CHAPTER I

INTRODUCTION

Helicobacter pylori infection in humans

Helicobacter pylori is a spiral-shaped gram-negative organism that persistently colonizes the stomach in about half of the human population. Infection is typically acquired in childhood and in the absence of antibiotic therapy often persists for the lifetime of the individual. The prevalence of *H. pylori* is higher in developing countries than in developed countries, with risk factors for acquisition including lower socioeconomic status, poor sanitation, and household crowding (*1-5*).

While most *H. pylori* colonize the gastric mucus layer and do not interact with the gastric cells, some organisms do adhere to gastric epithelial cells and occasionally are internalized. Additionally, some strains of *H. pylori* can secrete toxins or other effector molecules, causing further changes in the gastric environment and eliciting a host immune response and inflammation (gastritis). Despite these alterations in the host environment, the bacteria often colonize the human stomach for long periods of time without any adverse effects. Thus, infection and gastritis are asymptomatic in most persons (reviewed in (*2*)).

Although *H. pylori* typically colonizes the human stomach for decades without detrimental effects, its presence has been linked to an increased risk of developing peptic ulcers, gastric adenocarcinoma, and gastric mucosa-associated lymphoid tissue (MALT) lymphoma (reviewed in (*2-4, 6-10*)). The risks of developing these diseases are based, in part, on the characteristics of the *H. pylori* strain with which an individual is colonized. Bacterial polymorphisms in genes encoding host-interacting proteins can often influence the severity of symptoms in the host. Host and environmental factors can also contribute to infection-associated risk of gastroduodenal disease. For example, host factors such as polymorphisms in cytokine genes are important

determinants of the inflammatory response to infection and can predispose some individuals to malignancy (reviewed in (*2, 11*)).

Infection in most individuals causes superficial chronic gastritis which is usually clinically asymptomatic, however the changes in the gastric milieu are histologically apparent. In a subset of individuals with *H. pylori* infection, the chronic inflammation and associated oxidative stress can disrupt gastric mucosal homeostasis and promote progression from atrophic gastritis to dysplasia and eventually present as gastric cancer, the second major cause of cancer-related death worldwide. Progression through the various histologic cancer precursors is accompanied by the accumulation of mutations, typical of a multi-step process of carcinogenesis. There was sufficient evidence for an association between *H. pylori* infection and gastric cancer even in 1994 when the Working Group of the International Agency for Research on Cancer classified it as a class I carcinogen, the only bacterial agent on this list. Since then, data from animal models and human studies have only strengthened this association and provided insight into *H. pylori*'s ability to inhabit the gastric niche and its mechanisms of pathogenesis (reviewed in (*6, 11-14*)).

H. pylori and its ability to colonize the human stomach

While many microorganisms transit through the human stomach, most cannot successfully colonize this inhospitable niche. *H. pylori* is a neutrophilic bacterium that is uniquely adapted to colonize the human stomach and is therefore considered the dominant microbiota of the human stomach. Gastric colonization requires overcoming the stomach's many antibacterial properties such as acidity, peristalsis, scarcity of nutrient availability, and host innate and adaptive immunity (reviewed in (*2, 15, 16*)).

Gastric acidity is a notable stress in the human stomach that requires continuous monitoring of environmental pH and appropriate adaptation by the bacterium. This constant management is necessary because within the human stomach, *H. pylori* endures a range of acidic pH conditions (*17-21*). Upon entry of *H. pylori* into the stomach, the bacterium encounters gastric luminal pH

values that range from 1 while fasting to 5 after a meal. After the initial transit of *H. pylori* through the gastric lumen, the bacterium thrives in the gastric mucus layer, where the bacterium is still subject to considerable pH fluctuations, ranging from 4.5-6.5, as a consequence of the changing luminal pH (Figure I-1, adapted from (*15*)) (reviewed in (*22*)). Therefore, *H. pylori* has evolved mechanisms to survive severe acid shocks and multiply under moderately acidic conditions.

One of the mechanisms by which *H. pylori* responds to acidic pH is by taking shelter in the gastric mucus layer, where it is protected from the severe changes in luminal pH. Its flagella confer motility enabling the organisms to penetrate and colonize the viscous environment of the mucus layer (reviewed in (*2, 16*)). Another adaptive mechanism involves producing buffering compounds by enzymes such as urease, amidases, α-carbonic anhydrase, and arginase (*23-26*). In particular, urease is produced in abundance making up about 6% of total bacterial protein, and is an important colonization factor in *H. pylori* (*27, 28*). The nickel-containing enzyme hydrolyses urea into ammonia and CO_2. After the conversion of CO_2 to HCO_3^- by the periplasmic α-carbonic anhydrase, both ammonia and bicarbonate contribute to buffering the cytoplasm and the periplasm of *H. pylori.*

Expression of the acid-adaptive components of *H. pylori* is regulated by an intricate interplay of transcriptional regulators responding to different primary stimuli including pH changes in the bacterial environment ((*25*), and reviewed in (*22*)). The major players involved in regulating the acid response include ArsRS, CrdRS, FlgRS, NikR, and Fur, and it is likely that many additional regulators remain to be identified. Under acidic conditions, solubility and thus bioavailability of metal ions is increased and hence, metal-dependent regulators CrdRS, NikR and Fur indirectly monitor and respond to changes in pH (*19, 29*). Recent studies (Wen *et al.* (*30*)) indicate that the FlgRS system is directly involved in monitoring cytoplasmic pH and regulating gene expression in response to low pH. Targets of this regulatory system overlap with targets of another well

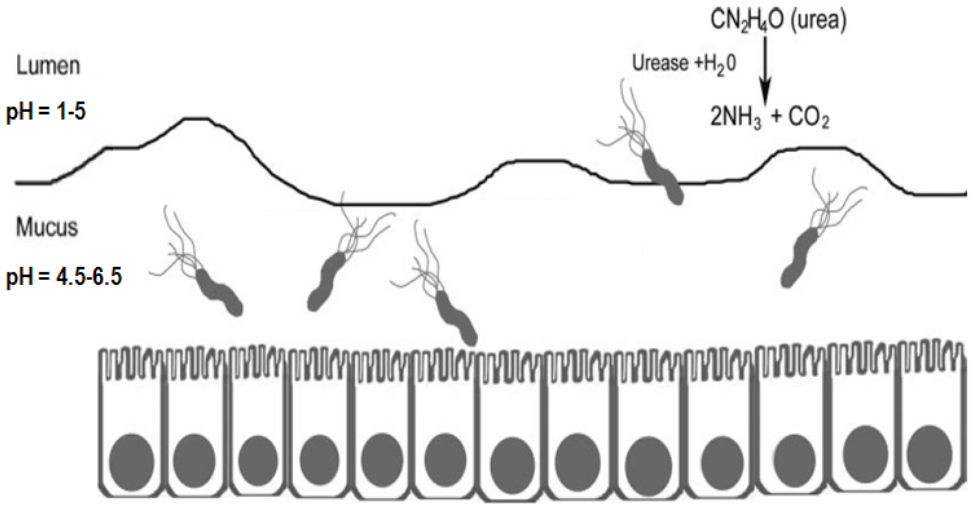


Figure I-1. *H. pylori* and gastric acidity. Within the human stomach, *H. pylori* encounters variable pH conditions ranging from 1-5 in the gastric lumen, and 4.5-6.5 in the protective mucus layer. *H. pylori* employs acid adaptive mechanisms such as urease-mediated generation of ammonia to neutralize its acidic environment (figure adapted from (*15*)).

characterized regulatory system, ArsRS (reviewed in (*22*)). CrdRS, FlgRS, and ArsRS are examples of regulatory schemes known as two-component systems.

Two component systems

Two-component systems (TCSs) are stimulus-response coupling communication modules. They are used primarily by prokaryotic organisms to regulate cellular functions in response to changing environmental conditions (*31*). TCSs were discovered by Hall and Silhavy in *Escherichia coli*. They described bacterial sensing and relay of changes in environmental osmolarity by a membrane protein, EnvZ, and a cytoplasmic protein, OmpR, which regulated the expression of outer-membrane protein genes, *ompF* and *ompC* (*31*). Since then, about 4000 such systems have been identified from the 145 sequenced bacterial genomes (*32*).

TCSs couple signaling with a transfer of a phosphate group from a sensor histidine kinase (HK) protein to a response regulator (RR) protein. The HK typically monitors specific environmental stimuli and autophosphorylates in response. The following phosphotransfer event relays information to a cognate RR. Both HKs and RRs are modular proteins, each comprised of several domains. The HK consists of a sensing domain and a transmitter domain. Sensor domains of HKs are modulated by environmental signals. Transmitter domains of multiple HKs autophosphorylate and relay information to the RRs. The RR has an information receiving regulatory domain, and a response-mediating effector domain. HKs can vary greatly in their sensing domains, but share a highly conserved kinase core in the transmitter domain that has a unique fold.

In a classic TCS (Figure I-2), stimuli sensed lead to autophosphorylation of the HK at a conserved histidine residue in the kinase core in its transmitter domain. This autophosphorylation is an ATP-dependent cross-phosphorylation reaction between homodimeric subunits of HK, where each subunit catalyzes the phosphorylation of the conserved His residue in the other subunit. The catalytic site in the RR's N-terminal regulatory domain transfers the phosphate

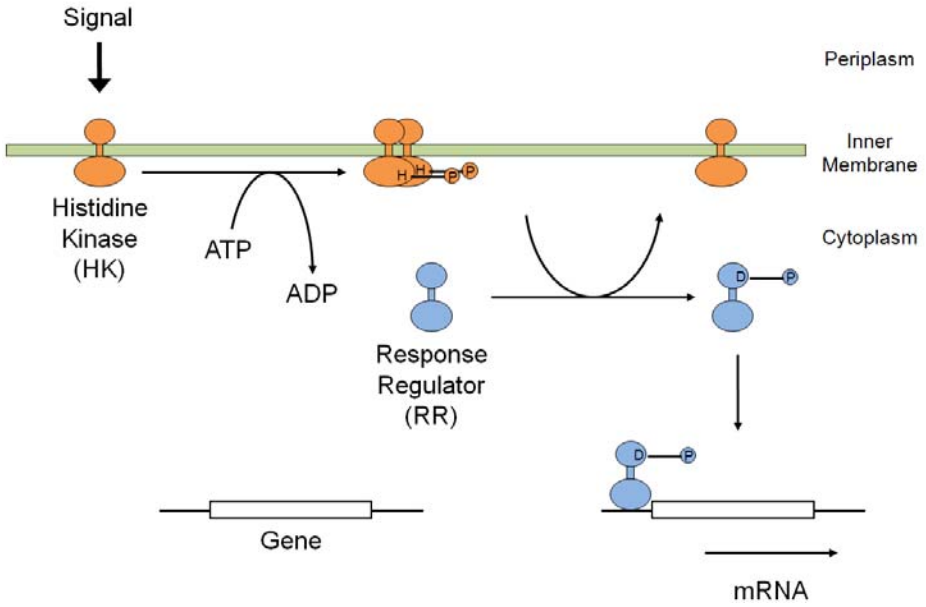


Figure I-2. Two component signal transduction system. The figure presents a model of a TCS in a gram-negative organism. A membrane bound HK detects environmental stimuli via its sensing domain that leads to autophosphorylation of the HK at a conserved histidine residue. The cognate RR catalytically transfers the phosphate group from phospho-HK to a conserved aspartate residue in its N-terminal regulatory domain. The activated RR, in this case a transcription factor, initiates a cellular response by activating and/or repressing transcription of specific genes.

group from phospho-HK to a conserved aspartate residue within the site. This phosphate-receiving aspartate lies adjacent to other acidic residues whose carboxylic side chains are involved in coordination of the Mg^{2+} that is required for phosphotransfer and dephosphorylation (Figure I-3, adapted from (*33*)). In most TCSs, RRs are the terminal component of the pathway, eliciting the adaptive responses when activated by phosphorylation through its C-terminal effector domain. While RRs can vary greatly in their effector domains, their regulatory domains containing the phosphate-receiving aspartate and the surrounding cluster of acidic residues is highly conserved (reviewed in (*34*)).

Small molecules, such as acetyl phosphate, carbamoyl phosphate, and phosphoramidate, can serve as phosphate donors to RRs *in vitro*, suggesting that RRs can catalyze phosphotransfer independent of the HK. Recent data suggest that phosphorylation of the RR induces a conformational change in the regulatory domain, which allows for a distinct set of inter- and intra-molecular interactions. The resulting RR activation could be through a variety of mechanisms: relief of inhibition, formation of dimers or higher order oligomers, or interaction with DNA. The TCS-mediated cellular response is elicited by the effector domain of the RR (reviewed in (*34*)).

The majority of RRs are transcription factors with DNA-binding effector domains (DBDs) that activate and/or repress transcription of specific genes. Based on the conservation of DNA-binding folds in their DBDs, the RRs can be classified into three major subfamilies: OmpR/PhoB, NarL, and NtrC. Despite performing similar functions, there is great variability in the actions of RRs, even within the same subfamily. RRs vary in their preferences for DNA sequences they recognize, the arrangement of their binding sites, and the specific mechanism of transcriptional regulation (reviewed in (*34*)). While most RRs regulate gene expression, not all the effector domains have DNA-binding functions. Some RRs have effector domains that function as enzymes, while others lack effector domains altogether, functioning as phosphate shuttles, or components of protein complexes (reviewed in (*34*)).

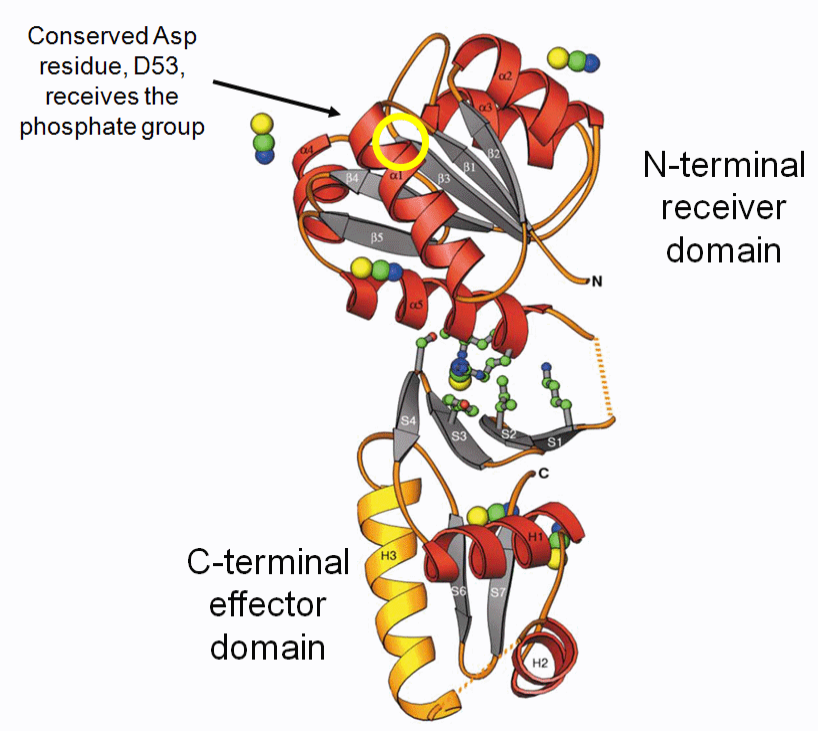


Figure I-3. Structure of DrrD, an example of an RR. The structure of DrrD, a RR from *Thermotoga maritima* serves as an example to understand the modular design of RRs. It is comprised of two domains: a phosphate receiving N-terminal regulatory domain, and a response mediating C-terminal effector domain. The regulatory domain contains a conserved aspartate residue, D53 in DrrD, that is involved in the phosphorelay between the HK and the RR. The effector domain is frequently involved in binding DNA, allowing the RR to regulate expression of specific genes (figure adapted from (*33*)).

H. pylori and two component systems

Bacterial TCSs control a variety of physiological processes in response to changing environmental conditions. *H. pylori* encodes a remarkably small number of TCS proteins in its genome. In contrast to the *E. coli* genome, which contains 62 open reading frames (ORFs) encoding TCS proteins (*35*), the *H. pylori* genome contains only 10 ORFs encoding TCS proteins, with only 3 HKs and 5 RRs that are likely to be involved in transcriptional regulation (Table I-1) (*36-38*). This paucity of regulatory genes in the *H. pylori* genome could reflect the pathogen's tight adaptation to its restricted ecological niche, and also the lack of competition from other microorganisms in the acidic gastric environment (*37, 39*).

Work by Beier and Frank (*37*) indicates that HK *arsS* and RR *arsR,* HK *flgS* and RR *flgR,* and most likely HK *crdS* and RR *crdR* are cognate HK-RR pairs, based on sequence similarities to well-characterized TCSs, the tandem organization of corresponding genes, or direct analysis of phosphotransfer between purified proteins. Furthermore, attempts to generate *H. pylori* mutant strains with disruptions in genes encoding any of the three RRs *arsR, hp1043,* and *hp1021* were unsuccessful, suggesting that these RRs are probably essential for *H. pylori* viability. Two of these RRs, HP1043 and HP1021, are unique in harboring variations in their regulatory domains, with the conserved phosphate-receiving aspartate residues replaced by lysine and serine, respectively (*37*). These RRs are not phosphorylated *in vitro* by any of the known *H. pylori* HKs (*37*) or by the low-molecular-weight phosphate donor acetylphosphate (*40*). Additionally, *H. pylori* strains expressing derivatives of these RRs with mutations in putative phosphate-receiving residues were viable. These data suggest that phosphorylation of HP1043 and HP1021 is not needed for the essential functions of these RRs, and may not occur at all (*40*). Structural data corroborate that the conserved canonical phosphorylation site is unnecessary in HP1043 by demonstrating that even without phosphorylation the protein maintains active conformation (*41*).

Table I-1. Open reading frames in the *H. pylori* genome encoding putative TCS proteins that are most likely involved in transcriptional regulation[a].

Gene encoding RR protein	Gene encoding HK protein[b]
arsR/hp0166 (*jhp0152*)	*arsS/hp0165* (*jhp0151*)
hp1043 (*jhp0381*)	?
hp1021 (*jhp0403*)	?
flgR/hp0703 (*jhp0643*)	*flgS/hp0244* (*jhp0229*)
crdR/hp1365 (*jhp1283*)	*crdS/hp1364* (*jhp1282*)

[a] Nomenclature in the table corresponds to gene names frequently used in literature (*30, 42, 43*); hp# names referring to corresponding genes from *H. pylori* strain 26695; jhp# names referring to corresponding genes from *H. pylori* strain J99. Even though we used *H. pylori* strain J99 for all the our studies, nomenclature based on the *H. pylori* strain 26695 is frequently used in literature, and thus this designation style will be used for the remainder of the thesis, unless indicated otherwise.

[b] Putative cognate pairs of RR-HK are depicted adjacently in the table. *Hp1043* and *hp1021* encode atypical RRs for which cognate HKs have not been identified.

ArsRS two component system

Several recent studies have suggested that ArsS-ArsR (acid responsive signaling sensor/response regulator), responds to acidic conditions and mediates pH responsive transcriptional regulation (*24, 25, 29, 37, 42, 44-47*). First, an *H. pylori arsS* mutant strain exhibits impaired growth at pH 5.0 compared to a wild type *H. pylori* strain (*29*). Second, an *H. pylori* mutant strain with an *arsS* deletion is unable to colonize the stomachs of mice (*45*). Finally, the ArsRS regulon is comprised of numerous genes encoding acid-acclimating proteins, such as subunits of the urease enzyme and many of its accessory proteins, amidases, arginase, and α-carbonic anhydrase (*24, 25, 29, 42, 44, 46, 47*).

Based on the evidence listed above, and current knowledge of other TCSs (*34*), the following model for ArsRS-mediated pH-responsive gene regulation in *H. pylori* has been described (reviewed in (*22*)). Low pH stimulus is sensed by the periplasmic domain of ArsS. This periplasmic domain contains several histidine residues that have a pK_a of ~ 6. These residues can change their protonation state as the periplasm pH falls from 7 to lower than 6, and thus may serve as indicators of environmental pH affecting the activation state of ArsS (*47*). Periplasmic signals that activate ArsS trigger an ATP-dependent autophosphorylation at the conserved histidine residue in the protein's cytoplasmic domain. The environmental signal sensed by ArsS is relayed to ArsR via a phosphotransfer reaction from the phospho-HK to a conserved aspartate residue, D52, in ArsR's N-terminal regulatory domain. Phosphorylation of ArsR induces conformational changes in the protein allowing for altered DNA-binding capabilities of its DNA-binding C-terminal effector domain (paradigm reviewed in (*22, 34*)).

While the *arsS* gene can be deleted without an effect on *in vitro* growth of *H. pylori* at neutral pH (reviewed in (*22*)), the gene for the RR ArsR is essential (*37*). Since a derivative of ArsR containing a mutation in the phosphate-receiving aspartate, D52N, is sufficient for *H. pylori* viability (*40*), it is predicted that ArsR regulates two sets of genes: (i) target genes that are regulated by unphosphorylated ArsR, at least one of which is essential for viability and (ii) target

genes that are non-essential and are regulated by phosphorylated ArsR in response to acidic pH (reviewed in (*22*)). The products of various genes in the latter group contribute to acid acclimation, and thus host colonization, by *H. pylori*. It should be noted that genes belonging to the former set (required for cell viability) have not yet been identified.

ArsRS regulon

One approach for identifying members of the ArsRS regulon has been to isolate DNA sequences that bind to ArsR. This approach resulted in the identification of two operons (*hp1408-hp1412* and *hp0427-hp0423*) and a family of paralogous genes (exemplified by *hp0119*) that are regulated by the ArsRS system (*44*). These genes encode proteins of unknown function.

As another approach to identify members of the ArsRS regulon, whole-genome transcriptional profiles of wild-type *H. pylori* and *arsS* mutant strains have been compared. Such studies have been done following growth of the bacteria at either neutral pH or acidic pH. Whole-genome transcriptional profiling of *H. pylori* strains cultured in low pH conditions identified over 100 genes that were differentially expressed in the wild-type strain compared to an ArsS-deficient mutant (*42*). Transcriptional profiling of *H. pylori* cultured in neutral pH conditions identified a smaller number of genes that were differentially expressed in the wild-type and *arsS* null mutant strains (*48*).

Acid-responsive *H. pylori* genes that are differentially expressed in wild-type and *arsS* mutant strains include amidases (*amiE*, *amiF*) and members of the urease gene cluster (*ureA*, *ureB*, *ureE*, *ureF*, *ureG*, *ureH* and *ureI*) (*20, 21, 24, 42, 46*). Amidases AmiE and AmiF and urease are members of major ammonia producing pathways that help maintain pH near neutrality. In addition to genes directly involved in acid acclimation, the ArsRS pH-responsive regulon is also implicated in the regulation of genes encoding several detoxifying enzymes. These enzymes may contribute to protection against toxic side effects of high intracellular concentrations of bivalent metal ions, whose solubility, and thus bioavailability, is increased under acidic

conditions (*42*). Gel-shift and DNA-footprinting analyses have confirmed that ArsR binds directly to the promoter regions of genes *ureA*, *ureI*, *amiE*, *amiF*, *rocF*, *hypA*, α-carbonic anhydrase, *hp1408*, *hp1119*, and *arsR* itself (*24, 26, 42, 44, 47*), emphasizing the role of this TCS in allowing *H. pylori* to adapt to changes in pH.

Research significance and specific aims

H. pylori relies on a small repertoire of transcriptional factors to regulate gene expression. TCSs have been implicated in regulation of mechanisms essential for *H. pylori* to efficiently adapt to the acidic environment encountered during its infectious cycle (*22, 24, 29, 37, 42, 44-47*). In particular, ArsRS control over expression of genes involved in gastric acid acclimation illustrates the significance of this TCS for *H. pylori*'s adaptation to its unique niche. Thus, interfering with the proper functioning of these systems can strongly attenuate bacterial capacity to inhabit the gastric mucosa, and thereby its capacity to cause disease.

While the importance of this TCS for *H. pylori*'s pathogenicity has been established, the ArsRS regulon has not yet been completely characterized and the mechanisms controlling ArsR mediated gene regulation and acid adaptation are not yet completely understood. Theories have been presented based on the activities of other TCSs (reviewed in (*22*)). However, there is a lack of experimental data to support the hypothesis that ArsR regulates expression of genes involved in acid adaptation in response to low pH stimulus detected by its cognate HK ArsS. Questions such as what are the target genes of unphosphorylated ArsR, and how phosphorylation alters the DNA-binding properties and regulatory effects of ArsR remain unanswered. The goals of our studies were to analyze biochemical and structural properties of the components of this TCS, and to elucidate the mechanisms by which these proteins allow *H. pylori* to adapt to changes in pH and colonize the inhospitable niche of the human stomach.

Aim 1: Characterize the biochemical features of ArsR

In a typical TCS, a phosphorylation-mediated conformational change affecting a large surface of the RR allows for altered inter- and/or intramolecular interactions between the RR and other molecules (*34, 49, 50*). An understanding of a two component system pathway and its biological mechanisms requires deciphering the biochemical features of its components. Thus, the goal of

this study was to express response regulators encoded in the *H. pylori* genome and, in particular, characterize the phosphorylation and oligomeric states of full-length ArsR and the DNA-binding domain of ArsR (ArsR-DBD).

Aim 2: Determine the structure of ArsR

Despite an expanding list of target genes identified for the ArsRS TCS, no structural information is available for either the HK or the RR. Structural analyses of a protein and comparisons with available structural data of related proteins can reveal shared functional elements, allowing us to make predictions about the protein's physiological roles. Additionally, these comparisons can also highlight novel features in the protein of interest that may be significant for its specific biological functions. For example, structural studies on RR HP1043 have verified that this RR contains a well characterized DNA-binding motif known as the winged helix-turn-helix (wHTH) motif, and reinforced the theory that the activity of this atypical RR is phosphorylation-independent (*41*). In this study, we wanted to look at the structural features of RR ArsR. We began by focusing on ArsR's effector domain, ArsR-DBD, that is involved in binding DNA to regulate gene expression.

Aim 3: Characterize the ArsRS regulon

Members of the ArsRS regulon have been identified by isolating DNA sequences that bind to ArsR (*44*) and by analyzing gene expression in *arsS* mutant strains compared to wild-type strains (*42, 48*). Despite these efforts, the ArsRS regulon has not yet been sufficiently characterized. Comparative proteomic analysis performed by our group for wild-type and *arsS* mutant strains of *H. pylori*, grown at either pH 5.0 or pH 7.0 has identified novel putative targets for the ArsRS regulon. In this study, we set out to further characterize this regulon by identifying genes that bind to ArsR, and are thus under direct control of this TCS.

CHAPTER II

EXPRESSION OF *HELICOBACTER PYLORI* RESPONSE REGULATORS AND BIOCHEMICAL CHARACTERIZATION OF ARSR AND ITS DNA-BINDING DOMAIN, ARSR-DBD

Introduction

Two component systems are widely used by prokaryotes as signal transduction mechanisms, which allow cells to adapt to changing environmental conditions. A histidine kinase protein, which is regulated by environmental stimuli, transmits information to a response regulator protein through a phosphorelay event. The majority of RRs are transcription factors, functioning as phosphorylation regulated switches to mediate the activated response of DNA-binding (*34*).

Sequence analysis of the *H. pylori* genome revealed few regulatory genes, with only four open reading frames with homologies to TCS histidine kinase proteins, and only six genes encoding response regulators (*36, 38*). This paucity of TCSs in *H. pylori* suggests tight adaptation of *H. pylori* to its gastric niche, and suggests that these few regulatory proteins ably mediate cellular responses to the changes in the gastric environment (*37*).

In particular, the ArsS-ArsR TCS has been implicated in allowing the bacterium to sense and respond to changing pH through transcriptional regulation of acid-responsive genes (*24, 25, 29, 37, 42, 44-47*). This TCS consists of the HK protein, ArsS, and the RR protein, ArsR. While *arsS* null mutants are viable, attempts to generate *arsR* null mutants have been unsuccessful, suggesting that ArsR is essential for *H. pylori* viability (*37*). Furthermore, ArsR binds directly to the promoter regions of genes encoding members of major ammonia producing pathways, such as amidases and urease, that help maintain pH near neutrality and aid bacterial adaptation to the acidic human stomach (*24, 42*).

In a classic TCS, phosphorylation of the RR induces a conformational change that affects a large surface of the RR. The altered molecular surface allows the RR to participate in a specific set of inter- or intramolecular interactions, sometimes leading to RR dimerization (*34, 49, 50*). An understanding of a two component system pathway requires analysis of the biochemical features of its components in conjunction with their biological activities.

The goal of this study was to express response regulators encoded in the *H. pylori* genome and, in particular, characterize the phosphorylation and oligomeric states of full-length ArsR and the DNA-binding domain of ArsR (ArsR-DBD). This chapter describes our results from these studies: His_6-tagged full length and DBDs of response regulators HP0166 (ArsR), HP1043, HP1021, CrdR, and the DBD of FlgR were overexpressed in *E. coli*. The full-length ArsR and its DBD were purified from cell lysate. ^{31}P NMR experiments conducted on full-length ArsR demonstrated that the majority of the purified full-length protein is in the unphosphorylated form. Furthermore, ArsR and ArsR-DBD behaved as monomers in size exclusion chromatography experiments.

Methods

Plasmid construction: Selected RRs encoded by *H. pylori* strain J99 and their corresponding DBDs, mapped based on comparison to other known bacterial response regulators, such as PhoB (UniProt KM accession number P0AFJ5) and OmpR (UniProt KM accession number P0AA16), are listed in Table II-1. The full-length RR genes and fragments encoding the corresponding DBDs were amplified by polymerase chain reaction (PCR) from *H. pylori* strain J99 using genomic DNA as template and pairs of primers including 5' BamHI and 3' KpnI restriction endonuclease sites (Table II-2). For a detailed list of nucleotide and amino acid sequences of selected RRs and their corresponding DBDs, see Appendix B. Purified PCR products were digested with KpnI and BamHI restriction enzymes (Promega, Madison, WI) and ligated into linearized pET-BNK, a modified pET vector (including fragments from plasmids pET32-LIC (Novagen, Madison, WI) and pQE-30 (Qiagen, Valencia, CA)) developed specifically for expressing NMR protein targets. The vector contains 5' coding sequence for an N-terminal purification tag MRGSHHHHHHGS in frame with the insert coding for desired proteins.

