY	HN	3.7	#peak	583
83	ASP-	HN	85	GLY	HN	4.83	#peak	584
83	ASP-	HA	85	GLY	HN	4.74	#peak	585
84	LEU	HB2	85	GLY	HN	4.62	#peak	589
84	LEU	HB3	85	GLY	HN	4.62	#peak	590
84	LEU	QD1	86	TYR	HN	4.81	#peak	591
84	LEU	HB3	86	TYR	HN	4.76	#peak	592

“alpha.upl, continued”

84	LEU	HB2	86	TYR	HN	4.76	#peak	593
86	TYR	HN	87	PRO	HD3	4.6	#peak	595
88	VAL	HN	89	GLY	HN	4.53	#peak	599
87	PRO	QB	88	VAL	HN	4.24	#peak	600
88	VAL	QG2	90	THR	HN	4.46	#peak	606
90	THR	HN	92	LEU	QD1	4.66	#peak	607
91	GLU-	HN	91	GLU-	HB3	4.09	#peak	608
91	GLU-	HA	92	LEU	HN	2.96	#peak	611
92	LEU	HN	92	LEU	HG	3.67	#peak	613
92	LEU	HN	92	LEU	HB2	4.13	#peak	614
95	GLU-	HN	95	GLU-	QG	4.36	#peak	619
95	GLU-	HN	95	GLU-	HB2	3.81	#peak	620
95	GLU-	HN	95	GLU-	HB3	3.81	#peak	621
94	PRO	QD	95	GLU-	HN	4.45	#peak	622
98	GLU-	HN	98	GLU-	QG	3.67	#peak	624
97	ARG+	HN	98	GLU-	HN	3.64	#peak	627
99	SER	HN	99	SER	HB2	3.5	#peak	629
99	SER	HN	99	SER	HB3	3.5	#peak	630
98	GLU-	QG	99	SER	HN	4.36	#peak	631
98	GLU-	QB	99	SER	HN	3.56	#peak	632
97	ARG+	HA	100	LEU	HN	4.14	#peak	635
100	LEU	HN	100	LEU	HB2	3.43	#peak	636
100	LEU	HN	100	LEU	HG	4.33	#peak	638
57	LEU	QD1	101	LYS+	HN	4.02	#peak	639
101	LYS+	HN	101	LYS+	QD	3.6	#peak	640
101	LYS+	HN	101	LYS+	QB	3.14	#peak	641
99	SER	HA	102	LYS+	HN	4.11	#peak	644
102	LYS+	HN	102	LYS+	HB2	3.27	#peak	647
102	LYS+	HN	102	LYS+	HB3	3.27	#peak	648
102	LYS+	HN	102	LYS+	QD	3.62	#peak	649
102	LYS+	HN	102	LYS+	QG	4	#peak	650
102	LYS+	QG	103	ARG+	HN	4.27	#peak	651
103	ARG+	HN	103	ARG+	QG	4	#peak	652
103	ARG+	HN	103	ARG+	HB2	3.31	#peak	653
100	LEU	HA	103	ARG+	HN	3.78	#peak	655
104	LEU	HN	104	LEU	QB	3.28	#peak	656
57	LEU	QD1	104	LEU	HN	4.42	#peak	657
104	LEU	HN	104	LEU	QD2	4.06	#peak	658
104	LEU	HN	104	LEU	QD1	3.95	#peak	659
104	LEU	QD1	105	GLU-	HN	4.6	#peak	660
104	LEU	QD2	105	GLU-	HN	4.8	#peak	661
104	LEU	HG	105	GLU-	HN	4.15	#peak	662
105	GLU-	HN	105	GLU-	QB	4.1	#peak	663
105	GLU-	HN	105	GLU-	QG	3.67	#peak	664
102	LYS+	HA	105	GLU-	HN	3.44	#peak	666

“alpha.upl, continued”

106	LYS+	HN	106	LYS+	HB3	3.5	#peak	670
106	LYS+	HN	106	LYS+	QD	4.02	#peak	671
106	LYS+	HN	106	LYS+	HG2	3.99	#peak	672
107	LEU	HN	107	LEU	QD1	3.77	#peak	259
107	LEU	HN	107	LEU	QD2	4.19	#peak	675
106	LYS+	HB3	107	LEU	HN	4.2	#peak	678
106	LYS+	HB2	107	LEU	HN	4.2	#peak	679
108	GLU-	HN	109	ASP-	HN	4.43	#peak	89
107	LEU	HB2	108	GLU-	HN	4.32	#peak	688
107	LEU	QD2	108	GLU-	HN	4.71	#peak	690
106	LYS+	HA	109	ASP-	HN	4.78	#peak	691
110	LYS+	HN	110	LYS+	QG	4.65	#peak	696
69	MET	HA	70	GLY	HN	3.53	#peak	703
68	ASN	HA	70	GLY	HN	4.01	#peak	704
69	MET	HB2	70	GLY	HN	4.81	#peak	706
69	MET	HB3	70	GLY	HN	4.81	#peak	707
90	THR	HB	91	GLU-	HN	4.64	#peak	710
53	GLY	HN	104	LEU	QD2	4.85	#peak	713
97	ARG+	HA	99	SER	HN	4.84	#peak	714
54	GLU-	HN	104	LEU	QD1	4.53	#peak	247
79	GLU-	HN	79	GLU-	QB	4.02	#peak	716
27	LEU	HB3	53	GLY	HN	4.78	#peak	717
103	ARG+	HN	103	ARG+	HB3	3.31	#peak	718
91	GLU-	HN	91	GLU-	HB2	4.09	#peak	719
91	GLU-	HN	91	GLU-	HG2	5.05	#peak	720
91	GLU-	HN	91	GLU-	HG3	5.05	#peak	721
60	GLU-	QG	61	GLU-	HN	4.7	#peak	722
17	ASP-	HB2	18	TYR	HN	5.21	#peak	727
17	ASP-	HB3	18	TYR	HN	5.21	#peak	728
66	VAL	HN	69	MET	QE	3.54	#peak	276
35	ASN	HB3	36	LEU	HN	5.08	#peak	730
30	LYS+	HB3	32	GLU-	HN	4.2	#peak	732
36	LEU	HN	37	SER	HN	5.06	#peak	734
34	MET	HB2	36	LEU	HN	4.75	#peak	735
38	ALA	QB	42	ASN	HN	4.8	#peak	736
46	LYS+	HB2	47	ILE	HN	4.53	#peak	737
49	ILE	QG2	51	TYR	HN	3.87	#peak	740
57	LEU	HB2	58	MET	HN	4.62	#peak	742
57	LEU	HG	58	MET	HN	4.88	#peak	743
59	SER	HN	62	GLU-	HN	4.48	#peak	744
83	ASP-	HN	83	ASP-	HB2	3.59	#peak	745
88	VAL	HN	88	VAL	QG1	3.22	#peak	746
88	VAL	QG2	89	GLY	HN	3.79	#peak	749
90	THR	QG2	91	GLU-	HN	4.02	#peak	750
92	LEU	HN	92	LEU	QD2	4.14	#peak	751

“alpha.upl, continued”

98	GLU-	HN	98	GLU-	QB	3.13	#peak	752
105	GLU-	HA	108	GLU-	HN	4.69	#peak	755
38	ALA	QB	42	ASN	HD22	4.3	#peak	358
82	ASN	HA	82	ASN	HD21	4.74	#peak	773
82	ASN	HA	82	ASN	HD22	4.74	#peak	774
79	GLU-	HA	81	LEU	HN	4.53	#peak	780
28	SER	HA	53	GLY	HN	5.2	#peak	783
95	GLU-	HA	98	GLU-	HN	3.99	#peak	784
58	MET	HG2	59	SER	HN	4.74	#peak	786
76	GLU-	HN	76	GLU-	HG2	4.16	#peak	787
55	LEU	QD2	88	VAL	HN	4.65	#peak	788
27	LEU	HA	27	LEU	QD2	3.72	#peak	1
27	LEU	HB2	27	LEU	QD2	4.17	#peak	3
27	LEU	QD2	52	VAL	HB	4.23	#peak	7
26	ASP-	QB	27	LEU	QD2	4.93	#peak	8
27	LEU	QD2	52	VAL	HA	5.34	#peak	10
26	ASP-	HA	27	LEU	QD2	4.97	#peak	12
27	LEU	QD2	86	TYR	HD1	4.51	#peak	13
27	LEU	QD2	29	ALA	HN	4.75	#peak	14
27	LEU	QD2	54	GLU-	HN	5.32	#peak	15
27	LEU	QD2	28	SER	HN	4.89	#peak	16
27	LEU	HN	27	LEU	QD2	4.35	#peak	17
26	ASP-	HN	27	LEU	QD2	5.19	#peak	18
27	LEU	QD2	53	GLY	HN	4.53	#peak	19
27	LEU	QD2	52	VAL	HN	5.46	#peak	20
27	LEU	HB2	27	LEU	QD1	3.72	#peak	22
54	GLU-	QG	104	LEU	QD2	4.03	#peak	24
104	LEU	HA	104	LEU	QD2	3.68	#peak	26
27	LEU	HA	27	LEU	QD1	3.79	#peak	27
27	LEU	QD1	86	TYR	HD1	4.21	#peak	28
51	TYR	QD	104	LEU	QD2	4.15	#peak	29
27	LEU	HN	27	LEU	QD1	3.99	#peak	31
28	SER	QB	107	LEU	QD1	3.74	#peak	254
28	SER	QB	107	LEU	QD2	3.81	#peak	34
29	ALA	HN	29	ALA	QB	2.96	#peak	203
28	SER	HN	29	ALA	QB	4.16	#peak	36
29	ALA	QB	33	SER	QB	3.93	#peak	39
29	ALA	QB	52	VAL	HB	4.03	#peak	40
29	ALA	QB	52	VAL	QG2	3.23	#peak	320
29	ALA	QB	52	VAL	QG1	3.64	#peak	43
31	ILE	QD1	55	LEU	QD1	3.1	#peak	45
31	ILE	QD1	49	ILE	HB	3.64	#peak	48
31	ILE	HA	31	ILE	QD1	3.89	#peak	51
31	ILE	QD1	44	LEU	HA	4.06	#peak	52
31	ILE	QD1	55	LEU	HA	4.31	#peak	53

“alpha.upl, continued”

31	ILE	QD1	50	LYS+	HA	4.24	#peak	54
30	LYS+	HA	31	ILE	QD1	5	#peak	55
31	ILE	QD1	51	TYR	HA	5.33	#peak	56
31	ILE	QD1	41	PHE	HD2	3.75	#peak	57
31	ILE	QD1	41	PHE	HE2	3.82	#peak	58
31	ILE	QD1	55	LEU	HN	4.59	#peak	59
31	ILE	QD1	54	GLU-	HN	4.74	#peak	60
31	ILE	HN	31	ILE	QD1	4.27	#peak	61
31	ILE	QD1	51	TYR	HN	5.01	#peak	62
31	ILE	QD1	32	GLU-	HN	4.61	#peak	63
31	ILE	QG2	31	ILE	HG13	3.83	#peak	66
31	ILE	QG2	31	ILE	QD1	2.86	#peak	44
31	ILE	QG2	31	ILE	HG12	3.83	#peak	71
31	ILE	QG2	52	VAL	HA	3.91	#peak	73
31	ILE	QG2	41	PHE	HD2	4.03	#peak	76
31	ILE	QG2	41	PHE	HE2	4.23	#peak	77
90	THR	HA	90	THR	QG2	3.87	#peak	80
33	SER	HN	33	SER	QB	3.47	#peak	245
90	THR	HA	91	GLU-	HN	3.32	#peak	609
38	ALA	QB	42	ASN	HD21	4.3	#peak	86
38	ALA	HN	38	ALA	QB	3.27	#peak	88
38	ALA	QB	41	PHE	HN	4.32	#peak	286
43	CYS	HB3	66	VAL	QG2	4.06	#peak	90
43	CYS	HB2	66	VAL	QG2	4.06	#peak	91
43	CYS	HB2	44	LEU	HG	4.43	#peak	92
43	CYS	HB3	44	LEU	HG	4.43	#peak	93
43	CYS	HB3	69	MET	QE	3.86	#peak	95
40	CYS	HA	43	CYS	HB2	4.03	#peak	96
40	CYS	HA	43	CYS	HB3	4.03	#peak	97
43	CYS	HB3	44	LEU	HN	4.24	#peak	101
43	CYS	HN	43	CYS	HB3	3.63	#peak	102
42	ASN	HN	43	CYS	HB2	4.91	#peak	104
42	ASN	HN	43	CYS	HB3	4.91	#peak	105
47	ILE	QD1	69	MET	QE	3.59	#peak	108
47	ILE	HB	47	ILE	QD1	3.24	#peak	109
47	ILE	QD1	66	VAL	HA	4.06	#peak	111
44	LEU	HA	47	ILE	QD1	3.79	#peak	112
47	ILE	QD1	66	VAL	HN	4.3	#peak	114
47	ILE	QD1	49	ILE	HN	4.13	#peak	115
47	ILE	QD1	48	GLY	HN	4.31	#peak	116
46	LYS+	HN	47	ILE	QD1	4.5	#peak	117
47	ILE	QD1	67	LYS+	HN	4.63	#peak	118
46	LYS+	HN	47	ILE	QG2	4.3	#peak	121
47	ILE	QG2	49	ILE	HN	4.26	#peak	123
47	ILE	HN	47	ILE	QG2	3.55	#peak	124

