Biochemical and Structural Studies on PrfA, the Transcriptional Regulator of Virulence

in *Listeria monocytogenes*

by

Keri Lynn Kimes Hamilton

Department of Biochemistry
Duke University

Date:_______________________

Approved:

Richard G. Brennan, Supervisor

Maria A. Schumacher

Kenneth N. Kreuzer

Raphael H. Valdivia

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry in the Graduate School of Duke University

2016
ABSTRACT

Biochemical and Structural Studies on PrfA, the Transcriptional Regulator of Virulence in Listeria monocytogenes

by

Keri Lynn Kimes Hamilton

Department of Biochemistry
Duke University

Date:_______________________

Approved:

___________________________
Richard G Brennan, Supervisor

___________________________
Maria A. Schumacher

___________________________
Kenneth N. Kreuzer

___________________________
Raphael H. Valdivia

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry in the Graduate School of Duke University

2016
Abstract

*Listeria monocytogenes* is a gram-positive soil saprophytic bacterium that is capable of causing fatal infection in humans. The main virulence regulator PrfA, a member of the Crp/FNR family of transcriptional regulators, activates the expression of essential proteins required for host cell invasion and cell-to-cell spread. The mechanism of PrfA activation and the identity of its small molecule coactivator have remained a mystery for more than 20 years, but it is hypothesized that PrfA shares mechanistic similarity to the *E. coli* cAMP binding protein, Crp. Crp activates gene expression by binding cAMP, increasing the DNA binding affinity of the protein and causing a significant DNA bend that facilitates RNA polymerase binding and downstream gene activation. Our data suggests PrfA activates virulence protein expression through a mechanism distinct from the canonical Crp activation mechanism that involves a combination of cysteine residue reduction and glutathione (GSH) binding.

*Listeria* lacking glutathione synthase (*ΔgshF*) is avirulent in mice; however virulence is rescued when the bacterium expresses the constitutively active PrfA mutant G145S. Interestingly, *Listeria* expressing a PrfA mutant in which its four cysteines are mutated to alanine (Quad PrfA), demonstrate a 30-fold decrease in virulence. The Quad and *ΔgshF* double mutant strains are avirulent. DNA-binding affinity, measured through fluorescence polarization assays, indicate reduction of the cysteine side chains
is sufficient to allow PrfA to bind its physiological promoters Phly and PactA with low nanomolar affinity. Oxidized PrfA binds the promoters poorly.

Unexpectedly, Quad also binds promoter DNA with nanomolar affinity, suggesting that the cysteines play a role in transcription efficiency in addition to DNA binding. Both PrfA and Quad bind GSH at physiologically relevant and comparable affinities, however GSH did not affect DNA binding in either case. Thermal denaturation assays suggest that Quad and wild-type PrfA differ structurally upon binding GSH, which supports the in vivo difference in infection between the regulator and its mutant.

Structures of PrfA in complex with cognate DNA, determined through X-ray crystallography, further support the disparity between PrfA and Crp activation mechanisms as two structures of reduced PrfA bound to Phly (PrfA-Phly$_{30}$ and PrfA-Phly$_{24}$) suggest the DNA adopts a less bent DNA conformation when compared to Crp-cAMP-DNA. The structure of Quad-Phly$_{30}$ confirms the DNA-binding data as the protein-DNA complex adopts the same overall conformation as PrfA-Phly.

From these results, we hypothesize a two-step activation mechanism wherein PrfA, oxidized upon cell entry and unable to bind DNA, is reduced upon its intracellular release and binds DNA, causing a slight bend in the promoter and small increase in transcription of PrfA-regulated genes. The structures of PrfA-Phly$_{30}$ and PrfA-Phly$_{24}$ likely visualize this intermediate complex. Increasing concentrations of GSH shift the
protein to a (PrfA-GSH)-DNA complex which is fully active transcriptionally and is hypothesized to resemble closely the transcriptionally active structure of the cAMP-(Crp)-DNA complex. Thermal denaturation results suggest Quad PrfA is deficient in this second step, which explains the decrease in virulence and implicates the cysteine residues as critical for transcription efficiency. Further structural and biochemical studies are on-going to clarify this mechanism of activation.
Contents

Abstract ................................................................................................................................. iv

List of Tables ........................................................................................................................ xi

List of Figures ....................................................................................................................... xii

List of Abbreviations ........................................................................................................... xiv

Acknowledgements ............................................................................................................. xvi

Chapter 1: Listeria monocytogenes- Virulence and Significance ........................................... 1

1.1 The Biological Relevance of Listeria monocytogenes ..................................................... 1

1.2 Listeria monocytogenes: Genus and Environmental Niche ........................................... 1

1.3 Mechanism of Listerial Host Cell Invasion and Virulence ............................................ 2

Chapter 2. PrfA is the Transcriptional Regulator that Mediates Virulence in Listeria
monocytogenes ....................................................................................................................... 8

2.1 The PrfA Box: PrfA-Dependent Gene Activation ............................................................... 9

2.2 Deviation from the Consensus PrfA Box: Not All PrfA-Dependent Promoters are Equal ................................................................................................................................. 10

2.3 PrfA Transcriptional Control by Alternative Sigma Factors, Thermal Regulation, and
Autoregulation ...................................................................................................................... 12

2.4 Post-Translational Control of PrfA Activity .................................................................... 14

2.5 PrfA Independent Virulence Regulation in Listeria monocytogenes ................................. 16

Chapter 3. Insight into the PrfA Activating Ligand ............................................................... 20

3.1 PrfA is Related to E. coli Crp .......................................................................................... 20

3.1.1 The Canonical Crp Mechanism .................................................................................. 21

3.1.2 Important Residues in Crp ....................................................................................... 23
3.1.3 Other Crp/FNR Transcriptional Regulators .......................................................24
3.1.3.1 *E. coli* FNR ............................................................................................................25
3.1.3.2 *Rhodospirillum rubrum* CooA ........................................................................25
3.1.3.3 *Desulfitobacterium hafniense* CprK ................................................................28
3.1.3.4 Cyanobacteria NtcA ............................................................................................29
3.2 The Structures of PrfA and Constitutively Active PrfA* ..............................................31
3.2.1 Mutations in PrfA and Effects on DNA Binding and *Listerial* Virulence ........33
3.2.2 Mutation of the Four Cysteine Residues of PrfA to Alanine ..........................36
3.3 Is Glutathione the Small Molecule Activator of PrfA? ............................................37

Chapter 4. Elucidation of PrfA-Mediated Transcription Mechanism Using X-Ray
Crystallography and Biochemical Assays ........................................................................51
4.1 An Overview of X-Ray Crystallography ................................................................51
4.1.1 Diffraction from the Crystal Lattice ....................................................................53
4.1.2 Using Molecular Replacement to Solve the Phase Problem ............................56
4.1.3 The Translation and Rotation Functions ..............................................................57
4.1.4 Model Refinement ................................................................................................59
4.2 Assessment of DNA-Binding Affinity Using Fluorescence Polarization ........61

Chapter 5. PrfA Binding to Cognate DNA ..................................................................69
5.1 DNA Binding Results .............................................................................................69
5.2 Discussion ................................................................................................................70
5.3 Materials and Methods ............................................................................................73
5.3.1 PrfA Protein Purification .......................................................................................73
Chapter 9. The Structure of the Quad-\textit{Phly}_{30} Complex is the Same as PrfA-\textit{Phly}_{30}........ 121

9.1 Quad-\textit{Phly}_{30} Crystal Conditions and Methods .................................................... 122

Chapter 10. Conclusions and Future Directions ................................................................ 126

10.1 Future Directions ...................................................................................................... 131

10.2 Summary .................................................................................................................. 133

References ..................................................................................................................... 136

Biography ....................................................................................................................... 151
List of Tables

Table 1: PrfA Box Sequences of Various Virulence Genes ..................................................... 18
Table 2: PrfA* and Transcriptionally Inactive PrfA Mutants .................................................. 48
Table 3: DNA Binding Affinities of PrfA (Kₐ ± s.e.m)................................................................. 76
Table 4: Fluoresceinated DNA Oligomers for FP ....................................................................... 78
Table 5: Glutathione Binding Affinity (Kₐ ± s.e.m.).............................................................. 88
Table 6: DNA Binding Affinities of Selected PrfA Point Mutants (Kₐ ± s.e.m.) ................. 110
Table 7: Mutagenesis Primers for PrfA Alanine Mutants ......................................................... 115
Table 8: Select Crystallographic Statistics ............................................................................... 125
List of Figures

Figure 1: Infectivity Cycle of *Listeria monocytogenes* .............................................. 7
Figure 2: PrfA Regulon ........................................................................................................ 19
Figure 3: Crp Family Alignment ....................................................................................... 41
Figure 4: Structures of Crp ............................................................................................... 42
Figure 5: Crp-Dependent Promoter Assembly ................................................................. 43
Figure 6: Structures of CooA .......................................................................................... 44
Figure 7: Structures of CprK .......................................................................................... 45
Figure 8: Structures of NtcA .......................................................................................... 46
Figure 9: Structures of PrfA .......................................................................................... 47
Figure 10: Structural Location of PrfA* and Transcriptionally Inactivate Mutants .... 49
Figure 11: Location of Cysteine Residues in PrfA ........................................................... 50
Figure 12: Crystal Phase Diagram ................................................................................... 65
Figure 13: Hanging Drop-Vapor Diffusion Crystal Screening ........................................ 66
Figure 14: Crystallographic Basis of Bragg’s Law ............................................................ 67
Figure 15: Theory of Fluorescence Polarization ............................................................... 68
Figure 16: Representative FP Curves for DNA Binding .................................................. 77
Figure 17: “Best” GSH Binding Modes to PrfA as Calculated by Autodock Vina ........ 89
Figure 18: Thermal Denaturation Curves for PrfA and Quad .......................................... 90
Figure 19: Two-Step PrfA Mechanism Hypothesis ......................................................... 91
Figure 20: Asymmetric Unit of Apo-PrfA ...................................................................... 105
List of Abbreviations

2-OG : 2-oxoglutarate
ActA : actin polymerizing protein A
αCTD : C-terminal domain of the α subunit of RNA polymerase
ADV : AutoDock Vina
AR1 : Activating Region 1
AR3: Activating Region 3
ASU: Asymmetric unit
ATP : adenosine triphosphate
bp : base pair
BME : 2-beta mercaptoethanol
B. subtilis : Bacillus subtilis
cAMP : cyclic adenosine monophosphate
CCDs : charge coupled detectors
CO : carbon monoxide
Crp : cAMP Receptor Protein
DBD : DNA Binding Domain
DNA : deoxyribonucleic acid
DTT : dithiothreitol
ds : double stranded
E. coli : Escherichia coli
EMSA : electrophoretic mobility shift assay
FNR : Fumarate/nitrogen reductase regulator
FP : fluorescence polarization
GSH : glutathione
GSSG : oxidized glutathione
H-bonded : Hydrogen bonded
HGF : Hepatocyte Growth Factor
hly : the gene encoding LLO
Hpt : hexose phosphate transporter
(w)HTH : winged helix-turn-helix
inlA,B,C : internalin protein A, B, or C
IPTG : isopropyl β-D-1-thiogalactopyranoside
Kd : dissociation constant
lacZ : Beta-galactosidase
LIP-1 : Listeria pathogenicity island 1
lisA : original designation of hly gene
L. monocytogenes : Listeria monocytogenes
LLO: Listeriolysin O, encoded by hly (formerly lisA)
Met: receptor tyrosine kinase
mP: Millipolarization
mpl: zinc dependent metalloprotease required to process PlcB
MR: molecular replacement
NADH: nicotinamide adenine dinucleotide
NBD: no Binding Detected
NS: Non-specific (binding)
OCPA: o-chlorophenolacetic acid
PBS: phosphate buffered saline
PDB: protein database
PlcA: also PI-PLC, phosphatidylinositol specific phospholipase
PlcB: also PL-PLC phospholipase C (nonspecific)
PrfA: positive regulatory factor of lisA
PrfA*: a constitutively active PrfA mutant; most commonly G145S
RBS: ribosome binding site
RMSD: root mean squared deviation
RNA: ribonucleic acid
mRNA: messenger ribonucleic acid
RNA: PRNA polymerase
SPR: surface plasmon resonance
TCEP: tris(2-carboxyethyl)phosphine
TD: thermal denaturation
UTR: untranslated region
Acknowledgements

This document represents the culmination of effort from a large number of people who provided both scientific and personal support. First, I would like to thank my mentor Richard Brennan for his constant guidance. I would also like to thank Maria Schumacher, who provided selfless assistance at pivotal points in my research. Working with and learning from Dick and Maria, two incredible giants in the crystallography field, has been an indescribable experience.

Kenneth Kreuzer and Raphael Valdivia, the last but certainly not least members of my committee, have also been instrumental in providing guidance and alternative perspectives on my research, especially when I would become structure-focused. Kenneth provided invaluable and unique biochemical suggestions, and Raphael was initially responsible for introducing our lab to the Portnoy lab; without him, this project would have never existed.

I would like to thank all the members of the Brennan and Schumacher labs, past and present, for providing scientific discussion, constructive criticism, and experimental suggestions during stalls. Specifically I would like to thank Yi Miao, who collected the data for the apo-PrfA structure when I was out of town as well as for showing me the ropes of data collection, Alex Kovach for making jokes, and Jill Orans for teaching a novice graduate student the ropes.
I would also like to thank our excellent collaborators, the Portnoy lab at UC Berkeley. Especially, I would like to recognize Michelle Reniere who collected the GSH binding data and performed the in vivo and tissue assays. Their work and discussion has been instrumental in progressing this research, and we look forward to further collaborations.

Briefly, I would also like to acknowledge my college advisor Dr. Robin McCann, who has shaped my love of learning, teaching, and science.

Finally, I want to thank my family- my mother, Debbie, my father, Kyle, and my sister, KB- as well as my husband Tony. They have been a constant and unyielding support network, and without them, I would have never made it this far. Also, Tony has provided many a cold homebrew that has gotten me through rough days.
Chapter 1: *Listeria monocytogenes*- Virulence and Significance

1.1 The Biological Relevance of *Listeria monocytogenes*

*Listeria monocytogenes* is the main causative agent of listeriosis, a systemic bacterial tissue invasion, which is particularly dangerous to immunocompromised individuals and neonates. In adults, listeriosis generally affects the central nervous system, leading to meningoencephalitis, paralysis, sepsis, and death in about 25 -30% of victims (Gellin and Broome, 1989). In pregnant women, infection of the fetus leads to chorioamnionitis and spontaneous abortion in late term pregnancies (Vázquez-Boland et al., 2001). *L. monocytogenes* is easily spread through contaminated food, causing significant outbreaks yearly in the United States (Centers for Disease Control and Prevention, 2014; Schlech et al., 1983). The nearly ubiquitous presence of this bacterium in the environment and its danger to infants, the elderly, and the immunocompromised makes the pathogenesis of *L. monocytogenes* an important topic for scientific investigation.

1.2 *Listeria monocytogenes*: Genus and Environmental Niche

The *Listeria* genus consists of six subspecies of gram-positive non-sporulating facultative anaerobes: *L. grayi, L. innocua, L. ivanovii, L. monocytogenes, L. seeligeri, and L. welshimeri*. Of these, *L. monocytogenes* and *L. ivanovii* are the only species with an intact *Listeria* pathogenicity island (LIP-1) (Vázquez-Boland et al., 2001) and are therefore
considered the only pathogenic subspecies although other subspecies are occasionally isolated from active infections (Gellin and Broome, 1989; Vázquez-Boland et al., 2001). L. ivanovii commonly infects animals while L. monocytogenes acts as the pathogenic agent in humans (Wong and Freitag, 2004).

*L. monocytogenes* is capable of thriving in the environment in addition to its ability to replicate in human hosts. In the environmental niche, it exists as a saprotrophic bacterium and has been isolated from pasture soil, fruits (Szymczak et al., 2013), unpasteurized dairy (Osman et al., 2014), and vegetation (Welshimer and Donker-Voet, 1971). During the environmental life cycle, *L. monocytogenes* upregulates genes responsible for carbohydrate and amino acid metabolism, cell wall maintenance, and phage protection while simultaneously suppressing transcription of virulence genes required for human tissue colonization (Piveteau et al., 2011). The bacterium generally infects humans though gastrointestinal absorption of the epithelial cells in the digestive tract and is spread through contaminated food (Pamer, 2004). Upon entry into host cells, *L. monocytogenes* mediates the switch to opportunistic pathogen (Gray et al., 2006). This replicative success in two significantly different niches makes the bacterium incredibly interesting and complex.

1.3 **Mechanism of Listerial Host Cell Invasion and Virulence**

Conversion of *L. monocytogenes* from benign soil saprophyte to dangerous intracellular pathogen requires the coordinated expression of multiple virulence
proteins that mediate bacterial absorption, replication, and cell-to-cell propagation. *L. monocytogenes* is capable of replicating in various cell types including macrophages (Hanawa *et al*., 1995), epithelial cells, endothelial cells, hepatocytes, and fibroblasts (Hamon *et al*., 2006). Similar to other pathogenic bacteria, the barriers to active infection require invasion into host cells, successful proliferation, and propagation to adjacent cells. This process is summarized in Figure 1 and requires the virulence proteins InlA, InlB, LLO, PlcA, PlcB, Hpt (not shown), and ActA.

Cell entry is mediated by the internalin proteins, InlA and InlB, which are required for uptake of *Listeria* into epithelial cells (Gaillard *et al*., 1991; Lecuit *et al*., 2001; Shen *et al*., 2000; Veiga and Cossart, 2005). Enterocyte uptake is especially important for this gastrointestinally spread bacteria, and interruption of the *inlAB* regulon prevents *Listeria* entry to enterocytes and epithelial cells (Gaillard *et al*., 1991; Lecuit *et al*., 2001). InlA and InlB, both surface expressed proteins with significant sequence homology (Gaillard *et al*., 1991), interact with distinct surface exposed receptors: InlA to E-cadherin (Mengaud *et al*., 1996) and InlB to Met, a receptor tyrosine kinase whose natural ligand is Hepatocyte Growth Factor (HGF) (Shen *et al*., 2000). These interactions induce endocytosis of *Listeria* in a similar fashion to clathrin-dependent endocytosis (Veiga and Cossart, 2005).

To escape the host vacuole, *L. monocytogenes* relies on listeriolysin O (LLO), a phosphatidylinositol-specific phospholipase C (PI-PLC or PlcA), and a nonspecific
phospholipase C (PC-PLC or PlcB). LLO is a cholesterol-dependent cytolysin required for macrophage survival and productive human infection (Michel et al., 1990; Portnoy et al., 1988), and it induces small membrane perforations, which disrupt vacuole maturation and lead to increased chance of bacterial escape (Shaughnessy et al., 2006). This escape is facilitated by a decrease in membrane integrity generated from lipase cleavage by PlcA and PlcB (Camilli et al., 1993; Goldfine et al., 1993). Listeria deficient in PlcA or PlcB have decreased vacuolar escape and deficient cell-to-cell spreading (Smith et al., 1995), indicating this membrane weakening is necessary to assist the bacterium in escaping into the host cytosol. Interestingly, LLO has also recently been implicated in suppressing the oxidative burst from NADPH oxidase, an antimicrobial activity in phagocytes (Lam et al., 2011), which may contribute to Listerial survival during endocytosis.

Once Listeria has escaped the host vacuole and entered the cytosol, it upregulates a hexose phosphate transporter (Hpt), a homolog to the mammalian glucose-6-phosphate translocase that allows the bacteria to import glucose-1-phosphate, glucose-6-phosphate, fructose-6-phosphate and mannose-6-phosphate from the host (Chico-Calero et al., 2002). Hpt is required for efficient and rapid replication of the bacteria, and mutation of Hpt leads to severe attenuation of proliferation and virulence (Chico-Calero et al., 2002).
As bacterial load in the cell increases, *Listeria* then induces cell-to-cell transfer allowing the infection to propagate. Intercellular spread depends on the actin polymerizing protein, ActA (Domann *et al.*, 1992), the mechanism of which has been extensively studied and summarized elsewhere (Hamon *et al.*, 2006; Ireton, 2013; Pizarro-Cerdá *et al.*, 2012; Vázquez-Boland *et al.*, 2001). The brief mechanistic overview of ActA induced cell-to-cell spread is that ActA is phosphorylated by a serine/threonine kinase (Chong *et al.*, 2009) which allows for activation of Arp2/3-dependent actin nucleation (Boujemaa-Paterski *et al.*, 2001; Zalevsky *et al.*, 2001) that mobilizes *Listeria* through the host cytosol and into adjacent cells through membrane protrusions (Domann *et al.*, 1992; Kocks *et al.*, 1992). The spreading process is facilitated by the internalin InlC, a secreted protein that promotes weakening of epithelial apical junction integrity and increases the productivity of ActA generated protrusions (Rajabian *et al.*, 2009). Deletion of ActA from *Listeria* prevents actin accumulation (Domann *et al.*, 1992) and inhibition of the debranching of actin polymers subsequently prevents intercellular propagation of *Listeria* (Talman *et al.*, 2014).

