

ABSTRACT

Title of Document: LOCAL AND GLOBAL GENE REGULATION
ANALYSIS OF THE AUTOINDUCER-2
MEDIATED QUORUM SENSING
MECHANISM IN *ESCHERICHIA COLI*

Christopher Matthew Byrd,
Doctor of Philosophy, 2011

Directed By: Dr. William E. Bentley, Professor and Chair,
Department of Bioengineering

The term ‘quorum sensing’ (QS) is used to define a population density based communication mechanism which uses chemical signal molecules called autoinducers to trigger unique and varied changes in gene expression. Although several communication methods have been identified in bacteria that are unique to a particular species, one type of signal molecule, autoinducer-2 (AI-2) is linked to interspecies communication, indicating its potential as a universal signal for cueing a QS response among multiple bacterial types. In *E. coli*, AI-2 acts as an effector by binding to the QS repressor LsrR. As a result, LsrR unbinds and relieves repression of the *lsr* regulon, stimulating a subsequent QS gene expression cascade.

In this dissertation, LsrR structure and *in vitro* binding activity are examined. Genomic binding and DNA microarray analyses are conducted and three novel sites putatively regulated by LsrR, *yegE-udk*, *mppA* and *yihF*, are revealed. Two cAMP

receptor protein (CRP) binding locations in intergenic region of the *lsr* regulon are also confirmed. The role of each CRP site in divergent expression is qualified, indicating the *lsr* intergenic region to be a class III CRP-dependent promoter. Also, four specific DNA binding sites for LsrR in the *lsr* intergenic region are proposed, and reliance upon simultaneous binding to these various sites and the resulting effects on LsrR repression is presented. Finally, a complex model for regulation of the *lsr* regulon is depicted incorporating LsrR, CRP, DNA looping, and a predicted secondary layer of repression by an integration host factor (IHF)-like protein. Further understanding of this QS genetic mechanism may potentially be used for inhibiting bacterial proliferation and infection, modifying the natural genetic system to elicit alternate desired responses, or extracted and applied to a highly customizable and sensitive *in vitro* biosensor.

LOCAL AND GLOBAL GENE REGULATION ANALYSIS OF THE
AUTOINDUCER-2 MEDIATED QUORUM SENSING MECHANISM IN
ESCHERICHIA COLI

By

Christopher Matthew Byrd

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2011

Advisory Committee:
Professor William E. Bentley, Chair
Professor Douglas Julin
Professor Peter Kofinas
Professor Gregory Payne
Dr. James Sumner

© Copyright by
Christopher Matthew Byrd
2011

Dedication

To my family, who have made great sacrifices for me and for whom I have the utmost respect and love, and to my wife, Monica. Because of her I am a far better person.

Acknowledgements

I would like to acknowledge the people and organizations without which I would not have had the opportunity to pursue a doctorate. Thanks to the United States Army and the founders of Uniformed Army Scientist and Engineer Program, who saw the potential for having Army officers who are also capable scientists.

I would like to thank all the members of my dissertation committee for their time and assistance throughout the PhD process. I would particularly like to acknowledge my advisor, Dr. Bentley, who has been a true mentor, and for being gracious and open-minded enough to give me the opportunity to be part of the PhD program in the Fischell Department of Bioengineering despite the unusual circumstances and requests with which I first approached him.

I have much gratitude for the time, effort, and accommodation shown by scientists and administrators at the Army Research Laboratory who helped fund, support, and facilitate my work in a world-class research facility. Specifically, I would like to acknowledge Dr. James Sumner, who supported me and my program from quite literally day one, and Dr. Matthew Servinsky and Dr. Christian Sund with whom I collaborated on a daily basis and always made time for insight, guidance and discussion.

I want to acknowledge the students in the Bentley lab, past and present, who have been a joy to know and an incredible source of learning and inspiration. Finally, I recognize the often unheralded but always unrelenting efforts of other scientists and fellow soldiers. I am honored and humbled to be considered among your ranks.

Table of Contents

Dedication	ii
Acknowledgements	iii
Table of Contents	iii
List of Tables	v
List of Figures	vi
List of Figures	vi
List of Abbreviations	vii
Chapter 1 : Communication in Bacteria.....	1
1.1 Introduction.....	1
1.2 Background: Bacterial Communication.....	5
1.3 Research Motivation	9
1.4 Literature Review of LsrR	10
1.5 Global Objective, Global Hypothesis, and Specific Aims.....	12
1.6 Dissertation Outline	12
Chapter 2 : Expression, <i>In Vitro</i> Activity, and Structure Analysis of LsrR.....	14
2.1 Abstract.....	14
2.2 Introduction.....	15
2.3 Materials and Methods.....	17
2.4 Results.....	20
2.5 Discussion.....	27
Chapter 3 : Genomic binding site analysis of the AI-2 quorum sensing regulator LsrR in <i>Escherichia coli</i>	31
3.1 Abstract.....	31
3.2 Introduction.....	32
3.3 Materials and Methods.....	34
3.4 Results.....	42
3.5 Discussion.....	56
3.6 Supplemental Data	62
Chapter 4 : Expression Regulation in the Divergent Quorum Sensing <i>lsr</i> Regulon in <i>Escherichia coli</i>	66
4.1 Abstract.....	66
4.2 Introduction.....	67
4.3 Materials and Methods.....	69
4.4 Results.....	71
4.5 Discussion.....	80
4.6 Supplemental Data	91
Chapter 5 : Summary	100
5.1 Summary of Results.....	100
5.2 Future Directions	101
References.....	104

List of Tables

Table 2-1. <i>E. coli</i> strains, plasmids and primers used in this study.	18
Table 3-1. <i>E. coli</i> strains, plasmids and oligonucleotides used in this study.	36
Table 4-1. <i>E. coli</i> strains and plasmids used in this study.	70
Table 4-2. Bacterial strains used in this study.	92

List of Figures

Figure 1-1. Intraspecies communication molecules.....	6
Figure 1-2. Synthesis of autoinducer-2 from S-adenosylmethionine (SAM).....	7
Figure 1-3. Autoinducer-based quorum sensing mechanism in <i>Escherichia coli</i>	8
Figure 2-1. EMSA analysis of CRP.....	21
Figure 2-2. LsrR purification analysis.	24
Figure 2-3. EMSA with LsrR and CRP.	25
Figure 3-1. ChIP-Chip process.	1
Figure 3-2. Purification and computer modeling analysis of LsrR in <i>E. coli</i>	43
Figure 3-3. ChIP-Chip in <i>E. coli</i> K12MG1655.....	48
Figure 3-4. Promoter region orientation and directional expression analysis of the <i>E. coli</i> <i>lsrA-lsrR</i> intergenic region.	53
Figure 3-5. Chromatin shearing.	63
Figure 3-6. Whole Genome Amplification.	65
Figure 4-1. Schematic of <i>lsr</i> intergenic region.	72
Figure 4-2. Effects of site mutations on <i>lsr</i> divergent gene expression.....	75
Figure 4-3. Schematic model of CRP-RNAP complex in the <i>lsr</i> regulon.	82
Figure 4-4. Proposed model for LsrR repression.....	87
Figure 4-5. L-ascorbate utilization in <i>E. coli</i>	1
Figure 4-6. Cross-species <i>lsr</i> intergenic sequence alignment.....	93
Figure 4-7. Preliminary structural comparison of DeoR, UlaR, and LsrR.	97

List of Abbreviations

AB = autoinducer bioassay
AHL = acylated homoserine lactone
AI-1 = autoinducer-1
AI-2 = autoinducer-2
ChIP-Chip = chromatin immunoprecipitation in conjunction with microarrays
CRP = cyclic-AMP receptor protein
 α CTD = RNA polymerase holoenzyme α subunit C-terminal domain
DPD = 4,5-dihydroxy-2,3-pentanedione
EMSA = electrophoretic mobility shift assay
HTH = helix turn helix
IB = inclusion body
IHF = integration host factor
IPTG = isopropyl β -D-thiogalactopyranoside (IPTG)
LB = Luria Bertani
Lsr = LuxS responsive
 α NTD = RNA polymerase holoenzyme α subunit N-terminal domain
OD = optical density
ONPG = o-nitrophenyl- β -D-galactopyranoside
QS = quorum sensing
RNAP = RNA polymerase holoenzyme
SAH = S-(5'-deoxyadenosin-5')-L-homocysteine
SAM = S-adenosylmethionine
SRH = S-ribosylhomocysteine
SDS-PAGE – sodium dodecyl sulfate-polyacrylamide gel electrophoresis
WGA = whole genome amplification
WT = wild type

Chapter 1: Communication in Bacteria

1.1 Introduction

The bacterial signal transduction process through which gene expression may be directly or indirectly affected has gained increasing focus since the mid-1980s when previously unrelated proteins were shown to be expressed during a sophisticated subsequent genetic response [1, 2]. These complex reactions in what were historically considered 'simple' prokaryotic organisms in a wide variety of environmental conditions range from nearly undetectable to drastic [3]. For example, this response cascade permits the marine bacteria, *Vibrio fischeri*, which can inhabit light organs of some marine animals, to bioluminate; however, this effect only triggers after the bacterial cell density reaches 10^{10} - 10^{11} CFU/ml [4, 5].

The term quorum sensing (QS) was first used by Fuqua *et al.* [5] to define a population density based communication mechanism which uses chemical signal molecules called autoinducers to trigger unique and varied changes in gene expression. Various communication methods have since been identified which employ a variety of autoinducers to effect such changes within species. Furthermore, autoinducer-2 (AI-2) has been linked to interspecies communication, indicating its potential as a 'universal' signal for QS circuitry among multiple bacterial species or, potentially, between prokaryotes and eukaryotes [6-8]. This universality suggests not only the prospect for adjusting individual parts to achieve or prevent specific

outcomes *in vivo*, but also for extracting desired components of a versatile system and using them in their native state, or potentially applying them for *in vitro* use [9].

In the health industry, antibiotic-resistant bacteria continue to emerge at alarming rates and all aspects of the infection process are being re-examined so that new procedures for prevention, diagnosis, and treatment of bacterial infections can be developed that reduce their occurrence and severity as well as the economic impact on health care systems. One of the most problematic pathogens is methicillin-resistant *Staphylococcus aureus* (MRSA). First reported among intravenous drug users in Detroit, MI (USA) in 1981 and later associated with the deaths of four children in Minnesota in 1997, MRSA was reported to be the cause of a staggering 85% of health care-associated infections in 2007 as well as the most common origin of skin and tissue infections presented to hospital emergency rooms in the United States [10]. These statistics illustrate the scale and impact that bacteria can have on a population and our health care system, and the importance of understanding the QS process through which pathogens such as MRSA may proliferate [9].

Contrary to the once common perception that bacteria operate as autonomous unicellular entities, recent work has demonstrated that they actually employ highly specific intercellular communication networks. The promising potential exists for modification of the QS systems; for example, multi-component QS networks, when targeted by drugs, become disabled so bacteria can survive but in a less lethal mode [11-13]. Still, the extremely selective and often staggeringly complex nature of the QS response cascades has resulted in many incompletely understood intracellular mechanisms. In many cases, even if each individual component of a transduction

cascade is nominally identified, its exact function and significance often remains unknown [9].

As previously discussed, autoinducer-2 (AI-2) signaling is regarded as a ‘universal’ bacterial communication system due to its capability to prompt population based changes in multiple species of both Gram-positive and Gram-negative bacteria. First examined after mid-exponential phase cell supernatant from *Salmonella typhimurium* was shown to trigger a QS response (in the form of bioluminescence) in *Vibrio harveyi*, AI-2 later gained recognition as a sort of ‘bacterial Esperanto’, synthesized by LuxS and its homologues in multiple bacterial strains and evoking considerable changes in gene expression [14-16]. Taga *et al.* [17, 18] linked LuxS and AI-2 in *Salmonella typhimurium* to the expression of the *lsr* (LuxS-regulated) operon, encoding an ATP binding cassette (ABC)-type transporter determined to be responsible for the uptake of AI-2 into the cells. The divergent gene set was also shown to encode its cognate repressor, LsrR which was, in turn, ineffective in the presence of AI-2 [9].

However, low levels of native AI-2 detected in cell lysates led to the speculation that the signal molecule was somehow modified after uptake and that native AI-2 did not affect LsrR repression. It was later demonstrated in *Escherichia coli* that the kinase, LsrK, is directly responsible for the post uptake modification of AI-2, and the resulting intracellular phospho-AI-2 provoked de-repression of the *lsr* operon via LsrR [19, 20]. In deciphering expression of the *lsr* operon and its subsequent impact throughout the bacterial genome, focus has shifted to a systems level approach that includes the interplay between signaling, uptake, and a variety of

other physical processes. Li *et al.* [21] showed the importance of LsrR as not only a determinant of *lsr* expression, but as a regulator of hundreds of genes, many of which would likely be involved in virulence processes such as motility and biofilm formation [9].

With the significance of LsrR emerging, Xue *et al.* [22] evaluated the exact binding site for the repressor via DNA footprinting and protein binding assays. LsrR was shown to bind two separate sites within the intergenic region of the *lsr* operon, repressing transcription in opposite directions. Also, LsrR binding was apparently unaffected by the presence of native AI-2; increasing levels of phospho-AI-2, however, resulted in proportionate decreases in LsrR binding [22]. These results confirmed that the AI-2 must be modified by LsrK prior to antagonizing the cognate repressor. The direct regulation of the *lsr* operon by LsrR reaffirms the suspicion from Li *et al.* [21] that LsrR serves as an important ‘global’ regulator of AI-2 quorum sensing in bacteria [9].

Cued as a single signal regulator but with wide genetic impact, LsrR represents a promising new avenue for deciphering the complex nature of QS. Moreover, it represents another target for drug therapies which fight bacterial infection by preventing intercellular communication. Further progress will depend on a full understanding of LsrR and its effects on gene expression as well as the subsequent physiological changes spurred by its regulation. By continuing to dissect and understand biological signal transduction cascades, tools may then be devised increasing efficacy of tasks ranging from reducing pathogenicity of common bacterial infections or increasing commercial protein yield of production strains to the abiotic

extraction and optimization of mechanisms for environmental sensing and detection devices [9].

1.2 Background: Bacterial Communication

In the last 15 years, it has become evident that prokaryotes actually employ several modes of highly selective and relatively refined intercellular communication methods [23]. Two common types of signal systems in bacteria, two-component regulatory systems and QS systems, effect a signal transfer across the cell membrane [24]. Both systems are composed of a transmembrane signal transfer protein and correlating regulator protein which reacts to a specific trigger. However, in two-component systems the extracellular signal molecule is not normally taken up by the cell, while in QS systems the signal molecule will typically be transferred across the cell membrane and initiates a change in gene expression through direct interaction with regulatory proteins [24, 25].

In QS systems, the signal molecule is both produced and taken up by the cell, varying its intracellular and extracellular concentration [5, 20, 26, 27]. As cell density increases in a confined space, the extracellular concentration of the signaling molecule reaches a stimulating threshold, triggering the transduction cascade and resulting in a population-dependent shift in gene expression [19, 21].

Three types of QS systems have been defined which are based upon the type of signaling molecule involved:

Acyl-homoserine lactone (AHL) signaling, also known as autoinducer-1 (AI-1) signaling, is the predominant communication method among Gram-negative bacteria and is synthesized by a LuxI-type protein in most systems, although specific AHL

forms will vary among species (Figure 1-1A) [28-30]. Synthesized AHL begins to accumulate in the bacterial environment, passively or, more rarely, actively diffusing through the cell wall and eventually reaching a threshold concentration, stimulating binding to a LuxR-type sensor kinase protein. The LuxR-AHL complex then binds to and promotes expression of the *lux* (or analogous) operon [11, 31].

Oligopeptide signaling is utilized by Gram-positive bacteria, during which a peptide approximately 5 to 17 amino acids in length (Figure 1-1B) is synthesized and actively exported by ATP-binding cassette type transporters [4, 8, 32]. The extracellular peptide then initiates a chain reaction of phosphorylations via additional systemic proteins, until finally a DNA binding protein is modified and activated, controlling gene transcription [23].

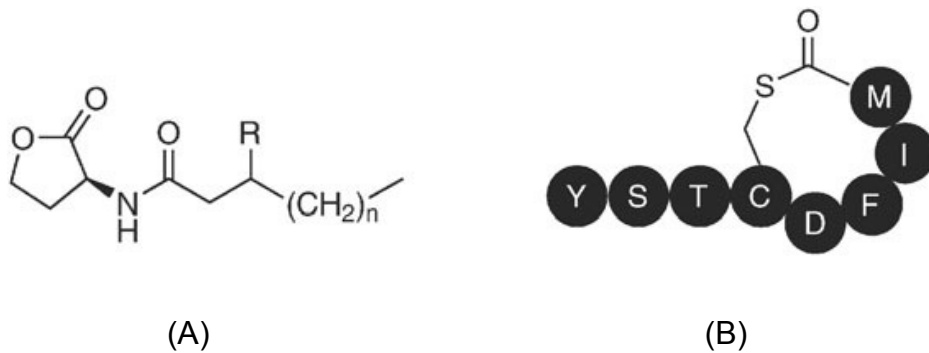


Figure 1-1. Intraspecies communication molecules.

(A) Acyl-homoserine lactone (AHL) molecule. R: -H, -OH, or =O; n: 3-11. (B) Oligopeptide structure. Source: Lowery *et al.* [8]

Autoinducer-2 (AI-2) signaling is regarded as a ‘universal’ bacterial communication system due to its capability to instigate population-based changes in genetic expression among multiple species of both Gram-positive and Gram-negative bacteria [8, 15, 33]. Able to trigger cross-species QS responses, homologues of AI-2 have since been identified in more than 70 bacterial species [34]. AI-2 is synthesized from S-adenosylmethionine (SAM) via three enzymatic steps (Figure 1-2):

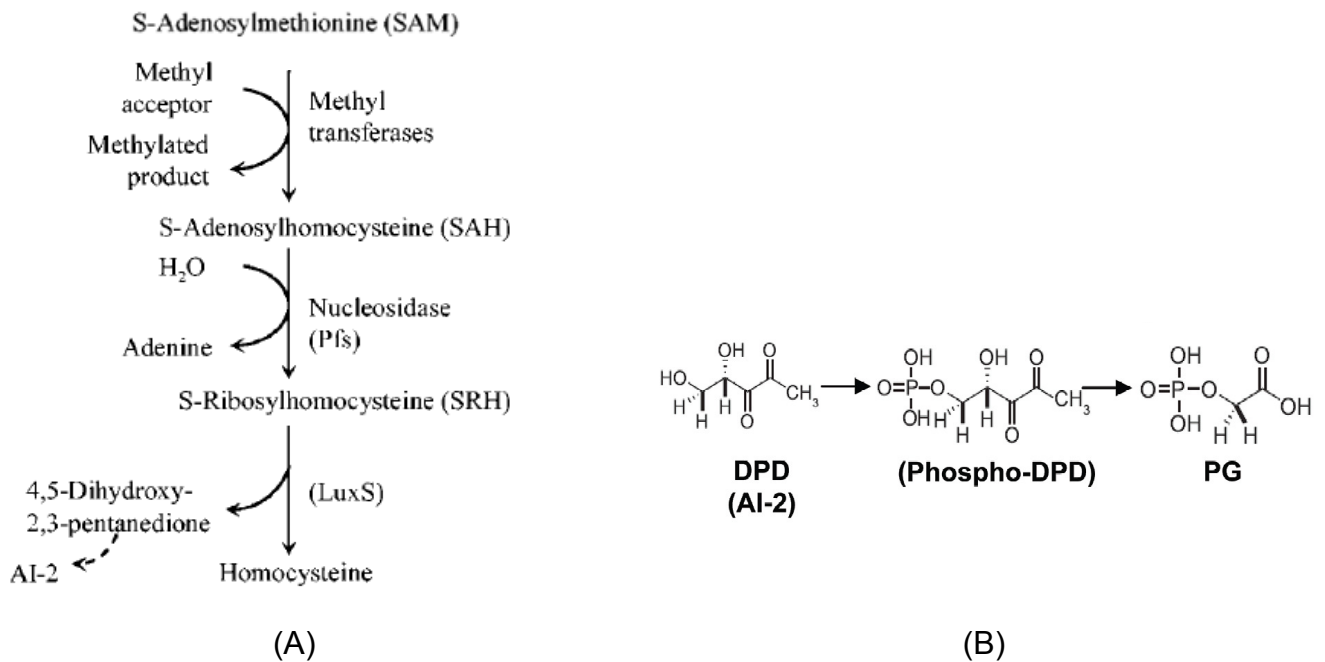


Figure 1-2. Synthesis of autoinducer-2 from S-adenosylmethionine (SAM).

(A) SAM is converted to SAH and SRH to synthesize DPD. Adapted from Wang *et al.* [27] (B) DPD, the precursor to AI-2, is phosphorylated via LsrK to form

phospho-DPD, but quickly degrades in solution to form 2-phosphoglycolic acid (PG) [12].

In *E. coli*, AI-2 is secreted from the cell and accumulates in the extracellular environment, where it is in turn transported back into the cell passively via diffusion or actively via LsrACDBFG (Figure 1-3). Upon reentry into the cell, AI-2 is phosphorylated by the kinase LsrK, and the resulting phospho-AI-2 complex purportedly antagonizes the *lsr* repressor protein LsrR, de-repressing transcription of the *lsr* operon [21, 22, 27]. Once repressor activity is negated, transcription of genes in the *lsr* regulon increases, promoted by the cAMP-CRP complex [27].

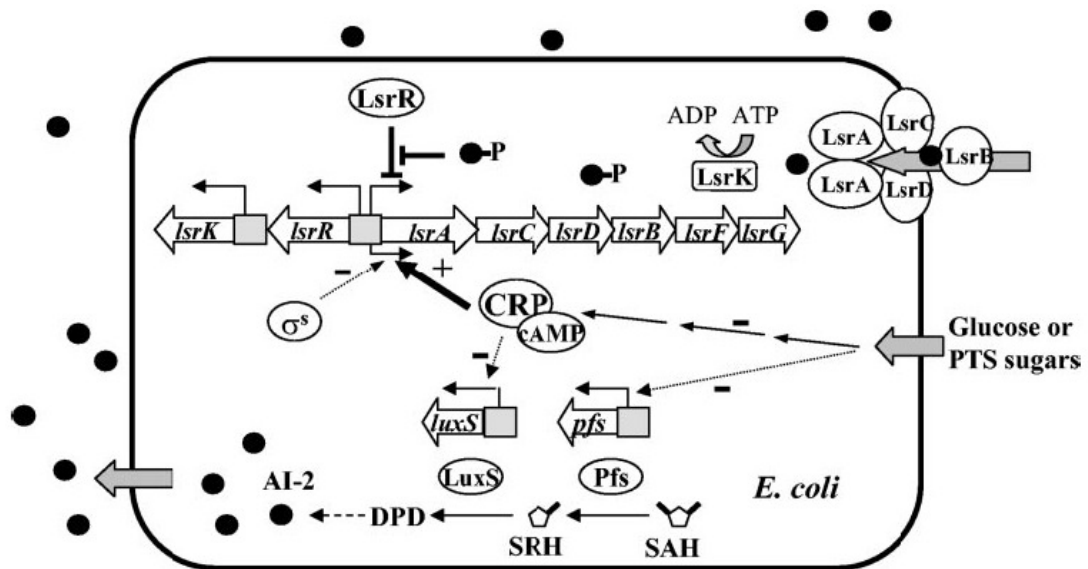


Figure 1-3. Autoinducer-based quorum sensing mechanism in *Escherichia coli*.

SAH is converted to SRH and DPD by Pfs and LuxS. DPD then spontaneously cyclizes to form AI-2, which is secreted and imported via LsrACDB. After phosphorylation by LsrK, the phospho-AI-2 complex antagonizes the repressor LsrR, permitting expression in the *lsr* operon. *Source: Wang et al. [27]*

1.3 Research Motivation

The AI-2 based QS has garnered significant attention in the last twenty years and progress in understanding the genetic circuitry involved in the QS process revealed its connection to myriad genetic and phenotypic attributes including virulence [32, 35], sporulation [11, 29], motility [36, 37], biofilm formation [38, 39], and cell growth [13, 40]. Extensive studies of enterohemorrhagic *E. coli* (EHEC) O157:H7 showed that AI-2 directly induced the virulence genes *LEE1-LEE3* and *tir*, and that roughly 10% of the genome is subsequently differentially transcribed as a response to the QS process [4, 37]. Similar studies in *E. coli* K12, *S. typhimurium*, and *V. harveyi* have also shown the potential global genetic impacts of QS [17, 18, 20, 21, 41].

Examining bacterial virulence as a result of quorum sensing is progressing in tandem with the study of biofilm formation. Involved in approximately two-thirds of bacterial infections and offering a safe haven for cellular proliferation in an otherwise hazardous environment, biofilms are now understood to often host multiple bacterial species simultaneously in an intricate and interdependent physical matrix [42, 43]. The role of AI-2 in eliciting biofilm formation was examined by Barrios *et al.* [38] through the addition of the quorum sensing molecule to mutant strains.

Much has been accomplished in elucidating specific elements of the QS circuit such as the involvement of LuxS, LsrK, and LsrACDB in the synthesis,

uptake, and phosphorylation of AI-2 [12, 44, 45], or the potential role of LsrFG in the breakdown of phospho-AI-2 [18]. The current model of the QS signal transduction circuit shows the *lsr* repressor, LsrR, to be a single point of action that, once antagonized, prompts the QS transcription cascade to begin. However, though the basic function of LsrR is known, the exact mechanisms and characterization of the repressor are still being evaluated [9, 22].

The QS process holds great potential for use in various fields. For example, artificially ceasing QS may help to fight infection, reducing the need for the use of antibiotics. Also, the food production industry seeks to harness the QS detection process in order to provide for the detection of food-borne pathogens [4, 29, 46, 47]. A more thorough understanding and successful application of the QS circuit demands further investigation of LsrR and its exact mechanism of repression in the *lsr* intergenic region.

1.4 Literature Review of LsrR

Although LsrR was shown, in the absence of AI-2, to repress transcription of the *lsr* operon in 2001 [17], it was not until 2005 that Taga *et al.* [18] first suggested that AI-2 requires phosphorylation prior to antagonizing LsrR. After this catalyst for relieving the repression of the *lsr* regulon by LsrR was explored, in 2007 the global impact of the repressor was examined using DNA expression microarrays. These data suggested that LsrR did, as seen in other organisms, have a widespread effect during the QS process, with 67 genes differentially regulated by LsrR and an additional 79 genes co-regulated by LsrR and LsrK [21].

Still, by 2009 the binding sites and regulatory mechanism by which LsrR effected repression in the *lsr* regulon or elsewhere had not been explored. While DNA binding analyses and motif confirmation are commonly investigated using *in vitro* assays, the challenges of *in vitro* work with LsrR (discussed further in Chapter 2) may have inhibited rapid discovery of binding locations. Xue *et al.* [22] reported success with localizing LsrR binding to two general locations in the *lsr* intergenic region, each within 50 bp upstream of the *lsrR* and *lsrA* start sites. Also reported was the first *in vitro* response of LsrR to phospho-AI-2 and its subsequent dissociation from DNA [22].

In 2010, although LsrR or its homolog had been indicated to have widespread impact in multiple species including *E. coli* [21], *S. typhimurium* [18], and *V. harveyi* [48], Thijs *et al.* [49] used chromatin immunoprecipitation in conjunction with microarrays (ChIP-Chip) in *S. typhimurium* to show, surprisingly, a single LsrR binding location within the *lsr* set of genes. These data, however, may be vulnerable to over- or underrepresentation of actual binding events for multiple reasons [50], and are discussed further in Chapter 3.

LsrR continues to be characterized, and in 2010 preliminary crystallographic analysis of the repressor was reported, estimating its structure to contain two protein molecules per asymmetric unit [51, 52]. Multimerization of a DNA-binding protein is not uncommon [53-55], and while the dimeric form reported for an overexpressed, his-tagged LsrR is unsurprising, the *in vivo* form during which full repression is enabled requires further investigation (Chapter 4).

1.5 Global Objective, Global Hypothesis, and Specific Aims

The global objective of this dissertation is to further investigate the regulation of expression as a result of the QS cue in *E. coli*.

Global Hypothesis: LsrR displays complex global and local regulatory binding.

Specific Aim 1: Assess LsrR activity and potential use *in vitro*.

Specific Aim 2: Determine specific binding sites for LsrR outside of the *lsr* region *in vivo*.

Specific Aim 3: Describe the mechanism of LsrR binding and repression and assess the role of cAMP-CRP in the *lsr* intergenic region.