Testing protein expression of response regulators and their DNA-binding domains: E. coli BL21 (DE3) cells were transformed with plasmids designed to overexpress response regulators and their DBDs. Cells were grown in 5 mL of LB media containing glucose (2 g/L) and ampicillin (50 mg/L). When the culture reached an $O.D._{600}$ of 1.0, 1 mL sample of culture was collected as the 'before induction' sample. Protein expression was induced for 2 h with 0.8 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and 1 mL sample of culture was collected as 'after induction' sample. Cells were collected by 1 min centrifugation at 14,000 g, suspended in B-PER II (bacterial protein extraction reagent, Pierce, Rockford, IL), and then lysed by sonication, 3 times for 5 s on ice. Lysate samples were electrophoresed on a 10% SDS-polyacrylamide gel and stained with Coomassie blue R-250 stain.

Table II-1. Response regulators of *H. pylori* overexpressed in *E. coli* as His_6-tagged recombinant proteins.

Gene encoding RR protein	Number of amino acids in full length protein	UniProtKM Accession Number	Number of amino acids in corresponding DBD[a]
arsR (*hp0166*)	225	Q9ZMR6	103
hp1043	223	Q9ZM42	100[b]
hp1021	298	Q9ZM20	167
flgR	381[c]	Q9ZLD7	151
crdR	213	Q9Z371	93

[a] DBD proteins comprise the final residues of the full length proteins as listed. For example, ArsR-DBD sequence corresponds to the C-terminal 103 residues of the full length ArsR.
[b] HP1043-DBD comprised of residues 115-214 of the full length protein.
[c] Full length FlgR was not overexpressed in *E. coli*.

Table II-2. Oligonucleotide primers used to amplify RR genes of interest from *H. pylori* strain J99. Pairs of primers used to amplify full length proteins are designated by gene number (J99 strain nomenclature (*36, 38*)) and Forw or Rev indicating forward or reverse primers, respectively. Primers used to amplify DBDs include the Rev primer used for the full length protein (except HP1043 – reverse primer is DBD381Rev) and the forward primer designated with DBD marked preceding the gene number as before.

Gene of interest	Primer	Primer sequence (5' to 3')
arsR (*hp0166*)	152Forw	GCGGGATCCATGATAGAAGTTTTAATGATA GAAGATGATATAG
	152Rev	GCGGGTACCTCAGTATTCTAATTTATAACC AATCCCTCT
arsR-DBD (*hp0166-DBD*)	DBD152Forw	GCGGGATCCGAAGAGGTGAGTGAGCCAGG C
hp1043	381Forw	GCGGGATCCCGGGTTCTACTGATTGAAAAA AATTCTGTTTT
	381Rev	GCGGGTACCAAAAATTACGAAGTTTATTCT TCACACG
hp1043-DBD	DBD381Forw	GCGGGATCCTTTTGGGGTTCTAATGTGATTG AA
	DBD381Rev	GCGGGTACCTTAAAAACGATAGCCTCTGCG
hp1021	403Forw	GCGGGATCCAAAATCTTAATCATTGAAGAT GATTTAGC
	403Rev	GCGGGTACCGGGCTTAAGGATTTTCTACTT GC
hp1021-DBD	DBD403Forw	GCGGGATCCTTGCCCTTGCCTAGAGATTTTT
flgR-DBD	DBD643Forw	GCGGGATCCACGATCTTTTTAGATGAAATC GCT
	643Rev	GCGGGTACCCTACCTTTCCAAAAACAAATC TTTCT
crdR	1283Forw	GCGGGATCCCAGAAAAAGATTTTTTTACTA GAAGACGATTAC
	1283Rev	GCGGGTACCTTTTCATAGTGGGTTAAAGCG ATAG
crdR-DBD	DBD1283Forw	GCGGGATCCGATGATCCGATAGAAATCATG CC

Expression and purification of Ars and ArsR-DBD: Transformed *E. coli* BL21 (DE3) cells were grown in 4.2 L of LB media containing glucose (2 g/L) and ampicillin (50 mg/L). When the culture reached an $O.D._{600}$ of 1.0, protein expression was induced for 2 h with 0.8 mM IPTG. Cells were collected by 8 min centrifugation at 12,000 g, suspended in 0.02 M Na_2HPO_4, 0.5 M NaCl, pH 7.6, and then lysed by sonication, 6 times for 30 s on ice. Both 1 mM phenylmethylsulfonylfluoride (PMSF) and 5 mM Tris(2-carboxyethyl)phosphine (TCEP) hydrochloride were added prior to sonication. Following sonication, the preparation was centrifuged at 31,000 g for 20 min. The supernatant (~100 mL) was collected and applied to a 30 mL metal affinity chromatography column (His-Bind, Novagen) charged with Ni^{2+}. The column was washed with a 0-0.08 M imidazole gradient to remove proteins bound non-specifically to the column, and the protein of interest was eluted with a 0.08-1.0 M imidazole gradient over 45 mL. Five mL fractions corresponding to peaks of interest were collected, pooled, and concentrated to a volume of 1 mL in Amicon Ultra-15 centrifugal filters [10 kDa molecular weight cut-off (MWCO) membrane for the full length proteins and a 5 kDa MWCO membrane for the DBDs]. Proteins were further purified by applying concentrated samples to size exclusion chromatography column, Sephacryl S-100 FPLC column (GE Healthcare, Piscataway, NJ) that was run at 4 ˚C in 0.02 M Na_2HPO_4, 0.5 M NaCl, 1 mM TCEP (pH 7.6) and a flow rate of 0.4 mL/min, with ultraviolet (UV) absorption measured at 214 nm. Fractions with proteins of interest were collected, pooled and concentrated to a volume of 1 mL in Amicon Ultra-15 centrifugal filters, as described above. The purity of the proteins was assessed by Tricine sodium-dodecyl-sulfate polyacrylamide gel electrophoresis (SDS–PAGE) on 10% gels. ^{15}N- and $^{15}N/^{13}C$-labeled samples were produced by a process similar to that described above, but cells were grown in M9 medium with the addition of ^{13}C-labeled glucose and ^{15}N-labeled ammonium chloride (CIL, Andover, MA).

NMR Analysis of ArsR and ArsR-DBD: The one-dimensional (1D) ^{1}H NMR spectra for ArsR and ArsR-DBD were acquired on an Avance 600 Bruker spectrometer with a triple resonance gradient probe and a cryoprobe. 1D ^{31}P NMR spectra for ArsR and inorganic phosphate were acquired on an Avance 600 Bruker spectrometer with a room temperature quadruple resonance gradient probe (Bruker, Billerica, MA). Sample protein concentrations were ~0.2 mM (full-length protein) and ~0.4 mM (DBD) in 0.5 M NaCl, 0.02 M Na_2HPO_4, 1 mM TCEP, pH 7.6 for 1D ^{1}H NMR, and ~0.5 mM (full-length protein) and ~ 20 mM (inorganic phosphate) in 10 mM Tris, pH 7.5 for 1D ^{31}P NMR. Spectra were collected at 27 °C for unlabeled (natural abundance); ^{15}N-labeled; or $^{15}N/^{13}C$-labeled protein samples. The data were processed using XWINNMR and TOPSPIN software (Bruker).

Analytical Size Exclusion Chromatography: ArsR and ArsR-DBD were analyzed by gel filtration on a Sephacryl S-100 FPLC column (GE Healthcare, Piscataway, NJ) run at 4 ˚C in 0.02 M Na_2HPO_4, 0.5 M NaCl, 1 mM TCEP (pH 7.6) and a flow rate of 0.4 mL/min, with ultraviolet (UV) absorption measured at 214 nm. Molecular weights of the proteins (M_r) were determined by calculating their partition coefficient (K_{av}) values and using a calibration curve, plotting logarithmic values of M_r against calculated K_{av} values for a set of standard globular monomeric proteins [bovine serum albumin (66 kDa), ovalbumin (45 kDa), and myoglobin (16.7 kDa)], run under identical conditions. K_{av} values were calculated using the following formula:

$K_{av} = (V_e - V_o)/(V_t - V_o)$, where V_e = elution volume, V_o = column void volume, V_t = total column volume.

Results

Expression of *H. pylori* response regulators in *E. coli*

DNA sequences encoding response regulators ArsR, HP1043, HP1021, CrdR, and the sequences corresponding to their DBDs, and the DBD of FlgR were PCR amplified from *H. pylori* strain J99 and ligated into the expression vector, a linearized plasmid, pET-BNK. To verify correct cloning of the gene of interest, plasmid DNA purified from liquid cultures was sequenced using primers flanking the insert. *E. coli* clones positive for transformation with plasmid with gene of interest were tested for protein expression as described above (Figures II-1, II-2). Clones expressing the highest amounts of soluble protein were selected. For our studies, we focused on RR ArsR and its DNA-binding domain, ArsR-DBD.

The structure of isolated ArsR-DBD is similar to that of the domain in full-length ArsR

The full-length His_6-tagged ArsR protein (237 residues) and a His_6-tagged ArsR fragment corresponding to the DBD (115 residues) were overexpressed in *E. coli,* recovered from soluble fractions of the cell lysate, and purified by metal affinity chromatography, followed by size exclusion chromatography. ArsR and ArsR-DBD protein samples were concentrated to ~0.2 mM and ~0.4 mM, respectively, and analyzed by NMR. ArsR was produced in natural abundance, and ^{15}N-enriched forms. ArsR-DBD was produced in natural abundance, and in ^{15}N- and $^{15}N/^{13}C$-enriched forms. One-dimensional ^{1}H-NMR spectra of ArsR and ArsR-DBD (Figures II-3, II-4) showed well-dispersed peaks in regions characteristic of amide protons (6-10 ppm) and aliphatic protons (0-4 ppm), indicative of folded protein. The ^{1}H-^{15}N correlation NMR spectrum of ArsR-DBD exhibited good chemical shift dispersion. A comparison of ^{1}H-^{15}N HSQCs of full-length ArsR and ArsR-DBD showed that ArsR-DBD amide proton peaks aligned with ArsR amide peaks, indicating that the structure of the isolated ArsR-DBD is very similar to that domain in the full-length protein (Figure II-5).

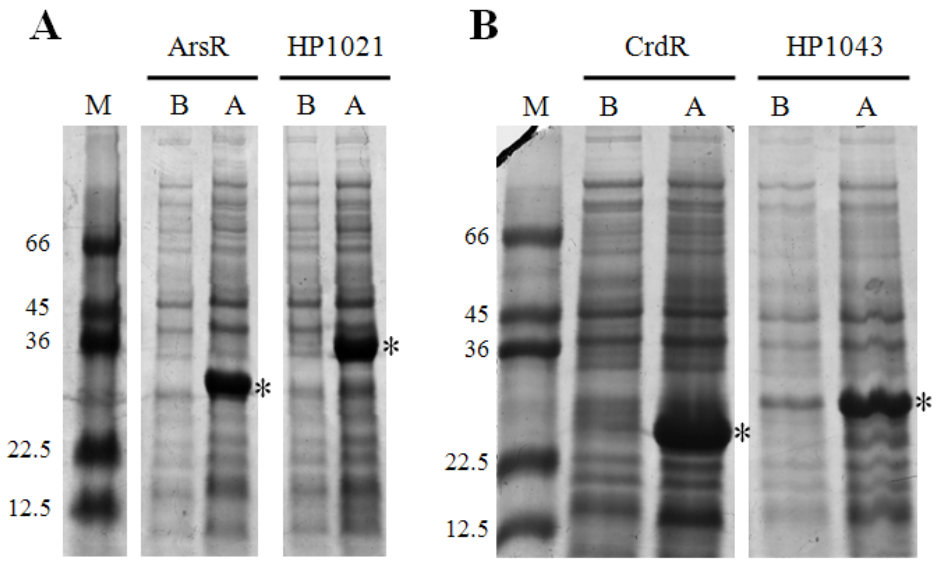


Figure II-1. Overexpression of full length RRs from *H. pylori*. Cell lysates from cultures of *E. coli* transformed with plasmids to overexpress specific *H. pylori* response regulators were collected before and after induction of protein production with IPTG. Lysates before (lane B) and after (lane A) induction were electrophoresed with a protein size marker (lane M) on a 10% SDS-polyacrylamide gel and stained with Coomassie blue stain. Proteins of interest overexpressed are indicated with an asterisk. **A.** Panel A shows the overexpression of response regulators ArsR, and HP1021. **B.** Panel B shows the overexpression of response regulators CrdR, and HP1043.

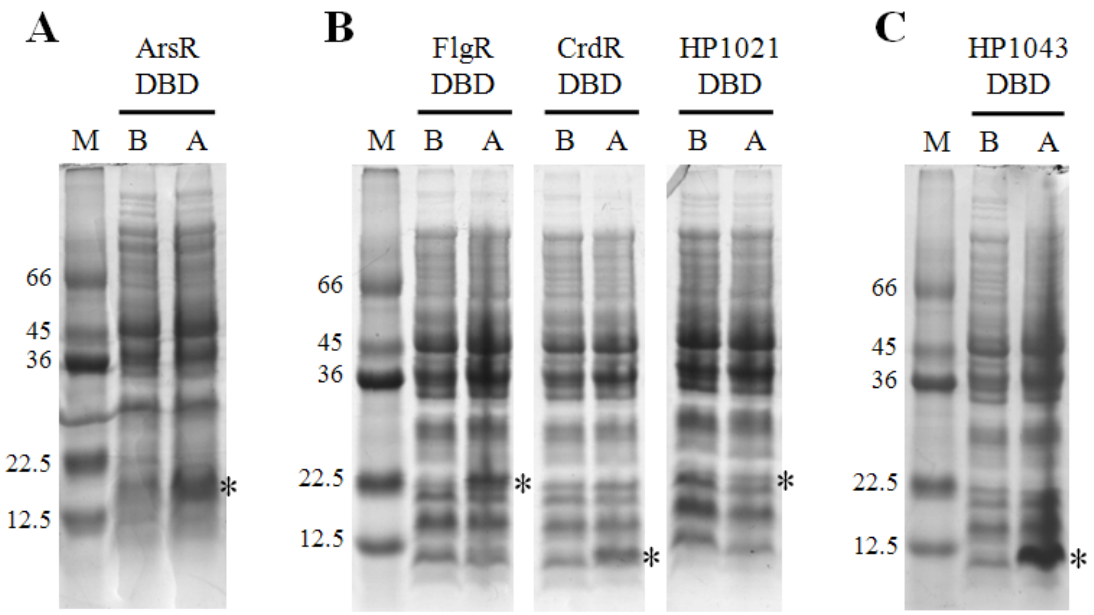


Figure II-2. Overexpression of DNA-binding domains of RRs from *H. pylori*. Cell lysates from cultures of *E. coli* transformed with plasmids to overexpress DBDs of specific *H. pylori* response regulators were collected before and after induction of protein production with IPTG. Lysates before (lane B) and after (lane A) induction were electrophoresed with a protein size marker (lane M) on a 10% SDS-polyacrylamide gel and stained with Coomassie blue stain. Proteins of interest overexpressed are indicated with an asterisk. **A.** Panel A shows the overexpression of ArsR-DBD. **B.** Panel B shows the overexpression of DBDs of response regulators FlgR, CrdR, and HP1021. **C.** Panel C shows the overexpression of HP1043-DBD.

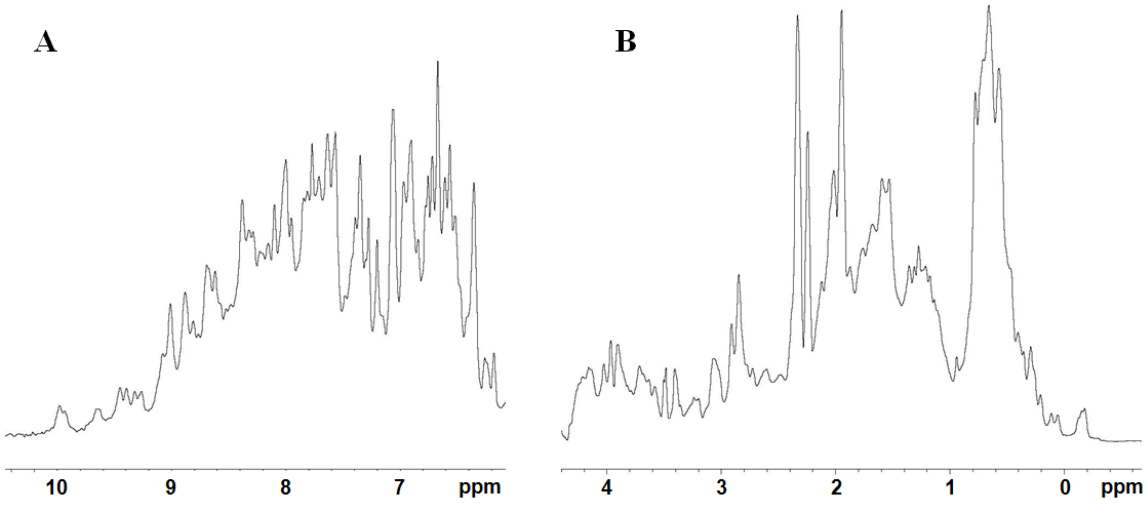


Figure II-3. One-dimensional ^{1}H NMR spectra of ArsR showing amide and aliphatic peaks. 256 scans on a 600 MHz spectrometer were acquired for full length ArsR (~0.2 mM) in 0.5 M NaCl, 0.02 M Na_2HPO_4, 1 mM TCEP, pH 7.6, 27 °C. **A.** Peaks in region characteristic of amide protons (6-10 ppm) are shown. **B.** Peaks in region characteristic of aliphatic protons (0-4 ppm) are shown.

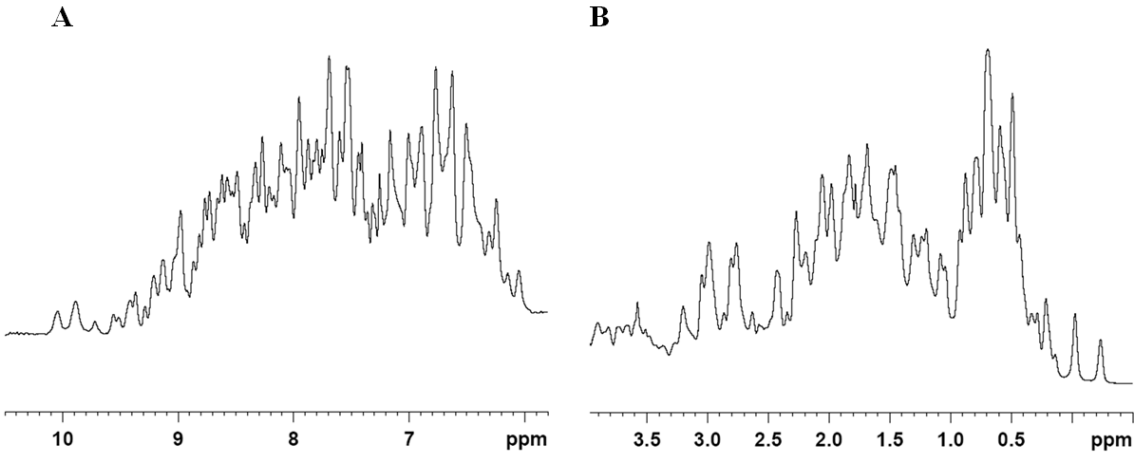


Figure II-4. One-dimensional ^{1}H NMR spectra of ArsR-DBD showing amide and aliphatic peaks. 16 scans on a 600 MHz spectrometer were acquired for ArsR-DBD (~0.4 mM) in 0.5 M NaCl, 0.02 M Na_2HPO_4, pH 7.6, 27 °C. **A.** Peaks in region characteristic of amide protons (6-10 ppm) are shown. **B.** Peaks in region characteristic of aliphatic protons (0-4 ppm) are shown.

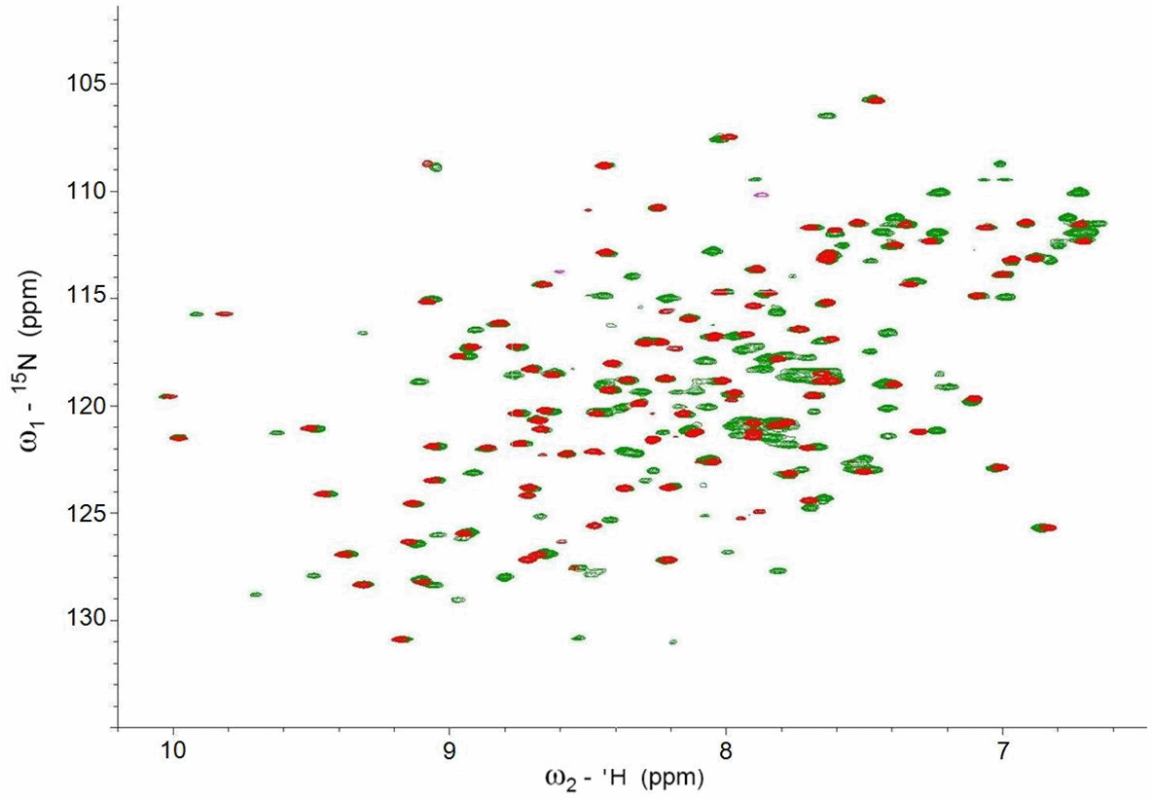


Figure II-5. Two-dimensional ^{1}H-^{15}N NMR spectra of ArsR and ArsR-DBD. Comparison of two dimensional ^{1}H-^{15}N HSQC spectra of ArsR (green) with ArsR-DBD (red). Protein concentrations were ~0.2 mM in 0.5 M NaCl, 0.02 M Na_2HPO_4, 1 mM TCEP, pH 7.6 and ~0.4 mM in 0.5 M NaCl, 0.02 M Na_2HPO_4, pH 7.6 for full length ArsR and ArsR-DBD, respectively. Spectra were collected at 800.50 MHz, with 256 scans at 27 °C for ArsR, and at 600.13 MHz, with 4 scans at 27 °C for ArsR-DBD.

Purified ArsR is not phosphorylated

Full length ArsR protein contains a putative phosphate-receiving aspartate residue, D52. Thus, ^{31}P NMR experiments were conducted on purified full-length ArsR to determine whether the recombinant protein isolated from *E. coli* lysate was phosphorylated or not. Initial studies conducted on a control sample of 20 mM inorganic phosphate in 10 mM Tris, pH 7.5 buffer revealed a peak at ~ 1 ppm, corresponding to the phosphorus in the sample (Figure II-6A). Data for this sample were acquired with 16 scans collected on an Avance 600 Bruker spectrometer with a room temperature quadruple resonance gradient probe. The signal to noise ratio (SNR) calculated for this sample was ~153. Based on the analysis of numerous spectra acquired on this instrument, we determined that an unequivocally identifiable peak would have a minimum SNR of 3 for samples observed under similar conditions. Thus, a signal for a 51-fold weaker than a 20 mM phosphate sample (~0.4 mM phosphate) would be detected above the noise level by this instrument. Furthermore, SNR for a peak is proportional to the square-root of the number of scans acquired. Thus, if the number of scans were increased from 16 to 1024, a further 8-fold reduction of the 0.4 mM peak, or ~0.05 mM phosphate, would be detected experimentally.

The directly detected ^{31}P NMR experiments conducted on purified full-length ArsR, ~0.5 mM in 10 mM Tris, pH 7.5 buffer, did not reveal any detectable phosphate signal (Figure II-6B). Data for this sample were acquired with 1024 scans on the same instrument, and processed as for the inorganic phosphate sample. Based on the calculated sensitivity of the instrument, and the absence of any phosphorus peaks from the ArsR sample, we determined that at least 90% of the purified ArsR sample is in the unphosphorylated form.

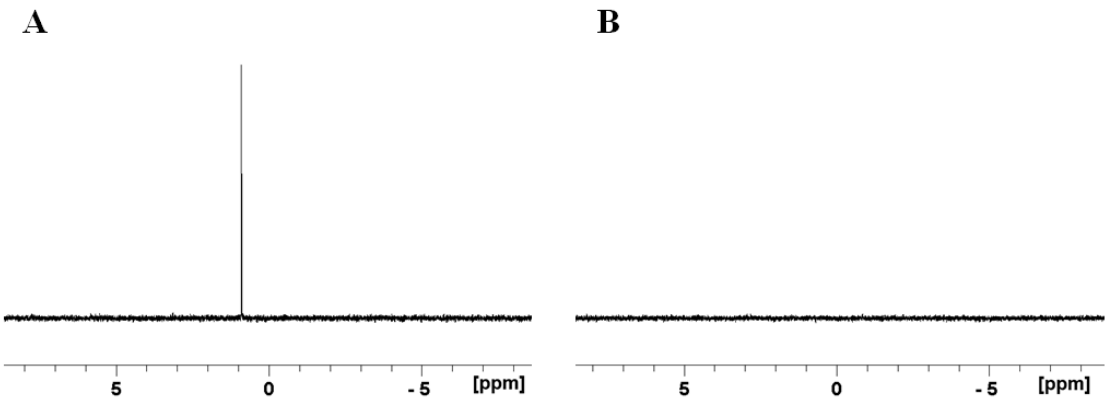


Figure II-6. One-dimensional ^{31}P NMR spectra of inorganic phosphate and ArsR. 242.94 MHz one dimensional ^{31}P NMR spectra were acquired for inorganic phosphate and ArsR in 10 mM Tris, pH 7.5. Spectra were collected at 27 °C. **A.** 1D ^{31}P NMR spectrum of inorganic phosphate (20 mM) displays a phosphorus signal at ~1 ppm, 16 scans. **B.** 1D ^{31}P NMR spectrum of natural abundance ArsR (~0.5 mM) demonstrates no detectable phosphorus signal, 1024 scans.

ArsR and ArsR-DBD behave as monomers in solution

The full-length His_6-tagged ArsR protein (237 residues) and a His_6-tagged ArsR fragment corresponding to the DBD (115 residues) were overexpressed in *E. coli*, recovered from soluble fractions of the cell lysate, and purified by metal affinity chromatography. Proteins were analyzed by gel filtration experiments on a Sephacryl S-100 FPLC column. A calibration curve was prepared with globular monomeric proteins, BSA, ovalbumin, and myoglobin run under identical conditions as the proteins of interest, and plotting the logarithmic values of their molecular weights against the calculated K_{av} values:

$K_{av} = (-0.377 * \log MW) + 1.999$

Based on their respective calculated K_{av} values and the calibration curve, ArsR and ArsR-DBD migrated with molecular masses of ~32 kDa (expected monomer size ~27 kDa) and ~19 kDa (expected monomer size ~13 kDa), respectively (Figure II-7). The larger than expected apparent molecular sizes were most likely caused by the 17-residue disordered N-termini of proteins. These data suggest that both ArsR and ArsR-DBD are monomeric in solution at concentrations below 0.1 mM.

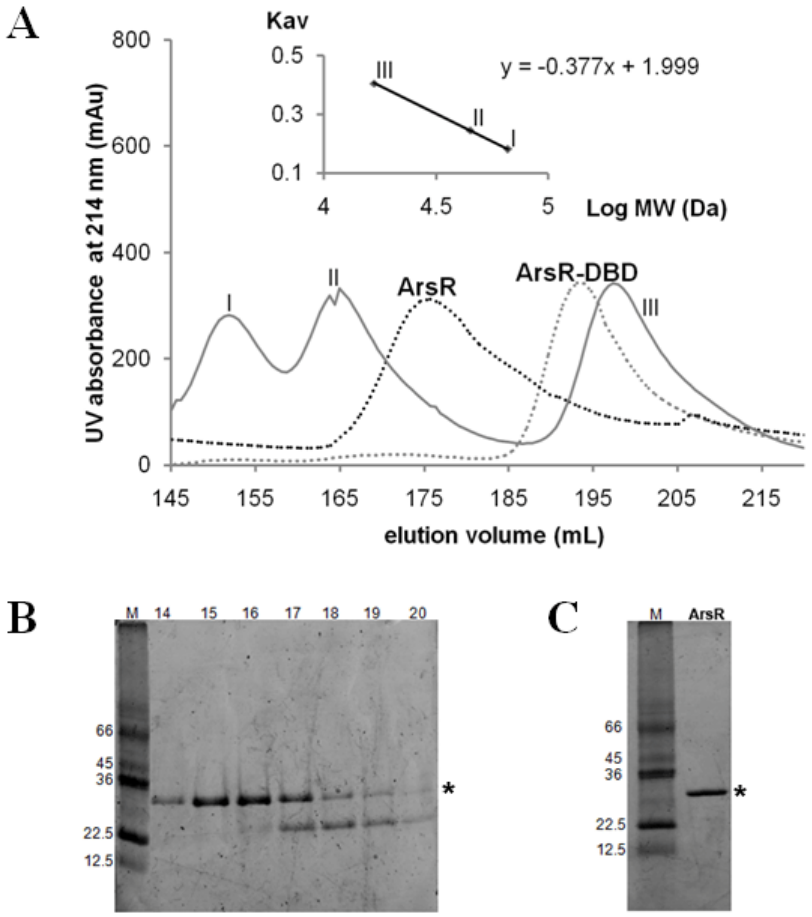


Figure II-7. Size exclusion chromatography of ArsR and ArsR-DBD. A. A mixture of globular monomeric standard proteins (elution profile shown as grey curve) was run on a Sephacryl S-100 FPLC column: (I) BSA (66 kDa), (II) ovalbumin (45 kDa) and (III) myoglobin (16.7 kDa). The ultraviolet (UV) absorption of the elution was measured at 214 nm. ArsR (black dotted curve) and ArsR-DBD (grey dotted curve) purified from the Ni-column were concentrated and run separately on the FPLC column. Logs of protein molecular weights were plotted against K_{av} values, calculated based on elution profiles of the molecular mass standards. A calibration curve was generated (inset) and used to determine the mass of purified ArsR and ArsR-DBD.
B. Panel B shows a representative gel demonstrating the elution of ArsR from the size-exclusion chormatogrpahy column. 10 µL samples of the 5 mL fractions were electrophoresed with a protein size marker (lane M) on a 10% SDS-polyacrylamide gel and stained with Coomassie blue stain. Band corresponding to ArsR is indicated with an asterisk. **C.** Panel C shows a 2 µL sample of ~0.2 mM ArsR electrophoresed with a protein size marker (lane M) on a 10% SDS-polyacrylamide gel and stained with Coomassie blue stain. Band corresponding to ArsR is indicated with an asterisk. The concentrated sample was generated by pooling fractions from the size exclusion chromatography column in an Amicon Ultra-15 centrifugal filter [10 kDa molecular weight cut-off (MWCO) membrane].