“alpha.upl, continued”

47	ILE	HA	47	ILE	QG2	3.11	#peak	125
47	ILE	QG2	58	MET	QE	3.56	#peak	127
47	ILE	QG2	47	ILE	QG1	3.14	#peak	128
47	ILE	QG2	49	ILE	QD1	3.42	#peak	129
47	ILE	HN	48	GLY	HA2	5.03	#peak	135
47	ILE	QD1	49	ILE	QD1	2.75	#peak	138
47	ILE	QG1	49	ILE	QD1	3.49	#peak	140
49	ILE	HB	49	ILE	QD1	3.43	#peak	141
49	ILE	QD1	58	MET	QE	2.64	#peak	272
47	ILE	HB	49	ILE	QD1	3.76	#peak	143
49	ILE	QD1	58	MET	HG2	4.32	#peak	144
49	ILE	QD1	58	MET	HG3	4.32	#peak	145
49	ILE	HA	49	ILE	QD1	3.4	#peak	146
49	ILE	QD1	55	LEU	HA	4.01	#peak	147
48	GLY	HN	49	ILE	QD1	4.13	#peak	149
49	ILE	QD1	50	LYS+	HN	4.23	#peak	150
49	ILE	HA	49	ILE	QG2	3.69	#peak	151
49	ILE	HN	49	ILE	QG2	3.96	#peak	152
52	VAL	HA	52	VAL	QG1	3.57	#peak	430
27	LEU	HA	52	VAL	QG1	3.7	#peak	156
27	LEU	QD2	52	VAL	QG1	3.58	#peak	157
27	LEU	QD2	52	VAL	QG2	4.36	#peak	158
34	MET	HG3	52	VAL	QG2	3.89	#peak	159
34	MET	HG2	52	VAL	QG2	3.89	#peak	160
52	VAL	HA	52	VAL	QG2	3.56	#peak	161
31	ILE	HA	52	VAL	QG2	3.73	#peak	162
29	ALA	HA	52	VAL	QG2	4.23	#peak	163
51	TYR	HA	52	VAL	QG2	4.22	#peak	164
27	LEU	HB2	53	GLY	HA1	5.34	#peak	169
56	VAL	HA	56	VAL	QG2	3.49	#peak	173
27	LEU	QD2	56	VAL	QG2	3.91	#peak	176
56	VAL	QG2	86	TYR	HD1	3.89	#peak	177
56	VAL	HN	56	VAL	QG2	4.01	#peak	178
66	VAL	QG1	67	LYS+	HN	3.84	#peak	184
66	VAL	HA	66	VAL	QG1	3.5	#peak	187
47	ILE	QG1	66	VAL	QG1	3.91	#peak	192
44	LEU	QD2	66	VAL	QG1	3.55	#peak	193
66	VAL	QG2	67	LYS+	HN	3.62	#peak	488
66	VAL	HN	66	VAL	HB	3.17	#peak	484
77	ILE	HA	77	ILE	QD1	3.74	#peak	201
74	TYR	HA	77	ILE	QD1	3.66	#peak	204
77	ILE	HN	77	ILE	QD1	3.96	#peak	205
77	ILE	QD1	78	ALA	HN	4.38	#peak	206
77	ILE	HA	77	ILE	QG2	3.63	#peak	207
74	TYR	HA	77	ILE	QG2	3.55	#peak	208

“alpha.upl, continued”

74	TYR	QD	77	ILE	QG2	4.16	#peak	210
74	TYR	QD	78	ALA	QB	4.32	#peak	214
74	TYR	QE	78	ALA	QB	4.12	#peak	215
78	ALA	QB	88	VAL	HB	3.97	#peak	218
90	THR	HN	90	THR	QG2	3.98	#peak	605
86	TYR	HD2	90	THR	QG2	4.05	#peak	222
90	THR	QG2	92	LEU	HG	4.05	#peak	225
90	THR	QG2	92	LEU	QD2	3.39	#peak	226
100	LEU	HA	100	LEU	QD1	3.34	#peak	227
81	LEU	QD1	86	TYR	HD1	4.1	#peak	228
100	LEU	HN	100	LEU	QD1	3.84	#peak	229
92	LEU	HN	92	LEU	QD1	3.99	#peak	230
57	LEU	QD1	104	LEU	QD1	3.24	#peak	233
104	LEU	QB	104	LEU	QD1	3.43	#peak	235
54	GLU-	QG	104	LEU	QD1	4.36	#peak	236
51	TYR	HB3	104	LEU	QD1	4.47	#peak	237
54	GLU-	QB	104	LEU	QD1	4.5	#peak	238
51	TYR	HB2	104	LEU	QD1	4.47	#peak	239
104	LEU	HA	104	LEU	QD1	3.47	#peak	241
28	SER	QB	104	LEU	QD1	3.87	#peak	242
28	SER	HA	104	LEU	QD1	3.93	#peak	243
51	TYR	QE	104	LEU	QD1	4.47	#peak	244
51	TYR	QD	104	LEU	QD1	3.95	#peak	245
104	LEU	QD1	107	LEU	HN	4.68	#peak	246
53	GLY	HN	104	LEU	QD1	4.54	#peak	248
104	LEU	HA	107	LEU	QD1	3.69	#peak	253
107	LEU	HA	107	LEU	QD1	3.71	#peak	255
28	SER	HA	107	LEU	QD1	3.88	#peak	256
51	TYR	QE	107	LEU	QD1	3.98	#peak	257
51	TYR	QD	107	LEU	QD1	4.12	#peak	258
107	LEU	QD1	108	GLU-	HN	4.28	#peak	260
57	LEU	QD1	101	LYS+	HA	3.56	#peak	262
34	MET	HN	34	MET	QE	3.72	#peak	264
58	MET	QE	63	LEU	HN	3.6	#peak	265
58	MET	QE	66	VAL	HN	3.72	#peak	266
58	MET	QE	74	TYR	QE	4.53	#peak	267
58	MET	QE	63	LEU	HA	2.99	#peak	268
58	MET	QE	62	GLU-	HA	3.34	#peak	269
58	MET	QE	66	VAL	HA	3.82	#peak	270
58	MET	HG3	58	MET	QE	3.23	#peak	271
47	ILE	QD1	58	MET	QE	2.45	#peak	273
58	MET	QE	66	VAL	QG1	2.76	#peak	274
49	ILE	QG1	58	MET	QE	3.39	#peak	275
69	MET	QE	74	TYR	QE	4.21	#peak	277
69	MET	HA	69	MET	QE	3.44	#peak	278

“alpha.upl, continued”

63	LEU	HA	69	MET	QE	3.38	#peak	279
69	MET	QE	74	TYR	HA	3.32	#peak	280
43	CYS	HB2	69	MET	QE	3.86	#peak	281
69	MET	HG3	69	MET	QE	3.27	#peak	282
66	VAL	QG1	69	MET	QE	3.02	#peak	283
44	LEU	QD2	69	MET	QE	2.79	#peak	284
33	SER	HN	34	MET	QE	3.7	#peak	285
31	ILE	HN	34	MET	QE	4.34	#peak	286
31	ILE	HA	34	MET	QE	3.7	#peak	287
34	MET	QE	52	VAL	HA	4.07	#peak	288
34	MET	HG2	34	MET	QE	3.32	#peak	289
34	MET	HG3	34	MET	QE	3.32	#peak	290
27	LEU	HB3	27	LEU	QD1	3.29	#peak	292
63	LEU	QD2	74	TYR	QE	3.73	#peak	293
63	LEU	QD2	74	TYR	QD	3.89	#peak	294
63	LEU	QD1	74	TYR	QD	3.61	#peak	295
63	LEU	QD1	74	TYR	QE	3.51	#peak	296
27	LEU	QD2	100	LEU	QD1	3.87	#peak	297
90	THR	QG2	92	LEU	QD1	3.31	#peak	298
107	LEU	HN	107	LEU	HG	3.98	#peak	305
63	LEU	HA	63	LEU	QD2	3.31	#peak	310
63	LEU	HN	63	LEU	QD2	3.94	#peak	311
63	LEU	HN	63	LEU	QD1	3.93	#peak	312
63	LEU	QD2	64	LYS+	HN	4.26	#peak	313
58	MET	HB3	63	LEU	QD2	3.79	#peak	314
58	MET	HB2	63	LEU	QD2	3.79	#peak	315
63	LEU	QD2	69	MET	QE	3.04	#peak	316
106	LYS+	HN	106	LYS+	HG3	3.99	#peak	324
88	VAL	HN	88	VAL	QG2	3.32	#peak	325
86	TYR	HE2	88	VAL	QG2	4.03	#peak	326
60	GLU-	QG	88	VAL	QG1	3.61	#peak	328
55	LEU	QD1	88	VAL	QG2	3.09	#peak	329
88	VAL	HA	88	VAL	QG1	3.08	#peak	330
100	LEU	QD1	104	LEU	QD1	3.72	#peak	332
59	SER	QB	62	GLU-	QG	5.22	#peak	337
56	VAL	HA	81	LEU	QD1	3.26	#peak	339
88	VAL	HA	90	THR	HN	4.79	#peak	344
31	ILE	HA	34	MET	HG2	4.6	#peak	345
31	ILE	HA	34	MET	HG3	4.6	#peak	346
31	ILE	HA	31	ILE	QG2	3.57	#peak	74
66	VAL	HA	67	LYS+	QB	4.17	#peak	353
66	VAL	HA	66	VAL	QG2	3.62	#peak	194
69	MET	HN	69	MET	HG3	4.43	#peak	357
57	LEU	QD2	97	ARG+	HA	3.72	#peak	361
57	LEU	QD2	97	ARG+	QG	3.33	#peak	362

“alpha.upl, continued”