1.6 Dissertation Outline

Chapter 2 describes the expression and purification of LsrR and *in vitro* assays used in order to assess CRP and LsrR binding activity. Various conditions for LsrR growth, purification methods, and plasmid constructs are described. cAMP receptor protein (CRP) binding analyses and methods are also shown.

Chapter 3 includes the global binding analysis for LsrR using DNA microarrays and discusses the potential impacts of the various binding sites.

Directional expression analysis of the divergent *lsr* regulon is also included, and directional bias is evaluated.

Chapter 4 describes and examines putative LsrR binding sites and describes a model for repression in the *lsr* intergenic region, incorporating LsrR, CRP, and DNA as a regulatory complex with an introduction to an additional integration host factor (IHF)-like regulatory mechanism in the *lsr* regulon.

Chapter 5 discusses current and potential future work applicable to LsrR and QS regulation and summarizes the previous chapters.

Chapter 2: Expression, *In Vitro* Activity, and Structure Analysis of LsrR

2.1 Abstract

In current models of the quorum sensing (QS) system, the repressor LsrR is the single point of relief for negative regulation of the *lsr* regulon and dissociates from the intergenic DNA after antagonization by the phosphorylated form of the communication molecule autoinducer-2 (AI-2). Once the *lsr* regulon is de-repressed, cyclic-AMP receptor protein (CRP) activates transcription of the divergent *lsr* gene set by binding upstream from transcription start sites and recruiting RNA polymerase holoenzyme (RNAP) to the region. In this study we construct a plasmid vector encoding LsrR in order to harvest the repressor for *in vitro* analysis, and use electrophoretic mobility shift assays (EMSA) to evaluate LsrR and CRP binding affinity to the *lsr* intergenic region. While LsrR demonstrated minimal *in vitro* binding activity under the tested conditions, CRP was confirmed to have two distinct binding sites in the *lsr* intergenic region with similar binding affinity despite the difference in binding site sequence.

2.2 Introduction

The QS repressor LsrR has been shown to negatively regulate expression in the *lsr* regulon of *E. coli* and *S. typhimurium* by binding to and inhibiting expression of the divergent gene set [17, 18, 21, 22]. After the QS molecule AI-2 is taken up by the cell and phosphorylated by LsrK, the resulting phospho-AI-2 complex antagonizes LsrR, resulting in its dissociation from the *lsr* intergenic region and subsequent expression of the *lsr* genes [18, 27, 44].

The impacts of LsrR on global genetic expression have been evaluated in *E. coli* K12, *S. typhimurium*, and *V. harveyi*, and numerous genes are confirmed to be differentially regulated as a result of LsrR activity [17, 18, 20, 21, 41]. However, in *S. typhimurium* chromatin immunoprecipitation in conjunction with DNA microarrays (ChIP-Chip) data showed only a single genomic binding location for LsrR, located in the *lsr* operon [49]. This discrepancy in binding locations versus differential regulation may be due to various causes. For example, the dissociation of LsrR from the *lsr* intergenic region may instigate a secondary regulatory cascade with alternate transcriptional regulators contributing to gene activation [56, 57]. Alternatively, inherent error in the ChIP-Chip process or complex protein-DNA interactions during LsrR repression may prevent harnessing of the repressor and underrepresent actual *in vivo* binding events [50, 58, 59].

The *E. coli* cyclic-AMP reception protein (CRP) activates gene transcription at over 100 promoters by binding to regions upstream of gene start sites and recruiting RNA polymerase holoenzyme (RNAP) to the location, subsequently increasing transcription of the target gene [60-62]. In some locations, however, CRP

binds to regions which overlap promoter segments, resulting in RNAP inhibition and acting to negatively regulate expression in those genes [63-65]. CRP changes conformation by interacting with the effector cyclic-AMP, after which it may bind to its consensus sequence (primary binding sites in bold):

(5'-AAAT**GTGATCTAGATCACATTT**-3') [60]. Two separate CRP binding locations have been previously identified in the *lsr* intergenic region and expression analysis of the divergent genes in the absence of CRP (by preventing intracellular cAMP availability through the addition of glucose) demonstrated a precipitous drop in expression of both gene sets [20, 27].

Studies involving DNA-binding proteins commonly use electrophoretic mobility shift assays (EMSA) in order to test protein binding affinity to various native or mutated DNA fragments [66-68]. Although two CRP sites were identified in the *lsr* intergenic region, neither binding location region carries the complete consensus sequence indicated above. In this study, we evaluate CRP binding to the two regions and compare binding affinity to the differing sequences and search the remainder of the intergenic region for possible additional CRP binding sites. Additionally, because the data for LsrR binding and global impact in *S. typhimurium* are apparently deviant, in this study we also construct a plasmid for overexpression and purification of LsrR, and focus on DNA binding activity by using various sections of the *lsr* intergenic region. Finally, we conduct preliminary analysis of LsrR secondary structure using similar repressors in the same regulatory family.

2.3 Materials and Methods

Bacterial strains, plasmids, and growth conditions.

E. coli strains used in this work are described in Table 2-1. All cultures were grown aerobically at 37°C in Luria-Bertani (LB) media or as specified. Media were supplemented as necessary with antibiotics at the following concentrations: ampicillin, 50µg ml⁻¹; kanamycin, 50µg ml⁻¹.

Plasmid construction. Plasmids and primers used in this study are listed in Table 2-1. Chromosomal DNA was purified from *E. coli* MG1655 (DNeasy tissue kit, Qiagen, CA, US) and used as the template for all PCR reactions. LsrR was expressed and purified by constructing a TOPO® clone using the pET200 vector (Invitrogen, CA, US). Briefly, the 954 bp open reading frame (ORF) for *lsrR* (NC_000913, 1598312...1599265) was amplified via PCR from an *E. coli* K-12 whole genome sample using SYBR green PCR master mix (Applied Biosystems, CA, US). Both plasmid and PCR product insert were digested with NheI and MfeI restriction enzymes (NEB, MA, US), and pET200LsrR was assembled using a quick ligation kit (NEB). Ligated plasmids were transformed into chemically competent TOP10 *E. coli* (Invitrogen) and cultured as specified previously. Construct sequences were confirmed via sequencing at the DNA core facility of the Center for Biosystems Research (University of Maryland Biotechnology Institute).

Protein Purification. CRP samples were supplied by Fred Schwarz, University of Maryland Biotechnology Institute. For LsrR production, plasmid pET200LsrR was

Table 2-1. *E. coli* strains, plasmids and primers used in this study.

Name	Description ^a	Source or Ref.
<i>E. coli</i> strains		
K12 MG1655	Wild type	Laboratory stock
BL21DE3	1. <i>fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS</i>	NEB
TOP10	F ⁻ <i>mcrA</i> Δ(<i>mrr-hsdRMS-mcrBC</i>) φ80 <i>lacZ</i> ΔM15 Δ <i>lacX74 recA1 araD139</i> Δ(<i>ara-</i>	Invitrogen
Plasmids		
pET200LsrR	expression vector encoding <i>lsrR</i> with N-terminus 6xHis-tag, Kan ^r	Invitrogen
Primers		
plsRFWD	CACCATGACAATCAACGATTCGGCAATTC	This study
plsRREV	TGCTGTGTCCTGATCGGTAACCAGTGCGTT	This study
lsrIGF	ATTTCCCCCGTTCAGTTTTGCAGGTGAG	This study
lsrIGR	TAAATCTTTAATGCAATTGTTCAAGTTCT	This study
crpbsAF	ATTTCCCCCGTTCAGTTTTGCAGGTGAGTTTTGAACAAATGTATTTCTGC	This study
crpbsAR	GCAGAAATACATTTGTTCAAAACTCACCTGCAAAACTGAACGGGGGAA	This study
crpbsBF	TTTTAATTTGTTTATAACCTTAGGTGGACATTGCACATATTTCCGACGA	This study
crpbsBR	TCGTTCGAAATATGTGCAATGTCCACCTAAGGTTATGAACAAATTTAAA	This study
crpbsCF	TAGATCACAATTTATGCTATTTTGATTTTACGGTTGCGTTTGTTCATGC	This study
crpbsCR	GCATGAACAAACGCAACCGTGAAAATCAAAATAGCATAAATTGTGATC	This study
crpbsDF	TCGTAGAGTCAAAGTGTGGTTGCCATCACAGATATAAATGAGCAAGAA	This study
crpbsDR	AGTTCTTGCTCATTATATCTGTGATGGCAACCACAGTTTGACTCTACG	This study
crpbsEF	GAACAATTGCATTAAAGATTTAAATATGTTCAAAGTGAAGAATGAATT	This study
crpbsER	AATTCATTCTTCACTTTGAACATATTTAAATCTTTAATGCAATTGTTT	This study
crpbs1F	GAGTTTTGAACAAATGTATTTCTGCTTTTAATTTGTTTATAACCTTAGGT	This study
crpbs1R	ACCTAAGGTTATGAACAAATTTAAAAGCAGAAATACATTTGTTCAAAC	This study
crpbs2F	GGACATTGCACATATTTCCGACGAATAGATCACAATTTATGCTATTTTG	This study
crpbs2R	TCAAATAGCATAAATTGTGATCTATTCGTTCGAAATATGTGCAATGTC	This study
crpbs3F	TTTTACGGTTGCGTTTGTTCATGCTCGTAGAGTCAAAGTGTGGTTGCC	This study
crpbs3R	TGGCAACCACAGTTTGACTCTACGAGCATGAACAAACGCAACCGTGAA	This study
crpbs4F	TCACAGATATAAATGAGCAAGAACTGAACAATTGCATTAAAGATTTAA	This study
crpbs4R	ATTTAAATCTTTAATGCAATTGTTCAAGTTCTTGCTCATTATATCTGTGA	This study

purified from an overnight culture of TOP10+pET200LsrR using a SpinPrep Miniprep Kit (Qiagen, CA, US). The pET200LsrR vector was then transformed into chemically competent *E. coli* BL21DE3 (NEB) and cultured as previously described. When the optical densities (OD600) of the cell cultures reached 0.4, growth media were supplemented with 1mM IPTG in order to induce protein production. After 4 hours of growth, cells were collected by centrifugation at 5,000 xg for 20 minutes at 4°C. The cell pellet was resuspended in the Bugbuster protein extraction reagent (Merck, GE) and shaken at 100 rpm at room temperature for 20 minutes in order to lyse cells. Samples were centrifuged at 1500xg at 4°C for 20 minutes in order to separate the inclusion body. Supernatant was collected, and 25 units of Benzonase (Merck) were added per ml of supernatant in order to reduce viscosity. The pET200LsrR plasmid encodes a hexahistidine tag (6xHis) onto the N-terminus of LsrR. Purification of His-LsrR was done using Ni-charged His-bind resin kit (Novagen) with affinity to the 6xHis tag. Briefly, resin beads were washed (0.5M NaCl, 120mM imidazole, 40mM Tris-HCl, pH 7.9) and charged (50mM NiSO₄) and resulting slurry was added to cell lysate and allowed to incubate on a rotating shaker at 4°C for 4 hours. Resin beads were separated by centrifugation at 100 rpm at room temperature. Beads were then washed thoroughly (0.5M NaCl, 120mM imidazole, 40mM Tris-HCl, pH 7.9) and protein was eluted using the supplied elution buffer (0.5M NaCl, 1M imidazole, 20mM Tris-HCl, pH 7.9). For gradient elution analysis, washes of LsrR-bound beads were conducted three times with wash buffer (0.5M NaCl, 40mM Tris-HCl, pH 7.9) with 120mM, 300mM, and 600mM of imidazole in washes 1-3, respectively. After washes, beads were eluted as discussed above.

Electrophoretic Mobility Shift Assays (EMSA). For EMSA with LsrR, the entire 248 bp *lsr* intergenic region was amplified via PCR using primers lsrIGF and lsrIGR (IDTDNA, IA, US). DNA was purified using a Quick PCR purification kit (Qiagen). For DNA preparation for EMSA with CRP, the *lsr* intergenic region was segmented into 50 bp overlapping sections (Figure 2-1A).

To construct dsDNA corresponding to each section, appropriate ‘crpbs’ primers listed in Table 2-1 were annealed by adding 10mM of each forward and reverse primer to 100 ul of annealing buffer (10mM Tris-HCl, 100 mM NaCl, 1mM EDTA, pH 7.5). Binding reactions included 10 μ M of CRP or 10-100 μ M of His-LsrR with 300 ng of appropriate DNA segments in a binding buffer containing 10mM Tris-HCl, 50mM KCl, 1mM EDTA, 1mM DTT. EMSA assays with CRP also included 100 μ M cAMP, and assays with LsrR included 30 μ M of unphosphorylated AI-2. Additionally, LsrR was grown in BL21DE3 Δ *luxS* in order to eliminate the presence of endogenous AI-2, preventing any phospho-AI-2 from binding to the purified LsrR during growth. Samples were run on a non-denaturing 8% Tris-HCl polyacrylamide gel (Bio-rad) in Tris-glycine buffer (25 mM Tris, 190 mM Glycine, pH 8.3) at 120V for 40 minutes.

2.4 Results

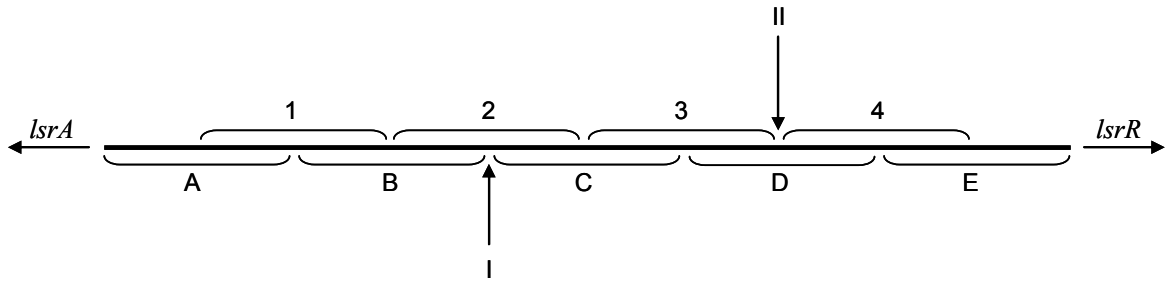
Intergenic region carries two CRP binding sites with similar affinity. EMSA was conducted with CRP and DNA segments A-E and 1-4 shown in Figure 2-1A. CRP

combined with segments '2' and 'D' showed the most significant shift and corresponds to previously reported CRP binding sites [20, 27]. Each of these two duplexes carry the entire CRP binding motif indicated for that site and display comparable magnitudes of shift, indicating a similarity in binding affinity for CRP despite the variation in sequence between the two (Figure 2-1B). Band intensity analysis using Quantity One software Version 4.6.5 shows magnitudes of shifted DNA bands are comparable to within 4%.

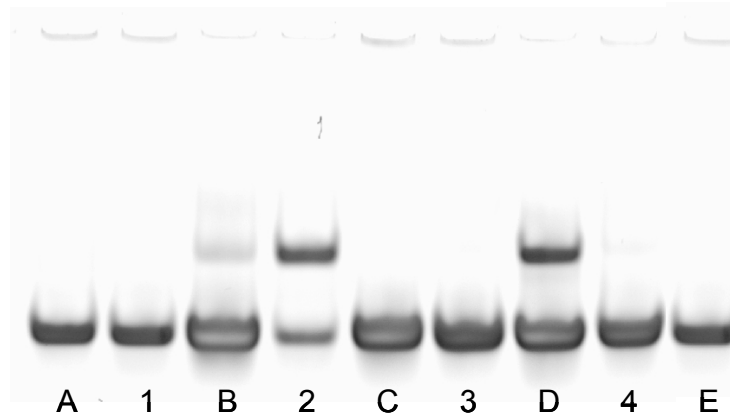
Figure 2-1. EMSA analysis of CRP.

(A) Each bracketed segment includes the 50 bp duplex section of the *lsr* intergenic region used in EMSA with CRP. Vertical arrows indicate locations of CRP binding sites I and II [20, 27]. For EMSA, 300 ng of intergenic DNA segments from (A) were incubated with 10 μ M of CRP in 20 μ l of EMSA binding buffer (10mM Tris-HCl, 50mM KCl, 1mM EDTA, 1mM DTT) for 10 minutes at room temperature. (B) Non-denaturing PAGE gel analysis shows significant, similar shifts of DNA segments '2' and 'D' confirming presence of two CRP binding sites with similar affinity. Less prominent shifts seen with segments 'B' and '4' are presumed to be due to the presence of half of the CRP binding site. Gel was stained with SYBR Green nucleic acid gel stain. (C) CRP binding motifs within segments '2' and 'D'. Bold letters indicate sequences that are also included on adjacent (Segments 'B' and '4') duplexes, resulting in faint bands seen in those lanes.

A



B



C

Duplex '2'
-92
TCCGAcgaatagaTCACA

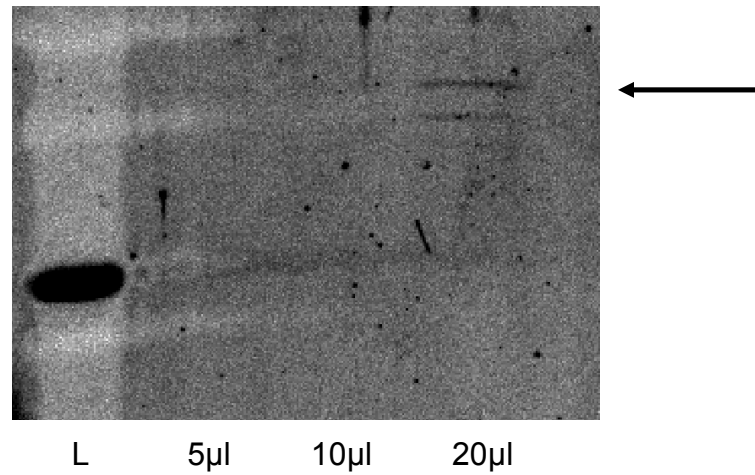
Duplex 'D'
-165
TGTGGttgccaTCACA

Additionally, a lesser, but still noticeable DNA shift is present in lanes 'B' and '4'. Each of these sections carries half of the adjacent CRP binding site, with sections 'B' and '4' having sequence 'TCCGAcgaa' and 'TCACA', respectively (Figure 2-1C) and likely minimal CRP binding results in the partial observed shift.

More than 95% LsrR remains in inclusion body after purification. LsrR was grown and purified as described in Materials and Methods, and various volumes of eluate were examined using SDS-PAGE. Flamingo protein stain of gel shows undetectable levels of LsrR in eluate volumes of 5 μ l and 10 μ l, with only minimal levels of LsrR present in the 20 μ l column (Figure 2-2A).

Since LsrR purification yielded low levels of protein, gradient elutions using wash buffers with increasing amounts of imidazole were conducted in order to assess whether the bound LsrR was dissociating from the purification beads during washing. Also, the inclusion body formed during purification was analyzed using PAGE-SDS in order to see if LsrR was precipitating out of solution. Figure 2-2B shows gradient washes 1-3, elution 'E', and inclusion body (IB) of the LsrR purification process. In each sample 1-3, low levels of LsrR are detectable in solution but, as compared to the final elution, demonstrate a significant loss of overall soluble protein during washing. Prominent band at the theoretical LsrR size of 37kDa in inclusion body column (horizontal arrows in Figure 2-2B) also shows a loss of the majority of LsrR during purification.

A



B

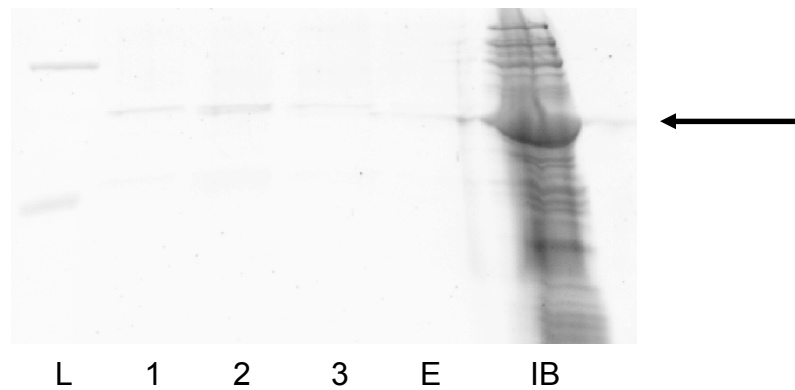


Figure 2-2. LsrR purification analysis.

(A) Flamingo protein stain of 5μl, 10μl, and 20μl volumes of LsrR purification elution in denaturing SDS-PAGE gel. LsrR only becomes detectable with 20μl of eluate. (B) 20μl of washes 1-3 and elution 'E' of protein purification beads. 'IB' shows 1μl of inclusion body resuspended in 20μl PBS. Prominent band in 'IB' column shows presence of majority of His-LsrR. Theoretical size of LsrR is 37kDa (horizontal arrows). 'L' in each gel is 8μl is Precision Plus Protein standard (Biorad).

No DNA binding activity detectable *in vitro* with N-terminus His-tagged LsrR.

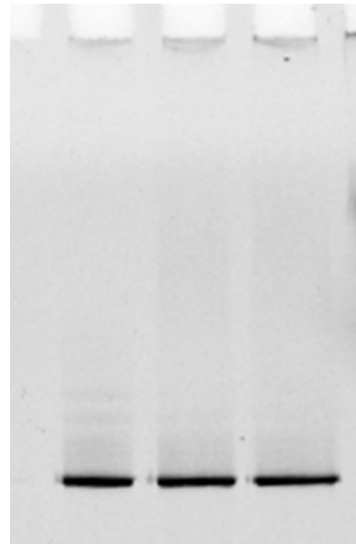
Although low levels of soluble LsrR were available after purification, sensitivity of EMSA assays permit *in vitro* analysis using similar samples [67, 69]. EMSA with LsrR was conducted under the same conditions used for CRP. Due to its potential role as an activator for LsrR, 30 μ M AI-2 was also added [18, 27, 34, 70].

Increasing levels of LsrR showed no *in vitro* binding activity when combined with the *lsr* intergenic region (Figure 2-3A). Nucleic acid stains of EMSA PAGE gels displayed no shift or other drop in magnitude as compared to the 'DNA only' lane indicating a lack of affinity for the tested DNA segments under conditions used. In order to evaluate the possibility that CRP binds to the *lsr* intergenic region and recruits or facilitates LsrR binding, we conducted another EMSA using the *lsr* intergenic region, 10 μ M LsrR, and 8 μ M CRP (Figure 2-3B). Results again show no shift with the DNA-LsrR only sample and a significant shift in the CRP-DNA lane (lanes 3 and 4 in Figure 2-3B, respectively). However, there was no apparent change between the CRP-DNA sample and the CRP-LsrR-DNA sample (lanes 4 and 5 in Figure 2-3B, respectively), implying no CRP facilitation of LsrR binding to DNA.

Figure 2-3. EMSA with LsrR and CRP.

(A) Nucleic acid stain of EMSA with increasing amounts of LsrR showing no visible shift in bands. (B) Nucleic acid and protein stain of EMSA. Shift of DNA in lane 4 confirms *in vitro* activity of CRP. Lack of variation between lanes 4 and 5 indicates no apparent affect of CRP on LsrR activity. Protein concentrations in (B) are: LsrR - 10 μ M, CRP - 8 μ M. 300 ng of DNA was used in all samples.

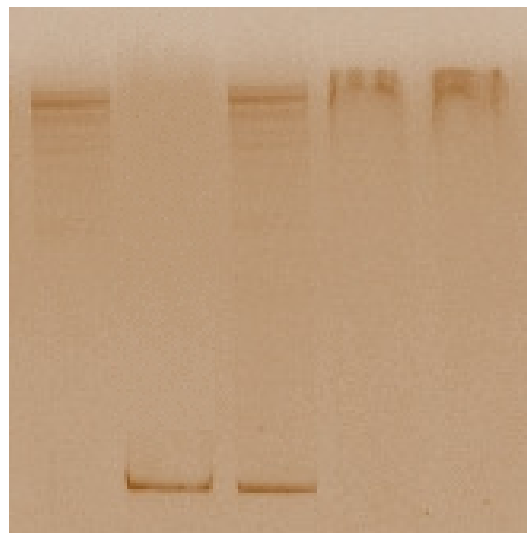
A



1: DNA only
2: DNA+10µM LsrR
3: DNA+100µM LsrR

1 2 3

B



1: LsrR only
2: DNA only
3: DNA+100µM LsrR
4: DNA+10µM CRP
5: DNA+10µM CRP
+100µM LsrR

1 2 3 4 5

2.5 Discussion

The current model of transcriptional regulation in the *lsr* regulon incorporates CRP as an activator and LsrR as the single repressor of expression [27, 44]. CRP is a well-characterized activator in *E. coli*, functioning by binding in upstream promoter regions and subsequently recruiting RNA polymerase holoenzyme (RNAP), resulting in transcription initiation [27, 71-73]. Although two CRP binding sites have been identified in the *lsr* intergenic region, the sites do not share the same consensus sequence and no study has examined CRP binding affinity to these motifs or used *in vitro* binding assays to search for additional sites in the region.

Additionally, the implications of LsrR as the single point of relief of *lsr* transcriptional repression during the QS process demonstrate the importance of understanding the regulator and its specific function [18, 19, 44]. At the time of this writing, however, the binding sites for LsrR in the *lsr* regulon are still undefined, and LsrR activity *in vitro* is elusive. In this study we examined CRP binding to the *lsr* intergenic region, its affinity for the two previously identified binding sites, and searched for additional sites that may contribute to control of the *lsr* regulon. We also constructed and purified a His-tagged LsrR in order to assess activity *in vitro* in an effort to define its consensus motif and precise binding location in the *lsr* region.

In vitro analysis of CRP binding using the segmented *lsr* intergenic region (Figure 2-1) validated the two previously identified binding sites and confirmed that only two CRP sites exist in this region. Additionally, the EMSA results show a similar magnitude in DNA shift from CRP site I to II, indicating a similar affinity to the sites regardless of their differences in sequence. The locations of each CRP site at

-91 and -68 from the *lsrA* and *lsrR* start sites, respectively, and the similar binding affinity to each site suggest that transcription in each direction is effectively a separate CRP-dependent promoter [60, 74]. However, each CRP protein while bound has the potential for affecting distal RNAP binding due to DNA bending and conformation and classification as a type I (single point of interaction between RNAP and one CRP protein), II (two points of interaction between RNAP and one CRP protein), or III (two points of interaction between RNAP and two CRP proteins) CRP-dependent region can only be determined through individual site mutation and expression analysis [60, 75, 76]. CRP analysis and classification in the *lsr* regulon is discussed further in Chapter 4.

The purification process for LsrR displayed a striking loss in recoverable protein, partially during the washing process but more obviously due to the majority of LsrR remaining in the insoluble inclusion body. Protein insolubility is the focus of a staggering number of published studies, and causes of insolubility may include protein surface charge, size, rate of formation, or unfavorable features of three dimensional structure, among others [77-80].

While the recoverable portion of LsrR did not show detectable activity *in vitro* it is possible that the most active protein is also hydrophobic, resulting in the apparent lack of DNA binding during EMSA analysis. Multiple modifications to protein growth were attempted in order to improve soluble LsrR yields. These include lowering growth temperature in order to slow misfolding, reducing protein induction rates, growth in *E. coli* strains which enhance disulfide bond formation (hence increasing stability), growth in various media types which may help slow or speed

protein production, resuspension of eluted LsrR in buffers with varying pH in order to increase stability, and denaturing and refolding of LsrR that had precipitated into the inclusion body. However, none of these approaches yielded significant increases in protein yield or activity (data not shown).

In order to progress with understanding LsrR characteristics and refining the repressor for *in vitro* use, we searched protein secondary structure databases NCBI (<http://www.ncbi.nlm.nih.gov>) and UniProt (<http://www.uniprot.org/>) for homologous proteins. LsrR is a member of the SorC family of transcriptional regulators, which feature a putative sugar binding domain and an N-terminus helix-turn-helix DNA-binding domain. DeoR, also in the SorC family, regulates transcription *in vivo* as a tetramer, binding to at least two operator sites simultaneously and holding the upstream region of DNA in a looped state [67, 68, 81, 82]. Similar multimeric form and multiple binding site activity are found in other regulators in the SorC family as well [69, 83, 84].