While the above list of proteins involved in infection is not exhaustive, and there are additional proteins that are expressed, the internalins InlA and InlB, listeriolysin, phospholipases A and B, Hpt, and ActA are all essential for virulence. The deletion of any of these proteins significantly decreases or completely abolishes the infectivity of *L. monocytogenes* as well as attenuating the severity of the disease. While each individual
protein would therefore be an attractive therapeutic target, it is important to note all described virulence proteins are under the transcriptional control of the same regulator, PrfA. To understand *Listerial* conversion to its human-pathogen form, we must delve into the mechanism of genetic control by PrfA, the single transcriptional control point for these critical virulence proteins.
Figure 1: Infectivity Cycle of *Listeria monocytogenes*

The infectivity cycle of *Listeria monocytogenes* in the human host in pictorial depiction (center) and micrographic images (outside). *L. monocytogenes* is endocytosed to endothelial cells through InlA and InlB interactions with surface proteins. LLO and PlcA/PlcB degrade the host vacuole, allowing for bacterial escape to the host cytosol. Finally, ActA induces actin polymerization for cell-to-cell spread, generating the famous “comet tails” visible by microscopy. (Figure by Portnoy *et al.*, 2002).
Chapter 2. PrfA is the Transcriptional Regulator that Mediates Virulence in *Listeria monocytogenes*

In 1990, an avirulent and listeriolysin-negative strain of *Listeria* was generated by interrupting a gene upstream of *hly* (also named *lisA*, the gene encoding LLO). This was found to be a 27 kDa transcriptional regulator that induced expression of LLO and was subsequently named positive regulatory factor of listeriolysin (*lisA*) or PrfA (Leimeister-Wachter *et al.*, 1990). Further experimentation indicated PrfA was also required for expression of PlcA and PlcB (Chakraborty *et al.*, 1992), a zinc dependent metalloprotease termed Mpl required for processing PlcB to its mature form (Chakraborty *et al.*, 1992; Poyart *et al.*, 1993), ActA (Freitag *et al.*, 1993; Sheehan *et al.*, 1995), Hpt (Chico-Calero *et al.*, 2002), and InlC (Engelbrecht *et al.*, 1996). PrfA also plays a role in *inlA* and *inlB* transcription, although these proteins are controlled by multiple promoters of which only one is PrfA-dependent (Lingnau *et al.*, 1995). The deletion or mutation of PrfA from *L. monocytogenes* leads to significant attenuation in virulence, deficiencies in vacuole escape, and prevention of actin nucleation and cell-to-cell spreading (Chakraborty *et al.*, 1992; Leimeister-Wachter *et al.*, 1990; Freitag *et al.*, 1993).

The *prfA* gene itself is located on the *Listeria* pathogenicity island 1 (LIPI-1) and regulates LIPI-1 genes (*plcA, hly, mpl, actA, plcB*) in addition to chromosomal genes (*inlAB, inlC*, and *hpt*) (Vázquez-Boland *et al.*, 2001). Untangling the web of PrfA-mediated gene expression has been relatively complex considering *Listeria monocytogenes* needs to coordinate gene expression in separate unique environments. In addition to
PrfA-dependence, many virulence genes have additional promoter elements, placing them under regulation of alternative sigma factors. A summary of the PrfA regulon can be seen in Figure 2, where black boxes and dashes represent gene promoter elements, and the details of virulence promoter regulation are discussed below.

### 2.1 The PrfA Box: PrfA-Dependent Gene Activation

The first indication of transcriptional regulation in *Listeria* was the identification of a 14 basepair (bp) palindromic sequence in the upstream promoter region of *hly* and the then designated open reading frames ORF D and ORF U (later discovered *plcA* and *mpl*) (Mengaud *et al.*, 1989). Further experimentation indicated that PrfA was a DNA binding protein that recognized the palindrome in a sequence specific manner. Base pair mutations that disrupted the sequence or removal of the upstream regulatory element abolished PrfA regulation (Freitag *et al.*, 1993, 1992). This conserved palindromic sequence, located in the -41.5 region upstream of the transcriptional start site, was found to exist in all of the virulence genes regulated by PrfA (Mengaud *et al.*, 1991) with varying degrees of conservation and was thus deemed the “PrfA box” to represent the critical point of transcriptional control. A list of the PrfA boxes of the *Listerial* virulence genes discussed in Chapter 1 shows that the boxes are either perfectly palindromic (*Phly* or *PplcA*) or deviate slightly (Table 1). The PrfA boxes at promoters *Pmpl*, *PactA*, *PinlC* and *Phpt* have single nucleotide mismatches, while *PinlA* and *PprfA* contain two or more deviations from the ideal palindrome.
2.2 Deviation from the Consensus PrfA Box: Not All PrfA-Dependent Promoters are Equal

Deviations in the PrfA box of the various virulence genes was an early observation, and the first hypotheses proposed that the regulator interacted differently with each, giving the protein an additional layer of control over virulence expression instead of simply being “on” (upregulating protein expression) or “off” (inactive) (Freitag et al., 1993). This proposal sought to rationalize how LLO was expressed early in Listeria infection (cellular uptake and escape) while ActA was expressed later in infection (to stimulate intercellular transfer). Thus it was hypothesized that the deviation from the consensus PrfA box would allow the proteins to be temporally regulated.

To analyze transcription from PrfA-dependent promoters, promoter-lacZ fusions in B. subtilis were created to elucidate the variations in expression. Beta-galactosidase (LacZ) is a protein that hydrolyzes glycosidic bonds and can be used to quantify gene expression, thereby measuring activation or inhibition. Fusions of Phly, PplcA, PactA, Pmpl, and PinlA to a lacZ reporter gene indicated that when PrfA was introduced in B. subtilis, PplcA and Phly upregulated expression the earliest and with substantial increases in protein production. PactA and Pmpl had intermediate increases in protein production and were activated later in infection, while PinlA had poor activation throughout the time course of the experiment (Sheehan et al., 1995). This led to the proposal that PplcA was the strongest promoter, followed closely by Phly, and both were
much more potent than *PactA* (which was about equal to *Pmpl*). *PinlA* is additionally regulated by other non-PrfA factors (Bohne *et al.*, 1996; Sheehan *et al.*, 1995), and thus unlikely to be fully dependent on PrfA for activation during infection. Further experimentation demonstrated that replacing *Phly* with *Pmpl* significantly reduced reporter gene expression, and mutating *Pmpl* to *Phly* significantly increased protein production in the presence of PrfA (Williams *et al.*, 2000), supporting the established hierarchy of promoter strength.

Interestingly, mutating the *actA* promoter region to contain a *Phly* PrfA box was not sufficient to increase LacZ expression in *B. subtilis* (Williams *et al.*, 2000). This, coupled to the poor induction of *PactA* in *B. subtilis*, may indicate that an additional activating factor is required in *Listeria* as ActA protein production is upregulated more than 200-fold in active *Listerial* infections, once the bacterium has invaded the host cell (Bohne *et al.*, 1994; Brundage *et al.*, 1993).

The PrfA box deviations from the consensus motif set up an intrinsic mechanism for control that allows virulence expression to progress temporally with infection. As *L. monocytogenes* enters the host cell, a small amount of PrfA can activate production of listeriolysin and phospholipase early, allowing for vacuolar escape and release into the host cytoplasm. As infection progresses, the increase in PrfA concentration from the *PplcA* promoter (discussed below in Section 2.2.3) increases the expression of late stage proteins, including Hpt, ActA, and InlC. InlA, which initiates endocytosis, is likely
maintained by alternative promoters so it is present when \textit{Listeria} encounters endothelial cells.

\textbf{2.3 PrfA Transcriptional Control by Alternative Sigma Factors, Thermal Regulation, and Autoregulation}

The concentration of PrfA is critical for mediating gene expression through interaction with the different PrfA boxes of the \textit{Listerial} virulence genes, and thus its intracellular concentration is tightly regulated. PrfA can be expressed from three separate promoters: \textit{PrfAP1}, \textit{PrfAP2}, and \textit{PplcA}. Both \textit{PrfAP1} and \textit{PrfAP2} generate monocistronic mRNAs (Freitag and Portnoy, 1994), while the PlcA promoter (\textit{PplcA}) allows for read-through through a predicted Rho-independent terminator to produce a bicistronic PlcA-PrfA mRNA (Camilli \textit{et al.}, 1993) (Figure 2).

Transcription from the three PrfA promoters are controlled in three mechanistically different ways. The first is through the association of multiple sigma (\(\sigma\)) factors to the promoter region. Association of \(\sigma\) to the core RNA polymerase (RNAP) allows cells to coordinate gene expression in different environments. Both \textit{PrfAP1} and \textit{PrfAP2} are recognized by \(\sigma^A\), the primarily expressed sigma factor of RNAP under active non-stressed growth conditions. \textit{PrfAP2} is additionally regulated by \(\sigma^8\), an alternative sigma factor associated with growth in stress conditions (Nadon \textit{et al.}, 2002; Ollinger \textit{et al.}, 2009; Rauch and Goebel, 2005).

Secondly, \textit{PrfAP1} has a thermoregulation switch encoded in the 5’ untranslated region (UTR) of the mRNA transcript. When the mRNA is formed, the 5’ UTR has a
sequence that can anneal into a hairpin that obscures the ribosome binding site (RBS) and is stable below 30 °C (Johansson et al., 2002). When the bacteria enters the host cell and the environmental temperature increases to 37 °C, the hairpin melts, exposing the RBS and allowing for rapid PrfA translation.

Coupling the thermoswitch with $\sigma^A/\sigma^B$ control in the monocistronic transcription promoters maintains a basal pool of prfA mRNAs, which can be rapidly translated upon host cell entry. Indeed this is critical, as it has been shown that PrfA expression from PrfAP1 and PrfAP2 generates a sufficient protein concentration to promote vacuole escape but is insufficient for propagation or cell-to-cell spread (Freitag et al., 1993; Mengaud et al., 1991).

In addition to sigma factor recognition and thermal control, PrfA also has two autoregulatory steps to mediate its own expression: a non-canonical PrfA box at PrfaP2 (PprfA in Table 1) and a perfect palindromic PrfA box at PplcA. Production of sufficient protein quantities for actin polymerization and intercellular spread is dependent on the bicistronic plcA-prfA transcript, and deletion of this regulatory sequence prevents the progression of infection (Camilli et al., 1993; Mengaud et al., 1991). As shown in Table 1 and Figure 2, the PrfA box upstream of PplcA is a perfect palindrome and strongly increases the concentration of PrfA (Mengaud et al., 1991). The PrfA box at PrfAP2 likely works in a negative autoregulatory mechanism, as, instead of increasing basal transcription of PrfA, there is evidence that PrfA represses transcription at this promoter.
(Freitag and Portnoy, 1994; Freitag et al., 1993). It is hypothesized this weak repression may play a role in the saprotrophic lifecycle instead of contributing to infection as abolishing repression is dispensable for virulence (Greene and Freitag, 2003).

Combining these mechanisms of control over PrfA transcription shows how intricately *Listeria monocytogenes* regulates virulence gene expression. Using costly nutrients to transcribe and translate unnecessary virulence genes in soil and plant environments would be detrimental to the bacterium. By maintaining a low basal concentration of PrfA mRNAs, which are held in limbo by their thermoswitch, and a low concentration of the protein under control of the $\sigma^A$-dependent promoters, *L.* *monocytogenes* can be prepared to upregulate the PrfA regulator in response to vacuolar endocytosis. As the bacterium enters the host cell and undergoes stress, an increase in temperature and presence of $\sigma^B$ can increase the concentration of PrfA, beginning the cycle of auto-upregulation from the *PplcA* promoter and, by extension, the infection cycle.

As infection progresses from bacterial endocytosis to propagation and spread, the relative consensus of the PrfA box dictates when and how strongly additional virulence genes are expressed.

### 2.4 Post-Translational Control of PrfA Activity

While it might be assumed that the presence of PrfA leads to immediate up-regulation of dependent genes, translation of the protein does not necessarily produce
an active transcriptional regulator as soon as the protein departs from the ribosome. In addition to transcriptional and translational control, PrfA is regulated post-translationally.

The evidence for the PrfA “switch” to turn on gene activation was proposed when PrfA purified from *Escherichia coli* (*E. coli*) needed supplementation with PrfA-free cell extract from *Listeria monocytogenes* in order to bind DNA in gel electrophoretic mobility shift assays (EMSAs) (Böckmann *et al.*, 1996). Experimentation in hyperhemolytic and hyperlecithinase (overly expressed PlcB) *Listeria* strains led to the discovery of a constitutively active PrfA with a glycine to serine mutation at amino acid 145 (PrfA* or G145S) (Ripio *et al.*, 1997b). The G145S mutant was the first to identify an on/off state conformational change that likely occurred through small molecule binding during infection. Subsequently additional constitutive activators were elucidated by further work with PrfA (Miner *et al.*, 2008; Shetron-Rama *et al.*, 2003; Vega *et al.*, 2004; K. Wong and Freitag, 2004; Xayarath *et al.*, 2011b), and while there was some evidence that iron (Böckmann *et al.*, 1996), activated charcoal (Ripio *et al.*, 1997a, 1997b), and growth in different media (Bohne *et al.*, 1994) affected PrfA activity, all hypotheses over small molecule activation remained mostly conjecture based on the structure (solved in 2005 (Eiting *et al.*, 2005)) and the protein’s similarity to cAMP receptor protein (Crp) of *E. coli*. The structural research and additional insight into PrfA activation will be discussed in Chapter 3.
2.5 PrfA Independent Virulence Regulation in Listeria monocytogenes

While PrfA regulation described above is an exquisitely controlled method of gene activation allowing for protein expression at different stages of infection, there are additional levels of control that increase the complexity of transcription. Regulation of PrfA expression by the Listeria sigma factors $\sigma^A$ and $\sigma^B$ has already been discussed in section 2.3, but other virulence genes also have PrfA-independent mechanisms of control.

InlA and InlB can be produced from a bicistronic mRNA or from monocistronic messages. While InlB is fairly strongly dependent on the presence of PrfA, InlA has two PrfA-independent promoters in addition to the PrfA-dependent promoter that are capable of expressing InlA in extracellular media. A PrfA-independent promoter also exists to generate the monocistronic InlB mRNA (Lingnau et al., 1995; Sheehan et al., 1995). Both InlA and InlB are controlled by $\sigma^B$ (Greene and Freitag, 2003; Kim et al., 2005). The listeriolysin gene hly additionally has a weak PrfA-independent promoter that expresses low, but significant, amounts of LLO in PrfA-deficient cells (Domann et al., 1993).

This additional level of regulation makes logical sense for Listeria virulence as InlA, InlB and LLO are temporally early stage virulence proteins required for cell-mediated entry. While PrfA would be expressed and begin autoregulation once the bacterium experienced the host temperature change, keeping low concentrations of InlA
and, to a lesser extent, InlB and LLO constitutively present would allow for immediate endocytosis and potential vacuole escape while PrfA mRNA was translated, preventing the bacterium from being passed benignly before efficient activation of productive infection.

While the mechanism of regulation for PrfA association to the PrfA-box of various virulence genes is relatively well elucidated from transcriptional assays throughout the 1990s and early 2000s, the post-translational factor that switches PrfA from inactive protein to transcriptional activator of virulence has remained elusive. Early studies attempted to look at various complex formations on EMSAs of PrfA, DNA and the mysterious “activating ligand” (Dickneite et al., 1998), however the mechanism and ligand identity have as yet remained unidentified. Chapter 3 will explore the insight offered by structural studies on PrfA and mutational effects on DNA binding.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Sequence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phly</td>
<td>TTAACA TT TGTTAA</td>
<td>(Sheehan et al., 1995)</td>
</tr>
<tr>
<td>PplcA</td>
<td>TTAACA AA TGTTAA</td>
<td>(Sheehan et al., 1995)</td>
</tr>
<tr>
<td>Pmpl</td>
<td>TTAACA AA TGTTAA</td>
<td>(Sheehan et al., 1995)</td>
</tr>
<tr>
<td>PactA</td>
<td>TTAACA AA TGTTAG</td>
<td>(Sheehan et al., 1995)</td>
</tr>
<tr>
<td>Phpt</td>
<td>ATAACA AG TGTTAA</td>
<td>(Chico-Calero et al., 2002)</td>
</tr>
<tr>
<td>PinlC</td>
<td>TTAAGG CT TGTTAA</td>
<td>(Engelbrecht et al., 1996)</td>
</tr>
<tr>
<td>PinlA</td>
<td>ATAAACA TA AGTTAT</td>
<td>(Sheehan et al., 1995)</td>
</tr>
<tr>
<td>PprfA</td>
<td>CTAACA AT TGTTGT</td>
<td>(Sheehan et al., 1995)</td>
</tr>
<tr>
<td>Consensus</td>
<td>TTAACA NN TGTTAA</td>
<td></td>
</tr>
</tbody>
</table>
The PrfA operon including chromosomal genes \textit{inlA}, \textit{inlB}, \textit{inlC} and \textit{hpt} and LIPI-1 genes \textit{prfA}, \textit{plcA}, \textit{hly}, \textit{mpl}, \textit{actA}, \textit{plcB}, and \textit{orfX}. The PrfA boxes are indicated by black boxes, and PrfA independent promoters are represented by black lines. The \textit{PrfAP1} and \textit{PrfAP2} promoters are labeled 1 and 2 (note the presence of a non-canonical PrfA box at \textit{PrfAP2} not indicated by a black box). The mono- and multi-cistronic mRNA messages are indicated by dotted lines under the expressed gene.
Chapter 3. Insight into the PrfA Activating Ligand

The PrfA control point of virulence genes is incredibly important for transforming Listeria monocytogenes from benign soil saprophyte to intracellular pathogen. If a small molecule does mediate the switch from inactive to active PrfA, this would be ideal for novel targeted therapeutics, which could compete and prevent PrfA activation. As mentioned in Chapter 2, PrfA has an inactive form that is incapable of binding DNA unless supplemented with some mysterious cofactor present in Listerial lysate (Böckmann et al., 1996). Later discovery of the constitutively active PrfA mutant, G145S or PrfA*, further supported this switch from inactive to active regulator (Ripio et al., 1997b) although the lysate supplied ligand was never identified.

Speculation on mechanism has emerged based on several aspects of the protein: a) the relation to E. coli cAMP receptor protein (Crp), b) the elucidated structure of PrfA and PrfA G145S, and c) the effects of point mutations on DNA binding and \textit{in vivo} virulence.

3.1 PrfA is Related to E. coli Crp

Initially, PrfA was assigned to the Crp family of transcriptional regulators by sequence alignments, which showed PrfA from \textit{L. monocytogenes} and \textit{L. ivanovii} shared homology with the Crp-FNR family of global transcriptional regulators (Lampidis et al., 1994). This homology was confirmed when a solution structure of LexA, a DNA damage response repressor, demonstrated structural similarity to Crp as PrfA also
shares sequence homology with LexA (Holm et al., 1994). An amino acid sequence alignment of Crp from *E. coli*, NtcA from *Synechococcus elongatus* (a Crp family member), and PrfA from *L. monocytogenes* and *L. ivanovii* is shown in Figure 3.

Before analyzing PrfA as a Crp-FNR transcriptional regulator, it would be beneficial to analyze the known mechanisms of Crp gene activation (the canonical family member) as well as several other structurally characterized proteins, including CooA, CprK, and NtcA, which deviate slightly from the canonical family mechanism.

### 3.1.1 The Canonical Crp Mechanism

Cyclic AMP receptor protein (Crp) or catabolite gene activator protein (CAP) is a transcriptional regulator that activates over one hundred promoters in the *E. coli* genome in the presence of the allosteric ligand cAMP by increasing the transcription efficiency of RNA polymerase (RNAP). Crp activation has been extensively studied and summarized in other articles (Lawson et al., 2004; Savery et al., 1996; Won et al., 2009), as has new insight into the differences in Crp homologues in different bacterial species (Green et al., 2014). This section will briefly address the important mechanistic aspects of *E. coli* Crp that correlate to PrfA.