27	LEU	QD2	56	VAL	QG1	3.02	#peak	4
57	LEU	HN	57	LEU	QD2	3.87	#peak	365
57	LEU	HN	57	LEU	QD1	3.57	#peak	421
55	LEU	HG	77	ILE	QD1	3.68	#peak	367
77	ILE	QG2	77	ILE	QG1	3.18	#peak	368
77	ILE	QG2	78	ALA	QB	3.44	#peak	370
64	LYS+	QD	74	TYR	QE	4.91	#peak	375
64	LYS+	QD	74	TYR	QD	5.24	#peak	376
31	ILE	QG2	51	TYR	HA	4.56	#peak	378
31	ILE	HG13	51	TYR	HA	5.06	#peak	379
31	ILE	HG12	51	TYR	HA	5.06	#peak	380
31	ILE	HB	51	TYR	HA	4.88	#peak	381
30	LYS+	HA	51	TYR	HA	4.43	#peak	421
51	TYR	HA	51	TYR	QD	4.45	#peak	385
31	ILE	HN	51	TYR	HA	3.87	#peak	223
51	TYR	HA	51	TYR	QE	5.34	#peak	388
69	MET	QE	74	TYR	QD	3.66	#peak	389
36	LEU	QD2	80	LYS+	QE	3.7	#peak	390
36	LEU	HA	36	LEU	QD2	3.57	#peak	391
65	GLY	HN	66	VAL	QG1	4.31	#peak	393
28	SER	HN	107	LEU	QD2	3.8	#peak	394
94	PRO	HA	97	ARG+	HN	4.64	#peak	397
86	TYR	HN	87	PRO	HD2	4.6	#peak	596
64	LYS+	HA	64	LYS+	QG	3.93	#peak	408
64	LYS+	HA	64	LYS+	QD	3.86	#peak	409
64	LYS+	HA	74	TYR	QD	4.47	#peak	412
55	LEU	QD1	81	LEU	HA	3.56	#peak	416
30	LYS+	HA	31	ILE	HN	3.26	#peak	222
30	LYS+	HA	51	TYR	QD	4.68	#peak	419
30	LYS+	HA	51	TYR	QE	4.86	#peak	420
84	LEU	HB2	86	TYR	HD1	5.12	#peak	423
84	LEU	HB2	84	LEU	QD2	3.89	#peak	425
84	LEU	HB3	84	LEU	QD2	3.89	#peak	428
31	ILE	QD1	52	VAL	HA	4.62	#peak	50
104	LEU	HA	107	LEU	HN	4.29	#peak	680
104	LEU	HA	107	LEU	QD2	4.39	#peak	440
49	ILE	QG2	55	LEU	HA	3.61	#peak	443
75	ASP-	HA	78	ALA	QB	3.49	#peak	216
107	LEU	HA	107	LEU	QD2	2.87	#peak	322
50	LYS+	HA	50	LYS+	QG	3.5	#peak	449
77	ILE	HA	80	LYS+	QB	4.2	#peak	452
27	LEU	HB3	27	LEU	QD2	3.97	#peak	6
107	LEU	HB2	107	LEU	QD1	3.93	#peak	460
107	LEU	HB3	107	LEU	QD1	3.93	#peak	250
51	TYR	QE	107	LEU	HB2	5.33	#peak	462

“alpha.upl, continued”

107	LEU	HB3	108	GLU-	HN	4.32	#peak	689
107	LEU	HB2	107	LEU	QD2	3.78	#peak	467
107	LEU	HB3	107	LEU	QD2	3.78	#peak	468
27	LEU	HB2	53	GLY	HA2	5.34	#peak	474
55	LEU	QD1	81	LEU	HB2	4.27	#peak	476
55	LEU	QD1	81	LEU	HB3	4.27	#peak	477
84	LEU	HB3	86	TYR	HD1	5.12	#peak	481
84	LEU	HN	84	LEU	HB2	3.74	#peak	574
84	LEU	HN	84	LEU	HB3	3.74	#peak	573
44	LEU	HN	44	LEU	QD1	3.99	#peak	312
44	LEU	HN	44	LEU	QD2	4.14	#peak	781
44	LEU	HA	44	LEU	QD2	3.58	#peak	486
92	LEU	HN	92	LEU	HB3	4.13	#peak	615
92	LEU	HB2	92	LEU	QD2	4.07	#peak	490
92	LEU	HB3	92	LEU	QD2	4.07	#peak	491
55	LEU	HN	55	LEU	HB3	4.05	#peak	408
55	LEU	HB2	55	LEU	QD1	4.08	#peak	495
49	ILE	QG2	55	LEU	HB2	4.29	#peak	496
55	LEU	HB3	55	LEU	QD1	4.08	#peak	306
55	LEU	HN	55	LEU	HB2	4.05	#peak	409
100	LEU	HN	100	LEU	HB3	3.43	#peak	637
63	LEU	QB	63	LEU	QD1	3.64	#peak	504
57	LEU	QD1	100	LEU	HB3	3.92	#peak	507
57	LEU	QD1	100	LEU	HB2	3.92	#peak	508
51	TYR	HA	104	LEU	QD1	5.05	#peak	509
50	LYS+	QD	51	TYR	QE	4.87	#peak	511
50	LYS+	QD	51	TYR	QD	5.07	#peak	512
50	LYS+	QD	51	TYR	HN	5.02	#peak	513
50	LYS+	HA	50	LYS+	QD	4.05	#peak	515
106	LYS+	HN	106	LYS+	HB2	3.5	#peak	669
51	TYR	QE	107	LEU	HB3	5.33	#peak	519
47	ILE	HB	49	ILE	HN	4.88	#peak	526
47	ILE	HB	48	GLY	HN	4.78	#peak	527
77	ILE	HB	78	ALA	HN	3.59	#peak	539
77	ILE	HN	77	ILE	HB	3.36	#peak	126
74	TYR	HA	77	ILE	HB	4.38	#peak	530
77	ILE	HB	77	ILE	QD1	3.18	#peak	532
31	ILE	HB	31	ILE	QD1	3.4	#peak	49
31	ILE	HB	41	PHE	HD2	5.14	#peak	537
31	ILE	HB	41	PHE	HE2	5.15	#peak	538
39	ARG+	HA	42	ASN	HB2	3.88	#peak	540
39	ARG+	HA	42	ASN	HB3	3.88	#peak	542
42	ASN	HN	42	ASN	HB2	3.55	#peak	297
42	ASN	HN	42	ASN	HB3	3.55	#peak	299
54	GLU-	QG	57	LEU	QD1	4.02	#peak	554

“alpha.upl, continued”

58	MET	HG2	58	MET	QE	3.23	#peak	556
66	VAL	QG2	69	MET	QE	3.22	#peak	566
29	ALA	HA	30	LYS+	HN	2.96	#peak	211
32	GLU-	HA	41	PHE	HE2	4.26	#peak	576
34	MET	QE	52	VAL	HN	5.08	#peak	580
34	MET	QE	55	LEU	QD1	2.4	#peak	581
31	ILE	QG2	34	MET	QE	2.64	#peak	582
40	CYS	HA	44	LEU	QD1	3.73	#peak	583
52	VAL	QG1	56	VAL	HB	3.37	#peak	586
51	TYR	HB2	107	LEU	QD1	5.03	#peak	588
51	TYR	HB3	107	LEU	QD1	5.03	#peak	589
56	VAL	HN	56	VAL	HB	3.47	#peak	417
64	LYS+	HA	74	TYR	QE	4.69	#peak	593
97	ARG+	HA	97	ARG+	QD	4.69	#peak	598
101	LYS+	HA	104	LEU	HN	3.86	#peak	754
103	ARG+	HA	106	LYS+	HN	3.97	#peak	602
27	LEU	QD2	86	TYR	HE1	4.59	#peak	604
27	LEU	QD1	86	TYR	HE1	4.28	#peak	605
50	LYS+	HB2	51	TYR	QE	4.21	#peak	606
50	LYS+	HB3	51	TYR	QE	4.21	#peak	607
56	VAL	QG2	86	TYR	HE1	4.28	#peak	608
72	LYS+	HA	75	ASP-	HN	4.29	#peak	610
69	MET	HG2	69	MET	QE	3.27	#peak	612
31	ILE	QG2	52	VAL	QG2	3.11	#peak	613
31	ILE	QD1	49	ILE	QG2	2.86	#peak	632
58	MET	QE	69	MET	QE	2.99	#peak	615
91	GLU-	HA	94	PRO	QD	4.55	#peak	626
78	ALA	QB	88	VAL	QG1	3.09	#peak	628
92	LEU	HA	92	LEU	QD2	3.62	#peak	630
36	LEU	HA	36	LEU	QQD	3.63	#peak	631
44	LEU	QD2	77	ILE	QD1	2.87	#peak	633
74	TYR	QE	77	ILE	QG2	4.35	#peak	635
55	LEU	QD2	88	VAL	QG1	3.16	#peak	636
82	ASN	HN	82	ASN	HB3	3.45	#peak	564
82	ASN	HN	82	ASN	HB2	3.45	#peak	563
90	THR	HB	92	LEU	QD1	4.63	#peak	642
90	THR	HN	90	THR	HB	3.3	#peak	115
51	TYR	HA	52	VAL	HB	5.31	#peak	648
34	MET	QE	52	VAL	QG2	3.17	#peak	649
35	ASN	HA	36	LEU	HN	3.51	#peak	261
68	ASN	HB3	69	MET	HN	4.67	#peak	659
32	GLU-	HA	41	PHE	HD2	4.57	#peak	663
96	GLN	HA	96	GLN	HG3	4.19	#peak	666
96	GLN	HA	96	GLN	HG2	4.19	#peak	667
47	ILE	QG2	47	ILE	QD1	2.69	#peak	669

“alpha.upl, continued”

100	LEU	HG	101	LYS+	HA	3.81	#peak	671
60	GLU-	HA	88	VAL	QG1	4.01	#peak	677
64	LYS+	QE	74	TYR	QD	4.68	#peak	678
64	LYS+	QE	74	TYR	QE	4.25	#peak	679
24	ALA	QB	56	VAL	QG1	3.74	#peak	683
24	ALA	QB	27	LEU	QD1	3.82	#peak	684
24	ALA	HN	24	ALA	QB	3.45	#peak	176
24	ALA	HA	27	LEU	QD1	4.5	#peak	687
41	PHE	HE1	45	ASP-	HB3	5.5	#peak	691
41	PHE	HE1	45	ASP-	HB2	5.5	#peak	692
36	LEU	QD2	76	GLU-	HG3	4.23	#peak	694
61	GLU-	HA	61	GLU-	QG	3.18	#peak	696
57	LEU	QD1	104	LEU	QD2	2.73	#peak	701
47	ILE	QD1	66	VAL	QG1	3.04	#peak	702
55	LEU	QD2	81	LEU	QD1	2.87	#peak	703
44	LEU	QD1	55	LEU	QD1	2.71	#peak	706
47	ILE	HA	47	ILE	QG1	4.25	#peak	708
94	PRO	QD	95	GLU-	QG	4.4	#peak	714
64	LYS+	QB	64	LYS+	QD	3.65	#peak	373
64	LYS+	QG	64	LYS+	QE	3.62	#peak	723
64	LYS+	QB	64	LYS+	QE	4.01	#peak	725
49	ILE	QG2	55	LEU	HB3	4.29	#peak	726
27	LEU	QD1	56	VAL	QG2	2.91	#peak	727
56	VAL	QG2	84	LEU	QD2	3.36	#peak	728
63	LEU	QD1	88	VAL	QG1	3.1	#peak	729
55	LEU	QD1	88	VAL	QG1	3.25	#peak	730
81	LEU	QD1	86	TYR	HE1	4.13	#peak	731
36	LEU	QD2	76	GLU-	HB2	4.66	#peak	735
76	GLU-	HN	76	GLU-	HB3	3.52	#peak	119
76	GLU-	HB2	77	ILE	HN	4.2	#peak	533
36	LEU	QD2	76	GLU-	HB3	4.66	#peak	739
36	LEU	QD2	76	GLU-	HG2	4.23	#peak	740
57	LEU	QD2	97	ARG+	QD	3.7	#peak	741
57	LEU	HA	97	ARG+	QG	4.95	#peak	742
97	ARG+	QG	98	GLU-	HN	5.04	#peak	743
26	ASP-	HA	29	ALA	QB	3.94	#peak	745
31	ILE	QG2	36	LEU	QD1	2.97	#peak	746
44	LEU	QD2	47	ILE	QD1	2.86	#peak	748
27	LEU	QD2	84	LEU	QD1	4.29	#peak	749
63	LEU	HA	66	VAL	QG1	3.56	#peak	186
60	GLU-	HA	60	GLU-	QG	3.98	#peak	751
55	LEU	QD2	63	LEU	QD1	3.02	#peak	752
63	LEU	QD1	88	VAL	QG2	3.03	#peak	753
9	ILE	HN	9	ILE	QG1	3.84	#peak	144
27	LEU	HB2	53	GLY	QA	4.66	#peak	169

“alpha.upl, continued”