In conclusion, if LsrR has a DNA-binding domain proximal to the N-terminus his-tag encoded by pET200LsrR, the tag may interfere with DNA binding activity *in vitro* and prevent the shift of DNA during EMSA. While unusual, his-tag interference has been shown to affect protein activity previously [85]. Analysis of LsrR with an oppositely located C-terminus his-tag and resulting activity are further described in Chapter 3. The analysis of the CRP sites in the *lsr* intergenic region has also confirmed the presence of two binding locations having similar affinity. Various models for CRP activation include not only the recruitment of RNAP, but in cases of multiple sites may involve DNA looping as part of the regulatory mechanism [86-89].

Also, in some cases CRP can bind near to and overlapping the transcription start site, acting effectively as a repressor of transcription [90, 91]. The distinct impact of each CRP site and the model for regulation of the *lsr* regulon are presented in Chapter 4.

Chapter 3: Genomic binding site analysis of the AI-2 quorum sensing regulator LsrR in *Escherichia coli*

3.1 Abstract

The quorum sensing (QS) process evokes significant phenotypic changes among multiple species of bacteria as a response to the ‘universal’ signal molecule autoinducer-2 (AI-2). Recent studies in *S. typhimurium*, *V. harveyi*, and *E. coli* have shown high conservation of the LuxS-regulated (*lsr*) set of genes which are expressed divergently as a direct result of phospho-AI-2 antagonizing the QS repressor LsrR. Given the widespread impact of QS on bacterial genetic expression, we used chromatin immunoprecipitation in conjunction with microarrays (ChIP-Chip) to discover genomic binding sites for LsrR. Our results show four putative binding locations: *lsrR-lsrA*, *yegE-udk*, *mppA* and *yihF*. Of these genes or gene sets, only *yihF* has not been characterized and would not logically be associated with the QS response. Also, because of the apparently unique role of LsrR in repressing the divergent *lsr* regulon, we used reporter plasmids with bidirectional promoter segments to further examine expression using β -galactosidase assays. We report for the first time a natural expression bias, independent of LsrR repression, of the AI-2 transporter *lsrACDBFG*.

3.2 Introduction

The term quorum sensing (QS) is used to define the communication mechanism in bacteria which uses chemical signal molecules called autoinducers to induce distinct genetic and physiological changes as a response to local population density [5, 29, 33]. Several QS systems have been identified within distinct bacterial populations that mediate intraspecies communication and are characterized by signaling molecules such as acyl-homoserine lactones or oligopeptides [92, 93]. Autoinducer-2 (AI-2) based QS prompts population-dependent genetic changes between many species of both Gram-positive and Gram-negative bacteria indicating its potential as a ‘universal’ bacterial communication system [8]. Supernatant from mid-exponential cell culture of *Salmonella typhimurium* was first documented to generate a QS response in *Vibrio harveyi* [14, 16]; since then, AI-2-dependent QS genes have been identified in more than 70 bacterial species and the QS signal molecule is now linked to multiple phenotypic changes including virulence, motility, sporulation and biofilm formation [8, 20, 29, 94].

In *E. coli*, the enzymatic synthesis steps of AI-2 include the metabolic conversion of S-adenosylmethionine (SAM) to S-adenosylhomocysteine (SAH) and then S-ribosylhomocysteine (SRH) via the enzyme Pfs. SRH then interacts with LuxS to form the precursor 4,5-dihydroxy-2,3-pentanedione (DPD) which spontaneously cyclizes to form two epimeric furanoses, (2*R*,4*S*)- and (2*S*,4*S*)-2,4-dihydroxy-2-methyldihydrofuran-3-one (*R*- and *S*-DHMF, respectively) [44]. After synthesis AI-2 is secreted from the cell via a putative exporter and accumulates in the extracellular environment [32, 95]. Once it reaches a threshold level it is transported

back into the cell via the *lsr* (LuxS-regulated) ATP binding cassette transporter, LsrACDB, and is phosphorylated by the kinase, LsrK. The resulting phospho-AI-2 complex purportedly antagonizes the *lsr* repressor protein LsrR, de-repressing transcription of the *lsr* operon and is further processed by LsrFG in order to prevent excess intracellular accumulation of the signal molecule [19].

Genes involved in QS are highly conserved, and exploration of the QS systems in *S. typhimurium*, *V. harveyi*, and *E. coli* are demonstrating functional similarities between species as well [5, 6, 96, 97]. The exact function of the *lsr* operon genes as well as the divergently transcribed repressor *lsrR* and kinase *lsrK* continue to be elucidated; recent work has described the phosphorylation kinetics of AI-2 with LsrK and the affinity of the product of this reaction with LsrR as well as AI-2 quorum quenching between populations via extracellular phosphorylation and subsequent reduction of the QS response due to bacterial inability to transport phospho-AI-2 [12, 44].

While DNA binding sites and *in vitro* LsrR-DNA binding activity have been reported in *S. typhimurium* and effects of its deletion on gene expression in *E. coli* have been revealed, the extent to which LsrR serves as a direct transcriptional regulator in the QS regulon has not been fully elucidated [21, 22]. Although in various bacterial species the genetic impact of the QS process appears to be considerable, specific binding site analysis for LsrR in *S. typhimurium* using chromatin immunoprecipitation in conjunction with microarrays (ChIP-Chip) showed only one binding location positively controlled by the QS repressor [49]. In this study we use ChIP-Chip to discover potential binding sites of the QS repressor LsrR

in *E. coli* MG1655 as a basis for comparison of other bacterial strains utilizing AI-2 QS. We also examine LsrR activity in the *lsrA-lsrR* divergent promoter region and measure its direct effect on expression levels using β -galactosidase activity assays. Finally, we more closely examine the *lsr* intergenic region and divergent expression effects of LsrR and promoter bias.

3.3 Materials and Methods

Bacterial strains, plasmids, and growth conditions. *E. coli* strains used in this work are described in Table 3-1. Cultures were grown aerobically at 37°C in Luria-Bertani (LB) medium to OD 0.35 and supplemented with 1mM IPTG. Cells used in microarray experiments were harvested at mid-exponential (OD 0.8) and stationary (OD 1.5) phases. Media were supplemented as necessary with antibiotics at the following concentrations: ampicillin, 50 μ g ml⁻¹; kanamycin, 50 μ g ml⁻¹; chloramphenicol 25 μ g ml⁻¹.

Plasmid construction. Plasmids and primers used in this study are listed in Table 3-1. Chromosomal DNA was purified from *E. coli* MG1655 (DNeasy tissue kit, Qiagen, CA, US) and used as the template for all PCR reactions. For pTH2lsrR, *lsrR* was amplified by PCR using oligonucleotide primers th2lsrRF and th2lsrRR. The PCR product and pTrcHis2 vector (Invitrogen, CA, US) were digested with BamHI and HindIII and purified (QIAquick PCR purification kit, Qiagen). Samples were ligated (Quick Ligation Kit, NEB, MA, US), transformed into TOP10 chemically

competent cells (Invitrogen) and spread onto LB-agar plates with ampicillin.

pLsrR26 and pLsrA14 were constructed by amplification of the *lsr* intergenic region (-262 to +12 relative to the *lsrR* start site) using lsrRP26F/ lsrRP26R and lsrAP14F/lsrAP14R, respectively, and digesting the resulting PCR product with BamHI and EcoRI. Plasmid pFZY1, a low-copy mini-F derivative with MCS upstream of a promoterless *galK'-lacZYA* reporter segment [98] was also digested with BamHI and EcoRI and purified (Gel Purification Kit, Qiagen). Ligation and selections were performed as described above. All plasmids were validated by DNA sequencing, performed at the DNA Core Facility of the Institute for Bioscience and Biotechnology Research (IBBR).

Table 3-1. *E. coli* strains, plasmids and oligonucleotides used in this study.

Name	Description or Sequence	Source
K12 Strains		
MG1655	Wild type	Lab stock
CB11	MG1655 Δ lsrR::Cm	This study
CB13	MG1655 Δ lsrR Δ lsrK::Cm, Kan	This study
Plasmid		
pTHlsrR	pTrcHis (Invitrogen) derivative, (N-terminus His-tag) containing <i>lsrR</i> , Ap ^r	This study
pTH2lsrR	pTrcHis2 (Invitrogen) derivative, (C-terminus His-tag) containing <i>lsrR</i> , Ap ^r	This study
pFZY1	<i>galK'-lacZYA</i> transcriptional fusion vector, Ap ^r	[98]
pLsrA14	pFZY1 derivative, containing <i>lsr</i> intergenic region expressed in direction of <i>lsrA</i> , Ap ^r	This study
pLsrR26	pFZY1 derivative, containing <i>lsr</i> intergenic region expressed in direction of <i>lsrR</i> , Ap ^r	This study
Oligonucleotides		
th2lsrRF	TTTGGATCCATGACAATCAACGATTCGGCAATTTTCAG	This study
th2lsrRR	TTTAAGCTTTTAACTACGTAAAATCGCCGCTGCTGTGTC	This study
lsrRP26F	TCCGAATTCTCACTCGTTTGCATATTTCCCC	This study
lsrRP26R	CTCGGATCCCGTTGATTGTCATAATTCATTCTTCACT	This study
lsrAP14F	TCCGAATTCTCGTTGATTGTCATAATTCATTCTTAC	This study
lsrAP14R	CTCGGATCCTCACTCGTTTGCATATTTCCCC	This study
yegEPF	AATGGATCCGATAATGACGCACTGATGAGACTGATCAGTCAT	This study
yegEPR	TTTGTCGACGGCAATTAATACATGCTGTGATTGTTTGCTCAT	This study
udkPF	AATGGATCCGGCAATTAATACATGCTGTGATTGTTTGCTCAT	This study
udkPR	TTTGTCGACGATAATGACGCACTGATGAGACTGATCAGTCAT	This study
lsrRHP1	ATCAACGATTCGGCAATTTTCAGAACAGGGAATGT <u>GTGTAGGCTGGAGCTGCTT</u>	[27]
lsrRHP2	GTAACCAGTGCGTTGATATAACCGCCTTTCATTGCATATGAATATCCTCCTTAG	[27]
lsrKHP1	AGGCATTGTTTTATATAACAATGAAGGAACACCGTGTAGGCTGGAGCTGCTTC	[27]
lsrKHP2	TCCGCCTGCAAAGACTAACGATGAAGGATGAATAGTCGAATTCCGGGGATCCG	[27]

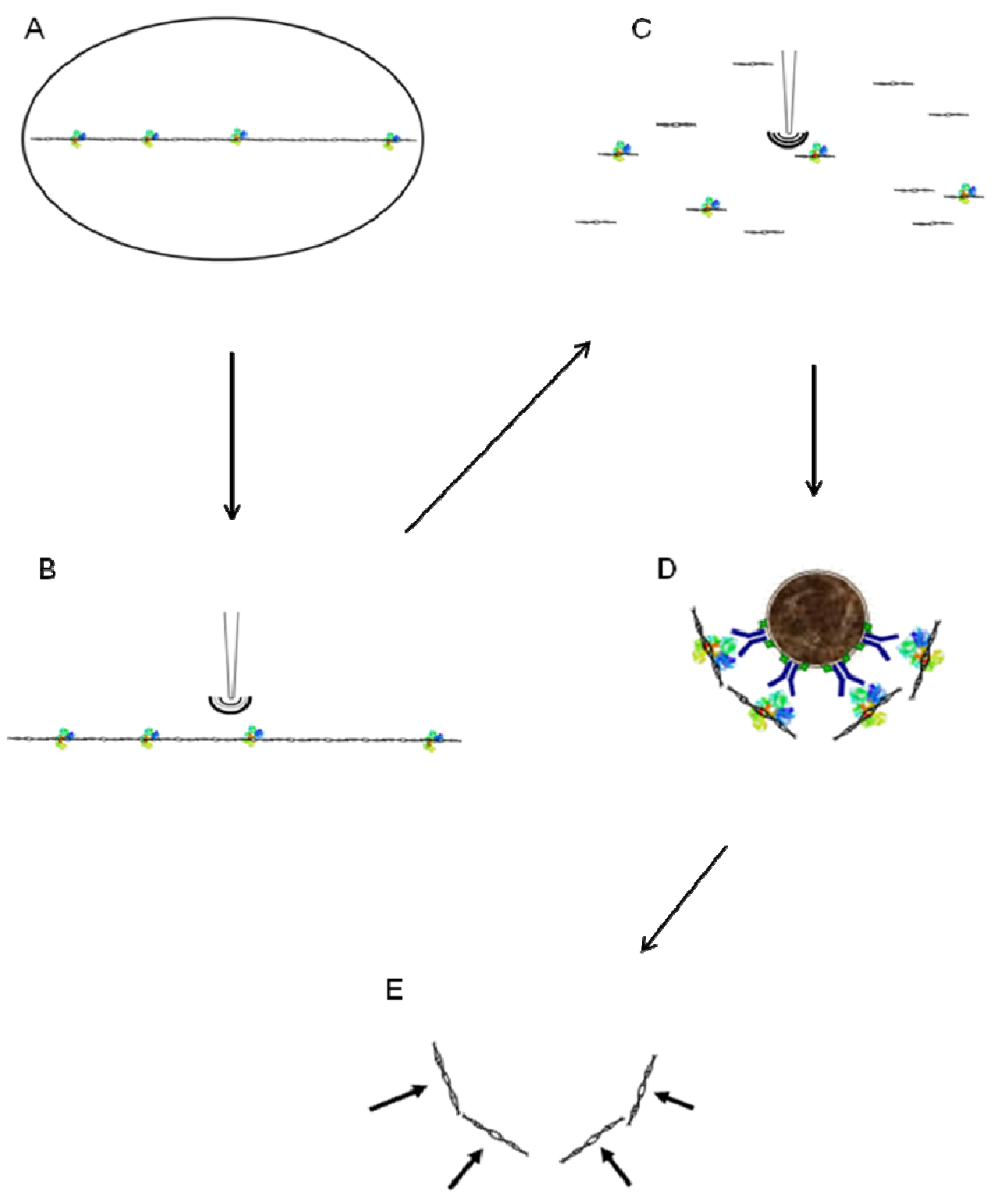
Underlined oligonucleotide segments anneal to template plasmids, while remaining sequences anneal to flanking region

Chromosomal deletions of *lsrR* and *lsrK*. *E. coli* strains, plasmids, and oligonucleotides used in this study are listed in Table 3-1. The one-step replacement method described by Datsenko and Wanner was used to construct CB13 (*E. coli* MG1655 Δ *lsrR* Δ *lsrK*). The phage Red recombination system was first used to construct CB11 (*E. coli* MG1655 Δ *lsrR*) by replacing *lsrR* with an *lsrR::cat* PCR fragment. PCR fragments were amplified using *lsrRHP1* and *lsrRHP2* using pKD3 as template. The PCR products were purified, incubated with DpnI and introduced by electroporation into *E. coli* MG1655 bearing the Red recombinase helper plasmid pKD46. Recombinants were incubated at 37°C and selected on LB-agar supplemented with chloramphenicol. Colonies were grown at 37°C to cure the Red recombinase plasmid and the *cat* insert was verified by PCR. Once verified, pKD46 was re-introduced into CB11 and the phage Red system was repeated in order to replace *lsrK* with an *lsrK::kan* segment generated by PCR with primers *lsrKHP1* and *lsrKHP2* and template plasmid pKD13. After selection on LB-agar plates supplemented with chloramphenicol and kanamycin and curing of the Red recombinase plasmid, the deletion of both *lsrR* and *lsrK* was verified by PCR tests.

Chromatin immunoprecipitation. *E. coli* strain CB13 was transformed with pTH2*lsrR* and grown as previously described. Chromatin immunoprecipitation (ChIP) was conducted as previously described[99] with some variations. Briefly, 1L cells were grown to desired OD and cross-linked with 1% freshly prepared formaldehyde for 25 minutes at room temperature, after which 125mM glycine was added in order to quench the fixation reaction. Fixed cells were harvested by

Figure 3-1. ChIP-Chip process.

(A) Cells are grown under desired conditions and fixed with formaldehyde, covalently binding DNA to proteins. (B) Cells are lysed and lysate is sonicated, shearing DNA (with proteins attached) into 300-1000 bp segments (C). (D) Protein-DNA complexes are precipitated using Protein G-coated magnetic beads with anti-histidine antibody. (E) IP is processed to remove beads, antibody, and protein of interest, resulting in DNA segments corresponding to protein-bound regions [58-59].



centrifugation and washed twice with 50 mL ice-cold TBS. Washed cells were resuspended in 5 mL IP buffer composed of 20mM Tris (pH 7.4), 200mM KCl, 1mM EDTA, protease inhibitor cocktail (Sigma-Aldrich, St. Louis, MO, US). Cell suspension was separated into 1 mL aliquots and sonicated 4 x 20 seconds on ice in order to lyse cells and shear chromatin to an average size of 500-600 bp (see Section 3.6, Supplemental Data). Cell debris was separated by centrifugation at 20,000 ref for 15 minutes at 4°C and supernatant was combined. To immunoprecipitate LsrR-DNA complexes, ChIP-grade 6x His tag antibodies (Abcam, Cambridge, MA, US) were added and incubated overnight at 4°C on a rotator. Two separate mock-IPs were conducted under the same conditions using anti-rabbit IgG antibody and sterile Rabbit serum (Abcam). 50µl of Protein G magnetic beads (Thermo-Fisher, Rockford, IL, US) was added to each solution and rotated for 4 hours at 4°C, after which the beads were washed and protein-DNA complexes were eluted and purified as previously described [50]. Five biological replicate positive chromatin-IP and mock-IP samples were collected and used for array hybridization.

Quantitative PCR. Differences in quantity of immunoprecipitated versus mock-IP DNA were measured by qPCR using a SmartCycler System (Cepheid, Sunnyvale, CA, US) and iQ SYBR Green Supermix (Bio-Rad). All samples were evaluated in triplicate, and threshold cycle (Ct) values were calculated using SmartCycler 2.0 software. Ct values were normalized between samples using three random sites inside coding regions in the *E. coli* genome as the reference set against which two suspected binding sites in the *lsr* intergenic region samples were compared.

DNA amplification, labeling and microarray hybridization. IP (test) and mock-IP (background) samples were amplified using a GenomePlex Whole Genome Amplification (WGA2) Kit (Sigma-Aldrich) and purified using QIAquick PCR purification kit (Qiagen). Test and background samples were then labeled and hybridized onto *E. coli* K12 MG1655 whole genome tiling arrays following manufacturer protocols (Nimblegen, Madison, WI, US). Arrays were scanned with a G2505C DNA microarray scanner at 2 micron resolution (Agilent, Santa Clara, CA, US). Arrays were repeated five times using replicate immunoprecipitation sets as previously discussed.

Data Analysis. Array data were processed using NimbleScan v2.4 software and visualized on SignalMap v1.9 (Nimblegen). Peak locations were selected using a false discovery rate (FDR) of 0.01 resulting in a 99% probability of valid binding site identification. Additionally, to further increase accuracy of binding site locations and to eliminate false positives due to background, peak data were manually evaluated based upon \log_2 value, width, shape, and location relative to transcriptional start site. Peaks that were not within the theoretical width estimate of 500-2500 bp, displayed erratic or noisy curve shape, or showed less than 4-fold enrichment over mock-IP were disregarded.

β -galactosidase activity assays. To determine the effect of LsrR on expression levels, pFZY1-derived plasmids (Table 3-1) were transformed into both *E. coli* K12MG1655 and CB11 strains. Cultures of *E. coli* were grown overnight in LB using

appropriate antibiotic, diluted 100-fold into fresh media, and grown to mid-exponential phase (OD 0.8). To determine β -galactosidase activity cells were processed in triplicate in 96-well microplates as previously described[40]. A_{420} , A_{550} and A_{600} values were collected on a Synergy HT Multi-Mode microplate reader (Biotek Instruments, VT, US).

Computer Modeling of LsrR

Secondary structure analysis for LsrR from *E. coli* K12MG1655 was retrieved from UniProt[100]. The amino acid sequence for K12MG1655 (NCBI Gene ID: 946070) was submitted to the I-TASSER tertiary protein structure prediction server, Center for Computational Medicine and Bioinformatics at the University of Michigan (<http://zhanglab.ccmb.med.umich.edu/I-TASSER>). The resulting data were further analyzed using Jmol (<http://www.jmol.org/>).

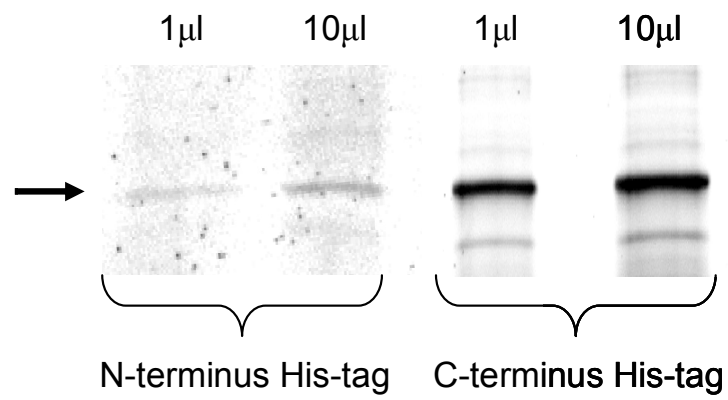
3.4 Results

LsrR shows limited binding throughout *E. coli* K12 MG1655 genome. Initial experiments with pTHlsrR (N-terminus His-tagged LsrR construct) yielded low quantities of purified protein and poor *in vivo* immunoprecipitation results (Figure 3-1A,C). A common challenge with DNA-binding proteins, we found the majority (>95%) of overproduced LsrR remained in inclusion bodies during expression and purification; denaturing and refolding LsrR did not improve *in vitro* activity (data not shown). Tertiary structure predictions and computer modeling analysis of LsrR

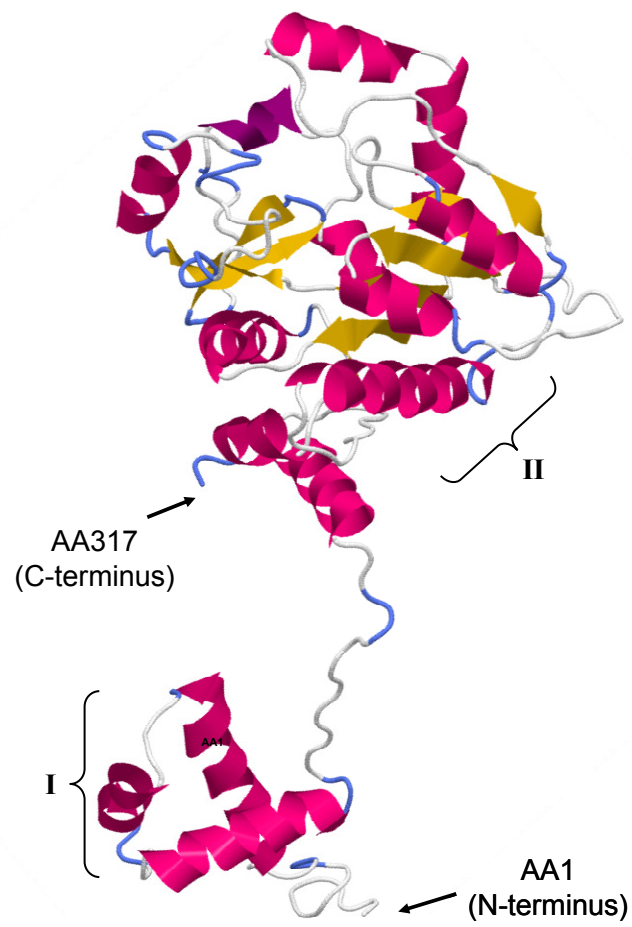
Figure 3-2. Purification and computer modeling analysis of LsrR in *E. coli*.

Recombinant LsrR was constructed by inserting *lsrR* from K12MG1655 into separate vector encoding an N-terminus or C-terminus His-tag and purified using Ni-NTA beads. (A) Parallel protein purification eluates of each recombinant LsrR and PAGE analysis using 0.1 μ l, 1 μ l, and 10 μ l aliquots show significantly higher yields of LsrR with the C-terminus His-tag. (B) Computer modeling and tertiary structure analysis of LsrR revealed the putative helix-turn-helix DNA-binding site (region I) and sugar binding domain (region II). The close proximity of the N-terminus His-tag (immediately adjacent to AA1) to DNA-binding site I indicates possible inaccessibility of the His-tag during DNA binding. (C) qPCR results of immunoprecipitation of *lsr* intergenic DNA for C-terminus and N-terminus His-tagged LsrR versus randomly selected negative controls *ydiV*, *yjeP*, and *ftsQ*. Values represent fold difference in quantity of immunoprecipitated DNA in positive versus mock-IP samples.

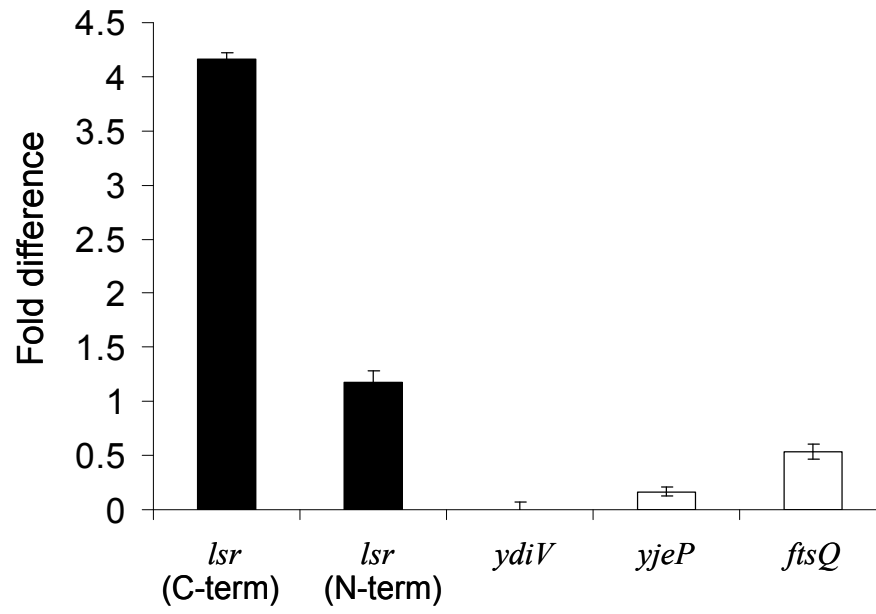
A



B



C



indicate two likely DNA binding regions (Figure 3-1B). Region I, identified by I-TASSER as a likely helix-turn-helix DNA binding region, is directly adjacent to the N-terminus His-tag encoded by pTHlsrR. Region II, closer to the C-terminus for LsrR, is a predicted ligand binding site, and is a structural analog to the effector binding domain for members of the SorC family of transcriptional regulators. Given the difficulty of purifying the N-terminus His-tagged LsrR and the poor immunoprecipitation results as well as the proximity of the His-tag to binding region I, a high-copy inducible vector encoding a C-terminus 6xHis-tag encoding *lsrR* was constructed and provided much better results with immunoprecipitation *in vivo* as well as significant increases in purified protein (Figure 3-1A,C); however, *in vitro* activity was still undetectable when purified recombinant LsrR was used in electrophoretic mobility shift assays (EMSA) (data not shown).

Increased yields of DNA during immunoprecipitation with the C-terminus His-LsrR permitted ChIP-Chip in conjunction with Nimblegen microarrays to be used for determining direct LsrR binding sites. Figure 3-1C depicts quantitative PCR results demonstrating the relative amounts of *lsr* intergenic DNA present after immunoprecipitation using C-terminus and N-terminus his-tagged LsrR as compared to randomly selected negative control regions in *ydiV*, *yjePI*, and *ftsQ*. Relative enrichment values clearly show an increased capacity for DNA binding and capture using the C-terminus LsrR. Microarray data indicate two putative binding sites for LsrR with at least 8-fold enrichment over mock-IP DNA, the *lsrR-lsrA* and *yegE-udk* intergenic regions, as well as two other sites with at least 4-fold enrichment, *mppA* and *yihF* (Figure 3-2B-E). For each individual peak represented in Figures 3-2B-E,

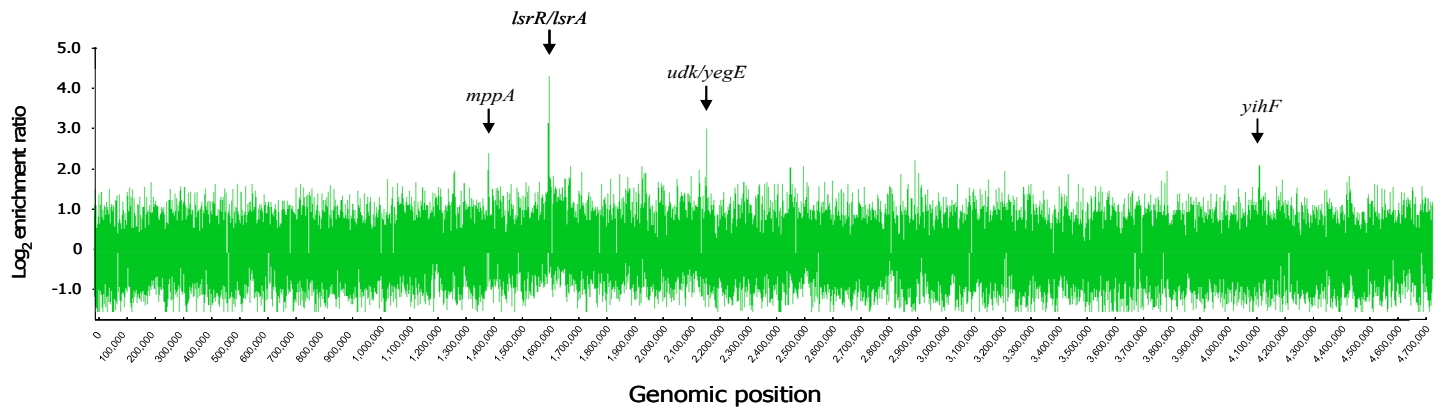
binding regions are localized at the approximate center of the normal bell-shaped distribution of ratios.