The structure of Crp bound to cAMP (McKay and Steitz, 1981; Passner et al., 2000) revealed the protein is a homodimer with two domains: an N-terminal cAMP binding region and a C-terminal winged helix-turn-helix (wHTH) DNA binding domain (Figure 4A). The N-terminal domain is comprised of eight anti-parallel β-strands
flanked by two α-helices. The dimer is coordinated through the long dimerization helix αC, leading to the C-terminal DNA-binding region, which is composed of three α-helices and two sets of two anti-parallel β-strands. In most of the currently solved crystal structures, Crp is in the active conformation with cAMP bound in the N-terminal effector site, however an NMR solution structure and apo-Crp crystal structure indicate the DNA-binding domain is conformationally flexible when the ligand is absent and that the cAMP-induced DNA binding requires a large conformational shift (Popovych et al., 2009; Sharma et al., 2009). The C-terminal HTH motif interacts as a dimer with a symmetric DNA site termed the Crp-box (Ebright et al., 1989).

Crp has additionally been crystallized in complex with DNA containing the Crp box (Figure 4B) (Parkinson et al., 1996; Passner and Steitz, 1997; Passner et al., 2000) as well as in complex with DNA and the αCTD of RNAP (Figure 4C) (Benoff et al., 2002). These structural complexes combined with biochemical and biophysical experiments have significantly contributed to understanding the mechanistic nuances of Crp-dependent gene activation. Briefly, cAMP binds the N-terminal domain of Crp leading to the formation of an activated Crp-cAMP complex (Parkinson et al., 1996; Passner and Steitz, 1997; Passner et al., 2000) with an increased affinity for DNA containing the Crp box (Fried and Crothers, 1984). The binding of cAMP to Crp causes the protein to induce a significant bend in the DNA through two kinks, totaling an overall bend angle of 87° to 90° (Parkinson et al., 1996; Passner and Steitz, 1997). The rearrangement of the
DNA positions Crp residues 156 – 164 (also known as activating region 1 or AR1, labelled in Figure 4C) to interact with the αCTD of RNAP, which leads to effective recruitment of the RNAP holoenzyme and an increase in gene activation (Zhou et al., 1993).

Crp-dependent promoters can be split into two separate classes that depend on the position of the Crp box in relation to the transcription start site (Busby and Ebright, 1999). Class I Crp-dependent promoters have a Crp box centered at -61.5 bp upstream of the transcription start site and solely interact with RNAP through the AR1. Class II Crp-dependent promoters have the Crp box centered around -41.5 bp upstream of the transcription start site, overlapping the -35 promoter element, and contain additional protein-protein interactions to the N-terminal domain of the α-subunit and the σ subunit of RNAP. The regions on Crp that interact with αNTD and sigma are termed the AR2 and AR3, respectively (Niu et al., 1996; Rhodius and Busby, 2000; Lawson et al., 2004).

Figure 5 shows schematic models of the fully assembled class I and class II promoter complexes at Crp-dependent promoter sites.

### 3.1.2 Important Residues in Crp

There are three categories of residues critical for Crp to function as a transcriptional activator: cAMP binding residues, the DNA binding/bending residues, and the AR1. The AR1 is critical for protein-protein interactions with RNAP, and
mutation of the amino acids 156 – 162 will generate Crp that is still capable of binding DNA but incapable of activating transcription (Zhou et al., 1993).

The cAMP binding pocket is located in the N-terminal domain between the β-barrel and the dimerization α-helix. Amino acid residues Glu81, Gln125, Ser83, and Thr127 in the N-terminal domain are critical for coordinating cAMP interaction, as are several water molecules in the binding pocket (Passner et al., 2000).

In the Crp-cAMP-DNA structure, the residues Arg180, Glu181, and Arg185 were noted as being responsible for DNA base contact specificity, and it had been discovered previously that mutation of Glu181 leads to a preference for a different Crp box sequence (Ebright et al., 1984). Additionally, Lys22, Lys26, Lys44, Lys166, His199, and Lys201 have been indicated as important residues for electrostatic interactions that stabilize the DNA bend when Crp-cAMP binds DNA (Hardwidge et al., 2002; Kapanidis et al., 2001; Schultz et al., 1991). These residues will be important to remember while discussing the structure of PrfA bound to DNA in later chapters.

### 3.1.3 Other Crp/FNR Transcriptional Regulators

*E. coli* Crp is an extremely well characterized member of the Crp/FNR family of transcriptional regulators, however additional members of this global regulator family function similarly but distinctly from Crp.
3.1.3.1 *E. coli* FNR

The fumarate and nitrogen reduction regulator (FNR) is responsible for allowing *E. coli*, a facultative anaerobe, to utilize nitrate and fumarate as final electron acceptors and thus respire in anaerobic conditions. Structurally, FNR is similar to Crp with an N-terminal β-barrel domain, dimerization helix, and C-terminal wHTH motif that recognizes a dyadic DNA sequence (Mazoch and Kucera, 2002). Similarly to Crp, FNR is active in the dimeric form, which allows binding to DNA and gene expression, however the dimeric form is stabilized by a [4Fe-4S]$^{2+}$ cluster, coordinated by four critical cysteine residues. Exposure to oxygen leads to the dissolution of the [4Fe-4S]$^{2+}$ cluster to a [2Fe-2S]$^{2+}$ cluster, which causes FNR to adopt a conformation that stabilizes the monomer and results in a loss of DNA binding affinity (Beinert and Kiley, 1999; Crack *et al.*, 2014). This deviation from cAMP recognition shows the versatility of ligand recognition that the overall family structure is capable of, as well as highlighting that key residues can vary in the family and significantly alter the protein function.

3.1.3.2 *Rhodospirillum rubrum* CooA

Composition of the atmosphere in the environment can be important for signaling protein expression, especially when bacteria need to coordinate oxygen and nitrogen fixing enzymes. Being able to rapidly respond to an oxygen-rich or depleted environment can be critical for survival. The same is true of other gaseous molecules in the atmosphere including carbon monoxide, ethylene, or nitric oxide (Aono, 2003).
Rhodospirillum rubrum is a photosynthetic bacteria that can grow in the presence of carbon monoxide (CO) and the absence of oxygen (Kerby et al., 1995), and one of the critical transcriptional regulators, CooA, is responsible for sensing CO availability to coordinate expression of the coo operon which encodes a CO dehydrogenase and hydrogenase (Aono et al., 1996; Shelver et al., 1997). While CooA is a member of the Crp/FNR family, it uniquely has a heme cofactor to sense CO binding. Biochemical experiments suggest in a reducing environment, the heme is coordinated by His77 and Pro2 of the adjacent monomer (Nakajima et al., 2001). Binding of CO to the heme group displaces the proline, leading to a large conformational change that allows CooA to bind DNA and activate expression of CO metabolizing proteins.

This hypothesis is supported by the current available structures of CooA: an inactive, CO-free form (Figure 6A) (Lanzilotta et al., 2000), a non-DNA binding intermediate in which imidazole has displaced coordination of the proline (Figure 6B) (Komori et al., 2007), and an activated mutant without CO bound (Figure 6C) (Borjigin et al., 2007). The inactive, reduced, CO-free structure of CooA indicates that the heme binds the N-terminal effector binding domain in the same localized region as the cAMP binding pocket of Crp, and again is coordinated by His77 and Pro2 (Lanzilotta et al., 2000). While the protein is dimeric in the absence of CO, the DNA binding domains are asymmetric with the wHTH motif of one subunit almost straight in orientation to the N-terminal domain (Figure 6A, wHTH motif orientation represented by arrows).
(Lanzilotta et al., 2000). Neither subunit is arranged in a way that would be permissive to DNA binding, however the difference between the two domains indicates significant flexibility in the DNA-binding domain.

When imidazole binds and displaces Pro2, partially mimicking the binding of CO to the heme, large overall structural changes occur including a 30° rotation of the heme and a significant shift of the DNA binding domains. Instead of being in line with the N-terminal domain, the wHTH motif is now perpendicular to the effector domain in both subunits, a conformation that more closely resembles the DNA-bound Crp structure. However, this shift does not align the recognition helices in a position capable of recognizing DNA and represents what is likely an intermediate (and inactive) conformation (Komori et al., 2007). The constitutively active CooA mutant (N127L/S128L) gives insight into the final active conformation (Figure 6C). The N-terminal tail that would otherwise occupy the heme pocket relocates between the heme and DNA binding domains, stabilizing the active DNA binding conformation, and the recognition helix of the wHTH motif rotates forward in an orientation that would permit DNA interaction (Figure 6D) (Borjigin et al., 2007). It is interesting to note that in the active CooA structure, only one subunit is in the active conformation. The second subunit has not bound the heme cofactor, and the DNA binding domain is disordered.

CooA is a unique addition to the Crp family. Like FNR, it utilizes a cofactor (the heme) to recognize its activating ligand (CO), however CooA contains an additional N-
terminal extension that appears to be required for stabilization of the DNA binding
domain. The CooA DNA binding domain also demonstrates the flexibility and ability to
sample structural conformations of these regulators.

3.1.3.3 *Desulfitobacterium hafniense* CprK

*D. hafniense* is a strict anaerobe that is capable of utilizing halogen compounds as
final electron acceptors in respiration (halorespiration). More specifically, they utilize
the reductive degradation of ortho-chlorophenol. The proteins responsible for this
dehalogenation are found at the *cpr* locus and are regulated by CprK, a Crp/FNR-type
regulator. Similarly to the mechanism of Crp, o-chlorophenolacetic acid (OCPA) acts as
the activating ligand that binds CprK causing an increase in DNA binding affinity for a
palindromic region of the promoter dubbed the “dehalo box”, leading to the increased
transcription of downstream genes (Gábor et al., 2006; Joyce et al., 2006; Pop et al., 2004).
The structure of CprK bound to OCPA (Joyce et al., 2006; Levy et al., 2008) shows the
ligand binds to the homodimer in the (now familiar) location that is spatially correlated
to the cAMP-binding pocket of Crp and the heme-CO binding region in CooA (Figure
7A).

CprK is also one of two Crp/FNR family regulators with a structure in complex
with cognate DNA (Levy et al., 2008) (Figure 7B). In agreement with the Crp
mechanism, the active CprK-OCPA complex bends the dehalo box DNA in a significant
angle to mimic the Crp-cAMP-DNA structure, which would likely place CprK in proximity to αCTD for recruitment of RNAP.

The interesting aspect of CprK that distinguishes it from *E. coli* Crp is the use of a redox sensor in addition to the modulation of DNA binding induced by OCPA. The determination of multiple crystal structures of reduced and oxidized CprK has helped elucidate the mechanism of CprK gene activation (Levy et al., 2008). In this model, Cys11 and Cys200 are located proximally, which allows for the formation of a disulfide bond under oxidizing conditions. The protein can be oxidized in both the ligand-free and ligand-bound form, and in both cases, the disulfide prevents rearrangement of the DNA binding domain to associate with the dehalo box, subsequently preventing gene activation in the presence of oxygen (Joyce et al., 2006; Levy et al., 2008). This disulfide bond formation adds an extra layer of control so CprK is only transcriptionally active in a reducing environment, assuring that costly downstream genes are only transcribed under the necessary conditions.

### 3.1.3.4 Cyanobacteria NtcA

Nitrogen control factor (NtcA) is a global transcriptional regulator in cyanobacteria that controls expression of enzymes required for nitrogen assimilation, the most important being nitrogen permease and glutamine synthetase, which are expressed in response to low levels of ammonium (Herrero et al., 2004).
NtcA recognizes the NtcA box with a consensus palindrome sequence GTA(N)sTAC (Luque et al., 1994). Similarly to class II Crp-dependent promoters, the NtcA box is generally localized around -41.5 bp upstream of the transcription start site, and promoter expression varies as NtcA binds with higher affinity to the perfectly palindromic NtcA boxes and less tightly to those with nucleotide mismatches (Forcada-Nadal et al., 2014). Intriguingly, in a deviation from the Crp model, NtcA is capable of binding its target DNA sequence in the absence of its activator ligand, 2-oxoglutarate (2-OG), a small molecule that signals nitrogen levels in cyanobacteria (Muro-Pastor et al., 2001), although 2-OG does increase the binding affinity of NtcA for its target promoters, especially in NtcA boxes that deviate from the ideal palindrome (Forcada-Nadal et al., 2014). It is important to note that experimental evidence indicates that NtcA is transcriptionally inactive in the absence of 2-OG (Tanigawa et al., 2002), which may indicate that the DNA-binding conformation and the transcriptionally active conformation are not mutually exclusive in this protein.

Several crystal structures of NtcA in the apo and 2-OG bound form have been obtained (Llacer et al., 2010; Zhao et al., 2010) (Figure 8). In the apo form (Figure 8B), the DNA binding domains of NtcA are widened in a conformation that prevents association with the DNA promoter. The binding of 2-OG to the β-barrel of the N-terminal effector domain (Figure 8A) in roughly equivalent location to cAMP in Crp leads to a structural
rearrangement to the active conformation and overlays well with the bent DNA form in the Crp-cAMP-DNA crystal structure (Llacer et al., 2010).

NtcA further deviates from the Crp model of regulation by the presence of a protein activator, the PII interacting protein, PipX. PipX is a small 89 residue protein that binds to the NtcA-2-OG complex and increases gene expression further than the ligand bound active form, although in conditions of high nitrogen concentration, PipX is sequestered by PII to prevent this interaction (Camargo et al., 2014; Llacer et al., 2010). A crystal structure of the PipX-NtcA-2-OG complex (Llacer et al., 2010) aligned to the Crp-cAMP-DNA-αCTD crystal structure from E. coli (Figure 8C) indicates that PipX may be located proximally to the αCTD, which could stabilize the interaction with RNAP and thus explain the increase in transcription efficiency.

NtcA is an especially interesting look at a global regulator with multiple levels of post-translational control for gene activation including DNA binding without gene activation, transcriptional activation by a small cofactor, and additional transcriptional activation by the protein binding partner PipX.

3.2 The Structures of PrfA and Constitutively Active PrfA*

The structures of PrfA and the constitutively activated mutant G145S (PrfA*) were determined in 2005, supporting the hypothesis that PrfA belonged to the Crp-FNR family of transcriptional regulators (Eiting et al., 2005). Like Crp, PrfA has an N-terminal domain consisting of eight anti-parallel β-sheets and two α-helices connected to a
wHTH motif through a long dimerization helix. The C-terminal DNA binding domain has six α-helices and four anti-parallel β-sheets, which represents a significant extension to the C-terminal domain that has no structural equivalent in Crp.

The structures of PrfA (Figure 9A) and PrfA* (Figure 9B, aligned to PrfA) are essentially the inactive vs active conformation for this regulator. While the overall structure of inactive (or “apo”) PrfA is remarkably similar to PrfA*, there is a significant shift in the C-terminal HTH helix αE, as well as a stabilization of the αF in the HTH motif, so that PrfA* aligns more closely with Crp-DNA (Figure 9C). If the structures of apo-PrfA, G145S PrfA, and Crp-DNA are aligned, the wHTH motif of the apo-PrfA structure is in an orientation that would clash with the DNA backbone (Figure 9D). The wHTH motif of G145S is better oriented to bind DNA, likely representing the active binding conformation PrfA would adopt in complex with its ligand in order to associate with the PrfA box (Eiting et al., 2005).

In occasional fortuitous circumstances, proteins will co-purify with a tightly binding ligand, and sometimes the crystal will reveal the ligand once the structure has been determined. While this was not the case for PrfA, the binding cavity is important to discuss from the determined structure. PrfA does not have conserved key cAMP binding residues in comparison to Crp (Eiting et al., 2005; Vega et al., 1998), however the overall structure of the binding pocket is similar as seen in CooA, CprK, and NtcA. In the G145S structure, the buffer reductant DTT (dithiothreitol) was found disulfide
bonded to Cys229 and, in two subunits of the asymmetric unit, was localized at the entrance of the ligand-binding cavity. DTT was found in proximity to the channel in the apo-PrfA structure as well (Eiting et al., 2005). This solvent-exposed binding pocket does suggest that like other Crp family proteins, PrfA might be able to bind a small molecule that could signal its conformational change.

3.2.1 Mutations in PrfA and Effects on DNA Binding and Listerial Virulence

The association of PrfA with the Crp/FNR family of transcriptional regulators led to an exhaustive search for the activating cofactor that initiates the shift from the inactive apo-PrfA to the active conformation represented by PrfA G145S. While the idea of a necessary activator was hypothesized in 1996 (Böckmann et al., 1996), no actual identity for the molecule has been presently proposed. Instead, mechanistic details about PrfA activation have proceeded in the form of cause and effect: creating protein point mutants and evaluating the effect on DNA binding and in vivo virulence.

These mutations generally can be categorized as either virulence gene overexpressers (PrfA* mutants) or transcriptionally inactive mutants, which either do not bind DNA or do not cause infection. A summary of PrfA mutants can be found in Table 2 and are highlighted on the wild-type structure in Figure 10.

Most of the PrfA* and inactive PrfA mutants have been assessed for several characteristics: their ability to activate PrfA-dependent genes (typically actA and plcA/hly) in vivo and in vitro, their ability to bind PrfA-box containing promoters (usually
*PactA* and *Phly*), the ability to dimerize, virulence in cell tissue and murine models, and the ability to form RNAP-PrfA-DNA complexes (which can be visualized by EMSA).

While ideally, these mutations would hint at ligand identity and subsequent PrfA activation, the combination of mutants is relatively informationally ambiguous. They do not localize to any particular key area, and mutation of the same residue (Y63C vs Y63W) or neighboring residues (S183A vs S184A) can lead to either over-expressing or inactivated mutants.

Most often, PrfA* mutants simply increase DNA binding affinity which leads to an increase in virulence and only demonstrates the protein is now in the active conformation, for example, I45S (Vega *et al.*, 2004), Y63C (Miner *et al.*, 2008), E77K (Miner *et al.*, 2008), L140F (Miner *et al.*, 2008), G145S (Eiting *et al.*, 2005), and S183A (Sheehan *et al.*, 1996), which does not impart mechanistic information in addition to the G145S structure solved in 2005 (Eiting *et al.*, 2005). A unique case of PrfA activation appears to be the P219S mutant, which significantly increases virulence gene expression with only a moderate increase in DNA binding affinity, however limited proteolytic digestion suggests the mutant adopts a conformation similar to G145S (Xayarath *et al.*, 2011a).

The inactivated PrfA mutants provide similarly vague mechanistic information. Again, most mutants decrease or abolish DNA binding including L48F (Deshayes *et al.*, 2012), K64Q (Xayarath *et al.*, 2011b), K122Q (Xayarath *et al.*, 2011b), K130Q (Xayarath *et al.*, 2011b), S184A (Sheehan *et al.*, 1996), R188I (Herler *et al.*, 2001), A218* (Herler *et al.*, 2011b),
2001), K220T (Velge et al., 2007), which naturally would decrease virulence as PrfA requires DNA binding to activate gene expression. In the case of M58I, the mutation is located in a region that maps to the AR3 region of Crp, which is responsible for interacting with the sigma subunit of RNAP, and subsequently is deficient in formation of the RNAP-PrfA-DNA complex (Herler et al., 2001). While this mutant does not attest to a small molecule cofactor, it does provide supporting information for mechanistic similarity to Crp.

The potentially most interesting inactive PrfA mutants are Y63W and Y154C. Y63W is located in what is predicted to be the ligand binding pocket in the N-terminal β-barrel of PrfA and is deficient in actin polymerization and accumulation, however the DNA binding affinity was comparable to WT in surface plasmon resonance (SPR) binding assays (Deshayes et al., 2012). This potentially could indicate the mutation interferes with ligand binding and subsequent gene activation, although it does not explain why only actin polymerization would be affected with no measurable effect on LLO production.

The mutation Y154C is located on the surface of the protein near the anti-parallel β-sheet in the C-terminal region of the protein. This mutation causes a slight increase in DNA binding affinity, however the protein is deficient in virulence with a decrease in LLO expression and a temporal delay in actin accumulation and polymerization. Listeria containing the Y154C PrfA show two-fold attenuation in mouse models of virulence
(Miner et al., 2008). Unfortunately, this mutant was not assessed for its ability to form
RNAP-PrfA-DNA complexes, so it can only be hypothesized that the protein is in some
way deficient in gene activation.

PrfA mutants M58I, Y63W and Y154C likely represent the only mutants in the
literature that are activation-deficient mutants, however they unfortunately do little to
contribute to a detailed mechanistic model of PrfA activation aside from highlighting
protein regions that were already hypothesized to contain importance. The mutants do
seem to suggest that PrfA activation is complex and in need of a more detailed approach
than scanning mutational analysis to elucidate the molecular mechanism of activation.