27	LEU	HB3	53	GLY	QA	4.98	#peak	168
27	LEU	QD1	53	GLY	QA	3.88	#peak	25
27	LEU	QD2	53	GLY	QA	4.09	#peak	11
27	LEU	QD2	86	TYR	QB	5.03	#peak	9
29	ALA	HA	30	LYS+	QB	4.18	#peak	572
29	ALA	QB	30	LYS+	QB	3.79	#peak	41
30	LYS+	HA	31	ILE	QG1	4.36	#peak	422
31	ILE	HN	31	ILE	QG1	3.34	#peak	217
31	ILE	HA	34	MET	QG	3.97	#peak	345
31	ILE	QG2	31	ILE	QG1	3.28	#peak	71
31	ILE	QG1	32	GLU-	HN	3.76	#peak	236
31	ILE	QG1	49	ILE	QG2	3.69	#peak	611
31	ILE	QG1	50	LYS+	HA	5.04	#peak	69
31	ILE	QG1	51	TYR	HA	4.33	#peak	379
31	ILE	QG1	52	VAL	HA	4.96	#peak	433
33	SER	HN	34	MET	QG	4.55	#peak	242
34	MET	HN	34	MET	QB	3.37	#peak	736
34	MET	HN	34	MET	QG	3.43	#peak	253
34	MET	QB	36	LEU	HN	4.07	#peak	263
34	MET	QG	34	MET	QE	2.89	#peak	290
35	ASN	QB	35	ASN	QD2	3.22	#peak	765
35	ASN	QB	36	LEU	HN	4.22	#peak	262
36	LEU	HN	36	LEU	QB	3.47	#peak	264
36	LEU	QB	37	SER	HN	3.69	#peak	269
36	LEU	QB	41	PHE	HN	4.3	#peak	287
36	LEU	QD2	76	GLU-	QG	3.67	#peak	694
38	ALA	HA	41	PHE	QB	3.89	#peak	713
38	ALA	QB	42	ASN	QB	3.86	#peak	85
39	ARG+	HA	42	ASN	QB	3.39	#peak	540
40	CYS	HA	43	CYS	QB	3.48	#peak	97
41	PHE	HN	41	PHE	QB	3.06	#peak	288
41	PHE	QB	42	ASN	HN	3.97	#peak	298
41	PHE	HD1	45	ASP-	QB	5.35	#peak	693
41	PHE	HE1	45	ASP-	QB	4.63	#peak	691
42	ASN	HN	42	ASN	QB	2.88	#peak	297
42	ASN	HN	43	CYS	QB	4.3	#peak	105
43	CYS	HN	43	CYS	QB	3.17	#peak	302
43	CYS	QB	66	VAL	QG2	3.48	#peak	90
43	CYS	QB	69	MET	QE	3.33	#peak	281
45	ASP-	HN	45	ASP-	QB	2.99	#peak	557
45	ASP-	QB	46	LYS+	HN	3.67	#peak	326
46	LYS+	HN	46	LYS+	QB	2.95	#peak	327
46	LYS+	QB	47	ILE	HN	3.93	#peak	332
47	ILE	HN	48	GLY	QA	4.32	#peak	334
47	ILE	QD1	58	MET	QG	4.17	#peak	110

“alpha.upl, continued”

49	ILE	HN	49	ILE	QG1	3.65	#peak	347
49	ILE	QG2	55	LEU	QB	3.71	#peak	726
49	ILE	QD1	58	MET	QG	3.74	#peak	144
50	LYS+	QB	51	TYR	QD	4.32	#peak	716
50	LYS+	QB	51	TYR	QE	3.66	#peak	606
51	TYR	QB	52	VAL	HN	3.78	#peak	380
51	TYR	QB	54	GLU-	HN	4.57	#peak	401
51	TYR	QB	107	LEU	QD1	4.42	#peak	588
52	VAL	QG1	53	GLY	QA	4.03	#peak	587
53	GLY	QA	56	VAL	HN	4.11	#peak	415
53	GLY	QA	56	VAL	HB	4.92	#peak	167
53	GLY	QA	56	VAL	QG1	3.52	#peak	261
53	GLY	QA	104	LEU	QD1	3.97	#peak	240
53	GLY	QA	104	LEU	QD2	4.38	#peak	172
53	GLY	QA	107	LEU	QD1	3.86	#peak	252
55	LEU	HN	55	LEU	QB	3.45	#peak	409
55	LEU	QB	55	LEU	QD1	3.23	#peak	306
55	LEU	QB	56	VAL	HN	3.73	#peak	416
56	VAL	QG2	86	TYR	QB	4.16	#peak	174
57	LEU	HN	57	LEU	QB	3.07	#peak	424
57	LEU	QB	58	MET	HN	3.81	#peak	435
58	MET	HN	58	MET	QB	3.19	#peak	434
58	MET	HN	58	MET	QG	3.52	#peak	432
58	MET	QB	59	SER	HN	3.94	#peak	437
58	MET	QB	63	LEU	HN	4.29	#peak	463
58	MET	QB	63	LEU	QD2	3.24	#peak	314
58	MET	QG	58	MET	QE	2.8	#peak	556
58	MET	QG	63	LEU	QD2	3.74	#peak	555
58	MET	QG	66	VAL	QG1	4.26	#peak	188
64	LYS+	HA	69	MET	QG	4.46	#peak	411
66	VAL	HN	69	MET	QB	4.25	#peak	485
69	MET	HN	69	MET	QB	3	#peak	501
69	MET	HN	69	MET	QG	3.87	#peak	357
69	MET	QG	69	MET	QE	2.83	#peak	612
69	MET	QG	70	GLY	HN	4.69	#peak	616
69	MET	QG	74	TYR	HN	4.47	#peak	514
69	MET	QG	74	TYR	QD	4.63	#peak	200
74	TYR	QB	75	ASP-	HN	3.39	#peak	522
74	TYR	QB	77	ILE	QG2	4.02	#peak	209
74	TYR	QB	77	ILE	QD1	3.9	#peak	202
74	TYR	QB	78	ALA	QB	4.55	#peak	359
75	ASP-	HN	75	ASP-	QB	3.15	#peak	524
75	ASP-	QB	76	GLU-	HN	3.57	#peak	528
76	GLU-	HN	76	GLU-	QG	3.63	#peak	787
76	GLU-	HA	76	GLU-	QG	3.58	#peak	695

“alpha.upl, continued”

76	GLU-	QB	77	ILE	HN	3.63	#peak	533
78	ALA	QB	81	LEU	QB	3.93	#peak	219
78	ALA	QB	82	ASN	QB	4.45	#peak	360
80	LYS+	HN	80	LYS+	QG	3.43	#peak	545
80	LYS+	HN	80	LYS+	QD	3.74	#peak	546
81	LEU	HN	81	LEU	QB	3.01	#peak	558
81	LEU	QB	82	ASN	HN	3.58	#peak	561
82	ASN	HN	82	ASN	QB	3.01	#peak	563
82	ASN	HA	82	ASN	QD2	4.07	#peak	773
82	ASN	QB	82	ASN	QD2	3.18	#peak	756
82	ASN	QB	83	ASP-	HN	3.51	#peak	570
83	ASP-	HN	83	ASP-	QB	2.94	#peak	745
83	ASP-	QB	84	LEU	HN	3.71	#peak	576
84	LEU	HN	84	LEU	QB	2.96	#peak	574
84	LEU	QB	84	LEU	QD1	2.82	#peak	308
84	LEU	QB	84	LEU	QD2	3.37	#peak	428
84	LEU	QB	85	GLY	HN	3.97	#peak	589
84	LEU	QB	86	TYR	HN	3.96	#peak	593
84	LEU	QB	86	TYR	HD1	4.46	#peak	481
86	TYR	HN	87	PRO	QD	4.03	#peak	595
87	PRO	QD	90	THR	QG2	4.48	#peak	224
91	GLU-	HN	91	GLU-	QB	3.59	#peak	608
91	GLU-	HN	91	GLU-	QG	4.39	#peak	720
91	GLU-	QB	92	LEU	HN	4.29	#peak	612
92	LEU	HN	92	LEU	QB	3.46	#peak	614
92	LEU	QB	93	SER	HN	3.7	#peak	363
93	SER	HN	94	PRO	QB	3.64	#peak	365
94	PRO	QB	95	GLU-	HN	3.95	#peak	618
94	PRO	QB	97	ARG+	HN	3.76	#peak	317
94	PRO	QB	97	ARG+	QB	3.47	#peak	398
96	GLN	HN	96	GLN	QB	3.4	#peak	499
96	GLN	HN	96	GLN	QG	3.08	#peak	502
96	GLN	QG	97	ARG+	HN	4.85	#peak	665
99	SER	HN	99	SER	QB	3.02	#peak	629
99	SER	QB	100	LEU	HN	3.85	#peak	634
100	LEU	HN	100	LEU	QB	3	#peak	636
102	LYS+	HN	102	LYS+	QB	2.82	#peak	647
104	LEU	HA	107	LEU	QB	4	#peak	442
106	LYS+	HN	106	LYS+	QB	3	#peak	670
106	LYS+	HN	106	LYS+	QG	3.37	#peak	324
106	LYS+	HA	106	LYS+	QG	3.03	#peak	323
107	LEU	HN	107	LEU	QB	2.92	#peak	677
107	LEU	QB	107	LEU	QD1	3.25	#peak	250
107	LEU	QB	108	GLU-	HN	3.67	#peak	688
108	GLU-	HN	108	GLU-	QB	3.69	#peak	687

CYANA Output File

alpha.ovw – contains target functions, violation statistics for 50 best structures out of 500

Structural statistics:

str	target	upper limits			lower limits			van der Waals			torsion angles		
	function	#	rms	max	#	rms	max	#	sum	max	#	rms	max
1	1.19	2	0.0169	0.20	0	0.0097	0.05	1	3.4	0.20	0	0.4544	2.40
2	1.24	2	0.0171	0.24	0	0.0098	0.04	1	3.5	0.20	0	0.5605	3.08
3	1.26	4	0.0179	0.26	0	0.0091	0.04	1	3.2	0.20	0	0.4279	2.11
4	1.34	1	0.0170	0.26	0	0.0144	0.09	0	3.9	0.20	0	0.8409	4.83
5	1.36	2	0.0174	0.24	0	0.0098	0.04	1	3.9	0.20	0	0.7289	3.99
6	1.37	2	0.0178	0.22	0	0.0085	0.03	1	3.9	0.20	0	0.5368	2.67
7	1.40	1	0.0170	0.21	0	0.0113	0.05	1	4.1	0.20	0	0.7846	4.40
8	1.41	2	0.0179	0.21	0	0.0107	0.05	0	4.1	0.20	0	0.7470	4.24
9	1.41	2	0.0179	0.28	0	0.0074	0.03	0	3.9	0.20	0	0.7413	4.72
10	1.45	2	0.0180	0.21	0	0.0131	0.07	1	4.3	0.20	2	0.8456	5.62
11	1.46	2	0.0184	0.28	0	0.0113	0.05	0	4.1	0.20	0	0.7207	4.69
12	1.47	2	0.0181	0.23	0	0.0126	0.06	0	4.4	0.19	0	0.7108	3.84
13	1.49	2	0.0187	0.22	0	0.0081	0.04	0	4.0	0.19	0	0.8171	4.00
14	1.49	2	0.0184	0.21	0	0.0100	0.05	1	4.3	0.20	0	0.8070	4.78
15	1.49	2	0.0177	0.22	0	0.0081	0.04	1	4.5	0.20	0	0.7473	4.30
16	1.50	1	0.0179	0.22	0	0.0119	0.07	0	4.4	0.19	1	0.8666	5.21
17	1.51	3	0.0198	0.24	0	0.0112	0.06	0	3.8	0.19	0	0.5446	3.22
18	1.51	0	0.0178	0.20	0	0.0086	0.05	0	4.1	0.20	0	0.7190	3.79
19	1.53	2	0.0185	0.23	0	0.0098	0.05	1	4.4	0.20	0	0.6449	3.08
20	1.53	1	0.0177	0.22	0	0.0103	0.04	0	4.8	0.19	0	0.7456	4.52
21	1.53	3	0.0184	0.22	0	0.0082	0.04	1	4.2	0.20	0	0.5880	3.10
22	1.53	1	0.0180	0.26	0	0.0091	0.05	1	4.1	0.20	0	0.7097	4.41
23	1.54	1	0.0181	0.22	0	0.0110	0.05	1	4.7	0.20	0	0.7200	4.24
24	1.54	4	0.0199	0.27	0	0.0107	0.05	1	3.9	0.20	0	0.7442	3.38
25	1.55	1	0.0177	0.22	0	0.0095	0.04	0	4.7	0.19	0	0.7486	4.68
26	1.56	2	0.0183	0.25	0	0.0095	0.06	0	4.3	0.19	0	0.7841	4.62
27	1.57	4	0.0191	0.30	0	0.0158	0.08	1	4.6	0.20	0	0.6105	3.45
28	1.57	4	0.0184	0.24	0	0.0108	0.05	0	4.6	0.19	0	0.5795	2.90
29	1.57	1	0.0186	0.30	0	0.0096	0.04	0	4.6	0.19	0	0.6238	3.52
30	1.58	2	0.0192	0.24	0	0.0117	0.06	1	4.4	0.20	0	0.7689	4.24
31	1.58	3	0.0191	0.28	0	0.0077	0.04	0	4.4	0.20	0	0.7704	3.89
32	1.58	3	0.0190	0.34	0	0.0105	0.05	0	4.0	0.19	0	0.7249	3.99
33	1.59	3	0.0182	0.27	0	0.0120	0.05	0	4.6	0.19	0	0.7967	4.68
34	1.59	2	0.0190	0.26	0	0.0147	0.08	0	4.5	0.19	0	0.8096	4.51
35	1.59	4	0.0196	0.34	0	0.0113	0.05	0	4.0	0.20	0	0.7569	4.63
36	1.59	1	0.0184	0.26	0	0.0100	0.06	1	5.0	0.21	0	0.6827	3.88
37	1.59	3	0.0195	0.26	0	0.0090	0.03	0	4.2	0.20	2	0.9318	5.29
38	1.59	2	0.0184	0.20	0	0.0136	0.07	1	4.2	0.20	0	0.7546	3.99
39	1.60	4	0.0195	0.31	0	0.0087	0.04	1	4.1	0.20	0	0.7841	3.97
40	1.60	3	0.0188	0.23	0	0.0111	0.05	2	4.4	0.21	0	0.8273	4.74
41	1.60	2	0.0192	0.24	0	0.0104	0.05	1	4.4	0.20	0	0.7212	4.21
42	1.61	2	0.0191	0.38	0	0.0123	0.06	0	4.4	0.20	0	0.6496	3.52
43	1.61	2	0.0193	0.24	0	0.0073	0.03	1	4.5	0.20	0	0.7023	3.55
44	1.61	3	0.0191	0.24	0	0.0105	0.06	0	4.6	0.20	0	0.6789	3.63
45	1.61	1	0.0192	0.24	0	0.0140	0.08	0	4.3	0.18	0	0.6552	3.07
46	1.62	1	0.0176	0.20	0	0.0112	0.05	2	4.8	0.21	0	0.7725	4.88
47	1.62	5	0.0209	0.38	0	0.0087	0.04	0	3.7	0.20	0	0.4967	2.78

```
 48   1.63  3 0.0191  0.29   0 0.0124  0.05   1   4.8  0.20   0 0.6695  3.94
 49   1.63  2 0.0181  0.21   0 0.0135  0.06   1   4.7  0.20   2 0.9096  5.34
 50   1.63  3 0.0201  0.24   0 0.0134  0.06   1   4.1  0.21   0 0.8678  4.99

Ave   1.52  2 0.0185  0.25   0 0.0107  0.05   1   4.3  0.20   0 0.7166  4.03
+/-  0.10  1 0.0008  0.04   0 0.0020  0.01   1   0.4  0.01   0 0.1094  0.79
Min   1.19  0 0.0169  0.20   0 0.0073  0.03   0   3.2  0.18   0 0.4279  2.11
Max   1.63  5 0.0209  0.38   0 0.0158  0.09   2   5.0  0.21   2 0.9318  5.62
Cut              0.20           0.20           0.20           5.00
```

Constraints violated in 16 or more structures:

```
                                  # mean max.1   5   10   15   20   25   30   35   40   45   50
Upper HN ALA 29 - HN LYS 30 4.33 21 .20 .25   + ++  +   +++ + +    *  ++ +  ++  +    + +   +++
VdW  CG2 ILE 77 - C  ILE 77 2.78 16 .20 .21   +      +   +      + ++  +  +       ++++ +  +  +*
```

1 violated distance constraint.
0 violated angle constraints.

```
Hydrogen bonds:                    # 1   5   10   15   20   25   30   35   40   45   50
HN    SER   33 - O     LYS+  30  38 ++++++++++ + ++  ++ + ++++ +++++  +++++++++ ++  ++
HN    MET   34 - O     ILE   31  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    SER   37 - OE1   GLU-  76  36 ++++++++++++ + + ++ + + +++++ ++   ++ ++++ + ++ ++
HN    SER   37 - OE2   GLU-  76  41 ++  +++ ++ +++++++++ + ++++++++++++++++++++  ++
HN    ASN   42 - O     ALA   38  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    CYS   43 - O     ARG+  39  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    LEU   44 - O     CYS   40  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    ASP-  45 - O     PHE   41  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    ILE   47 - O     CYS   43  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    GLY   48 - O     LEU   44  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    VAL   52 - O     ALA   29  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    LEU   55 - O     TYR   51  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    VAL   56 - O     VAL   52  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    LEU   57 - O     GLY   53  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    SER   59 - O     LEU   57  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    LEU   63 - O     SER   59  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    LYS+  64 - O     GLU-  60  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    GLY   65 - O     LEU   63  48 ++++++++++++++++++++++++++++++++++++++++++++++++ ++
HN    MET   69 - O     LYS+  67  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    TYR   74 - O     GLY   70  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    ASP-  75 - O     LYS+  71  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    GLU-  76 - O     LYS+  72  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    ILE   77 - O     SER   73  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    ALA   78 - O     TYR   74  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    GLU-  79 - O     ASP-  75  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    LYS+  80 - O     GLU-  76  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    LEU   81 - O     ILE   77  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    ASN   82 - O     ALA   78  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    ASP-  83 - O     GLU-  79  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    LEU   84 - O     LYS+  80  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    GLU-  95 - O     SER   93  38 +++ ++ +++++ +  +  ++++ ++++++ + ++++ + ++++++++++
HN    GLU-  98 - O     GLU-  95  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    LEU  100 - O     GLN   96  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    LYS+ 101 - O     ARG+  97  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    LYS+ 102 - O     GLU-  98  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    ARG+ 103 - O     SER   99  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    LEU  104 - O     LEU  100  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    GLU- 105 - O     LYS+ 101  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
HN    LYS+ 106 - O     LYS+ 102  50 ++++++++++++++++++++++++++++++++++++++++++++++++++++
```

39 hydrogen bonds.

RMSDs for residues 24..107:
Average backbone RMSD to mean : 0.32 +/- 0.06 A (0.19..0.50 A; 50 structures)
Average heavy atom RMSD to mean : 0.99 +/- 0.08 A (0.82..1.16 A; 50 structures)

AMBER Input File

min.in – contains parameters for running the sander program of the AMBER 9 package

```
#sander minimize structure
&cntrl
 imin=1, maxcyc=3000,
 cut=300.0, igb=2, saltcon=0.2, gbsa=1,
 ntpr=100, ntx=1,  ntb=0, nmropt=1,
&end
&wt type='REST', istep1=0, istep2=3000, value1=1, value2=1, &end
&wt type='END' &end
DISANG=./dist.in
LISTOUT = POUT
```

REFERENCES

Alm, R. A., L. S. Ling, D. T. Moir, B. L. King, E. D. Brown, P. C. Doig, D. R. Smith, B. Noonan, B. C. Guild, B. L. deJonge, G. Carmel, P. J. Tummino, A. Caruso, M. Uria-Nickelsen, D. M. Mills, C. Ives, R. Gibson, D. Merberg, S. D. Mills, Q. Jiang, D. E. Taylor, G. F. Vovis and T. J. Trust (1999). "Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori." Nature **397**(6715): 176-80.

Amieva, M. R. and E. M. El-Omar (2008). "Host-bacterial interactions in Helicobacter pylori infection." Gastroenterology **134**(1): 306-23.

Andersen-Nissen, E., K. D. Smith, K. L. Strobe, S. L. Barrett, B. T. Cookson, S. M. Logan and A. Aderem (2005). "Evasion of Toll-like receptor 5 by flagellated bacteria." Proc Natl Acad Sci U S A **102**(26): 9247-52.

Andersen, L. P. (2007). "Colonization and infection by Helicobacter pylori in humans." Helicobacter **12 Suppl 2**: 12-5.

Ang, S., C. Z. Lee, K. Peck, M. Sindici, U. Matrubutham, M. A. Gleeson and J. T. Wang (2001). "Acid-induced gene expression in Helicobacter pylori: study in genomic scale by microarray." Infect Immun **69**(3): 1679-86.

Armache, K. J., S. Mitterweger, A. Meinhart and P. Cramer (2005). "Structures of complete RNA polymerase II and its subcomplex, Rpb4/7." J Biol Chem **280**(8): 7131-4.

Artsimovitch, I. and D. G. Vassylyev (2006). "Is it easy to stop RNA polymerase?" Cell Cycle **5**(4): 399-404.

Baar, C., M. Eppinger, G. Raddatz, J. Simon, C. Lanz, O. Klimmek, R. Nandakumar, R. Gross, A. Rosinus, H. Keller, P. Jagtap, B. Linke, F. Meyer, H. Lederer and S. C. Schuster (2003). "Complete genome sequence and analysis of Wolinella succinogenes." Proc Natl Acad Sci U S A **100**(20): 11690-5.

Baynham, P. J., A. L. Brown, L. L. Hall and D. J. Wozniak (1999). "Pseudomonas aeruginosa AlgZ, a ribbon-helix-helix DNA-binding protein, is essential for alginate synthesis and algD transcriptional activation." Mol Microbiol **33**(5): 1069-80.

Beier, D. and R. Frank (2000). "Molecular characterization of two-component systems of Helicobacter pylori." J Bacteriol **182**(8): 2068-76.

Beier, D., G. Spohn, R. Rappuoli and V. Scarlato (1998). "Functional analysis of the Helicobacter pylori principal sigma subunit of RNA polymerase reveals that the spacer region is important for efficient transcription." Mol Microbiol **30**(1): 121-34.

Benoff, B., H. Yang, C. L. Lawson, G. Parkinson, J. Liu, E. Blatter, Y. W. Ebright, H. M. Berman and R. H. Ebright (2002). "Structural basis of transcription activation: the CAP-alpha CTD-DNA complex." Science **297**(5586): 1562-6.

Berezovski, M., A. Drabovich, S. M. Krylova, M. Musheev, V. Okhonin, A. Petrov and S. N. Krylov (2005). "Nonequilibrium capillary electrophoresis of equilibrium mixtures: a universal tool for development of aptamers." J Am Chem Soc **127**(9): 3165-71.

Berezovski, M. V., M. U. Musheev, A. P. Drabovich, J. V. Jitkova and S. N. Krylov (2006). "Non-SELEX: selection of aptamers without intermediate amplification of candidate oligonucleotides." Nat Protoc **1**(3): 1359-69.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne (2000). "The Protein Data Bank." Nucleic Acids Res **28**(1): 235-42.

Bernstein, F. C., T. F. Koetzle, G. J. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi (1977). "The Protein Data Bank: a computer-based archival file for macromolecular structures." J Mol Biol **112**(3): 535-42.

Bijlsma, J. J., B. Waidner, A. H. Vliet, N. J. Hughes, S. Hag, S. Bereswill, D. J. Kelly, C. M. Vandenbroucke-Grauls, M. Kist and J. G. Kusters (2002). "The Helicobacter pylori homologue of the ferric uptake regulator is involved in acid resistance." Infect Immun **70**(2): 606-11.

Bland, M. V., S. Ismail, J. A. Heinemann and J. I. Keenan (2004). "The action of bismuth against Helicobacter pylori mimics but is not caused by intracellular iron deprivation." Antimicrob Agents Chemother **48**(6): 1983-8.

Bloch, F., W. W. Hansen and M. E. Packard (1946). "Nuclear induction." Physical Review **69**: 127.

Borin, B. N. and A. M. Krezel (2008). "Structure of HP0564 from Helicobacter pylori identifies it as a new transcriptional regulator." Proteins **73**(1): 265-8.