Each distribution consists of microarray probes representing 50-bp, overlapping sections of the *E. coli* K12MG1655 genome and curve shapes represent binding of immunoprecipitation products of 500-1000 bps. The ChIP-Chip process and peak curve analysis were further described by Waldminghaus and Skarstad [50]. The artifact in Figure 3-2B showing a sharp drop in enrichment ratio on the left of the distribution is a result of the deletion of *lsrR*. Quantitative PCR was used to validate array data after immunoprecipitation using two sets of primers corresponding to the indicated binding region and three additional sets matching random sites in the *E. coli* genome. Figure 3-2F shows relative quantities of DNA present after immunoprecipitation indicating increased levels of precipitated DNA fragments corresponding to indicated LsrR binding sites *lsrA-lsrR*, *yegE-udk*, *mppA*, and *yihF* (black bars) as compared to negative controls (white bars). Next, ChIP-Chip data were manually evaluated for peak magnitude, curve shape, and width. Peaks that did not fall within the theoretical width projection of 500-2500 bp, displayed erratic or noisy curve shape, or showed less than 4-fold enrichment over mock-IP were disregarded.

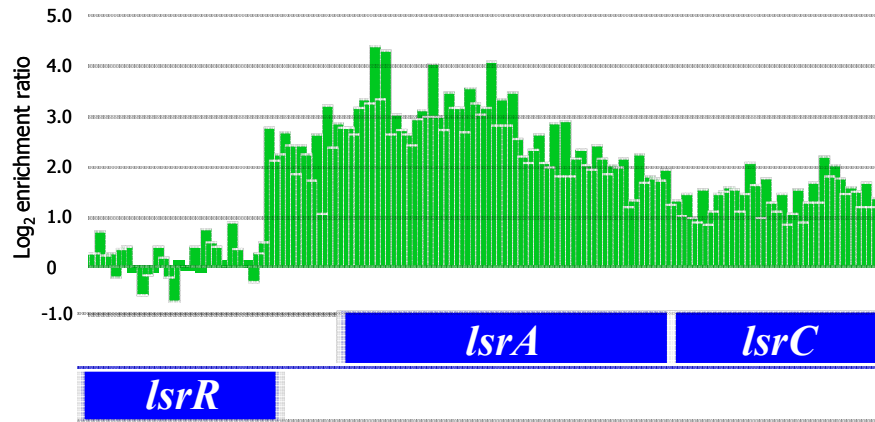
Figure 3-3. ChIP-Chip in *E. coli* K12MG1655.

Genomic distribution (A) and localized probe ratios (B-E) of putative LsrR binding regions. Annotated bars below localized sections display regional gene locations with direct and complementary genes located above and below the mid-line, respectively. Peak values represent the \log_2 ratio of control vs. mock-IP DNA for each individual probe. See discussion for more details. (F) qPCR comparison of ChIP product measuring indicated peak regions (black) compared to randomly selected negative controls (white) showing significantly greater precipitation of *lsr* intergenic DNA using the C-terminus his-tagged LsrR when compared to the N-terminus his-tagged LsrR and randomly selected negative controls *ydiV*, *yjeP*, and *ftsQ*. Values represent fold difference in quantity of immunoprecipitated DNA in positive versus mock-IP samples.

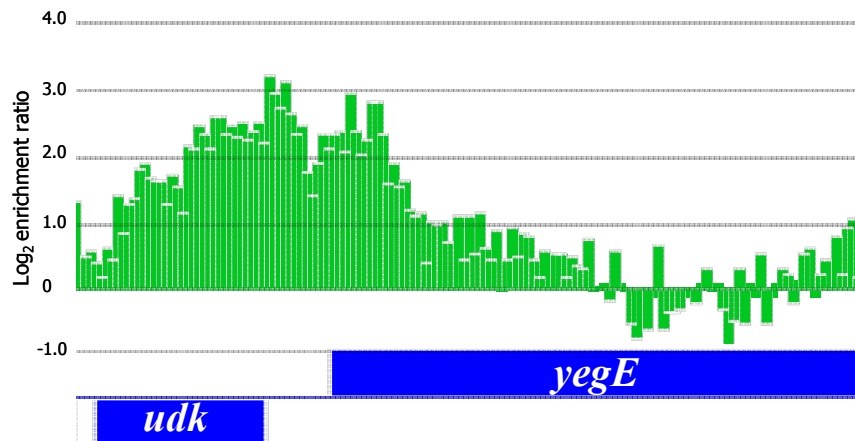
A



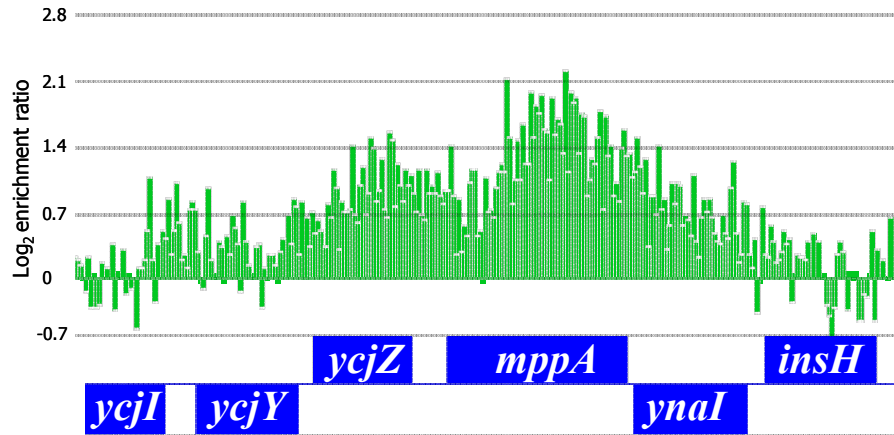
B



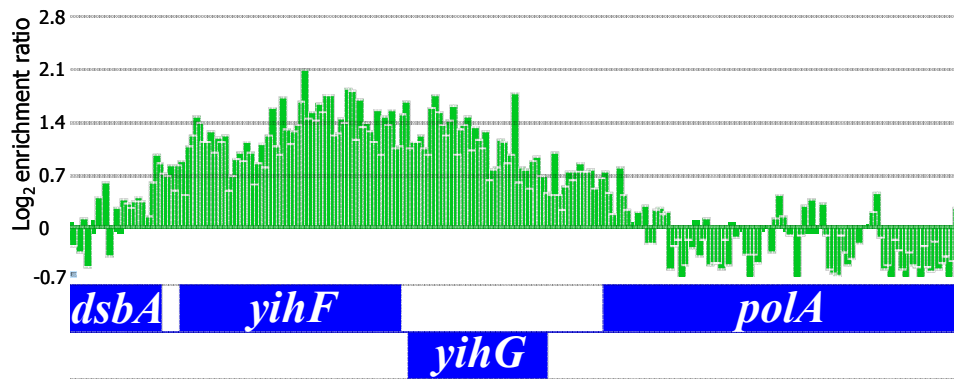
C



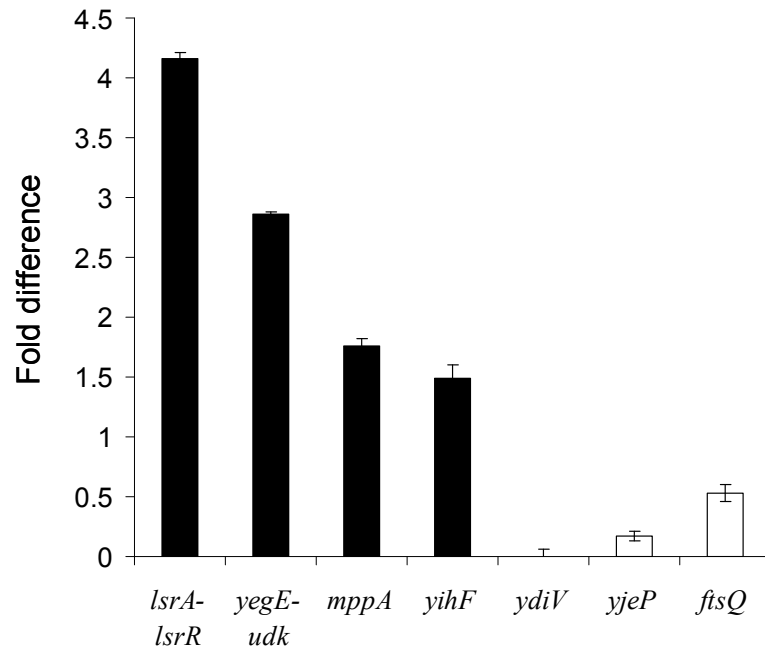
D



E



F



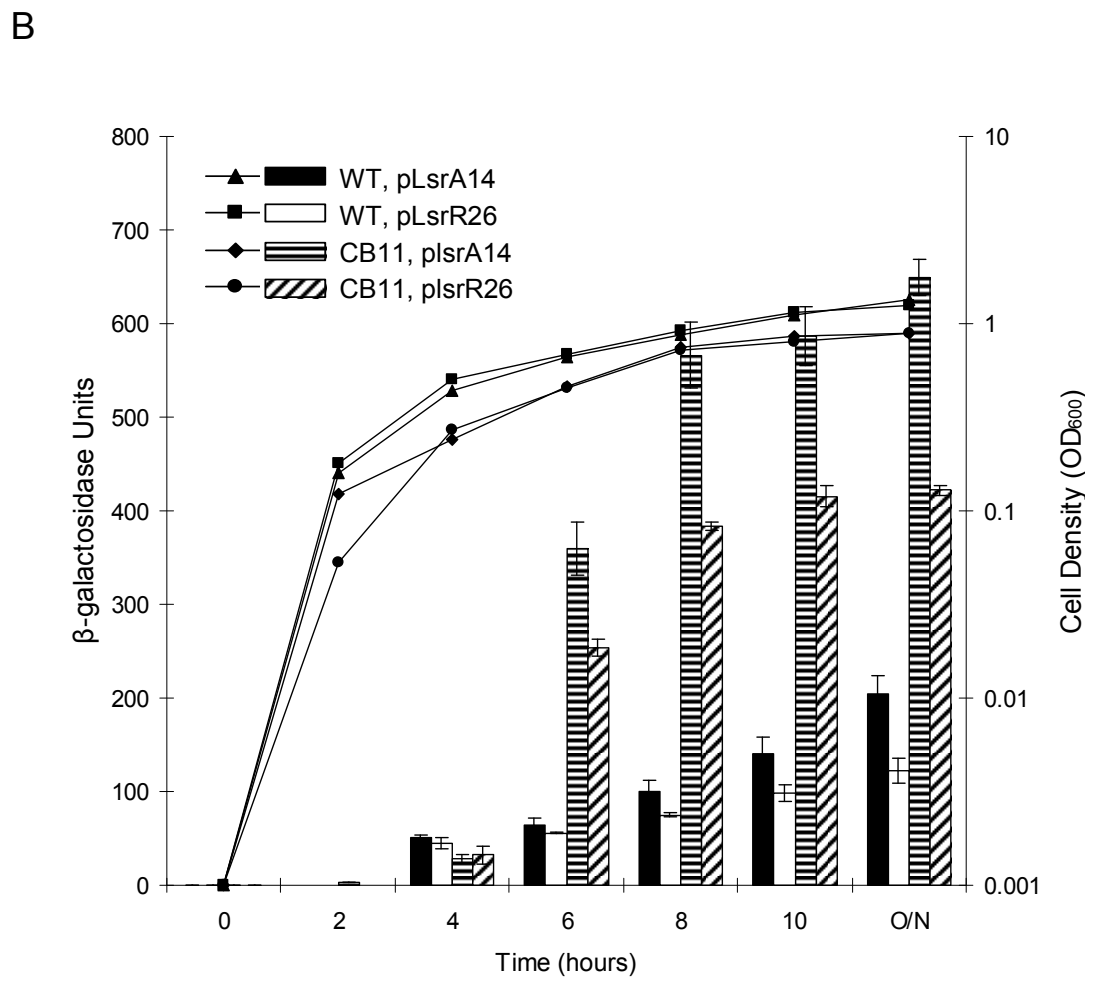
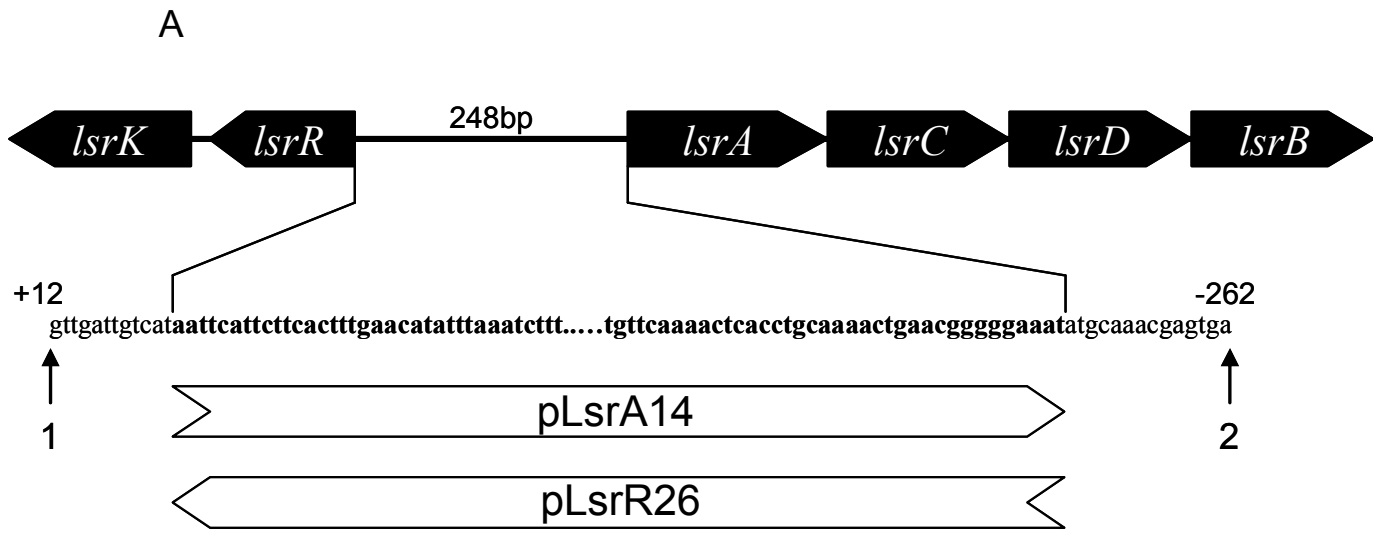
Transcriptional expression bias exists towards the *lsr* operon. Our ChIP-Chip data indicate that LsrR has several putative regulatory sites throughout the *E. coli* genome; however, in order to begin examining these novel sites more understanding about specific LsrR regulatory activity is required. The divergent structure of *lsr* regulon adds complexity when examining LsrR activity and while the effect of LsrR on expression levels in both directions has been previously measured [20, 27], promoter effects independent of LsrR have not been shown. To further explore directionality in the *lsr* intergenic region, we isolated and inserted the promoter segment in opposite directions into pFZY1 plasmids to evaluate the effect of LsrR on expression of *lacZ* in wild-type *E. coli* K12MG1655 and the *lsrR* mutant (CB11) using β -galactosidase reporter assays (Figure 3-3A).

Growth curves (OD_{600}) and *lacZ* expression levels using pLsrA14 and pLsrR26 in WT and CB11 strains are shown in Figure 3-3B and indicate obvious repression effects of LsrR were in both directions (cross hatched bars). These variations are further illustrated in Figures 3-3C and 3-3D. As expected, QS gene expression is observed to be switched on in the late exponential phase. In Figure 3-3C the effects of LsrR derepression result in an up to 5.5-fold increase in expression in CB11.

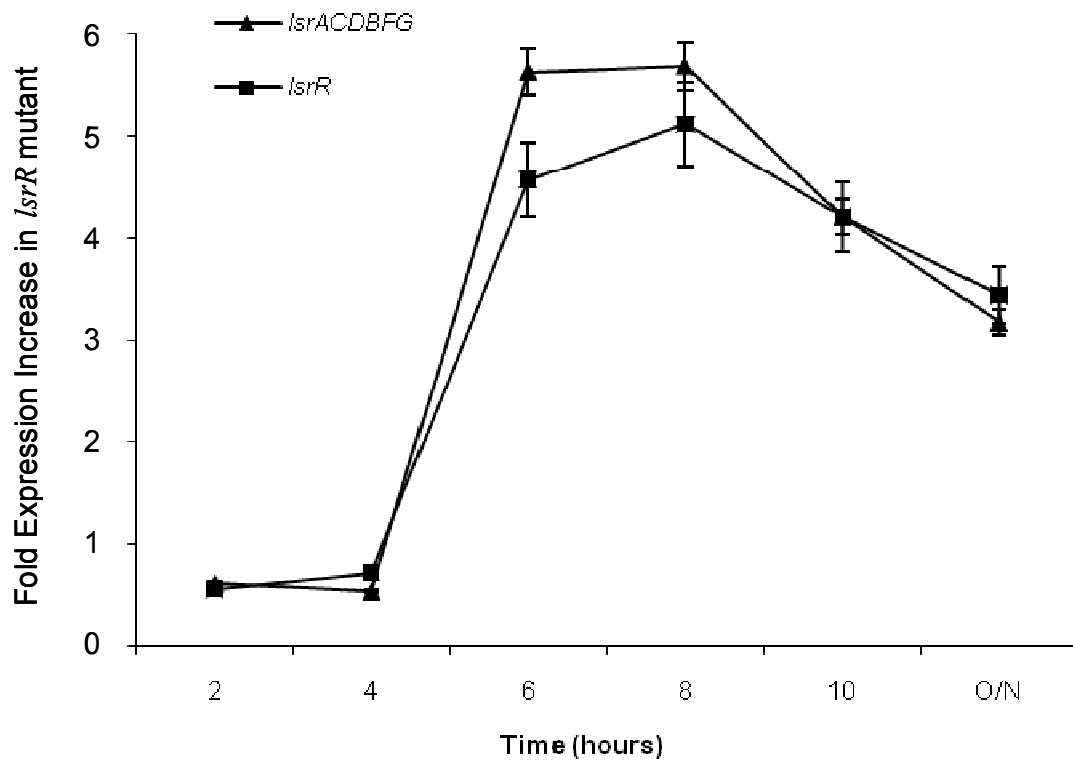
Also, it was apparent that either with or without the presence of LsrR, gene expression in the direction of *lsrA* was always greater than that in the direction of *lsrR*. We refer to this as promoter bias in the direction of *lsrA*. Figure 3-3D illustrates this difference in expression levels in the direction of *lsrA* (pLsrA14) versus *lsrR* (pLsrR26) over time, showing a steady increase in expression difference, from .4-fold

Figure 3-4. Promoter region orientation and directional expression analysis of the *E. coli* *lsrA-lsrR* intergenic region.

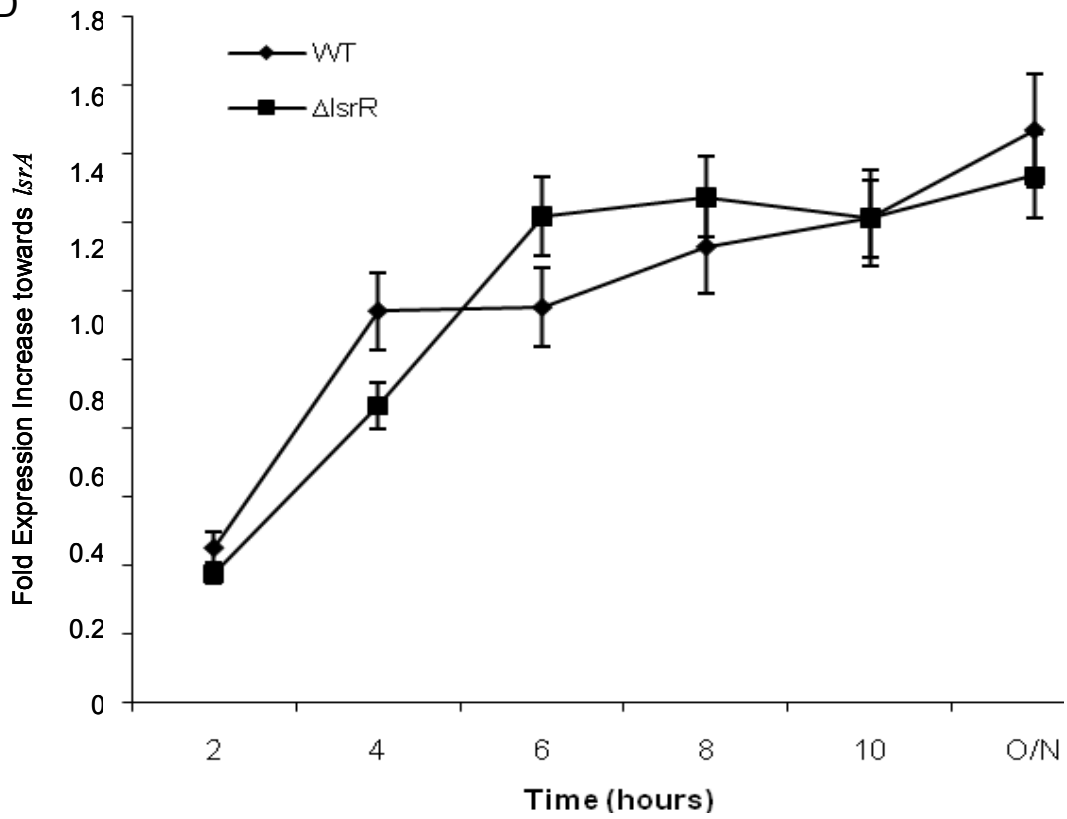
The *E. coli* K12MG1655 intergenic region (from -262 to +12 relative to the *lsrR* start site) between *lsrA* and *lsrR* was inserted in opposite directions into the MCS of the low-copy mini-F derivative pFZY1. (A) pLsrA14 and pLsrR26 express *lacZ* using the *lsr* promoter region oriented toward *lsrA* and *lsrR*, respectively. (B) *E. coli* strains K12MG1655 wild-type (WT) and K12MG1655 Δ *lsrR* (CB11) each carrying *lacZ* expression plasmids pLsrA14 and pLsrR26. Cultures were grown at 37°C in LB and samples were collected at time points shown from 0-10 hours and overnight for measuring cell density (connected lines) and β -galactosidase activity (bars). (C) Fold increase in expression levels in an *lsrR* mutant using pLsrA14 (triangles) and pLsrR26 (squares). Point values represent average expression difference between CB11 and WT strains for each time point (\pm 5% error). (D) Fold increase in expression bias toward the *lsr* operon in CB11 (squares) and WT (diamonds).



C



D



greater expression towards *lsrA* in early exponential growth (4 hours) to 1.6-fold greater during stationary phase (overnight). The trend showing a steadily growing difference in expression of pLsrA14 versus pLsrR26 in both the WT and *lsrR* indicates that the expression bias toward *lsrA* is inherent to the *lsr* intergenic region and is not influenced by LsrR.

3.5 Discussion

In this study we investigated genomic binding sites for LsrR, as well as divergent transcriptional activity within the *lsr* regulon. Although some investigation has been performed on the binding activity and expression results of LsrR in *E. coli* [22, 27], very little is known about the molecular kinetics of the QS regulator such as exact binding cues, interactions with secondary proteins or activators, or even whether it exists and functions in monomeric or multimeric form *in vivo*. This may be demonstrated by the greatly reduced purification and *in vivo* activity found with an N-terminus as compared to C-terminus His-LsrR (Figure 3-1A,C) and our subsequent difficulty with *in vitro* EMSA. While LsrR was shown to be antagonized by the phosphorylated AI-2 molecule [22, 27] and extracellular AI-2 phosphorylation reduced the QS response [12], the exact DNA binding cue for the QS regulator has not yet been determined and LsrR activation may be based upon secondary protein interactions, unphosphorylated AI-2 availability [21], or other activators [19, 27, 101].

lsrR belongs to the *sorC* transcriptional regulatory family which has an N-terminus helix-turn-helix DNA binding region and is regulated by a putative sugar

binding domain [100, 102]. In addition, both SorC and DeoR, another transcriptional regulator with structural similarities to LsrR, are biologically active in tetrameric form with DNA binding domains which extend divergently from the center, suggesting operator binding in at least two distinct regions [53, 66, 103, 104]. Recent crystallographic analysis indicated the presence of two molecules per asymmetric unit for LsrR, and quaternary structure for DNA-binding proteins with helix-turn-helix domains is common [52, 66]. However, given the domain similarities to the *sorC* family and the concentration-dependent transition from dimer to tetramer states noted during crystallography [54, 105], it is probable that LsrR is not only activated prior to binding DNA but functions *in vivo* as a tetramer. Still, further work is needed to confirm the *in vivo* structure and binding kinetics of the QS regulator.

The ChIP-Chip process is inherently vulnerable to background noise, generating false-positive or false-negative peaks as a result of several factors including non-specific antibody binding, incomplete reversion of crosslinks, or insufficient RNase treatment, among others [50]. To maximize signal-to-noise ratios, ChIP protocols used here were identical for both the test and background samples, differing only in the type of antibody added, and both samples were amplified using whole genome amplification (WGA). This method reduces error due to non-specific antibody-protein binding, which we found to be the major contributor to false-positive microarray results, as well as reduces potential loci amplification bias. Also, LsrR is antagonized by phospho-AI-2, causing the regulator to unbind DNA [22, 27, 44]. Clearly having the potential to affect ChIP by reducing LsrR binding persistence, we eliminated the intracellular availability of phospho-AI-2 by mutating

the AI-2 kinase, *lsrK*. Additionally, in order to remove the possibility of native LsrR outcompeting the His-tagged LsrR *in vivo*, we also mutated *lsrR* in K12MG1655.

While the *lsr* intergenic region has been documented for direct LsrR binding and regulation in *E. coli* [19, 22, 27], the three other putative LsrR binding sites indicated in Figure 2A have not been shown. The uridine/cytidine kinase *udk* and the predicted diguanylate cyclase *yegE* (Figure 2C) are also expressed from a divergent promoter similar to that found in the *lsr* region. In *E. coli*, *udk* mutants were shown to accumulate the endogenous inducer cytidine, while *yegE* has been directly linked to the transition from motility to sessile behavior [36, 106, 107]. Studies have clearly shown the linkage of the QS response to biofilm formation and motility [21, 37, 94, 108] and given the metabolic and phenotypic implications of the *udk-yegE* divergent gene set, it is feasible that they are regulated in conjunction with the bacterial QS system through LsrR. In order to build an exact DNA binding motif for LsrR we compared the *lsr* and *udk-yegE* intergenic regions using MEME motif discovery tool and previously indicated LsrR binding sites [22, 109]; while attractive hypotheses for binding locations could be made, given the limited sample size and lack of apparent motif homology we could not consider similarities significant (data not shown).