Discussion

Most prokaryotic signal transduction schemes use a phosphotransfer system involving a sensor histidine kinase and a response regulator. These components possess modular domains that can be integrated into proteins and pathways in a variety of ways to mediate signal specific responses in organisms such as *H. pylori* (*34*). The *H. pylori* genome encodes few TCS proteins that can allow the bacteria to adapt to changes in the gastric environment (*36, 38*). Within these TCSs, the response regulators are the terminal component of the pathway, executing the adaptive response, most commonly transcriptional regulation, to environmental stimuli. Response regulators differ in their mechanisms of activation. For example, RRs vary in their dependence on phosphorylation and oligomerization for transcriptional regulation. Upon phosphorylation, some RRs dimerize in solution, whereas others remain monomeric and subsequently dimerize after binding to specific DNA sequences (*49-52*). At present, relatively little is known about the effects of phosphorylation and oligomerization on the activity of RRs from *H. pylori*, in particular, ArsR. RRs have been shown to bind DNA as monomers, dimers, or even higher-order oligomers (*34, 53, 54*). Furthermore, while there is evidence for the requirement of phosphorylation of ArsR for its DNA-binding functions, the experimental conditions used by different groups for the analysis vary significantly. For example, Beier and coworkers (*42, 44*) phosphorylate purified recombinant protein with the phosphate donor, acetylphosphate, prior to their DNA-binding studies. Sachs and coworkers (*47*) conducted DNA-binding assays comparing binding of a promoter sequence from a known ArsR target *hp1408* (*44*) with wild type purified ArsR, wild type purified ArsR phosphorylated *in vitro* by carbamoyl phosphate, and a non-phosphorylatable ArsR mutant, ArsR-D52N. Similar DNA-binding patterns were observed between wild type purified ArsR and wild type purified ArsR phosphorylated *in vitro* by carbamoyl phosphate, versus ArsR-D52N. Based on these findings, Sachs and coworkers (*47*) claim that the purified recombinant protein is functionally phosphorylated and requires no pre-

incubation with a phosphate donor for their DNA-binding studies. Inconsistencies in these data could be a consequence of the variability in the lifetimes of phospho-Asp in RRs, which can range from seconds to hours. Frequently, autophosphatase activity within the protein decreases the lifetime of phospho-RR (reviewed in (*34*)). Thus, despite successfully phosphorylating a RR, the inability to preserve the protein in its phospho-RR state can hinder the analysis of phosphorylation induced changes in RR functions.

In view of the variability in biochemical characteristics of RRs and their effects on protein function, we wanted to express *H. pylori* RRs and characterize the biochemical features of proteins of interest prior to further structural and functional analysis. In this study, we successfully cloned response regulators encoded in the *H. pylori* genome into pET-BNK and transformed *E. coli* cells to overexpress His_6-tagged full length and DBDs of ArsR, HP1043, HP1021, CrdR, and the DBD of FlgR (Figures II-1, II-2). The full-length ArsR and its DBD were purified from cell lysate via metal affinity chromatography followed by size exclusion chromatography. The proteins were generated in natural abundance, ^{15}N- and $^{15}N/^{13}C$-labeled forms. Protein folding was tested via 1D ^{1}H NMR. Well dispersed peaks for both ArsR and ArsR-DBD in both upfield (0-4 ppm, characteristic of aliphatic peaks) and downfield (6-10 ppm, characteristic of amide peaks) regions of the spectra indicate that the purified protein is folded, with peak separation as a marker for the diverse microenvironments experienced by the proteins' atoms (Figures II-3, II-4). Isolated methyl group signals seen upfield of 0.5 ppm are sensitive and reliable probes for folded protein.

A comparison of the 2D ^{1}H-^{15}N HSQC spectra of ArsR and ArsR-DBD showed that ArsR-DBD amide proton peaks aligned with the corresponding amide peaks of ArsR (Figure II-5). This near perfect overlap of backbone signals indicates that the structure of the isolated ArsR-DBD is very similar to that domain in the full-length protein and expression of the DBD alone does not seem to alter protein folding. These data are in agreement with the consensus view that

while signaling occurs via conserved core structures involved in phosphorelay, RRs are modular proteins with distinct domains maintaining independent structural and functional features (*34*).

Prokaryotes employ numerous variations of the basic scheme of TCS phosphorelay to execute specific responses corresponding to a range of stimuli. While the regulatory domains of RRs contain a conserved phosphate receiving core, the effects of phosphorylation vary. Phosphorylation of the aspartate residue in the N-terminal receiver domain produces a high-energy acyl phosphate. Energy within this acyl-phosphate bond may be used to drive conformational changes in the RR and alter its surface, allowing for specific inter- or intramolecular interactions through which the response is elicited. These interactions include protein-DNA contacts, and protein-protein contacts, such as formation of RR dimers or higher-order oligomers (reviewed in (*34*)). Although typically members of the OmpR/PhoB subfamily of RRs bind to target DNA as dimers (*41, 53*), this dimerization may not be required for RR-target DNA interactions, as suggested by Rhee and coworkers (*54*). Additional variability within the TCS phosphorelay scheme arises from the different lifetimes of phospho-Asp in RRs, which can range from seconds to hours. Furthermore, while generally accepted as the "switch on" event, phosphorylation does not necessarily correspond to activation, but in fact may serve the opposite role of turning a function off (reviewed in (*34*)).

Based on these data, protein activity can vary greatly based on its phosphorylation and oligomeric state. Thus, we conducted NMR and gel filtration experiments to characterize ArsR and ArsR-DBD's phosphorylation and oligomeric state. Since ArsR-DBD does not contain the phosphate receiving residue, D52, ^{31}P NMR experiments were conducted on full-length ArsR alone. The absence of a phosphorus signal from purified ArsR sample (Figure II-6) demonstrated that the majority of the purified protein is in the unphosphorylated form. These data suggest that the cellular conditions in *E. coli*, in which our proteins were overexpressed, may not promote phosphorylation of ArsR. For example, the *E. coli* strain in which our proteins were overexpressed may not express an appropriate HK protein that could phosphorylate ArsR.

Alternatively, the protein may be phosphorylated in *E. coli*, but the lifetime of the phosphorylated state may be short due to intrinsic autophosphatase activity. Thus, at least 90% of the purified ArsR is unphosphorylated.

In order to characterize the oligomeric states of ArsR and ArsR-DBD, the proteins were applied to a size exclusion chromatography column. ArsR and ArsR-DBD migrated with molecular masses of ~32 kDa (expected monomer size ~27 kDa) and ~19 kDa (expected monomer size ~13 kDa), respectively (Figure II-7). These data suggest that both ArsR and ArsR-DBD behave as monomers in solution at concentrations below 0.1 mM. The observation that ArsR is monomeric in solution in its unphosphorylated state allows us to hypothesize that phosphorylation induced changes in oligomeric state could regulate ArsR's DNA-binding activity, as is suggested for PhoB (*53*). Alternatively, it is possible that one molecule of ArsR binds to its target DNA sites as a monomer, and subsequently recruits another ArsR molecule to form a dimer on the target DNA, as is suggested for OmpR (*54*). Phosphorylation mediated regulation of ArsR could still involve promotion of a dimeric interaction between RR molecules, but in this model, the monomeric form of the RR would not depend on phosphorylation for its ability to bind DNA.

CHAPTER III

STRUCTURAL ANALYSIS OF ARSR-DBD

Introduction

The *H. pylori* ArsR-ArsS two-component signal transduction system, comprised of a sensor histidine kinase (ArsS) and a response regulator (ArsR), has been implicated in allowing the bacteria to regulate gene expression in response to acidic pH (*24, 25, 29, 37, 42, 44-47*). A low pH stimulus is likely to be transduced to ArsR via a phosphotransfer reaction from phospho-ArsS to a conserved aspartate residue, D52, in ArsR's N-terminal regulatory domain. This phosphorylation of ArsR is thought to induce conformational changes in the protein, allowing for altered DNA-binding capabilities of its DNA-binding C-terminal effector domain (paradigm reviewed in (*22, 34*)).

A derivative of ArsR with a D52N mutation is not phosphorylated *in vitro* by ArsS (*40*). In contrast to an *arsR* null mutant (which is non-viable), a mutant *H. pylori* strain expressing the D52N form of ArsR is viable (*37, 40*). These findings suggest that this mutant form of ArsR has a function that is sufficient for cell viability. Based on these data, it seems likely that there are two sets of target genes for ArsR. One set of genes in the ArsR regulon is presumed to be essential for cell viability, and regulation of these genes can be accomplished by a non-phosphorylated form of ArsR (exemplified by the D52N mutant protein) (*37, 40*). A second set of genes in the ArsR regulon is not required for cell viability, and the regulation of these genes occurs by a pathway involving a phosphorylated form of ArsR (i.e. requiring the cognate HK, ArsS) (*24, 26, 42, 44, 47, 48*).

Despite an expanding list of target genes identified for the ArsRS TCS, no structural information is available for either the HK or the RR. In this study, we sought to determine the structural features of the RR, ArsR. We focused on ArsR's effector domain, ArsR-DBD,

involved in mediating DNA contacts to regulate gene expression. This chapter describes the expression and purification of the full-length ArsR protein and the DNA-binding domain of ArsR (ArsR-DBD), and the analysis of the tertiary structure of the ArsR-DBD using solution nuclear magnetic resonance (NMR) methods. As described in this chapter, the structure of ArsR-DBD consists of an N-terminal four-stranded β-sheet, a helical core, and a C-terminal β-hairpin. The overall tertiary fold of the ArsR-DBD is most closely related to DBD structures of the OmpR/PhoB subfamily of bacterial response regulators. However, the orientation of the N-terminal β-sheet with respect to the rest of the DNA-binding domain is substantially different in ArsR compared to the orientation in related response regulators. Molecular modeling of an ArsR-DBD-DNA complex permits identification of protein elements that are predicted to bind target DNA sequences and thereby regulate gene transcription in *H. pylori*.

Methods

Plasmid construction: The ArsR protein encoded by *H. pylori* strain J99 comprises 225 amino acids. The DNA binding domain of this protein, mapped based on comparison to other known response regulators, such as PhoB and OmpR, comprises 103 amino acids, beginning at E123 and ending at Y225. The full-length *arsR* protein coding sequence and a fragment encoding the DNA-binding domain were amplified from *H. pylori* strain J99 and ligated into a linearized pET-BNK *E. coli* expression vector as described in Chapter II.

Expression and purification of ArsR and ArsR-DBD: Protein samples were prepared from transformed *E. coli* BL21 (DE3) cells as described in Chapter II.

NMR Structure Calculation: NMR spectra were acquired for ArsR and ArsR-DBD on an Avance 600 Bruker spectrometer with a triple resonance gradient probe and a cryoprobe (Bruker, Billerica, MA). Protein samples were ~0.2 mM (full-length protein) and ~0.4 mM (DBD) in 0.5 M NaCl, 0.02 M Na_2HPO_4, 1 mM TCEP, pH 7.6. Spectra were collected at 27 °C for unlabeled (natural abundance); ^{15}N-labeled; and $^{15}N/^{13}C$-labeled protein samples. The data were processed using XWINNMR and TOPSPIN software (Bruker) and analyzed with the Sparky suite (Goddard TD and Kneller DG, Sparky 3, UC, San Francisco). The assignment of backbone resonances was completed using data from 2D ^{1}H-^{15}N HSQC, 2D NOESY, 3D CBCANH, and 3D CBCA(CO)NH experiments. The side chain resonance assignments were completed using data from 3D HCC-TOCSY, 3D HHC-TOCSY, 3D H(CC)(CO)NH, 3D HCC(CO)NH, 3D HBHA(CO)NH, 3D ^{15}N–edited NOESY, and 3D ^{13}C–edited NOESY experiments (for references see (*55*)). The chemical shifts of H_α, C_α, C_β and C' were analyzed with chemical shift index software (*56*) to produce a prediction of secondary structure elements. J-coupling constants calculated from HNHA experiments were used to determine φ angle constraints for structure

calculations. The chemical shifts of H_α, C_α, C_β, C' and N were also analyzed to calculate angle constraints to be used for structure calculations using TALOS software (*57*). The structures were calculated using the CYANA version 2.1 software package (*58*). Automatic calibration was used to convert the NOE peak intensities into distance constraints. The final calculations were performed for 1000 structures with 40,000 annealing steps for each. The 50 structures with the lowest target functions (≤ 0.6) were minimized with AMBER (version 9) (*59*). 20 structures with the lowest energy were visualized with InsightII (Accelrys, San Diego, CA), Chimera (*60*), and MOLMOL (*61*). Electrostatic potentials calculated using Delphi program (*62, 63*) were used to generate a surface potential map for ArsR-DBD in Chimera. The stereochemistry of the structures was analyzed with PROCHECK–NMR (*64*).

NMR analysis of ArsR-DBD in complex with the promoter region of a target gene: 2D ^{1}H-^{15}N HSQC spectra were collected for purified ^{15}N-labeled ArsR–DBD alone (0.1 mM), and ArsR-DBD combined with a 13 bp dsDNA fragment (5'CGCATCATTAACC) (0.1 mM) from the promoter region of a well-characterized ArsR target gene, *hp1408* (*44*). This DNA fragment corresponds to the 5' half of a DNA-binding region identified by footprinting analysis (*44*).

Sequence comparison of ArsR and related proteins—Sequence alignment of ArsR and several structurally characterized response regulators [OmpR (GenPept accession number AAC76430), PhoB (GenPept accession number AAC73502), DrrD (GenPept accession number 1KGS_A)] was carried out using CLUSTALW (*65*). The sequence of ArsR from *H. pylori* strain J99 was also compared to ArsR from other *H. pylori* strains (26695 and HPAG1) and orthologs from *Helicobacter acinonychis* (Sheeba), *Helicobacter hepaticus* (ATCC 51449), *Wolinella succinogenes* (strain DSM 1740), *Campylobacter lari* (RM 2100), *Campylobacter curvus* (525.92), and *Campylobacter concisus* (13826).

Results

ArsR-DBD structural determination

Backbone NH resonance assignments were obtained (Figure III-1) for all residues of the ArsR-DBD, except for the first 11 residues of the 12-residue N-terminal His_6 tag, four proline residues, and two isoleucine residues (I176 and I192, numbered based on the sequence of untagged full-length ArsR). The solution structure of ArsR-DBD was determined by NMR based on 1265 distance constraints, including 264 intraresidue, 329 sequential, 360 medium range, and 312 long range distance constraints, and 155 dihedral angle constraints calculated using TALOS (*57*) and J-coupling constants calculated from HNHA experiments. Table III-1 summarizes the structural statistics for the calculations. The first 17 residues of the protein (including the 12 residues from the His_6-tag) showed very few NOE interactions and a clustering of chemical shifts, findings that are characteristic of disordered peptides. Residues 128-225 form a well-structured domain. The 20 structures presented in the final ensemble (Figure III-2) conform to the average structure with atomic root mean-square deviations (r.m.s.d.s) about the mean coordinate positions of 0.55 (±0.07) Å for the backbone atoms, and 1.22 (±0.12) Å for all the heavy atoms of residues 129-224. The structural quality of the ensemble of the calculated structures was analyzed by using PROCHECK-NMR. 98.4% of the residues fall into the allowed regions (84.3%, 11.4%, 2.7% in the most favorable, additionally allowed, and generously allowed regions, respectively) of the Ramachandran plot (Figure III-3).

Overview of ArsR-DBD structure. The ArsR-DBD is composed of an N-terminal four-stranded antiparallel β-sheet [β1(G129-A131); β2(F134-D137); β3(E142-M145); β4(K148-D151)], three α-helices that form the core of the protein [α1(A155-K166); α2(R173-E179); α3(K190-K205)], and a C-terminal β-hairpin [β6(I213-V216); β7(G220-E224)] (Figure III-4). In addition, the domain contains a short β-strand connecting helices α1 and α2 [β5 (V170-S172)], which interacts with the C-terminal hairpin. The final topology of this domain, from N- to C-

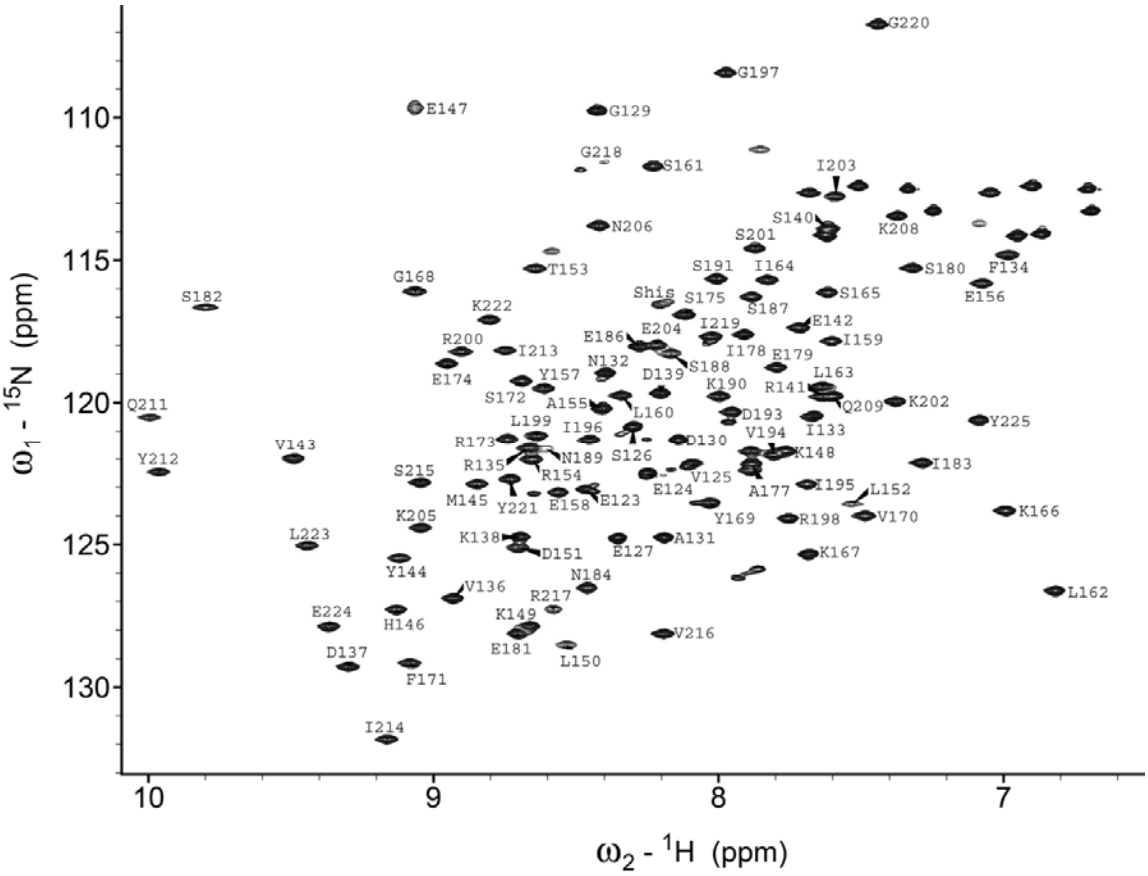


Figure III-1. Two-dimensional ^{1}H- ^{15}N NMR spectrum of ArsR-DBD. ^{1}H-^{15}N HSQC spectrum of ArsR-DBD. Peaks are labeled with one-letter amino acid code and residue numbers (Shis refers to a peak corresponding to a serine residue in the His_6-tag of the recombinant ArsR-DBD protein).

Table III-1. Structural statistics of the final ensemble of 20 energy minimized structures of ArsR-DBD.

Number of NOE distance constraints	
All	1265
Intraresidue	264
Sequential (|*i-j*|=1)	329
Medium range (2 ≤|i-j|≤5)	360
Long range (|i-j|>5)	312
NOE violations present in all 20 structures greater than 0.2 Å	0
Hydrogen bonds distance restraints[a]	48
Dihedral angle constraints	155
Mean root mean square deviations from the average coordinate (Å)[b]	
Backbone atoms (N,C_{α},C')	0.55 (±0.07)
Heavy atoms	1.22 (±0.12)
Ramachandran plot (%)	
Most favored regions	84.3
Additional allowed regions	11.4
Generously allowed regions	2.7
Disallowed regions	1.6

[a] Two restraints per one hydrogen bond
[b] Root mean square deviation values for residues 129-224

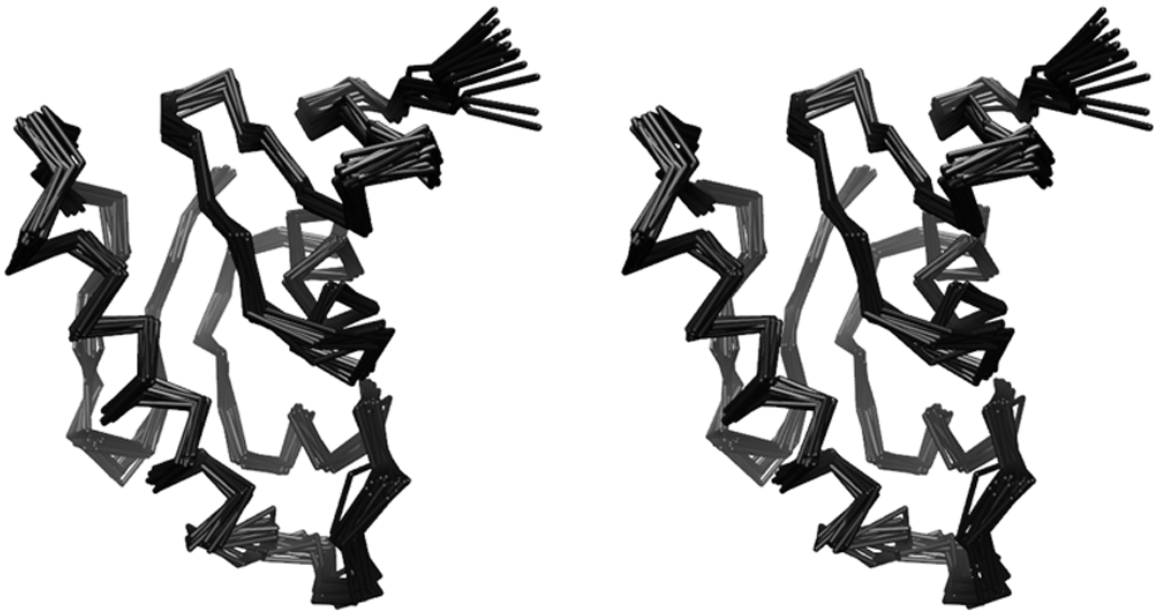
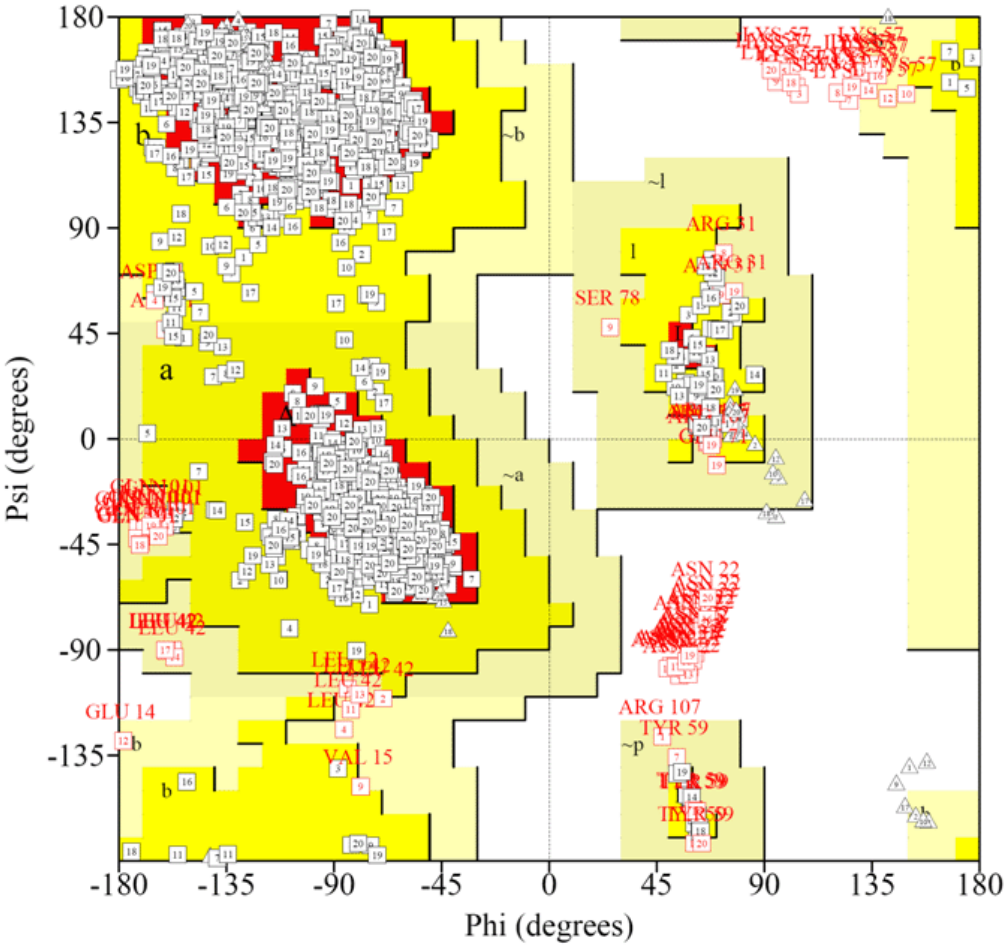


Figure III-2. Ensemble of 20 structures of ArsR-DBD. A stereo view of the best-fit superposition of backbone C_α atoms of the final 20 energy minimized structures of ArsR-DBD (residues 128-225).



Plot statistics

Residues in most favoured regions [A,B,L]	1584	84.3%
Residues in additional allowed regions [a,b,l,p]	215	11.4%
Residues in generously allowed regions [~a,~b,~l,~p]	50	2.7%
Residues in disallowed regions	31	1.6%
	----	------
Number of non-glycine and non-proline residues	1880	100.0%
Number of end-residues (excl. Gly and Pro)	20	
Number of glycine residues (shown as triangles)	120	
Number of proline residues	80	

Total number of residues	2100	

Figure III-3. Ramachandran plot for the residues belonging to the 20 structures of the ArsR-DBD. The quality of the covalent geometry of the calculated structures in the ensemble was analyzed by using PROCHECK-NMR. Residues belonging to the 20 structures fall into the allowed, favorable, additionally allowed, and generously allowed regions as indicated, with numbered squares representing residues in individual structures. Residues lying in the additionally allowed and the generously allowed regions are indicated by the three-letter amino acid code and the residue numbers.

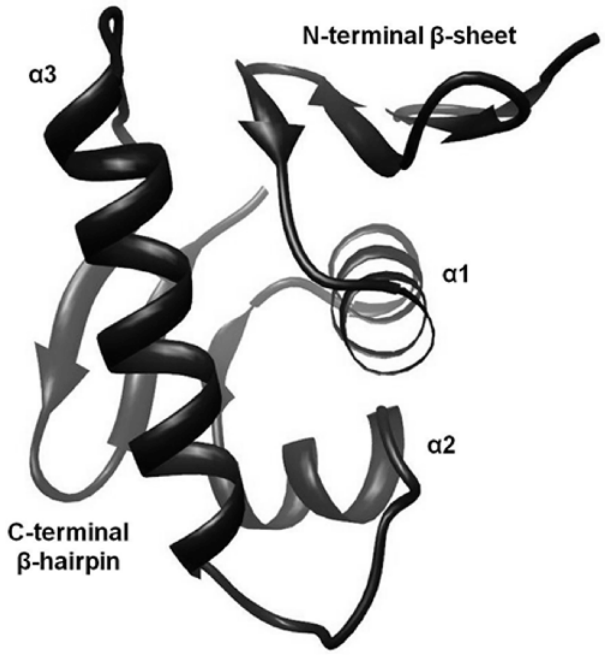


Figure III-4. Structure of ArsR-DBD. Ribbon diagram of ArsR-DBD (residues 128-225). Secondary structural elements are labeled.

terminus, is β1-β2-β3-β4-α1-β5-α2-α3-β6-β7. The structured regions of the 20 structures in the final ensemble (residues G129-A131, F134-D137, E142-M145, K148-D151, A155-K166, V170-S172, R173-E179, K190-K205, I213-V216, and G220-E224) converge with a backbone r.m.s.d. value of 0.43 Å. As discussed further below, the ArsR-DBD contains a winged helix-turn-helix (wHTH) fold that is predicted to mediate binding of the protein to DNA. This wHTH fold is formed by the α2 and α3 helices, the loop connecting them, and the loop connecting β strands 6 and 7 as the "wing".