Borin, B. N., A. Popescu and A. M. Krezel (2005). "NMR assignment of the novel Helicobacter pylori protein JHP1348." J Biomol NMR **32**(3): 262.

Breg, J. N., J. H. van Opheusden, M. J. Burgering, R. Boelens and R. Kaptein (1990). "Structure of Arc repressor in solution: evidence for a family of beta-sheet DNA-binding proteins." Nature **346**(6284): 586-9.

Burgess, R. R. (1969). "Separation and characterization of the subunits of ribonucleic acid polymerase." J Biol Chem **244**(22): 6168-76.

Bury-Mone, S., J. M. Thiberge, M. Contreras, A. Maitournam, A. Labigne and H. De Reuse (2004). "Responsiveness to acidity via metal ion regulators mediates virulence in the gastric pathogen Helicobacter pylori." Mol Microbiol **53**(2): 623-38.

Case, D. A., T. E. Cheatham, 3rd, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang and R. J. Woods (2005). "The Amber biomolecular simulation programs." J Comput Chem **26**(16): 1668-88.

Chivers, P. T. and R. T. Sauer (1999). "NikR is a ribbon-helix-helix DNA-binding protein." Protein Sci **8**(11): 2494-500.

Contreras, M., J. M. Thiberge, M. A. Mandrand-Berthelot and A. Labigne (2003). "Characterization of the roles of NikR, a nickel-responsive pleiotropic autoregulator of Helicobacter pylori." Mol Microbiol **49**(4): 947-63.

Cornilescu, G., F. Delaglio and A. Bax (1999). "Protein backbone angle restraints from searching a database for chemical shift and sequence homology." J Biomol NMR **13**(3): 289-302.

Correa, P. (2004). "The biological model of gastric carcinogenesis." IARC Sci Publ(157): 301-10.

Cramer, P. (2002). "Multisubunit RNA polymerases." Curr Opin Struct Biol **12**(1): 89-97.

Cramer, P., K. J. Armache, S. Baumli, S. Benkert, F. Brueckner, C. Buchen, G. E. Damsma, S. Dengl, S. R. Geiger, A. J. Jasiak, A. Jawhari, S. Jennebach, T. Kamenski, H.

Kettenberger, C. D. Kuhn, E. Lehmann, K. Leike, J. F. Sydow and A. Vannini (2008). "Structure of eukaryotic RNA polymerases." Annu Rev Biophys **37**: 337-52.

Cramer, P., D. A. Bushnell and R. D. Kornberg (2001). "Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution." Science **292**(5523): 1863-76.

Crane-Robinson, C., A. I. Dragan and P. L. Privalov (2006). "The extended arms of DNA-binding domains: a tale of tails." Trends Biochem Sci **31**(10): 547-52.

Croxen, M. A., G. Sisson, R. Melano and P. S. Hoffman (2006). "The Helicobacter pylori chemotaxis receptor TlpB (HP0103) is required for pH taxis and for colonization of the gastric mucosa." J Bacteriol **188**(7): 2656-65.

Dailidiene, D., S. Tan, K. Ogura, M. Zhang, A. H. Lee, K. Severinov and D. E. Berg (2007). "Urea sensitization caused by separation of Helicobacter pylori RNA polymerase beta and beta' subunits." Helicobacter **12**(2): 103-11.

Dangi, B., A. M. Gronenborn, J. L. Rosner and R. G. Martin (2004). "Versatility of the carboxy-terminal domain of the alpha subunit of RNA polymerase in transcriptional activation: use of the DNA contact site as a protein contact site for MarA." Mol Microbiol **54**(1): 45-59.

Davidson, B. E. and I. Saint Girons (1989). "The Escherichia coli regulatory protein MetJ binds to a tandemly repeated 8 bp palindrome." Mol Microbiol **3**(11): 1639-48.

de la Hoz, A. B., S. Ayora, I. Sitkiewicz, S. Fernandez, R. Pankiewicz, J. C. Alonso and P. Ceglowski (2000). "Plasmid copy-number control and better-than-random segregation genes of pSM19035 share a common regulator." Proc Natl Acad Sci U S A **97**(2): 728-33.

De Pina, K., V. Desjardin, M. A. Mandrand-Berthelot, G. Giordano and L. F. Wu (1999). "Isolation and characterization of the nikR gene encoding a nickel-responsive regulator in Escherichia coli." J Bacteriol **181**(2): 670-4.

del Solar, G., P. Acebo and M. Espinosa (1995). "Replication control of plasmid pLS1: efficient regulation of plasmid copy number is exerted by the combined action of two plasmid components, CopG and RNA II." Mol Microbiol **18**(5): 913-24.

Dietz, P., G. Gerlach and D. Beier (2002). "Identification of target genes regulated by the two-component system HP166-HP165 of Helicobacter pylori." J Bacteriol **184**(2): 350-62.

Djordjevic, M. (2007). "SELEX experiments: new prospects, applications and data analysis in inferring regulatory pathways." Biomol Eng **24**(2): 179-89.

El-Omar, E. M., M. Carrington, W. H. Chow, K. E. McColl, J. H. Bream, H. A. Young, J. Herrera, J. Lissowska, C. C. Yuan, N. Rothman, G. Lanyon, M. Martin, J. F. Fraumeni, Jr. and C. S. Rabkin (2000). "Interleukin-1 polymorphisms associated with increased risk of gastric cancer." Nature **404**(6776): 398-402.

Eppinger, M., C. Baar, B. Linz, G. Raddatz, C. Lanz, H. Keller, G. Morelli, H. Gressmann, M. Achtman and S. C. Schuster (2006). "Who ate whom? Adaptive Helicobacter genomic changes that accompanied a host jump from early humans to large felines." PLoS Genet **2**(7): e120.

Ernst, F. D., S. Bereswill, B. Waidner, J. Stoof, U. Mader, J. G. Kusters, E. J. Kuipers, M. Kist, A. H. van Vliet and G. Homuth (2005). "Transcriptional profiling of Helicobacter pylori Fur- and iron-regulated gene expression." Microbiology **151**(Pt 2): 533-46.

Ernst, F. D., E. J. Kuipers, A. Heijens, R. Sarwari, J. Stoof, C. W. Penn, J. G. Kusters and A. H. van Vliet (2005). "The nickel-responsive regulator NikR controls activation and repression of gene transcription in Helicobacter pylori." Infect Immun **73**(11): 7252-8.

Ernst, P. B. and B. D. Gold (2000). "The disease spectrum of Helicobacter pylori: the immunopathogenesis of gastroduodenal ulcer and gastric cancer." Annu Rev Microbiol **54**: 615-40.

Estrem, S. T., W. Ross, T. Gaal, Z. W. Chen, W. Niu, R. H. Ebright and R. L. Gourse (1999). "Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit." Genes Dev **13**(16): 2134-47.

Finn, R. D., J. Tate, J. Mistry, P. C. Coggill, S. J. Sammut, H. R. Hotz, G. Ceric, K. Forslund, S. R. Eddy, E. L. Sonnhammer and A. Bateman (2008). "The Pfam protein families database." Nucleic Acids Res **36**(Database issue): D281-8.

Forsyth, M. H., P. Cao, P. P. Garcia, J. D. Hall and T. L. Cover (2002). "Genome-wide transcriptional profiling in a histidine kinase mutant of Helicobacter pylori identifies members of a regulon." J Bacteriol **184**(16): 4630-5.

Galkin, V. E., X. Yu, J. Bielnicki, J. Heuser, C. P. Ewing, P. Guerry and E. H. Egelman (2008). "Divergence of quaternary structures among bacterial flagellar filaments." Science **320**(5874): 382-5.

Gauger, E. J., M. P. Leatham, R. Mercado-Lubo, D. C. Laux, T. Conway and P. S. Cohen (2007). "Role of motility and the flhDC Operon in Escherichia coli MG1655 colonization of the mouse intestine." Infect Immun **75**(7): 3315-24.

Gaynor, E. C., D. H. Wells, J. K. MacKichan and S. Falkow (2005). "The Campylobacter jejuni stringent response controls specific stress survival and virulence-associated phenotypes." Mol Microbiol **56**(1): 8-27.

Gebert, B., W. Fischer, E. Weiss, R. Hoffmann and R. Haas (2003). "Helicobacter pylori vacuolating cytotoxin inhibits T lymphocyte activation." Science **301**(5636): 1099-102.

Gerlach, W. and O. Stern (1922). "Das magnetische Moment des Silberatoms." Zeitschrift fur Physik **9**: 353-355.

Gerrits, M. M., M. R. de Zoete, N. L. Arents, E. J. Kuipers and J. G. Kusters (2002). "16S rRNA mutation-mediated tetracycline resistance in Helicobacter pylori." Antimicrob Agents Chemother **46**(9): 2996-3000.

Gerrits, M. M., D. Schuijffel, A. A. van Zwet, E. J. Kuipers, C. M. Vandenbroucke-Grauls and J. G. Kusters (2002). "Alterations in penicillin-binding protein 1A confer resistance to beta-lactam antibiotics in Helicobacter pylori." Antimicrob Agents Chemother **46**(7): 2229-33.

Ghosh, P., A. Ishihama and D. Chatterji (2001). "Escherichia coli RNA polymerase subunit omega and its N-terminal domain bind full-length beta' to facilitate incorporation into the alpha2beta subassembly." Eur J Biochem **268**(17): 4621-7.

Gilson, M. K., Sharp, K., Honig, B. (1987). "Calculating the Electrostatic Potential of Molecules in Solution: Method and error assessment." J Comp Chem **9**: 327-335.

Gopinath, S. C. (2007). "Methods developed for SELEX." Anal Bioanal Chem **387**(1): 171-82.

Graham, D. Y. and A. Shiotani (2008). "New concepts of resistance in the treatment of Helicobacter pylori infections." Nat Clin Pract Gastroenterol Hepatol **5**(6): 321-31.

Gressmann, H., B. Linz, R. Ghai, K. P. Pleissner, R. Schlapbach, Y. Yamaoka, C. Kraft, S. Suerbaum, T. F. Meyer and M. Achtman (2005). "Gain and loss of multiple genes during the evolution of Helicobacter pylori." PLoS Genet **1**(4): e43.

Guntert, P., C. Mumenthaler and K. Wuthrich (1997). "Torsion angle dynamics for NMR structure calculation with the new program DYANA." J Mol Biol **273**(1): 283-98.

Hanson, L. G. (2008). "Is Quantum Mechanics Necessary for Understanding Magnetic Resonance?" Concepts in Magnetic Resonance **32(A)**(5): 329-340.

Haugen, B. J., S. Pellett, P. Redford, H. L. Hamilton, P. L. Roesch and R. A. Welch (2007). "In vivo gene expression analysis identifies genes required for enhanced colonization of the mouse urinary tract by uropathogenic Escherichia coli strain CFT073 dsdA." Infect Immun **75**(1): 278-89.

Hautbergue, G. M. and A. P. Golovanov (2008). "Increasing the sensitivity of cryoprobe protein NMR experiments by using the sole low-conductivity arginine glutamate salt." J Magn Reson **191**(2): 335-9.

Havel, T. F., I. D. Kuntz and G. M. Crippen (1983). "The combinatorial distance geometry method for the calculation of molecular conformation. I. A new approach to an old problem." J Theor Biol **104**(3): 359-81.

Hayward, R. S., K. Igarashi and A. Ishihama (1991). "Functional specialization within the alpha-subunit of Escherichia coli RNA polymerase." J Mol Biol **221**(1): 23-9.

Herbert, K. M., W. J. Greenleaf and S. M. Block (2008). "Single-molecule studies of RNA polymerase: motoring along." Annu Rev Biochem **77**: 149-76.

Higashi, H., R. Tsutsumi, S. Muto, T. Sugiyama, T. Azuma, M. Asaka and M. Hatakeyama (2002). "SHP-2 tyrosine phosphatase as an intracellular target of Helicobacter pylori CagA protein." Science **295**(5555): 683-6.

Hirata, A., B. J. Klein and K. S. Murakami (2008). "The X-ray crystal structure of RNA polymerase from Archaea." Nature **451**(7180): 851-4.

Holm, L. and C. Sander (1996). "Mapping the protein universe." Science **273**(5275): 595-603.