3.2.2 Mutation of the Four Cysteine Residues of PrfA to Alanine

PrfA has four solvent accessible cysteines (highlighted in yellow on apo-PrfA
and PrfA*, Figure 11 A and B), which is unusual for a cytosolic protein, and the crystal
structure of G145S (PDB ID: 2BGC (Eiting et al., 2005)) has DTT disulfide bonded to
C229, which hints that the cysteines may be available to form disulfides in solution.
Mutation of all four cysteines to alanine (Quad PrfA or Quad) yields a mutant that is
poorly auto-upregulated, induces less ActA expression than wild-type, and is 30-fold
less virulent than wild-type PrfA in murine infection models (Reniere et al., 2015). This
suggests the cysteine residues in PrfA are important for gene expression, although it is
unclear how. The residues do not align with the mechanistically important residues in
FNR or CprK and are not proximal to each other based on the crystal structure, ruling
out the coordination of an undiscovered iron-sulfur cluster, such as in the case of FNR, or intramolecular disulfide mechanism, such as in CprK. Additionally, the decrease in virulence instead of complete avirulence of Quad indicates the cysteines are not likely activated through modification. It has been noted that DNA binding can be affected by iron concentration (Böckmann et al., 1996), which may indicate change in the sulfur oxidation state plays a role in gene activation.

### 3.3 Is Glutathione the Small Molecule Activator of PrfA?

Mutational studies on PrfA have culminated in the hypothesis that PrfA has two states - active and inactive (each structurally determined in 2005 (Eiting et al., 2005)) - and that the conversion of the inactive to the active conformation is likely induced by a negatively charged small molecule co-factor (Xayarath et al., 2011b) that fits into the predicted ligand tunnel in the β-barrel region of the N-terminal (Deshayes et al., 2012; Eiting et al., 2005; Xayarath et al., 2011b) domain of the protein similar in location to the cAMP binding pocket of Crp, the heme binding pocket of CooA, the OCPA binding pocket of CprK, and the 2-OG binding pocket of NtcA.

The first breakthrough on the identity of the PrfA-activating ligand emerged from the Portnoy lab in UC Berkeley. They found that mutations inactivating glutathione (GSH) synthase, an enzyme that synthesizes the low molecular weight thiol used as a reductant in *Listeria monocytogenes* and humans, could rescue a suicide bacterial strain controlled by the *actA* promoter (Reniere et al., 2015). Deletion of *gshF*,
the *Listerial* gene encoding GSH synthase (Gopal et al., 2005), resulted in poor cell-to-cell spread caused by the absence of ActA expression and infection that was 150-fold less virulent in mouse models. Interestingly, this phenotype could be rescued by the supplementation of G145S, the constitutive PrfA* mutant (Reniere et al., 2015). These results suggest that GSH is necessary for the activation of PrfA, and once PrfA is constitutively active, GSH synthesis is expendable.

The requirement of GSH for PrfA-dependent gene activation is incredibly exciting in the PrfA field of research as it is the first instance of a potential identity for the small molecule required for transcriptional activation. It is even more plausible, given that a molecule of DTT, a small thiol containing chemical, was found near the ligand channel in the PrfA crystal structures. The main questions then become how is GSH acting to induce transcription, and what role do the cysteine residues play in this mechanism?

With a candidate for the small molecule activator responsible for PrfA-dependent gene regulation, important mechanistic details remain to be elucidated as *Listeria* move from soil saprophyte to intracellular pathogen. There are three potential models for this activation:

1). PrfA mimics the Crp-cAMP mechanism through GSH allosteric binding. In this model, PrfA would be unable to bind DNA in its inactive conformation. GSH would then bind allosterically in the predicted ligand tunnel, causing a conformation
shift to the DNA-permissive form (resembling the G145S structure), which would then allow for RNAP recruitment to the gene promoter and efficient transcription. It is unclear what the function of the cysteine residues are in this model, although they could be important for additional RNAP interactions.

2). Through a mechanism more akin to CprK, GSH forms a mixed disulfide with one or more of the cysteine residues, glutathionylating the protein and causing the shift to the DNA-binding conformation that allows efficient transcription. This model may have been suggested by the coordination of DTT to C229 in the G145S structure (Eiting et al., 2005).

3). Some combination of 1 and 2.

In either case, GSH could also play an additional role in coordinating interactions with RNAP and assist in the formation of the final RNAP-DNA-PrfA complex. This also adds the question as to whether the DNA of the (PrfA-GSH)-DNA complex would adopt the significant bend of the Crp-cAMP-DNA or CprK-OCPA-DNA complex, which allows for the interaction between the AR1 and the α-CTD of RNAP.

The overall goal of this research will be to clarify the mechanistic details of PrfA as a transcriptional activator from a structure-function approach. X-ray crystallography will be employed to elucidate the structures of the PrfA-DNA complex and the Quad PrfA structure. Additionally, biochemical assays will be used to validate the obtained structural models. Fluorescence polarization will assess the DNA binding capability of
Quad and WT PrfA in the presence and absence of GSH. Other biochemical assays including bio-layer interferometry and thermal denaturation will suggest if GSH is binding as a ligand or forming a mixed disulfide with one or more cysteine residues. This combination of structural models and biochemical data will help elucidate the mechanism of PrfA-mediated gene activation in *Listeria monocytogenes*. 
Figure 3: Crp Family Alignment

Sequence alignment of *E. coli* Crp, *S. elongates* NtcA, *L. monocytogenes* PrfA, and *L. ivanovii* PrfA (PrfA_1).
Figure 4: Structures of Crp

Structure models of A Crp-cAMP (PDB: 1G6N), B Crp-cAMP-DNA (PDB: 2CGP), and C Crp-cAMP-DNA complexed with αCTD (PDB: 3N4M). In A, the α-helices of the N- and C-terminal domains are labelled, and cAMP is shown as a stick model. In B, the location of the primary and secondary kinks induced in DNA-bending are noted. The two subunits are colored in different shades of red. C, the AR1 region of Crp (red) interacts with αCTD (purple).
Figure 5: Crp-Dependent Promoter Assembly

The overall architecture of Class I and Class II Crp-dependent promoters. In Class I promoters, Crp binds the Crp box around the -61.5 bp region of the promoter and contacts the αCTD of RNAP through the AR1. In Class II promoters, the Crp box is localized around the -41.5 region relative to the promoter and contacts the αCTD through the AR1 region on the opposing side of the protein when compared to Class I. Additionally, Crp contacts the N-terminal domain of the RNAP α-subunit through the AR2 region and the sigma subunit through the AR3 (not shown). (Figure adapted from Meng et al., 2000).
Figure 6: Structures of CooA

Structures of CooA. A inactive (PDB: 1FT9), B inactive with bound imidazole (PDB: 2FMY), and C active mutant (PDB: 2HKX), also rotated 90° (D). The orientation of the DNA binding domains is represented by the blue arrows. Imidazole binding (B) causes a significant shift in the orientation of the wHTH. In C, the recognition helix and β sheets (circled) have shifted. D is a structural alignment of the imidazole structure with the active mutant, demonstrating the rearrangement that occurs to position the DNA binding domain in a binding accessible conformation.
Figure 7: Structures of CprK

The structure of A OCPA-bound CprK (PDB: 3E5X) and B CrpK-OCPA-DNA (PDB: 3E6C).
The structure of NtcA A bound to 2-OG (PDB: 2XHK) and B structurally aligned to the active structure with the apo protein (dark purple, PDB: 3LA7). The active structure shows a slight widening in the DNA binding domain (indicated by the purple arrows). C An alignment of active NtcA in complex with PipX (light green) to the Crp DNA with bound αCTD (purple) show PipX exists in a conformation to act as a bridge from NtcA to the αCTD (PDB: 2XKO and PDB: 3N4M)
Figure 9: Structures of PrfA

Structures of A apo-PrfA (PDB: 2BEO), B G145S (dark green aligned to apo-PrfA, PDB: 2BGC), C G145S aligned to Crp-DNA (red, PDB: 2CGP). The alignment of apo-PrfA, G145S, and Crp-DNA complex (D) indicates the recognition helix of PrfA in the apo conformation would clash with the DNA backbone and is likely in a non-DNA binding conformation (red arrow).
Table 2: PrfA* and Transcriptionally Inactive PrfA Mutants

<table>
<thead>
<tr>
<th>PrfA*</th>
<th>Inactive PrfA</th>
</tr>
</thead>
<tbody>
<tr>
<td>I45S</td>
<td>L48F</td>
</tr>
<tr>
<td></td>
<td>(Vega et al., 2004)</td>
</tr>
<tr>
<td>Y63C</td>
<td>M58I</td>
</tr>
<tr>
<td></td>
<td>(Deshayes et al., 2012)</td>
</tr>
<tr>
<td>E77K</td>
<td>Y63W</td>
</tr>
<tr>
<td></td>
<td>(Deshayes et al., 2012)</td>
</tr>
<tr>
<td>L140F</td>
<td>K64Q</td>
</tr>
<tr>
<td></td>
<td>(Xayarath et al., 2011b)</td>
</tr>
<tr>
<td>G145S</td>
<td>K122Q</td>
</tr>
<tr>
<td></td>
<td>(Xayarath et al., 2011b)</td>
</tr>
<tr>
<td>G155S</td>
<td>K130Q</td>
</tr>
<tr>
<td></td>
<td>(Xayarath et al., 2011b)</td>
</tr>
<tr>
<td>S183A</td>
<td>Y154C</td>
</tr>
<tr>
<td></td>
<td>(Miner et al., 2008)</td>
</tr>
<tr>
<td>P219S</td>
<td>S184A</td>
</tr>
<tr>
<td></td>
<td>(Sheehan et al., 1996)</td>
</tr>
<tr>
<td></td>
<td>R188I</td>
</tr>
<tr>
<td></td>
<td>(Herler et al., 2001)</td>
</tr>
<tr>
<td></td>
<td>K220T</td>
</tr>
<tr>
<td></td>
<td>(Velge et al., 2007)</td>
</tr>
<tr>
<td></td>
<td>A218*</td>
</tr>
<tr>
<td></td>
<td>(Herler et al., 2001)</td>
</tr>
</tbody>
</table>
Figure 10: Structural Location of PrfA* and Transcriptionally Inactivate Mutants

Location of PrfA* and transcriptionally inactive mutants on the wild-type PrfA structure. PrfA* mutants are green, and transcriptionally inactive mutants are colored red. Tyr63, which shows both a PrfA* and transcription-inactive phenotype, is colored in pink.
Figure 11: Location of Cysteine Residues in PrfA

Location of the four cysteine residues (yellow) in A apo-PrfA and B G145S PrfA
Chapter 4. Elucidation of PrfA-Mediated Transcription Mechanism Using X-Ray Crystallography and Biochemical Assays

4.1 An Overview of X-Ray Crystallography

As should be evident from the collection of Crp, CooA, CprK, and NtcA structures in complex with their subsequent activating ligands, protein structural determination can be an incredibly powerful tool to model mechanistic detail at the molecular level. However, generation of a structural model for biomolecules is dependent on the formation of high quality protein crystals, which is no simple feat.

Protein crystals are formed by the ordered precipitation of protein into a crystalline lattice, which is stabilized by noncovalent intermolecular interactions. Crystals can form as the protein changes solubility in a precipitation solution (shown by the crystal phase diagram in Figure 12). In the crystal phase diagram, a protein is mixed with a precipitant (polyethylene glycol for example) and starts at an unsaturated concentration. Through slow water evaporation, the protein and precipitant concentration will increase until the protein reaches the nucleation stage. At this stage, protein-protein association reaches a critical point where molecules associate more than they dissociate, which generates a basis for growth. The crystals slowly start to grow as more molecules are added to the nucleate, decreasing the concentration of the protein so that it reaches the growth phase. In this phase, no new nucleation steps occur. Instead, protein molecules associate with established crystal nuclei until association and
dissociation stabilize. If the researcher is lucky, the crystal becomes large enough to be subjected to X-rays to search for conditions that diffract well. It is critical in crystal formation that the protein remains in its native folded state, otherwise disordered precipitate will form.

While science would be simplified if there was one pre-established precipitant mix for growing protein crystals, this remains a dream for the future. Phase separation, denatured protein precipitate, salt crystals, or nothing are all possible results of a crystal screen set-up, and thus high throughput grid screening is used to vary the precipitant, buffer, and additional counter ions in an exhaustive search for diffraction quality crystals.

A common form of screening is hanging drop-vapor diffusion (Figure 13). In this type of screen, the protein (in a buffered solution) is mixed with the precipitating solution and suspended over that solution in a sealed environment (most commonly on a glass cover slip suspended over a well in a multi-well plate). Since the concentration of the precipitating solution is diluted by mixing with the protein, water evaporation occurs to equilibrate the concentration in the suspended drop to that of the well. This evaporation slowly increases the protein concentration, ideally leading to the formation of a well-ordered crystal.

Once a promising protein crystal has been formed, the next step is to optimize the crystal volume and evaluate cryo-protectant solutions, which help reduce damage
due to X-ray irradiation, with the goal of obtaining the highest resolution diffraction.
While inevitably crystal formation is the first step to determining a structural model, obtaining and interpreting crystal diffraction is a critical step.

4.1.1 Diffraction from the Crystal Lattice

When a protein crystal is exposed to X-rays, the electrons of the amino acids in the protein scatter the X-rays, yielding discrete spots that can be captured by any one of several devices including imaging plates and charge coupled detectors (CCDs). These images are referred to as the crystallographic diffraction pattern and can be utilized to discern the electron density composition of the crystal. However, understanding the intrinsic symmetry of the protein crystal is critical for interpreting the diffraction pattern.

Protein crystals pack in a repeating array of molecules in three-dimensional space (the crystal lattice). The smallest repeating subunit of the crystal is defined as the unit cell, composed of lengths a, b, and c and angles α, β, and γ. In most cases, the unit cell contains multiple protein molecules related to each other by symmetry elements including translation and rotation. These symmetry elements define the crystal space group, and the asymmetric unit (ASU) is the smallest structure needed to reconstruct the unit cell with the symmetry elements described by the space group. Unit cells that contain no internal symmetry are triclinic (space group P1) but are rarer.
Crystal space groups are conventionally written as the lattice type with numbers that represent the intrinsic symmetry operations. For example, a crystal in space group P422 is a primitive lattice (all indices shared by neighboring unit cells) with four-fold rotational symmetry along the c-axis, which is the principle axis to this space group, and two-fold rotational symmetry around the b- and a-axes. A screw axis, which is generated by rotation followed by translation of the molecule, would be indicated by a subscript on the symmetry element. The space group P43212, for example, has two screw axes. Along the 43 screw axis (parallel to the c-axis), the protein would be rotated 90° and translated three-fourths the length of the unit cell. Along the 21 screw edge, the molecules would be rotated 180° and translated half the length of the unit cell.

The unit cell can be divided by planes which are given numerical indices (h,k,l). These are known as Miller indices. Each index describes the number of parts created by the plane on the a, b, and c cell edges respectively. These parallel planes are responsible for the diffraction pattern of a crystal when the X-rays are scattered in a constructive pattern as described by Bragg’s law (Figure 14) and is represented mathematically by the equation

$$2d_{hkl}\sin\theta = n\lambda$$

In short, when parallel planes with indices (h,k,l) have an interplanar spacing of d (distance between A1 and A2), the X-rays will only be in phase and therefore constructive if the two incident beams (ray 1, R1, and ray 2, R2) are whole integer values
of \(\lambda\). If \(n\) is not an integer value, the emerging X-rays (R1 and R2) will interfere destructively leading to lower or non-detectible diffraction and therefore a weak or missing corresponding spot in the diffraction pattern. It is also important to note that the intensity of the emerging beam depends on the electron density in this plane as the X-ray scattering is increased by higher electron density.

The X-ray scattering from the unit cell is generated by all contributing atoms and yields a complex wave that represents all \((n)\) atoms in the unit cell. For a discrete spot on the X-ray detector, the contributions to that spot can be represented as a structure factor, \(F\), defined mathematically as a Fourier summation of all atoms \((1 \ldots n)\). In addition to the atom location contribution, the type of atom contributes as well through the atomic scattering factor \((f_j)\) which is determined by the atom identity (for example, scattering factor of carbon will be different compared to oxygen). The structure factor equation is

\[
F_{hkl} = \sum_{(j=1)} f_j e^{2\pi i (hx_j + ky_j + lz_j)}
\]

Calculated structure factors can then be used in a Fourier sum to generate electron density \((\rho)\) representing atomic distribution throughout the cell for any given \((x,y,z)\) coordinate in a unit cell of volume \(V\)

\[
\rho(x,y,z) = (V^{-1}) \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i (hx + ky + lz)}
\]

However, there is an inherit snag in the generation of structure factors from experimental X-ray diffraction. Structure factors must include amplitude and phase to
describe the scattered waves generated by the unit cell, however the diffraction
measurements yield only the intensity (and by direct relation, amplitude). Since no lens
is capable of focusing an X-ray beam yet, the phase information of the emitted X-ray is
lost which prevents calculation of structure factors directly from a single native crystal
diffraction experiment. This ambiguity in phases is the famous crystallographic “phase
problem”, which needs to be solved before electron density can be modelled.

4.1.2 Using Molecular Replacement to Solve the Phase Problem

There are several ways to determine wave phases in order to calculate structure
factors. de novo phasing methods allow for experimental determination of initial phases
for calculating electron density and often use heavy atom labelling (seleno-methionine
labelling or heavy atom soaking with mercury or platinum). However this method can
be difficult as the addition of heavy atoms can perturb the crystal symmetry and
degrade diffraction.

Another method of obtaining the initial phases is by using the structure(s) of a
similar protein that has already been determined. In this case, the phases of the model
are generated for the native data set of the unsolved protein crystal structure in a
method called molecular replacement (MR). Generally, the best search models have
>20% sequence identity to the protein of interest and usually a similar overall structure.
While this method of phase determination is considered experimentally simpler, it is
limited by the availability of proteins that have similar structures. MR methods are
usually ideal in cases where the structure of interest is a mutant, has different ligand complexes of an already structurally solved protein, or for protein families.

The largest obstacle to determining the phases for the native dataset from the search model is orientation in the unit cell. Since the atomic positions generate the phases, the search model needs to superimpose on the protein of interest. Knowing the unit cell space group allows for imposition of required symmetry operators, however the exact location of the target model is still unknown. MR then requires a brute force approach that tries multiple positions and orientations in the unit cell. This step is completed usually through a computer program such as those found in Phenix (Adams et al., 2010) or in CCP4 (Project, 1994; Winn et al., 2011). The computer program will calculate structure factors for each search position with the amplitudes provided from the target diffraction data, yielding phases from the best matching model.

4.1.3 The Translation and Rotation Functions

The computer memory and power required for determining the proper position and orientation of the search model in the unit cell simultaneously would be immense. In order to reduce computer system demands, MR programs separate the position and orientation searches into two parts. First, the program determines the proper orientation using projected Patterson maps (in essence, a Fourier sum without phases). Patterson maps are independent of the \((x,y,z)\) coordinates of the unit cell and depend instead on orientation. This fact means that Patterson functions can be calculated from the
intensities obtained from experimental data and can decouple orientation from location. Therefore, a rotation function, \( R(\Phi, \phi, \chi) \), calculates predicted Patterson maps from different orientations of the MR model and chooses the best option based on the Patterson map that most closely matches the experimental.

The rotation function value, \( R(\Phi, \phi, \chi) \)

\[
R(\Phi, \phi, \chi) = \int_{u,v,w} P_{\text{target}}(u,v,w) P_{\text{model}}[(u,v,w) \times [\Phi, \phi, \chi]] \, du \, dv \, dw
\]

is calculated as the MR model is rotated over angles \( \Phi, \phi, \chi \) to produce the Patterson functions for the target \( P_{\text{target}}(u,v,w) \) and the MR model \( P_{\text{model}}(u,v,w) \). The higher the number of coinciding peaks in the Patterson maps, the higher the value for \( R(\Phi, \phi, \chi) \) will be, and thus the best orientation will be the maxima when orientations are plotted vs the rotation function value.

Once a best orientation has been calculated by the MR program, this new model can be used in a translational search to properly place the orientation in the unit cell, which is commonly evaluated by either an R factor or standard linear correlation coefficient. The standard linear correlation coefficient is a calculation that measures agreement between observed and calculated intensities or amplitudes. The mathematically simpler R factor measures the relation between calculated and expected structure factors (given the phases) with the equation

\[
R = \left( \sum |F_{\text{obs}}| - |F_{\text{calc}}| \right) / \left( \sum |F_{\text{obs}}| \right)
\]
Ideally, the observed and calculated structure factors would be identical and thus R would equal zero, however MR searches are rarely ideal. Therefore, the smallest value of R usually indicates the best-predicted model location in the unit cell.

Once a MR model has been placed in the unit cell, it is important to analyze the predicted crystal packing generated by application of the space group symmetry operations. The protein placement generated by MR is a calculated prediction, and it is critical for the researcher to verify that the packing is viable and that the protein subunits are not occupying the same space, which is physically impossible. Once the MR prediction has been verified by assessing the crystal packing, structural refinement can proceed.