The murein tripeptide transporter subunit *mppA* also has potential linkage to the QS system due to its functional binding to heme, a primary iron source for Gram-negative bacteria [110, 111]. During the process of infection, iron levels are directly connected to formation, and possibly regulation, of virulence factors and toxins in bacteria [112, 113] and pathogenicity has also been directly linked to the QS process and cellular communication [13, 94, 114]. The uncharacterized hypothetical protein,

yihF, requires further functional analysis in order to evaluate its potential role in the QS process. Although *mppA* and *yihF* yielded lower relative enrichment versus mock-IP than the *lsrA-lsrR* and *udk-yegE* regions, the lack of data regarding exact LsrR binding activators, potential co-regulators or *in vivo* form preclude the *mppA* and *yihF* sites from being disregarded. Therefore, these regions await additional study to confirm direct LsrR regulation.

A more complete understanding of the regulatory effect of LsrR and structure of the complex *lsr* regulon would greatly contribute to further examination of the additional putative LsrR regulatory sites indicated during ChIP-Chip. Therefore, we focused on divergent expression in the *lsr* regulon examining the effect of LsrR as well as promoter strength in the intergenic region. Previous studies of LsrR from *E. coli* indicate potential binding sites for the regulator in the *lsr* operon [22] and were shown *in vivo* to have a direct effect on expression of the *lsr* operon under normal conditions [27]. In order to further analyze *in vivo* LsrR binding and regulation in the *lsr* intergenic region, we used β -galactosidase activity assays with reporter plasmids containing the *lsr* intergenic promoter region, oriented in opposite directions, in *E. coli* K12MG1655 and K12MG1655 Δ *lsrR* (CB11) (Figure 3-3A). Our results indicate that LsrR has a comparable bi-directional regulatory effect in the *lsr* regulon (Figure 3-3C). That is, LsrR regulates expression in both the *lsrA* and *lsrR* directions by direct interaction with the *lsr* intergenic region, and has similar impact on divergent expression regardless of promoter strength.

However, we also noted that the intergenic promoter region yielded an expression bias towards the *lsr* operon regardless of stage of growth. Figure 3-3D

demonstrates the steadily increasing expression level differences when using pLsrA14 and pLsrR26 in both the WT and CB11 strains. This relative linearity in percent increase of expression suggests a stronger innate promoter region for *lsrA* than for *lsrR*, but similar increases in expression measured in both WT and CB11 strains indicate that LsrR binding and repression is not affected by the stronger promoter. The natural bias also indicates that the AI-2 transport protein LsrACDB may be more available intracellularly than LsrR throughout the cell cycle providing an early and constant supply of AI-2 and increased sensitivity to exogenous AI-2. Also, we postulate that the AI-2 kinase *lsrK*, which is immediately downstream of *lsrR*, is either expressed at a lower level than the *lsr* operon and is still metabolically effective, is a limiting factor in the phosphorylation of available AI-2 and therefore essentially a QS regulator itself, or has its own promoter as previously suggested [20] and is unaffected by the *lsr* intergenic directional bias. Our current work is exploring further the binding dynamics of the QS repressor LsrR in this operon.

In summary, we have shown that while QS in bacteria has widespread genetic and phenotypic implications, LsrR in *E. coli* MG1655 appears under normal conditions to directly associate with only four binding sites. However, three of these indicated sites, *lsrR-lsrACDBFG*, *yegE-udk*, and *mppA* have direct potential ties to the QS process and require further analysis. Since genomic binding analysis assays such as ChIP-Chip or ChIP-Seq are dependent upon multiple preparatory factors which may introduce error into the resulting data including fixation, DNA shearing, antibody specificity, and non-specific binding, we maintain that there may be other sites for LsrR regulation by direct binding. These are possibly indicated with our

data, but might also be sufficiently transient or weak to provide positive identification using the tools as described here. That is, given the dynamic nature of bacterial QS, numerous possible internal and environmental cues may affect the *lsr* system and subsequently LsrR binding events. These considerations, therefore, lead us to use these results as primary indicators of regions in which to begin further analysis. Our current efforts continue work on binding specificity and activity for LsrR in the *lsr* intergenic region and other potential binding sites.

3.6 Supplemental Data

Chromatin Shearing and Whole Genome Amplification (WGA). The ideal size for sheared chromatin to be applied to microarrays for ChIP-Chip analysis is 400-1000 bp [50, 58, 59]. Excessive shearing may cause protein-DNA dissociation, DNA over-fragmentation, or denaturing, and under-fragmentation will yield fragments too large for the amplification process [50]. In order to ensure our samples were within the ideal size range and usable for proper hybridization, various cultures of *E. coli* K12 MG1655 were grown to stationary phase and harvested via centrifugation for resuspension and shearing. Shearing was conducted using a Misonix Microson XL2000 sonicator. Conditions evaluated during sonication were: culture resuspension density, culture resuspension volume, time of sonication, and sonication intensity.

Cultures were resuspended in volumes of resuspension buffer (50 mM Tris-Cl, pH 8.0, 10 mM EDTA) corresponding to ratios of original culture. That is, three final cell density samples were tested with 1:1000, 1:100 and 1:10 ratios of resuspension buffer to original culture volume. For each sample, 5 ml of solution was sonicated for 10x30 seconds in an ice bath at 50% power and DNA was collected using a QIAquick PCR purification kit (Qiagen). Agarose gel analysis shows no detectable size difference in DNA between samples with varying cell density (data not shown).

Resuspended cultures were then separated into 10ml, 5ml, and 1ml samples and again sonicated for 10x30 seconds in an ice bath at 50% power. DNA was again collected as described above and analyzed using agarose gel electrophoresis. Results

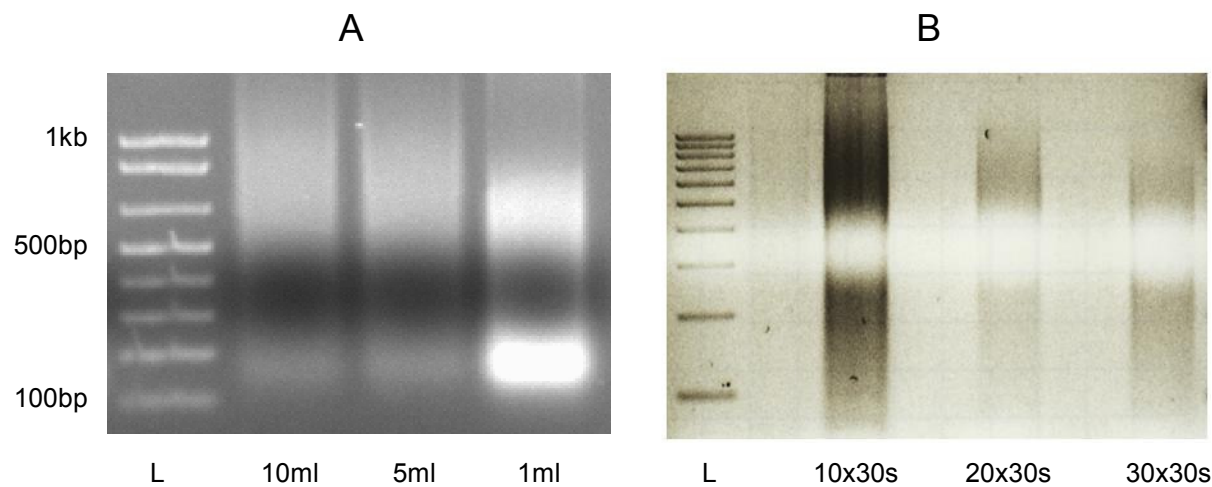


Figure 3-5. Chromatin shearing.

(A) 10ml, 5ml, and 1ml samples of resuspended cell culture showing difference in shearing of DNA as a result of sample volume. 1ml sample also demonstrates overshearing of chromatin sample. (B) 1 ml samples of resuspended cell culture sonicated at 50% power for 10x30, 20x30, and 30x30 seconds. Sample sonicated for 20x30s displays best size range without overshearing. ‘L’ in both Figures is 1kb Plus DNA Ladder (Invitrogen). Band causing loss of image saturation in approximate center of each image is due to absorption by running marker.

demonstrate significant difference in 1ml sample size versus 5ml and 10ml sets (Figure 3-4A), indicating a much greater dependence upon sample volume than sample density for shearing effectiveness. In order to assess sonication time and intensity simultaneously, 1 ml samples of resuspended culture were sonicated at 100% and 50% power for 10x30, 20x30, and 30x30 second pulses. 50% power shearing produced DNA fragments that were well above the 1000 bp upper limit and were deemed unusable (data not shown). Samples sheared at 100% power for 20x30 seconds yielded DNA fragments that were usable for whole genome amplification (WGA) (Figure 3-4B).

Samples sonicated to within the desirable size range for WGA were purified using a QIAquick PCR purification kit (Qiagen). A GenomePlex Whole Genome Amplification Kit (WGA2, Sigma) was used with various amounts of *E. coli* genomic DNA in order to determine the ideal starting quantity of DNA for the WGA process. 1, 10, and 25 ng samples of DNA were isolated, purified, and amplified using the WGA2 kit (Figure 3-5). Agarose gel analysis of purified WGA samples indicates that a minimum of 10ng of immunoprecipitated DNA would be necessary for amplification via WGA for subsequent microarray analysis.

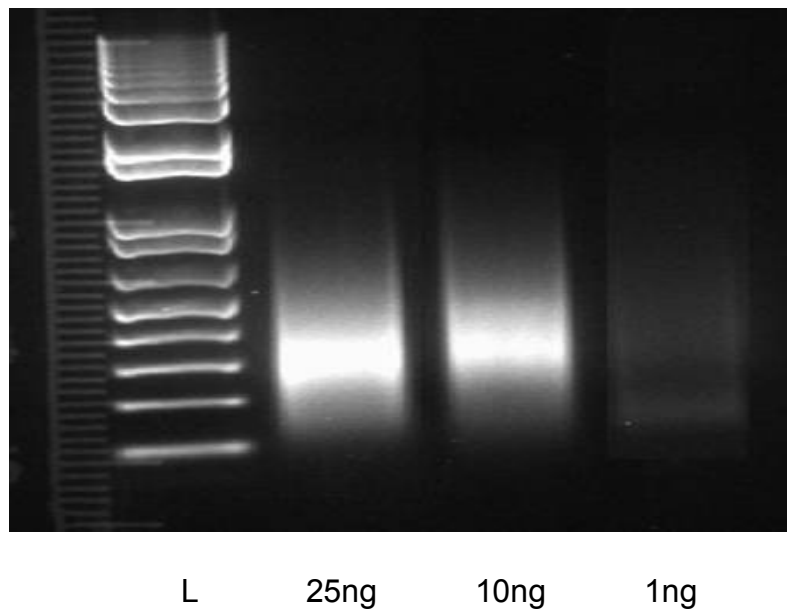


Figure 3-6. Whole Genome Amplification.

25ng, 10ng, and 1ng starting quantities of sheared DNA used in WGA process.

Results indicate a minimum of 10ng of DNA for effective amplification. 'L' is 1kb

Plus DNA Ladder (Invitrogen).

Chapter 4: Expression Regulation in the Divergent Quorum Sensing *lsr* Regulon in *Escherichia coli*

4.1 Abstract

The quorum sensing (QS) process induces shifts in genetic expression as a response to signal molecules termed ‘autoinducers’, queueing diverse phenotypic changes in various species of bacteria. The *lsr* (LuxS-regulated) regulon is negatively regulated by the transcriptional repressor LsrR via a direct response to threshold levels of extracellular autoinducer-2 (AI-2). This autoinducer elicits responses both within and between a multitude of bacterial species. In this study we show that expression in both directions of the divergent *lsr* gene set is dependent upon simultaneous binding of two cyclic-AMP receptor proteins (CRP). We also demonstrate that full *lsr* repression occurs via concurrent LsrR binding to opposite ends of the intergenic region, implying the formation of a looped DNA complex. The previously unreported potential for high expression of *lsrR*, the more stringent LsrR regulation of *lsrR* than *lsrACDB*, and an apparent secondary integration host factor (IHF)-like regulatory mechanism providing additional *lsrR* expression control are also described.

4.2 Introduction

Several quorum sensing (QS) systems have been identified among bacteria which serve as communication mechanisms that induce various phenotypic or genetic shifts in response to local cell density [5, 29, 33]. These systems include acyl-homoserine lactone (AHL) signaling, more common in Gram-negative species, and oligopeptide signaling, found in Gram-positive strains, both of which utilize highly species-specific molecules to draw a population-based response [23, 28, 29, 32]. The autoinducer-2 (AI-2) QS system, however, has been noted for its capacity to elicit marked responses both within and across species by employing a common synthase, LuxS, to form 4,5-dihydroxy-2,3-pentanedione (DPD), the precursor to AI-2 [44]. While *luxS* homologs have been identified in at least 55 species, QS genes and their functional similarities in model organisms such as *V. harveyi*, *S. typhimurium*, and *E. coli* continue to be explored for the exact genetic circuitry and dynamics involved in the AI-2 response [6, 9, 20, 34, 44, 97].

In *E. coli*, AI-2 is produced enzymatically via the methyltransferase conversion of S-adenosylmethionine (SAM) to S-adenosylhomocysteine (SAH), which then interacts with the nucleosidase Pfs to form S-ribosylhomocysteine (SRH). LuxS then converts SRH to DPD, which spontaneously cyclizes forming AI-2 (40, 47). Although any one of several specific structures of AI-2 may result, the signal molecule interconverts rapidly between equilibrating forms permitting detection by various bacteria with homologous QS circuitry [44, 97, 101, 115]. In *E. coli* QS circuitry, AI-2 is synthesized and exported, accumulating extracellularly in proportion with local bacterial populations. AI-2 is then re-imported via the *lsr* (*luxS*-regulated)

ATP binding cassette transporter LsrACDB. After import and phosphorylation by the kinase LsrK, AI-2 increases expression of the *lsr* operon by binding to the QS regulator LsrR, derepressing transcription [19, 27, 44].

While much focus has been placed upon global impacts of the bacterial QS response, mechanistic details of the *lsr* set of genes remain obscure. AI-2 transport, phosphorylation, and binding to LsrR are well documented in *E. coli* and *S. enterica* [44] but only recently have a putative DNA binding site motif and *in vitro* response to phospho-AI-2 been described for LsrR [22]. Current models of the QS circuit assign LsrR the role as sole transcriptional regulator for both the *lsr* and *lsrRK* operons, implying a single point of genetic activation during a QS response [18, 27, 116]. Interestingly, despite the connection of the QS mechanism to responses such as bioluminescence [14, 117], virulence [114, 118], motility [119, 120], and biofilm formation [38, 42], ChIP-Chip analysis with LsrR in *S. typhimurium* and *E. coli* have shown unexpectedly few direct binding sites [49]. This may suggest hierarchical modes of regulation including complex and transient genetic assemblies that may not be revealed in ChIP-Chip assays. That is, genomic binding assay conditions may over- or underrepresent binding activity [50, 59], and understanding the molecular basis for the global impact of AI-2 QS warrants further and more detailed examination of LsrR regulation.

In this study, we focus on transcriptional regulation in the divergent promoter region of the *lsr* and *lsrRK* operons and evaluate the effects of various DNA mutations on putative LsrR and cAMP-reactive protein (CRP) binding sites on genetic expression. We use multiple sets of mutations with bi-directional expression

analysis in order to describe LsrR repression. We assemble our observations in the context of similar repressor systems and present a novel model for LsrR regulation incorporating CRP, DNA looping, and a secondary layer of repression involving an integration host factor (IHF)-like protein as part of the *lsr* regulatory mechanism.

4.3 Materials and Methods

Bacterial strains, plasmids, and growth conditions. All strains and plasmids used in this work are described in Table 4-1. Cultures were grown aerobically at 37°C in LB medium to OD 0.8 or as noted. Media were supplemented with antibiotics at the following concentrations: ampicillin, 50µg ml⁻¹; chloramphenicol 25µg ml⁻¹.

Plasmid construction. Plasmids pLsrR26 and pLsrA14 were described previously (Byrd *et al.*, submitted for publication) and contain the entire native *lsr* intergenic region (-262 to +12 relative to the *lsrR* start site) in opposite orientations. All modified promoter segments were designed and submitted for synthesis by IDTDNA (IA, US). Modified sequences were received in pIDTSmart vectors and transformed into TOP10 chemically competent cells (Invitrogen, CA, US) for amplification. After purification (QiaPrep Spin Miniprep kit, Qiagen) plasmids were digested with BamHI and EcoRI and regions of interest were extracted using agarose gel purification (Gel Purification Kit, Qiagen). pFZY1, a low-copy mini-F derivative with MCS upstream of a promoterless *galK'-lacZYA* reporter segment [98] was also digested with BamHI and EcoRI and gel purified. Ligation and selections were performed using standard methods [121]. All plasmids were validated by DNA sequencing, performed at the

Table 4-1. *E. coli* strains and plasmids used in this study.

Name	Description ^a	Source or Reference
K12 Strains		
MG1655	Wild type	Laboratory stock
CB11	MG1655 Δ <i>lsrR</i> ::Cm	Byrd <i>et al.</i> ^b
Plasmids		
plsrR26	native <i>lsr</i> intergenic region expressed in direction of <i>lsrR</i> , Ap ^r	Byrd <i>et al.</i> ^b
plsrA14	native <i>lsr</i> intergenic region expressed in direction of <i>lsrA</i> , Ap ^r	Byrd <i>et al.</i> ^b
pcrp22	mutation of CRP binding site C1, expressed in direction of <i>lsrR</i> , Ap ^r	This study
pcrp9	mutation of CRP binding site C1, expressed in direction of <i>lsrA</i> , Ap ^r	This study
pcrp23	mutation of CRP binding site C2, expressed in direction of <i>lsrR</i> , Ap ^r	This study
pcrp8	mutation of CRP binding site C2, expressed in direction of <i>lsrA</i> , Ap ^r	This study
pcrp4b	mutation of CRP binding sites C1 and C2, expressed in direction of <i>lsrR</i> , Ap ^r	This study
pcrp10	mutation of CRP binding sites C1 and C2, expressed in direction of <i>lsrA</i> , Ap ^r	This study
plsrR24	mutation of operator sites O1 and O2, expressed in direction of <i>lsrR</i> , Ap ^r	This study
plsrA13	mutation of operator sites O1 and O2, expressed in direction of <i>lsrA</i> , Ap ^r	This study
plsrR27	mutation of operator sites O1 and O3, expressed in direction of <i>lsrR</i> , Ap ^r	This study
plsrA11b	mutation of operator sites O1 and O3, expressed in direction of <i>lsrA</i> , Ap ^r	This study
plsrR33	mutation of operator sites O1 and O4, expressed in direction of <i>lsrR</i> , Ap ^r	This study
plsrA34	mutation of operator sites O1 and O4, expressed in direction of <i>lsrA</i> , Ap ^r	This study
plsrR31	mutation of operator sites O2 and O3, expressed in direction of <i>lsrA</i> , Ap ^r	This study
plsrA32	mutation of operator sites O2 and O3, expressed in direction of <i>lsrR</i> , Ap ^r	This study
plsrR30	mutation of operator sites O2 and O4, expressed in direction of <i>lsrA</i> , Ap ^r	This study
plsrA11	mutation of operator sites O2 and O4, expressed in direction of <i>lsrR</i> , Ap ^r	This study
plsrR25	mutation of operator sites O3 and O4, expressed in direction of <i>lsrA</i> , Ap ^r	This study
plsrA12	mutation of operator sites O3 and O4, expressed in direction of <i>lsrR</i> , Ap ^r	This study
plsrR39	mutation of site P1, expressed in direction of <i>lsrR</i> , Ap ^r	This study
plsrA40	mutation of site P1, expressed in direction of <i>lsrA</i> , Ap ^r	This study
plsrR37	mutation of site P1 and O1, expressed in direction of <i>lsrR</i> , Ap ^r	This study
plsrA38	mutation of site P1 and O1, expressed in direction of <i>lsrA</i> , Ap ^r	This study
plsrR35	mutation of site P1 and O3, expressed in direction of <i>lsrR</i> , Ap ^r	This study
plsrA36	mutation of site P1 and O3, expressed in direction of <i>lsrA</i> , Ap ^r	This study
plsrR41	mutation of site P2, expressed in direction of <i>lsrA</i> , Ap ^r	This study
plsrA42	mutation of site P2, expressed in direction of <i>lsrR</i> , Ap ^r	This study
plsrR43	mutation of site O2 and P2, expressed in direction of <i>lsrA</i> , Ap ^r	This study
plsrA44	mutation of site O2 and P2, expressed in direction of <i>lsrR</i> , Ap ^r	This study
plsrR45	mutation of site O3 and P2, expressed in direction of <i>lsrA</i> , Ap ^r	This study
plsrA46	mutation of site O3 and P2, expressed in direction of <i>lsrR</i> , Ap ^r	This study

^a All plasmids are derivatives of the promoterless *galk'-lacZYA* fusion vector pFZY1. See Materials and Methods.

^b C. M. Byrd, C. Y. Tsao, W. M. Bentley, submitted for publication

DNA Core Facility of the Institute for Bioscience and Biotechnology Research (IBBR).

β -galactosidase activity assays. All pFZY1-derived plasmids (Table 4-1) were transformed into both *E. coli* K12MG1655 and CB11 strains as described previously [122]. Cultures of *E. coli* were grown overnight in LB using appropriate antibiotic, diluted 100-fold into fresh media, and grown to mid-exponential phase (OD 0.8). Cells were processed in triplicate in 96-well microplates as previously described [40]. A_{420} , A_{550} and A_{600} values were collected on a Synergy HT Multi-Mode microplate reader (Biotek Instruments, VT, US). All expression activity values are in Miller Units (MU).

4.4 Results

Both CRP sites are required for full *lsr* expression. Two CRP consensus sites were previously identified in the *lsr* intergenic promoter region, and the prevention of CRP binding *in vivo* via the addition of glucose demonstrated a precipitous drop in overall expression of the *lsr* operon [20, 27]. In order to examine the impact of each CRP site separately, expression analysis was conducted using reporter plasmids containing directionally oriented *lsr* intergenic promoter regions with mutated CRP binding sequences (annotated as C1 and C2 in Figure 4-1A). Because certain repressors, including LsrR, have been shown also to activate gene expression [21, 123, 124], each set of plasmids were therefore transformed into both *E. coli*

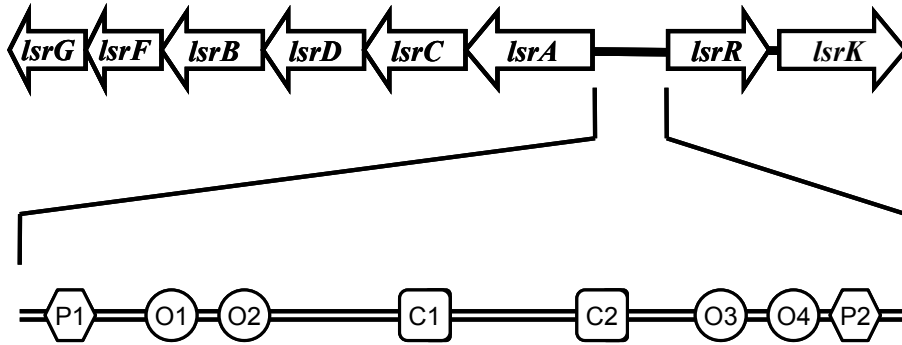
K12MG1655 wild-type (WT) and CB11 ($\Delta lsrR$) strains in order to determine the specific role of the effector, LsrR.

In Figure 4-1B, we list the specific mutations introduced into each site. The mutant form is indicated above that of the wild type sequence. In Figure 4-2, we list the β -galactosidase expression for each vector in the direction indicated. For example, pcrp9 is a vector containing the *lsr* intergenic region with the C1 mutation as the promoter segment for *lacZYA*. The expression levels for pcrp9 measured in the direction of *lsrA* was 6 and 4 MU in the WT and CB11 strains, respectively. Vector pcrp22, containing the same promoter region but oriented in the opposite direction (towards *lsrR*) resulted in 8 and 59 MU in the WT and CB11 strains, respectively.

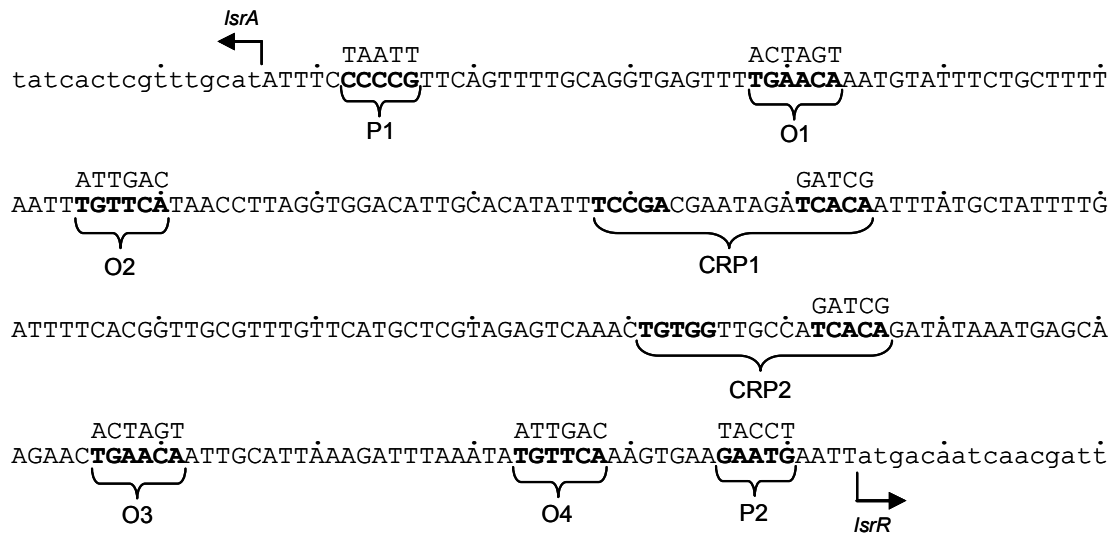
Figure 4-1. Schematic of *lsr* intergenic region.

(A) Map of region showing sites selected for mutation analysis. Sites O1 through O4 (circles) represent putative LsrR binding regions; C1 and C2 (boxes) are cAMP-CRP binding sites; P1 and P2 (hexagons) are negative controls. (B) Intergenic region for *lsr* regulon (uppercase letters) showing sequences of sites from (A). Segments located above indicated sites display mutated sequences for that region. For each CRP binding region [27] only consensus sequence (TCACA) was mutated. Arrows represent transcription start sites for divergent gene (lowercase letters). Points above sequences measure every tenth base.

A



B



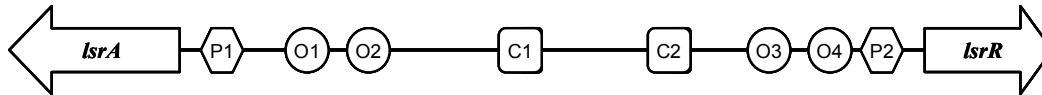
As expected, mutation of each CRP binding site resulted in a marked drop in expression of its corresponding adjacent gene. Mutation of C1 abolished expression of *lsrA* (by >96%) and mutation of C2 resulted in a similar drop in *lsrR* expression (Figure 4-2). Incidentally, we found that the CRP mutations also impacted expression for those representative genes located oppositely and distally from mutated sites. Specifically, the C1 mutation reduced expression towards *lsrR* by 65% in WT and 50% in CB11 (Δ *lsrR*) strains, while C2 dropped levels measured toward *lsrA* by 83% in the WT and 46% in CB11.

Mutation of both C1 and C2 simultaneously returned bi-directional expression to minimal levels directly comparable to adjacent CRP-transcriptional start site schema (Figure 4-2). Although expression changes appeared to be more drastic in the WT strains, inherent error of β -galactosidase assays at very low expression levels are significant (up to $\pm 30\%$). As a result, the effect of CRP site mutations is likely more representative in CB11, in which LsrR repression is absent and expression levels are higher with less standard error. Additionally, expression patterns in the WT and CB11 strains were similar, and no cases were observed in which expression levels increased in the presence of LsrR. That is, LsrR showed no activation effect under the conditions tested.

Notably, mutations in either CRP binding site did not reduce the effect of LsrR repression in the opposite direction. For example, mutation of C1 still resulted in an expression difference in WT and CB11 (Δ *lsrR*) strains toward *lsrR* (pcrp22 in Figure 4-2) correlating to approximately 80% repression due to LsrR, similar to that seen with the native promoter region (plsrR26 in Figure 4-2). LsrR repression of the

Figure 4-2. Effects of site mutations on *lsr* divergent gene expression.