The electrostatic surface potential map of ArsR-DBD (Figure III-5) reveals two distinct regions with opposite charge distribution. A surface with a largely positive electrostatic potential (including residues R173, K190, R198, R200, K202, K205 and R217) is shown in Figure III-5A. Based on comparisons with other RRs (*41, 53, 66*), this surface is predicted to bind to the negatively charged phosphate backbone of DNA. Figure III-5B displays a surface with a highly negative electrostatic potential (including residues D130, E158, E174, E179, E181, and E186), with the majority of the charge contribution from the residues at the C-terminal end of α2.

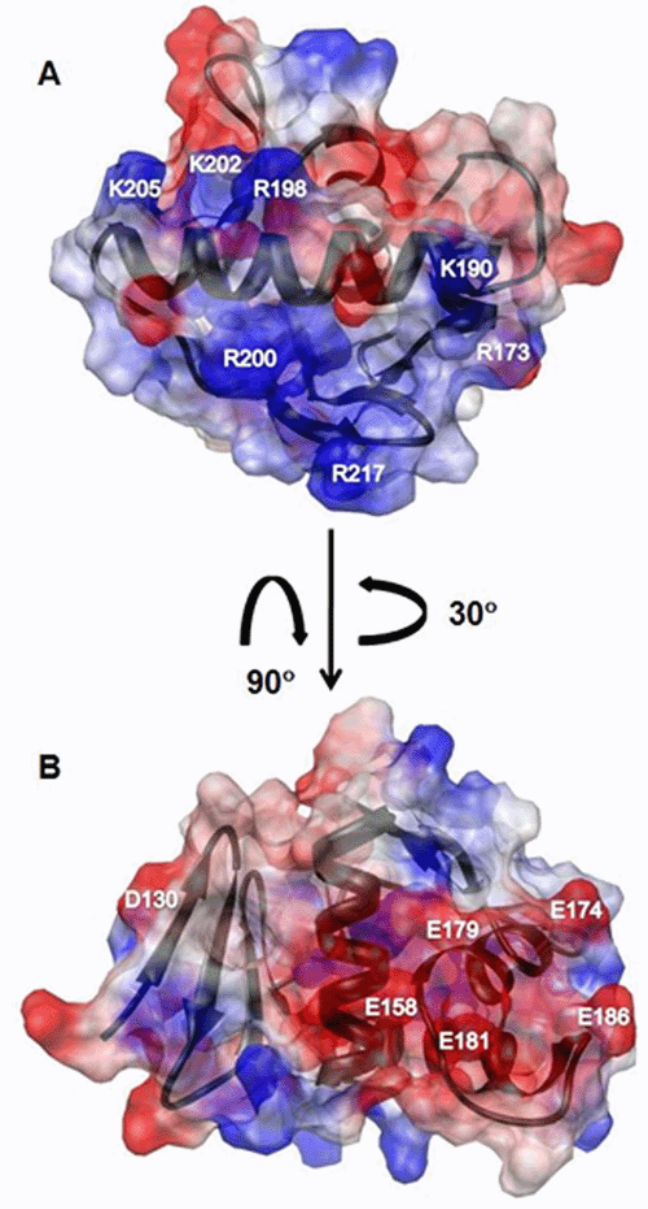


Figure III-5. Electrostatic surface potential maps of ArsR-DBD. Red indicates a negatively charged region, blue indicates a positively charged region, and white indicates a neutral or hydrophobic region. **A.** The molecule is oriented to display the predicted DNA-binding surface. Residues contributing to the positive surface potential are R173, K190, R198, R200, K202, K205, R217. **B.** The molecule was rotated to reveal a negatively charged surface, attributable to residues D130, E158, E174, E179, E181, and E186.

Structural comparison of ArsR-DBD with related structures

A BLAST search of GenBank (*67*), using the ArsR amino acid sequence, indicates that ArsR is most closely related to members of the OmpR/PhoB subfamily of response regulators, but the levels of amino acid sequence identity are fairly low [32% (OmpR), 28% (PhoB), and 32% (DrrD)] (Figure III-6). Despite these low levels of sequence identity, an analysis of ArsR with DALI software and the Families of Structurally Similar Proteins database (*68*), which catalogs known protein structures, showed that the ArsR-DBD structure was closely related to the structures of members of the OmpR/PhoB subfamily, including OmpR (PDB code 1opc) (*66*), PhoB (PDB code 1gxq) (*53*), and DrrD (PDB code 1kgs) (*33*), with Z-scores—depicting strength of structural similarity—of 3.8, 6.1, and 5.3, respectively. The secondary structural elements of ArsR-DBD are positioned by internally packed hydrophobic residues that stabilize the protein fold (Figure III-7). These core hydrophobic residues of ArsR-DBD (corresponding to residues F134, Vl136, L150, I159, L160, L163, I164, I176, I192, I196, L199, I203, and I213 in ArsR) are conserved across members of the OmpR/PhoB subfamily (Figure III-6, III-7) (*69*), which suggests that they are responsible for the shared structural features.

PhoB, a RR from *E. coli*, is a well-studied prototype of the OmpR/PhoB subfamily. Despite the low 28% amino acid sequence identity of PhoB with ArsR, the DNA binding domain of PhoB exhibits a high degree of structural similarity to the ArsR-DBD, as was shown by the DALI analysis (see above). Thus, we performed a detailed comparison of the ArsR and PhoB structures (Figure III-8). The arrangement of the three α-helical bundle is highly similar in the ArsR and PhoB DBDs, and was used to superimpose the two structures. When tracing along the N, C_α, C' and atoms, residues A155-K166 (α1), R173-E179 (α2), and K190-K205 (α3) of ArsR-DBD superimpose with residues of the PhoB-DBD helical core with an r.m.s.d value of 2.3 Å. The orientation of the loop between α2 and α3 (which is predicted to interact with subunits of RNA polymerase) and the length of the α3 helix (which is predicted to be involved in DNA recognition) are known to vary among members of the OmpR/PhoB subfamily, but these

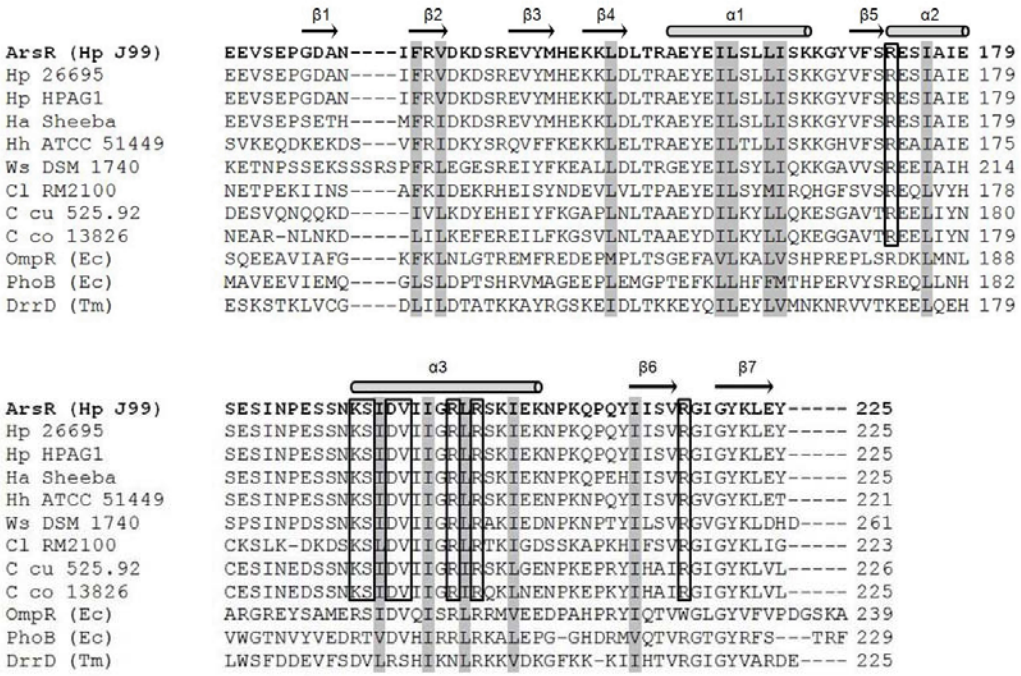


Figure III-6. Sequence comparison of ArsR-DBD with related proteins. Protein sequence of ArsR-DBD was alignment with protein sequences of orthologs from closely related species, and with closely related proteins from the OmpR/PhoB subfamily. The sequence of ArsR from *H. pylori* strain J99 was aligned with the sequence of ArsR from other *H. pylori* strains (26695 and HPAG1) and orthologs from *Helicobacter acinonychis* (Sheeba), *Helicobacter hepaticus* (ATCC 51449), *Wolinella succinogenes* (strain DSM 1740), *Campylobacter lari* (RM 2100), *Campylobacter curvus* (525.92), and *Campylobacter concisus* (13826) using CLUSTALW (*65*). Sequence of ArsR was also compared to sequences of several structurally characterized response regulators [OmpR (GenPept accession number AAC76430), PhoB (GenPept accession number AAC73502), DrrD (GenPept accession number 1KGS_A)]. Secondary structural elements of ArsR-DBD are displayed above the sequence with horizontal arrows indicating β strands and cylinders indicating α helices. Core hydrophobic residues that are conserved within the DBDs of the OmpR/PhoB subfamily are shaded grey. Surface-exposed residues in the wHTH motif that are identical in ArsR and orthologs from closely related species are boxed.

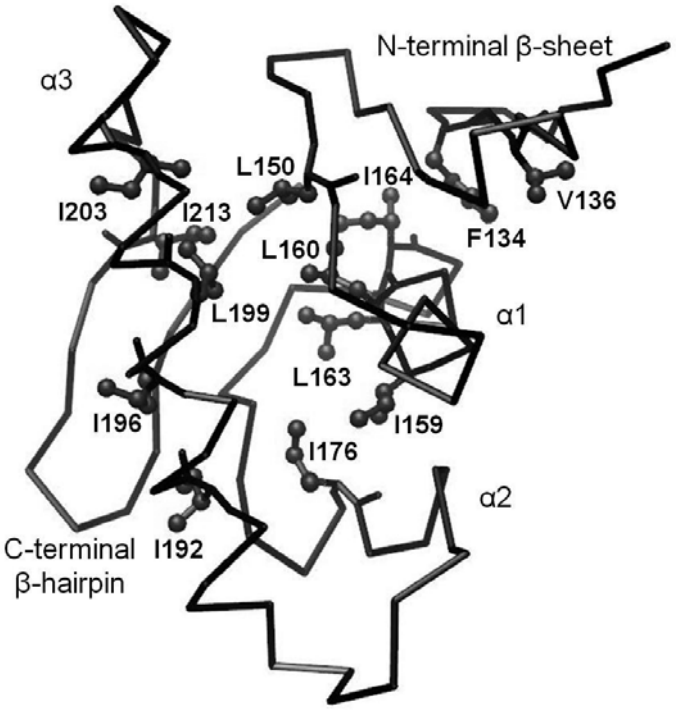


Figure III-7. Conserved core hydrophobic residues of ArsR-DBD and related proteins. Backbone C_α tracing of ArsR-DBD depicting side chains of conserved hydrophobic core residues. The hydrophobic residues conserved in ArsR-DBD and the DBDs of OmpR, PhoB and DrrD are indicated in ball-and-stick style on the line tracing of the ArsR-DBD backbone.

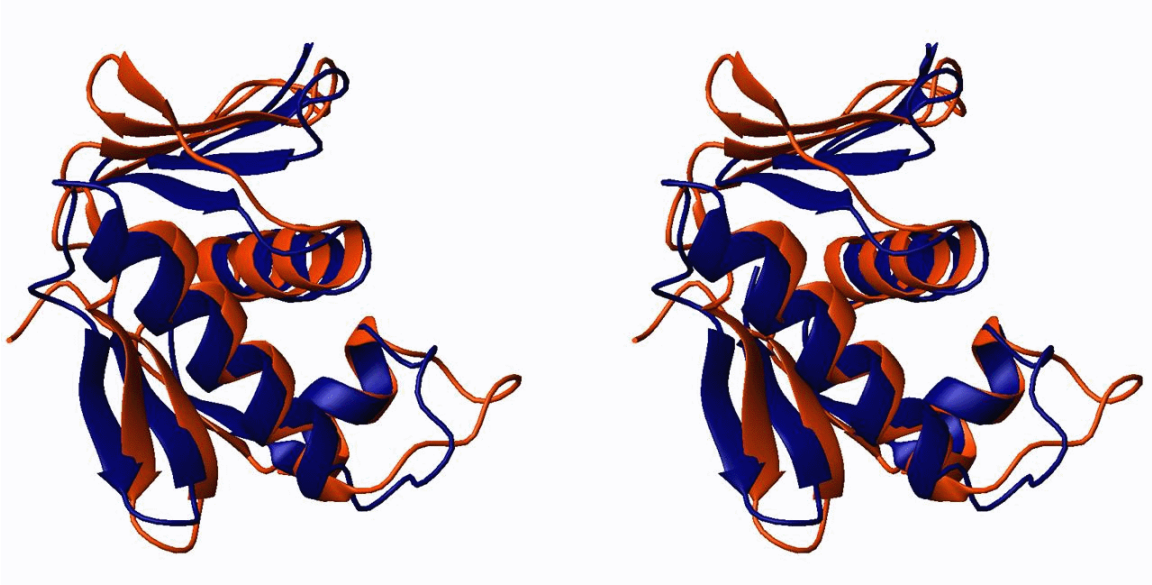


Figure III-8. Structural comparison of the DBDs of ArsR and PhoB. The images represent a stereo view of the superposition of the structures of ArsR-DBD (blue) and the DNA-binding domain of PhoB (orange) (PDB code 1gxq, (*53*)). The superimposition was based on the N, C_{α}, and C' atoms of the three core helices with an r.m.s.d. value of 2.3 Å.

structural features are similar in ArsR and PhoB. However, a notable difference is that the N-terminal β-sheet in ArsR-DBD is rotated almost 45º with respect to the N-terminal β-sheet in PhoB-DBD. This antiparallel β-sheet functions as a platform for interactions with the N-terminal receiver domain in many OmpR/PhoB RRs (*33, 70*).

Comparison of ArsR with closely related orthologs

A BLAST search, using the ArsR amino acid sequence, identified ArsR orthologs in all *H. pylori* strains for which genome sequences are available, as well as in closely related species. A pair-wise sequence comparison of ArsR from *H. pylori* strains J99, 26695, and HPAG1, as well as strains from related species, *H. acinonychis*, *H. hepaticus*, *Wolinella*, and *Campylobacter* species, is shown in Figure III-6. The ArsR proteins encoded by multiple *H. pylori* strains exhibited 99% pair-wise amino acid sequence identity. ArsR from another gastric *Helicobacter* species, *H. acinonychis*, also exhibited a high level of relatedness (94% amino acid identity) to *H. pylori* ArsR. ArsR from *H. hepaticus*, an intestinal *Helicobacter* species, was 71% identical to *H. pylori* ArsR. ArsR orthologs from *Wolinella* and *Campylobacter* were 54% to 67% identical to *H. pylori* ArsR (Figure III-6). Comparisons of the ArsR-DBD orthologs from these species showed a high degree of conservation in core hydrophobic residues belonging to the major secondary structural elements of the protein. Furthermore, a high degree of sequence identity was maintained in the wHTH motif, most notably in the surface-exposed residues of the α3 recognition helix and the wing of the C-terminal β-hairpin (corresponding to residues R173, K190, S191, D193, V194, R198, R200 and R217 in ArsR) (Figure III-6, III-9).

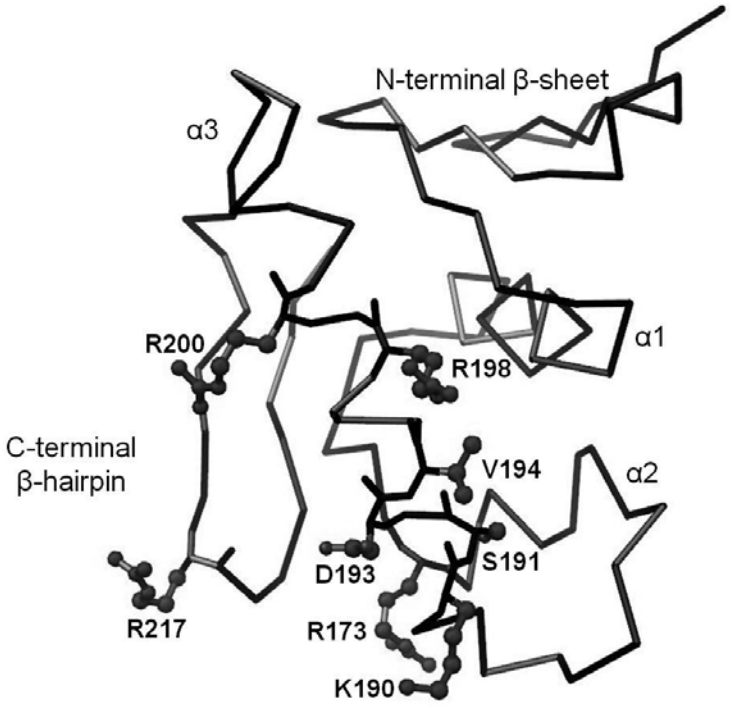


Figure III-9. Conserved surface-exposed residues in the wHTH of ArsR-DBD and related proteins. Backbone C_α tracing of ArsR-DBD and side chains of conserved surface exposed residues in the wHTH motif. Surface-exposed residues in the wHTH motif that are identical in ArsR and orthologs from closely related species are indicated in ball-and-stick style on the line tracing of the ArsR-DBD backbone.

Discussion

In this study, we present a three-dimensional structure of the ArsR DNA-binding domain, determined using NMR spectroscopy. The ArsR-DBD structure is most closely related to the structures of proteins classified in the OmpR/PhoB subfamily of response regulators. Like other members of this subfamily, the ArsR-DBD comprises two anti-parallel β-sheets flanking a core of three α-helices, and contains a winged HTH motif that is predicted to bind DNA.

The orientation of the N-terminal β-sheet in the ArsR-DBD differs markedly from the orientations of N-terminal β-sheets observed in previously solved effector domains of related RRs. Relative to the orientation of the N-terminal β-sheet in PhoB-DBD (*53*), the N-terminal β-sheet of ArsR-DBD is rotated about 45º to align in the direction of the α1 helix. As a consequence of this difference, we speculate that the interdomain interactions between the N-terminal receiver domain and the C-terminal DBD will differ in ArsR compared to what is observed in other members of the OmpR/PhoB subfamily. NTDs of RRs can regulate the activity of the corresponding DBDs (reviewed in (*34*)). Typically, NTDs of unphosphorylated RRs inhibit the activity of DBDs, often by providing steric inhibition to a functionally important site either in the NTD itself (as is suggested for PhoB (*53*)), or in the DBD (as for NarL (*71, 72*)). Phosphorylation-induced conformational changes in the NTD relieve this inhibition and allow the effector domain to carry out its cellular functions (reviewed in (*34*)). The novel orientation of the N-terminal β-sheet of the DBD in ArsR may allow unique DBD-NTD interdomain interactions that regulate ArsR's biological activities.

Based on the conservation of structural features in ArsR and members of the OmpR/PhoB subfamily of RRs, ArsR-DBD is predicted to interact with DNA in a manner similar to that of other members of the OmpR/PhoB subfamily. To test this hypothesis, the interaction of ArsR-DBD with DNA was analyzed by mapping spectral changes in ^{15}N-labeled ArsR-DBD in the presence of an equimolar concentration of a 13 bp fragment of the promoter sequence of *hp1408*

(*44*), a member of the ArsRS regulon (see Methods for details). Since OmpR/PhoB dimers bind approximately 24 bp of DNA, it is predicted that this 13 bp fragment would bind only one molecule of ArsR-DBD. A comparison of the 2D ^{1}H-^{15}N HSQC spectra of ArsR-DBD alone and combined with this DNA revealed changes in the intensities of peaks corresponding to specific amino acid residues (Figure III-10). This indicates moderately strong affinity between protein and DNA molecules, resulting in intermediate exchange regime in the NMR experiments. A detailed comparison of the peak intensities from the two spectra indicated that several residues exhibiting the most prominent reductions in peak intensities mapped to surface-exposed regions of the wHTH motif (residues R173, D193, and R200). Additionally, several core residues of the wHTH motif and other elements of the protein that make close contacts with the helices of the wHTH motif demonstrated diminished peak intensities (residues A155, I159, I183, I195, I196, G197, and G220) (Figure III-10). These data, combined with the analyses shown in Figures III-5, III-6, and III-8, support our prediction that ArsR-DBD interactions with DNA involve molecular surfaces homologous to those identified for other members of the OmpR/PhoB subfamily.

To date, the only solved structure of an OmpR/PhoB subfamily member in complex with its target DNA binding site corresponds to the DBD of an *E. coli* RR, PhoB-DBD, on the *pho* box of the *phoA* operon promoter (*53*). Analysis of the electrostatic surface potential map of the PhoB-DBD revealed a positively charged surface in the region that binds DNA (*53*), and the electrostatic surface potential map of ArsR-DBD reveals a similar basic surface (Figure III-5A). We used the structure of the PhoB-DBD in complex with a target promoter sequence (*53*) as a template to generate a model of the ArsR-DBD binding to DNA (Figures III-11, III-12). Thus far, a consensus sequence for ArsR binding sites has not been identified, and therefore, to generate this model, we retained the PhoB-specific binding site (*pho* box) as the target DNA; we did not attempt to model specific protein-DNA interactions when positioning ArsR-DBD on the DNA. Two ArsR-DBD domains were placed in tandem orientation (Figure III-11) in the major groove of the target DNA fragment, approximately 10 bp apart. The two ArsR-DBD molecules

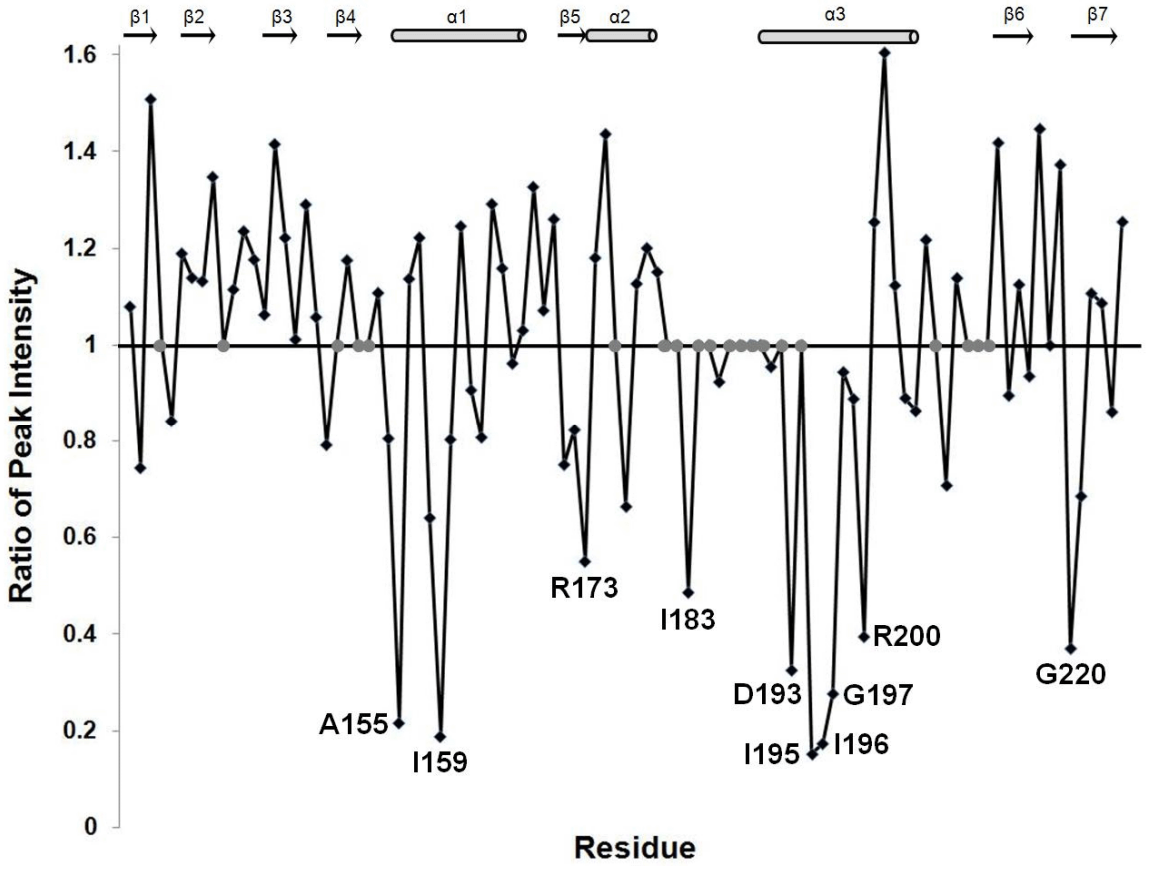


Figure III-10. Analysis of HSQC spectra of ArsR-DBD alone and ArsR-DBD combined with a 13 bp dsDNA fragment derived from the promoter region of an ArsR target gene (*hp1408*). For each residue, a ratio of peak intensity in the spectrum of ArsR-DBD with DNA to the peak intensity in the spectrum of ArsR-DBD without DNA is shown. Ratios substantially less than 1 indicate a decrease in peak intensity when the protein is combined with DNA. Residues for which intensity ratios were not available (due to absence of visible peaks or peak overlap) are depicted by grey symbols. Residues A155, I159, R173, I183, D193, I195, I196, G197, R200, and G220 showed the greatest decrease in intensity. ArsR-DBD secondary structure elements are shown above the plot.

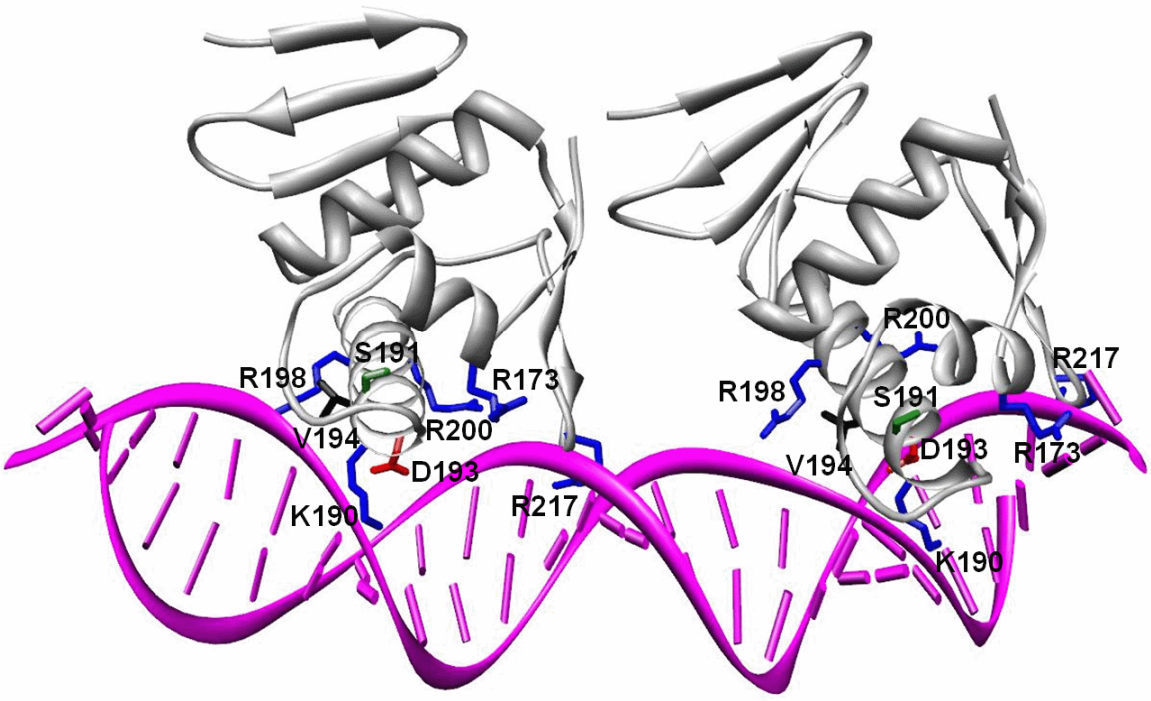


Figure III-11. Modeling of the ArsR-DBD-DNA interaction: tandem binding to a direct-repeat DNA sequence. Two ArsR-DBD molecules were positioned in tandem ("head to tail") on a 23 bp direct-repeat dsDNA sequence derived from the crystal structure of the PhoB-DNA complex (PDB entry 1gxp). Secondary structural elements of the proteins are shown as a grey ribbon. Residues in ArsR-DBD that potentially interact with DNA are presented as sticks with basic residues (R173, K190, R198, R200, R217) in blue, acidic residue D193 in red, polar uncharged residue S191 in green, and hydrophobic residue V194 in black.