### 4.1.4 Model Refinement

Once phases have been obtained from MR, structure factors and subsequently an electron density map can be calculated to give an overall look at the electron distribution in the unit cell. After this, a structural protein model is fit into the electron density and evaluated, utilizing known chemical parameters including atomic bond lengths, Ramachandran allowed peptide backbone angles, and most-likely side chain rotamers to guide protein model building in a biologically realistic fashion. The process of improving and evaluating the structure is known as refinement. It is an iterative process that constantly seeks to improve the molecular model and match it to the electron density map.
Visually, electron density (2Fo-Fc) and electron density difference maps (Fo-Fc) are excellent methods of monitoring refinement progression. The Fo–Fc map can be used to view the electron density map and model together. It gives peaks of positive and negative density where the model is either underbuilt (more model is needed for the density that is observed) or overbuilt (a model overfills the electron density), allowing for modification as necessary. The 2Fo–Fc electron density map shows the full distribution of the electron density. However, in the early stages of structure determination this map is highly biased from the search model and care must be taken not to allow this bias to persist.

The quality of the structural model is evaluated using R-factors: R_work and R_free. The R-factor equation is the same as that given in Section 4.1.3 and evaluates the deviation from the experimental structure factors based on the model (with R = 0 meaning perfect agreement to experimental data). The R factor for the entire model is referred to as the R_work.

R_free is a value used for measuring model bias and over interpretation of the map. It is calculated using the same equation as R_work, however it only samples a small percentage of the diffraction data excluded from refinement (usually 5 – 10%). This method of assessing how well the model predicts the diffraction intensities of an unrefined section is used as a validation tool to assess bias introduced during refinement.
Structure refinement is an iterative process that steadily improves the protein model through the use of tools such as simulated annealing omit maps, which can assist in highlighting bias introduced from the MR search model, with the intention of building a structure that best represents the diffraction data. In a good structural model, $R_{\text{work}}$ and $R_{\text{free}}$ will be within 5% of each other, and $R_{\text{free}}$ will be under 30% (or 0.3). However, this is not an unyielding rule as other factors influence both $R_{\text{work}}$ and $R_{\text{free}}$, resulting in higher values and larger percentage differences. Post refinement validation programs, such as Molprobity, can evaluate the reliability of the structural model.

### 4.2 Assessment of DNA-Binding Affinity Using Fluorescence Polarization

As a transcriptional regulator, the ability of PrfA to bind to DNA is critical to its ability to activate downstream gene expression. Previous experiments have used two methods for detecting PrfA binding to DNA: electrophoretic mobility shift assays (EMSAs) and surface plasmon resonance (SPR).

In an EMSA experiment, a protein of interest is mixed with DNA at a concentration that is less than the binding constant ($K_d$), electrophoresed on a polyacrylamide or agarose gel, and the DNA migration is detected with a radiolabeled DNA probe. If the protein has bound the DNA, it will cause retardation in the DNA migration (Carey et al., 2013; Kerr, 1995). While EMSAs can be a powerful tool for DNA-binding, it is not an equilibrium technique, and thus DNA can dissociate from the protein during electrophoresis, leading to an undetectable complex (Hellman and Fried,
In light of this limitation, EMSA may not accurately represent DNA binding affinity as it occurs in solution.

Surface plasmon resonance, or SPR, measures molecular interactions in real-time by monitoring the change in the refractive index of a molecule immobilized on a metal film as it interacts with other molecules in solution (for example, an immobilized protein interacting with DNA in solution). SPR experiments can yield quantitative information on DNA binding, however the DNA needs to be minimal in size and usually some characterization of protein-DNA interaction needs to be assessed by EMSAs or DNase footprinting before a productive SPR experiment is undertaken (Majka and Speck, 2007). Additionally, SPR can be poor at differentiating non-specific interactions, and the instruments are costly (Rich and Myszka, 2004).

In this work, fluorescence polarization (FP) will be utilized to quantitatively measure DNA-binding. It is an equilibrium-based method like SPR, however offers a less expensive alternative. The theory behind FP is presented in Figure 15. Briefly summarized, FP experiments measure polarized light emitted from the fluorophore attached to the 5’ or 3’ end of a dsDNA oligomer. Light is filtered through a monochromater then polarized and passed through a sample test tube containing the DNA (Figure 15A). The parallel and perpendicularly polarized light intensity is read and expressed by the formula

\[
P = \frac{(I_\parallel - I_\perp)}{(I_\parallel + I_\perp)}
\]
When DNA, a relatively small molecule, is alone in solution, it tumbles relatively evenly so the intensity of perpendicularly and parallel polarized light is close to equal. However, as a protein, which has a relatively larger molecular weight than the dsDNA, is titrated into the sample, and binding to the DNA occurs, the rotation of DNA and, by extension, the attached fluorophore is slowed (Figure 1B). This results in an increase of polarization that ultimately can be used to calculate a binding affinity of the protein for the DNA (Lundblad et al., 1996) by fitting the following equation to a plot of polarization vs protein concentration:

\[
mP = \frac{((mP_{\text{bound}} - mP_{\text{free}})[\text{protein}])}{(K_d + [\text{protein}])} + mP_{\text{free}}
\]

where \( mP \) is the millipolarization measured at a given protein concentration, \( mP_{\text{free}} \) is the initial millipolarization of free fluorescein-labelled DNA, \( mP_{\text{bound}} \) is the maximum millipolarization of specifically bound DNA, and \([\text{protein}]\) is the protein concentration.

FP experiments are invaluable for assessing the change in DNA binding affinity when amino acid residues are mutated, in the presence and absence of predicted coeffectors, or changes in the environment. As an example, FP was used to determine the affinity of the bacterial antibiotic resistance regulator, MtrR, to different DNA lengths and in changing concentrations of sodium chloride (Hoffmann et al., 2005).

In the context of PrfA, DNA binding affinity measured via FP-based experiments will help elucidate if GSH increases the protein's affinity for DNA (as cAMP does for
Crp), and if mutation of the cysteine residues leads to deficiency in association of PrfA with cognate DNA.
The crystal phase diagram illustrates how protein crystals grow. They start in an unsaturated phase, then as water evaporation occurs, the protein and precipitate concentration increases until nucleation can occur. As protein concentration decreases, a crystal moves to a growth phase. If the precipitant or the protein concentration is too high, the protein can become disordered aggregates or precipitate, which is not useful for X-ray crystallography.

Figure 12: Crystal Phase Diagram
A typical set up for hanging drop-vapor diffusion screening experiment. Multiple different precipitating solutions are screened in a multi-well plate. The protein and precipitating solution are mixed and suspended on a glass coverslip over the precipitating solution, which allows for water diffusion to change the concentration of the protein and precipitant, hopefully generating a crystal or good lead.
Two Miller planes, separated by a distance of $d_{hkl}$ will allow constructive interference of the “reflected” X-ray beams if the length of beam 1 ($R_1$) is an integer value of beam 2 ($R_2$). If the beams are not integer values, they will be destructive and no emerging X-rays will be determined. This is the basis of Bragg’s Law.
The instrumental and experimental theory behind fluorescence polarization. A The instrumental set up for a fluorescence polarization experiment which allows for the measurement of parallel (P∥) and perpendicularly polarized (P⊥) light. B A fluorophore is attached to DNA, which becomes polarized as a large protein molecule binds, slowing the tumbling of the DNA in solution. The change in light polarization is measured as shown in A.
Chapter 5. PrfA Binding to Cognate DNA

Since DNA binding is integral to the predicted PrfA mechanism, developing an assay to measure PrfA affinity for its cognate DNA was critical. Previously, EMSAs and SPR experiments had been used to determine PrfA affinity for Phly, however, for reasons described in Chapter 4.2, we constructed a DNA binding assay based on fluorescence polarization. Our experiments measured the affinity of PrfA for Phly (which previously had been measured with reported $K_d$ of 200 nM (Sheehan et al., 1996) and 900 nM (Eiting et al., 2005)) and PactA (reported as having lower binding affinity than Phly based on reporter fusion assays (Williams et al., 2000)). We additionally measured the affinity of G145S for Phly (reported nanomolar affinity (Eiting et al., 2005)) and PactA and of Quad PrfA for both Phly and PactA promoters. Affinities for all proteins were then measured in the presence of glutathione.

5.1 DNA Binding Results

Wild-type PrfA binding to promoter DNA varied by oxidation state. The reduced protein bound Phly with nanomolar affinity ($K_a = 34.2 \pm 4.8$ nM) and a three-fold diminished affinity for PactA ($K_a = 96.4 \pm 26.3$ nM). This binding was similar in the presence of three different reductants: tris(2-carboxyethyl)phosphine (TCEP), 2-beta-mercaptoethanol (BME), and DTT. When the protein was purified in the absence of reductant or allowed to air-oxidize (cysteine oxidation measured by Ellman’s reagent; data not shown), the DNA binding affinity for Phly decreased to the previously reported
Kₐ of 900 nM by Eiting et al., 2005 (measured 888.5 ± 140.3 nM). My data deviate from the literature results as PrfA has not previously been reported to be capable of binding DNA in the absence of its predicted ligand.

The constitutively active G145S mutant was also found to bind tightly to Phly with an affinity of 40.8 ± 3.3 nM which agrees well with the SPR literature value (Eiting et al., 2005). G145S bound with similar affinity to the PactA promoter (Kₐ of 45.4 ± 3.2 nM) despite the nucleotide mismatch.

Interestingly, the Quad mutant, which exhibits a defective infection phenotype in mouse models (Reniere et al., 2015), bound Phly with similar affinity to the G145S mutant and the reduced protein (Kₐ = 32.8 ± 5.5 nM). The affinity for PactA was three-fold lower compared to G145S, however matched the affinity of reduced PrfA (124.9 ± 26.3 nM). Results of FP binding assays are summarized in Table 3. Representative FP binding isotherms and their fitted curves are shown in Figure 16.

Intriguingly, the addition of GSH up to 2 mM in the FP binding buffer had no effect on DNA binding for PrfA (oxidized or reduced), G145S, or Quad. Higher concentrations of GSH could not be assessed due to the tendency of GSH to associate with the DNA, resulting in fluctuations in the polarization values.

5.2 Discussion

Analyzed in the context of Crp gene activation, the binding results of PrfA to Phly and PactA appear to indicate a deviation from the expected canonical mechanism.
If PrfA were to mimic the Crp binding, the protein would be less capable of binding DNA except in the presence of its activating ligand, however the reduction of the cysteine side chains was sufficient to initiate DNA-binding for both Phly and PactA. The addition of up to 2 mM GSH to the binding buffer had no effect on DNA binding for WT PrfA (either oxidized or reduced), a further deviation from the Crp model if GSH is indeed the small molecule activator. These results may indicate that PrfA is more similar to NtcA from cyanobacteria as NtcA can bind promoter DNA but not initiate transcription in the absence of its activating ligand, 2-OG (Muro-Pastor et al., 2001; Tanigawa et al., 2002).

It is not clear how PrfA oxidation interferes with DNA binding. Oxidation of the cysteine sulfhydryl groups may lead to one of two states: intermolecular disulfides (which could interrupt the natural dimer formation) or the formation of oxygenated-sulfur species such as sulfenic (R-SOH), sulfinic (R-SOOH) or sulfonic (R-SO$_2$OH) acid (Kim et al., 2015). Either modification could potentially interfere with DNA-binding by preventing the formation of the active conformation. This may indicate that PrfA is less active during endocytosis as the environment of the vacuole is oxidizing. Once the bacteria escapes to the reducing environment of the host cytosol, PrfA could become activated, although the change in oxidation state of PrfA through infection has not been addressed *in vivo.*
Another unexpected result is the ability of Quad PrfA to bind both \textit{Phly} and \textit{PactA} with affinities that resemble both G145S (slightly tighter for \textit{Phly} and three-fold weaker for \textit{PactA}) and reduced PrfA (comparable within experimental error). Almost all of the previously identified PrfA mutants with deficient infection phenotypes (Y63W and Y154C excluded) have resulted from a loss of DNA binding affinity. Quad PrfA is not deficient in DNA-binding, which may indicate that the cysteines, in addition to being important for DNA binding, may also assist in coordinating efficient RNAP interactions. It is unlikely that binding of Quad to DNA was mediated by reductants mimicking the binding ligand as the mutant was similarly capable of binding DNA in non-reducing buffer conditions.

Since glutathione, albeit at a relatively low concentration, had no measurable effect on DNA binding, the FP assay does not illuminate how GSH influences PrfA-mediated gene activation, although there are several options. The first is that reduced PrfA is the transcriptionally active form of PrfA. Since GSH is the major low-molecular weight thiol in \textit{Listeria monocytogenes} (Newton \textit{et al.}, 1996), it could potentially be required to maintain reduced PrfA during infection. If glutathione synthase were removed from the bacteria, \textit{Listeria} may not be able to maintain the necessary reduced state, leaving the regulator inactive, unable to bind DNA, and deficient for gene activation.
An alternative hypothesis is that GSH can bind PrfA as a ligand. In this model, transcriptionally active PrfA and DNA-binding PrfA would be separate and distinct entities. This would be described by a two-step mechanistic hypothesis where (1) in the first step, PrfA becomes reduced and capable of DNA binding but cannot interact efficiently with RNAP, therefore precluding its recruitment to PrfA-regulated promoters then (2) in the second step, GSH binds, forming a transcriptionally active complex and allowing for increased gene expression. In order to assess the likelihood of the active-reduced model compared to the two-step GSH binding model, we analyzed the ability of GSH to bind PrfA, the results of which are discussed in Chapter 6.

5.3 Materials and Methods

5.3.1 PrfA Protein Purification

The 237 amino acid-encoding *Listerial* *prfA*, *prfAG145S* and *quad prfA* genes were obtained as a gift from the Portnoy Laboratory at U.C. Berkeley in the expression vector pET100. The plasmids had intrinsic ampicillin resistance, and gene expression was controlled by an isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible promoter. The proteins were tagged with an N-terminal hexa-histidine (His6) tag, separated from PrfA by an enterokinase cleavable peptide sequence, which allowed for nickel NTA affinity chromatography purification.

PrfA was expressed in BL21-DE3 or BL21-DE3 pLysS (ThermoFisher) *E. coli* expression cell lines. Cells were grown to an optical density at 600 nM (OD600) of 0.6

73
absorbance units (AU) at 37 °C and induced with 1 mM IPTG for 1.5 hours before harvesting. The cell pellets were resuspended in a phosphate lysis buffer (10% glycerol, 50 mM NaPO₄ pH 7.9, 300 mM NaCl, 40 mM imidazole, 0.1% TritonX-100, and 1 mM BME) and lysed by using a microfluidizer. The lysate was centrifuged to remove cellular debris. The supernatant was then passed over a Ni²⁺-nitrilotriacetic affinity resin and washed with phosphate buffer saline (11.8 mM Na⁺/K⁺ phosphate, 2.7 mM KCl, 137mM NaCl) and PBS + 20 mM imidazole. PrfA was eluted in PBS + 1 M imidazole, and BME was added to the final concentration of 2 mM. PrfA was buffer exchanged over a Sephadex S75 size exclusion column chromatography (SEC) in PBS + 2 mM BME. SEC fractions were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis to assess protein purity before DNA binding assays and crystallization screens were conducted.

All PrfA mutants, including Quad and G145S, were expressed and purified using the above protocol. Oxidized PrfA was purified without BME in the lysis or exchange buffer, and concentration of the sulfhydryl groups was assessed with Ellman’s reagent.

5.3.2 Fluorescence Polarization DNA Binding Assays

A fluorescence polarization-based DNA-binding assay was used to determine the affinities of PrfA, PrfAG145S and Quad PrfA for Phly and PactA promoters. The sequences of the top strands of Phly and PactA used in this study are shown in Table 4 (where 6FAM represents the fluorescein fluorophore). The oligodeoxynucleotides were
purchased from IDT (Coralville, Iowa) with the fluorescein label covalently attached to the 5’ end. DNA binding was measured in PBS buffer at 25 °C using 2.5 nM (Phly) or 5 nM (PactA) fluoresceinated target dsDNA, and 1 μg poly(dI-dC) as a nonspecific DNA competitor. In some experiments, 1 mM BME, DTT, or TCEP was included to maintain a reducing environment. The reductant identity did not affect the $K_d$ for either binding site.

PrfA was titrated into buffer containing DNA until saturation as denoted by no further change in the millipolarization ($mP = \text{units of polarization} \times 10^{-3}$). The fluoresceinated DNA was excited at 490 nm and its parallel and perpendicular emission intensities measured at 530 nm and converted to units of $mP$ using a Beacon 2000 Variable Temperature Fluorescence Polarization System. Data were plotted and analyzed with the following equation

$$mP = \frac{((mP_{\text{bound}} - mP_{\text{free}})[\text{protein}]) / (K_d + [\text{protein}]) + mP_{\text{free}}}{\text{protein}}$$

where $mP$ is the millipolarization measured at a given protein concentration, $mP_{\text{free}}$ is the initial millipolarization of free fluorescein-labelled DNA, $mP_{\text{bound}}$ is the maximum millipolarization of specifically bound DNA, and [protein] is the protein concentration. The generated hyperbolic curves are fit by nonlinear least squares regression analysis, assuming a bimolecular model such that the $K_d$ values represent the protein concentration at half-maximal ligand binding and plotted by using the graphing program, Kaleidograph. The $K_d$ values are expressed in terms of PrfA dimer binding.
Table 3: DNA Binding Affinities of PrfA (Kd ± s.e.m)

<table>
<thead>
<tr>
<th></th>
<th>Phly (nM)</th>
<th>PactA (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G145S</td>
<td>40.8 ± 3.3</td>
<td>45.4 ± 3.2</td>
</tr>
<tr>
<td>PrfA (oxidized)</td>
<td>888.5 ± 140.3</td>
<td>ND</td>
</tr>
<tr>
<td>PrfA(reduced)</td>
<td>34.2 ± 4.8</td>
<td>96.4 ± 26.3</td>
</tr>
<tr>
<td>Quad PrfA</td>
<td>32.8 ± 5.5</td>
<td>124.9 ± 26.3</td>
</tr>
</tbody>
</table>
Figure 16: Representative FP Curves for DNA Binding

Representative FP curves for *Phly* binding assays. **A** Representative binding isotherms of reduced PrfA (red and blue), G145S (green), and Quad PrfA (black). **B** The binding of oxidized PrfA is shown in pink. Millipolarization (mP) has been normalized.
### Table 4: Fluorescinated DNA Oligomers for FP

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phly</strong></td>
<td>5’ 6FAM- TGAGGCATTAACATTTGTTAACGACGAT</td>
</tr>
<tr>
<td><strong>PactA</strong></td>
<td>5’ 6FAM- AACTGATTAACAAATGTTAGAGAAAACCT</td>
</tr>
</tbody>
</table>
Chapter 6. GSH Can Allosterically Bind PrfA

To analyze the ability of GSH to act as a ligand, PrfA affinity for GSH was measured by the Portnoy Laboratory with bio-layer interferometry (Reniere et al., 2015), and theoretical binding for GSH to PrfA was analyzed using AutoDock Vina, a molecular docking program that predicts molecular noncovalent interactions for small molecules to proteins in order to estimate binding conformations and affinities (Trott and Olson, 2010). The ability of GSH to interact with PrfA on DNA was then analyzed with thermal denaturation.

6.1 GSH Binds to PrfA with Millimolar Affinity

Bio-layer interferometry is an optical detection method that measures protein-ligand, protein-DNA, and protein-protein interactions (Ciesielski et al., 2016). Much like SPR, it detects changes in wavelengths of light that occur during the interaction of an analyte with protein immobilized on a biochemical probe (Rich and Myszka, 2007). For this experiment, GSH was used as the analyte to measure binding affinity to PrfA and was conducted by the Portnoy Laboratory using an Octet RED 384 instrument (Pall FortéBio) (Reniere et al., 2015). The buffer was maintained with reductant to isolate the ability of GSH to bind PrfA.

The results of the bio-layer interferometry experiments are given in Table 5. Reduced glutathione bound both PrfA and Quad PrfA with similar affinities (4.37 ± 1.2 mM and 4.74 ± 1.5 mM respectively). The binding was specific for reduced GSH as no
binding could be detected (NBD) for oxidized glutathione (GSSG). The affinity, in the millimolar concentration range, might appear to be low compared to many other protein-ligand interactions (Crp affinity for cAMP is low μM), however this is physiologically relevant for *Listeria monocytogenes* as the intracellular GSH concentration for prokaryotes and eukaryotes ranges from 0.1 – 10 mM (Masip *et al.*, 2006; Meister and Anderson, 1983). There is also evidence that *Listeria monocytogenes* can import GSH from the host cytosol, potentially increasing the intracellular GSH concentration further during infection (Reniere *et al.*, 2015). While GSH may be still be required for maintaining reduced pools of PrfA, it appears to be able to bind PrfA as a ligand as well. These results support the two-step model hypothesis proposed in Chapter 5, where PrfA binding to DNA and GSH binding to PrfA are separate steps, both of which are required for activation of transcription.