Mutations to selected locations in the intergenic region are indicated by an ‘X’ underneath the corresponding sites shown at the top. After mutations, entire region was inserted into the MCS of pFZY1, a low-copy mini-F derivative with a multiple cloning site located upstream of a promoterless *galK'-lacZYA* reporter segment. Each region was inserted into pFZY1 in both coding directions; plasmid names flanking each mutated region indicate the construct measuring expression in that direction. CRP site mutations (above dashed line) were measured in *E. coli* K12 WT and CB11 (Δ *lsrR*) strains in order to confirm LsrR activity as a repressor only. ‘% Exp.’ represents the percent of expression measured with each mutation set using the native intergenic region, represented by plsrA14 and plsrR26, as full expression levels. Operator and negative control (O1-O4 and P1-P2 sites, respectively; below dashed line) expression values were also measured in WT and CB11 strains. ‘% Rep.’ indicates the percentage of repression by LsrR for each individual region, measuring the difference in expression for WT and CB11 strains and comparing to complete expression (CB11). All values under WT and CB11 in each column are standard Miller Units. Expression assays were conducted in triplicate and are within $\pm 5\%$ error with the exception of values under 20 MU, which are $\pm 10-30\%$. Plasmid names in bold indicate constructs which reduce the amount of LsrR repression by $\geq 96\%$ in that direction.



WT			CB11			% Exp.						WT			CB11			% Exp.		
83	154	100	plsrA14	←	→	plsrR26	23	118	100											
6	4	3	pcrp9	←	×	→	pcrp22	8	59	50										
15	83	54	pcrp8	←	×	→	pcrp23	4	6	5										
6	3	2	pcrp10	←	×	×	→	pcrp4b	6	6	5									

WT			CB11			% Rep.						WT			CB11			% Rep.		
83	154	46	plsrA14	←	→	plsrR26	23	118	80											
147	149	1	plsrA34	←	×	×	→	plsrR33	119	122	2									
80	154	48	plsrA13	←	×	×	→	plsrR24	30	114	73									
61	146	42	plsrA12	←	×	×	→	plsrR25	183	380	52									
138	140	1	plsrA11b	←	×	×	→	plsrR27	260	489	47									
155	162	4	plsrA11	←	×	×	→	plsrR30	25	123	80									
112	141	20	plsrA32	←	×	×	→	plsrR31	249	476	48									
39	80	51	plsrA40	←	×	→	plsrR39	20	115	82										
67	137	51	plsrA38	←	×	×	→	plsrR37	33	123	73									
42	73	42	plsrA36	←	×	×	→	plsrR35	212	497	57									
81	148	45	plsrA42	←	×	×	→	plsrR41	12	45	73									
75	142	47	plsrA44	←	×	×	→	plsrR43	15	59	74									
68	149	54	plsrA46	←	×	×	→	plsrR45	85	125	68									

gene located proximally to the mutated CRP site is unclear. That is, reduced expression levels in, for example, the direction of *lsrA* using pcrp9 (Figure 4-2) may be due to LsrR repression or greatly reduced recruitment of RNAP because of lack of CRP due to mutation at C1.

Complete LsrR repression is dependent upon binding of two identical palindromic half-sites in the divergent *lsr* promoter region. In order to build a comprehensive model for gene expression and regulation in the *lsr* intergenic region we examined the upstream loci for *lsrR* and *lsrA* for potential repressor binding sites. Previous work indicates two 30-bp LsrR binding boxes exist in the upstream regions of *lsrR* and *lsrA*, and deletions and mutations of sections within each box affected LsrR binding in β -galactosidase activity assays [22]. However, the interspersed, A-T rich homology of the proposed 30-bp alignments suggests possible incongruity with a typical helix-turn-helix (H-T-H) binding motif that is predicted for LsrR. Specifically, the SorC family of transcriptional regulators, in which LsrR is categorized, has a H-T-H binding domain that binds to a typical consensus DNA motif of two conserved 6 to 9 base-pair (bp) segments separated by a 6 to 8 bp spacer [125-127].

After further examining the proposed LsrR binding regions, two identical palindromic sequences were discovered either overlapping or immediately adjacent to the sites indicated previously [22], with the sequence TGAACA-21-TGTTCA lying at -16 and -31 of the *lsrR* and *lsrA* start sites, respectively (Figure 4-1B, Section 4.6, Supplemental Data). This palindrome does not correspond to the typical length for

H-T-H consensus motifs; however, previous work has shown that regulators in the SorC family which also function in multimeric form *in vivo* are dependent upon adjacent operator sites up to 20 bp long, each containing a palindromic half-site [69, 128]. Using motif-based computer sequence analysis (<http://meme.nbcr.net>)[109] and manual sequence alignment we could find no other instances of this exact palindrome in the entire *E. coli* K12 genome.

Previous genomic binding studies using LsrR also report limited binding sites for the repressor in *S. typhimurium* and *E. coli* [49]. Given this apparent uniqueness and localization to the *lsr* intergenic region, we sought to evaluate the importance and effect of the palindrome on LsrR binding and divergent expression in the *lsr* regulon. When motifs for DNA-binding proteins are evaluated, one half of the binding site may be more highly conserved or specifically bound, and studies of multimeric repressors of divergent gene sets have shown simultaneous binding to two widely separated operator pairs [69, 83, 128, 129]. Therefore, in order to fully understand the importance of distinct segments of the palindrome on LsrR binding and subsequent expression we designated each palindromic half-site as a separate operator site (O1 through O4 in Figure 4-1A,B).

We constructed multiple pFZY1 fusions containing the *lsr* intergenic region with various sets of mutated operator sites and measured resulting bi-directional expression in both WT and CB11 (Δ *lsrR*) strains. Simultaneous mutation of the palindromic half-sites O1 and O4 (located proximally to each divergent gene start site) resulted in expression levels in WT comparable to those in CB11 when expression was measured in both the *lsrR* and *lsrA* directions (Figure 4-2, plsA34,

plsrR33), indicating a complete elimination of the effect of LsrR when these mutations were present.

Mutations of the O2-O3 half-sites located further from the transcription start sites showed a reduction in LsrR repression by 32% and 26% toward *lsrR* and *lsrA*, respectively; interestingly, however, although the repressive effect of LsrR observed in this mutation set measured in the direction of *lsrR* declined, overall *lsrR* expression increased drastically in both the WT and CB11 strains (Figure 4-2). Notably, the large increase in expression toward *lsrR* was also present in all plasmid constructs carrying a mutation in site O3. Additional mutation combinations resulted in either the complete or partial elimination of LsrR repression in a single direction.

Surprisingly, however, full palindrome mutations (O1-O2 or O3-O4) on either side of the *lsr* intergenic region had little apparent impact on LsrR regulation of expression towards *lsrA*. Mutation of O1-O2 also showed minimal effect in the direction of *lsrR*; however, mutation set O3-O4 did reduce LsrR repression toward *lsrR* from 80% to 52% (Figure 4-2, plsrR25).

Next, in order to show that other mutations outside the evaluated palindrome in the *lsr* operon promoter region did not have these same effects on LsrR binding, we constructed reporter plasmids carrying mutations from -5 to -10 from the *lsrA* and *lsrR* start sites (denoted as P1 and P2 in Figure 4-1B) and measured expression in both directions. We also constructed plasmids containing these mutations in conjunction with previously included half-site mutations (Figure 4-2, plsrR35-plsrA46). Results of each single mutation showed reduced expression levels in the proximal gene in both the WT and CB11 strains (plsrA40, plsrR41) but no apparent

reduction in the effect of LsrR binding. Combinations of P1 or P2 with operator site mutations resulted in generally lower expression levels when mutations were proximally located to gene start sites. For example, mutation of P1 and O3 (plsrA36, plsrR35) reduced *lsrA* expression levels in both the WT and CB11 strains by approximately half, with *lsrR* expression mirroring levels similar to those seen with other O3 mutations (plsrR31, plsrR27). Given the pattern of lower expression levels and maintained effect of LsrR we hypothesize that these mutation sets primarily interfere with RNA polymerase holoenzyme (RNAP) or sigma-factor binding.

4.5 Discussion

The repressor, LsrR, has been hypothesized as the single expression-regulating mechanism in the AI-2 QS ‘circuit’ and has recently garnered more focus for its local and global reach [18, 22, 49]. LsrR has been shown to regulate both the *lsr* operon and itself [21, 27] and our more recent work concerning the divergent nature of the regulatory network also revealed greater expression rates in the direction of the *lsr* operon (Byrd *et al.*, submitted for publication). However, no study has yet considered the dynamics of the disparate CRP or LsrR binding sites within the *lsr* intergenic region and whether regulation is a result of separate activation or repression of *lsrR* and the *lsr* operon, or if there is an interdependence of the entire region on expression of these QS genes. In this study we investigated the *lsr* intergenic region in its entirety, accounting for the directional impacts of CRP activation and LsrR repression in order to construct a model for expression control of the *lsr* regulon.

CRP is a global activator in *E. coli* and functions by binding in upstream promoter regions and increasing the ability of RNA polymerase to bind and initiate gene transcription [61, 72]. In the *lsr* intergenic region, the single and double mutations of CRP binding sites C1 and C2 show the involvement of both CRP proteins for expression of *lsrR* and the *lsr* operon (Figure 4-2). Mutation of the CRP binding site adjacent to the RNAP initiation site almost completely eliminated proximal protein expression. In the case of C1, mutation abolished expression of *lsrA*. Of note, this was irrespective of the presence or absence of LsrR. Thus, CRP is apparently required for transcription of the *lsr* operon. The exact converse scenario was also found. That is, mutation of C2 eliminated expression in the direction of *lsrRK*. Interestingly, the C1 mutation also reduced the transcription of *lsrRK* and C2 reduced that of the *lsr* operon. These sites were not originally thought to provide cooperativity. Also, it was noteworthy that LsrR was still functional with both of these mutations.

This combinative dependency on both sites qualifies the *lsr* regulon as a class III CRP-dependent promoter (Figure 4-3), at which one α subunit C-terminal domain (α CTD) of RNAP interacts with the downstream CRP dimer subunit of C1, one α CTD interacts with the upstream subunit of CRP at C2, and the downstream subunit of CRP at C2 interacts with RNAP α subunit N-terminal domain (α NTD). When bound, CRP can then coordinate to control regulation through multiple and sometimes remote operator sites to which their individual repressors bind [104, 130]. The presence of two or more CRP binding sites in a promoter region is not

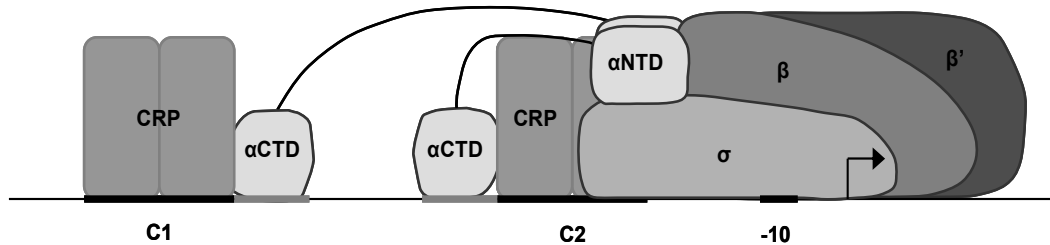


Figure 4-3. Schematic model of CRP-RNAP complex in the *lsr* regulon.

Dependency upon both CRP sites simultaneously for full divergent expression qualifies the *lsr* intergenic region as a class III CRP-dependent promoter. In this mechanism, the α subunit C-terminal domains (α CTD) of RNAP interact concurrently with the downstream CRP dimer subunit of C1 and the upstream subunit of CRP at C2, while RNAP α subunit N-terminal domain (α NTD) interacts with the downstream subunit of CRP at C2. (Adapted from Busby *et al.* [60, 131]).

uncommon, as illustrated in the models of the *lac*, *deo*, and *ara* operons [82, 132-134].

We also sought to thoroughly examine the binding motif for LsrR, and typical protein-DNA binding studies often include *in vitro* assays such as EMSA in order to evaluate changes in protein affinity to DNA sequences carrying mutations or deletions in the native binding region. However, the challenges of *in vitro* work with LsrR have been noted previously [22, 51] and continued efforts to modify conditions for LsrR purification and use have not been successful to date. As a result, we designed a series of *in vivo* reporter plasmids carrying the native *lsr* intergenic region and replicate sets with mutations at putative LsrR binding sites as well as other points outside the suspected binding region (Figure 4-2). In our previous study we showed that *lsrA* is more strongly expressed under normal conditions than *lsrR* (Byrd *et al.*, submitted for publication); therefore, we continued to factor directionality by inserting each mutated intergenic region in opposite directions into separate reporter plasmids, resulting in two plasmids for each mutation set. These sets were subsequently transformed into *E. coli* K12MG1655 and CB11 (Δ *lsrR*) in order to evaluate the effect of the carried promoter mutations on LsrR binding and repression.

Our results show an elimination of LsrR repression of *lsrA* when the staggered O1-O3 or O2-O4 operator mutations are present (Figure 4-2). In the direction of *lsrR* the O1-O3 mutation reduced LsrR repression by 33%; however, the O2-O4 mutation had no apparent effect in the same direction. In the medial O2-O3 mutation set, LsrR repression was reduced in both directions, with *lsrA* and *lsrR* repression dropping by

26% and 32%, respectively. Most notably, the effects of repression by LsrR are completely eliminated in both directions when operators O1 and O4 are mutated.

The observation that divergent LsrR repression is entirely eliminated when both O1 and O4, but not combinations thereof, are mutated suggests a dependence upon repressor interaction with both palindromic operator regions simultaneously. This, in addition to our findings that for normal expression in the *lsr* region both CRP proteins are bound at once, we propose here that transcriptional control of the *lsr* regulon incorporates DNA looping as part of its mechanism of regulation. DNA looping has become recognized as a centrally involved mechanism in expression control and is demonstrated in nearly every model of regulation in which CRP has a role; in addition to its function as an activator, CRP bends DNA from 90-125 degrees and facilitates the formation of DNA-repressor loops in other models [69, 74, 88].

Very surprisingly, mutations of O1-O2 did not show a significant reduction in LsrR repression in either direction, while the O3-O4 mutation set reduced the effect of LsrR by 28% only in the direction of *lsrR*. Like varying operator effects seen in the *lac* and *deo* operons [82, 135], we postulate that the *lsr* operator sites also vary in their impact on expression in the *lsr* regulon and the effect of DNA bending can recruit distal regions to a promoter site. As a result, while LsrR binding to one full palindromic site may be inhibited, DNA conformation due to CRP and other proximal interactions could maintain the *lsr* intergenic region in a looped state. Analogous to RepA repression, LsrR may impede RNAP-DNA binding events while in the vicinity of a promoter but not directly bound [90, 136], allowing, for example, LsrR bound at sites O1-O2 (vicinity of *lsrA*) to repress *lsrR* expression.

Although the effect of LsrR repression toward *lsrR* may be reduced using different mutation sets, we note again that in each plasmid construct carrying an O3 mutation *lsrR* expression levels in both the WT and CB11 strains were dramatically increased. These repeated expression level increases only toward *lsrR* suggest that a previously unreported layer of repression may also be present in addition to LsrR. In the divergently transcribed *ula* regulon and *hpt-gcd* gene set, each of which also incorporates various CRP and operator sites, similar increases in expression were noticed when the IHF binding site was mutated or expression activity was measured in cells deficient in IHF [63, 69]. Given our observation that expression values increase substantially beyond those measured in an *lsrR* mutant, we hypothesize that the *lsr* regulon also incorporates an additional IHF-like regulatory mechanism that is heavily dependent upon the operator O3 for expression control. Our search for IHF binding sequences in the *lsr* intergenic region produced several potential sites. However, the consensus sequence for IHF in *E. coli* is loosely conserved and difficult to predict and determining positive IHF binding sites in the *lsr* region awaits further study [137-139].

In our previous study we reported a native bias in expression resulting in approximately 1.5-fold greater facultative expression towards the *lsr* operon (Byrd *et al.*, submitted for publication). However, our current results illustrate that although *lsrA* is more highly expressed than *lsrR* under normal conditions, potential expression levels for *lsrR* are up to 11-fold higher in WT *E. coli* K12 when operator site O3 is mutated. We found no operator mutations that produced the same increased expression toward *lsrA*. Additionally, several of the mutation conditions resulted in a

complete elimination of the effect of LsrR in the direction of *lsrA* while only one produced the same effect toward *lsrR* (Figure 4-2, bold). The finding that LsrR repression is relieved from the *lsrA* promoter under various conditions, coupled with the newly realized potential strength of the *lsrR* promoter, suggests that LsrR affinity and subsequent repression is less specific in the *lsrA* promoter region than in that for *lsrR*. In addition, the observation that LsrR can continue to regulate expression in the direction of *lsrR* with different mutations present indicate a more specific affinity for the *lsrR* promoter region.

Considering the complete relief of LsrR repression seen with the O1-O4 mutation set, we postulate that these two operators are the primary binding sites for LsrR with the medial palindromic half-sites O2-O3, shown to partially relieve LsrR repression in each direction, being involved as secondary sites. Also, when the O1-O4 mutations were present, no drastic increases in expression were observed as previously noted with mutations of O3, indicating also that O1 and O4 are not involved in the secondary IHF-like regulation system we propose. LsrR and UlaR are both members of the SorC family of transcriptional regulators, and the *lsr* and *ula* regulons demonstrate significant structural similarities as well [17, 69, 128]. LsrR was previously determined to exist *in vivo* in dimeric form [52]; however, we offer that LsrR actively functions as a tetramer through interaction with flanking operator sites and itself simultaneously as is commonly found in repressor-DNA loop schema (Figure 4-4) [89, 124, 128, 131]. UlaR was demonstrated to maintain a stoichiometric ratio of one molecule of L-ascorbate-6-P per dimer [128]. We also

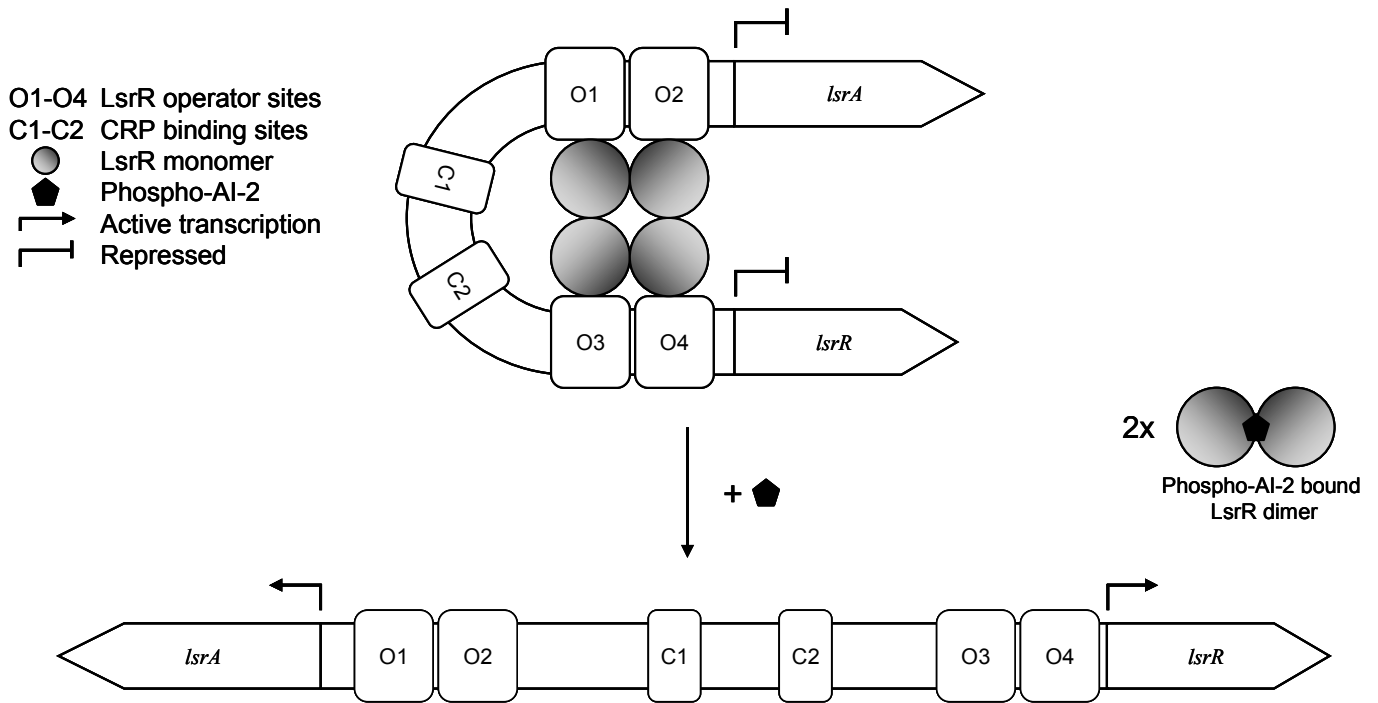


Figure 4-4. Proposed model for LsrR repression.

DNA interaction with CRP bound to sites C1 and C2 results in a looped DNA complex with each operator site bound by an LsrR tetramer, repressing transcription of *lsrR* and *lsrA*. After interaction with phospho-AI-2, LsrR dimerizes and dissociates from operator sites O1-O4, permitting active transcription of both divergent genes. Stoichiometric ratio of one molecule of phospho-AI-2 per LsrR dimer is based upon UlaR comparison (see Discussion). (Adapted from Garces *et al.* [128])

hypothesize that, given the functional and structural similarities to UlaR (Section 4.6, Supplemental Data), LsrR binds one molecule of phospho-AI-2 per dimer. However, further study is required to determine precise phospho-AI-2/LsrR stoichiometry and potential secondary protein interactions in the vicinity of operator sites O2 and O3 which may hinder or enhance LsrR binding.

E. coli K12 is able to use L-ascorbate as a carbon source through fermentation via the *ula* regulon [140]. The *ula* operon, consisting of six genes (*ulaA*, *B*, *C*, *D*, *E*, and *F*)(Figure 4-5A) comprise a PTS system reported to convert L-ascorbate to d-xylulose 5-phosphate, an intermediary in the pentose phosphate pathway. The divergently oriented gene *ulaG* encodes a metal-dependent hydrolase which further utilizes ascorbate, and *ulaR*, located downstream of *ulaG*, encodes the repressor for the *ula* regulon [140, 141]. Similarly to the *lsr* regulon, in which the cognate repressor LsrR is antagonized by the phosphorylated form of AI-2, L-ascorbate-6-phosphate, converted from L-ascorbate through interaction with UlaA, UlaB, and UlaC, binds to and prevents UlaR from repressing transcription of the *ula* regulon [142].

Because of the structural similarity of natural furanones such as L-ascorbate to that of AI-2 (Figure 4-5B), their application has been postulated to potentially inhibit the AI-2 QS response in bacteria [12, 143]. In a bioluminescence assay using *V. harveyi* and *C. perfringens*, increasing levels of both ascorbic acid and sodium ascorbate resulted in decreased luminescent response, likely due to their occupation of AI-2 receptors in the reporter strain [143]. However, a co-culture of five separate strains of *C. perfringens* grown on ground beef with increased levels of ascorbic acid

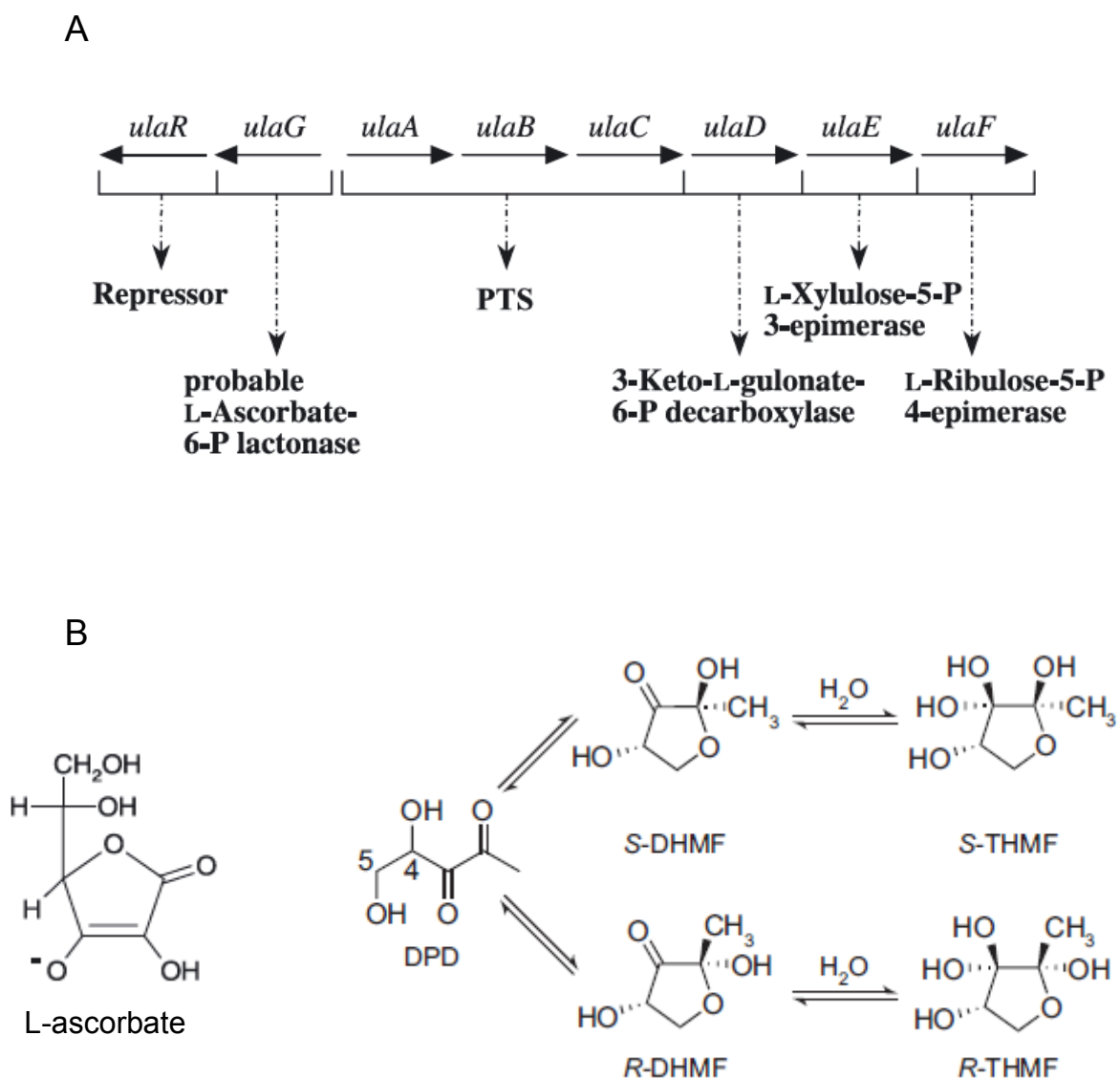


Figure 4-5. L-ascorbate utilization in *E. coli*.

(A) The *ula* regulon, indicated for the utilization of L-ascorbate in *E. coli* [142]. (B) Structural similarities of L-ascorbate and two epimeric furanoses formed after cyclization of DPD, (2R,4S)- and (2S,4S)-2,4-dihydroxy-2-methyldihydrofuran-3-one (R- and S-DHMF) which are then hydrated to form R- and S-THMF [44,142].

present showed little variation in growth versus a negative control, although [143]. While L-ascorbate did not demonstrate a drastic reduction in growth of *C. perfringens* in the studied case, the reduction of the luminescent response in the *V. harveyi* reporter strain shows the potential for ascorbate or other natural furanones as AI-2 analogs or potentially as cooperative sensing mechanisms in bacteria.

Although previous studies have approached the *lsr* intergenic region as two oppositely oriented promoters [22, 27], our results clearly show that dependencies upon interaction from remote portions of the region affect repression and activation of each divergent gene set and that the *lsr* quorum sensing genes are likely regulated as part of a DNA-CRP-LsrR complex, as is common in systems regulated by the SorC family of repressors [69, 87, 114]. The observation that various instances exist in which repression is completely relieved from *lsrACDB* but not *lsrR* imply a less stringent regulation of the AI-2 transport gene, potentially contributing to the native expression bias toward *lsrA* we reported previously (Byrd *et al.*, submitted for publication).

We have described for the first time the surprisingly strong potential expression of *lsrR* and its control by an additional putative layer of regulation by an IHF-like system separate from LsrR. However, conditions which stimulate *lsrR* expression and the exact system which specifically adds to its regulation require further examination. Also, that the putative nature of LsrR regulation may vary from dimer to tetramer is potentially interesting since this is a QS regulator, and members of this family are thought to have sharp or controllable ‘switching’ characteristics [144, 145]. The potential genomic and phenotypic impacts of high intracellular levels

of LsrR are significant, and understanding the mechanism and circumstances which facilitate such a response will help to better understand the QS process.