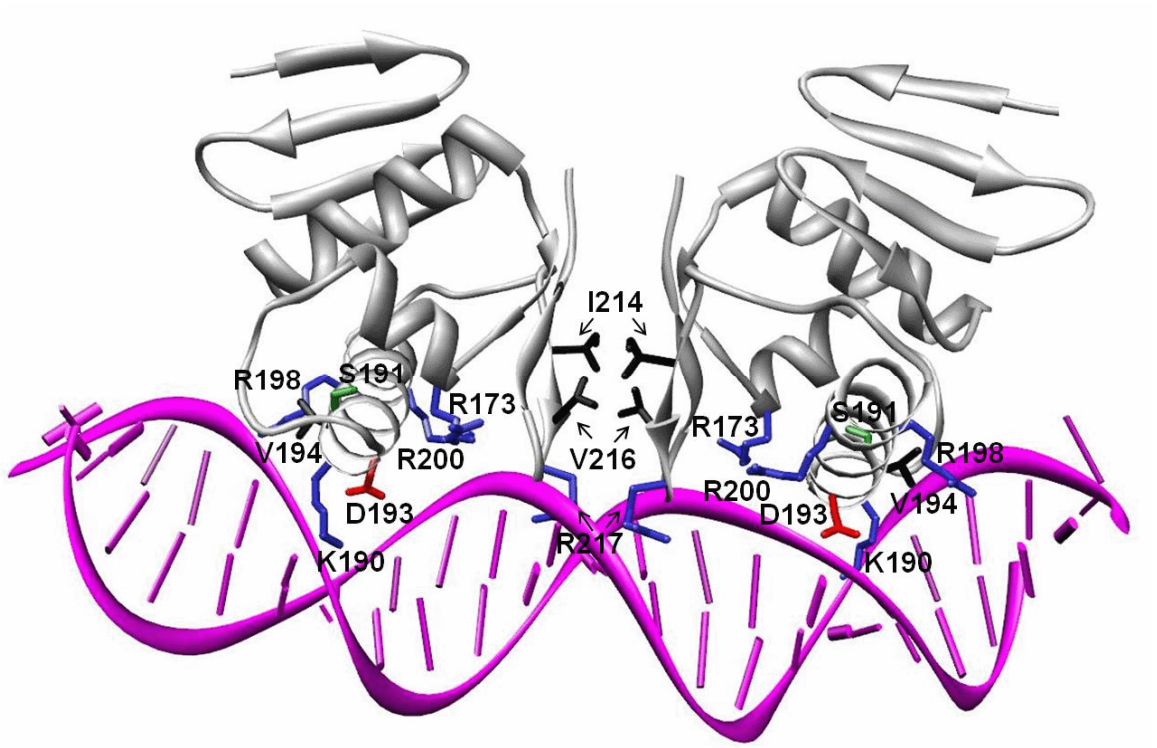


Figure III-12. Modeling of the ArsR-DBD-DNA interaction: symmetric binding to an inverted-repeat DNA sequence. Two ArsR-DBD molecules positioned "tail to tail" on a 23 bp ds DNA sequence containing an inverted repeat. Secondary structural and DNA binding elements are presented as in Figure III-11. Residues within the C-terminal β-hairpin that could interact with each other in this orientation, I214 and V216, are shown in black.

superimpose on the two PhoB-DBD molecules (tracing along the backbone N, C_α, C' atoms of the corresponding α2 and α3 helices), both with an r.m.s.d. value of 2.1 Å.

The PhoB-DBD binds to the direct repeat-containing *pho* box as a tandemly arranged ("head to tail") dimer (*53*). However, several binding sites of ArsR that have been reported thus far (from promoters of genes *arsR*, *ureA*, *ureI*, *amiE*, *amiF*, *rocF*, *hp1408*, and *hp1186* (carbonic anhydrase)) do not contain conserved symmetrical sequences (*24, 42, 44, 47*). The degeneracy of ArsR binding sites leaves open the possibility that ArsR may bind to DNA in a different manner than that observed with PhoB. In order to construct an alternative model of ArsR-DBD molecules arranged symmetrically on inverted repeat sequences, one half of the dimeric PhoB-DBD-DNA complex was inverted and superimposed on the original copy of DNA using only the DNA backbone phosphates for alignment, with an r.m.s.d. value of 0.96 Å. Two ArsR-DBD molecules were superimposed on the PhoB-DBD molecules (tracing along the C_α atoms of the corresponding α2 and α3 helices), with r.m.s.d. values of 2.1 Å (Figure III-12). In both orientations, residues R173, K190, S191, D193, V194, R198, and R200 from the ArsR HTH motif, and residue R217 from the 'wing' between β-strand 6 and β-strand 7 of the C-terminal hairpin, make contacts with the phosphate backbone and/or the bases of the target DNA. These interactions are supported by our NMR experimental data (Figure III-10). The functional groups of these ArsR residues are similar to those of the corresponding residues in PhoB, suggesting a conservation of their DNA-recognition functions.

In the models shown in Figures III-11 and III-12, residues K190, S191, V194, and R198 protrude into the major groove of the DNA to make specific contacts with bases. The side chains of lysine and arginine, and the hydroxyl group of S191 provide potential hydrogen-bonding partners. In addition, the methyl groups of valine can make specific van der Waals contacts with the methyl groups of thymines. The reported ArsR binding sites are A-T rich (*24, 26, 42, 44, 47*), suggesting that the valine-thymine contacts may be important determinants of sequence-specific protein-DNA interactions. A distinctive feature of the "tail to tail" symmetric orientation of the

ArsR-DBD-DNA complex model (Figure III-12) is the interaction of hydrophobic patches on the two ArsR-DBD molecules formed by residues I214 and V216 of each C-terminal β-hairpin. The surface-exposed hydrophobic residues on one ArsR-DBD molecule are stabilized by the hydrophobic residues exposed on the opposite ArsR-DBD molecule. This interaction could contribute to the overall stabilization of ArsR dimers on target DNA sequences.

In summary, the results of this study allow the classification of ArsR into a subfamily of DNA-binding proteins that contain a conserved wHTH motif, and yet exhibit diversity in their interactions with DNA and diversity in interdomain and protein-protein interactions. The structure of the ArsR-DBD provides a basis for future experimental studies designed to understand these interdomain, protein-protein and protein-DNA interactions.

CHAPTER IV

CHARACTERIZATION OF THE ARSRS REGULON

Introduction

The ArsRS two-component signal transduction system has been shown to be important in the ability of *H. pylori* to sense and regulate target genes in response to changes in pH, and is required for the colonization of the stomach (*24, 26, 29, 42, 45, 47*). Comprised of a sensor kinase ArsS and a response regulator ArsR, this TCS regulates multiple genes, including genes encoding members of the urease complex, amidases (*24, 42*), and carbonic anhydrase (*26*). The accepted model for pH-dependent regulation by this TCS involves detection of a change in the environmental pH by the sensor HK, which leads to autophosphorylation at a conserved histidine residue in the protein. This activated ArsS protein phosphorylates its cognate response regulator ArsR, which then binds to and modulates the expression of a specific set of target genes. In support of this model, several studies have detected direct binding of ArsR to the promoter regions of various acid-responsive target genes (*24, 26, 42, 47*). While phosphorylated ArsR is primarily thought to function in the regulation of pH-responsive genes, ArsR in its unphosphorylated state likely plays an essential role in the regulation of *H. pylori* genes. This view is supported by the observation that strains encoding a form of ArsR with a mutated phosphorylation site are viable, whereas *arsR* null mutant strains are non-viable (reviewed in (*22*)).

Members of the ArsRS regulon have been identified by isolating DNA sequences that bind to ArsR (*44*) and by analyzing gene expression in *arsS* mutant strains compared to wild-type strains (*42*). Comparative transcriptional profiling experiments have been performed using several different strains of *H. pylori*, grown at either pH 5.0 or pH 7.0, and several different types of *H. pylori* gene arrays (*29, 42, 47, 48*). Validation of the array results has been carried out for only a

few of the genes that are differentially expressed in wild-type and *arsS* mutant strains (*24, 26, 42, 47*), and the ArsRS regulon has not yet been completely characterized. In this study, we set out to further characterize the regulon controlled by the ArsRS two-component system.

As part of recent efforts to investigate the role of ArsS in regulating *H. pylori* gene expression, our lab collaborated with the Vanderbilt Proteomic Center and sought to identify *H. pylori* proteins that are differentially expressed in a wild type strain compared to an *arsS* mutant strain. Wild type *H. pylori* strain J99 and an isogenic *arsS* mutant strain were each cultured at pH 5.0 or pH 7.0. Quadruplicate samples were prepared independently and resolved on 8 coordinated difference gel electrophoresis (DIGE) gels for quantification as described (reviewed in (*73*)). 2/3rd of each of the 16 samples (WT and *arsS* mutant strains grown at pH 5.0 and 7.0, N=4 for each) were labeled with either Cy3 or Cy5. An internal standard was prepared with combined remaining portions of each of the samples labeled with Cy2. Randomized sets of two independent samples were co-resolved on each gel with first-dimension isoelectric point and second-dimension molecular weight based separations (representative gel shown in Figure IV-1). The normalized volume ratio of each individual protein spot-feature from a Cy3- or a Cy5-labeled sample was directly quantified relative to the Cy-2 internal standard signal corresponding to the same spot feature. Individual signals from the Cy-2 standard were used to normalize and compare Cy3:Cy2 and Cy5:Cy2 abundance ratios across the 8 gel set, enabling statistical confidence to be associated with each change in abundance.

639 resolved protein spot-features were matched across all eight gels, and the Cy3:Cy2 and Cy5:Cy2 ratios were normalized across all 16 samples using the Cy2 signal for each feature separately. Principal component analysis (PCA) was then used to assess the variation in expression patterns amongst these 639 features as it related to the characteristics of the sixteen individual samples. For all features, the first principal component (PC1) comprised 56.7% of the variation, distinguishing between the two bacterial stains, WT J99 and the *arsS* mutant. However, neither the second principal component, comprising the second greatest source of

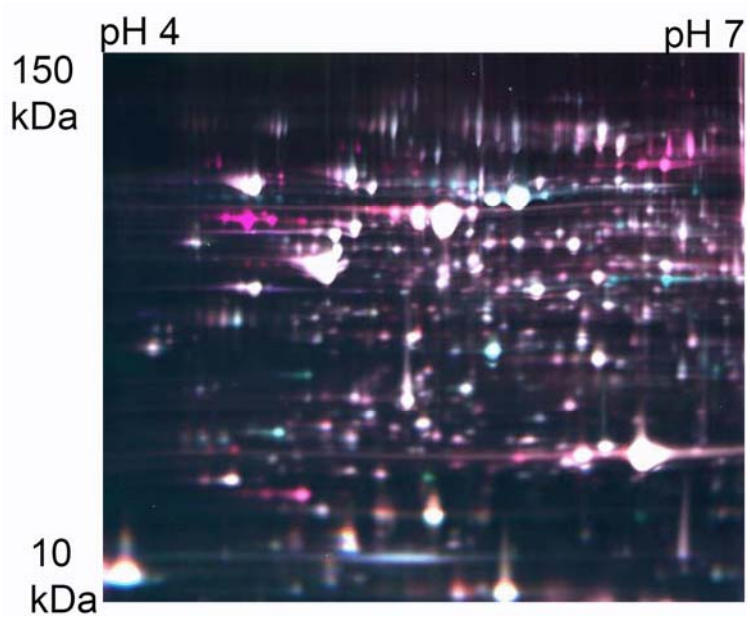


Figure IV-1. 2D-DIGE analysis of WT and *arsS* mutant *H. pylori* J99 strains. Proteomic profiles of an *H. pylori* WT strain J99 and an isogenic *arsS* mutant strain grown at pH 5.0 or pH 7.0 were compared by 2-dimensional differential gel electrophoresis (2D-DIGE). Presented in this figure is a representative 2D-DIGE gel with Cy2 (blue) internal standard, Cy3 (green) WT J99, pH 5.0, Cy5 (red) *arsS* mutant J99, pH 5.0 (false color overlay shown for demonstrative purposes). Cell lysates were simultaneously co-resolved with first-dimension isoelectric point and second-dimension molecular weight (12% SDS-PAGE) based separations.

variation (PC2=9.8%), nor any of the other principal components, distinguished between the two different growth states, i.e. pH 5.0 versus pH 7.0, (Figure IV-2A). This analysis demonstrated that the majority of the variation in protein expression reflected differences between the two strains. We predict that a reason for the paucity of pH-dependent changes in protein expression was that growth of *H. pylori* at pH 5.0 was not a sufficient stimulus for acid-induced alterations in protein expression. From the 639 protein spot-features, a subset of protein features that were significantly altered in expression in any of the four groups relative to the others was selected based on ANOVA *p*-values < 0.05. For this subset of 168 features, PC1 comprised over 81% of the variation, again distinguishing between the WT and the *arsS* mutant strains. PC2 comprised 6.4% of the variation, and this variation distinguished between bacteria grown at pH 5.0 and pH 7.0 (Figure IV-2B). Thus, PCA demonstrated distinct protein expression patterns in the four groups of samples with an extremely low level of technical noise.

A comparison of protein expression in WT J99 and the isogenic *arsS* mutant (grown at pH 5.0 and 7.0) revealed 25 proteins differentially expressed between the two strains, including proteins that contribute to acid resistance (urease, amidase), acetone metabolism (acetone carboxylase), and resistance to oxidative stress (thioredoxin reductase) (Table IV-1). The proteins were identified after in-gel digestion and mass spectrometric analysis of the resulting peptides.

Relatively few pH-responsive changes were observed through these proteomic analyses. AmiE was pH responsive in the WT strain, but not in the *arsS* mutant strain, providing evidence that the ArsRS TCS has a role in the regulation of the *amiE* gene. pH-responsive changes were observed in the expression of UreB in both WT and the *arsS* mutant strain, suggesting that acid-related changes in its expression may not be exclusively mediated by the ArsRS TCS.

The proteomic analysis conducted on WT and the isogenic *arsS* mutant *H. pylori* J99 strains grown at pH 5.0 and pH 7.0 revealed 25 proteins differentially regulated between the two strains (Table IV-1). In this study, we investigated which of these differentially expressed proteins are

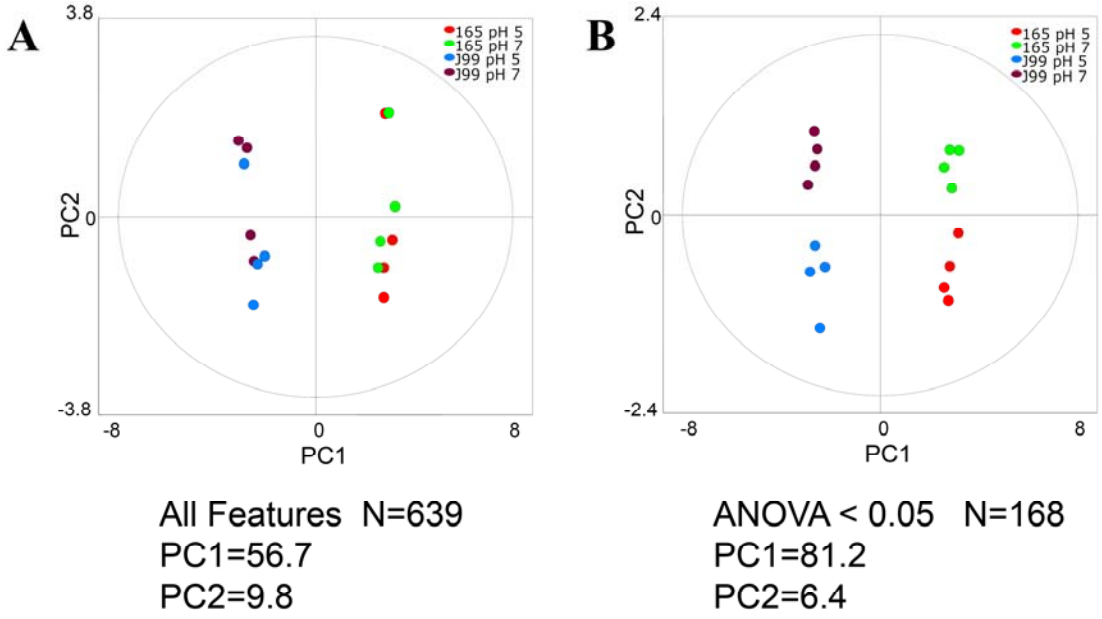


Figure IV-2. Principal component analysis of protein expression patterns in WT and *arsS* mutant *H. pylori* J99 strains grown at pH 5.0 and pH 7.0. A. PCA was used to access variation in protein expression patterns among 16 samples. PC1 comprised of 56.7% of the variation, distinguishing between the two bacterial strains, WT and the *arsS* mutant. PC2 comprising the second greatest source of variation did not distinguish between the two growth states, pH 5.0 and pH 7.0. **B.** PCA analyses of data sets identified based on ANOVA *p*-values < 0.05. PC1 comprised over 81% of the variation amongst the 168 features in this subset, distinguishing between the two strains. PC2 comprised of 6.4% of the variation, demonstrating distinction between the two growth states.

Table IV-1. Comparison of protein expression patterns in WT *H. pylori* strain J99 and an isogenic *arsS* mutant strain.

Proteins with decreased expression in *arsS* mutant strain vs. WT	**Proteins with increased expression in *arsS* mutant strain vs. WT**
Acid resistance	**Acetone metabolism**
AmiE	AcxA
HypB	AcxB
UreB	AcxC
UreG	**Motility**
UreH	FlaA
Oxidative stress resistance	FlaB
TrxR1	FlgE
TrxR2	FlgK
others	FlgL
AspB	FliD
Hypothetical (HP0162)	**others**
MetB	FabD
ProS	FutB
TatD	HomD
	Pfs

encoded by genes under direct regulation of the ArsRS TCS. This chapter describes the results of our DNA-binding analyses with full length ArsR and the DNA-binding domain of ArsR (ArsR-DBD), demonstrating that genes encoding at least 12 of the 25 differentially expressed proteins belong to the ArsRS regulon. Of these 12 genes, 9 are newly identified targets of the ArsRS TCS. The ArsRS-regulated proteins described in this study contribute to acid resistance (urease), acetone metabolism (acetone carboxylase), resistance to oxidative stress (thioredoxin reductase), and several other functions. These results provide further definition of the ArsRS regulon, and underscore the importance of the ArsRS system in regulating expression of *H. pylori* proteins during bacterial growth at both neutral pH and acidic pH.

Methods

Plasmid construction: The ArsR protein encoded by *H. pylori* strain J99 comprises 225 amino acids. The DNA binding domain of this protein, mapped based on comparison to other known response regulators, such as PhoB and OmpR, comprises 103 amino acids, beginning at E123 and ending at Y225. The full-length *arsR* gene and a fragment encoding the DNA-binding domain were amplified from *H. pylori* strain J99 and ligated into linearized pET-BNK as described in Chapter II.

Expression and purification of ArsR and ArsR-DBD: Protein samples were prepared from transformed *E. coli* BL21 (DE3) cells as described in Chapter II.

Electrophoretic Mobility Shift Assays (EMSA): Primers listed in Table IV-2 were used to PCR-amplify 100-150 bp regions located upstream of the translation initiation sites of the selected target genes. For each primer set, the forward primer was 5' biotinylated. To phosphorylate the full-length ArsR protein in vitro, the purified recombinant protein (3 μM) was incubated for 30 min at 25^{o}C in phosphorylation buffer [50 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 50 mM KCl, 1 mM dithiothreitol] containing 50 mM acetyl phosphate (*44*). To determine if the ArsR proteins bound to the gene of interest, the recombinant proteins were incubated for 20 min in a binding reaction mix [10 mM Tris, pH 7.5, 50 mM KCl, 1 mM DTT, 50 ng/ul poly(dI-dC), 0.05% NP-40, 2.5% glycerol, 5 mM $MgCl_2$] containing 100 pM of biotinylated probe. For competition assays, a 20-fold excess of non-biotinylated probe (compared to biotinylated probe) was included in the binding mixture. Following the incubation, loading buffer was added, and the samples were subjected to electrophoresis in a 6% polyacrylamide native gel in 0.5X TBE (50 mM Tris, 45 mM boric acid, 5 mM EDTA, pH 8.3). Samples were electrophoretically transferred to a nylon membrane (BioRad) and the transferred DNA was cross-linked to the membrane using

Table IV-2. Oligonucleotide primers used for the generation of promoter fragments of genes of interest from *H. pylori* strain J99 for EMSA assays.

Gene of interest	Primer	Primer Sequence 5' - 3'
ureI	jhp0066FB	GGATTTTAAGGAGCGTTCC
	jhp0066Rev	GCCTTATCCTTCCAAAAC
ureA	ureA12 FB	CAAATGATTTTAAAGGGTATTAAACGCGCTCC
	ureA12 Rev	ACATAACGCAATCAAGGTTGGAG
pfs	jhp0082FB	GCGTTATCAAGACTGCGTGA
	jhp0082Rev	TTGCACCATTTTCTCCCAAT
fabD	jhp0083FB	ATTCGAACCCTCGGTAACCT
	jhp0083Rev	GCGCGTATTGCATGATTATCC
metB	jhp0098FB	TTGCAAACTTGGATATTGATCG
	jhp0098Rev	GCATGCGCATAGTTTTTCC
flaB	jhp0107FB	TTTGCTGGGGATTCTACGAT
	jhp0107Rev	GTTTGCATCCTTTGCATAGATATTTGC
hp0162	jhp0149FB	AATGTGAGTGGGGAATACGC
	jhp0149Rev	CGTCCCATTCAATTCCTTTG
arsR	jhp0152 FB	CCATGAAAACAAAGCC
	jhp0152 Rev	CTTCAATTAACTCCTTCAATG
proS	jhp0223FB	GATCTTGCACAACGCTTTCA
	jhp0223Rev	GCATTATTCCTCGTAATATTCGC
hopF	jhp0237FB	CTCTCCCAACTGAGCTAACACCCC
	jhp0237Rev	CCTTTTTGCTTTGGTAGGTAAG
amiE	jhp0279FB	TTGCAAAGAGCGGCTAAAAT
	jhp0279Rev	GTTCCTAAAAACCCTCTATTGGA
flgL	jhp0280FB	ACTTAAAAACACCCCATTTAACC
	jhp0280Rev	CAAACAAGCTAAAAGTATTCCAAAA
flaA	jhp0548FB	TTATGGTATCCATTGGTGTTCG
	jhp0548Rev	GCCATTGTTGTAACTCCTTG
futB	jhp0596FB	ATGGAGCGAATGACTTGAGC
	jhp0596Rev	GGAACATGATTATCCTTTTAAAGG
hyuA (*acxA*)	jhp0633FB	TTTTGGCAATTTAAGATGAATATCAG
	jhp0633Rev	CTTGTATTAAAGTGTTGCTCAAGGCAACC
hyuB (*acxB*)	jhp0632FB	GGATGCAGATGTGTGGCACATG
	jhp0632Rev	TTTGAACTCCTTTATTTAATTTCTTTCAAGTGG
fliD	jhp0689FB	ACGCATTGGCACGGATATTA
	jhp0689Rev	CCTATTGCCATGTTTTACCTC
hp0813	jhp0749FB	CGATTTCCCTTTCATTCAAACTC
	jhp0749Rev	CTCTAAACTCCCTTCAAAC
trxR1	jhp0764FB	GCGAATATGAAGGTAAGGCTAAGA
	jhp0764Rev	GCCTAAAAGCTTGTTCAATTGCT

Table IV-2 (continued). Oligonucleotide primers used for the generation of promoter fragments of genes of interest from *H. pylori* strain J99 for EMSA assays.

Gene of interest	Primer	Primer Sequence 5' - 3'
trx1	jhp0763FB	AACGATCCGCTGTTATTTGTC
	jhp0763Rev	GTGACTCATTGGTTACTCC
flgE	jhp0804FB	TGGTGTTACAAAACCCCCTA
	jhp0804Rev	TTGACACCAGACCATAAAGACC
hypB	jhp0837FB	GGCGATAAGGGGTTAGGGTA
	jhp0837Rev	CGTTGTTCGCTCATGCTTGTTCCTTTC
flgK	jhp1047FB	TGCGAGAGTTCTTGCTTGTTT
	jhp1047Rev	GCCTGCACTCTTAAAAAATGACTCC
trxR2	jhp1091FB	TGCTTATGGGATTGGTAGGC
	jhp1091Rev	CTTGGTCCATTCCTATCC
hp1288	jhp1208FB	ACTTGGCATTCCAAAGTGCT
	jhp1208Rev	AAGGCATGTTTTGCACACCT
rocF	jhp1427FB	GTTAATCCTTTAAATAGTAGTGTC
	jhp1427Rev	AACTCCTTTAAATCCCCACTC
homD	jhp1346FB	AAGCATGGATTAAAACCCCTTT
	jhp1346Rev	CATGAAATCTCCTTAATGGTC

a UV Stralinker 1800 (Stratagene). Biotin-labeled DNA was detected using the LightShift Chemiluminescent EMSA kit (Pierce). First, the membranes were incubated with blocking buffer for 15 min, followed by an incubation of 15 min in blocking buffer containing a 1:300 dilution of streptavidin-horseradish peroxidase conjugate. The membranes were then washed with washing buffer and incubated in luminol based-substrate working solution to allow visualization of the DNA bands on X-ray film.

Results

Binding of full length ArsR protein to promoter regions of specific genes previously reported to be members of the ArsRS regulon

To determine whether the expression of proteins listed in Table IV-1 was directly regulated by the ArsRS system, we performed a series of experiments to investigate whether ArsR could bind to the promoter regions of genes of interest. The full-length ArsR protein was expressed and purified, and its binding to biotinylated DNA fragments (probes) was assessed in EMSAs. In initial experiments, we validated the ability of the full-length recombinant ArsR protein to bind to promoter regions of genes reported to be members of the ArsRS regulon (*24, 44, 47*). ArsR was initially tested at a concentration of 3 μM, as described previously (*37*). Consistent with previous results (*24, 37, 42*), ArsR retarded the migration of *ureI*, *rocF*, *ureA* and *arsR* promoter probes (Figure IV-3). The migration of at least 2 of these positive control promoter probes (i.e. *ureI* and *arsR*) was retarded to a greater extent if the full-length ArsR protein was pre-incubated with acetyl phosphate than if it was not pre-incubated with acetyl phosphate (Figure IV-3). This result suggests that interaction of ArsR with promoter regions of *ureI* and *arsR* is modulated by the phosphorylation state of ArsR. In contrast, the migration of the two other positive control promoter probes (*ureA* and *rocF)* was similar regardless whether or not ArsR was pre-incubated with acetyl phosphate.

We also examined the ability of ArsR to bind to promoter regions of several genes *(hopF*, *tlpC*, *flaA* and *flgK)* previously reported to be transcribed in an ArsRS-independent manner. Earlier studies showed that ArsR does not bind to the promoter regions upstream from these genes (*47*), and as expected, the migration of these negative control promoter probes was not retarded by ArsR (Figure IV-3).

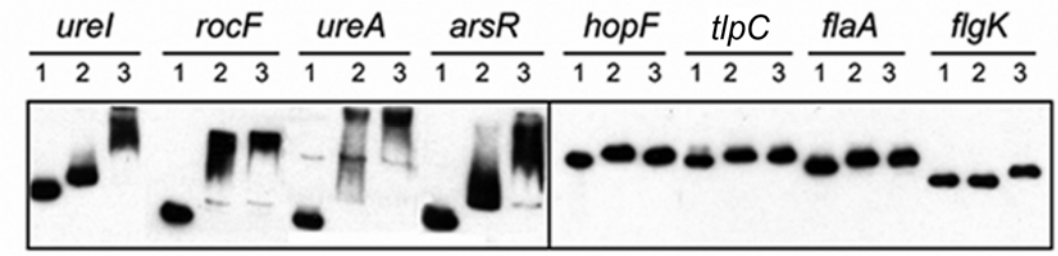


Figure IV-3. Binding of ArsR to control DNA promoter probes. 150 bp fragments upstream of the translational initiation sites of the target genes were PCR-amplified, using biotinylated primers. DNA fragments were incubated with buffer alone (lane 1), unphosphorylated ArsR (3 μM final concentration) (lane 2), or ArsR that was phosphorylated by incubation with acetyl phosphate (lane 3). See Methods for a detailed description of experimental conditions. Fragments tested included positive control promoter probes (*ureI*, *rocF*, *ureA, arsR*) known to be bound by ArsR, and negative controls (*hopF*, *tlpC*, *flaA*, *flgK*). Samples were subjected to non-denaturing PAGE, transferred to nylon membranes, and biotinylated DNA visualized using chemiluminescent techniques.

Binding of ArsR-DBD to promoter regions of specific genes previously reported to be members of the ArsRS regulon

In further experiments, we examined the binding of the ArsR-DBD to the promoter regions described above. ArsR-DBD protein bound to *ureA* and *arsR* promoter probes, with binding occurring in a concentration-dependent fashion (Figure IV-4). The minimum concentration of ArsR-DBD required for causing detectable retardation in migration of these promoter probes was 1.0 μM, and therefore, binding to all other promoter probes was subsequently carried out using 1.0 μM of ArsR-DBD. As shown in Figure IV-5, ArsR-DBD retarded the migration of promoter probes (*rocF, ureI*, *arsR, hp1288*) that were previously reported to be bound by the full-length ArsR protein (i.e. positive controls) (*47*). Moreover, ArsR-DBD did not bind to several promoter probes that were tested as negative controls (*hopF*, *hp0813*, *flaA*, *tlpC*). Thus, the recombinant ArsR-DBD demonstrated a DNA binding specificity similar to that of the full-length ArsR protein. As the full-length ArsR protein was less soluble and less stable in solution than the ArsR-DBD, all subsequent EMSAs were performed with the ArsR-DBD protein.

ArsR-DBD binding to promoter regions of specific genes identifies novel members of the ArsRS regulon

We next tested the ability of ArsR-DBD to bind to promoter regions of genes encoding proteins that were differentially expressed in the wild-type strain and the *arsS* mutant strain (Table IV-1). As shown in Figure IV-6, ArsR-DBD bound to biotinylated promoter probes for *acxA*, *fabD*, *homD*, *hp0162*, and *metB*. Binding of ArsR-DBD to these biotinylated promoter probes occurred in a concentration-dependent fashion, and binding was competitively inhibited by the addition of excess unlabeled probes (Figure IV-7). No binding of ArsR-DBD to DNA sequences upstream from *acxB*, or *pfs* was detected. This is probably attributable to the localization of these genes immediately downstream of *acxA* and *fabD*; we hypothesize that *acxA/acxB/acxC* and *fabD/pfs* are co-transcribed as polycistronic messages (Figure IV-7).

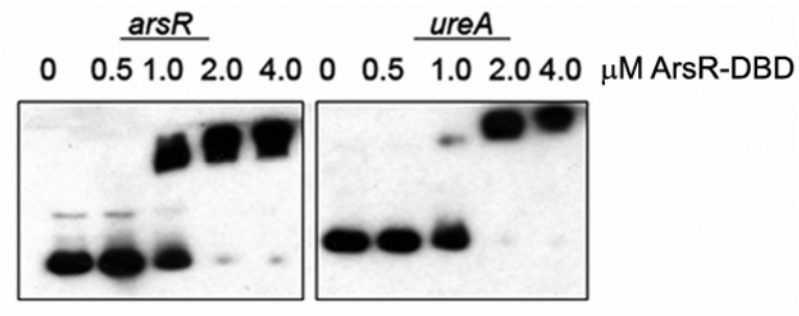


Figure IV-4. Binding of ArsR-DBD to control DNA promoter probes in a concentration dependent manner. Increasing concentrations of the ArsR-DBD protein were added to DNA fragments for 20 min, and the samples were then subjected to non-denaturing PAGE. Separation and visualization of biotinylated DNA was performed as described in the Methods.