The ability of GSH to bind Quad with comparable affinity to PrfA is perplexing. *Listeria* expressing Quad PrfA have a noticeably defective phenotype in mice infection models, yet there has been no indication of why as both DNA and GSH binding appear to be equivalent to the wild-type protein.

**6.1.1 Special Acknowledgement for the Bio-layer Interferometry Data**

We would like to acknowledge the Portnoy Laboratory at UC Berkeley and Dr. Michelle Reniere, who gathered the data for Table 5.
6.2 Docking Supports the Binding Potential of GSH as a Ligand

The results of the bio-layer interferometry studies indicate GSH is capable of binding PrfA, however the measured affinity is low. To further address the ability of GSH to act as a ligand, docking studies were conducted using AutoDock Vina.

AutoDock Vina (ADV) is a molecular docking program that uses molecular dynamic simulations to predict conformations and affinities of small molecules for a protein of interest. The mathematical theory and methods have been summarized elsewhere (Trott and Olson, 2010). Docking of GSH (a molecular model of GSH taken from glutathione s-transferase family protein, PDB: 4IEL) to PrfA (PDB: 2BEO) was performed using ADV Version 1.1.2 (Trott and Olson, 2010). The parameters for docking were set such that the program search encompassed one complete subunit, including the C-terminal DNA binding domain, to prevent biasing placement in the expected effector region.

The ADV search algorithms returned nine potential orientations for GSH with predicted binding free energy (ΔG) values ranging from -6.1 to -6.6 kcalmol⁻¹. Each of the nine docking solutions placed GSH in the N-terminal β-barrel region of the protein with a variety of orientations, two of which are shown in Figure 17 A and B.

It is important to note that ADV presents a theoretical placement of the ligand based on molecular simulations, so it is critical not to overdraw conclusions based on the structural model. However, the predicted GSH placement is proximal to residues...
Tyr154, Tyr63, Lys64, Lys122, Met58, and Lys130 (shown in red in Figure 17 C) that yield inactive protein when mutated (mutants discussed in Chapter 3.2.1). The proximity of critical PrfA residues for activity support the ADV predictions and further suggests that GSH can plausibly bind PrfA in the predicted N-terminal domain, supporting the bio-layer interferometry experiments. Unfortunately, it does not lend insight to the contribution of the cysteines to explain the mechanistic inadequacy of the Quad mutant. None of the cysteine residues are adjacent to any of the predicted GSH orientations, except for Cys144, although mutation of this residue does not exhibit any negative phenotypes (Reniere, personal communication). While ADV does support that GSH is a good ligand, it does not illuminate the role of the cysteine residues in ligand binding or genetic activation.

6.3 PrfA Stability Measured Through Thermal Denaturation

As a final approach to analyze the interaction between GSH and PrfA, a thermal denaturation assay was conducted. Thermal denaturation (TD) can give insight into changes in protein stability induced by ligand or nucleic acid interactions (Guédin et al., 2010; Senisterra et al., 2012; Waldron and Murphy, 2003). Denaturation is measured as temperature increases and comparisons can be made between melting temperatures (T_m, measured where the protein is half denatured) of protein and ligand-associated protein. An increase in T_m represents a stabilizing interaction.
Thermal denaturation has been used successfully to measure ligand interactions, such as in the case of the *E. coli* multi-drug tolerance regulator HipA. HipA, a serine protein kinase, binds ATP in order to autophosphorylate and modify its ligand glutamyl tRNA synthetase, and the binding of ATP to HipA and the kinase-dead mutant D309Q has been monitored with TD to follow the protein stabilization induced by ATP binding (Schumacher *et al.*, 2012). The HipA-ATP interaction serves as an excellent proof of concept for monitoring PrfA-GSH interactions through similar methods.

For PrfA specifically, TD could clarify the interaction between PrfA and Quad. FP, bio-layer interferometry, and ADV have all failed to provide plausible explanations for why Quad PrfA exhibits an *in vivo* decrease in *Listerial* infectivity yet binds DNA and GSH with similar affinity to WT PrfA. In an attempt to illuminate a difference between the WT and Quad mutant, TD experiments on PrfA or Quad on DNA in the presence and absence of GSH were conducted to analyze complex formation.

### 6.3.1 PrfA and Quad Exhibit Different Stabilities in the Presence of DNA and GSH

PrfA and Quad denatured at similar Tₘs (≈ 60 °C) in the absence of GSH and DNA. The addition of *Phly* or *PactA* increased the stability of both proteins with their Tₘs increasing 2 – 3 °C; *Phly* stabilized the proteins slightly more than *PactA* (Figure 18 A and C). This would be expected as DNA-protein interactions are usually stabilizing, and the difference in the shifts between the tighter binding *Phly* and weaker binding *PactA* parallels the previously determined binding affinities measured in Chapter 5.
There are several noteworthy results from the stability profiles of PrfA- and Quad-Phly complexes upon the addition of GSH. The first is that GSH has a destabilizing effect on the wild type PrfA-Phly complex when added in physiologically relevant concentrations (Figure 18 B). Biologically, protein-ligand interactions most commonly result in stabilized thermal denaturation profiles, such as in the case of ATP binding to HipA (Schumacher et al., 2012), however there are examples where protein destabilization is necessary for binding interactions (Walters et al., 2015; Zhang et al., 2013; Zhao et al., 2015). This destabilization is further supported by previous proteolytic digestions of PrfA and PrfA* as G145S is more susceptible to limited proteolysis than the inactive apo conformation. Other PrfA* mutants exhibit this pattern as well, suggesting that the “active” conformation of PrfA is less structured (Miner et al., 2008; Xayarath et al., 2011a, 2011b).

The second important result from this experiment is the difference in T\textsubscript{ms} of the PrfA-Phly and Quad-Phly in the presence of increasing concentrations of GSH. The presence of GSH decreased the T\textsubscript{m} of wild-type PrfA by ~5 °C in the presence of Phly while corresponding concentrations of GSH added to Quad had the opposite effect, increasing the T\textsubscript{m} by ~4 °C (Figure 18 B and D).

The difference between the WT and Quad protein support conformational differences between the (Quad-GSH)-DNA and the (PrfA-GSH)-DNA complexes. The thermal denaturation data in combination with the previously-reported proteolytic
digest data supports the hypothesis that Quad adopts a conformation that is less active (more stable) in the presence of GSH while PrfA needs to be destabilized to be transcriptionally active, perhaps to facilitate a productive interaction with RNAP.

The TD assays add further support to the bio-layer interferometry and AutoDock Vina experiments that GSH likely acts as an allosteric regulator of PrfA. It additionally indicates that GSH can interact with PrfA already bound to DNA, continuing to support the two-step activation hypothesis.

Furthermore, these results finally suggest there is a conformational difference between PrfA and Quad that could serve as the basis for the less virulent phenotype of Quad in Listeria. This incidentally further supports a two-step mechanism for PrfA gene activation where Quad and PrfA are both capable of binding DNA (the first step) but Quad is less able to transition to a transcriptional activator (the second step).

6.3.2 Thermal Denaturation Materials and Methods

Thermal denaturation experiments were performed with a Perkin Elmer Lambda 25 UV/VIS spectrophotometer equipped with PTP 6 Peltier system. PrfA was diluted to 2 µM in PBS with 2 mM BME and incubated with increasing concentrations of GSH for five minutes. The samples were mixed in quartz cuvettes and heated at a rate of 1˚C per minute. The UV absorbance was measured at 310 nm in the temperature range between 25˚C and 75˚C with data points obtained every 1˚C. Only one-direction denaturation could be monitored as PrfA was irreversibly denatured at high temperatures. In
experiments that included DNA, a 28 dsDNA encompassing the *Phly* or *PactA* site was used in 1:1 ratio to protein concentration. The promoter sequences were identical to the DNA sequences used for FP assays (given in Table 4). Non-specific dsDNA did not shift the PrfA $T_m$, and there was no contribution to the absorbance from the buffer, DNA, or GSH at 310 nM (data not shown). The 28 dsDNA had a predicted $T_m$ of 67 °C and 63 °C for *Phly* and *PactA* respectively. Absorbance at 310 nm is reported as normalized absorbance units (AU).

The thermal denaturation isotherms displayed in Figure 18 are representative isotherms. Quad-*Phly* and PrfA-*Phly* denaturation was repeated three and four times respectively. Quad-*Phly*-GSH and PrfA-*Phly*-GSH was performed once, however the $T_m$ of each protein was also measured in the presence of oxidized glutathione (GSSG) and was unaffected.

### 6.4 Proposed Mechanism of PrfA Activation Based on Biochemical Data

The DNA binding assay, bio-layer interferometry, and thermal denaturation data collectively contribute to a mechanism of PrfA activation based on both the oxidation state of its cysteines and formation of a complex with glutathione. The ability of GSH to bind to PrfA allows us to propose a two-step mechanism for PrfA-mediated gene activation (Figure 19).

Initially, when *Listeria* enters the host cell through endocytosis, PrfA is exposed to an oxidizing environment, leading to the formation of inactive apo-PrfA. Once the
bacterium escapes and enters the host cytosol, the reducing environment allows PrfA to switch to a DNA-binding yet transcriptionally less active conformation. As infection progresses, glutathione concentration increases, allowing GSH to bind to PrfA and transition the regulator to a transcriptionally active complex which facilitates RNAP binding to a number of virulence promoters. This begins the positive feedback cycle that proceeds to increase PrfA concentration and allow the infection to progress.

Quad is likely adequate in the DNA binding step of the model (allowing it to inefficiently escape the host vacuole) but deficient in the transition to the transcriptionally active conformation, which prevents it from increasing the PrfA concentration to sufficiently activate the autoregulatory loop and late stage virulence proteins such as ActA, explaining the 50% decrease in PrfA concentration and 30-fold decrease in virulence (Reniere et al., 2015). As Quad is capable of binding GSH with similar affinity to WT PrfA, this deficiency suggests the cysteines are important for coordinating interactions with RNAP.

The two-step hypothesis predicts that PrfA will shift from the apo (inactive) to the DNA-binding conformation then finally to an active complex that mimics the Crp-DNA-αCTD structure. In order to accurately assess this hypothesis, the next step is to determine the structures of PrfA in complex with Phly and the GSH-bound PrfA-Phly complex.
Table 5: Glutathione Binding Affinity (K_d ± s.e.m.)

<table>
<thead>
<tr>
<th></th>
<th>GSH (mM)</th>
<th>GSSG (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PrfA</td>
<td>4.37 ± 1.2</td>
<td>NBD</td>
</tr>
<tr>
<td>Quad PrfA</td>
<td>4.74 ± 1.5</td>
<td>NBD</td>
</tr>
</tbody>
</table>
Figure 17: “Best” GSH Binding Modes to PrfA as Calculated by Autodock Vina

A and B show two orientations out of the top nine found by AutoDock Vina for GSH binding to WT PrfA (2BEO), highlighted by red dashed circles. C Close up view of the residues proximal to the GSH shown in panel B. Residues near the putative GSH binding sites are shown that when mutated produce inactive PrfA (red) or activated PrfA (green). Many of the inactivating mutants are spatially close enough to potentially interact with bound GSH.
Thermal denaturation curves for A wild-type PrfA + DNA, B PrfA-Phly + glutathione, C Quad + DNA, and D Quad-Phly + GSH. For both PrfA and Quad, cognate promoters exhibited a stabilizing thermal shift. When GSH is added, PrfA-Phly is destabilized while Quad-Phly is further stabilized. (T_m shifts are indicated by black arrows for stabilizing interactions or the red arrow for destabilizing interactions).
Figure 19: Two-Step PrfA Mechanism Hypothesis

The hypothesized two-step mechanism of PrfA activation. In its own cellular environment, PrfA is reduced, capable of binding DNA, and slightly transcriptionally active. When *Listeria* is endocytosed by a host cell, the environment changes, and PrfA is likely oxidized in the vacuole thereby no longer able to bind DNA. Escape from the vacuole to the host cytosol allows the cysteine residues of PrfA to be reduced and PrfA to bind DNA. As glutathione concentration increases, PrfA also becomes transcriptionally active, allowing for an increased expression of the virulence genes required for propagation of the *Listerial* infection (Reniere *et al.*, 2015).
Chapter 7. Structure of the PrfA-Phly Complex

Obtaining the structure of reduced PrfA in complex with DNA is vital to evaluate the two-step gene activation hypothesis proposed in Chapter 6. If reduced and bound to GSH, PrfA should be in the transcriptionally active form of the protein, and the DNA will likely be bent in the PrfA-DNA structure similarly to the DNAs in the structures of Crp-cAMP-DNA (Schultz et al., 1991) and CprK-DNA (Levy et al., 2008). Therefore, the structure of PrfA was determined in complex with a 30-mer dsDNA encompassing the Phly promoter. This structure supplements interesting molecular details for the proposed PrfA-mediated gene activation hypothesis.

7.1 An Improvement in the Resolution of the Apo-PrfA Structure

The initially reported apo-PrfA structure was determined to 2.7 Å resolution (Eiting et al., 2005), however in hanging drop-vapor diffusion screening for PrfA-DNA crystals, we grew apo-PrfA that diffracted to 2.3 Å, which I have refined to $R_{\text{work}}$ and $R_{\text{free}}$ values of 19.6% and 25.7% respectively. Apo-PrfA crystalized in the space group P4$_1$2$_1$2$_1$ with unit cell parameters a = b = 89.24 Å, c = 448.84 Å ($\alpha$, $\beta$, $\gamma$ = 90°). Select crystallographic data is listed in Table 8. I shall use this higher resolution structure of apo-PrfA for all future structural comparisons.

There was one dimer and four subunits per asymmetric unit (Figure 20A) which allowed the generation of four additional dimers with symmetry mates. The generated dimers overall adopted similar conformations, especially in the N-terminal domain,
the greatest flexibility existing in the loops of the C-terminal domain between residues 156 and 161 (corresponding to the AR1 region of Crp), 173 and 182, and 201 and 206 (Figure 20B).

If one representative PrfA dimer (Figure 21A) is aligned with the apo-PrfA structure from 2005 (alignment shown in Figure 21B), the structures are similar with Cα alignment yielding an RMS deviation of 0.53 Å, again with the largest differences between our apo-PrfA and the 2005 structure localized to the C-terminal domain, which is likely a result of intrinsic protein flexibility in that region. This flexibility is not unusual for Crp family regulators as the DNA binding domain of CooA is highly flexible and does not even display electron density in one monomer of the active form (Borjigin et al., 2007; Lanzilotta et al., 2000). Additionally, the recognition helix in our apo-PrfA is slightly rotated and widened (Figure 21B, enlarged), again implying flexibility in the DNA-binding domains.

Similar to the original PrfA structure, a BME molecule is disulfide bonded to Cys205 in one of the 6 protomers, and a water molecule is bonded to C205 in another protomer. This may suggest, as the original structure did, that the cysteine residues are available to react when the protein is in an oxidizing environment, which may indicate that mixed disulfides stabilize the inactive form of the protein. It is particularly interesting that both DTT and BME bond to the cysteines as perhaps GSH can form a mixed disulfide with these residues in the Listerial cell. If this is the case, GSH may both
activate and inactivate PrfA; inactivating in oxidizing conditions when the concentration of GSH is lower in the cell and activating the protein later in infection when the pool of GSH is high.

### 7.2 Structure of PrfA-Phly$_{30}$

The structure of PrfA in complex with a 30bp DNA encompassing the Phly palindrome (PrfA-Phly$_{30}$) was determined to 2.90 Å ($R_{work}/R_{free} = 31.6/32.7\%$) with one dimer-DNA complex in the asymmetric unit. The overall structure is shown in Figure 22A and is more similar to the constitutively active G145S (PDB: 2BGC) structure than the apo-PrfA structure (PDB: 2BEO). Structural alignments of PrfA-Phly$_{30}$ (light green) to apo-PrfA (blue) and G145S (dark green) are shown in Figure 22B and C, and the root mean squared deviation (RMSD) for apo-PrfA is 1.61 Å compared to 0.69 Å for G145S.

The most noticeable variances between DNA-bound PrfA and apo-PrfA exist in the C-terminal domain of the protein (Figure 22B and C, indicated by black arrows) in both the loop regions and the HTH motif. The loops in PrfA-Phly$_{30}$ are more flexible than in apo-PrfA, and very little electron density is visible between residues Gly155 and Pro159, which prevents accurate model building in this region. The recognition helices rotate as well, straightening slightly and allowing them to fit in the major groove of the DNA with the center of the PrfA-box located under the dimerization helix. There is also a significant rearrangement of αE as previously seen in the G145S structure.
While the resolution prevents detailed analysis of the predicted GSH binding pocket, there are several key residues that are proximal to the DNA that may be responsible for base specificity and interactions with the phosphate backbone that can be discussed. These residues are shown both in cartoon form and as a schematic of interactions with the *Phly* palindrome. The forward DNA sequence is numbered T1 – A14, and the reverse sequence is T1’ – A14’ (Figure 23A and B).

The residues responsible for base interactions are Arg188 and Ser184 (Figure 24). Arg188 imparts base specificity to G10 and G10’, although interactions are asymmetric with the palindromic *Phly* sequence. To G10, atom NH₂ of Arg188 is proximal for interactions with the O₆ atom. NH₁ is slightly out of distance for optimal interaction with N₇ of G10, however this is potentially due to ambiguity in distances at this resolution. Additionally, the Nₑ atom of Arg188 is able to interact with the O₆ atom of T11. These interactions are shown in Figure 24B. To G10’, NH₂ and NH₁ of Arg188 coordinate as expected to O₆ and N₇ respectively (Figure 24A).

Ser184 interacts with bases T11, T12, T11’, and T12’. The Ser184 hydroxyl is proximal to both O₄ atoms of T11 and T12, however the Cβ is slightly too distant for an interaction to the T12 methyl group to be described with confidence (Figure 24B). Interaction between Ser184 and T12’ is specific however with interactions of the hydroxyl to O₄ of T11’ and T12’ and van der Waals interactions between Cβ and the Cγ methyl group of T12’ (Figure 24A).
Additionally, PrfA has a combination of residues that make symmetrical and asymmetrical contacts with the phosphate backbone. His182, Lys139, and Ser187 interact symmetrically by making electrostatic interactions with the phosphate backbone. Lys139 also makes van der Waals interactions by laying along the deoxyribose moiety of T9 (and T9’). The backbone amide nitrogen of Lys139 and Ile180 additionally make electrostatic phosphate contacts, although the interaction from Ile180 is asymmetric.

Outside the palindromic sequence of this PrfA box containing promoter (not shown), phosphate backbone interactions are made by the polypeptide chain backbone amide nitrogen of residue Met171 and hydroxyl of Tyr201 between the G0 and T(-1) and the A0’ and C(-1’) nucleotides. Additionally, Ser183 makes an asymmetric H-bonding interaction with the phosphate backbone between G0 and T(-1).

Previous research has indicated the S183A mutation results in a PrfA* mutant (Sheehan et al., 1996), although it is unclear from the structural model why S183A would prompt hyper-virulence. The backbone contact does not seem overly important to the protein-DNA connection in light of the numerous alternative backbone interactions from Lys139, His182, and Ser187, however the residue may be necessary for coordinating the overall bend of the Phly DNA.

The 30-mer dsDNA itself exhibits some flexibility at the ends of the oligomer. While there is positive density in the 2Fo-Fc electron density map at the 5’ ends of the
DNA, there is not significant enough occupancy to accurately model in the accompanying base pairs despite multiple efforts of building and refinement. Hence, six nucleotides (three each from the 5' DNA ends) are not modelled and their absence provides a solid explanation of the higher than normal values of the $R_{\text{work}}$ and $R_{\text{free}}$.

The dsDNA itself is in a slightly bent conformation compared to linear B-DNA, measuring an overall bend of about 35 - 40°. This bend is less than that of the crystallized Crp DNA by about 50° and may indicate that this PrfA-Phly$_{30}$ complex structure has captured an intermediate DNA conformation as compared to the DNA in a fully transcriptionally active complex. However, this intermediate bend should be scrutinized as a potential artifact of crystal packing. This possibility will be discussed further in light of the structure of the PrfA-Phly$_{24}$ complex in Chapter 8.