4.6 Supplemental Data

***lsr* promoter region sequence alignment.** Cross-species investigations of DNA homology in upstream promoter regions have been recognized to provide identification of conserved regulatory protein (primarily transcription factor) binding sites and, as a result, aid in identifying transcriptional regulatory networks and systems [91, 96, 146]. As discussed in section 4.4, the palindrome 5'-TGAACA-21-TGTTCA-3' is indicated to be a putative binding site for the QS regulator LsrR. In order to examine the phylogenetic conservation of this palindrome, we selected 12 alternate bacterial species which were previously reported to contain QS orthologs for all proteins encoded via the *lsr* operon in *E. coli* (Table 4-2) [147]. For each strain, the entire putative *lsr* intergenic region was extracted from GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>) and aligned using ClustalW alignment tool (ClustalW Ver. 2.1, <http://www.clustal.org/>). Resulting data were visualized using CINEMA (Colour INteractive Editor for Multiple Alignments v. 2.1).

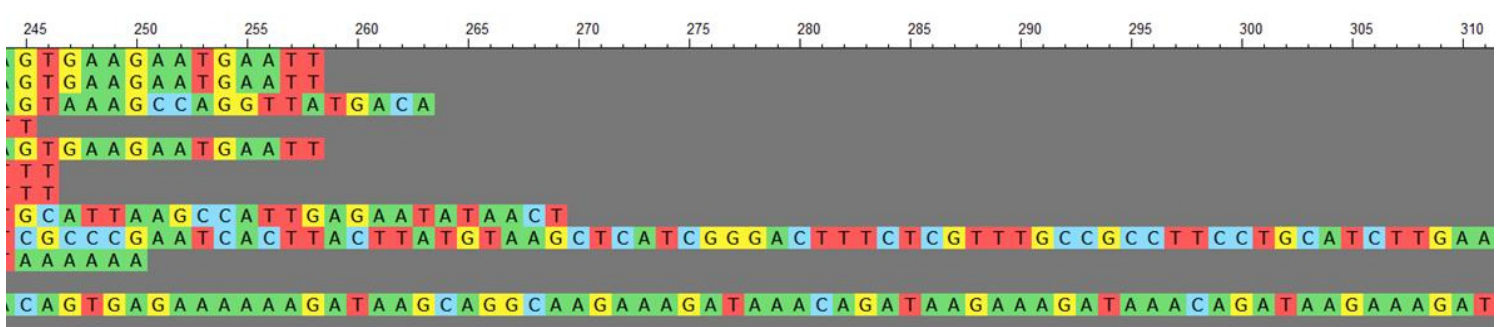
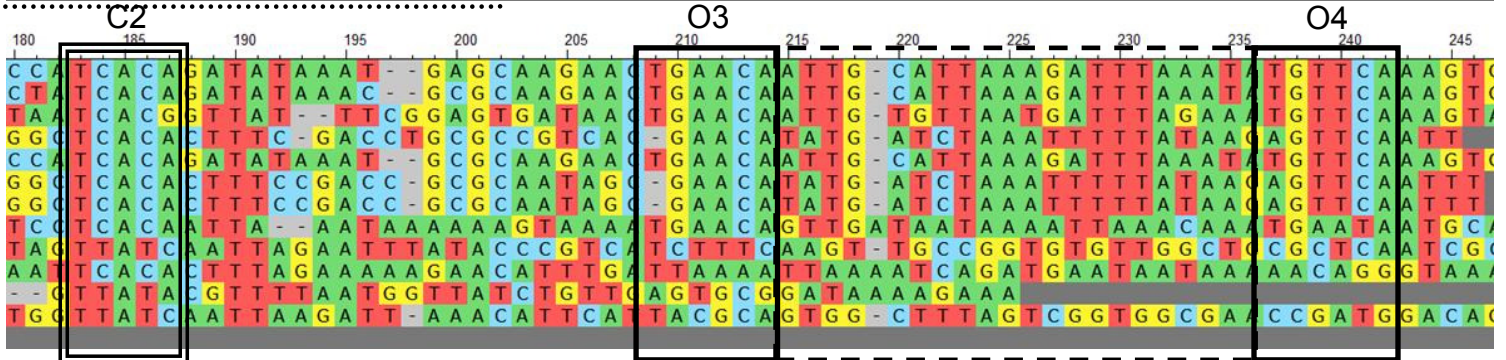
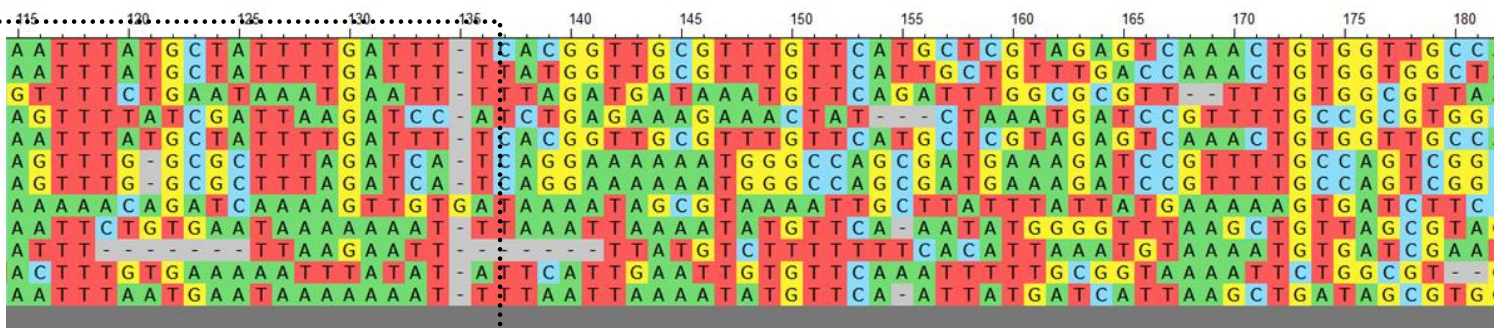
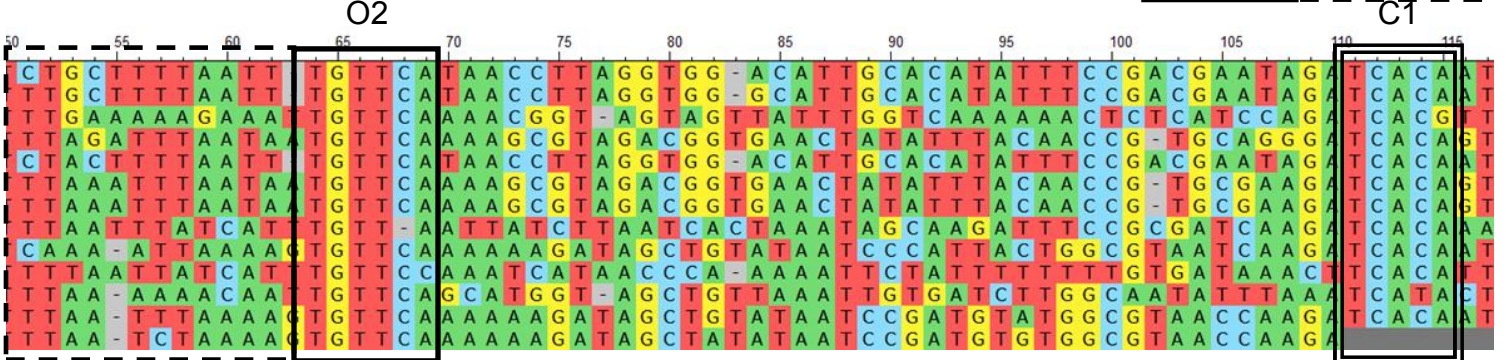
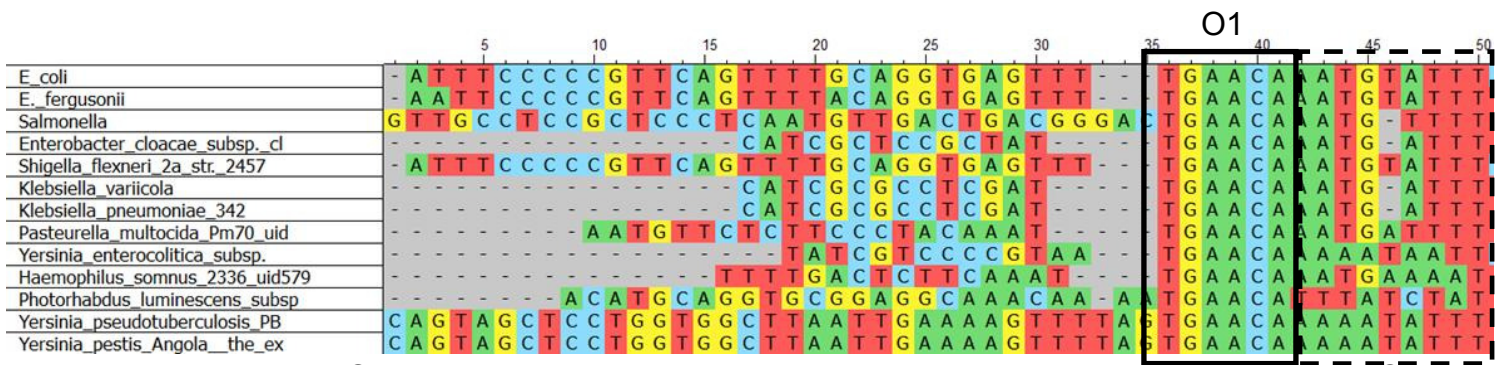
Sequence alignments show strong conservation of the palindromic LsrR binding sites previously designated as O1 through O4 in Figure 4-1A (single boxes in Figure 4-5). Operator site O1 is completely conserved across all species investigated, while O2 has complete conservation with the exception of *Pasteurella multocida* str. Pm70 which displays a single base-pair mismatch. Alignment of operator site O3

Table 4-2. Bacterial strains used in this study.

GenBank Accession Number	Description
U00096.2	<i>Escherichia coli</i> str. K-12 substr. MG1655
CU928158.2	<i>Escherichia fergusonii</i> ATCC 35469
NC003197.2	<i>Salmonella enterica</i> Typhimurium str. LT2
CP001918.1	<i>Enterobacter cloacae</i> ATCC 13047
NC004741	<i>Shigella flexneri</i> 2a str. 2457T
CP001891.1	<i>Klebsiella variicola</i> At-22
CP000964.1	<i>Klebsiella pneumoniae</i> 342
AE004439.1	<i>Pasteurella multocida</i> str. Pm70
AM286415.1	<i>Yersinia enterocolitica</i> str. 8081
CP000436.1	<i>Haemophilus somnus</i> 129PT
BX571869.1	<i>Photorhabdus luminescens</i> subsp. TTO1
CP001048.1	<i>Yersinia pseudotuberculosis</i> PB1/+
CP000901.1	<i>Yersinia pestis</i> str. Angola

Figure 4-6. Cross-species *lsr* intergenic sequence alignment.

Sequences corresponding to intergenic regions for twelve bacterial species shown to have orthologous genes to the *lsr* regulon in *E. coli* were compared. Species selected for comparison are listed to left of first base pair for each region. Base pair designations are: adenine - A (green), thymine - T (red), cytosine - C (blue), guanine - G (yellow). Dashes represent blank positions. CRP and putative LsrR binding regions show high conservation between species and are indicated by solid boxes (O1 through O4) and double boxes (C1 and C2). Additional sections were noted containing A-T rich (<80%) regions lying between palindromic half sites (dashed boxes) and flanking CRP site C1 (dotted box).



showed five strains with complete homology to *E. coli* K12, and another three with a single base-pair mismatch. Four strains, *Haemophilus somnus* 129PT, *Photothabdus luminescens* subsp. TTO1, *Yersinia pseudotuberculosis* PB1/+, and *Yersinia pestis* str. Angola, demonstrate little homology to operator site O3. Similarly, alignments for operator site O4 show four strains with complete matches and an additional three strains with a single base-pair mismatch. Strain *Yersinia enterocolitica* str. 8081, in addition to the four strains showing a complete base-pair mismatch with operator site O3, also had minimal conservation for O4.

cAMP-CRP binding sites are also well conserved between species, showing high homology between sequences for both C1 and C2 (double boxes in Figure 4-5). In C1, all evaluated species with documented CRP motifs show the consensus half-site 5'-TCACA-3' (5'-TCACG-3' in *Salmonella enterica*) with the exception of *Yersinia pestis*, for which the *lsr* intergenic sequence is limited, and *Photothabdus luminescens*, which shows a single base pair mismatch [86, 148-150]. Similarly to operator sites O3 and O4, site C2 also shows slightly lower conservation of DNA sequence between species than the corresponding site proximally located to *lsrA*. Only three strains, *Yersinia enterocolitica* str. 8081, *Haemophilus somnus* 129PT, and *Yersinia pseudotuberculosis* PB1/+ demonstrate variation from the consensus sequence, with three or fewer base-pair differences present.

Alignment of the *lsr* intergenic regions also demonstrates notable similarities in other segments. First, the sections between palindromic half-sites show A-T rich (approximately 80%) DNA segments (dashed boxes in Figure 4-5). While difficult to predict, such segments may indicate increased DNA flexibility or conformational changes including bending, hairpin, or cruciform formation [151-153]. Finally, an

additional 21-bp A-T rich area is located at approximately bp 115-136, flanking the CRP binding site C1.

Previous work has compared genes in the *lsr* regulon and the potential impact of orthologous proteins among various species for detection of AI-2 [147]. Although exact roles of these newly identified sections demonstrating high conservation are impossible to predict without further study, the intergenic regions are also a promising location to evaluate similarities among species in order to better understand the QS regulatory process and commonalities of AI-2 detection.

Preliminary structural comparison of LsrR, DeoR, and UlaR. As discussed in Sections 2.5 and 4.4, LsrR is categorized in the SorC family of transcriptional regulators, and according to I-TASSER structural prediction analysis holds high similarity to DeoR, regulator of distinct both sugar-binding and DNA-binding domains and transition from bound (repressive) to unbound upon binding their appropriate effectors [67, 128]. These effectors, L-ascorbate-6-phosphate and deoxyribose-5-phosphate for UlaR and DeoR, respectively, are phosphorylated before antagonizing the repressors similar to AI-2 prior to interaction with LsrR [22, 51].

Secondary structure prediction (UniProt) indicates domain similarities for LsrR to UlaR as well as DeoR. Given the strong predicted similarities, tertiary structural comparison was conducted by submitting the amino acid sequence for DeoR and UlaR to I-TASSER protein prediction server and visualizing the result using Jmol (See section 3.3, Materials and Methods). Results indicate clear structural similarities between DeoR, UlaR, and LsrR (Figure 4-6). Each predicted

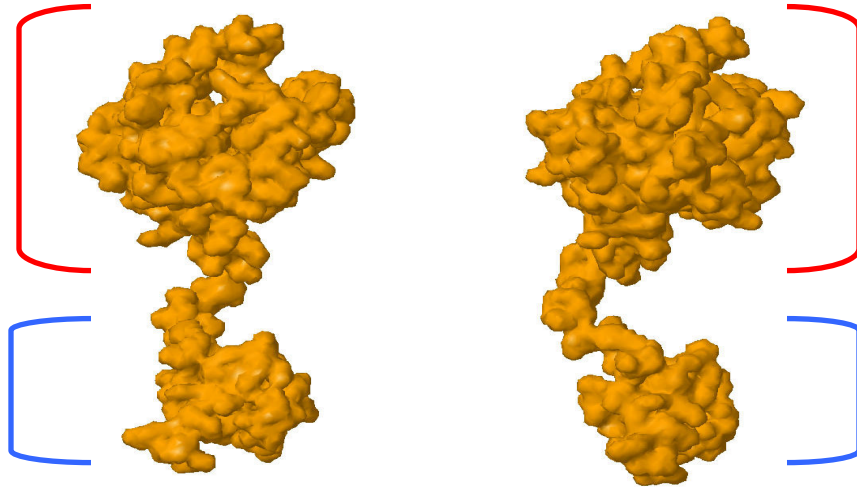
Figure 4-7. Preliminary structural comparison of DeoR, UlaR, and LsrR.

Front (left column) and side (right column) view of computed DeoR, UlaR, and LsrR tertiary structures. Predicted sugar-binding (red brackets) and DNA-binding (blue brackets) domains are indicated. Tertiary structure prediction for each protein was completed using I-TASSER Protein Structure and Function Prediction server (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>) and van der Waals surface was computed and visualized using Jmol (<http://jmol.sourceforge.net/>).

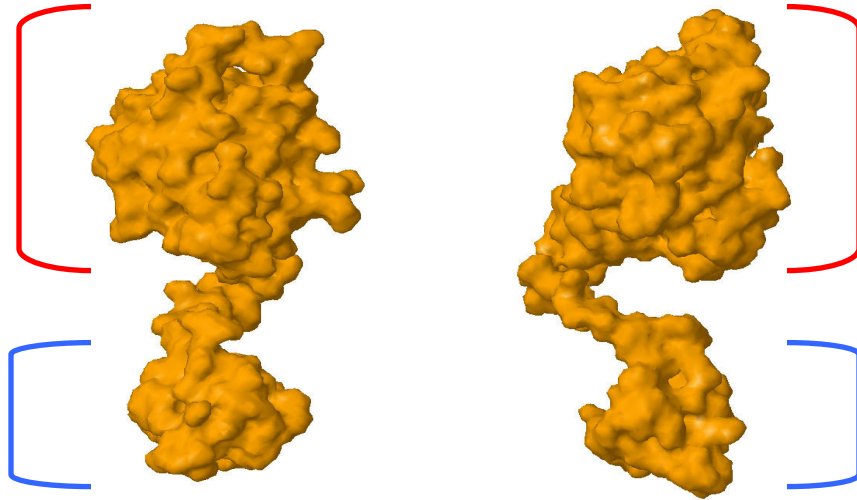
FRONT

SIDE

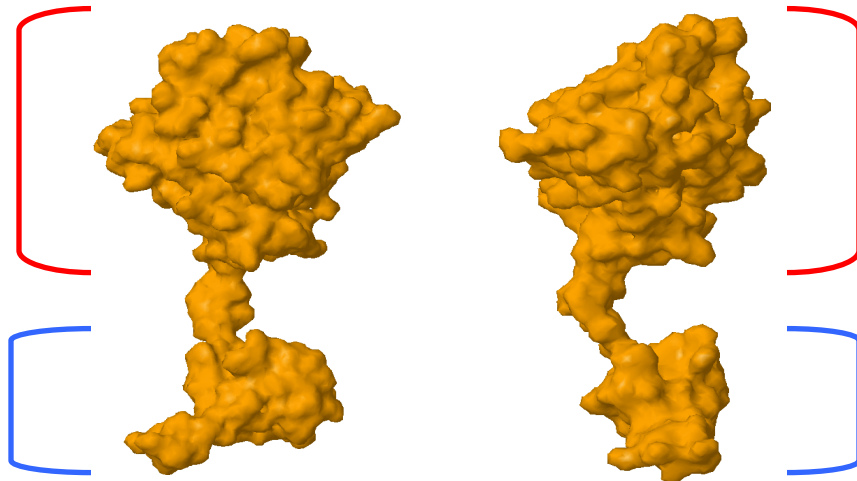
DeoR



UlaR



LsrR



protein structure indicates a sugar-binding domain distinct from the DNA-binding domain, and basic forms appear similar from front and side views.

Other proteins with similar functions have been previously characterized and demonstrate comparable compositions. For example, arabinose-binding protein (ABP), galactose binding protein (GBP), sulfate-binding protein (SBP), ribose-binding protein (RBP), and four others all have two distinct lobes connected by two to three connecting strands [154, 155]. LsrR has also demonstrated binding to multiple sites simultaneously *in vivo* similar to UlaR and DeoR (See section 4.5, Discussion). While conclusive functional similarities between proteins based solely upon structural analysis cannot be confirmed, these apparent similarities further support the hypothesis that LsrR, like UlaR and DeoR, functions *in vivo* as a tetramer and potentially uses DNA looping to aid in repression of the *lsr* regulon.

Chapter 5: Summary

5.1 Summary of Results

In this dissertation, the regulatory effects and scope of the QS repressor LsrR were examined and impacts of the activator CRP in the *lsr* intergenic region were explored. The focus was to better understand the genetic impact of QS circuitry and construct a more accurate model of regulation within the *lsr* regulon.

In Chapter 2, a plasmid vector encoding *lsrR* was constructed and an N-terminus His-tagged LsrR was cultured. The challenges of protein purification with the repressor due to insolubility under various conditions were discussed. *In vitro* analysis of harvested LsrR showed undetectable DNA binding activity to the *lsr* intergenic region, and the addition of the activator CRP did not enhance or otherwise promote LsrR binding under tested conditions. Two sites for CRP binding in the intergenic region were confirmed, and binding affinity was shown to be similar between the sites despite differences in specific DNA sequences.

In Chapter 3 genomic binding site analysis was conducted using DNA microarrays and constructing a C-terminus LsrR, which displayed better DNA binding activity and *in vitro* characteristics. The results showed four putative binding sites for LsrR in the *E. coli* genome, three of which (*lsrR-lsrA*, *yegE-udk*, *mppA*) have direct potential connection to the QS response. Also, divergent expression analysis in the *lsr* regulon revealed a natural expression bias in the direction of *lsrACDBFG*, which encodes the AI-2 import and degradation proteins. Implications of a consistently higher facultative expression toward *lsrA* were discussed.

Chapter 4 introduced specific putative binding sites for LsrR in the *lsr* regulon and utilized extensive mutation sets in the *lsr* intergenic region in conjunction with divergent expression analysis in order to narrow the impact of the various sites. Similar analysis of individual CRP binding sites and their impact upon divergent expression were conducted, indicating the *lsr* region to be a class III CRP-dependent promoter, in which both CRP proteins simultaneously contribute to transcriptional activation in each direction. Finally, a proposed model for regulation of the *lsr* regulon was presented incorporating LsrR, CRP, DNA looping, and a secondary layer of repression by an IHF-like protein.

5.2 Future Directions

In vivo activity and analysis of LsrR permitted the collection of data necessary to construct a putative model of regulation in the *lsr* operon. However, various aspects of LsrR characteristics and activity are still unclear. Efforts to stabilize LsrR and increase protein solubility and purification yields for *in vitro* analysis would permit more exhaustive analysis of LsrR binding sites and provide the ability to confirm a consensus motif.

There appears to be a discrepancy between the unexpectedly few genes directly regulated by LsrR and the wide genetic and phenotypic impacts of the QS process. If LsrR is the primary regulator for the QS circuit, further work may examine subsequent transcriptional activity beyond LsrR derepression in the *lsr* intergenic region and determine downstream impacts of the genes that were indicated to also be directly controlled during microarray analysis.

The proposed model for the LsrR-CRP-DNA complex presented in this dissertation also discusses for the first time an additional layer of regulation in the *lsr* regulon. When this layer of repression was relieved via DNA sequence mutations, expression of *lsrR* increased by up to 11-fold. Similar models of divergent transcriptional regulation in which drastic increases in expression were measured showed the involvement of integration host factor (IHF) as part of the regulatory mechanism. This putative model can be confirmed by obtaining IHF and conducting *in vitro* or *in vivo* analysis to explore the *lsr* intergenic region for potential binding sites. IHF can then be incorporated into the model for *lsr* regulation and its impacts on the QS process can be considered.

An alternate possibility for the additional layer of regulation in the *lsr* regulon is the formation of DNA hairpin, cruciform, or other structural DNA formations. If these structural changes exist, the palindrome explored in Chapter 4 will likely be involved and the resulting RNA formation and subsequent rates of transcription will be affected. Microbiological assays to explore the effect of these various DNA structures will help show their impact on the QS process and their potential contribution to the switch-like effect commonly seen in QS circuits.

Finally, alignment of the intergenic sequences of multiple strains of *E. coli* and the closely related *S. flexneri* showed conservation of the LsrR and CRP binding motifs that were evaluated in this dissertation. In addition to these species, however, the same motifs for the LsrR palindrome presented for *E. coli* were also discovered in *S. enterica* which is further removed phylogenetically but displays a similar AI-2 based QS process. Database searches for bacteria with AI-2 QS systems and gene

sets homologous to the *lsr* regulon in *E. coli* may allow a further refinement of the regulatory module which controls these genes as well as provide greater insight into the similarities of the QS process among various bacterial species.

References

1. Nixon BT, Ronson CW, Ausubel FM. 2-component regulatory systems responsive to environmental stimuli share strongly conserved domains with the nitrogen assimilation regulatory genes *ntrB* and *ntrC*. Proceedings of the National Academy of Sciences of the United States of America. 1986;83(20):7850-54.
2. Kofoed EC, Parkinson JS. Transmitter and receiver modules in bacterial signaling proteins. Proceedings of the National Academy of Sciences of the United States of America. 1988;85(14):4981-85.
3. Ortega F, Acerenza L, Westerhoff HV, Mas F, Cascante M. Product dependence and bifunctionality compromise the ultrasensitivity of signal transduction cascades. Proceedings of the National Academy of Sciences of the United States of America. 2002;99(3):1170-75.
4. Turovskiy Y, Kashtanov D, Paskhover B, Chikindas ML. Quorum sensing: Fact, fiction, and everything in between. Advances in Applied Microbiology, Vol 62. 2007;62:191-234.
5. Fuqua WC, Winans SC, Greenberg EP. Quorum sensing in bacteria - the *luxR-luxI* family of cell density-responsive transcriptional regulators. Journal of Bacteriology. 1994;176(2):269-75.
6. Xavier KB, Bassler BL. LuxS quorum sensing: more than just a numbers game. Current Opinion in Microbiology. 2003;6(2):191-97.
7. Jayaraman A, Wood TK. Bacterial quorum sensing: Signals, circuits, and implications for biofilms and disease. Annual Review of Biomedical Engineering. 2008;10:145-67.
8. Lowery CA, Dickerson TJ, Janda KD. Interspecies and interkingdom communication mediated by bacterial quorum sensing. Chemical Society Reviews. 2008;37(7):1337-46.
9. Byrd CM, Bentley WE. Quieting cross talk - the quorum sensing regulator LsrR as a possible target for fighting bacterial infections. Cell Research. 2009;19(11):1229-30.

10. Klevens RM, Morrison MA, Nadle J, Petit S, Gershman K, Ray S, *et al.* Invasive methicillin-resistant *Staphylococcus aureus* infections in the United States. *Jama-Journal of the American Medical Association*. 2007;298:1763-71.
11. Czajkowski R, Jafra S. Quenching of acyl-homoserine lactone-dependent quorum sensing by enzymatic disruption of signal molecules. *Acta Biochimica Polonica*. 2009;56(1):1-16.
12. Roy V, Fernandes R, Tsao CY, Bentley WE. Cross Species Quorum Quenching Using a Native AI-2 Processing Enzyme. *ACS Chemical Biology*. 2010;5(2):223-32.
13. Asad S, Opal SM. Bench-to-bedside review: Quorum sensing and the role of cell-to-cell communication during invasive bacterial infection. *Critical Care*. 2008;12(6).
14. Engebrecht J, Nealson K, Silverman M. Bacterial bioluminescence - isolation and genetic-analysis of functions from *Vibrio fischeri*. *Cell*. 1983;32(3):773-81.
15. Winans SC. Bacterial Esperanto. *Nature Structural Biology*. 2002;9(2):83-84.
16. Bassler BL, Wright M, Silverman MR. Multiple signaling systems controlling expression of luminescence in *Vibrio-harveyi* - sequence and function of genes encoding a 2nd sensory pathway. *Molecular Microbiology*. 1994;13(2):273-86.
17. Taga ME, Semmelhack JL, Bassler BL. The LuxS-dependent autoinducer AI-2 controls the expression of an ABC transporter that functions in AI-2 uptake in *Salmonella typhimurium*. *Molecular Microbiology*. 2001;42(3):777-93.
18. Taga ME, Miller ST, Bassler BL. Lsr-mediated transport and processing of AI-2 in *Salmonella typhimurium*. *Molecular Microbiology*. 2003;50(4):1411-27.
19. Xavier KB, Bassler BL. Regulation of uptake and processing of the quorum-sensing autoinducer AI-2 in *Escherichia coli*. *Journal of Bacteriology*. 2005;187(1):238-48.
20. Wang L, Li J, March JC, Valdes JJ, Bentley WE. *luxS*-Dependent gene regulation in *Escherichia coli* K-12 revealed by genomic expression profiling. *Journal of Bacteriology*. 2005;187(24):8350-60.
21. Li J, Attila C, Wang L, Wood TK, Valdes JJ, Bentley WE. Quorum sensing in *Escherichia coli* is signaled by AI-2/LsrR: Effects on small RNA and Biofilm architecture. *Journal of Bacteriology*. 2007;189(16):6011-20.