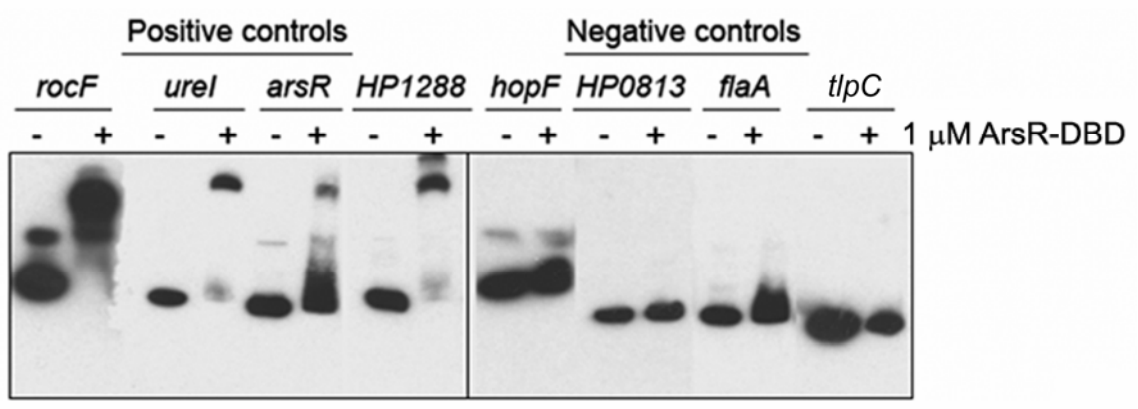
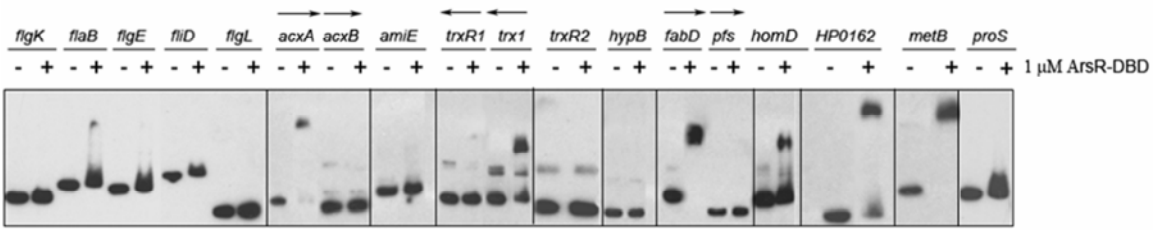


Figure IV-5. Binding of ArsR-DBD to control DNA promoter probes. PCR-amplified fragments of promoters previously tested for ArsR binding (*42, 47*) were incubated with 1.0 μM ArsR-DBD and then subjected to non-denaturing PAGE. Fragments tested included positive control promoter probes (*rocF*, *ureI*, *arsR*, HP1288) known to be bound by ArsR, and negative controls (e.g. *hopF*, HP0813, *flaA, tlpC*). Separation and visualization of biotinylated DNA was performed as described in the Methods.


flgK flaB flgE fliD flgL acxA acxB amiE trxR1 trx1 trxR2 hypB fabD pfs homD HP0162 metB proS
- + - + - + -

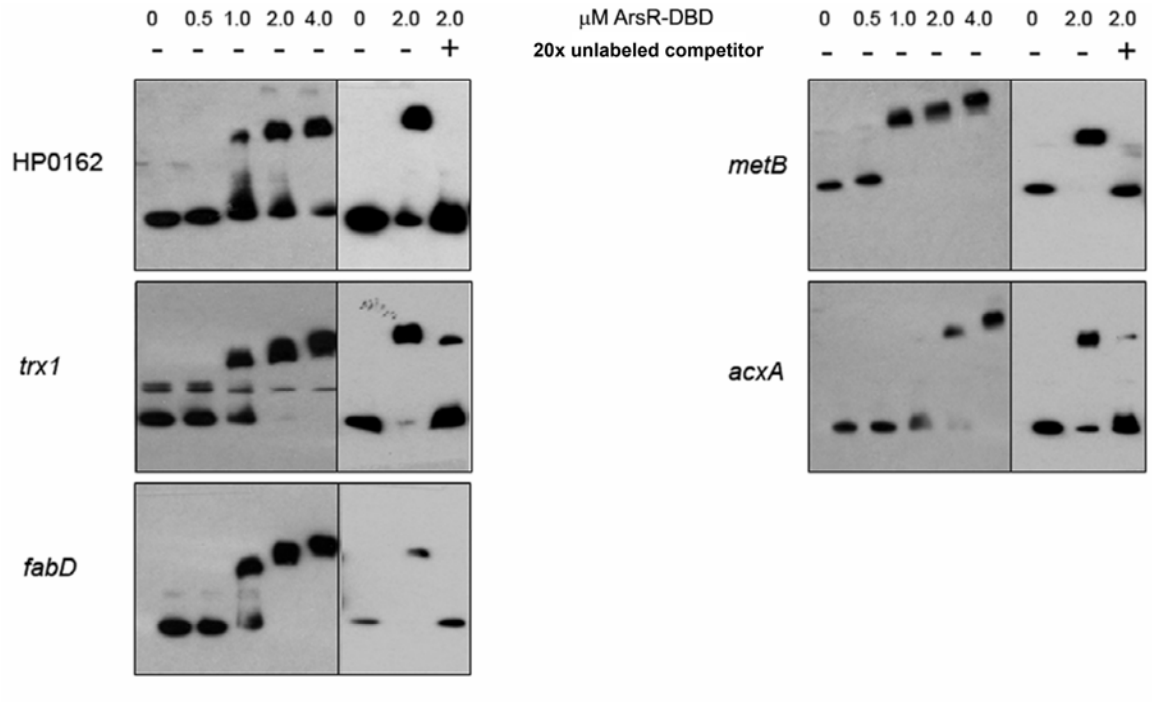


Figure IV-7. Specificity in the binding of ArsR-DBD to target DNA sequences. Varying concentrations (0 to 4.0 μM) of the DNA-binding domain of ArsR (ArsR-DBD) were added to promoter probes for genes encoding differentially expressed proteins (HP0162, *trx1*, *fabD*, *metB*, *acxA)*, and the ability of ArsR-DBD to bind these sequences was examined using non-denaturing gel electrophoresis. For competition assays, reaction mixtures contained 2.0 μM ArsR-DBD with or without a 20-fold excess of non-biotinylated probe.

Similarly, ArsR-DBD did not bind to DNA sequences upstream from *trxR1*, but it did bind to a promoter probe for *trx1*, a gene located immediately upstream of *trxR1* (Figures IV-6 & IV-7). We hypothesize that *trx1* and *trxR1* are co-transcribed. In summary, these EMSA results indicate that ArsR binds directly to the promoter regions located upstream from *acxA*, *acxB*, *acxC*, *trxR1*, *fabD*, *pfs*, *metB*, *homD*, and *hp0162*, and provide evidence that the ArsRS system directly regulates expression of these genes.

The ArsR-DBD did not bind to promoter regions located upstream from several of the genes listed in Table IV-1. For example, ArsR-DBD did not bind to the promoter regions located upstream of the flagellar genes *flgK*, *flgE*, *fliD* and *flgL* (Figure IV-6). This result is consistent with the results of transcriptome analyses (data not shown and (*42*)), which did not detect any *arsS*-dependent differences in transcript levels of these genes.

Discussion

Several previous studies have shown that the ArsRS two-component system is involved in the regulation of *H. pylori* gene expression in response to pH changes in the environment (*26, 42, 46, 47*). Analysis of proteomic profiles of *H. pylori* WT J99 strain and an isogenic *arsS* mutant strain resulted in the identification of 25 proteins that were differentially expressed between the two strains. Additionally, with DNA binding assays, we have provided evidence that the ArsRS system directly regulates the expression of at least 12 of these proteins. Of the genes encoding these 12 proteins, 9 genes are newly recognized members of the ArsRS regulon that were identified in this study. These new members include genes encoding subunits of acetone carboxylase (AcxABC), members of an operon [FabD (predicted malonyl CoA-acyl carrier protein transacylase) and Pfs (methyladenosine nucleosidase)], HomD (a predicted outer membrane protein, HP1453), and HP0162 (a protein of unknown function).

Several of the proteins listed in Table IV-1 (e.g. AmiE, UreB, UreH) are known to be important for survival of the bacteria in acidic pH. Members of the thioredoxin complex may contribute indirectly to *H. pylori* acid resistance by acting as chaperones for RocF (arginase) (*74*), an enzyme known to have an acid protective role in *H. pylori.* The set of differentially expressed proteins identified in the proteomic study included three subunits of acetone carboxylase (AcxA, AcxB, and AcxC), which are predicted to be transcribed in an operon. *H. pylori* is able to use acetone as an alternative carbon source, and mutagenesis of the *acx* operon reduced the ability of *H. pylori* to colonize mice (*75*). A previous study showed that the *acx* operon is regulated by an orphan response regulator, known as HP1021 (*76*). Similarly, the *fabD-pfs* operon also seems to be regulated by both the orphan response regulator HP1021 and the ArsRS system (this study, (*76*)). Thus, there seems to be overlap or cross-talk in the regulation of these target genes by HP1021 and the ArsRS TCS. Cross-talk among TCSs has previously been demonstrated in bacteria such as *E. coli*, and often involves cross-phosphorylation of a response

regulator by a non-cognate histidine kinase (e.g. phosphorylation of PhoB by the non-cognate VanS) (*77, 78*). As HP1021 does not possess a canonical phosphorylation site and is not phosphorylated in vitro by small molecule phosphate donors such as acetyl phosphate (*40*), cross-phosphorylation of HP1021 seems unlikely. Pflock and coworkers (*46*) demonstrated that regulation of HP1021 is not controlled by the ArsRS system, and analysis of gene expression in an HP1021 mutant did not identify *arsR*, or *arsS* genes as part of the HP1021 regulon (*76*). Therefore, a hierarchical organization of these two regulatory circuits, in which one two-component system regulates another, also seems unlikely. As both HP1021 and ArsR bind to the promoter regions of *acxA* (*76*), it is likely that the transcription of the *acxABC* operon is directly regulated by both of these systems. The direct involvement of multiple two-component systems in the regulation of shared target genes (e.g. *cpxR*/*ompR* in *E. coli*) (*79*) is thought to provide for finely regulated gene expression in response to the changing environmental stimuli. Besides the overlap of genes regulated by ArsRS and HP1021, there are also other examples of *H. pylori* genes that seem to be regulated by both the ArsRS system and other transcriptional regulatory systems. For instance, a previous study revealed overlap in the genes regulated by the CrdRS two-component system and members of the ArsRS regulon (*29*). More recently, Wen and coworkers (*30*) demonstrated that a group of acid-responsive genes are regulated by both the ArsRS and FlgRS two-component systems. With the paucity of regulatory genes in the *H. pylori* genome (*36, 38*), it is likely that interplay between the few regulatory circuits allows for meticulous control of gene expression (reviewed in (*22*)).

Members of the ArsRS regulon identified in the current study also included genes encoding Pfs and MetB. Pfs is a methyladenosine nucleosidase that plays a key role in the bacterial activated methyl cycle [reviewed in (*80*)] by converting the toxic S-adenosyl-L-homocysteine to S-ribosyl-L-homocysteine. This latter product can then be converted by the LuxS protein to the quorum signal AI-2 (*80*). The finding that MetB, another protein associated with the methyl cycle, is also a member of the ArsRS regulon (this study, (*42*)) suggests that there could

potentially be a link between acid-induced responses and quorum sensing phenomena in *H. pylori*.

In this study, we assessed binding of ArsR to target promoters by using either recombinant full-length ArsR or the DNA binding domain of ArsR. In general, we did not detect any differences in the specificity of target DNA binding by full-length ArsR and ArsR-DBD. The ability of the DNA-binding domain of a response regulator to bind to or activate target genes has been demonstrated in other two-component systems such as the PhoBR (*81*) or the CpxAR (*82*) systems in *E. coli*. In the case of the response regulator PhoB, the C-terminal DNA-binding domain bound to a target *pho* box sequence with an affinity 7 times higher than that of the unphosphorylated full-length protein (*81*).

Another interesting observation of our DNA-binding studies is that ArsR binds to promoters upstream from genes whose expression is upregulated and as well as genes whose expression is downregulated in the wild type compared to an *arsS* null mutant *H. pylori* strain. While RRs can bind DNA upstream from promoters of target genes and upregulate transcription via synergistic interactions with subunits of RNA polymerase, they can also compete for RNA polymerase binding sites in target gene promoters, resulting in downregulation of their transcription (reviewed in (*83*)). The ability of ArsR to act as an activator and a repressor of different target genes, a phenomenon also seen with RR BvgA in *Bordetella pertussis* (reviewed in (*83*)), exemplifies the efficiency of *H. pylori* in exploiting the few transcription factors encoded in its small genome (*36, 38*) to regulate a range of cellular functions.

We propose that ArsRS plays a role in both pH-dependent and pH-independent regulation of *H. pylori* gene expression as described below. Based on previous results indicating a clear role for ArsS in pH-responsive gene expression (*26, 42, 47*), ArsS responds to changes in pH by phosphorylating its cognate RR ArsR. Consequently, levels of phosphorylated ArsR are higher in the wild-type strain at pH 5.0 than at pH 7.0. ArsR can bind to the promoter regions of pH-responsive genes and essential genes, as well as the promoter region of *arsR* itself. Current

evidence suggests that the unphosphorylated form of ArsR modulates expression of essential genes (*44*) and that the phosphorylated form modulates expression of pH-responsive genes and *arsR* (*24, 42*). It is possible that the phosphorylated and unphosphorylated forms of ArsR interact with the same DNA targets with different affinities (*84, 85*), or that phosphorylated RR binds to target sites that are different from those of the unphosphorylated response regulator (*86*). Results of previous studies indicate that binding of phosphorylated ArsR to its own promoter region results in repression of *arsR* transcription (*42, 44*), suggesting that there are differences in the levels of ArsR in wild-type and mutant strains, and that levels of ArsR in the wild-type strain differ depending on whether the strain is grown at neutral pH or acidic pH (*19*).

In summary, these results provide new insight into the genes that are regulated by the ArsRS system. We confirm that several previously identified target genes are regulated by this system, and we identify multiple new members of the ArsRS regulon. The results of this study emphasize the importance of the ArsRS system in regulating gene expression in *H. pylori* under both neutral and acidic pH conditions.

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions

H. pylori infection of the human stomach has been linked to an increased risk of developing peptic ulcers, gastric adenocarcinoma, and gastric MALT lymphoma (reviewed in (1-8)). This neutrophilic organism is highly adapted to colonize the gastric niche and its harsh conditions such as acidic pH (reviewed in (*2, 15, 16*)). One of the mechanisms by which *H. pylori* responds to acidic pH is by regulating expression of acid-adaptive proteins such as the enzymes urease, amidases, α-carbonic anhydrase, and arginase, which produce buffering compounds to help neutralize the bacterial cytoplasm and periplasm (*23-26*). The regulation of these and other acid responsive genes occurs through an intricate interplay of transcriptional regulatory circuits, each responding to different primary stimuli including pH changes in the bacterial environment ((*25*), and reviewed in (*22*)).

One of the major players in regulating the expression of acid adaptive genes in *H. pylori* is the TCS ArsRS (*24, 25, 29, 37, 42, 44-47*). The significance of this TCS is highlighted by studies that show that the RR ArsR is essential for *H. pylori* viability (*37*), and deletions of the HK ArsS impairs bacterial ability to grow in acidic pH conditions (*29*). While several studies have contributed to an understanding of the ArsRS regulon (*20, 21, 24, 26, 42, 44, 46-48*), a comprehensive description of this regulon is still lacking. Furthermore, little is known about the mechanisms controlling ArsR mediated gene regulation and acid adaptation. Evidence suggests a model for differential gene regulation by ArsR in its phosphorylated versus unphosphorylated forms (*37, 40*). In this model, the unphosphorylated form of ArsR is predicted to modulate the expression of essential genes, while the phosphorylated form regulates expression of pH-

responsive genes. While the two sets of targets could be distinct genes, it is possible that both forms of ArsR interact with the same target genes but with different affinities.

Studies to date have largely limited their focus to describing members of the ArsRS regulon. Structural and biochemical analyses of ArsR and ArsS would be novel approaches to understanding the mechanisms underlying their physiological roles. The main goal of this thesis was to characterize the structure and function of members of the TCS—in particular, the response regulator ArsR.

Chapter II of this thesis describes the overexpression of His_6-tagged full length forms and DBDs of ArsR, HP1043, HP1021, CrdR, and the DBD of FlgR. The full-length ArsR and its DBD were purified and further analyzed via 1D ^{1}H NMR. Well dispersed peaks for both ArsR and ArsR-DBD in both upfield (0-4 ppm, characteristic of aliphatic peaks) and downfield (6-10 ppm, characteristic of amide peaks) regions of the spectra indicate that the purified protein is folded. A comparison of the 2D ^{1}H-^{15}N HSQC spectra of ArsR and ArsR-DBD indicates that the structure of the isolated ArsR-DBD is very similar to that domain in the full-length protein and expression of the DBD alone does not seem to alter protein folding. ^{31}P NMR experiments conducted on ArsR demonstrate that the majority of the purified protein is in the unphosphorylated form.

Since RRs can differentially regulate gene transcription depending on their phosphorylation and oligomeric states (*49-52*), it is important to characterize these biochemical features of ArsR. Data describing these biochemical features together with results from DNA-binding assays can be extrapolated to predict how changes in phosphorylation or oligomeric state can alter ArsR's ability to regulate gene expression. In previous studies, there has been discrepancy in the reported phosphorylation state of recombinant ArsR purified from *E. coli* (*42, 44, 47*), and nothing is known about its oligomeric state. Our data clearly demonstrate that majority of ArsR purified from *E. coli* is not phosphorylated and both ArsR and ArsR-DBD behave as monomers in solution at concentrations as high as 0.1 mM. Thus, results from future biochemical analyses

with ArsR and ArsR-DBD proteins (in conditions similar to ones described in this study) can be interpreted to correspond to the unphosphorylated, monomeric form of the protein, and monomeric form of the protein, respectively.

In Chapter III, we present the three-dimensional structure of the DBD of ArsR, determined by NMR spectroscopy. The ArsR-DBD comprises two anti-parallel β-sheets flanking a core of three α-helices, and contains a winged HTH motif that is predicted to bind DNA. Presence of this tertiary fold classifies ArsR with the OmpR/PhoB subfamily of response regulators, as has been predicted by sequence analysis (*37*). However, one striking difference between ArsR-DBD and the DBDs of related RRs of this subfamily is the orientation of the N-terminal β-sheet. Residues from this β-sheet may be involved in interdomain interactions between the N-terminal receiver domain (NTD) and the C-terminal DBD and thus may align the NTD with the DBD in a conformation distinct from what is observed in other members of the OmpR/PhoB subfamily. Conventionally, RRs' NTDs regulate the activity of the corresponding DBDs (reviewed in (*34*)), often by providing steric inhibition to a functionally important site either in the NTD itself (as is suggested for PhoB in (*53*)), or in the DBD (as for NarL, (*71, 72*)). Thus, interdomain interactions can play a significant role in determining the activity of an effector domain. In ArsR, a unique arrangement of the N-terminal β-sheet of the DBD could facilitate interdomain interactions that are significant for NTD-mediated regulation of effector domain function.

One of the conserved structural features between ArsR and members of the OmpR/PhoB subfamily of RRs is the presence of the DNA-binding winged helix-turn-helix motif. A comparison of NMR spectra of ArsR-DBD alone and ArsR-DBD in complex with a promoter fragment of target gene *hp1408* demonstrated several residues exhibiting prominent reductions in peak intensities. These residues mapped to surface-exposed regions of the wHTH motif (residues R173, D193, and R200) in addition to several core residues of the wHTH motif and other elements of the protein that make close contacts with the helices of the wHTH motif (residues A155, I159, I183, I195, I196, G197, and G220). Conservation of protein sequences in these

regions within ArsR and related proteins supports our prediction that ArsR-DBD interactions with DNA involve molecular surfaces homologous to those identified for other members of the OmpR/PhoB subfamily.

Based on the evidence presented above, we predict that ArsR binds to DNA comparably to related RRs from the OmpR/PhoB subfamily. Since most members of the OmpR/PhoB subfamily of RRs interact with DNA as dimers, we predict that ArsR binds to its target gene sequences in a dimeric manner as well. Using the structure of the PhoB-DBD in complex with a target promoter sequence (*53*) as a template, we modeled ArsR-DBD in complex with a putative target DNA fragment. While several binding sites of ArsR have been reported thus far, the predicted ArsR binding sites do not contain a conserved consensus binding sequence (*24, 42, 44, 47*). Since the predicted ArsR binding site sequence is highly degenerate, the protein-DNA interactions of ArsR may differ from those observed for PhoB. In this study, we present two different models for ArsR-DBD dimers (interacting in a head-to-tail or a tail-to-tail manner) binding to target DNA sequences. The residues of the wHTH motif make significant contacts with DNA in both our models. This observation is in agreement with data from our NMR analysis of ArsR-DBD in complex with the promoter of *hp1408*.

In Chapter IV, we present data to further characterize the ArsRS regulon. Proteomic profiles of *H. pylori* WT J99 and an isogenic *arsS* mutant strains were searched and we identified 25 proteins that were differentially expressed between the two strains. DNA binding assays verified that the ArsRS system directly regulates the expression of at least 12 of these proteins. Newly recognized members of the ArsRS regulon identified in this study include genes encoding subunits of acetone carboxylase (AcxABC), members of an operon [FabD (predicted malonyl CoA-acyl carrier protein transacylase) and Pfs (methyladenosine nucleosidase)], HomD (a predicted outer membrane protein), and HP0162 (a protein of unknown function). With the discovery of these new members of the ArsRS regulon, our studies underscore the importance of

this TCS in regulating expression of proteins that contribute to a variety of cellular functions in *H. pylori*.

Future directions

Even though our studies have expanded our knowledge of the ArsRS TCS and its role in regulating transcription in *H. pylori*, there are still many questions remaining regarding its biological functions and the interactions between this TCS and other regulatory proteins in *H. pylori* to synchronize a response to stimuli such as changes in environmental pH. Some of the questions that we find interesting are:

- Is ArsS a direct pH sensor?

  ArsS is predicted to be a pH-sensing HK. A low pH stimulus is putatively sensed by its periplasmic domain. This periplasmic domain contains several histidine residues that have a pK_a of ~ 6. These residues can change their protonation state as the periplasm pH falls from 7 to lower than 6, and thus may serve as indicators of environmental pH affecting the activation state of ArsS (*47*). There is, however, no direct evidence to confirm that ArsS is a pH-sensing protein. Future studies attempting to dissect the putative signals sensed by the periplasmic domain of ArsS would clarify the role of this TCS in mediating an acid-adaptive response of *H. pylori*.

- How does phosphorylation alter the DNA-binding activity of ArsR?

  Upon phosphorylation, some RRs dimerize in solution, whereas others remain monomeric and subsequently dimerize after binding to specific DNA sequences (*49-52*). We have shown that ArsR purified from *E. coli* is unphosphorylated and behaves as a monomer in solution. In addition, our structural data confirm that ArsR belongs to the OmpR/PhoB subfamily of RRs. Since most members of the OmpR/PhoB subfamily of RRs interact with DNA as dimers, we predict that ArsR binds to its target gene sequences in a dimeric manner as well. Changes in a protein's local structure and/or assembly of novel protein-protein complexes in

response to phosphorylation can be tested by structural methods, such as electron paramagnetic resonance (EPR).

- What are the essential targets of ArsR?

  Attempts to generate an *H. pylori* mutant strain with disruptions in the *arsR* gene have been unsuccessful, suggesting that this RR is essential for bacterial viability (*37*). However, an *H. pylori* mutant strain expressing a derivative of ArsR with a mutation in the phosphate-receiving aspartate residue (D52N) is viable (*40*). Furthermore, the gene encoding the cognate HK ArsS can be deleted without affecting *in vitro* growth of *H. pylori* at pH 7.0, but the growth of this mutant strain is impaired compared to a wild type strain at pH 5.0 (*29, 37*). These data suggest that unphosphorylated ArsR regulates the expression of a specific set of genes, at least one of which is essential for *H. pylori* viability, while phosphorylated ArsR regulates expression of acid-adaptive genes as a member of the ArsRS TCS. Many of ArsR's non-essential acid-adaptive targets have been described; however, the essential genes regulated by this RR have yet to be identified. Identification of these essential targets would be the first steps in unfolding a novel role for ArsR in regulating essential cellular functions.

- What is the ArsR target gene binding site?

  While several binding sites of ArsR have been reported thus far (from promoters of genes *arsR*, *ureA*, *ureI*, *amiE*, *amiF*, *rocF*, *hp1408*, and *hp1186* (carbonic anhydrase)), analysis of these data reveals that these sequences do not contain a consensus binding sequence (*24, 42, 44, 47*). The degeneracy of ArsR binding sites suggests that while the proteins may not bind to a conserved consensus site, they may bind with variable affinities to binding sites with preferred nucleotide arrangements. Such a pattern of binding DNA with variable affinities has been demonstrated for the *H. pylori* transcription factor NikR. NikR does not seem to bind a classic symmetric binding site, but instead binds to sites with conserved purine versus pyrimidine compositions (*87*).

In our preliminary efforts, we analyzed promoter sequences of known ArsR target genes to identify conserved consensus sequences. Our studies with DNA sequence analyzing programs such as CLUSTALW (*65*), and TRES (Transcription Regulatory Element Search) (*88*) did not reveal any obvious consensus DNA sequence present in the promoter regions of ArsR target genes. For future studies, we intend to use a more sophisticated sequence analysis approach in collaboration with Dr. Antonis Rokas, Vanderbilt University. DNA sequences in the upstream regions of genes clustered in the same gene expression pattern groups (for example, positively-regulated versus negatively-regulated) in the various strains of *H. pylori* can be analyzed to search for conserved regulatory motifs. Bioinformatic programs, for example BioProspector, that can search for gapped motifs and motifs with palindromic patterns, as are predicted for ArsR, can help propose a 'consensus sequence' of favored positions of preferred bases that are conserved within the promoter regions of target genes (*89*).

- Is ArsR involved in mediating bacterial responses to stimuli other than changes in environmental pH?

  ArsR regulates the expression of many genes, including the gene encoding ArsR itself (reviewed in (*22*)). Many of the genes belonging to the ArsRS regulon are also regulated by other *H. pylori* transcription systems, for example, HP1021, and the CrdRS and FlgRS TCSs (*29, 30, 76*). If *arsR* belongs to the regulons of other transcriptional systems, stimuli regulating the activities of these systems could mediate an indirect effect on expression of members of the ArsRS regulon by regulating the expression of ArsR. DNA-binding assays testing interactions between the promoter of *arsR* and transcription factors from *H. pylori* can verify whether *arsR* is regulated by transcription factors other than ArsR itself.

- Is ArsR-DBD alone sufficient for *H. pylori* viability?

  Evidence suggests that in its unphosphorylated state, the DNA binding functions of PhoB are inhibited by its receiver N-terminal domain (NTD) (*81*). This inhibition is relieved upon

phosphorylation of the protein, generating an activated form of PhoB with enhanced DNA binding abilities. In fact, if the NTD is removed completely, the PhoB-DBD binds to DNA with 7 times greater affinity than the unphosphorylated full length protein (*81*). To date, the only solved structure of a response regulator from the OmpR/PhoB subfamily in complex with a target DNA sequence is that of the DBD of PhoB bound to the promoter region of the *phoA* operon. Two PhoB-DBD molecules bind in tandem on a direct repeat sequence in the target DNA fragment. Based on their structural data, Blanco and coworkers (*53*) have proposed a mechanism for the phosphorylation-mediated activation of PhoB. In their model, in the absence of phosphorylation, the orientation of the NTD with respect to the DBD produces steric clashes should the protein attempt to fit in its binding site with another PhoB molecule in the adjacent binding site. Phosphorylation induces a conformational change in the NTD that relieves the intermolecular steric inhibition, allowing a stronger dimeric interaction between two PhoB molecules and DNA (*53*).

Our structural data suggest that ArsR-DBD is closely related to PhoB-DBD, with the two proteins sharing a high degree of structural similarity. Furthermore, through EMSA analyses, we have shown that ArsR-DBD mimics the full length protein in binding to specific DNA sequences. We hypothesize that like PhoB-DBD, ArsR-DBD when expressed alone is no longer inhibited by its NTD and thereby behaves as the domain would in the phosphorylated protein. Thus, a mutant *H. pylori* strain expressing ArsR-DBD alone should not only be viable, as the protein would retain its ability to regulate essential genes, but should be able to grow in low pH by regulating genes that full length ArsR would subsequent to phosphorylation by ArsS. This hypothesis makes the assumption that the regulatory functions of RR, including interactions with RNA polymerase, are mediated via the DBD.

To test this hypothesis, we are in the process of generating plasmids to be used to produce a mutant *H. pylori* strain in which the gene for full length ArsR has been replaced with a truncated version of the protein, a His_6-tagged ArsR-DBD. A schematic of the proposed

plasmid is presented in Figure V-1. The plasmid will be used to transform *H. pylori* J99 strains using natural transformation and positive recombinants will be selected by antibiotic resistance. Subsequent to obtaining viable mutant strains expressing the truncated form of ArsR, it would be of interest to test our hypothesis that ArsR-DBD alone is sufficient for *H. pylori*'s ability to survive acidic conditions via growth assays in low pH conditions as described in (*29*). Furthermore, differences in transcript and proteomic profiles of the bacterial strains (WT vs. ArsR-DBD alone mutant) grown at pH 5.0 vs. pH 7.0 could be compared to identify members of the ArsRS regulon whose gene expression is altered in the absence of the regulatory NTD of ArsR.