IARC (1994). "Schistosomes, liver flukes and Helicobacter pylori. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, 7-14 June 1994." IARC Monogr Eval Carcinog Risks Hum **61**: 1-241.

Irvine, D., C. Tuerk and L. Gold (1991). "SELEXION. Systematic evolution of ligands by exponential enrichment with integrated optimization by non-linear analysis." J Mol Biol **222**(3): 739-61.

Ishihama, A. (1981). "Subunit of assembly of Escherichia coli RNA polymerase." Adv Biophys **14**: 1-35.

Israel, D. A., N. Salama, U. Krishna, U. M. Rieger, J. C. Atherton, S. Falkow and R. M. Peek, Jr. (2001). "Helicobacter pylori genetic diversity within the gastric niche of a single human host." Proc Natl Acad Sci U S A **98**(25): 14625-30.

Iwamoto, H., D. M. Czajkowsky, T. L. Cover, G. Szabo and Z. Shao (1999). "VacA from Helicobacter pylori: a hexameric chloride channel." FEBS Lett **450**(1-2): 101-4.

Jain, V., M. Kumar and D. Chatterji (2006). "ppGpp: stringent response and survival." J Microbiol **44**(1): 1-10.

Jeon, Y. H., T. Negishi, M. Shirakawa, T. Yamazaki, N. Fujita, A. Ishihama and Y. Kyogoku (1995). "Solution structure of the activator contact domain of the RNA polymerase alpha subunit." Science **270**(5241): 1495-7.

Jeon, Y. H., T. Yamazaki, T. Otomo, A. Ishihama and Y. Kyogoku (1997). "Flexible linker in the RNA polymerase alpha subunit facilitates the independent motion of the C-terminal activator contact domain." J Mol Biol **267**(4): 953-62.

Jones, M. A., K. L. Marston, C. A. Woodall, D. J. Maskell, D. Linton, A. V. Karlyshev, N. Dorrell, B. W. Wren and P. A. Barrow (2004). "Adaptation of Campylobacter jejuni NCTC11168 to high-level colonization of the avian gastrointestinal tract." Infect Immun **72**(7): 3769-76.

Josenhans, C., D. Beier, B. Linz, T. F. Meyer and S. Suerbaum (2007). "Pathogenomics of helicobacter." Int J Med Microbiol **297**(7-8): 589-600.

Josenhans, C., K. A. Eaton, T. Thevenot and S. Suerbaum (2000). "Switching of flagellar motility in Helicobacter pylori by reversible length variation of a short homopolymeric

sequence repeat in fliP, a gene encoding a basal body protein." Infect Immun **68**(8): 4598-603.

Kanagawa, T. (2003). "Bias and artifacts in multitemplate polymerase chain reactions (PCR)." J Biosci Bioeng **96**(4): 317-23.

Kandulski, A., M. Selgrad and P. Malfertheiner (2008). "Helicobacter pylori infection: a clinical overview." Dig Liver Dis **40**(8): 619-26.

Kang, J. and M. J. Blaser (2006). "Bacterial populations as perfect gases: genomic integrity and diversification tensions in Helicobacter pylori." Nat Rev Microbiol **4**(11): 826-36.

Kersulyte, D., H. Chalkauskas and D. E. Berg (1999). "Emergence of recombinant strains of Helicobacter pylori during human infection." Mol Microbiol **31**(1): 31-43.

Kim, N., E. A. Marcus, Y. Wen, D. L. Weeks, D. R. Scott, H. C. Jung, I. S. Song and G. Sachs (2004). "Genes of Helicobacter pylori regulated by attachment to AGS cells." Infect Immun **72**(4): 2358-68.

Kivi, M. and Y. Tindberg (2006). "Helicobacter pylori occurrence and transmission: a family affair?" Scand J Infect Dis **38**(6-7): 407-17.

Kleinberg, R. (2001). "NMR Measurement of Petrophysical Properties." Concepts in Magnetic Resonance **13**(6): 404-406.

Kraft, C., A. Stack, C. Josenhans, E. Niehus, G. Dietrich, P. Correa, J. G. Fox, D. Falush and S. Suerbaum (2006). "Genomic changes during chronic Helicobacter pylori infection." J Bacteriol **188**(1): 249-54.

Krishnaswamy, R. and D. B. Wilson (2000). "Construction and characterization of an Escherichia coli strain genetically engineered for Ni(II) bioaccumulation." Appl Environ Microbiol **66**(12): 5383-6.

Laskowski, R. A., J. A. Rullmann, M. W. MacArthur, R. Kaptein and J. M. Thornton (1996). "AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR." J Biomol NMR **8**(4): 477-86.

Lee, D. J., H. J. Wing, N. J. Savery and S. J. Busby (2000). "Analysis of interactions between Activating Region 1 of Escherichia coli FNR protein and the C-terminal domain of the RNA polymerase alpha subunit: use of alanine scanning and suppression genetics." Mol Microbiol **37**(5): 1032-40.

Levine, H. A. and M. Nilsen-Hamilton (2007). "A mathematical analysis of SELEX." Comput Biol Chem **31**(1): 11-35.

Levine, M., S. Truesdell, T. Ramakrishnan and M. J. Bronson (1975). "Dual control of lysogeny by bacteriophage P22: an antirepressor locus and its controlling elements." J Mol Biol **91**(4): 421-38.

Linz, B., F. Balloux, Y. Moodley, A. Manica, H. Liu, P. Roumagnac, D. Falush, C. Stamer, F. Prugnolle, S. W. van der Merwe, Y. Yamaoka, D. Y. Graham, E. Perez-Trallero, T. Wadstrom, S. Suerbaum and M. Achtman (2007). "An African origin for the intimate association between humans and Helicobacter pylori." Nature **445**(7130): 915-8.

Lipsitz, R. S. and N. Tjandra (2004). "Residual dipolar couplings in NMR structure analysis." Annu Rev Biophys Biomol Struct **33**: 387-413.

Loh, J. T. and T. L. Cover (2006). "Requirement of histidine kinases HP0165 and HP1364 for acid resistance in Helicobacter pylori." Infect Immun **74**(5): 3052-9.

Loh, J. T., M. H. Forsyth and T. L. Cover (2004). "Growth phase regulation of flaA expression in Helicobacter pylori is luxS dependent." Infect Immun **72**(9): 5506-10.

Marais, A., G. L. Mendz, S. L. Hazell and F. Megraud (1999). "Metabolism and genetics of Helicobacter pylori: the genome era." Microbiol Mol Biol Rev **63**(3): 642-74.

Marcus, E. A., A. P. Moshfegh, G. Sachs and D. R. Scott (2005). "The periplasmic alpha-carbonic anhydrase activity of Helicobacter pylori is essential for acid acclimation." J Bacteriol **187**(2): 729-38.

Marshall, B. J. and J. R. Warren (1984). "Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration." Lancet **1**(8390): 1311-5.

Martinez-Granero, F., R. Rivilla and M. Martin (2006). "Rhizosphere selection of highly motile phenotypic variants of Pseudomonas fluorescens with enhanced competitive colonization ability." Appl Environ Microbiol **72**(5): 3429-34.

Mathew, R. and D. Chatterji (2006). "The evolving story of the omega subunit of bacterial RNA polymerase." Trends Microbiol **14**(10): 450-5.

McFarland, L. V. (2008). "Antibiotic-associated diarrhea: epidemiology, trends and treatment." Future Microbiol **3**: 563-78.

McLeod, S. M., S. E. Aiyar, R. L. Gourse and R. C. Johnson (2002). "The C-terminal domains of the RNA polymerase alpha subunits: contact site with Fis and localization during co-activation with CRP at the Escherichia coli proP P2 promoter." J Mol Biol **316**(3): 517-29.

Merrell, D. S., M. L. Goodrich, G. Otto, L. S. Tompkins and S. Falkow (2003). "pH-regulated gene expression of the gastric pathogen Helicobacter pylori." Infect Immun **71**(6): 3529-39.

Mouery, K., B. A. Rader, E. C. Gaynor and K. Guillemin (2006). "The stringent response is required for Helicobacter pylori survival of stationary phase, exposure to acid, and aerobic shock." J Bacteriol **188**(15): 5494-500.

Murakami, K. S. and S. A. Darst (2003). "Bacterial RNA polymerases: the wholo story." Curr Opin Struct Biol **13**(1): 31-9.

Musheev, M. U. and S. N. Krylov (2006). "Selection of aptamers by systematic evolution of ligands by exponential enrichment: addressing the polymerase chain reaction issue." Anal Chim Acta **564**(1): 91-6.

Nakagawa, S., Y. Takaki, S. Shimamura, A. L. Reysenbach, K. Takai and K. Horikoshi (2007). "Deep-sea vent epsilon-proteobacterial genomes provide insights into emergence of pathogens." Proc Natl Acad Sci U S A **104**(29): 12146-50.

Newberry, K. J., S. Nakano, P. Zuber and R. G. Brennan (2005). "Crystal structure of the Bacillus subtilis anti-alpha, global transcriptional regulator, Spx, in complex with the alpha C-terminal domain of RNA polymerase." Proc Natl Acad Sci U S A **102**(44): 15839-44.

Niehus, E., H. Gressmann, F. Ye, R. Schlapbach, M. Dehio, C. Dehio, A. Stack, T. F. Meyer, S. Suerbaum and C. Josenhans (2004). "Genome-wide analysis of transcriptional hierarchy and feedback regulation in the flagellar system of Helicobacter pylori." Mol Microbiol **52**(4): 947-61.

Niehus, E., F. Ye, S. Suerbaum and C. Josenhans (2002). "Growth phase-dependent and differential transcriptional control of flagellar genes in Helicobacter pylori." Microbiology **148**(Pt 12): 3827-37.

Oh, J. D., H. Kling-Backhed, M. Giannakis, J. Xu, R. S. Fulton, L. A. Fulton, H. S. Cordum, C. Wang, G. Elliott, J. Edwards, E. R. Mardis, L. G. Engstrand and J. I. Gordon (2006). "The complete genome sequence of a chronic atrophic gastritis Helicobacter pylori strain: evolution during disease progression." Proc Natl Acad Sci U S A **103**(26): 9999-10004.

Ottlecz, A., J. J. Romero and L. M. Lichtenberger (1999). "Effect of ranitidine bismuth citrate on the phospholipase A2 activity of Naja naja venom and Helicobacter pylori: a biochemical analysis." Aliment Pharmacol Ther **13**(7): 875-81.

Overhauser, A. (1953). "Polarization of Nuclei in Metals." Physical Review **92**(2): 411-415.

Papini, E., B. Satin, N. Norais, M. de Bernard, J. L. Telford, R. Rappuoli and C. Montecucco (1998). "Selective increase of the permeability of polarized epithelial cell monolayers by Helicobacter pylori vacuolating toxin." J Clin Invest **102**(4): 813-20.

Peek, R. M., Jr., G. G. Miller, K. T. Tham, G. I. Perez-Perez, X. Zhao, J. C. Atherton and M. J. Blaser (1995). "Heightened inflammatory response and cytokine expression in vivo to cagA+ Helicobacter pylori strains." Lab Invest **73**(6): 760-70.

Perez-Losada, M., E. B. Browne, A. Madsen, T. Wirth, R. P. Viscidi and K. A. Crandall (2006). "Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data." Infect Genet Evol **6**(2): 97-112.

Pervushin, K., R. Riek, G. Wider and K. Wuthrich (1997). "Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution." Proc Natl Acad Sci U S A **94**(23): 12366-71.

Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin (2004). "UCSF Chimera--a visualization system for exploratory research and analysis." J Comput Chem **25**(13): 1605-12.

Pflock, M., P. Dietz, J. Schar and D. Beier (2004). "Genetic evidence for histidine kinase HP165 being an acid sensor of Helicobacter pylori." FEMS Microbiol Lett **234**(1): 51-61.

Pflock, M., N. Finsterer, B. Joseph, H. Mollenkopf, T. F. Meyer and D. Beier (2006). "Characterization of the ArsRS regulon of Helicobacter pylori, involved in acid adaptation." J Bacteriol **188**(10): 3449-62.

Pflock, M., S. Kennard, I. Delany, V. Scarlato and D. Beier (2005). "Acid-induced activation of the urease promoters is mediated directly by the ArsRS two-component system of Helicobacter pylori." Infect Immun **73**(10): 6437-45.