22. Xue T, Zhao LP, Sun HP, Zhou XX, Sun BL. LsrR-binding site recognition and regulatory characteristics in *Escherichia coli* AI-2 quorum sensing. *Cell Research*. 2009;19(11):1258-68.
23. Waters CM, Bassler BL. Quorum sensing: Cell-to-cell communication in bacteria. *Annual Review of Cell and Developmental Biology*. 2005;21:319-46.
24. Hellingwerf KJ, Crielgaard WC, de Mattos MJT, Hoff WD, Kort R, Verhamme DT, et al. Current topics in signal transduction in bacteria. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology*. 1998;74(4):211-27.
25. Olivera BCL, Ugalde E, Martinez-Antonio A. Regulatory dynamics of standard two-component systems in bacteria. *Journal of Theoretical Biology*. 2010;264(2):560-69.
26. Bassler BL, Wright M, Showalter RE, Silverman MR. Intercellular signaling in *Vibrio-harveyi* - sequence and function of genes regulating expression of luminescence. *Molecular Microbiology*. 1993;9(4):773-86.
27. Wang L, Hashimoto Y, Tsao CY, Valdes JJ, Bentley WE. Cyclic AMP (cAMP) and cAMP receptor protein influence both synthesis and uptake of extracellular autoinducer 2 in *Escherichia coli*. *Journal of Bacteriology*. 2005;187(6):2066-76.
28. McNab R, Lamont RJ. Microbial dinner-party conversations: the role of LuxS in interspecies communication. *Journal of Medical Microbiology*. 2003;52(7):541-45.
29. Miller MB, Bassler BL. Quorum sensing in bacteria. *Annual Review of Microbiology*. 2001;55:165-99.
30. Farah C, Vera M, Morin D, Haras D, Jerez CA, Guiliani N. Evidence for a functional quorum-sensing type AI-1 system in the extremophilic bacterium *Acidithiobacillus ferrooxidans*. *Applied and Environmental Microbiology*. 2005;71(11):7033-40.
31. Steindler L, Venturi V. Detection of quorum-sensing N-acyl homoserine lactone signal molecules by bacterial biosensors. *Fems Microbiology Letters*. 2007;266(1):1-9.
32. Federle MJ, Bassler BL. Interspecies communication in bacteria. *Journal of Clinical Investigation*. 2003;112(9):1291-99.

33. Bassler BL. How bacteria talk to each other: regulation of gene expression by quorum sensing. *Current Opinion in Microbiology*. 1999;2(6):582-87.
34. Sun JB, Daniel R, Wagner-Dobler I, Zeng AP. Is autoinducer-2 a universal signal for interspecies communication: a comparative genomic and phylogenetic analysis of the synthesis and signal transduction pathways. *BMC Evolutionary Biology*. 2004;4.
35. Novick RP, Muir TW. Virulence gene regulation by peptides in *staphylococci* and other Gram-positive bacteria. *Current Opinion in Microbiology*. 1999;2(1):40-45.
36. Pesavento C, Becker G, Sommerfeldt N, Possling A, Tschowri N, Mehlig A, *et al*. Inverse regulatory coordination of motility and curli-mediated adhesion in *Escherichia coli*. *Genes & Development*. 2008;22(17):2434-46.
37. Sperandio V, Torres AG, Giron JA, Kaper JB. Quorum sensing is a global regulatory mechanism in enterohemorrhagic *Escherichia coli* O157 : H7. *Journal of Bacteriology*. 2001;183(17):5187-97.
38. Barrios AFG, Zuo RJ, Hashimoto Y, Yang L, Bentley WE, Wood TK. Autoinducer 2 controls biofilm formation in *Escherichia coli* through a novel motility quorum-sensing regulator (MqsR, B3022). *Journal of Bacteriology*. 2006;188(1):305-16.
39. Lee JT, Jayaraman A, Wood TK. Indole is an inter-species biofilm signal mediated by SdiA. *Bmc Microbiology*. 2007;7.
40. Griffith KL, Wolf RE. Measuring beta-galactosidase activity in bacteria: Cell growth, permeabilization, and enzyme assays in 96-well arrays (vol 290, pg 402, 2002). *Biochemical and Biophysical Research Communications*. 2002;292(1):292-92.
41. DeLisa MP, Wu CF, Wang L, Valdes JJ, Bentley WE. DNA microarray-based identification of genes controlled by autoinducer 2-stimulated quorum sensing in *Escherichia coli*. *Journal of Bacteriology*. 2001;183(18):5239-47.
42. Davies DG, Parsek MR, Pearson JP, Iglewski BH, Costerton JW, Greenberg EP. The involvement of cell-to-cell signals in the development of a bacterial biofilm. *Science*. 1998;280(5361):295-98.
43. Prouty AM, Schwesinger WH, Gunn JS. Biofilm formation and interaction with the surfaces of gallstones by *Salmonella* spp. *Infection and Immunity*. 2002;70(5):2640-49.

44. Rickard AH, Palmer RJ, Blehert DS, Campagna SR, Semmelhack MF, Eglund PG, et al. Autoinducer 2: a concentration-dependent signal for mutualistic bacterial biofilm growth. *Molecular Microbiology*. 2006;60(6):1446-56.
45. Xavier KB, Miller ST, Lu WY, Kim JH, Rabinowitz J, Pelczer I, et al. Phosphorylation and processing of the quorum-sensing molecule autoinducer-2 in enteric bacteria. *ACS Chemical Biology*. 2007;2(2):128-36.
46. Beeston AL, Surette MG. pfs-dependent regulation of autoinducer-2 production in *Salmonella enterica* serovar Typhimurium. *Journal of Bacteriology*. 2002;184(13):3450-56.
47. Heal RD, Parsons AT. Novel intercellular communication system in *Escherichia coli* that confers antibiotic resistance between physically separated populations. *Journal of Applied Microbiology*. 2002;92(6):1116-22.
48. Pompeani AJ, Irgon JJ, Berger MF, Bulyk ML, Wingreen NS, Bassler BL. The *Vibrio harveyi* master quorum-sensing regulator, LuxR, a TetR-type protein is both an activator and a repressor: DNA recognition and binding specificity at target promoters. *Molecular Microbiology*. 2008;70(1):76-88.
49. Thijs IM, Zhao H, De Weerd A, Engelen K, De Coster D, Schoofs G, et al. The AI-2-dependent regulator LsrR has a limited regulon in *Salmonella Typhimurium*. *Cell Res*. 2010;20(8):966-69.
50. Waldminghaus T, Skarstad K. ChIP on Chip: surprising results are often artifacts. *BMC Genomics*. 2010;11.
51. Li J, Xue XJ, Ruan JB, Wu MH, Zhu ZQ, Zang JY. Cloning, purification, crystallization and preliminary crystallographic analysis of LsrR from *Escherichia coli*. *Acta Crystallographica Section F-Structural Biology and Crystallization Communications*. 2010;66:902-04.
52. Xiaotian L, Minhao W, Demeng S, Jianye Z. Cloning, purification, crystallization and preliminary crystallographic analysis of LsrR from *Escherichia coli*. *Acta Crystallographica, Section F (Structural Biology and Crystallization Communications)*. 2010:913-15.
53. de Sanctis D, McVey CE, Enguita FJ, Carrondo MA. Crystal Structure of the Full-Length Sorbitol Operon Regulator SorC from *Klebsiella pneumoniae*: Structural

- Evidence for a Novel Transcriptional Regulation Mechanism. *Journal of Molecular Biology*. 2009;387(3):759-70.
54. Friedman AM, Fischmann TO, Steitz TA. Crystal-structure of *lac* repressor core tetramer and its implications for DNA looping. *Science*. 1995;268(5218):1721-27.
 55. Parkinson G, Wilson C, Gunasekera A, Ebright YW, Ebright RE, Berman HM. Structure of the CAP-DNA complex at 2.5 angstrom resolution: A complete picture of the protein-DNA interface. *Journal of Molecular Biology*. 1996;260(3):395-408.
 56. Lindahl L HA. Diversity of mechanisms in the regulation of translation in prokaryotes and lower eukaryotes. *Curr Op Gen & Dev*. 1992;2(5):720-26.
 57. Winkler WC, Breaker RR. Regulation of bacterial gene expression by riboswitches. *Annual Review of Microbiology*. 2005;59:487-517.
 58. Buck MJ, Lieb JD. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics*. 2004;83(3):349-60.
 59. Wu JJ, Smith LT, Plass C, Huang THM. ChIP-chip comes of age for genome-wide functional analysis. *Cancer Research*. 2006;66(14):6899-902.
 60. Busby S, Ebright RH. Transcription activation by catabolite activator protein (CAP). *Journal of Molecular Biology*. 1999;293(2):199-213.
 61. Kolb A, Busby S, Buc H, Garges S, Adhya S. Transcriptional regulation by cAMP and its receptor protein. *Annual Review of Biochemistry*. 1993;62:749-95.
 62. Merkel TJ, Dahl JL, Ebright RH, Kadner RJ. Transcription activation at the *Escherichia-coli uhpT* promoter by the catabolite gene activator protein. *Journal of Bacteriology*. 1995;177(7):1712-18.
 63. Izu H, Ito S, Elias MD, Yamada M. Differential control by IHF and cAMP of two oppositely oriented genes, *hpt* and *gcd*, in *Escherichia coli*: significance of their partially overlapping regulatory elements. *Molecular Genetics and Genomics*. 2002;266(5):865-72.
 64. Aiba H. Transcription of the *Escherichia-coli* adenylate-cyclase gene is negatively regulated by cAMP receptor protein. *Journal of Biological Chemistry*. 1985;260(5):3063-70.

65. Hanamura A, Aiba H. Molecular mechanism of negative autoregulation of *Escherichia-coli* CRP gene. *Nucleic Acids Research*. 1991;19(16):4413-19.
66. Mortensen L, Dandanell G, Hammer K. Purification and characterization of the DeoR repressor of *Escherichia-coli*. *Embo Journal*. 1989;8(1):325-31.
67. Zeng XM, Saxild HH, Switzer RL. Purification and characterization of the DeoR repressor of *Bacillus subtilis*. *Journal of Bacteriology*. 2000;182(7):1916-22.
68. Hammer K, Bech L, Hobolth P, Dandanell G. DNA specificity of *Escherichia-coli* *deoP1* operator-DeoR repressor recognition. *Molecular & General Genetics*. 1993;237(1-2):129-33.
69. Campos E, Baldoma L, Aguilar J, Badia J. Regulation of expression of the divergent *ulaG* and *ulaABCDEF* Operons involved in L-ascorbate dissimilation in *Escherichia coli*. *Journal of Bacteriology*. 2004;186(6):1720-28.
70. DeLisa MP, Valdes JJ, Bentley WE. Mapping stress-induced changes in autoinducer AI-2 production in chemostat-cultivated *Escherichia coli* K-12. *Journal of Bacteriology*. 2001;183(9):2918-28.
71. Tagami H, Aiba H. Role of CRP in transcription activation at *Escherichia-coli lac* promoter - CRP is dispensable after the formation of open complex. *Nucleic Acids Research*. 1995;23(4):599-605.
72. Hammerjerspersen K, Nygaard P. Multiple regulation of nucleoside catabolizing enzymes in *Escherichia-coli* - effects of 3-5' cyclic-amp and CRP protein. *Molecular & General Genetics*. 1976;148(1):49-55.
73. Tagami H, Aiba H. A common role of CRP in transcription activation: CRP acts transiently to stimulate events leading to open complex formation at a diverse set of promoters. *Embo Journal*. 1998;17(6):1759-67.
74. Lin SH, Lee JC. Determinants of DNA bending in the DNA-cyclic AMP receptor protein complexes in *Escherichia coli*. *Biochemistry*. 2003;42(17):4809-18.
75. Perros M, Stietz TA. DNA looping and *lac* repressor CAP interaction. *Science*. 1996;274(5294):1929-30.
76. Zhang X, Schleif R. Catabolite gene activator protein mutations affecting activity of the *araBAD* promoter. *Journal of Bacteriology*. 1998;180(2):195-200.

77. Bondos SE, Bicknell A. Detection and prevention of protein aggregation before, during, and after purification. *Analytical Biochemistry*. 2003;316(2):223-31.
78. Maxwell KL, Mittermaier AK, Forman-Kay JD, Davidson AR. A simple *in vivo* assay for increased protein solubility. *Protein Science*. 1999;8(9):1908-11.
79. Schein CH. Solubility as a function of protein-structure and solvent components. *Bio-Technology*. 1990;8(4):308-15.
80. Dale GE, Broger C, Langen H, Darcy A, Stuber D. Improving protein solubility through rationally designed amino-acid replacements - solubilization of the trimethoprim-resistant type S1 dihydrofolate-reductase. *Protein Engineering*. 1994;7(7):933-39.
81. Hammerjespersen K, Munchpetersen A. Multiple regulation of nucleoside catabolizing enzymes - regulation of *deo* operon by CytR and DeoR gene products. *Molecular & General Genetics*. 1975;137(4):327-35.
82. Amouyal M, Mortensen L, Buc H, Hammer K. Single and double loop formation when DeoR repressor binds to its natural operator sites. *Cell*. 1989;58(3):545-51.
83. Plumbridge J, Kolb A. *nag* repressor-operator interactions - protein-DNA contacts cover more than 2 turns of the DNA helix. *Journal of Molecular Biology*. 1995;249(5):890-902.
84. Krin E, Laurent-Winter C, Bertin PN, Danchin A, Kolb A. Transcription regulation coupling of the divergent *argG* and *metY* promoters in *Escherichia coli* K-12. *Journal of Bacteriology*. 2003;185(10):3139-46.
85. Freydank AC, Brandt W, Drager B. Protein structure modeling indicates hexahistidine-tag interference with enzyme activity. *Proteins-Structure Function and Bioinformatics*. 2008;72(1):173-83.
86. Barber AM, Zhurkin VB, Adhya S. CRP-binding sites - evidence for 2 structural classes with 6-bp and 8-bp spacers. *Gene*. 1993;130(1):1-8.
87. Larson TJ, Cantwell JS, Vanloobhattacharya AT. Interaction at a distance between multiple operators controls the adjacent, divergently transcribed *glpTQ-glpACB* operons of *Escherichia-coli* K-12. *Journal of Biological Chemistry*. 1992;267(9):6114-21.

88. Youn H, Kerby RL, Conrad M, Roberts GP. Study of highly constitutively active mutants suggests how cAMP activates cAMP receptor protein. *Journal of Biological Chemistry*. 2006;281(2):1119-27.
89. ValentinHansen P, SogaardAndersen L, Pedersen H. A flexible partnership: The CytR anti-activator and the cAMP-CRP activator protein, comrades in transcription control. *Molecular Microbiology*. 1996;20(3):461-66.
90. Perezmartin J, Espinosa M. The RepA repressor can act as a transcriptional activator by inducing DNA bends. *Embo Journal*. 1991;10(6):1375-82.
91. Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J, *et al*. Transcriptional regulation by the numbers: models. *Current Opinion in Genetics & Development*. 2005;15(2):116-24.
92. Parsek MR, Val DL, Hanzelka BL, Cronan JE, Greenberg EP. Acyl homoserine-lactone quorum-sensing signal generation. *Proceedings of the National Academy of Sciences of the United States of America*. 1999;96(8):4360-65.
93. Magnuson R, Solomon J, Grossman AD. Biochemical and genetic-characterization of a competence pheromone from *Bacillus-subtilis*. *Cell*. 1994;77(2):207-16.
94. Zhu J, Miller MB, Vance RE, Dziejman M, Bassler BL, Mekalanos JJ. Quorum-sensing regulators control virulence gene expression in *Vibrio cholerae*. *Proceedings of the National Academy of Sciences of the United States of America*. 2002;99(5):3129-34.
95. Herzberg M, Kaye IK, Peti W, Wood TK. YdgG (TqsA) controls biofilm formation in *Escherichia coli* K-12 through autoinducer 2 transport. *Journal of Bacteriology*. 2006;188(2):587-98.
96. MacIsaac KD, Lo KA, Gordon W, Motola S, Mazor T, Fraenkel E. A Quantitative Model of Transcriptional Regulation Reveals the Influence of Binding Location on Expression. *Plos Computational Biology*. 2010;6(4).
97. Miller ST, Xavier KB, Campagna SR, Taga ME, Semmelhack MF, Bassler BL, *et al*. *Salmonella typhimurium* recognizes a chemically distinct form of the bacterial quorum-sensing signal AI-2. *Molecular Cell*. 2004;15(5):677-87.

98. Koop AH, Hartley ME, Bourgeois S. A low-copy-number vector utilizing beta-galactosidase for the analysis of gene-control elements. *Gene*. 1987;52(2-3):245-56.
99. Cho BK, Barrett CL, Knight EM, Park YS, Palsson BO. Genome-scale reconstruction of the Lrp regulatory network in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;105(49):19462-67.
100. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, *et al.* UniProt: the Universal Protein knowledgebase. *Nucleic Acids Research*. 2004;32:D115-D19.
101. Xavier KB, Bassler BL. Interference with AI-2-mediated bacterial cell-cell communication. *Nature*. 2005;437(7059):750-53.
102. Wehmeier UF, Lengeler JW. Sequence of the *sor*-operon for l-sorbose utilization from *Klebsiella-pneumoniae* kay2026. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology*. 1994;1208(2):348-51.
103. Wohrl BM, Wehmeier UF, Lengeler JW. Positive and negative regulation of expression of the l-sorbose (*sor*) operon by SorC in *Klebsiella-pneumoniae*. *Molecular & General Genetics*. 1990;224(2):193-200.
104. Dandanell G, Hammer K. 2 operator sites separated by 599 base-pairs are required for DeoR repression of the *deo* operon of *Escherichia-coli*. *Embo Journal*. 1985;4(12):3333-38.
105. Pazy Y, Eisenberg-Donlovich Y, Laitinen OH, Kulomaa MS, Bayer EA, Wilchek M, *et al.* Dimer-Tetramer transition between solution and crystalline states of streptavidin and avidin mutants. *Journal of Bacteriology*. 2003;185(14):4050-56.
106. Sikdar MSI, Kim JS. Expression of a natural fusion gene for uracil phosphoribosyltransferase and uridine kinase from rice shows growth retardation by 5-fluorouridine or 5-fluorouracil in *Escherichia coli*. *African Journal of Biotechnology*. 2010;9(9):1295-303.
107. Hammerje.K, Munchpet.A. Mutants of *Escherichia-coli* unable to metabolize cytidine - isolation and characterization. *Molecular & General Genetics*. 1973;126(2):177-86.

108. Sperandio V, Mellies JL, Nguyen W, Shin S, Kaper JB. Quorum sensing controls expression of the type III secretion gene transcription and protein secretion in enterohemorrhagic and enteropathogenic *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America. 1999;96(26):15196-201.
109. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, *et al*. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Research. 2009;37:W202-W08.
110. Park JT, Raychaudhuri D, Li HS, Normark S, Mengin-Lecreulx D. MppA, a periplasmic binding protein essential for import of the bacterial cell wall peptide L-alanyl-gamma-D-glutamyl-meso-diaminopimelate. Journal of Bacteriology. 1998;180(5):1215-23.
111. Letoffe S, Delepelaire P, Wandersman C. The housekeeping dipeptide permease is the *Escherichia coli* heme transporter and functions with two optional peptide binding proteins. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(34):12891-96.
112. Griffiths E. Iron and bacterial virulence - a brief overview. Biology of Metals. 1991;4(1):7-13.
113. Griffiths E. Iron-regulated membrane-proteins and bacterial virulence. Journal of Biosciences. 1990;15(3):173-77.
114. Miller MB, Skorupski K, Lenz DH, Taylor RK, Bassler BL. Parallel quorum sensing systems converge to regulate virulence in *Vibrio cholerae*. Cell. 2002;110(3):303-14.
115. Vendeville A, Winzer K, Heurlier K, Tang CM, Hardie KR. Making 'sense' of metabolism: Autoinducer-2, LuxS and pathogenic bacteria. Nature Reviews Microbiology. 2005;3(5):383-96.
116. Ahmer BMM. Cell-to-cell signalling in *Escherichia coli* and *Salmonella enterica*. Molecular Microbiology. 2004;52(4):933-45.
117. Lilley BN, Bassler BL. Regulation of quorum sensing in *Vibrio harveyi* by LuxO and Sigma-54. Molecular Microbiology. 2000;36(4):940-54.

118. Passador L, Cook JM, Gambello MJ, Rust L, Iglewski BH. Expression of *Pseudomonas-aeruginosa* virulence genes requires cell-to-cell communication. *Science*. 1993;260(5111):1127-30.
119. Elvers KT, Park SF. Quorum sensing in *Campylobacter jejuni*: detection of a *luxS* encoded signalling molecule. *Microbiology-Sgm*. 2002;148:1475-81.
120. Giron JA, Torres AG, Freer E, Kaper JB. The flagella of enteropathogenic *Escherichia coli* mediate adherence to epithelial cells. *Molecular Microbiology*. 2002;44(2):361-79.
121. Sambrook J, Fritsch EF, Maniatis T. *Molecular cloning: a laboratory manual*. 2d ed. Cold Spring Harbor, N.Y. : Cold Spring Harbor Laboratory Press, 1989.
122. Inoue H, Nojima H, Okayama H. High-efficiency transformation of *Escherichia-coli* with plasmids. *Gene*. 1990;96(1):23-28.
123. Gedeon T, Mischaikow K, Patterson K, Traldi E. When activators repress and repressors activate: A qualitative analysis of the Shea-Ackers model. *Bulletin of Mathematical Biology*. 2008;70(6):1660-83.
124. Shin M, Kang S, Hyun SJ, Fujita N, Ishihama A, Valentin-Hansen P, et al. Repression of deoP2 in *Escherichia coli* by CytR: conversion of a transcription activator into a repressor. *Embo Journal*. 2001;20(19):5392-99.
125. Harrison SC, Aggarwal AK. DNA recognition by proteins with the helix-turn-helix motif. *Annual Review of Biochemistry*. 1990;59:933-69.
126. Brennan R. Interactions of the helix-turn-helix binding domain. *Curr Op Str Bio*. 1991;1:80-88.
127. Huffman JL, Brennan RG. Prokaryotic transcription regulators: more than just the helix-turn-helix motif. *Current Opinion in Structural Biology*. 2002;12(1):98-106.
128. Garces F, Fernandez FJ, Gomez AM, Perez-Luque R, Campos E, Prohens R, et al. Quaternary Structural Transitions in the DeoR-Type Repressor UlaR Control Transcriptional Readout from the L-Ascorbate Utilization Regulon in *Escherichia coli*. *Biochemistry*. 2008;47(44):11424-33.
129. Plumbridge J, Kolb A. DNA bending and expression of the divergent *nagE-B* operons. *Nucleic Acids Research*. 1998;26(5):1254-60.

130. Carra JH, Schleif RF. Variation of half-site organization and DNA looping by AraC protein. *EMBO Journal*. 1993;12(1):35-44.
131. Meibom KL, Kallipolitis BH, Ebright RH, Valentin-Hansen P. Identification of the subunit of cAMP receptor protein (CRP) that functionally interacts with CytR in CRP-CytR-mediated transcriptional repression. *Journal of Biological Chemistry*. 2000;275(16):11951-56.
132. Lobell RB, Schleif RF. *araC* DNA looping - orientation and distance-dependent loop breaking by the cyclic-AMP receptor protein. *Journal of Molecular Biology*. 1991;218(1):45-54.
133. Irani MH, Orosz L, Adhya S. A control element within a structural gene - the gal operon of *Escherichia-coli*. *Cell*. 1983;32(3):783-88.
134. Majumdar A, Adhya S. Demonstration of 2 operator elements in gal - *in vitro* repressor binding-studies. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*. 1984;81(19):6100-04.
135. Oehler S, Eismann ER, Kramer H, Mullerhill B. The 3 operators of the *lac* operon cooperate in repression. *Embo Journal*. 1990;9(4):973-79.
136. Perezmartin J, Rojo F, Delorenzo V. Promoters responsive to DNA bending - a common theme in prokaryotic gene-expression. *Microbiological Reviews*. 1994;58(2):268-90.
137. Goodman SD, Velten NJ, Gao QA, Robinson S, Segall AM. *In vitro* selection of integration host factor binding sites. *Journal of Bacteriology*. 1999;181(10):3246-55.
138. Goodrich JA, Schwartz ML, McClure WR. Searching for and predicting the activity of sites for DNA-binding proteins - compilation and analysis of the binding-sites for *Escherichia-coli* integration host factor (IHF). *Nucleic Acids Research*. 1990;18(17):4993-5000.
139. Kur J, Hasan N, Szybalski W. Physical and biological consequences of interactions between integration host factor (IHF) and coliphage lambda late p'r promoter and its mutants. *Gene*. 1989;81(1):1-15.
140. Yew WS, Gerlt JA. Utilization of L-ascorbate by *Escherichia coli* K-12: Assignments of functions to products of the *yjf-sga* and *yia-sgb* operons. *Journal of Bacteriology*. 2002;184(1):302-06.

141. Campos E, Montella C, Garces F, Baldoma L, Aguilar J, Badia J. Aerobic L-ascorbate metabolism and associated oxidative stress in *Escherichia coli*. *Microbiology-Sgm*. 2007;153:3399-408.
142. Linster CL, Van Schaftingen E. Vitamin C - Biosynthesis, recycling and degradation in mammals. *Febs Journal*. 2007;274(1):1-22.
143. Novak JS, Fratamico PM. Evaluation of ascorbic acid as a quorum-sensing analogue to control growth, sporulation, and enterotoxin production in *Clostridium perfringens*. *Journal of Food Science*. 2004;69(3):M72-M78.
144. Hooshangi S, Bentley WE. From unicellular properties to multicellular behavior: bacteria quorum sensing circuitry and applications. *Current Opinion in Biotechnology*. 2008;19(6):550-55.
145. Yao Y, Martinez-Yamout MA, Dickerson TJ, Brogan AP, Wright PE, Dyson HJ. Structure of the *Escherichia coli* quorum sensing protein SdiA: Activation of the folding switch by acyl homoserine lactones. *Journal of Molecular Biology*. 2006;355(2):262-73.
146. Zhu Z, Pilpel Y, Church GM. Computational identification of transcription factor binding sites via a transcription-factor-centric clustering (TFCC) algorithm. *Journal of Molecular Biology*. 2002;318(1):71-81.
147. Pereira CS, de Regt AK, Brito PH, Miller ST, Xavier KB. Identification of Functional LsrB-Like Autoinducer-2 Receptors. *Journal of Bacteriology*. 2009;191(22):6975-87.
148. Pyles EA, Chin AJ, Lee JC. *Escherichia coli* cAMP receptor protein-DNA complexes: energetic contributions of half-sites and flanking sequences in DNA recognition. *Biochemistry*. 1998;37(15):5194-200.
149. Barber AM, Zhurkin VB. CAP binding-sites reveal pyrimidine-purine pattern characteristic of DNA bending. *Journal of Biomolecular Structure & Dynamics*. 1990;8(2):213-32.
150. Yanagihara S, Iyoda S, Ohnishi K, Iino T, Kutsukake K. Structure and transcriptional control of the flagellar master operon of *Salmonella typhimurium*. *Genes & Genetic Systems*. 1999;74(3):105-11.

151. Wilson DP, Tkachenko AV, Meiners JC. A generalized theory of DNA looping and cyclization. *Epl.* 2010;89(5).
152. Popov YO, Tkachenko AV. Effects of sequence disorder on DNA looping and cyclization. *Physical Review E.* 2007;76(2).
153. Bikard D, Loot C, Baharoglu Z, Mazel D. Folded DNA in Action: Hairpin Formation and Biological Functions in Prokaryotes. *Microbiology and Molecular Biology Reviews.* 2010;74(4):570-88.
154. Oh B-H, Kang C-H, Hendrik DB, Sung-Hu K, Kishiko N, Joshi A, Ames, Giovanna Ferro-Luzzi. The Bacterial Periplasmic Histidine-binding Protein. *Journal of Biological Chemistry.* 1994;269(6):4135-43.
155. Vyas NK, Vyas M, Quiococho F. Comparison of the Periplasmic Receptors for L-Arabinose, D-Glucose/D-Galactose, and D-Ribose. *The Journal of Biological Chemistry.* 1991;266(8):5226-37.