- Are FlgR, CrdR, and HP1021 structurally related to any other known RRs?

  In this thesis work, we successfully characterized the biochemical features of ArsR that are predicted to influence its biological functions. We also presented a detailed three-dimensional structure of its DBD that allowed us to classify the RR into a well-described subfamily of RRs, and to predict DNA-binding modes based on comparison with related proteins. Such an analysis could be extended to other members of the *H. pylori* RR repertoire, including RRs FlgR, CrdR, and HP1021.

  The few transcription factors encoded in the *H. pylori* genome are predicted to interact in a precise interplay to regulate common cellular processes in response to different physiological stimuli. Evidence demonstrates overlap in genes regulated by ArsRS and those regulated by other two-component system proteins (our study, (*29, 30*)). Our biochemical and structural studies have provided clues to understanding how ArsR interacts with target DNA sequences. Analyses of the structures of FlgR, CrdR, and HP1021 can be used to identify related proteins. Based on conserved features between the RRs and corresponding related proteins, we can predict modes of protein-protein interactions, characteristics of target DNA sequences, and the biochemical processes, such as phosphorylation, that regulate the functions of these proteins.

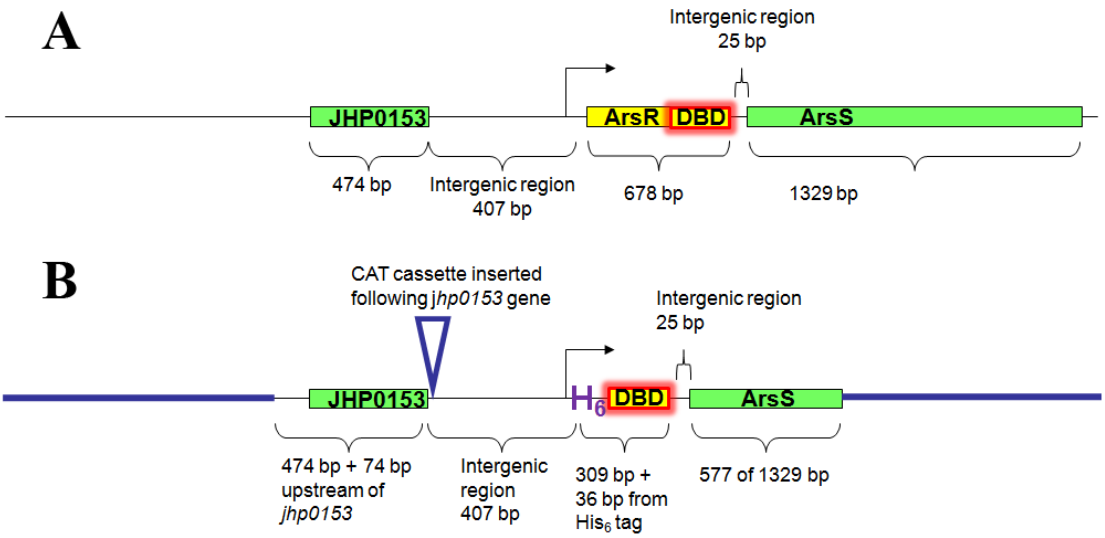


Figure V-1. Experimental design to generate an *H. pylori* strain J99 mutant expressing ArsR-DBD alone. A. Schematic of the genomic region of ArsR in *H. pylori* J99 strain that will be manipulated to remove the native *arsR* gene and insert a fragment encoding ArsR-DBD alone. Shown are *arsR* and flanking genes, including *arsS* and the corresponding intergenic regions as indicated. **B.** Schematic of a portion of a plasmid to be used to transform *H. pylori* strain J99. Via homologous recombination, the plasmid will replace the *arsR* gene with a gene fragment that allows expression of the His_6-tagged DBD alone. Our experimental design includes the insertion of a CAT cassette in the construct for chloramphenicol-based selection of recombinants. The cassette will be inserted following the *jhp0153* gene, maintaining the region upstream of the *arsR* transcription start site, including its promoter region, intact. A similar construct to express a His_6-tagged full length ArsR protein will be designed to be used as a positive control in our experiments.

We have already shown that we can overexpress His_6-tagged full length CrdR, HP1021 and DBDs of CrdR, FlgR, and HP1021. These proteins can be purified and analyzed by size exclusion chromatography and NMR as described in this thesis. Structural relatedness to known proteins can provide clues to the DNA-binding and gene regulatory characteristics of RRs that are still relatively unknown.

- What are the essential target genes of the RRs HP1043 and HP1021?

  Attempts to generate an *H. pylori* mutant strain with disruptions in the *hp1043* and the *hp1021* genes have been unsuccessful, suggesting that these RRs are essential for bacterial viability (*37*). The conserved phosphate receiving aspartate residues in HP1043 and HP1021 are replaced by lysine and serine, respectively (*37*), and *H. pylori* strains expressing derivatives of these RRs with mutations in alternate putative phosphate receiving residues were viable (*40*). Furthermore, these RRs are not phosphorylated *in vitro* by the *H. pylori* HKs (*37*) or by phosphate donor acetylphosphate (*40*). These data suggest that phosphorylation of HP1043 and HP1021 is not needed for the essential functions of these RRs, and may not occur at all (*40*). The essential genes regulated by this RR have yet to be identified. With increasing evidence of overlap between regulons of various transcription factors in *H. pylori* (*29, 30, 76*), essential targets identified for HP1043 and/or HP1021 could potentially be the essential targets regulated by RR ArsR.

- If phosphorylation is not required for the essential functioning of the atypical RRs HP1043 and HP1021, are their DBDs alone sufficient for *H. pylori* viability?

  Evidence suggests that HP1043 and HP1021 are atypical phosphorylation-independent RRs (*37, 40, 41*). Since their activity is predicted to be independent of phosphorylation-based NTD regulation, we predict that the corresponding DBDs should be sufficient for *H. pylori* viability. This hypothesis can be verified by generating viable mutant *H. pylori* strains expressing HP1043-DBD or HP1021-DBD alone.

- If multiple transcription factors are regulating the expression of the same genes, are they interacting on the promoter sites of target genes? If so, is this interaction crucial to their biological activity?

  The ArsRS regulon overlaps with regulons of other *H. pylori* transcription systems such as RR HP1021, and the CrdRS and FlgRS TCSs (*29, 30, 76*). A study comparing the binding affinities of various RRs regulating the same gene could shed light on the interplay among TCSs to regulate gene transcription. While different RRs could be competing for common binding sites, it might be worthwhile to consider the possibility of interactions between different RRs as a mechanism to stabilize their interactions with DNA. For example, in the case of ArsRS and FlgRS, the corresponding RRs are predicted to bind promoters of a common set of genes in response to the same stimulus—changes in pH—and thus, may interact with the same promoter sequences concurrently (*30, 42*). Protein-protein interactions can be mapped via well characterized assays such as yeast two hybrid analyses. Site directed mutagenesis could be used to introduce mutations in predicted surface residues of interacting proteins. Comparing *in vitro* growth and acid adaptive capabilities of wild type strains versus mutants expressing these protein derivatives could verify the significance of these protein-protein interactions for *H. pylori* viability and acid response.

Expanding our knowledge about the TCSs of *H. pylori* can help us understand how this organism adapts to its very specific niche of the human stomach. Despite the prevalence of gastric infection in over half the population of the world, the occurence of *H. pylori* related gastric disease is limited to a subset of individuals. Strain variations play a significant role in determining whether diseases such as peptic ulceration or gastric cancer occur. Knowledge of features that render some strains more pathogenic than others, together with an understanding of biological functions crucial to *H. pylori*'s ability to survive gastric stresses can be used to design tailored therapeutics for its eradication in individuals deemed to have high-risk infections.

Current projects in our lab are directed at answering some of the questions listed above. In the following section, we present the experimental design and preliminary results for one of these projects.

How does phosphorylation alter the DNA-binding activity of ArsR?

Response regulators differ in their mechanisms of activation, including their dependence on phosphorylation and oligomerization for transcriptional regulation. Upon phosphorylation, some RRs dimerize in solution, whereas others remain monomeric and subsequently dimerize after binding to specific DNA sequences (*49-52*). At present, relatively little is known about the effects of phosphorylation and oligomerization on the activity of RRs from *H. pylori*, in particular, ArsR. We have shown that ArsR purified from *E. coli* is unphosphorylated and behaves as a monomer in solution. Based on the structure of its DBD, we have verified that ArsR belongs to the OmpR/PhoB subfamily of RRs. Since most members of the OmpR/PhoB subfamily of RRs interact with DNA as dimers, we predict that ArsR binds to its target gene sequences in a dimeric manner as well.

We have proposed two models for ArsR-DBD-DNA interactions. One of the models was generated by placing two ArsR-DBD domains in tandem orientation (head to tail) in the major groove of a DNA fragment containing putative direct repeat sequences. The lack of consensus in reported ArsR binding sites allowed us to explore the possibilities of novel protein-DNA interaction modes. Thus, we arrived at a second model in which ArsR-DBD molecules are arranged symmetrically (tail to tail) on putative inverted repeat sequences in the DNA fragment. In both orientations, the same residues from the DBD molecules make contacts with the phosphate backbone and/or the bases of the target DNA. However, in the "tail to tail" model, an interaction between hydrophobic patches on the two ArsR-DBD molecules formed by residues I214 and V216 of each C-terminal β-hairpin is possible.

The validity of our proposed models can be tested via structural methods such as EPR. EPR based experiments can be used to identify any changes in ArsR's tertiary structure and

oligomerization state upon phosphorylation, and/or upon binding promoter sequences of target genes. Spectroscopic probes, or spin-labels, are introduced at selected sites on the protein that mark regions of interest. Interacting pairs of spin labels can be detected by EPR as they undergo dipolar coupling in the 5-20 Å range, which is manifested as a broadening of the EPR lineshape (*90*). This allows detection of local structural changes upon complex formation (Figure V-2, panels A and C adapted from (*72*)).

ArsR contains a cysteine residue, C64, in its native form. In our preliminary efforts, we have been successful in labeling this native cysteine residue with reagent (**I**), kindly provided by Dr. Hassane Mchaourab, Vanderbilt University Medical Center, as confirmed by EPR (Figure V-3A). Upon adding a phosphomimetic compound, BeF_3^- (*91*), to the protein sample, no changes were observed in the spectral line shape (Figure V-3B). A similar result was observed when the protein was incubated with a target DNA sequence from the promoter of the gene *hp1408*, the same DNA fragment used for our NMR analysis discussed in Chapter III (Figure V-3D). Even when DNA was added to the mix of protein with BeF_3^-, no changes to the spectral line shape were observed (Figures V-3C, V-3E). While it is possible that the protein is unaffected by the introduction of a phosphomimetic agent, our NMR studies provide evidence that the DBD interacts with the target DNA fragment. However, any changes in the protein's structure and any possible resulting intra- or inter-molecular interactions are not being detected by EPR. A possible explanation for this result is that even if phosphorylation and/or DNA binding is inducing changes in protein structure, the regions undergoing conformational change and possibly participating in intra- or inter-molecular interactions, are beyond of the range of detection by the spin-label.

Current experimental settings allow a detection of changes in the local environment of the spin-label within a range of ~5-20 Å. The DNA-binding recognition helix, α3, is at least 40 Å from the spin-labeled C64 residue, and it is very likely that the surfaces involved in phosphorylation mediated conformational changes lie at distances greater than 20 Å from the

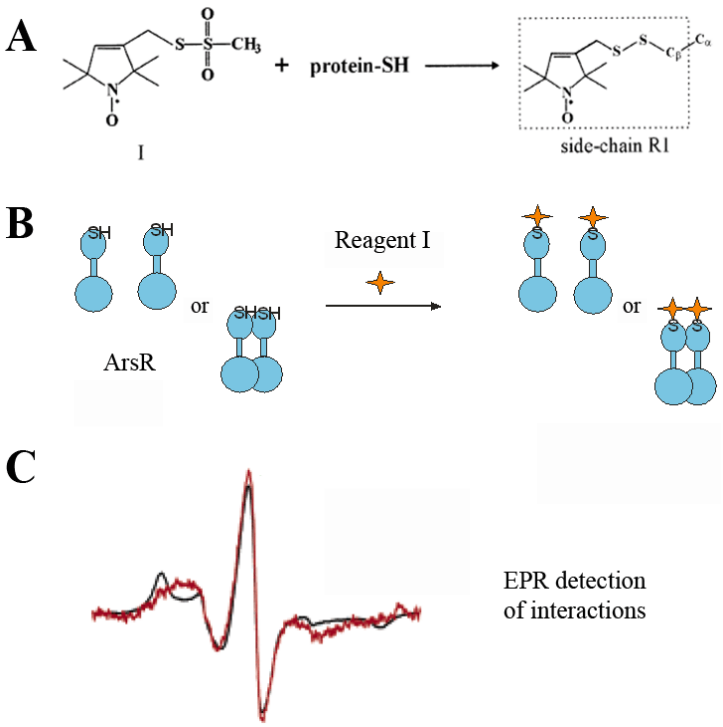


Figure V-2. EPR based approach to study protein-protein interactions. **A.** Reaction of (1-oxy-2,2,5,5-tetramethylpyrrolinyl-3-methyl)methanethiosulfonate, or reagent (**I**), a thiol specific spin label, with a protein sulfhydryl group to generate the nitroxide side chain R1. **B.** Schematic of an approach to detect protein-protein interactions via EPR. Changes in the local environment of the spin-labels can be monitored by observing changes in the mobilities of single nitroxide side chains and/or observing nitroxide-nitroxide interactions (reviewed in (*92*)). **C.** Comparison of spectral line shapes for spin-labeled proteins in different conditions can reveal the effects that change had on the local environment of the spin-label. Conformational changes that result in a buried spin-label becoming surface exposed, or that allow intermolecular interactions bringing spin-labels in close proximity can be detected via EPR. For example, comparing the line shapes for the protein before (black tracing) and after (red tracing) phosphorylation reveals that the label whose mobility was previously constrained is relieved of those restrictions after phosphorylation. Thus, the residue that was spin-labeled is in a region that was putatively buried and thus immobilized due to the dense packing in the interior of the protein. However, phosphorylation-induced conformational changes induced surface exposure of this buried site, leading to increased mobility of the spin-label, and therefore the observed broadening in the corresponding lineshape (panels **A** and **C** adapted from (*72*)).

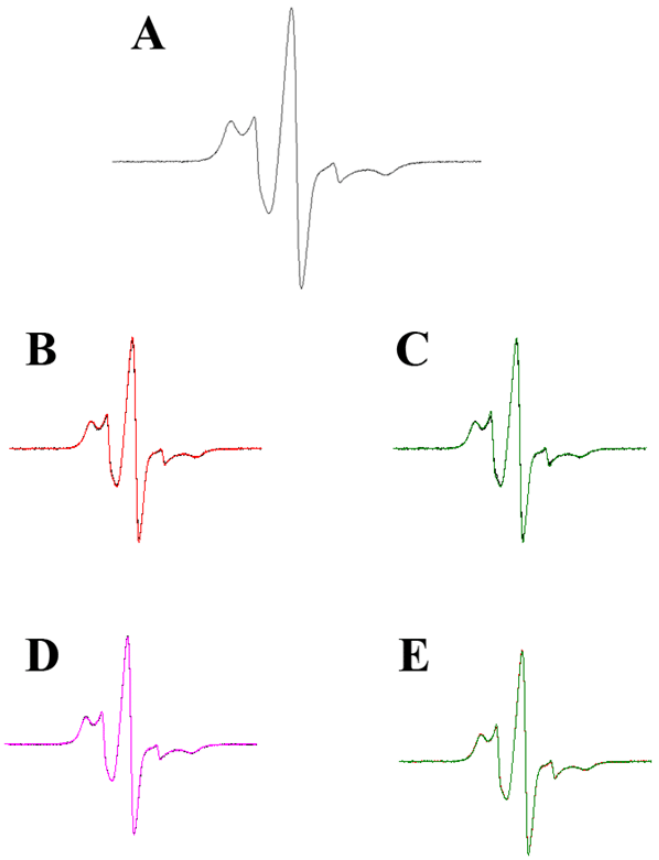


Figure V-3. EPR analysis of ArsR. **A.** EPR spectrum of WT ArsR spin-labeled at residue C64 (C64(**I**)-ArsR) protein. **B.** Comparison of spectral line shapes of C64I-ArsR protein in buffer containing $MgCl_2$ without (black tracing) and with BeF_3^- (red tracing). **C.** Comparison of spectral line shapes of C64I-ArsR protein in buffer containing $MgCl_2$ without (black tracing) and with BeF_3^- and a 38 bp DNA fragment from the *hp1408* promoter (green tracing). **D.** Comparison of spectral line shapes of C64I-ArsR protein in buffer containing $MgCl_2$ without (black tracing) and with a 38 bp DNA fragment from the *hp1408* promoter (pink tracing). **E.** Comparison of spectral line shapes of C64I-ArsR protein in buffer containing $MgCl_2$ and BeF_3^- without (red tracing) and with a 38 bp DNA fragment from the *hp1408* promoter (green).

spin-label, and out of its range of detection. To rectify this shortcoming, we want to apply site-directed mutagenesis to introduce novel cysteine residues in ArsR that can be spin-labeled and observed with EPR under conditions similar to the ones described above for WT C64(**I**) ArsR above. Such a site-directed spin labeling (SDSL) approach has been employed successfully by Hubbell and coworkers to study protein folding and to explore structure dynamics (reviewed in (*92*)) in proteins including a response regulator NarL (*72*).

The pET-BNK plasmid construct containing the full-length *arsR* protein coding sequence (described in Chapter II) has been mutated to encode a derivative of ArsR with the native cysteine residue mutated to a serine (C64S) that cannot be spin-labeled. This plasmid was used as the template to introduce novel cysteine residues into the protein coding sequence of *arsR*. We have successfully generated constructs expressing ArsR C64S derivatives with novel cysteine residues at sites we believe may undergo phosphorylation-mediated changes, and/or may be involved in intra- or intermolecular interactions subsequent to phosphorylation and/or DNA binding. For example, by introducing an E224C mutation, we have produced an ArsR mutant that can be spin-labeled at the C-terminal surface of the DBD (Figure V-4). This region is predicted to be a dimeric interface for the DBD in one of our ArsR-DBD-DNA models, described in Chapter III. In this model, the two glutamate residues lie within ~7 Å of each other, and should this dimeric interaction occur when the protein binds to a target DNA sequence, the spin-spin interaction can be detected by EPR. Additional mutants designed in our lab emphasize the various predicted biochemically relevant surfaces on ArsR, including surfaces on the NTD based on homology to related proteins (*93*). Spin-labels introduced at these sites can monitor structural changes in the protein and novel intra- and intermolecular interactions.

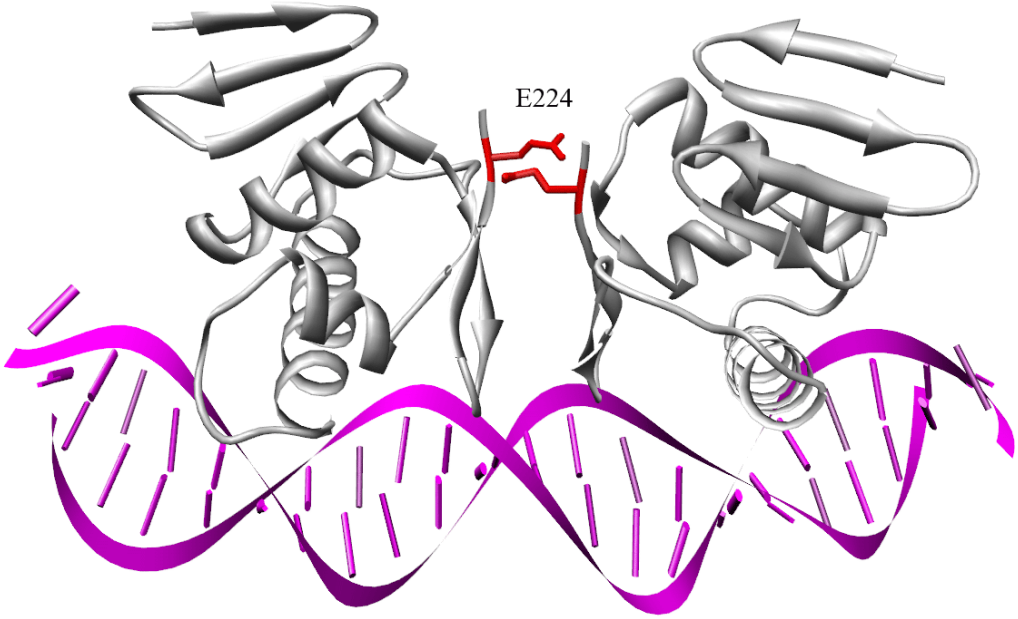


Figure V-4. Analysis of protein-protien interactions by EPR: a putative model of the ArsR-DBD-DNA interaction. Two ArsR-DBD molecules are positioned in a "tail to tail" manner on a 23 bp ds DNA sequence containing an inverted repeat. For details on model, refer to Figure III-12. Highlighted in red is residue E224 that has been mutated to a cysteine in an ArsR C64S, E224C derivative to be used for EPR analysis. The predicted distance between spin-labels introduced at this novel cysteine position in this conformation is ~7 Å.

APPENDIX A

LIST OF PUBLICATIONS

Gupta, S. S., Borin, B. N., Cover, T. L., and Krezel, A. M. (2009) Structural analysis of the *Helicobacter pylori* response regulator ArsR, *The Journal of Biological Chemistry 284*, 6536-6545.

Loh, J. T., Gupta, S. S., Friedman, D. B., Krezel, A. M., and Cover, T. L. Analysis of protein expression regulated by the *Helicobacter pylori* acid-responsive two-component signal transduction system (ArsRS), (submitted for publication).

APPENDIX B

NUCLEOTIDE AND AMINO ACID SEQUENCES OF SELECTED RESPONSE REGULATORS AND CORRESPONDING DNA-BINDING DOMAINS

ArsR

```
atg ata gaa gtt tta atg ata gaa gat gat ata gaa tta gcc gag   45
M   I   E   V   L   M   I   E   D   D   I   E   L   A   E
                5                   10                  15

ttt ttg agc gag ttt ttg ctc caa cat ggc att cat gta atc aat   90
F   L   S   E   F   L   L   Q   H   G   I   H   V   I   N
                20                  25                  30

tac gat gag cca tac acc ggc att agt gcg gct aac aca caa aat   135
Y   D   E   P   Y   T   G   I   S   A   A   N   T   Q   N
                35                  40                  45

tat gat ttg ttg tta ttg gat ttg act ttg cct aat tta gac ggg   180
Y   D   L   L   L   L   D   L   T   L   P   N   L   D   G
                50                  55                  60

ctt gaa gtg tgt agg cgc att tcc aaa caa aaa cat atc cct att   225
L   E   V   C   R   R   I   S   K   Q   K   H   I   P   I
                65                  70                  75

att att tct tca gcg aga agt gat gtg gaa gat aag att aaa gcg   270
I   I   S   S   A   R   S   D   V   E   D   K   I   K   A
                80                  85                  90

cta gat tat ggg gct gat gat tac ctc cct aaa ccc tat gat cct   315
L   D   Y   G   A   D   D   Y   L   P   K   P   Y   D   P
                95                  100                 105

aag gaa tta tta gct cgt atc caa tcc ttg ctc agg cgt tct cat   360
K   E   L   L   A   R   I   Q   S   L   L   R   R   S   H
                110                 115                 120

aaa aaa gaa gag gtg agt gag cca ggc gat gcg aat atc ttt agg   405
K   K   E   E   V   S   E   P   G   D   A   N   I   F   R
                125                 130                 135
```

```
gta gat aag gat agc cga gaa gtg tat atg cat gaa aaa aag cta   450
V   D   K   D   S   R   E   V   Y   M   H   E   K   K   L
                140                 145                 150

gac tta act agg gct gaa tat gaa atc ctt tcg ctt ctc atc agc   495
D   L   T   R   A   E   Y   E   I   L   S   L   L   I   S
                155                 160                 165

aaa aaa ggt tat gtg ttt agc cgt gaa agc att gcg att gag agc   540
K   K   G   Y   V   F   S   R   E   S   I   A   I   E   S
                170                 175                 180

gag agc atc aac cct gaa agc tct aat aaa agc att gat gtg atc   585
E   S   I   N   P   E   S   S   N   K   S   I   D   V   I
                185                 190                 195

att ggc cgt ttg cga tcc aag att gaa aaa aat cct aaa caa ccg   630
I   G   R   L   R   S   K   I   E   K   N   P   K   Q   P
                200                 205                 210

caa tac atc atc tct gtt aga ggg att ggt tat aaa tta gaa tac   675
Q   Y   I   I   S   V   R   G   I   G   Y   K   L   E   Y
                215                 220                 225
```

ArsR-DBD

```
gaa gag gtg agt gag cca ggc gat gcg aat atc ttt agg gta gat   411
E   E   V   S   E   P   G   D   A   N   I   F   R   V   D
        125                 130                 135

aag gat agc cga gaa gtg tat atg cat gaa aaa aag cta gac tta   456
K   D   S   R   E   V   Y   M   H   E   K   K   L   D   L
        140                 145                 150

act agg gct gaa tat gaa atc ctt tcg ctt ctc atc agc aaa aaa   501
T   R   A   E   Y   E   I   L   S   L   L   I   S   K   K
        155                 160                 165

ggt tat gtg ttt agc cgt gaa agc att gcg att gag agc gag agc   546
G   Y   V   F   S   R   E   S   I   A   I   E   S   E   S
        170                 175                 180

atc aac cct gaa agc tct aat aaa agc att gat gtg atc att ggc   591
I   N   P   E   S   S   N   K   S   I   D   V   I   I   G
        185                 190                 195
```

```
cgt ttg cga tcc aag att gaa aaa aat cct aaa caa ccg caa tac   636
R   L   R   S   K   I   E   K   N   P   K   Q   P   Q   Y
        200                 205                 210

atc atc tct gtt aga ggg att ggt tat aaa tta gaa tac   675
I   I   S   V   R   G   I   G   Y   K   L   E   Y
        215                 220                 225
```

HP1043

```
atg cgc gtt cta ctg att gaa aaa aat tct gtt tta ggt gga gaa   45
M   R   V   L   L   I   E   K   N   S   V   L   G   G   E
                5                   10                  15

att gaa aag ggc tta aat gtt aaa ggc ttt atg gct gat gta aca   90
I   E   K   G   L   N   V   K   G   F   M   A   D   V   T
                20                  25                  30

gag agt tta gag gat ggg gaa tac ctt atg gat att agg aat tat   135
E   S   L   E   D   G   E   Y   L   M   D   I   R   N   Y
                35                  40                  45

gac tta gtt atg gtt agc gat aaa aac gct tta agt ttt gtt tct   180
D   L   V   M   V   S   D   K   N   A   L   S   F   V   S
                50                  55                  60

aga atc aag gag aaa cat tct tct att gtt gtt tta gtt tct tct   225
R   I   K   E   K   H   S   S   I   V   V   L   V   S   S
                65                  70                  75

gat aat cct aca agc gaa gaa gaa gtc cat gcg ttt gag caa ggc   270
D   N   P   T   S   E   E   E   V   H   A   F   E   Q   G
                80                  85                  90

gcg gac gat tat atc gct aag cct tac cgc agc att aaa gct tta   315
A   D   D   Y   I   A   K   P   Y   R   S   I   K   A   L
                95                  100                 105

gtc gca agg att gag gct cgt ttg agg ttt tgg ggt tct aat gtg   360
V   A   R   I   E   A   R   L   R   F   W   G   S   N   V
                110                 115                 120

att gaa att ggg gat ttg acc att agc cct gat gaa gaa aag att   405
I   E   I   G   D   L   T   I   S   P   D   E   E   K   I
                125                 130                 135
```

```
att tac aag ggg cgt gaa gtt gag gtt aaa ggg aag ccc ttt gaa   450
I   Y   K   G   R   E   V   E   V   K   G   K   P   F   E
                140                 145                 150

gta ctg acc cat ctt gcc agg cat agg gat caa atc gtc tcc aaa   495
V   L   T   H   L   A   R   H   R   D   Q   I   V   S   K
                155                 160                 165

gaa cag ctt tta gac gct att tgg gaa gag cct gaa atg gtt acc   540
E   Q   L   L   D   A   I   W   E   E   P   E   M   V   T
                170                 175                 180

cct aat gtg att gaa gtg gct atc aat caa atc cgc caa aaa atg   585
P   N   V   I   E   V   A   I   N   Q   I   R   Q   K   M
                185                 190                 195

gat aaa ccc ttg ggg att tcc acg gtt gaa acc gta agg cgc aga   630
D   K   P   L   G   I   S   T   V   E   T   V   R   R   R
                200                 205                 210

ggc tat cgt ttt tgt tac ccc aaa ccg gcg tgt gaa gaa   669
G   Y   R   F   C   Y   P   K   P   A   C   E   E
                215                 220
```

HP1043-DBD

```
ttt tgg ggt tct aat gtg att gaa att ggg gat ttg acc att agc   387
F   W   G   S   N   V   I   E   I   G   D   L   T   I   S
115                 120                 125

cct gat gaa gaa aag att att tac aag ggg cgt gaa gtt gag gtt   432
P   D   E   E   K   I   I   Y   K   G   R   E   V   E   V
130                 135                 140

aaa ggg aag ccc ttt gaa gta ctg acc cat ctt gcc agg cat agg   477
K   G   K   P   F   E   V   L   T   H   L   A   R   H   R
145                 150                 155

gat caa atc gtc tcc aaa gaa cag ctt tta gac gct att tgg gaa   522
D   Q   I   V   S   K   E   Q   L   L   D   A   I   W   E
160                 165                 170

gag cct gaa atg gtt acc cct aat gtg att gaa gtg gct atc aat   567
E   P   E   M   V   T   P   N   V   I   E   V   A   I   N
175                 180                 185
```

```
caa atc cgc caa aaa atg gat aaa ccc ttg ggg att tcc acg gtt   612
Q   I   R   Q   K   M   D   K   P   L   G   I   S   T   V
190                 195                 200

gaa acc gta agg cgc aga ggc tat cgt ttt   642
E   T   V   R   R   R   G   Y   R   F
205                 210
```