### 7.2.1 Validation of the PrfA-Phly DNA Interactions

In order to validate the structural model, key residues that make base and backbone contacts between PrfA and the Phly dsDNA were substituted singly with alanine. The mutations targeted base-interacting residues Arg188 and Ser184 and backbone contacting residues Lys139, His182, and Ser187. The binding affinities to Phly and PactA were measured using the FP-based DNA binding assay described in Chapter 5 and are summarized in Table 6 (where N.S. stands for non-specific binding). Representative binding isotherms are given in Figure 25.
Mutation of the residues responsible for base interactions negatively impacted the protein binding affinity. The R188A mutation resulted in non-specific binding to both Phly and PactA (Figure 25A) while S184A decreased the affinity of the reduced protein four-fold for Phly (Figure 25B) and abolished specific binding for PactA. The experimental results for Ser184 concur with previously conducted EMSA analysis that indicated this alanine mutant alone was enough to induce Listerial avirulence (Sheehan et al., 1996). The difference in binding affinities of S184A for Phly and PactA is interesting as the structure of PrfA-Phly30 does not suggest Ser184 would be affected by the nucleotide change in PactA and supports previous experiments that suggests PrfA interacts differently with PrfA box sequences that deviate from the palindrome.

It is also worth noting that previous observations indicate that the PrfA box only requires invariant base-pairs at the 2 – 5 positions (Scortti et al., 2007), which is supported by the PrfA-Phly30 structure where Arg188 and Ser184 make critical connections to this stretch of DNA.

The K139A and H182A mutants had similar results to R188A as alanine substitution destroyed specific binding for both promoters (Figure 25A), implying these backbone contacts are critical for DNA-association. In light of the number of interactions between Lys139 and His182 to the DNA backbone including side chain and amide nitrogen electrostatic interactions, it is understandable that both of these mutations would be particularly devastating.
The S187A PrfA mutant had little effect on DNA binding as the affinities for Phly and PactA were comparable to reduced PrfA and G145S (Figure 25B), suggesting that either the backbone contact was accounted for by other residues or the proximity of the Ser187 side chain to the backbone in this structure is not critical to PrfA binding to DNA.

Thus the analysis of these point mutants suggest that contributions from residues Arg188 and Ser184 are required for base interactions, Lys139 and His182 are sufficient for DNA binding, while Ser187 may be necessary for DNA bending.

### 7.2.2 Structural Comparison of the PrfA-Phly<sub>30</sub> and Crp-cAMP-DNA Complexes

When the structure of the PrfA-Phly<sub>30</sub> complex is aligned with the Crp-cAMP-DNA complex structure, several critical differences between these two family members are observed that are localized around the protein and the DNA.

First, the amino acid residues of PrfA required for DNA recognition are different from the corresponding residues required for specificity in Crp (see protein sequence alignment in Chapter 3, Figure 3). The Crp residues that impart specificity for the Crp box are Arg180, Glu181, and Arg185. While Arg188 of PrfA is comparable to Crp Arg185 and Ser184 is one residue away from Glu181, PrfA His182 (equivalent to the base-contacting residue to Crp Arg180) only interacts with the DNA backbone. The differences in base contact coordination between Crp and PrfA may originate from the difference in sequences between the Crp box and PrfA box (shown in Figure 26).
While protein sequence alignment does not indicate comparable residues in Crp and PrfA make the same interactions, the spatial coordination of the palindrome is similar in the second half of the PrfA and Crp boxes. However, PrfA makes asymmetric base contacts (Arg188 contacts T11 while Ser184 contacts T11’) and only makes base specific interactions with half of the palindrome while Crp base interactions are symmetrical and observed to occur on both halves of the Crp box.

Another key difference between PrfA-Phly30 and the Crp-cAMP-DNA structure is the overall bending of the dsDNA oligomer (Figure 27A). While the DNA in the center of the palindrome overlays well, the Crp DNA kinks sharply, yielding an 80 - 90° bend. The Phly oligomer on the other hand bends 35 - 40°, which results in the 5’ ends being displaced about 8 Å from the corresponding 5’ ends of the Crp-cAMP-DNA structure. Additionally, the wing loop in the wHTH motif of PrfA when bound to Phly30 does not insert into the minor groove of the DNA as observed in the Crp-cAMP-DNA structure (Figure 27A, enlarged). This is likely due to the difference in bending; if Phly were to adopt a more significantly bent conformation (such as the Crp structure), the wing would be more proximal to the minor groove and able to form backbone interactions or even insert into the groove to contact the bases.

Variations in DNA bending are even more prevalent when PrfA-Phly30 is aligned with the Crp-cAMP-DNA-αCTD structure (Figure 27B). The 5’ end of the 30-mer Phly is now about 10 Å displaced from its spatial counter-part in the 44 dsDNA oligomer of the
Crp-cAMP-DNA-αCTD complex. Furthermore, the putative AR1 region in PrfA is disordered compared to the Crp AR1, which is an ordered loop (yellow in Figure 27B, enlarged), although it is unclear if this flexibility is necessary for interaction with the αCTD.

It is critical to note that the lack of DNA bending in the Phly structure suggests the αCTD of RNAP would be unable to interact with the proposed AR1 region of PrfA in the crystallized structure. If the Crp DNA were to unbend to the same angle as Phly, the αCTD would shift farther away from the protein and it would be unlikely to make the necessary protein-protein interactions, which would decouple the associations needed for efficient transcription and gene activation.

7.2.3 Reduced PrfA is Likely not the Transcriptionally Active Form

Superposition of the structures of the PrfA-Phly30 and Crp-cAMP-DNA complexes leads us to conclude that reduced PrfA is not transcriptionally active. The moderate DNA bend in comparison to Crp would prevent recruitment of the αCTD, which in turn would not allow for the efficient transcription that is observed when PrfA activates downstream virulence genes. Instead, PrfA-Phly30 is likely an intermediate conformation in the transition from apo (inactive) PrfA to the active GSH-bound conformation, perhaps indicating that GSH binding is required for the full conversion to the bent-DNA conformation predicted by the Crp model. The PrfA-DNA structure additionally supports the two-step mechanism proposed in Chapter 6.4 with the PrfA-
Phly\textsubscript{30} conformation being the first step that occurs when \textit{Listeria} enter the reducing environment of the host cytosol.

**7.3 PrfA Crystallization Methods and Conditions**

**7.3.1 Apo-PrfA Crystal Conditions and Methods**

PrfA was crystallized in drops containing 1 μL precipitating solution and 1 μL protein-DNA mixture at a protein concentration of 5 mg/mL in 48-well hanging drop-vapor diffusion plates at room temperature. The precipitating solution was a mixture of 1.0 M ammonium phosphate and 0.1 M imidazole pH 8.0. Cryo-protection was accomplished by transferring protein crystals to mother liquor containing 30% ethylene glycol for 1 – 3 minutes before freezing. Images of the crystals and diffraction pattern can be seen in Figure 28A and Figure 28C. The protein-DNA mixture was a 1:1.1 ratio of PrfA to a double stranded, perfectly palindromic \textit{Phly}-encompassing 30 nucleotide oligomer (synthesized by IDT, Coralville, Iowa, top strand sequence 5’ TTGAGGCATTAACATATGTTAATGCCTCAA). It is worth noting that while PrfA was screened as a protein-DNA mixture, the determined structure yielded no DNA density, and protein crystals did not form in the absence of DNA. The DNA oligomer likely acted as an additional precipitant or assisted in crystal packing. This crystal was designated apo-PrfA.

Apo-PrfA crystals adopted the space group P4\textsubscript{i2}2\textsubscript{i} with unit cell dimensions a = b = 89.24 Å, c = 448.84 Å and α = β = γ = 90°. X-ray intensity data were collected at the
ID22 line at APS (Chicago, IL) on MarCCD detectors and processed with HKL2000 to a resolution of 2.31 Å. The structure was determined via molecular replacement (MR) with wild-type PrfA (PDB: 2BEO) as the search model using Phenix (Adams et al., 2010). There was one PrfA dimer and four additional protomers per asymmetric unit, which results in a solvent content ($V_s$) = 48.8%. The structural model was refined to a final $R_{work}/R_{free}$ of 19.6/25.7 %, respectively. Selected crystal statistics are given in Table 8. All structural models were generated using Pymol (DeLano, 2002a, 2002b).

### 7.3.2 PrfA-Phly$_{30}$ Crystallization and Other Methods

The PrfA-Phly$_{30}$ complex was crystallized in drops containing 4 μL precipitating solution and 5 μL protein-DNA mixture at a protein concentration of 6.5 mg/mL in 48-well hanging drop-vapor diffusion plates at room temperature. The precipitating solution was a mixture of 25% PEG 400, 0.1 M sodium acetate pH 4.5, and 0.3 M calcium acetate, which was sufficient for cryo-protection. Images of the crystals and diffraction pattern can be seen in Figure 28B and 28D. The protein-DNA mixture was a 1:1.1 ratio of PrfA to a double stranded 30 oligomer (synthesized by IDT, Coralville, Iowa) encompassing the Phly palindrome where the top strand was 5’ TTGAGGCATTAACATTTGTTAACGACGATA. PrfA-Phly$_{30}$ crystals assumed the space group P4$_3$2$_1$2 with unit cell parameters $a = b = 71.6$ Å, $c = 260.4$ Å, and $\alpha = \beta = \gamma = 90^\circ$. There was one PrfA dimer and dsDNA oligomer per ASU, which results in a solvent content of $V_s = 43.1\%$. X-ray intensity data were collected at the BM22 line at
APS on MarCCD detectors and processed with HKL2000. The structure was determined via MR with G145S (PDB: 2BGC) as the search model using Phenix (Adams et al., 2010) and refined to a final resolution of 2.9 Å. Selected crystal statistics are given in Table 8. All structural models were generated using Pymol (DeLano, 2002a, 2002b).

### 7.3.3 Mutagenesis of PrfA

PrfA alanine mutants were generated using site-directed mutagenesis. The wild-type prfA gene in pET100 was mutated in the presence of PFU Turbo DNA polymerase (Agilent Technologies) and mutagenesis primers, the sequences of which are given in Table 7, by polymerase chain reaction in an Eppendorf thermocycler. Mutant sequences were verified by sequencing (Eton Bioscience) before being expressed and purified by metal affinity chromatography as described in Chapter 5.3.1.
Figure 20: Asymmetric Unit of Apo-PrfA

A View of the ASU for apo-PrfA with each individual protomer shown as cartoons and in a different color. B When the dimers of each PrfA of the ASU (generated by symmetry mates) are aligned, the highest flexibility in the structure is observed in the loop regions between residues 156 – 161, 173 – 182, and 201 – 206, indicated by black arrows.
Figure 21: Apo-PrfA

A One representative apo-PrfA dimer determined at 2.3 Å resolution. B The higher resolution structure described in this thesis overlays closely with the one determined in 2005 (dark blue). The largest structural differences are in the C-terminal domain (B, enlarged) and likely a result of intrinsic flexibility of the DNA binding domain.
Figure 22: The PrfA-Phly30 Complex

A The overall structure of PrfA-Phly30 (left) and rotated 90° to look down the DNA axis (right). In B and C, PrfA-Phly30 (lime green) is overlaid onto apo-PrfA (B, dark blue) and to G145S (C, dark green), which shows significant flexibility in the loops (noted by black arrows). Structural differences were seen only between apo-PrfA and PrfA-Phly30 and concentrated in the recognition helices.
Figure 23: Protein-DNA Contacts between PrfA and Phly

The residues making DNA base and backbone contacts shown as a ribbon diagram (A) and schematic representation (B). The base interaction residues are Arg188 and Ser184 while the residues interacting with the phosphate backbone are Ser187, Lys139, His182, and the amide nitrogen of Ile180. These are the interactions for the palindromic sequence only (labeled in the forward direction as 1 – 14 and in the reverse as 1’ - 14’).
Figure 24: Interactions between PrfA Residues R188A and S184 and the DNA Bases of Phly

A stick representation of the protein-DNA interactions between Arg188 and Ser184 with G10', T11', and T12 (A) and G10, T11, and T12 (B) of the Phly palindrome. The base interactions are asymmetric and may fluctuate depending on the intrinsic flexibility of the DNA binding domain.
Table 6: DNA Binding Affinities of Selected PrfA Point Mutants (K<sub>d</sub> ± s.e.m.)

<table>
<thead>
<tr>
<th></th>
<th>Phy (nM)</th>
<th>PactA (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K139A</td>
<td>N.S.</td>
<td>N.S.</td>
</tr>
<tr>
<td>H182A</td>
<td>N.S.</td>
<td>N.S.</td>
</tr>
<tr>
<td>S184A</td>
<td>133.8 ± 17.8</td>
<td>N.S.</td>
</tr>
<tr>
<td>S187A</td>
<td>59.6 ± 3.5</td>
<td>74.8 ± 11.5</td>
</tr>
<tr>
<td>R188A</td>
<td>N.S.</td>
<td>N.S.</td>
</tr>
</tbody>
</table>
Figure 25: Binding Isotherms for Selected PrfA Point Mutants

Binding isotherms for the PrfA point mutants K139A, H182A, and R188A, which bind DNA non-specifically (A), and S184A and S187A, which are still able to bind DNA. Isotherms are representative curves for binding to *Phly*.
Figure 26: Comparison of the DNA Contacts from Crp and PrfA

Alignment of the sequences of the Crp and PrfA boxes with base-interacting residues indicated by the arrows. The required base-pairs for DNA binding are boxed. The spatial recognition is similar between Crp and PrfA although the sequences are different which may explain the differences in the residues required for specificity.
A Alignment of PrfA-Phly30 (lime green) with Crp-cAMP-DNA (red) on DNA (Phly orange, Crp DNA grey). The overall bend angles of the DNA are significantly different between the structures and the wing of the wHTH motif of PrfA does not insert in the minor groove (enlarged, right). B Alignment of the complexes in A with Crp-cAMP-DNA-αCTD complex (Crp red, DNA gray, αCTD purple) which shows that while αCTD would be able to interact with PrfA if the DNA were to adopt the same bent conformation, the Phly DNA would not allow the two proteins to interact. Additionally, the enlarged window of B shows that the corresponding AR1 region (yellow) of PrfA is disordered or highly flexible in the PrfA-Phly30 structure resulting in no discernible electron density for much of this region.
Figure 28: Crystal and Diffraction Images of Apo- and DNA-Bound PrfA

Representative crystals of apo-PrfA (A) and PrfA-Phly30 (B). Example diffraction images are shown under their respective crystals with apo-PrfA in C and PrfA-Phly30 in D.
Table 7: Mutagenesis Primers for PrfA Alanine Mutants

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Forward Sequence</th>
<th>Reverse Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>K139A</td>
<td>5’ TCG ATT AAC GGG <strong>GCG</strong> CTT GGC TCT ATT TGC</td>
<td>5’ GCA AAT AGA GCC AAG <strong>CGC</strong> CCC GTT AAT CGA</td>
</tr>
<tr>
<td>H182A</td>
<td>5’ GGC ATC GCA GCT AGC TCA <strong>GCT</strong> GTT AGC AGA</td>
<td>5’ TCT GCT AAC <strong>AGC</strong> TGA GCT AGC TGC GAT GCC</td>
</tr>
<tr>
<td>S184A</td>
<td>5’ GGC ATC GCA CAT AGC <strong>GCA</strong> GCT GTT AGC AGA</td>
<td>5’ TCT GCT AAC <strong>AGC</strong> TGC GCT ATG TGC GAT GCC</td>
</tr>
<tr>
<td>R188A</td>
<td>5’ GCA CAT AGC TCA GCT GTT <strong>GCC</strong> AGA ATT ATT TCC</td>
<td>5’ GGA AAT AAT TCT <strong>GGC</strong> AAC AGC TGA GCT ATG TGC</td>
</tr>
<tr>
<td>S187A</td>
<td>5’ GCA CAT AGC TCA GCT GTT <strong>GCC</strong> AGA ATT ATT TCC</td>
<td>5’ GGA AAT AAT TCT <strong>GTC</strong> AAC AGC TGA GCT ATG TGC</td>
</tr>
</tbody>
</table>
Chapter 8. Structure of PrfA-Phly<sub>24</sub> Complex

The two-step activation hypothesis for PrfA predicts that once reduced PrfA has bound DNA, GSH interacts and generates a transcriptionally active (PrfA-GSH)-DNA complex. Obtaining the complex of GSH-bound PrfA on DNA is the final critical structure to generate a snapshot of the molecular evolution from apo-PrfA to fully active PrfA.

The structure of PrfA in complex with a 24 oligomer of dsDNA encompassing the Phly palindrome (PrfA-Phly<sub>24</sub>) was determined to 3.6 Å resolution (R<sub>work</sub>/R<sub>free</sub> = 23.5/27.4 %). The crystals grew in a solution that contained high concentrations (7.5 mM) of glutathione with the goal of obtaining a GSH-bound PrfA-DNA structure.

The crystallographic unit cell space group was C2 and contained 2.5 PrfA dimers bound to Phly<sub>24</sub> in the ASU (the protomer and half DNA site generated a full DNA-complexed dimer with subsequent symmetry mates). Alignment of all three dimers (the last generated by crystallographic symmetry) indicates they adopt the same overall conformation and same DNA shape. One representative dimer-DNA complex is shown in Figure 29A.

Interestingly, structural alignment of PrfA-Phly<sub>30</sub> and PrfA-Phly<sub>24</sub> indicates no significant structural differences between the proteins (Figure 29B, RMSD of 0.7 Å). While the resolution of PrfA-Phly<sub>24</sub> is modest, precluding the analysis of protein-DNA interactions, the overall bend in the DNA can be assessed, and it aligns with the Phly<sub>30</sub>
dsDNA as opposed to the more bent Crp DNA. In order to prevent biasing the DNA conformation, initial MR searches used only the PrfA protein and built DNA into the electron density, which reduces the likelihood of this DNA bend occurring due to structure bias from the MR placement.

The PrfA-Phly24 complex structure would indicate a deviation from the Crp model mechanism if GSH was bound to the protein, however there does not appear to be positive electron density for a glutathione to be placed within the N-terminal ligand binding domain. It is possible that GSH is indeed bound but not visible at the current resolution or that it has multiple binding modes thereby precluding model building.

By contrast, there is significant positive electron density around residue Cys205 in all independent protomers. This density could be BME or GSH forming a mixed disulfide with the sulfhydryl side chain. BME can be modelled in to the density, which slightly improves the $R_{work}$, however the resolution of the data makes identification of the molecular density around Cys205 difficult. This density is interesting as it implies that either the mixed disulfide itself does not necessarily preclude a DNA-binding conformation or that if PrfA is oxidized while on the DNA, the protein does not automatically dissociate.

Without apparent density for glutathione in the effector binding domain, it is more probable that the PrfA-Phly24 does not have GSH bound and does not represent the
final active complex structure. It is possible that the protein, when oxidized on DNA, prevents conversion to the transcriptionally active complex.

The PrfA-Phly\textsubscript{24} structure does however contribute additional support to the partially bent DNA conformation induced by PrfA binding. Obtaining two structures with similar DNA bends using two different Phly lengths (30-mer vs 24-mer), in different environments (PrfA-Phly\textsubscript{30} crystallized at pH 4.5 compared to PrfA-Phly\textsubscript{24} which crystallized at pH 8.5), and in different space groups (P4\textsubscript{3}2\textsubscript{1}2 compared to C2), suggests that the obtained deviation from linear DNA and the Crp-bent DNA is not simply a crystallographic artifact from packing but likely a conformational change induced by PrfA binding to the DNA. This supports the likelihood that the structure of reduced PrfA on DNA is an intermediate conformation and the first of the two steps to Listerial gene activation.

\textbf{8.1 PrfA-Phly\textsubscript{24} Crystallization Conditions and Methods}

PrfA-Phly\textsubscript{24} was crystallized in drops containing 1 μL precipitating solution and 1 μL protein-DNA-GSH mixture at a protein concentration of 5 mg/mL in 48-well hanging drop-vapor diffusion plates at room temperature. The precipitating solution was a mixture of 1.5 M ammonium sulfate, 0.1 M Tris base pH 8.5, and 12% glycerol which acted as sufficient cryo-protection. The protein-DNA-GSH mixture was a 1:1:1 ratio of PrfA to a double stranded 24 oligomer (synthesized by IDT, Coralville, Iowa) encompassing the Phly palindrome where the top strand was 5′
AGGCATTAACATTTGTTAACGACG and 7.5 mM GSH. PrfA-Phly24 crystals assumed the space group C1 2 1 (C2) with unit cell parameters a = 283.9 Å, b = 102.7 Å, c = 147.4 Å and α = γ = 90°, β = 104.2°. There were two and a half PrfA dimers and DNA oligomers per asymmetric unit, which gives a solvent content of about 75%. X-ray intensity data were collected at the ID22 line at APS with a MarCCD detector and processed with HKL2000. The structure was determined via MR using protein protomers from the PrfA-Phly30 complex as the search model and Phenix (Adams et al., 2010) and refined to a final resolution of 3.65 Å. Selected crystal statistics are given in Table 8. All structural models were generated using Pymol (DeLano, 2002a, 2002b).
Figure 29: Structure of the PrfA-Phly24 Complex

A The overall structure of PrfA-Phly24 (left) including a rotated view looking down the DNA axis (right). B The overall alignment with PrfA-Phly30 (light green) shows the proteins adopt the same overall conformation, and the different length oligomers adopt the same overall bend.
Chapter 9. The Structure of the Quad-Phly\textsubscript{30} Complex is the Same as PrfA-Phly\textsubscript{30}

The structure of Quad PrfA in complex with the 30 bp oligomer in the PrfA-Phly\textsubscript{30} structure was also determined to 2.95 Å resolution to assess potential molecular differences from the DNA-bound wild-type structure.