Pflock, M., S. Muller and D. Beier (2007). "The CrdRS (HP1365-HP1364) Two-Component System Is Not Involved in pH-Responsive Gene Regulation in the Helicobacter pylori Strains 26695 and G27." Curr Microbiol **54**(4): 320-4.

Popescu, A. (2004). Structure-Based Discovery of Protein Function: HP0222 of *Helicobacter pylori*. Department of Biological Sciences. Nashville, Vanderbilt University. **Ph.D.:** 110.

Popescu, A., A. Karpay, D. A. Israel, R. M. Peek, Jr. and A. M. Krezel (2005). "Helicobacter pylori protein HP0222 belongs to Arc/MetJ family of transcriptional regulators." Proteins **59**(2): 303-11.

Purcell, E. M., H. C. Torrey and R. V. Pound (1946). "Resonance Absorption by Nuclear Magnetic Moments in a Solid." Physical Review **69**: 37-38.

Rabi, I. I. M., S.; P. Kusch P., Zacharias, J.R. (1939). "The Molecular Beam Resonance Method for Measuring Nuclear Magnetic Moments. The Magnetic Moments of 3Li6, 3Li7 and 9F19." Physical Review **55**: 526-535.

Rader, B. A., S. R. Campagna, M. F. Semmelhack, B. L. Bassler and K. Guillemin (2007). "The quorum-sensing molecule autoinducer 2 regulates motility and flagellar morphogenesis in Helicobacter pylori." J Bacteriol **189**(17): 6109-17.

Rafferty, J. B., W. S. Somers, I. Saint-Girons and S. E. Phillips (1989). "Three-dimensional crystal structures of Escherichia coli met repressor with and without corepressor." Nature **341**(6244): 705-10.

Raumann, B. E., M. A. Rould, C. O. Pabo and R. T. Sauer (1994). "DNA recognition by beta-sheets in the Arc repressor-operator crystal structure." Nature **367**(6465): 754-7.

Rice, P., I. Longden and A. Bleasby (2000). "EMBOSS: the European Molecular Biology Open Software Suite." Trends Genet **16**(6): 276-7.

Ross, W., K. K. Gosink, J. Salomon, K. Igarashi, C. Zou, A. Ishihama, K. Severinov and R. L. Gourse (1993). "A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase." Science **262**(5138): 1407-13.

Rozen, S. and H. J. Skaletsky (2000). Primer3 on the WWW for general users and for biologist programmers. Bioinformatics Methods and Protocols: Methods in Molecular Biology. S. Krawetz and S. Misener. Totowa, NJ, Humana Press**:** 365-386.

Saadat, I., H. Higashi, C. Obuse, M. Umeda, N. Murata-Kamiya, Y. Saito, H. Lu, N. Ohnishi, T. Azuma, A. Suzuki, S. Ohno and M. Hatakeyama (2007). "Helicobacter pylori CagA targets PAR1/MARK kinase to disrupt epithelial cell polarity." Nature **447**(7142): 330-3.

Sachs, G., D. L. Weeks, Y. Wen, E. A. Marcus, D. R. Scott and K. Melchers (2005). "Acid acclimation by Helicobacter pylori." Physiology (Bethesda) **20**: 429-38.

Saint-Girons, I., C. Parsot, M. M. Zakin, O. Barzu and G. N. Cohen (1988). "Methionine biosynthesis in Enterobacteriaceae: biochemical, regulatory, and evolutionary aspects." CRC Crit Rev Biochem **23 Suppl 1**: S1-42.

Salama, N., K. Guillemin, T. K. McDaniel, G. Sherlock, L. Tompkins and S. Falkow (2000). "A whole-genome microarray reveals genetic diversity among Helicobacter pylori strains." Proc Natl Acad Sci U S A **97**(26): 14668-73.

Salcedo, J. A. and F. Al-Kawas (1998). "Treatment of Helicobacter pylori infection." Arch Intern Med **158**(8): 842-51.

Sattler, M., J. Schleucher and C. Griesinger (1999). "Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients." Progress in Nuclear Magnetic Resonance Spectroscopy **34**: 93-158.

Schreiter, E. R. and C. L. Drennan (2007). "Ribbon-helix-helix transcription factors: variations on a theme." Nat Rev Microbiol **5**(9): 710-20.

Scoarughi, G. L., C. Cimmino and P. Donini (1999). "Helicobacter pylori: a eubacterium lacking the stringent response." J Bacteriol **181**(2): 552-5.

Shirai, M., R. Fujinaga, J. K. Akada and T. Nakazawa (1999). "Activation of Helicobacter pylori ureA promoter by a hybrid Escherichia coli-H. pylori rpoD gene in E. coli." Gene **239**(2): 351-9.

Snelling, W. J., M. Matsuda, J. E. Moore and J. S. Dooley (2006). "Under the microscope: Arcobacter." Lett Appl Microbiol **42**(1): 7-14.

Solnick, J. V., L. M. Hansen and M. Syvanen (1997). "The major sigma factor (RpoD) from Helicobacter pylori and other gram-negative bacteria shows an enhanced rate of divergence." J Bacteriol **179**(19): 6196-200.

Somers, W. S., J. B. Rafferty, K. Phillips, S. Strathdee, Y. Y. He, T. McNally, I. Manfield, O. Navratil, I. G. Old, I. Saint-Girons and et al. (1994). "The Met repressor-operator complex: DNA recognition by beta-strands." Ann N Y Acad Sci **726**: 105-17.

Speksnijder, A. G., G. A. Kowalchuk, S. De Jong, E. Kline, J. R. Stephen and H. J. Laanbroek (2001). "Microvariation artifacts introduced by PCR and cloning of closely related 16S rRNA gene sequences." Appl Environ Microbiol **67**(1): 469-72.

Srivatsan, A. and J. D. Wang (2008). "Control of bacterial transcription, translation and replication by (p)ppGpp." Curr Opin Microbiol **11**(2): 100-5.

Stratton, C. W., R. R. Warner, P. E. Coudron and N. A. Lilly (1999). "Bismuth-mediated disruption of the glycocalyx-cell wall of Helicobacter pylori: ultrastructural evidence for a mechanism of action for bismuth salts." J Antimicrob Chemother **43**(5): 659-66.

Suerbaum, S. and C. Josenhans (2007). "Helicobacter pylori evolution and phenotypic diversification in a changing host." Nat Rev Microbiol **5**(6): 441-52.

Susskind, M. M. Y., P. (1983). Bacteriophage P22 antirepressor and its control. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.

Swisher, S. C. and A. J. Barbati (2007). "Helicobacter pylori strikes again: gastric mucosa-associated lymphoid tissue (MALT) lymphoma." Gastroenterol Nurs **30**(5): 348-54; quiz 355-6.

Thompson, J. R., L. A. Marcelino and M. F. Polz (2002). "Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'." Nucleic Acids Res **30**(9): 2083-8.

Thompson, L. J., D. S. Merrell, B. A. Neilan, H. Mitchell, A. Lee and S. Falkow (2003). "Gene expression profiling of Helicobacter pylori reveals a growth-phase-dependent switch in virulence gene expression." Infect Immun **71**(5): 2643-55.

Tomb, J. F., O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H. G. Khalak, A. Glodek, K. McKenney, L. M. Fitzegerald, N. Lee, M. D. Adams, E. K. Hickey, D. E. Berg, J. D. Gocayne, T. R. Utterback, J. D. Peterson, J. M. Kelley, M. D. Cotton, J. M. Weidman, C. Fujii, C. Bowman, L. Watthey, E. Wallin, W. S. Hayes, M. Borodovsky, P. D. Karp, H. O. Smith, C. M. Fraser and J. C. Venter (1997). "The complete genome sequence of the gastric pathogen Helicobacter pylori." Nature **388**(6642): 539-47.

Tuerk, C. and L. Gold (1990). "Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase." Science **249**(4968): 505-10.

van Nimwegen, E. (2003). "Scaling laws in the functional content of genomes." Trends Genet **19**(9): 479-84.

Vassylyev, D. G., S. Sekine, O. Laptenko, J. Lee, M. N. Vassylyeva, S. Borukhov and S. Yokoyama (2002). "Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution." Nature **417**(6890): 712-9.

Waidner, B., K. Melchers, I. Ivanov, H. Loferer, K. W. Bensch, M. Kist and S. Bereswill (2002). "Identification by RNA profiling and mutational analysis of the novel copper resistance determinants CrdA (HP1326), CrdB (HP1327), and CzcB (HP1328) in Helicobacter pylori." J Bacteriol **184**(23): 6700-8.

Waidner, B., K. Melchers, F. N. Stahler, M. Kist and S. Bereswill (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**(13): 4683-8.

Wang, G., M. Z. Humayun and D. E. Taylor (1999). "Mutation as an origin of genetic variability in Helicobacter pylori." Trends Microbiol **7**(12): 488-93.

Warren, J. R. and B. J. Marshall (1983). "Unidentified Curved Bacilli on Gastric Epithelium in Active Chronic Gastritis." The Lancet **321**(8336): 1273-1275.

Wells, D. H. and E. C. Gaynor (2006). "Helicobacter pylori initiates the stringent response upon nutrient and pH downshift." J Bacteriol **188**(10): 3726-9.

Wen, Y., J. Feng, D. R. Scott, E. A. Marcus and G. Sachs (2006). "Involvement of the HP0165-HP0166 two-component system in expression of some acidic-pH-upregulated genes of Helicobacter pylori." J Bacteriol **188**(5): 1750-61.

Wen, Y., J. Feng, D. R. Scott, E. A. Marcus and G. Sachs (2007). "The HP0165-HP0166 two-component system (ArsRS) regulates acid-induced expression of HP1186 alpha-carbonic anhydrase in Helicobacter pylori by activating the pH-dependent promoter." J Bacteriol **189**(6): 2426-34.

Wen, Y., E. A. Marcus, U. Matrubutham, M. A. Gleeson, D. R. Scott and G. Sachs (2003). "Acid-adaptive genes of Helicobacter pylori." Infect Immun **71**(10): 5921-39.

Williamson, M. P., T. F. Havel and K. Wuthrich (1985). "Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry." J Mol Biol **182**(2): 295-315.

Wirth, T., X. Wang, B. Linz, R. P. Novick, J. K. Lum, M. Blaser, G. Morelli, D. Falush and M. Achtman (2004). "Distinguishing human ethnic groups by means of sequences from Helicobacter pylori: lessons from Ladakh." Proc Natl Acad Sci U S A **101**(14): 4746-51.

Wishart, D. S., C. G. Bigam, J. Yao, F. Abildgaard, H. J. Dyson, E. Oldfield, J. L. Markley and B. D. Sykes (1995). "1H, 13C and 15N chemical shift referencing in biomolecular NMR." J Biomol NMR **6**(2): 135-40.

Wishart, D. S. and B. D. Sykes (1994). "The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data." J Biomol NMR **4**(2): 171-80.

Wolanin, P. M., P. A. Thomason and J. B. Stock (2002). "Histidine protein kinases: key signal transducers outside the animal kingdom." Genome Biol **3**(10): REVIEWS3013.

Wüthrich, K. (1986). NMR of Proteins and Nucleic Acids. New York, NY, John Wiley & Sons, Inc.

Xiang, Z., S. Censini, P. F. Bayeli, J. L. Telford, N. Figura, R. Rappuoli and A. Covacci (1995). "Analysis of expression of CagA and VacA virulence factors in 43 strains of Helicobacter pylori reveals that clinical isolates can be divided into two major types and that CagA is not necessary for expression of the vacuolating cytotoxin." Infect Immun **63**(1): 94-8.

Young, K. T., L. M. Davis and V. J. Dirita (2007). "Campylobacter jejuni: molecular biology and pathogenesis." Nat Rev Microbiol **5**(9): 665-79.

Zakharova, N., B. J. Paster, I. Wesley, F. E. Dewhirst, D. E. Berg and K. V. Severinov (1999). "Fused and overlapping rpoB and rpoC genes in Helicobacters, Campylobacters, and related bacteria." J Bacteriol **181**(12): 3857-9.

Zhang, G., E. A. Campbell, L. Minakhin, C. Richter, K. Severinov and S. A. Darst (1999). "Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution." Cell **98**(6): 811-24.