HP1021

```
atg aaa atc tta atc att gaa gat gat tta gca cta gct agg agt   45
M   K   I   L   I   I   E   D   D   L   A   L   A   R   S
                5                   10                  15

atc tct cat aat ttg cat gat tta ggg cat ttt tgc gag atc atc   90
I   S   H   N   L   H   D   L   G   H   F   C   E   I   I
                20                  25                  30

tct agc att tca gag gaa aat aaa gag cct tat gat gtg att tta   135
S   S   I   S   E   E   N   K   E   P   Y   D   V   I   L
                35                  40                  45

gtt tct tct aaa gtt tgc act caa ggg cgt tgc gaa cat ttt gtg   180
V   S   S   K   V   C   T   Q   G   R   C   E   H   F   V
                50                  55                  60

cgt tat aat tcc aag caa atc att atc atg atg gct tcg cat gtc   225
R   Y   N   S   K   Q   I   I   I   M   M   A   S   H   V
                65                  70                  75

aat gaa gat ggt gtg aat aaa ccc att caa gcg gga gcg aga gat   270
N   E   D   G   V   N   K   P   I   Q   A   G   A   R   D
                80                  85                  90

tat att cta aag cct ttt aaa atg gac gaa ttg ttg cgt aag atc   315
Y   I   L   K   P   F   K   M   D   E   L   L   R   K   I
                95                  100                 105

caa tac cac aga gcc tac caa gaa atg acc gct cgc ttg gga ttt   360
Q   Y   H   R   A   Y   Q   E   M   T   A   R   L   G   F
                110                 115                 120
```

```
tat gaa aat tac ttg gac ttt atc cat gcg gaa ttg ccc ttg cct   405
Y   E   N   Y   L   D   F   I   H   A   E   L   P   L   P
                125                 130                 135

aga gat ttt tct tac aga ccg cct ttt atc atc cac gca ccc tct   450
R   D   F   S   Y   R   P   P   F   I   I   H   A   P   S
                140                 145                 150

caa gag ctt gcg aac gct tat tta ttg caa tac gct aaa gaa agg   495
Q   E   L   A   N   A   Y   L   L   Q   Y   A   K   E   R
                155                 160                 165

caa atg gat ttt tct ttt ttc tct tta aag gat acc act tgg aaa   540
Q   M   D   F   S   F   F   S   L   K   D   T   T   W   K
                170                 175                 180

gag cta tac aag aat aaa gac aaa tta gaa cgc cct ttt tac atc   585
E   L   Y   K   N   K   D   K   L   E   R   P   F   Y   I
                185                 190                 195

atg cat tta gaa gag ctt aag aaa gac gag caa ttg aaa ttg cta   630
M   H   L   E   E   L   K   K   D   E   Q   L   K   L   L
                200                 205                 210

gaa ttg gcc cgt tca tgc ccc att gtt ttg tcc tat acc cat aaa   675
E   L   A   R   S   C   P   I   V   L   S   Y   T   H   K
                215                 220                 225

gaa cca tta gaa ttt cct aaa att atg agc att gaa tgc ggc aat   720
E   P   L   E   F   P   K   I   M   S   I   E   C   G   N
                230                 235                 240

aga ccc cta tct ttg ttt agc gat aac acg act ttc ctt tcc att   765
R   P   L   S   L   F   S   D   N   T   T   F   L   S   I
                245                 250                 255

caa gag tat gaa aaa gaa gcc att agg cat ttt tct tct act tgc   810
Q   E   Y   E   K   E   A   I   R   H   F   S   S   T   C
                260                 265                 270

acg gac aca gaa tta gcc agc aag ctt ggc att agc cgt aaa agc   855
T   D   T   E   L   A   S   K   L   G   I   S   R   K   S
                275                 280                 285
```

```
ctt tgg gaa aaa cgc cgg aaa tat aac tta ccg cgc aag   894
  L   W   E   K   R   R   K   Y   N   L   P   R   K
                  290                 295
```

HP1021-DBD

```
  ttg ccc ttg cct aga gat ttt tct tac aga ccg cct ttt atc atc   438
  L   P   L   P   R   D   F   S   Y   R   P   P   F   I   I
              135                 140                 145

  cac gca ccc tct caa gag ctt gcg aac gct tat tta ttg caa tac   483
  H   A   P   S   Q   E   L   A   N   A   Y   L   L   Q   Y
              150                 155                 160

  gct aaa gaa agg caa atg gat ttt tct ttt ttc tct tta aag gat   528
  A   K   E   R   Q   M   D   F   S   F   F   S   L   K   D
              165                 170                 175

  acc act tgg aaa gag cta tac aag aat aaa gac aaa tta gaa cgc   573
  T   T   W   K   E   L   Y   K   N   K   D   K   L   E   R
              180                 185                 190

  cct ttt tac atc atg cat tta gaa gag ctt aag aaa gac gag caa   618
  P   F   Y   I   M   H   L   E   E   L   K   K   D   E   Q
              195                 200                 205

  ttg aaa ttg cta gaa ttg gcc cgt tca tgc ccc att gtt ttg tcc   663
  L   K   L   L   E   L   A   R   S   C   P   I   V   L   S
              210                 215                 220

  tat acc cat aaa gaa cca tta gaa ttt cct aaa att atg agc att   708
  Y   T   H   K   E   P   L   E   F   P   K   I   M   S   I
              225                 230                 235

  gaa tgc ggc aat aga ccc cta tct ttg ttt agc gat aac acg act   753
  E   C   G   N   R   P   L   S   L   F   S   D   N   T   T
              240                 245                 250

  ttc ctt tcc att caa gag tat gaa aaa gaa gcc att agg cat ttt   798
  F   L   S   I   Q   E   Y   E   K   E   A   I   R   H   F
              255                 260                 265

  tct tct act tgc acg gac aca gaa tta gcc agc aag ctt ggc att   843
  S   S   T   C   T   D   T   E   L   A   S   K   L   G   I
              270                 275                 280
```

```
agc cgt aaa agc ctt tgg gaa aaa cgc cgg aaa tat aac tta ccg   888
S   R   K   S   L   W   E   K   R   R   K   Y   N   L   P
            285                 290                 295

cgc aag   894
R   K
```

FlgR-DBD

```
acg atc ttt tta gat gaa atc gct gaa atg ccc ctt caa ctg caa   735
T   I   F   L   D   E   I   A   E   M   P   L   Q   L   Q
                235                 240                 245

agc aaa ctt tta aga gtg gtt caa gaa aaa gaa atc acg cgc ctt   780
S   K   L   L   R   V   V   Q   E   K   E   I   T   R   L
                250                 255                 260

ggg gat aat aaa agc gtt aaa att gat gtt cgt ttc att tcc gcc   825
G   D   N   K   S   V   K   I   D   V   R   F   I   S   A
                265                 270                 275

acc aac gcc aac atg aaa gaa aaa atc gct tca aaa gaa ttc aga   870
T   N   A   N   M   K   E   K   I   A   S   K   E   F   R
                280                 285                 290

gaa gat ttg ttt ttc cgc ttg caa atc gtg cct ata act atc gcg   915
E   D   L   F   F   R   L   Q   I   V   P   I   T   I   A
                295                 300                 305

ccc tta aga gag agg gta gaa gag att ttg ccc att gct gaa atc   960
P   L   R   E   R   V   E   E   I   L   P   I   A   E   I
                310                 315                 320

aag ctt aaa gaa gtg tgc gag gcg tat cat ttg ggg cca aaa tct   1005
K   L   K   E   V   C   E   A   Y   H   L   G   P   K   S
                325                 330                 335

ttt tca aaa aac gcc gca aaa cgc ctt tta gaa tac tct tgg cat   1050
F   S   K   N   A   A   K   R   L   L   E   Y   S   W   H
                340                 345                 350

ggg aat gtg cga gag ctt tta ggc gtc gtg gaa aga gcg gcg att   1095
G   N   V   R   E   L   L   G   V   V   E   R   A   A   I
                355                 360                 365
```

```
tta agc gaa gaa aca gaa atc caa gag aaa gat ttg ttt ttg gaa   1140
L   S   E   E   T   E   I   Q   E   K   D   L   F   L   E
                370                 375                 380

agg   1143
R
```

CrdR

```
atg caa aaa aag att ttt tta cta gaa gac gat tac ctt tta agc   45
M   Q   K   K   I   F   L   L   E   D   D   Y   L   L   S
                5                   10                  15

gag agt atc aag gag ttc ttg gag cat tta ggc tat gaa gtg ttt   90
E   S   I   K   E   F   L   E   H   L   G   Y   E   V   F
                20                  25                  30

tgc gct ttt aac ggg aaa gag gct cat gaa agg ctc tct gtt gag   135
C   A   F   N   G   K   E   A   H   E   R   L   S   V   E
                35                  40                  45

cgc ttt aac ctc ttg ctt tta gac gtg caa gtg cct gaa atg aat   180
R   F   N   L   L   L   L   D   V   Q   V   P   E   M   N
                50                  55                  60

agc ttg gaa tta ttc aag cgc atc aaa aac gat ttt tta atc tct   225
S   L   E   L   F   K   R   I   K   N   D   F   L   I   S
                65                  70                  75

acg cct gtg att ttt atc acc gcc tta cag gat aac gct acc tta   270
T   P   V   I   F   I   T   A   L   Q   D   N   A   T   L
                80                  85                  90

aaa aac gct ttt aat tta ggg gcg agc gat tat ttg aaa aag cct   315
K   N   A   F   N   L   G   A   S   D   Y   L   K   K   P
                95                  100                 105

ttt gat ttg gac gaa ttg gaa gcg cgc att aaa agg ttt ttc aat   360
F   D   L   D   E   L   E   A   R   I   K   R   F   F   N
                110                 115                 120
```

```
gat gat ccg ata gaa atc atg cct aac att ttt tac cac caa cac   405
 D   D   P   I   E   I   M   P   N   I   F   Y   H   Q   H
                 125                 130                 135

 gct ttg aac gtt aaa ggg aaa aag gaa atc tta gcg ccc aaa acc   450
 A   L   N   V   K   G   K   K   E   I   L   A   P   K   T
                 140                 145                 150

 gcc caa ctt tta gaa tat ttt tta gag cat aaa ggg caa atc atc   495
 A   Q   L   L   E   Y   F   L   E   H   K   G   Q   I   I
                 155                 160                 165

 agc tct caa gcg tta gaa aac aac tta tgg gag caa gcg att gat   540
 S   S   Q   A   L   E   N   N   L   W   E   Q   A   I   D
                 170                 175                 180

 gat tcc acc tta cgc act tac att aaa gtg ttg cgc aag ctt ttg   585
 D   S   T   L   R   T   Y   I   K   V   L   R   K   L   L
                 185                 190                 195

 ggt aaa aat tgc ata gaa acg cat aag ggg gtg ggc tat cgc ttt   630
 G   K   N   C   I   E   T   H   K   G   V   G   Y   R   F
                 200                 205                 210

 aac cca cta   639
 N   P   L
```

CrdR-DBD

```
 gat gat ccg ata gaa atc atg cct aac att ttt tac cac caa cac   405
 D   D   P   I   E   I   M   P   N   I   F   Y   H   Q   H
                 125                 130                 135

 gct ttg aac gtt aaa ggg aaa aag gaa atc tta gcg ccc aaa acc   450
 A   L   N   V   K   G   K   K   E   I   L   A   P   K   T
                 140                 145                 150

 gcc caa ctt tta gaa tat ttt tta gag cat aaa ggg caa atc atc   495
 A   Q   L   L   E   Y   F   L   E   H   K   G   Q   I   I
                 155                 160                 165

 agc tct caa gcg tta gaa aac aac tta tgg gag caa gcg att gat   540
 S   S   Q   A   L   E   N   N   L   W   E   Q   A   I   D
                 170                 175                 180
```

```
gat tcc acc tta cgc act tac att aaa gtg ttg cgc aag ctt ttg   585
D   S   T   L   R   T   Y   I   K   V   L   R   K   L   L
                185                 190                 195

ggt aaa aat tgc ata gaa acg cat aag ggg gtg ggc tat cgc ttt   630
G   K   N   C   I   E   T   H   K   G   V   G   Y   R   F
                200                 205                 210

aac cca cta   639
N   P   L
```

BIBLIOGRAPHY

1. Cave, D. R. (1996) Transmission and epidemiology of *Helicobacter pylori*, *Am J Med 100*, 12S-17S; discussion 17S-18S.

2. Cover, T. L., and Blaser, M. J. (2009) *Helicobacter pylori* in Health and Disease, *Gastroenterology 136*, 1863-1873.

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

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

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

6. Monack, D. M., Mueller, A., and Falkow, S. (2004) Persistent bacterial infections: the interface of the pathogen and the host immune system, *Nat Rev Microbiol 2*, 747-765.

7. Mueller, A., Falkow, S., and Amieva, M. R. (2005) *Helicobacter pylori* and gastric cancer: what can be learned by studying the response of gastric epithelial cells to the infection?, *Cancer Epidemiol Biomarkers Prev 14*, 1859-1864.

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

9. Peterson, W. L. (1991) *Helicobacter pylori* and peptic ulcer disease, *N Engl J Med 324*, 1043-1048.

10. Suermaum, S., and Michetti, P. (2002) *Helicobacter pylori* infection, *New England Journal of Medicine 347*, 1175-1186.

11. Crowe, S. E. (2005) Helicobacter infection, chronic inflammation, and the development of malignancy, *Current Opinion in Gastroenterology 21*, 32-38.

12. Moss, S. F., and Blaser, M. J. (2005) Mechanisms of disease: Inflammation and the origins of cancer, *Nat Clin Pract Oncol 2*, 90-97; quiz 91 p following 113.

13. Peek, R. M., Jr., and Blaser, M. J. (2002) *Helicobacter pylori* and gastrointestinal tract adenocarcinomas, *Nat Rev Cancer 2*, 28-37.

14. Suerbaum, S., and Josenhans, C. (1999) Virulence factors of *Helicobacter pylori*: implications for vaccine development, *Mol Med Today 5*, 32-39.

15. Algood, H. M., and Cover, T. L. (2006) *Helicobacter pylori* persistence: an overview of interactions between H. pylori and host immune defenses, *Clin Microbiol Rev 19*, 597-613.

16. Dick, J. D. (1990) Helicobacter (Campylobacter) pylori: a new twist to an old disease, *Annu Rev Microbiol 44*, 249-269.

17. Ang, S., Lee, C. Z., Peck, K., Sindici, M., Matrubutham, U., Gleeson, M. A., and Wang, J. T. (2001) Acid-induced gene expression in *Helicobacter pylori*: study in genomic scale by microarray, *Infect Immun 69*, 1679-1686.

18. Audia, J. P., Webb, C. C., and Foster, J. W. (2001) Breaking through the acid barrier: an orchestrated response to proton stress by enteric bacteria, *Int J Med Microbiol 291*, 97-106.

19. Bury-Mone, S., Thiberge, J. M., Contreras, M., Maitournam, A., Labigne, A., and De Reuse, H. (2004) Responsiveness to acidity via metal ion regulators mediates virulence in the gastric pathogen *Helicobacter pylori*, *Mol Microbiol 53*, 623-638.

20. Merrell, D. S., Goodrich, M. L., Otto, G., Tompkins, L. S., and Falkow, S. (2003) pH-regulated gene expression of the gastric pathogen *Helicobacter pylori*, *Infect Immun 71*, 3529-3539.

21. Wen, Y., Marcus, E. A., Matrubutham, U., Gleeson, M. A., Scott, D. R., and Sachs, G. (2003) Acid-adaptive genes of *Helicobacter pylori*, *Infect Immun 71*, 5921-5939.

22. Pflock, M., Kennard, S., Finsterer, N., and Beier, D. (2006) Acid-responsive gene regulation in the human pathogen *Helicobacter pylori*, *J Biotechnol 126*, 52-60.

23. Marcus, E. A., Moshfegh, A. P., Sachs, G., and Scott, D. R. (2005) The Periplasmic α-Carbonic Anhydrase Activity of *Helicobacter pylori* is Essential for Acid Acclimation, *Journal of Bacteriology 187*, 729-738.

24. Pflock, M., Kennard, S., Delany, I., Scarlato, V., and Beier, D. (2005) Acid-Induced Activation of the Urease Promoters Is Mediated Directly by the ArsRS Two-Component System of *Helicobacter pylori*, *Infect Immun 73*, 6437-6445.

25. Scott, D. R., Marcus, E. A., Wen, Y., Oh, J., and Sachs, G. (2007) Gene expression in vivo shows that *Helicobacter pylori* colonizes an acidic niche on the gastric surface, *Proc Natl Acad Sci U S A 104*, 7235-7240.

26. Wen, Y., Feng, J., Scott, D. R., Marcus, E. A., and Sachs, G. (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*, 2426-2434.

27. Marais, A., Mendz, G. L., Hazell, S. L., and Megraud, F. (1999) Metabolism and genetics of *Helicobacter pylori*: the genome era, *Microbiol Mol Biol Rev 63*, 642-674.

28. Stingl, K., Altendorf, K., and Bakker, E. P. (2002) Acid survival of *Helicobacter pylori*: how does urease activity trigger cytoplasmic pH homeostasis?, *Trends Microbiol 10*, 70-74.

29. Loh, J. T., and Cover, T. L. (2006) Requirement of histidine kinases HP0165 and HP1364 for acid resistance in *Helicobacter pylori*, *Infect Immun 74*, 3052-3059.

30. Wen, Y., Feng, J., Scott, D. R., Marcus, E. A., and Sachs, G. (2009) The pH-responsive regulon of HP0244 (FlgS), the cytoplasmic histidine kinase of *Helicobacter pylori*, *J Bacteriol 191*, 449-460.

31. Hall, M. N., and Silhavy, T. J. (1981) The ompB locus and the regulation of the major outer membrane porin proteins of Escherichia coli K12, *J Mol Biol 146*, 23-43.

32. Ulrich, L. E., Koonin, E. V., and Zhulin, I. B. (2005) One-component systems dominate signal transduction in prokaryotes, *Trends Microbiol 13*, 52-56.

33. Buckler, D. R., Zhou, Y., and Stock, A. M. (2002) Evidence of intradomain and interdomain flexibility in an OmpR/PhoB homolog from Thermotoga maritima, *Structure 10*, 153-164.

34. Stock, A. M., Robinson, V. L., and Goudreau, P. N. (2000) Two-component signal transduction, *Annu Rev Biochem 69*, 183-215.

35. Mizuno, T. (1997) Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of Escherichia coli, *DNA Res 4*, 161-168.

36. Alm, R. A., Ling, L. S., Moir, D. T., King, B. L., Brown, E. D., Doig, P. C., Smith, D. R., Noonan, B., Guild, B. C., deJonge, B. L., Carmel, G., Tummino, P. J., Caruso, A., Uria-Nickelsen, M., Mills, D. M., Ives, C., Gibson, R., Merberg, D., Mills, S. D., Jiang, Q., Taylor, D. E., Vovis, G. F., and Trust, T. J. (1999) Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*, *Nature 397*, 176-180.

37. Beier, D., and Frank, R. (2000) Molecular characterization of two-component systems of *Helicobacter pylori*, *J Bacteriol 182*, 2068-2076.

38. Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E. F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H. G., Glodek, A., McKenney, K., Fitzegerald, L. M., Lee, N., Adams, M. D., Hickey, E. K., Berg, D. E., Gocayne, J. D., Utterback, T. R., Peterson, J. D., Kelley, J. M., Cotton, M. D., Weidman, J. M., Fujii, C., Bowman, C., Watthey, L., Wallin, E., Hayes, W. S., Borodovsky, M., Karp, P. D., Smith, H. O., Fraser, C. M., and Venter, J. C. (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori*, *Nature 388*, 539-547.

39. Eguchi, Y., and Utsumi, R. (2005) A novel mechanism for connecting bacterial two-component signal-transduction systems, *Trends Biochem Sci 30*, 70-72.

40. Schar, J., Sickmann, A., and Beier, D. (2005) Phosphorylation-independent activity of atypical response regulators of *Helicobacter pylori*, *J Bacteriol 187*, 3100-3109.

41. Hong, E., Lee, H. M., Ko, H., Kim, D. U., Jeon, B. Y., Jung, J., Shin, J., Lee, S. A., Kim, Y., Jeon, Y. H., Cheong, C., Cho, H. S., and Lee, W. (2007) Structure of an atypical orphan response regulator protein supports a new phosphorylation-independent regulatory mechanism, *J Biol Chem 282*, 20667-20675.

42. Pflock, M., Finsterer, N., Joseph, B., Mollenkopf, H., Meyer, T. F., and Beier, D. (2006) Characterization of the ArsRS regulon of *Helicobacter pylori*, involved in acid adaptation, *J Bacteriol 188*, 3449-3462.

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

44. Dietz, P., Gerlach, G., and Beier, D. (2002) Identification of target genes regulated by the two-component system HP166-HP165 of *Helicobacter pylori*, *J Bacteriol 184*, 350-362.

45. Panthel, K., Dietz, P., Haas, R., and Beier, D. (2003) Two-component systems of *Helicobacter pylori* contribute to virulence in a mouse infection model, *Infect Immun 71*, 5381-5385.

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

47. Wen, Y., Feng, J., Scott, D. R., Marcus, E. A., and Sachs, G. (2006) Involvement of the HP0165-HP0166 two-component system in expression of some acidic-pH-upregulated genes of *Helicobacter pylori*, *J Bacteriol 188*, 1750-1761.

48. Forsyth, M. H., Cao, P., Garcia, P. P., Hall, J. D., and Cover, T. L. (2002) Genome-wide transcriptional profiling in a histidine kinase mutant of *Helicobacter pylori* identifies members of a regulon, *J Bacteriol 184*, 4630-4635.

49. Fiedler, U., and Weiss, V. (1995) A common switch in activation of the response regulators NtrC and PhoB: phosphorylation induces dimerization of the receiver modules, *Embo J 14*, 3696-3705.

50. McCleary, W. R. (1996) The activation of PhoB by acetylphosphate, *Mol Microbiol 20*, 1155-1163.

51. Aiba, H., Nakasai, F., Mizushima, S., and Mizuno, T. (1989) Phosphorylation of a bacterial activator protein, OmpR, by a protein kinase, EnvZ, results in stimulation of its DNA-binding ability, *J Biochem (Tokyo) 106*, 5-7.

52. Jo, Y. L., Nara, F., Ichihara, S., Mizuno, T., and Mizushima, S. (1986) Purification and characterization of the OmpR protein, a positive regulator involved in osmoregulatory expression of the ompF and ompC genes in Escherichia coli, *J Biol Chem 261*, 15252-15256.

53. Blanco, A. G., Sola, M., Gomis-Ruth, F. X., and Coll, M. (2002) Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator, *Structure 10*, 701-713.

54. Rhee, J. E., Sheng, W., Morgan, L. K., Nolet, R., Liao, X., and Kenney, L. J. (2008) Amino acids important for DNA recognition by the response regulator OmpR, *J Biol Chem 283*, 8664-8677.

55. Sattler, M., Schleucher, J., and Griesinger, C. (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.

56. Wishart, D. S., and Sykes, B. D. (1994) Chemical shifts as a tool for structure determination, *Methods Enzymol 239*, 363-392.

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

58. Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA, *J Mol Biol 273*, 283-298.

59. D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D.H. Mathews, C. Schafmeister, W.S. Ross, and P.A. Kollman. (2004) Amber 9.

60. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *J Comput Chem 25*, 1605-1612.

61. Koradi, R., Billeter, M., and Wuthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures, *J Mol Graph 14*, 51-55, 29-32.

62. Rocchia, W., Alexov, E., Honig, B. (2001) Extending the applicability of the nonlinear Poisson-Boltzmann equation: Multiple dielectric constants and multivalent ions, *J Phys. Chem. B*, 6507-6514.

63. Rocchia, W., Sridharan, S., Nicholls, A., Alexov, E., Chiabrera, A., Honig, B. (2002) Rapid Grid-based Construction of the Molecular Surface for both Molecules and Geometric Objects: Applications to the Fininte Difference Poisson-Boltzmann Method, *J Comput Chem 23*, 128-137.

64. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, *J Biomol NMR 8*, 477-486.

65. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting,

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

66. Martinez-Hackert, E., and Stock, A. M. (1997) The DNA-binding domain of OmpR: crystal structures of a winged helix transcription factor, *Structure 5*, 109-124.

67. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. L. (2008) GenBank, *Nucleic Acids Res 36*, D25-30.

68. Holm, L., Kaariainen, S., Rosenstrom, P., Schenkel, A. . (2008) Searching protein structure databases with DaliLite v.3., *Bioinformatics 24*, 2780-2781.

69. Martinez-Hackert, E., and Stock, A. M. (1997) Structural relationships in the OmpR family of winged-helix transcription factors, *J Mol Biol 269*, 301-312.

70. Robinson, V. L., Wu, T., and Stock, A. M. (2003) Structural analysis of the domain interface in DrrB, a response regulator of the OmpR/PhoB subfamily, *J Bacteriol 185*, 4186-4194.

71. Eldridge, A. M., Kang, H. S., Johnson, E., Gunsalus, R., and Dahlquist, F. W. (2002) Effect of phosphorylation on the interdomain interaction of the response regulator, NarL, *Biochemistry 41*, 15173-15180.

72. Zhang, J. H., Xiao, G., Gunsalus, R. P., and Hubbell, W. L. (2003) Phosphorylation triggers domain separation in the DNA binding response regulator NarL, *Biochemistry 42*, 2552-2559.

73. Lilley, K. S., and Friedman, D. B. (2004) All about DIGE: quantification technology for differential-display 2D-gel proteomics, *Expert Rev Proteomics 1*, 401-409.

74. McGee, D. J., Kumar, S., Viator, R. J., Bolland, J. R., Ruiz, J., Spadafora, D., Testerman, T. L., Kelly, D. J., Pannell, L. K., and Windle, H. J. (2006) *Helicobacter pylori* thioredoxin is an arginase chaperone and guardian against oxidative and nitrosative stresses, *J Biol Chem 281*, 3290-3296.

75. Brahmachary, P., Wang, G., Benoit, S. L., Weinberg, M. V., Maier, R. J., and Hoover, T. R. (2008) The human gastric pathogen *Helicobacter pylori* has a potential acetone carboxylase that enhances its ability to colonize mice, *BMC Microbiol 8*, 14.

76. Pflock, M., Bathon, M., Schar, J., Muller, S., Mollenkopf, H., Meyer, T. F., and Beier, D. (2007) The orphan response regulator HP1021 of *Helicobacter pylori* regulates transcription of a gene cluster presumably involved in acetone metabolism, *J Bacteriol 189*, 2339-2349.

77. Fisher, S. L., Jiang, W., Wanner, B. L., and Walsh, C. T. (1995) Cross-talk between the histidine protein kinase VanS and the response regulator PhoB. Characterization and identification of a VanS domain that inhibits activation of PhoB, *J Biol Chem 270*, 23143-23149.

78. Laub, M. T., and Goulian, M. (2007) Specificity in two-component signal transduction pathways, *Annu Rev Genet 41*, 121-145.

79. Jubelin, G., Vianney, A., Beloin, C., Ghigo, J. M., Lazzaroni, J. C., Lejeune, P., and Dorel, C. (2005) CpxR/OmpR interplay regulates curli gene expression in response to osmolarity in Escherichia coli, *J Bacteriol 187*, 2038-2049.

80. Vendeville, A., Winzer, K., Heurlier, K., Tang, C. M., and Hardie, K. R. (2005) Making 'sense' of metabolism: autoinducer-2, LuxS and pathogenic bacteria, *Nat Rev Microbiol 3*, 383-396.

81. Ellison, D. W., and McCleary, W. R. (2000) The unphosphorylated receiver domain of PhoB silences the activity of its output domain, *J Bacteriol 182*, 6592-6597.

82. Tapparel, C., Monod, A., and Kelley, W. L. (2006) The DNA-binding domain of the Escherichia coli CpxR two-component response regulator is constitutively active and cannot be fully attenuated by fused adjacent heterologous regulatory domains, *Microbiology 152*, 431-441.

83. Beier, D., and Gross, R. (2006) Regulation of bacterial virulence by two-component systems, *Curr Opin Microbiol 9*, 143-152.

84. Bogel, G., Schrempf, H., and Ortiz de Orue Lucana, D. (2007) DNA-binding characteristics of the regulator SenR in response to phosphorylation by the sensor histidine autokinase SenS from Streptomyces reticuli, *Febs J 274*, 3900-3913.

85. Schaaf, S., and Bott, M. (2007) Target genes and DNA-binding sites of the response regulator PhoR from Corynebacterium glutamicum, *J Bacteriol 189*, 5002-5011.

86. Boucher, P. E., Murakami, K., Ishihama, A., and Stibitz, S. (1997) Nature of DNA binding and RNA polymerase interaction of the Bordetella pertussis BvgA transcriptional activator at the fha promoter, *J Bacteriol 179*, 1755-1763.

87. Dosanjh, N. S., West, A. L., and Michel, S. L. (2009) *Helicobacter pylori* NikR's interaction with DNA: a two-tiered mode of recognition, *Biochemistry 48*, 527-536.

88. Katti, M. V., Sakharkar, M. K., Ranjekar, P. K., and Gupta, V. S. (2000) TRES: comparative promoter sequence analysis, *Bioinformatics (Oxford, England) 16*, 739-740.

89. Liu, X., Brutlag, D. L., and Liu, J. S. (2001) BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes, *Pacific Symposium on Biocomputing*, 127-138.

90. Rabenstein, M. D., and Shin, Y. K. (1995) Determination of the distance between two spin labels attached to a macromolecule, *Proc Natl Acad Sci U S A 92*, 8239-8243.

91. Wemmer, D. E., and Kern, D. (2005) Beryllofluoride binding mimics phosphorylation of aspartate in response regulators, *J Bacteriol 187*, 8229-8230.

92. Hubbell, W. L., McHaourab, H. S., Altenbach, C., and Lietzow, M. A. (1996) Watching proteins move using site-directed spin labeling, *Structure 4*, 779-783.

93. Bachhawat, P., Swapna, G. V., Montelione, G. T., and Stock, A. M. (2005) Mechanism of activation for transcription factor PhoB suggested by different modes of dimerization in the inactive and active states, *Structure 13*, 1353-1363.