Overall, the structures of the Quad- Phly\textsubscript{30} and PrfA- Phly\textsubscript{30} complexes are essentially identical. Quad- Phly\textsubscript{30} crystallized in the same space group (P4_2\textsubscript{1}2) with similar unit cell dimensions (a = 72.1 Å, b = 72.1 Å, c = 259.8 Å) and one dimer-DNA complex in the asymmetric unit. A ribbon diagram representation of refined Quad-Phly\textsubscript{30} (R\textsubscript{work}/R\textsubscript{free} = 31.6/34.3 %) is shown in Figure 30A. Alignment of the corresponding C\textalpha s of each protein Quad-Phly\textsubscript{30} to PrfA-Phly\textsubscript{30} (Figure 30B) returns an RMS deviation of 0.47 Å.

As with PrfA, there is some flexibility in the AR1 region of the C-terminal domain represented by a dearth of electron density in that area. The Phly DNA adopts the same overall bent conformation seen in PrfA-Phly\textsubscript{30} and PrfA-Phly\textsubscript{24}, although the 5' ends of the oligomer are less flexible in this structure which allows for the full DNA strand to be built. The main conclusion from the Quad-Phly\textsubscript{30} structure is that there are no significant differences between the protein-DNA complexes.

This conclusion was implied previously by the DNA-binding results (discussed in Chapter 5) and the thermal denaturation curves (discussed in Chapter 6.3), which indicated that the quadruple cysteine:alanine mutations did not have a deleterious effect.
on DNA binding and that the presence of Phly had roughly the same stabilizing effect when PrfA and Quad were subjected to thermal denaturation. The Quad-Phly30 structure confirms that the protein is capable of adopting the same DNA-binding conformation as WT.

Furthermore, the structure of the Quad-Phly30 complex continues to support the hypothesis that the DNA-binding conformation of PrfA is distinct from the transcriptionally active conformation. The ability of Quad to adopt the same structural conformation as the WT protein in the absence of GSH implies that GSH induces a transcriptionally active complex that has not yet been structurally elucidated and must exist as the functional difference between Quad and WT PrfA.

### 9.1 Quad-Phly30 Crystal Conditions and Methods

The Quad-Phly30 complex was crystallized in drops containing 3 μL precipitating solution and 5 μL protein-DNA mixture at a protein concentration of 5 mg/mL in 48-well hanging drop-vapor diffusion plates at room temperature. The precipitating agent was a solution of 25% PEG 400, 0.1 M sodium acetate pH 4.5, and 0.3 M calcium acetate, which was sufficient for cryo-protection without further supplementation. The protein-DNA mixture was a 1:1.1 ratio of PrfA to a double stranded 30-mer (synthesized by IDT, Coralville, Iowa) encompassing the Phly palindrome where the top strand was 5’ TTGAGGCATTAACATTGTTAACGACGATA. Quad-Phly30 crystals took the space group P4_2_2_2 with unit cell parameters a = b = 72.1 Å, c = 259.8 Å, and α = β = γ = 90°.
There was one PrfA dimer and DNA oligomer per ASU, which gives a solvent content $V_s = 43\%$. X-ray intensity data were collected at the ID22 line at APS on MarCCD detectors and processed with HKL2000. The structure was determined via MR with subunits from PrfA-Phly$_{30}$ as the search model using Phenix (Adams et al., 2010) and refined to a final resolution of 2.95 Å. Selected crystal statistics are given in Table 8. All structural models were generated using Pymol (DeLano, 2002a, 2002b).
Figure 30: Structure of Quad-Phly$_{30}$

A The structure of Quad PrfA in complex with a 30-mer encompassing Phly. B Alignment of the Quad-Phly$_{30}$ (violet, orange DNA) and PrfA-Phly$_{30}$ (lime green, grey DNA) complex structures reveal the overall conformations are very similar.
## Table 8: Select Crystallographic Statistics

<table>
<thead>
<tr>
<th></th>
<th>Data Collection Statistics</th>
<th>Refinement Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Apo-PrfA</td>
<td>PrfA-Phly$_{30}$</td>
</tr>
<tr>
<td><strong>Unit Cell</strong></td>
<td>a = 89.2 Å</td>
<td>a = 71.6 Å</td>
</tr>
<tr>
<td><strong>Dimensions</strong></td>
<td>b = 89.2 Å</td>
<td>b = 71.6 Å</td>
</tr>
<tr>
<td></td>
<td>c = 448.8 Å</td>
<td>c = 260.3 Å</td>
</tr>
<tr>
<td><strong>Unit Cell Angles</strong></td>
<td>(α β γ) 90° 90° 90°</td>
<td>(α β γ) 90° 90° 90°</td>
</tr>
<tr>
<td><strong>Space Group</strong></td>
<td>P4:2:2</td>
<td>P4:2:2</td>
</tr>
<tr>
<td><strong>X-Ray Source</strong></td>
<td>ID22</td>
<td>BM22</td>
</tr>
<tr>
<td><strong>Resolution (Å)</strong></td>
<td>2.31 (2.35 – 2.31)</td>
<td>2.90 (2.95 – 2.90)</td>
</tr>
<tr>
<td><strong>Number of reflections (#)</strong></td>
<td>80818</td>
<td>15880</td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
<td>99.9 (100)</td>
<td>99.4 (99.7)</td>
</tr>
<tr>
<td><strong>R$_{merge}$</strong></td>
<td>12.0</td>
<td>7.7</td>
</tr>
<tr>
<td><strong>I/σ</strong></td>
<td>37.7 (3.9)</td>
<td>18.6 (1.6)</td>
</tr>
<tr>
<td><strong>Redundancy</strong></td>
<td>10.6 (10.6)</td>
<td>4.4 (4.2)</td>
</tr>
<tr>
<td><strong>R$<em>{work}$/R$</em>{free}$ (%)</strong></td>
<td>20.6/25.4</td>
<td>31.6/32.7</td>
</tr>
<tr>
<td><strong>RMS Bonds (Å)</strong></td>
<td>0.010</td>
<td>0.006</td>
</tr>
<tr>
<td><strong>RMS Angles (°)</strong></td>
<td>1.304</td>
<td>0.938</td>
</tr>
<tr>
<td><strong>Ramachandran (%)</strong></td>
<td>96.7</td>
<td>87.1</td>
</tr>
</tbody>
</table>

---

*a* $R_{merge} = \Sigma |I - <I>| / \Sigma |<I>|$, where $I$ is the observed intensity and $<I>$ is the average intensity

*b* $R_{work} = \Sigma ||F_o - F_c|| / \Sigma |F_o|$, where $F_o$ and $F_c$ are the observed and calculated structure factors respectively

*c* $R_{free} = \Sigma |F_o - F_c|| / \Sigma |F_o|$ for 5% of the data not used at any stage of the structural refinement

* Numbers in parenthesis represent the value in the highest shell
Chapter 10. Conclusions and Future Directions

PrfA, a Crp/FNR family transcriptional regulator in *Listeria monocytogenes*, is responsible for the transcriptional activation of virulence genes that initiate the conversion of *Listeria* from benign environmental bacterium to intracellular pathogen (Scortti *et al.*, 2007). The inactivation of PrfA prevents the ability of *Listeria* to invade or spread between host cells (Gray *et al.*, 2006). While it was previously assumed that PrfA functioned similarly (Vega *et al.*, 1998) to the extensively studied canonical family member, Crp (Benoff *et al.*, 2002; Won *et al.*, 2009), there had been no identity for the proposed small molecule coactivator that would be needed to convert PrfA into a transcriptional activator.

Experimental results emerging from the Portnoy laboratory at UC Berkeley finally proposed that glutathione, the major low-molecular weight thiol of *Listeria monocytogenes* (Gopal *et al.*, 2005; Newton *et al.*, 1996), was required for PrfA-mediated virulence, and quadruple mutation of the cysteine residues to alanines (Quad PrfA) also exhibited a deficient but not avirulent phenotype (Reniere *et al.*, 2015). The emerging hypothesis was therefore that glutathione (GSH) was the small molecule allosteric activator for PrfA and that the cysteine residues somehow were required for efficient transcription that did not involve glutathionylation.

The experiments described in this dissertation have contributed further to the understanding of PrfA-mediated transcriptional activation in *Listeria monocytogenes* by
evaluating the role of the oxidation state of PrfA on DNA-binding and the ability of GSH to allosterically interact with PrfA. Additionally, we have elucidated the structure of reduced PrfA in complex with Phly in two different crystal forms as well as the structure of Quad PrfA in complex with DNA.

Using a fluorescence-polarization based DNA binding assay, we showed that the oxidation state of PrfA contributed to its ability to bind DNA (Chapter 5). Oxidized PrfA bound Phly very poorly in the near micromolar range, while reduced PrfA was capable of binding Phly and PactA with nanomolar affinity. The binding of reduced PrfA was comparable to the constitutively active mutant, G145S (PrfA*). Interestingly, Quad PrfA bound DNA with an affinity similar to reduced PrfA and PrfA* despite being 30-fold less virulent in mouse infection models. The addition of GSH had no effect on PrfA (oxidized or reduced), PrfA*, or Quad binding to DNA.

Despite its lack of effect on DNA binding, glutathione was determined to be capable of allosteric ligand binding (Chapter 6). Bio-layer interferometry conducted by the Portnoy laboratory (UC Berkeley) indicated GSH binding to PrfA was specific and that affinity was in the millimolar concentration range, which is physiologically relevant for *Listeria*. This interaction was supported by molecular docking with AutoDock Vina (Trott and Olson, 2010), which localized GSH near many residues that have been shown to be important contributors to PrfA function through prior mutational studies. Quad bound GSH with similar affinity to WT PrfA, indicating the cysteine residues were not
important for ligand interaction. Thermal denaturation, however, implicated that GSH did not interact equally between the PrfA-DNA and Quad-DNA complexes, suggesting a conformational difference exists between (PrfA-GSH)-DNA and (Quad-GSH)-DNA that might explain the decreased infection phenotype of *Listeria* expressing Quad PrfA.

The DNA- and GSH-binding studies challenge the canonical Crp model and indicates PrfA exhibits a closer mechanistic resemblance to NtcA (Muro-Pastor *et al.*, 2001; Tanigawa *et al.*, 2002). The biochemical assays suggested a two-step mechanism where, upon cytosolic entry, PrfA becomes reduced and capable of binding DNA and then increasing intracellular concentrations of GSH convert the protein to a transcriptionally active complex.

Structural determination of the PrfA-Phly\textsubscript{30} (Chapter 7) and PrfA-Phly\textsubscript{24} (Chapter 8) complexes supports the two-step mechanism hypothesis. The Phly DNA in both PrfA-Phly structures adopts a conformation that is more bent than linear DNA (around 35 - 40°) but is appreciably less bent than the activated Crp-DNA structure (~90°). Alignment with the Crp-DNA-\alpha\text{CTD} structure demonstrates that Phly in the less bent conformation would not allow \alpha\text{CTD} to be proximal enough to form protein-protein interactions with the AR1 region of PrfA, further indicating that PrfA-Phly\textsubscript{30} and PrfA-Phly\textsubscript{24} are an intermediate, DNA bound but transcriptionally less active conformation.

The PrfA-Phly\textsubscript{30} structure was validated by mutation of key DNA-binding residues responsible for base interactions with the PrfA box (Arg188 and Ser184) and
phosphate backbone contacts (Lys139, His182, and Ser187). Mutation of Arg188, Lys139, or His182 to alanine resulted in non-specific binding to Phly and PactA while the S184A substitution abolished specific binding to PactA and decreased the affinity for Phly four-fold. The S187A mutation did not appreciably change PrfA affinity for either Phly or PactA, and this residue may instead be responsible for coordinating the overall DNA bend.

Based on the biochemical and structural results described in this work, our proposed overall hypothesis for PrfA-mediated gene activation is shown in Figure 31.

In the environment, PrfA is expressed in low concentrations from the PrfAP1 and PrfAP2 promoters which are controlled by σA-bound RNAP (Freitag and Portnoy, 1994; Freitag et al., 1993), and PrfA is maintained as untranslated mRNA transcripts with the previously discussed thermoswitch obscuring the ribosome binding site (Johansson et al., 2002). The cysteine side chains of PrfA would be reduced in the environment, and the protein overall will adopt the intermediate, transcriptionally less active structure as shown by PrfA-Phly30 and PrfA-Phly24. This DNA binding conformation may actually explain the previous suggestion that PrfA can negatively regulate its own promoter, as RNAP may not be able to associate when the protein is present at this promoter (Greene and Freitag, 2003).

Once Listeria is transferred to a suitable host, the thermoswitch melts, allowing for rapid translation of PrfA and activation of InlB, LLO, PlcA and PlcB expression. The
alternative sigma factor, $\sigma^B$, also increases its intracellular concentration, leading to the expression of InlA and InlB and a small amount of LLO (Kim et al., 2005; Nadon et al., 2002; Ollinger et al., 2009). *Listeria* will be endocytosed to a host cell, exposing PrfA to an oxidizing environment in the vacuole and inactivating the protein. In this step of infection, PrfA will exist in the apo-structure as described in Chapter 7.1, preventing full activation of the PrfA autoregulatory loop.

Upon vacuolar escape, PrfA becomes exposed to the highly reducing environment of the host cytosol, allowing for it to bind DNA as it did in the environment. However, now the increasing concentrations of GSH from both *Listerial* synthesis (Gopal et al., 2005) and host import (Reniere et al., 2015) allow for the formation of a transcriptionally active complex and efficient recruitment of RNAP. We propose that this complex has more significantly bent DNA, and hence is the fully active transcriptional complex resembling Crp-DNA (Schultz et al., 1991) and CprK-DNA (Joyce et al., 2006; Pop et al., 2004).

Full activation of PrfA allows for the positive regulation of the transcriptional feedback loop, increasing PrfA concentrations to the point where gene expression increases from late stage promoters with poorer intrinsic DNA binding affinity such as *Phpt* and *PactA*. Actin polymerization occurs, spreading *Listeria* to neighboring cells and allows progression to an active infection.
The Quad PrfA mutant is flawed in the conformational shift from DNA-binding to the active complex. The Quad-Phly30 structure indicates there is no significant difference compared to PrfA-Phly30 which, coupled with the biochemical assays that indicate no deleterious change in DNA or GSH binding, implicates that infection deficiency must be a result of decreased transcriptional efficiency.

This two-step hypothesis demonstrates that unlike other Crp/FNR family members, PrfA does not solely exist in an on/off state. Instead, PrfA evolves between three conformations: the apo non-DNA binding conformation, the reduced DNA binding conformation, and the GSH-bound transcriptionally active conformation. PrfA is only the second Crp/FNR family regulator that can bind DNA in the absence of its proposed ligand (NtcA is also capable of this (Tanigawa et al., 2002)), and to our knowledge the first structure of an intermediately bent DNA conformation for the Crp/FNR transcriptional family.

10.1 Future Directions

There are several future experimental contributions that can solidify the details of the hypothesized two-step PrfA mechanism of gene activation. First, while we propose an active structure with a bent DNA conformation in the presence of GSH, this structure has not yet been conclusively determined. Future crystal screening to determine the (PrfA-GSH)-DNA complex can confirm or reject the proposed model. We
have obtained initial crystal hits with GSH saturated PrfA-DNA mixtures in search of this complex, and crystal optimization will likely yield improved results.

Biochemical assays may allow for the measurement of DNA bending in solution to provide additional support for a more bent structure of the (PrfA-GSH)-DNA complex. There has previously been success in measuring Crp-cAMP-DNA bending in solution using FRET (fluorescence resonance energy transfer) to quantitate changes in distance of the promoter DNA ends (Kapanidis et al., 2001). EMSAs can also be used to determine bend angle by measuring changes in DNA motility when the promoter is located at the middle or end of a set length of DNA. The pBend vector has been commercially created for this exact purpose and has been used successfully to calculate bend angles in the Crp-controlled lac, malT, gal, and crp promoters (Kim et al., 1989; Zwieb and Adhya, 1994).

Additionally, the PrfA-Phly30 structure did not appear to indicate why PactA would have a lower binding affinity than Phly as the base-specific residues do not interact with the changed nucleotide in PactA. Determining the structure of PrfA in complex with PactA to high resolution may provide supplementary insight into the interactions with PrfA boxes that deviate from the ideal palindromic sequence.

Experimental studies on larger complex formation will also contribute to the PrfA mechanism. Biochemically, the FP-based DNA assay may be modified to assess binding of the αCTD of RNAP to a longer oligomer of promoter DNA. Our preliminary
experiments designed a 44-mer fluorescein-labelled dsPhly based on the length of the Crp-DNA-αCTD structure, which PrfA binds with the same affinity as the 28-mer used in Chapter 5. Purification of *Bacillus subtilis* αCTD, which is nearly identical to the *Listerial αCTD*, may allow for binding studies in the presence of glutathione to assess if GSH mediates PrfA interactions with the αCTD. Once binding conditions with αCTD are honed in the FP assay, crystal screening for a (PrfA-GSH)-DNA-αCTD complex can be conducted to determine if the overall promoter assembly for the PrfA box is similar to Crp.

It is worth noting that crystal screening for a (PrfA-GSH)-DNA-αCTD complex is based on the known mechanistic details of Crp interactions with αCTD, and we have described deviations from the canonical mechanism in PrfA-mediated gene activation. It is therefore possible that PrfA interacts differently with *Listerial RNAP* than Crp with the *E. coli* RNAP. In this case, the studies described above with αCTD may need to be repeated with intact *Listerial RNAP* to isolate interacting subunits.

### 10.2 Summary

Overall, the results presented in this dissertation provide biochemical and structural contributions that take steps in elucidating the mechanism of PrfA-mediated gene activation in *Listeria monocytogenes*. While there are still questions that need to be answered, the results presented here support the two-step mechanism that is relatively unique compared to other characterized members of the Crp/FNR family of
transcriptional regulators. Additionally, the structures described in this thesis could provide insight into the design of novel antibiotics that target Listerial PrfA in its less transcriptionally active conformation thereby reducing or preventing infections.
The two-step hypothesis for PrfA gene activation. In the environment, reduced PrfA adopts a DNA-binding but transcriptionally inactive conformation represented by the PrfA-Phly30 (lime green) and PrfA-Phly24 (light blue) structures. Once engulfed by the host cell, PrfA is oxidized in the vacuole, adopting the “apo” conformation (dark blue), which binds cognate DNA poorly, before being released to the host cytosol. In the cytosol, PrfA binds GSH and bends its target DNA significantly, allowing increased transcription of virulence genes (transparent, proposed but not determined). The bent-DNA conformation was generated by aligning PrfA with DNA from the Crp-cAMP-DNA complex (PDB: 2CGP).
References


Gray, M.J., Gray, M.J., Boor, K.J., Boor, K.J., 2006. How the Bacterial Pathogen Listeria monocytogenes Mediates the Switch from Environmental Dr. Jekyll to Pathogenic Mr. Hyde. Society 74, 2505–2512. doi:10.1128/IAI74.5.2505


Biography

Keri Hamilton, nee Kimes, was born in Pittsburgh, Pennsylvania. She started undergraduate classes at Shippensburg University in 2006 where she graduated with a Bachelors of Science in Chemistry with a concentration in Biochemistry in May 2010. In the summer of 2010, Keri started an early rotation with Dr. John York, funded by the Duke University Kamin Foundation, then class at Duke University in Fall 2010. In 2011, she joined the lab of Dr. Richard Brennan to pursue and eventually obtain her doctoral degree in Biochemistry. Keri’s thesis work was funded in part by the Duke University Program for Structural Biology and Biophysics (SBB Training Grant 5T32GM008487-18).

Keri also obtained a certificate in College Teaching from Duke University and was awarded the Duke University Bass Fellowships for Instructor of Record for an advanced biochemistry class and Online Apprentice for the Introduction to Genetics and Evolution Massively Open Online Class pioneered by Dr. Mohamed Noor.

Publications
